Ferguson, Jamie (2020) Investigating perceptual congruence between information and sensory parameters in auditory and vibrotactile displays. PhD thesis.

Investigating Perceptual Congruence Between Information and Sensory Parameters in Auditory and Vibrotactile Displays

Jamie Ferguson

Submitted in fulfilment of the requirements for the Degree of Doctor of Philosophy

School of Computing Science
College of Science and Engineering
University of Glasgow

Abstract

A fundamental interaction between a computer and its user(s) is the transmission of information between the two and there are many situations where it is necessary for this interaction to occur non-visually, such as using sound or vibration. To design successful interactions in these modalities, it is necessary to understand how users perceive mappings between information and acoustic or vibration parameters, so that these parameters can be designed such that they are perceived as congruent. This thesis investigates several data-sound and data-vibration mappings by using psychophysical scaling to understand how users perceive the mappings. It also investigates the impact that using these methods during design has when they are integrated into an auditory or vibrotactile display.

To investigate acoustic parameters that may provide more perceptually congruent data-sound mappings, Experiments 1 and 2 explored several psychoacoustic parameters for use in a mapping. These studies found that applying amplitude modulation — or *roughness* — to a signal, or applying broadband noise to it resulted in performance which were similar to conducting the task visually.

Experiments 3 and 4 used scaling methods to map how a user perceived a change in an information parameter, for a given change in an acoustic or vibrotactile parameter. Experiment 3 showed that increases in acoustic parameters that are generally considered undesirable in music were perceived as congruent with information parameters with negative valence such as stress or danger. Experiment 4 found that data-vibration mappings were more generalised — a given increase in a vibrotactile parameter was almost always perceived as an increase in an information parameter — regardless of the valence of the information parameter.

Experiments 5 and 6 investigated the impact that using results from the scaling methods used in Experiments 3 and 4 had on users' performance when using an auditory or vibrotactile display. These experiments also explored the impact that the complexity of the context which the display was placed had on user performance. These studies found that using mappings based on scaling results did not significantly impact user's performance with a simple auditory display, but it did reduce response times in a more complex use-case.

Acknowledgements

Firstly, I would like to thank my supervisor, Steve Brewster, for giving me the opportunity to do this PhD and whose expertise in the field was so valuable. I would also like to thank my second supervisor, John Williamson, whose input and fresh perspective helped improve this research.

I would also like to thank my examiners Paul Vickers and Mohamed Khamis for taking the time to read and critique this thesis. Their comments have helped improve this work in many ways.

I would like to thank the incredibly friendly and supportive colleagues, both past and present, that I have had in the Multimodal Interaction Group and the School of Computing Science. In particular, my fellow students: Aris, Paddy, Gözel, Joseph, Alberto, Shaun, Lovisa, Julio and Andrii. Also, special thanks to Mark, Euan and Graham who mentored me throughout. I would like to thank Helen Purchase and Mary Ellen Foster for their invaluable feedback and encouragement during my annual exams.

I first discovered an interest in this field in my final undergraduate year at the University of Aberdeen, so I would like to thank George Coghill, Suk-Jun Kim and Wanda Diaz-Merced for their enthusiasm as supervisors.

I would like to thank my Mum and Dad for their support. Thanks also to my friends - thanks to Calum and Chris for being there since we lived in Aberdeen, and thanks to Harry and Caroline for always being there for a pint and for being willing participants in hours of pilot testing.

Additionally, I would like to thank: Kate Meredith, Bruce Walker, Olivia at KM Woods Physiotherapy, Laura Bell, everyone at Glasgow Rollerhoc, Carol, David, Penny and Nell.

Finally, I would like to thank Alice for her love, support and patience. It is not an understatement to say that she has been there for me every single day and I am incredibly lucky to have shared this experience with such a wonderful person.

This research was funded by the Engineering and Physical Sciences Research Council.

Declaration

The research presented in this thesis is entirely the author's own work. Research in this thesis has been published at the following venues, using only the parts of these papers that are directly attributable to the author:

- The research in Chapter 3 has been published at ICMI 2017: Ferguson, Jamie, and Stephen A. Brewster. "Evaluation of psychoacoustic sound parameters for sonification." Proceedings of the 19th ACM International Conference on Multimodal Interaction. 2017.

- The research in Chapter 4 has been published at CHI 2018: Ferguson, Jamie, and Stephen A. Brewster. "Investigating perceptual congruence between data and display dimensions in sonification." Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems. 2018.

- The research in Chapter 5 has been published at NordiCHI 2018: Ferguson, Jamie, John Williamson, and Stephen Brewster. "Evaluating mapping designs for conveying data through tactons." Proceedings of the 10th Nordic Conference on Human-Computer Interaction. 2018.

- The research in Chapter 6 has been published at ICAD 2019: Ferguson, Jamie, and Stephen Brewster. "Evaluating the magnitude estimation approach for designing sonification mapping topologies." Proceedings of the 2019 International Conference on Auditory Displays. 2019.

- A summary of the research conducted in this thesis was the basis for a doctoral consortium paper at BCS 2018: Ferguson, Jamie. "Evaluating mapping designs for conveying data using auditory and tactile displays." Proceedings of the 32nd International BCS Human Computer Interaction Conference 32. 2018.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

One of the primary utilities of a computing device is the transmission of information to the user. Information can be conveyed from a computer either through language or non-linguistically — the latter being advantageous in many situations because it is independent of the languages the user understands and affords the potential for multiple dimensions of information to be conveyed simultaneously. The majority of human-computer interfaces rely on the visual sense, however there are many situations where non-visual information transmission is preferable or necessary such as: in situations where the visual sense is occupied whilst driving or operating complex machinery, or the user having limited or no vision. One strategy for conveying information using such a non-visual representation is *parameter mapping*, where some information parameter is used to manipulate some sensory parameter, which in turn facilitates the communication of the information [65]. Two common methods for conveying information non-visually that can utilise parameter mapping are *sonification* — "the use of non-speech audio to convey information" [87] and *tactons* — "structured, abstract, [vibro] tactile messages which can be used to communicate information" [16], which are the cutaneous equivalent of *earcons* [9] — themselves a form of sonification. In their chapter on parameter mapping in *The Sonfication Handbook* [58], Grond & Berger posit that "effective parameter mapping sonification often involves some compromise between intuitive, pleasant and precise display characteristics". However, there is little theory or evidence to guide designers toward establishing the most effective parameter to convey a particular data value — for either the auditory or vibrotactile modalities.

This lack of theory is a problem that designers face when creating a parameter mapping based sonification or tacton system — finding an acoustic or tactile parameter to represent a particular value of data in such a way that the user perceives the mapping between the parameters as perceptually congruent. Here, perceptually congruent meaning that the relationship between the acoustic or tactile parameter and the information parameter which it is representing is perceived as congruent by the user. For example, artificial and natural alarm sounds like a fire alarm or a person screaming are typically rough or disharmonious sounding [5], therefore it is reasonable to assume that users may expect a sonification relating to some value of danger to sound rougher as the value increases. Similarly, many non-linguistic representations of information are designed by using an analogy from a similar domain or context. For example, the a Geiger counter, which is an instrument for measuring ionising radiation, outputs increasing numbers of audible clicks as the detector gets closer to a source of radiation which is a similar interaction to that found in modern vehicle parking sensors in which the closer the vehicle is to some object, the number of tones played increases. Often in these types of non-visual displays, such expectations based on previous experiences or analogy from other contexts are not considered and even more rarely are they investigated, resulting in interaction designs that may be in conflict with how a user expects data to sound, feel or look like (depending on the modality in use) [155,157].

In sonification systems, pitch is the most commonly used acoustic parameter to represent information [36] and has even been described as the “Hello World" of sonification [64]. This is to be expected, as pitch is one of the most salient characteristics of music and sonification designers commonly employ musical structures and aesthetics [65]. Regardless of the positive attributes of pitch as an acoustic parameter, there are many situations where multiple data parameters must be conveyed at once and pitch alone cannot fulfil this need. The range of acoustic parameters available to sonification designers is limited, as only a small subset of parameters have been empirically evaluated in a sonification context. The situation is similar for vibrotactile displays, with the majority of vibrotactile feedback given from mobile phones and wearable devices, as well as video game controllers, consisting of very simple vibrations, commonly using only a binary on/off cue. However, increasing numbers of recent devices contain more complex actuators such as: Apple’s *Taptic Engine* [166] and the linear resonant actuators (LRA) found in video game console controllers such as the Nintendo Switch Joy Con [89] and the the DualSense controller for the Playstation 5 [131]. These allow designers to control a number of vibrotactile parameters such as frequency, amplitude and rhythm. The reliance on using simple parameters for feedback is limiting for designers thus making it difficult to reach a balance between Grond & Berger’s “intuitive, pleasant and precise" characteristics [58].

Psychoacoustic parameters such as *fluctuation strength*, *roughness*, *sharpness* and *loudness* [39] can be used as methods of measuring the subjective qualities of sound and are responsible for a substantial part of the effect a sound has on a listener — combined, they can model the relative degree of noise annoyance [34], and roughness independently has been found to be a primary component in both natural and artificial alarm sounds [5]. Therefore, these parameters may be more effective than pitch at conveying certain types of data, as they are closely linked with certain experiences (such as roughness and alarm, or noise and annoyance/unpleasantness). Similarly, exploring more expressive vibrotactile feedback using parameters such as frequency and amplitude can allow more information to be encoded within the vibration feedback itself.

This can reduce the need to repeatedly look at a device for information transmission and also improve the perceptual congruence of a data-to-vibration mapping.

Exploring acoustic and vibrotactile parameters beyond the commonly used ones is a potential solution to the problem of perceptual congruence in auditory and vibrotactile displays (i.e. what the listener perceives is complimentary with how they expect a data value to sound or feel when it is represented through sound or vibration). This congruence is key to a successful mapping, because if a user is confused by, annoyed by or unsure of a mapping, this can lead to a acoustic or vibrotactile display that is difficult to use. Such a difficult-to-use display could result in misinterpreting the information being conveyed and slower responses to the information — which in a safety critical context where these displays can be found such as surgery [8, 42, 94], air-traffic control [120] or driving [129] would have serious and potentially fatal consequences.

Therefore, the research conducted in this thesis investigated how to create more perceptually congruent data-sound and data-vibration mappings. Specifically, it focused on the use of psychophysical methods to understand how mappings were perceived by participants and then, based on these results, these mappings were implemented into systems and evaluated to understand the impact of the use of psychophysical measurements in the design of an auditory or vibrotactile display. Overall, the goal of this thesis was to provide results which give designers or researchers some basis for the choices they make during the design process when creating a data-sound or data-vibration mapping, maximising the likelihood of creating a mapping which users perceive as perceptually congruent and therefore reduce the risk of confusion, misinterpreted information and slower responses.

1.1 Thesis Statement

> Which sensory parameter to use to convey a particular information parameter in an auditory or vibrotactile display is a significant design problem, as the perceived meaning and connotations of auditory and haptic stimuli are complex. This thesis presents experiments which used psychophysical scaling to establish how several data-sound and data-vibration mappings were perceived and subsequently investigated the impact that using results from these experiments had on users' performance when they were implemented into the design of a parameter mapping display. This thesis presents a number of design recommendations based on results from these studies which designers and researchers can use to guide their decisions on which acoustic or vibrotactile parameter is appropriate for a particular information parameter.

1.2 Research Questions

This thesis answers the following questions:

RQ1: How may acoustic parameters that are derived from the field of psychoacoustics encode and transmit information to a user?

RQ2: To what extent are mappings between acoustic sonification parameters and information parameters with clear positive or negative valence perceived as congruent?

RQ3: To what extent are mappings between vibrotactile parameters and information parameters perceived as congruent?

RQ4: How does the complexity of the usage context impact how effective a data-sound or data-vibration parameter mapping is?

1.3 Thesis Structure

This section briefly summarises the contents of each chapter. Table 1.1 shows the purpose of each experiment conducted in the thesis and how each experiment relates to the Research Questions.

Chapter 2 — *Literature Review*

This chapter presents a review of literature and related work on the topics subsequently covered in this thesis. The review begins by discussing the auditory modality, covering: auditory displays, psychoacoustics and perceptual factors involved in representing information using sound. The review then moves on to discussing the tactile modality: beginning with a discussion of the physiology of the sense of touch, the dimensions that can be manipulated to covey information through touch and related work. Finally, to better understand the role of mental models and analogy-making in the perception of auditory and vibrotactile display, the review covers some relevant theories such as semiotics and analogy.

Chapter 3 — *Psychoacoustic Parameters for Sonification*

This chapter investigates how several acoustic parameters that are derived from psychoacoustic research may be implemented into an auditory display, addressing **RQ1**. This chapter presents two experiments which explore how these parameters can be manipulated to convey information, how they compare with each other and investigates some of their advantages and disadvantages. Later chapters build upon these results by further exploring these parameters.

Chapter 4 — *Perceptual Congruence in Data-Sound Mappings*

This chapter investigates how several acoustic parameters that were derived from the preceding chapter perceptually relate to information parameters to understand what pairings of information parameters and acoustic parameters is perceived as congruent, addressing **RQ2**. This chapter presents an experiment which uses magnitude estimation to map how much participants perceive a change in an information parameter such as danger, for a given change in an acoustic parameter that represents it.

Chapter 5 — *Perceptual Congruence in Data-Vibration Mappings*

This chapter investigates how several vibrotactile parameters are perceptually related to several information parameters, addressing **RQ3**. This chapter presents an experiment which uses the same magnitude estimation paradigm as the preceding chapter to explore how to pair vibrotactile parameters and information parameters such that they are perceived as congruent by users.

Chapter 6 — *Evaluating the Magnitude Estimation Approach*

This chapter explores the question of whether data-sound or data-vibration mappings that are designed based on prior magnitude estimation experiment results impact a user's performance when using the mappings, addressing **RQ4**. This chapter presents an experiment which uses the simple use case of ranking WiFi networks based on how secure they are, conveyed using auditory or tactile cues. These cues are designed either based on results from Chapters 4 and 5, or are designed in an arbitrary manner, to explore the question of how much this design process impacts performance.

Chapter 7 — *Parameter Mappings Under Cognitive Load*

This chapter further explores the main question of Chapter 6 by building on the experiment presented in that study, investigating how the design of a data-sound or data-vibration mappings impacts how well users can use them, again addressing **RQ4**. This chapter presents an experiment where participants were tasked with attending to auditory or vibrotactile alarms which, again, were either designed based on prior results or arbitrarily. Alongside attending to these alarms, participants also conducted a cognitively demanding memory task, to explore how cognitive load impacts user performance with data-sound or data-vibration mappings.

Chapter 8 — *Conclusions*

This chapter summarises the research contained in the thesis and how it answered the Research Questions set out at the beginning. This chapter also summarises the novel contributions that are made by the thesis, some of the limitations of the research and how this could be addressed in future work.

1.4 Overview of Experiments

Experiment	RQs	Purpose
Experiment 1 Experiment 2	RQ1	Evaluate psychoacoustic parameters for sonification
Experiment 3	RQ2	Investigate perceptual congruence in data-sound mappings
Experiment 4	RQ3	Investigate perceptual congruence in data-vibration mappings
Experiment 5	RQ4	— Investigate mappings in a simple use-case scenario — Evaluate the impact of using a magnitude estimation approach
Experiment 6	RQ4	— Investigate the impact of a complex use-case scenario — Further investigate the impact of using a magnitude estimation approach

Table 1.1: Summary of experiments presented in this thesis and Research Question the studies contributed to answering.

Chapter 2

Literature Review

2.1 Introduction

The opening chapter introduced the fundamental problem that this thesis addresses: that there is a lack of research around how to convey information through sound or vibration in a way that both the information dimension and the sensory dimension are perceived as congruent by the user. This chapter reviews existing literature surrounding this problem, which includes work relating to the auditory and tactile modalities and the transmission of information using these modalities.

The review begins with an overview of the auditory modality which covers: techniques for encoding information in sound; notable relevant works through the history of this area of research; and finally, psychoacoustics and the human perception of sound. These sections explain how information can be conveyed using sound with a focus on how parameter mapping can be used to facilitate this information transmission. As will be discussed, the overview of existing works in this particular area of auditory display shows that mapping information to acoustic parameters in a way that a listener perceives as congruent it is a significant problem in the design of parameter mapping sonification systems. To understand these problems more thoroughly, an understanding of how people perceive sound is necessary, this therefore is the purpose of the subsequent sections on psychoacoustics. These sections provide literature that is particularly relevant to the following research questions:

RQ1: How may acoustic parameters that are derived from the field of psychoacoustics encode and transmit information to a user?

RQ2: To what extent are mappings between acoustic sonification parameters and information parameters with clear positive or negative valence perceived as congruent?

RQ4: How does the complexity of the usage context impact how effective a data-sound or data-vibration parameter mapping is?

Following this introduction to the auditory modality, literature relevant to the tactile modality will be discussed. This will include a review of literature relating to: the fundamentals of the human sense of touch and its perceptual limitations; the dimensions of this sense that can be manipulated; and how information can be encoded in and conveyed using these manipulations of vibrations. This literature shows how the sense of touch is an effective method of information transmission, however this area is less researched than auditory displays and it faces similar problems such as designing mappings in which the data and vibration dimensions are perceived as congruent by a user. These sections provide literature that is particularly relevant to the following research questions:

RQ4: How does the complexity of the usage context impact how effective a data-sound or data-vibration parameter mapping is?

Finally, this review looks at some of the theoretical background work relating to conveying and receiving information. These sections include: a brief overview of semiotics — the study of signs and symbols — and how it relates to auditory and tactile displays; discussion of analogy and how making analogies between media and domains may be a key part of how mappings between information and sensory parameters are processed. This section aims to use semiotics and other related theory to understand the fundamental processes involved in one actor transmitting information to another and how that information is processed by the receiving actor.

2.2 The Auditory Modality

An auditory display can be defined generally as any display that uses sound to communicate information [158], more specifically, *Sonification* is the use of non-speech audio to convey information [87]. One of the earliest and most discussed examples of a successful implementation of sonification is the Geiger counter, which was designed by Hans Geiger in 1908 [122]. The device works by producing an audible click through a loudspeaker every time a particle of ionising radiation is detected. The use of auditory feedback allows the user to navigate their surroundings with their other senses and to not rely on monitoring a gauge visually. The auditory feedback of the Geiger counter was purely functional and no research was conducted on the counter's effectiveness as a display tool. However, early research into auditory display in itself began with Pollack & Ficks' [117] early exploratory investigations of auditory displays in 1954. This work demonstrated that multiple dimensions of information could be effectively transmitted to a listener through an auditory display. Another early example of sonification research was published by Speeth in 1961, in which he used sonification to classify different types of seismic events [132]. A more comprehensive investigation of auditory display was conducted by Bly in 1982 [10, 11], which showed that a multitude of data types (multivariate, logarithmic and time-varying) could be effectively conveyed through sound.

Since sonification has been recognised as a field of research in its own right, there have been a number of attempts to establish guidelines for effective sonification design. Early in the development of the field, Kramer [86] discussed some general principles for designing auditory displays. In a more extensive work, Barrass [6] proposed a number of "golden rules" for sonification design. Based on these two works, Anderson [4] identified some of the crucial elements that future research in this area must address — the mapping of data to parameters being one of these. This is summarised in the following quote from Anderson:

> Sonification designers need empirical evidence to guide them in what strategies increase intelligibility when there are multiple sound dimensions changing. Research is needed on the nature and causes of interactions between dimensions and needs to include a wide range of dimensions. [4] (p. 395)

These works provide a thorough view of what is required to provide a framework for sonification designers, however they were all theoretical and did not attempt to provide any experimental evidence on which to base sonification design choices. This gap in the literature is part of the motivation for Research Question 1, which aims to evaluate several novel parameters for use in sonification and therefore widen the range of acoustic dimensions which have been studied in this context:

> **RQ1:** How may acoustic parameters that are derived from the field of psychoacoustics encode and transmit information to a user?

2.2.1 Subtypes of Auditory Displays

There are a number of subtypes of auditory displays including *Audification* — the direct translation of a data waveform into sound [154] and *Model-Based Sonification* — a system mediating between data and sound by means of a dynamic model [65]. Within sonification there are a number of different techniques including: *Auditory Icons*, which are sounds used to convey information about computer events by analogy with everyday sound-producing events [50] and the similar *earcons* which are abstract, synthetic tones that can be used in structured combinations to create auditory messages [15].

2.2.2 Parameter Mapping Sonification

The particular technique for sonification which this thesis is concerned with is *Parameter Mapping Sonification* (PMSon). In a parameter-mapping sonification system, data values are used to manipulate an acoustic parameter, such as frequency (pitch) or tempo, which facilitate the communication of the data value [65]. Simply put, this means that a data parameter is used to control some auditory parameter. Consider a number of readings from a thermometer which ranges from 0 to 100°C. A parameter mapping sonification for this data set could be a mapping of temperature to the pitch of a tone, such that as the temperature increases, the pitch of the tone also increases. Dubus & Bresin [36] conducted a systematic review of 179 sonification publications and found that parameter mapping was the most commonly used method of sonification. Grond & Berger's chapter in *The Sonification Handbook* [58] provides the most thorough overview of PMSon and defines one of the biggest challenges of this technique for sonifying data: "establishing perceptually valid mappings between data and signal synthesis and/or processing parameters" [58] (p. 392)

Regarding which acoustic parameters are found most often in PMSon systems, pitch is unequivocally the most used. According to Dubus & Bresin's systematic review of sonification papers, pitch was found to be used significantly more often than all other acoustic parameters. In their review, pitch was found in 23.8% of studies and the next most frequent parameter was loudness, found in 15.2% of studies [36]. A pitch-based parameter mapping has even been dubbed the "Hello World" of sonification [64]. This is to be expected, as pitch is one of the most salient characteristics of a musical sound and sonification designers commonly employ musical structures and aesthetics into the designs of sonification systems [65]. Only a small subset of acoustic parameters have been empirically evaluated in a sonification context, meaning that the parameter space in which designers can work in is limited when the wealth of potential auditory phenomena that could potentially be used as an acoustic parameter in a PMSon system is taken into account. Other possible parameters have been posited as potential candidates for use in sonification such as psychoacoustic parameters [41] but these have not been implemented into a PMSon system and evaluated experimentally

In Kramer's original principles for sonification design [86], a number of perceptual factors were established that may be practically implemented in a sonification system. Among these factors were *affective associations* and *metaphorical associations*. These notions began to explore how the importance of human perception in sonification design, especially affective associations, which were described as "the association of feelings about data (if such feelings exist) with feelings aroused by changes in the sound" [153] (p. 88). An example of such an association was given in the context of an ecologist: to such a researcher, data indicating an increase in rain acidity would generally be described as undesirable, and therefore may cause a subtle negative affect. Therefore, a sonification mapping that may utilise this effect could be described as: an increase in "auditory ugliness" = an increase in an undesirable information parameter. These affective associations were not empirically tested however, they were precursors to further research by Kramer and others which would put perception and more specifically perception of parameter mappings at the forefront of their approach to auditory display design.

The first work to directly investigate data-sound mappings experimentally was Walker & Kramer's study in which a number of acoustic parameters were used to represent a number of data dimensions in a simple process monitoring task (originally presented in 1996 [153], published in 2005 [155]). Mappings were grouped based on how effective sound designers perceived them to be *a priori*. Results showed that mappings which were predicted to perform well (such as temperature-pitch) did not always result in increased performance over mappings that were predicted to perform badly, or even mappings in which the data and acoustic parameters were paired randomly. Walker & Kramer posited the explanation that a listener's perceptual models of data-sound relationships had a large part to play in their surprising results. For example, slow acoustic changes like slow tempos and onsets were perceived by participants to be representative of larger objects — as would be the case with larger objects in the physical world — larger objects are often slower due to inertia. Therefore, participants may have called on their experiences of the physical world to form their mental model of a totally non-physical relationship in a data-sound mapping. As the authors say, this not only further underlines the need for iterative prototyping and verification of interface designs, but also demonstrated that listeners brought expectations about how certain types of information "ought to" sound.

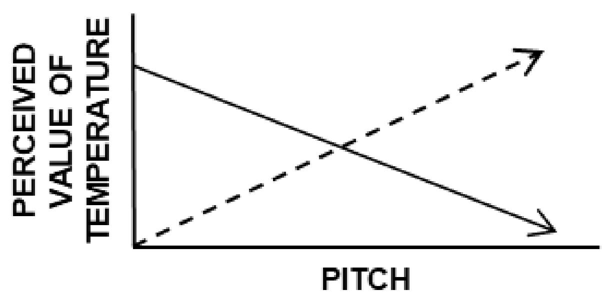

Figure 2.1: Example magnitude estimation results for a mapping between pitch and temperature. Dotted line shows a *positive* polarity, solid line shows a *negative* polarity. The angle of the *slope* of these lines shows the scale of the relationship

To attempt to find an empirical method of investigating parameter mappings Walker [151, 152] proposed *magnitude estimation* as a tool for empirically testing data-sound mappings. Magnitude estimation is a standard method of psychophysical scaling, which maps the relationship between the magnitude of a sensory stimulus and its associated perceived intensity [137], resulting in a power function between the actual stimulus magnitude and the perceived magnitude. Today, magnitude estimation is a frequently used tool in studying the perception of auditory stimuli in a wide array of contexts including ergonomics [74] and medicine [82]. Walker's results [151, 152] showed that magnitude estimation provides reliable measures for both the scaling function (what increase in temperature is perceived for a 100Hz increase in pitch?) and polarity (is increasing pitch perceived as increasing temperature?) for a data-sound mapping. Figure 2.1 shows a graphical representation of scale and polarity. Walker's first studies used a number of data dimensions such as pressure, velocity, size, dollars, urgency and attractiveness and a number of simple acoustic parameters: pitch, tempo and modulation index. As this was the first proper exploration of how data-sound parameter mappings are perceived it was advantageous that the data dimensions used in the study were relatively non-specific — that is to say, things like temperature and pressure can represent a number of different types of temperature or pressure. Furthermore, the acoustic parameters used in the study were simple in that there was no emphasis put into pairing data dimensions and parameters in a way that would be perceived as congruent by the listener. This gap in the literature regarding perceptual congruence between data dimensions and acoustic parameters is the motivation for research question 2:

> **RQ2:** To what extent are mappings between acoustic sonification parameters and information parameters with clear positive or negative valence perceived as congruent?

Polarity Preference of a Parameter Mapping

The magnitude estimation technique was used by Walker to establish polarity preferences among participants, which can be thought of as the number of participants that perceive a data-sound relationship in a particular direction. If a given polarity obtained a majority of responses, it was deemed to be a "good" mapping. Walker suggested that differences in polarity preference for a given mapping between participants may suggest that they use differing mental models — "a representation of some domain or situation that supports understanding, reasoning, and prediction" [52] (p. 9683) — to guide their preferences. Previous research has shown that mental models used to inform polarity preference for a data sound mapping can differ significantly between sighted and visually impaired listeners [157]. Additionally, Edworthy *et al.* [37] found that the polarity of a data-sound mapping was an important factor in perception of auditory alarms for helicopter warnings — a situation where the ability to clearly interpret auditory information is safety-critical. They found that increases in a stimulus value had more effect on listeners perception of a mapping than a decrease in the stimulus value. A quote from Edworthy *et al.* quite succinctly summarises the importance of understanding parameter mappings in general, but specifically the polarity (or direction as it is called in the paper) of the mapping:

> It could be argued that there is little benefit in seeking to delineate precisely the relationship between changes in an acoustic parameter and changes in meaning without establishing first that a particular meaning can be communicated effectively by a particular acoustic parameter. In addition, it may be of more practical importance for an operator to be able to accurately identify which monitored parameter is changing and in what direction than to minutely quantify those changes. [37] (p. 217)

In this work Edworthy *et al*. [37] discuss polarity preference in terms of *denotative*, *connotative* and *perceptual* meaning. A denotative interpretation of rising pitch would be that it could be the actual sound made by an object as some value increases (e.g. an engine working harder) — i.e. the rising pitch literally *denotes* that an engine is working harder. A connotative (or analogical) interpretation would be that the rising pitch indicating there is increasing work by or strain on the engine, based on other situations with high pitch sounds, like a child crying. This can be thought of as additional meanings or concepts being used to interpret the sound. Finally, they describe a perceptual interpretation as the rising pitch meaning some value that is getting higher. This can be thought of the rising pitch *signifying* the value getting higher. The complexity of possible interpretations of sonification mappings is exemplary of the importance of investigating a mapping during its design.

Individual differences in how sounds are perceived are a persistent challenge to designing polarities of data-sound mappings [106]. For example, in a visual display, "up" represents "more" in most cases and generally more of some visual parameter means more of whatever value it is representing. However, in the auditory modality it is almost always not as clear, as can be seen from some of the work mentioned so far like Walker and Kramer's findings [155] where *decreasing* pitch was perceived to represent an *increase* in size. Similarly, the context in which a mapping is situated can impact how it is perceived. Consider a simple metaphor often used in data-sound and data-vibration mapping designs — the Geiger counter metaphor — where the frequency of repetition of a sound or cue is used to convey some information. This can be perceived differently in different contexts in that the information being represented may be positive or negative in terms of valence. For example, in a physical Geiger counter, the more "clicks" heard, the closer the user is to a source of ionising radiation which could be considered to have *negative* valence, due to the fact that radiation can be dangerous. Conversely, this metaphor is also used successfully in contexts like gesture recognition [45] and navigation [73], where increasing "clicks" (and vibration pulses in the case of [45]) are indicative of the user completing an action correctly, thus can be considered a situation with *positive* valence.

2.2.3 Psychoacoustics and Perception of Sound

In order to establish and evaluate acoustic parameters to be used in an auditory display, a designer or researcher needs to understand some of the factors relating to how sound is perceived. If a cue or system is not designed with perceptual factors in mind, then there are many potential problems that may occur, such as: the sound being inaudible, sounds not being distinguishable from one another and sounds being uncomfortable, irritating or even painful. Therefore, an understanding of psychoacoustics can act as a guide during design. Psychoacoustics is the branch of psychophysics that deals with the perception of sound and aims to model the complexity of auditory sensation in physical parameters [174]. The study of psychoacoustics essentially affords an understanding of how a sound is comprised and how these components are perceived, allowing designers to potentially manipulate these base components to evoke or incite a specific affect, or more pertinent to this thesis, evoke certain sensations or "musical universals" like harmony and disharmony [20]. The history of psychoacoustic research spans the length of scientific inquiry itself. Historic polymaths such as Galileo Galilei and Robert Hooke experimented with auditory perception and Gustav Fechner and Hermann von Helmholtz — who many consider to be the founders of modern experimental psychology — spent much of their careers investigating the psychology of hearing. As the history of the field is so vast, this review will focus primarily on contemporary research. See Yost's account [168] for a concise historical overview of psychoacoustics research.

Zwicker and Fastl's work [174] collects theoretical and experimental research that presents algorithms which physically quantify auditory sensations such as loudness, roughness and sharpness. Application of modelled psychoacoustic sensations is common in industrial acoustics design to optimise the acoustic qualities of products and appliances [96, 124, 125]. Psychoacoustic models applied in this context are often defined as *sound quality metrics*. Sound quality being defined as "a perceptual reaction to the sound of a product that reflects the listener's reaction to how acceptable the sound of that product is" [95]. Psychoacoustic parameters are an important factor in product design, as the aesthetic and cognitive effects of a product's sound emission may be important to a user's experience with a product. Although these parameters are a common consideration in product design, they have been rarely employed overall in sonification research, aside from a limited number of works which will be discussed later in this section. The following sections will discuss several acoustic parameters and their psychoacoustic properties that influence how they are heard and processed.

Pure Tones and Noise

Before discussing particular psychoacoustic sensations, a brief introduction to the types of sound that are used in the definition of psychoacoustic parameters is necessary. In Zwicker and Fastl's standard text *Psychoacoustics: Facts and Models* [174], they primarily discuss psychoacoustics in terms of pure tones, narrow band noise or broadband noise. A pure tone is simply a single sinusoidal wave of any frequency, phase or amplitude. Furthermore the tone must remain unmodulated throughout the duration it is presented — unmodulated meaning that no amplitude or frequency modulation has been applied to the signal [126]. This can be thought of as the most basic tone possible to create. A pure tone contains one frequency, whereas broadband noise is a stochastically produced signal which contains equal intensity of a range of frequencies [55]. Some examples of broadband noise are *white* noise where all frequencies in the human hearing spectrum are at equal intensity and *pink* noise where each octave rather than frequency is of equal intensity and therefore has more energy at lower frequencies than higher ones.

Pitch

There are a number of definitions of pitch, however the most commonly accepted is: "that attribute of auditory sensation in terms of which sounds may be ordered on a musical scale" [116] (p. 1). It is the mechanism that we use to perceive melody in music and several sounds with different pitches together allow us to perceive harmony in the form of chords [126]. Fundamentally, pitch is a sensation in which a listener assigns musical tones based on a scale, in which this scale is based on the vibration frequency of the sound being heard [116] — pitch is the subjective measurement, whereas frequency is the physical measurement. As will be discussed in the section on loudness, how loud a sound is perceived to be is dependent on the pitch or frequency of that sound, with lower frequencies being perceived as quieter than higher frequencies. Furthermore, the intensity of a sound also impacts the perceived frequency of a sound. Therefore intensity affects perceived pitch and frequency affects perceived loudness. Note here the difference between intensity-loudness and frequency-pitch, with the former in both cases being the objective measure and the latter being the psychoacoustic sensation that is perceived by a listener.

Roughness

Roughness quantifies the subjective perception of rapid amplitude or frequency modulation (Figures 2.2 and 2.3) of a sound (modulation frequencies from around 15 Hz to 300 Hz, reaching its maximum near 70 Hz) [174]. The same rapid amplitude modulation that elicits the psychoacoustic sensation of roughness is also used to explain the acoustic component of dissonance [79]. The unit of roughness is the *asper* and one asper is defined as a 60 dB(SPL), 1 kHz tone that is 100% amplitude modulated at a frequency of 70 Hz [174].

Fluctuation Strength

Fluctuation strength is similar to roughness, however it quantifies the subjective perception of slower amplitude or frequency modulation of a sound (modulation frequencies of up to 20 Hz, reaching its maximum near 4 Hz) [174]. The border between the sensations of roughness and fluctuation strength is fuzzy. The unit for fluctuation strength is the *vacil* and one vacil is defined as a 60 dB(SPL), 1 kHz tone 100% amplitude modulated at 4 Hz [174].

Loudness

The loudness of a sound is the perceived effect of sound pressure on the ear [174]. Note that loudness is a subjective measure and it is related to the physical measure of sound pressure (commonly seen as a decibel measurement). Loudness depends not only on the energy of a sound, but on frequency. For example a low frequency tone would be perceived as quieter than a tone of 1 kHz even if they have the same physical energy. Figure 2.4 shows how frequency

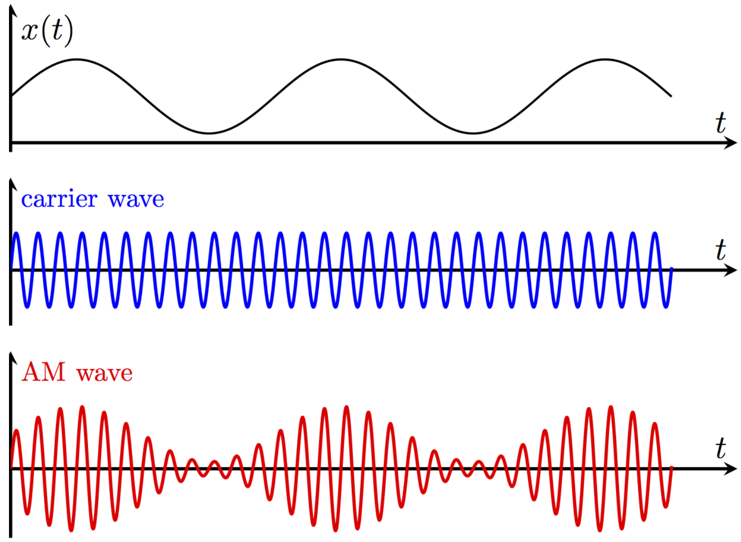

Figure 2.2: Visual representation of amplitude modulation where the modulating wave (black) is modulating amplitude of the carrier wave (blue), resulting in an amplitude modulated (AM) wave (red) [142].

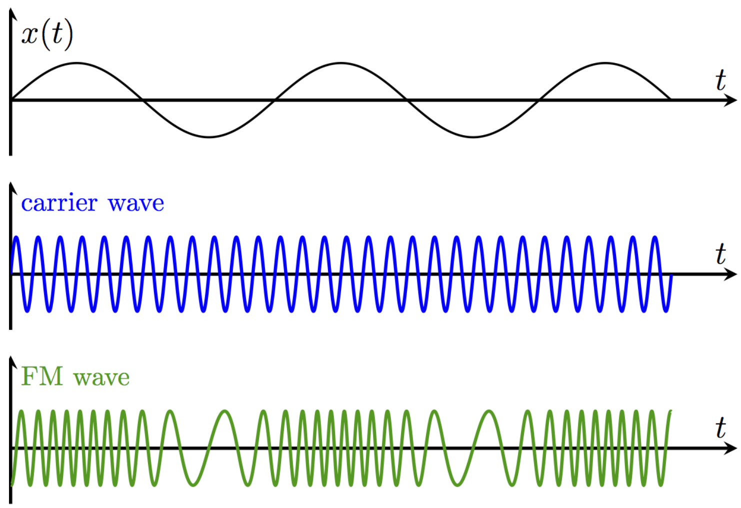

Figure 2.3: Visual representation of frequency modulation where the modulating wave (black) is modulating the frequency of the carrier wave (blue), resulting in a frequency modulated (FM) wave (green) [143].

relates to perceived loudness. However loudness is often more complex, as ‘critical bands’ of the cochlear affect how loud a sound is perceived to be. For a given frequency, a critical band is the smallest range of frequencies around it that activate the same area of the basilar membrane within the cochlear — the primary sensory apparatus of the ear [126]. The cochlear could be thought of as a crude frequency analyser as when a sound of a single frequency is presented to the cochlear, the basilar membrane maximally vibrates in a particular location (Figure 2.5). If two sounds that are less than a critical bandwidth apart they will partly mask each other, making the calculation of loudness more difficult. There are 24 critical bands, this scale being the Bark scale [173].

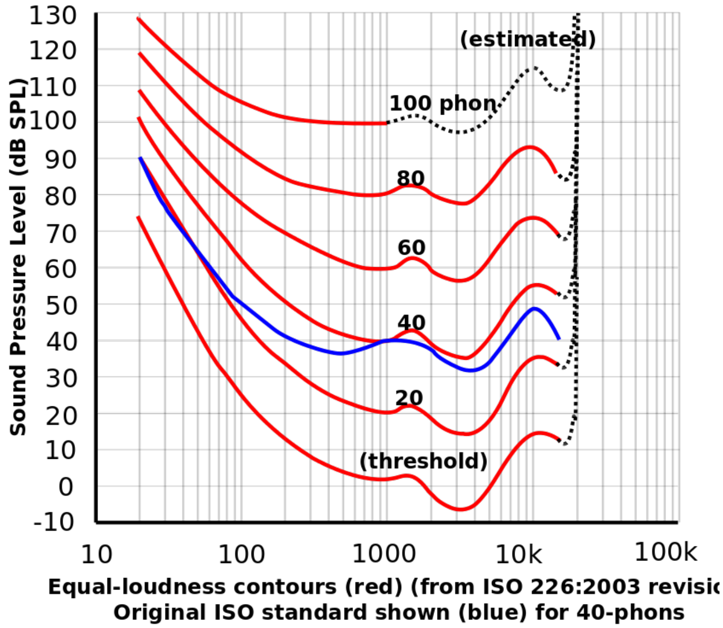

Figure 2.4: Representation of equal-loudness curves for various frequencies [93].

Sharpness and Booming

Sharpness is determined the high-frequency content of a sound — a larger proportion of higher frequencies equates to a sharper sound [174]. Compared to other psychoacoustic sensations, sharpness is less standardised and there are multiple methods of defining it, however the commonly used definition is again by Fastl & Zwicker who defined the unit of sharpness — the *acum*. One acum is a narrow band noise, one critical band (or Bark) wide at a centre frequency of 1 kHz having a level of 60 dB(SPL). Booming is similar to sharpness, however, as the name implies it quantifies the low frequency content of a sound as opposed to the high frequency content and thus could be considered the opposite of sharpness [63].

2.2.4 Psychoacoustic Parameters in Auditory Displays

Neuhoff and Heller [107] proposed a method of auditory display that uses data-sound mappings that utilise a "pre-existing cognitive structure in which to interpret the sonified changes in a variable" [107] (p. 418). This method of auditory display was motivated by the fact that when a listener is evaluating some fundamental acoustic dimensions such as pitch, they can encounter interference from other such dimensions like loudness, as the two are perceptually integrated [56]. Peres and Lane [113] conducted experiments which utilised these integrated acoustic dimensions in an auditory display. Perhaps the most fundamental work in terms of

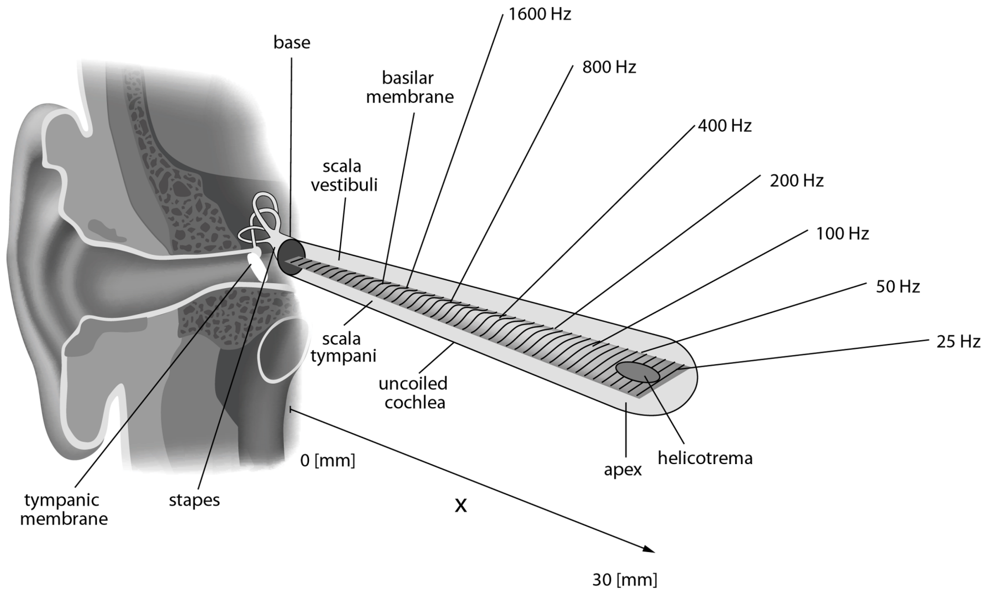

Figure 2.5: Visual representation of the unrolled basilar membrane, showing the approximate best frequencies of various locations from base to apex [83]

psychoacoustics use in auditory display however is by Ferguson *et al*. [41] which proposed the use of psychoacoustic models for sonifcation and discussed the potential benefits and laid out a research agenda for future research in this area, however no experimental assessment was conducted.

Ziemer and a number of other researchers at the Bremen Spatial Cognition Center have conducted a number of studies which have attempted to directly use psychoacoustic parameters to represent information in an auditory display [127, 170, 171]. For example, a study by Ziemer & Schultheis used roughness and fluctuation strength (which is called *beating* in their paper) for orientation and navigation in three dimensional space [171] (Figure 2.6) and applied to navigating surgical tools toward an insertion point, such as a drill toward a burr hole for a craniotomy [170]. In another study, Schwarz & Ziemer used roughness and noise to indicate whether a patient's oxygen saturation was within a healthy range or not [127]. The primary motivation for their work was that auditory displays can be very beneficial when the visual sense is committed to another task — such as conducting surgery on a patient [170] or tending to a neonate [127] — and more specifically that using psychoacoustics to motivate their sound design may make their auditory display as easily understood as possible in such high cognitive load situations.

