A First and Second Longitudinal Study of Haptic Icon Learnability

The Impact of Rhythm and Melody

by

Bradley Adam Swerdfeger

B.Cmp.H, Queen's University, 2007

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE

in

The Faculty of Graduate Studies

(Computer Science)

THE UNIVERSITY OF BRITISH COLUMBIA

(Vancouver)

December 2009

ABSTRACT

The design and evaluation of haptic icons – brief, meaningful tactile stimuli – has been studied extensively in the research community. Haptic icons are designed to support communication of information through the often-underutilized haptic modality. However, the learnability of haptic icons has not been evaluated in an ecologically plausible, longitudinal deployment scenario.

This thesis endeavours to evaluate the learnability of haptic icons in a realistic context. We assign abstract meanings based on a realistic context to a large, previously developed set of rhythmic haptic stimuli. Then, during a period of 12 sessions over 4 weeks, we train users to recognize these icons and observe identification performance under workload using a Tetris game interruption task. Icons are presented to users in sets of 7. Upon the mastery of their current 7 icons, the user graduates to a new set, but must remember previously learned icons.

We discover that perceptual discriminability dominates learnability – the semantics of the icons have very little effect. We also find evidence that design based on multidimensional scaling (MDS) is adequate for developing haptic stimulus sets, but can be quite conservative in its identification performance predictions during deployment. Haptic icon learning is characterized by a peak in difficulty after learning progresses past a single group 7 icons, which may be explained by cognitive long-term encoding and an increase in perceptual sensitivity. In addition, we present a series of heuristics for designing rhythmic haptic icons, as well as guidelines for haptic icon training and advice for hardware designers.

In an attempt to increase the expressiveness and learnability of rhythmic haptic icons, we explore the addition of melody. We iteratively develop a second set of 30 melodic haptic icons using an MDS methodology. We discover that rhythm dominates user categorization of melodies. This work also results in a set of heuristics for designing melodic icons.

Finally, we evaluate the learnability of this new melodic set using our previous longitudinal methodology. Our results indicate that purely rhythmic haptic icons are easier to learn than melodic haptic icons that are grouped by rhythm and are thus more viable for deployment.

TABLE OF CONTENTS

Appendices

LIST OF TABLES

LIST OF FIGURES

ACKNOWLEDGEMENTS

The learning process experienced while performing research is a rare and irreplaceable one. Every facet of this work has resulted in some kind of maturation where I feel like I am better able to be a successful contributor to society. I did not do this alone.

First and foremost, I would like to thank my supervisor Karon MacLean for her guidance, support and encouragement. I never left a meeting without feeling a little bit more on top of things – this is was entirely her doing, as she gave me the advice and drive needed to continue progressing on a very challenging thesis.

I would also like to thank my second reader, Bob Woodham for his help and feedback in preparing this manuscript.

There are many other people who made this thesis possible, including family, friends and collaborators (which are not mutually exclusive groups).

First, I'd like to thank Jennifer Fernquist and Thomas Hazelton for their immensely important collaboration in this thesis. I had a blast doing the melody work. Tom's friendship and lunch times at the OnlyU were great for my sanity although not for my dietary health :P. Jen's companionship has made this entire experience much, much more fun that I'd imagined was possible. Her supportive and caring manner always brought me back up when I was feeling overwhelmed.

Ilia Pak was also extremely helpful in his work administering the second longitudinal study. His enthusiasm, reliability and intelligence made this work much easier to do.

Thanks to the good people at Nokia who made this work possible by providing funding, support and hardware. Most notably, Patrick Wong, Mihaela Zahariev and Topi Kaaresoja. I'd also like to thank NSERC for their generous support through the Canadian Graduate Scholarship fund.

I'd also like to thank everyone in the 'UBC Sushi and Other Fun Things Friends' group. Your friendship made these past could of years extremely memorable and enjoyable. Thanks to my boys, Mike Ho, James Relyea and David Liimatainen for always being rock solid.

My family played no small part, as well. My brother Derek and stepfather Stephen Headlam deserve recognition for their support. Finally, I must thank my mother, Elaine Casselman, whose encouragement, pride and love has brought me this far. Thanks for all of the cookies and gum, as well.

STATEMENT OF CO-AUTHORSHIP

I hereby declare that this thesis incorporates material that is the result of joint research.

Some initial experimental design and implementation for the work described in Chapter 3 was completed by David Richard Ternes.

All of the work described in Chapter 4, including experimental design, implementation, experiment administration, data analyses and manuscript preparation was completed in collaboration with Jennifer Fernquist and Thomas W. Hazelton under the supervision of Karon E. MacLean. Written permission to use this work in this thesis has been obtained by the authors. An adaptation of this work was published as *Exploring melodic variance in rhythmic haptic stimulus design* in Graphics Interface 2009, pages 133 – 140.

The experiment described in Chapter 5 was administered and executed by Ilia Pak, with direction from the author.

All work described in this thesis was performed under the supervision of Karon E. MacLean. We collaborated extensively in defining the research plan, experimental design, data analyses and manuscript preparation.

CHAPTER 1

INTRODUCTION

Imagine that you're on a bus on your way home from work. Naturally, the bus is crowded due to the fact that it is rush hour, and all of the seats are occupied – you have to stand while holding onto the pole. It's raining outside, so your other hand is encumbered with your umbrella, the groceries that you picked up after work, and your day bag. A colleague is going the same way as you, so you strike up a conversation.

This mundane situation features many visual and auditory tasks, in both the background and the foreground. To make matters worse, your hands are also completely occupied, as well as your feet and proprioceptive system with information about bus progress, starts, stops and turns. Your foreground task is the conversation that you are having with your colleague. Holding a conversation is a cognitively demanding visual and auditory task. You must pay attention to your colleague's facial expressions and maintain eye contact in order to increase engagement. Furthermore, you must listen to and understand the auditory linguistic content in the presence of noise from the bus' engine, as well as the noise from outside and inside. Your background task is to make sure that you get off at the right stop. Due to the rain, the windows are foggy, so it's extremely difficult to tell where you are. In addition, the bus is too loud to hear the bus driver announcing the stops. This background task is mission critical. If you miss your stop then you have to walk an increased amount of time in the rain, while already encumbered. However, you also do not want to offend your colleague by not paying attention to your conversation.

This task has huge demands on attention. Both your visual and auditory modalities are receiving a great deal of stimulation and your attention is occupied and divided among multiple tasks. Now, what happens if we add a mission critical notification to the task? Let us say that your spouse is trying to call or text you on your cellular phone in order to notify you that he/she is running late, and you will have to find a way to pick up the kids from school, thus altering your route. Assuming that you even manage to hear your phone over all of the noise, you may not bother to dig your phone out of your pocket to view the message because your hands are full and you are holding on to the pole for safety.

One can imagine many similar situations where a person is performing multiple visual and auditory, foreground and background tasks among important notifications. Piloting an airplane is a commonly explored high workload task due to the mission critical nature of a pilot's reactions [36].

What if we could administer this notification as a vibratory pattern through your cellphone, or another device situated on your person? Through this pattern of vibrations, you will be able to identify the sender and contents of the message, all without occupying your hands, eyes or ears. In addition, the message can be understood in the background, with very little additional demands on attention or as an interruption to your primary task. Finally, the message is transmitted privately and you are not required to share its arrival or contents with others.

We anticipate that *haptic icons* – brief tactile or force stimuli associated with a meaning – will demonstrate the greatest utility in situations where other senses are occupied, such as those outlined above.

Many researchers have explored haptic icons extensively on subjects such as their perceptual properties [6, 32, 42–44], design [10, 12, 32, 43], structure [6, 8, 17] and short term learnability [6, 7, 10, 17]. However, the long term potential of haptic icons have not been evaluated. In order to understand the viability of haptic icons in mission critical, deployment scenarios, they must be understood in a realistic, longitudinal context. In addition, melodic variation of haptic stimuli as a design parameter has yet to be explored due to its daunting design space [43].

Thesis Objectives

The main goal of this thesis is to answer the following question: How effective are haptic icons in a realistic deployment scenario, and what is the learning trajectory for users? We attempt to address these questions through a longitudinal evaluation of the largest perceptually evaluated set of haptic stimuli that currently exists. Due to the richness of the data collected during this process, we can also provide heuristics for effective design, guidelines for training and advice for hardware designers.

Another important goal of this thesis is to understand the role of melody in haptic icon design. Melody offers benefits to the expressiveness and articulation of haptic icons but was previously unexplored. We endeavour to understand how melodic icons should be designed, as well as assessing their viability in a deployment scenario through longitudinal evaluation with a comparison to our previously determined baseline.

1.1 Background

Throughout this thesis, we assume some background knowledge and an intuitive understanding of certain concepts relating to haptic icons. In this section, we will give a brief overview of these concepts in order to aid comprehension.

1.1.1 Icons vs. Stimuli

There are two aspects of a haptic icon: a *stimulus* and a *meaning*. As such, in order to create a haptic icon one must first develop the haptic perturbations that a user perceives (the stimulus) and following that, assign semantics to the stimulus (the meaning).

This thesis deals with both aspects of the haptic icon design process. When we refer to an *icon*, we are referring to a haptic stimulus that is laden with a specified meaning. For instance, an icon might be a series of vibrations which signify the semantics *'Call me'*. When we refer to a *stimulus*, we are referring to a haptic stimulus *without* any meaning. This is simply an actuation that has not been assigned any semantics. Chapter 4 deals strictly with haptic *stimuli*, while Chapters 3 and 5 deal with complete *icons*.

1.1.2 Actuation Technique

Haptic stimuli are rendered by applying forces to a user's body. There are many methods for producing haptic stimuli. One very common and well-known actuation technique exists in most modern mobile phones. In these phones, there is a rotary motor with an offset weight on the armature. This offset weight causes an oscillation of forces about the armature, resulting in a vibration that is felt by the user.

For the creation of all stimuli described throughout this thesis, we utilize a Nokia 770T. For an apt description of the device and its actuation technique, we refer you to the following quotation that is taken verbatim from David Ternes' master's thesis [43]:

> The Nokia 770 (Figure 1.1) is a handheld internet tablet, with a large (90x54 mm) high resolution (800x480) screen, ARM-based processor, and runs a modified version of the Debian Linux distribution. While the 770 is already commercially available, Nokia has added haptic feedback to a prototype model, identified as the 770T (see [27] for details). Though visually identical to the 770, the 770T has a piezo-mounted touchscreen, which allows the screen to be pulsed with small displacements in the axis orthogonal to the screen, giving the sensation of a single 'click' when done once, and of a continuous vibration when done repeatedly at tightly spaced intervals. This technique can give quite convincing and satisfying haptic feedback, all within the context of a handheld device.

Figure 1.1: The Nokia 770T Internet Tablet

For a discussion of the hardware platform's advantages and disadvantages, please consult [43].

A *piezoelectric material* changes its shape when a voltage is applied to it. In order to create the perception of a continuous vibration, we administer a series of voltaic impulses through the use of a scripting language that controls the piezo controller (depicted in Figure 1.2). The details of the scripting technique for haptic feedback control is described in [43].

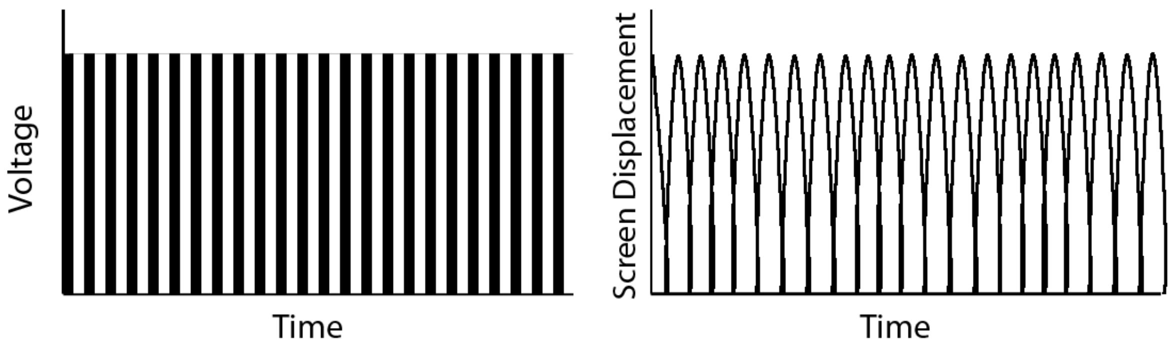

Figure 1.2: Depiction of the actuation technique for administering a vibratory stimulus on a piezo-mounted touchscreen. Values are approximated.

The degree of the piezoelectric material's shape change is proportional to the amplitude of the voltage applied.

1.1.3 Frequency and Amplitude

Throughout this thesis, we refer to variations in the *frequency* and *amplitude* of a vibration or stimulus. In the auditory perception domain, a change in frequency of a wave corresponds to a change in perceived pitch, and a change in amplitude corresponds to a change in loudness. These dimensions are similar in the haptic modality.

In order to increase/decrease the frequency of a vibration, we increase/decrease the number of voltaic impulses that occur per second. The usable output range of the device utilized (Section 1.1.2) is 100Hz to 800Hz. An increase in frequency from the waveform in Figure 1.2 is depicted in Figure 1.3.

In order to increase/decrease the amplitude of a vibration, increase/decrease the voltage of an impulse, corresponding to a proportional change in maximal screen displacement. A decrease in amplitude from the waveform in Figure 1.2 is depicted in Figure 1.4.

For an intuitive understanding of the perceptual differences corresponding to variations of these parameters, consider touching the cage of a household electric fan. As the fan spins faster, its vibration increases in frequency. Now consider a larger fan, operating at the same frequency; the spinning of this fan will result in vibrations with a larger amplitude.

Figure 1.3: Depiction of a high frequency vibration.

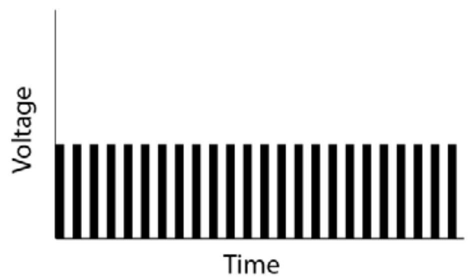

Figure 1.4: Depiction of a low amplitude vibration.

1.1.4 Rhythm vs. Melody

In the context of this thesis, the term *rhythm* refers to a periodic expression of vibrations whose constituent notes have *constant* frequency and amplitude. The term *melody* refers to a periodic haptic stimulus whose constituent notes *vary* in frequency and amplitude. Note that a melody contains a rhythmic component.

Figure 1.5 demonstrates this distinction pictorially. It also demonstrates the visualizations for periodic haptic stimuli that we utilize throughout this thesis.

1.1.5 Perceptual Multidimensional Scaling

Perceptual Multidimensional Scaling (MDS) is an established technique for visualizing how users perceptually organize a set of stimuli. The algorithm takes a dissimilarity matrix of the stimuli and reduces this large dimensional space to a specified number of dimensions where the variance along these dimensions is maximized as much as possible. The version of MDS that we use in Chapter 4 is identical to principal components analysis (PCA). In other words, the first dimension accounts for as much variability in the dissimilarity matrix as possible, while subsequent dimensions account for as much of the remaining variability as possible. These new, composite dimensions may give insight into complex, perceptual

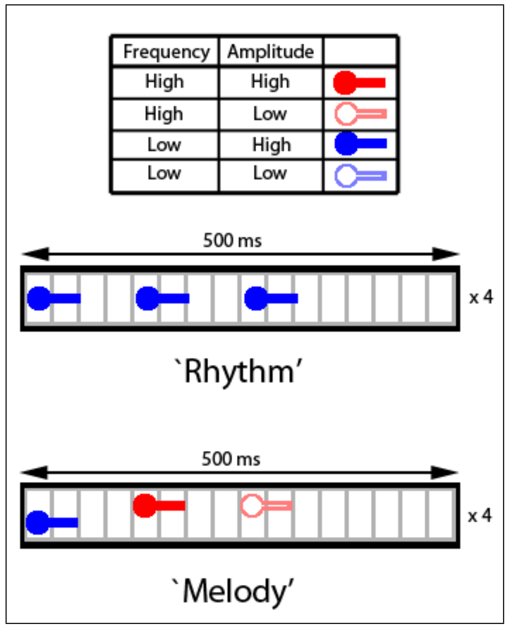

Figure 1.5: Graphical depiction of a *rhythm* vs a *melody*. For our periodic stimuli, each bar is 500ms in duration and is repeated 4 times. These parameters are explained in detail in Section 3.1.1.

dimensions. For instance, Hollins *et al.* [23] found dimensions such as hard/soft and slippery/sticky for real tactile surface textures. In our work we use the relatively efficient and accurate cluster-sorting method adapted by MacLean and Enriquez [32] and further analyzed and validated by Pasquero *et al.* [38] and Luk *et al.* [30].

1.2 Approach and Overview

The primary aim of this thesis is to evaluate the potential effectiveness of haptic icons in a deployment scenario, as well as observing the haptic icon learning trajectory if users are given a significant amount of time. In addition, we wish to explore the use of melody for designing haptic icons in order to ascertain whether the increased expressiveness provided by a larger design space results in improved effectiveness over previously existing icon sets. In support of these goals, the novel work described in this thesis is divided into three separate, but related components.

We begin by outlining previous, related work pertaining to iconography, haptic icons, and longitudinal evaluations of haptic learning in Chapter 2.

In the novel component of the thesis (Chapter 3), we perform a longitudinal evaluation of icons derived from the largest haptic stimulus set to date. Ternes and MacLean [43] developed a set of 84 rhythmic stimuli using heuristic design, as well as MDS as a perceptual validation technique. Due to the size of the set, we consider it to be a viable candidate for evaluating the limits of haptic icons in a deployment scenario. We assign meanings to these stimuli within a deployment scenario of cellular messaging and train users to identify these icons over a period of one month. Our experiment is a classical immersed task with interruptions, where we evaluate icon identification performance while users are immersed in an unrelated, primary task. Through this experiment, we can understand a great deal about the effectiveness of haptic icons with respect to their learnability, as well as providing heuristics for future design, guidelines for training and advice for hardware designers.

In the next part (Chapter 4), we explore the use of melody as a design parameter for haptic stimulus design. Similarly to [43], we utilize a perceptual MDS evaluation methodology and perform an iterative design process. We begin by defining design heuristics based on musical theory to define an initial set of melodic haptic stimuli. This set is evaluated through a user study where we have users perform a cluster sorting task and we observe the perceptual properties of the stimuli through MDS. Using the results obtained from this evaluation, we perform an additional design and evaluation iteration in order to arrive at a set of 36 rhythm-based melodic haptic stimuli with desirable perceptual properties. This design process also results in a set of design heuristics for perceptually groupable melodic haptic stimuli.

In the final novel component (Chapter 5), we evaluate this rhythm-based melodic stimulus set by assigning meanings within our cellular messaging scenario and perform another longitudinal evaluation that is identical to the study conducted previously. In this part, we focus on a comparison of the effectiveness of purely rhythmic icons versus rhythm-based melodic icons in a deployment scenario.

Finally, in Chapter 6 we conclude the thesis by providing a summary of contributions and directions for future work. At the conclusion of this thesis, we will have provided a strong justification for the viability of rhythmic haptic icons in a deployment scenario. In addition, we provide icon designers with actionable heuristics, as well as guidelines for training and advice for hardware designers.

CHAPTER 2

RELATED WORK

Iconography has been used in computer systems since the advent of WIMP (Windows, Icons, Menus and Pointers) interfaces. By taking advantage of the incredible human capacity for symbols [14], icons serve as a natural way to *represent information* in complex environments. With the recent advances in haptic display technology, researchers have been building on previous work in visual and primarily auditory icons to develop sets of *haptic icons*. A comprehensive overview can be found in [31]; here, we summarize the most relevant details.

2.1 Auditory Icon Design Approaches

Auditory and haptic icon design share a key attribute since both modalities are temporally sequential [22, 32]. Auditory icon design can be divided into two philosophies of design: *representational/metaphorical* and *abstract*.

Gaver [19] introduced auditory icons by representing information with a specific sensory experience that is directly related to the item being symbolized so that the link between stimulus and meaning is as intuitive and natural as possible. For instance, dragging an object might be accompanied with a 'scraping' sound. Unfortunately, this approach suffers from poor salience control. An unimportant event may be perceived as more salient or similar to a critical notification. This problem lies in conflict with Weiser's ideal of 'calm technology' [47] which serves as a widely accepted philosophy for non-visual interaction design.

Blattner *et al.* [2] take an abstract approach to designing structured 'Earcons' where their 'motives' (a series of notes that differ in pitch and amplitude) can be combined to create compound icons. For example, one can combine the motives sequentially for 'destroy' and 'file' to represent 'delete file' abstractly. Brewster *et al.* [5] extended this work by examining how people can perceptually differentiate 'Earcons'. They found that the structured approach aided in differentiation, as did varying timbre rather than restricting stimuli to simple tones.

Although Blattner and Brewster both provide a theoretical evaluation and describe heuristics for increasing the learnability of earcons, an empirical evaluation was not conducted.

2.2 Haptic Icon Design Approaches

A variety of approaches to haptic icon design have been attempted since this is new ground that must be validated by research before it is applied commercially. These strategies include perceptual, musical and structural design. In addition, we explore the past use of melody in haptic design.

2.2.1 Perceptual Design

MacLean and Enriquez [32] emphasize that the stimuli in haptic icon sets should be designed by first understanding how synthetic haptic signals are perceived and then later assigning meanings to these perceptually validated stimulus sets. Their design process involves fully exploring the output space of their device and then performing quick, iterative user studies to plot the perceptual space of their icons using MDS (**M**ulti**d**imensional **S**caling, elaborated upon in Section 1.1.5) in order to make their icon sets as distinguishable as possible. Using this method, they created 36 icons that vary in waveform, amplitude and frequency.

Enriquez and MacLean [16] hypothesized that when users can choose the haptic signals that represent particular concepts, the learnability of these icons would be increased. Contrary to this hypothesis, they found no significant difference in recall performance between conditions arbitrary and participant-chosen stimulus-meaning associations. In addition, after 20 minutes of learning, users were able to recall icons at 80% accuracy and then subsequently recall 86% of the icons that were initially identified correctly after two weeks. This study is a first result in the longitudinal recall of haptic icons, revealing a great deal of potential for success in deployment, as well as a surprising ability for users to develop helpful mnemonics despite an arbitrary meaning assignment strategy.

Using a similar method to [32], Ternes and MacLean developed the largest stimulus set to date with 84 perceptually distinguishable tactile stimuli [44]. They created these stimuli by first using heuristics to choose 21 *rhythms*, then expanding this set with two variants each of amplitude and frequency, applied as a constant to the entire rhythm (4×21 monotone stimuli). This set, which is used as the basis of the first longitudinal study in this thesis, is elaborated upon in Section 3.1.1. We are not aware at this time of other examples of icon design that includes an explicit component of perceptual validation of stimuli, independent of meaning association.

2.2.2 Structural Design

Attempts at creating structured sets of haptic icons have focused on family-based approaches: icons in each family share haptic features, and conversely share semantic components, thus increasing set learnability by allowing users to 'chunk' groups of items [34].

Chan *et al.* [10, 11] create a representational set of haptic icons in the context of remote collaboration, where the metaphor used for design reveals the family of the icon. They created seven icons by varying the order of tones, number of pulses and magnitude for families representing *changes* in control, *being* in control and *waiting* for control, respectively. Their set achieved 95% recognition rates under workload after three minutes of learning.

In Enriquez *et al.*, stimulus frequency corresponded to the icon's family and the waveform to its function [17]. They demonstrated a 76% recognition rate in completely arbitrary meaning-matches (for a conservative test) after 20 minutes of practice for a set of nine icons encoded as a 3^2 matrix.

Brown [6] found 73% user accuracy in identifying nine two-dimensional 'Tactons', where dimensions of priority and message type (3 priorities, 3 types) were encoded as roughness and rhythm.

Two of these studies [10, 17] also employed perceptual MDS to validate and refine the icons within their sets.

Although the family-based approaches appear to be effective, they are limited to relatively few families. Enriquez *et al.*'s approach limits its expressiveness for families to one dimension (frequency) and is therefore limited by perceptual acuity along that dimension. For instance, if the haptic device is capable of displaying a frequency range of 500 Hz and humans can only perceive differences of 100 Hz reliably, then the technique is limited to at most six families. Representational approaches illustrated by Chan *et al.* [10] and (in the auditory domain) Gaver [19] are weak both in repeatability (through reliance on designer creativity in generating good metaphors, which is particularly difficult for more abstract concepts) and salience management; and consequently in scalability.

Ternes and MacLean found a strong perceptual axis that differentiates even ('continuous') rhythms versus uneven ('lurching') rhythms. We use this perceptual dimension as a basis for structural design in Chapter 3.

2.2.3 The Use of Melody

Van Erp and Spape [46] created a set of 59 haptic stimuli by translating music sequences from the auditory to the vibrotactile domain on the basis of note tone. They found that users distinguish these melodies on dimensions of *intrusiveness* and *tempo*. However, this investigation did not extend to meaning assignment or learnability, leaving open the question of whether designers can reassign arbitrary semantic associations to stimuli that might already have meaning for the user.

2.3 Haptic Perception and Learning in General

In the book Plasticity in the Human Nervous System: Investigations with Transcranial Magnetic Stimulation, the authors cite a study where unsighted people learning braille show a gigantic increase in the cortical representation of the dominant reading finger after 2 months of learning [3]. There is also a slower, steady increase in representation that is believed to be caused by recruitment of other structures.

These results indicate that when people partake in symbolic haptic learning, the brain areas associated with sensitivity in the reading finger increase dramatically in size. For the longitudinal studies that we conduct, we expect that an increase in sensitivity due to an increase in cortical representation will occur, resulting in an increase in discriminability.

Newman et al. [35] show that the braille letters A-J are more discriminable than the letters K-T, and this resulted in a very significant effect of learnability: people were able to learn the letters A-J much faster than they could learn K-T. Sighted participants were given five trials to learn the haptic stimuli. It was also shown that the number of errors was directly related to complexity of a braille letter with respect to the number of dots that it contained.

2.4 The Motivation for our Work

The use of perceptual design techniques such as perceptual multidimensional scaling has been established by researchers in our field as a viable haptic icon design technique. Its efficacy has been demonstrated in semi-deployed situations, which are examples of the potential value in practical applications [10, 11, 17].

In the domain of haptic learning, we believe that humans have great potential to learn a large number of haptically expressed symbols. Research by Enriquez and MacLean showed that people have a surprising ability to remember abstract haptic stimuli even after very little exposure [16]. Brain imaging and braille learning studies performed on non-sighted participants has revealed impressive feats of symbolic haptic learning [3, 35]. Evelyn Glennie, a Scottish percussion virtuoso, learned – and thrived – as a solo percussionist by using haptic cues throughout her body [48].

We believe that there is a need for the communication of information through the haptic channel since today's computing device users are inundated with visual and aural notifications, with the consequence of useful information becoming an irritating interruption. The haptic sense has the potential to support background communication that can be designed to reduce disruption in portable and embedded applications. With the use of larger icon sets, we can communicate a more diverse set of information to users.

However, we have some uncertainty whether the effort involved to learn these large sets of haptic icons is practical for non-impaired users.

In this thesis, we address how difficult it is for non-impaired users to learn haptic icons, as well as how long it might take to become proficient with a sufficiently large set. We also wish to understand how users can be assisted in this effort through training techniques, as well as design heuristics. Furthermore, we aim to determine the level of performance that they can achieve over a period of time that is comparable to practical scenarios, such as mobile messaging.

CHAPTER 3

LONGITUDINAL EVALUATION OF RHYTHMIC HAPTIC ICONS

Previous work in haptic icon design and evaluation is based upon relatively small icon sets. These icon sets are primarily evaluated through relatively short-duration user studies, where users typically have less than an hour to learn an icon set, and then their recognition performance is tested using a short, one-off quiz [6, 10, 11, 17]. Although this evaluation technique provides a great deal of insight about the immediate perception of haptic icons, as well as their suitability for short-term recall situations, it is difficult to conclude whether these results will generalize to extended usage situations, which both offer greater opportunities for learning, and greater challenges for remembering. These types of situations would occur in the event that haptic icons are deployed within embedded devices and other products.

Enriquez and MacLean reveal initial evidence that people are considerably more adept at identifying a set of haptic icons, after a two week interval without any exposure, than self-reported confidence levels would predict [16]. This result may indicate that haptic icons may be more effective than one would intuitively believe in a deployment situation.

Despite this initial result, there have been no longitudinal evaluations regarding the effectiveness of haptic icons. Haptic icons were developed for a purpose: to provide sub-attentive, background information to users through an under-utilized modality. They exist to provide information through background channels to users who are interacting with this notification system throughout their daily lives or professions. Currently, we have no definitive indications if haptic icons would scale well or be effective in deployment scenarios.

This chapter aims to fill that need. We endeavor to develop an understanding of the effectiveness of haptic icons in a longitudinal, deployment scenario. With the rich data that we collect over this relatively long period of time, we also strive to develop a series of heuristics for haptic icon designers, provide guidelines for haptic icon training, and advise hardware designers in order to make haptic icons as effective as possible.

In this chapter, we will describe a month-long training and evaluation of haptic icons. We will begin by summarizing the haptic icon design process, including a summary of the previously completed stimulus design process [43] and the meaning attribution process (Section 3.1). Next, we will describe an experiment that follows an immersed task with interruptions paradigm (Section 3.3). Users will be immersed in a task, and they must identify haptic icons that occur periodically as interruptions. This experiment runs for the period of one month, and features numerous sessions of training and evaluation. Next, we will present our results and discuss them with the goal of determining the effectiveness of haptic icons in a longitudinal scenario, as well as providing heuristics for haptic icon designers, guidelines for haptic icon training and advice for hardware designers (Sections 3.4 to 3.7). Finally, we will summarize the primary contributions of this chapter (Section 3.8).

3.1 Haptic Icon Design Process

Recall that a haptic icon is a brief, tactile or force feedback stimulus associated with a meaning. Therefore, the creation of haptic icons is a two-step process: first, design a haptic stimulus; second, assign a meaning to the stimulus. These two steps can be approached using a multitude of techniques [6, 10, 11, 16, 17]. In this section, we describe our haptic icon design process, which is largely a continuation of a previous body of work.

3.1.1 Haptic Stimulus Design: Summary of Previous Work

The haptic stimuli utilized in this work were designed by David Ternes and Karon MacLean [44]. Although the development process for these stimuli is explained in detail through their work, we will provide a summary in this section since understanding the origin of our stimuli is essential for coherence. Substantially more detail, background and rationale can be found in David Ternes' master's thesis [43] or their Eurohaptics paper [44].

Ternes and MacLean's stimulus set was designed by utilizing *rhythm* as a design parameter. They define a rhythm as a repeated, non-melodic (or monotone) pattern of notes arranged relative to a beat (4/4) and played at a constant tempo. Variation between rhythms is obtained solely by arranging the *length* and *number* of notes, as well as the *spacing* between them (rests). For an intuitive exemplification, imagine an auditory tone with a constant amplitude and frequency, controlled by an on/off switch. Any repeated pattern of 'on' and 'off' durations – with a consistent time signature and tempo – is defined as a rhythm.

They designed their rhythmic stimulus set with an *include/exclude* heuristic and constraint-based design procedure. These heuristics (informed choices) and constraints (measured necessities) were developed through a detailed analysis of the tactile rhythm space, which was based on extensive informal user testing. They first performed an *exclude* step, where they identified the entire stimulus space expressible by their hardware and eliminated possible stimuli based on practical or perceptual constraints. Next, they performed an *include* step with the remaining stimulus space, where they selected the most auspicious stimuli based on positive heuristics, up to a target size of 21 rhythms.

Next, they expanded the set to 84 stimuli by multiplying the 21 rhythms by *two amplitudes* (200Hz and 300Hz −) and *two frequencies*, which were determined to be perceptually optimized for acuity through piloting and [45]. The two frequencies used were 200 Hz and 300 Hz, which are consistent with findings related to the nominal frequency for human sensitivity to vibrations at 250 Hz [33]. Table 3.1 enumerates their 84 stimuli in terms of frequency and amplitude. Each group of 21 stimuli utilizes the same 21 heuristically determined rhythms.

Stimulus Group	Amplitude	Frequency
1–21	High	High
22–42	High	Low
43–63	Low	High
64–84	Low	Low

Table 3.1: Frequency and amplitude groupings developed by David Ternes and Karon MacLean in [44].

Finally, they performed perceptual evaluation on the set using a variant of a cluster sorting [32, 43] and Perceptual Multidimensional Scaling (MDS) methodology [23].

From our perspective, they determined one very surprising, but important result: *The "evenness/unevenness" of a rhythm is the most clearly delineated perceptual axis.* Even rhythms have a regular, consistent repeating pattern, where uneven rhythms have an 'irregular, lurching feel with emphasis emerging on the first part' [44]. The perceptual difference between these two types of rhythms is extremely distinctive.

Their MDS analysis also revealed that the rhythms devised were perceptually distinguishable due to a general lack of clustering, therefore we sought to determine the learnability of this set with an abstract meaning assignment paradigm.

3.1.2 Meaning Attribution

Although our primary aim was to determine the limits of abstract haptic icon learnability, we found it necessary to make some design decisions during the meaning attribution process:

- **Ecological Validity and Message Contents:** Should our icons be completely abstract, similar to [17], where a stimulus might represent a fruit or plant; or should they emulate a deployment situation, where haptic stimuli are assigned meaningful messages that one might encounter in a mobile messaging application? If they are messages, what kind should the contents be? For instance, are they system notifications or messages sent by contacts?

- **Grouping Strategy:** Should perceptually similar stimuli be grouped meaningfully, or should each stimulus have orthogonal semantics?

- **Role of Amplitude/Frequency:** Should stimuli with identical rhythms but different amplitudes/frequencies be assigned completely distinct meanings? If rhythm is defined as the only abstract unit, what role does frequency and amplitude play?

Ecological validity and message contents

There are two possible methodologies that we could have pursued for abstract meaning assignment: 1. purely and completely abstract, whereby the stimulus has absolutely no relation to the meaning; or 2. an ecologically plausible context which frames the assignment of abstract meanings.

A purely abstract assignment methodology would be for instance, assigning a stimulus to be *Banana*, and another *Ficus*. This meaning attribution strategy would reveal insights into the abilities of the human mind with respect to learning completely abstract haptic icons. What we may learn using this methodology may be generalizable to any context and represents a worst case – in this sense, it is pure, theoretical research, similar to what is pursued in the field of cognitive psychology.

The alternative strategy would be to adopt a philosophy more in line with applied research. We could choose a deployment context and assign related meanings in order to understand the viability of this technology within a certain deployment scenario. We would ensure that our meaning assignment strategy is as abstract as possible within reasonable parameters determined by our deployment context. Results obtained using this methodology could potentially pertain only to the chosen context, but they would reveal what is possible even when performing very little actual design, which is an expensive process. The resulting methodology would be a systematic and repeatable process, potentially within any context.

As computer scientists interested in mobile deployment scenarios, we chose the applied strategy. We seek to learn about the application of a technology within a real deployment scenario. *Haptic icons are not a vehicle through which to understand perceptual psychology, they are a potential solution to a clearly identified need.* We wish to investigate their efficacy.

For this work, we avoid performing metaphorical design. In metaphorical design, there is a meaningful relationship between the semantics of an icon and its stimulus. We do not believe that the metaphorical design process is scalable to the size and nature of our stimulus set. Metaphorical design is non-repeatable, domain-specific, very costly, and prone to perceptual artifacts. Furthermore, there is evidence that careful, metaphorical meaning assignment has no effect on the learnability of haptic icons [16]. The confirmation of this claim would be a strong argument for the viability of abstract meaning assignment. This assertion will be untestable with purely abstract stimuli since there are no available sources of comparison.

Deployment context

For this work, we chose a cellular messaging deployment context. We imagine an application where a sender can send a receiver a predefined message. The receiver's mobile phone administers a series of vibrations which the receiver can then interpret and understand the sender's message and identity. For instance, a user might receive a message from their spouse saying, 'I'm late, I will be there'. Since the message is communicated through the haptic modality, the receiver will not even have to glance at their phone in order to assimilate the message contents.

This type of deployment scenario is becoming exceedingly more probable as haptic feedback implemented within mobile devices improves. Since mobility is a rapidly changing field, subject to a great deal of publicly generated content and open APIs, we seek to understand this context before the technology is utilized naively.

In order to better understand our deployment context, we will examine some of its properties. A typical mobile phone user:

- **Receives many different messages from few senders.** According to a news posting by Reuters in 2005, approximately 88 million mobile phone subscribers in the United States send text messages on a regular basis. Users sent approximately 130 billion text messages during that year [9]. This means that the ratio between messages and senders is very high.

- **Receives more messages from some contacts than others.** Anecdotal evidence and intuition leads us to believe that most mobile phone users have a few contacts who they receive text messages to on a regular basis, and many whom they communicate with less frequently. The distribution of messages to senders may be something approximating a Zipfian distribution.

- **Prioritizes senders and messages.** From a user's perspective, some contacts are more important than others. In addition, messages with identical content can be urgent or low priority. For instance, your spouse might send you a message that states, 'Call me at home.' This message might mean that your spouse wants you to call home when you have a chance to talk about your day, or it might mean that the dog is sick and you need to call home as soon as possible. In addition, a message from your spouse might be much more important to you than a message from your cousin. For important contacts, a mobile phone user is more likely to use special features such as custom ringtones.

Grouping strategy

Within any set of stimuli, there are some groups of stimuli that are perceptually more similar than others. Should these icons have similar meanings? Should they share similar attributes or properties? Alternatively, should we design these icons to be semantically orthogonal so that there are absolutely no meaningful relationships between stimuli? The latter case would imply a purely abstract meaning attribution methodology. Within a cellular messaging deployment context, it seems very difficult to imagine this scenario.

Appropriate to the properties of typical mobile phone usage stated above, we chose to group similar stimuli based on the sender of the message. In Ternes and MacLean's work, outlined in Section 3.1.1, they identified two very clearly delineated groups of rhythms: *even* and *uneven*. For this reason, we chose two different senders: *Spouse* and *Boss*. Even rhythms are sent by *Spouse*, while uneven rhythms are sent by *Boss*. The practice of assigning a particular property to a particular group of similar stimuli is suggested by previous work and seems intuitive from a design standpoint [6, 10, 11, 17, 34]; therefore this seems like a realistic deployment scenario. We chose only two groupings due to the perceptual nature of the stimuli, but believe that more groupings are possible if stimuli with different grouping characteristics can be designed.

The idea of using few senders for haptic messaging seems to be in line with patterns of typical mobile phone usage. It seems unlikely that users will enable haptic messaging for contacts with whom they do not communicate with frequently. However, if they commonly receive messages from a few specific contacts, they may enable haptic messaging to reduce workload and distraction.

Role of amplitude/frequency

Due to the reuse of rhythms employed to expand Ternes and MacLean's set of haptic stimuli, we face a very difficult design decision: In a deployment scenario, what should be the role of variation in amplitude and frequency?

Within their set of 84 stimuli, there are 4 copies of all 21 rhythms, each played with combinations of two levels of amplitude and two frequencies (see Table 3.1).

Despite the fact that Ternes and MacLean found a very clear perceptual separation between stimuli with high and low amplitudes, when the same-rhythm stimuli are felt in succession, they convey the impression of being multiple versions of the same stimulus. The monotone variation in frequency and amplitude gives the impression that an *attribute* of the rhythm has been changed.

In the context of Ternes and MacLean's experiment, the separation between high and low amplitude stimuli seems natural: the stimuli are arranged randomly; half are high amplitude and half are low amplitude. Since the variation in amplitude is very easy to perceive [43], this is a natural partition – the set is divided in half. After this simple division it is more likely that participants will focus on frequency or properties of the rhythms.

In the context of a set of icons, the variation in frequency and amplitude is very perceivable, but the rhythm appears to be the defining cognitive unit. For this reason, we utilize amplitude and frequency to modulate attributes of the haptic icons. Each rhythm has a specific message associated with it. The amplitude modifies the message's priority. This seems in line with the properties of typical mobile phone usage outlined above. The frequency denotes the number of times that a message has been transmitted to the user. Our stimulus set has two different frequencies, therefore we distinguish between the first sending of a message (low frequency) and the second sending of a message (high frequency).

Informal user testing revealed that the high amplitude, low frequency icons result in the most intense sensation. Since we are primarily concerned with the limits of abstract learnability, all participants must learn these 21 rhythms before any attributes are added. We would like to understand the limits of symbolic haptic memory, and then to assess the cost of learning slight variations. In our deployment context, the addition of these attributes is realistic.

Table 3.2 summarizes the assignment of amplitude and frequency to groups of stimuli.

		Frequency	
		High	Low
Amplitude	High	22–42	1–21
	Low	64–84	43–63

Table 3.2: Assignment of frequency and amplitude to each group of 21 rhythms. Amplitude modifies the priority of a message while the frequency denotes a first (low) or second (high) sending of a message.

Meaning assignment technique

As a result of the above design decisions, we have 21 stimulus units (rhythms) that require an abstract meaning assignment. Of these rhythms, 11 are uneven and 10 are even. There are 4 combinations of high/low amplitude and high/low frequency. Table 3.3 summarizes our design parameters.

Parameter	Semantics
'Evenness'	Sender
Amplitude	Priority
Frequency	Repeated

Table 3.3: Summary of design parameters determined for the meaning attribution process.

By following our design decisions, we must create 11 messages that a boss might send and 10 messages that a spouse might send. These were developed by examining pre-made text messages in mobile phones and informal design processes.

To keep the meaning assignment process as abstract as possible, we randomly assigned each message to a stimulus (while staying within the parameters defined by our stimulus grouping strategy).

The next section (3.2) details the final result of the icon set.

3.2 Haptic Icon Set

In Figure 3.1, we display a visual representation of the first 21 rhythmic stimuli and their associated meanings. As you can see, they are grouped into 3 groups of 7 and each group contains an assortment of messages from *Boss* and *Spouse* (this will be further detailed in Section 3.3.5). The contents groups are based solely on the order in which the stimuli in the original set were presented [43]. In total, there are 11 messages from *Boss* (paired with 'uneven' stimuli) and 10 messages from *Spouse* (paired with 'even' stimuli). They are all high amplitude, low frequency stimuli.

Each icon has a unique and representative 4 character code which is used by participants to identify the icon in the experiment (Section 3.3). For instance, the high priority message '*!Boss!: Check your email*' has the code '*!BCE*'. This method of identification was chosen to make the recall task as realistic as possible while simplifying the meaning input task. This is elaborated upon further in Section 3.3.7.

For the sake of space and usefulness, we will not enumerate all of the icons representing variations in frequency and amplitude for attribute modification (explained in Section 3.1.2. Figure 3.2 displays a representative sample from each attribute group.

There are a total of 84 icons. However, for the purposes of this chapter, only icons 1–63 are relevant. This will be explained in further sections.

Table 3.4 enumerates all of the icon groupings used during the study. It is helpful to examine Figure 3.1 at this time. Recall that icons 22–63 have identical rhythms and meanings to previous icons, however the attributes of the message are modified. Recall that the low priority messages have low amplitude and the repeated messages have a high frequency.

Grouping	Icons	Base Group	Attribute	Amplitude	Frequency
1	1–7	-	Base Group	High	Low
2	8–14	-			
3	15–21	-			
4	22–28	1	Repeated Message	High	High
5	29–35	2			
6	36–42	3			
7	43–49	1	Low Priority	Low	Low
8	50–56	2			
9	57–63	3			

Table 3.4: Icon groupings utilized in this work. Icons are divided into groups of 7 for batch presentation to participants.

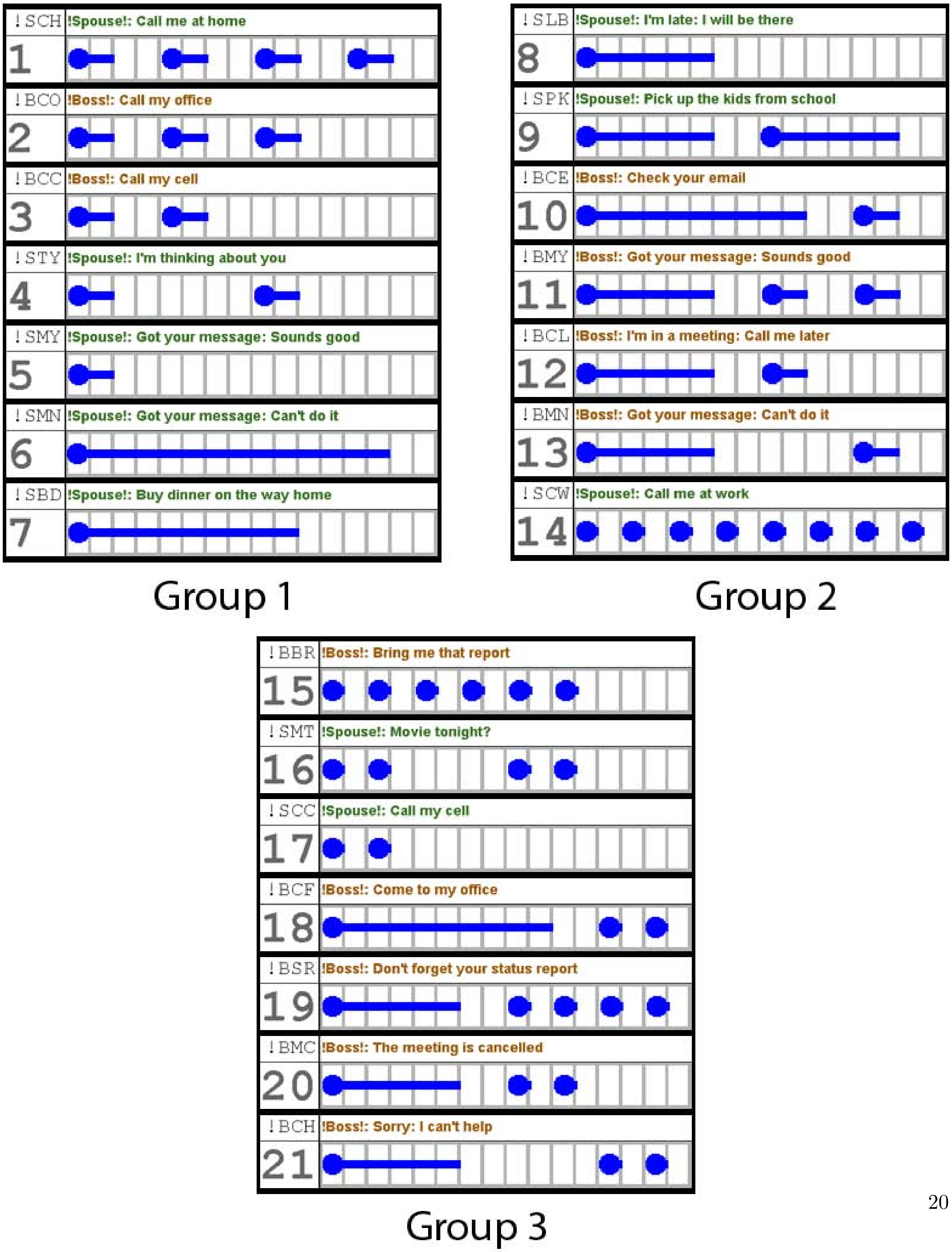

Figure 3.1: Visual representation of the 21 rhythmic stimuli and their associated meanings, separated by group.

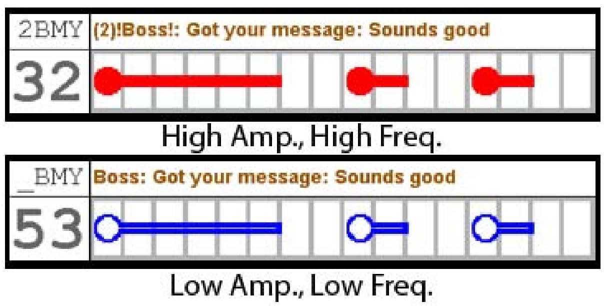

Figure 3.2: Representative samples of relevant attribute groups. Red notes denote high frequency, blue notes denote low frequency, hollow notes denote low amplitude and solid notes denote high amplitude. Thus, message 32 is high priority and being repeated, while message 53 is low priority and being heard the first time

3.3 Experiment

In order to assess the learnability of the haptic icon set outlined in 3.2 and of rhythmic haptic icons in general, we administered a longitudinal learning study which had a duration of one month and required three 20 minute sessions per week (for a total of 12 sessions) in addition to three 20 minute interviews. The details of the experiment are outlined in this section.

3.3.1 Objectives and Research Questions

Our objective for this experiment is to understand the limits of haptic icon learnability in a deployment scenario. Previously, learning has only been assessed very small numbers of icons (7–9) with very short learning periods (up to 1 hour) [6, 10, 11, 16, 17]. Now that we have a large and perceptually segregated set of stimuli, we wish to assess how easily learning can occur if users are given a long period of time for learning.

In addition, through this work, we would like to understand how we might design haptic icons to be more learnable and how we might train users to learn haptic icons so that assimilation is as fast as possible.

In order to guide our investigation, we have formulated the following research questions:

- Pertaining to icons:
  - What makes icons hard to learn?
  - What makes icons easy to learn?
  - What learning techniques do people use to remember icons?
  - What happens when we add modifiers/attributes to icons?
  - Does perception-based design work? Does it accurately predict learnability?
  - What are some generalizable haptic icon design heuristics?
- Pertaining to users:
  - How many icons can people learn?
  - How well can people retain learned icons?
  - What does the learning curve look like?
  - What makes some people better than others at learning haptic icons?
  - What are some generalizable methodologies for successful haptic icon training?

To answer the above research questions, we have developed a heavily instrumented experimental procedure so that we can capture complex, time-varying data. Our data analysis will explore this data exhaustively.

3.3.2 Hypotheses

In response to the above research questions (Section 3.3.1), for each question that does not rely on our results *a priori*, we have developed testable hypothesis that will provide structure for our experimental design.

How many icons can people learn?

Based on our incredible feats with language and other symbols such as mathematics, it is widely believed that human brains are adapted to learning symbolic referents [14], and there is no known limit to long term symbolic memory – it is believed that it is potentially infinite.

In this book, Deacon dismisses Chomsky's universal grammar module [13] and states that language and grammar have evolved to fit the architecture of human brains. Essentially, we are specialized symbolic memorizers, so there is no reason to believe that there is a limit to the number of symbolic stimuli that we can recognize. The special feats of language can mostly be attributed to a person's massive exposure and practice.

For this reason, we do not believe that there is a limit to the number of symbolic stimuli that a person can learn. However, we would like to assess the limits within a reasonable deployment scenario. We postulate that participants will learn at least 21 haptic icons with high proficiency over period of one month (for a total of 4 hours), assuming that all of the haptic stimuli are easy to distinguish perceptually.

As mentioned, the above hypothesis will be tempered by the perceptual distinctiveness of the haptic stimuli. If the stimuli are very difficult to distinguish, then the learnability will be reduced. One must first be able to perceive the differences between stimuli before they can be recognized.

What does the learning curve look like?

Since we present icons in 'batches' of 7 icons (see Section 3.3.5), we believe that participants will have fairly low performance on new icons encountered with a new batch. However, this performance will increase subject to a standard symbolic item learning curve [40].

