

Investigation of Alzheimer's Disease EEG Frequency Components with Lempel-Ziv Complexity

Samantha Simons, Daniel Abasolo, and Michael Hughes

Centre for Biomedical Engineering, Department of Mechanical Engineering Sciences,
Faculty of Engineering and Physical Sciences, University of Surrey, Guildford, UK

Abstract— This pilot study applied Lempel-Ziv Complexity (LZC) to 22 resting EEG signals, collected using the 10-20 international system, from 11 patients with Alzheimer's disease (AD) and 11 age-matched controls. This allowed for frequency band analysis as the EEG signals were first pre-filtered with a third order Hamming window in the ranges F to F+WHz with both F and W equal to 1-30Hz respectively. Control subjects were found to have a greater signal complexity than AD patients with statistically significant bands seen at various ranges in all 16 electrodes. The maximum statistical significance (Student's t test, $p < 0.01$) was increased over the findings with traditional signal filtering techniques allowing the whole range, with a maximum significance of $p = 3.50 \times 10^{-6}$ at electrode T4 between 7-18Hz. Electrode F4 also showed significantly high statistically significant differences. The maximum accuracy, both controls and AD patients correctly identified, found with Receiver Operating Characteristic Curves was 95.45% (21 of 22 subjects correctly classified) at T4 (7-18Hz and 7-20Hz), Fp2 (8-32Hz) and F4 (6-21Hz), which is significantly more accurate than the most accurate methods previously applied to this dataset. The beta band (13-30Hz) was found to be most influential in separating the two test groups in this study with the best range suggested to be 5-26Hz, combining traditional theta, alpha and beta bands. These findings suggest pre-filtering has a significant effect on method outcomes and can be successfully tailored to improve the statistical effectiveness of LZC at distinguishing between these two EEG groups. However, more testing is required to investigate the effectiveness at distinguishing other signal dynamics.

Keywords— Alzheimer's disease, Electroencephalogram, Non-linear analysis, Lempel-ziv complexity, Frequency component analysis.

I. INTRODUCTION

Alzheimer's disease (AD) is a prevalent, degenerative neurological disorder of which, currently, little is known about the development of the disease.

With the increasing multitude of linear and non-linear signal processing techniques much research has been recently completed into the difference in electroencephalogram (EEG) signals from patients suffering from AD and controls [1]. AD signals have been found to contain a lower percentage of alpha and beta rhythms (8-12Hz and 13-30Hz respectively) and a higher percentage of delta and theta rhythms (0-3Hz and 4-7Hz respectively) [2], an increase in regularity and a decrease in complexity [3].

These studies often use the whole signal, filtered to ensure DC components and environmental noise is removed. In this paper, the change in frequency components of the EEG is investigated using the non-linear method of Lempel-Ziv complexity (LZC) [4]. LZC has been chosen for this pilot study due to its rigorous scientific basis [4] and previously published success in differentiating EEG AD and control signals [5]. Furthermore, the reduction in signal frequency range has shown an increase in statistically significant changes in a paper by Gallego-Jutgla *et al.* [6].

It is hypothesized from the large body of work in AD EEG signal analysis and LZC that controls will show a more complex signal than those from AD patients. Furthermore, it is suggested that the decreasing frequency ranges under analysis will identify greater statistical significance between the two groups.

The paper is arranged as follows. In the second section the methods are introduced including the signal database investigated in this study. Results are detailed in the third section and discussed in the fourth section. References are presented at the end of the paper after discussions and conclusions.

II. METHODS

A. EEG Signal Database

This database has been used in a number of studies, e.g. [5]. The subject group consisted of 22 subjects, 11 probable AD patients (5 males, 72.5 ± 8.3 yrs, mean \pm standard deviation (SD)) and 11 age-matched controls (7 males, 72.8 ± 6.1 yrs, mean \pm SD). The two groups had a Mini-Mental State Examination score of 13.1 ± 5.9 (mean \pm SD), indicating moderate to severe dementia, and 30 ± 0 (mean \pm SD), indicating no traits of dementia, respectively and had undergone clinical examination including clinical history, brain scans and physical and neurological exam.

Signals were recorded using Profile study room EEG equipment 2.3.411 (Oxford Instruments). The signal was recorded at 256Hz with a 12 bit analog-to-digital conversion using the international 10-20 electrode system of 19 channels with electrodes referenced to the chin. This resulted in signals from 16 electrode positions for analysis, Fp1, Fp2, F3, F4, C3, C4, P3, P4, O1, O2, F7, F8, T3, T4, T5 and T6. Over five minutes of recordings were taken with the

subjects in a resting but awake state with closed eyes. A clinician then isolated 5s epochs (1280 data points) clear of movement and electrooculographic artifacts and showing no signs of sleep. These signals, of which 30.0 ± 12.5 (mean \pm SD) epochs per subject per electrode were collected, were then saved offline for further analysis as ASCII files. Two dementia patients were receiving treatment including lorazepam but no prominent change in the beta activity of the EEG was observed.

