Danmarks Tekniske Universitet

DTU

Evaluation and Testing of a Gesture-Based Self-Tracking Prototype: A Pilot Study

Author

Alexander Nils Tommy Johansson - s222994

July 4, 2024

Contents

1 Introduction

The current iteration of Subjective Self Tracking [1] features a macOS app built in Swift [2]. The app is accompanied by a 3D-printed prototype with a microcontroller running CircuitPython [3], which is programmed to get sensor readings from a Adafruit BNO085 sensor board [4]. The prototype interacts with the app through a web server to gather experiment data.

Focus is on performing a pilot study with user testing of the prototype and app. The purpose is to find, evaluate and address potential issues, bugs and crashes. In addition, the study intends to gather feedback and insights on practical use of the app, with availability and usability in primary focus.

Participants engage in hands-on activities to test and give feedback on the prototype and app. Based on evaluated feedback, appropriate changes are made to affected parts. The study offers direct feedback for the prototype and app, where changes are made to adjust and tweak the experience prior to experimental use.

2 Method

To collect feedback on the app and prototype, participants are tasked to complete available experiments in the app using the device prototype attached to their wrist.

Participants All participants are required to not have previous experience with hands-on use of the device prototype or app. All participants are selected at random. Any age or gender allowed.

Apparatus A MacBook with the Subjective Self Tracking app installed. The device prototype fastened around the participant's wrist.

Procedure The participants are asked to be seated at a table with a MacBook, with the app installed and opened. Participants have the device prototype attached to their wrist of choice. Participants are asked to test the app including the available experiments as instructed in the app. Participants are tasked to provide feedback on instructions or functionality they deem uncertain. All comments and questions regarding the experience are considered feedback and are noted. At the start of the test run, a timer is started, and stopped at the end of the test, once the participant has completed the experiments as instructed. Datasets are reviewed after the experiments have been completed.

3 Pilot

In this pilot study, there was a total of four participants. All participants had no previous experience using the device prototype. Participants are anonymous, where only comments and feedback is recorded and summarised per participant. In Figure 1, a participant can

Figure 1: Participant in Pilot Study

be seen performing testing as instructed by the app. The participants are listed below in random order.

Participant 1 The participant's first impression of the main experiment screen immediately followed with questions regarding the clarity of instructions. Notably, they mentioned phrasing of the first paragraph on the screen and instructions prior to unlocking the experiments. Once starting the experiments, feedback on the task context was indicated as clear. The participant expressed annoyance regarding the delay between prototype inputs, as indicated with animations in the app. The participant was asked once to repeat stimuli, which they indicated to be an unexpected action. A suggestion to reduce the trial stimuli count was made. The total time spent was 16 minutes and 35 seconds.

Participant 2 The participant initial thoughts on the prototype referred to the large size of the prototype. The participant repeatedly found they were unable to click the prototype button to indicate a response, with multiple attempts for calibration. Finally, they experienced an app crash on the final screen after completing the experiments. The total time spent was 14 minutes and 19 seconds.

Participant 3 The participant expressed that the wrist band felt large on their wrist. The participant mentioned phrasing of a description text. The participant expressed they found it hard to estimate the greyness without visual guidance. Participant found the experiments to be quite repetitive, and experienced delays with device input related to calibration. A suggestion to indicate more clearly when trial stimuli transitions into the real experiment stimuli was made. The total time spent was 18 minutes and 2 seconds.

Participant 4 The participant indicated that one of the experiment task contexts and the input delivered contradicting instructions, where the input slider had white to black labels instead of 0 to 100 as intended. The participant experienced the device prototype

not correctly delivering input to the app server, requiring restart of hotspot and prototype. Additionally, the device prototype needed calibration. The participant suggested to limit greyscale trial images to white, grey and black shades to ensure that the participant would try diversified shades before the experiment. The participant was not timed during the test.

Datasets When reviewing the datasets, it was found that datasets were missing the truth value of the recorded stimuli. The data only contained the participants input to the stimuli, not the value of the stimuli itself.

