The Haptic Crayola Effect: Exploring the Role of Naming in Learning Haptic Stimuli

Inwook Hwang,[1] Karon E. MacLean,[2] Matthew Brehmer,[2] Jeff Hendy,[2] Andreas Sotirakopoulos,[2] Seungmoon Choi[1]*

[1] Pohang University of Science and Technology (POSTECH) [2] University of British Columbia

Abstract

A haptic icon is a short physical stimulus attached to a simple meaning, which provides information and feedback to a user. To scale the utility demonstrated for small icon sets to larger ones, we need efficient strategies to help users learn subtle distinctions among stimuli, in a modality for which they may not hold detailed descriptive percepts. This paper investigates the effect of *naming* haptic stimuli – i.e. explicitly creating a linguistic marker – on the accuracy with which users are able to identify, distinguish, and recall stimuli.

We conducted a between-subjects experiment using 60 participants equally divided among three naming conditions: no names, pre-selected non-descriptive names, and self-selected names. The experiment examined the impact of naming strategy on the ability of participants to identify stimuli in a nonverbal matching test, and on remembering stimulus names. For this challenging task and the degree of learning afforded, naming did not significantly impact accuracy of matching stimuli to meanings for all participants. However, more than twice of many of those allowed to choose names reported the ability to remember and distinguish stimuli than those required to use non-descriptive names, and many participants felt that the names were useful. Of middle-performing participants, the self-selected names group performed significantly better than the non-descriptive names group, and appeared to progress more quickly in learning. We summarize evidence for a trend that might widen with refined naming strategies and more extensive learning.

Index Terms: H.5.2 [Information Interfaces and Presentation]: User Interfaces—Haptic I/O

1 Introduction

Touch is a promising channel for conveying information with minimal attentional load, in environments where vision, audition and other mental resources are heavily utilized. Haptic icons (HIs: short tactile or force stimuli attached to simple meanings) provide information and feedback on event notification, system and application state, or other contextual user cues. Haptic icons can peripheralize information that would otherwise be distracting, as demonstrated by high recognition rates for peripheral identification under varying amounts of workload [4, 19]. HIs and other similar feedback have also shown benefits in mobile contexts, e.g. through assistance during text entry tasks [2], resulting in subjective user reports of workload reduction and less need for other feedback.

To use any kind of icon, users must be able to efficiently map each stimulus to its meaning, and remember this association. This requires easy-to-learn strategies to help users differentiate, identify, and remember the stimuli themselves. However, most people have less linguistic specificity for nuanced touch sensations than for visual or auditory experiences (each of which benefits, for example, from vocabularies derived from music and visual arts).

*E-mail: inux@postech.ac.kr, maclean@cs.ubc.ca, brehmer@cs.ubc.ca, jchendy@cs.ubc.ca, andreass@ece.ubc.ca, choism@postech.ac.kr

IEEE World Haptics Conference 2011
21-24 June, Istanbul, Turkey
978-1-4577-0298-3/11/$26.00 ©2011 IEEE

We thus hypothesized that descriptive naming of the stimuli would help users differentiate, identify, and remember the resulting icons. This notion was inspired by the memorable names that the Crayola Company has given to hundreds of crayon colors [5], and Crayola users' apparent ability to differentiate and remember the colors without practice, often by reference to personal experience (e.g. "Sunset Orange" versus "Atomic Tangerine"). We call our hypothesized impact of stimuli naming on signal learnability the Haptic Crayola Effect. The aim of this paper is to explore the Haptic Crayola Effect with three naming conditions: no stimulus names, preselected non-descriptive stimulus names, and freely chosen stimulus names. Our hypotheses were:

H1: Identification accuracy will be higher in the self-selected names condition than the other two conditions.

H2: Ability to remember names will be higher in the self-selected names condition than the non-descriptive names condition.

H3: Participants who report being able to remember the names better will have higher identification accuracy.

In our experiment, 60 participants divided equally among the three naming conditions first completed a series of exercises to gain familiarity with the stimuli and, if applicable, their names, with comparable exposure duration. They completed a series of tasks in which they were asked to replicate a series of target stimuli, and finally provided subjective feedback on the conditions through a questionnaire. The rest of this paper introduces relevant literature and then describes and discusses the experiment and its results.

2 Related Work

In designing haptic icons, to begin with stimuli obviously must be discernible. Associations between stimulus and meaning must be well understood and easily learned. Blattner et al. summarizes amodal techniques for achieving stimulus saliency and learnability [1]; MacLean does so specifically for haptic icons [13].

