EEG Analysis via Multiscale Lempel-Ziv Complexity for Seizure Detection

N. Sertac Artan
New York Institute of Technology
School of Engineering and Computing Sciences
New York, New York 10023
Email: nartan@nyit.edu

Abstract—Robust seizure detection and seizure prediction continues to be a challenge. Lempel-Ziv Complexity (LZC) is one of the features that has shown to be relevant in seizure detection. Recent work has shown that augmenting LZC can be beneficial to emphasize variations in amplitude or frequency when analyzing biomedical signals. In this paper, we present a first look into evaluating the feasibility of using a recently proposed feature stemmed from LZC, namely the Multiscale Lempel-Ziv Complexity (MLZC) for seizure detection. MLZC does not allow the high-frequency signal components to be overwhelmed by the low frequency signal components when calculating complexity values. We compare MLZC and LZC for identifying seizures for three cases and show MLZC can provide a clear separation between non-ictal and ictal periods for all three cases using a single threshold over 7 recordings and 7 seizures per patient, whereas LZC provided such a clear separation for only one of the patients.

I. INTRODUCTION

Epilepsy is the fourth most common neurological disorder in United States and it is reported as of 2012 to affect 65 million people worldwide [1]. Various methods have been proposed for detecting and/or predicting epileptic seizures [2]. Most of these methods rely on one or more features [3] to ultimately classify the EEG activity into non-seizure and seizure periods.

One family of non-linear features is based on information theory, and includes Sample Entropy, Approximate Entropy, and Lempel-Ziv Complexity (LZC) [4]–[7], among others [2], [8]. Jouny and Bergey listed sample entropy and LZC among the five most reliable features to assess early seizure onset [5] within the set of 18 features they evaluated. LZC [9] has also been used successfully as a feature for assessing the complexity in other types of biomedical signals [10] for a variety of modalities and applications including diagnosis of depression [11] and distinguishing between sleep and awake cycles [12].

Recent studies identified potential cases where LZC can be augmented for better analysis in the domain of biomedical signal processing [4], [13], [14]. Sarlabous et al. [13] observed that LZC cannot discern between signals with similar random components but with different amplitude variations. They pointed out that this is due to the binary quantization of the signal in LZC. When LZC is applied in biomedical signal processing, the signal is traditionally first quantized into only two amplitude levels via a single threshold. Usually, the median of the whole recording is selected as this threshold, although mean and mode have also been applied successfully. Sarlabous et al. [13] addressed this issue by further increasing the quantizer resolution in an approach called Multistate Lempel-Ziv Complexity (MSTLZC). They stated that with a small number of quantization levels, MSTLZC can only distinguish complexity changes in the signal, whereas for high quantization levels MSTLZC can only distinguish the amplitude variations. It is also noted in their paper that if an intermediate number of quantization levels is used, MSTLZC can distinguish both complexity and amplitude variations. The work of Sarlabous et al. focuses on diaphragmatic mechanomyographic signals. Kamath used MSTLZC for distinguishing epileptic subjects from healthy subjects [15].

Another augmentation to LZC, which is also the focus of this paper is Multiscale Lempel-Ziv Complexity (MLZC). MLZC was recently proposed by Ibáñez-Molina et al. [14]. Ibáñez-Molina and colleagues have shown that the impact of high frequency components on the complexity value can be shadowed by the presence of high amplitude, low frequency components in EEG analysis. To illustrate their idea, they showed that LZC failed to distinguish between eyes open and eyes closed states for healthy subjects, whereas MLZC successfully showed high and low complexity, respectively. Later, Kalev et al. reported that MLZC has higher diagnosis accuracy in depression compared to LZC by incorporating the high frequency components [11].

In this paper, we present a brief evaluation of the seizure detection performance of MLZC with respect to LZC. This study is a first look at seizure detection performance of MLZC with a small number of cases, rather than an in-depth evaluation, which is left as a future work. For three cases, we compare MLZC and LZC for identifying seizures and show MLZC can provide a clear separation between non-ictal and ictal periods for all three cases using a single threshold over 7 recordings and 7 seizures per patient, whereas LZC provided such a clear separation for only one of the patients. More specifically, in the two cases, no single threshold is feasible for LZC that can separate non-ictal and ictal periods for all the 7 recordings, and all the 7 seizures for either of the patients. On the other hand, MLZC have 7 and 1 channels
for the two patients, respectively, where ictal values are all distinct from the non-ictal values and a single threshold can separate non-ictal periods from ictal periods for all the seizures in all the recordings. In the third patient’s test, both LZC and MLZC have many channels where ictal values are all distinct from the non-ictal values.

The rest of this paper is organized as follows. In Section II, we briefly summarize the Lempel-Ziv Complexity, whereas in Section III, Multiscale Lempel Ziv Complexity is introduced. We provide performance evaluation of MLZC for seizure detection in Section IV. Section V concludes the paper.

II. LEMPEL-ZIV COMPLEXITY (LZC)

The LZC calculation consists of three steps: (1) quantization, (2) counting unique pattern occurrences, and (3) normalization.

To calculate the LZC, first the time series data should be quantized into two levels. This is typically achieved by comparing the sample value with a threshold value such as the median of the overall signal as mentioned above.

