Sensory Feedback and Movement on User Error Detection with a Touchscreen Mobile Device

[redacted] Paul Khawaja, [redacted]
University of Toronto, St. George
40 St.George st, Toronto, Canada
[[redacted] paul.khawaja, [redacted]]@mail.utoronto.ca

ABSTRACT
This paper seeks to build upon previous investigations on how auditory, visual, and haptic feedback influence user error detection when filling forms on a touchscreen mobile device. Specifically, our paper seeks to understand how feedback methods affect a form-filling task when the user performing the task is walking versus sitting. We had participants perform a series of simple captcha-typing tasks on an Android mobile device, and collected error-response time data as well subjective data from the trials. We hypothesized that haptic and visual feedback would be preferred when sitting, and that auditory and haptic would be preferred for when walking. Our experimental results conclude that users have no preference towards which feedback method they prefer whether sitting or walking, but our results also show that in terms of performance, haptic feedback is the most optimal for sitting and walking, since haptic feedback elicits the fastest error-response reaction from the user.

Author Keywords
Sensory feedback; mobile; haptic; auditory; visual; error detection; touchscreen; text entry; finger input.

INTRODUCTION
With the onset of touchscreen technology, mobile devices have seen increasing public adoption for use in performing tasks that were typically done using desktop computers. Using a touchscreen mobile device to perform tasks such as writing a form or web-browsing come with numerous advantages. Among these advantages are portability, ability to rely solely on fingers for input (as opposed to a mouse and keyboard), and having applications that can be configured to offer short transaction times by eliminating unnecessary buttons and tools conventional to desktop software.

While these advantages, among others, have led to widespread public adoption, there are still drawbacks, namely, in the form of sensory feedback limitations, which limit user interactions. First of all, they have small interface sizes, and hence, small keys, which, as found in an experiment by Lewis et al. [6], contributes to an increased error rate for touchscreen keyboards compared to desktop computer keyboards. Also, due to limited visual space and the use of fingers as input, the visual space can be occluded leading to error [2]. In addition to visual issues are tactile ones, as existing touchscreen devices do not offer natural haptic feedback like press-able buttons on a conventional keyboard do, which, as Barrettand et al. found [1], is responsible for decreased typing speed for touchscreen devices in comparison to non-touch devices. Finally, auditory feedback can be undesirable given that these devices are often used in public spaces where it is either too loud to detect or the sound can be disrupting to other people [5].

The issue of which feedback is best for performing specific tasks in specific environments is an ongoing issue in human-computer interaction. An experiment by Ghosal et al. [4] found that both auditory and tactile feedback on keystroke in a mobile device, when finger positioning was controlled, lead to increased typing speed. Another experiment by Dang et al. [3] found that visual coupled with haptic feedback decreased user errors in a tapping and sliding tasks, while visual coupled with auditory feedback increased user errors in a sliding task.

Our experiment builds on some of the issues highlighted by past research by investigating how auditory, visual, and haptic feedback can influence user error detection and correction when filling forms on a touchscreen mobile device. Rather than vary the participant's task performed on the device we asked how different movement tasks done while using a mobile device, namely, walking and sitting, are impacted by these 3 different feedback methods when filling in forms.. We measured the time it took participants to detect and respond to the stimuli to correct their error, as well as the total time it took to complete each task and the number of errors made, under different sensory feedback and sitting or walking conditions. Our research questions were: Which feedback method resulted in the fastest response to errors in the sitting condition? Which feedback method resulted in the fastest response to errors in the walking condition? Finally, which feedback methods did participants favor subjectively, based on our qualitative assessment following each task? We hypothesized that visual and haptic feedback will yield the faster responses in the sitting condition, whereas auditory and haptic feedback will yield faster responses in the walking condition. We also hypothesized that participants will favor their experience with visual feedback under all conditions.

METHODOLOGY
To help us analyze what sensory feedback methods result in the fastest error response time while walking or sitting, and to analyze which methods participants favor, we developed the following experimental methodology.

Participants

There were 25 volunteer participants, all between the ages of 19 to 26, 15 female and 10 male. All participants described themselves as experienced users of touchscreen mobile devices. None of the participants identified as having any sensory or motor disability that they could not mitigate with devices (for example, some participants had poor eyesight but mitigated the effect by wearing their glasses or contact lenses). All participants used two hands when texting on a mobile phone. Most participants were students from the University of Toronto, though some were from other universities or alumni.