Icons can range on a continuous scale of abstraction from representational to abstract. The former is a caricature of a real-world stimulus, while the latter is (possibly arbitrary) semantic link between meaning and stimulus. The approach used has implications for scalability and learnability. Metaphorical approaches driven by application-relevant meanings have been effective for small stimulus sets [2, 4]. Although metaphoric stimulus-meaning associations may seem initially intuitive, they do not incorporate a systematic basis for determining relative stimulus salience or discriminability, instead relying on designer creativity in generating good metaphors – a difficult task for abstract concepts and larger icon sets [13]. The strategy further depends on the discriminability and saliency of stimulus sets in combination [14], and is subject to pre-existing associations the user may have with a particular metaphor, which may impose conflicts with a new meaning [21].

2.1 Stimulus Differentiability

For abstract stimulus sets, there has been considerable examination of stimulus design parameters such as frequency, amplitude, waveform, duration and rhythm. Brown and Brewster sought parameters that were both easily distinguishable and memorable [3], and proposed that rhythm and roughness were most appropriate; Ternes and MacLean found rhythm to dominate other factors, with evenness emerging as crucial in the perceptual scaling analysis [21]. The process of creating large abstract stimulus sets with maximum differentiability through perceptual optimization [21], i.e. maximizing the perceived "distance" between each pair of stimuli, has been mapped and maximized. Explained in [12, 14], this technique has been used to validate stimulus sets prior to assigning meanings, for parameters varying on frequency, force amplitude, wave shape [14], rhythm [21], and melodic variants [19, 20, 23].

2.2 Stimulus-Meaning Learnability

Studying the ability of users to learn arbitrary stimulus-meaning associations for 9 haptic icons felt through a force-feedback knob, Enriquez et al. found that participants recalled up to 81% of learned associations during a later test [7]. In another experiment, participants learned a set of abstract icons over a two-week period [6]. No significant difference in identification performance was noted in a within-subject comparison between arbitrarily assigned and participant-defined stimulus-meaning pairs. Some participants preferred making their own associations and others reported that arbitrary pairs seemed better matched than their own, supporting the notion that whether or not the designer has *intended* a semantic stimulus-meaning association, users will likely develop their own mnemonics. It was also noted that participants generally doubted their ability to recall stimulus-meaning pairs after a two-week period, despite their demonstrated ability to do so.

Swerdfeger et al. conducted two longitudinal studies wherein notifying icons (piezo screen vibrations displayed through a stylus) interrupted a foreground game task on a mobile device [19]. The first study utilized stimuli described in [21] which varied primarily varied in rhythm. Pairing similar and non-intuitive meanings with stimuli that are, in contrast, perceptually very distinct did not adversely impact learning relative to a policy of associating perceptually similar stimuli with similar meanings, contrary to predictions in [1]. Participants generally reported effort focused on distinguishing similar stimuli, an irrelevance of the stimulus-meaning association in learning the icons, and meanings that acted simply as a *name* for each stimulus. Many participants developed unique adhoc mnemonics for distinguishing and learning the stimuli. A second study based on melodic stimuli tested whether a melodic component would enhance stimulus expressiveness [20]. When many users struggled to perceive and learn melodic variation, researchers theorized that the evocation of a possible emotional response to a stimulus may have hindered the ability to form new abstract semantic associations, and concluded that purely rhythmic stimuli currently support more predictable learning.

2.3 Stimulus Naming and Learnability

Within these studies are indications that participants may have used "internal" descriptors for the physical stimuli to help them remember abstract icon meanings. This has been explicitly observed: Brown et al. suggested a need to label the perceptual parameters of rhythm as a means of learning rhythmic haptic icons [3], independent of the meaning association. In a psycho-acoustical experiment in which musician and non-musician participants were asked to rate simple rhythms from a variety of musical styles along 92 adjective scales, Gabrielsson found that people primarily use three classes of words to describe and label rhythms: (1) structurally descriptive terms for the rhythms, describing note accentuation or pattern complexity; (2) terms that are characteristic of perceptual or movement actions; and (3) emotional terms such as *vital*, *dull*, *excited*, or *calm* [8]. Musical-linguistic tags for timbre have also been precisely specified among musicians and acousticians: HLF von Helmholtz's timbre descriptions based on the presence of specific harmonics in the audio signal are still used today [9]. For musicians, subtle and consistent descriptions of rhythm and timbre are imperative for communication and training.

This paper explores whether such terms are of similar benefit to non-musicians and in other modalities, for the purpose of learning a similar *kind* of distinction.

Descriptive words are also used for taste stimuli; however, a consistent language of terms beyond combinations of the taste primaries *salty*, *sweet*, *bitter*, *sour*, and *umami* does not exist other than idiosyncratic and anecdotal descriptions [10]. An exception is among taste panelists, whose training includes a greater commonality in verbal taste concepts.

