New-onset Alzheimer's Disease and Normal Subjects 100% Differentiated by P300

Barrie Jervis, C Bigan, M Jervis, M Besleaga

New-onset Alzheimer's Disease and Normal Subjects 100% Differentiated by P300

B. W. Jervis[1] PhD, C. Bigan[2] PhD, M. W. Jervis[1] BSc, M. Besleaga[3] Clinical Neurologist.

Affiliations**:** [1]Personal contribution, [2]Ecological University of Bucharest, Bucharest, Romania, [3]Neuroptics Med Clinic, Bucharest, Romania.

Corresponding author: B. W. Jervis, 115 Button Hill, Sheffield, S11 9HG, England, Tel. 44 (0)114 2962170, e-mail barrie.jervis@cantab.net

Disclosure: The authors have reported no conflicts of interest.

Abstract

Previous work has suggested that evoked potential analysis might allow the detection of subjects with new onset Alzheimer's Disease, which would be useful clinically and personally. Here, it is described how new-onset Alzheimer's disease subjects have been differentiated from healthy, normal subjects to 100% accuracy, based on the back-projected independent components (BICs) of the P300 peak at the EEG electrodes in the response to an oddball, auditory evoked potential paradigm. After artefact removal, clustering, selection, and normalisation processes the BICs were classified using a neural network, a Bayes classifier, and a voting strategy. The technique is general and might be applied for pre-symptomatic detection, and to other conditions and evoked potentials, although further validation with more subjects, preferably in multi-centre studies is recommended.

Keywords
Alzheimer's disease, biomarkers, P300, auditory evoked potential, independent components analysis, artificial neural network, PSFAM, Bayes classifier

Introduction

It is important to be able to diagnose new onset Alzheimer's Disease (AD) sufferers both for their care and their personal planning. Evoked potential analysis might provide a relatively inexpensive, quick, and non-invasive technique for this and has therefore been investigated. A method of distinguishing with 100% accuracy between early-stage AD patients and normal, healthy subjects (normals) based upon the non-oscillatory, independent components of the P300 peak in the P300 waveform elicited by an auditory oddball paradigm is described in this paper. Because averaging is not used, potentially significant components, unsynchronised to the stimulus, are not reduced, and results may be obtained using fewer trials per subject. This method could be a useful tool to aid diagnosis, and the selected

independent components may be regarded as biomarkers. The method might also be useful for pre-symptomatic testing for Alzheimer's Disease (AD).

Since the review and description of work previous to 2011[1], there have been further publications on the topic in question. In a review of 2011 [2] it was concluded that the sensitivities of a number of ERP components have great promise for the detection of the stages of Alzheimer's disease. Another review in 2014 [3] was focussed on the progression from mild cognitive impairment (MCI) to AD. All the studies quoted followed changes in amplitude and latency of the P300 peak, but on an averaged basis. In reference[4] trial averaging and statistical analysis of the peak ERP amplitudes and latencies derived from a three-stimulus auditory oddball paradigm showed that the P3a and P3b peaks produced the most sensitive and reliable measures of the cognitive deficits associated with early Alzheimer's disease. None of this work [2, 3, 4] addressed the analysis on a single trial basis as described here. However, Ouyang et al [5] have analysed single trials by applying the technique of residue iteration decomposition (RIDE) to identify the latencies of the different ERP peaks in different trials. It seems that this technique, though, does require some averaging of trials to obtain the initial most likely latency of the ERP peaks, after which the individual latencies are found by an iterative method. No application which differentiated between different subject groups was presented. By contrast, in our work we derive the individual components of the individual ERP peaks comprising each individual single trial using independent components analysis (ICA), and apply this knowledge to differentiate between normals and ADs. It seems none of the authors [2, 3, 4, 5] were aware of the previous work by both ourselves and those we quoted[1], although the work of Jung et al using ICA is mentioned in one paper[5]. In another review[6] it was concluded that MCI patients had prolonged P300 latencies compared to controls, but shortened P300 latencies when compared to AD patients meaning that ADs had longer latencies than normals.

