# Inferring User Engagement from Interaction Data

**Jonathan Carlton**
The University of Manchester, UK
BBC Research & Development, UK
jonathan.carlton@bbc.co.uk

**Andy Brown**
BBC Research & Development, UK
andy.brown01@bbc.co.uk

**Caroline Jay**
The University of Manchester, UK
caroline.jay@manchester.ac.uk

**John Keane**
The University of Manchester, UK
john.keane@manchester.ac.uk

## ABSTRACT

This paper presents preliminary results of a study designed to quantify users' engagement levels with interactive media content, through self-reported measures and interaction data. The broad hypothesis of the study is that interaction data can be used to predict the level of engagement felt by the user.

The challenge addressed in this work is to explore the effectiveness of interaction data to act as a proxy for engagement levels and reveal what that data shows about engagement with media content. Preliminary results suggest several interesting insights about participants engagement and behaviour. Crucially, temporal statistics support the hypothesis that the participant making use of the controls in the interactive, video-based experience positively correlates with higher engagement.

## CCS CONCEPTS

• **Information systems** → **Web mining**; • **Human-centered computing** → **Web-based interaction**; *User models*; *User studies*;

Permission to make digital or hard copies of part or all 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 third-party components of this work must be honored. For all other uses, contact the owner/author(s).

*CHI'19 Extended Abstracts, May 4–9, 2019, Glasgow, Scotland UK*

© 2019 Copyright held by the owner/author(s).

ACM ISBN 978-1-4503-5971-9/19/05.
https://doi.org/10.1145/3290607.3313009

## KEYWORDS

Interaction Data; User Engagement; Click; Media

**ACM Reference Format:**

Jonathan Carlton, Andy Brown, Caroline Jay, and John Keane. 2019. Inferring User Engagement from Interaction Data. In *CHI Conference on Human Factors in Computing Systems Extended Abstracts (CHI'19 Extended Abstracts), May 4–9, 2019, Glasgow, Scotland UK*. ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3290607.3313009


## INTRODUCTION

Creating media experiences that resonate with an audience poses design and engineering challenges for producers, who need to understand what engages audiences, what constitutes good user interface design, and how this varies between individuals. To address these challenges, and determine where to target production resources, one approach is to gather information from current experiences, and feed resulting insights into future productions. Collecting and analysing interaction data in particular would enables us to glean insights into user behaviours and form a clearer understanding of their engagement.

This paper considers preliminary results of a study designed to measure users' levels of engagement with a novel interactive and adaptive media experience – *Make-along: Origami Frog*, created by the BBC – using user-generated data. Both self-reported levels of user engagement and interaction data (cursor movement, touch gestures, keystrokes, and application-level button clicks) are recorded while using the *Make-along: Origami Frog*.

The paper is structured in the following way: *Background* presents related and previous work, the study components are presented in *Study Design*, while *Preliminary Results* discusses the collected data and an initial analysis, and finally *Conclusion & Future Work* concludes and presents future work.

### Sidebar 1: Research Questions

**R1** Do users differ between low, medium, and high levels of engagement, when considering summary statistics about their interactions with *Make-along: Origami Frog*?

**R2** How do interaction sequences differ between low, medium, and high levels of user engagement?

**R3** How can sequences of interactions effectively differentiate between low, medium, and high levels of user engagement?

## BACKGROUND

User engagement can be thought of as a user's cognitive involvement with a technology which generally relates to the quality of their experience [6]. The User Engagement Scale (UES) [7] is a survey frequently used to measure user engagement. UES measures the multi-dimensional nature of engagement by presenting the participant with a set of statements that measures four factors: focused attention, perceived usability, aesthetics, and reward. Interaction data can also be used to measure engagement, such as in [4] where scroll data is used to measure and predict future levels of engagement with news articles, where the depth of the user's reading plus their interactions with an article are used to determine the level of engagement.

