Learning and Identifying Haptic Icons under Workload

Technical Report TR-2004-15

Andrew Chan, Karon MacLean, Joanna McGrenere
Department of Computer Science, University of British Columbia, Canada
chana@cs.ubc.ca, maclean@cs.ubc.ca, joanna@cs.ubc.ca

Abstract

This work addresses the use of vibrotactile haptic feedback to transmit background information with variable intrusiveness, when recipients are engrossed in a primary visual and/or auditory task. Our testbed will be a novel urgency-based turn-taking protocol for remote collaboration, and our setup uses inexpensive off-the-shelf technology. We describe two studies designed to (a) perceptually optimize a set of vibrotactile "icons" for our protocol and (b) evaluate users' ability to identify them in the presence of varying degrees of workload.

We found that 7 icons learned in approximately 3 minutes were each typically identified within 2.5 s and at 95% accuracy in the absence of workload. With added visual and auditory distractor tasks, the time required to detect a change in haptic icon increased from 1.9 s to an average of 4.3 s. We further provide initial parameters to help designers intelligently balance the need to support communication while minimizing disruption.

1. Introduction

Communicating information through haptic feedback is a young but growing resource for user interface designers. Force feedback plays a direct role in many virtual and teleoperated environments; e.g. in laparoscopic surgery simulators, it is used to render a haptic version of a scene or task also displayed visually [26]. More recently, vibrotactile displays are being used to transmit abstract information in consumer applications. Often, these displays convey information to a user whose attention is focused on a different primary task. Cell phone manufacturers are experimenting with vibrotactile "touch tones", hypothesizing that different vibrations can carry specific meanings like their auditory counterparts, but without their intrusiveness. Auto manufacturers have begun introducing haptics into vehicle cockpits: BMW's iDrive was first to market [1].

Touch, which supports our rapid and integrated perception of texture, temperature, frequency, contour and heft, should be a viable alternate communication medium when other sensory modalities are engaged. However, the mixed success of attempts to date suggest we would benefit from a more comprehensive understanding of how easily people can learn and utilize information conveyed by haptic feedback in the presence of significant cognitive workload. This need grows more urgent as we bring haptic feedback into time- and safety-critical settings.

1.1 Context: Haptic Support of Turn-Taking

While the questions addressed in this work are general, we needed a specific design and evaluation context to explore how haptic feedback might be used with varying levels of peripheral awareness under different amounts of workload. Our research in application sharing by distributed users provided us with such a testbed and guided our design decisions.

In the current commercial state-of-the-art for application sharing, users select a running application and permit others to view it from their own workstations while communicating via text chat or audio conferencing [2, 3]. Technological constraints permit only one user to interact with the shared application at a time, necessitating a protocol for sharing control.

One typical approach is the *give* protocol: when a user requests control, the in-control user is expected to relinquish it. Requests are presented visually through modal dialog boxes or tool-tips. The former is disruptive, and the latter can be missed if the user is engrossed in the task. If the user in control chooses to temporarily retain control, there is no reminder that a request was made.

We have developed a new urgency-based turn-taking protocol that uses "haptic icons," haptic stimuli to which meaning has been associated [17]. Periodic haptic feedback from a vibrotactile display tells users whether they are in control of a shared application, waiting to take control, or simply observing their collaborators' actions. It also tells an in-control user whether others want control urgently, gently or not at all, and periodically reminds him of his collaborators' (possibly changing) intentions without disrupting his actions.

1.2 Haptic Icons for Workload Conditions

Designing our protocol raised several questions: What set of haptic sensations will be easiest for the user to distinguish? How can we maximize users' ability to recall their assigned meanings when preoccupied? To of knowledge, these questions have not yet been addressed for the haptic sense. Here, we explore them before integrating haptic communication into our proposed system, in the hope of both improving its effectiveness and moving towards a general process for designing flexible haptic icons that will be robust to varying workload.

This paper describes two studies designed to address these questions, using the context of our testbed application as an exploratory vehicle. The first study helped us to optimize the perceptual composition of three families of haptic stimuli that support our new turn-taking protocol. In the second study, we tested subjects' ability to associate these haptic stimuli with a set of arbitrary labels, and recall the labels while performing distractor tasks that simulated the cognitive effort of working in a collaborative environment. We employed inexpensive off-the-shelf haptic technology to broaden the applicability of our investigation.

2. Related Work

2.1 Haptic Communication

Designing haptic feedback to convey messages requires an understanding of our tactile psychophysical capabilities. A comprehensive understanding is emerging through the work of researchers such as Klatzky and Lederman, who have for example documented our exquisite sensitivity to texture felt through a probe [15, 16]. Research results have been used to develop basic guidelines for using vibrotactile displays [25]. Tan et al. measured information transfer rates of 2-3 items/second for vibrotactile stimuli independent of duration [24], suggesting that appreciable content can be conveyed through this channel.

