

# Space-time ICA versus Ensemble ICA for ictal EEG analysis with component differentiation via Lempel-Ziv complexity

Christopher J. James\*, *Senior Member, IEEE*, Daniel Ab solo, *Member, IEEE*, and Disha Gupta, *Student Member, IEEE*

**Abstract**—In this proof-of-principle study we analyzed intracranial electroencephalogram recordings in patients with intractable focal epilepsy. We contrast two implementations of Independent Component Analysis (ICA) – Ensemble (or spatial) ICA (E-ICA) and Space-Time ICA (ST-ICA) in separating out the ictal components underlying the measurements. In each case we assess the outputs of the ICA algorithms by means of a non-linear method known as the Lempel-Ziv (*LZ*) complexity. *LZ* complexity quantifies the complexity of a time series and is well suited to the analysis of non-stationarity biomedical signals of short length. Our results show that for small numbers of intracranial recordings, standard E-ICA results in marginal improvements in the separation as measured by the *LZ* complexity changes. ST-ICA using just 2 recording channels both near and far from the epileptic focus result in more distinct ictal components – although at this stage there is a subjective element to the separation process for ST-ICA. Our results are promising showing that it is possible to extract meaningful information from just 2 recording electrodes through ST-ICA, even if they are not directly over the seizure focus. This work is being further expanded for seizure onset analysis.

## I. INTRODUCTION

THE electroencephalogram (EEG) is particularly useful in the diagnosis of epilepsy, which is a potentially debilitating disorder characterized by sudden and recurrent brain dysfunction called epileptic seizures. Whilst most seizures can be controlled with antiepileptic drugs, for most intractable cases surgery is possibly the only option. Synchronized ictal EEG activity is generally observed during epileptic seizures. This activity may be focal, multi-focal or generalized and may change in focus as the seizure progresses. Seizure onset is of particular interest as it holds clues to the epileptogenesis and so studies of seizure onset are of immense value in understanding *what* is happening and *where*; early on in the occurrence of a seizure. Any spread of activity to other brain areas is of clinical relevance [1]. Due to the large amount of multi-channel recordings generated in dedicated Epilepsy Monitoring Units, automated methods of detecting seizures and seizure onset within digitally recorded multi-channel EEG is of immense value to clinicians. Many

\*C.J. James and D. Gupta are with the Signal Processing and Control Group, ISVR, University of Southampton, SO17 1BJ, UK, (Tel: +44 (0)2380 593043; Fax: +44 (0) 23 8059 3190; e-mail: C.James@soton.ac.uk).

D. Ab solo is with the Biomedical Engineering Group, Department of Signal Theory and Communications, E.T.S.I. de Telecomunicaci n, University of Valladolid, Camino del Cementerio s/n, 47011, Valladolid, Spain.

different criteria have been evaluated in the past and recent progress in the theory of non-linear dynamics has provided new methods for the study of the EEG in this manner. Non-linear dynamical analysis techniques may be a better approach than traditional linear methods to obtain a better understanding of EEG dynamics, especially for ictal EEG.

In previous work we have decomposed ictal EEG recordings using the Blind Source Separation (BSS) technique of Independent Component Analysis (ICA) [2]. This method decomposes a set of measurements into a corresponding set of underlying sources or components using the assumption of statistical independence between the underlying sources. Ensemble ICA (E-ICA) derives a series of spatial filters from multi-channel recordings. Where single channel recordings only are available or desirable we introduced a Single-Channel ICA (SC-ICA) algorithm that could extract multiple underlying sources [3]. Here we introduce the use of Space-Time ICA (ST-ICA) where the SC-ICA process is applied to a small number of simultaneously recorded channels.

The present proof-of-principle study examines intracranial EEG recordings of epilepsy patients with intractable focal seizures. We introduce the ST-ICA algorithm applied to two recording electrodes both within and without the seizure focus, and compare performance between the raw recordings, E-ICA and ST-ICA. We assess the performance by computing a particular non-linear method suitable for biomedical signal processing called the Lempel-Ziv (*LZ*) complexity. The *LZ* complexity is a non-parametric measure of complexity for one-dimensional signals [4], [5].

