Multimodal Feedback for Finger-Based Interaction in Mobile Augmented Reality

Wolfgang Hürst[1]
[1] Department of Information & Computing Sciences
Utrecht University, Utrecht, The Netherlands
huerst@uu.nl

Kevin Vriens[1,2]
[2] TWNKLS
Rotterdam, The Netherlands
kchj.vriens@gmail.com

ABSTRACT
Mobile or handheld augmented reality uses a smartphone's live video stream and enriches it with superimposed graphics. In such scenarios, tracking one's fingers in front of the camera and interpreting these traces as gestures offers interesting perspectives for interaction. Yet, the lack of haptic feedback provides challenges that need to be overcome. We present a pilot study where three types of feedback (audio, visual, haptic) and combinations thereof are used to support basic finger-based gestures (grab, release). A comparative study with 26 subjects shows an advantage in providing combined, multimodal feedback. In addition, it suggests high potential of haptic feedback via phone vibration, which is surprising given the fact that it is held with the other, non-interacting hand.

CCS Concepts
• **Human-centered computing~Mixed / augmented reality**

Keywords
Handheld AR; AR interaction; multimodal feedback.

1. INTRODUCTION
Modern smartphones offer the opportunity to create simple, yet powerful augmented reality (AR) where the video stream of the away facing camera creates a live snapshot of the user's surrounding world (representing reality) and enriches it with superimposed graphics in real-time (representing an augmented reality). Yet, interaction in such a setup remains cumbersome; for example, touch screens are small, only allow operations in 2D, and your finger covers large parts of the actual scene during interaction. Researchers have therefore started exploring the usage of finger tracking for mobile AR interaction. Tracking the motions of one's fingers in front of the mobile's camera and interpreting them as input gestures enables users to directly interact with the AR scene. While at first sight, this resembles a more natural, realistic interaction, problems occur when trying to touch and manipulate the superimposed virtual graphical objects. A lack of haptic feedback makes interaction appear unreal and adds uncertainty ("Did I touch it now or not?"). In this research, we explore the potential of using different modalities, in particular sound, visual feedback, and haptic in form of vibrations of the phone, to improve finger-based interaction in a mobile AR setting. After addressing the general context in Section 2, we describe our scenario in Section 3 and experiment design in Section 4, the present and discuss the results in Section 5, before concluding in Section 6.

ICMI'16, November 12–16, 2016, Tokyo, Japan
© 2016 ACM. 978-1-4503-4556-9/16/11...$15.00
http://dx.doi.org/10.1145/2993148.2993163

2. CONTEXT AND RELATED WORK
Common approaches for AR interaction include tangible user interfaces (UIs) and freehand gesture-based interaction. With tangible UIs, physical objects from the real environment (e.g., cups [10] or cards [9]) are recognized by the AR system and can be used to manipulate virtual parts of the AR environment – thus providing a bridge between the "touchable" physical and abstract virtual world. Direct manipulation of virtual objects via, for example, finger or hand tracking suggests a more natural, real-world like interaction but lacks this "feeling of touch". In a mobile context, utilization of the touch screen is also commonplace, yet, suffers from issues, too – such as occlusion of the screen and ergonomics [7]. Researchers therefore started exploring finger tracking for handheld AR interaction as well (e.g., [5,6,7]; see [13] for an overview of different scenarios, including but not limited to handheld AR). When comparing touch versus finger tracking, Hürst et al. [7] showed that the latter often suffers in performance, likely due to a lack of haptic feedback. In particular, this lack produces a feeling of uncertainty if virtual objects have been touched or not (or if this touch has been recognized by the system or not), which in turn has a negative impact on interaction time. Multimodal feedback provides a means to deal with this problem. For example, Chang et al. [4] state that "multisensory presentations may be effective measures to provide feedback" in the context of handheld AR games. Sound and visuals are obvious choices applicable to a handheld AR scenario. Haptics in AR is often provided via gloves [3]. Such sophisticated solutions requiring additional hardware seem unsuitable in many handheld AR setting relying on mobile phones. Unfortunately, at the time being, the only means of tactile feedback provided by such devices are integrated vibration motors commonly used for notifications and alerts. Therefore, they cannot provide direct feedback at the location of touch, but only remote one on the other hand holding the phone. Richter et al. [12] evaluated the benefit of both direct and remote haptic feedback in context with interactive surfaces. Their work suggests that the latter can still provide a benefit, and thus served as a motivation for our research, i.e., investigating if such a remote vibration feedback via the handheld phone can be beneficial in finger-based mobile AR interaction. In particular, we are interested in comparing three modalities: vision via the phone's display, audio via its speakers, and haptics via phone vibrations (cf. Fig. 1).

