Evaluating Engagement in Digital Narratives from Facial Data

Rui Huan
University of Glasgow
Glasgow, UK
r.huan.1@research.gla.ac.uk

ABSTRACT

Engagement researchers indicate that the engagement level of people in a narrative has an influence on people's subsequent story-related attitudes and beliefs[6], which helps psychologists understand people's social behaviours and personal experience. With the arrival of multimedia, the digital narrative combines multimedia features (e.g. varying images, music and voiceover) with traditional storytelling. Research on digital narratives has been widely used in helping students gain problem-solving and presentation skills [4] as well as supporting child psychologists investigating children's social understanding such as family/peer relationships [18] through completing their digital narratives. However, there is little study on the effect of multimedia features in digital narratives on the engagement level of people.

This research focuses on measuring the levels of engagement of people in digital narratives and specifically on understanding the media effect of digital narratives on people's engagement levels. Measurement tools are developed and validated through analyses of facial data from different age groups (children and young adults) in watching stories with different media features of digital narratives. Data sources used in this research include a questionnaire with Smileyometer scale[10] and the observation of each participant's facial behaviours.

CCS CONCEPTS

• **Human-centered computing → Human computer interaction (HCI)**;

KEYWORDS

Engagement; Automated measures; Digital narrative; MCAST

ACM Reference Format:
Rui Huan. 2017. Evaluating Engagement in Digital Narratives from Facial Data. In *Proceedings of 19th ACM International Conference on Multimodal Interaction (ICMI'17).* ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/3136755.3137036

1 INTRODUCTION

Recent studies [3, 17–20, 23, 24] have emphasised that engagement is a key factor in understanding a user's psychology and behaviour in many areas. For instance, researchers have been working on implementing a conversational agent that adapts conversations with a user according to the user's engagement level to improve naturalness in human-agent communications [20]. They have also been exploring automated recognition of student engagement which may help teachers constantly evaluate the level of their students' engagement to adjust the learning process appropriately [23]. In these studies, there is not a general definition of the term "engagement" and it is interpreted based on the purposes of the research. For example, it can be described as a person's willingness to participate in a task, a person's emotional attitude towards tasks and a person's focused attention as well as creative thinking.

One important area related to engagement which we would like to focus on is measuring the engagement levels in narratives. A narrative is composed of the unique sequence of events, mental states and occurrences that involves human beings as characters or actors, which can be "real" or "imaginary" [12]. With the arrival of multimedia, the idea of merging traditional storytelling with digital tools is now common. *Digital narrative* is a narrative tool with which life stories are reconstructed using multimedia features such as varying images, text, photos, audio, voiceover, hypertext, animation and video[10, 16]. It can help students gain problem-solving and presentation skills through completing their digital narratives [4] as well as support child psychologists investigating children's social understanding [13, 18]. Although there are many studies for measuring people's engagement of digital narratives and researchers already indicated many media features for designing digital narratives, we have not found any study on the effect of these media features of digital narratives on the level of engagement of people.

Therefore, the first problem of my research is to evaluate if people's levels of engagement can be affected by different media types of digital narratives. The test on which we have based our initial work is the Manchester Child Attachment Story Task (MCAST), which is a structured doll play methodology by using short stories to assess the Attachment status of children [8]. Children in this study were asked to listen the beginning of a story with a situation of specific anxiety and distress (e.g. The child awakes at night alone with a nightmare.) and then asked to act out what happens in the next part of the story with symbolic dolls. A key feature of this task is that the child should be repeatedly engaged in the distress situation until he/she is generally able to complete the story in a spontaneous play. An engaged child means that he/she focuses on increasing attention to play and materials as well as feeling empathy with the symbolic dolls in the story. MCAST assessors can evaluate child attachment status based on the story he/she completed and the child's behaviours [8]. In order to improve children's engagement levels, I chose several media types of digital narratives to display MCAST stories, including different voiceover (e.g. female/male voice), animation and video clips recorded by ourselves (i.e. an actor holds two dolls and displays MCAST stories).