Apparatus

All experimental trials were performed on an LG G3 android touchscreen mobile phone with 1440x2560 resolution. We developed an android software application to conduct our experiment. It opened first to a page for the experimenters to enter a unique ID for the participant, followed by 4 pages that consisted of the participant's tasks. Each page contained 6 text boxes to fill, with instructions above each box asking participants to fill it with a unique captcha (a string of 7 random numbers and alphabet characters). When filling the boxes, the characters would appear as dots, resembling typing a password. The first of the four pages did not give any sensory feedback when the participant entered a captcha that did not match what was requested. The remaining three were presented to participants in a random order: one with visual feedback (see figure 1), one with auditory feedback in the form of a beep, and one with tactile feedback in the form of a short vibration. For each page, the total time spent on the page is recorded, as well as the number of errors (incorrectly written captchas). For the pages with sensory feedback, error response time for each error is recorded by starting from when the stimuli is given and ends when the user taps on the text-box. All data is stored in plain-text on the phone's internal memory.

Participants were to complete these 4 pages twice, once while sitting, and once while walking in a fixed area. The fixed walking area was a 125x65cm table which participants were asked to walk around at a steady pace. For a qualitative assessment, we used a computer version of NASA-TLX developed by David Sharek [7], to analyze how participants perceived the workload of each task.

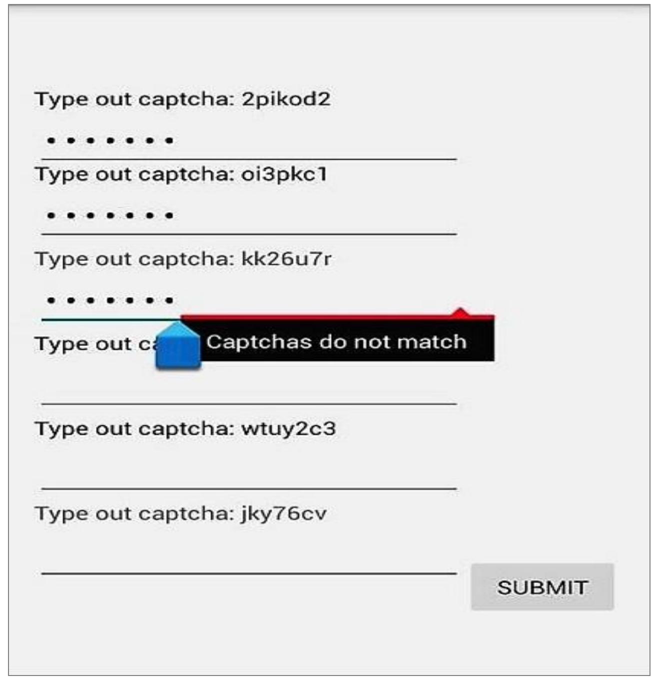

Figure 1. Example screen from the application. Participants tap on each of the 6 textbox lines to fill in the captcha as it is written above. These captchas are randomly generated 7-character strings with alphabet or number symbols. When participants enter the captcha they appear as dots. In this example, a participant as incorrectly copied "kk26u7r" and has received visual feedback to indicate that they have made an error.

Procedure

Each experiment was conducted with one participant at a time. Before beginning the experiment, participants were given a detailed informed consent form to read and be guided through in order for them to understand what is expected if they agree to participate. If participants agreed, they were asked to fill in a pre-questionnaire (on paper). The purpose of this questionnaire was to provide us with demographic information as well as to reference it in case any participants produced data outliers. The questionnaire asked participants for their name, age, gender, and the following more experiment-specific questions:

- On a scale of 1 to 5, rate your proficiency with touchscreen mobile devices.
- Do you most often use 1 or 2 hands to text when using your phone?
- Please describe any disabilities/limitations you may have with vision, hearing, tactile sensation, or fine-motor abilities. Are any of these limitations mitigatable through basic devices (e.g. eyewear, ear-piece, etc. . .)? Be as specific as possible.

Participants were told that they could ask if they needed any clarifications on any of the questions. Reading and signing the consent form and filling the pre-questionnaire took no more than 10 minutes per participant.