## II. MATERIAL AND METHODS

### A. Ictal EEG

The data were recorded during pre-surgical evaluation at the Epilepsy Center of the University Hospital of Freiburg, Germany. Intracranial grid-, strip-, and depth-electrodes were used. The EEG data were acquired using a Neurofile NT digital video EEG system with 128 channels, 256 Hz sampling rate, and a 16 bit analogue-to-digital converter. 23 EEG recordings with simple partial, complex partial and generalized tonic-clonic seizures from patients with focal epilepsy originated in the temporal region were recorded. For this study we depict the results on two of these patients. The subjective nature of the feature selection – as described in the next section – makes automated analysis of many patients

impractical at this stage.

### B. Ensemble (Spatial) ICA

In the standard, noise free, formulation of the ICA problem, the observed signals  $\mathbf{x}(t)$  are assumed to be a linear mixture of an equal number of unknown but statistically independent source signals  $\mathbf{s}(t)$ , i.e.  $\mathbf{x}(t)=\mathbf{A}\mathbf{s}(t)$ , where  $\mathbf{A}$  is the square mixing matrix which is unknown but invertible. The problem is solvable up to a permutation, and sign and power indeterminacy of the sources, by finding an appropriate de-mixing matrix  $\mathbf{W}=\mathbf{A}^{-1}$  which allows estimation of the source waveforms by  $\mathbf{s}(t) = \mathbf{W}\mathbf{x}(t)$ . In this way the E-ICA algorithm derives a series of spatial filters in the form of the columns of the mixing matrix  $\mathbf{A}$  through the multi-channel observations.

Different ICA algorithms can be derived based on different theoretic techniques. Infomax ICA and Fast ICA are two of the more popular and referenced ICA techniques in the literature, more information about these algorithms can be found in [2]. Here we apply the Fast ICA algorithm.

### C. Single Channel (Temporal) ICA

For the SC-ICA formulation a ‘multi-channel’ data representation of the single data channel is required. This is obtained by generating a series of delay vectors to form a matrix of delays [3]. Assuming a single data channel with  $N$  elements:  $\{\mathbf{x}_t\}_{t=1,\dots,N}$ , then delayed vectors in the constructed matrix are obtained. The delay matrix  $\mathbf{v}$  is formed by obtaining  $\mathbf{v}_t$  for successive values of  $t$ , and combining these to form

$$\begin{aligned} \mathbf{v}_t &= [x_t, x_{t+1} \dots x_{t+m-1}]^T, \\ \mathbf{v} &= [\mathbf{v}_1, \mathbf{v}_{m+1}, \dots, \mathbf{v}_{m\tau+1}], \end{aligned} \quad (1)$$

where  $\tau$  is the lag term, and  $m$  is the number of lags (or the embedding dimension). We set the lag  $\tau$  to 1 and  $m=95$  based on previous research on this subject [3]. After Fast ICA is applied to the delay matrix the significance of each IC in turn can be assessed by projecting each back to the measurement space in isolation such that  $\mathbf{Y}^i = \mathbf{a}_i \mathbf{u}_i^T$ , where  $\mathbf{u}_i$  is the  $i$ th IC ( $i = 1, 2, \dots, p$ ),  $p$  is the number of ICs,  $\mathbf{a}_i$  the corresponding column of the mixing matrix  $\mathbf{A}$  and  $\mathbf{Y}^i$  the resulting ‘matrix of delay vectors’. From  $\mathbf{Y}^i$  it now becomes possible to extract the projected time series,  $y_i(t)$ , by performing an average of the rows of the matrix  $\mathbf{Y}^i$ , in order to recover the time series (see [3] for more details on this technique). In SC-ICA the columns of the mixing matrix are interpreted as shifted versions of the mixing filters.

