The distribution of information for sEMG signals in the rectal cancer treatment process

Paulina Trybek\textsuperscript{a}, Michal Nowakowski\textsuperscript{b}, Jerzy Salowka\textsuperscript{c}, Łukasz Machura\textsuperscript{a,∗}

\textsuperscript{a}Division of Computational Physics and Electronics, Institute of Physics, Silesian Centre for Education and Interdisciplinary Research, University of Silesia in Katowice, Poland
\textsuperscript{b}Department of Medical Education, Jagiellonian University Medical College, Krakow, Poland
\textsuperscript{c}Department of Surgery, Stanley Dudrick Memorial Hospital, Skawina, Poland

Abstract

The electrical activity of external anal sphincter can be registered with surface electromyography. This signals are known to be highly complex and nonlinear. This work aims in characterisation of the information carried in the signals by harvesting the concept of information entropy. We will focus of two classical measures of the complexity. Firstly the Shannon entropy is addressed. It is related to the probability spectrum of the possible states. Secondly the Spectral entropy is described, as a simple frequency-domain analog of the time-domain Shannon characteristics. We discuss the power spectra for separate time scales and present the characteristics which can represent the dynamics of electrical activity of this specific muscle group. We find that the rest and maximum contraction states represent rather different spectral characteristic of entropy, with close-to-normal contraction and negatively skewed rest state.

Keywords: surface electromyography, colorectal cancer, entropy

1. Introduction

The colorectal cancer (CRC) remains one of the most common cancer with a high mortality rate [1, 2]. The appropriate diagnostic procedures are required for both prior to treatment and during the multimodal therapy. In
order to implement the most effective healing process while increasing the probability of fast recovery, not only the tumor diagnostic techniques are important but also the monitoring of surrounding anatomical structures that may be potentially affected by the invasive therapy including surgery or radiation [3, 4]. On top of that the understanding of the evolution and mutation of the CRC cells are crucial for more effective future treatment [5]. In this work we study the results of some less conventional method of diagnosis of an external anal sphincter (EAS) activity before and after the surgery [6]. The electrical activity of this muscle group is investigated through the surface electromyography technique (sEMG) before and after the anterior resection (AR) or the lower anterior resection (LAR). The registered time series indicate a high degree of complexity, hence it is necessary to apply methods that adequately characterise the dynamics of the complex process hidden in the high-dimensional data sets [7]. Quite recently, considerable attention has been paid to the entropy-based techniques with a very promising potential in bio-medical signal processing [8, 9, 10, 11]. The idea of implementing the entropy notions for the measure of biodiversity is present in an increasing number of the studied biosystems, from the microscopic level [12, 13] to the whole body response that can be investigated by the spectrum of electrophysiological signals as electrocardiography (ECG) [14, 15, 16], encephalography (EEG) [17] or magnetoencephalography (MEG) [18].

This work presents the application of basics ideas taken from the information theory for the characterisation of a specific neuromuscular activity of EAS. To better understand of features of analysed signals and how complexity is assessed from them the standard Shannon entropy ($H_x$) with their frequency analogue – spectral entropy ($H_f$) were applied. A thorough spectral analysis was also carried out for the full characterisation of the examined time series.

2. Material and methods

Patients. The study group include 20 subjects. Among them were 6 female and 14 male with the average age 64.6. All patients were diagnosed with a rectal cancer and undergone one of a surgical procedure (LAR or AR).

Data acquisition. The series were recorded before surgery, and two times after, respectively 6 months and 1 year in the postoperative period. In addition the two extreme cases of muscle tension, relaxation and maximum voluntary contraction (MVC) were considered separately. The measuring
device developed at the Politecnico di Torino in collaboration with the OT-Bioelettronica consist of 3 rings of 16 silver/silver oxide bar electrodes. The proper measurement took 10 seconds each. The sampling frequency was set to 2048 Hz. Low and high pass filters were used at 10 and 500 Hz respectively and resulted in typical 3dB bandwidth for the Analog to Digital Converter.

