Multistate Lempel-Ziv (MLZ) Index Interpretation as a Measure of Amplitude and Complexity Changes

Leonardo Sarlabous, Abel Torres, *Member, IEEE*,
José A. Fiz, Joaquim Gea, Juan B. Galdiz, and Raimon Jané, *Member, IEEE*

Abstract— The Lempel-Ziv complexity (LZ) has been widely used to evaluate the randomness of finite sequences. In general, the LZ complexity has been used to determine the complexity grade present in biomedical signals. The LZ complexity is not able to discern between signals with different amplitude variations and similar random components. On the other hand, amplitude parameters, as the root mean square (RMS), are not able to discern between signals with similar power distributions and different random components. In this work, we present a novel method to quantify amplitude and complexity variations in biomedical signals by means of the computation of the LZ coefficient using more than two quantification states, and with thresholds fixed and independent of the dynamic range or standard deviation of the analyzed signal: the Multistate Lempel-Ziv (MLZ) index. Our results indicate that MLZ index with few quantification levels only evaluate the complexity changes of the signal, with high number of levels, the amplitude variations, and with an intermediate number of levels informs about both amplitude and complexity variations. The study performed in diaphragmatic mechanomyographic signals shows that the amplitude variations of this signal are more correlated with the respiratory effort than the complexity variations. Furthermore, it has been observed that the MLZ index with high number of levels practically is not affected by the existence of impulsive, sinusoidal, constant and Gaussian noises compared with the RMS amplitude parameter.

I. INTRODUCTION

THE complexity measure proposed by Lempel and Ziv (LZ) [1] has been extensively used in order to evaluate the randomness present in finite sequences. Different LZ variants have been used in order to: solve information theory problems [2], data codification applications [3], and data compression without losses purposes [4]. Recently, the LZ has been applied to different biomedical signals, such as: electrocardiographic signals [5], electroencephalographic signals [6]–[8], electromyographic signals [9], intracranial pressure signals [10], and mechanomyographic signals (MMG) [11].

L. Sarlabous, A. Torres and R. Jané are with Dept. ESAII, Universitat Politècnica de Catalunya, Institut de Bioenginyeria de Catalunya (IBEC) and CIBER de Bioingeniería, Biomateriales y Nanomedicina (CIBER-BBN), Barcelona, Spain (e-mail: leonardo.sarlabous@upc.edu, abel.torres@upc.edu, raimon.jane@upc.edu).

J. A. Fiz is with the Department of Pneumologie of the Navarra Hospital, Pamplona, Spain (e-mail: jfizfern@navarra.es).

J. Gea is with the Department of Respiratory Medicine at the IMIM-Hospital del Mar, UPF, CIBERES, Barcelona, Spain (e-mail: jgea@imim.es).

J. B. Galdiz is with the Department of Pneumologie of Cruces Hospital, Bilbao, Spain (e-mail: jbgaldiz@hcru.osakidetza.net).

To date, the studies performed in biomedical signals only try to quantify complexity changes, being far to inform about the amplitude variations. In most of these studies, the biomedical signal is transformed into a two symbols sequence by comparison with a threshold. In some cases the threshold corresponds to the median value [6], [7], [9], [10], [12], or mean value [5], [8], or modified mean value using dispersion parameters of the signal [5]. In few studies, the biomedical signal is transformed into a three symbols sequence using two thresholds obtained through median and dynamic range of the signal [7], [8].

In order to quantify not only the complexity but also the amplitude variations present in the signals, in this work we propose to use the LZ coefficient as proposed in [11], i.e., with more than two quantification levels, and with quantification thresholds that are fixed and independent of the dynamic range or standard deviation of the signal. With the intention of differentiate between this use of the LZ coefficient and the traditional application of the LZ coefficient to biomedical signals, we define the parameter utilized in this work as the Multistate Lempel-Ziv (MLZ) coefficient.

In [11], the correlation coefficient between MLZ complexity parameter obtained in diaphragmatic MMG signals and inspiratory pressure parameter was bigger than the correlation coefficient between the RMS amplitude parameter obtained in the MMG signal (that only evaluates the amplitude changes) and inspiratory pressure parameter.