Figure 1. Modalities evaluated in our study.

3. SCENARIO AND IMPLEMENTATION

Hürst et al. [7] identified gestures comparable to a board game setting on a table as most suitable for finger-based mobile AR interaction. They also evaluated different types of gestures, with simple grab operation utilizing two fingers (thumb and index finger) as intuitive and appropriate (Fig. 2). Uncertainty due to lack of haptic feedback mostly comes into play when grabbing and releasing an object. Thus, in this work, we are focusing on two basic gestures that serve as building blocks of more complex ones: selection via grabbing and deselection via releasing, using the two-finger-gestures illustrated in Fig. 2 (steps 1&4).

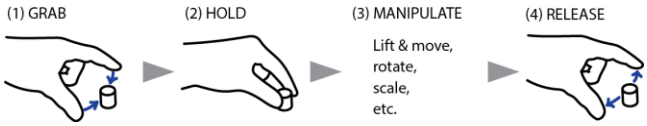

Figure 2. Basic finger-based interaction gestures.

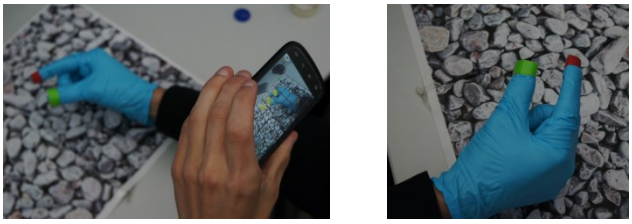

Figure 3. Setup and markers used in the evaluation.

For the evaluation, we implemented a simple setup using the Qualcomm AR SDK (now Vuforia), an AR library for natural feature tracking (Fig. 3). For finger tracking, we used one marker on the tip of the thumb and index finger, respectively. Subjects were asked to wear a blue medical glove in order to improve tracking accuracy by creating higher contrast in the images. Despite recent improvements in marker-less tracking (see, e.g., [1] and [2] for examples in handheld and non-handheld AR scenarios) we purposely decided for such an artificial setup. Using a simple, but robust color tracking eliminated possible influences of noise or inaccurate tracking results. Thus, we assume that our results can be applied to any reliable tracking mechanism implemented on mobile phones. Likewise, gestures were recognized via a simple thresholding approach, where grab and release actions are recognized by both color markers entering or leaving a bounding box around the object. Tests were done with a Google Nexus S smartphone featuring a 4 inch, 800×480 pixels screen and a Linear Resonant Actuator (LRA) as vibration unit. Given the pilot study character of our work, we purposely opted for such rather simple, but common specifications to get more general results applying to multiple setups. Further studies should include more complex scenarios, for example, with advanced technologies (e.g., piezoelectric actuators) that might become more common in future generations of phones.

Virtual objects are integrated into the AR environment in the form of yellow barrels on a grid that was aligned with the real world markers placed on the table (Fig. 4). Our goal was to evaluate the potential of all three kinds of multimodal feedback such a setup can provide: visuals, sound, and haptic via vibration of the phone. Visual feedback was implemented via a small bounding box that appeared once an object was selected (Fig. 4, right). Audio was provided via neutral standard beeps from the phone. For haptic feedback, the integrated vibration unit was used. In all three cases, feedback was either constant, i.e., started with a detected select gesture and ended with a detected release gesture, or temporary, i.e., only active for 500 milliseconds. In case of audio, this means that three beeps were played during this time interval.

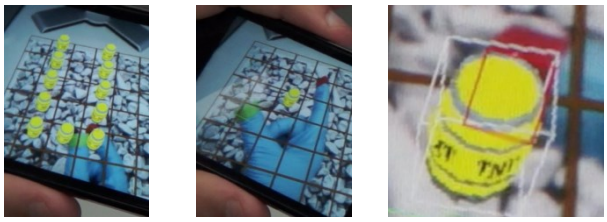

Figure 4. Setups used in experiments and visual feedback.

