Experience Design: Video Without Faces Increases Engagement But Not Empathy

Carina E.I. Westling
School of Media Film & Music
University of Sussex
c.e.i.westling@sussex.ac.uk

James K. Ackah
Brighton & Sussex Medical School
Brighton BN1 9PX, UK
J.Ackah1@uni.bsms.ac.uk

Carlos P. Santos
Brighton & Sussex Medical School
Brighton BN1 9PX, UK
C.PaulinoDosSantos1@uni.bsms.ac.uk

Nachiappan Chockalingam
CSHER
Staffordshire University
n.chockalingam@staffs.ac.uk

Harry J. Witchel
Brighton & Sussex Medical School
Brighton BN1 9PX, UK
h.witchel@bsms.ac.uk
+44 1273 873 549

ABSTRACT
Counter to prior claims that empathy is required for higher levels of engagement in human-computer interaction, our team has previously found that, in an analysis of 844 stimulus presentations, empathy is sufficient for high engagement, but is not necessary. Here, we ran a carefully controlled study of human-computer interactions with musical stimuli — with and without visuals, and with and without recognizable people – to directly test whether we could design an engaging stimulus that did not elicit empathy, by avoiding human faces or personal interaction. We measured subjective responses by visual analogue scale and found that the faceless stimulus was as engaging as the face-containing stimulus, but much less empathy-provoking. Therefore, we propose that empathy and engagement be considered independently during interaction design, because they are not monotonically related.

CCS Concepts
• *Information systems~Multimedia information systems* • *Human-centered computing~HCI design and evaluation methods* • *Human-centered computing~Laboratory experiments*

Keywords
Engagement; empathy; multimodality; interest; faceless; OK Go; music video.

1. INTRODUCTION
Researchers have created a range of validated instruments (i.e. subjective questionnaires) to quantify the engagement of end-users during human computer interaction (HCI), including the different types of engagement [13, 10], immersion [7], presence, and others. Interactive learning systems [5] and gaming platforms [2, 3, 4, 9] have implemented these engagement measurements.

ECCE '16, September 05-08, 2016, Nottingham, United Kingdom
 ISBN 978-1-4503-4244-5/16/09...$15.00
DOI: http://dx.doi.org/10.1145/2970930.2970936

1.1 Empathy within Engagement
The composite state of engagement may cause or result from empathy (caring or emotional involvement), as well as influencing decision-making and driving interaction [13]. Brown and Cairns [2] propose that empathy is an essential precursor to immersion, the highest level of game engagement, and it has since been considered a key component of self-directed engagement with digital content [8]. We have previously defined engagement as a family of related cognitive states geared toward extended interaction and/or a purposeful outcome, operationalized by a collection of behaviors, none of which are absolutely necessary at a given point in time, including: attendance, attention, memory, caring, emotion, taking action, making an effort, and (like the exclusion in attention) inhibition of irrelevant activities [13], with intellectual and emotional engagements being seen as separable [13, 15, 16].

In a recent study our team showed that, in 844 stimulus presentations, subjective empathy could be disassociated from interest in some cases [17]. In those rare cases where empathy was high but interest was low, 5 of the 9 instances were related to two stimuli: one showed many people's faces, while the other was photographed from the first person vantage point of "being there" (from behind the person/subject).

	Multimodal Stimulus	Audio Only Stimulus
Paired Test A	Do What You Want – music video	Here It Goes Again – music with black screen
Paired Test B	Here It Goes Again –- music video	Do What You Want – music with black screen

Figure 1. Table of the two counterbalanced stimulus pairs that volunteers would experience. Each volunteer would experience either paired test A, or paired test B, but not both.

The current study investigated the effect on empathy and on subjectively reported engagement when comparing a faceless to a face-containing music video. We also wanted to test directly

whether empathy is truly a prerequisite to high engagement in HCI, as presented by Brown and Cairns [2]. Neuroscientists testing autistic spectrum adolescents have shown that activity in facial recognition regions of the brain is important for empathy [6]. Here we designed a set of four new stimuli for this series of experiments with either music and human faces (and bodies), or music without visuals, to test three hypotheses about empathy in interface design and its effect on engagement:

a) It is possible to design an engaging stimulus that does not elicit empathy,

b) Seeing faces contributes to increasing empathy, and

c) Empathy is not necessary for higher levels of engagement.

