2021 43rd Annual International Conference of the
IEEE Engineering in Medicine & Biology Society (EMBC)
Oct 31 - Nov 4, 2021. Virtual Conference

Analysis of eyewitness testimony using electroencephalogram signals

B.V. Mendes[1], A.M. Tomé[2], I.M. Santos[3], P. Bem-Haja[4]

Abstract— Face recognition and related psychological phenomenon have been the subject of neurocognitive studies during last decades. More recently the problem of face identification is also addressed to test the possibility of finding markers on the electroencephalogram signals. To this end, this work presents an experimental study where Brain Computer Interface strategies were implemented to find features on the signals that could discriminate between culprit and innocent. The feature extraction block comprises time domain and frequency domain characteristics of single-trial signals. The classification block is based on a support vector machine and its performance for the best ranked features. The data analysis comprises the signals of a cohort of 28 participants.

I. INTRODUCTION

Brain signals provide relevant information regarding the mental state of a human subject. There are different ways to acquire brain signals but the electroencephalogram (EEG) is the most used technique, specially because it is non-invasive and possible using portable devices. Therefore, EEG signal recordings are widely used in neuroscience and psychology for the study the brain and its functioning. Applications such as cognitive and affective monitoring are very promising as they could allow unbiased measures of levels of fatigue, mental workload, mood or emotions [1]. Finally, EEG is widely used in brain-computer interfaces to allow brain activity to be directly translated into commands to output devices that carry out desired actions [2]. Although EEG has proven to be a critical tool in many domains, it still suffers from a few drawbacks that require further developments. First, EEG has a low signal-to-noise ratio (SNR) as the relevant brain activity measure is often buried under other physiological or environmental signals. Neurocognitive studies address this problem by performing time-locked averages of repeated trials while brain computer interfaces rely on the application of signal processing methods (like independent component) to extract the task related components of the signal. Another important issue is the high inter-subject variability. In most Brain Computer Interfaces (BCIs) to obtain a reliable system an off-line calibration step is often used. For this step a training data set is pre-recorded from the user. In this stage the classification algorithm is calibrated and often the optimal features are selected.

In this work, we propose to follow the BCI Pipeline which consists in feature extraction, feature ranking and classification to select the most relevant features of EEG signals of participants performing a face recognition task [2]. The extracted features provide time and frequency domain signal's characteristics and they were calculated on single-trials signals. Before classification an univariate test (ANOVA) was applied to select the highest ranked features. Finally, a linear and non-linear classification scheme based on the support vector machine (SVM) was applied. The classification methodology was applied for each participant and the more frequent selected features were studied.

[1] University of Aveiro, DETI/ IEETA, 3810-193 Aveiro, Portugal brunovm@ua.pt
[2] University of Aveiro, DETI/ IEETA, 3810-193 Aveiro, Portugal ana@ua.pt
[3] University of Aveiro, Department of Education and Psychology, William James Center for Research, 3810-193 Aveiro, Portugal isabel.santos@ua.pt
[4] University of Aveiro, Department of Education and Psychology, CINTESIS, 3810-193 Aveiro, Portugal pedro.bem-haja@ua.pt

II. MATERIALS AND METHODS

All of the processing steps were performed at an intra-participant level [3] and at trial level. The analysis was performed on a set of signals of a cohort of 28 participants.

A. Dataset

The existing database was the result of an experiment performed in [4] following the principles outlined in the Helsinki Declaration. A total of eight theft videos of twenty seconds were displayed to 28 participants in which the culprit was presented in frontal view during four seconds and not in frontal view the rest of the time. Afterwards, the EEG was recorded while participants watch face images (of culpits and of distractors) in gray-scale and emotionally neutral. The lineup of six images includes five distractors and a culpit, in random order, were presented 10 times for each participant. Each face was displayed for one and a half seconds and its appearance is the stimulus considered for the experiment. Then, each participant was asked to classify each face as a culprit or a distractor. The EEG data was recorded from 32 electrodes mounted on a waveguard cap (ANT™) according to the 10-20 system at a sampling rate of 2048 Hz.