For the entire experiment, we expect to see a sawtooth pattern for the learning curve, as participants are exposed to 7 new icons at a time, after having learned the previous set.

Which icons are easy/hard to learn?

We hypothesize that perceptually very distinct icons (with respect to the stimulus) will be easy to learn since they will be more recognizable amongst other icons.

Conversely, we hypothesize that the more perceptually proximal icons will be difficult to learn since there will be a great deal of confusion between these perceptually similar stimuli.

Does the meaning of an icon have any effect?

We do not believe that the meaning of an icon will have much of an effect on learning, if any. Previous research has shown that humans are very adept at developing learning techniques and construct cognitive scaffolding for an arbitrary assignment of meanings to haptic icons [16].

We believe that perceptual factors will be the main variable affecting performance.

What kinds of learners are there?

We believe that there will be significant individual differences in performance between users. This is due to the fact that some people are more suited to learning symbolic or abstract concepts due to the application of elaboration techniques('meaning-enhancing additions, constructions, or generations that improve one's memory for what is being learned') [29]. Users who actively and effectively use elaboration techniques will perform significantly better than those who do not.

3.3.3 Participants

Participants were recruited from 4 different labs in the Department of Computer Science at the University of British Columbia. There were 2-5 participants per lab and a total of 15 participants. Three participants were female and 12 were male. They are all aged 20 - 35 and come from various ethnic backgrounds.

All participants were graduate students, with ages ranging from 23–32. Nine participants were raised in North America, five in the Middle-East, and one in Northern Europe. Nine participants reported that they had <2 years of experience playing a musical instrument, three reported between 3–9 years and three reported >10 years. Five participants self-reported a very good sense of rhythm, eight reported a decent sense of rhythm and two reported that they did not have a good sense of rhythm.

Although we would have liked to recruit a more heterogeneous group of participants in terms of age and cognitive ability, this was logistically very difficult since the participants were required to share devices for the period of an entire month. It was not feasible to request public participants to enter a secure building three times per week for 20 minutes.

3.3.4 Compensation

Participants were paid $65 at completion of the 4 week long study. They were also given a $10 bonus if they scored within the top 50% for haptic icon learning and if they maintained within the 50% highest average score in the foreground task (Section 3.3.7). This means that they could earn a total of $85. Participants were given the option to withdraw at any point and would be paid 10$ per hour for their participation. No one withdrew.

3.3.5 Icon Group Ordering and Counterbalancing

For clarification, we will define the terms *group* and *batch* in the context of this chapter. A *group* is an experimenter-determined grouping of icons. The initial set of 21 rhythms was split up evenly in order to introduce icons to participants progressively. As you can see in Figure 3.1, icons 1–7 belong to Group 1, icons 8–14 belong to Group 2, and so on. A *batch* is more complicated, and arises due to the need to counterbalance the groups defined above. A *batch* is a group of icons presented during the experiment, relative to the user. The batch number is the order in which a participant encounters a group of icons during the experiment. Since the order of icon group presentation is counterbalanced between 3 *sets* of users, we must have a different term to refer to the ordering of icon groups that a particular participant experienced in sequence. From a participant's perspective, they proceed from Batch 1 to Batch 2, to Batch 3 and so on – the ordering of icon groupings from the experimenter's and reader's perspective is not sequential for each participant.

During the experiment, icons were presented to participants in 'batches' of seven, including up to 2 of the previous batches encountered. Once the experiment has assessed that a user has learned the new icons and retained the old icons to satisfactory performance (Section 3.3.7), they move on to a new batch while having to remember the previous icons learned. The first three icon groups and batches contain the 21 unique rhythms. The next 6 batches contain either low priority or second message versions of previous rhythms.

Icon groups were presented to participants in three different orders. Each counterbalancing *set* contained 5 people.

Table 3.5 enumerates the ordering in which the 3 different sets of participants encountered the icon groupings. It is helpful to examine Table 3.4 while viewing this table. We do not enumerate batches 7–9 since they were not reached by any participant.

Batch	User Set 1		User Set 2		User Set 3	
	New Group	Old Groups	New Group	Old Groups	New Group	Old Groups
1	1		2		3	
2	3	1	1	2	2	3
3	2	1, 3	3	1, 2	1	2, 3
4	7	1, 2	4	1, 3	8	1, 2
5	8	2, 3	7	1, 3	4	1, 8
6	4	1, 7	8	2, 3	7	1, 3

Table 3.5: Icon group ordering presentations for each counterbalancing arrangement of users. We do not enumerate batches 7–9 since they were not reached by any participant by the end of the experiment.

Notice that all participants had to learn groups 1–3 first, which contain all 21 rhythms at a high amplitude and low frequency. After completing the first 3 groups, participants moved on to learn icons with modified attributes (groups 4–9).

Notice also that when a participant is presented with an attribute-modified group, we attempt to provide as much confusion as possible by presenting as many of the previously experienced icons with the same base group as possible.

We have also counterbalanced which attribute modification is presented to participants first. Arrangements 1 and 3 experience second messages first (in batch 4), while arrangement 2 experiences low priority messages first.

3.3.6 Apparatus

As stated previously, we use the device described in Section 1.1.2 for administering haptic feedback, as well as acting as the computing platform for our experiment. Participants were instructed to hold the device in their non-dominant hand and interact with the device using a stylus.

Participants also wore passive noise reduction headphones rated at 25 DRR. These headphones successfully dampened most of the audio component coming from the devices, although some participants reported a slight audio component if they strained to hear it.

3.3.7 Procedure

Participants had the handheld devices located in their labs for a period of four weeks. During this period, they performed three 20 minute sessions per week on the device for a total of four hours (a total of 12 sessions). Three interviews were conducted throughout the course of the study: one at the beginning, one during the third week, and one at the conclusion of the experiment.

Participants were provided a manual for the experiment which they could consult if they ever experienced any confusion regarding the procedure. It is attached in Appendix B.

Each session lasts 20 minutes and is divided into 5 sections:

1. Login
2. Training
3. Quiz
4. Game
5. Questionnaire

Each session always begins with the Login, Training and Quiz for the participant's current batch. If the participant passes the quiz, then the remaining time is spent playing the game.

If at any point the experiment program determines that the user has completed a batch during the game, they will fill out the questionnaire and then begin training for the next batch. Sessions are capped at 20 minutes and may end during the game or during the training/quiz.

A flowchart that represents a user's possible procession through each of the procedural elements during each session is shown in Figure 3.3. It is helpful to have this on hand when reviewing the procedure.

The user has the ability to pause the experiment at any time by pressing one of the device's hardware buttons. This has been enabled in the event of a short interruption such as a phone call or short conversation. In the event of a pause, its duration is added to the length of the session to ensure that 20 minutes is spent interacting with the experiment.

Login

Before starting, the user must enter his/her login name. Once the name is accepted, they are given a prompt to put their headphones on (Figure 3.5). Headphones are passive noise reduction headphones rated at 25 DRR. The login screen is displayed in Figure 3.4

After closing the prompt the timer starts, and they are shown a window that describes and introduces their current batch. At this point they can press a button to 'Proceed to Training'. This screen is displayed in Figure 3.6.

Training

At the beginning of each session, a training screen (Figure 3.7) comes up for the user's current batch. The 7 new icons in their current batch are listed, and they can feel each icon as many times as they want. In addition, the old icons for the current batch are accessed by pressing a button, bringing them to a new screen (Figure 3.8). Users can also filter the icons by sender to feel all of the icons from each sender in isolation.

Once the participant feels that they have memorized the icons adequately, they can press the 'Start Quiz' button to test their learning.

Quiz

The quiz consists of 7 (only for batch 1) or 14 icons from the batch. All 7 new icons are present in the quiz as well as 7 randomly chosen icons from the older batches. The quiz presents an icon to participants one at a time and they can feel it as many times as they would like. They identify the icon by selecting from the list of 21 possible answers (7 distractors are not included in the quiz). The quiz screen is displayed in Figure 3.9

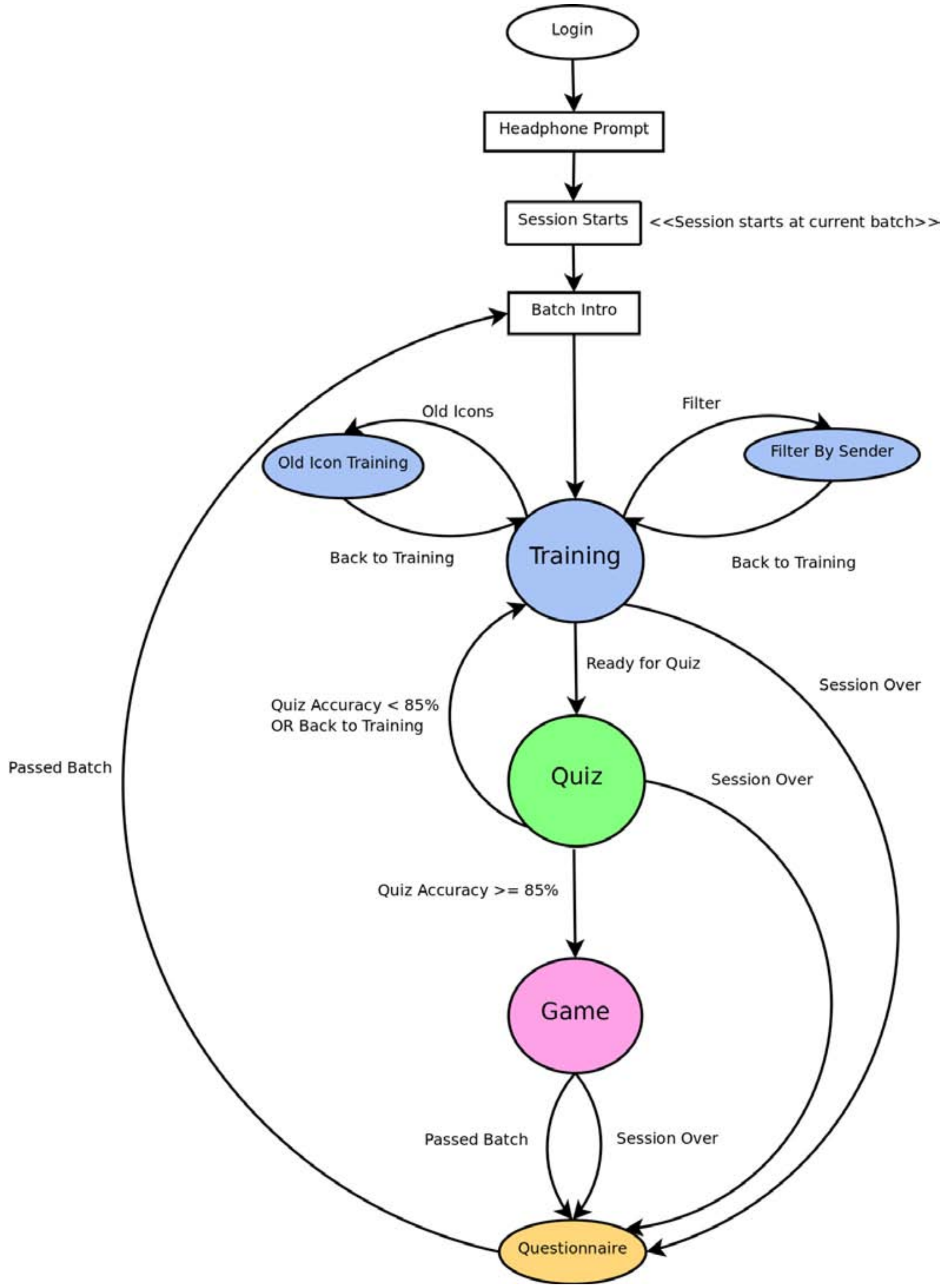

Figure 3.3: A flowchart that represents a user's possible procession through each of the procedural elements during each session.

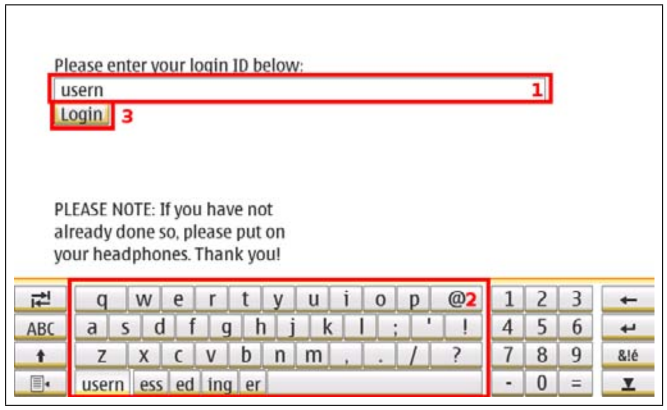

Figure 3.4: The Login screen. The numbers indicate the user's order of operations.

Figure 3.5: The prompt that reminds users to wear their headphones.

Once the participant achieves 85% or higher, they may go on to the Tetris game, otherwise they return to the training and must take the quiz again. Participants can exit the quiz and return back to training at any time.

We choose a performance level of 85% since this indicates a high level of recall in a training situation. We would like to simulate the participants having some experience with the icons before going into the more cognitively demanding and ecologically plausible situation.

Game: Foreground Task

The game task is simple, the participant must play Tetris [37] using only the stylus. They can move the block in any direction other than up, and they can flip the block by pressing the flip button at the bottom right of the screen (Figure 3.10.

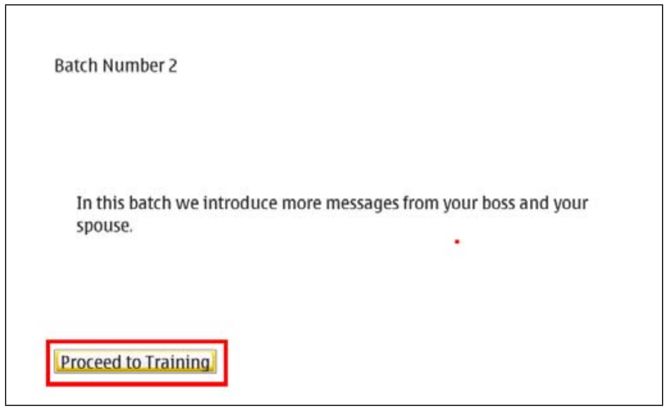

Figure 3.6: The Batch Introduction screen.

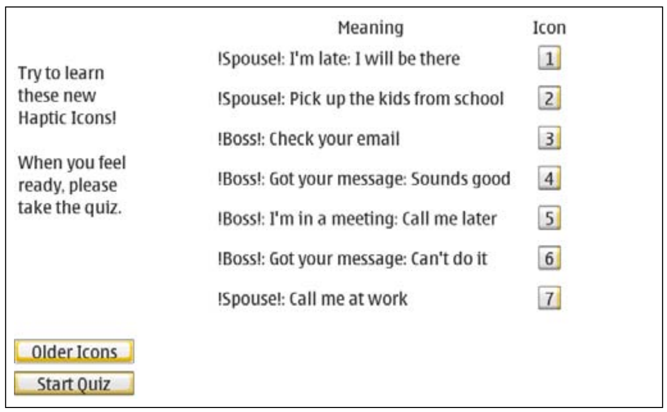

Figure 3.7: The Training screen.

- Points are awarded based on the Game Boy ™ system:
  - 1 lines = 10*level
  - 2 lines = 20*level
  - 3 lines = 40*level
  - 4 lines = 80*level
- Every 10 lines, the user goes up a level.
- If the user experiences a game over, they will reset to level 1, causing a penalty in their score.

Participants were encouraged to perform as well as possible during the Tetris game through verbal encouragement during interviews and monetary compensation 3.3.4.

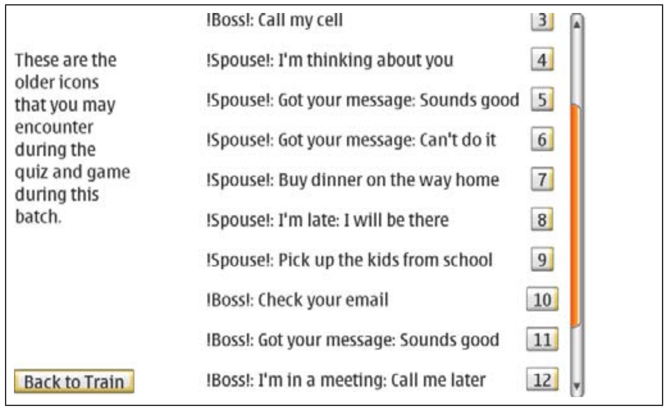

Figure 3.8: The Older Icons screen.

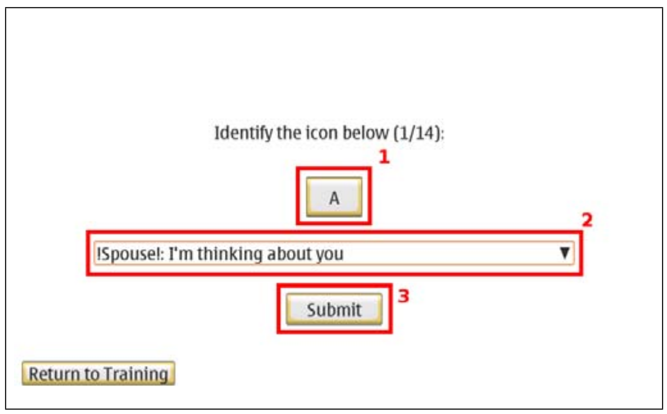

Figure 3.9: The Quiz screen.

Our experiment program is a heavily customized and instrumented version of the third-party and open source application Maemoblocks [15], which is an isomorphic version of the original Tetris game.

Game: Background Task

At 15-45 second intervals during the game, a haptic icon will play. This icon is chosen randomly from the pool of old and new icons in the current batch. The icon will play once a block drag is detected. Here is how the user responds:

1. Respond button comes up in response area (Figure 3.11).

2. User presses response button.

3. Game pauses.

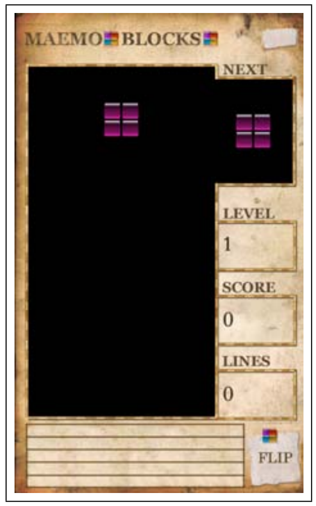

Figure 3.10: The foreground task: Tetris.

4. A prompt comes up telling user to identify haptic icon.

5. The user must enter a meaningful 4-character code that represents the message on the keyboard that is displayed on the screen (Figure 3.12). They are given a list of all of the codes (see Figures 3.1 and 3.2) on a separate piece of paper that they may refer to at any time. For the actual list, see Appendix A.

6. The user enters the code, then presses the 'Answer' button on the screen. Until they press the 'Answer' button, they can change their answer by deleting characters. Game play pauses for 2.5 seconds for the user to observe the result of their response. If the user selected the wrong answer, the correct answer will be displayed (Figure 3.13. If they select the right answer, they are told so.

7. Game play resumes.

If the participant does not press the Response button within 6 seconds, it is marked as a miss and the game continues.

The size of the interval was chosen such that we could test a significant number of responses in a short amount of time while keeping the notifications somewhat unpredictable.

A batch is considered complete if the following conditions are met:

- The user has responded correctly to all of the new icons at least once from the current batch during gameplay. This includes previous sessions.

- The user has scored 85% on the last 14 responses in the current session of gameplay. That means that a user must respond to at least 14 haptic icons during game play to advance to a new batch.

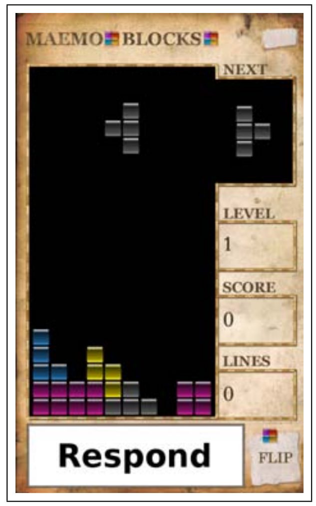

Figure 3.11: The Response prompt.

We chose the above criterion because it represents a breadth of knowledge over all of the new icons and high retention of the older icons. This high level of proficiency demonstrates that the participant has actually *learned* the icons.

Questionnaire

Once a participant completes a batch or a session, they are asked to fill out a questionnaire. Its details can be observed in Figure 3.14.

Interviews

Three interviews are conducted throughout the course of the study. An introductory interview is conducted before any sessions begin, an interim interview is conducted as close as possible to the beginning of the third week, and a debrief interview is conducted after all sessions have been completed.

Introductory Interview: During this interview, the device and experiment is introduced to the participant. It is in the form of a short tutorial which shows participants what to expect during the study. This is also where we receive consent to conduct the experiment.

Interim Interview: This interview is fairly free-form and asks participants about:

- Any mappings between vibrations and messages that seem semantically coherent.
- Strategies that the participant is using to remember icons.

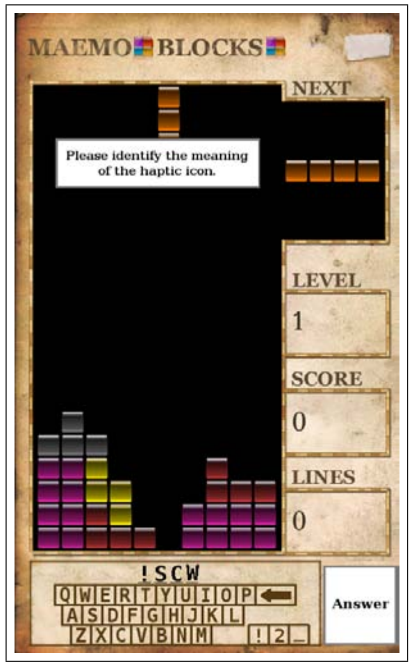

Figure 3.12: Entering a response using the on-screen keyboard.

- Thoughts about how the learning curve is going.
- Perceptual distinguishability of icons.
- Icons which are perceived as difficult.
- Icons which are perceived as easy.

Debrief Interview: This interview is identical to the Interim Interview, but a quiz which tests recall of the first 21 icons is administered. The participants are also given their compensation.

3.4 Results

In this section, we describe the results and statistics obtained from the experiment illustrated above. Since the amount of data collected is vast and varied, we will present them in separate subsections for readability. First, we will explain the data visualization techniques utilized in order to aid in the comprehension of results. In the next subsection (3.4.2), we will present results relating exclusively to the properties of the icons (without consideration of participant differences). Results relating to the

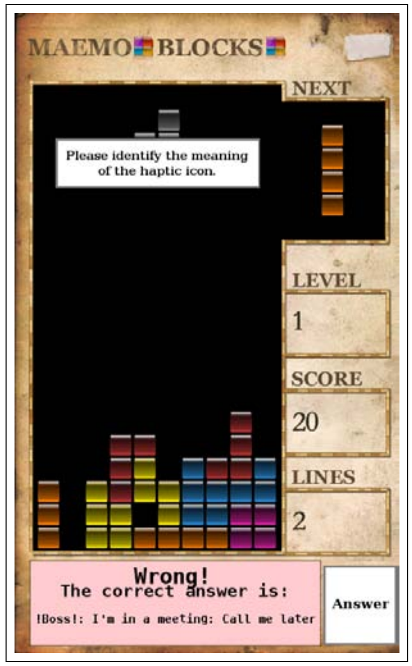

Figure 3.13: The answer observation time after submitting a response.

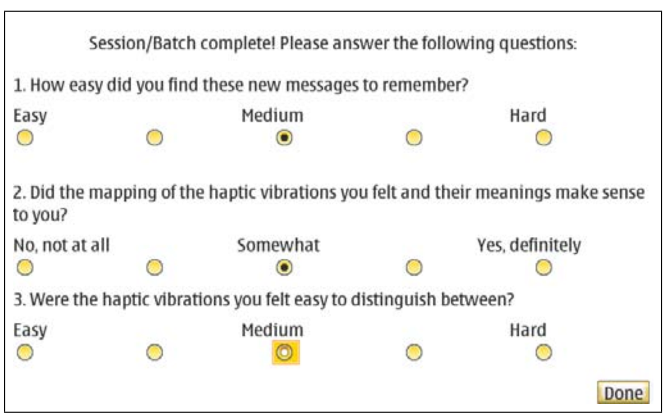

Figure 3.14: The Questionnaire.

differences and properties between different users and the learning strategies that they employed will be presented in the following subsection (3.4.3).

For the purpose of this section, all of the results displayed are based on performance during the icon identification task while playing the Tetris (Section 3.3.7) game unless stated otherwise.

An analysis and interpretation of the results will follow in Sections 3.5 and 3.6.

3.4.1 Explanation of Visualization Techniques

In Figure 3.15, we display an example of the visualization technique that we utilize to show statistically significant differences. The rectangles indicate homogeneous subsets that are not statistically significant at the 0.05 level using Tukey's HSD. Data points located outside of a given subset are significantly different from the points within that subset. The within-group significance level is indicated within each rectangle. For example, the mean for Group A is significantly different from the means for Groups C and D, but not Group B. Groups A and B are significantly different at 0.124, which is not lower than 0.05. The mean for Group B is significantly different from the mean for Group D, but not Groups A and C. Take note that the rectangles are hand-specified in order to encompass the related means in a visually aesthetic manner. As opposed to the confidence intervals in the plots, they are not exact specifications.

In Figure 3.16, we show the same relationships, but give a different visualization for the significant differences discovered. The arcs indicate a significant difference at the 0.05 level between the two connected means. For instance, Group A is significantly greater than Groups C and D. Group D is significantly lesser than Groups A and B. This visualization is used when the homogeneous subsets are complex, resulting in a plot that is difficult to read using the technique displayed in Figure 3.15 and describe above. We use a red arc to signify that a value is significantly larger than another, and blue arcs to signify that a value is significantly lower than another.

In Figure 3.17, we show an example of a visualization technique, for a single user (7), that we utilized in order to gain an understanding of how users experienced sessions. Each row of the figure represents one session, while each column represents a second in the session. The meaning of each colour represented is explained in the legend. Although we were not able to gain any vitally important insights about users by using these visualizations, they provided us with a better understanding of a typical session, as well as revealing some common patterns. We have attached these visualizations for all users in Appendix C in order to provide the reader with an in-depth understanding of user sessions.

3.4.2 Results Pertaining to Icon Differences

Icon accuracy

In Figure 3.18, we show the 95% confidence intervals for the accuracy on the first 21 (base rhythm) icons, averaged by participant. The mean icon accuracy is 79.2%, with a standard deviation of 11.8%

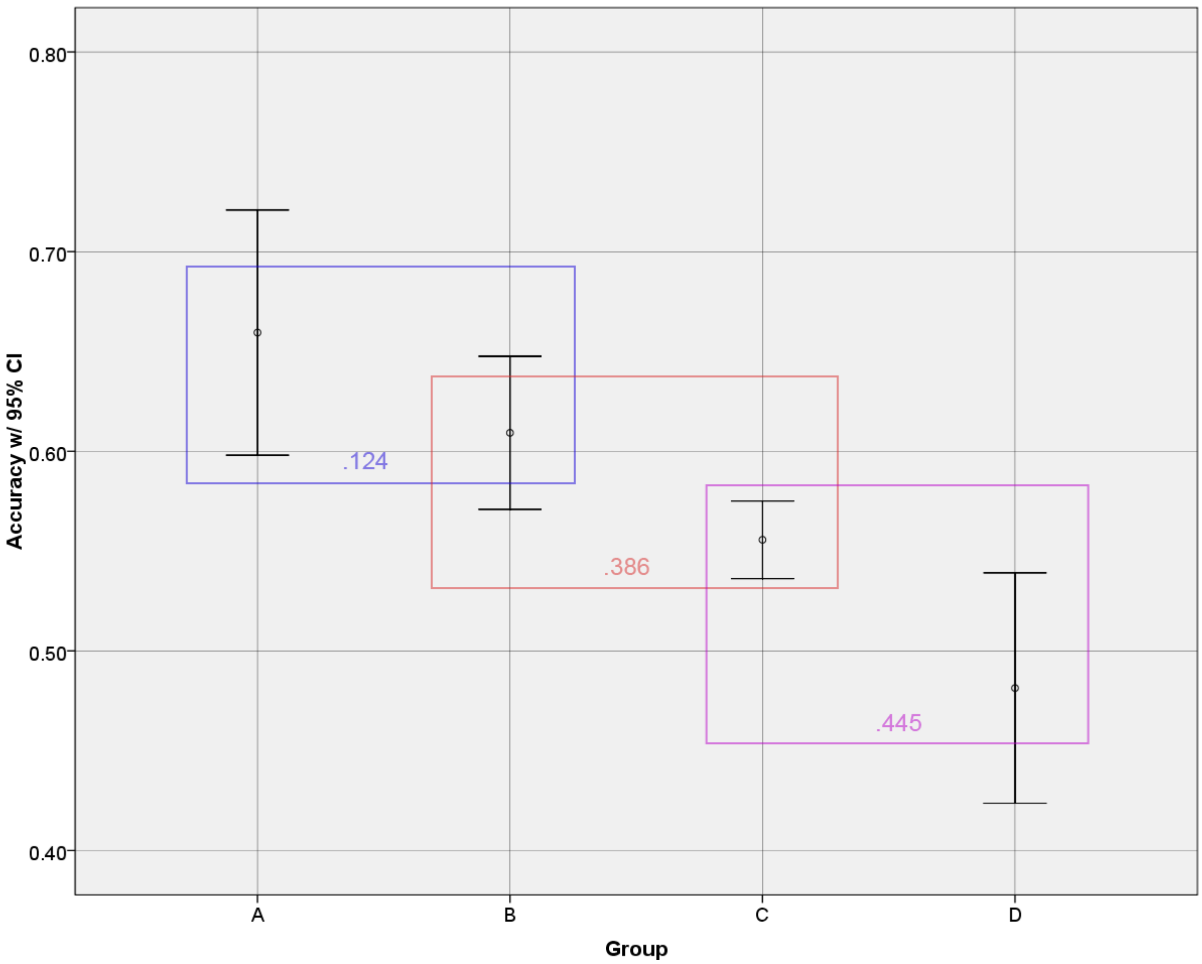

Figure 3.15: Example plot for the visualization of statistical significance utilized within this section. The rectangles indicate homogeneous subsets that are not significant at the 0.05 level using Tukey's HSD. Data points located outside of a given subset are significantly different from the points within that subset. The within-group significance level is indicated within each rectangle. The bars shown are 95% confidence intervals.

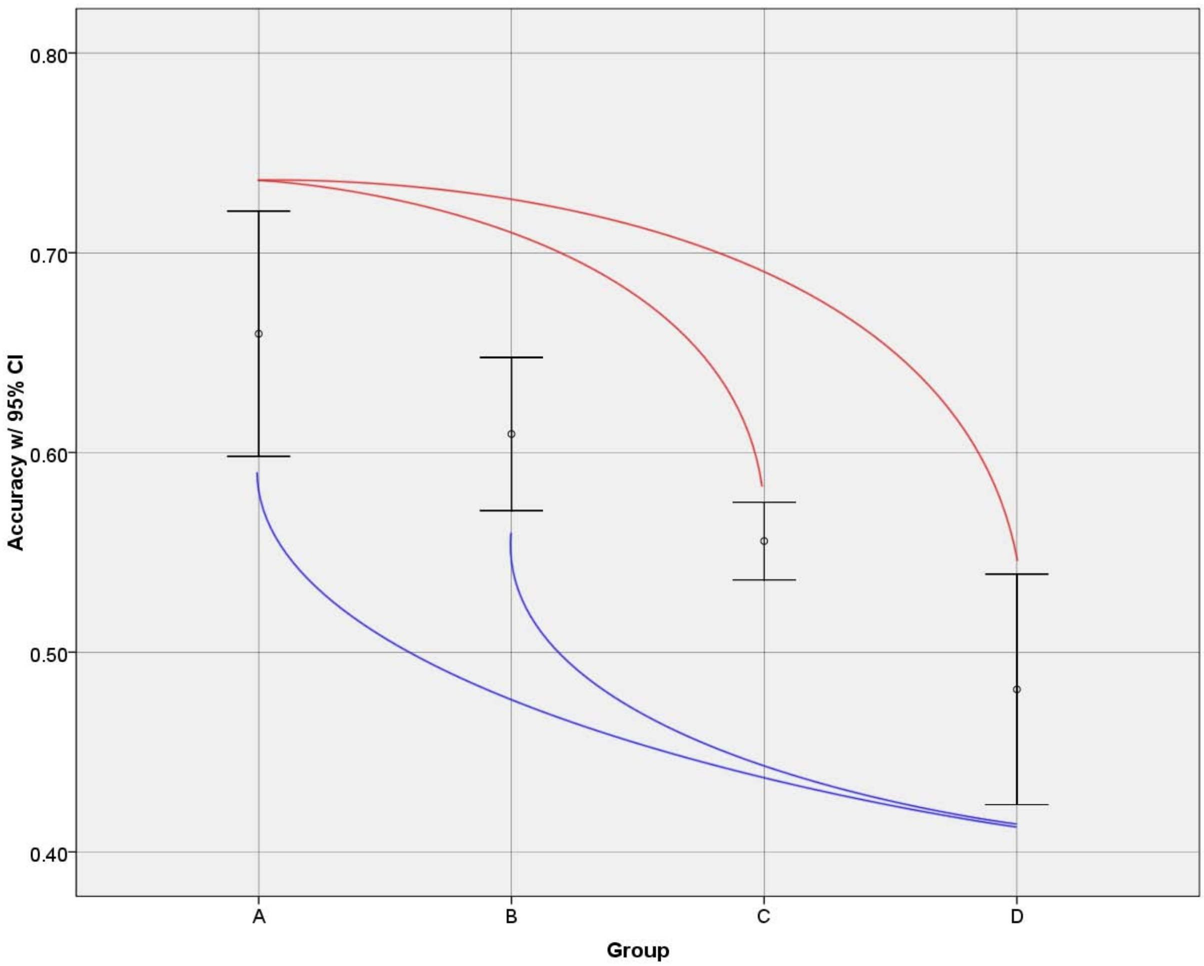

Figure 3.16: Second example plot for the visualization of statistical significance utilized within this section. The arcs indicate a significant difference at the 0.05 level using Tukey's HSD. We use a red arc to signify that a value is significantly larger than another, and blue arcs to signify that a value is significantly lower than another.

($N = 21$). Icon 21 had the minimum accuracy with 50.5% and icon 14 had the maximum with 99.1%.

In order to test for statistically significant differences between the accuracies attained for each icon, we conducted a One-Way Analysis of Variance (ANOVA) with multiple comparisons using Tukey's HSD test. The ANOVA found significant differences between the mean icon accuracies; $F(20,255) = 4.464$, $MS = 0.178$, $p < 0.05$, uses harmonic mean sample size of 13.1. The results of the multiple comparison tests are displayed in Figure 3.18 using the technique outlined in Section 3.4.1.

We can see that the mean accuracy attained for icon 14 is significantly higher than every icon with a mean accuracy less than or equal to 76.8% (icon 12). The mean accuracy for icon 21 is significantly lower than every icon with a mean accuracy greater than or equal to 80.1% (icon 8).

Learning curves for each icon

In Figure 3.19 we see a different metric for determining the difficulty of each icon. We have plotted the cumulative average of the performance on each individual icon. The highlighted icons are emphasized solely for the purpose of readability. The remaining icons will not be discussed and are displayed to show the distribution of the learning curves, as well as the final averages.

This cumulative average was calculated by summing the performance of all users for each progressive encounter of an icon and then taking the cumulative average of these progressive instances. Here, we have expressed this formally as:

$$P(i,j) = \frac{\sum_{x=1}^{j} \sum_{\forall u \in \mathbf{U}} c(u,i,x)}{\sum_{x=1}^{j} \sum_{\forall u \in \mathbf{U}} n(u,i,x)} \tag{3.1}$$

where $P(i,j)$ is the cumulative accuracy for icon i at instance j; $\mathbf{U}$ is the set of all users; $c(u,i,x)$ is the function that returns 1 if user u responded correctly on the xth instance encountering icon i, 0 otherwise; and $n(u,i,x)$ is the function that returns 1 if user u encountered icon i x number of times, 0 otherwise.

As one can see from Figure 3.19 the difficulty of learning each rhythmic icon varies considerably. Icons 14, 11, 15 and 1 appear to be quite easy to learn, whereas icons 21 and 18 are considerably more difficult than other icons.

Performance during the game vs. quiz

In order to determine whether there is a significant difference in the icon identification performance during the game versus during the quiz, we conducted a One-Way ANOVA. We found no significant difference between the mean response accuracy during the game and the mean accuracy during the quiz; $F(1,28) = 0.259$, $MS = 0.001$, $p = 0.61$.

Confusion between icons

In Figure 3.20, we show the confusion matrix for all of the responses on the first 21 icons. The results are summed over all users and all sessions, with each response instance acting as a data point; thus, these gross statistics do not take into account differences between users and their progression through the experiment. Due to this difference in the calculation method, there are inconsistencies between the averages presented in this confusion matrix and those presented in Figure 3.18. Those accuracies represent means calculated for each individual user and then averaged, rather than a gross sum.

The target icons are listed along the rows of the matrix, while the user's response is listed along the columns. For instance, users mistakenly identified icon 8 as icon 17 11% of the time. Icons within each thick blue box belong to the same group of icons (7 per group).

Confusion between senders

In Figure 3.21, we plot the confusion between the icons belonging to each sender. We plot 95% confidence intervals for the proportion of responses answered correctly for each sender; the proportion of responses where the target sender was *Boss* and the participant answered incorrectly but the sender was identified correctly (mistook a *Boss* icon for another *Boss* icon), and likewise for *Spouse*; and the proportion of responses where the target sender was identified incorrectly (mistook a *Boss* icon for a *Spouse* icon and vice-versa).

In order to test for statistically significant differences between these classes of responses, we conducted a One-Way ANOVA with multiple comparisons using Tukey's HSD test. The ANOVA found significant differences between the mean proportion of responses belonging to each response class; $F(5,84) = 245.773$, $MS = 1.823$, $p < 0.05$. In Figure 3.21, we can observe that there were significant differences between each type of response class, but the identity of the sender had no effect. In other words, there were significant differences between the proportion of responses answered correctly (73.3% for *Boss*, 80.4% for *Spouse*), the proportion of responses where the sender was identified correctly (24.5% for *Boss*, 16.9% for *Spouse*), and the proportion of responses where the sender was identified incorrectly (2.1% for *Boss*, 2.7% for *Spouse*). Icons where *Boss* was the sender were not significantly more difficult to identify than icons where *Spouse* was the sender. To summarize, participants identified the sender correctly approximately 97% of the time, but they misidentified the exact message 17-25% of the time.

Confusion between modified rhythms and their base

In Figure 3.22 we plot the confusion between the base icons (batches 1–3), and the icons where attributes have been added (low amplitude and high frequency rhythms; batches 4 onwards). We plot the proportion of responses answered correctly for each class (base, attribute); the proportion of responses where the participant mistook a base icon for its modified counterpart; and the proportion of responses where the participant mistook an icon with a modified attribute for its base icon. Thus, we exclude errors where users mistook an icon for an icon with different semantic content. Here, we only use results collected from the six users who completed batch 3.

To test for the presence of statistically significant differences, we conducted a One-Way ANOVA with multiple comparisons (Tukey's HSD). We found significant differences between the proportion of responses for each response class; $F(3,20) = 267.739$, $MS = 1.332$, $p < 0.05$, uses harmonic mean sample size of 6.0. In Figure 3.22, we can observe that participants very rarely mistook a modified icon for its base (1.8%), and never mistook a base icon for one with modified attributes.

Learning techniques employed by participants, sorted by icon

In Table 3.6, we display a list of mnemonic learning techniques that participants self-reported, sorted by icon. As one can see, there is very little consistency to the techniques used, and none are truly metaphorical. At best, *some* are semi-abstract and most are purely mnemonic.

Most participants also reported that most of their effort was devoted to distinguishing the various stimuli, using the meanings as a 'name' for each stimulus. In other words, they reported that the actual meaning had very little bearing on whether or not they could learn an icon. However, they reported that the fact that messages had been grouped by sender based on the stimulus' 'evenness' was extremely helpful.

Participants also reported that extremely distinctive stimuli were easier to learn. In this context, they made no mention of the meaning of the icon.

Many participants also related the rapidity (or density of notes) in a stimulus to a feeling of urgency. These notes would impose a feeling of urgency to each meaning, thus aiding in learning. Note that this urgency assignment is completely arbitrary and not related to our meaning assignment strategy (Section 3.1.2). One participant even tried to force himself to dislike the feeling of all of the rhythms associated with negative messages. Conversely, he tried to train himself to like the feeling of the rhythms with positive messages.

A few participants also reported that they would first try to vocalize the rhythm and associate it somehow with a word in the meaning.

Relation to previously obtained perceptual map

In Figure 3.23, we show a reproduction of Ternes and MacLean's MDS results for the stimuli of this icon set [43].

In Figure 3.24, we display the difference between Ternes and MacLean's stimulus similarity statistics obtained through their sorting study [43] and the confusion matrix obtained through our identification task (Figure 3.20). Since they publish dissimilarity results at a different scale, we normalize their results between 0 and 1, and subtract that number from 1; thus calculating an experimentally determined

Icon	User	Stimulus Description (Meaning Relationship)
1	4	Relates stimulus to something that his wife does to annoy him (N/A)
1	1	Stimulus feels urgent due to its rapidity (Message could be urgent)
1, 14	3	Stimuli 1 and 14 feel similar (Have similar meanings – request to call)
2, 3	3, 6, 9	Number of notes ('Office' has more letters than 'Cell')
2, 3	7	Stimuli 2 and 3 feel similar (Have similar meanings – request to call)
4	6, 13	Feels like a heartbeat (Message is sentimental)
6	1	Feels negative (Message is negative)
6	14	Reminded of a busy signal (Message is negative)
6	13	Visualizes a woman shaking her head (Message conveys the meaning 'no')
8	13	Feels slow (Message relates to tardiness)
8	3	Feels like 4 words [1] (I-will-be-there)
9	4	Reminded of running track at school (Message contains the word 'school')
10	9, 14	Vocalization ('Ee-mail')
11	1, 3, 4, 7, 9, 10, 11, 13, 14, 15	Rhythm feels happy (Conveys a positive message)
11	2	Reminded of prototypical rhythm for Morse code – distinctive rhythm (N/A)
12	6, 7	Vocalization ('LA-ter')
12	11	Morse code for 'm' (Message relates to a meeting)
13	15	Describes stimulus as 'a let down' (Message is negative)
14	10	Reminded of a telephone ringing (Message relates to telephoning)
14	14	Feels urgent (Message can be construed as urgent)
15	2	Feels like a burst-fire machine gun (N/A)
15	4	Feels like a gallop (Relates to bringing something)
15	14	Feels urgent (Message can convey a sense of urgency)
19	10	Stressful rhythm (Message reminds user of work to do)
20	8	Rhythm feels military – like cavalry (N/A)
20	9, 14	Vocalization ('CAN-celled, CAN-celled')
20	1	Rhythm feels happy (Conveys a message that causes happiness)
-	2, 9, 11, 12, 15	Subdivides similar stimuli to partition learning (N/A)
-	4	Visualizes spouse performing rhythms haptically (Visualizes spouse communicating message)

Table 3.6: List of mnemonics employed by participants, sorted by icon. It is helpful to observe Figure 3.1 while examining this table.

similarity statistic. Then, we take the value of the difference between their similarity results and our confusion matrix. By using this difference measure, we can identify discrepancies between the predicted perceptual similarity and actual confusion in deployment. *Large values indicate where the MDS analysis predicted perceptual similarity, but it was not observed as confusion during our study.* Values greater than 0.5 are highlighted in red. There were no instances where the MDS analysis predicted perceptual confusion and none was observed (negative values).

3.4.3 Results Pertaining to Participant Differences

Icon retention

In Figure 3.25, we can observe that people correctly identified, on average, about 16.4 icons under no workload about one week after twelve 20 minute sessions over four weeks.

The exact number of icons learned is difficult to express from experimental results since participants were constantly learning. Participants experienced between 14 and 42 icons, with overall cumulative average accuracy ranging between 63% to 89%.

To identify if there was a statistically significant difference between the mean response accuracy for old versus new icons, we conducted a One-Way ANOVA. We define *old icons* as icons belonging to a batch that the participant has already completed. Conversely, *new icons* are icons belonging to the participant's current batch. We found no significant difference between the mean accuracy on the old icons (76.2%) and the mean accuracy on new icons (76.3%); $F(1,28) = 0.001$, $MS = 0.000$, $p = 0.979$.

We also conducted a One-Way ANOVA to check for statistically significant differences between the mean response accuracy for old versus new icons during the quiz. Recall that, during the quiz, the user must identify a subset of 7 icons from previous batches (Section 3.3.7). We found that the mean response accuracy for old icons (81%) was significantly higher than the mean response accuracy for new icons (75%); $F(1,28) = 4.190$, $MS = 0.029$, $p = 0.050$.

Batch progression

For a general idea of what the learning curve was for haptic icon learning, we examine Figure 3.26. This figure shows how many sessions it took users, on average, to complete each batch.

Recall that past batch 3, icons are simply modified versions of previously learned rhythmic bases. To test for statistically significant differences between the number of sessions needed to complete each batch, we conducted a One-Way ANOVA with multiple comparisons using Tukey's HSD. We observe that batch 2 took significantly longer than batches 1, 3 and 4 to complete; $F(3,35) = 14.724$, $MS = 22.870$, $p < 0.05$, uses harmonic mean sample size of 7.2.

Another view of this phenomenon can be seen in Figure 3.27. As you can see, most participants required about 2 or 3 sessions to complete batches 1 and 3. However, most people required about 5 sessions or above to complete batch 2 – one person required 9 sessions.

As you can see from Figure 3.26, only 6 out of 15 participants finished batch 3, therefore it is difficult to determine if the least proficient users bring the average up on batch 2, causing this 'bump' in the learning curve. In Figure 3.28, we show the average number of sessions required to complete a batch for the top 6 users.

As you can see, the same learning pattern exists, although the 'bump' is slightly less prominent; F(3,18) = 12.923, MS = 7.957, $p < 0.05$, $N = 6$. Batch 1 required less sessions to complete. The exact same pattern as Figure 3.28 exists for the top 4 and top 2 learners, but their plots are not shown in the interest of space.

Table 3.7 shows the number of batches completed by each user.

User	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Batches Completed	2	3	2	5	5	2	2	1	4	2	2	2	3	2	4

Table 3.7: The number of batches completed by each user.

Figure 5.7 shows the questionnaire results, averaged by user, over all 12 sessions of the study. Recall that the questionnaire questions are as follows:

1. How easy did you find the new messages to remember? (Easy – Hard)
2. Did the mapping of the haptic vibrations you felt and their meanings make sense to you? (No – Yes)
3. Were the haptic vibrations you felt easy to distinguish between? (Easy – Hard)

From this plot, we can see that there is a slight downward trend in the subjective opinions of how easy participants find new messages to remember. There is no apparent trend for whether participants throught that the mapping between the stimulus and the meaning made sense. We observe a slight 'U-shaped' trend for how easy participants thought it was to distinguish between the vibrations. However, in general, the responses are quite sporadic and we do not believe that there are any novel revelations to be gained from these questionnaire results, thus we will not discuss them further.

Learning curves for each participant

Figure 3.30 shows the cumulative average accuracy for each participant over each response instance. The highlighted users are the subject of discussion and are emphasized for the purposes of readability. The curves for the remaining users are shown to display the range of learning abilities and final cumulative accuracies. Take note that these curves represent a cumulative average, therefore there is a great deal of inertia in the exact value of average – individual responses have less of an effect when they are averaged among many other responses as opposed to averaged among few responses.

As one can see from Figure 3.30, there is much variation in final accuracy between users; however, the distribution is fairly even in the range of 63% to 87%. User 8 is also notable since he/she completes the study with the lowest cumulative average performance: 63%.

It is also important to note the effect on learning once a user has completed a batch. In many cases, performance drops after a new batch is reached. This phenomenon is most easily exemplified by User 9's learning curve.

Effect of counterbalancing

To test if there was a statistically significant effect of the batch presentation ordering on account of the counterbalancing, another One-Way ANOVA with multiple comparisons was conducted. No significant difference between the mean response accuracies for each counterbalancing arrangement was found; $F(2,12) = 0.197$, $MS = 0.001$, $p = 0.824$, $N = 5$.

Demographics

We did not find any significant differences between any demographic groups relating to age, gender, musical experience, country of origin or self-reported sense of rhythm.

User groups

There are no obvious segmentations between user types based solely on accuracy, due to the even spread of the learning curves notable in Figure 3.30. Despite this fact, we will divide the users into 3 groups to understand differences between proficient, average and below average participants. Here we define proficient users as participants who maintained a higher level of accuracy than other participants. In all cases, a higher accuracy resulted in a faster progression through the icon set. For the sake of analysis, we will split the participants into the following groups:

Proficient: 80% accuracy and above (4, 5, 9, 15; N = 4).

Average: Between 71% and 79% accuracy (1, 3, 7, 10, 12, 13, 14; N = 7).

Below Average: 70% accuracy and below (2, 6, 8, 11; N = 4).

A One-Way ANOVA with multiple comparisons using Tukey's HSD reports significant differences between the mean accuracies of all user groups; $F(2,12) = 35.408$, $MS = 0.034$, $p < 0.05$, harmonic mean sample size of 4.7.

Performance of user groups

In Figure 3.32, we show the mean accuracy on the 7 easiest, middle and most difficult icons for the user groups specified, averaged by icon and user.

From Figure 3.32, we can observe that all user groups are accurate at identifying the easiest 7 icons. However, a pattern begins to emerge for the more difficult icons. For this middle 7 icons, performance for the Average participants drops slightly from 87% to 75%, but accuracy for the below average users drops drastically from 82% to 66%. For the most difficult icons, the performance of the average users drops to 64%, and the performance for the below average participants drops to 53%. Furthermore, the proficient users do not experience a drop in accuracy except for the most difficult icons, where performance drops from 91% to 77%.

Examining statistical significance, we see that there are no significant differences found in the mean performance for all 3 icon groups for the Proficient user group; and a significant drop in performance between the easiest icon group and both more difficult icon groups, but no significant difference between the middle and difficult icon groups for the Average and Below Average user groups; $F(8,278) = 8.680$, $MS = 0.510$, $p < 0.05$.

3.5 Discussion of Results Pertaining to Icon Differences

In this section, we will discuss the results of the experiment pertaining to the icons in order to understand how icons are perceived and learned outside of the context of individual differences, allowing for more generalizable findings. We aim to provide insights related to how haptic icons can be designed so that learning and recognition is as effective as possible. In the following section (3.6), we will discuss the results pertaining to inter-participant differences.

The discussion is organized by using the research questions stated in Section 3.3.1 so that we can discuss our findings within the context of each question.

3.5.1 What Makes Icons Difficult to Learn?

By understanding what makes icons difficult to learn, we can avoid designing icons with these properties in order to enable efficient learning. In addition, we can examine whether most difficulties were caused primarily by the perceptual properties of the icons or their semantic properties. An understanding of this distinction can aid in determining whether a perceptual approach or a metaphorical approach is more appropriate for icon design.