2. METHODS

2.1 Experimental volunteers and protocol

27 healthy volunteers, 16 female, age range 18-28, were recruited from the university community via advertisements and emails. Ethical approval was obtained from the local university ethics committee. Participants experienced a range of auditory, visual and audiovisual stimuli, each lasting 170 seconds. After each stimulus, participants completed a questionnaire. All stimuli were presented in a counterbalanced order (see Figure 1); for a complete description of the counterbalance and stimuli, see the supplementary material in [16]. The experimenters left the room prior to starting the stimulus, so the volunteer was alone during the presentation.

2.2 Stimuli and subjective rating scales

The stimuli were a pair of comparable audio tracks (see below), and the same audio tracks with added video. The goals when designing these stimuli were:

1) To create paired stimuli to allow for direct comparisons of stimuli on the same persons

2) The paired stimuli should not be identical, to prevent habituation/repetition effects, but

3) The paired stimuli had to be approximately equal in how exciting and engaging they were

4) The paired stimuli had to differ in terms of whether they contain faces, but not in terms of showing bodies and activities. That is, one of the videos had the band's faces completely covered by cloth.

For stimuli we chose two similarly engaging OK Go music videos: *Here It Goes Again* (HIGA, where the band dances on treadmills) [11] and Do What You Want (DWYW) – Wallpaper Version [12], where the band and acrobats are completely covered in wallpaper (see design goal 4). These music videos elicit nearly identical amounts of interest and engagement (design goal 3, see Results), yet they are different enough to not elicit habituation/repetition effects (design goal 2); as a result they fulfill design goal 1, and can be tested on the same person. Each video was cut to 170 seconds, and a version with a blank, black screen was created for each video. Each volunteer was randomized into one of two paired test conditions, with each condition having a multimodal stimulus and an audio-only stimulus (see Fig. 1). The order that the two music stimuli were presented to the volunteers was counterbalanced (i.e. half the participants experienced paired condition 1, and of those people, half experienced the multimodal stimulus first, while the other half experienced the audio-only stimulus first).

Each stimulus was preceded with 45 seconds of white noise and 'TV snow' to establish a baseline. Each stimulus was rated immediately after watching, using a questionnaire with 6 adjective statements to be rated on a visual analogue scale (VAS), subdivided into 10 steps from 0 (not at all) to 100 (extremely). The VAS statements were: "I felt interested", "I felt bored", "I was totally engaged by the experience", "I wanted it to end earlier", "I was engrossed by the experience", and "I felt empathy or emotional attachment to what I saw". All statistics reported here are paired T tests, regressions or ANOVAs calculated in Matlab. Box-and-whisker plots were made in Matlab.

3. RESULTS

3.1 Comparably engaging music and music videos

When comparing the responses to two versions of the same stimulus in the same person (e.g. music with vs. without video) there is a likelihood of habituation (with increased boredom) over time [1]. Our team has previously demonstrated that volunteers' aversion-related movement patterns increase (i.e. non-instrumental movements increase) as time progresses while watching the same stimulus. Thus, instead of using two versions of the same stimulus (audio only vs. video + audio), we started with two equally interesting (i.e. similar but not identical) OK Go music videos, which were *Here It Goes Again (on treadmills)* and *Do What You Want (wallpaper version)*. There were four stimuli for this music experiment, but each volunteer was only exposed to two (see Fig. 1). Each volunteer saw/heard each of the songs only once. We compared the elicited subjective engagement VAS responses using a linear regression with song and modality as independent variables (table 1). The model was significant (R-squared = 0.39, $F(3,55) = 15.4$, $P < 0.0001$); as shown in the box plot (Fig. 2, right), the beta for song was not significant, but the beta for modality (video) was.

Table 1. Linear regression model of "engagement" rating vs. song and modality.

ENGAGEMENT	CI (-95%)	Beta	CI (+95%)
HIGA (vs. DWYW)	- 9.61	0.81	11.23
Video (vs. music only)	18.39	28.81	39.23
Constant	22.66	31.89	41.12

Thus, the music pieces themselves were equally engaging, but seeing the videos (either one) strongly enhanced engagement.

3.2 Adding video (with/without faces): effects on empathy

To determine whether adding video (i.e. multimodality) also affects empathy (which has been defined as a prerequisite for engagement [2]), we compared the VAS empathy responses using a regression of song and modality, which resulted in a statistically significant model (R-squared = 0.12, $F(3,55) = 3.69$, $P < 0.05$).