Signal collection and investigation was approved by local ethical committees. Informed consent was collected from the control group and the caregivers of demented patients.

B. Signal Filtering

Previous studies with this database have used a Hamming window finite impulse response band-pass filter between 0.5-40Hz. As the interest in this study is the investigation into the key frequency changes in the AD signals in comparison to the control signals, this has not been carried out as it may unduly influence the findings.

Instead, the filtering carried out was as follows. A third order Butterworth filter, chosen due to its clear transition properties [7] was used to filter the signal in the range F to F+W Hz with F in the range 1,2,...,30Hz and W in the range 1,2,...,30Hz. This method has already been applied to another signal database [6]. This and all other signal manipulation, calculation and result analysis presented in this study was carried out using Matlab® [8].

C. Lempel-Ziv Complexity

LZC is a univariate, non-linear method that measures the complexity of that signal [4]. It has been found to be reliable with short data sets [5]. This method has been widely applied to medical studies including investigations into AD [5], Parkinson's disease [9] and coma [10] along with many others. In order to apply LZC the signal must first be parsed.

Parsing converts the signal into a pattern of reduced symbols given the location of each value in the dataset to a set of thresholds. In this study, the signal was parsed into a binary sequence given the location of the data values to the threshold (T_d) of the mean, chosen as it is robust to outliers [11]. Given a signal $X = x(1), x(2), \dots, x(n)$, the converted sequence $H = h(1), h(2), \dots, h(n)$ with elements $h(i)$ defined by [12]:

$$h(i) = \begin{cases} 0 & \text{if } x(i) < T_d \\ 1 & \text{if } x(i) \geq T_d \end{cases} \quad (1)$$

This new sequence H is then scanned from left to right to identify the different subsequences held within the sequence. The LZC, $c(n)$, is the number of these unique subsequences, identified as follows [4]:

1. S and Q are two subsequences of H and SQ the concatenation. The last character is removed is denoted $SQ\pi$ and $v(SQ\pi)$ the vocabulary of all subsequences of $SQ\pi$. Initially $c(n) = 1$, $S = s(1)$ and $Q = s(2)$, so $SQ\pi = s(1)$
2. Generally $S = s(1), s(2), \dots, s(r)$ and $Q = s(r+1)$, so $SQ\pi = s(1), s(2), \dots, s(r)$. If Q belongs to $v(SQ\pi)$ then it is not a new sequence but a subsequence of $SQ\pi$
3. With $Q = s(r+1), s(r+2)$ again judge if Q does or does not belong to $v(SQ\pi)$
4. Repeat until Q is a new sequence, then $c(n) = c(n)+1$. In this case $Q = s(r+1), s(r+2), \dots, s(r+i)$ is not a subsequence of $SQ\pi = s(1), s(2), \dots, s(r+i-1)$
5. Thereafter, $S = s(1), s(2), \dots, s(r+i)$ and $Q = s(r+i+1)$. This is repeated until Q is the last character of H

For example, $H = 110101101010$ parses giving 1.10.1011.01010 i.e. LZC = 4. This method, therefore, gives a higher LZC to a more complex signal. A less complex signal would be denoted by a lower LZC.

D. Statistical Analysis

Results were initially tested for normal distribution to allow for the identification of the most relevant statistical analysis methods. If a normal distribution was established, then Student's t test ($p < 0.01$) was chosen to evaluate statistically significant differences between patients and controls. If the data was found not to follow a normal distribution then a Kruskal Wallis test ($p < 0.01$) was to be used over a Student's t test. Statistically significant pairs were further investigated using Receiver Operating Characteristic (ROC) Curves [13]. ROC curves are a plot of the sensitivity (ability to identify AD patients correctly) against 1-specificity (where specificity is the ability to correctly identify controls). These can then be combined to find the accuracy (correctly identified patients and controls) and the area under the ROC curve (AROC), which shows the data distribution and some indication of the separation of the two data groups. A higher AROC describes a more separated distribution than a lower AROC distribution.

III. RESULTS

EEG recordings from AD patients were found to have a lower LZC than those from controls. Data was found to predominantly follow a normal distribution (96.76% with a Lillifors test with a null hypothesis threshold of 0.01) and so Student's t test was used.

Signals starting from low frequencies (F from 1-12Hz) identified most of the statistically significant differences between the EEG signals from AD patient's EEGs and those from controls. Small bandwidth signals also struggled to differentiate statistically between the two groups. Widths of W=1-4Hz showed very few statistically

significant differences. The most successfully distinguishing bands were produced with $F=4-6\text{Hz}$ and $W=17-30\text{Hz}$. The most successful band was that of $5-26\text{Hz}$.

