

Analysis of Soft Data for Mass Provision of Stereoacuity Testing Through a Serious Game for Health

Gary Ushaw¹, Craig Sharp¹, Jessica Hugill², Sheima Rafiq², Carla Black², Therese Casanova², Kathleen Vancleef², Jenny Read² and Graham Morgan¹

¹School of Computing Science,²Institute of Neuroscience

Newcastle University, Newcastle upon Tyne, UK

(gary.ushaw,craig.sharp,jessica.hugill,sheima.rafiq,carla.black,therese.casanova,kathleen.vancleef,jenny.read,graham.morgan}@newcastle.ac.uk

ABSTRACT

Mass provision of healthcare through a digital medium can be greatly enhanced by the use of serious games. The accessibility and engagement provided by a serious game to the subject can significantly increase participation. The commercial games industry employs numerous techniques to analyse soft data collected from early users of an application to evolve the application itself and improve the experience of playing it. A game for mass stereoacuity testing of young children is used as a case study in this paper, to illustrate how soft feedback can be used to improve the effectiveness of a clinical trial. The key to the approach is identified as rapid incremental evolution of the application and trial protocol in a manner which increases the amount and usefulness of soft data collected, and reacts to issues identified in the soft data in a timely fashion. It is hoped that the approach can be adopted for a wide range of digital applications for mass health provision.

CCS CONCEPTS

• Applied computing → Consumer health;

KEYWORDS

Digital Health, Serious Gaming

ACM Reference format:

Gary Ushaw¹, Craig Sharp¹, Jessica Hugill², Sheima Rafiq², Carla Black², Therese Casanova², Kathleen Vancleef², Jenny Read² and Graham Morgan¹. 2017. Analysis of Soft Data for Mass Provision of Stereoacuity Testing Through a Serious Game for Health. In *Proceedings of DH '17, London, United Kingdom, July 02-05, 2017*, 5 pages. <https://doi.org/http://dx.doi.org/10.1145/3079452.3079496>

1 INTRODUCTION

A scheme for mass provision of health testing requires accessibility for the intended users to achieve maximum efficacy. Medical trials of such an application provide the opportunity to assess the accessibility of the process. Much of the feedback that can be gathered

from such trials is in the form of soft data, often represented by anecdotal reporting or preferential rating. The interpretation of this data is key to improving the efficacy of the medical process. However, soft data relating to protocols for mass provision of health intervention techniques are often difficult to acquire due to lack of engagement by the target users.

Measurement of stereoacuity in young children is vital in preventing longer term visual defects. However, current measurement techniques tend to be repetitive and uninvolved, so children do not engage with the process, leading to very sparse soft feedback. Employing a serious game for health, on a commercial tablet device, greatly increases the accessibility of the application, and the engagement of the children with the process. This engagement significantly increases the amount of soft data that is generated, due to the enthusiasm of the users. Processes for accommodating large amounts of soft data are therefore required to adequately assess and react to the soft data collected during a set of trials.

The commercial games industry is well versed in collecting and analysing soft data feedback, which is then used to evolve their product to provide greater enjoyment and engagement for the target users. Soft data is collected from focus groups and online communities throughout the life cycle of a commercial game. A vital part of this life cycle, when soft data must be reacted to in a timely manner, is immediately prior to release. The targets for commercial game developers at this stage are twofold. Firstly to gather as much useful soft data feedback as possible, and secondly to react to that data in an agile and timely fashion so as to create the largest impact on the final released version of the software.

In this paper, a serious game for measurement of stereoacuity is used as a case study to demonstrate how the analysis of soft data from clinical trials of the application was used to evolve the format and content of the medical protocol. Techniques employed in the games industry are adopted to evolve the testing protocol and increase patient enjoyment, and therefore engagement, leading to more reliable results for the protocol.

2 BACKGROUND AND RELATED WORK

2.1 Soft Data Analysis

While much medical research is based on quantitative biometric data that can be measured or observed in the patient, it is often the case that the available methods for collecting data are more qualitative. In such cases, medical markers must be identified within

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

DH '17, July 02-05, 2017, London, United Kingdom

© 2017 Association for Computing Machinery.