Upon a participant finishing the pre-questionnaire, we entered a unique participant ID into the application on the

LG G3 phone and handed it to the participant. The participants were reminded, as per the information on the consent form that they will be filling in text-box captchas. They were told that the application may notify them when they make an error, and that they should try to fix it immediately. They were also instructed to fill in the forms as fast as possible. Some participants began with the walking condition, and were asked to walk around our designated walking space (as described in the apparatus section, a 125x65cm table) at a steady pace. Others were instructed to do the sitting tasks firt. As per the details in the apparatus section, both the walking and sitting conditions consisted of 4 screens. The first screen for both conditions did not give any sensory feedback when participants made errors. As the participants were not notified of their errors, data on error-detection time (collection described in apparatus section) was not collected here. Following the first screen were 3 screens each presenting another 6 text-boxes to fill with 7-character captcha strings (see Figure 1). One screen gave participants auditory feedback (a short beep) when they made an error, another screen gave participants tactile feedback upon making an error (a short vibration), and another screen gave visual feedback indicating that they made an error (see Figure 1). These three screens were given to participants in a random order from the application itself.
After completing each of the feedback screens (i.e. not including the first screen), participants were asked to fill in post-questionnaire on our laptop. This questionnaire was a computer-based NASA-TLX [7] workload assessment tool. We entered the participant ID along with the type of task condition the participant just completed, and had the participant answer the form. The purpose of the NASA-TLX form was to help us evaluate how the participants perceived the workload and general experience with each feedback method for both walking and sitting conditions. The assessments asked participants to rate the tasks mental demand, physical demand, temporal demand, performance, effort, and frustration (see Figure 2). The software recorded this data specific to each participant ID and task ID and generated an overall assessment score. Once they completed the NASA-TLX assessment they were given the phone back to begin the next task (and subsequent NASA-TLX questionnaire). Once participants finished all screens for the sitting or walking condition (whichever they began with) they were instructed to do the same with the condition they did not do yet. When participant finally completed their last task and submitted the final NASA-TLX assessment, they were offered a debriefing of the experiment. This experimental process overall took about 20 minutes for each participant. Throughout the experimental trials, we recorded any note-worthy observations about the participants (such as whether their walking pace was irregular, or if they used only a single hand to type, etc. . .).

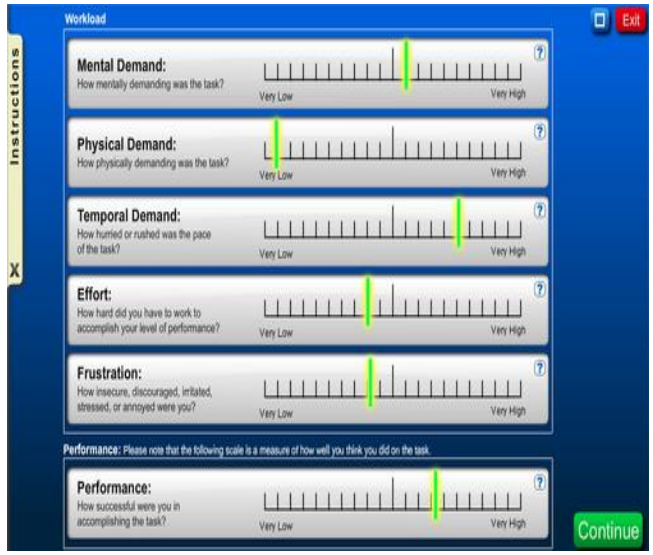

Figure 2. Screenshot of NASA-TLX 2.1.2 computer-version by David Sharek [7]. Participants were asked click on the slider for each section so we could evaluate how participants perceived the workload for each task. The instruction panel could be expanded so participants could read more elaborate information on what each section was asking.