While we can generally expected that feedback has a positive effect on interaction, the impact of individual modalities is unclear. Visual feedback is generally the standard in such scenarios and thus might be considered most intuitive. Yet, it does not resemble a natural realistic situation and, in particular on small displays, might easily get overlooked. Similarly, audio feedback is well-known and established in general human-computer interaction, but does not resemble a natural situation and, especially when provided only temporary, might get missed. Haptic is the only kind of feedback that generally appears in a comparable real life situation. Yet, the implementation via vibration on the phone is neither realistic nor common. Most importantly, in this setup, it is not provided at the actual point of interaction, but remotely at the hand holding the phone. The related slight shaking of the device might also have a negative impact on the recognition of visual feedback and user comfort.

4. EXPERIMENT DESIGN

In our experiment, we focus on select and deselect gestures (cf. "grab" and "release" in Fig. 2), first, because these are the most basic ones and building blocks of more complex interactions. Second, they are the most likely to benefit from additional feedback, especially with respect to uncertainty ("Did I grab/release it yet?"). We can split these two basic operations into (a) the moment a selection is recognized by the system, (b) the moment that the user realizes that the object is selected, (c) the moment a deselection is recognized by the system, and (d) the moment that the user realizes that the object is deselected. By providing feedback, we aim at reducing the time intervals between (a)-(b) and (c)-(d). In addition to such performance improvements, we are interested in the qualitative experience, which is partly influenced by performance (e.g., people feel more confident), and partly subjective (e.g., people like certain feedbacks more or less).

To investigate such quantitative and qualitative influences, we set up two experiments. The first one purely focused on selection. Users were asked to select eleven virtual objects shown on the table. No specific order was given, because searching for the next one would have impacted interaction time. Instead, they were arranged in a U-shape (cf. Fig. 4, left) and participants were asked to perform this task as quickly as possible, thus resulting in an

obvious order and similar distances between two selection steps. Users had to do this test eight times, once for each feedback type: none, audio, visual, haptic, three pairwise combinations, and all three together. Because there was no deselection, feedback was only provided temporarily for 500 msec.

In the second experiment participants had to select and deselect a single object eleven times (cf. Fig. 4, center). Provided feedbacks were similar as in the first one, but this time also included constant feedback between selection and deselection in addition to the previously used temporal feedback of 500 msec, resulting in 27 different feedback/duration options.

Experiments took place in a neutral room with a test person that instructed the subjects, interviewed them, and took notes during the tests, but did not interfere in any way during the actual tasks. Quantitative data was gathered via logging on the phone. Possible outliers in the data were removed before the analysis using the Median Absolute Deviation method. Qualitative information was gained via questionnaires, an informal discussion at the end, and observations made by the test person. Each evaluation started with a training session where subjects saw three virtual objects placed next to each other. Each provided a different type of feedback modality (audio, visual, and haptic, respectively). They were instructed on how to do the gestures and then had to perform them several times to understand and gain experience with the respective feedback types.

A total of 26 subjects took place in the two experiments. Tests were done anonymously. Participants were students from the local computer science program ages 21 to 30 (average 24, standard deviation 2.69) with 25 males and only one female. We decided to go for such a specific user group to gain higher statistical power for this particular subset, which also represents early adopters and thus target audiences of the tested technologies. Evaluations for other populations are an interesting aspect to address in future work. Due to the basic characteristic of the task, we do not expect a gender bias, and thus did not aim for a gender balance.

5. EVALUATION AND DISCUSSION

In both experiments, we opted for eleven virtual objects, so we can measure ten individual interactions, i.e., logging of time on the phone started after the first object was selected. Fig. 5 illustrates the times between two selections for experiment 1, averaged over subjects and selected objects. The dark colored column on the left represents the case of feedback from all modalities. Medium dark colors show pairwise combinations of modalities, and light ones illustrate a single modality feedback. While the experiment hypothesized an equality of the means for different conditions, it was expected that more modalities lead to the desired decrease in reaction time. A one-way repeated measures ANOVA with a Greenhouse-Geisser showed a statistical significance ($F(2.628, 63.077) = 10.688$, $p<0.05$). A Bonferroni post-hoc test showed that the "triple modality" feedback as well as the pairwise feedback options were all significantly faster than 'no feedback' and pure visual feedback ($p<0.05$). Pure haptic feedback proved to be significantly faster than pure visual one ($p<0.05$). While the positive result for the multimodal cases are kind of expected and what we were hoping for, the outcome for feedback with a single modality comes a bit surprising. While there is not much of a difference between audio and no feedback, visual was much slower than no feedback at all. A possible explanation could be that it got sometimes overlooked and therefore actually added to the level of uncertainty instead of decreasing it. In addition, it is known from literature that humans react faster to sound than light [11], which could explain the difference between audio and visual feedback. Noteworthy though is the relatively good performance of pure haptic feedback, especially compared to the other singular modality feedbacks.