Similarly in [1], mouse cursor data is used as a measure of engagement with direct displays within search engine result pages. A combination of UES and log data is used in [10] to predict users'

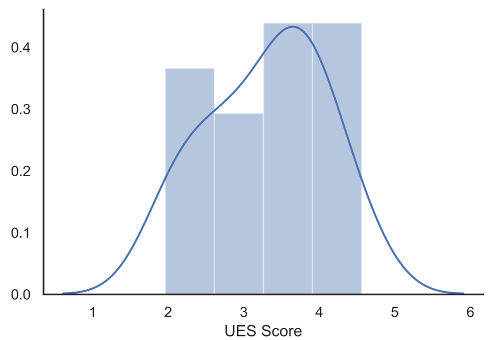
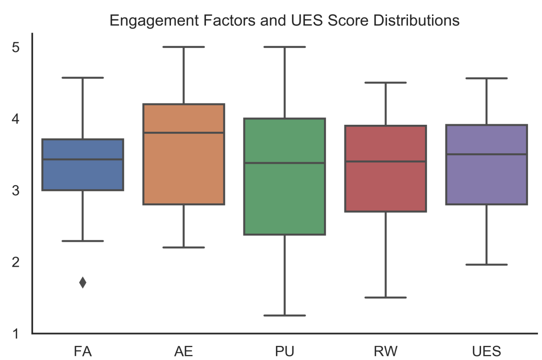



**Figure 1: The distribution of UES scores for the participants.**

**Figure 2: Focused Attention (FA), Aesthetic Appeal (AE) Perceived Usability (PU), Reward (RW), and UES score distributions.**




perceived engagement in a search-based environment, and they find that time- and query-related features (from logs) perform best in predicting engagement.

We have previously performed secondary analysis on interaction data collected in a separate study, which uncovered indications of differences in behaviour between groups of users and how engaged they were [2, 3]. However, this dataset did not contain ground truth about engagement levels. The current study aims to address this issue by collecting interaction data from users for whom we have established self-perceived engagement levels.

The generalisability of measuring engagement from interaction data across media-types is a long-term goal, but a first step is explore a single media instance. As such, the study and preliminary results presented here are an attempt to provide a deeper understanding of how interaction relates to engagement levels while using the video-based, interactive media experience *Make-along: Origami Frog* (an adaptive experience where users are guided through the steps to create a frog).

## STUDY DESIGN

The aim of this work is to test the broad hypothesis that: *interaction data can be used to predict the level of engagement felt by the user*. In order to test the hypothesis, three research questions will be explored (see Sidebar 1). To ensure ecological validity, the study was designed to be remote with the participants free to complete it at their leisure, and consisted of three components: a demographic survey, the origami experience itself, and the UES. To ensure that all components worked correctly and that the study was clear to participants, two separate pilots were run. In each pilot, two participants performed the study tasks, and their feedback allowed refinement of the process.

### Data

Four datasets were collected during the study: demographic, low-level interactions (including cursor movements and keystrokes), application-level (interface button clicks), and Likert scale UES data. The demographic data was collected to sample the population taking part in the study; seven questions were asked: the participants' age, gender, level of education, employment status, technology competence, previous origami making experience, and preferred method of consuming video-based content. Built-in analytics for the experience recorded the application-level button clicks, which log the following data when a participant clicks on any button present in the interface: *userid* - identification string, *timestamp* - time of event at a millisecond granularity, *item* - type of event that occurred, *action* - the button that was clicked, and *message* - metadata about the event. Low-level data was collected only for study participants, using EvTrack [5] to collect all events that can be triggered while interacting with a website. As output, CSV files are produced containing the following data: *cursor* - whether a computer mouse or touch device is being used, *timestamp* - time of event at a millisecond granularity, *x* and *y-axis* cursor positions, *event* - browser's event name, *xpath* - target element, and