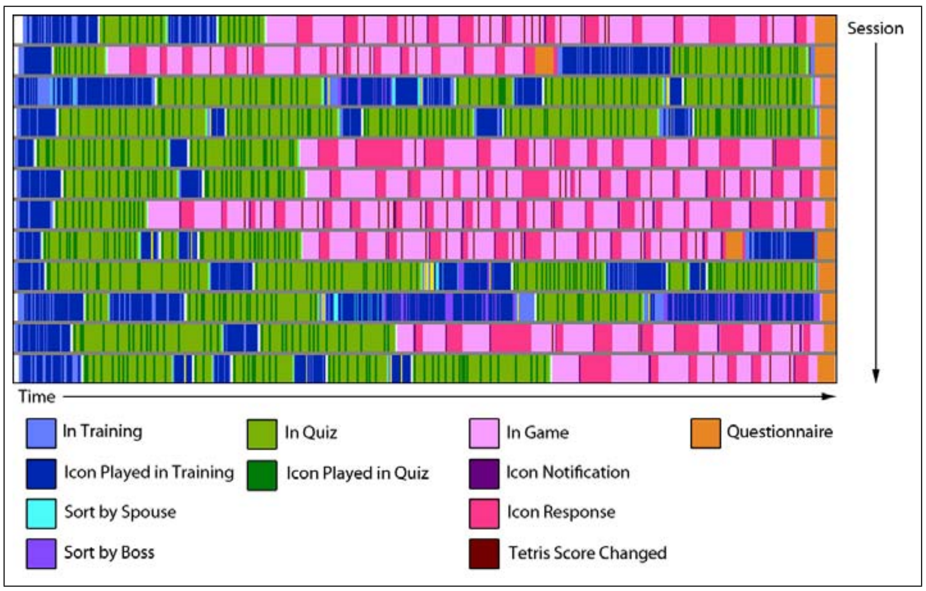

Figure 3.17: Example visualization of all sessions for user 7. Each row of the figure represents one session, while each column represents a second in the session. The meaning of each colour represented is explained in the legend.

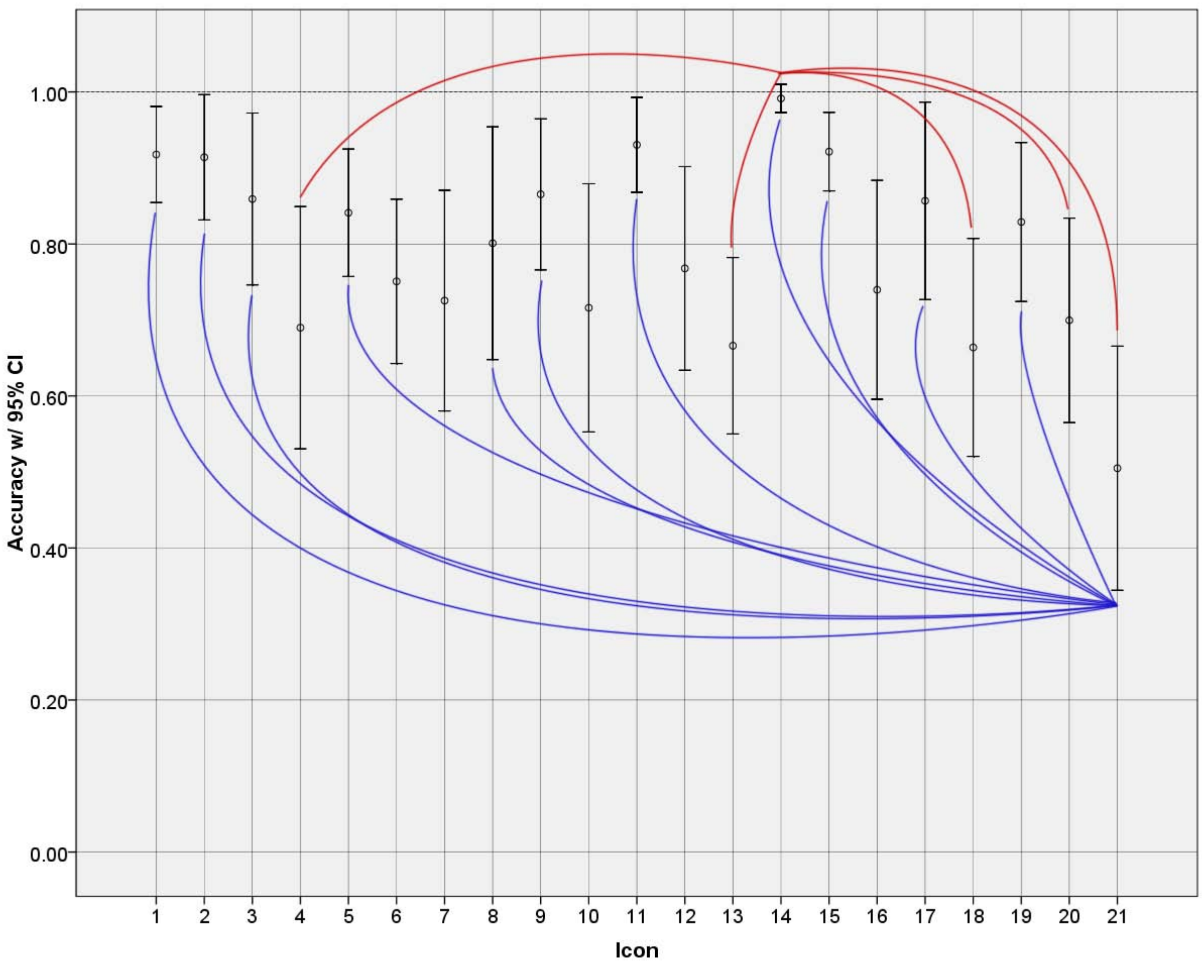

Figure 3.18: 95 % confidence intervals for accuracy on base 21 icons, averaged by user. Tukey's HSD test uses harmonic mean sample size of 13.1.

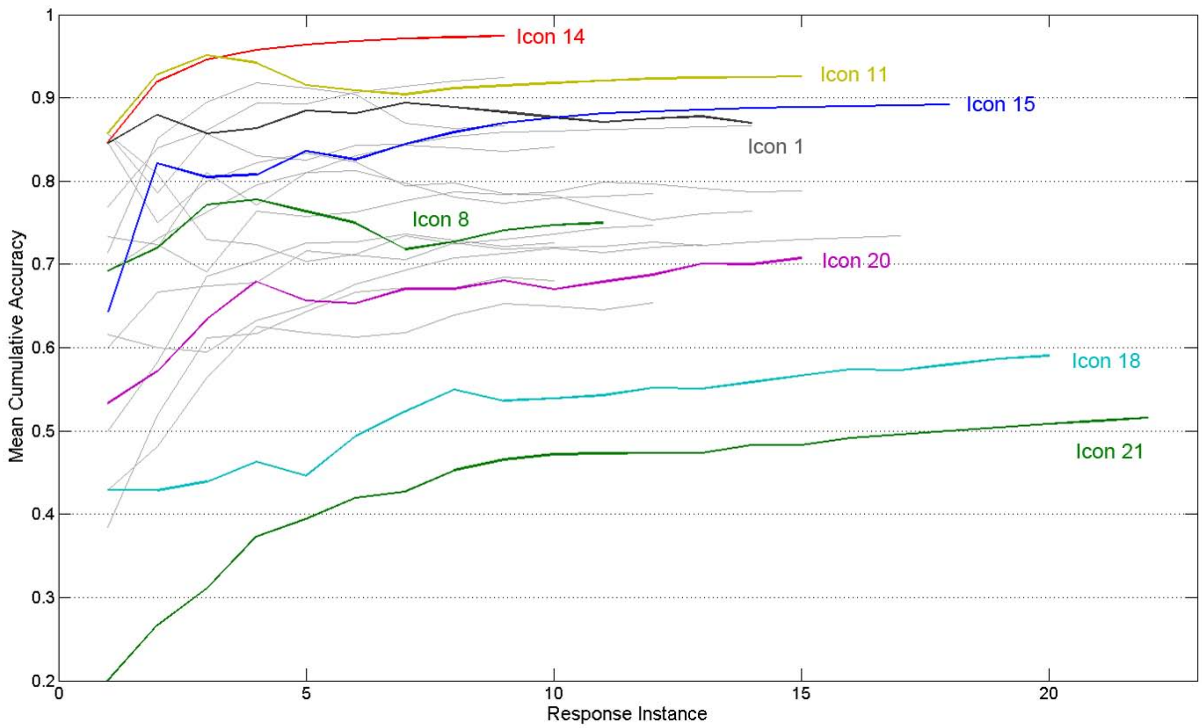

Figure 3.19: Learning curves for individual icons. The cumulative performance of individual icons for all users during the game is plotted. Performance for a given icon instance is averaged over all participants who encountered the icon at least that many times. Highlighted icons are subjects of discussion and are emphasized for the purposes of readability.

Icon Responded

Target Icon	N	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
1	92	87%			1%				1%	3%					7%							1%
2	85		88%	5%								1%	1%	2%		1%			1%			
3	92		7%	77%	2%	3%							1%	1%				8%				1%
4	75	9%	1%		68%	8%				7%							4%	3%				
5	85	1%			5%	84%	4%		4%			1%						2%				
6	84					2%	74%	19%	2%	1%				1%								
7	62					2%	6%	73%	18%												2%	
8	83	1%				5%		5%	76%	2%								11%				
9	79	1%			1%		1%	3%	5%	85%	1%		1%				1%					
10	94						1%				72%	6%	5%	12%					3%			
11	95					1%					5%	92%		1%						1%		
12	81			1%			1%				9%	2%	73%	9%	2%						2%	
13	83										10%	1%	18%	65%							2%	4%
14	76														100%							
15	112		2%													89%	1%	3%	1%		3%	2%
16	113	1%	1%	2%	4%				4%	1%						4%	73%	7%			3%	1%
17	89			3%		4%											4%	88%				
18	122		2%								1%			1%		7%			59%	23%	2%	6%
19	123															6%			14%	79%	2%	
20	126												5%	3%		2%	1%		7%	1%	71%	10%
21	127		1%								1%			5%		2%			25%	6%	9%	52%

Figure 3.20: Confusion matrix for all 21 rhythms for responses during the game. Calculated by summing over all response instances, irrespective of the user. Icons within each thick blue box belong the same group of icons (7 per group).

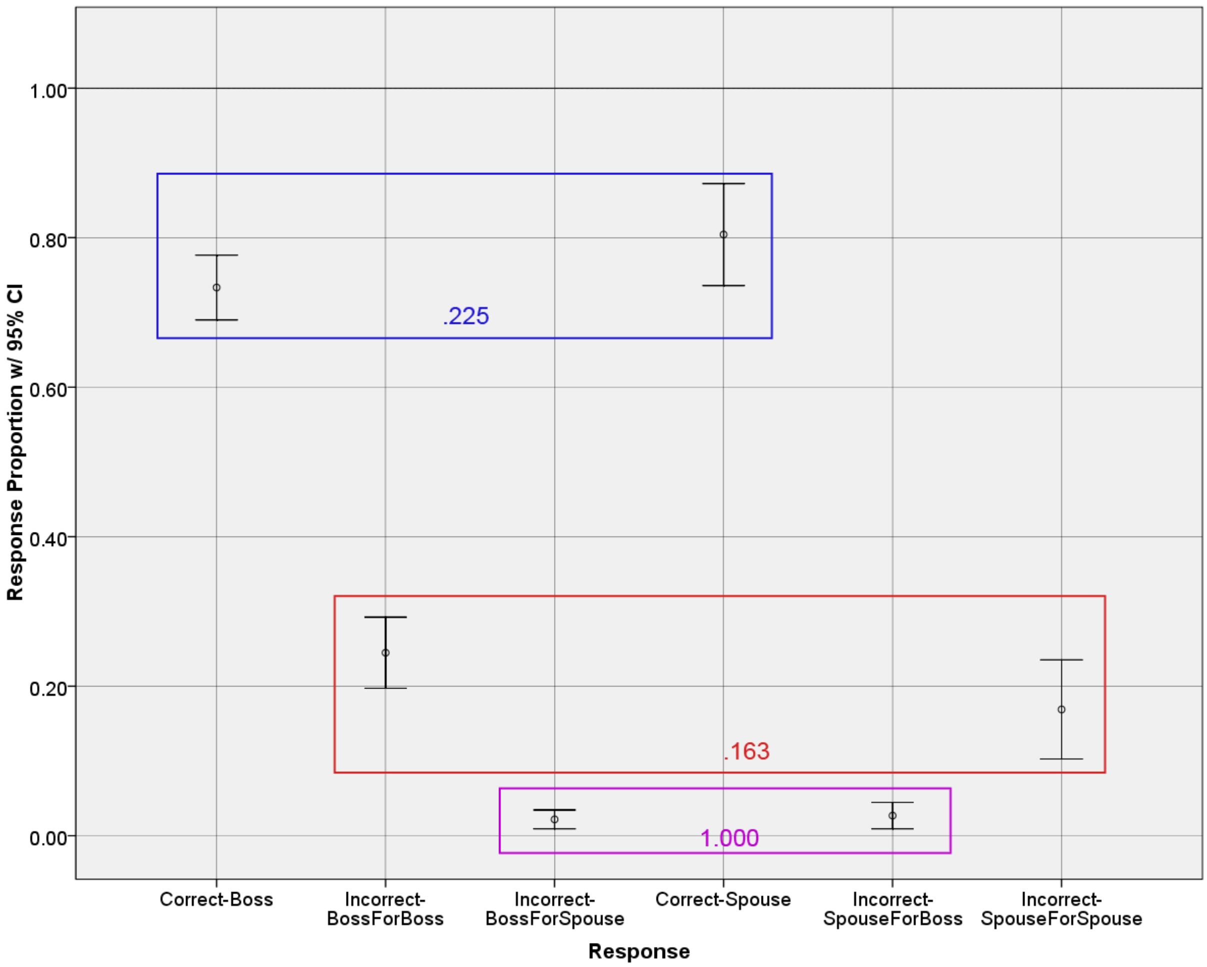

Figure 3.21: Proportion of confusion between icons belonging to each sender. We plot the proportion of responses answered correctly; the proportion of responses answered incorrectly, but the sender was identified correctly; and the proportion of responses where the sender was identified incorrectly.

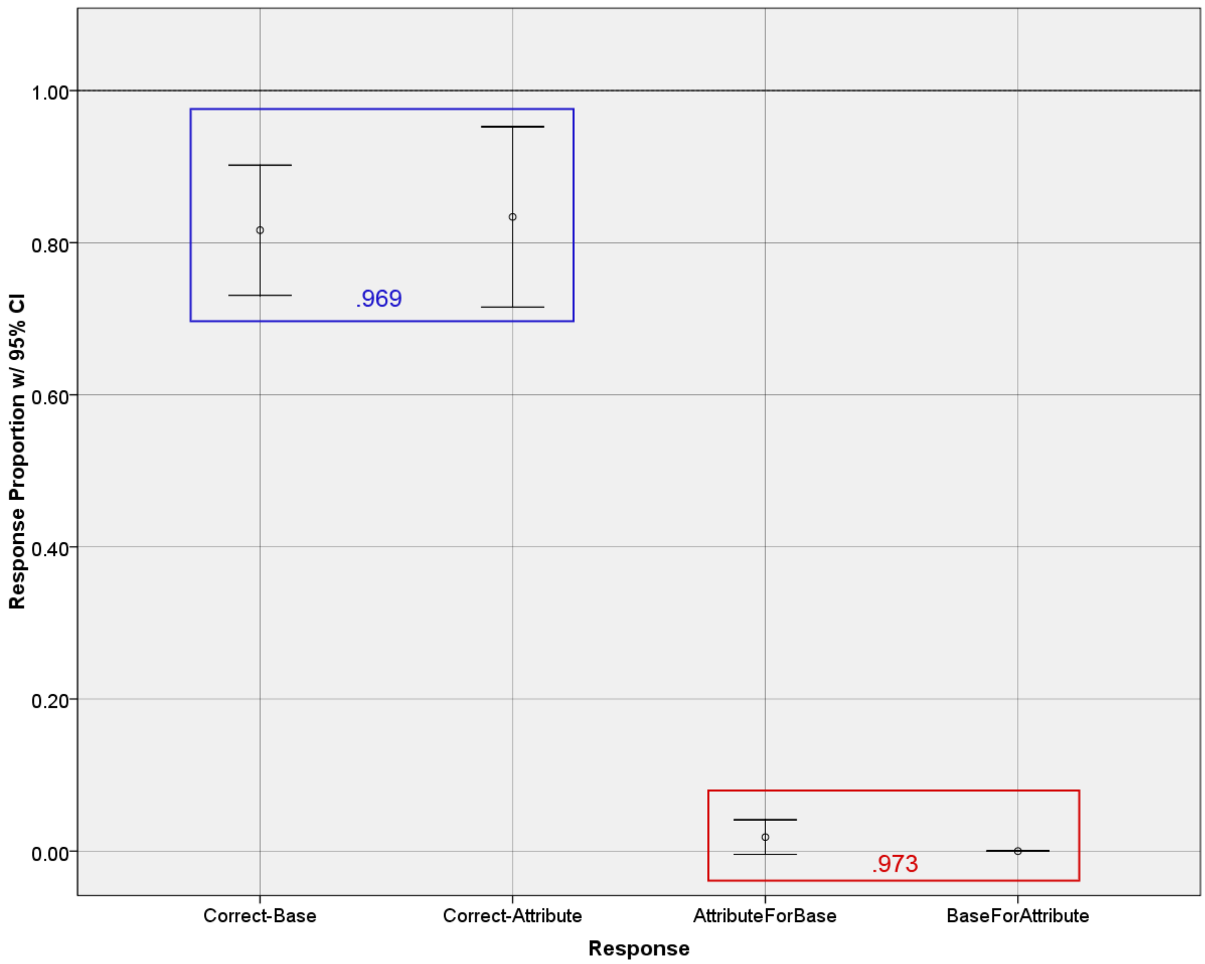

Figure 3.22: Proportion of confusion between base icons and their counterpart with added attributes for users that completed batch 3. We plot the proportion of responses answered correctly for each class (base, attribute); the proportion of responses where the participant mistook a base icon for its modified counterpart; and the proportion of responses where the participant mistook an icon with a modified attribute for its base icon. We exclude errors where users mistook an icon for an icon with different semantic content Tukey's HSD test uses harmonic mean sample size of 6.0.

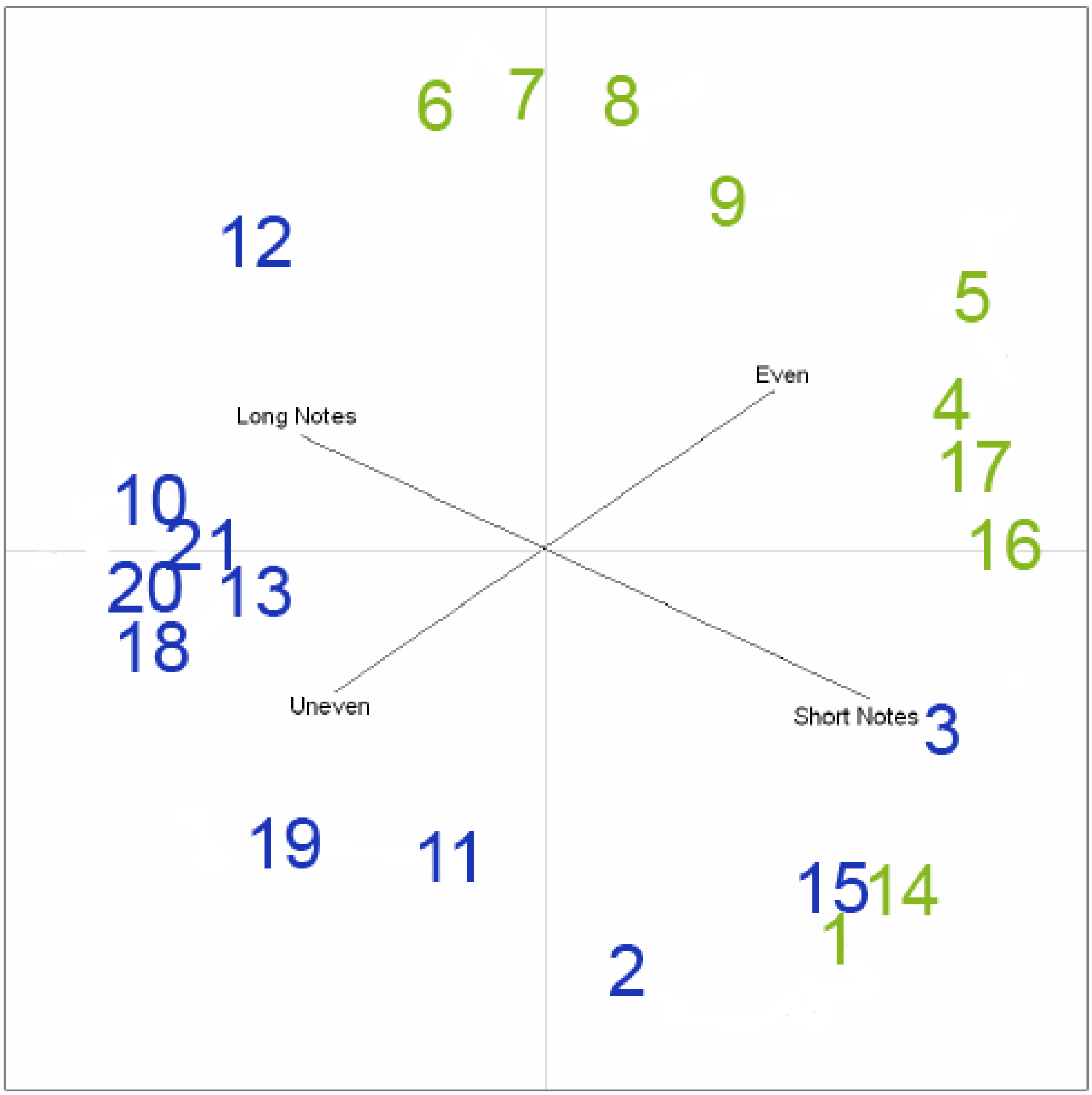

Figure 3.23: Interpretation of Ternes and MacLean's subset MDS results for the rhythmic stimuli of our icon set. Icons from *Boss* are coloured blue and icons from *Spouse* are coloured green.

Target Icon \ Response Icon	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
1	0.1	0.3	0.2	0.3	0.3	0.1	0.1	0.1	0	0.1	0.1	0.2	0.1	0.3	0.4	**0.5**	0.3	0.1	0	0.1	0.3
2	0.3	0.1	0.4	0.3	0.3	0.1	0	0.2	0	0.1	0.1	0.1	0.1	0	**0.6**	0.4	0.4	0.1	0.4	0.1	0.1
3	0.2	0.3	0.2	0.3	0.4			0.2		0.2	0.1	0	0.1	0	**0.6**	0.3	**0.6**	0.1	0.2	0.1	0.3
4	0.2	0.3	0.3	0.3	**0.6**	0.1	0.1	0.2	0.1	0		0.2	0	0	0.3	**0.5**	0.4		0	0.1	0.1
5	0.3	0.3	0.4	**0.6**	0.2	0.1	0.1	0.4	0		0	0.1	0	0.2	0.3	**0.6**	**0.5**		0	0.1	0.1
6	0.1	0.1		0.1	0.1	0.3	0.4	0.3	0.4	0.1	0.1	0.3	0.1			0	0	0.1	0.2	0.2	0.2
7	0.1	0		0.1	0.1	0.5	0.3	0.4	0.4	0.1	0.3	0.3	0.2				0	0.1	0.1	0.2	0.2
8	0.1	0.2	0.2	0.2	0.3	0.4	0.5	0.2	0.2	0		0	0.1	0.2	0.1	0.3	0.2		0	0	0.1
9	0.1	0		0.1	0	0.4	0.4	0.2	0.2	0.2	0.4	0.2	0.1	0	0	0.1	0	0.3	0.1	0.3	0.3
10	0.1	0.1	0.2	0		0.1	0.1	0	0.2	0.3	**0.5**	0.3	**0.6**		0.2	0	0.1	0.3	0.1	0.4	0.5
11	0.1	0.1	0.1		0	0.1	0.3		0.4	**0.6**	0.1	0.5	**0.6**		0.1	0		**0.5**	0.2	**0.6**	**0.7**
12	0.2	0.1	0	0.2	0.1	0.3	0.3	0	0.2	0.3	0.4	0.3	0.3	0	0.1	0.1	0	0.3	0.3	0.4	**0.6**
13	0.1	0.2	0.1	0	0	0.1	0.2	0.1	0.1	**0.7**	**0.6**	0.2	0.3	0.1	0.1	0	0.1	0.4	0.3	0.4	0.4
14	0.4	0	0	0	0.2			0.2	0			0	0.1		0.1	0.3	0.1	0.1	0.1	0	
15	0.4	**0.6**	**0.6**	0.3	0.3			0.1	0	0.2	0.1	0.1	0.1	0.1	0.1	0.4	0.4	0.1	0.3	0.1	0.1
16	**0.5**	0.4	0.3	**0.5**	**0.6**	0		0.3	0.1	0	0	0.1	0	0.3	0.4	0.3	0.4	0	0		0
17	0.3	0.4	**0.7**	0.5	**0.5**	0	0	0.3	0	0.1		0	0.1	0.1	0.5	0.4	0.1		0	0	0.2
18	0.1	0.1	0.1			0.1	0.1		0.3	0.3	**0.5**	0.3	0.4	0.1	0.1	0		0.4	0.1	**0.6**	0.5
19	0	0.4	0.2	0	0	0.2	0.1	0	0.1	0.1	0.2	0.3	0.3	0.1	0.2	0	0		0.2	0.1	0.1
20	0.1	0.1	0.1	0.1	0.1	0.2	0.2	0	0.3	0.4	**0.6**	0.4	0.3	0	0.1	0	0	**0.5**	0.1	0.3	0.5
21	0.3	0.1	0.3	0.1	0.1	0.2	0.2	0.1	0.3	0.5	**0.7**	**0.6**	0.4		0.1	0.1	0.2	0.3	0	**0.5**	0.5

Figure 3.24: Difference between the normalized similarity statistics experimentally determined by Ternes and MacLean [43] and the confusion matrix obtained through our icon identification task. High values indicate where the MDS analysis predicted perceptual similarity, but it was not observed as confusion during our study.

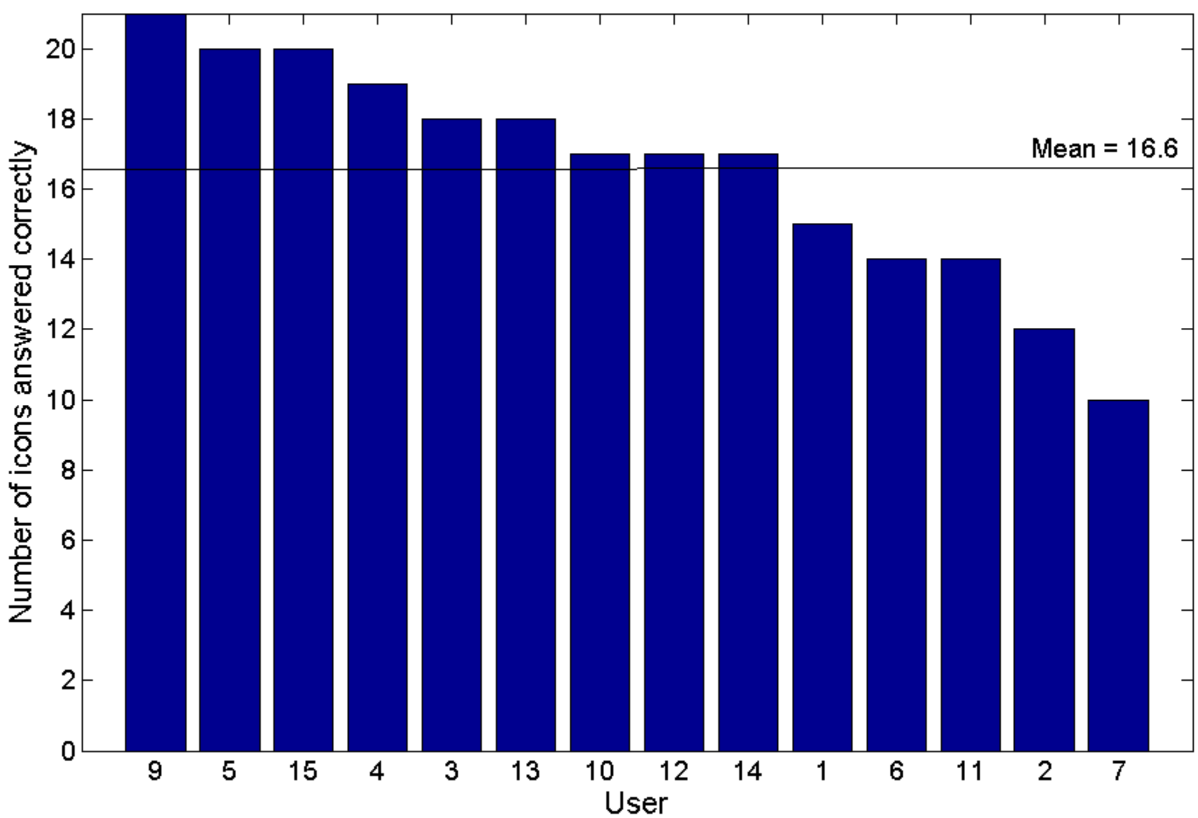

Figure 3.25: Number of icons answered correctly during final quiz for each user. Sorted by descending performance. Participant 8 is removed since he/she did not experience all 21 icons presented in the quiz.

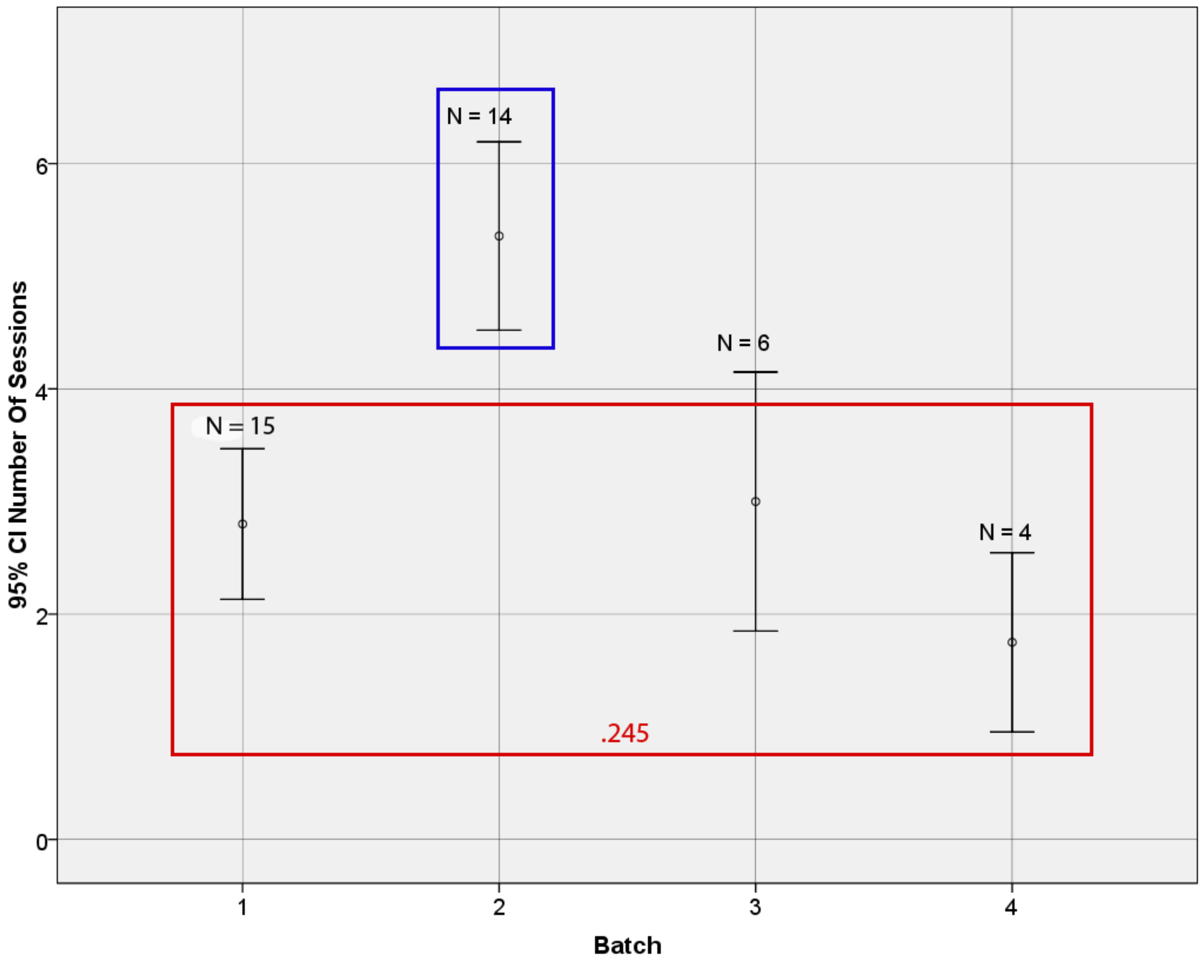

Figure 3.26: Average number of sessions required to complete each batch. Tukey's HSD test uses harmonic mean sample size of 7.2.

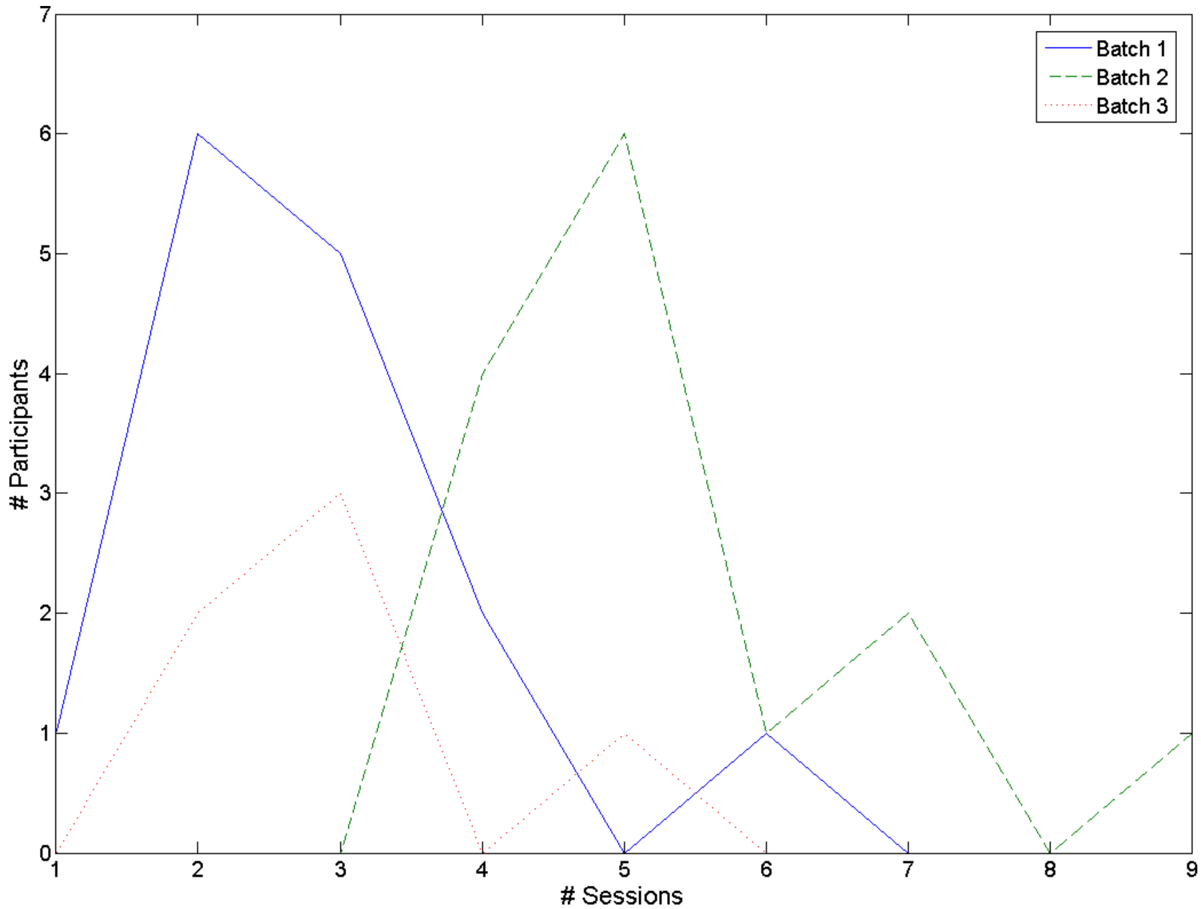

Figure 3.27: Number of participants finishing a batch in a certain number of sessions.

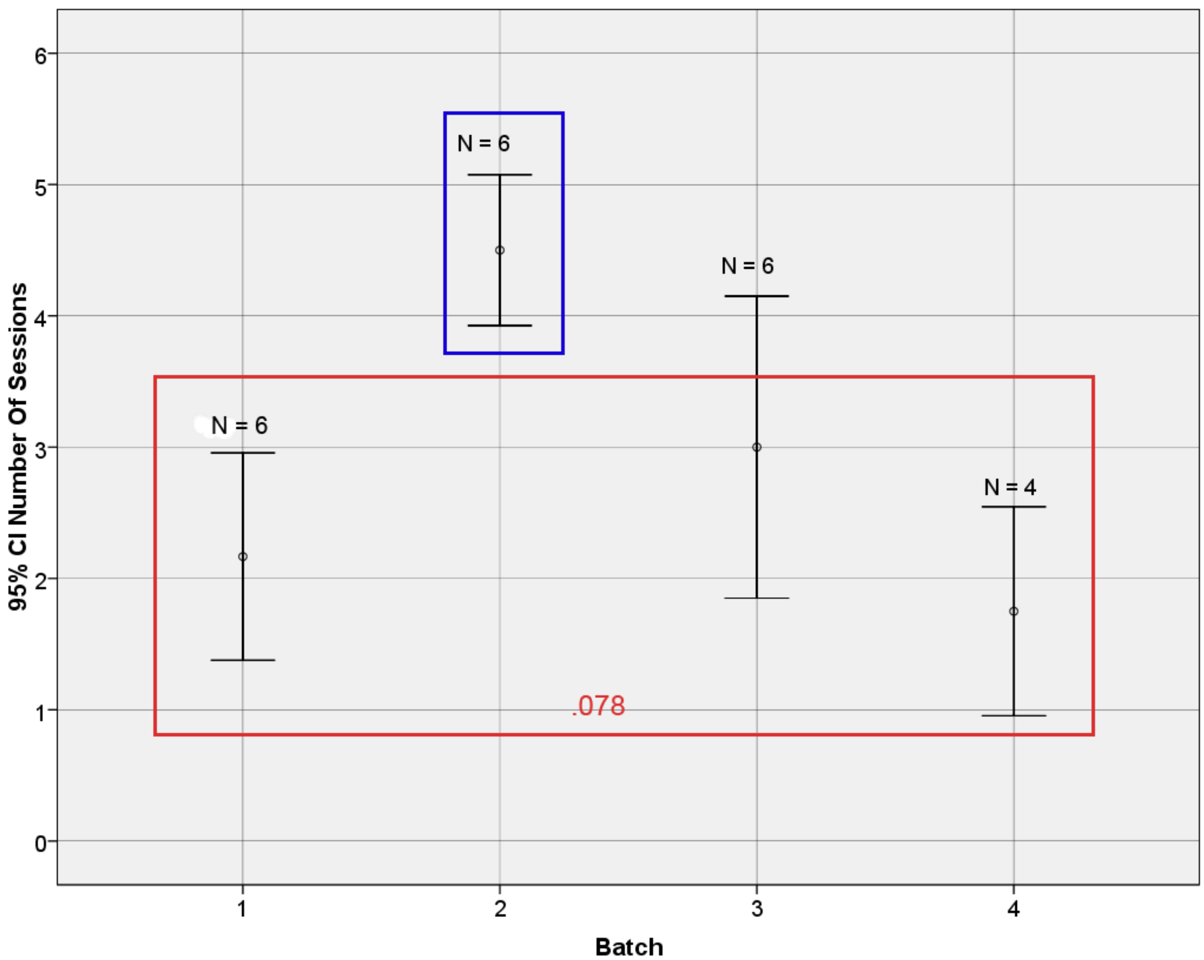

Figure 3.28: Average number of sessions required to complete each batch. Only the top 6 participants are plotted.

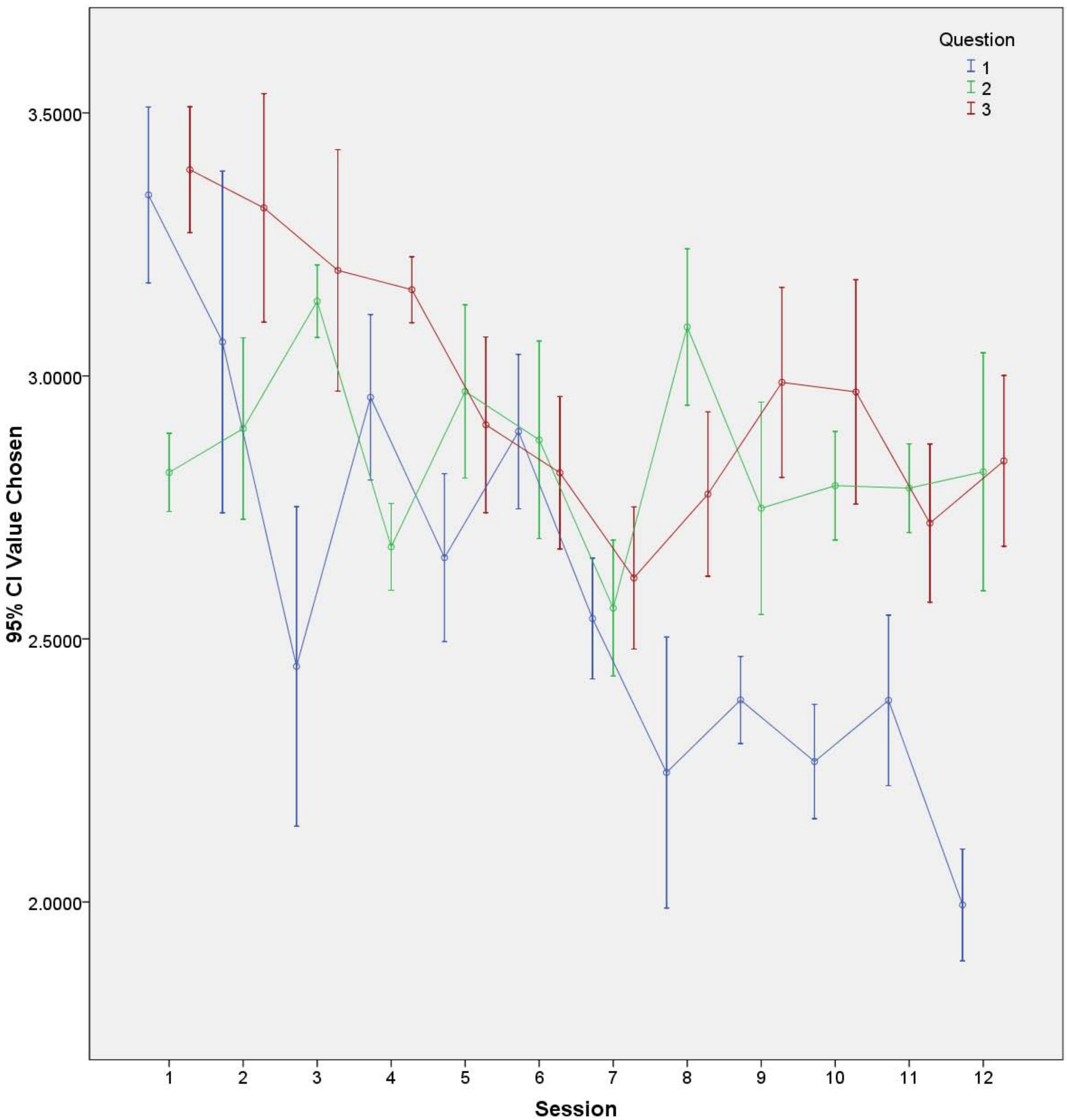

Figure 3.29: Questionnaire results, averaged by user. 95% confidence intervals are shown, N = 15.

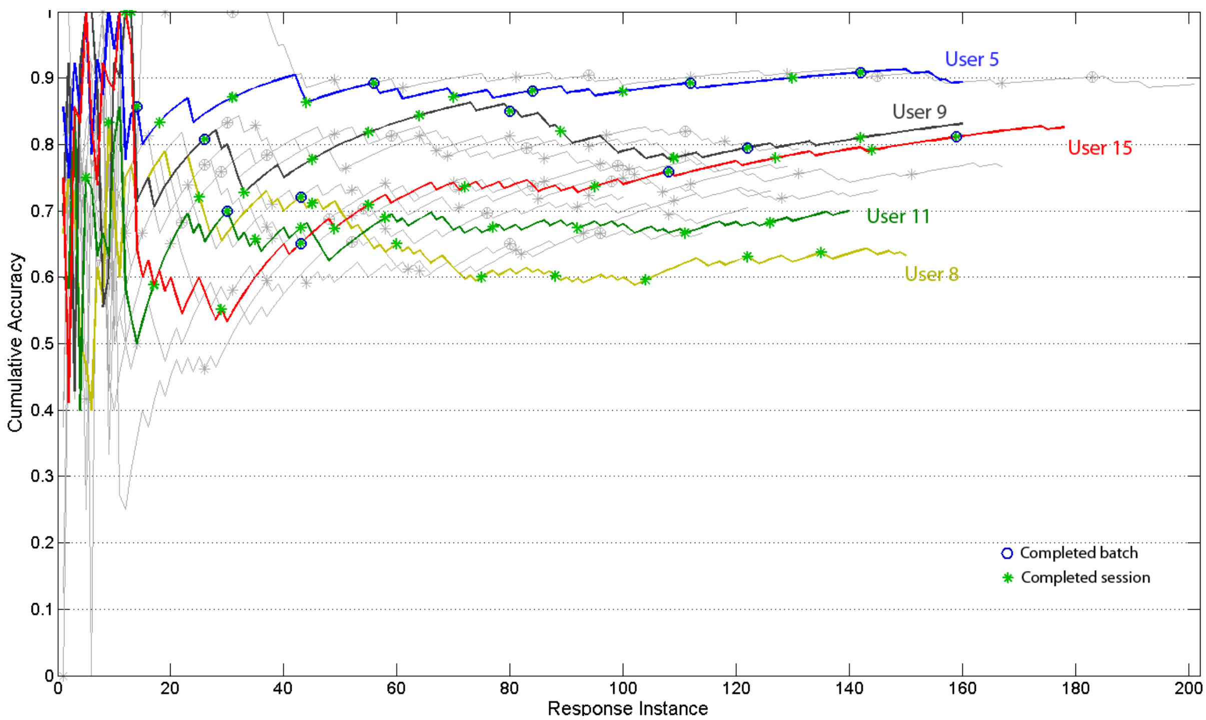

Figure 3.30: Learning curves for each participant. The cumulative accuracy is calculated for each response instance. Highlighted users are the subject of discussion and are emphasized for the purposes of readability. Blue open circles indicate where a participant completed a batch and green stars indicate where a participant completed a session.

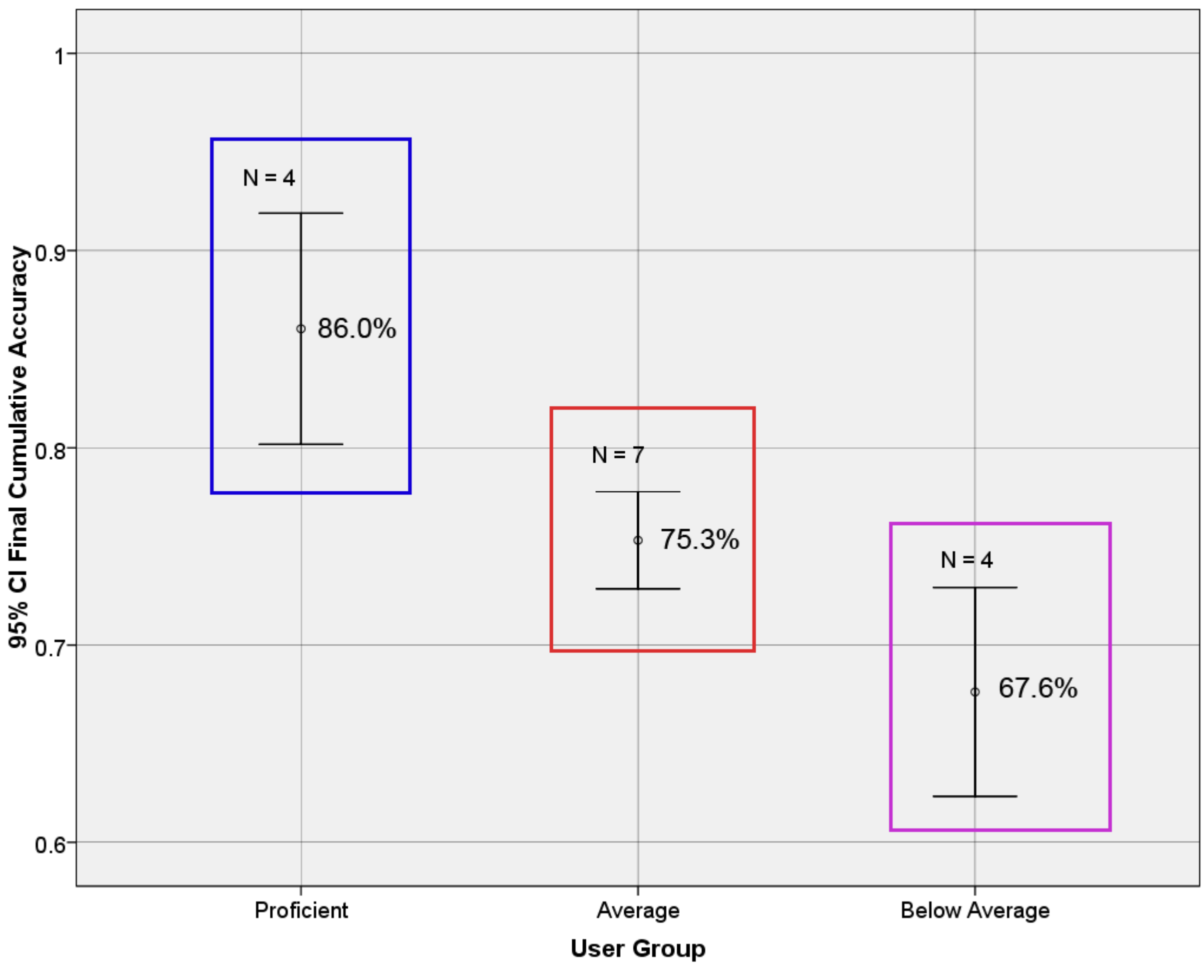

Figure 3.31: 95% Confidence intervals for the mean accuracy of each user group (Proficient, Average and Below Average). Tukey's HSD uses a harmonic mean samples size of 4.7.

