Evaluation of a Novel Eye Track Set-up for Playtesting Handheld Games in a Lab Setting

Jeroen van der Heijden
Valsplat usability research
22 Prins Hendrikkade
1012 TM, Amsterdam
+31 20 330 5454
jeroen@valsplat.nl

Collin van Ginkel
Two Tribes
Arnhemseweg 6
3817 CH Amersfoort
+31 33 432 8832
collin@twotribes.com

ABSTRACT
With heavy competition between iPhone games, proper playtesting is vital in making an easy to use, fun game. Eye tracking can give valuable insights in player behavior but current handheld eye tracking set-ups suffer technologial limitations, inhibiting normal play. This study aims to identify the merits and shortcomings of a new handheld eyetracking set-up for qualitative user research. It is part of a series of ongoing tests to improve the set-up. In this study, seven participants played an iPhone puzzle game using the new set-up. Results indicated the set-up was suited for simple tasks like browsing, but interfered with normal gaming too much for most players. Factors contributing to interference were: Lack of depth perception, unnatural handling, uncomfortable posture and enlarged display of hands. Solutions for improvement are discussed: With longer practice for players and with tweaks to the set-up, interference can be reduced or partly removed. Accurate depth perception remains a challenge, however.

Keywords
Game user testing, playtest, mobile gaming, handheld device, usability, playability, iPhone, eye tracking.

INTRODUCTION

Playtesting gives insight in player behavior, which can improve the overall quality and playability of a game [1]. This is essential when making iPhone games since competition is fierce: There are over 21.000 iPhone games [2], which is more than all games on all other platforms combined [3]. Playtesting can help developers in making games which players quickly understand, and like.

A playtest can be defined as each opportunity developers have to gain info and feedback from a non-developer playing the game. Within this paper, we limit ourselves to user tests: playtests aiming to gauge player understanding and appreciation of the game and interface, not on finding and fixing bugs. More specifically, within this paper, a playtest seeks to improve gameplay, usability, playability and engagement.

Reducing Bias in Playtests

A playtest, like any test, needs to be properly set up and executed in order to get valid results. A playtest organized by and at the gamestudio developing the game can generate player feedback quickly and relatively inexpensive.

However, the set-up biases both participants and developers. Social desirability bias [4] may arise when participants know the researcher (friends or family). Players recruited through game forums or social media may also be prone to social desirability, since they are directly recruited by the company and are interviewed and tested by the makers themselves. When tests are held within the gamestudio, at the developer's desk, the environment may cause response bias [5]. Also, developers researching their own game can suffer from observer bias [6], not seeing the results objectively.

Outsourcing a playtest to research companies can be more expensive, but likely has less biased results with experienced user researchers who are independent of the game and gamestudio. Also, participant social desirability bias [4] is reduced when participants are recruited by the research company and interviewed by independent researchers. Furthermore, home labs are specifically designed to reduce response bias [5] and are also equipped for observing and measuring player behavior, with tools like screen capture, hidden camera, microphone and eye tracker and other biometrics.

Eye Tracking Handheld Devices

Eye tracking reveals where somebody is looking at. With a qualitative playtest, this means being able to observe how a player scans the screen, what elements are seen and missed. This helps when analyzing player behavior and interaction with the game and interface [7].

Eye tracking has been been troublesome when testing the usability of handheld devices like smartphones, tablets and gaming devices. The problem is that current handheld eye tracking set-ups alllow limited hand and/or head

movement, resulting in reduced ecological validity [8]. Head mounted eye trackers inhibit head movement and therefore normal behavior. Set-ups using eye track monitors allow head movement, but have the handheld device attached to a fixed place, inhibiting natural handling and interaction [9], especially when the device is controlled using motion sensor.

Goals of this Paper

This paper aims to investigate the merits and shortcomings of a novel eye tracking set-up for qualitative user research of handheld devices. The study is part of an ongoing series of tests (mainly unpublished) to improve the set-up.

The set-up is developed by Dutch usability research company Valsplat. It is based on prior work done by the eye tracking manufacterer, Tobii [10] and usability research company Simple Usability [11].