**Figure 3: The previous origami-making experience of participants compared to the recorded levels of engagement.**

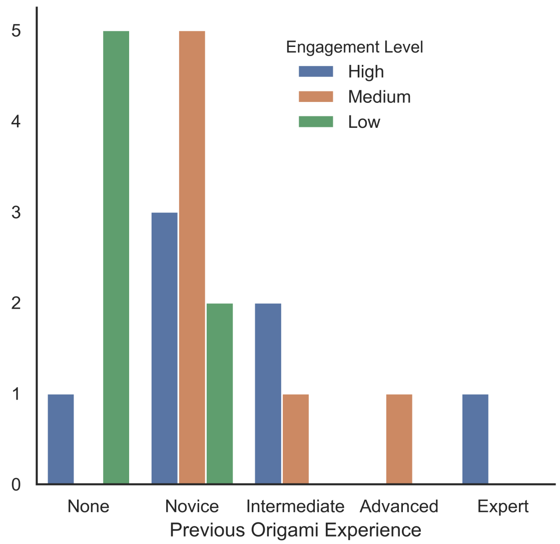



**Table 1: Overall statistics from the collected interaction data.**

| | Total # of events | $\mu$ | $\sigma$ |
|---|---|---|---|
| **AL Data** | 1,458 | 69.42 | 26.92 |
| **LL Data** | 135,127 | 6,439.61 | 3,972.11 |

*attrs* - elements attribute. The long-form version of UES [7] was used, consisting of 30 statements where participants rate their dis/agreement with each, resulting in Likert scale data. Engagement scores are calculated (the mean of the factor means) and percentiles created from these scores to group participants into categories of low, medium, and high levels of engagement [8]. A question was appended to the end of the survey, asking participants if they had completed making the frog.

## PRELIMINARY RESULTS

To date, 21 participants (11 females, 9 males, and 1 did not specify), recruited through the advertisement of the study, have taken part in the study; data collection is ongoing. A majority had a master's degree (42%), while others had either bachelor's (19%), doctoral (14%), college (19%), or high-school (4%) educations. They were primarily pursuing further education (57%) while the remainder were in full-time employment (43%). Technical competency was reported in the following way: 43% expert, 38% advanced, 9% intermediate, and 9% novice. They were also asked to rate their previous experience in making origami; 48% of participants were novices, 28% had no previous experience, 14% had intermediate experience, 5% were advanced and 5% were expert. Finally, laptops were reported to be the majority preference for media consumption (43%), with 19% for desktop computers, 14% for both television, 14% for mobile phone, and 10% for tablet devices.

### Initial Analysis of the Data

Participants are split into three quantiles of low, medium, and high engagement [8]. Those in the low quantile have scores below 2.89, while those in the high quantile have scores above 3.72, and participants in between the two are in the medium category. The participants recorded a mean UES score of 3.29 ($\sigma$ = 0.77), and there were an equal number of seven participants in each of the low, medium, and high engagement buckets. The distribution of UES scores recorded by participants is shown in Figure 1, and demonstrates that scores are trending more towards the higher end of the scale, indicating that on the whole participants were positively engaged with the content. Figure 2 shows the distribution of scores across all of the factors and overall UES score. Perceived Usability (PU) is the most distributed, while Focused Attention (FA) is the least indicating that participants are in agreement about the absorption but exhibited mixed feelings about usability. Two points are indicated by Figure 3: 1) no previous experience in origami overwhelmingly produced low engagement scores, and 2) those with a novice experience level have a good mix of engagement levels. These two points establish that participants who have some experience, and have an idea of what to expect when making origami, find the experience engaging. While those that have no previous experience found the process difficult, leading to a low engagement score - the data records these participants a low PU score ($\mu$ = 1.96, $\sigma$ = 0.56).

