Exploring the Interplay of Visual and Haptic Modalities in a Pattern-Matching Task

Katie Seaborn, Bernhard E. Riecke, Alissa N. Antle
Simon Fraser University, School of Interactive Arts and Technology, Surrey, British Columbia, Canada
{kseaborn, ber1, aantle}@sfu.ca

Abstract— **It is not well understood how working memory deals with coupled haptic and visual presentation modes. Present theoretical understandings of human cognition indicate that these modes are processed by the visuo-spatial sketchpad. If this is accurate, then there may be no efficiency in distributing information between the haptic and visual modalities in situations of visual overload [1]. However, this needs to be empirically explored. In this paper, we describe an evaluation of human performance in a pattern-matching task involving a fingertip interface that can present both haptic and visual information. Our purpose was to explore the interplay of visual and haptic processing in working memory, in particular how presentation mode affects performance. We designed a comparative study involving a pattern-matching task. Users were presented with a sequence of two patterns through different modalities using a fingertip interface and asked to differentiate between them. While no significant difference was found between the visual and visual+haptic presentation modes, the results indicate a strong partiality for the coupling of visual and haptic modalities. This suggests that working memory is not hampered by a using both visual and haptic channels, and that recall may be strengthened by dual-coding both visual and haptic modes.**

Tactile displays, user interfaces, user interface human factors

I. INTRODUCTION

Representing information through tactile or haptic channels is a relatively untouched area of study in multimodal interface design. By and large, multimodal research has looked at how to design for the more common modalities of sight and sound. One focus of multimodal research has been on dual-coding modalities, or presenting linked information in different modalities with the effect of reducing cognitive load [1]; [2]. Cognitive load refers to the demand placed on working memory when processing sensory information. In Baddeley's model of working memory [5], visual and auditory sensory information are processed separately. The dual-coding of visual and auditory modalities has been supported by empirical research, particularly with regard to multimedia presentations [3]. Mayer [4] devised his Multimedia and Modality Principles, which include dual-coding visual and auditory modalities. However, research has been limited to these modalities, perhaps because the model upon which they are based—Baddeley's model of working memory—accounts only for these modalities. Research is needed to discover if the haptic channel can be used to reduce cognitive load in situations of visual overload, or excessive visual stimuli [6]. Two ways to achieve this are by distributing some of that information to another modality or dual-coding modalities. Current theory suggests that the visuo-spatial sketchpad, which is responsible for processing visual information in working memory, may also process haptic information; making use of haptic modalities in cases of visual overload may then be inconsequential [6]. Sweller's Cognitive Load Theory [7] would suggest a conflict between visual and haptic modalities in working memory, if these channels are processed by the same system. Indeed, dual-coding visual and haptic modalities may introduce a bottleneck in working memory. Empirical research is needed to resolve these unknowns.

Current research on how haptic and visual modalities work together is incongruent. At least one study suggests that working memory may blend haptic and visual information [8]. Haptic augmentations in a mobile touch screen device were found to be beneficial for feedback but had no statistically significant effect with varying cognitive load [9]. Another study promotes the use of haptic modalities to lessen cognitive load when the visual modality is stressed. This study looked at the use of haptic cues in a multimodal driving-like task; the authors concluded that haptic feedback is "… a robust, intuitive and non-intrusive way to communicate information to a user performing a visual primary task" [10]. Recent studies also advance making use of haptic modalities for increased performance and quality of experience [11]. But whether or not a bottleneck in processing occurs is unknown.

This research informs our understanding of the interplay of visual and haptic modalities in working memory through a pattern-matching task that dual-codes sequence information in both modalities simultaneously. The results add to understanding of human performance with tasks involving working memory and these presentation modes. Further, it shows how the pairing of these presentation modes is a worthwhile and potentially useful alternative to traditionally paired presentation modes, like visual and audio. Not only could this lead to novel applications of these presentation modes, but it could provide new solutions to the problem of visual information overload or be used to reinforce attention to vital information (in the case of a paired visual+haptic presentation mode). Ignoring the potentials of the results is to ignore the theoretical and applied implications for haptic presentation modes, particularly with respect to cases of visual overload. While further research is needed to provide empirical support for pursuing these applications, the results expound the potential of pairing these presentation modes.

