Author(s) Rantanen, Ville; Verho, Jarmo; Lekkala, Jukka; Tuisku, Outi; Surakka, Veikko; Vanhala, Toni

Title The effect of clicking by smiling on the accuracy of head-mounted gaze tracking

Citation Rantanen, Ville; Verho, Jarmo; Lekkala, Jukka; Tuisku, Outi; Surakka, Veikko; Vanhala, Toni 2012. The effect of clicking by smiling on the accuracy of head-mounted gaze tracking. In: Stephen N. Spencer, (ed.) . Proceedings ETRA 2012, Eye Tracking Research and Applications Symposium, Santa Barbara, CA, USA, March 28-30, 2012. ACM Symposium on Eye Tracking Research & Applications New York, 343-348.

Year 2012

DOI http://dx.doi.org/10.1145/2168556.2168633

Version Post-print

URN http://URN.fi/URN:NBN:fi:tty-201409151430

The Effect of Clicking by Smiling on the Accuracy of Head-Mounted Gaze Tracking

Ville Rantanen*
Jarmo Verho†
Jukka Lekkala‡
Sensor Technology and Biomeasurements
Department of Automation Science and Engineering
Tampere University of Technology, Finland

Outi Tuisku§
Veikko Surakka¶
Research Group for Emotions Sociality and Computing
Tampere Unit for Computer–Human Interaction (TAUCHI)
School of Information Sciences
University of Tampere, Tampere, Finland

Toni Vanhala‖
VTT Technical Research Centre of Finland, Tampere, Finland

Abstract

The effect of facial behaviour on gaze tracking accuracy was studied while using a prototype system that integrated head-mounted, video-based gaze tracking and a capacitive facial movement detection for respective pointing and selecting objects in a simple graphical user interface. Experiments were carried out to determine how voluntary smiling movements that were used to indicate clicks affect the accuracy of gaze tracking due to the combination of user eye movement behaviour and the operation of gaze tracking algorithms. The results showed no observable degradation of the gaze tracking accuracy when using voluntary smiling for object selections.

Keywords: capacitive facial movement detection, gaze tracking, human-computer interaction, multimodal interaction, voluntary facial movements

1 Introduction

Gaze tracking has been used in human-computer interaction (HCI) already for some decades [Ware and Mikaelian 1987]. It is commonly used for two-dimensional pointing, which makes the accuracy of the tracking an important measure of its performance. The accuracies of different video-based gaze trackers have been considered in several studies. The video-based trackers can be categorized into head-mounted and remote trackers based on the location of their cameras. Head-mounted devices have had accuracies better than 2° [Li et al. 2006; Ryan et al. 2008; Rantanen et al. 2011]. Similar degree of accuracy has been achieved with a single remote camera [Coutinho and Morimoto 2010]. Commercial gaze trackers with multiple remote cameras, for example those by Tobii [2011], report accuracies up to 0.5°.

Conventionally two-dimensional pointing in HCI is carried out with the mouse that also includes another interaction modality, i.e. indicating the selections. However, in some cases using only information measurable from the head area to implement both pointing and selecting may be beneficial, e.g. when the user has physical disabilities. New interaction methods different from the conventional ones may also be experienced more interesting and fun to use. Originally, when gaze tracking was used as a method to control an on-screen cursor, so called dwell time was introduced to make the selections [Ware and Mikaelian 1987]. With the dwell time, the selections are made when the gaze has dwelled on the same location for a certain time period. Some studies have considered dwell time problematic and using another modality for the selections has been suggested. Using blinking, winking and gaze gestures and detecting them by, for example, imaging the eyes or measuring electro-oculogram are some alternatives [Park and Lee 1996; Gips 1998]. After electromyography (EMG) was introduced as a method to use voluntary facial muscle activity in the interaction [Barreto et al. 2000], it has also been used for the clicking task when gaze tracking is used for controlling the cursor [Partala et al. 2001; Surakka et al. 2004; Surakka et al. 2005; San Agustin et al. 2009]. These studies considered frowning, smiling, and tightening the jaw movements in the interaction, with the statement that the smiling is the fastest one in this context [Surakka et al. 2005]. Later on, Rantanen et al. [2010] introduced a capacitive facial movement measurement to detect the facial movements instead of the EMG measurement to avoid using the electrodes with galvanic contact to the skin. The method has been used to detect frowning and raising eyebrows for clicking when a gaze tracker has been used for cursor control [Rantanen et al. 2011; Tuisku et al. 2011; Tuisku et al. 2012].

