An Evaluation of Lower Facial Micro Expressions as an Implicit QoE Metric for an Augmented Reality Procedure Assistance Application.

Eoghan Hynes
Dept. of Computer & Software Engineering
Athlone Institute of Technology
Athlone, Ireland.
e.hynes@research.ait.ie

Ronan Flynn
Dept. of Computer & Software Engineering
Athlone Institute of Technology
Athlone, Ireland.
rflynn@ait.ie

Brian Lee
Dept. of Computer & Software Engineering
Athlone Institute of Technology
Athlone, Ireland.
blee@ait.ie

Niall Murray
Dept. of Computer & Software Engineering
Athlone Institute of Technology
Athlone, Ireland.
nmurray@research.ait.ie

Abstract— Augmented reality (AR) has been identified as a key technology to enhance worker utility in the context of increasing automation of repeatable procedures. AR can achieve this by assisting the user in performing complex and frequently changing procedures. Crucial to the success of procedure assistance AR applications is user acceptability, which can be measured by user quality of experience (QoE). An active research topic in QoE is the identification of implicit metrics that can be used to continuously infer user QoE during a multimedia experience. A user's QoE is linked to their affective state. Affective state is reflected in facial expressions. Emotions shown in micro facial expressions resemble those expressed in normal expressions but are distinguished from them by their brief duration. The novelty of this work lies in the evaluation of micro facial expressions as a continuous QoE metric by means of correlation analysis to the more traditional and accepted post-experience self-reporting.

In this work, an optimal Rubik's Cube solver AR application was used as a proof of concept for complex procedure assistance. This was compared with a paper-based procedure assistance control. QoE expressed by affect in normal and micro facial expressions was evaluated through correlation analysis with post-experience reports. The results show that the AR application yielded higher task success rates and shorter task durations. Micro facial expressions reflecting disgust correlated moderately to the questionnaire responses for instruction disinterest in the AR application.

Keywords—quality of experience, augmented reality, micro facial expression, affective state

I. INTRODUCTION

As automation of repetitive procedures increases, augmented reality (AR) has been identified as an important technology to enable humans to have key roles in the workforce addressing frequently changing and complex procedures [1]. Examples of such perpetually novel procedures in the literature include mass customisation [2], optimisation of warehouse and distribution logistics [3] and predictive maintenance [4]. These procedures often involve bimanual visuomotor interventions such as alignment, adjustment and orientation manipulations, combined with visual identification, comparison, inspection and verification [5]. To this end, an optimal Rubik's Cube solver AR application was evaluated as a proof of concept for complex procedure assistance and compared against a paper-based assistance control group.

Quality of experience (QoE) reflects quality judgments of multimedia experiences and these quality judgments can in turn influence acceptability [6] of the media under consideration. An important factor in the acceptability of AR for procedure assistance roles is a positive user QoE. Traditional approaches of determining user QoE have relied upon post-experience reports, which have recently been shown to be skewed by primacy, recency and peak experience stimuli [7]. For this reason, recent research has focused on identifying implicit metrics of QoE such as skin temperature, facial expressions [8], gait [9], heart rate, electrodermal activity [10]–[13], motion sickness, electroencephalogram [14] and eye gaze [15]. Similarly, the focus of this work is on evaluatig the relationship between micro facial expressions and QoE. User QoE can influence the user's emotional state [16], which is reflected in facial expression [17]. Micro facial expressions are described as spontaneous and subtle movements [18] that occur when a person experiences emotion [19]. They reveal true and potential expression [18] and are more accurate indicators of a train of thought or even subtle, passive or involuntary thoughts [20], particularly when the person is trying to conceal or repress that emotion [21]. Emotions encoded in these micro facial expressions resemble those in normal expression [21] but are distinguished by their brief duration [19].

Normal facial expressions have previously been used to determine user QoE via affect [8]. However, categorisation of emotion from normal expressions has been shown to be highly dependent upon context [22], which can result in incorrect emotion categorisation. QoE is also influenced by context (personal, social, application) [6]. Any context influencing factors that are common to both QoE and facial expressions will be reflected in both the user's QoE and in their facial expressions. Normal facial expressions are recorded in this work for comparison to micro facial expressions by means of cross-correlation to subjective reports. In doing so, the main contribution of this work is therefore an evaluation of the influence of context on normal and micro facial expression of QoE via affect.

