Creativity, Exploration and Control in Musical Parameter Spaces

by

Robert H. Tubb

Submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy

School of Electronic Engineering and Computer Science
Queen Mary University of London
United Kingdom

January 2016

Statement of Originality

I, Robert Tubb, confirm that the research included within this thesis is my own work or that where it has been carried out in collaboration with, or supported by others, that this is duly acknowledged below and my contribution indicated. Previously published material is also acknowledged below.

I attest that I have exercised reasonable care to ensure that the work is original, and does not to the best of my knowledge break any UK law, infringe any third party's copyright or other Intellectual Property Right, or contain any confidential material.

I accept that the College has the right to use plagiarism detection software to check the electronic version of the thesis.

I confirm that this thesis has not been previously submitted for the award of a degree by this or any other university.

Signature:

R. Tubb

Date: 23/1/2016

Details of Collaboration and Publications

The experimental design, experimental work, analysis and writing of the following published works were carried out by Robert Tubb, under supervision of Simon Dixon.

Material in Chapter 5 was previously published in

Tubb, R. and Dixon, S. (2014). A Four Strategy Model of Creative Parameter Space Interaction. In *Proceedings of the Fifth International Conference on Computational Creativity*, pages 16-23.

Material in Chapter 6 was previously published in

Tubb, R. and Dixon, S. (2014). The Divergent Interface: Supporting Creative Exploration of Parameter Spaces. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, pages 227-232.

Tubb, R. and Dixon, S. (2014). A zoomable mapping of a musical parameter space using Hilbert curves. *Computer music journal*, 38(3):23-33.

Material in Chapter 7 was previously published in

Tubb, R. and Dixon, S. (2015). An Evaluation of Multidimensional Controllers for Sound Design Tasks. *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, 47-56.

For my wife

Abstract

This thesis investigates the use of multidimensional control of synthesis parameters in electronic music, and the impact of controller mapping techniques on creativity. The theoretical contribution of this work, the EARS model, provides a rigorous application of creative cognition research to this topic. EARS provides a cognitive model of creative interaction with technology, retrodicting numerous prior findings in musical interaction research. The model proposes four interaction modes, and characterises them in terms of parameter-space traversal mechanisms. Recommendations for properties of controller-synthesiser mappings that support each of the modes are given.

This thesis proposes a generalisation of Fitts' law that enables throughput-based evaluation of multi-dimensional control devices.

Three experiments were run that studied musicians performing sound design tasks with various interfaces. Mappings suited to three of the four EARS modes were quantitatively evaluated.

Experiment one investigated the notion of a 'divergent interface'. A mapping geometry that caters to early-stage exploratory creativity was developed, and evaluated via a publicly available tablet application. Dimension reduction of a 10D synthesiser parameter space to 2D surface was achieved using Hilbert space-filling curves. Interaction data indicated that this divergent mapping was used for early-stage creativity, and that the traditional sliders were used for late-stage fine tuning.

Experiment two established a 'minimal experimental paradigm' for sound design interface evaluation. This experiment showed that multidimensional controllers were faster than 1D sliders for locating a target sound in two and three timbre dimensions.

The final study tested a novel embodied interaction technique: ViBEAMP. This system utilised a hand tracker and a 3D visualisation to train users to control 6 synthesis parameters simultaneously. Throughput was recorded as triple that of six sliders, and working memory load was significantly reduced. This experiment revealed that musical, time-targeted interactions obey a different speed-accuracy trade-off law from accuracy-targeted interactions.

Acknowledgments

I would like to sincerely thank my supervisor Simon. Throughout the research he consistently asked challenging questions, and presented exactly the right balance of critique and open mindedness. I was very lucky to be allowed a large amount of freedom to pursue my interests, and he tolerated many tangents along the journey. Nevertheless he had an excellent eye for spotting potential dead-ends and woolly thinking.

I would like to thank my family, who have been an unfailing source of support throughout my entire life, and have been the source of love, emotional stability and even occasionally income. I have my mother to thank for my love of art, creativity and non-conformism, and my father to thank for my love of engineering, science, and rational thinking. I'm grateful to both of them for providing me with a good education, but more importantly the love of learning and discovery that remains with me.

Warm regards go to my long-time friends from Keele university, who are my second family, and among whom I feel most myself.

Lastly, my greatest thanks goes to my wife Barbara. Thanks for staying in London so that I could complete my studies, and providing a huge amount of support during the toughest times. Thanks for being very tolerant of me being a little absent minded at times, absorbed in music, technology or abstract thought! Thanks also go to Barbara for excellent advice about English grammar and spelling, despite her being a non-native speaker.

I'd like to thank all the musicians, DJs and event promoters that I have played music with and have had discussions with over the years about music and technology.

Respect goes out to you for keeping the flame of creativity alive during challenging times.

Thanks also go to the online community of developers who provide open-source code and helpful information voluntarily, for example the Pure Data and openFrameworks communities. Without these resources the development of the software used in this thesis would have been considerably harder.

Many thanks go to all those who gave up their time to participate in the experiments, and all those who gave considered and insightful comments on the work.

I'm very grateful to the School of Electronic Engineering and Computer Science at Queen Mary for providing the funding for this work, and providing an extremely stimulating multi-disciplinary environment to work and think.

Contents

List of Figures

List of Tables

CHAPTER 1

Introduction

1.1 The Controller

Can we control more than one thing at once? More specifically, can a musician simultaneously manipulate multiple aspects of a sound in a spontaneous yet controlled fashion? This question is the central thread that runs through this thesis. Addressing this question will raise further, more complex questions, which will require ideas taken from a wide range of scientific and creative disciplines to answer. What are the limits of the human brain's ability to control multiple parameters in parallel, and do these cognitive processes differ from controlling a single parameter? Are there speed (or other) gains to be made using this parallel mode? Are there certain ways of presenting parameters that our brains can more easily process? How can we measure the amount of control achievable with a controller? When is control *creative*? Can something be *out* of control but still be useful, creatively speaking?

In this thesis, we will tackle these questions by focusing the design of interfaces for the control of sound synthesiser timbre and the creation of electronic music. We shall investigate how the design of these interfaces has an effect on the creative process, how different types of interfaces may suit different stages of the creative process, and how interaction in a creative or musical fashion differs from everyday computer use. In particular, this work focuses on the difference between using single dimensional controls such as sliders and knobs, and multidimensional controllers with many degrees of freedom.

The overwhelming majority of commercial musical interfaces still involve editing musical parameters in a serial, one-at-a-time fashion. This is quite different from more traditional musical instruments, where performers play many notes simultaneously, and can often control multiple aspects of timbre in real-time. The difference between these control paradigms is quite marked. Not only in the practical senses, such as the way the musician learns their operation, how long the parameters take to adjust, or how accurately the musical details can be specified; but also the subjective 'feel' of these interfaces, what musicians are consciously attending to when using them, and even the kind of music that emerges from them. Do these more subjective experiences relate to the objective dimensional structure of the interface? Is it impossible to relate quantitative measures such as speed and accuracy to rich and complex artistic experiences, or can we construct a theory that bridges this semantic gap? With the increasing availability of new multidimensional input devices such as multi-touch screens, the Kinect, and the Leap Motion hand tracker these issues are becoming increasingly relevant for musical instrument designers, electronic musicians, digital artists and the HCI field in general.

The principal goals of this research were to:

1. Theorise about some of the mental processes that underlie the navigation of musical parameter spaces, and propose a cognitive model of these creative

search strategies.

2. In view of the above model, establish a methodology to evaluate interfaces with respect to their creativity-supporting ability.

3. Design a variety of novel interfaces using new multidimensional input devices to augment these different control strategies.

4. Quantitatively test these interfaces, via user studies of synthesiser sound design tasks.

5. Relate these results to existing issues in NIME (New Interfaces for Musical Expression) research.

First, we shall discuss the general motivations behind this thesis topic. If the above questions can be answered, is it just electronic musicians that benefit? What other knowledge is to be gained that may have benefits further afield? What draws people to study musical interaction?

1.2 Motivating Musical Interaction Research

It could be thought that the study of electronic musical interfaces is something of a niche subject. Put bluntly, why should we devote time and energy studying the workflow of a small minority of individuals producing music that, on average at least, hardly anyone will ever hear? Why should universities spend their limited budget on the research into exotic noise making contraptions that few people will ever use[1]? Will we increase the productivity of the world economy? Will it make humankind happier? Will we be delving into any fundamental scientific mysteries?

[1] Digital Musical Instruments often suffer from the "problem of the second performer" [McPherson and Kim, 2012; Jordà, 2004]

Surprisingly perhaps, I think the answer on all three counts is yes. In this section I argue that the New Interfaces for Musical Expression (NIME) research field has unique potential with regard to transforming the way we interact with computers, and potentially of great interest to researchers in the wider field of Human-Computer Interaction (HCI).

1.2.1 Computer Music as "Extreme HCI"

Musical performance is consistently mentioned as being a very special form of interaction by researchers in other fields. In psychology literature it is cited as being one of the pinnacles of human cognitive and motor abilities [Penhune and Steele, 2012; Barrett, 1998; Limb and Braun, 2008; Ericsson, 2006]. It is also an aspirational, and even therapeutic endeavour in that it may give rise to peak experience states such as Flow [Wrigley and Emmerson, 2013; O'Neill, 1999], and self actualisation [Maslow, 1968]. In the HCI literature it represents an example of a highly optimal interaction, in terms of speed, accuracy, engagement and embodiment [Buxton, 1997; Kirsh, 2013].

Reasons why electronic music interaction is an exciting proving ground for novel interaction technologies include the following:

1. *Extreme technological reliance*: without electronics, electronic music does not exist. Therefore the technology's benefits and disadvantages, and the effects of interface design on the creative process are brought into sharp relief.

2. *High user motivation*: The electronic music community are enthusiastic, motivated, novelty-seeking, and willing to put in time and effort in testing new technologies.

3. *Ability of users to reflect and critique*: The electronic music community is technologically literate, highly self-aware, and take a great interest in how

technology affects their creative process, as attested by responses to the surveyB.

4. *A single individual takes responsibility for the entire creative process.* The proliferation of easily accessible music software and hardware has empowered a large number of solo music producers, who alternately take on the roles of composer, producer, arranger, stage performer and audio engineer. This makes the study of the creative life-cycle simpler than for, say, orchestral music.

5. *Altered subjective states*: The creative process is very different from casual computer use. The brain is calling on all its resources: imagination, technical skill, emotion, experience and the desire for self expression. Mental states are heightened, but also quite delicate and fleeting, even indescribable. The way the mental states of the performer are conveyed to the audience is still mysterious, some would say impossible to investigate scientifically. What happens when these high-level mental states run up against the nuts and bolts of technology? How is it possible to design for psychological processes about which so little is known?

6. *Complex, high-dimensional spaces*: The parameter space of electronic music features many interacting perceptual dimensions. A single sound object may be the result of setting tens or even hundreds of controls. Other creative domains, such as writing, seem well served by current serial input devices such as the keyboard. Music poses far tougher questions about how best to navigate these higher-dimensional spaces.

7. *Obvious catastrophic failure*: If technology falls short in terms of speed, accuracy or flexibility, this becomes painfully highlighted when in musical situations. Any short-fall in this regard leads to real-time music becoming im-

possible. Music is an extreme use case: where interface ideas can be tested to destruction. Buxton [1997] noted:

> "There are three levels of design: standard spec., military spec., and artist spec. Most significantly, I learned that the third was the hardest (and most important), but if you could nail it, then everything else was easy. After my work with artists, my research career at the University of Toronto and Xerox PARC was relatively simple."

In addition, music technology research is a unique area in terms of the skills of the individual researchers that are drawn to it. As we shall see, an individual's ability to span multiple creative domains is considered an important factor contributing to transformational creativity [Csíkszentmíhályi, 2009; Simonton, 1996]. NIME research — being located at the nexus of art, music, philosophy, computer science, signal processing and cognitive science — seems to attract those individuals with both the explicit knowledge and implicit intuitions that could give rise to radical synthesis between these disparate disciplines. There is potential to narrow the divide between the "two cultures" [Snow, 2012] of art and science. Not only does this research field have a unique set of problems, but also a unique set of skills with which to solve them. The solutions to these problems may be of great benefit in other fields, as we shall discuss next.

1.2.2 Digital Productivity and Input Device Throughput

Much of the global economy is fuelled by human creativity. That is, by the human brain's ability to generate novel, useful concepts and artefacts. These novel ideas can then be replicated across society and utilised, ideally, to increase quality of life. Much of this creativity may go on in the minds of people having conversations, walking in the countryside, or mulling over a book. However, an increasing amount of creativity occurs whilst using a computer. Computers are a means — like writing, diagrams, and mathematical symbols — of extending the abilities of the mind [Clark and Chalmers, 1998]. The flexible externalisation of words, images, and numerical quantities can significantly augment cognitive abilities such as memory, spatial reasoning and numerical calculations.

Creativity is already augmented by computers to certain extent, but could it be augmented *better*? What is different about augmenting creativity compared to other cognitive processes? First of all it would seem we need a definition of what creativity is, and a model of the processes of which it comprises (this is the subject of Chapter 3). Creativity is complex, and our understanding of it still poor, but even given a most prosaic definition — that creativity is the production of some new information — we already see that the interface is crucial. The computer's interface is the means of transforming mental information into digital form. Enhancing the speed and fidelity of the connection between the creator's intentions and the data within the computer is an essential prerequisite to the development of digital creative artefacts. Therefore the bandwidth, or *throughput* of the human-computer input channel will have a significant effect on how long it takes for an idea to be realised, and quite possibly whether it is realised at all.

In recent years we have come to expect constant improvement in information technology. What progress are we making in increasing the throughput of these input channels? Some futurists, most famously Kurzweil [2005], extrapolate Moore's law (the exponential increase in information processing power throughout history) to apply to machine intelligence in general, and propose a "singularity", where the speed of technological intelligence outstrips our ability to understand it. Vinge [1993] claimed:

Figure 1.1: Fifty years of interface research has not resulted in many revolutionary designs reaching the mainstream, though perhaps a tendency to construct devices from plastic instead of wood. Top left: Douglas Engelbart's mouse [English et al., 1967]. Bottom left: a modern mouse (2014). Top right: the Minimoog (1970)[Pinch, 2003]. Bottom right: the Moog Sub 37 (2014).

> "Within thirty years, we will have the technological means to create superhuman intelligence. Shortly after, the human era will be ended."

But Moore's law is not universally applicable to all technologies, not even all information technologies. The speed of many non-informational processes appear stagnant, or even declining, for example the speed of inner city traffic or transatlantic passenger flights. But more surprisingly, human-computer input devices would appear to number among these stagnant technologies. Whilst the responsiveness and expressiveness of the graphical user interface (GUI) has undoubtedly improved, the physical channel between our hands and our machines has barely changed since the advent of the era of personal computing[2].

Why should interface throughput be stagnant? Due to the inflexibility of procedural knowledge, humans demand interfaces that are consistent with learned skills

[2] Will we have to wait Vinge's remaining "eight years" for a superhuman intelligence to arrive and redesign our input devices for us?

[John and Kieras, 1996; Shneiderman and Plaisant, 2004]. Therefore user interface metaphors can become "locked in", for example the QWERTY keyboard is locked in to preserve existing typing skills. In the musical domain, the multi-track recording studio paradigm expressed in most DAWs[3], or the rotary dials used to control software synthesis (soft-synth) parameters, seem also to have become locked in. Granted, making use of existing interface skill is important [Antle et al., 2009]; but it is a surprising fact that both the rotary knob synthesiser interface format and the computer mouse [Myers, 1998] are half a century old (Fig. 1.1). Despite the hundred million-fold increase in CPU processing power since the 60's, human to computer throughput has barely changed.

It does not seem to be the case that high-throughput devices exist and have merely been overlooked, or failed to gain mainstream acceptance. Rather, it appears that nobody has yet designed a device, or even a theoretical approach that will significantly increase throughput. In a recent study the mouse was compared to two more recent input technologies: a touch screen and a hand tracker. The mouse and touch-screen performance showed comparable throughput, free gesture was worse [Sambrooks and Wilkinson, 2013]. Are we even going backwards? Synthesiser interfaces made a move toward being virtualised in the computer in the late 90's, but consensus seems to be that this was a bad move: substantial numbers of musicians and manufacturers having back-pedalled to analogue technology [Barlindhaug, 2007]. Take, for example, promotional material such as this, from synthesiser manufacturer Korg[4], persuading us to return to 60's style analogue step sequencing using rotary pitch controls:

> "Liberate yourself from the numerically-bound parameter editing that's typical on a DAW; you'll enjoy truly musical inspiration as you train

[3]Digital Audio Workstations

[4]`http://www.korg.com/us/products/dj/sq_1/`

> your ears and concentrate on what your fingertips are doing. Don't miss the experience of music that's driven by your instinct."

Whether or not you buy these claims, what is interesting is that no less than four cognitive processes are mentioned: inspiration, training, concentration and instinct. Can we actually use cognitive science to investigate claims like these? Why would seemingly more advanced technology be a step backwards? Is there a mistake that interface designers keep making, or is increasing throughput impossible for some fundamental cognitive reason?

Witnessing even a moderately competent musical performance would persuade us that higher throughput is possible. To watch the performance of a concert pianist is to witness a virtuoso display of "space-multiplexed" [Fitzmaurice and Buxton, 1997] user input. In their discussion of high-performance interfaces, Despain and Westervelt [1997] estimate the throughput of a virtuoso pianist at 300 bits/s, contrasting with about 50 bits/s for a good typist. Estimates of throughput values for a mouse range from 4 to 5 bps [MacKenzie, 2015]. Epworth [2013] cites more conservative figures of around 25 bits/s for piano playing (and similar figures for the fastest ever typing speed). This would imply that the rate at which a musician can produce information, i.e. their potential for 'digital productivity' is much higher than that of the average computer user, and higher even than conscious processing itself, which has been estimated around 15-18 bits/s [Epworth, 2013]. The increase in speed that comes with virtuosity is obvious, but usually comes at a cost in training time: some tens of thousands of hours in the pianist's case [Ericsson, 2006]. So a key question is how much practice is required to reach a throughput greater than that of a standard 'serial' computer interface. The second and third experiments in this thesis tackle this question.

I argue that there are two main factors retarding interface throughput:

1. The lack of a methodology to measure the throughput of high-dimensional controllers.

2. Over reliance on conscious thought. A failure to design for implicit brain systems, and a refusal to accept technologies that require training of implicit skill.

In fact, almost all our thoughtlessly performed everyday movements, such as reaching, grasping an object, talking to a friend, or walking through a crowd, are likely to have an information rate exceeding that of a WIMP interface. "Thoughtlessly performed" is the key phrase here, however. There is considerable evidence that explicit thought processes have surprisingly limited bandwidth, consciousness itself is an information processing bottleneck [Epworth, 2013]. The fact that we constantly mistake our conscious abilities for our complete abilities is considered to have had a negative effect on interaction design [Nørretranders, 1991; Norman, 2002].

When using an overly complex and analytical interface a musician feels the bottleneck instinctively: almost like a "barrier" to their musical instinct. One of the goals of this thesis is to show that this barrier is not some eternally mysterious incompatibility between the fiery artistic temperament and cold digital technology; it is a failure to utilise high-bandwidth, subconscious cognitive[5] machinery: those sensorimotor brain modules that we use every day to turn our intentions into reality. Musical interaction may be an ideal experimental arena to investigate the bandwidth of conscious and unconscious control, and look at the effect of changes in throughput on the subjective experience of expressive performance. By studying and improving musical interaction — such that this invisible barrier to expression is removed — we

[5] Some psychology fields take "cognitive" to imply conscious explicit processing of symbolic information, and would say subconscious movement control is non-cognitive. I will use the word in its broadest sense, to refer to any information processing in the brain.

may discover secrets to unlocking the bandwidth and productivity of user interfaces across many other domains.

1.2.3 Towards More Rewarding Interactions

Digital productivity may or may not be a worthwhile goal, but there is a deeper, more philosophical slant to musical interaction research, which relates to what our technology is ultimately for. One of the interesting aspects of NIME research, as opposed to 21^{st} century technological culture in general, is that it raises questions about the motivation for computer use.

A large amount of computer science research seems to be driven by a tacit goal of automating human cognitive processes. The goal of a machine is to automate some laborious task that the human would have done, and do it faster and more accurately. According to this goal, the ideal musical interface would be a single button labelled "make music for me". But what if the task is not laborious, what if it is fun? What if the task has no fixed objective? What if the goal is self-expression? In these cases the assumptions that motivate automisation are undermined. Asking what interface designs produce most engagement between human and machine changes our perspective somewhat: suddenly the human is not something to imitate and replace, but someone to assist and inspire.

This thesis draws on a number of models of creativity taken from the computational creativity field, but is less focused on creating an *artificial* musician, and more in using these models to *assist* the musician. Artificial creativity is certainly a fascinating exploit, and one that will produce a great deal of valuable knowledge — some of which I have drawn on heavily in this thesis. But it is an endeavour that would not seem to contribute directly to our quality of life. Philosophically, my approach is to consider the expression of the human creative drive as the self evident good. This generates a very different stance toward the role of the computer. Rather than "doing the work of a man", the technology should be considered as a tool or a medium for the unfolding of the creative act. In modern society, we see the "work" as the goal, and plan our temporal experiences in order to achieve the completion of the work. In reality, peak experience is the goal, and we should plan our work such that we can enter and reside within this optimal cognitive state. Automating physically embodied, skilled work is therefore suspect [Morris, 2002]. By doing so, we are undermining the preconditions for the peak experience of Flow [Csíkszentmíhályi, 1991].

In this regard, the current work shares many sympathies with the "third wave" of HCI research [Harrison et al., 2007]. Third wave HCI deals with more affective concerns: embodiment, intimacy, and a reduced emphasis on work. On the other hand, the methodology presented in his thesis is firmly in the 2nd wave camp. With its emphasis on information processing, cognitive models and and quantitative measurement, this work may seem to hark back to an earlier, more aggressively reductionist era. I take the view that whilst newer approaches certainly have huge relevance for musical interaction, the second wave programme is far from complete[6], and far from becoming obsolete. Just because artistic concerns *appear* to be unamenable to a reductionist analysis does not mean they aren't. On the contrary, considering the accelerating progress in cognitive neuroscience, one might predict that the application of neuroscientific concepts to third wave concerns will become increasingly fruitful over the coming decades.

One novel approach taken in this thesis is that some concerns that may have been considered marginal in second wave HCI (see Harrison et al. [2007]) are addressed in a quantitative fashion, including indirect, fluid and multiple goals; an emphasis on skilled, embodied cognition; and an acknowledgement of curiosity, exploration

[6]See for example, the debate concerning Fitts' law and device evaluation in Section 2.4.1

and fun[7]. One of the theoretical contributions of this thesis is to propose how exploratory interaction may be thought of in information processing terms. In short, this work concurs with the critiques of task-based HCI methodologies that arise in the NIME literature [Johnston, 2011; Stowell et al., 2009; Dobrian and Koppelman, 2006] but seeks to extend reductionist quantitative models and methods such that these critiques can be addressed.

Millions of people now use digital technology to make music. Most of these musicians have no real chance of making a living from it. For some reason, they feel that it is is worth their while spending many hours working on intricately wrought sonic artefacts with little or no obvious utility. Obviously the activity is rewarding. Why? Perhaps if we knew why musical interaction was so motivating, other interactions could be designed to be equally rewarding. Consider the impact on, not just productivity, but job satisfaction if the office workers of the world were interacting with their computers with the virtuosity of a concert pianist. What if all our interactions in supermarkets, public transport networks and political systems could take on the qualities that make peak musical experience so intrinsically magical, joyful and frictionless?

Music, therefore, represents a way to interact with technology that *all* system design should aspire to emulate. If people are tremendously productive when they are doing it, and they feel good while doing it, then it seems a very good idea to research why this should be so. A greater understanding of the cognitive science of artistic interaction may one day uncover interaction design strategies that help bring about a radical transformation of computer use during working life.

[7] It may seem that the concept of fun is incommensurate with the information processing paradigm, but as we shall see in Chapter 3, information-theoretic considerations may shed considerable light on what fun really is, e.g. [Schmidhuber, 2010].

1.2.4 Scientific Questions

The third motivation for studying musical interaction is to answer some more fundamental scientific questions about the operations of the human brain.

The development of music throughout the ages is a fascinating case of technology inspiring and enabling creativity. First via new acoustic instruments with richer tones and increased playability, and then through increasingly sophisticated audio manipulation technology. The carved bones of neolithic flutes, the cast iron frame of the piano, magnetic tape recorders, digital samplers, and the computer studios of today have all utilised and reflected the science and technology of the day. All have transformed the very music we produce. To say that technology is just a tool by which we realise our internal ideas is therefore very wrong: it provides entirely new conceptual spaces to explore.

Scientifically, the observation that the technological means of realisation massively affects the creative process should immediately alert us to a potential 'way in' to investigating the mysteries of creativity. The experimenter can set up a number of experimental conditions by providing an artist with different technologies, and then can observe creative outcomes as a result of manipulating these conditions. Furthermore, if that technology can record data-traces of the path the artist takes through the space of possible solutions, then the researcher is given quantitative data to analyse for each experimental condition. They can then investigate in detail the smaller individual actions that make up the larger creative process.

If musical interaction experiments are carefully designed to test hypotheses about the structure of creative thought, they not only have the potential to find out which interface is better for creativity, but more importantly to find out *why*, and perhaps even shed light on the mental processes involved. This thesis sits alongside other work such as Nash [2012] and Jennings et al. [2011] in attempting this type of research programme.

1.3 Scope and Outline

'Deep not wide' is a guideline often heard with regard to approaching doctorate level research, with the implication that it is better to investigate a narrow field in depth than be spread too thinly. Nevertheless, as we shall see in the review of creativity research, diversity is essential to any creative endeavour. Diversity of influences and recombination of ideas are a vital aspects of creative cognition and the progress of human culture. Some research topics are necessarily multidisciplinary, and the NIME field is one. Diversity on its own is not useful however, rather it is the *synthesis* of remote concepts that provides insight, by connecting and explaining previously unrelated phenomena. In order to investigate how interfaces and creativity interact, I regard it necessary to draw on research from further up the hierarchy of scientific knowledge. Figure 1.2 illustrates the scope of both the background literature and the thesis. The principal focus of the experimental work is timbre control during a sound design task. This topic principally falls under the umbrella of New Interfaces for Musical Expression (NIME): the investigation of how an interface affects the production of music, and how technology can be used to enable musical expression and creativity. Next up is the wider field of HCI and interaction design: the art and science of producing interfaces that enable humans to accomplish their information manipulation goals; in particular content creation software, or "creativity support tools". These are computer systems that are designed to enable creative content composition and enhance or augment creativity. This in turn can draw on psychological research into creativity, which in turn requires background knowledge of cognitive science, the study of information processing in the brain.

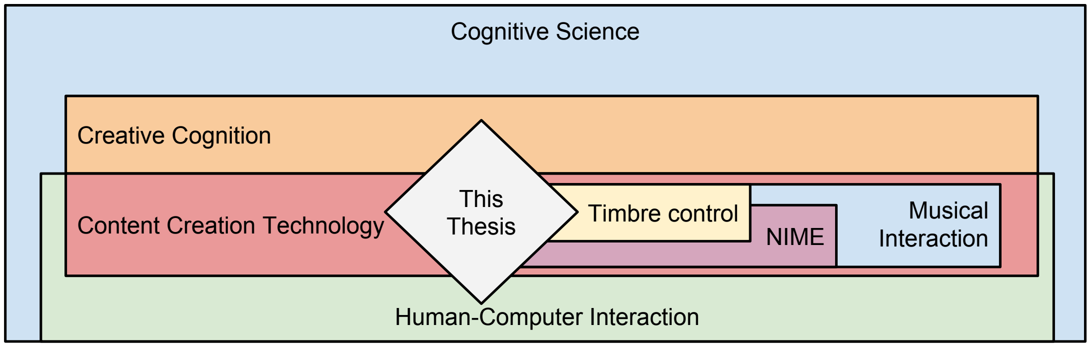


Figure 1.2: Illustration of the scope of the literature review, and the contribution of this thesis. This work, though ostensibly about timbre control, contains some contributions to HCI and creative cognition.

1.3.1 Thesis Outline

This thesis starts by presenting a broad overview of some fundamental concepts in cognitive science and creative cognition in Chapter 2. This includes the distinctions between implicit and explicit processing in the brain; perception, action and movement control; and embodied cognition. Also presented are information-theoretic approaches to Human-Computer Interaction such as Fitts' law.

In Chapter 3, models of *creative* cognition are discussed. Stage models such as the incubation-illumination model are reviewed. Complementary processes such as divergent and convergent thinking are outlined. Computational creativity models are presented, such as the Creative Systems Framework, blind variation and selective retention (BVSR), and the notion of creativity resulting from a drive to predictively code experience data.

A number of founding notions for the rest of the argument are distilled from the cognitive science literature:

1. The free-energy principle.

2. Dual Process theory, Global Workspace theory.

3. Embodied Cognition, affordances and active inference.

From HCI literature:

1. The cognitive mirroring principle.
2. The cognitive pipelining principle.
3. Interface bandwidth maximisation principle.

From the creative cognition literature:

1. PSVSR model of creative cognition.
2. The Creative Systems Framework.
3. Creativity as data compression.

Some literature regarding Digital Musical Instruments and interaction with music technology is reviewed in Chapter 4. In particular, design frameworks, evaluation methodologies, and work that has investigated the geometry of mappings between gestural controllers and the synthesiser engine parameters.

In the theoretical contribution, Chapter 5, a model of the interacting agent in a perception-action loop is developed. The creative process is portrayed in terms of a search through parameter space. The connection between a technological parameter space, a mental conceptual space, and a hypothetical fitness function is outlined. The role of entropy reduction in creative interactions is proposed. Artistic creativity is connected to the free-energy principle, and described as as an extension of an agent's desire to reduce surprise in their environment via creation of sensory data. Portraying the musical instrument as a communications channel leads to a role for Shannon information in musical expression. Consideration of information flow through the perception-action loop leads to the proposal that the potential expressivity of an instrument is closely related to the throughput achievable with the interface.

Fitts' law can be used to measure the effectiveness of interfaces for 1D and 2D target based interaction, but the methodology is the subject of some controversy. I therefore propose an alternative measure of throughput, one applicable to arbitrary n-dimensional search spaces. This is termed the "Index of Search Space Reduction", or ISSR. If a target is specified in advance, the rate of convergence on that target is best measured by the amount of search-space volume reduction that is achievable in a certain time. Given a large ensemble of recordings of users' search trajectories, the entropy of this distribution of search points can be calculated. The reduction of the entropy of this distribution as the searches converge on the target provides a measure of the average amount of information that 'flowed' from the participants through the interface. If an interface demonstrates higher throughput than another at an identical sound search task, it can then be said to be more effective. This approach deviates somewhat from the ISO standard for Fitts' law experiments. Whilst it may sacrifice the predictive aspect of Fitts' law, it has a number of useful advantages over the current standard for device evaluations.

In live performance, this communications channel connects to the audience, but perhaps more importantly also feeds back to the artist, and becomes a perception-action loop. I claim that the tight coupling of perception and action (via affordances, or active inference) results in the interface having a substantial influence on the interaction strategy. This has knock-on effects on the route the artist takes through solution space. Given that this route through parameter space is intimately related to the route through conceptual space, the geometry of the parameter mappings should reflect the geometry of the creative process in mental space. In other words, the interface should augment whatever creative strategy the user seeks to employ at a given time. This is an extension of the cognitive mirroring principle.

The second part of the theory chapter then looks at the nature of four creative strategies. In order to build a usable model of the geometry of creative thought processes, I draw a distinction between divergent and convergent interaction. I attempt to define these terms in a more rigorous way, with appeal to the cognitive models discussed in Chapter 3. Convergence is defined as solution space traversal driven by a prediction of increasing value. Divergence is traversal independent of these predictions. I then claim that both explicit and implicit thought can generate divergent and convergent strategies. This results in four 'quadrants', or parameter-traversal strategies, that have fundamentally different properties. This forms the EARS model of creative interaction. The four quadrants consist of Exploratory (implicit-divergent), Algorithmic (explicit-convergent), Reflective (explicit-divergent), and Skilled (implicit-convergent).

Once the connection between creativity and parameter space is made, this opens up the possibility to investigate the actual paths that musicians take through parameter space, and design navigation strategies (via different mapping geometries) conducive to creative results. I then offer a critique[8] of the current standard of uni-dimensional knobs and sliders or WIMP interfaces. I propose that these interfaces can be understood to cater for only one quadrant of the EARS model (the algorithmic/analytic). Analytic thought places excessive demands on working memory. This cognitive load may inhibit reflective quadrant processes, leading to the subjective experience of "loss of perspective" and an interference with high-level aesthetic goals.

The theoretical background established, we then move onto the design and implementation of some novel interfaces, and experimental studies that aim to test their effectiveness. The investigation of alternative interfaces, ones that are design for the "exploratory" and "skilled" quadrants, in direct comparison to standard "analytic"

[8] A development of the critique found in Hunt and Kirk [2000]

interfaces, then forms the methodological basis for the three experiments.

Experiment 1, presented in Chapter 6, investigated the idea that divergent and convergent phases of creative interaction require very different mapping strategies. For this experiment, a novel, exploratory interface (Sonic Zoom) was developed, which enables lossless dimension reduction of a synthesiser parameter-space to two dimensional surface. The mapping is lossless in that it preserves access to the entire combinatorial space of the high dimensional interface: this is achieved using Hilbert space-filling curves. This 2D, zoomable interface is then provided alongside a more traditional slider-based interface and released into the wild. The exploratory interactions of over 400 users were logged, and a survey conducted. This experiment revealed that the divergent mapping was indeed preferred for early-stage creativity, and that the traditional sliders were used for late stage honing. Survey responses revealed that users felt this interface did enhance the exploratory aspect of sound design.

Experiment 2, in Chapter 7, attempted to establish a 'minimal experimental paradigm' for investigating the difference between skilled and algorithmic interaction quadrants. The principal aim was to compare separate, 1D parameter controls (touchscreen sliders) to multidimensional controllers (an XY touchpad for 2D, the Leap Motion for 3D), and determine whether the multidimensional controllers were more suited to skilled interaction. This was carried out via a target matching sound design task. Subjects had to match randomly generated target sounds as quickly and accurately as possible by setting the parameters to the right values. Results showed that after about two hours of practice, the XY pad was 9% faster than two sliders for no accuracy loss, and the Leap was 17% faster than 3 sliders with 9% accuracy loss. The results of this experiment were analysed using the Fitts' law based ISSR methodology presented in Chapter 5.

The final experiment (Chapter 8) built on the second, in that it was again a sound matching task, and again tried to ascertain if multidimensional controllers were more effective for skilled interaction. This study, however, altered the task to be tempo-based: users had to match targets in time to a metronome. It also used 6 degrees of freedom, and only 8 target sounds, in order to more specifically investigate well practised movements in a higher dimensional space. A system for learning high-throughput interactions by means of visually matching hand poses was developed: ViBEAMP[9]. This system allowed participants to rehearse matching 6 parameters at once by lining their hand up with a visual target; subjects would then perform the target sequence again without guides to test how well they could be memorised. This task revealed large differences between interface types, with the leap motion being far more effective at fast matches than 6 sliders, in faster conditions demonstrating over three times the throughput. The leap also exhibited far less working memory load, as tested by participants having to perform a secondary task of memorising and recalling sequences of up to 3 sound targets. Plotting entropy reduction against movement time reveals that tempo-based interactions do not conform to Fitts' law. Rather they conform to a linear relationship between movement velocity and absolute accuracy (sometimes referred to as the "Schmidt paradigm"). This resulted in a throughput peak at a particular tempo.

In Chapter 9 I discuss the experimental results as a whole. I propose that exploratory, algorithmic and skilled modes have different quantitative signatures when analysed in terms of entropy reduction. Detection of these signatures may enable researchers to investigate which modes are being used at any one time within interaction data at longer time-scales for real creative projects. Detection of steady algorithmic progress, exploratory wandering, and even moments of 'inspiration' may be possible. I reflect upon how the experimentally untested 'reflective' mode of creative interaction may be designed for and evaluated. I sketch out how all 4

[9]Virtual Body Emulation Assisted Multidimensional Performance

quadrants, and the transitions between them, might be provided for in a single system, and open research challenges in doing this. This leads to recommendations of how designers can provide technology that better supports musicians' creativity.

Finally, in a more speculative discussion, I extrapolate from the 9 cognitive principles, and investigate the idea that these modes might display these distinctive throughput signatures even *within* the brain: in conceptual spaces of vastly higher dimensionality. This may have interesting implications for creative cognition, and the nature of inspired states such as Flow.

The thesis concludes with a discussion of future directions for this research project.

As background motivation for this work, a survey of 45 electronic musicians was conducted. This investigated their attitudes to how technological interactions relate to the creative process. The results of this survey serve to link the abstract problems addressed by the theoretical work in Chapter 5 to the real-world problems faced by musicians. The responses are analysed in terms of the EARS theory. The responses showed that throughput and cognitive load are important issues to address, and that EARS is a useful framework with which to investigate musical interaction issues. Results from this questionnaire are presented in Appendix B.

1.3.2 Novel Contributions

This thesis makes a theoretical contribution to musical instrument research by:

1. Providing a condensed and concise theory of creative interaction with music technology: the EARS model. This draws on some state-of-the-art developments in the cognitive science of creativity, and presents them in a way so as to be useful for the design of interactive systems. This theory retrodicts and clarifies some of the issues in the NIME research field.

2. Providing some recommendations for design of musical interfaces that follow from the above theory. This is done by describing four interaction modes that need to be supported in music production software, detailing the relationships between them, and proposing geometrical properties that controller to synthesis engine parameter mappings should possess in order to support them.

3. Proposing a method for measuring the speed, accuracy and flexibility of multidimensional musical interfaces in terms of information throughput.

Empirical results that contribute to the NIME field include:

1. Implementing and testing an interface specifically designed for divergent, exploratory interaction. The value of this approach is tested 'in the wild' via a publicly available mobile app.

2. Establishing a minimal experimental paradigm for multidimensional controller evaluation, revealing differences between standard and multidimensional controllers.

3. Proposing a high-bandwidth embodied interaction technique: the ViBEAMP system, and experimentally confirm its effectiveness for timbre performance.

4. Revealing that musical performance, as in timely, rhythmic interaction, is unique in the HCI field. Time-based interactions obey a different speed/accuracy trade-off law from accuracy-based interactions such as mouse pointing.

This thesis also makes contributions to the more general HCI field by:

1. Proposing an way to measure the effectiveness of high-dimensional continuous parameter interfaces. This approach provides a new perspective on Fitts' law,

and a clarification of some of the issues surrounding its application to interface evaluation.

2. Investigating some causes of interface-induced working memory load.

3. Demonstrating that multidimensional controllers such as hand trackers provide a route to higher bandwidth interaction, and proposing an easier way to train users to make effective use of them, via virtual hand pose imitation.

CHAPTER 2

The Interacting Brain: Cognitive Science and HCI

2.1 The Interacting Brain

This chapter reviews necessary elements of psychology and cognitive science, as they apply to Human-Computer Interaction, creativity support technology, and electronic music production. Attempting to characterise the human brain in a simple way poses some immediate problems. Firstly, being the most complex system in the known universe, the brain's operations are still full of mystery. Even with modern brain imaging techniques such as fMRI, and increasingly sophisticated neurologically inspired computational models, researchers are still just beginning to understand how the brain supports complex human behaviour. Secondly, the literature is vast and diverse: there are many ways to approach the study of behaviour, and many levels of analysis; from individual neural computations—already a immensely intricate system of molecular biology—right up to the study of culture and society: where interactions between many minds over human history have formed systems of immense complexity.

A third difficulty is that restricting the scope to research that has relevance to music production scarcely narrows the focus at all. Musical creativity is an extremely complex behaviour and makes use of almost all human faculties. Music is a 'whole brain' phenomenon [Ball, 2010; Limb and Braun, 2008]. Likewise, interacting with computers requires a huge variety of cognitive processes [Card et al., 1983]. Almost all of what is omitted here may surely be relevant for musical interaction: such topics as emotion [Juslin and Sloboda, 2010], music perception [Pearce, 2005], linguistic and semantic thought [Koelsch et al., 2004], social bonding [Freeman III, 1998], and many more. However, as mentioned in the introduction, we restrict the scope by focusing on a particular time-scale, ranging from the motor actions involved in musical performance (around 100ms), to the completion of moderately complex musical components (less than an hour). The eliminates consideration of the behaviour of neurons and low-level brain architecture, and also wider considerations of long term artistic development, cultural value systems and so on.

The caveats above do not imply that a broad review of cognitive science is useless for the study and design of creative musical interaction. On the contrary, the field of Human-Computer Interaction (HCI) is necessarily based on knowledge of human cognitive abilities [Card et al., 1983]. A wide range of key concepts in HCI such as embodied cognition, affordances, working memory load and the speed-accuracy trade-off can be directly motivated by considerations of cognitive architecture. In fact, almost all areas of design can benefit from better understanding of how our the mind works [Forsythe et al., 2014; Norman, 2004].

Whilst studying the mind is an immense challenge, the past few decades have seen an astounding rate of progress in the cognitive sciences. The various disparate strands of psychology are, as one would hope from any scientific endeavour, increasingly converging into a more unified picture. This review attempts to present this unified picture, and make it accessible for digital music researchers. Cognitive science and neuroscience are perhaps not as intimidating as might be imagined outside the field. The real challenge in cognitive science is obtaining significant evidence for theories, rather than the conceptual difficulty of the theories themselves.

This chapter is more of a broad 'review of reviews', rather than a presentation of the state of the art in any particular specialism. Little attention will be paid to the history and development of psychological theories. The experimental evidence for these models, from brain lesion studies, brain imaging, animal and human behavioural studies and so on is not discussed. Also omitted are the open controversies and scientific debates within the field. The aim here is to provide a brief overview of the most relevant, well established, and concise theoretical constructs in order to inform the theory of creative musical interaction in Chapter 5. As these theories are presented, their relevance to Human-Computer Interaction and musical situations will be considered.

Readers more familiar with cognitive science may proceed to Section 2.4, which looks at Human-Computer interface design principles and evaluation methods in more detail. Section 2.4.1 will investigate information-theoretic ideas of human action such as Fitts' law. Fitts' law finds wide application in HCI, and will be relevant to the proposed methods of quantitatively evaluating musical interfaces in later chapters. Prior work with multidimensional controllers will be discussed.

Finally, we summarise by outlining and justifying the assumptions and simplifications we will use for the rest of the thesis, outlining 3 principles of cognition, and 3 principles of cognitively-inspired interaction design. These form the basis for the review of creative cognition in Chapter 3, and the creative interaction model in Chapter 5.

2.2 Perceiving and Acting

2.2.1 What is the Brain For?

Fundamentally, what distinguishes brains from other biological organs is that it is primarily responsible for controlling actions [Wolpert et al., 2001]. Actions are complex, coordinated physical movements that maximise the chances of the survival and flourishing of an animal[1]. In order to generate meaningful actions, the organism must obtain relevant information from the environment, by means of sensory perception. It must process and transform that information: in order to generate the optimal actions given the state of the world. Action is costly, in that it requires energy, which is in limited supply. Also costly is information processing, for example, the human brain consumes around twenty percent of the body's energy budget [Laughlin et al., 1998]. Therefore there is considerable selection pressure for brains to process information in the most efficient way possible, as well as generating the most efficient actions [Attwell and Laughlin, 2001].

Once the ability to sense and act upon the environment is established, it becomes advantageous to evolve more sophisticated processing, to address such issues as:

- What aspects of the sensory information are most relevant?
- How can I move in order to better sense the environment?
- How can I predict the environment without direct access to the relevant sensory data?
- If I act in a certain way, what will the effects be?
- How can I do all this using as few resources as possible?

[1] Or rather the genes of the animal [Dawkins, 2006]

Consideration of possible solutions to the above problems already hints at the complexity that can arise from the simple ability to sense and act, and the importance of predictive abilities. The continuity of evolution gives us hope that higher forms of cognition can be understood better by looking at how the brain carries out simpler actions. From basic cognitive functions such as memory and reinforcement learning, to more sophisticated human behaviours such as curiosity, exploration, science, art, and even electronic music: all are likely to be elaborations of this fundamental perception-action loop, indeed models of musical interaction often feature action-perception feedback [Leman, 2008; Armstrong, 2006; Nash, 2012; Wessel and Wright, 2002].

Efficiency of information processing is closely related to predictive coding and information theory. In the same way that it would be inefficient to email an uncompressed video file, it would be highly inefficient to perceive or act upon uncompressed data. Data with causal regularities can be compressed by using predictive coding [Bar, 2007], and hence also be used to anticipate future events in an efficient manner Schmidhuber [2009]. The picture of the brain as a hierarchical, predictive model generator is seen as an important unifying principle in modern cognitive science [Dietrich and Haider, 2014; Friston, 2010; Clark, 2013], and is returned to in Section 2.2.7.

2.2.2 Brain Modules and Networks

The brain is not an amorphous mass of thinking material. The modular mind hypothesis [Fodor, 1983; Meunier et al., 2010], which, in various forms, has been part of psychology since its earliest days, is the conjecture that the brain is composed of various subsystems, each with its own functional specialisation. It is the concerted action of these systems working in unison that gives rise to more complex forms of

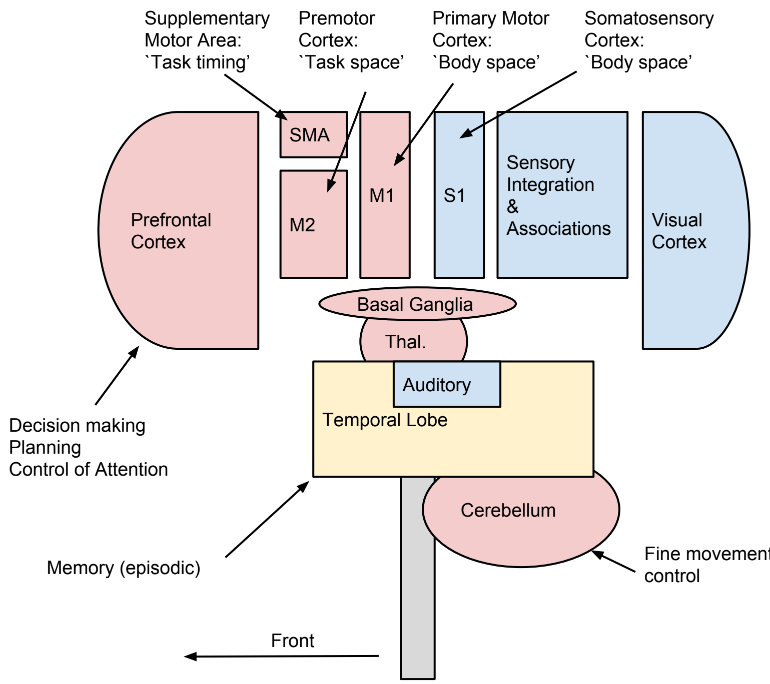


Figure 2.1: Highly simplified layout of some important brain structures. The eyes would face to the left. Some regions with a perceptual role are shown in blue, some regions with a role in producing actions in red. The temporal lobe is mainly devoted to episodic and semantic memory.

thought. Whilst not all cognition is encapsulated and modular, there certainly exist regions of the brain devoted to specialised processing, and a number of aspects of behaviour are well localised in specific anatomical structures. Figure 2.1 shows a highly simplified version of the brain, illustrating some regions that will be referred to later. A number of modules are shown in their approximate anatomical layout, principally the main regions relevant to perception and action.

The emerging science of brain networks investigates the connections between brain regions. Techniques such as Diffusion Tensor Imaging (DTI) enables researchers to map the longer range connectivity (white matter) between regions of the cortex (grey matter) [Le Bihan et al., 2001]. The complete map of these connections is the so called ‘human connectome’ [Sporns et al., 2005]. The network architecture of the connectome resembles a 'small world' network: a graph structure that features many local connections but fewer long range connections. Hierarchically modular, small world topology provides resilience in complex systems [Meunier et al., 2010]. Minimising the number of long distance connections is important for efficiency reasons, as they are costly to maintain, and introduce signalling latencies. Modular processes, then, are anatomically localised, encapsulated, fast, specialised, and resource-light. Conversely, brain-wide processing utilises many modules, involves long-range connections, and is slow and resource-hungry. However, this connectivity results in far more flexible and generalised processing due to the ability to integrate information from very different memory sources and sensory modalities. The encapsulated nature of smaller, less connected modules enables them to run independently of one another, whereas the central 'hubs' and multi-module network operations necessarily suffer more bottlenecks in their operation [Marois and Ivanoff, 2005]. Nørretranders estimates that the ratio of subconscious to conscious throughput may be as much as 10^{10} to 16 bits per second, and claims that user interfaces being limited to the speed of conscious processing is therefore highly inefficient [Nørretranders, 1991].

One of the most significant brain network findings in the last few decades is that there exist two major networks in the brain that associate with conscious processing. These networks are based around the 'rich-club' of some highly interconnected nodes in the brain network, principally those in frontal regions (attentional control), temporal regions (episodic and semantic memory) and parietal regions (associative multi-modal processing and self-perception). Consciously reportable mental states are strongly associated with these large networks [Dehaene, 2014]. The Default Mode Network (DMN) is associated with the brain's resting state [Buckner et al., 2008]. The wakeful brain is rarely 'resting' as such, rather, it tends to be engaged in 'mind wandering' [Mooneyham and Schooler, 2013; Schooler et al., 2011; Baird et al., 2012]. When the subject is not engaged with any particular task, they will often be ruminating over past events, simulating future events, conducting an internal monologue, or merely wandering through an associative chain of thoughts: what we would call a 'stream of consciousness'. When subjects are engaged with an external task and attending to their environment, another network is active: the Task Positive Network (TPN). Activation in these two networks are generally anti-correlated: in other words we are either engaged with carrying out a task, or we are mind-wandering, but not both at the same time. Intriguingly though, some brain imaging experiments appear to show that creative improvisation simultaneously utilises some components of both networks [Ellamil et al., 2012].

This distinction between localised, unconscious modules running in parallel, and wide, conscious networks being serial will surface repeatedly in later discussions of Dual Process Theory (Section 2.3.5), consciousness, attention and Global Workspace Theory (Section 2.3.4), Creative insight (Section 3.3), and indeed the rest of this thesis.

2.2.3 The Perceptual Hierarchy

An obvious example of a specialised, hierarchical set of brain modules—one of the most well mapped areas within the brain—is the visual system. Visual perception is the most information rich of our senses, and demands the most processing power: the visual cortex accounts for around 30% of cortical neurons (as opposed to just 3% for the auditory cortex). The optic nerve projects from the back of the eye onto the cortex at the rear of the brain, the primary visual cortex. It then joins groups of neurons, arranged in a retinotopic map of the visual field: that is, the relative positions in the retinal image are mapped on to the physical arrangement of neurons in a highly organised, geometry-preserving fashion. It then, as it passes further towards the front of the brain, goes through various stages of further refinement: from basic feature extraction, such as edge and movement detection [Hubel and Wiesel, 1968], to more sophisticated functions, such as object recognition, which occurs via the lower path (the 'ventral stream'). Object recognition has been shown to use distinct regions for distinct categories of stimuli, such as faces, tools and animals [Perani et al., 1995]. Hence, some of the neural architecture of the brain actually resembles the structure of conceptual categories. The upper path (the 'dorsal stream') undergoes completely different processing, more related to spatial awareness, movement and actions [Wilson, 2002]. In addition to propagating up the hierarchy, information also propagates downwards: that is, processing of the perceptual information is dependent on higher level expectations in addition to low-level input features [Bar et al., 2006; Clark, 2013]. Therefore, sensory processing involves a combination of top-down and bottom-up information flow: resulting in highly non-linear feedback processes.

Topographical maps are found in numerous regions of the brain [Kaas, 1997], for instance both the somatosensory and the primary motor cortex contain maps of the body[2], and the auditory cortex contains a 'tonotopic' map of the frequency spectrum. The other sensory modalities also undergo similar processing, filtering to remove irrelevant information, and extraction of distinct meaningful objects.

Beyond these specialised sensory pathways, perception becomes increasingly multi-modal. Areas in the parietal cortex bind sensory information into representations that begin to resemble concepts. These concepts form an associative network. Multi-sensory experiences are stored as 'episodic' memory. Concepts are associated with symbolic languages and grammars and stored as 'semantic' memory. The binding of complex multimodal information into individual, efficiently coded units is

[2] often referred to as 'homunculi' due to having the structure of a miniature, somewhat distorted human being.

called ‘chunking’ [Miller, 1956; Bor and Seth, 2012; Mathy and Feldman, 2012]. New chunks are being formed all the time, both explicitly, via attentional processes, and implicitly, via consolidation. Related chunks form an associative, semantic network [Wickelgren, 1979]. This associative network is thought to be traversed in a fast, massively parallel fashion by the subconscious process of ‘spreading activation’ [Anderson, 1983]. This associative traversal will be discussed in relation to mechanisms underlying creativity and insight 3.3.

Therefore, much of the brain is devoted to taking in sensory information, and inferring from it a concise explanation of what it observes in terms of well coded concepts. Thus the world, as presented to higher level awareness, tends to arrive already chunked into action-meaningful units. As a corollary of this, most of the information that is potentially available to us through our senses is simply thrown away, or at least never reaches the higher levels of conscious awareness, resulting in the phenomenon of ‘inattentional blindness’ [Simons and Chabris, 1999].

2.2.4 Perceptual Dimensions

When electronic musicians interact with synthesis algorithms, they are adjusting parameters that affect the resulting perceptual qualities of the sound. These qualities could be thought of as ‘perceptual dimensions’ [Garner, 1970]. Perceptual dimensions are qualities that can differentiate perceptions and conceptualisations of perceived objects. For instance, whilst colour is physically just a mixture of photons of differing wavelengths, perceptually it can be expressed as a point within a three-dimensional “perceptual space” (often represented computationally as red-green-blue (RGB) or hue-saturation-brightness (HSB) space).

Physical properties of sound map onto perceptual dimensions. The physical quantity of frequency maps to the perceptual dimension of pitch, amplitude maps to loudness, and so on. The exact nature of this mapping can be rather complex. Pitch, one of the most salient dimensions, has already a dual topological nature. Pitch is both a linear-continuous (low to high pitch) and circular-discrete (chroma-key) quantity. Distance metrics can therefore be difficult to ascertain: should notes an octave apart be considered closer together (chroma-similar) than two notes separated by a semitone (frequency-similar)? When timbre is considered, the situation becomes even more challenging. In-depth discussions of timbre can be found in [Stowell, 2010; Osaka, 2004; Grey, 1977].

Perceptual dimensions have been conjectured to fall along an axis of separable to integral [Garner, 1970]. Separable dimensions are those dimensions that are easily manipulated separately, for instance size and colour. Integral dimensions tend to be perceived and processed holistically, for instance hue and brightness. Distance metrics for these different qualities are also assigned differently. Separable dimensions are judged by a city-block (or Manhattan distance) similarity metric, whereas integral ones seem to be estimated by a Euclidean metric [Dunn, 1983].

This distinction has been shown to be important for HCI by Jacob et al. [1994]. They conducted an experiment to test the different performance of various control devices, and found that the integral dimensions were best controlled by multidimensional controllers (see Section 2.4.2), and the separable dimensions best controlled by independent, one-dimensional controllers. This indicates that the structure of the interface should match the perceptual nature of the task. This is a key result for our current investigation, as it implies that multi-dimensional controllers will be more suitable for timbre navigation [Vertegaal and Eaglestone, 1996]. Timbre dimensions are almost certainly integral. Even pitch and loudness show evidence of integrality [Grau and Nelson, 1988; Melara and Marks, 1990], however there is evidence that musical training influences how well the dimensions can be separated [Pitt, 1994].

Perceptual dimensions have also been used as a foundation for more general models of cognition. Gärdenfors' notion of 'conceptual space' [Gärdenfors, 1996, 2004], is an important one, as it makes the connection between mental concepts (chunks), and the dimensions of perception. A conceptual space is a space described by perceptual dimensions. A concept is a convex volume within this space: a shape without holes or indents. For instance, the color blue is a volume in HSB space. These spaces have certain topological structure, and this topology can be mapped onto different conceptual spaces: analogical thinking can be thought of in this way. This clearly has import for HCI, for example, hue is a circular dimension, therefore can be mapped on a circle in space, hence the use of a 'color wheel' when designing color selection interfaces. The fact that we do refer to these other spaces as 'spaces', and make use spatial metaphors for a huge range of phenomena (e.g. the future is 'ahead' of us, prices have gone 'up' etc.) indicates that spatial dimensions are probably the most fundamental of all. This is to be expected from an embodied brain designed to produce movement. Gärdenfors' claim is that many cognitive processes are more akin to geometrical manipulations than to serial, symbolic and logical operations[3].

The notion of conceptual spaces provides a basis for Wiggins' Creative Systems Framework, described in Section 3.4.1. Later, this link between perceptual dimensions and conceptual space enables us to propose a link between the artist's mental model of a creative artefact, and the dimensions of control of a musical instrument.

2.2.5 The Motor Hierarchy

We must necessarily interact with our environment and our technology via the body. The body is controlled by the motor system. Like the perceptual systems, the motor

[3]The symbolic manipulations promoted by what he calls the 'sentential paradigm' of earlier cognitive science and GOFAI (Good Old-Fashioned Artificial Intelligence)

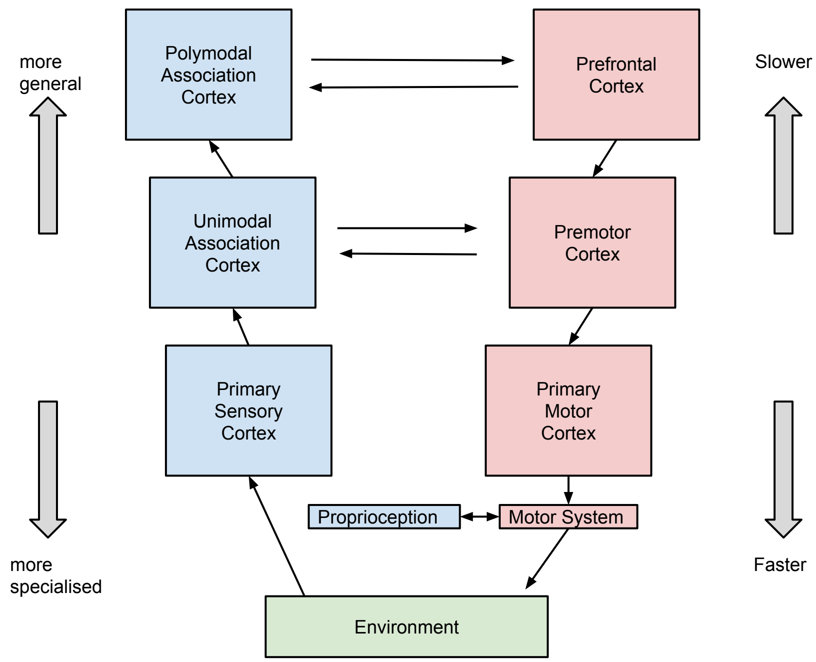


Figure 2.2: Perception-action cycles of increasing generality, and operating over increasing time-scales (adapted from [Fuster, 2001]).

system is hierarchical in nature. Figure 2.2, adapted from [Fuster, 2001], illustrates this hierarchy. Actions, in general, are motor outputs that are intended to affect future events in the environment, therefore they extend over a given time-scale. Fuster portrays progressively more frontal regions of the cortex (ascending up the left hand side in Fig. 2.2, or from centre to left in Fig. 2.1) as associated with ever larger time-scale action planning (temporal integration windows). The various levels of motor control act in concert with increasingly integrated, generalised and flexible forms of sensory information (right hand side).

The simplest, and fastest, responses are spinal reflexes: these connect simple stimuli, such as pain, to basic muscular actions, such as withdrawing a limb. These processes operate on the scale of tens of milliseconds. Next up the simplified motor hierarchy in Fig. 2.2 is the primary motor cortex. Stimulating parts of this region electrically will cause simple movements in various parts of the body [Banich and

Compton, 2011].

At the mid-level, the Premotor Cortex is believed to contain representations of more complex, goal-directed actions. In addition to sending movement commands to the primary motor cortex, predictions of the feedback that the senses *should* be recieving (an 'efference copy' of the movement, or 'corollary discharge') are sent to the relevant sensory areas, enabling faster error correction on the basis of sensory feedback [Miall and Wolpert, 1996]. By this means we gain the ability to distinguish external from internal causes [Wolpert and Flanagan, 2001]. The unimodal association areas (the corresponding level in the perceptual hierarchy) also associate certain sensory inputs with relevant motor goals: for instance, when a certain shape of object is perceived, it triggers similar neural activity as when actually grasping that object [Rizzolatti and Luppino, 2001]. This type of predictive circuitry may be the neural basis for 'affordances' [Gibson, 1977] and Embodied Cognition, a key concept for HCI and interaction design; this is further discussed in Section 2.2.6. It should be noted that such predictive signals will be present at all levels of the hierarchy [Kiebel et al., 2008].

The top level of this hierarchy, associated with the Prefrontal Cortex, consists of higher level thought, such as the ability to plan and act according to longer term goals, and the ability to coordinate and evaluate our own thoughts. Higher level goal processing can involve recalling episodic information in the distant past, and projecting hypothetical situations into the distant future. These processes work with more complex multi-modal information, however there is considerable evidence that even our most abstract and sophisticated behaviours are still grounded in bodily actions, having evolved from them [Wilson, 2002]. These higher level thought processes are the hardest to model, but are considered indispensable for creative cognition [Dietrich, 2004a; Limb and Braun, 2008].

One useful concept from control theory is that of closed loop and open loop control. Open loop, or feed-forward control consists of a pre-packaged command being sent to the actuator. This action just plays out like a recording, and once triggered cannot be adjusted. ‘Motor program’ is one term used for this open loop control signal [Schmidt et al., 1979]. Closed-loop, or feedback control, in contrast, uses information returning from sensors to continually correct for errors in the movement. The advantage of closed-loop control is that by using information from sensory feedback, more accuracy and more robustness to noise and unexpected changes can be achieved. This adaptability comes at a cost of speed however: it takes time to process the sensory feedback. Therefore open-loop control is more suited to rapid tasks, and closed-loop to accurate ones. The only way open-loop performance of a movement can be trained is by repetition and error based reinforcement learning. This idea will be familiar to any performer learning a rapid passage of notes on an instrument: one simply has to run through the sequence repeatedly until no more errors are made. There is simply not enough time to notice and correct errors whilst playing [Zatorre et al., 2007].

A theoretical result from control theory [Touchette and Lloyd, 2000; Klyubin et al., 2008] states that the information that an agent can use to find a target in a parameter space is the sum of that available to the open loop and the amount of information that is sensed from the environment. Thus, to achieve an accuracy greater than the existing internal representation of a target, feedback must be used. This has relevance to a musician seeking to create a certain sound: either they must know to certain accuracy how to move through parameter space to where that sound is located, or they need to recourse to a slower, trial-and-error mode of interaction based on evaluating auditory feedback.

One issue with the idea of ‘motor programs’ is that there are an infinite number of possible movements, even an infinite number of ways to achieve a single outcome (the ‘degrees of freedom’ problem [Bernstein, 1967]). So whilst there is evidence that a motor program must play out to completion regardless of feedback, clearly it is not simply a static recording of muscle activation amounts: motor programs must be extremely adaptable. One example of this adaptability is how we can write a much larger version of our signature on a blackboard, the rhythms of the movements are the same as writing on paper, but are scaled to use the whole arm rather than the wrist and fingers, i.e. a completely different set of muscle activations [MacKay, 1982]. 'Motor primitives' are one proposed solution to this problem [Wolpert et al., 2011; Thoroughman and Shadmehr, 2000]. A motor primitive is a basic scalable unit in a generalised coordinate system that forms the basic building block of more complex actions. Stored primitives can be combined in various amounts and at various times, and parameters such as size, speed, rotation and muscle groups, and can therefore be adjusted according to context, rather like a vector basis. They then act in concert to produce a large repertoire of actions.

There is increasing evidence that motor control is not as simple as an either-or firing of a closed or open-loop program. A highly influential theory [Wolpert and Kawato, 1998] claims that competing populations of forward and inverse models are generated. The forward model is a prediction of the sensory feedback that is expected to result from the movement. An inverse model is a movement instruction 'reverse engineered' from the desired sensory state. In reaching to grasp a cup, the initial intention (grasping certain shaped object in a certain position) is converted into both a model of the movement sequence that is required, and also the expected visual, kinaesthetic, haptic and proprioceptive sensory data that is expected to result throughout the movement (an 'efference copy'). This enables lower level feedback loops between the motor and sensory modules to quickly compare the forward model with the actual sensory data, minimise a cost function, select the current best performing models, and hence fluidly adapt the movement. Reinforcement learning mechanisms then learn from the errors encountered to reinforce or downgrade the relative contributions from amongst the population of models appropriately.

Further refinements of this picture seem to do away with the inverse models, and simply have the movement proceed via a continual reduction of the gap between predicted and actual sensory feedback [Friston, 2011]. Kiebel et al. [2008] dispense with the need for two separate hierarchies all together, as action and perception are tightly interlinked. This is justified on the basis of the 'free energy principle' (see Section 2.2.7); which states that actions are also a means to predict, i.e. minimise the 'surprise' from, sensory data. They state: "Generators of motor output simply predict sensory consequences of anticipated movements". This idea will be returned to when discussing the Free-Energy Principle in Section 2.2.7, and embodied cognition in Section 2.2.6.

If we wish to study musical, time-based interaction, it is advantageous to have estimates for the time-scales predictive feedback loops of various levels in the cognitive hierarchy: the 'temporal integration window' over which the predictor operates. One way the brain's time-scale hierarchy can be investigated is by studying Electroencephalography (EEG) results, in particular Event Related Potentials (ERP) [Banich and Compton, 2011, p. 73]. ERPs are spikes in the electrical activity of the brain emitted by large numbers of neurons firing on response to some instantaneous sensory stimulus. Spikes occurring sooner than 100ms are related to low-level early sensory processing. The P_{100} and N_{100}[4] are also related to specific sensory processes, but these peaks are modulated by attention: if a stimulus is subject to attention, then the P_{100} and N_{100} will have greater amplitude. The N_{200} is known as 'mismatch negativity' and can be elicited by an *unexpected* stimulus. The P_{300} is associated with an attended stimulus that necessitates an update in working memory. Without conscious attention, the P_{300} is not seen at all. N_{400} is thought to occur when

[4] P and N refer to positive and negative peaks in the electrical potential, the subscript refers to the time in milliseconds after the stimulus.

some *semantic* anomaly is detected. For expectation related ERPs relating to more musically relevant stimuli, see Pearce and Wiggins [2012]. It seems that these different time delays really are associated with different levels in the predictive hierarchy [Wacongne et al., 2011].

Similar time-scales apply to outgoing motor control. Therefore for simple stimulus-response tasks, the reaction time is around 200ms, whereas for more complex semantic processing, the round trip could be more than 800ms. If we contrast this with the time-scales of musical events such as a fast sequence of notes, which can occur at intervals around 100ms, we reach the somewhat perplexing conclusion that volitional control cannot be exerted at these time-scales. The conscious action-perception feedback loop could span more than a whole beat. Therefore, improvising musicians must be explicitly working with higher level phrases, and not individual notes [Johnson-Laird, 2002].

2.2.6 Embodied Cognition

Given the likelihood that higher order cognition evolved from motor control, we might expect that more sophisticated planning and executive functions will be extensions of this fundamental architecture. That is, given a successful and supremely flexible solution to the problem of movement control, evolutionary pressures meant the brain reused ("exapted") the same mechanisms to enable *cognitive* control. Processes that effectively orchestrated coordination of the limbs were re-routed amongst disparate brain modules in order to orchestrate mental processes. This would lead to the idea of 'forward models' of sensory predictions applying to conscious thought also.

Embodied cognition, a topic with an interesting history [Chemero, 2011], is increasingly seen as both a means to explain certain aspects of human cognition, and an approach to creating more natural and engaging computer interfaces [Klemmer et al., 2006; Dourish, 2004]. Philosophers such as Heidegger and Merleau-Ponty came to acknowledge that mental existence was intimately tied up with the body and the environment, these themes being taken up by others such as Dreyfrus and Varela and applied to our relationship with technology [Dourish, 2004]. Gibson [1977] introduced the notion of "affordances": perceived action possibilities, which found their way into the HCI literature via Norman [2002].

The literature is diverse and features various degrees of departure from the cognitive science mainstream [Goldman, 2012], in some cases being directly opposed to the idea of the brain as a computational system processing Shannon-Weaver type information [Gibson, 2014]. It is worth describing a clarifying review conducted by Wilson [2002], which critically investigates six claims of embodied cognition:

1. Cognition is situated.
2. Cognition is time pressured.
3. We off-load cognitive work onto the environment.
4. The environment is part of the cognitive system.
5. Cognition is for action.
6. Off-line cognition is body-based.

The claims Wilson treats with caution are 1, 3 and 4. They are sometimes interpreted to mean that we cannot analyse cognition at all without taking the environment and the context into account. We certainly do off-load computation onto the environment, notably to assist with working and long term memory. However, Wilson points out that cognition in the brain is fairly consistent and independently ongoing, whilst cognition using the environment is intermittent. It therefore makes sense to study the brain as the fundamental unit of cognition. I would make a further point that it must not be forgotten just how much computation the brain is performing: surely the activity carried out by 10^{15} synaptic junctions vastly outweighs anything we can outsource to the environment in real-time, even with extremely generous estimates for the throughput of physical interactions and sensory inputs. Nevertheless, in the case of musical interaction, it will be essential to incorporate the immediate environment: the interface and synthesis algorithm that the user is controlling and listening to, which forms part of the perception-action loop (see Section 5.3).

The view on Embodied Cognition taken for the purposes of this thesis is that designing for the body is, in fact, designing for the implicit sensorimotor system: the vast majority of the computational power of which resides in the cortex. The great contribution of embodied approach is that it highlights just how much processing does go on in the implicit system: using embodiment to downgrade the computational significance of the brain is probably unwise.

Another aspect of embodied cognition that is particularly important for creativity, is that it is becoming increasingly apparent that perceptual and motor areas make a substantial contribution to the ability to *imagine* [Lotze et al., 1999]. When imagining carrying out an action, imagining a visual image or imagining music, the areas of the brain that would have been involved in actually performing or sensing those phenomena become active. So 'virtual' actions and perceptions utilise the same mid-level modules as 'real' perception and action, but the connections to the lower level modules that would actually sense or act have been inhibited. Therefore imagination may make use of the 'top-down' paths that are used for sensory prediction [Stokes et al., 2009].

Embodied cognition is frequently cited as being an important principle for the future direction of HCI research [Kirsh, 2013]. Many see a historical neglect of consideration the body [Djajadiningrat et al., 2007]. In [Klemmer et al., 2006], five ways that embodiment can integrate the physical and computational worlds are described:

1. *Thinking through doing:* manipulating concrete physical objects enable off-loading of thought processes into the environment, make learning a more active and exploratory process, and make evaluation more tangible.

2. *Performance:* "the intimate incorporation of an artefact into bodily practice to the point where people perceive that artefact as an extension of themselves; they act through it rather than on it". This is the claim that action-centred skills and motor memory vastly speed up and lighten the cognitive load of dealing with external information.

3. *Visibility:* the ability to access information at a glance is a huge cognitive aid. Spatial arrangements contain large amounts of information that can be quickly gleaned.

4. *Risk:* Our physical bodies provide an element of vulnerability and immersion that changes our relationship to and responsibility for elements of the environment.

5. *Thick Practice:* in some situations leveraging existing 'real world' skills is important. Kelmmer et. al. cite the musical example of Final Scratch: time-coded vinyl records that enable DJs to manipulate digital audio with their existing turntable skills.

Despite these clear advantages to embodied interactions, it is hard to find quantitative proof that designing for the body can bring measurable benefit in terms of the rate at which tasks can be accomplished. Very often novel interfaces claim to be embodied and tangible, but do not provide any experimental comparison of their effectiveness against a 'non-embodied' counterpart. Perhaps this is one reason why "technologies... continue to place demands on our cognitive abilities, and deny us the opportunity of building bodily skill" [Djajadiningrat et al., 2007]. Whilst embodied interaction is a hot research topic, there is perhaps a perception that embodied interaction methods as mere novelties, or somehow un-serious. Heavyweight, professional office or creative work will still be carried out using a standard computer interface. But if embodied cognition is as powerful as it seems, there are very serious gains to be made in productivity by 'pipelining' the autonomous modules in the brain that deal with complex sensorimotor skills. If these gains could be proved, perhaps embodiment would be taken more seriously.

2.2.7 Unifying Perception and Action via Hierarchical Prediction

The free-energy principle[5] is one of the most exciting developments in recent cognitive research. It shows great potential as a unifying principle, making many testable predictions for which evidence is accumulating rapidly. Friston summarises the free energy principle thus:

> "In brief, the motivation for this minimization [of free-energy] is to explain how biological systems maintain their biophysical states within bounds and thereby resist the second law of thermodynamics, in other words, to explain how they maintain a homeostasis. They can do this by minimizing the long-term average of surprise, which implicitly minimizes the entropy of their sensory states. Surprise is just the negative

[5] Similar models have gone by various names over the years, the 'Helmholz machine' [Dayan et al., 1995], the 'Bayesian brain' [Knill and Pouget, 2004], or the 'predictive brain' [Wacongne et al., 2011].

> log probability of the sensory signals encountered by an agent. In information theory, surprise is called self information, while in statistics it is the negative log model evidence or marginal likelihood. Although agents cannot minimize surprise directly, they can minimize a free energy that is always greater than surprise; hence the free-energy principle. Under some simplifying assumptions, this free energy can be thought of as prediction error. This means that perception can reduce prediction errors by changing predictions, while action reduces prediction errors by changing sensations."

This last point is intriguing: the agent can either update their model to align with patterns in the sensory environment, or they can act so as to produce patterns in sensory data that align with their internal model. The latter may be a possible starting point for artistic behaviour, and will be explored further in Section 5.3.1.

For an excellent review of theoretical, experimental and philosophical aspects of this paradigm see [Clark, 2013].

2.3 Implicit and Explicit Processes

From the earliest beginnings of the field, the distinction between conscious and subconscious thought has been one of the most essential, but yet most perplexing distinctions in psychology. Why are we aware of some thought processes and not others? The unconscious mind is clearly important for creative endeavours. Many great works of art and music and discoveries in science are claimed to have emerged from the unconscious [Zhong et al., 2008].

So what is the nature of the distinction? What can the conscious mind do that the unconscious cannot, and vice versa? The question of what mechanisms and correlates distinguish the two is the so called 'easy' problem of consciousness. The ‘hard’ problem of consciousness is how exactly ‘qualia’ (the ineffable qualities of sensation, awareness, the feeling of what it is like to *be* a conscious entity) can arise from these mechanisms [Chalmers, 1995]. There has been considerable progress with regards to the easy problem, but arguably very little with regard to the hard problem, so we shall not deal with it here.

Skills can be transferred between these realms via repetitive practice [Anderson, 1981]. When initially faced with learning a skill, one will follow explicit instructions, one’s actions will be deliberate, slow and consciously controlled. With practice, the actions become more fluid and automatic, but can still be accessed and adjusted if needed. With very well practised skills, particularly those learned from an early age, the actions become so automatic that they cannot be accessed at all: one knows how to do something, but one doesn’t know how one knows. This is sometimes referred to as ‘overlearning’, and is not always desirable [Langer and Imber, 1979]. This process of skill consolidation may be related to chunking.

2.3.1 Correlates of Consciousness

The reason that we should possess consciousness at all is similarly mysterious, but there are a number of functions consciousness seems to be associated with that may provide clues as to its purpose, if not its essence:

1. *The spotlight of attention:* attention is the process by which certain information is selected for further processing. The focus of attentional processes tends to be what we are ‘conscious of’, often the most important or most surprising information in the environment[6].

2. *Integration of information* [Tononi, 2012]: The flexibility of explicit thought

[6]However, there is evidence for attentional selection occurring unconsciously [Bor and Seth, 2012].

implies that information in working memory can be transferred across domains, i.e. is flexibly routed across many different areas of the brain.

3. The contents of consciousness (i.e. working memory) tends to be that which requires *sustained preservation of mental states*: multiple steps of reasoning over time-scales of more than a second.

4. Consciousness is a *gateway to episodic memory*: all the memories and facts that we can explicitly recall once occurred as conscious moments.

5. Stimuli that make it into consciousness are often more unexpected that those processed by lower-level processes. The explicit system is adept at dealing with *unexpected situations*.

6. The stream of consciousness can be analysed in a *meta-cognitive* fashion, and can be *reported* to other people.

There are also 'neural correlates of consciousness' (NCC) [Dehaene et al., 2014]. For instance:

1. *Integrated brain-wide processing:* conscious processing seems to correlate with the activation of the largest brain networks involving frontal and parietal hubs [Chennu et al., 2014; Bor and Seth, 2012].

2. A *non-linear cascade of electrical activity* occurs after 300ms of a consciously reportable stimulus, but not a subliminal one. This is known as the $P3/P_{300}$ wave [Dehaene et al., 2014].

3. Late, sudden bursts of *gamma oscillations* located in the area associated with the item that is raised to conscious access [Hameroff, 2010].

So there do indeed seem to be 'neural correlates of consciousness' [Sergent and Dehaene, 2004] that can be seen whenever the subject reports awareness of a stimulus, and not seen when they cannot [Dehaene, 2014].

2.3.2 Attention and The Frontal Lobes

Another cognitive process related to consciousness is attention. Attention is the process by which certain information is selected for further processing and other information is discarded. This selectivity is needed to avoid sensory overload. The relevance of sensory information depends very much on what kind of task we are engaged in. Whilst a surprising stimulus, such as a sudden loud noise, will immediately command our attention ('bottom up attentional selection'), often it is 'top down', goal-directed actions that determine the objects of our attention. The focus of attention is clearly related to the contents of consciousness, however it is more the mechanism by which salient information is selected and prioritised [Baars, 1997]. The brain does not have the capacity to fully process all the information it receives. Nor would it be efficient for it to do so. As such, attention is often likened to a filter, or a bottleneck in processing [Marois and Ivanoff, 2005]. To attend to a particularly demanding stream of information requires ignoring and actively suppressing others [Payne and Sekuler, 2014].

As mentioned in Section 2.2.5 on the motor hierarchy, one of the functions of the Pre-Frontal Cortex (PFC, see Fig. 2.1) is to control attention and manage higher level goals. These areas at the front of the brain are thought to direct attention so as to carry out structured actions. The PFC is one of the most highly connected brain regions, and the region that is most highly developed in humans relative to other species. It appears able to integrate many sources of information, internal states such as emotions, sensory input, and long term memory. On the basis of this integrated information it orchestrates other brain regions in order to carry out planned actions, hence the term 'executive' functions often used in reference to the frontal lobes. It is possible to function with damage in this area, so it should not be considered as controlling *everything*—that would lead to infinite regress—but regulation of behaviour, decision making, personality and social judgement seem to suffer in patients with PFC damage (specifically ventromedial prefrontal cortex (VMPFC)). One interesting condition is 'environmental dependency', where patients' actions become overly dependent on the objects in their environment: the affordances of the external world seem to completely determine their actions.

Dietrich highlights the importance of the PFC for creative behaviour [Dietrich, 2004a], and lists the following essential functions, along with references to experimental evidence for them:

1. self construct and self-reflective consciousness,
2. complex social function,
3. abstract thinking,
4. cognitive flexibility,
5. planning,
6. willed action,
7. temporal integration,
8. sustained and directed attention,
9. working memory.

The last function is deemed particularly important. Working memory provides

> "...the infrastructure to compute these complex complex cognitive functions by providing a buffer to hold information in mind in order to order it in space-time. It is this superimposing of already highly complex mental constructs that dramatically increases cognitive flexibility." [Dietrich and Haider, 2014]

Without this buffer, a chain of reasoning is incapable of extending over more than the time it takes a excitatory signal to traverse the brain, less than a second or so [Dehaene, 2014]. In order to sustain a controlled train of thought, and put multiple concepts together, the explicit system must be used.

Whilst attention and consciousness may not be synonymous, explicit conscious processing and working memory are increasingly being thought of as tightly interlinked. In the next section we focus on working memory, and its relevance to HCI.

2.3.3 Working Memory

Working memory (or the now lesser used term short-term memory) is a form of memory that is preserved over short time-scales, is limited in capacity: the so-called 'magic number' of chunks: 7 ± 2 [Miller, 1956]. The number of simultaneous items has been revised down to 4 by some researchers. Working memory is effortful to maintain, and overloading it causes discomfort and frustration. Rather than these memories being stored in a particular region of the brain, like procedural or episodic memory, more recent theories propose that it is an attentional process, whereby memories stored in various modules are kept activated [Banich and Compton, 2011]. So rather than the PFC being a 'storage area' for chunks, it is more like a juggler of chunks, and is engaged in top-down activation of particular information in other brain modules [Banich and Compton, 2011, p. 295]. This activation fades over time if not maintained by executive processes.

Other short-term memory stores are associated with individual senses. There seems to be a short-term auditory store, for example, that can maintain a few seconds of sound information ready to be accessed as the need arises. Our ability to suddenly understand a previously misheard, or non-attended sentence indicates that the phonemes from the start of this sentence must have been stored in this auditory buffer: thus they can be reinterpreted and retrospectively raised to consciousness and understood. These sensory buffers fade after a short period, and can be overwritten by new salient information.

Another aspect of short-term memory that was considered as separate from generalised working memory is the 'visuospatial sketchpad' [Baddeley, 1992], where non-verbal visual and spatial imagery can be manipulated. There may be more such sketchpads however, with the possibility that top down attentional processes may use a number of the brain's mid-level modules to simulate, or emulate, hypothetical realities [Stokes et al., 2009; Dietrich and Haider, 2014].

Working memory is an essential cognitive aspect of Human-Computer Interaction. Whilst computers obviously shoulder a huge amount of the cognitive burden, they may also place quite high demands on explicit thought. Interface design is very much the art of minimising working memory demand[7]. In HCI research the maintenance of task relevant information can be studied in multi-tasking and task interruption experiments [Salvucci and Taatgen, 2010]. An important determinant of what makes an interruption disruptive is 'problem state'. Problem state consists of the temporary contextual information needed to be maintained in working memory in order to complete the task. It appears that the speed at which a user can resume a task after an interruption is affected by three things: firstly how complex

[7] It is now possible to monitor working memory demand by measuring blood flow in frontal regions; hence design human-machine interfaces in order to maintain suitable levels of mental load [Ayaz et al., 2012].

the problem state is, secondly how long the interruption is; and finally how task-relevant the interruption was [Borst et al., 2015]. For example, if one interrupts a timbre design task by experimenting with that same sound on a keyboard, this will be less disruptive than, say, answering an email. Experiments reveal that the activation of the task context fades more or less exponentially, and requires time and effort to reactivate in proportion to the amount of fading. When designing musical interfaces, if we take the pessimistic view that manipulating the interface is a form of 'interruption', it would be vital to design for interactions that have a low amount of problem state; are as rapid as possible; and are as 'musical' as possible.

Human beings tend to overestimate how much of their actions rely on conscious abilities. After all, this is where conscious, executive control tends to reside, and these processes are the only ones we have meta-cognitive access to. Our self-model is a model of the conscious self. But it must be stressed that the explicit system is really the tip of the cognitive iceberg, the implicit modules maintain the vast majority of everyday behaviour, and without them we would be helpless. Even highly analytical tasks such as mathematics, chess or computer programming must rely on a vast amount of tacit knowledge [Wagner and Sternberg, 1985]. Some even claim that decisions about complex matters can be better approached with unconscious thought [Dijksterhuis and Nordgren, 2006].

2.3.4 Global Workspace Theory

Baars' global workspace theory (GWT) is one of the most widely accepted theories of consciousness [Baars, 2005]. It attempts to explain the division between conscious and unconscious thought by proposing that there is a central workspace in the brain that can flexibly work with information (see Fig. 2.3). GWT is often illustrated by a theatre metaphor. The objects/actors on stage are the contents of working memory.

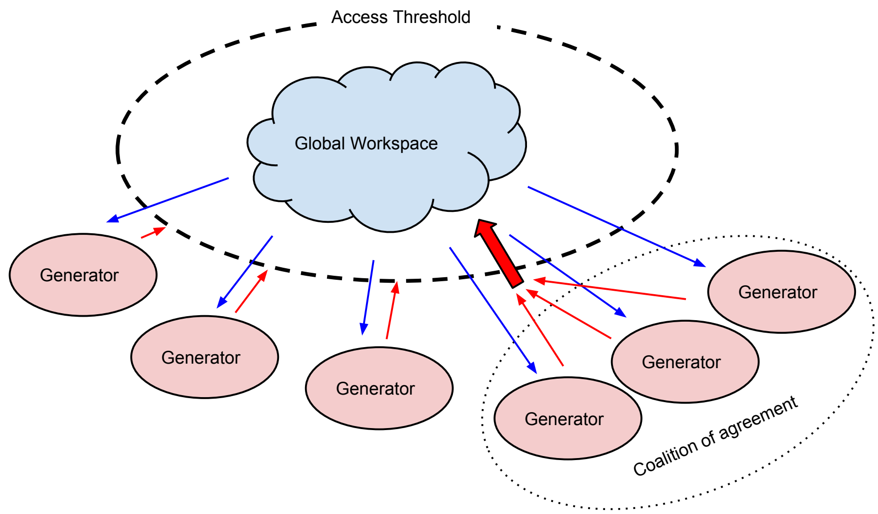


Figure 2.3: Global Workspace Theory (GWT): an illustration of the threshold paradox (adapted from [Wiggins, 2012]). The global workspace broadcasts its content to generators (blue arrows). Generators (mid-level modules in the predictive hierarchy), constantly suggest ideas that are associative variations on the content of the workspace (small red arrows). The only way a generator can gain access to the global workspace (large red arrow) and get their ideas heard is by recruiting other generators into a coalition of agreement. This seems to be a Catch-22 situation, because the only way to recruit other generators is by knowing about other generators, which requires a broadcast mechanism such as the GW.

The stage of the theatre is lit up by the spotlight of attention. Backstage, there are many actors and technicians playing a vital role in what transpires on stage, but are unseen, i.e. carrying out subconscious processing. The activity on stage is being broadcast, in that everyone in the theatre is aware of the action on stage (hence the 'global' workspace). Conversely, behind the scenes activity goes unnoticed, therefore exhibits the encapsulated and modular properties discussed in Section 2.2.2.

In recent years this model has been supported in neuroscientific work, namely in the Global Neuronal Workspace model [Dehaene et al., 1998], which integrates Baars' theory with brain network-topological and ERP phenomena.

One thing missing from Baars' model is a mechanism by which items are selected and raised into consciousness. If every actor is capable of taking the stage as needed, how is it decided who actually does? This is known as the 'threshold paradox' (see Fig. 2.3). Wiggins and Bhattacharya [2014] propose a solution to this using an information-theoretic notion of surprise. Surprise can be thought of as a violation of expectation. This leads us neatly to the idea that spontaneous creativity can be explained by the ability of subconscious processes to produce a concept that is surprising, and hence raised into consciousness by exactly the same mechanisms that maintain the implicit-explicit threshold in everyday waking cognition:

> "...non-conscious creativity is happening all the time as a result of ongoing anticipation in all sensory (and other) modalities. When conditions are right, this essential survival mechanism is not so much exapted for creativity, but gives rise to creativity as a side effect."[Wiggins, 2012]

This ties in neatly with Friston's cascade of prediction error 'surprisal'. Violation of expectation has indeed been used as an experimental variable to trigger conscious processing [Dehaene et al., 2014].

2.3.5 Dual Process Theory

A word heard again and again in discussion of user interfaces and particularly music technology is "intuitive". Obviously if a device conforms to Shneiderman's consistency rule (see section 2.4), or makes use of an obvious similarity to an existing device, then users will be familiar with the way it works and may declare it intuitive. However the notion seems to express something deeper than mere familiarity. The truly valuable intuitive interface is one that may be new, but nevertheless immediately satisfies our most basic expectations of what kind of results will follow our gestures, and not require time and effort to "figure out". The formal definition states that intuition is the ability to acquire knowledge without the use of reason. This is a rather negative definition. So the question must be asked: what mechanisms are present in the brain *apart* from reason?

Much of the literature regarding intuition occurs in the context of "dual process theories" of reasoning [Evans, 2003; Kahneman, 2011]. There are a variety of different views on this topic. The principal source for our purposes is Stanovich [2011]. The hypothesis is that two systems, of different evolutionary origin and different capabilities, are present in the brain. The first ('implicit' or System 1) is fast, parallel and associative, but can suffer from inflexibility and bias. The second ('Explicit' or System 2) is more rational and analytical but is slower, requires more effort and makes use of limited working memory. It is often used by social psychologists to explain why many decisions that humans take (under, for example time constraints) seem to be irrational, however the theory is also relevant to a great deal of other cognitive behaviour.

Memory also has implicit and explicit flavours. Implicit memory is knowledge of 'how' e.g. how to ride a bike, this is known as procedural memory. Explicit is knowledge of 'what', e.g. Paris is the capital of France, this is also known as declarative memory. Evaluation, or judgement, also comes in explicit and implicit versions. Explicit, analytical judgements tend to take time, and are not good at dealing with large amount of information or uncertainty. Therefore intuitive hunches, gut feelings and judgements based on affect are often used instead [Sadler-Smith, 2012]. One particularly relevant circumstance where implicit judgement may be more effective is in problems that involve multidimensional data [Dietrich, 2004b], leading us directly to the one of the principal hypotheses of this thesis: that multidimensional controllers should be more suited to implicit processing.

The 'cognitive miser' hypothesis is that the brain will tend to process information in the cheapest way possible [Stanovich, 2011, 98-100]. Therefore the biases of the implicit system can be explained in terms of people taking mental shortcuts

(Heuristics) that reduce the attentional cost of arriving at satisfactory, rather than optimal, solutions.

Characterisations of these dual processes are summarised in Section 2.3.7.

2.3.6 Reflective Meta-cognition

Attention is not just focused on external stimuli. Attention can also shift to internal, mental states. Stanovich proposes the need for a tri-process model [Stanovich, 2009], in which the explicit system is divided into two systems: the 'analytic' and the 'reflective'. A separate 'reflective' or 'metacognitive' system must flag up the need for the explicit algorithmic system to re-analyse the low effort solution and check that it is not in error, in other words acting as an advisor to encourage the 'cognitive miser' to invest a little more. This ties in with the multiple functions of the frontal lobes: orchestrating cognitive operations, but also switching tasks and attentional set if necessary. The act of realisation is known as 'decoupling', as it decouples attention from the algorithmic task that it is involved in. The fact that the ability to carry out complex abstract thinking (e.g. in standard IQ tests) does not correlate well with the ability to overcome implicit biases provides evidence that these are in fact two separate brain systems that contribute to an individual's rationality [Stanovich, 2011, p. 154]. Stanovich also proposes that, rather than consideration of an single 'implicit system', Type 1 thinking should be thought of as being carried out by many different specialised systems, christened 'The Autonomous Set of Subsystems', or TASS. Figure 2.4 shows this model of cognition. What is not clear is if the reflective and algorithmic systems can operate in parallel. We shall assume that they can interfere with each other, as they may both require access to limited working memory[8]. It also seems justified to promote

[8]The parallels between the reflective and algorithmic mind and the Default and Task-Positive Network are certainly extremely suggestive: one might speculate that these non-correlated cogni-

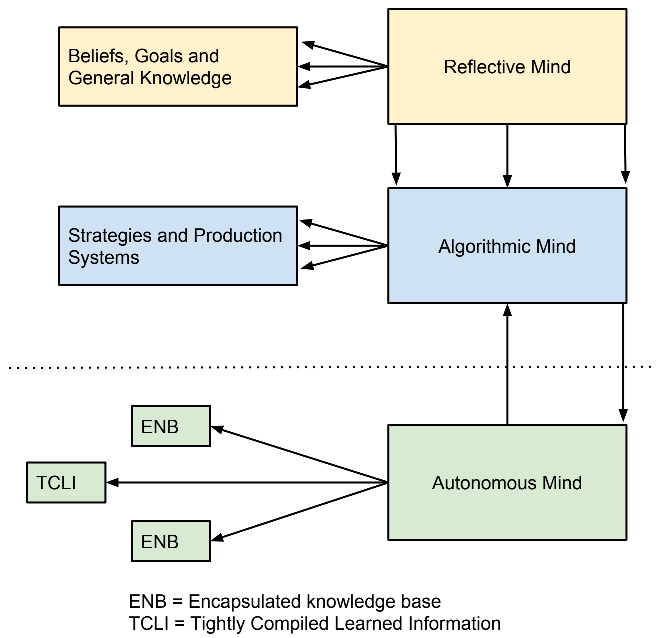


Figure 2.4: The tri-process model. Adapted from Stanovich [2011, p. 96]. Stanovich posits another system higher in the hierarchy than the algorithmic mind, which is responsible for questioning the 'quick and dirty' results of autonomous processing.

the reflective system as being a contributor to creative thinking. In fact I will argue in Section 5.5.3 that it shows a number of hallmarks for being a mechanism for divergent thinking and transformational meta-exploration.

In [Baumer, 2015] the literature on reflection in philosophy, cognition and design is reviewed, and the following three conceptual dimensions are discussed:

1. *Breakdown:* reflection is called for when existing solutions prove inadequate.
2. *Inquiry:* questioning, re-examining and hypothesis forming.
3. *Transformation:* transforming understanding by changing the rules by which other thought systems act.

The first two seem to relate to problem finding stage of creativity, the latter with transformational creativity (see Section 3.1). From a predictive coding perspective,

tive abilities to have a counterpart in the high-level functional anatomy of the brain.

it seems that the purpose of reflection is to go over existing knowledge and experience and attempt to recode it in a more effective form. Baumer summarises Kant's notion of reflective judgement like so:

> "For Kant, reflective judgement is that which reorders our conceptual schema. Reflection occurs precisely when our existing conceptual schema do not apply (or do not apply well) and thus we need to reschematize nature in order to come to a (better) understanding of it." [Baumer, 2015]

As we shall see, this 'reschematizing', or coding of the world is thought by some to be an essential component of creativity. Pearce and Wiggins [2002] also contend that reflective strategies — as in the construction of hierarchies of abstractions — are an essential component of music composers' creative cognitive processes.