4 Evaluation

In order to assess which findings need to be addressed and considered, context and evaluation needs to be completed. This section provides investigated context and suggests changes based on the context. The findings are categorised as follows: bugs, negative user experiences and suggestions.

Bugs Identified issues that are considered fatal to the experiment are evaluated based on severity. All listed bugs need to be addressed.

Stimuli truth value not saved in dataset Without the stimuli truth value, the data is incomplete and the dataset cannot be validated. Action to include missing stimuli truth value.

App crash on experiment end screen The participant potentially not reaching the end screen would result in the loss of final experiment data, and the user not being prompted to complete the participant survey. An index out of bounds exception is the cause of this crash. Action to address call-site where the exception occurs.

Experiment incorrectly showing white-to-black labels instead of number labels Wrongly indicating the input labels would confuse the participant, and skew indicated responses to the stimuli. Slider is incorrectly setup with wrong labels. Action to correct labels.

Participant unable to indicate a response to stimuli using prototype Being repeatedly unable to indicate a response disrupts the experiment, potentially causing fatigue in later experiment stimuli responses. Unhandled exception in the prototype code causes the code to stop running, entering a mode where the device no longer listens to button clicks. Action to catch the exception and restore the prototype to a state where input can be indicated.

Negative user experiences Identified issues that impact user experience. All listed items are not to be addressed.

Large prototype device The size constraints are limited with this iteration of the prototype. No action.

Uncomfortable wrist band (not fitted) The wrist band can feel large for people with smaller wrists. As an option for people that experience this, a suggestion to provide a sweat band to fill in excess space is made.

Prototype needing calibration Device calibration may sometimes be needed. The sensor board self-calibrates on movement. The app accepts responses with a certain calibration status. Action to reduce calibration status threshold and user action to perform an infinity-eight movement to calibrate the device, if needed.

Waiting time after successful prototype response Waiting time after a successful response (from response indicated to waiting for response) is 2 seconds. When repeating this action 20 times for each experiment, the waiting time is noticeably long. Action to reduce animation and waiting times (inclusive) to 1 second.

Suggestions Participant-suggested changes to be considered.

Review phrasing of instruction and descriptions Phrasing of description texts and instructions can be optimised. Action to revise prominent texts, and clarify context and intent with in-person explanations.

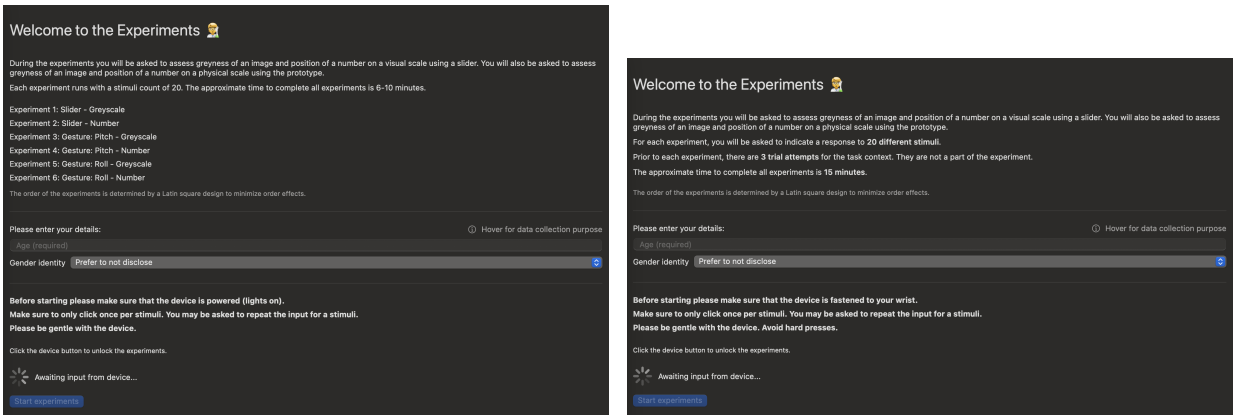
Reduce trial stimuli count With six experiments running with expected 20 stimuli responses, the amount of responses is very high already. The trial attempts are not always needed for each participant, if they already understand the assignment. A reduction in trial stimuli count could reduce likeliness of fatigue. Action to reduce trial stimuli count from 5 to 3.