**Table 2: Pause count statistics from both interaction datasets: $\mu$ ($\sigma$)**

| | Application-Level | Low-Level |
|---|---|---|
| **Short** | 16.61 (9.06) | 35 (20.67) |
| **Medium** | 6.80 (4.04) | 13.95 (9.54) |
| **Long** | 7.57 (4.31) | 12.04 (5.65) |
| **Very Long** | 31.09 (9.51) | 29.71 (9.25) |

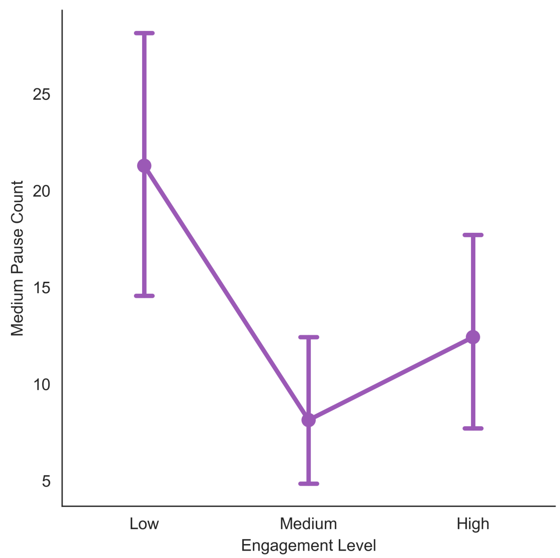



**Figure 4: An estimation of central tendency for the medium pause count metric (collected from low-level data) across the engagement levels ($p$ = 0.02).**

Descriptive and temporal statistics are extracted from the datasets, and the collected data breakdown is shown in Table 1. For the application-level data, 43 features were extracted: individual button click counts and relative frequencies, the total time spent on the experience, and the time between events (max/min, total, $\mu$, $\sigma$, and $\sigma^2$). The low-level data yielded 87 features, including cursor speed, distance, and acceleration metrics, event counts and relative frequencies, and the time between event metrics. Included in these features, and similar to [9], are pause-based metrics. These are counts, but are defined as follows: *short* (paused for between one and five seconds), *medium* (between six and 15), *long* (16 and 30), and *very long* (more than 30). We found on average participants spend 42 minutes ($\sigma$ = 16) completing the origami component of the study. *Very long* pauses are the majority pause-type for the application-level data, while in the low-level data there is a mix between the types (shown in Table 2).

The application-level data showed that a participant switching their camera angle view positively correlated with a high engagement score ($\rho$ = 0.5). In contrast, an increased usage of the repeat button (to replay the step) and the play/pause button is negatively correlated with the engagement scores (-0.41 and -0.44), establishing that those with low engagement scores find themselves pausing or replaying the step more frequently. The total number of events (in both datasets) typically increased when the participant's engagement score is lower ($\rho$ = -0.31 ($\mu$)), with *mousemove* events also increasing when the score is lower ($\rho$ = -0.30), suggesting a possible 'in-the-moment' observation of engagement. Pauses recorded in low-level data produced the strongest correlation, with both short and medium pauses negatively correlating (-0.38) suggesting how difficult participants found making the origami. Applying a one-way ANOVA to the medium pause count metric demonstrates a significant difference between engagement levels ($p$ = 0.02, Figure 4), and a similar difference is found for short pauses ($p$ = 0.04, Figure 5), with participants who experienced low engagement levels differing the most. These two results show that the proposed hypothesis may be supported by the interaction data and that it can highlight differences between engagement levels and the users that exhibit them when using online media.

## CONCLUSION & FUTURE WORK

The engagement scores have suggested insights into how participants find the origami experience, with a distributed set of scores being produced. Participants found the experience aesthetically pleasing, while those with no previous origami experience typically found the experience less engaging and usable. An initial analysis of the data has demonstrated that the proposed hypothesis could be supported and that there are differences between the three engagement levels, particularly from a temporal context. However, deeper analysis and further data collection is required to provide conclusive evidence to support the hypothesis. Exploring the importance of features (training a rule-based classification model) in the statistical data should yield further insights and identify predictive