2.3.7 Summary of Consciousness

Looking at the literature concerning heuristics and biases, analysis of brain network architectures, and global workspace style cognitive models, clear convergence can be seen when it comes to the problem of how the conscious and unconscious relate to behaviour, hierarchical brain architecture, and information processing. Table 2.1 gives a list of properties for the two systems. The issue of whether 'System 2' explicit thought equates precisely with that which is conscious, that which is the subject of attention, and the contents of working memory is perhaps not settled, but will not be deal with further. We shall assume that these various strands of research are converging on a unified picture, and that the distinct constructs correlate strongly enough to treated as identical [Bor and Seth, 2012], or at least to the extent that results from musical interaction studies will not be precise enough to reveal differences to any statistically significant degree. In any case, any 'two-system' model is likely to be a gross oversimplification. To simplify terminology the difference between

	Implicit	Explicit
Dual Process Theory:	**System 1**	**System 2**
	associative system	rule-based system
	heuristic processing	analytic processing
	tacit thought processes	explicit thought processes
	interactional intelligence	analytic intelligence
	experiential system	rational system
	quick and inflexible modules	intellection
	intuitive cognition	analytical cognition
	recognition-primed decisions	rational choice strategy
	automatic processing	controlled processing
Properties:	associative	rule-based
	holistic	analytic
	automatic	controlled
	relatively undemanding of cognitive capacity	demanding of cognitive capacity
	fast acquisition by biology	slow acquisition by formal tuition
	slow acquisition via reinforcement learning	fast one-shot learning
	procedural memory	episodic and declarative memory
Task Construal	highly contextualized	decontextualized
	personalized, conversational and socialized	asocial
	interactional	analytic (psychometric IQ)
Neuroscience	encapsulated modules	brain-wide networks
	unconscious	conscious
	unreportable, unavailable to introspection	reportable, available to introspection
	parallel	serial
	P_{100},N_{100},P_{200} waves	P_{300} wave
Evolution	early	recently, particularly in humans

Table 2.1: Summary of distinctions between explicit and implicit cognition. Dual process terms are taken from Stanovich and West [2000], Cognitive Neuroscience from Dehaene [2014].

unconscious/conscious System 1/System 2 thought shall henceforth be referred to as Implicit/Explicit.

2.4 Human-Computer Interaction

Many of the pioneers in HCI research were concerned with applying cognitive science to interface and interaction design. Foundational texts by Norman [1986], Shneiderman [1982], Card et al. [1983], Anderson et al. [1997], Winograd and Flores [1986], and Suchman [1987], whilst focussing on various different approaches, all stress the importance of understanding how our cognitive processes shape our interactions with artificial systems. Indeed, the discipline has been referred to as 'Cognitive Engineering' which

> "...is a type of applied Cognitive Science, trying to apply what is known from science to the design and construction of machines ... the goal of Cognitive Engineering is to come to understand the issues, to show how to make better choices when they exist, and to show what the trade-offs are when, as is the usual case, an improvement in one domain leads to deficits in another. "[Norman, 1986]

A good, accessible and current overview of topics from cognitive neuroscience as they apply to design of everyday interactions can be found in [Forsythe et al., 2014]. If a single focus of investigation sums up the field, it would be the effective transfer of information between human and machine. According to [Hinckley et al., 2004]:

> "The fundamental task of human-computer interaction is to shuttle information between the brain of the user and the silicon world of the computer. Progress in this area attempts to increase the useful bandwidth across that interface by seeking faster, more natural, and more

> convenient means for users to transmit information to computers, as well as efficient, salient, and pleasant mechanisms to provide feedback to the user."

With that in mind, we next discuss how measuring information transmission has been used to evaluate and design interfaces.

2.4.1 Input Devices, Movement Laws and Human-Computer Interaction

Entropy is a measure of uncertainty, and is also a measure of information [Shannon and Weaver, 1949]. When an agent reduces the uncertainty of a quantity, i.e reduces the spread of its probability distribution, it reduces its entropy. The entropy of a discrete probability distribution, or information in bits, is given by the formula

$$H = -\sum_{n=1}^{N} p_n \log_2(p_n), \tag{2.1}$$

where p_n is the probability of the n^{th} alternative in the distribution. For example, if all the possibilities are equally likely, $p_n = \frac{1}{N}$ for all n and the entropy is

$$H = \log_2(N), \tag{2.2}$$

which states the fact that an H bit number can express 2^H possibilities.

Information can be used to reduce uncertainty, and hence make predictions. Conversely, to decrease the entropy in a system (for example, creating a piece of music on a hard drive), requires information to be input from the environment, i.e. work must be done by the person operating the computer.

Information theory has been applied to many musically relevant areas, for instance perceptual laws [Norwich, 1993], music perception, and creativity [Wiggins and Forth, 2015; North and Hargreaves, 1995]. It has also motivated two laws relating to Human-Computer Interaction, discussed in the next section.

Quantifying the variability of human cognitive and motor processes in terms of information theory became popular in the 50's, shortly after Shannon introduced the idea. Miller stated the case as follows:

> "The advantages of this new way of talking about variance are simple enough. Variance is always stated in terms of the unit of measurement — inches, pounds, volts, etc. — whereas the amount of information is a dimensionless quantity. Since the information in a discrete statistical distribution does not depend upon the unit of measurement, we can extend the concept to situations where we have no metric and we would not ordinarily think of using the variance. And it also enables us to compare results obtained in quite different experimental situations where it would be meaningless to compare variances based on different metrics. So there are some good reasons for adopting the newer concept." [Miller, 1956]

In other words, measuring uncertainty in bits is an excellent way to investigate how well a certain task can be achieved.

Fitts' Law and Rapid Aimed Movement

Fitts' law [Fitts, 1954; MacKenzie, 1992a] applies to rapid aimed movements in a single dimension towards a visible target. The original experiment studied subjects moving a stylus between two target strips as fast as they could [Fitts, 1954]. It is a linear relation between movement time, MT, and an "index of difficulty", ID:

$$MT = a + b \times ID. \tag{2.3}$$

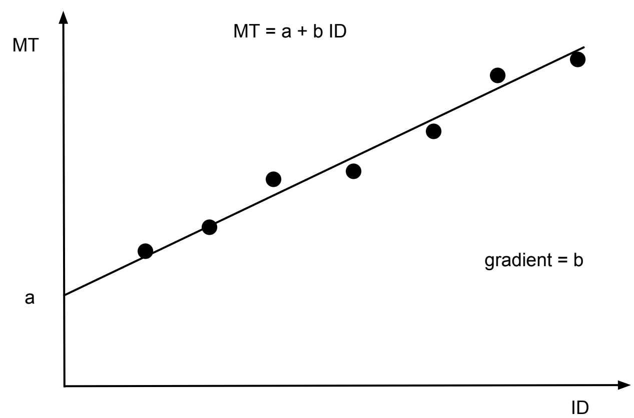


Figure 2.5: A typical Fitts' law plot. Movement time (MT) is linearly related to the index of difficulty ID. The intercept a could be thought of as reaction time, and the gradient b as the reciprocal of throughput, in bits per second, however this interpretation is the subject of some debate.

The ID is a measure of task difficulty in bits. It is calculated from the target width W, and the distance moved to reach the target D. Fitts' law can be used to predict the time taken for various common interaction tasks, such as moving a cursor to a GUI button. It can also be used to compare the effectiveness of input devices, via the "throughput" (TP): the rate at which a user can input information to the system, in bits per second, calculated as $TP = ID/MT$. A typical Fitts' style interface evaluation proceeds as follows:

1. Establish a number of movement distances and target sizes to achieve a range of values for ID.

2. Participants then move a pointing device to hit to these targets in the quickest time possible. Two or more different interface devices are provided as different experimental conditions.

3. Movement times for each ID are averaged, and a regression line fitted to these points in order to obtain constants a and b (see Fig. 2.5)

The device with the highest value for $TP = \frac{1}{b}$ is considered more effective.

Fitts' original formula for *ID* [Fitts, 1954] can be derived by considering the movement as a series of smaller movements with iterative corrections [Card et al., 1983, p. 53]. However, there are alternative formulae, around seven different derivations [Hoffmann, 2013], and even power laws fit the data well in many cases [Goldberg et al., 2013]. For around 20 years the accepted ISO standard has been that of [MacKenzie, 1992a]:

$$ID = \log_2 \left(\frac{D}{W_e} + 1 \right), \tag{2.4}$$

which is the so-called 'Shannon formulation' obtained by considering the nervous system as a noisy communication channel, and making an analogy to Shannon's Theorem 17 [Shannon and Weaver, 1949]. In this formula W_e is the 'effective width': the standard deviation of the distribution of end points multiplied by 4.11[9].

However the debate continues, causing some frustration for those who simply want to carry out interface evaluations. A glaring issue is that the HCI community have followed MacKenzie [Soukoreff and MacKenzie, 2004], whilst the psychology and ergonomics communities have continued to use Fitts' original law. Some see this as a sign of an endemic failure of the HCI community to critique [Drewes, 2010; Hoffmann, 2013].

Another point of debate is the exact relationship between the law and 'information' [Soukoreff et al., 2011; Zhai, 2004; Hoffmann, 2013]. Whilst it is clear that the 'noise' in MacKenzie's noisy channel analogy is the end point variability, it is not clear what the signal is. A discrete movement is not a continuous signal [Hoffmann, 2013]. Another discrepancy in this analogy is that Shannon uses signal *power*, not amplitude [Drewes, 2010; Hoffmann, 2013].

A further issue is how exactly one calculates throughput from the obtained

[9] This adjustment is because the distribution of the end points is generally Gaussian, rather than a uniform distribution over the target width.

straight lines. One way to do this is to use the gradient of the line (b). The ISO standard however, recommends to average $\frac{MT}{ID}$ across the whole line. This has issues, because if a is non-zero, the TP value will vary linearly across the line, and the final average then depends on the experimental range chosen for ID [Zhai, 2004]. This is precisely the sort of thing using a predictive law is supposed to eliminate: why even debate the formula giving the best regression line if the data all along the line are to be averaged? One argument is that a should be kept near zero, and "A large intercept value in the absence of an explanation indicates a problem with the methodology." [Soukoreff and MacKenzie, 2004]. However Zhai [2004] argues for the gradient approach. Zhai claims that $1/b$ is the 'informational' component, and a is the 'non-informational' component of the task. Input device evaluations should consider both. This makes far more sense. If a should for some reason be very large, then surely this is a sign that the device has a high time cost associated with any movement. It is perfectly possible that some devices are more effective at high IDs and others at low IDs: reporting both a and b characterises this behaviour. The accuracy of the end point distribution will eventually saturate at high ID, due to the maximum accuracy with which people can see and point to very small targets. Movement time will saturate for very low ID. Therefore attempting a perfect linear fit for all conditions may be an impossible task.

A number of attempts have been made to extend Fitts' law to more than one dimension. The first, [MacKenzie and Buxton, 1992], considers the effect of rectangular targets. They approach the problem by using the diagonal width of the rectangle, but preserving the one dimensionality of the law. Fitts' law has also been investigated in 3D [Murata and Iwase, 2001; Grossman and Balakrishnan, 2004; Cha and Myung, 2010]. The most recent expression for the ID is dependent on the elevation (θ_1) and the cosine of the azimuth (θ_2) angles to the target:

$$ID_{3D} = c\theta_1 + d\cos\theta_2 + \log_2\left(\frac{D}{W_e} + 1\right). \tag{2.5}$$

All these extensions take the approach that the goal of the ID formulation is to predict movement times. Whilst this certainly results in useful findings, and recommendations for arrangement of interface elements, the underlying theoretical approach is questionable. No attempt is made to consider how these models relate to information theory. How is the above formula to be understood with relation to the amount of information that can be achieved with a 3D interface? Shouldn't a 3D space be providing more information than a 2D one? Using the 1D equation with constant terms added for the difficulty of angle will not tell us this, and might miss the potential to make throughput gains with high-dimensional control. Furthermore this approach will become increasingly unwieldy as dimensionality increases, and probably impossible to extend to $n > 20$ dimensional spaces such as hand and body pose parameters [Rautaray and Agrawal, 2015]. Will new experiments have to be conducted in every single dimensionality, resulting in n regression constants?

There is also a slight circularity in this whole process, if we are *defining* throughput using an *ID* obtained from experimental data conducted with a certain device in a certain dimensionality, then how can throughput be considered a device/dimension independent quantity? There is a danger that establishing regression constants on the basis of movement time will rule out the detection of any cognitive speed-up due to the parallelism of skilled multidimensional movements. Ideally, throughput should be defined theoretically first, and then measured experimentally.

If the goal is to compare input devices, predicting movement times is a secondary consideration to measuring effectiveness. In this regard, the approach taken by Soukoreff et al. [2011] shall be considered most promising for our purposes. Here, the informational component of the task is defined with regard to the probability distributions of the start and end points of the movement. It advances a strong theoretical argument, makes few assumptions, and abstracts away the complexities of the human motor system. In section 5.4.1 we extend this probability distribution approach to generalised n-dimensional search paths. This enables us to analyse arbitrary parameter adjustments in target based tasks (including sound design and timbre performance tasks) in terms of information throughput. Further distinctions are drawn between the predictive and comparative uses of the index of difficulty.

Other Movement Paradigms

Further complexity is added to the debate when considering 'ballistic' movements, and when considering movements where allowed movement time is the independent experimental variable rather than accuracy. In the former case, where the hand is 'thrown' towards the target with no corrective behaviour, end point variability W_e is proportional to distance. In the latter case, referred to by Guiard and Olafsdottir [2011] as the "Schmidt paradigm" [Schmidt et al., 1979],

$$W_e = K_1 + K_2 \left(\frac{D}{MT} \right), \tag{2.6}$$

where W_e is again effective target width, calculated as the standard deviation of the finishing position (in more than one dimension, this can be calculated as the square root of the mean squared Euclidean distance to the target), D is the initial distance to target, MT is the (pre-specified) movement time, and K_1 and K_2 are constants. In other words, accuracy is linearly related to movement *velocity*. The underlying reason for this is thought to be that noise in the muscle activation signals is proportional to the force [Schmidt et al., 1979; Meyer et al., 1988]. Clearly, this is a very different relationship from Fitts' law.

In addition to time constraints resulting in non-Fitts style behaviour, there is evidence that cyclical movements also utilise different mechanisms, and result in more effective and efficient movements [Boyle et al., 2012a]. Guiard [1997] argues that cyclical movements, such as walking, hitting, throwing and reaching should be considered the general case and discrete movements are in fact a 'degenerate' case. In one sense, this throws doubt on the information processing account of movement analysis. However, in my view, just because Fitts' law is not universally applicable, this does not mean that throughput becomes a meaningless quantity. Information is still flowing into the computer, it just has a different dependence on movement parameters. As we shall see in Experiment 3, throughput actually peaks at a certain speed in the Schmidt paradigm. To date, there has been very little research into the implication of the Schmidt linear relation with respect to input devices. In my view, rather than being a theoretical anomaly, this throughput peak may in fact be a good opportunity for designing high-throughput interaction.

Fitts-style tasks seem to show little improvement with practice [Boyle et al., 2012b], presumably because reaching movements are carried out in many everyday situations, and are therefore already well optimised. This indicates that the learning curve (the amount of time it takes a novice to gain enough skill with the device that the experience of using it is rewarding [Levitin et al., 2002]) is quite fast for these reaching and pointing based interactions.

Finally we should note that "performance and user satisfaction are not necessarily correlated" [Macleod et al., 1997]. Therefore, just because an interface has high average throughput, this doesn't mean it will be more pleasant to use. For instance, higher throughput at the cost of higher mental load would not be satisfactory. This is one reason why subjective questionnaires are often used alongside Fitts' law style investigations [Bachmann et al., 2014]. Experiment 3 specifically tests for both throughput and working memory, and also for subjective workload using the NASA TLX questionnaire [Hart and Staveland, 1988].

2.4.2 Multidimensional Controllers

How best to control the multiple parameters provided by digital content creation software? Multidimensional, or high Degree-of-Freedom (high-DOF) control devices seem an obvious candidate. Work in this area goes back at least 25 years[10], with much activity in the early 90's relating to Virtual Reality [Cruz-Neira et al., 1993; Conner et al., 1992] and hence devices for controlling items in 3D space [Jacob and Sibert, 1992; Jacob et al., 1994; Zhai, 1993]. This spawned corresponding investigations using these VR devices for musical interaction [Mulder, 1994; Bargar et al., 1994; Vertegaal and Eaglestone, 1996; Choi, 2000].

Many device evaluations relate to navigation, pointing and object manipulation in 3D worlds. For instance [Zhai and Milgram, 1998a] look at 6-DOF input devices, and propose that one measure of efficiency is 'coordination': how well users can travel diagonally through the space, as opposed to moving one dimension at a time (via 'Manhattan' or 'city block' style navigation). Diagonality can be calculated by several methods:

1. Coordination [Zhai and Milgram, 1998a]: the difference between the shortest path and the recorded path between start and end points.
2. The time series correlation between movements in the different dimensions.
3. Diagonal thresholding [Jacob et al., 1994]: i.e. dividing the trajectory into small sections and counting how many exceeded an angular threshold.

[10] Though multidimensional control was the focus of research during WW2, when a pressing concern was how to rapidly and accurately control the two degrees of freedom of gun turrets [Ellson, 1947].

4. Time on Target (TOT): how much of the path was pointed in the target direction [Ellson, 1947].

Early attempts to provide multidimensional interaction made little impact on the mainstream, probably because of the poor accuracy and reliability of the devices [Vertegaal and Eaglestone, 1996]. Recent years have seen a resurgence of interest however, with very high-DOF input devices such as the Kinekt [Sambrooks and Wilkinson, 2013; Pino et al., 2013; Zeng et al., 2012] which enables skeletal tracking of the entire human body (albeit with latency and accuracy that is a little less than desirable for fast musical interaction), and the Leap Motion [Weichert et al., 2013; Bachmann et al., 2014], which tracks hand pose in a fast and accurate manner.

Whilst the Leap motion has been evaluated via Fitts' law based method, these studies only assessed its ability for 2D pointing tasks [Seixas et al., 2015; Bachmann et al., 2014], where it was found to perform poorly in comparison to a mouse (approximately doubling movement times). Using a 20 plus DOF controller for 2D interaction seems to miss the point somewhat. Given the lack of a Fitts evaluation technique for higher dimensions, perhaps this may be understandable. In Experiment 3 we demonstrate that using the Leap for 6D musical interaction can show a factor of three increase in throughput compared to the usual figures reported with a mouse.

Musical research using the Leap Motion [Hantrakul and Kaczmarek, 2014; Ritter and Aska, 2014; Han and Gold, 2014; Mandanici and Canazza, 2014] has seen some interesting musical applications, and revealed a number of important issues (such as occasional, but musically catastrophic loss of tracking), but are yet to follow any rigorous quantitative evaluation method.

2.4.3 Usability: Some HCI design Principles

In this section we investigate some design principles from the HCI literature. These often consist of lists of recommendations based on user studies. It is worth attempting to distil these in reference to the above review of cognitive science principles. In fact, most of them relate to a single principle: that of off-loading as much of the cognitive burden from the brain's explicit system as possible.

There have been a number of attempts to provide concise, general guidelines for creating user interfaces. One set is the eight golden rules [Shneiderman and Plaisant, 2004], a good starting point for any user interface designer:

1. *Strive for consistency.* Similar tasks should require similar actions. Terminology in text should be the same across labels, menus etc. Copy and Paste is a good example of successful, widely applicable and consistent interaction. Cognitively, this relates to the fact that implicit systems are fast and automatic, but are slow to learn and inflexible. Consistency encourages procedural memories to form, hence relieving cognitive load.

2. *Enable frequent users to use shortcuts.* As a user's expertise increases, so does their desire to access frequent commands via keyboard short-cuts. This results from the same implicit-offloading principle. It is quicker for the motor system to press Ctrl-C than to move a cursor to an icon, not just because of the time taken to move the mouse, but also due to the speed difference of visuo-motor vs. haptic-motor feedback. There is an implied cost-benefit analysis in which the effort of learning lots of key commands is weighed against the efficiency gain of faster interactions in future use. It goes without saying that a musician who spends years using a software package will benefit from shortcuts, and will be willing and able to develop virtuosity in their motor skills. High throughput from keyboard typing skill is one argument for tracker style sequencers [Nash

and Blackwell, 2011] and also for the practice of 'live coding' [Nilson, 2007].

3. *Offer informative feedback.* When a user enters some instruction, it is necessary that some form of feedback occurs to indicate that this was performed. This is usually visual, but obviously audible feedback occurs in music software. This aspect is crucial in music and is related to the idea of "liveness".

4. *Design dialogue to yield closure.* This relates to an overall task composed of subtasks. When something is completed it should feel as such. The state of the brain should mirror the state of the task data. The computer is 'reporting back' the state of its processes, such that the brain's predictions are either confirmed or an error is noted. A simple example is a 'progress bar'. This principle of the human and computer needing mutual updating of their progress through a task-space will be referred to as 'cognitive mirroring'.

5. *Offer simple error handling.* Design the system so the user cannot make a serious error. If something does go wrong, offer clear and unobtrusive notifications and remedies.

6. *Provide an undo option.* Reversibility is important in all user interfaces. This can get complex in a music system that is designed for both off-line serial editing and live parallel performance. This and the previous point seem to be related to the cognitive mirroring principle, in that an error is a deviation from the solution trajectory that both human and machine seek to follow. If either party has deviated from that path, they need to inform the other.

7. *Support internal locus of control.* The feeling of autonomy is extremely important for a computer user. It increases motivation, and has been shown to aid creativity [Amabile, 1998]. Whilst the computer and human are cognitive partners, the human ultimately has the goals and intentions. The explicit sys-

tem is at the top of the cognitive hierarchy. Attempts to interrupt the user's thought processes with error messages or even well-meaning 'recommendations' can be incredibly distracting.

8. *Reduce short-term memory load.* Multiple windows and complex dependencies should be minimised. This is essential for creative software too, as a creator may have a complex overall goal, or a sudden fleeting idea in mind. An excess of interface related material in working memory may interfere with this [Tano et al., 2012]. This is a restatement of the explicit offloading principle.

Direct Manipulation [Shneiderman, 1982; Beaudouin-Lafon, 2000] is another one of Sheiderman's influential interaction models based on the above principles. This model proposes a number of principles which clearly relate to the embodied cognition approach, and to reducing cognitive load. These include: continuous representation of objects of interest; fast incremental and reversible operations with an immediately apparent effect; and "physical" actions on objects rather than complex syntax.

Whilst the WIMP GUI[11] model is clearly based on ideas of direct interaction, perhaps it does not go far enough. Beaudouin-Lafon uses direct manipulation principles to critique this model and propose a post-WIMP model known as "instrumental interaction". This is inspired by tool use in the physical world, where instruments (such as pens and hammers) are used to affect change on "domain objects" (paper, nails). This paradigm is made use of in music software that provide different tools for selection, drawing, zooming, slicing and so on.

Cognitive Dimensions of Notation

How musical information is represented is immensely important for its creation and manipulation. The "Cognitive Dimensions of Notations" [Green, 1989], is a

[11] Windows, Icons, Menus, Pointers, Graphical User Interface

useful set of criteria for evaluating how demanding notations are to understand and manipulate. They originally pertained to software development, but are applicable to all forms of abstract notation, and in particular have been applied to musical representations in software by Nash [Nash, 2012; Duignan, 2008].

The 14 dimensions run as follows (quoted from [Duignan, 2008]):

1. *Hidden dependencies* occur when important links between entities are not visible.

2. *Premature commitment* places constraints on the order of doing things.

3. *Provisionality* is the degree to which users are committed to actions or marks.

4. *Secondary notations* allow extra information to be added using means other than the formal syntax, such as notes and comments.

5. *Viscosity* is the resistance of aspects of the notation to change.

6. *Visibility* is the extent to which components can be easily viewed.

7. *Consistency* occurs when similar semantics are expressed in similar syntactic forms.

8. *Diffuseness* of the notation is determined by the verbosity or terseness of the notation.

9. *Error-proneness* is defined by how difficult it is for the user to avoid mistakes.

10. *Hard mental operations* occur when a high demand is placed on the users cognitive resources.

11. *Progressive evaluation* allows work to be checked at any time.

12. *Role-expressiveness* is how easily the purpose of a component can be inferred.

13. *Closeness of mapping* is how well a representation fits to a specific domain.

14. *Abstraction* is the extent to which higher level abstractions are provided in the notation.

These dimensions relate to how well the notation functions as an extension of the user's cognitive processes, or how well the software functions as part of the user's 'extended mind' [Clark and Chalmers, 1998]. Ideally one would design for all these dimensions to be optimised, but in practice trade-offs exist between them. For example, increasing the amount of abstractions available may increase the number of hidden dependencies. In the case of sequencers it might be desirable to have a main theme clip and then variations on that theme. If one decided to change the main theme and have those changes cascade down to the variants, this would introduce a dependency that may result in unpleasant surprises. Shneiderman's principles also feature some trade-offs. A consistency in style may interfere with the expressiveness of graphical icons. There may also be trade-offs between Norman's emotional and behavioural levels [Norman, 2004]. For example, the folders in Apple's OSX have been forced to all look identical for aesthetic reasons, making visual recognition of particular locations harder.

2.5 Summary: Seven Principles

This chapter has tried to present a unified account of some of the most important aspects of human cognition, and how they relate to our interactions with information technology. Here we summarise the central ideas that will be utilised throughout the rest of this thesis.

2.5.1 Three Cognitive Principles

The three main cognitive theories, along with key texts providing a background to the subsequent discussion, are as follows:

- *The Free-Energy Principle* [Friston, 2010] The brain is a hierarchically organised, modular predictive system. It attempts to encode/explain the world in as efficient a manner as possible, and act on the world so as to minimise surprise, and hence keep its entropy down. The agent does not passively receive and encode sensory information, it actively seeks out information by which to advance the sophistication of its prediction mechanisms.

- *Dual Process Theory [Stanovich, 2011] and Global (Neuronal) Workspace Theory [Baars, 2005; Dehaene, 2014].* The higher levels of the predictive hierarchy are thought to be associated with conscious processing: 'explicit' thought. The brain-wide fronto-parietal networks are able to utilise lower level modules to process complex information in an immensely flexible, concerted manner. The wealth of episodic memory enables the explicit system to carry out 'mental time-travel' in order to establish predictions over much larger time-scales than the implicit system. However, brain-wide processing can suffer from bottlenecks at network hubs, and is therefore slower, serial and more effortful. It also necessarily requires information to be manipulated as encoded chunks, rather than probabilistically. Limited working memory is one limitation of this 'explicit' system, meaning that high-level information used in the 'Global Workspace' is easily displaced. Dual Process Theory specifically highlights the difference between fast implicit processes and slow explicit processes and how they affect everyday reasoning. For our purposes, we shall assume that both dual process and global workspace theories are referring to the same dichotomy.

- *Embodied Cognition [Wilson, 2002]*: Action and perception are tightly interlinked. The motor system produces predictions of sensory feedback when initiating motor actions and vice versa. Similarly the brain actively infers possible motor activities in sensing its environment. A stronger claim is that higher level cognitive process are 'exapted' from more basic action-perception mechanisms. This has significant implications for tangible and embodied interaction design.

2.5.2 Human Interface Design

We can attempt to condense the rules for creative human-computer interface design into a more concise form, with reference to both the principles in Section 2.4.3, and the cognitive principles above.

1. *Reduce cognitive load* on the explicit brain system. Quite possibly the reason we started using writing, algebra, musical notation and digital computers in the first place is to offload cognition onto the environment. This is done in two ways: cognitive mirroring and cognitive pipelining.

2. *Cognitive Mirroring Principle*: The computer can relieve the burden on the explicit brain by carrying out some of its computation. In effect the computer attempts to mirror what the brain *would have done* itself, but faster and more accurately. This means that the computer should 'know' what the human intends to accomplish, as if it were receiving the top-down prediction information from the brain directly. The principle works both ways, the human too needs to be able to predict the consequences of a computational action, know what the computer is doing, and be able to evaluate the current state of the data. The human or machine may deviate from one another, or deviate from progress toward the goal, therefore these discrepancies need to be easily

correctable. How fast the human and the machine can synchronise themselves is important (this leads to the fourth principle of throughput maximisation, below).

3. *Cognitive Pipelining Principle*:

   The process of instructing the computer of one's intentions is itself associated with cognitive load. Complex interfaces can demand a lot from the human's explicit processing capacity. If the implicit modules in the brain carry out interface manipulation beneath conscious awareness, cognitive load is reduced. The lower level, and the more encapsulated the module entrusted with these manipulations, the better. For every interface action one needs to ask how it can be presented in a way so as to be processed with as little dependency on context, and within as fast a perception-action loop as possible. This is reminiscent of pipelining instructions to a computer processor, and features in models of multitasking such as 'threaded cognition' [Salvucci and Taatgen, 2010]. We have evolved for natural interactions with the environment: therefore our brains possess sophisticated machinery for dealing with physical materials in 3D space. These built-in modules can be utilised by interface designers to provide skilful and intuitive interactions. If physical intuitions are inapplicable, a means for the user to train themselves to attain skilled automaticity should be provided.

4. *Throughput maximisation:* How fast does the data in the machine change to mirror an idea in the mind, and vice versa? How easy is it to faithfully express our ideas digitally? This can be measured using throughput—the rate at which information enters the input device. Expressive *output* from the machine to the user is also important. This corresponds to maximising throughput in the return path of the perception-action loop, and involves the human inferring

> the computer's state from its output. How well does the visual systems interpretation of the display relate to the structure of the data, the current task, and the potential actions one can take? How quickly can one evaluate the musical components one is working on? An expressive display offloads cognition because complex information is accessible 'at a glance'. Proprioceptive and haptic feedback contain less detailed information, but are far lower-level and hence faster than visual feedback.

There is still a long way to go before our everyday computing practice is fully embodied. One of the aims of this thesis is to show that throughput, as in the number of bits successfully specified per second, can be measurably improved using ideas of designing for the implicit sensorimotor system.

CHAPTER 3

Creativity

3.1 Introduction

Creativity is one of our most exalted and, some would say, mysterious behaviours. Creativity has acquired an almost mythical status in our society. Our artists, physicists, pop musicians and film stars seem to be allocated far higher social kudos than politicians, bankers and estate agents; jobs that are perhaps not seen as quite so creative. And yet, the study of the mechanisms by which creativity happens is still in its nascent stages. Not even artists would appear to have a clear grasp of their own creative process. It is no surprise that it represents a major challenge to research.

There is, at least, a general consensus on what one could call the 'minimal definition' of creativity. Creativity is defined as the generation of new, original ideas that are also useful. As the term 'useful' is perhaps biased toward utilitarian disciplines, the more general word ‘valuable’ will be used here.

The first necessary clarification here is, as Boden [1992] points out, that the mere novelty or improbability of an idea is not sufficient to capture true originality. She states that “A merely novel idea is one that can be described and/or produced by the same set of generative rules as are other, familiar ideas. A genuinely original idea is one that cannot.”. So true originality not only generates something different, but that also alters the very rules, assumptions and methods of a domain. Lesser forms of creativity are divided by Boden into “Combinatory”: ideas obtained by combining previously unrelated ideas, and “Exploratory”: ideas arrived at by randomly exploring parameters of the creative artefact. The production of truly revolutionary ideas that cannot be arrived at by merely exploratory processes is termed “Transformational Creativity”.

Another distinction found in Boden’s work is that between creative products that are completely new to human culture, and widely regarded as valuable by experts within a domain (referred to as historical, or H-creativity), and those which are merely new for a certain individual (referred to as psychological, or P-creativity). The former is obviously accorded higher status, involves a large amount of domain knowledge on the part of the creative person, and is highly dependent on the social and historical context of the work.

One might also draw a line between scientific and artistic creativity. Science and engineering address ‘well-structured problems’, where there are clear criteria for determining the success of a solution. In contrast, artistic endeavours tend to deal with ‘ill-structured’ or ‘under-constrained’ problems [Simon, 1974]. Here the rules are culturally generated, negotiable and provisional. This gives the artist more freedom to invent their own rules of evaluation, but makes objective assessment and generalisation difficult for researchers.

The line between artistic and scientific creativity is not as severe as it might appear however, and modern treatments tend not to treat them separately. Some of the most important ideas in science rewrote the rules of the domain, and artistic developments may often proceed via analytic, incremental processes [Root-Bernstein and Root-Bernstein, 2004].

Creativity theorists, according to Shneiderman [2000], fall into three general categories: *Structuralists*, those who hold that creative accomplishments can be attained by following some ordered process; *Inspirationalists*, who focus more on the 'Eureka moment' and who encourage breaking away from familiar thought processes and taking a more free-form approach; and *Situationalists*: those who emphasise social context and collaborative networks within a domain. This thesis, whilst acknowledging its importance, does not consider the Situationalist approach in any great depth. We will not claim to tackle the question of assisting truly original H-creativity, as this involves far too many wider issues of personality, historical or social context [Csíkszentmíhályi, 2009], and a consistently underrated amount of luck [Taleb, 2005]. What is attempted here is enhancement of the process of discovering, or performing, new, original and valuable material in a P-creative sense. I consider, along with many other researchers, that P-creativity involves a combination of both structuralist and inspirationalist approaches.

In this chapter we look at those cognitive aspects of creativity that are useful and interesting for designers of, and researchers into, digital musical interactive technology. We shall discuss models that involve two opposing but complementary mechanisms, in particular divergent and convergent thinking. Next we look at the phenomenon of insight, and the spontaneous emergence of ideas from the unconscious. We then describe Wiggins' computation models of creativity, including the 'suprisal threshold' for conscious access, and the Creative Systems Framework (CSF).

Some compelling results from the cognitive neuroscience of creativity will be mentioned, along with a particularly useful cognitive model (partially sighted variation and selection) that shall be used in later chapters. The subtle, but essential topic of how constraints can affect innovation is investigated in Section 3.6. The phenomenon of Flow is discussed in Section 3.7, along with its possible relation to the idea of creativity as a form of data compression.

Finally we look at some of the research into creativity support tools, and how interface design can affect creativity.

3.2 Divergence and Convergence, Variation and Selection

Guilford [1967] was one of the first prominent psychologists to draw attention to creativity as something that could be studied scientifically. The creative process was characterised as a combination of "convergent" and "divergent" thinking. Divergent production is the generation of many provisional candidate solutions to a problem, whereas convergent thinking is the generation of *the* unique solution. Others have taken convergence to mean the narrowing of the options to find the most appropriate solution, in other words idea selection. Whilst this is a simple model of a complex process, and throws up just as many questions as it answers (why should the mechanism for generation of a single solution be different from the generation of many?), most modern theories have these two processes present in some form. Sometimes the distinction is made between "Generative" and "Evaluative" instead. The fundamental ideas, that creativity is not 'magic', is exhibited by all human brains, and may be broken down into simpler, computable sub-processes is now fairly mainstream amongst researchers, and ultimately forms the foundations of the "Creative Cognition" [Sternberg, 1999; Smith et al., 1995] and"Computational

Creativity" fields [Colton et al., 2012].

Campbell [1960] and Simonton [1999] considered creativity as a Darwinian process, and propose a process of idea variation and selection. The most extreme version of this theory is Blind Variation and Selective Retention (BVSR) [Simonton, 2012]. BVSR is a process in which large numbers of alternative solutions to a given problem are generated blindly, that is to say that the variations are made without knowing whether they progress toward more valuable solutions or not. These solutions are then tested, or evaluated, and then the most valuable solutions are selected. The process can then iterate by varying the most successful selections from the previous iterations until some value criteria is met, or a definitive solution is found. The advantage of this model is that it can easily be modelled formally and numerically. However, it seems unlikely that idea generation is entirely blind, and the theory has been criticised on this and other accounts [Gabora, 2005]. Nevertheless, the key concepts of variation (by whatever means), evaluation and selection are certainly important for all theories of creativity. A Darwinian approach also provides a bridge to many powerful and illuminating ideas from evolutionary biology, such as the 'fitness landscape' [Wright, 1931]: the hypothetical assignment of a value function across all points in 'design space' [Dennett, 1995].

In my view, BVSR should be considered as a form of 'null hypothesis': a cheap fall-back when other more sophisticated processes are not available. We should not be surprised to find it operating in human creativity: Darwinian evolution is a mechanism that operates whether designed for or not. It is a fundamental and inevitable result of the passage of time in a complex system of mutating, replicating information. However, BVSR on its own would be immensely inefficient. Whilst a Darwinian process will find optimal solutions given enough time, there are limits to the size of its leaps through solution space. The presence of variation and selection does not rule out other mechanisms, particularly if those mechanisms are more efficient. Indeed, it is likely that natural selection will have selected precisely those cognitive mechanisms that *were* more efficient. The brain demonstrates a supreme aptitude both at predicting the direction of increasing fitness in solution space, and predicting the effectiveness of purely hypothetical solutions without testing them [Dietrich and Haider, 2014]. Simonton thus reformulates BVSR to incorporate the ability of the creative agent to predict value, and hence increase the sightedness of the variation mechanism [Simonton, 2012]. Therefore we could refer to a slightly sighted version of BVSR as 'partially sighted variation and selective retention' (PSVSR), this will be described in more detail in Section 3.5.

The Geneplore model [Ward et al., 1999] also features two complementary processes. The first is the generation of so called "pre-inventive structures", fluid collections of provisional, experimental concepts. The "-plore" stage is the exploration of those concepts in further depth. Generation of pre-inventive structures can happen in a variety of ways. *Combination* is where existing concepts are conjoined, and their properties merged in some way. *Transformation* is where properties of an existing concept are altered. *Analogy* is where a concept is transferred from one domain to another. A successfully applied analogy results if some properties that solved the problem in the first domain also apply in the new one. Analogy and conceptual blending are seen by Fauconnier [2001] as inherent in all our thought processes. A third component of the Geneplore model is problem constraints: these may be applied to constrain the type of ideas in the generation stage, or be used to guide the exploration stage. Constraints will be investigated in more depth in Section 3.6.

In all of these models, flexible, fast alternation between the processes is seen as crucial: innovation is often an iterative cycle.

3.3 Insight: The Role of the Unconscious

Stage models break down the creative problem-solving process into distinct phases. Helmholtz, Wallas [1926] and Hadamard suggest that the creative process breaks down into four main stages:

1. *Preparation*: involves researching the problem in question and trying, consciously, to solve it using existing techniques.

2. *Incubation*: in which the problem is left alone for a time, but the unconscious mind is hypothesised to be still working on the problem.

3. *Illumination*: where a sudden flash of insight occurs and the solution presents itself—seemingly out of nowhere.

4. Finally, *Verification* is when the solution is checked for its suitability.

Illumination is said to often happen when the mind is on other things, having just woken up in the morning, taking a walk or otherwise engaged in some non-demanding task. Illumination is more or less synonymous with “insight”, and is characterised by a sudden “Aha!” or “Eureka!” moment [Kounios and Beeman, 2009]. Insight problems are a tool that psychologists have used to study this phenomenon. These are puzzles that no amount of step-by-step reasoning can solve. These problems often involve setting up some “functional fixedness” [Duncker and Lees, 1945], commonly known as a “mental block”. The insight occurs when the problem is suddenly seen from a different angle. The “special process” model holds that these problems require different brain processes from logical or verbal problems, and are non-conscious or at least indescribable. This is evinced by the fact that verbalisation of the problem solving process hampers progress in insight problems but not in non-insight problems [Schooler et al., 1993].

There may also be preliminary stages to problem solving, namely problem finding and problem definition [Csíkszentmíhályi, 2009]. In fact asking the right questions may be the key to a breakthrough. As Einstein, quoted by Getzels [1975] put it:

> "The mere formulation of a problem is far more often essential than its solution, which may be merely a matter of mathematical or experimental skill. To raise new questions, new possibilities, to regard old problems from a new angle requires creative imagination and marks real advances in science."

This is strongly reminiscent of reflective cognition, discussed in Section 2.3.6.

3.3.1 Remote Associations and Spreading Activation

One mechanism that has been proposed for the operation of insight is 'spreading activation' [Anderson, 1983]. The properties of associative memory—the network of associated concepts, where one can pass from concept to concept if they are closely related—can be employed to explain the effect of priming [Meyer and Schvaneveldt, 1971] and insight [Langley and Jones, 1988]. Priming is the phenomenon where if a subject is presented with a word or image, even subliminally, related concepts are processed faster. Therefore there must be some means by which heightened availability of concepts resonates outward through the associative network. This is clearly related to the idea of 'divergence' in the previous section.

An individual's ability to utilise this branching network can be tested using 'Remote Associates Tests' (RATs) [Mednick, 1962]. In these studies, subjects are given pairs or triplets of seemingly unrelated words, and asked to find another word that links them, so as to form compound words. For example, given the triple: 'safety', 'cushion' and 'point', the correct answer would be 'pin'. There are more prosaic methods of solving these problems, for instance enumerating every possible word that could go before or after 'safety', then checking each in turn with the other two words. Therefore in some experiments subjects are asked to self-report their feeling of having solved the problem in a flash, via insight. The solving of insight problems has been found to be facilitated by exposing subjects to subliminal cues [Hattori et al., 2013], further bolstering the case for unconscious associative processing being involved.

The associative nature of memory does not always assist in generating original solutions, however. Smith et al. [1993] demonstrate that prompting subjects with examples of potential solutions to a design problem diminished the originality of their suggestions, because they tended to closely resemble the examples.

Given the importance of fluently generating a large quantity of varied ideas, a proposed test for creative abilities is the 'Unusual Uses' test [Wilson et al., 1953]. This evaluates a subject's ability to generate ideas via divergent production. The subject is required to come up with as many unusual uses as possible for an everyday object such as a brick or a cardboard box. The ideas are then simply counted, and can also be rated for originality. The unusual uses test has been used to study the effects of various externally imposed conditions on divergent abilities. One relevant example is an experiment to investigate incubation [Baird et al., 2012]. After an initial unusual uses session, different groups of subjects carried out an undemanding task, a cognitively demanding task, and a period of complete rest. The subjects that carried out the undemanding task performed the best when returning to the task, despite not having consciously thought about it. This suggests that incubation can be assisted by mind wandering. This finding clearly has ramifications for creative interfaces, as a demanding interface will be less supportive of mind wandering than a simple one (see Section 5.5.4).

The quality of attentional processes has also been regarded as important for creativity [Ansburg and Hill, 2003; Vartanian et al., 2007]. In particular, the spectrum between diffuse and narrowed attention. When focused on a task, executive control activates those regions of the brain needed to carry out task-relevant processes. It is also inhibiting other regions, and discarding sensory input deemed to be irrelevant. Whilst this is a good thing for carrying out convergent, step-by-step tasks, it is hypothesised to have negative effects on remote associations: hence inhibiting creativity. The opposite of this narrowed focus is 'diffuse' or 'distributed' attention, where the mind is more open to extraneous information, both internal and external [Takeuchi et al., 2011]. This relates back to an earlier idea, that of a 'stimulus generalisation gradient' [Mednick, 1962]: the idea that the width of the associative hierarchy is important for remote associations, and indeed can be measured. An individual with a narrow associative spread will tend to only activate concepts that are closely related to a stimulus, whereas a highly creative individual may have a shallower gradient that will activate many, less related concepts. Gabora [2002] describes a process by which the creative agent moves from an *associative* mode, where the activation function is widely spread across many related concepts, to a *causation* based mode of thinking, which enables the investigation of the ramifications and implementations of the creative work. Again, it is the ability to combine 'conceptual fluidity'[1], with analytic reasoning that is seen as the key to creative ability. More recent investigations [Benedek and Neubauer, 2013] of this idea seem to indicate that creative people are not so much characterised by a flat associative hierarchy, rather by 'fluency', or the speed with which they can traverse the network.

Whilst insight and illumination may be synonymous, inspiration can refer to more of an affective state than a particular mechanism [Oleynick et al., 2014]. One may deliver an 'inspired performance', where the mind seems to be in a particularly productive, motivated, receptive and heightened state. In this state the probability of insight is higher, or feels to be so subjectively. One can also be inspired by

[1] Otherwise known as defocused/diffuse attention or 'reduced latent inhibition'

some external influence (evocation), and hence be put into some state where one wants to be creative, and feels enabled to be so. Inspiration may be thought of as a disinhibition mechanism, as proposed in improvisation experiments [Limb and Braun, 2008; Liu et al., 2012], where, given the right conditions, the artist seems to let go of normal everyday inhibitions and error correction processes in order to enhance top-down creative expression.

3.3.2 The Informational Dynamics of Insight

An important question is whether insight requires both generation and evaluation to occur beneath the level of consciousness. Does the subconscious mind need to post all its productions into consciousness, which is forced to evaluate them all explicitly? Or is there a preconscious selection process operating? To evaluate the combinatorial possibilities of all connected concepts consciously would take a huge amount of time, so it seems likely that some form of evaluation and selection must be conducted subconsciously. Wiggins [2012] draws on Global Workspace Theory to propose a mechanism for how unconscious probabilistic processing can give rise to the phenomenon of insight; highlighting the importance of the 'informational dynamics' of the interplay between the implicit and explicit systems. His hypothesis is that the generating units of GWT monitor their own information-theoretical expectation violation i.e. 'surprise'. The likelihood of obtaining access into the GW is proportional to this surprise[2]. The formulae for the threshold of consciousness is given as the product of the number of generators making a certain prediction, p, and the surprisal/novelty of the prediction h, divided by the entropy of the distribution of possible events, prediction-H: which is the uncertainty a predictor has about its

[2]Strongly reminiscent of the prediction errors in the Free-Energy principle described in Section 2.2.7.

prediction:

$$T = \frac{p \times h}{H}. \tag{3.1}$$

There are therefore three ways of increasing the chance of crossing the threshold:

- If lots of predictors have the same idea, and the likelihood p is large.
- If the idea is very surprising, and h is large.
- If prediction-H is low, the generator is very certain that its idea is 'right'. This also encourages admission into explicit awareness.

It should be noted that this is a highly competitive process: the idea will be in competition with myriad other ideas and with incoming sensory data. The idea that internally generated ideas are in competition with externally generated ideas is returned to in Section 5.3.1

3.4 Computational Models of Creative Cognition

3.4.1 The CSF

Creativity is also studied in the context of artificial intelligence: a field known as Computational Creativity. By attempting to build artificial systems that exhibit creative behaviour, we may form models of how creativity might function in our own minds. Wiggins' Creative Systems Framework (CSF) [Wiggins, 2006] is a more formal descendent of Boden's theories of artificial creativity [Boden, 1992]. In this framework creativity is seen as a way of extending conceptual space: using a traversal mechanism that produces a concept falling outside of the existing space (an "aberration"), but is nevertheless seen as valuable and appropriate according to the evaluation function of the domain.

Wiggins describes creativity in terms of the exploration of conceptual space. It consists of the universe of all possible concepts $\mathscr{U}$, an existing conceptual space (for example domain knowledge) $\mathscr{C}$, rules (domain constraints) that define this conceptual space $\mathscr{R}$, a set of techniques to traverse the space $\mathscr{T}$, and an evaluation method $\mathscr{E}$: a way to assign value to a location c that yields a "fitness function".

Exploratory creativity is said to proceed as follows: if a traversal rule takes us outside the space of existing concepts, this results in an "aberration": a novel concept. If the aberration proves valuable according to $\mathscr{E}$, then the new point is included in the domain, and the conceptual space is extended to include this point. Wiggins claims that transformational creativity (a fundamental shift in the rules of the domain) can be viewed as no different from exploratory creativity but on a meta-level. This is to say that a transformation of conceptual space can be achieved by exploring the conceptual space of conceptual spaces. Clearly there is no limit to how 'meta' this process can get, giving rise to a creative ratchet effect [Leman, 2008, p. 54]. Later we attempt to adapt this model to apply to a parameter space, to propose what creativity might mean in the case of adjusting continuous controls of a sound synthesis engine (Section 5.2.1).

3.4.2 Novelty Based Search

An interesting critique of convergence oriented algorithms is found in neuroevolution literature [Lehman and Stanley, 2011; Nguyen et al., 2015]. By evolving different neural networks with reward based on either progress or novelty, this research demonstrated that novelty-oriented search can significantly outperform objective-based search. Even well-defined problem spaces—such as maze navigation and simulated biped walking tasks—seem to be performed better by selecting for novelty-generating behaviour. The reason for this is that novelty producing behaviours are more generalisable than progress oriented behaviours. An agent that has discovered how to walk, crawl or hop in any direction, and 'enjoys' exploring new areas will do better in a maze than an agent that can only crawl towards the goal. This point needs to be emphasised. According to Stanley and Lehman [2015], planned progress may in fact be a 'myth'. By enforcing a constant need for measurable progress toward an explicitly defined goal, society may be stifling innovation in the most important and complex of domains. Therefore creativity is not just 'evolutionary cheesecake' [Pinker, 1999; Bown, 2012]: the drive to produce novelty is fundamentally adaptive.

In domains where originality is often seen as *intrinsically* desirable, such as art and music, novelty driven exploration should be even more appropriate. Evaluating musical interfaces by exclusively testing goal directed behaviour may therefore miss the most important part of the picture. User studies that test for unplanned behaviour may be more revealing (e.g. Gurevich et al. [2012]; Zappi and McPherson [2014b]).

3.4.3 Evaluating Creative Output

The cognitive processes by which artists and audience evaluate cultural artefacts are quite possibly the most complex of all the aspects of creativity. It is perhaps the biggest challenge facing the computational creativity field: how to get a computer to judge its own work in a sophisticated, cultured way [Cardoso et al., 2009; Galanter, 2012]?

How can we tell 'how much' creativity has happened? Empirical studies of creativity may require that a participant's output must be judged for its originality and appropriateness. Of course, it is a tough prospect to reduce creativity to a scientifically or computationally measurable quantity [Jordanous, 2012]. It may necessitate the 'consensual assessment technique' where a panel of (human) experts in the field are called upon to judge the outputs of the experimental participants [Hennessey and Amabile, 1999]. This may be an effective method of evaluating creativity in terms of its final products, but is quite costly in terms of time and resources.

For these reasons, the full cultural and aesthetic complexity of how creative works are evaluated lies outside the scope of this thesis. In general we assume that musicians must evaluate their own output as they create it. This self-evaluation lies outside the design of music technology, as the full complexity of the fitness function is latent in the artist's or listener's brain. Nevertheless, it is important to acknowledge that the musician's evaluative process may be a complex and demanding one, and an interface that induces high cognitive load may interfere with this [Mycroft et al., 2013]. We must also consider that evaluation may also proceed in a predictive fashion: the adjustment of musical parameters may proceed by means of a prediction of the direction of increasing value in some conceptual space. In Section 5.5.2 we propose that the difference between divergent and convergent interactions is entirely characterised by the engagement and disengagement of these predictive and evaluative processes. Whilst acknowledging the importance of evaluation, the actual details of evaluative processes will be spared.

3.5 The Cognitive Neuroscience of Creativity

There have been a number of studies investigating creative cognition using brain imaging techniques. For recent overviews, see Dietrich and Kanso [2010] and Sawyer [2011]. Reviews such as this point out that much of the literature is inconsistent, and approaches the topic in too vague a way to produce clear results. For instance distinctions such as divergent and convergent thinking, or generative or evaluative thought, are not sufficiently low-level constructs to produce distinct activation patterns in neuroimaging studies. No 'creativity region' or 'divergence module' is likely to be found in the brain. Nevertheless, some studies contain some interesting and suggestive findings.

Ellamil et al. [2012] have conducted fMRI scans of people engaged in creative activity, and report on the activation of different brain regions during generative and evaluative stages. Their findings indicate that the generative stage makes use of associative processing. Perhaps more interesting is that the evaluative stage appears to activate components of two brain systems thought to be seldom used in conjunction: the default (task-negative) network and the executive (task-positive) network. They speculate that this is because creative evaluation involves a mixture of introspective and analytical thought.

Another method to separately investigate the generative and evaluative stages is to study improvisation: which presumably emphasises generation rather than evaluation. There is some evidence that improvised, spontaneous creativity involves the disinhibition of certain evaluative, monitoring processes. In an fMRI brain imaging study of jazz improvisation on the piano, Limb and Braun [2008] found that:

> "Improvised performance is characterized by dissociated activity in medial and dorsolateral prefrontal cortices, providing a context in which stimulus-independent behaviors may unfold in the absence of conscious monitoring and volitional control."

The dorsolateral prefrontal cortex (DLPFC) is one of the main nodes in the 'task positive network', and is implicated in the maintenance of working memory and control of attention. It seems that reducing cognitive control allows faster, more fluent generation of content. These results were replicated in another study of improvisation, this time lyrical (freestyle rap):

> "Improvisation, contrasted with conventional performance, was in general associated with relative decreases in activity in supervisory attentional and executive systems... An alternative, direct route through cingulate pathways into the motor system may allow the medial frontal regions to generate novel, exploratory behaviours, bypassing conventional executive controls and thereby providing the cognitive flexibility necessary for successful improvisation."[Liu et al., 2012]

It seems action commands are sent *down* the motor hierarchy but not *across* to the evaluative systems at the top level, therefore in some ways this mechanism resembles an 'open-loop' motor command, but operating at a higher level in the predictive hierarchy. Of course, this requires that the lower-level systems are already highly trained to produce appropriate behaviours: both the studies above were conducted with expert practitioners. The idea of implicit modules autonomously generating skilled behaviour recurs in the discussion of Flow (Section 3.7).

Flying in the face of this compelling idea, Bengtsson et al. [2007] found that DLPFC regions *were* more activated during piano improvisation, one of the many contradictory findings in this area. These disagreements indicate that greater rigour is needed in constructing computational models of creative thought, and mapping them onto brain processes [Dietrich, 2007]. One preliminary candidate for such a model is proposed in Dietrich and Haider [2014]. Here, an evolutionary (but partially sighted) approach is proposed. Like Wiggins, they stress the contribution of—and interplay between—implicit and explicit cognition. The free-wheeling generative, associative mode of unconscious 'noise' can fuel creative thought, therefore provide an element of blind variation. On the other hand, the ability of the explicit system to chain together distantly related concepts and to explore and evaluate hypothetical ‘thought trials’ may vastly accelerate this generate and test process. They make reference to embodied cognition, and the connection between sensory prediction and action:

> “The fact that our cognition is embodied implies that movements, or the emulation of movements, are important in finding new ways to solve problems. Creativity, in this case, is to generate a predictive goal representation as well as to emulate our way to it; that is, creativity is to find the evolutionary algorithm that binds the goal state to the problem by way of a series of motor plans.”

By preserving purely hypothetical, imaginary world states in working memory (‘predictive goal representations’), the explicit system can then work backwards to construct the means to this end, thus leaping over possibly non-viable regions of solution space via ‘cognitive scaffolding’.

This PVRSR model features three of the four strategies of the EARS theory of creative interaction in Section 5.5.3. What I believe is missing from this picture is the reflective component of the explicit system. The ability to introspect, question and alter patterns in one’s own creative behaviour has a huge impact on creativity. ‘Cognitive scaffolding’, whilst immensely important, is not the only role that the explicit system can play. Therefore in Section 5.5.3 we will appeal to Stanovich’s notion of the reflective mind [Stanovich, 2009], and posit a second, more divergent, contribution from consciousness.

3.6 The Constraint Paradox

The phrase ‘creative freedom’ seems to imply that the less constraints one is bound by, the more creative one can be, but in reality the situation is more complex.

One thing that the proliferation of music technology has made clear is that it is often necessary to restrict one's options in order to be more creative [Magnusson, 2010; Gurevich et al., 2012]. Whilst there is obviously no such thing as completely unconstrained creativity—we are operating within a universe with physical laws, and we have cognitive limitations—there are also constraints that culture imposes, and rules that artists agree to submit to. More tellingly, many artists choose to deliberately constrain themselves. Stravinsky claimed:

> "My freedom will be so much the greater and more meaningful the more narrowly I limit my field of action, and the more I surround myself with obstacles. Whatever diminishes constraint diminishes strength. The more constraints one imposes, the more one frees one's self of the chains that shackle the spirit." [Stravinsky, 1970].

Stokes relates how Picasso, Monet, Schoenberg and many others used constraints to guide the creative process [Stokes, 2006]. She touches upon the notion of creativity being a strategic search through solution space, and refers to constraints as

> "...barriers that lead to breakthroughs. One constraint precludes (or limits search among) low-variability tried and true responses. It acts as a barrier that allows the other constraint to promote (or direct search among) high-variability novel responses."

A comprehensive study of creativity in organisations was conducted by Joyce [2009]. She found "that the degree of constraint imposed on a creative task affects individuals' creative outcomes in a curvilinear fashion, such that a moderate degree of constraint was optimal", and "the benefits of choice to creative outcomes quickly drop off as the number of choices becomes overwhelming". Her conclusion was:

> "Paradoxically, the freedom of creating with very little constraint can result in a narrow-minded creative process. The logistical overwhelm and

> confusion resulting from unfocused search can actually restrict teams' open mindedness... members increase their reliance on their own assumptions."

Whilst some studies have found evidence for constraints leading to enhanced creativity, some have found the exact opposite. There are many examples of situations in which *loosening* constraints leads to creative breakthrough. For example, Amabile and Gitomer [1984] showed that children who were allowed to choose their own materials produced more creative collages. Polya descibes the "Inventor's paradox" [Polya, 2004], where the inventor has to free themselves from the ostensible constraints of the problem. Amabile also found that overly strict constraints imposed from the outside can inhibit creativity [Amabile, 1998]. This is principally because intrinsic motivation is an important determinant of creativity. Intrinsic motivation is the willingness to do something for its own sake, and has also been linked to Flow (see section 3.7).

Constraints are a recurring topic in the literature relating to musical interaction. Later, in Section 5.5.4 we propose an explanation of why constraints should encourage creativity. This explanation should help to make the situation seem less paradoxical, and make designing and interacting with these constraints more productive.

It could be that there is an optimal level of constraint. Elster [2000] states "inspiration — defined as the rate at which ideas move from the subconscious into the conscious mind — can be defined as an inversely U shaped function of the tightness of the constraints". Next, we will discuss how, if the balancing act between constraint and freedom is achieved successfully in a challenging activity, it can have a significant positive effect on the subjective well-being of an individual, contributing to an phenomenon known as "Flow".

3.7 Flow, Complexity and Creativity as Data Compression

3.7.1 Flow Theory

Flow was a term coined by Csíkszentmíhályi [1991] to mean a state of complete absorption in an activity, and is particularly associated with creative work. Another popular term, for instance in sports psychology, is "being in the zone" [Young and Pain, 1999]. Flow is "an almost automatic, effortless, yet highly focused state of consciousness" [Csíkszentmíhályi, 2009]. Flow goes beyond mere enjoyment. This state of mind is both a productive one—in that people experiencing Flow feel that they are operating in a highly effective and creative manner—and one that also generates very positive affective states, counting alongside some of the best experiences in people's lives. It frequently arises in discussions of music listening and performance [Armstrong, 2006; Diaz, 2013; Pachet et al., 2013; Wrigley and Emmerson, 2013; Nash, 2012].

The defining characteristics of the Flow experience are:

- Being engaged in a *challenging activity*, but having the skills to meet that challenge.
- *The merging of action and awareness:* the person feels 'at one' with the task.
- *Clear goals,* and immediate and unambiguous feedback.
- *Concentration:* irrelevant information does not impinge on consciousness.
- *The sense of control:* somewhat paradoxically, despite the task being challenging and perhaps unpredictable, the person feels as though they are in control, and fear of failure does not arise.

- *Loss of self-consciousness:* the person ceases to concentrate on egoistic concerns.

- *Transformation of time:* a single task can seem to pass very slowly, on the other hand large amounts of time seem to fly past quickly.

- *Autotelic experience:* the activity becomes intrinsically rewarding, and concerns about external threats or rewards diminish.

Presumably the tools we use can also hinder or encourage the Flow experience [Selker, 2005]. Whilst Csíkszentmíhalyi himself draws no distinction between preconditions for this state and the characteristics of it (it is a psychological process that feeds into itself, in that state is interdependent with process), he also encourages changes to be made in our working environment, and society in general, such that flow becomes more natural. For the purposes of 'designing for flow' [Pearce and Howard, 2004], it seems useful to separate the eight dimensions into preconditions and resultant subjective experiences. For instance, the transformation of time and the merging of action and awareness are more emergent mental qualia, but clear goals and feedback are more external preconditions. The sense of control, and the necessity of applying skills are also properties of interactive systems that can be designed for. HCI research has been carried out in this area [Ghani and Deshpande, 1994; Webster et al., 1994; Bederson, 2004; Van Schaik and Ling, 2012]. Typically, Flow is measured by applying the 'experience sampling form' relating to the eight dimensions of Flow, and correlating these dimensions with the various experimental conditions [Bakker, 2005; Nash and Blackwell, 2011; MacDonald et al., 2006].

Despite Flow being a highly influential notion, the theory[3] does not have much to say about the cognitive processing underlying the state. Dietrich claims "Next to nothing is known about the brain mechanisms that give rise to such exceptional

[3] It may be more of an observation than a "theory".

human functioning" [Dietrich, 2004b]. His hypothesis is that the implicit system becomes so effective at performing skilled tasks that the frontal lobes do not need to process any un-dealt with information. The successful completion of tasks activate reward mechanisms, therefore the experience is pleasurable. Dietrich describes the flow state as:

> "...a period during which a highly practised skill that is represented in the implicit system's knowledge base is implemented without interference from the explicit system... a necessary prerequisite to the experience of flow is a state of transient hypofrontality that enables the temporary suppression of the analytical and meta-conscious capacities of the explicit system."

In other words, the lower levels of the brain's hierarchy are anticipating events so well that the explicit system receives no top-level prediction errors at all: and is free to simply sit back and enjoy the spectacle unfolding. This state of 'hypofrontality' (low activation of the Prefrontal Cortex, see Section 2.3.2) has interesting parallels with altered states of consciousness [Dietrich, 2003].

Flow has been found to correlate with musical creativity [MacDonald et al., 2006]. Reassuringly, it seems that optimal subjective experience of the user is very much related to optimal quality of the created artefact, bolstering the "research through design" approach to creative composition software [Zimmerman et al., 2007]: if the user feels good, then one would hope they are producing valuable artefacts. On the other hand, some musicians report that bouts of Flow experiences do not always produce valuable creative works. There is a danger that excessive absorption in the task may inhibit the essential reflective, meta-cognitive component of creativity. Dietrich's 'hypofrontality' might imply reduction of critical thinking abilities. This would perhaps explain the widespread "it didn't sound so great the next day"

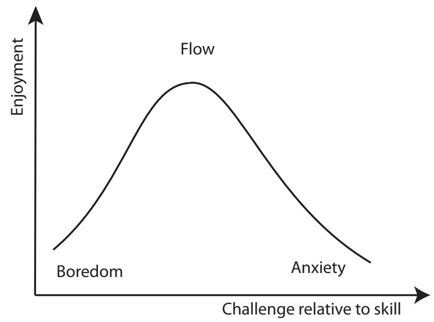


Figure 3.1: Optimal complexity: the right amount of challenge produces the conditions for flow.

phenomenon that musicians find all too familiar.

3.7.2 Flow and Predictive Coding

Returning to the question of complexity vs. constraint, Csíkszentmíhalyi also indicates that "It is likely that both too little and too much freedom... are inimical to creativity" [Csíkszentmíhályi, 1999]. Fig. 3.1 sketches this enjoyment peak at the happy medium between boredom and anxiety. This mirrors the situation described in Section 3.6, and similar results can be found throughout psychological research. Many inverse-U type figures resemble both in shape and conceptual basis the Wundt curve of novelty vs. affect Berlyne [1970], whereby the pleasure derived from a stimulus varies as an inverse-U curve with its complexity, and also Hebb's plot of arousal vs. rate of response and learning [Hebb, 1955]. It also recurs in a study of enjoyment vs. familiarity in popular music [North and Hargreaves, 1995], and studies of anxiety vs. performance in sports [Raglin and Turner, 1993]. Of course, just because all exhibit a similar shaped curve doesn't mean they are truly related, or stem from the same mechanisms. However all are related to complexity, i.e. entropy, and all are related to reward mechanisms, i.e. reinforcement learning.

This leads us to the provocative and powerful idea put forward by Schmidhuber [2009, 2010], which may be a route to establishing Flow theory on a more rigorous computational foundation. Schmidhuber posits that this optimal complexity point represents the point of maximum rate of change of experience-data compression. The brain's desire is to encode and predict the environment, and therefore seeks out experiences that involve active reduction of the subjective complexity of the world. The faster the brain reduces the complexity, the more the reward mechanisms of the brain make us feel good. This is a highly adaptive trait: better coders/predictors of their environment will tend to thrive within it. If the input is too complex, the brain lacks the necessary tools to encode it: it will feel subjectively anxious and seek to move toward less challenging sense data. If the input is too simple, there is no subjective complexity reduction potential: in the extreme case that it is uniformly ordered, there is no initial subjective complexity, and in the opposite extreme case that the data is uniformly random, there is also no chance of reducing that complexity. In these cases the brain experiences the subjective feeling of boredom, and attention will switch to more 'interesting' data. Interestingness being defined as having more compression-potential.

This elegantly explains exactly why we are curious and novelty-seeking creatures. It also chimes well with the scientific endeavour as a whole: science is the drive to explain the world using simple mathematical models[4]. Most relevantly to our current analysis it may also explain why we enjoy art and music: it is a way to experience artefacts that often appear intricately complex, but have some underlying patterned structure that reveals itself to us over time. It also explains the intrinsic reward that artists get when they *produce* art that 'expresses themselves': their own experiences are complex and subtle, therefore to express/compress them in something as simple as a 3 minute pop song, or a 17 syllable Haiku [Buchanan, 2001], renders their self-knowledge more compact and activates these reward mechanisms [Schmidhuber, 2012]. Clearly this idea chimes neatly with the Free-Energy principle discussed

[4]Perhaps the most surprising thing of all is how well these simple models work [Wigner, 1960]

earlier: our creative drives emerge naturally out of our predictive drives.

Csíkszentmíhalyi defines the opposite of Flow to be 'psychic entropy', a term borrowed from Jung [Butz, 1992]. This is far from being a formally defined concept, but is described as a disordered state of mind, involving conflicting internal goals, or 'noise' in the mind. There seems to be no attempt to utilise entropy reduction as an actual measure of Flow, perhaps because the idea smacks of naive reductionism. Nevertheless, a reduction of entropy does equate to information flow. And as we saw in Section 2.2.7, reduction of relative entropy in both an organism's internal model and sensed environment is a fundamental behaviour. So could we have, if not a rigorous proof, at least a strong connection between the ability to reduce entropy in a parameter space, and the resulting psychological state of the interacting agent?

This potential link between Flow, entropy and information raises many tantalising questions. Is there an optimal level of subjective complexity for music making? Is it possible to relate the information-theoretic processes involved in a creative perception action loop to some of the more enigmatic subjective aspects of the Flow state? Is there a way to measure the rate of reduction in entropy that the creator is achieving? This last question will be tackled in the next chapter.

3.8 Creativity Support Tools

Computers obviously provide a great extension to our creative powers [Shneiderman, 2000]. Not only this, but the possibilities they open up may increase the ability to think creatively: for example, Hanna [2012] monitored two groups, 'intensive' users of Computer Aided Design (CAD) software and 'casual' users. This study found significant positive correlation between increased CAD use and individuals' ideational fluency (see Section 3.3.1).

Despite the advantages, there are many points of friction between computer systems and creative practice. According to Selker [2005], "Possibly the most important creativity enhancers include recognizing and promoting a state of flow", but he points out that this can be broken by interruptions, shallow inflexible undo options, and working memory load: "interface design can carelessly add extra mental effort when it is not necessary".

How can software systems be designed to aid creativity? According to Lubart [2005], computers may facilitate:

1. the management of creative work,
2. communication between individuals collaborating on creative projects,
3. the use of creativity enhancement techniques,
4. assisting the creative act through integrated human-computer cooperation during idea production.

Many studies look at item 1 and 2 in Lubart's list, investigating organisational creativity, and the potential for information technology to enhance creativity in groups [Nunamaker Jr et al., 1987]. However, the last two of these are of particular interest here.

One set of design principles for creativity support tools specifically relates to interfaces. Shneiderman [2007] and Resnick et al. [2005] propose the following guidelines, listed with some relevant connections to musical interaction design:

- *Support exploratory search.* Divergent navigation through the solution space is essential.
- *Low thresholds, high ceilings, wide walls.* The barrier to entry for novices must be low, but there must be enough power for experts to engage in highly skilled interaction. There must be a large space to explore. These notions recur in the musical interaction literature time and again (see Section 4.2).

- *Support many paths and styles.* Creativity can be a highly individual activity, therefore unique ways of using a tool must be allowed. There are multiple routes to attain a creative goal, and they may be highly indirect.

- *Enable collaboration.* Collaboration is certainly important for musicians, and more research in this area (e.g. [Hattwick and Wanderley, 2012; Xambó et al., 2013; Bengler and Bryan-Kinns, 2014; Murray-Browne et al., 2014]) is surely needed, given the currently poor provision for groups in commercial music software. However, collaboration lies outside the scope of this thesis.

- *Support open interchange.* This relates to the idea of being able to move information around, place it in different contexts, and transform it in different ways.

- *Make it simple.* Reducing cognitive load is again essential.

- *Choose black boxes carefully.* Whilst encapsulated processing of information is good for reducing cognitive load, it may reduce the feeling of autonomy and inhibit the artist's ability to experiment with the processes.

- *Invent things you would want to use yourself.* A personal, implicit understanding of the creative community's practices is immensely helpful in designing 'cultural' artefacts such as musical instruments.

- *Provide rich history keeping.* Session histories can promote new ways of thinking by providing an overview of one's own creative process [Shneiderman, 2000]. This may relate to reflective meta-cognition (Section 5.5.3).

- *Balance user suggestions with observation.* Users know what they want, but don't always know their own mind, especially with regard to the unconscious aspects of cognition that are essential for creativity.

- *Evaluate the creative product.* Ultimately the success of a tool is measured by the success of what is produced with it. However, in creative fields evaluation is complex and somewhat subjective. Evaluation of musical instruments will be discussed in Section 4.2.1, however we intentionally leave aside the issue of evaluating musical output.

Each of these topics is extremely relevant to music creation technology and surfaces in a number of places in the computer music literature (see Chapter 4). They also tie in well with the HCI principles in the last chapter. Many of these guidelines make sense in view of the four-strategy model detailed in section 5.5.

These considerations have been formalised into a psychometric test to measure how well creativity is supported by computer systems: the Creativity Support Index (CSI) [Cherry and Latulipe, 2014]. It provides six carefully de-correlated[5] dimensions by which to assess the various aspects of creativity: Exploration, Expressiveness, Immersion, Enjoyment, Results Worth Effort, and Collaboration.

Cognitive models of creativity have been used for the design of creativity augmentation systems. The distinction between divergent and convergent thinking has been used to this effect. Work in this area includes software that generates 'interactive suggestions', where the designer can see both divergent variations and convergent suggestions for improvement alongside their current work [O'Donovan et al., 2015]. Drawing software that elaborates on the artist's input [Davis et al., 2014] has been inspired by an enactive, embodied approach to computer use: here the computer becomes an 'artistic colleague'. The injection of randomness into the creative process has been used for a number of 'divergent' systems, but as André et al. [2009] point out, there may be more structure to serendipitous discoveries than mere random variation. Darwin's Gaze [DiPaola et al., 2013] is a visual art system

[5] These six dimensions were distilled down from more, using a similar methodology to the NASA TLX workload questionnaire [Hart and Staveland, 1988] used in Experiment 3.

inspired by genetic programming and dual process theories that provides associative variations of portrait images.

Despite the fact that almost all creative disciplines now use computers in some way, and that almost all content creation work could benefit from increased creativity, specific research into how technology can support or encourage creativity is not as widespread as one might expect. In their review of the creativity support system literature, Müller-Wienbergen et al. [2011] identify:

> "...a lack of research on how to design IT systems that support both convergent and divergent thinking in creative work. Existing research either postulates rather generic design requirements lacking any detailed specification on how to address these requirements, or it focuses on a specific IT system that only supports divergent thinking. Due to the central role of both modes of creative cognition, and their intimate relationship in solving creative problems, we contend that there is a need for a detailed design specification considering both 'levers' to support creativity. "

They also propose a number of design principles and experimental hypotheses, reminiscent of those in Experiment 1, but for discrete 'knowledge item' search systems, rather than continuous parameter spaces.

Finally, the interactive process itself can be used to study the creative process: the 'studio as laboratory' approach [Edmonds et al., 2005]. How people interact with a parameter space can be logged, and their search trajectories studied [Jennings et al., 2011].

3.9 Summary

To design for creativity, augment it, or even just to avoid interfering with it, it would seem to be useful to have a model of how creativity operates. Whilst these models are far from the level of sophistication required to artificially generate transformational creative works, there seems to be consensus on a number of themes. These we can utilise for the purpose of informing musical interaction design. There are a variety of models of creative cognition with varying terminology, but all necessitate a mechanism for combining or transforming existing ideas, a means of evaluating and selecting the best ideas, and all posit a crucial role for the interplay between conscious and subconscious mind.

Creativity is a complex phenomenon, and one that is unlikely to be pinpointed as emerging from a single brain region. Rather, this discussion of recent creativity models such as Dietrich's, Schmidthuber's, and Wiggins', points towards creative principles being built into the fundamental workings of the mind. Even our basic everyday perceptions of the world could be considered, by any reasonable definition, creative, as they project onto sensory data an internally generated predictive model of the world. The modularity we find in biological structures such as the brain has evolved for the purposes of extreme flexibility, and this implies we are capable of continuously generating novel behaviour. Therefore, great H-creative works differ from everyday P-creativity in degree, not kind.

Flow is a remarkable mental state that may accompany the optimal use of creative tools. Designers of technologies where interaction in the Flow state is desirable should think equally about the state of mind of the user, as well as the goals of the user. Use of a content creation system should perhaps be considered less as a means by which a brain turns its desires into reality, and more as a facilitator of a constantly unfolding, delicate mental balancing act; in which the artist maintains themselves in an optimal state for creativity to occur. Therefore, rather than seeking to enhance creativity by targeting some 'special' aspect of human computer interaction, in a way we simply need to enable the brain to do what it does best, and enjoys most. However the role of implicit skill seems to be a crucial precondition to Flow, and thus should be designed for carefully.

The following characterisations of creative cognition inspire the EARS theory in Chapter 5:

1. *The Creative Systems Framework (CSF)* [Wiggins, 2006] Exploratory creativity occurs when a solution space traversal mechanism produces a useful 'aberration': a concept that falls outside the current conceptual space. Transformational creativity can occur if conceptual space is extended on the meta-level.

2. *PSVSR theory* [Dietrich and Haider, 2014] Creativity features variational processes that can vary between blind and sighted, hence 'Partially Sighted Variation and Selective Retention'. The implicit system provides short term predictions and evaluations unconsciously; the explicit system can leap over large unviable regions of solution space via 'cognitive scaffolding'.

3. *Everyday insight from the 'surprisal' threshold.* [Wiggins, 2012] This is the claim that creative cognition differs from our everyday cognition in degree but not in kind. Artistic behaviour may have evolved from more basic action-perception coupling. Imagination — and ultimately creativity — can emerge naturally from the cognitive principles outlined in the previous chapter. Novelty, being defined as something surprising or unexpected, can be formalised as having high information content relative to some agent's predictive model. Creativity may be an inherent property of implicit-explicit thresholding, due to the selection mechanism's dependence on information-theoretic surprisal. Truly sudden and momentous insights progress via a similar process, but would involve far more preparation and training of implicit solution generators and solution recognisers.

4. *Expression as Compression* [Schmidhuber, 2010] An important corollary of

> the free-energy principle is that it is vital for the brain to encode predictive models of the world in an efficient form, i.e. in concise rules of broad explanatory power. This means that the brain has evolved to be exceptionally good at finding patterns. Finding a pattern (coding progress) generates reward, and proceeds via reinforcement learning. The brain actively seeks out high information content in order to improve the adaptability of its world model. Encountering unexpected events in the world can cause anxiety if the agent is unable to deal with it, but at the right level of complexity (the mid point of the Wundt curve), the new information can be encoded and understood, and this is rewarding. As well as seeking novelty in external events, the brain also seeks novelty internally, for the same reason: the vast wealth of memories the mind possesses may still be able to be encoded in a more efficient form, or be used to simulate hypothetical events. Creative works could be seen as attempts by agents to encode their complex experiences in a compressed format, for the benefit of themselves and other members of society.

With regards to Human-Computer interaction, there seems to be huge, and largely untapped potential in applying these state-of-the-art creativity models to the design of creativity support systems. There is also great potential for researchers intimately connected with creative technological practice (such as NIME designers and users) to 'join the dots' between their findings and these models. Computational formalisations of the creative mind may provide a gateway for designers and researchers to start analysing their interactive systems more quantitatively with respect to their ultimate purpose: which is to generate novel and valuable artefacts, and enable users and audiences to achieve the peak experiences that creativity offers.

CHAPTER 4

Interfaces for Musical Expression

In this section we survey the relevant literature concerning musical interaction and Digital Musical Instruments (DMIs).

A DMI [Wanderley, 2001; Miranda and Wanderley, 2006] is a synthesis engine controlled by a gestural controller. The design of the control device and the sound synthesis engine are both clearly important for the playability and sound quality of a DMI, but also important is the mapping between them, this will be discussed in Section 4.3. A DAW (Digital Audio Workstation) on the other hand, is a piece of software geared towards the construction of finished musical works. Most are software equivalents of the analogue multi-track recording studio, and inherit the abstractions, metaphors and workflow of the studio [Duignan et al., 2010]. DAWs are best suited to complex non real-time editing and arrangement, whereas DMIs are usually designed with solo, expressive performance in mind.

In this review I usually refer to DMIs, but with the view that DAW-style composition technology may one day benefit from similar support for faster, more creative interaction modes. Certainly for many musicians, the production process features rapid switching between instruments and editing tools, and in many cases the distinction blurs[1].

In this review we tend to focus more on the background and frameworks behind design of music making tools, rather than looking at specific devices, systems, or instruments. For a comprehensive review see Miranda and Wanderley [2006].

4.1 Digital Musical Instruments

Until relatively recently, the means to control musical sound was fundamentally bound to the physical mechanisms of sound production. Musicians necessarily learned to accommodate their instrument, developing their motor skills through years of practice to enable them to play complex music. The advent of analogue electronic devices increased the amount of abstraction available for musical control. Any voltage control signal could now be routed so as to modulate a wide variety of synthesis parameters and signal processors; this also brought the possibility of automating musical sequences. For various reasons—principally the influence of the first widely available consumer synthesisers such as the Minimoog [Pinch et al., 2009]—the most common arrangement for synthesiser interfaces has been a piano style keyboard to control pitch, accompanied by rotary potentiometers to control timbre. In the last few decades the power of digital processing units and the variety of control devices has vastly expanded, and with it (as a combinatorial explosion of controller-synthesis parameter connections) the number of possible instruments. However, it seems few specific controller-synthesis pairings appear to have achieved

[1] There do exist hybrids that sit between these two extremes, Native Instruments' Maschine and Ableton Push, for example, are both physical controllers that enable real-time improvisation and a degree of instrumental control, but act as physical interfaces for DAW-type software.

wider acceptance, or attracted communities of virtuoso performers. This has been called the 'problem of the second performer' [McPherson and Kim, 2012]. A number of factors may be behind this:

1. Performers like to customise and create their own instruments, but there are so many possible mappings to explore, it is tempting to choose exploration of a new mapping over development of skill with an existing one.

2. Without visible proof of other performers having attained virtuosity (role models), 'investment of play' [Cannon and Favilla, 2012] in a particular instrument can be perceived as risky. There is simply no way to know how high the ceiling on virtuosity is before having invested large amounts of time in both practice and exploration.

3. Musical material that is too complex to play can be edited together off line, so motor skill is now no longer a prerequisite to composition.

4. Technology becomes obsolescent too fast for the slow development of social institutions, teaching structures, or the development of a musical canon. Obsolescence further adds to the risk of investment of play.

Traditional instruments are sometimes seen as having expressive properties which modern technology lacks. A survey of musicians by Thor Magnusson [2007], investigated contrasts between computer music software and acoustic instruments. The themes that emerged from these surveys are summed up in Table 4.1, describing the advantages and disadvantages of acoustic instruments, and Table 4.2 listing the same for digital software[2]. Why these properties are seen as positive and negative with regard to creative processes may seem obvious to musicians, but it will

[2] A number of these issues are marked as being incorporated with in the EARS model of Chapter 5 (where it is assumed that acoustic instruments fall into the 'skilled' quadrant, and digital fall into the 'analytic'), which aims to propose underlying cognitive explanations for some of these phenomena.

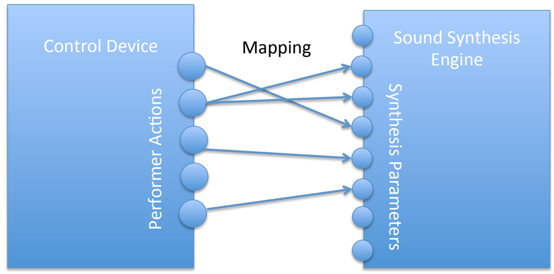


Figure 4.1: The mapping between a gestural controller and a synthesis engine. This is the standard definition of a DMI (digital musical instrument) [Wanderley et al., 2000].

be worth exploring these in more depth. For instance tactile feedback may seem clearly advantageous, but what is special about it that enables better performance? Exploration too seems obviously good for creativity, but how do we define it and how do we design for it? If we can quantify these concepts then we stand a better chance of discovering which aspects are unavoidable trade-offs, and which can be combined to construct systems with the advantages of both digital and traditional acoustic instruments[3].

4.1.1 Musical Expression

Research into expressive performance often focuses on the subtle nuance that an instrumental performer puts into their playing in order to convey, say, emotion. One working definition of expression is the difference between the mechanical playing of a score by a computer and a performance by a human involving variations in "tempo, sound level, timing, intonation, articulation, timbre, vibrato, tone attacks, tone decays and pauses" [Poepel, 2005]. An expressive performance adds humanity and

[3]One powerful way to achieve this is with augmented instruments [Newton and Marshall, 2011], which are DMIs based on incorporating technology into existing instruments. These leverage existing musical skills, but offer further control over novel transformations of their sounds.

Acoustic - positive		**Acoustic - negative**	
Tactile feedback	✓	Lacking in range	✓
Limitations inspiring	✓	No editing out of mistakes	✓
Traditions and legacy		No memory or intelligence	
Musician reaches depth		Prone to cliche playing	✓
Instrument becomes 2nd nature	✓	Too much tradition/history	
Each instrument is unique		No experimentation in design	
No latency		Inflexible no dialog	✓
Easier to express mood	✓	No microtonality or tunings	
Extrovert state when playing	✓	No inharmonic spectra	

Table 4.1: Advantages and disadvantages of acoustic instrument, taken from [Thor Magnusson, 2007]. Marked issues are discussed in Chapter 5, with regard to their contribution to the creative process.

Digital - positive		**Digital - negative**	
Free from musical traditions		Lacking in substance	
Experimental explorative	✓	No legacy or continuation	
Any sound and any interface		No haptic feedback	✓
Designed for specific needs		Lacking social conventions	
Freedom in mapping	✓	Latency frequently a problem	
Automation, intelligence		Disembodied experience	✓
Good for composing with	✓	Slave of the historical/acoustic	
Easier to get into	✓	Imitation of the acoustic	
Not as limited to tonal music		Introvert state when playing	✓

Table 4.2: Advantages and disadvantages of digital music software Thor Magnusson [2007].

brings the music to life. Of course there are performance variations that result naturally from the noisiness of the human nervous system, but more important are the variations that result from the musicians expressive *intent* [Palmer, 1996].

Studies of expressive performance may focus on particular aspects of expression, such as timing and tempo changes and how they relate to the structure of the score [Repp, 1998]. Alternatively they might focus on the ability of the musician to effectively convey a particular concept or emotion to the listener by means of these variations [Gabrielsson and Juslin, 1996]. In the latter case there appears to be some 'coding' process occurring, whereby a complex mental state can be encoded in the music by the performer and then decoded by listeners [Poepel, 2005]. The idea that the audience decodes the intention of the performer is an important one. One of the most frequent critiques of the computer music performance is that the mappings between intention, action and sound are not 'transparent' [Murray-Browne et al., 2011]. In extreme cases it may be impossible for the audience to see what the performer is doing, or even if they are performing at all, leading to suspicions of inauthenticity [Cascone, 2002; Keith, 2010].

Digital instrument designers and researchers are particularly concerned with expression, hence the umbrella term "New Instruments for Musical Expression" (NIME) for this research field [Poupyrev et al., 2001; Caramiaux et al., 2014]. Technology may encourage expression both in the way it provides access to sonic variation and nuance, but also in its physicality, stage presence and the bodily forms of the gestures that are required [Bergeron and Lopes, 2009]. The physical movements of the performer, both those that directly contribute to the sound and those that do not (e.g. body poses, exaggerated motions, running around the stage etc.) may also be important aspects of expressive performance. However, here we focus solely on control of sound.

Matthews investigated the idea of performing *only* these expressive dimensions of music using the Radio Baton [Mathews, 1991]. By moving the baton in space, and mapping it to the tempo and dynamics of a MIDI score playback, the performer (or perhaps conductor) is able to control only those aspects of the music deemed expressive in this traditional sense. Automation is sometimes seen as removing humanity and character from our lives, but this approach illustrates the potential for automation to relieve us of the difficult low-level mechanics of musical performance, and purely concentrate on expressive aspects. This highlights the potential for technology to provide what is known as *meta-control* of higher-level aspects of the performance [Brown and Sorensen, 2009; Van Nort and Wanderley, 2007; De Campo, 2014; Jordà et al., 2007; Sheridan, 2004]. However, there seems to be an awkward tension between the freedom of being relived of the drudgery of playing every individual note, and the alienation that results from losing the visceral, moment by moment connection of the mind to sound events via the body. Both approaches can be valuable in different artistic contexts [Gurevich and Cavan Fyans, 2011].

Arfib et al. [2005] state that nuance requires both flexibility and precision. Jordà et al. [2007] say "music performance outstandingly combines precision with freedom". In other words, musicians require ability to move freely through a rich parameter space, and also the ability to quickly and accurately locate points within that space. Arfib et. al. highlight that expression may extend over various timescales. At the note or sound-object level expressiveness may occur via vibrato or damping, whereas at the phrasing level, dynamic and rubato trajectories may be used to impart expressive structure. They also describe four methods of learning gestures: imitating gestures, performing gestures to copy specific sounds, interpreting a score or a gestural notation, and inventing new gestures.

Whilst dynamics and timing form the principal dimensions of expression (particularly for semi-mechanical instruments such as the piano), in general any musical parameter could count as expressive, including melodic improvisation or sound object selection. Many styles of electronic music raise challenges to the notion that timing and dynamics are essential expressive dimensions. For instance, much of electronic dance music (EDM) is quantised to a precise metrical grid, and often amplitude is 'brick wall' limited for maximum loudness. In fact, the "expressive" aspect is often to be found in the timbral changes that drive the musical structure. Also less important in EDM is the distinction between the composers score and the performers rendition of that score. Whilst there is a certain division between the composition (or "production") of a track and the live performance of it (e.g. playback by a DJ or live hardware performance), for many live electronic performances with an improvised component, the line between composition and performance blurs somewhat. These distinctions notwithstanding, the idea that meaning, intention or emotion is encoded into sound and then decoded by the audience is still very much applicable.

Relevant to this change of emphasis is the critique of the 'dominant model' of expression found in Gurevich and Treviño [2007]. They see the NIME community's model as being too firmly rooted in the western classical tradition of separating score from performance (text vs. act). They argue for a more ecological view, accounting for the diversity of aesthetic goals, and complex interrelation between performer and technology. Another very relevant aspect of this paper is the reference to Norman's three levels of processing in the human brain: the visceral, the behavioural and the reflective. These map onto performance of musical expression as:

- low-level nuance (visceral),
- Practised gesture and phrasing (behavioural),
- Large-scale form (reflective).

Norman's model fits well with the hierarchical structure of increasing time-scale predictors discussed in the last chapter, and also with the cognitive model utilised by [Malloch et al., 2006] [4]. Gurevich and Treviño [2007] note that many artists choose to avoid expression altogether, and focus on the reflective level (John Cage being a prominent example).

This thesis agrees with many of the points in this critique, with the exception of this final provocative statement:

[4] For the purposes of this thesis we only use 2 levels. The visceral and behavioural are both lumped into 'implicit' processes, whereas 'reflective' becomes 'explicit', in line with the majority of the reviewed cognitive science literature. A three level model may be more useful in some circumstances however.

> As an ecological model of musical creation prohibits the isolation of musical interfaces from their artistic contexts, it is meaningless for the authors to make prescriptive statements regarding [future directions for the design of new instruments].

As mentioned in Section 1.2.3, the current work takes quite the opposite attitude. If one cannot isolate NIME's from their artistic contexts, then this would seem to undermine the entire existence of the field. Conducting and disseminating research findings that do not generalise seems an unsatisfactory endeavour. In this thesis therefore, I attempt to both generalise and distil the notion of expression somewhat. Whilst many consider expression to be "a concept that is unquantifiable and dynamically subjective" [Malloch et al., 2006], I will consider expression as a kind of coding: encoding high-level representations in the brain into audio data on the computer. The artists nervous system, body, DMI interface and synthesis software form a communications channel, as Wessel and Wright [2002] put it:

> "Our human performer has intentions to produce a certain musical result. These intentions are communicated to the bodys sensorimotor system (motor program). Parameters are sensed from the body at the gestural interface. These parameters are then passed to controller software that conditions, tracks, and maps them to the algorithms that generate the musical material."

Throughput, or bandwidth, of interaction has been mentioned a number of times in the music interaction literature [Pennycook, 1985; Jordà et al., 2007; Pachet, 2012], but seems never to have been actually measured for any musical task. There does seem a general reluctance to equate controllability to expressivity [Dobrian and Koppelman, 2006] but the exact difference, and what is left when controllability is eliminated from expression seems unclear. If it relates to what is being expressed, then surely this is the artist's responsibility. If it relates to the finer sonic qualities of the output, then this is the synthesis engine's responsibility, something not dealt with in this work. A final possibility is that it relates to the effectiveness of the embodied metaphor [Antle et al., 2009] used in the gesture: how well the 'perceptual structure' of the gesture maps onto the qualities of the sound. Metaphor is certainly a powerful tool for aiding the musician's learning of the instrument, and also for the audiences perception of the connection between gesture and sound [Fels et al., 2002]. However, metaphorical mappings run into a problem in parameter spaces large enough to support exploration. It would seem immensely challenging to construct a mapping with multiple (and mutually consistent) gestural metaphors that apply to the whole space, therefore mapping via metaphor runs the risk of producing one-off use instruments that do not provide the longevity of continual discovery.

4.2 Instrument Design and Evaluation Frameworks

In this section we discuss proposals for frameworks that assist in designing and evaluating music technology, DMIs and DAWs. Frameworks provide a a dual purpose in design: to both inform the design process via a set of principles/guidelines, and also provide metrics by which to measure the success of the result. Therefore the design and evaluation of DMIs is often an circular, iterative process.

Three desirable DMI properties are a 'low threshold', a 'high ceiling' and 'wide walls'. These are criteria borrowed from HCI [Resnick et al., 2005]. The first refers to a low barrier to entry for novices, the second an unbounded potential complexity/virtuosity for experts. The last refers to having a wide variety of possibilities to explore. Virtuosity is difficult to test, because it takes such a long time for users to develop. Longitudinal studies such as [Cannon and Favilla, 2012; Nash and

Blackwell, 2011; Zappi and McPherson, 2014a] attempt to address this problem. Given the potential complexity and variety in creative outcomes over these long time-scales, quantitative comparisons become challenging.

Jordà [2004] looks at these three criteria, and synthesises them into a heuristic formula for the "efficiency" of a DMI:

$$\text{Musical Instument Efficiency} = \frac{\text{Musical Input Complexity}}{\text{Musical Output Complexity}} \times \text{Performer freedom}$$

Efficiency relates to the ability to produce as complex and expressive sound as possible with minimum input effort. However there must be an extra 'freedom' term introduced to specify the variability, or liveness of the output (otherwise the most efficient approach is to just press play on a recording!). Jordà states "Good musical instruments must strike the right balance between challenge, frustration and boredom.", harking back to Flow theory (Section 3.7). Complexity is a tricky thing to define, however we shall reconsider this efficiency formula in terms of information throughput in Section 5.4.2.

Mooney [Mooney, 2011] presents a model in which the 'frameworks' — both the physical instruments and the conceptual tools of music making—are viewed in terms of what they allow us to do, i.e. their affordances:

> The frameworks used have an influence upon the musical results. This cannot be over-emphasized. If a composer chooses to write for violin, for example, (s)he is buying into a certain set of affordances and therefore the musical results will be infused with violin-ness. If the composer chooses to use the five line staff framework to notate the composition, then the musical results will be constrained to those attainable using that system of representation.

Mooney goes on to look at the frameworks and affordances of music technology in more detail. The fact that sequencers start up with a metrical grid — often in 4/4 time and with a default tempo of 120bpm — immediately means it is easier to create music within this time signature than any other. Also considered are the affordances of the fader: its one dimensionality, the necessity to travel through intervening positions to reach another setting, and its limited extent. This makes a very important point: despite the entire parameter space being accessible in theory, some regions are more accessible than others in practice. No matter how logically and innocuously the parameters seem to be presented, the affordances of the interface necessarily render some regions of the space more probable than others[5], and encourage certain methods of navigation. Thus, affordances set up certain probabilistic tendencies toward action in the implicit system of the perceiver. Interface design may subtly, and perhaps subliminally, guide the thinking of the artist along certain channels.

Malloch et al. [2006] uses the human information processing model of Rassmussen [1986] to propose three levels of control: model, rule and skill. These control methods operate on symbols, signs and signals respectively. These refer to different levels of abstraction. Skilful interaction will manipulate the audio signal itself at a low-level and in real-time. Rule based interaction "consists of the selection and execution of stored procedures in response to cues extracted from the system", and model based refers to reference to the artist's internal model of a conceptual goal. This model then generates rule and skill based actions. This framework is rather similar to that developed in Chapter 5, but the EARS model extends this by introducing a divergent thinking component.

[5]Taking control of the probability distribution over the parameter space is one theme of this thesis, in particular Experiment 1.

4.2.1 Evaluation of Musical Interfaces

How should a musical interface be evaluated? Some find that mappings form part of the composition [Doornbusch, 2002], in which case it is an art, and the instrument should be judged using aesthetic criteria. Others claim it as part of the synthesiser [Arfib et al., 2002b], which would imply that it is engineering, suggesting that quantitative measures of effectiveness should be used. Jorda [2005] maintains that it is both. The first challenge to the music technologist is to be very aware that engineering is a field adapted to solving well-structured problems, but to apply these techniques to art—a world of ill-structured problems—is to run the risk of applying an inappropriate methodology. Should we measure the efficiency of performing various pre-specified musical tasks and run the risk of missing the real point, which is the performing of as-yet-unspecified tasks? Or should the technology be judged as if it were a work of art itself, and then run the risk of losing objectivity and generalisability?

It has been noted before that rigorous evaluation is rare in the NIME literature[Poepel, 2005; Barbosa et al., 2015]. [Stowell et al., 2009] look at both quantitative and qualitative methods for evaluating computer music systems, and provide a useful breakdown of which methodology is applicable where. One recent approach to evaluating DMIs is a dimensional approach Cannon and Favilla [2012]. They propose 8 dimensions by which to assess instruments. The overall expressivity can then be given by looking at how well all of these dimensions are rated. This has the advantage that performers can assess many different aspects of the instrument and provide a more detailed view of its characteristics, but has the disadvantage of many of the dimensions being subjective judgements.