Meanwhile, the second problem of my research is how to measure people's levels of engagement automatically. Non-verbal behaviours such as eye movement, facial expression and gesture have been widely used to measure engagement. In many situations, these kinds of behaviours are relatively easy to collect. For example, not looking at the TV can be a good indicator of low engagement while looking at it can be recognised as high engagement in a viewing task [9]. I will explore a set of tools that are developed and validated through analyses of facial data from children when watching digital narratives in different media types. Several face features will be extracted and used them to automatically predict children's engagement levels of MCAST stories. Then, to develop a set of general tools for measuring engagement or to gain a deeper understanding of engagement, an experiment with young adult participants will be conducted.

To summarise, this research focuses on understanding the media effect of digital narratives on people's engagement levels and specifically on developing a set of tools measure the engagement. These tools are developed and validated through analyses of facial data from different age groups (children and young adults). The main research questions for my research are:

RQ1: What media types affect engagement in digital narratives in children and young adults? There are three types of digital narrative to be studied: animation, video clips and voiceover.

RQ2: What methods should be used to measure engagement in children and young adults?

In the following section, three popular tools to measure engagement will be described. Then the work done to date is presented along with plans for the next stage of this study.

2 METHODS

Based on the purpose of the research, a critical issue is choosing a proper method that can measure engagement. There are typically three methods for measuring engagement: self-reports, observation and physiological measures. The self-report is a method commonly used in social science which allows individuals to express their attitudes, feelings, beliefs about a narrative [11, 21]. It includes two types: verbal self-reports (interviews or surveys) and non-verbal self-reports (questionnaire). Another popular method to measure engagement relies on observations from some external observers. They may also be asked to follow checklists for objective measures that are supposed to indicate engagement. For instance, engagement in MCAST is measured by an assessor's observation of facial expressions, gestures, etc. [8]. Automated measures are based on the timing and participants' physiological information produced by heart, brain and skin [19, 23]. Another kind of automatic engagement recognition is based on computer vision, which provides an automatic estimation of engagement by analysing cues from the face and gestures [5, 7, 19, 23]. For instance, facial behaviour consists of facial landmark motion, head orientation and motion, facial expressions and eye gaze [2] and its analysis have been used in different applications to facilitate human-computer interaction [19], education [23] and entertainment [5]. More information about how can I use these method to measure the engagement levels will be introduced in the following section.

Figure 1: Smileyometer Scale [10]

2.1 Self-report measures

The self-report represents a robust, efficient and easy to implement approach for collecting valid, reliable data for assessing engagement in multiple areas such as video games-based environment and education [14, 19, 21, 24]. Researchers have suggested that the self-report differs from external annotation and automated measures as it provides a participant's perspective of a system based on his/her mental state to help researchers understand the participantâĂŹs engagement [19, 21].

We designed a questionnaire with the Smileyometer [10] instrument in which pictorial representations of emotional faces were used to depict the level of agreement with a question (based on a 1 to 5 Likert scale) as shown in Figure 1. The Smileyometer has been widely applied in many child studies as it is easy to complete and requires no writing on behalf of the children. In my questionnaire, there were ten questions to indicate participants' attitude towards the content and different media types (e.g. I liked the male/female voice used on this video) of the story. I will combine results from questionnaire with automated measures so that to gain a better understanding in children's engagement.

2.2 Data Annotation

A team of labellers viewed and rated the recorded videos for engagement. All recordings for participant's performance were split into 10-second segments, and the labellers were asked to give a single number to rate each video clip. In such case, I defined the engagement level as the extent to which the participant is focused on the story being shown on the screen. An annotation scheme was shown in Table 1 to distinguish four different levels of engagement, ranging from no engagement to full engagement.

The labellers were instructed to label engagement only based on appearance. They did not need to try to infer what a participant was "really" thinking at that time because this left the labelling problem too open-ended. For instance, a subject who looks very relaxed and has no expression can hardly be labelled as highly or fully engaged. Labellers have independently labelled the data so that each segment has been annotated by at least two labellers. At the end of the annotation process, I had each participant's level of engagement for every clip of video. Segments with consensual rating were firstly used to be compared primary eye-tracking features among different levels of engagement. Then I analysed the clips with low but not same levels (i.e. one labeller gave a label of 1 and another gave a label of 2, labelled as 1-2) to capture differences between low levels of engagement. Also, a same analysis with high neighbour levels (i.e. labelled as 3-4) was conducted. If the minimum and maximum label given to one segment differed by more than 1 (e.g., one labeller

Table 1: The engagement level annotation

Level	*Name*	*Characteristic*
1	No engaged	e.g. looking away from screen and focusing on something other than the video; eyes completely closed over 3 seconds
2	Rarely engaged	e.g. clearly not "into" the task; paying attention to something else (e.g.camera), but sometimes focusing on the video
3	Highly engaged	e.g. good enough to proceed to the task; participant requires no admonition to "stay on task"
4	Fully engaged	e.g. good quality engagement; participant could be "commended" for his/her level of engagement in task
X		The clip was very unclear or contains no person at all.

gave a label of 1 and another gave a label of 3), then the clip was discarded.