To count the unique occurrences, let $P$ show the quantized sequence from the time series in the quantization step. For each element $i$, $e_i$ of the sequence, this step checks whether any of the $i$ subsequences, which include $e_i$ is unique. If a unique sequence is found, a counter, $c$ is incremented by one. If not, the next element in the sequence and the corresponding subsequence is evaluated until the evaluation of the last element in $P$ is completed. To avoid bias due to the length of the sequence, $|P|$, the count value found in step 2 should be normalized to finalize the LZC calculation. The normalized value, $C$, i.e., the LZC for a sequence of $n$ elements can then be given as in (1).

$$ C = \frac{c}{n/\log_2(n)} $$ (1)

III. MULTISCALE LEMPEL-ZIV COMPLEXITY (MLZC)

Multiscale Lempel-Ziv Complexity [14] allows a finer representation of the impact of frequency on the LZC. This is achieved by having multiple sliding windows rather than a single fixed-size sliding window. Each of these sliding windows in MLZC is tuned to a different frequency band to independently evaluate each band’s contribution on the complexity.

More specifically, MLZC defines $k$ windows where each window has a different length given as $W = \{w_1, w_2, \ldots, w_k\}$. For each sample $x(n)$, $k$ thresholds are defined $T_{d1}(n)$ to $T_{dk}(n)$ corresponding to the $k$ windows. The $i$th threshold for the $n$th sample is then calculated as (Eq. (3) from [14] repeated below for completeness):

$$ T_{di}(n) = $$ (2)
\[
\text{median}(x \left( n - \frac{w_i - 1}{2} \right), \ldots, x \left( n + \frac{w_i - 1}{2} \right))
\]

Each window provides information about a low-pass band determined by the window size. More specifically, the window size slightly larger than \( f_s / f \) captures the band \([0, f]\), where \( f_s \) is the sampling frequency. To evaluate the impact of an arbitrary, non-DC band, the difference of complexities between two windows can be used.

Note that according to (3), for a window with length \( w_i \), no MLZC complexity can be calculated for the first and last \((w_i - 1)/2\) samples in a recording. Since, the typical frequencies of interest are above 1 Hz, the excluded regions will be brief, and possibly less than a second. Thus, these excluded regions can be neglected in practice for seizure detection.

### IV. Results

We considered three cases from the MIT-CHB Scalp EEG database [16], [17]. Each of these patients have 7 recordings with one or more seizures and other recordings with no seizures. We studied seizures with recordings and only the first seizure in each recording.

For each recording for a patient, the LZC, and MLZC for various windows are calculated for epochs of 1 second, each. Four MLZC windows are considered at window sizes of \( W = \{w_1 = 35, w_2 = 17, w_3 = 9, w_4 = 7\}\). Performance results are provided for 4 MLZC bands in Table I: MLZC1 corresponds to the first window \((w_1 = 35)\), whereas MLZC2, MLZC3, and MLZC4 are calculated as the differences of MLZC values in neighboring windows \(W_2-W_1, W_3-W_2,\) and \(W_4-W_3\), respectively.

The mean of the complexity values (LZC, and MLZC for \(k\) windows) of all recordings of a patient is calculated for non-ictal and ictal phases. The non-ictal phase is defined as the time before the first seizure in the recording. The ictal phases are defined by the MIT-CHB database seizure times.

In Table I, the non-overlapping channel counts are shown for each patient. For patient A, B, C, both LZC and MLZC has many non-overlapping channels.

### V. Conclusion

Easily identifiable properties of seizures differ between the seizures of one patient, or in between different patients or recordings, making robust seizure detection a challenge. Lempel-Ziv Complexity (LZC) is one of the features that has shown to be relevant in seizure detection. In this paper, we present a first look into evaluating the feasibility of using a recently proposed feature stemmed from LZC, namely the Multiscale Lempel-Ziv Complexity (MLZC) for seizure detection. We show that MLZC may provide a better separation between non-ictal and ictal phases for at least some of the patients with respect to LZC. In our future work, we will evaluate and compare MLZC to LZC for seizure detection in a large set of patients.

### Acknowledgment

This work is partially supported by an NYIT Institutional Support of Research and Creativity Grant to the author. We also like to thank Dr. Paolo Gasti for providing additional server time.

### REFERENCES


### Table I

<table>
<thead>
<tr>
<th>Patients</th>
<th>A(3)</th>
<th>B(1)</th>
<th>C(14)</th>
</tr>
</thead>
<tbody>
<tr>
<td>LZC (All)</td>
<td>0</td>
<td>0</td>
<td>15</td>
</tr>
<tr>
<td>MLZC1 (6-14 Hz)</td>
<td>0</td>
<td>0</td>
<td>14</td>
</tr>
<tr>
<td>MLZC2 (6-14 Hz)</td>
<td>1</td>
<td>1</td>
<td>8</td>
</tr>
<tr>
<td>MLZC3 (14-28 Hz)</td>
<td>7</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>MLZC4 (28-36 Hz)</td>
<td>0</td>
<td>1</td>
<td>5</td>
</tr>
</tbody>
</table>