The rest of this paper is structured as follows. A critique of the works that inform our method is given in Section II. The experimental setup and method used in obtaining the results are described in Section III. This is followed in Section IV with the results and a discussion of the key findings. The final conclusions are presented in Section V.

II. RELATED WORK

QoE is defined in [6] as "*the degree of delight or annoyance of a person whose experiencing involves an application, service, or system*". Delight and annoyance denote emotions [23]. While normal facial expression of emotion has previously been used to evaluate the user's degree of delight or annoyance [8], micro facial expressions are described in the literature as more accurate indicators of affect that occur during the experience of emotions [19]. If this is true, it should be possible to more accurately infer the degree of a user's delight or annoyance by means of micro facial expressions of affect rather than normal facial expressions of affect.

The literature has previously defined the upper threshold of micro facial expressions to last between 200ms and 333ms. However, the authors of [24] demonstrated that micro facial expressions not only have an upper duration threshold of 502.78 ms, but also a lower duration threshold of 169.07 ms. They arrived at these results by recording the facial expressions of 20 test subjects as they watched a randomised set of 17 videos containing positively or negatively valanced content. The test subjects were instructed not to show any facial expressions for a high stakes financial reward. From 1,000 leaked facial expressions, 245 lasted up to 1s and 109 lasted up to 0.5s. The distribution of durations fitted a Gamma model with a Kolmogorov-Smirnov static deviation from a normal distribution of 0.082, with the lower and upper bounds of 170-500ms (rounded) respectively. Micro facial expressions were also shown to feature rapid onset with an upper onset duration threshold of 260ms. The authors of [25] demonstrated that it is possible to detect micro facial expressions of this duration using a standard video camera recording at 25 frames per second (FPS) i.e. between 4.25–12.5 frames.

In [26], the authors created a database of compound facial expressions, which consisted of combinations of the six basic emotions (happiness, surprise, anger, sadness, fear and disgust) [27]. The authors of [26] combined 1,610 images of 230 individuals expressing the six basic emotions, plus neutral expression, to define a new set of 15 compound emotion expression categories. In doing so, the authors identified exclusive and consistently active (>70% of instances) action units (AUs) in the six basic emotions. They showed that AU10 occurred consistently in disgust, AU12 occurred consistently in happiness, AU15 occurred consistently in sadness, AU20 occurred consistently in fear, AU24 occurred consistently in anger and AU26 occurred consistently in surprise.

The authors of [28] independantly identified the same set of exclusively occuring AUs to categorise emotions. This was done as part of a multi-step selection process towards extending the existing Cohn-Kanade dataset for automatic detection of facial expression of emotion. They extended the existing dataset of 97 test subjects by analysing the expression sequences of a further 26 test subjects. This set of exclusively occurring constituent AUs [29] are used in the methodology of the work presented in this paper because they are lower facial AUs [27] that can be used to determine the emotion being expressed in an AR environment where the head mounted device (HMD) occludes upper facial AUs.

This work contributes to the state-of-the-art in QoE as the first work to evaluate micro facial expressions as a continuous implicit metric of QoE. This involved a correlation analysis of the affective signal transmitted in micro and normal facial expressions, to post experience subjective reports. In doing so, this work evaluates the influence of context on normal and micro facial expression of QoE via affect.

III. Experimental setup and Method

A. Participants

A gender balanced sample of 48 test subjects was used as recommended in ITU-T P.913 specifications for subjective and objective assessment methods. 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.

B. AR and paper-based procedure assistance modalities

In the AR condition, the instructions to solve the Rubik's Cube were presented using the Meta2[1] AR HMD. The Meta2 was a prototype AR HMD aimed at developers and was made never commercially available. The Meta2 had a 90-degree field of view; 2.5K screen resolution with a 60Hz refresh rate; a 720p front-facing RGB camera; a 9 ft (2.7 m) USB cable for video, data and power. The AR application was developed using the Kociemba algorithm [30] to solve the standard 3x3 Rubik's Cube in the fewest possible number of moves (optimally) from any scrambled state. The AR application was adapted from an online repository[2], originally developed for Android™[3] devices. This was implemented after an initial development approach using template matching with the Vuforia™ AR SDK was found to be ineffective for fiducial-less Rubik's Cube tracking. This repository was translated from Java to C#, using a C# wrapper for the OpenCV library in Unity 3D™ for use with the META2 AR HMD. At the beginning of each test, the front-facing camera on the Meta2 AR HMD was first used to scan all faces of the scrambled cube. The AR application tracked the Rubik's Cube using a combination of OpenCV filters, colour and shape detection algorithms. C# algorithms were used to match the affine features of the standard 3x3 Rubik's Cube for real-time detection from the input video feed. Once the Rubik's Cube was successfully scanned at the beginning of each test in the AR condition, the AR application proceeded to heuristically solve the cube. Once solved, the AR application guided the user step-by-step by displaying the shortest path to the solved state directly in the user's field of view on the META2's screen.

