VR & EEG – Mixed Computer Science Methods in the evaluation of Facial Emotion Expression Recognition

Sean Haddick[1], David J. Brown[1], James Lewis[1], Bonnie Connor[1,2] & Simon Schofield[1]

[1]Department of Computer Science, Nottingham Trent University, Clifton Campus, College Drive, Clifton, Nottingham, United Kingdom | *[2]1105 Kennedy Place Suite 6, Davis CA 95616*

[1]sean.haddick@ntu.ac.uk, [2]bonnie@bonnieconnor.com

[1]https://www.ntu.ac.uk/

DOI: 10.54337/aau.icdvrat26.26

Abstract

Facial Emotion Expression Recognition is widely agreed to be one aspect of social function that is impaired within the usual characterisation of Autism Spectrum Disorder. The ”Reading the Mind in the Eyes” test is used to this day in the diagnostic process. However, it disregards temporal data, has potential test format effects, and only observes performance rather than underlying behavioral triggers. With the increasing prevalence of VR and EEG tools, including this temporal data and investigating brain activity during testing may result in novel methods to investigate impairments in social function. This paper outlines a study protocol to be used to investigate the application of said tools to this purpose.

1. Introduction

Autism Spectrum Disorder (ASD) is a neuro developmental disorder characterised in part by persistent and recurring deficits in social interaction, including impaired social skills, weak communication skills, and repetitive behaviours (American Psychiatric Association (2022)).

An aspect that is applied both in the diagnosis and in the care of people with Autism is social understanding. A long standing canon of research has been performed investigating theory of mind, emotional understanding, and emotion identification, from developmental stages through adolescence and into adulthood. Common broad stroke diagnostics include self-reporting of a patient’s experiences with emotional connection to others, and measurements of the ability to observe, process, analyse and identify emotions in the face of other people. The latter is a key component of one of the formative tools for the diagnostic process regarding ASD, the “Reading the Mind in the Eyes” test or the RMET. This, alongside other tests of its kind, are named tests of “Facial Emotion Expression Recognition”, or FEER. However, some research contests the consistency and reliability of these tests, specifically when attempting to diagnose ASD (Vellante et al. (2013), Higgins et al. (2026)).

Other research also investigates the use of Electroencephalography (EEG) to examine and study the causes of deficits in FEER – however, there is a persistent problem with low power and poor repeatability. In particular, a verifiable and easily communicated standard is needed.

This paper will describe an ongoing study that expands the investigation of the technologies examined in the previous study, investigating the impact of spatial-temporal information on brain mimicry processes. It will also specifically examine the use of VR and High Realism Human Avatars in the standardisation of FEER assessment tools, utilising low-cost motion capture solutions to create a replicable research tool.

2. Background

EEG can be used to capture social-action / observation differences relative to Autism; a significant amount of research has already found some connection between the "Mu rhythm" and Autism, but the evidence is not yet strong enough to consistently differentiate ASD from typically developing persons consistently (Bernier et al. (2007)). The Mu rhythm is an alpha-band sensorimotor rhythm that is often suppressed during action observation, and some studies found reduced mu attenuation in Autism during observation tasks, with links to imitation ability (Fox et al. (2016)).

Some research has found that there are anomalies in Mu Rhythm when a person with Autism completes a motor mimicry task (Bernier et al. (2007)). Other studies show that there is a significant difference in cognitive processing detected by EEG (Wagner et al. (2013)) in general. However, it remains unclear whether these signals have diagnostic sensitivity (Bender et al. (2025)), in part due to the poor power in most studies on the subject (Hobson & Bishop (2017)).

For these reasons there is a need to develop a standardised methodology for the creation of FEER tasks – whilst there appears to be some consensus that there is a potential for a medium-sized effect relating ASD and Mu Rhythm Attenuation during emotional recognition tasks, there are few studies with sufficiently high power in and of themselves – and in general studies are not using sufficiently comparable methods to allow for inter-study validation. Using a method which can precisely measure or define the strength of the stimuli – say, highly realistic avatars, such as that in Epic Games' MetaHuman framework (Fang et al. (2021)) – will improve the ongoing quality of research into social-emotional recognition.