The most distinguishing frequency band, with $p=3.50e^{-6}$, was frequency band $7-18\text{Hz}$ at electrode T4. Further statistically significant bands include electrode T4 in frequency bands $7-16\text{Hz}$, $7-17\text{Hz}$ and $7-19\text{Hz}$ ($9.07e^{-5}$, $6.78e^{-5}$ and $1.35e^{-5}$ respectively) and electrode F4 in frequency bands $5-29\text{Hz}$, $6-27\text{Hz}$, $6-29\text{Hz}$, $6-30\text{Hz}$ and $6-32\text{Hz}$ ($9.04e^{-5}$, $8.34e^{-5}$, $6.59e^{-5}$, $3.54e^{-5}$ and $5.96e^{-5}$ respectively). In all, 9.89% of tested frequency bands showed statistically significant differences between the EEG signals of AD patients and control subjects. All electrodes showed statistically significant differences for a number of frequency ranges.

The maximum accuracy of 95.45% (21 subjects correctly identified) was seen at a number of electrodes, identified in table 1 below.

Table 1 Statistical and ROC analysis results for the four instances of 95.45% accuracy

Elec.	Freq. band	p	Sensitivity	Specificity	Accuracy	AROC
T4	7-18Hz	$3.50e^{-6}$	90.91%	100.0%	95.45%	0.9587
T4	7-20Hz	$2.54e^{-4}$	90.91%	100.0%	95.45%	0.9421
Fp2	8-32Hz	$1.24e^{-4}$	90.91%	100.0%	95.45%	0.9669
F4	6-31Hz	$5.58e^{-5}$	100.0%	90.91%	95.45%	0.9669

Also seen in the table are a number of 100.0% sensitivity and specificity tests, showing signal frequency bands from differing electrodes which are able to completely differentiate AD patients from controls, given the hypothesis stated in section 1. Further, an accuracy of 90.91% (20 subjects correctly identified) was seen in 8.50% of statistically significant EEG comparisons.

IV. DISCUSSION AND CONCLUSIONS

In this pilot study we have looked at the changes in specific ranges of frequencies within an EEG signal in comparison to the usual application of non-linear processing to the whole signal, to identify key changes in the resting EEGs of AD patients in comparison to age-matched controls. This has been facilitated by reapplying LZC to a dataset of 22 subjects previously investigated over a more traditional frequency range [5]. Controls were found to have a more complex signal than that of the AD patients, many with clear statistically significant differences, supporting the body of findings with this and other methods with AD EEG recordings which have been published.

When looking at the optimal range of frequencies in comparison to spectral findings, it is noted that the optimal band of $5-26\text{Hz}$ covers theta, alpha and beta signal ranges.

One possibility of this technique was to identify those spectral changes with a non-linear technique, thus bringing together the findings from the two, often separately investigated, methods. However, findings show that the most influential range in this test those that fell within the beta range of signal frequencies ($13-30\text{Hz}$) with delta ranges ($0-3\text{Hz}$) contributing no statistically significant changes.

Beyond the 40Hz range, there was only one incidence of statistical significance. As this is beyond the normal frequency range seen in EEG signals when investigating background activity and is instead associated with contamination frequencies, this identifies the lack of useful information within these recordings.

In comparison to the previously published findings with LZC on this database statistically significant differences, sensitivities, specificities, accuracies and AROC values at all electrodes were significantly increased in comparison to both the two symbol (binary) and three symbol conversion methods applied previously to these signals, filtered only to reduce background electrical interference (Hamming window between $0.5-40\text{Hz}$) [5]. Electrodes found to be significant in this original study were P3 and O1 with binary conversion and T5, P3, P4 and O1 with three symbol conversion. The maximum accuracy was 81.82% (18 correctly classified subjects) with a maximum sensitivity/specificity pair of 90.91%/72.73% or 72.73%/90.91% respectively. Therefore the method of sectioning the signal into smaller frequency bands allows for a greater resolution of the LZC method in some frequency regions. However, the electrodes with the most statistically significant differences in this study, T4 and F4 did not show any differences in the original study.

Previously to this study, the most accurate method applied to this pilot database was that of the slope of a multiscale entropy (MSE) graph for grains of $\tau > 6$ [14] and auto mutual information (AMI) [15]. Both, at electrode Fp1 for MSE and P3 for AMI respectively, gave an accuracy of 90.91%. Both papers also gave a maximum AROC of 0.9339 at these electrodes. The most statistically significant findings in the study detailed in this paper also significantly increase the identified differences in the EEG signals of the two groups over these two, previously most accurate, methods.