Also, in studies performed in lung sounds signals (of vibratory nature, the same as the MMG signals) [13] and in MMG signals [14], the Shannon entropy by means of a histogram estimate of the signal was determined. In the same way, in these studies fixed and independent thresholds of the dynamic range of the corresponding signal were used. Thus, the Shannon entropy evaluates amplitude changes of the signal as well as changes in the probability density function, just as the MLZ coefficient.

Therefore, the objectives of this study are to analyze the MLZ capacity to monitor both complexity and amplitude changes in MMG signals, and to compare the MLZ and RMS performance in presence of different kind of noises.

II. METHODOLOGY

A. Biomedical signals

The performance of the MLZ coefficient has been tested using an example of 30 s of duration of a real diaphragmatic MMG signal extracted from a signal database acquired in an

animal model (dogs) [14]. In this data base two biomedical signals were acquired: the diaphragmatic MMG signal and the inspiratory pressure. The dogs performed an inspiratory progressive resistive load respiratory test during the acquisition. This test was performed with the intention to increase the respiratory muscular force performed by the dogs. The diaphragmatic MMG signal was acquired with a Kistler 8302A capacitive accelerometer placed on the surface of the thoracic cage. The placement of the sensor was between the seventh and eighth intercostal spaces in the anterior axillary line. Inspiratory pressure (P_{ins}) was measured with a pressure transducer placed in the trachea.

All analog signals were amplified, analog filtered, digitized with a 12 bit A/D system at a sampling rate of 4 kHz, and decimated at a new sampling rate (MMG: 200 Hz; P_{ins}: 100 Hz). An example of the P_{ins} and diaphragmatic MMG signal corresponding to six respiratory cycles (with increasing level of respiratory effort) is shown in the first two rows of Fig.1.

B. Multistate Lempel-Ziv coefficient (MLZ)

The Lempel-Ziv complexity analysis is based on a quantification of the signal to be analyzed, i.e., the signal is transformed into a sequence of symbols S with β quantization levels. Then the number of distinct patterns P contained in a given sequence is determinated.

The sequence $S = s(1), s(2),\ldots s(N)$ is scanned from left to right and the complexity counter P is increased by one unit every time a new subsequence of consecutive characters is found in the scanning process.

Once the number of distinct patterns P has been found, the LZ complexity value of the signal C_β can be obtained by normalizing P in function of the length of the analyzed sequence N and the number of quantization levels β [1],

$$C_\beta = \frac{P \log_\beta(N)}{N}$$

where $\log_\beta(N)$ is the logarithm in base β of N.
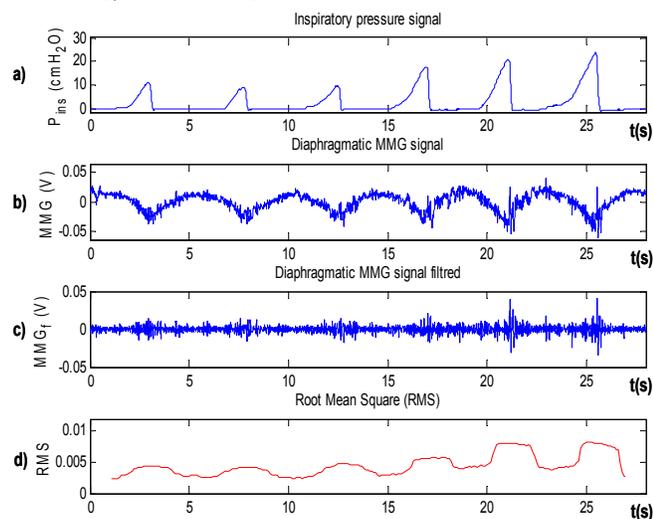
Fig. 1. (a) Inspiratory pressure signal, (b) diaphragmatic MMG signal without filtering, (c) diaphragmatic MMG signal band-pass filtered between 5 Hz and 25 Hz, and (d) RMS of the band-pass filtered MMG signal using a moving window of 2 s of duration

In the two states LZ version, the signal to be analyzed is transformed into a 0-1 sequence using as threshold U_d the median value [6], [7], [9], [10], [12], or mean value [5], [8], or modified mean value using dispersion parameters of the signal [5]. In the three states LZ version, the signal is transformed into a 0-1-2 sequence using two thresholds U_{d1} and U_{d2} that are function of the median and the dynamic range of the signal [7], [8].