Most relevant for the current work, Wanderley and Orio [2002] suggest borrowing input device evaluation techniques from the HCI community. In this approach various tasks are set for users, for example the performance of a specific musical melody, and objective numerical results such as speed and accuracy are measured. In this way devices can be compared against one another in performing the same task. They also discuss the notion of "Usability", breaking this into four desirable features: learnability, explorability, feature controllability, and timing controllability. The last is somewhat different in DMI research as opposed to normal HCI research: rather than judging a device on just how quickly a task can be accomplished it is also essential that input events possess high timing accuracy. One important suggestion in this paper is using the quantitative laws such as Fitts' law for interface evaluation. However it seems that no one has attempted this, perhaps due to lack of a suitable extension to high dimensional control spaces. The current work heeds the following advice:

> "From HCI research, it appears that musical tasks should in general strive for maximum simplicity. Even though it may seem entirely non-musical, the use of a few simple tasks may help..."

By reducing musical interaction to its most basic components, clearer results can be obtained. This approach may not always 'scale up' to real-world, long-term creative use however [Gelineck and Serafin, 2012].

One of the most basic evaluation methodologies is that of target finding, or sound imitation. In this approach, a sound is provided by the experimenter, and the user has to imitate that sound with a variety of different control devices. This is the approach taken by [Hunt and Kirk, 1999], where sounds produced by various controller trajectories were imitated using various different control devices. Target matching was also the approach taken by [Vertegaal and Eaglestone, 1996], where a target timbre had to be found in a 4D space using various input devices, including a 2-dimensional controller (a mouse) and a 4-dimensional controller: the Nintendo

Power Glove. This study is the closest antecedent to the target matching experiments in chapters 7 and 8 of this thesis, in that it looked at speed, accuracy and control integration (simultaneous alteration of multiple dimensions) when locating a desired timbre. It was found that the mouse performed better for this search task, but this was attributed to the unreliability of the Power Glove, rather than inherent difficulty in the multidimensional tasks. We aim to revisit this type of study by using more reliable, state-of-the-art controllers and applying a Fitts' law based methodology to the analysis of the results.

4.3 Representation and Mapping of Musical Parameters

Much research has been carried out into reducing the difficulty of navigating large timbre parameter space. Indeed, almost all music technology must address this problem in one way or another, as it is impossibly laborious to specify by hand an entire musical signal [Smith, 1991].

The mapping of physical controllers to sound synthesis parameters has been an active research topic for at least twenty years [Winkler, 1995; Wanderley and Depalle, 2004]. The question addressed is, what aspects of physical movement should be detected (control parameters), and how should those control signals affect the parameters of some sound synthesis engine (synthesis engine parameters)? Mapping has a significant effect not only on what sounds are easy or difficult to create, but also the subjective experience of the user: the 'feel' of the interface.

The main focus of mapping strategies has been in the realm of real-time control of DMIs. A good review of mapping techniques can be found in [Wanderley and Depalle, 2004]. Mapping is a topic amenable to mathematical and geometrical analysis. The geometry of the mappings, as in how one space is embedded in the other, and the smoothness of the resulting hypersurface, is seen as extremely important for the playability and the identity of the resulting instrument [Van Nort et al., 2004]. Considered mathematically, the control space is a manifold embedded in the parameter space. An interesting treatment of this concept, describing control of vowel formant synthesis in a virtual environment can be found in [Choi, 2000].

The principal distinctions between types of mappings are as follows [Hunt et al., 2000].

- One-to-many: one control dimension is mapped to many synthesis parameters.
- Many-to-one: many control parameters affect one synth parameter.
- Many-to-many: a combination of the above (also known as 'complex' mappings).

[Hunt and Kirk, 2000] complex many-to-many mappings appear to be more effective for expressive performance, and may lead to greater performance improvements with practice. This may seem counter-intuitive, as a complex mapping would appear to be less understandable or predictable than the alternative.

> "The sliders interface, whilst it physically allowed people to control multiple parameters, forced the user to mentally 'strip apart' the control task into separate control streams. This caused a form of cognitive overload which users generally found restricting and frustrating."

[Hunt and Wanderley, 2002]

This particular claim is more thoroughly investigated in this thesis, both theoretically and empirically. Hunt's work is worth revisiting for several reasons:

1. More recent work in cognitive science and cognitive neuroscience may shed more light on the underlying cognitive processes that explain this finding. For instance, Hunt speculates that it arises from a distinction between left and right brained thinking, which may not be the case.

2. The ratings of how well subjects managed to match the target sounds relied on subjective judgements by the experimenter. It would be useful to develop more objective ways of measuring target matching ability.

3. The HCI-based methodology (of comparing multiple interface types being used for the same task) revealed some important and counterintuitive results. However this methodology is very rarely used in NIME evaluations [Stowell et al., 2009]. It seems more work could be done in this area.

Hunt also found that movements requiring greater energy should map onto sounds with greater perceived energy, reflecting intrinsically embodied nature of sound perception. He also suggested that preventing the explicit system from concentrating on individual dimensions frees up explicit resources to work on other things. In later sections we propose what these other things might be, and why they are so important for creative interaction.

Other properties that have been noted as desirable for controller mappings are (after [Van Nort et al., 2004]):

1. *Low dimensionality:* Control devices often have fewer parameters than synthesis engines. Given the brain's limited conscious multi-tasking abilities and working memory capacity, simple controllers are preferable. This has led to a variety of dimensionality-reduction algorithms being used for this purpose (see Chapter 6).

2. *Locality, or distance preservation:* Having travelled a certain distance in control space, we want that to be reflected in the distance travelled in parameter space, and ideally perceptual distance too.

3. *Revisitability:* If we return to the same point, we wish it to sound the same. The location of preset points should be stable.

4. *Continuity:* If a point is adjacent to another point on the low-dimensional surface, they should be adjacent in the high-dimensional space.

5. *Smoothness:* Continuous higher derivatives are desirable to eliminate sudden changes in direction, this has relevance to the predictability of a control.

6. *Linearity:* When a gesture, such as a scroll, occurs it will have a certain effect on that sound, more extreme versions of this gesture should produce more of the same effect. This property is hard to achieve with any dimensionality reduction method, however smoothness implies some linearity in the immediate neighbourhood.

It should also be noted that manually constructing these mappings is a demanding process in itself. "Programmability is a curse" Cook [2001] and "There are simply way too many combinations of features and parameters to manually think about trying too many decisions to make and too many combinations that are useless. Its a process that invariably takes way too much time" [Fiebrink et al., 2010]. If the parameter space is large, the space of possible mappings is magnitudes larger. The interfaces usually provided that map controllers to synthesis parameters are themselves not considered 'musician-friendly'.

Machine learning seems to offer great potential in this regard. Interesting work is being carried out where the musician can perform a control gesture along to a parameter trajectory, and a mapping between the two is then learned automatically [Fiebrink et al., 2009; Françoise, 2013]. This approach greatly simplifies the programming of mappings[6].

Another important distinction is dynamic vs. static mappings [Arfib et al., 2005]. A static mapping is one for which visiting a certain point in contol-space always maps directly to a point in parameter-space. A dynamic mapping may rely on temporal variations in the control movements to determine the location in parameter-space. For example, the parameter space position could be attached to the control-space position via a model of a spring, and have momentum or friction, resulting in complex and emergent dynamic behaviour based on the time derivatives and history of the controller's movement. Arbitrarily complex spatio-temporal responsive behaviours can be established, for instance by the use of recurrent neural networks [Bown and Lexer, 2006; Kiefer, 2014]. For simplicity, the experiments in this thesis only investigate static mappings.

4.3.1 Preset Interpolation

The simplest, and most widely used way to make a parameter space quickly navigable is simply to save the coordinates of preferred points (these are referred to here as "presets"). Once a set of presets has been created, a low dimensional subspace can be created from them (e.g. the simplest being a line that interpolates between two preset points). The presets can then be 'morphed' by navigating the subspace using a gestural controller. Given a D dimensional controller, and a P dimensional parameter space, $D+1$ presets can be used to form a D dimensional subspace within $\mathbb{R}^P$. In-depth treatments of the geometry of these interpolation-based mappings can be found in Goudeseune [2002] and Van Nort et al. [2004]. Applications that

[6] Chapter 8 demonstrates that this idea can be inverted, such that the user can be shown a gesture (via an animated 3D representation of a hand) and then can imitate and learn it.

have implemented these ideas include SYTER [Allouis and Bernier, 1982], Bencina's metasurface [Bencina, 2005], and the "nodes" object in Max/MSP.

There are a number of issues with this approach. Firstly, the piecewise-linear P-space interpolation between two favoured sounds will not necessarily be pleasing, or conform to a linear route through perceptual space [Van Nort and Wanderley, 2007]. For instance, interpolating between two 'realistic' FM synthesised instrument sounds may produce decidedly unrealistic movements in the individual partials. Furthermore, many synthesis parameters are discrete switches and discontinuities may result. Nevertheless, preset interpolation is an intuitive and computationally simple way of creating useful subspaces in which to perform and improvise. It is perhaps surprising that more commercially available software and hardware does not implement preset interpolators.

4.3.2 Timbre space approaches

Researchers have tried to create dimensions that correspond to high-level perceptual descriptors of the character of the sound, using techniques such as multidimensional scaling to create a "timbre space" [Grey, 1977; Wessel, 1979; Arfib et al., 2002a]. This approach certainly fits many users expectations of an intuitive control space, but the nature of timbre is extremely hard to quantify [Pachet and Aucouturier, 2004]. Useful dimensions may vary widely between musical styles and different users. This timbre space approach has been used in zoomable interfaces, where large parameter spaces can be zoomed into to provide smaller variations. Examples include SoundExplorer by [Yee-King, 2011] uses an MFCC based timbre similarity metric and multidimensional scaling to create a 2-D zoomable timbre map, and the ISEE (Intuitive Sound Editing Environment) by [Vertegaal and Bonis, 1994], where zooming in to a region of timbre space would take the user further down a hierarchy of instrument categories.

Most techniques that reduce dimensionality in order to provide lower numbers of control parameters throw away a large proportion of the space. Even with the best subspace finding algorithm, many interesting settings will become inaccessible. In Chapter 6 an interface is proposed that reduces dimensionality whilst maintaining access to all possibilities.

4.4 Cognitive Approaches, Flow and Liveness

Many musical interaction researchers have highlighted the importance of the perception-action loop as a basis for cognitive models of musical interaction [Armstrong, 2006; Leman, 2008; Jones et al., 2012]. That the loop feeds back on itself is vital for any description of musical interaction, as unexpected sonic results may influence the performers subsequent gestural input [Wessel and Wright, 2002].

A further important aspect of this loop is the speed of feedback, otherwise referred to as the 'liveness'. An important influence on this thesis is Nash's study of computer sequencer and tracker use [Nash, 2012]. Nash finds a correlation between Flow (Section 3.7) and liveness [Nash and Blackwell, 2012]. Liveness is associated with the delay in receiving feedback about ones alterations of musical parameters; Nash found that the faster the feedback the greater the chance of experiencing Flow. Scientifically investigating Flow experiences is difficult, as they are somewhat rare, spontaneous, and may not occur at all under controlled test conditions involving simple tasks. Nash overcomes this by conducting a longitudinal study of real-world use of music tracker software 'in the wild'; logging all interaction with a musical sequencer over a period of 2 years. This thesis also contains an useful review of creativity literature. The importance of the speed of interaction is backed up by many other researchers [Wanderley and Orio, 2002], particularly the tactile and haptic aspects that provide fast feedback. The experiments in this thesis do not investigate haptic controllers, but note that the haptic channel is obviously one of the fastest and most effective ways for the body to recieve feedback about its actions. For a good studies of haptic controls see Marshall [2009], Wanderley et al. [2000], and Bongers [2000].

Relatively little work has been carried out on how cognitive load might affect musical interaction. However one important result was that interface complexity could have a negative effect on critical listening skills [Mycroft et al., 2013]. The necessity to drag more windows during a mixing task inhibited people's ability to detect changes in the volume of particular tracks. Maes et al. [2015] found that the accuracy of continuous cello bowing gestures were less affected by a dual task than discrete gestures, and that faster movements were less disrupted than slow movements. This indicates that fast rhythmic interaction may involve different timing mechanisms[7], and hence result in less cognitive load than discrete, slow controlled gestures.

Embodied cognition has been very influential in recent studies of musical interactions. One comprehensive discussion of this is found in Leman [2008]. Music performance is clearly an embodied activity, as one must move to make sound happen (with the exception of brain-computer interfaces [Moore, 2003]). However the relationship between embodiment and music runs deeper than that: even when just listening to music, structures such as metric levels may be parsed using real or imagined bodily movements [Toiviainen et al., 2010]. The body and the mind entrain to audible rhythm, presumably for some evolutionary adaptive purpose such as social bonding [Bispham, 2006]. Therefore embodiment is not just a necessary means to generate the sounds that make up music, it is also a way to listen, process and understand music. The power of embodied cognition is a justification to

[7]Probably originating in the motor cortex rather than fronto-parietal networks

develop ‘tangible’ musical interfaces [Jordà et al., 2007; Newton-Dunn et al., 2003]. However, as mentioned in Section 2.2.6, there has yet to be a study that measures any quantitative increase in effectiveness for an ‘embodied’ way to control synthesis parameters. There is the possible exception of Hunt and Kirk [1999], where it could be argued that complex mappings are more embodied, but it is not clear how.

4.5 Exploration, Appropriation, and Meta-control

Exploration and serendipity frequently emerge as desirable aspects of musical interaction in user surveys [Fiebrink et al., 2010; Doornbusch, 2002; Kiefer, 2010].

The notion that evolutionary processes can generate creative artefacts (see Section 3.2) has been investigated by many [Johnson, 1999; Yee-King, 2007]. [Dahlstedt, 2001] developed an ‘interactive evolution’ interface to explore large parameter spaces. This acknowledges that the user must perform the evaluation of the sound, but uses genetic algorithms to iteratively generate novel sounds based on mutations and offspring of previous user selected favourites.

The serendipitous results of interacting with complex content creation systems is often claimed as as primary motivation for using them [Fiebrink et al., 2010; Pease et al., 2013].

“Appropriation” could be seen as a special case of exploration. Musicians do not simply use their technology according to the intentions of the designers. On a trivial level, it is usually impossible for the designers of a flexible synthesis engine to investigate the entire parameter space of their creation. It is also impossible to foresee the extra processing that the synthesised signal may undergo, or to imagine the musical contexts it may be placed in. Therefore exploratory creativity is a given. On another level, it is impossible to imagine how unintended capabilities inherent in an artefact may be ‘appropriated’ by artists [McPherson and Zappi, 2015]. The guitar amplifier was not originally intended to be distorted via huge gains, the Roland TB-303 was never intended to have its timbre controls altered live as part of a performance [Barlindhaug, 2007]), and the turntable was never intended to be ‘scratched’, or to synchronise and mix two records [Smith, 2000]. Indeed, some the first electronically synthesised sounds of all were an appropriation of the optical encoding of film sound tracks: these were waveforms recorded onto the side of cellulose film [Levin, 2003].

What we will refer to as ‘reflective’ creative interaction — the ability to step outside of a constraining parameter space via some creative misuse — is therefore a vital part of the history of music technology.

Magnusson [2010] looks at DMI design as being the art of constructing constraints:

> Composing an instrument therefore implies some degree of affordance design, but the core activity typically involves the iterative process of experiencing and adopting the system’s constraints.

There is a basic fact that synthesis engines provide such a huge range of options that the instrument must attempt to constrain them in order to be playable. But there is also an acknowledgement of constraints as being good for creativity for more subtle reasons. Experiments have investigated the interesting question is how the complexity (or dimensionality) of an instrument affects the likelihood of appropriation and diversity of behaviour [Gurevich et al., 2012; Zappi and McPherson, 2014b]. These studies reveal that constraint (in this case low numbers of controllable parameters) encourages unusual uses. The studies implement a ‘minimal experimental paradigm’ by providing two groups of artists with instruments with one or two controls. The counter-intuitive result was that the box with a single control was perceived as having more possibilities that the box with two. The more minimal device also resulted in more diversity of behaviour. Section 5.5.4 suggests a possible explanation of this behaviour in terms of a cognitive model of creativity.

Musical structure is not 'flat', rather it is in general hierarchical. This hierarchy is very often represented in music software, and therefore imposed on the artist. In many cases the composer themselves may want to generate alternative conceptual hierarchy, and use it as a framework for creativity. Algorithmic and generative music involves the artist/programmer creating computer algorithms that can generate many of the low level details of the music themselves. In this case the performer can either play along with the emerging music, control high-level meta-parameters or simply let the music unfold autonomously [Collins, 2008]. Livecoding, on the other hand, is the generation of generative music in near-real time via a musical programming language [Collins et al., 2003; Brown and Sorensen, 2009]. These programming languages enable the artist to construct their own music abstractions, and hence free themselves of the affordances and constraints provided by commercial software.

Leman [2008, p. 54] describes the ability to reflect on previous creations and focus on fruitful or interesting zones as a 'ratchet' effect. By means of building a repertoire of tricks and meta-tricks and then automating them, the artist, and indeed culture as a whole, is able to continually build on past successes and create new complexity-generating interactions amongst existing cultural artefacts. This ratchet mechanism is present in musical creativity, and also scientific and technological progress: in that technology builds and recombines what comes before it, producing new complexities and emergent behaviours as it goes. By incorporating useful abstractions into our tools we can more efficiently navigate solution space. Alternatively, the tricks may be practised until they become automatic skills. This ability to reflect upon past creative behaviour and extract effective meta-strategies is an important component of the EARS model (the reflective quadrant).

4.6 Summary

Many of the consistent themes in music interaction research tie up with those in the previous chapters. The nature of "expressivity" is the subject of considerable debate, but it seems certain that the more mainstream HCI goals of responsiveness, speed and accuracy are important contributors to expressive range. However, for creative situations where foals are fluid, increased flexibility and exploratory ability is required.

The NIME research field has highlighted many areas in which work-oriented computer interfaces are lacking. Musical interaction design often throws up extremely radical new ways to interact with digital information. Studies of these new interaction methods have highlighted the following points:

1. The speed the speed and accuracy of input devices, and the speed of feedback on the effects of one's actions are both essential.
2. The geometry of the gesture to parameter mapping has a significant effect on the feel of the instrument, and an effect on the creative process.
3. Embodiment and tangibility adds richness to interactions.
4. Exploration is vital.
5. Constraints are often good for creativity.
6. The artist often wants to radically misuse, customise or transform the instrument they are working with.

Surveys of the electronic music community reveal a number of consensus opinions and recurrent themes. The existence of recurrent themes in qualitative research raises the possibility that there are underlying cognitive principles at work. In the next chapter we attempt to connect themes in NIME research with the cognitive principles discussed in the preceding two chapters. It is clear that more work could be done that more explicitly attempts to link theories of creative cognition with musical interaction research. There may be many aspects of the former that can be tested by experiments involving computer interfaces, and many issues in the latter that could be better explained with reference to cognitive science.

This may help to address the following three outstanding challenges in musical interaction research:

1. *The need for a clearer definition of the goal of DMI and DAW design.* Creativity is often a tacit goal underlying NIME research, but without a clarification of the processes involved, progress may be difficult. A grounding in creativity theories such as those in Section 3.9 can clarify the types of processes to be assisted/augmented.

2. The need for a methodology that distils the simplest set of empirical tools with which to investigate musical interaction: a *minimal experimental paradigm* with which to compare interface types for quantitative and qualitative differences.

3. The need for connections between objectively measurable quantities and subjective mental states. A good example of this is the connection between the speed of feedback and the experience of Flow made by [Nash, 2012].

The next chapter proposes a theoretical account of how this may be achieved.

CHAPTER 5

The EARS Theory of Creative Interaction

5.1 Introduction

A musician creatively engaged in music making with a computer forms an immensely complex system. There will never be any hard and fast rules governing this system, indeed if some could be established, some artist somewhere would immediately set about subverting them. Every instrument is different, every musician is different. Can we hope for a theory of creative interaction design that applies generally? Or must we always look at artistic interactions individually, and accumulate a body of knowledge that is simply a mass of disparate subjective opinions? This chapter aims to set out a theory of creative interaction. It aims to begin the process of building a bridge between two seemingly incompatible worlds: the computational, information-theoretic and probabilistic models of cognition that were outlined in Chapter 2, and the electronic musician's subjective experience of the production process.

First of all, it is helpful to establish what is to be gained by theorising, and what may be the dangers. Theory can bring benefits to many of the stakeholders in digital music. The designers of systems can benefit from a more structured design and evaluation methodology, and artists may benefit from greater insight into their own creative process and how technology may affect it. Researchers can gain from a more structured approach to looking at musical interaction, such as a more rigorous experimental method driven by testing of competing hypotheses.

Theorising may have its downsides however: it can excessively narrow the focus on testing only those phenomena that the theoretician assumes necessary to explain. This danger becomes particularly acute in rich and varied social contexts such as the creative arts; historically this has resulted in a widespread wariness regarding the reductionist "agenda"[1]. Another danger is the large resource overhead in evidence based hypothesis testing. If every aspect of a theoretical model needs to be carefully tested in highly controlled experimental settings, then the design and implementation of fully fledged music creation systems—ones that are actually complex enough to make music with—may get indefinitely postponed. Therefore there is often a trade-off between rigour and relevance [Fallman and Stolterman, 2010].

In an ideal world, a theoretical account of creative musical interaction would achieve the following:

1. Be based on more fundamental underlying *cognitive principles*.

2. Have *explanatory power*: explain, unite, and connect disparate observations.

3. Encompass a *wide range of creative behaviour*.

[1] E.g. "Contrary to an information technologist's reductionist perspective... creation is more than the mere movement and manipulation of bits and bytes of so-called 'information'." [Gouzouasis, 2005]

4. Be *simple*, concise, and elegant, and communicable to both designers and artists.

5. Generate *design recommendations* for future technologies.

6. Enable *quantitative comparisons* between designs.

7. Make *testable, falsifiable hypotheses* for future experiments.

This chapter makes an attempt to address these concerns. Whilst it is unlikely to be definitive, it aims to be the most comprehensive attempt to date to put musical parameter control on a more quantitative basis.

In Section 5.3, a method is outlined for quantitatively evaluating the effectiveness of multidimensional controllers (in Chapters 7 and 8 this methodology is used to evaluate the control of timbre in sound design and performance scenarios). Next we look at the properties of creative strategies themselves, and relate them to the geometry of parameter mappings. By building on the cognitive, creative, and HCI principles discussed in Chapters 2 and 3, a four-quadrant model of creative interaction is developed. This describes four cognitive strategies: Exploratory, Algorithmic, Reflective, and Skilled (EARS).

Much of the EARS theory is intended to apply to a wide range of creative interaction behaviour (not necessarily musical). These four modes could be observed in any Human-Computer hybrid creative systems, and be applicable to to thought, movement, interaction or data manipulation. Nevertheless, a considerable number of simplifying assumptions and simplifications are made for the purpose of establishing a 'minimal paradigm' for investigating creative interaction. These are as follows:

1. Assume a single solo artist interacting with a single piece of technology.

2. Assume that the evaluation of the creative artefact is carried out by the musician whilst engaged with the interface.

3. Assume a finite, continuous parameter space, for example some audio synthesis parameters that can be manipulated via a controller in real time.

4. Assume these parameters correspond roughly to musical, perceptual attributes of the sound e.g. musical time, pitch, amplitude, decay time, filter frequency and so on. Distances between synthesis parameter settings are assumed to be roughly equal to that of the perceptual differences of the sound itself[2].

These assumptions obviously apply to some musical practices better than others. In particular, the electronic dance music production process is an example of this kind of human-computer creative system. The respondents for Appendix B's survey were mainly drawn from this community.

If a theory does not simplify matters, it is not doing its job. Although inspired by cognitive research, and capable of being deepened and formalised further, the concepts EARS introduces need not be described in overly complex or mathematical language, and do not necessitate detailed knowledge of brain anatomy or low-level computation to understand. Hence the EARS model should should be conveyable to moderately technologically competent artists and software designers. The ingredients of EARS are certainly not entirely novel, many of the ideas it deals with will be very familiar to computer music researchers (see, for example, Jones et al. [2012]). However, it is intended to provide these ideas in a more condensed form: hopefully drawing somewhat tighter connections between brain processes, feedback, prediction, movement, unconscious skills, information and technology.

The theory generates several hypotheses. By no means all of these hypotheses have been addressed in the experimental work, therefore the tested and untested predictions will be explicitly stated in Section 5.6.

[2]This is probably the most questionable assumption. For the simple synthesiser in Experiment 2 it holds well enough, but one might doubt that this assumption would scale up to entire musical projects.

5.2 The Human-Computer Hybrid Creative System

In this section, a model of the artist interacting with content creation technology is developed. This model establishes a framework within which to analyse the information flow in a creative perception-action loop. The human-computer hybrid [Burleson, 2005; Jones et al., 2012] is depicted as a distributed creative system in its own right. Various aspects of the creative process can occur at different points in the system, therefore novel information flows around the loop in various ways, not just from the artist to the machine. I propose that the traversal through the parameter space of a piece of software or hardware is a 'mirror' of the traversal through conceptual space occurring within the brain of the artist, and vice-versa. How fast, and how faithfully the movements in the two spaces reflect each other is an important determinant of how effective the creative system will be, and the interface design is crucial in determining the speed and accuracy of this mirroring. Designing an interface for a DMI or DAW establishes the probabilistic geometry of the solution space navigation strategies. It is argued that the traversal mechanisms provided by the interface should mirror those strategies the brain uses to be creative.

5.2.1 Linking Conceptual Space to Parameter Space via Cognitive Mirroring

At the end of Chapter 2, an underlying principle of human-computer interaction was suggested, that of "cognitive mirroring". This principle is based around the idea that the human and the computer are partners on an expedition through solution space. The goal of the expedition is to discover and ascend peaks of high value in the fitness landscape. Like mountain climbers, they need to stick together. Without the technology, the actual audio data of electronic music could not be produced. Without the human, the data could not be evaluated or imbued with cultural meaning. The two explorers have very different skills: the computer has a immense and precise memory, and can process large amounts of raw data. The human, on the other hand, is rather good at seeing the way ahead and guessing routes through challenging terrain, and is the only party who can evaluate the height (value) of the point in the landscape. The human cannot actually see the territory in detail unless the computer renders it in concrete form. Loosely speaking, the faster and more faithfully the computer can be told where the human is trying to get to, the better a climbing companion it will be. Likewise, the more descriptive the computer's 'display' (auditory, visual or haptic) can be about the current state of the data, the better. The role of the interface is to yoke the two explorers together such that they can communicate, collaborate, and utilise their differing skills to their best effect.

It is worth discussing the geography of conceptual space. The space of possible pieces of electronic music is, of course, combinatorially vast, and high-dimensional[3]. The space is also rather foggy, in that it suffers from low "sightedness" [Simonton, 2012]. The human may be able to judge their immediate surroundings and get a sense of the gradient of the terrain, but predicting what will happen beyond that requires either knowledge from a previous expedition, or some tricks of the trade such as musical (geographical?) expertise. In this space, there are vastly more boring, random pieces of noise than good music: most of the landscape is flat 'desert'. Therefore the agent might need to learn or invent some techniques to limit the space to well structured pieces. Even then, the artists who have gone before are likely to have discovered many of the obvious regions of interest already. The most accessible peaks will have been claimed. If there exists an unconquered mountain there must be some reason why: it is probably the other side of a huge desert,

[3]Our spatial intuitions can be very wrong in high dimensional spaces [Aggarwal et al., 2001].

CSF		Electronic Musical Interactive Systems	
$\mathscr{U}$	Space of all possible concepts	$\mathscr{U}_{param}$	All possible electronic music pieces
$\mathscr{C}$	Space of all existing concepts	$\mathscr{C}_{param}$	All existing electronic music pieces
$\mathscr{T}$	Traversal mechanism	$\mathscr{T}_{param}$	Interface manipulations
$\mathscr{E}$	Evaluation function	$\mathscr{E}_{param}$	Currently unavailable

Table 5.1: Analogies between components of the Creative Systems Framework, and creative interaction with music technology. Without a way for the machine to evaluate its own progress, high bandwidth communication with the human is still essential.

beyond a deep chasm, where previous explorers have turned back or perished, or on another planet entirely. Clearly, simple strategies such as hill climbing or map following will not find novel, or particularly lofty peaks.

The CSF terminology (Section 3.4.1) is useful for asking what creativity might mean when navigating such a parameter space as that provided by a DAW or music synthesiser. The various components of the CSF constitute a powerful analogy for the various elements of the human-computer system. As the musician is interacting with the parameter space, and is constrained by it, it is ostensibly a space of possible compositions $\mathscr{U}_{param}$, and the interface provides $\mathscr{T}_{param}$: the mechanisms to navigate the space (for example, a knob provides the means to travel in a single dimension across the width of the space of possible compositions). $\mathscr{C}_{param}$ corresponds to the conceptual space of existing, non-novel compositions. Table 5.1 summarises these correspondences.

Obviously there are cultural and emotional associations that sounds may possess that are not represented in the very reduced domain of their digital representations. This means conceptual space possesses a far richer and more complex structure than parameter space. Parameters such as pitch, filter cut-off frequency, and amplitude envelopes only represent the lowest levels in the hierarchical conceptual space of music[4]; nevertheless, data structures such as MIDI and DAW project files are far

[4]For this work we assume that higher level aesthetic concepts are outside the domain of user interface design, they are still the responsibility of the human artist. By assuming that the aesthetic evaluation of the fitness of a given point in parameter space is carried out by the user, difficult

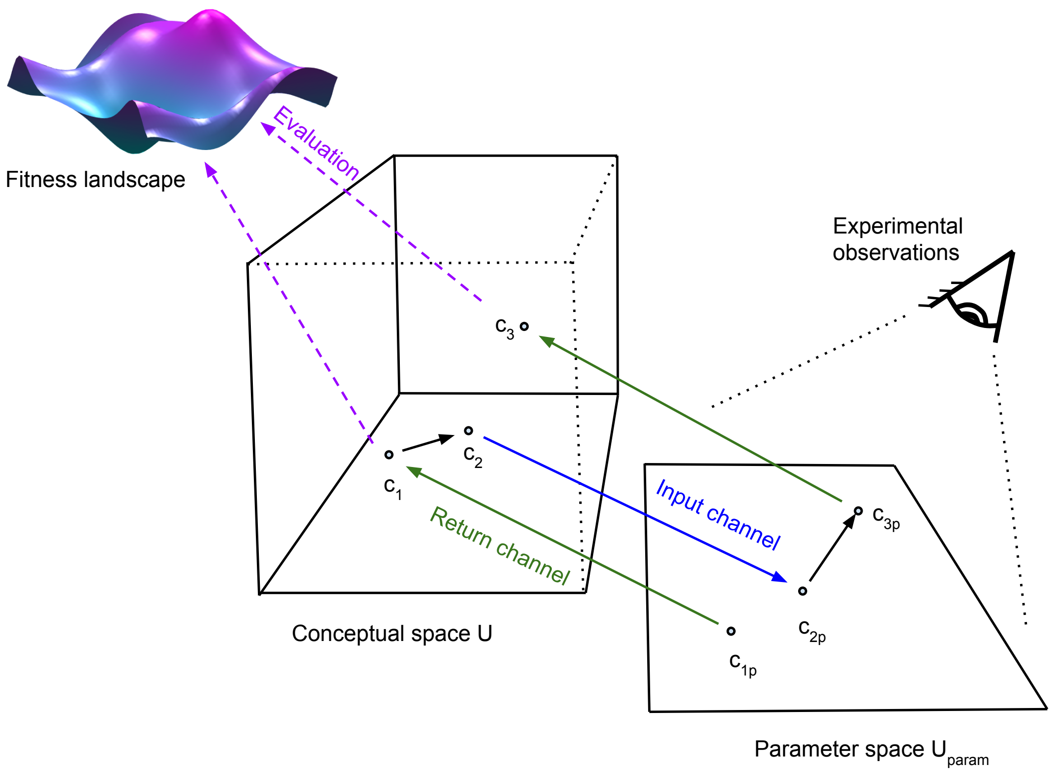


Figure 5.1: An illustration of the cognitive mirroring principle. Some initial parameter settings c_{1p} are perceived by the artist via the return channel (the audio 'waveform space', not shown), giving rise to the point in conceptual space c_1. The artist then evaluates c_1 to give $\mathscr{E}(c_1)$, some point on the fitness surface. The artist then varies the idea in order to increase its value, this gives c_2, which then needs to be input via the interface to give c_{2p}. At this point, instead of the idea occurring in the artists mind, the computer generates a variation c_{3p}, which is then listened to and evaluated. All these transfers of information take place via band-limited channels of varying degrees of lossiness — this lossiness will determine the amount of mutual information between the two spaces. An experimenter cannot observe $\mathscr{U}$ or $\mathscr{E}$, but they can observe what is happening in $\mathscr{U}_{param}$. If mutual information is high, creative behaviour in $\mathscr{U}_{param}$ may leave distinctive signatures in $\mathscr{U}_{param}$.

questions such as the cultural conceptual dimensions of particular musical sounds can be side-stepped. For now, we assume some complex fitness function is being optimised, without needing to know its exact form (though interesting work has been done both tracking users' paths through solution space and obtaining value ratings [Jennings et al., 2011]). This does not mean that the *navigation* of solution space is exclusively carried out within the brain, however. The constraints and affordances of the tools, notations and abstractions used for composition have a significant effect on the routes the artist takes through solution space, and thereby on the form of the finished from being the 'ground floor' of conceptual space, as they include a large amount of structure based on music-theoretical conceptual frameworks such as western tonal scales, metrical time signatures, conceptual units such as 'instruments' and 'effects', and so on[5]. Creating a piece of electronic music could be seen as a form of projection operation: projecting the high-level ideas in the artist's head down onto numerical representations of musical data. One possible depiction of the artistic endeavour is that another listener experiencing the creative artefact will be able to faithfully reconstruct the high-level aesthetic dimensions by the act of perception, analogous to how shining laser light on a 2D pattern can reproduce a 3D form holographically. Indeed, the artist may need to simulate their audience's de-projection process when evaluating the current state of their work.

We shall therefore assume there is some mapping, or coupling, between conceptual space, parameter space and waveform space. Given effective communication between these domains, movements in one will be reflected in movements in the other. Fig. 5.1 illustrates this exchange of information between the conceptual space within the artists brain, the evaluation function, and the parameter space of the technology. The mapping from physical gesture to synthesis parameters is a crucial link in this larger mapping process.

As a concrete example of how the technology may take on a role in the creative process, consider the two scenarios below:

1. The composer has a clear idea in mind, and will therefore need to optimise parameter settings such that the idea is realised.

2. The composer does not have anything specific in mind, and is looking to engage in an exploratory process that may produce a moment of inspiration.

product [Mooney, 2011; Magnusson, 2010]. This makes the traversal mechanisms $\mathscr{T}_{param}$ provided by the interface an essential topic of enquiry.

[5] Flexibly navigating the depth of the abstraction hierarchy is an interesting avenue of research [Duignan et al., 2010].

These two scenarios seem to map very well to notions of convergent and divergent thinking (see Section 5.5.1), however, in the latter case the divergent traversal of solutions is being carried out *within* the machine, *via* the interface. In scenario 1 the creative act has already occurred in the brain of the composer, and all that is necessary is an interface that enables the user to adjust parameters such that the data *converges* to match the idea. This can be seen in Fig. 5.1 in the steps from c_1 to c_{2p}: the move occurs in conceptual space, and then information flowing through the input channel produces a reflection in parameter space. The goal would be to find the point in $\mathscr{U}_{param}$ that, when rendered and perceived, best approximates c_2. Scenario 2 is just as important: the composer embarks on an interactive journey, and unpredictability is a key ingredient. The artist is using the data generation capabilities of the technology to *diverge* from their previous works. In Fig. 5.1 this process is seen in the route from c_{2p} to c_3. The move occurs in parameter space, via some exploratory interface manipulation, and is then reflected in conceptual space upon perception of the results. Therefore, it would appear that some divergent thought has been outsourced to the technology. This leaves the human as the evaluator rather than the generator of ideas. These technological flukes are analogous to aberrations in the CSF. A technological aberration is a movement in *parameter* space that takes one out of one's existing *conceptual* space. Thus, the interface may mirror exploratory creativity to a greater or lesser extent, depending on its provision of a traversal strategy that might produce these aberrations.

Perhaps more than any other creative domain, the electronic musician's conceptual space is closely tied to the parameter space of tools such as sequencers, synthesisers and effects. Many electronic musicians carry out their creative work whilst actually manipulating the interface. In addition, abstractions of musical data are handled within the machine, having a large effect on the mental representation of the material [Duignan et al., 2010]. This means $\mathscr{C}_{param}$ will possess high mutual information with $\mathscr{C}$. The tighter the feedback loop, the more cognition may be mirrored in the trajectories through $\mathscr{C}$. Study of these parameter space trajectories may reveal more about the trajectory through $\mathscr{C}$. By logging interactions with a certain set of controls, some clues as to the artist's navigation strategies may be obtained [Jennings et al., 2011].

The input and output communication channels between $\mathscr{C}$ and $\mathscr{C}_{param}$ are of vital importance. They determine how well the creative agent can realise their concepts. The bandwidth of the input channel will determine how fast changes in conceptual space (the musicians own ideas) can be implemented in the creative artefact. The bandwidth of the return channel will determine how fast the artist can hear the effect of those implementations, and hear the results of transformations carried out within the computer. Therefore the information-theoretic properties of the input and return channel are crucial areas of study. In the next section information flow in various subsections of this communications loop is described in detail, and methods are proposed to calculate and measure throughput experimentally.

5.3 Entropy and Information Flow in the Creative Perception-Action Loop

A blank canvas, a blank word processor document, or an empty DAW project may appear to be a highly ordered, low entropy piece of information. Appearances can be deceptive. At the start of the process, each part of the blank canvas has the potential to take on many different possible colour values. Due to this uncertainty, it possesses a much higher entropy than a finished work, where all degrees of freedom have been specified. By making many large and small decisions, the artist gradually reduces the entropy of the canvas until each area is specified to their satisfaction,

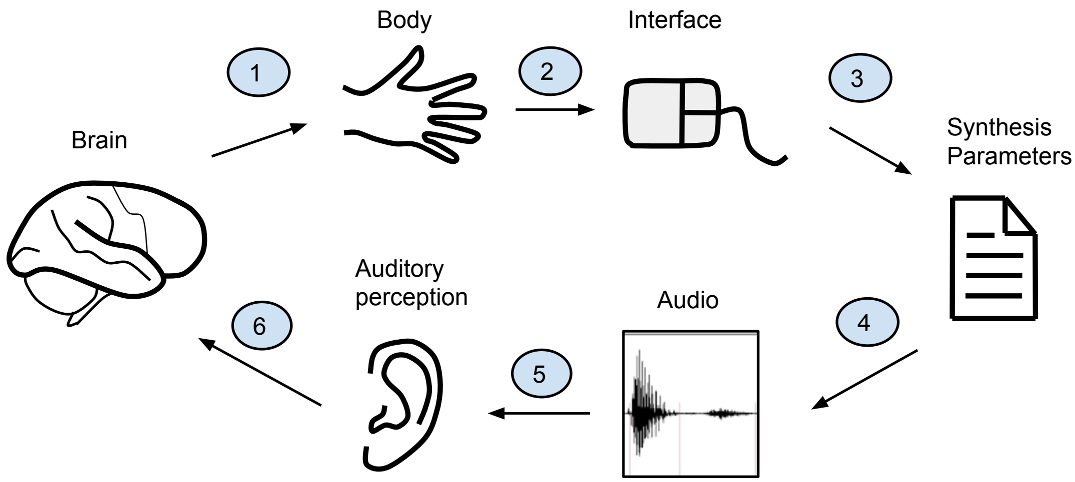


Figure 5.2: Information flow around the perception-action loop.

and the work is finished. The electronic musician, by altering the values of synthesis parameters of various events in time, traces a path through a space of possibilities, a trajectory that, after many hours reaches a target point: the finished piece. Whilst the decisions the artist makes may be unobserved, tell-tale traces of these decisions are left in the interaction data. By logging the search trajectory, and measuring how it progresses towards the target, it may be possible to infer such things as the rate of entropy reduction. This can tell us how much information has been transmitted from the artist to the artefact as a function of time.

Figure 5.2 illustrates how information flows from the artist's brain to the creative artefact (and back). Similarities should be noted with Leman's perception-reaction cycle [Leman, 2008, p. 54], Wessel and Wright's conceptual framework for controller research [Wessel and Wright, 2002], the 'extended composer' [Jones et al., 2012], Pressing's cybernetic perspective [Pressing, 1990], and Nash's feedback loops [Nash, 2012, p. 101][6]. In this loop, six points of information transfer are identified:

[6] Nash investigates in detail the relationship between the user, the notation and the music (audio), using a three node network with bi-directional feedback loops between all three.

1. *Intention to Gesture:* This is the transition between the user's goals and their bodily movements. The information loss depends on things such as the noise in the efferent motor nerve signals, amount of practice, and basic physical limitations.

2. *Action to Interface:* Movements of the body will translate, via sensors, to a data stream emerging from the input device. The information loss will depend on the accuracy of the sensor devices.

3. *Interface to Synthesis Parameters:* Control device parameters can be mapped to synthesis parameters in a variety of ways. This transfer point is the subject of DMI mapping research, and the focus of the three experimental studies in later chapters.

4. *Synthesis Parameters to Audio:* The synthesis engine will render audio on the basis of the parameter settings. In general, the amount of information expands many-fold here, as the synthesiser will produce complex waveforms (e.g. 44.1KHz 16-bit audio) on the basis of a smaller number of parameters [Smith, 1991]. The design of the synthesis engine is of course incredibly important, and will have knock-on effects throughout the loop, but is out of the scope of this work.

5. *Audio to Perceptual Dimensions:* The amount of musical information the artist consciously perceives is, in general, far smaller than that contained in the audio information. It is also highly dependent on what aspect of the sound is the current focus of attention, which may in turn depend on the artist's current sub-goals and associated interface operations.

6. *Perceptual Dimensions to Musical Evaluations:* Finally, the perception is evaluated in some way, and possibly compared to the artists original intent. The

artist may then produce a prediction as to how to increase the quality of the sound, form the intention to realise it via gestural interface manipulations, and the cycle continues.

Ideally, the interface exhibits an interaction bandwidth (Channel 2) at least equal to the rate at which the brain can control the motor system (Channel 1). This is far from the case for the majority of computer interfaces, and this results in a certain amount of frustration, and many claims that the computer is not expressive or sufficiently 'embodied'. On the other hand, many instruments (e.g. a piano) have a far greater potential bandwidth than the rate at which a novice can control their movements, which can also result in frustration.

To complicate matters, information can also cascade backwards around some parts of this loop. Mini-feedback paths may well have an effect on the overall throughput. For instance, if there is fast[7], informative haptic or visual feedback it may significantly speed up interaction [Cockburn and Brewster, 2005]. The user will gradually learn the peculiarities of the interface, therefore a representation of the mappings at points (2), (3) and (4) will gradually be formed in the mapping from intention to body movement (1).

As [Wessel and Wright, 2002] mention, in relation to the perception-action loop:

> "Admittedly this diagram is schematic and incomplete. One aspect that is not well captured by it is the way in which performers' intentions are elaborated upon by discovery of new possibilities afforded by the instrument."

If the majority of the content is emergent, and goals are fluid, then how do we measure effectiveness? Any attempt to measure information flow must define the

[7]Perhaps counter-intuitively, the fastest feedback of all is *no* feedback, i.e. 'open loop' control (see Section 2.2.5). This does require the user to have an accurate internal model of the exact movements required to achieve a certain result. Experiment 3 will reveal evidence to support this claim.

input and output, and what is signal and noise. This is not obvious. If the artist's goal is ill-defined, extremely abstract and high level, or the emerging audio is evaluated as actually being *better* than the artists original intent, then at any moment the noise could become the signal. In fact many electronic musicians estimate that this 'noise' generates the majority of their material (see Appendix B). The noise could be generated anywhere within this loop. Even data produced as a result of the artist's incompetence may occasionally be novel and valuable!

In view of the potential for the sudden redefinition of noise, it seems necessary to develop two evaluation methodologies to separately tackle 'divergent' and 'convergent' interaction modes. The effectiveness of convergent interaction can be measured by comparing the current data to a pre-specified target, and timing how fast the user can achieve a suitable match. This will be the approach taken in experiments 2 and 3, and is similar to the traditional approach taken in HCI research: in particular by use of Fitts' law for pointing tasks. Put more specifically, throughput can only be measured if we know (or have specified) the user's prior intent, and the user sticks to that intent for the duration of the interaction. Developing a similar model for divergent processes involves the acknowledgement of more fluid goal states. Next, a model will be sketched out of how an information-theoretic approach may go towards dealing with these situations, and how the creative loop may be a phenomenon that emerges from the Free-Energy principle.

5.3.1 Towards A Free-Energy Account of Fluid Artistic Goal Hierarchies

A probabilistic, free-energy-minimising view of the artists underlying intent leads to quite a different balance of information processing in the human-machine hybrid system. Let us say that the musician has an internal, probabilistic predictive model

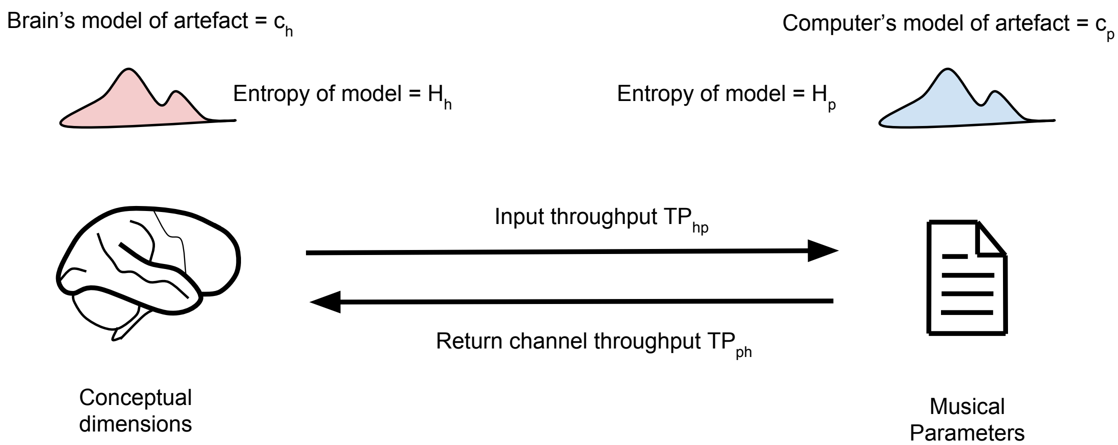


Figure 5.3: Notation for this section. The artist has some vague probabilistic model of their goal state c_h, with entropy H_h. The computer has the current specification of the data c_p, but there is a degree of uncertainty over whether this is the final artefact or not, and degrees of freedom remaining to be specified, giving rise to a hypothetical entropy H_p.

of what they desire to the creative artefact to be: the human's conceptual model, c_h. If c_h is extremely vague, then we could say the model has high entropy (H_h is large). The computer has a data representation of what the artefact currently is: the computer's parameter model, c_p. If data is yet to be specified, c_p too will possess many potential degrees of freedom and hence has high entropy: H_p is large[8].

Recall that Friston states there are two ways for an agent to minimise free-energy [Friston, 2010]. The first is to act so as to change the sense data to match the internal model. In this case the human has an 'idea' that is better specified (lower entropy) than anything yet present in the machine: $H_h < H_p$. The human then seeks to change the data on the machine such that it comes to resemble c_h. The human then needs to act, by manipulating the interface, to reduce H_p, the entropy of c_p. We shall see in later sections that the amount of *search space reduction* can

[8]Of course, the data in a digital computer is perforce precisely specified, however considered *relative* to the user's final goal state there are many parameters that exist in a yet-to-be-specified state.

be used to measure how fast the data converges on the desired result, and hence the throughput of the input channel $TP_{h \to p} = \frac{dH_p}{dt}$.

The second route to free-energy minimisation is to update the internal model to match the sensory environment. For instance, if some accidentally discovered aspect of c_p resonates with artist's higher level aesthetic goals more than the current c_h, then the artist updates their internal model of mid-level structure accordingly. In this case $H_h > H_p$, as some pattern in the data in the computer has been utilised to reduce H_h, information has flowed in the opposite direction: from computer to human, via the return channel. This process, more akin to discovery than invention, is a far quicker update process due to the higher bandwidth of the return channel.

This leads us to a number of interesting predictions. If the interface is indeed a highly restrictive bottleneck, as some electronic musicians feel, then one of the things we would expect to see is that the artist increasingly relies on the exploratory aspects of the interaction to generate material. If the random traversal of the parameter space tends to generate interesting material at a rate faster than competing ideas in the mind can be input via the interface, i.e. $TP_{p \to h} \gg TP_{h \to p}$, then the optimum strategy is increased reliance on the return channel. Put simply, there is no use having an excellent, intricately detailed idea if the chances of realising it are very small, or prohibitively laborious. Much better to explore the actual achievable options, and select that which meets your evaluation criteria. In the exploratory mode, the high bandwidth of the return channel is making up for the low bandwidth of the input channel. This may indeed be what we observe in electronic musicians behaviour. The ability to 'talent spot' parts of the parameter space with potential becomes more important than the ability to 'think up' musical ideas[9]. If we believe Schmidthuber that *rate of coding progress* is the generator of subjective reward in

[9]Indeed the most lauded 'musicians' in electronic dance music tend to be DJs, whose artistic expression is more or less exclusively related to their ability to select music, rather than to create it. Their status as artists is founded on their evaluation abilities.

artistic domains, then the sudden discovery of a valuable pattern in the return channel data will actually be more rewarding than a gradual and effortful reduction in entropy via the interface. The suddenness of this experience in some way resembles the suddenness of insight, and may induce a similar subjective 'aha!' experience.

A second prediction relates to when the return channel bandwidth decreases for some reason, as for instance when the musician is attempting to structure a whole piece, when quite small control changes may take minutes to evaluate rather than seconds. In this case, the exploratory mode becomes non-viable, and far more predictive effort is required from the musician. Serendipitous discovery of song structure will be rare. This may be one contributor to the observation that musicians feel that structuring a track is hard mental work (see Appendix B).

So how do we measure the effectiveness of an interface for the 'divergent' mode? Calculating $\frac{dH_h}{dt}$ does not seem possible: this would entail witnessing some model update in the artist's mind. The only ways to do this would be to

1. Instruct the artists to self-report moments of discovery. The user saves all the discovered points. The interface with the highest average number of discoveries in a given amount of time is deemed more effective.

2. Infer a discovery via some tell-tale signature in H_p.

The second approach must be done in light of the goals that emerged during the process. If we know the final state of the data, then throughput can be calculated retrospectively: post-hoc analysis may reveal information surges when a discovery was made, or plateaus when the musician was making no progress. This may have a different time series signature from more monotonic, convergent processes such as the linear throughput-time relation in Fitts' law. Seeing as discovery is a rewarding, and often sudden thing, these discrete events might be countable as both spikes in throughput plots, and subjective 'aha!' moments reported by the user, and these could be correlated. Alternatively, survey questions could be used, asking the user which interface they preferred for various correlates of divergent behaviour: properties such as being able to generate and evaluate lots of novel sounds, traversing the space quickly, or feelings of surprise. We will return to the question of how exactly an interface might provide an effective divergent traversal strategy in Section 5.5. In the mean time, we focus on the 'convergent' mode, and propose a way to measure throughput for multidimensional parameter space searches by looking at how the point in the parameter space c_p evolves over time.

5.4 Measuring Throughput in Synthesiser Parameter Space

Wanderley and Orio [2002] make recommendations are made for improving DMI research by borrowing tools from HCI. Fitts' law is mentioned as having potential, but has not yet been seriously investigated, perhaps due to lack of a easily applicable methodology for either multidimensional controllers or time-based rhythmic interaction. Information "throughput" has certainly been mentioned in relation to synthesiser interfaces [Pennycook, 1985] and musical virtuosity [Pachet, 2012], but has yet to be seriously investigated experimentally. Whilst there are many analogies between visual target pointing and sound target matching tasks, there are a number of differences and extra challenges with an auditory search. The following must be considered when finding an analogue of Fitts' law for sound target acquisition:

1. *Delayed Assessment.* Visual differences in position can be assessed extremely quickly. Sounds, however, take more time to listen to. Some control adjustments may have delayed effects, particularly time envelope controls. Differences between two sounds cannot be easily assessed with them playing simul-

taneously.

2. *Anisotropy.* Timbre space does not look the same in all directions. Pitch, timbre and temporal features are all very different perceptual qualities. In contrast, 3D space can be considered isotropic, although evidence for some anisotropy in pointing tasks has been found [Murata and Iwase, 2001].

3. *Low sightedness*: Differences in sounds are harder to judge, and take far more effort to process that differences in position. Parameter spaces could be said to vary between being “sighted”, where the distance and direction to the target is predictable, and “blind” where it is impossible to know which direction to move in, or how far away one is from the target. Sightedness will have an effect on the extent to which information about the target position can inform the movement. Sightedness depends on the user’s expertise with an instrument: how well do they know the parameter space?

4. Due to this low sightedness, it is hard to specify a target size, therefore movement accuracy needs to be inferred post hoc.

5. Timbre space can possess a dimensionality far higher than that of ordinary space.

This section is devoted to developing a method of measuring throughput that can deal with these considerations. This requires several new approaches with regards to the use of Fitts’ law in device evaluation. In fact, this technique is not music-specific, and may clear up a number of points of debate in the wider HCI field.

5.4.1 Sidestepping the Fitts' debate via Index of Search Space Reduction (ISSR)

In Section 2.4.1 we encountered several issues with the existing Fitts' law-based device evaluation methodology:

1. Uncertainty over which formula to use.
2. Uncertainty of how to calculate throughput from plots of MT vs. ID.
3. Uncertainty over the exact relationship of ID to Shannon information.
4. The necessity to devise new experiments and movement laws for increasing dimensionality.

The Index of Search Space Reduction (ISSR) approach aims to resolve (or at least circumvent) these issues.

The first novel aspect of this approach is a change of perspective on exactly what we are trying to achieve by using Fitts' law in HCI. Considerable debate has gone into which formula gives the straightest regression lines. In other words, the assumed goal is to establish a predictive scientific law. Granted, in some cases, an HCI researcher is interested in estimating how long tasks will take given different interface arrangements, and requires Fitts' law to provide a predictive model of human abilities [MacKenzie, 1992b]. This model can then be used to optimise interface layouts, without the laborious process of user-testing each one, as in [Zhai et al., 2002]. However, a far more common use of Fitts' law is to establish which devices are most effective *via* user testing [MacKenzie, 1992a]. I argue that these two scenarios are distinct, and propose that many of the issues with Fitts' law in the former disappear for the latter if an ISSR-based methodology is used.

In fact, the ISSR approach dispenses with the notion that we must base throughput measurements on a linear time-ID relationship. A constant rate of information

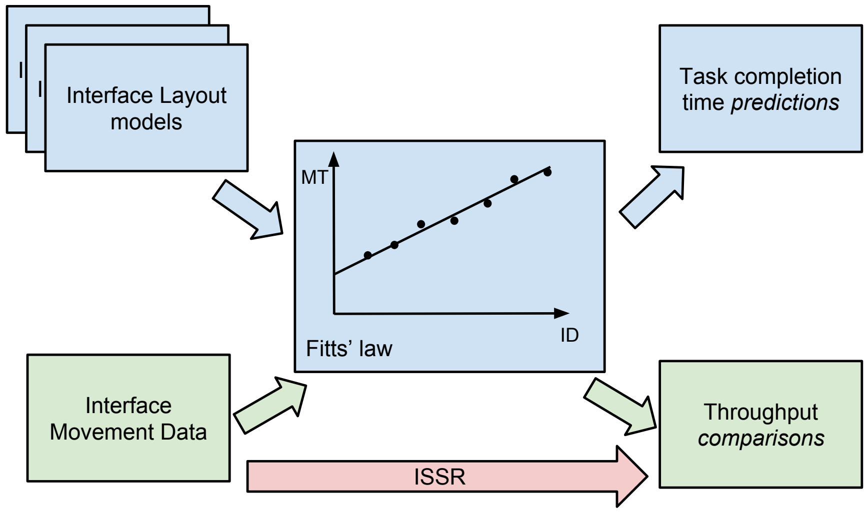


Figure 5.4: Two uses of Fitts' law in HCI. The top paradigm (blue) uses Fitts' law as a model of human behaviour to predict task completion times. The bottom paradigm (green) uses Fitts' law to estimate throughput (TP) for the purpose of comparatively evaluating input devices. ISSR sidesteps the debate over the relationship between throughput and movement laws, by redefining TP in terms of search space reduction.

flow is not a prerequisite for measuring the rate or amount of information transmitted. Even if we would abandon Fitts' law altogether, Fitt's idea of information transferral via aimed movement should still prove very useful for interface evaluation.

Evaluation is methodologically distinct from establishing predictive scientific laws. In essence, evaluation is to measure some property X in a variety of experimental conditions, with the relative values of X reflecting how well the various experimental conditions contribute to the user being able to achieve their goals. If the user's goal is to adjust some data within the memory of the computer so as to align with a target value, then surely the convergence of this data on the target value is the process of interest. So the the point at which X should be measured is the change in this parameter data.

This evaluative quantity X should have the property that if, say, $X_a > X_b$, then we know interface a is more effective than interface b. The units of X should be comparable in a variety of situations, because the input device may be mapped to all kinds of different tasks: the rate of transmission of Shannon information is certainly an ideal candidate, for the reasons given in Section 2.4.1. Whilst it would be a tremendous boon if there were indeed a linear relationship between X and task completion time T, such that $X_a = \gamma X_b$ entails $T_a = \alpha + \gamma T_b$, this is by no means guaranteed. If we do know the relationship, we then get an idea of how much faster interface a really is for an arbitrary task. Nevertheless, nowhere in our desiderata for the effectiveness measure X do we require that it should be able to precisely predict *absolute* movement times, only that it provides a way to compare input devices via some quantitative ratio. Fig. 5.4 illustrates this methodological 'short cut', from experimental data directly to throughput comparison.

So, rather than defining throughput on the basis of an empirical fit to movement data, with all the debate that involves, let us instead recall the definition of information gain: the entropy reduction of a probability distribution. The details of this calculation are established in the next section. The first approach is to look at the probability distributions of an ensemble of search paths. The second approach is to make some simplifying assumptions about these probability distributions. This results in a simple formula for calculating the throughput of a target search in n dimensions.

Probability Distribution Approach

Fig. 5.5 shows the distributions of points obtained from search path data from Experiment 2. Participants were searching for a target sound randomly positioned in a 2D parameter space. Each point is positioned according to the 2 parameter settings obtained from an individual trial. Six points in time are shown. The target position has been normalised to the centre of the image. It can be seen that,

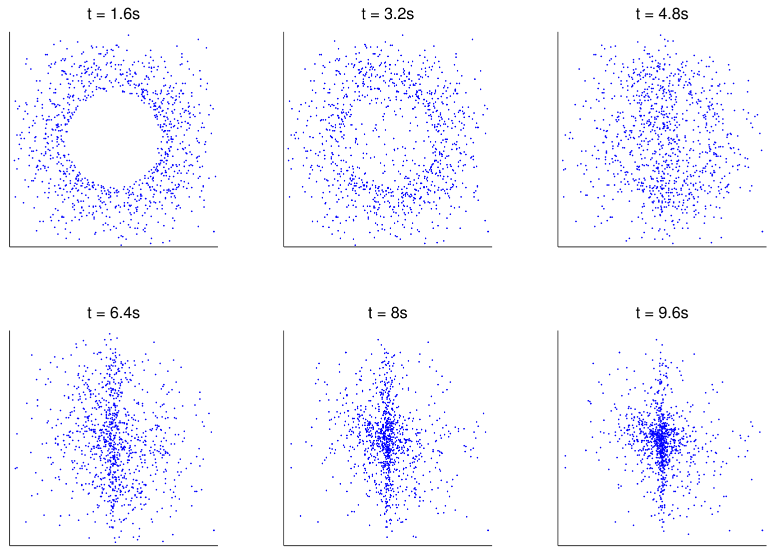


Figure 5.5: The "search ensemble" for the 2D sound target matching trials in Experiment 2. The target sound has been normalised to the centre of the image. The gradually coalescing probability distributions have decreasing entropy.

although individual search paths can be quite random, the overall tendency over time is for the distribution to converge on the target. Given a large enough number of these search paths, this ensemble approximates a probability distribution. Like Soukoreff et al. [2011], we take the approach that it is the reduction in the entropy of this distribution that reveals information gain, and hence throughput:

> "to calculate entropy, we must first identify the set of possible outcomes, and their probabilities. The ultimate goal of a rapid aimed movement is to enter a specific area of space, and so the movement endpoints seem to be the natural quantity of interest that reflects the result of the movement task."

Task information gain, or index of difficulty (ID), is given as the difference between the starting and ending entropy: $ID = H_{start} - H_{end}$.

Not only this, but if we have trajectory data for the whole path, entropy can be measured at all points in time throughout the movement, potentially giving greater statistical power than simply looking at the end points. If Fitts' law is indeed a result of a constant information processing rate, then presumably this can be observed in the course of the movement as well as for its end points (with the possible exception of the initial acceleration phase).

The entropy of this probability distribution at each point in time $H(t)$, discretised into B equally large bins can be calculated for each point in time like so:

$$H(t) = -\sum_{i=1}^{B} p_i(t) \log_2(p_i(t)), \tag{5.1}$$

where $p_i(t)$ is the probability of a search point being found in bin i at time t. This can be calculated from a distribution of N points via $p_i(t) = \frac{n_i(t)}{N}$, where $n_i(t)$ is the number of points in bin i at time t.

Fig. 5.6 shows the decline in entropy for the distributions shown in Fig. 5.5. The narrower and more focused the distribution becomes, the lower the entropy.

Unfortunately this does not give us a generalisable means of calculating the throughput. Imagine every search path started at the same distance to the target. This would be an extremely narrow starting distribution, resulting in a very low starting entropy. As the users set off, their paths will more randomly disperse and the entropy will *increase*. This would give us very strange results. This problem can be seen in the first 4 seconds of the plot in Fig. 5.6, where the entropy initially increases due to the hole caused by there being a minimum absolute distance from starting point to target (see Fig. 5.5, top left). Because this bunches up the starting points, the entropy of the initial distribution is artificially low.

The attitude to this issue taken in [Soukoreff et al., 2011] is justified on the basis of

> "...the principle of maximum entropy, which states that the probability

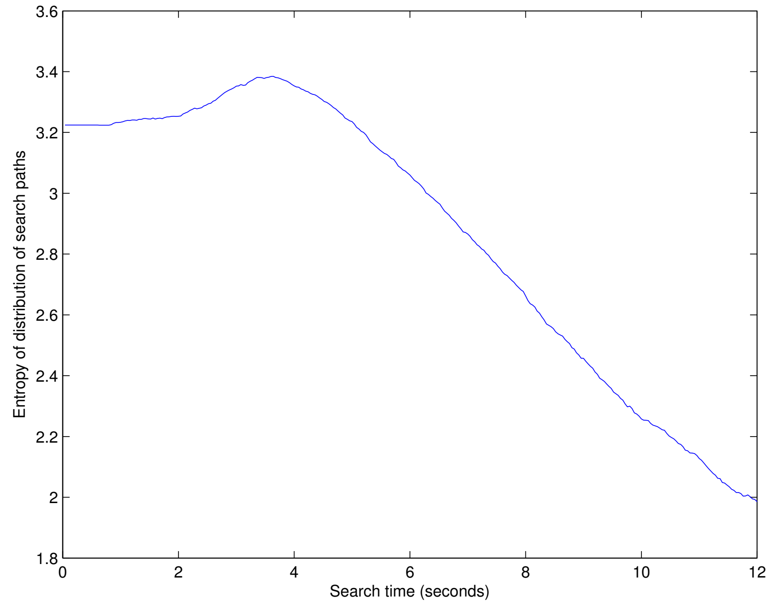


Figure 5.6: Entropy decreasing with time, calculated for the search ensemble in Fig. 5.5. The initial increase is due to the hole in the middle of the search paths: this low starting entropy does not accurately reflect the participants' uncertainty of the target position. Once the distribution has had time to settle out and more approximates a Gaussian, the entropy decreases approximately linearly. The gradient of this linear region represent throughput.

> distribution that best represents a given phenomena is the distribution that gives the largest entropy *subject to what is known about the phenomena*".

In other words, it is not the experimenter's knowledge of the distribution that determines the entropy, it is the subject's. Therefore the starting distribution should reflect the participant's knowledge/ignorance of the target just before the particular trial is presented. They therefore assume that the initial distribution is uniform across the range of possible target positions. The formula for the entropy of a 1D uniform distribution of width d_1 is

$$H_{unif} = \log_2(d_1). \tag{5.2}$$

They also assume that the distribution of the end points is Gaussian. The formula for the entropy of a 1D normal distribution of standard deviation σ is

$$H_{norm} = \frac{1}{2}\log_2(2\pi e\sigma^2) = \frac{1}{2}\log_2(\pi e\frac{d_2^2}{8}), \tag{5.3}$$

where the effective target width d_2 has been substituted for σ. By subtracting the final entropy from the starting entropy one can get a formula for the ID of a movement ensemble

$$ID = \log_2(d_1) - \frac{1}{2}\log_2(\pi e\frac{d_2^2}{8}). \tag{5.4}$$

What about the case of the our 2D trials in Fig. 5.5? We could assume that all the subject knows is that the target is somewhere within the range of the 2D square. Unfortunately however, this would then raise the question of what point in time the 'maximum entropy' distribution no longer applies and the actual data distribution begins to accurately reflect the subjects' uncertainty. It seems we need to make some simplifying assumptions. If we assume that the distribution is symmetrical around the target, and the same (let's say uniform) for both start and end points, then we start with a uniform distribution of width $2d_1$ and end with a uniform distribution of width $2d_2$. We then obtain

$$ID = \log_2(2d_1) - \log_2(2d_2) = \log_2(\frac{d_1}{d_2}). \tag{5.5}$$

Gaussian distributions give the same result. Therefore, if the distribution maintains its shape and symmetry, only the ratio of start and end distances matter. Is this assumption justified? Obviously the end distribution in Fig. 5.5 is not perfectly symmetrical. But judging the exact initial state of knowledge of the participants prior to the commencement of a search is hard to do. They may not be aware of the extent of the search space (for example when using a 3D hand tracker as in Experiment 2). They may have implicitly noticed some underlying pattern in the distance to the end points. A further point of confusion is that different trials take different amounts of time, and users submit their trials at different points—how should the multiple paths be time aligned? Should $t = 0$ be aligned, the commencement of the trial, or should the end points, when the user felt they had finished their attempt? With all these "unknown unknowns", assuming the distributions maintain their shape seems fairly innocuous.

Next, we build on this 'log ratio of distances' idea in order to define the Index of Search Space Reduction (ISSR), a informational quantity of search volume reduction for a *single* search path in an n-dimensional parameter space.

Index of Search Space Reduction (ISSR) Approach

In this analysis, we look at a single search trajectory, as opposed to an ensemble. Assume a target point $\mathbf{x}_t$ has been given to the user to find in an n-dimensional parameter space, P^n (refer to Chapter 7 for the specific experimental method). They make a movement in order to progress toward $\mathbf{x}_t$, and have reached a current position $\mathbf{x}_2$, having started at a point $\mathbf{x}_1$ (see Fig. 5.7).

We first calculate the distances to the target before and after the movement,

$$d_1 = \|\mathbf{x}_t - \mathbf{x}_1\|, d_2 = \|\mathbf{x}_t - \mathbf{x}_2\|.$$

We define the search space reduction factor R as the ratio of the n-volumes corresponding to radii of the distances. Volume is calculated as $V = Cd^n$. The exact shape of the volume, or even whether it has some probabilistic fuzziness, does not in fact matter, as the multiplier C cancels:

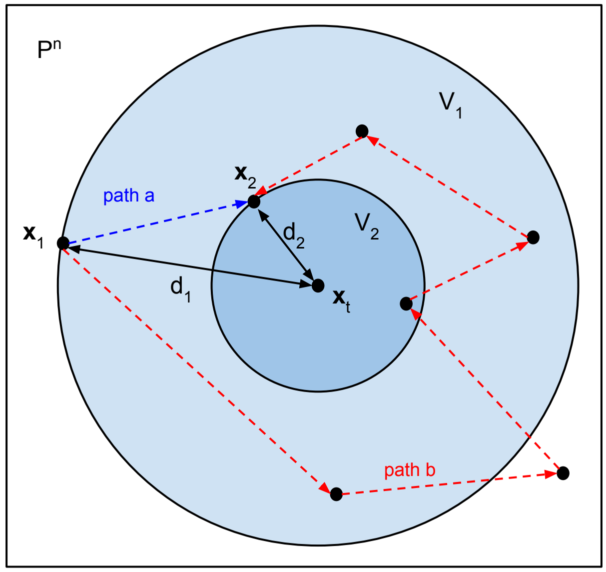


Figure 5.7: Search space reduction. The target is x_t, x_1 and x_2 are the start and end points of the movement along path a. V_1 and V_2 are the volumes associated with distances to target d_1 and d_2. The logarithm of the ratio between these two volumes gives a measure of information gain. Path b is an alternative, more convoluted search route to achieve the same result. Summing over all the steps of b gives the same amount of information gain as a (Eq. 5.9).

$$R = \frac{V_1}{V_2} = \frac{Cd_1^n}{Cd_2^n} = \frac{{d_1}^n}{{d_2}^n}.$$

The fact that parameters have finite extent, and these limits may truncate some larger volumes is assumed to have no significant effect.

In general, any search task can be said to be a reduction of a set of possibilities. For a task involving a fixed number of options, the entropy reduction (or information conveyed) by a choice of a subset of these possibilities will be the logarithm of the ratio of the number of possible states before the choice was made and the number of possible states afterwards. If the remaining search volume is reduced by a factor of two, then we have successfully completed one bit of the search. This relationship between parameter space volume and information in bits gives the "index of search space reduction",

$$ISSR = log_2\left(R\right) = \log_2\left(\frac{{d_1}^n}{{d_2}^n}\right) = n\log_2\left(\frac{d_1}{d_2}\right). \tag{5.6}$$

This makes intuitive sense. The proportion of the search space one can select for a given radial accuracy scales exponentially with dimensionality n, therefore the informational content of that selection scales proportionally to n. This hints at the promise of multidimensional controllers: if they achieve even roughly similar levels of accuracy in absolute terms, their throughput will be considerably greater. Whilst a negative "difficulty" seemed meaningless, it seems reasonable to say that moving away from the target results in lost information, and $ISSR < 0$ when $d_1 < d_2$. If no progress is made and $d_1 = d_2$, then $ISSR = 0$.

In complex search spaces such as for synthesis parameters, there may be no guarantee of a linear relationship between ISSR and movement time. But if Fitts' law holds, and information processing speed is constant, the movement time (MT) will have a linear relationship with ISSR. For MT, dependence on dimensionality would be simple: a constant multiplier of the gradient,

$$MT = a + bn\log_2\left(\frac{d_1}{d_2}\right). \tag{5.7}$$

Where does this leave the Fitts' law formula debate? In fact this alternative derivation, for the one dimensional case, gives us Fitts' original equation [Fitts, 1954]: by substituting $n = 1$, $D = d_1$ and taking the target width as twice the final distance to the target centre, $W = 2d_2$ we get

$$MT = a + b\log_2\left(\frac{2D}{W}\right). \tag{5.8}$$

Fig. 5.8 illustrates the relationship between 1D search "volume" and Fitts' original variables.

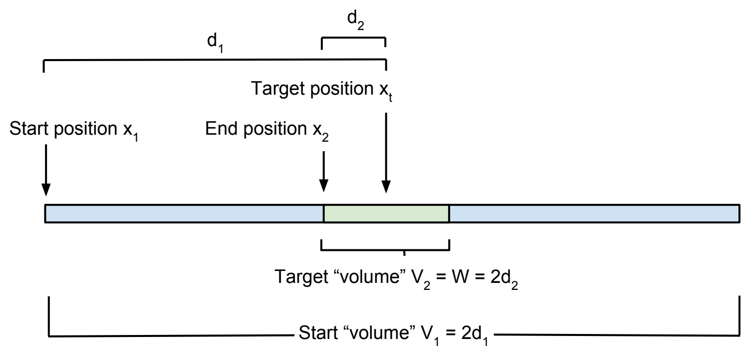


Figure 5.8: In 1D, these substitutions give an *ISSR* identical to Fitts' original *ID*.

A further reassuring property of the ISSR is that it conserves information, i.e. the total information gain of a search path can be considered as the sum of the information of all its sub-paths, irrespective of how it is divided. The sum of information gain for M steps is

$$\begin{aligned} ISSR &= \sum_{m=1}^{M-1} n \log_2 \left(\frac{d_m}{d_{m+1}} \right) \\ &= n \sum_{m=1}^{M-1} \Big(\log_2 (d_m) - \log_2 (d_{m+1}) \Big). \end{aligned}$$

All the terms cancel except the first d_m and the last d_{m+1} term, giving

$$ISSR = n \log_2 \left(\frac{d_1}{d_M} \right). \tag{5.9}$$

This is identical to Eq. 5.6 for the start and end points of the whole path. It is difficult to see how the Shannon formulation in the ISO standard *ID* (Eq. 2.4) can conserve information in this way.

Finally, as noted by Jacob et al. [1994], we must acknowledge a the possibility that a position close to the target was attained by sheer chance. In cases like these, thresholding might be desirable. The remaining search space will be whatever volume the remaining search path is constrained within. In other words the thresholded value for the current remaining search radius should be taken as the maximum value of all subsequent search radii,

$$d_{thresh}(t_c) = \max(d(t) : \{t_c < t \leq t_{final}\}). \tag{5.10}$$

This will restrict ISSR to be a monotonically decreasing function of time.

Advantages of ISSR

In summary, the proposed ISSR characterisation of Fitts' law proves useful for the following reasons:

1. It provides a theoretical baseline of how throughput should scale with dimensionality.

2. It measures information throughput at the point of interest: the effectiveness of the search.

3. Where varying accuracy levels cannot be specified in advance, it enables us to extract a range of difficulty values from the trajectory data, giving a large number of "retroactively simulated" experiments, as in Jacob et al. [1994].

4. It has a simple and generalisable definition, and can be applied to a wide variety of search task situations.

5. Information is always conserved, no matter how convoluted the search path.

6. It avoids the difficulty of having to empirically establish a predictive relationship. By measuring throughput by looking at the reduction of entropy of an ensemble of search paths, we avoid the question of whether the human nervous

system really does approximate a noisy channel, or which formula is the most accurate fit to the data. Thus the usefulness of the ISSR quantity does not rely on the experimental validity of any particular version of Fitts' law[10].

As mentioned, plots of movement time against ISSR may not be straight. Nevertheless such plots may prove revealing. For example, if the gradient becomes more shallow as time progresses, this would indicate the search is becoming easier as it progresses and that more information is being successfully input. These kind of plots will reveal a number of interesting results in Experiment 2.

One potential theoretical problem is that ISSR, averaged across an ensemble of search paths, will be the mean of the log distance ratios, whereas the probability distribution approach calculates the log of the ratio of mean distances. This is potentially a different quantity. Further work is needed to more carefully analyse the relationship between the two approaches, more carefully deal with changes in the shape of the probability distributions, establishing the exact state of the participant's uncertainty before, during and after the search.

5.4.2 Throughput, Expression and Jordà's Efficiency

Whilst throughput should by no means be the *only* measure used to evaluate musical interfaces, it nevertheless captures three important ideas in one single quantity: flexibility, speed and precision. Precision, in that it reflects how small the target achieved was. Speed is expressed through how long it takes to achieve a given accuracy. Flexibility is expressed via the fact that the calculation takes into account what the user *could have* done: the size of the search space that was selected from.

Throughput based analysis may be one useful step towards measuring Jordà's notion of musical instrument efficiency. This relied on three terms, output complex-

[10]Indeed, in Experiment 3 8 we shall find that Fitts' linear relationship does not hold for rhythmical interaction, nevertheless ISSR comparisons remain revealing.

ity C_{out} (the complexity of the musical output), input complexity C_{in} (how difficult the movement is) and performer freedom F (the size of the possibility space).

$$efficiency = \frac{C_{out} \times F}{C_{in}} \tag{5.11}$$

Performer freedom could be related to the total size of the combinatorial space V_{total}, perhaps the maximum amount of bits achievable $I_{max} = \log_2 V_{total}$. Complexity (at least in the sense of the potential intricacy of the control trajectories) might be considered the amount of fine grained specification achieved in unit time i.e. ISSR-throughput. Input complexity seems related to movement difficulty, as in the original conception of Fitts' *ID*-throughput. Therefore rather than being 'competing theories', *ID* and *ISSR* may be quantities that are useful to compare to find out the ratio between how challenging the movement was, and how much that movement actually achieved in the task-space. Referring back to Fig. 5.2, *ID* would seem to measure information flow at point (1), whereas ISSR would measure information flow through the whole input channel (1),(2) and (3). The difference between the two, $ID - ISSR$, might be able to tell us how much information was lost in the device (2) and mapping (3) channels.

For high dimensional spaces, *ID* and *ISSR* may actually diverge: the movement may be easy but the search space reduction could be large. Further work is needed to establish how throughput may be measured at all points along this loop. For instance, the "liveness" of feedback is crucial [Nash, 2012]. This aspect could be measured by return channel throughput.

Another important aspect of Jordà's conception of instrument effectiveness more closely looks at the notion of diversity, and how it extends over various time scales. The nuance control of fine sonic structure is termed 'micro-diversity'. 'Mid-diversity' pertains to how different two performances can be, and 'Macro-diversity' refers to how well the instrument adapts to stylistically different contexts across different genres, performers or even cultures. This conception of diversity is probably not addressed by simply measuring throughput—which seems to apply more to short time-scales of micro-diversity. However one could argue that a synthesiser with a larger, more controllable parameter space is likely to be better adaptable to different stylistic contexts.

Another essential question is how the complexity of the audio output changes as a function of the parameter changes. Optimising this measure, and testing it with relation to all the other points in the loop may be a potentially fruitful direction for future sound synthesis research.

For real-time musical performance the ability to sustain a high bit-rate is essential, if the performer cannot realise a musical event to a certain accuracy in a certain time, many types of music will be unplayable. However, musical expression is a far more subtle concept than mere controllability. Like all artistic notions it is multidimensional, and perhaps should be left to artists to define. The instrument designer, straddling engineering and art, must deal with both the artistic qualities of their instrument (the sonic qualities, the visual aesthetics, the stylistic and cultural resonances) and the engineering aspects (the controllability, the reliability the range of sounds that can produced). Expressivity is a notion that seems to span this spectrum. Nevertheless, approaching from an engineering perspective, if we were to demand a single, objective means to assess expressiveness, then throughput would appear to be a strong candidate. The more complexity and nuance that can be imparted to a musical event, the greater the expressive range. The larger the parameter space that can be reliably selected from, the more range the instrument will have, thus throughput is a at least a *precondition* for expressiveness [Dobrian and Koppelman, 2006].

For off-line interaction, for example in sound design or studio production, the need for high throughput is less urgent, nevertheless the speed of interaction is very important. Obviously it is unpleasant to use a slow interface, but more importantly, ideas are fleeting and easily lost: therefore the speed of their realisation can affect whether they get realised at all, and the longer they need to be preserved in working memory, the higher the cognitive load (see the survey responses in Appendix B). There is also an argument for viewing live music as a form of communication between the performer and the audience, in which case the clearer and faster a musical intent can be expressed to an audience the better.

I am not advocating that throughput must be maximised at all times during the actual performance with an instrument. That would be quite antithetical to the idea of artistic expression, and would rather be a meaningless competitive maximisation of virtuosity. Besides, the perception of the music (Channel (6) in Fig. 5.2) is also a limited bandwidth process — it may be counter productive to generate musical content that is far too complex for the performer or audience to assimilate [North and Hargreaves, 1995]. The point is that the bandwidth should be there as and when the musician requires it. The artist is then free to express themselves in as 'minimal' or 'maximal' fashion as they see fit.

One could argue that the artistic aspects of expressivity pertain to "what is expressed" i.e. the information content; whereas the engineering aspects pertain to "how the content becomes expressed" or how efficiently this content can be transmitted from the mind into sound. By this argument one can see that throughput seems an ideal quantity for the interface engineer to investigate if musical content is to be abstracted away, however in the design of a real instrument the two aspects are inevitably much intertwined. For instance, the sonic qualities of the instrument will require many engineering decisions to alter, but inevitably possess aesthetic and evocative qualities which will change the what seems to be expressed when it is played.

5.4.3 Information Flow Summary

In this section we have seen how throughput, measured via entropy reduction, may be a good correlate for the effectiveness of a content creation interface. By side-stepping the necessity for Fitts' law to be established using experimental data for each and every increase in dimensionality, a simple formula was established for calculating input information for high-DOF controllers.

$$ISSR = n \log_2 \left(\frac{d_1}{d_2} \right). \tag{5.12}$$

This quantity seems to encapsulate many of the desiderata established in prior Digital Musical Instrument research. However, the vital 'divergent' aspect of creative interaction is missing. ISSR can only measure the rate of convergence upon a pre-identified target. Thus it is only applicable to the two convergent modes in the EARS model described in the remaining parts of this chapter.

Also, yet to be established is a method of actually *designing* high-throughput interfaces, which necessitates due consideration of people's cognitive and motor control abilities. Is it really possible to control multiple dimensions at once, fast enough to take advantage of the dimensional multiplier in equation 5.7? If so what disadvantages are there to this form of interaction? What other modes of interaction are necessary in order to fully augment creative cognition?

A further question is that of cognitive load. If higher throughput comes at the cost of greater working memory demands, then it may not be a worthwhile goal. As we have seen in Chapter 2, the explicit brain system is also thought to be a processing bottleneck [Marois and Ivanoff, 2005]. According to the cognitive pipelining principle, it is vital for the throughput of the loop as a whole if low-level short time-scale interface operations are designed to bypass the explicit system when at all possible. Do we really need to consciously keep track of the value of every single parameter, or is it enough to just evaluate the music as a whole? Is it only when consciously noticeable errors occur that it becomes necessary to explicitly drill down to individual parameters and correct them? What properties of interaction are suited to low cognitive demand? The next section attempts to address these questions.

5.5 A Four-Strategy Model of Creative Interaction

This theory details how a simple two stage model of creativity (divergent vs. convergent thinking) and dual process theory (implicit vs. explicit cognition) can be combined to construct a more principled approach the design of creative composition interfaces. It is worth setting out the exact scope of this model. It only addresses what Boden [1992] terms P (psychological) creativity, rather than the H (historical) creativity found in culturally significant achievements, therefore ignores ecological factors that contribute to true artistic breakthroughs. It is not intended to be a model of how information is processed in separate modules within the brain. It is not computationally detailed enough to be implemented as a genuine artificial creative system. Specifically, it is intended to be a categorisation of parameter search strategies, a summary of how those strategies work together (or not) to create novelty and value, and how parameters could be mapped to interface gestures to assist each of these processes. This design methodology should prevent the designer forcing the user into the wrong creative problem solving strategy at the wrong time.

At the time of the original formulation of this theory, the links between implicit and explicit thinking, predictability, creative insight and solution spaces were novel. However since then several other publications have made a similar connection [Sowden et al., 2015; Wiggins and Bhattacharya, 2014; Dietrich and Haider, 2014]. However I believe this characterisation of 4 unique strategies is unique, simple and useful. It is intended to be a basic minimal amount of cognitive creativity required for effective design of a creative tool, i.e. a necessary (if not sufficient) model of human-computer creative interaction.

After the four quadrants are described, Section 5.5.5 reveals how this model sheds light on a number of issues in NIME research.

5.5.1 Divergent and Convergent Solution-Space Traversal

To start this section, it is worth tackling Dietrich's criticism that the notions of divergent and convergent thinking are 'intellectual duds'. In reference to the lack of progress in cognitive neuroscience of creativity, Dietrich states

> "...divergent thinking is way too broad a construct to be of any real use as a process. It is a compound construct that must be dissolved into its constituent processes before meaningful research can be done. In short, the concept of divergent thinking doesn't do any explanatory work for the study of creativity and it is high time that we heave it into the dustbin of outdated ideas."[Dietrich, 2007]

His other objections are firstly, that divergent or convergent processes on their own can result in creativity; and secondly, that convergent thinking tests, such as RAT (see Section 3.3.1), don't measure what they claim to measure [Dietrich and Haider, 2014]. The argument that RAT tests measure convergent thinking rests on two things: first, the fact that they have a single valid answer as opposed to many possible solutions (stemming from Guilford's definition of convergent [Guilford, 1967]), and second the fact that peoples RAT scores correlate better with other convergent tests than other divergent tests. I would argue that the RAT tests measure both divergent and convergent thinking[11], if we take divergence in this case to be spreading activation in the associative network, and convergence to be the eventual selection of the correct item. Clearly the latter ability is crucial, but that doesn't preclude a initial divergent component. The fact that these tests do not measure what they purport to measure does not mean the *terms* are meaningless[12].

[11] The unusual uses test on the other hand does seem to measure pure divergence, i.e. fluent idea generation abilities.

[12] Utilisation of the terms divergence/convergence or the Darwinian terms variation/selection probably boils down to preference. If a definition of a term turns out to be inadequate do we

Given the massive parallelism and inherent noisiness of the brain, it is almost certain that divergent thinking is a fundamental component of all cognition, not just high-level creativity. For instance, recall that movement may create a parallel populations of forward models, which are then whittled down as the movement progresses [Wolpert and Kawato, 1998]. Multiple parallel ideas may be constantly competing for entry into in the Global Workspace, and may be selected on the basis of surprisal [Wiggins and Bhattacharya, 2014]. This does indeed make divergence a non-starter for neuroscientists aiming to isolate the "secret sauce" that highly creative people possess. The creativity support tool designer, on the other hand, can simply acknowledge that this compound construct of divergence may be augmented.

Nevertheless, it is certainly worth heeding Dietrich's warnings, firstly that these concepts as they stand are not sufficiently well defined to produce unambiguous scientific results, and secondly that it is worth applying the principles of predictive, embodied cognition to investigate these thought mechanisms.

So next we shall attempt to define divergence and convergence in more detail, with references to their relation to:

1. Conceptual space traversal strategies.
2. The structure of the fitness function.
3. Their relation to prediction mechanisms.
4. Their relation to evaluation mechanisms.

scrap the terms or just refine the definition? My preference is to keep these terms for the following reasons. Convergence is a widespread term in optimisation literature, meaning the settling of an algorithm on a preferred value, or the gradual narrowing of the distribution of some population. This is both descriptive and clear. Convergence implies a more anticipatory, value-driven traversal of solution space than mere "selection". Divergence is its ideal counterpart, being linguistically its opposite, and also highly descriptive of a parallel spreading network of associations, or a spreading probability distribution.

Definition of a Convergent Process:

Convergent processes are conceptual/parameter space traversal mechanisms that work upon improving the fitness of solutions. This could either be a *selection* of a discrete option, for example selecting the best sound from a number of candidates, or the *honing in* on an optimal point along a continuum, for example finding the preferred setting for a synthesis parameter via incremental adjustments.

In the continuous case, convergence requires both fitness evaluation $\mathscr{E}$, and some prediction of what changes will increase value, which yields a conceptual space traversal strategy $\mathscr{T}$. Prediction of increasing value is therefore actively employed in guiding $\mathscr{T}$, analogous to a gradient descent algorithm. So whilst some models of creativity postulate generative and evaluative stages, where the latter is just selection of the best solutions, an important aspect of the EARS model is that convergence can still change the solution via incremental, predictive improvement (c.f. the 'honing' theory of creativity [Gabora, 2005]).

If the fitness function is very rough and unpredictable, convergent processes will not prove very effective at finding optimal solutions. Convergence by itself will rarely produce novelty, as multiple runs will have a tendency to follow the same paths. Therefore creativity requires a mechanism to extract itself from local optima: hence the need for divergent strategies.

Definition of a Divergent Process

Divergent processes are different in that they set aside questions of fitness, and generate candidate solutions via traversal mechanisms that ignore any prediction of increasing value, e.g. creating lots of scattered points. These points might occur according to some probability distribution, the variance of which may be controlled by the creator to a greater or lesser extent[13]. This would entail that some distance metric may still be important for divergent processes, but the direction of traversal is not. $\mathscr{E}$ may still operate in the background in order to spot promising new ideas, but predictions of value are disengaged from directly determining $\mathscr{T}$, in order to prevent it revisiting unoriginal ideas.

Divergence by itself will produce useless noise, unless evaluation and/or selection processes are applied at some point. So it is the careful blending of these processes that yields progress. Of course, the exact way to combine these processes in the optimal way is an entire research field in itself. Examples abound from AI machine learning that combine both divergent and convergent behaviours, such as random forests, genetic algorithms and particle swarm optimisation. Balancing the two tendencies is sometimes known as the exploration-exploitation trade-off [Barto, 1998], or as diversification vs. intensification [Blum and Roli, 2003]. The brain, presumably, has attained a masterful blending of these two tendencies. Luckily, we do not necessarily need to know the exact details of how the brain achieves this balance, however the implicit-parallel and explicit-serial distinction seems extremely suggestive of a mechanism of variation (unconscious-parallel) and selection (entry into the global workspace). In the next section we incorporate this distinction into our creativity model.

As noted before, sightedness, or the ability of the agent to predict the value of a local region of the solution space, may vary. Absolute certainty is rare, absolute uncertainty is impossible. Therefore the divergence/convergence distinction is a continuum, rather than two truly distinct states. Lest this confuse the issue, for descriptive purposes we shall portray the two processes as distinct, whilst ac-

[13]Defining divergence in a probabilistic manner maintains the spirit of Guilford's initial conception, but removes the misleading idea that divergent thinking necessarily produces multiple solutions to a problem, it is rather the extent to which predictions of value drive the generation process that is the key factor.

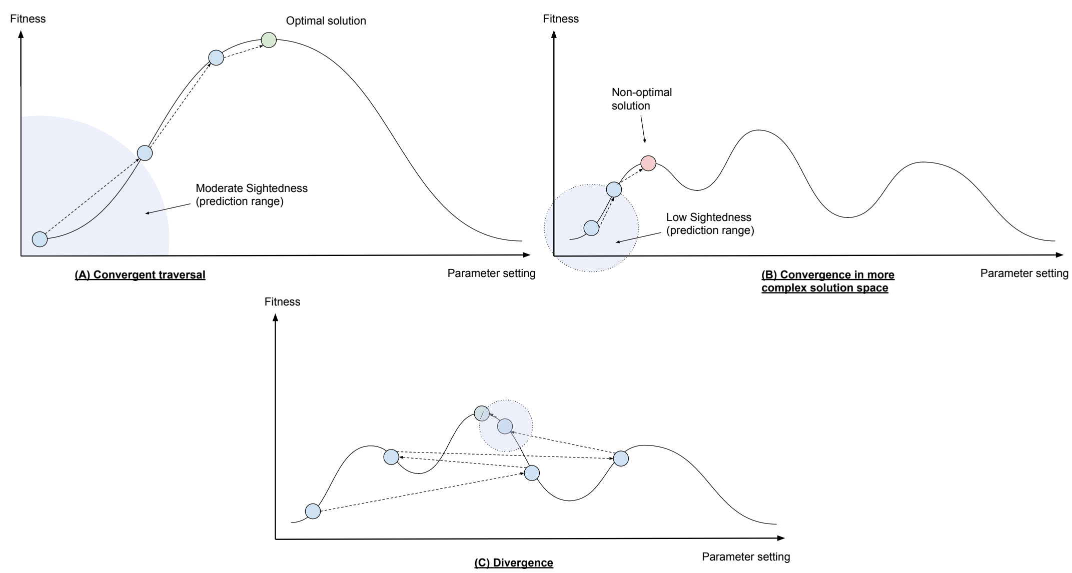


Figure 5.9: An illustration of how sightedness and fitness function complexity affect the appropriateness of divergent or convergent strategies. A successful convergent process is show on the top left. The process is able to make improvements by predicting the direction of increasing fitness. For a more complex fitness function, or when sightedness is reduced, convergence becomes less effective, and has a tendency to get stuck in local maxima (top right). An injection of exploratory randomness (bottom), can overcome "barriers" in the fitness function, but this comes at the expense of less efficient search paths.

knowledging that it would be good to design for *varying* amounts of divergence, so that the user can 'adjust the novelty thermostat' [Jennings, 2008]. Stochastic optimisation algorithms often progressively reduce the diversity component as the search progresses, a trait seen in human creativity too.

Therefore, how well the artist can predict the effect of their interactions (sightedness), how complex the search space (and the fitness function) is, both affect whether convergent or divergent strategies are appropriate. Figure 5.9 illustrates this.

5.5.2 The Importance of Explicit, Strategic Divergence: Evaluative Decoupling

In this section we claim that by intentionally inhibiting predictions of increasing value, the artist can disrupt their own convergent thought processes. Therefore, inhibiting evaluation can be an important creative strategy. This leads to a discussion of how both unconscious and conscious brain processes can generate divergent and convergent thought.

It should be noted that divergent processes may be considerably cheaper, cognitively speaking, than convergent processes. It should not take as much 'effort' to generate a random sojourn through parameters than to generate an accurate prediction of an optimal route. As we saw in Section 2.2.5, the further into the future the predictions extend, the more demanding they are, and the higher the hierarchical level employed to deal with them. However, this does not mean divergent behaviour emerges automatically. In fact, if forward models drive all human actions, this entails that prediction of the value function $\mathscr{E}$ drives $\mathscr{T}$ at a very fundamental level[14]. It would seem that our current definition of divergence becomes impossible under this scheme. If prediction *always* drives action, then how do we ever produce exploratory, unpredictable behaviour? It seems to only become possible if other levels in the hierarchy play an inhibitory role. Here are three divergence inducing strategies:

1. Predictive mechanisms have not yet been trained for the parameter space in question. The agent has no choice but to fall back on blind variation.

[14]The cost/value function can in fact be shown to be equivalent to a Bayesian prior: "the replacement of value and cost functions with prior beliefs about movements removes the optimal control problem completely" [Friston, 2011]. In other words, the probabilistic forward models underlying movement *are* a value function, albeit one with some uncertainty. Without wishing to belabour the point, it could also be argued that evaluation is a form of predictive simulation too: the producer in the studio may not be asking themselves 'how much am I enjoying this music right now?', rather 'how much would I be enjoying this music if I was hearing it as part of an audience?'.

2. Higher levels in the cognitive hierarchy can somehow interfere with lower level predictions by top-down inhibitory processes. If an action-generating brain module can be decoupled from its error-correcting feedback, it may gain freedom to generate more divergent behaviour. Some researchers claim to have found evidence for this decoupling in improvisatory creativity (see Section 3.5).