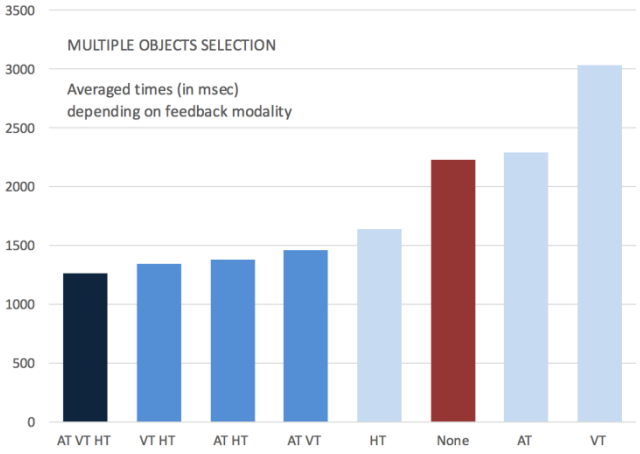

Figure 5. Experiment 1 (multiple objects selection): times (in msec, averaged over all subjects and tasks) depending on feedback modality (A/V/H = audio/visual/haptic) and implementation (T = temporal feedback for 500 msec).

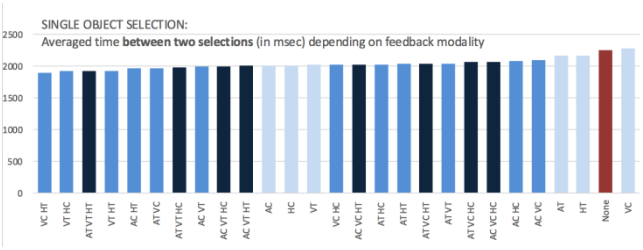

Figure 6. Experiment 2 (single object selection & deselection): times (in msec, averaged over all subjects & tasks) depending on feedback modality (A/V/H = audio/visual/haptic) and implementation (T = temporal / C = constant feedback).

Unfortunately, such a clear trend cannot be observed from the corresponding average times in experiment 2, where interaction included selection and deselection of an object (Fig. 6). Likewise, although multiple modalities often performed faster, no general conclusion can be made here (cf. Fig. 6, color coding as in Fig. 5, i.e., dark blue = three modalities, blue = two modalities, light blue = one modality, red = none). Not surprisingly, a one-way repeated measure ANOVA analysis did not reveal any statistical significance. Yet, a direct comparison of the best performer, i.e., the visual-haptic combination with constant visual and temporary haptic feedback (VC HT) with the no feedback case showed a significant difference using the Wilcoxon signed ranked test.

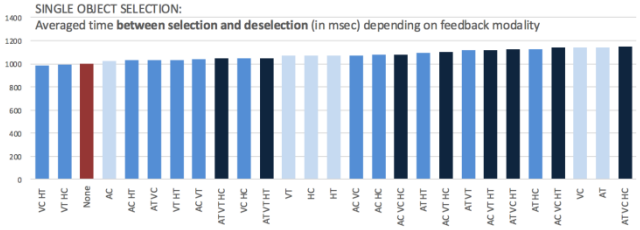

Figure 7. Exp. 2: Times between selection & deselection.

To further investigate this result, we split times in intervals (a)-(b), i.e., the time between a selection and deselection, and (c)-(d), i.e., the time between a deselection and selection (cf. first paragraph in section 4). Fig. 7 shows the results for (c)-(d), which again do not show a trend in favor of any kind of feedback.