Our research was carried out using the data obtained in earlier work, which has been thoroughly described in two previous publications[1, 7]. Thus, only the essentials of that work are repeated here. A selective analysis of that data using an artificial neural network, the Probabilistic Simplified Fuzzy ARTMAP (PSFAM) [8], and a voting strategy is presented.

The undulatory P300 waveform includes a number of positive and negative peaks[1]. By using independent components analysis (ICA) and back-projecting the non-oscillatory, independent, source signals to the scalp electrodes, it was found that the peaks in the P300 waveform consisted of many short duration, randomly occurring, and randomly positive or negative half-sinusoidal pulses[1]. Here, attention is focussed upon the positive back-projected independent components (BICs) centred on the P300 peak, because the shape of the peak is primarily determined by these, and the latency of this peak is delayed in ADs compared to normals[1, 6]. Therefore, these

BICs were deemed the most likely to be useful for differentiating between ADs and normals.

Theoretical aspects

The voltage measured at each electrode depends upon the contributions there from all the independent cortical signal sources. These depend upon the unknown source signals and their unknown transmission paths from the sources to the electrodes. Fortunately, the individual source signals may be computed from the measured scalp voltages using ICA[1], where it was explained that if $\mathbf{S}$ be a matrix of temporally independent source signals and $\mathbf{Y}$ be the matrix of measured signals at the electrodes, which are assumed to consist of linear sums of the source signals ($\boldsymbol{S}$), which have passed through an unknown, linear transmission system characterised by an m x m mixing matrix, $\mathbf{A}$, then we may write

$$\mathbf{Y} = \mathbf{A}\boldsymbol{S} \qquad (1)$$

and

$$\hat{\mathbf{S}} = \mathbf{A}^{-1}\mathbf{Y} \qquad (2)$$

Thus, the estimated source signals ($\hat{\mathbf{S}}$) may be found since $\mathbf{A}^{-1}$ can be found. Selected estimated source signals may then be multiplied by the mixing matrix to obtain their estimated contributions at the measurement electrodes, $\hat{\mathbf{Y}}$**.** These are referred to as the back-projected independent components (BICs). Thus,

$$\hat{\mathbf{Y}} = \mathbf{A}\hat{\boldsymbol{S}} \qquad (3)$$

The BICs are correct in both magnitude and sign, and so may be compared.

The PSFAM, used to classify the data, consisted of a Simplified Fuzzy ARTMAP (SFAM) and a Bayes classifier [8]. The latter produced the Bayes posterior probability $P(\mathbf{A} \mid \mathbf{X})$ that the test vector $\mathbf{X}$ belonged to the class AD or class normal.

Measurements

Six male and three female normal, healthy subjects and two male and seven female newly diagnosed, early stage, mildly cognitively impaired AD subjects participated in auditory-evoked oddball P300 recordings as fully detailed previously [1, 7]. The ADs were under various drug treatments[1], where age effects are also discussed.

Scalp voltages were recorded at 27 standard electrode sites (see below). The voltage waveforms were sampled at 1024 Hz, the high pass cut-off frequency was 0.016Hz and the low pass cut-off was at 60Hz. A notch filter eliminated the mains frequency of 50 Hz. There were 40 target tones of 2 kHz and 160 non-target tones at

1 kHz. The inter-stimulus interval was 1.29 s. The subjects had closed eyes, were relaxed, and responded to the target tones by button-pressing. For each subject 360 target stimuli were recorded, with 600 pre-stimulus samples and 700 post-stimulus samples.

Procedures

The following signal processing was performed as fully detailed before[1, 7].The independent components (ICs) of the P300 waveforms were obtained by applying Principal Components Analysis (PCA) first and then Independent Components Analysis (ICA)[1]. These ICs were then back-projected to the measurement electrodes as the BICs. These were separated into separate bins centred around the P300 peaks. The highest variance BICs were selected for further processing. The BICs in each bin were clustered in two stages using the k-means clustering algorithm[1, 7]. In the primary stage clustering was by amplitude and latency; in the secondary stage by the scalp topographies[1, 7]. Noise components were eliminated by filtering out ICs according to the number of zero crossings in their waveform and their largest and smallest amplitudes[1, 7]. Within each bin the peak amplitudes, latencies, and the scalp topographies of the BICs were saved for analysis[1, 7].

