Review and Testing of Entropy Metrics Applied to surface Electromyography

By: Yiyang Shi
Supervisor: D MacIsaac

EE6000 Report
Fall 2014

Contents

Review and Testing of Entropy Metrics Applied to surface Electromyography

By: Yiyang Shi

1 Introduction

Biomedical signals are highly complicated and usually difficult to interpret directly. As the complexity of such signals may vary with changes in biological status, entropy can be a useful tool for interpretation. Entropy has already been applied in biosignal processing to the assessment of depth of sedation [1], and the measurement of the electroencephalographic effects of desflurane [2]. Statistics derived from the entropy in dynamic systems are becoming popular in signal processing and are being recommended as powerful tools in biomedical signal analysis. Such statistical metrics, including Approximate Entropy, Sample Entropy and Fuzzy Entropy, have been applied to various clinical datasets including respiratory patterns, heart rate variability, hormone pulsatility, electromyography (EMG), electrocardiography (ECG) and electroencephalography (EEG) [3] [4] [5] [6] [7] [8].

This paper reviews Shannon Entropy, Approximate Entropy, Sample Entropy and Fuzzy Entropy, to determine their efficacy for application to the surface EMG (sEMG) signal. Each metric's characteristics were examined, including the impact of internal parameter selection on each metric. Then these entropy metrics were implemented on simulated sEMG to examine metric consistency when applied to sEMG records with roughly the same complexity.

2 Entropy Estimation

2.1 Shannon Entropy

Probably the most intuitive way to examine entropy is by introducing it from the perspective of information theory. This was first done comprehensively by Shannon, in his 1948 technical communication, "A Mathematical Theory of Communication" [9]. Shannon introduced the concept of

'information size' in terms of bits required to represent a random variable in a data stream. He used the term 'entropy' to represent this 'size' and noted that it was inherently related to the uncertainty of the data being streamed as described in the equations below:

X is completely deterministic $\longleftrightarrow$ X is completely random

$$H_0(X) = 0 \quad (1a) \qquad H(X) = \sum_{i=1}^{N} p_i \log_2\left(\frac{1}{p_i}\right) \quad (1b) \qquad H_N(X) = \log_2(N) \quad (1c)$$

In equations 1a-c, X represents the random variable being streamed which can take on N possible values and $H(X)$ represents entropy. When the data being streamed is completely certain, its entropy is considered to be 0. When the data being streamed is completely random, that is all possible values for a sample are equally probable, its entropy is given by the number of bits required to represent all the values in binary[1] (recall that n bits can encode $2^n = N$ distinct values). The interesting cases take place when the probabilities of each possible value are not equal. When this is true, we know something about data stream and this allows us to represent it with less bits. Its uncertainty decreases, and so does its information size. To quantify this, we can use equation (1b), where p_i represents the probability of occurrence of the i^{th} possibility. Intuitively, we can think of possibilities that occur regularly as low information occurrences (they happen all the time). Contrastingly, possibilities that occur rarely are high information occurrences (when they happen, it is important). Additionally, the high information occurrences don't balance out the low information occurrences because they occur less often in comparison. Thus random variables with unbalanced probabilities are less uncertain yielding lower entropy, and lower information sizes. In this paper, the normalized form of $H(X)$ was adopted, known as Shannon's Relative Entropy (ShEn):

$$ShEn = \frac{H(X)}{H_N(X)} \quad (2)$$

[1] Entropy is described here in binary, but the concepts can extend to b-ary representation (in our case substituting b=2)

This normalized form gives the amount of information as a ratio of the highest information a random variable of N possible values can carry, thus ranges from 0 to 1.

Typically for a given dataset of length L, the probability density function is unknown and thus needs to be estimated. The simplest way of estimating is to first list all unique values within this dataset, then count the times that each value appears within the data set, and then divide the counts with the total number of samples in the dataset. The resulting quotients can be used as the probabilities for each value and the estimate of Entropy provides the average number of bits required to represent a sample (or the average level of uncertainty in a sample). However, this estimation biases every probability by a leading order of $-\frac{N-1}{2L}$ [10], which can be significant for small L and thus result in inaccurate ShEn.

2.2 Approximate Entropy

Approximate Entropy (ApEn) [11] is actually an approximation of a measure of *complexity*, first introduced by Kolmogorov when he introduced 'Complexity Theory' in 1965 [12]. Kolmogorov's aim was similar to Shannon's in that they were both trying to provide a means for 'measuring the size of information' [13], but Kolmogorov used a non-probabilistic approach, and focused on the message (a sequence in a data stream) itself, rather than its source (the random variable X, in Shannon's Theory). Kolmogorov's notion of information size takes into consideration the regularity of data within a sequence and complexity is a measure of this regularity – Complexity increases as regularity decreases (or as the sequence becomes disorderly, which is often described as increasing entropy). Intuitively, to consider complexity in this way, it is useful to consider two example messages as shown in Figure 1:

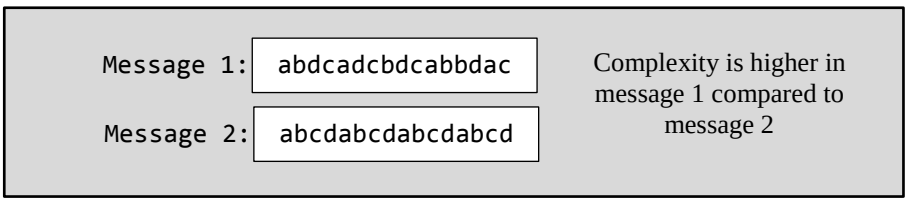

Figure 1: Example Comparing Complexities in Simple Messages

Based on Kolmogorov's work, the Kolmogorov-Sinai (K-S) Entropy was introduced by Sinai [13]. It basically measures the regularity between subsequences in a message. As regularity increases, K-S Entropy decreases. One of the widely-used algorithms for the calculation of the K-S Entropy was given by Eckmann and Ruelle and is called the E-R Entropy [14]. This calculation is summarised below.

Consider a dataset of L contiguous samples $\{u(n): 1 \le n \le L\}$ as depicted in the example in Figure 2.