B. Preprocessing

The EEG signals were downsampled to 512 Hz after applying a zero-phase filtering strategy. Afterwards the signals were epoched around the stimulus (-0.5 seconds before and 3 seconds after). Trials where the participant did not answer or gave multiple answers were discarded and the remaining were split according to four different classes: culprit and right answer, culprit and wrong answer, distractor and right answer, and distractor and wrong answer.

C. Feature Extraction

Two different types of signal characteristics can be extracted from brain signals: frequencies related features which describe the energy in distinct frequency bands and the time related features which describe the stimulus elicited evoked potential related components. Frequency analysis and time domain using MATLAB tools, specially EEGLAB [7].

1) Frequency domain: Three rhythmic activities (theta, alpha and beta) bands were selected for this study. The theta waves correlate with mental operations such as focused attention, learning or memory recalls. The beta waves are related to active, busy or anxious thinking [5]. Finally, studies like [6] have been made concerning the asymmetry of the alpha waves in questions of emotion and motivation.

The spectrogram methodology (available in MATLAB) was applied for the single-trial signal divided into segments of 0.25s seconds with an overlap of 0.125s. Threfore, a total of 27 subsegments (the first 3 pre-stimulus and the last 23 post-stimulus) were obtained. The frequency analysis of each sub-segment, multiplied by a Hamming window, is calculated for the range [4, 30)Hz with a resolution of 0.1Hz. Further processing of the frequency analysis comprises calculation of the energy in the three bands of the EEG signal: theta, alpha and beta. The energy in theta, alpha and beta band were calculated adding the energy of the energy values belonging to the [4, 8)Hz, [8, 13)Hz and [13, 30)Hz respectively.

2) Time domain: In order to study the EEG signals through a time perspective point of view, the signal trial signal is read regarding the cognitive operations that are happening and mapped on peaks.

These peaks are enhanced and visible in epoch-averaged signals and are called Event Related Potentials (ERPs) [1]. An ERP provides two kinds of characteristics: whether the peak is positive or negative, and the latency of the peak. For example, a P200 provides the amplitude value for the Positive-going peak that occurs around 200 milliseconds after a given stimulus. According to [8], a stimulus is the prime independent variable of a psychological experiment. According to the literature [9] [10] [4], the most relevant ERPs components in face recognition are the P100 (80ms to 120ms) which is mostly related to any visual evoked potential, the N170 (150ms to 190ms) associated with visual processing of stimuli and the P300 (300ms to 600ms) which relates to the occurrence of rare events. Study [10] also takes into account the P200 (180ms to 220ms). To extend this characterization to single-trials, the signals were smoothed by an infinite impulse response (IIR) filter. The 4th order IIR filter was passband filter between 2 and 12 Hz designed using the Butterworth approach. The zero-phase filtering strategy was used to assure synchronization between the filter's output signal and the stimulus. Note that the global attenuation (in dB) is the double of the magnitude of the filter frequency response. All 30 EEG channels were taken into consideration and the amplitude and latency of all four ERP components were extracted.

D. Feature Vectors

Each feature extraction method will produce a feature vector from one of two types: a frequency feature vector or a time feature vector. A total of 3 frequency bands of interest, 27 time intervals and 30 channels were taken into account make 2430 features per trial which can be concatenated into a frequency vector with dimension $D = 3 \times 27 \times 30$. In the time domain, 4 amplitude and 4 latency values related to the four accounted single trial ERPs per each 30 EEG channels make a total of 240 features which can also be concatenated into a time vector with dimension $D = 4 \times 4 \times 30$.

E. Balancing the Dataset

The present dataset is unbalanced with only 10% of the answers provided were wrong and the experiment itself consisted in 5 distractor face for each culprit and this problem usually affects the behavior of machine learning models. Because of that, this study proceeded with only the correctly identified trials and therefore it is a binary classification problem.

To balance the data of the binary classification problem, the number of trials related with distractors was reduced to match the number of trials related with culpits. The priority in this last point was given to the first trials based on the fact that, when a subject performs a task, over time the detection rate tends to decrease [11]. After balance, the total trial number was around 140 per participant.