In the previous paper[1], the BIC results obtained at this stage of processing were discussed, and have also been briefly reviewed in the above Introduction. Here we describe that data in detail and how it has been processed further to allow identification of the individuals with newly diagnosed Alzheimer's disease. This data may be requested from the corresponding author.

The data spreadsheet was 39 columns wide and contained 5,302 rows. The data in each row included subject details, subject class (AD or normal), trial number, BIC information (which bin, which cluster, positive or negative, amplitude, latency), and the voltages of the BICs at the twenty seven measurement electrodes used which were Fp1, Fp2, F7, F8, F3, F4, FC5, FC6, FC1, FC2, T7, T8, C3, C4, CP5, CP6, CP1, CP2, P7, P8, P3, P4, O1, O2, Fz, Cz, and Pz. These twenty seven BIC voltages, taken in the above order, comprise the BIC topology vector. This spreadsheet was divided into separate spreadsheets for the AD and normal subjects.

Data preparation for PSFAM

It was intended to use the data in the above two spreadsheets to train classifiers to distinguish between the AD and normal subjects. These 5,302 x 39 data contained some personal details which were irrelevant to this training, and so these columns were ultimately deleted. Since it had been established[1] that those positive BICs associated with the P300 peak, and their latencies, were the most significant in distinguishing the two classes, the amount of data could be considerably reduced by

using only that for positive BICs found close to the P300 peak, i.e. those in bin 5. This reduced the data arrays to 520 and 581 rows for the AD and normal subjects respectively. Since latency was more important than amplitude, the amplitude column was also deleted. The training and test vectors then consisted of the latency vector and the topology vector, making a 1 x 28 row vector per trial.

The PSFAM neural network required input data normalised to between 0 and 1. The latency column was normalised by dividing all values by the largest. Some of the elements of a topology vector could be positive, while others were negative, reflecting the scalp voltage topography. A simple formula was applied to convert the values over the positive and negative voltage range of the topology vector array to values between 0 and 1. A spreadsheet function was used to set all data to numerical values, because the classifiers could not accept exponentials. A sample section of a resulting data array up to column F4 is shown in Table 1. The values are as calculated, but of course were not measured to the accuracy shown. The contents of the first column define the bin and cluster numbers of the BIC in a particular row. Subj is the subject number. Trial is the trial number for that subject. Latencies are in ms, and voltages in mV. Files in training and validation data formats were derived from the two data arrays by deleting columns 1 to 3, and replacing column 4 (trial) by the subject class; 0 for a normal, and 1 for an AD subject. Test vectors may be formed by deleting the first four columns.

Insert Table 1

A training file was constructed consisting of the data for the first five normal subjects and for the first five AD subjects. A validation file was constructed from the last four normal subjects and the last four AD subjects. Thus all the data was used. The structures of these two files were identical, and so their training and validation roles were reversible. These files contained the data obtained from all 40 trials. In clinical practice, where patients may not be sufficiently co-operative, it is necessary to use fewer trials. For this reason we also explored the use of just 10 trials and of just five trials. We had observed that five trials were necessary to ensure that a BIC was found (remember the random nature of the appearance of the BICs). Thus additional training and validation files were produced by eliminating the row data for trial numbers greater than 11 and greater than 6 using functions of the spreadsheet.

