Beyond user-centred design. Crowdsourcing with Serious Games for Design.

Edward Oates

Cranfield University, UK

ABSTRACT

Designers rely on direct access to 'users' (those who will use the product) to assist in their design process. User-centred design strongly emphasises the full involvement of users in the design process; but what happens when they aren't available? This study investigates the extent to which Serious Games may offer an asynchronous remote alternative to 'face-to-face' design processes through Crowdsourcing. A design process completes with summative usability testing of the product. Again, a lack of access to users is a serious limitation and one that may be ameliorated by remote unmoderated usability testing. The extent to which Serious Games may be the vehicle for remote usability testing is also explored in this research. Results from the Crowdsourcing activity show, from contributed design ideas, that a Serious Game may provide a credible tool for Crowdsourced Design. Remote unmoderated testing has known limitations and the use of Serious Games provides some mitigation, with careful implementation being required. This is a mid-study report on UK Ministry of Defence sponsored research under the Royal Navy 'DARE Innovation' initiative.

KEYWORDS

Crowdsourcing, serious games, remote usability testing

Introduction

The process of taking a product from a concept to a fielded product is a multi-disciplinary task. Within the UK Ministry of Defence (MOD) the areas of responsibility are broken down into nine domains, referred to as Defence Lines of Development (DLOD):



Figure 1: Defence Lines of Development (DLOD)

Each of these nine domains have their own design process and work in conjunction with the others. Training, Personnel and Concepts & Doctrine have a close connection with the Equipment DLOD that this research focuses on, and all require access to the developing Human-Computer Interface (HCI). At the early stages of defence procurement between Initial Gate (start project Assessment) and Main Gate (start Demonstration) it's envisaged that Crowdsourced Serious Games could be used as a rapid prototyping tool. In this role it would supplement the current text-based System Requirements Document to provide a dynamic graphical representation of the HCI. Once at the Demonstration phase, the Serious Game would be passed to the Solution Provider to help inform the final HCI implementation. At this stage, the analysis of game play data would provide objective assessments on the effectiveness of HCI options.

As part of the equipment demonstration and development process, the contracted Solution Provider will typically simulate the defence equipment HCI in a test laboratory. This will be supported by a small number of Subject Matter Experts (SME) from the users' organisation as directed by the users' representative in the Procurement Team. The test laboratory will act as a human-factors usability testing facility as well as a prototyping tool which is a well-documented route for HCI design (The Ergonomist 560 12-13, Lazar 20-21, Sauro Ch2).

The issue to be addressed is the user organisation's inability to attend the Solution Provider's test laboratory in sufficient numbers to fully support the design and testing activities. In the UK MOD context, this may be due to operational commitments, travel restrictions or by part-time Reservists' employment contract duration. A further point is that the users' perspectives will be limited to their own domain experiences but during the assessment and early demonstration phases, a wider set of ideas are required to explore a range of HCI options before narrowing down to the optimal delivered solution. As the Training, Personnel and Concepts & Doctrine DLODs also rely on access to the prototyping activity, an on-line Serious Games offers further benefits.

The terms: 'Serious Games' and 'Crowdsourcing' may be new to some so the following paragraphs provide a brief introduction.

Serious Games have developed since the 1950s out of War Gaming (Abt xiii) based on the Prussian Army's 'Kriegsspiel' but became wide spread with computerisation. Browser games are being used for education, entertainment and customer relations marketing. They capture the elements of a scenario, provide a HCI and, if well designed are hugely engaging (Rigby). The term 'Serious Games' is used to describe a computer game used for a business purpose as opposed to entertainment. When used for data capture and analysis they can also be useful (Mitroff 1-5) and it's for this purpose that a Serious Game was built within this research.

Crowdsourcing (Howe) was termed out of the business process of 'outsourcing', moving business tasks to other companies and often other nations. Where a task may be performed on-line, then a self-selected workforce may be engaged either for fun or for payment. For Crowdsourcing to be effective Surowiecki (Surowiecki 10-11) offers four essential criteria for the process to have: Diverse opinions, Independence, Decentralisation, and an Aggregation mechanism.

Method

The research seeks to use an on-line Serious Game as a method to connect designers with users. The synthetic environment within the game was subject to verification and validation methods using live system data where available. The use of Crowdsourcing is the novel step beyond usercentred design to connect with expertise outside of the users' domain experience to obtain innovative design options. An Electronic Warfare game was built around a maritime search task and the HCI was developed by Crowdsourcing from an initial design based on the Merlin Mk2 helicopter. Following the design activity, the game was then used for remote unmoderated testing to identify functions and displays correlated with mission success. The on-line game competition formed the aggregation mechanism to resolve which design elements might be taken forward to a production design. As the MOD intranet prevents write-back to the server, all game results had to be recorded in log files on the client computer and emailed in by the player.