For the test, Valsplat collaborated with Dutch handheld game developers Two Tribes whose game Swap This was used. Findings from the test served as input for improving the set-up, as well as the playability of Swap This.

MATERIALS AND METHOD

The iPhone Game Swap This

Swap This is a puzzel game for iPhone, developed by Two Tribes. The game has a 7.5 score on metacritic [12].

The basic gameplay is as follows: connect and match fish of the same color. On the screen, fish in six different colors are randomly spread in a 6x8 grid (see figure 1). By connecting at least four fish of the same color you set them free and new fish appear. Fish can be connected by exchanging places in the grid: tap a fish on the touchscreen, then a second and the two exchange places. Bonus points are gained for longer combinations, involving more colors.

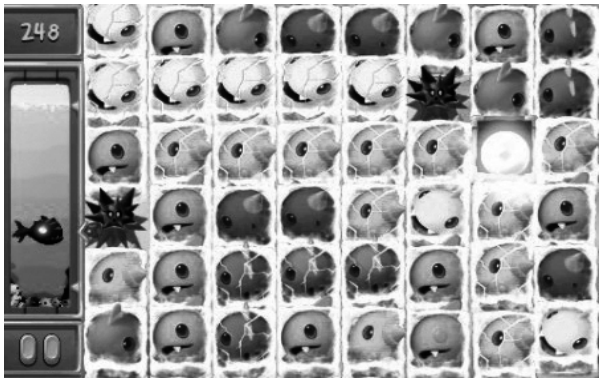

Figure 1. Screenshot of the iPhone game Swap This. Goal of the game is to connect fish of the same color. This is done by swapping fish. Tap a fish, then a second and the two exchange places.

Set up of the Novel Handheld Eye Tracker

Users saw their hands and handheld device displayed in real time on the eye track monitor, and interact looking at the monitor (see figure 2). The handheld device and user's hands were inside a box, with a high definition video camera focusing on both hands and handheld. In the box, light conditions were controlled, and the device was kept in viewport of camera, while allowing some freedom of movement. Camera output was displayed immediately in real time on a Tobii XL60 eye track monitor [13]. Participants looked at this video feed and interacted with their device.

For the test observers, the eye track monitor captured the user's gaze, which was overlayed on the video feed in real time, and visible in an observation room, together with a video feed of the user's face and a live audio feed from the lab.

Figure 2. Set-up for eye tracking use of handheld devices. Users (right) interacted with the device, looking at the eye track monitor which displayed hands and handheld. The researcher (left) saw the same image, enhanced with gaze overlay.

Natural Handling of the Device

It was vital to make interacting with the device feel as natural as possible. In pilotstudies held at Valsplat (not published), users reported handling the device while looking at the screen felt unnatural at first. However, after a few minutes they got used to the set-up.

Therefore, in the test players were asked to practice for five minutes, using their own iPhone. In those minutes, they were asked to perform some neutral tasks: planning a route, type some notes, browse the internet. Next, players were asked whether handling the device felt comfortable and natural. If so, players were asked to go to the app store, search, install and play Swap This. Otherwise, players were asked to browse the iPhone some more before installing and playing the game.

Participants Characteristics and Recruitment

The test was held with seven participants. Eight were invited; one was not present during the test. Previous pilotstudies (not published) with 6-8 participants gave valuable results for evaluating the set-up. Further, based on web usability tests [14] and previous Valsplat playtests [15], this number of participants was deemed enough to identify playability issues for Swap This.

Target group characteristics were as follows: owning and using an iPhone and regularly playing games, including puzzle games, but not familiair with Swap This. Active on Facebook and/or Twitter and sharing game scores on social media.

Players were recruited from the Valsplat database. They received an incentive equal to 20 euro, plus Swap This on their iPhone.

Test Sessions of 45 Minutes

Each session took 45 minutes. The first five minutes were used for introductions, instructions and setting up the equipment. Next, the player used his/her iPhone to get used to the box. When handling felt comfortable, players installed and played Swap This. They played for 20 minutes in the box and then alternated between playing inside and outside the box. Players were asked to compare their experiences.