In a similar way, in the β states LZ version (Multistate Lempel-Ziv: MLZ), before the MLZ coefficient is computed, the signal to be analyzed X must be transformed into a 0-1-2,…, β -1 sequence, where β is a integer number bigger than three. In this case, the thresholds corresponding (U_{d1}, U_{d2},…,U_{dK}, where $K=\beta+1$) are conformed as follows: 1) the median, mean, mode or other central tendency parameter of the signal X is estimated (x_m), 2) the x_m value is subtracted to the original signal, i.e., $X_m = X - x_m$, 3) the maximum absolute value of the signal X_m is sought out: x_{abs}, 4) the lower and upper bounds, $l_i = - x_{abs} - c$ and $l_s = x_{abs} + c$, respectively, are fixed, where c is a positive real number, 5) K thresholds equally spaced between l_i and l_s are determined. Finally, the sequence S is conformed by the following rule:

$$s(i) = \begin{cases} \text{"0"} & \text{if} \quad U_{d1} < x(i) \le U_{d2} \\ \text{"1"} & \text{if} \quad U_{d2} < x(i) \le U_{d3} \\ \vdots & \qquad \vdots \\ \text{"}\beta\text{-1"} & \text{if} \quad U_{dK-1} < x(i) \le U_{dK} \end{cases}, \quad \text{for } i = 1, 2,\ldots, N$$

Once the sequence S has been conformed, the MLZ complexity is determined by means of the LZ algorithm. If K is an odd number, one of the quantification thresholds coincides with the central tendency parameter x_m, while if K is an even number, the x_m value is centered between two thresholds. For $K = 3$ and $K = 4$ quantification levels ($\beta = 2$ and $\beta =3$), using appropriates values of x_m and c, the MLZ coefficient coincides with the standard two and three states LZ complexity analysis, respectively.

In this work the mode value has been proposed to be used instead of median or mean value in the conformation of the thresholds. The mode value is the most probable value and, in this way, maximizes the number of crossings or number of state changes produced. In general, for two quantification levels the LZ complexity value is maximized when the mode value is used.

C. Signal processing

A diaphragmatic MMG signal corresponding to six respiratory cycles before and after to adding different noise components, was analyzed. These noise components of 2 s of duration were added in each cycle of the signal and they are: a signal composed by zeros, a constant value signal, a sinusoidal component signal, a rectangular pulse and random component signal.
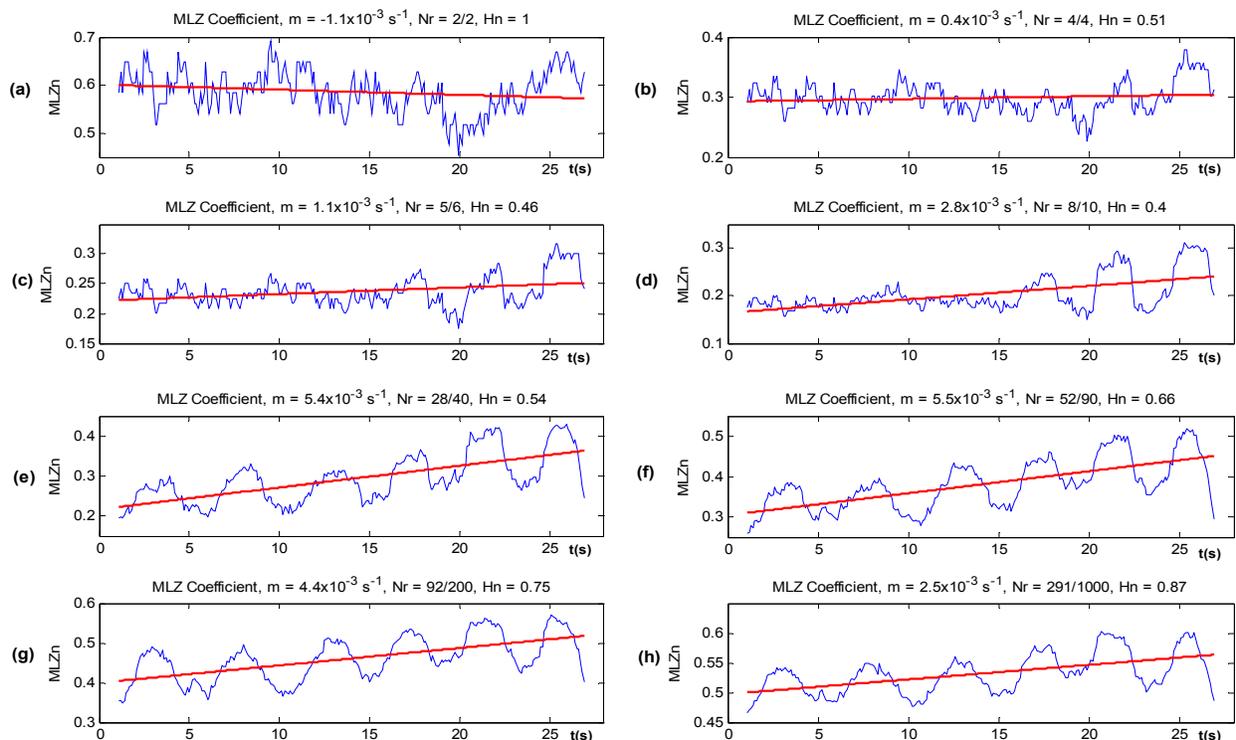
Fig. 2. MLZ of the diaphragmatic MMG signal using moving windows of 2 s of duration with 2, 4, 6, 10, 40, 90, 200 and 1000 quantification levels (a, b, c, d, e, f, g and h, respectively). For each realization it is shown: the slope of the lineal regression line (m), the ratio between the number of utilized quantification levels versus the total number of quantification levels (Nr), and the normalized entropy value of the Nr distribution.