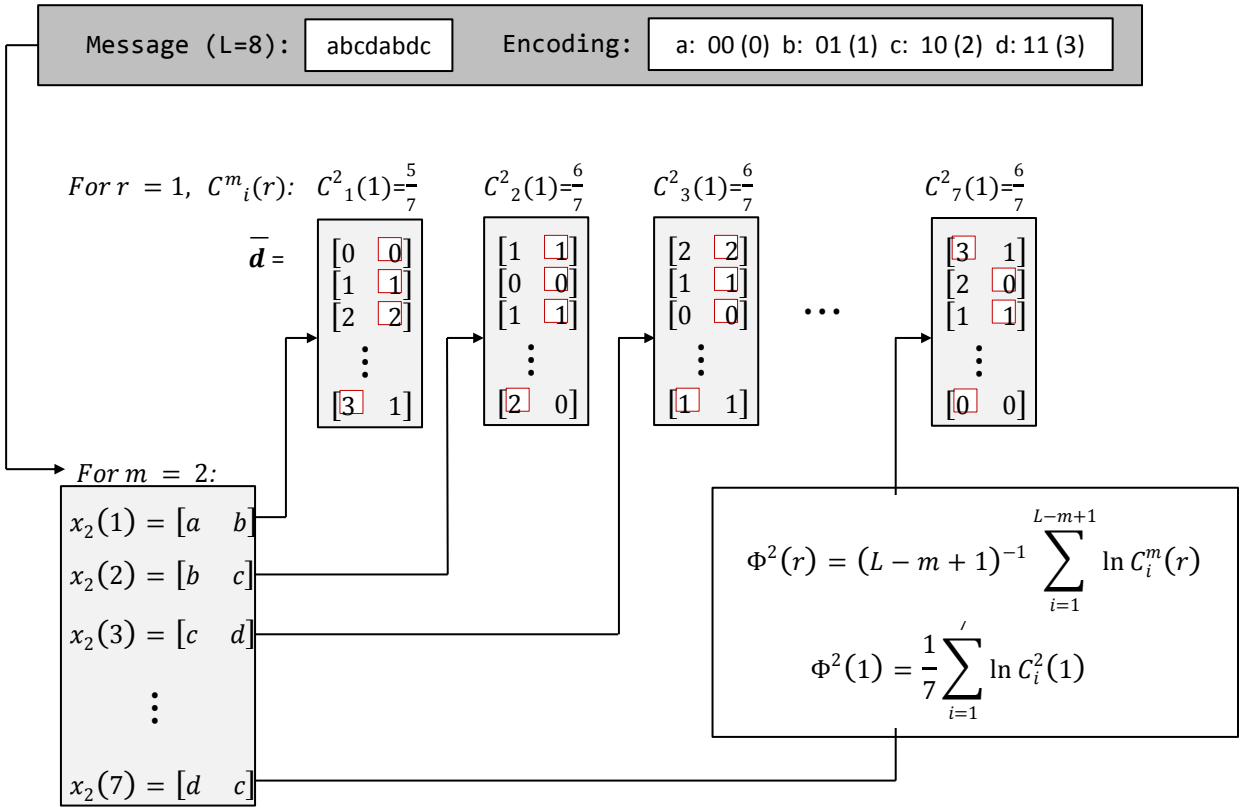

Figure 2: Example of the foundation of Approximate Entropy Calculations, bease of E-R Entropy

Given a positive integer m representing the compared window length (also known as the embedding dimension), form a set of subsequences:

$$\boldsymbol{x}_m(i), 1 \le i \le L - m + 1 \qquad (3)$$

where $\boldsymbol{x}_m(i) = \{u(i), u(i+1), \ldots, u(i+m-1)\}$. This creates $L - m + 1$ subsequences, each containing m contiguous data points.

Define the distance between two subsequences as:

$$d[\boldsymbol{x}_m(i), \boldsymbol{x}_m(j)] = \max\{|u(i+k) - u(j+k)|\}, 0 \le k \le m-1 \qquad (4)$$

This represents the maximum difference between any two corresponding elements in the subsequences, indicated in the figure as the circled elements. Now, define $C_i^m(r)$ to be the number of matching subsequences normalized to the total number of subsequences:

$$C_i^m(r) = \frac{\text{number of } \boldsymbol{x}_m(j) \text{ such that } d[\boldsymbol{x}_m(i), \boldsymbol{x}_m(j)] \le r}{L-m+1} \quad 1 \le j \le L-m+1 \qquad (5)$$

In ((5), matching subsequences are defined as those whose distances are less than a fixed threshold r. In the calculation of $C_i^m(r)$, subsequence $\boldsymbol{x}_m(i)$ is called the template, the cases such that $d[\boldsymbol{x}_m(i), \boldsymbol{x}_m(j)] \le r$ are called template matches, and r is called the filtering dimension. Finally, the function:

$$\Phi^m(r) = (L-m+1)^{-1} \sum_{i=1}^{L-m+1} \ln C_i^m(r) \qquad (6)$$

can be calculated to estimate the average natural logarithm of the probability that two subsequences of length m are similar. Repeating this with $m+1$ instead of m provides the E-R Entropy:

$$\text{E} - \text{R Entropy} = \lim_{r \to 0} \lim_{m \to \infty} \lim_{L \to \infty} [\Phi^m(r) - \Phi^{m+1}(r)] \qquad (7)$$

As indicated in equation ((7), the E-R Entropy requires a set of infinitely long data, which is impractical. Also, an r that approaches zero implies that any noise may cause two identical pattern to be considered not identical and thus results in inaccurate results. To overcome these impracticalities, Approximate Entropy fixes the values m and r to some reasonable choice to approximate E-R Entropy. With the parameters fixed, ApEn is defined as:

$$\mathrm{ApEn}(m,r) = \lim_{L\to\infty}[\Phi^m(r) - \Phi^{m+1}(r)] \quad (8)$$

For a given dataset of L points, ApEn is implemented by defining the statistic:

$$\mathrm{ApEn}(m,r,L) = \Phi^m(r) - \Phi^{m+1}(r) \quad (9)$$

The calculation of ApEn requires pre-determined values of parameters m and r. As concluded by Pincus and Goldberger [15], large values of m allows the joint probabilistic dynamics of the process to be reconstructed more strictly, and a data length of at least 10^m, or preferably 20^m is needed. Moreover, r is recommended to be a specific percentage of the standard deviation (SD) of the subject dataset. A frequent choice of these two parameters is $m = 2$ and $0.1\ \mathrm{SD} \leq r \leq 0.2\ \mathrm{SD}$.

The ApEn is approximately the negative average natural logarithm of the conditional probability that if two subsequences of length m are similar, the subsequences of length $m + 1$ from the same dataset will remain similar. It inherited the ability of reflecting system complexity by using measured data from the Kolmogorov complexity. However, this statistic is biased by self-matching - that is, the cases where $j = i$ in equation (5) are included. Removing the self-matches is not trivial, since it would introduce the possible problem caused by calculating the natural logarithm of zero. Self-matching causes two major limitations. First, the value of ApEn is highly dependent on the length of dataset and decreases uniformly with the decreasing of the data length. Thus, the effects of self-matching become more influential with smaller datasets. Second, it negatively influences relative consistency. That is, if the ApEn of one dataset is higher than that of another dataset, ApEn is supposed to remain higher when the parameters (m , r)change. In general ApEn does not, and self-matching makes worsens this inconsistency.

The lack of consistency mainly happens when there are changes in parameter r. This limitation was discussed in depth by Chon et al. [16]. They demonstrated that ApEn of a signal could be either higher or lower than another signal as r changes, no matter if the other signal has higher complexity or not. To

accommodate this, the concept of maximum ApEn was introduced, which is the maximum value of ApEn with respect to the change of r. Instead of calculating ApEn for a series of r values, Chon et al. also introduced an approach for calculating a common value $r_{\max}$ to estimate maximum ApEn, based on a relationship between the ratio of short-term and long-term variability.

2.3 Sample Entropy

Since ApEn is highly dependent on the length of dataset and lacks relative consistency, Richman and Moorman developed Sample Entropy (SampEn) [17] to improve on it. It is different from the ApEn mainly in two ways. First, self-matches are excluded. This cancels the bias caused by self-matching in ApEn. However, this may result in a zero count of matches which would introduce $ln(0)$ into the intermediate calculations of ApEn. This is dealt with by taking the logarithm only at the last step of calculation. Second, the template-wise approach is not employed to estimate the conditional probabilities.

The first step of SampEn calculation is to form a set of subsequences:

$$\boldsymbol{x}_m(i) = \{u(i), u(i+1), \dots, u(i+m-1)\}, 1 \le i \le L-m \quad (10)$$

Note that there are $L-m$ subsequences formed instead of $L-m+1$ as in the calculation of ApEn, in order to ensure both $\boldsymbol{x}_m(i)$ and $\boldsymbol{x}_{m+1}(i)$ are defined for $1 \le i \le L-m$. Then $B_i^m(r)$ and $A_i^m(r)$ are defined as:

$$B_i^m(r) = \frac{\text{number of } \boldsymbol{x}_m(j) \text{ such that } d[\boldsymbol{x}_m(i), \boldsymbol{x}_m(j)] \le r}{L-m-1} \quad 1 \le j \le L-m, j \ne i \quad (11)$$

$$A_i^m(r) = \frac{\text{number of } \boldsymbol{x}_{m+1}(j) \text{ such that } d[\boldsymbol{x}_{m+1}(i), \boldsymbol{x}_{m+1}(j)] \le r}{L-m-1} \quad 1 \le j \le L-m, j \ne i \quad (12)$$

In the above two equations, self-matches are not counted. Then the averages of template matches for m and $m+1$ are defined as:

$$B^m(r) = (L-m)^{-1} \sum_{i=1}^{L-m} B_i^m(r) \quad (13)$$

$$A^m(r) = (L-m)^{-1} \sum_{i=1}^{L-m} A_i^m(r) \quad (14)$$

and finally the SampEn can be estimated by following statistic:

$$\text{SampEn}(m, r, L) = -\ln \frac{A^m(r)}{B^m(r)} \quad (15)$$

SampEn shows better relative consistency and less dependent on the length of datasets compared to ApEn [17].

