Comparing Gesture and Touch for Notification System Interactions

Maria Karam
Ryerson University
Centre For Learning Technologies
Toronto, Ontario, Canada
maria.karam@ryerson.ca

Jason Chong Lee
Virginia Tech.
Department of Computer Science
Blacksburg VA, USA
chonglee@vt.edu

Travis Rose
Virginia Tech.
Dept. of Computer Science
Blacksburg VA, USA
rtrose@vt.edu

Francis Quek
Virginia Tech.
Dept. of Computer Science
Blacksburg VA, USA
quek@cs.vt.edu

Scott McCrickard
Virginia Tech.
Dept. of Computer Science
Blacksburg VA
USA
dmccrick@cs.vt.edu

Abstract

We explore some of the characteristics of multimodal input interaction spaces for notification systems within a multi-tasking environment like a command and control center using two promising interaction methods: gestures and touch based input through a laboratory experiment comparing both techniques. Results of our study suggest that gestures are better suited for multi-tasking situations because they are less interruptive than touch interaction to users' primary tasks and are subjectively preferred by users in certain situations.

1 Introduction

Notification systems are designed to provide users with often critical information to assist with their everyday tasks, while causing minimal interruption to their current tasks. Interactions with notification systems often only require users to acknowledge the notification, which takes the form of an acceptance or rejection of the notification alert. Such notifications can be executed as a secondary task to the user's primary attention focus [12]. But even with the simplicity of the commands associated with notification system interactions, acknowledging or otherwise responding to a notification system alert can result in more disruptive interruptions to a user's primary task. For example, switching one's interaction context from the keyboard to the mouse to respond to a notification alert can lead to unnecessary distractions to the user's primary task [12]. Gestures represent a natural way in which humans can dismiss or acknowledge notifications, both in real and digital worlds. They are an intuitive form of interaction that can readily be implemented using computer vision technology [10]. Although gestures offer a less precise form of interaction than direct input methods like the mouse or keyboard, most systems do provide over 90% accuracy in detecting gestures [11]. While gestures are not a replacement for the mouse or keyboard, they hold the potential to effectively serve as an additional input modality that can easily be integrated into most existing systems [10]. Although we only see vision-based gestures used in a few gaming applications [1, 2], web cameras have effectively become ubiquitous components on personal computers and laptops, providing the infrastructure necessary to implement gestures as another form of input that can be geared specifically at controlling notification system interactions. In this paper, we report on an experiment that compares semaphoric hand gestures to touch screen input for notification system interactions, suggesting that gestures offer a less distracting method of responding to simple notification system alerts than is possible using touch input. We measure the effectiveness of gestures over touch screen interactions within a multi-tasking situation, similar to a command and control center, for minimising the level of attention required to respond to a secondary task, while optimising utility for the user in maintaining focus on their primary task. Details and results of the experiment are presented, along with a discussion addressing the relevance of these findings towards promoting the adoption of gestures as a valuable contributor to multimodal interactions for improving notification system interactions.