The work from the Bremen Spatial Cognition Center showed that psychoacoustic parameters can be used in an auditory display effectively, however the scope of this work did not include emphasis on investigating how the psychoacoustic parameters used could be perceptu-

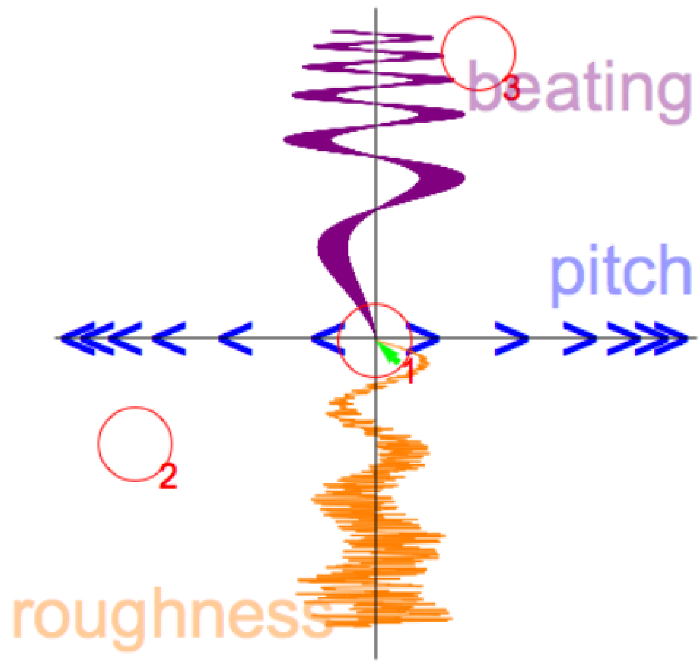

Figure 2.6: Visual representation of Ziemer & Schultheis' psychoacoustic auditory display for navigation [171].

ally congruent with the data dimensions that they were conveying. The motivation for using these parameters in this work was that they could be perceived as one *auditory stream* — that is, they are perceived by the listener as one, combined group, as opposed to multiple parameters [14, 105], meaning that the cognitive load that the auditory display incurs may be reduced. The work showed that psychoacoustic parameters could be used in an auditory display, however the question of how they may be used to represent data dimensions which share some of their connotations has not been explored. For example, the concept of *rough* would generally be considered to be undesirable (the roughness of a sound has been shown to be an integral component of how humans perceive danger through cries or screams [5]). Therefore using roughness to encode some value of data relating to some data dimension which may be considered undesirable such as danger, error or stress may be more aligned with how a listener would expect a value of this kind to sound when it is sonified. Psychoacoustic parameters such as roughness and sharpness are multimodal descriptions of texture or quality, as they apply in the visual and haptic modalities as well as the auditory. Therefore, the cross-modal nature of these parameters, in addition to other perceptual factors such as the importance of roughness to the perception of dissonance [79] may offer novel data-sound mappings that are more aligned with perceptions of the data parameter being sonified. This thesis aims to address this under-researched area by investigating how psychoacoustic parameters can encode and transmit information to a user and how congruent these mappings are perceived to be.

2.2.5 Summary

An auditory display is a system which uses sound to communicate information [158] and *sonification* is the use of non-speech audio to convey information [87]. *Parameter Mapping Sonification* (PMSon) is a technique for displaying information where data values are used to manipulate an acoustic parameter. A significant problem in PMSon design is finding an acoustic parameter to represent a particular dimension of data in such a way that the user perceives the mapping between the parameters as perceptually congruent and, therefore, alignment between how the listener interprets a sonification and how the designer of a sonification intended it to be interpreted. A number of data-sound relationships have been investigated in terms of how they are perceived by listeners [151, 152], however these were fundamental studies were conducted at a time when the field of auditory display was relatively nascent and therefore the data-sound mappings were not designed with any emphasis on pairing data dimensions with acoustic parameters in a way that is perceived as congruent by a listener, as one would expect from a fundamental study.

Furthermore, the acoustic parameter space that has been researched is limited, with pitch being the most investigated parameter, found in 23.8% of studies included in Dubus & Bresin's systematic review of sonification papers. The field of psychoacoustics aims to put the sensations of hearing into physical parameters and can therefore be a useful tool in investigating acoustic parameters in data-sound mappings. There are many parameters or sensations that have been derived, such as roughness, loudness and sharpness. These parameters are commonly used to optimise the acoustic qualities of products and appliances. These parameters have been used in auditory displays to aid movement in three dimensional space [170, 171] and to convey the oxygen saturation of a patient [127]. These works showed that psychoacoustic parameters could be used to convey information, but the parameters were paired with data dimensions in a fairly arbitrary way — meaning that that emphasis was not put on pairing the psychoacoustics parameters with data dimensions in a way that is perceptually congruent.

This thesis addresses these gaps in the literature by: (1) investigating a number of acoustic parameters for use in sonification beyond common simple parameters by utilising psychoacoustics; (2) investigating whether data-sound mappings which are designed with the goal of intentionally being perceived as congruent by a listener are actually perceived as such; and (3) investigating the usability of these mappings when placed in a cognitively demanding situation. This leads to the following research questions:

> **RQ1:** How may acoustic parameters that are derived from the field of psychoacoustics encode and transmit information to a user?

RQ2: To what extent are mappings between acoustic sonification parameters and information parameters with clear positive or negative valence perceived as congruent?

RQ4: How does the complexity of the usage context impact how effective a data-sound or data-vibration parameter mapping is?

2.3 The Tactile Modality

As the introduction discussed, this thesis focuses not only on the auditory modality, but the tactile or haptic modality as well. The sense of touch can be a powerful method for transmitting information, yet it is relatively under-researched when compared with other modalities. Therefore, a goal of this thesis is to answer some questions about how information is perceived through the sense of touch with and to therefore learn more about how this sense can be utilised to communicate information. This section introduces the tactile modality and related work on how it can be used in information transmission.

Tha haptic sense is integral to human communication — it is one of the earliest developed senses [43] and is a powerful tool for information transmission, both between people and between a user and a system. Haptics is defined as “sensory and/or motor activity based in the skin, muscles, joints and tendons” [77] (p. 2). Haptic sensation can broadly be split into two categories: the cutaneous sense which is perception through/from the skin [80] and proprioception which is perceived through/from the muscles and joints [139]. There are two primary types of skin which perceive haptic information differently: *hairy* skin which covers the entire body aside from the sole of the foot and the palm of the hand which have hairless, *glabrous* skin [80].

Within the tactile sense there are four sub-senses: touch, temperature, pain and itch. Additionally, there is a growing body of research that proposes a fifth modality conveying the sense of pleasure [99]. It is also worth noting the slight distinction in terminology between cutaneous, haptic and tactile. Cutaneous refers to anything affecting the skin and haptic/tactile refers to anything relating to the sense of touch. There are four types of receptors that are responsible for communicating information to the brain about events on the skin’s surface which are collectively called low-threshold mechanoreceptors. These are: Meissner’s corpuscles, Pacinian corpuscles, Merkel’s disks (also seen as Merkel’s cells) and Ruffini endings [99]. Meissner’s and Pacinian corpuscles are classified as fast adapting receptors (FA) as they respond to the initial contact and the final contact of a stimulus on the skin. Merkel’s disks and Ruffini endings are classified as slowly adapting (SA) as they respond continuously while a stimulus is present on the skin [99]. Figure 2.7 shows a diagram of these receptor fibres. These receptors are further classified based on their *receptive field* (RF), that is the area of the skin in which they are responsive. There are two classifications of RF: type I which is a diameter of 2-8mm or type II which is a diam-

Receptor Type	Pacinian	RA Meissner's	SAII Ruffini	SAI Merkel's
Frequency Response	40 — 800 Hz	10 — 100 Hz	15 — 400 Hz	0.4 — 100 Hz
Sensation	Vibration	Flutter	Not Known	Pressure
Temporal Summation	Yes	No	Yes	No
Spatial Summation	Yes	No	Not Known	No
Depth	Deep Tissue	Superficial	Deep Tissue	Superficial

Table 2.1: Summary of Bolanowski's findings [12]. Table adapted from [69].

eter of 10-100mm [80]. Further to this, Bolanowski [12] carried out a number of experiments which resulted in the proposal that there are four distinct psychophysical channels that control tactile perception in glabrous skin and that each channel is represented by one of the anatomical receptors just mentioned and nerve fiber subtypes (FAI, FAII, SAI, SAII) [12, 99]. Bolanowski found the maximum sensitivity for glabrous skin to be between 200 and 300 Hz [12] and Verrillo found the maximum sensitivity for hairy skin to be around 220 Hz [144]. Table 2.1 shows these receptors and some of their properties.

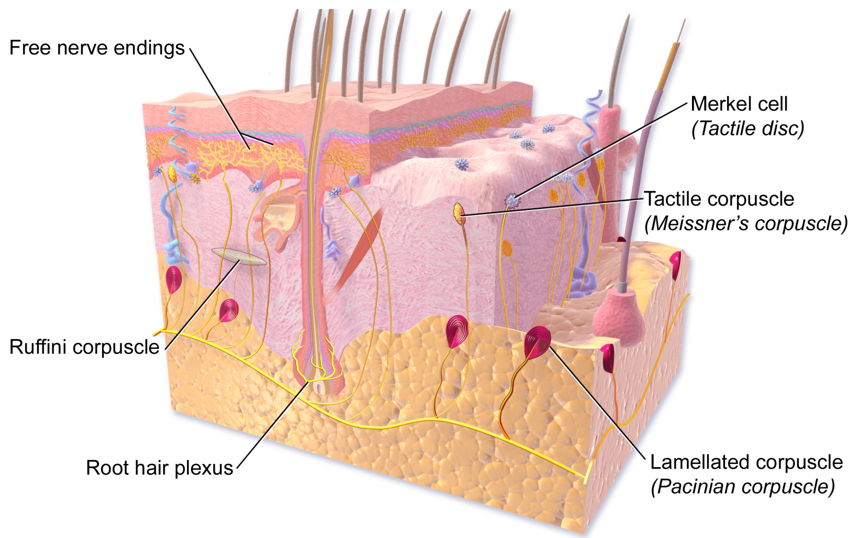

Figure 2.7: Diagram of the receptor fibres responsible for sensing touch [133].

2.3.1 Dimensions of Vibration

There are a number of dimensions which can be perceived through the sense of touch and therefore, similarly to the auditory and visual modalities, manipulations of these dimensions can encode information. This section will briefly introduce a number of dimensions of vibrotactile feedback. This section focuses particularly on vibrotactile dimensions as the rest of the thesis only uses the vibrotactile sensation and not other aspects of touch like pressure or temperature.

Frequency

The range of frequencies that are perceptible to humans and the ranges of maximum sensitivity were discussed in the previous sections, but like audible frequencies, the perception of frequency is affected by the intensity of the stimulus [145]. Figure 2.8 shows the contours of equal sensation for vibrations at different frequencies that were derived from studies by Verillo [145]. Additionally, in these contours the areas of maximum sensation as found by Bolanowski [12] and Verillo [144] can be clearly seen between 200 and 300 Hz. The minimum difference in frequency between two stimuli such that those two stimuli are perceived to be different — the just noticeable difference (JND) — is around 18% [118].

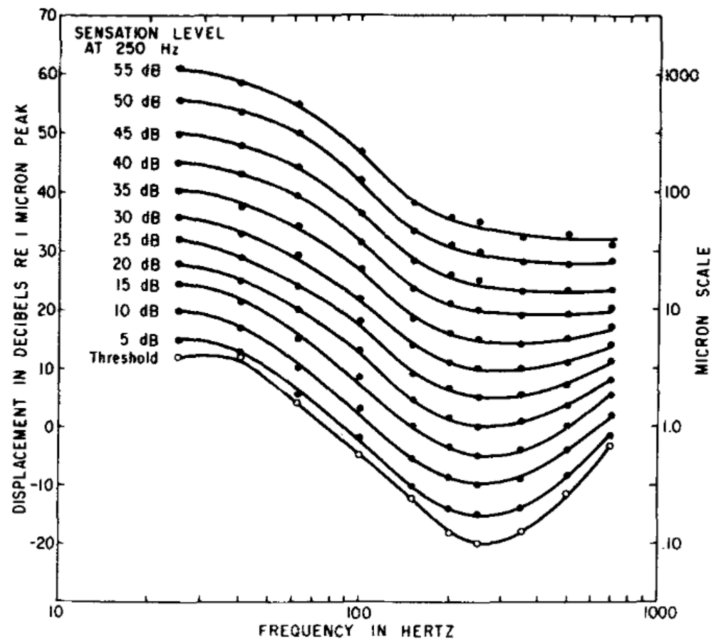

Figure 2.8: Verillo's contours of equal sensation of vibration [146].

Duration

Another dimension which can encode information is the temporal variation or duration of a stimulus. The majority of research on the duration of vibrotactile stimuli has used 0.1 ms as the lower limit for stimulus duration [146]. Geldard [51] conducted a study to attempt to find the just noticeable difference for vibration duration where they presented 60 Hz vibrations to the ventral region of the thorax (the chest) and found that the discrimination threshold increased from 0.05 seconds to 0.15 seconds as duration increased from 1 to 2 seconds and that there are 25 distinguishable levels in this range. When discussing duration, another important factor is how the duration *between* two stimuli is perceived. Gescheider [53, 54] measured the just noticeable difference between two vibrotactile stimuli such that they are perceived as distinct parts (i.e. not perceived as one, combined stimulus) and found that the threshold for gap detection was at around 10 ms. Furthermore, Gescheider found that the gap detection threshold improves as a function of the intensity of the stimuli and the duration of the stimuli, with gap detection improving as these dimensions increase.

Intensity

As was mentioned, the perceived intensity of a vibration stimulus depends on the frequency of that vibration [149], therefore using these two parameters to convey information simultaneously could be problematic. The just noticeable difference in intensity of a vibration stimulus that can be reliably discriminated does depend on the amplitude of the vibration, but it does not follow Weber's law [81] — that is to say that the JND is *not* directly related to the initial stimulus magnitude — the JND is smaller at moderate-high intensities [31].

Temporal Patterns

Combining various vibrotactile pulses in patterns that change over time can create rhythms akin to those found in music and therefore information can be encoded in these rhythms — an analogue from the auditory modality of such information encoding is Morse code. In order to encode information in temporal patterns, it is necessary to understand the number of pulses that can be perceived in a given duration. Lechelt [92] found that beyond 5 pulses within a 700 ms duration, the skin begins to struggle to perceive the pulses, meaning that the cutaneous modality is approximately between the visual and auditory modalities in this respect.

Location of Stimulation (Locus)

Since the location of a stimulus on the body can generally be easily distinguished — for example, distinguishing the difference between being tapped on the left or right shoulder — information can be encoded in the locus of where stimuli are presented. Different parts of the body have differing levels of sensitivity, with two-point limens (the smallest distance between two points of pressure that is still distinguishable as two distinct points) which, generally speaking, show that the fingertip is most sensitive, followed by the hands and face and the rest of the body having lower sensitivity to varying degrees [159]. Cholewiak & Collins found that, in general, it is easier to locate a point of stimulation on the body when it is located near an anatomical reference point or “anchor-point" such as the elbow or wrist [29]. They conducted a second study which applied stimulation to the abdominal region which similarly found that localisation was improved at the anatomical reference points of the navel and spine [28].

Waveform

The shape of the vibration wave, such as sine, saw or square (Figure 2.9) is often used in the auditory modality, especially music, to elicit different timbres, however research on this aspect of vibrotactile stimulation it is underdeveloped compared to the other parameters mentioned here. Based on early mentions of waveform in relation to tactile information, some research has been carried out looking at the difference in perception between vibrations comprised of pure tone sine waves versus broadband noise [88, 140] which found that sinusoids were perceived as smooth and noise as rough. A more recent study by Hoggan [70] also attempted to investigate the correlation between roughness and waveform, and found that sine, sawtooth and square waves were perceived as three distinct sensations. As well as using simple waveforms (e.g. sine or square), information can be encoded using complex waveforms. For example, Weisenberger [160] found that when participants were presented with an amplitude modulated sine tone and an unmodulated tone, they perceived the modulated tone as being rougher than the unmodulated tone.

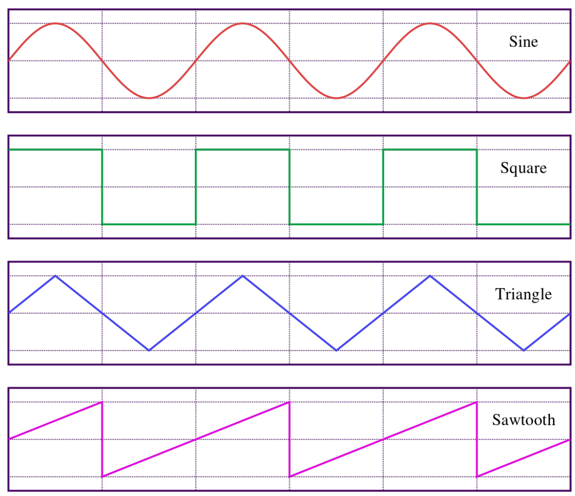

Figure 2.9: Visual representations of sine, square, triangle and sawtooth waveforms [109]

2.3.2 Vibration for Interaction and Information Transmission

Vibrotactile displays have been utilised in communication and human-computer interaction contexts for a number of years. In Geldard's 1957 paper *Adventures in Tactile Literacy* [51], they presented a vibrotactile language *Vibratese* which three individuals learned in only 12 hours and could interpret a maximum of 38 words per minute. Long before computing technology was generally available, especially systems with the hardware to provide vibrotactile stimulation, *Vibratese* showed vibration's information transmission potential. Vibrotactile displays are now ubiquitous in computing and communication devices, becoming a useful interaction modality for a number of contexts including: navigation [141], in-car interaction [108, 123], conveying emotion in mobile communications [162, 167] and accessibility [172].

In addition to these general examples, vibrotactile stimulation is the most common form of haptic interaction in consumer computing devices, with tactile feedback being present in the majority of currently available mobile phones and wearable devices, as well as video game and virtual reality controllers. The majority of vibrotactile feedback given from these devices consists of very simple vibrations produced by eccentric rotating mass (ERM) motors that afford the designer little control of the feedback parameters. However, increasing numbers of recent devices contain more complex actuators such as the linear resonant actuators (LRA) found in video game console and virtual reality controllers such as the Playstation *DualSense* [131], the Nintendo Switch Joy Con [89] and Apple's *Taptic Engine* [166]. Apple have gone as far as providing so-called "Core Haptics" guidelines to help developers integrate vibrotactile feedback into their designs.[1] These actuators allow designers to control a number of vibrotactile parameters such as frequency, amplitude and rhythm. Most vibrotactile feedback given from current devices alerts the user to a notification, but then forces them to visually attend to the device directly to acquire the information they are being notified about. More expressive vibrotactile feedback can allow more information to be encoded within the vibration feedback itself, reducing the need to repeatedly visually attend to the device for information transmission.

Tactons

In many cases, vibrotactile feedback is simply used to attract attention (e.g. to notify users to a text message or incoming call), or to give interaction feedback (such as confirming a key press on a touchscreen device). However, vibration dimensions can be structured into messages or *tactons* (tactile icons) which can afford more complex information to be encoded in vibration. tactons are the cutaneous equivalent of the aforementioned audio encoding strategy *earcons*. The advantage of tactons is that they not only notify the user about information but also convey it. In some contexts this may be all that is required, or it may provide some information and the user may choose to engage another modality to access more. Brewster & Brown [17] first described tactons and outlined a design space comprising seven properties that could be used for encoding information: frequency, amplitude, waveform, duration, rhythmic patterns, body location, and spatiotemporal patterns. Brown & Brewster then introduced roughness — amplitude modulation — as an additional property [18]. Hoggan and Brewster further investigated modulating other parameters such as frequency and waveform to convey roughness through vibrotactile stimulation. tactons have been explored unidimensionally [17, 18], multidimensionally (as in modulating more than one vibrotactile parameter simultaneously to convey more complex information) [18] and in crossmodal contexts in combination with audio feedback [70, 71]. Brown & Brewster's initial study that evaluated two-dimensional tactons comprising three levels each of roughness (amplitude modulated) and temporal patterns showed an overall identification rate of 71%. In a later study, Brown & Brewster. evaluated three-dimensional tactons comprising three

[1] https://developer.apple.com/documentation/corehaptics

Figure 2.10: The C2 Actuator [38].

Figure 2.11: The Haptuator Mk II [136].

levels of roughness, rhythm and spatial location (three contacts spread across the length of the forearm). In this study, overall identification was lower than the previous study (48%) however this increased to 81% when the roughness parameter was reduced to only 2 levels. These results show that even with the cutaneous sense's low resolution when compared with the visual and auditory modalities, complex information can still be transmitted and received.

2.3.3 Hardware

The design of tactons is affected by the type of actuator that is used to present the stimuli. The previous studies on the fundamentals of tactons discussed [18, 19, 70] used acoustic signals passed through an EAI C2 voice-coil actuator (Figure 2.10).[2] However, more recent studies such as [163, 167] use a Tactile Labs Haptuator Mk II (Figure 2.11) which has improved acceleration and bandwidth over the C2.[3]

2.3.4 Analogy and Affect in Vibrotactile Stimulation

The previous sections on parameter mapping in sonification show that the choice of acoustic parameter used to convey a certain dimension of information can have a significant impact on how that information is perceived and that many users employ some form of mental model in their interpretation of data-sound mappings. The situation with vibrotactile information transmission, or data-vibration mappings, seems to be similar. There are examples from vibrotactile research that suggests that users invoke analogy and mental models in their interpretations. For example, rougher patterns as well as stronger vibrations were perceived as more alarming [128]; analogously, a strong, rough vibration in the real-world may indicate an alarming event such as an earthquake. Furthermore, Hameed *et al.* [60] mapped a high importance cue designed to divert attention to a rapidly on/off fluctuating vibration which "equated to a rapid tapping of

[2]https://www.eaiinfo.com/product/c2/
[3]http://tactilelabs.com/products/haptics/haptuator-mark-ii-v2/

the skin" (p. 378). This is similar to tapping someone rapidly on the shoulder to gain their attention. Although these works indicate that there are associations made between vibrotactile stimuli and various concepts and meanings such as roughness/intensity = alarming [128], "tapping" = interruption [60] or increasing intensity = unpleasantness [162, 167], there has not been any attempt to evaluate data-vibration parameter mappings. These examples are indicative that the receiver of vibrotacitle stimuli apply mental models in their processing of a data-vibration mapping, similar to the findings from data-sound mappings. However, research on the perception of data-vibration mappings is far less advanced that its auditory counterpart.

2.3.5 Summary

Information can be encoded and conveyed using vibrotactile feedback, by manipulating a number of dimensions such as frequency, duration and intensity using tactons, which are structured vibrotactile messages that transmit information solely through the tactile modality. These are particularly useful in situations where visual or auditory displays are unavailable or inappropriate. Work by Brown & Brewster. [18, 19] and Hoggan [70] has shown that this technique is an effective method of information transmission. The relations between vibrations and their perceived meaning are complex and, in many cases, context dependent. Therefore, it is vital to consider this mapping when designing successful tactons. Similarly to the aforementioned literature from auditory displays, receivers of vibrotactile stimuli that represents information (a data-vibration mapping) may apply analogy and mental models in their processing of the mapping, for example intense/rough vibrations being associated with alarm [128]. However, research on how data-vibration parameters are perceived is significantly less investigated that their equivalent auditory mappings and the literature that has been discussed here does not put emphasis on the perceptual congruence of the data-vibration mapping.

This thesis addresses these gaps in the literature by: (1) investigating a number of mappings between vibrotactile parameters and data dimensions to establish how congruent they are perceived as being; and (2) investigating the usability of these mappings when placed in a cognitively demanding situation. Furthermore, this leads to the following research question:

RQ3: To what extent are mappings between vibrotactile parameters and information parameters perceived as congruent?

RQ4: How does the complexity of the usage context impact how effective a data-sound or data-vibration parameter mapping is?

2.4 How is a Parameter Mapping Conveyed and Received?

As the previous sections on the auditory and tactile modalities have discussed, the mechanisms that are involved in the reception, perception and processing of information are complex and at times hard to predict, meaning that it can be difficult for a designer to ensure that the meaning they intended to convey is actually received. Looking to theories such as semiotics, cognition and other areas where information is already effectively conveyed non-visually may help understanding how to approach the problems that this thesis addresses.

The studies by Kramer [153], Walker [151, 152] and Edworthy *et al.* [37] underline how the complexities and intricacies of human perception mean that information which is transmitted via parameter mapping may not necessarily be received by the listener in the way in which it was intended. A clear example of the kind of disparity can be found in Walker & Lane's 2001 study [157], in which they found that participants who were visually impaired perceived a mapping of dollars to pitch to possess a *negative polarity* (meaning they perceived lower pitch tones to represent more dollars), whereas participants with no visual impairment perceived the mapping oppositely (with a *positive* polarity). The authors discuss that some visually impaired participants noted that they used the analogy of money dropping on a table to form their polarity preference; for example, a single coin makes a high-pitch *clink* sound, whereas a large bag of notes makes a dull *thud* sound.

This disparity between groups exemplifies the subjectivity of mental models and analogies a listener may employ to understand a mapping and how this may not correlate with how designers may expect a mapping to be perceived *a priori*, that is before it has been tested in a manner like the above study. Additionally, the surprising disparity between mappings that sound designers believed to be effective and mappings which actually resulted in better performance in Walker & Kramer's 1996 study [153, 155] further support the notion that mappings designed solely from designer preference are problematic. The situation in the vibrotactile domain is a significant distance behind in terms of this knowledge, with research on this area of data-vibration mappings being near non-existent. However, the works discussed in the section 2.3.4 touch on this, but the results only go so far as to indicate that analogy and mental model seem to be involved in how these mappings are processed.

The Semiotic Approach

Semiotics is the study of signs and symbols and their use or interpretation [27]. In semiotic terms, a sign can be a number of things: words, images, sounds and objects to name a few. If a data-sound or data-vibration mapping's most basic intention is considered, it is that it intends to *signify* information, therefore a number of researchers have argued for a *semiotic* approach to designing auditory displays [104, 114, 115]. However, before semiotics was introduced to auditory display research, Kramer [153] designed a continuum for how data can be represented using sound which ranges from *symbolic* to *analogic* (Figure 2.12). This continuum wasn't strictly rooted in the language or theory of semiotics, however it is important to note that many of the ideas that would later be introduced by researchers hoping to reconcile auditory displays and semiotics are present in Kramer's initial discussion of a and analogic/symbolic continuum for auditory displays.

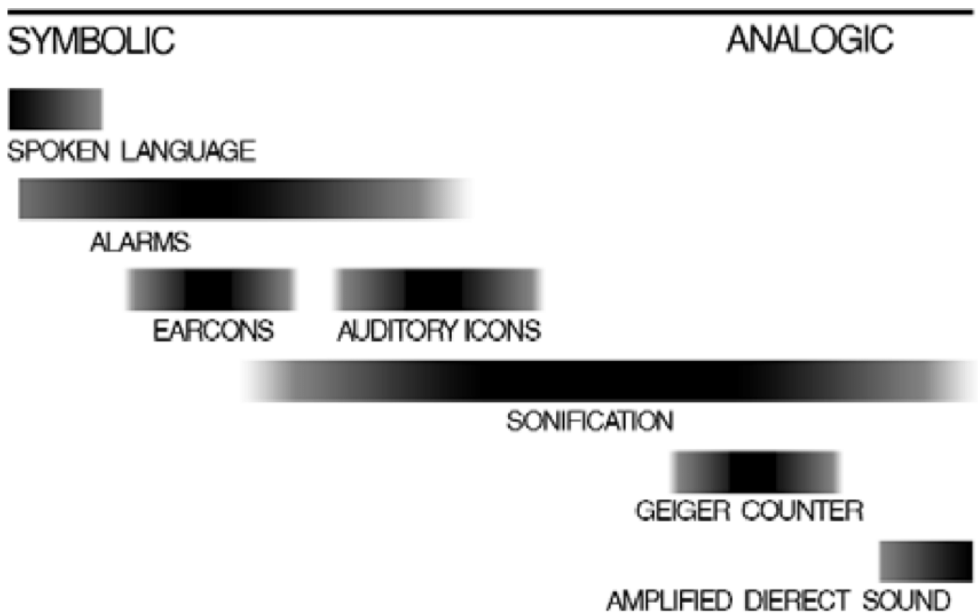

Figure 2.12: Kramer's analogic/symbolic continuum [86].

Moving forward, Pirhonen *et al*. [115] posited an approach based in semiotic theory, in response to the problem discussed in the previous section — the dichotomy between how designers intend an auditory display to be received by a user and how it is actually received. They claim that auditory display design methods are commonly "derived from random selection or the personal preferences of the designer" or "*ad hoc* empirical settings" [115] (p. 134). Pirhonen *et al*.'s approach focused on using use-cases with experts and/or users who are representative of the domain in which a system is being designed, thus gathering their input on how information should be signified in a particular context.

A Brief Overview of Semiotic Theory

To better put into context how semiotics can be a useful way to understand mappings, this section will give a brief overview of semiotic theory, introducing some key terms. This overview is primarily based on Chandler's textbook *Semiotics: The Basics* [27] and articles by Vickers [147], Oswald [110] and Jeon [78]. Semiotic theory is generally split into two traditions: The

Saussurean Model, developed by the Swiss linguist Ferdinand de Saussure and the Peircian Model developed by the American logician Charles Sanders Peirce [27].

The Saussurean Model

Saussure's model of a sign is *dyadic* in that it is comprised of a 'Signifier' (*Signifié*) and the 'Signified' (*Signifiant*) (Figure 2.13).

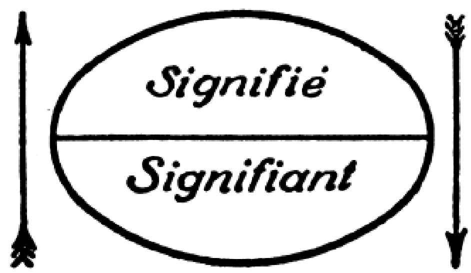

Figure 2.13: Ferdinand de Saussure's model of the sign [33].

A linguistic example from [27] is the word 'open' when it is encountered on some shop front's door is a *sign* consisting of two parts:

- *A Signifier*: the word 'open'.
- *A Signified Concept*: that the shop is open for business

The Saussurean tradition was primarily developed in the system of language, either the spoken or written word, therefore almost all research on semiotics has been based in the Peircean model, as it is not rooted in a specific semiotic system (such as language in the Saussurean Model) and therefore lends itself better to non-speech sounds.

The Peircean Model

As opposed to the Saussurean Model's dyadic structure, the Peircean Model is triadic and requires all three aspects to be present in order for a sign to exist (Figure 2.14):

1. *The Sign (or Representamen)*: The carrier of the sign, the perceptible signal;
2. *The Interpretant*: The interpretation in the mind of the receiver of the signal;
3. *The Object*: The thing that the sign refers to.

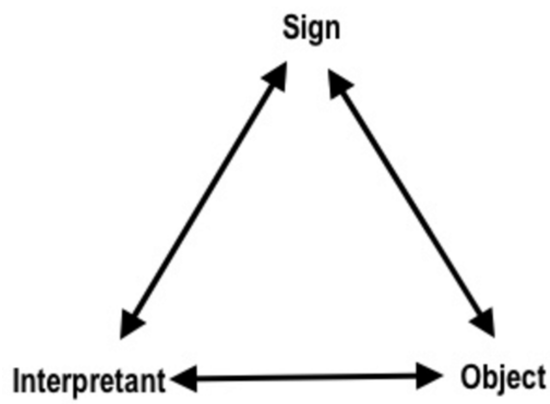

Figure 2.14: Charles Sanders Peirce's model of the sign [27].

In Peirce's terminology, the *sign* refers not only to the overall sign, but to the physically existing evidence of the sign, such as some kind of visual, haptic, acoustic or other sensory signal. Some call this component the 'sign-carrier', 'sign-vehicle' or 'representamen', but rather confusingly the word 'sign' is most commonly used for this component of a sign. The *object* can be a physical object or something intangible like the concept of 'justice', ultimately it is absolutely anything which is knowable or discussable and is a concept that a sign refers to. The final component of a sign in Peirce's model is the *interpretant*, the interpreter's conception of the meaning of the sign.

Typology of Signs

An advantage of the Peircean Model over the Saussurean Model is that Peirce offered a typology of signs, a way to break down signs into different categories. Peirce divided sign types into three categories of relationships between signs and the object to which they refer:

1. *Symbol*: The relationship between the sign and object is arbitrary or based on convention, that is to say that it has been learned or agreed upon in some way. For example, a national flag is an arbitrary representations of a nation, there is no physical or causal link between the concept of Norway and the Norwegian flag;

2. *Index*: The relationship is *not arbitrary* and is directly connected causally or physically. A classic example of this is smoke and fire. Smoke is directly caused by fire, therefore smoke in an index of fire;

3. *Icon*: The sign has some similarity to the object through some form of resemblance or imitation. For example: a portrait of a person visually resembles that person, or a scale model of the international space station shares some qualities with the real station.

The Semiotics of Mappings

A quote from Vickers' *Ways of Listening and Modes of Being: Electroacoustic Auditory Display* posits how auditory displays can fit into Peircian semiotics.

> Auditory display, then, is a semiotic activity in which a given data set (object) is represented by a sound signal (representamen) resulting in an awareness and understanding (interpretant) of that object in the mind of the listener. Furthermore, the relationship between the listener and the object is mediated by the auditory display and the ground upon which it sits. [147] (p. 5)

As pointed out by Oswald [110], semiotic theory is deeply rooted in linguistics, therefore its application to auditory communication is underutilised. As mentioned briefly at the beginning of this literature review, Auditory Icons and earcons are two of the standard approaches to conveying information through sound in a human-computer interface. Stereotypically, Auditory Icons have been regarded as *iconic* in that they are based on "everyday" sounds, whereas earcons are *symbolic* since they are based on musical principles [78], however Oswald points out the many semiotic inconsistencies with this stereotype [110].

For example, earcons are not always arbitrary and therefore symbolic, as there are many metaphorical associations that can be evoked by music. Consider the earcon played when a USB device is added in Microsoft Windows — a two tone earcon is played which ascends in pitch and the inverse (two tone descending in pitch) is played when the device is removed. Even this simple two tone earcon indicates tempo, direction of motion (ascending pitch = "in" and *vice versa*) and a melodic gestalt (the combined form of tempo and pitch) with metaphorical attributes like "calm/erratic", "fast/slow", "flowing/jarring", etc. [110]. Similarly for Auditory Icons, not all are strictly *iconic*. For example, in Gaver's Sonic Finder research which utilised Auditory Icons in a computer desktop [49] they used a "carpenter metaphor", that is the *materials of work* in a computer desktop — files and folders — sounded like wood and the *tools* — the applications — sounded metallic as a tool may do. However, Oswald questions how *iconic* these signs are, for in what way does a wooden sound resemble or indicate a digital file? Therefore, Auditory Icons may in many instances be more arbitrary and therefore *symbolic* than may have been previously considered.

In conclusion, it seems that many signs used in auditory displays are inherently polysemic, polysemy being a sign which is ambiguous or has multiple meanings [110]. For example, "beetle" can refer to an insect, a Volkswagen car or a member of "The Beatles" and in an auditory display. Therefore, due to this inherent poysemy, it is difficult to predict how a listener is going to interpret an auditory sign — as is exemplified in the previously discussed research by Kramer, Walker and Edworthy *et al*. Yet, this does not mean that meaning is lost; one listener may interpret a sign as an *icon* and another may interpret is as *symbol*, but either way the sign is still interpreted only in different ways. The structure of semiotics gives researchers and designers a language and framework with which to discuss, design and interpret mappings and, therefore, a tool to use in the struggle with the dichotomy between how an data-sound or data-vibration mapping is intended to be received by a user and how it is actually received.

2.4.1 The Importance of Analogy

Some researchers have used the basic ideas of Saussure and Peirce's semiotics to attempt to understand not only specific instances of signs, but using how we interpret'signs'as an important tool in understanding perception and cognition in general. Douglas Hofstader and colleagues at the Center for Research on Concepts and Cognition at Indiana University who investigate computational models of fundamental human thought, put analogy-making at the core of this work. Their contention is that analogy is the core of cognition [66] and that mechanisms such as isomorphism and recursion of one concept to another is how complex thought, and therefore systems which model thought, are achieved [68]. Although their work differs in focus from this thesis in that they are in the domain of cognitive science and not human-computer interaction, the perceptual problems faced can be strikingly similar. A quote from a textbook published by the group *Fluid Concepts and Creative Analogies* concisely shows this problem:

> The quality of an analogy between two situations depends almost entirely on one's perception of the situations. If Ronald Reagan were to evaluate the validity of an analogy between the U.S. role in Nicaragua and the Soviet Union's role in Afghanistan, he would undoubtedly see it as a poor one. Others might consider the analogy excellent. The difference would come from different perceptions and this representations, of the situations themselves. Reagan's internal representation of the Nicaraguan situation is certainly quite different from [Nicaraguan leader] Daniel Ortega's. [67] (p. 179)

This problem is incredibly similar to the problem of alignment between how a designer expects or intends a data-sound or data-vibration mapping to be perceived by the user. Just as Reagan and Ortega may have differing internal representations on the U.S/Nicaragua situation, two individuals may differ in their internal representations of a mapping between temperature and pitch, or pressure and vibration frequency. The fact that this problem can be found in multiple domains speaks to how a person's own internal representation of something is fundamental to how they perceive stimuli and therefore perceive information.

In a similar, but more objectivist approach, in *Philosophy in the Flesh* [90] Lakoff & Johnson posit that concrete experiences ground abstract concepts. For example, in the context of audio, a concrete experience such as hearing a fire alarm can help us ground and process the concept of danger or alarm. Furthermore, such abstract concept-grounding experiences may also come from media — books, films, television, video games, etc. To return to the Geiger counter metaphor, even if someone has not experienced a Geiger counter (or some other system which invokes a similar data-sound mapping by analogy such as a parking sensor), it is more likely that they have come across this metaphor being used in media. For example, in the film *Aliens* from 1986 [24] and the more recent video game *Alien: Isolation* from 2014 [32], where a number of key scenes involve a tracker which outputs increasing audio cues as moving objects get closer to the device. Additionally, in *Alien: Isolation* the controller outputs corresponding vibrotactile cues (Figure 2.15). From experiencing this parameter mapping, whether it be via a real world event such as using a Geiger counter or parking sensor or through media, the core abstract concept of the mapping is grounded, here being: *increasing proximity of some object results in increasing frequency of repetition of audio cues*. This mapping can then be extended by analogy to other, similar situations.

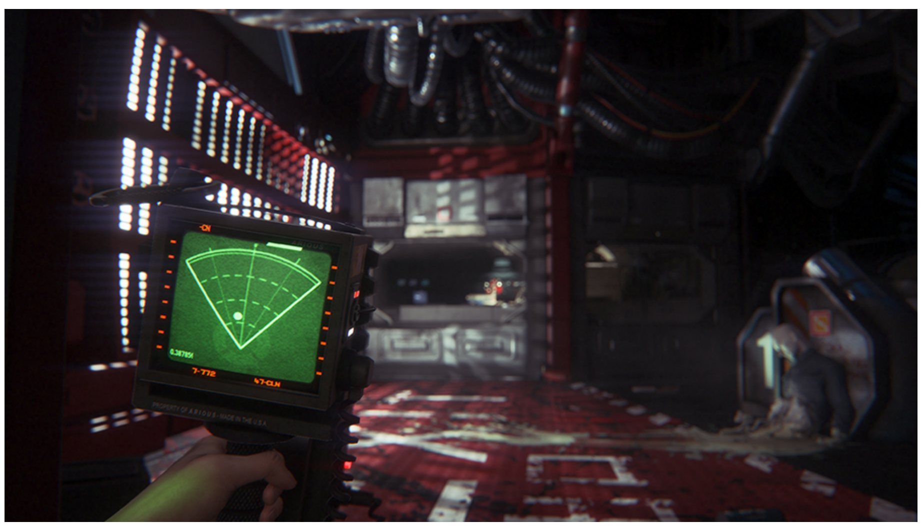

Figure 2.15: The tracker device from *Alien: Isolation* [32].

2.4.2 Summary

Aligning how a designer intends a mapping to be received with how a user actually receives it is a significant challenge in the design of parameter mappings. Users may have differing mental models of a mapping based on a variety of factors including past experiences, analogies from other contexts, or exposure to media, making it challenging for the designer to predict how a mapping will be perceived. Theories of semiotics and analogy-making can provide a useful framework for how these internal models and their relationship with the information being transmitted can be potentially applied in the design of data-sound and data-vibration mappings. A number of researchers have advocated for a semiotic approach to sonification design, however there is little to no mention of semiotics in vibrotactile feedback literature therefore it cannot be determined if there is a similar sentiment from researchers in the field. Some researchers have taken the fundamentals of semiotics and applied it broadly to thinking in general. Douglas Hofstader *et al*. [67] use analogy as a way of understanding human thought and to inform computational models of human thought and Lakoff & Johnson [90] say that all abstract concepts are built from concrete experiences, such as fire alarm = danger. Such experiences can not only be obtained through direct experience, but also be from many different media like films, books or video games and the this experience-concept relationship can be applied by analogy to other situations. This theoretical background shows the potential depth of how people process information. This depth means that predicting how someone will receive and process data-sound or data-vibration mappings is challenging, but more positively it shows that it may be possible to utilise this phenomena to design richer and more usable auditory and vibrotactile display systems.

2.5 Summary of Literature Review

This chapter reviewed existing literature on three areas: (1) auditory displays and the problems with parameter mapping sonification systems; (2) vibrotactile displays and (3) theory relating to conveying, receiving and processing information. Section 2.2 showed that the perceptual factors and processes involved in a data-sound mapping and complex varied, with different people having their own individual models and perceptions of a mapping. Although this complexity is noted throughout the literature little more has been done — there is a need for research to actually investigate this area more. This leads to the following research questions:

RQ1: How may acoustic parameters that are derived from the field of psychoacoustics encode and transmit information to a user?

RQ2: To what extent are mappings between acoustic sonification parameters and information parameters with clear positive or negative valence perceived as congruent?

RQ4: How does the complexity of the usage context impact how effective a data-sound or data-vibration parameter mapping is?

The collection of research relating to similar vibrotactile displays is lacking further still. Section 2.3 showed that there are multiple dimensions of vibration that can be perceived through the skin and research has shown that these dimension can effectively be manipulated to create a data-vibration mapping (akin to parameter mapping in the auditory displays literature).

However, also like the literature reviewed in section 2.2 on auditory displays, there is a lack of research that focuses on how to design these mappings to be more congruent with a user's mental model(s) of the particular data-vibration relationship. This leads to the following research questions:

> **RQ3:** To what extent are mappings between vibrotactile parameters and information parameters perceived as congruent?
>
> **RQ4:** How does the complexity of the usage context impact how effective a data-sound or data-vibration parameter mapping is?

Chapter 3

Psychoacoustic Parameters for Sonification

3.1 Introduction

Sonification designers have little theory or experimental evidence to guide the design of data-sound mappings. Many mappings use acoustic representations of information which may not correspond with the listener's perception of how that information should sound during sonification, with most systems defaulting to use pitch as a mapping parameter when it may not necessarily be the best choice for the particular use case. Taking guidance from psychoacoustic research when designing sonification mappings may provide data-sound mappings that are better aligned with listeners' perception of the data type being sonified. For example, many people would describe the sounds relating to a thunderstorm to be low, rumbling and booming sounds, therefore in some contexts, using an increasing pitch of a tone to sonify data relating to thunderstorm intensity may be conflicting with the listener's perception of what they think a thunderstorm sounds like. Whereas, in this case, a more perceptually congruent mapping could implement the psychoacoustic sensation of *booming* (the presence of low frequencies in a sound [63]). The first research question in this thesis, therefore, aims to address this problem by asking:

> **RQ1:** How may acoustic parameters that are derived from the field of psychoacoustics encode and transmit information to a user?

Two experiments are reported in this chapter which evaluate data-sound mappings that are based on psychoacoustic sensations, to move towards data-sound mappings that are aligned with the listener's perception and expectation of how a particular information parameter should sound when it is sonified. Multiple psychoacoustic parameters were evaluated, which were designed in the context of a domain-specific problem, that of detecting the level of focus of an astronomical image using an auditory display. Recommendations for designing sonification systems with psychoacoustic sound parameters are presented based on the results.

3.2 Application

A domain-specific problem in which these ideas may be implemented was developed with the Innovators Developing Accessible Tools for Astronomy (IDATA) research initiative.[1] IDATA aimed to use astronomy as a tool to develop transferable skills in computational thinking, software design and computer accessibility in both sighted and visually impaired students. Students taking part accessed the Skynet Robotic Telescope Network (Figure 3.1) to remotely take images of astronomical objects using a large telescope. [2] They can then analyse and share these images using the Skynet Junior Scholars portal and its related analysis software.[3] One of the main goals of the project was to provide a platform where visually impaired students can carry out these tasks unaided by a sighted person. Furthermore, IDATA aimed to ensure that the accessibility tools they provide for visually impaired students are as close to equal as possible to the tools available to sighted students in terms of functionality and accuracy.

