AAAI Technical Report WS-11-19

Towards a Nondisruptive, Practical, and Objective Automated Playtesting Process

Chek Tien Tan
Games Studio,
Human Centred Technology Design,
School of Software,
Faculty of Engineering and IT,
University of Technology, Sydney
chektien.tan@uts.edu.au

Andrew Johnston
Creativity and Cognition Studios,
Human Centred Technology Design,
School of Software,
Faculty of Engineering and IT,
University of Technology, Sydney
andrew.johnston@uts.edu.au

Abstract

Playtesting is the primary process that allows a game designer to access game quality. Current playtesting methods are often intrusive to play, involves much manual labor, and might not even portray the player's true feedback. This paper aims to alleviate these shortcomings by presenting the position that state of the art artificial intelligence techniques can construct automated playtesting systems that supplement or even substitute this process to a certain extent. Several potential research directions are proposed in this theme. A work-in-progress report is also included to demonstrate the conceptual feasibility of the potentials of this research area.

Introduction

A playtest describes the process in which a game is played by a participant with the primary purpose of obtaining feedback from the participant in order to resolve design flaws and improve the game in general. For the sake of clarity, this paper only refers to digital games.

It is well-known that playtesting plays a crucial role in the game design process. In a play-centric design process (Fullerton 2008), playtests are conducted in all stages of the process from conceptualization to after the game is shipped. It is only during playtests that the game designers get external feedback of their game quality. Hence it is important that playtests are conducted frequently enough to eliminate early mistakes in early stages as well as to get the game polished in later stages.

The problem with current playtesting methods is that they mostly follow traditional social research methods which involves much manual labor, is intrusive to play, and might not portray the player's true feedback (as will be elaborated in the next section). Firstly, the time and money needed to conduct a playtest session will affect the producer's budget and might deter the team from conducting enough playtests. Secondly, if the designers probe the players with questions during play, it might interfere with their true experience. Thirdly, even if the designers record a video of the play sessions, the post-articulation of the experience might not be accurate, and double time is spent deciphering the videos manually.

Therefore, this paper proposes the employment of artificial intelligence (AI) techniques in alleviating these problems. The contributions of this paper includes

- highlight shortcomings in current playtesting methods,
- articulating potential research directions in automated playtesting, and
- presenting a work-in-progress report demonstrating the feasibility potentials of this research area.

In the remainder of this paper, the next section reviews current methods used in playtesting, then the section that follows articulates several potential research directions. The succeeding section then describes our current research progress and the last section concludes with suggestions on future work.

Current Playtesting Methods

There are a vast number of ways playtests are conducted but traditional social research methods like post-game reports and direct observations currently dominate. In-game data hooks are getting more common whilst physiological measurements are just starting to get noticed. The following subsections summarizes each type of method and reviews the advantages and disadvantages of each.

Post-game Reports

This category of methods includes interviews, open discussions and questionnaires. Interviews are oral quizzes given to individuals after they play the game. Open discussions are also oral but have a free format and normally conducted with groups of players that aims to facilitate more participation with each other. Questionnaires are fixed sets of written questions given to players after playing the game. The main advantage of these methods is the ease of implementation that allows for playtests of larger scales.

A shortcoming of these methods is that there might be hidden experiences that might not be captured by the preformulated questions. Open discussions alleviate this to a certain extent but still cannot capture everything that actually happened during play. These methods suffer from results that are skewed towards the players' memory of the experience. The findings are also very much dependent on what the players choose to report, as well as how the players want to report it.

Direct Observations

Observing players during play is the most straightforward method to conduct playtests. This method is popular as the actual reactions can be recorded by the observers. Video recordings is another way of implementing this method. Sometimes the players are also asked to provide ongoing commentary of their experiences during play. In comparison with the post-game reports this might be more accurate as the findings represent the actual experiences during the game and will not be totally biased towards the reports of the players.

One problem with direct observations is that player experiences might be affected by the presence of observers or recording tools. When asked to provide an ongoing commentary, the players' experiences might also be affected with this additional role imposed. Results might also be biased according to the observers' interpretation of the behaviors. Also, if video recordings are required, it might deter participants who have privacy concerns or are basically uncomfortable with others obtaining a video of them.

In-game Data Hooks

Modern video games sometimes also use data hooks in the game software to automatically track play statistics. Examples of data include the time taken to clear a level as well as the number of kills and deaths. The advantage of this method is that it provides a record of the activities during actual play, much like the direct observations method. Additionally, it allows for much

The main shortcoming of this method is that it is highly limited by the software actions available to the player in the virtual game world. Hence these "virtual observations" might not be representative of the actual experiences of the player in real life. For example, a player might take a long time to clear a level, but he might be smiling in real life having fun exploring more areas of the level.