2.3 Automated measures

Facial data are collected while the participant is watching the stories. Using eye-tracking [22] and facial expression [2] analysis, we can collect some data without disrupting participants' attention from the narrative.

2.3.1 Eye-tracking Measures. Eye-tracking is being used as one method of measuring child engagement to investigate the first research question. Information gathered from eye-tracking technique can help measure engagement in the field of user-system interactions [1, 15, 20]. For instance, researchers have investigated fixations in eye movement to measure the degree of engagement to improve naturalness in human-agent communication according to the participant's level of engagement [20]. Other research has identified several features of eye-movements: fixations per area of interest (the number of fixations on a particular area); fixation duration; gaze/fixation clusters [16, 22]. In my research, I will use these features to measure childrenâĂŹs engagement with digital narratives.

2.3.2 Facial Expression Recognition. Besides eye-tracking, I also focus on recognition of detailed facial movements as a method for evaluating engagement. In MCAST assessments, engagement is measured by an assessor's observation of facial expressions, gestures, etc. [8]. However, human observation is not accurate and time-consuming because the assessors may lose some details of childrenâĂŹs facial expression and they may need to check the video recordings several times. Therefore, I would like to use the Facial Action Unit (AU) classification [2, 19, 23], which measures the intensity of over 40 distinct facial muscles to describe people's facial expression frame-by-frame.

3 THE WORK DONE TO DATE

My first study has investigated the effect of voiceover quality on child engagement using animations of four MCAST stories as the source material. I hypothesised the voiceover in digital storytelling affects the engagement level of people because it may increase mood state around a specific emotion (e.g. distress in MCAST stories [8]) represented in the story. There were four conditions: monotonous male voice, expressive male voice, monotonous female voice and expressive female voice. The aim of this study was to measure if children's level of engagement can be affected by different voiceover quality in a digital narrative and to develop if two automated facial measures (see Section 2.3) can be used to evaluate child engagement.

The study was run on 40 children from primary schools located in Glasgow area. Children in this study were guided by an on-screen avatar to watch the beginning of the MCAST story and were asked to act out what happens in the next part of the story. During the watching session, data of children's eye-movements and facial expression were collected by a Tobii eye tracker and a Logitech web camera so that we can detect if the child is engaged enough in this story to complete/represent it. The questionnaire (Section 2.1) was used after the child finished the story.

I have followed two ways to analyse the relationship between engagement and gaze behaviours. Firstly, gaze data has been analysed to compare the primary eye-tracking features (see Section 2.3.1) among different levels of engagement based on results from human annotation. Secondly, I conducted classification with two classes: class A for the low engagement levels (i.e. including no engaged at all and rarely engaged) and class B for high engagement levels (i.e. including engaged in task and fully engaged). The aim of the classifier is to estimate the level of engagement in a video of a child watching a story to understand whether the child is engaged or not.

My results show that gaze behaviour differs during distinct levels of engagement in digital narratives from comparing primary eye-tracking features. With high levels of engagement from human annotation, the mean fixation duration increases and there is also an increase in the count of fixation durations. Fixation duration also strongly suggests a significant difference in the gender of the speaker. The female voice used draws more attention from participants than our male voice, which results in a longer mean fixation duration. However, for the voice quality (i.e. monotonous and expressive voices), there is no significant effect of the male's quality voice on engagement and a slight difference between female monotonous and expressive voice. We also show that gaze behaviours contain information related to levels of engagement. It was demonstrated by creating an SVM classifier [3] using fixation features we can predict engagement correctly in 70% of cases, which is a promising result for further applications of automatic engagement recognition. In the SVM classification task, participants are more engaged in stories with female animated voice displayed than other voice conditions.