2.4 Fuzzy Entropy

Developed by Chen et al. [18], Fuzzy Entropy (FuzzyEn) utilises the concept of "fuzzy sets" proposed by Zadeh [19]. In an ordinary set, the membership of its elements is definite. However, in a fuzzy set, every element is allocated with a degree of membership, i.e. a factor describes how likely this element may belong to the fuzzy set. Given the details in the previous sections, it is clear that both ApEn and SampEn determine the similarity of subsequences based on the Heaviside functions. That is, for any two subsequences, they can only be considered as absolutely matched or unmatched based on a pre-filter-level r. In FuzzyEn, the degree of match between two subsequences is computed by an exponential function, which allows the similarity to have continuous values, and has maximum value when two subsequences are identical.

Like SampEn, the first step of FuzzyEn is to form subsequences:

$$\boldsymbol{x}_m(i) = \{u(i), u(i+1), \dots, u(i+m-1)\} - u_0(i), 1 \le i \le L-m \quad (16)$$

It is worth noting that the baseline of each subsequence $u_0(i)$ is removed so that the comparison is based on the shape of two subsequences rather than their absolute values. The baseline is defined as:

$$u_0(i) = \frac{1}{m}\sum_{j=0}^{m-1} u(i+j) \quad (17)$$

Given the removal of baseline, the distance between two subsequences is now defined as:

$$d_{ij}^m = d[\boldsymbol{x}_m(i), \boldsymbol{x}_m(j)] = \max\{|[u(i+k) - u_0(i)] - [u(j+k) - u_0(j)]|\}, 0 \le k \le m-1 \quad (18)$$

The next step is where the most significant difference compared to SampEn exists. The similarity of subsequences in FuzzyEn is not determined by Heaviside function. Instead, an exponential function is employed for calculating the continuous similarity degree:

$$D_{ij}^m(n, r) = \exp\left[-\frac{\left(d_{ij}^m\right)^n}{r}\right] \quad (19)$$

In (19), n and r determine the shape of the exponential function. To be exact, the represent the position and gradient of its boundary respectively as depicted in Figure 3.

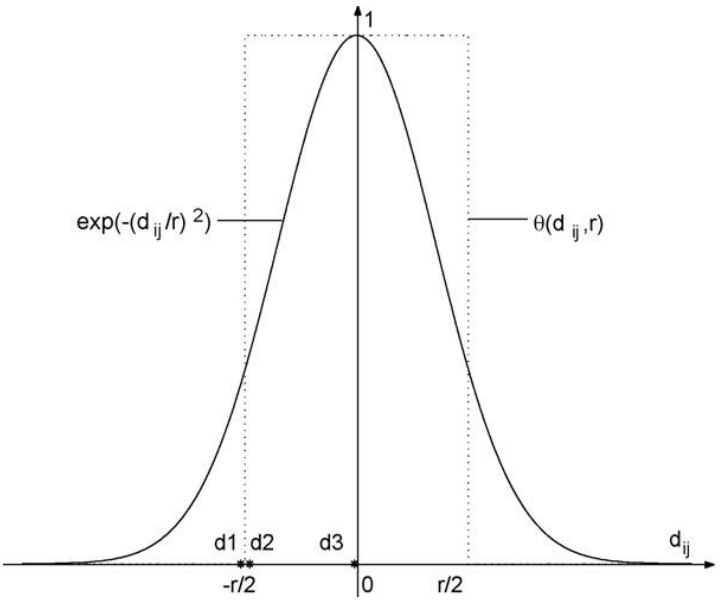

Figure 3: Plot of Heaviside function and exponential function $\boldsymbol{\theta(d[x_m(i), x_m(j)], r)}$ and exponential function $\boldsymbol{\exp[-\left(d_{ij}^m\right)^n/r]}$ when calculating the degree of similarity versus parameter $\boldsymbol{r}$ [20]

In ApEn and SampEn, the similarity of two subsequences is determined by Heaviside function, with a filter-level r specified. This rule can be expressed as:

$$\theta(d[\boldsymbol{x}_m(i), \boldsymbol{x}_m(j)], r) = \begin{cases} 1, & \text{if } d[\boldsymbol{x}_m(i), \boldsymbol{x}_m(j)] \leq r \\ 0, & \text{if } d[\boldsymbol{x}_m(i), \boldsymbol{x}_m(j)] > r \end{cases} \quad (20)$$

The Heaviside function has a rigid boundary. That is, all distances less than r have the same contribution to the membership of similarity, while all distances larger than r have no contribution. However, this may lead to completely different contributions for two highly similar subsequences while two subsequences with apparently different distances may have the same contributions.

Take for example, the contributions of three distances d1, d2 and d3 in Figure 3. The figure shows a scenario in which d1 has no contribution according to the Heaviside approach, while a distance d2 that is very close to d1 has maximum contribution, and d3 has the same maximum contribution as d1 has although the difference between them is obviously larger than that between d1 and d2. The rigid boundary causes both ApEn and SampEn to be highly sensitive to the change of r. A small change in parameter r can possibly lead to a significant change in the result.

In practical scenarios, processes are often full of imprecision and uncertainty, so the boundary of a set and the membership of an element should not be considered rigid. The concept of "fuzzy sets" proposed by Zadeh in [19] introduced "membership degree", which has the ability to handle this kind of membership by introducing a degree of membership that describes how likely a pattern may belong to a certain set. The exponential function as delineated in equation 19 employed in the calculation of similarity is a simple fuzzy function. It calculates the degree that a distance can be considered as a member of class "similarity", and then the contributions of d1, d2 and d3 in Figure 3 become more logical. Using the exponential function, the contribution of each distance changes continuously when r changes. As a result, a minor change in r will not have huge impact on the result of FuzzyEn.

The remaining steps for calculating FuzzyEn are similar to those of SampEn, which involve taking the average of the average probability that one subsequence is similar to the other ones, as shown below:

$$\Phi^m(n,r) = (L-m)^{-1}\sum_{i=1}^{L-m}\left[(L-m-1)^{-1}\sum_{j=1,j\neq i}^{L-m} D_{ij}^m\right] \quad (21)$$

Like SampEn, FuzzyEn excludes self-matches. Finally, the FuzzyEn can be estimated by the statistic:

$$\text{FuzzyEn}(m,n,r,L) = \ln\Phi^m(n,r) - \ln\Phi^{m+1}(n,r) \quad (22)$$

which is equivalent to:

$$\text{FuzzyEn}(m,n,r,L) = -\ln\frac{\Phi^{m+1}(n,r)}{\Phi^m(n,r)} \quad (23)$$

Chen et al. [18] [20], experimented with FuzzyEn to various data streams including (1) independent, identically distributed (i.i.d) random numbers of different lengths, (2) sinusoidal signals of different frequencies, (3) MIX processes, (4) the distinguishing of Logistic systems contaminated by different noise levels, and (5) the characterization of sEMG signal records of various forearm actions. They demonstrated that FuzzyEn possesses stronger relative consistency, less dependence on data length, freer parameter selection, and better tolerance to noise.

3 Entropy applied to sEMG

The purpose of this investigation was to explore the potential of entropy for use in sEMG assessment. As a first step towards this exploration, the reliability of entropy metrics was examined. Reliability in this context means test-retest repeatability. If we are to use changes in entropy as a means for interpreting changes in the state of sEMG, the entropy metric we choose must be relatively repeatable to across muscles in the same state.

3.1 Generation of sEMG Records

The generation of sEMG records was achieved using the simulation tool Myosim, which was developed by MacIsaac et al. [21], and made available from the Institute of Biomedical Engineering, University of New Brunswick.