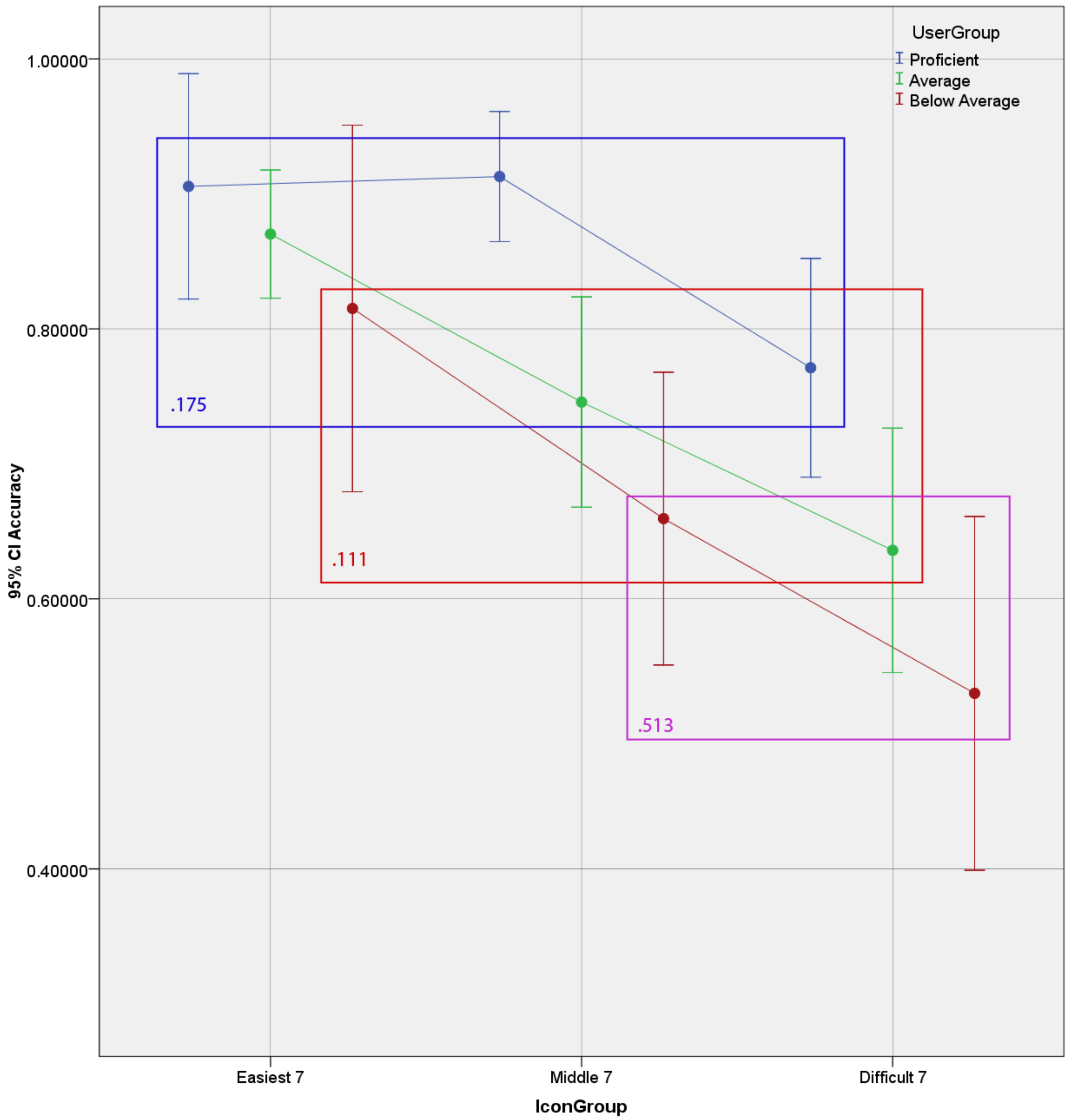

Figure 3.32: Average accuracy on 7 easiest, middle and most difficult icons for each participant group defined. Tukey's HSD uses a harmonic mean sample size of 29.6

Confusion with additional icons

As one can see from Figure 3.18, there are few significant differences between the mean performance of icons, except within the most extreme cases (icons 14 and 21). The large 95% confidence intervals reveal that there is a large amount of variation in overall mean accuracy for most icons. Examining the mean accuracy of the icons in isolation reveals very little, instead we will examine the learning process for each individual icon in order to best understand what might be responsible for learning difficulties.

In Figure 3.19, we examine the mean cumulative accuracy of each icon. In many instances, we can observe a sudden increase until between the 2nd and 7th encounter up to a peak, followed by a sudden decrease, followed by a slower climb to the final mean accuracy. This curve is best exemplified by highlighted icons 8, 11 and 20. This peak occurs at around the point where most participants complete their first batch. After this point, participants must remember 14 icons instead of recognizing and identifying only 7 icons. The introduction of additional icons increases the number of potential sources of confusion, causing a decrease in performance for icons encountered previously. Other learning effects pertaining to the increase from 7 to 14 icons are discussed extensively in Section 3.6.3.

In order to gain a more detailed and specific understanding of what causes icons to be difficult to learn, we will examine what kinds of mistakes people frequently make and identify possible explanations.

Frequently observed mistakes

To fully understand what kinds of mistakes people make, we will examine common mistakes made by participants. In this case, we inspect cases where participants made a mistake, on average, more than 10% of the time. We chose this threshold since it represents a frequency that is more than twice as likely as chance (5% for 21 icons) and may represent a chronic confusion that might be indicative of learning difficulties experienced by users. Table 3.8 enumerates the most common confusions made, sorted by decreasing frequency [2].

After examining Table 3.8 and Figure 3.20, we have identified 2 consistent and recurring classes of mistakes made by participants. These 2 classes account for 49% of all mistakes made by participants. They are confusions of:

- Note Length (36% of all mistakes)
- Rest Placement (13% of all mistakes)

In the following subsections will explain, in detail, the specific results that lend evidence to the formation of these confusion classes. We ignore cases where the type of mistake goes over the even-uneven boundary. This distinction is strong and does not result in these common mistakes – consistent with the results obtained by Ternes and MacLean [43].

[2] It is helpful to have Figure 3.1 (page 20) on hand while examining these results.

Target	Response	Frequency	Class
21	18	25%	Note Length
18	19	23%	Note Length
6	7	19%	Note Length
7	8	18%	Note Length
13	12	18%	Rest Placement
	10	10%	Note Length
19	18	14%	Note Length
10	13	12%	Note Length
8	17	11%	Meaning Assignment
20	21	10%	Rest Placement

Table 3.8: Common confusions made by participants during the Tetris game. These are all of the confusions that occurred with frequency greater than 10%. The classification of each type of error is also displayed. For clarification, 'frequency' refers to the observed proportion of a particular icon identification response out of every response made by every user for a specific target icon. 'Note length' means that the icons only differed in the length of notes. 'Rest Placement' means that the icons only differed in the timing of notes.

Confusions of note length

Icon 6 is often confused with icon 7, icon 7 is often confused for icon 8, and icon 8 is often confused for icon 17. This is an interesting non-symmetric and transitive relationship. Icons 6, 7 and 8 are all single-note rhythms that vary in length only. The target icon is often confused with the rhythm whose note length is a quarter of the bar shorter. Icons are not as often confused for their longer counterpart. The confusion between icons 8 and 17 fairly mild (11%) and will be explained in an upcoming subsection.

Icons 10 and 13 also differ only by the note length of the first note in the rhythm (Three Quarter vs. Half). Although the relationship is symmetrical, 10 is more often mistaken for its shorter-note counterpart, 13.

However, icon 21 is very often mistaken for icon 18 (25% of the time). These rhythms vary only by note length, and icon 18's first note is longer than 21's first note by a quarter of a bar. This relationship is not symmetrical (21 being 18's shorter-note counterpart).

Icons 18 and 19 are often confused for one another. Although the relationship is not symmetric since icon 18 is mistaken for icon 19 23% of the time, their confusion is quite prevalent. Both of these rhythms feature a full bar, with a long note followed by a sequence of eighth notes which feel 'shaky', but continuous. The similarity between these two stimuli is difficult to perceive visually. However, haptically, there appears to be a difference in note length between the 'shaky' part and the initial note. This relationship is similar to that between icons 6 and 7 – the rhythm with the longer note is more often mistaken for its shorter counterpart. In this case, the 'shaky' part is similar to the resting part of icons 6 and 7 since the longer note dominates the perception of the rhythm.

In general, icons are most often mistaken for their shorter-note counterpart. However, there are instances where the shorter version is mistaken for the longer version. It is difficult to determine the exact reason for this bias. We speculate that the perceptual system may experience a sort of 'resonance', where a note's perceived length might be extended. The onset of a note might mask this resonance and impose an abrupt ending to longer notes. Further investigation is required to identify the exact cause of this bias.

We hypothesize that the confusion of note length occurs most often because tactile perception is dominated by the onset of notes. Our tactile receptors respond to changes in sensation [20], therefore it may be difficult for humans to easily identify the exact length of a particular vibratory note. In addition, the resonance factor speculated above may cause further ambiguity. For this reason, we believe that it is easier for participants to identify differences in the number and density of notes, as these parameters are defined by the number and spacing of note onsets. In all of the above cases, each set of rhythms has the same number and density of perceived notes, lending further evidence to the fact that perceptual similarity is gauged based on these parameters.

Confusions of rest placement

Icon 13 is more often mistaken for icon 12, which varies simply by the placement of the quarter note in its rhythm. These rhythms differ by the placement of a rest. Icon 12 is mistaken for icon 13 9% of the time.

Icons 20 and 21 also differ only by the placement of a rest. They are also confused symmetrically about 10% of the time.

We speculate that this mistake occurs fairly often, once again, due to the fact that the number and density of notes is what dominates the recognition of a rhythm, rather than the exact timing of the notes. In the above two cases, both sets have the same number and density of notes. It is possible that the perception of the rest is also masked by the resonance factor explained above.

Notice that the confusions of rest placements and the confusions of note length carry the same explanation: that the identification of rhythmic tactile stimuli is dominated by the number and density of notes, not by the specific timing.

Confusions of icon meaning

By examining the common mistakes made in Table 3.8, we can observe that the assignment of similar meanings to perceptually distinct stimuli does not seem to have an obvious negative effect on learning. Most common mistakes are caused by similarities between stimuli. However, there is one instance of a common mistake that can be explained by meaning assignment:

Icon 8 (*'!Spouse!: I'm late, I will be there*) is mistaken for icon 17 (*'!Spouse!: Call my cell'*) 11% of the time. This is the most common confusion for icon 8 (although it is still very mild compared to the rest of the mistakes identified). The cause of this is likely due to the meaning assignment of a similar stimulus, icon 3. From personal observation, in isolation, icon 3 (*'!Boss!: Call my cell'*) seems to be perceptually more similar to icon 8 than icon 17. Since icon 17 has the message *Call my cell*, if the participant is using organizations or mnemonics based on the stimulus, it may be the case that participants are making an inference to identify icon 8 as icon 17 (*'!Spouse!: Call my cell'*) if icon 3 is their prototype for stimuli communicating the message contents *'Call my cell'*.

Although this source of error is fairly uncommon, it may lend evidence to the fact that icons with similar stimuli should have similar meanings in order to avoid confusion caused by previously existing cognitive organizations.

Performance after the removal of difficult icons

To gain an idea of how this icon set could perform without the presence of the more difficult icons, we can plot the confusion matrix with those items removed. In Figure 3.33, we show the resulting confusion matrix if we had removed the five most difficult icons (4, 13, 18, 21).

Icon Responded

Target Icon	N	1	2	3	5	6	7	8	9	10	11	12	14	15	16	17	19	20
1	90	89%						1%	3%				7%					
2	82		91%	5%							1%	1%		1%				
3	88		7%	81%	3%							1%				8%		
5	81	1%			88%	4%		4%			1%					2%		
6	83				2%	75%	19%	2%	1%									
7	62				2%	6%	73%	18%										2%
8	83	1%			5%		5%	76%	2%							11%		
9	78	1%				1%	3%	5%	86%	1%		1%			1%			
10	80					1%				85%	8%	6%						
11	94				1%					5%	93%						1%	
12	74			1%		1%				9%	3%	80%	3%					3%
14	76												100%					
15	109		2%											92%	1%	3%		3%
16	108	1%	1%	2%				4%	1%					5%	77%	7%		3%
17	89			3%	4%										4%	88%		
19	106													7%			92%	2%
20	100											6%		3%	1%		1%	89%

Figure 3.33: Confusion matrix with the five most difficult icons removed (4, 13, 18, 21). Icons within each thick blue box belong the same group of icons (7 per group).

As you can see from Figure 3.33, most of the confusion is removed, and the mean icon accuracy increases from 79% to 86%. Notice how icon 7 is the cause of most of the most chronic confusion in this matrix. As you may recall, it was the source of a three-way confusion of note length. If we remove it and plot the results (Figure 3.34), then even more of the confusion in the matrix is removed, and the mean icon accuracy is 88%. The largest source of error is from the icon meaning described earlier in this section. This is a relatively mild source of error.

Icon Responded

Target Icon

	N	1	2	3	5	6	8	9	10	11	12	14	15	16	17	19	20
1	90	89%					1%	3%				7%					
2	82		91%	5%						1%	1%		1%				
3	88		7%	81%	3%						1%				8%		
5	81	1%			88%	4%	4%			1%					2%		
6	67				3%	93%	3%	1%									
8	79	1%			5%		80%	3%							11%		
9	76	1%				1%	5%	88%	1%		1%			1%			
10	80					1%			85%	8%	6%						
11	94				1%				5%	93%						1%	
12	74			1%		1%			9%	3%	80%	3%					3%
14	76											100%					
15	109		2%										92%	1%	3%		3%
16	108	1%	1%	2%			4%	1%					5%	77%	7%		3%
17	89			3%	4%									4%	88%		
19	106												7%			92%	2%
20	100										6%		3%	1%		1%	89%

Figure 3.34: Confusion matrix with the five most difficult icons removed (4, 13, 18, 21), as well as icon 7. Icons within each thick blue box belong the same group of icons (7 per group).

These results are promising. If we can deploy the modified set, or redesign it in order to avoid the perceptual confusions observed, we may attain fairly recognition accuracy (88%).

3.5.2 What Makes Icons Easy to Learn?

From Figure 3.18, we can see that icons 14, 11, 15, 1 and 2 had a mean accuracy of above 90%. We will examine the properties of these icons further in order to understand how we should design icons to be as learnable as possible.

Simple rhythms

Icon 1 can be considered the simplest rhythm within the parameters defined for our stimuli (Section 3.1.1), and most participants expressed this fact during interviews. Quarter notes are very perceptually distinct from eighth and half notes, and they feel very natural within the time signature of our rhythms [43]. Since icon 1 consists of a continuous stream of quarter notes, it is very easy to distinguish.

Icons 2 and 3 can also be considered to be very simple rhythms since they consist only of quarter notes and the number of notes within each rhythm is extremely easy to perceive.

Even though icons 4 and 5 consist of quarter notes, we would not consider them to be simple rhythms since the notes in these rhythms are not perceived as holistic groups of notes, like in icons 1,2 and 3, and are thus more difficult to learn.

Given our time signature, the expressiveness of the rhythms that consist only of the most natural-feeling type of note (quarter notes) is limited. However, designers of haptic icons should strive to maximize the number of rhythms that they can obtain, using only simple arrangements of notes, before adding complexity through variations in timing and note length.

Recognizable rhythms

Icon 11 is an extremely interesting case. Ten out of fifteen participants explicitly reported that it feels ‘happy’ or ‘jolly’, one participant also noted that it is extremely distinctive. In addition, this message is paired with a very positive message: *‘!Boss!: Got your message - Sounds good’*.

When feeling this rhythm, its effect is palpable – it feels very uplifting. We sought to explain this phenomenon. After some discussion with a self-identified guitarist, she identified this rhythm as a popular strumming pattern utilized by musicians for happy songs [18].

Although these recognizable rhythms are very rare, it would be very helpful to identify ones that have the same effect across multiple cultures. These stimuli should be assigned meanings evoking similar emotions.

Distinctive rhythms

Icons 14 and 15 are extremely distinct within the context of the other rhythms. Since they consist entirely of eighth notes, they feel extremely intense, and most participants expressed identified these rhythms as being the most distinctive rhythms in the set.

We predict that if there were other icons in the set that were perceptually very similar to these icons, they would not have been identified as accurately. For instance, icons 6, and 7 are extremely similar, although any one in isolation would be extremely distinct from other icons (notable by the lack of confusion for icons other than 6, 7 and 8 in Figure 3.20). Without the presence of icons 7 and 8, we believe that icon 6 would have been identified with very high accuracy.

Stimulus grouping

In Figure 3.21, we showed that the confusion between different senders was extremely low. Participants mistook *Spouse* icons for *Boss* very rarely and vice-versa (below 3%).

We believe that this effect is due to the fact that the 'even/uneven' (*Spouse/Boss*) perceptual axis is extremely effective at perceptual segregation for haptic rhythms [44]. The lack of confusion between senders means that participants rarely mistook an even rhythm for an uneven rhythm.

This effective grouping property aids in the learning task considerably by partitioning the response space. Instead of choosing between 21 icons, a user only needs to decide between 10 or 11 icons depending on the sender's group.

Since this grouping characteristic is so strong and aids in the identification task considerably, we believe that future work should examine what additional grouping characteristics could be used in order to increase the size of the icon set and expressiveness without increasing the difficulty of learning.

3.5.3 What Learning Techniques do People use to Remember Icons?

In Section 3.4.2, we described the various learning techniques and mnemonics that were described by users during interviews. There were a few techniques or observations that many users agreed upon. However, for the most part, the mnemonics developed were *ad hoc* and varied considerably between users and between icons.

In this section, we will discuss the common techniques employed by participants in order to understand how we can facilitate them in the future. In addition, we will discuss the haphazardness of many of the mnemonics in order to assess its implications on haptic icon design.

Common Technique: Focus on stimulus

During interviews, most participants stated that the meanings of the stimuli were fairly unimportant, and they focused the majority of their efforts on distinguishing and remembering perceptual differences between stimuli. The meanings essentially acted as 'name' for each icon, and the association between stimulus and meaning was very simple to develop. Problems with learning arose when *stimuli* were similar.

For this reason, participants utilized the technique of focusing most of their effort on distinguishing between stimuli, rather than making the association between meaning and stimulus – this was quite easy.

This might imply that the meaning attribution process for haptic icons is not as vital to learning as the stimulus design process. Designers should focus on creating extremely distinctive and recognizable stimuli and rely on the symbolic powers of the human brain to perform the meaning association.

This does not mean that the meaning attribution process is unimportant. It is extremely important to assign meaning sensibly and with good reasoning in order to facilitate good salience management and the formation of strong mnemonic learning techniques. We believe that participants would have performed better if we did not perform a randomized meaning attribution process. However, without perceptually distinct and recognizable stimuli, a haptic icon set will be very difficult to learn. Our results confirm that good, perceptual design is paramount.

Common Technique: Vocalization

Many participants reported that they would vocalize an aspect of the message with the same rhythm as its associated stimulus. For instance, icon 12 has the meaning *'!Boss!: I'm in a meeting: Call me later'* and the stimulus has a rhythm consisting of a half note followed by a quarter note. Two participants would vocalize this as 'LA-ter'. Participants 9 and 14 reported vocalization as a primary technique.

Most participants would also rehearse the rhythms vocally in order to get a better sense for the timing and composition of the haptic rhythms by utilizing a more familiar channel.

These results indicate that participants utilize the auditory channel in order to help them identify or distinguish the tactile rhythms. The auditory channel is very closely linked to the haptic channel since they are both primarily temporal in nature [22, 32] – stimuli in both of these modalities are transient. In terms of communicating information, humans utilize the auditory channel extensively through speech, music, etc. It is not surprising that participants rehearse and make use of the auditory channel in order to learn tactile stimuli. It is very difficult to reproduce a tactile rhythm with the body, as tapping does not produce the same vibratory sensation.

As an implication for design, tactile rhythms should facilitate vocalization. For instance, if one has two meanings where the subject of one has 2 syllables and the subject of another has 3 syllables, one should assign haptic rhythms with 2 and 3 notes respectively. In addition, longer notes should represent emphasized syllables.

Although this is a heuristic for design, it has limited use as a global design technique and should not be used as a primary design parameter for large sets of icons. However, its implications for learning facilitation should be kept in mind while assigning meanings to these large sets.

Common Technique: Rapidity = Urgency

Most participants reported that they associated rapidity, or note density with urgency. Participants stated that they perceived certain messages associated with rapid stimuli as being more urgent than others. Participants did not state that any messages associated with languid rhythms were urgent.

This effect was so strong that many participants referred to rapid rhythms as 'urgent rhythms'.

Haptic icon designers should take advantage of this observation while designing icon sets. The more rapid stimuli should be assigned urgent meanings. In addition, one could modify the tempo dimension of a rhythm to portray urgency, rather than using amplitude or frequency. Further research in examining the implications of utilizing tempo as a design parameter would be a worthwhile pursuit.

Common Technique: Partitioning

Many participants reported that they would subdivide similar stimuli in order to partition learning. For instance, icons 6,7, and 8 have extremely similar rhythms. Some participants identified that they thought of the three rhythms as a group, and this allowed them to avoid confusion with all other icons. The task then, became identifying which of the three icons was being played. If the stimuli were too similar, this was a very difficult task.

Another instance of this partitioning was imposed by assigning the sender of the message based on a stimulus' evenness. In Section 3.5.2, we identified that this design decision made the icons easier to learn.

This has a very strong implication for design. Haptic icon sets should take advantage of the similarities between stimuli by making their semantics similar as well. This way, users can take advantage of the natural grouping imposed by the perception of the stimuli. However, if the stimuli are too similar, the identification task will be too difficult for this learning aid to be beneficial.

Haphazard Mnemonics

From Table 3.6, we can notice that there is very little consistency between the mnemonics utilized for learning icons, except for the special cases that we have already discussed.

This lack of consistency has a very powerful implication: People are good at coming up with mnemonics and learning techniques no matter what the association between the stimulus and meaning are. This implication is consistent with results obtained by Enriquez and MacLean [16], where they found that identification performance for haptic icons was not improved when users were allowed to perform the meaning attribution process compared to a completely random meaning attribution process.

These results may indicate that, given the very limited expressiveness of modern tactile devices, metaphorical design may not be any better than abstract, perceptual design. In addition, metaphorical design is more difficult to execute, is subject to limitations imposed by the creativity of the designer and is not repeatable. Through abstract design, we can focus our efforts on creating a large, reusable stimulus set with desirable perceptual qualities and assign meanings in a sensible manner that facilitates learning.

Summary: Implications of learning techniques for design

The discussion in this section identified a number of implications for design that were identified by the learning techniques that participants exhibit. We give a summary here:

- Tactile rhythms should facilitate *vocalization*.
- Well-performed *perceptual design* enables the learnability of a haptic icon set. However, meanings should still be assigned sensibly in order to facilitate learning.
- Utilize the *rapidity* of a note to portray *urgency*.
- Take advantage of *perceptual similarity* by assigning *semantically related* meanings.
- Avoid the temptation to perform metaphorical design. There is very little evidence to suggest that it is better than abstract, perceptual design since humans are very adept at learning symbolic stimuli.

3.5.4 What Happens when we Add Modifiers/Attributes to Icons?

After participants complete Batch 3, they encounter icons with identical rhythms and meanings as before, but have a different amplitude and/or frequency. Recall that this change in amplitude and/or frequency denotes an attribute of the icon. High frequency rhythms denote a second sending of a message and low amplitude rhythms denote a low priority message (Section 3.1.2).

Figure 3.22 shows the average accuracy on icons with attributes added as well as the accuracy on the base icons, for users who completed batch 3. We also display the proportion of responses where the modified icon is mistaken for its default base and vice-versa.

As you can see, performance on icons with attributes added continue to be identified with high accuracy by the users who completed batch 3. In addition, there is rarely any confusion between an icon with an attribute added and its rhythmic base and vice-versa.

We believe that this phenomenon is due to the fact that the *rhythm* of the stimulus dominates perception. It is apparent that the addition of attributes is easily perceptible since there is little confusion between icons that have an attribute added and one that does not. However, participants do not perceive these icons as completely distinct. Since the accuracy is so high, it is apparent that they make the association between the rhythm and its meaning, since both of those remain the same, despite the addition of attributes.

We do not believe that accuracy would have been nearly as high if we gave entirely new meanings to these modified rhythms. Since the association between a rhythm and its meaning has already been made, we predict that there would be a great deal of confusion between the new, modified rhythm and the old, unmodified rhythm.

This has implications for design. If one were to design a set of icons and they wished to add additional attributes to the icon delivery, they could utilize frequency or amplitude to convey that attribute. Future work would have to examine the limitations of this phenomenon. For instance, was learning so easy because the stimulus attribute (amplitude) is salient and well matched to the semantic attribute (priority) or would we get a similar effect with an attribute that is less compatible? We believe that the effect would be the same as long as the attribute's perceived intensity correlates properly with the direction of the semantic attribute. For instance, it would be difficult for users if low priority messages were high amplitude and high priority messages were low amplitude.

We are unsure whether the perception of rhythm is strong enough to require a designer to assign one meaning per rhythm. Perhaps a designer could assign different meanings to stimuli with the same rhythm, but for instance, a different melody (explored in Chapter 5).

3.5.5 Does Perception-based Design Work? Does it Accurately Predict Learnability?

In Figure 3.23, we showed a reproduction of the perceptual MDS map obtained by Ternes and MacLean [43], with only the high amplitude, low frequency stimuli displayed.

In our study, we observed that icons 6, 7 and 8 are often confused for one another. In Ternes and MacLean's results, these rhythms are indeed clustered. In this case, the MDS proximity supports our hypothesis that the perceptual similarity of the stimuli caused the confusion between the icons.

In Ternes and MacLean's results, stimuli 10, 13, 18, 20 and 21 are clustered. Our results do show confusion between 10 and 13 and between 18, 20 and 21, with very little confusion between the two groups (10, 13 in group 2; 18, 20, 21 in group 3). We can come up with a couple of explanations for this fact. First, the groups of icons exist in different batches and were learned separately. The method of introducing icons by batch imposes a cognitive organization to icon learning. This imposed organization may overcome confusion caused by stimulus similarity. Second, the task in Ternes' study is quite different. Participants were asked to sort stimuli based on similarity. Creating a group containing of a long note followed by short notes is a natural organization within his context. This kind of organization may not be immediately obvious when portions of the set are revealed in sequence rather than all at once.

The task in MDS is similarity, from which we are inferring dissimilarity. However, stimuli can have strong points of similarity, but be very dissimilar at the same time. These discrepancies may reveal that some of the stimuli in the MDS results are somewhat similar, but the analysis did not capture an important dimension (such as perceived rapidity).

Another interesting result from Ternes and MacLean is the similarity between stimuli 1, 14, and 15. In our study, these icons are all identified with extremely high accuracy – in fact, the entire quadrant is identified with high accuracy. The lack of confusion may demonstrate that Ternes' results are confounded by the task. In the context of his task, it makes sense to group rhythms with a large number of short notes together (high density). However, these stimuli could be extremely perceptually distinct despite sharing a similar property. For an intuitive example, consider classifying a sponge and gelatin amongst a large group of materials with very little compliance. One might group the sponge and gelatin together since they are both 'squishy'. However, they have entirely different textures and are perceptually very distinct from each other. However, in the presence of many materials which are different by a single dimension, they would be grouped as similar.

Another instance of the above phenomenon may be observed with icons 4, 16, and 17. In Ternes and MacLean's results, they are grouped closely together. However, in our study, there is relatively little confusion between these icons (although there is some), likely due to the difference in task and the fact that stimuli are presented in groups rather than all at once.

It may also be the case that adding a unique meaning to each stimulus makes them more perceptually segregated. The imposition of meaning as a dimension can further disambiguate the differences between the stimuli.

The similarity between all of these inconsistencies is the fact that Ternes and MacLean *overestimate* confusion caused by perceptually similar stimuli. In Figure 3.24, we plot the difference between the normalized similarity values obtained by Ternes and MacLean and our confusion matrix (Figure 3.20). In this figure, the cells with high brightness indicate where the results from Ternes and MacLean overestimate perceptual similarity with respect to observed confusion.

There are no cases where Ternes and MacLean's similarity matrix does not predict confusion, whereas we observe it here. In hindsight, we should have removed some stimuli that they found to be perceptually proximal using MDS in order to reduce the perceptual confusion, and thus, the number of mistakes made by participants. However, we wished to evaluate their complete set, and, given over-conservatism, this would have resulted in removing some items that are actually perceptually distinctive.

These results have a very important implication. Perceptual, MDS-based design is not likely to underestimate confusion between icons, but rather overestimate it. Results obtained from perceptual similarity experiments are much more conservative than results in deployment.

It is safe to draw conclusions about identification performance from MDS results since, in the worst case, the designer will only lose expressiveness by eliminating stimuli that are clustered together in an MDS map. The MDS technique accurately predicts instances of close perceptual similarity, but does not do as well at predicting dissimilarity in deployment.

There is a need for a perceptual design tool that can overcome the shortcomings of perceptual MDS. For instance, the task should be more similar to conditions encountered in deployment such that stimuli should be revealed in sequence rather than all at once. In addition, if the tool could capture the *differences* between stimuli, it would be very useful as well. Since cluster sorting studies have users sort stimuli based on similarity, and from this, we infer dissimilarity. However, this only captures how stimuli are similar and now how stimuli are different.

3.6 Discussion of Results Pertaining to Participant Differences

In this section we will discuss the results of the experiment pertaining to differences between participants in order to understand how people perform during the haptic icon learning process. We attempt to provide insight into how haptic icons can be learned as effectively as possible in deployment through our analysis of the results obtained.

This section is also organized by our research questions, similar to Section 3.5, but we discuss the questions pertaining to users.

3.6.1 How Many Icons can People Learn?

In Figure 3.25, we can observe that people correctly identified, after one week without any exposure to haptic icons, on average about 16 icons under no workload after about three 20 minute sessions for four weeks. This final quiz was administered without any warning to participants.

The exact number of icons learned is difficult to express from experimental results since participants were constantly learning. Participants experienced between 14 and 42 icons, with overall cumulative average accuracy ranging between 63% to 89%. In general, the most accurate users experienced the most icons, while the least accurate ones experienced the least.

It is difficult to determine whether the large differences in ability were due to differences in haptic sensitivity, the ability to form a semantic link with a haptic stimulus, or symbolic learning abilities in general. In hindsight, we should have administered tests to collect baselines for all of these measures.

However, there is no reason to believe that there is a limit on the number of learnable icons if their stimuli are perceptually distinct. Since these are simply symbolic and abstract stimuli, humans have the ability to learn an infinite number of them [14].

Consider how many words a human knows: our vocabulary is essentially limitless. We have symbolic identities for a seemingly infinite number of referents. Given enough exposure to haptic icons people should be able to learn a very large number, as long as they are perceptually distinguishable.

We also believe that the speed of icon learning will depend on how the icons are designed. If we design icon sets that take advantage of the symbolic learning abilities of the human brain, learning will be facilitated. For instance, one could create groups of similar, but distinctive icons in order to allow users to partition the learning set into groups, or 'chunks' [34].

3.6.2 How Well can People Retain Learned Icons?

In Section 3.4.3, we showed that there was a significant difference between the performance on new icons (75%) vs. old icons (81%) during the quiz. This result indicate that the identity of the old icons are retained with high accuracy in later batches. Recall that users were required to identify a subset of 7 icons from previously encountered batches during the quiz (Section 3.3.7).

In addition, we tested participants on the first 21 icons, without warning, approximately one week after all of the training and identification sessions had been completed. During this quiz, users were able to identify, on average, 78% of the 21 icons correctly (excluding participant 8). Recall that 8 of these participants did not successfully complete batch 3, which would have qualified their mastery of the first 21 icons. Despite this fact, participants were still able to identify almost 80% of the icons.

If we remove the effects of the 5 most difficult icons, as well as icon 7 (identified in Section 3.5.1), participants had a mean final quiz accuracy of 85%.

We believe that these results show promise. After one month of training, and a week of no exposure, people can correctly identify, on average, 16 out of 21 icons. We believe that this performance level and number of learned icons could increase significantly with the constant exposure that one would experience in a deployment scenario. In addition, if the icons were designed with more care by avoiding perceptual similarity (Section 3.5.1) and maximizing aspects that make icons easy to learn such as grouping and stimulus distinctiveness (Section 3.5.2), then we believe that user could learn even more icons with even greater accuracy.

3.6.3 What Does the Learning Curve Look Like?

In Section 3.4.3, we showed that Batch 2 required significantly more sessions (approximately 2) to complete than batches 1, 3 and 4. This pattern was apparent for all 15 users, as well as the top 6 users. We postulate that this 'bump' in the learning curve may be explained by two factors, or a combination of both:

1. An increase in haptic or rhythmic sensitivity resulting in the increased ability to distinguish similar rhythms.

2. The formation of elaborated, deep cognitive structures for long-term storage and retrieval.

We will argue that this 'bump' is likely caused by a combination of cognitive scaffolding and an increase in sensitivity. We will first explain why we believe that this phenomenon may caused by an increase in sensitivity, and then explain why we believe that cognitive changes are also taking place.

Increase in haptic or rhythmic sensitivity

One explanation for the 'bump' in the learning curve would be that the progression to learning 14 rather than 7 icons increases the potential sources for perceptual confusion. This may also be an explanation for the temporary decrease in performance for already-learned icons after the progression to a new batch (identified in Section 3.5.1). The 'bump' may be caused by the need for the development of increased haptic or rhythmic sensitivity, explained by the slower learning in Batch 2. Once this sensitivity has been developed, the participant might have better acuity and perception during the progression past the third batch, which is learned much faster.

There is evidence in research in the field of cognitive neuroscience that the development of perceptual sensitivity, resulting from synapse development, neuronal recruitment and nerve development, can occur within a timespan of months. In the book *Plasticity in the Human Nervous System: Investigations with Transcranial Magnetic Stimulation*, the authors cite a study where unsighted people learning braille show a gigantic increase in the cortical representation of the dominant reading finger after 2 months of learning [3]. There is also a slower, steady increase in representation that is believed to be caused by recruitment of other cerebral structures.

Unfortunately, in the above study, the first cortical imaging session occurred after 2 months. Since the increase in cortical representation is so drastic, from interpolation, we might predict that the increase in sensitivity was taking place very early in the process – even from the first training session. We believe that at least some increase in sensitivity was occurring through synapse development since participants were taking part in a novel form of learning.

Participants may also have learned how to pay more attention to perceiving stimuli using their tactile sense since haptic perception usually occurs sub-attentively. Our task trains the participants to be more aware of receiving information through this background channel.

Formation of cognitive scaffolding

In his seminal paper, George A. Miller reports that the capacity of working memory is 7 'chunks' of information, plus or minus two [34]. In other words, people are able to hold approximately 7 chunks of information in memory for identification; any more and deeper cognitive structures must be developed for reliable recall [1].

This phenomenon appears to be apparent in our results. In Section 3.4.3, we showed that Batch 1 was fairly easy for participants to learn, while Batch 2 was significantly more difficult. Batch 1 only contains 7 icons, therefore one could make the argument that their identity could be retained in working memory during the experiment and recalled without entering the information into long term memory. Once the user progresses to Batch 2, they have 14 icons to remember. This exceeds the limits of working memory, therefore the participants must encode the information into long term memory through elaboration, the formation of mnemonics, rehearsal, etc. This process takes a great deal of time and effort, which could explain the increase in the number of sessions required to learn 14 icons. This theory is also supported by the fact that we observed a degradation in performance for many icons once participants completed Batch 1. The identity of these icons can no longer be held in working memory, therefore they must encode them into long term memory before the icons can be recalled reliably.

We observe that in Batch 3, learning occurs significantly faster than in Batch 2, and approximately as fast as in Batch 1. This phenomenon may be due to the fact that once a user has cognitive scaffolding in place that will support encoding new icons into long term memory, long term learning occurs with much greater ease. This explanation may be akin to learning the basics of programming, which takes a great deal of time and practice, but after the basics are learned, it is generally much easier to grasp more advanced techniques. In addition, learning supplementary programming languages becomes mostly just a matter of syntax.

In order to increase the speed of the haptic icon learning process, designers should focus on facilitating the formation of this cognitive scaffolding so that icons can be learned with greater ease. Meaning assignments should be predictable, and take advantage of cognitive aids such as grouping stimuli to partition the learning effort, as described in Section 3.5.3

Individual learning curves

In order to glean additional information about the learning process experienced by users, we will examine some individual learning curves plotted in Section 3.30.

Many users experienced a significant drop in performance after completing a batch. We have highlighted User 9 as an obvious exemplar of this phenomenon. This may be caused by two things: first, a redefinition of the user's cognitive structure for recognizing haptic icons; and secondly, that the training and quiz portion of the experiment program were not sufficient for reliable learning at the 80% level.

Once a user obtains a new set of haptic icons to learn, they must integrate it into their currently existing cognitive structures for long term storage and recall. If they are progressing past Batch 1, then they likely go through the process explained in the previous subsection. If they are progressing past Batch 2, then they must integrate the new icons with their existing cognitive structure. The process of elaboration and long term learning does take time, no matter how strong cognitive scaffolding is. Users will likely not display proficiency immediately after being introduced to a new batch. Potential conflicts may also arise if previous mnemonics are no longer applicable. A few users complained of this.

In addition, this lends evidence to the fact that, for the most part, the training and quiz were not sufficient for reliable recall at the 80% level. The fact that most users performed below the 80% level during the game even after passing the quiz was evidence of this as well. For practice to be effective, it is important that it occurs within a realistic setting. In addition, the practice of attempting a response and then receiving feedback on the response's correctness is important to the icon learning process.

User 8 was the participant with the lowest accuracy and the least amount of progression through the icon set. After learning Batch 1, we notice a very significant drop in accuracy, and then a long plateau, followed by an increase in accuracy. This particular user complained that he was unable to feel the icons very well during the game, thus the differences between them were very difficult to perceive. During the quiz, he was able to concentrate on the stimulus and perceive them much more easily. The progression to Batch 2 may have introduced far too many stimuli for the user to distinguish with his current level of sensitivity and the addition of workload present in the game may have also diverted too much attention away from his already straining haptic modality. Eventually we notice a rise in the identification performance. This may be due to an increase in tactile sensitivity, and/or an increased ability to divert more attention to the haptic channel.

User 15 is also an interesting case. This participant reported that he did not pay much attention to or put much effort into the experiment until after around session 3. This is evident by his fairly low accuracy during the first three sessions of the experiment after around response instance 15. At around response instance 30, his accuracy becomes extremely high due to the fact that he is actually paying attention and making an effort. In fact, for the last 100 response instances, his accuracy is 99% – higher than any other participant. This phenomenon is evidence that effort is extremely important to learning ability. If a person does not make an effort to elaborate on stimuli in order to encode them into long term memory, then their ability to learn will suffer dramatically.

3.6.4 What Makes some People Better than Others at Learning Haptic Icons?

In Section 3.4.3, we divided the participants into three different groups: proficient, average and below average. We showed that these groups are all significantly different from each other, with mean accuracies of 86% for proficient users, 75.3% for average users and 67.6% for below average users. Based on this result, we believe that our segmentation is adequate to draw conclusions about groups of users who achieve different levels of accuracy.

Based on the results obtained, we were only able to observe one consistent difference between proficient and less proficient users: robustness to difficult icons.

Robustness to difficult icons for proficient users

In Figure 3.32, we showed that there was no significant difference between the performance for the 7 easiest, moderate and most difficult icons for users in the proficient group. In fact, the mean accuracy remained almost constant for the easiest and moderate 7 icons, while a drop of 14% for the mean accuracy was observed for the most difficult icons.

Users in both the average and below average groups experienced a linear degradation of performance between each successively more difficult icon group. The mean accuracy for average users dropped by about 11.5% between successive groups, while the mean accuracy for below average users dropped by approximately 14.5% between icon groups. There was a significant drop in performance between the easiest 7 icons and the 14 more difficult icons for both groups.

It is not completely clear whether proficient users have a better sense of rhythm in general, have an increased ability to recall abstract content, or have greater acuity for differentiating tactile rhythms.

The self-reports for musical ability and sense of rhythm give no indication that these users might have a better sense of rhythm. We did not find any correlation between these measures and the accuracy obtained by participants.

We also do not believe that this effect was caused by an increased ability to recall abstract content. All users are drawn from a fairly homogeneous, highly intelligent group of individuals. Also, most participants complained, during interviews, of problems with distinguishing similar stimuli, not with making a link between the meaning and stimulus. In particular participant 8, who was the user with the lowest mean accuracy, reported that he found it very difficult to feel the vibrations coming from the device. For him, perceptual similarities were more pronounced due to the low intensity of the vibrations. The proficient users did not report these problems with perceptual similarity as often.

These results indicate that the proficient users were simply more robust to confusion imposed by more difficult icons. Since most errors in identification were caused by perceptual similarity (Section 3.5.1), it appears that the more proficient users have greater acuity for differentiating tactile rhythms.

A related finding by Newman et al. supports this hypothesis [35]. They show that the braille letters A-J are more discriminable than the letters K-T, and this resulted in a very significant effect of learnability: people were able to learn the letters A-J much faster than they could learn K-T.

Although the proficient users may have had a stronger baseline ability for perceptual discrimination of tactile rhythms, it is possible that all of the users could obtain additional sensitivity through training. As explained in Section 3.6.3, a very significant increase in the cortical representation of the dominant reading finger was observed after 2 months of learning braille for unsigned users.

It may be beneficial to future attempts at training people to learn haptic icons if they begin by attempting to increase the user's sensitivity to tactile rhythms by introducing a wide range of stimuli and requesting the user to focus on the differences between them before introducing meanings. Naturally, in deployment, the meanings should be attached to the most distinguishable stimuli.

3.7 Heuristics for Design, Guidelines for Training and Advice for Hardware Designers

In Sections 3.5 and 3.6, we examined a great number of different results, developed explanations, and discussed prescriptive implications for haptic icon design and training. In this section, we will summarize these findings and list them as heuristics for rhythmic haptic icon design and provide guidelines for haptic icon training. We anticipate that these heuristics and guidelines will aid designers in creating expressive haptic icon sets that are more easily learned, and present them to users in a way that will be most beneficial to the learning process.

In addition, we wish to make suggestions to hardware designers with the goal of making their devices as useful as possible for haptic icon transmission.

3.7.1 Heuristics for Haptic Icon Design

The haptic icon design heuristics developed throughout this chapter are not all exclusive to rhythmic stimuli. In addition, some heuristics are concerned with increasing the perceptual distinctiveness of a set, while some relate to increasing the icon set's ability to support the learning of the association between meaning and stimulus. For this reason, we divide the design heuristics into three different groups: heuristics for perceptual distinctiveness; heuristics for associability; and heuristics specific to rhythmic icons.

Heuristics for perceptual distinctiveness

We have argued that haptic icon designers should design stimuli to be perceptually distinctive in order for the resultant icon sets to be more learnable. The following is a list of heuristics that will aid in developing distinctive stimulus sets:

- *Utilize MDS analysis to give accurate predictions of perceptual confusion.* MDS analysis is a relatively quick and efficient design and evaluation methodology for creating distinguishable sets of stimuli [32]. However, be aware that although MDS accurately predicts perceptual similarity between stimuli, it can underestimate dissimilarity in a deployment scenario involving meaning assignment.
- *Avoid the temptation to perform metaphorical design.* There is a widespread belief that metaphorical icon design is better than abstract icon design; however, there is very little evidence to support this fact. Metaphorical design is non-repeatable, relies on the creativity of the designer, and could result in a stimulus set with undesirable perceptual properties. With abstract design, haptic stimulus sets are adaptable to any domain, and can be more learnable of the stimuli are distinctive.
- *Modify the number and density of notes for perceptual distinctiveness. Utilize note length and timing to accentuate differences.* We have evidence that variations in the note length and/or the timing of rhythms are the main causes of perceptual confusion during icon identification. However, these differences are perceptible; instead, note length and timing should be utilized to accentuate differences between stimuli that already vary in number and density of notes. We believe that this heuristic applies to more than just rhythmic stimuli since our findings relate to the haptic modality in general.

Heuristics for associability

Although we emphasize designing perceptually distinct stimuli, the meaning attribution strategy employed while designing haptic icons is extremely important. Here we list strategies that may increase a user's ability to form associations between semantics and stimuli:

- *Similar stimuli should have similar meanings.* Perceptual similarity imposes a natural grouping to stimuli. Designers should take advantage of this fact by assigning similar meanings to groups of stimuli. This makes the icon set more predictable and structured as well as facilitating partitioning in order to make learning occur more easily. However, be careful not to make stimuli too similar in order to avoid perceptual confusion.
- *Utilize the rapidity of a note to portray urgency.* Rapid icons were strongly perceived as being more urgent than their languid counterparts. Icons with urgent semantics should be assigned highly dense stimuli. Further work should investigate whether a modification of tempo can have the same effect.
- *Take advantage of vocalization when assigning meanings to haptic icons whenever possible.* Although it is likely not useful as a primary design parameter for large sets of icons, vocalization is a powerful learning tool. If the stimulus corresponds with an important aspect of the icon's semantics, it may be easier to learn. If this design choice is made clear to the user, it will likely increase the chance of success.

Heuristics specific to rhythmic icons

Although we believe that most of our heuristics can be applied to haptic icons in general, these heuristics need more support before they can be classified as general:

- *Utilize extremely recognizable and emotion-evoking rhythms.* Rhythms that are commonly used in popular music may be extremely recognizable and evoke emotions from users. Use these rhythms to your advantage by assigning semantics that evoke similar emotions. Further study is required to determine exactly which rhythms evoke such familiarity.
- *Design stimuli with simple rhythms before adding complexity through variations in note timing and types.* Simple rhythms make use of the most natural note type within a time signature and are more likely to contain variations in note number. People consider simple rhythms to be very distinctive and agreeable.
- *Utilize display parameters such as frequency and amplitude to represent modifications on already learned haptic icons.* Modifications of frequency and amplitude to rhythms, representing different semantic attributes, were adopted effortlessly by participants that encountered these icons. Further study is required to investigate the limitations of this heuristic.

3.7.2 Guidelines for Haptic Icon Training

In deployment, users will need to be trained to learn haptic icons in order for them to useful. If the presentation and training strategy is not effective, then the learning process will be jeopardized. In this section, we list some guidelines for haptic icon training.

- *Begin training by focusing on increasing haptic sensitivity.* Sensitivity to perceptual differences between stimuli appeared to separate the proficient users from the less proficient users. To increase the haptic sensitivity of users, expose them to a wide array of stimuli while pointing out the differences. We believe that this will increase their ability to pay attention to haptic information, as well as increasing the cortical representation devoted to haptic sensors.
- *Encourage helpful techniques such as rehearsal through vocalization and stimulus partitioning.* Both of these techniques were shown to be beneficial when utilized by participants. If icons are designed so that similar stimuli have similar semantics, then partitioning will have an even greater effect. Learning will also be facilitated if the meaning attribution takes advantage of this vocalization.
- *Have users perform realistic practice.* We have shown that workload can affect the ability to perceive the difference between haptic stimuli. In order to ensure that users can identify icons in a realistic situation, their training should include high workload situations.
- *Expose your design parameters to users.* If users are aware of your stimulus design and meaning attribution strategies, then they will be more able to develop sensible mnemonics and learning strategies. In addition, unless one is running an experiment, there is very little to lose by doing this.
- *Provide incentives or sources of enjoyment in order to keep motivation high.* Motivation and effort increase learning ability. The training process should be fun and encouraging in order to increase the desire to learn haptic icons.

3.7.3 Advice for Hardware Designers

In our current experiments, it is important to note that we are using a particular snapshot of tactile display technology, which has a set degree of expressiveness. With advances in hardware display technology, perhaps a greater number of perceptually distinct stimuli can be created. In this section, we enumerate a few guidelines for hardware designers seeking to create more expressive displays.

- *Focus on the intensity of the attack of a vibration.* In our case, the perception of rhythmic, vibratory stimuli was primarily distinguished by the number and density of notes. The timing and length of notes was less important. For this reason, by making the attack of a note more perceptible, users will be more easily able to identify individual notes.

- *Aim for a quick recovery.* Since the density of notes is an important factor in the perception of rhythms, increasing the display capabilities along this dimension is desirable. If notes are spaced too closely together, there is a risk that they will 'bleed' together, causing them to be perceived as a single vibration. If the hardware is able to recover more quickly, then the individual notes will be much more distinct.
- *Support variations in frequency/amplitude.* Our results showed that variation of frequency and amplitude over an entire stimulus was very successful at expressing varying attributes on icons. This capability can be very powerful. Control over the global frequency/amplitude of notes over any given series of actuations is desirable. We do not have evidence at this point whether it is necessary to allow for frequency/amplitude variation of individual notes. This capability is explored extensively in Chapters 4 and 5.

3.8 Primary Contributions

In this chapter, we described and analyzed a longitudinal study designed to ascertain the effectiveness of haptic icons in a deployment situation, as well as developing a deep understanding of what makes haptic icons most easily learnable. In this final section of the chapter, we will summarize the primary contributions of this work.

A common theme was prevalent throughout this chapter: the perceptual distinctiveness of icons is paramount. Our results support the fact that the semantics of any particular icon is not as important as perceptual distinctiveness for optimal learnability. Unless a user can perceive the differences between haptic stimuli, they cannot learn them. Haptic icon designers should focus on the stimulus, and use helpful tools such as perceptual multidimensional scaling in order to create distinctive haptic icons sets. The semantic attribution task for icon design should be performed carefully, using some of the heuristics outlined in this chapter.

Specifically, stimuli should not express variation through the timing and length of notes. Instead, designers should focus on the number and density of notes in order to ensure optimal perceptual distinctiveness. A global modification of monotone rhythms though amplitude and frequency variation can enable the robust addition of attributes to haptic icons. This technique was extremely effective, and did not result in any additional confusion between icons. In addition to these findings, in Section 3.7, we outlined a series of design heuristics, guidelines for haptic icon training, and advice for hardware designers in order to optimize the learnability of haptic icons in deployment.

Based on the findings described during this chapter, we believe that haptic icons have the potential to be quite effective in deployment. Users were able to learn, on average, about 16 icons after a period of one month. However, despite the fact that individual users tended to learn consistently over time, there were much variability in performance between users. This phenomenon may be due to differences in haptic sensitivity between users and further study is required in order to determine how we can best utilize individual sensitivity parameters. Despite this fact, most users were able to successfully learn and identify a larger number of haptic icons than previously reported by related research, therefore we believe that, given adequately distinctive haptic stimuli, haptic icons will be extremely effective in deployment.

CHAPTER 4

PERCEPTUAL ANALYSIS OF MELODIC ADDITION

Despite the encouraging results of the previous chapter, the maximum number of available semantic units – individual, perceptually distinctive rhythms – was limited to approximately 16 icons. However, the device utilized to render our haptic stimuli (described in Section 1.1.2) could express variations in frequency and amplitude, but the original design effort of the stimulus set did not attempt melodic variation due to the exponential explosion in design space [43].