Design

The independent variables of the experiment were feedback mode (haptic, visual or auditory) and participant's interaction with their immediate environment (walking or sitting). Error response time, as well as participant's perception of task workload (evaluated using NASA-TLX [7]), were the conditions that these independent variables were subject to. The experiment was within-subjects, since we aimed for the participant's to have a perspective of the workload for each task that stems from comparing different feedback methods to each other. Within-subject design was also necessary in order to collect meaningful error-response time data to compare against the different feedback trials, since some participants have generally faster reaction times than others, and this would cause significant result variation due to participants. There were a total of 8 experimental conditions each participant did:

- Filling in the form while sitting, with no sensory response to error making.
- Filling in the form while walking around a table, with no sensory response to error making.
- Filling in the form while sitting, with auditory response to error making
- Filling in the form while walking, with auditory response to error making
- Filling in the form while sitting, with haptic (vibrational) response to error making.
- Filling in the form while walking, with haptic (vibrational) response to error making.

- Filling in the form while sitting, with visual response to error making.
- Filling in the form while walking, with visual response to error making.

The purpose of the first two condition (no sensory response) was to provide participants with a brief period to familiarize with the application tasks, and also to provide us as researchers with a reference for total task completion time and number of errors made in comparison to the sensory feedback conditions. The order of the sensory condition tasks (following the no-sensory response task) were also randomized within both the walking condition and sitting condition. Of the 25 participants, 13 began with all the sitting conditions, and the remaining 12 did the walking ones first. This randomization was used to prevent any learning bias that could result from within-participant data had the order not been randomized.

Measures

Error response time and participant's perception of workload were measured against the independent variables of feedback method and participant's interaction with immediate surroundings. Our research question sought to investigate the optimal feedback response method for form-filling tasks on mobile devices, where the *optimal* method is determined as the feedback method that elicits the quickest response from the participant, and is also deemed as non-intrusive, intuitive and preferred by the participant. Error response time was thus measured as the time taken for the participant to, after receiving feedback indicating them of an error, to refocus on the field containing the error in order to fix the error. Qualitative data was gathered through observation and post-experiment NASA-TLX questionnaires.

RESULTS

In order to analyze the optimal feedback method, the application recorded the overall average amount of time, in milliseconds, it took each participant to respond to errors they made in each task (if they made any errors). It also recoded the overall time each participant took, and the number of errors each participant made. The average response time, total time, and errors made across all participants can be seen in figures 3, 4, and 5 below respectively.

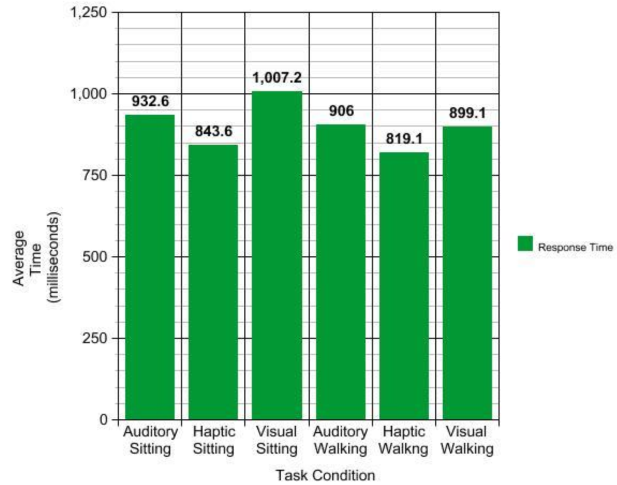

Figure 3. Average response times (in milliseconds) across participants for each task condition. Few participants did not make any errors in some tasks and made errors in others. Error reaction time averages were calculated in each condition based on the participants who had made errors.

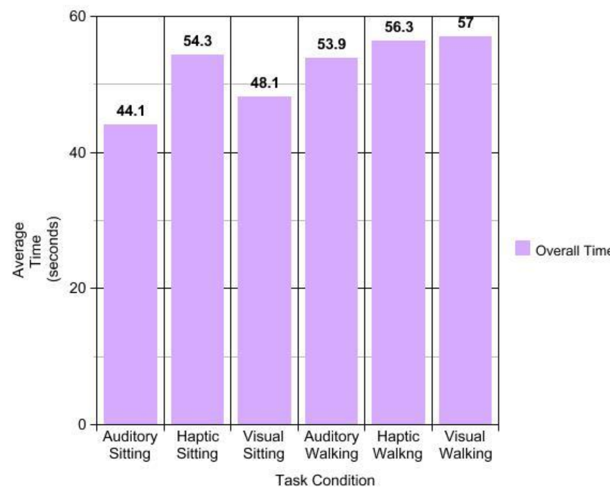

Figure 4. Average Overall Time it took for participants to complete each task. Unlike response time, all users were included at every level of analysis as they all had an overall time spent performing each task.