### D. Space-Time ICA

In ST-ICA the method of delays as for SC-ICA is used and is repeated for each channel of interest. In this way a series of delay matrices are ‘stacked’ to form a complete data matrix such that

$$\mathbf{v}^{Tot} = \left[ \begin{pmatrix} \mathbf{v}^1 \end{pmatrix}^T \quad \dots \quad \begin{pmatrix} \mathbf{v}^n \end{pmatrix}^T \right]^T \quad (2)$$

for an  $n$  channel system of  $N$  samples, such that  $\mathbf{v}^{Tot}$  is an  $nm \times N$  matrix. Fast ICA is then applied to this matrix and a number of ICs are derived. For each IC, each column of the mixing matrix  $\mathbf{A}$  is composed of  $n$  mixing filters – i.e. a mixing filter per measurement channel.

### E. Component Selection

Like most ICA algorithms the current bottle-neck in their application is the selection of relevant components, as for the most part this is a subjective process. In SC-ICA and ST-ICA the mixing filters generated usually represent shifted copies of the same filters and so some clustering must take place to group similar filters together [3]. This is still ongoing work but for this study we manually group mixing filters into 3 or 4 clusters based on the similarity between magnitude responses of each filter. Once the filters have been grouped the separated brain activity is obtained by summing the grouped ICs projected back to the measurement space as described in C above.

### F. Lempel-Ziv complexity

LZ complexity analysis is based on a coarse-graining of the measurements. Before calculating the LZ complexity measure  $c(n)$ , the signal must be transformed into a finite symbol sequence. In the context of biomedical signal analysis, typically the discrete-time biomedical signal  $\{x(n)\}$  is converted into a binary sequence. By comparison with a threshold  $T_d$ , the original signal samples are converted into a 0–1 sequence  $P = s(1), s(2), \dots, s(n)$ , with  $s(i)$  defined by,

$$s(i) = \begin{cases} 0 & \text{if } x(i) < T_d \\ 1 & \text{if } x(i) \geq T_d \end{cases} \quad (3)$$

We used the median as the threshold  $T_d$  because of its robustness to outliers. Previous studies [6] have shown that 0–1 conversion is adequate to estimate the LZ complexity in biomedical signals. This complexity measure can be estimated using a suitable algorithm as seen in [6], [7]. Larger values correspond to more complexity.

Here the LZ complexity is used to quantify changes in complexity of both the raw recordings as well as the transformed signals following E-ICA and ST-ICA.

## III. RESULTS

Here we apply both E-ICA on all 6 measurement channels, as well as ST-ICA on two groups of two channels. The data consists of 3 channels over the focus (1, 2 and 3) and 3 extra-focal (4, 5 and 6). Figs. 1 and 2 depict two recordings, one for each patient with focal temporal lobe epilepsy; the vertical lines in each figure represent the start and stop of the seizure as indicated by an epileptologist. After ICA the evolution over time using LZ complexity before, during and after the seizure is analysed using a sliding window of 10s, with a 1s shift for each window. Fig. 1a shows the LZ complexity values for each recording and changes in complexity can be observed over the 3 focal channels during the seizure (greatest over 2 & 3), there is no visible change in the 3 non-focal channels, other than a slight deviation on