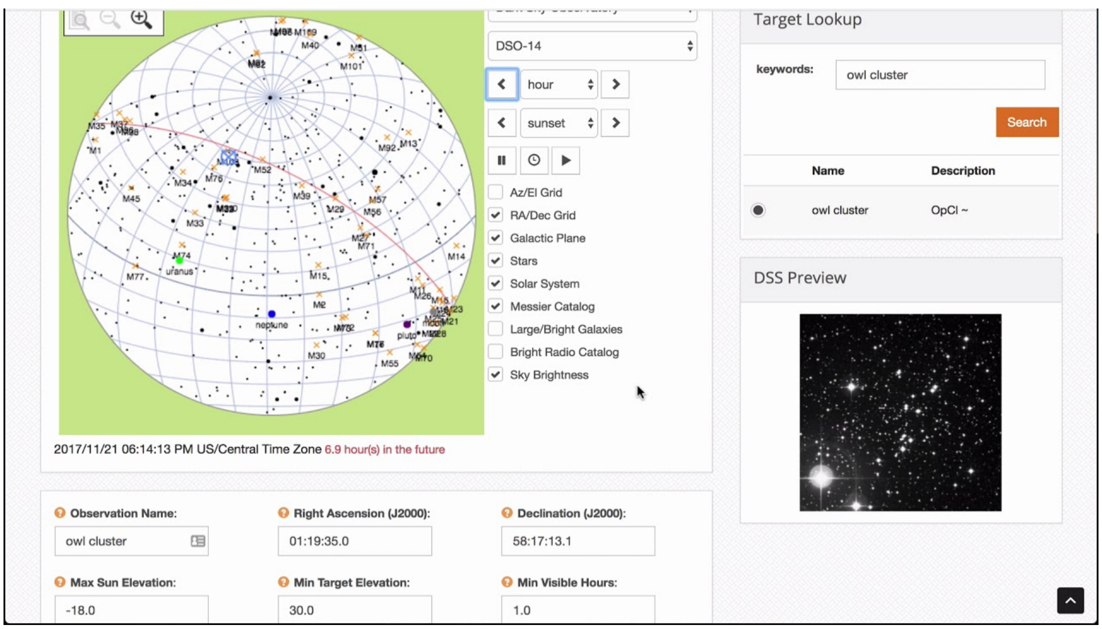

Figure 3.1: The Skynet Junior Scholars portal[4]

The application described in this section was formed during informal discussions with IDATA researchers on topics such as: information on their user group, domain-specific problems they are facing and how sonification may provide a solution to them. One of the main outcomes from these discussions was how currently it is almost impossible for their visually-impaired students to assess if an image taken from a telescope is in focus and if the tracking was set correctly, without the help of a sighted person. IDATA expressed an interest in attempting to solve this

[1] http://www.yerkesoutreach.org/idata.html
[2] https://skynet.unc.edu/
[3] https://skynetjuniorscholars.org/

problem using an auditory solution, therefore a sonification-based approach was formulated in which acoustic parameters to display the level of focus of an image were designed based on psychoacoustic sensations. As discussed above, the connotations of quality and texture present in psychoacoustic parameters may provide a data-sound mapping for this application that is more aligned with a user's preconceived perceptions of image focus.

3.3 A Brief Background of Sonification in Astronomy

Until recently, the space exploration community rarely made use of auditory displays to visualise and analyse data, however in the literature there are a few important instances where auditory display has been key in a discovery. The first example of this is NASA's Voyager 2 mission in 1982, in which scientists directly translated plasma detection data from the spacecraft to audio (using audification) to detect micrometeorite impacts, which lead to the discovery that the spacecraft had entered Saturn's ring system. A similar approach was taken by the Cassini mission in 2004, where scientists, again through audification, directly translated plasma data from the spacecraft to an audio signal to determine that it had crossed Saturn's bow shock [59].[5] These examples were simple explorations into the use of sound by astronomers and astrophysicists and were not in collaboration with auditory display researchers outright. However, since then there has been an increase in research from sound design, computing science and astronomy/astrophysics to investigate the use of auditory display in the field more thoroughly.

Alexander's 2015 PhD thesis [3] showed that using an auditory display in data analysis — using audification in particular — significantly improved professional solar physicists' ability to detect and classify solar events from heliospheric data. In this work the solar physicists used a system developed by Alexander to distinguish an underlying "hum" in the data that could be used to ascertain if certain features were present. Furthermore, these studies found that on a number of occasions, researchers detected solar events through auditory means that were missed during traditional visual analysis. Work by Diaz-Merced [101], used a NASA developed sonification tool (Xsonify [25]) to sonify a number of different types of space data using parameter mapping sonification. Diaz-Merced found that certain events of interest in the data were picked up through sonification that had been missed during visual analysis. Furthermore, perceptual experiments conducted with space scientists as a part of this work showed that using sonification as an adjunct to visual display aided scientists in analysing numerical data.

Although these examples utilise sonification for scientific reasons, the majority of sonification applications in astronomy have been in the area of public communication and outreach. In

[5] Don Gurnett, the lead author of this paper also hosts one of the best archives of space related audio (mainly audifications) at http://www-pw.physics.uiowa.edu/space-audio/

these instances, sonification is often used to convey some astronomical discovery, event or phenomenon through sound with the goal of a providing a more engaging and accessible science communication or education experience than if it were presented only using the visual modality. One of the most successful examples of this is the LIGO Collaboration's use of sonification to communicate the detection of gravitational waves [23] where they used pitch to represent the energy detected by the instrument. The European Space Agency's (ESA) Rosetta mission used audification [102] to represent data recorded by the spacecraft's magnetometer by scaling it up to be within the human hearing range. ESA's Gaia missions also made use of sonification during their public outreach [103] by using a more complex mapping with multiple parameters to sonify seven different star catalogues from history to show how they have developed. A quote from the article shows how the most recent and complex star catalogue was sonified:

> Parameters used to create the sound: Pitch (star brightness) and location in the sound space (2D position: right ascension and declination); volume (distance); hiss (colour); length of hiss (proper motion). The very high accuracy of this catalogue is conveyed by there being no background noise. Groups of ten stars are played simultaneously. ([103])

This work was a continuation of the *Bell3D* system [40] which was a preliminary attempt to provide an audio-based star map viewer, primarily intended for visually impaired users in an educational setting. Similarly to ESA's work, there have been examples of more complex and multi-parameter sonification being used in the public engagement of exoplanet science such as García Riber's *Sonigraphier* that used an interactive interface to allow a user to interact with a sonic representation of light-curves that used additive synthesis in C-Sound [48]. Furthermore, sonification has also been used in a planetarium setting, such as the system developed by Quinton *et al.* [119] which primarily used pitch to represent the size of planets, rhythm to represent their radio wave emissions and this was diffused over a quadrophonic system to show the movement of the planets. A similar approach was taken by Tomlinson *et al.* [138] in their planetarium system.

3.4 Experiment 1

An experiment was conducted to provide an initial evaluation of utilising several psychoacoustic parameters in a data-sound mapping. The experiment was designed with the IDATA application discussed earlier in mind and the experiment was planned with input from the IDATA team, primarily in the form of advice regarding what aspects of an astronomical image could be useful to sonify. The aim of the experiment was to identify how accurately participants could determine specific points on a 10-point scale for each sound parameter, thus giving an initial overview of the acuity of each psychoacoustic parameter in a sonification context. This 10-point scale was framed in terms of the image focus application mentioned previously and throughout the experiment, the acoustic parameters were discussed as being "mapped to" or "representing" the level of focus of an image.

3.4.1 Participants

Twenty-one participants took part in the study (12 female, 9 male; mean age = 24.3 years, SD of age = 7.7 years). All participants were university students and staff. All participants reported no uncorrected vision impairment (i.e. if they need glasses or contact lenses then they use them), no hearing impairment and no music/sound related neurological condition such as amusia [112].

3.4.2 Design

Four conditions were examined in which the single independent variable in each condition was the acoustic parameter used to represent the level of focus of an image. Roughness, sharpness and a combination of both were chosen; these are defined in the following section on stimuli design. The latter was used to investigate if a combination of psychoacoustic parameters result in improved performance than each alone, as some research suggested that parameters used redundantly can result in improved performance [113]. In addition, pitch was included as a condition to provide a comparison to a more traditional data mapping. The experiment used a within-subjects design. Dependent variables collected during the experiment included: number of correct responses, mean distance from correct response, reaction time for each response and perceived polarity of data-sound parameter mapping.

3.4.3 Stimuli

Ten stimuli were used for each condition. Each stimulus was two seconds in length, this was chosen as work by Brewster [15] showed that information encoded in sounds can be obtained from a stimulus between one and two seconds in duration. Each stimulus had an amplitude envelope with a 0.2 second linear ramp onset (attack) and offset (release) — shown in Figure 3.2. An amplitude envelope was included in the sound design, as an abrupt start or stop of a sound can be perceived as unpleasant [7] and could therefore impact participants' perception of the stimuli and thus their responses. All stimuli were created in the SuperCollider programming language.[6] The acoustic design of these stimuli is described below.

Roughness

This condition comprised of sinusoidally amplitude modulated broad band noise with modulation frequencies of 14, 17, 21, 25, 30, 36, 43, 52, 62 and 75 Hz. This range of modulation

[6] http://supercollider.github.io

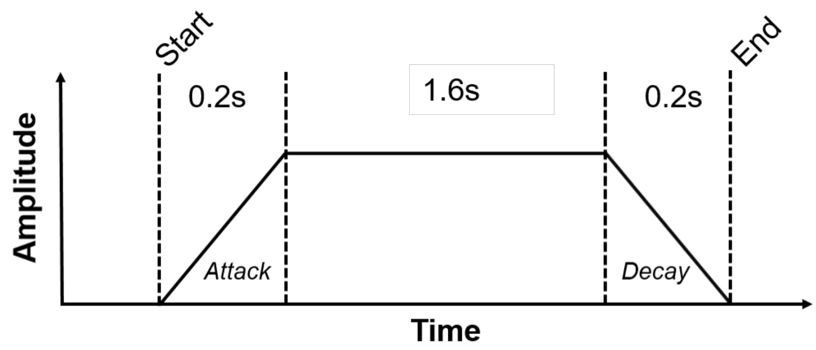

Figure 3.2: The amplitude envelope used for all stimuli in Chapter 3.

frequencies was chosen based on Fastl & Zwicker's estimation that, in the case of amplitude-modulated sine tones, a step in roughness is perceptible in steps of 17%. Informal pre-testing was carried out to determine an approximate number of roughness steps that could be perceived for corresponding steps in modulation frequency. These tests suggested that between modulation frequencies of 14 and 75 Hz, differences in roughness are just-noticeable in steps of around 17%, similar to Fastl & Zwicker's results. In this experiment, broad band noise was used as opposed to sine tones as the carrier signal for roughness. This was because the sharpness condition required the use of a complex noise signal, therefore noise was used in the roughness condition to mitigate any performance differences based on the carrier signal type. Modulation frequency was chosen as the variable of roughness to vary instead of degree of modulation, as modulation frequency provides a larger range of potential levels of roughness.

Sharpness

This condition used critical band wide white noise. For a given frequency, a critical band is the smallest range of frequencies around it that activates the same area of the basilar membrane, the primary sensory apparatus of the ear [126]. There are 24 critical bands, this scale being the Bark scale [173]. Stimuli in this condition were at Bark levels 2, 4, 6, 8, 10, 12, 14, 16, 18 and 20. This range of critical bands was chosen based on informal pre-testing which suggested that an increment in sharpness was perceptible per increment in Bark level. As increments of one Bark in the range of 24 Bark were perceptible and a resolution of only ten was required for this study, increments of two Bark were used to ensure the maximum possible distinction between levels. The range began from two Bark, as some pre-test participants reported difficulty in hearing one Bark. Loudness for each stimulus was normalised using SuperCollider's built in library for basic psychoacoustic amplitude compensation, AmpComp.[7]

[7] http://doc.sccode.org/Classes/AmpComp.html

Combined Roughness and Sharpness

This condition was a direct pairing of corresponding roughness and sharpness stimuli. This condition was included to assess how a combined psychoacoustic parameter mapping faired in comparison to mappings consisting of individual parameters.

Pitch

In this condition, sine tones with frequencies of 100, 200, 300, 400, 800, 1000, 1400, 1800, 2400 and 3200 Hz were used. This frequency range and the choice of using a sine tone was based on the stimuli used by Walker [151]. Loudness for each stimulus was normalised, again using AmpComp inside of SuperCollider.

3.4.4 Procedure

The experiment consisted of four blocks of 30 trials (one block for each condition). In each trial, participants were presented with a stimulus which lasted for two seconds and after a two second pause, was repeated. Based on this stimulus, participants were asked to respond on a 10-point scale, how clear and in focus they believed the image represented by the sound was (10 being a perfectly clear image, one being very blurry and out of focus), they did not see the image, they only heard the sound cue representing its level of focus. They indicated their response via a mouse, on a scale shown on-screen. The system was built using PsychoPy.[8] Each stimulus was presented three times, with the order of presentation randomised and the order in which each block was presented also randomised. Before starting the experiment, participants were trained to use the interface by performing two mock trials. They received no feedback on their performance during this training and received no training on the acoustic definitions of any of the parameters that were used.

Polarity Choice

At the beginning of each block, participants were presented with the 10-step scale of each acoustic parameter in full three times. After hearing all ten stimuli for a given parameter, the participant indicated on-screen whether their polarity preference was *positive*, that is an increase in the attribute meant an increase in blur or, *negative*, that an increase in the attribute meant a decrease in blur. An example of a positive mapping would be increasing pitch = increasing blur and a negative example could be increasing roughness = decreasing blur. There is a not a standard way to name these polarities, therefore positive and negative have been used, as in Walker's previous polarity work [151, 152].

[8] https://www.psychopy.org/

3.5 Experiment 1 Results

Three participants were eliminated from the evaluation as they confused their polarity choices in all conditions — meaning that in each condition they deviated from the polarity choice they initially indicated. The mean Pearson correlation was calculated between each sound level and the participant's response at that level. The criterion for elimination was: If a participant's mean Pearson correlation over all levels showed a negative correlation, this indicated a significant diversion from their selected polarity for all conditions and they were eliminated from the evaluation. Therefore, the results reported are for the remaining 18 participants.

Condition	**Negative Polarities**	**Positive Polarities**
Combined Rough + Sharp	7	11
Pitch	14	4
Roughness	12	6
Sharpness	8	10

Table 3.1: Breakdown of polarity choices for each condition in Experiment 1.

Responses for all conditions were not normally-distributed (Shapiro-Wilk test, $p < 0.05$), therefore the Aligned Rank Transform [165] was applied before analysis. The Aligned Rank Transform (ART) was developed by HCI researchers for "use in circumstances similar to the parametric ANOVA, except that the response variable may be continuous or ordinal, and is not required to be normally distributed." [165] (p. 144). The ART can therefore be used to reformat non-normal data for use in traditional factorial analysis.

First, looking at the percentage of responses that were correct, with a correct response being, for example, roughness level 7 being presented and the participant indicated that this cue represented an image with a focus level of 7. Results showed an overall correct response rate of 24%, with the pitch condition having the highest percentage of correct responses of 31.9% and roughness having the lowest of 19.3%. A one-way analysis of variance (ANOVA) looking at the effect of acoustic parameter on the percentage of correct responses showed a significant effect of acoustic parameter on the total number of correct responses ($p = 0.0208$, Table 3.2). Of the psychoacoustic parameters, the combined and sharpness conditions showed the highest percentage of correct responses (22.6%). Figure 3.3 shows the percentage of correct responses for each condition in boxplots and Table 3.3 shows percentage of correct responses for each condition numerically.

Post hoc Tukey HSD tests (with adjusted p-values being reported) showed that the pitch results were significantly better than the roughness condition ($p = 0.03$). No other significant

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Condition	3	1606.79	535.60	3.47	0.0208
Residuals	68	10496.30	154.36		

Table 3.2: ANOVA results for the effect of acoustic parameter on the percentage of responses that were correct in Experiment 1.

pairwise comparisons were found. Table 3.1 shows the split in polarity choice for each condition. To reiterate, a positive polarity suggests that an increase in the acoustic parameter is perceived as an *increase* in image blur. So for roughness here, six participants thought that a rougher sound indicated a blurrier image, or considering the pitch condition, 12 participants thought that lower pitches indicated a blurrier image.

Figure 3.4 shows the mean response error for each level of each condition in Experiment 1. This means that for a given condition level, say the first level of roughness, if a participant's response indicated that this stimulus represented focus level of two then the error would be one, as their response is one value away from the correct response. A one-way ANOVA did not show a significant effect of acoustic parameter on mean response error (p = 0.199).

Condition	**Percentage**
Combined Rough + Sharp	22.6%
Pitch	31.9%
Roughness	19.3%
Sharpness	22.2%

Table 3.3: Percentage of responses that were correct for each condition in Experiment 1

3.6 Discussion

The pitch condition resulted in the highest percentage of correct results, however it is difficult to attribute particular factors that may have resulted in this condition's increased performance. Perhaps participants are more acclimatised to hearing changes in pitch than say, roughness, as they are used to pitch from music. If the stimulus design alone is considered, the pitch condition was the only one that used a sine tone whereas the others were based on noise signals. Figure 3.4 shows that the mean error of responses (or the distance a response is from the level) is increased towards level ten (a completely blurry image) in the sharpness and combined conditions and this may also be because they are based on manipulations of noise as opposed to sine tones. One reason for this could be that using solely noise at as a parameter to convey image focus may

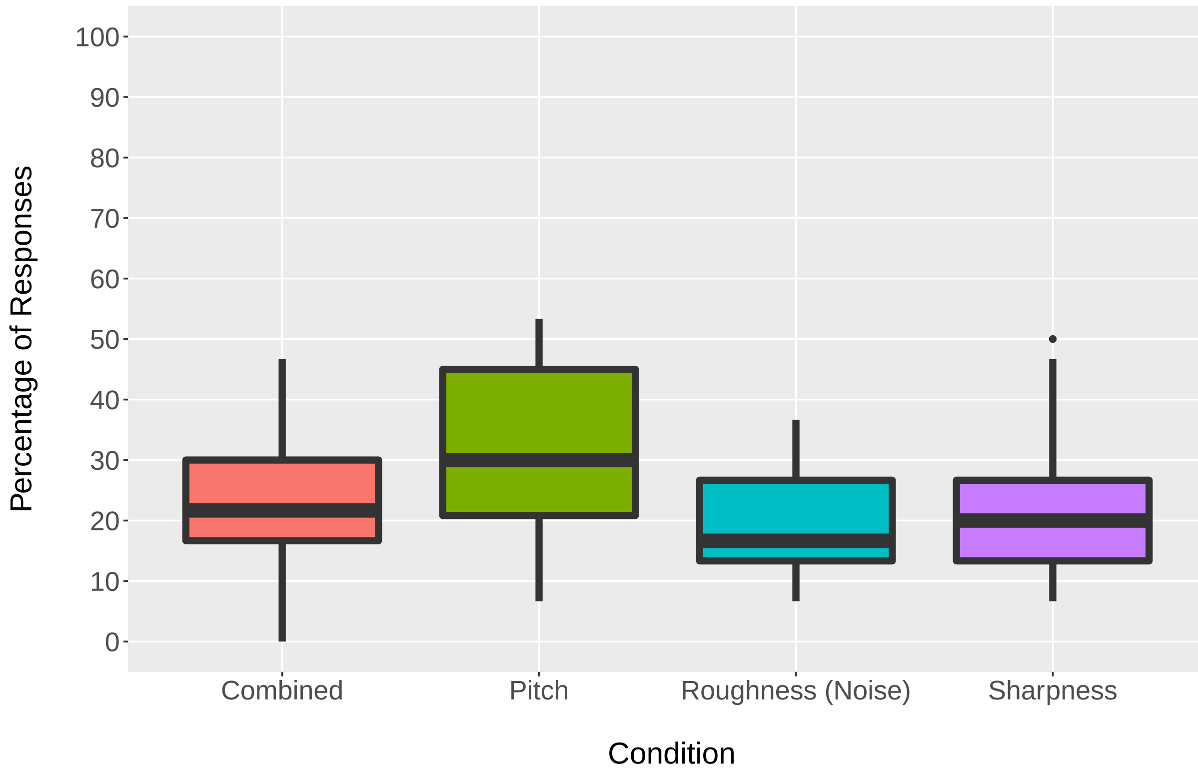

Figure 3.3: Boxplots showing the percentage of responses that were correct for each condition in Experiment 1.

not be congruent with how a listener might expect in focus, or out of focus to sound. In terms of very basic associations and analogy, noise is generally associated with negative attributes in music, sound design and product design, i.e. noise is often something to be reduced or removed. Therefore, participants may have struggled to process this parameter when used in the context of conveying focus. Experiment 2 explores the use of both sine tones and noise in order to investigate this further see if the use of sine tones versus noise or a mixture of the two results in better performance.

3.7 Experiment 2

Results from Experiment 1 indicated that the sound design of the psychoacoustic sound parameters was not the most effective and could potentially be improved. Therefore, a second experiment was conducted under the same experimental design and procedure, with a new set of psychoacoustic sound parameters. Furthermore, as IDATA aims to ensure that the accessibility tools they provide are equal to the tools available to sighted students, a visual condition was included as a comparison. This was to evaluate how auditory analysis of image focus compared to visual.

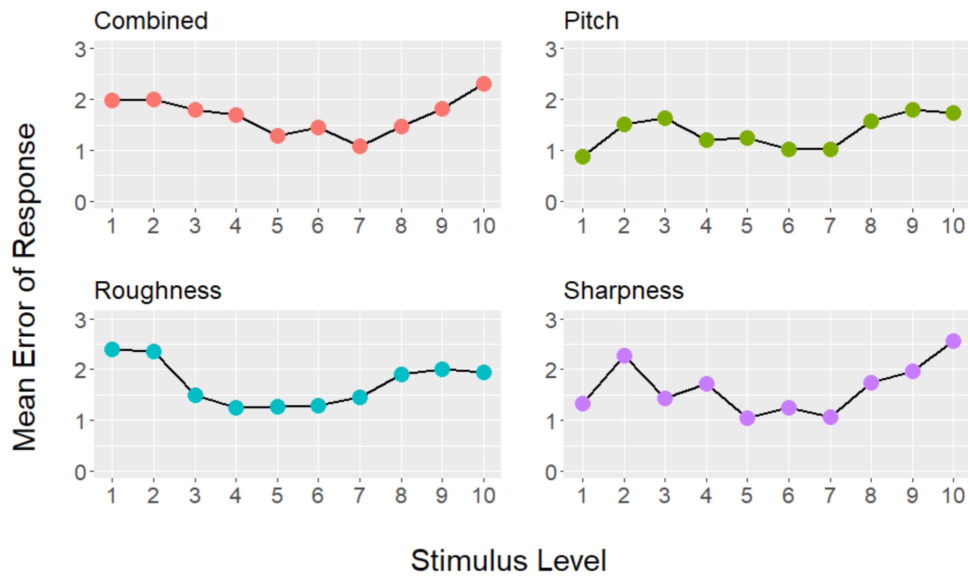

Figure 3.4: Mean error of responses from correct response for each condition and stimulus level in Experiment 1.

3.7.1 Participants

Nineteen new participants took part in Experiment 2 (11 female, 8 male; mean age = 23 years, SD of age = 3.4 years). Participants were recruited under the same criteria as Experiment 1.

3.7.2 Design

Five conditions were examined in which the single independent variable was the parameter used to represent the level of focus of an image (four acoustic parameters and one visual parameter). Roughness (of a sine tone), noise, combined roughness and noise and pitch were chosen as auditory conditions. In this experiment, the pitch range was based on a musical scale, to investigate if it was more effective than the frequencies chosen in Experiment 1. The visual condition consisted of an astronomical image being artificially blurred to various degrees. During the experiment, the same data were collected as in Experiment 1.

3.7.3 Stimuli

Auditory stimuli were designed in the same fashion as Experiment 1 and the visual stimuli were artificially blurred using OpenCV,[9] The design of these stimuli is described below.

[9]http://opencv.org

Figure 3.5: The image of the dwarf galaxy M110 used in the visual condition in Experiment 2 (NOAO/AURA/NSF).

Roughness

This condition used sinusoid amplitude modulated 1 kHz sine tones with modulation frequencies of 0, 2, 4, 7, 11, 16, 23, 34, 49 and 70 Hz. The poor results for the roughness condition in the previous experiment may have been impacted by the fact that the range of roughness did not begin on a clean, unmodulated sound at level 1. Therefore, in designing a range of roughness sounds for this experiment, a sine tone was used as a carrier instead of noise. The use of noise as a carrier in the previous experiment may have conflicted with participants' perceptions of a clear and in focus image, as noise may be associated with notions unclear or out of focus. As the transition between fluctuation strength and roughness is considered to be smooth rather than a strict border [174], the lower modulation frequency ranges of fluctuation strength were included in this experiment. These lower modulation frequencies down to an unmodulated sine tone were included so that the range of roughness converged on a sine tone. The step size of this range was based on the same informal pre-testing as the roughness condition in Experiment 1, however the step size was increased, due to the larger range of available frequencies afforded by including the range of fluctuation strength.

Noise

This condition consisted of a 1 kHz sine tone and broad band white noise, with the sine tone beginning at 100% amplitude and noise at 0%, going through noise contents of 0, 10, 20, 30, 40, 50, 60, 70, 80 and 100%. This essentially adds noise to the sine tone signal in increments of 10% until the signal is completely noise. The aim of this condition was to utilise the effect suggested by the results of Experiment 1, that noise may be associated with blurriness or lack of clarity.

Combined Roughness and Noise

This condition used a direct pairing of corresponding roughness and noise stimuli. As the combined condition in the previous experiment showed slightly increased performance over the individual parameters, a combined condition was used again.

Pitch

In this condition pure sine tones in a C-major scale (plus two extra notes to make the range ten notes long) beginning at middle C (C4, freq = 261.63 Hz) and ending on E6 (freq = 1318.51 Hz) were used. Loudness for each stimulus was normalised, again using AmpComp inside of SuperCollider. As mentioned, a musical scale was used in this experiment instead of the frequencies used by Walker [151] to see if one may be more effective than the other.

Visual

In this condition, a 1700 $\times$ 2000 pixel, grey scale image of the dwarf galaxy M110 (Figure 3.5) taken on a telescope typical of the kind used in the IDATA project was used. The image was artificially blurred using OpenCV 's *blur* function. This takes as input a window (kernel) size in pixels and each pixel in the output is the mean of all of the pixels in the window around the corresponding pixel in the input [13]. The kernel sizes used were 0 (image with no blur), 10, 20, 40, 70, 110, 160, 220 and 440 pixels squared and additionally a blank image to indicate a completely unusable image, resulting in ten images total. The experimental procedure was identical to Experiment 1. For the visual condition, the stimuli were presented for the same amount of time as in the auditory conditions.

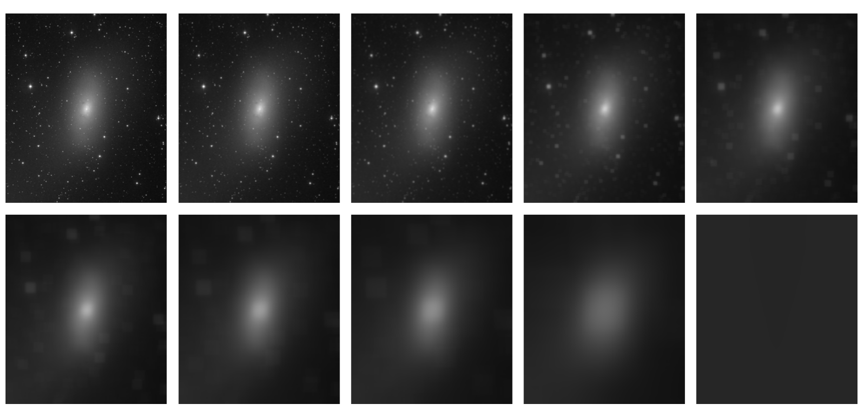

Figure 3.6: The levels of blur of the image of M110.

3.8 Experiment 2 Results

Responses for all conditions were not normally-distributed (Shapiro-Wilk test, $p < 0.05$), therefore the Aligned Rank Transform [165] was applied before analysis. A one-factor ANOVA looking at the effect of sensory parameter on the percentage of correct responses showed a significant effect of sensory parameter (acoustic or visual) on the percentage of responses that were correct ($p = 0.0118$, Table 3.4). *Post hoc* Tukey HSD tests (with adjusted p-values being reported) showed that performance in the visual condition was significantly better than the roughness condition ($p = 0.025$). No other significant differences were found.

The visual condition resulted in the highest percentage of correct responses with 53.9%, followed by the combined roughness and noise condition with 49%. The poorest performing conditions were roughness at 33.7% and pitch at 35.8%. Figure 3.7 shows boxplots of the perentage of correct for each condition and Table 3.5 shows the mean and median percentage of correct responses for each condition numerically. Figure 3.8 shows the mean response error for each level of each condition in Experiment 2. Table 3.6 shows the split in polarity choice for each condition. The combined roughness and noise condition had a unanimous polarity split, with all participants responding in a positive polarity, the only other unanimous polarity was the visual condition.

	Df	**Sum Sq**	**Mean Sq**	**F value**	**Pr(>F)**
Condition	4	5727.02	1431.75	3.43	0.0118
Residuals	90	37611.70	417.91		

Table 3.4: ANOVA results for the effect of acoustic parameter on the percentage of responses that were correct in Experiment 2.

Condition	**Mean**	**Median**
Roughness	10.1 (33.7%)	10 (33.3%)
Noise	14 (46.5%)	15 (50%)
Combined Rough + Noise	14.68 (49%)	17 (56.7%)
Pitch	10.7 (35.8%)	10 (33.3%)
Visual	16.2 (53.9%)	15 (50%)

Table 3.5: Mean and median total number of correct responses (out of 30) for all conditions in Experiment 2

Condition	Negative Polarities	Positive Polarities
Roughness	3	16
Noise	3	16
Combined Rough + Noise	0	19
Pitch	16	3
Visual	0	19

Table 3.6: Breakdown of polarity choices for each condition in Experiment 2.

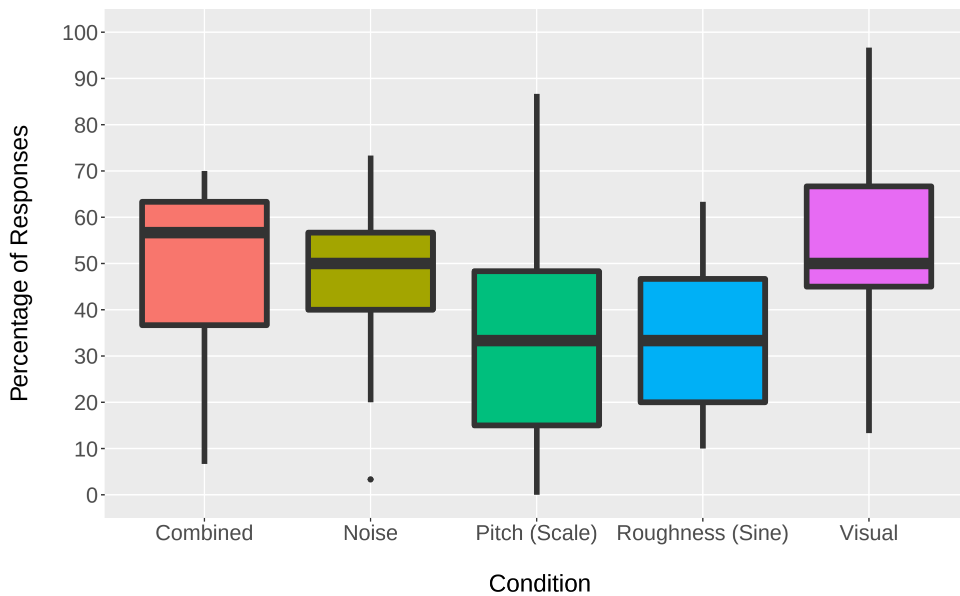

Figure 3.7: Boxplots showing the percentage of responses that were correct for each condition in Experiment 2.

An ANOVA was conducted which compated the percentage of correct responses for the roughness conditions in each experiment — where the carrier signal in Experiment 1 was noise and was a pure tone in Experiment 2 — found that the roughness condition in Experiment 2 showed significantly more correct responses than the roughness used in Experiment 1 (p = 0.003, Table 4.2).

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Experiment	1	1923.33	1923.33	10.08	0.0031
Residuals	35	6676.67	190.76		

Table 3.7: ANOVA results for the effect of roughness designs in Experiments 1 and 2 on the percentage of responses that were correct in Experiment 2.

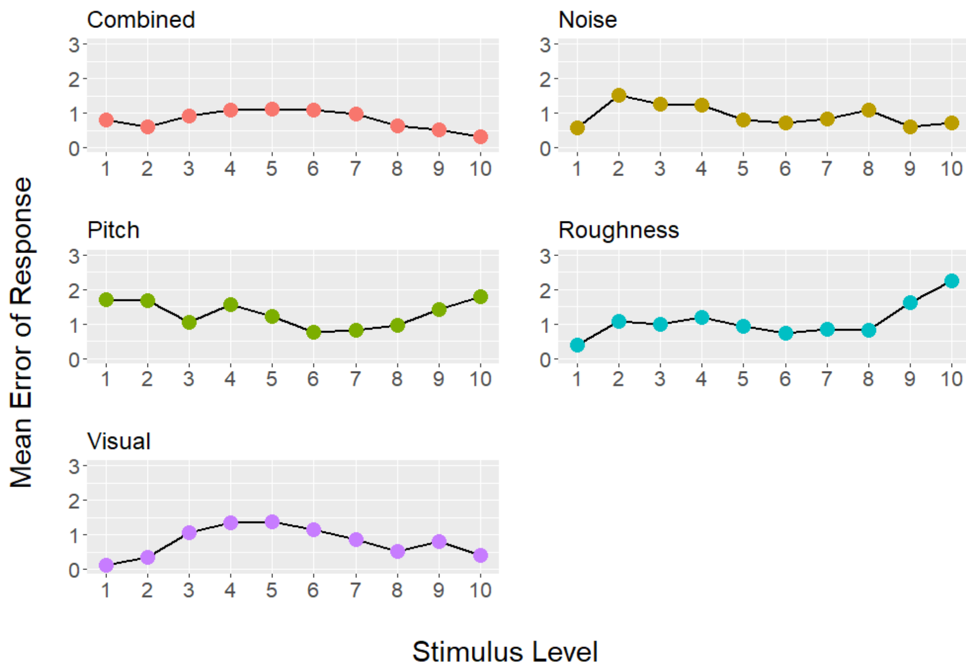

Figure 3.8: Mean error of responses from correct response for each condition and stimulus level in Experiment 2.

3.9 Discussion

Polarity

The breakdown of polarity choices in Figure 3.6, shows that all participants chose a positive polarity for this mapping, for example increasing roughness and noise in combination results in increasing blur. The roughness and noise conditions had a similar breakdown of 16 to 3. In all of these conditions (roughness, noise and the two combined), since the carrier signal was a sine tone, the first stimulus was an unmodulated tone, that is to say a sine wave with an absence of roughness or noise. The polarity choice breakdown for Experiment 1 (Figure 3.1) for the conditions where noise was the carrier signal (roughness, sharpness and combined roughness and sharpness) shows a much more mixed result.

This may indicate that using a sine tone as a carrier signal results not only in better performance, but this agreement among participants could suggest that this mapping is appropriate for the type of information being conveyed. One cause of the consensus among participants may be the fact that in these conditions in Experiment 2, the first stimulus level was an unmodulated tone and it may have represented a lack of something to the participants, similar to how a perfect image could be said to represent a lack of blur. This can be particularly seen in the noise condition in Figure 3.8 where, similar to the visual condition, the mean response error dips at the first level before rising. This dip at the first level is also present in the roughness condition.

The Zero Value in an Auditory Display

In auditory displays, particularly in PMSon systems, it is often a challenge to understand values absolutely as there are no axis labels as there are in visual displays. For example, if pitch is being used to represent temperature how does a listener know where 88°C is? Specifically, it is difficult to represent the bottom of a scale, or a zero value. In visual media, zero values can be encoded somehow and represented. For example the numerical 0, or even a blank graph is still a representation that can be perceived and understood. However, in auditory media this is not so simple, it is almost always not appropriate to use a lack of sound to represent zero because how does a listener know that the sonification is still active, or if there is some system failure that has resulted in the audio stopping? The fact that the mean error response results for the roughness and noise conditions in Experiment 2 (Figure 3.8) both show reduced errors at level one may mean that this sound design may be used to encode a zero value or state into the auditory display. As mentioned, the unmodulated tone that the roughness and noise conditions used at level one may have effectively encoded no-blur or in focus. This design could be extrapolated to other situations whereby an unmodulated sound can represent a zero state and is then subsequently modulated to convey a change in an information variable where some zero state or absence of some variable is necessary in the data-sound mapping. Figure 3.9 shows a visual representation of how this could be done.

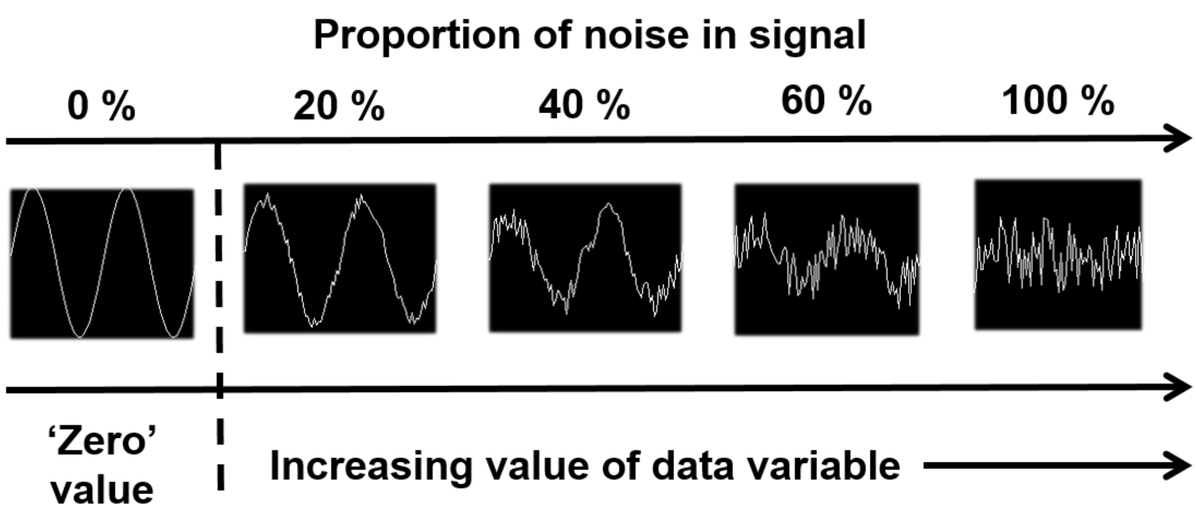

Figure 3.9: Representation of how a zero value can be encoded in a signal where noise is systematically applied.

Comparing Auditory and Visual Displays

The total number of correct responses for each condition in Experiment 2 (Figure 3.7) show that the combined roughness and noise and the singular noise conditions performed closely to the visual condition — within 10% — with mean scores for noise being 46.5%, combined being 49% and visual being 53.9%. Looking at the distribution of responses for the combined and noise conditions in Experiment 2, both are heavily negatively skewed and this can be seen from the boxplot in Figure 3.7. However the skew is more clearly seen in a density plot of these conditions (Figure 3.10) — with Pearson skewness values (using the R moments package) of -0.91 for the combined condition and -1.03 for the noise condition. Given this skew, considering the median of the responses may also give a clearer insight into these results than the mean (Table 3.5). When the median is take into account, the combined condition performs better than the visual condition by almost 7% and the noise condition performs equally to the visual.

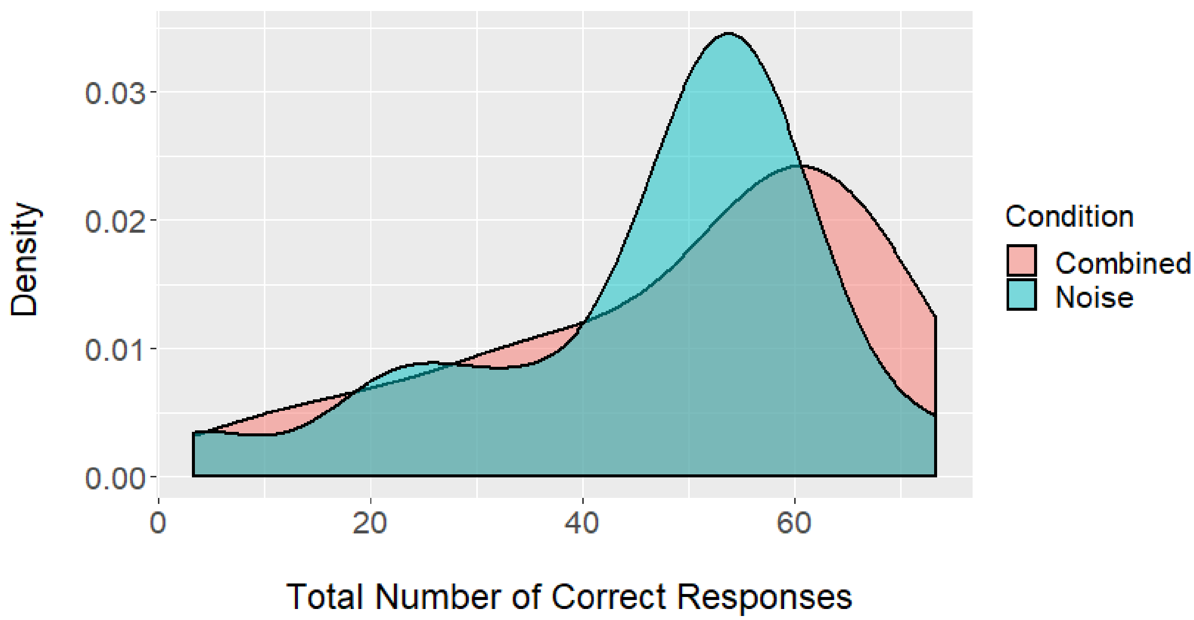

Figure 3.10: Distribution of total number of correct responses for the combined and noise conditions in Experiment 2.

Therefore, for similar applications to the one presented here, these auditory parameters may provide a satisfactory substitute or supplement for vision. This would be beneficial in contexts where the users are visually impaired, such as the context of the IDATA project, or where the visual modality is already occupied by some other task. Furthermore, as discussed in section 3.2, one of the IDATA projects goals was to attempt, as much as possible, to ensure that the tools and systems created as part of the project are usable by both sighted and visually impaired students. The closeness of some of the acoustic parameters to the visual condition in Experiment 2 indicate that this goal may be attainable, at least for simple systems and contexts like the experiment described in this chapter.

3.10 Conclusions

Over two experiments, this chapter investigated the use of a number of psychoacoustic parameters to encode information that could then be conveyed to a user in a sonification context. The experiments were designed around a domain-specific sonification problem, sonifying the level of focus of an astronomical image. The first study evaluated the use of roughness (broadband noise), sharpness, and a combination of these two as acoustic parameters alongside the pitch mapping used by Walker [151]. The second study evaluated roughness (sine tone), noise, a combination of the two and pitch (using a C-Major scale) as acoustic parameters, as well as a visual condition.

Conditions in Experiment 2 resulted in improved performance over Experiment 1. There was also more of a consensus among preferences of polarity in Experiment 2 that could suggest that the mapping design used was more appropriate, or more closely aligned with what participants may have expected. The roughness and noise conditions in Experiment 2 showed decreased errors at the first level of the spectrum of focused to unfocused/blurry, where the first level represents a totally clear image. For both of these conditions the first level stimulus was a sine tone, with no modulation (roughness) or noise added, therefore this seemed to act as a zero value, i.e. a lack of blur and participants appeared to find this level particularly easy to identify. In Experiment 2, the combined roughness and noise and the noise alone conditions performed very closely to the visual condition. Furthermore, the combined condition resulted in better performance than the visual condition and the noise condition was equal to the visual when the median values were considered, which here is not trivial due the significant negative skew of the distribution of the combined and noise results.

3.10.1 Research Question 1

The experiments discussed in this chapter show that psychoacoustic parameters can be used to encode and convey information and do hold advantages over other acoustic parameters that are more commonly used in PMSon systems. The main outcomes of these experiments are summarised and presented in the form of the following design recommendations that address Research Question 1, which asked:

> **RQ1:** How may acoustic parameters that are derived from the field of psychoacoustics encode and transmit information to a user?

Psychoacoustic Parameters Can Encode and Convey Information

This research has shown that roughness, noise and the combination of both of these parameters can be effective at conveying a simple information parameter, similar to the focus value used in this study. However, the particular information parameter used may impact the users perception. Experiment 2 showed that using a carrier signal comprising of an unmodulated sine tone which subsequently increases a given psychoacoustic parameter in an additive manner resulted in higher performance than when broadband noise was used as a carrier signal in Experiment 1. In summary, going from a clean, unmodulated tone to a rough or noise-filled signal to represent the scale of focused to blurry could be more perceptually congruent, here meaning the mapping between the acoustic parameter and the information parameter is perceived as congruent to a listener, unlike a mapping which only uses manipulations of noise.

A Sine Tone Carrier and Extended Amplitude Modulation Range Improve Roughness

The roughness parameter used in Experiment 2, which used a sine tone instead of noise, in addition to an expanded range of modulation frequencies over the range used in Experiment 1 resulted in a significantly higher percentage of correct responses. Therefore, designers intending to use roughness in a data-sound mapping may find better results if a sine tone signal is used as a carrier signal and a modulation range of 0-70 Hz is used.