4 THE PLAN FOR NEXT STAGE

The next step in my research is to analyse the relationship between engagement and facial expressions. For eye-tracking measures, we ask labellers to give a single number of the engagement level to label 10-second segments because eye movement is difficult to capture from static frame judgments. However, this time, labellers will be asked to view static images and give a single number to rate each frame independently as Facial Action Unit (AU) classification [2, 19, 23] is measured people's facial expression frame-by-frame. The labellers will still be instructed to label engagement only according to participants' appearance. For instance, if a subject blinked, then he/she would be labelled as no engaged (Engagement = 1) in that frame because he/she would appear to be no engagement at that moment. A tool called OpenFace [2] will be used to analyse children's facial expression. It is a fully open source real-time facial behaviour analysis tool for computer vision, machine learning and affective computing communities. After the facial expression recognition, I will analyse children's attitude towards the content and media type of story from the questionnaire, which can reflect their emotional and cognitive engagement.

The next study will then be conducted with young adults as participants to detect if adult engagement can be affected by different voice quality. We will try to use same measurement methods from the first experiment and find out if these measurement tools are also successful in measuring engagement in young adults.

In order to gain a deeper understanding of engagement, I will also test if video formats (i.e. animated video and live action video) of digital narrative influence people's levels of engagement. Previous research indicated that animation can attract users' attention to a certain part of the screen. Once viewers' attention is captured, they can be quickly focused on the item and may enhance their memory of the emphasised content. I suppose that different formats of video (animated video and live action videos) affect people's engagement levels and two studies with different age groups will be conducted. Lastly, we will compare the methods we used between children and young adults to find out if the measurement tools are successful in measuring engagement in the two different groups.

REFERENCES

[1] Elgiz Bal, Emily Harden, Damon Lamb, Amy Vaughan Van Hecke, John W. Denver, and Stephen W. Porges. 2010. Emotion recognition in children with autism spectrum disorders: Relations to eye gaze and autonomic state. *Journal of Autism and Developmental Disorders* 40, 3 (2010), 358–370. https://doi.org/10.1007/s10803-009-0884-3

[2] Tadas Baltrušaitis, Peter Robinson, and Louis Philippe Morency. 2016. OpenFace: An open source facial behavior analysis toolkit. *2016 IEEE Winter Conference on Applications of Computer Vision, WACV 2016* (2016). https://doi.org/10.1109/WACV.2016.7477553

[3] Roman Bednarik, Shahram Eivazi, and Michal Hradis. 2012. Gaze and conversational engagement in multiparty video conversation. *Proceedings of the 4th Workshop on Eye Gaze in Intelligent Human Machine Interaction - Gaze-In '12* (2012), 1–6. https://doi.org/10.1145/2401836.2401846

[4] Robin R Bernard. 2011. The Educational Uses of Digital Storytelling. (2011). https://digitalliteracyintheclassroom.pbworks.com/f/Educ-Uses-DS.pdf

[5] Paris Mavromoustakos Blom, Sander Bakkes, Chek Tien Tan, Shimon Whiteson, Diederik Roijers, Roberto Valenti, and Theo Gevers. 2014. Towards Personalised Gaming via Facial Expression Recognition. In *Proceedings of the Tenth AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment (AIIDE'14)*. AAAI Press, 30–36. http://dl.acm.org/citation.cfm?id=3014752.3014758

[6] Rick Busselle and Helena Bilandzic. 2009. Measuring narrative engagement. *Media Psychology* 12, 4 (2009), 321–347. https://doi.org/10.1080/15213260903287259

[7] Joseph F Grafsgaard, Joseph B Wiggins, Kristy Elizabeth Boyer, Eric N Wiebe, and James C Lester. 2013. Automatically recognizing facial expression: Predicting engagement and frustration. *International Conference on Educational Data Mining* (2013).

[8] Jonathan Green, Charlie Stanley, Vicky Smith, and Ruth Goldwyn. 2000. A new method of evaluating attachment representations in young school-age children : The Manchester Child Attachment Story Task. *Attachment & Human Development* 2, 1 (2000), 48–70. https://doi.org/10.1080/146167300361318

[9] Javier Hernandez, Zicheng Liu, Geoff Hulten, Dave Debarr, Kyle Krum, and Zhengyou Zhang. 2013. Measuring the engagement level of TV viewers. *2013 10th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition, FG 2013* (2013). https://doi.org/10.1109/FG.2013.6553742

[10] Arafeh Karimi and Yan Peng Lim. 2010. Children, engagement and enjoyment in digital narrative, In Proceedings ascilite Sydney 2010. *ASCILITE 2010 - The Australasian Society for Computers in Learning in Tertiary Education*, 475–483.