PSFAM procedures and considerations

A number of tests were carried out using the training and validation files to determine the optimal values of the PSFAM parameters. These were found to be: vigilance, ρ = 0.65; global smoothing parameter, σ = 0.02; the remaining parameters set to 1. In the training mode the data were presented in random order using the “shuffle” button. The normal subjects were assigned to the reference class 0, and the AD subjects to class 1. Each subject was represented by several row vectors, which

contained the values calculated for different trials and clusters. The classification of an individual row for a given subject could be correct or incorrect. When the overall accuracy per subject was investigated the SFAM was found to result in the higher accuracy for the normal subjects, but the AD subjects were most accurately classified by the Bayes method. The reason for the difference lies in the different classification techniques used by these classifiers. In the SFAM the degree of fuzzy membership of the test input vector $\mathbf{I}$ to the fuzzy power set of the weight vector $\mathbf{W}_\mathrm{J}$ is calculated using the match function $\mathrm{MF}(\mathbf{I}, \mathbf{W}_\mathrm{J})$ [8] where

$$\mathrm{MF}\left(\mathbf{I}, \mathbf{W}_\mathrm{J}\right) = \frac{\mathbf{I} \wedge \mathbf{W}_\mathrm{J}}{\mathbf{I}} \tag{4}$$

and ∧ is the fuzzy logic operator. Thus, each input vector's similarity to each weight vector is determined and the test input vector is assigned to the class of the weight vector for which $\mathrm{MF} > \rho$, i.e. for which it is acceptably close. In the Bayes classifier the summed differences of the test input vector to all the training vectors of the normal class is compared to those of the AD class, according to

$$\sum_{i=1}^{i=n_N} exp\{-(\mathbf{X} - \mathbf{Y}_\mathrm{Ni})^t(\mathbf{X} - \mathbf{Y}_\mathrm{Ni})/2\sigma^2\} \geq \sum_{i=1}^{i=m_{AD}} exp\{-(\mathbf{X} - \mathbf{Y}_\mathrm{ADi})^t(\mathbf{X} - \mathbf{Y}_\mathrm{ADi})/2\sigma^2\} \tag{5}$$

where n, m are the numbers of training vectors for the normal and AD subjects respectively, the indices N, AD refer to the normal and AD classes, $\mathbf{X}$ is the assumed normal input test vector, the $\mathbf{Y}_{Ni}$ and the $\mathbf{Y}_{ADi}$ are the training vectors, σ is the smoothing parameter, and t denotes the vector transpose. If the equality is satisfied, the test input vector is assigned to the class normal. This difference in the classification methods explains why in general the two classifiers do not always assign the same class to a test vector. The discrepancies between the two classifiers can be expected to be greater when the training and test vectors have more random properties as in the case for the BICs of both the normal and the AD subjects obtained in this work. Further, it was also suspected that the ADs' BICs were more random than those for the normal subjects. In fact it was found that the SFAM gave the higher classification accuracy for the normal subjects, while the Bayes classifier gave the higher classification accuracy for the AD subjects. The correct class for a subject was indicated by the classifier which classed the most input vectors for that subject as being of the same class. The classification of a subject was therefore decided by adopting this voting strategy. When the number of test vectors assigned to the two classes was equal, they both indicated the correct class.

Another test was undertaken to establish whether fewer, carefully chosen electrodes might be used. Thus classification using only columns Lat, Fp1, Fp^2, P3, P4, Fz, and Pz was attempted, but the results were worse than when all 28 columns were used. In another test, the training and validation files were exchanged, when equally good results were obtained.

PSFAM results

Forty trials

Table 2 gives the overall percentage classification accuracies for training and testing the PSFAM with the full set of data from 40 trials when the training and validation files were prepared as described above. It is seen that the higher percentage correct classification of normal subjects is achieved by the SFAM (77%:45%), and that of the AD subjects by the Bayes classifier (62%:40%).

Insert Table 2

The percentage classification accuracies of the individual subjects, by their individual input test vectors, are shown in Table 3. It is seen that, if the percentage correct in the SFAM column is greater than that in the Bayes column, the subject is normal; if the converse is true, the subject is in the AD class. This is the basis of the voting strategy, which yields 100% correct classification of the newly diagnosed AD subjects in this study from the positive BICs at the 27 EEG electrodes centred on the P300 peak. Of course, slightly different numerical values are obtained when the training is repeated, owing to the shuffling of the training data, but the conclusions remain unaltered.