Haptic feedback has been used to aid collaboration in virtual environments: Sallnas et al. had users jointly arrange virtual cubes [22], and Basdogan et al. asked them to jointly move a virtual ring along a wire without touching the wire [4]. Both found that haptic feedback significantly improved task performance and aided cooperation. In a different approach, Oakley et al. added haptic feedback to a synchronous shared editor so that users could locate and move each other around [21]. Guiding forces have been shown to positively and nonintrusively influence behavior during an engrossing task [11, 23].

Other work has been directed towards building an abstract haptic language. One approach is to load haptic stimuli with comprehensible information [8, 17, 24]. However, we are not aware of any studies that specifically address the encoding of information in families of haptic icons, nor evaluation of user's comprehension of the icons under workload.

2.2 Turn-Taking

A variety of protocols have been proposed [6, 13, 20] to mediate control over a shared artifact. Two protocols commonly cited are *give* (a user who requests control must wait for the user in control to voluntarily relinquish it) and *take* (the user requesting control is immediately given control). However, only a few comparisons of the effectiveness of these protocols have been conducted (e.g., studies by Inkpen et al. [14] and Mackinlay et al. [18]) with results that are inconclusive.

Conversational analysis reveals that when physically co-located, our body language communicates to our partners information such as how urgently we wish to speak [10]; and that non-verbal communication is crucial for smooth and efficient information transfer [7]. We expect similar non-verbal communication needs for collaborative work; in particular, its use to express a desire to assume control with varying degrees of urgency. However, none of the turn-taking protocols we have observed provide flexibility in the means of requesting application control, nor do their implementations support user management of a sent signal's intrusiveness – features we take for granted in everyday conversation. Visual elements such as dialog boxes and tool-tips are not ideal for this function: by sharing the medium of the primary task, they can grossly interfere when intrusive, and even subtle status updates can impinge undesirably on visual attention.

3. Approach

We followed a 3-step process in designing haptic icons for our turn-taking protocol: we prototyped families of icons using common metaphors, optimized the perceptual spacing of the icons in Study 1, and evaluated subjects' ability to learn and recall the icons under workload in Study 2.

3.1 Icon Set Design: Families and Metaphors

The use of haptic icons in general can be related to past research in synthesizing non-verbal auditory cues. We chose to create three families of icons to support our turn-taking protocol rather than a perceptually ‘flat’ stimuli set: icons in a family shared tactile features and had related meanings. The families and icons are described in Table 1 and below:

- The first family had two transient stimuli associated with *Changes in Control*, indicating that the receiving user has just (1) gained or (2) lost control of the application.
- The second family consisted of three periodic, ongoing sensations delivered to the user *In Control*: (3) no outstanding requests, or someone has (4) gently or (5) urgently requested control.
- The final family included two sensations delivered periodically to a user *Waiting for Control*, following a (6) gentle or (7) urgent request.

These families mapped well to the different states in our turn-taking protocol, and we expected that users would be able to cognitively chunk [19] the icons, thereby facilitating learning. This approach was first suggested in work on earcons [5, 9].

Users must be able to learn the haptic icons easily. While earcons used chunking, the association between an earcon and its meaning is abstract and possibly difficult to learn. In contrast, auditory icons [12] use sounds from real-world objects to create intuitive mappings between an object and its icon. Likewise, we designed each of our icon families using common metaphors that we expected users would understand:

- The two icons in the Change in Control family used the haptic equivalent of the two-tone sound made when a PCMCIA card is inserted into and removed from a Windows laptop. We hypothesized that making the icons mirror opposites would make them as intuitive to learn as the audio versions on which they were based.
- The icons in the In Control family used a heartbeat metaphor. A gentle sensation was used for the no outstanding request icon, but progressively more intrusive sensations were used for the gentle and urgent request icons.
- A “pulse” sensation suggesting a person tapping or drumming her fingers while waiting in line was used for the Waiting for Control icons.

Since feedback was provided to a user in control or waiting for control, we decided it would be appropriate not to provide feedback when a user is simply observing the collaboration.

3.2 Setup: Haptic Display

We used Logitech iFeel mice to deliver haptic feedback in our studies. These are standard optical mice with the addition of an embedded vibrotactile display (technology licensed from Immersion Corp.). Although the range of frequencies and amplitudes the iFeel supports are limited, we wanted to find how much we could accomplish using off-the-shelf technology. This consideration also motivated the choice of a mouse rather than introduction of a new hardware element (e.g. a glove). We created haptic sensations using Immersion Studio [11], a GUI-based haptic editor, and integrated the sensations into our test program using an Immersion API.

The iFeel has a frequency response of 0.01-500 Hz. Immersion Studio employs a “magnitude” from 0-10,000. It is difficult to interpret this quantity, particularly since it interacts with the stimulus’ frequency: a 2000 magnitude vibration at 100 Hz feels significantly stronger than a 2000-magnitude vibration at 20 Hz. In general, we observed that if two stimuli are the same frequency and above 10 Hz, they must have a magnitude difference of at least 1000 to feel perceptually different.

Subjects used Pentium IV 2.67 GHz computers with 512 MB of RAM, running Windows XP Professional. To mask audible noise from the vibrotactile display that might influence their perception of the haptic feedback, subjects wore Bose noise-canceling headphones and listened to recorded white noise.