F. Classification Pipeline

A complete diagram describing all the implementation steps for the classification with feature ranking is displayed in figure 1. The system was implemented using scikit-learn packages [13]. A detailed description can be found in [14].

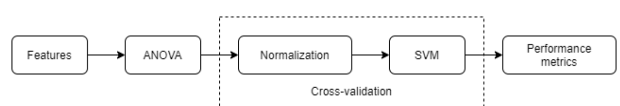

Fig. 1. Classification using feature ranking.

1) Feature Ranking: The selection of the most important features was done by ranking all the features using the ANOVA technique. The ANOVA is a statistical test that can be used to analyze differences between two or more groups of data. It usually uses the f-test score and makes a single, overall decision as to whether

a significant difference is present among sample means of the groups [12]. The discriminating power of each feature is evaluated individually and assigned a value which will be used for ranking purposes. After being tested, each feature is assigned a rank value and it is that value that will be used to rank the features by their importance.

2) Normalization: After feature selection, the next block performs feature normalization (see figure 1). In this work, the z-score methodology to remove the mean and scaling to unit the variance the features [13] was applied. The normalization parameters (mean and standard deviation) were calculated in the training set and applied also in the test sets. That way, normalization is also included in the cross-validation strategy.

3) Support Vector Machines: Having few trials, it was decided to use the SVM for the classification since it presents good performance with small training datasets [15]. Support Vector Machines were developed in the 90s but are still very popular nowadays in solving supervised learning problems. They tackle this problem by trying to find a hyperplane that can distinctly classify each example either in input space (linear kernel) or in the mapped feature space (for instance RBF kernel). In other words, the parameters of the classifier determine a hyperplane decision boundary either in the input space or the mapped feature space [16].

4) Cross-Validation: All models were evaluated, at participant level, by a cross-validation ($k = 5$) and the preferential measure chosen was the accuracy score [13]. In a cross-validation, the data is divided into k partitions and there are k iterations with training and test phases. In each iteration, one partition is the test set while the remaining ones are used as training set. The final accuracy is the average of all the partial accuracies. Note that this way all examples are used as test examples.

III. RESULTS

The inter-participant accuracy results were obtained by performing an average across all the participants' accuracy values. And the relevant features identified for each subject were compared between participants. After some initial testing and observation, the ideal number of selected features in each participant using the ANOVA ranking ended up being around 100 frequency features and 50 time features.

A. Classification Performance

The table I summarizes the classification accuracy of the performed experiments. The input of the classifier comprises frequency features or time domain features or both. The SVM uses either RBF or linear with the exception of the feature fusion case. In general better performance is achieved with RBF kernel using as input frequency domain features, note that a slight improvement is achieved with both types of features.

TABLE I
Classification results: Accuracy values between 0 and 1.

Feature vector w/ SVM kernel	Acc. (mean)	Acc. (std)	Acc. (range)
freq. w/ RBF	0.853	0.073	0.717 - 0.983
freq. w/ linear	0.819	0.097	0.633 - 1.000
time w/ RBF	0.713	0.051	0.591 - 0.801
time w/ linear	0.657	0.066	0.495 - 0.783
freq. and time w/ RBF	0.864	0.073	0.700 - 1.000

B. Feature Relevance

The inter-participant univariate analysis of the features was studied. With that goal, the choices of the four scalp regions were compared. The EEG channels were aggregated as follows:

- Frontal region: Fp1, Fp2, Fpz, F1, F3, F2, F4, Fz, FC5, FC1, FC2 and FC6.
- Parietal region: Cz, C3, C4, CP1, CP2, CP5, CP6, Pz, P3, P4, P7 and P8.
- Temporal region: T7 and T8.
- Occipital region: POz, O1, O2 and Oz.

Then for each region the number of times one feature was selected in the cohort of participants is calculated.