Physiological Measurements

Methods in this category primarily uses specialized equipment to measure physical responses of players during play. This includes measuring heart rate, skin conductance level, electroencephalography, body temperature, pupil dilations, and facial expressions. Initial experimental use of skin conductance level for playtesting can be seen in Valve's talk at the Game Developers Conference 2011 [1].

The position taken by this paper is that they provide unbiased data of actual play that can be used in more objective analysis methods. However, use of these methods are generally scarce because of the equipment and personnel cost required to conduct these experiments. Much more research needs to be done to justify the viability of such methods as well.

Towards Automated Playtesting

As mentioned in the previous section, this paper holds the view that there exists vast potential in physiological methods. Other than the lacking in knowledge, these methods are however currently impractical to implement especially with budget and situational constraints.

However, amongst the many forms of measurable physiological aspects, facial expressions, voice and gestures seem to be the most accessible and promising. This is because:

- The advent of motion detection game consoles like Microsoft's Kinect [2] and Nintendo's 3DS [3], video and voice feeds are naturally incorporated into gameplay. This has the advantages of direct observations but eliminates intrusiveness and biased observer analysis.
- Webcams and microphones are relatively cheap. This eliminates the cost factor when compared to specialized physiological equipment.
- These capturing devices are commonly found on consumer desktops, laptops and mobile devices which allow for continuous playtesting to be incorporated after the game ships. This has similar advantages to in-game data hooks but eliminates the "virtual observation" shortcoming as described in the previous section.
- Fairly mature computer vision techniques exist in research literature with regards to facial expression (Buenaposada, Muñoz, and Baumela 2007), voice (Rabiner 1989) and gesture (Wu and Huang 1999) recognition. This provides confidence in utilizing these approaches.

The crucial step that is lacking is determining the feasibility in using face, voice and gesture data to answer playtest questions like useability, flow, and ultimately fun in games. This is where the research opportunities lie.

Current Work

As an initial proof of concept, we target motion detection games as this seems to be currently most suitable platform genre. As mentioned, it is the least intrusive due to video and voice recording being essential to the gameplay itself. In particular we will be using the Microsoft Kinect platform as an initial testbed. Using the Kinect hardware also has additional advantages because of the depth data available from the camera which allows for facial expression recognition

[1] http://www.valvesoftware.com/publications/2011/ValveBiofeedback-Ambinder.pdf

[2] http://www.xbox.com/en-US/kinect/

[3] http://www.nintendo.com/3ds

enhancements in the future. Moreover, facial expression detection has been shown to work well with noisy inputs from Microsoft's Kinect (Weise et al.). The initial approach and experimental plan is described as follows.

Approach

The proposed framework is a supervised learning system that maps the physiological features (face, voice and gesture features) to playtest questions (for example useability, flow, and ultimately fun in games). An overview of our framework is as shown in Figure 1. We will also initially limit our scope to only facial expressions as it intuitively contains the most information. Moreover, voice and gestures might not be always available during gameplay.

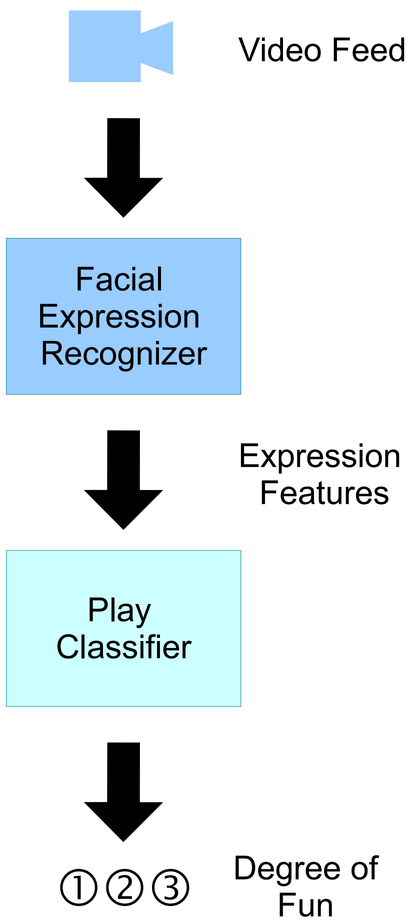

Figure 1: Overview of the automated playtesting system framework. The player's captured video is fed into the face recognizer which outputs coded expressions and durations as features. Then these features are input into the play classifier which determines the degree of fun, which, at this initial stage, is simply 3 classes.