ACM ISBN 978-1-4503-5249-9/17/07...\$15.00

<https://doi.org/http://dx.doi.org/10.1145/3079452.3079496>

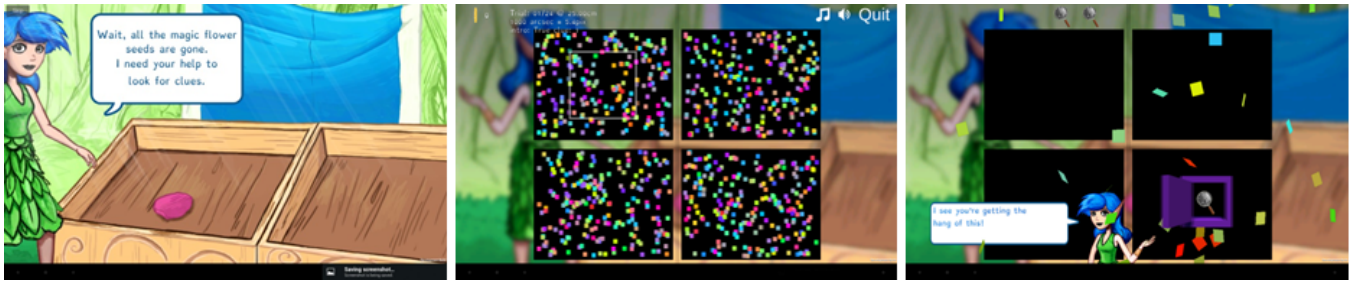


Figure 1: Three screenshots from the Asteroid game showing the macro game loop (character asks player for help, and congratulates success) and micro game loop (player selects the 3D image).

the soft data that is accrued [4], [2]. The analysis of soft data requires techniques for understanding qualitative results, which are by definition open to some interpretation [6].

The analysis of soft data is explored in [3], in the context of assumptions implicitly made about the reliability and importance of soft, or anecdotal, evidence in comparison with directly measurable parameters. A study on the use of soft data from randomised clinical trials [8] concludes that intensive design of the trials is crucial for maximising the efficacy of test procedures. In our approach we apply an early agile approach to the trials, whereby feedback in the form of soft data is used to evolve the design of the trials iteratively.

2.2 Commercial Games Development

As the appeal of video games has broadened into the population as a whole, the commercial games industry has increasingly adopted the analysis of soft data from potential and existing customers as a source of market analysis. Serious games for health are an attractive proposition as they offer an appealing method for the target user to engage with the serious intention using familiar domestically available technology. Consequently commercial techniques for the design and development of video games have been adopted into the development of serious games for health [11].

While there has been minimal literature published on the techniques employed by commercial game publishers in assessing soft data feedback (due to the commercial sensitivity of such data), the authors' own experience in commercial games development suggests that there is much that can be applied to the development of serious games for health. Commercially, soft data is accrued from a number of different sources both before and after the release of a major video game. Broadly speaking, the two sources of soft data for commercial games development are focus groups and online data gathering.

Focus groups are utilised to gain feedback on a game as its development nears completion. The process generally involves gathering a group of game players (ie the target market), and allowing them to play the game in a relaxing environment. The players' behaviour is usually observed throughout (sometimes through a two-way mirror), and any questions, comments or difficulties are noted. After a play session the groups are asked questions to assess how much they enjoyed the experience. Typically, as the focus groups consist of individuals with a keen interest in gaming, the soft data accrued is extensive and detailed.

The ubiquity of online access to games vastly increases the amount of soft data that can be accrued for commercial video games. It is common to release a *Beta* version of a game to selected registered users, allowing them to play the pre-release version of the game in their homes and give feedback. In-game telemetry data is also commonly uploaded to the game publisher giving minute-by-minute information on which aspects of a game are the most played, and most challenging [10]. Further to this, game development companies routinely analyse online forums, social media, and similar communities for soft data on every aspect of their games.