Standardised testing necessitated that the procedure be the same in each test. To ensure this, each test began with the Rubik's Cube in the same initial position. The superflip position [31] was used for this, requiring the maximum (20) steps to solve using the optimal solution algorithm. In this way, the test subjects in the paper-based instruction (control) group were provided with the same set of procedure instructions as the AR group, printed in a 22-page, A4 paper, instruction manual. To further standardise testing, instruction progression was user-controlled in both test conditions as in [32]. This afforded both test groups the same level of control over instruction progression. The control group (CG) instruction manual consisted of one instruction per page. The CG progressed through their instructions by turning each page of the instruction manual as required. The AR test subjects were presented with one instruction at a time in their

[1] https://metavision.com/
[2] https://github.com/AndroidSteve/Rubik-Cube-Wizard
[3] **Android** is a **trademark** of Google LLC.

field of view in the HMD. They progressed through the instructions using keyboard input.

C. Experimental methodology

The experimental methodology consisted of six phases. The first phase (1) was the information sharing phase. 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). Upon giving informed consent, test subjects were screened (2) for visual acuity and spatial cognition. No test subjects were excluded from testing during this screening phase. Next came a five-minute baseline phase (3). In this phase, the test subjects were seated at a table in a controlled lab environment. A desk mounted Logitech 1,080p camera began recording the test subjects' AUs using the OpenFace face recognition application [33], from which normal and micro facial expressions were determined and categorised into expressions of emotions. The percentage of expression of these emotion categories during this five-minute period was considered as their baseline normal and micro facial expression rating. Recording of AUs continued throughout the rest of the test. Next, the test subjects were trained (4) in using the Rubik's Cube manipulation instructions. This was followed by the practice phase (5), in which the test subjects' understanding of the Rubik's Cube manipulation instructions was evaluated. Upon successful demonstration of understanding, the test subjects proceeded to the testing phase (6). Upon completion of the Rubik's Cube task, the test subjects responded to Likert and SAM questionnaires as described in the following section.

D. QoE metrics

Normal and micro facial expressions of emotion were correlated to post experience subjective reports. The explicit and implicit QoE metrics used in this evaluation are described in this section.

1) Explicit QoE metrics:

a) The self-assessment manikin (SAM) questionnaire.

The test subjects explicitly reported their affective state using a post-experience SAM questionnaire [34]. The SAM questionnaire consists of three scales, one for each dimension of affect (arousal, valence and dominance). The test subjects completed the questionnaire by circling one manikin on each scale, representing the level of the dimension of affect that they felt upon task completion.

b) The Likert scale questionnaire

A five-point Likert scale questionnaire consisting of 14 statements was created to allow the test subjects to report their quality judgments on utility, interaction, aesthetics, usability and efficiency criteria, as well as their acceptability of the procedure assistance media post-experience. The authors of [35] recommended using such subjective reports in combination with objective performance metrics to holistically evaluate user QoE. The 14 statements used are listed in Table I in section IV. Test subjects used the five-point scale to agree or disagree with the 14 statements on a scale from 1 (strong disagreement) to 5 (strong agreement).

2) Implicit QoE metrics:

Normal and micro facial expression of emotion were considered in this work in the following way. The percentage of expression of the basic emotions during the five-minute baseline phase were considered as each test subject's baseline rating. Each test subject's task duration was normalised by percentage of completion in increments of 10%. The deviation of emotions expressed at each of these intervals from the test subject's baseline rating was calculated. Each test subject's overall average deviation was then used for statistical analysis within and between the test groups.

