Longitudinal study on text entry by gazing and smiling

Longitudinal Study on Text Entry by Gazing and Smiling

Outi Tuisku[1], Ville Rantanen[2], Veikko Surakka[1]
[1]Research Group for Emotions, Sociality, and Computing, Tampere Unit for Computer-Human Interaction, School of Information Sciences, University of Tampere, Finland
{firstname.lastname}@sis.uta.fi
[2]Sensor Technology and Biomeasurements, Department of Automation Science and Engineering, Tampere University of Technology, Finland
{firstname.lastname}@tut.fi

Abstract

This study presents the results of a longitudinal study on multimodal text entry where objects were selected by gazing and smiling. Gaze was used to point at the desired characters and smiling movements were performed to select them. Participants (N=12) took part in the experiments where they entered text for a total of 2.5 hours in ten 15-minute-long sessions during one-month time period. The results showed that the text entry rate improved with practice from 4.1 to 6.7 words per minute. However, the learning curve had not reached its plateau phase at the end of the experiment. Subjective ratings showed that the participants appreciated this multimodal technique.

Keywords: text entry, gaze direction, facial muscle activity

Concepts: • **Human Centered Computing ~ Interaction devices;** *Pointing devices*

1 Introduction

Gaze has often been used as an input method so that user points the object by looking at it and selects it by holding the gaze on it for a pre-defined period of time (i.e., dwell time). To avoid unintentional selections caused by dwell time, different gaze-based selection techniques has been applied, for example, context switching [Morimoto and Amir 2010]. Further, an added modality in conjunction with the gaze has provided researchers a possibility to make gaze-based interaction even more natural. The aim in multimodal human-computer interaction (HCI) is to create interaction techniques that imitate the natural behavior of humans and thus, use their full capacity when interacting with computers [Turk 2013]. One possibility is to use facial activations (e.g., smiling, frowning) as a selection technique with gaze pointing [Tuisku et al., 2012; 2016]. They are assumed to be natural to use in HCI as they are already used in everyday human-human communication, for example, by looking at the person that is communicated with and smiling at them.

Today, gaze-based text entry has been studied for over three decades [Majaranta and Räihä 2007]. During that time, the text entry techniques have evolved from on-screen keyboards modelled after physical QWERTY layout to different types of technical solutions where the opportunities and limitations of the gaze has been taken into consideration, such as, Dasher [Tuisku et al. 2008]. New HCI and text entry methods need to be favorable for the users in order for them to be adopted for wider use. The favorability needs to be evaluated, and cross-sectional studies do not reveal the learning curves of the techniques and their impact on the subjective experiences of the users. For these reasons, longitudinal studies on gaze-based text entry have been conducted more and more frequently [Majaranta et al. 2009; Tuisku et al. 2008] but no such studies exist on multimodal techniques that utilize gaze direction for pointing and facial muscle activations for selecting.

The aim of this study was to evaluate how participants learn to enter text using a multimodal technique that uses gaze for pointing and selecting with facial activations, and gain insight on their experiences to develop the technique further. The speed and accuracy of the text entry was also under investigation.

ETRA '16, March 14 - 17, 2016, Charleston, SC, USA

ACM 978-1-4503-4125-7/16/03…$15.00
DOI: http://dx.doi.org/10.1145/2857491.2857501

2 Methods

2.1 Participants

Twelve voluntary and naïve participants (2 male, 10 female) took part in the experiment. Their mean age was 27 years (range 19-37 years) and they were native Finnish speakers. All had normal or corrected-to-normal (i.e., with contact lenses) vision by their own report. Each participant attended ten 15-minute-long experimental sessions during one-month time-period. The sessions were arranged so that there would not be more than two consecutive days in between the sessions. Participants were rewarded with four movie tickets after the last session.

2.2 Apparatus

A wearable head-worn prototype system was used for pointing and selecting technique (see Figure 1). The prototype is described in [Rantanen et al. 2012]. The head-worn device built on the frames of protective glasses includes two cameras, an infrared (IR) light emitting diode (LED), and sensors and electronics for detecting facial movements using a capacitive method. The used cameras were low-cost, commercial cameras. The eye camera (placed near user's left eye) was a greyscale camera with a resolution of 352 × 288 pixels that was modified to image IR wavelengths. The scene camera (placed in front of user's forehead) was a color camera with a resolution of 597 × 537 pixels. The frame rate for both of the cameras was 25 fps. The IR LED was placed next to the eye camera to provide illumination for the eye and to produce a corneal reflection for the eye tracking. The scene camera was used to head-movement compensation [Rantanen et al., 2011]. The viewing angle of the scene camera was 70°. The facial movement sensors (i.e., placed in front of both cheeks in the frames) were based on measuring capacitances with a programmable controller for capacitance touch sensors (AD7147 by Analog Devices).