2 Related Work

The increasing use and pervasiveness of computer systems in command and control environments in our everyday lives often result in people managing multiple information streams and engaging in several tasks simultaneously [5]. In spite of the inherent problems with the interruptions these systems can cause, they are often vital aspects of the overall goals people are accomplishing. For example, large screen displays have been integrated into office and classroom environments to allow coworkers to maintain awareness of each other's tasks and schedules while engaged in a primary work task [9, 14]. However, while ubiquitous systems can assist in managing multiple streams of information and supporting communication and group activities, users must also manage their attentional resources to glean the benefits of such systems. One of the key problems for attention management in multitasking situations is the disruption caused to a current task when an interruption from a different system occurs. For instance, Czerwinski et al. [7] observed the harmful effects of instant messaging notifications to the performance of a list evaluation task. Fogarty et al. [8] and Bailey et al. [4] have explored when and how interruptions should be delivered, but our work specifically deals with how different interaction techniques affect users' primary task during secondary task interactions. This tension between maximising utility for users and managing their attention is a key theme of McCrickard et al's [13] work in attention and system design, which can effectively guide the development of notification systems that deliver current, important information in a variety of platforms and modes without excessively disrupting users from their primary task. For example, notification systems can be described using three critical parameters — interruption, reaction and comprehension (IRC). This IRC framework has been used to guide the design of notification systems, as well as create benchmarks and performance goals for improving existing systems [15, 13, 6, 7]. Using the IRC framework, we can describe a notification system in terms of the level of interruption it provides, the type of reaction required by the user, and the amount of comprehension necessary to appropriately notify the user of the underlying event. Thus, an instant messaging window that indicates a new message has arrived could have a low interruption level, low reaction requirements, and low levels of comprehension to address the notification. More critical notification systems would present high levels of interruption, require immediate reaction, and a higher level of comprehension to appropriately respond, as when monitoring system critical information. With the emergence of notification systems within ubiquitous computing scenarios, we find novel input methods such as touch screens, voice input, and gestures, that support richer interactions between people and systems. These input methods are designed to support more intuitive, expressive interactions within environments that would be difficult or impossible to control using traditional desktop input modes such as a mouse and keyboard [3, 16]. However, understanding when specific characteristics of input devices are most suitable to a given task can assist users in better managing their attention and minimising disruption from primary tasks. For example, gestures provide a natural way to interact with computer systems at a distance; in particular, semaphoric gestures, which involve specific hand motions and poses to communicate symbols may be ideal for interacting with secondary tasks supported by notification systems [11]. In this paper, we claim that gestures offer an effective mode of interaction for notification systems that is less disruptive, and as effective as touch based interactions. We next present the details of our experiment and the context in which we compared gestures to touch input for notification system interactions.

3 Experiment

To test the effectiveness of using gestures as an input modality for notification system interactions, we created a dual-task situation that simulated a command and control environment requiring users to monitor and interact with multiple information systems simultaneously. The primary task required participants to search through a satellite photo, and locate a specific area on the satellite within a specified time limit. The satellite image was displayed at a resolution of 1024x768 on a 136cm x 101cm SmartBoard[TM] large-screen touch screen as the primary display. Users had to locate the target area within the satellite image, which was displayed on the right side of the screen as a 45x45 pixel image (see Figure 1 for experiment the set up). For the secondary task, participants monitored a notification system animation presented on a peripheral secondary display located to the side of the primary display.

We simulated user interactions on the peripheral display using a Wizard of Oz (WoZ) technique [17], and provided two levels of interruption, low and high. The wizard used a wireless keyboard to control the peripheral display input in response to participant actions (gestures or touch). Both the gestures and touch interactions were simulated by the wizard during the experiment to further control for differences in the two systems and to reduce the confounds arising from user variation in experience and in learning to use these two input techniques. Participants performed single right-handed gestures and touch input for the secondary task, which was located to the right of the SmartBoard[TM] but were free to use either hand to select regions on the main display and resume the primary task. We recruited 20 undergraduate and graduate students from a variety of disciplines at Virginia Tech to participate in the study.

Figure 1. Participant conducting focal search task (left) with secondary notification task (right).

Figure 2. Overhead view of experiment layout with [c]amera, [p]articipant, [w]izard, kb = keyboard.

Our experiment tested the following hypotheses in the context of the IRC framework:

1. Interruption: Semaphoric gestures support less interruptive interactions with notification systems than touch screen interactions, resulting in higher primary task performance.

2. Reaction: Semaphoric gestures support more efficient reaction to both high and low interruption notifications, resulting in higher secondary task performance.

3. Comprehension: Gesture interactions do not decrease comprehension levels of the content on the notifications, leading to similar success rates for secondary tasks completed.

4. Satisfaction: Semaphoric gestures are as intuitive, easy, and efficient to use as touch-screens for interacting with notifications, resulting in higher subjective ratings and user preference for gestures.