Emotion expression was determined using the presence of facial AUs. To achieve this, OpenFace was used to detect and record the presence of the test subjects' AUs by means of the desk mounted video camera. OpenFace can detect eighteen AUs. Ten of these ARs are lower facial AUs. Lower facial AUs were used because the AR HMD partially occluded some of the upper facial AUs in the AR environment. The set of lower facial AUs available was AU 10, 12, 14, 15, 17, 20, 23, 25, 26 and 28. Five of these AUs have been shown in the literature to be consistently active constituent AUs in five mutually exclusive emotions [26], [28]. The five AUs used in this methodology to determine emotion expression were AU10 for disgust, AU12 for happy, AU15 for sad, AU20 for fear, AU26 for surprise, and neutral expression. These are shown in Table II. AU24 (anger) is not detected by OpenFace. Contiguous constituent AU presence between 170-500ms [24] was counted as a micro facial expression of emotion. Contiguous constituent AU presence above 500ms was counted as a normal facial expressions.

IV. Results and Discussion

Statistical results reported herein emanate from nonparametric independent samples Mann-Whitney U-tests and Kendall's Tau (τ) correlations at 95% confidence levels. Objective performance results showed 91.67% and 95.83% successful task completion for the CG and AR groups respectively (p=0.555). The mean task completion times were 162.29s and 142.21s for the CG and AR groups respectively (p=0.040). These results show that AR yielded pragmatic gains over alternative assistance modalities in line with findings reported in the literature [2], [3].

TABLE II. THE CONSTITUENT LOWER FACIAL AUS ASSOCIATED WITH THE NEUTRAL STATE AND THE EMOTIONS DISGUST, HAPPY, SAD, FEAR AND SURPRISE [26], [28], [33].

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

Fig. 1. shows a bar chart of the groups' average deviation from baseline of normal expression for the five emotion categories, plus neutral expression. The difference between the groups was significant for normal expression of surprise (p=0.002), neutral (p=0.032) and happy (p=0.046). The AR group expressed more positive deviation of happiness. The paper-based CG expressed more positive deviation of surprise. This would suggest that the AR group was significantly happier with its procedure assistance medium and the CG was significantly more surprised by its procedure assistance medium. Fig. 2. shows the groups deviation from baseline of micro expressions of the same six categories. In micro expressions, a significant difference between the groups was only seen for neutral expression (p=0.001). The only polarity change of deviation between normal and micro facial expression was the AR group's positive mean deviation of micro expression of surprise compared to a negative mean deviation of normal expression of surprise. The inter-group significance (p values) for normal and micro expressions are given in Table III for ease of comparison. The only inter-group majority difference between normal and micro facial expression was in expression of AU10 (classified as disgust emotion), where the CG expressed more disgust emotion than the AR group in normal expression and less disgust than the AR group in micro expressions. The AR group's expression of disgust was consistent in normal and micro facial expressions (Δ 0.002%).

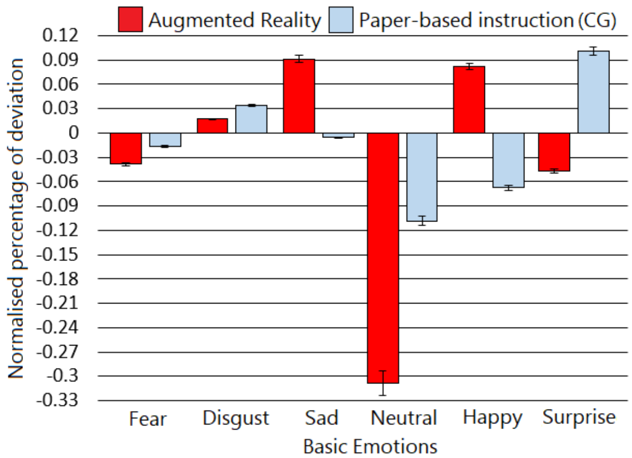

Fig. 1. Deviations from baseline of normal expression of five basic emotions for CG and AR test groups.

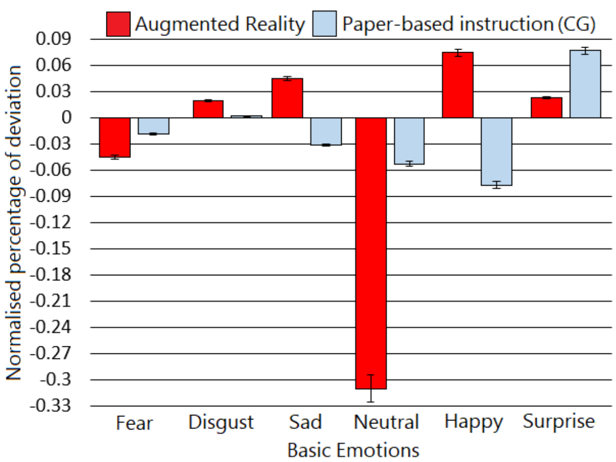

Fig. 2. Deviations from baseline of micro expression of five basic emotions for the CG and AR test groups.