Previous study [Rantanen et al. 2011] already discussed the angular accuracy of making selections with a head-mounted gaze tracker that has an integrated facial movement measurement for the clicking. While the use of voluntary facial movements in combination of gaze tracking have been well justified and studied, the current study provides information about the effect that the voluntary movements have on the gaze tracking accuracy. The movement whose effect on the gaze tracking accuracy was studied was the smiling movement. Smiling was selected because it has previously been argued to be the fastest one in this context [Surakka et al. 2005]. It is evident that the dwell time should not compromise gaze tracking accuracy, but the use of facial movements might have some effect due to the combination of user eye movement behavior and the way the gaze tracking algorithms operate. For the study, we constructed a prototype system that carries out head-mounted gaze tracking in combination with a capacitive measurement that detects the facial movements. The contactless, capacitive measurement was chosen because the measurement shouldn't be obtrusive as EMG is, and, thus, possibly affect facial behaviour. The processing of the measured data to achieve the point-and-click operation is also presented. The prototype system was built because head-mounted systems for the tasks

*e-mail: ville.rantanen@tut.fi
†e-mail: jarmo.verho@tut.fi
‡e-mail: jukka.lekkala@tut.fi
§e-mail: outi.tuisku@uta.fi
¶e-mail: veikko.surakka@uta.fi
‖e-mail: toni.vanhala@vtt.fi

are not available if not built from separate components. Our system contains many low-cost, off-the-shelf components but some of the electronics is custom made. Off-the-shelf equipment for capacitive facial movement detection is not available at all even if the technological solutions are.

2 Methods

2.1 Prototype System

Our prototype system has a similar design as previous head-mounted prototype systems [Barreto et al. 2000; Li et al. 2006; Ryan et al. 2008; Franchak et al. 2010; Rantanen et al. 2011]. The constructed prototype device is seen in Figure 1. The head-mounted unit consists of the measurement and imaging parts, and a separate unit provides the wireless functionality. The two are connected via wires. The wireless unit includes 4 AA batteries, power supply circuitry, and wireless transmitters for the video signals. The computer that is used with the prototype needs receivers and frame grabbers for the videos as well as Bluetooth functionality for the capacitive facial movement measurement.

(a) *Head-mounted unit.* **(b)** *Unit for wireless operation.*

Figure 1: *The wearable prototype device.*

The software for the interpretation of the data from the prototype device to on-screen cursor movements and clicks was made using Microsoft Visual C++, Microsoft Foundation Classes (MFC), the OpenCV and Boost libraries, and the OpenEyes software. The software is designed for processing the data online with common desktop or laptop computers.

2.1.1 Gaze Tracking

The gaze tracking is based on low-cost CMOS cameras with analog video output in PAL format. The eye camera is a grayscale pinhole camera with a resolution of 352 x 288 pixels and the scene camera is a color camera with a resolution of 597 x 537 pixels. The viewing angles of the cameras are approximately 50° and 85° for the eye and scene ones, respectively. The cameras are automatic except for the focusing. The eye camera was slightly modified to image in the infrared wavelengths. An infrared light emitting diode (LED) is used to provide illumination and a corneal reflection for gaze tracking. Thus, the tracking is based on so-called dark pupil technique [Li et al. 2006]. The used video transmitter and receiver modules are AWM630TX and AWM634RX by Airwave Technologies. The transmission uses the common free 2.4 GHz band. Frame grabbers are low-cost ones by MSI.

Before applying gaze tracking, the intrinsic camera parameters (e.g., focal point) and distortions (e.g., lens distortion) need to be accounted for. The commonly used chessboard calibration procedure was used for the task [Bradski and Kaehler 2008].

The first task in the gaze tracking with the head-mounted system is estimating the eye orientation. To find the locations of the pupil and the corneal reflection for the estimation, a feature-based approach similar to the one described in [Rantanen et al. 2011] is used. The image is preprocessed, pupil is searched from the darkest pixels of the image, and the reflection from the lightest. Different criteria are set to identify the two from possible candidates, and the eye orientation is considered to be the 2D vector between their centers [Rantanen et al. 2011].