In this chapter, we explore the use of melodic addition in rhythmic haptic stimulus design in an effort to increase the usable set size of perceptually distinctive haptic icons. Up to this point, the design and evaluation of haptic icons that are designed through the variation of frequency and amplitude between individual notes has not been attempted. We attempt to make this large design space more manageable through the introduction of heuristics inspired by musical theory. In addition, we wish to design 'families' of distinct stimulus groups. In the previous chapter, we were only able to support two senders in our cellular messaging deployment scenario since Ternes and MacLean identified one very prominent perceptual axis (evenness). We wish to utilize this larger, more complex design space in order to discover additional perceptual axes on which to develop groups of perceptually similar, but mutually distinctive stimuli.

Melodic variation would also allow for an increase in the potential expressiveness of haptic icons over monotone stimuli since designers can develop richer, more elaborate stimuli. The use of melody in haptic stimuli may allow for an increased likelihood of an emotion-evoking stimulation, resulting in a more memorable set. However, the use of melody may increase the potential of evoking a memory, which comes with its own and highly individual meaning or semantics. We fear that the evocation of an emotion may make it harder to create new semantic associations.

In order to design a usable melodic stimulus, in addition to developing an initial understanding of melodic haptic stimuli, we performed a series of iterative perceptual MDS analyses. This exploration process will be described in the following sections.

4.1 Approach and Overview

The primary goals of this work are to explore the use of melody in increasing the expressiveness of stimulus sets, while increasing usable set size (number and size of viable stimulus groups). In addition, we wish to determine and generalize the criteria that people use to group such stimuli, in support of future design.

To deal with the explosive increase in design space with the addition of melodic variation, we used heuristics derived from musical theory to compose our initial groups of stimuli, which were displayed on a mobile device with a piezo-actuated touchscreen (described in Section 1.1.2).

Music composition rules are melodically or rhythmically based (Table 4.1, page 90). Ternes [43] states that for perceptibility, stimuli should be limited to a small frequency range on the Nokia platform used here, and differences between frequencies must be relatively large. Thus, to enable creation of a large number of groups, rhythm was our main basis for group design. Most of our initial groups had related but not identical rhythms, while amplitude and frequency varied within groups. To understand how users organize these stimuli (relevant to our secondary goal), we conducted user studies in which participants sorted stimuli into varying numbers of groups.

In the context of this melodic design space defined by rhythm, tone and intensity, we pose the following hypotheses: (H1) user-study participants' first-order stimulus groupings will follow *rhythm* in spite of frequency and amplitude variation; (H2) participants' first-order groupings will *not* follow either amplitude or frequency; and (H3) participants will demonstrate ability to discriminate stimuli which vary in amplitude and/or frequency but have consistent rhythm. Experimental validation of all three would confirm that rhythm should be used for primary groupings, and further imply that amplitude and frequency are suitable for within-group variation.

We also predict that participants' first-order stimulus groupings may be based upon ideas transferred from musical theory; although this conjecture is less certain and considered a path of investigation.

In the future, we also would like to test whether melodic variation enhances icon usability beyond that achievable with monotone rhythm; however, it will be challenging to devise an unbiased comparison.

In the remainder, we first outline our design approach used to design a haptic stimulus set. Through an iterative design and evaluation sequence, we then show that even in the presence of melodic variation, users employ rhythm as a first-order grouping property of haptic stimuli; whereas melody is suitable (and important) for within-group stimulus variation. The following sections describe our stimulus design process, the studies and their analysis, and conclusions drawn from them.

4.2 Stimulus Design

4.2.1 Design Parameters

Creative experimentation cannot be the sole basis for designing reusable sets of haptic melodies, limited as it is by designer creativity and situation-specific concerns. Instead, with an approach of 'perceptual design' we seek to understand how humans classify, compare and respond to melodies [31] by designing stimuli with informed heuristics and validating stimulus sets through user participation.

4.2.2 Stimulus Design Space

Haptic melodies are defined by rendering characteristics such as duration, tempo, frequency, note density and amplitude ranges, which in turn are constrained by hardware capabilities and human perception. Our haptic melodies are displayed on the modified Nokia 770T Internet Tablet, described in Section 1.1.2. This device restricts vibration rendering to ≈ 50 discretely timed vibrations each at uniform frequencies and amplitudes per programmed script; therefore continuously varying amplitudes (e.g. complex sinusoids) are not feasible.

Constraints due to the need for a consistent and usable signal structure (e.g. overall duration, repetition to achieve rhythmic sense, empty space for note definition, etc) are taken from [44]. To review, Ternes and MacLean's 21 haptic rhythms for the Nokia 770T had the following characteristics:

- Each stimulus is 2 seconds long, with 4 identical, consecutive 500 ms repetitions.
- Each iteration is divided into 16 equally spaced segments.
- Each iteration is comprised of notes which occupy 16 (*whole*), 12 (*three-quarter*), 8 (*half*), 4 (*quarter*), 2 (*eighth*) and any number (rests) of segments.
- Spacing between notes is 31.25 ms (after eighth and quarter notes) or 62.5 ms (all others).
- Distinct 500 ms stimuli were devised based on heuristics.

Ternes expanded the 21 rhythms to 84 stimuli by playing each at one of 2 amplitudes and 2 frequencies (i.e. every note in a given 2-second stimulus has the same tone and intensity); and demonstrated set discriminability with an MDS visualization[44]. The mutual uniqueness of these stimuli is based solely on rhythmic variance.

To move from haptic *rhythms* to *melodies*, we developed new heuristics to systematically apply melodic variation (note-by-note changes in tone and intensity) to this base set of heuristically determined, perceptually validated rhythms. Ternes [43] suggests that two amplitudes and two frequencies best suit the capabilities of the human perceptual system.

4.2.3 Design Heuristics

The inherently rhythmic nature of haptic melodies suggests a potential for design guidance from musical theory. Evidence that audio and haptic signals are correlated and complementary [22, 28] suggests that musical techniques for eliciting human affect might transfer to the haptic sense.

Characteristic	**Classification**	**Response(s)**
Major tonal scale	Melodic	Happiness
Minor tonal scale	Melodic	Sadness
Slow tempo	Rhythmic	Serenity, Fear
Rapid tempo	Rhythmic	Jollity, Anger
Regular Spacing	Rhythmic	Calmness
Syncopation	Rhythmic	'Jarred' feelings

Table 4.1: Common associations between musical characteristics and human affective responses

Associations between melodic and rhythmic composition techniques (e.g., specific variations of pitch, tempo, use of syncopation, etc.) and human affective response are well documented. Table 4.1 lists associations derived e.g. from [21, 39, 41], which we used as a starting point for our own heuristic set. Since many rely on differences in rhythm, we also used them to produce new groups of stimuli centered around rhythms that depart from the original 21. Altogether, our initial heuristics were:

- **Ensure syncopation differences between melodies** (i.e., ensure that some rhythmic bases are perceived as 'off the beat' as opposed to being regularly spaced)
- **Ensure note density differences between melodies** (i.e., rapidity vs. languidness – resulting from the number of notes per melody – should differ between melodies)
- **Ensure differences in frequencies and amplitudes between melodies** (i.e., strength and variance of vibrations should differ between distinct melodies)
- **Ensure rhythmic differences between melodies** (i.e., rhythmic base of melodies should vary between groups)

4.2.4 Melodic Stimulus Design Tool

A haptic melody visualization, authoring and playback tool (called the *Melodic Stimulus Design Tool* or MSDT) was developed using Java 1.6 and the related Swing Project. It supports drag-and-drop note placement according to the constraints of the design space and melody structure (Figure 4.1). We used MSDT to produce haptic stimuli according to the heuristics above, by modifying frequency and amplitude composition of the 21 starting rhythms as well as producing new rhythms in line with the heuristics not captured in the original set.

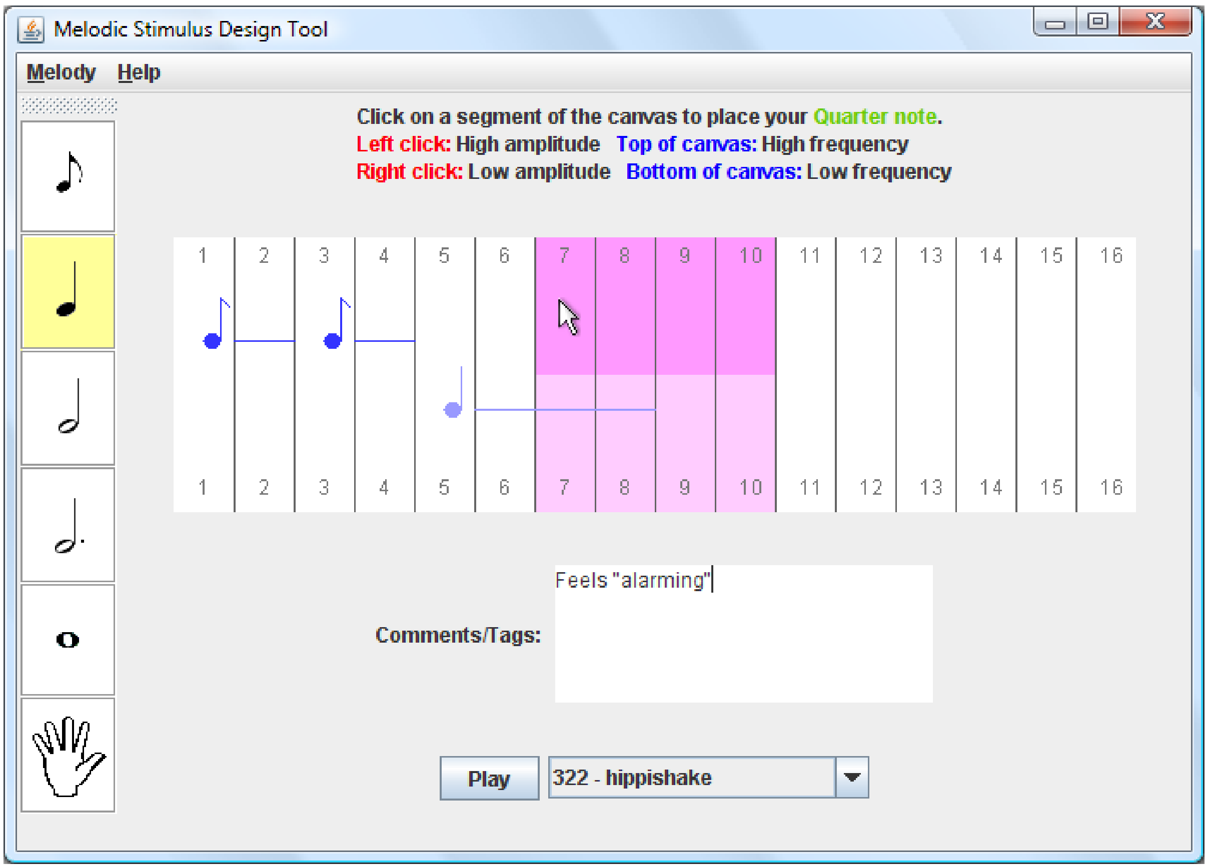

Figure 4.1: The *Melodic Stimulus Design Tool* showing a haptic melody consisting of two high frequency, high amplitude eighth notes followed by a low frequency, low amplitude quarter note.

MSDT's note canvas has two vertically stacked segments. The medium gray horizontal bar in Figure 4.1 under the cursor indicates that a mouse click will place a *high frequency* quarter note (4 segments) starting in the 7th segment of the bar. If the user moves the mouse downward to the light gray horizontal bar, the colours will swap and clicking will result in *low frequency* note placement. In both cases, a left click will place a *high amplitude* note, and a right click a *low amplitude* note. Opacity differences visually indicate note amplitude. Because a note starting in segment 7 would overlap with the quarter note already starting in segment 5, the pre-existing quarter note would be automatically removed to make room. The 'Comments/Tags' box has been used to note that the melody 'feels alarming', a tag which will be saved with the stimulus. Visual and auditory playback allowed stimulus preview; encoding was required to play them on the Nokia display.

4.3 Initial Stimulus Set

We ultimately produced 6 stimulus groups, each containing 6 stimuli. Based on the design heuristics laid out in Section 4.2.3, we developed perceptual distance metrics to mutually compare the initial 21 rhythmic bases. Results from this process provided a space from which we could evenly sample rhythmic bases. We chose 4 bases from the initial set of 21 to provide adequate coverage, allowing the two remaining groups to explore heuristics not captured in the initial set. Specifically, the groups based on concepts (syncopation or descending amplitude) employ new rhythmic bases not represented in the original 21 stimuli. The rhythm-based groups have identical note composition, varying only in frequency and amplitude of individual notes. The concept-based groups vary in rhythm, fullness (the number of segments in which a note is present), frequency and/or amplitude. The initial set is described as follows and illustrated in Figure 4.2:

Rhythm-based groups:

- **3Q**: 3 quarter notes followed by a rest (1–6)
- **1H2Q**: 1 half note followed by 2 quarter notes. (7–12)
- **2Q**: 2 quarter notes followed by 2 rests. (13–18)
- **4-6E**: 4 or 6 eighth notes; either 6 eighth notes; or 2 eighth notes, 2 rests, and 2 eighth notes. In the case that 6 consecutive eighth notes are used, the center two are low amplitude to reduce their impact. (19–24)

Concept-based groups:

- **S**: all melodies are uneven or syncopated, with varying note densities and types. (25–30)

- **HLA**: high-to-low amplitude; the whole bar is filled with various note types. The first half always has high amplitude notes, while the second half always has low amplitude. The melody is always low frequency despite changes in note composition. (31–36)

4.4 Study 1

4.4.1 Hypotheses

Group Number. Because we designed our stimuli in 6 groups, we hypothesized that users would also find 6 the most natural number of groupings for the full set.
Rhythmic Grouping. We also hypothesized that users would sort melodies based on our rhythmic grouping ideas, outlined in Section 4.3.
Concept Grouping. Finally, since this work is exploratory in nature, we hypothesized — tentatively — that users would likely find syncopation and consistent amplitude composition to be groupable criteria. Although this lies in conflict with the hypotheses laid out in Section 4.1, this seems somewhat reasonable based on musical theory transfer. The reconciliation of this hypothesis will provide additional design guidance, reduce potential sources of Type II error and avoid confirmation bias.

4.4.2 Participants

The initial study had 7 participants, a number in line with recommendations for MDS analyses [31]. All were Computer Science graduate students aged 18–25 (6 male).

4.4.3 Apparatus, Task and Design

Using the Nokia N770T tablet described in Section [16], participants were asked to group the melodies together into a specified number of bins based on their own similarity criteria. An application was created on the Nokia device to facilitate the grouping task (Figure 4.3). In the application, all 36 melodies appeared at the bottom of the screen as buttons. Their order was randomized each time a new grouping screen was loaded. The buttons could be clicked on to feel their melody, or dragged into a bin. Participants were asked to hold the device in their non-dominant hand and use the stylus for interaction.

Due to limitations of the device, only 16 of the buttons (indicated by an '!' symbol beside their number) had haptic feedback loaded at any given time. To load a button's feedback, a user would click on an unmarked button and, after a few seconds, feedback would load for that and the surrounding buttons. While this inconvenience increased the difficulty of the task, we have no reason to believe that it biased the final outcome.

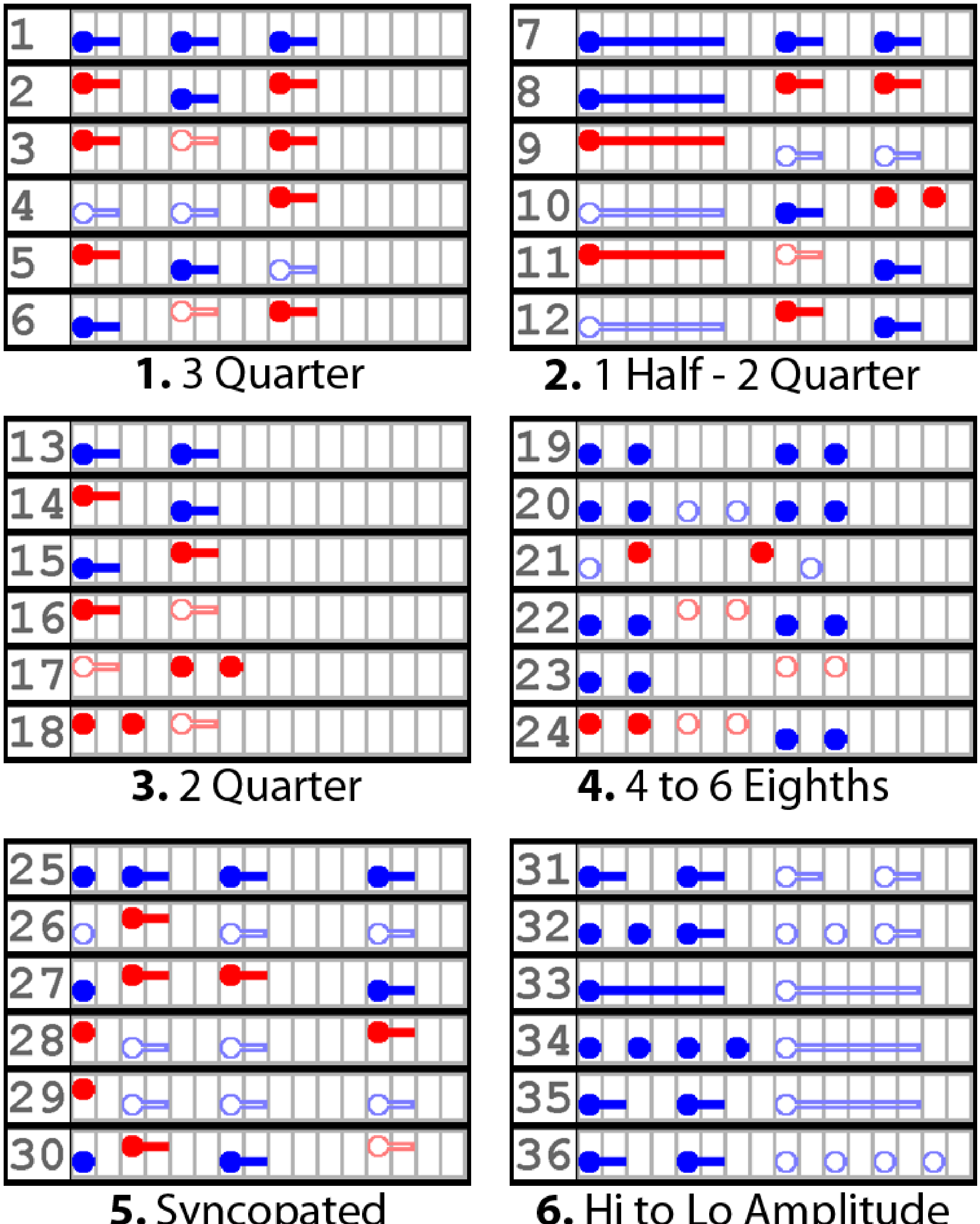

Figure 4.2: All 36 melodies from the initial six groups. Hollow notes are low amplitude, raised, red notes are high frequency.

Each participant performed 4 grouping tasks for the same 36 stimuli. The group sizes were 3, 6, 9, 12, presented in random order. The sorting data was recorded as well as subjective information obtained through various questionnaires.

4.4.4 Procedure

Participants completed a pre-session questionnaire to report demographic information such as country of origin, years of experience playing a musical instrument, and sense of rhythm.

Before completing the sorting tasks, participants were given a quick demonstration of the software and task. They were encouraged to develop their own similarity criteria with which to sort the stimuli, and asked to keep the bins filled at similar levels. After each of the four sorts, participants filled out a post-group questionnaire where they indicated the difficulty of choosing melody groups given the number of groups allowed in that sort.

After all four sorts, participants completed a post-session questionnaire asking for their preferred number of groups, the overall difficulty choosing a group for stimuli, sorting strategies, and meanings that came to mind for any of the stimuli.

Dissimilarity values for each user and stimulus pair were calculated identically to [32].

4.4.5 Results and Preliminary Findings

Qualitative MDS analysis

The first study's results are shown in Figure 4.4. The most groupable characteristic appears to be consistent, non-syncopated rhythm, which is mostly in line with our Rhythmic Grouping hypothesis. Most melodies in the rhythm-based groups are clustered fairly close together in while those in the concept-based groups are not.

We carried out a detailed qualitative analysis by closely examining the MDS stimulus clusters.

Group 3Q. 3Q melodies 1, 2, and 4 were sorted together, but 5, 6 and 3 were not. While the 3Q melodies all have the same sequence of notes with varying frequency and amplitude, 3 and 6 have a varying amplitude that participants described as 'rolling' or 'pulse-y'. Conversely, 1, 2, and 4 had either no or one amplitude change so they did not feel the same as 6 and 3. These melodies were described as 'abrupt' or 'hard'. Melodies 1,2,3,4 and 6 ended in a high frequency or maintained a consistent frequency, in contrast to 5 which ended in a low frequency, possibly explaining why it was not sorted near other 3Q melodies.

Group 1H2Q. Melodies 7, 8, 10 and 12 were sorted together. As for 3Q melodies, similarity was based on a change in amplitude and frequency that ended in a high or consistent selection of each. One outlier, 11, had a low amplitude in the middle of the melody giving it a more 'rolling' feeling than the

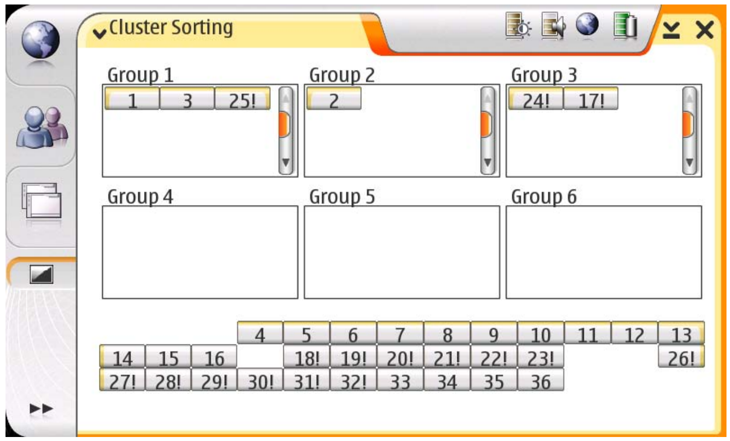

Figure 4.3: Screenshot of the sorting application on the Nokia 770T. A user is sorting the 36 stimuli into 6 groups.

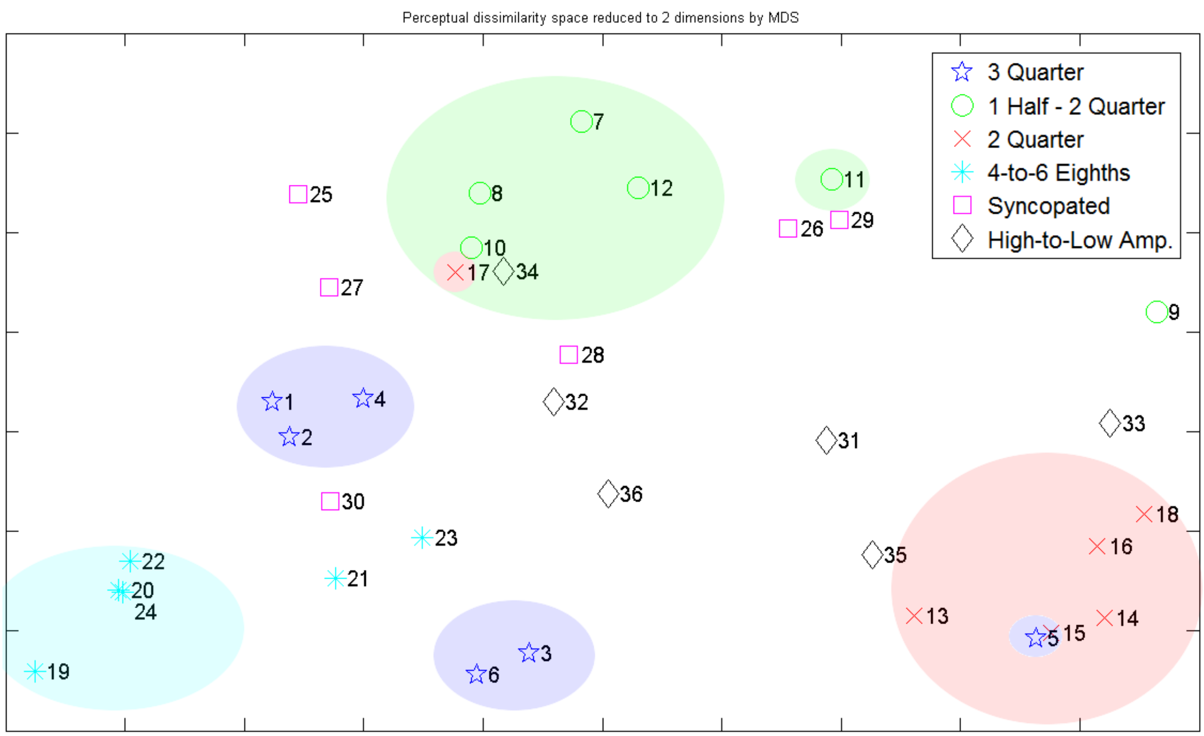

Figure 4.4: Visual representation of the icon sortings for all group sizes. The exact nature of the dimensions is unclear and complex. However, the horizontal axis seems to reflect changes mostly in fullness and the vertical axis seems to reflect changes in note length.

others. The other, 9, was the only one in this group to begin with a high amplitude and frequency and end with a low value of each.

Group 2Q. All 2Q melodies except 17 were sorted together. Their simple, consistent rhythm made them easy to distinguish from other melodies while feeling similar to each other. Melodies 13 – 16 had 2 quarter notes. Melody 17 had 1 quarter then 2 eighth notes, and 18 had 2 eighth and then 1 quarter note. 17 probably felt the most dissimilar due to its 2 concluding eighth notes which were emphasized by their high amplitude. This accentuated the different number of notes between this melody and the others in group 2Q.

Group 4-6E. Melodies 20, 22, and 24 were most closely grouped. Most of these stimuli felt 'rolling' as they had either no or low amplitude notes in their middle. Even though two melodies ended in a low amplitude and/or frequency, the 'rolling' characteristic was strong enough for most participants to group them together.

Group S. Participants did not group the syncopated melodies together. Although the sequence of notes was consistent, we suspect that the unevenness of the melodies coupled with widely varying amplitudes and frequencies made them feel dissimilar.

Group HLA. This group, based on the concept of descending amplitude, had widely varying rhythms while amplitude, frequency, and fullness were held constant. No melodies in this group were found near one another, suggesting that rhythm and note type may be a more groupable characteristic.

Questionnaire results

The post-group questionnaires indicated that sorting the melodies into 12 groups was the most difficult (average=1.7, on a scale of 1-5 with 1 being most difficult) and 3 groups easiest (average=2.6).

The results for the post-session data revealed the average difficulty overall to be 1.9. 3/7 participants thought 6 was the best number of groupings, 3/7 preferred 3, and 1/7 preferred 9 groups. All participants indicated that they correlated stimuli based on the rhythm or tempo of notes. 3 said that amplitude similarity played a factor, and 3 participants also sorted based on frequency or intensity of notes.

Although demographic information such as musical experience and country of origin was collected, due to the small number of participants, it was not feasible to analyze the effect of these variables on the task.

Preliminary findings

From this study, we summarize several observations of stimulus grouping:

- Changes in both frequency and amplitude were perceived qualitatively as changes in *intensity* – consistent with [10].

- A high amplitude note followed by a low amplitude note of the same frequency is perceived as 'rolling' or 'pulse-y' – in contrast to 'hard' or 'abrupt' sequence. Such notes feel continuous and seem to 'roll' into one another rather than feeling distinct.
- Low amplitude sustained notes surrounded by *staccato* notes are often grouped with those that have rests in the same position as the sustained notes.
- Syncopated melodies are not grouped together.
- Melodies consistent in amplitude composition but varying in rhythm and note density are not grouped.
- Stimuli are generally not grouped by amplitude or frequency.
- The most groupable stimulus characteristics are note density and rhythm, as long as the rhythms are not syncopated.

With these observations in mind, we redesigned the stimulus set (Section 4.5) and conducted a new study with the goal of formulating more definite design heuristics.

4.5 Icon Redesign

To create a better set of stimuli and learn more about how melodies are perceptually grouped, we develop a new set of melodies by iterating on the previous set with our new information.

First, we removed the two concept-based groups. We then modified the anomalous melodies in the other four groups to make them more perceptually uniform (e.g. making Group 1 more 'abrupt' rather than 'rolling' by avoiding low-then-high amplitude notes on the same frequency).

Finally, we created two new groups in an aim to learn more about potential design heuristics. Both were derived from qualitative feedback from the first study. For the first new group [**All eighths (AllE)**], the entire bar is filled with eighth notes since dense stimuli were perceived as urgent and we wanted to maximize this characteristic for discriminability. The second group was created to feel like a horse 'gallop' [**Gallop (G)**] since some previous participants associated stimuli with patterns of locomotion ('walking', 'limping', 'running', etc.)

Figure 4.5 illustrates representative stimuli for these six groups.

4.6 Study 2

Hypotheses, apparatus, task, design and procedure are identical to Study 1 (Section 4.4). Six Computer Science graduate students (5 male) participated in our second study. None of the participants were involved in Study 1.

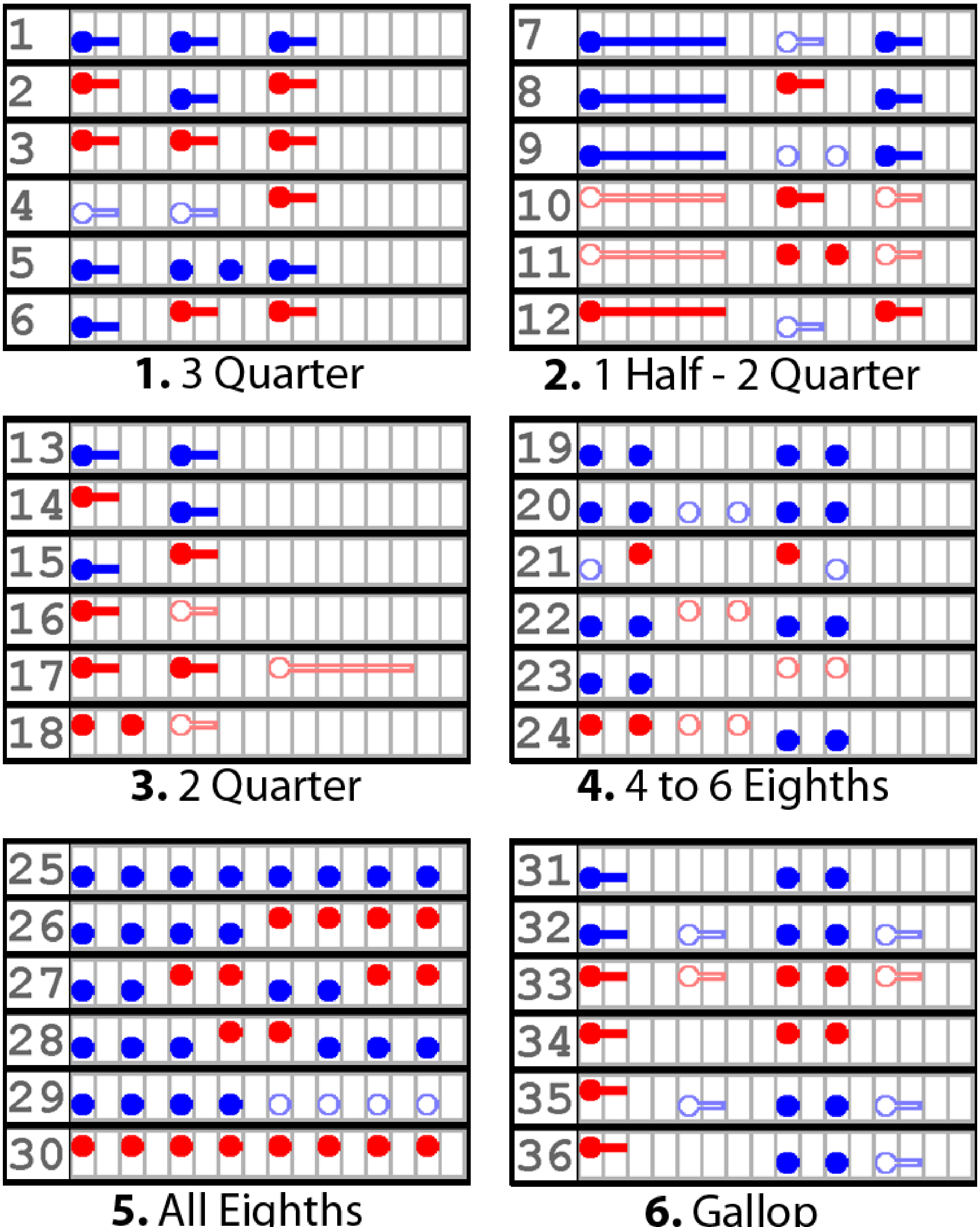

Figure 4.5: All 36 melodies from the re-designed six groups. Hollow notes are low amplitude, raised, red notes are high frequency.

4.6.1 Results

Inspection of the MDS visualization from the second study (Figure 4.6) show that the melodies of group AllE, with one exception, were highly distinguished from all other melodies.

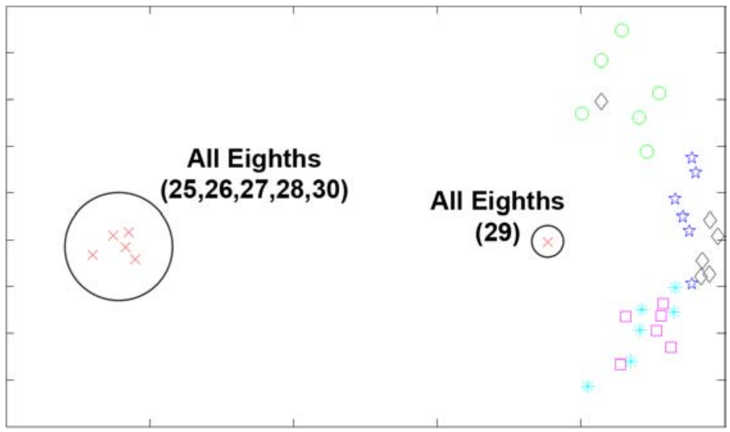

Figure 4.6: Thumbnail of MDS map of the stimulus sortings for all group sizes showing group position. 5/6 stimuli in group AllE are displayed in the left of the diagram, while all other stimuli are displayed on the right.

A closer inspection – without the dominating AllE group – was desired to analyze the results more thoroughly. Pasquero *et al.* [38] state that hidden patterns within the data can be uncovered by performing MDS on sub-matrices of the original dissimilarity matrix. A visual representation of the results excluding group AllE was produced (Figure 4.7).

Qualitative MDS analysis

A detailed qualitative analysis was again carried out on the stimulus clustering in the MDS perceptual visualization.

Group 3Q. Every melody in this group was sorted together except for 5, which was the only one with 2 emphasized eighth notes in place of the center quarter note.

Group 1H2Q. All melodies were grouped together.

Group 2Q. All melodies except for 17 were sorted together. 17 was the only one with a half note taking up the whole last half of the bar. While designing this group we predicted that the half note would feel like a rest, but here it clearly did not. Instead, this rhythm (2 quarter notes followed by the single half note) was more often sorted with notes of group 1H2Q (one half note followed by 2 quarter notes); stimulus 17 was distinct from the stimuli from group 1H2Q only by *phase*.

Perceptual dissimilarity space reduced to 2 dimensions by MDS for second set of stimuli with All Eighths group removed

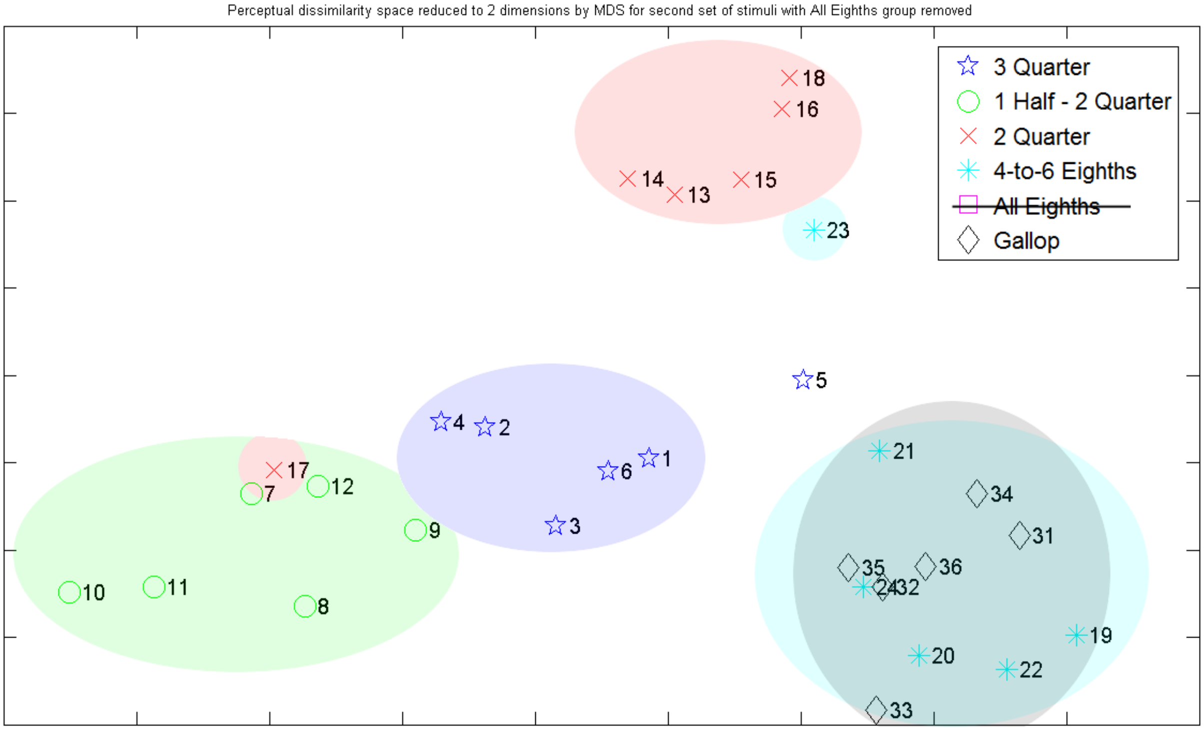

Figure 4.7: Subset MDS map stimulus sortings for groups 3Q, 1H2Q, 2Q, 4-6E and G, to better illustrate mutual differences between groups. The horizontal axis seems to reflect changes mostly in note length and the vertical axis seems to reflect changes in fullness.

Group 4-6E. The melodies in this and group G were sorted together with little distinction between the two. Melody 23 is close to group 2Q in these results because, we suspect, the two sets of 2 eighth notes with a rest in between feels like just 2 notes (the defining rhythm of group 2Q *was* just 2 notes). This suggests that participants were sensitive to the number of notes in a melody.

Group AllE. This group was nearly always sorted together by every participant. The rhythm for this group was very distinctive likely due to the saturation of notes in the bar and its continuous buzzing feeling. Melody 29 stands alone since it felt 'rolling' as opposed to 'abrupt'.

Group G. As mentioned above, these melodies were indistinguishable from group 4-6E. We suspect that the inclusion of eighth notes in the second half of the rhythm may have resulted in one of two things: groups consisting of rhythms with eighth notes are perceptually salient, lending them to be sorted with other groups consisting of such notes; or the inclusion of eighth notes in a rhythm makes it perceptually too complex to distinguish among its different melodies, therefore if more than one group is considered complex their melodies are sorted together.

Questionnaire results

The post-group questionnaires for this second study indicated that, like the first, sorting the melodies into 12 groups was the most difficult (average=2.0), 6 and 9 groups were the same (average=2.7), and 3 groups was the easiest (average=4.0).

Overall the average difficulty was reduced to 2.7 from 1.9 in the initial study. 4/6 participants thought that 6 was the best number of groups while 2/6 thought 3 was best. The strategies to sort melodies in this study centered around note rhythm. Only one user said amplitude played a factor and there were no indications that frequency or intensity of notes was used as a grouping strategy. Indeed, the number of notes, rhythm, pacing, and unevenness [44] were all considerations used by participants to sort the melodies.

As in study 1, demographic information was not considered.

4.7 Discussion

Generalizable results can be inferred from the detailed analyses of the two studies. Although the analysis is based primarily on a visual analysis of the plots, and is therefore subjective, this work is exploratory in nature and geared toward design guidance and the development of a usable stimulus set. In particular, the following trends appeared to dominate participants' perception of melodic haptic stimuli:

- *Rhythmic differences between melodies dominate other distinctions.* Our results show that participants consistently grouped stimuli based on non-syncopated rhythm, even if such stimuli have different amplitudes and/or frequencies (H1, H2).

- *Perceived quantity of notes is a major grouping factor.* Despite variance in rhythm, amplitude and frequency, melodies which contain a similar number of notes were often grouped together.
- *Groups of rapid (eighth) notes are perceptually salient in rhythms.* Distinct groups in our second study that contained eighth notes were often confused by participants. The authors conjecture that the complexity of these rhythms makes it difficult to identify mutual differences, leading to confusion. This supports Newman's claim that simple haptic stimuli are preferable [35]. The rhythmic domination is also supported as complex melodies contain quantities of notes that could be difficult to count.
- *'Abrupt' melodies are perceptually segregated from 'rolling' melodies even if they are devised from the same rhythm.* The holistic feeling of a note depends on its internal composition. Alternating between high and low amplitude notes on the same frequency makes a stimulus feel 'rolling' (Section 4.4.5).
- *Items that only differ in phase are grouped together.* The rhythm of a stimulus is perceived rather than its onset.
- *In some circumstances, replacing a quarter note with two eighth notes can increase expressiveness while maintaining groupability.* When we replaced a non-emphasized quarter note with two eighth notes, it was often grouped with its counterpart. This is not true for the case where the emphasis is on the two eighth notes.
- *Stimuli that only vary by frequency and amplitude are perceptually discriminable.* These stimuli were often, but not always grouped together (no occluding points in MDS plot). This indicates that people can perceive differences between these stimuli. This lends evidence to H3, however the exact degree is not conclusive and further investigation is required.

These trends suggest the following design heuristics for family-based design of melodic haptic icons:

1. **Rhythm primary, amplitude and frequency secondary.** Group stimuli based on simple and distinctive non-syncopated rhythms, and modify the amplitude and frequency of individual notes for within-group variation.
2. **Abruptness**. Design groups to be either *abrupt* or *rolling*.
3. Be mindful of **periodicity** when dealing with rhythmic stimuli.
4. As long as they are not emphasized, **quarter notes can be replaced with two eighth notes** for within-group variation.

This work provides the community with guidance for developing melodic haptic stimuli, as well as a perceptually validated, groupable, and expressive set of haptic stimuli for which learnability can now be determined.

4.8 Contributions

In this chapter we presented a process for the design of melodic haptic stimuli with the goal of making haptic stimulus sets larger, more expressive, and more learnable. To assess how the designed stimuli were perceptually grouped, we described two iterative user studies.

From these studies, we suggest several design heuristics for family-based icon design. Stimuli from the same family should have a common, non-syncopated rhythm and should be perceptually similar in terms of abruptness (versus being 'rolling'). Rhythm is fairly powerful in the melodic design space as a grouping characteristic, though syncopation should be avoided since it lessens the resulting grouping effect. Designers should also be mindful of periodicity and can use un-emphasized groups of eighth notes to expand the expressiveness of their melody.

Despite these findings, the learnability and the exact degree of within-group discriminability of melodic haptic stimuli has not been evaluated. In the next chapter, we will attempt to evaluate the viability of rhythm-grouped melodic haptic stimuli in a longitudinal learning scenario, similarly to Chapter 3. Performance during a learning task will provide insight into within-group discriminability since learnability relies on a unique and distinct percept.

CHAPTER 5

LONGITUDINAL EVALUATION OF RHYTHM-GROUPED MELODIC HAPTIC ICONS

Our discussion thus far has revealed that rhythm-grouped melodic haptic icons have three advantages over strictly rhythmic haptic icons: icon grouping, increased expressiveness, and a large design space.

First, melodic haptic icons are more easily groupable, which has better support for family-based design. As described in Section 2.2, family-based design allows users to partition their learning into more easily manageable 'chunks' [34], thus increasing learnability. Ternes and MacLean revealed one very effective grouping characteristic in rhythmic icons; however this particular perceptual axis divides the stimulus set into two discernible groups [43]. In Chapter 4, we discovered that the grouping characteristic of rhythm is fairly powerful in the melodic design space. Thus, grouping stimuli by rhythm and then varying the melody through individual frequency and amplitude variation is a promising technique for creating perceptually similar groups of icons that can be assigned a related semantic component for enhanced learnability.

Secondly, melodic haptic icons are more expressive than monotone, strictly rhythmic haptic icons. As described in Chapter 4, melodic icons are desirable since designers can develop richer, more elaborate stimuli. The use of melody in haptic stimuli may allow for an increased likelihood of an emotion-evoking stimulation, resulting in a more memorable set. In Section 3.5.2, we showed that recognizable, emotion-evoking rhythms are more memorable. Based on musical theory, we believe that melodic haptic icons will have a similar, emotion evoking effect since certain scales or patterns in music can evoke very palpable emotions [41].

Finally, the design space of melodic haptic icons is exponentially larger than the strictly rhythmic design space due to between-note variations in frequency and amplitude. Through heuristic design, Ternes and MacLean were able to identify 21 distinctive stimuli to which semantic attribution could be performed. If we constrain ourselves to their 21 rhythms and 2 amplitudes and frequencies, melodic variation allows for 71,616 individual melodies, an exponential increase.

Before exposing new design parameters for haptic icons to users through deployment, one must first evaluate their efficacy in a controlled experiment that is able to approximate the target environment. In this chapter, we will evaluate the efficacy of the rhythm-grouped melodic haptic icons that we designed in Chapter 4 in a realistic deployment scenario identical to Chapter 3. Our analysis of this second longitudinal study of haptic icons will focus around a comparison between the two design paradigms in an effort to understand which type of icon is more easily learnable, and which type of icon is currently more viable for deployment.

This chapter has a similar structure to Chapter 3. We will first describe the haptic icon design process, then we will outline the procedure of the longitudinal experiment conducted. Next, we present our results and analyze these results with an emphasis on a comparison between the results obtained in this chapter, and those presented in Chapter 3. Finally, we summarize the contributions of the research described in this chapter.

5.1 Haptic Icon Design Process

Our approach to designing the haptic icon set used within this study is very similar to the process outlined in 3.1, except we employ melody as a design parameter and use rhythm as the primary grouping characteristic. In other words, melody was used to distinguish otherwise identical rhythms.

5.1.1 Haptic Stimulus Design

Our stimulus design process is described in detail in Chapter 4. For this work, we use all of the stimuli created there, with a few modifications to enhance perceptual coherence with respect to icon groupability.

First, we remove the ‘Gallop’ group entirely since it conflicts strongly with the ‘4-to-6 Eighths’ group. This leaves us with 5 groups of 6 stimuli, for a total of 30.

Next, we modify the remaining perceptual anomalies found in our modified set in Section 4.6.1, Figure 4.7.

Stimulus 17 from the ‘2 Quarter’ group is found near the ‘1 Half 2 Quarter’ group since it differed from icons within that group only by *phase*, and the low amplitude half note was not perceived as a rest. To replace this stimulus, we designed a melody that employs heuristic 4 (4.7), and replace a quarter note with two un-emphasized eighth notes. The new stimulus is shown in Figure 5.1. The high frequency, low amplitude note paired with the low frequency, high amplitude note has the perception of a single quarter note that rises in intensity.

Stimulus 23 from the ‘4-to-6 Eighths’ group is perceptually located next to the ‘2 Quarter’ group since, we suspect, the two sets of un-emphasized eighth notes feel like two distinct quarter notes. We modified this stimulus using the converse of heuristic 4 (4.7)and replace the two sets of un-emphasized eighth notes with a quarter note and a set of *emphasized* eighth notes. This stimuli should feel like it has 4 distinct eighth notes. It is shown in Figure 5.2.

With the removal of the 'Gallop' group and the perceptual anomalies found, we are ready to perform the meaning attribution process with confidence that our stimuli are well-designed from a perceptual standpoint, despite the fact that we have not re-tested our stimuli since we followed the heuristics outlined in Chapter 4.

5.1.2 Meaning Attribution

For the meaning attribution process, we take a similar approach as described previously in Section 3.1.2. However, the icons in a group share a similar rhythm, where previously groups were defined by a rhythm's evenness. In addition, we have 5 distinct perceptual rhythm-based clusters (Section4.6.1, Figure 4.7) instead of only 2. We utilize the same deployment context and grouping strategy, while the role of amplitude and frequency becomes moot since we are now using melody as a design parameter. Due to the increase in the number of perceptual groupings, we must introduce new senders. Our meaning assignment strategy is outlined in the following subsection.

Meaning assignment strategy

As a result of our stimulus design process (Section 5.1.1), we have 5 perceptual groupings, each with 6 constituent stimuli for a total of 30. As described above, we choose 5 different ecologically plausible senders and assigned each a group of stimuli. For our group assignment technique, we made our best approximation of which senders would send the most urgent messages, and assigned the most rapid rhythms to the more urgent senders. Conversely, we assigned the least urgent senders to the least rapid messages. This assignment technique is in line with the heuristics described in Section 3.7.1. The 5 different senders are listed in Table 5.1 along with their associated perceptual group (Chapter 4, Section 4.5):

Sender	Rhythm Group
Spouse	3 Quarter
Child	1 Half 2 Quarter
Friend	2 Quarter
Babysitter	4-to-6 Eighths
Boss	All Eighths

Table 5.1: List of haptic icon senders and their associated rhythm-based perceptual grouping.

5.2 Haptic Icon Set

In Figure 5.3, we display a visual representation of the 30 melodic stimuli and their associated meanings. As you can see, they are grouped into 5 groups of 6 and each group contains an assortment of messages from *Spouse*, *Child*, *Friend*, *Babysitter* and *Boss* (this will be further detailed in Section 5.3.5). In total, there are 6 messages from each sender, with each sender being associated with a specific rhythmic grouping (Section 5.1.2).

Please note that the numbering is different from the stimuli displayed in Figure 4.6. The stimuli have been shuffled so that the base rhythms of the icons are distributed between two different icon groups.

We arrange the batches (defined in Section 3.3.5) so that there are two different senders in each batch, with 3 randomly chosen messages from each sender. The introduction of each sender is distributed such that each sender is encountered as soon as possible for both counterbalancing arrangements (explained in Section 5.3.5).

Each icon has a unique and representative 4 character code which is used to identify the icon in the experiment (Section 5.3). For instance, the message '*Child: I aced the test!*' has the code '*CIAT*'. Note that the codes contain only alphabetic characters (as opposed to symbolic and alphanumeric in Section 3.2) since there are no attribute modifications (Section 3.1.2).

5.3 Experiment

To assess the learnability of the melodic set we performed another longitudinal study that is nearly identical to the one outlined in Section 3.3 for purely rhythmic icons. Now that we have obtained additional information about the learnability of large set of icons, we have new objectives and research questions. Obviously, we also have a new set of participants and the icon presentation ordering has changed.

This experiment was conducted before the results from the study conducted in Chapter 3 was fully analyzed, therefore we were unable to apply all of the heuristics outlined in Section 3.7 to this set. However, this fact makes the comparison less biased.