The experiment was a full factorial, repeated measures design with interaction modes (gesture and touch) as the within-participant factor, and the interruption level (high or low) as the between-participant factor. The interaction modes were counter-balanced within participants, consisting of one set of trials using gestures, and the other using a touch screen. For the low interruption condition, a visual notification was used, and in the high interruption condition, both an audio and visual notification were presented to the participants. We tested 4 conditions, and they are coded based on the interruption level of the notification (0=low, 1=high) and counterbalancing (G=gestures seen first, T=Touch screen seen first). The four conditions were thus 0G (low interruption, gestures first), 0T (low interruption, touch first), 1G (high interruption, gestures first), and 1T (high interruption, gestures first). Our within-participant variable was the interaction mode, where each participant used both the touch screen and the gesture interaction. The
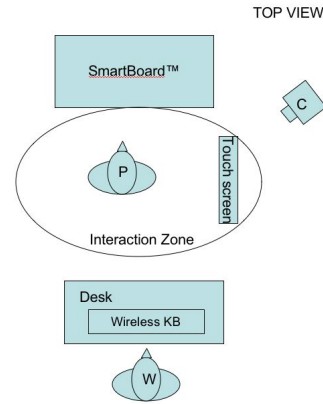
between-participant variable used in the experiment consisted of altering the interruption level so that in the low interruption condition, participants only received a visual notification for the secondary task, and in the high interruption condition, participants received both an audio and the visual notifications.

3.1 Method

To investigate different settings of the IRC values for notification system interactions, we created two interaction contexts to compare gesture and touch input for supporting effective notification system interactions. In this paper, we define interaction zone as the area in which a user can conduct purposeful interactions with the system; as illustrated in Figure 2, this area coincides with the camera's field of view. In practice sessions, participants were trained to use a computer-vision enabled recognition system for semaphoric gesture interactions [10], along with the touch screen interface. The gestures we used in the experiment consisted of three single right-handed waves in the direction of the camera, which were monitored by the wizard. The participants stood in front of the SmartBoard™, and the camera was positioned behind and to the right of the secondary task display to capture the users motions during the trials. This positioning enabled gestures and the touch interactions to be performed using similar motions in a similar interaction zone. The notifications and secondary tasks were all presented on the secondary touch screen, which was positioned within the periphery of participants sight line. Our prototypical gesture set consisted of leftward, vertical, and rightward hand motions. We chose gestures that were easy to perform and had straightforward mappings to concepts (left, middle, right). We wished to minimise the cognitive effort required to perform the correct semaphore in response to notifications. In this case, the gesture path encoded the relative positions of the coloured bars present in the notification display.

The gestures were understood to serve the function of acknowledgement as well as having specificity for the particular notification issued. During the course of the primary search task, participants were presented with a series of notifications at 20-25 second intervals. In the low interruption mode, notifications consisted of a visual strobe lasting 0.333 seconds each second (1/3 duty cycle). In the high interruption condition, the notification was issued as a visual strobe (1/2 duty cycle) with an initial beep lasting 1 second, available from the Java 2, Standard Edition, v 1.4.2 sample code (TicTacToe applet). Sound was issued from speakers positioned near the secondary display for a spatially appropriate cue. Figure 1 shows a participant working on the primary task during the experiment.

To determine the degree of interruption and efficiency of semaphoric gesture vs. touch screen, we measured reaction time, recovery time, and secondary task time (defined below). Times when a notification was issued and when the primary task resumed were recorded by software. The time of the response to the notification was logged via keypress at the wizard's station. Regarding the relative ease of execution of gesture vs. touch, we looked for significant differences in reaction and recovery times. To infer effects on attention, we calculated the search success rates and measured the search times (primary task). We also recorded the number of secondary tasks completed and the total time to perform each (secondary task time). A detailed definition of the dependent variables is given below:

- Reaction time: time taken to respond to a notification. Measured from time notification is issued to time of user's response to notification.

- Recovery time: time taken for user to resume primary task. Measured from end time of secondary task to resume time for primary task.

- Primary task success rate: number of times an image was found during the focal search task.

- Secondary task success rate: number of successful responses to alerts before timing out.

- Primary task time: total amount of time participants spent searching for an image.

- Secondary task time: the total amount of time participants spent responding to notifications.