Myosim is a matlab tool which implements a model based on the one proposed by Gonzalex-Cueto and Parker [22], in which the generation of a single fibre action potential (SFAP) measured at skin surface is described as the convolution of a source with a tissue filter applied. As suggested by Plonsey [23], a bi-directional propagating double-layar differential source is adopted. The filtering function employed in Myosim is based on the work of Dimitriv and Dimitrova [24], with field distribution properties, conduction velocity, depth of fibre, location of innervation point, and location of left and right terminations of fibre with respect to electrode location taken into consideration. The geometric parameters are visually depicted in Figure 4:

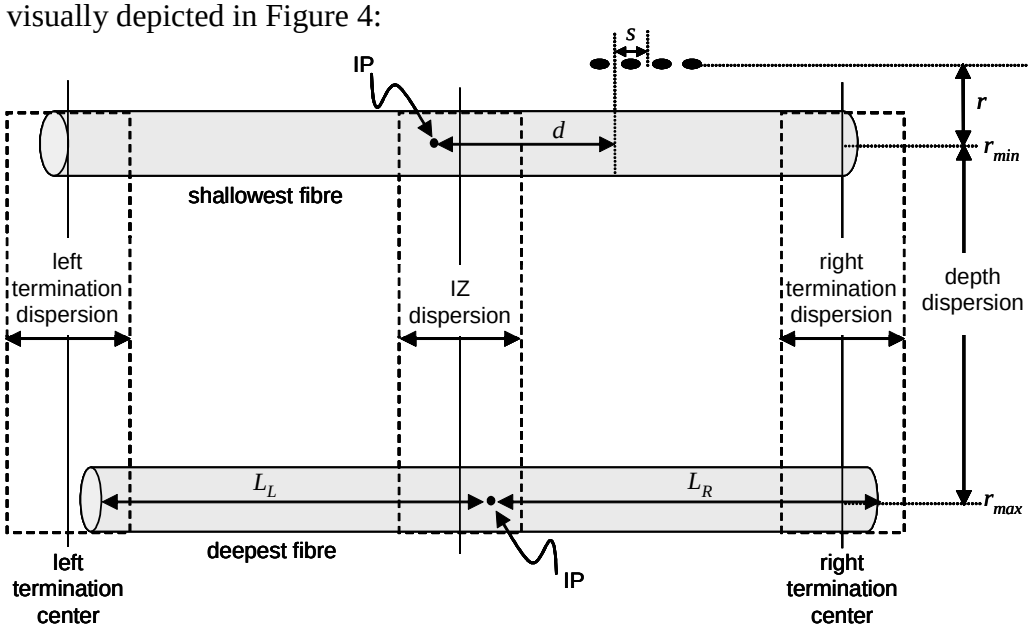

Figure 4: Parameters used in modeling sEMG: (IP: motor neuron inervation point, IZ: innervation zone defining the limits of IP, d: channel distance from innervation point, s: electrode spacing, r: fibre depth, [r_{min}, r_{max}]: depth dispersion boundaries defining the limits of r, L_L: fibre length from proximal end to innervation point (bounded by left termination dispersion), L_R: fibre length from distal end to innervation point (bounded by right termination dispersion)

Once the SFAPs are generated, the motor unit action potentials measured at the skin surface are generated by summing multiple SFAPs. Since the signal generated by a specific motor unit is identical

during each innervation process, an MUAP train can be formed by convolving the same MUAP with a series of impulses, which simulates the timing of innervation process. Finally, to generate the sEMG record, a number of independently generated MUAP trains are summed. Figure 5 shows a dialog box provided in Myosim for setting all necessary parameters to simulate a sEMG recording.

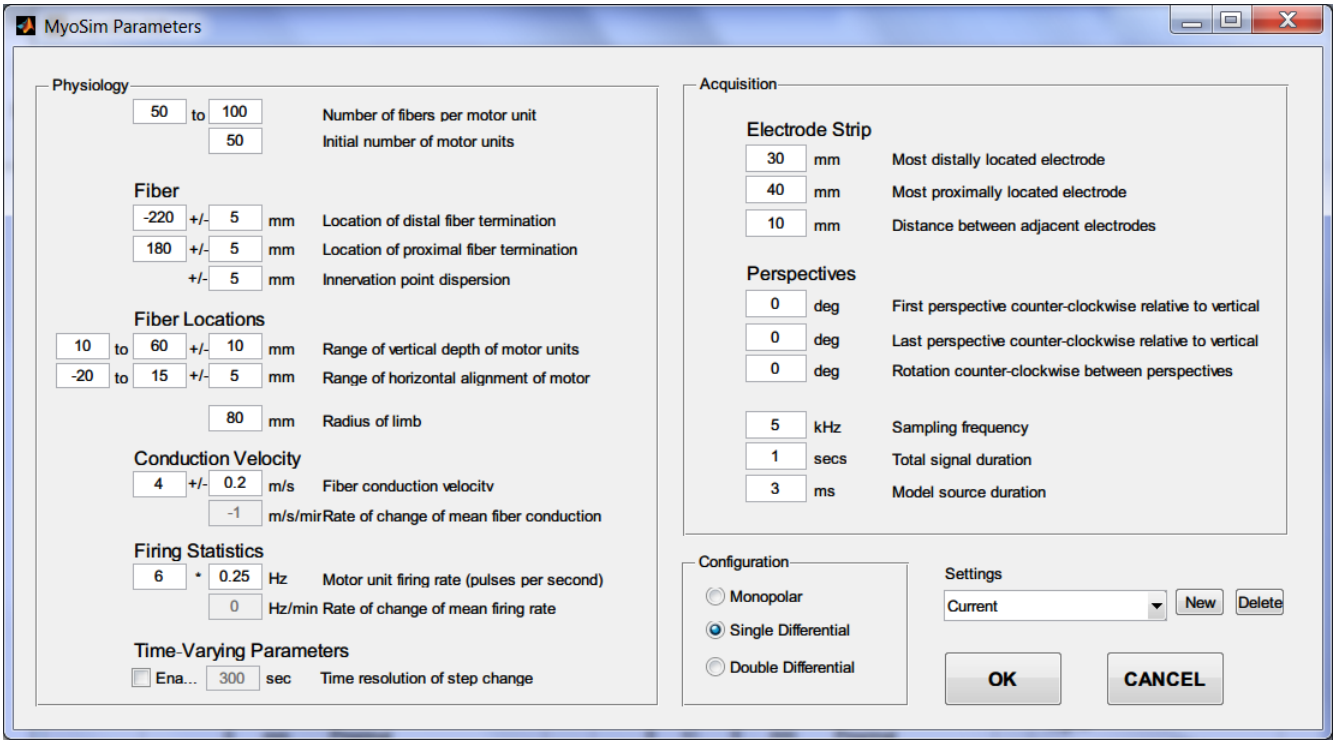

Figure 5: Screenshot of Myosim Parameters for sEMG Generation

Fifty sEMG records simulating contractions of the brachial biceps (muscle in the upper arm) were generated for testing the performance of the entropy metrics under investigation. Model parameters used for generating the simulated recordings are shown in Figure 5. The physiology parameters illustrated in the left-half of Figure 5 are typical parameters for a brachial bicep [25] [26]. All physiology parameters were allowed to vary within a specified range, thus the 50 sEMG records were treated as records of 50 different participants, each of them having similar but unique physiology parameters.

As illustrated in the right-half of Figure 5, a single differential (or bipolar) configuration was adopted for the electrodes, which were placed at 30mm and 40mm respectively, measured from the mid-point of innervation zone. By placing the electrode channel at this location, end-effects were minimized and the possibility of electrodes being placed over innervation zone was avoided. Simulated recordings were meant to be taken from the perspective of the surface of the bicep, so the perspective parameters were set to simulate placing the measuring channel at 0 degrees. Sampling frequency and signal duration were set to 5kHz and 1s respectively. As a result, each generated record has a data length of $L = 5000$.

Figure 6 and Figure 7 depict records 28 and 30. These figures demonstrate the expected similarities and differences between simulated signals with parameters set as described. Note that the shapes and amplitudes may vary heavily from record to record.