TABLE III. STATISTICAL DIFFERENCE BETWEEN THE TEST GROUPS FOR NORMAL AND MICRO EXPRESSION OF FIVE BASIC EMOTIONS.

Emotion	Normal Expression	Micro expression
Fear	0.085	0.185
Disgust	0.081	0.613
Sad	0.884	0.621
Neutral	0.032	0.001
Happy	0.046	0.053
Surprise	0.002	0.070

A within group analysis of the differences between normal and micro facial expressions showed no significant differences. The closest to significant difference was between the CG's normal and micro expression of sad emotion (p=0.152). The next most significant difference was between the AR group's normal and micro expression of surprise (p=0.167). The difference between the sum of positive and negative deviation of positive and negative valence showed that the CG expressed a net increase of 1.37% more positive valence than the AR group in normal facial expression while the AR group expressed a net increase of 3.01% more positive valence than the CG in micro facial expressions.

Fig. 3. shows the distribution of SAM questionnaire responses for both groups. There were no significant differences between the groups for valence (p=0.161), arousal (p=0.561) or dominance (p=0.620). The CG reported higher mean valence than the AR group. The strongest correlation of SAM responses to facial expression was between the AR group's lower mean SAM valence and its lower deviation micro expression of surprise with τ=0.418. The second strongest correlation was a negative correlation between the CG's higher SAM valence responses and its higher normal expression of disgust with τ=-0.410. Even though this correlation was weak, the fact that was negative may demonstrate the influence of cultural context [22] on categorisation of disgust in normal facial expressions.

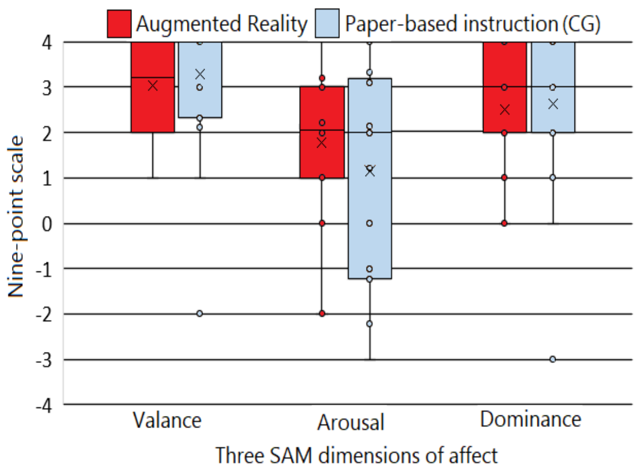

Fig. 3. Box plot of distribution of SAM questionnaire responses for the CG and AR test groups. x = median, bar = mean.

Table I shows the significance of the Likert scale questionnaire responses including the Mann-Whitney U-test mean rank (MR) values. This demonstrated statistically significant differences between the groups for questionnaire statements 1, 3 and 8, highlighted in grey. Statement 1 related to usefulness of the instructions. The difference between the groups was significant with p=0.043. The U-test mean rank (MR) for interaction was higher for the CG (27.00) than for the AR group (22.00). For this evaluation, the instructions were text-only to standardise modalities for both groups. This might be considered a limitation of this evaluation. AR test subjects could have expected more interactive instructions in an AR experience. This result may reflect a lack of hedonic expectation fulfilment in the AR environment, where pragmatic needs were met as reflected in better performance results for the AR group. Statement 3 related to aesthetics of the assistance media in terms of comfort. This was significantly different between the groups with p=0.006 in favour of the paper-based assistance medium. The MR of question 3 was higher for the AR group (29.42) compared to the CG (19.58). As this question was posed negatively, it signifies that the AR group reported experiencing higher discomfort. The difference in reported comfort is partially due to bespectacled test subjects reporting less comfort with the HMD in the AR group, where 6 and 5 test subjects wore glasses in the CG and AR groups respectively. The META2 AR HMD was designed for use with spectacles but caused some pressure at the sides of the head. This left a temporarily visible mark on some bespectacled test subjects after use. Statement 8 related