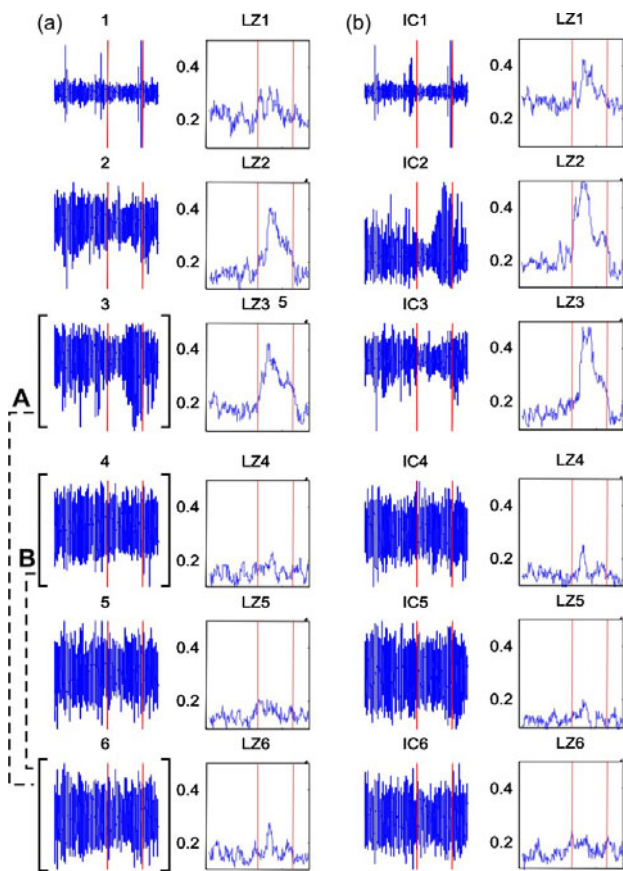


Fig. 1. (a) 6 channels (around 7 mins) of intracranial EEG for patient 21, channels 1-3 are focal and 4-6 are extra-focal. The LZ complexity value for each channel is depicted. Channels 3 & 6 and channels 4 & 6 are used for ST-ICA analysis. (b) 6 ICs extracted following E-ICA and their corresponding LZ complexity values.

channel 6. In Fig. 2a a similar picture emerges, changes in LZ complexity values over the focal channels (mainly 1 & 2) with very little evident changes in the non-focal channels. Figs. 1b and 2b depict the outputs of E-ICA and their corresponding LZ complexity values. It can be seen that the ICA process has succeeded in slightly unmixing seizure from background EEG with slightly elevated LZ complexity values over seizure onset. However, the unmixing process has not completely isolated seizure activity, most probably because the recordings are already fairly independent of each – this is apparent in the mixing matrices for each seizure shown in Fig. 3 where the seizure components map almost 1:1 onto their responding recording channels. Next, the two pairs of electrodes A:3&6 and B:4&6 (Fig.1a) for patient 21 and A:3&4 and B:4&6 (Fig. 2a) for patient 12 are analysed with

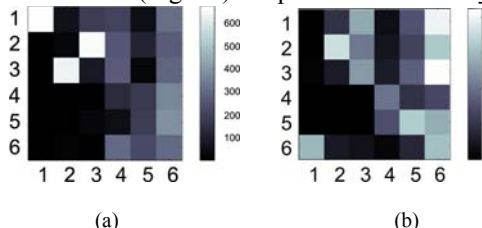


Fig. 3. Mixing matrix A following E-ICA for (a) patient 21 and (b) patient 12. The raw data and ICs are depicted in Figs. 1 & 2.