Roughness and Noise Can Encode a Zero Value

An unmodulated signal — here, specifically referring to a signal with no roughness or noise is applied to it — can represent or encode a zero value, or simply put, an absence of whatever information parameter is being sonified. This is not possible when using more commonly-seen acoustic parameters in sonification such as pitch or tempo, as it is not possible to have a sine tone with no frequeny or a beat with no tempo.

Conditions Using Noise as a Parameter Rivalled Vision

In Experiment 2, the combined noise and roughness and noise alone conditions performed similarly to the visual condition. Therefore these parameters may be useful in systems where the intent is to substitute an auditory display for a visual one, with the caveat that the data-sound mapping used here was very simple and therefore should be considered when using these parameters without further research.

Chapter 4

Perceptual Congruence in Data-Sound Mappings

4.1 Introduction

The previous chapter reported results of two experiments that presented a first attempt to evaluate how psychoacoustic parameters could be used as the acoustic component of a data-sound mapping within a PMson system. Perhaps the most salient finding from the preceding chapter is that the polarity preferences for the roughness and noise conditions where the carrier signal was a pure tone were predominantly positive (16 out of 19 participants). Positive polarity here meaning an increase in the acoustic parameter is perceived as an increase in the data parameter, so more noise or roughness applied to a pure tone was perceived as representing more 'blur' in an image. This polarity preference would seem to be expected, given that in music and sound design of the Western tradition, extraneous noise and dissonance (of which roughness is a key part [79]) is commonly considered to be undesirable and are mitigated or removed where possible. Experiments 1 and 2, however, were in the context of a very specific use-case, distinguishing how in or out of focus an astronomical image was. To better understand how to use these psychoacoustic parameters in an auditory display, it would be beneficial to understand how these parameters are perceived in data-sound mappings that are more contextually generalised, so that designers and researchers of PMSon systems may use the results in mappings with other, similar information parameters.

Generally speaking, the mappings discussed from the previous chapter can be discussed in terms of *valence*, which in psychological terminology refers to "events, objects or situation that possess intrinsic attractiveness or aversiveness" [46] (p. 207). The finding from the previous section that increasing a musically undesirable parameter like noise was perceived by most participants to represent an increase in blur could be considered to represent *negative* valence, because one undesirable parameter, or one with negative valence, represents another parameter with negative valence. It is difficult to specifically identify why this mapping polarity was the most preferred, however one potential cause could be the listener's mental model of blur in the context of this parameter being presented through sound. For example, one potential conjecture could be that, in the noise condition, the pure tone signal was increasingly 'obscured' by the added noise which may have been considered analogous to how to how blur obscures a visual image and was therefore perceived as perceptually congruent and thus the preferred polarity.

This chapter presents an experiment which further investigates the idea of perceptual congruence in data-sound mappings by exploring how roughness, noise and pitch relate to the perceived magnitude of stress, error and danger. The relationships between sounds and their perceived meaning and connotations are complex, making auditory perception an important factor to consider when designing sonification systems. Listeners often have a mental model of how an information parameter should sound during sonification and this model is not considered in most data-sound mappings. This can lead to mappings that are difficult to use and can cause confusion. This problem is effectively conveyed in Vickers *et al.*'s *Aesthetics of Sonification: Taking the Subject-Position*:

> A major design challenge is to create sonifications that are not only effective at communicating information but which are sufficiently engaging to engender sustained attention. Sonification may be ineffective if the rendered sound appears arbitrary to the listener in relation to the underlying data. The design task then becomes about finding a suitable fit between communicational efficacy and appropriate aesthetic character. [148] (p. 2).

Therefore, understanding how a variety of acoustic and information parameters with negative valence (or positive valence when they are inverted) are perceived when the two are combined in a data-sound mapping may provide a first step to addressing this problem of creating perceptually congruent auditory displays. The second research question of this thesis aims to address this by asking:

> **RQ2:** To what extent are mappings between acoustic sonification parameters and information parameters with clear positive or negative valence perceived as congruent?

There is little theory or evidence to guide sonification designers in the most effective data-sound mapping to use in particular contexts and there is even less research on evaluating the perceptual reactions of listeners to acoustic parameters in sonification, beyond pitch, tempo and loudness [151, 152, 155]. In this experiment, relationships between a number of psychoacoustic parameters and conceptual variables, such as the relationship between perceived danger and auditory roughness [5] are utilised and new ones are proposed in an attempt to expand the number of effective sonification parameters and to design data-sound mappings that are complementary to the listener's mental model of how an information parameter should sound during sonification. The experiment described here uses magnitude estimation to map how the magnitudes of acoustic parameters relate to the perceived magnitudes of information parameters. Magnitude estimation has been shown as an effective method of establishing if data-sound mappings are conflicting with how the listener expects the information parameter to sound [151, 152].

4.2 Experiment 3

In this experiment, magnitude estimation was used to evaluate the perception of the magnitude of three information parameters: *stress*, *error* and *danger* based on four acoustic parameters: *roughness*, *noise*, *combined roughness & noise* and *pitch*. The information parameters were chosen because they are semantically broad, allowing for various use-cases to utilise this study's findings. Such scenarios could include:

- Biofeedback of physiological measures to reduce stress/anxiety such as the *Sonic Respiration* system developed by Harris *et al.* [61].
- Sonifying error values in non-visual graph presentation, such as in the *TableVis* system developed by Kildal & Brewster [84].
- Monitoring or alarm systems where multiple levels of danger, urgency or severity must be conveyed.

The experiment investigated how the perceived magnitude of the information parameter changed as the magnitude of the auditory stimuli changed and in which polarities the participants perceived this mapping to be. Polarity information is a useful indication of when a data-sound mapping is conflicting with the participants' mental models [151], as agreement in the polarities between participants suggests correlation between their mental models and the data-sound mapping. Furthermore, magnitude estimation produces the slope values between each data and acoustic dimension, allowing future designers to scale their mappings appropriately.

4.2.1 Participants

Sixteen participants took part in the study (11 female, 5 male; mean age = 25 years, SD = 5.1 years). All were university students and staff. All participants reported no uncorrected vision impairment, no hearing impairment and no music/sound related neurological condition such as amusia [112].

4.2.2 Design

A magnitude estimation experiment was conducted based on the design originally suggested by Stevens [134] and further described by Marks & Gescheider [97]. Applying this design to a sonification context was based on previous experiments by Walker [151, 152] and Walker *et al.* [156]. Twelve conditions were examined in which the independent variables in each condition was the information parameter used in the data-sound mapping: error, danger, stress and the acoustic parameter that was used to convey these information parameters: roughness, noise, and a combination of both. Pitch was also included to compare to a more traditional parameter mapping.

Error, danger and stress were chosen as information parameters, due to previous studies' suggestions that the perception of danger can be dependent on acoustic qualities [5] and the previous chapter's findings that the blur of an image (generalised to error, as blur could be described as an error in an image) was strongly correlated with roughness and noise, in terms of the polarity that participants preferred in a mapping between these parameters. Furthermore, these information parameters were chosen because they are a general set of potential use cases that are semantically similar. All of these are generally considered negative variables with negative valence in that they are generally considered undesirable. For example danger is generally avoided where possible and error in any form is generally mitigated or reduced. The experiment used a within-subjects design. Each condition was a pairing of an information parameter and acoustic parameter into a data-sound mapping, dependent measures collected were the participants estimations of the magnitude of the information parameter.

4.2.3 Stimuli

Ten stimuli were used for each condition, and the design was similar to the stimuli used in Experiment 2, as it was desirable to try and leverage the effects found in the previous experiment. Each stimulus was two seconds in length. Brewster [15] showed that information encoded in sounds can be obtained from a stimulus between one and two seconds. Each stimulus had an amplitude envelope with a 0.2 second linear ramp onset (attack) and offset (release). An amplitude envelope was included in the sound design, as an abrupt start or stop of a sound can be perceived as unpleasant [7]. All stimuli were created in SuperCollider. Pitch stimuli were based on the frequencies used by Walker [151].

Roughness

Stimuli in this condition consisted of 100% sinusoidally amplitude modulated 1 kHz sine tones with modulation frequencies of 0, 2, 4, 7, 11, 16, 23, 34, 49 and 70 Hz.

Noise

This condition consisted of a 1 kHz pure tone and broadband white noise, with the pure tone beginning at 100% amplitude and noise at 0%, going through noise contents of 0, 10, 20, 30, 40, 50, 60, 70, 80 and 100%

Combined Roughness and Noise

This condition used a direct pairing of corresponding roughness and noise stimuli.

Pitch

This condition used sine tones with frequencies of 100, 200, 300, 400, 800, 1000, 1400, 1800, 2400 and 3200 Hz. Loudness for each stimulus was normalised using Supercollider's built in library for basic psychoacoustic amplitude compensation, AmpComp. Frequencies used here were based on those used by Walker [151].

4.2.4 Procedure

Before the first block of trials, the experimenter read aloud the following text (adapted from [151] and [97]) and the participant read along on-screen.

> You will hear a series of sounds, one at a time, in random order. Your task is to indicate what magnitude of the variable the sounds seem to represent, by assigning numbers to them. For the first sound, assign it any number of your choosing that represents a value of the variable (e.g. stress). Then, for each of the remaining sounds, estimate its 'stress', relative to your subjective impression of the first sound.
>
> For example in the case of stress as the variable, if the second sound seems to represent a stress level that is ten times as 'stressful' as the first, then assign it a number that is tentimes bigger than the first number. If the sound seems to represent a stress level that is one-fifth as 'stressful', assign it a number that is one-fifth as large as the first number, and so on. You can use any range of numbers, fractions, or decimals that seem appropriate, so long as they are greater than zero.

Here the first sound is used as a comparison value, it is the middle level of all the stimuli (level five of ten). The participant indicated the comparison magnitude once at the beginning of the block. In each trial, participants were presented with the comparison sound with its user-defined value simultaneously displayed on-screen and then after a one second pause, one of the nine remaining stimuli were presented. The participant was then asked to respond with a subjective value (i.e. the magnitude of stress they perceived the sound cue to represent). Each block of the experiment consisted of 27 trials. In each of these blocks, each stimulua was presented three times in random order and after a brief break, the next condition was presented, with new instructions that presented the mapping to be used. The order in which each block was presented was also randomised.

Taking the danger-roughness condition as an example, in this block, roughness level 5 would be used as the comparison value. Firstly, participants are presented with this stimulus and asked to indicate what magnitude of danger this stimulus represented. This is then used as the comparison value. The remainder of the block consisted of each of the remaining nine roughness levels, which would be presented and the participant would compare these to their comparison value. Each of these remaining nine stimuli are presented three times, in random order.

4.3 Results

First, data were split on polarity response for each data-sound mapping. Then, as the scale of the responses could vary widely between each participant, responses for each data-sound mapping and polarity were normalised using the geometric mean. Stevens *et al.* in their textbook on psychophysics [137] indicate that the distribution of the log magnitude estimations is approximately normal and therefore an effective averaging for such a log-normal distribution is using the geometric mean of responses. Therefore geometric mean averaging was chosen (also due to this method's use in Walker's sonification magnitude estimation work [151, 152] and more recent HCI specific magnitude estimation studies [135]).

Consistent with Walker's findings [151], for a given data-sound mapping most participants responded in a consistent polarity. Some participants responded in an erratic polarity, therefore it was necessary to group the data into *positive*, *negative* or *no* polarity. This was done by calculating the Pearson coefficient for each participant in each block between the log of the responses and the log of the actual stimulus value. A participant's data were considered to have *no* polarity and therefore not included in further analysis if the absolute coefficient for that block did not reach statistical significance. Therefore, with nine stimuli presented three times ($df = 27$), data sets with a correlation coefficient of less than $r_{critical} = 0.367$, $p = <0.05$) were removed. Data reaching statistical significance were grouped into polarities based on the sign of their correlation coefficient (positive or negative). Table 1 shows the number of participants whose data remained in the analysis of each condition, compared to the number of participants who originally carried out the experiment.

After this preparation, a simple linear regression was carried out for each data-sound pairing, using the logs of both the geometric means of responses for each stimulus and the actual stimulus values (as per [151]). The slope value was calculated using the geometric mean of all participant responses for a given polarity group within a condition. These mean response values

	Error	**Stress**	**Danger**
Noise	15/16	12/16	10/16
Roughness	10/16	13/16	13/16
Combined	12/16	15/16	14/16
Pitch	11/16	13/16	12/16

Table 4.1: Proportion of participants whose data was of a high enough collinearity to be used in analysis for Experiment 3.

were plotted against the actual stimulus values in log-log coordinates and fitted with the power function $y = bx^m$ where the slope of the fit line m indicates how much the perceived, or estimated magnitude changes as the actual acoustic parameter changes [151]. The slope value of this regression indicates how much the perceived value changes based on changes in the stimulus. Figures 4.1 to 4.4 show the results for each data-sound pairing grouped by acoustic parameter.

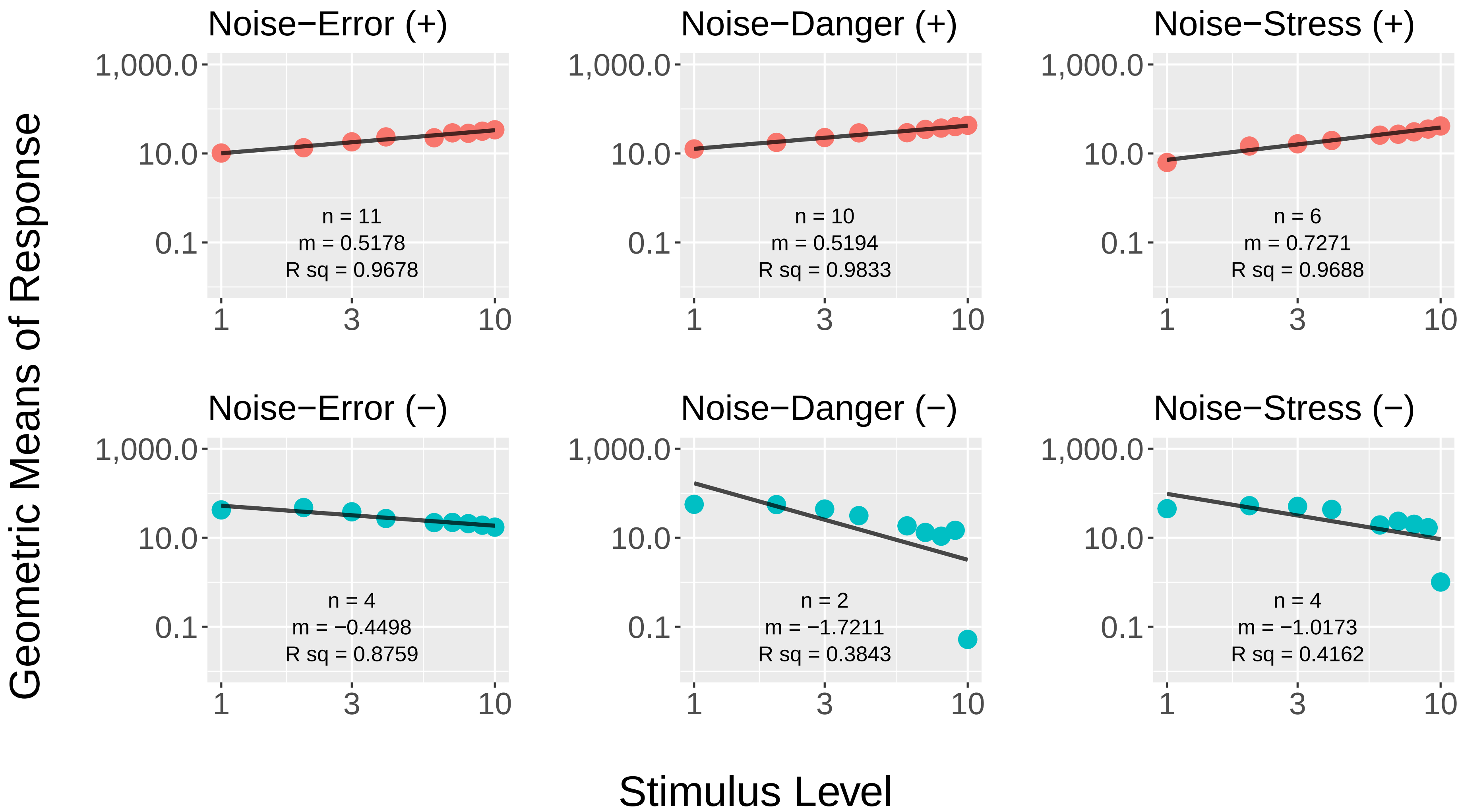

Figure 4.1: Plots for the geometric mean of estimated values against the actual stimulus level for the noise conditions in Experiment 3. Axes are in log-log scale. The first row in red shows positive polarities and the second row in blue shows the negative polarities. *n* denotes the number of participants that responded in that polarity, *m* denotes the slope value and *R sq* denotes the coefficient of determination (R^2)

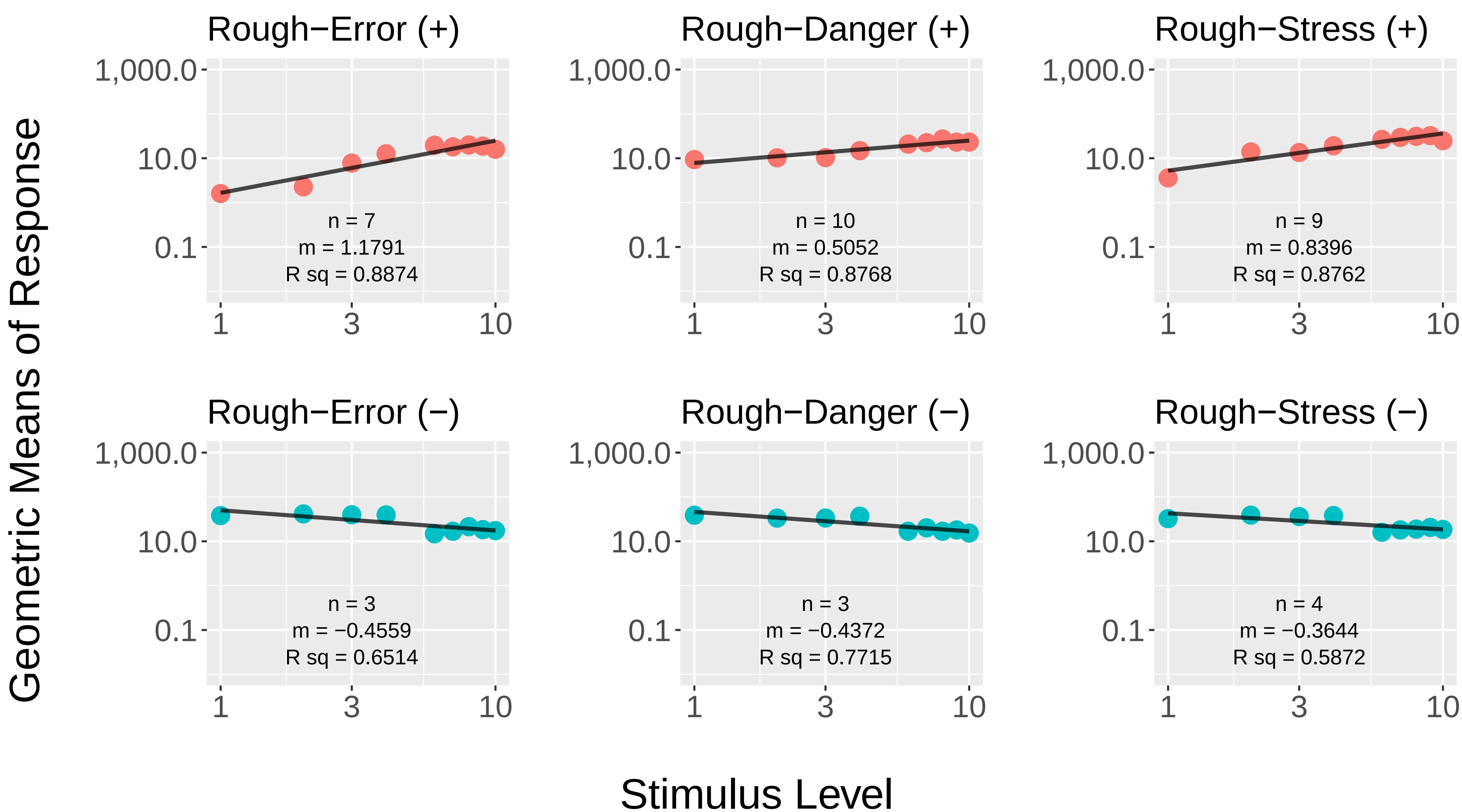

Figure 4.2: Plots for the geometric mean of estimated values against the actual stimulus level for the roughness conditions in Experiment 3. Axes are in log-log scale. The first row in red shows positive polarities and the second row in blue shows the negative polarities. *n* denotes the number of participants that responded in that polarity, *m* denotes the slope value and *R sq* denotes the coefficient of determination (R^2).

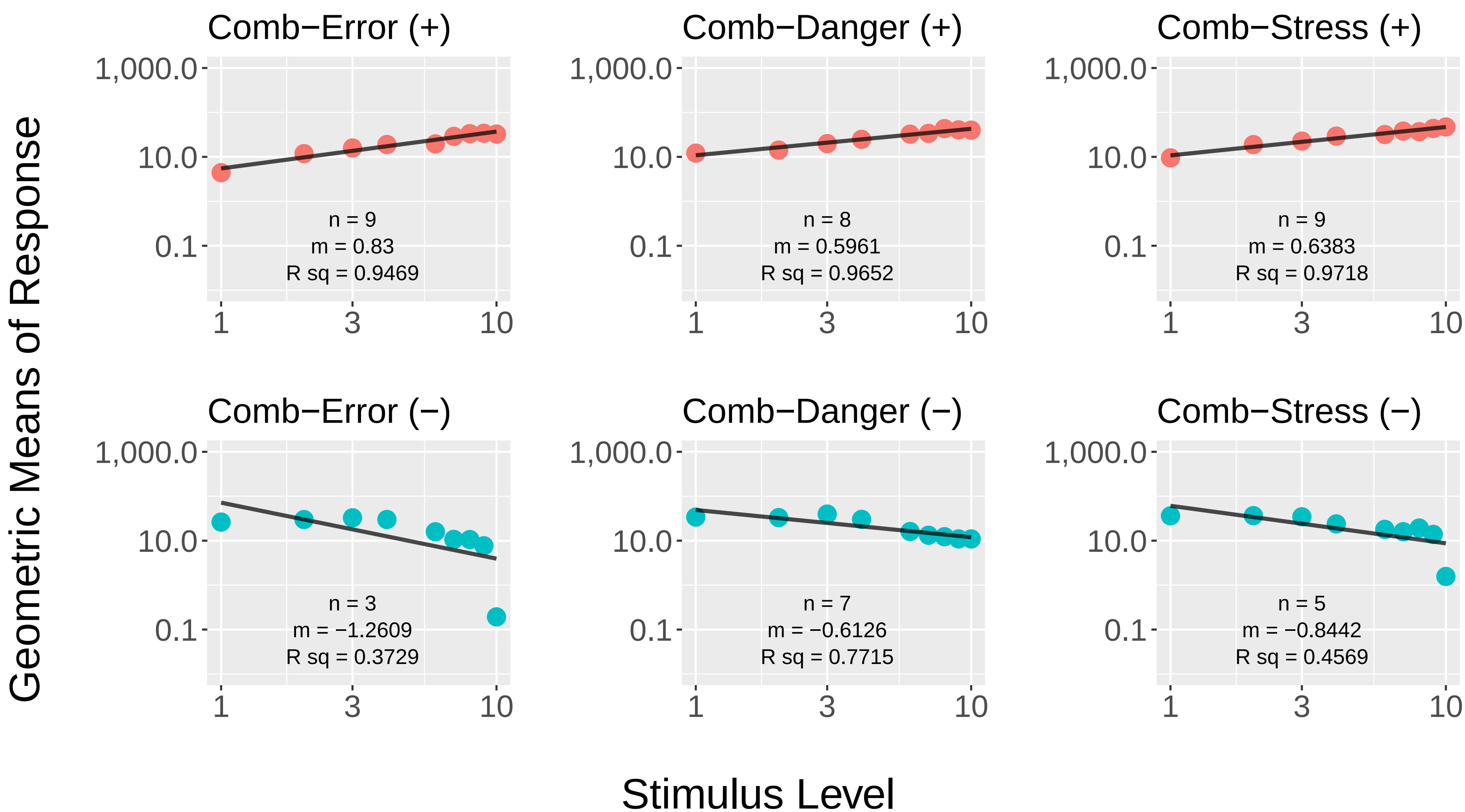

Figure 4.3:]
Plots for the geometric mean of estimated values against the actual stimulus level for the combined roughness and noise conditions in Experiment 3. Axes are in log-log scale. The first row in red shows positive polarities and the second row in blue shows the negative polarities. *n* denotes the number of participants that responded in that polarity, *m* denotes the slope value and *R sq* denotes the coefficient of determination (R^2).

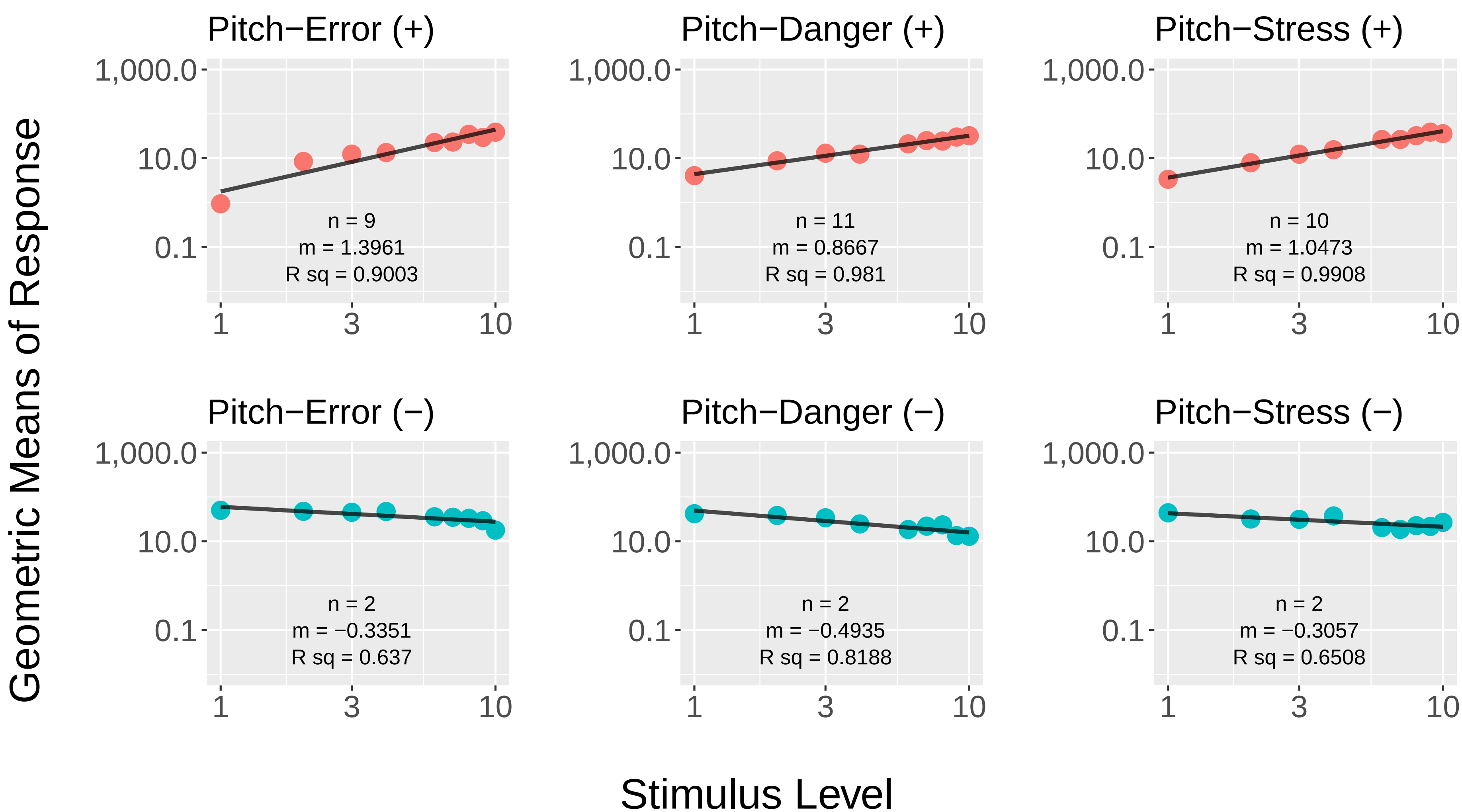

Figure 4.4: Plots for the geometric mean of estimated values against the actual stimulus level for the pitch conditions in Experiment 3. Axes are in log-log scale. The first row in red shows positive polarities and the second row in blue shows the negative polarities. *n* denotes the number of participants that responded in that polarity, *m* denotes the slope value and *R sq* denotes the coefficient of determination (R^2).

4.4 Discussion

Polarity

For the noise (Figure 4.1), roughness (Figure 4.2) and combined roughness and noise (Figure 4.3) conditions, the majority of participants responded in a positive polarity. As the stimuli at level 0 for all of these conditions was an unmodulated pure tone, participants may have perceived this clean tone as very low or no error/stress/danger, due to the absence of any additional acoustic attribute in the sound. Furthermore, at the opposite end of the stimuli levels, the roughness condition at level ten consisted of a very dissonant sound (roughness being a key part of the perception of dissonance [79]) and the noise and combined conditions at level purely of white noise. Therefore, participants may have associated these sounds with error/stress/danger, as these attributes of sound are normally considered unpleasant, just as error, stress and danger are generally considered unpleasant variables. The majority of the responses for pitch mapping (Figure 4.4) being in a positive polarity may be explained by higher pitches resulting in a higher sense of urgency, such as an infant crying [169]. These results suggest that for data sets relating to unpleasant or undesirable qualities such as the information parameters used in this study possess, the acoustic parameters used in this experiment may provide auditory displays that are consistent with listener mental models of similar variables.

Mapping Stress to Noise

There appears to be less consensus in polarity preference for the noise-stress mapping compared with other data-sound mappings, with 6 participants responding in a positive polarity and four responding in a negative polarity. One explanation for this could be the fact that, although noise is generally undesirable in the fields of music and sound design, it can be desirable in others, for example its use as a sleep aid [1]. This may also be reflected in the responses in the negative polarity for the noise-danger and noise-stress mappings (Figure 4.1) where the estimated value of danger/stress at stimulus level ten — that consisted solely of white noise, was far lower than the trend of the other stimulus levels. The fact that the results for the noise condition in the noise-error and noise-danger mappings trended more toward the positive polarity may lead to the assumption that noise is then always going to be associated with negative valence, however the findings for the noise-stress mapping are not as clearly divided. This does not detract from the fact that based on these results there is a general sense that more noise equates to a higher value of an information parameter with negative valence, but it highlights the importance of the listeners mental model and expectation in the processing of data-sound mappings and how they can change significantly when the information parameter in data-sound mapping is changed.

Slope

The results show that how a listener perceives the magnitude of an information parameter varies depending on the data-sound mapping used, participants heard the same sounds but their responses depended on the information parameter being used. Similar to the previous discussion regarding the polarity of responses, the slope values that were obtained for each data-sound mapping depended on the information parameter being used. This further underlines the findings by Walker [151] and Walker & Kramer [155] that the mapping topology used in a sonification system has a significant effect on how the user interacts with it. Using a magnitude estimation paradigm to estimate the polarities and slopes, like the one used in this study and Walker's initial studies [151, 152] can point towards mappings that are consistent with the listener's mental model of how an information parameter should sound.

4.5 Conclusions

This chapter investigated the effect of acoustic parameter choice on a number of data-sound mappings in which the acoustic parameter was based on research from the field of psychoacoustics. This was done by carrying out an experiment in which magnitude estimation was used to map how the perceived magnitude of an information parameter changed based on a change in an acoustic parameter. Polarities and scales were derived for each mapping. Results from this experiment show that the acoustic parameter used to convey an information parameter impacted the listener's perception of the magnitude of that information parameter, as each data-sound mapping produced unique polarity preference consensus and slope values. Furthermore, the majority of participants tended to prefer a positive mapping in all data-sound parameters. This indicates that most participants perceived an increase in an information parameter with negative valence to be better represented by an increase in an acoustic parameter that could be perceived as musically undesirable. This is complementary to the results from Experiments 1 and 2 which showed that most participants perceived a parameter mapping between the quality of an image and acoustic roughness or acoustic noise to be in a similar polarity as the results found here — increasing noise/roughness was perceived as representing an increase in image blur (and thus a decrease in quality). Overall, this suggests that to achieve an effective and successful data-sound mapping for a PMson system then the data-sound mapping must be congruent with the listener's mental model of how they expect the data value to sound in a sonification system.

4.5.1 Research Question 2

The findings from this experiment suggest that acoustic parameters which may be considered undesirable in terms of Western music tradition are perceived by listeners as congruent with information parameters which have negative valence. These findings can be used as a starting point for designers of PMson systems that wish to utilise such perceptual congruence in order to design data-sound mappings in that align with their expectations of how such a mapping should be presented in terms of its polarity and scale (or slope). Furthermore, this study provide polarity and slope information for a number of data-sound mappings that may be used by designers to inform their mappings using similar information parameters. The main outcomes of this experiment is summarised and presented in the form of the following design recommendations that address research question 2, which asked:

> **RQ2:** To what extent are mappings between acoustic sonification parameters and information parameters with clear positive or negative valence perceived as congruent?

Roughness, Noise and Pitch Map to Information Parameters with Negative Valence

In all conditions, the most popular polarities were positive (such as, an increase in the acoustic parameter was perceived as an increase in the information parameter). This suggests that using increasing roughness, noise or pitch to represent an increase in similar undesirable information parameters to the ones used here, may be a mapping that is aligned with a listener's expectations of how that information parameter should sound when it is sonified.

Certain Data-Sound Mappings Have More Complex Valence Relationships

The noise-stress condition had a much more split consensus regarding polarity preference, compared to not only the other conditions which used noise as an acoustic parameter, but any other condition. Although the trend from this experiment and those discussed in the previous chapter is that musically undesirable variables are perceived as congruent with information parameters with negative valence, the noise-stress mapping is an example of a mapping where the relationship between the valence of the acoustic parameter and information parameter is more complex than positive = positive and negative = negative. This is an important notion to keep in mind when choosing acoustic parameters to pair with information parameters in a PMSon system and adds more weight to the importance of understanding a data-sound mapping on a perceptual level before it is integrated into a PMSon system, to avoid confusing mappings.

An Unmodulated Signal Can Convey a Minimum Value

For the roughness, noise and combined conditions, the level 0 stimulus was an unmodulated tone (i.e. no roughness or no noise applied to it) and responses for all data-sound pairings for these conditions tended toward a positive polarity. This suggests that an unmodulated tone can convey an absence or a very low value of an undesirable information parameter such as stress or danger. This is reflective of the findings in the previous chapter's experiments.

Chapter 5

Perceptual Congruence in Data-Vibration Mappings

5.1 Introduction

The previous chapter discussed how acoustic parameters that are based on psychoacoustics are perceived in relation to information parameters with negative valence in the context of a parameter mapping sonification system. The most salient finding from this experiment — which echoes the findings from the Experiments 1 and 2 , is that musically undesirable acoustic parameters such as roughness and noise and were perceived to be congruent with information parameters with negative valence such as danger or stress. The findings from the preceding these experiments, alongside other magnitude estimation work by Walker [151, 152] represent the first steps towards exploring the nuances of how data-sound mappings are perceived by listeners, however no such steps have been taken to similarly investigate data-vibration mappings.

Data-vibration mappings, particularly tactons — “structured, abstract, [vibro] tactile messages which can be used to communicate information" [16] (p. 1), have applications in many areas, however listed below are some particular use-cases which were a motivation for researching this area:

- ***Process monitoring —*** In a context where vision and hearing are already in use, there may still be critical information that needs to be transmitted without interrupting vision and hearing, such as a warning with an indication of severity (low, med, high) when monitoring a patient. tactons can provide this without interrupting the other senses.
- ***Feedback in hazardous environments —*** When in a dangerous environment such as a location with high quantities of radiation or unstable ground, tactons can convey information such as distance to a particular location while keeping vision and hearing free for navigation and communication.

- ***Feedback in loud environments —*** Hoggan *et al.* [72] found vibrotactile feedback to be an effective non-visual feedback modality in loud environments. An example of such an environment where tactile feedback may be beneficial could be a helicopter pilot receiving vibration cues relating to the accuracy of orientation for landing.

Such use cases can not only provide some context for why it is necessary to study and understand how data-vibration mappings are perceived, but also highlight the unique potential of tactons.

As mentioned in the previous chapter, both the acoustic parameter and the information parameter in a data-sound mapping, can be discussed in terms of valence. The valence of the information parameter may be quite trivial where there is a clear good or bad connotation of the information, for example — increasing values of danger would have a negative valence, whereas decreasing numbers of cancerous cells could be considered to have a positive valence. The valence of the acoustic parameter is more complex, however the results reported so far would suggest that for many listeners, the desirability of an acoustic parameter in terms of Western musical traditions may offer a guide. For example roughness (disharmony) was associated with increasing values of danger and therefore would be considered to have a negative valence. However, it is unknown if such examples of perceptual congruence exist in data-vibration mappings, as there is a lack of research in this particular aspect of vibrotactile feedback.

As with data-sound mappings, it is not possible to predict what mappings of information parameters to vibrotactile parameters are optimum *a priori*, there are many cognitive, perceptual and cultural factors that may influence how a user expects a particular data value *should* feel when it is presented using vibration. Such a mental model driven preference as found in data-sound displays may be a useful tool in establishing the optimum vibrotactile parameter to convey a particular information parameter in a data-vibration mapping. This chapter presents an experiment, which is a first in the area of vibrotactile research, to investigate a number of data-vibration mappings using a magnitude estimation paradigm to map how the magnitudes of vibrotactile parameters like frequency and duration relate to the perception of the magnitude of a number of information parameters such as error, size and distance. Using magnitude estimation to understand how these data-vibration mappings are perceived can aid in creating data-vibration mappings that are more congruent with a user's expectations and mental models of how a particular information parameter *should* feel to them and therefore make more usable, less confusing vibrotactile displays. The third research question of this thesis addresses this by asking:

> **RQ3:** To what extent are mappings between vibrotactile parameters and information parameters perceived as congruent?

5.2 Experiment 4

This section details the design of the magnitude estimation experiment that was conducted. The design is very similar to that of Experiment 3 that looked at data-sound mappings. In a similar fashion to Experiment 3, this magnitude estimation study provides information regarding the polarities (e.g. is increasing frequency perceived as increasing values of size?) and slope values (e.g. for each value of frequency, what is the perceived value of size?) that are produced for each data-vibration mapping.

5.2.1 Participants

18 participants took part in the study, demographic information gathered included gender (self-identified), age and handedness. Participants were: 9 female, 9 male, mean age = 24.6 years, SD = 5.4 years, 15 right-handed, 2 left-handed, one ambidextrous. All were university students and staff and reported no uncorrected vision impairment and no hearing impairment. These participants were unique to this study and hadn't taken part in any of the previous studies reported in this thesis.

5.2.2 Design

The magnitude estimation experiment was based on the design originally suggested by Stevens [134] and further described by Marks & Gescheider [97]. Applying this design to investigate parameter mapping of tactons was based on the same methodology used in the previous chapter on data-sound mappings, which itself was based on previous experiments which used a similar methodology to investigate parameter mapping of sonifications [151,152]. This method has also been used to study the perception of textures via vibrotactile stimulation [135].

Twenty-eight conditions were examined in which the two independent variables in each condition were the vibrotactile parameter used and the information parameter which the vibration stimuli ostensibly represented. These information parameters were: accuracy, error, stress, danger, size, distance and current and the vibrotactile parameters used were frequency, roughness, duration and tempo (rate of on/off fluctuation of a signal).

Choice of Information Parameters

The information parameters were chosen to provide a general set of potential use cases that future tacton designers may encounter. Also, they were selected in an attempt to provide a set of parameters that may be generalised such that these results could be applied in a variety of situations. For example, size could be used in the context of physical size, or the size of a network. The inclusion of error, stress and danger as information parameters was in part based on Experiment 3 which suggested that information parameters with negative valence are associated with acoustic parameters which are often considered undesirable or negative in a musical context (such as roughness and noise). To investigate if these associations are present in similar vibrotactile mappings these information parameters were included in this experiment, to ask questions such as: does increasing vibrotactile roughness equate to a perceived increase in the magnitude of stress?

5.2.3 Apparatus

The actuator used in this study was a Tactile Labs Haptuator Mk II. [1] This actuator was chosen based both on its high acceleration and bandwidth. The actuator was fixed to the inside wall of a cylindrical steel housing to ensure maximum contact with the participant's hand, as this wall was placed facing upward and thus facing where the participant placed their hand. This housing was filled with sound-proof foam to secure the actuator and reduce audible emissions. The housing was then placed in soundproof foam to reduce vibration of the surrounding equipment and the table. Participants placed their non-dominant palm onto the actuator during the experiment and white noise was played to them over Beyerdynamic DT-100 headphones to mask any audible noise from the vibration actuator.

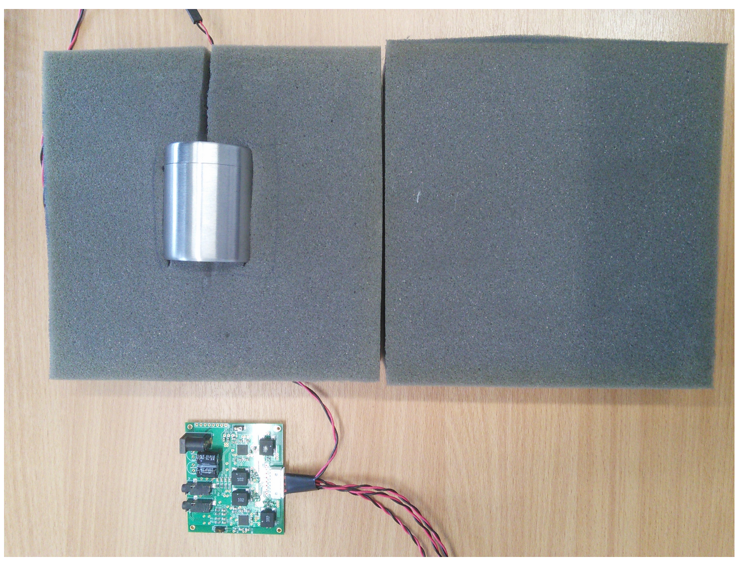

Figure 5.1: Haptuator in steel housing and placed in soundproof foam, with additional foam for participants to rest their forearm.

5.2.4 Stimuli

Three stimuli were used for each condition and were created in the SuperCollider programming language. Stimuli were designed to provide three of the most easily distinguishable cues for each

[1] http://tactilelabs.com/products/haptics/haptuator-mark-ii-v2/

vibrotactile parameter, therefore the cues were designed based on previous perceptual studies on the particular parameter, or related studies that successfully applied them.

Frequency

Stimuli consisted of one second long cues at 90, 200 and 300 Hz. Similarly to audible frequency [121], perception of frequency via the skin is also affected by the intensity of the stimulus [51, 145], therefore each frequency cue's amplitude was adjusted such that intensity was equal across each stimulus. These were based on the stimuli used by Wilson *et al.* [162], as they used the same actuator as this experiment and derived equal-intensity cues for the frequencies used in this experiment.

Roughness

Previous studies have shown that a sensation of roughness can be conveyed via vibrotactile stimuli using sinusoidally amplitude-modulated waveforms [18,70]. Furthermore, these studies have shown that differing levels of amplitude modulation can be perceived when they are used one dimensionally [18] or multidimensionally with other parameters being simultaneously modulated [19]. Roughness cues were one second long consisting of a 200 Hz sine wave modulated at 20 and 50 Hz, as well as an unmodulated waveform. These cues were based on the findings from Brown's investigations into vibrotactile roughness [18] which found that modulation frequencies of 0 (no modulation), 20 Hz and 50 Hz are significantly distinguishable from each other.

Duration

Duration cues consisted of 200 Hz waveforms that were 100, 500 and 1000 ms in duration, also based on [162].

Tempo

Tempo cues consisted of two second long 200 Hz waveforms in which each cue rapidly fluctuated on and off at different frequencies. Each cue's *on* period was 50 ms and the *off* periods were 15, 50 and 800 ms.

5.2.5 Procedure

Before the first block of trials, the experimenter read allowed the following text (adapted from [97, 135, 151]) and the participant read along on-screen.