TABLE I. LIKERT SCALE QUESTIONNAIRE STATEMENT, U-TEST MEAN RANK VALUES AND STATISTICAL SIGNIFICANCE

Statement	AR MR [a]	CG MR	Sig. [a]
1. The instructions were useful.	22.00	27.00	.043
2. Following the instructions was not interesting.	26.50	22.50	.286
3. I became physically uncomfortable during the experience.	29.42	19.58	.006
4. My experience was not frustrating.	21.90	27.10	.128
5. I felt confident in my ability to follow the instructions.	23.00	26.00	.355
6. Learning to use the instructions correctly was not easy.	24.81	24.19	.859
7. I really enjoyed my experience.	23.77	25.23	.682
8. The instructions were distracting.	28.33	20.67	.032
9. My experience was stressful.	25.60	23.40	.543
10. I would like to experience this form of instruction again.	24.79	24.21	.876
11. Attempting to solve a Rubik's Cube was an enjoyable experience.	23.63	25.38	.577
12. Moving on to the next instruction was easy.	22.38	26.63	.227
13. Using the instructions felt intuitive.	23.92	25.08	.753
14. The mode of instruction was not natural.	26.67	22.33	.253

[a] MR: U-test mean rank, Sig: Statistical p-value (α=0.05).

to interaction quality in terms of the instructions being distracting. This was significantly different between the groups with p=0.032. The MR was 20.67 and 28.33 for the CG and AR groups respectively. This statement was posed negatively, signifying that the AR group reported that the AR-based instructions were more distracting than the paper-based group. In the AR environment, the AR instructions remained in the AR user's field of view throughout the experience. In the control environment, the CG test subjects were free to focus their full attention on the Rubik's Cube once they had read each instruction from the instruction manual. If AR augmentations are not carefully designed, they can in fact result in increased distraction from the workpiece. The trade-off of not having to commit instructions to short term memory as in the control environment could be increased distraction by the instructions in the AR user's FOV while manipulating the workpiece.

A Kendall's correlation analysis revealed one moderate positive correlation of τ=0.517 between the AR groups Likert questionnaire response for statement 2 and their micro facial expressions of disgust (0.02% > CG). Statement 2 related to interest in following the instructions. The AR group reported more disinterest in following their instructions while expressing more AU10 (disgust) in micro facial expressions than the CG.

V. CONCLUSIONS

This paper evaluated the relationship between QoE and micro facial expressions of emotion during AR and paper-based assistance in a Rubik's Cube solving procedure. In doing so, micro facial expression of AU10 was shown to be a moderate correlate of disinterest in the AR condition. The AR group expressed more micro facial expression of AU10 (disgust) than the CG, while reporting greater disinterest in following their instructions. This might have been because of lack of expectation fulfilment in the AR environment due to the text-only nature of the instructions. The AR group's higher deviation of micro facial expression of disgust emotion may have reflected this potential lack of expectation fulfilment, which correlated moderately to post experience reports of disinterest by them.

AR yielded higher task success rates and significantly shorter task completion durations than the paper-based procedure assistance medium. However, the CG reported higher mean positive valence in the SAM questionnaire, which was reflected in their 1.37% higher net positive valence in normal facial expressions. The AR group expressed 3.01% more net positive valence in micro facial expressions than the CG. This might have implied higher QoE in the AR environment where micro facial expressions are described as more accurate than normal facial expressions throughout the literature, but this was not reflected in the AR group's lower mean SAM responses for valence. The reasons for lower subjective valence reported by the AR group in the SAM questionnaire despite objective performance gains was likely due to the significant lack of perceived usefulness, increased discomfort and greater distraction caused by the AR instructions in the AR environment as reported by them in the Likert scale questionnaire.

In conclusion, despite evident objective performance gains, AR requires careful design to minimise discomfort and distraction and to increase perceived usefulness. In arriving at these conclusions, the subjective SAM questionnaire was used to provide a ground truth for affective state. However, micro facial expressions are subconscious and may not be reflected in subjectively perceived and reported affect. For example, while the CG's higher mean subjective (SAM) valance was reflected in the higher net positive valance expressed in their normal facial expressions, the AR groups lower mean subjective valance was not reflected in the higher net positive valance expressed in their micro facial expressions. Nevertheless, the negative correlation between the CG's higher subjective valance and their higher normal facial expression of disgust may have served to highlight a greater influence of context on disgust categorisation in normal facial expression than in micro facial expression. Future work will continue to evaluate micro facial expressions as an implicit metric of QoE during AR usage.

ACKNOWLEDGMENT

This publication has emanated from research conducted with the financial support of Science Foundation Ireland (SFI) under grant number SFI/16/RC/3918 and the AIT President's Seed Fund.