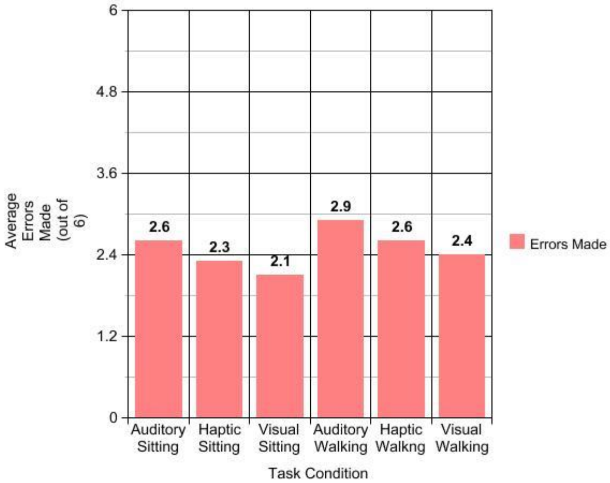

Figure 5. Average Overall Errors made by each participant. Participants were asked to enter captchas into textboxes as fast as they could. Each task had 6 7-character long captchas to kill in, so the average is out of 6 possible errors in total.

According to the data in figure 3, haptic feedback seems to have given the shortest (and therefore, optimal) response time of all the conditions. In order to analyze whether this difference is significant or not, we performed a repeated-measures ANOVA test with a significance level α of 0.05. We chose a repeated measures ANOVA as we used a within-participant design and have two independent variables being measured. For the overall analysis, the F_{crit} value for our data was 2.2899, and the F value from our analysis was 2.91, with a p value of 0.023. Since of F value is less than the F statistic and the p value is less than α, the overall Anova test yielded statistically significant results ($F_{5, 120} = 2.2899$, $p < 0.5$). Therefore, at least two conditions show a significant difference. There were also a few statistically significant pairwise comparisons of mean differences (where $p < 0.5$). All non-haptic conditions had a significantly higher average error response time than both haptic walking and haptic sitting conditions. The visual sitting condition had a significantly higher average error response time than all conditions except auditory sitting. For our analysis of qualitative data, we chose the NASA-TLX assessment to measure how user perceive the overall "task load" or, in other words, the burden of performing each task. The computer-version of NASA-TLX stored data for each participant on how they rated each of 6 workload categories and calculated a "task load index" for each based on the categories [7]. Figure 6 shows the averages across all 25 participants for each task load category for each test condition.

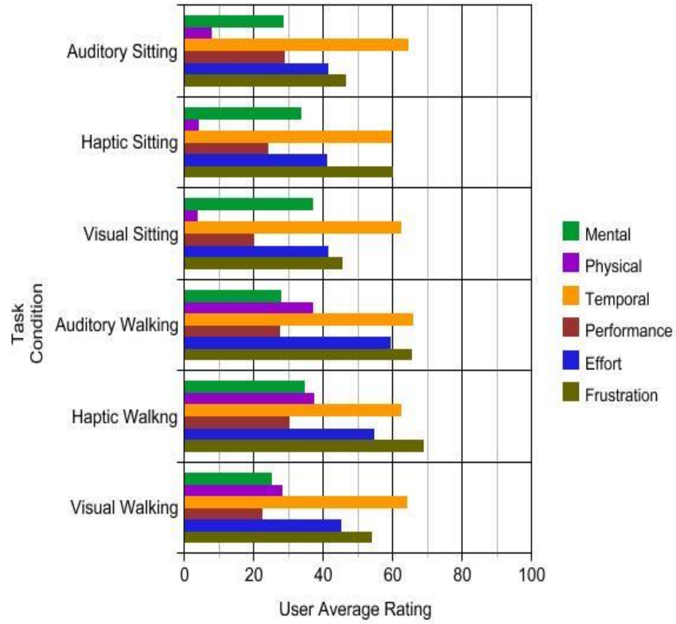

Figure 6. NASA-TLX averages participant rating for each task load category for each task condition. Participant rated the task's mental demand, physical demand, temporal demand (how much time it took or perceived time constraint), how the perceived their overall performance, the amount of effort the task took, and their sense of frustration.