> You will feel a series of vibrations, one at a time, in random order. Your task is to indicate what magnitude of the concept the vibrations seem to represent, by

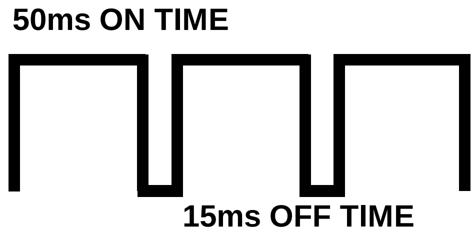

Figure 5.2: Waveform of one stimulus from the tempo condition in Experiment 4 with a 50 ms *on* phase and 15 ms *off* phase

> assigning numbers to them. For the first vibration, assign it any number of your choosing that represents a value of the concept (e.g. distance). Then, for each of the remaining vibrations, estimate its 'distance', relative to your subjective impression of the first vibration.
>
> For example in the case of 'distance' as the concept, if the second vibration seems to represent a distance level that is 10 times as 'close' or 'far' as the first, then assign it a number that is 10 times bigger than the first number. If the vibration seems to represent a distance that is one-fifth as 'close' or 'far', assign it a number that is one-fifth as large as the first number, and so on. You can use any range of numbers, fractions, or decimals that seem appropriate.

Here the first vibration is used as a comparison value, it is the middle level of all the stimuli (level two of 3). The participant indicated the comparison magnitude once at the beginning of the block. In each trial, participants were presented with the comparison vibration with its user-defined value simultaneously displayed on-screen and then after a one second pause, one of the two remaining stimuli were presented. The participant was then asked to respond with a subjective value (i.e. the magnitude of distance they perceived). In a block of six trials, each stimulus was presented three times in random order and after a brief break, the next condition was presented, with new instructions that presented the mapping to be used. The order in which each block was presented was also randomised.

5.3 Results

First, data were split based on polarity response for each data-vibration mapping. Then, as the scale of the responses could vary widely between each participant, responses for each mapping and polarity were normalised using the geometric mean, as was done in the previous experiment and as per Walker [151, 152] and Stevens *et al.* [137]. Consistent with the audio related findings from Walker [151] and Experiment 3, for a given data-vibration mapping most participants responded in a consistent polarity (either positive or negative), such that, for example, increasing vibrotactile roughness represented increasing danger. Some participants responded in an erratic polarity, therefore it was necessary to group the data into *positive*, *negative* or *no* polarity. This was done by calculating the Pearson coefficient for each participant in each block between the log of the responses and the log of the actual stimulus value. A participant's data were considered to have *no* polarity if the absolute coefficient for that block did not reach statistical significance and were removed from analysis.

Therefore, with two stimuli presented three times (df = 6), participant data sets with a correlation coefficient of less than $r_{critical}$ = 0.707 (p = <0.05) were classed as having *no* polarity. Note that the middle level of all the stimuli (level two of 3) was used as a comparison with which to compare the other two levels' perceived magnitude, therefore it is not included in the coefficient calculation. This is a fairly strict requirement due to the low number of stimuli. Experiment 3 reported in the previous chapter and Walker [151] used more generous limits of $r_{critical}$ = 0.367 & 0.444 respectively. Such a strict $r_{critical}$ ensures that the results truly reflect the participants' polarity choices. Table 5.1 shows the number of participants whose data remained in the analysis of each condition, compared to the number of participants who originally carried out the experiment.

	Accuracy	**Error**	**Danger**	**Size**	**Distance**	**Current**	**Stress**	**Total**
Frequency	14 / 18	16 / 18	13 / 18	15 / 18	15 / 18	16 / 18	16 / 18	105 / 126
Roughness	12 / 18	10 / 18	11 / 18	10 / 18	15 / 18	9 / 18	10 / 18	75 / 126
Duration	18 / 18	18 / 18	17 / 18	18 / 18	16 / 18	17 / 18	18 / 18	122 / 126
Tempo	14 / 18	18 / 18	18 / 18	18 / 18	17 / 18	17 / 18	18 / 18	120 / 126

Table 5.1: Proportion of participants whose data was of a high enough collinearity to be used in analysis for Experiment 4.

After this preparation, a simple linear regression was carried out for each data-vibration mapping, using the logs of both the geometric means of responses for each stimulus and the actual stimulus values (as per [151]). Conversely to the analysis for the magnitude estimation carried out in Experiment 3 as described in previous Chapter where the middle comparison value (stimulus level two of three) was not included in the linear regression, here it was due to the fact that at least three values were required to calculate a slope. The slope value was calculated using the geometric mean of all participant responses for a given polarity group within a condition. These mean response values were plotted against the actual stimulus values in log-log coordinates and fitted with the power function $y = bx^m$ where the slope of the fit line m indicates how much the perceived, or estimated magnitude changes as the actual vibrotactile parameter changes [151]. The slope value of this regression indicates how much the perceived value changes based on changes in the stimulus. Figures 5.3 to 5.9 show the results for each data-vibration mapping grouped by information parameter.

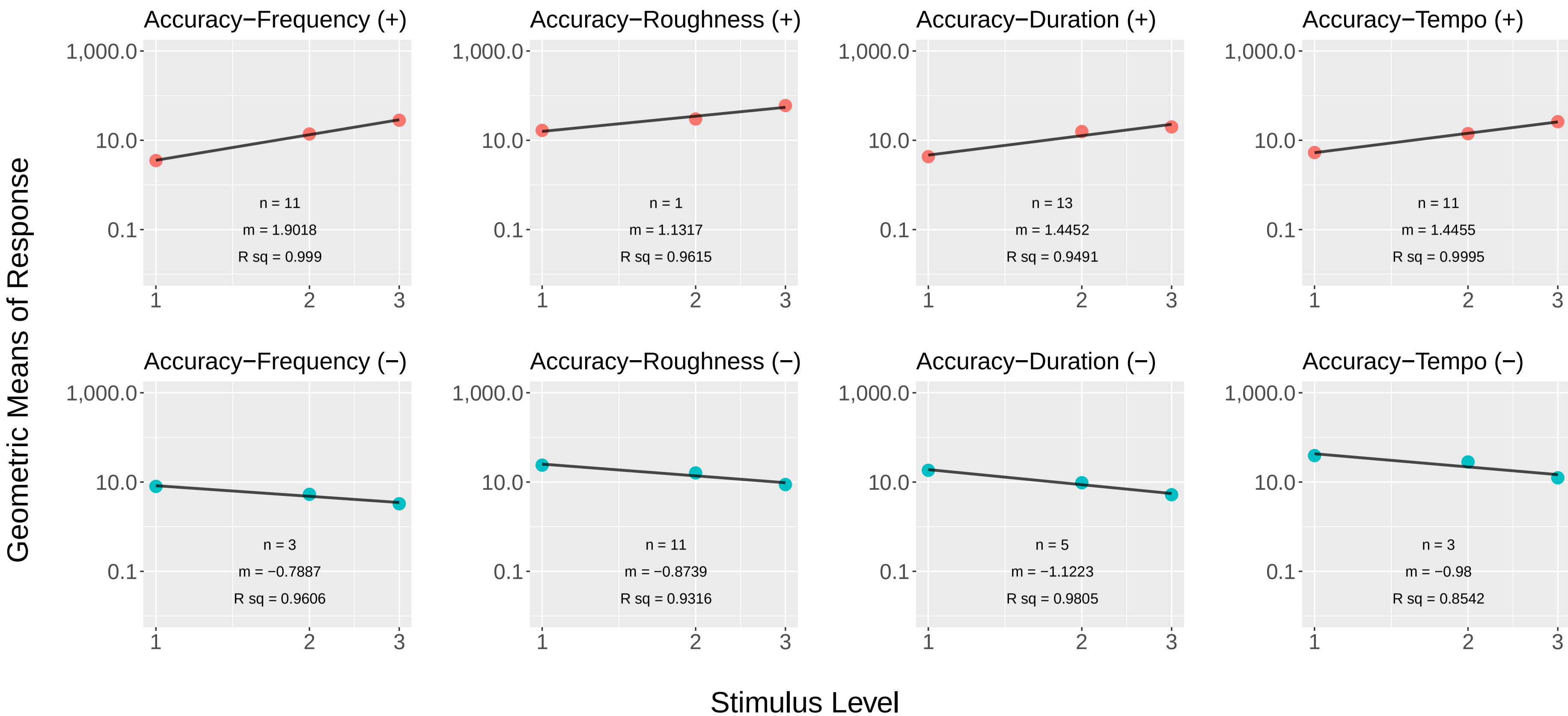

Figure 5.3: Plots for the geometric mean of estimated values against the actual stimulus level for the accuracy conditions in Experiment 4. Axes are in log-log scale. The first row in red shows red polarities and the second row shows the negative polarities. *n* denotes the number of participants that responded in that polarity, *m* denotes the slope value and *R sq* denotes the coefficient of determination (R^2).

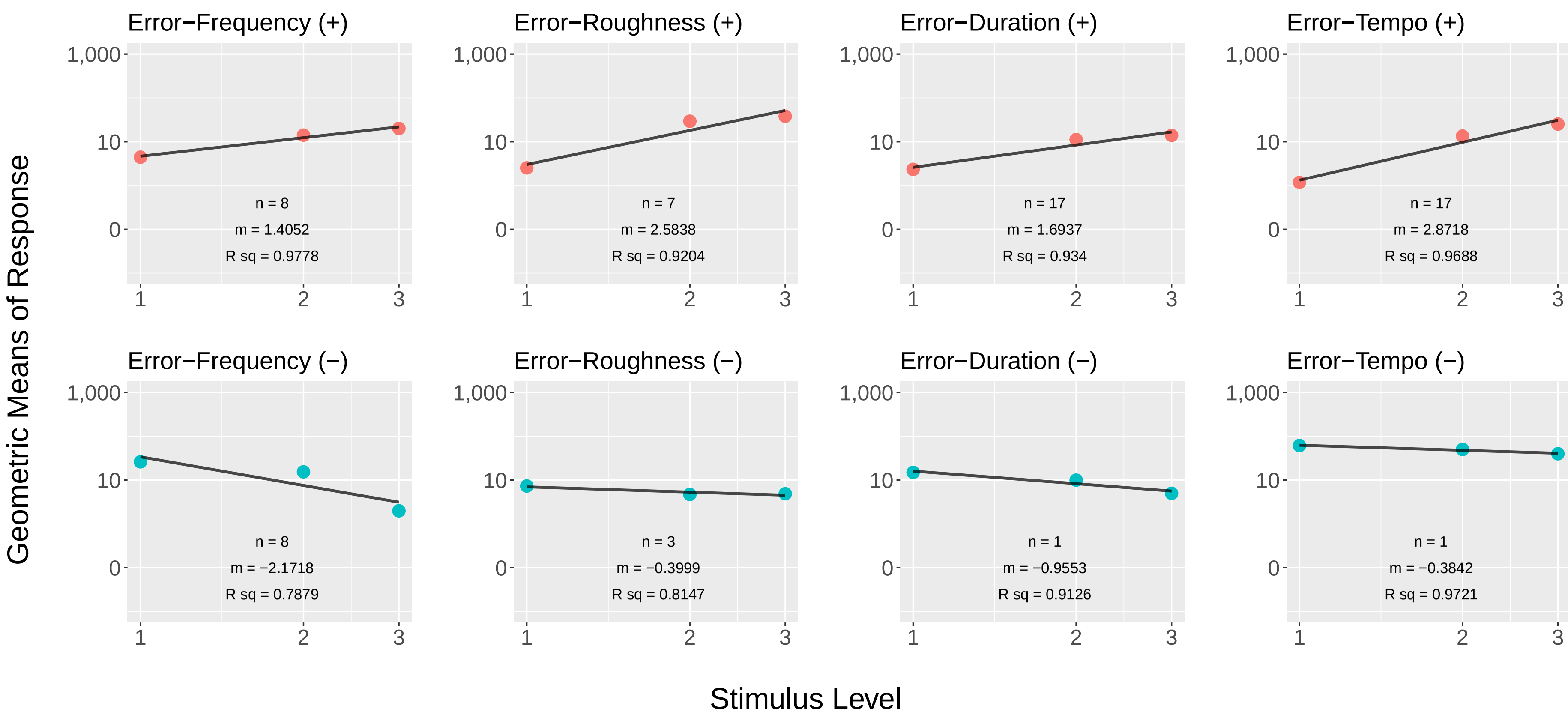

Figure 5.4: Plots for the geometric mean of estimated values against the actual stimulus level for the error conditions in Experiment 4. Axes are in log-log scale. The first row in red shows red polarities and the second row shows the negative polarities. *n* denotes the number of participants that responded in that polarity, *m* denotes the slope value and *R sq* denotes the coefficient of determination (R^2).

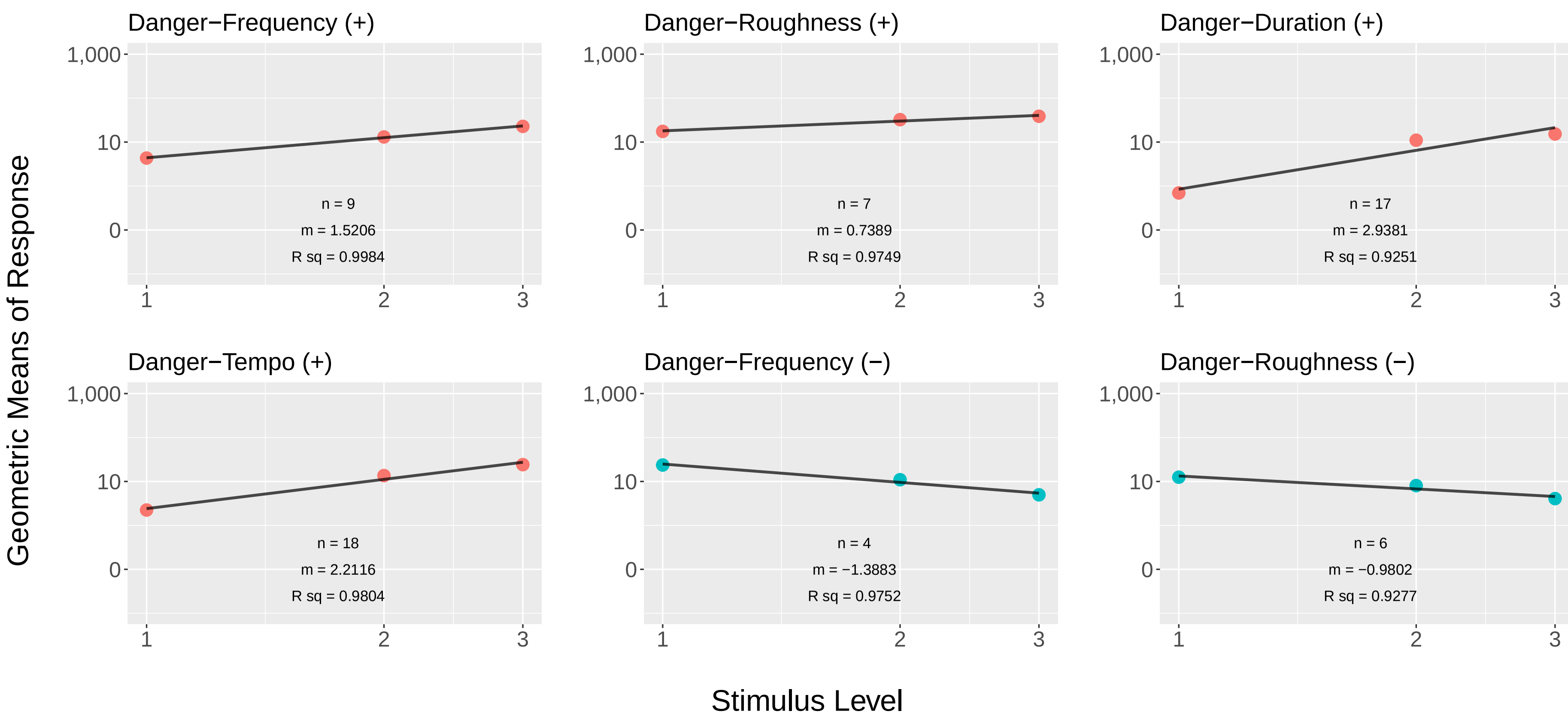

Figure 5.5: Plots for the geometric mean of estimated values against the actual stimulus level for the danger conditions in Experiment 4. Axes are in log-log scale. The first row in red shows red polarities and the second row shows the negative polarities. *n* denotes the number of participants that responded in that polarity, *m* denotes the slope value and *R sq* denotes the coefficient of determination (R^2)

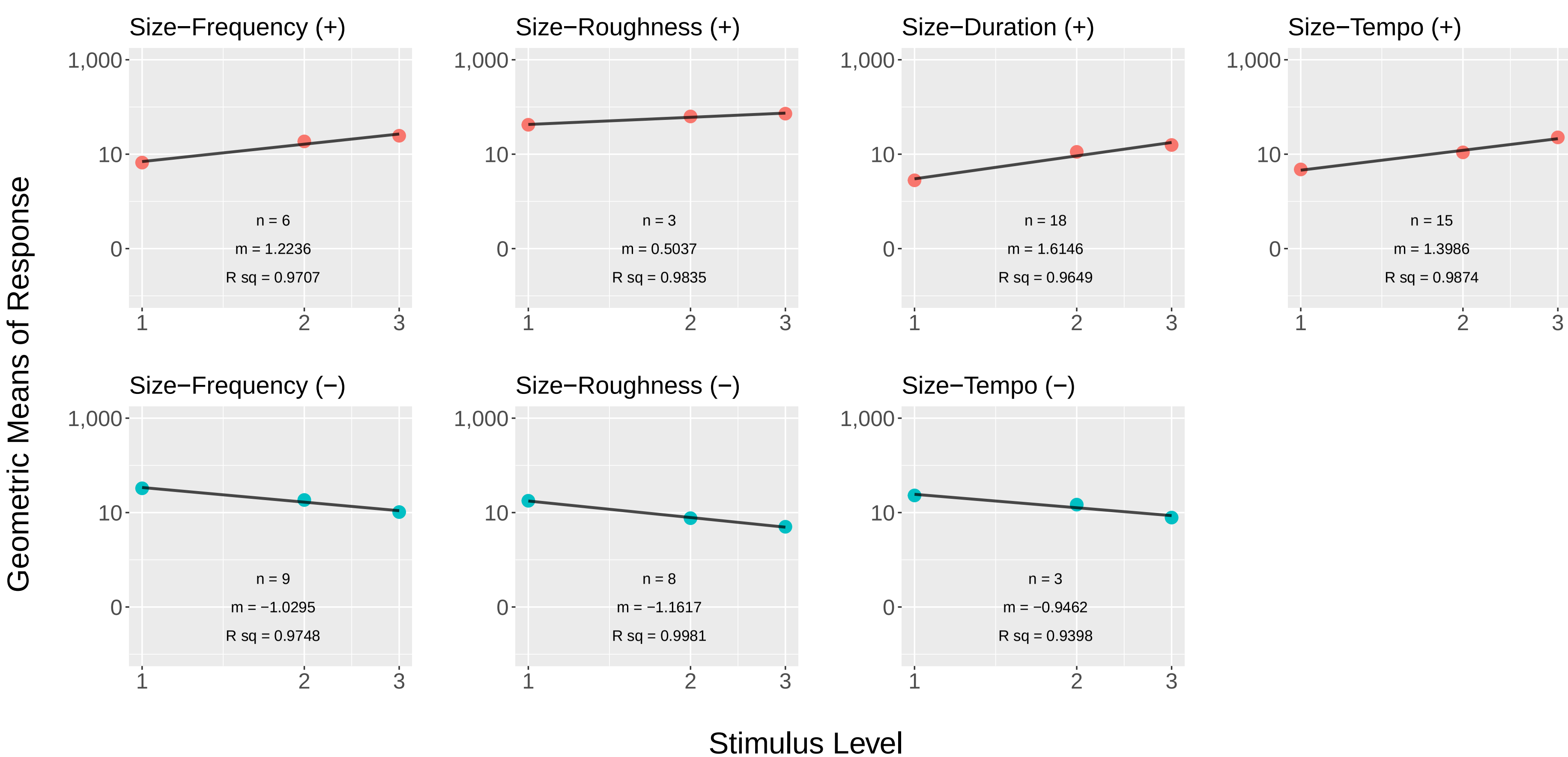

Figure 5.6: Plots for the geometric mean of estimated values against the actual stimulus level for the size conditions in Experiment 4. Axes are in log-log scale. The first row in red shows red polarities and the second row shows the negative polarities. *n* denotes the number of participants that responded in that polarity, *m* denotes the slope value and *R sq* denotes the coefficient of determination (R^2)

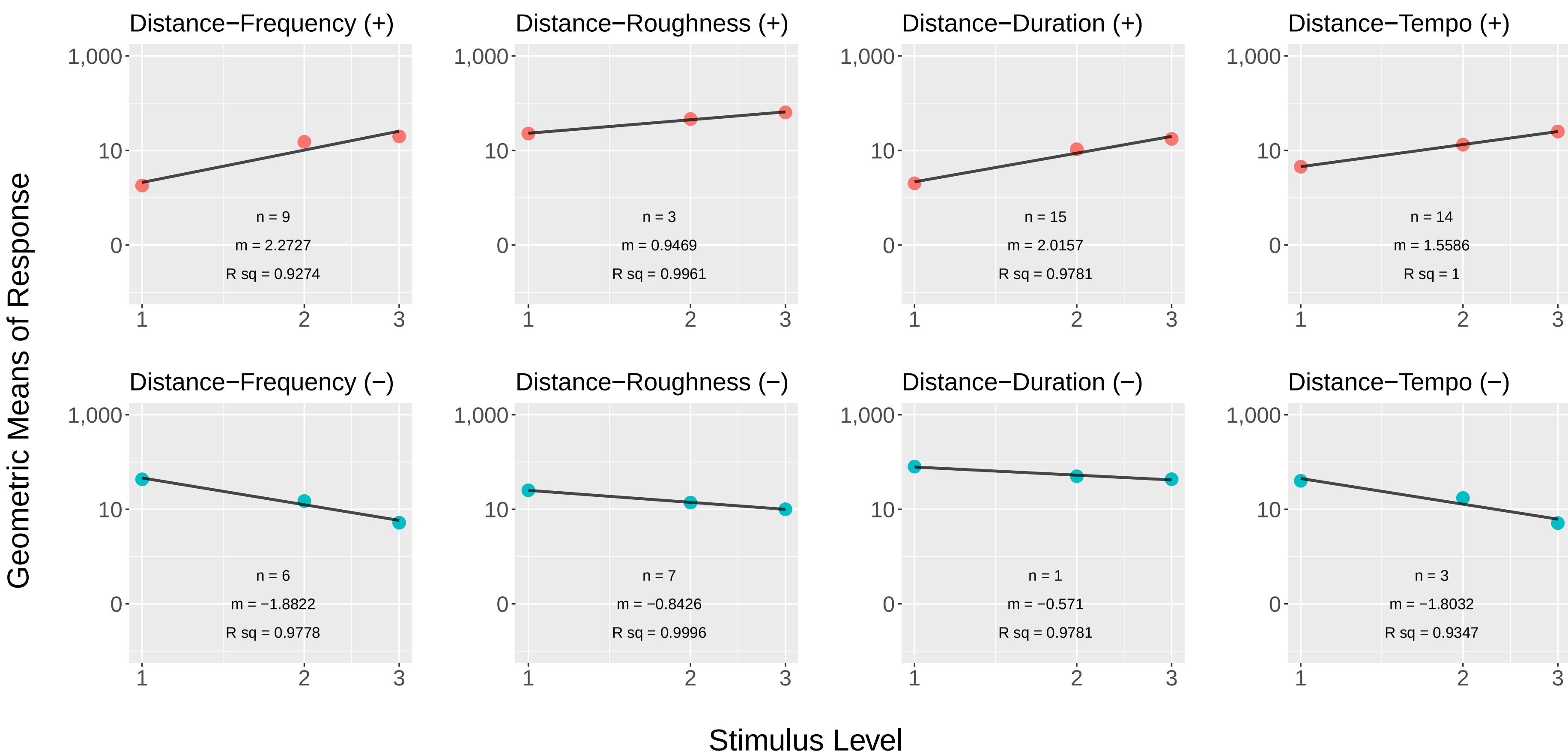

Figure 5.7: Plots for the geometric mean of estimated values against the actual stimulus level for the distance conditions in Experiment 4. Axes are in log-log scale. The first row in red shows red polarities and the second row shows the negative polarities. *n* denotes the number of participants that responded in that polarity, *m* denotes the slope value and *R sq* denotes the coefficient of determination (R^2)

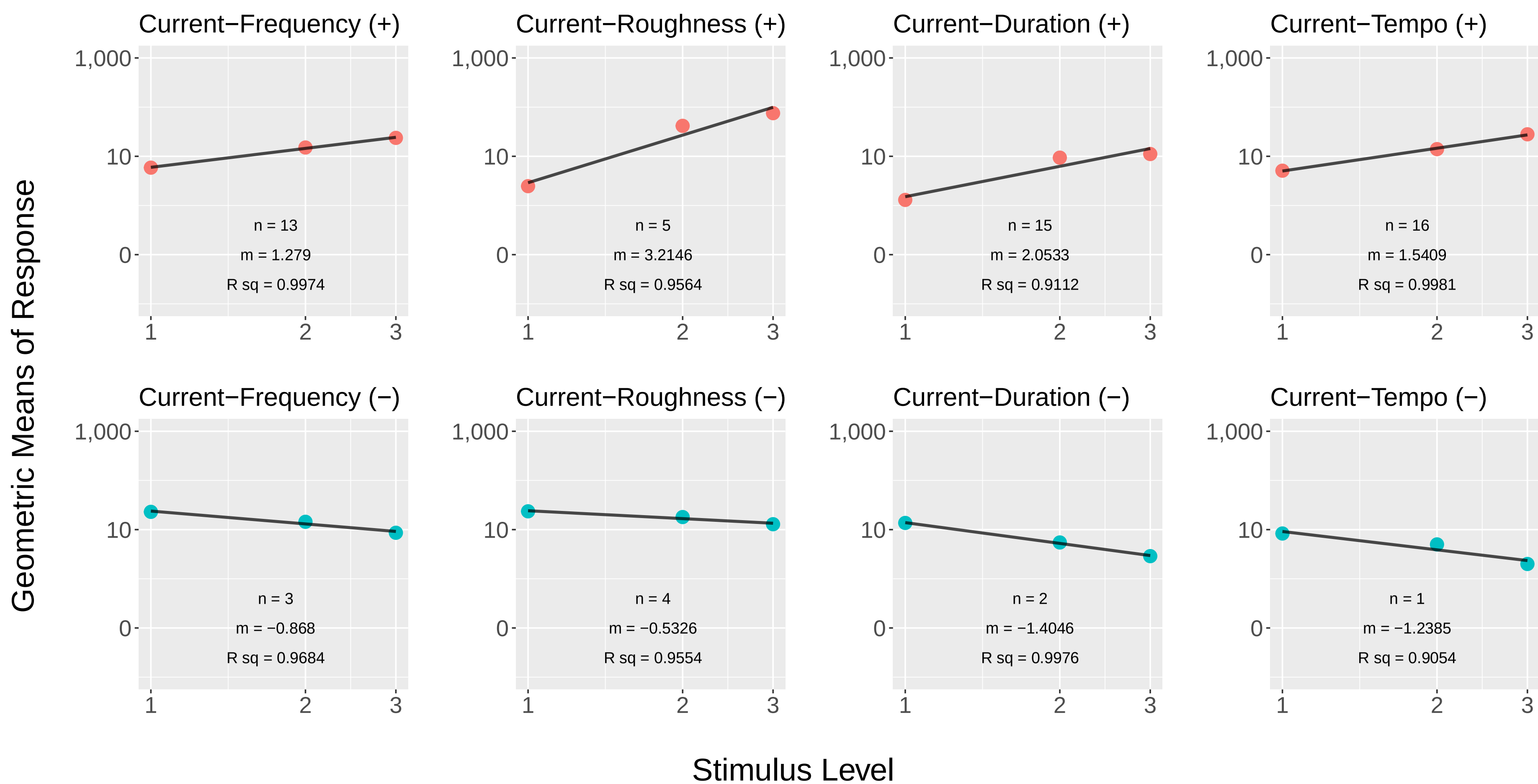

Figure 5.8: Plots for the geometric mean of estimated values against the actual stimulus level for the current conditions in Experiment 4. Axes are in log-log scale. The first row in red shows red polarities and the second row shows the negative polarities. *n* denotes the number of participants that responded in that polarity, *m* denotes the slope value and *R sq* denotes the coefficient of determination (R^2)

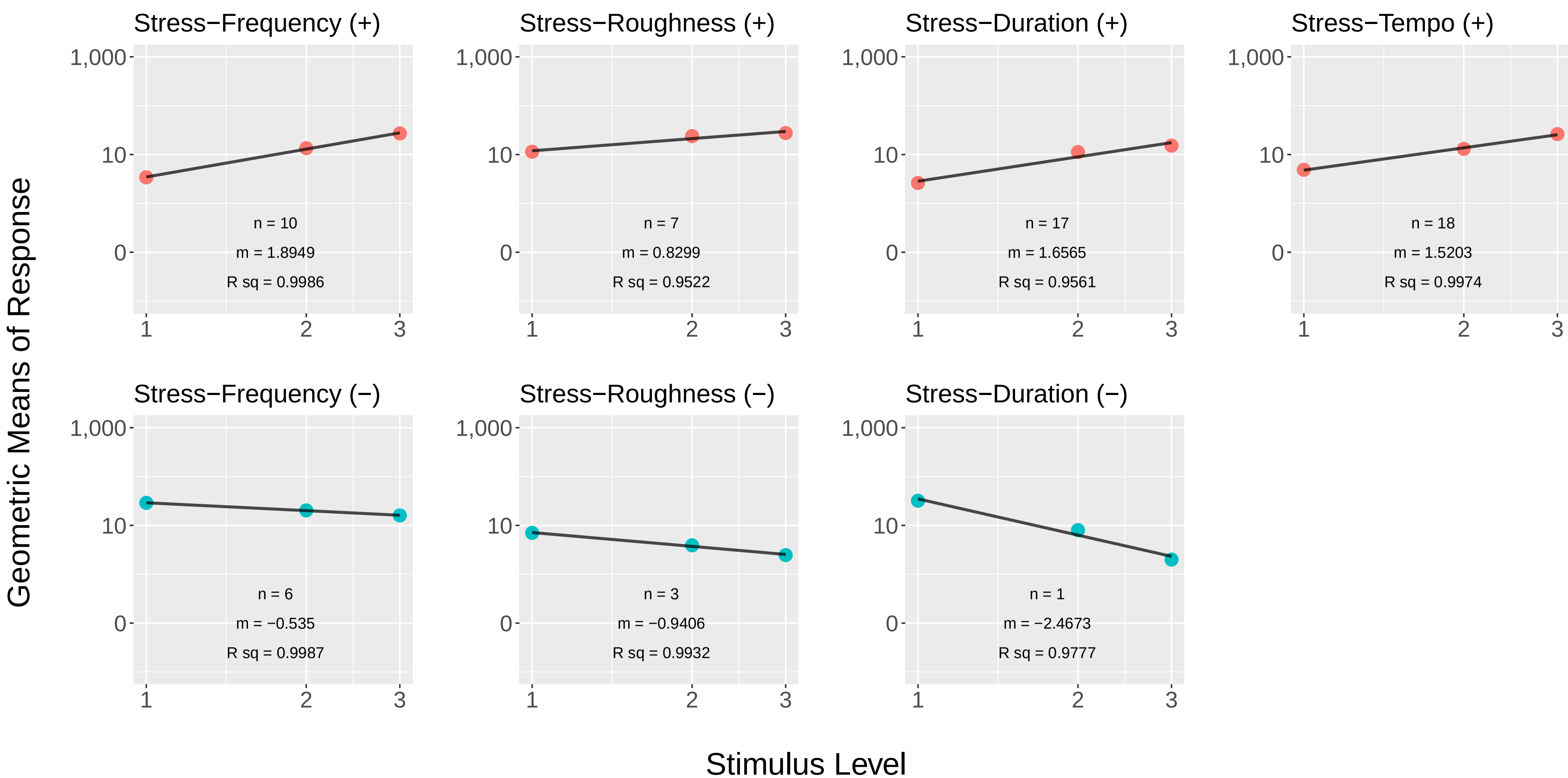

Figure 5.9: Plots for the geometric mean of estimated values against the actual stimulus level for the stress conditions in Experiment 4. Axes are in log-log scale. The first row in red shows red polarities and the second row shows the negative polarities. *n* denotes the number of participants that responded in that polarity, *m* denotes the slope value and *R sq* denotes the coefficient of determination (R^2)

5.4 Discussion

Polarity — Frequency and Roughness Conditions

First, Table 5.1 shows one of the most obvious findings is that the duration and tempo mappings resulted in more participants responses reaching or passing the $r_{critical}$ = 0.707 threshold for significance (122 and 120 out of a possible 126 respectively) than the roughness and frequency conditions, with the roughness conditions having a lower number of participants reaching significance (75 out a possible 126). Looking at the split in participants' polarity preferences in terms of positive or negative polarities, frequency and roughness conditions show a higher mixture of positive and negative polarity consensus, such as the size-frequency mapping being split $n = 6$ for positive and $n = 9$ for negative and the current-roughness mapping being split $n = 5$ for positive and $n = 4$ for positive. These unclear polarity preferences may indicate that there was a lot of confusion when choosing a polarity for this mapping and that they were difficult to process or to relate to a participant's mental model of how they thought size or current, to use the two mappings picked as examples, should feel when presented in a data-vibration mapping.

Polarity — Duration and Tempo Conditions

Conversely, participants responded with a much clearer consensus in the duration and tempo conditions, with the danger-duration, danger-tempo, and size-duration mappings obtaining unanimous responses from participants, here meaning that they all responded in a positive polarity. Aside from that unanimous polarity preferences, for all mappings participants received the most popular polarity was positive, suggesting that for the mappings studied here, increasing the duration or tempo of a tacton are both perceived as an increase in the magnitude of an information parameter, regardless of the concept. These results are particularly interesting for the accuracy and error information parameters. These concepts are the opposite of each other in terms of valence, accuracy having positive valence and error having negative valence, therefore it would be reasonable to assume *a priori* that their polarity results would generally reflect this opposition, however these results show their polarities and slope values to be similar. The duration and tempo conditions also both contain two unanimous polarities (i.e. meaning one or more polarity group's n was 0), lending further weight to the suggestion that increasing these vibrotactile parameters results in increased perception of the magnitude of a information parameter. The duration and tempo conditions also contain four mappings in which all 18 participants response reached the $r_{critical}$ suggesting there was little confusion amongst participants in choosing which polarity to respond. Particularly, the error, size and stress condition across both polarities obtained responses above $r_{critical}$ for all 18 out of a possible 18 participants, suggesting there was little confusion amongst participants in choosing which polarity to respond.

These results show that the polarity and perceived magnitude of information parameters have less consensus among participants and thus more confusion about how to process some mappings (frequency, roughness), or it may be consistent across mappings (duration, tempo), suggesting a more clear understanding of processing the mapping and therefore further underlining the importance of understanding data-vibration mappings in vibrotactile feedback design. Furthermore, as can be seen from the conditions where tempo and duration were used to convey both error and accuracy, even though these two concepts are opposing in terms of valence, they both elicited a positive polarity preference among participants, contrary to the findings from the previous chapter's experiment on the auditory modality. This suggests that using opposing feedback parameters to represent opposing concepts in terms of valence, as may be the *a priori* expectation, may not be the optimum method for conveying these types of data using vibration. Using a magnitude estimation paradigm to estimate the polarities and slopes, like the one used in this study can guide designers towards mappings that are consistent with the user's perceptions of how a information parameter *should* feel when it is conveyed through a data-vibration mappings.

Slope

Similar to the results in the previous chapter about magnitude estimation results for data-sound mappings, these results for data-vibration mappings show that how a user perceives the magnitude of an information parameter varies depending on the data-vibration mapping used, participants were presented with the same vibrotactile stimuli but their responses varied depending on the information parameter being used in the given data-vibration mapping. Again, this echoes the findings from the auditory modality such as: the previous chapter, Walker [151] and Walker & Kramer [155] in that the mapping topology used in a tacton system has a significant effect on how the user interacts with it. Using a magnitude estimation paradigm to estimate the polarities and slopes for a data-vibration mapping can aid designers in creating mappings that are more consistent with user mental models of how an information parameter should feel when it presented using vibrotactile feedback in a tacton system.

5.5 Conclusions

This chapter investigated the effects of vibrotactile parameter choice on a number of data-vibration mappings for tactons. A study was presented in which magnitude estimation was used to map how the perceived magnitude of an information parameter changed based on a change in a vibrotactile parameter used in a tacton. Polarity and scale information was derived for each mapping. Results from this experiment indicated that the vibrotactile parameter used to convey an information parameter affected participants' perception of the magnitude of the information parameter. Overall, the mappings that used duration or tempo as the vibrotactile parameter in the data-vibration mapping resulted in more clear agreement about which polarity was most preferred among participants. Also for these vibrotactile parameters, all mappings were perceived in a positive polarity, even though the valence of some mappings were opposing (accuracy and error). This suggests that for mappings using these vibrotactile parameters, increasing the vibrotactile parameter leads to a perceived increase in the information parameter, regardless of the valence of the information parameter. These findings highlight the importance of understanding user perceptions of data-vibration mappings in order to design tactons that are effective and to reduce the potential for users to be confused or frustrated by them. These results can be applied in a number of contexts: either by augmenting devices with external haptic actuators like the ones used in this study, or using consumer hardware like video game and virtual reality controllers, smartphones and wearable technologies as more powerful and flexible embedded vibration actuators are becoming more common in such devices.

5.5.1 Research Question 3

The study presented in this chapter presents a first step in understanding data-vibration mappings from a psychophysical perspective, by using magnitude estimation to establish polarity and slope information for several mappings. The main outcomes of this study are summarised and presented in the form of the following design recommendations that address research question 3 which asked:

> **RQ3:** To what extent are mappings between vibrotactile parameters and information parameters perceived as congruent?

Duration and Tempo tactons Are Not Mapping Dependent

Results from this study suggest that if using duration or tempo to convey a information parameter through a tacton, the perceived magnitude of the information parameter will increase as duration or tempo increases, regardless of what the particular information parameter may be.

Information Parameters with Opposing Valence Are Not Necessarily Perceived in Opposing Polarities

The results for the duration and tempo mappings showed both accuracy and error polarity choices to be almost totally positive polarities. This suggests that information parameters which have opposing valence may not necessarily be perceived in opposing vibrotactile polarities. From these results it can be suggested that designers test their mappings when tactons are used for representing information parameters with opposing valence, to ensure that the vibrotactile feedback is designed such that the user can understand the intended meaning of each tacton.

Chapter 6

Evaluating the Magnitude Estimation Approach

6.1 Introduction

The previous two chapters each reported results from magnitude experiments which obtained polarity and slope information for a number of data-sound (Experiment 3) and data-vibration mappings (Experiment 4), which gave an insight into how those mappings were perceived by users. These studies have further highlighted a significant challenge in tacton and PMson design: mapping information parameters onto acoustic or vibration parameters in a way that aligns with a listener's mental models of how a given data parameter should sound or feel. These studies used the psychophysical scaling method of magnitude estimation to systematically evaluate how participants perceive mappings between data and sound parameters, giving data on perceived polarity and the scale of the relationship between the data and sound parameters.

These studies and others [151, 152] have shown that magnitude estimation can be a useful predictor of how a data-sound or data-vibration mapping may be perceived in that it tells the researcher how ambiguous the mapping polarity is amongst participants. However, there has been no research investigating the effect of using results from these studies has in an actual task in which the mapping is applied. This chapter describes an experiment which investigates the effect that using parameter mapping polarities based on results from prior magnitude estimation experiments has on the performance of a simple auditory and vibrotactile feedback task. In this experiment the parameter mapping polarities based on results from prior studies are compared with using mappings with arbitrary polarities. These arbitrary polarities in this case were the inverse polarity of the most popular one derived from the magnitude estimation results. The goal of this study is to establish to what extent the data from magnitude estimation experiments are generalisable when used in actual auditory or vibrotactile display tasks. The fourth research question of this thesis aim to address this by asking:

RQ4: How does the complexity of the usage context impact the effectiveness of a data-sound or data-vibration parameter mapping?

Chapter 7 further investigates this problem by looking specifically how the complexity of the context of the auditory or vibrotactile display impacts the user's performance. Therefore, this chapter and Chapter 7 can be considered as a pair, as they are both addressing RQ4. These research questions will be concluded at the end of Chapter 7.

6.2 Experiment 5

An experiment was conducted to investigate if using mapping polarities based on results from a prior magnitude estimation experiment has any effect on performance during a simple task that used auditory and vibrotactile display. In this experiment, a task consisting of ranking three WiFi networks based on their security level was used. The level of security for each network (low, medium, high) was conveyed using either an auditory or vibrotactile cue. Results were obtained on correctness of these rankings and the time it took participants to complete them. The data-sound mappings and the polarities were based on results from Experiment 3 and the data-vibration mappings were based on results from Experiment 4, here using the mappings of *danger-roughness* and *danger-noise* for the data-sound mappings and *danger-duration* and *danger-tempo* for the data-vibration mappings. In this context danger meant how insecure or dangerous a WiFi network may be. The magnitude estimation study for the data-sound mappings found that when noise was used to represent danger, 13 of 15 participants perceived this mapping in a positive polarity (increasing noise = increasing danger). Similarly, when roughness was used, 12 of 13 participants responded in a positive polarity. The magnitude estimation study for the data-vibration mappings found that when duration was used to represent values of danger, 17 of 17 participants responded in a positive polarity and when tempo was used, 18 of 18 participants responded in a positive polarity. Therefore these acoustic and vibrotactile parameters were chosen, based on the clear consensus of polarity preference they elicited from participants.

6.2.1 Participants

Twenty-four participants took part in the study. Participants were: 12 female, 11 male, 1 non-binary, mean age = 28.2 years, SD = 6 years, 23 right-handed, 1 left-handed. All participants reported no uncorrected vision impairment and no hearing impairments.

6.2.2 Design

Eight conditions were investigated in which the two independent variables were the acoustic or vibrotactile parameter (acoustic roughness, acoustic noise, vibrotactile tempo and vibrotactile duration) used and the polarity in which it was mapped to the security of the network. The polarity of each mapping was either based on results from Experiment 3 or 4 or was the inverse of these polarities, meaning the polarity was inverted such as increasing roughness = increasing danger would become increasing roughness = decreasing danger. Moving forward from here, all polarities based on results from the previous chapters will be referred to as *aligned polarities* and the 'arbitrary' polarities, as *inverted polarities*. Therefore, four acoustic and vibrotactile parameters presented in aligned and inverted polarities resulted in eight conditions (4 parameters $\times$ 2 polarities = 8 conditions). The dependent variables collected during the experiment were completion time, correctness of responses and NASA Task Load Index (TLX) [62]. The experiment used a within-subjects design. For this experiment, it was important to consider the order in which participants were presented with each polarity, as a participant may favour whichever polarity they were exposed to first, thus reducing the quality of the data gathered. Therefore, counterbalancing was used to account for this.

6.2.3 Stimuli

Roughness and noise were used as acoustic parameters and duration and tempo as vibrotactile parameters in this study, based on the stimuli used in Experiments 3 and 4. The roughness and noise parameters were chosen due to the effect of roughness on the perception of danger [5] and noise's effect on the perception of image focus (see Experiments 1 and 2) (with lack of focus or bluriness also being an undesirable data concept or could be said to have negative valence). Duration and tempo were chosen as vibrotactile parameters because in Experiment 4, they elicited consistent polarity preferences and, compared to the other vibrotactile parameters used in that study (frequency and roughness), the duration and tempo conditions elicited more responses that reached statistical significance.

The auditory magnitude estimation experiment in Experiment 3 used ten levels for each acoustic parameter, an example being the noise condition in that study containing ten sound cues ranging from a clean tone to white noise, whereas the vibrotactile magnitude estimation study in Experiment 4 used three levels. Therefore, for this task, the number of levels for the acoustic parameters was reduced to three to both ensure that there were equal numbers of stimuli for the auditory and vibrotactile conditions and to ensure that the task would be simple and the sound cues could easily distinguished from each other. As in Experiment 3, each auditory stimulus was two seconds in length. Each stimulus had an amplitude envelope with a 0.2 second linear ramp onset (attack) and offset (release). An amplitude envelope was included in the sound design, as an abrupt start or stop of a sound can be perceived as unpleasant [7]. The vibrotactile parameters were identical to Experiment 4. All stimuli were created in the SuperCollider programming language. The design of each stimuli is described below.

Roughness

This condition consisted of a 1 kHz sine tone that was 100% amplitude modulated with modulation frequencies of 0 (i.e. an unmodulated sine tone, 11 and 70 Hz.

Noise

This condition consisted of a 1 kHz sine tone for the first level, and equal blend of a pure tone and broadband and white noise for the second level and the final level was solely broadband noise.

Duration

Duration cues consisted of 200 Hz waveforms that were 100, 500 and 1000 ms in duration, based on cues used by Wilson [162].

Tempo

Tempo cues consisted of two-second long 200 Hz waveforms in which each cue rapidly fluctuated on and off at different rates. Each cue's *on* period was 50 ms and the *off* periods were 15, 50 and 800 ms.