5.3.1 Objectives and Research Questions

The main research objective of this experiment is to compare rhythm-grouped melodic haptic icons against entirely rhythmic haptic icons, which were discussed extensively in Chapter 3. We seek to determine which type of icon is more easily learnable in order to provide suggestions for the utilization of haptic icons in a deployment scenario.

In order to guide our investigation, as in Section 3.3, we have formulated the following research questions:

- Which type of icon is more easily learnable?
- Which type of icon is more viable for deployment?

The question of learnability has two parts. Since perceptual distinctiveness is a prerequisite to learnability, we would like to determine whether rhythm-group melodic haptic icons are more easily distinguishable than purely rhythmic icons. In addition, we would like determine whether the addition of improved groupability for rhythm-grouped melodic icons contributes to the formation of cognitive scaffolding, thus resulting in faster learning over purely rhythmic icons.

5.3.2 Hypotheses

In the following subsections, we will elaborate on our hypotheses for each of the research questions outlined in Section 5.3.1.

Which type of icon is more easily learnable?

It is very difficult to predict which type of icon will be more easily learnable. We predict one of three outcomes will result from our experiment:

1. *Rhythm-based melodic icons will be drastically easier to learn than rhythmic icons.* This outcome may occur since the cognitive partitioning caused by easily recognizable stimulus groups will overcome the 'bump' in learning observed for purely rhythmic icons and discussed in Section 3.6.3. Grouping icons based on rhythm will cause a clear separation between different senders, and participants will only have to learn to recognize between 3 different icons for each sender. Participants will require less elaboration to learn the icons due to this partitioning and will learn icons much more quickly.

2. *The differences between rhythm-based melodic icons from the same sender will be too difficult to perceive, causing learning difficulty to increase.* As discussed in Sections 3.5.5 and 4.7, perceptual MDS does not give a very clear indication of perceptual differences within stimulus groups. Although we believe that the technique is well suited to mapping perceptual space based on the differences between fairly unique stimuli, mapping the differences between similar stimuli is a much more difficult task. In a cluster sorting study, participants are asked to sort stimuli based on similarity, thus similar stimuli will naturally be often sorted together. However, these studies do not test if participants will be able to perceive the differences between stimuli in isolation. As discussed extensively in Chapter 3, ease of learning relies heavily on perceptual distinctiveness. If rhythm dominates perception too heavily, then the differences between melodies with the same rhythm will not be perceivable.

3. *A mixture of the above two.* Since we observed some individual differences in the ability to distinguish haptic icons in Chapter 3, we predict that some users might be able to easily distinguish between icons with the same rhythm, and thus their learning speed will increase due to the groupable nature of the icons. Conversely, some users may not be as sensitive to the differences between the icons, and may have difficulties progressing due to this perceptual handicap.

Which type of icon is more viable for deployment?

The exact nature of our hypothesis for this research question depends heavily on the outcome of the learnability comparison. If rhythm-based melodic icons are more easily learnable, then they are definitely more viable for deployment due to the reasons outlined in the introduction of this section. As a reminder, these reasons are: icon grouping; expressiveness; and a larger design space. We hope that the resulting diversity expressed by our medium – the haptic feedback device (Section 1.1.2)– is perceptually salient to humans.

If the differences between rhythm-based melodic icons are too difficult to perceive, causing a problematic learning process, then we would conclude that purely rhythmic icons are more viable for deployment until the research community can ascertain how to design rhythm-based melodic icons that have perceptually distinctive melodic components.

5.3.3 Participants

Participants were recruited from the Department of Computer Science at the University of British Columbia. There were 10 participants in total, all aged from 18 – 31. Eight participants were male and two participants were female. All participants were undergraduate or graduate students in the Department of Computer Science. There was one participant that grew up in each of the following countries: Russia, Bangladesh, India, Brazil, and Taiwan, while two participants grew up in Iran and three in Canada. One participant had <2 years experience playing a musical instrument, six had 3–9 years of experience and three had >20 years of experience. Six participants reported having a very good sense of rhythm and four participants reported having a decent sense of rhythm.

5.3.4 Compensation

The compensation scheme is identical to the one described in Section 3.3.4.

5.3.5 Icon Group Ordering

Icon groups were presented to participants in two different orders, thus each counterbalancing set of users contained 5 people.

Table 5.2 enumerates the ordering in which the 2 different arrangements of participants encountered the icon groupings. It is helpful to examine Figure 5.3 while viewing this table.

Batch	User Set 1		User Set 2	
	New Group	Old Groups	New Group	Old Groups
1	1		5	
2	2	1	4	5
3	3	1, 2	3	4, 5
4	4	2, 3	2	3, 4
5	5	3, 4	1	2, 3

Table 5.2: Icon group ordering presentations for each counterbalancing arrangement of users in longitudinal study for melodic icons.

5.3.6 Apparatus

The apparatus is identical to that described Section 3.3.6.

5.3.7 Procedure

The entire procedure is identical to that described in Section 3.3.7.

However, the user now must achieve a performance level of 80% or higher. This is due to the fact that batches only contain 6, rather than 7 icons – the number of icons that a participant is permitted to answer wrongly remains constant. In addition, the history of icons responded to during the game has changed from 14 to 12 so that the number of permissible mistakes remains constant between studies.

The participants were provided with the two code lists shown in Appendix D based on their counterbalancing arrangement.

5.4 Results

In this section, we describe the results and statistics obtained from the experiment illustrated above. The layout and structure of this section is different from Section 3.4 since the goal of the study is not to explore the properties of icons that make them learnable and the properties of users that make them proficient, but rather to compare the learnability of rhythm-based melodic icons with purely rhythm-based icons. For this reason, we will separate results pertaining to perceptual differences between icons from results pertaining to the learning process. We use the same visualization techniques that are explained in Section 3.4.1.

For the purpose of this section, all of the results displayed are based on performance during the icon identification task while playing the Tetris (Section 3.3.7) game unless stated otherwise.

Unfortunately during this experiment, we experienced a ceiling effect with some users. Users 7, 8 and 10 passed the final batch (batch 5) before the end of the experiment, which lasts for 12 sessions. Users 7 and 8 passed batch 5 in session 11 and user 10 passed batch 5 in session 8. In the event that a user passes the final batch, they repeat the same batch until the experiment has ended. For all results presented in this section, we have removed all response instances during sessions after a user has completed batch 5. In other words, session 12 is removed for users 7 and 8, and sessions 9-12 are removed for user 10.

An analysis and interpretation of the results will follow in Section 5.5.

5.4.1 Results Pertaining to Perceptual Differences Between Icons

Confusion between icons

In Figure 5.4, we show the confusion matrix for all of the responses on all 30 icons. The results are summed over all users and all sessions, with each response instance acting as a data point; thus, these gross statistics do not visualize differences between users and their progression through the experiment (except for the responses omitted due to the ceiling effect). This matrix is read the same way as in Section 3.4.2.

As one can observe, many icons – such as icons 13 and 30 – had very few responses instances recorded for them. We believe that this effect was caused by normal inconsistencies in sampling due to random number generation. The average number of encounters for each icon is 50.

The mean icon accuracy is 79% with a standard deviation of 15% (Icons = 30), comparable to a 79% mean icon accuracy and standard deviation of 12% in Chapter 3. Icon 27 had the lowest mean accuracy with 41.5% (comparable to 50.5% in Chapter 3) and icons 4 and 24 had the highest with 100% (comparable to 99.1% in Chapter 3.

Confusion between senders

In order to determine what kinds of mistakes participants make with respect to identifying the sender of an icon, we average the proportion of the time that a user identifies the sender correctly in the event of a mistake and the proportion of the time that the user identifies a sender incorrectly. Note that this measure does not take into account instances where the user responded correctly. We found no significant differences between the average likelihood of each class of misidentification; $F(1,18) = 1.644$, $MS = 0.066$, $p = 0.216$. In other words, in the case where the user made a mistake, we found no significant difference in the likelihood of erring in their rhythm identification compared to their melody differentiation.

5.4.2 Results Pertaining to the Learning Process

Batch progression

For a general idea of what the learning curve was for haptic icon learning, we examine Figure 5.5. This figure shows how many sessions it took users, on average, to complete each batch.

To test for statistically significant differences between the number of sessions needed to complete each batch, we conducted a One-Way ANOVA with multiple comparisons using Tukey's HSD. We observe no significant differences between the average number of sessions required to complete each batch; F(4,27) = 1.284, MS = 7.671, $p < 0.301$, uses harmonic mean sample size of 5.3.

As you can see from Figure 5.5, only 8 out of 10 participants finished batch 1, therefore it is difficult to determine the leaning curve for more proficient users. In Figure 5.6, we show the average number of sessions required to complete a batch for the top 8 users.

As you can see from Figure 5.6, the average number of sessions required to complete batch 1 is reduced to 2.75, and the variance decreases. However, we do not find any significant differences between the average number of sessions to complete each batch; F(4,25) = 1.419, MS = 2.863, $p < 0.257$, uses harmonic mean sample size of 5.1.

Table 5.3 shows the number of batches completed by each user.

User	1	2	3	4	5	6	7	8	9	10
Batches Completed	2	4	3	3	1	3	5	5	1	5

Table 5.3: The number of batches completed by each user.

Figure 5.7 shows the questionnaire results, averaged by user, over all 12 sessions of the study. Recall that the questionnaire questions are as follows:

1. How easy did you find the new messages to remember? (Easy – Hard)
2. Did the mapping of the haptic vibrations you felt and their meanings make sense to you? (No – Yes)
3. Were the haptic vibrations you felt easy to distinguish between? (Easy – Hard)

From this plot, we cannot observe any trends for any of the questions, therefore we will not discuss them further.

Learning curves for each participant

Figure 5.8 shows the cumulative average accuracy for each participant over each response instance. The highlighted users are the subject of discussion and are emphasized for the purposes of readability. The curves for the remaining users are shown to display the range of learning abilities and final cumulative accuracies.

As one can see from Figure 5.8, there is much variation in final accuracy between users; however, the distribution is fairly even in the range of 62% to 85%. User 5 is also notable since he/she completed the study with the lowest cumulative average performance: 46%. User 10 completed the study with the highest cumulative average performance of 90%.

Users 4, 8 and 10 experience a drastic decrease in performance after they have completed a certain number of batches.

It is also important to note the effect on learning once a user has completed a batch. In many cases, performance drops after a new batch is reached. This is phenomenon is most easily exemplified by user 4 and 7's learning curve.

User 5's learning curve stayed relatively constant at around 45% performance.

Icon retention

In Figure 5.9, we plot the number of icons answered correctly by users during the final quiz, administered approximately one week after all participants have completed their 12 independent sessions. We have removed participants 1, 5 and 9 since they did not experience all 24 icons tested due to the fact that they did not pass batch 3. Those users experienced between 6 and 12 icons.

In Figure 5.9, we can observe that people correctly identified, on average, about 16.7 icons under no workload about one week after twelve 20 minute sessions over four weeks.

The exact number of icons learned is difficult to express from experimental results since participants were constantly learning. Participants experienced between 6 and 30 icons, with overall cumulative average accuracy ranging between 46% to 90%.

Performance during the game vs. quiz

In order to determine whether there is a significant difference in the icon identification performance during the game versus during the quiz, we conducted a One-Way ANOVA. We found that the mean response accuracy during the quiz (80.4%) was marginally larger than the mean response accuracy during the game (71.9%); $F(1,18) = 3.075$, $MS = 0.036$, $p = 0.097$.

Effect of counterbalancing

To test if there was a statistically significant effect of the batch presentation ordering on account of the counterbalancing, another One-Way ANOVA with multiple comparisons was conducted. No significant difference between the mean response accuracies for each counterbalancing arrangement was found; $F(1,8) = 0.081$, $MS = 0.001$, $p = 0.783$, $N = 5$.

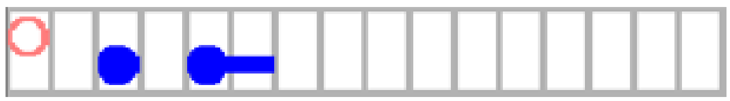

Figure 5.1: New stimulus #17. The hollow notes are low amplitude and the blue notes are low frequency. Conversely, the filled notes are high amplitude and the red notes are high frequency.

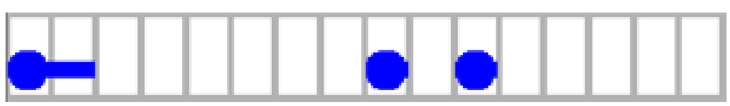

Figure 5.2: New stimulus #23. The hollow notes are low amplitude and the blue notes are low frequency. Conversely, the filled notes are high amplitude and the red notes are high frequency.

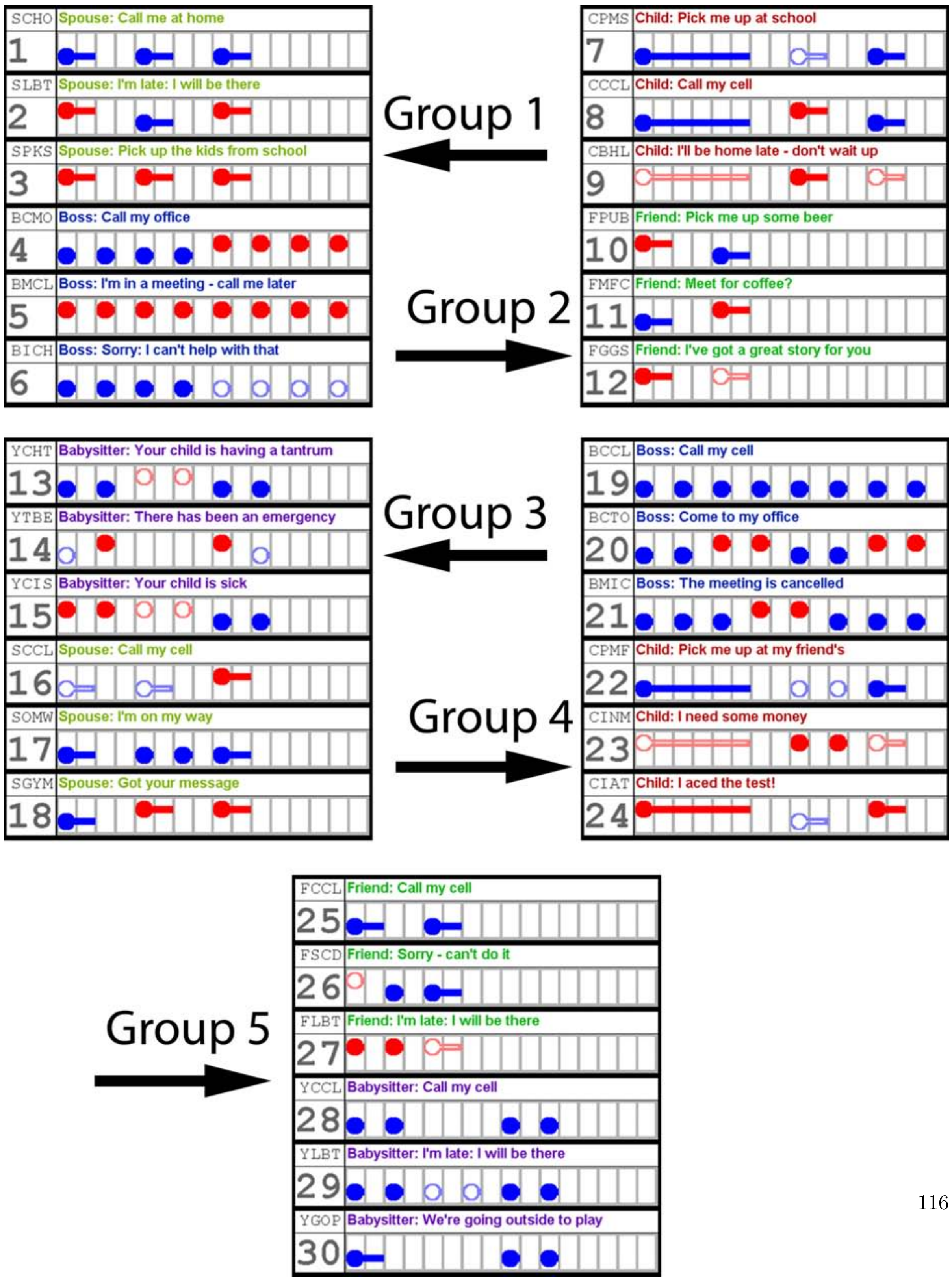

Figure 5.3: Visual representation of the 30 melodic stimuli and their associated meanings, separated by group.

	N	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
1	64	61%	20%	13%																	5%			2%							
2	69	36%	55%	7%																	1%										
3	63	29%	33%	30%											2%						3%			3%							
4	21				100%																										
5	77					83%	1%					3%											8%					5%			
6	94					2%	91%																4%					2%			
7	23	4%						74%			13%				4%									4%							
8	21		10%						90%																						
9	33									73%		9%															3%	15%			
10	41							2%			88%			2%												2%		5%			
11	105					7%	1%			1%	2%	54%											2%	1%		19%		13%			
12	19											5%	68%					5%				11%							5%	5%	
13	33													82%				3%	9%	6%											
14	29														86%	10%	3%														
15	42						7%					2%				81%	7%										2%				
16	42						2%				2%				5%	2%	86%							2%							
17	66																	56%			12%			30%	2%						
18	29										3%		3%	14%					76%	3%											
19	34													3%			3%		26%	68%											
20	75	3%	7%	1%														4%			81%			4%							
21	14												7%									93%									
22	83				2%	14%	1%				1%												71%	2%		5%		2%			
23	70	1%																29%			6%			64%							
24	20																								100%						
25	91					1%				2%		11%											3%			53%	2%	27%			
26	29												3%														86%		3%	7%	
27	95					5%				2%		18%											2%	1%		36%		36%			
28	30					3%	3%						3%																83%	7%	
29	35			6%																							3%			91%	
30	38									3%														5%							92%

Figure 5.4: Confusion matrix for all 30 rhythms for responses during the game. Calculated by summing over all response instances, irrespective of the user. Icons within each thick blue box belong the same group of icons (6 per group). The color of the icon indicates its sender with the following mapping: Yellow-Spouse; Blue-Boss; Red-Child; Green-Friend; Purple-Babysitter. N is the total number of responses counted.

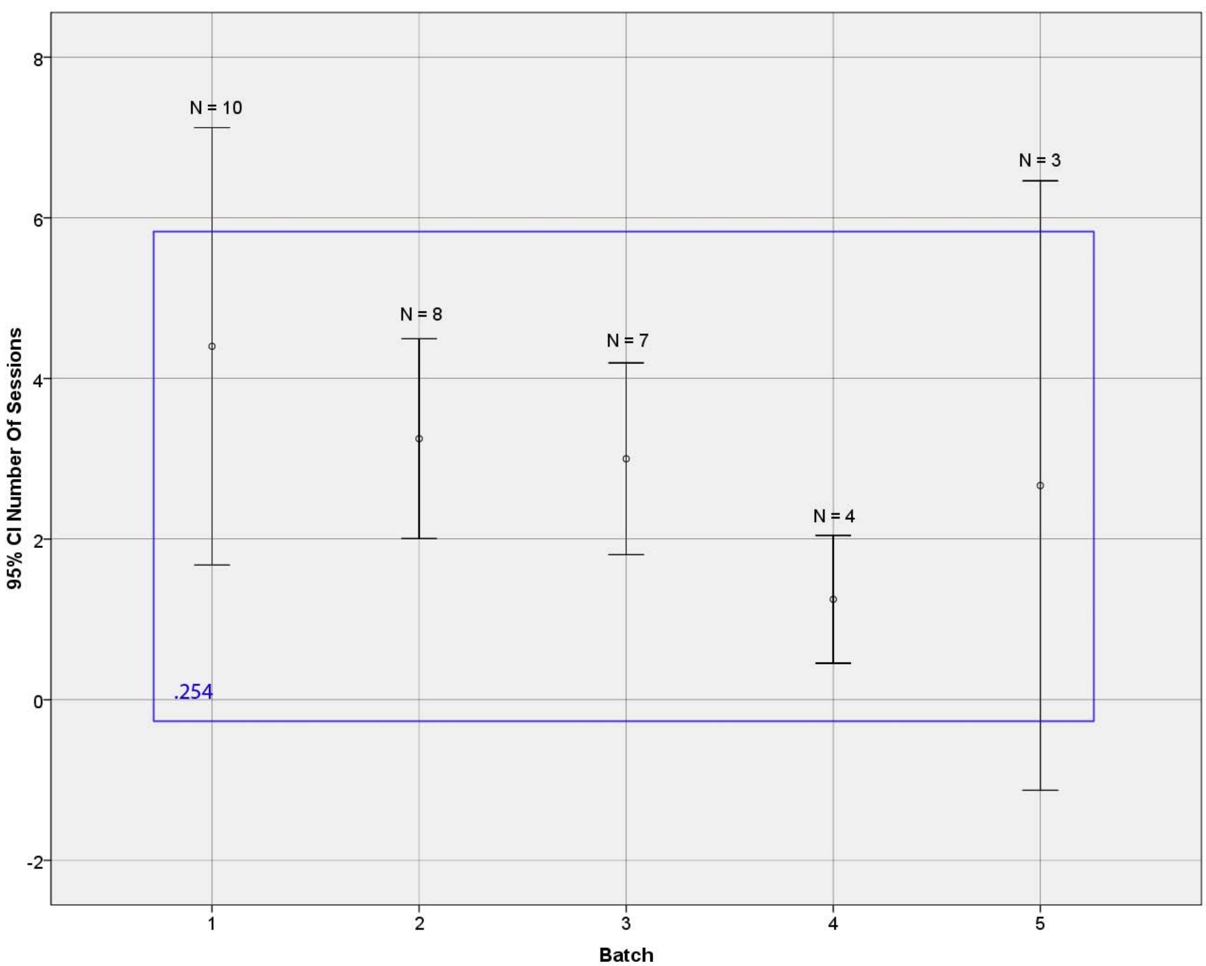

Figure 5.5: Average number of sessions required to complete each batch. Tukey's HSD test uses harmonic mean sample size of 5.3.

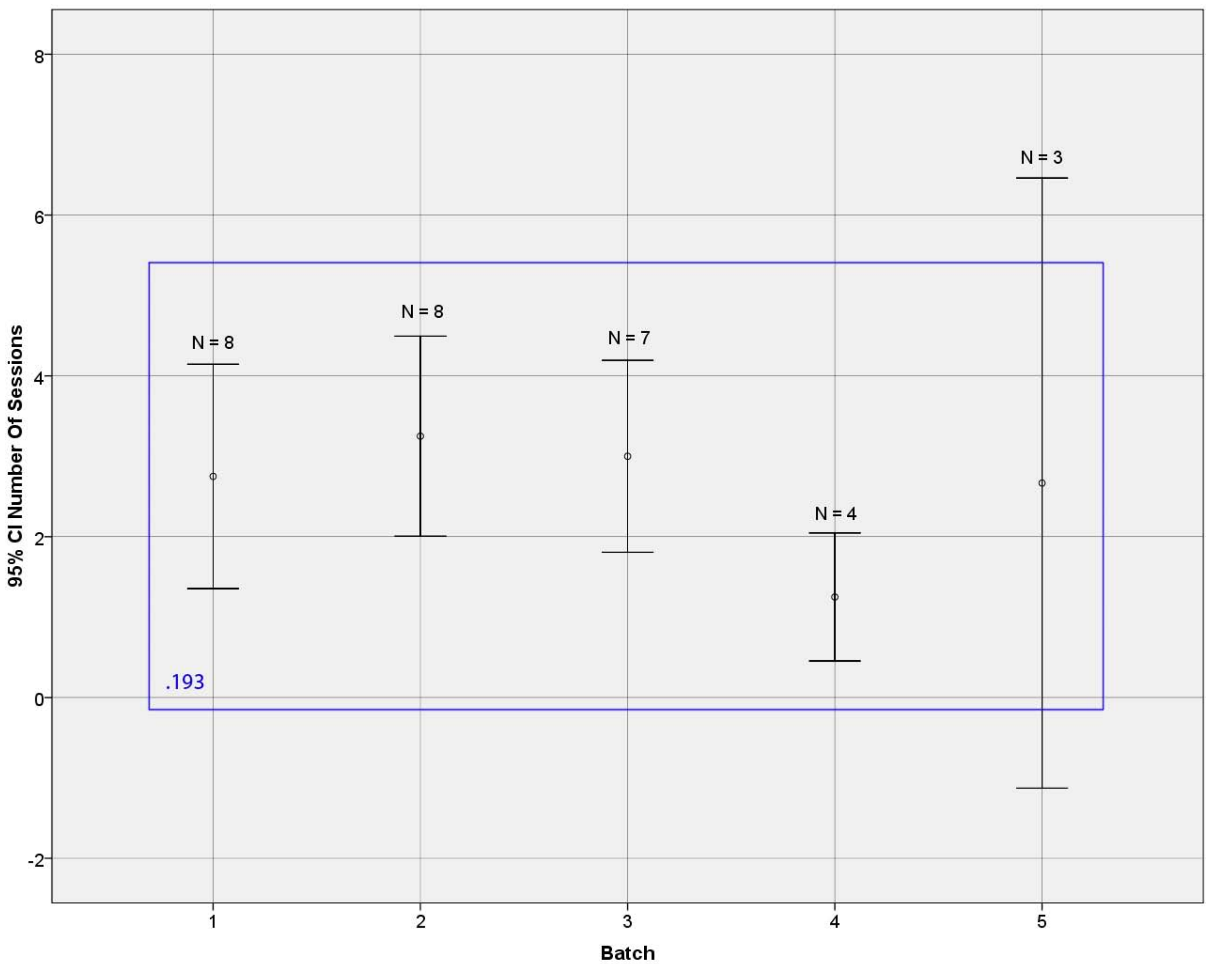

Figure 5.6: Average number of sessions required to complete each batch. Only the top 8 participants are plotted.

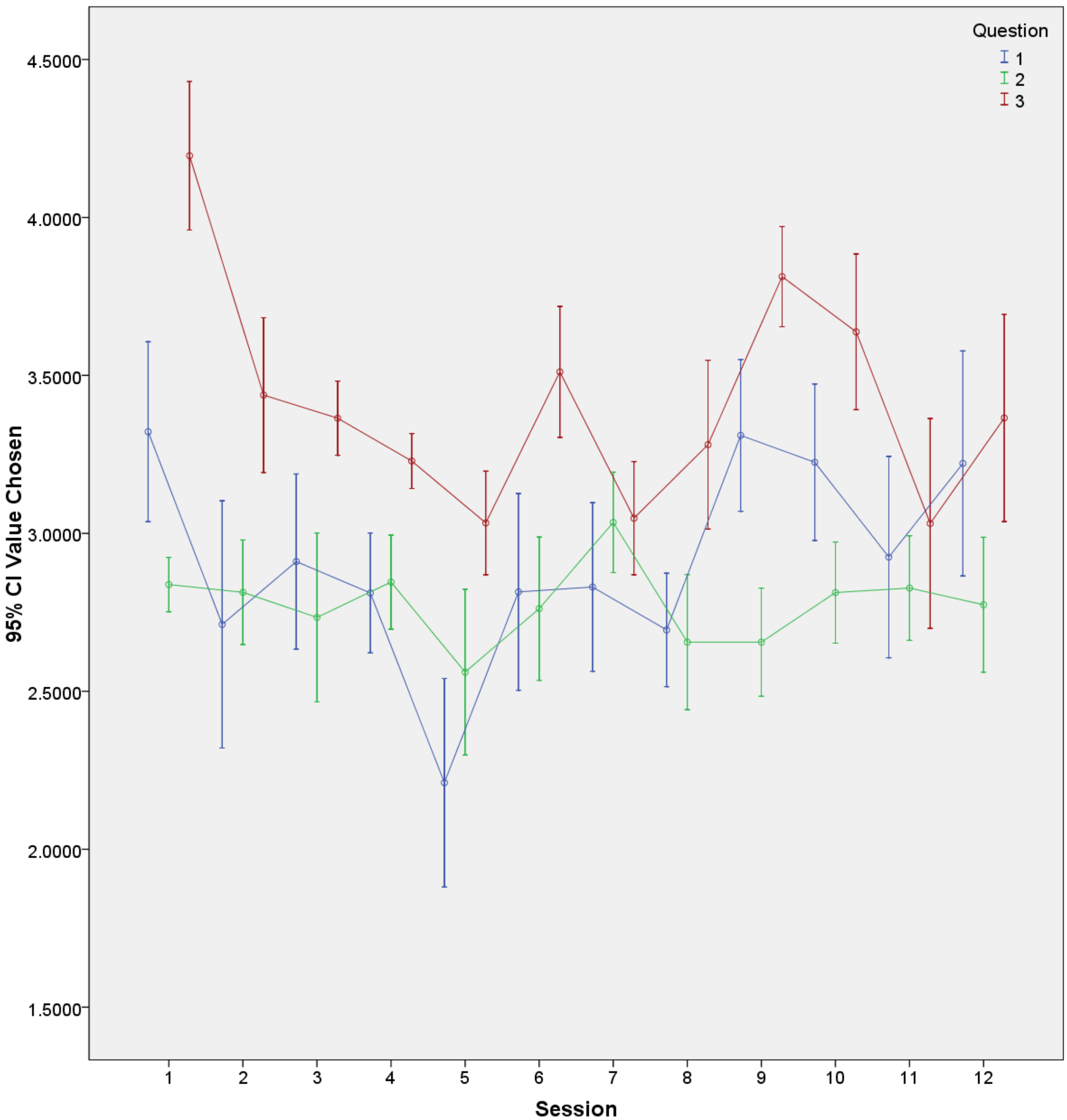

Figure 5.7: Questionnaire results, averaged by user. 95% confidence intervals are shown, N = 10.

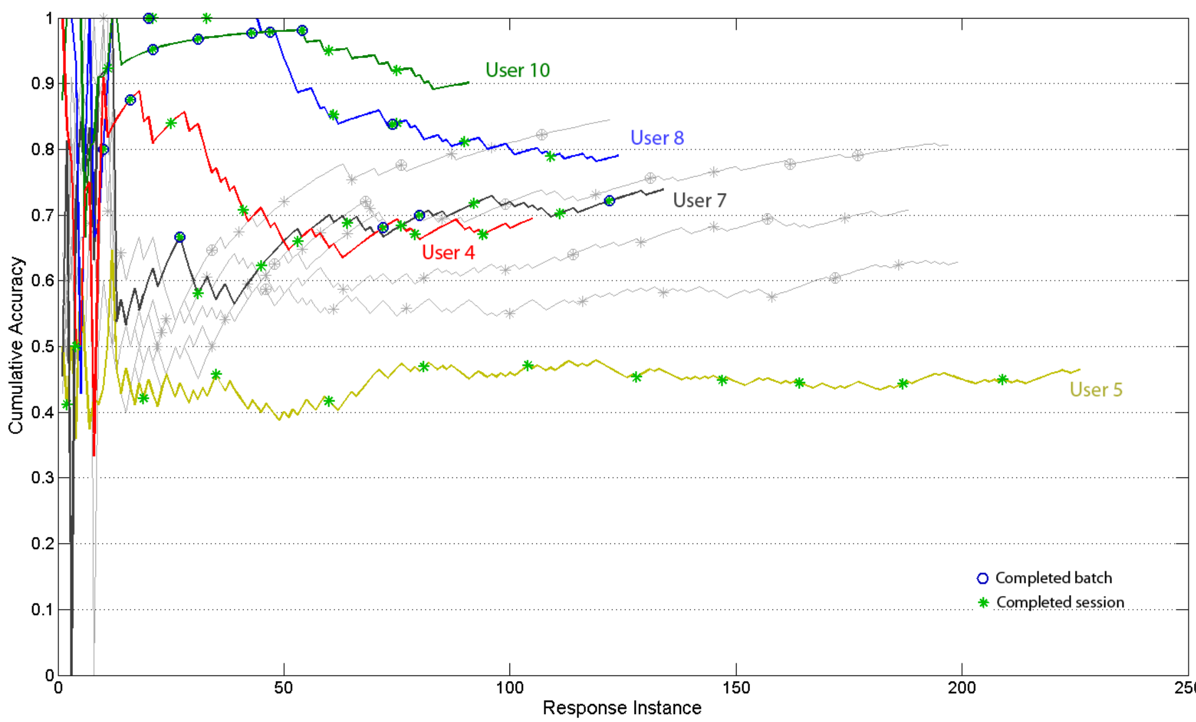

Figure 5.8: Learning curves for each participant. The cumulative accuracy is calculated for each response instance. Highlighted users are the subject of discussion and are emphasized for the purposes of readability. Blue circles indicate where a participant completed a batch and green stars indicate where a participant completed a session.

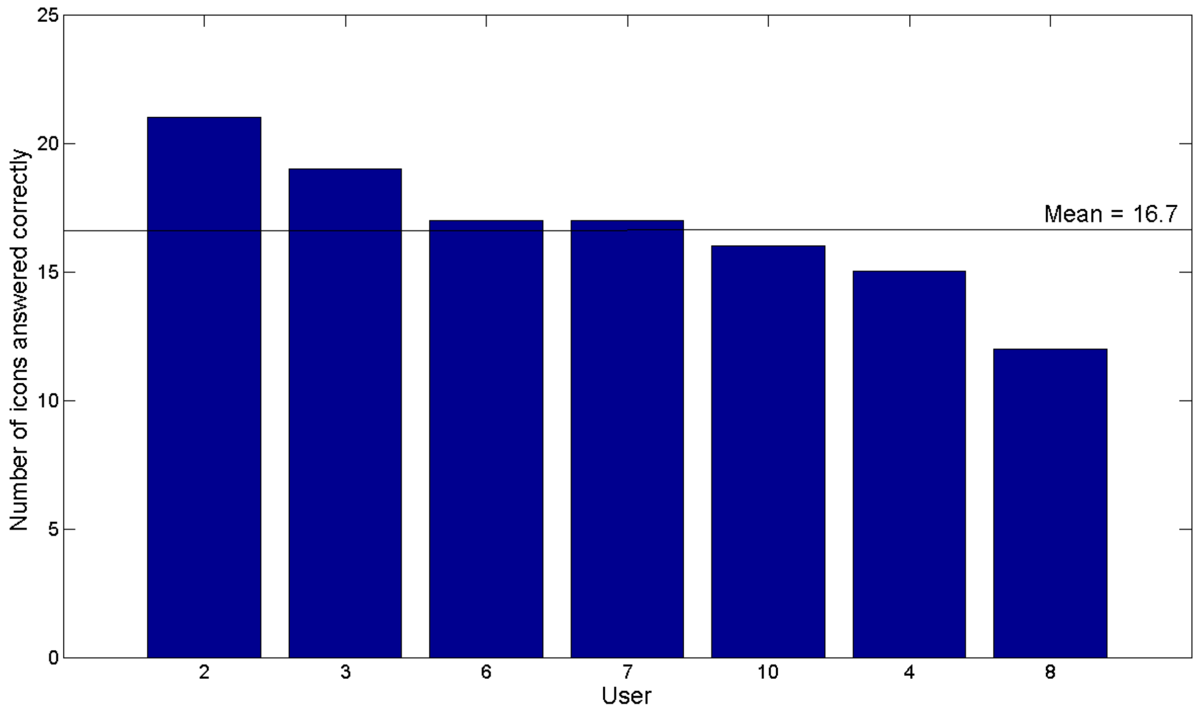

Figure 5.9: Number of icons answered correctly during final quiz for each user, sorted by descending accuracy. Participants 1, 5 and 9 are removed.

Demographics

We did not find any significant differences between any demographic groups relating to age, gender, musical experience, country of origin or self-reported sense of rhythm.

5.5 Discussion

In this section, we will analyze and discuss the results outlined in Section 5.4. We aim to understand the differences between purely rhythmic and rhythm-based, melodic icons, with a focus on learnability and their viability in deployment. Although we could provide an in-depth analysis of rhythm-based melodic icons similar to in Chapter 3, given our results, we believe that a comparison-based approach would be more informative.

Similarly to Sections 3.5, and 3.6, we will frame our analysis and discussion around the research questions revealed in Section 5.3.1.

5.5.1 Which Type of Icon is more Easily Learnable?

We believe that the success of haptic icons in deployment relies most heavily on their learnability. Here, we compare the results obtained through the longitudinal study for melodic icons with the results discussed previously (Sections 3.4, 3.5 and 3.6) to determine which of the two icon sets are more easily learnable.

As discussed in Chapter 3, the learnability of haptic icons depends heavily on their mutual perceptual distinctiveness. In order to organize our discussion of learnability, we will first discuss the perceptual characteristics of the rhythm-based melodic icons in comparison to the purely rhythmic icons and then discuss the learning process observed.

Perceptual distinctiveness

In Figure 5.4, we displayed the confusion matrix for all of the responses made by participants during the course of the experiment.

If we were to make a table similar to Table 3.8 (page 64), where we list all of the mistakes that were made over 10% of the time, it would have 22 rows covering 15 target icons as opposed to 9 rows covering 9 target icons. If we normalize for the number of icons in the set (22/30 vs. 9/21), this is still 1.7 times as many common mistakes.

Although roughly the same proportion of the set has an instance of a commonly made error, 8 of the mistakes in the melodic set have a larger frequency than the highest observed confusion frequency in the purely rhythmic set [3]. However, the mean frequency of these common mistakes is approximately equal at 18% for the melodic set and 17% for the rhythmic set.

The summary statistics expressed above have two implications for the perceptual distinctiveness of rhythm-based melodic haptic icons.

First, in the event that icons are perceptually similar, the confusion between them can be severe. For instance, there is severe confusion between icons 1, 2 and 3, which all share the same rhythm. Icon 3's accuracy is only 30%, which is approximately chance if we consider the rhythm of the group to be salient (chance is $\frac{1}{3}$ since there are three icons that share a rhythm in each group). Icon 25 is mistaken for icon 27 27% of the time, and the converse is true 36% of the time. As one can see due to their frequency, these mistakes are much more problematic than the worst mistake observed in Section 3.5.1.

Second, there are more total instances of common confusions between icons. This means that it is more difficult to design icons that are perceptually distinct within the context of rhythm-based melodic icons than it is with purely rhythmic icons. As outlined in Chapter 4, the melodic icons were carefully and iteratively developed using heuristics and a series of two separate cluster sorting studies. Despite this fact, our methodology did not capture numerous sources of common confusions, as well as some extremely significant sources of confusion. The purely rhythmic icons used in Chapter 3 were designed using a similar methodology.

The common confusion between icons 17 and 23 is an interesting case. Icon 17 is mistaken for icon 23 and vice-versa about 30% of the time. Recall from Section 3.5.2 that the base rhythm for icons from *Child* is extremely recognizable. Unfortunately, when we add the two eighth notes, the rhythm's unique effect is lost and it becomes a variation on 17 based on timing and note length, as well as amplitude for the eighth notes. As explained previously, this experiment was conducted before the results from Chapter 3 were fully analyzed; therefore did not predict that varying rhythms based solely on timing and note length would cause a significant source of confusion. It is apparent that the amplitude variation on the eighth notes was not sufficient to provide adequate disambiguation. In this case, we failed to follow a design guideline described by Andrew Chan in [11]. When using MDS analysis as a design tool, always be sure to evaluate the final stimulus set after any changes have been made.

In Section 5.4.1, we showed that there was not a significant difference between the mean frequency where users misidentified an icon but identified the sender correctly (same rhythm), and the mean frequency where users identified the sender incorrectly (different rhythms). Since there are 5 senders, if the grouping characteristic of rhythm did not have any effect, then the frequency where users identified the sender incorrectly (different rhythms) would be larger by a factor of 4. Based on this result, we can conclude that the grouping characteristic of rhythm did have some effect. However, due to the fact that mistakes where the user identified the sender correctly (same rhythms) did not dominate perception, we conclude that the rhythmic grouping characteristic did not dominate perception to the level predicted

[3] For clarification, 'frequency' refers to the observed proportion of a particular icon identification response out of every response made by every user for a specific target icon.

– as a grouping dimension, rhythm was not nearly as effective as the evenness/unevenness dimension utilized in Chapter 3.

Despite the above result, the grouping characteristic of rhythm was *too* effective for some groups such that performance was significantly reduced. From Figure 5.4, we can observe that the icons from *Spouse* and *Friend* (yellow and green) were sources of a great deal of within-group confusion. Note that the rhythms for these two groups consist of two and three quarter notes: the least number of notes for all rhythms. We believe that the fewer number of notes may reduce the user's ability to get a sense for the melody of the stimuli. There are also fewer possible melodic variations available due to the smaller number of notes, therefore the variations chosen may not be as distinctive.

In summary, designing rhythm-based melodic icons is an extremely difficult process. If the stimuli are too similar, then the group identification works extremely well, but the exact icon identification performance is reduced, often with a very high level of confusion. If the stimuli are too distinct, then the group is not identified correctly as often, but the exact identity of the icon is easier to recall. For these reasons, rhythm-based melodic design is extremely risky compared to purely rhythmic design with our current level of knowledge. Further research is required to examine how to maximize the perceptual distinctiveness of haptic stimuli through the exclusive variation of amplitude and frequency. Perhaps we can utilize melodic variation to reinforce and increase the perceptual difference between similar rhythms. What if designers were to allow users apply their melody of choice to rhythms that they are using? This might help users formulate mnemonic associations, while giving haptic icon designers a great deal of control over the primary differentiating feature (rhythm).

Learning process

Despite the perceptual issues stated above, we would like to examine the learning process in order to understand the differences between purely rhythmic icon learning process.

In Figure 5.5, we show the average number of sessions required to complete each batch for all 10 users. As one can see, this measure is quite flat from batch to batch. A notable data point is for batch 1, where there is a large standard error due to two participants who were stuck on that batch for a very long time. The truth of this is revealed when we plot the top 8 users in Figure 5.6, since the mean number of sessions to complete batch 1, as well as its standard error, are reduced. The values in this plot are extremely similar to those in Figure 3.26, with the exception of batch 2, which is reduced from 5 to approximately 3 sessions required to complete it. However, in batch 5 shows a contradictory trend. The three best users complete it, but require an increased and widely varying number of sessions. We can see in Figure 5.8 that users 10 and 8 experience a severe drop in performance in the later sessions of the experiment. This can explain the increase in the number of sessions required to complete batch 5. In addition, this phenomenon shows that melodic icons may not be as scalable as we previously thought possible. Even the most adept users experience confusion in later batches. There are a few potential explanations for this observation:

- The participants may be experiencing boredom or fatigue. Before experiencing the drop in performance, these participants had extremely high performance. It is possible that they found the task too easy and put less effort into maintaining a high level of performance.
- This phenomenon may reveal a critical point, in terms of the number of icons, when a user loses the ability to perform reliable discrimination of haptic stimuli. In order to overcome this critical point, users may need to further increase their sensitivity to discriminating haptic stimuli through continued training in order to develop the appropriate cognitive mechanisms (see Section 3.6.3).
- Due to a lack of elaboration and training, these icons may have failed to be encoded into long term memory. These users progressed extremely quickly through the batches of icons. There is a possibility that they may have progressed too quickly to develop the cognitive scaffolding and increases in sensitivity required for reliable icon identification amongst many confusing stimuli.

It is difficult to determine the exact cause of this drastic decrease in performance and no additional information was gained through interviews with participants. Although it would be extremely costly to perform, a user study with an even longer duration and larger number of rhythm-based melodic icons may reveal the source of this scalability issue.

If it is the case that the effect is caused by a lack of elaboration and the need for increased haptic sensitivity, then perhaps the performance may increase for these users if longer-term training is performed since this will provide these users with the elaboration needed for long-term encoding and an increase in cortical representation.

In Figure 5.8, we display the learning curves for each user throughout the duration of the experiment. A couple of participants, such as users 4 and 7 experience the characteristic 'dip' in the learning curve after passing the second batch. However, this pattern is not as prevalent as in the results from the purely rhythmic set (Figure 3.30. This is another way of viewing the flatter learning curve observed earlier.

A few users, such as user 7, experience a fairly consistent climb in performance in later sessions. This may be due to slow progression at the beginning of the experiment, possibly resulting in an improved formation of cognitive scaffolding. Unfortunately, we cannot predict if they will encounter the same scalability that was observed with other users without understanding the source of that problem.

In Figure 3.30, the vast majority of users experienced an upward climb in performance, while in Figure 5.8, we observe that some users either did not improve, or experienced decreased performance as the study progressed. This issue may be caused by an increase in perceptual confusion caused by additional stimuli that are extremely similar, or perhaps the lack of elaboration explained above.

As one can see, user 5 did not improve at all throughout the course of the experiment. This particular user complained that he/she could not feel the difference between the melodic stimuli. Rhythm dominated his/her perception entirely. The most challenged user in the experiment conducted in Chapter 3 at least began to improve after a great deal of training with the rhythmic icons. We are unsure how long it would take user 5 to experience the melodic variation, or if they would be able to perceive a different range of variation in frequency/amplitude.

In Figure 5.9, we show the number of icons identified correctly during the final quiz for all participants except users 1, 5 and 9 due to the fact that they did not progress enough to experience 24 icons. The mean number of correctly identified icons is 16.7, which is in the same range as the results for the purely rhythmic set, except their quiz only had 21 icons (Figure 3.25. Note that user 10 may have performed poorly since he/she was redoing batch 5 for 4 sessions. This user was our top performer during the course of the experiment and he/she may have performed better if more of the icons had been experienced more recently. It is also possible that this user did not successfully encode the identify of the icons into long term memory due to a lack of elaboration – he/she may have relied heavily on training to recall the identity of icons throughout the experiment through short-term encoding.

Although the performance during the quiz is roughly similar to the purely rhythmic study, three users instead of one were removed due to inadequate performance. In addition, the users who progressed the most were not the top performers in the quiz. This may give evidence to the theory that the users who experienced a steady climb in performance had adequate time to develop a solid cognitive scaffolding with which they could perform long-term encoding.

In order to encourage the development of cognitive structures that aid in long-term encoding in deployment, training should allow users more exposure to haptic icons and have higher requirements on progression.

We now wish to conclude this section and give our assessment regarding the learnability of rhythm-based melodic haptic icons versus purely rhythmic haptic icons. Although our results reveal that there may be some desirable qualities with respect to the potential learnability of rhythm-based melodic haptic icons, there are some issues that must be explored further and resolved before we can conclude that they are easier to learn than purely rhythmic icons.

The flatter learning curve observed for some users represents a very desirable quality. We wish for the learning process to be as efficient and easy as possible through haptic icon design that utilizes the strengths of our cognitive structures. This benefit is evident for some users who were able to perceive the differences between melodies sufficiently enough to progress through the experiment.

Unfortunately, there are some very significant problems that temper this result. First, there are some users who either struggle to, or simply cannot perceive the melodic variation in the stimuli. These users were unable to progress through the experiment and could not learn the haptic icons presented to them. Second, we observed a decrease in performance in later sessions for some proficient users. Until we can explore this phenomenon enough in order to understand the source of this scalability issue, it may be difficult to design rhythm-based melodic haptic icons that have desirable scalability properties. Unfortunately, we believe that this problem must be explored through longitudinal studies with a greater duration, or within deployment.

Given the current state of our understanding of these results, we would conclude that purely rhythmic haptic icons are more *consistently* learnable than rhythm-based melodic haptic icons.

5.5.2 Which Type of Icon is more Viable for Deployment?

In this section we will discuss the advantages and disadvantages of both rhythm-based melodic icons, as well as purely rhythmic icons in an effort to determine which type of icon is more viable for deployment.

Arguments for purely rhythmic haptic icons

In this section we will list and discuss the advantages of purely rhythmic haptic icons while referring to the analysis developed in Section 5.5.1, as well as some novel arguments that were not discussed in Chapter 3.

- *They result in a more consistent sensation under workload.* In Section 5.4.2, we reported a marginally significant difference between the mean accuracy during the quiz (80%) versus the mean accuracy during the game (71%). However, in Section 3.4.2, we reported no significant difference for purely rhythmic icons. These results may indicate that icons with a melodic component require more conscious attention to perceive effectively. Intuitively, this makes a great deal of sense. Purely rhythmic icons only deliver one type of sensation (in terms of frequency and amplitude) to the user. With this method, users do not have to attend to the exact nature of each note in the stimulus – they must only recognize the presence of the stimulus and get a sense of the rhythm being expressed. With melodic icons, the user must attend to each specific note in order to understand the sequence of the melody. In this case, they must attend to two separate dimensions of differentiation throughout time, as opposed to a single dimension on the purely rhythmic case. Purely rhythmic icons have much more promise in a deployment scenario since we wish to transmit messages without imposing much additional workload. Using a medium that requires very little conscious attention to the stimulus itself is preferrable.
- *The learnability is more predictable.* From our discussion in Section 5.5.1, we argued that, with our current knowledge about rhythm-based melodic haptic icons, their use in deployment is quite risky due to the fact that some users simply cannot perceive the variations in amplitude and frequency. In addition, we have not developed a sufficient explanation or understanding for the reduction in performance experienced by some very proficient users (Figure 5.8). On the other hand, in the learning curves for the purely rhythmic set (Figure 3.30), every user experienced a consistent improvement in performance throughout the course of the experiment, despite the differences in learning speed. For these reasons, we would argue that all users have a better chance of being able to learn purely rhythmic icons given a sufficient learning period.

- *The addition of modifiers is supported and effective.* Although each rhythm can only be assigned one meaning – as opposed to several with the addition of melody – there will be less confusion regarding the exact semantics of each icon. Modifiers can be applied to each rhythm in order to express various dimensions such as urgency or emotional valence. Our discussion in Section 3.5.4 reveals that performance does not suffer when modifiers are added to rhythms for the users who experienced them. We do not believe that this technique can be applied to melodic icons since a transposition may render some notes extremely difficult to perceive. However, further study is required regarding the addition of modifiers to melodic icons before a certain conclusion can be made.
- *The main grouping characteristic is well-understood and effective.* In Section 3.5.2, we showed that there was very little confusion between the two senders. In Section 5.5.1, there was a great deal of confusion between the senders, although the grouping characteristic of rhythm did have an observable effect. The even/uneven distinction is strong enough to avoid mostly all confusion. Despite our results in Chapter 4, utilizing rhythm as a grouping characteristic did not result in effective results, in this case.
- *They are better supported by current hardware.* Most current hardware in the marketplace is able to administer a series of timed on/off commands to an actuator that acts at a constant frequency – this is all that is required for purely rhythmic icons. In order to administer melodic stimuli, a designer must have control over the amplitude and frequency of each individual note. There are few devices that allow this in the market today.

Arguments for rhythm-based melodic haptic icons

Despite the advantages of purely rhythmic icons explained above, there are some advantages to rhythm-based melodic haptic icons that were discussed in the introduction to this chapter (page 105). We will briefly summarize those advantages here.

- *It is easier to formulate icon groups.* By using rhythm as a grouping characteristic, it is fairly easy to develop icons for within-group variation. One simply uses the base rhythm as a starting point and modifies the frequency and amplitude of its constituent notes – keeping in mind the heuristics from Section 4.7. However, it is advisable to perform an iterative perceptual design methodology such as MDS to refine the icon groups. This technique allows for the development of more icon groups than purely rhythmic icons. For instance, in a deployment scenario such as ours, it is useful to have a stimulus set that supports a large number of stimulus groupings (sender, in our case), rather than a limited number. In addition, by partitioning learning into more manageable, smaller families, the learning speed of the icon set may be increased. Instead of focusing on the differences between all of the stimuli, with perceptually segregated groups, users only need to learn the differences between the defining characteristics of each group, and then the differences between each icon within the context of each group. Unfortunately, our results do not indicate that rhythm is as strong of a grouping characteristic as predicted. Further study is required to develop a deep understanding of this grouping characteristic.