3. Information entropy

In statistical thermodynamics the macroscopic state of the microsystem is characterised by the distribution of the possible states. The entropy of distribution $p_i$ of a discrete set $\{X_i\}_{i=1}^n$ of $n$ microstates is given by the Gibbs formula $S = -k_B T \sum_i p(X_i) \log p(X_i)$, where $T$ is a temperature and $k_B$ is the Boltzmann constant. This definition works also for systems far away from equilibrium. An early idea of Gibbs was transmitted to the information theory in 1948 by Claude Shannon [19] and since then serves as a popular definition of the complexity of signals [20].

3.1. Shannon Entropy

The average Shannon information content $H_x$ of an outcome $x$ characterises the average information contained in the signal. It calculates the number of possible states that the system represented by a respective time
series can take. $H_x$ is often associated with the degree of uncertainty present in the signal or with the information that is inversely proportional to its predictability. The basic formula for calculating $H_x$ of time series of $N$ samples $\{y_i\}_{i=1}^N$ with the probability distribution function $p(x_i)$ is given as a average number of the logarithm of the probability of a certain state $x_i$.

$$H_x = -\sum_{i=1}^{N} p(x_i) \log p(x_i)$$  \hspace{1cm} (1)

The higher values of $H_x$ are determined by more unpredictable character of the series. The maximum of $H_x$ occurs in the case when all the states are equally probable. The base for logarithm in (1) defines the unit of entropy. Throughout this paper we will use the original formulation of Shannon [19], where base equals 2 and the unit of entropy is called shannon $Sh$ or bit.

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure2.png}
\caption{The probability distributions of amplitudes for 16 electrodes, shifted along abscissa for visibility. State $x$ represent the voltage registered with sEMG.}
\end{figure}

### 3.2. Power Spectral Entropy

The Power Spectral Entropy ($H_f$) is a simple analog of the $H_x$ in the frequency–domain. It tells us how broadly the power in the signal is distributed across the available frequencies. It is used as a measure of the width
and uniformity of the power spectrum. The frequency components of the power spectrum are always found to be linearly independent, regardless of the presence or absence of correlations. It gives an advantage over the standard method of creating a histogram where the correlations present in the signal are lost. The main limitation of this method lies in the fact that the determination of the power spectrum filters out all the nonlinearities that might be present in the signal, thus is an insensitive method for more subtle features of the signals [21]. The first step in the procedure of estimation of the $H_f$ is the calculation of the Power Spectral Density $S_{xx}$ function. Next, the $S_{xx}$ is normalised by the sum of all its spectral components $s_i$.

$$S_{xx}(f_i) = \frac{1}{N} |x(f_i)|^2, \quad s_i = \frac{S_{xx}(f_i)}{\sum_i S_{xx}(f_i)}$$

(2)

Above $x(f_i)$ is the Fourier transform of the signal $x(t)$ and $f$ is the frequency in Hz. The Power Spectral Entropy is defined as

$$H_f = -\sum_i s_i \log s_i$$

(3)

If only few frequencies are present the $H_f$ will naturally be low valued, so deterministic systems with a single frequency, such as a sine wave, have an entropy value of 0. For the case of wide distribution of frequencies, like for the white or pink noises, the value of $H_f$ will be much higher. For instance white Gaussian noise have the maximal $H_f$ on a given scale.

4. Results

Shannon entropy. We start with the estimation of the probability distribution $p(x)$ of the amplitudes of the signal. The simplest way to do so is to use the values of the signal itself and calculate their frequency. It is done by creating histograms of the amplitudes and normalise them by total number of samples $N$. Exemplary plots of the histograms for 133 bins can be seen in Fig. 2 for all 16 channels at 5cm depth. At first glance all the characteristics look normal, although about 85% of the analysed signals fail Lilliefors test [22] and more than 90% of signals fail Anderson-Darling normality test [23] at the significance level $\alpha = 0.05$.