1) Frequency domain: The alpha band was the less predominant one of all three, with only 18% of all selected features. The theta band was expected to dominate the graph due to its relation to memory, specially with episodic memory as shown in studies like [17], but that place was claimed by the beta band with 53% against the 29% from the theta band. This is specially evident in the temporal channels (T7 and T8) which can be seen in figure 2 which shows the 148 most selected features with each one appearing in at least 4 participants.

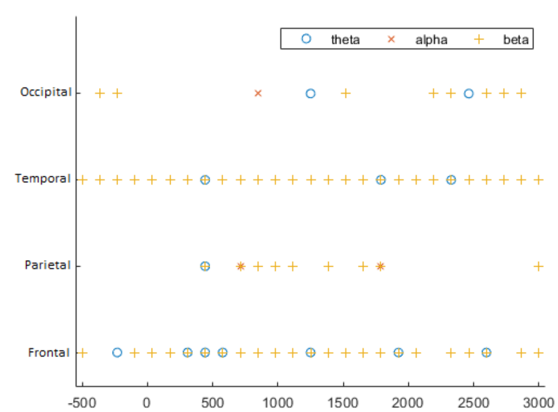

Fig. 2. Most common frequency features selected . The horizontal axis represents time in milliseconds being 0 the instant of the stimulus.

One can also point out that less features are chosen as time goes by as 39% of the features selected were from the first second post-stimulus, 29% from the second and 23% from the third and last second. Around 9% of the features selected were from a time before the stimulus.

After analysing the feature relevance, the scalp temporal features in the beta frequency band were noticed to be present in great number across all the participants. Therefore, an SVM RBF model, having as input beta band features of the temporal channels, was evaluated. The performance results aren't as good as the one having as input the first hundred highest ranked features (by the ANOVA technique). However, if it is taken into account the extreme simplicity of the procedure, just two electrodes and their energy points between [13, 30)Hz producing a total of only 58 features, the accuracy values were very promising ($68,9 \pm 12\%$). Despite the average accuracy value, the range and even the maximum accuracy value show promising results for possible future works.

2) Time domain: Figure 3 shows the 36 most predominant features that appear in at least 12 participants when using ANOVA for feature selection. In this study the features of temporal region were the less relevant and the P100 component related features are also not selected. The absence of the P100 makes sense since it is an ERP that is present in every visual stimulus regardless of its type. And note that the ERP components were detected in smoothed signals, e.g, without the beta band contribution. The most relevant components are related with N170, P200 and P300 amplitudes localized in frontal and parietal regions.

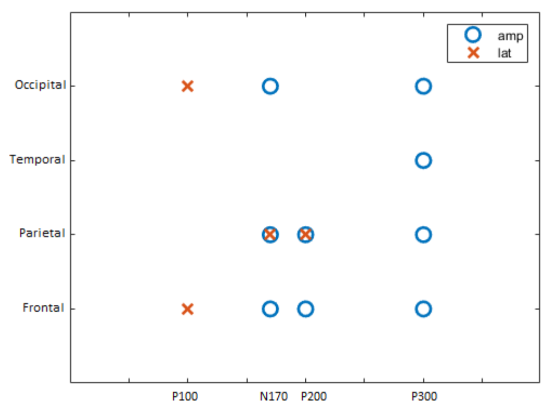

Fig. 3. Most common time features selected.

IV. Conclusion

In this paper, the BCI's methodologies were used in order to study the feature relevance related to eyewitness testimony. It is worth noting that this study follows an intra-participant methodology. Applying a feature selection and a classifier optimized for a participant and using it on another one does not achieve a good performance (between 45% and 60% of accuracy). This proves that, despite the common selected features between participants, this type of methodology should be followed. The most interesting conclusion was the relevance of the beta band on the temporal regions due to its simplicity, since a very decent level of accuracy was obtained using only two EEG channels on one frequency band.

ACKNOWLEDGMENT

This work was funded by National Funds through the FCT - Foundation for Science and Technology, in the context of the project/R&D Unit UIDB/00127/2020 (IEETA), UIDB/4255/2020 (CINTESIS), and UIDB/04810/2020 (William James Center for Research).