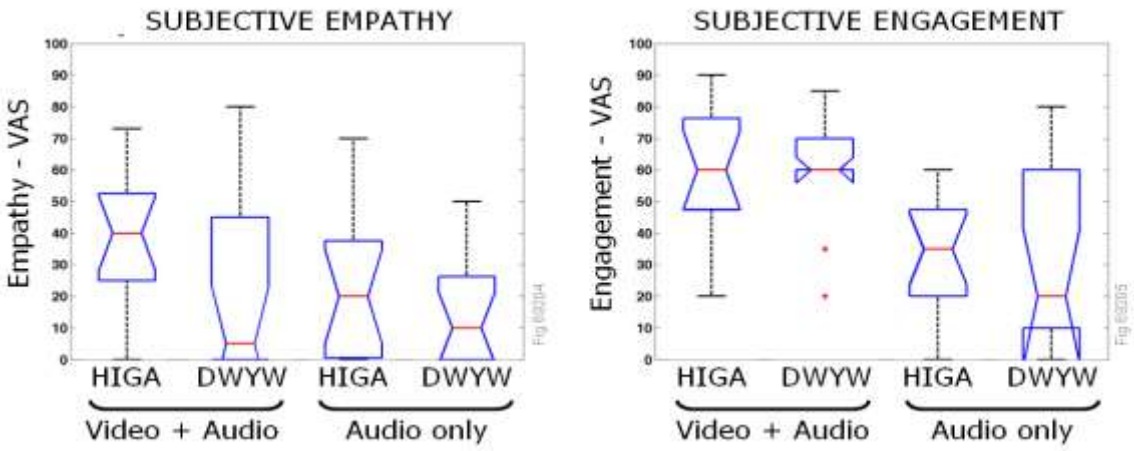

Figure 1. Box plots comparing subjective responses to the stimuli with and without video. Left panel compares empathy ratings in video + audio stimuli vs. the audio only stimuli. It compares "Here It Goes Again (Treadmills)" (HIGA) to "Do What You Want (Wallpaper)" (DWYW). Right panel compares the ratings for subjective engagement in the same volunteers. Outliers are data with values beyond the ends of the whiskers; the maximum whisker length is 1.5 × the inter-quartile range. Boxes whose notches do not overlap indicate that the medians of the two groups differ at the 5% significance level. N = 15.

There was a significant contribution of song to the empathy ratings, while the trend for modality was *reversed* – no additional empathy was engendered by adding video Fig.2, left panel). In an unplanned regression analysis of one-vs.-all of the effect of video + audio version of HIGA vs. the other three stimuli (Table 2), the R-squared was 0.13 ($F(1,55) = 7.82$, $P < 0.01$), which explains effectively all of the predictive power of the song vs. modality empathy model. We conclude that the strongest effect on empathy involved seeing faces (Figure 2, left box plot, left-most column).

Table 2. Linear regression model of "empathy" rating vs. song and modality.

EMPATHY	CI (-95%)	Beta	CI (+95%)
HIGA (vs. DWYW)	1.05	13.08	25.11
Video (vs. music only)	-1.41	10.61	22.64
Constant	1.16	11.81	22.47

4. DISCUSSION AND CONCLUSION

Empathy has previously been proposed as essential to potent engagement [2], but the findings of our team suggest that this relationship is not monotonic. While empathy can be supportive to engagement, it should be considered as a part of a portfolio of possible, desirable contributors to user experience. Other options may include narratives of complex emotional states: i.e. combinations of frustration, disorientation, empathy and even seemingly counter-productive states, e.g. boredom, as well as 'simple' task-oriented satisfaction.

In the two limited but similar examples we tested (from non-narrative music videos), as one might expect, seeing faces does increase people's (lay-defined) ratings of empathy, but does not seem to be a key contributor to engagement. While empathy may increase engagement in many contexts (e.g. narrative films), there are other HCI contexts where attempts to include faces (and amplify emotional empathy) have backfired, e.g. Microsoft's Office Assistant named "Clippy".

We have previously pointed out that the word "immersion" has been used confusingly by different HCI groups as either 1) a structural property of the interaction or 2) as an emotional response of the end-user [14]; eliminating this confusion became the springboard for an entire interdisciplinary conference called Inputs/Outputs [13]. We conclude that the same confusion has been created with the word "empathy". Brown and Cairns [2] use "empathy" in game design to mean a first-person point-of-view (i.e. it is a structural property of the interaction), and explicitly use this theoretical construct (and the word "empathy") to generalise these design principles beyond game design.

The word "empathy" has established definitions in other fields (e.g. media studies, literature) that differ from this meaning in HCI. For example in psychology and among the lay population, empathy is an emotional response in the audience, not a structural property of the stimulus/interaction. If we use the lay understanding of empathy when looking at films, empathy clearly is not sufficient to increase engagement. We suggest that the HCI field extend the influence of its findings by conforming to established, common terminology from the other fields, and that it be scrupulous when selecting and differentiating which terms refer to the structure of interaction vs. which terms refer to emotions in the end-user. Enhanced disambiguation and context sensitivity within HCI will help prevent future errors like "Clippy".