3. The 'objective driven' value function is abandoned, and replaced by a value function based purely on novelty. The agent is no longer searching for anything other than surprising things [Lehman and Stanley, 2011].

Divergent strategy 1 implies that to produce exploratory interaction one must *force* oneself to venture into unpredictable parameter spaces, such that the brain's prediction machinery fails. This is the motivation behind the Sonic Zoom experiment (Chapter 6), where deliberately unpredictable mappings are introduced. Here a higher level explicit strategy has deliberately put a 'spanner in the works' of the lower level perception-action couplings.

Regarding strategy 2, a 'convergence interrupt' could be one of the functions of the reflective system, which prevents conscious attention from 'collapsing' the parallelism of implicit thoughts to a single solution. Alternatively, by engaging in a skilled interaction whose speed is too fast to pass judgement on consciously (see Section 2.3.1), the musician may be forced into a state of evaluative decoupling, hence subconsciously generated novelty may increase. This may explain why some musicians are so drawn to live, improvisatory forms of interaction [Barrett, 1998; Johnson-Laird, 2002].

If we assume that conscious selection is a discrete process (i.e. a 'sampling' from the implicit, parallel distributions [Dehaene, 2014, chap. 3]), the *rate* at which selection occurs will determine the spread of the parallel solutions. Fig. 5.10 sketches

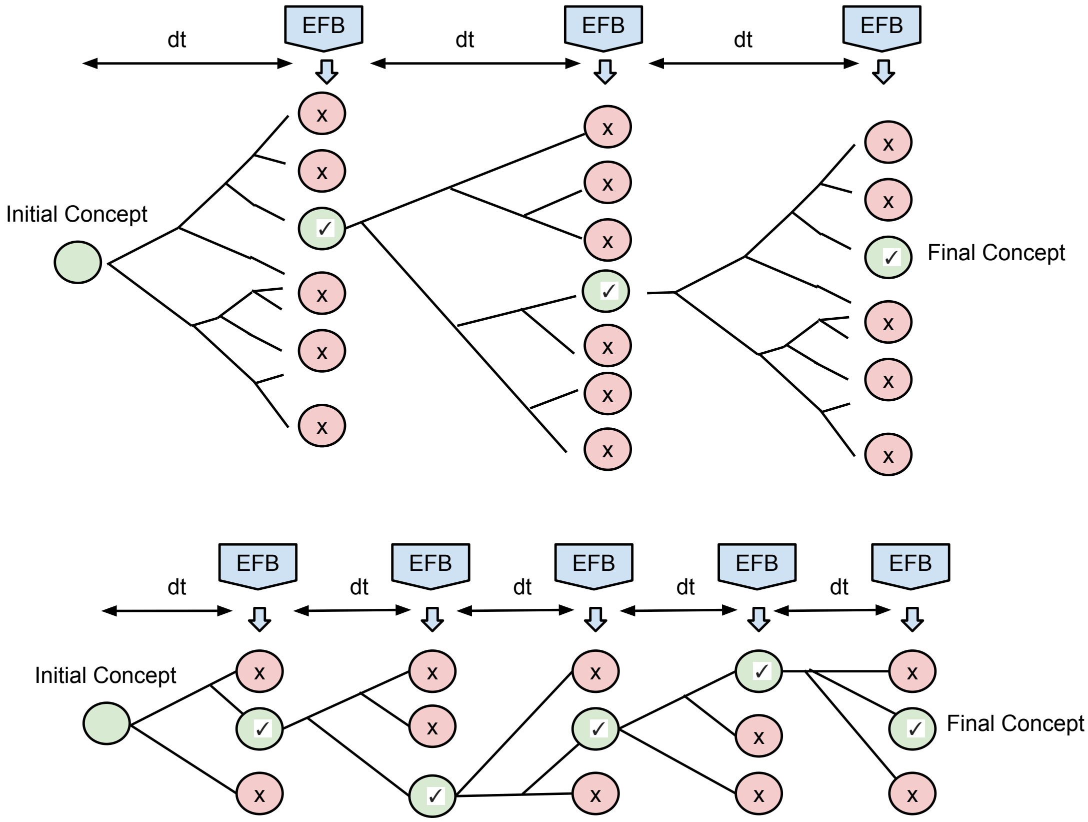


Figure 5.10: An illustration of how different evaluative feedback (shown as EFB) rates could give rise to differing amounts of spread amongst creative solutions. The less frequent selection events produce wider associative trees (top). The shorter the evaluation cycle, the narrower the spread of parallel concepts (bottom).

this process. The longer the time period *dt* between evaluations, the more diverse the associative tree. Perhaps, by suspending judgement, more novel solutions may be reached. This bears a strong similarity to the incubation process, which can be inhibited by top-down conscious effort [Baird et al., 2012]. It also bears a distinct similarity to the fact that use of slow feedback paths can slow down movements: it is the feedback that whittles down the population of forward models to make them more accurate, but at the cost of parallelism and speed[15].

Strategy 3 seems also to be a explicit, reflective strategy. An artist can con-

[15] In later speculations, we take this idea to its logical conclusion. Evaluative decoupling is linked to the concept of open loop control, and the subjective experience of inspiration and Flow.

sciously take the attitude of engaging in a curiosity driven process rather than an objective driven process.

As we have seen in Chapter 2, the distinction between implicit and explicit cognitive processes is crucial for studies of behaviour in both HCI and creative cognition. Therefore it seems pertinent to ask how these two systems relate to divergent and convergent thought, and how this will impact creative technological interactions. How do these two axes interact? There are a number of possibilities. First, we could consider that the unconscious is the source of divergent thought, and consciousness handles convergence, via selection of the best solution from a number of parallel unconscious ideas. Our unconscious appears very capable of generating divergent thought. When the autonomous modules are left to themselves they seem to 'freewheel' and generate many ideas in parallel, via recombination and associative chaining [Wiggins, 2012]. But it is clearly not the case that implicit cognition cannot carry out convergent thought. Our implicit knowledge clearly excels at selecting 'correct' solutions to the various everyday problems of life. Thoughtlessly reaching and picking up a cup of coffee clearly involves unconscious control, which is carrying out processing to converge on a specific outcome. The explicit mind can converge on solutions too, via semantic reasoning, episodic recall, and algorithmic problem solving procedures.

Furthermore, it seems that consciousness can also carry out divergent thought — i.e. produce genuinely new concepts — in a fundamentally different way from the implicit system. The implicit system can carry out exploratory and recombinatory creativity via freewheeling associative generation. What it cannot do is introduce new, higher levels in the abstraction hierarchy. Noticing and encoding a new pattern, or 'meta-concept' seems to be the exclusive preserve of conscious reflective thought[16].

[16] It could be argued that 'consolidation'—the more efficient recoding of recently formed memory—occurs unconsciously. However, there is evidence that consolidation occurs during dreams [Wamsley et al., 2010]. Dreams are very much conscious states.

So it seems that we can consciously decide to generate novel things purely for the sake of it. We can ask questions with no answer. We can take an existing set of concepts and abstract from them new meta-concepts. We can take an overview of our existing solutions, declare them not good enough, and intend to think differently. We can notice connections and patterns that no one has noticed before, and then use them generatively. This explicit-divergent ability has clear parallels with Stanovich's reflective mind [Stanovich, 2009].

So having made the argument that both implicit and explicit systems can carry out both convergent and divergent thinking—giving four quite distinct traversal strategies—the next section establishes what properties these four quadrants display, including how they may relate to the geometry of parameter mappings.

5.5.3 The EARS Model

The central hypothesis in this section is that both fast-parallel-unconscious and slow-serial-conscious brain systems may conduct convergent or divergent searches. This results in four conceptual space traversal strategies. Different representations (mappings) of the parameter space suit these different strategies to a greater or lesser extent.

Figure 5.11 shows the four quadrants: divergent-implicit (Exploratory), divergent-explicit (Reflective), convergent-implicit (Skilled) and convergent-explicit (Algorithmic). These may be strategies carried out within the brain (conceptual space traversal), or actual manipulations of the controls of an instrument (parameter space traversal). Below, each quadrant is described in more detail, both in terms of cognitive processes and musical interaction styles. We also discuss their relation to various speeds of feedback from the interface. These various aspects of the four quadrants are summarised in Table 5.2.

	Exploratory	**Algorithmic**	**Reflective**	**Skilled**
Cognition	Spreading activation, recombination, dreams	Analytical thought	Reflective meta-cognition	Subconscious automatic actions and perceptions
Working memory demand	Low	High	High	Low
Computational processes	Stochastic	Means-ends analysis, rule based	Compression-reward-based learning?	Supervised learning of input-output mappings
Musical context	Early stage composition	Production	Composition, time away from interface	Live performance
Electronic musician behaviour	Experimentation with random adjustments	Setting parameters to achieve an explicit goal	Stepping back to get an overview, thinking outside the box, switching approaches	Virtuosity
Appropriate Interface	Low demand, unpredictable diverse output	Separate predictable parameters	Supports abstraction, programmable	Learnable, multi-dimensional, high-throughput
Mapping geometry	Few-to-many	One-to-one	Few-to-many?	Many-to-many
Feedback	A fast way to evaluate many options	Individual perceptual dimensions changing	Chunked/abstracted overviews	Rapid audio, haptic and proprioceptive responses

Table 5.2: Summary of the various aspects of the EARS quadrants, in terms of cognitive processes, suitable interface design, and musical behaviour. The examples given in this table are not intended to be exclusively related to the EARS modes, rather they just provide typical instances. For example, real-time performance provides an archetypical example of the skilled mode, but surely other modes may come into play during performance.

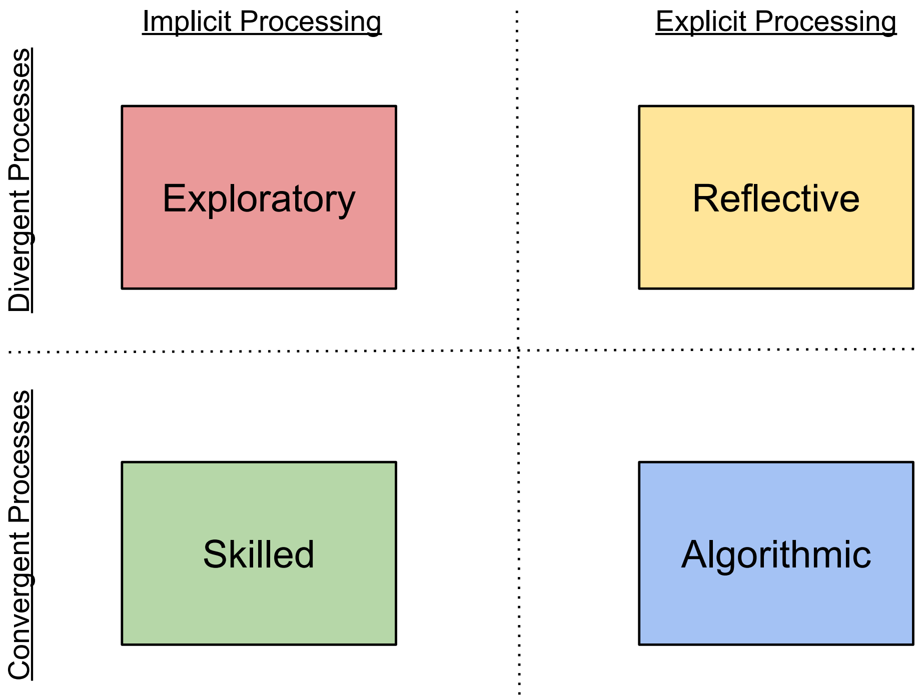


Figure 5.11: The four quadrants formed by drawing distinctions between implicit vs. explicit thinking (left/right) and divergent vs. convergent thinking (top/bottom).

Skilled (Implicit-Convergent)

Skill is intended to refer to those instinctive or learned techniques that quickly produce a valuable, but probably unoriginal local solution to a problem. These could be instinctive, or learned well enough to become automatic. If the gestural location of a target is stored as chunked unit, the motor system can proceed there in a diagonal fashion: taking the shortest path between solutions. In this mode the separate parameters of a problem are hypothesised to be treated in an integral fashion, and processed in parallel.

The appropriate interface is a well learned complex, multi-dimensional, space-multiplexed interface, but could also be an interaction metaphor such as a physical model that makes use of instinctive understanding of the physical world. Skilled interaction often involves a large amount of training: such that the solution space becomes predictable, and hence navigable, by subconscious encapsulated processes.

The feedback for this mode of interaction should be as rapid as possible, therefore be processed low down in the predictive hierarchy. For instance, haptic or proprioceptive feedback is important for skilled interaction. The information required for location of a correct solution is a by-product of errors in a previous reinforcement learning scenario, therefore the feedback need not be consciously evaluated in real time, hence gestural control can proceed as 'open-loop' motor programs. Any interaction that requires a temporary goal state to be stored in working memory (e.g. a means-ends analysis) is therefore not possible to process in a skilled fashion.

Whilst live performance may be the archetypal example of this mode, implicit-convergent interaction is not exclusively real-time musical performance. Any fast, automatic action—for example our everyday use of keyboard short-cuts—could be referred to as skilled.

Exploratory (Implicit-Divergent)

Exploration consists of experimentation with random adjustments of parameters. Such randomisation may extract an agent from local optima and overcome the barriers in the cost function formed by regions of poor solutions. These barriers ensure that convergence oriented processes tend to get channelled toward non-novel solutions.

Cognitive examples are the unconscious process of spreading activation, conceptual recombination, techniques such as brainstorming, dreaming, or simple uninhibited playfulness.

Computers effectively generate random, transformed and recombined data, therefore for an electronic musician exploration is quite easily augmented, and to a certain extent is well catered for already. Nevertheless there may be room for improving this process, as attempted with the Sonic Zoom application in Chapter 6.

In terms of interaction, the traversal strategy could provide a way to alter the unpredictability of the results, and also select meaningful subsets of the parameters to explore, whilst keeping previously converged parts of the solution intact.

The feedback required during exploration is simply a way to evaluate the results as quickly as possible, and pick the most promising candidates for further refinement. Providing responsive feedback for the lower-level brain modules is not particularly important because they are not required to generate predictions. In fact, it may be important to thwart automatic skilled responses in order to generate more novel exploratory behaviour. Dimensions can be treated as integral, as there is no need for specific prediction of changes in individual perceptual dimensions.

Algorithmic (Explicit-Convergent)

Analytic/algorithmic[17] processes break a search into separate components, and solve them in a sequential way. In the solution space they would tend to proceed in a city-block fashion, one dimension at a time.

In a more predictable, sighted solution space it may be possible to converge on solutions via some explicit step by step processes. If the individual operations are independent, they may commute[18]. If I wish to produce a short synthesised sub-bass note, then I can proceed via a recipe: I simply set an oscillator to a sine wave, set its frequency low, and the amplitude envelope to decay quickly. As long as each dimension is set correctly it does not matter what order these steps occur in.

[17]Stanovich calls the analytic part of the explicit system 'algorithmic'. An unfortunate aspect of this term is that it suggests a thoughtless, inflexible and formulaic process. Analysis, 'resolution of anything complex into simple elements' better reflects sub-goaling and the 'taking apart' of the individual steps in a complex action that is hypothesised to be the function of the cognitive control network.

[18]As in the mathematical term: the order of applying the operations does not change the result.

In more complex spaces, without independence of operations, this strategy becomes much less effective. A more demanding method of step-by-step convergence is possible: where again, the route can be broken down into smaller individual success criteria, but this time each may depend on previous and future steps in some complex way. These kind of situations are surprisingly common in music software use. As a simple example, if a GUI has many floating windows one first has to find the correct window, and then find the parameter one wished to adjust, swapping the order these operations will fail. A more complex example is obtaining real time control of a parameter via a hardware control surface: One first has to plug in the controller, then select the parameter, enable mapping mode, turn the physical control, enable automation recording mode, de-select other record enabled tracks, and so on and so forth. Many of these steps depend on previous steps. In all likelihood some error is made and one needs to work backwards through the chain to troubleshoot the problem. This is referred to as "mean-ends analysis", as the traversal strategy steps have to be calculated by working backwards from the goal to generate a sequence of sub-goals. Note that, as discussed by Sweller [1988], "conventional problem solving in the form of means-ends analysis requires a relatively large amount of cognitive processing capacity". Obviously this type of solution finding is heavily reliant on explicit thought and working memory.

An example of an analytic interface is a DAW that provides individual parameters as knobs and sliders, may have many views and sub-views, and where parameters are accessible one at a time via serial, time-multiplexed input devices such as the mouse. Hardware synthesisers are also fairly analytically oriented. The great advantage of this mode is that complex problems can be broken down into simpler parts. With well defined goals and predictably behaved parameters, accurate location of desired solutions can be achieved in linear time, despite the exponential increase in the size of the space.

Reflective (explicit-divergent)

There are no doubt a wide variety of means by which the explicit system might produce novel concepts. However the following seem to be the two main deliberate reflective strategies:

- Introducing *new levels of abstraction*: generating concepts on a meta-level.
- *Top-down inhibition* of previous best solutions: encouraging novelty by precluding non-novelty.

As opposed to barriers or walls, which are just regions of poor or non-viable solutions in an existing conceptual or parameter-space, ceilings represent the barriers caused by the lack of tools to traverse a space, lack of correct dimensions to describe that space, or even the lack of rules for evaluation of points in the space. To overcome such ceilings requires some form of 'vertical divergence': mere horizontal exploration will not suffice. One way this can be achieved is to introduce new levels of abstraction. Abstraction could be thought of as a way to more efficiently code some data, either by creating a new chunk in order to manipulate large quantities of data as a single unit, or by defining a new generative technique or pattern that is considered as a way to produce both the existing data, and new data of a similar type [Schmidhuber, 2012]. This process may heavily rely on 'meta-cognition': the ability to think about ones own thoughts, and hence investigate and alter one's own creative processes. Reflection provides the capability for exploration on the meta-level that Wiggins sees as the essence of transformational creativity:

> "For true transformational creativity to take place, as described in my framework, above, the creator needs to be in some sense aware of the rules he/she/it is applying. This follows from the need to explore the space of possible rule sets defining the conceptual space... Self-awareness

> is generally cited as the property which distinguishes the artist from the craftsperson." [Wiggins, 2006]

As an example, imagine a child hitting random notes on a piano. They may happen to notice that it becomes easier to create pleasing sounds using only the black notes. By introducing a new higher level rule, 'stick to black notes' they have delineated a subspace within which exploration is more effective than it was in the space of all notes. This may seem like a convergent strategy, after all it was driven by an expectation of increased value for an average note, however in the meta-space of note-subspaces it was a exploratory strategy. This is because it was found by exploring the space of note-subspaces. The note-space 'all notes' was tried and found rather discordant, so then the note subspace 'black notes' was tried and found to be better. The divergence was being carried out on the meta-level (exploration of the conceptual space of conceptual spaces), which then drives a more efficient convergence mechanism on the lower level: that of the individual notes. Whilst this is an exploratory process, exploration on its own would not be enough, as a reflective abstraction mechanism that can generate meta-concepts is required.

A reflective musical interface might be one that offers the ability to create new musical abstractions, for example a musical programming language [Blackwell and Collins, 2005; Bresson et al., 2011]. The phenomenon of 'livecoding' [Collins et al., 2003] seems a particularly salient example of reflective strategies generating new meta-parameter spaces, and even managing to do so in a skilled and rapid fashion. Other examples of reflective meta-control in the mapping literature include [De Campo, 2014; Van Nort and Wanderley, 2007], where the parameter mappings themselves are altered in real-time. NIME research as a whole could be considered a reflective endeavour: the aim to find an elegantly represented set of principles for designing the tools to navigate musical space.

A related strategy, but somewhat different in emphasis, is inhibiting of lower

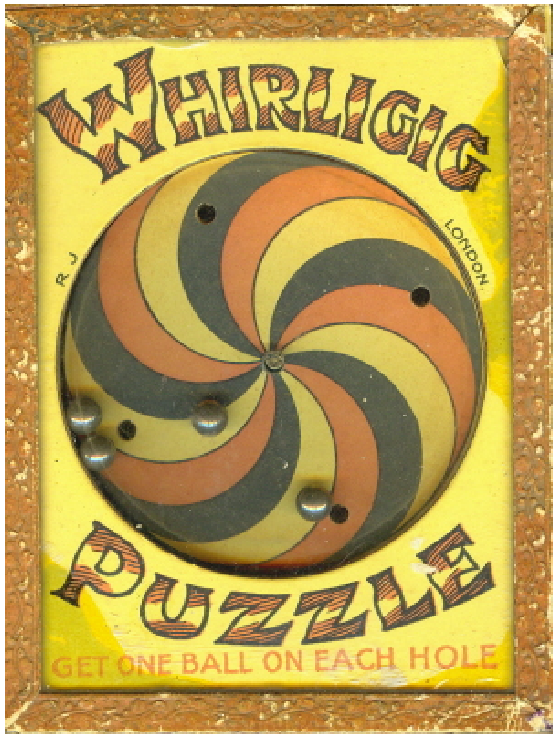


Figure 5.12: An analogy for a reflective strategy. In this solution space, the 2D coordinates of the balls cannot be directly manipulated. Rather, the gradient of the fitness function itself has to be manipulated in order to encourage convergence on optimal locations.

level convergent processes in order to eliminate their non-novel tendencies. Habits are tremendously powerful forces for generating of non-original behaviour [Duhigg, 2012]. Explicit strategies that can prevent habitual patterns from playing out can therefore generate novelty.

A reflective strategy that bears similarity to both inhibition and coding is to enforce stringent, arbitrary or even bizarre constraints or rules (see Section 3.6). This eliminates previous solutions by fiat. Looked at in terms of the solution space, self-imposed constraints can be regarded as creating a subspace: one that throws away regions of the space likely to contain non-novel material *a priori* in order to more efficiently traverse the more novel regions. This 'lossy code' could be viewed as a new concept in itself, these codes can in turn be explored, learned as skills and formalised as algorithmic techniques. Reflective divergence could be considered 'vertical divergence', in that it introduces a new level to the conceptual hierarchy, and a new meta-parameter space to explore[19].

5.5.4 Interplay between EARS Quadrants: Illuminating perceived interface problems in electronic music

All four quadrants play their part in creativity. Take incubation-illumination style problem solving as an illustration. Preparation is the process of asking a new question, or finding a new problem (reflective), and attempting to solve it, consciously via the (algorithmic) methods of the past. Applying methods based on past rules and concepts leads to repeated failure, but this process is both activating concepts in the subconscious for recombination (a process known as priming), and tacitly learning how to quickly and skilfully apply methods that seem as though they should work (constructing a neural fitness landscape that will function as an unconscious solution generator). At some point one of the many divergent (exploratory) subconscious combinations will be implicitly recognised according to some surprise/likelihood criteria, and then "miraculously" provided to the conscious mind[20]. In this way the implicit system can be set to work exploring large regions of a complex solution space in a parallel fashion.

Insight may be an example of when the four EARS strategies gel, however there are also numerous inhibition effects (some are shown as red arrows in figure 5.13), when they actually trade-off against one other. The single most important inhibition

[19]However, the more novel the parameter space, the less widespread will be the social acceptance of this novel, and perhaps arbitrary, fitness landscape. Indeed one could hypothesise that value systems cannot exist without inter-subjective consensus.

[20]Recall that Wiggins proposes that the criterion for admission into consciousness is not only the certainty of the idea as a good solution, but also an information theoretic measure of surprise: implying that novelty generation is practically hard-wired into the threshold between implicit and explicit thought [Wiggins, 2012].

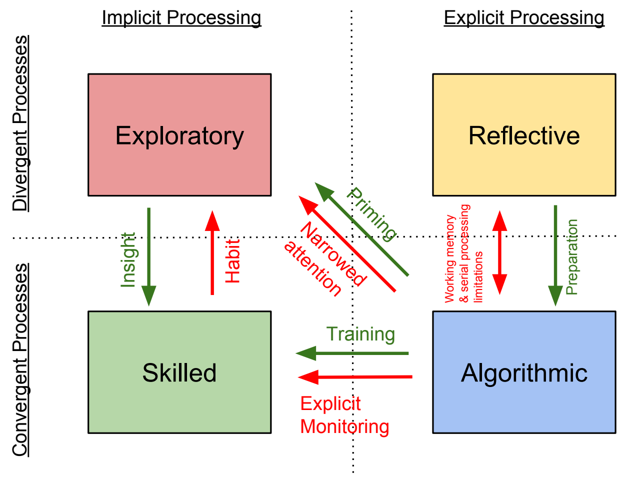


Figure 5.13: The four quadrants of implicit vs. explicit thinking (left/right) and divergent and convergent thinking (top/bottom). Examples of useful mode swaps are shown in green. Examples of detrimental interference effects shown in red.

effect is that explicit processing is a serial bottleneck, with limited working memory. Therefore if it is fully engaged with analytic processing, there will be less resources available for meta-cognition and high-level reflection.

The way virtuosity is presented in much of the literature suggests that it is more a precondition for deeper aspects of musical expression and creativity, than a goal in itself [Pachet, 2012]. The above model would seem to explain why high-throughput skilled interaction is considered so vital: it has to do with the resolutely serial nature of consciousness. If the interface is not virtuosity-supporting then this implies that the lower level modules of the motor hierarchy cannot be trained well enough to relieve the cognitive burden on the explicit system. This means working memory is occupied with interface-related tasks rather than higher level artistic goals. The reason analytic interfaces are not 'expressive' is not only that they lack nuance with which to express mental states, but also that the artist may have no cognitive capacity left for forming any spontaneous mental states to express! Furthermore, if the Task-Positive algorithmic network is occupied with interface tasks, then this will inhibit the reflective mind-wandering that may be essential for remote associations and creative incubation to occur. In order for a brain to carry out a coherent connected sequence of operations, it must necessarily inhibit the ideas that 'bubble up' from remote associations. Even assuming the incubation process could still run in the background, illumination's 'Aha! moment' may be damped out by narrow attentional processes focussing on manipulating the interface. Therefore everything points to the fact that the interface designer must at all costs relieve the burden on the explicit system. Thus the cognitive pipelining principle, already vital for HCI in general, becomes even more essential for creativity support tools.

Ideally, any interface that involves the user working backwards from a goal to a series of sub-goal related actions via a means-ends analysis should be completely redesigned. Instances include the dreaded 'menu-diving' style interfaces that musicians tend to revile, where one has to mentally work upwards and then physically work downwards through a conceptual hierarchy of sub-menus. Hierarchical organisation is certainly a way to speed up navigation through items ($t = O(\log(n_{items}))$, but it does rely on explicit, semantic reasoning.

The artist's musical goals will tend to form hierarchies too. Higher level goals (e.g. make a piece of music) are made up of lower-level sub-goals (e.g. make a bass line) which eventually are made up of even lower-level operations (e.g. alter the pitch of a single note). The lower level these goals, the more clearly defined they are, the shorter time-scales they extend over, and the easier they are to express in terms of interface components and manipulations (cf. GOMS [John and Kieras, 1996]). An important consideration is how much of this goal hierarchy must be maintained in working memory. Some goals can be assumed from context—if you are seated at a musical keyboard, you're probably making music—but other goals may be easily displaced from working memory by complex sub-goaling, interface tasks, serendipitous discoveries, or ideas spontaneously emerging from the subconscious. Given that the highest level aesthetic goals are the top of the 'stack' and probably the least well defined, it is likely that they are particularly delicate, and prone to interference when working memory is taxed. This is investigated in the survey by asking artists about 'losing perspective'—the idea that one might be so embroiled in low-level editing tasks that one's higher level artistic goals are compromised. As we shall see, artists state that low-level editing tasks do indeed have the most negative impact on maintaining perspective (see Survey, Appendix B).

Traversing the goal stack takes cognitive effort. The deeper the goal stack, the more disconnected the low-level actions can become from the higher level goals. So another prediction here would be that it is easier to express a high-level aesthetic concept or emotion using a simple, well learned instrument than a complex piece of technology. This is because there are less intermediate layers between one's aesthetic goals and one's physical actions. In terms of our 'projection' analogy, despite the fact that with a simple instrument we are projecting into a more restricted space, the projection operation itself may be easier, thus more faithful when communicating the original intention. As we have seen, the motor hierarchy is designed precisely for astonishingly fluent execution of exactly this type of projection operation: the conversion of a high-level intention into physical parameter changes. The cognitive pipelining principle would urge interface designers to utilise this ability. Rather than time-multiplexed algorithmic interface operations, which require a complex goal stack, space-multiplexed skilled interactions should be designed that can utilise the human nervous system's ability to convert intentions into data streams.

Another inhibition effect of the algorithmic mode is also shown in Fig. 5.13: narrowed attention. Carrying out a sequence of one dimensional tasks involves narrowed perceptual focus: users may be less open to peripheral cues and remote associations emerging from exploratory processes if they are highly focused on interacting with one perceptual dimension of the sound [Ansburg and Hill, 2003]. This may be a contributor to Hunt and Kirk's findings concerning complex mappings [Hunt and Kirk, 1999]. This prediction seems to align with many users' reports of using computers to make music. Evaluation of one's own work requires taking a step back to get a "big picture perspective" of musical structure at longer time-scales [Nash and Blackwell, 2012], and is interfered with if one is too focused on detail:

> Participants voiced strong feelings that computer-music systems encouraged endless experimentation and fine-tuning of the minutiae of sound design, in conflict with pushing forward and working on higher-level compositional decisions and creating finished works. [Duignan et al., 2010]

Ideally, the musician would realise this was happening and try another approach. Unfortunately, the reflective attention-monitoring system may *itself* be inhibited, therefore preventing the meta-awareness that perspective has been lost [Schooler et al., 2011]. This would mean that not only is one mired in low-level editing, one is also incapable of realising that it may not be the best strategy to employ at this current stage, and incapable of noticing some way to proceed more efficiently. So there seems to be a high risk that interfaces that overly rely on explicit-convergent processes may inhibit meta-level transformational creativity.

Analytic thought can interfere with skilled performance too. "Explicit monitoring", also known as "analysis paralysis" is a phenomenon where if an attempt is made to consciously control an automatic action, performance suffers [Masters, 1992; Wan and Huon, 2005]. This is for two reasons: firstly that much implicit procedural knowledge is simply not available to the explicit system, and secondly that the explicit system features a much longer round-trip processing time than the implicit.

The algorithmic mode is not the exclusive inhibitor of other processes. Habit, the 'dark side' of skill, naturally inhibits exploration. An automatic action will tend to be repetitive and inflexible, requiring conscious effort to suppress [Barrett, 1998].

Another problem, particularly for instrument mappings, is that the reflective transformation of a parameter space may render the musician's inflexible, implicitly trained mappings useless. Whilst the ability to remap a controller to multiple synthesis parameters might be a bonus for exploratory and reflective strategies, every time the mapping is changed, the time consuming process of building procedural knowledge has to start again. With luck, there may be at least some transferable skills to the new domain. For example, when using an unusual guitar tuning, knowledge of the positions of notes becomes obsolete (perhaps encouraging novel chord exploration), but the dexterity and strength to apply the fingers to produce rapid and clean notes is still very much applicable.

Finally, does this model have anything to say about why constraints are perceived as good for creativity? Two factors suggest themselves. First, a smaller parameter space is more easily automatisable, therefore interaction places less load on working memory, freeing up cognitive resources. Second, in a small parameter space finding novelty via exploration becomes less viable, prohibiting divergence via mere random variations. Quite possibly the narrowness of the walls forces the artist to attempt escape via the roof: with the reflective system freed up, the artist can utilise more conscious reflection, introduce higher levels of abstraction, and carry out exploration on this meta-level. The artist is forced to think outside of the restricted parameter space, resulting in more transformational creativity.

5.5.5 Mapping Quadrants to Mapping Techniques

The first objection to this model that might be raised is, what real difference does the interface make? Doesn't the human adapt such that they know how to use the interface to get what they want regardless? Do properties of a tool really affect the cognitive functioning of its user? The amount of time, money and effort musicians spend on their equipment would imply that it does, but it requires testing more thoroughly. From a purely theoretical standpoint we can say this: every computer interface expects a certain form of input; the interface is a presentation of a number of options. Therefore the creative question has already been framed in a certain way. In the case of separate controls for timbre parameters, there is an immediate question posed to the user by the technology: which control needs altering? In the case that the user has no precise idea (yet) of what kind of sound they wish to create, this is already the wrong question, and a glaring case of premature specification. The precise direction does not matter, only that one can explore sounds in as effortless a way as possible, and leave the higher cognitive functions free to listen, evaluate and possibly be inspired. This scenario requires a divergent-implicit exploratory interface, not a skilled or analytic one. In the case where the user *does* have an idea, this idea must be either (a) broken down in the user's mind into its separate properties and then built up step by step (explicit-convergent) or (b) performed real-time using the performer's expert implicit-convergent skills. The more the musician can rely on (b) the more high-level creative reflection their frontal lobes can engage in.

Most interfaces render some regions of the space more probable than others. For instance a a pair of 1D linear controls render a cross shaped region more probable than the other regions, for the reason that swapping controls takes time. Hence regions of the space with many different control alterations will come at a 'switch-cost' in time and effort. This reduces the explorability of the space. Fig. 5.14 shows a 'heat map' of explorability diminishing towards the corners of the space. Multidimensional controllers seem promising for exploratory interaction, because they should decrease the cost of diagonal movements.

A more fundamental argument for the effect of an interface on cognition is obtained by considering free-energy and embodied cognitive principles that link action to perception. An array of 1D controls are tools that imply a certain action by means of its affordances [Mooney, 2011]. When a musician engages with a control, they will implicitly attempt to create a prediction of what the effects of their actions will be: with a 1D control, this will be a 1D perceptual prediction. Therefore the attentional processes will be directed towards only that single aspect of the sound, and inhibit everything else happening in the music. Hence 'narrowed attention'. Action-perception coupling cascades back round the loop from the interface to affect the actual perception of the creative artefact. A parameter space 'sliced' by 1D controls might entail a sliced perceptual window on the music. On the other hand, if the effects of their actions are unpredictable and multidimensional, users' perceptual processes are decoupled from their actions, and hence the focus of attention is broader and more flexible. This retrodicts the findings in studies such as [De Campo, 2014], where complex relationships between action and perception were

> "...a far cry from telltale single slider/single parameter movements; and players appear quite absorbed in listening, so possibly the very opacity of the mapping does free players to listen more attentively to the changes their actions induce"

Since it is difficult to provide any mapping that satisfies all criteria for all situations, it is useful to indicate how different geometrical properties become more or less important for the four EARS modes. Table 5.3 enumerates the four creative

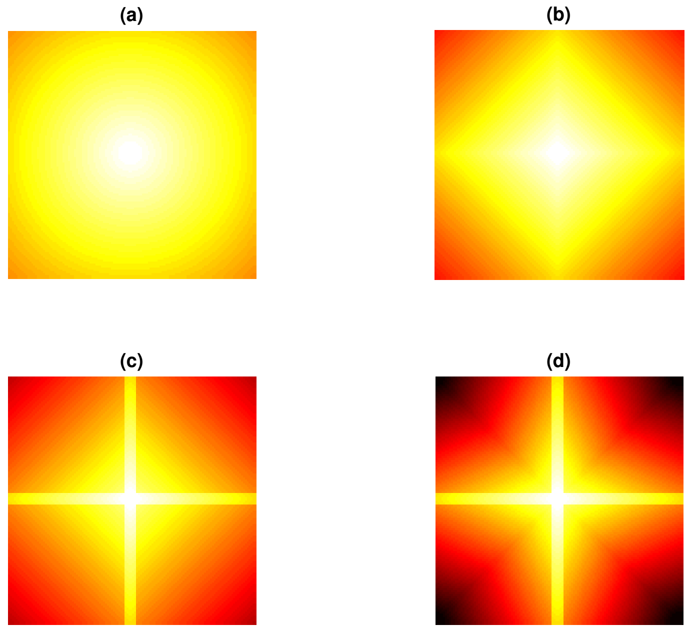


Figure 5.14: An illustration of how separate 1D controls can lower the explorability of a 2D parameter space. Increasing darkness indicates a larger time taken to travel from the centre to that part of the space. Explorability could be quantified as the inverse of the average amount of time to reach an arbitrary point, in other words the average brightness of the image. In figure (a), point in all directions are equally accessible, for example with a multidimensional controller (time increases according to a Euclidean distance metric). (b) shows what happens when controls must be adjusted one at a time (time increases with city-block distance). The corners of the space get darker: they become less accessible. What happens due to a 'switch cost' is shown in (c). Here there is some time delay associated with swapping between the dimensions: everything off the central cross becomes harder to reach. Finally, (d) shows what happens when highly diagonal movements incur repeated control swaps due to parameter interdependence. In higher dimensional spaces explorability will be even more compromised, as the ratio between Euclidean and city-block distance increases by a factor of $\sqrt{n}$ as the dimensionality n increases. Large volumes necessitate n control swaps with the zero-swap cross becoming an ever smaller proportion of the space. Even the switch cost for a single control swap may increase according to Hick's law ($\propto \log(n)$), given the increased number of controls to decide between [Hick, 1952].

stages, and which mapping properties may suit them best.

Predictability of a mapping (a result of geometrical properties such as linearity and smoothness) is clearly important for both analytic step by step construction of solutions and live, skilled performance; but becomes less necessary for exploratory interaction.

Separability of parameters can have a negative effect on performance due to encouraging slow, 1D sequential thinking [Hunt and Kirk, 2000], but certainly becomes useful when fine tuning details. Explorability should be better with integral controls, to avoid swap-cost (Fig. 5.14).

Distance Preservation seems to be useful in most mappings. Even in exploratory mode one may wish to produce larger or smaller variations.

Similarly *location preservation* seems desirable generally: the ability to get back to known favourite points when needed.

Space Elimination: A gigantic parameter space may be undesirable for performance, because it would take too long to learn. However for exploration and algorithmic construction of sounds, freedom is desirable.

Dimension reduction is useful for reducing complexity, but again if fine tuning details one wants specific parameters to be accessible.

Speed: How long does it take to do something? How fast can one traverse the space? This is a 'nice-to-have' in most modes, but is particularly relevant for skilled performance. Reflection, on the other hand, may actually be aided by things happening slowly [Hallnäs and Redström, 2001]. The size of jumps in the exploratory mode can be as high as one likes, but in convergent modes the size of the jumps must be balanced with predictability.

Hierarchical Structure: Structured information aids explicit comprehension, and new hierarchical levels are essential for abstraction and reflective divergence. However complex data structures may inhibit skilled performance.

There are probably many more mapping properties that can be considered and allied to EARS quadrants. The above examples show that EARS can be a useful conceptual model for approaching mapping geometries. Consideration of the com-

	Exploratory	Algorithmic	Skilled	Reflective
Predictability & Control	×	✓	✓	×
Separability	×	✓	×	×
Distance Preservation	✓	✓	✓	×
Location Preservation	✓	✓	✓	×
Space Elimination	×	×	✓	✓
Dimension Reduction	✓	×	✓	✓
Traversal speed	✓	✓	✓	×
Hierarchical Structure	×	✓	×	✓

Table 5.3: Desirable properties of controller mappings for different creative stages.

ponents of the model can inform designers as to exactly when certain features are necessary, and when they are not. Later chapters concerning the three experiments will test some mapping geometries in more detail.

5.6 Testable Predictions: How EARS informs the Experiments

The EARS model generates many hypotheses, not all could be investigated during the course of this work. Some of the more speculative notions are perhaps too high-level to test without a considerable amount of further methodological work. Below is an enumeration of the hypotheses tested in each experiment, and the remaining untested hypotheses. Table 5.4 provides an overview of the experiments.

5.6.1 Experiment 1

Experiment 1 investigates the relationship between exploratory and algorithmic modes. A mapping that provides divergent traversal is contrasted to one that provides convergent. The hypothesis is that an unpredictable multidimensional controller will be preferred for exploratory early-stage creativity such as fast idea discovery and generation, but separate 1D controls will be preferred for fine-tuning

Expt.	App name	Novel interface/mapping	Musical task	Task goal	EARS modes	No. params	Evaluation measure
1	Sonic Zoom	Hilbert curve 2D to 10D mapping	Sound design	Open ended	Exploratory, Algorithmic	10	Rate of saving favourites, survey responses
2	Tweakathon	Leap motion	Sound design	Target matching	Skilled, Algorithmic	1,2 and 3	ISSR, survey responses
3	ViBEAMP	Leap motion with virtual hand guides	Rhythmic performance	Target matching	Skilled, Algorithmic	6	ISSR, survey responses, NASA TLX questionnaire

Table 5.4: An overview of the experiments. Shown is what was the novel aspect of the tested interfaces; the type of musical task; whether this task was open ended or consisted of matching a goal set by the experimenter; the EARS modes thought to be involved; the number of synthesis parameters and the dimensionality of the controller; and the observed quantity used to evaluate the effectiveness of the interface.

during late-stage creativity. A further prediction is that the combination of these two interfaces provided together will be preferred than either individually, and produce more ideas. This study also investigates the notion of measuring explorability by counting the 'aha!' moments: the number of favourites that are saved in a certain time spent exploring.

5.6.2 Experiment 2

This experiment attempts to set up a minimal paradigm for testing interfaces for synthesis parameter control. The increase in difficulty for target sound location in differing numbers of dimensions (1, 2 and 3) was tested, the hypothesis being that difficulty of the search would scale disproportionately with dimensionality.

The single dimensional controls were hypothesised to be more suited to analytic control, whereas the multi-dimensional controllers were thought to be more suited for automatic, skilled control, therefore evidence of diagonal movement i.e. parallel processing of dimensions was sought.

A further hypothesis was that multidimensional controller (an XY pad and a 3D hand tracker) would be faster for locating a target sound than single dimensional controllers, but only with skill development via practice.

Another tested claim was that a time and memory constraint would lead to a higher throughput, and a more diagonal navigation strategy (i.e. more parallelism).

5.6.3 Experiment 3

This experiment looked in more depth at attaining a skilled mode of interaction using the Leap Motion hand tracker. The hypothesis was that for a small number of well practised locations in a 6D space, the multidimensional controller would be considerably faster, and all dimensions would be processed in parallel. This experiment also tested the notion that associating a preset location with a visual representation of the hand in space would provide an undemanding way to learn skilled, high-dimensional interaction. The second hypothesis was that imposing a time constraint (by conducting sound matches to a metronome) would further reveal the throughput differences between skilled and algorithmic control methods. The other hypothesis was that skilled interaction with a multidimensional controller would place less working memory load on the user. This was tested by participants having to memorise sequences of matches. Lower working memory load was expected to result in a more pleasant, flowing experience.

5.6.4 Untested Hypotheses

The reflective mode is yet to be designed for in any detail, implemented or tested. The experiments do not provide much evidence for the claim that reducing working memory load will lead to more reflective thought and hence higher levels of transformation-type creativity. The considerable challenge of measuring interface dependent occurrences of transformational creativity is left for further work; some speculations as to how to achieve this can be found in Section 9.2.2.

The increase in exploratory behaviour as a result of input bandwidth decrease (as discussed in Section 5.3.1) is not tested, though is investigated informally in the survey in Appendix B.

The claim that "evaluative feedback decoupling" (Section 5.5.2) leads to increased creativity is not tested, at least in terms of high-level cognitive processes that generate creative products. However, at a lower, motor control level, preventing subjects defaulting to slow evaluative feedback modes is shown to increase performance in both Experiment 2 and Experiment 3.

The claim that throughput is a good measure of the expressivity of an instrument is not confirmed with any rigour. However, questionnaire feedback from the participants seems to support this notion.

Finally, creating a system that supports all 4 modes, and showing it enhances creativity has not been attempted. A brief discussion of how this might be attempted can be found in Section 9.2.2.

5.7 Summary

In this chapter, a theoretical analysis of the structure of creative thought has provided a platform from which to investigate the relationship between the artist and the interface.

In the first half of the chapter, by looking at the creative perception-action loop in terms of information flow and entropy reduction, we made a number of predictions about creative interaction. Inspired by Fitts' law, we proposed a way to measure the rate at which information flows through the interface to shape the creative artefact. In the second half of the chapter, four types of creative strategy were identified. These four quadrants were related to the design of interfaces by asking how the cognitive strategies traverse *conceptual* space, and therefore how an interface should enable an artist to traverse *parameter* space. By asking how these four EARS modes interact, a number of clarifying explanations have been proposed for some of the peculiarities of the behaviour of a human-computer creative system in the particular case of electronic music.

The principal practical application of the above framework is to generate a number of guidelines by which to design and evaluate creative interfaces. Some of these already correspond to those put forward within the HCI and DMI literature, some may be novel. For a discussion of design recommendations see Section 9.2. However, one underlying principle is proposed: just as the dimensional structure of the interface (how the parameters are presented and mapped) must match the perceptual nature of the task [Jacob et al., 1994], so also the structure of the interface must be able to match the current creative strategy of the artist. The computer interface should follow the human thought process as closely as possible, not only in terms of the steps required to *render* a final product, but also in terms of the different geometries of the search strategies employed to *discover* that final product. Therefore the interface must support exploratory, reflective, skilled and analytic modes.

This analysis has provided considerable motivation for the further investigation of multidimensional controllers. The explorability of a parameter space can be expected to be enhanced, and high-throughput skilled interaction may be enabled via high-DOF input. This is the motivation for the experimental studies.

Having established a theoretical framework for creative musical interaction, this research can go in a number of directions. The first is to test the predictions of the model, and attempt to confirm its hypotheses. This would entail trying to isolate individual cognitive processes. The second is to actually use the model to design a fully functioning musical interface, and provide a complex and engaging experience for participants. The experiments to follow aim to strike a balance between these two directions by providing extremely simple, low dimensional systems, but that present complex and novel enough tasks to engage participants with rewarding, and ecologically valid interactions.

CHAPTER 6

Sonic Zoom: A Divergent Mapping Strategy Using Hilbert Curves

6.1 Introduction

This chapter presents a novel interface for navigating a musical parameter space, based around the Chapter 5's definition of a exploratory traversal strategy. The entire combinatorial space of a ten parameter synthesiser is laid out as a two-dimensional surface on a multi-touch screen. The surface can be scrolled and zoomed using touchscreen swipe and pinch gestures, reminiscent of a maps application. The user can place markers on the surface to flag favourites, and explore different sized regions around these points. The mapping from the two dimensional surface to the high dimensional parameter space uses a space-filling curve. Hilbert curves constructed from Gray codes with long bit runs can be used to preserve locality in the mapping, whilst still maintaining access to all timbral possibilities. A crowd-

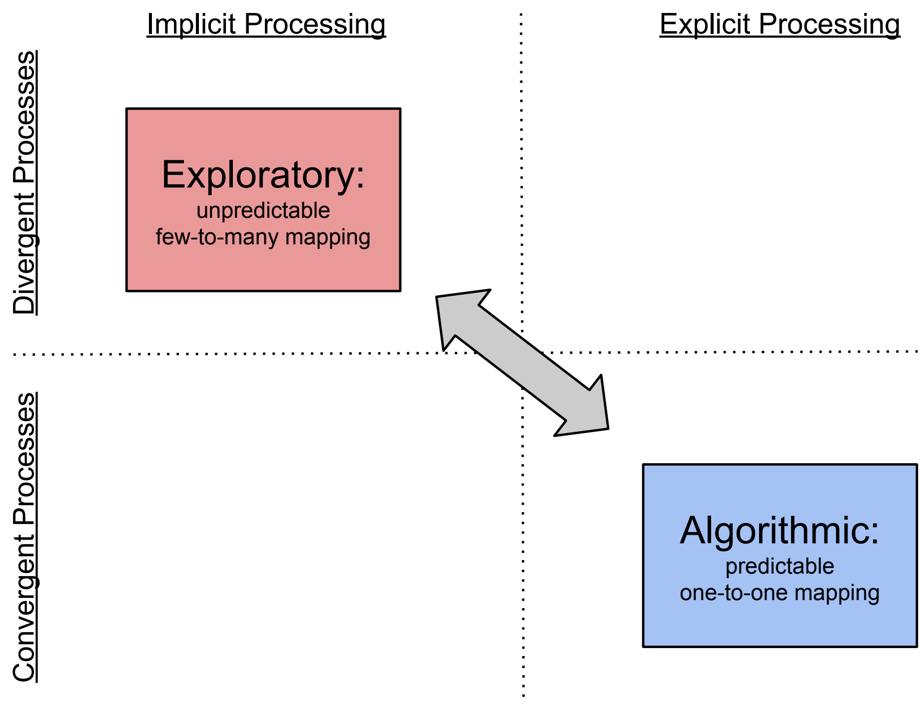


Figure 6.1: Sonic Zoom compares two interfaces, hypothesised to be suited to exploratory and algorithmic modes of the EARS model.

sourced user study was performed to compare with a more traditional one-slider-per-parameter interface. Over 300 users competed a 15 minute interaction session and questionnaire.

The experiment was conducted using a publicly available iPad app: "Sonic Zoom" (Fig. 6.2). Participants were encouraged to conduct an open-ended exploration of the different timbres available from a 10 parameter FM-subtractive synthesiser, using a combination of two different interfaces. The first was a standard interface with ten sliders, hypothesised to be suited to the "convergent" stage of creation. The second was a 2D surface with a space filling curve mapping, intended to facilitate "divergent" exploration. In terms of the EARS quadrants, these relate to exploratory and algorithmic modes (Fig. 8.1). Both the interaction data and questionnaire results show that the different interfaces did tend to be used for different stages of the sound creation process. The combination of the two interfaces was deemed more useful than either individually, reinforcing the notion that a combination of divergent and convergent processes is important for creative tasks.

In the next section, we discuss some prior work on dimension reduction for synthesis parameter mappings, and recap on some ideas in Chapter 5 relating to the exploratory interaction mode. In Section 6.3, we discuss the Hilbert space filling curve, and the particular variant used in this application. The implementation and the evaluation experiment are described in section 6.4. Results from the questionnaire, analysis of the interaction logs, and users' comments are reported in section 6.5.

6.2 Divergent Strategies, Mapping and Dimensionality Reduction

In the synthesis mapping literature relating to controlling synthesis algorithms via physical controllers, there tends to be a focus on expressive musical performance. But for this study, we investigate a mapping technique for a different stage of the creative process: namely the idea generation phase. Here, predictability is deemed less important, and access to the full range of sonic possibilities more so. We attempt to reveal whether speeding up the transitions between the exploration, evaluation and refinement stages is conducive to the creative process.

Recall from section 3.4.1, the Creative Systems Framework (CSF, [Wiggins, 2006]). A solution space traversal mechanism can occasionally produce a novel

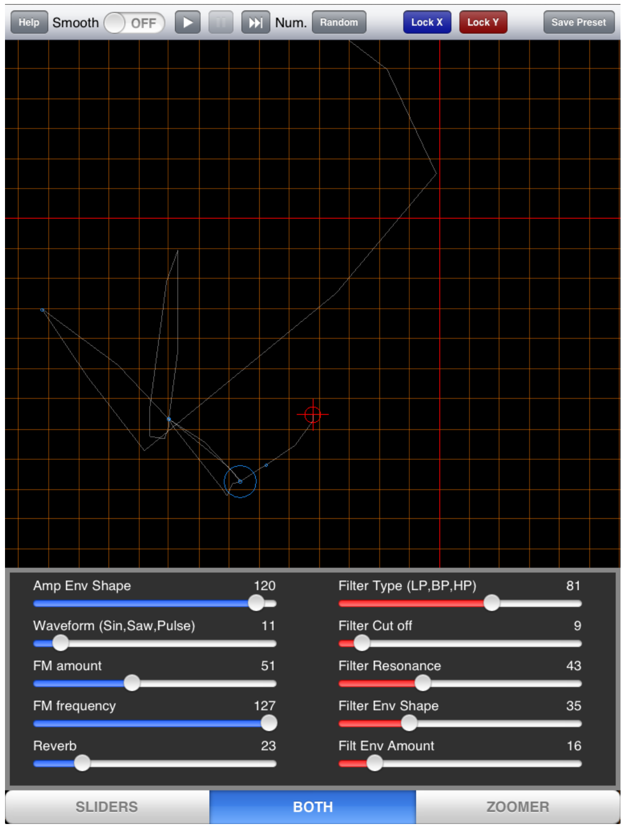


Figure 6.2: Sonic Zoom screenshot. The red cross-hairs show the point corresponding to the slider settings, so scrolling the surface alters the synth timbre. The users path can be seen as a white line, with blue circles indicating points previously listened to.

concept falling outside of the existing conceptual space. This is termed an "aberration", and can sometimes be seen as useful according to the artist's evaluation criteria. However, given the huge amount of cultural exploration that has gone before, it is perhaps unlikely that purely sighted, predictive mechanisms will yield a point outside the existing domain: rather they will tend to lead to into non-original local optima. Therefore it is an interesting question whether increasing the blindness of the variation mechanism can actually encourage aberrations, and hence the likelihood of novel solutions. It is possible that certain mapping geometries may enable a user to traverse and evaluate large parameter spaces faster, bypassing their existing predictive bias, such that local optima can be avoided.

In the survey in Appendix B, it was found that musicians, particularly those working in the purely electronic domain, say that a large proportion of the raw material they use comes from unplanned, emergent phenomena. Many claim that in the process of making music, accidents will occur and sounds will be discovered that were never intended, but prove invaluable. It seems that a surprising amount of the divergent process has already been outsourced to the computer. The aim behind the Sonic Zoom investigation was to propose and test a parameter exploration mechanism that would be effective for generating this kind of aberrant material, and hence accelerate innovative sound design, albeit in a very reduced creative domain: a simple 10 parameter synthesiser.

Dimension reducing mappings are potentially useful for both exploration and performance. In particular, there are numerous advantages to *two*-dimensional representations of data. The main advantage is that activity within the space is easily visualised. Favourite presets can appear as points arranged on the plane, and can be recalled in a more integrated way than using a drop down text list: scrolling to a preset utilising exactly the same gestural context as scrolling to explore. This has the potential to build a memorable “geography” of the sound-space, and may take advantage of spatial memory [Cockburn and McKenzie, 2002]. There is a compelling metaphor of exploration of physical terrain. The path that has been explored to date can be displayed, providing a history that can be in turn be explored. Multitouch control can be made completely consistent with maps applications: a widespread, familiar interaction paradigm. Interaction with large 2D surfaces can be made very efficient by the use of zooming [Bederson and Meyer, 1998; Guiard et al., 1999].

Prior work with explorable 2D surfaces frequently takes a timbre space approach. For example, SoundExplorer [Yee-King, 2011] constructs a zoomable 2D surface using multi-dimensional scaling on the MFCC distance. CataRT [Schwarz, 2012] enables rapid exploration of concatenative synthesis via a 2D surface that arranges the sample corpus according to various descriptor axes. SoundTorch [Heise et al., 2008] enables browsing of audio samples via a 2D surface, the sounds being arranged by timbral distance using a Self-Organising Map. Alternatively, a 2-D subspace can be generated by interpolating between existing preset points (e.g. Bencina's metasurface [Bencina, 2005], the preset explorer in [Van Wijk and van Overveld, 2003] and the "nodes" object in Max/MSP).

One criticism of preset interpolation techniques from an exploratory point of view is that they build the low-dimensional space from pre-selected favourites. This leaves open the question of how the favorites can best be discovered in the first place—presumably by manipulating the individual parameters in the traditional fashion. Furthermore, dimension reduction techniques that eliminate large regions of the parameter space on the basis of prior discoveries may in fact lower the probability of generating aberrant points that are essential for novelty generation. MDS and SOM based approaches will have the advantage that timbral distances are better preserved, but the computational cost of creating a surface from a fine grained timbre analysis of a large combinatorial space might be enormous. Therefore there is scope for developing low dimensional interfaces for the initial exploratory phase.

Reduced dimensionality may also be suited to the exploratory mode for cognitive reasons. It can confound the separability of perceptual dimensions. That is, the user is not forced to make a decision about which parameter to change, and is not prematurely encouraged into optimising this individual aspect of the sound. This might have a number of interrelated beneficial effects:

- Encourage users to take a holistic rather than analytic perceptual stance [Hunt and Kirk, 2000]. Unpredictability in the mapping may decouple the evaluation of the sound from the actions required by the interface, hence the subject's perception of the sound will not be biased by action-oriented efference predictions.

- Enable defocused rather than narrowed attention [Gabora, 2002].
- This interaction style appeals to the musicians sense of curiosity. Discoveries become more surprising, and hence may feel more rewarding from the musicians perspective.
- A simple, low dimensional interface may free up working memory for other tasks (e.g. critical listening [Mycroft et al., 2013]).
- A single, large control surface eliminates both physical and cognitive switch costs of constantly swapping between small controls.

If the musician is looking to the instrument for discoveries and inspiration, it makes no sense to enforce predictability. Put simply, if you do not yet know where you want to go it scarcely matters if you don't know how to get there.

So, referring to the mapping recommendations in Section 5.5.5, let us define an "exploratory interface". It should enable traversal speed, low dimensionality and repeatability, whilst preserving access to all possibilities. It intentionally sacrifices some degree of predictability and transparency in the action-perception coupling. Preserving locality would be useful, such that users can deliberately explore the closer neighbourhood of a sound to create larger or smaller variations according to preference. Another vital consideration is that once an interesting sound has been discovered, it should not incur too much effort to swap to a convergent interaction mode to hone it in a more predictable, separable fashion. The next section describes an intriguing mathematical object that may be a good candidate for this type of mapping.

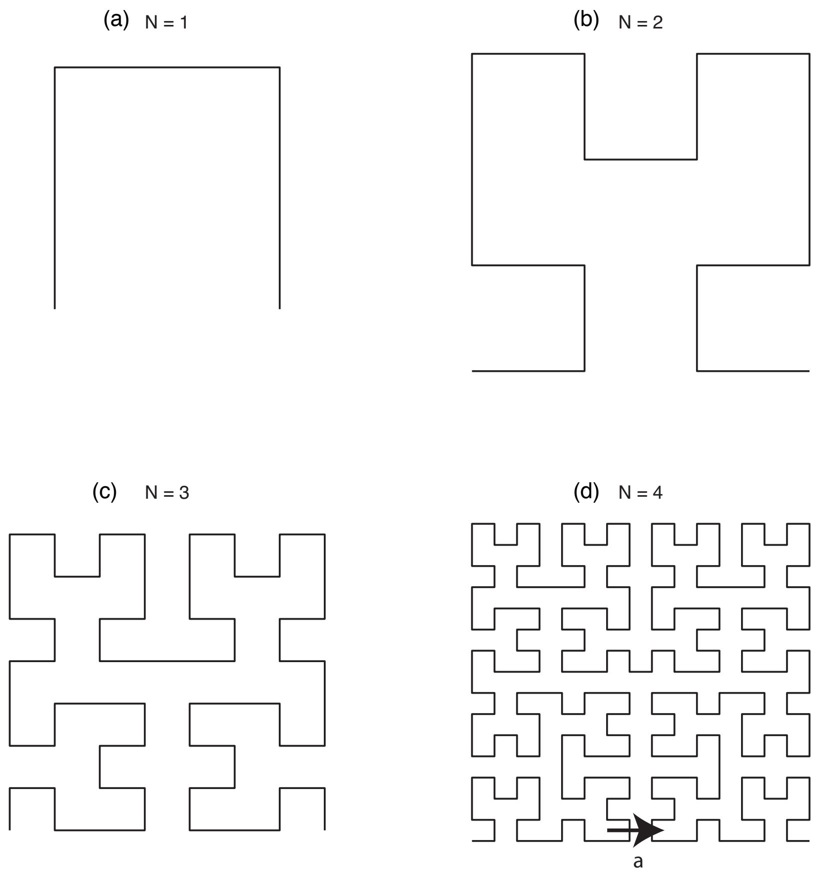


Figure 6.3: Four iterations of Hilbert curve construction in two dimensions. Arrow a demonstrates a locality violation: a small movement in the 2D plane can result in a large 1D distance along the curve.

6.3 The Hilbert Curve

How can we navigate continuously through a high dimensional space with a low dimensional controller, without rendering some regions inaccessible? A space-filling curve [Gotsman and Lindenbaum, 1995] is a continuous parameterised function that maps a line segment to a continuous path in a higher dimensional space. The curve is usually constructed recursively, is self-similar, and can approach any point in the space arbitrarily closely as the iteration parameter is increased. These mappings have proved to be useful in all kinds of applications such as clustering, data indexing, parallel computing and even a computationally cheap solution to the travelling salesman problem [Bartholdi and Platzman, 1982], due to their locality preservation properties. They have also been used for data visualisation, for instance a 3-D colour space can be distributed onto a 2-D swatch chart [Jaffer, 2005], a good example of dimension reduction increasing usability. Hilbert curves in particular are easy to construct using binary operations, and have good locality [Gotsman and Lindenbaum, 1996]. The locality property relates very well to a zoomable interface: the further the user zooms in, the smaller the accessible sonic neighbourhood will become. A zooming function will be essential, because the length of the curve will become huge as the dimensionality increases.

Fig. 6.3 shows four iterations of a Hilbert curve in $\mathbb{R}^2$. In its first iteration, the curve simply visits each corner of a square. For the next iteration, sub-squares are formed at these 4 corners, and the corners of each of these are visited in a similar manner, but the bottom left and right sub-squares are rotated to ensure the continuity of the line. The process is iterated until the plane is filled to some desired resolution. In this way, a 1D line visits every point in a 2D plane. A single control can move a point along this line: thereby altering 2 parameters.

For an N-dimensional Hilbert curve, there are always upper bounds for distance in $\mathbb{R}^N$ given a certain distance in $\mathbb{R}^1$. This locality property is illustrated in Fig. 6.4. Unfortunately, distance preservation in the opposite direction can be worse, as can be seen if the x direction is traversed at the bottom of Fig. 6.3 (d). In the application this corresponds to moving a slider a small amount, but yet jumping to a distant point on the Zoomer surface.

For a higher dimensional curve, the basic unit is an N-dimensional hypercube. Each hypercube, at each level of iteration, must have all of its corners visited exactly once, and the path must only run along the edges of the cube. Such paths are

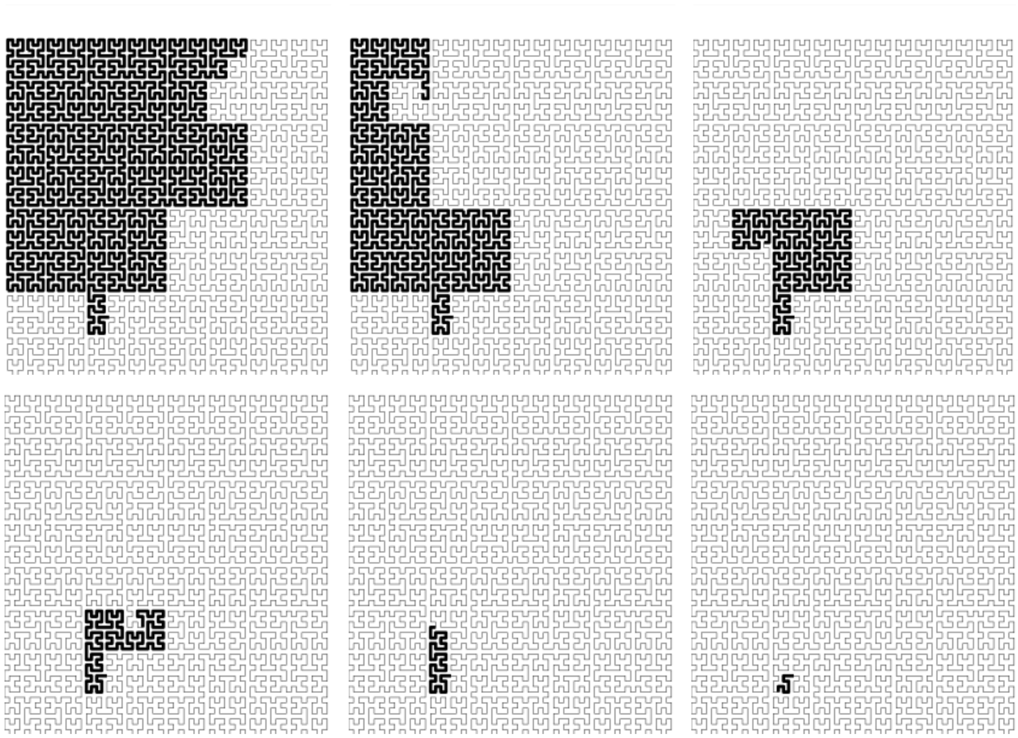


Figure 6.4: The locality property in the high dimensional space as a result of zooming into a segment of the low dimensional space. Assume, for some current zoom magnification level, the bold portion of the line is visible in some "1D view". Zooming in will shorten the line segment: this is in the progression from the top left to bottom right figures. This results in a smaller compact region of the 2D space being visible/accessible.

called "Hamiltonian paths", and if the corners of the cube are labelled with binary coordinates the sequence forms a "Gray code" (Fig. 6.5 shows this for 3 dimensions). Gray codes are binary numeral systems where only one digit changes at a time [Gray, 1953]. Fig. 6.6 (a) shows the binary numerals for five digits, and Fig. 6.6 (b) shows the standard 'reflected' Gray code.

For the purposes of parameter mapping, there is a problem with this Gray code: the rightmost bit flips sixteen times whereas the leftmost only flips once. This would be rather frustrating for the user, as the user scrolls, one parameter will flip extremely fast and another will hardly ever change. We wish to achieve some kind

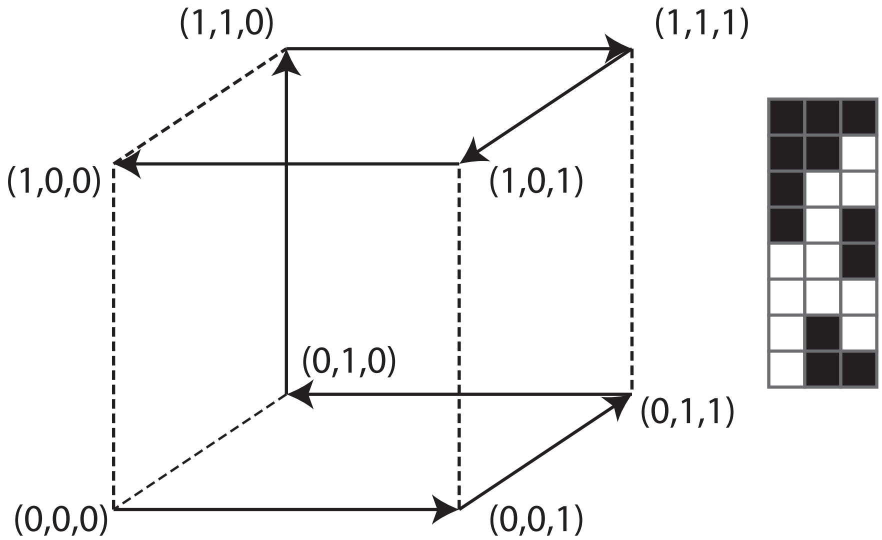


Figure 6.5: The Hamiltonian path visiting the 8 corners of a cube. The sequence of coordinates is a Gray code (right).

of balance in the speed of the changes by distributing these transitions more evenly. Gray codes with two particular properties can mitigate this issue. The first is a "balanced" Gray code [Bhat et al., 1996]. Here the transitions are spread as evenly as possible between dimensions (Fig. 6.6 (c)). The second property is "minimum run length" or MRL [Goddyn et al., 2003]. If the MRL is maximised then the bit swaps of any specific bit never occur within a certain distance of one another (Fig. 6.6 (d)). This is also desirable: when one parameter goes high we would not want it to immediately flip low again.

Beyond three dimensions there are rapidly increasing numbers of alternative Gray codes and Hilbert curves. It is beyond the scope of this paper to investigate these in detail, as their construction seems to follow no simple method. For practical purposes we assume that the long MRL code shown in Fig. 6.6 (d) will be indistinguishable from better balanced codes.

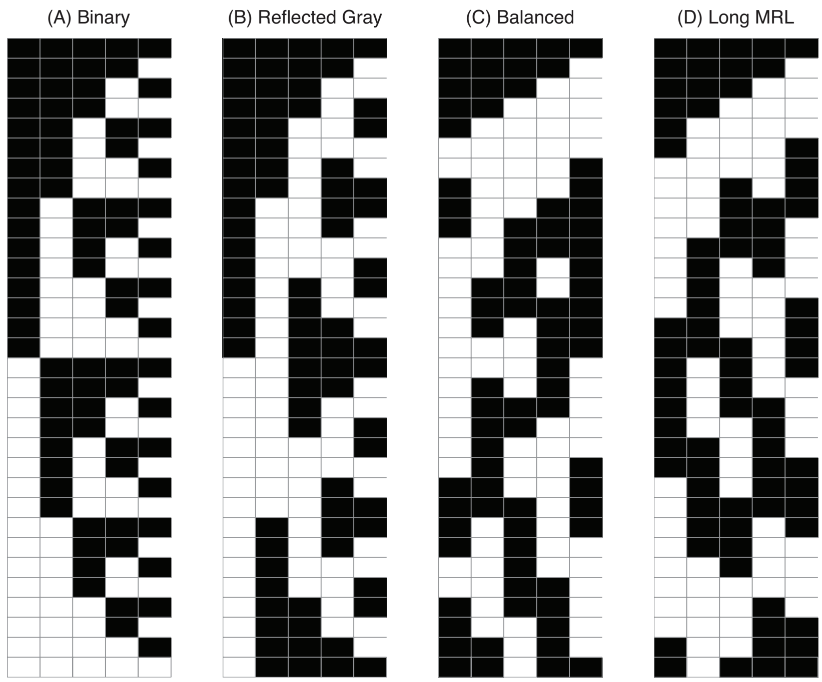


Figure 6.6: 5 digit binary codes. (a): Standard binary numbers in ascending order. (b): The Gray code: a single bit flips from row to row. (c): "Balanced" Gray code: all the digits transition either 6 or 7 times. (d): "High MRL" Gray code: The minimum distance between flips is 4 steps.

6.3.1 The Mapping Algorithm

The 2D surface is displayed as a grid, and the ten individual parameters as 1D sliders (see Fig. 6.2). Consideration of the 2-D curves in Fig. 6.3 may help imagine the result. In this case, the winding path would be traversed by a single 1D control, and cause two parameters to change: altering according to the horizontal and vertical coordinates the path visits. In the case of Sonic Zoom, each dimension of the zoomable surface uses a separate 5D Hilbert curve: moving in the x direction will change the first five sliders, and moving in the y direction will change the other five. When zoomed in fully, the smallest subdivisions of the grid become visible. These correspond to sliders changing by a single unit. Zooming out fades these low level grid lines, revealing lines corresponding to the entry and exit points of larger and larger hypercubes. When zoomed out fully, the grid divisions delineate hypercubes of 64 units per side.

The algorithm for converting an x or y coordinate to P N-bit parameters, giving $2P$ parameters for each point on the surface, is as follows:

Each coordinate is first expressed as a base-2^P N-tuple

$$a = (a_{N-1}, ..., a_1, a_0), \tag{6.1}$$

where the individual base 2^P digits are calculated like so

$$a_n = \left\lfloor \frac{x}{2^{Pn}} \right\rfloor \mod 2^P. \tag{6.2}$$

Each of the a_n are then converted to P-digit binary numbers $(b_{n,P-1}, ..., b_{n,1}, b_{n,0})$ using the Gray code $G()$, via a look-up table such as the code in Fig. 6.6 (d).

$$b_{n,p} = G(a_n) \tag{6.3}$$

Then the parameter control values c_p can be built up by treating the N different scales as standard binary digits.

$$c_p = \sum_{n=0}^{N-1} b_{n,p} 2^n \tag{6.4}$$

The number of points along one coordinate necessary for full resolution is $(2^P)^N$. In our implementation we require ten 7-bit MIDI control parameters (five per axis), so $P = 5$ and $0 \leq x < 32^7$.

6.3.2 Sub-Cube Permutations for Hilbert Curves with arbitrary Gray Codes

For the pilot study implementation [Tubb and Dixon, 2014], there was no way found to orient the 5D hypercubes such that the ends of the Hamiltonian paths always connected between adjacent cubes. This would cause large discontinuities at major divisions. If the user happened to be zoomed in to a great extent, and cross over one of the highest-level divisions, the sliders may have jumped 64 units instead of 1. There seems to be no algorithm to determine these rotations for arbitrary Gray codes. Hamilton and Rau-Chaplin [2008] provide a method for aligning N dimensional cubes in Hilbert curves, but the proof relies on the symmetry of the *reflected* Gray code in his Lemma 2.6, and therefore Theorem 2.9 relating to the intra-sub-cube dimensions, and 2.10 giving the entry points do not hold.

Extending this method by mathematical proof is beyond the scope of this research. However, as dimensionality increases there are more and more ways to construct continuous paths through the cube, therefore there are likely to be many valid rotations that satisfy any given Gray code. A brute force search was applied combined with the following three heuristics to limit the search space.

Due to the cyclical nature of Gray codes, entry and exit points of any sub-cube only change a single bit of a single dimension. This direction of travel is referred to as the "internal direction" for that sub-cube. Therefore given an N-dimensional cube and an entry point there are only $N-1$ valid exit points (ruling out $2^N - N + 1$ other corners). Sub-cube orientations for a 3D Hilbert curve are shown in Fig. 6.7.

Fig. 6.7 also shows that sub-cubes must be aligned such that entry and exit points in the super-cube are at extremal points. This is so that the super-cubes can connect to other super-cubes in the same manner as the sub-cubes, enabling recursion.

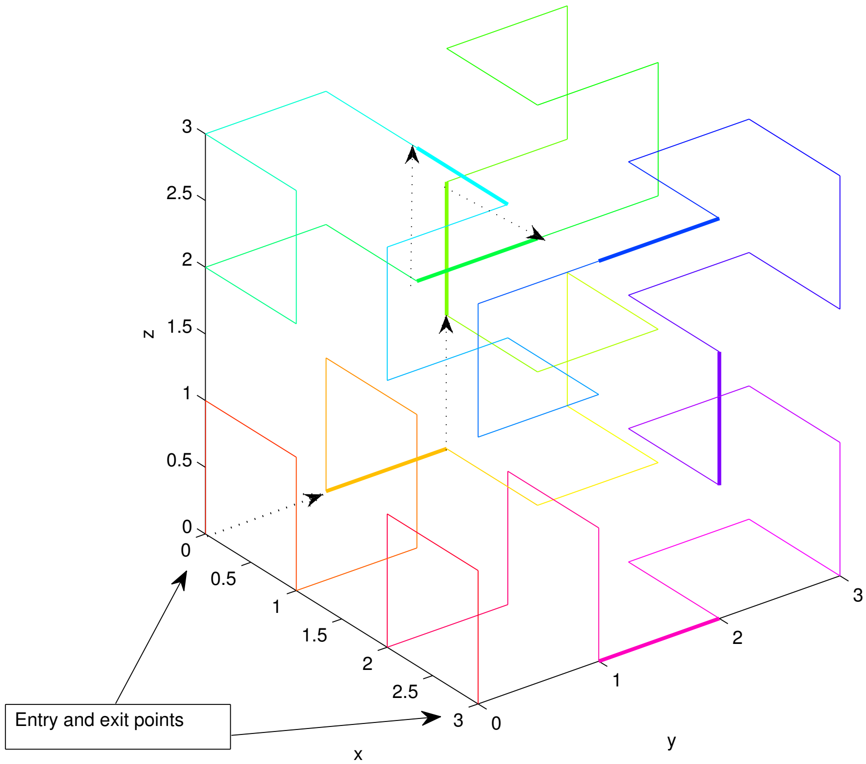


Figure 6.7: A 3-D Hilbert curve with 2 iterations. Entry and exit points of a super-cube must be in the extreme corners (bottom left). Black dotted arrows show the first four sub-cubes' internal directions. Coloured bold lines joining these black lines are the *intra*-sub-cube directions.

The final rule is that if the entry point is on the opposite side of the sub-cube from the next sub-cube, we *must* travel in that direction, otherwise the exit point cannot connect with the next cube. This can also be seen in Fig. 6.7: the first and second internal directions necessarily traverse the y and z directions respectively, whereas the third, having its entry point already adjacent to the next sub-cube, is actually free to travel in the x or z direction.

Given these three constraints, a brute force algorithm can then try a variety of super-cube Hamiltonian paths. Faced with a choice of orientation such as the ones above, a random direction is chosen. If we get to the end of the path and cannot satisfy the extremal exit point, simply start again. Of course, this is highly inefficient, but was found for 5D that after 100 to 1000 runs (a few seconds of computation time) a path was found. Repeated runs showed that there were many such paths. The specific internal direction sequence used for the released software was:

$$[2, 1, 0, 4, 0, 3, 0, 2, 2, 3, 4, 3, 0, 4, 0, 2, 2, 4, 4, 3, 1, 3, 4, 1, 1, 4, 4, 1, 1, 2, 3, 0]$$

With entry vertices:

$$[0, 0, 0, 0, 0, 5, 5, 5, 3, 3, 27, 10, 0, 5, 29, 29, 9, 9, 27, 10, 18, 20, 20, 5, 5, 3, 3, 18, 18, 20, 24, 17]$$

Where these decimal numbers are converted to 5-bit binary numbers and used as 5-D coordinates for the corners of the sub-cubes. This was for the Gray code with transition bits:

$$[2, 3, 4, 0, 2, 1, 4, 3, 2, 0, 4, 3, 2, 1, 4, 0, 2, 3, 4, 0, 2, 1, 4, 3, 2, 0, 4, 3, 2, 1, 4, 0]$$

If dimensionality was increased it is likely that this method would become intractable, in which case a depth-first search tree would be recommended [Cormen et al., 2001, p. 540-549], providing a way of back-tracking and eliminating options. It is likely that this approach too would eventually become unfeasable for even higher dimensions.

6.4 Implementation: The "Sonic Zoom" app

A scrollable, zoomable 2-D surface is well suited to implementation on a multitouch screen. Sonic Zoom is an iPad app made publicly available on the Apple App Store. Fig. 6.2 shows the interface. In the application two interfaces are provided. The first is a standard set of ten sliders (sending 7-bit MIDI continuous control (CC) values), used to control the timbre of a subtractive synthesiser. The second interface is the scrollable, zoomable surface: a map of every possible slider combination (referred to from now on as the "Zoomer").

The sliders appear as an overlay at the bottom of the screen. When both the sliders and the Zoomer are on screen together, movements with one interface are immediately reflected in the other. The "listen point" location is represented as a cross-hair in the middle of the screen. The absolute locations of touch points have no bearing on the sound. A single dragged touch point scrolls the surface: changing the coordinates of the listen point, and hence the positions of the sliders. Of course, the space is huge: in the case of ten 7-bit midi parameters, each axis contains $2^{5\times7} \simeq 10^{10}$ points. Zooming functionality is therefore essential. A two finger pinch-out gesture will zoom into an area around the listen point, whilst keeping the listen point stationary. As the user zooms, smaller sub-divisions of the grid become visible. Due to the Hilbert curve's locality properties, these smaller grid squares will correspond to smaller 10-D hypercubes which can then be explored in further detail. The sub-divisions are coloured according to their Hilbert curve iteration level. Both scrolling and zooming have momentum (i.e. the surface will keep moving at the speed the finger was travelling when it left the screen) and a small amount of friction, to enable fast navigation. The "lock sequencer" and "lock synth" buttons constrain the surface to move in only the x or y direction, respectively. Once a preset is saved, it appears as a coloured dot on the surface. If the listen point moves near to a preset, it will snap to the preset coordinates. Without snapping it is impossible to line the preset up precisely.

On opening the application, participants were simply instructed to search for sounds they liked, or thought were useful or interesting. They were told to make sure to save favourites as presets. This introductory text is given in Appendix B.

The different interfaces were presented individually and in combination for 5 minutes each, in a randomised order. After the timed session a questionnaire was presented, and on completion further features were unlocked: such as the ability to show and hide the two interfaces, and MIDI connectivity. Users agreed to a statement of consent before their interactions were logged.

One important point about doing research via an App Store distribution model, is that it is critical that the implementation and the data collection is done correctly first time. Fig. 6.8 shows the number of events generated by day. Clearly most activity takes place within the first week, so if an update is necessary, due to a bug or unanticipated problem, it will then take another week to get the app approved, and vital user data would be lost. This initial flurry is possibly due to appearance on the "what's new" part of the service, but also the rapid spread of the news of the release through a small community of keen iPad musicians.

6.4.1 Extensions After the Pilot Study

During a pilot study (14 participants), many users suggested smoother transitions, both in a general sense, and between specific presets. Therefore a 'smooth' mode was implemented that, rather than exposing the full 7-iteration Hilbert curve, interpolated between the different iterations. This made it possible to move between various complexity levels: detail will only reveal itself as one zooms in. The drawback of this mode was that the vast majority of the saved locations could not be

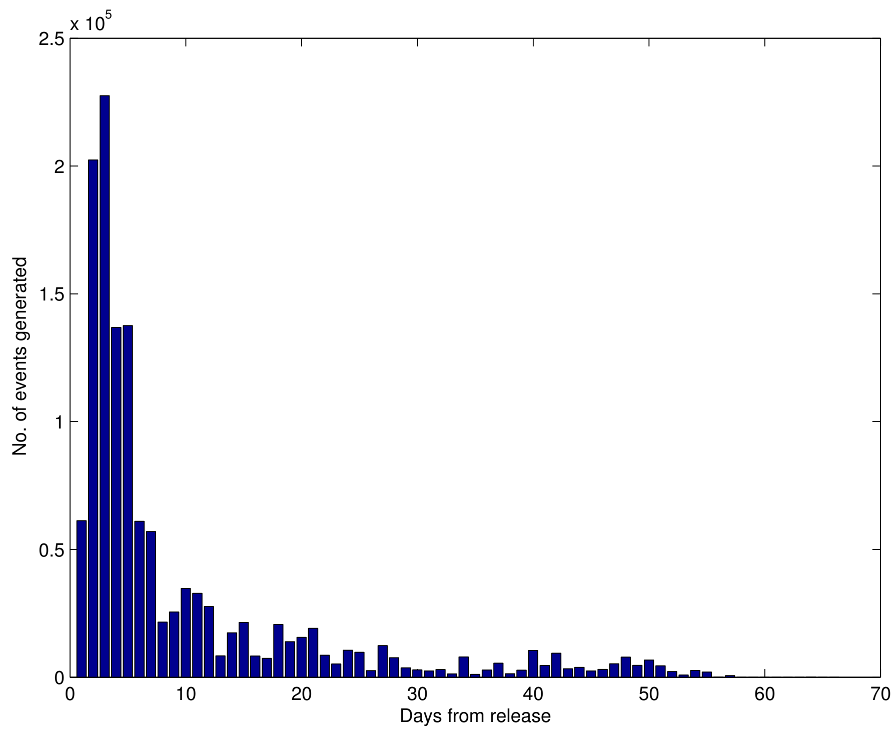


Figure 6.8: Number of interaction log events generated by users per day.

seen on the surface.

There were frequent requests for an easier way to return to where you had visited before, but had not saved. It is too easy to accidentally overshoot something that caught your ear. Two users came up with the intriguing suggestion of a visual, scrub-able undo trail for these situations. Session histories can promote new ways of thinking by providing an overview of one's own creative process [Shneiderman, 2000]. This was implemented for the public release. Points that had been listened to before showed up as blue circles, the diameter of the circle indicating how long you had listened to that point. The crosshairs would snap to these 'evaluation points' similarly to the presets, so that previously evaluation points could be returned to easily.

A featureless grid was not enough to make use of visual spatial memory. Colour coded lines were not sufficient for people to instantly grasp what scale they were at, though this can be learnt with use. One user suggested using audio feature extraction techniques to create a texture that would convey the nature of the sounds underneath. Alternatively, patterns or shapes could be associated with presets and then morphed, as in [Van Wijk and van Overveld, 2003].

In the pilot study, five of the interface's ten parameters were used for a simple melodic pattern generator. This sequencer creates a 16-step sequence based on five sine waves of integer frequencies; this "Frequencer" is detailed in Tubb [2015]. However most participants felt these parameters were confusing, so for the public release, the 10 parameters only controlled timbre. The melodic sequences were randomly selected from those saved in the pilot study. Buttons for play, pause and skip to next sequence were provided (see top of Fig. 6.2), but the sequences themselves were not editable, so as to restrict interaction to timbre adjustments.

In summary, from the conclusions and suggestions from the pilot study, a number of changes were implemented:

- The sequencer controls were removed and five new timbre controls were added.
- The sequencer now just played saved sequences from the pilot study, these could be skipped through.
- The user's path across the grid surface was shown as a white line, and previously evaluated points (hovered over for more than 300ms) were shown and could be snapped to. This enabled easy reversal of actions.
- Double tapping on a preset will move the listen point to that preset, and over one second smoothly interpolate all parameters.
- Sliders were moved to the bottom of the screen and made iOS generic..

- As a reward for completing the experiment, users were allowed to swap interface type, and an interpolation mode was implemented. This mode truncated the Hilbert curve at the current zoom level, creating a much smoother surface. A further interpolation was employed between levels when zooming in.

6.5 Results

At the time of writing, the app has been available for over two years, and the total number of downloads now stands at 1970. The number of successfully completed experiments is 384, with a total interaction time of over 100 hours. The number of valid questionnaire responses was 361. Unfortunately, there were over 1000 started but uncompleted experiments. This may have been due to the inconvenience of finding a solid 15 minutes to participate, or may indicate a lack of interest in the application. It is important to note that in the absence of supervision, only those who are actually interested in such an interface will complete the experiment. This may bias the questionnaire results in favour of the Zoomer.

6.5.1 Questionnaire

Tables 6.1 and 6.2 show the questions asked at the end of the timed sessions. The former (question numbers prefixed by AD) required Likert style agree-disagree responses, the latter (question numbers prefixed by SZ) required respondents to rate how strongly different interfaces were preferred for various tasks. Fig. 6.9 and 6.10 show the results as diverging stacked bar charts [Robbins and Heiberger, 2011]. Results where the user had answered every question identically were discarded.

Most respondents were clearly very familiar with electronic music (AD 1). The participants self select, so some bias in favour of novel interfaces can be expected. Highly positive responses to this application include the ability to see the presets as

AD 1	I am familiar with music software and sound synthesis.
AD 2	The ability to retrace my steps using the history path was useful.
AD 3	The correspondence between the sliders and the grid was understandable.
AD 4	Scrolling a greater distance on the grid seemed to correspond to larger difference in the sound.
AD 5	The ability to see other presets laid on the grid was useful.
AD 6	The range of sounds was too limited/poor quality to be able to judge the eventual usefulness of the interface.
AD 7	The Zoomer was an improvement on just using a randomiser.
AD 8	The combination of Zoomer and Sliders was better than either individually.
AD 9	I enjoy "happy accidents" in the creative process

Table 6.1: Questions requiring a 5 point agree/disagree answer.

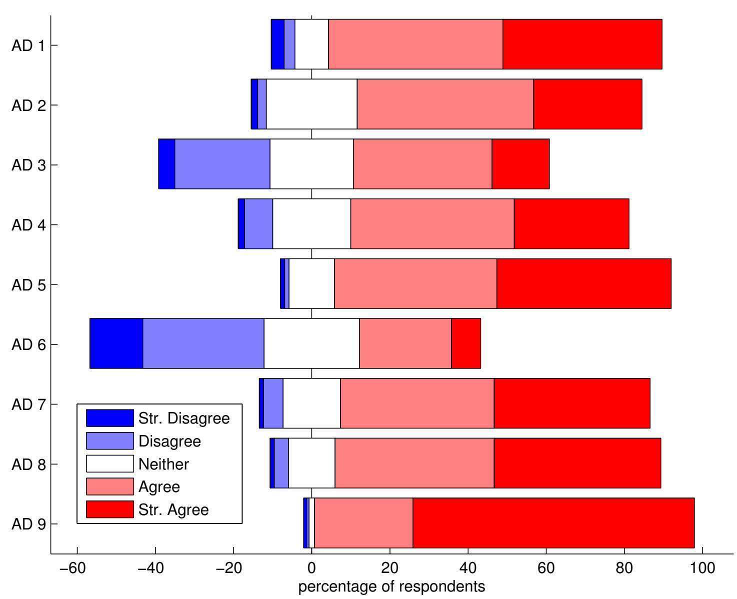


Figure 6.9: Questionnaire responses to agree/disagree (AD) Likert items. Neutral response is centred.

SZ 10	The best interface for discovering interesting sounds quickly was...
SZ 11	The best interface for fine tuning a sound was...
SZ 12	Interface that I felt more in control using...
SZ 13	The interface that felt more creative was...
SZ 14	Interface better for generating new ideas...
SZ 15	Interface better for performing live would be...
SZ 16	Overall, the interface I preferred using was...

Table 6.2: Questions requiring a 5 point sliders vs. Zoomer answer.

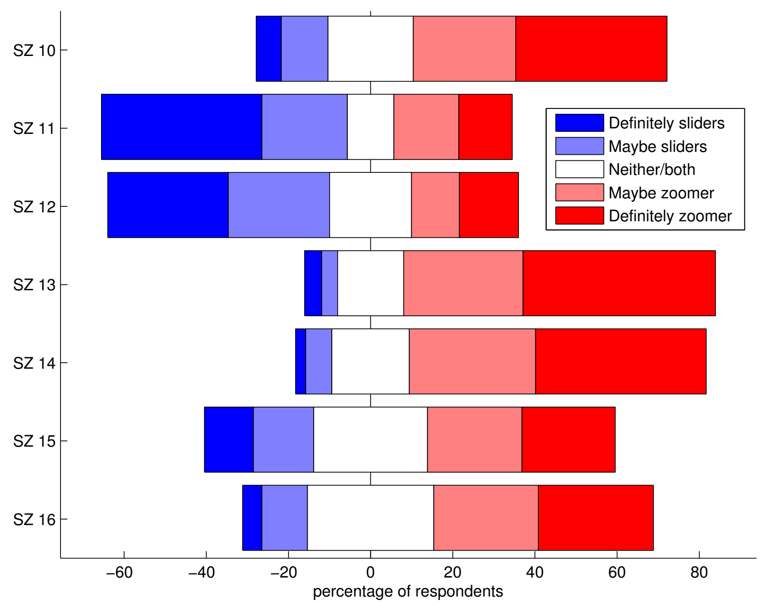


Figure 6.10: Interface preference responses. Blue bars indicate preference for sliders, red for the Zoomer. "No preference" is centred. 11 and 12 reveal slider preference for convergent properties, 10 and 14 show Zoomer preference for divergent.

points in space, and to see the "undo" path (AD 2 and 5). The question of whether the mapping was understandable was less conclusive (AD 3), but most users did get a sense of the locality property (AD 4). The Zoomer was deemed considerably more useful than a simple randomiser (AD 7). The strongest response of all (albeit to a heavily loaded question!) was that people highly value happy accidents in the creative process (AD 9).

Particularly of interest was the hypothesis that sliders would be preferred for convergent tasks and the Zoomer preferred for divergent. Responses to SZ 10 and 14 (divergent aspects) contrast sharply with SZ 11 and 12 (convergent aspects). There was a large significant difference between the means of these two properties (difference $= -2.6, t(768) = 3.2, p < 0.01$), confirming this hypothesis. Most participants felt that the Zoomer was the more creative (SZ 13) which may reflect the popular identification of creativity with novelty and divergent thinking, or simply the fact that new experiences with novel technologies can be inspiring in themselves. The Zoomer was slightly favoured as an interface for performing live (SZ 15). This is surprising considering its unpredictability. Nevertheless, at moderate zoom levels, interesting variations can be performed that are always kept within certain bounds, and a visible cluster of presets can make revisiting regions known to be performable easy.

6.5.2 Sound Discovery Rates

Interface	Sliders	Combination	Zoomer
No. saves (timed)	693	630	762
No. saves (free)	15	505	103

Table 6.3: Total number of presets saved for the three interface views, during timed stages and after the completed experiment. The Zoomer proved most prolific.

One hypothesis was that if more presets were saved in a particular mode, it might indicate that this interface was best for locating good sounds quickly. The total numbers of presets saved in each different session are shown in Table 6.3. The upper row of values show the totals when the users spent 5 minutes on each interface, the lower shows the number of saves during the subsequent free-use period. For timed sessions, the most presets were saved in the Zoomer-only mode, indicating that this may have been the fastest interface for sound discovery. However these results are not statistically significant (2 sample t-test, $t(384) = 0.34, p > 0.05$), as the number of saves per user is rather low and highly variable. Greater incentive to find as many sounds as possible may have improved the experiment in this regard. The large number of saves in the combination interface after the experiment reveal that people much preferred the combination, given the choice. Unfortunately, the fact that the combination interface has the lowest number of sound discoveries in the timed session seemingly contradicts the participants' preference for the combination, and rather undermines save rate as a reliable measure of effectiveness.

6.5.3 Interface Preference for Divergent and Convergent Traversal

Was divergent or convergent behaviour detectable from the interaction data? One indication of this was the average zoom level at which people scrolled around compared to the average zoom level at which they saved a preset. The hypothesis would be that people zoomed in to hone the sound before saving. Therefore the prediction was that the average level for scrolling would be higher than the average level at which they saved a preset. This was indeed the case, although the difference was small. The total amount of time users spent scrolling at 7 different zoom scales is shown in Fig. 6.11. The zoom levels are the logarithm of the scale factor rounded to the nearest integer. Data from before the zoom functionality was first used (i.e. the default zoom level when the app loaded) were omitted from the calculation. Users showed a clear preference for larger scales, despite the unpredictable timbre changes: they spent 200 minutes in total scrolling at the largest scale (where sliders change by 64 MIDI CC units per grid division), and only 30 minutes at the lowest scale (1 MIDI CC unit per division). The mean scroll and save levels are marked.

With both Zoomer and sliders present, there was less of a tendency to zoom

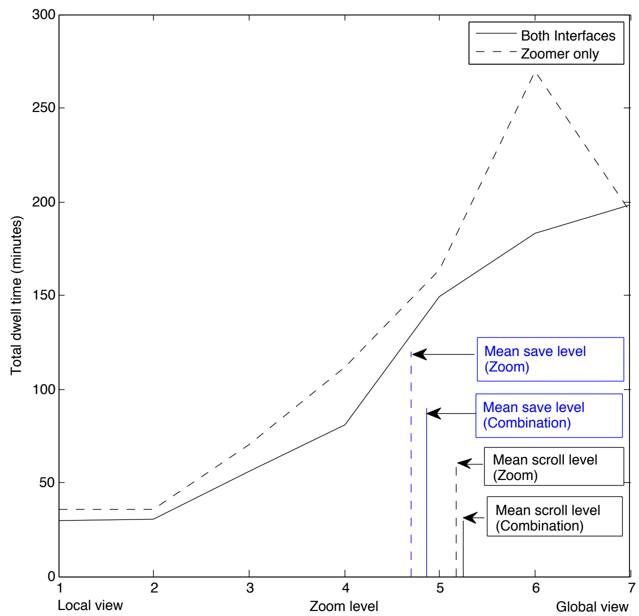


Figure 6.11: Histogram of the time users spent at each of seven zoom levels. Vertical lines show the means of zoom levels when presets were saved and when scrolling the grid for both the Zoomer and Combination stages. These values show that people would zoom in before saving a preset. However with the sliders also present, this tendency was reduced slightly.

in in order to save a sound: this indicated that when the sliders were present the Zoomer's convergent functionality was eschewed in favour of convergence using separate parameters. However, this difference between mean zoom level in Zoomer-only and combined mode was not significant (2 sample t-test, $t(384) = 1.18, p = 0.12$).