Another study where this signal filtering was applied was that by Gallego-Jutgla *et al* [6]. Their findings too showed the improvement in distinction that this filtering method could obtain with a range of linear methods including synchrony and coherence methods. Statistical significances were within the range seen in this study (e.g. $2.6e^{-6}$) in their optimized frequency band of $5-6\text{Hz}$. This is a significantly smaller frequency and than that identified as the most statistically distinguishing in this study; indeed this range in this

study produced no statistically significant differences between the two groups. This difference in findings is probably due to the nature of the LZC used in this study in comparison to the linear methods applied by Gallego-Jutgla *et al.* [6], where more resolution may be gained from the greater variety of possible data points and patterns.

While this method has supported the findings of Gallego-Jutgla *et al.* [6], there are some notable differences. Given the small sample size of the data set tested here, further investigation using other ranges of frequencies, other methods and greater data groups will give more insight into the key ranges of frequencies which show changes and the ability to combine both spectral and non-linear research techniques to increase our understanding of the disease. Furthermore, with similar findings from those with AD coming from pathologies such as Parkinson's disease [9] and the precursor to AD, Mild Cognitive Impairment (MCI) [16], other pathologies must also be investigated to determine if reducing the frequency may identify differences specific to changes in the EEG due to AD.

In conclusion, this pilot study has identified increased complexity in control subject's EEG recordings in comparison to those of AD patients when the signal is split into smaller frequency ranges than usually tested. This method shows a significant increase in the statistical significance of those differences between the two groups over traditionally filtered signals in a wide range of frequency bands, with the optimum band being that of 5-26Hz. However, caution should be taken due to the limited size and scope of the study group.

V. CONFLICT OF INTEREST

The authors declare that they have no conflict of interest.

REFERENCES

1. Dauwels J, Vialatte F, Cichocki A (2010) Diagnosis of Alzheimer's disease from EEG signals: where are we standing? *Curr Alzheimer Res* 7:487-505
2. Jeong J (2002) Non-linear dynamics of EEG in Alzheimer's disease. *Drug Develop Res* 56:57-66
3. Jeong J (2004) EEG dynamics in patients with Alzheimer's disease. *Clin Neurophys* 115:1490-1505
4. Lempel A, Ziv J (1976) Complexity of finite sequences. *IEEE T Inform Theory* 22:75-81
5. Abasolo D, Hornero R, Gomez et al. (2006) Analysis of EEG background activity in Alzheimer's disease patients with Lempel-Ziv complexity and Central Tendency Measure. *Med Eng Phys* 28:315-322
6. Gallego-Jutgla E, Elgendi M, Vialatte F et al. (2012) Diagnosis of Alzheimer's disease from EEG by means of synchrony measures in optimized frequency bands. *IEEE Ann Int Conf EMBC, San Diego, California, 2012*, pp 4266-4270
7. Oppenheim A, Shafer R (1989) *Discrete time signal processing*, NJ, USA:Prentice Hall
8. Matlab 2013, The MathWorks Inc., Natick, MA, 2000
9. Gomez C, Olde Dubbelink K, Stam C et al. (2011) Complexity analysis of resting state MEG activity in early-stage Parkinson's disease patients. *Ann Biomed Eng* 39:2935-2944
10. Wu D, Ca G, Yuan Y et al. (2011) application of non-linear dynamics analysis in assessing unconsciousness: A preliminary study. *Clin Neurophysiol* 122:490-498
11. Nagarajan R (2002) Quantifying physiological data with Lempel-Ziv complexity- Certain issues. *IEEE Trans Biomed Eng* 49:1371-1373
12. Zhang X, Roy R, Jensen E (2001) EEG complexity as a measure of depth of anesthesia for patients. *IEEE Trans Biomed Eng* 48:1424-1433
13. Fawcett T (2006) An introduction to RC analysis. *Pattern Recogn Lett* 27:861-874
14. Escudero J, Abasolo D, Hornero R et al. (2006) Analysis of electroencephalograms in Alzheimer's disease patients with multiscale entropy. *Physiol Meas* 27:1091-1106
15. Abasolo D, Escudero J, Hornero R et al. (2008) Approximate entropy and auto mutual information analysis of the electroencephalogram in Alzheimer's disease patients. *Med Biol Eng Comput* 46:1019-1028
16. Fernandez A, Hornero R, Gomez C et al. (2010) Complexity analysis of spontaneous brain activity in Alzheimer's disease and mild cognitive impairment: An MEG study. *Alzheimer Dis Assoc Discord* 24:182-189

Author: S Simons
 Institute: University of Surrey
 City: Guildford
 Country: UK
 Email: s.simons@surrey.ac.uk