[11] Mounia Lalmas, Heather L. O'Brien, and Elad Yom-Tov. 2013. Measuring User Engagement. (2013). https://doi.org/10.2200/S00605ED1V01Y201410ICR038

[12] J Lambert. 2002. *Digital Storytelling: Capturing Lives, Creating Community.* Digital Diner Press, Berkeley, CA. 11 pages.

[13] Umasschi Bers Marina and Cassell Justine. 1998. Interactive Storytelling Systems for children. *Journal of Interactive Learning Research* 9, 2 (1998), 183–215. http://www.media.mit.edu/gnl/publications/JILR-sept-98.pdf

[14] Kevin Marshall, Gavin Wood, Janet C Read, Lana Yarosh Svetlana, Madeline Balaam, and Jung-Joo Lee. 2015. Supporting Children to Engage in Play for Wellbeing. In *CHI'15 Extended Abstracts*. Seoul, Republic of Korea, 2445–2448. https://doi.org/10.1145/2702613.2702658

[15] Rhonda N. Mcewen and Adam K. Dubé. 2015. Engaging or Distracting : Children's Tablet Computer Use in Education. 18, 4 (2015), 9–23.

[16] Daniel Meadows. 2003. Digital Storytelling: research-based practice in new media. *Visual Communication* 2, 2 (2003), 189–193. https://doi.org/10.1177/1470357203002002004

[17] Brian W. Miller. 2015. Using reading times and eye-movements to measure cognitive engagement. *Educational Psychologist* 50, 1 (2015), 31–42. https://doi.org/10.1080/00461520.2015.1004068

[18] Helen Minnis, Warren Read, Brenda Connolly, Amanda Burston, Tanja-Sabine Schumm, Suren Putter-Lareman, and Jonathan Green. 2010. The Computerised Manchester Child Attachment Story Task: a novel medium for assessing attachment patterns. *International Journal of Methods in Psychiatric Research* 19, 4 (2010), 233–242. https://doi.org/10.1002/mpr

[19] Hamed Monkaresi, Nigel Bosch, Rafael A. Calvo, and Sidney K. D'Mello. 2017. Automated Detection of Engagement Using Video-Based Estimation of Facial Expressions and Heart Rate. *IEEE Transactions on Affective Computing* 8, 1 (2017), 15–28. https://doi.org/10.1109/TAFFC.2016.2515084

[20] Yukiko I. Nakano and Ryo Ishii. 2010. Estimating User's Engagement from Eye-gaze Behaviors in Human-agent Conversations. In *Proceedings of the 15th International Conference on Intelligent User Interfaces (IUI '10)*. ACM, New York, NY, USA, 139–148. https://doi.org/10.1145/1719970.1719990

[21] Heather L. O'Brien, Elaine G. Toms, E. Kevin Kelloway, and Elizabeth Kelley. 2009. Developing and evaluating a reliable measure of user engagement. In *Proceedings of the American Society for Information Science and Technology*, Vol. 45. 1–10. https://doi.org/10.1002/meet.2008.1450450258

[22] Alex Poole and Linden J Ball. 2005. Eye Tracking in Human-Computer Interaction and Usability Research: Current Status and Future Prospects. *Encyclopedia of Human-Computer Interaction* (2005), 211–219. https://doi.org/10.4018/978-1-59140-562-7

[23] Jacob Whitehill, Zewelanji Serpell, Yi-Ching Lin, Aysha Foster, and Javier R. Movellan. 2014. The faces of engagement: Automatic recognition of student engagement from facial expressions. *IEEE Transactions on Affective Computing* 5, 1 (2014), 86–98. https://doi.org/10.1109/TAFFC.2014.2316163

[24] Eric N. Wiebe, Allison Lamb, Megan Hardy, and David Sharek. 2014. Measuring engagement in video game-based environments: Investigation of the User Engagement Scale. *Computers in Human Behavior* 32 (2014), 123–132. https://doi.org/10.1016/j.chb.2013.12.001