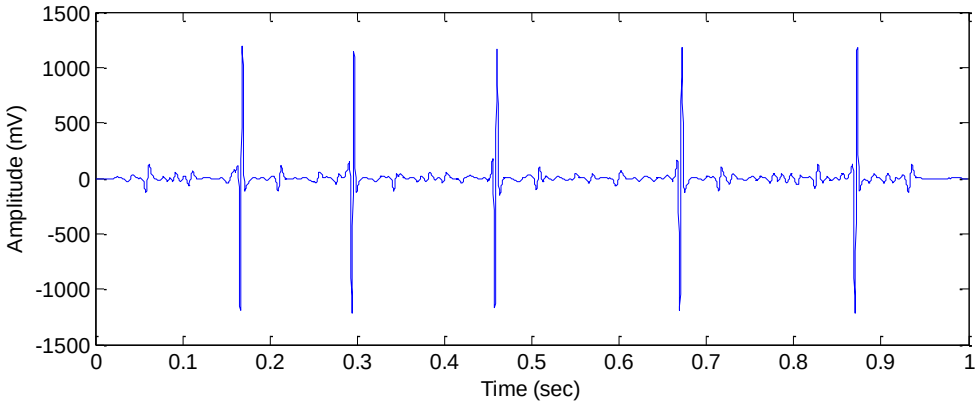

Figure 6: Plot of sEMG Record 28

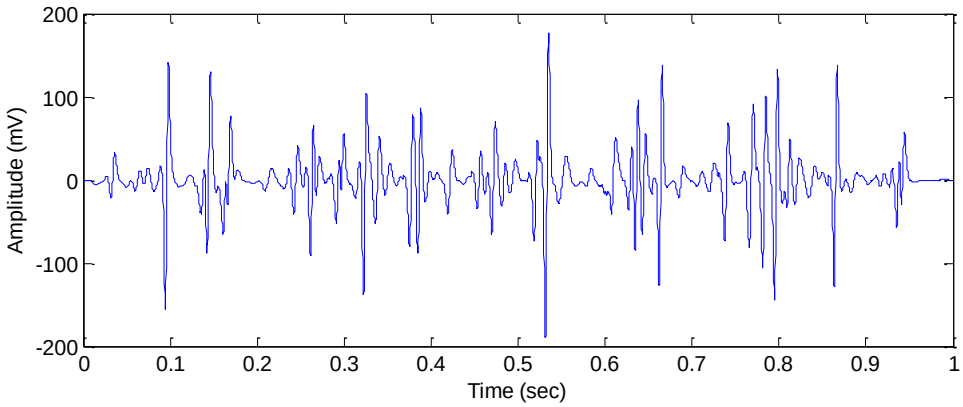

Figure 7: Plot of sEMG Record 30

Records generated by Myosim have data points of continuous magnitudes. However, practical recorded data is usually available in digital form which has been quantised into a particular resolution. Since a resolution of 12-bit is commonly used for sEMG, the data was quantised before any entropy metrics were applied with a resolution of 12-bit. This was achieved by pre-defining a list of $(2^{12} - 1)$ levels uniformly distributed between the limits of the full-scale range and then rounding each sample to the closest value in this list. The full-scale range was considered as the full dynamic range of a specific sEMG record, which simulates the common practice of setting gains in the acquisition instrumentation in a way that the sEMG can vary within its complete dynamic range to improve the signal-to-noise ratio.

3.2 Performance of the Entropy Metrics

Although each of the 50 sEMG records used in this work had unique parameters, they were all records modeling the brachial biceps, thus the amount of information contained in the records should not differ largely from each other. That is, the results among different records should be relatively repeatable. Thus, an entropy metric can be considered reliable if the variance across the 50 records is low. In this section, a comparison of ShEn, ApEn, SampEn and FuzzyEn applied to the 50 sEMG records is presented.

3.2.1 Preliminaries

Table 1 lists the parameters required to compute each of the entropy metrics together with the default values used in the estimations for each metric.

ShEn	---	
ApEn SampEn FuzzyEn	embedding dimension	$m = 2$
	filter-level/boundary position†	$r = 0.4$
	record length	$L = 5000$
	boundary gradient†	$n = 2$

†FuzzyEn only

Table 1: Default Parameter Values used in Investigation

As concluded in [15], the choice of m should ensure that the data length is greater than 10^m. As the length of sEMG records is 5000, the value of m cannot exceed 3 since 5000 is less than 10^4. However, preliminary studies indicated that performance achieved when $m = 3$ does not show significant improvement though the computation time was doubled compared to when $m = 2$. Thus, the value of m was chosen to be 2 for the purpose of this exploration. Parameter r (and n for FuzzyEn) were selected through trial and error.

In addition to applying ApEn with a constant value of parameter r, the concept of ApEn_{max} introduced by Chon et al. in [16] was tested. Values for ApEn_{max} were obtained using two methods: the first one simply calculated ApEn across the 50 sEMG records with incrementing values of r and then for each sEMG chose the result (ApEn_{max}) from the r-value which produced the maximum result. The other method chose r by calculating an r_{max} and used that as the value in the ApEn_{max} calculation for each sEMG. The r_{max} value was calculated utilising the variance of differences between all two adjacent elements of the dataset (SD_1, short-term variability), the variance of the whole dataset SD_2 (long-term variability) and the length of dataset L. The r_{max} is calculated with

$$r = \frac{(-0.036 + 0.26\sqrt{\text{SD1}/\text{SD2}})}{\sqrt[4]{N/1000}} \text{ when } m = 2 \qquad \textbf{(24)}$$

3.3 Results

Figure 7(a) shows variance in entropy metrics ApEn, SampEn, FuzzyEn across all 50 records for different values of r. Since the variance of FuzzyEn did not decrease significantly when $r > 0.4$, $r = 0.4$ was chosen for FuzzyEn. The value of $r = 0.4$ for both ApEn and SampEn was also chosen. According to the results, continuing to increase r can further decrease the variance, but this is not preferred since the larger r gets, the more chance we have of mistakenly identifying two subsequences as a match. Using the same line of reasoning and the results depicted in Figure 7(b), $n = 2$ was set.

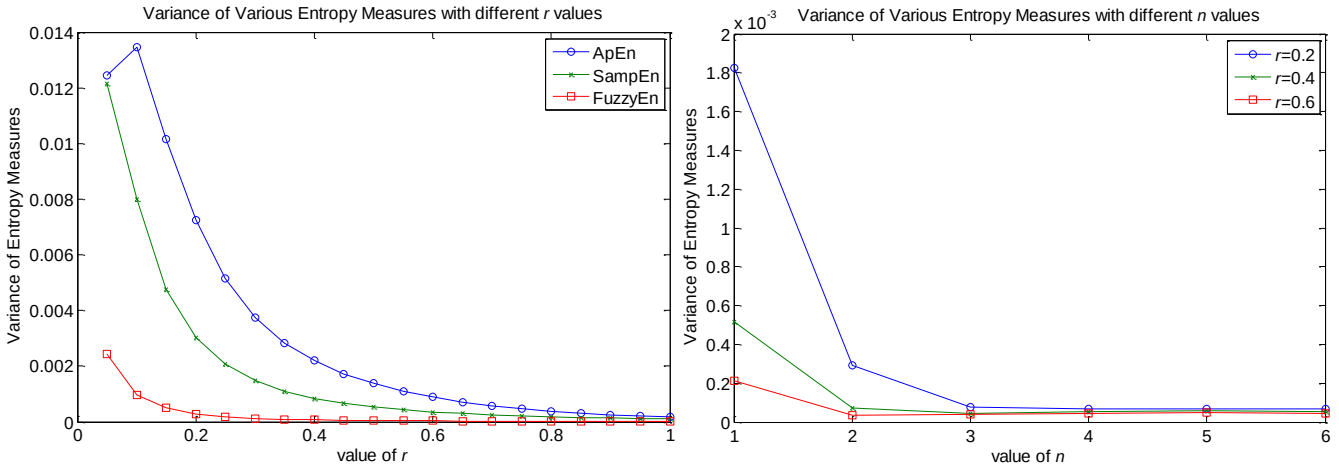

Figure 8: Variance of ApEn, SampEn and FuzzyEn applied to the 50 sEMG records with various n and r

Table 2 gives the means, variances and coefficient of variation across the 50 records for all the entropy metrics investigated. The coefficient of variation is a normalised measure of variation. This normalisation made it possible to compare metrics. In this sense, ShEn provides best performance, followed closely by ApEn_{max} estimated with r varying across sEMG.