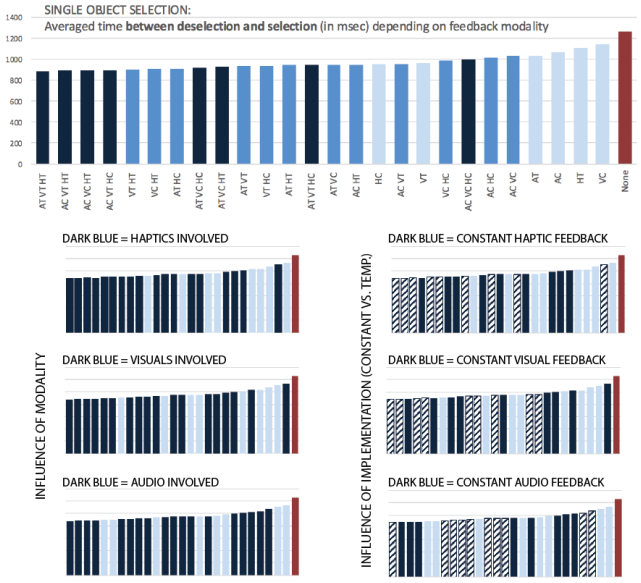

Figure 8. Exp. 2: Times between deselection & selection.

Considering that the deselect action relates to the gesture where users just have to put their fingers apart, and thus do not really need that much confirmation, this result does not come surprising. And indeed, the other time interval, i.e., the time between deselection and selection does show a similar trend as in the first experiment; more modalities generally result in faster performance (Fig. 8, top). The diagrams on the left below represent the very same data, but color encoded to illustrate the influence of different modalities. We see that haptics (encoded dark blue in the first one) in most cases contributes to a better performance. For visuals (encoded dark blue in the second one), this trend is still existing but less distinct. For audio on the other hand (encoded dark blue in the last one), there is hardly any trend recognizable. The three diagrams on the right show the same data as the ones on the left, but also illustrate the difference between constant feedback (dark blue) and temporary feedback (dark blue line pattern). With the exception of audio, it suggests a minor trend in favor of temporary feedback.

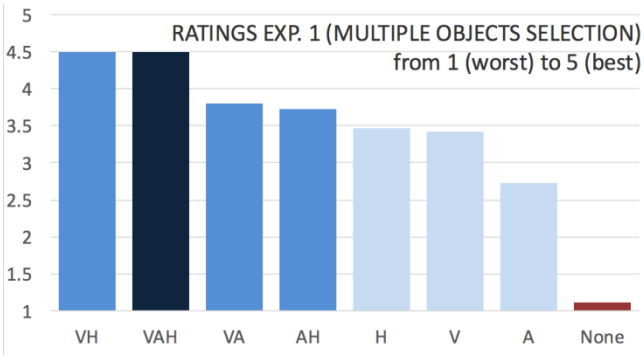

Figure 9. Subjective ratings for modalities in experiment 1.

In addition to these quantitative results, user experience is another important aspect of any kind of interaction. Fig. 9 and 10 show the ratings given by the subjects for modalities in each experiment on a five-point Likert scale (with 1 being worst and 5 being best). Results for experiment 1 are in line with the performance observations. Subjective judgements for experiment 2 show a similar trend although not as distinct. It seems noteworthy though that pure haptic feedback was actually preferred over the two options with two-modal feedback that did not include haptics.

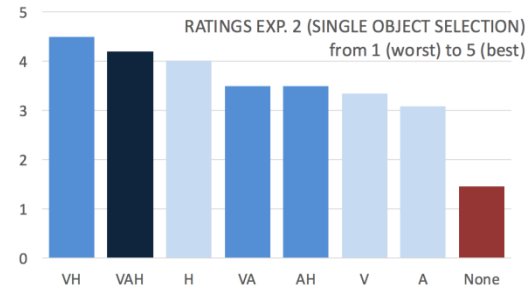

Figure 10. Subjective ratings for modalities in experiment 2.

As expected, in the informal interviews, subjects often said that a constant visual feedback should be given, likely because this is in line with common approaches and graphical user interface design. Additional feedback was appreciated and described as useful by many, for example, in case the visual feedback gets missed or is hard to see in a specific scene. Audio was generally less appreciated, and especially in case of constant feedback sometimes even considered annoying. Some also characterized it as unnatural, since grabbing objects in the real world usually does not make a sound either. Interestingly though, others characterized both visual and haptics as natural and adding to the 'realness' of the experience, which is technically not true; neither is there a natural equivalent to highlighting the touched object nor does the remote haptic feedback on the phone resemble any realistic situation. Yet, for some users it did feel that way. The ratings of experiment 2 (Fig. 10) combine both implementations; constant and temporary feedback. When asked about these options explicitly, constant audio feedback not preferred, nor was constant haptic feedback. Visual feedback on the other hand was considered helpful when displayed permanently.