In both signals (before and after noise addition), a band-pass Butterworth filter with the band frequency of interest in MMG signals (lower cutoff frequency of 5 Hz and upper cutoff frequency of 25 Hz) was applied, in order to eliminate the low frequency component produced by the thoracic cage movement during respiration and other biological sound sources of higher frequencies (mainly heart sounds) [15].

The RMS and MLZ parameters were determined over a moving window of 2 s of duration in both signals. The MLZ complexity was estimated using 2, 4, 6, 10, 20, 40, 200 and 1000 quantification levels.

The relationship among the inspiratory pressure signal and both MLZ and RMS parameters determined in MMG signals was analyzed by means of the *Pearson* correlation coefficient.

III. RESULTS

In Fig. 2 it is shown the evolution of the MLZ index determined in the MMG signal before adding the noise components, for all the quantification levels. In each case, the slope of the lineal regression line (m), the number of quantification levels used (Nr) and the normalized Shannon entropy value of the Nr distribution (Hn) were computed. The waveforms described by 2 (in this case the MLZ coincides with the LZ complexity index with two symbols) and 4 quantification levels (Fig. 2. a and b, respectively) are very similar (with different absolute values due to the normalization) and they do not follow a definite pattern. For these quantification levels the MLZ coefficient gives only information about the complexity measure, and do not

correlate well with the respiratory activity. With the increase of the number of quantification levels the waveform of the MLZ is more affected by the amplitude variations of the signal. This behavior is observed starting from 10 quantification levels. With 40 quantification levels all respiratory cycles are already perfectly defined.

Fig. 3 shows on one hand, the results of the correlation between P_{ins} signal with: the RMS, and the MLZ parameter with 2, 4 and 200 quantification levels determined in the MMG signal without noise components added (Fig 3. a, b, c and d, respectively). An increment of the correlation index when it increases the number of quantification levels of the MLZ was observed. In this sense, the correlation value reported by MLZ with 200 quantification levels is higher than the one reported by the RMS. The lowest correlation index was obtained with 2 quantification levels. On the other hand, Fig.3 shows also the results of correlation between P_{ins} with: the RMS and the MLZ parameter with 2, 4 and 200 quantification levels, for the MMG signal with noise components added (Fig 3. e, f, g and h, respectively). In this case, lower correlation coefficients were obtained. However, the decrease in the correlation coefficient between the RMS and the P_{ins} signal is higher than the decrease produced in the correlation index between MLZ with 200 quantification levels and P_{ins} signal.