REFERENCES

[1] S. Egger-Lampl, C. Gerdenitsch, L. Deinhard, R. Schatz, and P. Hold, 'Assembly Instructions with AR: Towards measuring Interactive Assistance Experience in an Industry 4.0 Context', in 2019 Eleventh International Conference on Quality of Multimedia Experience (QoMEX), Jun. 2019.

[2] F. Loch, F. Quint, and I. Brishtel, 'Comparing Video and Augmented Reality Assistance in Manual Assembly', in 2016 12th International Conference on Intelligent Environments (IE), Sep. 2016, pp. 147–150.

[3] B. Mueck, M. Höwer, W. Franke, and W. Dangelmaier, 'Augmented Reality applications for Warehouse Logistics', in Soft Computing as Transdisciplinary Science and Technology, vol. 29, Springer Berlin Heidelberg, 2005, pp. 1053–1062.

[4] R. Schlagowski, et al., 'Design of an assistant system for industrial maintenance tasks and implementation of a prototype using augmented reality', in 2017 IEEE Int. Conference on Industrial Engineering and Engineering Management (IEEM), Dec. 2017, pp. 294–298.

[5] U. Neumann and A. Majoros, 'Cognitive, performance, and systems issues for augmented reality applications in manufacturing and maintenance', in Proceedings. IEEE 1998 Virtual Reality Annual International Symposium (Cat. No.98CB36180), Mar. 1998, pp. 4–11.

[6] S. Möller and A. Raake, Quality of Experience, Advanced Concepts, Applications and Methods. Springer, 2013.

[7] S. S. Sabet, C. Griwodz, and S. Möller, 'Influence of primacy, recency and peak effects on the game experience questionnaire', in Proceedings of the 11th ACM Workshop on Immersive Mixed and Virtual Environment Systems, Amherst, Massachusetts, Jun. 2019, pp. 22–27.

[8] E. Hynes, R. Flynn, B. Lee, and N. Murray, 'A Quality of Experience Evaluation Comparing Augmented Reality and Paper Based Instruction for Complex Task Assistance', in 2019 IEEE 21st International Workshop on Multimedia Signal Processing, Sep. 2019.

[9] T. B. Rodrigues, C. Ó. Catháin, D. Devine, K. Moran, N. E. O'Connor, and N. Murray, 'An evaluation of a 3D multimodal marker-less motion analysis system', in Proceedings of the 10th ACM Multimedia Systems Conference, Amherst, Massachusetts, Jun. 2019, pp. 213–221.

[10] D. Egan, S. Brennan, J. Barrett, Y. Qiao, C. Timmerer, and N. Murray, 'An evaluation of Heart Rate and ElectroDermal Activity as an objective QoE evaluation method for immersive virtual reality environments', in 2016 Eighth International Conference on Quality of Multimedia Experience (QoMEX), Jun. 2016.

[11] C. Keighrey, R. Flynn, S. Murray, S. Brennan, and N. Murray, 'Comparing user QoE via physiological and interaction measurements of immersive AR and VR speech language therapy applications.', 25th ACM Int. Conf. Multimed. ACM MM 2017, vol. Thematic Workshop.

[12] C. Keighrey, R. Flynn, S. Murray, and N. Murray, 'A QoE evaluation of immersive augmented and virtual reality speech language assessment applications', in 2017 Ninth International Conference on Quality of Multimedia Experience (QoMEX), May 2017.

[13] C. Keighrey, R. Flynn, S. Murray, and N. Murray, 'A Physiology-based QoE Comparison of Interactive Augmented Reality, Virtual Reality and Tablet-based Applications', IEEE Trans. Multimed, 2020.

[14] D. P. Salgado et al., 'A QoE assessment method based on EDA, heart rate and EEG of a virtual reality assistive technology system', in Proceedings of the 9th ACM Multimedia Systems Conference on - MMSys '18, Amsterdam, Netherlands, 2018, pp. 517–520.

[15] D. Concannon, R. Flynn, and N. Murray, 'A quality of experience evaluation system and research challenges for networked virtual reality-based teleoperation applications', in Proceedings of the 11th ACM Workshop on Immersive Mixed and Virtual Environment Systems, Amherst, Massachusetts, Jun. 2019, pp. 10–12.

[16] P. Reichl et al., 'Towards a comprehensive framework for QOE and user behavior modelling', in Seventh International Workshop on Quality of Multimedia Experience (QoMEX), Pylos-Nestoras, 2015.