Results

The results offered in this paper are at a preliminary stage. The new elements of HCI design gathered during the Crowdsourced design phase are known and listed in Figure 2 below, but their effect on game-play and scoring are not yet confirmed.

The ability to Crowdsource on the MOD intranet was constrained by the inability to advertise the game's presence to a wide audience. This lead to a grass-roots approach, avoiding the command structure, where word of mouth and emails to those with a known interest in the subject grew the audience incrementally over a two-and-a-half-year period. Even with these constraints, novel concepts for functions and displays were offered. As the browser game was written in simple script-code: HTML, CSS and jQuery, it was relatively easy to implement the new ideas. For example, the most complicated request for an Emitter Library (a database of known radar emitters) only took a week to build.

To illustrate the development rate, the following time-line (Figure 2) shows the Crowdsourced Design feedback over time.



Figure 2: Crowdsourced Design Feedback Rate

From the players' point of view, having requested a new function or display they then had the opportunity to re-play the game with the new functionality. Whether the innovation came from a 'user' or from someone outside the Maritime or Electronic Warfare domain, the new function or display was coded into the game within a week, offered as a player-selectable option and left to

them to choose whether to use it or not. Whether these new functions proved to be useful was assessed in a correlation analysis.



Suggestions for HCI development came from a variety of locations within MOD.

Figure 3: Crowdsourced Design Feedback

Although the MOD is a single government department, it is formed of many discrete units which, in the author's experience tend not to interact at the lower levels of the hierarchy. Figure 3 shows a broad breakdown of the sources of players who contributed and the suggestions offered. Of note is that in the traditional 'user-centred' procurement process, only the 6 Players (8 Suggestions) from the Royal Navy Merlin Helicopter group would have been included.

The next stage of the Crowdsourcing process is to establish an Aggregation mechanism, and an online game competition was used as the first part of this method. As this paper is a mid-study report, only the results from one competition have been used to illustrate the remote unmoderated testing. The game log files provided data on game score (points being gained for the correct identification of ships), and the time taken to play the game in seconds. The log files also recorded the number of times function buttons had been clicked (selected) and the duration that displays had been shown. As this was remote unmoderated testing, there was no further data to understand why a function was selected, or if a display on show was being monitored or actively used in the game-play. A within-group randomised trial was organised with experienced aircrew selected to play two versions of the game. Each player was asked to play either the Crowdsourced game four times and then the Original game four times, or the Original game four times and then the Crowdsourced game four times (though one player only completed four Original interface games). Only experienced users were selected at this stage, as through their game-play tactics they would be selecting the design elements for their own future HCI. The two versions of the game are labelled 'Crowdsourced' (the fully developed game interface from Crowdsourced Design activity), and 'Original' (the original



game interface with only the real equipment functions and displays). Figure 4 provides an overview of all scores and times achieved.

Figure 4: Game Level 3 Results Overview

Table 1 shows the statistical information obtained from the game log files.

Table 1: Game Log Files statistical information

Game Interface	Score (points)		Duration (sec)	
	Mean	SD	Mean	SD
ALL	1286.80	271.18	1116.43	1138.18
Original	1320.93	246.29	1419.00	1390.76
Crowdsourced	1246.74	298.33	761.24	595.99

From Table 1, the Crowdsourced displays have a lower mean game score than those games with the Original displays. This was not expected as the extra functions and displays built into the Crowdsourced game were derived from players' requests for improvements. It was expected that these improvements would lead to higher scores. Games played with the Crowdsourced version were, however completed faster on average. The standard deviations show that the Crowdsourced displays have a much smaller spread of game-play times than the Original displays, but a slightly greater spread of scores.

The Serious Game was played by seven similarly experienced operators from the user organisation, all with at least 10 years' experience of Electronic Warfare equipment. An ANOVA showed that 'Player' was the greatest source of variation. Given the level of player skill, the variation in results was not expected.

A Principal Component Analysis of 30 functions and displays did not show a clear distinction between Original game play data and Crowdsourced game play data. To help focus the investigation, a cluster analysis based on a balanced combination of Scores and Times (termed 'Utility where Score and Time are given equal weighting) showed the following groupings.



Figure 5: Game Level 3 Utility Cluster Analysis

By re-analysing correlation of the functions and displays used by 'Utility' groups, it was discovered that the difference in the use of the Inverse Synthetic Aperture Radar (imaging radar) display and the use of the Electro-Optic Camera gave some players a distinct advantage. The imaging radar gave a much greater identification range than the Electro-Optic Camera, and those players who had discovered this feature (the High Utility group) produced much better results. The objective to isolate Electronic Warfare controls and displays was being masked. Another factor that may have contributed to a wide spread of results was that the game was built with a true random selection for initial ships' locations, course and speeds creating a wide range of scenario challenges. Having reviewed the results from Game Level 3 it was decided that seeded random selections would control the variables more effectively and offer a known factor for analysis, and the relative detection ranges of imaging radar and electro-optic camera should be re-balanced.