The amount of soft data that can be gathered from commercial game users is therefore considerably larger than can typically be gathered for a research project. While the success of analysis of soft data in the commercial sector can primarily be measured in terms of sales of product, the performance indication for serious games is less straightforward. Soft data can be used to assess how much a user group enjoys the experience of the serious game, and to modify the content and protocol accordingly, however there are further key performance indicators related to the efficacy of delivering the serious purpose of the game.

3 METHODOLOGY

In this section we briefly introduce the serious game for health (named *Asteroid*) which was developed for the assessment of stereoacuity in young children, and then describe the protocol for the use of *Asteroid* in the field tests. The structure of the soft data feedback from the user groups is then presented. Finally we describe in detail the manner in which the protocol and application were evolved based on the soft data feedback accrued. The approach is inspired by commercial practices in readying a video game for release to the general public, and the analysis techniques are presented in the context of applying such techniques to serious games for intervention in public health issues.

3.1 Asteroid: A Serious Game for Health

Asteroid is a serious game developed in conjunction with a professional game studio (Fluid Pixel) to provide stereoacuity testing of young children.

The game is developed for hand-held tablet devices which incorporate a glasses-free stereoscopic display. The choice of platform has two benefits in provisioning the testing application for the general public: firstly it is developed on commodity hardware which is

cheaply available and familiar to users; secondly, as the display incorporates the 3D element, there is no requirement for the children to wear 3D glasses (which some are reluctant to do for the duration of a test).

Game design typically consists of two levels of engagement with the player [7]. The *micro game loop* describes what the player does from moment to moment (also sometimes referred to as *flow* [1]), while the *macro game loop* gives the gameplay some wider context in the form of progression of story or ability. In the case of *Asteroid*, the micro-game loop requires the player to correctly identify an image which includes an element in 3D perspective from a set of four images (the other three images are presented entirely in 2D format). The macro-game loop entails a scenario whereby the player is searching for items that are revealed by correctly choosing the 3D image in order to progress a storyline (for example, a farmer has lost some chickens and must find them). Figure 1 shows some screenshots from the game. In screen A the character is asking for help (macro loop), in screen B the player must identify the 3D image (micro game loop), and in screen C the player is congratulated on success (macro game loop) so the game moves to the next challenge.

The macro game loop of *Asteroid* is also purposed to provide progression of the stereoacuity test. The success or failure of the user to identify the 3D images provides the input for changing the amount of 3D perspective that is applied in the next test according to a typical staircase structure. This approach allows a much finer level of fidelity in the testing of stereoacuity than is provided by previous methods where the thresholds of perspective are not adaptable from one test to the next.

The distance from the user's eyes to the display device is a required parameter in assessing stereoacuity. *Asteroid* utilises the in-built camera in the tablet device to measure the distance at which the device is held from the child's face. This is achieved through attaching a high contrast black-and-white sticker to the child's forehead, which can then be detected in the camera image.

3.2 Testing Protocol

To assess the efficacy of *Asteroid* as a tool for stereoacuity testing, a set of trials were carried out whereby children were asked to take part in four different types of vision testing:

- Preschool Randot (Stereo Optical Company) is a pre-school stereotest book which utilises polarised glasses to enable children to see 3D pictures emerging from a random dot pattern in 3 out of 4 boxes at each stereoacuity level that is tested. It is made up of 6 stereoacuity levels over 3 pages, with stereoacuity reducing (i.e. images becoming harder to see) as the task progresses. It is usually tested at a viewing distance of 40 centimetres. The final stereoacuity attainable is 40 seconds of arc.
- The Crowded LogMar linear test (Keeler Ophthalmic Instruments UK) is a flipbook which consists of a series of 4 letters on each page. The visual acuity level decreases (i.e. the letters become smaller) as it progresses. It is usually tested at a viewing distance of 3 metres. The lowest visual acuity attainable is -0.3 logmar units.
- During the Cover test, an occluder is moved between both eyes and each eye individually, with the subject fixating

a target, i.e. on a fixation stick, to determine whether the individual has a latent or manifest deviation. The recovery and fixation of the eyes are noted. This test is performed both for near (1 to 3 metres) and distance (> 4 metres).

