Recognizing Faces: Pz classification using a research grade system

Ryan McCardle, Stephen Jacobs, Amir H. Moslehi, T. Claire Davies and Shane D. Pinder

Abstract— **Objective: Our overall research goal is to use electroencephalography (EEG) to identify recognition of faces by patients with dementia to improve reminiscence therapy. The objective of this paper was to evaluate the performance of a g.tec EEG system using a facial recognition event related potential (ERP) task. Previous work has used the Emotiv off-the-shelf system which resulted in lack of classification. For this work, we sought to use a g.tec system with the same electrode placement as the Emotiv Insight to evaluate whether a reduced electrode set would provide sufficient classification accuracy. Methods: EEG was recorded with the same electrode montage as the Insight while obscure and famous facial stimuli were presented; participants confirmed recognition with a button press and oral confirmation. ERPs were averaged across five presentations of each facial stimuli to improve the signal to noise ratio. Classifiers were trained on one session, tested on another session. Results: Using the same electrode montage as the Insight, the g.tec achieved an average accuracy of 80%. Conclusion and Significance: A new research quality system with only the Pz electrode may provide sufficient classification to develop a system for reminiscence therapy.**

Index Terms—Emotiv Insight, brain computer interface (BCI), electroencephalography (EEG), event related potential (ERP), recognition memory

I. Introduction

Reminiscence therapy (RT) is an effective therapy for patients with dementia that requires a caregiver to encourage and facilitate patient recall of intact memories from their past [1]. There is clinical potential for a system supported by electroencephalography (EEG) to enhance RT; one component of which would be to determine which topics best stimulate memory recall.

We have shown that it is not feasible to use an off-the-shelf system, specifically the Emotiv Insight with 5 electrodes, to accurately assess recognition [2]. We further found too many artefacts present in many of the electrodes of the Insight. However, the Pz electrode location offered the best possibility for detecting recognition. A research system that focuses on only the Pz electrode might allow classification of recognition.

We hypothesised that classifying facial recognition ERPs using the g.tec with a focus on the Pz electrode location would provide classification results better than the Insight.

II. EEG System Comparison

In EEG BCI research, signals are usually under 30 Hz thus the Emotiv Insight sampling rate of 128 Hz should be reasonable to capture all relevant brain activity. However, a much higher internal sampling rate allows the g.tec system to oversample resulting in improved resolution and reduced noise.

*R. McCardle is with the Department of Mechanical Engineering, Faculty of Engineering, The University of Auckland, Auckland, New Zealand.

S. Jacobs is with the Department of Nursing, Faculty of Medical and Health Sciences, The University of Auckland, Auckland, New Zealand.

T. C. Davies and A. H. Moslehi are with the Department of Mechanical and Materials Engineering, Faculty of Engineering and Applied Science, Queen’s University, Kingston, Ontario, Canada (correspondence e-mail: claire.davies@queensu.ca).

III. Methods

A. Participants

Ten healthy participants (3 females), aged 18–50 yrs were initially recruited [2]. All participants had normal or corrected vision and watched popular TV shows and movies.

All participants tolerated the sessions well, with only occasional evidence of fatigue towards the end of sessions that did not appear to affect their performance. Data from one participant was removed due to a technical issue. Informed consent was obtained for all participants and the experiment was approved by the University of Auckland Human Participants Ethics Committee 8895.

B. Experimental Paradigm

The trials consisted of two 1.5h sessions for each participant. Before the first session, participants identified potentially recognisable images based on written excerpts of 377 famous persons in the database. The stimuli database was made up of obscure (people the participant would have never seen before), and famous (people in the media the participant may have seen before) (see [2] for more detail).

Each facial stimulus was repeated five times in each session and in random order. Figure 1 shows the flow and timing of a single trial. To reduce artefacts, participants were asked to fixate on a cross and refrain from blinking, making large eye movements, or moving their body or face (except for during the 1s response window or break periods).

If the participants responded too early, the response was not recorded. If the participant responded that they recognised a facial stimulus, verbal confirmation was required. The experimenter waited until the EEG artefacts caused by speaking were no longer visible and indicated with a key press if the person was correctly identified.

C. Experimental Protocols with EEG System

For a valid comparison of the accuracy in identifying facial recognition responses between the g.tec system and the Insight, only the Pz was recorded. EEG was recorded at 1200 Hz via BCI2000 [3] with a g.GAMMAbox (g.tec, Austria) active electrode amplifier connected to a g.USBamp. No filters were used during recording. g.ACTIVEelectrode Ag/AgCl electrodes were used, with a g.GAMMAearclip for the reference electrode.