A far clearer asymmetry between the two interfaces is seen when both interfaces were on screen, by investigating which interface was being used immediately before and after saving a preset. The hypothesis was that users would exhibit a repeating diverge - converge - save approach, therefore the interface used immediately after saving would be the one preferred for exploration, and the one immediately prior would be the one preferred for honing. Table 6.4 shows the results, indicating that people were about five times more likely to follow a Sliders - Save - Zoomer pattern than the reverse, strongly supporting this hypothesis. It is hard to confirm actual divergent or convergent behaviour by analysing the search trajectories, as they are hugely different for the different interfaces[1]. Path properties are easier to analyse for a single interface: Fig. 6.12 shows that, for the Zoomer only sessions, the average speed of scrolling tends to reduce by about a factor of two as the user converges on a saved point. Evidence for convergent behaviour using the sliders is given in section 6.5.5.

Another point to note is that presets were around five times more likely to be saved during Zoomer manipulation than during Slider manipulation when both were on-screen. It could be argued that this merely shows an overall preference for using the Zoomer, but the total interaction time was only 2:1 in favour of the Zoomer. Overall, engaging with the Hilbert mapping seems more conducive to discovering and saving sounds.

6.5.4 Distribution of Saved Presets and Evaluation Points

Many saved presets (4006) begin to make possible the analysis of the distribution of favourites in the parameter space. Fig. 6.13 shows these presets, arrayed as they would be on the 2D zoomer, and coloured according to whether they were saved

[1] The average parameter space distance-per-event was 60 CC for the Zoomer and 10 CC for the sliders, so obviously the Zoomer was intrinsically more rapid, random and divergent.

Before Save	Zoomer	Sliders	Zoomer	Sliders
After save	Zoomer	Sliders	Sliders	Zoomer
Total	1020	175	93	446

Table 6.4: Which interface was used immediately before and after saving favourites. This includes free interaction after the timed sessions. Note the difference between the last two figures, showing a strong asymmetry as to which interface was used in the early and late stages of a search. By far the most common sequence was Zoomer-Save-Zoomer, reflecting the overall popularity of this interface.

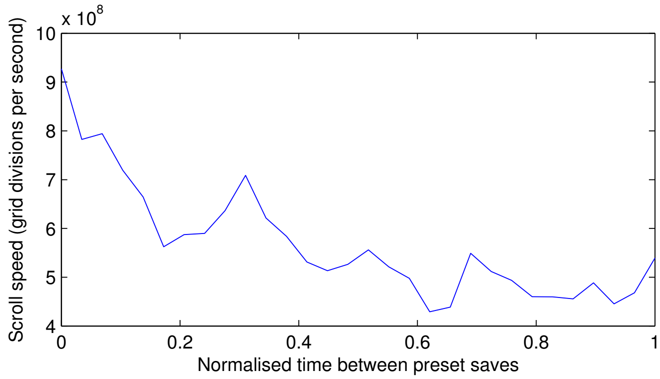


Figure 6.12: Scrolling speed between preset saves, averaged across all pairs of consecutive saves. Time is normalised such that the time of the previous save $t = 0$ and time of next save $t = 1$.

in slider, zoomer or combination mode. It is clear that the presets are not evenly distributed. Two sparser horizontal bands show up, these correspond to regions of the space that were likely to generate no sound whatsoever, due to the combination of filter parameters. For instance if a high pass filter was set above the range of human hearing with no envelope amount to sweep it down. It is worth noting that these bands occur in all three interface modes, indicating that this distribution is interface-independent.

One could argue that the presence of very sparse regions indicates a 'sub-optimal' synthesiser. Large areas of uninteresting sounds will have a negative effect on the effectiveness of random exploration. This distribution could be used to either: re-design the parameters so that unusable settings are not achievable at all, or to warp the space so that the likelihood of finding a 'good' sound is equal over the whole

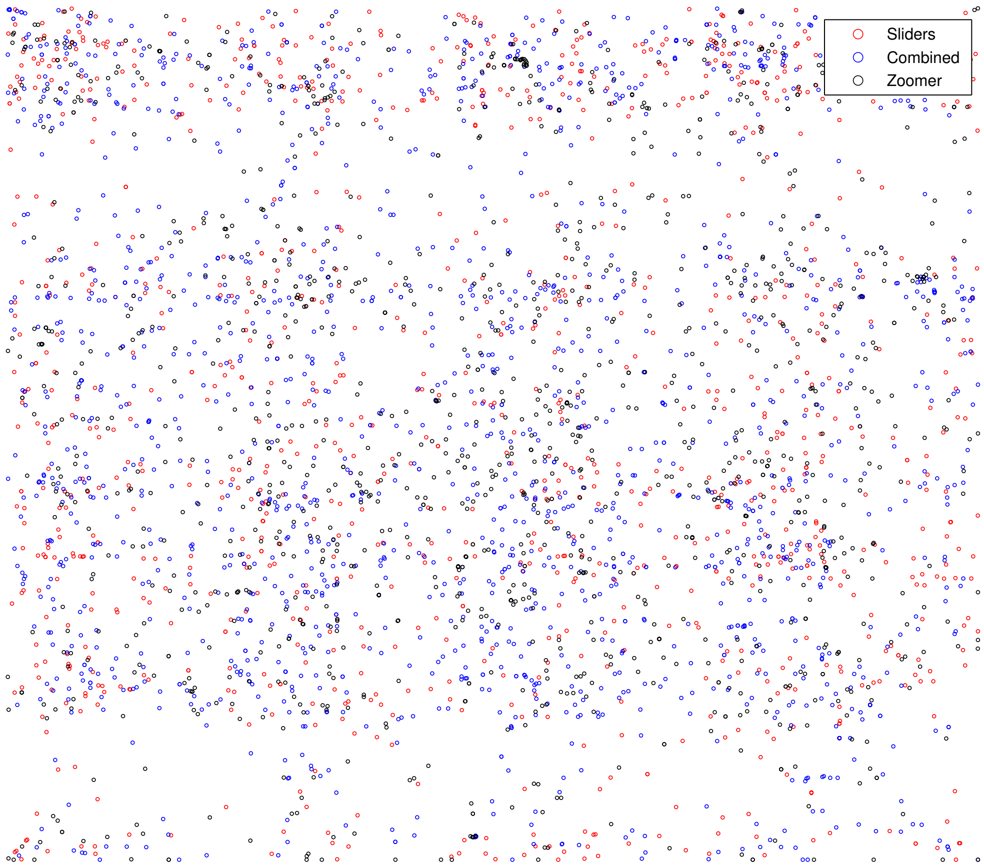


Figure 6.13: All 4006 saved presets for all users and all interface stages. The arrangement is exactly as they appear on the 2-D surface.

space. This would be a similar approach to that of [Loviscach, 2008]. However there may be interesting oases of novel sounds in these deserted regions.

The distribution of "evaluation points" was obtained (Fig. 6.14). A total of around 50,000 evaluation points were recorded in the successfully completed experiments, and 122,000 overall. These show an extremely similar distribution to the presets, but with far greater resolution due to more data points. This is probably due to the tendency of people to be zoomed in around the location of an existing preset, as can be seen with the 3 "factory" presets that were provided to everyone in the application, which have dense clusters surrounding them. Further inspection (not shown) showed that any perceived clustering is due to individual users exploring a small area in detail, rather than that area being particularly popular with all users.

Evaluation points for the slider mode were obtained in a similar fashion as the Zoomer by noting a setting that has not been changed for 300ms. Note that given the smoother nature of the sliders it is easier to evaluate the sound whilst moving them (on account of their linearity and predictability).

Distributions can also be taken by individual parameter. Histograms of slider settings for saved presets are shown in Fig. 6.15. The sliders tend to get set to the end points; this makes sense for some parameters that have particularly useful extremal values. For example reverb may often be set to completely dry, but for other parameters this might indicate an insufficient range for that parameter.

Whilst there are some patterns in these distributions, the chances of designing radically more effective instruments by means of analysing the positions of these points and extracting what constitutes a 'good' sound seems slim. It is likely that more sophisticated analysis of the actual audio output from these points would be needed to attempt this.

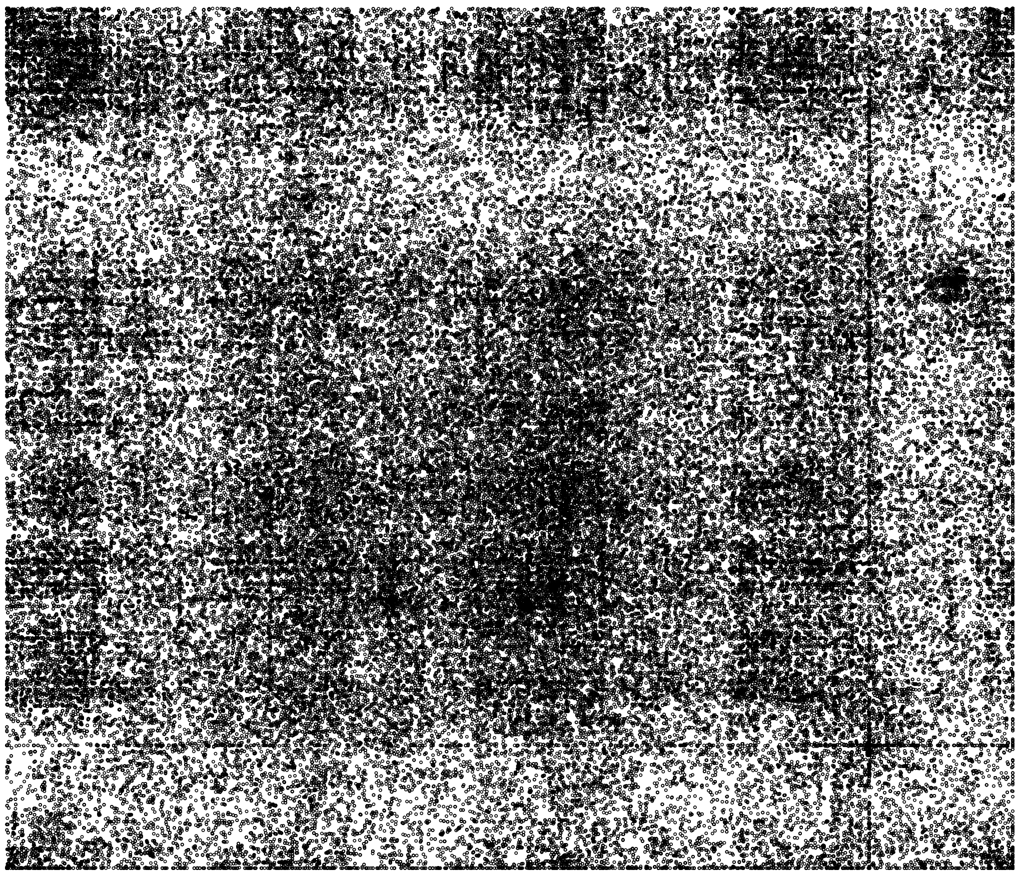


Figure 6.14: Distribution of all users evaluation points, combined interface stage. Dark regions are presumably sounds of more interest to the participants, light regions less interesting. The distribution is similar to the saved presets.

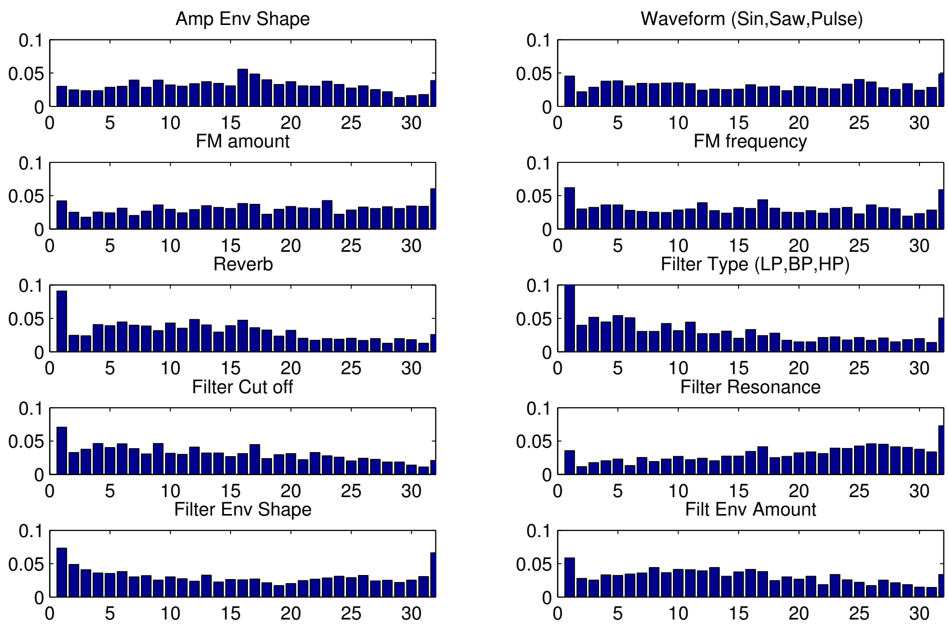


Figure 6.15: Histograms of slider settings for all saved presets.

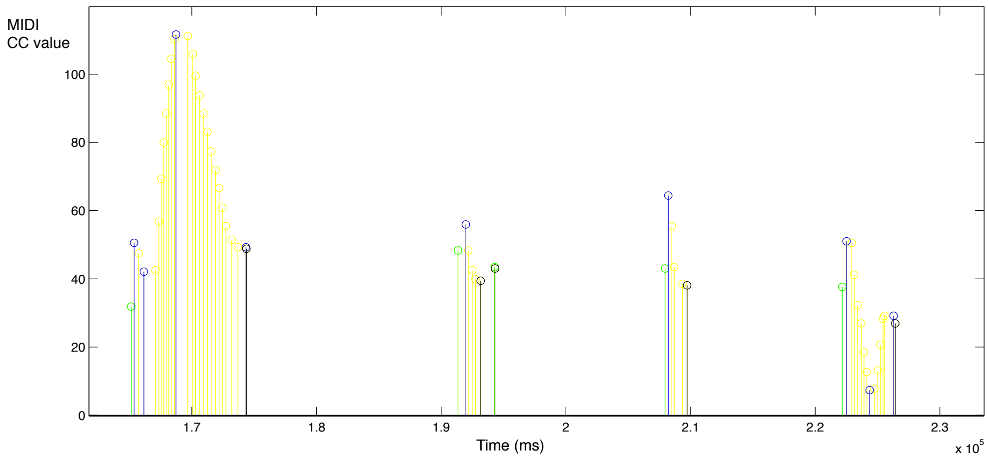


Figure 6.16: Example of how an interaction with a single slider is divided into blocks and adjustments. Yellow stems are the raw slider events. Green indicates start of block, blue indicates a change in direction, and black indicates the end of a block.

6.5.5 Analysis of Slider Adjustments

To analyse navigation strategies in the slider-only case, slider movement data were separated into discrete “interaction blocks”. This was defined as a series of movement events on a single slider. Blocks were defined as being separated by a swap to another slider, or a gap of at least a second. These blocks were then divided into “adjustments” by segmenting them according to changes in direction. Fig. 6.16 shows how an example of how a series of slider events is grouped into blocks and adjustments. Slider values run from 0 to 127 CC units.

The first question is: how many adjustments does it take before people are happy with their engagement with a particular slider? Fig. 6.17 shows this, broken down by each parameter. A single movement or one double-back are the most common, with higher numbers falling off more or less exponentially. An interesting thing occurs in the tail of this distribution (Fig. 6.18). It does not fall off towards high values as much as expected. Very high numbers of adjustments probably don’t indicate that it took more than ten adjustments to find a sound, they probably indicate that the user was just playing in an instrumental, expressive fashion with these parameters. The way these instrumental interactions break down by different parameter type is revealing, if not especially surprising. For low numbers of consecutive adjustments, interactions are fairly evenly distributed (Fig. 6.17). By contrast, for high numbers of adjustments the distribution is not even. Filter cut off is the most popular, followed by FM amount and filter envelope amount. Filter type (which morphed between low-pass band-pass and high-pass) is least popular, followed by filter resonance and reverb amount. This result has a fairly obvious interpretation in terms of the "standard" synth parameters that people may want to perform in real-time. However, one could imagine that given a novel synthesis technique, and the question of which parameters would be best for gestural control, by observing this type of user behaviour, one could obtain an indication as to which controls people feel are most 'performable'.

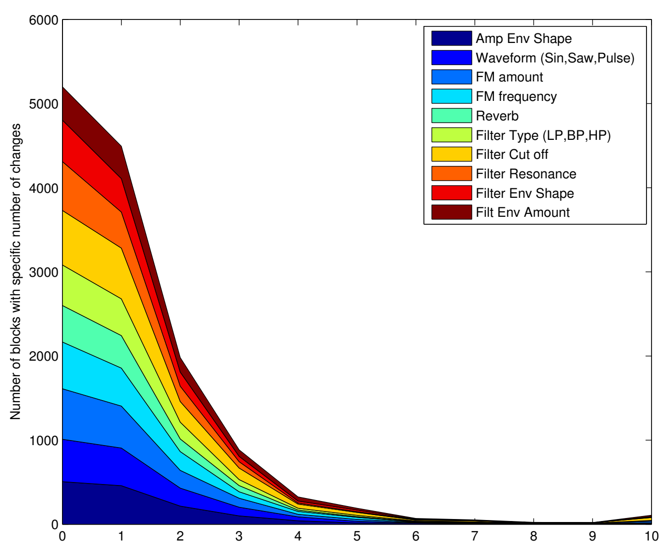


Figure 6.17: Histogram of number of blocks with different numbers of direction changes.

Did slider use show any indication of convergent optimisation-style behaviour?

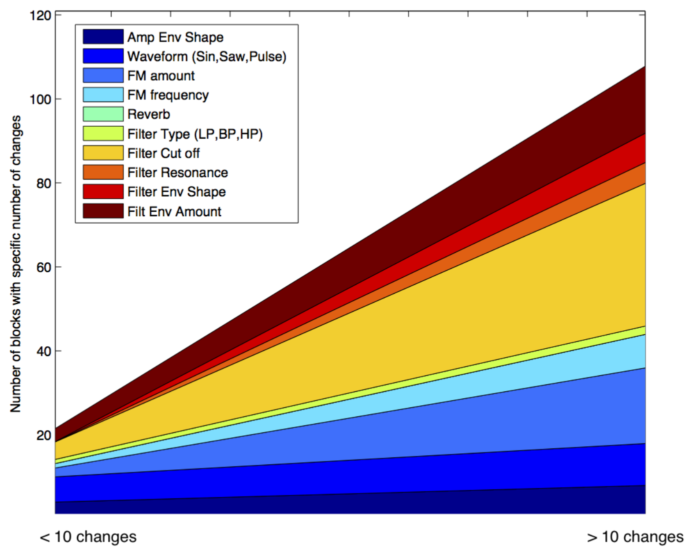


Figure 6.18: Histogram of number of blocks with different numbers of direction changes greater than 9. Detail showing that large numbers of consecutive reversals only occur with those parameters typically associated with performance-style interaction, for example filter cutoff.

Fig. 6.19 shows the absolute size of consecutive adjustments in the case where 3 direction changes were made before the interaction with this slider was ended. The different distribution shapes do seem to indicate different interaction stages. From the second movement onwards, the average sizes of adjustments get progressively smaller, rather similar to gradient descent algorithm exhibiting some overshoot. A tentative model to explain this would consist of three basic stages:

1. *Initial effect query*: the humped distribution, and the fact that there are hardly any small adjustments indicates that this might be just an exploratory enquiry: "what happens when I move this?".

2. *Exploratory scan*: the second movement possesses a fairly uniform distribution, indicating that all sizes of change are equally likely. Given that moving the entire length of the slider is less likely a priori (if the starting position is uniform), this might indicate that the previous move takes on to an extreme

value and now people are scanning the length of the slider. This stage usually takes more time (See the middle distribution in Fig. 6.20).

3. *Honing*: there can be several of these stages becoming smaller and more focused on the eventual preference. The last movement is usually very small ($M < 30$CC) indicating that the desired setting has just been found, but has been overshot somewhat.

This tendency to progress toward smaller adjustments can be seen for all interaction blocks, up to around 7 direction changes. So, one could speculate that slider interaction may reveal a smaller, one dimensional microcosm of divergent and convergent behaviour, with an initial exploratory scan of the length of the control, followed by convergence on the favoured setting. Most of the time however, little exploration is necessary with a slider: the most common number of direction changes is one or none.

Of course, investigating the actual slider plots such as Fig. 6.16 we see that this is far from being the "rule". Notwithstanding this the presence of these distributions in all blocks from 2 to 7 adjustments long indicates that it is a definite pattern, and appears to confirm that behaviour using sliders is convergence oriented.

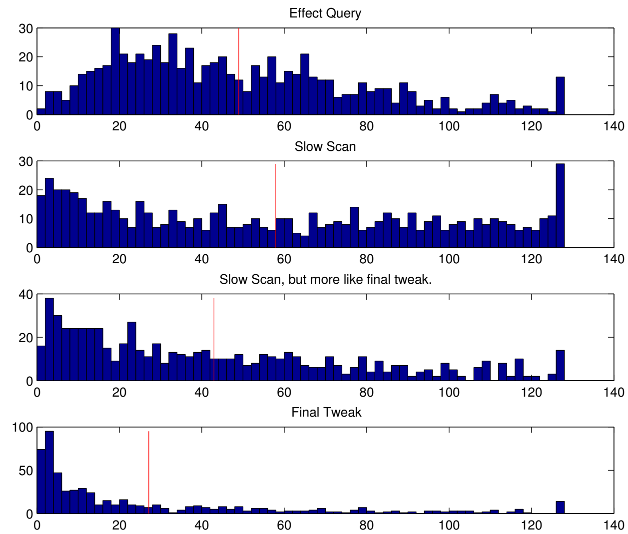


Figure 6.19: Histogram of the size of consecutive slider adjustments, in the case of 3 changes of direction (first top, last bottom). Thin vertical line shows the mean.

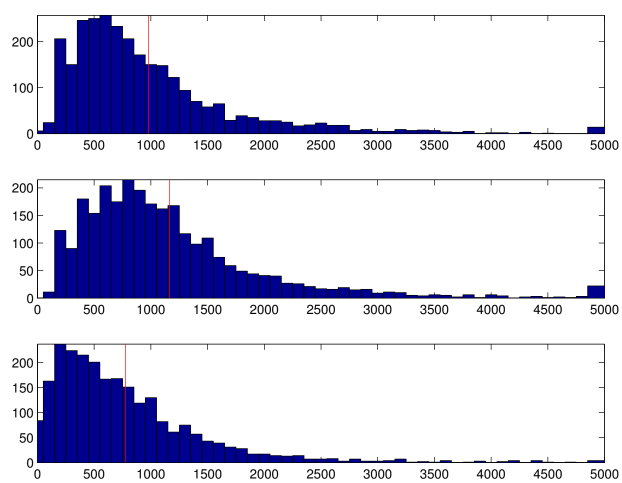


Figure 6.20: Histogram of the time taken for consecutive slider adjustments, in the case of 2 changes of direction (first top, last bottom). Red vertical line shows the mean. The second adjustment takes the most time, hypothesised to be a slow exploratory scan of a range of values.

6.5.6 User Comments

It should be noted that participants were aware of the focus being on "creativity" so many of the comments to the effect that this was a "creative" interface could have been primed by the introductory text. Users were told in the instructions that zooming in was meant to "hone in on the sound", but not that the hypothesis was that the sliders would be preferred for fine tuning.

Some users enjoyed the zoomer for both divergent and convergent strategies:

> "The zoomer interface was really great for being able to try out different sounds and then be able to hone in on variations. I've used a lot of music apps and haven't run across any that enable you to do this sort of creative exploration. The zoom really allows you to fine tune and focus in on a particular sound in a way that sliders or other input controls can't do."

> "I have many, many iOS synths, and this interface is the best for exploration and creativity, and then dialing it in."

One user mentions disillusionment with standard interfaces that imitate classic devices:

> "I think I'm going to love it. I really found regular emulations of synths increasingly discouraging in order to find new sounds and expressions."

There were specific mentions of importance of using a combination of both interfaces:

> "I will use combined mode from now on as the experience with both interfaces brought completely different ways to find and manipulate sounds."

> “Nice to see new experiments in sound creation, but any useful synthesis must rely on both methods for fine tuning sounds or for small changes if using it in a live performance”

The strength of combining the two interfaces seemed to pursuade this user that a randomiser was crucial when the zoomer wasn’t present:

> “At one point in the Zoomer I felt I was close to the sound I wanted but not having the slider made me feel like I couldn’t experiment with exactly what I wanted. The randomizer button was crucial for just using the slider to find creative sounds, just having the Zoomer only really helped when looking for something that catches your attention and then refining on a macro level.”

Some appreciated the very tiny changes available when using high zoom levels:

> “I especially enjoyed being able to go hi-res, so as to change the sound very, very slowly.”

> “The Zoomer is brilliant. Being able go that deep was good.”

> “Amazing when you can zoom so deep into the sound to edit.”

Some comments relate to parts of the EARS theory that were not mentioned in the help text. This remark reveals what was discussed in Section 3.6, but also perhaps reveals the subjectivity of what constitutes a constraint:

> I’ve currently moved from digital to analog synths for the greater feeling of control, and feel that some restrictions are useful or even required for creativity. Less is more and stricter bounds make one work harder and more innovatively to break them. Thus my preference of the slider

> interface, but on the other hand I love the simple and quick intuitiveness of touch zoom interface.

Some people specifically enjoyed the unintentional result of the sounds that emerged when moving across the surface:

> "My favourite sounds came from moving the cursor around on the grid, not when it is sitting in one spot."

> "I really enjoyed the sound created when moving the curser in zoom mode ie the sounds that were generated in between the stationary points."

This user was not experienced, and does not make music, but still enjoyed the exploratory process:

> "I have no musical training so am hard pressed to make anything that resembles a song. Yet, as a hobby, I am immensely entertained just coming up with new sounds. So this experiment app was extremely fun for me... I found the grid to be more for larger changes in sound, and then use the sliders to fine tune."

There were a number of more negative comments. Many comments highlighted how important connectivity is for computer musicians: e.g. requests for MIDI (8 mentions) and Audiobus[2] (ten mentions). If users cannot easily integrate a tool into their existing work-flow, with synchronisation and audio transfer capabilities, its utility is severely compromised. There were also more than 13 complaints/requests for a sequencer. However all these features would have interfered with the experiment, as they would have distracted users from timbre adjustments. One user suggested using two zoomers, one for the sequence and one for the synth timbre.

[2]Audiobus allows iOS to transfer audio and MIDI between different music apps.

As anticipated, many (ten) comments featured complaints about the unpredictability and inconsistency of the surface. Chiming with results across HCI, perceived control is important:

> "The grid seems too random"

> "No obvious correspondence between movement on the graph and sound changes. It all felt random, which is ok, but precludes using it as an instrument in its own right, which would need more predictable and useful changes which would be learnable."

> "Zoomer was creative and good for inspiration but a bit unpredictable"

> "Only at the deep zoom level could I hear a change and associate the change with my hand motion."

> "The zoomer doesn't feel like it maps a space except locally, and there the mapping feels like it doesn't mean the same as it does elsewhere in the grid. I would have preferred finding a happy accident with the zoomer and then be able to hunt around in parameter space with the axes corresponding more directly to parameters in a consistent way."

This next comment indicates that a plain grid was not enough to make use of spatial memory:

> "The mapping of the parameters into the 2d plane seemed very arbitrary, so it was difficult to get any real grip on the process. Doing the same thing to a greater number of variables would likely exacerbate the problem. The zoomer was slightly interesting when the sliders were also present, but on its own was just too featureless and abstracted."

There were many requests for the recording, playback and looping of movements across the 2D surface, indicating that certain paths produced interesting dynamic changes and that meta-control and abstraction methods would be useful.

6.6 Conclusion: No Need to Leave Serendipity to Chance

This study strongly indicates that different ways of navigating parameter space are suited to different stages of the creative process. For exploration the Zoomer was preferred, for fine tuning the sliders were preferred. Users responded very positively to the assertion that the combination of the two interfaces was better than either individually. This indicates that being able to alternate navigation styles is valuable.

In terms of the EARS model, we have compared interfaces presumed to suit the exploratory and algorithmic modes, and found that they were used roughly as expected. The strongest result from the interaction data was that, when a predictable one-to-one mapping interface is combined with an unpredictable, exploratory interface, a clear asymmetry in interface preference is seen before and after the locating of favourites. Users were five times more likely to be using a "Zoomer, then sliders, then save" strategy than the reverse. This asymmetry seems well explained with reference to divergent and convergent search strategies.

Whilst there are many experimental variables at play in the comparison of these interfaces, there seems to be a clear link between the predictability of a mapping and its use for the differing creative stages.

The fact that people tended to use very high zoom levels (i.e. they were zoomed out to levels where movements had large effects on the sound) indicates that during the exploratory phase, an overview of the entire space is a very useful feature, but access to every level of detail at this stage is not. Given that users prefer to fine-tune with the sliders, we could eliminate access to some of parameter space with no noticeable ill effect. Further testing should be done to establish what users feel is the ideal scale, but certainly the lowest two levels seem disposable: leading to at least a millionfold reduction in the area of the surface.

Combining this finding with the questionnaire feedback we can claim the following:

- Divergent exploration and convergent honing behaviour can be detected in interaction logs.
- Different parameter navigation strategies are suited to different stages of the creative process.
- Users will naturally gravitate toward the most suitable interface for these strategies, given a choice.
- The ability to switch between navigation styles is important.

So, even in an uncontrolled experiment such as this, some clues as to musicians' creative processes can be obtained.

The Hilbert curve is far from an ideal instrumental mapping *per se*, due to its unpredictability, lack of smoothness and nonlinearity. It can however be very useful in cases where a low dimensional representation of a complete parameter space is desired, and was shown to be preferred to a randomiser. Even with more sophisticated dimensional reduction techniques such as MDS and SOMs, collapsing a 10D space down to 2D will cause some unpredictable twists in the subspace. This experiment shows that this is not necessarily a bad thing. One advantage of the Hilbert approach over these other methods is that it is extremely cheap computationally[3], and requires no audio rendering or analysis time. Further work could directly compare these approaches to generating 2D surfaces: the collected preset points from all the participants could be used as input to an MDS or SOM algorithm to generate another surface, provided as a new interaction mode in the application.

The use of a balanced Gray codes greatly improves the usability of the Hilbert mapping. Problems can be expected when trying to scale this method up to higher dimensionality, however. Doubling the dimensionality to 10D per axis may be feasable, but beyond this finding balanced Gray codes and their associated sub-cube orientations becomes considerably harder. The size of the scrollable surface may become prohibitive. Given that many soft-synths possess over 100 parameters, this is certainly not an easily applicable technique for a general commercial synth, unless some subset of the parameters are chosen—which would rather defeat the object of the exercise. One solution to the exponentially increasing size of the space may be abandoning the idea of exploring every corner of a cube, and rather explore every corner of a simplex. The number of corners will then be proportional to P rather than 2^P.

Evaluating genuine creativity is a hard task, and this study has not addressed many issues, for instance whether this type of exploration can improve the value of the final musical results. What is missing is some attempt to obtain user evaluations of the discovered sounds, and to ascertain whether adding a divergent component had a positive effect on the *quality*, as well as the quantity of the discoveries. A social media aspect could be introduced to enable users to rate each others presets (in the manner of Amabile's consensual assessment technique [Hennessey and Amabile, 1999; Amabile, 1996]). Alternatively users could listen back to their exploration

[3] The grid to slider coordinate transformation took negligable computation time compared to, say, graphically displaying the grid lines on the screen.

session and use a slider to evaluate the sounds they had just navigated through. This approach to revealing the fitness function would be reminiscent of the work of Jennings et al. [2011].

The exact motivation for the users to save the sounds in this experiment was probably too unclear, reflected in the very low average number of presets saved. Perhaps if the app was better integrated into musicians' everyday workflow, and less of a one-off experiment, the saved sounds would be more indicative of presets though genuinely valuable.

A promising avenue to analyse this interaction data would be to use more sophisticated statistical behavioural analysis tools. For example, when analysing large amounts of animal location data, techniques exist to detect discrete hidden states (such as feeding, resting or migrating) from the statistical properties of the animal's movement time series [Jonsen et al., 2005]. It would be interesting to apply these to the navigation paths across the Zoomer surface, and attempt to detect if divergent and convergent interaction really are statistically distinct states. Is a two state model really the most appropriate, or are there perhaps other interaction modes?

It might be objected that the creation of a finished piece of music is a truly creative act, but sound design is not. A very reduced domain, such as this ten parameter synthesiser, might not be considered powerful or high-dimensional enough system enough to study anything as sophisticated as creativity. If this is true, then the results from studying sound design will not 'scale up' to more complex creative works. Certainly this experiment has not investigated transformational creativity. However, in Appendix B's survey, many artists claimed that sound design discoveries could indeed inspire entire tracks. In addition, the solution space of the entire ensemble is a combinatorial superset of the parameters of the separate devices. Therefore the individual adjustments of instruments are still navigations within this larger space, and can presumably exhibit more or less innovation. Considerations of small-scale exploratory creativity in the reduced domain should, in theory, still apply in the larger one. According to the everyday creativity principle, even in very basic interactions small amounts of creativity may be present.

Musicians freely admit to a role for unpredictability and serendipity in their work. Due to music technology's roots in the recording studio, and the designers' tendency to think in terms of objective-oriented search strategies, there has perhaps been a lack of acknowledgement of the more serendipitous aspects of creation when designing interfaces and controller mappings. The happy accidents that do emerge are often seen as uncontrollable by-products, and not something possible to design for. The results of this experiment indicate that, whilst completely removing convergent control of individual parameters would certainly be a bad idea, deliberate design according to the considerations in Section 6.2 may unlock divergent traversal strategies and yield increased engagement and innovation.

A video of Sonic Zoom in use can be found at https://youtu.be/485FnfJOuhI, and the app can be obtained from http://appstore.com/soniczoom.

CHAPTER 7

Evaluating Multidimensional Controllers for Sound Design Tasks

7.1 Introduction

This experiment investigated the differences between single dimensional controllers (touchscreen sliders) and multidimensional controllers (an XY touchpad for 2D, the Leap Motion hand tracker for 3D) for matching a target sound. The original hypothesis behind this experiment was that the Leap would suit the 'skilled' mode and the sliders would suit 'algorithmic' interaction mode of the EARS model. Thus, the intention was to search for evidence that, after enough practise, using the Leap would result in faster and more intuitive manipulation of multiple dimensions simultaneously, but when using the sliders this behaviour would be absent. The corollary of this hypothesis is that the Leap should provide higher throughput measurements.

Both skilled and algorithmic modes are 'convergent' in that they seek to locate

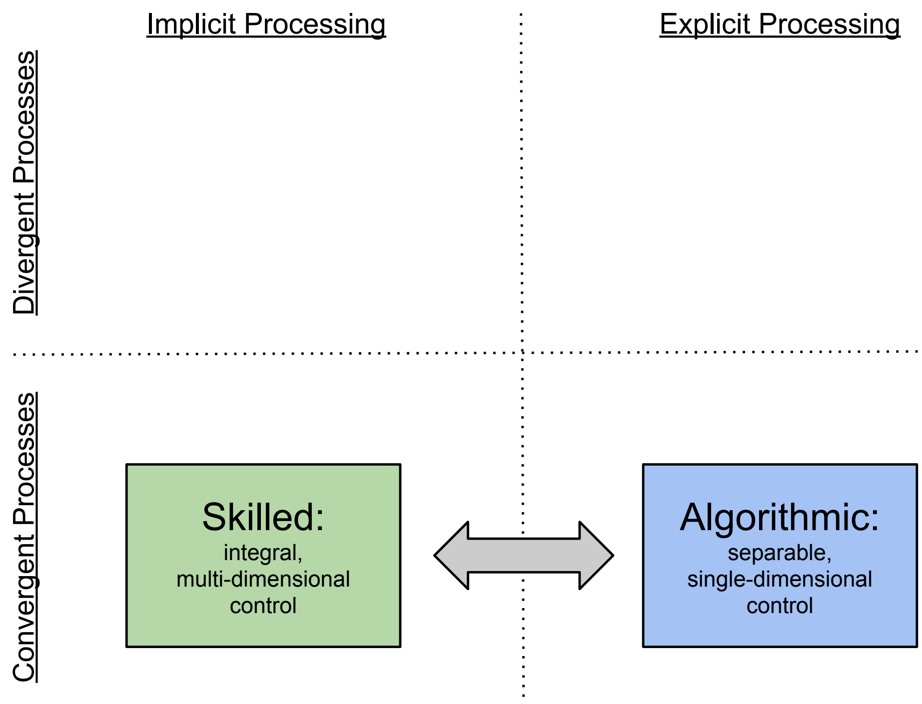


Figure 7.1: This experiment contrasted two methods of controlling synthesis parameters during a sound design task, hypothesised to be suited to skilled and algorithmic modes of the EARS model.

some predefined, optimal target point in the parameter space. This study assumes that if we provide the user with a target sound, which they need to alter the controls to match, this will resemble the process of trying to find an imagined sound. The interface that is fastest for finding the specified target will presumably be the interface that is faster for realising a well specified internal goal. Subjects had to alter the timbre of their controllable sound match a target sound as quickly and accurately as possible. This task was 'gamified' [Deterding et al., 2011] in that participants were provided with real time feedback as to their performance, a running score based on speed and accuracy, and a small prize for the best result.

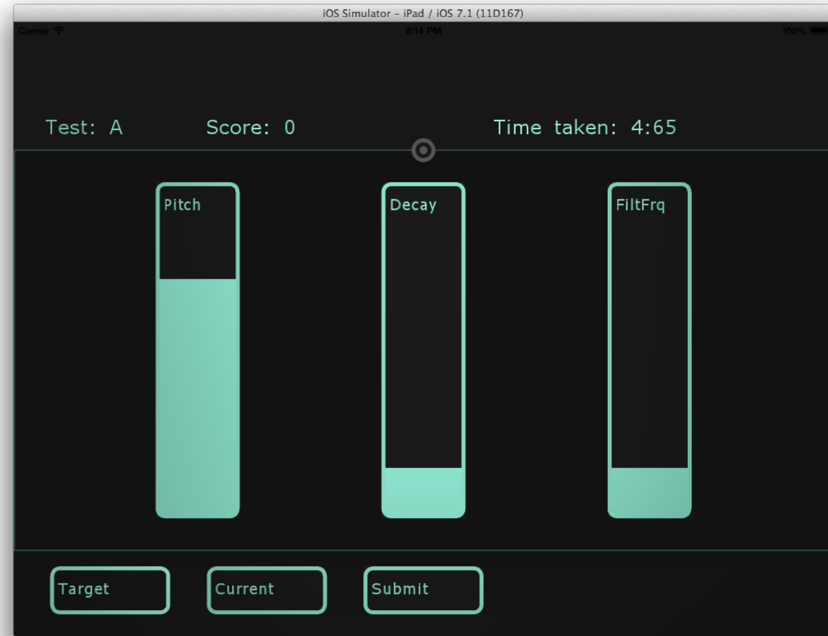


Figure 7.2: Screen shot of the 3 slider interface during the search task. The "Target" button plays the target sound, the "Current" button plays the sound that is being adjusted using the sliders. When the user has matched the two sounds, "Submit" is pressed.

The methodology used for analysing the results of this experiment is the one based on Fitts' law described in Section 5.4. The amount of convergence on a target is quantified using ISSR: the logarithm of the ratio of start distance to current distance to the target, multiplied by the dimensionality. One thing the study aims to establish is if Fitts' law, or rather our ISSR version of it, holds for auditory target matching as well as visual target matching. In other words, does the rate at which people hone in on a synthesiser sound indicate a constant rate of information processing in the sensorimotor loop? Can we use ISSR based analysis to determine any difference between multidimensional and separate 1D controls, and if there is one, determine the cognitive or physical reason for it? If there is a difference between single and multi dimensional controllers, does it increase with dimensionality? By carrying out sound matching tasks in 1, 2 and 3 dimensions, we investigate how the speed of information processing scales with increasing numbers of parameters.

Another phenomenon of interest in this experiment is how different types of feedback will affect the target matching process. How does the location of a visual target (i.e. the standard Fitts' paradigm) differ from location of an auditory one? Is the difference between interfaces in the auditory case *explained* by the differences in the visual target case, or are there more subtle ways in which the perceptual qualities of timbre interact with the affordances of multidimensional controllers? We also compare the case where users can repeatedly compare their sound to the target, to the case where they have to memorise a target.

Figure 7.3 shows a possible cognitive model for the sound matching process. If an implicit, holistic way to compare the differences in sounds exists, then it should show up as diagonal movements in the parameter space. However, there are two possible mechanisms for this diagonal movement. The first is the prediction of a diagonal direction that makes the current sound more like the target, the second being a prediction of the absolute location in the space.

In the next section, the experimental method is described. Then, in Section 7.3 we discuss how the data was analysed using the Index of Search Space Reduction (ISSR) method. Then the results are reported, first in terms of the speed and absolute accuracy of the search end points (Section 7.4.1), then in terms of ISSR for the entire search trajectories (Section 7.4.2).

7.2 Experimental Method

The study was a within-subjects repeated measure design. 8 subjects carried out 8 blocks of 94 sound matches. Whilst it is generally better to use more subjects for less trials, a pilot study revealed that performance was still improving after numerous runs, so a more longitudinal study was required. 'Expert' participants were selected,

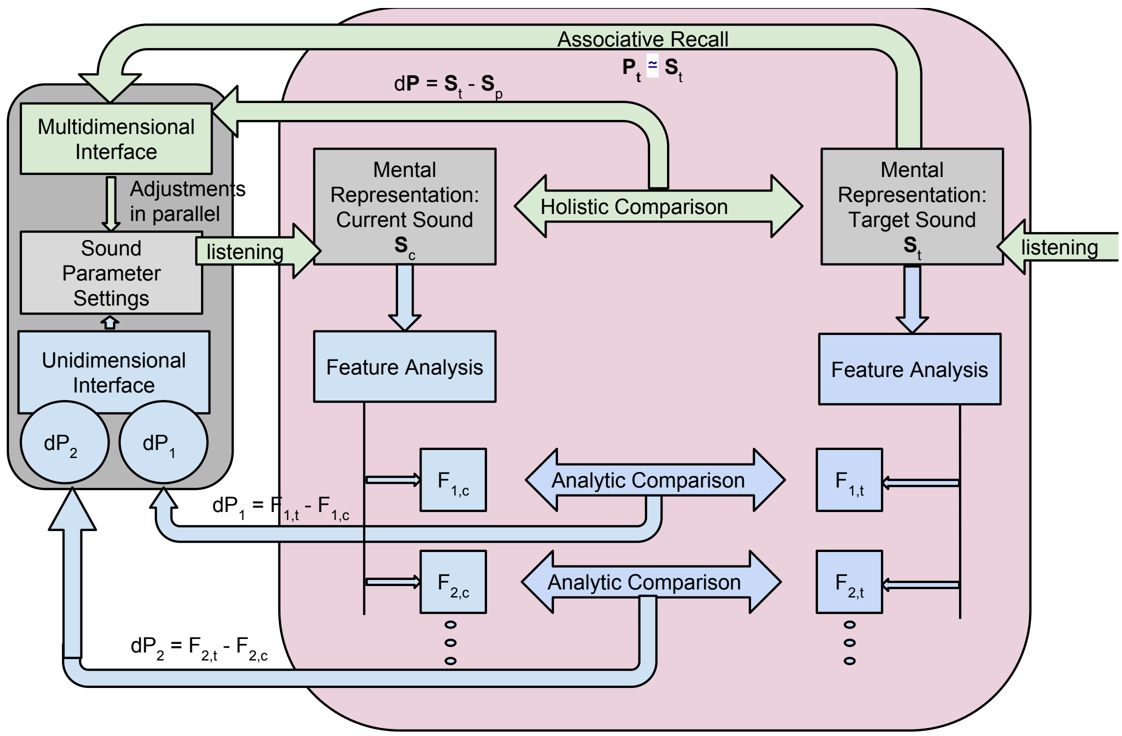


Figure 7.3: A speculative cognitive model of the user's sound matching strategy. The grey box is the technology, the pink box contains cognitive processes. For analytic comparisons (lower, blue processes), each feature F_n needs to be compared by being in working memory separately for both the target $F_{n,t}$ and the current controllable sound $F_{n,c}$, resulting in a single-dimensional parameter adjustment dP_n. If a way to holistically compare sounds using the implicit system exists (upper, green processes), then it will result in parallel manipulation of the interface parameters according to the difference vector $d\mathbf{P}$. A more likely method of implicit interaction is associative recall of the (approximate) absolute position $\mathbf{P}_t$ of the target sound $\mathbf{S}_t$.

with at least 5 years experience of music, sound synthesis or working with audio. They were paid 30 GBP for participating. To avoid fatigue, participants completed four blocks on one day, and four the following day. Table 7.1 shows the sequence of trials for a single block. All users conducted the trials in this order, which was designed to ramp up in difficulty, whilst balancing the multi-D and separate slider conditions. The sequence could have been better balanced or randomised, but at the expense of a coherently gamified user experience. It was assumed that after

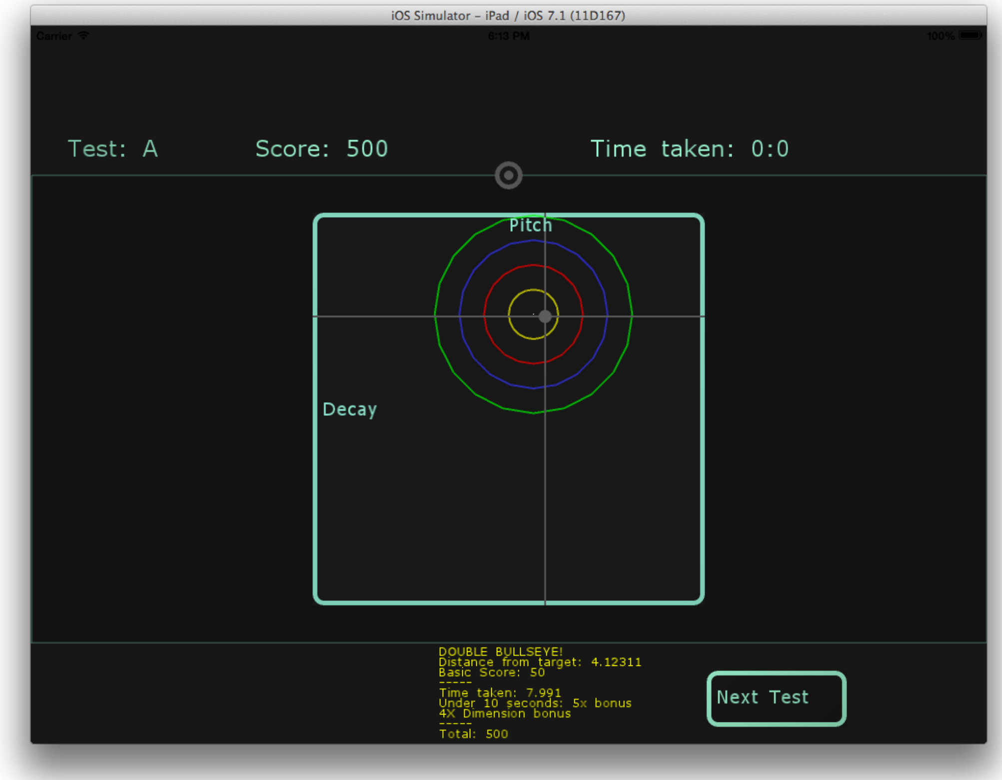


Figure 7.4: XY pad trial. A successfully located sound has been submitted.

8 blocks the effects of the ordering would have balanced out, but this could be a limitation of the study.

The sound generator was a basic digital subtractive synthesiser, constructed in Pure-Data [Puckette, 1996]. The sound could be described as a short "pluck" with varying pitch, duration and brightness, as often heard from classic synths such as the Minimoog. The application ran on an iPad multi-touch tablet, the hand's coordinates being sent from the Leap via a MIDI connection. The following parameters were sent to the synth as 7-bit MIDI CC[1] values:

1. Pitch: a one octave range, midi note 40 (E2) to 52 (E3).

[1]Musical Instrument Digital Interface, Continuous Control

REP	DIM	UI	PRC	VIS	MEM	PIT	DEC	FLT
1	1	Slidr	Y	-	-	1	-	-
1	2	XY	Y	-	Y	1	2	-
1	3	Leap	Y	Y	-	1	2	3
2	1	Slider	-	-	-	1	-	-
2	1	Slider	-	-	-	-	-	1
2	1	Slider	-	-	-	-	1	-
4	2	Slider	-	-	-	1	2	-
4	2	XY	-	-	-	1	2	-
4	2	XY	-	-	-	-	1	2
4	2	Slider	-	-	-	-	1	2
8	3	Leap	-	-	-	1	2	3
8	3	Slider	-	-	-	1	2	3
1	1	Slider	-	Y	-	1	-	-
1	1	Slider	-	Y	-	-	-	1
1	1	Slider	-	Y	-	-	1	-
2	2	XY	-	Y	-	1	2	-
2	2	Slider	-	Y	-	1	2	-
4	3	Slider	-	Y	-	1	2	3
4	3	Leap	-	Y	-	1	2	3
2	1	Slider	-	-	Y	1	-	-
2	1	Slider	-	-	Y	-	-	1
2	1	Slider	-	-	Y	-	1	-
4	2	XY	-	-	Y	1	2	-
4	2	Slider	-	-	Y	1	2	-
4	2	Slider	-	-	Y	-	1	2
4	2	XY	-	-	Y	-	1	2
8	3	Slider	-	-	Y	1	2	3
8	3	Leap	-	-	Y	1	2	3

Table 7.1: The trial sequence for one block. REP column gives the number of repetitions of this trial. PRC indicates a practice run, not scored and not included in results. Controlled conditions were: DIM: number of parameters, UI: interface type, VIS: Visible target, MEM: only one listen to target sound. PIT: indicates which control (if any) operated pitch, DEC: decay time, FLT: filter cut-off. For Multi-D controls 1, 2 and 3 correspond to X,Y and Z dimensions respectively.

2. Decay time: this affected both the decay of the amplitude, and also the rate of decay of high frequencies. The maximum note length was 500ms.

3. Filter cut-off: the cut-off frequency for the resonant low-pass filter.

The multidimensional controller used for the 3D hand tracking was the Leap

Motion (Section 2.4.2). Skeletal hand tracking can generate at least 20 DOF, however the number of parameters was limited to 3: the XYZ position of the hand. More parameters would likely have increased the difficulty of the search beyond most people's capabilities.

For each trial, after an initial 3 second countdown the user was presented with two sounds: the "target" and the "adjustable" sound, with parameters set to random values and a required minimum Euclidean distance between the two[2]. The task was to alter the adjustable sound so that it matched the target sound. Participants were told that speed and accuracy were equally important, and this was reflected in the scoring system. Controls were adjusted with the right hand, and the results heard by retriggering the sounds with the left[3]. In the standard test, either sound could be triggered whenever the user wished. In the target sound memorisation test (Table 7.1, MEM condition) the target button would disappear after a single listen. The intention behind this test was to more closely approximate a realistic sound design task, where the user may have a sound "in their head" that they wish to create, but was assumed to be a more difficult condition due to memory fade.

When the user was happy that their settings matched the target, they would press the "submit" button (centre bottom Fig. 7.2) and were given a score and a visual indication of where the target really was (e.g. Fig. 7.4). A small prize was offered for the best score for one block. Participants stated that 'gamification' of the task increased their motivation and engagement.

Figure 7.5 shows how the various experimental conditions feature in the perception-action loop (discussed in Section 5.3).

A number of tests were control tests with a visual target (Table 7.1, VIS con-

[2]For the Leap, the initial settings would correspond to wherever the user's hand was when the test started. This start position was taken into account when calculating ISSR from distance ratios.

[3]In the pilot test the sounds played automatically in alternation (reducing variability in this part of the task), but people found it too hard to determine which sound was which.

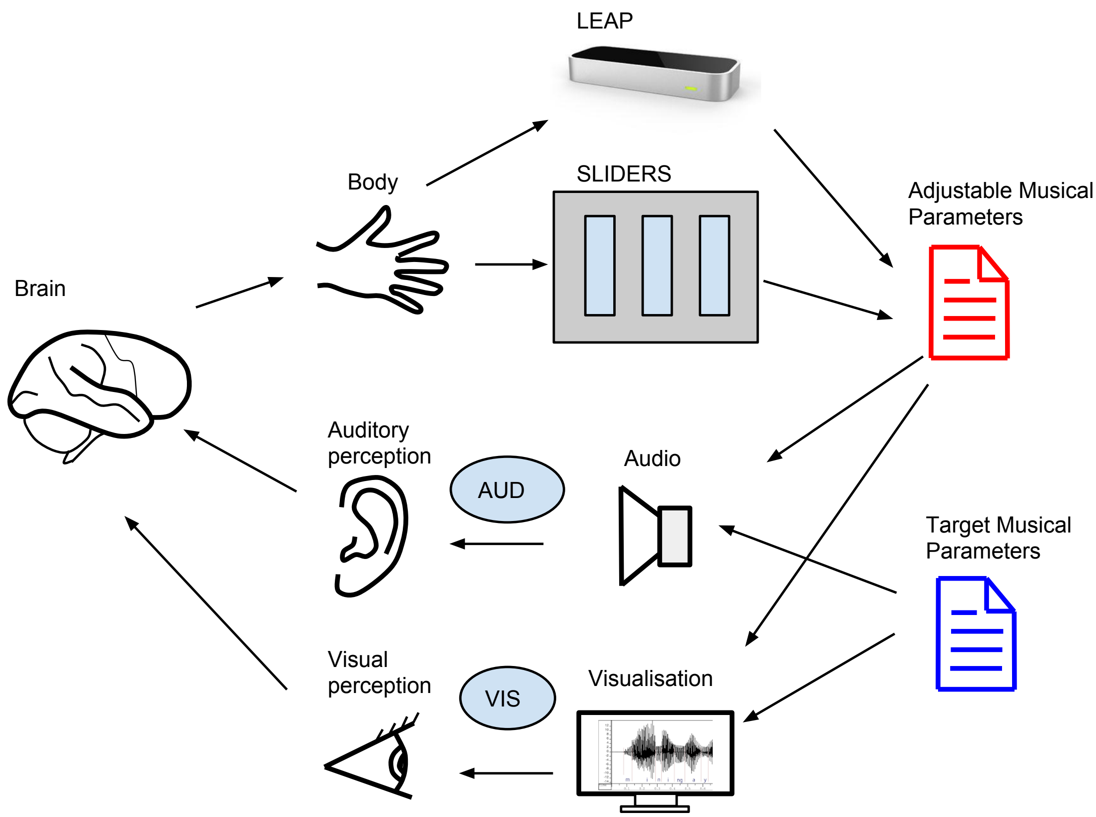


Figure 7.5: Different experimental conditions in the perception-action loop. The interface can be multidimensional (Leap) or unidimensional (sliders). The return channel is swapped between a visual target (VIS condition) and auditory target (AUD). The VIS condition is a more accurate and immediate feedback modality, so revealed interface differences more clearly.

dition). The user simply had to line the controls up with this visual indicator, the sound being irrelevant. This was to test for interface effectiveness independent of the more complex perceptual aspects of sound matching. In the Leap motion case a 3D scene was displayed on the touch-screen, with "jack" style crosshairs to be aligned.

For the 3D trials, parameters 1 to 3 were always assigned to the x (left/right), y (forward/backward) and z (up/down) axes respectively. The 2D tests alternated between pairs of parameters 1 & 2 and 2 & 3. The 1D tests alternated between all 3 parameters. There were an equal number of trials for 2D and 3D tasks, sliders/Leap control type and normal/memorise conditions.

The 1D controls were 10cm vertical sliders on the tablet screen. The 2D XY pad's height and width was also 10cm. Users did not have to pick up the position indicator from its current position before moving it. Unfortunately this meant losing data in the VIS scenario, as users could just tap the target and hardly any of the trajectory would be recorded. The iPad was directly in front of the user, and the Leap was positioned 20cm to the right of the top right corner of the iPad. The size of the Leap's active volume was 30cm cubed, 15cm above the device/table. All interaction movements and events were logged at a sample rate of 50Hz.

7.3 Interpreting the data using ISSR

Throughput (TP) seems like it should be a useful measure of progress in this target acquisition task. One question is if the prerequisites for Fitts' law apply for this experiment. The search is certainly not "rapid", and may not be "aimed", due to low sightedness. The size of a sonic target is impossible to specify to the user, therefore they cannot implement different accuracy levels to provide a range of values for a regression line. One can calculate W from the standard deviation of the results to obtain the "effective width" $W_e = 4.133\sigma$. However, the high variance in accuracy generates extremely low ID values (for the 3D search in this study $\sigma \approx 10, D/W_e \approx 64/40, ID \approx 2bits$), and this single error distribution would not provide a *range* of difficulties, which is necessary for ascertaining any linear relationship. On the other hand, we carried out a large number of trials, and have a record of all the search paths, many resulting in high accuracies. Therefore, it would be helpful to have a method of extracting useful information from these trajectories.

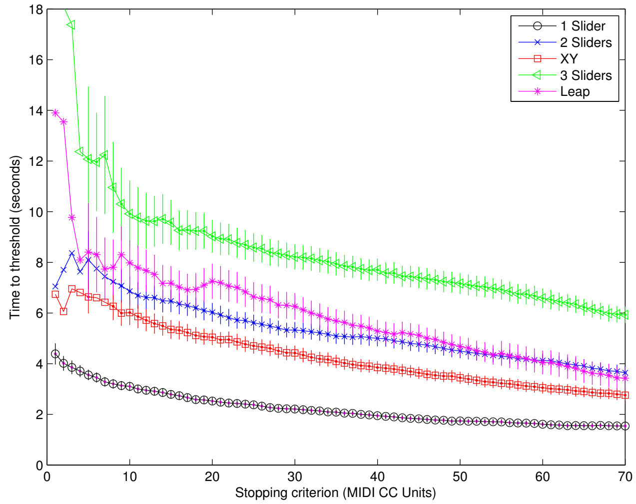


Figure 7.6: Average time taken to reach a given Euclidean distance threshold for each interface condition, day 2, obtained using the method in Jacob et al. [1994] . Whilst different dimensionalities may not be directly comparable here, they are shown on the same plot for brevity. Whiskers display 95% confidence ratios at points where difference between interfaces is significant.

As mentioned in Section 5.4.1, Jacob et al. [1994] performed a retroactive analysis of the search trajectory that measured the time taken to reach various accuracy thresholds (or stopping criteria). This produces a series of simulated experiments with different target sizes. We can set as many of these levels as we wish, and average many trials to get a mean time-to-threshold value. One can then produce plots of time against accuracy (e.g. Fig. 7.6). For our purposes, these plots have a number of issues:

1. The lines often curve up steeply at smaller thresholds. Straight lines would be preferable, as would expressing the accuracy in terms of information gain, in order to investigate Fitts' law.

2. The starting point is not taken into account: if the user starts close to the target, then achieving an absolute distance threshold will be easier.

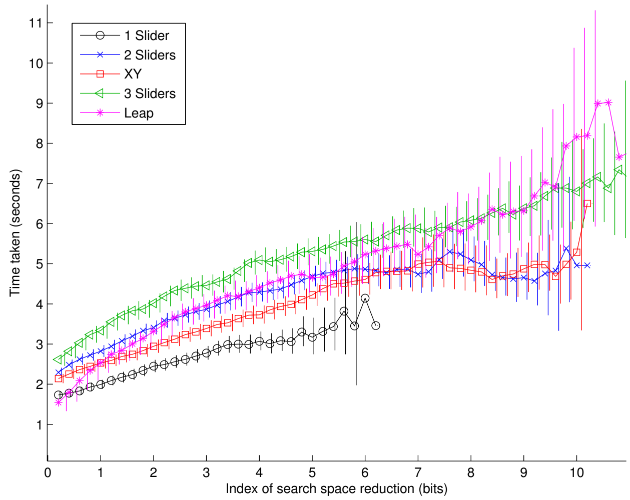


Figure 7.7: Movement time for ISSR thresholds, day 2. The lines are much straighter than for Fig. 7.6 which plots the time to achieve absolute accuracies. Participants rarely achieved accuracies over 8 bits, therefore variance gets much larger towards higher $ISSR$.

3. In more dimensions, the search space is larger, therefore achieving a given threshold will be harder.

We can avoid these issues by expressing the simulated stopping criterion in terms of ISSR. If the sensorimotor loop is processing information at a constant rate, then plotting average MT against ISSR should give straight lines. Their gradients should reflect the relative difficulties in different dimensionalities.

There is a statistical dilemma with this multiple threshold technique, however. One can include all the trials, but poor performances never reach high bit levels, and will not be represented towards the right of the plot. This will tend to make the lines curve downwards, due to higher accuracies only resulting from "luckier" trials or more skilled users. High ISSR values will also display less reliable averages, due to the averaging of fewer observations. On the other hand, if the tests where the threshold was never reached are omitted entirely, the good performances are over-represented and statistical significance decreases due to the smaller sample

	Pitch	Decay	Cut-off
1S	5.28	9.21	9.75
2S	4.49	13.94	12.99
XY	6.13	13.76	13.34
3S	6.67	14.79	14.24
LM	7.93	16.07	15.21

Table 7.2: Inaccuracies (standard deviation from the target in CC units) of individual parameters for all trials. Pitch is always most accurate.

size. The policy here is to use the best half of all the trials for a given condition, i.e. set a threshold at the median ISSR achieved. Any trial that did not reach the median bit threshold are categorised as 'failed searches' and discarded[4]. Whilst this means the final *TP* values may seriously underestimate the task difficulty as a whole, they should at least provide a *relative* comparison between experimental conditions. Higher ISSRs for the successful tests are not featured on the plot, therefore sample size is the same for every point along the line. This should not unfairly favour any particular control device, though it will favour the results from users more comfortable with the task. If the ISSR version of Fitts' law holds, then this technique should give straight lines across a range of bit values.

7.4 Results

7.4.1 Speed and Absolute Accuracy

Scatter plots of speed (time to submit) and accuracy (Euclidean distance to target at submission) for all 2D and 3D trials are shown in figures 7.8 and 7.9. Both axes display approximately log-normal distributions. No correlation between speed and accuracy is seen.

Overall, the decrease in completion time for the multidimensional controllers

[4]Note that this discarding of sub-median trials only applies to ISSR plots in Section 7.4.2. All the trials were used in the scatter plots and speed/accuracy results in Section 7.4.1

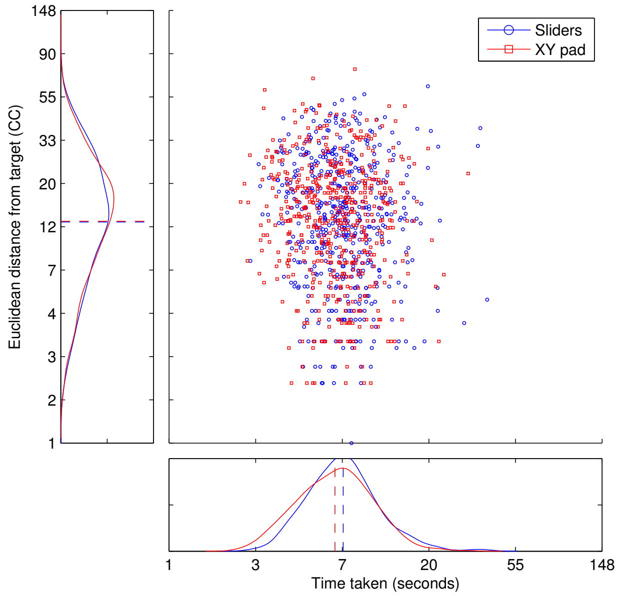


Figure 7.8: Distributions of log speed and accuracy results for the two parameter case, on the second day of the test. The XY-pad is as accurate as the sliders, but has more results under 5 seconds, resulting in a small but significant difference.

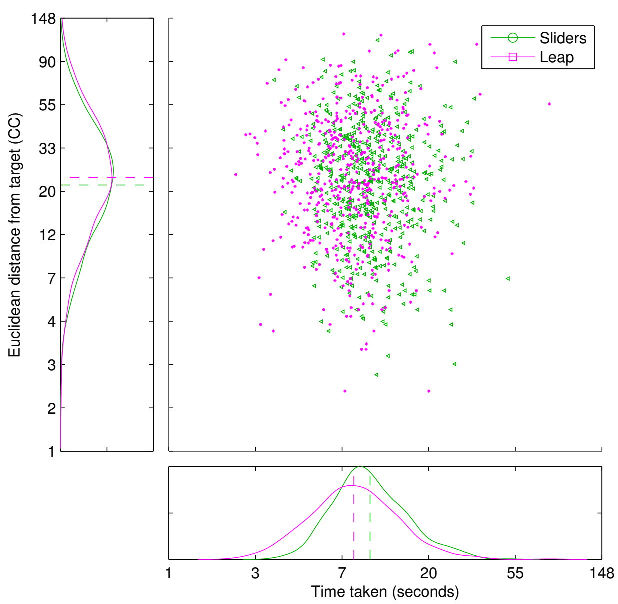


Figure 7.9: Speed and accuracy results for the three parameter case (day 2). Accuracy is slightly less with the leap but it yields many more results faster than 7 seconds.

compared to equivalent numbers of sliders is around 8 percent for the XY and 13 percent for the Leap. However, people significantly improved across the two days (see later). If we look at results for the last 4 blocks (day 2), post practice the XY was 9% faster (paired T-test between interfaces $t(527) = 5.22, p < 0.01$). The leap was 17% faster than 3 sliders ($t(527) = 9.61, p < 0.01$), for an accuracy reduction of 9% ($t(527) = -2.36, p < 0.05$). Individual analyses for each user reveal similar patterns. Here we assume that different dimensionalities are not comparable, but if 2-way ANOVA is run for both dimensionality and interface type, the speed-up due to interface type is still significant ($F(1,1) = 192.4, p < .01$) and there is a significant interaction ($F(1) = 5.58, p < 0.05$).

Accuracy errors for all trials are given in Table 7.2, in the form of the standard deviation of the difference between the target value and the value of the parameter at submission. Not surprisingly, the accuracy for each parameter decreases the more sliders need to be set (the one exception being the good result for pitch in the 2D case). Timbre errors were around twice the size of pitch, despite a pitch range of only 1 octave, illustrating the "anisotropy" mentioned earlier.

We may already conclude that the higher DOF controllers are marginally more effective, but it would be preferable to have a single measure of throughput and trajectory progress plots giving more insight into the cause of the differences.

7.4.2 Throughput

Figure 7.6 shows the average time taken to reach a given Euclidean distance threshold for all 2 and 3 dimensional trials. The Leap and XY pad are faster than the corresponding number of sliders for thresholds $> 5CC$. Figure 7.10 shows ISSR plots for day 1 and day 2. Most lines now appear straighter than in Fig. 7.6, supporting the idea that a Fitts-style law applies. Table 7.3 shows that if a regression is fit to the raw data, the wide distributions generate low R^2 values, but confidence bounds for the slope and intercept are reasonable.

On day 1 the leap was faster than the 3 sliders up to 3 bits, but the gradient b_{LM} is obviously steeper. Day two, the gradients b_{LM} and b_{3S} appear the same, so the Leap's throughput has improved with practice much more than the sliders'. The intercept a is lower for the Leap. This pattern is not seen in the 2D case, here a_{XY} and a_{2S} appear equal but b_{XY} is shallower than b_{2S}. The XY pad is faster even on day 1. Throughput values on the plots are calculated by averaging $ISSR/MT$ for all data points.

The lines appear straight, the high R^2 values for the averaged points indicate a good linear fit. This indicates a Fitts-style constant information processing rate applies for auditory search as well as visual target pointing. However, the lines for the 3D controllers are slightly sub-linear. This would indicate that the search is slightly harder when further away from the target. This would make sense if people were conducting the search one parameter at a time, because the first parameter adjustments will tend to be slightly oblique to the target direction. This may also be the reason that throughput decreases with dimensionality (see line 3, Table 7.3).

	1S	**2S**	**XY**	**3S**	**LM**
Intercept	1.6±0.2	1.8±0.1	2.0 ± 0.1	2.2±0.1	1.2±0.15
Slope (b)	.51±.04	.79±.03	.70±.02	.85±.02	.87±.02
TP (1/b)	1.96	1.26	1.42	1.17	1.14
R^2**(all)**	0.123	0.204	0.156	0.175	0.156
R^2**(mean)**	0.984	0.999	0.997	0.990	0.994

Table 7.3: Results of regression line fitting for each interface on day 2. Throughput (TP) here is taken as the reciprocal of the slope. R^2(all) refers to the goodness of fit for a linear relationship to the points for all trials, R^2(mean) refers to the goodness of fit when trials are averaged.

The intercepts can be largely explained by calculating reaction times (RT), these are shown in Table 7.4. Firstly, RT is the average time from the presentation of the

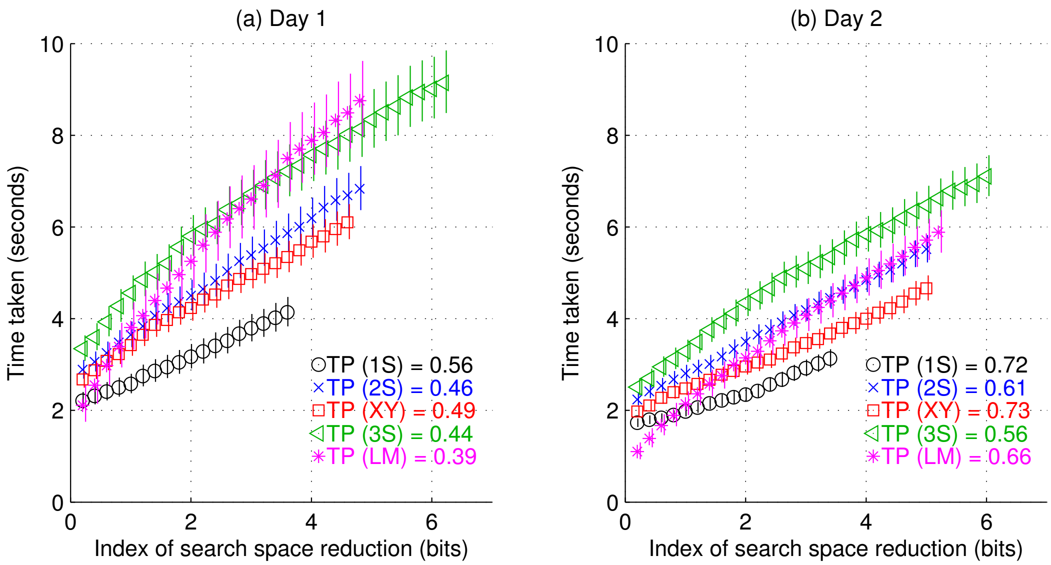


Figure 7.10: ISSR vs. MT, day 1 and day 2. The points are averaged across all participants and all trials that reached the median accuracy (i.e. the accuracy of the rightmost point on the line). Colours and markers are consistent with Fig. 7.6. The gradient for the Leap improves (becomes shallower) with practice to match the sliders, but is about 1 second faster at all bit levels. Throughput (TP) values are calculated from the average of $MT/ISSR$ for every point along the line.

test until the sound is triggered. Second, listening time, LT, is taken as the time taken from the first sound trigger until the first significant control adjustment[5]. RTs are the same for all interfaces (around 1s). LT is more variable. With the Leap, people start moving within 0.25s, even before they have time to listen to the sound they are adjusting. This could be just random hand waver triggering the movement threshold, but the advantage carries through to higher accuracies, so it would appear to be real progress. The quick start also seems to explain the lower intercept on the Leap's plots. The question then becomes: what was it about the Leap that enabled people to start making progress sooner? One hypothesis is that people can *categorise* a sound quickly, and associate it with an approximate region in 3D space. On hearing the target, they can move in roughly the right direction

[5]A significant adjustment was defined as being a movement with a velocity of over 10CC/s

	1S	2S	XY	3S	LM
RT	0.99	0.99	0.98	0.96	1.03
LT	0.85	1.26	1.05	1.39	0.24

Table 7.4: Reaction times (RT) and initial listening times (LT) for different interfaces. People seem to start moving much faster with the leap, explaining the lower intercepts.

	1S	2S	XY	3S	LM
Time	$-22^{\star}$	$-21^{\dagger}$	$-28^{\dagger}$	$-21^{\dagger}$	$-37^{\dagger}$
Median Acc.	-6	4	$9^{\star}$	-3	8
Throughput	27	$32^{\star}$	$48^{\dagger}$	$28^{\star}$	$70^{\dagger}$

Table 7.5: Percentage difference for time taken, accuracy, and throughput between day 1 and day 2 (See Fig. 7.10). Two sample t-test, $\star p < 0.05$, $\dagger p < 0.01$.

without even listening to their current position or considering individual parameters. This would indicate a completely distinct learning process from that occurring with separate controls. This will be investigated further in the next experiment (Chapter 8). Alternatively, one could argue that differences in reaction times reveal a flaw in the methodology, in which case some way of eliminating this effect should be found.

Table 7.5 summarises the effects of practice. The sliders show around a 21% speed improvement from day 1 to day 2, the XY improves by 28%, the Leap improves 37%. Participants keep their accuracy threshold relatively steady.

People quite often needed to revisit a slider once the others were closer to the correct values. In theory, setting 3 parameters necessitates 2 slider swaps. In fact the mean number of swaps was 3.3: indicating that adjustments became less accurate if the other parameters were not set. The mean time for a swap was 0.9s. In the 2D case, number of swaps = 1.7 and swap time = 0.86s. When a visible target was present the swap times were faster: 0.65s. So an extra 0.2s was required to re-orient to another perceptual dimension in the sound task, probably to re-compare the sounds and listen out for that particular difference. These swap-cost issues will get worse in higher dimensions, leading to increased difficulty of navigation. This bears out the hypothesis behind the 'explorability' heat maps in Section 5.5.5.

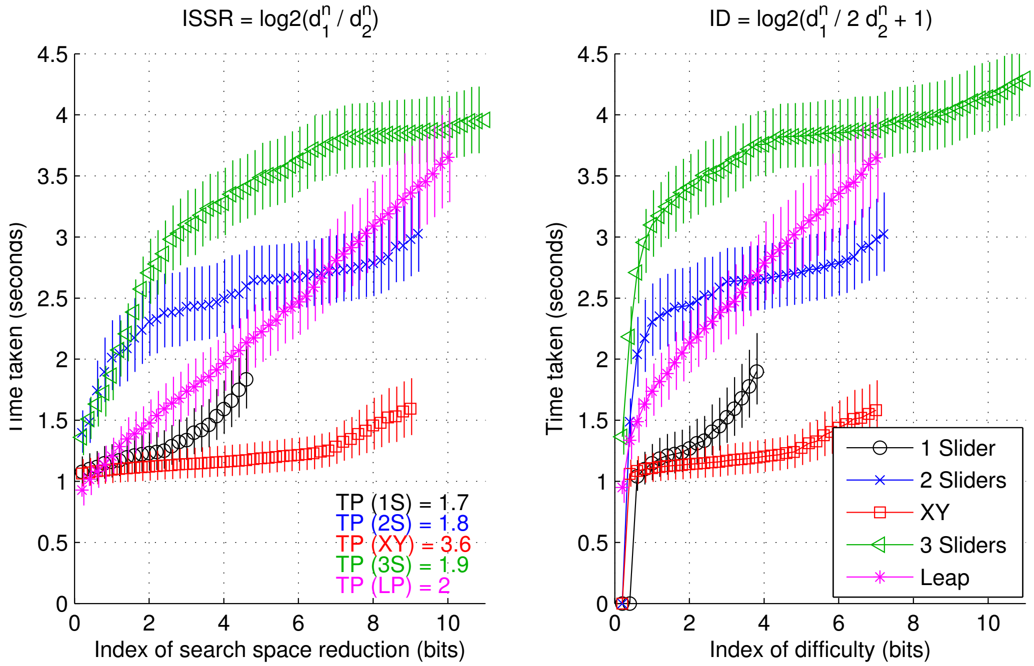


Figure 7.11: ISSR plot for the visible target condition (left) the kinks in the plots for the sliders are caused by having to swap controls. If we imagine the curves extrapolated onwards to higher accuracies, it seems that the 3 sliders will overtake the Leap. The right hand plot uses MacKenzie's ID, introducing sharp drops at low IDs.

7.4.3 Comparisons with Visual Target Acquisition

Figure 7.11 shows the results for acquisition of the visual targets. Around twice the speed and twice the bit accuracy was achieved compared to the sound task. The only interface that gives a straight line is the Leap. The mostly flat lines for the 1 slider and XY plots are because users could simply tap the target, so unfortunately the movement data was not recorded until their finger was on the screen for the final adjustments. However, the straightness of the Leap's plot does seem to indicate a linear relation between ISSR and movement time.

In the left plot the difference between the 3 sliders and Leap is more apparent. The sliders plot has two kinks in it. This is a result of the two control swaps, and the fact that the first slider will be moving at a tangent to the target direction, and will therefore be far less efficient than the last slider: this lower initial efficiency results in a steeper initial slope. Reaction times are similar to the auditory feedback task (but without the extra listen time).

The second plot shows the lines when a +1 is incorporated in the ID calculation (i.e. Eq. 2.4). This results in a sharp curve when $ID < 1$bit. If a regression line is fitted to the data, this reduces R^2 from 0.32 to 0.24 in the Leap's case. So the ISSR formula does seem more appropriate for handling this time-to-threshold data.

By subtracting these results from the results for the sound matching, we can obtain an estimate for a residual measure: the difficulty of searching the sonic parameter space *independent* of the physical control element. We may then be able to see if the interface has an effect on the cognitive processing of the dimensions, in addition to its effect on the physical manipulations of those dimensions. Fig. 7.12(a) shows that the lines seem to coincide, and the intercepts and gradients of the lines become very similar. This suggests that differences in performance were mostly attributable to the physical issues, rather than any specific cognitive suitability to sound design. Fig. 7.12(b) shows that, if the volumetric multiplier n is omitted from Equation 5.7, gradients become inversely proportional to dimensionality. This lends support to the volume reduction derivation. One could claim this as evidence against the perceptual structure matching theory of Jacob et al. [1994], but it is likely that the effect due to perceptual structure is too small to detect in such high variance task.

.

7.4.4 Integration: Diagonal Movement

Another quantity of interest is whether people really did operate more than one dimension at a time, i.e. move diagonally in the integral controller case. Diagonal

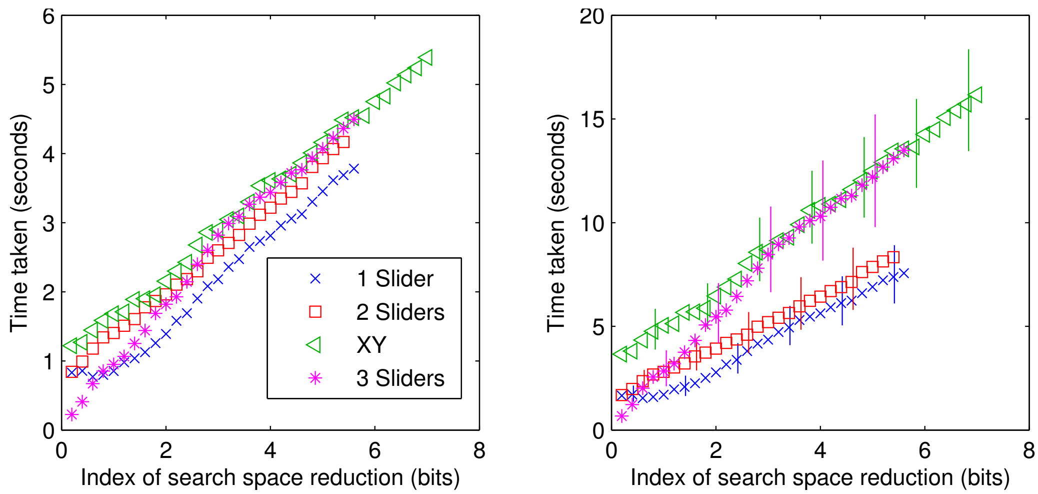


Figure 7.12: ISSR plot for the "interface independent" component of the task (nonVIS - VIS). Slopes and intercepts become very similar, confidence intervals are larger than differences (left plot omits confidence intervals for clarity). Right hand plot shows that omitting the dimensionality multiplier from the ISSR equation destroys this coincidence, lending support for Equation 5.7.

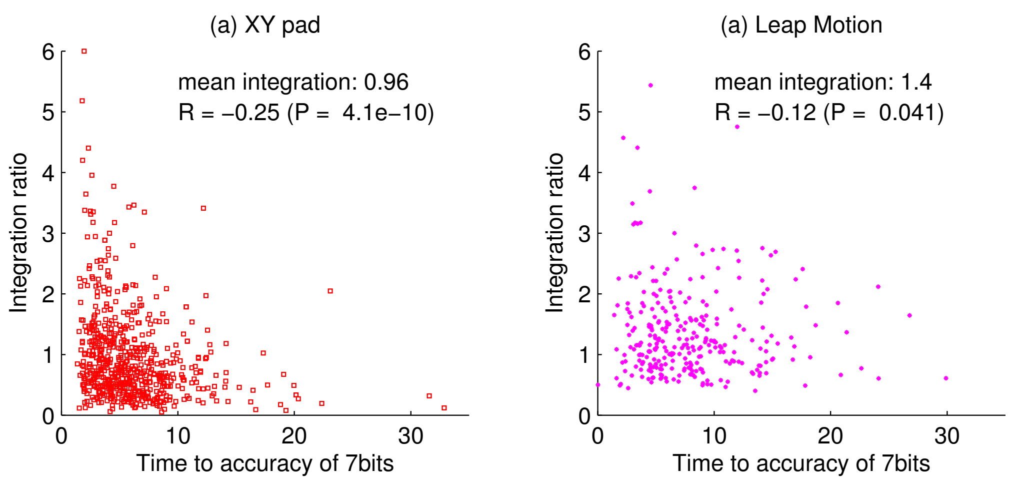


Figure 7.13: A small but significant correlation between amount of diagonal movement and speed for the XY pad (left) and Leap (right).

travel is also referred to as "coordination" [Zhai and Milgram, 1998a] and "controller integration" [Vertegaal and Eaglestone, 1996]. The former is calculated from the correlation of different dimensions, but here, as in [Jacob et al., 1994], integration was calculated as being the ratio between the amount of time that more than one dimen-

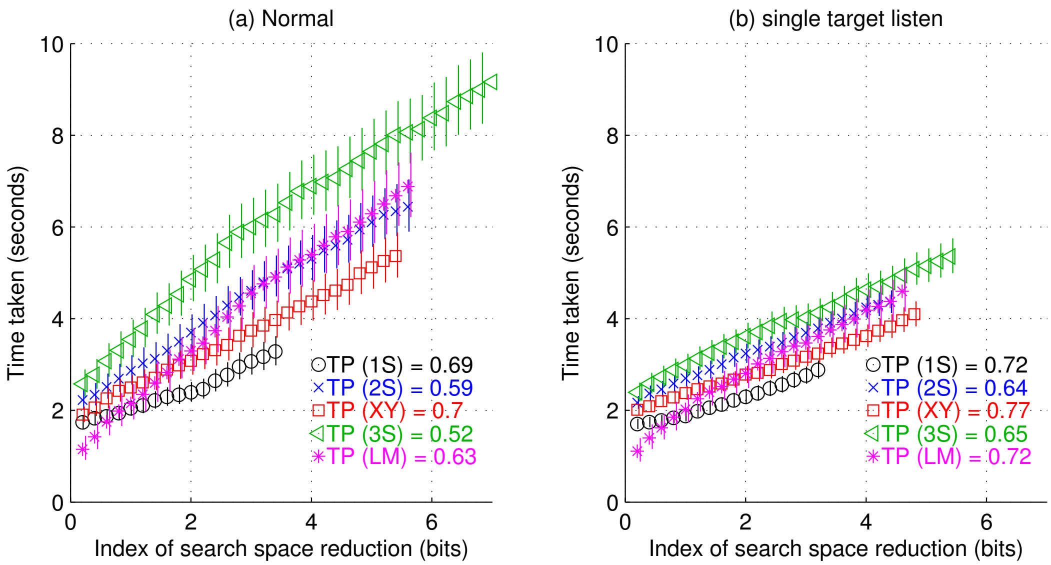


Figure 7.14: ISSR vs. MT, for the normal (left) and MEM (right) case, day 2 only. Not surprisingly, final accuracy has decreased, but the speed up for a given accuracy is quite surprising.

	1S	2S	XY	3S	LM
Mean Accuracy	-6*	-15†	-10†	-21†	-18†
Time to mean acc.	$-14^\star$	$-26^\dagger$	$-17^\dagger$	$-33^\dagger$	$-29^\dagger$
Throughput	8	17	8	26*	22

Table 7.6: Percentage change from normal to MEM conditions (See Figure 7.14). $\star P < 0.05$, $\dagger P < 0.01$

sion was moving to the time only one dimension was moving. The speed threshold distinguishing a moving/stationary dimension was set at 10CC/s. Integration values were heavily dependent on the threshold value, but results comparing experimental conditions were not. A scatter plot of diagonality vs. completion speed (Fig. 7.13) shows that the amount of diagonal travel did slightly correlate with speed, however most navigation was being carried out in a city-block fashion, with integration ratios < 1.

7.4.5 Target Memorisation Test

Fig. 7.14 shows the differences in the MEM case, where only a single listen to the target sound was allowed. Accuracy, not surprisingly, worsens. Participants said that auditioning the sound they were controlling degraded their memory of the target. However it is interesting that the actual time to a given bit threshold is much faster. Table 7.6 shows these differences as percentages. So for rough matches it is actually faster to not keep re-listening to the target. Nevertheless, participants failed to implement this strategy when they were given the choice, indicating that they underrate their own ability to either memorise a target or predict the effect of parameter adjustments. This would indicate that deliberate practice of a feedback-free strategy would result in better performance.

Another interesting aspect of the MEM condition was that search trajectories were more diagonal. For the XY pad, the integration ratio was 1.2 (MEM), vs. 0.8 (non-MEM), $t(1022) = 6.95, p < 0.01$. For the Leap it was 2.2 (MEM) vs. 1.3 (non-MEM), $t(1022) = 7.4, p < 0.01$. It seems that if people are forced not to repeatedly compare the two sounds, they treat the dimensions in a more integral fashion. Could this be because a back and forth comparison encourages a slower, analytical mode of thinking, whereas a sound stored in a short term auditory buffer is treated in a more holistic fashion? This result shows that the type of feedback the user has access to can change their interaction strategy in quite subtle ways.

7.5 User Survey

7.5.1 Workload

Figure 7.15 shows how people rated interface difficulty. Subjects were asked to "rate the following aspects of the task in terms of difficulty":

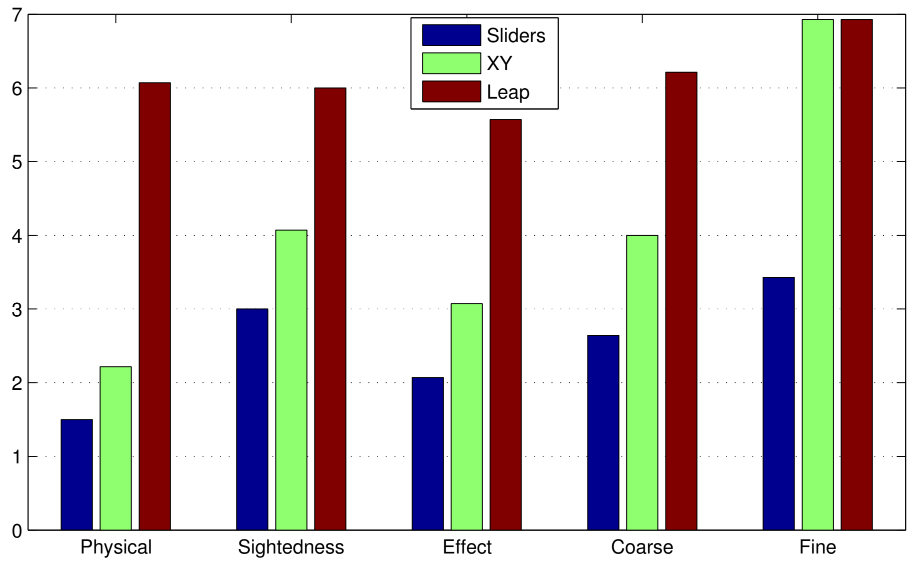


Figure 7.15: User ratings for difficulty of various aspects of the task: Physical effort, ability to know which direction to travel in (i.e. sightedness), ascertaining the effect of movements on timbre, locating a coarse match, and locating a fine match.

1. Physically operating the interface.
2. Deciding which dimension/direction to move (when there were multiple dimensions)
3. Working out what effects my movements were having (when using sliders/ XY / Leap)
4. Locating an approximate match
5. Fine tuning an exact match

Despite the fact that the multidimensional controllers were quantitatively determined to be more effective, both for absolute speed and throughput, users felt they were harder to use. The Leap in particular rated as extremely difficult. So, for this particular task it seems that the gain in speed has come at a cost in both physical effort and cognitive load.

7.5.2 Self-reported Strategies

Next we look at participants' responses to questions that asked how they approached the task.

With the multi-dimensional controllers (XY / Leap) did you try to campare and adjust each dimension separately, or could you begin to associate sound type with locations in the space and move to them diagonally? Did this change over time?

This question revealed that with practice most people did begin to use more diagonal movements with the multidimensional controllers.

> "In the last 2/3 runs I felt ability in 3d environment improve greatly and even began to feel instinctual on final few attempts, rather than using individual parameter thinking. Started using lots of diagonal movements. This might have happened earlier on if I had had more personal experience with the controller before doing the study as part of the improvement might have been down to familiarity with how the device behaves. I tend to be very sensitive to pitch so was good to be able to adjust that at same time as homing in on filter parameters - made doing the two tasks together quite easy."

> "With XY control I found I could work with both parameters at the same time but with the Leap I found myself working with two parameters primarily and then trying to adjust the third one separately. I think the effect was the same over time"

This participant tried to remember absolute 3D locations to move diagonally to sounds, but found thinking about separate dimensions easier:

> "I think I would have gotten to that [skilled/automatic] state eventually with the leap controller... it may have been that i was trying to associate sound type locations within the space and just wasn't very good at it. when i started to think about the dimensions separately, i think i performed much better."

Similarly, this user also said that they had not reached the stage at which diagonal movements would be possible:

> "I felt like was still adjusting the dimensions individually, and I think I always started with pitch in all cases (the easiest and largest effect, so removing that first leads to easier homing in on the other differences). I may have moved diagonally in specific cases, but I didn't feel like I was doing it very systematically - perhaps I was more so near the end when I was getting much more confident on the system. I reckon I'd need 100 hours to get really good at it though, much like driving a rally car on mud for the first time - the first half hour you go from not driving to driving, but you don't get the fluidity until you've done a week of it."

Several users reported that they did use diagonal movements, but would switch back to individual dimensions for fine tuning:

> "Movement in XY pad became more free later on (last 3 runs) but throughout I was generally thinking about individual parameters when using XY. Did begin to use diagonal movement in this for approximation but reverted to individual parameter thinking for fine tuning. Found it

> easy to get confused between which parameter the x and y axis were controlling, despite being labelled, so sometimes free / diagonal movement left me not knowing what was going on."

> "I used the space more, generally moved around to find the right sort of area and then fine tune by using the parameters separately."

> "With the XY pad I tended to listen to each dimension separately. With the Leap I did a lot of diagonal movements and this greatly improved my ability to find the target sound I think. When making precise adjustments I always switched back to thinking one dimension at a time."

This seems to support the fact that fine tuning is an analytic process. Perhaps this is because small differences are easier to break down into separate dimensions than large ones?

Was the task when you had to memorise the target sound from a single listen easier or harder? Why? What difference did it make to your strategy?

This question revealed that most participants found the task harder, but many noticed that they made faster progress toward the approximate location of the sound.

> "I found the memory task easier; I didnt have to think about matching the sound while toggling back and forth between the target and current sample. I think it was easier because i could store the sound in memory and not toggle the sounds. I was also able to find the sounds in the space of the controllers better, albeit for time after hearing the target sound. if I took too long, or had to think of where in the space the sound was, I lost the sound in my memory."

> "Actually, I was pretty convinced I was better at the memory version of the task."

> "Both easier and harder. Easier because I was forced to listen and concentrate more which helped focus on parameter analysis before making lots of noise with user sound. Harder because memory of target degraded with every triggered version of user sound. Overall it made the matching task quicker as I had to be more economical and pragmatic with initial adjustments. I think it improved my performance at approximation but made it harder to fine tune the sound. The strategy used for the memory test involved far fewer test of the user sound."

> "Harder. I paid much more attention to the sound on the single listen as I couldn't keep going back to compare it. I probably submitted my sound quicker."

> "I found it harder overall as I could not carry on checking how close I was to the target sound. However I surprised myself at how well I managed to do on it. The XY graph i found the easiest to to the memory test."

> "Harder. I found it easier to converge to the target sound by switching rapidly between target and current sound. While fine-tuning the sound I tended to switch more rapidly. On the other hand, the memorising task made it easier to find the approximate space."

Did you feel at any point that you were conducting the task "without thinking about it?" Were certain interface types better for this?

Some users did report that "not thinking" seemed to make the task easier, particularly toward the end of the last session.

> "The less I second-guessed my positioning, and the more I made it intuitive and immediate, the better I think I did. I.e., I felt like, after I'd got into the swing of it, my first stab was usually spot on, and quite often I observed that my 'last minute adjustment' moved me right out of the bullseye!"

> "Occasionally early on I found this with the XY pad but mostly in first 4 runs I was thinking about and adjusting 1 parameter at a time in all 3 environments. Later on I found myself being more free with the 3D controller and using more instinct, thinking more about the space and movement rather than about 3 individual parameters."

> "There were moments when it felt like I was intuitively seeking out a note. This was more acute with the 3D hand tracker, then as the X/Y box and finally the sliders gave less room for feeling where the note was rather than working it out / constructing it. "

Other Comments

Whilst most reported the Leap motion as being harder to use, several users really took to it. There seem to be quite large individual differences with regard to how comfortable people feel in the 3D space.

> "Just really enjoyed using the Leap, I felt a lot more comfortable with it after the 8 sessions. Each time I felt like I knew exactly where to place my hands to create the sound I wanted."

> "After playing the 3D notes, it often felt clumsy and mechanical to return to using the sliders."

7.6 Discussion

This experiment provides evidence that multidimensional controllers are more effective, though not by a huge margin. However, they were showing greater improvements with practice, so may be expected to become faster still. The reasons for the speed improvement appeared different for the different devices, however. The XY pad showed a greater throughput due to a shallower gradient: it was faster traversing the space. The speed gains with the Leap, on the other hand, seemed to be a result of faster reaction time: for some reason people felt they could start the search quicker, without waiting to compare the sounds first. Speculatively, this is the result of associating regions of the space with approximate sound characters.