There are also novel technologies within Brain-Computer-Interfaces that provide some new avenues for research. Research headsets such as the Galea (OpenBCI (2026)) couple visualization and multi-modal data collection – including EEG, with the Mu rhythm in theory able to be captured. However, as these headsets use dry electrodes, there are complications with obtaining consistently clean data – which could complicate group separation if we only look at Mu. However, due to the multi-modal aspects of the headset, as well as recent advances in the construction of dry EEG electrodes, it may be more plausible to filter out unclean data and artifacts resulting from head movement, eye movement and arm movement.

Other research proposes that Mu Rhythm and EEG in general, could provide a highly sensitive perspective on understanding internal perception of emotion, especially if integrated into multimodal systems (Huang et al. (2017)). As such, later research could combine Mu band metrics with eye tracking, response time, task completion accuracy, facial action captured through Electromyography (EMG), and broader EEG features, many of which are also shown to have strong links to emotion processing (Wall et al. (2026)).

The main intent of the planned study is to identify if such a multi-modal system is viable, starting with an intensive evaluation of the presence of the Mu Rhythm, doing so by proposing a standard method of FEER evaluation creation.

3. Methodology

3.1. Participants

The recruitment of participants for this study is ongoing. 24 have been recruited so far (April 2026) from an open recruitment call within Nottingham Trent University, UK. More participants will be recruited from recruitment calls at two partner specialist schools, recruiting a range of students with ASD, with the school itself providing verification of a diagnosis of ASD in these cases. Four further participants have been recruited this way to date. The remainder will be recruited from a wider recruitment call in mainstream education throughout Nottinghamshire. Where a verified diagnosis is not available, self-reporting will be utilized.

In addition, 10 volunteers were recruited from Nottingham Trent University to aid in the creation of the testing software prior to data collection – specifically generating the example avatars and animations themselves. These were recruited to try and ensure a reasonable range of age and sex, with any self-reported ASD diagnoses noted.

3.2. Avatar Creation

The MetaHuman toolkit Fang et al. (2021) allows for humanoid models to be generated using a facial scan generated by a depth sensing camera. In this case, the TrueDepth camera and the LiDAR scanner on Apple's iPhone products were used, utilising Epic Games' integrated solution for generating MetaHuman assets and motion capture, Live Link Face (Games (2023)). This was connected to a target device running Unreal Engine 5.6 via a local wireless network connection. The subject who was to be made an avatar of was handed the recording device, and asked to hold it at approximately chin height, and then stood with their back to a plain, clear, consistent depth background, about a meter away to avoid depth sensing artifacts. The connected PC device was then told to start recording and the participant was taken through a standardised procedure in which they were asked to follow the researcher's finger with their face and gaze while keeping their body steady. The process is adapted from official guidance by Epic Games.

The participant was first to look straight ahead so their gaze passed over the recording device, then slowly (over about 3 seconds in each direction) directed to turn their head to the left and right. Once this is done, the connected device stops recording. Then, a specific frame was chosen based on the visibility of the participant's ears – when both ears were equally visible (or presumed to be so if obstructed by hair), that frame was chosen to form the neutral pose. This was then used to generate a facial mesh, a 3D model of the face, which was then sent to Epic Games for full processing, before being made into a MetaHuman properly.

For the volunteers, the same application was used to record a set of seven facial emotion expressions. These were translated into motion capture using other tools within Unreal Engine. These were invoked using neuropsychologist approved prompting, and based on the eight options used in the test as decribed below. These eight options are common forms of the seven emotional adjectives used in the Facial Action Coding System (Ekman (1992)) plus Neutral. Five of the seven active expressions were chosen at random. This is based on the theory that, rather than strictly impaired emotional processing, people with ASD have a different emotional model of understanding. Therefore, beyond using the adjectives as a baseline, it was decided to avoid strictly validating the animations for this experiment to capture the breadth of potential data sequences. This then avoids one of the primary criticisms of the RMET – namely that is is soley based in a ”Neurotypical” understanding of emotion.

3.3. RA-FEER-VR

The Realistic Avatar based assessment of Facial Emotion Expression Recognition utilising Virtual Reality (the RA-FEER-VR application) was developed for use in this study. In this application, the participant is placed into a plain game world – a box room with neutral white walls, and a white table in the centre. Projected from this table is a simple interface of eight purple buttons with white text. On these buttons were listed Happiness, Sadness, Fear, Surprise, Anger, Disgust, Contempt and Neutral. In the case of non verbal participants, a version with small pictograms of the relevant expressions was used. A short way behind this table is a small plinth, on which a Purple button with START written on it in white text is projected.