Figure 1: *Left: Wearable prototype for pointing by gaze and selecting by smiling. Right: Examples of smiling movements (from neutral to smile) for producing a click by smiling.*

A 24" widescreen display was used at the viewing distance of approximately 60 cm. A PC with Windows XP operating system was used to run the experiment. The software for online processing of the data from the prototype was implemented with Microsoft Visual C++ 2008. The software transformed the obtained gaze information to cursor movements and smiling movements as the selections on the computer screen.

The keyboard was implemented with Visual Basic 2008 programming language. The layout of the keyboard was introduced by Tuisku et al. [2013] (see Figure 2). The letters were placed so that the most frequent letters in Finnish language were placed in the middle of the screen, in order for them to be more easily selectable and less frequent letter were placed at the edges of the layout.

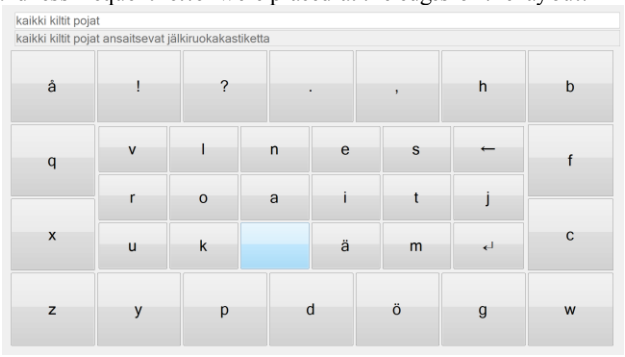

Figure 2: *The on-screen keyboard used in the experiment. The space key is currently selected.*

After typing a character, a 'click' sound was played in order to indicate a successful selection of character. The typed text appeared in the white text box above the keyboard and the target text (i.e., text to be typed) was shown below it on a grey background color. The Enter key updated the target text with a new one and cleared the typed text box if the length of the typed text was at least 70% of the target text. The cursor was not visible during the experiment, instead the key that was selected was highlighted (in Figure 2, the Space key highlighted).

2.3 Experimental Task

The task of the participant in each session was to enter text as fast and as accurately as possible for 15 minutes. If participant noticed an error during the text entry right after making it, they were advised to correct it. However, if they noticed the error later, they were instructed to ignore it. Target phrases were chosen randomly from Finnish translations [Isokoski and Linden 2004] of a phrase set for text entry experiments [MacKenzie and Soukoreff 2003].

2.4 Procedure

When a participant arrived to the laboratory, the laboratory was introduced to her/him. Then the aim of study was described to the participant and she/he was seated in a chair and the prototype device and its functionality was introduced. The participant was explained that the task would be to enter text by pointing the correct character by gazing at it and selecting it by performing a smiling movement. Participant was also told that there was a short practice task before the actual text entry task to familiarize the participant with the pointing and selecting technique. Then, the participant wore the prototype and was allowed to move in front of the computer to see how much head movements would be possible during the experiment. Then the eye tracker was calibrated. Next, a 5-minute practice task where the task was to point and select pairs of circles and squares appearing on the screen was run.

After the practice, the keyboard and the experimental task were introduced to the participant. The eye tracker was re-calibrated, and the participant started the 15-minute-long task. Once the final phrase was typed, the on-screen keyboard was hidden to indicate ending of the task. During the task, the eye tracker was recalibrated when needed (on average < 1 times/participant).

After first, fifth, and tenth session participants filled a slightly modified ISO rating scales about the functionality of the technique [ISO 9241-9:2000]. The scales were 7-point Likert scalers that varied from 1 to 7. After the tenth session, they were shortly interviewed using a semi-structural technique to find out more about how the participants experienced the multimodal technique and typing with it. Participant visited the laboratory ten times within one-month period. The first session lasted approximately an hour, sessions 2-9 approximately 30 minutes, and tenth session approximately 40 minutes. In total, each participant entered text for 2.5 hours (i.e., 10 * 15 min).