6.2.4 Procedure

The experiment consisted of eight blocks with each parameter (roughness, noise, duration, tempo) mapped in each polarity (*aligned*, *inverted*). Each block consisted of three trials. At the beginning of each condition, participants were presented with a screen which explained the acoustic or vibrotactile parameter being used in that block and how it was mapped to each level of network (Figure 6.1). In this screen, participants could use three buttons to hear the sound cues for each level of security for the given condition's mapping. Participants could not press the continue button until the button for each level of security was pressed at least once, however they could play each sound or vibration cue as many times as they needed. In each trial, participants were presented with a screen showing three WiFi networks, each with a button to play the cue to convey their level of security (Figure 6.2).

Sounds were presented using a pair of Beyerdynamic DT100 headphones. The participants were tasked with rearranging the networks based on their security (top = most secure, bottom = least secure) as conveyed by the sound cues. Similar to the first screen, participants had to play each stimuli at least once before being able to submit their ranking, however there was no upper limit to how many times they could repeat each sound or vibration cue. For each

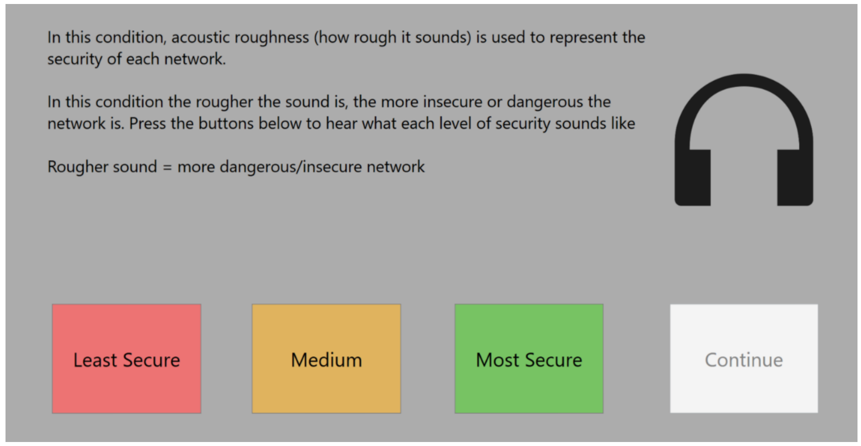

Figure 6.1: Condition introduction page where the mapping is explained, including the polarity. Three coloured buttons allow the participant to hear the sound for each level of security.

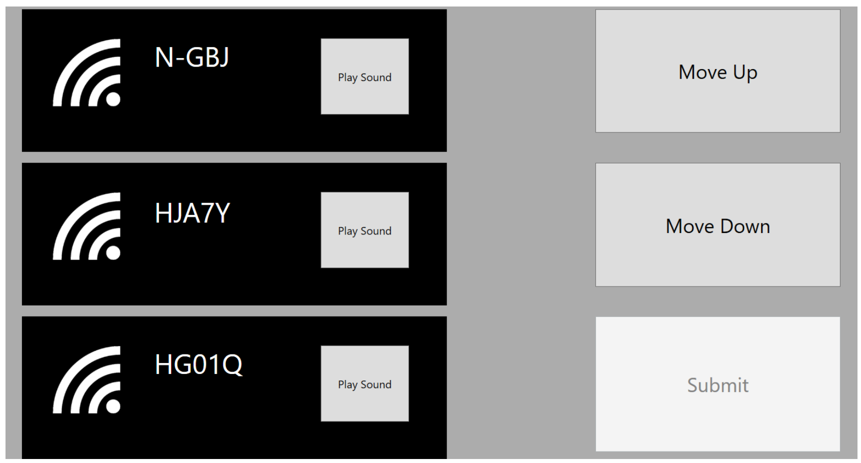

Figure 6.2: Screen for each trial showing three networks, each with a button to play their associated sound stimuli.

condition participants completed three rankings, each time the ordering and name of the networks were randomised. The randomisation was implemented such that the network ordering was always mixed, therefore the ordering was not correct at the beginning of each trial and thus ensuring that the participant had to rearrange the networks. At the beginning of the experiment participants completed a practice condition to familiarise them with the experiment. This practice condition was identical in procedure to all the subsequent conditions, except participants were required to rank the networks based on colour rather than using sound or vibration cues. Sound and vibration was omitted from the practice to ensure that participants were not influenced in any way which may have an effect their future responses. At the end of each condition, participants completed the NASA Task Load Index Questionnaire.

6.3 Results

As can be seen from Figure 6.3, which shows boxplots of ranking completion times for each participant, there is a large number of outliers in participant response times that do not reflect the participant's overall performance (for example participant 8 in Figure 6.3 has two outliers which are more than twice as slow as their mean response time). As the main goal of this study was to investigate the difference between the two polarity types, not specifically how well each participant performed, these observations were removed to reduce these outliers capacity to skew the subsequent analysis. any observation beyond 1.5 times the interquartile range threshold was removed. This resulted in the removal of 29 observations from a total of 576 equating to 547 total observations. Figure 6.4 shows the reaction times for each participant after this removal of outliers.

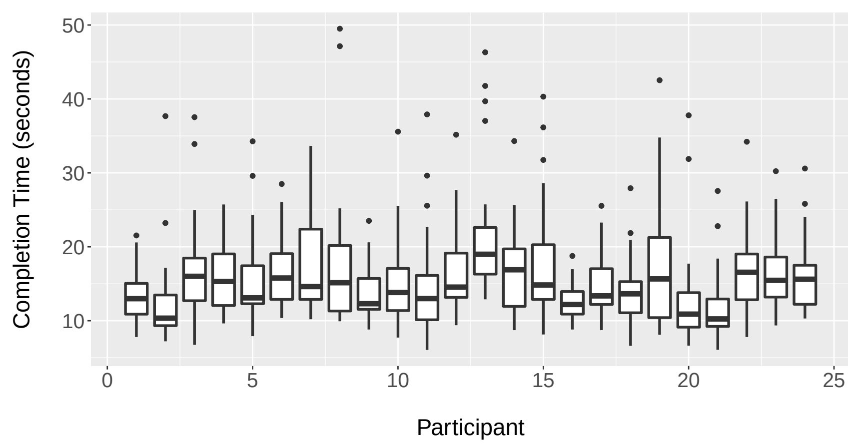

Figure 6.3: Boxplots showing completion times for each participant before outlier removal in Experiment 5.

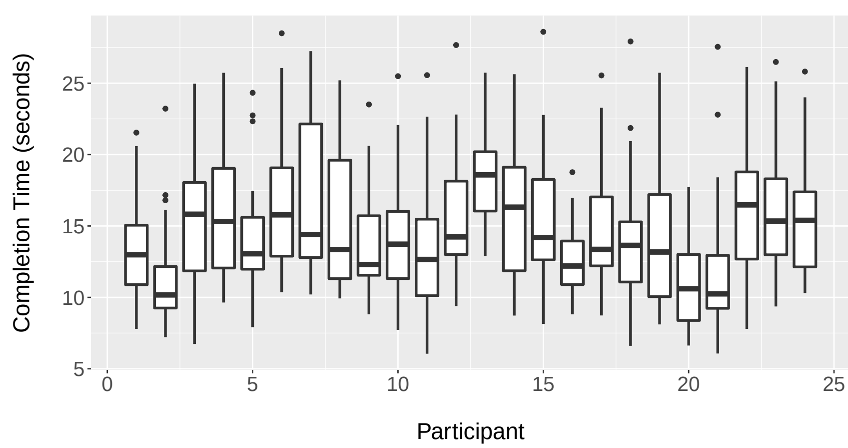

Figure 6.4: Boxplots showing completion times for each participant after outlier removal.

6.3.1 Completion Time

First, since each ranking was completed three times, the mean completion time was calculated for each participant, for each sound/polarity combination. Looking at each mapping individually, ANOVAs investigating the effect of polarity on completion time showed no statistically significant difference between the completion times for both polarities in all conditions: roughness ($p = 0.245$, Table 6.1), noise ($p = 0.495$, Table 6.2), duration ($p = 0.651$, Table 6.3) and tempo ($p = 0.535$, Table 6.4). Figure 6.5 shows boxplots for each condition's completion times, for each polarity.

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Polarity	1	13348388.67	13348388.67	1.39	0.2447
Residuals	46	442282314.22	9614832.92		

Table 6.1: ANOVA results for the effect of polarity on completion time in the roughness conditions of Experiment 2.

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Polarity	1	4358355.76	4358355.76	0.47	0.4947
Residuals	46	423197665.97	9199949.26		

Table 6.2: ANOVA results for the effect of polarity on completion time in the noise conditions of Experiment 2.

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Polarity	1	2507606.42	2507606.42	0.21	0.6506
Residuals	46	554956950.70	12064281.54		

Table 6.3: ANOVA results for the effect of polarity on completion time in the duration conditions of Experiment 2.

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Polarity	1	4.18	4.18	0.39	0.5347
Residuals	46	491.81	10.69		

Table 6.4: ANOVA results for the effect of polarity on completion time in the tempo conditions of Experiment 2.

In this case, attempting to gain a more thorough insight into how little polarity choice impacted the participants' completion time may be advantageous. For example, it may be useful for an auditory or vibrotactile interface designer to understand whether carrying out a resource-intensive magnitude estimation experiment to evaluate a particular data-sound or data-vibration mapping is necessary for their particular use-case. Therefore, effect sizes were calculated (using the recommendations set out by the Transparent Statistics in Human-Computer Interaction

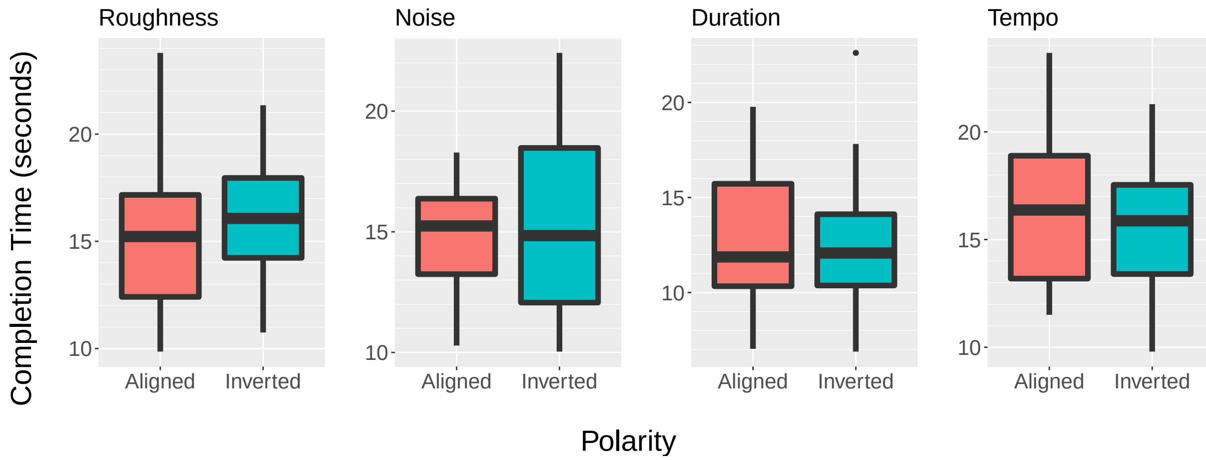

Figure 6.5: Boxplots for ranking completion times for each condition and polarity in Experiment 5.

Group [76]). To investigate this, a t-test was used to find the estimated difference in the mean completion time between the *aligned* and *inverted* polarities. Cohen's *d* test was also used to estimate effect sizes. In the roughness condition the inverted polarity had a greater mean completion time than the aligned polarity by 1055ms (95% CI: [-2858, 749]), Cohen's $d = 0.34$. In the noise condition the inverted polarity had a greater mean completion time than the aligned polarity by 603ms (95% CI: [-2376, 1170]), Cohen's $d = 0.2$. In the duration condition the aligned polarity had a greater mean completion time than the inverted polarity by 457ms (95% CI: [-1561, 2475]), Cohen's $d = 0.13$. Finally, in the tempo condition the aligned polarity had a greater mean completion time than the inverted polarity by 590ms (95% CI: [-1310, 2491]), Cohen's $d = 0.18$.

6.3.2 Correctness of Rankings

To understand how correctly participants completed the ranking task for each condition, the number of incorrect rankings was calculated for each mapping/polarity pair. An incorrect ranking here is any submitted ranking in which at least one of the WiFi networks was ranked in the incorrect position. Of all 547 rankings completed throughout the entire study, only 50 were incorrect (around 91% correct). Since this variable is count data and count data tends to be non-normally distributed and highly positively skewed, a standard analysis of variance may not be an effect method of analysis. Poisson regression is a more appropriate form of analysis and is commonly used for count data [30, 75, 91], therefore this was used in the analysis of this data. Table 6.5 shows these results. The roughness conditions were the only ones which showed significant results, with the aligned polarity resulting in significantly more incorrect rankings than the inverted polarity.

Condition	Polarity	Total Incorrect Rankings	χ^2	*p*
Roughness	*Aligned*	17	4.3	0.0382*
	Inverted	7		
Noise	*Aligned*	2	2.94	0.086
	Inverted	7		
Duration	*Aligned*	6	2.09	0.148
	Inverted	2		
Tempo	*Aligned*	4	0.11	0.739
	Inverted	5		

Table 6.5: Number of incorrect rankings for each parameter and polarity in Experiment 5. χ^2 shows the chi-squared value from the Poisson regression and p indicates the p-value of the regression.

6.3.3 NASA TLX Workload

A Wilcoxon signed rank test found no statistically significant differences between NASA Task Load Index overall workloads for both polarities in the roughness and noise conditions (Table 6.6 shows TLX results).

Condition	Polarity	Mean (Workload)	SD (Workload)
Roughness	*Aligned*	22.6	16.7
	Inverted	23.8	18.8
Noise	*Aligned*	18	16.6
	Inverted	11.5	12.3
Duration	*Aligned*	13.3	10.4
	Inverted	15.5	13.5
Tempo	*Aligned*	14.7	13.6
	Inverted	16.9	18.1

Table 6.6: Summary of NASA Task Load Index results from Experiment 5 including mean workload and standard deviations.

6.4 Analysis

The Cohen's *d* values and estimated differences between polarities for both acoustic and vibrotactile mappings reported in section 6.3 is surprising. It would be a reasonable expectation *a priori* to assume that the mapping design based on results from a previous magnitude experiment would result in a faster completion time. This is because the polarity which participants preferred in the previous study could be assumed to have lower cognitive load, as there is less confusion about what they mapping conveys.

The largest estimated differences found in all the conditions is 1.055 seconds, which was found in the roughness conditions, with all other conditions having estimated differences in the means of less than a second. For many applications this very small difference may be acceptable, meaning that carrying out a magnitude estimation experiment to gather polarity data may not be necessary in some cases. In order to further explore this, *Akaike Information Criterion* was used to investigate whether a model in which the completion times for both polarities are *equal* is more representative of the data gathered from this study, than a model in which the completion times for each polarity is assumed to be different. The following section introduces this method and describes its application to the results from this experiment.

6.4.1 A Brief Overview of the Akaike Information Criterion

As this method is uncommon in auditory and vibrotacitle display literature (with the notable exception of Frohmann et al. [47]), a brief overview of the process of using it is given. Hirotugu Akaike's information criterion (*AIC*) [2] is a method to estimate the relative quality of statistical models for a given data set. AIC estimates the amount of information lost when data is fitted to a given model, thus when comparing two potential models, the model with the lowest information loss is more representative of the data in question. Null hypothesis testing cannot allow acceptance of the null hypothesis (in this case being "the completion times for mappings that are based either on polarity data from a prior experiment or inverted polarities are equal") however AIC can be used to reframe the question to be "is there more support for a model in which the completion times for both polarities are equivalent than one in which they are not". The following equation is used to estimate the AIC of a model [2,21]:

$$AIC = 2k - 2\log(\hat{L}) \tag{6.1}$$

where k is the degrees of freedom and $\hat{L}$ is the maximum value of the likelihood function of the model. To quantify the quality of each model, the raw AIC score must be converted to weighted scores. The first step is to calculate the differences in AIC for each model with the respect to the AIC of the best candidate model [21,22]:

$$\Delta_i(AIC) = AIC_i - AIC_{min} \tag{6.2}$$

where AIC_{min} is the minimum of the AIC values. This transformation causes the best model to have $\Delta_i(AIC) = 0$, while the rest of the models have positive values. The next step is to establish the relative likelihood L for each model i given the data

$$L(M_i|data) \propto exp\{-\frac{1}{2}\Delta_i(AIC)\} \tag{6.3}$$

where $\propto$ denotes "is proportional to". Finally, the relative likelihoods for each model are normalised to obtain weighted AIC scores for each model (w_i). Here each model's relative likelihood is divided by the sum of the likelihoods of all other models being compared, like so:

$$w_i(AIC) = \frac{\exp\{-\frac{1}{2}\Delta_i(AIC)\}}{\sum_{k=1}^{k} \exp\{-\frac{1}{2}\Delta_k(AIC)\}} \tag{6.4}$$

Finally, the weighted AIC scores can infer the best fitting model. For example, if two models A and B are being compared, with weighted AIC scores of $w_a(AIC) = 0.6094$ and $w_b(AIC) = 0.2242$, their weighted AIC scores would be used to show that model A is around 2.7 times more likely to be a better fit for the data than model B. [1]:

$$\frac{w_a(AIC)}{w_b(AIC)} = \frac{0.6094}{0.2242} = 2.7$$

6.4.2 Applying AIC to the Current Experiment

As discussed in the prior sections, we use the Akaike Information Criterion to answer the question: Is there more support for a model in which the completion times for both polarities are equivalent than one in which they are not? Or, put more simply, this is assuming that the polarities have absolutely no effect on completion times.

First, two models are fitted: a linear model in which the completion times for both polarities is forced to be equivalent and effectively treating the data as if there were only one polarity category (henceforth written as *Equivalent Model*), and a linear model in which they are assumed to be not equal (here named *Unequal Model*). By using the process described in the previous section, the quality of these models can be compared. Table 6.7 shows the results of the AIC analysis.

Condition	**Model**	**log($\hat{L}$)**	*AIC*	ΔAIC	$w_i(AIC)$	$\frac{w_a(AIC)}{w_b(AIC)}$
Roughness	*Equivalent*	-454	911.4	0	0.571	**1.332**
	Unequal	-453	922	0.573	0.429	
Noise	*Equivalent*	-452	908.3	0	0.68	**2.126**
	Unequal	-451	909.8	1.508	0.32	
Duration	*Equivalent*	-459	921.1	0	0.709	**2.44**
	Unequal	-458	922.9	1.784	0.291	
Tempo	*Equivalent*	-456	915.5	0	0.689	**2.218**
	Unequal	-456	915.1	1.593	0.311	

Table 6.7: Summary of results from AIC analysis in Experiment 5.

[1]example taken from [150]

6.5 Discussion

From the AIC analysis results in Table 6.7 we can see that for all conditions, the Equivalent Model, i.e. the one in which the completion times for each polarity are assumed to be equivalent, is the most parsimonious for the given data, that is to say the better fit of the two models. The AIC results in combination with the small effect sizes reported earlier support the argument that for this task, the polarity of the data-sound or data-vibration mapping did not have a substantial effect on the time it took participants to complete the task. For the auditory mappings, these results are surprising as the previous magnitude estimation experiments carried out in Experiment 3 showed that nearly all the tested participants perceived *increasing* noise or roughness as *increasing* danger, therefore it was expected that these mappings would result in faster completion times due to reduced cognitive effort needed to process the mapping. However, this was not found to be true, the AIC analysis showed that the Equivalent Model, that is a model assuming there is *no* difference between the aligned and inverted polarities, was a better fit for the data than the Unequal Model which assumed each polarity was different. The AIC analysis for the roughness condition found the equivalent model to be 1.332 times more likely to be a better fit to the data than the unequal model. Similar results were found for the noise condition where the Equivalent Model was 2.126 times more likely to be a better fit to the data than the Unequal Model. Furthermore, the roughness condition elicited significantly more incorrect rankings in the aligned polarity than in the inverted polarity, which is again incongruous with the results from Experiment 3. As roughness was found to be representative of danger in this study, it would follow that this mapping, when used here, would result in at least a similar number of incorrect rankings to the inverted polarity if not less incorrect rankings.

A similar picture appears for the vibrotactile parameters, where results from Experiment 4 showed that increasing the duration or tempo of a tacton were perceived as being representative of an increase in danger. In this study, however, the aligned polarity, that is, the polarity which was most popular in Experiment 4, did not result in better performance over the inverted polarity in terms of completion time, correctness or perceived workload. Similar to the data-sound mappings the AIC analysis for the duration condition found the equivalent model to be 2.44 times more likely to be a better fit to the data than the unequal model, and the same for the tempo condition (2.218 time more so). Overall, these findings suggests that for simple displays using these acoustic or vibrotactile parameters such as the application used here, the polarity in which the information parameter is mapped to the acoustic or vibrotactile parameter does not have a substantial effect. This means that for designers working in a similar space, the expenditure of resources to carry out a magnitude estimation experiment to establish polarities may not be necessary.

6.6 Limitations

The generalisability of data-sound and data-vibration mappings polarities obtained from magnitude estimation studies like Experiments 3 and 4 and previous work by other researchers such as [151, 152] has yet be fully determined. This study provided insight into the generalisability of polarities obtained for two data-sound mappings: *danger-roughness* and *danger-noise* and two data-vibration mappings- *danger-duration*, *danger-tempo*, but it is only a first step toward understanding how generalisable data from these magnitude estimation experiments are in practice.

6.6.1 Difficulty of the Task

The primary limitation of this study is that participants could potentially work through a ‘bad’ data-sound or data-vibration mapping because of all 547 rankings completed throughout the entire study, only 50, or 9%, were incorrect. For example, even if a participant thinks that a more natural representation of increasing danger for them is using increasing roughness to convey this increase, they may still be able to complete the task quickly using a conflicting representation (i.e. increasing danger conveyed by *decreasing* roughness) due to the simplicity of the task. The task was intentionally designed to be easy to carry out, both to account for participants who may be new to the notion of an auditory or vibrotactile display and so that investigation into this area could began with a simple auditory and vibrotactile display. Many situations where such displays are commonly encountered are quite simple such as mobile phone notifications or in-car displays, etc. To reflect that in this study, a simple display was used before moving onto more complex auditory and vibrotactile feedback systems.

6.6.2 Specificity of Context

This experiment focused solely on data-sound and data-vibration mappings conveying danger, specifically the danger posed by an insecure WiFi network. The previous magnitude estimation studies in Experiments 3 and 4 presented danger in general and contextually agnostic terms. Therefore it may be useful for future works attempting to investigate the generalisability of polarities gathered from magnitude estimations to evaluate multiple contexts for a given data-sound mapping. For example, a sonification of a value of danger in terms of WiFi security may be perceived quite differently than a much more severe context such as a process-monitoring sonification system in a nuclear power station. Therefore, evaluating a broader range of contexts may afford a more well-rounded view of how generally polarities and scales from magnitude estimation experiments may be applied.

6.6.3 Number of Levels

The mappings used in this study used only three levels to ensure that the task was simple. However, this led to the problem that no matter which polarity was used, the middle stimulus was going to be the same for both polarities. This meant that participants only had to make a binary choice, reducing the overall difficulty of the task.

6.7 In Defence of Magnitude Estimation

The findings from this study must be taken with the qualifications that this analysis focuses only on quantitative measures like completion time and correctness and does not consider other factors in assessing what an ideal mapping would be such as aesthetic preferences. This could mean, for example, that the mapping polarities derived from Experiment 3 or 4 are more preferable, but participants are able to carry out the task effectively even if the mapping polarity is not their preferred one. Therefore, this is not to say that there is no place for the use of the magnitude estimation paradigm in the PMson and tacton mapping design process; it may be that it becomes important when tasks and situations are more complex. In such situations, the ability to understand how a PMson or tacton is processed at a low perceptual level may be advantageous, because misunderstanding or becoming frustrated by the mapping design of an auditory or vibrotactile cue used in such a system could result in dire consequences if the use-case is safety critical. In summary, the findings from this study show that users can satisfactorily conduct a simple task even if the mappings used in the task are designed in contrast to findings from a magnitude estimation experiment. However, this does not necessarily speak to more complicated or safety-critical systems where accurate perception may be critical.

6.8 Conclusions

The research presented in this chapter presents a first attempt to investigate the effect of designing data-sound parameter mapping polarities based on data from magnitude estimation experiments. A study was presented in which the time it took participants to complete an auditory or tactile display based ranking task using two data-sound mappings in a simple auditory display task was compared between two mapping designs: one mapping in which the polarity was based on results from a previous magnitude estimation experiment and one mapping in which the polarity was inverted. From this experiment the Akaike Information Criterion was used to show that the polarity of the data-sound mappings did not have a significant effect on the time it took participants to complete a ranking. Finally, limitations of this study were discussed in addition to some strategies that may address them in the following chapter. This work represents a first step toward researching how data obtained from magnitude estimation experiments can be appropriately applied in real-world sonification tasks and results from this study underline the need for further research in this area. With this in mind, the following chapter reports a study involving a more complex task, using a similar auditory and vibrotactile display but in a more cognitively demanding situation, to investigate how this affects mapping use. As the subsequent chapter also addresses RQ4, this question will be discussed at the end of the next chapter to combine the discussion.

Chapter 7

Parameter Mappings Under Cognitive Load

The previous chapter reported Experiment 5 which presented a first attempt at investigating the impact of using magnitude estimation based mappings. Findings suggested that for simple auditory or vibrotactile display tasks, whether the polarity was derived from magnitude estimation preferences or was arbitrary did not significantly affect user performance. However, this study was limited as it used a simple task of ranking three WiFi networks based on their security. Auditory and vibrotactile displays may be most useful for tasks where the visual modality is already engaged, such as during surgery [8, 42, 94], controlling air-traffic [120] or driving [129]. Crucially, no research has looked at data-sound or data-vibration mappings in a divided-attention setting which may be more representative of these use-cases.

The study presented in this chapter addresses the limitations of Experiment 5 by measuring how the polarity of a mapping impacts classification of audio or vibration cues whilst under higher cognitive load. Results from Experiment 5 suggest that for simple contexts, the polarity has little effect on performance. However, this has not been explored in a more ecologically valid context where attention is divided between some cognitively demanding task and receiving non-visual information. Furthermore, this has not been explored for data-vibration mappings. Understanding how polarity impacts the usability of a system using data-sound or data-vibration mappings under different levels of cognitive load can inform further research and design of mappings that have the user's perception at the forefront. The study presented in this chapter, in tandem with Experiment 5 address Research Question 4:

> **RQ4:** How does the complexity of the usage context impact how effective a data-sound or data-vibration parameter mapping is?

7.1 Experiment 6

A study was conducted to investigate the impact of the polarity of a data-sound and a data-vibration mapping on its ability to convey information, when the user is under two different degrees of cognitive load. This study used a simple working memory task with two difficulties to induce low and high levels of cognitive load on participants. Simultaneously, participants were tasked with interpreting and classifying auditory and vibrotactile alarms. These alarms were simple parameter mappings where the sensory parameter (e.g. acoustic noise) was mapped to four levels of danger and when participants are presented with an alarm, they must respond with what level the alarm represents.

As in Experiment 5, the polarity of these mappings were either based on results from Experiments 3 and 4 (here called *aligned* polarities) or were the inverse of these results (here called *inverted* polarities). For example, in Experiment 3, the most popular polarity choice for a mapping between acoustic noise and danger was a *positive* polarity, meaning the more noise present in an audio signal, the higher the perceived value of danger. Whereas, the inverse of this polarity would be the *less* noise present in an audio signal, the higher the perceived value of danger. The reason for using the inverse of the polarities found to be most popular in previous works is to investigate a 'badly designed' parameter mapping, allowing an investigation of the impact that the polarity of a mapping has on how well the mapping can be understood and used.

7.1.1 Manipulating Cognitive Load

To induce cognitive load a visual N-back task [100] with two levels of difficulty ($N = 0$ and $N = 2$) was used. This is a common paradigm for inducing cognitive load in neuroimaging [111] and has been successfully applied to human-computer interaction studies [57,85]. In the N-back task, participants are presented with a series of symbols such as letters or numbers, one at a time. On each presentation, participants are tasked with responding with whether the current stimulus matches a stimulus that they saw N presentations back in the sequence. For example, in the 2-back version of the task, participants must indicate whether the current number they see matches the one they saw two trials ago. Figure 7.1 shows a visual representation of the process for the 2-back task. For the 0-back task, participants need to compare each stimulus with the one that they saw first in the sequence, therefore this is a matching task, compared to the 2-back task where participants' working memory is being constantly updated. The advantage of the N-back task is that the perceptual and motor demand remains constant across difficulty levels.

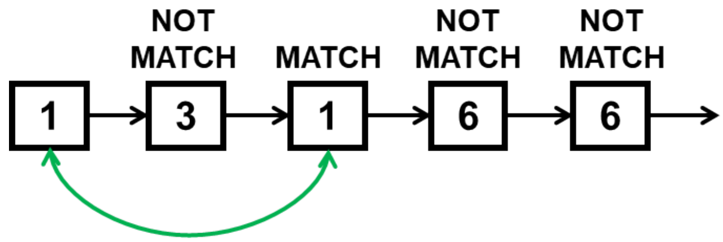

Figure 7.1: Visual representation of the 2-back N-back task used in Experiment 6.

7.1.2 Participants

Sixteen new participants took part in the study (8 female, 7 male and 1 non-binary; mean = 28.4 years, SD = 5 years, 15 right-handed and 1 left-handed). All participants reported no uncorrected vision impairment and no hearing impairment.

7.1.3 Design

Sixteen conditions were examined in which the three independent variables were the audio or vibration parameter used (acoustic roughness, acoustic noise, vibration pulse tempo and vibration duration), the polarity of the data-sound or data-vibration mapping (*aligned* or *inverted*) and the level of the N-back task (0-back or 2-back). Therefore four acoustic and vibrotactile parameters presented in aligned and inverted polarities, presented in combination with two N-back levels resulted in 16 conditions (4 parameters $\times$ 2 polarities $\times$ 2 N-back levels = 16 conditions). Dependent variables were: accuracy of response to alarm classification, accuracy of response to N-Back score and reaction times. Additionally, the NASA Task Load Index Questionnaire (TLX) [62] was used to collect subjectively perceived workload for each condition. The study used a within subjects design. In this study, the order in which polarities were presented could have a confounding effect, as participants may be biased towards whichever polarity they were presented with first. Therefore, the order that each polarity-acoustic/vibration parameter pairing was presented was counterbalanced.

Danger was chosen as the information parameter to convey using the acoustic and vibration stimuli for two reasons: first, there is a number of potential use cases in which some danger could conceivably be represented through some auditory or vibrotactile alarm, such as: hospitals, surgery, in-car, in-aircraft or process monitoring. Second, the results from Experiments 3 and 4 contain polarity data for several data-sound and data-vibration mappings where danger was used as the information parameter in the mapping.

7.1.4 Stimuli

The acoustic and vibrotactile parameters used were based on previous magnitude estimation studies which obtained polarity preference data for a number of data-sound (Experiment 3) and data-vibration (Experiment 4) mappings. Each parameter had four levels. The stimuli in Experiment 5 used only three levels, this led to the problem that no matter which polarity a mapping was in, the middle stimulus was the same for both polarities. This means that participants only had to make a binary choice. Therefore, four levels were used here to ensure participants had to make a more nuanced classification to better test the mapping strategy.

7.1.5 Acoustic Parameters

Roughness and noise were chosen as acoustic parameters, based on results from Experiment 3 which evaluated the use of these parameters in an auditory display. Roughness quantifies the subjective perception of rapid amplitude modulation of a sound, ranging from modulation frequencies of 15 Hz to 300 Hz with the maximum level of perceived roughness being at 70 Hz [174]. More commonly, acoustic roughness is related to the sensation of dissonance or a sound being ‘out of tune’ [79]. Stimuli were two seconds long and included an amplitude envelope with a 0.2 second linear onset (attack) and offset (release). Stimuli were created in the SuperCollider programming language and were designed as follows:

Roughness

Cues were 1 kHz sine tones, amplitude modulated at 0, 7, 23 and 70 Hz. These frequencies were chosen based on results from Experiment 2 that established ten levels of roughness for use in an auditory display. Here the first, fourth, seventh and tenth levels of that range were used to ensure a maximum difference between cues. It is worth noting again that the roughness levels which have been used in all experiments so far are based on Zwicker & Fastl’s work [174] where they established a range of *perceptually equally distributed* levels for roughness. So, even though 0, 7, 23 and 70 Hz are unevenly distributed numerically, according to Zwicker & Fastl’s findings, they are equally distributed in terms of listener perception of roughness.

Noise

This condition was a 1 kHz sine tone and white noise, mixed together to varying degrees, with level one being purely a sine tone and level four being purely white noise with the levels in between being blends of the two signals, Table 7.1 shows the amplitude levels for each stimulus level. Similar to roughness, the range of stimuli used here was extracted from the range of ten evaluated in Experiment 2

	Sine Tone (A)	**White Noise (A)**
Level 1	100%	0%
Level 2	60%	30%
Level 3	30%	60%
Level 4	0%	100%

Table 7.1: The amplitude (A) of the sine tone and broadband noise for each of the four stimulus levels used in Experiment 6.

Vibrotactile Stimuli

The duration and pulse tempo of a vibration cue were chosen as vibrotactile parameters, based on results from Experiment 4. In that previous magnitude estimation experiment, three vibration cues were used and here the range was adjusted to provide four cues. As with the acoustic parameters, the vibrotactile parameters were designed in SuperCollider and were designed as follows.

Duration

Duration cues were 200 Hz waveforms that were 100, 500, 1000 and 2000 ms in duration.

Tempo

Tempo cues were also 200 Hz waveforms which pulsed on and off at varying frequencies. The *on* period for all cues was 50 ms and the *off* periods were: 15, 150, 400 and 800 ms.

7.1.6 Procedure

First, the research was explained to the participants, then consent and demographics information were gathered. Each block consisted of a combination of acoustic or vibrotactile parameter, polarity and N-back level. The ordering of the acoustic or vibrotactile parameter and polarity combination was counterbalanced and for each of these combinations, participants completed both the levels of the N-back before continuing to the next combination. The N-back level ordering was randomised. For example, if the current combination was roughness in an aligned polarity, the participant would complete both the 0-back and 2-back tasks for that combination before moving on to the next. Participants were presented with each stimulus twice.

At the beginning of each condition participants were presented with a screen which explained the acoustic or vibrotactile parameter that was being used in that block and how it was mapped to levels of danger (Figure 7.3). On this screen participants could press buttons to hear/feel the cues for each level. They needed to hear/feel each level at least once before continuing, but could repeat the cues as many times as they wished. Furthermore, before each N-back task, participants were shown a screen telling them which *N*-back level they were about to undertake (Figure 7.2). At the end of each *N*-back task, participants filled in a NASA TLX questionnaire. For the sake of readability, the specific procedures for the *N*-back task and the audio/vibration alarm classification task will be discussed separately as follows. Before the experiment trials began, participants carried out a practice trial of the alarm classification and *N*-back task separately, followed by a practice trial of the two combined.

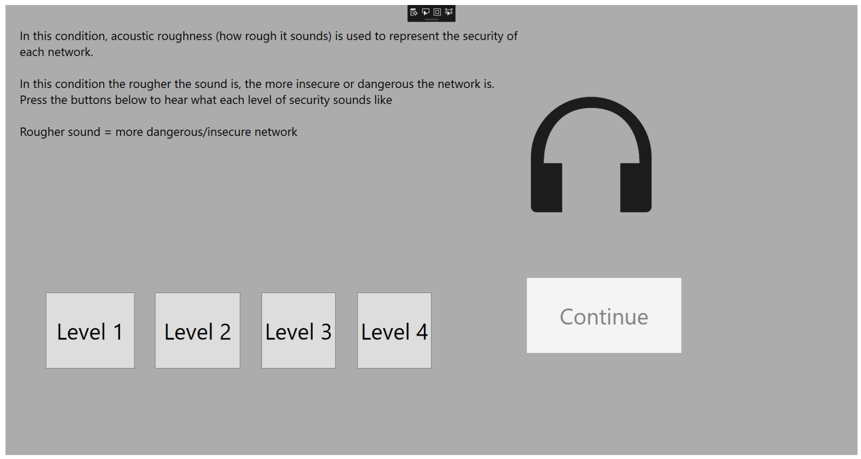

Figure 7.2: Condition introduction page where the mapping for the upcoming trial is explained, including the polarity. The buttons at the bottom of the page allow the participant to hear the sound for each level of security.

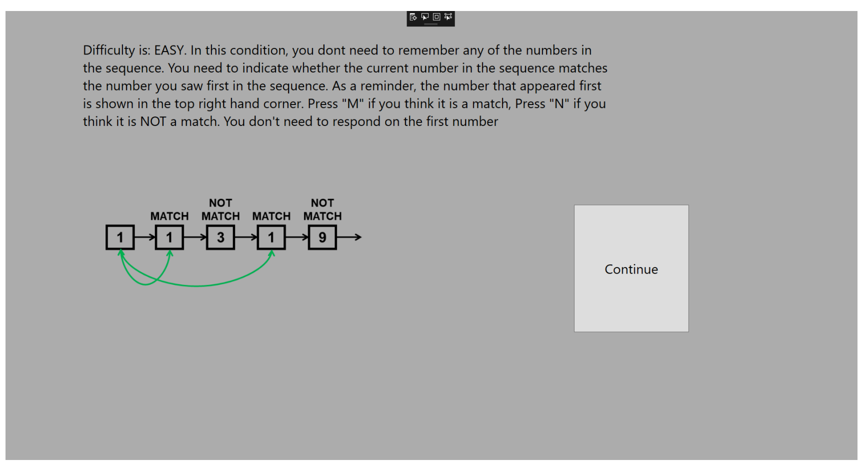

Figure 7.3: Screen where the participant is explained what *N*-back task they will need to complete in the upcoming trial

N-back task

In this study, the range 1-9 was used for the displayed symbols which appeared in random order. Each number was displayed in the centre of the screen for one second and then the screen went blank for three seconds before the next number in the sequence appeared (as per [57]).

Participants indicated via a keypress if a given number on-screen was a match or non-match. Each *N*-back task had a sequence which was 21 numbers long with the constraint that one-third of these sequence (7 trials) were matches.

Audio and Vibration Alarm Classification Task

Concurrently with the *N*-back task, the audio and vibration alarms were presented every ten seconds. The time at which they were presented within that ten second window was randomised (Figure 7.4). There was a one second gap at the start and end of each stimulus to ensure there was a minimum interval between stimuli. To ensure that participants had enough time to process and respond to the alarm, they were given *at least* five seconds from the onset of the cue. Therefore, this left a three second interval in which the stimulus could be randomly presented. When participants were presented with an alarm, they responded with the appropriate number key to indicate what level they believed the alarm to represent. For a given condition, each of the four stimulus levels was presented four times, resulting in 16 trials.

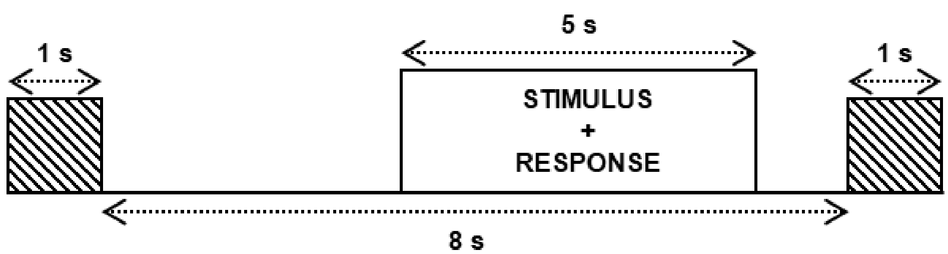

Figure 7.4: A visual representation of the time frame of each stimuli's presentation in Experiment 6.

7.1.7 Apparatus

Audio cues were presented using a pair of Beyerdynamic DT-100 headphones. The vibration actuator used was a Tactile Labs Haptuator MkII. [1] The actuator was attached to the posterior of the participants wrist, in a similar fashion to a watch.

7.2 Results and Discussion

This section is split by each dependent variable, reporting each of these measurements in its own section. Discussion of these results is split into two: one section discussing the results pertaining to the alarm classification task and another discussing the *N*-back results.

[1] http://tactilelabs.com/products/haptics/haptuator-mark-ii-v2

7.2.1 Total Errors in Alarm Classifications

Responses for all conditions were not normally-distributed (Shapiro-Wilk test, $p < 0.05$), therefore the Aligned Rank Transform [165] was applied before analysis. A mixed-effects model analysis was used, looking at the effect of polarity, stimulus level, N-back level in addition to the interactions between these variables effect on the total errors in alarm classification. To investigate pairwise differences between each stimulus level, *Post hoc* tests were carried out using least-squared means in the R emmeans package.[2] For pairwise comparisons of interaction effects, the testInteractions function of the phia package was used.[3] A Bonferroni corrected chi-squared test was used. Figure 7.5 shows the mean number of alarm classification errors for each stimulus level in each condition and polarity. Note these plots combine both N-back levels. Figure 7.6 shows the mean number of alarm classification errors for each stimulus level in each condition and N-back level.

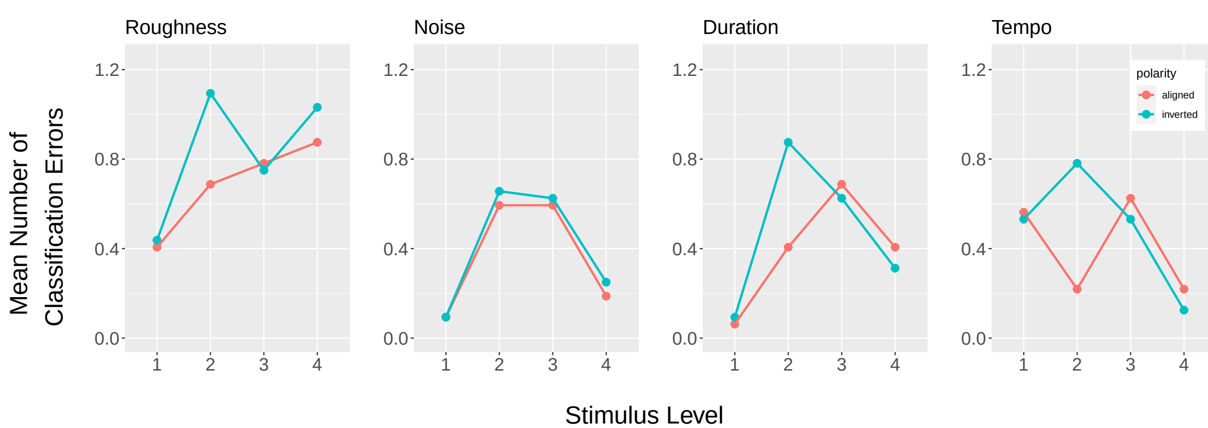

Figure 7.5: Plots showing the mean number of classification errors for each stimulus level and polarity in each condition in Experiment 6. Red plots show the aligned polarity, blue show the inverted polarity.

Roughness

Results for the roughness conditions showed that the stimulus level (or the amount of roughness present in the signal) was the only factor that significantly affected the number of erroneous alarm classifications ($p < 0.0001$). Table 7.2 shows more details of these results. Pairwise differences were found between stimulus levels 1 and 2 ($p = 0.0005$), levels 1 and 3 ($p = 0.009$) and levels 1 and 4 ($p < 0.0001$).

[2]https://cran.r-project.org/web/packages/emmeans/index.html
[3]https://cran.r-project.org/web/packages/phia/phia.pdf

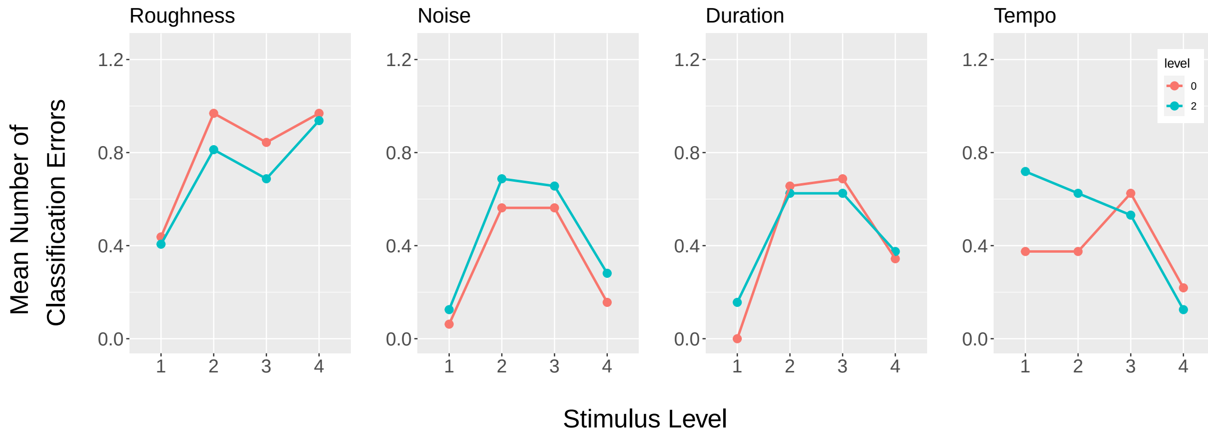

Figure 7.6: Plots showing the mean number classification errors for each stimulus level and *N*-back level in each condition in Experiment 6. Red plots show responses for *N*-back level 0, blue plots show *N*-back level 2.