- *They are potentially more expressive.* Melodic icons may have the potential to evoke a range of responses that may be elicited through melodic music. For instance, a melody utilizing the major scale may elicit a happy emotion. This ability allows for a wide range of potential techniques for increasing learnability through the facilitation of mnemonic generation. Unfortunately, this is very difficult to accomplish within the constraints of our stimuli and hardware platform. Future work may examine how haptic melodies can be used to elicit specific emotions or responses, given fewer constraints on the design of the stimuli, as well as a hardware platform that permits complete control over the actuation.

- *The design space is much larger.* The melodic design space is exponentially larger than the rhythmic design space since every melody has a rhythm. With this expanded design space, we can create a much larger number of haptic icons – the number of potential icons is virtually limitless. Unfortunately, this large design space makes perceptual design extremely difficult to pursue. Care must be taken to avoid designing stimuli that do not follow the characteristics outlined in [31].

5.6 Contributions

This chapter had two primary research goals: to determine which type of icon supports increased leanability, between purely rhythmic icons and rhythm-grouped melodic icons; and to determine which type of icon is currently more viable for deployment.

We showed that, although rhythm-grouped melodic icons are more learnable for some users, generally, performance is unpredictable. Some users decreased in performance as the study progressed, and one in particular was not able to perceive melodic variation. Conversely, all users showed steady improvement in learning purely rhythmic icons, although the learning occurred at different speeds.

Due to the arguments outlined in Section 5.5.2, we believe that purely rhythmic haptic icons are *currently* more viable for deployment. They are much more predictable, better understood, easier to perceive, and better supported by currently existing hardware.

The design of perceptually distinct rhythm-based melodic icons needs to be better understood if we wish to utilize the desirable properties that we have listed above. In order to conclude this comparison with certainty, further work must be conducted in order to reconcile a couple of the phenomena that we observed. First, recall that some users simply cannot perceive the differences between melodic stimuli. Perhaps these users are not sensitive to the frequency band that we used and they would perform better if a different range were utilized. Future work could aim to understand the exact cause behind this phenomenon and design melodic icons accordingly, through calibration or other techniques. Second, we still need to determine what the optimal learning process is for rhythm-based melodic icons, as well as haptic icon in general. Should we have presented all of the icons that share the same rhythm in a single batch in order to increase our users' sensitivity to melody? Perhaps users need to first develop a sophistication for differentiating between melodies with the same rhythm before they can use the grouping characteristic effectively. If we had solved these open questions, perhaps our results and outcome would have been drastically different.

CHAPTER 6

CONCLUSIONS

Our work had two primary goals: to discover how effective haptic icons would be in a realistic, deployment-based scenario and to understand the learning trajectory that users would undergo when they are given a significant amount of time to learn. In other words, how *far* can we take haptic icons if people have a long time to learn them? Due to the richness of the data collected during our studies, we were able to develop a prescriptive set of guidelines for haptic icon design – encompassing which design parameters to use, haptic icon training and hardware design.

We developed this deep understanding of the potential of haptic icons through three separate, but progressive user studies – two of which were longitudinal and unprecedented in scale in the haptic icon community. In the following sections, we will summarize the primary contributions that our work provides to our field and provide directions for future work, given the findings of our research.

6.1 Summary of Contributions

To motivate this thesis, we advocated for a need to understand the effectiveness of haptic icons in a deployment scenario. Previously, studies assessing the design and learnability of haptic icons were restricted to short, one session user studies with small sets of icons. We aimed to take haptic icons from 'toy' research to practical application.

In order to fulfill this goal, we performed a first longitudinal study of haptic icons, designed to train and evaluate users on the largest stimulus set designed to date [43] with regular use over a period of one month. An in depth analysis of the results obtained from this user study resulted in many interesting contributions.

We observed a few properties of haptic icons that can make them more difficult to learn. First, simply having more icons increases the difficulty of learning. Naturally, as the set size increases, there are more potential sources of confusion. The main sources of confusion are caused by stimuli that only differ by note length or by timing – the number and density of notes dominates perception. Perceptual phenomena account for most of the difficulties in learning, however the meaning of an icon does have an effect in some isolated cases.

Conversely, we observed a few properties of haptic icons that make can them easier to learn. Icons with simple, recognizable and/or distinctive rhythms, within the context of the set, are much easier to learn. In addition, the ability to group icons by similarity based on a particular parameter makes icons easier to learn since learning is partitioned into smaller, more manageable 'chunks'.

Participants utilize many common techniques while learning haptic icons. Most prominently, participants focused on perceiving the differences between stimuli, while the semantics of the icons were easy to learn and considered simply a 'name' for the icon. Participants also used vocalization techniques to rehearse and remember icons. The rapidity of a rhythm was almost always associated with its urgency. Participants would impose a false sense of urgency on rapid rhythms in order to learn them more easily. Another common theme in learning arises in the haphazard mnemonics that users employ. They develop creative, one-off learning aids for many of the icons in the set, revealing that even with random meaning assignment, humans are very adept at developing useful mnemonics.

Later in our experiment, some users encountered icons that had the same base rhythm and semantics of previous icons, but had modified attributes relating priority to the amplitude of the stimulus and a repeated message to the frequency of the stimulus. Our results indicated that this the technique for adding attribute modifications to icons is extremely effective since users were rarely confused by these modifications.

Our analysis of the effectiveness of perceptual MDS as a design tool revealed that MDS is often very conservative in its estimates of confusion, predicting confusion when we did not observe any in our study. There were no instances where we observed confusion and MDS did not predict any. This result leads us to conclude that MDS is adequate as a perceptual design tool, however it might reduce the resulting icon set unnecessarily. In addition, a tool that focuses on the *differences* rather than the similarity between haptic stimuli is desired.

After twelve, twenty minute sessions over the period of one month and an interval of one week with no exposure, participants were able to correctly identify 16 out of 21 icons, on average. Although the learning speeds and performance between participants varied widely, identification performance improved consistently after some initial hurdles involving the development of cognitive scaffolding and an increase in haptic or rhythmic sensitivity. We believe that there are no limits to the number of haptic icons that a person can learn, as long as the stimuli are sufficiently perceptually distinctive given a user's current level of sensitivity, as well as the amount of expressiveness capable by the hardware platform.

Proficient haptic icon learners are separated from average and below average learners due to their increased sensitivity to the perceptual differences of rhythmic tactile stimuli. Indeed, most of the mistakes made by users were caused by perceptual confusion, thus the proficient users are defined by their robustness to perceptual similarities.

6.1.1 List of Heuristics, Guidelines and Advice for Rhythmic Icons

Based on a synthesis of the results obtained in Chapter 3 and summarized above, we developed a series of heuristics for haptic icon design, guidelines for haptic icon training, and advice for hardware designers.

We summarize them in the following lists:

Heuristics for perceptual distinctiveness:

- Utilize MDS analysis to give accurate predictions of perceptual confusion.
- Avoid the temptation to perform metaphorical design.
- Modify the number and density of notes for perceptual distinctiveness.

Heuristics for associability:

- Similar stimuli should have similar meanings.
- Utilize the rapidity of a note to portray urgency.
- Take advantage of vocalization when assigning meanings to haptic icons whenever possible.

Heuristics specific to rhythmic icons:

- Utilize extremely recognizable and emotion-evoking rhythms.
- Design stimuli with simple rhythms before adding complexity through variations in note timing and types.
- Utilize display parameters such as frequency and amplitude to represent modifications on already learned haptic icons.

Guidelines for Haptic Icon Training:

- Begin training by focusing on increasing haptic sensitivity.
- Encourage helpful techniques such as rehearsal through vocalization and stimulus partitioning.
- Have users perform realistic practice.
- Expose your design parameters to users.
- Provide incentives or sources of enjoyment in order to keep motivation high.

Advice for Hardware Designers:

- Focus on the intensity of the attack of a vibration.
- Aim for a quick recovery.
- Support variations in frequency/amplitude.

The results in Chapter 3 shared a common theme that we consider vital to haptic icon learnability: the perceptual distinctiveness of an stimuli is paramount. We believe that the majority of haptic icon design should be devoted to the design of perceptually desirable and distinctive stimulus sets, while the assignment of meaning should be performed sensibly, after the stimuli have been designed. There is a natural inclination to believe that the design of metaphorical icons would be best for learnability, however, we have no evidence that the meaning of an icon has much of an effect, and metaphorical design suffers from poor repeatability, is limited by the creativity of designers and cannot scale as well as perceptual, abstract design.

6.1.2 The Addition of Melodic Variation

Despite the promising results found in Chapter 3, we found that the addition of melodic variation through note-by-note variation in frequency and amplitude had the potential to provide even greater benefits since the design space is exponentially larger, allowing designers to draw from an extremely large pool of stimuli, and allows for increased expressiveness due to the similarity of melodic haptic icons to melodic music.

In order to design a set of melodic haptic stimuli, a design process similar to Enriquez and MacLean and [32] Ternes and MacLean [43]. We performed a set of iterative cluster sorting studies and plotted the perceptual space of the stimuli using perceptual MDS.

Through an extensive analysis of the results obtained from this iterative design process, we discovered that rhythmic differences between melodies dominate other distinctions. One can make use of this contribution by designing stimulus groups around a base rhythm, and then expressing within-group variation through note-by-note variation in frequency and amplitude.

In addition to this primary finding, we observed some other patterns relating to the use of melody in tactile rhythms: the perceived quantity of notes is a major grouping factor; groups of rapid eighth notes are perceptually salient in rhythms; and ‘abrupt’ melodies are perceptually segregated from ‘rolling’ melodies, even if they are devised from the same base rhythm.

In order to understand whether these rhythm-grouped melodic haptic stimuli are as effective as purely rhythmic icons, we performed another longitudinal study that is almost identical to the study described in Chapter 3, but we used the stimulus set devised in Chapter 4.

Our analysis of this second longitudinal study of haptic icon learnability focused around the differences between rhythm-grouped melodic icons and purely rhythmic icons, in an effort to understand which type of icon is more learnable and which type of icon is currently more viable for deployment. Our results showed that purely rhythmic icons are more predictable in terms of their learnability, despite the fact that some users showed promise with the rhythm-grouped melodic icons. Due to the predictability of purely rhythmic icons, as well as the fact that they are better understood, easier to perceive and better supported by current hardware, we concluded that rhythmic icons are currently more viable for deployment.

6.2 Directions for Future Work

Despite our conclusion that purely rhythmic haptic icons are ready for deployment, there is much work left to be conducted in order to further improve the suitability of haptic icons in practical applications.

At this point, it is obvious that haptic icons should be tested in an *actual* deployment setting. Haptic feedback is becoming much more common in mobile devices, and researchers should take it upon themselves to test the efficacy of haptic icons on these devices in the real world, over long periods of time. If researchers do not take action, then naive developers might begin creating haptic icons, increasing the risk that lower quality haptic icons that do not have desirable perceptual or associability characteristics will be made. An initial negative outcome of a technology drastically decreases the likelihood of acceptance, even if the technology is improved significantly. For instance, 'Clippy', or the Microsoft Office Assistant, had a dismal reception by the public due to the overzealous implementation of its adaptive help behavior (the initial prototype is published in [24]). To this day, over 10 years later, users are still extremely resistant to accepting adaptive interfaces because they fear a resurrection of 'Clippy'. The first haptic icons to reach the mobile community should be backed by good science in order to increase the likelihood of their acceptance.

Although we advocate heavily that designing a good stimulus set is the most important task for haptic icon designers, and that metaphorical design should be avoided due to its problems with scalability and repeatability, we did not conduct a comparative study. Future work should examine the performance differences between abstract, perceptual design and metaphorical design in an unbiased comparison that examines expressiveness, learnability, repeatability and scalability to larger sets.

Both of the studies conducted in Chapters 3 and 5 showed that some people were better than others at perceiving the differences between the haptic stimuli. We are uncertain whether this difference is due to a difference in overall haptic sensitivity, or in the sensitivity to the particular physical parameters that we utilized. Would the results have been different if we had used frequencies of 150Hz and 250Hz instead of 200Hz and 300Hz? We find these results surprising since our chosen range was not chosen with good reason due to the fact that it is consistent previous observations of human sensitivity and was confirmed through pilot testing [43]. Future work could examine the range of individual sensitivities for the various physical parameters of tactile, vibratory feedback. This work could identify the existence (or lack) of a heterogeneous sensitivity ranges between users. Designers could then improve the overall perception of haptic icons by calibrating the stimuli to take advantage of a user's optimal sensitivity range. One could even evaluate whether users are able to self-calibrate, allowing hardware designers to generate a single haptic icon set that is calibrated by the user upon deployment.

As we alluded to in Section 3.5.4, future work should examine the limitations of the use of hardware parameters such as frequency and amplitude to add modifiers to monotone icons. Would we have the same effect if the stimulus attributed is not as compatible to the semantic attribute? Is the phenomenon robust enough to maintain its efficacy even if the relationships were drastically incongruous? The exact parameters of this design heuristic should be studied further in order to increase our understanding of the link between the haptic modality and memory.

We utilized melody as a design parameter in a particular way, where our melodies were all periodic, with four repeated 500 millisecond components in 4/4 time to create a 2 second stimulus. This repetitive structure tends to highlight the rhythmic aspect of the stimulus, yet might not be the optimal way to express a melody, since they are often more unique and may require a longer duration to convey. How might melody be used as a design parameter if we remove these constraints? Are there additional grouping characteristics that we can exploit, such as tonal scale or tempo? Recall that in Section 5.5.2, we observed that rhythm-grouped melodic icons require more attention since their differentiation occurs over two dimensions. Perhaps if melody were a unique, redundant differentiator, or alternatively a highly common one with only a small number of values detectable from the first few moments – for instance, ascending vs. descending vs random tonality. – then their perception may require less cognitive load. Much of the melodic design space remains unexplored and we believe that it is still a rich vehicle through which to further understand haptic communication.

In the same vein as the previous paragraph, an interesting path of future work would be to examine how haptic stimuli can elicit specific emotions or responses. For instance, can we create a vibratory stimulus that makes someone cringe, or can make someone feel happy? Auditory music is well understood in its ability to affect the mood of a listener [41]. How can we effect emotion through the haptic modality, and as an extension, how can we take advantage of these findings for haptic communication?

In closing, we hope that this thesis has improved your understanding of the usefulness of the haptic modality for the communication of information, while simultaneously raising thought-provoking and inspirational questions.

BIBLIOGRAPHY

[1] Alan Baddeley. *The Essential of Human Memory*. Psychology Press, 1999.

[2] Meera M. Blattner, Denise A. Sumikawa, and Robert M. Greenberg. Earcons and icons: Their structure and common design principles. *Human-Computer Interaction*, 4(1):11–44, 1989.

[3] Simon J. Boniface and Ulf Ziemann, editors. *Plasticity in the Human Nervous System: Investigations with Transcranial Magnetic Stimulation*. Cambridge University Press, 2003.

[4] Stehen A. Brewster and Lorna Brown. Non-visual information display using tactons. In *Extended Abstracts of the 2004 Conference on Human Factors and Computing Systems*, 2004.

[5] Stehen A. Brewster, Peter C. Wright, and Alistair D.N. Edwards. Experimentally derived guidelines for the creation of earcons. In *HCI'95*, Huddersfield, UK, 1995.

[6] Lorna M. Brown. *Tactons: Structured Vibrotactile Messages for Non-Visual Information Display*. PhD thesis, University of Glasgow, April 2007.

[7] Lorna M. Brown, Stephen A. Brewster, and Helen C. Purchase. A first investigation into the effectiveness of tactons. In *1st Worldhaptics Conference (WHC 2005)*, pages 167–176, Pisa, Italy, 2005.

[8] Lorna M. Brown, Stephen A. Brewster, and Helen C. Purchase. Multidimensional tactons for non-visual information presentation in mobile devices. In *8th Conference on Human-Computer Interaction with Mobile Devices and Services*, pages 231–238, 2006.

[9] Antony Bruno and Chris M. Walsh. *Text messaging catches on in music business*. `http://www.reuters.com/article/industryNews/idUSN3030999220061101`.

[10] Andrew Chan, Karon E. MacLean, and Joanna McGrenere. Learning and identifying haptic icons under workload. In *WHC*, pages 432–439, 2005.

[11] Andrew Chan, Karon E. MacLean, and Joanna McGrenere. Designing haptic icons to support collaborative turn taking. *Int'l J Human Computer Studies*, 66:432–439, 2008.

[12] Angela Chang and Conor O'Sullivan. An audio-haptic aesthetic framework influenced by visual theory. In *HAID*, pages 70–80, 2008.

[13] Noam Chomsky. *Aspects of the Theory of Syntax*. MIT Press, 1965.

[14] Terrence W. Deacon. *The Symbolic Species*. W.W. Norton & Company Ltd., 1997.

[15] Marcelo Eduardo. *Maemoblocks*. `http://maemo.org/community/oldwiki/applicationcatalog/`.

[16] Mario Enriquez and Karon E. MacLean. The role of choice in longitudinal recall of meaningful tactile signals. In *in Proceedings of the IEEE Symposium on Haptic Interfaces for Virtual Environments and Teleoperator Systems*, pages 49–56, March 2008.

[17] Mario Enriquez, Karon E. MacLean, and Christian Chita. Haptic phonemes: Basic building blocks of haptic communication. In *8th Int'l Conference on Multimodal Interfaces*, pages 302–309, 2006.

[18] Fretjam. *Guitar Strumming Technique*. `http://www.fretjam.com/guitar-strumming-technique.html`.

[19] William W. Gaver. Auditory icons: Using sound in computer interfaces. *Human-Computer Interaction*, 2:167–177, 1986.

[20] Bruce E. Goldstein. *Sensation and Perception*, chapter The Cutaneous Senses. Wadsworth, sixth edition, 2002.

[21] Kate Hevner. The affective character of major and minor modes in music. *American Journal of Psychology*, 47:103–118, 1935.

[22] Eve E. Hoggan and Stephen A. Brewster. Crossmodal icons for information display. In *CHI '06: CHI '06 extended abstracts on Human factors in computing systems*, pages 857–862, 2006.

[23] Mark Hollins, Richard Faldowski, Suman Rao, and Forrest Young. Perceptual dimensions of tactile surface textures: A multidimensional scaling analysis. *Perception and Psychophysics*, 54:697–705, 1993.

[24] Eric Horvitz, Jack Breese, David Heckerman, David Hovel, and Koos Rommelse. The lumiere project: Bayesian user modeling for inferring the goals and needs of software users. In *Conference on Uncertainty in Artificial Intelligence*, 1998.

[25] Edwin L. Hutchins, James D. Hollan, and Donald A. Norman. Direct manipulation interfaces. *Human-Computer Interaction*, 1(4):311–338, February 1985.

[26] Topi Kaaresoja, Lorna M. Brown, and Jukka Linjama. Snap-Crackle-Pop: Tactile Feedback for Mobile Touch Screens. *Proceedings of Eurohaptics 2006*, pages 565–566, 2006.

[27] Pauli Laitinen and Jani Maenpaa. Enabling mobile haptic design: Piezoelectric actuator technology properties in hand held devices. In *HAVE2006 - IEEE Workshop on Haptic Audio Visual Environments and their Applications*, Ottawa, Canada, 2006.

[28] Susan J. Lederman. Auditory texture perception. *Perception*, 8(1):93–103, 1979.

[29] Joel R. Levin. Elaboration-based learning strategies: Powerful theory = powerful application. *Contemporary Educational Psychology*, 13(3):191 – 205, 1988.

[30] Joseph Luk, Jerome Pasquero, Shannon Little, Karon E. MacLean, Vincent Levesque, and Vincent Hayward. A role for haptics in mobile interaction: Initial design using a handheld tactile display prototype. In *ACM Conference on Human Factors in Computing Systems*, Montreal, Canada, 2006.

[31] Karon E. MacLean. Foundations of transparency in tactile information design. *IEEE Transactions on Haptics*, page 12, 2008.

[32] Karon E. MacLean and Mario Enriquez. Perceptual design of haptic icons. In *Eurohaptics*, Dublin, UK, July 2003.

[33] J.C. Makous, R.M. Friedman, and C.J. Vierck. A critical band filter in touch. *Neuroscience*, 15:2808–2818, 1995.

[34] George Miller. The magical number seven, plus or minus two: some limits on our capability for processing information. *Psychological Review*, 63:81–97, 1956.

[35] Slater E. Newman, Anthony D. Hall, Darryl J. Foster, and Vasudha Gupta. Learning as a function of haptic discriminability among items. *The American Journal of Psychology*, 97(3):359–372, 1984.

[36] Takuya Nojima and Kohei Funabiki. Cockpit display using tactile sensation. In *Eurohaptics*, pages 501–502, 2005.

[37] Alexey Pajitnov. Tetris. `http://www.tetris.com/`.

[38] Jerome Pasquero, Joseph Luk, Shannon Little, and Karon E. MacLean. Perceptual analysis of haptic icons: an investigation into the validity of cluster sorted mds. In *Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems (IEEE-VR)*, 2006.

[39] Isabelle Peretz, Lise Gagnon, and Bernard Bouchard. Music and emotion: perceptual determinants, immediacy and isolation after brain damage. *Cognition*, 68:111–141, 1998.

[40] Peter L. Pirolli and John R. Anderson. The role of practice in fact retrieval. *Journal of experimental psychology. Learning, memory, and cognition*, 11(1):136–153, 1985.

[41] M.G. Rigg. The mood effects of music: A comparison of data from four investigators. *Journal of Psychology*, 58:427–438, 1964.

[42] Anthony Tang, Peter McLachlan, Karen Lowe, Chalapati Rao Saka, and Karon MacLean. Perceiving ordinal data haptically under workload. In *ICMI*, 2005.

[43] David Ternes. Building large sets of haptic icons: Rhythm as a design parameter, and between-subjects mds for evaluation. Master's thesis, University of British Columbia, 2007.

[44] David Ternes and Karon E. MacLean. Designing large sets of haptic icons with rhythm. In *Eurohaptics*, pages 199–208, 2008.

[45] Ville Tikka and Pauli Laitinen. *Haptic and Audio Interaction Design*, volume Volume 4129/2006 of *Lecture Notes in Computer Science*, chapter Designing Haptic Feedback for Touch Display: Experimental Study of Perceived Intensity and Integration of Haptic Audio, pages 36–44. Springer Berlin/Heidelberg, 2006.

[46] Jan B.F. van Erp and Michiel M.A. Spape. Distilling the underlying dimensions of tactile melodies. In *Eurohaptics*, pages 111–120, 2003.

[47] Mark Weiser and John Brown. Designing calm technology. *Powergrid Journal*, 1.01:94–110, 1996.

[48] Wikipedia. *Evelyn Glennie*. `http://en.wikipedia.org/wiki/Evelyn_Glennie`.

[49] Wikipedia. *Zipf's law*. `http://en.wikipedia.org/wiki/Zipf's_law`.

APPENDIX A

CODE LIST FOR LONGITUDINAL RHYTHM STUDY

```
!SCH - "!Spouse!: Call me at home"
!BCO - "!Boss!: Call my office"
!BCC - "!Boss!: Call my cell"
!STY - "!Spouse!: I'm thinking about you"
!SMY - "!Spouse!: Got your message: Sounds good"
!SMN - "!Spouse!: Got your message: Can't do it"
!SBD - "!Spouse!: Buy dinner on the way home"

!SLB - "!Spouse!: I'm late: I will be there"
!SPK - "!Spouse!: Pick up the kids from school"
!BCE - "!Boss!: Check your email"
!BMY - "!Boss!: Got your message: Sounds good"
!BCL - "!Boss!: I'm in a meeting: Call me later"
!BMN - "!Boss!: Got your message: Can't do it"
!SCW - "!Spouse!: Call me at work"

!BBR - "!Boss!: Bring me that report"
!SMT - "!Spouse!: Movie tonight?"
!SCC - "!Spouse!: Call my cell"
!BCF - "!Boss!: Come to my office"
!BSR - "!Boss!: Don't forget your status report"
!BMC - "!Boss!: The meeting is cancelled"
!BCH - "!Boss!: Sorry: I can't help"

2SCH - "(2)!Spouse!: Call me at home"
2BCO - "(2)!Boss!: Call my office"
2BCC - "(2)!Boss!: Call my cell"
2STY - "(2)!Spouse!: I'm thinking about you"
2SMY - "(2)!Spouse!: Got your message: Sounds good"
2SMN - "(2)!Spouse!: Got your message: Can't do it"
2SBD - "(2)!Spouse!: Buy dinner on the way home"

2SLB - "(2)!Spouse!: I'm late: I will be there"
2SPK - "(2)!Spouse!: Pick up the kids from school"
2BCE - "(2)!Boss!: Check your email"
2BMY - "(2)!Boss!: Got your message: Sounds good"
2BCL - "(2)!Boss!: I'm in a meeting: Call me later"
2BMN - "(2)!Boss!: Got your message: Can't do it"
2SCW - "(2)!Spouse!: Call me at work"

2BBR - "(2)!Boss!: Bring me that report"
2SMT - "(2)!Spouse!: Movie tonight?"
2SCC - "(2)!Spouse!: Call my cell"
2BCF - "(2)!Boss!: Come to my office"
2BSR - "(2)!Boss!: Don't forget your status report"
2BMC - "(2)!Boss!: The meeting is cancelled"
2BCH - "(2)!Boss!: Sorry: I can't help"

_SCH - "Spouse: Call me at home"
_BCO - "Boss: Call my office"
_BCC - "Boss: Call my cell"
_STY - "Spouse: I'm thinking about you"
_SMY - "Spouse: Got your message: Sounds good"
_SMN - "Spouse: Got your message: Can't do it"
_SBD - "Spouse: Buy dinner on the way home"

_SLB - "Spouse: I'm late: I will be there"
_SPK - "Spouse: Pick up the kids from school"
_BCE - "Boss: Check your email"
_BMY - "Boss: Got your message: Sounds good"
_BCL - "Boss: I'm in a meeting: Call me later"
_BMN - "Boss: Got your message: Can't do it"
_SCW - "Spouse: Call me at work"

_BBR - "Boss: Bring me that report"
_SMT - "Spouse: Movie tonight?"
_SCC - "Spouse: Call my cell"
_BCF - "Boss: Come to my office"
_BSR - "Boss: Don't forget your status report"
_BMC - "Boss: The meeting is cancelled"
_BCH - "Boss: Sorry: I can't help"
```

APPENDIX B

EXPERIMENT MANUAL FOR LONGITUDINAL RHYTHM STUDY

Haptic Icon Study: Participant Procedures

Bradley A. Swerdfeger

October 6, 2008

Contents

1 Pausing your session

During any of the screens, you can press the 'Escape' button (Figure 1) to pause your session. The experiment will add the duration of the pause to your session. *Please only use this in the event of an interruption.*

Figure 1: The Escape key.

You cannot pause during the following times:

- During the login screen.
- While pop-up boxes are displayed.
- During the questionnaire.
- During a haptic icon response in the Tetris game.
- Pausing will not increase the length of time you have to respond to a haptic icon during the Tetris game.

2 Login

At the login screen (Figure 2), click on the text-entry box with your stylus and then enter your username with the on-screen keyboard. Click the login screen.

If your login was correct, you will get a prompt to put on your headphones. **Please put on your headphones if you have not already.**

If your login was incorrect, you will be prompted to enter your login name again. Logins are not case-sensitive.

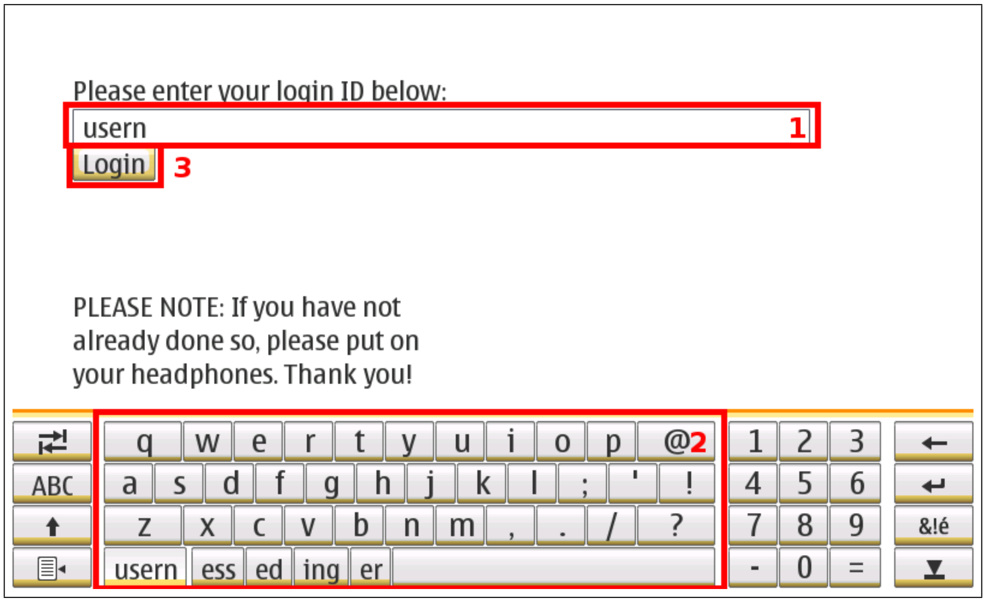

Figure 2: The Login screen.

3 Training

Each session begins with a training and quiz portion, with the remaining time being spent playing Tetris.

3.1 Training Introduction Screen

After closing the prompt to put your headphones on, you will be introducted to your current batch. Press the 'Proceed to Training Button' to begin training (Figure 3).

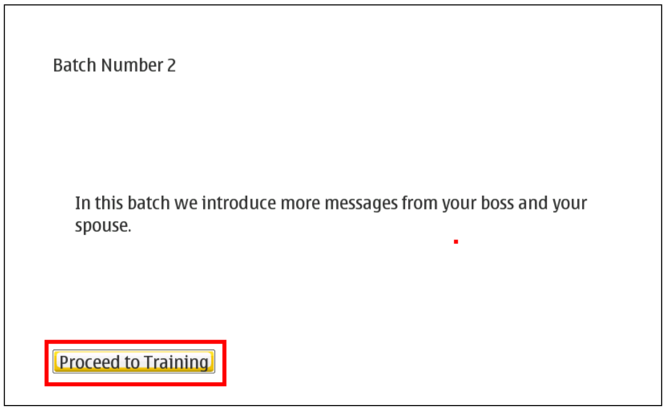

Figure 3: The Batch Introduction screen.

Batches are groups of 7 new haptic icons, including up to 14 icons from previous batches. Once the experiment calculates that you have learned a batch adequately during the Tetris game, you will move on to a new batch with its own set of older icons.

3.2 Training Screen

Figure 4 shows the training screen. A haptic icon is a meaningful tactile signal. In this case, there are messages from your boss and your spouse that correspond to a set of vibrations. The meaning of a haptic icon is on the left, and you can feel the icon by pressing the button on the right. For instance, a repetition of four quarter notes might correspond to the message '!Spouse!: Call me at home'. You can feel each of the icons as many times as you like by pressing on the numbered button.

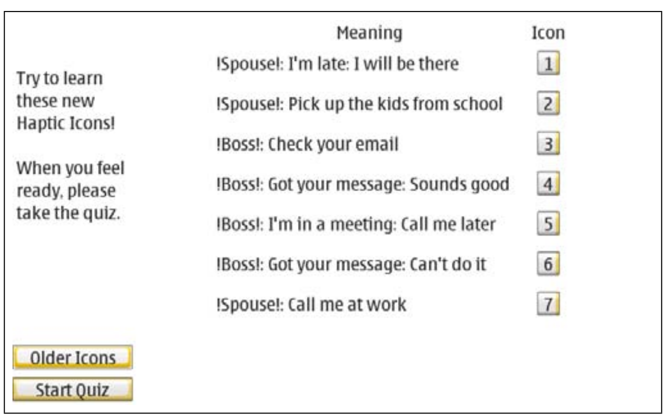

Figure 4: The Training screen.

There are three types of haptic icons that may be presented during the course of the experiment:

High Priority These have the sender's name surrounded by exclamation marks (ie. "!Spouse!: Call me"). These are high amplitude rhythms.

Low Priority These do not have exclamation marks (ie. "Spouse: Call me"). These are low amplitude rhythms.

Second Messages These have a '(2)' before the message (ie. "(2) !Spouse!: Call me"). These are high frequency rhythms.

When you feel ready, please press the 'Start Quiz' button to begin the quiz. If you need to feel the icons from previous batches, please press the 'Older Icons' button. You can also filter by 'Spouse' or 'Boss' by using the drop-down menu on any of the training screens and pressing 'Filter'. This is useful if you wish to compare similar icons from the same sender.

3.3 Older Icons

To access the older icons screen (Figure 5), you can press the 'Older Icons' button in the Training Screen (Figure 4). This will take you to a screen that contains up to 14 icons from previous batches that you may encounter during the quiz and the game. You can feel these as many times as you would like.

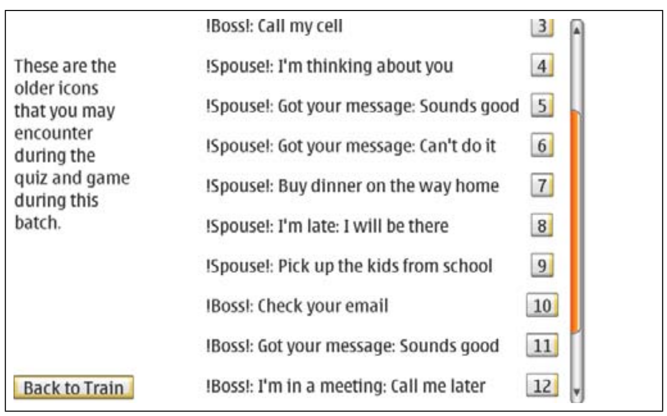

Figure 5: The Older Icons screen.

Please Note: There is a 20 minute refresh timer when moving from screen-to-screen. This is due to the fact that quick changing can un-sync the haptics.

When you are ready, please press the 'Back to Train' button.

4 Quiz

Before going on to the game, you must score 85% or better on the quiz. This tests your knowledge of the new icons so that the experiment knows you are ready for the game.

To start the quiz, press the 'Start Quiz' button from the training screen (Figure 4).

The quiz screen (Figure 6) will have a large button with a letter on it. When you click this button, the target haptic icon will play. You can feel this button as many times as you would like. Once you feel like you know which icon it is, select the proper meaning from the drop-down list, then press the 'Submit' button. A prompt will pop up, telling you whether or not you got the correct answer. If you got it wrong, the prompt will display the correct meaning for that icon. Once you close the prompt, you will be taken to the next question.

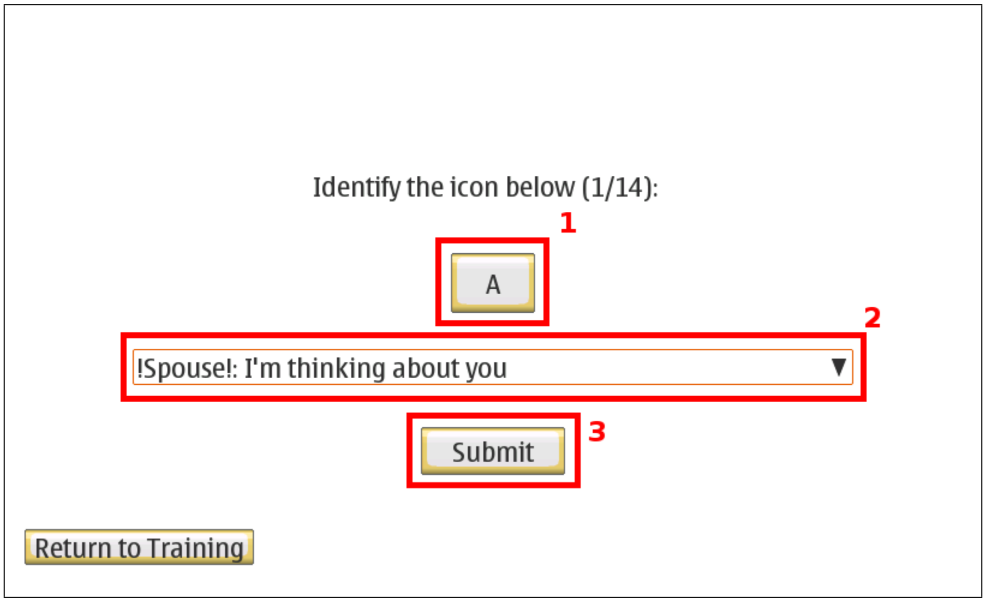

Figure 6: The Quiz screen.

At any point, you can press the 'Return to Train' button to feel all of the icons again.

Once you are done the quiz, a prompt will pop up telling you if you got a high enough score. If your score was not high enough, you will be returned to the training screen after closing the prompt. If your score was high enough, you will start the game after closing the prompt.

5 Game

Once you have completed the quiz (Figure 6) and scored 85% or better, the Tetris game will start.

At this point you should flip the device so that the long end is vertical.

5.1 Tetris

Our implementation of Tetris (Figure 7) uses only the stylus. You can drag the blocks using the stylus, and flip the blocks by pressing the 'Flip' button on the bottom right of the screen. The arrow buttons are disabled.

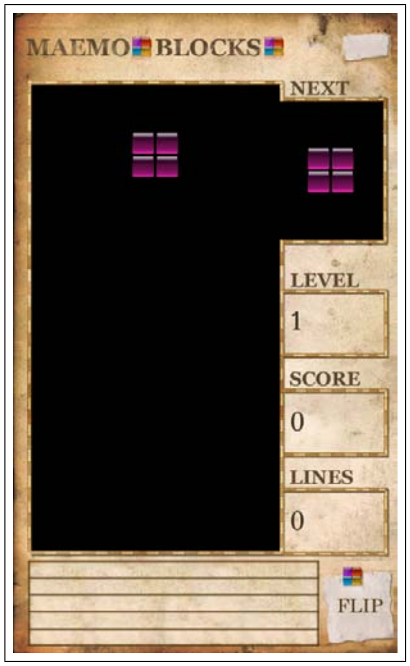

Figure 7: Tetris.

The object of the game is to get the highest score possible. You gain points by completely filling in a line with blocks. After 10 lines, the level will increase, which results in increased speed and points per line.

The game uses Game Boy scoring, where getting more lines at once gives you a higher multiplier on your score. They are as follows:

- 1 lines $= 10 * level$
- 2 lines $= 20 * level$
- 3 lines $= 40 * level$
- 4 lines $= 80 * level$

If the blocks reach the top of the screen, the level will return to 1 and the screen will be cleared. You cannot ever have a game over, but the penalty is a lower score multiplier.

5.2 Haptic Messages

At random intervals a haptic icon will play (Figure 8). At this point, a 'Respond' button will cover the bottom of the screen and you have 7 seconds to click on the button. If you do not hit the 'Respond' button, it will disappear, and the icon instance will be marked as incorrect.

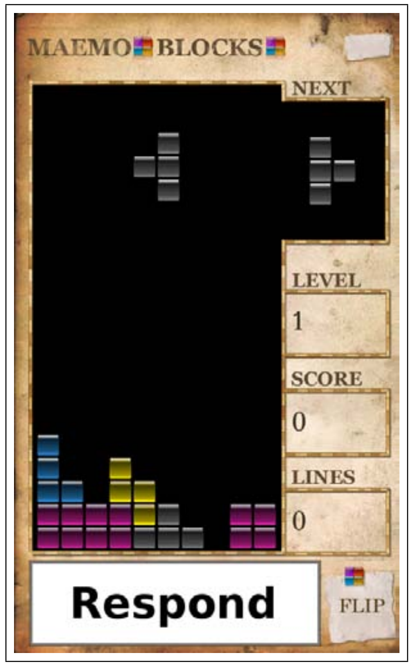

Figure 8: The Response prompt.

Once you press the 'Respond' button, you will be able to identify the haptic icon. A keyboard will be displayed and you must enter the 4-character code and press 'Answer' to identify the haptic icon. Please see the attached code sheet for the various codes (this means that you don't have to memorize them!)

If you answer the message correctly, the answer area will be covered in green. If you answer the message incorrectly, the response area will be covered in pink and you will be able to observe the correct answer 9. You will have 2 seconds to observe the correct answer.

5.3 Moving on to a New Batch

To move on to a new batch, the following two conditions must be met during the Tetris game:

1. All 7 new icons must have been answered correctly. This persists over multiple sessions.
2. You must have achieved 85% accuracy or better in the last 14 haptic messages.

Once these criteria have been met, a questionnaire will pop up. Fill out the questionnaire, submit it, and you will begin training for the next batch.

6 Questionnaire

Even if you do not complete a batch, at the end of your session, a questionnaire will pop up (Figure 10). After submitting the questionnaire, your session ends and the login screen is

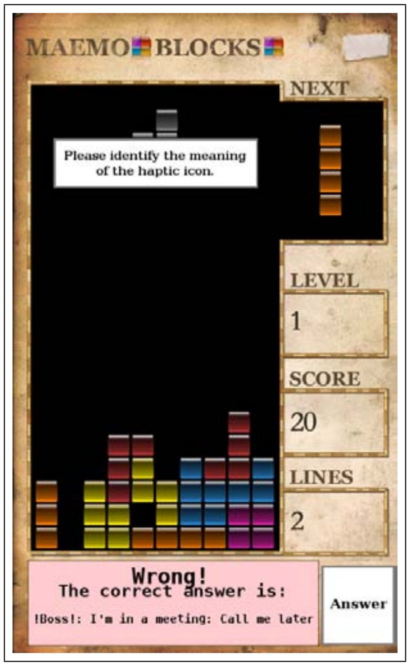

Figure 9: The answer observation time after submitting a response.

shown again.

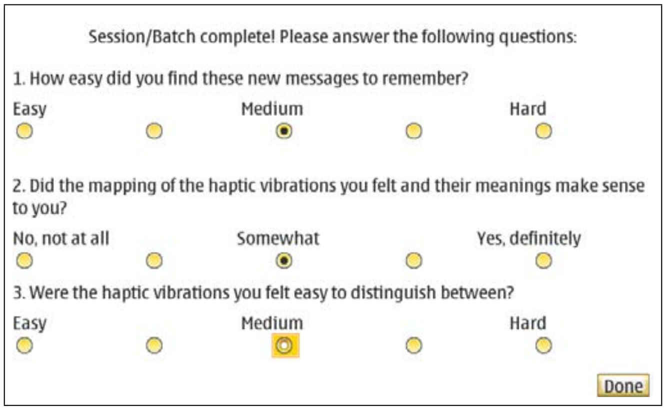

Figure 10: The Questionnaire.

7 Ending your session

19 minutes after you have logged in, your session will end (if you did not pause the session). At this point, a prompt will pop up telling you that your session has ended. Once you close the prompt, a questionnaire will pop-up. Please submit the questionnaire and your session will end.

The session can end during the following times:

- On loading the training screen.
- On finishing the quiz.
- At any time during the Tetris game except during a response.

Please make sure you do three sessions per week. Do not do more than one session per day.

APPENDIX C

VISUALIZATIONS OF PARTICIPANT SESSIONS FOR LONGITUDINAL RHYTHM STUDY

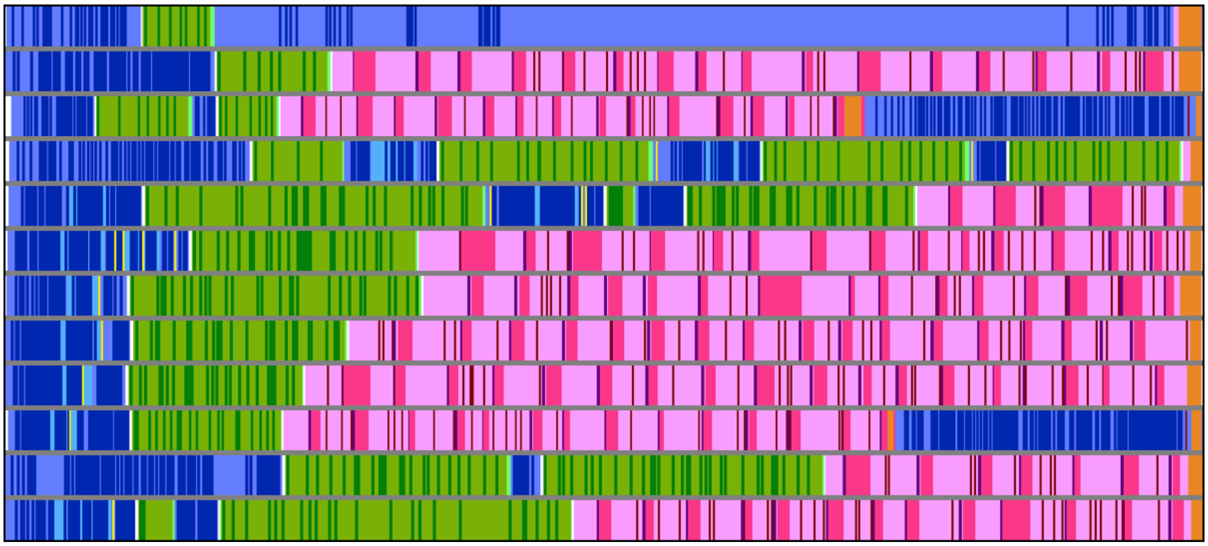

Figure C.1: Session activity visualization for User 1.

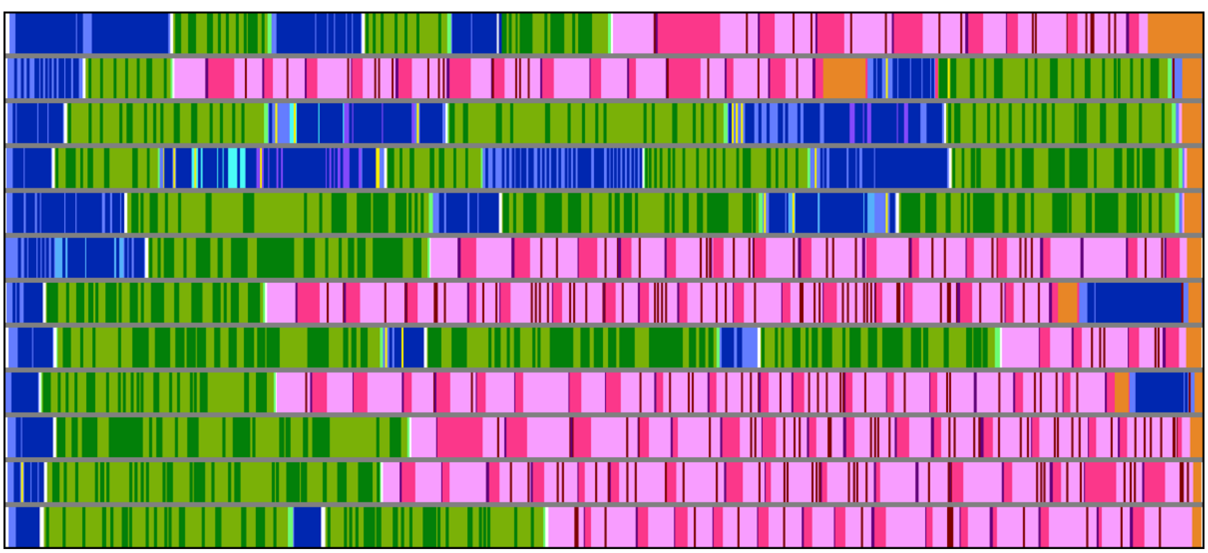

Figure C.2: Session activity visualization for User 2.

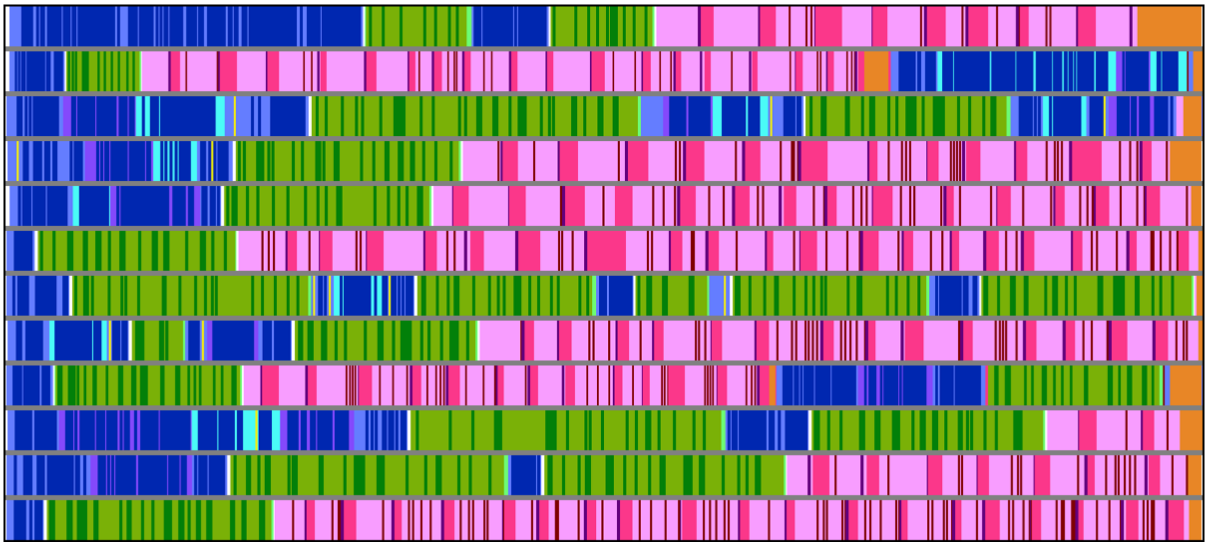

Figure C.3: Session activity visualization for User 3.

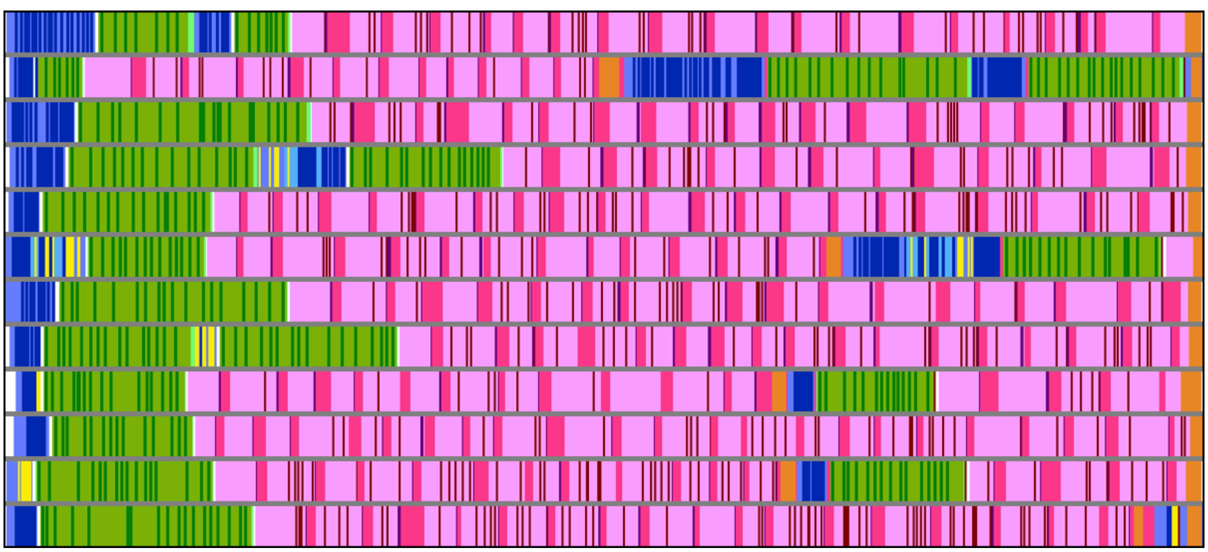

Figure C.4: Session activity visualization for User 4.