Real Time Observation of Player Behavior

One valsplat researcher was present in the lab during the test. He observed gameplay with gaze overlay (see figure 2). From the observation room, gameplay with gaze overlay and the player's face was observed, plus all audio from the lab. A Valsplat researcher and a Two Tribes developer were in the observation room. All observers wrote down findings relating to the ecological validity of the eye track set-up: does the box notably affect player actions and intentions? How can the set-up improve? Additionally, they marked findings regarding the playability of Swap This, partly using available heuristics [16].

After the sessions, observations, issues and their possible solutions were compared and discussed. This resulted in two reports: one describing the pros and cons of the eye track set-up, the other describing the playability of Swap This.

RESULTS

Interference Varying Between Players

The seven players varied in the amount of interference they encountered and the time it took them to get used to the set-up.

Three players indicated they could not get completely used to playing in the box. This was supported by their performance: In the box, they played slow and repeatedly touched the wrong part of the screen. Outside the box, they played faster and touched the correct part of the screen. The three players each reported a different physical discomfort: light headache, strained eyes and strained neck.

Furthermore, two others played slow and made errors in the first 15 minutes. After that, their performance inside the box was equal to outside the box. They reported no discomfort.

Finally, two players immediately adjusted to the set-up and encountered no interference.

Factors Interfering Normal Play in the Box

Reduced Depth Perception

Five players reported trouble gauging the exact distance between their fingers and the device. Also, they had trouble landing their finger exactly on the intended place.

Enlarged Display of Hands

To maintain readability, the eye track monitor displayed an enlarged image of the device. Depending on the distance between device and camera, the magnification was about 300%. Consequently, players' hands and fingers were also magnified, which three players found unnatural. It took them a lot of effort to accurately control their large fingers.

Uncomfortable Posture

The set-up required smaller players to slightly tilt their head upwards. After 20 minutes their necks were strained.

Limited room for handling in the box

Playing in the box, players held the iPhone in both hands and played with their thumbs. Outside the box, some players held the phone in one hand and played with index and middle finger.

Interference Varying Between Tasks

There was more interference when speed and accuracy were required: When the five players had to quickly swap the fish, they made many more errors than when there was no time pressure. When browsing the menu, they made almost no mistaktes.

DISCUSSION

Current Use of the Set-Up

The eye tracking set-up was useful for testing a handheld game's interface and menu structure. For finding playability issues, however, five players encountered too much interference from the box to get useful results. In order to get usable test results for all players from the current eye tracking set-up, the inhibiting factors have to be removed or reduced.

Solving Depth Perception Issues

The errors -touching the wrong part of the screen- were partly due to a lack of accurate depth perception. Since players looked at a 2D screen, stereoscopic view was impossible. The issues concerning lack of accurate depth perception can be solved when the player looks directly at his device. Set-ups with a separate eye tracking bar, detached from the monitor, allow very limited head movement and the user's arms or hands are prone to interfere with the eye tracking equipment [10]. However, new hardware from the manufacterer, Tobii, has promising possibilities [17].

Alternatively, Tobii has eye tracking glasses [18]. They are suited for eye tracking in real-word environments (like in supermarkets) but not for handheld devices: when you look down, the bottom eyelids interfere with measurements and a lot of data is lost.

Another possibility for depth perception would be an eye tracking monitor which presents images in 3D. Presently, these are not available.

Using the currently available 2D eye tracking monitor, the depth perception issues can be reduced by adding visual depth cues in the box, like a grid. Also, better lighting conditions in which the player sees the shadow of his fingers might reduce the issues. This can be a challenge, however, because the device's screen emits light, cancelling out ambient light sources.

The lack of buttons on the touchscreen likely contributed to the problem of accurately touching the correct part of the screen. A previous pilotstudy (in press) which tested a game on a handheld device with buttons, revealed no depth perception issues. This suggests that the tactile feedback of buttons gives players a reference point.

Set-up Without a Desktop

A set-up without a desktop in which the player holds the device in his hands (figure 3) meets some of the challenges.