The second task for the gaze tracking is estimating the head orientation. It is done by detecting the computer screen from the scene camera image. Preprocessing is done with Gaussian filtering. Then, threshold filtering is done to extract the dark screen border before contours are searched from the resulting image. The contours are classified as rectangles if their shape is rectangular enough. Possible candidates for the outer and inner contour of the screen border are selected from the rectangles based on the colors of the neighboring pixels of the rectangle corners. The screen is considered to be found if an inner contour candidate has its corners close enough to the diagonals of an outer contour candidate. If such criterion is not met, the largest inner contour candidate is considered as the screen's inner border. Based on the found screen, the translation and rotation of the camera is computed.

Calibration procedure is required to solve what point on the scene image does each eye orientation vector correspond to. Gazing at on-screen points is used to collect information for solving a homography matrix that can be used to transform eye orientation vectors to coordinates on the scene image [Li 2006; Rantanen et al. 2011].

The coordinates on the scene image still need to be projected to physical coordinates on the computer screen to use the gaze for pointing. Values for the scene image coordinates, the intrinsic scene camera parameters, and the rotation and translation of the scene camera are used in the projection. The physical coordinate is mapped to a pixel coordinate to obtain the on-screen location of the gaze. The pixel coordinates are filtered with a simple moving average of 10 samples to smoothen the cursor movements.

2.1.2 Capacitive Facial Movement Detection

The measurement method for detecting facial movements capacitively in human-computer interaction uses the same principle as capacitive push-buttons and has been introduced in [Rantanen et al. 2010]. Due to the advancements in the capacitive measurement technology, the capacitive measurement can be implemented with a single electrode that introduces an electric field. When a conducting object shunts the electric field to the ground, the shunting can be measured. We chose a programmable controller for capacitance touch sensors (AD7147 by Analog Devices) for the task. The prototype device in Figure 1 has one electrode in front of the forehead behind the scene camera, and one in front of both of the eyebrows and cheekbones. This placement allows to target frowning, raising eyebrows, and smiling movements. The measured signals are transmitted to the computer via a Bluetooth transceiver module (RN-41 by Roving Networks).

The algorithm used for detecting the movements from the measured signals is similar to the one described in [Rantanen et al. 2011]. The signal is pre-processed before it's fed to a constant false alarm rate (CFAR) processor to produce an adaptive threshold. The threshold can be used to produce a binary signal that is integrated, and the output is compared with a second threshold to decide if a movement has occurred or not.

2.2 Experiments

Point-and-click experiments were carried out in laboratory conditions to collect data. Five participants, females between ages 21 to 34, were selected from volunteers. They all had normal vision. On-screen targets, a pair at a time, were shown to the participants, and participants had to select the targets by pointing them with the gaze and carrying out the voluntary smiling movement as many times as required to hit the target. After each pair was chosen there was a few second pause to allow the participants to relax briefly. The locations of the targets were varied so that the targets of the pairs appeared in all the possible horizontal, vertical, and diagonal directions with respect to each other. The sizes and distances of the targets were also varied, so that the locations of the targets covered the screen quite extensively. Each participant had to select 72 pairs of targets. The described experimental procedure was chosen to make the experiments similar to actual use in a point-and-click computer interface. The cursor was kept visible during the experiments. The computer screen was a 24-inch wide screen, and the participants sat at a distance of 60 cm from it.