4. Study 1: Optimizing Families of Haptic Icons

Our family-based, metaphor-driven approach yielded a preliminary set of haptic icons to support our turn-taking protocol. However, we noted that our set of icons representing the In Control states was only one of many possibilities that would satisfy the heartbeat metaphor. We also wanted to verify that the Change in Control icons would not be confused with the In Control icons, and that all the icons were mutually distinguishable.

To address these issues, we used a technique based on Multidimensional Scaling (MDS) that provides an estimate of the perceptual similarity of haptic stimuli [17]. MDS takes as input an $n \times n$ matrix of data indicating the dissimilarity of a set of n stimuli, and outputs their relative locations in the Euclidean space of an m-dimensional graph, where m is user-specified (typically, $m << n$). In our usage, the distance between stimuli was an estimate of their perceptual similarity.

Family	Icon ID	State	Haptic Sensation	Study 2 Label
Change of Control	CH+	User has gained control of the shared application	0.25 s, 3000-magnitude, 100 Hz vibration, followed by a 0.05 s pause, followed by a 0.25 s, 8000-magnitude, 200 Hz vibration (modified in Study 2) *Feel:* short, weak buzz followed by short, strong buzz	Awake
	CH-	User has lost control of the shared application	0.25 s, 8000-magnitude, 200 Hz vibration, followed by a 0.05 s pause, followed by a 0.25 s, 3000-magnitude, 100 Hz vibration (modified in Study 2) *Feel:* short, strong buzz followed by short, weak buzz	Asleep
In Control	IN	User is in control of the shared application	1 s, 500-magnitude, 60 Hz vibration; 1 s delay between iterations *Feel:* just-noticeable vibration	Low Stress
	IN+	User is in control, but someone has gently requested control	1 s, 5000-magnitude, 60 Hz vibration; 1 s delay between iterations *Feel:* noticeable, but not unpleasant vibration	Medium Stress
	IN++	User is in control, but someone has urgently requested control	0.7 s, 5000-magnitude, 100 Hz vibration, followed by a 0.1 s pause, followed by a second identical vibration; 0.6 s delay between iterations *Feel:* very noticeable, somewhat unpleasant vibration	High Stress
Waiting for Control	WAIT	User has gently requested control	Single pulse; 1 s delay between iterations *Feel:* a light, quick tap	Bored
	WAIT+	User has urgently requested control	Two pulses, separated by a 0.15 s pause; 1 s delay between iterations *Feel:* two light, quick taps	Really Bored

Table 1: Icon Set Composed after Study 1

First, we created 24 candidates for the IN icons, varying along three parameters. We used three frequencies (20, 60, and 100 Hz); the iFeel motor produced unmaskable confounding auditory noise above 100 Hz, and insufficient magnitude below 20 Hz. We sampled the iFeel's magnitude range with four values (500, 2000, 5000, 8000). Two levels of a temporal variable were introduced by playing each of these 12 stimuli either singly for 1000 ms, or in two 700 ms bursts separated by a 100 ms delay.

To these 24 stimuli we added the CH+ and CH- icons we had prototyped. We opted not to test the WAIT icons, as we were confident their pulse-based nature would make them distinct from the other vibration-based stimuli and each other, and their presence could have distorted our MDS results.

Ten subjects (6 male, 4 female) participated in the study. They ranged in age from 21 to 31 and were each paid $10 for a one-hour session. Six of the subjects had little exposure to haptic feedback prior to the study. Subjects were asked to sort the 26 stimuli into various numbers of categories based on their similarity (see [17] for a complete description of the method), after which MDS was used to analyze the data. Subjects were also asked to rate the intrusiveness and pleasantness of the stimuli on Likert scales.

4.1 Results

Unlike [17], we found that a 3D MDS analysis substantially reduced the goodness of fit over a 2D analysis. Based on this analysis, we selected a set of IN icons with maximal perceptual distinctiveness Likert-scale ratings that conformed to our heartbeat metaphor; IN was considered very pleasant and not at all intrusive, IN++ was considered somewhat unpleasant but quite noticeable, and IN+ fell in between. The CH icons appeared to be distinguishable from these icons and from each other. The icons selected after Study 1 are described in Table 1.

In summary, in this step of the design process we were able to quickly adjust and validate the perceptual spacing of a set of seven haptic icons clustered into three metaphor-based families. To learn whether the spacing achieved in this way was suitable required deployment in a more realistic context, which we examined in Study 2.

5. Study 2: Identifying Haptic Icons

The second study addressed the ability to learn and recall meanings for haptic icons, a crucial aspect of utility in any application. Specifically, its focus was to examine the ease with which the stimuli composed in the first study could be learned and recalled even while subjects were engaged in distractor tasks. Our questions here were threefold:

1. How long will it take subjects to learn the meanings associated with each of the stimuli?
2. How quickly can users detect and identify an icon change? This corresponds, for example, to recognizing an externally driven change in a state that is communicated by periodic haptic feedback.
3. How will the ability to detect and identify changes be affected by workload? Users will often need to notice and recognize icons while another task absorbs their attention, and ideally workload will affect response to urgent and non-urgent icons differently

We thus designed the second study to include a learning phase, followed by an evaluation phase where we measured detection and identification performance in the presence of varying degrees of visual and auditory workload. Both phases were completed by every subject during a 1.5 hour session. We used the 7 icons from Study 1 (see Table 1), with slight modifications to the CH family[1].