For achieving high accuracies, the sliders were still preferable to the Leap, which was 9% less accurate. Therefore, in terms of sound production work-flow, high-DOF controllers may be better for early stage exploratory creativity and live performance, but individual controls better for late stage creativity and fine tuning, as hypothesised. However it seems doubtful that these small speed gains in 2D and 3D would be worth the extra effort. It is unlikely that this interaction method fits painlessly into the music production workflow. There are a number of other costs apart from the during-sound-search physical effort: for instance the initial practice time, the pre-session effort of mapping the parameters, and the effort of raising the hand into the air to engage with the Leap.

There is a small correlation between diagonal movement and speed, but not yet enough to be the sole cause of significant speed up for multidimensional control. Far more practice seems to be needed to be able to be completely comfortable taking the shortest path through the parameter space. The text responses to the questionnaire seemed to indicate that people did notice themselves beginning to use a more intuitive, unconscious movement strategy, but seemingly not to the extent that it would have unambiguously appeared in the quantitative analysis. One user guessed that about 100 hours would be required before they had learned the perceptual space well enough to move directly to the target in 3D. Indeed, one of the most striking findings is how hard the perceptual component of this task really is. Even with three simple audio parameters, experienced participants, and elimination of the worst half of the results, throughput is only around 0.5bit/s, around a quarter of that for the visual target pointing tasks.

As regards the initial hypothesis, this experiment has not generated any definitive indication that multidimensional controllers are best for the skilled quadrant of the EARS model. For one thing, participants never consistently achieved this level of skill (only one participant felt they had achieved accurate manipulation of more than one parameter at a time by the end of the experiment). For another, even if the participants had moved diagonally through parameter space this would not have indicated automatic subconscious processing taking place. In the next experiment we propose a methodology that addresses these problems.

The proposed ISSR characterisation of Fitts' law proved useful for the following reasons:

1. It enabled us to plot and compare information throughput for interfaces of different dimensionality.

2. Varying accuracy levels could not be specified in advance, but ISSR enabled us to extract a range of difficulty values from the trajectory data.

3. For the multi-DOF controllers, ISSR generated straight lines on movement time plots. This leads to the conclusion that there is a constant rate of information processing in the perception-action loop when engaged in a convergent sound design task. Unlike the Shannon formulation of Fitts' law, the thresholded ISSR plots were straight near the intercept, and these intercepts appear

to agree well with reaction time measurements.

This experiment was probably not precise enough to expose subtle cognitive effects such as the integrality or separability of timbre parameters. One of the main problems in this experiment was the variability in performance. Whilst the main results are significant, the high variance means that detailed analysis of the shape of the plots is probably not meaningful. In the case where subjects were matching visual targets, the differences between separate and integral controllers can be seen more clearly. Another unanswered question is whether it becomes possible to predict the effects of diagonal movement, or whether the only way to effectively process dimensions in parallel is to associate certain timbres with absolute positions in the space.

Given that the differences between controller types seems to increase with dimensionality, it seems the best way to obtain a definitive answer is to increase the number of timbre parameters. However, as it stands, this experimental method will probably become unviable for higher numbers of parameters, simply because it will take subjects too long to find a match. Furthermore it seems that subjects were not using the fastest strategy: the memorisation test showed that there was a sub-optimal reliance on slow comparative feedback. It seems a new approach is needed—can we 'cognitively pipeline' a task such that participants are *forced* to use skilled, diagonal traversal in a higher dimensional space? Can we provide them with an easier way to train themselves using a fast, associative, location based navigation strategy, instead of a city-block predictive one? This is the goal of the next experiment.

CHAPTER 8

Experiment 3: Evaluating 6-DOF hand tracking for Rhythmic Timbre Performance.

8.1 Introduction

The final experiment was similar to the second experiment, in that it compared the Leap motion and touch screen sliders for target matching. In this case however, the task is modelled on a practised performance scenario, instead of a studio-based sound design task. The user, rather than trying to locate an arbitrary sound in the parameter space, is assumed to have specified a number of preferred locations (presets) which they wish to access in a timely fashion during a performance. Nevertheless, they may desire to maintain scope for improvisation, and require the entire parameter space to be accessible if needed.

Shifting from off-line sound design to live performance alters the sound matching task in the following ways:

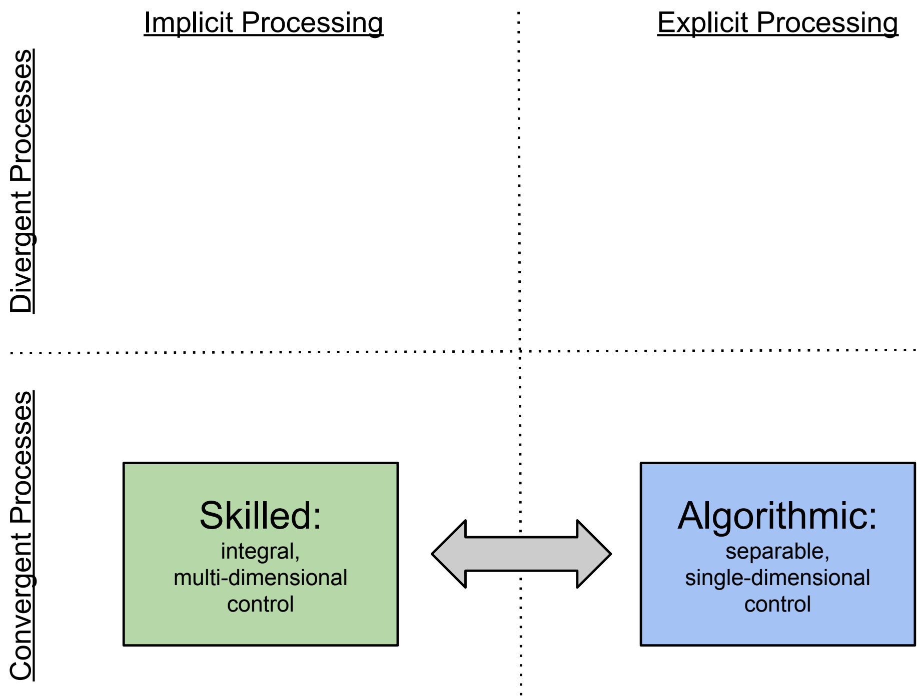


Figure 8.1: Similarly to experiment 2, this experiment contrasted two methods of controlling synthesis parameters hypothesised to be suited to skilled and algorithmic modes of the EARS model. The difference in this case was that the task was far more rapid, easier to learn, and more akin to a practised performance. This task was expected to accentuate the advantages of multi-dimensional control.

1. *Strict time constraints:* there is no longer an arbitrary amount of time to find a target. Most music must be performed at a set tempo, so the most pressing accuracy demand is temporal rather than spatial.

2. *Memorisation and sequence recall:* rather than a one-off search for a sound that is then saved, now parameter settings must be memorised, strung together into sequences, and reliably retrieved many times.

3. *Higher mental load:* during a live performance there is more sensory input from

the environment (e.g. the audience, other musicians), and perhaps more anxiety. Larger temporal structures may need to be attended to and performed, introducing further demands on explicit cognition.

Whilst the task presented here is a little more artificial than a genuine live performance, it features all the above characteristics. It should utilise similar cognitive mechanisms, and the method could easily be adapted to create a genuine performance tool. The findings from the survey in Appendix B provide motivation that real-time performance of timbre variation is a crucial aspect of electronic musical interaction. The software developed for this study makes a design contribution by proposing a system for visualising and learning stored parameter sets: Virtual Body Emulation Assisted Multidimensional Performance (or "ViBEAMP"). By displaying presets as virtual hand positions, this system enables the user to train themselves to perform these sounds via a gamified sequence-matching task. By utilising some of the cognitive principles expounded in the EARS model, a framework has been developed where the user can quickly recall and locate a position in a high dimensional space.

In terms of which aspects of the EARS model are being investigated, this interface is specifically designed for skilled (convergent-implicit) interaction. Therefore we are aiming to find evidence of fast, open loop control for movements toward well-known locations in the space. The experiment specifically tests the hypothesis that multidimensional control is more suited to the skilled interaction mode, and furthermore tests the hypothesis that this frees up working memory for other musically relevant tasks. In addition, Part 1 of the session featured a short sound design task to briefly assess this interface's applicability to exploratory and algorithmic modes. Incorporation of all-quadrant interaction is not yet implemented or tested, due to lack of specific provision for the reflective mode; however, relevant observations from the user survey will be touched upon in section 8.5.5.

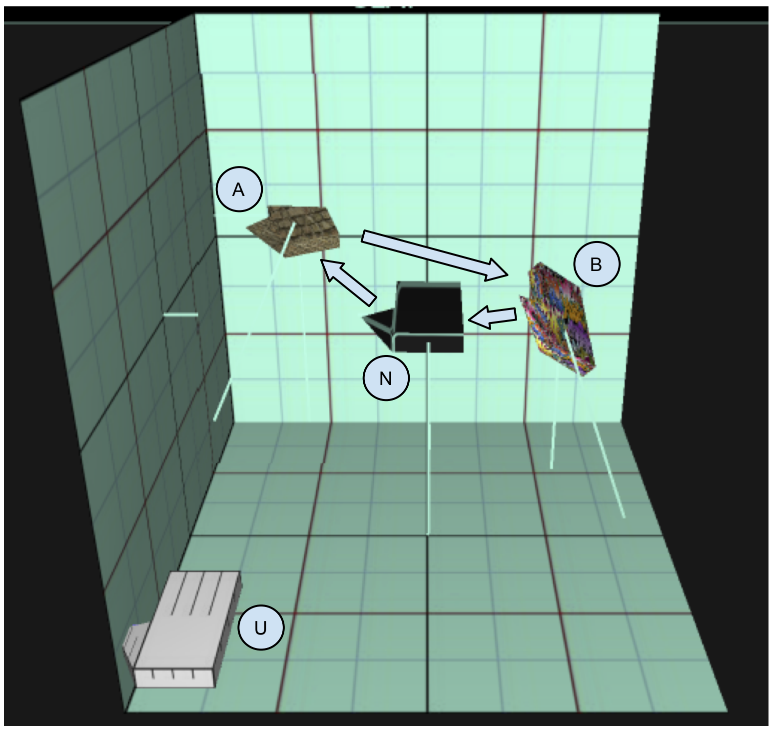


Figure 8.2: Illustration of the two sound sequence matching task for the 6-DOF controller. Subjects need to move their controllable block hand (U) to match the neutral setting (N), then to match the position and rotation of preset A (in this case a kick drum with a “bricks” texture), then preset B (a bongo with a “graffiti” texture), and then back to N. This same sequence repeats four times. Note that during the real task only the next preset polyhedron would be showing at any one time. The kick drum is embedded in the back wall — this type of relative reference point made position memorisation easier.

There were a number of issues that prevented the results of Experiment 2 being entirely conclusive, including high task variance and long and variable reaction times. The requirement that the user triggered the sounds to obtain feedback on progress also introduced variability. The metronome based performance task, described in detail in section 8.3, addresses these issues in the following ways:

1. Any difference between uni- and multi-dimensional control should increase with dimension. Increasing the dimensionality to six should produce a greater effect size.

2. There was no explicit test of working memory load in the previous experiments.

By using six parameters, it should be possible to saturate working memory in the slider condition. Resulting interference with a secondary task can then be measured.

3. Large variances resulted from the difficulty of the perceptual component of the task i.e. the analysis of audio feedback. The previous experiment failed to reveal any significant interface dependence for this perceptual component, so this aspect has been eliminated. By using a limited repertoire of preset targets, and increasing reliance on associative and visuospatial memory, timbre analysis is eliminated from the perception-action loop.

4. There was no way to specify the independent variable for Fitts' law style regression, for either index of difficulty or movement time. However, musicians can accurately predict how long they have to complete a task by synchronising to a metronome. By varying the tempo, a range of movement times can be specified.

5. Reaction times were both a source of variance, and also of a possible bias toward the hand tracker. By using continuous, repeating target sequences, the reaction time due to the unpredictability of the next target is reduced. A cyclical sequence of matches also conforms better to the multi-directional tapping task detailed in the ISO standard for pointing devices [ISO, 2002].

6. The sounds are now triggered automatically in time to a beat. This eliminates the button presses required to trigger the sounds and to submit the matched sound. This should result in less variability in subjects task completion strategies.

7. Participants were, in general, still approaching Experiment 2 in a slow, analytical way. If the hypothesised faster modes of cognition exist, we should be

able to expose them by using short time constraints and well practised targets.

8. Enforcing time constraints whilst engaging subjects in a continuously unfolding sequence of events has the added benefit of generating more data points for a given amount of participants' time.

9. Tasks on shorter time-scales tend to be easier to analyse, simply because there far fewer cognitive processes that can occur.

Therefore by introducing the idea of performing sequences to a beat, many of the uncertainties in the last experiment can be avoided.

The following section describes the design of the ViBEAMP Leap Motion interface, and how it solves one of the main problems with hand tracking based instruments. Next, in section 8.3, the experimental method is discussed, the task that participants performed is detailed, and the methods for analysis of the data are described. In section 8.5 the results of the experiment are presented. Finally, in section 8.6 we summarise and draw conclusions, and then suggest avenues for further research, including extending the design of this instrument.

8.2 Interface Design: ViBEAMP

8.2.1 Design Principles

The motivating principle behind ViBEAMP is that there exist dedicated systems in the brain to process hand pose information quickly and holistically (i.e. the motor cortex). Furthermore there are specific systems dedicated to chaining sequences of movements (the "supplementary motor cortex"), and systems that respond to seeing others perform certain actions and mapping them onto ones own body (the "mirror system" [Rizzolatti et al., 1996]). In order to increase the bandwidth of the

Figure 8.3: Slider interface for 6-DOF performance. Visual target settings ("guides") appear as textured bars on the slider. Similarly to Experiment 2, a tap anywhere along the slider would set the bar to that location.

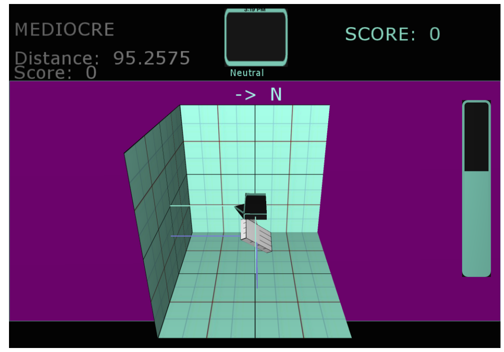


Figure 8.4: figure
Visualisation for 6-DOF hand tracker. The white polyhedron indicates the users current hand position, the guide for the neutral preset is the black polyhedron in the centre.

connection between the high-level artistic goals of the musician and the parameters of the audio engine, these brain systems can be utilised: this is done by encoding the parameters in a form that is well suited for the motor cortex to process. The innate ability to imitate actions carried out by another body naturally leads us to the concept of mapping parameter settings to a hand-space, displaying to the user a virtual hand carrying out gestures, and then getting the user to imitate that movement. How well the user can imitate the virtual hand can then be measured experimentally, and compared to an equivalent imitation task using a less 'embodied' representation of the parameters, i.e. sliders or knobs.

One of the main problems with free hand tracked performance is the lack of any visual or haptic reference points. This system goes some way to addressing the visual aspect, if not the haptic. Recalling a given sound with the current implementation of ViBEAMP relies on a basic motor task, that of matching a position and rotation of a hand-like object in space. The 6-DOF[1] alignment task is often referred to as "docking" [Zhai and Milgram, 1998b]. These preset hand positions are shown

[1] In this study we only use 6 dimensions, but the skeletal structure of the arm/hand has at least 20 degrees of freedom, and this design could be easily extended to incorporate them.

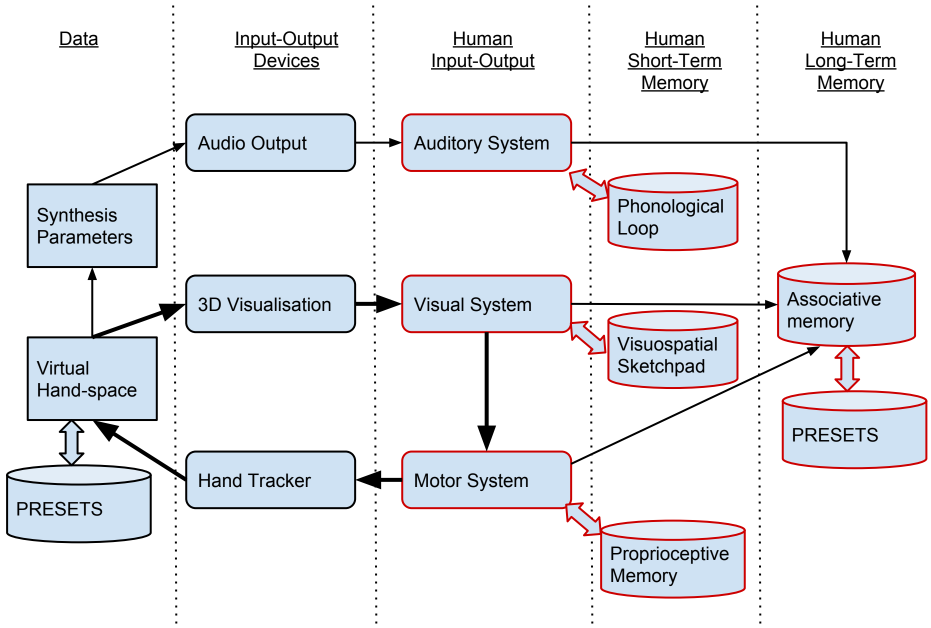


Figure 8.5: ViBEAMP's "cognitive engineering'. Information flows in the direction of arrows. The main loop (shown with bold arrows) involving the visual system, motor system and hand-space is fast and high-bandwidth. The multidimensional data always remains "chunked". Much of the cognitive processing on the right can operate without interfering with generalised working memory.

as virtual block-hands in a 3D scene (see Fig. 8.4). The user's hand appears as a similar shape, but animated according to the hand tracking data. Hence, setting the parameters to play a desired preset sound is then a matter of aligning the controlled hand with the displayed preset hand. The matching process intentionally bypasses any individual processing of dimensions, in order that the user can concentrate on remembering sequences of movements, improving their performance, or listening to the resulting sounds.

Fig. 8.5 illustrates the design of the information flow through both the computer and the brain. At no point in the perception-action loop does the information about presets need to be de-chunked. Furthermore, the system is able to use a number of short-term memory systems that may operate independently of general working memory.

Recalling the position of a given sound will involve associative memory. Associative memory is aided by concrete mental imagery [Paivio, 1969]. A mental image of a hand positioned in a space should be far more easily recalled than an abstract set of parameter values. The more distinctive features a location possesses, the easier will be its recall, therefore associative memory should be aided further by giving the preset hand polyhedron a distinctive visual texture (e.g. bricks, water, sweets etc.), therefore the user learns to associate a given sound with a given appearance of the hand, as well as the hand position. Texture is recommended rather than colours, as there are vastly more distinct and recognisable textures than colours. A saved preset is now a multi-modal associative cluster, rather than a text item in a drop down menu, a series of slider or knob positions, or a step by step sound design algorithm. This is a generalisable approach to learning multidimensional control settings as a chunked unit.

8.2.2 Implementation

As in experiment 2, the iPad was used for the touch screen slider controls and rendering the 3D scene. The Leap Motion was again used for hand tracking. As fast simultaneous manipulation of 6 sliders was rather challenging, their width was increased to 3cm (Fig. 8.3). For interaction with the Leap, the user's hand position, and saved preset positions, were conveyed graphically using an animated polyhedron in a 3D box with 3 walls (Fig. 8.4). The walls of this box featured grid lines to help with orientation. A triangular outcrop on the block-hand indicated the thumb position, and lines indicated the finger end of the block, in order to make the orientation clear.

The synthesiser used was a simple drum synthesiser, modelled on analogue drum machines of the 80's. These machines use a simple sine wave for the pitched component, and a noise source for the stochastic components. Similar drum sounds are still widely used today despite, or perhaps because of, a lack of realism. The noise component was generated from two oscillators, frequency modulating each other via feedback, producing a range of timbres from inharmonic, metallic tones to coloured noise. The three noise controls were highly interdependent, therefore deemed more appropriate for the 3 rotation dimensions.

The reason for using a drum synthesiser, rather than the pitched notes of the last experiment, were as follows:

1. Drum sounds are short, therefore can be performed and listened to rapidly without overlapping.

2. The previous experiment indicated that the pitch dimension was prioritised, and more accurately adjusted than timbre. Drums tend to be distinguished more by timbre, so pitch should be treated in a more integral fashion.

3. There are a number of widely recognised categories of synthesised drum sounds.

4. For rhythmic tasks that needed to be conducted in time to a metronome, percussive sounds may help the user maintain timing accuracy.

For now, no dimension reduction or complex mapping was carried out. The one-to-one mappings for the Leaps degrees of freedom are detailed in Table 8.1.

8.3 Method

The different interfaces (Leap and Sliders) were used in two separate sessions of about 1 hour and 20 minutes. Interface order was counterbalanced. There were two parts to each session, the first being a short exploratory task, and the second a performance training session. Twelve subjects participated, graduate students aged

DOF	Synthesis Parameter	Physical Range
x (left/right)	Sine pitch	-150mm to + 150mm
y (up/down)	Sine pitch decay time	100mm to 400mm
z (in/out)	Sine amplitude decay time	-150mm to + 150mm
roll	FM noise 1 pitch	$-45°$ to $+45°$
pitch	FM noise 2 pitch	$-45°$ to $+45°$
yaw	Noise amplitude decay time	$-45°$ to $+45°$

Table 8.1: Mapping from the 6 degrees of freedom of the hand to the drum synthesis parameters. Physical position range is relative to the coordinates of the Leap Motion device. Angular ranges are relative to the direction the subject was facing.

between 26 and 41. 11 were male and 1 female. Three of the participants had completed Experiment 2, no others had experience of using a Leap Motion. All had at least 3 years experience of playing a musical instrument ($M = 14$, $SD = 7.7$), but 5 had little experience of synthesisers or sound design.

Task 1: Sound Design

This part of the experiment was a sound design task, aiming for a middle ground between the completely open-ended exploration of experiment 1 and the precisely specified target location of experiment 2. Participants had to search the parameter space for 8 drum sounds, conforming to the following drum categories: kick drum, snare drum, tom-tom, hand clap, closed hi-hat, cymbal, bongo and cowbell. They were told to search for a sound that they liked, but that was also appropriate to the category. Only 6 participants felt experienced enough to carry out this task, therefore the other 6 were provided with examples randomly chosen from the ones saved by other users. When they had located the sound they would save it, and the parameter settings would be stored for use in part 2. Participants were told to take about 10 minutes to find these 8 sounds. All interactions were logged.

Task 2: Performance

This task took the form of a training exercise, where the goal was to match targets in time to a metronome. These targets were selected from the 8 presets saved with the same interface in part 1. The target would appear as a floating block-hand for the Leap, or indicator bars on the sliders (Fig. 8.3). These guides were textured with the relevant texture image, in a position determined by the 6 parameters of the relevant preset. Hence, by continually moving the hand to align with the next guide in time for the next downbeat, the user would perform a rhythmic sequence of drum hits. Users received immediate feedback as to their accuracy, in the form of their block-hand momentarily flashing a colour based on the accuracy of the match. A running total score was displayed in the top right hand corner.

First, participants were shown a demonstration performed by the experimenter. They could then practice at a slow tempo, until they felt they fully understood the task. During the Leap demonstration, an indicator displayed the current Euclidean distance to target to assist the user in learning how to align the blocks precisely.

Table 8.2 summarises the terminology used for the series of events. Whilst the sequence of events was somewhat complex, in fact it could be reasonably summed up by the heuristic "watch the preview of the sequence, then imitate that repeatedly

Table 8.2: Summary of terminology used regarding the order and grouping of trials.

Target	A single trial, featuring a target to be matched within a given time.
Sequence	A randomly selected sequence of 1-3 targets to be matched consecutively. The neutral position bookended these sequences.
Tempo Level	A set of 5 repetitions of the same sequence at the same tempo (1 preview, 2 guided and 2 unguided). Each level was slightly faster that the last.
Run	A continuous 5 minute training sequence featuring 15 tempo levels.
Set	3 runs of sequence length 1, 2 and 3.
Session	Consists of the sound design task, and 4 training sets. There was one session per interface.

until the next preview".

A session with one interface consisted of 4 'sets' of 3 'runs', a run being about 5 minutes of continuous interaction. A run was performed to a metronome click, increasing in tempo for each level. Participants would need to continuously keep up with the targets during this run, but could rest between runs. For each run there were 15 different sequences, at 15 different speeds: these are referred to as 'tempo levels'. A sequence always started and ended at a neutral setting (preset N), where all parameters needed to be set to the middle position (64CC). In the Leap's case, preset N was the hand being held horizontally in the middle of the space. Subjects were told that the neutral position was scored, and important to match accurately, but results for this target were not used in any data analysis. The neutral position both provided a reference event for people to know when the sequence was starting and ending, and also ensured that participants would not start moving towards preset A before the allowed movement period. The neutral preset sounded like a low snare drum.

The sequence for the first run of a set was only 1 sound long, therefore involved alternating between the neutral and a random selection from the 8 preset sounds (N,A, N,A...N). The second run featured two preset sounds (N,A,B, N,A,B...,N), and the third run of a set featured a sequence length of 3 (N,A,B,C, N,A,B,C...,N). The selection of the presets was random for each new level. There were no repeating sounds in one sequence. Figure 8.2 illustrates how a two sound sequence would proceed in the 3D visualisation.

Figure 8.6 shows the format of a single level in more detail. The user would be played a preview of the sequence, and shown the control settings for each one in turn. To help the user to judge the movements required, an animated block hand was shown moving from target to target during the preview. For the sliders, dimmed slider bars would animate between the guides. Subjects were allowed to rehearse their movements along to the preview, but were advised they could rest at this point. The animation was an exponential curve adjusting each dimension simultaneously. The preview also offered an opportunity to mentally adjust to a slight increase in tempo. Subjects would then have to recreate this sequence 4 times, by moving the controls to the right settings: twice through the sequence with visual guides to align with, and then twice with no guides. These unguided sequences were to test how well they could memorise and recall both the control settings and the sequence ordering. By the time users reached the unguided stage they had seen the target sequence 3 times, this repetition was intended to eliminate a large amount of uncertainty in reaction times, visual search and so on. The participant's own block-hand was still visible during the unguided runs[2].

The metronome played in 4/4 time, with the drum sounds being triggered on the first beat of each bar. This beat was also the point in time at which the distance from the target was measured. Visual guides for a target needed to be shown the bar *previously*, so that they appeared in time for users to match them. Unfortunately, this caused a little "cognitive dissonance" for some users, as the icon for the next sound would appear as the sound for the previous target was playing (for illustration, refer to the difference between the sound output and visual guide rows in Fig. 8.6). This clash is largely unavoidable: alternative orderings of events were tested and found to be even more confusing. A potential visual reference that could have helped orient people in the sequence would be to display the whole sequence as a DAW style timeline, with a play-head marker scrolling through it.

Some care was put into specifying the tempo range. On one hand it is useful to test the limits of performance: generating an intercept on the time/ISSR plot, and hence deriving an estimate for processing latency. This would imply taking the

[2]For future work it may be interesting to remove the visualisation entirely, in order to investigate proprioceptive memory, and how well the instrument would function on-stage without an obtrusive screen.

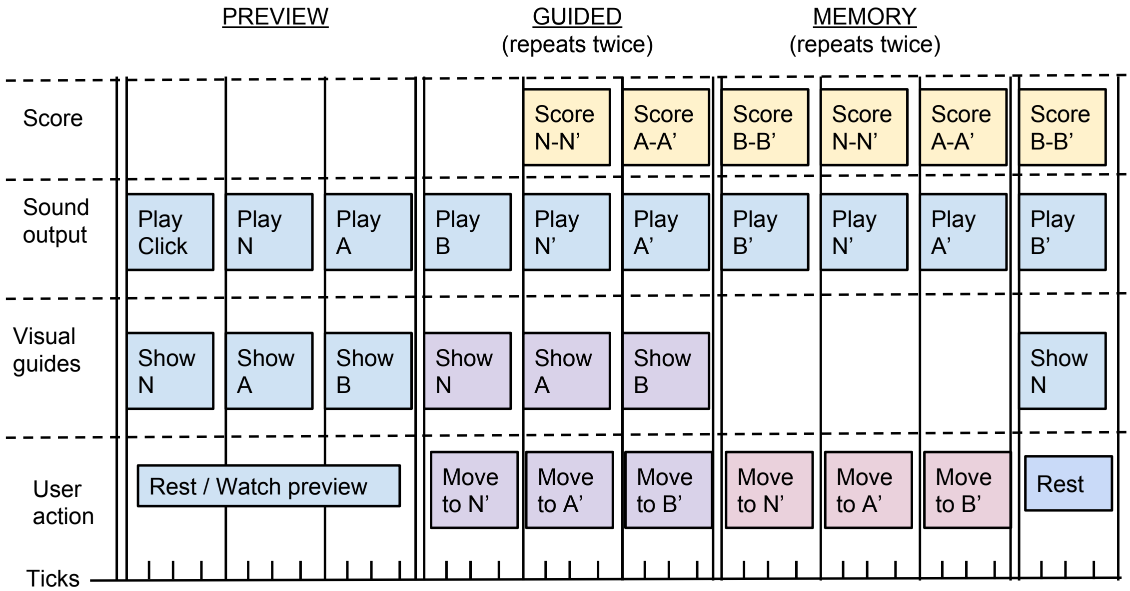


Figure 8.6: Order of events for a single tempo level, for a 2 preset sequence. N refers to the neutral preset, A and B the first and second randomly chosen presets. A' refers to the user's attempt to match the settings of A. Time runs from left to right. Long vertical lines represent bars, with beats (audible as metronome ticks) marked along the bottom. At the end of this series, the next tempo level starts, and this series repeats.

tempo up to the limits of people's abilities to make any progress at all. On the other hand it is unwise to provide subjects with a task so bewildering that they become discouraged and demotivated. In preliminary tests it was found that an allowed movement time (MT) of 2.5 seconds was certainly enough time to reach a saturation point for accuracy for both interfaces, and a comfortable speed to get used to the process. $MT = 0.5s$ seemed to be the point at which the slider task became impossible. Therefore the initial tempo was 96 bpm, and the final tempo was 480 bpm[3]. The demo runs ran considerably slower for explanatory purposes ($MT = 5$ seconds). The tempo increased linearly, therefore MT was spaced inverse-linearly. A linear increase in tempo is preferable to a linear decrease in movement time; for example, the transition from 2.5 to 2.4 seconds is almost imperceptible,

[3]Note the matches were only every 4 beats. At the upper end of the tempo range (> 300bpm) the clicks would probably be interpreted as 16th notes.

but one from 600ms to 500ms is abrupt and hard to adjust to. This spacing also generates more data points in the region of interest: where performance starts to degrade significantly.

To alleviate unnecessary anxiety, participants were forewarned that the tempo would reach a point beyond their abilities, and to try to keep their motivation relatively steady throughout the task.

Ideally, we would have had some way to measure performance on the secondary task, which was to remember the drum category sequence. This could have been self reported by pressing an extra button to indicate a "blank mind" sensation, or an extra stage where the user would to enter the drum sound categories in the previous sequence via another interface. It was felt that this extra demand would have overloaded the user still further, and interfered with the rhythmic nature of the interaction. We assume that forgotten sequences will be distinguishable in the data nevertheless, as by the end of the test the 8 sound positions will be fairly well memorised, but the random sequences will not. Therefore poor performance on the longer sequences will be mostly due to failure to recall the sequence, rather than inaccuracies in recalling the control settings.

After the two sessions subjects completed an online questionnaire (see Section 8.5.5).

8.4 Data Analysis Methods

As in the previous experiment, accuracy results are transformed into search space reduction (ISSR) in bits, measured relative to the start point of the movements. For trajectory analysis, the thresholding technique is again performed, to ensure accidental movements towards the target do not count as real search space reductions. With highly sighted aimed movements this becomes less necessary, and thresholding only affected the trajectory shape of the slowest trials and the exploratory sessions.

For calculation of throughput (TP), ISSR is divided by movement time and averaged across all trials. As we shall see, TP values could be very different for different tempi: this makes averaging the TP across all results and drawing statistical conclusions problematic. For instance, if the experimenter carries out more trials at slower tempos, averages and relative results could be very different. The average of a quantity that changes under different experimental conditions is not very meaningful. Despite this problem, it is still worth reporting average TP, if only for the sake of clarity in summarising the results. However for determining statistical significance, only comparisons between distributions at a given tempo level can be made. No confidence intervals are shown on the averaged throughput bar charts for this reason. Instead, significance levels for all reported effects are given in the ANOVA results, and confidence intervals for each tempo condition are shown on the ISSR plots.

8.4.1 Hypothesised Results

Here follows a summary of the main hypotheses in terms of the expected observations.

1. The operation of the sliders will be generally slower than the Leap. This will result in lower values for throughput for all tempo levels.

2. The latency to process 6 separate dimensions, visually, cognitively and physically, will be higher than for hand poses. For the sliders, at a certain tempo the latency will be too great, and motor productions will simply not be ready in time. This will result in slider throughput falling off more steeply towards high tempo levels.

3. Chunking of hand poses will result in spare cognitive capacity, enabling better memorisation of preset sequences. Hence, with the Leap, longer sequences will show less information loss, relative to their guided counterparts, than the separate controls. Furthermore, reduced cognitive load is expected to have other beneficial side effects, including enabling participants to pay more attention to the audible results, and producing a greater sense of rhythm and flow. The slider task will feel unpleasant and frustrating.

4. Reaching is one of the most essential and innate motor tasks, therefore is hypothesised to make use of a well coded and highly efficient "motor program". The fact that the task is time constrained will encourage this motor program to be executed as an "open loop" or "ballistic" motion, rather than as a "closed loop", iterative motion. Therefore trajectory analysis may produce differently shaped curves from those derived from the iterative correction model and Fitts' law. Rather, results are more likely to show the linear relationship given in [Schmidt et al., 1979].

5. Due to the holistic nature of the position-rotation chunk, and parallel processing of hand pose dimensions, a high degree of coordination between the dimensions (direct diagonal movement through the 6 DOF) is expected.

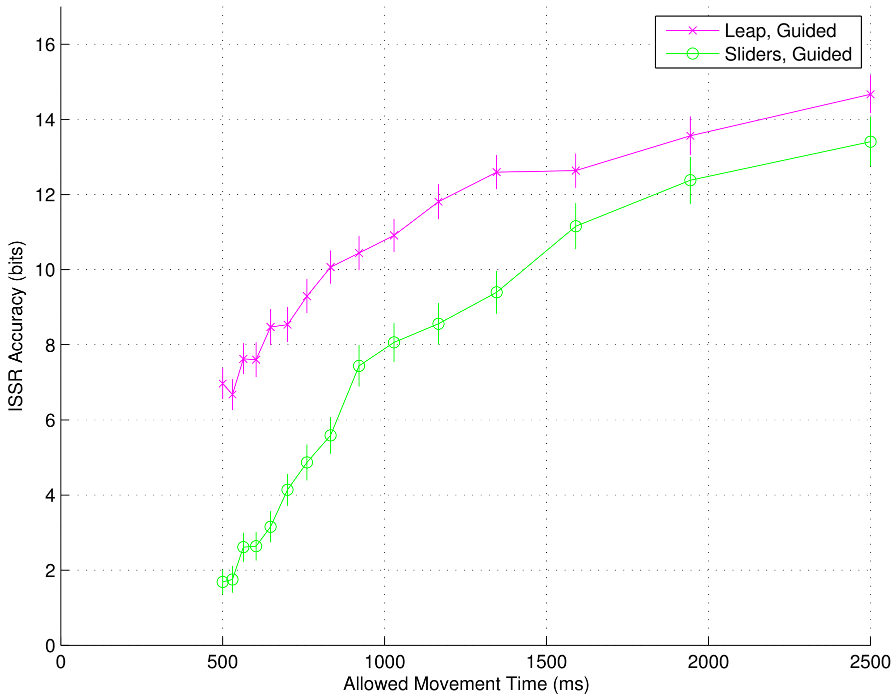


Figure 8.7: Averaged ISSR values against movement time, for all guided trials. Whiskers show 95% confidence interval. Differences were significant for all tempo levels.

8.5 Results

An 11-way ANOVA was run on the matching accuracy values (in ISSR bits) for all the trials. Details can be found in Table 8.3. Highly significant effects were found for interface (Leap/Sliders), user (1-12), tempo level (15 steps from 2500-500ms), guidedness (visual guides present or not), order (order in which users performed interface sessions), set (practice effects), preset (which drum type was being matched), location in sequence (1st, 2nd or 3rd sound), and repetition (1st or 2nd repetition of either guided or memorised sequence). Clearly the visually guided, tempo based scenario produces more statistically robust results compared to the unguided search in Experiment 2.

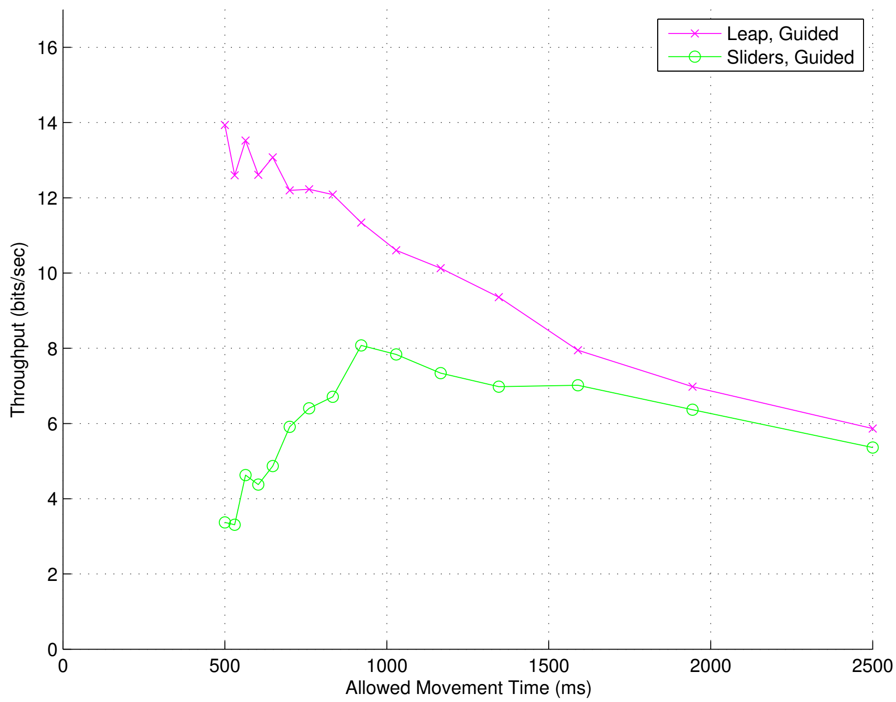


Figure 8.8: Throughput against movement time, for all guided trials. The Leap's throughput keeps rising for increasing tempos, but the sliders reach a peak at 900ms.

8.5.1 Throughput Results for Guided Trials

For guided matches, the average accuracy did not differ significantly between different sequence lengths[4]. Therefore, ISSRs for all guided trials at a given tempo level were averaged.

The most important results of this experiment can be seen in Fig. 8.7. This plots search space reduction (in bits) against the time allowed for the movement—with faster tempos to the left[5]—for both interface types. In this, and all subsequent ISSR plots, vertical whiskers indicate 95% confidence intervals, the slider interface

[4]The only exceptions were the very fastest trials, at these speeds the Leap showed a 2 bit decrease in accuracy for longer sequences, the sliders a 3 bit decrease. These were still averaged across.

[5]To avoid confusion, note that as the trials were performed in order of decreasing movement time, so curves with positive gradient may be referred to in the text as "decreasing towards faster tempos".

Source	d.f.	F	Prob F
Interface	1	6457	< 0.001
User	11	438	< 0.001
TempoLevel	14	430	< 0.001
Guided	1	2635	< 0.001
Order	1	100	< 0.001
Set	3	267	< 0.001
Preset	7	33.1	< 0.001
Position in sequence	2	26.3	< 0.001
Seq. length	2	232	< 0.001
Seq. repetition	1	20.06	< 0.001

Table 8.3: 11-way analysis of variance of ISSR for the various experimental conditions. All sources produced significant effects. Many cross terms (not shown for clarity) were also significant.

results will be shown in green, and the Leap's in magenta. The plot shows that for movements towards visual targets, the Leap outperformed the sliders for all tempo levels, but in particular the faster trials. At 500ms the sliders performance is less than 2 bits, while users can still achieve 7 bits of search space reduction with the Leap. This result is highly significant (e.g. T-test carried out on ISSR values for the 500ms trials, $t(574) = 21.1, p < 0.001$).

If we divide these results by the allowed time, we get values for throughput. Fig. 8.8 shows how throughput alters with allowed movement time. According to Fitts' law, the line should be straight, and approximately horizontal, however this is not the case. The Leap's throughput significantly increases as the task gets faster, from 6b/s at the slowest tempo, to a peak of around 14b/s at the fastest tempo. The sliders throughput begins to increase slightly, but hits a peak far earlier than the Leap, achieving 8b/s at around 900ms and then rapidly declines. The total average throughput for the Leap was 11b/s, and for the sliders 5.9b/s.

The sliders' ISSR curve would appear to be intercepting the time axis at around 400ms, for both guided and memorised targets. Data for sub-500ms movements was not recorded, so the zero bit intercept for the Leap is harder to predict. In order to

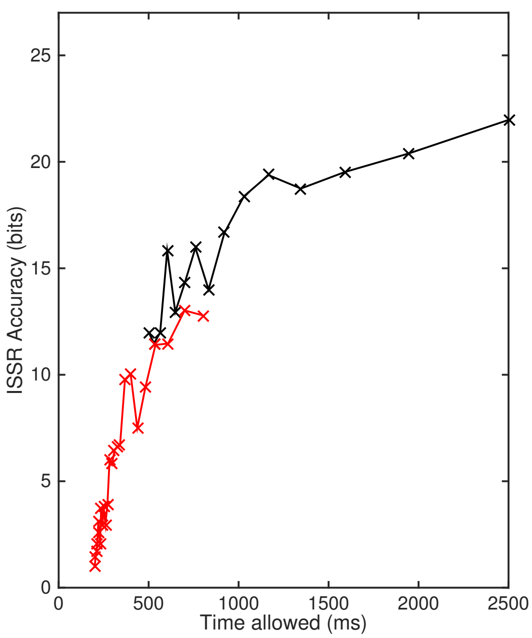


Figure 8.9: ISSR (left) and TP (right) plotted against movement time, for the author's high tempo trials with the Leap (shown in red), along with the author's session at the standard speeds (black). Throughput peaks at around 400ms. The intercept is predicted to be around 150ms.

produce an estimate, the author carried out a number of faster trials and the results are shown in figure 8.9. Here we can see that the intercept is around 150ms, with peak throughput being at around 400ms. So, as expected, the latency is higher for the sliders, by nearly a factor of three[6]. Values quoted in Kieras and Meyer [1997] estimate that each motor command 'feature' takes about 50ms to prepare, with another 50ms to send to the muscles. Therefore it appears that the hand pose requires two prepared features (position and rotation?), whereas the sliders require 7 (i.e. approximately siz, the number of features). It should be noted that this plot shows what kind of performance would be expected with approximately 5 hours more practice.

[6]The author's fast, unguided results (not shown) are roughly 5 bits less accurate at 2500ms, stay flat until joining up with the guided results at around 750ms. The intercept is again around 150ms.

8.5.2 Throughput Results for Unguided Trials

Figure 8.10: Throughput decreasing with the length of the memorised sequences. The sliders suffer more for long memorised sequences.

Fig. 8.10 shows a bar chart of average throughput for the different lengths of guided and memorised sequences. When sequences were recalled from memory, performance suffered for both interfaces, but the sliders significantly more so. For single sound sequences, the Leap performed almost as well as for the guided task (11.4b/s guided, 10.9b/s memorised). The sliders did not suffer too much for a sequence length of 1 either (5.9b/s guided, 5.6b/s unguided), but for two and three sound sequences average slider TP declined by 3.0 and 4.5 bits per second, compared to a decline of 1.6 and 3.4 b/s for the Leap.

In the sliders case, this loss of information is rather more catastrophic, as there was far less information there to begin with. Stated in relative terms, the Leap's decline for lengths 1, 2 and 3 compared to guided sequences was 5%, 11% and 26% . Using the sliders, TP was reduced by 21%, 52% and 73% respectively. So the sliders suffer disproportionately for longer sequences, supporting the hypothesis that by placing more demand on working memory, concurrent tasks suffer as a result.

Fig. 8.11 shows the memorised trial results, again plotting ISSR against MT. The upper plot reveals that the Leap's ISSR stays very flat for most of the tempo range. This implies that the accuracy of the remembered positions was the limiting

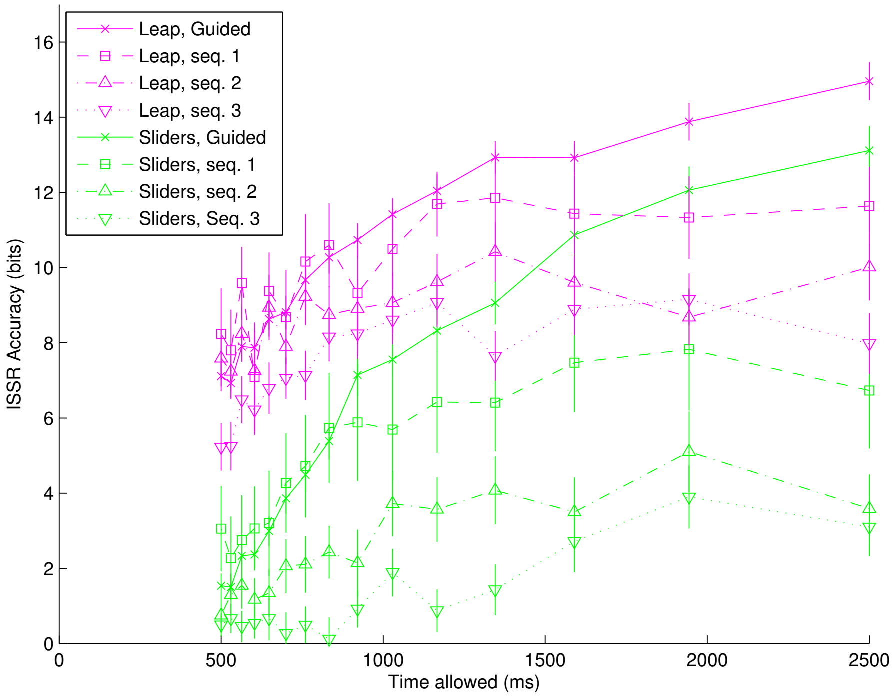


Figure 8.11: Accuracy decreasing with the length of the memorised sequences. For the Leap, the accuracy level of longer sequences remain flat until the intercept with the guided results, at around 800ms/9 bits. This suggests that it is the accuracy of the memory that is the limiting factor, rather than speed of recall or manipulation.

factor for the slower speeds, rather than the speed of recall or manipulation. For example, the 2 sound sequence was achieved to an accuracy of about 9 bits given 760ms. About the same accuracy was achieved for all MT between this and 2.5s (9.5 bits). Therefore it was the memory itself that, on average, contained 9 bits; only with movement speeds below 760ms would this show any deterioration. Whilst the sliders curve shows a similar phenomenon for the 1 sound sequence, the curves for lengths 2 and 3 tend to decrease more steadily as tempo increases. This leads to the conclusion that the loss of information was related to the latency of recall, as opposed to the accuracy of the memories themselves, or the speed of physical manipulation.

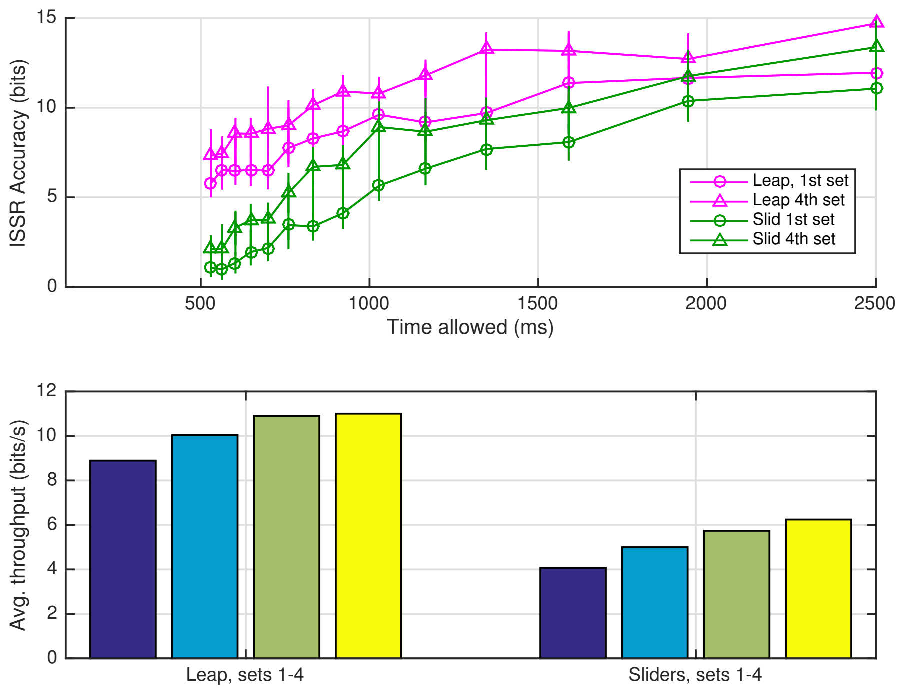


Figure 8.12: Performance improvements with practice, averaged across all guided trials for each of the 4 sets. Upper plot shows detailed differences between first and last set. The lower bar chart shows that, in absolute terms, the two interfaces improved by roughly the same amount: about 2 bits/s. The sliders improved more in relative terms.

Fig. 8.12 shows how performance improved with practice, with throughput averaged for all trials for each of the four sets. In absolute terms, the two interfaces improved by roughly the same amount: about 2 bits/s. The sliders improved more in relative terms however, from 4 to 6 bits/s. For the memorised sequences (Fig. 8.13), progress was very similar: improvements of 2 bit/s for both interfaces. However, closer analysis of the upper plot indicates that the sliders tended to show larger improvements at slower tempos—in fact, almost catching up with the Leap for the 2500ms trials—but this 5 bit improvement did not translate to the faster results. Again, this implies that it wasn't simply that slider positions are intrinsically harder to encode in memory (given enough exposure, almost anything can be memorised),

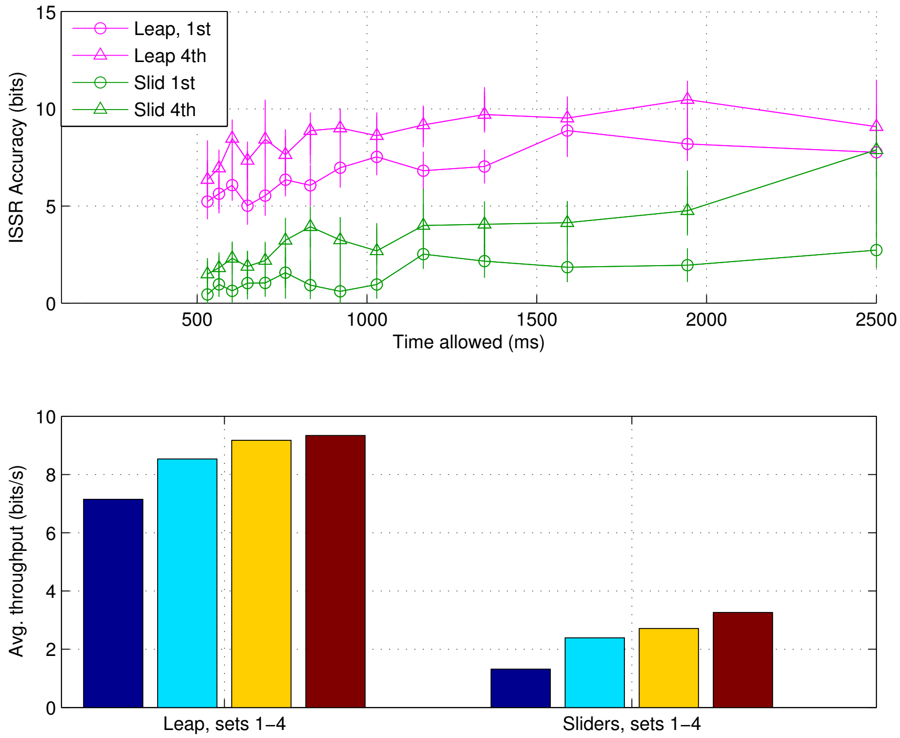


Figure 8.13: Performance improvements with practice, averaged across all memorised trials. The upper plot shows detailed differences between the first and last set. The upper green curve (triangle markers) reveals that the sliders improved most at slow tempos, but this improvement in accuracy did not translate to a similar increase at fast tempos. The Leap improved more evenly across the tempo range.

it was also that recall took longer, due to how the information was presented and encoded. These results support the hypothesis that using hand pose information significantly decreases the time taken for the participants to access memorised parameters.

If participants had already completed a session with one interface, then presumably this would improve their performance with the other interface in the second session, due to familiarity with the format of the task. Surprisingly, this was an asymmetrical effect. Figure 8.14 shows carry over effects due to interface order. Experience with the sliders helped the subsequent Leap session by 1bits/s (for guided trials), however experience with the leap did not help with the next day's guided slider performance. Obviously this effect is small when compared to the overall difference between interfaces (5bit/s), so controlling for this effect is not essential. Unguided sequences improved by around 0.5bits/s for the second session regardless of interface order.

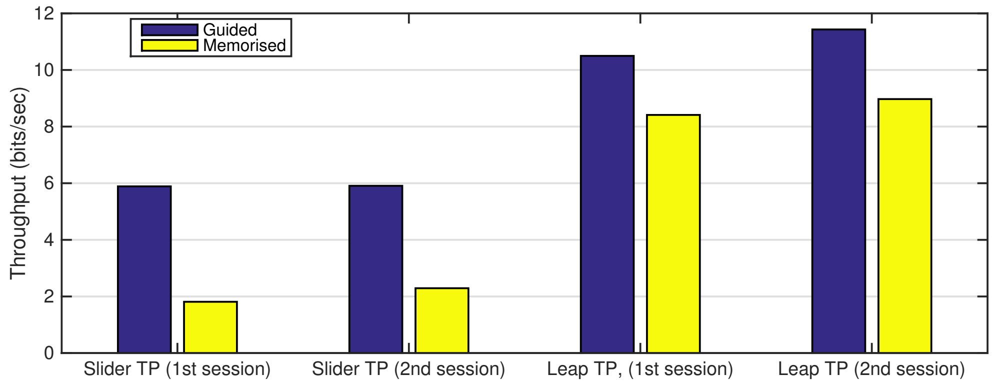


Figure 8.14: Effects of session ordering, in terms of average TP. Participants who used the Leap first showed no advantage over those who used the sliders first in terms of their slider performance (first 2 blue bars), but those who used the sliders first showed a slight advantage in the Leap task (blue bars 3 & 4).

It was expected that the reduced cognitive load when using the hand tracker would result in users being better able to correct mistakes when repeating the same sequence. In other words, they would be able to reflect upon their results and improve them the next time round. In fact both interfaces showed similar improvements for the second repetition of a guided sequence, and slight worsening when repeating a memorised sequence (Fig. 8.15).

Finally, we look at the effect of preset positioning on performance. A number of users remarked that some locations were harder to physically reach than others, and some were harder or easier to remember. The effects of this can be seen in Fig. 8.16, which shows the average absolute distance to target for each preset type, for guided and memorised trials. Kick drums, which tended to be located in the rear left corner

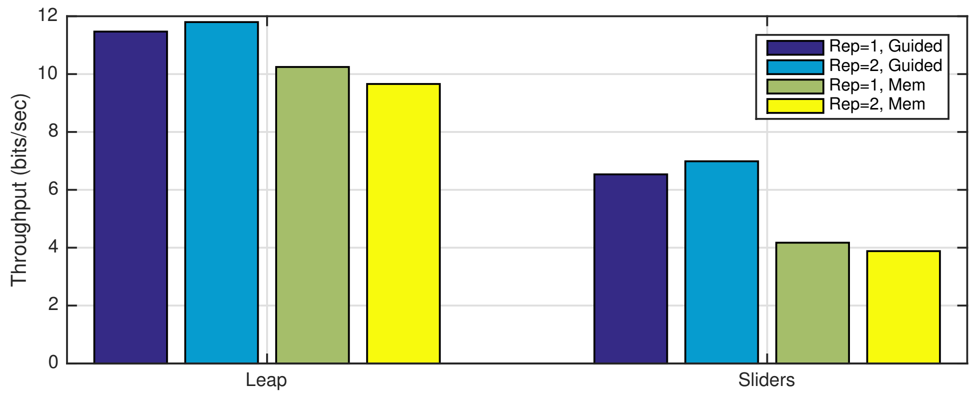


Figure 8.15: Difference in TP for repetitions of the same sequence. Both interfaces show slight improvements when repeating guided sequences, but a small decrease for repetition of memorised sequences.

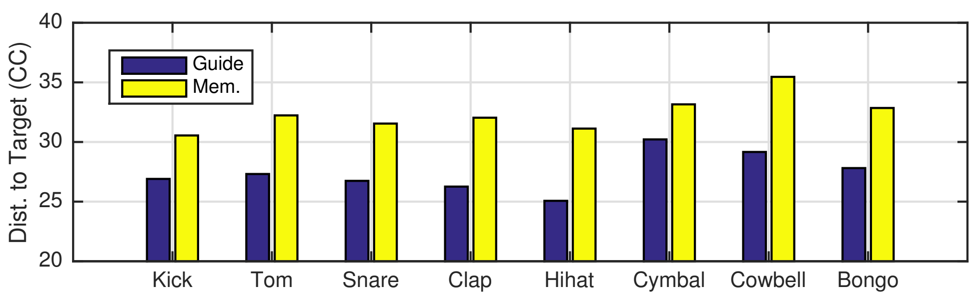


Figure 8.16: Different drum sound categories had somewhat different average accuracy levels and different recall accuracies, due to their location in the hand pose space. Accuracy is given as absolute Euclidean distance (lower is better).

of the space (see Fig. 8.2), tended to be close to the left and rear walls: how the hand abutted these walls could be recalled fairly accurately. This leads to the difference between guided and memorised accuracy being small (3.5CC distance increase when guides removed). The cowbell, in contrast, tended to be at the front right, with no features behind it to act as a reference, therefore recall suffered (6.3CC worse when guides removed). The lowest absolute guided accuracy was for the cymbal, this tended to be right at the front of the space with a large clockwise yaw (a physically awkward manouvre), and also near the edge of the tracking volume and pitched toward the line of sight of the tracker (yielding unreliable hand pose estimation).

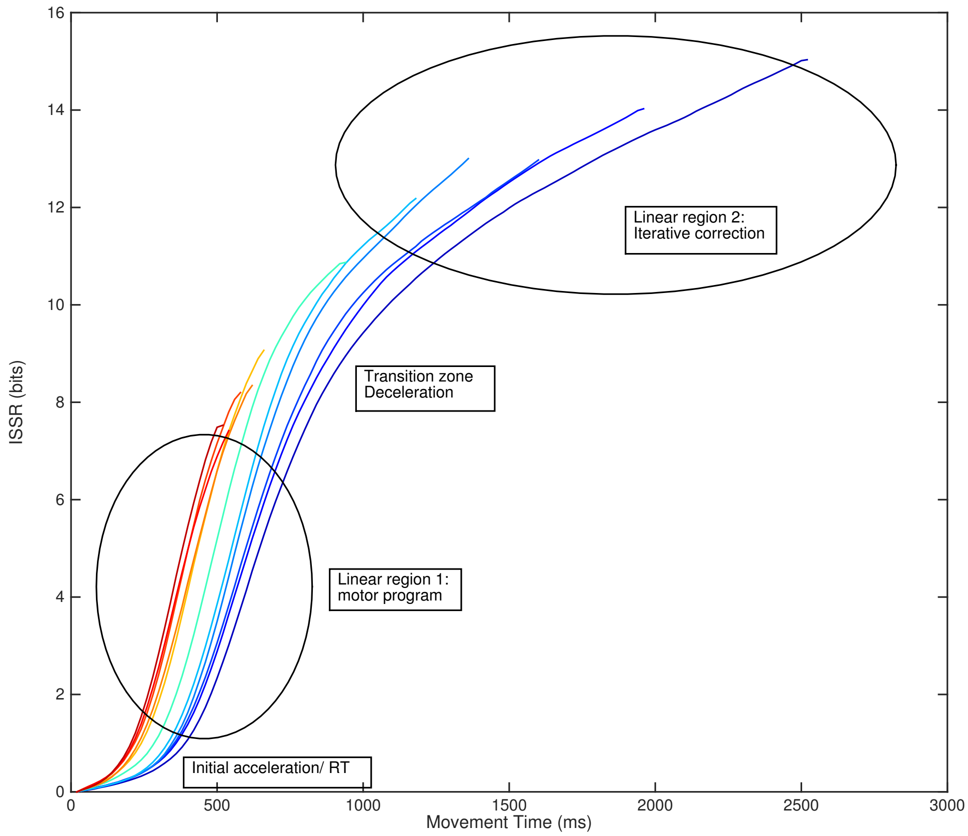


Figure 8.17: Averaged trajectory curves, colour indicates tempo (blue slowest, red fastest). There seems to be two approximately linear regions. The first (region 1), corresponds to fast ballistic motion toward the approximate target. This seems to be adequate for accuracies of up to around 8 bits. The second linear region only appears for movement times slower than 1 second and accuracies beyond 10 bits.

Hence this region's accuracy suffered irrespective of the presence of a visible guide.

8.5.3 Trajectory Analysis

In this section we use the trajectory data to investigate the physical movements in more detail. As in Expt. 2, the trajectory data was thresholded and converted to ISSR. All the different trajectories for a given tempo level were then resampled and averaged for each time point. Unfortunately, the early trajectories for the sliders are often not present, due to the fingers not contacting the screen; nevertheless, the curves were surprisingly similar to the Leap trajectories. Results for the guided trials for the Leap are shown in Figure 8.17. Note that the shapes of these curves result from and average of many trajectories, so some features may be have more of a statistical origin rather than being features present in every trajectory. Lines are not straight, so a regression is not applicable, however there do seem to be two zones that are approximately linear. The first is a rapid increase, achieving a gradient (throughput) of about 25b/s. The second, towards the end of the slowest tempo level's curves, is slower with a gradient of about 3b/s. These plots seem to reveal two different phases, the first being fast ballistic movement, the second using slower iterative corrections. Faster tempos yield different trajectories: the acceleration is sharper, the gradients of the curves increase, and the final corrective phase is eliminated entirely. This shows that movements were being adjusted according to the allowed time. Note that the end points of the curves—the ISSR at the end of the movements—give exactly the data points for the plots in the previous section (e.g. the end points of the lines in Fig. 8.17 give the shape of the Leap's plot in Fig. 8.7).

Figure 8.18 shows the trajectories for the memorised targets (all trials of sequence lengths 2 and 3 were included). Here, we see the very similar final accuracies for the slower trials noted in the previous section, around 9 bits. Subjects respond to this limitation by stretching out the trajectory curve along the time axis, presumably to minimise the force needed to carry out the movement. In theory, one could use the fastest curves to start the movement and then begin corrections earlier, but since there is not sufficient knowledge of the target, this would be inefficient. It appears that even at the start of the movement there is already an estimate of the accuracy of the memory, and the movement is scaled in time accordingly. This scaling is seen down to approximately 900ms. At this point, there is no longer enough time to reach that accuracy level. The solid red line is the curve for the fastest guided trajectory,

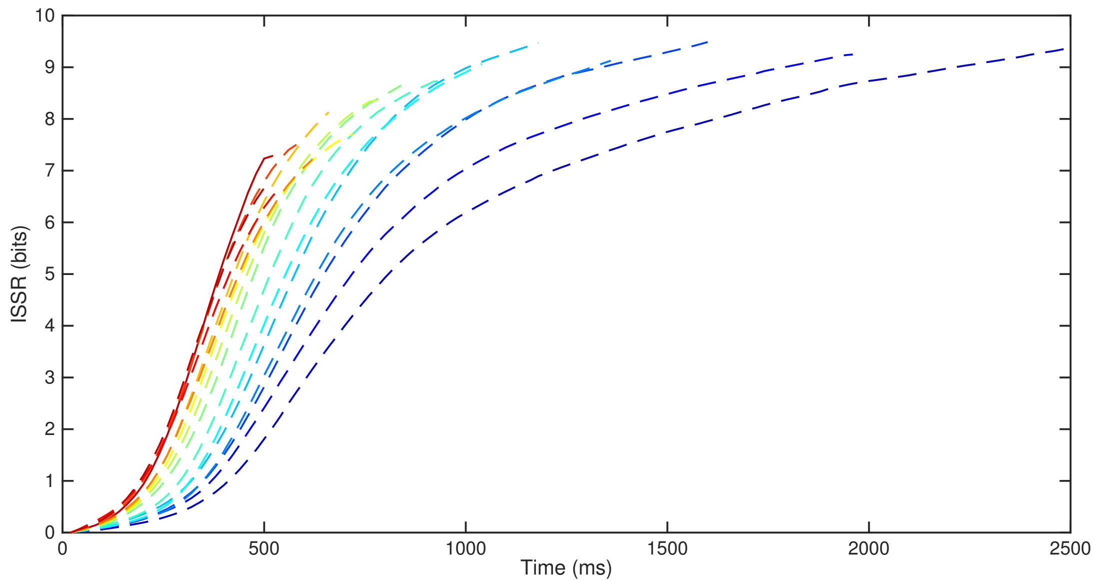


Figure 8.18: Averaged trajectory curves for memorised sequences (lengths 2 and 3 averaged). Participants appear to stretch the movement curve in time. This suggests a prediction of a realistically achievable accuracy is taken into account during the movement preparation stage, and effort is minimised accordingly.

showing that the fastest guided and memorised trajectories are virtually identical (at least when averaged), with less than 1 bit difference, only diverging right at the end of the curve. This is interesting as it would imply that either the time for recalling the memorised target and processing the visual information is exactly the same; or, more likely, that using proprioceptive target memory was faster, and therefore subjects were not using much detailed visual information at all[7].

A number of users mentioned that they found rotation harder to match than position. To see this effect, average ISSR trajectories for each 3D quantity separately are plotted in Fig. 8.19. Rotation achieves just 60% of the information accuracy

[7]A third explanation is that physical limitations are dominant here, and cognitive differences are not important: this would mean the 150ms time axis intercept discussed in the previous section should not be interpreted as cognitive latency: rather it is an upper limit for the frequency of arm movements. However the 400ms intercept for the sliders is surely not related to this same limitation, and as noted these values agree well with the motor program preparation time based on the number of features.

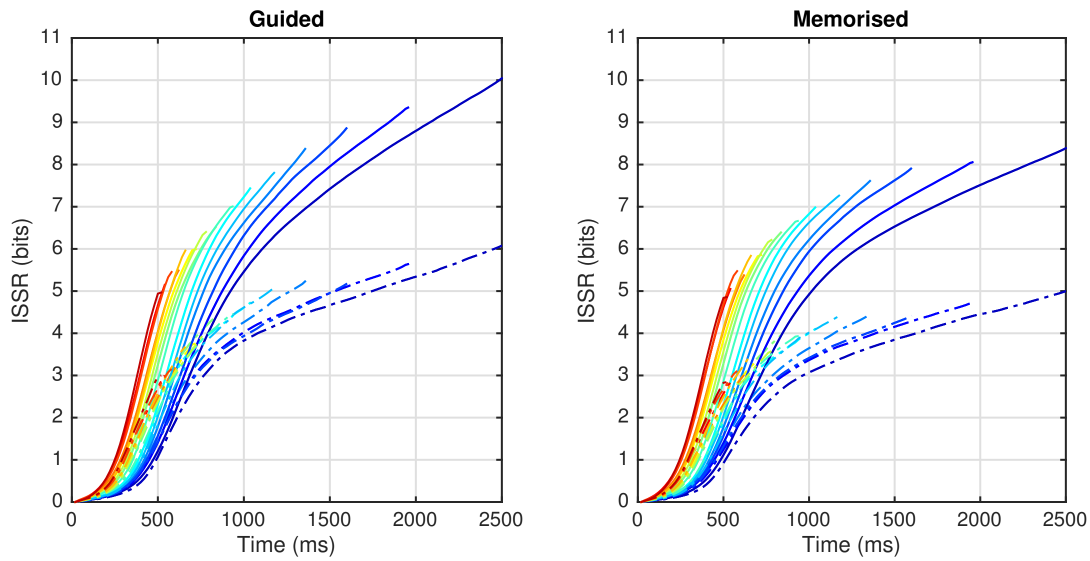


Figure 8.19: 3D search space reduction trajectories for position (solid lines) and rotation (dotted lines). Rotation is consistently about 60% less accurate than position.

of position. This ratio seems largely independent of movement speed. This effect could have been due to the difficulty of aligning the visual blocks, but presumably visual feedback becomes less important as the 8 positions are learned. Alternatively there may be a tendency to prioritise position over rotation, certainly hand position is a more essential quantity for common tasks. It maybe simply that the movements were smaller to control rotation, and subject to more noise. Another factor could be a number of extremely poor results due to tracking loss: where the location was correct but the orientation of the hand would be flipped. Another interesting feature of this plot is that the linear corrective phase seems to be entered slightly earlier for rotation.

If the hypothesised switch between fast ballistic movement and slower corrective behaviour exists, then we would also predict that the coordination between the different dimensions decreases during this phase: it is harder to make deliberate corrections in 6 dimensions simultaneously. To affirm this, the trajectories were split into 6 time windows, and the average correlation between all pairs of dimensions

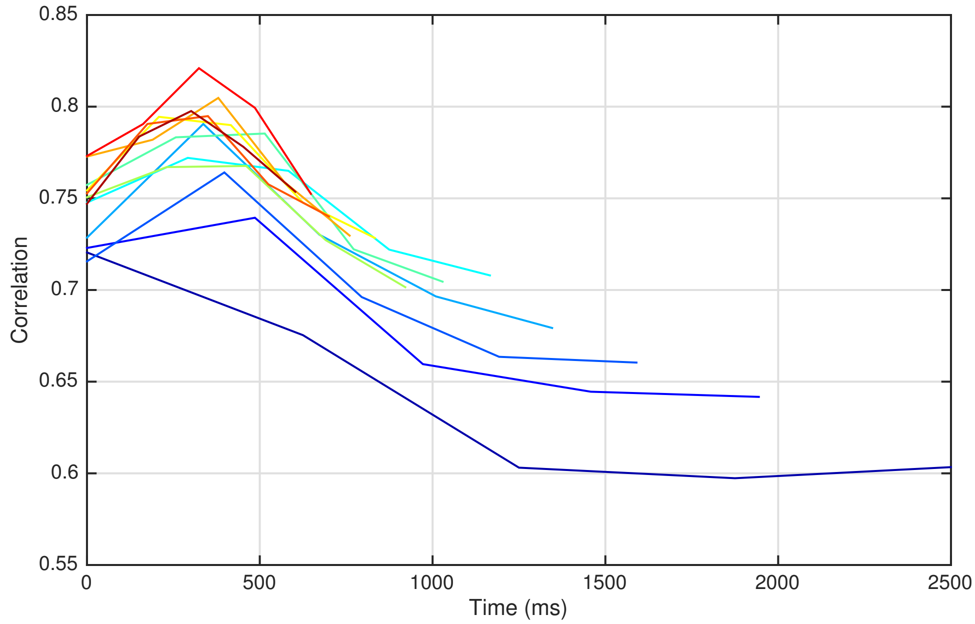


Figure 8.20: Average correlation between all pairs of dimensions for Leap trials. Trajectories were split into 6 time windows, and correlation between dimensions calculated within each window. The time axis indicates window start time. Colour indicates tempo level (blue slowest, red fastest). As expected, correlation (diagonality) tends to be high during the early ballistic phase of the movement, and then reduces in the corrective phase. The faster the tempo level, the more coordination.

calculated within each window. These correlations were then averaged across all trials of a given tempo. Thus, a correlation of 1 would indicate perfect synchronisation of progress in all dimensions, i.e. a perfectly straight 6D diagonal line. Fig. 8.20 shows that coordination does indeed vary as expected, from around 0.8 for the beginning of movements (this peak consistently occurring around 400ms) to around 0.65 for the end of slower movements. Furthermore the faster the movements, the higher the coordination.

8.5.4 Schmidt or Fitts?

The plots of ISSR against MT presented in the previous section are clearly not straight. This could be due to saturation effects, reaction times at fast speeds and

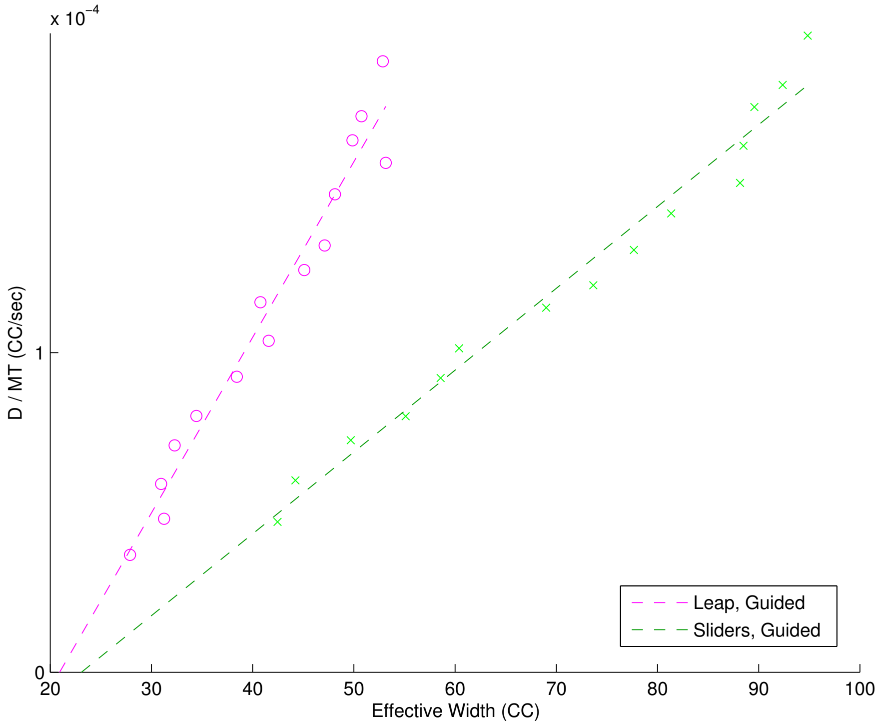


Figure 8.21: Plot of the linear relationship between effective target width and movement velocity (for all guided trial results). Faster tempo trials are towards the upper right. This confirms that tempo based movements are better modelled by the "Schmidt paradigm" than Fitts' law.

accuracy limitations at slow speeds. Or it could be a statistical artefact of averaging of many curves with two linear regions: these regions transitioning at different times according to some distribution. More likely, however, is that movements conducted in a fixed time interval require a different model, due to a different movement control strategy i.e. open loop control results in the linear relationship as discussed in Section 2.4.1.

This relationship is formulated as

$$W_e = K_1 + K_2 \frac{D}{MT}, \tag{8.1}$$

where W_e is the effective target width, calculated as the standard deviation of the finishing position (in more than one dimension, this can be calculated as the square root of the mean squared Euclidean distance to the target), D is the initial distance to target, MT is the movement time, and K_1 and K_2 are constants. In other words, accuracy is proportional to movement velocity. Plotting W_e against D/MT should therefore give straight lines. Figure 8.21 shows that this model is indeed well supported by the data. For the sliders it seems particularly surprising that this rule would hold.

One would also expect, given the seemingly two stage process observed in the trajectories, that the slower trials nearer the origin would deviate from the straight line (curving upwards), but this seems not to be the case. Values for the regression constants were $K_1 = 20.9$CC, $K_2 = 1.82 \times 10^5 s^{-1}$ for the Leap, and $K_1 = 23.1$CC, $K_2 = 3.91 \times 10^5 s^{-1}$ for the sliders. It is interesting to note that the K_1 intercept is almost identical: this value might be interpreted as the final accuracy for an infinitely slow trial. The gradient K_2 expresses how the accuracy varies with absolute velocity: the Leap's inaccuracy increases at half the rate of the Sliders.

ISSR now depends on d_2. This is somewhat unfortunate as *ISSR* now depends on the absolute results of the trials, rather than being a relative, scale-free quantity. This dependence also means that we cannot switch back from this linear relationship to extrapolate what throughput curves would look like outside this time range, or with different constants.

8.5.5 Subjective Experiences and Questionnaire Responses

The slider task was immediately perceived by participants as far harder, both physically and cognitively, than the Leap. Many users expressed a certain amount of trepidation as to what they were expected to perform. This contrasts with the previous experiment where the participants initially seemed more comfortable with the

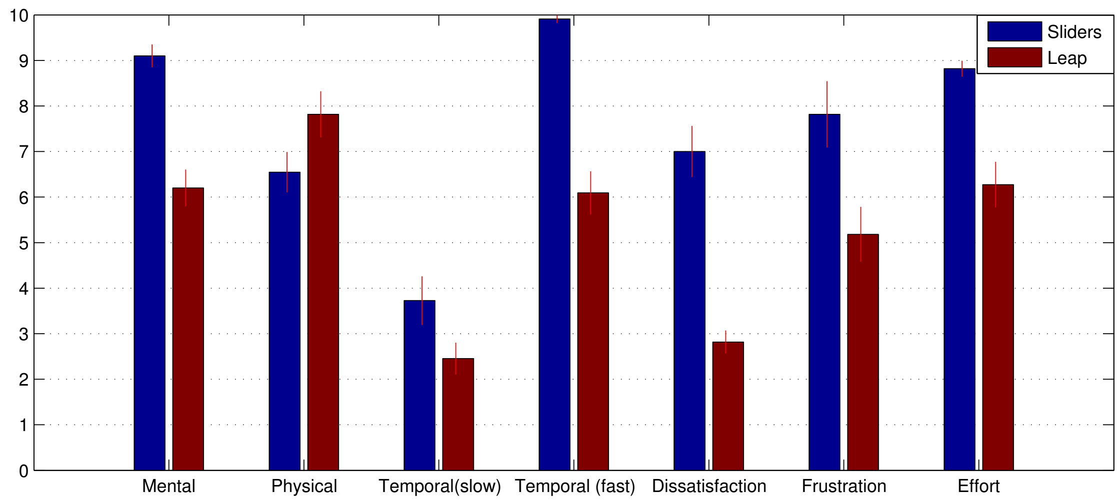


Figure 8.22: Averaged responses for estimates of workload. Standard errors are shown as red whiskers. The only aspect rated higher for the Leap was physical demand.

sliders, despite comparable measured performance with the Leap.

Other advantages and disadvantages of the two control types, not immediately apparent from the data, were noted during this experiment. Many people noted that some slider positions were much harder than others, particularly those with large variations from the centre line, and those that involved a high-low-high shape for the index, middle and ring fingers. One female participant felt that the sliders were too large for her hands. There was also an issue that some hand positions could obscure the position of the guide bars. Therefore a portion of the sliders' performance deficit should be attributed to their unsuitability for simultaneous manipulation. The Leap also suffered from some issues, including arm fatigue, and occasional, but serious, loss of tracking at the extremes of the tracked region.

For assessing subjective experiences of workload, the standardised NASA TLX questionnaire was used [Hart and Staveland, 1988], all questions being asked for both interface conditions. The one alteration was to separate the "temporal demand" into two questions, one relating to the slowest tempo and one to the fastest. This was to see if the subjective results matched the quantitative result that the Leap showed an increased relative performance advantage for fast tempos. Each aspect was rated from 1 to 10, and are treated as continuous variables. Mean scores for the sliders are reported as M_S, and for the Leap as M_L.

Figure 8.22 shows a bar plot of the average response to the TLX questions. As expected, the Leap was rated as lower demand across almost all aspects, resulting in average workload score of 7.6 for the sliders and 5.2 for the Leap ($t(11) = 3.3$, $p < 0.01$). The one exception was for physical demand ($M_S = 6.5$, $M_L = 7.8$), not surprising as it is difficult to hold the arm in the air for long periods of time. Interestingly, the biggest difference was for how happy the participants felt with their own performance ($M_S = 3.0, M_L = 7.1$)[8].

The other large differences were for mental demand ($M_S = 9.1$, $M_L = 6.2$), and temporal demand at the fast tempo ($M_S = 9.9$, $M_L = 6.1$). Temporal demand for the slowest tempo ($M_S = 3.7, M_L = 2.5$) gave a smaller difference than the fastest tempo, agreeing well with the throughput results.

The leap was rated as being fairly frustrating ($M_L = 5$) despite better performance than the sliders ($M_S = 7.6$). This is probably mainly due to issues with the errors in rotation tracking mentioned above. An intermittent catastrophic error can often be more frustrating than a predictable performance deficit, particularly for musical interaction. "Effort" was still rated reasonably high for the Leap ($M_S = 8.8$,$M_L = 6.2$).

To what extent were people aware of the sound aspect of the task? In theory, one could mute the drum sounds and perform perfectly well along with just a metronome click, so it was necessary to ask the participants to report their level of attention to the audio. Ability to engage with the sounds should also be an indicator of spare

[8]Note that this feature is inverted for the plot and the overall average, i.e. happiness is subtracted from 10 and becomes a 'dissatisfaction' rating

cognitive capacity. Users were asked "How often did you notice the sounds you were making with the Sliders/Leap?" and rated this from 1-10. As expected, the sounds were noticed far more when using the Leap ($M_L = 6.8$, $M_S = 3.2$). Another event not deducible from from the data was if participants forgot the order of the sounds entirely (the secondary task). To ascertain this, the question asked was: "With the Sliders/Leap, what proportion of the time do you think you had forgotten the 2/3 sound sequence of sounds entirely?". The Sliders were estimated as causing forgotten sequences 71% of the time, contrasting with 35% for the Leap.

An important aspect of this training task which was not tested was how well the better short term memory results would translate to long-term memory. Participants were asked "If you tried to remember the settings of your sounds now, which interface would be easier to visualise?". Responses were rated on a scale of 1 (definitely the Leap), to 10 (definitely the sliders). The ViBEAMP visualisation was strongly favoured ($M = 2.1$). A similar scale was used for the question "Which interface felt more rhythmical/flowing?". Again users responded strongly in favour of the Leap ($M = 1.8$) with all but two respondents selecting the extremal value.

Further questions featured text responses, to attempt to find out how people reacted to the task on a higher, reflective level. Users were asked what strategies they developed to improve their performance. Many strategies were reported for the sliders, but less were reported for the Leap, probably because there was less need for them. With the sliders, users immediately discovered that setting each control one by one was not a viable strategy, and some kind of simultaneous operation was necessary. Most strategies involved learning the shapes that the bar heights formed: a good example of attempting to chunk separate quantities into a holistic unit. Many users reported grouping the positions into one shape per hand, each hand being responsible for 3 sliders. When the tempo became too fast some said they would simply ignore some of the sliders. Several users tried to identify the controls that moved the most from the central neutral position, and only concentrate on those. One user, a pianist, and the best performer on the sliders by some margin, sometimes rotated the hand to approach the sliders from the top of the screen in order to set the high-low-high groups, and sometimes used groupings of 2 and 4 sliders per hand, instead of 3 and 3.

With the Leap the strategies tended to be less about chunking, and more about finding analogies for the gestures. For instance, "the change from the neutral position to the tom position was like a plane taking off and banking to the left. Such patterns were easy to remember", and "I also found trying to imagine 'hitting' the gesture helpful with my timing as it made more of a dance—as opposed to move my hand into position as soon as possible and holding it until the beat.", this comment would suggest that trajectory stretching results in section 8.5.3 could be explained by this rhythmic interaction strategy. One user found that it helped to interact with the hand tracker standing up (it was less fatiguing if the hand was lower relative to the shoulder). Another made use of their "phonological loop" to repeat the names of the drum sounds to themselves in order to remember the sequence.

Participants were also asked "Did you notice any differences in your strategies between the slowest and fastest tempos?". Ideally we would like to find a subjective correlate of the increasing bandwidth for the Leap at higher tempos, whether there was any further chunking going on for the whole sequence, or whether people could feel their proprioceptive memory coming into play. One response seems to indicate that a short-term proprioceptive memory might play a part when gesturing rapidly with the Leap: "when the tempo was slow, I paid more attention to what I saw. when the tempo was faster, I paid more attention to where the previous position of my body was.". Other users remarked on the fact that they were starting to connect the discrete targets into continuous gestures: "I was more able to remember the sounds as a pattern which I then started to remember how to 'play'. At some of the faster points I started to keep my hand moving so that it travelled through the measurement point at the right time.", and another participant: "Faster tempos had a more fluid movement, and it was then that I started to think about how to play the musical pattern rather than where to put my hand.". This respondent noted the more direct link between body movement and rhythm: "With the leap, at higher tempos it felt more like playing a 'traditional' drum controller, where the movement was directly related to the rhythm. This was not the case with the sliders". One participant noted that using the Leap at slower tempos left enough spare cognitive capacity to be able to plan a more effective strategy for the task: "I was more creative with the tasks in the slower Leap sessions (i.e. increased use of 'ping pong technique'), which may have led to a better score in the faster tempos" (note the use of another movement analogy). This provides support for spare cognitive capacity enabling reflective divergence. One participant identified the transition to slower corrective adjustments during longer movement times: "At the slowest tempo there was more time for small adjustments, whereas faster tempos where thought of as one continuous movement.". Most subjects remarked they ceased to consider what individual controls were doing to the sound when carrying out the matching task, though one user claimed they could specifically correct for pitch when repeating sequences using the Leap.

Another aspect that was hard to ascertain was which interface was better for finding sounds in the exploratory session. Opinion was more mixed here. Participants were slightly in favour of the leap (M = 5.8), but three users strongly favoured the sliders. A text response question revealed that it was previous knowledge of how to construct certain drum sounds that the users felt made the sliders preferable for this task. For instance, "Some of the sounds like kick drum and hi-hats I have produced many times on conventional/software synths. It was quite easy to make those using sliders. For exploration however I found leap more interesting since the space is not explored yet. The problems arise with leap once small changes need to be made. For me it will take more time to adjust to 6 degrees of freedom using the hand.". One participant remarked that the predictable parameters (pitch, decay etc.) were better controlled by a slider, but the unpredictable controls (FM noise) seemed easier to explore with hand poses. Another difficulty noted with the FM feedback component was that small adjustments sometimes had a very large effect on the sound. In situations like this it was impossible to reach a stable sound with the Leap, or find a good sound again once the hand had drifted away from it. This problem particularly applied to metallic sounds, like the cowbell, which were often sought on the narrow borderline between pitched and chaotic oscillation. It should be noted that some participants failed to realise that certain dimensions existed (particularly yaw) when exploring the space. Perhaps hand poses are not good for exploratory search, as they tend to produce somewhat stereotyped behaviour.

A final question was if the participants themselves felt that the task was a fair comparison between the interfaces. Answers were on a 1-10 scale from strongly agree to strongly disagree. Opinion was quite ambiguous here ($M = 5.8$), and 4 responses were strongly negative. This is probably because the task was heavily weighted towards performance, and inappropriate for other stages of music making. The final question read "Do you think this type of task could be a helpful way of practising and improving musical performance with a hand tracker?". Answers ranged from 1 (Strongly disagree) to 10 (Strongly agree). An average of 7.5 indicates that users generally perceived the task as musically relevant.

A final text box was provided for any other comments, one user remarked: "The sliders provided a pretty hectic and unpleasant experience whereas the Leap was fun and engaging to use. I wanted to play more after the experiment whereas after the sliders I NEVER wanted to use sliders for anything ever again.".

8.6 Conclusion

By designing a musical task based around fast, automatic motor control, the true strengths of multidimensional controllers are revealed. The distinction between slow and fast thinking drawn in the EARS model has proved useful in motivating this experimental design. By demanding the interaction be fast, and providing a number of parameters in a way the sensorimotor system can understand, we have utilised a number of autonomous brain systems to carry out a search space reduction problem much faster than standard interface paradigms allow. So, this experiment has demonstrated that the "cognitive pipelining" design principle can work well for musical interfaces. It provided qualitative measurement of the resulting speed up, and both the data and the survey responses fitted well with the hypotheses.

The study also provided evidence that it is the linear, Schmidt model that governs the speed-accuracy tradeoff for rhythmic, timely musical interactions. This is in contrast to Experiment 2, which seemed to support the logarithmic relationship of Fitts' law. So it seems that, depending on whether the goal constraint is time or accuracy, the brain proceeds using a different strategy: i.e. fires a different motor program. If the task is time constrained, and an appropriate learned motor program is available, an open loop production will be fired. If, on the other hand, the subject is faced with an accuracy goal constraint, or no previously practiced motor program is available, then a closed loop production is fired, and sensory feedback is used to guide the system to its target. The analysis of the trajectories seemed to show that, if the time constraint is long enough, it is possible to start open loop and finish closed loop.

Judging by the effectiveness of Fitts' law in describing many rapid one and two dimensional tasks, it may be that closed loop interaction is usually optimal in low-dimensional situations. Figures as low as 200ms have been claimed as being the threshold for switching between the two strategies [Meyer et al., 1982]. On the other hand, figures as high as 600ms have been quoted when there is a time rather than accuracy constraint. For this 6-DOF study, the transition between the two approaches occurs at around 1 second, so it could be that increasing the dimensionality of the task lengthens this threshold. This would make sense if closed loop interaction is less effective at optimising many dimensions simultaneously: the higher the dimensionality, the longer the feedback loop, and the greater the benefit gained from eliminating it. The decrease of the coordination between dimensions over the duration of the slower trials, where users switched from ballistic to corrective motion, seems to further support that iterative corrections focus on fewer dimensions.

The current version of the Leap Motion proves a useful tool for investigating this form of interaction. However, it is far from the perfect hand tracking solution: the chief problems being occlusion of its line of sight making hand pose unpredictable, and a decrease in reliability at the limits of the interaction space. This could be alleviated by using multiple devices set up orthogonally, and the data integrated according to each stream's confidence rating (currently impossible without using multiple computers). It would open up many further gestures if the two hands could reliably make contact with themselves and each other. Future devices will no doubt show improved tracking performance [Rautaray and Agrawal, 2015].

Other problems may be unavoidable with free-space gestures. The first problem is disengaging with the device. It is hard to remove your hand from the interaction volume without that movement being interpreted as a gesture itself. There is currently no way of switching between coupled and decoupled control. There may be ways to deal with this, for instance by using a foot-switch, or perhaps closing the hand. Another design issue related to disengagement is how to provide the option of locking parameters that you do not wish to change. With 6-DOF interaction it is rather difficult to swap to a fine-tuning mode in lower dimensionality, as you are constantly engaged with all 6 parameters. The major advantage of the sliders in the exploratory stage is how easy it is to focus on a subset of the parameters and ignore the rest. Their stability when making very small adjustments is also essential. Perhaps future versions of ViBEAMP could implement a zoom function, and switches available to the other hand to mute and solo different parameters. The second major problem is fatigue, this was the only aspect rated worse than the sliders in the questionnaire. All participants experienced discomfort from an hour of interaction, despite taking breaks during the preview sequence and between runs. Hand pose interaction probably compares unfavourably with keyboards or physical controllers in this regard, but perhaps no worse than supported instruments such as violin and flute. Standing up seemed to alleviate this problem slightly. Twisting the hand caused issues for both exploration and performance, therefore it is recommended that yaw should be avoided, or at least assigned to a less important parameter.

It is interesting to note that if the experiment was conducted without very fast tempos, and with no working memory task, hardly any difference between the devices would have been observed. Fig. 8.11 shows that at slow tempos (2.5 seconds per match) and single sound sequences, the sliders achieve 13 bits of accuracy and the Leap 14 bits. In contrast, at around 1 second matching time and a 3 sound sequence to memorise, the sliders are reduced to 1 bit and the Leap maintains 9 bits. Therefore it is only when introducing cognitive demand and temporal urgency—precisely the conditions pertaining to musical performance—that dramatic discrepancies are seen. This explains why DMI research might reach conclusions at odds to other HCI user studies: in many HCI studies the user is never forced out of the "comfort zone" where they can rely on visual feedback, and hence operate within the Fitts paradigm. Nevertheless, these results are also expected to apply to hand-pose matching for other, non-musical applications, in particular fields where time based interaction is crucial: such as animation or computer games.

In conclusion, this experiment illustrates that it is possible to conduct research that both solves some of the problems in musical interaction design, as well as testing and confirming deeper hypotheses about the underlying cognitive principles involved.

8.6.1 Future Directions, Limitations of The Study

Presumably, if participants were asked to repeat the exploratory stage after their performance training sessions, they would be able to find appropriate drum sounds much faster. In a longer experiment one could imagine revisiting the first task, to detect any change in exploratory strategies with the presence of tacit experience. Another similar open question is whether practice at the metronome based performance task would speed up an unconstrained search for an arbitrary sound in the space, as in Experiment 2. In theory it should help, enabling a fast starting movement toward the approximate sound location, but after this the search would probably proceed at a similar speed as before. For completeness, it would be good to compare six-dimensional sound search results with those from Experiment 2, and also to perform 1, 2 and 3 dimensional matches along to a metronome.

Due to the possibility of using 6 fingers to set the sliders simultaneously, it is claimed that the performance differences to the Leap are principally cognitive, rather than physical in nature. However, as noted, some slider settings were rather awkward, and necessitated further strategies to overcome, such as reallocating fingers to different sliders etc. As well as the physical awkwardness, recall and selection of these alternative strategies may have further slowed the sliders down. This is a limitation of the experiment, in that we cannot be absolutely sure what proportion of the sliders' poor performance was due to this particular physical problem.

It may have been a mistake to show an animated hand during the preview stage, as participants may have unconsciously been imitating that movement, rather than generating their own strategy. Whilst this would be an interesting result (and encouraging in terms of basing our instrument design on the power of imitation), we cannot say for sure, but seeing as the results agreed with prior Schmidt paradigm experiments, presumably the movement was still natural. It would be interesting to carry out the experiment with no animation, or perhaps a number of different animation curves, to see if any imitation effect can be seen.

Another limitation of this experiment was that many design features were not be tested in isolation, for instance it is not clear how much the association of the visual texture with the presets aided recall, or how much performance would suffer if all visual feedback was taken away for the unguided matches.

A final concern is that, due to the Schmidt paradigm of a linear relation between W_e and $\frac{D}{MT}$ it is not clear that $ISSR$ is, statistically speaking, the best approach to analyse the results. Whilst the peak throughput is an interesting effect, and this information transfer is perfectly translatable into usable data, the resulting dependence of ISSR with absolute distance is concerning, as one is effectively averaging over different experimental conditions.

