Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230225

Eye-Tracking on Touch Screen - Evaluating User Interaction

Camilla Panduro NIELSEN[a,1], Jacob NIELSEN[b], Chris Bang SØRENSEN[b] and Ellen Raben PEDERSEN[a]
[a] *SDU Health Informatics and Technology, University of Southern Denmark, Denmark*
[b] *SDU Game Development and Learning Technology, University of Southern Denmark, Denmark*
ORCiD ID: Camilla Panduro Nielsen 0000-0001-7547-5455, Jacob Nielsen 0000-0002-5134-8950, Chris Bang Sørensen 0000-0002-2279-2091, Ellen Raben Pedersen 0000-0001-9917-9837

Abstract. This paper suggests a setup for using remote eye-tracking on a touchscreen tablet to evaluate user interaction for older adults interacting with a user-driven hearing test. By using video recordings to support the eye-tracking data, it was possible to evaluate quantitative usability metrics that could be compared to other research findings. The video recordings revealed useful information to distinguish between reasons for gaps in data and missing data and to inform future similar studies of human-computer interaction on a touch screen. Using only portable equipment allows researchers to move to the location of the user and investigate the user interaction of devices in real-world scenarios.

Keywords. Eye-tracking, touch screen, video, usability, human-computer interaction

1. Introduction

Methods like Concurrent Think Aloud (CTA) and Retrospective Think Aloud (RTA), are often used for testing and evaluating user interaction [1]. These methods can have limitations for investigating user interaction, as CTA can cause the test subject to be distracted from the task at hand and make them perform differently, and RTA relies on the test subject to remember everything they thought while the test was happening, which is not always obtainable, especially when tasks become complex [1, 2]. Eye-tracking can be used to give quantitative measurements of user interaction to compensate for the limitations of the CTA and RTA, and it is especially useful for people with a decreased capacity in working memory, as it allows for investigation of user behavior without placing an additional cognitive burden on the participants [1].

This study aimed to set up a method for remote eye-tracking to evaluate user interaction on a User Interface (UI) of a user-driven hearing test (pure tone audiometry), displayed on a touch screen. The focus group of the study was older adults, i.e., people aged 50 years and above. Older adults fit the age of most first-time hearing aid users in

[1] Corresponding Author: Camilla Panduro Nielsen, E-mail: camilla.panduro.nielsen@rsyd.dk.

Denmark, and they are relevant for eye-tracking studies, as their working memory is lower than that of younger adults, thus making them less ideal for CTA and RTA [3, 4].

2. Methodology

To compose comparable quantitative measures for user interaction, usability metrics were set up, according to the ISO 25022:2016 standard on measurement of quality in use. The ISO standard suggests that quality in use can be measured by evaluating effectiveness, efficiency, satisfaction, freedom from risk, and context coverage [5]. In this study, effectiveness and efficiency were measured using eye-tracking and video recordings; satisfaction by assessing the user's response to the System Usability Scale questionnaire [6]; freedom from risk by observing potentially dangerous situations from the video recordings; context coverage was omitted, as this study only assessed the intended context of use. As only effectiveness and efficiency were assessed using eye-tracking, only these measures will be reported in this paper.

2.1. Experimental Setup

In this study, 17 people in the age group 51 to 85 years of age participated. None of the participants had interacted with the hearing test prior to the study. The Tobii EyeX eye-tracking controller (EyeX) was used for the experiment. It has previously been evaluated to potentially be useful for research applications [7]. The EyeX does not provide an accessible data stream of numbers, and therefore the software GazeViewer from Tobii Dynavox (GazeViewer) was used to display screen recordings with heat maps and gaze plots overlayed [8, 9]. Two GoPro cameras were used to capture the participant's movements during the hearing test. One was set behind the user to the opposite side of their dominant hand, and one was set facing the user, so their face and upper body were visible.

The hearing test used for the study was based on the Automated Method for Testing Auditory Sensitivity (AMTAS), integrated into the Interacoustics AMTAS (IA-AMTAS) solution Affinity Compact Suite. The IA-AMTAS was an alternate implementation for research purposes in the User Operated Audiometry project, with differences in UI and instructional text [10]. The UI was displayed on a 15.6-inch ZenScreen touch MB16AMT tablet, and the UI comprised several screen images that the user was interacting with. The eye-tracking controller was mounted at the bottom of the tablet, following the manufacturer's instructions [11]. The test stimuli (pure tones) were presented to the participants through RadioEar DD450 headphones.

The participants were seated at a table with the tablet at a distance so they could comfortably touch the screen, while the eye-tracker could detect the reflections from their eyes. The acceptability of the participant placement was investigated using the built-in seven-point calibration routine from the EyeX software, following the manufacturer's instructions while changing the participant's seating position, the light conditions, placement of the tablet, etc., and redoing the calibration until acceptable or abortion of the test if no acceptable calibration could be reached [9, 12].