- *Asteroid* is our serious game on the hand-held tablet, as described in section 3.1.

On commencement of the protocol, the child is issued with a colourful treasure map on paper. Taking part in each test gives them a tick next to the image of each piece of treasure. The child can take each test in whichever order they choose. At the end of the session, if all treasures are ticked, the child receives a certificate and a prize sticker. Each child is also issued with a Case Report Form (CRF) to be completed by the researchers at each testing station.

At the *Asteroid* testing station, a sticker is attached to the child's forehead so that the software can automatically measure the distance from the 3D image to the eyes. The child is then asked to choose which of the five different *Asteroid* games they would like to play. *Shape World* and *Simple Test* are more suitable for very young children (nursery and reception age), while *Chicken Farm*, *Football World* and *Fairytale Forest* are suitable for older children. Each game performs identically in terms of the stereoacuity test and the micro game loop, but the context of the macro game loop is different in each. A trial starts with some puzzles which do not include 3D perspective, so the solution is indicated through a 2D highlight on the screen - this approach familiarises the child with the application and the macro game loop before the stereoacuity testing commences.

The *Asteroid* session involves around 28 trials (some leeway is allowed for mistakes in the initial practice screens). The application adjusts the degree of stereo perspective applied to the image dynamically as the game progresses. Correct or incorrect identification of the 3D image causes the degree of stereo perspective to be adjusted according to a staircase structure. The application therefore gradually identifies the threshold level of stereo perspective that is detectable by the subject. This means that the last few trials are likely to be focussed around that threshold leading to the child feeling less comfortable with their ability to find the correct solution. The researcher administering the trials is briefed to offer words of encouragement whenever required (For example "it's okay to make a mistake now and then" or "where do you think the prize is hiding now?"). Any such comment is noted by the trial administrator. Any comments made by the child while playing the *Asteroid* game are also noted, as part of the soft data accrued from the trials.

3.3 Soft Data Structure

The trials took place in schools, nurseries and museums with various sized groups of children. In total 580 children were assessed using the protocol. The soft feedback is structured in two ways. Firstly, each child is asked their preference for how much they liked two of the tests (*Asteroid* and the Randot test). The level of enjoyment was indicated by the child choosing from a range of five increasingly smiley faces. Secondly, any comments made by the child during the tests were noted down. All feedback is voluntary, so not all children expressed a preference.

3.4 Evolution of Protocol

The games industry relies heavily on soft data feedback to hone the enjoyment factor of their product. The key to success in this area is twofold: the accrual of as much soft data as possible, and a rapid analysis and reaction to that data. Consequently our description of the evolution of our protocol falls into these two categories. Firstly we address how we modified the protocol in order to increase the amount of soft data which is gathered from the user groups, and secondly we address how the rapid assessment of that data was used to evolve the application and the protocol with the aim of increased engagement and efficacy of the stereoacuity testing.

3.4.1 Soft Data Quantity and Quality. The focus of the project was the collection of hard data relating to the stereoacuity of the subject's vision, which is recorded by the application itself in the case of *Asteroid* and through the protocol of testing in the case of Randot and the other tests. Initially the soft data gathered was anecdotal - i.e. noting any comments made by the children without prompting them for an opinion.

In order to increase the amount of soft data accumulated three changes were made to the protocol after the initial set of trials. Firstly, the child was asked whether they preferred *Asteroid* or Randot. Secondly, the children were prompted for comments on how enjoyable they found the experience, with all feedback noted on the CRF. Thirdly a quantitative feedback scheme was introduced in the form of a Likert scale [9]. Each child was asked to rate *Asteroid* and Randot on a scale of 1 to 5, where the lowest score indicates a strong dislike for the game, and the highest a strong liking of the game.













1. Did you enjoy playing the Randot game? 	    
2. Did you enjoy playing the Asteroid game? 	    
3. Tell us what you think about ASTEROID:	

Figure 2: The pictorial Likert scale utilised to assess young children's enjoyment of the stereoacuity testing approaches.