As an example, Fig. 4 shows the RMS (upper plot) and the MLZ (lower plot) with 200 quantification levels determined in the MMG signals with and without noise addition. It can be seen that the MLZ is less affected by the noise than the RMS.
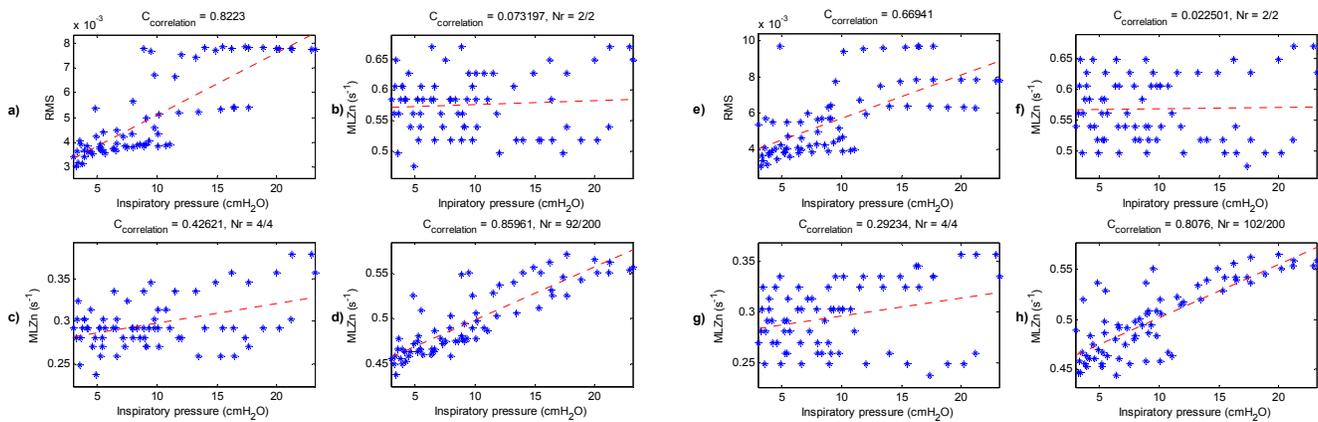
Fig. 3. Correlation plots obtained in diaphragmatic MMG signal without (a, b, c and d) and with (e, f, g and h) noise components added between inspiratory pressure and: RMS parameter (a and e), MLZ with 2 (b and f), MZL with 4 (c and g) and MLZ with 200 quantification levels (d and h).
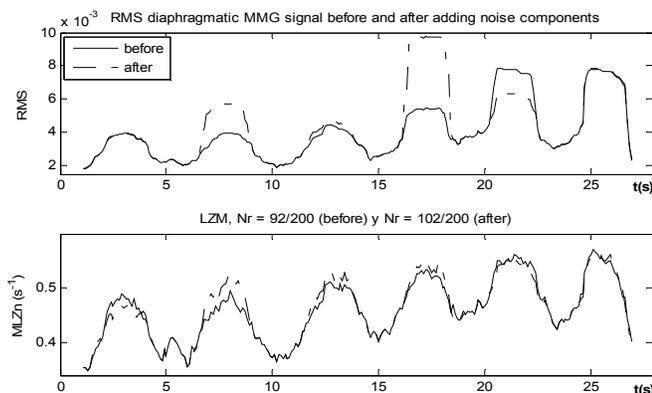
Fig. 4. RMS (a) and MLZ with 200 quantification levels (b) evolution curves determined in diaphragmatic MMG signal before and after adding noise components obtained using a moving window of 2 s of duration

IV. DISCUSSION AND CONCLUSIONS

In this work LZ coefficient using more than two quantification states with fixed and independent thresholds of the dynamic range or standard deviation to the signal (Multistate Lempel-Ziv: MLZ) was presented. We have observed that the MLZ with 2 quantification levels only evaluates complexity (LZ classic method). The MLZ with a high number of quantification levels gives information about the amplitude changes of the signal. Finally, the MLZ with an intermediate number of quantification levels provides information about both complexity and amplitude changes.

