A Quality of Experience Evaluation of Augmented Reality for Procedure Training and Assistance in Industry 4.0.

Eoghan Hynes1, Dr. Ronan Flynn1, Dr. Niall Murray1, Dr. Brian Lee2
Department of Electronics & Informatics1, SRI2, Athlone Institute of Technology

INTRODUCTION

Augmented Reality (AR) is a technology that superimposes computer generated images over the user's real world view. AR is a key technology to increase worker utility in Industry 4.0 in two ways; training for repeatable procedures, and assistance for frequently changing procedures. Crucial to the success of AR for these roles is good user Quality of Experience (QoE).

QoE is a measure of application, service or system performance. It considers how well these technologies fulfill both pragmatic and hedonic needs and expectations of the user. This is measured in terms of the user's degree of delight or annoyance with the technology. Delight and annoyance are emotions. Emotions change in response to stimuli (i.e. the technology), manifesting as changes in physiology and thought patterns. These changes are measurable using sensors and questionnaires. Although the individual human experience is subjective, it is underpinned by this common physiological mechanism. However, it is influenced by personality, context, technical, social and psychological factors.

Fig. 1. QoE influencing factors.

Many recent QoE research works aim to identify novel implicit QoE metrics that can be used to measure QoE continuously during usage of the technology under evaluation. This research involves correlating the changes in user physiology to objective performance results and subjective questionnaire responses to identify how physiological features relate to QoE.

EXPERIMENTAL METHODOLOGY

In work done to date, AR was evaluated against a paper-based instruction medium in a between groups study design. An optimal Rubik's cube solving procedure was used as a proof of concept for the manual orientation and visual verification interventions common to perpetually novel mass customisation, predictive maintenance and distribution optimisation procedures of Industry 4.0. Solving the Rubik's Cube optimally is a least moves solution from any one of the 4.5 quintillion possible Cube states, requiring assistance for successful completion from any non trivial starting position.

The testing protocol was informed by ITU-T P.913 recommendations for subjective and objective assessment methods. This included a gender balanced sample of 48 test subjects, which was procured by convenience sampling. The sample group had an age range from 20 to 64 years old with a mean age of 32 (σ 10 years). The test subjects were divided into two groups of 24, with 12 males and 12 females in each group. The sample group had no prior optimal Rubik's Cube solver experience.

To date, test subject QoE has been evaluated via emotional state as expressed in heart rate, skin temperature & conductance, frequency of head rotation and facial expression for the AR procedure assistance role. The physiological ratings were correlated to Self Assessment Manikin (SAM) affect questionnaire and Likert scale questionnaire responses. The SAM questionnaire allowed the test subjects to report their affective state on three dimensions; valance, arousal and dominance. The Likert scale questionnaire allowed the test subjects to report their quality judgments on fourteen aspects of the procedure assistance medium including usability, utility, aesthetics, interaction and acceptability. The test subjects' physiological ratings were recorded using the Empatica E4 wrist device. The test subject's facial expressions and head rotations were recorded using a 1080p Logitech video camera and OpenFace software.

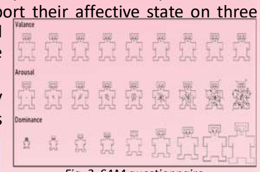

Fig. 2. SAM questionnaire

Fig.3. The Empatica E4.

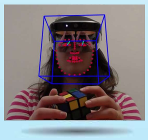

Fig.4. OpenFace estimation.

Methodology phases

1. Information sharing - In this phase, volunteer test subjects were informed that they would be required to solve a Rubik's Cube under one of the test conditions (AR or paper-based instruction).

2. Screening - Upon giving informed consent, test subjects were screened for visual acuity and spatial cognition. No test subjects were excluded from testing during this screening phase.

3. Baseline - The test subjects were seated at a table in a controlled lab environment. The E4 began recording their baseline physiological ratings. The video camera began recording their head rotation and facial Action Units (AUs) using OpenFace. Normal and micro (<.5s) facial expressions were categorised into emotions using consistent and exclusive AUs.

Table.1. Lower Facial AUs.

AU	Full Name	Emotion	Image of AU
AU10	Upper lip raiser	Disgust	
AU12	Lip corner puller	Happy	AU12
AU15	Lip corner depressor	Sad	
AU20	Lip stretched	Fear	
AU26	Jaw drop	Surprise	
Neutral	Lips relaxed and closed	Neutral	

4. Training - The test subjects were trained in using the Rubik's Cube manipulation instructions.

5. Practice - Demonstration of understanding of the Rubik's Cube manipulation instructions was evaluated. Upon successful demonstration of understanding, the test subjects proceeded to testing.