Next, for the estimated probabilities we calculate the Shannon entropy. In Fig. 3 we depict the distribution of the obtained values. They range
Figure 3: The distribution of Shannon entropy $p(H_x)$ calculated for all signals. Light solid lines represent histograms obtained separately for pre/post surgery recordings. Red dots mark the respective position of (left-to-right) mean, median and most probable value (max) of $H_x$. 
from 0.9492Sh to 0.9639Sh. The most common value is 0.9625Sh, while the average and median read 0.9602Sh and 0.9609Sh, respectively. All values are accurate to four significant digits. This characteristics stay similar for the pre/post stages of surgical treatment, c.f. light blue lines in Fig. 3.

**Spectral entropy.** Similarly we start with the calculation of power spectra. For sEMG signals we found them rather complex where certain timescales can be described with different distribution functions, see Fig. 4 for details. Slow dynamics ($f < 2^{7}$, dashed grey) can be represented by the stretched exponential distribution of the Weibull form $\propto (x/\lambda_1)^{\beta_1} \exp[-(x/\lambda_1)^{\beta_1}]$, with the average values of $\lambda_1 = 0.0316 \pm 0.0095$ and $\beta_1 = 1.5708 \pm 0.4988$. On the other hand relatively slow dynamics ($50 < f < 200$, dashed orange) can also be represented by the simple exponential distribution $\propto \exp(-x/\lambda_2)$ with the average $\lambda_2 = 0.0325 \pm 0.0105$. $S_{xx}$ for the fast and intermediate frequencies ($f > 2^{8}$, dashed green) is given by the stretched exponential distribution $\propto \exp(-x/\lambda_3)^{\beta_3}$ with the average values of $\lambda_3 = 0.1109 \pm 0.0124$ and $\beta_3 = 2.6261 \pm 0.2037$.

In comparison with the probability based $H_x$, the typical values of $H_f$ are drastically lower, see Fig. 5. The average value of the spectral entropy for the total number of signals equals $\langle H_f \rangle = 0.0179 \pm 0.0017$. The distribution of the spectral entropy over the whole samples in question is bi-modal, c.f. grey histogram in Fig. 5. This can be explained once we consider separately the rest (blue) and MVC (orange) states. For the rest state the shape of the distribution of spectral entropy reminds the distribution of the Shannon entropy, with specific most common (maximum) state and the negative skewness, see Tab. 1. The distribution of the $H_f$ for MVC is much closer to normal with the kurtosis 3.2969 and skewnes 0.2556. In contrast to the $H_x$ its frequency domain equivalent also indicates a significant decrease of the average value $H_f$ for MVC. Both distribution are rather leptokurtic as both produce rather above average number of outliers than the standard normal distribution.

This situation is partially justified by the concept of entropy as a measure of the diversity of the available states in the system. The reduction of these states is a consequence of the muscle contraction phenomenon itself. During the propagation of the action potential within the functional motor units the specific direction of the process is dominated, which automatically entails the decrease of the number of possible states that the system can choose.
Figure 4: Exemplary power spectrum calculated for the raw sEMG data. Black dots denote periodogram estimated with FFT. Red dots mark power spectrum calculated with the Welch’s method. $S_{xx}$ for the slow dynamics can be approximated by the stretched exponential distribution of the Weibull form (dashed grey) or simple exponential (dashed orange) distribution. Intermediate and fast dynamics is represented by the stretched exponential distribution (dashed green), see text for details.
Figure 5: The distributions of the spectral entropy $p(H_f)$ over the whole examined samples (gray) and separately for MVC (orange) and rest states (blue). Bar plots represent the histograms, while solid lines depict the kernel density estimation. Note shifted distribution of the MVC and rest states. See text for physiological explanation.