Hair was moved aside, and the scalp was lightly abraded before application of electrode gel. Since electrode impedance to measure contact quality cannot be measured with active electrodes, care was taken to ensure consistent application of electrodes. EEG signals were monitored throughout the experiment using the BCI2000 interface and electrodes were reapplied if necessary.

978-1-5090-0617-5/19/$31.00

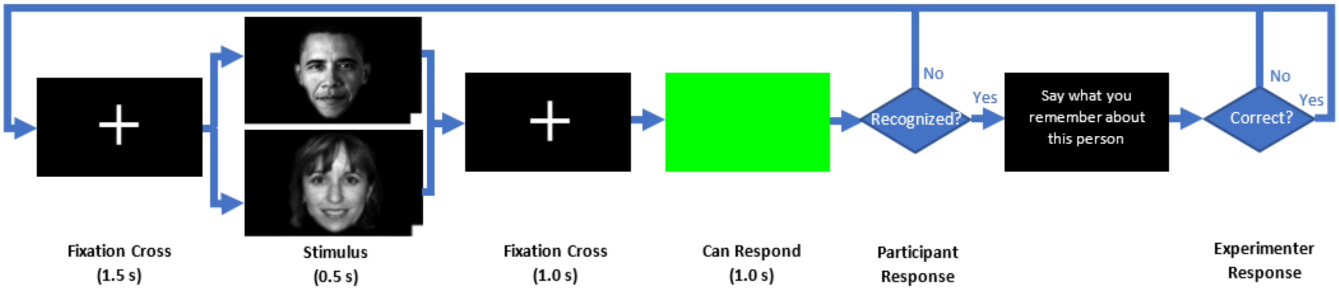


Figure 1. Experimental paradigm, showing screenshots of the participants' screen with flow of each trial and timing.

D. Experimental Setup

Trial flow and stimulus presentation was handled via a custom MATLAB program, using the Psychophysics Toolbox extensions [4], [5]. To ensure ERPs were recorded accurately, precise timing of events was essential. This can be difficult to achieve with common desktop computers and operating systems [6], and additional measures were taken to reduce timing errors, as follows:

A micro-controller was used to precisely record stimulus presentation and participant response times, with a trial proceeding as follows:

i) A stimulus was presented to the participant.
ii) The micro-controller recorded the presentation time via a light sensor and caused the digital input on the EEG amplifier to register as high.
iii) When the screen went green, the participant responded via a button connected to the micro-controller.
iv) The micro-controller caused the digital input on the EEG amplifier to register as low and passed the response type to the stimulus-presentation computer via serial [2].

The response buttons used by the participant were small push-buttons, red for unfamiliar (pressed with the right index finger), black for recognized (middle finger). The EEG recording was combined with the recorded stimulus and response types later with no need to synchronise the clocks between the presentation computer and the EEG amplifier.

E. Data Preprocessing

Recorded data were loaded into EEGLab for preprocessing. Presentation times were extracted from leading edges of the digital input output (DIO) channel (g.tec) and linked with the corresponding responses' information. Break periods (counted as gaps between trials greater than 8s in length) and the start and end of the recording were rejected from the data. Removing break periods from the data ensured that large artefacts from the participants' unrestricted movement in breaks, or the reapplication of electrodes did not bleed over into the rest of the data during filtering.

EEGLab is designed to accept timestamps that monotonically increase by the inverse of the sampling rate which occurs with the g.tec system. To correct for samples appearing out of order, the timestamps were first sorted. Samples with duplicate timestamps were sorted by the sample counter and then the timestamps were interpolated between the first sample and the next sample with a higher timestamp. The EEG recordings and their timestamps were then spline interpolated to generate a constant 256 Hz sampling rate before processing.

To correct for baseline shifts, a 1 Hz high-pass zero-phase finite impulse response (FIR) filter was applied. To remove 50Hz New Zealand powerline noise and other high frequency noise, a 30Hz low-pass filter was then applied. Trials were epoched to 1.5s after stimulus presentation and the baselines were removed by subtracting the average of the EEG signal 200ms before stimulus presentation. Trials were then cut to a window of 300–1000ms after the stimulus presentation before being used for classification.