4   Results

We initially ran a repeated-measures analysis of variance (ANOVA) to test the effects of interaction mode within-participants for each of the two input modes (gesture and touch). Observations of mean values for recovery time/reaction time are shown in Table 1, with mean

		Mode			
		Gestures		Touch Screen	
		Reaction	Recovery	Reaction	Recovery
		Mean	Mean	Mean	Mean
Condition	OG	2.984	1.868	3.238	1.921
	OT	2.362	1.999	2.286	2.282
	1G	2.410	2.082	2.424	1.914
	1T	2.103	2.018	2.281	2.256

Table 1. Mean times, in seconds, for reaction and recovery for gestures and touch input.

values for primary and secondary tasks shown in Table 2. Reaction time was significant in the ANOVA, with the mean reaction time for gestures faster in all but the 0T condition. We next ran a multivariate-analysis of variance (MANOVA) to examine the effects due to the independent variables of mode and interruption level and counterbalancing in isolation of the four conditions between participants. While our ANOVA showed no significant results for the within-participants factor of interaction mode (gesture or touch), significant results were found in the between-participant MANOVA, suggesting differences in primary and secondary tasks success rates, reaction time and secondary task time for the different conditions.

The MANOVA revealed that an interaction effect exists between modes and counterbalancing, suggesting a learning effect was present. The MANOVA also revealed that interruption level yields a faster reaction time with the high interruption notification than for low interruption ($F_{(1,32)}$ = 6.383, p <.05). While no interaction effect was shown between gestures and interruption level in the MANOVA, reaction times for gestures tended to be faster than those for touch screen (see Table 1). An interaction effect is shown for reaction time, between mode and order of presentation, suggesting that a learning effect positively affects reaction times for gestures in certain cases.

In condition 0G, reaction time for touch actually increased, while in the 1T condigion, reaction time for gesture decreased during the second trial with 0T and 1G having similar reaction times ($F_{(1,32)}$ = 9.583, p<.005). Our analysis did not show any significant differences in the recovery times, however, there is a definite trend towards faster recovery using gestures over the touch interaction (see Table 1). This applies to all but the 1G condition, where gestures appear to lead to a slower recovery time than for the touch interaction. While the slower recovery time may be due to increased interruption level, and to the novelty of using gestures, the differences are not shown to be significant in this model.

For primary and secondary task completion rates, significant results were seen due to interaction mode in the ANOVA for primary task in the 0G condition with the gesture interactions showing a greater number of primary tasks completed ($F_{(1,8)}$ = 6.733, p<.05). Secondary tasks were

		Mode			
		Gestures		Touch Screen	
		Primary Tasks	Secondary Tasks	Primary Tasks	Secondary Tasks
		Mean	Mean	Mean	Mean
Condition	OG	9	25	4	31
	OT	8	30	10	22
	1G	8	25	9	29
	1T	11	30	7	26

Table 2. Mean values for primary and secondary tasks completed for gesture and touch conditions.

also significant in this model ($F_{(1,8)} = 5.829$, p<.05), however there were fewer secondary tasks completed for the gestures (Table 2). Significant results were found in the 0T condition, however gestures led to lower success rates for primary tasks ($F_{(1,8)} = 6.897$, p<.05) but a greater success rate for secondary tasks ($F_{(1,8)} = 9.218$, p<.05). The 1T condition suggest that gestures also support a higher success rate for primary tasks ($F_{(1,8)} = 9.529$, p<.05). No significant results were found for the 1G condition. The MANOVA results show that overall, there were significantly more secondary tasks completed in the second set of trials then in the first ($F_{(1,32)} = 14.286$, p<.001), which may explain the difference in primary task success rates between input mode in the ANOVA. There is also a significant interaction effect present for primary task success rate ($F_{(1,32)} = 16.711$, p<.001) due to the factors interruption level and presentation order.

In the first set of trials, more primary tasks were completed for low interruption than for high interruption. In the second set of trials, there were also fewer primary tasks completed in the high interruption condition than for low interruption. This suggests performance degradation occurs with higher interruption since it draws attention away from the primary task.