The results obtained indicate that the highest correlation index was reported between P_{ins} signal and MLZ with 200 quantification levels. The lowest correlation index was reported between P_{ins} signal and MLZ with 2 quantification levels, indicating that there is not correlation between P_{ins} and the complexity of the MMG signal. On the other hand, the correlation index between MLZ coefficient and P_{ins} signal was bigger that the correlation index between RMS parameter and P_{ins} signal. Furthermore, the RMS parameter is more affected by the presence of noise than the MLZ coefficient. These results suggest that the MLZ coefficient with a high number of states is more robust to noise than the RMS parameter.

REFERENCES

[1] A. Lempel and J. Ziv, "On the complexity of finite sequences," IEEE Trans. Inform. Theory, vol. IT-22, no.1 pp. 75-81, 1976

[2] S. Savari, "Redundancy of the Lempel-Ziv incremental parsing rule", IEEE Trans. Inf. Theory, vol. 43, no. 1, pp. 9–21, 1997

[3] G. Louchard and W. Szpankowski, "On the average redundancy rate of the Lempel-Ziv code," IEEE Trans. Inf. Theory, vol. 43, no. 1, pp. 2–8, 1997

[4] J. Adiego, G. Navarro and P. De la Fuente, "Lempel-Ziv Compression of Structured Text", Proc. of the Data Compression Conf. IEEE, pp. 112-121, 2004

[5] X. Zhang, Y. Zhu, N.V. Thakor and Z. Wang, "Detection Ventricular Tachycardia and Fibrillation by Complexity Measure", IEEE Trans. Biomed. Eng., vol. 46, no. 5, pp. 548-555, 1999

[6] J. Hu, J. Gao and J.C. Principe, "Analysis of biomedical signals by the Lempel-Ziv complexity: the effect of finite data size", IEEE Trans. Biomed. Eng., vol. 53, no. 12, pp. 2606-2609, 2006

[7] D. Abásolo, R. Hornero, C. Gómez, M. García and M. López, "Analysis of EEG Background Activity in Alzheimer's Disease Patients with Lempel-Ziv Complexity and Central Tendency Measure," Medical Engineering & Physics, vol. 28, no. 4, pp. 315-322, 2006

[8] X. Zhang, R.J. Roy and E. Weber, "EEG complexity as a measure of depth of anesthesia for patients", IEEE Trans. Biomed. Eng., vol. 48, no. 12, pp. 1424-1433, 2001

[9] R. Nagarajan, "Quantifying physiological data with Lempel-Ziv complexity-certain issues", IEEE Trans. Biomed. Eng., vol. 49, no. 11, pp. 1371-1373, 2002

[10] M. Aboy, R. Hornero, D. Abásolo and D. Álvarez, "Interpretation of the Lempel-Ziv Complexity Measure in the Context of biomedical signal analysis", IEEE Trans. Biomed. Eng., vol. 53, no. 11, pp. 2282-2288, 2006

[11] A. Torres, J.A. Fiz, R. Jané, E. Laciar, J.B. Galdiz, J. Gea and J. Morera, "Rényi Entropy and Lempel-Ziv Complexity of Mechanomyographic Recordings of Diaphragm Muscle as Indexes of Respiratory Effort", Proc. 30th Annual Int. Conf. IEEE-EMBS, pp. 2112-2115, 2008

[12] R. Nagarajan, J. Szczepanski and E. Wajnryb, "Interpreting non-random signatures in biomedical signals with Lempel-Ziv complexity", Physica D, vol. 237, pp. 359-364, 2007

[13] A. Yadollahi and Z. Moussavi, "A Robust Method for Heart Sound Localization using Lung Sound Entropy", IEEE Trans. Biomed. Eng., vol. 53, no. 3, pp. 497-502, 2006

[14] A. Torres, J.A. Fiz, B. Galdiz, L. Gea, J. Morera and R. Jané, "Inspiratory Pressure Evaluation by means of the Entropy of Respiratory Mechanomyographic Signals," 28th Ann. Conf. IEEE-EMBS, pp: 5735-5738, 2006

[15] C. Orizio, "Muscle Sound: Bases for the introduction of a mechanomyographic signal in muscle studies", Critical reviews in biomedical engineering. pp 201-243, 1993