For future research, with higher degrees of freedom, it is recommended not to make comparisons with the 'current standard' (sliders or knobs), as it seems fairly certain they will not perform as well, and will simply take up the participants time with an overly frustrating task. Rather, two different high-DOF controller designs should be used to explicitly test hypotheses, and refine the design of multidimensional control systems. For instance, it is not clear if there is a particular advantage to using hand pose, or whether other high-DOF tactile physical systems would show similar performance. Other adjustable experimental conditions could be the provision/non-provision of visual, haptic and auditory feedback, and implementing the system in Virtual Reality (VR)[9]. A ViBEAMP system, ideally, would be rendered in a stereoscopic VR environment, meaning that the user's real hand would be perceived as exactly coinciding with their virtual hand. It is not clear how much extra performance would really be gained in VR, but it is likely that depth alignment would improve [Boritz and Booth, 1998]. It may become possible to construct more informative reference points for the middle of the space, and the ability to move the head around the space may alleviate some problems caused by occlusion in some cases. With the current implementation, users have to perform a body-space to screen-space coordinate transformation[10], in VR the coordinate system would be perfectly body-centric. Full hand tracking, and rendering accurate hand models could make accurately performing more than 20 parameters possible. One wonders how well this method of learning hand poses could be applied to such tasks as learning the Glove-talk system [Fels and Hinton, 1993], where hand gestures were mapped to 10 parameters of a parallel formant speech synthesizer. For the expert user (who was a pianist), around 100 hours were required to speak intelligibly, could this be sped up using virtual hand emulation?

Compositions have traditionally been stored and recalled using music notation. The ViBEAMP technique has the potential to store musical gestures as 3D recordings of those gestures themselves, which could be replayed in virtual reality in the body-space of the user. This is a visual notation that is as tightly related to the motor actions as one could hope for. Whilst it may take many months to learn how to associate dots on a stave to a finger actions on an instrument, even a novice can match their hand to a virtual hand without much effort. Given a particular mapping, one could imagine experts recording their skilled actions for beginners to

[9] The Oculus Rift VR headset was tested for this system, but proved to be bulky and unreliable, and could have caused motion sickness. The main experimental results were obtainable without this extra complexity.

[10] In fact, this transformation seems to be cognitively undemanding: mapping mouse movement to an on-screen cursor involves a similar transformation, and is performed effortlessly.

imitate and learn from.

Another logical extension to this investigation is testing the ability to follow more complex, continuous gestures. Instead of sequences of discrete hand poses, subjects could be shown an animated hand movement, and have to imitate this gesture from start to finish. Again, the guided and non-guided conditions could be tested. The performance evaluation would have to be adjusted so that the distance of the subject's hand to the animated hand was constantly being assessed. A continuous information-theoretic measure of throughput would need to be used, similar to that developed in [Accot and Zhai, 1999].

What about attempting to create meaningful improvisations within the full degrees-of-freedom hand pose space? For two hands the number of dimensions is approximately 40 (potentially realising Pressing's "Imaginary Superinstrument" [Pressing, 1990]). With such high dimensionality, the effects of an arbitrary spatial variation in gesture would be impossible to predict, and effectively random, however, as we have seen, even in low dimensions timbre is still quite unpredictable. Ultimately, exploratory random variations may be valuable: as long as they can be stored, recalled and incorporated into a repertoire. So hand tracking might cater well to both the skilled and exploratory modes. The challenge is the alternation between the skilled and algorithmic modes, which boils down to the ability to select subsets of the parameters to be frozen, or manipulated separately. This is one design challenge facing any further work on multidimensional interaction. For accurate adjustments, the lack of stability for an unsupported arm is likely to be a problem for hand trackers.

It could be argued that for this experiment, only 8 sounds were being used, therefore the actual information content of a target selection was only 3 bits. Whilst this is a fair objection, and in a way we have "cheated" both by providing preset positions and visual targets, there is actually no limit to the number of potential guide positions, and, given practice, no theoretical limit to the number of items that can be stored in procedural or associative memory. Furthermore, the other regions of the parameter space are instantly available for exploration, therefore improvisatory flexibility is there if desired. Since exploration is the principle reason why one would want a large parameter space, it seems justified to include the entire volume in the measure of effectiveness. Memorising large numbers of different hand poses is a challenge, but in fact the guides can be used during performance whenever needed. The major limitation is probably be the distinguishability of very similar hand poses. Possibly guides could be semi-transparent, only becoming more visible as the euclidean distance becomes small, in this way only the approximate position would be need to be memorised, and more precise alignment could be visually guided.

Hand tracking technology will develop, but it will presumably only become more accurate and more reliable, rather than being reconfigured entirely. This gives practised free space gestures more longevity, potentially solving the problem of skill obsolescence that dogs existing controllers. If there is no physical device, then the musician's device manipulation skills will not become obsolete. The mappings themselves may still be brittle, in that whenever a new instrument is mapped with the system the preset space must be created and learned anew; however the skill of hand pose matching should be transferable to new synthesisers, and even other computing tasks.

This kind of system may finally realise the dream of making a direct physical connection to truly complex, yet precisely specified, musical output a reality. Unfortunately, the dream of single handedly *consciously* controlling every aspect of that output in real time is almost certainly an impossible one, due to the apparently serial nature of conscious cognition.

CHAPTER 9

Discussion

In this Chapter, we summarise and tie together the results of the three experiments. In Section 9.2 we discuss some design recommendations that emerge from this investigation. We then discuss where this research could lead in the future, and speculate further on the relationship between throughput, Flow and creativity.

9.1 Summary of Experimental Methodology and Results

The three experiments provide strong evidence that the way musical parameters are presented to an artist does indeed affect the cognitive strategies used to locate target states. If the creative process is considered as nothing more than a sophisticated parameter-space search, this entails that how the interface represents the parameters will indeed have significant effects on the creative process. The brain's internal predictive model of the parameter space is used to generate gestures that will adjust the artefact into a desired state. Thus, the geometry of the mapping between conceptual space, gestural space and the synthesised output has an impact on not just how easy or quick it is to achieve the desired results, but also how parameters are processed mentally, and the forms of the paths through solution space. These search paths can be analysed in terms of search space reduction and information flow, and this analysis reveals differently shaped throughput plots for the different interaction modes.

9.1.1 Evidence for Distinct Modes in Creative Interaction

The EARS model makes claims for there being four main 'modes of thought' behind creative interactions. By dividing creative thought along a divergent-convergent axis and an implicit-explicit axis, it provides an argument about what underlying cognitive principles these modes emerge from, and predictions of how they translate into interactive behaviour. The three experiments fall short of providing comprehensive evidence for all of Chapter 5's theory, but they do support the claim that there are indeed interaction modes that fundamentally differ from the analytic, accuracy-oriented mode. The experiments reveal in more detail what these alternative modes might be, and make significant progress towards quantitatively investigating the parameter space traversal strategies associated with each mode.

The first experiment revealed that the exploratory mode is very different from targeted search. Exploration was hypothesised not to rely on any prediction of the effect of one's actions, therefore parameters designed to be separable and predictable are less appropriate, and indeed may be harmful due to their reduction of explorability[1]. Experiment 1 confirmed this by showing that an unpredictable,

[1]Explorability being loosely defined as the inverse of the amount of effort required to access the entire parameter space.

low dimensional mapping was preferred for exploration over larger numbers of predictable, one dimensional controls.

Experiment 2 revealed that despite their appearance of being predictable and separable, typical synthesis parameters are far harder to predict than one might suppose. Finding a target sound by setting three of these parameters was a challenge even for experienced musicians. Increased numbers of dimensions increases the time to find desired sounds to such an extent that some amount of exploratory interaction becomes necessary. Therefore exploratory interaction is both a creative strategy that encourages novelty, and also a response to a large and unpredictable search space.

Experiment 2 also revealed that the hypothesis underlying Fitts' law — that the human nervous system processes information at a constant rate — still seems to be applicable in target finding situations that are very different to one or two dimensional pointing tasks. For sound design tasks, auditioning of target and current sounds is needed for feedback on progress. The target search is far slower and the search trajectories are far less direct than for pointing, nevertheless averaged plots of bit-accuracy versus time were still approximately linear.

The third experiment revealed that skilled, rhythmic interaction is also fundamentally different from accuracy based target finding movements. Rather than obeying Fitts' law, timely movements obey Schmitt's law — a linear relation between absolute accuracy and movement velocity. This means that when analysed in terms of information input, throughput peaks at a certain tempo, and this peak throughput is considerably higher than the values for slow, accuracy based movements. This experiment showed that it is possible to train people to rapidly select targets in high-dimensional control spaces.

Unfortunately there was no experiment to compare all the creative modes for various numbers of dimensions. If this was done, then the throughput plots could be overlaid and more definitive answers obtained concerning which mode is more

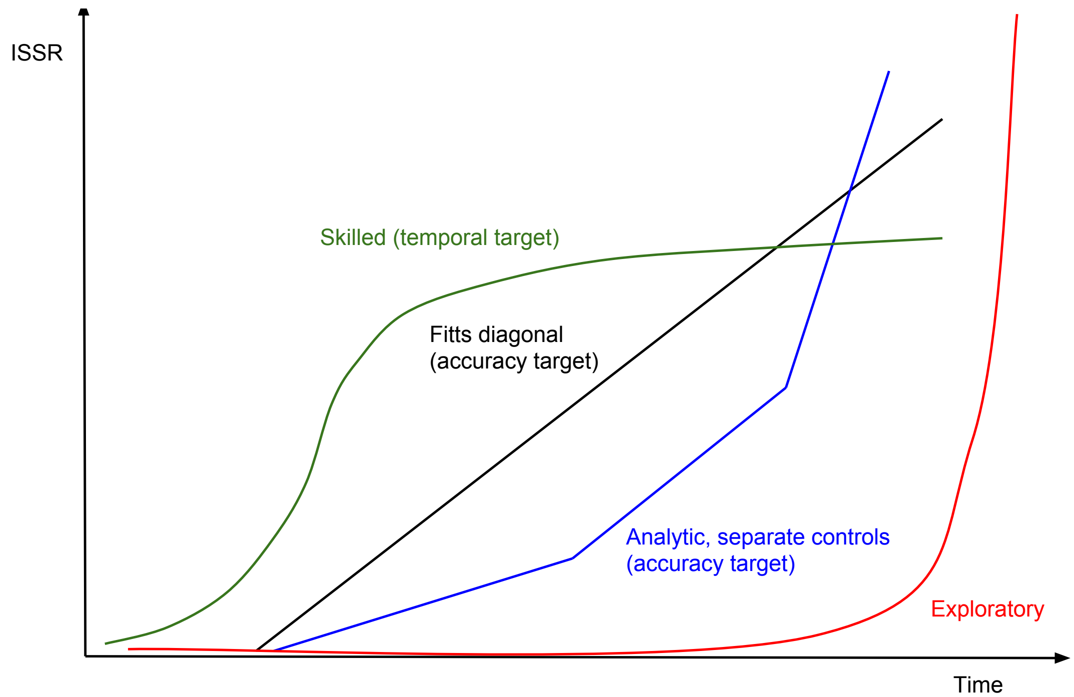


Figure 9.1: Sketch of the information flow signatures of various interaction modes from all three experiments. Gradient gives throughput. Skilled, rhythmic interaction makes rapid initial progress, but levels off at low accuracy (green). Fitts-style movements demonstrate constant information progress (black). If controls are separated, city block navigation means the first movements make slow progress, but the final ones may be more effective and reach high accuracy (blue). Random exploratory interaction makes little progress until the target is stumbled upon, whereupon a large ISSR spike is seen (red). The precise scaling and crossover points of these plots will depend on dimensionality, and probably the specifics of the experimental setup.

effective when, and how this effectiveness scales with dimensionality. Nevertheless we can sketch some speculative estimates. Figure 9.1 shows a rough summary of the shapes of throughput versus time plots[2] for a sound search in 3 dimensions. The curves suggest that different interaction modes have different information processing 'signatures'. For rhythmic open-loop interaction the curve is sub-linear (see Section 8.5.1). For accuracy based Fitts-style interaction the curve is linear for

[2]Note that in experiment 2, time was the dependent variable and was on the y axis, whereas in experiment 3 time was the independent variable so was on the x axis. Therefore experiment 2's curves (separable, analytic navigation) are flipped.

multi-dimensional controllers, but introducing separate controls for separate dimensions introduces kinks due to city-block paths and control swap times. Hence the curve becomes super-linear (see Section 7.4.3). Finally, for exploratory search it is likely that no progress is made for a considerable amount of time, but then the sudden discovery of a suitable target sound results in a large information spike. An interesting topic of further research is whether one could detect the presence of these different modes in a more extended data record of creative interaction, for instance across the creation of an entire piece of music.

It is too strong a claim to make that any of these curves really represent an informational signature distinguishing fast-implicit and slow-explicit brain processes, but this possibility is certainly worth investigating further.

One objection to the EARS theory in relation to experimental evidence is that it would be extremely hard to show that the four proposed modes are exhaustive. They may be neither necessary nor sufficient to form a complete model of creative thought or interaction. It is hard to guarantee that there is no remaining "secret sauce" to creativity that we have not considered. Even assuming the four modes are sufficient, one might ask what additional process is it that decides which mode to use in which situation. This line of questioning easily leads to infinite regress, however. One possible avenue to investigate the sufficiency of the model would be to try to implement these computational processes in software, and let the system try to produce its own music. However this would require implementing a means for the software to evaluate the results of its actions.

9.1.2 Implicit and Explicit Thought Processes

One of the divisions in the EARS model is the difference between explicit and implicit thinking. One of our principal goals was to show that encouraging the use of unconscious, embodied processing can result in measurable speed-ups in the control of digital information. Experiment 3 showed that by utilising a cognitive process thought to be implicit (i.e. the ability to align hand poses) one can design a way to perform parameter adjustments that shows increased speed and lower working memory use. The ability to memorise and perform longer sequences of sound matches using the ViBEAMP technique showed that hand tracking technology can potentially yield interactions that are less intrusive on explicit thought processes. This was reflected in both quantitative measures of throughput and qualitative user feedback.

Comparing these results to Experiment 2, it seems clear that merely providing the potential for adjusting multiple parameters simultaneously will not necessarily result in users doing so. Initially, the searches with the Leap in Experiment 2 were conducted in a city-block fashion. It was only by redesigning the task, and representing the target as a chunked unit (a position/orientation) that this tendency could be bypassed. The major caveat here is that the ability to keep track of and individually adjust single parameters was then lost. It appears difficult to reconcile fast performance-style interaction and detailed sound design in a single interaction style. Either one is using chunked multidimensional processes, or one is explicitly focussing on a single parameter, but not both.

Future work could involve questioning the participant as to what they were consciously aware of during the sound search. Whilst self-report can be unreliable, this may be one way to find out what the explicit system was occupied with, and hence infer what other aspects of the task were being carried out unconsciously.

A more interesting dual task could be used, where the second task could be a more creative one, such as a Remote Associates Test (RAT) [Mednick, 1962]. In this way interference of interface use on more creative thought might be tested.

9.1.3 Quantifying Creative Interface Effectiveness

This thesis has attempted to establish a methodology for evaluating musical interfaces quantitatively and objectively. The main quantities of interest, throughput, explorability and working memory load are summarised below. Throughput, as measured by the rate of search space reduction, could be thought of as a measure of the artist's ability to shape their artistic work by a certain amount in a certain time. Whilst this quantity is probably not a direct substitute for the quality of "expressiveness", it should be clear that it is a good correlate, and certainly a prerequisite, for expressive and fluent interaction. It may be excessively reductionist to rate musical devices by a single number, but if a single number *is* required then surely throughput is a leading candidate.

The proposed ISSR method of measuring the amount of information flow through a device solves some of the stumbling blocks of using Fitts' law for synthesiser interfaces: the unnecessary overhead of attempting to establish movement laws for increasing numbers of dimensions, and the difficulty of establishing a "target size" in a timbre space. ISSR should be applicable to many other input device evaluation scenarios.

Throughput is only valid as a measure of how easily a pre-existing target can be obtained using a device. However, this form of interaction may in fact be a minority case for creative pursuits. Exploratory scenarios without predefined goals are equally important. Whilst speed of manipulation is important here too, measuring progress towards a target is impossible. Therefore evaluating 'explorability' seems essential, but to do this objectively is rather challenging. Explorability was measured in Experiment 1 as being the number of interesting or useful sounds discovered in a certain time using a given interface, however the notion of 'useful sound' is somewhat subjective and may vary across participants, and indeed may vary for a given participant across time.

Another important quantity evaluated was working memory (WM) load. This must be tested alongside throughput, as an interface that provides high throughput at the expense of saturating attentional bandwidth will not count as expressive. With a demanding interface, the artist may have no spare cognitive capacity with which to form intentions to express. In Experiment 3, this was measured as the fall-off in performance when trying to remember sequences of different lengths. For a low WM demand interface, longer sequences should be easier to remember, and hence be performed with higher accuracy. For experimental data analysed using ISSR, this accuracy deficit can also be given in bits. Comparing this deficit for different interfaces gives us the amount of information that was lost from memory as a result of the interface-related cognitive overhead.

This methodology may be extended to investigate other important quantities not specifically looked at in this thesis, such as learnability and retainability of instrument skill. For learnability, one can track how quickly throughput increases with practice. For skill retainability, one would perform repeated experiments, separated by a number of weeks, and investigate how well the practised gestures from the first session were retained for the next. Any drop in performance could again be measured in bits (similarly to the working memory test), giving figures for how well skill can be maintained over longer time periods.

9.2 Design Recommendations

In this section we look at the implications that the EARS theory and these experimental results have on the design of musical interfaces, and make some recommendations for designing them.

For multidimensional skill acquisition, use associative chunks, not predictive dimensions

Access to individual parameters is only required by the algorithmic mode. Neither the exploratory nor the skilled modes should require the user to adjust one thing at a time. Therefore this interaction style should surely be much less prevalent than it is at the moment. This is not a new claim, but Experiment 3 revealed that by deliberately designing an interaction that used 'chunked' information that obscured the individual parameter values, throughput could be increased significantly.

None of the experimental work specifically tested people's abilities to predict the effects of travelling diagonally through arbitrary parameter spaces, but Experiment 2 revealed how hard this was. It seems that *associative* processing is far faster and easier than predictive multidimensional processing. Experiment 3 showed that it is quite easy to learn specific points in the parameter space via discrete hand shapes that are associated with particular sound and visual textures. Therefore we can make a recommendation that a musician's repertoire should be built up from these chunked *locations* in the space, not from knowledge of individual *directions* in the space.

This is not such a radical proposal as it may seem. For example, beginner guitarists learn chords as visual chunks. Predicting how to move their fingers individually so as to change each of up to 6 notes to fit with the next chord would be far too difficult. Instead, they simply recall the shape of the chord as an arrangement of dots on a grid. This shape provides the visuospatial scaffolding by means of which a single hand shape chunk can be established in memory. Whilst practising using the visual aid, procedural memory is established, and eventually the visual notation is no longer necessary. Experiment 3 indicates that hand tracking and virtual reality technology could provide a powerful tool to establish a closer connection between visuospatial notation and body movements.

Consider the brain's time-scale hierarchy: the speed of feedback paths

For musical interactions, the ability to work with multiple time-scales is essential. The presence of a number of time-scale systems in the brain should be considered when attempting to provide an interface that works efficiently: there is no use providing an interface that requires the top level cognitive systems to deal with round-trip events on time-scales shorter than around 500ms. Likewise, one cannot expect intuitive lower-level movement control processes to know about or predict the effect of operations that extend over longer time-scales, or that feature explicit thought processes such as branching logic or means-ends analysis. If operations are consistent and state-free, they can be performed automatically, but if the user has to consciously make a decision between multiple scenarios held in working memory, then this takes both time and attentional resources.

One concrete recommendation is to avoid 'modes' in which the same physical controls have different effects depending on some mode which is set elsewhere. This will be the cause of constant error as the implicit system will most likely fail to consider complex modal dependencies. For example, one may automatically reach and change a control and only afterwards will the explicit system realise that the control was not mapped to the correct parameter in the current mode.

Informative feedback is a staple of HCI guidelines, but it is also important to consider how informative feedback might *negatively* affect instrumental skill acquisition. Given the brain being a 'cognitive miser', it will tend to settle for least-effort interaction styles. Experiment 2 showed that if 'slow and easy' feedback is available, users may rely on this excessively, and not be training their internal models to operate the device without feedback. This is one aspect of the trade-off between the 'barrier to entry' and the 'virtuosity ceiling'. More consideration should be applied to the training stage for electronic music input devices, not just telling users what the device does, but also helping people to train themselves to use it optimally. Again, cognitive science can inform this research.

A few simple tests can establish if skilled interaction style has been achieved. These tests involve removing various means of feedback. Can the device be used to achieve a specific result:

1. With the eyes shut?

2. With the sound off?

3. Without the device present?

The latter case will test how strong the user's mental model of the parameter space is. If the artist can imagine their way to the solution without any feedback at all, this will make interaction more efficient when the device really is present.

Avoid forcing the user to constantly customise and alter mappings

The space of mappings between control and parameter spaces is magnitudes larger than the two spaces themselves. Therefore the construction of a mapping will be far more demanding than using an existing mapping. Therefore the typical interfaces that are used for one-to-one mapping (requiring slow analytical thought) can be critiqued using the above model. It seems that many designers over-estimate the predictive abilities of the user — in this case their ability to predict the instrumental capabilities of the mappings they are constructing. The user is almost always forced to specify exactly which control maps to what, despite the unpredictability of the results. Is it for the artist to worry about the intractable problem of what the 'best' mapping is? Yes, some mappings will be better than others, but not by such a large amount that practice times will be reduced hugely.

Given the rather unpredictable nature of the resulting instrument, there is a strong argument that construction of control mappings should fall into the exploratory interaction mode, not the algorithmic one. Here the approach of Van Nort and Wanderley [2007] seems to be promising: providing meta-controls for the mappings themselves, which can be explored and saved in the same manner as presets.

Machine learning techniques may also significantly speed up the mapping process [Fiebrink et al., 2009]. However, given the importance of the skilled interaction quadrant, building the mapping between intention and gesture in the brain of the artist is at least as important as any mapping between gesture and sound. No matter how compelling the mapping, the information required to reliably reproduce a repertoire of complex gestures must be built up in the artist's memory. Again, this would imply that training is an important unsolved problem in NIME research. It may be that greater benefits are to be found in changing the way the space is learned and visualised (see Experiment 3), not in the specific dimensional arrangement of the space itself.

There is also the obvious point that if the user is able to constantly change the mapping, this will interfere with their acquired skills.

9.2.1 Interfaces for the Reflective Mode

How to design for the reflective mode? As argued in Chapter 5, cognitively pipelining the interface such that it can stay out of working memory is an essential prerequisite for reflective thought. Beyond this, how could an interface actually provide the means to *augment* reflective thought? The reflective mode needs some way to handle increasing layers of abstraction. Today's DAWs have a rather inflexible ceiling in terms of abstraction hierarchy [Duignan, 2008]. It is difficult to notice a pattern in one's own compositional practice and turn that into a script or macro that implements that pattern in a more efficient way.

One way to achieve this is to provide scripting languages that operate on top of existing sequencers. Good candidates for reflective interfaces are musical programming languages such as Max/MSP or Supercollider. Here one can define one's own abstractions and transformations, but at the cost of considerable cognitive demand and technical skill.

There is potential for software to provide meta-views of one's creative history. Sonic Zoom showed that users found a visible overview of their path through parameter space useful. A way to meaningfully extract patterns from an interaction history, or a way to find patterns in a large corpus of DAW projects could provide an artist with a higher level view of their own creative process and hence a way to make their creative process more flexible or efficient.

Machine learning techniques may provide a means to reflective interaction by providing a way to automatically encode existing musical structures in a simpler form. For instance, consider a piece of music whose structure consists of many complex automation curves for timbre parameters. If an approximate lower-dimensional subspace of these parameter adjustments could be found (for instance using Principle Component Analysis), then a multidimensional interface could be automatically generated that would enable more direct, more flexible, and less demanding improvisational control of that track's structure.

9.2.2 Implementing the EARS cycle

A full 4-quadrant supporting interface has not been developed for this work. Rather, the approach has been to directly test the interplay between the exploratory and the algorithmic mode (Experiment 1), and the skilled and analytic mode (Experiments 2 and 3). This was in order to more carefully test some of the predictions of the model with regard to mappings and controller types. But eventually, the recommendation would be to provide an integrated system that enables fast switching between all four modes. This is easier said than done, considering the radical differences in mapping geometries, and the inter-quadrant interference and trade-off effects mentioned in Section 5.5.4.

The incubation-illumination cycle is already somewhat mirrored in creative technological interaction, by virtue of random explorations producing sudden 'Aha!' moments of discovery. However, to date this cycle has not been specifically designed for. Sonic Zoom provided one example of how an exploratory and an analytic interface could be coupled and presented together such that these modes could be alternated quickly, and user feedback suggested that this ability to transition was seen as valuable. However, given the four EARS modes, there are many other transitions that may need to occur. For example, switching between instrumental play (skilled) to computer-based editing (analytic) is currently awkward.

How could all four modes be provided without merely increasing the complexity of the system? How, specifically, are these twelve possible transitions to be carried out (Fig. 9.2)? Enumeration of all of these transitions is left for future work. But a hypothetical 4-quadrant workflow might proceed as follows: The artist should deliberately set aside time to reflect, plan and imagine an approach before engaging with any interface (reflective); then they may attempt to design a number of dimensions that define an interesting space to explore (reflective → algorithmic). Next the artist proceeds to explore the new space (algorithmic → exploratory), discovering many useless regions, but some useful ones. Next, some way to reflect upon, abstract and condense these interesting regions into a smaller space is required (exploratory → reflective). Next comes a process of practice within this new smaller space to ensure expertise (reflective → skilled). Then, using these skills a performance can be recorded, and then perhaps edited in fine detail (skilled → algorithmic). At each

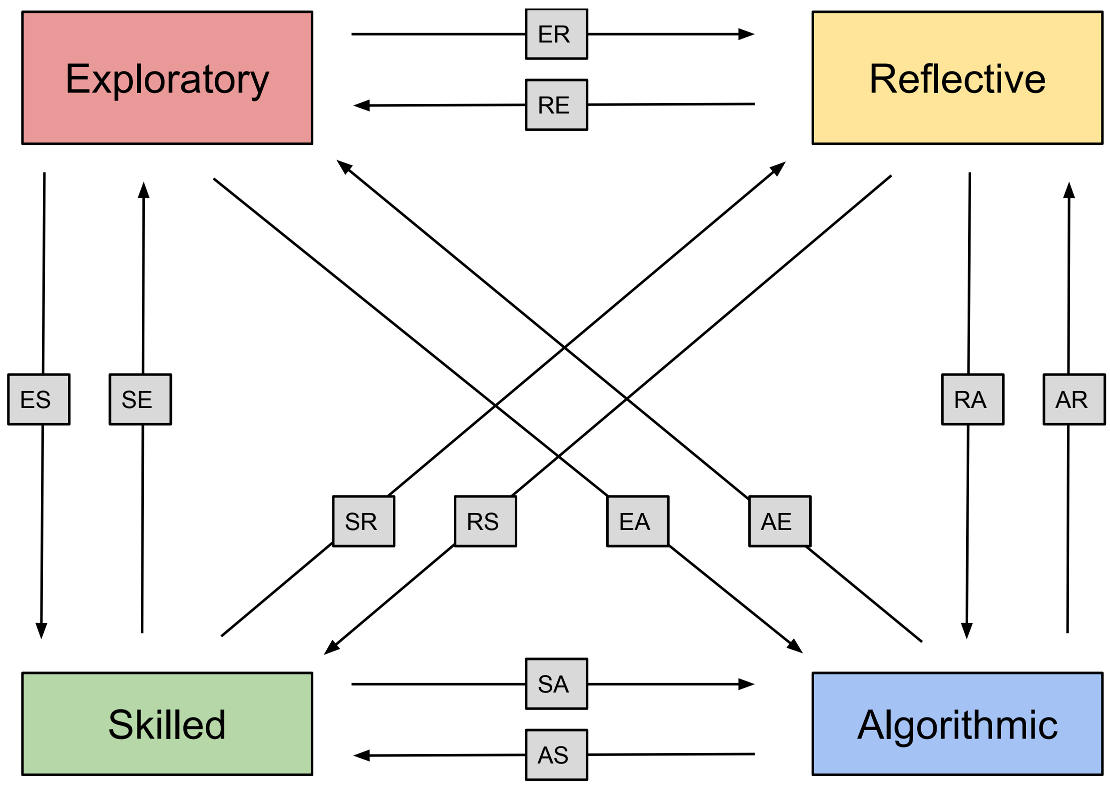


Figure 9.2: There are twelve possible transitions between EARS modes, currently many of these are difficult to achieve. For example, the transitions between detailed editing and skilled performance (EA and AE) are rather difficult with current DAW/controller set-ups.

stage an interface is needed that will suit that cognitive mode, but also at each stage the existing musical data should be made available to the new interface in as painless a way as possible.

9.3 Future Directions

This section discusses some of the many directions this research could lead in the future. These range from pursuing deeper questions about creative cognition, to designing more complete and practically usable musical interfaces. There is also the question of where this investigation leaves the debate over Fitts' law based evaluations and the quest for higher throughput human-computer interfaces.

9.3.1 The Future of Fitts: Beyond 100b/s Interaction

Some participants in experiment 3 showed peak throughput values of around 22b/s. One would expect most users to be able to attain this with practice. Compare this to the commonly reported values of around 5b/s with a mouse, and the apparently "remarkable" value of 6.2b/s obtained with touch screens [MacKenzie, 2015]. A four-fold speed increase over traditional computer interfaces is therefore something quite dramatic[3]. It is furthermore not unreasonable to suppose that two fully tracked hands, with 20 DOF each, can be fairly well lined up with virtual guides, these guides now being fully rendered hand models. Given haptic feedback to improve accuracy and stability, throughput values may well be able to exceed 100b/s, which would be a truly remarkable rate of search space reduction[4]. These estimates should make us take hand-tracking very seriously as a future interaction method.

Given that the standard Fitts' law methodology would not reveal these high throughput values, where does this leave the debate over which formula and methodology to use? Does this mean all the 1D and 2D evaluations are invalid? I would argue that the 2D evaluations do in fact underestimate TP values. In [MacKenzie, 2015] the 2D tasks give a lower TP value (6.39bps) than the 1D task (7.52). Considering that the *usable* information received by the UI is almost certainly higher in the 2D case due to the larger potential number of selection options, this seems suspect. If these results were re-evaluated using ISSR, the 2D task would likely give a higher throughput, because the area of the target is much smaller relative to the area of po-

[3]To put this in a one-dimensional context, this rate of search space reduction would be the equivalent of locating the correct key once every second on a piano keyboard 70km long. This raises the possibility of more effectively controlling a 1D quantity, such as pitch, with a multi-dimensional controller.

[4]The equivalent of selecting a note every second from a keyboard the width of the visible universe

tential targets. Nevertheless, it may be unwise to change the 1D and 2D standard, as there is a great deal of previous work that would need to be fairly compared. However, no amount of experimentation and line fitting will tell us how throughput should scale with dimensionality, because the multiplicative constant is essentially arbitrary for these plots. Only theoretical considerations can reveal this, and these firmly point to information gain being proportional to dimensionality. Hence the need for the dimensional multiplier in the ISSR equation (Eq. 5.12). Therefore I would recommend that researchers working with high-DOF controllers use the ISSR metric (or the entropy of the distribution of search paths) for evaluation purposes.

Further work could apply these high throughput opportunities to other, more everyday HCI scenarios. Even just the corners of the 20D hypercube of hand poses constitute over a million different hand shapes, and these could provide a practically unlimited repertoire of gestural shortcuts for instant access to common computing operations. If augmented reality becomes widespread, learning and recalling these hand poses could be significantly aided by virtual body emulation.

Another interesting direction for future research, that this thesis has only touched on, is closing the gap between state of-the-art models of embodied cognition and well-established movement laws currently used in the HCI field. Can the latter be derived from the former? Can empirically discovered movement laws such as Fitts' and Schmidt's law [Fitts, 1954; Schmidt et al., 1979] be explained by theories of hierarchical, predictive motor control [Friston, 2010; Wolpert and Flanagan, 2001]? Can we explain why the two different movement laws emerge in the two different situations of accuracy-targeted and time-targeted aimed movements? Our hypothesis here is that the difference between the two laws is due to use of, or decoupling from, corrective feedback, but this would need to be researched more thoroughly.

9.3.2 Throughput Optimality: A Recipe for Inspiration and Flow?

Whilst flow is a widely recognised phenomenon, and is well documented by Csíkszentmíhályi and others, Flow theory could be criticised for not being a theory at all, more of an observation [Dietrich, 2004b]. With the exception of Dietrich, few have attempted to propose underlying cognitive mechanisms that generate the Flow state. The eight dimensions of Flow as yet have no more fundamental characterisation in terms of information processing that might explain and connect them. In this section we speculate whether the throughput peak found in Experiment 3 might be one instance of a more fundamental phenomenon of open-loop versus closed-loop information processing, and whether this phenomenon, when extrapolated to higher-level thought processes, may underlie the experience of Flow and artistic inspiration.

The different shaped entropy reduction curve of Experiment 3 revealed a non-linear relationship between the information output of the nervous system and time. One explanation of this could be that movements intended to realise a target in a pre-specified time (e.g. rhythmic interaction) occur open-loop, because temporal targets do not allow the luxury of time spent processing sensory feedback. Open-loop movements mean that the information needed to specify the movement is pre-programmed, and released as a single burst, rather than being accumulated over the duration of the movement by means of error correction.

One of the tenets of Embodied Cognition is that higher level cognitive processes utilise the same mechanisms as movement control processes (see Section 2.2.6). If this is true, then we could ask: is there a correlate of this open-loop throughput peak for higher-level creative thought as well as low-level motor control? Assuming this mode of thought exists, it would display the following two characteristics:

1. It does not involve error correction on the basis of feedback.

2. It would result in a higher rate of information production than 'normal' closed loop cognition.

The lack of error-correcting feedback could then have a number of effects:

1. The predictive probabilistic model being used to generate action becomes fuzzier and more widely distributed, producing more divergent ideas. These ideas may be more surprising, and have a higher novel information content when compared to closed-loop responses to normal everyday situations.

2. The decoupling of prediction and corrective feedback may generate strange subjective effects, such as the merging of action and awareness. The distinction between one's own volitional actions and those occurring due to outside influences depends on error-correcting feedback. In a 'decoupled' state this distinction may dissolve. In turn the distinction between deliberately searched for sounds and those sounds found by accident may blur in the artist's mind.

These effects begin to sound very much like the characteristics of Flow, or artistic inspiration. This may also explain what it is about music that is so conducive to these types of experience. By enforcing time constraints, we force ourselves to go into an open-loop mode, i.e. we decouple ourselves from correcting ourselves on the basis of slower sensory feedback, or critical thought. The faster and less consciously analysed the feedback becomes, the more it allows our 'inner critic' to switch off and allow things to unfold on the basis of learned skills, or spontaneous ideation[5]. In addition, allowing ourselves "artistic freedom" might enable risk-free decoupling from error correction. So music seems to positively encourage open-loop behaviour, and may thus produce states where our minds seem to be functioning more efficiently

[5]It should be noted that this state requires skills to be built up beforehand, via a more critical process. To generate genuinely H-creative [Boden, 1992] artefacts there would need to be considerable amount of prior time and effort spent in other non-flow activities: reflecting and questioning and updating internal models on the basis of errors and feedback.

than normal: it may feel as though information is being processed/produced at a greater rate, but with less effort. Thus "Flow" may actually be linked to the "flow" of information. As for the question of why Flow is so intrinsically rewarding, we can appeal to a Schmidhuber-type argument [Schmidhuber, 2010] and surmise that rapidly reducing entropy triggers reward mechanisms.

So an interesting research programme could be to try to link Flow more firmly to information theory and predictive brain models. Experiments could be done to measure the correlation between throughput, rate of idea production, and the subjective feeling of Flow (as measured by the experience sampling questionnaire [Nash and Blackwell, 2011]). This would entail an artist using an interface for several hours, rating how inspired they were at each point, and looking for corresponding regions of rapid entropy decrease as measured via the ISSR metric. It could turn out that throughput is more than just a useful methodology for evaluating interfaces; it may be fundamental to the peak creative experience itself. Carrying out these types of experiments with brain imaging techniques would also be fascinating. If brain images were obtained, the above theory of Flow could be tested against Dietrich's hypo-frontality theory.

One objection to this throughput-based approach, and a potential weakness in this thesis, is that the link between throughput and the artistic *quality* of the creative artefact has not yet been established. Indeed, it may be possible that rapid progress towards the completion of an artefact may be completely irrelevant to the value of that artefact, as it would be judged by the artist themselves or other domain experts. Much further work could be done here, for instance by getting people to rate the results of sound design sessions with different interfaces for novelty and value. Would the results of a session using the Hilbert curve mapping result in higher novelty or value ratings than that with the Leap Motion or the sliders? One could also imagine imposing different time constraints on the creative process and judging whether fast sessions produced more creative output than slow ones.

9.4 Final Thoughts: Towards a Multidimensional Future?

Almost all user interfaces for creative software provide parameters such that features are edited in a separate, serial fashion, or at most on a two-dimensional plane using a mouse. These interfaces are used to create music, animation, industrial design, architecture and computer games. The artefacts created with these tools define a huge proportion of 21st century culture. If this interaction paradigm really does change the way in which people are creative, this seemingly innocent and logical arrangement may already have had significant consequences for the form and quality of artistic innovations. Will new multidimensional interaction devices encourage different modes of being creative? Will we see a return to more embodied and intuitive forms of expression in the coming decades?

Transferring to the use of high-dimensional controllers is a huge step for the computer industry to take. It may be that contemporary input devices are too firmly locked in, or that users are unwilling to invest time and effort to develop highly skilled actions. If a transformation is to occur, it may well be driven by virtual reality, in particular, computer games. VR Game environments necessitate richer, faster and more natural interactions with virtual worlds. However, I believe the potential of new input devices is greater than just mimicking the world that we already inhabit. There is also potential to take control of more abstract data in a far more effective and rewarding way. There is a great deal of brain processing power lying idle in our motor cortex when we are sitting at our desks. What will be the key to utilising this motor skill to enable heightened levels of control of digital artefacts? The key will be the successful mapping of complex gestures to manipulate more abstract parameter spaces. Thus the "mapping problem" of NIME research, where gesture must be mapped to the abstract realm of sound synthesis parameters, has a lot to contribute to the future of computer use, and the way humans will augment their own cognition and creativity in years to come. Music is one of the most intimate and spectacular real-time interactions that humans can partake in. It is exciting to think that this mapping problem, which began as some musicians' dissatisfaction with computer technology and has now developed into a small but vibrant research field, may eventually inform a completely new, deeper and more intimate relationship between humanity and information.

Technology affects cognition, shapes it, and provides a space in which it operates. This thesis has only shown this in a few limited task domains, but if cognitive principles generalise — and research into embodied cognition would persuade us that they do — then interface technology may shape our very attitude towards cultural artefacts, and our relationship with the world of information we inhabit. Therefore as musicians, designers, technologists, and members of human society, we need to ask ourselves: do we want a one-dimensional culture?

Appendices

APPENDIX A

Studio Technology and the Creative Process: A Survey of Electronic Musicians

A.1 Introduction

Whereas a writer's words will probably not be affected to a great extent by the specific word-processing software they are using, or even whether they use a typewriter or a pen, the sonic material of electronic music is deeply wedded to the mechanisms of its production. Electronic musicians are therefore uniquely positioned to assess how technology impacts their creative process. This survey was designed to assess musician's own experiences of using technology, specifically targeting the themes of divergence/convergence and implicit/explicit thinking presented in Chapter 5. This survey, whilst not conclusive enough to form part of the main body of this thesis, shows that the EARS model can be a useful and simplifying framework: in that it can generate revealing questions, and provide a structured analysis of the artists' responses. The questions are presented in sections relating to the following themes:

- Motivating the research topic (Section A.2).
- Constraints and complexity (Section 3.6).
- Exploratory interaction and serendipitous discovery (Section A.4).
- Mental load and interference from demanding interfaces (Section A.5).
- Skilled interaction and intuitive decision making (Section A.6).
- Reflective and evaluative processes (Section A.7).

A.1.1 Questionnaire Format

Most questions were presented as a sliding 11 point scale (0-10) between two extremes of opinion, with 5 representing the neutral response. One problem with this format is that the middle response of 5 can possibly be interpreted as "definitely equal weighting between the two extremes", "I don't know", "I don't care" or even "I don't understand the question". Participants were told not to select any answer if they were very uncertain of their opinion, however questions with a large number of neutral responses may still indicate a problem.

Whilst Likert responses are technically considered ordinal, we assume here that 11 points are enough to assume an interval scale, and that means and standard deviations are meaningful[1]. Again, these results are more motivation than evidence, therefore no specific hypotheses are tested and statistical significance is not claimed or reported.

The survey was presented online, via Google Forms[2]. The questions were not presented in the order below, rather they were grouped around the topics "work-

[1]This is a controversial practice however [Jamieson, 2004]
[2]`https://www.google.co.uk/forms/about/`

flow", "sound design" and "live interaction". This was for two reasons, firstly to not bias responses by imposing the EARS model's conceptual categorisations on the respondents. Secondly, to widely space questions that were re-phrasings of one another, or complementary in some way. The number relating to the actual positioning of the question will be given at the start of the heading, and in the figure caption.

Means and standard deviations will be reported as ($M = 3.4$, $SD = 2.1$). Modes will be reported as (mode = 8, $N(8) = 12$) where 12 respondents selected response 8. Numbers of responses with a particular value will be reported as, for example, ($N(5) = 10$) in the case that 10 people selected a response of 5. The labelling of histogram axes always refers to number of respondents who selected that option, except when marked as a percentage.

At the end of each section, participants were invited to comment further on the questions. Some selections from these text responses are quoted alongside the most relevant question.

A.1.2 Respondent Information

The respondents were generally quite mature ($M = 33$ years, $SD = 5.5$), with a large number of years experience with both electronic production ($M = 15$ years, $SD = 6.6$), and traditional instruments ($M = 9.8$ years, $SD = 11$), although tended to be self taught in this regard ($M = 2.3$ years of formal musical training, $SD = 3.7$). The majority described themselves as "semi-professional" musicians, with a large number having officially released records ($M = 8.2$ releases, $SD = 8.2$). Therefore, in general we can expect a high level of proficiency with these tools, and a high degree of awareness and critical thinking when it comes to reflecting on creative practices (e.g. question 19 and 40). The different types of music technology that the

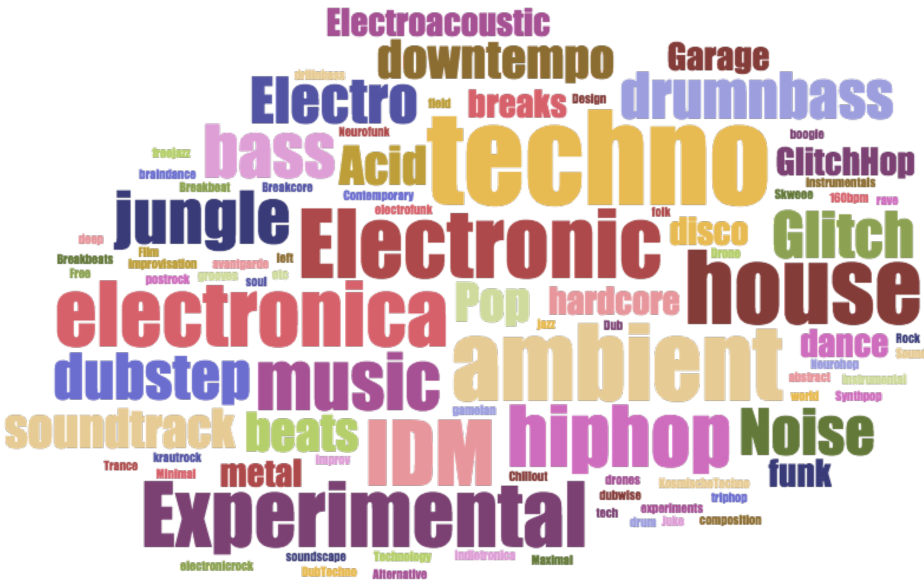
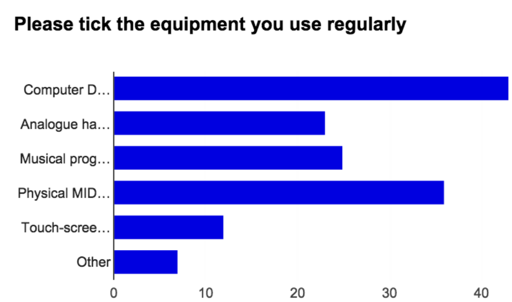


Figure A.1: Relative frequency of musical genres reported by respondents.



Computer DAW & Plug-ins	43	95.6%
Analogue hardware	23	51.1%
Musical programming environment e.g. Max, Reaktor	25	55.6%
Physical MIDI controllers	36	80%
Touch-screen/mobile	12	26.7%
Other	7	15.6%

Figure A.2: Proportion of respondents using various types of music technology. Respondents could select more than one option. Equipment reported as 'other' included custom speakers, contact microphones, circuit bent electronics, C++ DSP code, and paint pots.

respondents reported working with are shown in Fig. A.2.

A wide range of working styles and genres were reported A.1. However, due to the author's strong ties to this community, and the use of social networking sites to recruit the sample of respondents, there is a strong bias toward underground, experimental dance music. Other stylistic groups may reveal different results, therefore these results may not be generalisable to all electronic music practitioners.

A.2 Motivating the Thesis

This section motivates the topic of research. The questions aim to support the following assumptions:

1. Timbre is an essential component in electronic music, for both low-level sound objects and long term dynamic structure.

2. Speed, liveness and throughput are vital for effective musical interaction, both for performance and for studio work.

3. The design of the interface is of the utmost importance for electronic music producers.

27. How important to you is shaping the timbre of the sounds you use?

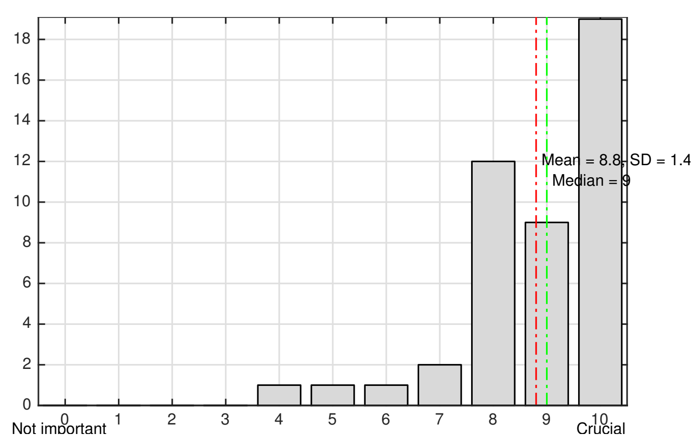


Figure A.3: Q27: Importance of timbre.

To counter the potential objection that this thesis overly focuses on timbre design, as opposed to melody, harmony or rhythm, this extremely strong response ($M = 8.8$, $SD = 1.4$) reveals that timbre design is an absolutely essential aspect of the creative process for electronic musicians. Hence, manipulation of large numbers of timbre parameters is one of the most important things an interface must enable them to accomplish.

36. How important is changing timbre for the structure of your tracks (e.g. do you "tweak" or automate parameters to generate tension/development over larger time scales)?

Very positive responses ($M = 7.7$, $SD = 2.4$) serve to emphasise that timbre design is not just surface ornamentation or production "fairy dust", but an integral part of the structure and dynamics of electronic music. Thus, manipulation of timbre is vital not just for instrumentation, but also improvisation and performance.

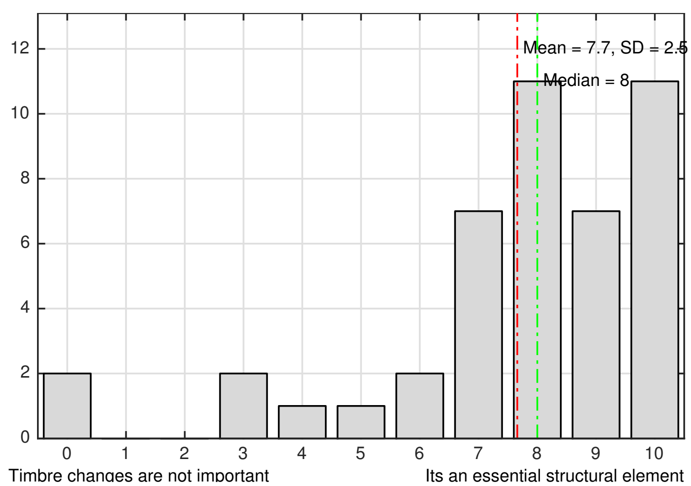


Figure A.4: Q36: Structural importance of timbre.

28. If you imagine a sound, can you then realise it faithfully using your technology?

Here, the hypothesis was that musicians may find it rather difficult to construct imagined timbres. However the majority of the responses were 7 and 8. So, subjectively, respondents feel able to accurately realise what is in their heads. Evidence presented in Experiment 2 might contradict this assertion, however.

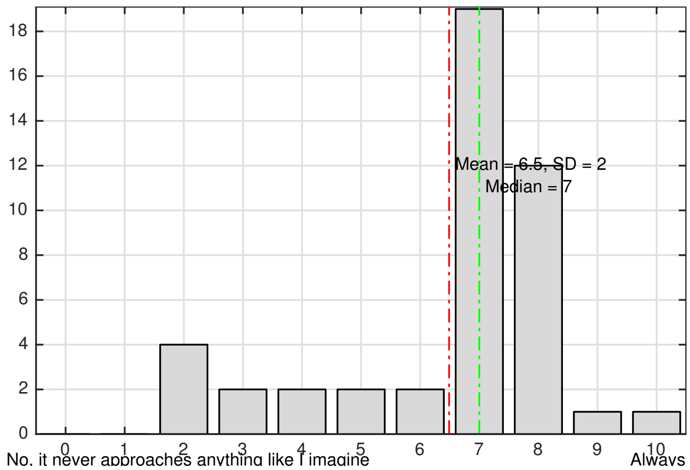


Figure A.5: Q28: Ease of realisation.

30. What proportion of the sounds you use are presets or samples, and how much are custom-made by adjusting parameters?

Again, the response motivates the investigation into timbre design. It may be imagined that the large libraries of samples and pre-programmed presets that come bundled with many commercial software and hardware systems are sufficient for musicians to create music with. This response reveals a strong tendency toward customisation[3], and that sound design is an integral part of electronic music making.

31. Do you need to design (or select) the sounds in detail before working on the structure of the track, or sketch out rough high-level ideas and improve the details later?

The pre-digital recording studio process leaned toward late stage timbre design, with the song written first, the instrumentation and signal processing applied later.

[3] If we invert the scale we could estimate the proportion of customised elements as being 77%

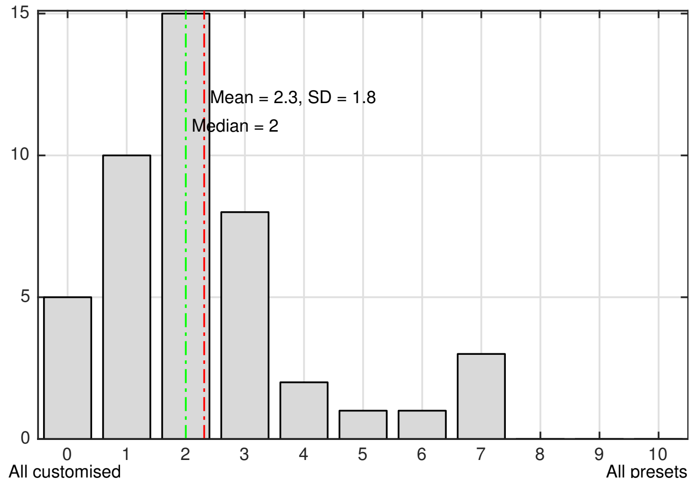


Figure A.6: Q30: Timbre customisation extent.

At what stage do modern producers create the details of timbre? Is it during the early "blank slate" stage, or later on, when the track has already taken shape? The expectation here was that there would be a tendency to start with sound design: due to the nature of production software/hardware, it seems one needs to instantiate one's own instrumentation before doing anything else. Many had a strong tendency toward early stage sound design (mode = 2, $N(2) = 10$), but other opinion was fairly mixed. A large number of neutral responses probably indicates what later questions reveal: that the value of timbre is highly context-sensitive, therefore sound design can be necessary at any point.

34. Do sounds inspire tracks, or vice versa?

This question was related to question 32 and 35, but with altered emphasis to find out how important sounds are for generating further ideas, producing feelings of inspiration, and forming the basis for entire pieces of music. By far the majority answered neutral (mode = 5, N = 15), indicating both, neither, or a confusing question. However those who did not answer neutrally were disposed toward sounds inspiring tracks ($M = 3.4$, $SD = 2.1$). This again highlights exploratory timbre

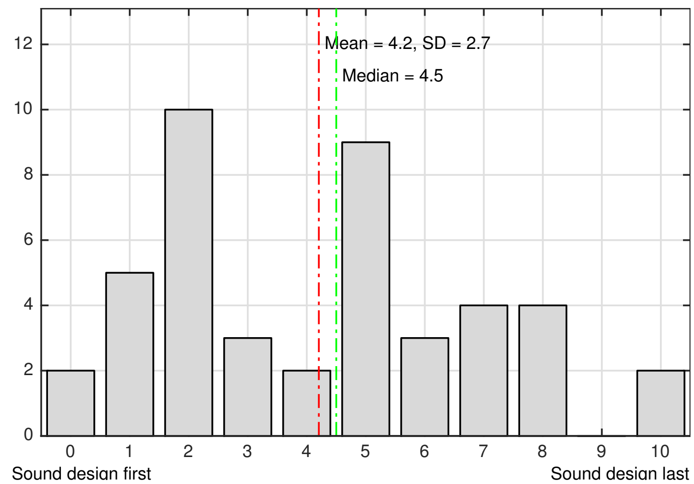


Figure A.7: Q31: Early/late stage sound design.

design being an important component of early stage creativity.

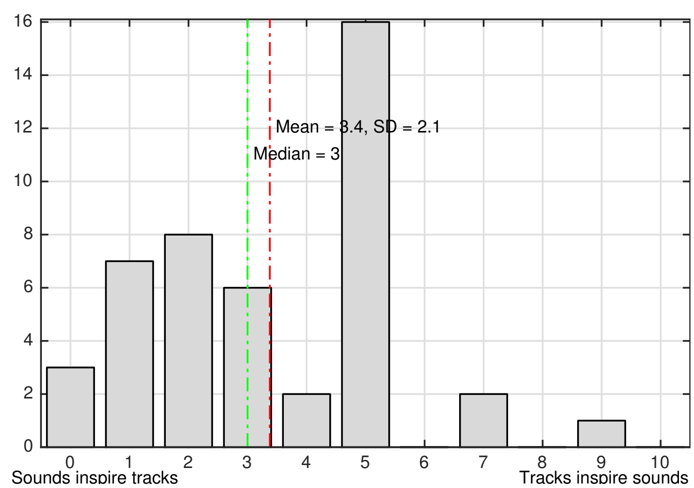


Figure A.8: Q34:

35. Do you separate sound design from composition?

Here the tendency was towards creating sounds "on the fly" (mode = 8, N = 12). This may be because the appropriateness of timbre depends so much on the surrounding musical elements, and each sound object has to be optimised to fit in with the current context.

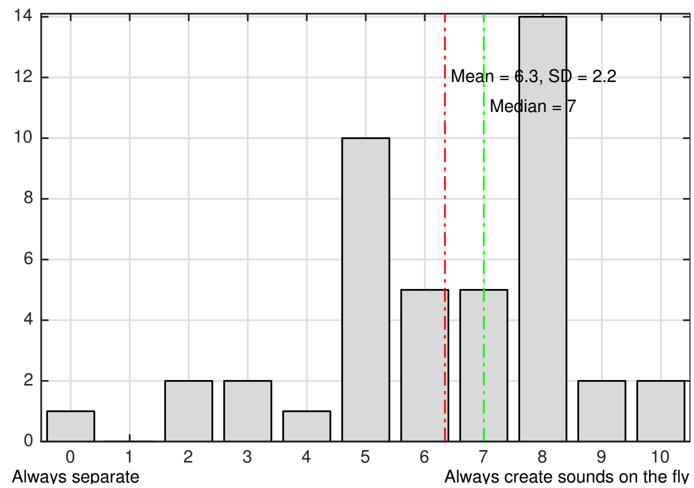


Figure A.9: Q35: Separating sound design from composition.

63. I really do perform fully "live" with my technology

The majority of respondents confessed that they do not feel as though they are genuinely performing live[4] ($M = 3.7$, $SD = 3.1$).

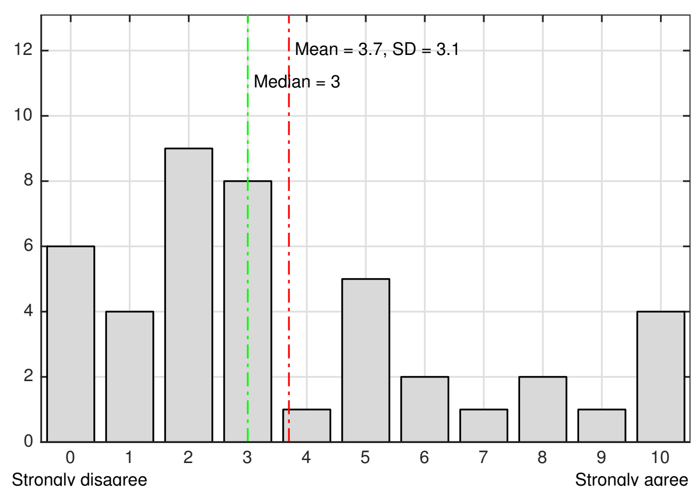


Figure A.10: Q63: I perform fully "live" with technology.

[4]A note indicated that if the respondent did not play in front of an audience they should not respond to these questions about performance.

60. If you perform live, do you want to perform more components of the music than you currently feel able to?

Another strong positive response ($M = 7.1$,$SD = 3$). This motivates investigations into how to enable greater feeling of control over more musical parameters.

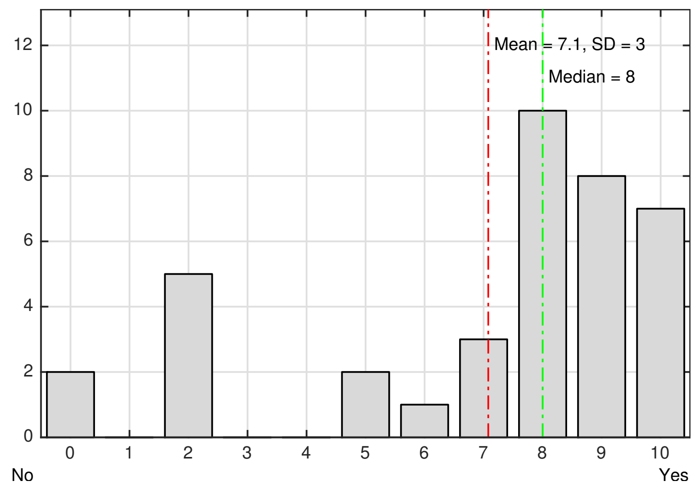


Figure A.11: Q60: Desire for more live control over musical parameters.

51. Does better technology make for better music?

"Better" may be a complex notion, nevertheless it was hypothesised that people would say that the quality of the music was not dependent on technology. This question produced many ambiguous responses, but ones that slightly tended toward the negative.

65. Do you feel that you can make better music with a better user interface?

This question produced a very strong positive response (Mode = 8,10, N(8,10) = 13). This contrasts markedly with the previous question (despite the fact that the interface *is* technology!). This discrepancy should persuade us that improving musical interaction is potentially of more benefit for practising electronic musicians than any other research regarding synthesis algorithms, DSP, audio engineering, or any software "inside the box".

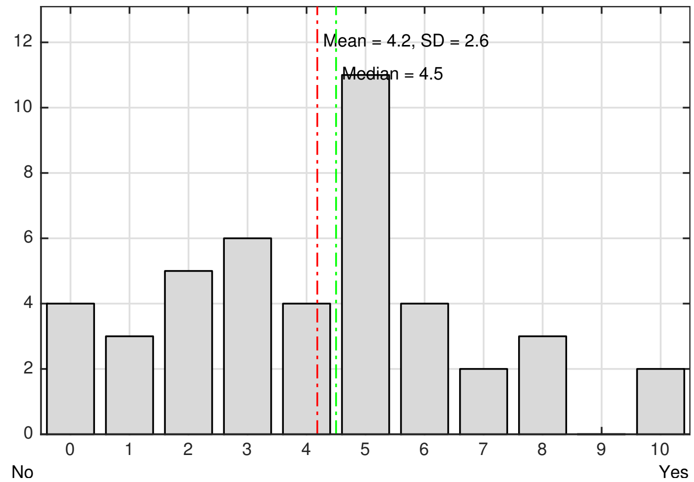


Figure A.12: Q51: Better technology leads to better music.

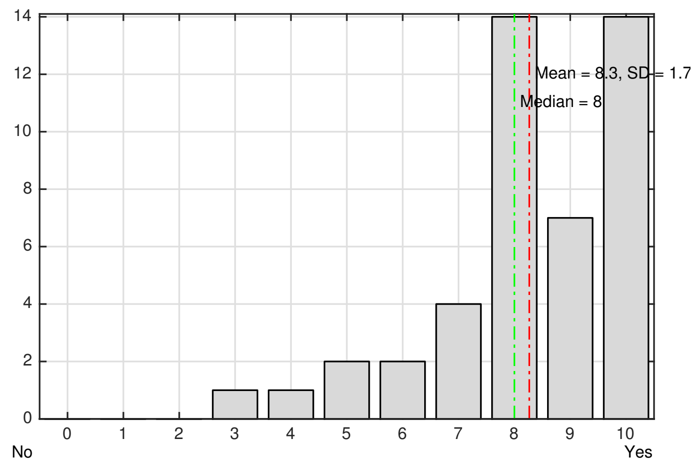


Figure A.13: Q65: Better interfaces lead to better music.

59. Is the speed at which you can turn your ideas into sound important?

This question is important, as it motivates the investigation of throughput in the theory chapter, and experiments 2 and 3. It is hypothesised that musicians find speed extremely important, even when not in time pressured live situations, for reasons of Flow, liveness, and maintaining ideas in working memory. There was general agreement with this statement ($M = 6.6$, $SD = 2.6$).

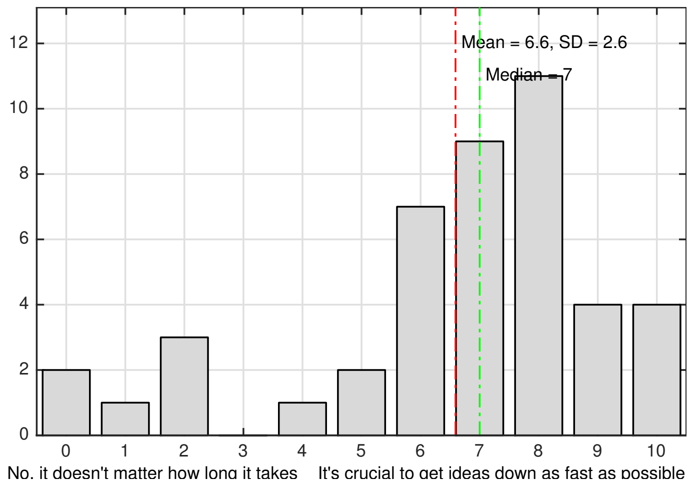


Figure A.14: Q59: Is the speed at which you can turn your ideas into sound important?

A.3 Constraints and Complexity

These questions related to the idea that constraints can sometimes encourage creativity. However externally imposed constraints, or complex interdependencies may increase cognitive load.

41. Constraints encourage creativity

Participants overwhelmingly agreed with this statement, however in the text comments, many pointed out the subtlety of the relationship between creativity and constraints. One respondent noted that it is easy to get lost in the exploratory mode, and that constraints encourage meta-level divergence by 'breaking' the rules:

> "Yes and No. Restraints can force you to push past boundaries—I was much more technically experimental when using limited equipment. Now that technology is limitless it's easy to get lost in it. You have lots of

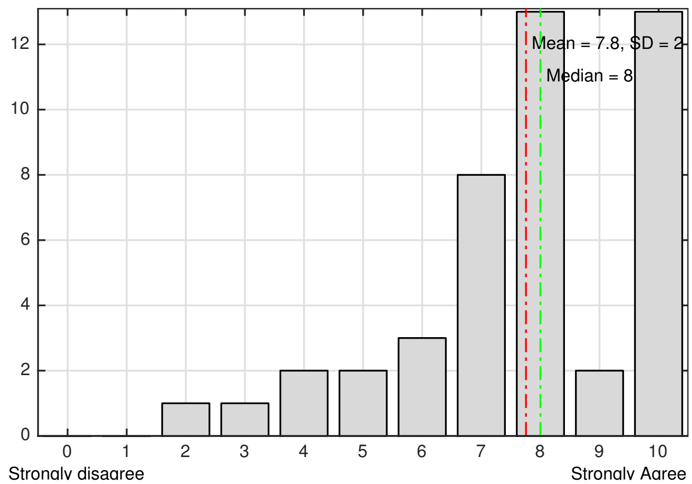


Figure A.15: Q41: Constraints.

options but you're not forced to 'break' your system and do something truly radical.

For this reason, I deliberately don't learn some things, and use old software and techniques, to try and put more emphasis on doing something weird from these restrictions."

Similarly:

"The more options available, the more avenues you can explore. However, once a track has started, limitations are required in order to provide boundaries to construct within, as you're forced to push against those restrictions in order to generate novelty."

"I think that having unlimited options is almost always bad, as it's the act of constrained problem-solving which forces unusual and interesting solutions to problems. Having access to every option allows each musical problem to be solved in the most obvious (hence musically uninteresting) way."

Other responses point out that restricted parameter spaces make decisions quicker and less effortful:

> "When you have too many options, most of the time this ends overwhelming you and making you spend too much time deciding on which option to use."

> "If I have an idea for a sound but no facility to create it, I will be frustrated. Having everything I *might* need ready to go can be liberating. But also, if I sit down and have to create something with no real plan, the breadth of choice can be daunting and I might end up f**king about and never deciding on anything for ages. With fewer options I find what I feel is best far quicker."

This respondent provided a neat summary of several of the considerations in the chapter 5:

> "Times when lots of options good:
>
> - Already have a clear idea in mind and know how to use all parameters to achieve it.
> - Experimenting wildly with unfamiliar parameters - unfamiliarity aids my creativity greatly as you are forced to listen more than if you know what each control does.
>
> Fewer options good because:
>
> - Less things to manipulate to get where you need / want to go, so less distraction.
> - Less room for perfectionism.

- Forces creativity in problem solving, say, when finishing a track or finding a new sound to fit.
- Encouraged to keep things simple. Simple ideas usually good from listener perspective."

47. Technological limitations restrict my ability to be creative.

Again, there was a strong tendency to disagree. Again it seems that constraints (this time technological ones) do not restrict creativity.

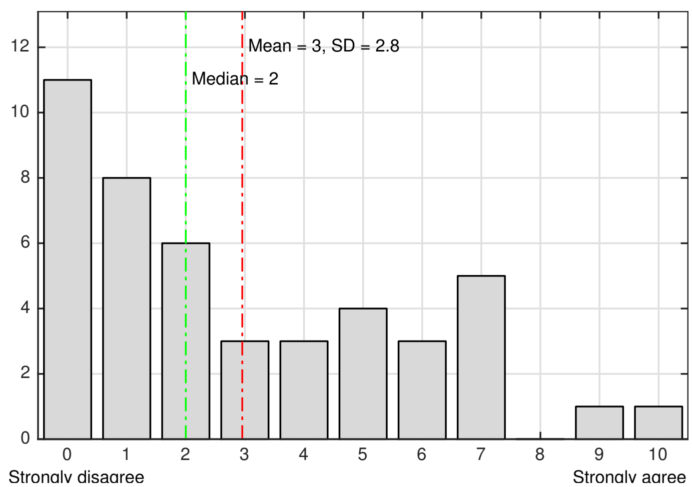


Figure A.16: Q47: Tech limits inhibit creativity.

32. How much does the value of a sound depend on its musical context?

A wide range of opinions, but a tendency toward context dependence.

Q18. Which is harder, generating ideas/sounds or fitting them together to make a track?

Another question related to constraints, this generated a majority of responses in favour of it being harder to fit ideas together. Generating and evaluating a single

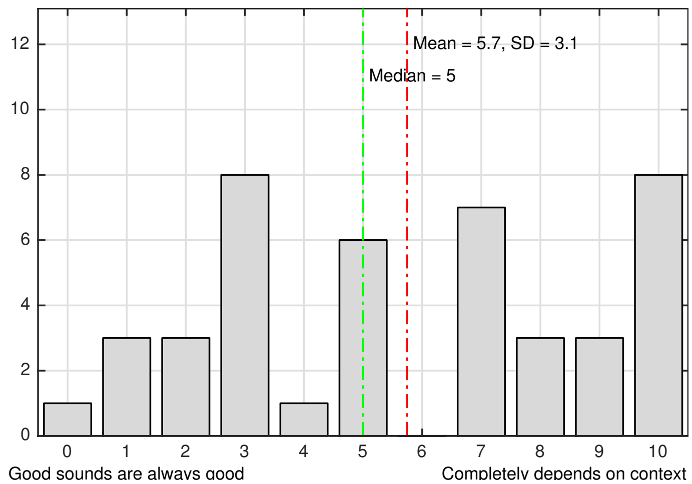


Figure A.17: Q32: Context sensitivity of timbre.

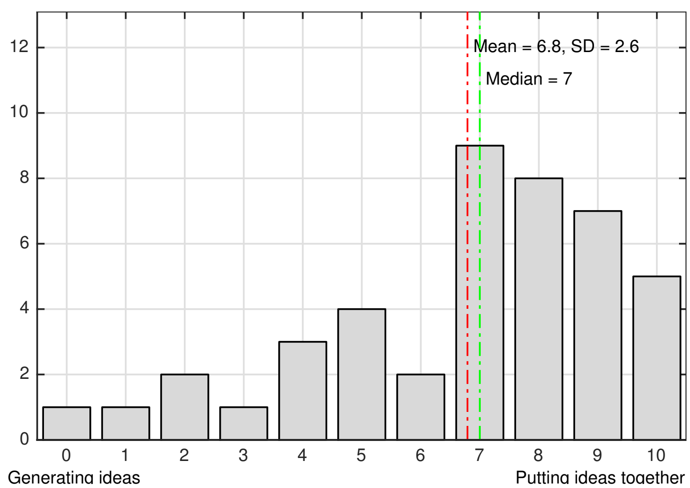


Figure A.18: Q18: Which is harder, generating concepts or fitting them together?

component on its own is not difficult, but thinking about interrelationships, and evaluating over longer time-scales is seen to be more of a challenge.

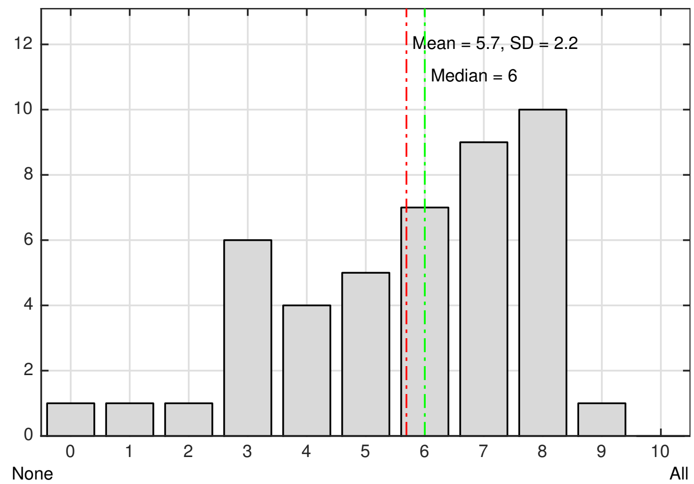


Figure A.19: Q33: What proportion of the ideas or sounds you use do you discover by accident? 0 = 0%, 10 = 100%.

A.4 Exploratory Interaction

33. What proportion of the ideas or sounds you use do you discover by accident?

This question relates to exploratory behaviour. The mean percentage of material estimated as generated via technological aberrations was 63%, SD = 22%.

48. Estimate what proportion of your material comes from interacting with technology, and what proportion is generated entirely in your own mind.

This question was intended to roughly assess the amount of information generated as part of the interactive loop, compared to that coming from top-down artistic goals. This could be considered as an alternative wording of question 34. It produced many neutral responses, but still the average tended toward a technological origin of ideas.

In the text comments, a creator of their own technology claimed:

"The scale doesn't really make sense here, my technology is handmade,

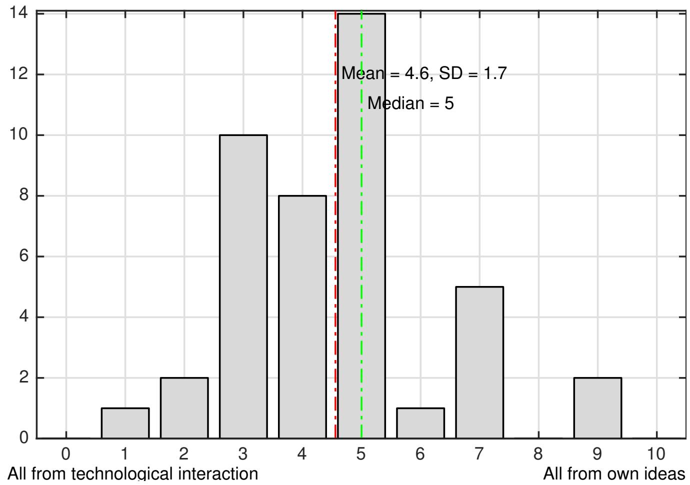


Figure A.20: Q48: Proportion of material generated top-down. 0 = 0%, 10 = 100%.

> so in a way it all comes from my mind"

Another respondent noted the fact that ideas in the mind will be shaped by past engagements with technologies and other musical experience:

> "Ideas generated in my mind but from past interactions with technology, imagined interactions with technology or from what I hear or observe with other performers/recordings."

Another points out that the proportion of the material that is genuinely novel may be quite small:

> "I'd probably go so far as to say that almost NO artists do stuff from their own mind 99% of the time. It's all got to originate somewhere."

58. Do you have an overall idea of what you will make from the start? Or does the track emerge during interaction with technology?

The hypothesis here was that electronic music tends to be an emergent process, and this seemed to be the general opinion.

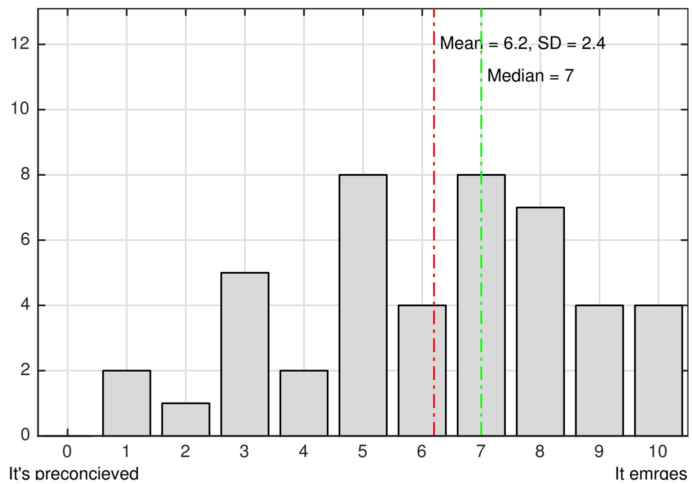


Figure A.21: Q58: Is the track preconceived or does it emerge during interaction?

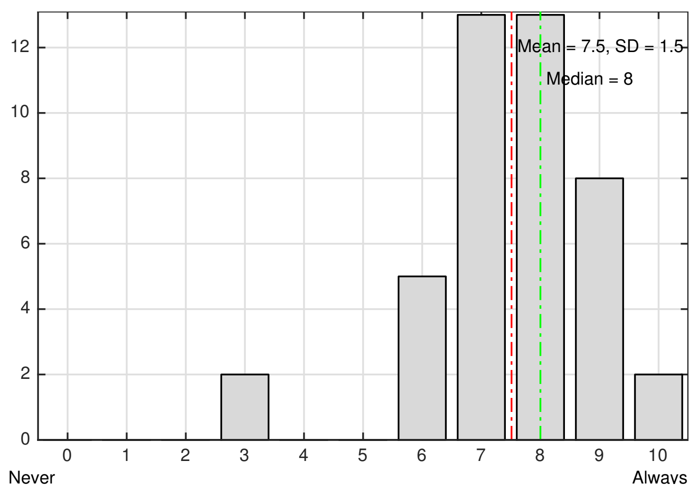


Figure A.22: Q61: Proportion of curiosity driven interaction.

61. How much of the time would you say that curiosity and exploration drive your interaction?

This was phrased as an estimate of time spent exploring. This turned out to average 75%. Due to the fact that this proportion is very similar to the proportion of ideas that were deemed serendipitous in question 33, it seems that the high accidental discovery rate is not something that is forced on the user by lack of competence or ineffective input devices, rather it is a deliberate strategy. This again indicates that exploratory interaction is vital for interface designers to consider.

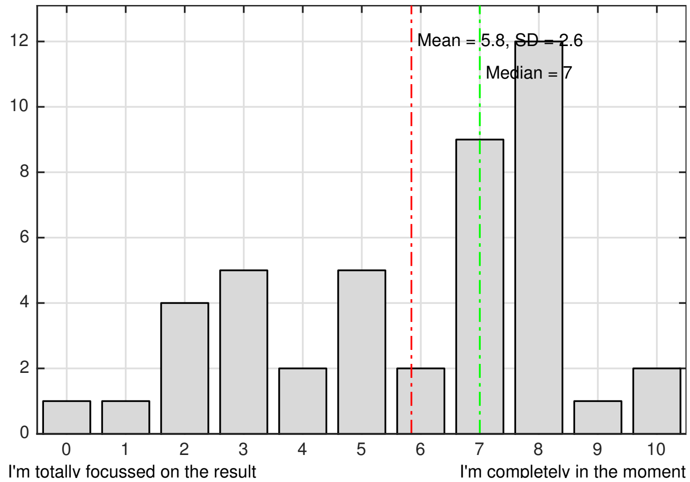


Figure A.23: Q13: How much of the time do you focus on reaching an end product, and how much are you absorbed in the process?

13. How much of the time do you focus on reaching an end product, and how much are you absorbed in the process?

Here the hypothesis was that respondents would answer very much in favour of "absorption", and this was the case. There are multiple explanations of this tendency however. The positive ones would be:

1. Musicians enter a state of flow, where they become one with the task.

2. The creative process is intrinsically interactive, and emergent (supported by question 58). Pre-planning one's goal is unnecessarily restrictive.

A more cynical assessment would be that attempting to reaching a pre-planned end product is far more demanding than exploratory interaction, or perhaps that the complexity of interaction occupies so much of the working goal-hierarchy that carrying out long term plans becomes impossible.

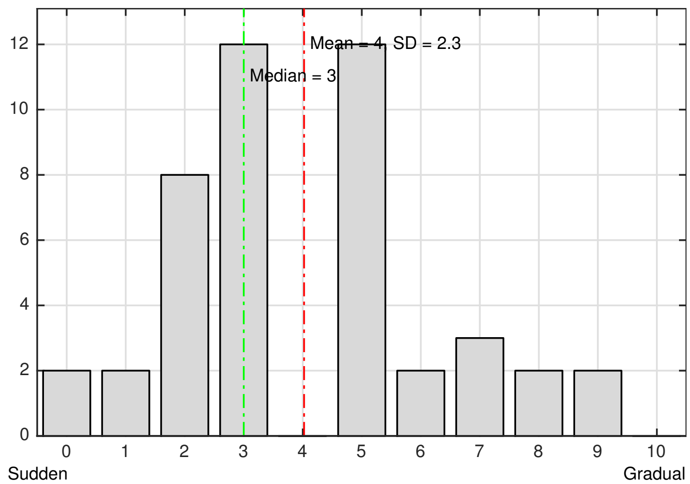


Figure A.24: Q37: Good ideas occur suddenly.

37. Do good ideas seem to occur suddenly? Or do they build up gradually?

An interesting issue is whether constant engagement with technology changes the sudden 'aha!' moment into a more gradual process, or if the serendipitous nature of the interaction produces sudden step changes. Most answers were neutral, but many erred on the side of suddenness.

This comment sums up the situation quite well:

> "Good ideas appear suddenly, but do continue to grow. So both answers are valid again in that respect. A track's theme however can sometimes take time to emerge. That counts as an idea too. So it depends whether you are talking about individual sounds / parts (which tend to appear quite suddenly) or overall compositional balance or throughline, which is sometime sudden, or sometimes very gradual. Usually for the overall structure / composition, there is a 'click' point, where things suddenly make sense, but there has been a gradual curve to get there.".

Another user points out that both fast and slow progress can yield good final products:

> "Some of the best tunes I've done have happened in moments and have been very simple in structure and composition. Similarly, there have been tunes I've worked on for ages that have finally come good. I consider these to be different ways of working."

.

A.5 Mental Load

These questions were intended to investigate which aspects of the creative process induce the most mental load, and in turn how that cognitive load impacts on various other aspects of the creative process.

12. Which stages of making a track would you say were hard mental work?

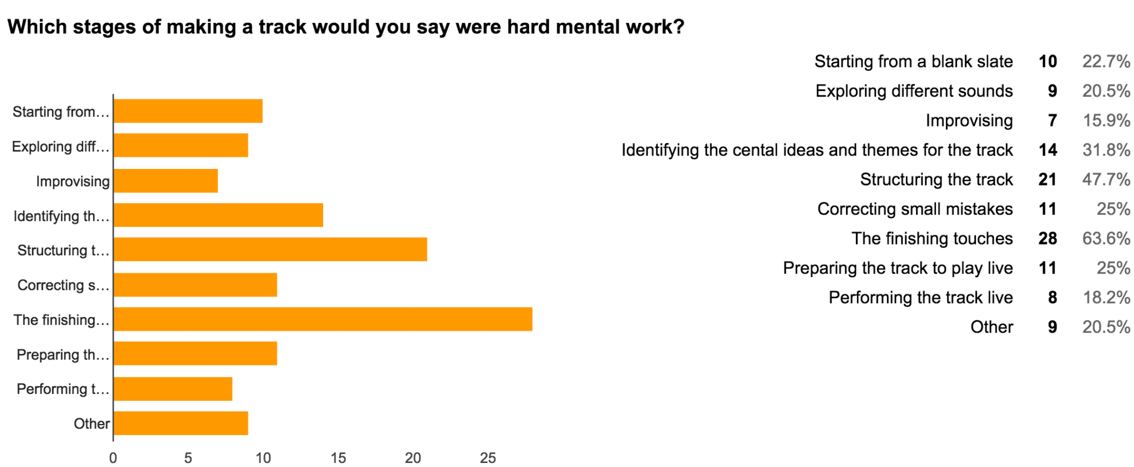


Figure A.25: This question aimed to ascertain which stages of the creative process were associated with high cognitive load. The responses suggest that the mental load increases as the amount of completed material increases. Interestingly, performance and improvisation are rated as the least demanding.

This question aimed to find out which stage of the creative process musicians found hardest in terms of mental load, and by inference, the one that which placed most demands on working memory.

Here, the most popular answer was "The finishing touches" (26 votes). In fact, the responses seemed to indicate a steady increase in difficulty from early stage exploration to final stage honing. Structuring the track was the second highest voted (19 votes). There may be a number of reasons for this: increasing numbers of constraints, increasing amounts of comparative judgements, longer evaluation times scales and more complex units to manipulate would all lead to more mental load.

Surprisingly, fast interactions such as improvisation and performing the track live were not selected as hard work by many (6 and 7 votes respectively). This might suggest the fact that skilled interactions are less cognitively demanding, despite being more demanding in terms of time constraints.

"Starting from a blank state" received only 8 votes. This indicates that the terror sometimes associated with a "blank canvas" does not apply in electronic music.

One respondent notes the difficulty of high-level evaluation:

> "[The] most challenging thing is questioning myself 'is this good enough'?"

If a piece of music technology is complex, fiddly or demanding, how much does it interfere with the following tasks?

This was a series of questions designed to assess the impact of interface design on various aspects of music-making. The three responses were "Not at all", "Somewhat" and "A lot". In general, interference was rated as medium or high for all tasks apart from "Deciding if something is good or not". Another lower interference task was "Exploring and discovering sounds". These two responses seem to rather contradict the "narrowed attention" hypothesis of inhibited evaluative abilities. It

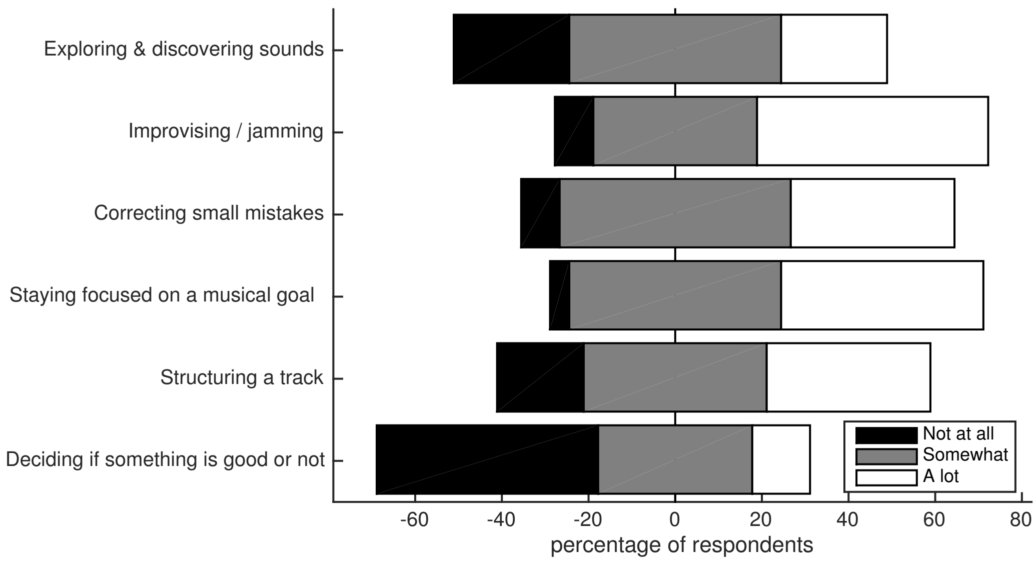


Figure A.26: If a piece of music technology is complex, fiddly or demanding, how much does it interfere with the tasks listed on the left? Results are shown as a stacked bar chart [Robbins and Heiberger, 2011], with middle responses centred.

is however possible that narrowed attention *does* inhibit other abilities, but the user would never realise, and would therefore be unable to report the fact. In other words, studying narrowed attention requires gathering experimental, not anecdotal, evidence.

Most interference was shown with "improvising/jamming", which is not surprising given its real-time nature. This is interesting to contrast with the results in question 12, which show that improvising is not rated as particularly demanding in itself.

The high interference reported for "Staying focused on a musical goal" fits with the hypothesis that the top levels of the artist's goal hierarchy are most sensitive to disruption by cognitive load.

Two respondents objected to this question: "Complexity vs interference seems like a non-question to me. For me, when something is demanding, it will always interfere with your results, unless the result us purely the exploration of it. I have

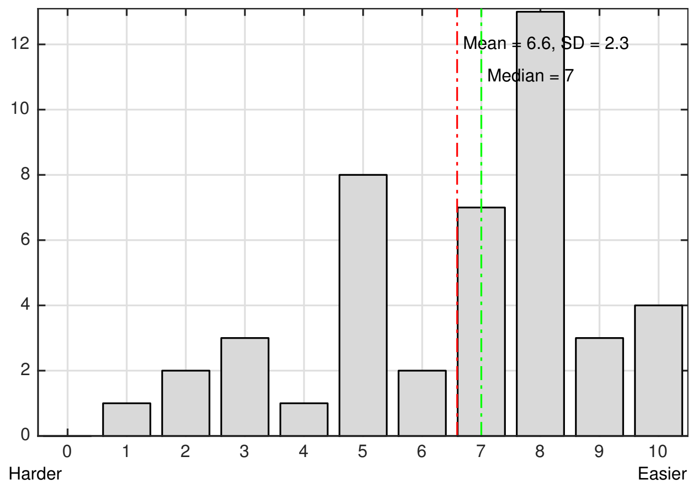


Figure A.27: Q49: Better tracks are easier.

trouble understanding how anyone would feel otherwise.”. And another respondent: “This is a leading question so it answers itself. If a piece of tech[nology] is too fiddly or demanding then of course it will interfere with completing any task.”.

49. Are your best tracks harder work or easier to make than your poorest ones?

Perhaps relating to the idea that a state of Flow feels effortless, and that the feeling of mental load may well be associated with less creative cognitive states, musicians felt that their better tracks actually felt easier to make. An alternative explanation is that sheer chance occurrence of multiple serendipitous events may contribute to the best tracks.

66. Being able to perfect every detail of the music is important to me

In general, respondents agreed with this statement (mode = 7). This relates to the desire for sonic precision and fine detail that is particularly prevalent in electronic music. Indeed, unlimited control may be the chief attraction of software production tools. This result should make us wary of any trade-offs between algorithmic and

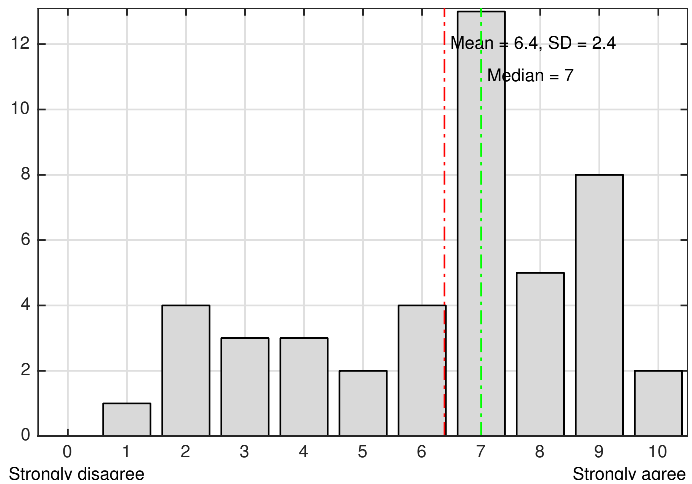


Figure A.28: Q66: Perfection of details.

other EARS modes: if providing for the other three modes comes at the cost of not being able to analytically specify individual aspects of the sound, the design will not find favour in this particular community.

One comment expressed dissatisfaction with this perfectionist approach:

> "The process of *finalising* a track is totally different thing from *creating* it. This is at least how I've always seen it in the past. But I'm beginning to feel as though the last 5% of quality/whatever that I spend 95% of my time trying to achieve is not worth the effort..."

67. How often do you think you lose ideas that are too hard to make happen?

Here, the hypothesis was that a large amount of human creativity can be lost in the transition from mind to machine, due to the complexity of realisation. Whilst there was a wide range of opinion, overall there was an average estimate of around 55% of ideas being lost. This does not seem to agree with the answers for question 29, where most people felt they could realise a sound faithfully. Perhaps this response relates to higher level, more aesthetic ideas rather than the sound design stage? Either way, this lossiness seems quite an alarming waste of creative potential. Would doubling the input channel throughput alleviate the issue?

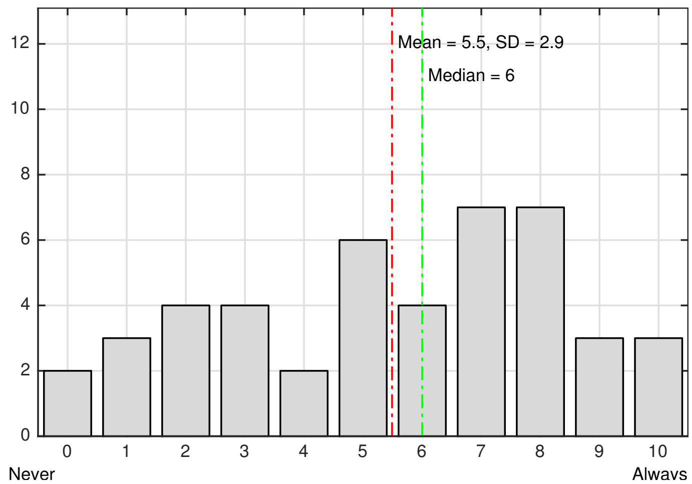


Figure A.29: Q67: Amount of lost ideas due to complexity (0 = 0%, 10 = 100%).

68. Which tasks most interfere with having a good perspective on what you are making?

"Perspective" is a high-level evaluative judgement that artists often refer to as being important, but easily lost. To avoid ambiguity, it was defined in this question as being "knowing what you want to achieve and how well you are progressing towards that goal". It may be related to working memory load, and the depth of the goal stack. Here people could select multiple answers.

Highest rated for interference was "detailed editing" (43%), closely followed by "exploring/experimenting" (41%). Exploration comes in surprisingly high, given its earlier rating as low demand. This may be due to the tendency to get distracted by novel ideas which do not contribute to the music as a whole. "Structuring the track" scores highly (39%)[5]. Other perspective interference suggestions included the inevitable crashes and technical problems, "final production/mixing", and even

[5]Presumably maintaining perspective would be most useful precisely at this stage.

collaboration.

Lowest rated for interference "Improvising/Jamming" (7%). Perhaps a surprising result considering that improvisation is a temporally demanding real-time activity. Could something about in-the-moment live interaction be particularly conducive to successful evaluation? Perhaps it is the fact that improvising gives the musician a direct insight into how a listener would be experiencing the music in real-time: as an unfolding narrative.

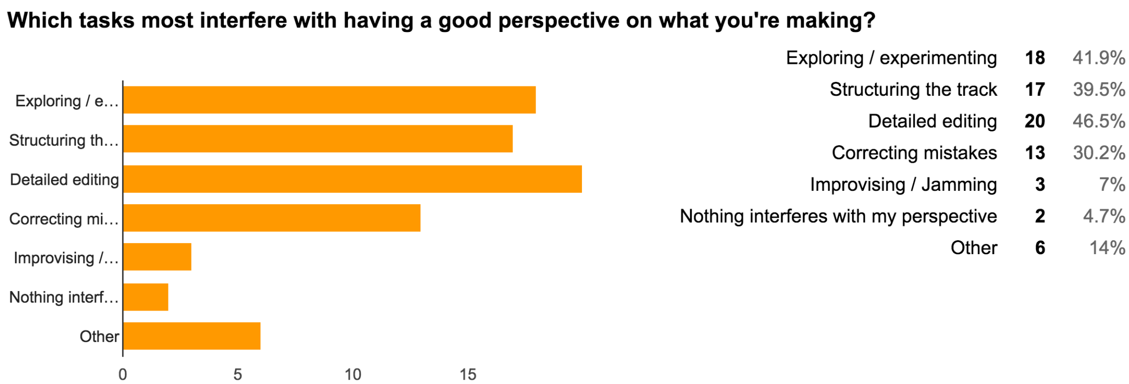


Figure A.30: Q68: Which of these tasks most interfere with having a good perspective on what you are making?