Diversify trial stimuli greyscale colors Trial attempts can end up delivering very similar shades of grey sequentially. The diversity of these shades does not contribute to a good learning experience. Action to introduce chunks of shades, where all 50 shades are divided into equal chunks. Trial attempts should now indicate at least one shade of white, one shade of grey and one shade of black.

Indicate transition from trial into real experiment The current transition is assessed to be indicative. Action to introduce text addition to experiment screen that each experiment features trial stimuli which is not part of the experiment.

5 Changes

Based on the evaluations made in section 4, a number of code changes have been made to both the prototype and macOS app.



(a) Before (b) After

Figure 2: Experiments Screen Comparison

In Figure 2, we see a comparison where (a) is before text changes have been made, and (b) is after text changes were made. This change is the only change that affects content displayed in the app. A version to be used with experimental testing has been released as 1.0.0 [5] on Github.

A full list of changes include:

- Fixed crash where an array would cause an index out of bounds exception.
- Added stimuli truth values to saved datasets.
- Changed incorrect black-to-white labels to correctly reflect desired number labels.
- Introduced additional try-catching of unexpected exceptions. Microcontroller will gracefully restart and restore state if an exception occurs.
- Reduced calibration status threshold from equal or greater than 2 to equal or greater than 1.
- Reduced waiting time between successful prototype response from 2 seconds to 1 second.
- Adjusted descriptions and instructions on Experiments screen. Added additional text about trial stimuli.
- Reduced total trial stimuli count from 5 to 3.
- Diversified trial stimuli colors: trial colors are now chunked, displaying one shade of white, one shade of grey and one shade of black.

6 Discussion

Findings indicate that multiple participants experienced negative user experience, particularly when it came to delays. Delays can come from multiple sources, but during the pilot runs, the participants experienced delays with device calibration and device input. Unfortunately, this contributed to a higher fatigue, where it was perceived that the participants became less detailed in their response as the test dragged on.

Fatigue in this sense can be problematic if the participant is asked to repeat the same input numerous times. Addressing this by reducing the waiting time between inputs, specifically referring to animation and code-defined delays, can noticably reduce likelihood of fatigue. In addition, reducing the calibration status threshold allows for a reduction in pauses for recalibration needed. Changing the calibration status does reduce the overall quality of the collected sensor reading, but maintains it at an acceptable integrity.

Reduction of trial stimuli contributes to an improved experiment completion speed as well, with no drawback for initial learning. With the introduction of diversified trial colors, a more varied learning phase is created.

It became apparent that less text, and more comprehensive phrasing was needed. Complementary supervision and in-person explanations will supplement the text descriptions available in the experiment app. This will help clarify intent and give proper context when future participants try the experiments for data collection.

Suggestions regarding prototype size and wristband are things to consider in future prototypes or iterations. There is an opportunity to expand into other types of devices, to evaluate if they may be better suited for the purpose.

7 Conclusion

In this study we have found, evaluated and addressed a number of issues prior to experimental use of the solution. The pilot proved to be effective at identifying critical issues not addressed during development of the prototype. The study additionally helped address some negative user experiences prior to experimental use. Finally, potential future considerations regarding device size are mentioned.

References

[1] "Github repository: Subjective self tracking." https://github.com/johalexander/subjective-self-tracking, 2024. Accessed: July 3, 2024.

[2] "Swift.org." https://www.swift.org/, 2024. Accessed: July 3, 2024.

[3] "Circuitpython." https://circuitpython.org/, 2024. Accessed: July 3, 2024.

[4] "Adafruit 9-dof orientation imu fusion breakout - bno085 (bno080)." https://www.adafruit.com/product/4754, 2024. Accessed: July 3, 2024.

[5] "Github repository: Subjective self tracking - release 1.0.0." https://github.com/johalexander/subjective-self-tracking/releases/tag/1.0.0, 2024. Accessed: July 3, 2024.