Entropy Measures / Statistics	ShEn	ApEn (2, 0.4, 5000)	Maximum ApEn (various r) (2, r, 5000)	Maximum ApEn (calculated r) (2, r, 5000)	SampEn (2, 0.4, 5000)	FuzzyEn (2, 2, 0.4, 5000)
Mean	0.9033	0.1327	0.5324	0.4553	0.0590	0.0409
Variance	0.0028	0.0022	0.0018	0.0141	8.4804×10^{-04}	7.1731×10^{-5}
Coefficient of Variation	0.0583	0.3537	0.0803	0.2607	0.4934	0.2069

Table 2: Means, Variance and Coefficient of Variantion of Investigated Entropy Metrics when Applied to the 50 sEMG Records

Figure 6 to 11 gives the bar charts showing the results of these metrics applied to the 50 sEMG records. All of their scales range from 0 to 1 so that variation across the 50 data sets can be compared between metrics visually. The only exception is Figure 9. As a result of the baseline removal during the formation of subsequences, the values obtained with FuzzyEn is markedly lower than other entropy metrics, thus the variation among results may not be illustrated in comparison to the others.

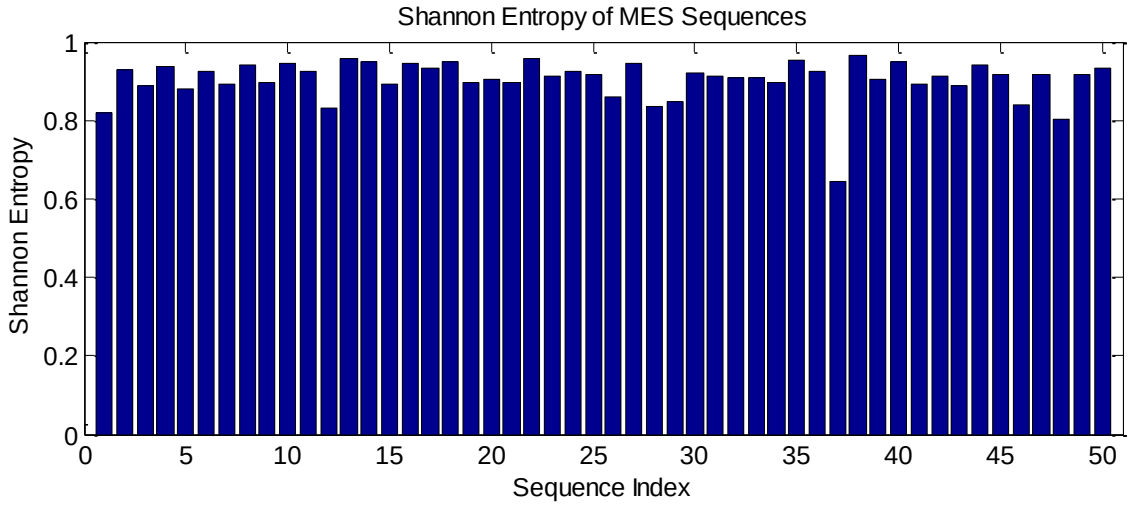

Figure 9: ShEn of the 50 sEMG Records

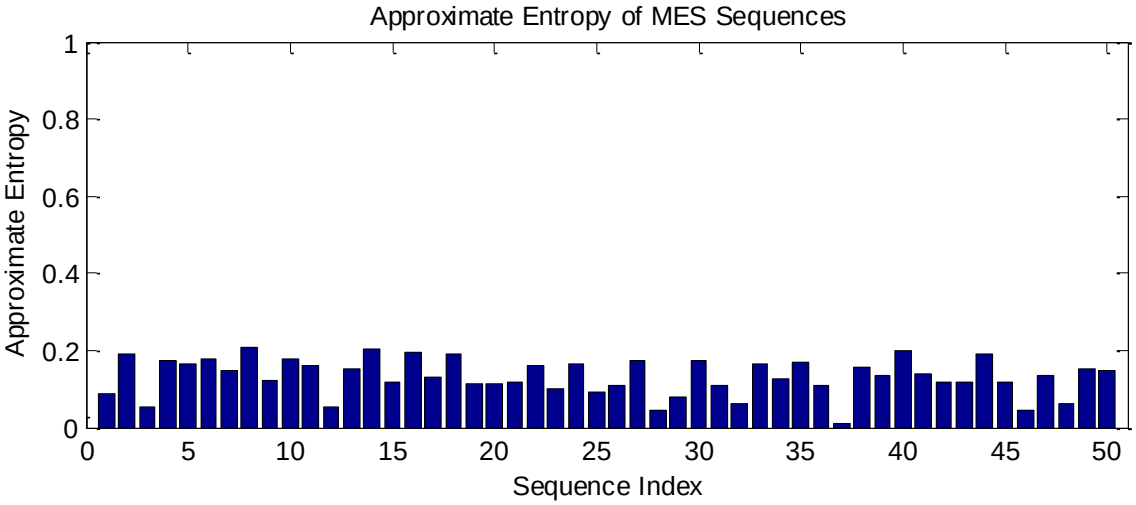

Figure 10: ApEn(2, 0.4, 5000) of the 50 sEMG Records

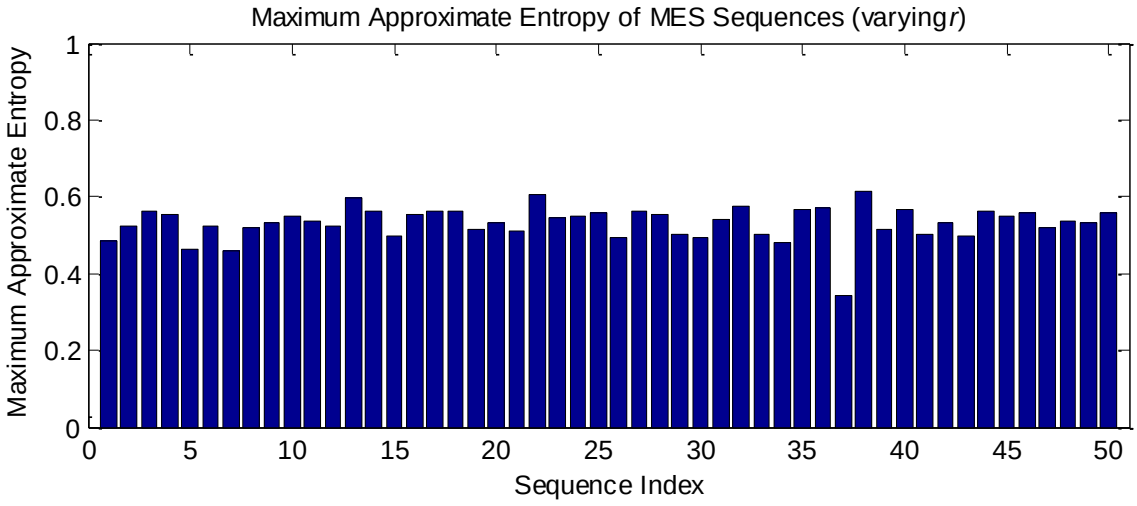

Figure 11: Maximum ApEn (varying *r*) (2, r, 5000) of the 50 sEMG Records

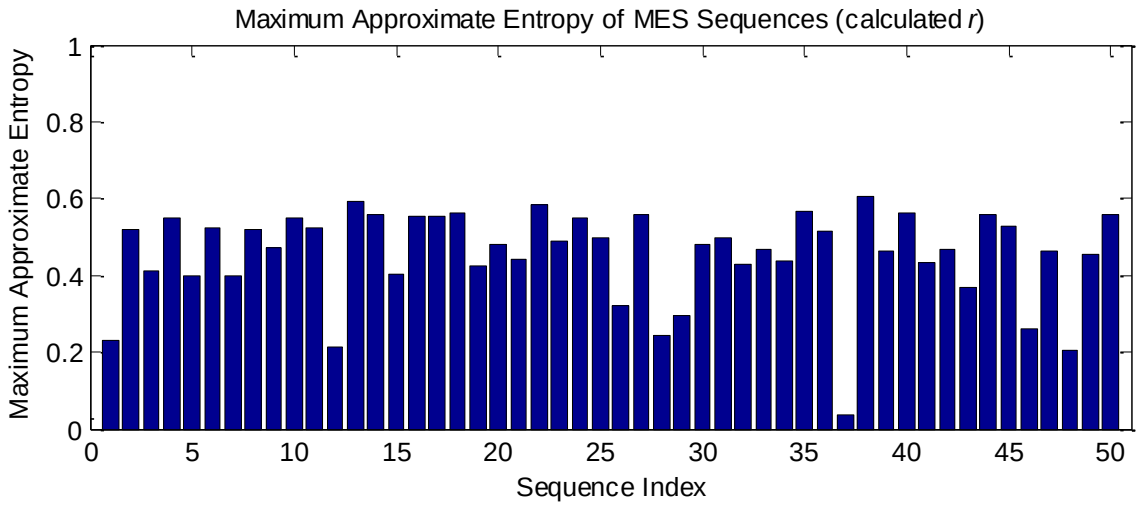

Figure 12: Maximum ApEn (calculated *r*) (2, r, 5000) of the 50 sEMG Records

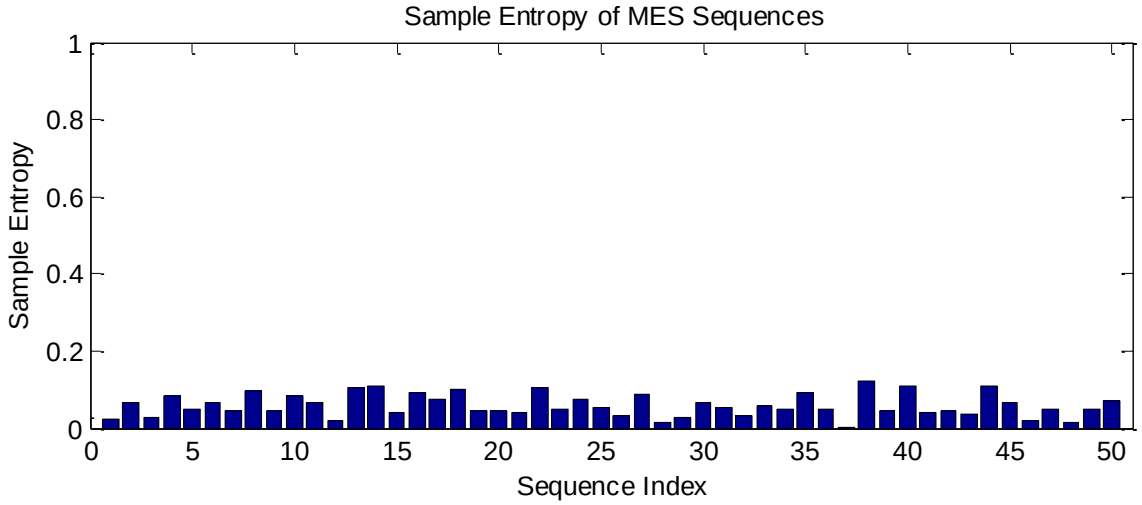

Figure 13: SampEn(2, 0.4, 5000) of the 50 sEMG Records

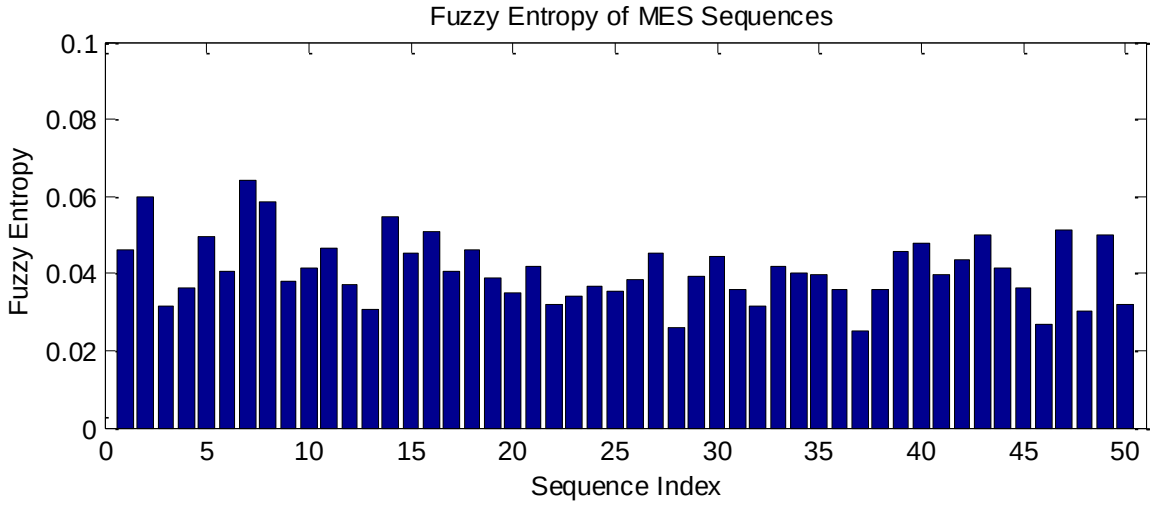

Figure 14: FuzzyEn(2, 2, 0.4, 5000) of the 50 sEMG Records

Figure 9 depicts ShEn measured in each of the 50 simulated sEMG recordings. Since the data length $L = 5000$ is relatively long, it can be inferred that the error caused by the estimation of probabilities is negligible and the values depicted correctly reflect the entropy contained in a record, interpreted as a measure of the quantity of information. The mean across these data is high ($\hat{u}_{SH} = 0.9033$ which is close to $max(u_{SH}) = 1$), indicating that the source of these signals is highly random. As expected for signals derived from the same random source, the variance of results obtained with ShEn is low.