While a consistent terminology does not exist for odors among non-experts, odor memory and odor naming appear to be strongly related, with consistency of odor label use the main predictor of odor memory [11]. While personally generated labels for odors were somewhat arbitrary, odor recognition may evoke both perceptual and semantic memories. Contrary to these results, experiments comparing odor recognition between wine experts and novices revealed superior recognition by wine experts, but failed to show better performance of experts than novices on related verbal memory tasks, indicating that expert's superior recognition was not reliant on linguistic and semantic memory for wine odor terms [16]. One possibility is that trained wine experts may rely solely on their sensory-perceptual memory, ignoring inhibitory linguistic-semantic memory. This would be an example of *verbal overshadowing*, a memory illusion occurring when individuals must name complex non-verbal stimuli which are difficult to capture in words, including wine odors [15] or faces [18]. In essence, one's perceptual memory is overshadowed by attempts to verbally communicate that memory following perceptual encoding. Novices are particularly susceptible, because low verbal expertise impinges on perceptual expertise.

Supposing that naming a rhythmic stimulus can promote long-term stimulus-meaning learning and is *not* susceptible to verbal overshadowing, it becomes unclear which type of name to use; rhythms seem to be linguistically overloaded with structural, perceptual-anthropomorphic, and emotional words. We anticipate that naming and remembering haptic or auditory stimuli based on their physical characteristics may be analogous to naming and remembering colors. The process by which children learn color words involves several mental mappings [17]; the non-obvious links between words and stimuli are learned if they are used consistently, and eventually the perceived property of a stimulus is abstracted to its color word. Unfortunately, there are very few widely-agreed upon color names beyond the basic color terms, and our ability for learning additional colors and their names is poor [24]. In controlled conditions, a trained colorist can distinguish between a million colors [22] and can remember 1,000 color names, but many such names are idiosyncratic or imprecisely defined [24]. The crayon manufacturer Crayola uses a wide array of names, many of which evoke a colorful material familiar to the user [5].

How effective in comparison are individual users at naming colors themselves, then later remembering these name-stimulus associations? This question can also be asked of haptic stimuli. To our knowledge, in no previous experimental research did participants give haptic stimuli their own descriptive names to aid in discrimination and learning. Exploring the effect of this learning adjunct is the basis of our experiment.

3 Experiment

To explore the hypothesized Haptic Crayola Effect, we tested participants' ability to distinguish and recall touch stimuli under three conditions: the stimuli had no names, predetermined non-descriptive names, or participant-selected names.

3.1 Tasks

The experimental had two phases: learning, in which participants gained familiarity with the stimuli and their names if applicable, and testing, where they recalled stimuli sequences in a nonverbal

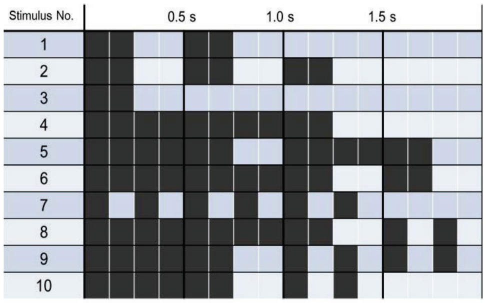

Figure 1: The ten rhythmic stimuli used in this study. Dark cells indicate "on" period of 250Hz vibration.

Figure 2: Experimental setup. The inset shows the vibrotactile actuator used in this experiment.

matching test. The three conditions differed only in the learning phase. Ten vibratory stimuli, each 2 seconds in length (0.5 second sequence repeated four times), were chosen randomly from the 84-item set of rhythmic stimuli designed by [21] (Fig. 1). The same stimuli were used for all participants. This source is the only large set we are aware of that has been perceptually optimized; and rhythmic variation is known as an expressive and easily manipulated parameter. The stimuli were displayed as 250 Hz sinusoidal vibrations using a Tactaid VBW32 voice coil installed on a layer of vibration-isolation foam within a plastic box on which participants' left index finger rested (Fig. 2). Control code (Microsoft Visual C++ and OpenGL) ran on an IBM X41 laptop with a 12.1" display with optical mouse, and drove the tactor through the laptop's audio port with amplitude maximized.

3.1.1 Learning Phase

All participants experienced the stimulus set in preparation for test-phase identification; those in the two "name" conditions also learned and/or assigned stimulus names. Regardless of condition, participants had equal exposure to stimuli. for an initial understanding of the set's range, all participants first felt each stimulus in sequence twice, advancing by clicking on a button labeled "NEXT". All participants then completed a sorting task [12][14], successively arranging stimuli into a different number of bins (here, six sorts ranging from two to seven bins). During a sort, each of the ten stimuli was shown as a blue circle on the screen and could be played freely via a mouse click, then dragged and placed with similar stimuli into a bin at the bottom of the screen.