Our experiment addresses this theoretical uncertainty through a pattern-matching task using three representations of information—visual, haptic, and dual-coded visual+haptic—under varying difficulty of complexity of a pattern. We look at the following research questions and hypotheses: First, how does presentation mode (visual, haptic, and visual+haptic) affect users' performance in a pattern-matching task? Do dual-coded haptic and visual channels negatively affect performance in a pattern-matching task that relies on working memory? We hypothesize participants' accuracy will be lower in the visual+haptic mode than either visual or haptic modes for all complexity levels of the pattern-matching task (H1). We further hypothesize participants' task times will be longer in the visual+haptic mode than either visual or haptic modes for all complexity levels of the pattern-matching task (H2). Our expectations are shaped by current understandings of how working memory processes visual and haptic modalities of information and the implied attentional bottleneck when both these modes are used simultaneously. Second, which presentation mode (visual, haptic, and visual+haptic) do users prefer in a pattern-matching task? We anticipate users will prefer either haptic or visual over visual+haptic modes in a pattern-matching task (H3). We assume that when performance is stressed, users will not prefer the presentation mode that causes this stress; hence our prediction that should users perform worse with the visual+haptic presentation mode, they will also not prefer it.

II. Methods

We conducted a controlled lab experiment to investigate our hypotheses about the effects and interplay of visual and haptic modalities on working memory performance.

A. Experimental Design

The experiment has a 2x3 within-subjects design: all participants experience all levels of both independent variables. Crossover effects were accounted for by counterbalancing all levels of mode and complexity, and randomly attributing counterbalanced trials to each participant.

1) Independent Variables

The primary independent variable was the presentation mode (hereafter referred to as "mode"), and the conditions were visual (V), haptic (H), and visual+haptic (V+H). The secondary independent variable was complexity as modulated by the length of pattern: lengths of three (3), five (5), and seven (7). These lengths were chosen based on recent research on the limits of capacity in working memory [12].

2) Trials

Each session per participant was divided into three sets of trials, one for each of the three conditions (presentation modes). Each set totaled 42 trials and was comprised of 14 repetitions of the three complexity levels. In total, there were 126 trials per participant. One half (50%) of the 42 repetitions per condition had the same pattern, and the other half had a different pattern. The order of trials was randomized within each block of 42 trials. Trials were presented in uninterrupted sequence with a short break at the halfway point. The average duration for all trials per participant was roughly 45 minutes, and varied due to response time.

3) Internal Validity

To control for subjective variability, a within-subjects design was used whose conditions were counterbalanced and randomly assigned to participants. Each participant was presented with a random selection of order of mode and a random order of complexity levels per trial. For example, a participant could be randomly assigned visual, then visual+haptic, then haptic conditions.

B. Participants

Nineteen (19) undergraduate students were recruited. One participant experienced a programming glitch during the experiment, and their results had to be removed; thus there are eighteen (18) accounts of performance and nineteen (19) accounts of preference. Participants' ages ranged between eighteen and thirty years old. Eleven participants were female and eight male. Each participant was screened for right-handedness, not being a student of the researchers, and having no known working memory difficulties. Participants were given bonus course credit in their respective design courses as compensation for participating in the experiment.

C. Tasks

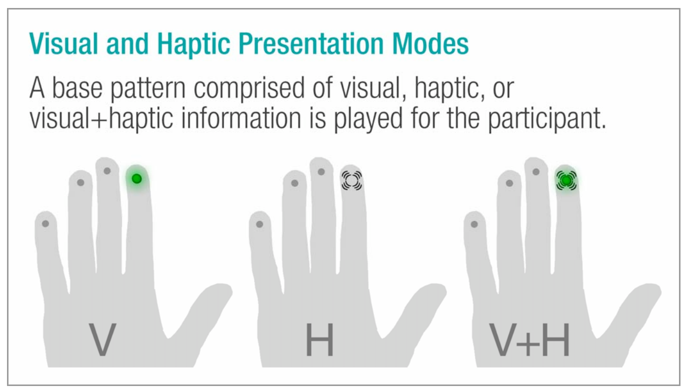

Figure 1. The task presentation modes (independent variable).