Figure 7 shows the averages of the task load index for each condition.

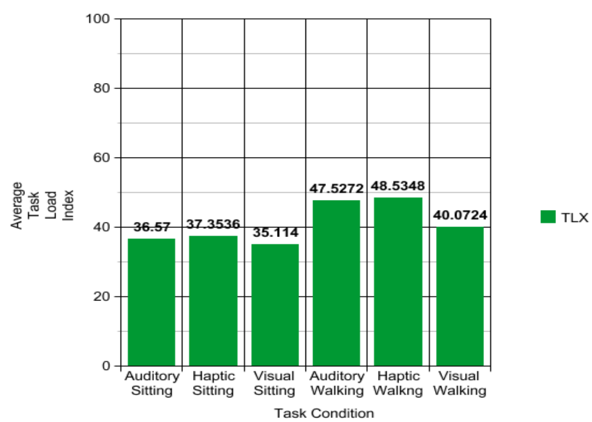

Figure 7. Average Task Load Index for each experimental condition. The TLX is calculated by the NASA-TLX software and is a weighted value based on the different categories [7].

Under both walking and sitting conditions, visual feedback seems to have the best results, more significantly so in the walking condition. In order to analyze the significance of this data, we performed a Friedman statistical analysis on the ranked "task load index" data generated by the NASA-TLX software for each condition [7]. As the NASA-TLX data is also within-participants and has the same set of conditions as the error detection data set, we opted to use the non-parametric equivalent to the previous data set's repeated measure ANOVA, which is the Friedman test.

We selected our significance level α = 0.05. For the overall analysis, the F_{crit} value for our data was 2.2899, and the F value from our analysis was 5.441, with a p value of 0.00015. Since of F value is less than the F statistic and the p value is less than α, the overall Friedman test yielded statistically significant results ($F_{5, 120} = 5.44$, $p < 0.5$). Therefore, at least two conditions show a significant difference. Multiple comparisons analysis required a mean rank difference of at least 0.9648 for significant difference between conditions (see figure 8 for mean rank difference per condition). Based on this data, the auditory sitting and

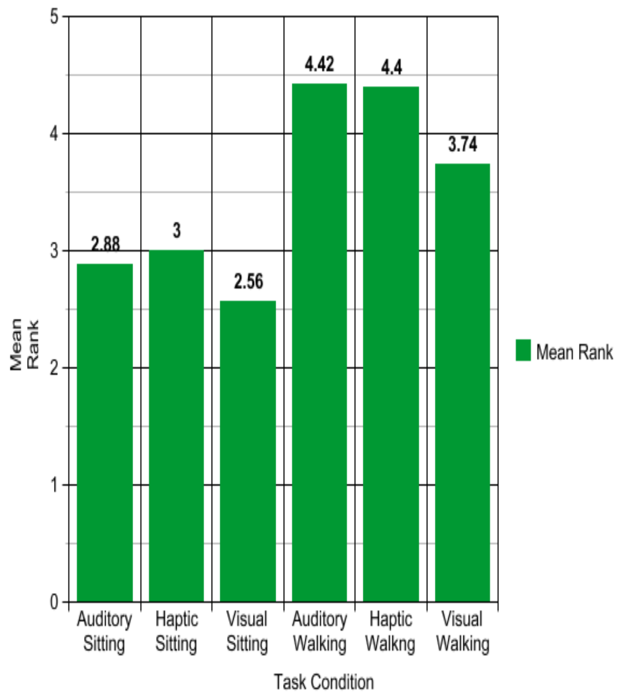

Figure 8. Friedman Test multiple comparison analysis. Each bar represents the mean rank of each condition. A statistically significant difference between ranks must be at least 0.9648.

haptic sitting conditions ranked significantly lower than both the auditory and haptic walking conditions, but not the visual walking condition (i.e. participants rated task load of the sitting conditions significantly lower). Task load of the visual sitting condition was rated significantly lower than all three walking conditions.

Finally, we performed another ANOVA significance test to compare the mean errors made between the experimental conditions and the no-feedback condition. The results did not yield anything significant. We did not analyze the overall time between no-feedback and feedback conditions as we believed any significance would be due to the time it would take for participants to correct the errors they made.