[17] Z. Zeng et al., 'A Survey of Affect Recognition Methods: Audio, Visual, and Spontaneous Expressions', IEEE Trans. Pattern Anal. Mach. Intell., vol. 31, no. 1, pp. 39–58, Jan. 2009.

[18] Y. Huang, F. Chen, S. Lv, and X. Wang, 'Facial Expression Recognition: A Survey', Symmetry, vol. 11, no. 10, p. 1189, Oct. 2019.

[19] A. K. Davison, C. Lansley, N. Costen, K. Tan, and M. H. Yap, 'SAMM: A Spontaneous Micro-Facial Movement Dataset', IEEE Trans. Affect. Comput., vol. 9, no. 1, pp. 116–129, Jan. 2018.

[20] M. Takalkar et al., 'A survey: facial micro-expression recognition', Multimed. Tools Appl., vol. 77, no. 15, pp. 19301–19325, Aug. 2018.

[21] S. Polikovsky et al., 'Facial micro-expressions recognition using high speed camera and 3D-gradient descriptor', pp. 16–16, Jan. 2009.

[22] 'Context in Emotion Perception - Lisa Feldman Barrett, Batja Mesquita, Maria Gendron, 2011'. https://journals.sagepub.com/doi/abs/10.1177/0963721411422522 (accessed Feb. 14, 2020).

[23] J. A. Russell, 'A circumplex model of affect', J. Pers. Soc. Psychol., vol. 39, no. 6, pp. 1161–1178, Dec. 1980.

[24] W.-J. Yan, Q. Wu, J. Liang, Y.-H. Chen, and X. Fu, 'How Fast are the Leaked Facial Expressions: The Duration of Micro-Expressions', J. Nonverbal Behav., vol. 37, no. 4, pp. 217–230, Dec. 2013.

[25] T. Pfister, Xiaobai Li, G. Zhao, and M. Pietikäinen, 'Recognising spontaneous facial micro-expressions', in 2011 International Conference on Computer Vision, Nov. 2011, pp. 1449–1456.

[26] S. Du et al., 'Compound facial expressions of emotion', Proc. Natl. Acad. Sci., vol. 111, no. 15, pp. E1454–E1462, Apr. 2014.

[27] G. Donato, M. S. Bartlett, J. C. Hager, P. Ekman, and T. J. Sejnowski, 'Classifying facial actions', IEEE Trans. Pattern Anal. Mach. Intell., vol. 21, no. 10, pp. 974–989, Oct. 1999.

[28] P. Lucey, J. F. Cohn, T. Kanade, J. Saragih, Z. Ambadar, and I. Matthews, 'The Extended Cohn-Kanade Dataset (CK+): A complete dataset for action unit and emotion-specified expression', in 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Workshops, Jun. 2010, pp. 94–101.

[29] Y.-I. Tian, T. Kanade, and J. F. Cohn, 'Recognizing action units for facial expression analysis', IEEE Trans. Pattern Anal. Mach. Intell., vol. 23, no. 2, pp. 97–115, Feb. 2001.

[30] T. Rokicki, H. Kociemba, M. Davidson, and J. Dethridge, 'The Diameter of the Rubik's Cube Group Is Twenty', SIAM Rev., vol. 56, no. 4, pp. 645–670, Jan. 2014 .

[31] J. Pride, 'From Superflip to Solved', Academic Excellence Showcase Schedule, Jun. 01, 2017. https://digitalcommons.wou.edu/aes_event/2017/all/226.

[32] S. J. Henderson, 'Augmented reality in the psychomotor phase of a procedural task', in 2011 10th IEEE International Symposium on Mixed and Augmented Reality, Oct. 2011, pp. 191–200.

[33] T. Baltrušaitis et al., 'OpenFace: An open source facial behavior analysis toolkit', in 2016 IEEE Winter Conference on Applications of Computer Vision (WACV), Mar. 2016.

[34] J. D. Morris, 'Observations: SAM: The Self-Assessment Manikin An Efficient Cross-Cultural Measurement Of Emotional Response',1995.

[35] I. Wechsung, K.-P. Engelbrecht, C. Kühnel, S. Möller, and B. Weiss, 'Measuring the Quality of Service and Quality of Experience of multimodal human–machine interaction', J. Multimodal User Interfaces, vol. 6, no. 1, pp. 73–85, Jul. 2012.