F. Classification

Individualised classification for each participant included training of the classifier on one session and testing on the other. An averaged event related potential (aERP) for each stimulus was made by averaging across the number of stimulations (1–5). EEG was then normalised by dividing it by the root mean square (RMS) of the training session.

For classifiers utilising features, the following feature vector was generated: the average response latency [1x1], the prediction of the pattern match classifier [1x1], the mean squared error (MSE) between the trial and the Unfamiliar and Recognised aERPs [2x1], the z-score [7] with a window length of 300ms and a p-value threshold of 0.01 [1x19], the area under the curve [1x1], the average (calculated for each presentation of the stimulus and then averaged) power spectral density up to 40 Hz given by MATLAB's implementation of Welch's power spectral density [41x1].

All features were individually normalised before being combined into the final feature vector (final length of 1541) by dividing by the RMS of the training set.

G. Classifiers

A number of different classifiers were trained on the dataset and their accuracies were compared. These included: a pattern matching classifier, which compared the MSE between the current trial and the aERP for each class, a linear discriminant analysis (LDA) classifier (MATLAB 2014a implementation) with principal component analysis (PCA) used to decrease the dimensionality to 40 components, a class-wise principal component analysis (CPCA) classifier using the MATLAB library created for electrocorticography (ECoG) and EEG classification, using approximate information discriminant analysis (AIDA) for dimensionality reduction.

To provide a baseline against which the other classifiers could be measured, two baseline classifiers were also created: a response-latency classifier, which was an LDA classifier with the average response-latency as the only input and a shuffled classifier, which was the CPCA classifier except the training output was shuffled randomly before any of the feature extraction and normalisation steps.

IV. Recorded EEG and Classification Results

All participants performed the task successfully and kept electrooculography (EOG) and electromyography (EMG) artefacts at acceptable levels. However, although the data from all participants appeared usable during the data collection phase, post-experiment processing revealed this was not the case.

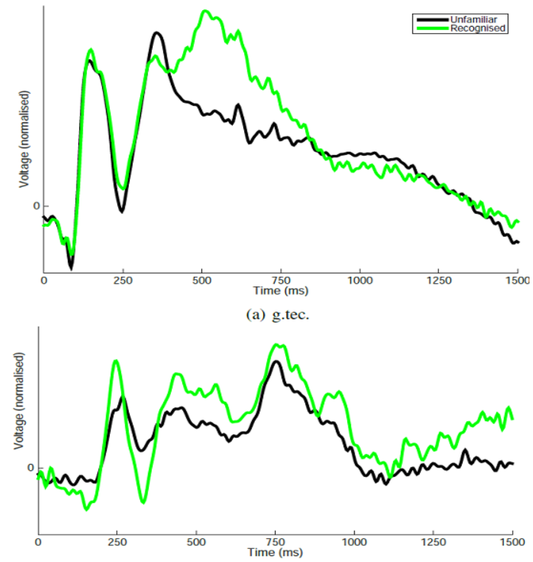
Data from two participants were rejected for having a high correlation between response latency and class. To compare directly to the Insight study [2], an additional three participants were rejected from the g.tec dataset. These were randomly deleted based on the identifier used in the study (rather than picking and choosing the best or worst data). Additionally, the high number of trials rejected due to artefacts in the Insight recordings reduced the size of the training set and compromised classification accuracy. Therefore, a random 10% of the g.tec trials for each participant were rejected to better compare with the Insight.

Figure 2 shows a comparison between the aERPs averaged across participants for both systems. The g.tec aERP (fig. 2a) is similar between the two classes from 1–400ms, then diverges before coming back together at around 900ms.

This divergence is consistent with the FN400 (or old/new effect) and appears similar to aERPs from other studies [8]–[12]. The Insight aERP (fig. 2b) however is much less similar between classes across the entire range, with the recognised class showing a much more pronounced N170 spike than the unfamiliar class.

The difference between the Insight and g.tec results is more apparent when considering the change in p-value between the unfamiliar/recognised classes over time, fig. 3. To generate fig. 3, a two-sample t-test was taken at each time point—comparing the values at this time point for every unfamiliar and recognised trial. The p-value at each time point was then taken and graphed on a log scale. As expected, for the g.tec system the difference between the unfamiliar and recognised aERPs tracks the change in p-value fairly closely, with a large spike in the 500–750ms range. For the Insight system the changes in p-value are much less consistent, with spikes occurring across the entire time. Low p-values indicate there was a large difference between the recognised and unrecognised classes in the g.tec recordings, but much higher p-values indicate that this difference is less pronounced in the Insight recordings.