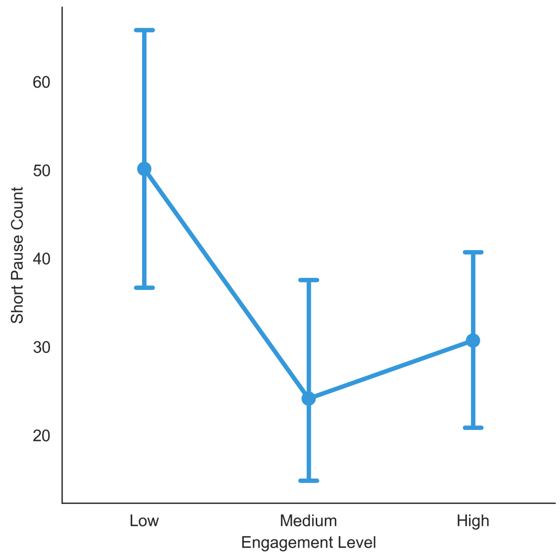



**Figure 5: An estimation of central tendency for the short pause count metric (collected from low-level data) across the engagement levels ($p = 0.04$).**

features for each engagement level. Further, the training of a regression model could provide additional insights about the relationship between correlated features and the engagement level. Addressing *R2* and *R3* is crucial; to find differences between interaction sequences an N-gram analysis will be performed. While for *R3*, exploring the feasibility of a sequence-based model to predict which category a user belongs to, similar to [9], is critical. More generally, identification of fine-grained behaviours in the data is key, such as cursor movements that indicate the participant is thinking about using an element on the interface, and will supplement both datasets. Finally, a model accurately trained on predicting user engagement levels can be deployed in a real-time environment to nudge users towards a more engaging experience. While, through feature selection and model deployment, insights can be garnered to inform the creators of media about what is engaging to users.

## ACKNOWLEDGMENTS

Jonathan Carlton would like thank the EPSRC and the BBC for iCASE PhD funding support (16000156).

## REFERENCES

[1] I. Arapakis and L. A. Leiva. 2016. Predicting user engagement with direct displays using mouse cursor information. In *Proc of the 39th International ACM SIGIR Conference on Research and Development in Information Retrieval*. ACM, 599–608.

[2] J. Carlton, A. Brown, J. Keane, and C. Jay. 2018. Using Low-Level Interaction Data to Explore User Behaviour in Interactive-Media Experiences. In *Procs of the 11th International Conference: Measuring Behaviour 2018*.

[3] J. Carlton, J. Woodcock, A. Brown, C. Jay, and J. Keane. 2018. Identifying latent indicators of technical difficulties from interaction data. In *ACM KDD Workshop on Data Science, Journalism and Media (DSJM)*.

[4] D. Lagun and M. Lalmas. 2016. Understanding user attention and engagement in online news reading. In *Proc of the 9th ACM International Conference on Web Search and Data Mining*. ACM, 113–122.

[5] L. A. Leiva and R. Vivó. 2013. Web Browsing Behavior Analysis and Interactive Hypervideo. *ACM Trans on the Web* 7, 4 (2013).

[6] B. Miroglio and et. al. 2018. The Effect of Ad Blocking on User Engagement with the Web. In *Proc of the 2018 World Wide Web Conference*. 813–821.

[7] H.L. O'Brien, P. Cairns, and M. Hall. 2018. A practical approach to measuring user engagement with the refined user engagement scale (UES) and new UES short form. *International Journal of Human-Computer Studies* 112 (2018), 28–39.

[8] H.L. O'Brien and M. Lebow. 2013. Mixed-methods approach to measuring user experience in online news interactions. *Journal of the American Society for Information Science and Technology* 64, 8 (2013), 1543–1556.

[9] K. Williams and I. Zitouni. 2017. Does That Mean You're Happy? RNN-based Modeling of User Interaction Sequences to Detect Good Abandonment. In *Procs of the 2017 ACM on Conference on Information and Knowledge Management*. ACM, 727–736.

[10] M. Zhuang, G. Demartini, and E. G. Toms. 2017. Understanding engagement through search behaviour. In *Proc of the 2017 ACM on Conference on Information and Knowledge Management*. ACM, 1957–1966.