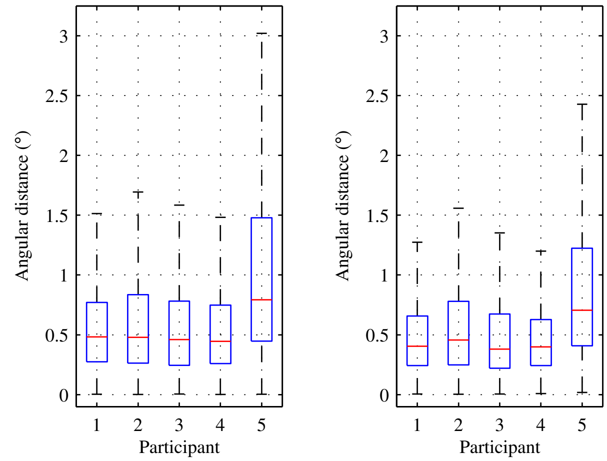
2.3 Measurement Data Analysis

Angular distances of consecutive gaze coordinates were considered as an indicator of the accuracy for this purpose. While they actually represent the precision, they also contribute to the accuracy. Using them as the inspected measure in the analysis provides a bit more information about the behavior of the computed gaze coordinates compared to using the angular error distances of the selections. Each angular distance was estimated by calculating the arc tangent of the ratio of the on-screen distance of the coordinates and the distance of the user to the computer screen. The distances were calculated separately for two time frames: the entire experiment and during the smiles. For the latter time frame, short intervals right before each selection that was registered by the facial movement detection algorithm were considered. The lengths of the intervals were 10 samples, i.e. 0.4 seconds, based on the length of the smoothing filter of the gaze coordinates. This makes the intervals include the samples that contribute to the gaze coordinates right at the selection. As this analysis would include all angular distances including those resulting from saccadic eye movements between the on-screen targets and possible eye movements resulting from gazing somewhere else than at the targets, distances beyond 1.5 times the interquartile range from the upper quartile were considered outliers to leave only the data about the fixations. However, outliers were included while comparing the distribution medians of the two time frames. The comparison was done by calculating the Mann-Whitney U test for the distances of the two time frames with each participant individually. Removing also the resting periods between consecutive pairs of targets as outliers was considered, but it was not done as it had an insignificant impact the results.

3 Results and Discussion

The distributions of the computed angular distances are shown in Figure 2. Table 1 shows the comparison between the two time frames and also lists the number of the angular distances they included. The angular distances with participant 5 were larger throughout the experiments due to distortions in the wireless transmission resulting possibly from low battery charge levels of the wearable prototype.

The results show that with the participants there was no observable degradation of the gaze tracking accuracy in terms of the observed angular distances due to the voluntary smiling movements that were used for the selections. The angular distances of the gaze coordinates are even slightly smaller right before each selection. However, this may be affected by the fact that the entire experiment includes all the eye movements that are not considered outliers by the chosen criterion. The criterion makes it possible to remove the effect of saccadic eye movements, but other types of eye movements might still affect the results.



(a) *Angular distances of the entire experiment.* **(b)** *Angular distances right before each selection.*

Figure 2: *The distributions of the angular distances between consecutive gaze coordinates. Horizontal lines are the medians, boxes extend from lower to upper quartiles, and whiskers to the maximum angle that is not considered an outlier.*

Partic.	n_{tot}, n_{sel}	z	p
1	8820, 1500	-7.65	< 0.001
2	11250, 1730	-3.12	< 0.01
3	9130, 1740	-7.27	< 0.001
4	19450, 2270	-6.96	< 0.001
5	16400, 2600	-6.87	< 0.001

Table 1: *Results of the Mann-Whitney U test that compared the angular distances during the two time frames. The test reveals a statistically significant difference with each participant. n_{tot} is the number of angular distances in the entire experiment and n_{sel} right before each selection. The values for the normalized test statistic z and the p value of the two-tailed test are also shown.*

The number of participants in our experiments was low, but we collected relatively significant amount of data from the few female participants we had. Bollen et al. [1993] have previously shown that eye movements have large intra-individual variation which is why a large group of participants may not even be needed to meet different variations. The same study didn't report any differences between male and female participants.

4 Conclusion

Previous studies have reported the use of voluntary facial movements in human-computer interaction to indicate selections when gaze tracking has been used for cursor control. The effect of the voluntary movements on the gaze tracking accuracy has not been studied, although it has been stated that a multimodal system similar to the one in this study can achieve an accuracy of selections comparable to that of pure gaze trackers [Rantanen et al. 2011]. Our results with the constructed prototype system do not show that the gaze tracking accuracy would be compromised during the voluntary smiling movements. In other words, the possible effect of the movements to eye movement behavior and the operation of gaze tracking algorithms could not be observed.

The limitations of our study include the fact that the accuracy and the precision of the constructed point-and-click prototype system limit the smallest observable degradation of accuracy. As gaze trackers improve, similar studies should be carried out to find out if the effect we were looking for could be registered. Additionally, the study could be expanded with different facial movements in addition to the smiling one even if smiling has previously been suggested as the fastest movement for the selections. A comparison of accuracies between selecting with the facial movement detection and with the dwell time could be carried out. Additional objective measures and subjective ratings could be considered and compared to provide more information for extending the current study.

Acknowledgments

This work was funded by the Academy of Finland, funding decision nos. 115997, and 116913. Part of the work was carried out in the Finnish Doctoral Programme in User-Centered Technology (http://www.sis.uta.fi/~ucit/contact.html). We would like to thank Pekka-Henrik Niemenlehto and Oleg Špakov for their contribution to the study, and also the reviewers for their comments.