Figure 10-Figure 12 show the results for estimating ApEn. The plots (and statistics in Table 2) show that the criteria used to choose r may have an effect on the values we measure, and their variance. Comparing Figure 10 (simple ApEn) with Figures Figure 11-Figure 12 (maximum ApEn), we see that entropy, interpreted as a measure of complexity in the data, is 3-4x lower in the simple metric. We also see that variance is about 8x higher when estimating maximum ApEn with an r calculated by formula (Figure 12) vs choosing an r which yields a maximum result (Figure 11) for each sEMG, even though their mean values are relatively similar. As a result, the latter approach to calculating ApEn hails as the most repeatable ApEn metric for estimating entropy, as confirmed by the lowest coefficient of variation (4x lower than the simple ApEn). This is not surprising since choosing an r specifically to produce a maximal result for each particular data set ensures that the mean will be relatively high and adds an extra condition to constrain the values, ensuring that the variance will be relatively low. However, searching for $r_{\max}$ is very time-consuming. Using 0.05 as the searching step and searching from 0.05 to 1 increases computational time by a factor of 20 when compared to computing simple ApEn.

SampEn and ApEn are both measures of complexity, but their approach to estimating complexity differs so a direct comparison between means cannot be made. Typically, these estimators are compared in terms of their sensitivity to changes in their own internal parameters (m and r) and data length (L). SampEn generally performs better in this regard as evidenced by the work of Richman et al [17], but this kind of comparison was beyond the scope of this work. Instead, the focus here was on repeatability

across data streams of similar complexity. The results depicted in Table 2 indicate that SampEn and ApEn are similar in this context, when comparing coefficients of variation for SampEn with the simple ApEn metic. However, SampEn shows a reduction in performance similar to the simple ApEn metric (6x for SampEn) when comparing coefficients to that of maximum ApEn calculated with the varying r value.