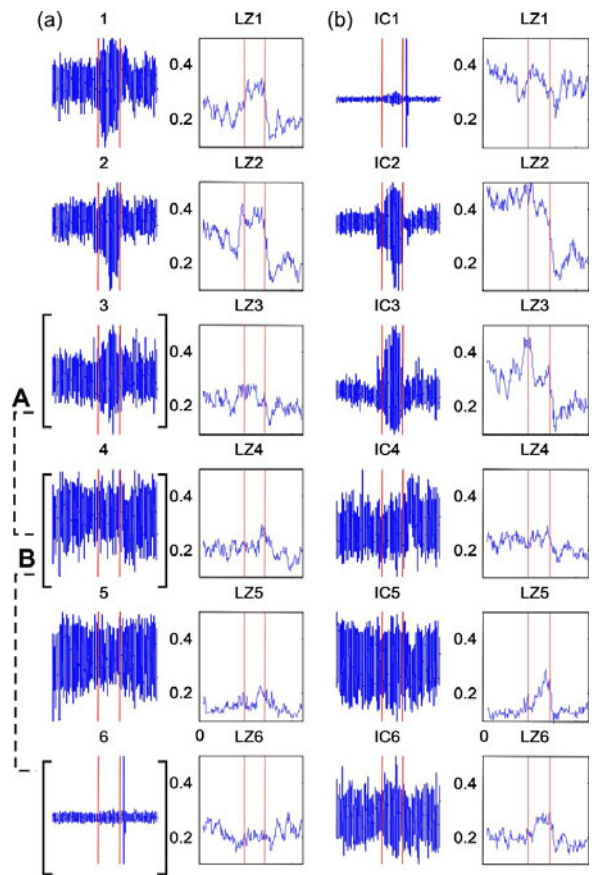


Fig. 2. (a) 6 channels (around 7 mins) of intracranial EEG for patient 12, channels 1-3 are focal and 4-6 are extra-focal. The LZ complexity value for each channel is depicted. Channels 3 & 4 and channels 4 & 6 are used for ST-ICA analysis. (b) 6 ICs extracted following E-ICA and their corresponding LZ complexity values.

ST-ICA. Figs. 4 and 5, a and c, depict the mixing filters learned by the ST-ICA process. These can be clustered into similar groups of shifted filters. Figs. 4 and 5, b and d, depict the extracted components after the mixing filters were manually clustered into 4 and 3 groups respectively. For patient 21, the LZ complexity values for the focal/non-focal pair reflect 2 distinct seizure related components with particular complexity profiles around the seizure. For the non-focal pair, 3 components emerge, one of which shows larger LZ complexity values during the seizure. For patient 12 the focal/non-focal pair yields 3 clusters with a less distinct increase in LZ complexity value over seizure for one cluster. The non-focal pair yields 4 clusters with non-distinct LZ complexity values. Of interest is that the LZ complexity value level of some extracted components is high throughout (pre- and post-ictally).

#### IV. DISCUSSION AND CONCLUSION

In this proof-of-principle study, we analyzed intracranial EEG recordings from 2 patients with focal epilepsy, contrasting E-ICA and ST-ICA. We used LZ complexity values to quantify the complexity in each IC for each case. Our results, although preliminary, show that; (i) although

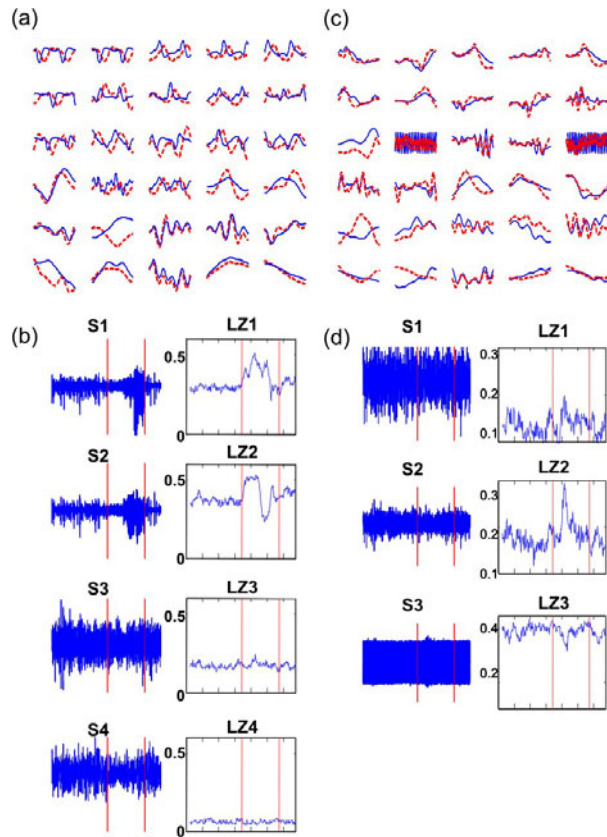


Fig. 4. (a) & (c) mixing filters following ST-ICA on data of patient 21 for focal/ non-focal (a) and non-focal (c) pairs. (b) & (d) the clustered and re-covered components and the LZ complexity values for each.