Insert Table

Ten trials

Similar results were obtained as for the case of the 40 trial data when only the first 10 trials were used. Table 4 gives the subject classification accuracies. The same conclusions apply as for 40 trials, with the addition that in two cases, N14 and AD40, the percentage accuracies are the same for both classifiers, but they both predict the same correct class. We also see that the accuracy of classification of the normal subjects by the SFAM, and that of the AD subjects by the Bayes classifier has increased with the reduced number of trials. Some of this may be attributed to the shuffling of the input vectors, but it seems more likely it could be owing to a reduction in the degree of randomness associated with using fewer vectors.

Insert Table 4

Table 5 presents a sample of the numbers of 0s and 1s output by the classifiers to represent the classes of the input vectors for the subjects, normal and AD respectively, and the classification voted, which is correct in each case.

Insert Table 5

Five trials

Similar results were obtained again in the case of 5 trials as shown in Table 6. Apart from subject N15, the classification accuracies were again improved by this further reduction in the number of trials included. However, the use of fewer trials still is likely to lead to more classification failures when there may be too few input vectors for reliable testing or even the absence of any vector.

Insert Table 6

Discussion

It has been clearly demonstrated in this research that the positive voltage BICs associated with the P300 peak may constitute an excellent biomarker for new onset AD, since 100% accurate differentiation between new onset ADs and normals was achieved. It is quite possible that they could indicate AD pre-symptomatically. This could be tested by making measurements on subjects at-risk of AD, such as carriers of the apolipoprotein E4 gene with a family history of AD, or subjects for whom synaptic dysfunction has been detected by elevated CSF phosphor-tau[2]. The technique is non-invasive, requires a reduced number of trials, is inexpensive, and can be employed in any hospital EEG department. Because of the small sample size it is desirable that far more subjects be tested, and preferably in multi-centre studies, to validate it, if it be considered useful. The digitised multi-centre recordings could be processed centrally to ensure conformity. This validation could take place during the clinical studies. Such studies might also be used to investigate the effects of drug treatment, the relationship of the results to those of similar studies on subjects with other neurological diseases to reduce the risk of misdiagnosis, and the possible usefulness of the BICs associated with other peaks in the waveform. Extension to other conditions such as Parkinson's disease and other evoked potentials is also a possibility.

Conclusions

New-onset ADs were differentiated from normals to 100% accuracy by classifying the positive voltage, back-projected, independent components centred on the P300 waveform peak response to an oddball task, auditory evoked potential using a simplified fuzzy ARTMAP neural network, a Bayes classifier and a voting strategy. It may also be possible to detect AD pre-symptomatically, but more, preferably multi-centre research on more subjects is necessary to validate the technique. Extension to other conditions and evoked potentials is a possibility.

References

1. Jervis BW, Belal S, Cassar T, Besleaga M, Bigan C, Michalopoulos K. et al. Waveform Analysis of Non-Oscillatory Independent Components in Single-Trial Auditory Event-Related Activity in Healthy Subjects and Alzheimer's Disease Patients. Current Alzheimer Research 2010;7:2:1-14.
DOI: https://doi.org/10.2174/156720510791162412

2. Olichney JM, Yang J-C, Taylor J, and Kutas M. Cognitive Event-Related Potentials: Biomarkers of Synaptic Dysfunction Across the Stages of Alzheimer's Disease. J Alzheimers Dis. 2011;26;s3:215–228.
DOI: 10.3233/JAD-2011-0047

3. Cintra MTG, Tavares MCS, Gomes SA, Gonçalves TO, da Cunha LCM, Gonçalves DU, et al. P300 evoked potential and risk of mild cognitive impairment progression to AD: a literature review. J Neurol Neurophysiol. 2014;6:5.
http://dx.doi.org/10.4172/2155-9562.1000322

4. Cecchi M, Moore DK, Sadowsky CH, Solomon PR, Doraiswamy PM, Smith, CD et al. A Clinical Trial to Validate Event-Related Potential Markers of Alzheimer's Disease in Outpatient Settings. Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring. 2015;1:4:387–394.
https://doi.org/10.1016/j.dadm.2015.08.004