Because we felt subjects would find it difficult to learn the turn-taking protocol meanings without an elaborate explanation, we substituted simpler labels that still reflected the icons' familial relations (shown in Table 1).

Sessions were automated. To avoid subtle strategic bias from variations in instruction delivery [11], subjects read instructions on-screen and in a booklet provided at the beginning of the session.

5.1 Learning Phase

Subjects were instructed to learn the meanings associated with the 7 haptic stimuli "as quickly as possible". To proceed to the evaluation phase, subjects had to score over 90% on a test.

Subjects were first presented with a simple GUI that allowed them to play back the 7 icons as many times as they wanted in any order, by clicking on its graphical trigger; they chose without penalty when to proceed to the test. The icons' triggers were labeled and arranged by family, but no other aids were provided to help subjects learn the icons.

In the learning test, subjects felt a haptic icon and identified it by selecting the correspondingly labeled radio button. Each icon was presented three times for a total of 21 trials, randomized with the constraint that the same icon was never presented twice in a row. To prevent positional memorization, the labeled radio button response choices were randomly re-ordered on each trial. As well, subjects were only told whether they had passed or failed the test, without specific feedback. When subjects correctly identified 19 or more icons (allowing 2 errors), they proceeded to the evaluation phase; otherwise, they returned to the initial screen for more practice, before repeating the test.

5.2 Evaluation Phase

During the evaluation phase, subjects' ability to recall the meanings they learned in the learning phase were tested under three increasingly difficult workload conditions: control, visual distractor, and visual+auditory distractor. Since our turn-taking protocol is intended for use with a visual task, we did not include an auditory-only distractor condition in this study. The order of the conditions was counterbalanced across subjects.

In the control condition, icons were presented as a sequential pair. The transition from the first icon to the second occurred after a randomly chosen delay of 10, 15, or 20 seconds. Icons not periodic by design (CH-, CH+) were repeated every 2 s. Subjects were instructed to press the space bar as soon as they noticed the change. Although not specifically instructed to do so, all subjects used their non-mouse hand to press the space bar. A modal dialog box then appeared that listed the 7 icons, again grouped by family. Subjects identified the second icon by selecting a radio button, pressed an OK button, and then proceeded to the next pair. If the subject had not pressed the space bar 10 seconds after the transition, this was counted as a "missed transition" and the dialog box was displayed, forcing the user to identify the second icon.

Each subject responded to a total of 35 pairs, consisting of 5 transitions to each of the 7 icons. Testing all 42 possible transitions would have made the duration of the study unreasonable, so we selected the subset possible in our turn-taking protocol to help us predict performance. Transitions were presented in random order.

In the visual distractor condition, subjects had to perform a visual task of solving a picture puzzle while performing the icon identification described for the control condition. An image was subdivided into a grid of 12 pieces and the pieces were randomly rearranged. Subjects were instructed to rearrange the pieces to restore the original image, which was displayed beside the scrambled puzzle. A puzzle piece could be swapped with any other piece by dragging one piece on top of

[1] Pilot subjects reported utilizing audible noise from the iFeel to distinguish between these stimuli, despite wearing noise-canceling headphones playing white noise. As a result, the CH+ signal was changed to a 0.4 s, 1000-magnitude 100 Hz vibration, followed by a 0.2 s, 8000-magnitude, 100 Hz vibration. The CH- signal was the mirror opposite.

the other. When the subject had successfully solved a puzzle, a new puzzle was presented. Images were randomly selected from a set of 65 images until all icon transitions had been presented, and the same image was never repeated during a session.

In the visual+auditory distractor condition, subjects had to listen for a keyword to be spoken while also performing the tasks described in the visual distractor condition. The keyword "blue" was spoken 30 times at random intervals interspersed with approximately 120 enunciations of 14 other colors in this condition, thus requiring subjects to attend to the audio stream. When subjects heard the keyword, they pressed the "b" key on the keyboard with their non-mouse hand.

The haptic, visual, and auditory tasks in our study were completely unrelated, whereas in our intended application and many others, the visual and auditory channels would be used in concert, with the haptic feedback used to mediate turn-taking. This study design allowed us to predict the performance of our haptic icons conservatively.

In each condition, subjects first practiced on five pairs of icon transitions to familiarize themselves with the user interface for that condition and thereby mitigate learning effects. Subjects were also given an opportunity before each condition to review the 7 icons, using the same UI as in the learning phase of the experiment. This was done because pilot subjects reported becoming unsure over time as to whether they had associated the stimuli with their meanings correctly. This is probably because subjects never received reinforcement in stimulus identification in either the learning or the evaluation phase.