2.5 Metrics

The evaluation how the participants learned during the experiments was done with objective metrics. Text entry rate was measured in words per minute (wpm), where one word is defined as five characters (including space). Error rates were measured in two different ways: the minimum string distance (MSD) error rate and keystrokes per character (KSPC). The MSD error rate was measured with the improved MSD error rates as suggested by Soukoreff and MacKenzie [2003]. MSD error rate was calculated by comparing the transcribed text (i.e., the text that was written by the participant) with the presented text, using minimum string distance. The KSPC value indicates how often the participant had cancelled characters during writing process [Soukoreff and MacKenzie 2003]. If KSPC is 1.0, it indicates that each key press produced a correct character. If a participant makes a correction during text entry (i.e., presses Backspace key and chooses another letter), the value of KSPC is larger than one. MSD error rate only compares the transcribed text to the presented text, whereas KSPC takes into account the whole procedure.

3 Results

Outliers were removed from the data by applying Grubb's test for the MSD error rates for each session. That is, if a single value

exceeded the three standard deviations in MSD error rate analysis, the corresponding data was removed from the analysis for all of the metrics. This lead to the removal of the data of a single participant in sessions 1, 3, and 6-10, that is, 5.8% of the data.

3.1 Text Entry Rate

Figure 3 shows the mean text entry rate averaged over the participants for all of the sessions. The average text entry rate ± standard error of the mean (S.E.M.) for the first session was 4.11 ± 0.35 wpm, for the fifth session 5.36 ± 0.43 wpm, and finally, for the tenth session 6.64 ± 0.41 wpm.

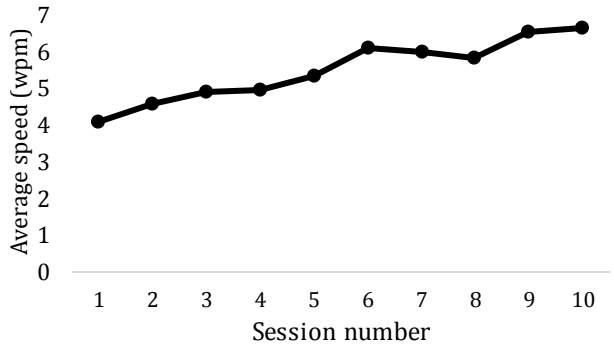

Figure 3: *The text entry rate.*

3.2 Error Rates

Figure 4 shows the mean MSD error rates throughout the experiment. The average MSD error rates ± S.E.M.s for the first session was 0.22 ± 0.12, for the fifth session 0.11 ± 0.03, and finally, for the tenth session 0.02 ± 0.01.

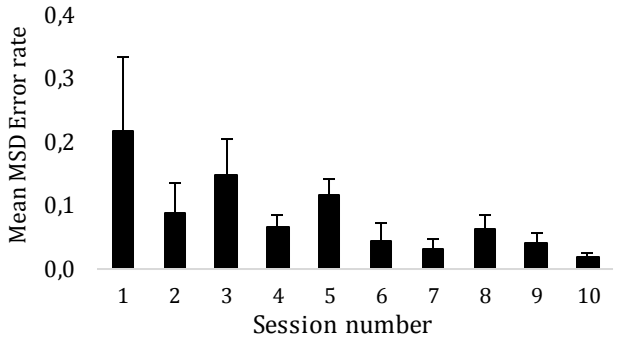

Figure 4: *MSD error rate.*

Figure 5 shows the mean KSPC values throughout the experiment. The average KSPC ± S.E.M.s for the first session was 1.24 ± 0.05, for the fifth session 1.17 ± 0.04, and finally, for the tenth session 1.20 ± 0.12.

3.3 Subjective Ratings

The subjective ratings are shown in Figure 6. In the scale, the left hand side (1) represent poorer evaluations and right hand side (7) represent better evaluations.

The Friedman test showed statistically significant differences in the overall rating ($\chi^2(2) = 12.38$, $p < 0.05$), and in the rating target selection ($\chi^2(2) = 9.18$, $p < 0.05$). To further analyze where the statistically significant effects resulted from, the Wilcoxon signed-rank test was used for pairwise comparisons.