5 Ouyang G, Sommer W, Zhou C. Reconstructing ERP amplitude effects after compensating for trial-to-trial latency jitter: A solution based on a novel application of

residue iteration decomposition. International Journal of Psychophysiology. 2016; 109:9–20.
https://doi.org/10.1016/j.ijpsycho.2016.09.015

6. Vecchio F and Määttä S, The Use of Auditory Event-Related Potentials in Alzheimer's Disease Diagnosis. International Journal of Alzheimer’s Disease. 2011; 2011:Article ID 653173, 7 pages
http://dx.doi.org/10.4061/2011/653173

7. Jervis BW, Belal S, Camilleri K, Cassar T, Bigan C, Linden DEJ, et al. The independent components of auditory P300 and CNV evoked potentials derived from single trial recordings. Physiol. Meas. 2007;28:745-771.
DOI https://doi.org/10.1088/0967-3334/28/8/001

8. Jervis BW, Garcia T, Giahnakis, EP. The Probabilistic Simplified Fuzzy ARTMAP (PSFAM). IEE Proc. Science, Measurement and Technology. 1999;146:165-169.
DOI: 10.1049/ip-smt:19990383

Tables

Bin and Cluster	Sex	Subj	trial	latency	Fp1	Fp2	F7	F8	F3	F4	
B5C7	M	5	5	0.98242	0.50569	0.50308	0.50855	0.49351	0.50064	0.51222	
B5C7	M	5	9	0.945306	0.48338	0.48203	0.50514	0.50355	0.50223	0.49200	
B5C4	F	14	6	0.790989	0.52205	0.51503	0.50927	0.51248	0.51003	0.51407	
B5C5	F	14	3	0.84373	0.49514	0.50459	0.49695	0.48690	0.49842	0.50018	
B5C5	F	14	6	0.83201	0.49751	0.49771	0.49366	0.50117	0.49390	0.49727	
B5C5	F	14	8	0.83787	0.51013	0.51636	0.51359	0.52257	0.51380	0.52197	
B5C7	F	14	10	0.958979	0.50815	0.50307	0.50486	0.49801	0.51677	0.50936	
B5C2	F	15	8	0.61128	0.50710	0.51153	0.51494	0.50784	0.50121	0.50973	
B5C2	F	15	9	0.595653	0.51948	0.51007	0.52694	0.48678	0.52015	0.51536	
B5C4	F	15	3	0.800756	0.51126	0.51315	0.50458	0.50959	0.51277	0.51684	
B5C4	F	15	6	0.763642	0.48685	0.48498	0.48810	0.49188	0.49468	0.49973	

Table 1 Sample section of data array

SFAM		**Bayes**	
Normal subject	AD subject	Normal subject	AD subject
77%	40%	45%	62%

Table 2 Overall percentage classification accuracies by the two classifiers for normal and AD subjects.

Subject	**SFAM % correct**	**Bayes % correct**
N14	61	53
N15	69	41
N16	75	39
N20062	84	41
AD36	52	90
AD38	39	86
AD40	34	82
AD43	43	77

Table 3 Percentage classification accuracies for individual test subjects based on their test vectors using 40 trials.

	SFAM % correct	Bayes % correct
N14	80	80
N15	100	44
N16	67	50
N20062	81	57
AD36	54	100
AD38	50	88
AD40	53	53
AD43	50	57

Table 4 Percentage classification accuracies for individual test subjects based on their test vectors using 10 trials.

		SFAM		Bayes		
Subject	**Number of vectors**	**Number of 0s**	**Number of 1s**	**Number of 0s**	**Number of 1s**	**Class by voting**
N14	5	4	1	4	1	N
AD40	19	9	10	9	10	AD
AD43	14	7	7	6	8	AD
N20062	16	13	3	9	7	N

Table 5 Numbers of vectors per subject classified as 0 or 1 and the class of the subject by voting.

Subject	SFAM % correct	Bayes % correct
N14	100	100
N15	50	25
N16	86	57
N20062	89	64
AD36	57	100
AD38	71	100
AD40	30	60
AD43	80	80

Table 6 Percentage classification accuracies for individual test subjects based on their test vectors using 5 trials.