In summary, the evaluation phase was a 3 conditions x 7 stimuli x 5 transitions design, where all factors were within-subjects. The order of the 3 conditions was counterbalanced, and icon transitions were delivered randomly within each condition. Thus, subjects each completed a total of 105 trials.

5.3 Performance Metrics

We measured several aspects of subjects' performance, including:

- Time spent **learning** the associations between stimuli and their meanings
- Time required to **detect** icon transition each trial
- Time required to further **identify** the second icon in the pair, once the transition had been detected
- The number of **missed** transitions
- The number of **correctly identified** icons
- The number of puzzle pieces correctly placed in the visual distractor task and the number of audio keywords correctly identified in the auditory distractor task

We note that by collecting the icon identification data through a modal dialog box, subjects did not have to multi-task while identifying icons. Other methods of gathering the data (such as verbal reports) would neither have provided accurate identification times nor forced subjects to identify the second icon in each pair. As a result, our data on identification time and identification accuracy reflects how subjects' ability to "context switch" to the identification task was affected by different levels of workload.

5.4 Results

Six males and six females participated in the study. Subjects ranged from 17 – 28 years old; four subjects had no exposure to haptic feedback prior to the study, and the remaining eight were novice users. Due to an oversight, one subject had also participated in Study 1; his data was included only after verifying that he was not an outlier. Subjects were paid $10 for a 1.5 hour session.

To encourage brisk execution, subjects were informed that the four subjects with the best overall "score" would receive an additional $10. To avoid biasing any one task, instructions explicitly directed subjects to pay equal attention to the haptic identification and the visual and auditory distractor tasks (when present) in order to maximize their score. On average, during the evaluation phase the control, visual distractor, and visual+auditory distractor conditions were completed in 11, 12, and 13 minutes respectively.

A series of repeated-measures ANOVAs with an alpha level of 0.05 was run. When the data failed Mauchly's Test of Sphericity, the Huynh-Feldt correction was applied, reducing the degrees of freedom in several *F*-tests. In keeping with the exploratory nature of this work, we conducted post-hoc pair-wise comparisons liberally, but for protection used a Bonferroni adjustment, also at a 0.05 alpha level.

Learning Time: Learning time was calculated as the amount of time spent exploring the set of haptic icons. The learning test time was not included, as this would have unfairly penalized subjects who required multiple attempts to pass the test. Subjects spent approximately three minutes learning the 7 haptic icons (mean 177 s, std. dev. 114 s). Learning times ranged from 56 to 446 s. This result is lower than we expected, given the unfamiliarity of the medium, and suggests deployment in consumer applications would be feasible.

Condition	Mean	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
control	1815	1186	2444
v distractor	3507	2421	4594
v+a distractor	4269	2998	5540

Table 2: Mean Detection Times for each Condition (ms)

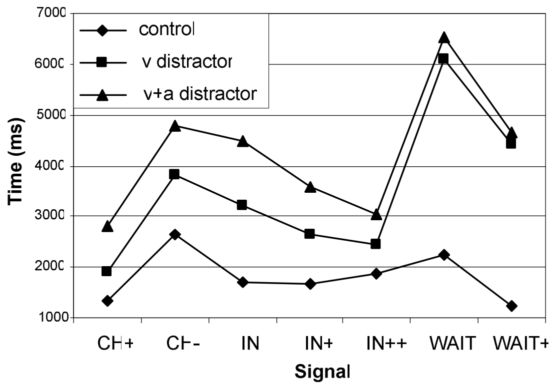

Figure 1: Mean Detection Times for each Condition

Icon	Condition Pairs: control vs. v distractor	v distractor vs. v+a distractor	control vs. v+a distractor
CH+	0.884	0.010*	0.015*
CH-	0.022 *	0.435	0.039*
IN	0.128	0.218	0.020*
IN+	0.338	0.168	0.028*
IN++	0.797	0.743	0.452
WAIT	< 0.001*	1.000	< 0.001*
WAIT+	0.002*	1.000	0.001*

Table 3: *p*-values for Sig. Differences in Detection Times across Conditions (* means *p* < 0.05)

*Detection Time***:** Detection time was calculated from the time the second icon in a pair began playing to when the subject pressed the space bar, with missed transitions set to 10 seconds. Table 2 shows mean detection times by condition. Figure 1 shows a graph of the mean detection times in each condition for each icon.

Overall, there was a significant impact of condition on the detection time ($F_{1.297,\ 14.270} = 20.359$, $p < 0.001$, partial $\eta^2 = 0.649$). Mean detection time in the visual distractor condition was nearly double that of the control condition, and that of the visual+auditory distractor condition was 22% longer again than the visual distractor condition. Both comparisons were significant.

The results also revealed a significant interaction between condition and icon, indicating that the detection times of some icons were more sensitive to condition than others ($F_{7.940,\ 87.342} = 4.472$, $p < 0.001$, partial $\eta^2 = 0.289$). We compared detection times for each icon across different condition pairs; the significant differences are summarized in Table 3. As shown in the third column of that table, detection times for all except IN++ in the highest-workload condition (visual+auditory distractor) were significantly greater than in the control condition. With respect to the comparisons between the two other condition pairs, the results were not as strong: only three icons required a longer detection time for the visual distractor condition compared to the control, and one icon for visual+auditory compared to the visual distractor.