1) Pattern-Matching Task

The psychophysics-inspired task involved comparing presentation modes (Fig. 1). Participants were presented with a series of two patterns (Fig. 2). The task was derived from a study which compared visual representations of forms of varying complexity in a sequence matching task [13]. Each pattern was a sequence of discrete stimuli, or items, spatially distributed across fingers.

Participants were asked to indicate whether or not the two patterns presented were the same by pressing dedicated buttons on a keyboard. Each item in a given pattern was presented for 500 ms, with a pause of 500 milliseconds between items (these times were determined in the pilot study). Task complexity was varied by modulating the length (number of items) of the pattern. A pause of 3000 milliseconds was introduced to distinguish the two patterns played in series. Following user input, there was a pause of 6000 milliseconds before the next series of patterns began. Participants were able to input a response after the second pattern began; they were told to respond as soon as they knew the answer.

2) Auditory Task

To prevent the encoding of haptic information to the auditory channel, participants were asked to continuously recite the English alphabet throughout the pattern-matching task. This task is taken from another articulatory suppression task involving whistling [14]. It was chosen based on its relatively low cognitive load – the task procedure remains in working memory while it is repeated – which was preferred in order to avoid increasing the perceived difficulty of the pattern-matching task through stressing more than one modality. In our pilot study, we found that the haptic actuators emit a distracting sound when activated; to account for this, we added noise-canceling headphones and a techno loop track.

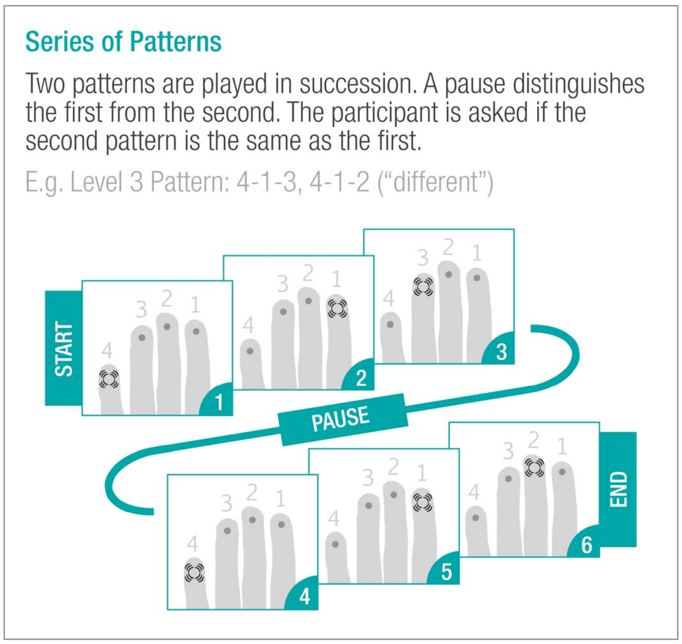

Figure 2. A series of patterns comprised of a sequence of stimuli, or items.

A possible confounding factor is the auditory task. Current understandings of working memory suggest we are able to encode one information type into another, e.g. visual information to auditory (or articulatory). To account for this, we introduced an auditory task whose goal was to occupy the phonological loop and therefore prevent cross-modality encoding. However, performing the task along with the main pattern-matching task could increase cognitive load and therefore affect performance and preference data.

D. Research Instrument

1) Description

A fingertip interface (Fig. 3) was constructed for the pattern-matching task. It consists of a wood board with four glove tips attached such that a hand could sit comfortably on the board with all fingers resting inside the tips. On the tips are green LEDs; inside are pager motors which provide vibration actuation and are shaped like small discs. Participants were asked to rest their fingertips comfortably on these discs throughout the duration of the experiment. The physical prototype is controlled by an Arduino board that receives commands from a MaxMSP patch in a laptop computer. The MaxMSP patch is comprised of the Arduino controller, pre-recorded patterns, a trial facilitator, and a data collection mechanism. Participants used a standard keyboard with clearly marked keys to input their responses.