Finally, in terms of relative variability as measured by the coefficient of variation, FuzzyEn performed on par with the simple ApEn and SampEn. However, a major dropback is its computation time. For the 50 sEMG records tested in this paper, FuzzyEn requires roughly twice the time required to calculate simple ApEn.

4 Conclusion & Future Works

In this paper, several entropy metrics were described and applied to sEMG. Entropy as a measure of information size was explored in the context of ShEn, and entropy as a measure of complexity was explored in the context of ApEn, SampEn, and FuzzyEn.

By applying each metric to 50 brachial bicep sEMG records generated with a simulation tool, repeatability could be compared among metrics. In doing so, ShEn and maximum ApEn emerged as the most repeatable metrics and therefore merit further investigation. Caution must be applied when considering the utility of maximum ApEn however. If the intention of applying the entropy estimator is to distinguish between states in a muscle, letting r vary between data sets may adversely affect the estimator's sensitivity to differences of interest. Thus, estimating maximum ApEn with a calculated r value, may be a better alternative when entropy as a measure of complexity is relevant.

This work represents a first step in exploring the potential of entropy for use in sEMG assessment. Results indicate that entropy can be estimated with sufficient repeatability in similar sEMG data to warrant further investigation. As a next step, its ability to distinguish between differing sEMG data will

be explored. In particular, we are interested in examining entropy for its ability to detect crosstalk. Crosstalk results from unavoidably measuring the electrical activity of surrounding muscles along with the activity from the muscle of interest. As such, it is difficult to distinguish between these electrical sources. We are hoping that entropy increases when the number of muscles contributing to the measurement increases. If this is the case, we may be able to use high entropy as an indicator for crosstalk.

Reference

[1] R. Ferenets, T. Lipping, A. Anier, V. Jäntti, S. Melto and S. Hovilehto, "Comparison of Entropy and Complexity Measures for the Assessment of Depth of Sedation," *IEEE Transactions on Biomedical Engineering,* vol. 53, no. No. 6, pp. 1067-1077, Jun. 2006.

[2] J. Bruhn, L. E. Lehmann, H. Röpcke, T. W. Bouillon and A. Hoeft, "Shannon Entropy Applied to the Measurement of the Electroencephalographic Effects of Desflurane," *Anesthesiology,* vol. 95, no. 1, pp. 30-35, Jul. 2001.

[3] D. Caldirola, L. Bellodi, A. Caumo, G. Migliarese and G. Perna, "Approximate Entropy of Respiratory Patterns in Panic Disorder," *American Journal of Psychiatry,* vol. 161, no. 1, pp. 79-87, 2004.

[4] S. M. Pincus, T. R. Cummins and G. G. Haddad, "Heart Rate Control in Normal and Aborted-SIDS Infants," *American Journal of Physiology: Regulatory Integrative and Comparative Physiology,* vol. 264, pp. 638-646, 1993.

[5] S. M. Pincus and D. L. Keefe, "Quantification of Hormone Pulsatility via an Approximate Entropy Algorithm," *American Journal of Physiology: Endocrinology and Metabolism,* vol. 262, pp. 741-754, 1992.

[6] N. Radhakrishnan, J. D. Wilson, C. Lowery, P. Murphy and H. Eswaran, "Testing for Nonlinearity of the Contraction Segments in Uterine Electromyography," *International Journal of Bifurcation and Chaos in Applied Sciences and Engineering,* vol. 10, no. 12, pp. 2785-2790, 2000.

[7] X. Ning, Y. Xu, J. Wang and X. Ma, "Approximate Entropy Analysis of Short-term HFECG Based on Wave Mode," *Physica A: Statistical Mechanics and its Applications,* vol. 346, no. 3-4, pp. 475-483, 2005.

[8] J. Bruhn, "Approximate Entropy as an Electroencephalographic Measure of Anesthetic Drug Effect during Desflurane Anesthesia," *Anesthesiology,* vol. 92, no. 3, pp. 715-726, 2000.

[9] C. E. Shannon, "A Mathematical Theory of Communication," *The Bell System Technical Journal,* vol. 27, pp. 379–423, 623-656, Jul., Oct. 1948.

[10] J. A. Bonachela, H. Hinrichsen and M. A. Muñoz, "Entropy Estimates of Small Data Sets," *Journal of Physics A: Mathematical and Theoretical,* vol. 41, no. 20 , pp. 202001-202009, 2008.

[11] S. M. Pincus, "Approximate Entropy as a Measure of System Complexity," *Proceedings of the National Academy of Sciences,* vol. 88, pp. 2297-2301, Mar. 1991.

[12] A. N. Kolomogorov, "Three approaches to the quantitative definition of information," *Problems of Information Transmission,* vol. 1, no. 1, 1965.

[13] P. Grunwald and P. Vaitanyi, "Shannon Information and Kolmogorov Complexity," Cornell University Online Library, 2010.

[14] P. Grassberger and I. Procaccia, "Estimation of the Kolmogorov Entropy from a Chaotic Signal," *Physical Review A,* vol. 28, pp. 2591-2593, 1983.

[15] S. M. Pincus and A. L. Goldberger, "Physiological time-series analysis: What does regularity quantify?," *American Journal of Physiology - Heart Circulatory Physiology,* vol. 266, p. 1643–1656, 1994.

[16] K. H. Chon, C. G. Scully and S. Lu, "Approximate Entropy for All Signals: Is the Recommended Threshold Value r Appropriate," *IEEE Engineering in Medicine and Biology Magazine,* vol. 28, no. 6, pp. 18-23, Nov./Dec. 2009.

[17] J. S. Richman and J. R. Moorman, "Physiological Time-series Analysis using Approximate and Sample Entropy," *American Journal of Physiology - Heart Circulatory Physiology,* vol. 278, p. 2039–2049, 2000.

[18] W. Chen, Z. Wang, H. Xie and W. Yu, "Characterization of Surface EMG Signal Based on Fuzzy Entropy," *IEEE Transactions on Neural Systems and Rehabilitation Engineering,* vol. 15, no. 2, pp. 266-272, Jun. 2007.

[19] L. A. Zadeh, "Fuzzy sets," *Information Control,* vol. 8, p. 338–353, 1965.

[20] W. Chen, J. Zhuang, W. Yu and Z. Wang, "Measuring Complexity using FuzzyEn, ApEn, and SampEn," *Medical Engineering & Physics,* vol. 31, pp. 61-68, 2009.

[21] D. MacIsaac, D. Rogers and A. Bhandarkar, "Simulating Myoelectric Signals with a Finite-Length Model of Muscle," Fredericton, 2007.

[22] J. A. Gonzalex-Cueto and P. A. Parker, "Deconvolution Estimation of Motor Unit Conduction Velocity Distribution," *IEEE Transactions on Biomedical Engineering,* vol. 49, no. 9, pp. 955-962, 2002.

[23] R. Plonsey, "The Active Fiber in a Volume Conductor," *IEEE Transactions on Biomedical Engineering,* vol. 12, no. 2, pp. 371-381, 1974.

[24] G. V. Dimitriv and N. A. Dimitrova, "Fundamentals of Power Spectral of Extracellular Potentials Produced by A Skeletal Muscle Fibre of Finite Length," *Medical Engineering & Physics,* vol. 20, no. 8, pp. 580-587, 1998.

[25] C. E. Clauster, J. C. McConville and J. W. Young, "Weight, Volume, and Center of Mass of Segments of the Human Body," Wright Patterson Air Force Base, Ohio, 1969.

[26] E. J. van Zulen, A. van Velzen and J. J. Denier van der Gon, "A Biomechanical Model for Flexion Torques of Human Arm Muscles as a Function of Elbow Angle," *Journal of Biomechanics,* vol. 21, no. 3, pp. 193-190, 1988.