References

BARRETO, A. B., SCARGLE, S. D., AND ADJOUADI, M. 2000. A practical emg-based human-computer interface for users with motor disabilities. *Journal of Rehabilitation Research & Development 37*, 1, 53–64.

BOLLEN, E., BAX, J., VAN DIJK, J. G., KONING, M., BOS, J. E., KRAMER, C. G. S., AND VAN DER VELDE, E. A. 1993. Variability of the main sequence. *Investigative Ophthalmology & Visual Science 34*, 13, 3700–3704.

BRADSKI, G., AND KAEHLER, A. 2008. *Learning OpenCV: Computer Vision with the OpenCV Library*. O'Reilly Media, Sebastopol, California, USA, Septemper.

COUTINHO, F. L., AND MORIMOTO, C. H. 2010. A depth compensation method for cross-ratio based eye tracking. In *Proceedings of the 2010 Symposium on Eye-Tracking Research & Applications*, ACM, Austin, Texas, USA, University of São Paulo, 137–140.

FRANCHAK, J. M., KRETCH, K. S., SOSKA, K. C., BABCOCK, J. S., AND ADOLPH, K. E. 2010. Head-mounted eye-tracking of infants' natural interactions: a new method. In *Proceedings of the 2010 Symposium on Eye-Tracking Research & Applications*, ACM, Austin, Texas, USA, 21–27.

GIPS, J. 1998. On building intelligence into eagleeyes. In *Assistive Technology and Artificial Intelligence*. Springer Berlin / Heidelberg, 50–58.

LI, D., BABCOCK, J., AND PARKHURST, D. J. 2006. openeyes: a low-cost head-mounted eye-tracking solution. In *Proceedings of the 2006 symposium on Eye tracking research & applications*, ACM, San Diego, California, USA, 95–100.

LI, D. 2006. *Low-cost eye-tracking for human computer interaction*. Master's thesis, Iowa State University, Iowa, USA.

PARK, K. S., AND LEE, K. T. 1996. Eye-controlled human/computer interface using the line-of-sight and the intentional blink. *Computers & Industrial Engineering 30*, 3 (July), 463–473.

PARTALA, T., AULA, A., AND SURAKKA, V. 2001. Combined voluntary gaze direction and facial muscle activity as a new pointing technique. In *Proceedings of IFIP INTERACT'01*, IOS Press, Tokyo, Japan, 100–107.

RANTANEN, V., NIEMENLEHTO, P.-H., VERHO, J., AND LEKKALA, J. 2010. Capacitive facial movement detection for human-computer interaction to click by frowning and lifting eyebrows. *Medical and Biological Engineering and Computing 48*, 1, 39–47.

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.

RYAN, W. J., DUCHOWSKI, A. T., AND BIRCHFIELD, S. T. 2008. Limbus/pupil switching for wearable eye tracking under variable lighting conditions. In *Proceedings of the 2008 symposium on Eye tracking research & applications*, ACM, Savannah, Georgia, USA, Eye Tracking Research & Application, 61–64.

SAN AGUSTIN, J., HANSEN, J. P., HANSEN, J. P., HANSEN, D. W., AND SKOVSGAARD, H. 2009. Low-cost gaze pointing and emg clicking. In *Proceedings of the 27th international conference extended abstracts on Human factors in computing systems*, ACM, Boston, Massachusetts, USA, 3247–3252.

SURAKKA, V., ILLI, M., AND ISOKOSKI, P. 2004. Gazing and frowning as a new human-computer interaction technique. *ACM Transactions on Applied Perception 1*, 1, 40–56.

SURAKKA, V., ISOKOSKI, P., ILLI, M., AND SALMINEN, K. 2005. Is it better to gaze and frown or gaze and smile when controlling user interfaces? In *Proceedings of HCI International 2005*.

TOBII, 2011. Tobii – world leader in eye tracking & eye control. http://www.tobii.com.

TUISKU, O., SURAKKA, V., GIZATDINOVA, Y., VANHALA, T., RANTANEN, V., VERHO, J., AND LEKKALA, J. 2011. Gazing and frowning to computers can be enjoyable. In *Proceedings of the Third International Conference on Knowledge and Systems Engineering (KSE)*, IEEE Computer Society, Hanoi, Vietnam, 211–218.

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.

WARE, C., AND MIKAELIAN, H. H. 1987. An evaluation of an eye tracker as a device for computer input. In *Proceedings of the SIGCHI/GI conference on Human factors in computing systems and graphics interface*, 183–188.