An examination of specific icons helps to interpret these trends. The IN++ icon was designed to be the most intrusive, and it is therefore not surprising (and indeed desirable) that there was no difference in detection times across conditions. By contrast, the Waiting for Control icons were designed to be the least intrusive; we see that in two out of the three condition pairs there were differences in detection times, suggesting that this icon is in fact more amenable to 'backgrounding' than the others. In other words, we found that increasing workload impacted the detection of a nonintrusive icon, but did not impact that of an intrusive one. However, the results for the Change in Control icons were counter to our expectations: both were intended to be intrusive but, the detection time analysis revealed that they behaved more like the nonintrusive icons.

Identification Time: Identification time was measured from the appearance of the modal dialog box listing the 7 icons (whether the subject had detected the change or missed it) to the subject pressing the OK button. Table 4 shows the mean identification times for each condition. Figure 2 shows a graph of the mean identification times in each condition for each icon. The data revealed only a marginally significant main effect of condition ($F_{2,\ 22} = 3.175$, $p = 0.061$, partial $\eta 2 = 0.224$). This result, combined with the smaller difference in mean times for each condition, suggests that subjects' ability to context switch and identify the haptic icons is more robust to workload than their ability to detect a change in the icons.

There was also a significant main effect of icon, indicating that some icons took longer to identify than others ($F_{6,\ 66} = 20.993$, $p < 0.001$, partial $\eta 2 = 0.656$). Comparisons revealed that identification of the Change in Control icons, and in particular of CH-, took significantly longer than for the others, suggesting that subjects found CH- the most difficult to identify. This was confirmed by our data; CH- was mistaken for CH+ or IN+ four times more often than those icons were mistaken for CH-.

Condition	Mean	95% Confidence Interval	
		Lower Bound	Upper Bound
control	2548	2260	2836
v distractor	2698	2293	3103
v+a distractor	3022	2769	3276

Table 4: Mean Identification Times for each Condition (ms)

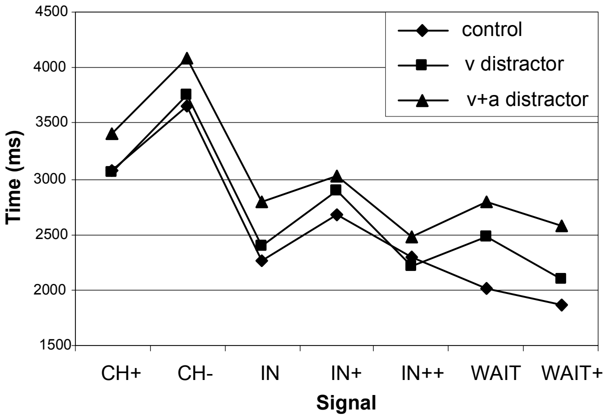

Figure 2: Mean Identification Times for each Condition

Condition	% Missed	95% Confidence Interval	
		Lower Bound	Upper Bound
control	1.7	0.2	3.1
v distractor	10.5	4.4	16.6
v+a distractor	18.8	9.8	27.8

Table 5: Percentage of Missed Transitions by Condition

We unexpectedly found a significant main effect of trial ($F_{4, 44}$ = 3.325, p = 0.018, partial $\eta 2$ = 0.232). Comparisons showed that identification times for the fifth trial in a set of icon transitions were 13% faster than for the first trial. This indicated that despite our practice transitions in each condition, subjects were still learning and improving as the study progressed.

Missed Transitions: Overall, condition did have a significant impact on the number of missed transitions ($F_{2,22}$ = 13.822, p < 0.001, partial η^2 = 0.557). Table 5 summarizes the percentage of missed transitions in each condition. Comparisons showed that the visual+auditory distractor condition had significantly more missed transitions (19%) than the visual distractor (11%) and the control (2%) conditions.

The results also revealed a significant interaction between condition and icon, indicating that transitions to some icons were missed more in some conditions than in other conditions ($F_{12, 132}$ = 3.402, p < 0.001, partial η^2 = 0.236). Comparisons revealed a difference between the control and the visual+auditory distractor condition for the IN icon. There were also differences between the control and the visual distractor conditions, as well as between the control and the visual+auditory distractor conditions for both of the Waiting for Control icons. These results show that transitions to the three subtlest icons were often overlooked as workload increased.

Correct Identification: Contrary to our expectations, condition did not significantly impact the rate of correct identification. On average, subjects identified icons correctly 95% of the time in each of the three conditions.

Distractor Task Performance: On average, subjects correctly placed 387 puzzle pieces in the two conditions with the visual distractor task. No significant difference was found between the number of pieces placed in those two conditions, which subjects finished in a combined total of approximately 25 minutes; one piece was placed approximately every 4 seconds. In the visual+auditory distractor condition, subjects also identified 27 out of the 30 keywords on average. Together, these results suggest that subjects were well engaged by the distractor tasks.