No names condition: As described above.

Non-descriptive names condition: Beginning with the 2nd stimulus display cycle, a pre-determined, randomly chosen name appeared on the screen whenever its stimulus was played. Names were nouns such as *Tiger*, *Leopard*, *Cheetah*, *Cougar*, *Hockey*, *Skiing*, *Skating*, *Spoon*, and *Knife*.

Self-selected names condition: Participants chose their own stimulus names during the 2nd sequential exposure, being prompted to type in a name. Participants were offered paper sheets listing potential stimulus names (words describing emotions, physical properties of objects, and rhythms). They were told that the words on the sheets were suggestions, and they could use any desired names with the restriction that no two stimuli could have the same name. During sorting, assigned names were then graphically displayed as for the non-descriptive names condition.

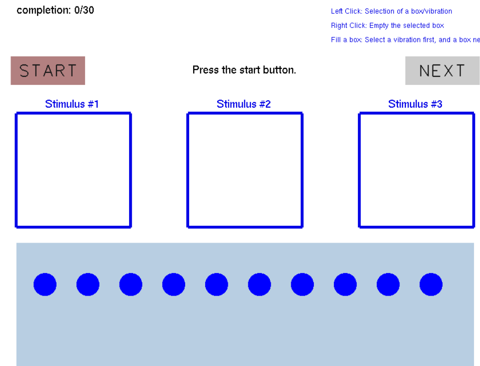

Figure 3: Screenshot of the test phase task.

The learning-phase tasks were a compromise, in that they are not optimized for learning either the stimuli or their names. MacLean [13] recommended the use of practical, regular-use learning scenarios, the inclusion of reinforcement when possible, and avoidance of long, stressful experimental sessions. Enriquez et al. suggested that it may be beneficial to allow the user to return to an exploratory learning phase if performance in reinforced learning is low [7]. However, our priority here was to attain equal exposure time across conditions. This precluded tasks that related specifically to the names, nor could we test the learning of the names, as this would not have an equivalent in the no-name group.

3.1.2 Testing Phase

The testing phase, identical for all conditions, was designed to determine participants' ability to recognize and recall sequences of three target stimuli; Fig. 3 shows a screenshot of the interface. A trial began when the participant clicked the "START" button. Each of three 2-second target stimuli was then played once separated by 1.5-second delays, while the corresponding graphical box highlighted. Participants then searched for the three targets through the full 10-item stimulus set, represented as clickable circles at the bottom of the screen in a spatial order chosen randomly for each trial. Participants could not see the name of each stimulus and indicated each choice by dragging it to the corresponding target box. Participants could feel the full set without limit, but the target stimulus was displayed once. Each participant completed 30 trials with a 1-minute break after every 10 trials.

3.2 Participants

60 participants (32 female; aged from 19-44 with median 25; 20 randomly assigned to each condition) were recruited on the UBC campus and paid $10 for their time. 53 of them were undergraduate or graduate students of UBC. To motivate performance, an additional $5 prize was promised and paid to the top 25% achievers.

3.3 Design, Measures and Procedure

The experiment employed a three (conditions) $\times$ 30 (trials) mixed-factors design. Condition was a between-subjects factor, and trial was within-subjects. Each participant completed a single session of approximately one hour. After prebriefing, participants completed the learning and testing phases, then a questionnaire asking for basic background information and about their experience based on condition. The latter included ability to remember stimulus names,

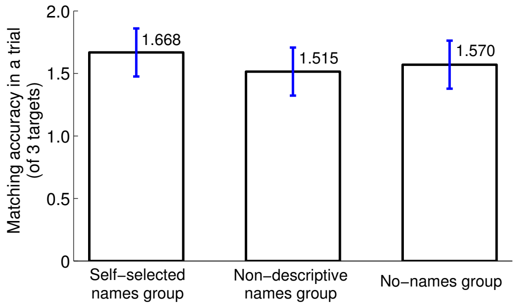

Figure 4: Matching accuracy (N=60). Error bars show standard error.

whether they felt that the names were helpful, and use of memory "tricks" in addition to names.

Accuracy was measured as the number of targets correctly identified out of the three for each trial. Ability to remember stimulus names was self-reported. Completion time was calculated from the time that the final target stimulus was presented until the "NEXT" button was clicked in a trial.

4 Results and Analysis

All 60 participants reported at least an intermediate level of computer proficiency. 92% reported using a mobile device with vibratory notification signals; of these, 29/55 were aware of or made use of distinct vibratory signals for different notifications. 30% of participants were familiar with vibratory signals or force feedback used in video game controllers.