For the duration of play for each player, the system will continuously detect important common expressions like happiness, sadness and anger as depicted in the Facial Action Coding System (FACS) (Ekman and Friesen 1978), as well as the duration of these expressions. These features will then be inputs into the classifier in which an answer class will be determined as the output. Then the player will be required to indicate the actual degree of fun they had. This will act as the supervision information used to train the classifier. As training convergence is reached, the system can then be used reliably to automatically obtain playtest results.

At this proof of concept stage, we have initially chose a neural network as the play classifier due to its success in prior facial expression work (Dailey et al. 2002). However, it is too early to decide on a classifier that is best suited for this novel task and we will leave it as future work which will be discussed in the conclusion section.

Experimental Plan

We are currently in the process of developing a Kinect game with game balance mechanics that can be skewed. Once that is completed, the automated learning system will be incorporated within the game. The game will then be available in our research lab for any students to play. The condition to play is an agreement to participate in our study. The deterrence is minimal as the only task for the player is to indicate the level of fun at the end of the game.

This process will be ran until we obtain about 100 participants' data. If necessary, the game balance will be tweaked to collect a wider range of results (for example if we receive too many players indicating maximal fun). There is no need to segregate the training stage as it is built into the process, and the system simply learns continuously until convergence and then performs maximally. The results will then be plotted for an analysis of the convergence point as well as to cater for further iterations of the system.

Conclusion

This paper has presented a position that more research needs to be performed in the area of automated methods of playtesting. The advantages to pursuing this research is multi-fold. Other than not interfering with the player during play, these automated methods can also eliminate the manual labor involved, collect more accurately unbiased data. In addition, these techniques can also be applied to adaptive gameplay, by making the game adapt to different user moods. The prospects of pursuing this research is hence advantageous to the game design community.

At this point however, the goal of this research proposal is not to totally replace social research methods of playtesting. It can be used as non-intrusive supplementary data that can enhance traditional methods like direct observations. It is also a way to enable better continuous feedback gathering in non-formal playtest sessions, like natural play situations in the players' own home. An analogy might be that this is similar to user agreements in sending bug reports which are common in general software applications. When compared to the reluctance of players agreeing to being video recorded for traditional playtest sessions, this automated way simply send post-processed statistical data and not the raw video feed, so there is no issue of any stranger seeing the players video.

In terms of limitations, situations in which this technology cannot be used are in games which involve facial expressions, voice and gestures in the gameplay. Currently usage of voice and gestures are getting common but usage of facial expressions would be odd in most game genres and hence would be extremely rare. This is also the motivation in starting with facial expressions first. However, we do not rule of voice and gestures totally as they can still be used in game interludes like loading screens, and of course, for games do not use voice and gestures in gameplay, they can be fully utilized if capture is allowed.

It should also be noted that motion-detection games are just the most appropriate starting points for this research. With the advent of novel multi-modal gaming platforms that use video and sound captures, this automated approach might also be used in other game platforms. Even for games which do not utilize these modes, camera and microphones are becoming increasingly mainstream on all forms of static and mobile computing devices, which means possibilities exist to motivate their use in playtesting in the future.

After the results of this work has been obtained and presented, a planned future work is to investigate the transferability of the learner system across game genres. Other than that, the comparative feasibility of different classifiers will also be investigated. Beyond that, the next step is to include voice and gesture features into the research project.

References

Buenaposada, J. M.; Muñoz, E.; and Baumela, L. 2007. Recognising facial expressions in video sequences. *Pattern Analysis and Applications* 11(1):101–116.

Dailey, M. N.; Cottrell, G. W.; Padgett, C.; and Adolphs, R. 2002. Empath: A neural network that categorizes facial expressions. *Journal of cognitive neuroscience* 1158–1173.

Ekman, P., and Friesen, W. 1978. *Facial Action Coding System: A Technique for the Measurement of Facial Movement.* Palo Alto: Consulting Psychologists Press.

Fullerton, T. 2008. *Game Design Workshop*. United States: Morgan Kaufmann, second edition.

Rabiner, L. R. 1989. A tutorial on hidden markov models and selected applications in speech recognition. In *Proceedings of the IEEE*, 257–286.

Weise, T.; Bouaziz, S.; Li, H.; and Pauly, M. Realtime Performance-Based Facial Animation. *lgg.epfl.ch*.

Wu, Y., and Huang, T. 1999. Vision-based gesture recognition: A review. *Gesture-based communication in human-computer interaction* 103–115.