The player interacts with this environment through a HTC Vive VR controller – for simplicity, only the trigger is used, used as a pointer. A representation of the player’s hand is displayed in the environment, with a red laser projecting from it. The player can interact with the buttons in the environment by pointing the laser at it and pressing the trigger.

After pressing the start button, a purple wall with a white fixation cross is raised. After approximately a second, it is lowered again, revealing one of the captured MetaHumans, performing one of the captured animations. The player, in their own time, points to and selects the word that they think best fits the expression shown. The wall is then raised again, and the process is repeated – 5 times per model, for 10 models, for a total of 50 tasks.

3.4. Experimental Protocol

After initial briefing and the collection of consent, the Galea headset was shown to the participants. The Galea Beta headset is an integrated ExG and VR system that can be utilised for rapid collection of electrographical data – including EEG data – during VR experiments such as these. Once the participant was comfortable, they were encouraged to put the headset on themselves before the researcher fastens it, ensuring the participant’s comfort at all times.

At this point, following a short test to ensure that the EEG sensors were collecting clean data, the test application was started, and the participant was encouraged to move to a place in the game world where they are comfortable, whilst the test was explained. Then they were encouraged to select the ”Start” button and begin the test.

They were then encouraged to complete the task on their own whilst the headset recorded data. The system also captured the response times. Once this task was completed, the participants were debriefed.

Participants also took a Medical Survey. This was a short questionnaire asking the participant to self-report diagnoses of ASD, ADHD, DCD, Prosopognosia and other relavant cognitive affects, as well as handedness and gender (to compensate for artefacts these can cause in EEG (Darvishi-Bayazi et al. (2025) Ramos et al. (2025))). This questionaire also included the AQ50 (Baron-Cohen et al. (2001)) and the Social Phobia – Adult (LeBeau et al. (2016)) diagnostic tools.

4. Discussion

Data are currently being collected and are expected to be completed and analysed prior to presentation at the conference. However, we have the following expectations, based on previous research:

- In prior research, subjects in a general population see a drop-off in the presence of EEG signals at a specific frequency – the Mu Wavelength – during emotional recognition tasks. We expect this to persist broadly in VR.
- Participants with ASD in previous research see less Mu attenuation. We expect this to persist broadly in VR.
- Previous research used only models that have been validated within a neuro-typical emotional framework. We expect to see some changes in task performance (assessed by responses and response time) and mu attenuation within each group for ASD face models.
- However, it is worth noting in advance the following limitations:
- Aside from diagnoses confirmed by the partner institutions, this study relies on the AQ-50 and self reporting. The self reported ASD phenotype varies substantially from that which is clinically diagnosed. However, if we discover large within-groups variation but small between-groups variation, we can more accurately monitor these differences.
- Dry EEG electrodes have issues with poorer signal quality, and can be disturbed by hand and head movement. The Galea does integrate other forms of ExG, and this can be used to filter some movement artefacts. In many ways, this study is attempting to validate the technology and identify potential improvements and best practice, as outlined elsewhere in these proceedings.

Based on prior research, we are expecting to see a somewhat large effect of ASD on Mu Desynchronisation. Prior to the study, we estimated approximately 26 participants per group would be needed for this study to have sufficient power ($\alpha = 0.05$, $\beta = 0.2$, $d = 0.7$, right-sided test). This estimate appears to be supported by early analysis of the data.

5. Conclusion

Depending on the outcomes of this and other studies performed in the same project, the use of a highly standardised system to create, pose and evaluate social expessions on a humanoid model can be validated as useful for this kind of work. If this is the case, it opens up possibilities for wide-reaching multi-institutional research, enabling much faster progression in investigating various aspects of social perception and interaction. If face tracking can be done whilst a VR headset is being worn, this could extend this approach even further by opening up the possibility of a fully immersive social interactive experience, with facial gestures of all participants being monitored at all times – enabling objective evaluation of facial action, over the time-consuming and unreliable methods of expert witness currently used.

As an alternative, a multimodal approach to understanding FEER deficits could be investigated, analysing EEG, heart rate, broader ExG, etc., in order to gain a fuller image of the nature and variety of these deficits. This could assist in differentiation between often confused disorders or symptoms, increasing availability, and improving the nature of support. Multimodal systems such as this, in this case, would be a very clear avenue for future research within BCI.