DISCUSSION

The NASA TLX results conclude that participants showed no significant preference for any feedback, regardless of the whether they were filling the forms out while walking or while sitting. The walking trials perceived workload was significantly higher than sitting trials, but the results show no significant preference for any feedback type within the sitting or walking trials. The significant difference between walking trials and sitting trials must thus be attributed to the added physical workload that walking demanded of the participants. The reason for this is perhaps due to the overall few errors participants made on average, which could have resulted in their perspective being mostly shaped by the task of filling in the forms and not the actual stimuli, since participants had very few experiences with each sensory feedback method.

Error-response time data analysis shows that haptic feedback yielded most significant results across walking and sitting trials. This result is partially in line with our hypothesis and also aligns with similar research done by Hoggan, et al.[3], which concludes that haptic feedback, although limited in the range, elicits an immediately perceivable feedback to the user which thus translates to the quickest reactionary response from the user [5]. In this case, the reactionary response being re-focusing on the text field with an error in order to correct it. A possible reason why we did not see the expected results for auditory and visual feedback could be due to the way they were implemented. Auditory feedback, as noted by Hoggan et al [5], could be disturbing to users, and we observed that our participants often had an initial shock reaction to the beeps which may have disrupted their reaction time. The visual cues may have not resulted in fast reaction times as they take longer to appear and users had to read the cue.

CONCLUSION

At the start of our current experiment we hypothesized that users would favor the task load of the visual feedback to others, and that the optimal feedback method that would produce the shortest error-response time would be haptic feedback as well as visual feedback (roughly tied). Our analysis yielded no significant data from qualitative analysis about the preferred feedback method, so we cannot draw conclusions about which method is best preferred by users. However, we did find that haptic feedback was in fact the optimal feedback method and produced the shortest error response time, however, visual was significantly worse. While what we hypothesized may not be completely in line with our findings, we believe this data can provide insights into future research. However, any future research should attempt to alleviate at least some of the limitations that our methodology had. First of all, we were unable to design an application that encouraged or heightened the amount of errors produced, as participants overall averaged making an error less than half the time across all conditions. This resulted in collecting far less error response data than we would have wanted. Additionally, the walking trial was an attempt to mimic a user walking in real life, however, it was far too controlled to really replicate such a scenario as users simply paced around a fixed and safe area with little interference or distraction that could result in a user having difficulty detecting certain types of feedback while finding other types easier to detect. Any future research could attempt to mitigate these limitations by developing more error-prone software, targeting a much less experienced audience than university students, as well as attempting to replicate a less controlled environment to see how distraction plays a role in error response.

REFERENCES

1. Barrettand J., Krueger H., Performance Effects of Reduced Proprioceptive Feedback on Touch Typists and Casual Users in a Typing Task, *Behavior and Information technology*, *13*, 1994, 373- 381.
2. Brewster, S., Overcoming the Lack of Screen Space on Mobile Computers, *Personal and Ubiquitous Comput, 6*, 2002, 188-205.
3. Dang, N. T., Perrot, V., Mestre, D., Effects of Sensory Feedback While Interacting with Graphical Menus in Virtual Environments. *IEEE Virtual Reality Conference – VR 2011*, (Singapore: IEEE, 2011), 199-200
4. Ghosal, S., Shringi K., Rattan, P., Gudhainiya, K., Verma, A., Comparison and Evaluation of Factors Affecting Typing Speeds on Touchscreen Keyboards, *Intl Journal of Mechanical Engineering and Technology, 6*, 2015, 61-72.
5. Hoggan, E., Crossan, A., Brewster, S., Kaaresoja, T., Audio or Tactile Feedback: Which Modality When? *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems – CHI 2009*, (New York: ACM, 2009), 2253-2256.
6. Lewis J. R., Potosnak K. M., Magyar R. L., Keys and Keyboards, *Handbook of Human Computer Interaction*, (M. Helander and T.L.P. Prabhu, Eds.). Rochester, NY: Elsevier Science, 1997, 1285-1315.
7. Sharek, D. *NASA-TLX 2.1.2*, Psychology Department, North Carolina State University, Raleigh NC, 2011.