6. CONCLUSION

In this paper, we presented an initial user study investigating different types and implementations of multimodal feedback for finger-based interaction in mobile augmented reality. An evaluation with a basic interaction task showed that multimodal feedback was not only preferred by users but has the potential to improve interaction speed. In particular, constant visual feedback as standard method can benefit from haptic feedback at the begin and end of an action (e.g., in our case selection and deselection). The benefit of haptic feedback is particularly interesting and noteworthy because it did not, as one would commonly expect, appear at the location of the actual action, but remotely on the other hand by vibration of the phone. Adding this simple, yet effective type of feedback seems promising and worth further investigation. Interesting aspects to study include variations of the vibration signal (duration, intensity, etc.) and if these can be recognized by a user and used in a beneficial way for interaction design. The fact that some users characterized it as natural despite the remote location suggests further potential with respect to user engagement. Finally, although we expect our results to generalize to more complex gestures, additional studies, also in relation to a concrete application case, are worth pursuing.

7. REFERENCES

[1] H. Bai, L. Gao, J. El-Sana, and M. Billinghurst. 2013. Markerless 3D gesture-based interaction for handheld augmented reality interfaces. In *2013 IEEE International Symposium on Mixed and Augmented Reality (ISMAR),* pp. 1-6, IEEE.

[2] M. Bikos, Y. Itoh, G. Klinker, and K. Moustakas. 2015. An interactive augmented reality chess game using bare-hand pinch gestures. In *2015 International Conference on Cyberworlds (CW)*, pp. 355-358, IEEE.

[3] V. Buchmann, S. Violich, M. Billinghurst, and A. Cockburn. 2004. FingARtips: gesture based direct manipulation in Augmented Reality. In *Proceedings of the 2nd international conference on Computer graphics and interactive techniques in Australasia and South East Asia* (GRAPHITE '04), Stephen N. Spencer (Ed.), pp. 212-221, ACM, New York, NY, USA.

[4] Y. N. Chang, R. K. C. Koh and H. Been-Lirn Duh. 2011. Handheld AR games – A triarchic conceptual design framework. *2011 IEEE International Symposium on Mixed and Augmented Reality - Arts, Media, and Humanities*, Basel, 2011, pp. 29-36, IEEE.

[5] Wendy H. Chun and Tobias Höllerer. 2013. Real-time hand interaction for augmented reality on mobile phones. In *Proceedings of the 2013 international conference on Intelligent user interfaces* (IUI '13), pp. 307-314, ACM, New York, NY, USA.

[6] K. Dorfmuller-Ulhaas and D. Schmalstieg. 2001. Finger tracking for interaction in augmented environments. In *Proceedings IEEE and ACM International Symposium on Augmented Reality, 2001*, pp. 55-64, IEEE.

[7] W. Hürst and C. van Wezel. 2013. Gesture-based interaction via finger tracking for mobile augmented reality. *Multimedia Tools and Applications*, Vol. 62(1), pp. 233-258, January 2013, Springer.

[8] W. Hürst and K. Vriens. 2013. Mobile Augmented Reality Interaction via Finger Tracking in a Board Game Setting. *Proceedings of MobileHCI2013 AR-workshop "Designing Mobile Augmented Reality"*, 4 pages, 2013.

[9] H. Kato, M. Billinghurst, I. Poupyrev, K. Imamoto and K. Tachibana. 2000. Virtual object manipulation on a table-top AR environment. . In *Proceedings IEEE and ACM International Symposium on Augmented Reality, 2000,* pp. 111-119, IEEE.

[10] H. Kato, K. Tachibana, M. Tanabe, T. Nakajima and Y. Fukuda. 2003. MagicCup: a tangible interface for virtual objects manipulation in table-top augmented reality. *Augmented Reality Toolkit Workshop, 2003. IEEE International*, 2003, pp. 75-76.

[11] Kosinski, Robert J. "A literature review on reaction time." *Clemson University* 10 (2008); http://archive.is/jw9W

[12] H. Richter, S. Loehmann, F. Weinhart, and A. Butz. 2012. Comparing direct and remote tactile feedback on interactive surfaces. *Lecture Notes in Computer Science*, Vol 7282, pp 301-313, Springer.

[13] K.N. Shah, K.R. Rathod, and S.J. Agravat. 2014. A survey on human computer interaction mechanism using finger tracking. *International Journal of Computer Trends and Technology (IJCTT)*, Vol.7(3), pp. 174-177, January 2014, Seventh Sense Research Group.