Subjective results show that despite the limited differences observed for the performance measures, participants showed an overall preference for gestures over the touch screen interaction (mean 7.15/10 for gesture). Participants rated gestures slightly higher than touch-screen for ease of interaction (mean of gestures=7.41; touch=7.22) but performance was slightly lower (mean of gestures=8.52; touch=8.78). Ratings for ease of resuming the primary task were higher for gestures (mean gesture=7.11; touch=5.22), and lower for attention required for secondary task (mean gesture=5.37; touch=6.74), and distraction caused (mean gesture=4.67; touch=6.63).

When we examine the results based on the different interruption levels, participants' perceptions were that gestures required less attention in the low interruption condition than in the high interruption condition. Touch screen interaction was roughly equivalent for both high and low interruption conditions. Gesture and touch were both rated as less disruptive in the high interruption condition while resuming tasks in general was rated as easier in the low interruption condition.

5 Discussion

Results from this study suggest that gestures were less interruptive than touch for interacting with notification systems. While our results show that the significant differences in performance are primarily affected by condition rather than the interaction modes, we observe that semaphoric gestures support less interruptive interactions with notification systems than touch-based interactions for low interruption secondary tasks. Our first hypothesis is thus supported for non-critical (less interruptive) secondary tasks. Overall results show that performance in the gesture condition does not significantly differ from performance results in the touch screen condition. This suggests that our second hypothesis is not supported when the interaction space for gestures and touch are similar. Since our interaction scenario uses similar interaction zones for both gestures and touch screen, we observe that the increased interaction zone possible with gestures would thus be more effective in a pervasive or ubiquitous computing environment where interactions at a distance are required.

Results also support our third hypothesis, as secondary task completion rates or comprehension levels were not affected by input mode, but rather, familiarity with the tasks and the interruption level of the notification played a more significant role. Gestures caused less distraction to primary tasks in the low interruption conditions, however, no differences were seen for gesture or touch in the high interruption condition suggesting that in a multi-tasking situation the choice of input mode can also have an impact on the user's reaction to secondary tasks. In particular, using gestures to respond to notifications leads to lower levels of distraction, whereas touch screen often resulted in loss of context in the focal search task. Our fourth hypothesis is thus supported for non-critical notification tasks. We observed that gesture performed as well as touch-screen, and was subjectively the preferred mode of interaction for the majority of users, further supporting our fourth hypothesis.

Most participants experienced gestures as a less disruptive interaction, and it permitted easier resumption of their primary task. For gestures, several users reported an ability to search without feeling tied to the secondary display. The benefit of gesturing in-place avoids re-purposing ones hands to carry out a secondary task such as responding to a notification, and avoids changing ones physical position such that the current context in the focal search task is lost. Thus, after a period of adjustment, gesturing was seen as easier by the majority of participants.

Our results show the most benefit for gesture occurs

when attention draw is low. The benefit appeared to be masked in the high interruption group, as there was no significant difference in performance. We note that the higher interruption was an increased rate of visual strobe with an accompanying sound; as expected, this was sufficient to reduce the reaction times. However, the greater level of distraction meant that participants lost the benefit of gesture over touch-screen interaction. We provided a spatially appropriate cue, which caused eye gaze to be diverted to the notification task. This effect interfered with the eyes free benefit of gestures.

6 Conclusions and Future Work

Using the IRC framework, we were able to select two different styles of notification, knowing in advance there was a salient difference in the level of interruption. The efficacy of gestures versus touch were compared in a multitasking situation meant to simulate command and control environments. We found that gestures can reduce distraction in such situations, are less interruptive to primary tasks during low-interruption notifications, are better for managing attention-utility tradeoffs, and support similar levels of reaction times and comprehension levels as touch interfaces. Gestures appear to be a better interaction choice for secondary notification tasks based on the tasks we evaluated.

Results highlight the importance of addressing different input modalities in the context of the IRC framework. Interaction design typically influences specifics of interfaces, such as their form and function; an area for future work is to investigate how gestures can be consistently woven into a more general theory of interaction design. One particularly striking aspect of gesture, for example, is that it can be employed in scenarios where there is no visible interface. Building on this work, we will continue to explore how gestures can optimally be used in secondary task contexts.