4.1 Matching Task Accuracy

On average over the 30 trials, the self-selected names group correctly matched 1.67, the non-descriptive names 1.52, and the no-names group 1.57 of the three target stimuli (Fig. 4). This corresponds to 56%, 51% and 52% accuracy respectively, where chance selection (three tries for different targets with no replacement) would achieve $\frac{1}{10\times9\times8}$ or 0.13% probability of matching all 3 targets correctly (100% accuracy), and $(\frac{1\times7\times7}{10\times9\times8}+\frac{1}{10\times9})\times 3$ or 23.75% chance of matching 1 of 3 (33% accuracy).

A three (conditions) $\times$ 30 (trials, repeated measures) mixed ANOVA failed to show a significant difference for condition (between subjects $p = 0.524$, $F(2,57) = 0.654$), but did show a significant difference for trial (within subjects $p < 0.001$, $F(1,57) = 815.993$, partial $\eta^2 = 0.935$); post-hoc comparisons between trial mean accuracy scores across the three conditions using the Bonferroni adjustment indicated many significant differences between trial pairs, but their general trend as a function of trial order was not apparent. Overall mean accuracy scores for all participants from trials #11 and #22 (immediately or shortly after 1-minute breaks) were typically higher than the mean scores of most other trials.

4.2 Matching Task Completion Time

Averaged over 30 trials, the self-selected, non-descriptive and no-names groups completed matching tasks in 33.1, 32.0 and 35.0s respectively (average std. error = 2.380). A three (conditions) $\times$ 30 (trials, repeated measures) mixed ANOVA did not show significance for condition (between subjects $p = 0.667$, $F(2,57) = 0.408$), but did show a significant difference for trial (within subjects $p < 0.001$, $F(1,57) = 589.407$, partial $\eta^2 = 0.912$); post-hoc comparisons between mean completion times across conditions using the Bonferroni adjustment indicated many significant differences between trial pairs, with the first trial taking longest to complete.

4.3 Subjective Results

The post-experiment questionnaire queried self-selected and non-descriptive names group members about their ability to distinguish and remember stimuli throughout the experiment, on a 5-point Likert scale. For example, "By the end of the experiment, I remembered the icon names." was given as a statement and participants rated it from *strongly agree* to *strongly disagree*. All participants were asked to describe any tricks or mnemonics used for either stimulus discrimination or learning, aside from the names.

Self-selected names group: Many participants reported use of vibration spacing or pause length as additional memory cues. Nine participants agreed at the statement: "After the learning phase, I was able to associate the icons with their names.", four undecided, and the rest seven disagreed. The names were reported as effective ($N = 10$), neutral ($N = 1$) or ineffective ($N = 9$) in helping to discriminate stimuli during testing phase. Individuals claimed that the names they chose described the stimuli accurately (7), undecided (8) or inaccurately (5). Participants in this group were confident (8), neutral (6), or unconfident (6) in remembering the names by the end of the experiment. The majority (12) agreed that more descriptive names would have helped in discriminating and remembering the stimuli in the testing phase.

Non-descriptive names group: Participants mentioned counting and distinguishing between short and long pulses. Other tactics included tapping, associating sounds, visualizing vibrations spatially as "peaks on a graph", or attempting to map them to Morse code. Nine participants agreed at the statement: "After the learning phase, I was able to associate the icons with their names.", five undecided, and the rest six disagreed. Individuals claimed that the pre-selected names described the stimuli accurately (2), undecided (4) or inaccurately (14). Participants in this group were confident (3), neutral (12), or unconfident (5) in remembering the names by the end of the experiment. The majority (11) of participants did not feel that the non-descriptive names helped them distinguish or remember the stimuli during testing. They believed more descriptive names would have improved their performance (12), and that they would have selected better names themselves (13) if given the opportunity.

No-names group: Participants reported a similar variety of tricks or mnemonics for distinguishing and learning stimuli throughout the experiment; many paid attention to the number and spacing of long and short pulses or pauses. Others attempted to visualize the sequence as the lines and dots of Morse code, moved their heads along with the vibration, hummed along with the sequence, or counted the number of pulses on the fingers of their free hand. The majority of these participants (14) agreed that giving the stimuli names would have assisted them in the process of distinguishing and remembering the stimuli.