References

American Psychiatric Association (2022), *Diagnostic and Statistical Manual of Mental Disorders*, American Psychiatric Association Publishing.

Baron-Cohen, S., Wheelwright, S., Skinner, R., Martin, J. & Clubley, E. (2001), 'The autism-spectrum quotient (aq): Evidence from asperger syndrome/high-functioning autism, males and females, scientists and mathematicians', *Journal of Autism and Developmental Disorders* 31, 5–17.

Bender, A., Voytek, B. & Schaworonkow, N. (2025), 'Resting-state alpha and mu rhythms change shape across development but lack diagnostic sensitivity for attention-deficit/hyperactivity disorder and autism', *Journal of Cognitive Neuroscience* 37, 1581–1615.

Bernier, R., Dawson, G., Webb, S. & Murias, M. (2007), 'Eeg mu rhythm and imitation impairments in individuals with autism spectrum disorder', *Brain and Cognition* 64, 228–237.

Darvishi-Bayazi, M.-J., Ghaemi, M. S., Rish, I. & Faubert, J. (2025), 'Critical role of eeg signals in assessment of sex-specific insights in neurological diagnostics via machine learning approach', *Scientific Reports* 16, 1060.

Ekman, P. (1992), 'Facial expressions of emotion: New findings, new questions', *Psychological Science* 3, 34–38.

Fang, Z., Cai, L. & Wang, G. (2021), Metahuman creator the starting point of the metaverse, in '*2021 International Symposium on Computer Technology and Information Science (ISCTIS)*', IEEE, pp. 154–157.

Fox, N. A., Bakermans-Kranenburg, M. J., Yoo, K. H., Bowman, L. C., Cannon, E. N., Vanderwert, R. E., Ferrari, P. F. & van IJzendoorn, M. H. (2016), 'Assessing human mirror activity with eeg mu rhythm: A meta-analysis.', *Psychological bulletin* 142, 291–313.

Games, E. (2023), 'Scan yourself for mesh to metahuman'. https://dev.epicgames.com/community/learning/tutorials/Ovd2/capturing-reality-scan-yourself-for-mesh-to-metahuman

Higgins, W. C., Savalei, V., Polito, V. & Ross, R. M. (2026), 'Reading the mind in the eyes test scores demonstrate poor structural properties in nine large non-clinical samples', *Assessment* 33, 204–220.

Hobson, H. M. & Bishop, D. V. M. (2017), 'The interpretation of mu suppression as an index of mirror neuron activity: past, present and future', *Royal Society Open Science* 4, 160662.

Huang, Y., Yang, J., Liao, P. & Pan, J. (2017), 'Fusion of facial expressions and eeg for multimodal emotion recognition', *Computational Intelligence and Neuroscience* 2017, 1–8.

LeBeau, R. T., Mesri, B. & Craske, M. G. (2016), 'The dsm-5 social anxiety disorder severity scale: Evidence of validity and reliability in a clinical sample', *Psychiatry Research* 244, 94–96. OpenBCI (2026), 'Galea'.

Ramos, Y. E., Santos, M. T., Yamamoto, I. N. R., da Costa Accioly, C. B., Daneault, J.-F., de Almeida Filho, D. G. & Miranda, J. G. V. (2025), 'Handedness and brain lateralization: A nonlinear motor approach combined with eeg', *Human Movement Science* 104, 103425.

Vellante, M., Baron-Cohen, S., Melis, M., Marrone, M., Petretto, D. R., Masala, C. & Preti, A. (2013), 'The "reading the mind in the eyes" test: Systematic review of psychometric properties and a validation study in italy', *Cognitive Neuropsychiatry* 18, 326–354.

Wagner, J. B., Hirsch, S. B., Vogel-Farley, V. K., Redcay, E. & Nelson, C. A. (2013), 'Eye-tracking, autonomic, and electrophysiological correlates of emotional face processing in adolescents with autism spectrum disorder', *Journal of Autism and Developmental Disorders* 43, 188–199.

Wall, N. G., Smith, O., Campbell, L., Loughland, C. & Schall, U. (2026), 'Using eeg and eye tracking to evaluate an emotion recognition ipad app for autistic children', *Clinical EEG and Neuroscience* 57, 130–140.