Table 1: The basic statistical description of the distributions of the spectral entropy in sEMG signals presented in Fig. 5 separately for rest and MVC states as well as for all the signals. All values are accurate to four significant digits. Note similar average values for all sets. Close-to-Gaussian distribution of $H_f$ for MVC state can be expressed with rather low skewness and similar kurtosis.

|      | total | rest  | MVC   |
|------|-------|-------|-------|
| mean | 0.0179| 0.0190| 0.0167|
| std  | 0.0017| 0.0015| 0.0011|
| skewness | -0.0894| -1.4040| 0.2556|
| kurtosis  | -0.2013| 4.3347 | 3.2969|
5. Discussion and Conclusions

The purpose of this work is to characterise the average information content of sEMG signals registered from external anal sphincter during multimodal rectal cancer therapy. The data were examined in terms of application of the basic notions of information theory with a special effort devoted to the frequency distribution and the spectral entropy which is built upon it. The spectral analysis itself brings a valuable information about the character of data. The power spectrum density for frequencies of the electrical activity of EAS is relatively wide, with the power that decrease exponentially for slow and even faster for the intermediate and higher frequency components. For the slow dynamics there is clear frequency of around 50Hz for which $S_{xx}$ is maximal for all signals.

The standard probability based Shannon entropy did not provide any significant discrepancy for the signals measured before and after the surgery or for the relaxation vs MVC stages. This is partially a consequence of noisy character of sEMG data and similar probability distributions obtained for all the compared states. The broad power spectrum identified below the Nyquist limit should result in the relatively large values of $H_f$, which is not confirmed by our results. The effect of the drastically decreasing of $H_f$ and its very small values can be justified by the dominance of the narrow-banded slow dynamics and fast decreasing power spectrum for fast dynamics. Despite the a little amount of spectral information contained by the EAS activity, in contrast to the Shannon estimator, $H_f$ parameter clearly distinguishes the extreme states of muscle tension (relax vs MVC). The anal sphincter muscle group resting activity contains a valuable amount of information which in frequency domain is greater than for the case of forced activity. This feature can of importance for the clinicians in case of pre-surgery survey with sEMG.

The main limitations of this study is due to the problem of intersubject variability. The large diversity in distribution of EAS innervation zones, mainly caused by the high level of the individual asymmetries significantly affects the differences between the compared groups. That effect potentially has the greatest impact on very weak ability to distinguish the different states of treatment. Also the computational technique used has its internal limitations such as insensitivity to non-linear character of data. However, to our knowledge this is the first study which implements the information theory to the complex neuromuscular EAS activity and further investigations of the issue is still required. In our future research we intend to particularly
concentrate on the methods that next to the statistical properties of data also taking into consideration the signal dynamics.

References

[1] J. Ferlay, I. Soerjomataram, R. Dikshit, S. Eser, C. Mathers, M. Rebele, D. M. Parkin, D. Forman, F. Bray, Cancer incidence and mortality worldwide: sources, methods and major patterns in globocan 2012, International journal of cancer 136 (5).

[2] M. Arnold, M. S. Sierra, M. Laversanne, I. Soerjomataram, A. Jemal, F. Bray, Global patterns and trends in colorectal cancer incidence and mortality, Gut (2016) gutjnl–2015.

[3] K. Ammann, W. Kirchmayr, A. Klaus, G. Mühlmann, R. Kafka, M. Oberwalder, A. De Vries, D. Öfner, H. Weiss, Impact of neoadjuvant chemoradiation on anal sphincter function in patients with carcinoma of the midrectum and low rectum, Archives of Surgery 138 (3) (2003) 257–261.

[4] T. J. Ridolfi, N. Berger, K. A. Ludwig, Low anterior resection syndrome: Current management and future directions, Clinics in colon and rectal surgery 29 (03) (2016) 239–245.