Figure 3. Proposed improvement on the current set-up in which the device is held in hand.

More Room for Natural Handling the Device

One can hold the phone with one or two hands, move it around and rest it in the lap.

More Comfort for Small Players

The set-up has the monitor lower so players don't have to look upward and strain their neck. Instead, they can look straight ahead or slighlty downward.

Less Maginified Image of Hands

The set-up allows for more distance between camera and device. The resulting display of device and hand is less magnified, allowing for more intuitive handling. Especially when a camera with a wider zoom is used.

Longer Practice Time

Some players may need longer to get used to the set-up. Performance may be more natural and intuitive if players can practice in a separate room, in the improved box, where they can take all the time they need to adjust.

ABOUT THE AUTHORS

Collin van Ginkel is one of the founders of Two Tribes and was part of the development team for Swap This.

Jeroen van der Heijden is a playability specialist at valsplat and has worked with various game companies including Guerrilla Games and SPIL games.

ACKNOWLEDGMENTS

We thank Joris Leker, Hessel Bonenkamp and Laurens Hoogeveen for their support.

REFERENCES

1. EDGE magazine, available online: http://www.next-gen.biz/features/the-science-of-usability-testing
2. EDGE magazine, available online: http://www.next-gen.biz/news/app-store-now-houses-over-21000-games.
3. Gizmocrunch, available online: http://www.gizmocrunch.com/apple/4067-apple-app-store-games
4. Grimm, P. 2010. Social Desirability Bias. *Wiley International Encyclopedia of Marketing.*
5. De Ruyter, B. and Aarts, E. Experience Research: a Methodoloy for Developing Human-centered Interfaces. *Handbook of Ambient Intelligence and Smart Environments*. Springer, 2009.
6. James R. Lewis. Introduction: Current Issues in Usability Evaluation. *International Journal of Human-Computer Interaction, 13,* 4, 2001, 343 – 349.
7. Johansen, S. Norgaard, M. Rau, J. Can eye tracking boost usability evaluation of computer games? Proceedings of Chi, 2008, available online: http://www.itu.dk/docadm/docs/1706.pdf.gz.
8. Goldstein, G. Ecological validity of neuropsychological testing. *Functional considerations in neuropsychology.* Delray Beach, FL, England: Gr Press/St Lucie Press.
9. World News. Eye tracking an Android Mobile Phone. Video available on world news http://wn.com/Eye_Tracking_an_Android_mobile_phone_with_a_Tobii_eye_tracker
10. Tobii. Using Eye Tracking to Test Mobile Devices. Avaliable online: http://www.tobii.com/Global/Analysis/Training/WhitePapers/Tobii_Using_EyeTracking_to_Test_Mobile_Devices_WhitePaper.pdf
11. Simple usability. M&S Xmas Multi-channel – Our iPhone Eye Tracking Review, 2011. Available online:
12. Metacritic. Swap This! Review. Availabe online: http://www.metacritic.com/game/ios/swap-this!
13. http://www.tobii.com/Global/Analysis/Downloads/User_Manuals_and_Guides/Tobii_T60XL_EyeTracker_UserManual.pdf
14. Nielsen, J. Why You Only Need to Test with Five Users. http://www.useit.com/alertbox/20000319.html
15. Van der Heijden, J. Succesful Playtesting in Swords & Soldiers. *Gamasutra.com.* Available onlie: http://www.gamasutra.com/view/feature/5939/successful_playtesting_in_swords__.php
16. Desurvire, H., Caplan, M., and Toth, J.A. Using heuristics to evaluate the playability of games. In *Extended Abstracts of the 2004 Conference on Human Factors in Computing Systems*. ACM Press, New York, 1509-1512.
17. Tobii. Tobii Mobile Device Stand. Available online: http://www.tobii.com/analysis-and-research/global/library/videos/all/mobile-device-setups-with-tobii-mobile-device-stand/
18. Tobii. Tobii Glasses Eye Tracker. Available online: http://www.tobii.com/analysis-and-research/global/products/hardware/tobii-glasses-eye-tracker/