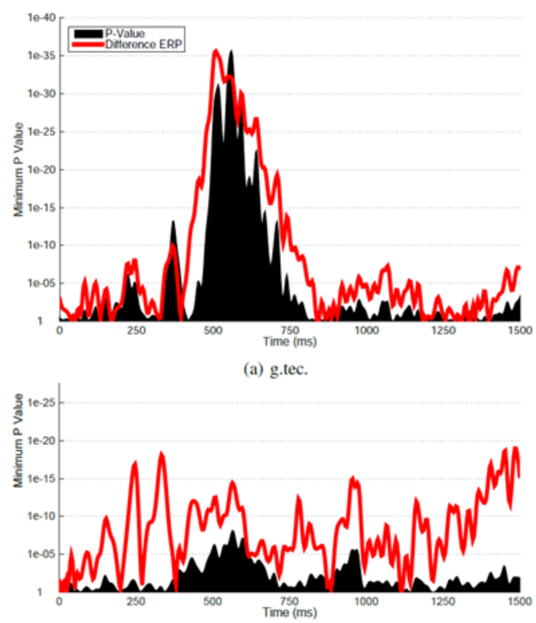
Overall, electrode artefacts are less frequent in the g.tec recordings. This is likely because the g.tec electrodes are more firmly attached to the scalp (by both the EEG cap and electrode gel) so electrode movements are reduced, and when they do occur, the active electrodes reduce the size of the resulting artefacts. This firm attachment, electrode gel, and the scalp preparation also reduce impedance at the electrode which can improve the overall signal to noise ratio. Another reason for the high number of trials rejected due to artefacts in the Insight may be the onboard filtering in this device.



(a) g.tec.

(b) Insight.

Figure 2. Cross-participant ERPs from Pz showing differences between classes for the two systems (a) g.tec and (b) Insight.



(a) g.tec.

(b) Insight.

Figure 3. Change in p-value between classes over time, overlaid with the difference between the normalized aERPs of each class for the two systems (a) g.tec and (b) Insight.

The ERP images for both systems also exhibit the difference in recording quality between the Insight and g.tec systems. The g.tec system exhibits excellent temporal alignment of the ERPs with very little variation between trials which leads to a sharp aERP. In the Insight, there is less consistent temporal alignment due to jitter, which leads to a more blurred aERP. The ERPs for the Insight also start around

100ms later than the g.tec system, suggesting there is an extra delay somewhere in the experiment hardware.

Previous work using all trials and a full electrode array [2] resulted in an average classification accuracy of 89%. To help compare the classification results between classifiers, 95% confidence intervals were calculated and plotted in fig. 4. In this study, the g.tec results had an average classification accuracy of 80%.

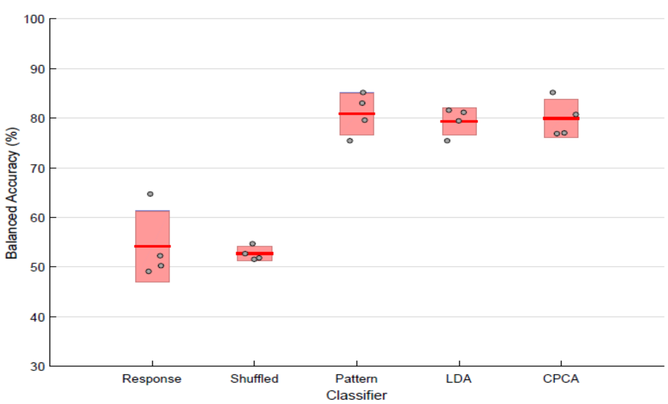
The g.tec system produces fairly tight 95% confidence intervals for all classifiers (fig. 4a) and an average accuracy of 80% for the CPCA classifier. The LDA, CPCA, and Pattern classifiers are all well above the 95% confidence intervals for the Response and Shuffled baseline classifiers. Therefore, we can be fairly sure that the non-brain signal of response latency is not being used to artificially increase accuracies of the classifiers and there are unlikely to be any train/test separation problems.

V. Discussion

The g.tec system produced classifications with high accuracy using only the Pz electrode whereas the Insight classification was similar to random chance. The only significant difference between the Insight and g.tec experimental paradigms was the event integration, due to the Insight lacking a hardware event integration feature. The g.tec results show that repeating the stimulus multiple times greatly improves the accuracy (for the Insight, no effect).