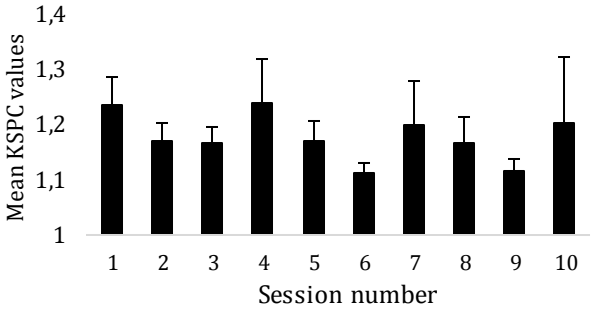

Figure 5: *KSPC values.*

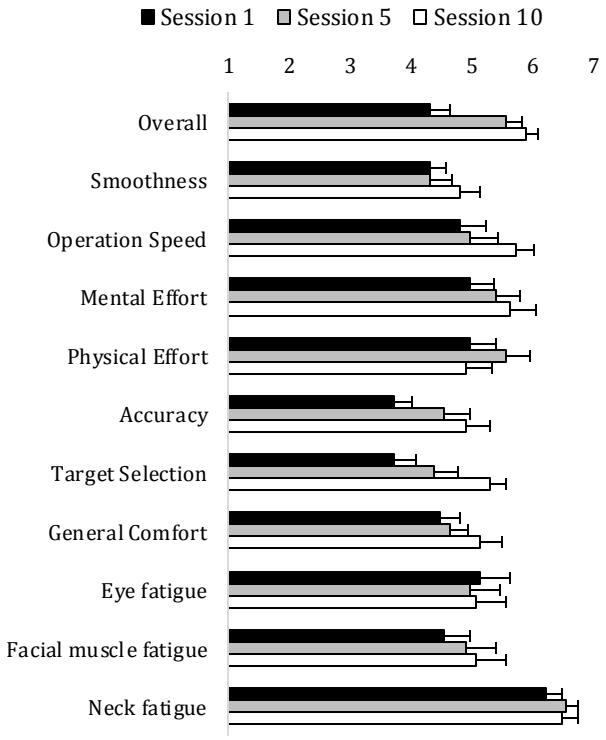

Figure 6: *Subjective ratings*

For overall rating, Wilcoxon signed-rank test showed that participants rated the system as overall better in Session 10 ($Z = 2.68$, $p < 0.05$) and Session 5 ($Z = 2.54$, $p < 0.05$) than in Session 1. The difference in ratings was not significant between Sessions 5 and 10.

For the rating of target selection, Wilcoxon signed-rank test showed that participants rated the target selection to be easier in Session 10 than in Session 5 ($Z = 2.03$, $p < 0.05$) or in Session 1 ($Z = 2.54$, $p < 0.05$). The difference in ratings was not significant between Sessions 1 and 5.

3.4 Interviews

Participants clearly appreciated the potential of the head-worn device as nine participants commented it positively. One participant, for example, mentioned that this is something that "is the future". They mentioned also that the smiling was quite natural to use. All of the participants mentioned that the use of the combination of gaze direction and smiling movement was very easy to learn and use. They particularly liked the fact that only a

small smiling movement was required for the selection.

Participants also mostly felt that they learned the places of the keys in the keyboard during the experiment. However, some of them still wondered why not to use a traditional QWERTY layout, which would (in their opinion) make the text entry even faster.

4 Discussion

The text entry rate in this study at the end of the experiment was on average 6.6 wpm. This is somewhat slower that has been reported in the previous gaze-based text entry experiments (e.g., 17.3 wpm [Tuisku et al. 2008], and 19.9 wpm [Majaranta et al. 2009]), although, completely similar study using these two modalities does not exist. Despite this, participants rated the operation speed as quite fast. Further, the text entry rate is growing at the end of the experiment. Usually the learning curve grows rapidly during first few sessions and then reaches its plateau phase [Tuisku et al. 2008]. In this case, the learning curve appeared to be still growing at the end of the experiment. Moreover, it should be noted that typing itself was a task that all the participants were familiar with, which is why the observed learning curve does not start from a completely unlearnt state. It can only be speculated on how long time would have been needed for the learning curve to reach its plateau phase. Tuisku et al. [2008] reported a similar finding in their Dasher experiment. Thus, on the whole, it seems that novel interaction techniques in entering text requires more time than 2.5 hours in order to it to gain its full potential.