E. Procedure

Participants were greeted with a brief introduction and signed the consent form. A screening test (pre-task questionnaire) was then administered to ensure the participant was right-handed, not a student of the researchers, and had no known working memory problems. The participant then did a learning trial, during which time they gained a feel for the research instrument and an understanding of the main pattern-matching task. Participants were told to answer as quickly but as accurately as possible. They were given as much time as they needed in the learning trial, and were asked if they had any questions about what they were expected to do in the experiment. After completion of the learning trial, participants were introduced to the noise-canceling headphones and auditory task. The main trial (see B.1.) then began, during which performance data was collected. After each condition, the participant was allowed to take a short break. After the conclusion of the main trial, participants were asked to fill out a post-task questionnaire. Participants were then asked if they had any remaining questions or feedback. Finally, they were thanked for their time.

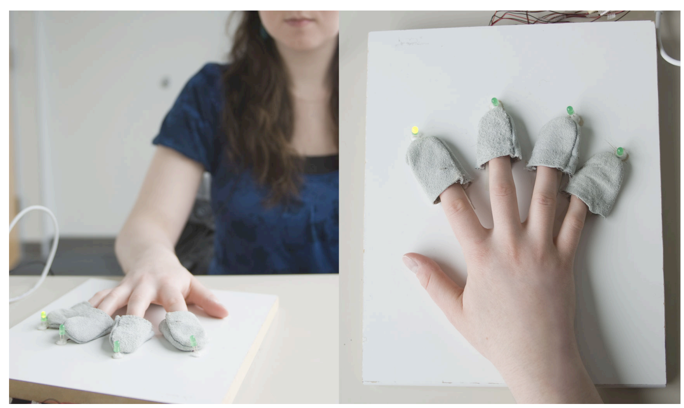

Figure 3. Fingertip interface for the pattern-matching task.

F. Measures

1) Dependent Variables

The primary dependent variable is performance operationalized as accuracy and task time. Accuracy is defined as score per trial, either wrong (0) or right (1), and is nominal. Task time was recorded in milliseconds, and is ratio. The research instrument recorded this data via a text file generated per participant.

2) Self-Reports of Performance and Preference

The secondary dependent variable is preference for presentation mode. Preference was collected through a post-task questionnaire. The questionnaire had ten questions: self-reports of performance using a 7-point Likert scale (Q1-8), preference for presentation mode (Q9), and an open-ended question for qualitative feedback (Fig. 6). Self-reports of performance were collected in order to validate and support the performance data recorded by the research instrument. We expected that comparing self-reports of performance with the recorded performance data would be interesting regardless of whether or not they matched. If they matched, the former could be said to support the latter; and if they did not match, this would prompt further research to discover the cause of this discrepancy.

III. Results

A. *Performance*

1) *Accuracy*

Means and standard error for accuracy are in Fig. 4. Mauchly's Test of Sphericity showed no significant difference between groups, allowing for the use of a repeated measures ANOVA parametric test for all three within-subjects effects (mode, complexity, and mode × complexity).

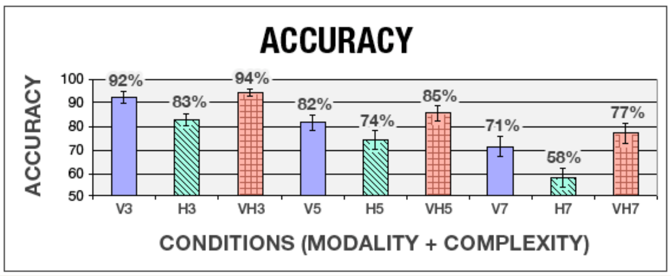

Figure 4. Means and std. error for accuracy across mode and complexity (chance level is 50%).

Repeated measures ANOVA results for the dependent variable accuracy showed significant main effects for both presentation mode ($F(2) = 14.10$, $p = .0001$) and complexity ($F(2) = 43.66$, $p = .0001$). The interaction between mode and complexity was not significant. Pairwise comparison results for mode indicated that participants performed significantly more accurately at the $p < .05$ level in the visual mode as compared to the haptic mode ($p = .02$) and in the visual+haptic mode compared to the haptic presentation mode ($p = .05$). There was no significant difference in accuracy between the visual and visual+haptic presentation modes. Pairwise comparison results for complexity indicated that participants performed significantly more accurately at the $p < .001$ level for all complexity levels.