6. Testing – The test subjects completed the Rubik's Cube task. Their physiological ratings were recorded continuously throughout. Upon task completion the test subjects completed the Likert and SAM questionnaires.

RESULTS

Objective Performance results:

The AR group had a significantly shorter mean task completion time than the control group (CG) (p=0.040) with 4.16% higher success rates (p=0.555).

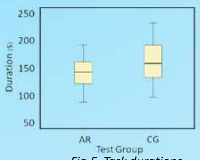

Fig.5. Task durations.

Subjective results:

There were no significant differences between the groups in SAM questionnaire responses for valance (p=0.161), arousal (p=0.561) or dominance (p=0.620). The CG had higher mean positive valance.

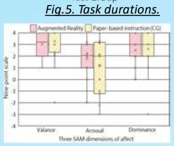

Fig.6. SAM responses.

There were significant differences between the groups on three of the Likert scale questionnaire responses. Table 2 shows the Mann-Whitney U-Test mean rank values and statistical significance for these three statements.

Table 2. Significant differences on three Likert Scale questionnaire responses.

Statement	AR	CG	Sig.
1. The instructions were useful.	22.00	27.00	0.043
3. I became physically uncomfortable during the experience.	29.42	19.58	0.006
8. The instructions were distracting.	28.33	20.67	0.032
2. Following the instructions was not interesting	26.50	22.50	0.286

Implicit results:

Table 3. Physiological Results

	AR mean	CG mean	Sig.
EDA (µS)	-0.29	0.27	0.000
BPM	0.65	1.45	0.613
ST (⁰C)	0.58	0.72	0.273

The CG had significantly higher deviation in skin conductance (EDA) than the AR group.

The CG exhibited twice the mean amplitudes of high frequency (10-13.5 Hz) head rotations then the AR group on three axes of freedom. These are identified in the as exclusively expressing anger emotion.

Fig.7. Axes of head rotation.

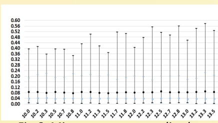

Fig.8. Min, mean, max amplitudes.

Longer task duration accounted for a small increase in these anger frequencies. The CG had longer task durations and higher amplitudes of anger head rotation frequencies.

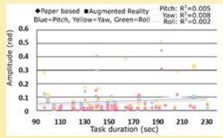

Fig.9. Radian amplitudes over time .

The AR group expressed a significantly higher mean positive deviation in normal expression of happy emotion (p=0.046). The CG expressed a significantly higher mean positive deviation in surprise (p=0.002) emotion.

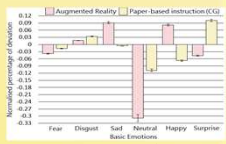

Fig.10. Deviations in normal expressions.

The only majority difference between normal and micro expression was in expression of disgust emotion. The only polarity deviation difference was the AR groups deviation of surprise, from negative in normal expression and positive in micro facial expression.

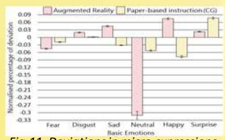

Fig.11. Deviations in micro expressions .

Correlations:

The AR group's mean heart rate deviation correlated moderately and negatively to their Likert scale questionnaire responses for statement #8 (r_s=-0.522).

A positive association was observed between longer task durations and higher mean amplitudes of anger head rotation frequencies in both test groups, accounted for mostly in yaw rotation (R^2=0.008).

The AR groups higher positive micro facial expression deviation of AU10 (disgust) correlated moderately (τ=0.517) to their higher Likert scale response for disinterest in statement #2.

The CG's higher positive deviation of normal facial expression of disgust correlated a negatively (τ=-0.410) to their higher mean subjectively reported (SAM) valance.

CONCLUSION

Despite evident objective performance gains, AR requires careful design in order to fulfil both pragmatic and hedonic needs and expectations of the user. Hardware design must focus on minimising discomfort. Augmentations must be designed to minimise distraction and to maximise perceived usefulness.

By reducing task durations, AR may also reduce the expression of anger emotion in head rotation. Micro facial expression of disgust was a moderate correlate to disinterest in the AR environment. A negative correlation between normal facial expression of disgust and higher mean subjective valance in the control condition may suggest a greater influence of context on classification of disgust in normal expression than in micro expressions.

ACKNOWLEDGMENT

This research was supported by Science Foundation Ireland (SFI) under grant number SFI/16/RC/3918.