Term	**F**	**Df**	**Df.res**	**Pr(>F)**
Polarity	3.66	1	225	0.0570
N-back Level	1.65	1	225	0.2005
Stimulus Level	9.81	3	225	<0.0001
Polarity:*N*-back Level	0.25	1	225	0.6196
Polarity:Stimulus Level	1.79	3	225	0.1498
N-back Level:Stimulus Level	0.40	3	225	0.7530
Polarity:*N*-back Level:Stimulus Level	1.28	3	225	0.2831

Table 7.2: ANOVA results for the effect of polarity, *N*-back level and stimulus level on number of errors in the roughness condition in Experiment 6.

Noise

For the noise conditions, both the level of the *N*-back task ($p = 0.0133$) and the stimulus level ($p < 0.0001$) and significantly affected the number of incorrect classifications. Table 7.3 shows more details of these results. Pairwise differences were found between stimulus levels 1 and 2 ($p < 0.0001$), levels 1 and 3 ($p < 0.0001$) and levels 1 and 4 ($p = 0.036$), levels 2 and 4 ($p = 0.044$) and levels 3 and 4 ($p = 0.004$).

Duration

For the vibrotactile duration conditions, the stimulus level ($p < 0.001$) and the interaction between the polarity of the mapping and the stimulus level ($p = 0.006$) significantly affected the total number of errors in alarm classification. Table 7.4 shows more details of these results. Pairwise differences were found between stimulus levels 1 and 2 ($p < 0.0001$), levels 1 and 3 ($p < 0.0001$), levels 1 and 4 ($p = 0.0003$) and levels 2 and 4 ($p = 0.028$). Looking at *post hoc* tests for the interaction between polarity and the stimulus level, the following pairwise differences were found: the interaction between the aligned-inverted polarities and stimulus levels 2-3 (p = 0.013) and the interaction between the aligned-inverted polarities and stimulus levels 2-4 (p = 0.012).

Term	**F**	**Df**	**Df.res**	**Pr(>F)**
Polarity	1.56	1	225	0.2123
N-back Level	6.23	1	225	0.0133
Stimulus Level	15.43	3	225	<0.0001
Polarity:*N*-back Level	0.74	1	225	0.3915
Polarity:Stimulus Level	1.14	3	225	0.3348
N-back Level:Stimulus Level	1.36	3	225	0.2557
Polarity:*N*-back Level:Stimulus Level	1.45	3	225	0.2284

Table 7.3: ANOVA results for the effect of polarity, *N*-back level and stimulus level on number of errors in the noise condition in Experiment 6.

Term	**F**	**Df**	**Df.res**	**Pr(>F)**
Polarity	1.69	1	225	0.1946
N-back Level	0.31	1	225	0.5774
Stimulus Level	20.41	3	225	<0.0001
Polarity:*N*-back Level	0.10	1	225	0.7578
Polarity:Stimulus Level	4.25	3	225	0.0061
N-back Level:Stimulus Level	0.17	3	225	0.9180
Polarity:*N*-back Level:Stimulus Level	0.90	3	225	0.4424

Table 7.4: ANOVA results for the effect of polarity, *N*-back level and stimulus level on number of errors in the duration condition in Experiment 6.

Tempo

For the vibration pulse tempo conditions, the level of the *N*-back task ($p < 0.031$), the stimulus level ($p < 0.001$), and the interaction between the mapping polarity and stimulus level ($p = 0.001$) and the interaction between the level of the *N*-back task and stimulus level ($p = 0.015$) significantly affected the number of classification errors. Table 7.5 shows more details of these results. Pairwise differences were found between stimulus levels 1 and 4 ($p = 0.0001$), levels 2 and 4 ($p = 0.0002$) and levels 3 and 4 ($p = 0.0002$). Looking at *post hoc* tests for the interaction between polarity and the stimulus level, the following pairwise differences were found: the interaction between the aligned-inverted polarities and stimulus levels 1-2 ($p = 0.002$), the interaction between the aligned-inverted polarities and stimulus levels 2-3 ($p = 0.004$) and the interaction between the aligned-inverted polarities and stimulus levels 2-3 ($p = 0.035$).

Term	F	Df	Df.res	Pr(>F)
Polarity	2.95	1	225	0.0871
N-back Level	4.67	1	225	0.0317
Stimulus Level	9.39	3	225	<0.0001
Polarity:*N*-back Level	3.73	1	225	0.0548
Polarity:Stimulus Level	5.59	3	225	0.0010
N-back Level:Stimulus Level	3.58	3	225	0.0147
Polarity:*N*-back Level:Stimulus Level	1.91	3	225	0.1286

Table 7.5: ANOVA results for the effect of polarity, *N*-back level and stimulus level on number of errors in the tempo condition in Experiment 6.

7.2.2 Alarm Classification Reaction Time

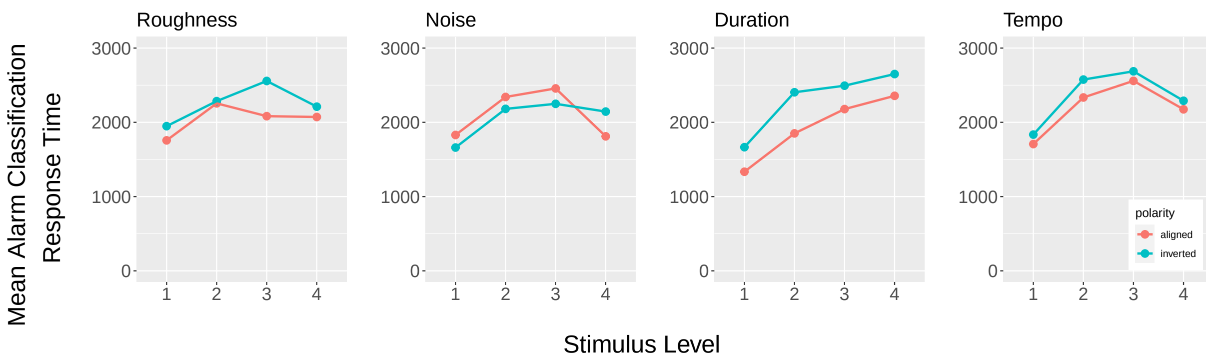

Figure 7.7: Plots showing the mean classification response times in milliseconds for each stimulus level and polarity in each condition in Experiment 6. Red plots show the aligned polarity, blue show the inverted polarity.

Responses for all conditions were not normally-distributed (Shapiro-Wilk test, $p < 0.05$), therefore the Aligned Rank Transform [165] was applied before analysis. A mixed-effects model analysis was used, looking at the effect of polarity, stimulus level, *N*-back level in addition to the interactions between these variables effect on the alarm classification reaction time. Figure 7.7 shows the mean classification response time for each stimulus level in each condition and polarity. Note these plots combine both *N*-back levels. Figure 7.8 shows the mean classification response time for each stimulus level in each condition and *N*-back level.

Roughness

For the roughness condition, the *aligned* polarity resulted in significantly faster response times than the *inverted* polarity ($p = 0.0033$) the level of the *N*-back task ($p < 0.0001$) and the stimulus level ($p < 0.0001$) significantly affected the response times. Table 7.6 shows more details of these results. Pairwise differences were found between stimulus levels 1 and 2 ($p = 0.02$), levels 1 and 3 ($p = 0.009$) and levels 1 and 4 ($p < 0.0001$).

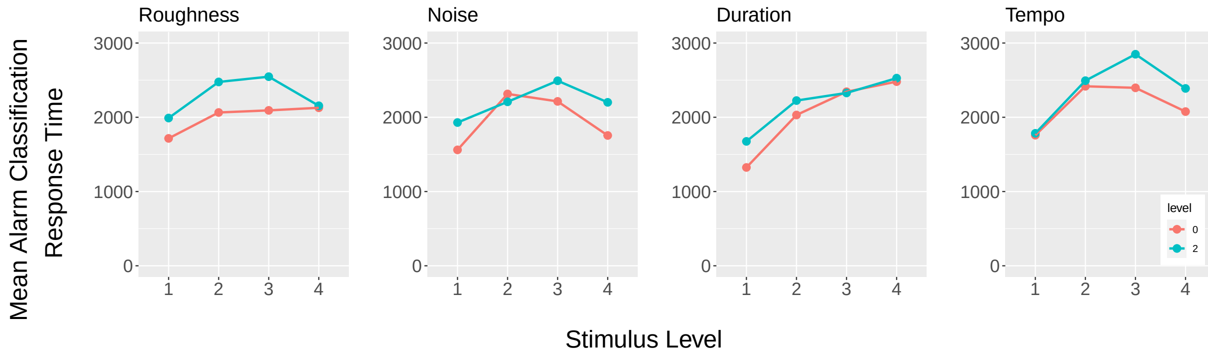

Figure 7.8: Plots showing the mean classification response times for each stimulus level and *N*-back level in each condition in Experiment 6. Red plots show responses for *N*-back level 0, blue plots show *N*-back level 2.

Term	F	Df	Df.res	Pr(>F)
Polarity	8.82	1	225	0.0033
N-back Level	15.03	1	225	0.0001
Stimulus Level	11.78	3	225	<0.0001
Polarity:*N*-back Level	1.43	1	225	0.2325
Polarity:Stimulus Level	1.31	3	225	0.2721
N-back Level:Stimulus Level	0.83	3	225	0.4761
Polarity:*N*-back Level:Stimulus Level	1.72	3	225	0.1644

Table 7.6: ANOVA results for the effect of polarity, *N*-back level and stimulus level on reaction time in the roughness condition in Experiment 6.

Noise

For the noise condition, both the level of the *N*-back task ($p < 0.034$) and the stimulus level ($p < 0.0001$) significantly affected the response times. Table 7.7 shows more details of these results. Pairwise differences were found between stimulus levels 1 and 2 ($p = < 0.0001$), levels 1 and 3 ($p < 0.0001$), levels 2 and 4 ($p = 0.001$) and levels 3 and 4 ($p = 0.0005$).

Term	F	Df	Df.res	Pr(>F)
Polarity	0.02	1	225	0.8870
N-back Level	4.53	1	225	0.0345
Stimulus Level	14.44	3	225	<0.0001
Polarity:N-back Level	0.79	1	225	0.3765
Polarity:Stimulus Level	2.08	3	225	0.1043
N-back Level:Stimulus Level	1.16	3	225	0.3250
Polarity:N-back Level:Stimulus Level	0.85	3	225	0.4687

Table 7.7: ANOVA results for the effect of polarity, *N*-back level and stimulus level on reaction time in the noise condition in Experiment 6.

Duration

For the vibrotactile duration condition, the *aligned* polarity resulted in significantly faster response times than the *inverted* polarity ($p < 0.0001$). Additionally, the stimulus level significantly affected classification response times ($p < 0.0001$). Table 7.8 shows more details of these results. Pairwise differences were found between levels 1 and 2 ($p < 0.0001$), levels 1 and 3 ($p < 0.0001$), levels 1 and 4 ($p < 0.0001$), and levels 2 and 4 ($p = 0.0005$).

Term	F	Df	Df.res	Pr(>F)
Polarity	20.30	1	225	<0.0001
N-back Level	1.89	1	225	0.1706
Stimulus Level	49.52	3	225	<0.0001
Polarity:N-back Level	1.13	1	225	0.2899
Polarity:Stimulus Level	1.41	3	225	0.2409
N-back Level:Stimulus Level	1.36	3	225	0.2570
Polarity:N-back Level:Stimulus Level	1.15	3	225	0.3299

Table 7.8: ANOVA results for the effect of polarity, N-back level and stimulus level on reaction time in the duration condition in Experiment 6.

Tempo

For the pulse tempo condition, the *aligned* polarity resulted in significantly faster response times than the *inverted* polarity ($p = 0.0273$). Additionally, the level of the N-back task ($p < 0.012$) and the stimulus level ($p < 0.001$) significantly affected the response times. Table 7.9 shows more details of these results. Pairwise differences were found between levels 1 and 2 ($p < 0.0001$), levels 1 and 3 ($p < 0.0001$), levels 1 and 4 ($p < 0.0002$), and levels 3 and 4 ($p = 0.006$).

Term	F	Df	Df.res	Pr(>F)
Polarity	4.93	1	225	0.0273
N-back Level	6.40	1	225	0.0121
Stimulus Level	21.80	3	225	<0.0001
Polarity:N-back Level	3.56	1	225	0.0606
Polarity:Stimulus Level	0.48	3	225	0.6934
N-back Level:Stimulus Level	0.53	3	225	0.6627
Polarity:N-back Level:Stimulus Level	0.74	3	225	0.5279

Table 7.9: ANOVA results for the effect of polarity, N-back level and stimulus level on reaction time in the tempo condition in Experiment 6.

7.2.3 Alarm Classification Discussion

In none of the conditions did the polarity alone affect the mean number of classification errors. However, in both the vibrotactile conditions (duration and tempo) the interaction between the polarity and the stimulus level significantly affected the mean number of classification errors. The *post hoc* tests for polarity:stimulus level in both of these parameters showed significant pairwise differences between stimulus levels 2-3 and 2-4 for each polarity. It is interesting that stimulus level 2 for both parameters is where the largest disparity between stimulus levels occur. The evidence from these tests, in addition to observing Figure 7.5, suggest that participants' ability to classify alarms that were represented using stimulus level 2, was increased in the aligned polarity over the inverted polarity. Stimulus level 1 seemed to be easily distinguishable in both polarities, in all the parameters, stimulus level 1 responses had a similar number of errors for each polarity — but when the inverted polarity was used, stimulus level 2 became more difficult to distinguish. This could be caused by participants preferring a certain sequence of stimuli to represent increasing levels of danger (the aligned polarity would be expected as this preferred sequence). Therefore if stimulus level 2 is presented in the non-preferred sequence (the inverted polarity) participants found this more difficult to classify.

Considering which stimulus levels resulted in the lowest mean number of classification errors may provide some interesting insight into how the design of each vibrotactile or acoustic parameter impacts the correctness of alarm classifications. Looking at the noise conditions, levels 1 and 4 resulted in significantly fewer classification errors in both polarities (Figure 7.5), this would suggest that participants found cues that were a totally unmodulated tone (level 1) or white noise (level 4) easier to classify than levels 2 and 3 which were mixes of tone and noise. This can also be seen in the roughness conditions where level 1 (which was also a totally unmodulated tone) resulted in significantly fewer classification errors than the rest of the levels. This means that the difference between some kind of complete modulation, either a complete lack of modulation, as in a pure tone, or a completely modulated tone, as in white noise, is easier to classify than degrees between modulations. This could be exploited in systems where easily distinguishable cues are key, especially in safety-critical contexts.

This echoes the findings from Experiment 2, where level 1 of the roughness and noise ranges used (which was also an unmodulated tone) was found to be a good representation of a zero value. The same can be found the vibrotactile results where level 1 of the duration conditions resulted in significantly fewer errors than the other levels and level 4 of the tempo conditions resulted in significantly fewer errors than all the other levels. For the vibrotactile parameters, these results suggest that certain extremes of the range, the start in the case of duration and the end in the case of tempo, were easier to classify.

For all parameters aside from noise, the aligned polarity resulted in significantly faster alarm classifications than the inverted polarity, meaning that these mappings based on prior magnitude estimation experiments did result in quicker responses. Similar to the results for the total errors in alarm classification, particular stimuli levels resulted in faster performance, based on pairwise comparison tests: level 1 of roughness, levels 1 and 4 of noise, level 1 of duration and level 4 of tempo. Again, this indicates that the extremes of the ranges of stimulus levels are easier to classify and therefore they were classified faster than the in-between levels of the range. Additionally, for the acoustic parameters (roughness and noise) the *N*-back level significantly affected the alarm classification reaction time, with responses at *N*-back level 0 being faster than those at *N*-back level 2

7.2.4 Total Errors in *N*-back Response

Responses for all conditions were not normally-distributed (Shapiro-Wilk test, $p < 0.05$), therefore the Aligned Rank Transform [165] was applied before analysis. A mixed-effects model analysis was used, looking at the effect of polarity and *N*-back level in addition to the interactions between these variables effect on the total errors in *N*-back response. Figure 7.9 shows the mean number of incorrect *N*-back responses for each condition. For all conditions, the *N*-back level was the only factor that significantly affected the number of incorrect *N*-back task responses. roughness ($p < 0.001$), noise ($p < 0.001$), duration ($p < 0.001$) and tempo ($p < 0.001$). Tables 7.10 - 7.13 show more details of these results.

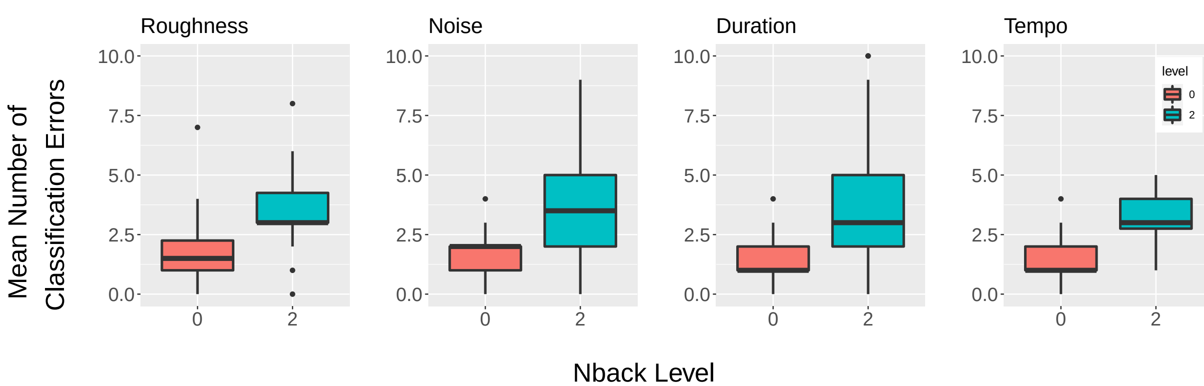

Figure 7.9: Plots showing the mean number of *N*-back response errors for each condition in Experiment 6. Red plots show responses for *N*-back level 0, blue plots show *N*-back level 2.

7.2.5 *N*-back Response Reaction Time

Responses for all conditions were not normally-distributed (Shapiro-Wilk test, $p < 0.05$), therefore the Aligned Rank Transform [165] was applied before analysis. A mixed-effects model

Term	**F**	**Df**	**Df.res**	**Pr(>F)**
Polarity	1.55	1	45	0.2194
N-back Level	25.97	1	45	<0.0001
Polarity:*N*-back Level	0.83	1	45	0.3666

Table 7.10: ANOVA results for the effect of polarity, *N*-back level and stimulus level on the number of *N*-back errors in the roughness condition in Experiment 6.

Term	**F**	**Df**	**Df.res**	**Pr(>F)**
Polarity	0.02	1	45	0.8816
N-back Level	35.17	1	45	<0.0001
Polarity:*N*-back Level	0.03	1	45	0.8578

Table 7.11: ANOVA results for the effect of polarity, *N*-back level and stimulus level on the number of *N*-back errors in the noise condition in Experiment 6.

Term	**F**	**Df**	**Df.res**	**Pr(>F)**
Polarity	2.21	1	45	0.1439
N-back Level	40.26	1	45	<0.0001
Polarity:*N*-back Level	0.78	1	45	0.3813

Table 7.12: ANOVA results for the effect of polarity, *N*-back level and stimulus level on the number of *N*-back errors in the duration condition in Experiment 6.

Term	**F**	**Df**	**Df.res**	**Pr(>F)**
Polarity	2.52	1	45	0.1191
N-back Level	54.76	1	45	< 0 .0001
Polarity:*N*-back Level	3.34	1	45	0.0742

Table 7.13: ANOVA results for the effect of polarity, *N*-back level and stimulus level on the number of *N*-back errors in the tempo condition in Experiment 6.

analysis was used, looking at the effect of polarity and *N*-back level in addition to the interactions between these variables effect on the reaction time for *N*-back responses. For all conditions, the *N*-back level was the only factor that significantly affected the reaction time of responses. roughness (p = 0.0017), noise ($p < 0.001$), duration ($p < 0.001$) and tempo ($p < 0.001$). Figure 7.10 shows the mean reaction time for *N*-back responses in each condition. Tables 7.14 - 7.17 show more details of these results.

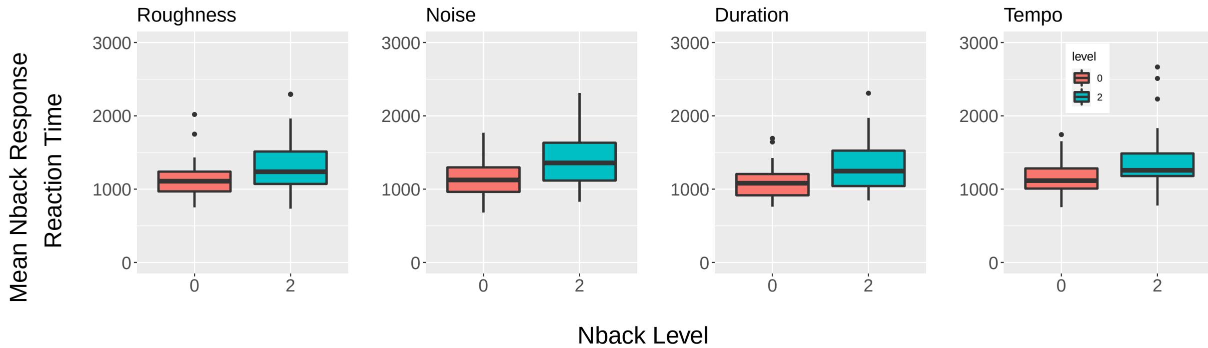

Figure 7.10: Plots showing the mean *N*-back response reaction time in milliseconds for each condition in Experiment 6. Red plots show responses for *N*-back level 0, blue plots show *N*-back level 2.

Term	**F**	**Df**	**Df.res**	**Pr(>F)**
Polarity	0.37	1	45	0.5472
N-back Level	11.15	1	45	0.0017
Polarity:*N*-back Level	0.12	1	45	0.7335

Table 7.14: ANOVA results for the effect of polarity, *N*-back level and stimulus level on the *N*-back reaction time in the roughness condition in Experiment 6.

Term	**F**	**Df**	**Df.res**	**Pr(>F)**
Polarity	0.00	1	45	0.9552
N-back Level	20.96	1	45	<0.0001
Polarity:*N*-back Level	0.01	1	45	0.9098

Table 7.15: ANOVA results for the effect of polarity, *N*-back level and stimulus level on the *N*-back reaction time in the noise condition in Experiment 6.

Term	**F**	**Df**	**Df.res**	**Pr(>F)**
Polarity	2.49	1	45	0.1214
N-back Level	31.44	1	45	< 0.0001
Polarity:*N*-back Level	0.00	1	45	0.9442

Table 7.16: ANOVA results for the effect of polarity, *N*-back level and stimulus level on the *N*-back reaction time in the duration condition in Experiment 6.

Term	F	Df	Df.res	Pr(>F)
Polarity	3.74	1	45	0.0596
N-back Level	22.06	1	45	< 0.0001
Polarity:*N*-back Level	0.08	1	45	0.7815

Table 7.17: ANOVA results for the effect of polarity, *N*-back level and stimulus level on the *N*-back reaction time in the tempo condition in Experiment 6.

7.2.6 NASA TLX Results

Responses for all conditions were normally-distributed (Shapiro-Wilk test, $p > 0.05$). An ANOVA was conducted that looked at the effect of polarity, stimulus level, *N*-back level in addition to the interactions between these variables effect on the subjective overall workload scores obtained. For all conditions, the level of the *N*-back task was the only factor that significantly affected the total subjective workload, as obtained from the NASA TLX questionnaires: roughness ($p = 0.0082$), noise (mbox$p = 0.001$), duration ($p = 0.0021$) and tempo ($p = 0.0019$). Figure 7.11 shows the mean subjective workload for each condition. Tables 7.18 - 7.21 show more details of these results.

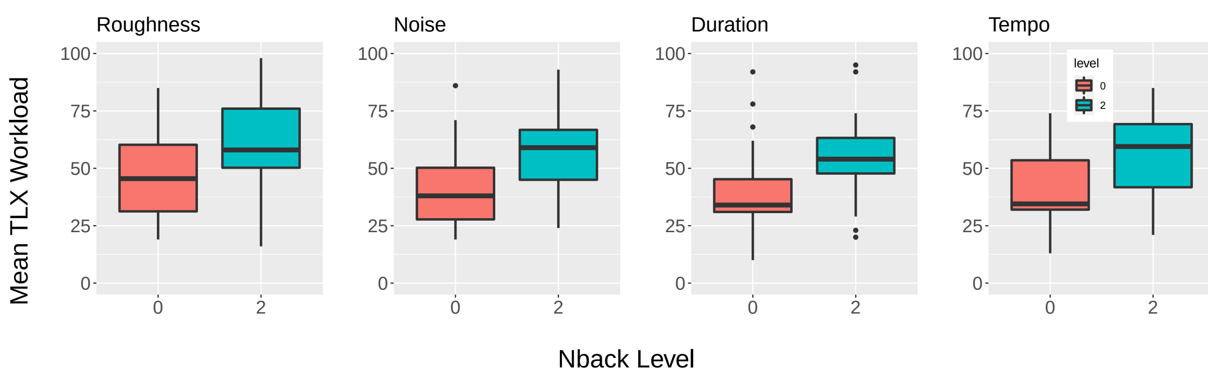

Figure 7.11: Plots showing the mean subjective workload for each condition in Experiment 6. Red plots show responses for *N*-back level 0, blue plots show *N*-back level 2.

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Polarity	1	2.16	2.16	0.01	0.9389
N-back Level	1	2758.02	2758.02	7.56	0.0082
Polarity:*N*-back Level	1	75.45	75.45	0.21	0.6512
Residuals	52	18968.21	364.77		

Table 7.18: ANOVA results for the effect of polarity and *N*-back level on perceived workload for the roughness condition in Experiment 6.

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Polarity	1	108.64	108.64	0.34	0.5616
N-back Level	1	3877.79	3877.79	12.18	0.0010
Polarity:*N*-back Level	1	434.57	434.57	1.37	0.2480
Residuals	52	16554.43	318.35		

Table 7.19: ANOVA results for the effect of polarity and *N*-back *N*-back Level on perceived workload for the noise condition in Experiment 6.

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Polarity	1	6.45	6.45	0.02	0.8900
N-back Level	1	3504.45	3504.45	10.51	0.0021
Polarity:*N*-back Level	1	5.16	5.16	0.02	0.9015
Residuals	52	17341.79	333.50		

Table 7.20: ANOVA results for the effect of polarity and *N*-back level on perceived workload for the duration condition in Experiment 6.

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Polarity	1	228.02	228.02	0.77	0.3849
N-back Level	1	3195.16	3195.16	10.76	0.0019
Polarity:*N*-back Level	1	0.45	0.45	0.00	0.9692
Residuals	52	15440.93	296.94		

Table 7.21: ANOVA results for the effect of polarity and *N*-back level on perceived workload for the tempo condition in Experiment 6.

7.2.7 *N*-Back Discussion

Generally, the conditions where the difficulty of the *N*-back task was 0-back as opposed to 2-back resulted in significantly increased performance: less errors, faster reaction time and lower subjective workload. No other variable significantly impacted the mean number of *N*-back errors, the response reaction time or NASA TLX results. From this it can be concluded that the overall difference in difficulty between the two *N*-back levels was significant and accomplished what they were designed to do, that is induce two levels of cognitive workload.

7.3 General Discussion

Generally, the stimulus levels within the mapping had more impact on the usability of the mapping than the polarity when it comes to the correctness of classifying the stimuli. This could suggest that that designing an appropriate range of stimuli, of which the constituent levels of the range are as distinguishable from each other as possible, is a more important factor in correctly classifying stimuli than the polarity of the range is presented in. However, considering reaction time of alarm classification, polarity was a significant factor for all acoustic and vibrotactile parameters except for noise, with reaction times for the aligned polarity being significantly faster. This is an important result, as throughout this thesis there have been many examples of auditory and vibrotactile displays that are used in situations where fast reactions are critical such as: during surgery [8, 42, 94], controlling air-traffic [120] or driving [129]. Therefore, in these situations, any design choice that can increase the speed of response is important.

Even though the polarity of the mapping significantly affected the alarm classification response time in the roughness, duration and tempo conditions, it did not affect any of the N-back measures, either the reaction time, number of errors in responses, or subjective workload. This could suggest that the design of the data-sound or data-vibration mapping may not have a significant impact on how the users perform a simultaneous task but it does impact how the user interacts with the mapping itself. Furthermore, this suggests that when a participant was faced with the choice of spending cognitive resources on the N-back task or processing an auditory or vibrotactile cue that was presented in a polarity that was incongruent to them, they would rather sacrifice the alarm classification than the N-back response. If the user was spending more cognitive resources on classifying the alarm cues, it would be expected to be reflected in the N-back results. It would be expected that the polarity that fared worse in the alarm classification results would also fare worse in the N-back results, but this is not the case from the results obtained here.

This is an important finding, as it means that future research and design of data-sound and data-vibration mappings needs to have an awareness of any concurrent tasks that may be conducted alongside the auditory or vibrotactile display, as these can have an effect on how the user interacts with the mapping. This is further highlighted in the finding that the N-back level significantly affected the mean number of alarm classification errors in the noise and tempo conditions, as well as the alarm classification reaction times for the roughness, noise and tempo conditions.

7.4 Conclusions

This chapter investigated how inducing two different levels of cognitive load impacted performance in an acoustic or vibrotactile alarm classification task. A study was presented in which two difficulty levels of the N-back task were used to induce cognitive load, while a user simultaneously attempted to classify acoustic or vibrotactile alarms relating to danger, presented using acoustic roughness, acoustic noise, vibrotactile duration and vibrotactile pulse tempo. The participants classified the alarms based on the level of danger they represented (level 1 to 4). The polarity of these data-sound and data-vibration mappings were designed either using results from a previous magnitude estimation experiment (Experiment 3 for the acoustic parameters and Experiment 4 for the vibrotactile parameter) or from the inverse of these results. This resulted in two sets of stimuli: one where the polarity is *aligned* with previous results or *inverted* from the previous results.

Results showed that the levels of the stimuli, not the polarity of the mapping, significantly affected the number of erroneous classifications. *Post hoc* tests showed that all the acoustic and vibrotactile parameters had significantly less errors at one or both of the extremes of their ranges, for example, level 1 of the roughness stimuli range or levels 1 and 4 of the noise conditions. The time it took a participant to respond to an alarm was significantly affected by the polarity of the mapping in the roughness, duration and tempo conditions. In these conditions, the aligned polarity resulted in faster response times than the inverted one. Furthermore, for the acoustic parameters, the conditions where the *N*-back level was 0 resulted in significantly faster alarm classifications than *N*-back level 2.

The polarity of the mapping did not significantly affect any of the measures relating to the *N*-back task, these measures were: the number of erroneous responses and the reaction time of these responses. Additionally, the only factor that significantly affected these measures and the total subjective workload, was the level of the *N*-back task, with level 0 resulting in significantly fewer errors, faster responses and lower subjective workload. These findings related to the *N*-back measures, in tandem with the fact that the aligned polarity resulted in faster responses in the roughness, duration and tempo conditions, suggest that when participants are faced with a dual-attention task, where one of the tasks is attending and responding to an auditory or vibrotactile display in an incongruent mapping, the participant will sacrifice expending cognitive resources on attending to the mapping in favour of completing the other task.

7.4.1 Research Question 4

The studies presented in this chapter and the preceding one, present a first attempt at investigating how the use of magnitude estimation to design data-sound and data-vibration mappings' polarities actually impacts how participants interact with them when they are when implemented into a system. The main outcomes from these two studies are presented in the form of the following design recommendations that address research question 4, which asked:

> **RQ4:** How does the complexity of the usage context impact how effective a data-sound or data-vibration parameter mapping is?

Polarity Does not Impact Simple Display Contexts

Experiment 5 found that the polarity of the data-sound or data-vibration mapping did not significantly impact how well participants could carry out a simple auditory and vibrotactile display task. Therefore, designers of such systems may be able to forego a resource intensive magnitude estimation experiment if it cannot be afforded.

Polarity Impacts Alarm Classification Reaction Time

A mapping designed using prior magnitude estimation results induced significantly faster alarm classification response times in the roughness, duration and tempo conditions. Therefore, conducting a magnitude estimation experiment prior to design can reduce how long it takes to process and respond to these cues.

Complexity of Secondary Task Impacts Mapping Reaction Time

Increased cognitive load from a secondary task affected the response time for classifying alarm cues. Therefore, a designer creating a mapping with these parameters should take into consideration the difficulty of the secondary task as it can impact how quickly users respond to cues from these mappings.

Extremes of the Mapping Range are Easier to Distinguish

The first or first and last stimuli in a mapping's range induced significantly better results than the other stimuli. Therefore it may be advantageous for designers to pay extra care to the design of the first and last cues in the range of a data-sound or data-vibration mapping to ensure that the extremes of the range are as easy to distinguish as possible.

Chapter 8

Conclusions

This thesis made the following statement in its introduction:

> Which sensory parameter to use to convey a particular information parameter in an auditory or vibrotactile display is a significant design problem, as the perceived meaning and connotations of auditory and haptic stimuli are complex. This thesis presents experiments which used psychophysical scaling to establish how several data-sound and data-vibration mappings were perceived and subsequently investigated the impact that using results from these experiments had on users' performance when they were implemented into the design of a parameter mapping display. This thesis presents a number of design recommendations based on results from these studies which designers and researchers can use to guide their decisions on which acoustic or vibrotactile parameter is appropriate for a particular information parameter.

In the subsequent chapters, research was presented in support of this thesis statement, which in turn, addressed the research questions of the thesis:

RQ1: How may acoustic parameters that are derived from the field of psychoacoustics encode and transmit information to a user?

RQ2: To what extent are mappings between acoustic sonification parameters and information parameters with clear positive or negative valence perceived as congruent?

RQ3: To what extent are mappings between vibrotactile parameters and information parameters perceived as congruent?

RQ4: How does the complexity of the usage context impact how effective a data-sound or data-vibration parameter mapping is?

Chapter 2 presented a literature review which laid the basis for the problems that this thesis addressed and provided a summary of related work in the field. Chapter 3 presented and studied several psychoacoustic parameters, tested if they could be used in a sonification context and demonstrated some of their advantages in this context. This study was carried out in the real-world use-case of detecting the quality of an astronomical image via sonification, for use with visually impaired students. Following this, Chapter 4 further explored these psychoacoustic parameters in more general contexts by exploring how they could represent more generalised information parameters such as error and danger. This study used a magnitude estimation paradigm to establish how participants perceived these data-sound mappings and obtained information such as the most popular polarities for each mapping. Chapter 5 introduced the first vibrotactile study in the thesis, using a similar paradigm as in Chapter 4 to understand how several data-vibration mappings were perceived. Chapters 5 and 6 explored how the use of this magnitude estimation paradigm to inform the design of data-sound and data-vibration mappings impacted the usability of the systems in which they were implemented. Chapter 6 investigated a basic context, an auditory and vibrotactile display to convey the security of a WiFi network. Chapter 7 explored a more complex use-case, a dual attention task where participants completed a memory task while simultaneously attending to and classifying alarms relating to a level of danger. This chapter summarises the research and returns to each Research Question to discuss how they were addressed. This chapter also summarises the main contributions of this thesis and discusses possible limitations and areas of future work.

8.1 Research Questions

8.1.1 Research Question 1

> **RQ1:** How may acoustic parameters that are derived from the field of psychoacoustics encode and transmit information to a user?

Sonification designers have little evidence to guide them on the design of data-sound mappings. Many mappings use acoustic representations of data values which may not correspond with the listener's perception of how that data value should sound during sonification, with most systems defaulting to use pitch as a mapping parameter. Chapter 3 described two experiments which explored how acoustic parameters that were derived from psychoacoustics research could be used in sonification. In these experiments, participants were tasked with establishing the level of focus of an astronomical image through sonification. Experiment 1 explored the use of the psychoacoustic parameters of roughness, sharpness and a combination of both to convey image focus. This experiment found that using acoustic noise as a signal with which to apply roughness and sharpness to did not result in a usable interface.

Therefore, in Experiment 2, pure tones instead of noise were used as a carrier signal for roughness and this experiment also explored an extra parameter, the amount of white noise present in a signal. As with Experiment 1, a combined parameter was also explored: roughness and white noise amount combined. A visual condition was included to explore how well participants completed the task as a baseline. Experiment 2 found that the psychoacoustic parameters used could effectively convey the information, with the noise and combined roughness and noise parameters performing similarly to the visual condition. Another finding that can be a useful as a guide for designers, is that since the roughness and noise conditions both consisted of an unmodulated pure tone at level 1 and this level resulted in decreased errors compared to the other levels, having an unmodulated value can represent a ‘lack of’ whatever the variable may be.

8.1.2 Research Question 2

> **RQ2:** To what extent are mappings between acoustic sonification parameters and information parameters with clear positive or negative valence perceived as congruent?

Experiments 1 and 2 looked at a specific use-case, establishing how in or out of focus an astronomical image was, based on a sonification. This value of ‘out of focus-ness’ or, more generally speaking, blur, can be discussed in terms of its *valence*, meaning “events, objects or situations that possess intrinsic attractiveness or aversiveness” [46] (p. 207). The findings from these two experiments suggested that increasing a musically ‘undesirable’ parameter like noise is perceived by most participants to represent an increase in blur. This could be considered to represent a *negative* valence, because one undesirable parameter or parameter with negative valence represents another parameter with negative valence. Therefore, Experiment 3 explored these psychoacoustic parameters for more general information parameters which shared the property of negative valence. These parameters were error, stress and danger. To establish whether participants did in fact associate these psychoacoustic parameters with these information parameters, a magnitude estimation experiment was conducted. This experiment paradigm gives an indication of how magnitudes of acoustic parameters relate to the perceived magnitudes of information parameters.

This experiment found that the majority of participants responded in a positive polarity for all of the data-sound mappings used, meaning that, for example, as roughness or noise in a signal increased the magnitude of error, stress or danger they perceived also increased. This suggests that mappings between these acoustic parameters and information parameters were perceived as congruent by participants. This study also found more complex relationships such as the noise-stress mapping, where the polarity preference consensus among participants was less clear than the other parameters. This means that some participants perceived increasing noise as an increased value of stress, but some perceived it as a decrease. One reason for this could be that white noise is often used as a relaxation or sleep aid and therefore some participants may associate noise with the opposite of stress and there-fore responded in a negative polarity. This is indicative of a mapping where more care needs to be taken to establish the polarity of the mapping and how best it should be presented to users.

8.1.3 Research Question 3

> **RQ3:** To what extent are mappings between vibrotactile parameters and information parameters perceived as congruent?

This research question is addressed in Chapter 5, where Experiment 4 was presented, in which a similar magnitude estimation paradigm as Experiment 3 was used to explore several data-vibration mappings. There is less work looking at data-vibration mappings than there is for data-sound mappings. Therefore, in this experiment, a broader set of information parameters was investigated: error, stress and danger as in Experiment 3, as well as accuracy, size, distance and current. The vibrotactile parameters used were duration, pulse tempo, frequency and roughness (amplitude modulation).

This experiment found that information parameters with opposing valence are not necessarily perceived in opposing polarities. For example, in tempo-accuracy and the tempo-error conditions, the preferred polarity for both was positive, even though error has negative valence and accuracy has positive valence. Furthermore, this experiment found that polarity responses for mappings where duration and tempo were used did not depend on what the information parameter was, meaning that in every mapping the preferred polarity was positive. This suggests that for data-vibration mappings, the valence of the information parameter is less of a key factor in a user's decision about a mapping's congruence, but the vibrotactile parameter's design is more important. The findings here suggest that if an information parameter is conveyed by a vibrotactile parameter positively, that is an increase in the vibrotactile parameter represents an increase in the information parameter, then this is likely to be perceived as congruent, regardless of the information parameter's valence.

8.1.4 Research Question 4

> **RQ4:** How does the complexity of the usage context affect the effectiveness of a data-sound or data-vibration parameter mapping?

This research question is addressed in Chapters 6 and 7, where Experiments 5 and 6 were presented. Experiment 5 took the data-sound mappings of danger-roughness and danger-noise from Experiment 3, in addition to the data-vibration mappings of danger-duration and danger-tempo from Experiment 4, and applied them to a simple use case, identifying the security of a WiFi network (low, medium or high). The polarity of these mappings was either based on results from Experiments 3 and 4 or was the inverse of these polarities. This was to explore if using polarities derived from magnitude estimation experiments impacted the usability of a data-sound or data-vibration mapping. Participants had to rank WiFi networks according to their level of security, based on an auditory or vibrotactile cue.

This experiment found that the polarity of the data-sound or data-vibration mapping did not significantly impact users' performance with the system, in terms of correctness of rankings or the time it took them to complete rankings. The Akaike Information Criterion was used to provide a statistical basis for this. This suggests that for simple contexts using these parameters in a data-sound or data-vibration mapping, conducting a resource intensive magnitude estimation experiment may not be necessary. However, this study had a number of limitations, primarily that the task's simplicity was not an example of the more cognitively demanding situations where an auditory or tactile display might be used, such as surgery [8, 42, 94], controlling air-traffic [120] or driving [129].

Therefore, to explore these limitations and to further address research question 4, Experiment 6 was conducted. Here, the same data-sound and data-vibration mappings were used but they were in the context of an alarm that conveyed a level of danger. Similarly to Experiment 5, the polarities of these mappings were either based on results from Experiments 3 and 4 or their inverse. Simultaneously to attending to and processing these alarms, participants conducted a secondary task which was a cognitively demanding memory task with two levels of difficulty. This experiment found that the polarity of the data-sound or data-vibration mapping did impact participants reaction time in responding to alarms, with the polarities derived from the previous magnitude estimation experiments resulting in faster response times. The complexity of the secondary task also impacted how long it took participants to respond to both auditory and vibrotactile alarms, with the higher difficulty of the memory task resulting in slower response times. Finally, participants performed significantly better both in terms of correctness and response times when the auditory or vibrotactile alarm that was presented was at the extremes of their range, either the first, last or both first and last stimuli in the mapping range. This means that, for example, participants classified alarms using the danger-noise mappings significantly better at levels 1 and 4, than levels 2 and 3.

8.2 Impact of Results

The discussion of the results of this thesis so far have been self-contained, however, this section discusses how the conclusions of the research conducted can impact human-computer interaction design and research more broadly. The goal of this section is to highlight how findings from this research may benefit designers and researchers who are working in areas and contexts that may be beyond the direct scope of this thesis, but may still find use in the results.

8.2.1 Use-Cases

The latter part of this thesis focused on applications where danger was conveyed using an auditory or vibrotactile display, particularly in the context of network security and a generic alarm application. The reason for using a generic application of danger being conveyed in the alarm application was to allow the results to be generalisable, so that they can be relevant to several potential applications where some value of danger is required to be conveyed. There are several situations where such a non-visual alarm system to convey danger could be required: in the cockpit of a car or aircraft, process monitoring such as in a power station, controlling air-traffic or in medical setting such as surgery. As a non-specific context was used in the alarm application in Experiment 6, designers or researchers working on these use-cases that have more particular values of danger to convey may find benefit in these results. For example, in these contexts, there may be several systems transmitting sound, therefore designers may be constrained in terms of what acoustic parameters they can use, as some are already implemented. Therefore the finding from this thesis regarding perceptual congruence between musically undesirable parameters and information parameters with negative valence can provide a guide for designers to find new parameters that may not already be implemented in their systems. Furthermore, all of these use-cases are dual-attention tasks where users are conducting some cognitively demanding task while simultaneously attending to information being transmitted non-visually and this is an additional similarity shared with Experiment 6, further increasing the relevance of this thesis' findings to these potential use-cases.

8.2.2 Acoustic and Vibrotactile Parameters

This thesis focused on a subset of the acoustic and vibrotactile parameter space, however, results from the experiments conducted may be applicable to parameters beyond the ones studied here. For example, this thesis focused primarily on acoustic parameters that could be considered undesirable in Western music tradition, primarily roughness and noise, and investigated how they could be used to convey information with negative valence. This thesis found that pairing these types of acoustic and information parameters was perceived as congruent, therefore it could be suggested that other acoustic parameters that are considered undesirable, beyond those studied here, could be similarly perceived as congruent with information parameters with negative valence. For example, in music, instruments being out of time with each other is generally unwanted, therefore this could be leveraged in an auditory display. Furthermore, this thesis applied noise to a signal in attempt to give the effect that the signal is less clear, there are several other effects that could be used in a similar fashion such as reverb, delay, distortion, overdrive, fuzz, chorus, flanging or phase shifting.