4.4 Performance of Self-Reported "Rememberers"

We compared matching-task performance of the 11 participants who reported name remembrance with the 18 who reported forgetting with a non-parametric Mann-Whitney test. We compared overall accuracy, or total number of correct matches over 30 trials (90 matching trials), for the "rememberers" ($N = 11$, $M = 52.55$, $SD = 11.46$, mean rank $= 17.45$) with the "forgetters" ($N = 18, M = 46.33, SD = 15.33$, mean rank $= 13.50$), but found no significant difference between groups ($p = 0.225$). We also compared average total task completion time over 30 trials for rememberers ($M = 1025.46s$, $SD = 325.59s$, mean rank $= 16.36$) with forgetters ($M = 969.12s$, $SD = 265.45s$, mean rank $= 14.17$), but found no significant difference between groups ($p = 0.500$).

4.5 Self-Selected Stimulus Names

We analyzed the 200 names given to the ten stimuli by each of the 20 participants in the self-selected names group. Over half of these

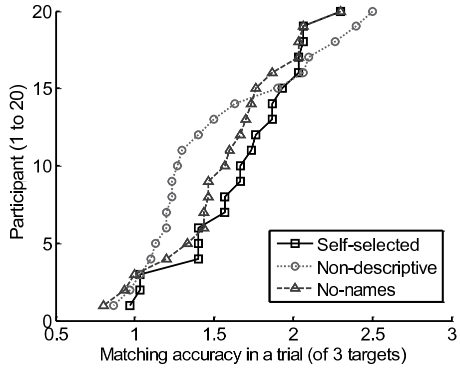

Figure 5: Cumulative distribution of average matching accuracies. Each point represents one participant's average performance.

names were purely rhythmic descriptions, derived from the pattern pulses; e.g. "one short, three quick." The next largest group of terms consisted of additional rhythm descriptors, often used in combination with the number of pulses, such as *scattered*, *elongated*, *punctuated*, and *staccato*. 12 participants in this group used solely rhythmic descriptions, including one who named most of stimuli using a literal coding strategy of small and lower case 'B's, presumably to indicate long and short tones, e.g. *bBbb*. Six participants employed a combination of rhythmic and other words, with rhythm-based names forming the majority. Other words included emotional adjectives, such as *angry*, *impatient*, or *talkative*, as well as actions or objects in the real world, such as *hop & skip*, *warning*, *mosquito*, *inkjet*, and *doorbell*. One participant used solely emotional words, and one used solely real objects as names.

4.6 Impact of Names on Slow, Middle and Fast Learners

Participants vary considerably in facility at learning icons (e.g. [19]). To explore differential impact of naming over the range of learning quickness, we examined the cumulative distribution of members of the three condition groups in participant's average matching accuracy. Fig. 5 shows curves for the three conditions, which separated in the middle of the performance range and met together in both ends (i.e. for participants with low or high matching performance). To investigate this in detail, we divided each condition group three ways based on percentile of matching accuracy within the group, using the top 25%, middle 50% and bottom 25%. One-way ANOVA tests for these three groups found a significant difference for the middle 50% group ($p = 0.002$, $F(2,27) = 7.944$). Tukey's HSD multiple comparison test for the middle-performing participants (most separated) showed that the self-selected names group had a significantly larger mean accuracy than the non-descriptive names group.

Thus, while neither the best nor worst learners were influenced by naming in early learning, the middle-performing group shows significant evidence of naming strategy on matching performance.

4.7 Patterns in Early Learning

We looked for evidence of early learning in the confusion matrices for each participant, with attention to both the "perfectly correct" frequencies observed along the main diagonal, and the regularity of errors. A common error in early, unreinforced learning is to mislearn a relationship and then repeat it; and overly similar stimuli will be consistently confused by multiple participants.

Error behaviors, and presumably identification difficulties, were common across the naming conditions. Correlation coefficients between average confusion matrices were high (0.974 to 0.985), confirming a lack of structural difference, with a similar result for middle-performance participants' data alone. Correlations between average confusion matrices and the identity matrix (perfect matching) are an indicator of general progress. When all participants' data were used, these coefficients were similar but ordered consistently with earlier results: 0.903, 0.887 and 0.898 for self-chosen, non-descriptive and no names. For the middle-performance data alone, correlations spread to 0.870, 0.795 and 0.860, suggesting a potential advantage of self-selected names and no names. Overall, this analysis suggests that the naming condition did not grossly alter learning patterns, but may have impacted learning rate for middle-performing subjects.

4.8 Results Summary

H1: Identification accuracy will be higher in the self-selected names condition than the other two conditions. *Not supported (See Section 4.1.).*

H2: Ability to remember names will be higher in the self-selected vs the non-descriptive names condition. *Supported by self-report responses (Section 4.3): 40% vs. 15%; and by significantly improved matching performance for middle learners (Section 4.6, and 4.7).*

H3: Participants who report being able to remember the names better will have higher identification accuracy. *Not supported (Section 4.4).*

5 Discussion

For the experiment tasks used here, we did not find definitive evidence for the hypothesized immediate and apparent difference between naming conditions, in either performance or subjective results. However, we did observe trends which suggest further exploration is needed. Possible interpretations of the statistical results include inadequate exposure (i.e. too early on the learning curve of a difficult task to see separation in performance); insufficiently meaningful names; or a true absence of effect.