As the trials were carried out with young children, some of whom needed further prompting on the meaning of the Likert scale, a more child-friendly approach was later adopted. In this approach, rather than indicating a score between 1 and 5 to represent their enjoyment, the child is asked to indicate a face emoji, ranging from unhappy to very happy [5], as shown in Figure 2. Speed of feedback strongly suggested that this approach was more understandable to the children, although in some cases the child had a tendency to choose the face they liked the most.

3.4.2 Process Evolution. A number of changes were made to the process whereby the *Asteroid* application was used in the trials after initial testing with groups of young children.

It was quickly observed that tablet alignment was crucial in allowing the participant to detect the 3D image on the screen. The

tablet must be held orthogonally to the viewing angle, whereas many children preferred to place it in their lap so that the viewing angle was much more oblique. Initial encouragement for the child to hold the tablet at arms length in front of their face greatly improved the reliability of identifying the 3D image this is a known issue with glasses-free 3D display hardware, which could not be addressed through adaptation of the application software.

The initial images in each trial do not use 3D perspective to indicate the target solution. In order to familiarise the child with the game's format and macro game loop, the first few tests highlight the correct image with a coloured background. While this was effective, it was found that colouring the background of the whole image led the child to concentrate more on the centre of the image; however the 3D perspective is better detected by considering the edges (ie where the change in perspective occurs). Consequently a coloured frame was introduced for the initial 2D tests, to encourage the child to attend to the edges of the image from the start of the trials.

This was replaced with a more obviously coloured framing highlight as the initial colouring was too subtle for some test subjects.

A tendency was observed in some nursery-aged children to indiscriminately tap at the screen until something happened. To reduce this phenomenon additional instructions and verbal prompts were introduced to the protocol, such as "look at each of the four squares before tapping". This modification to the protocol was observed to reduce the amount of indiscriminate tapping and improve the efficacy of the application.

The *Asteroid* application automatically measures the distance from the images on the screen to the user's eyes from image analysis techniques applied to the input taken from the tablet's in-built camera. As the tablet devices are of relatively low computational power, the software technique utilised has to be extremely efficient. Consequently a technique was adapted which detects brightly coloured stickers in the scene [12]. At the start of each trial the child is asked to place a high contrast sticker on their forehead. The orientation of the sticker does not matter as the software takes this into account.

Most of the children tested regarded wearing a sticker as part of the "fun" of the experience. It was found that a few children refused to wear the sticker. In these cases a tape measure was used to set the distance from the tablet to the eyes to 25cm, and noted in the CRF. In cases where the child's hair would cover the sticker on the forehead, it could be attached to the chin or cheek (not to the nose however, as the spatial displacement from the plane of the eyes is too great).

4 RESULTS AND EVALUATION

The figures in Table 1 represent the number of children that expressed a preference for either *Asteroid* or the Randot test. In total 196 children provided soft data feedback from these studies. Slightly over half of the respondents expressed a specific preference for one test over the other.

The stated preferences are overwhelmingly in favour of the *Asteroid* tablet game. Of those children who expressed a preference, 85% preferred *Asteroid*. Taking all respondents of soft data into account, 44% expressed a preference for *Asteroid* while only 8% preferred the Randot book test (a greater than 5:1 ratio).

Table 1: Comparison of user preference for *Asteroid* or Randot Test. Average score (out of 5) is taken only from those users who preferred the stated test.

	Subjects with soft data	Preferred Randot test	Preferred <i>Asteroid</i>
Average Score		4.55	4.63
Initial trial	149	8	53
Restests	47	8	34
Totals	196	16	87

The average score for each of the tests is taken from those subjects who preferred that test. The scores are roughly equivalent, with *Asteroid* scoring slightly higher. However the number of users taken into account for the *Asteroid* score is significantly higher than that for the Randot score.

The quality and quantity of the soft data was greatly improved as the trials progressed, as the protocol was evolved to reflect the analysis of the soft data from the early tests, as described in section 3.4.