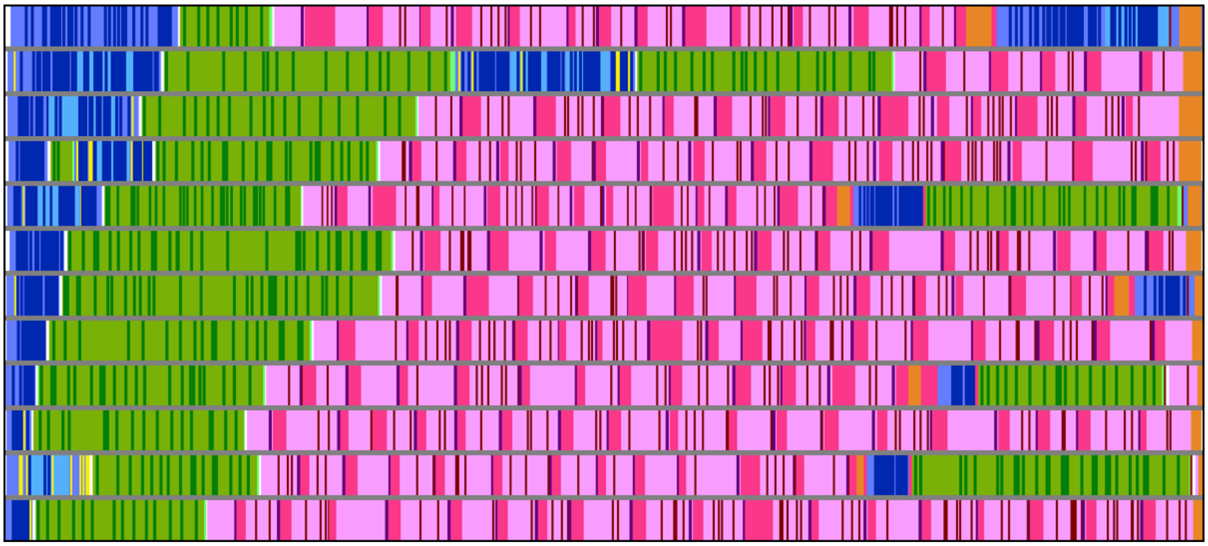

Figure C.5: Session activity visualization for User 5.

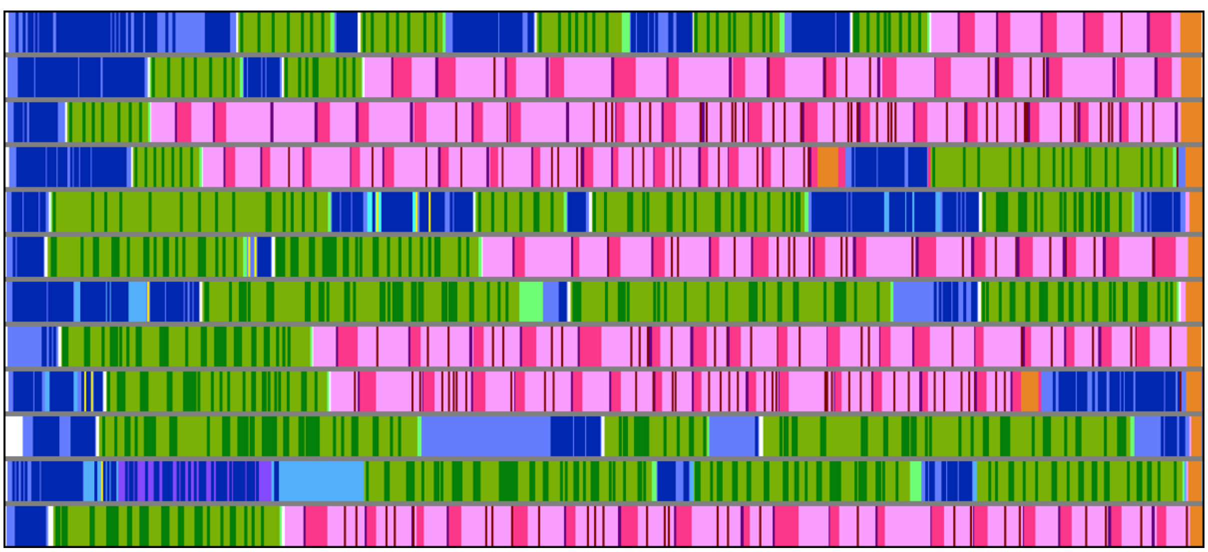

Figure C.6: Session activity visualization for User 6.

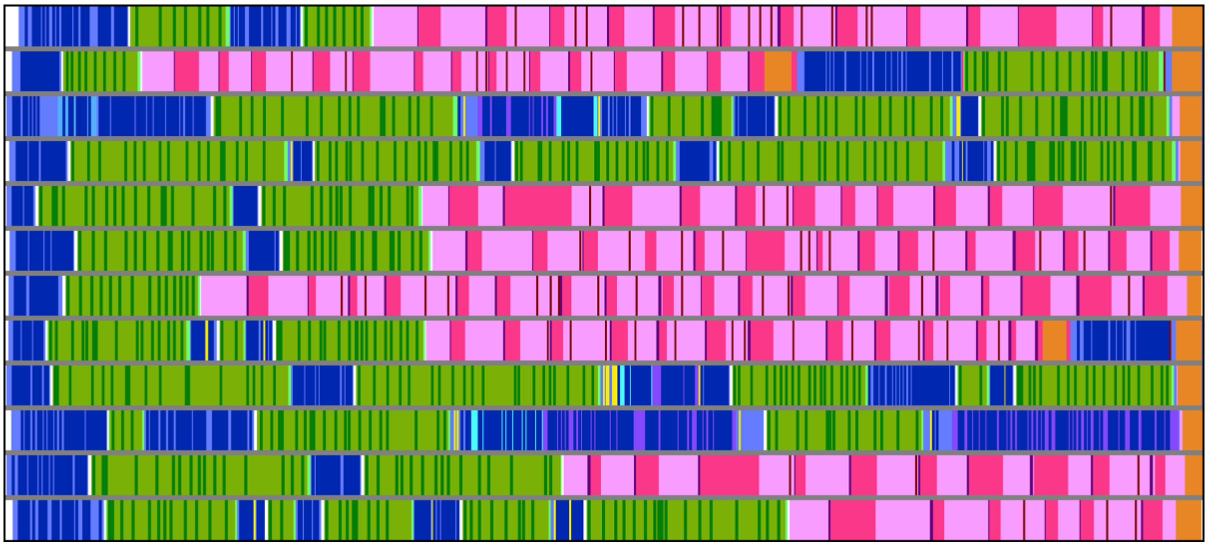

Figure C.7: Session activity visualization for User 7.

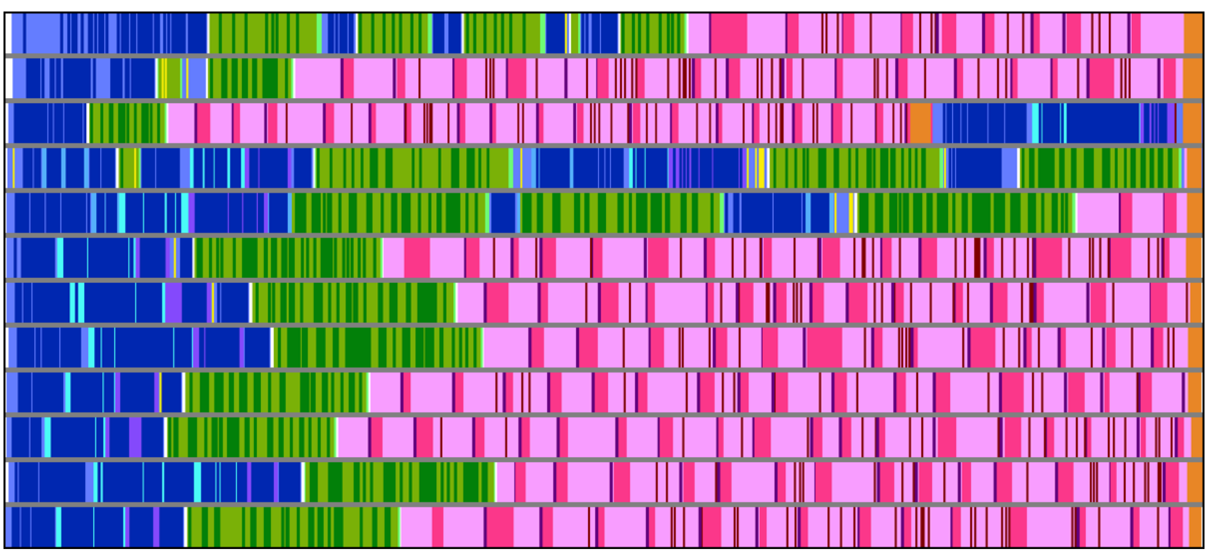

Figure C.8: Session activity visualization for User 8.

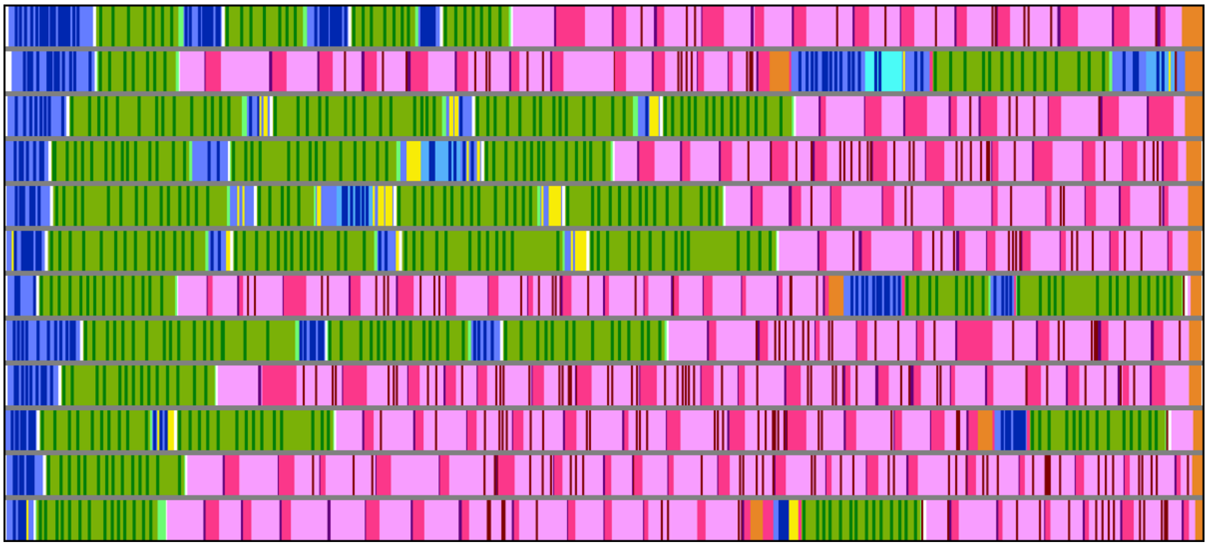

Figure C.9: Session activity visualization for User 9.

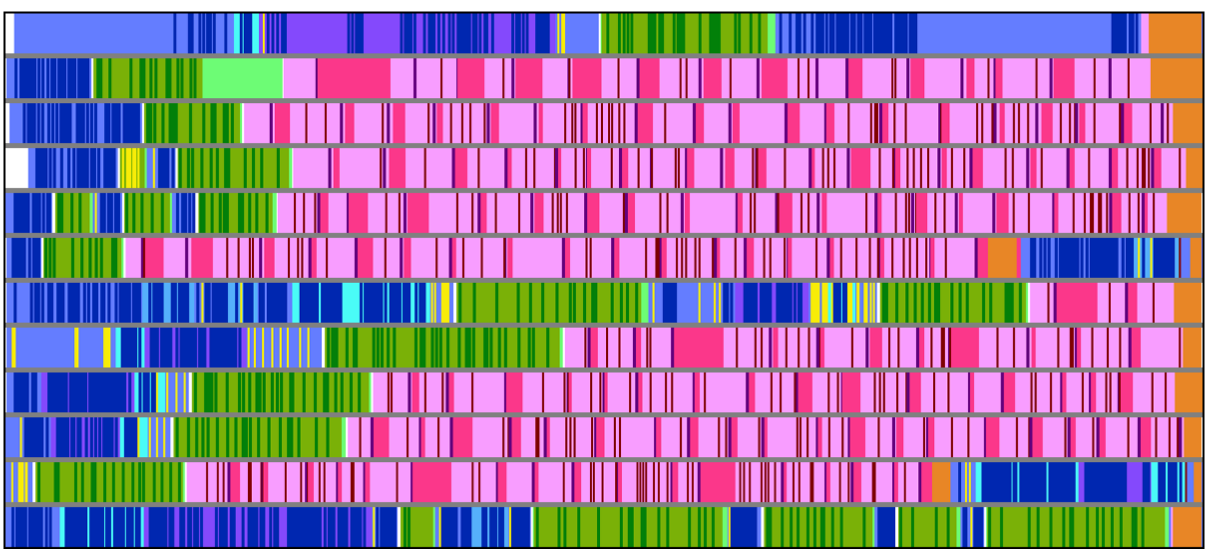

Figure C.10: Session activity visualization for User 10.

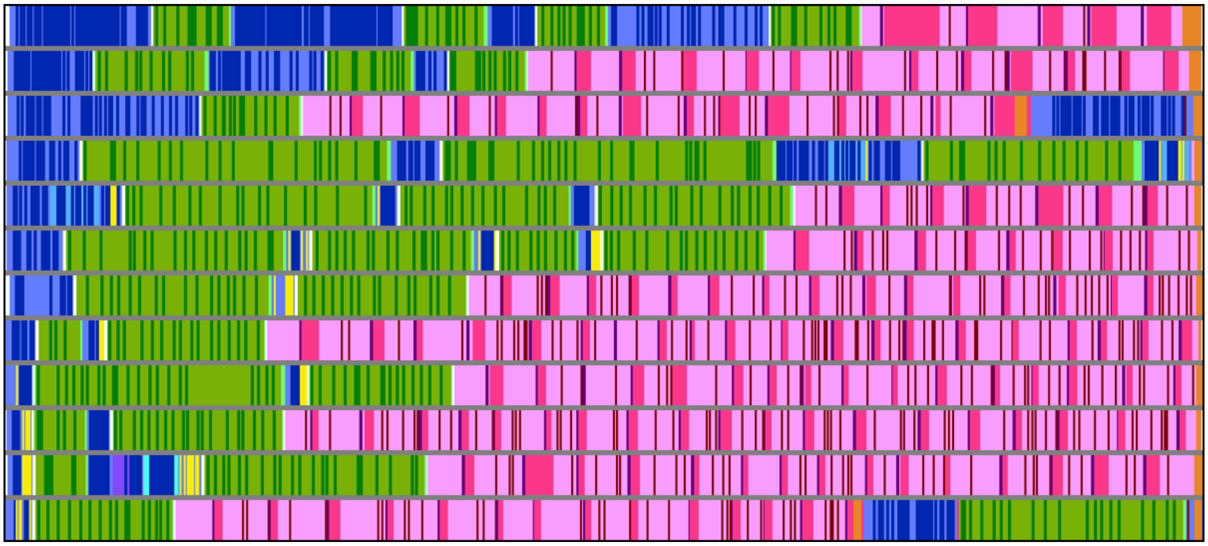

Figure C.11: Session activity visualization for User 11.

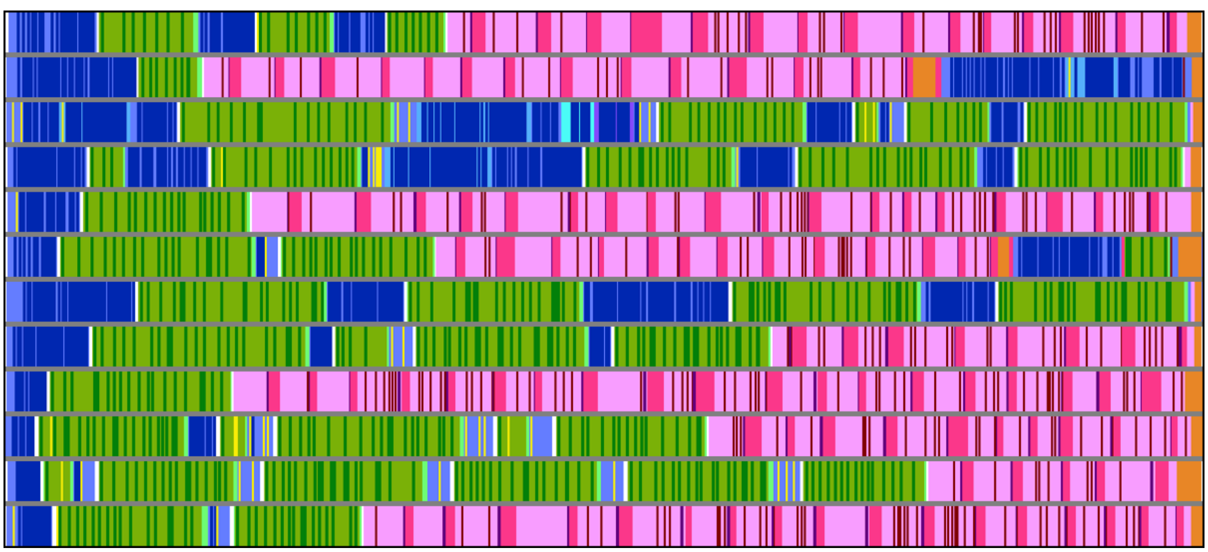

Figure C.12: Session activity visualization for User 12.

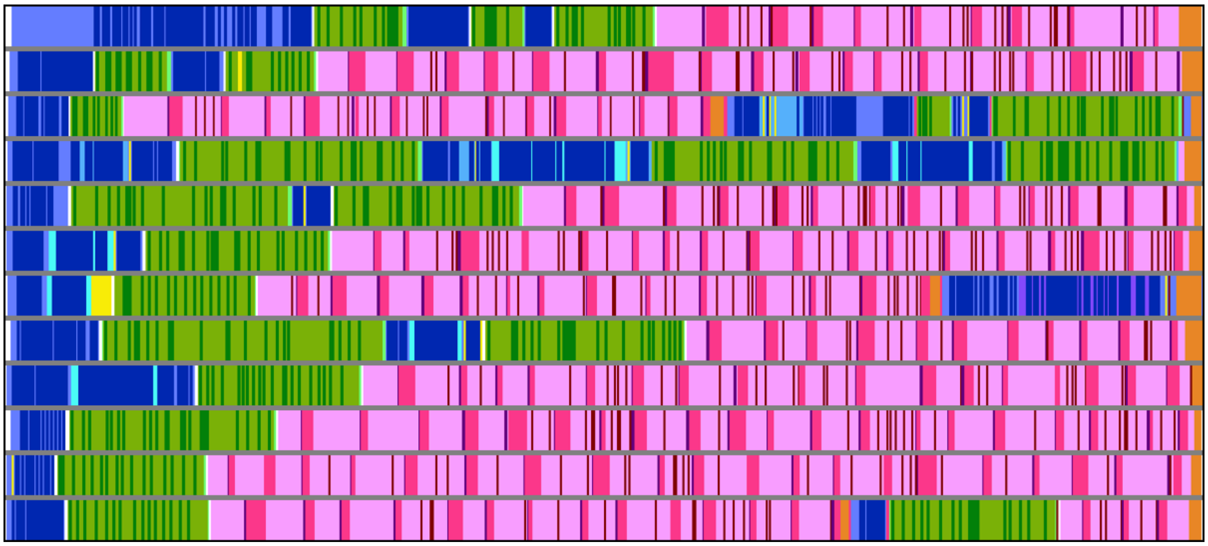

Figure C.13: Session activity visualization for User 13.

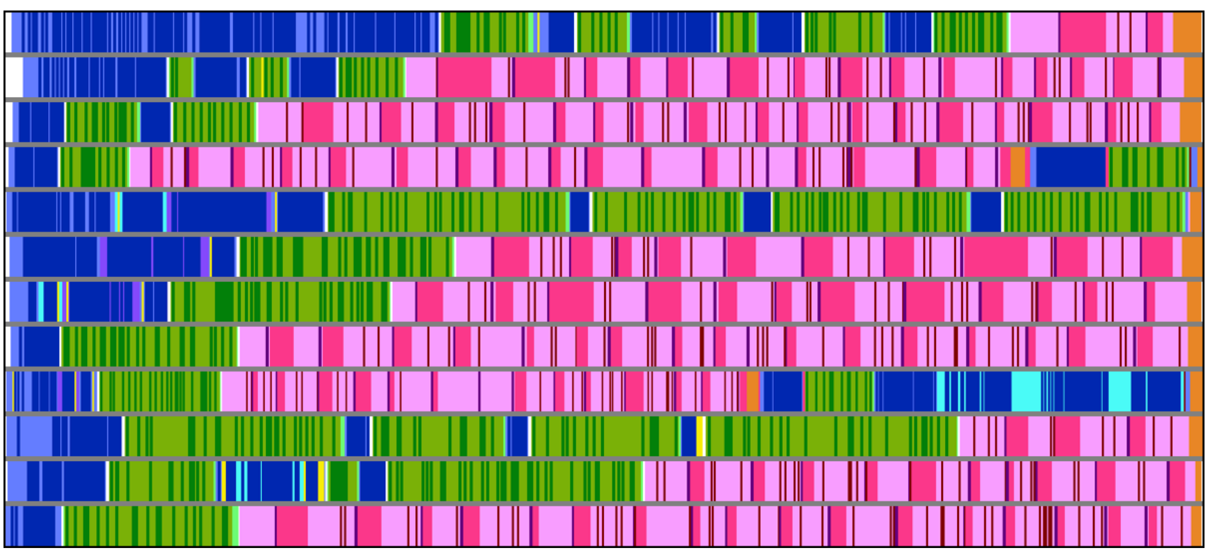

Figure C.14: Session activity visualization for User 14.

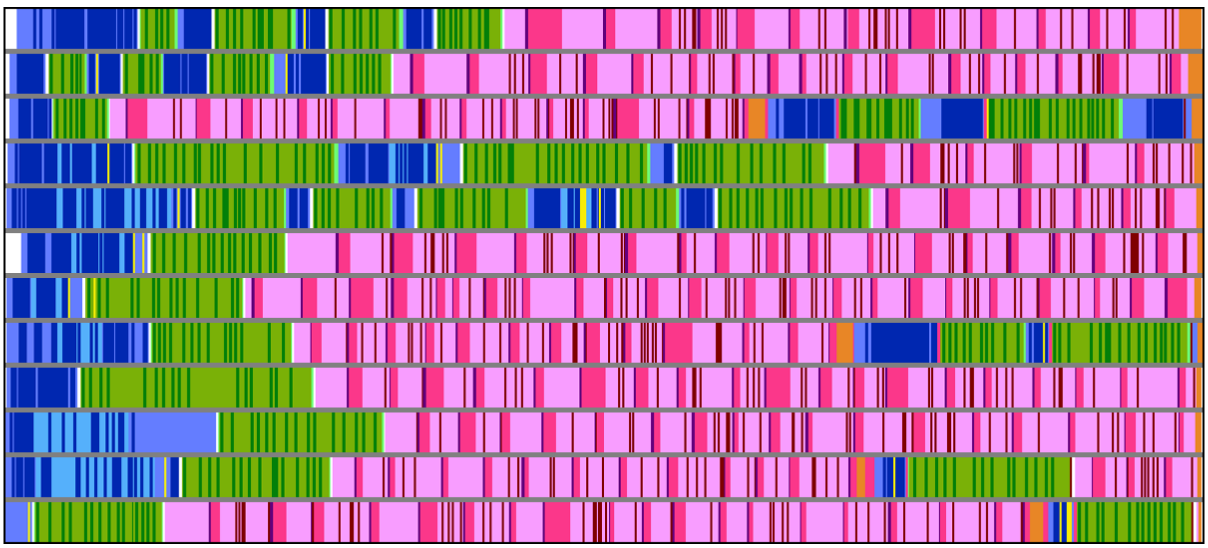

Figure C.15: Session activity visualization for User 15.

APPENDIX D

CODE LISTS FOR LONGITUDINAL MELODY STUDY

```
SCHO - "Spouse: Call me at home"
SLBT - "Spouse: I'm late: I will be there"
SPKS - "Spouse: Pick up the kids from school"
BCMO - "Boss: Call my office"
BMCL - "Boss: I'm in a meeting, call me later"
BICH - "Boss: Sorry: I can't help with that"


CPMS - "Child: Pick me up at school"
CCCL - "Child: Call my cell"
CBHL - "Child: I'll be home late, don't wait up"
FPUB - "Friend: Pick me up some beer"
FMFC - "Friend: Meet for coffee?"
FGGS - "Friend: I've got a great story for you"

YCHT - "Babysitter: Your child is having a tantrum"
YTBE - "Babysitter: There has been an emergency"
YCIS - "Babysitter: Your child is sick"
SCCL - "Spouse: Call my cell"
SOMW - "Spouse: I'm on my way"
SGYM - "Spouse: Got your message"


BCCL - "Boss: Call my cell"
BCTO - "Boss: Come to my office"
BMIC - "Boss: The meeting is cancelled"
CPMF - "Child: Pick me up at my friend's"
CINM - "Child: I need some money"
CIAT - "Child: I aced the test!"


FCCL - "Friend: Call my cell"
FSCD - "Friend: Sorry, can't do it"
FLBT - "Friend: I'm late: I will be there"
YCCL - "Babysitter: Call my cell"
YLBT - "Babysitter: I'm late: I will be there"
YGOP - "Babysitter: We're going outside to play"
```

```
FCCL - "Friend: Call my cell"
FSCD - "Friend: Sorry, can't do it"
FLBT - "Friend: I'm late: I will be there"
YCCL - "Babysitter: Call my cell"
YLBT - "Babysitter: I'm late: I will be there"
YGOP - "Babysitter: We're going outside to play"

BCCL - "Boss: Call my cell"
BCTO - "Boss: Come to my office"
BMIC - "Boss: The meeting is cancelled"
CPMF - "Child: Pick me up at my friend's"
CINM - "Child: I need some money"
CIAT - "Child: I aced the test!"

YCHT - "Babysitter: Your child is having a tantrum"
YTBE - "Babysitter: There has been an emergency"
YCIS - "Babysitter: Your child is sick"
SCCL - "Spouse: Call my cell"
SOMW - "Spouse: I'm on my way"
SGYM - "Spouse: Got your message"

CPMS - "Child: Pick me up at school"
CCCL - "Child: Call my cell"
CBHL - "Child: I'll be home late, don't wait up"
FPUB - "Friend: Pick me up some beer"
FMFC - "Friend: Meet for coffee?"
FGGS - "Friend: I've got a great story for you"

SCHO - "Spouse: Call me at home"
SLBT - "Spouse: I'm late: I will be there"
SPKS - "Spouse: Pick up the kids from school"
BCMO - "Boss: Call my office"
BMCL - "Boss: I'm in a meeting, call me later"
BICH - "Boss: Sorry: I can't help with that"
```

APPENDIX E

VISUALIZATIONS OF PARTICIPANT SESSIONS FOR LONGITUDINAL MELODY STUDY

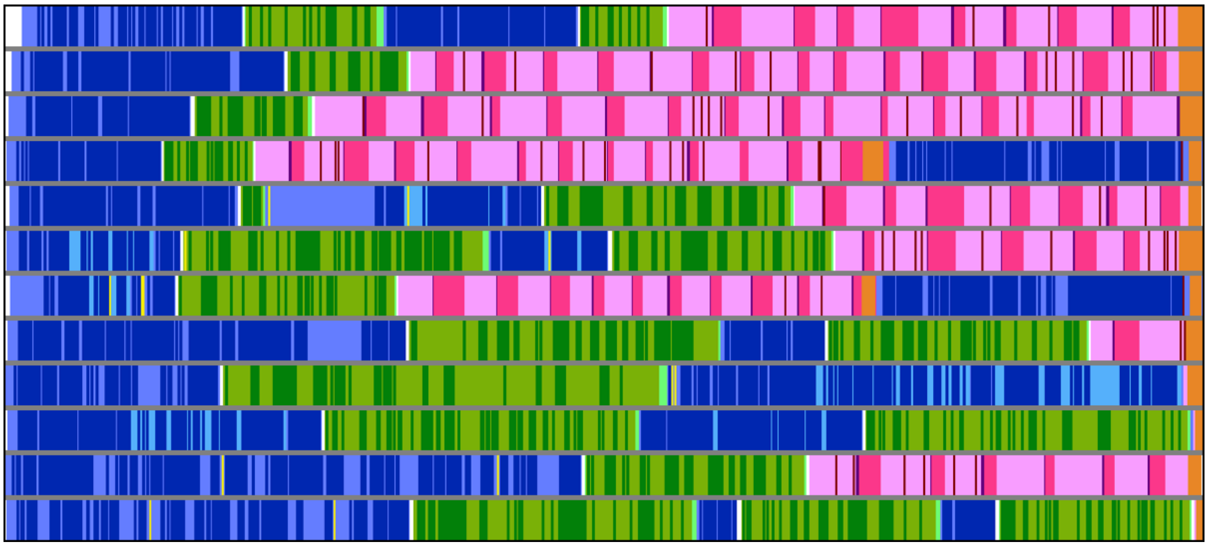

Figure E.1: Session activity visualization for User 1.

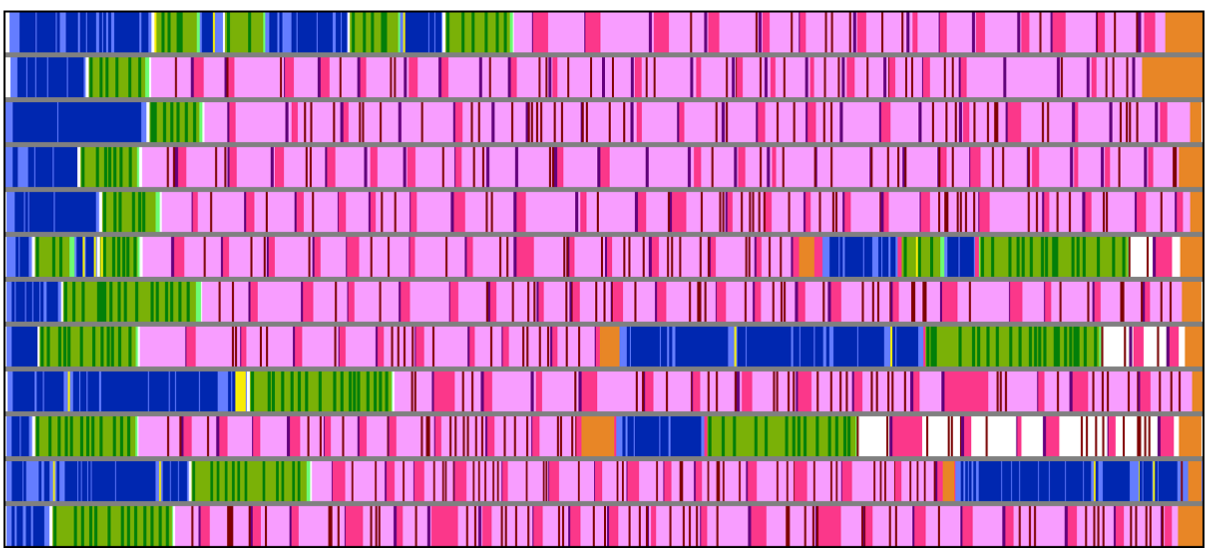

Figure E.2: Session activity visualization for User 2.

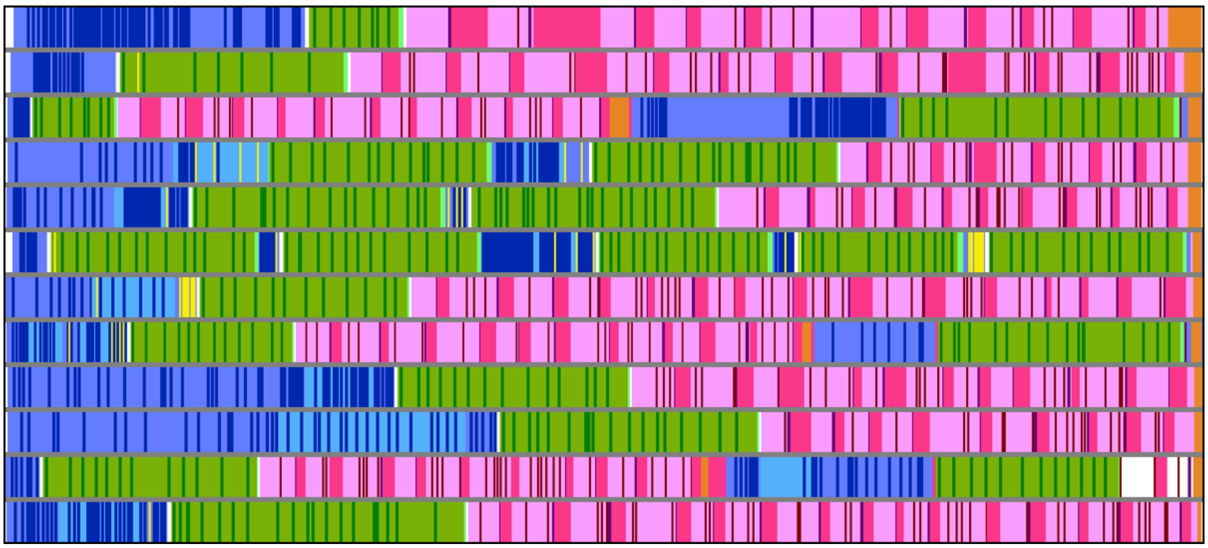

Figure E.3: Session activity visualization for User 3.

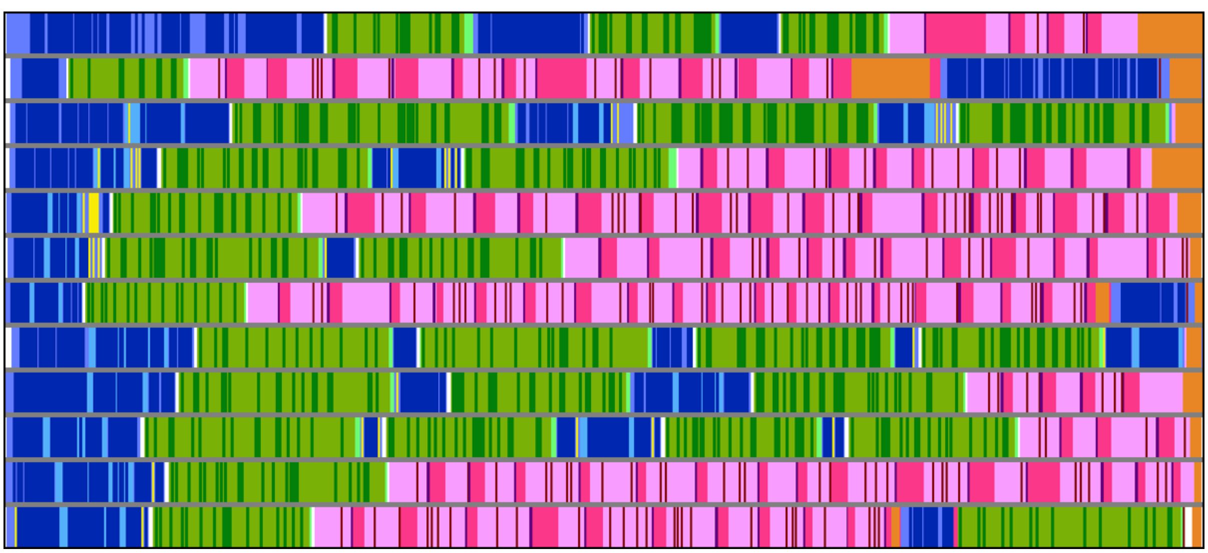

Figure E.4: Session activity visualization for User 4.

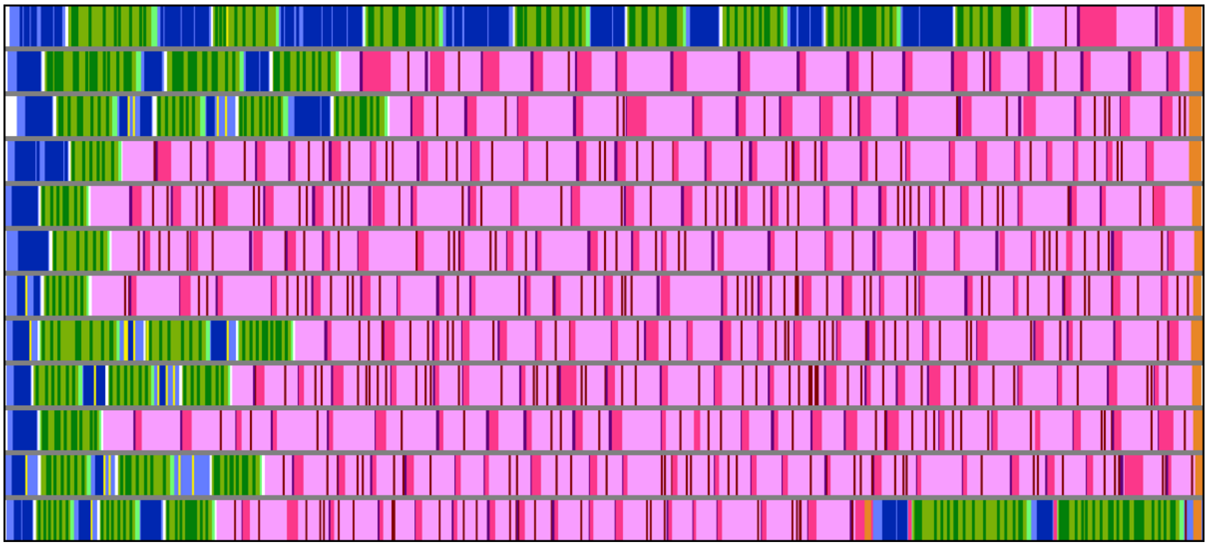

Figure E.5: Session activity visualization for User 5.

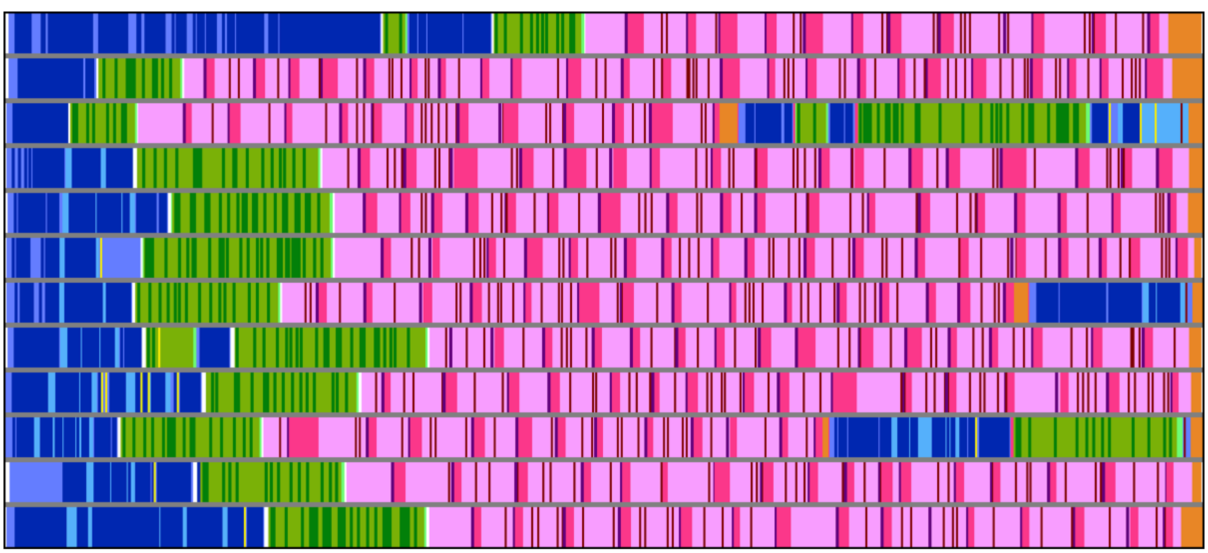

Figure E.6: Session activity visualization for User 6.

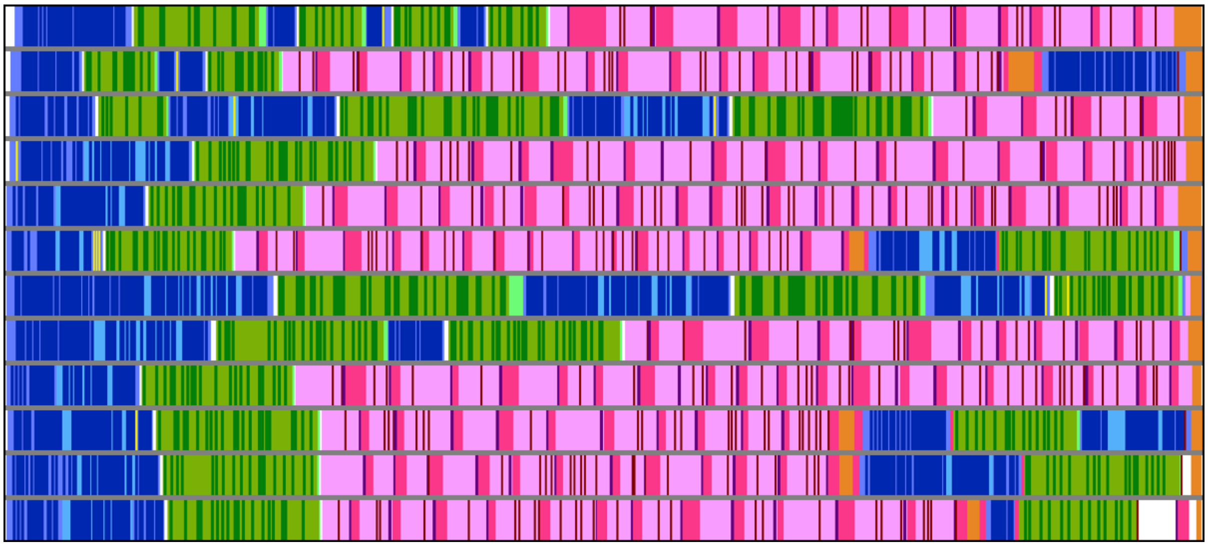

Figure E.7: Session activity visualization for User 7.

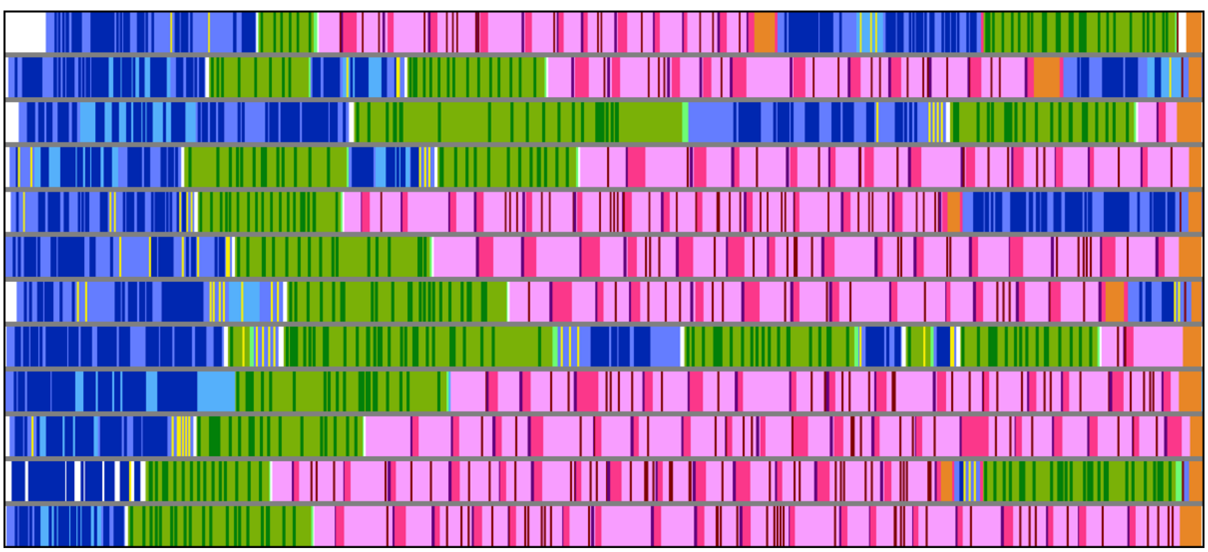

Figure E.8: Session activity visualization for User 8.

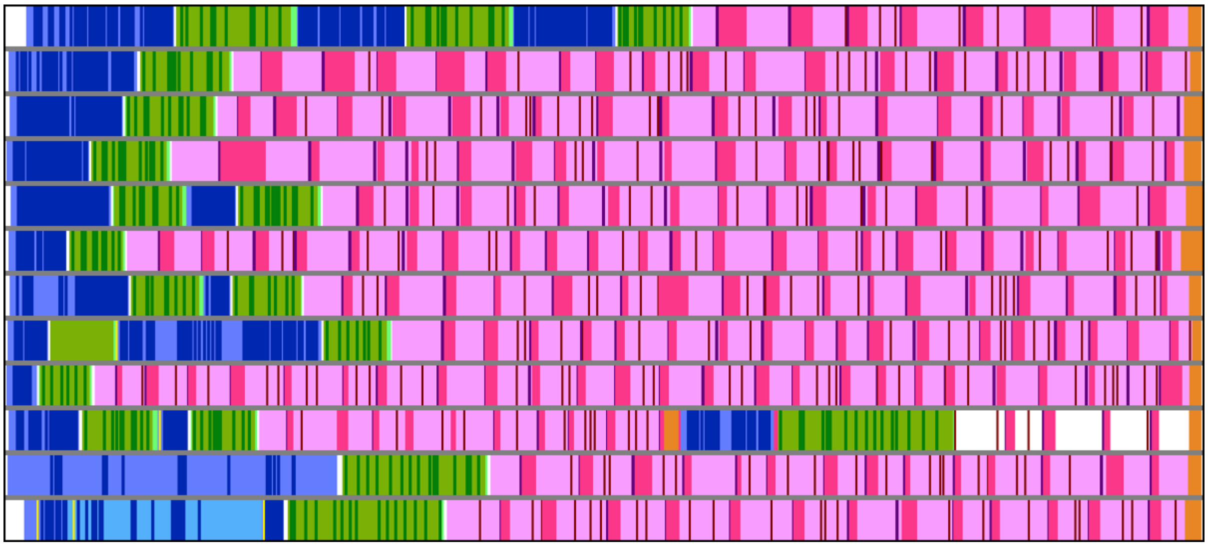

Figure E.9: Session activity visualization for User 9.

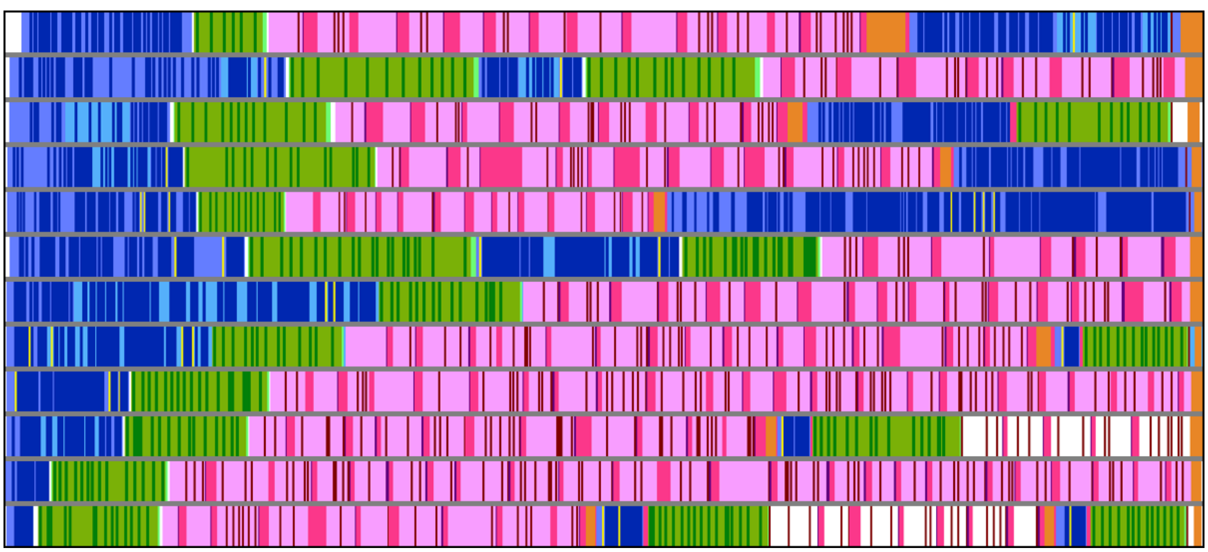

Figure E.10: Session activity visualization for User 10.

APPENDIX F

UBC RESEARCH ETHICS BOARD CERTIFICATES OF APPROVAL

The University of British Columbia
Office of Research Services
Behavioural Research Ethics Board
Suite 102, 6190 Agronomy Road, Vancouver, B.C. V6T 1Z3

CERTIFICATE OF APPROVAL - MINIMAL RISK AMENDMENT

PRINCIPAL INVESTIGATOR:	DEPARTMENT:	UBC BREB NUMBER:
Karon E. MacLean	UBC/Science/Computer Science	H01-80470

INSTITUTION(S) WHERE RESEARCH WILL BE CARRIED OUT:

Institution	Site
UBC	Vancouver (excludes UBC Hospital)

Other locations where the research will be conducted:
N/A

CO-INVESTIGATOR(S):
Ricardo Pedrosa
Susan Gerofsky
Bradley A. Swerdfeger
Idin Karuei
Matt Savage-LeBeau
Steve Yohanan

SPONSORING AGENCIES:
Natural Sciences and Engineering Research Council of Canada (NSERC)
Various Sources

PROJECT TITLE:
Orsil title - Low-Attention and Affective Communication Using Haptic Interfaces

Expiry Date - Approval of an amendment does not change the expiry date on the current UBC BREB approval of this study. An application for renewal is required on or before: June 23, 2009

AMENDMENT(S):	AMENDMENT APPROVAL DATE: July 8, 2008

Document Name	Version	Date

The amendment(s) and the document(s) listed above have been reviewed and the procedures were found to be acceptable on ethical grounds for research involving human subjects.

Approval is issued on behalf of the Behavioural Research Ethics Board

Dr. M. Judith Lynam, Chair
Dr. Ken Craig, Chair
Dr. Jim Rupert, Associate Chair
Dr. Laurie Ford, Associate Chair
Dr. Daniel Salhani, Associate Chair
Dr. Anita Ho, Associate Chair