6. Discussion

Learning Time: On average, subjects only required 177 s to learn the abstract associations between haptic stimuli and their assigned labels. It is reasonable to expect that times would have been even lower if the subjects knew the metaphors underlying the icons. These results are quite encouraging, as they show that a modest-sized, well-designed set of haptic icons can be learned with relatively low effort - especially important if haptics are to be used in mainstream consumer applications.

Detection Time: The significant increase in icon detection times across conditions as overall workload increased was expected; of more interest is its small absolute size. The mean detection time in the visual+auditory distractor condition was double that in the control, but at 4.3 seconds, still acceptable for many purposes.

It was particularly important for deliberately intrusive icons (here, Change of Control and IN++) to be quickly detected and identified regardless of workload. This proved true for IN++; but CH+ was detected quickly and not identified quickly, and CH- was neither detected nor identified quickly. We attribute the difficulty to the modifications we made to the Change of Control icons immediately before the study which inadvertently introduced the side effect of making them less distinguishable. Had we re-piloted the study, this would have been discovered. Despite the mixed results, it appears possible to specify the robustness of individual icon detection time to workload.

Effect of the Auditory Distractor: It is of interest that both the mean detection time (all icons combined) and the percentage of missed transitions in the visual+auditory distractor condition were significantly greater than in the visual distractor condition. The auditory distractor was specifically designed to be easy so as not to unduly overload the user, but we saw that the addition of even an easy auditory task made a significant difference. The values were acceptable for our intended use, particularly since we expect them to be a conservative prediction of performance. However, if haptic icons are to be used in time- or safety-critical environments, the ability of users to detect them would have to be tested under levels of cognitive load expected in those environments.

Identification Time and Accuracy: Although identification times (including the time required to physically select the icon in the dialog) did increase marginally across conditions, the average time of 3.0 seconds in the visual+auditory distractor is shorter than we expected . Subjects maintained 95% accuracy in haptic icon identification, regardless of condition. We anticipate that with further training, identification times would decrease further. Further study is required to see whether the identification of haptic icons could become nearly automatic and be performed in parallel with other tasks.

Haptic Display: Our results provide a general indication of the performance that can be expected if other sets of haptic icons were created for the iFeel mouse. For a haptic device capable of producing more intense vibrations, these values again could be lower; and likewise, a richer design palette may increase distinctiveness and aid in identification. However, designers of haptic icons would have to balance the need to communicate information to users quickly with the risk of user annoyance through overuse of intrusive sensations.

Icon Set Design Process: We have prototyped a 3-step haptic icon design process: (1) metaphor-based initial set design; (2) optimization of perceptual spacing; (3) verification of meaning learnability and robustness to workload. Ultimate validation lies in a long-term evaluation of the icon set in a realistic context. Based on our experience to date, this version has given us promising results, demonstrated lessons in its use and exhibited some limitations.

Our metaphor- and family-based approach appeared successful based on Study 2 performance. It will be interesting to compare it with abstract and/or flat set construction, and as a function of set size.

The entire process is fundamentally iteration-based, and the iterations need to be carried out fully. Study 1 told us that its final set of CH icons could be distinguished; however, these were modified to eliminate suspected auditory cuing in pilots, and the resulting set used in Study 3 was apparently not sufficiently distinguishable.

Our MDS procedure is good for efficiently visualizing a set's perceptual spacing, but it doesn't support quantitative specification of separations. This points to the need for a new evaluative/analytic tool which incorporates the desired workload response to icon set members into the icon design process.

Study 2 assesses short-term learnability, but it will be more difficult to efficiently determine learning both over time and in presence of larger or multiple sets. Our simple workload tasks do appear effective at exercising the icon set we designed, and seem to be a worthwhile preliminary to the effort and expense of deployment in a real application context.

7. Conclusions and Future Work

In this paper, we have presented results from two studies that prototype a process for designing and evaluating haptic icons for use in a cognitively intensive task; that is, for situations where it is important to understand the effect workload will have on users’ ability to detect and identify changes in haptic icons. The icon set derived here will support peripheral awareness in distributed collaboration in the form of a new turn-taking protocol; however, the process and our findings have broader implications as haptics are introduced into consumer and time- and safety-critical environments.

We found that the time required to detect a change in haptic icons approximately doubled when subjects were asked to carry out visual and auditory distractor tasks; but had a ‘worst case’ mean of 4.3 s. Stimuli designed to be subtle were affected more than those designed to be intrusive. The low identification time (2.5 s) and 95% accuracy in identifying icons in the absence of workload was encouraging – subjects achieved this after spending less than three minutes on average learning the 7 haptic stimuli and their abstract labels. Further study will be required to evaluate the effect of workload on identification time.

Possibilities for further research include examining the ability of users to remember the associations over time, and ascertaining optimal learning practices. It will be of value to compare our contextually-driven study results with a systematic examination of haptic icon recognition performance under workload, comparing all possible transitions between a set of stimuli that differ along only a single parameter. Workload itself can be controlled at a finer resolution: is haptic performance unaffected up to a certain level of workload? What is that level? It also remains to be seen whether the results shown here generalize to other kinds of haptic devices, such as actuated knobs.