5 CONCLUSIONS

Techniques from commercial games development for the accrual and analysis of soft data can be applied to the development of serious games for health. In this paper a game for measuring the stereoacuity of young children has been used as a case study in this area. In order to successfully evolve a product based on soft feedback from potential users, two areas have been identified as key.

Firstly a testing protocol should be designed in a manner which maximises the amount of useful soft data which is collected during the process. The approach to achieve this should take into account the demographic nature of the user group (in our case study, young children). Secondly the analysis of that data should be rapid and, most importantly, should be fed back in order to evolve the protocol and increase its efficacy in as timely a manner as possible.

The evolution implemented to the case study protocol for stereoacuity testing was successful in both of these contexts. Early tests contained very little useful soft data so the protocol was immediately changed to include a statement of preference by the user and then a Likert scale feedback system, which was again evolved to a more child-friendly scale based on smiley faces. Similarly changes to the application and the protocol, identified from early soft data, improved the quality of the feedback, and the engagement of the subjects in later trials.

The approach taken for the collection and analysis of soft data in the case study in this paper adopted strategies which are commonly used in the commercial game sector to assess the viability of a software application as it is released to the mass market. The authors hope that the recommendations made from the presented case study can be adopted for future development of digital applications for

the provision of mass health interventions, particularly in the field of serious games for health.

6 ACKNOWLEDGEMENTS

This paper presents independent research commissioned by the Health Innovation Challenge Fund (HICF-R8-442, WT102565/z/13/z), a parallel funding partnership between the Wellcome Trust and the Department of Health. The views expressed in this paper are those of the authors and not necessarily those of the Wellcome Trust or the Department of Health.

REFERENCES

- [1] Jenova Chen. 2007. Flow in games (and everything else). *Commun. ACM* 50, 4 (2007), 31–34.
- [2] Richard Davison, Sara Graziadio, Kholood Shalabi, Gary Ushaw, Graham Morgan, and Janet Eyre. 2014. Early response markers from video games for rehabilitation strategies. *ACM SIGAPP Applied Computing Review* 14, 3 (2014), 36–43.
- [3] Silvia Gherardi and Barry Turner. 2002. Real men don't collect soft data. *The qualitative researcher's companion* (2002), 81–100.
- [4] S Graziadio, R Davison, K Shalabi, KMA Sahota, G Ushaw, G Morgan, and JA Eyre. 2014. Bespoke video games to provide early response markers to identify the optimal strategies for maximizing rehabilitation. In *Proceedings of the 29th Annual ACM Symposium on Applied Computing*. ACM, 20–24.
- [5] Lynne Hall, Colette Hume, and Sarah Tazzyman. 2016. Five Degrees of Happiness: Effective Smiley Face Likert Scales for Evaluating with Children. In *Proceedings of the The 15th International Conference on Interaction Design and Children*. ACM, 311–321.
- [6] Michael Huberman and Matthew B Miles. 2002. *The qualitative researcher's companion*. Sage.
- [7] Raph Koster. 2013. *Theory of fun for game design*. "O'Reilly Media, Inc".
- [8] Helena C Kraemer and Sue Thiemann. 1989. A strategy to use soft data effectively in randomized controlled clinical trials. *Journal of Consulting and Clinical Psychology* 57, 1 (1989), 148.
- [9] Rensis Likert. 1932. A technique for the measurement of attitudes. *Archives of psychology* (1932).
- [10] Ben Medler, Michael John, and Jeff Lane. 2011. Data cracker: developing a visual game analytic tool for analyzing online gameplay. In *Proceedings of the SIGCHI conference on human factors in computing systems*. ACM, 2365–2374.
- [11] Gary Ushaw, Richard Davison, Janet Eyre, and Graham Morgan. 2015. Adopting Best Practices from the Games Industry in Development of Serious Games for Health. In *Proceedings of the 5th International Conference on Digital Health 2015*. ACM, 1–8.
- [12] G Ushaw, E Ziogas, J Eyre, and G Morgan. 2013. An efficient application of gesture recognition from a 2D camera for rehabilitation of patients with impaired dexterity. In *8th International Conference on Health Informatics*. Springer.