The MSD error rates have a decreasing trend throughout the sessions and the difference between the first and last one is statistically significant. The KSPC values are very low in the last session but also in the first sessions and the difference between the first and the last is not significant. Overall, the participants made only few errors during the text entry, and the learning in this regard happens gradually with practice. Thus, the findings are promising because it seems that the use of two modalities clearly contributes to the low level of errors produced while entering text. It is noteworthy to mention, that this study did not reveal outlier participants as the previous experiments did [Tuisku et al. 2008; Majaranta et al. 2009].

Participants seemed to appreciate this gazing and smiling technique. All the ratings were above the middle point of the scale after the last session. These findings are similar as Tuisku et al. [2012; 2016] have reported about the use of the multimodal technique in single-session experiments. There seemed to be a trend that all the ratings improved throughout the practice, although, the improvement was not statistically significant. Comments about the prototype and the technique itself were mainly positive. This is a good indication to further improve the prototype and text entry technique by gazing and smiling. The potential of this multimodal technique appears to be greatly influenced by its naturalness and ease of use. In the future, gazing and smiling should be compared to other gaze-based multimodal techniques.

Acknowledgements

Titta Rintamäki is thanked for acting as an experimenter. The used prototype and software were mostly developed in a project funded by the Academy of Finland (decision nos 115997 and 116913).

References

ISO 9241-9:2000. 2000. *Ergonomic requirements for office work with visual display terminals VDTs - Part 9: Requirements for non-keyboard input devices*, CEN.

ISOKOSKI, P., AND LINDEN, T. 2004. Effect of foreign language on text transcription performance: Finns writing english. In *Proceedings of NordiCHI 2004*, ACM Press, 105-108.

MACKENZIE, I. S., AND SOUKOREFF, R. W. 2003. Phrase sets for evaluating text entry techniques. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, Extended Abstracts, ACM Press, 754-755.

MAJARANTA, P., AHOLA, U.-K., AND ŠPAKOV, O. 2009. Fast gaze typing with an adjustable dwell time. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ACM Press, 357-360.

MAJARANTA, P., AND RÄIHÄ, K.-J. 2007. Text entry by gaze: utilizing eye-tracking. In I.S. MacKenzie and K. Tanaka-Ishii, (Eds.): *Text entry systems: Mobility, Accessibility, Universality*, San Francisco: Morgan Kaufmann, 175-187.

MORIMOTO, C. H., AND AMIR, A. 2010. Context switching for fast key selection in text entry applications. In *Proceedings of Eye Tracking Research & Applications*, ACM Press, 271-274.

RANTANEN, V., VANHALA, T., TUISKU, O., NIEMENLEHTO, P.-H., VERHO, J., SURAKKA, V., JUHOLA, M., AND LEKKALA, J. 2011. A wearable, wireless gaze tracker with integrated selection command source for human-computer interaction. *IEEE Transactions on Information Technology in BioMedicine* 15, 5, 795-801.

RANTANEN, V., VERHO, J., LEKKALA, J., TUISKU, O., SURAKKA, V., AND VANHALA, T. 2012. The effect of clicking by smiling on the accuracy of head-mounted gaze tracking. In *Proceedings of Eye Tracking Research & Applications*, ACM Press, 345-348.

SOUKOREFF, R. W., AND MACKENZIE, I. S. 2003. Metrics for text entry research: An evaluation of MSD and KSPC, and a new unified error metric. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ACM Press, 113-120.

TUISKU, O., MAJARANTA, P., ISOKOSKI, P., AND RÄIHÄ, K.-J. 2008. Now Dasher! Dash away! Longitudinal study of fast text entry by eye gaze. In *Proceedings of Eye Tracking Research & Applications*, ACM Press, 19-26.

TUISKU, O., RANTANEN, V., ŠPAKOV, O., SURAKKA, V., AND LEKKALA, J. 2016. Pointing and selecting with facial activity. *Interacting with Computers* 28, 1, 1-12.

TUISKU, O., SURAKKA, V., RANTANEN, V., VANHALA, T., AND LEKKALA, J. 2013. Text entry by gazing and smiling. *Advances in Human-Computer Interaction*, Article ID 218084, 13 pages.

TUISKU, O., SURAKKA, V., VANHALA, T., RANTANEN, V., AND LEKKALA, J. 2012. Wireless Face Interface: Using voluntary gaze direction and facial muscle activations for human-computer interaction. *Interacting with Computers* 24, 1, 1-9.

TURK, M. 2013. Multimodal interaction: A review. *Pattern Recognition Letters* 36, 15, 189-195.