Finally, the process we used is promising, and also illustrates the need for tools that support integration of desired workload performance into the design process.

8. References

[1] BMW AG. iDrive. Accessed: September 14, 2004. http://www.bmw.com/generic/com/en/fascination/technology/technologyguide/index.html.

[2] WebEx Communications Inc. WebEx. Accessed: April 6, 2004. http://www.webex.com/.

[3] Microsoft Corp. NetMeeting. Accessed: Apr. 6, 2004. http://www.microsoft.com/windows/netmeeting/.

[4] Basdogan, C., Ho, C., Srinivasan, M., and Slater, M. (2000). An experimental study on the role of touch in shared virtual environments. ACM Transactions on Computer-Human Interaction (TOCHI), 7 (4), 443-460.

[5] Blattner, M., Sumikawa, D., and Greenberg, R. (1989). Earcons and icons: Their structure and common design principles. Human-Computer Interaction, 4 (1), 11-44.

[6] Boyd, J. (1993). Floor control policies in multi-user applications. In INTERACT '93 and CHI '93 Conference Companion (pp. 107-108).

[7] Boyle, E., Anderson, A., and Newlands, A. (1994). The Effects of Visibility on Dialogue and Performance in a Cooperative Problem Solving Task. Language & Speech, 37 (1), 1-20.

[8] Brewster, S., and Brown, L. (2004). Non-visual information display using tactons. In Extended Abstracts of the CHI 2004 (pp. 787-788).

[9] Brewster, S., Wright, P., and Edwards, A. (1993). An evaluation of earcons for use in auditory human-computer interfaces. In Proceedings CHI '93 (pp. 222-227).

[10] Duncan, S., and Niederehe, G. (1974). On Signalling That It's Your Turn to Speak. Journal of Experimental Social Psychology, 10, 234-247.

[11] Enriquez, M., and MacLean, K. (2004). Impact of Haptic Warning Signal Reliability in a Time-and-Safety-Critical Task. In 12th Annual Symposium on Haptic Interfaces for Virtual Environments and Teleoperator Systems, IEEE-VR2004.

[12] Gaver, W. (1993). Synthesizing auditory icons. In Proceedings of CHI '93 (pp. 228-235).

[13] Greenberg, S. (1990). Sharing views and interactions with single-user applications. In ACM/IEEE Conference on Office Information Systems (pp. 227-237).

[14] Inkpen, K., McGrenere, J., Booth, K., and Klawe, M (1997). Turn-Taking Protocols for Mouse-Driven Collaborative Environments. In Graphics Interface '97 (pp. 138-145).

[15] Klatzky, R., and Lederman, S. (2003). Touch. In A.F. Healy, and Proctor, R.W. ed. Handbook of Psychology, (pp. 147-176). New York: John Wiley & Sons.

[16] Klatzky, R., Lederman, S., Hamilton, C., Grindley, M., and Swendsen, R. (2003). Feeling textures through a probe: Effects of probe and surface geometry and exploratory factors. Perception & Psychophysics, 65 (4), 613-631.

[17] MacLean, K., and Enriquez, M. (2003). Perceptual Design of Haptic Icons. In Proceedings of Eurohaptics.

[18] McKinlay, A., Arnott, J., Procter, R., Masting, O., and Woodburn, R. (1993). A Study of Turn-Taking in a Computer-Supported Group Task. In People and Computers VIII, Proceedings of the HCI '93 Conference (pp. 383-394).

[19] Miller, G. (1956). The magical number seven, plus or minus two: Some limits on our capability for processing information. Psychological Review, 63, 81-97.

[20] Myers, B., Chuang, Y., Tjandra, M., Chen, M., and Lee, C. Floor control in a Highly Collaborative Co-Located Task. Accessed: April 6, 2004. http://www-2.cs.cmu.edu/~pebbles/papers/pebblesfloorcontrol.pdf.

[21] Oakley, I., Brewster, S., and Gray, P. (2001). Can You Feel the Force? An investigation of haptic collaboration in shared editors. In Proceedings of Eurohaptics.

[22] Sallnas, E., Rassmus-Grohn, K., and Sjostrom, C. (2000). Supporting presence in collaborative environments by haptic force feedback. ACM Transactions on Computer-Human Interaction (TOCHI), 7 (4), 461-476.

[23] Steele, M., and Gillespie, B. (2001). Shared control between human and machine. In Human Factors and Ergonomics Society 45th Annual Meeting.

[24] Tan, H., Durlach, N., Reed, C., and Rabinowitz, W (1999). Information transmission with a multifinger tactual display. Perception & Psychophysics, 61 (6), 993-1008.

[25] van Erp, J. (2002). Guidelines for the Use of Vibro-Tactile Displays in Human Computer Interaction. In Proceedings of EuroHaptics.

[26] Wagner, C.R., Stylopoulos, N., and Howe, R. D (2002). The role of force feedback in surgery: analysis of blunt dissection. In 10th Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems.