Volume Conduction Effects on Bivariate Lempel-Ziv Complexity of Alzheimer’s Disease Electroencephalograms

Samantha Simons, Daniel Abásolo, Member, IEEE, and Paul Sauseng

Abstract—The spurious increase in coherence of electroencephalogram (EEG) signals between distant electrode points has long been understood to be due to volume conduction of the EEG signal. Reducing the volume conduction components of EEG recordings in pre-processing attenuates this. However, the effect of volume conduction on non-linear signal processing of EEG signals is yet to be fully described. This pilot study aimed to investigate the impact of volume conduction on results calculated with a distance based, bivariate form of Lempel-Ziv Complexity (dLZC) by analyzing EEG signals from Alzheimer’s disease (AD) patients and healthy age-matched controls with and without pre-processing with Current Source Density (CSD) transformation. Spurious statistically significant differences between AD patients and control’s EEG signals seen without CSD pre-processing were not seen with CSD volume conduction mitigation. There was, however, overlap in the region of electrodes which were seen to hold this statistically significant information. These results show that, while previously published findings are still valid, volume conduction mitigation is required to ensure non-linear signal processing methods identify changes in signals only due to the purely local signal alone.

I. INTRODUCTION

Alzheimer’s disease (AD) is the most common form of dementia [1], with a long period of progressive memory loss prior to death [2]. As the only current accurate method of diagnosis is necropsy [1] and increasing numbers of effected patients worldwide, the need to improve diagnostic procedures for AD is acute. To this end, clinical guidelines have recently been updated to recognize the importance of early and accurate diagnosis [2].

While updated guidelines have identified the diagnostic need for AD diagnosis, they have not identified how this may be achieved. Electroencephalogram (EEG) recordings have previously shown the ability to identify changing frequency spectrums with AD patients to lower frequency components in comparison to healthy controls [3]. The use of EEG is also supported both by the portability and cost of the equipment over other signal collection methods and non-invasive aspects over other medical procedures.

The EEG is a complex signal, arising from the interactions of electric fields caused by brain cells [4]. Given this complex, aperiodic, time series, non-linear techniques are appropriate for use in analyzing these signals and may provide complementary information to linear methods on the interregional communications of the brain and the effects of AD on those communications [5]. However, care must be taken that the signal processing method’s assumptions match the signal it is used to analyze. Given the short, noisy, non-stationary properties of an EEG signal, many traditional signal-processing techniques, therefore, are unsuitable [6].

Entropy, as defined by Shannon [7], identifies the amount of information a signal carries, with a higher entropy signal carrying more information than a lower entropy signal. Shannon’s entropy, however, is unsuitable for EEG analysis due to the need for large datasets [7]. Nevertheless it is still widely used in EEG research. Shannon’s entropy should be replaced by Lempel-Ziv complexity (LZC) for short, noisy, non-stationary datasets [8]. Recently, the method of LZC has been further expanded to enable it to be applied to bivariate signals [9].

Volume conduction is the act of a signal conducting away from its point of origin to another site, contaminating the signal from that site. This issue of the inverse problem is a significant issue with EEG recordings, i.e. it is not clear from the signals reaching the surface electrodes where those signals originated. When there is synchronous activity from neurons in a sulcus a dipole tangent to the surface of the scalp will result. Due to volume conduction in the brain, the electrical field generated by this source will then be detectable at relatively distant EEG sensors. However this will be at the opposite polarity [10]. When linear measures are applied to signals, which have not been corrected for volume conduction effects, spuriously high coherence and synchrony are seen between distant EEG sensors. It is crucial, with the increasing use of bivariate and multivariate non-linear signal processing techniques, to investigate whether these method types, e.g. LZC, are similarly vulnerable to the effects of volume conduction.

The aim of this pilot study was to investigate the impact of volume conduction on a bivariate LZC method when utilized to distinguish between control subjects and AD patients through their EEG’s. It is hypothesized that applying a correction for volume conduction on the EEG signals tested will significantly decrease distance measures, indicating significantly decreased signal similarities, based on bivariate LZC.

S. Simons is with the Centre for Biomedical Engineering, Faculty of Engineering and Physical Sciences, University of Surrey, Guildford, GU2 7XH, England (phone: +44 (0)1483 300800; fax: +44 (0)1483 300800; e-mail: s.simons@surrey.ac.uk).

D. Abásolo is with the Centre for Biomedical Engineering, Faculty of Engineering and Physical Sciences, University of Surrey, Guildford, GU2 7XH, England.

P. Sauseng was with the Department of Psychology, University of Surrey, Guildford, GU2 7XH, England. He is now with the Department of Phycology, Ludwig-Maximilians-University Munich, Leopoldstr.13, 80802 Munich, Germany.
II. SUBJECT DATABASE AND METHODS

A. Subject Database

The subject database consisted of 22 subjects, 11 AD patients and 11 age-matched control participants. The AD patients (6f/5m, 72.5 ± 8.3 years, mean ± standard deviation (SD)) achieved a Mini-Mental State Examination (MMSE) [11] of 13.3 ± 5.9 (mean ± SD), indicating moderate to severe changes to cognitive function. These subjects were recruited from the Alzheimer’s Patients’ Relatives Association of Valladolid. The control participants (4f/7m, 72.8 ± 6.1 years, mean ± SD) were screened for past or present neurological disorders. They obtained an MMSE score of 30 ± 0 (mean ± SD), indicating no change to cognitive function. Informed consent was obtained for all subjects. The local ethics committee approved this study.

B. Electroencephalogram Collection

EEG was recorded using a Profile Study Room 2.3.411 EEG Equipment (Oxford Instruments) using the international 10-20 electrode placement system with a sampling frequency of 256Hz and 12-bit A-to-D precision with subjects in a resting, eyes closed, but awake state. This resulted in over 5 minutes of data for each subject from 16 EEG channels (F3, F4, F7, F8, Fp1, Fp2, T3, T4, T5, T6, C3, C4, P3, P4, O1 and O2).

5s-epochs (1280 points) containing minimal contamination by artifacts were selected by a specialist clinician. These epochs were then saved offline for further analysis. 30 ±12.4 (mean ± SD) epochs from each electrode per subject were collected. The signal length of 5s was chosen to maximise the length of the signal to be analyzed while ensuring the effect of artefacts on findings were minimised. These signals were then pre-filtered using a Hamming band-pass filter between 0.5 and 40Hz to remove both DC components and any residual high frequency noise.

C. Volume Conduction Mitigation

Volume conduction arises through passive conduction of electrical signals through biological tissue from an underlying source [12]. By identifying this relationship, the recorded electric potentials can be mathematically modified to remove these volume conduction components. In this study, Current Source Density (CSD), the spherical Laplace operator [13] was applied to the dataset using VisionAnalyzer 2.0 software (BrainProducts®, Gilching, Germany) to remove volume conduction effects. Values chosen for CSD calculation were order of splines = 4, maximum degree of Legendre polynomials = 10 and λ = 1e-5.

The CSD algorithm spline interpolates any location on the scalp by subtracting a weighted mean of the signals recorded from all other electrode sites. Data from the scalp is mapped to a sphere and interpolated. This interpolation is then re-evaluated for potential at each point of interest, each of the original electrode locations. Thus recordings of the whole head are needed for each time point analyzed to create a reliable data interpolation. Depending on the distance from the estimated location to the point of signal detection contributes more or less to this mean signal with a nearer point of signal origin being weighted more heavily than one further away. The mathematical equations for this three-step process can be found in [13].

D. Bivariate Lempel-Ziv Complexity

A distance measure satisfies three criteria, of identity ($D(S,Q) = 0$ when $S = Q$), symmetry ($D(S,Q) = D(Q,S)$), and triangle inequality ($D(S,Q) \leq D(S,T) + D(T,Q)$). A range of distance-based bivariate LZC measures (dLZC) have been defined and tested by Otu and Sayood. These are [9]:

$$D_1(R,T) = \max[c(RT) - c(R), c(TR) - c(T)]$$

$$D_3(R,T) = \frac{\max[c(RT) - c(R), c(TR) - c(T)]}{\max[c(R), c(T)]}$$

$$D_4(R,T) = c(RT) - c(R) + c(TR) - c(T)$$

$$D_{alt}(R,T) = \frac{c(RT) + c(R) + c(TR) - c(T)}{\sqrt{2[c(RT) - c(R) + c(TR) - c(T)]}}$$

Where $c(X)$ and $c(XY)$ are the complexities of sequence $X$ and the concatenation of the two sequences, $X$ then $Y$, respectively, calculated with LZC. For the method for calculating LZC, please refer to [8]. Thus the distance-based bivariate LZC of signal pairs with few subsequences in common would be higher than in signal pairs with a large number of subsequences in common.

All possible electrode pairs were tested with each method, using the median as a threshold for two symbol LZC parsing, each pair of electrodes being separately investigated for their ability to distinguish between the two subject groups.

E. Statistical Analysis

The mean and SD for the 11 AD subjects and the 11 controls were initially calculated. Lilliefors test was also used to check for normal distribution of the results (p<0.01). If the results were found to follow a normal distribution then Student’s t test was chosen to evaluate statistically significant differences between patients and controls with a statistical significance threshold chosen to be p<0.01 [14].

Statistical analysis was also completed in relation to the ability of each electrode to be used as a testing electrode in relation to all others using two-way ANOVA with factors of diagnosis, either AD or healthy control, with two sets of 11 repeats and electrode pairings of 120 different electrode pairs. This was also given the statistical threshold of p<0.01. In order to acknowledge the large numbers of test carried out in this pilot study a multiple comparison test was also calculated as a correction method.

III. RESULTS

Equation (4) was found not to follow the criteria for distance measures as it did not satisfy the criterion of symmetry ($D(S,Q) = D(Q,S)$). Thus this method was removed from all further data analysis.
AD patients showed reduced dLZC than controls. The number of electrode pairs where AD patients were found to have a higher dLZC than controls changed with the distance measure applied. Equations (2) and (5), described as normalized metrics of (1) and (3) respectively, significantly differ from their non-normalized methods. Significantly, the numbers of electrode pairs where AD patients were found to have a higher dLZC than controls changed with the distance measure applied. Equations (2) and (5), described as normalized metrics of (1) and (3) respectively, significantly differ from their non-normalized methods. Significantly, the numbers of electrode pairs where AD patients showed increased dLZC in comparison to controls for the two normalized methods are 67.5% and 75.8% respectively but these are small in difference, with a maximum of 2.8%. The percentages with the non-normalized methods were only 3.3% and 0.08% respectively.

Data were found to be normally distributed. Statistically significant differences between AD patients and controls were found in all four methods, with electrodes from the anterior and the posterior of the head more likely to highlight statistically significant differences as shown in figure 1. Equation (1) was the most successful, with 21 statistically significant pairs identified and (3) gave the most statistically significant differences of \( p=0.0016 \) at Fp1-P3. These statistically significant electrode pairs are shown in figure 1.

ANOVA identified no significant interaction between patient diagnosis and electrode pairs. Equation (1) and (3) identified diagnosis as more significant than electrode pairs though both were extremely significant. This was reversed with (2) and (5) though again all were found to be significantly statistically significant. Multiple comparison analysis identified statistically significant \( (p<0.01) \) comparisons with (2) of 90 statistically comparisons (of 7140) and (5) of 104 statistically significant comparisons. No significant electrode pairs were found for (1) or (3) with multiple comparison corrections.

With CSD modified results AD patients had a lower dLZC than controls in 58% of all calculations, not including results from (4) as this again was found not to be a distance measure. Again (2) and (5) showed significantly higher percentages of AD patients with higher dLZCs than controls. Results were found to be more normally distributed with less electrode pairs identified as showing statistically significant differences between the two subject groups. Electrode pairs identified as statistically significant were Fp2-O2, F3-O1, F3-O2 \( (p=0.0035) \), F4-O2, F7-O2, P4-O2, T5-O2 and P3-P4 for (1), P3-O1, P3-O2 \( (p=0.0019) \), P4-O2, P4-T5 and T5-O2 for (3) and P3-T3 and P3-T4 \( (0.0003 \), the minimum \( p \) value seen in this study) for (5), reducing the statistical significance of the anterior part of the brain. The statistically significant electrode pairs for (1) and (3) are also shown in figure 1. It

![Figure 1](image-url)
can be seen from comparing the top two images and the bottom two images the significant impact on results of volume conduction.

ANOVA again identified no significant interactions in any cases other than (5) where \( p = 0.0009 \). All equations identified both electrode pairs and diagnosis as significant factors for analysis with a maximum of \( p = 8 \times 10^{-6} \) (Diagnostic significance for (2)) and a minimum of \( p = 4 \times 10^{-44} \) (electrode pair significance for (5)). Even with multiple comparison correction, methods (1), (2) and (5) showed statistically significant electrode pair comparisons in 6, 60 and 49 cases respectively.

IV. DISCUSSIONS

In this pilot study, a distance based bivariate measure of LZC was used to analyze the complexity of the EEG of 22 subjects, 11 AD patents and 11 age-matched controls, having first undergone two different preprocessing steps. Data calculated from EEG signals, which did not undergo CSD transformation, showed a greater number of statistically significant differences between the two diagnostic groups than data calculated from EEG signals that had undergone CSD transformation. As the CSD transformation reduces the impact of volume conduction on recorded signals, it is our belief that the larger significance of dLZC results caused by not mitigating volume conduction factors are spurious.

As mentioned previously, the impact of volume conduction on results from EEG study using linear signal processing techniques such as synchrony and coherence are already clearly understood in the research community. The results of this study show these effects are also seen when using non-linear methods, such as LZC, to analyze differences in signals across the brain.

Although, in figure 1, the number of statistically significant electrode pairs clearly reduces, it can also be seen that the focus of statistically significant differences between the two diagnostic groups remains towards the posterior part of the brain, where there is a clear overlapping of statistically significant brain regions. While our hypothesis was proven, it is clear that the impact of volume conduction was not to change the results of this study, which can be derived from the data calculated, but to affect the data created.

Thus, while the data of previously published bivariate and multivariate non-linear signal analysis is highly likely to have been affected by volume conduction, the results gained from those studies are still valid in light of our findings. Caution should remain, however, when investigating results from any bivariate or multivariate signal processing EEG study where volume conduction effects have not been mitigated.

V. CONCLUSION

In conclusion, this pilot study showed that non-linear signal processing techniques based on LZC are significantly influenced by volume conduction. Though the number of statistically significant electrode pairs was smaller after CSD transformation, there was still a substantial overlap between reliable and spurious findings, indicating that previously published bivariate and multivariate results remain valid in light of our findings. It is suggested, however, that CSD transform or other volume conduction mitigation be used in pre-processing for any bivariate or multivariate method to reduce spurious coupling.

ACKNOWLEDGMENT

We would like to thank Dr Pedro Espino (Hospital Clinico San Carlos, Madrid, Spain) for his help in the recording and selection of EEG epochs.

REFERENCES