Discussion

User-centred design can be thwarted in situations where the users are not available to take part in traditional processes. Face-to-face meetings and the use of human factors laboratories may offer the ideal solution but where domain expertise is widely geographically dispersed then this may not be achievable. Even when users are available, the early development of HCI may be restricted by previous experience on legacy designs. The concept of Crowdsourcing (Surowiecki 10-11) requires four criteria to capture innovative ideas:

- 1. Diverse Opinions. Where the members of the crowd do not all share the same view.
- 2. **Independence**. Where the members of the crowd are not affected by 'group-think' or influenced by hierarchy.
- 3. **Decentralisation**. Where the members of the crowd are from different specialities and areas of expertise so that they bring differing experiences and knowledge.
- 4. **Aggregation Mechanism**. Where the output from the crowd is brought together there needs to be some way of bringing the ideas together constructively.

By using the MOD intranet, remote groups may be accessed which, at least offers the possibility of incorporating 'Diverse Opinions'. The use of a Serious Game then allows players to test their perspectives in a representative synthetic environment. This 'play' combines thought and action with feedback to the player through the game scoring mechanism (Abt Ch1). The use of a Serious Game played anonymously as a single-player browser game allows the 'Independence' condition to be met. If it is accepted that the MOD is formed of many decentralised units comprising academia, scientific and operator groupings, then the third criteria may be met. Crowdsourced feedback for new designs then allows the HCI to be developed and then analysis of the data log files may provide an Aggregation Mechanism through a correlation study. This last point is the subject of further game play competitions following refinement of the imaging radar and electro-optic detection ranges and the seeding of random variables within the game.

Where the term 'user' is given the narrow definition of the 'equipment end user' then opportunities for HCI innovation may be missed. For pan-domain defence equipment such as Radar, Electronic Warfare or Unmanned Vehicles there will be decentralised expertise in land, air and maritime domains but not a cross-connection between them. User-centred design processes will not normally seek design opinions from outside the equipment end user community. By contrast, the use of Crowdsourced Serious Games for design becomes a form of corporate knowledge capture allowing expertise from across the enterprise to be sought.

The use of the intranet browser as a game platform gives an easily accessed method for dynamic graphical representation of defence systems' controls and displays. The MOD intranet has 120,000 users and a Serious Game has the potential to engage with a wide range of perspectives. As a default application on desk-top computers, the web browser requires no additional software to run a game other than the game files themselves. With computer coding part of the primary school national curriculum since 2014 (Department of Education), it may be expected that future employees will come with basic knowledge of web coding leading to a wider use of independently developed games.

This project shows that the Crowdsourced Design process using a Serious Game is effective in obtaining design ideas through engaging with a dispersed community. The innovative ideas for HCI realised within this project are being used and further data is now being gathered for usability assessment by correlation analysis.

References

Abt, Clark C. Serious Games. New York, The Viking Press, 1970

- Defence Lines Of Development Analysis With MODAF. UK Government, 2017, www.gov.uk/government/uploads/system/uploads/attachment_data/ file/36720/20090210 MODAFDLODAnalysis V1 0 U.pdf
- Department of Education, 2013, https://www.gov.uk/government/publications/national-curriculumin-england-computing-programmes-of-study/national-curriculum-in-england-computingprogrammes-of-study
- Gupta, Saransh. 2017, "A Day in the Life of A User Researcher". *The Ergonomist. Magazine of The Chartered Institute of Ergonomics & Human Factors*, (July-August 2017).
- Howe, Jeff. Crowdsourcing. London, Random House Business Books, 2009
- Lazar, Jonathan et al. *Research Methods In Human Computer Interaction*. Chichester, John Wiley & Sons Ltd, 2010
- Mitroff, Stephen R. et al. "What Can 1 Billion Trials Tell Us About Visual Search?" *Journal Of Experimental Psychology: Human Perception And Performance*, vol 41, no 1, 2015, pp1-5. American Psychological Association (APA), doi:10.1037/xhp0000012

Rigby, Scott, and Richard M Ryan. *Glued To Games*. Santa Barbara, Calif, Praeger, 2011 Sauro Jeff. and Lewis JR., 2016 *Quantifying the User Experience*. Morgan Kaufmann, USA. Surowiecki James. 2004 *The Wisdom of Crowds*, Abacus, UK