Live artefact rejection could improve results by repeating the stimulus until enough artefact free trials are obtained—at the cost of longer experimental sessions. The problems from electrode artefacts taking several seconds to subside could be improved simply by leaving longer breaks after any period that could introduce extra artefacts—however this would also increase session length.

A limitation of the current study is that it does not use the same subjects with both the g.tec and the Insight systems. Some of the differences may be due to the individual subject differences rather than the performance of the systems. However, all the g.tec recordings were of good quality whereas all of the Insight recordings were of lower quality suggesting that personal/human variance are not factors contributing to the recorded differences between the systems. A further limitation is that jitter in the EEG recordings was not compared for the Insight and g.tec devices. Hairston et al. [13] describes a measurement of the jitter in the Emotiv EPOC which was often around 50ms. As the recording electronics are likely similar in the Insight, it is not surprising that jitter was also observed in the current experiment. Improvements to the Insight firmware or MATLAB SDK by Emotiv Systems could reduce this jitter, but researchers should also investigate techniques to correct for this, such as that described by Hairston et al. [13].



(a) g.tec.

Figure 4. Classifier accuracy with 5 stimulus presentations showing 95% confidence interval (red rectangles), mean (thick red lines), and individual samples (grey dots).

Balanced Accuracy (%)

VI. Conclusions

A research grade EEG system with one electrode could provide enough of an ERP to use in RT. The advantages include: it only takes a few seconds to don, and a minute or two to achieve good contact quality. It is also fully self contained with no wires to restrict movement of the participant and looks less intimidating. These aspects are helpful for research with healthy participants, but essential for work with dementia patients.

References

[1] L. Pittiglio, "Use of reminiscence therapy in patients with Alzheimer's disease.," *Lippincotts. Case Manag.*, vol. 5, no. 6, pp. 216–220, 2000.

[2] R. D. McCardle et al., "Classifying Faces: Can an off-the-shelf system be effective?," BIOCAS, 2019, paper 8051.

[3] G. Schalk, D. J. McFarland, T. Hinterberger, N. Birbaumer, and J. R. Wolpaw, "BCI2000: a general-purpose brain-computer interface (BCI) system.," *IEEE Trans. Biomed. Eng.*, vol. 51, no. 6, pp. 1034–1043, Jun. 2004.

[4] M. Kleiner, D. Brainard, and D. Pelli, "What's new in Psychtoolbox-3?," in *Perception ECVP Abstract Supplement*, 2007, vol. 36, no. 0, p. 0.

[5] D. H. Brainard, "The Psychophysics Toolbox.," *Spat. Vis.*, vol. 10, no. 4, pp. 433–436, 1997.

[6] J. Wolpaw and E. W. Wolpaw, *Brain-Computer Interfaces: Principles and Practice*. Oxford University Press, USA, 2012.

[7] A. Lage-Castellanos, J. I. Nieto, I. Quinones, and E. Martinez-Montes, "A zero-training algorithm for EEG single-trial classification applied to a face recognition ERP experiment.," *Conf. Proc. ... Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. IEEE Eng. Med. Biol. Soc. Annu. Conf.*, vol. 2010, pp. 4209–4212, 2010.

[8] M. Eimer, "The Face-Sensitivity of the N170 Component," *Front. Hum. Neurosci.*, vol. 5, p. 119, Oct. 2011.

[9] Brian D. Earp; Jim A. C. Everett, "Is the N170 face specific? Controversy, context, and theory," *Neuropsychol Trends*, vol. 13, no. 1, pp. 7–26, 2013.

[10] C. Saavedra, J. Iglesias, and E. I. Olivares, "Event-related potentials elicited by face identity processing in elderly adults with cognitive impairment.," *Exp. Aging Res.*, vol. 38, no. 2, pp. 220–245, 2012.

[11] T. Curran and J. Hancock, "The FN400 indexes familiarity-based recognition of faces.," *Neuroimage*, vol. 36, no. 2, pp. 464–471, Jun. 2007.

[12] G. Yovel and K. A. Paller, "The neural basis of the butcher-on-the-bus phenomenon: when a face seems familiar but is not remembered.," *Neuroimage*, vol. 21, no. 2, pp. 789–800, 2004.

[13] W. D. Hairston, "Accounting for Timing Drift and Variability in Contemporary Electroencephlography (EEG) Systems." 01-Mar-2012.