A.6 Skill and Automaticity

55. I have deliberate practice sessions, where I aim to get better at physically controlling a particular device.

It was hypothesised that the ability to edit all aspects of a piece would inhibit the necessity and the motivation to practice motor skills. By far the majority of respondents (61%) do practice occasionally but not systematically. 23% never practised. 14% practised every few weeks and only 2% claimed to practice more than once a week. Unfortunately there was no question relating to whether musicians felt they would actually benefit from more motor skill based interaction.

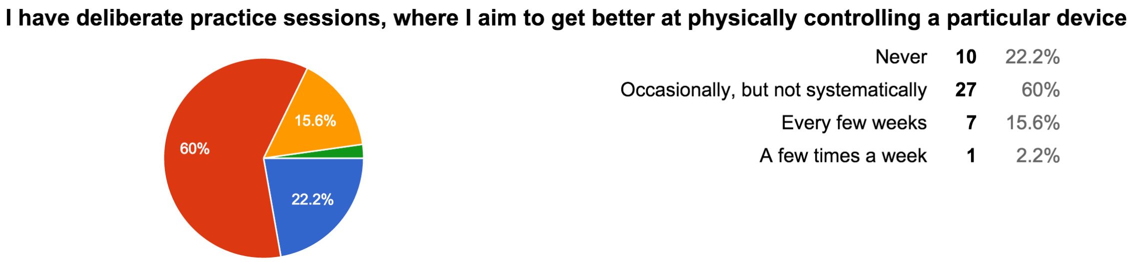


Figure A.31: Q55: Deliberate practise sessions.

15. I wish I could break out of my habits more often.

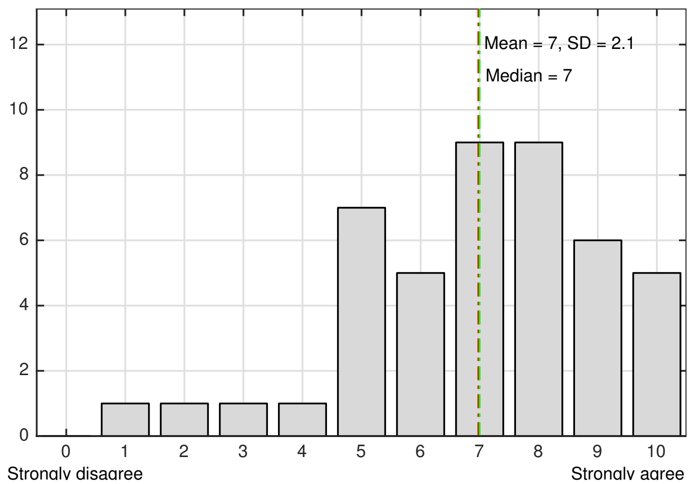


Figure A.32: Q15: Habits.

This question was intended to find out if musicians struggle to overcome their automatic responses to certain situations, and respondents almost all agreed with this statement. It is clear that skill can be a double edged sword. This observation motivates the “exploratory interface” presented in Chapter 5, which is intended to interfere with the artist’s automatic, predictive cognitive mechanisms.

This respondent noted that it may be habits relating to the high-level goals that may be the most restricting[6]:

> “I think the biggest bar to my creativity isn’t so much the technology,

[6]Low level exploratory interfaces such as Sonic Zoom (Chapter 5) would not address this problem. Perhaps there is room for developing a higher level, reflective divergence support software systems?

but more my preconceived ideas about the kind of tune I'm trying to write. I get stuck in certain expectations of what my own sound is, and sometimes end up effectively doing a pastiche of my own sound."

16. I find traditional instruments enable me to express a feeling better than computer-based instruments

People without experience of traditional instruments were told to skip this question. The hypothesis was that traditional instruments would be preferred, as being better for immediate expression of complex affective states. Opinion here was widely distributed, with a spike at the middle response, but with a tendency to disagree. So, in fact, this community feels able to express feelings very well with digital tools, and presumably find that having a wider sonic palette offsets any disadvantage in immediacy. One commented: "I find using traditional instruments strongly limits the potential for more obscure ideas and tracks". So, for many musicians, the exploratory benefit of having a bigger sonic space takes precedent over the skilled aspects of interaction.

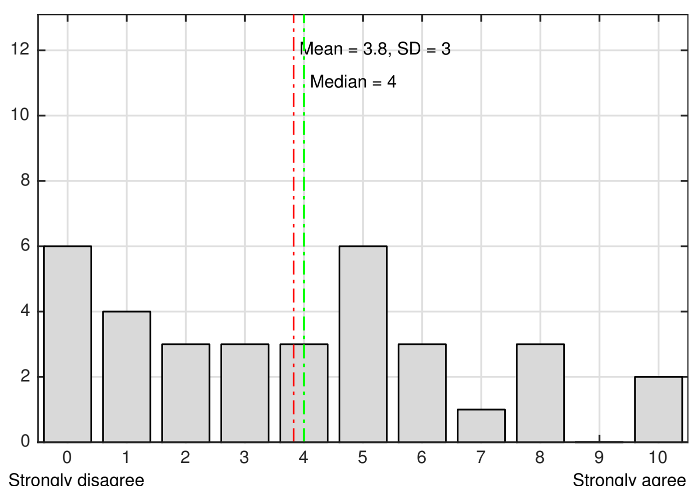


Figure A.33: Q16: Traditional instruments more expressive.

On the other hand, this respondent highlighted the desirability of the stable, predefined control mappings that hardware offers:

> "If by 'traditional instruments' you count hardware synths not connected to a computer or sequencer, then I would say these do allow for better expression than VST instruments. Even with midi controllers applied to VSTs you have to make the conscious effort to assign knobs to a premeditated subset of parameters, rather than being permanently restricted to whatever parameters the hardware synth presents you with (limitation is a good thing)."

46. How many of the creative decisions you make would you say are "intuitive"?

Here musicians clearly favour instinctive responses over careful deliberation. Intuitive thinking was deemed to determine around 70% of decisions. There is some ambiguity whether this is a choice or a necessity: is it because musicians want to use their gut feeling? Or are they forced to because explicit evaluation is too slow, or too difficult whilst engaged with complex technology?

One participant commented on the intuitive aspects in more depth, even claiming that the process may be outside conscious access, such that one cannot explicitly recall how the track was made. They also draw a distinction between the initial idea and the emergent content, and note the distortion of the passage of time that is one of the dimensions of Flow:

> "Actually, the (creative) workflow never requires brainwork, especially not hard (compared to my full time job as a creative director). In this case you may ask: 'then what makes it creative, if there's no or there is a mild conceptual level in it?'. Well, the answer is every track have some kind of a starting point [initial idea]. I think most of the magic

happens there OR if you carried away with the furore/flow you currently creating - you may never being able to subsequently explain or describe what happened in the past few hours. How you did this, or that...

...this is purely magical if you never experienced anything like this. Whenever you notice that it is already 3 am, and you spent more than 8 hours in the making (without noticing how's time passed), you finished the track, and you cannot remember every detail, how you made your sounds, etc... But if you listen it from the beginning, there's some minor mistakes, etc, but overall it is an absolute madness [good] in both technical and musical aspects. And you did that, but you cannot remember, it's not a conscious process!

In the other case, you may have a melody, or a sample keeps running in your head, an inspirational track you heard, or you are in a special mood that you think you could share with the rest of the world - the starting point could be anything. The challenge is how you can adapt this "brain" thing from scratch into a musical form. You need to "research, develop and build up" the rest of the track - within [that] context."

A.7 Reflection and Evaluation

19. Do you try to analyse your work-flow and make it more efficient?

The majority answered in the positive. This shows not only that "efficiency" is very important for musicians, but that they also take active steps to reflect on their own practice and improve it.

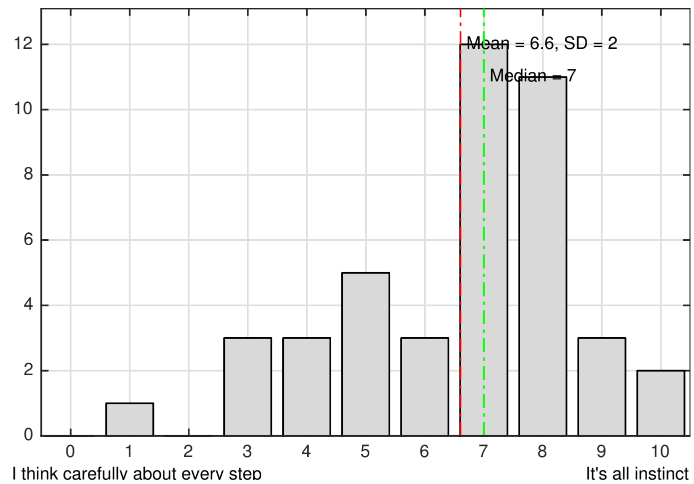


Figure A.34: Q46: Decisions intuitive.

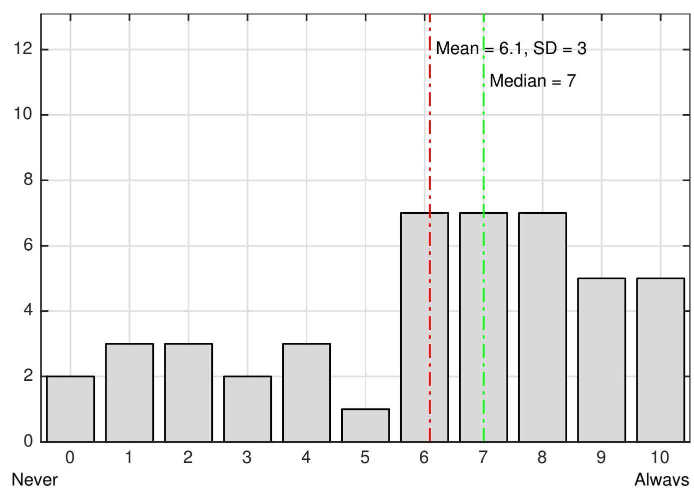


Figure A.35: Q19: Reflection on efficiency.

40. Do you improve your creative process by consciously analysing how you created your best tracks?

Another inquiry into reflective practice. The initial hypothesis was that because it is difficult to remember such a complex and absorbing interaction, reflection on the process would be infrequent. In fact, a large number of musicians did carry out this type of meta-analysis (Mode = 7,8 $N(7,8) = 10$).

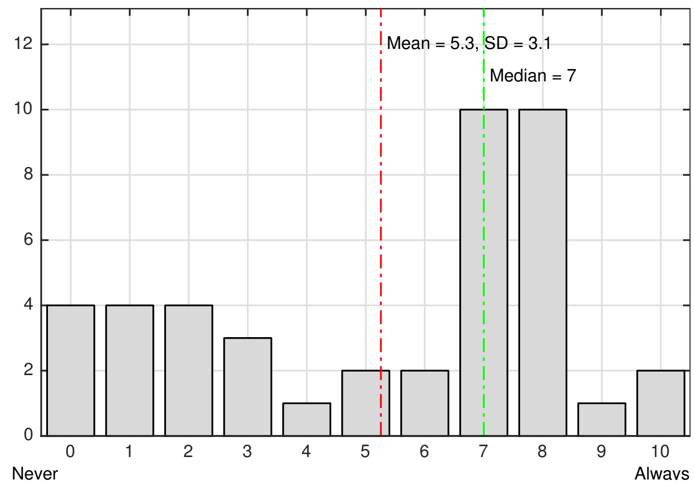


Figure A.36: Q40: Reflective practice.

38. Which is more important for producing good music, technical skill or musical taste?

Can one approach creativity solely on the basis of evaluation and selection? Or does one need technical skill and execution to produce valuable works? Interestingly, results here seem to be in favour of musical taste, perhaps related to the close ties between electronic music and DJ culture. This emphasises that the sophistication of the value system one uses for judging output may be at least as important, if not more, than the ability to generate the output effectively. This implies BVSR-type creativity models may be viable [Simonton, 2012].

44. Do you evaluate music by comparing to other music, or just go with how it makes you feel?

The slight tendency here was to go with intuition over comparative judgements. Many text comments indicated that it tends to be that early stage creativity is more based on intuitive, implicit judgements, late stage mixing and mastering is more analytic and comparative:

> "The comparative part is mostly by how well the music is mixed/mastered."

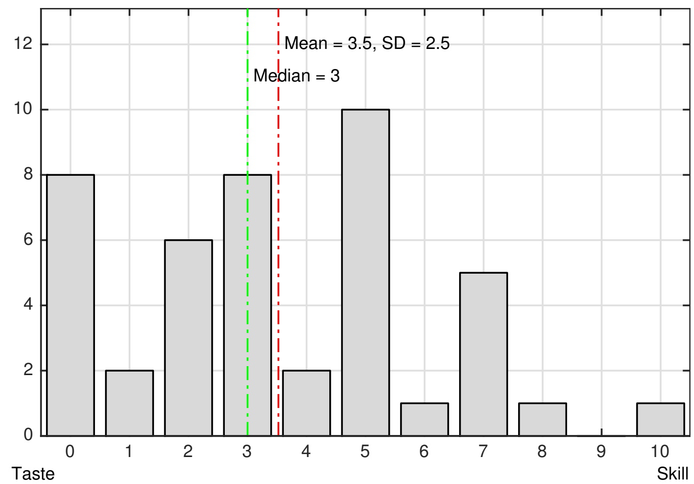


Figure A.37: Q38: Skill or taste.

"Again, different at different parts of process. A weakness I have personally is "mixdown-insecurity"—I tend to compare to other tracks at this stage. Earlier in the process it's almost entirely feeling."

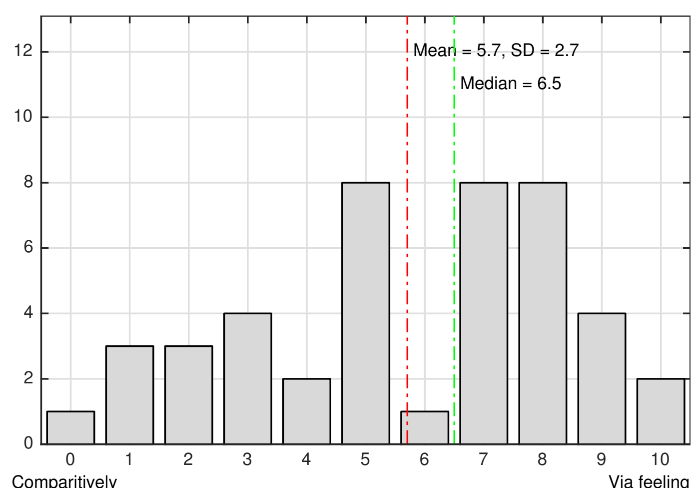


Figure A.38: Q44: Comparative / intuitive evaluation.

45. Is your judgement of the quality of a track when listening in different contexts quite consistent, or different each time?

Like question 32, this relates to how context-sensitive the evaluation process is, and again the answers were that judgements varied considerably based on the situation (Mode $= 8$, $N(8) = 15$). The fitness function would appear to be a highly flexible thing.

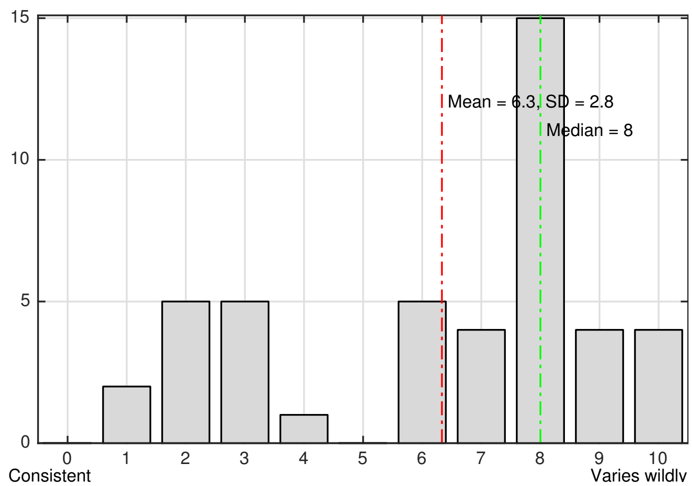


Figure A.39: Q45: Evaluation of a whole track is context-sensitive.

42. The more I think about my creative decisions, the more confident I am in them.

This question produced one of the widest distributions of all. The intent was to find out if deliberation, and analytical thinking was seen to decrease confidence in the value of the output. The hypothesis being that in such a complex underspecified problem such as musical creativity, over-analysis would be seen as harmful, and hence there would be disagreement with this statement, however the response was ambiguous. This response should be contrasted with the response to question 45 and 46, which indicated that decisions are more often made intuitively. Looked at together, this would indicate that more analytic evaluation would, in theory, lead to better end results, but in practise is too demanding. One user speculates that fatigue may be an issue here:

> "Over-thinking leads to fatigue, which changes the way you perceive. The more I consider decisions I've made the more I'm likely to bugger about and backtrack. The solution can either be to whip through everything like a lunatic, not looking sideways or backwards (flow?) or to delay decision making until an idea has been meticulously explored. The delay approach often leads to more confusion as you hold different contingencies and comparisons in mind. Beginning to think more and more that the first good idea wins, and move on. There is probably a happy medium, but how to enforce such an approach?"

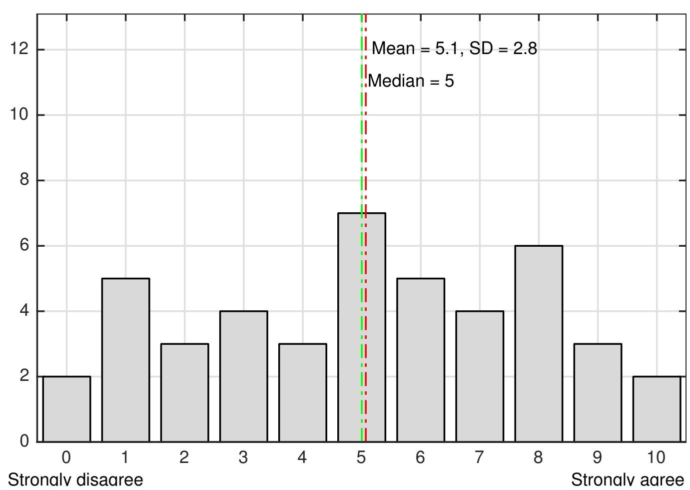


Figure A.40: Q42: Analysis increases confidence.

A.8 Conclusions

Based on the responses in Section A.2, it appears that real-time control of timbre is a key problem in electronic music production. If higher throughput, lower cognitive load interaction methods could be achieved, it would assist with sound design, structuring, live performance and quite possibly general levels of creativity.

Exploratory, curiosity-driven, emergent and serendipitous discoveries seem to constitute the bulk of the creative process, an estimated two thirds of interaction being of this type. Implicit decision-making also was rated as being an essential ingredient, again with two thirds of decisions being considered intuitive. Late stage creativity, such as long term structure, editing fine detail, and the final mix-down/mastering process are considered more demanding, comparative and analytical.

Subjects report significant levels of interference with many creative tasks if interface-induced cognitive load is high. Evaluation, however, was rated as being less interfered with, whilst rated as being more important than technical ability.

There is widespread recognition that self-imposed constraints are a good thing, and more options and features do not necessarily lead to better results: indeed may cause confusion and hamper progress. Comments indicate that there seems to be an instinctive understanding that constraints encourage a switch to reflective meta-exploration.

Some interesting differences between the questions related to mental load emerged. Small edits were not rated as particularly hard work (Q.12). But yet they were rated highly in terms of losing perspective (Q.68). Perhaps loss of perspective is not purely a question of mental load. Narrowed attention may play a part, unfortunately this survey was not precise enough to tease apart these constructs. The main challenge with assessing the "mental load" responses is establishing whether the work load is intrinsic to the task itself, or an artefact of the current state of interface design. For instance, "structuring a track" may be expected to be difficult, given the complexity and interdependence of the units being manipulated, the strong constraint that good existing components should not be damaged in any way, and the long feedback time required to evaluate the results of an alteration. On the other hand, very few DAW systems provide simple visualisation or easy manipulation of song structure, so perhaps the interface may be partially to blame.

Despite the fact that over 60 questions were presented to 45 respondents, many further questions remain, and in retrospect some quite useful information was not obtained. Some questions were intended to be provocative, but this may have backfired by encouraging people to sit on the fence. The optional text responses often justified this ambiguity in more detail, a great many starting with "Yes and no...". This strongly indicates that creativity is often a blending of opposing tendencies: almost any technique can be subverted or inverted and still constitute a valid tool in the artist's creative arsenal. Many questions with large spikes for the neutral value may have reflected the discomfort described by this participant:

> "The problem for me in a lot of the scenarios outlined by these questions is that at a specific moment I would favour a particular end of the spectrum whilst in another moment the exact opposite would be desirable."

Some questions may have been misinterpreted and would have benefited from less ambiguous wording. Even the various very clear responses could have multiple explanations, which would require a follow up survey to investigate fully.

Another difficulty with this kind of survey is that we only receive the subjective opinions of the respondents, and of course they may be biased or mistaken. If, as suggested, their reflective capacity is compromised in situations of high workload, then this will inhibit their ability to objectively self-report in precisely those situations we are most interested in. This survey was not conducted in as rigorous a fashion as would be required for reporting as scientific results, and therefore should more be considered as background motivation for this research programme, rather than providing concrete evidence for any of the theories advanced in Chapter 5.

Despite these reservations, the responses bear out a number of predictions from the theoretical part of this thesis, and serve to further motivate the experimental work.

APPENDIX B

Participant instruction texts

B.1 Introductory text for Sonic Zoom

The following test appeared on the iTunes App Store as an introduction to the SOnic Zoom application used for Experiment 1 (Chapter 6).

> Sonic Zoom is a PhD research project from Queen Mary University. The app aims to look at how people interact with music synthesisers: how they adjust parameters and explore the vast range of sounds on offer.
>
> There are two interfaces presented. The first is fairly standard: ten sliders that control each parameter. The second is more novel: a two-dimensional surface that can be scrolled and zoomed similar to a map. Every sound that can be made with the synth is located somewhere on this surface. If you find a sound you like, you can zoom in on it to explore smaller variations. You can save a sound and this will drop a marker

on the surface. These markers are easy to revisit and can be smoothly interpolated between. Your path through the sound space is visible, so you can retrace your steps.

The first 15 minutes is a timed experiment. Users will then be asked to complete a quick questionnaire. After this you are free to use the app, and a few extra features will be unlocked as a reward.

The following introductory text was shown to all participants in the Sonic Zoom experiment on starting the app.

> Welcome to “Sonic Zoom”. This is a Queen Mary University PhD research project aimed at finding out how people explore sound synthesis parameters.
>
> In this app there are just 10 parameters for a somewhat basic FM/subtractive synthesizer. However, even with only 10 parameters the amount of different sounds to explore is vast: in fact there are just over a billion trillion distinct settings! We aim to look at what paths people take in this huge space, what points they like and dislike, and use the data to create synths that are easier to navigate, and hopefully encourage creativity.
>
> There are two interfaces presented here. The first is one you will probably be used to if you are an electronic musician: 10 sliders for each of the parameters. The second is somewhat new: every one of the billion trillion points has been mapped to a 2D surface. But basically, the further you travel along this 2D surface, the more different the sound will become.
>
> When you click “save preset” you will drop a pin onto the surface, and you can revisit this point at anytime by scrolling to it. Think of it like a “Google Maps” for synth sounds...

You can use a pinch gesture to zoom in and out of the surface. In this way, you can explore the "neighbourhood" of a particular sound. Zooming in will enable you to explore smaller localities of the sound space. Zooming out will give you a bigger perspective, but of course the transitions will become more sudden and random as you move bigger distances.

What we are asking you, the participant, to do is use the Zoomer, the sliders and a combination of both for 5 minutes each, and save any sounds that you like. The interfaces will swap automatically. Also check out the "randomise" and "lock" buttons. After this a short questionnaire will appear that will help us assess the interface further. After that, please feel free to use it some more: the more data we get the better. As a bonus some extra features will be unlocked when you complete the questionnaire: MIDI out and smooth interpolation mode.

Please note: Your actions will be logged, and sent to a secure server here at Queen Mary University. However no personal data (name, email etc.) will be collected, associated with this ID or stored in the database. It is highly recommended that you are connected to WiFi internet whilst using this App, otherwise the log data will take up space on your device.

To consent to this please press "Agree" below. To decline just exit the app.

Bibliography

Accot, J. and Zhai, S. (1999). Performance evaluation of input devices in trajectory-based tasks: an application of the steering law. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*, pages 466–472. ACM.

Aggarwal, C. C., Hinneburg, A., and Keim, D. A. (2001). *On the surprising behavior of distance metrics in high dimensional space*. Springer.

Allouis, J. and Bernier, J. (1982). The syter project: Sound processor design and software overview. In *Proceedings of the 1982 International Computer Music Conference*, pages 232–240.

Amabile, T. M. (1996). *Creativity and innovation in organizations*. Harvard Business School, Boston, Massachusetts.

Amabile, T. M. (1998). *How to kill creativity*. Harvard Business School Publishing.

Amabile, T. M. and Gitomer, J. (1984). Children's artistic creativity effects of choice in task materials. *Personality and Social Psychology Bulletin*, 10(2):209–215.

Anderson, J. R. (1981). *Cognitive skills and their acquisition*. Psychology Press.

Anderson, J. R. (1983). A spreading activation theory of memory. *Journal of Verbal Learning and Verbal Behavior*, 22(3):261–295.

Anderson, J. R., Matessa, M., and Lebiere, C. (1997). ACT-R: A theory of higher level cognition and its relation to visual attention. *Human-Computer Interaction*, 12(4):439–462.

André, P., Teevan, J., Dumais, S. T., and Schraefel, M. C. (2009). Discovery is never by chance: Designing for (un) serendipity. In *Proceedings of the Seventh ACM Conference on Creativity and Cognition*, pages 305–314. ACM.

Ansburg, P. I. and Hill, K. (2003). Creative and analytic thinkers differ in their use of attentional resources. *Personality and Individual Differences*, 34(7):1141–1152.

Antle, A. N., Corness, G., and Droumeva, M. (2009). What the body knows: Exploring the benefits of embodied metaphors in hybrid physical digital environments. *Interacting with Computers*, 21(1):66–75.

Arfib, D., Couturier, J. M., Kessous, L., and Verfaille, V. (2002a). Strategies of mapping between gesture data and synthesis model parameters using perceptual spaces. *Organised Sound*, 7:127–144.

Arfib, D., Couturier, J.-M., Kessous, L., Verfaille, V., et al. (2002b). Strategies of mapping between gesture data and synthesis model parameters using perceptual spaces. *Organised Sound*, 7(2):127–144.

Arfib, D., Jean-Michel, C., and Loïc, K. (2005). Expressiveness and digital musical instrument design. *Journal of New Music Research*, 34(1):125–136.

Armstrong, N. (2006). *An Enactive Approach to Digital Musical Instrument Design*. PhD thesis, Princeton University.

Attwell, D. and Laughlin, S. B. (2001). An energy budget for signaling in the grey matter of the brain. *Journal of Cerebral Blood Flow & Metabolism*, 21(10):1133–1145.

Ayaz, H., Shewokis, P. A., Bunce, S., Izzetoglu, K., Willems, B., and Onaral, B. (2012). Optical brain monitoring for operator training and mental workload assessment. *NeuroImage*, 59(1):36 – 47.

Baars, B. J. (1997). Some essential differences between consciousness and attention, perception, and working memory. *Consciousness and cognition*, 6(2):363–371.

Baars, B. J. (2005). Global workspace theory of consciousness: toward a cognitive neuroscience of human experience. *Progress in Brain Research*, 150:45–53.

Bachmann, D., Weichert, F., and Rinkenauer, G. (2014). Evaluation of the Leap Motion controller as a new contact-free pointing device. *Sensors*, 15(1):214–233.

Baddeley, A. (1992). Working memory. *Science*, 255(5044):556–559.

Baird, B., Smallwood, J., Mrazek, M. D., Kam, J. W., Franklin, M. S., and Schooler, J. W. (2012). Inspired by distraction: Mind wandering facilitates creative incubation. *Psychological Science*, 23(10):1117–1122.

Bakker, A. B. (2005). Flow among music teachers and their students: The crossover of peak experiences. *Journal of Vocational Behavior*, 66(1):26–44.

Ball, P. (2010). *The music instinct: how music works and why we can't do without it*. Random House.

Banich, M. T. and Compton, R. (2011). *Cognitive neuroscience: Third Edition*. Cengage Learning.

Bar, M. (2007). The proactive brain: using analogies and associations to generate predictions. *Trends in Cognitive Sciences*, 11(7):280–289.

Bar, M., Kassam, K. S., Ghuman, A. S., Boshyan, J., Schmid, A. M., Dale, A. M., Hämäläinen, M. S., Marinkovic, K., Schacter, D. L., Rosen, B. R., et al. (2006). Top-down facilitation of visual recognition. *Proceedings of the National Academy of Sciences of the United States of America*, 103(2):449–454.

Barbosa, J., Malloch, J., Wanderley, M. M., and Huot, S. (2015). What does "Evaluation" mean for the NIME community? In *NIME 2015 - 15th International Conference on New Interfaces for Musical Expression*, page 6, Baton Rouge, United States. Louisiana State University.

Bargar, R., Choi, I., Das, S., and Goudeseune, C. (1994). Model based interactive sound for an immersive virtual environment. In *Proceedings of the International Computer Music Conference*, pages 471–471.

Barlindhaug, G. (2007). Analog sound in the age of digital tools. the story of the failure of digital technology. *A document (re) turn: Contributions from a research field in transition*, pages 73–93.

Barrett, F. J. (1998). Coda: creativity and improvisation in jazz and organizations: Implications for organizational learning. *Organization Science*, 9(5):605–622.

Bartholdi, J. and Platzman, L. (1982). An o(n log n) planar travelling salesman heuristic based on spacefilling curves. *Operations Research Letters*, 1(4):121 – 125.

Barto, A. G. (1998). *Reinforcement learning: An introduction*, pages 25–49. MIT press.

Baumer, E. P. (2015). Reflective informatics: Conceptual dimensions for designing technologies of reflection. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, pages 585–594. ACM.

Beaudouin-Lafon, M. (2000). Instrumental interaction: an interaction model for designing post-WIMP user interfaces. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 446–453. ACM.

Bederson, B. and Meyer, J. (1998). Implementing a zooming user interface: experience building pad++. *Softw., Pract. Exper.*, 28(10):1101–1135.

Bederson, B. B. (2004). Interfaces for staying in the flow. *Ubiquity*, 2004(September):1–1.

Bencina, R. (2005). The Metasurface - applying natural neighbour interpolation to two-to-many mapping. In *Proceedings of the 2005 International Conference on New Interfaces for Musical Expression (NIME05), Vancouver, BC, Canada*.

Benedek, M. and Neubauer, A. C. (2013). Revisiting Mednick's model on creativity-related differences in associative hierarchies. evidence for a common path to uncommon thought. *The Journal of Creative Behavior*, 47(4):273–289.

Bengler, B. and Bryan-Kinns, N. (2014). In the wild: evaluating collaborative interactive musical experiences in public settings. In *Interactive Experience in the Digital Age*, pages 169–186. Springer.

Bengtsson, S. L., Csíkszentmihályi, M., and Ullén, F. (2007). Cortical regions involved in the generation of musical structures during improvisation in pianists. *Journal of Cognitive Neuroscience*, 19(5):830–842.

Bergeron, V. and Lopes, D. M. (2009). Hearing and seeing musical expression. *Philosophy and Phenomenological Research*, 78(1):1–16.

Berlyne, D. E. (1970). Novelty, complexity, and hedonic value. *Perception & Psychophysics*, 8(5):279–286.

Bernstein, N. A. (1967). *The co-ordination and regulation of movements.* Pergamon Press Ltd.

Bhat, G. S., Savage, C. D., et al. (1996). Balanced gray codes. *The Electronic Journal of Combinatorics*, 3(1):25.

Bispham, J. (2006). Rhythm in music: What is it? Who has it? And why? *Music Perception*, 24(2):125.

Blackwell, A. and Collins, N. (2005). The programming language as a musical instrument. *Proceedings of PPIG05 (Psychology of Programming Interest Group).*

Blum, C. and Roli, A. (2003). Metaheuristics in combinatorial optimization: Overview and conceptual comparison. *ACM Computing Surveys (CSUR)*, 35(3):268–308.

Boden, M. A. (1992). *The Creative Mind: Myths and Mechanisms.* Abacus.

Bongers, B. (2000). Physical interfaces in the electronic arts. *Trends in gestural control of music*, pages 41–70.

Bor, D. and Seth, A. K. (2012). Consciousness and the prefrontal parietal network: insights from attention, working memory, and chunking. *Frontiers in Psychology*, 3.

Boritz, J. and Booth, K. (1998). A study of interactive 6 dof docking in a computerised virtual environment. In *Virtual Reality Annual International Symposium, 1998. Proceedings., IEEE 1998*, pages 139–146.

Borst, J. P., Taatgen, N. A., and van Rijn, H. (2015). What makes interruptions disruptive?: A process-model account of the effects of the problem state bottleneck on task interruption and resumption. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, CHI '15, pages 2971–2980, New York, NY, USA. ACM.

Bown, O. (2012). Generative and adaptive creativity: A unified approach to creativity in nature, humans and machines. In *Computers and Creativity*, pages 361–381. Springer.

Bown, O. and Lexer, S. (2006). Continuous-time recurrent neural networks for generative and interactive musical performance. In *Applications of Evolutionary Computing*, pages 652–663. Springer.

Boyle, J., Kennedy, D., and Shea, C. H. (2012a). Optimizing the control of high ID movements: rethinking the obvious. *Experimental Brain Research*, 223(3):377–387.

Boyle, J., Panzer, S., Wright, D., and Shea, C. H. (2012b). Extended practice of reciprocal wrist and arm movements of varying difficulties. *Acta Psychologica*, 140(2):142–153.

Bresson, J., Agon, C., and Assayag, G. (2011). Openmusic: Visual programming environment for music composition, analysis and research. In *Proceedings of the 19th ACM International Conference on Multimedia*, MM '11, pages 743–746, New York, NY, USA. ACM.

Brown, A. R. and Sorensen, A. (2009). Interacting with generative music through live coding. *Contemporary Music Review*, 28(1):17–29.

Buchanan, B. G. (2001). Creativity at the metalevel: AAAI-2000 presidential address. *AI Magazine*, 22(3):13.

Buckner, R. L., Andrews-Hanna, J. R., and Schacter, D. L. (2008). The brain's default network. *Annals of the New York Academy of Sciences*, 1124(1):1–38.

Burleson, W. (2005). Developing creativity, motivation, and self-actualization with learning systems. *International Journal of Human-Computer Studies*, 63(4–5):436 – 451. Computer support for creativity Computer support for creativity.

Butz, M. R. (1992). The fractal nature of the development of the self. *Psychological Reports*, 71(3f):1043–1063.

Buxton, B. (1997). Artists and the art of the luthier. *ACM SIGGRAPH Computer Graphics*, 31(1):10–11.

Campbell, D. T. (1960). Blind variation and selective retentions in creative thought as in other knowledge processes. *Psychological Review*, 67(6):380.

Cannon, J. and Favilla, S. (2012). *The Investment of Play: Expression and Affordances in Digital Musical Instrument Design*. Ann Arbor, MI: MPublishing, University of Michigan Library.

Caramiaux, B., Tahiroğlu, K., Fiebrink, R., and Tanaka, A., editors (2014). *Proceedings of the International Conference on New Interfaces for Musical Expression*. Goldsmiths, University of London, UK.

Card, S. K., Moran, T. P., Newell, A., et al. (1983). *The psychology of human-computer interaction*. Erlbaum Hillsdale, NJ.

Cardoso, A., Veale, T., and Wiggins, G. A. (2009). Converging on the divergent: The history (and future) of the international joint workshops in computational creativity. *AI Magazine*, 30(3):15.

Cascone, K. (2002). Laptop music-counterfeiting aura in the age of infinite reproduction. *Parachute*, pages 52–59.

Cha, Y. and Myung, R. (2010). Extended Fitts' law in three-dimensional pointing tasks. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, volume 54, pages 972–976. SAGE Publications.

Chalmers, D. J. (1995). Facing up to the problem of consciousness. *Journal of Consciousness Studies*, 2(3):200–219.

Chemero, A. (2011). *Radical embodied cognitive science*. MIT press.

Chennu, S., Finoia, P., Kamau, E., Allanson, J., Williams, G. B., Monti, M. M., Noreika, V., Arnatkeviciute, A., Canales-Johnson, A., Olivares, F., et al. (2014). Spectral signatures of reorganised brain networks in disorders of consciousness. *PLoS Computational Biology*, 10(10):e1003887.

Cherry, E. and Latulipe, C. (2014). Quantifying the creativity support of digital tools through the creativity support index. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 21(4):21.

Choi, I. (2000). A manifold interface for kinesthetic notation in high-dimensional systems. In Wanderley, M. and Battier, M., editors, *Trends In Gestural Control Of Music*. Ircam - Centre Pompidou.

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences*, 36(03):181–204.

Clark, A. and Chalmers, D. (1998). The extended mind. *Analysis*, pages 7–19.

Cockburn, A. and Brewster, S. (2005). Multimodal feedback for the acquisition of small targets. *Ergonomics*, 48(9):1129–1150.

Cockburn, A. and McKenzie, B. (2002). Evaluating the effectiveness of spatial memory in 2d and 3d physical and virtual environments. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '02, pages 203–210, New York, NY, USA. ACM.

Collins, N. (2008). The analysis of generative music programs. *Organised Sound*, 13(03):237–248.

Collins, N., McLean, A., Rohrhuber, J., and Ward, A. (2003). Live coding in laptop performance. *Organised Sound*, 8(03):321–330.

Colton, S., Wiggins, G. A., et al. (2012). Computational creativity: The final frontier? In *Proceedings of the European Conference on Artificial Intelligence*, pages 21–26.

Conner, B. D., Snibbe, S. S., Herndon, K. P., Robbins, D. C., Zeleznik, R. C., and Van Dam, A. (1992). Three-dimensional widgets. In *Proceedings of the 1992 Symposium on Interactive 3D Graphics*, pages 183–188. ACM.

Cook, P. (2001). Principles for designing computer music controllers. In *Proceedings of the conference on New Interfaces for Musical Expression*, pages 1–4. National University of Singapore.

Cormen, T. H., Leiserson, C. E., Rivest, R. L., Stein, C., et al. (2001). *Introduction to algorithms*, volume 2. MIT press Cambridge.

Cruz-Neira, C., Leigh, J., Papka, M., Barnes, C., Cohen, S. M., Das, S., Engelmann, R., Hudson, R., Roy, T., Siegel, L., et al. (1993). Scientists in wonderland: A report on visualization applications in the CAVE virtual reality environment. In *Proceedings of the IEEE 1993 Symposium on Research Frontiers in Virtual Reality.*, pages 59–66. IEEE.

Csíkszentmíhályi, M. (1991). *Flow: The Psychology of Optimal Experience*. Harper Perennial, New York, NY.

Csíkszentmíhályi, M. (1999). 16 implications of a systems perspective for the study of creativity. *Handbook of Creativity*, page 313.

Csíkszentmíhályi, M. (2009). *Creativity: Flow and the Psychology of Discovery and Invention*. HarperCollins.

Dahlstedt, P. (2001). Creating and exploring huge parameter spaces: Interactive evolution as a tool for sound generation. In *Proceedings of the 2001 International Computer Music Conference*, pages 235–242.

Davis, N., Popova, Y., Sysoev, I., Hsiao, C.-P., Zhang, D., and Magerko, B. (2014). Building artistic computer colleagues with an enactive model of creativity. In *Proceedings of the Fifth International Conference on Computational Creativity.*, pages 38–45.

Dawkins, R. (2006). *The selfish gene*. Oxford University Press.

Dayan, P., Hinton, G. E., Neal, R. M., and Zemel, R. S. (1995). The Helmholtz machine. *Neural Computation*, 7(5):889–904.

De Campo, A. (2014). Lose control, gain influence–concepts for metacontrol. In *Proceedings of the Sound and Music Conference (SMC)*.

Dehaene, S. (2014). *Consciousness and the brain: Deciphering how the brain codes our thoughts*. Penguin.

Dehaene, S., Charles, L., King, J.-R., and Marti, S. (2014). Toward a computational theory of conscious processing. *Current Opinion in Neurobiology*, 25:76–84.

Dehaene, S., Kerszberg, M., and Changeux, J.-P. (1998). A neuronal model of a global workspace in effortful cognitive tasks. *Proceedings of the National Academy of Sciences*, 95(24):14529–14534.

Dennett, D. C. (1995). Darwin's dangerous idea. *The Sciences*, 35(3):34–40.

Despain, A. and Westervelt, R. (1997). High performance human-computer interfaces. Technical Report JSR-96-130, Advanced Research Projects Agency (ARPA).

Deterding, S., Sicart, M., Nacke, L., O'Hara, K., and Dixon, D. (2011). Gamification. using game-design elements in non-gaming contexts. In *CHI'11 Extended Abstracts on Human Factors in Computing Systems*, pages 2425–2428. ACM.

Diaz, F. M. (2013). Mindfulness, attention, and flow during music listening: An empirical investigation. *Psychology of Music*, 41(1):42–58.

Dietrich, A. (2003). Functional neuroanatomy of altered states of consciousness: the transient hypofrontality hypothesis. *Consciousness and Cognition*, 12(2):231–256.

Dietrich, A. (2004a). The cognitive neuroscience of creativity. *Psychonomic Bulletin & Review*, 11(6):1011–1026.

Dietrich, A. (2004b). Neurocognitive mechanisms underlying the experience of flow. *Consciousness and Cognition*, 13(4):746–761.

Dietrich, A. (2007). Who's afraid of a cognitive neuroscience of creativity? *Methods*, 42(1):22–27.

Dietrich, A. and Haider, H. (2014). Human creativity, evolutionary algorithms, and predictive representations: The mechanics of thought trials. *Psychonomic Bulletin & Review*, pages 1–19.

Dietrich, A. and Kanso, R. (2010). A review of EEG, ERP, and neuroimaging studies of creativity and insight. *Psychological Bulletin*, 136(5):822.

Dijksterhuis, A. and Nordgren, L. F. (2006). A theory of unconscious thought. *Perspectives on Psychological Science*, 1(2):95–109.

DiPaola, S., McCaig, G., Carlson, K., Salevati, S., and Sorenson, N. (2013). Adaptation of an autonomous creative evolutionary system for real-world design application based on creative cognition. In *Proceedings of the Fourth International Conference on Computational Creativity*, volume 40.

Djajadiningrat, T., Matthews, B., and Stienstra, M. (2007). Easy doesn't do it: skill and expression in tangible aesthetics. *Personal and Ubiquitous Computing*, 11(8):657–676.

Dobrian, C. and Koppelman, D. (2006). The 'E' in nime: Musical expression with new computer interfaces. In *Proceedings of the 2006 conference on New Interfaces for Musical Expression*, pages 277–282. IRCAM—Centre Pompidou.

Doornbusch, P. (2002). Composers' views on mapping in algorithmic composition. *Organised Sound*, 7(2):145–156.

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

Drewes, H. (2010). Only one Fitts' law formula please! In *CHI '10 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '10, pages 2813–2822. ACM.

Duhigg, C. (2012). *The power of habit: Why we do what we do in life and business*, volume 34. Random House.

Duignan, M. (2008). *Computer mediated music production: A study of abstraction and activity*. PhD thesis, Victoria University of Wellington.

Duignan, M., Noble, J., and Biddle, R. (2010). Abstraction and activity in computer-mediated music production. *Computer Music Journal*, 34(4):22–33.

Duncker, K. and Lees, L. S. (1945). On problem-solving. *Psychological Monographs*, 58(5):i.

Dunn, J. C. (1983). Spatial metrics of integral and separable dimensions. *Journal of Experimental Psychology: Human Perception and Performance*, 9(2):242.

Edmonds, E. A., Weakley, A., Candy, L., Fell, M., Knott, R., and Pauletto, S. (2005). The studio as laboratory: combining creative practice and digital technology research. *International Journal of Human-Computer Studies*, 63(4):452–481.

Ellamil, M., Dobson, C., Beeman, M., and Christoff, K. (2012). Evaluative and generative modes of thought during the creative process. *Neuroimage*, 59(2):1783–1794.

Ellson, D. (1947). The independence of tracking in two and three dimensions with the b-29 pedestal sight. Technical report, TSEAA-694-2G, Aero Medical Laboratory.

Elster, J. (2000). *Ulysses Unbound: Studies in Rationality, Precommitment, and Constraints*. Cambridge University Press.

English, W. K., Engelbart, D. C., and Berman, M. L. (1967). Display-selection techniques for text manipulation. *Human Factors in Electronics, IEEE Transactions on*, (1):5–15.

Epworth, R. (2013). *Bottleneck: Our Human Interface with Reality*. Goforich Publications.

Ericsson, K. A. (2006). The influence of experience and deliberate practice on the development of superior expert performance. *The Cambridge handbook of expertise and expert performance*, pages 683–703.

Evans, J. S. B. (2003). In two minds: dual-process accounts of reasoning. *Trends in Cognitive Sciences*, 7(10):454–459.

Fallman, D. and Stolterman, E. (2010). Establishing criteria of rigour and relevance in interaction design research. *Digital Creativity*, 21(4):265–272.

Fauconnier, G. (2001). Conceptual blending and analogy. *The Analogical Mind: Perspectives from Cognitive Science*, pages 255–286.

Fels, S., Gadd, A., and Mulder, A. (2002). Mapping transparency through metaphor: towards more expressive musical instruments. *Organised Sound*, 7(02):109–126.

Fels, S. S. and Hinton, G. E. (1993). Glove-talk: A neural network interface between a data-glove and a speech synthesizer. *Neural Networks, IEEE Transactions on*, 4(1):2–8.

Fiebrink, R., Cook, P. R., and Trueman, D. (2009). *Play-along mapping of musical controllers.* Proceedings of the International Computer Music Conference, Montreal, Quebec.

Fiebrink, R., Trueman, D., Britt, C., Nagai, M., Kaczmarek, K., Early, M., Daniel, M., Hege, A., and Cook, P. (2010). *Toward understanding human-computer interaction in composing the instrument.* Citeseer.

Fitts, P. M. (1954). The information capacity of the human motor system in controlling the amplitude of movement. *Journal of Experimental Psychology*, 47(6):381.

Fitzmaurice, G. W. and Buxton, W. (1997). An empirical evaluation of graspable user interfaces: towards specialized, space-multiplexed input. In *Proceedings of* *the ACM SIGCHI Conference on Human Factors in Computing Cystems*, pages 43–50. ACM.

Fodor, J. A. (1983). *The modularity of mind: An essay on faculty psychology.* MIT press.

Forsythe, C., Liao, H., Trumbo, M. C. S., and Cardona-Rivera, R. E. (2014). *Cognitive Neuroscience of Human Systems: Work and Everyday Life*. CRC Press.

Françoise, J. (2013). Gesture–sound mapping by demonstration in interactive music systems. In *Proceedings of the 21st ACM international conference on Multimedia*, pages 1051–1054. ACM.

Freeman III, W. J. (1998). A neurobiological role of music in social bonding. pages 405–415. MIT Press.

Friston, K. (2010). The free-energy principle: a unified brain theory? *Nature Reviews Neuroscience*, 11(2):127–138.

Friston, K. (2011). What is optimal about motor control? *Neuron*, 72(3):488 – 498.

Fuster, J. M. (2001). The prefrontal cortex—an update: time is of the essence. *Neuron*, 30(2):319–333.

Gabora, L. (2002). Cognitive mechanisms underlying the creative process. In *Proceedings of the 4th Conference on Creativity & Cognition*, pages 126–133. ACM.

Gabora, L. (2005). Creative thought as a non Darwinian evolutionary process. *The Journal of Creative Behavior*, 39(4):262–283.

Gabrielsson, A. and Juslin, P. N. (1996). Emotional expression in music performance: Between the performer's intention and the listener's experience. *Psychology of music*, 24(1):68–91.

Galanter, P. (2012). Computational aesthetic evaluation: past and future. In *Computers and Creativity*, pages 255–293. Springer.

Gärdenfors, P. (1996). Mental representation, conceptual spaces and metaphors. *Synthese*, 106(1):21–47.

Gärdenfors, P. (2004). *Conceptual Spaces: The Geometry of Thought*. MIT press.

Garner, W. R. (1970). The stimulus in information processing. *American Psychologist*, 25(4):350.

Gelineck, S. and Serafin, S. (2012). Longitudinal evaluation of the integration of digital musical instruments into existing compositional work processes. *Journal of New Music Research*, 41(3):259–276.

Getzels, J. W. (1975). Problem-finding and the inventiveness of solutions. *The Journal of Creative Behavior*, 9(1):12–18.

Ghani, J. A. and Deshpande, S. P. (1994). Task characteristics and the experience of optimal flow in human—computer interaction. *The Journal of Psychology*, 128(4):381–391.

Gibson, J. J. (1977). The theory of affordances. *Hilldale, USA*.

Gibson, J. J. (2014). *The Ecological Approach to Visual Perception: Classic Edition*. Psychology Press.

Goddyn, L., Gvozdjak, P., et al. (2003). Binary gray codes with long bit runs. *The Electronic Journal of Combinatorics*, 10(27):1.

Goldberg, K., Faridani, S., and Alterovitz, R. (2013). A new derivation and dataset for Fitts' law of human motion. *EECS Dept., Univ. California, Berkeley, CA, USA, Tech. Rep. UCB/EECS-2013-171*.

Goldman, A. I. (2012). A moderate approach to embodied cognitive science. *Review of Philosophy and Psychology*, 3(1):71–88.

Gotsman, C. and Lindenbaum, M. (1995). On the metric properties of discrete space-filling curves. *IEEE transactions on image processing: a publication of the IEEE Signal Processing Society*, 5(5):794–797.

Gotsman, C. and Lindenbaum, M. (1996). On the metric properties of discrete space-filling curves. *IEEE Transactions on Image Processing*, 5(5):794–797.

Goudeseune, C. (2002). Interpolated mappings for musical instruments. *Organised Sound*, 7(2):85–96.

Gouzouasis, P. (2005). Fluency in general music and arts technologies: Is the future of music a garage band mentality?. *Action, Criticism, and Theory for Music Education*, 4(2):n2.

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

Gray, F. (1953). Pulse code communication. US Patent 2,632,058, March 17.

Green, T. R. (1989). Cognitive dimensions of notations. *People and computers V*, pages 443–460.

Grey, J. M. (1977). Multidimensional perceptual scaling of musical timbres. *Journal of the Acoustical Society of America*, 61(5):1270–1277.

Grossman, T. and Balakrishnan, R. (2004). Pointing at trivariate targets in 3d environments. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 447–454. ACM.

Guiard, Y. (1997). Fitts' law in the discrete vs. cyclical paradigm. *Human Movement Science*, 16(1):97–131.

Guiard, Y., Beaudouin-Lafon, M., and Mottet, D. (1999). Navigation as multiscale pointing: extending Fitts' model to very high precision tasks. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 450–457. ACM.

Guiard, Y. and Olafsdottir, H. B. (2011). On the measurement of movement difficulty in the standard approach to Fitts' law. *PloS One*, 6(10):e24389.

Guilford, J. P. (1967). *The Nature of Human Intelligence.* McGraw-Hill (New York).

Gurevich, M. and Cavan Fyans, A. (2011). Digital musical interactions: Performer–system relationships and their perception by spectators. *Organised Sound*, 16(02):166–175.

Gurevich, M., Marquez-Borbon, A., and Stapleton, P. (2012). Playing with constraints: Stylistic variation with a simple electronic instrument. *Computer Music Journal*, 36(1):23–41.

Gurevich, M. and Treviño, J. (2007). Expression and its discontents: toward an ecology of musical creation. In *Proceedings of the 7th International Conference on New Interfaces for Musical Expression*, pages 106–111. ACM.

Hallnäs, L. and Redström, J. (2001). Slow technology–designing for reflection. *Personal and ubiquitous computing*, 5(3):201–212.

Hameroff, S. (2010). The "conscious pilot"—dendritic synchrony moves through the brain to mediate consciousness. *Journal of Biological Physics*, 36(1):71–93.

Hamilton, C. H. and Rau-Chaplin, A. (2008). Compact Hilbert indices: Space-filling curves for domains with unequal side lengths. *Information Processing Letters*, 105(5):155–163.

Han, J. and Gold, N. (2014). Lessons learned in exploring the Leap Motion sensor for gesture-based instrument design. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, pages 371–374.

Hanna, R. (2012). Computer aided cognition and creativity: a three year monitoring exercise. *Physical Digitality.*

Hantrakul, L. and Kaczmarek, K. (2014). Implementations of the Leap Motion in sound synthesis, effects modulation and assistive performance tools. In *Proceedings of the Sound and Music Computing Conference.*

Harrison, S., Tatar, D., and Sengers, P. (2007). The three paradigms of HCI. In *Alt. Chi. Session at the SIGCHI Conference on Human Factors in Computing Systems San Jose, California, USA*, pages 1–18.

Hart, S. G. and Staveland, L. E. (1988). Development of nasa-tlx (task load index): Results of empirical and theoretical research. *Advances in psychology*, 52:139–183.

Hattori, M., Sloman, S., and Orita, R. (2013). Effects of subliminal hints on insight problem solving. *Psychonomic Bulletin & Review*, 20(4):790–797.

Hattwick, I. and Wanderley, M. M. (2012). A dimension space for evaluating collaborative musical performance systems. In *Proceedings of the 2012 Conference on New Interfaces for Musical Expression.*

Hebb, D. O. (1955). Drives and the CNS (conceptual nervous system). *Psychological Review*, 62(4):243.

Heise, S., Hlatky, M., and Loviscach, J. (2008). Soundtorch: Quick browsing in large audio collections. In *Audio Engineering Society Convention 125*. Audio Engineering Society.

Hennessey, B. A. and Amabile, T. M. (1999). Consensual assessment. *Encyclopedia of Creativity*, 1:347–359.

Hick, W. E. (1952). On the rate of gain of information. *Quarterly Journal of Experimental Psychology*, 4(1):11–26.

Hinckley, K., Jacob, R., and Ware, C. (2004). Input/output devices and interaction techniques.

Hoffmann, E. R. (2013). Which version/variation of Fitts' law? A critique of information-theory models. *Journal of Motor Behavior*, 45(3):205–215. PMID: 23581725.

Hubel, D. H. and Wiesel, T. N. (1968). Receptive fields and functional architecture of monkey striate cortex. *The Journal of Physiology*, 195(1):215–243.

Hunt, A. and Kirk, R. (1999). Radical user interfaces for real-time control. In *EUROMICRO Conference, 1999. Proceedings. 25th*, volume 2, pages 6–12. IEEE.

Hunt, A. and Kirk, R. (2000). Mapping strategies for musical performance. *Trends in Gestural Control of Music*, 21.

Hunt, A., Wanderley, M., and Kirk, R. (2000). Towards a model for instrumental mapping in expert musical interaction. In *Proceedings of the 2000 International Computer Music Conference*, pages 209–212.

Hunt, A. and Wanderley, M. M. (2002). Mapping performer parameters to synthesis engines. *Organised Sound*, 7:97–108.

ISO (2002). Ergonomic requirements for office work with visual display terminals (vdts) part requirements for non-keyboard input devices (iso 9241-9).

Jacob, R. J. and Sibert, L. E. (1992). The perceptual structure of multidimensional input device selection. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 211–218. ACM.

Jacob, R. J., Sibert, L. E., McFarlane, D. C., and Mullen Jr, M. P. (1994). Integrality and separability of input devices. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 1(1):3–26.

Jaffer, A. (2005). Color-space dimension reduction. [Online; accessed 12-February-2014].

Jamieson, S. (2004). Likert scales: how to (ab) use them. *Medical education*, 38(12):1217–1218.

Jennings, K. E. (2008). Adjusting the novelty thermostat: Courting creative success through judicious randomness. In *AAAI Spring Symposium: Creative Intelligent Systems*, pages 26–33.

Jennings, K. E., Simonton, D. K., and Palmer, S. E. (2011). Understanding exploratory creativity in a visual domain. In *Proceedings of the 8th ACM conference on Creativity and Cognition*, pages 223–232. ACM.

John, B. E. and Kieras, D. E. (1996). The GOMS family of user interface analysis techniques: Comparison and contrast. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 3(4):320–351.

Johnson, C. G. (1999). Exploring the sound-space of synthesis algorithms using interactive genetic algorithms. In *Proceedings of the AISB Workshop on Artificial Intelligence and Musical Creativity, Edinburgh*.

Johnson-Laird, P. N. (2002). How jazz musicians improvise. *Music Perception*, 19(3):415–442.

Johnston, A. (2011). Beyond evaluation: Linking practice and theory in new musical interface design. *Interpretation*, 14:105.

Jones, D., Brown, A. R., and d'Inverno, M. (2012). The extended composer. In *Computers and Creativity*, pages 175–203. Springer.

Jonsen, I. D., Flemming, J. M., and Myers, R. A. (2005). Robust state-space modeling of animal movement data. *Ecology*, 86(11):2874–2880.

Jordà, S. (2004). Instruments and players: Some thoughts on digital lutherie. *Journal of New Music Research*, 33(3):321–341.

Jorda, S. (2005). *Digital Lutherie: Crafting musical computers for new musics' performance and improvisation*. PhD thesis, Departament de Tecnologia, Universitat Pompeu Fabra, Spain.

Jordà, S., Geiger, G., Alonso, M., and Kaltenbrunner, M. (2007). The reacTable: exploring the synergy between live music performance and tabletop tangible interfaces. In *Proceedings of the 1st international conference on Tangible and embedded interaction*, pages 139–146. ACM.

Jordanous, A. (2012). A standardised procedure for evaluating creative systems: Computational creativity evaluation based on what it is to be creative. *Cognitive Computation*, 4(3):246–279.

Joyce, C. (2009). *The blank page: Effects of constraint on creativity*. PhD thesis, University of California, Berkeley.

Juslin, P. N. and Sloboda, J. A. (2010). *Handbook of music and emotion: Theory, research, applications*. Oxford University Press.

Kaas, J. H. (1997). Topographic maps are fundamental to sensory processing. *Brain Research Bulletin*, 44(2):107 – 112.

Kahneman, D. (2011). *Thinking, fast and slow*. Farrar, Straus and Giroux.

Keith, S. (2010). Bridging the gap: Thoughts on computer music and contemporary (popular) electronic music. In *2010 Australasian Computer Music Conference Proceedings*.

Kiebel, S. J., Daunizeau, J., and Friston, K. J. (2008). A hierarchy of time-scales and the brain. *PLoS Computational Biology*, 4(11):e1000209.

Kiefer, C. (2010). A malleable interface for sonic exploration. In *Proceedings of the International Conference on New Interfaces for Musical Expression (NIME).*, pages 291–296.

Kiefer, C. (2014). Musical instrument mapping design with echo state networks. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, pages 293–298.

Kieras, D. E. and Meyer, D. E. (1997). An overview of the epic architecture for cognition and performance with application to human-computer interaction. *Human-computer interaction*, 12(4):391–438.

Kirsh, D. (2013). Embodied cognition and the magical future of interaction design. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 20(1):3.

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

Klyubin, A. S., Polani, D., and Nehaniv, C. L. (2008). Keep your options open: an information-based driving principle for sensorimotor systems. *PloS One*, 3(12):e4018.

Knill, D. C. and Pouget, A. (2004). The Bayesian brain: the role of uncertainty in neural coding and computation. *Trends in Neurosciences*, 27(12):712–719.

Koelsch, S., Kasper, E., Sammler, D., Schulze, K., Gunter, T., and Friederici, A. D. (2004). Music, language and meaning: brain signatures of semantic processing. *Nature neuroscience*, 7(3):302–307.

Kounios, J. and Beeman, M. (2009). The aha! moment: The cognitive neuroscience of insight. *Current Directions in Psychological Science*, 18(4):210–216.

Kurzweil, R. (2005). *The Singularity is Near: When Humans Transcend Biology*. Penguin.

Langer, E. J. and Imber, L. G. (1979). When practice makes imperfect: debilitating effects of overlearning. *Journal of Personality and Social Psychology*, 37(11):2014.

Langley, P. and Jones, R. (1988). A computational model of scientific insight. *The Nature of Creativity: Contemporary Psychological Perspectives*, 177:201.

Laughlin, S. B., van Steveninck, R. R. d. R., and Anderson, J. C. (1998). The metabolic cost of neural information. *Nature neuroscience*, 1(1):36–41.

Le Bihan, D., Mangin, J.-F., Poupon, C., Clark, C. A., Pappata, S., Molko, N., and Chabriat, H. (2001). Diffusion Tensor Imaging: concepts and applications. *Journal of Magnetic Resonance Imaging*, 13(4):534–546.

Lehman, J. and Stanley, K. O. (2011). Abandoning objectives: Evolution through the search for novelty alone. *Evolutionary Computation*, 19(2):189–223.

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

Levin, T. Y. (2003). "Tones from out of nowhere": Rudolph pfenninger and the archaeology of synthetic sound. *Grey Room*, 12:32–79.

Levitin, D. J., McAdams, S., and Adams, R. L. (2002). Control parameters for musical instruments: a foundation for new mappings of gesture to sound. *Organised Sound*, 7(2):171–189.

Limb, C. J. and Braun, A. R. (2008). Neural substrates of spontaneous musical performance: An fMRI study of jazz improvisation. *PLoS One*, 3(2):e1679.

Liu, S., Chow, H. M., Xu, Y., Erkkinen, M. G., Swett, K. E., Eagle, M. W., Rizik-Baer, D. A., and Braun, A. R. (2012). Neural correlates of lyrical improvisation: an fMRI study of freestyle rap. *Scientific Reports*, 2.

Lotze, M., Montoya, P., Erb, M., Hülsmann, E., Flor, H., Klose, U., Birbaumer, N., and Grodd, W. (1999). Activation of cortical and cerebellar motor areas during executed and imagined hand movements: an fmri study. *Journal of cognitive neuroscience*, 11(5):491–501.

Loviscach, J. (2008). Programming a Music Synthesizer through Data Mining. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, pages 221–224.

Lubart, T. (2005). How can computers be partners in the creative process: classification and commentary on the special issue. *International Journal of Human-Computer Studies*, 63(4):365–369.

MacDonald, R., Byrne, C., and Carlton, L. (2006). Creativity and flow in musical composition: An empirical investigation. *Psychology of Music*, 34(3):292–306.

MacKay, D. G. (1982). The problems of flexibility, fluency, and speed–accuracy trade-off in skilled behavior. *Psychological Review*, 89(5):483.

MacKenzie, I. S. (1992a). Fitts' law as a research and design tool in human-computer interaction. *Human-computer interaction*, 7(1):91–139.

MacKenzie, I. S. (1992b). Movement time prediction in human-computer interfaces. In *Proceedings of Graphics Interface*, volume 92, page 1.

MacKenzie, I. S. (2015). Fitts throughput and the remarkable case of touch-based target selection. In *Human-Computer Interaction: Interaction Technologies*, pages 238–249. Springer.

MacKenzie, I. S. and Buxton, W. (1992). Extending Fitts' law to two-dimensional tasks. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 219–226. ACM.

Macleod, M., Bowden, R., Bevan, N., and Curson, I. (1997). The music performance measurement method. *Behaviour & Information Technology*, 16(4-5):279–293.

Maes, P.-J., Wanderley, M. M., and Palmer, C. (2015). The role of working memory in the temporal control of discrete and continuous movements. *Experimental Brain Research*, 233(1):263–273.

Magnusson, T. (2010). Designing constraints: Composing and performing with digital musical systems. *Computer Music Journal*, 34(4):62–73.

Malloch, J., Birnbaum, D., Sinyor, E., and Wanderley, M. M. (2006). Towards a new conceptual framework for digital musical instruments. In *Proceedings of the 9th international conference on digital audio effects*, pages 49–52.

Mandanici, M. and Canazza, S. (2014). The "hand composer": gesture-driven music composition machines. In *Proceedings of 13th International Conference on Intelligent Autonomous Systems*.

Marois, R. and Ivanoff, J. (2005). Capacity limits of information processing in the brain. *Trends in Cognitive Sciences*, 9(6):296–305.

Marshall, M. T. (2009). *Physical Interface Design for Digital Musical Instruments*. PhD thesis, McGill University, Montreal, QC, Canada.

Maslow, A. H. (1968). Music education and peak experience. *Music Educators Journal*, 54(6):72–171.

Masters, R. S. (1992). Knowledge, knerves and know-how: The role of explicit versus implicit knowledge in the breakdown of a complex motor skill under pressure. *British journal of psychology*, 83(3):343–358.

Mathews, M. V. (1991). The radio baton and conductor program, or: Pitch, the most important and least expressive part of music. *Computer Music Journal*, 15(4):37–46.

Mathy, F. and Feldman, J. (2012). What's magic about magic numbers? chunking and data compression in short-term memory. *Cognition*, 122(3):346–362.

McPherson, A. and Kim, Y. (2012). The problem of the second performer: Building a community around an augmented piano. *Computer Music Journal*, 36(4):10–27.

McPherson, A. and Zappi, V. (2015). Exposing the scaffolding of digital instruments with hardware-software feedback loops. In Berdahl, E. and Allison, J., editors, *Proceedings of the International Conference on New Interfaces for Musical Expression*, pages 162–167, Baton Rouge, Louisiana, USA. Louisiana State University.

Mednick, S. (1962). The associative basis of the creative process. *Psychological Review*, 69(3):220.

Melara, R. D. and Marks, L. E. (1990). Interaction among auditory dimensions: Timbre, pitch, and loudness. *Perception & Psychophysics*, 48(2):169–178.

Meunier, D., Lambiotte, R., and Bullmore, E. T. (2010). Modular and hierarchically modular organization of brain networks. *Frontiers in Neuroscience*, 4.

Meyer, D. E., Abrams, R. A., Kornblum, S., Wright, C. E., and Keith Smith, J. (1988). Optimality in human motor performance: ideal control of rapid aimed movements. *Psychological Review*, 95(3):340.

Meyer, D. E. and Schvaneveldt, R. W. (1971). Facilitation in recognizing pairs of words: Evidence of a dependence between retrieval operations. *Journal of Experimental Psychology*, 90(2):227.

Meyer, D. E., Smith, J., and Wright, C. E. (1982). Models for the speed and accuracy of aimed movements. *Psychological Review*, 89(5):449.

Miall, R. C. and Wolpert, D. M. (1996). Forward models for physiological motor control. *Neural networks*, 9(8):1265–1279.

Miller, G. A. (1956). The magical number seven, plus or minus two: some limits on our capacity for processing information. *Psychological Review*, 63(2):81.

Miranda, E. R. and Wanderley, M. M. (2006). *New digital musical instruments: control and interaction beyond the keyboard*, volume 21. AR Editions, Inc.

Mooney, J. (2011). Frameworks and affordances: Understanding the tools of music-making. *Journal of Music, Technology & Education*, 3(2-3):141–154.

Mooneyham, B. W. and Schooler, J. W. (2013). The costs and benefits of mind-wandering: a review. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale*, 67(1):11.

Moore, M. M. (2003). Real-world applications for brain-computer interface technology. *Neural Systems and Rehabilitation Engineering, IEEE Transactions on*, 11(2):162–165.

Morris, W. (2002). *News From Nowhere*. Broadview Press.

Mulder, A. (1994). Virtual musical instruments: Accessing the sound synthesis universe as a performer. In *Proceedings of the First Brazilian Symposium on Computer Music*, pages 243–250.

Müller-Wienbergen, F., Müller, O., Seidel, S., and Becker, J. (2011). Leaving the beaten tracks in creative work–a design theory for systems that support convergent and divergent thinking. *Journal of the Association for Information Systems*, 12(11):714–740.

Murata, A. and Iwase, H. (2001). Extending Fitts' law to a three-dimensional pointing task. *Human Movement Science*, 20(6):791–805.

Murray-Browne, T., Dom Aversano, S., Garcia, W. H., Lopez, D., Tadeo Sendon, P., Tigas, K. Z., and Chapman, D. (2014). The cave of sounds: An interactive installation exploring how we create music together. In *Proceedings of the International Conference on New Interfaces for Musical Expression (NIME).*

Murray-Browne, T., Mainstone, D., Bryan-Kinns, N., and Plumbley, M. D. (2011). The medium is the message: Composing instruments and performing mappings. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, pages 56–59.

Mycroft, J., Reiss, J. D., and Stockman, T. (2013). The influence of graphical user interface design on critical listening skills. In *Proceedings of the Sound and Music Computing (SMC) Conference, Stockholm*.

Myers, B. A. (1998). A brief history of human-computer interaction technology. *interactions*, 5(2):44–54.

Nash, C. (2012). *Supporting Virtuosity and Flow in Computer Music*. PhD thesis, University of Cambridge.

Nash, C. and Blackwell, A. (2011). Tracking virtuosity and flow in computer music. In *Proceedings of International Computer Music Conference*.

Nash, C. and Blackwell, A. (2012). Liveness and flow in notation use. In Essl, G., Gillespie, B., Gurevich, M., and O'Modhrain, S., editors, *Proceedings of the International Conference on New Interfaces for Musical Expression (NIME)*, Ann Arbor, Michigan. University of Michigan.

Newton, D. and Marshall, M. T. (2011). Examining how musicians create augmented musical instruments. *Proceedings of the International Conference on New Interfaces for Musical Expression (NIME)*.

Newton-Dunn, H., Nakano, H., and Gibson, J. (2003). Block jam: a tangible interface for interactive music. In *Proceedings of the 2003 conference on New interfaces for musical expression*, pages 170–177. National University of Singapore.

Nguyen, A., Yosinski, J., and Clune, J. (2015). Innovation engines: Automated creativity and improved stochastic optimization via deep learning. In *Proceedings of the Genetic and Evolutionary Computation Conference*.

Nilson, C. (2007). Live coding practice. In *Proceedings of the 7th International Conference on New interfaces for Musical Expression*, pages 112–117. ACM.

Norman, D. A. (1986). Cognitive engineering. *User centered system design: New perspectives on human-computer interaction*, pages 31–61.

Norman, D. A. (2002). *The Design of Everyday Things*. Basic books.

Norman, D. A. (2004). *Emotional design: Why we love (or hate) everyday things*. Basic books.

Nørretranders, T. (1991). *The user illusion: Cutting consciousness down to size.* Viking.

North, A. C. and Hargreaves, D. J. (1995). Subjective complexity, familiarity, and liking for popular music. *Psychomusicology: Music, Mind & Brain*, 14(1):77–93.

Norwich, K. H. (1993). *Information, sensation, and perception*. Academic Press San Diego.

Nunamaker Jr, J. F., Applegate, L. M., and Konsynski, B. R. (1987). Facilitating group creativity: Experience with a group decision support system. *Journal of Management Information Systems*, pages 5–19.

O'Donovan, P., Agarwala, A., and Hertzmann, A. (2015). Designscape: Design with interactive layout suggestions. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, pages 1221–1224. ACM.

Oleynick, V. C., Thrash, T. M., LeFew, M. C., Moldovan, E. G., and Kieffaber, P. D. (2014). The scientific study of inspiration in the creative process: challenges and opportunities. *Frontiers in Human Neuroscience*, 8.

O'Neill, S. (1999). Flow theory and the development of musical performance skills. *Bulletin of the Council for Research in Music Education*, pages 129–134.

Osaka, N. (2004). Toward Construction of a Timbre Theory for Music Composition. In *Proceedings of ICMC*.

Pachet, F. (2012). Musical virtuosity and creativity. In *Computers and creativity*, pages 115–146. Springer.

Pachet, F. and Aucouturier, J.-J. (2004). Improving timbre similarity: How high is the sky. *Journal of negative results in speech and audio sciences*, 1(1):1–13.

Pachet, F., Roy, P., and Ghedini, F. (2013). Creativity through style manipulation: the flow machines project. In *Proceedings of the Marconi Institute for Creativity Conference (Bologna, Italy*, volume 80.

Paivio, A. (1969). Mental imagery in associative learning and memory. *Psychological review*, 76(3):241.

Palmer, C. (1996). Anatomy of a performance: Sources of musical expression. *Music Perception*, pages 433–453.

Payne, L. and Sekuler, R. (2014). The importance of ignoring alpha oscillations protect selectivity. *Current Directions in Psychological Science*, 23(3):171–177.

Pearce, J. M. and Howard, S. (2004). Designing for flow in a complex activity. In *Computer Human Interaction*, pages 349–358. Springer.

Pearce, M. and Wiggins, G. A. (2002). Aspects of a cognitive theory of creativity in musical composition. In *Proceedings of the ECAI02 Workshop on Creative Systems*.

Pearce, M. T. (2005). *The construction and evaluation of statistical models of melodic structure in music perception and composition*. PhD thesis, City University London.

Pearce, M. T. and Wiggins, G. A. (2012). Auditory expectation: The information dynamics of music perception and cognition. *Topics in Cognitive Science*, 4(4):625–652.

Pease, A., Colton, S., Ramezani, R., Charnley, J., and Reed, K. (2013). A discussion on serendipity in creative systems. In *Proceedings of the Fourth International Conference on Computational Creativity*, page 64.

Penhune, V. B. and Steele, C. J. (2012). Parallel contributions of cerebellar, striatal and m1 mechanisms to motor sequence learning. *Behavioural brain research*, 226(2):579–591.

Pennycook, B. W. (1985). Computer-music interfaces: a survey. *ACM Computing Surveys (CSUR)*, 17(2):267–289.

Perani, D., Cappa, S., Bettinardi, V., Bressi, S., Gorno-Tempini, M., Matarrese, M., and Fazio, F. (1995). Different neural systems for the recognition of animals and man-made tools. *Neuroreport*, 6(12):1637–1641.

Pinch, T. (2003). Giving birth to new users: How the minimoog was sold to rock and roll. *How users matter: the co-construction of users and technologies*, pages 247–270.

Pinch, T. J., Trocco, F., and Pinch, T. (2009). *Analog days: The invention and impact of the Moog synthesizer*. Harvard University Press.

Pinker, S. (1999). How the mind works. *Annals of the New York Academy of Sciences*, 882(1):119–127.

Pino, A., Tzemis, E., Ioannou, N., and Kouroupetroglou, G. (2013). Using Kinect for 2d and 3d pointing tasks: Performance evaluation. In *Human-Computer Interaction. Interaction Modalities and Techniques*, pages 358–367. Springer.

Pitt, M. A. (1994). Perception of pitch and timbre by musically trained and untrained listeners. *Journal of experimental psychology: human perception and performance*, 20(5):976.

Poepel, C. (2005). On interface expressivity: a player-based study. In *Proceedings of the 2005 conference on New interfaces for musical expression*, pages 228–231. National University of Singapore.

Polya, G. (2004). *How to solve it: A new aspect of mathematical method*. Princeton University Press.

Poupyrev, I., Lyons, M. J., and Fels, S. (2001). New interfaces for musical expression. In *CHI'01 Extended Abstracts on Human Factors in Computing Systems*, pages 491–492. ACM.

Pressing, J. (1990). Cybernetic issues in interactive performance systems. *Computer music journal*, 14(1):12–25.

Puckette, M. (1996). Pure Data: another integrated computer music environment. *Proceedings of the Second Intercollege Computer Music Concerts*, pages 37–41.

Raglin, J. S. and Turner, P. E. (1993). Anxiety and performance in track and field athletes: A comparison of the inverted-u hypothesis with zone of optimal function theory. *Personality and Individual Differences*, 14(1):163 – 171.

Rassmussen, J. (1986). Information processing and human-machine interaction. *An Aproach to Cognitive Engineering, Amsterdam, Nom Holland.*

Rautaray, S. S. and Agrawal, A. (2015). Vision based hand gesture recognition for human computer interaction: a survey. *Artificial Intelligence Review*, 43(1):1–54.

Repp, B. H. (1998). A microcosm of musical expression. i. quantitative analysis of pianists timing in the initial measures of chopins etude in e major. *The Journal of the Acoustical Society of America*, 104(2):1085–1100.

Resnick, M., Myers, B., Nakakoji, K., Shneiderman, B., Pausch, R., Selker, T., and Eisenberg, M. (2005). Design principles for tools to support creative thinking. In *In Proceedings of the NSF Workshop on Creativity Support Tools*, pages 25–36.

Ritter, M. and Aska, A. (2014). Leap Motion as expressive gestural interface. *Proceedings of the International Computer Music Conference (ICMC).*

Rizzolatti, G., Fadiga, L., Gallese, V., and Fogassi, L. (1996). Premotor cortex and the recognition of motor actions. *Cognitive brain research*, 3(2):131–141.

Rizzolatti, G. and Luppino, G. (2001). The cortical motor system. *Neuron*, 31(6):889 – 901.

Robbins, N. B. and Heiberger, R. M. (2011). Plotting Likert and other rating scales. *JSM Proceedings, Section on Survey Research Methods, Alexandria, VA*, pages 1058–1066.

Root-Bernstein, R. and Root-Bernstein, M. (2004). Artistic scientists and scientific artists: The link between polymathy and creativity. In *Creativity: From potential to realization*. American Psychological Association.

Sadler-Smith, E. (2012). *Inside intuition*. Routledge.

Salvucci, D. D. and Taatgen, N. A. (2010). *The multitasking mind*. Oxford University Press.

Sambrooks, L. and Wilkinson, B. (2013). Comparison of gestural, touch, and mouse interaction with Fitts' law. In *Proceedings of the 25th Australian Computer-*

Human Interaction Conference: Augmentation, Application, Innovation, Collaboration, pages 119–122. ACM.

Sawyer, K. (2011). The cognitive neuroscience of creativity: a critical review. *Creativity Research Journal*, 23(2):137–154.

Schmidhuber, J. (2009). Driven by compression progress: A simple principle explains essential aspects of subjective beauty, novelty, surprise, interestingness, attention, curiosity, creativity, art, science, music, jokes. In *Anticipatory Behavior in Adaptive Learning Systems*, pages 48–76. Springer.

Schmidhuber, J. (2010). Formal theory of creativity, fun, and intrinsic motivation (1990–2010). *IEEE Transactions on Autonomous Mental Development*, 2(3):230–247.

Schmidhuber, J. (2012). A formal theory of creativity to model the creation of art. In *Computers and Creativity*, pages 323–337. Springer.

Schmidt, R. A., Zelaznik, H., Hawkins, B., Frank, J. S., and Quinn Jr, J. T. (1979). Motor-output variability: a theory for the accuracy of rapid motor acts. *Psychological Review*, 86(5):415.

Schooler, J. W., Ohlsson, S., and Brooks, K. (1993). Thoughts beyond words: When language overshadows insight. *Journal of Experimental Psychology: General*, 122(2):166.

Schooler, J. W., Smallwood, J., Christoff, K., Handy, T. C., Reichle, E. D., and Sayette, M. A. (2011). Meta-awareness, perceptual decoupling and the wandering mind. *Trends in Cognitive Sciences*, 15(7):319–326.

Schwarz, D. (2012). The sound space as musical instrument: Playing corpus-based concatenative synthesis. *New Interfaces for Musical Expression (NIME)*, pages 250–253.

Seixas, M., Cardoso, J., and Dias, M. T. G. (2015). The Leap Motion movement for 2d pointing tasks: Characterisation and comparison to other devices.

Selker, T. (2005). Fostering motivation and creativity for computer users. *International Journal of Human-Computer Studies*, 63(4):410 – 421.

Sergent, C. and Dehaene, S. (2004). Neural processes underlying conscious perception: experimental findings and a global neuronal workspace framework. *Journal of Physiology*, 98(4):374–384.

Shannon, C. E. and Weaver, W. (1949). The mathematical theory of information.

Sheridan, T. B. (2004). Musings on music making and listening: supervisory control and virtual reality. *Proceedings of the IEEE*, 92(4):601–605.

Shneiderman, B. (1982). The future of interactive systems and the emergence of direct manipulation. *Behaviour & Information Technology*, 1(3):237–256.

Shneiderman, B. (2000). Creating creativity: user interfaces for supporting innovation. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 7(1):114–138.

Shneiderman, B. (2007). Creativity support tools: Accelerating discovery and innovation. *Communications of the ACM*, 50(12):20–32.

Shneiderman, B. and Plaisant, C. (2004). *Designing the User Interface: Strategies for Effective Human-Computer Interaction (4th Edition)*. Pearson Addison Wesley.

Simon, H. A. (1974). The structure of ill structured problems. *Artificial Intelligence*, 4(3):181–201.

Simons, D. J. and Chabris, C. F. (1999). Gorillas in our midst: Sustained inattentional blindness for dynamic events. *Perception*, 28(9):1059–1074.

Simonton, D. K. (1996). Creative expertise: A life-span developmental perspective. *The road to excellence: The acquisition of expert performance in the arts and sciences, sports, and games*, pages 227–253.

Simonton, D. K. (1999). *Origins of genius: Darwinian perspectives on creativity.* Oxford University Press.

Simonton, D. K. (2012). Creativity, problem solving, and solution set sightedness: Radically reformulating BVSR. *The Journal of Creative Behavior*, 46(1):48–65.

Smith, J. O. (1991). Viewpoints on the history of digital synthesis. In *Proceedings of the International Computer Music Conference (ICMC-91, Montreal)*, pages 1–1.

Smith, S. (2000). Compositional strategies of the hip-hop turntablist. *Organised Sound*, 5(02):75–79.

Smith, S., Ward, T., and Schumacher, J. (1993). Constraining effects of examples in a creative generation task. *Memory & Cognition*, 21(6):837–845.

Smith, S. M., Ward, T. B., and Finke, R. A. (1995). *The creative cognition approach.* MIT Press.

Snow, C. P. (2012). *The Two Cultures.* Cambridge University Press.

Soukoreff, R. W. and MacKenzie, I. S. (2004). Towards a standard for pointing device evaluation, perspectives on 27 years of Fitts' law research in HCI. *International Journal of Human-Computer Studies*, 61(6):751–789.

Soukoreff, R. W., Zhao, J., and Ren, X. (2011). The entropy of a rapid aimed movement: Fitts' index of difficulty versus Shannon's entropy. In *Human-Computer Interaction–INTERACT 2011*, pages 222–239. Springer.

Sowden, P. T., Pringle, A., and Gabora, L. (2015). The shifting sands of creative thinking: Connections to dual-process theory. *Thinking & Reasoning*, 21(1):40–60.

Sporns, O., Tononi, G., and Kötter, R. (2005). The human connectome: A structural description of the human brain. *PLoS Computational Biology*, 1(4):e42.

Stanley, K. O. and Lehman, J. (2015). *Why Greatness Cannot Be Planned*. Springer.

Stanovich, K. (2011). *Rationality and the reflective mind*. Oxford University Press.

Stanovich, K. E. (2009). Distinguishing the reflective, algorithmic, and autonomous minds: Is it time for a tri-process theory? *In Two Minds: Dual Processes and Beyond*, pages 55–88.

Stanovich, K. E. and West, R. F. (2000). Individual differences in reasoning: Implications for the rationality debate? *Behavioral and Brain Sciences*, 23(5):645–665.

Sternberg, R. J. (1999). *Handbook of Creativity*. Cambridge University Press.

Stokes, M., Thompson, R., Cusack, R., and Duncan, J. (2009). Top-down activation of shape-specific population codes in visual cortex during mental imagery. *The Journal of Neuroscience*, 29(5):1565–1572.

Stokes, P. (2006). *Creativity from Constraints: The Psychology of Breakthrough*. Springer.

Stowell, D. (2010). *Making music through real-time voice timbre analysis: machine learning and timbral control*. PhD thesis, School of Electronic Engineering and Computer Science, Queen Mary University of London.

Stowell, D., Robertson, A., Bryan-Kinns, N., and Plumbley, M. D. (2009). Evaluation of live human–computer music-making: Quantitative and qualitative approaches. *International Journal of Human-Computer Studies*, 67(11):960–975.

Stravinsky, I. (1970). *Poetics of music: In the form of six lessons*. Harvard University Press.

Suchman, L. A. (1987). *Plans and situated actions: The problem of human-machine communication*. Cambridge University Press.

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive science*, 12(2):257–285.

Takeuchi, H., Taki, Y., Hashizume, H., Sassa, Y., Nagase, T., Nouchi, R., and Kawashima, R. (2011). Failing to deactivate: the association between brain activity during a working memory task and creativity. *Neuroimage*, 55(2):681–687.

Taleb, N. N. (2005). The roots of unfairness: The black swan in arts and literature. *Literary Research/Recherche Litteraire*, 21(41-42):241–254.

Tano, S., Yamamoto, S., Dzulkhiflee, M., Ichino, J., and Hashiyama, T. (2012). Three design principles learned through developing a series of 3d sketch systems: “memory capacity”, “cognitive mode”, and “life-size and operability”. In *Proceedings of the IEEE International Conference on Systems, Man, and Cybernetics October 14-17, 2012, COEX, Seoul, Korea*.

Thor Magnusson, E. H. M. (2007). The acoustic, the digital and the body: A survey on musical instruments. In *NIME 2007 International Conference on New Interfaces for Musical Expression*, pages 94–99.

Thoroughman, K. A. and Shadmehr, R. (2000). Learning of action through adaptive combination of motor primitives. *Nature*, 407(6805):742–747.

Toiviainen, P., Luck, G., and Thompson, M. R. (2010). Embodied meter: Hierarchical eigenmodes in music-induced movement. *Music Perception: An Interdisciplinary Journal*, 28(1):59–70.

Tononi, G. (2012). Integrated information theory of consciousness: An updated account. *Archives Italiennes de Biologie*, 150(2-3):56–90.

Touchette, H. and Lloyd, S. (2000). Information-theoretic limits of control. *Physical Review Letters*, 84(6):1156.

Tubb, R. (2015). Controlling rhythm in the frequency domain. *Leonardo*, 48(3):282–283.

Tubb, R. and Dixon, S. (2014). A zoomable mapping of a musical parameter space using Hilbert curves. *Computer music journal*, 38(3):23–33.

Van Nort, D. and Wanderley, M. (2007). Control strategies for navigation of complex sonic spaces. In *Proceedings of the 7th International Conference on New Interfaces for Musical Expression*, pages 379–382. ACM.

Van Nort, D., Wanderley, M. M., and Depalle, P. (2004). On the choice of mappings based on geometric properties. In *Proceedings of the 2004 conference on New Interfaces for Musical Expression (NIME)*, pages 87–91. National University of Singapore.

Van Schaik, P. and Ling, J. (2012). A cognitive-experiential approach to modelling web navigation. *International Journal of Human-Computer Studies*, 70(9):630–651.

Van Wijk, J. J. and van Overveld, C. (2003). Preset based interaction with high dimensional parameter spaces. *Kluwer International Series in Engineering and Computer Science.*, pages 391–406.

Vartanian, O., Martindale, C., and Kwiatkowski, J. (2007). Creative potential, attention, and speed of information processing. *Personality and Individual Differences*, 43(6):1470 – 1480.

Vertegaal, R. and Bonis, E. (1994). ISEE: An intuitive sound editing environment. *Computer Music Journal*, 18:21–29.

Vertegaal, R. and Eaglestone, B. (1996). Comparison of input devices in an ISEE direct timbre manipulation task. *Interacting with Computers*, 8(1):13–30.

Vinge, V. (1993). The coming technological singularity: How to survive in the post-human era. *Vision-21: Interdisciplinary science and engineering in the era of cyberspace*, pages 11–22.

Wacongne, C., Labyt, E., van Wassenhove, V., Bekinschtein, T., Naccache, L., and Dehaene, S. (2011). Evidence for a hierarchy of predictions and prediction errors in human cortex. *Proceedings of the National Academy of Sciences*, 108(51):20754–20759.

Wagner, R. K. and Sternberg, R. J. (1985). Practical intelligence in real-world pursuits: The role of tacit knowledge. *Journal of Personality and Social Psychology*, 49(2):436.

Wallas, G. (1926). *The Art of Thought*. Solis Press.

Wamsley, E. J., Tucker, M., Payne, J. D., Benavides, J. A., and Stickgold, R. (2010). Dreaming of a learning task is associated with enhanced sleep-dependent memory consolidation. *Current Biology*, 20(9):850–855.

Wan, C. Y. and Huon, G. F. (2005). Performance degradation under pressure in music: An examination of attentional processes. *Psychology of Music*, 33(2):155–172.

Wanderley, M. M. (2001). Gestural control of music. In *International Workshop Human Supervision and Control in Engineering and Music*, pages 632–644.

Wanderley, M. M. and Depalle, P. (2004). Gestural control of sound synthesis. *Proceedings of the IEEE*, 92(4):632–644.

Wanderley, M. M. and Orio, N. (2002). Evaluation of input devices for musical expression: Borrowing tools from HCI. *Computer Music Journal*, 26(3):62–76.

Wanderley, M. M., Viollet, J.-P., Isart, F., and Rodet, X. (2000). On the choice of transducer technologies for specific musical functions. In *Proceedings of the 2000 International Computer Music Conference*, pages 244–247.

Ward, T. B., Smith, S. M., and Finke, R. A. (1999). *Handbook of Creativity*, chapter Creative cognition. In Sternberg [1999].

Webster, J., Trevino, L. K., and Ryan, L. (1994). The dimensionality and correlates of flow in human-computer interactions. *Computers in Human Behavior*, 9(4):411–426.

Weichert, F., Bachmann, D., Rudak, B., and Fisseler, D. (2013). Analysis of the accuracy and robustness of the Leap Motion controller. *Sensors*, 13(5):6380–6393.

Wessel, D. and Wright, M. (2002). Problems and prospects for intimate musical control of computers. *Computer Music Journal*, 26(3):11–22.

Wessel, D. L. (1979). Timbre space as a musical control structure. *Computer Music Journal*, 3(2):pp. 45–52.

Wickelgren, W. A. (1979). Chunking and consolidation: A theoretical synthesis of semantic networks, configuring in conditioning, SR versus cognitive learning, normal forgetting, the amnesic syndrome, and the hippocampal arousal system. *Psychological Review*, 86(1):44.

Wiggins, G. A. (2006). A preliminary framework for description, analysis and comparison of creative systems. *Knowledge-Based Systems*, 19(7):449–458.

Wiggins, G. A. (2012). The mind's chorus: Creativity before consciousness. *Cognitive Computation*, 4(3):306–319.

Wiggins, G. A. and Bhattacharya, J. (2014). Mind the gap: An attempt to bridge computational and neuroscientific approaches to study creativity. *Frontiers in Human Neuroscience*, 8.

Wiggins, G. A. and Forth, J. (2015). IDyOT: A computational theory of creativity as everyday reasoning from learned information. In *Computational Creativity Research: Towards Creative Machines*, pages 127–148. Springer.

Wigner, E. P. (1960). The unreasonable effectiveness of mathematics in the natural sciences. richard courant lecture in mathematical sciences delivered at new york university, may 11, 1959. *Communications on Pure and Applied Mathematics*, 13(1):1–14.

Wilson, M. (2002). Six views of embodied cognition. *Psychonomic Bulletin & Review*, 9(4):625–636.

Wilson, R. C., Guilford, J., and Christensen, P. R. (1953). The measurement of individual differences in originality. *Psychological Bulletin*, 50(5):362.

Winkler, T. (1995). Making motion musical: Gesture mapping strategies for interactive computer music. In *ICMC Proceedings*, pages 261–264.

Winograd, T. and Flores, F. (1986). *Understanding computers and cognition: A new foundation for design*. Intellect Books.

Wolpert, D. M., Diedrichsen, J., and Flanagan, J. R. (2011). Principles of sensorimotor learning. *Nature Reviews Neuroscience*, 12(12):739–751.

Wolpert, D. M. and Flanagan, J. (2001). Motor prediction. *Current Biology*, 11(18):R729 – R732.

Wolpert, D. M., Ghahramani, Z., and Flanagan, J. R. (2001). Perspectives and problems in motor learning. *Trends in Cognitive Sciences*, 5(11):487–494.

Wolpert, D. M. and Kawato, M. (1998). Multiple paired forward and inverse models for motor control. *Neural Networks*, 11(7):1317–1329.

Wright, S. (1931). Evolution in Mendelian populations. *Genetics*, 16(2):97.

Wrigley, W. J. and Emmerson, S. B. (2013). The experience of the flow state in live music performance. *Psychology of Music*, 41(3):292–305.

Xambó, A., Hornecker, E., Marshall, P., Jorda, S., Dobbyn, C., and Laney, R. (2013). Let's jam the Reactable: Peer learning during musical improvisation with a tabletop tangible interface. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 20(6):36.

Yee-King, M. J. (2007). An automated music improviser using a genetic algorithm driven synthesis engine. In *Applications of Evolutionary Computing*, pages 567–576. Springer.

Yee-King, M. J. (2011). *Automatic sound synthesizer programming: techniques and applications*. PhD thesis, University of Sussex.

Young, J. A. and Pain, M. D. (1999). The zone: Evidence of a universal phenomenon for athletes across sports. *Athletic Insight: The Online Journal of Sport Psychology*, 1(3):21–30.

Zappi, V. and McPherson, A. (2014a). Design and use of a hackable digital instrument. *Proceedings of the Live Interfaces Conference*.

Zappi, V. and McPherson, A. P. (2014b). Dimensionality and appropriation in digital musical instrument design. In *Proceedings of the International Conference on New Interfaces for Musical Expression*.

Zatorre, R. J., Chen, J. L., and Penhune, V. B. (2007). When the brain plays music: auditory–motor interactions in music perception and production. *Nature Reviews Neuroscience*, 8(7):547–558.

Zeng, X., Hedge, A., and Guimbretiere, F. (2012). Fitts' law in 3d space with coordinated hand movements. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, volume 56, pages 990–994. SAGE Publications.

Zhai, S. (1993). Investigation of feel for 6DOF inputs: Isometric and elastic rate control for manipulation in 3d environments. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, volume 37, pages 323–327. SAGE Publications.

Zhai, S. (2004). Characterizing computer input with Fitts' law parameters, the information and non-information aspects of pointing. *International Journal of Human-Computer Studies*, 61(6):791–809.

Zhai, S., Hunter, M., and Smith, B. A. (2002). Performance optimization of virtual keyboards. *Human–Computer Interaction*, 17(2-3):229–269.

Zhai, S. and Milgram, P. (1998a). Quantifying coordination in multiple DOF movement and its application to evaluating 6 DOF input devices. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 320–327. ACM.

Zhai, S. and Milgram, P. (1998b). Quantifying coordination in multiple dof movement and its application to evaluating 6 dof input devices. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '98, pages 320–327, New York, NY, USA. ACM Press/Addison-Wesley Publishing Co.

Zhong, C.-B., Dijksterhuis, A., and Galinsky, A. D. (2008). The merits of unconscious thought in creativity. *Psychological Science*, 19(9):912–918.

Zimmerman, J., Forlizzi, J., and Evenson, S. (2007). Research through design as a method for interaction design research in HCI. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 493–502. ACM.