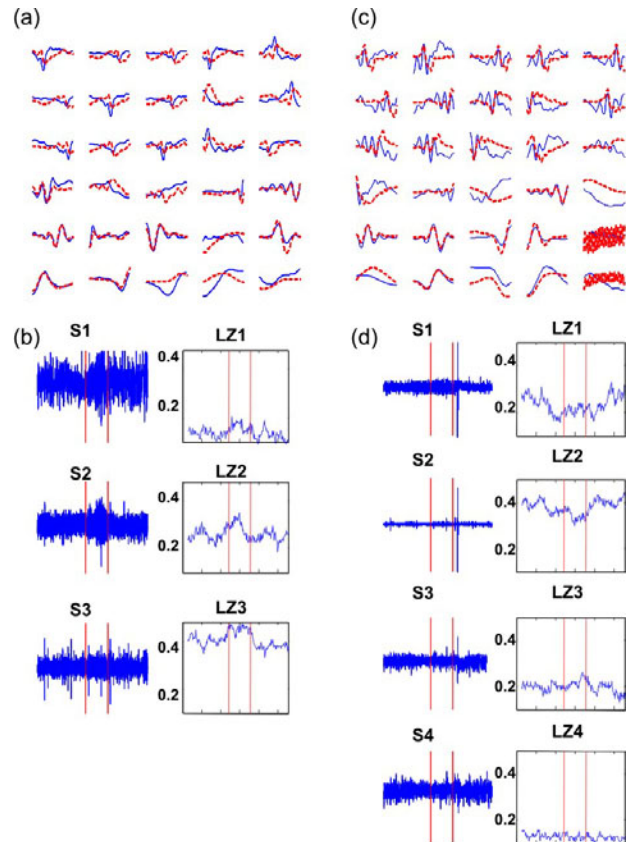


Fig. 5. (a) & (c) mixing filters following ST-ICA on data of patient 21 for focal/ non-focal (a) and non-focal (c) pairs. (b) & (d) the clustered and covered components and the LZ complexity values for each.

E-ICA performs a slight demixing of the measured cortical activity, the separation is only slightly improved – this is backed up by the changes observed in LZ complexity over the seizure period; (ii) ST-ICA can extract and de-mix meaningful information from as little as 2 recording channels even from channels not located over the focus – the current limitation is the requirement to manually cluster mixing filters, which is highly subjective (automating this process is the current focus of continued research); (iii) LZ complexity has already been shown as a useful measure of complexity for biomedical signals and in this context provides a useful indicator of extracted ictal activity. This technique shows promise as a method for extracting multiple sources underlying single or few channel recordings, as is generally the case with intracranial recordings for epilepsy analysis.

#### ACKNOWLEDGMENT

We are very grateful to the Epilepsy Center of the University Hospital of Freiburg, Germany for their permission to use the invasive EEG recordings in this study.

#### REFERENCES

- [1] B. Schelter, M. Winterhalder, T. Maiwald, A. Brandt, A. Schad, A. Schulze-Bonhage, and J. Timmer, "Testing statistical significance of multivariate time series analysis techniques for epileptic seizure prediction," *Chaos*, vol. 16, p. 013108, Jan. 2006.
- [2] C.J. James and C.W. Hesse, "Independent Component Analysis for Biomedical Signals," *Physiological Measurement*, 26, R15-R39, 2005.
- [3] M.E. Davis and C.J. James, "Source-separation using single-channel ICA," *Signal Processing*, vol. 87, issue 8, pp. 1819-1832, Aug 2007.
- [4] A. Lempel and J. Ziv, "On the complexity of finite sequences," *IEEE Trans. Inform. Theory*, vol. 22, pp. 75–81, Jan. 1976.
- [5] R. Nagarajan, "Quantifying physiological data with Lempel-Ziv complexity—certain issues," *IEEE Trans. Biomed. Eng.*, vol. 49, pp. 1371–1373, Nov. 2002.
- [6] X. S. Zhang, R. J. Roy, and E. W. Jensen, "EEG complexity as a measure of depth of anesthesia for patients," *IEEE Trans. Biomed. Eng.*, vol. 48, pp. 1424–1433, Dec. 2001.
- [7] X. S. Zhang, Y. S. Zhu, N. V. Thakor, and Z. Z. Wang, "Detecting ventricular tachycardia and fibrillation by complexity measure," *IEEE Trans. Biomed. Eng.*, vol. 46, pp. 548–555, May 1999.