References

[1] S. Sanei, J. A. Chambers, EEG Signal Processing. isbn: 9780470025819. doi: 10 . 1002/9780470511923. 2013.
[2] A. Bablani, D. R. Edla, D. Tripathi, R. Cheruku, Survey on Brain-Computer Interface: An Emerging Computational Intelligence Paradigm. ACM Comput. Surv. 52, 1, Article 20, 32 pages. doi: https://doi.org/10.1145/3297713. 2019.
[3] F. Lotte, L. Bougrain, A. Cichocki, M. Clerc, M. Congedo, A. Rakotomamonjy, F. Yger, A Review of Classification Algorithms for EEG-based Brain-Computer Interfaces: A 10-year Update, Journal of Neural Engineering. 15. 10.1088/1741-2552/aab2f2. 2018.
[4] P. Bem-Haja, Psychophysiology of Eyewitness Testimony, Ph.D. dissertation, Universidade de Aveiro, Departamento de Educação e Psicologia. Online: http://hdl.handle.net/10773/22797. 2018.
[5] IMotions, Electroencephalography - The Complete Pocket Guide, 2019.
[6] E. E. Smith, S. J. Reznik, J. L. Stewart, J. J. B. Allen, Assessing and Conceptualizing Frontal EEG Asymmetry: An Updated Primer on Recording, Processing, Analyzing, and Interpreting Frontal Alpha Asymmetry, International Journal of Psychophysiology. doi:10.1016/j.ijpsycho.2016.11.005. 2016.
[7] A. Delorme, S. Makeig, EEGLAB: an open-source toolbox for analysis of single-trial EEG dynamics, Journal of Neuroscience Methods 134:9-21. 2004.
[8] J. Gibson, The concept of the stimulus in psychology. American Psychologist, 16, 694-703. American Psychologist. 15. 694-703. 10.1037/h0047037. 1960.
[9] M. V. Martin, V. Cho, G. Aversano, "Detection of subconscious face recognition using consumer grade brain-computer interfaces", ACM Transactions on Applied Perception, vol. 14, no. 1. issn: 15443965. doi: 10.1145/2955097. 2016.
[10] M. Hoshiyama, R. Kakigi, S. Watanabe, K. Miki, Y. Takeshima, Brain responses for the subconscious recognition of faces, Neuroscience Research, vol. 46, no. 4, pp. 435–442, issn: 01680102. doi:10.1016/S0168-0102(03)00121-4. 2003.
[11] Verity J. Brown, Eric M. Bowman, Alertness in Encyclopedia of the Human Brain. 2002.
[12] P. Mishra, U. Singh, C. M. Pandey, P. Mishra, G. Pandey, Application of student's t-test, analysis of variance, and covariance. Ann Card Anaesth. 22(4):407-411. doi:10.4103/aca.ACA_94_19. 2019.
[13] Pedregosa et al., Scikit-learn: Machine Learning in Python, JMLR 12, pp. 2825-2830. 2011.
[14] B. Mendes, Analysis of eyewitness testimony using electroencephalogram signals, Master's Thesis, Universidade de Aveiro, Departamento de Eletrónica, Telecomunicações e Informática. Online: http://hdl.handle.net/10773/31348. 2021.
[15] X. Wu, V. Kumar, Q. J. Ross, J. Ghosh, Q. Yang, H. Motoda, G. J. McLachlan, A. Ng, B. Liu, P. S. Yu, Z. H. Zhou, M. Steinbach, D. J. Hand, D. Steinberg, Top 10 algorithms in data mining, Knowledge and Information Systems. issn: 02191377. doi: 10.1007/s10115-007-0114-2. 2008.
[16] T. Evgeniou, M. Pontil, Workshop on Support Vector Machines: Theory and Applications, Machine Learning and Its Applications: Advanced Lectures, pp. 249–257. doi: 10.1007/3-540-44673-7. 2001.
[17] S. C. Berens, A. J. Horner, Theta Rhythm: Temporal Glue for Episodic Memory. doi: 10.1016/j.cub.2017.08.048. 2017.