2.2. Annotation of data

Before starting the experiments, every possible task to conduct while interacting with the UI of the hearing test was denoted with a specific code. Mandatory tasks were marked as such. All participants were faced with the same mandatory tasks and allowed to spend as much time as they wanted to interact with the UI without interruptions. Examples of mandatory tasks were: looking at information text, tapping appropriate buttons on the screen, picking up headphones, etc. After the experiments, the video recordings from the cameras and the screen recording with heat maps and gaze plots were synchronized, and annotation of task occurrence and duration was done in the transcription program ELAN by looking at the video recordings and noting which task occurred for how long [13, 14]. Annotations were done following the procedure in **Figure 1**.

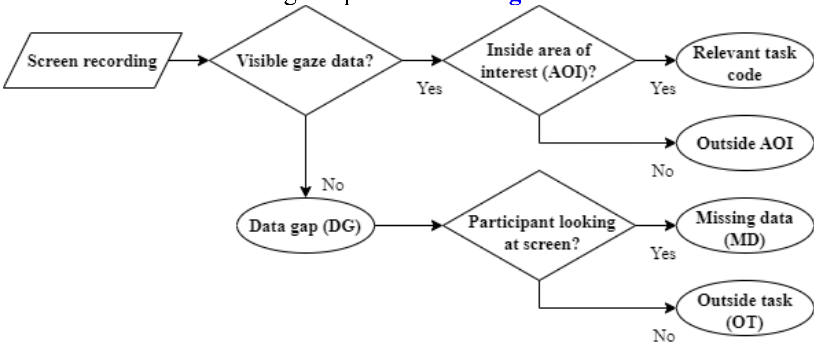

Figure 1. Flow chart describing the procedure for data annotation.

If gaze data appeared inside predefined Areas of Interest (AOI's), it was noted with the relevant task. If a mandatory task was not conducted, the duration of that task was left empty. If no gaze data were visible, it was denoted as data gaps (DG), and the DG's were further investigated by looking at the video recordings to distinguish between no data value stored from the eye tracker - missing data (MD), or the user looking outside the screen - performing outside tasks (OT's), which would not be detectable by the eye-tracker. In this study, the MD could be caused by random user behavior or prompted by the UI, so it was classified as data Missing At Random (MAR) [15]. If the recordings showed that MD occurred during the performance of a task, then all eye-tracking data were deleted for that task, using pairwise deletion [15].

The effectiveness was the percentage of mandatory tasks completed for each participant. The efficiency was reported as the percentage of time with available data when users were looking at AOI's. DG, MD, and OT were reported as percentages of the total time spent interacting with the UI for each participant.

3. Results

Table 1 shows a summary of the results. The average effectiveness is reported alongside the average percentage (%Average) of efficiency, DG, MD, and OT taking the different time spent pr. individual participant into consideration. The bottom row shows the average deviation from the average (Avgdev) for all measures. The video recordings revealed that DG, as well as OT, occurred for 16 out of the 17 participants.

Table 1. Average, percentage average, and average deviation for effectiveness, efficiency, data gaps (DG), missing data (MD), and outside tasks (OT).

	Effectiveness		Efficiency	DG	MD	OT
Average	97.6%	**%Average**	98.6%	6.8%	6.0%	0.8%
Avgdev	3.0		0.5%	7.3%	7.5%	0.6%

For six participants the DG was partly caused by MD. This is visible in **Table 1** as a higher Avgdev for both DG and MD. The reasons for MD were for three participants that their hands were blocking the eye tracker, one participant moved so that their eyes were out of range of the eye-tracker, one held the headphones so they blocked the eye-tracker, and for one participant the MD was assumed to be caused by direct sunlight.

4. Discussion

Table 1 shows that not all DG was MD, but instead, some DG was OT, accepted user behavior prompted by following the instructions on the UI. Without the cameras, all DG would have been assumed to be MD and treated as MAR, thus having to delete data for that participant. Being able to distinguish between MD and OT instead of assuming all DG to be MD, made it possible to avoid deleting relevant data while choosing the most appropriate methods for handling MD based on the true amount and cause [15].

Table 1 also shows examples of quantitative measures of user interaction like effectiveness and efficiency. Even though a data stream was not accessible as numbers, comparable measures could still be obtained by measuring from the screen- and video recordings, when following a predefined annotation procedure. It allowed for assessing some user performance metrics against other researchers' experiences. For example, the general average effectiveness of UI's has been assessed to 78% [16], and the effectiveness of the UI tested in this study was 97.6% on average. Furthermore, the efficiency shows that the UI prompted the participants to look where the researchers expected them to look 98.6% of the time on average.