The statistical results do not support hypothesis (H1). Participants performed most accurately with the dual-coded visual+haptic presentation mode. The addition of haptic information to visual information did not degrade accuracy. The presentation of haptic information alone was not effective.

2) *Task Time*

Means and standard error for task time are in Fig. 5. Mauchly's Test of Sphericity showed a significant difference at the $p < .001$ level for mode × complexity, but not mode or complexity alone, allowing for the use of a repeated measures ANOVA against only the first two within-subjects effects.

Repeated measures ANOVA results for the dependent variable task time showed significant main effects of both presentation mode ($F(2) = 4.02$, $p = .05$) and complexity ($F(2) = 247.32$, $p = .0001$). Pairwise comparison results for mode indicated that participants performed significantly faster in the visual+haptic mode as compared to the haptic mode ($p = .014$). There was no significant difference in participants' task times between the visual and visual+haptic modes, or the visual and haptic modes. Pairwise comparison results for complexity indicated that participants performed significantly more quickly in the second level of complexity (sequence of five) compared to the third (sequence of seven) ($p = .014$).

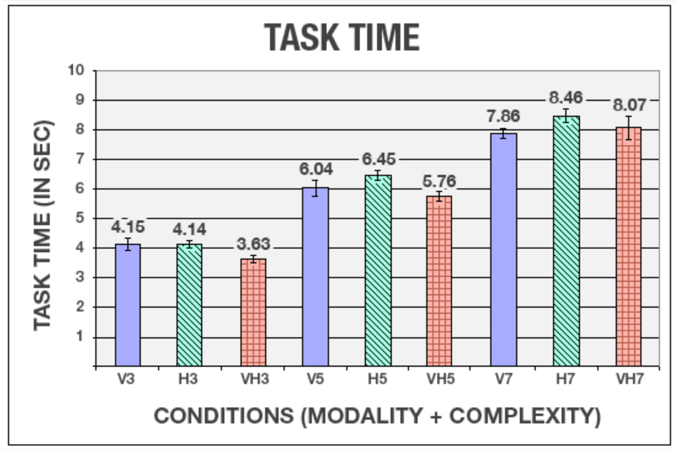

Figure 5. Means and std. error for task time across mode and complexity.

The statistical results do not support hypothesis (H2). Participants performed quickest with a presentation mode that included visual information. The addition of haptic information to visual information did not significantly degrade task time; in fact, this dual-coding of modalities significantly enhanced the efficiency of the haptic mode on its own.

3) *Self-Reports of Performance*

The first eight questions (Q1-8) in the post-task questionnaire addressed self-reports of performance (Fig. 6). Mode was used as a measure of central tendency for the 7-point Likert scale data (Fig. 6). On average, participants found the visual and visual+haptic modes somewhat easy to use, and the haptic mode difficult. Participants found the overall pattern-matching exercise neither difficult nor easy. Participants disagreed that they were able to repeat the alphabet (perform the auditory task) throughout the exercise without stopping. Participants somewhat agreed that they were able to follow the visual patterns with their eyes the entire exercise; somewhat disagreed with regard to the haptic patterns; and agreed with regard to the visual+haptic patterns.

The statistical results do not support hypothesis (H3). Participants preferred presentation modes that had a visual mode; this includes the visual+haptic presentation mode, which did not suffer in the performance sector as hypothesized in (H2). Further, the majority of participants preferred the visual+haptic mode, and no one preferred the haptic mode.

Figure 6. Modes for self-reports of performance.

Self-reports of performance support the accuracy and task time findings. Participants found the haptic mode on its own difficult to use. Participants found that they performed similarly with visual and visual+haptic modes. Regardless of presentation mode, participants felt neutrally about the difficulty or ease of the pattern-matching task and struggled to maintain the auditory task throughout the exercise.