5.1 Task Difficulty and Degree of Learning Achieved

One certainty based on observation and participant feedback is that participants found the experiment tasks, designed to generate a spread of results, very challenging despite piloting to target their level. Further, it was an accepted constraint of the experimental approach that the method of learning stimulus names used was not optimal. Tasks that are too difficult can distort early learning.

That learning opportunity was insufficient is supported by the combination of participants' self-reported lack of learning with other evidence that good learning is possible. Swerdfeger et al. gradually introduced a superset of the same stimuli; after a month's time, participants achieved identification accuracy averaging 76% for sets of 14 or 21 icons [19]. Subjectively, 27.5% of our participants from the two "name" conditions agreed that they remembered names. While confidence can lag performance (notably, [6]) this does suggest an early point on the learning curve for this task.

However, as hypothesized, more than twice as many participants (8 vs 3 /20) reported remembering stimulus names when they chose them rather than had them imposed. Furthermore, the middle range of learners benefited significantly in performance from self-selected relative to non-descriptive names at this early stage.

5.2 Implications of Stimulus Name Choices

Self-reports indicate a belief that adequately descriptive names should help: in the no-names condition, 70% of participants agreed with the statement, "Giving the stimulus names would have helped me remember the stimuli and tell them apart." Furthermore, 60% of participants in both name conditions agreed that "More descriptive names would have improved performance in the matching task." This suggests that names in both name conditions were not ideal.

The ten vibratory stimuli in this study differed only by rhythm because this is known to be a relatively expressive parameter. Most chosen names were literal descriptions of the rhythm, while names evoking emotion or physical characteristics were rare. We speculate that this literality hampered performance.

We have two explanations for participants' choice of rhythmic descriptions rather than abstract names. First, as noted in Section 2, Gabrielsson suggested that people usually use structural rhythmic properties in adjective rating of auditory stimuli [8]. It is reasonable to expect similar results here.

Secondly, the degree of verbal abstraction supported was low. In the learning phase, participants were asked to name each stimulus on only their second exposure to it. The verbal abstraction of vibratory sensation may have been at a primitive level, resulting in literal descriptions rather than emotional or cognitive metaphors to represent a vibratory stimulus.

Our learning phase was also relatively short for the sufficient memorization of each vibration stimulus and its name. Many participants remained unfamiliar with various vibration patterns and their discrimination. Future work in this area should include a longitudinal study with enough time and experience to fully memorize the stimuli and their names.

5.3 Other Factors

Verbal Overshadowing: Our results may be subject to verbal overshadowing [18]: when people verbally describe a complex sensation, identification performance on the stimulus can be lower than when the verbal step is omitted. It is not clear whether these stimuli were experienced as complex. If present, this effect could have contributed to weakened performance in the two named conditions.

Hierarchy in Naming: Many collections of object names have a clear hierarchical structure, as do many sets of stimuli themselves. Hierarchical or familial grouping appears to be helpful in mentally organizing haptic stimulus properties and name of a haptic stimulus, as evidenced in past work [1, 3, 4, 7].

6 Conclusions

Our results do not show a definitive impact of stimulus verbalization on performance, in a challenging single-session matching task requiring differentiation and memory, whether names are non-descriptive and imposed, or user-generated; neither do participants who claim to able to remember the names perform more accurately.

However, we found subjective divergences and evidence of inadequate learning which point to a larger story. More participants using self-selected names reported remembering stimulus names than those using non-descriptive names, with increased accuracy performance for the middle-range performers. Meanwhile, task difficulty, type of self-chosen names, the inefficient learning task used for experiment control, and a possible role of verbal overshadowing may have impeded learning. Thus, we cannot tell what *performance* spread might eventually emerge. For example, given evidence that confidence is a poor indication of recognition performance potential [6, 19], it is possible that performance gaps from using self-selected or non-descriptive names will never emerge, even as a confidence gap persists. Likewise, verbalization use or absence could play out in a variety of ways with more learning and larger sets.

Altogether, our observations are grounds to suspect that (a) more extensive learning and the challenge of larger sets could potentially generate larger differences between name and no-name conditions; and (b) there is room for better chosen self-selected names to widen the early confidence gap between meaningful and random name associations. Conversely, this could disappear with increased familiarity with either set of names, and it does not necessarily imply that a similar trend in actual performance would follow.