References

[1] Eyetoy web site. Electronic Resource, 2006.

[2] Sony play station. Electronic Resource, 2007.

[3] G. Abowd. Classroom 2000: An experiment with the instrumentation of a living educational environment. *IBM Systems Journal*, 38(4):508–530, 1999.

[4] B. P. Bailey and S. T. Iqbal. Understanding changes in mental workload during execution of goal-directed tasks and its application for interruption management. *ACM Trans. Comput.-Hum. Interact.*, 14(4):1–28, 2008.

[5] A. Blandford and W. Wong. Situation awareness in emergency medical dispatch. *International Journal of Human-Computer Studies.*, 61(4):421–452, 2004.

[6] C. M. Chewar, D. S. McCrickard, and A. G. Sutcliffe. Unpacking critical parameters for interface design: evaluating notification systems with the irc framework. In *DIS '04:*

Proceedings of the 2004 conference on Designing interactive systems, pages 279–288, New York, NY, USA, 2004. ACM Press.

[7] M. Czerwinski, E. Cutrell, and E. Horvitz. Instant messaging and interruption: Influence of task type on performance. In *OZCHI 2000*, pages 356–361, 2000.

[8] J. Fogarty, A. J. Ko, H. H. Aung, E. Golden, K. P. Tang, and S. E. Hudson. Examining task engagement in sensor-based statistical models of human interruptibility. In *CHI '05: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 331–340, New York, NY, USA, 2005. ACM.

[9] C. H. Ganoe, J. P. Somervell, D. C. Neale, P. L. Isenhour, J. M. Carroll, M. B. Rosson, and D. S. McCrickard. Classroom bridge: using collaborative public and desktop timelines to support activity awareness. In *UIST '03: Proceedings of the 16th annual ACM symposium on User interface software and technology*, pages 21–30, New York, NY, USA, 2003. ACM Press.

[10] J. Hare, M. Karam, P. Lewis, and m.c. schraefel. igesture: A platform for investigating multimodal, multimedia gesture-based interactions. Technical report, Electronics and Computer Science, University of Southampton, November 2005.

[11] M. Karam and m. c. schraefel. Investigating user tolerance for errors in vision-enabled gesture-based interactions. In *AVI '06: Proceedings of the working conference on Advanced visual interfaces*, pages 225–232, New York, NY, USA, 2006. ACM Press.

[12] D. McCrickard, R. Catrambone, C. Chewar, and J. Stasko. Establishing tradeoffs that leverage attention for utility: Empirically evaluating information display in notification systems. *International Journal of Human-Computer Studies*, 8(5):547–582, 2003.

[13] D. S. McCrickard, C. M. Chewar, J. P. Somervell, and A. Ndiwalana. A model for notification systems evaluationassessing user goals for multitasking activity. *ACM Trans. Comput.-Hum. Interact.*, 10(4):312–338, 2003.

[14] D. C. McFarlane. Comparison of four primary methods for coordinating the interruption of people in human-computer interaction. *Human-Computer Interaction*, 17:63–139, 2002.

[15] W. M. Newman. Better or just different? on the benefits of designing interactive systems in terms of critical parameters. In *DIS '97: Proceedings of the conference on Designing interactive systems*, pages 239–245, New York, NY, USA, 1997. ACM Press.

[16] S. Oviatt, P. Cohen, L. Wu, J. Vergo, L. Duncan, B. Suhm, J. Bers, T. Holzman, T. Winograd, J. Landay, J. Larson, and D. Ferro. Designing the user interface for multimodel speech and pen-based gesture applications: State of theart systems and future research directions. *Human-Computer Interaction*, 15:263–322, 2000.

[17] D. Salber and J. Coutaz. Applying the wizard of oz technique to the study of multimodal systems. In *EWHCI '93: Selected papers from the Third International Conference on Human-Computer Interaction*, pages 219–230, London, UK, 1993. Springer-Verlag.