Intrusive and head-mounted eye-tracking systems were excluded from the setup, as wearing the eye-trackers could make the participants more aware that they were being studied and then potentially cause them to change their behavior. The EyeX and GazeViewer alongside GoPro cameras provided measures for effectiveness and efficiency with a percentage average of 6.0% MD, which could be used for statistical analysis without affecting the bias of the results [17].

No matter which eye-tracking systems would be used, supporting the eye-tracking data with video recordings and potentially a questionnaire could reveal additional useful information about the experiments. As all the systems are portable, it allows the researcher to move to the location of the users, and this can give a more realistic experimental setup when investigating user interaction and usability on devices in real-world scenarios in any location. To determine if the usability is good/better or bad/worse, it is suggested that the researcher determines thresholds before starting the experiments.

The reasons for MD in this study were largely expected user behavior and environmental conditions that can reasonably occur in a realistic use case for the IA-AMTAS. Knowing the reasons for MD enables the researcher to evaluate if precautions should be taken in similar future setups to avoid MD, or if some of the reasons are expected user behavior and MD therefore should be accepted to keep the experimental conditions of the interaction as close to reality as possible.

5. Conclusion

This paper suggested a method for using eye-tracking alongside video recordings to evaluate quantitative usability metrics. It can provide a starting point for future studies using eye-tracking for research on user interaction on a touch screen. By supporting eye-tracking measurements with video recordings, the researcher can understand the reasons for DG and take necessary precautions to avoid them if desirable. Furthermore, the amount and cause of MD can be discovered to allow for appropriate handling of data for further statistical analyses.

Funding

This work was supported by Innovation Fund Denmark Grand Solutions 9090-00089B and the William Demant Foundation case number 19-1251.

References

[1] Bergstrom JR, Schall AJ. Eye tracking in user experience design: Elsevier; 2014.
[2] Prokop M, et al. Impact of Think-Aloud on Eye-Tracking: A Comparison of Concurrent and Retrospectiv Think-Aloud for Research on Decision-Making in the Game Environment. Sensors (Basel). 2020;20(10).
[3] Fandakova Y, Sander MC, et al. Age differences in short-term memory binding are related to working memory performance across the lifespan. Psychol Aging. 2014;29(1):140-9. doi: 10.1037/a0035347.
[4] Houmøller SS, et al. Prediction of successful hearing aid treatment in first-time and experienced hearing aid users: Using the International Outcome Inventory for Hearing Aids. International Journal of Audiology. 2022;61(2):119-29. doi: 10.1080/14992027.2021.1916632.
[5] ISO. ISO/IEC 25022:2016 Systems- and Softwareengineering - Systems and software quality requirements and evaluation (SQuaRE) - Measurement of quality in use. Dansk Standard. 2016.
[6] Brooke J. SUS: A quick and dirty usability scale. Usability Eval Ind. 1995;189.
[7] Gibaldi A, Vanegas M, Bex PJ, Maiello G. Evaluation of the Tobii EyeX Eye tracking controller and Matlab toolkit for research. Behav Res Methods. 2017;49(3):923-46. doi: 10.3758/s13428-016-0762-9.
[8] tobiidynavox. Gaze Viewer Tobii Dynavox [cited 2022]. Available from: https://www.tobiidynavox.com/products/gaze-viewer.
[9] Tobii. Eye Tracking Software Tobii Gaming2022 [cited 2022]. Available from: https://gaming.tobii.com/getstarted/.
[10] Sidiras C, et al. User-Operated Audiometry Project (UAud) - Introducing an Automated User-Operated System for Audiometric Testing Into Everyday Clinic Practice. Front Digit Health. 2021;3:724748.
[11] Tobii. Mounting Guide: Tobii Help; 2014 [cited 2022]. Available from: https://help.tobii.com/hc/en-us/articles/210252245-Mounting-guide.
[12] Holmqvist K. Eye tracking: a comprehensive guide to methods and measures. Paperback ed. Oxford: Oxford University Press; 2015.
[13] ELAN (Version 6.3). Max Planck Institute for Psycholinguistics, The Language Archive, Nijmegen, The Netherlands; 2022.
[14] Sloetjes H, & Wittenburg, P. Annotation by category - ELAN and ISO DCR. Proceedings of the 6th International Conference on Language Resources and Evaluation (LREC 2008). 2008.
[15] Kang H. The prevention and handling of the missing data. Korean journal of anesthesiology. 2013;64(5):402-6. doi: 10.4097/kjae.2013.64.5.402.
[16] Sauro J, Lewis JR. Chapter 2 - Quantifying User Research. In: Sauro J, Lewis JR, editors. Quantifying the User Experience. Boston: Morgan Kaufmann; 2012. p. 9-18.
[17] Dong Y, Peng CY. Principled missing data methods for researchers. Springerplus. 2013;2(1):222. doi: 10.1186/2193-1801-2-222.