5. ACKNOWLEDGMENTS

We acknowledge the Wellcome Trust for funding a Biomedical Vacation Scholarship for CS (Grant Number 105298/Z/14/Z). We acknowledge the BSMS Individual Research Project (IRP) programme for financing the IRP running costs. We are grateful to Chätrin Tolga for volunteer recruitment and data entry. Finally we are grateful to William Emmanuel Bevan for the initial research idea on the importance of music.

6. REFERENCES

[1] Bornstein, R. F., Kale, A. R., & Cornell, K. R. 1990. **B**oredom as a limiting condition on the mere exposure effect. Journal of Personality and Social Psychology, 58(5), 791-800.

[2] Brown, E & Cairns, P. 2004. A grounded investigation of game immersion. In CHI'04 extended abstract on Human Factors in computing systems (pp. 1297-1300). ACM.

[3] Cairns, P., Cox, A.L., Day, M., Martin, H. & Perryman, T. 2013. Who but not where: The effect of social play on the immersion in digital games. International Journal of Human-Computer Studies, 71(11), pp.1069-1077.

[4] Cairns, P., Cox, A. & Nordin, A.I. 2014. Immersion in digital games: A review of gaming experience research. Handbook of digital games. M.C. Angelides and H. Agius (eds.) Wiley-Blackwell, pp. 339-361.

[5] D'Mello, S., Graesser, A., & Picard, R. W. 2007. Toward an affect-sensitive AutoTutor. Intelligent Systems 22, 4, 53-61.

[6] Greimel, E., Schulte-Rüther, M., Kircher, T., Kamp-Becker, I., Remschmidt, H., Fink, G. R., et al, 2010. Neural mechanisms of empathy in adolescents with autism spectrum disorder and their fathers. Neuroimage, 49(1), 1055-1065.

[7] Jennett, C., Cox, A. L., Cairns, P., Dhoparee, S., Epps, A., Tijs, T., & Walton, A. 2008. Measuring and defining the experience of immersion in games. International Journal of Human-Computer Studies, 66(9), 641-661.

[8] Lee, M. & Kim, G.J. 2014. Empathetic video clip experience through timely multimodal interaction. Journal on Multimodal User Interfaces, 8(3), pp. 273-288.

[9] Mota, S., Picard, R. W. 2003. Automated Posture Analysis for Detecting Learner's Interest Level. Conf. on Computer Vision and Pattern Recognition, 5, 49-49.

[10] O'Brien, H.L. & Toms, E.G. 2010. The development and evaluation of a survey to measure user engagement in e-commerce environments. Journal of the American Society for Information Science & Technology, 61(1), 50-69.

[11] OK Go (2009). Here It Goes Again (music video). EMI Music. https://youtu.be/dTAAsCNK7RA

[12] OK Go (2009). Do What You Want Version 2 (Wallpaper Background) (music video). EMI Music. https://youtu.be/i00GDT9FuFM

[13] Witchel, H. J. 2013a. Engagement: the Inputs and the Outputs. Proceedings of Inputs-Outputs: An Interdisciplinary Conference on Engagement in HCI and Performance. ACM: New York. Article 1.

[14] Witchel HJ, Westling CEI, (2013b). Inputs and Outputs: Engagement in Digital Media from the Maker's Perspective. Excursions, 4(1). www.excursions-journal.org.uk/index.php/excursions/article/view/80

[15] Witchel, H. J., Westling, C., Tee, J., Healy, A., Needham, R., & Chockalingam, N. 2014a. What does not happen: Quantifying embodied engagement using NIMI and self-adaptors. Participations: Journal of Audience and Reception Studies, 11(1), 304-331.

[16] Witchel, H. J., Santos, C. P., Ackah, J. K., Westling, C.E.I., Chockalingam, N, 2016a. Non-Instrumental Movement Inhibition (NIMI) Differentially Suppresses Head and Thigh Movements during Screenic Engagement: Dependence on Interaction. Front. Psychol. 7:157.doi: 10.3389/fpsyg.2016.00157

[17] Witchel, H. J., Santos, C. P., Ackah, J. K., Chockalingam, N, Westling, C.E.I., 2016b (in press). The Complex Relationship between Empathy, Engagement and Boredom. European Conference on Cognitive Ergonomics 2016.