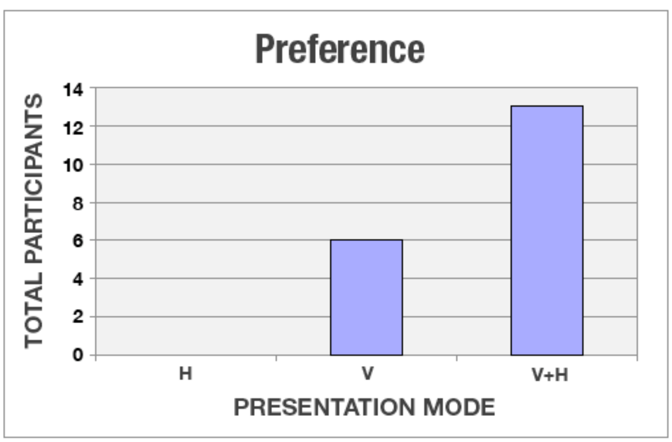

Figure 7. Preference for presentation modes.

B. Preference

In reply to: "Which presentation mode did you prefer?" no participants preferred the haptic (H) presentation mode, 6 out of 19 (32%) preferred the visual (V) presentation mode, and 13 out of 19 (68%) preferred the visual+haptic (V+H) presentation mode (Fig. 7).

In reply to an open-ended question, "Do you have any comments about your experience?" four participants reported a numbing to the haptic feedback over time. One participant said: "During the [haptic presentation mode] there was a point where all the vibrations felt the same no matter what finger. Could be because of longevity. At the get go it was easily distinguishable." Two reported discomfort with the fingertip interface. One participant was distracted by the music, and another by the articulatory task.

IV. DISCUSSION

The results indicate that visual presentation modes aid performance in a pattern-matching task. Further, the dual-coding of visual and haptic modes appears to benefit instead of degrade performance. While participants performed comparably with visual and visual+haptic modes, the reported quality of experience favours the coupling of visual and haptic modes. These results support findings by [9]; [10]; and [11].

No participant preferred the haptic mode. The final open-ended question in the post-task questionnaire (Q10) reveals a possible reason why: haptic adaptation, with respect to participants ultimately becoming less sensitive to the haptic mode. This has been noted previously, such as in a study involving a vibrotactile actuator [15]. Possible ways to counter this are: increasing the time between trials, allowing participants to break more often, or decreasing the magnitude and/or length of the haptic stimulus. Perhaps a different style of haptic feedback—for example, force—would be more suitable for repetitions and longevity, and thus would increase preference for that presentation mode. Further, other styles of haptic feedback, like force, would eliminate the need to block sound [16] emitted by certain actuators, like the pager motors used in this experiment. Participants' self-reports of performance with relation to the articulatory suppression task (reciting the alphabet continuously) suggest that this task may have increased cognitive load, despite what was found in [14]. Using another style of haptic feedback might reduce his issue; however, we suggest that further research is needed to discover what effect the articulatory suppression task has on cognitive load.

Haptic modes may be strengthened when paired with visual counterparts to provide the same information. However, the findings suggest haptic modes may also be salient in situations where different but simultaneous modes occur. Empirical research is needed to assess this possibility.

The high performance and preference scores for presentation modes that include a visual component could be accounted for by our natural predisposition for the visual modality. However, the findings suggest that even in cases of increased complexity, dual-coding visual and haptic modalities rate on a similar level of performance as the visual modality alone. Further, there is no indication of a bottleneck in working memory. The findings provide a basis for extending Mayer's [4] Multimedia and Modality principles to include the dual-coding of visual and haptic modalities for reducing cognitive load.

This research looked at a particular task (pattern-matching) and so the results may or may not be generalizable to other applications. More research is needed before these findings can be generalized for use in different contexts, for example adding haptic presentation modes into a visually-overloaded environment that also uses visual presentation modes.

V. CONCLUSION

The findings support the dual-coding of visual and haptic modalities and advocate this pairing for further research in a wider variety of interfaces and contexts. Further, the findings show no support for the notion of an attentional bottleneck in working memory, and do lend support the use of haptic channels in high cognitive load situations involving the visual modality, as per Sweller's [7] Cognitive Load Theory and Mayer's [4] Multimedia and Modality principles.

Novel and useful applications of these results are conceivable. This research can be used as a baseline for extending into the context of a visually complex multimodal gaming, training or simulation environments. In this context, haptic presentation modes can be evaluated against their visual counterparts or simultaneously occurring but different visual information. A comparison of dual-coded visual+haptic and auditory+haptic modes may yield interesting results. This research was limited to one type of haptic feedback—vibration—and calls for an evaluation of other types, such as force or movement.