[5] S. F. Roerink, N. Sasaki, H. Lee-Six, M. D. Young, L. B. Alexandrov, S. Behjati, T. J. Mitchell, S. Grossmann, H. Lightfoot, D. A. Egan, A. Pronk, N. Smakman, J. van Gorp, E. Anderson, S. J. Gamble, C. Alder, M. van de Wetering, P. J. Campbell, M. R. Stratton, H. Clevers, Intra-tumour diversification in colorectal cancer at the single-cell level, Nature 556 (7702) (2018) 457.

[6] M. Nowakowski, K. A. Tomaszewski, L. Machura, P. Trybek, R. M. Herman, Sensitivity and specificity of multichannel surface electromyography in diagnosing fecal incontinence, Folia Med Cracov 57 (1) (2017) 29–38.

[7] P. Trybek, M. Nowakowski, L. Machura, Multifractal characteristics of external anal sphincter based on semg signals, Medical engineering & physics 55 (2018) 9–15.
[8] V. V. Nikulin, T. Brismar, Comment on multiscale entropy analysis of complex physiologic time series, Physical review letters 92 (8) (2004) 089803.

[9] M. Costa, A. L. Goldberger, C.-K. Peng, Multiscale entropy analysis of biological signals, Physical review E 71 (2) (2005) 021906.

[10] P. A. Kaplanis, C. S. Pattichis, D. Zazula, et al., Multiscale entropy-based approach to automated surface emg classification of neuromuscular disorders, Medical & biological engineering & computing 48 (8) (2010) 773–781.

[11] M. Borowska, Entropy-based algorithms in the analysis of biomedical signals, Studies in Logic, Grammar and Rhetoric 43 (1) (2015) 21–32.

[12] J. J. Stewart, C. Y. Lee, S. Ibrahim, P. Watts, M. Shlomchik, M. Weigert, S. Litwin, A shannon entropy analysis of immunoglobulin and t cell receptor, Molecular immunology 34 (15) (1997) 1067–1082.

[13] Z. Zhang, R. Xiao, Empirical study on entropy models of cellular manufacturing systems, Progress in Natural Science 19 (3) (2009) 389–395.

[14] D. Cysarz, H. Bettermann, P. Van Leeuwen, Entropies of short binary sequences in heart period dynamics, American Journal of Physiology-Heart and Circulatory Physiology 278 (6) (2000) H2163–H2172.

[15] C. Kamath, Entropy-based algorithm to detect life threatening cardiac arrhythmias using raw electrocardiogram signals, Middle East J. Sci. Res 12 (2012) 1403–1412.

[16] D. Makowiec, A. Kaczkowska, D. Wejer, M. Żarczyńska-Buchowiecka, Z. R. Struzik, Entropic measures of complexity of short-term dynamics of nocturnal heartbeats in an aging population, Entropy 17 (3) (2015) 1253–1272.

[17] D. Q. Phung, D. Tran, W. Ma, P. Nguyen, T. Pham, Using shannon entropy as eeg signal feature for fast person identification., in: ESANN, Vol. 4, 2014, pp. 413–418.

[18] C. Gómez, R. Hornero, Entropy and complexity analyses in alzheimers disease: an meg study, The open biomedical engineering journal 4 (2010) 223.
[19] C. E. Shannon, A mathematical theory of communication, Bell system technical journal 27 (3) (1948) 379–423.

[20] D. J. MacKay, Information theory, inference and learning algorithms, Cambridge university press, 2003.

[21] J. L. Semmlow, B. Griffel, Biosignal and medical image processing, CRC press, 2014.

[22] H. W. Lilliefors, On the kolmogorov-smirnov test for normality with mean and variance unknown, Journal of the American statistical Association 62 (318) (1967) 399–402.

[23] T. W. Anderson, D. A. Darling, Asymptotic theory of certain "goodness of fit" criteria based on stochastic processes, The annals of mathematical statistics (1952) 193–212.