The findings from Experiment 6 regarding the vibrotactile parameters investigated concluded that an increase in a vibrotactile parameter was perceived as congruent with an increase in an information parameter, regardless of the information's valence. Therefore, it could be suggested that this could be seen for other vibrotactile parameters that were not studied here, such as location of the actuator on the body or more complicated temporal patterns.

8.3 Contributions

This section summarises the novel contributions made in this thesis. Its main contributions are: (1) a study of the use of psychoacoustic parameters for sonification (2) an investigation of how mappings between information and acoustic or vibration parameters are perceived as congruent and (3) how the complexity of a use-case impacts how well a data-sound or data-vibration mapping can be used.

8.3.1 Psychoacoustic Parameters for Sonification

This thesis evaluated several novel psychoacoustic parameters as candidates for use in a data-sound mapping. Findings showed some advantages of psychoacoustic parameters over traditional acoustic parameters commonly used in sonification in that they can convey a zero value or a lack of whatever the data variable may be, something which is difficult to convey using sound.

8.3.2 Perceptual Congruence in Data-Sound and Data-Vibration Mappings

This thesis conducted two studies which used magnitude estimation to establish how magnitudes of acoustic or vibrotactile parameters relate to the perceived magnitudes of information parameters. Findings from these studies found that the valence of an information parameter is key in a data-sound mapping, for example increasing a musically 'undesirable' acoustic parameter such as noise resulted in a perceived increase in an information parameter with a negative valence such as danger. This results in a relationship that could be described as: more negative sound properties = more of a negative information parameter. For data-vibration mappings, findings suggested that they were more generalised in that an increase in a vibrotactile parameter was generally perceived as an increase in the information parameter, regardless of the valence of that parameter.

8.3.3 Complexity of Use-Cases of Auditory or Vibrotactile Displays

This thesis investigated the as-yet unexplored question of how using mapping designs based on magnitude estimation experiments impacts systems where these mappings are used. Two studies were conducted, the first being a simple use-case and the second being a complex dual-attention task. Findings from these studies suggested that the polarity of the data-sound or data-vibration mapping had little effect in the simple use case, however it significantly impacted performance in the complex use case. Furthermore, the difficulty of the non-acoustic/vibration component of the dual-attention task significantly impacted user's performance.

8.4 Design Recommendations

This thesis contributed several design recommendations that could be used by future researchers and designers to inform their data-sound or data-vibration mapping designs. These are summarised in the following section:

DR1: Use Pure Tones as Carrier Signals for Applying Roughness

Users are able to better distinguish cues with a data-sound mapping that used roughness applied to a pure tone, as opposed to roughness applied to broad band noise.

DR2: Use an Unmodulated Tone to Convey a Zero Value

An advantage of acoustic parameters that are based on modulation, such as roughness, or based on some additive property (such as adding noise to a signal) is that they can have a state that has no modulation or additive property. This is something that is not possible with more commonly used acoustic parameters like pitch or tempo. This means that the first level of such a mapping can be an unmodulated tone, representing a zero value or a lack of the information parameter being conveyed.

DR3: Use Musically Undesirable Parameters such as Roughness or Noise to Convey Information Parameters with Negative Valence

Acoustic parameters which would generally be deemed to be undesirable in terms of Western musical tradition such as roughness and noise were perceived as correlated with information parameters with negative valence. Therefore these parameters can be effective at conveying such information.

DR4: Use an Increase in Vibrotactile Duration or Tempo to Convey an Increase in an Information Parameter

An increase in a vibration's duration or tempo is perceived as an increase in all data-vibration mappings that were tested, regardless of the valence of the information parameter being used in the mapping.

DR5: Polarity of Mappings has Little Impact in Simple Auditory or Vibrotactile Displays

In a simple network security ranking tasks, whether the polarity of a data-sound or data-vibration mapping was based on results from a magnitude estimation experiment or were completely arbitrary had little effect on how well participants could carry out the task.

DR6: Polarity and Complexity of Secondary Task Does Impact Reaction Time

Users respond to alarm cues more quickly when the polarity of a mapping is based on magnitude estimation results. They also respond more quickly to alarm when conducting a secondary task if this task is lower in cognitive demand. Therefore conducting a magnitude estimation study on potential mappings and being aware of potential secondary tasks which can induce cognitive workload can result in faster responses.

DR7: Pay Attention to Extremes of a Mapping Range

The first or first and last stimuli in a mapping's range induced significantly faster and more accurate classifications than the other stimuli in their range. Therefore it may be advantageous for designers to pay extra care to the design of the first and last cues in the range of a data-sound or data-vibration mapping to ensure that the extremes of the range are as easy to distinguish as possible.

8.5 Limitations and Future Work

This section will summarise some general limitations of the research as a whole, alongside a discussion of future work that could both address these limitations and further explore the themes brought up in this thesis.

8.5.1 Other Acoustic and Vibrotactile Parameters

This thesis focused on a specific subset of acoustic parameters and there are a wide range of other parameters that could be investigated in a parameter mapping setting. Similarly with vibrotactile parameters, this thesis covered several vibration parameters, but there are some that were not investigated such as location on the body, waveform and more complex temporal patterns. Investigating more parameters beyond the ones explored in this thesis would be a logical next step for this work, as any opportunity to widen the understanding of how different parameters are perceived in mappings would be welcome for designers.

8.5.2 Crossmodal Mappings

This thesis explored acoustic and vibrotactile parameter mappings distinctly, looking at auditory displays and vibrotactile displays separately. A potential avenue for future work could be to investigate crossmodal displays where audio and vibration are used in tandem. Such crossmodal interfaces have proven to be advantageous [69], but there has not been a substantial amount of work in this area which has explored perception of parameter mappings.

8.5.3 Other Non-Visual Modalities

This thesis focused on parameter mappings that used sound or vibration to convey information. However, there many other modalities, specifically within the sense of touch, that can also be used to convey information and how users perceive and experience parameter mappings in these modalities has not been explored.

Ultrasonic feedback

Ultrasonic feedback [26] has emerged in the last decade as a novel method for haptic interaction, which uses an array of ultrasound emitting speakers to provide multi-point haptic feedback (Figure 8.1). It can convey complex structures like textures [44] and has been used in cognitively demanding situations like in-car feedback [130]. However, there has not been much work in this nascent field around the parameter mapping potential of ultrasonic feedback.

Thermal Feedback

The capacity for thermal feedback to convey information has been touched upon in human-computer interaction research, notably for navigation [35, 161], and it has been more thoroughly explored for its capacity to assist in emotion-based affective interactions [162, 164]. Wilson *et al.* [162] defined *Thermal Icons* that can be used to convey information in a similar manor to tactons. Their research was a first investigation of these icons and established that they can in fact convey information. Particularly, this work investigated the valence and arousal qualities of different thermal stimuli, which makes the modality a good potential candidate to study in a similar way to the research carried out in this thesis.

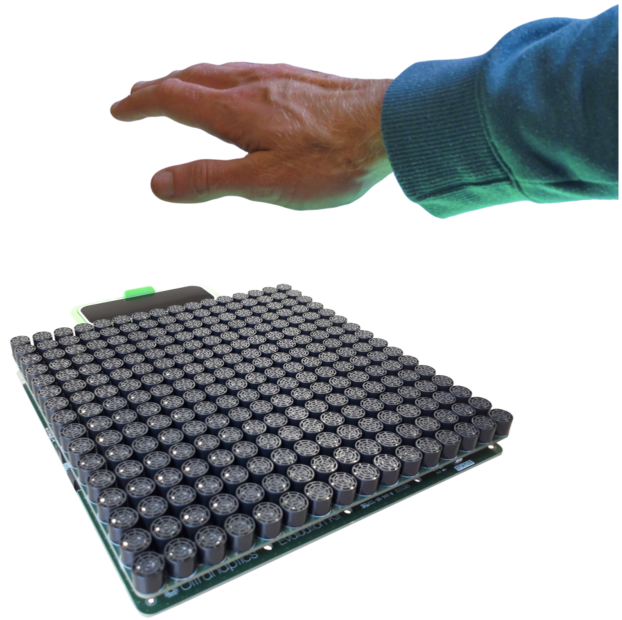

Figure 8.1: An UltraHaptics/Ultraleap Ultrasonic Array.

8.5.4 *In Situ* Research

This thesis focused on lab-based experiments, but a logical trajectory for this research would be studying data-sound or data-vibration mappings in the environment in which they are intended to be used. Experiment 6 looked at several mappings that conveyed alarms while participants attended to a secondary, cognitively demanding task. This basic scenario could be played out in several different settings, for example, a situation where some aspect(s) of a power plants condition is being monitored non-visually while the controller attends to some other complex task. Another typical situation would be driving, where the driver is carrying out the complex task of controlling the vehicle, but still requires feedback non-visually, relating to alarms or other information types. To explore a particular use-case of data-sound or data-vibration mappings, researching them in the location that they intend to be implemented in is necessary.

However, considering the power-plant monitoring situation as an example, it would be a hazard to safety to implement an untested system in a safety-critical context like this. Therefore, extended reality systems like virtual reality and augmented reality can provide a middle ground between highly controlled lab-based studies and studies which are fully integrated into the location where they are to be used. Of course, there may be situations where facilities have simulators that are used for training purposes, in which case that is an ideal setting for *in situ* experimentation, however if this is not available, extended reality can fill this gap.

8.5.5 Background and Culture of Participants

The majority of the participants that took part in the studies of this thesis were University of Glasgow students. Although there was a mixture in nationality, a large portion were from Europe. These participants bring their particular sensibilities about stimuli, music in particular, which leads to these results being biased towards people who have grown up being exposed to Western musical traditions. For example, acoustic roughness or disharmony was a factor studied throughout this thesis, and in the Western musical tradition disharmony is generally avoided because it is displeasing perceptually, this is not necessarily the case for some people who grew up elsewhere [98].

Further to the variety of sensibilities about stimuli and music, different participants may bring a differing mental models to processing auditory and vibrotactile displays and this is another potentially interesting area of future work. As was mentioned earlier in this thesis, research has shown that participants who are visually impaired used differing mental models than sighted participants in their understanding of a data-sound mapping [157], however this has not yet been explored for other demographics. To make more accurate assumptions about data-sound mappings in particular, but also data-vibration mappings, taking into account the user's background and culture is necessary and this is something which has not been explored in depth in human-computer interaction research.

8.6 Conclusions

Ensuring that the acoustic or vibrotactile parameter used to represent information in an auditory or vibrotactile parameter mapping system is perceptually congruent with the information being presented is key to a successful interaction between a user and the system. This thesis investigated several novel psychoacoustic parameters and how they could be advantageous in a data-sound mapping. Further to this, this thesis explored how these parameters in addition to several vibrotactile parameters were perceived by users when they were implemented in a mapping. Finally, this thesis investigated how the complexity of the use-case of an auditory or vibrotactile display impacted the usability of the display and studied what impact designing these systems based on prior perceptual experiment results had on user performance. This research allows designers to better understand how to design data-sound or data-vibration mappings in such a way that the users of the mappings find the relationships between the data and sensory parameters to be congruent. It also allows designers to understand what factors of user performance are impacted by the mapping in addition to the complexity of the context in which the mapping is implemented in.

Appendices

Appendices for this thesis can be found at **this link** or go to url directly: `https://www.dropbox.com/sh/6kuq8ad47d8dmni/AABdDdlWyr0QSJpPNGc8HYKJa?dl=0`. This is a temporary storage location for this data during the examination process. For the final submission, this data will be hosted on the University's Enlighten system.

For each study the appendices consists of the following folders:

- Data
- Forms
- R Files
- Stimuli

Bibliography

[1] AFSHAR, P. F., BAHRAMNEZHAD, F., ASGARI, P., AND SHIRI, M. Effect of white noise on sleep in patients admitted to a coronary care. *Journal of caring sciences 5*, 2 (2016), 103.

[2] AKAIKE, H. Information theory and an extension of the maximum likelihood principle. In *Selected Papers of Hirotugu Akaike*. Springer, 1998, pp. 199–213.

[3] ALEXANDER, R. L. *The Bird's Ear View: Audification for the Spectral Analysis of Heliospheric Time Series Data*. PhD thesis, University of Michigan, 2015.

[4] ANDERSON, J. Creating an empirical framework for sonification design. In *Proceedings of the International Conference on Auditory Display* (2005), Georgia Institute of Technology.

[5] ARNAL, L. H., FLINKER, A., KLEINSCHMIDT, A., GIRAUD, A.-L., AND POEPPEL, D. Human screams occupy a privileged niche in the communication soundscape. *Current Biology 25*, 15 (2015), 2051–2056.

[6] BARRASS, S. Some golden rules for designing auditory displays. *in Csound Textbook, B. Vercoe and R. Boulanger, eds., MIT Press, Cambridge, Mass* (1998).

[7] BERGMAN, P., SKÖLD, A., VÄSTFJÄLL, D., AND FRANSSON, N. Perceptual and emotional categorization of sound. *The Journal of the Acoustical Society of America 126*, 6 (2009), 3156–3167.

[8] BLACK, D., HANSEN, C., NABAVI, A., KIKINIS, R., AND HAHN, H. A survey of auditory display in image-guided interventions. *International Journal of Computer Assisted Radiology and Surgery 12*, 10 (2017), 1665–1676.

[9] BLATTNER, M. M., SUMIKAWA, D. A., AND GREENBERG, R. M. Earcons and icons: Their structure and common design principles. *Human–Computer Interaction 4*, 1 (1989), 11–44.

[10] BLY, S. Presenting information in sound. In *Proceedings of the 1982 Conference on Human Factors in Computing Systems* (1982), ACM, pp. 371–375.

[11] BLY, S. Sound and computer information presentation. Tech. rep., Lawrence Livermore National Lab., CA (USA); California Univ., Davis (USA), 1982.

[12] BOLANOWSKI JR, S. J., GESCHEIDER, G. A., VERRILLO, R. T., AND CHECKOSKY, C. M. Four channels mediate the mechanical aspects of touch. *The Journal of the Acoustical society of America 84*, 5 (1988), 1680–1694.

[13] BRADSKI, G., AND KAEHLER, A. *Learning OpenCV: Computer vision with the OpenCV library*. " O'Reilly Media, Inc.", 2008.

[14] BREGMAN, A. S. *Auditory scene analysis: The perceptual organization of sound*. MIT press, 1994.

[15] BREWSTER, S. A. *Providing a structured method for integrating non-speech audio into human-computer interfaces*. PhD thesis, University of York York, UK, 1994.

[16] BREWSTER, S. A., AND BROWN, L. M. Non-visual information display using tactons. In *CHI'04 extended abstracts on Human factors in computing systems* (2004), ACM, pp. 787–788.

[17] BREWSTER, S. A., AND BROWN, L. M. Non-visual information display using tactons. In *CHI '04 Extended Abstracts on Human Factors in Computing Systems* (New York, NY, USA, 2004), CHI EA '04, ACM, pp. 787–788.

[18] BROWN, L. M., BREWSTER, S. A., AND PURCHASE, H. C. A first investigation into the effectiveness of tactons. In *Eurohaptics Conference, 2005 and Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems, 2005. World Haptics 2005. First Joint* (2005), IEEE, pp. 167–176.

[19] BROWN, L. M., BREWSTER, S. A., AND PURCHASE, H. C. Multidimensional tactons for non-visual information presentation in mobile devices. In *Proceedings of the 8th conference on Human-computer interaction with mobile devices and services* (2006), ACM, pp. 231–238.

[20] BROWN, S., AND JORDANIA, J. Universals in the world's musics. *Psychology of Music 41*, 2 (2013), 229–248.

[21] BURNHAM, K. P., AND ANDERSON, D. R. *Model selection and multimodel inference: A practical information-theoretic approach*. Springer-Verlag, 2002.

[22] BURNHAM, K. P., AND ANDERSON, D. R. Multimodel inference: understanding aic and bic in model selection. *Sociological methods & research 33*, 2 (2004), 261–304.

[23] CALTECH/MIT/LIGO. The sound of two black holes colliding. In *LIGO/CALTECH.edu* (February 2016). available at https://www.ligo.caltech.edu/video/ligo20160211v2, [last visited 01/05/20].

[24] CAMERON, J (DIR). Aliens (1986). 20th Century Fox.

[25] CANDEY, R. M., SCHERTENLEIB, A. M., AND DIAZ MERCED, W. Xsonify sonification tool for space physics. In *Proceedings of the International Conference on Auditory Display* (2006), Georgia Institute of Technology.

[26] CARTER, T., SEAH, S. A., LONG, B., DRINKWATER, B., AND SUBRAMANIAN, S. Ultrahaptics: multi-point mid-air haptic feedback for touch surfaces. In *Proceedings of the 26th annual ACM symposium on User interface software and technology* (2013), pp. 505–514.

[27] CHANDLER, D. *Semiotics: the basics*. Routledge, 2007.

[28] CHOLEWIAK, R. W., BRILL, J. C., AND SCHWAB, A. Vibrotactile localization on the abdomen: Effects of place and space. *Perception & psychophysics 66*, 6 (2004), 970–987.

[29] CHOLEWIAK, R. W., AND COLLINS, A. A. Vibrotactile localization on the arm: Effects of place, space, and age. *Perception & psychophysics 65*, 7 (2003), 1058–1077.

[30] COXE, S., WEST, S. G., AND AIKEN, L. S. The analysis of count data: A gentle introduction to poisson regression and its alternatives. *Journal of personality assessment 91*, 2 (2009), 121–136.

[31] CRAIG, J. C. Difference threshold for intensity of tactile stimuli. *Perception & Psychophysics 11*, 2 (1972), 150–152.

[32] CREATIVE ASSEMBLY. Alien: Isolation (2014).

[33] DE SAUSSURE, F. Cours de linguistique générale (1916.

[34] DI, G.-Q., CHEN, X.-W., SONG, K., ZHOU, B., AND PEI, C.-M. Improvement of zwicker's psychoacoustic annoyance model aiming at tonal noises. *Applied Acoustics 105* (2016), 164–170.

[35] DI CAMPLI SAN VITO, P., BREWSTER, S., POLLICK, F., WHITE, S., SKRYPCHUK, L., AND MOUZAKITIS, A. Investigation of thermal stimuli for lane changes. In *Proceedings of the 10th International Conference on Automotive User Interfaces and Interactive Vehicular Applications* (2018), pp. 43–52.

[36] DUBUS, G., AND BRESIN, R. A systematic review of mapping strategies for the sonification of physical quantities. *PloS one 8*, 12 (2013), e82491.

[37] EDWORTHY, J., HELLIER, E., ALDRICH, K., AND LOXLEY, S. Designing trend-monitoring sounds for helicopters: methodological issues and an application. *Journal of Experimental Psychology: Applied 10*, 4 (2004), 203.

[38] ENGINEERING ACOUSTICS INC. C-2. available at https://www.eaiinfo.com/product/c2/, [last accessed 01/05/20].

[39] FASTL, H. Psychoacoustics and sound quality metrics. In *INTER-NOISE and NOISE-CON Congress and Conference Proceedings* (1998), vol. 1998, Institute of Noise Control Engineering, pp. 3–10.

[40] FERGUSON, J. Bell3d: An audio-based astronomy education system for visually-impaired students. *Ripples on Everyone's Lips 1* (2016), 35.

[41] FERGUSON, S., CABRERA, D., BEILHARZ, K., AND SONG, H.-J. Using psychoacoustical models for information sonification. In *Proceedings of the International Conference on Auditory Display* (2006), Georgia Institute of Technology.

[42] FICHERA, L., PACCHIEROTTI, C., OLIVIERI, E., PRATTICHIZZO, D., AND MATTOS, L. S. Kinesthetic and vibrotactile haptic feedback improves the performance of laser microsurgery. In *2016 IEEE Haptics Symposium (HAPTICS)* (2016), IEEE, pp. 59–64.

[43] FIELD, T. *Touch*. MIT press, 2014.

[44] FREEMAN, E., ANDERSON, R., WILLIAMSON, J., WILSON, G., AND BREWSTER, S. A. Textured surfaces for ultrasound haptic displays. In *Proceedings of the 19th ACM International Conference on Multimodal Interaction* (2017), pp. 491–492.

[45] FREEMAN, E., BREWSTER, S., AND LANTZ, V. Do that, there: an interaction technique for addressing in-air gesture systems. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (2016), ACM, pp. 2319–2331.

[46] FRIJDA, N. H., ET AL. *The emotions*. Cambridge University Press, 1986.

[47] FROHMANN, L., WEGER, M., AND HÖLDRICH, R. Recognizability and perceived urgency of bicycle bells. In *Proceedings of the International Conference on Auditory Displays* (2018), Georgia Institute of Technology.

[48] GARCÍA RIBER, A. Sonifigrapher: Sonified light curve synthesizer. In *Proceedings of the International Conference on Auditory Displays* (2019), Georgia Institute of Technology.

[49] GAVER, W. W. The sonicfinder: An interface that uses auditory icons. *Human–Computer Interaction 4*, 1 (1989), 67–94.

[50] GAVER, W. W. What in the world do we hear?: An ecological approach to auditory event perception. *Ecological psychology 5*, 1 (1993), 1–29.

[51] GELDARD, F. A. Adventures in tactile literacy. *American Psychologist 12*, 3 (1957), 115.

[52] GENTNER, D. Mental models, psychology of. In *International Encyclopedia of the Social & Behavioral Sciences*. Pergamon, 2001, pp. 9683 – 9687.

[53] GESCHEIDER, G. A. Resolving of successive clicks by the ears and skin. *Journal of experimental psychology 71*, 3 (1966), 378.

[54] GESCHEIDER, G. A. Auditory and cutaneous temporal resolution of successive brief stimuli. *Journal of Experimental Psychology 75*, 4 (1967), 570.

[55] GINSBERG, J. H. *Acoustics: A Textbook for Engineers and Physicists*. Springer, 2018.

[56] GRAU, J. W., AND NELSON, D. K. The distinction between integral and separable dimensions: Evidence for the integrality of pitch and loudness. *Journal of Experimental Psychology: General 117*, 4 (1988), 347.

[57] GRIMES, D., TAN, D. S., HUDSON, S. E., SHENOY, P., AND RAO, R. P. Feasibility and pragmatics of classifying working memory load with an electroencephalograph. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (2008), ACM, pp. 835–844.

[58] GROND, F., AND BERGER, J. Parameter mapping sonification. *The sonification handbook* (2011), 363–397.

[59] GURNETT, D., KURTH, W., KIRCHNER, D., HOSPODARSKY, G., AVERKAMP, T., ZARKA, P., LECACHEUX, A., MANNING, R., ROUX, A., CANU, P., ET AL. The cassini radio and plasma wave investigation. In *The Cassini-Huygens Mission*. Springer, 2004, pp. 395–463.

[60] HAMEED, S., FERRIS, T., JAYARAMAN, S., AND SARTER, N. Supporting interruption management through informative tactile and peripheral visual cues. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (2006), vol. 50, Sage Publications Sage CA: Los Angeles, CA, pp. 376–380.

[61] HARRIS, J., VANCE, S., FERNANDES, O., PARNANDI, A., AND GUTIERREZ-OSUNA, R. Sonic respiration: controlling respiration rate through auditory biofeedback. In *CHI'14 Extended Abstracts on Human Factors in Computing Systems* (2014), ACM, pp. 2383–2388.

[62] HART, S. G., AND STAVELAND, L. E. Development of nasa-tlx (task load index): Results of empirical and theoretical research. In *Advances in psychology*, vol. 52. Elsevier, 1988, pp. 139–183.

[63] HATANO, S., AND HASHIMOTO, T. Booming index as a measure for evaluating booming sensation. In *Proc. Inter-Noise* (2000), no. 233.

[64] HENKELMANN, C. Improving the aesthetic quality of realtime motion data sonification. *Computer Graphics Technical Report CG-2007-4. University of Bonn* (2007).

[65] HERMANN, T., HUNT, A., AND NEUHOFF, J. G. *The sonification handbook.* Logos Verlag Berlin, 2011.

[66] HOFSTADTER, D., AND SANDER, E. *Surfaces and essences: Analogy as the fuel and fire of thinking*. Basic Books, 2013.

[67] HOFSTADTER, D. R. *Fluid concepts and creative analogies: Computer models of the fundamental mechanisms of thought.* Basic books, 1995.

[68] HOFSTADTER, D. R., ET AL. *Gödel, Escher, Bach: an eternal golden braid*, vol. 20. Basic books New York, 1979.

[69] HOGGAN, E. Crossmodal audio and tactile interaction with mobile touchscreens. In *Social and Organizational Impacts of Emerging Mobile Devices: Evaluating Use*. IGI global, 2012, pp. 249–264.

[70] HOGGAN, E., AND BREWSTER, S. New parameters for tacton design. In *CHI'07 Extended Abstracts on Human Factors in Computing Systems* (2007), ACM, pp. 2417–2422.

[71] HOGGAN, E., AND BREWSTER, S. A. Crosstrainer: testing the use of multimodal interfaces in situ. In *Proceedings of the SIGCHI conference on human factors in computing systems* (2010), ACM, pp. 333–342.

[72] HOGGAN, E., CROSSAN, A., BREWSTER, S. A., AND KAARESOJA, T. Audio or tactile feedback: which modality when? In *Proceedings of the SIGCHI conference on human factors in computing systems* (2009), ACM, pp. 2253–2256.

[73] HOLLAND, S., MORSE, D. R., AND GEDENRYD, H. Audiogps: Spatial audio navigation with a minimal attention interface. *Personal and Ubiquitous computing 6*, 4 (2002), 253–259.

[74] HUANG, Y., AND JIANG, W. The effect of exposure duration on the subjective discomfort of aircraft cabin noise. *Ergonomics 60*, 1 (2017), 18–25.

[75] HUTCHINSON, M. K., AND HOLTMAN, M. C. Analysis of count data using poisson regression. *Research in nursing & health 28*, 5 (2005), 408–418.

[76] IN HUMAN–COMPUTER INTERACTION WORKING GROUP, T. S. Transparent Statistics Guidelines, Feb 2019. (Available at https://transparentstats.github.io/guidelines).

[77] Ergonomics of human-system interaction — part 910: Framework for tactile and haptic interaction. Standard, International Organization for Standardization, 2011.

[78] JEON, M. An exploration of semiotics of new auditory displays: A comparative analysis with visual displays.

[79] JOHNSON-LAIRD, P. N., KANG, O. E., AND LEONG, Y. C. On musical dissonance. *Music Perception: An Interdisciplinary Journal 30*, 1 (2012), 19–35.

[80] JONES, L. A. *Haptics*. The MIT Press, 2018.

[81] JONES, L. A., AND SARTER, N. B. Tactile displays: Guidance for their design and application. *Human factors 50*, 1 (2008), 90–111.

[82] KAN, A., LITOVSKY, R. Y., AND GOUPELL, M. J. Effects of interaural pitch-matching and auditory image centering on binaural sensitivity in cochlear-implant users. *Ear and hearing 36*, 3 (2015), e62.

[83] KERN A, HEID C, S. W.-H. S. N. S. R. Uncoiled cochlea with basilar membrane. In *WikiMedia* (August 2008). available at https://commons.wikimedia.org/wiki/File:Uncoiled_cochlea_with_basilar_membrane.png, [last accessed 01/05/20].

[84] KILDAL, J., AND BREWSTER, S. A. Non-visual overviews of complex data sets. In *CHI'06 Extended Abstracts on Human Factors in Computing Systems* (2006), ACM, pp. 947–952.

[85] KOSCH, T., FUNK, M., SCHMIDT, A., AND CHUANG, L. L. Identifying cognitive assistance with mobile electroencephalography: A case study with in-situ projections for manual assembly. *Proceedings of the ACM on Human-Computer Interaction 2*, EICS (2018), 11.

[86] KRAMER, G. Some organizing principles for representing data with sound. *Auditory Display-Sonification, Audification, and Auditory Interfaces* (1994), 185–221.

[87] KRAMER, G., WALKER, B., BONEBRIGHT, T., COOK, P., FLOWERS, J., MINER, N., NEUHOFF, J., BARGAR, R., BARRASS, S., BERGER, J., ET AL. 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* (1999).

[88] KRUEGER, L. E. Tactual perception in historical perspective: David katz's world of touch. *Tactual Perception; A Sourcebook* (1982), 1–55.

[89] KUCHERA, B. How the nintendo switch's hd rumble makes tumbleseed feel real. In *Polygon* (May 2017). available at https://www.polygon.com/2017/5/1/15499328/tumbleseed-hd-rumble-nintendo-switch, [last accessed 01/05/20].

[90] LAKOFF, G., AND JOHNSON, M. *Philosophy in the Flesh*, vol. 4. New york: Basic books, 1999.

[91] LAWLESS, J. F. Negative binomial and mixed poisson regression. *Canadian Journal of Statistics 15*, 3 (1987), 209–225.

[92] LECHELT, E. C. Temporal numerosity discrimination: Intermodal comparisons revisited. *British Journal of Psychology 66*, 1 (1975), 101–108.

[93] LINDOSLAND. Equal-loudness contours created by lindosland using openoffice draw by reference to iso 226 :2003 standard. In *WikiMedia* (December 2005). available at https://commons.wikimedia.org/wiki/File:Lindos1.svg, [last accessed 01/05/20].

[94] LOEB, R. G., BRECKNELL, B., AND SANDERSON, P. M. The sounds of desaturation: a survey of commercial pulse oximeter sonifications. *Anesthesia & Analgesia 122*, 5 (2016), 1395–1403.

[95] LYON, R. *Designing for product sound quality*. CRC Press, 2000.

[96] MALUSKI, S., CHURCHILL, C., AND COX, T. J. Sound quality testing and labelling of domestic appliances in the uk. In *Proceedings of Internoise* (2004).

[97] MARKS, L. E., AND GESCHEIDER, G. A. Psychophysical scaling. *Stevens' handbook of experimental psychology* (2002).

[98] MCDERMOTT, J. H., SCHULTZ, A. F., UNDURRAGA, E. A., AND GODOY, R. A. Indifference to dissonance in native amazonians reveals cultural variation in music perception. *Nature 535*, 7613 (2016), 547–550.

[99] MCGLONE, F., AND REILLY, D. The cutaneous sensory system. *Neuroscience & Biobehavioral Reviews 34*, 2 (2010), 148–159.

[100] MEHLER, B., REIMER, B., AND DUSEK, J. A. Mit agelab delayed digit recall task (n-back). *Cambridge, MA: Massachusetts Institute of Technology* (2011).

[101] MERCED, D., AND WANDA, L. *Sound for the exploration of space physics data*. PhD thesis, University of Glasgow, 2013.

[102] MIGNONE, C. The singing comet. In *European Space Agency Blog* (November 2014). available at http://blogs.esa.int/rosetta/2014/11/11/the-singing-comet/, [last visited 01/05/20].

[103] MIGNONE, C. From hipparchus to hipparcos: A sonification of stellar catalogues. In *European Space Agency Blog* (September 2016). available at https://sci.esa.int/web/gaia/-/58311-from-hipparchus-to-hipparcos-a-sonification-of-stellar-catalogues, [last visited 01/05/2020].

[104] NAM, Y., AND KIM, J. A semiotic analysis of sounds in personal computers: Toward a semiotic model of human-computer interaction. *Semiotica 2010*, 182 (2010), 269–284.

[105] NEUHOFF, J. G. Perception, cognition and action in auditory displays. In *The sonification handbook*. 2011.

[106] NEUHOFF, J. G. Is sonification doomed to fail? In *Proceedings of the International Conference on Auditory Displays* (2019), Georgia Institute of Technology.

[107] NEUHOFF, J. G., AND HELLER, L. M. One small step: Sound sources and events as the basis for auditory graphs. In *Proceedings of the International Conference on Auditory Displays* (2005), Georgia Institute of Technology.

[108] NG, A., AND BREWSTER, S. A. Investigating pressure input and haptic feedback for in-car touchscreens and touch surfaces. In *Proceedings of the 8th International Conference on Automotive User Interfaces and Interactive Vehicular Applications* (2016), ACM, pp. 121–128.

[109] OMEGATRON. Waveforms. In *WikiMedia*. available at https://en.wikipedia.org/wiki/File:Waveforms.svg, [last accessed 01/05/20].

[110] OSWALD, D. Non-speech audio-semiotics: A review and revision of auditory icon and earcon theory. Georgia Institute of Technology.

[111] OWEN, A. M., MCMILLAN, K. M., LAIRD, A. R., AND BULLMORE, E. N-back working memory paradigm: A meta-analysis of normative functional neuroimaging studies. *Human brain mapping 25*, 1 (2005), 46–59.

[112] PEARCE, J. Selected observations on amusia. *European neurology 54*, 3 (2005), 145–148.

[113] PERES, S. C., AND LANE, D. M. Auditory graphs: The effects of redundant dimensions and divided attention. In *Proceedings of the International Conference on Auditory Display* (2005), Georgia Institute of Technology.

[114] PETOCZ, A., KELLER, P. E., AND STEVENS, C. J. Auditory warnings, signal-referent relations, and natural indicators: Re-thinking theory and application. *Journal of experimental psychology: Applied 14*, 2 (2008), 165.

[115] PIRHONEN, A., MURPHY, E., MCALLISTER, G., AND YU, W. Non-speech sounds as elements of a use scenario: a semiotic perspective. In *Proceedings of the International Conference on Auditory Displays* (2006), Georgia Institute of Technology.

[116] PLACK, C. J., OXENHAM, A. J., AND FAY, R. R. *Pitch: neural coding and perception*, vol. 24. Springer Science & Business Media, 2006.

[117] POLLACK, I., AND FICKS, L. Information of elementary multidimensional auditory displays. *The Journal of the Acoustical Society of America 26*, 2 (1954), 155–158.

[118] PONGRAC, H. Vibrotactile perception: examining the coding of vibrations and the just noticeable difference under various conditions. *Multimedia systems 13*, 4 (2008), 297–307.

[119] QUINTON, M., MCGREGOR, I., AND BENYON, D. Sonifying the solar system. In *Proceedings of the International Conference on Auditory Displays* (2016), International Community on Auditory Display.

[120] REYNAL, M., IMBERT, J.-P., ARICÒ, P., TOUPILLIER, J., BORGHINI, G., AND HURTER, C. Audio focus: Interactive spatial sound coupled with haptics to improve sound source location in poor visibility. *International Journal of Human-Computer Studies 129* (2019), 116–128.

[121] ROBINSON, D. W., AND DADSON, R. S. A re-determination of the equal-loudness relations for pure tones. *British Journal of Applied Physics 7*, 5 (1956), 166.

[122] RUTHERFORD, E. An electrical method of counting the number of α-particles from radio-active substances. *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character 81*, 546 (1908), 141–161.

[123] RYU, J., CHUN, J., PARK, G., CHOI, S., AND HAN, S. H. Vibrotactile feedback for information delivery in the vehicle. *IEEE Transactions on Haptics 3*, 2 (2010), 138–149.

[124] SAGOO, H. S. *Reduction of noise levels in vaccum cleaners*. PhD thesis, Aston University, 1988.

[125] SATO, S.-I., YOU, J., AND JEON, J. Y. Sound quality characteristics of refrigerator noise in real living environments with relation to psychoacoustical and autocorrelation function parameters. *The Journal of the Acoustical Society of America 122*, 1 (2007), 314–325.

[126] SCHNUPP, J., NELKEN, I., AND KING, A. *Auditory neuroscience: Making sense of sound*. MIT press, 2011.

[127] SCHWARZ, S., AND ZIEMER, T. A psychoacoustic sound design for pulse oximetry. In *Proceedings of the International Conference on Auditory Displays* (2019), Georgia Institute of Technology.

[128] SEIFI, H., AND MACLEAN, K. E. A first look at individuals' affective ratings of vibrations. In *World Haptics Conference (WHC), 2013* (2013), IEEE, pp. 605–610.

[129] SHAKERI, G., WILLIAMSON, J. H., AND BREWSTER, S. Novel multimodal feedback techniques for in-car mid-air gesture interaction. In *Proceedings of the 9th International Conference on Automotive User Interfaces and Interactive Vehicular Applications* (2017), ACM, pp. 84–93.

[130] SHAKERI, G., WILLIAMSON, J. H., AND BREWSTER, S. May the force be with you: Ultrasound haptic feedback for mid-air gesture interaction in cars. In *Proceedings of the 10th International Conference on Automotive User Interfaces and Interactive Vehicular Applications* (2018), pp. 1–10.

[131] SONY. Introducing dualsense, the new wireless game controller for playstation 5. In *Playstation Blog* (April 2020). available at https://blog.us.playstation.com/2020/04/07/introducing-dualsense-the-new-wireless-game-controller-for-playstation-5/, [last visited 01/05/2020].

[132] SPEETH, S. D. Seismometer sounds. *The Journal of the Acoustical Society of America 33*, 7 (1961), 909–916.

[133] STAFF, B. File:blausen 0809 skin tactilereceptors.png in medical gallery of blausen medical 2014. *WikiJournal of Medicine 1 2* (2014).

[134] STEVENS, S. S. The measurement of loudness. *The Journal of the Acoustical Society of America 27*, 5 (1955), 815–829.

[135] STROHMEIER, P., AND HORNBÆK, K. Generating haptic textures with a vibrotactile actuator. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (2017), ACM, pp. 4994–5005.

[136] TACTILE LABS. Haptuator mark ii. http://tactilelabs.com/products/haptics/haptuator-mark-ii-v2/, [last accessed 01/05/20].

[137] TEGHTSOONIAN, R., STEVENS, S., AND STEVENS, G. Psychophysics: Introduction to its perceptual, neural, and social prospects. *The American Journal of Psychology 88*, 4 (1975), 677.

[138] TOMLINSON, B. J., WINTERS, R. M., LATINA, C., BHAT, S., RANE, M., AND WALKER, B. N. Solar system sonification: Exploring earth and its neighbors through sound. In *Proceedings of the International Conference on Auditory Displays* (2017), Georgia Institute of Technology.

[139] TUTHILL, J. C., AND AZIM, E. Proprioception. *Current Biology 28*, 5 (2018), R194–R203.

[140] VAN DOREN, C. L., GESCHEIDER, G. A., AND VERRILLO, R. T. Vibrotactile temporal gap detection as a function of age. *The Journal of the Acoustical Society of America 87*, 5 (1990), 2201–2206.

[141] VAN ERP, J. B., VAN VEEN, H. A., JANSEN, C., AND DOBBINS, T. Waypoint navigation with a vibrotactile waist belt. *ACM Transactions on Applied Perception (TAP) 2*, 2 (2005), 106–117.

[142] VELIČKOVIĆ, P. Amplitude modulation. In *GitHub* (July 2016). available at ttps://github.com/PetarV-/TikZ/tree/master/Amplitude modulation, [last accessed 01/05/20].

[143] VELIČKOVIĆ, P. Frequency modulation. In *GitHub* (July 2016). available at https://github.com/PetarV-/TikZ/tree/master/Frequency modulation, [last accessed 01/05/20].

[144] VERRILLO, R. T. Vibrotactile thresholds for hairy skin. *Journal of Experimental Psychology 72*, 1 (1966), 47.

[145] VERRILLO, R. T., FRAIOLI, A. J., AND SMITH, R. L. Sensation magnitude of vibrotactile stimuli. *Perception & Psychophysics 6*, 6 (1969), 366–372.

[146] VERRILLO, R. T., AND GESCHEIDER, G. A. Perception via the sense of touch. *Tactile aids for the hearing impaired* (1992), 1–36.

[147] VICKERS, P. Ways of listening and modes of being: Electroacoustic auditory display. *arXiv preprint arXiv:1311.5880* (2013).

[148] VICKERS, P., HOGG, B., WORRALL, D., AND WÖLLNER, C. The aesthetics of sonification. In *Body, sound and space in music and beyond: Multimodal explorations*. Routledge, 2017, pp. 89–109.

[149] VON BÉKÉSY, G. Synchronism of neural discharges and their demultiplication in pitch perception on the skin and in hearing. *The Journal of the Acoustical Society of America 31*, 3 (1959), 338–349.

[150] WAGENMAKERS, E.-J., AND FARRELL, S. Aic model selection using akaike weights. *Psychonomic bulletin & review 11*, 1 (2004), 192–196.

[151] WALKER, B. N. Magnitude estimation of conceptual data dimensions for use in sonification. *Journal of Experimental Psychology: Applied 8*, 4 (2002), 211–221.

[152] WALKER, B. N. Consistency of magnitude estimations with conceptual data dimensions used for sonification. *Applied Cognitive Psychology 21*, 5 (2007), 579–599.

[153] WALKER, B. N., AND KRAMER, G. Mappings and metaphors in auditory displays: An experimental assessment. In *Proceedings of the International Conference on Auditory Display* (1996), Georgia Institute of Technology.

[154] WALKER, B. N., AND KRAMER, G. Ecological psychoacoustics and auditory displays: Hearing, grouping, and meaning making. *Ecological psychoacoustics* (2004), 150–175.

[155] WALKER, B. N., AND KRAMER, G. Mappings and metaphors in auditory displays: An experimental assessment. *ACM Transactions on Applied Perception (TAP) 2*, 4 (2005), 407–412.

[156] WALKER, B. N., KRAMER, G., AND LANE, D. M. Psychophysical scaling of sonification mappings. In *Proceedings of the International Conference on Auditory Display* (2000), Georgia Institute of Technology.

[157] WALKER, B. N., AND LANE, D. M. Psychophysical scaling of sonification mappings: A comparison of visually impaired and sighted listeners. In *Proceedings of the International Conference on Auditory Display* (1996), Georgia Institute of Technology.

[158] WALKER, B. N., AND NEES, M. A. Theory of sonification. *The sonification handbook* (2011), 9–39.

[159] WEINSTEIN, S. Intensive and extensive aspects of tactile sensitivity as a function of body part, sex and laterality. *The skin senses* (1968).

[160] WEISENBERGER, J. M. Sensitivity to amplitude-modulated vibrotactile signals. *The Journal of the Acoustical Society of America 80*, 6 (1986), 1707–1715.

[161] WETTACH, R., BEHRENS, C., DANIELSSON, A., AND NESS, T. A thermal information display for mobile applications. In *Proceedings of the 9th international conference on Human computer interaction with mobile devices and services* (2007), pp. 182–185.

[162] WILSON, G., AND BREWSTER, S. A. Multi-moji: Combining thermal, vibrotactile & visual stimuli to expand the affective range of feedback. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (2017), ACM, pp. 1743–1755.

[163] WILSON, G., FREEMAN, E., AND BREWSTER, S. Multimodal affective feedback: combining thermal, vibrotactile, audio and visual signals. In *Proceedings of the 18th ACM International Conference on Multimodal Interaction* (2016), ACM, pp. 400–401.

[164] WILSON, G., HALVEY, M., BREWSTER, S. A., AND HUGHES, S. A. Some like it hot: thermal feedback for mobile devices. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (2011), pp. 2555–2564.

[165] WOBBROCK, J. O., FINDLATER, L., GERGLE, D., AND HIGGINS, J. J. The aligned rank transform for nonparametric factorial analyses using only anova procedures. In *Proceedings of the SIGCHI conference on human factors in computing systems* (2011), ACM, pp. 143–146.

[166] YE, S. The science behind force touch and the taptic engine, Apr 2015.

[167] YOO, Y., YOO, T., KONG, J., AND CHOI, S. Emotional responses of tactile icons: Effects of amplitude, frequency, duration, and envelope. In *World Haptics Conference (WHC), 2015 IEEE* (2015), IEEE, pp. 235–240.

[168] YOST, W. A. Psychoacoustics: A brief historical overview. *Acoustics Today: A Publication of the Acoustical Society of America 11*, 3 (2015), 46–53.

[169] ZESKIND, P. S., AND COLLINS, V. Pitch of infant crying and caregiver responses in a natural setting. *Infant Behavior and Development 10*, 4 (1987), 501–504.

[170] ZIEMER, T., BLACK, D., AND SCHULTHEIS, H. Psychoacoustic sonification design for navigation in surgical interventions. In *Proceedings of Meetings on Acoustics 173EAA* (2017), vol. 30, ASA, p. 050005.

[171] ZIEMER, T., AND SCHULTHEIS, H. Psychoacoustic auditory display for navigation: an auditory assistance system for spatial orientation tasks. *Journal on Multimodal User Interfaces 13*, 3 (2019), 205–218.

[172] ZIENTARA, P. A., LEE, S., SMITH, G. H., BRENNER, R., ITTI, L., ROSSON, M. B., CARROLL, J. M., IRICK, K. M., AND NARAYANAN, V. Third eye: a shopping assistant for the visually impaired. *Computer 50*, 2 (2017), 16–24.

[173] ZWICKER, E. Subdivision of the audible frequency range into critical bands (frequenzgruppen). *The Journal of the Acoustical Society of America 33*, 2 (1961), 248–248.

[174] ZWICKER, E., AND FASTL, H. *Psychoacoustics: Facts and models*, vol. 22. Springer Science & Business Media, 2013.