The theoretical implications of this research combined with potential applications make haptic and visual presentation modes a fertile ground for additional research. Further work with these dual-coded modes will open up testable solutions to the limited styles of presentation modes currently so widespread. We urge for a focus on these combined modalities and speculate that the empirical research to come will be of valuable in this area.

ACKNOWLEDGMENTS

The authors would like to thank their colleagues Andy Huang, Daniel Feuereissen, and Mahshid Zeinaly for their support during the experimentation and writing process. We thank Brian Fisher for providing guidance on the state of attentional bottlenecks in multimodal research. Finally, we thank Andrew Hawryshkewich for the photos in this paper.

REFERENCES

[1] S.Y. Mousavi, R. Low, and J. Sweller, "Reducing cognitive load by mixing auditory and visual presentation modes," *Journal of Educational Psychology*, vol. 87, 1995, pp. 319-334.

[2] Y. Cao, "Modality effects on cognitive load and performance in high-load information presentation," *International Conference on Intelligent User Interfaces*, Sanibel Island, Florida, USA: ACM, 2009, pp. 335-344.

[3] R.E. Mayer, *Multimedia Learning*, Cambridge, New York: Cambridge University Press, 2001.

[4] R. Mayer, *The Cambridge Handbook of Multimedia Learning*, Cambridge, New York: Cambridge University Press, 2005.

[5] A. Baddeley, "Working Memory," *Comptes Rendus de l'Académie des Sciences*, vol. 321, Mar. 1998, pp. 167-173.

[6] A.N. Antle and A.F. Wise, "Getting down to details: Using learning theory to inform tangibles research and design for children," *Interacting with Computers*, under review, available at http://www.antle.iat.sfu.ca/publications.php (accessed September 9, 2010).

[7] J. Sweller, "Cognitive Load during Problem Solving: Effects on Learning," *Cognitive Science*, vol. 12, 1988, pp. 257-285.

[8] D. Kerzel, "Visual short-term memory is influenced by haptic perception," *Journal of Experimental Psychology*, vol. 27, Jul. 2001, pp. 1101-1109.

[9] R. Leung, K. MacLean, M.B. Bertelsen, and M. Saubhasik, "Evaluation of haptically augmented touchscreen gui elements under cognitive load," *Proceedings of the 9th international conference on Multimodal interfaces*, Nagoya, Aichi, Japan: ACM, 2007, pp. 374-381.

[10] M. Enriquez, K. MacLean, and H. Neilson, "Interplay of tactile and visual guidance cues under multimodal workload," *Department of Computer Science, University of British Columbia, Vancouver*, 2007.

[11] M.F. Zaeh, B. Faerber, B. Deml, H. Egermeier, J. Schilp, and S. Clarke, "A study on visual, auditory, and haptic feedback for assembly tasks," *Presence*, vol. 13, Feb. 2004, pp. 16-21.

[12] N. Cowan, "The magical number 4 in short-term memory: A reconsideration of mental storage capacity," *Behavioral and Brain Sciences*, vol. 24, 2001, pp. 87-114.

[13] D.S. Klopfer, "Constructing mental representations of objects from successive views," *Journal of Experimental Psychology*, vol. 11, Oct. 1985, pp. 566-582.

[14] S. Saito, "Phonological loop and intermittent activity: A whistle task as articulatory suppression," *Canadian Journal of Experimental Psychology*, vol. 52, Mar. 1998, pp. 18-23.

[15] S. Holland, A.J. Bouwer, M. Dalgeish, and T.M. Hurtig, "Feeling the beat where it counts: Fostering multi-limb rhythm skills with the haptic drum kit," *Proceedings of the fourth international conference on Tangible, embedded, and embodied interaction*, Cambridge, Massachusetts, USA: ACM, 2010, pp. 21-28.

[16] C. Forlines and R. Balakrishnan, "Evaluating tactile feedback and direct vs. indirect stylus input in pointing and crossing selection tasks," *Proceedings of the twenty-sixth annual SIGCHI conference on Human factors in computing systems*, Florence, Italy: ACM, 2008, pp. 1563-1572.