Future work should explore the effects of different types of naming including hierarchical and non-verbal names, with a better optimized and longer learning phase involving a larger set of stimuli, to move participants higher up the learning curve.

Acknowledgements

This work was supported in parts by an NRL program R0A-2008-000-20087-0 from NRF and an ITRC program NIPA-2010-C1090-1011-0008 from NIPA, funded by the Korean government; and by the Canadian NSERC Discovery Grant program. Experiments were conducted under UBC Ethics Approval H01-80470.

References

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

[2] S. Brewster, F. Chohan, and L. Brown. Tactile feedback for mobile interactions. In *Proc. of SIGCHI*, pages 159–162. ACM, 2007.

[3] L. Brown, S. Brewster, and H. Purchase. A First Investigation into the Effectiveness of Tactons. In *Proc. of World Haptics*, pages 167–176. IEEE Comp. Soc., 2005.

[4] A. Chan, K. MacLean, and J. McGrenere. Learning and Identifying Haptic Icons under Workload. In *Proc. of World Haptics*, pages 432–439. IEEE Comp. Soc., 2005.

[5] Crayola. Current crayola crayon colors. http://www.crayola.com/.

[6] M. Enriquez and K. MacLean. The Role of Choice in Longitudinal Recall of Meaningful Tactile Signals. In *Proc. of the Haptics Symposium*, pages 49–56. IEEE Comp. Soc., 2008.

[7] M. Enriquez, K. MacLean, and C. Chita. Haptic phonemes: basic building blocks of haptic communication. In *Proc. of Int'l. Conf. on Multimodal Interfaces*, pages 302–309. ACM, 2006.

[8] A. Gabrielsson. Adjective ratings and dimension analyses of auditory rhythm patterns. *Scand. Journ. of Psychology*, 14(1):244–260, 1973.

[9] D. Howard and J. Angus, editors. *Acoustics and Psychoacoustics (3rd ed)*. Focal Press, 2009.

[10] R. Ishii and M. O'Mahony. Taste sorting and naming: can taste concepts be misrepresented by traditional psychophysical labelling systems? *Chemical Senses*, 12(1):37–51, 1987.

[11] J. Lehrner, J. Glück, and M. Laska. Odor Identification, Consistency of Label Use, Olfactory Threshold and their Relationships to Odor Memory over the Human Lifespan. *Chemical Senses*, 24(3):337–346, 1999.

[12] J. Luk, J. Pasquero, S. Little, K. MacLean, V. Levesque, and V. Hayward. A role for haptics in mobile interaction: initial design using a handheld tactile display prototype. In *Proc. of SIGCHI*, pages 171–180. ACM, 2006.

[13] K. MacLean. Foundations of Transparency in Tactile Information Design. *Trans. on Haptics*, 1(2):84–95, 2008.

[14] K. MacLean and M. Enriquez. Perceptual design of haptic icons. In *Proc. of Eurohaptics*, pages 351–363, 2003.

[15] J. Melcher and J. Schooler. The misremembrance of wines past: Verbal and perceptual expertise differentially mediate verbal overshadowing of taste memory. *Journal of Memory and Language*, 35(2):231–245, 1996.

[16] W. Parr, D. Heatherbell, and K. White. Demystifying wine expertise: Olfactory threshold, perceptual skill and semantic memory in expert and novice wine judges. *Chemical Senses*, 27(8):747–755, 2002.

[17] C. Sandhofer and L. Smith. Learning color words involves learning a system of mappings. *Developmental Psychology*, 35(3):668–679, 1999.

[18] J. Schooler and T. Engstler-Schooler. Verbal overshadowing of visual memories: Some things are better left unsaid. *Cognitive Psychology*, 22(1):36–71, 1990.

[19] B. Swerdfeger. A First and Second Longitudinal Study of Haptic Icon Learnability. Master's thesis, Univ. of British Columbia, 2009.

[20] B. Swerdfeger, J. Fernquist, T. Hazelton, and K. MacLean. Exploring melodic variance in rhythmic haptic stimulus design. In *Proc. of Graphics Interface*, pages 133–140. Cdn. Info. Proc. Soc., 2009.

[21] D. Ternes and K. MacLean. Designing large sets of haptic icons with rhythm. In *Proc. of Eurohaptics*, pages 199–208, 2008.

[22] E. Tufte. *Envisioning information*. Graphics Press, 1990.

[23] J. van Erp and M. Spapé. Distilling the underlying dimensions of tactile melodies. In *Proc. of Eurohaptics*, pages 111–120, 2003.

[24] C. Ware. *Information Visualization: Perception for Design*. Morgan Kaufmann Publishers, 2004.