Baseline Noise and Model Parameters in Surface Electromyography

By

Shriram Tallam Puranam Raghu

B. Tech, JNTU-H, India, 2014

A Thesis Submitted in Partial Fulfilment
of the Requirements for the degree of

Master of Science in Engineering

In the Graduate Academic Unit of Electrical and Computer Engineering

Supervisor(s): Dawn T. MacIsaac, PhD, Dept. of Electrical & Computer Engineering

Philip A. Parker, PhD, Prof. Emeritus, Dept. of Electrical & Computer Engineering

Examining Board: Julian Meng, PhD, Dept. of Electrical & Computer Engineering

Michael W. Fleming, PhD, Faculty of Computer Science

Maryhelen Stevenson, PhD, Dept. of Electrical & Computer Engineering, Chair

This thesis is accepted by the
Dean of Graduate Studies

THE UNIVERSITY OF NEW BRUNSWICK

November 2017

ABSTRACT

The Pairwise Attribute Noise Detection Algorithm has been recommended by others as a way to classify surface electromyography signals as clean or noisy. To train the algorithm, simulation can be used to generate clean examples. Used in this way, the algorithm has been shown to perform well for classifying simulated test signals, but not for in vivo SEMG records. This work investigated the poor performance with in vivo SEMG records in order to improve it, if possible. Impact of introducing instrumentation effects into simulated signals was shown to be negligible. Impacts from judicious selection of simulation parameters, including both embedded and user specified values was shown to significantly improve falsely classifying clean records as noisy. A genetic algorithm was developed to provide support for choosing user specified values and using this technique, false positive rates (i.e. classifying clean signals as noisy) decreased from $95\ \%$ to $20\ \%$, without degradation of other classification rates (i.e. classifying noisy signals as noisy, clean signals as clean).

ACKNOWLEDGEMENTS

First and foremost, I would like to express my deepest gratitude to Dr Dawn MacIsaac and Dr Phil Parker for providing me with an opportunity to participate in their research projects. Their vast knowledge, support and seemingly never ending patience in every aspect of my research has provided me with a wonderful experience. I could not have asked for anything better. I look forward to working with them again in the future!

I would also like to thank my officemate Yiyang Shi for all the days he spent collecting data from the awesome volunteers who were kind enough to participate in our experiments! I had a great time collaborating with him. I thank the faculty and staff of the Department of The Department of Electrical and Computer Engineering here at UNB, especially: Shelly, Denise and Dawne for remaining patient and understanding despite my constant barrage of questions and requests, you guys are awesome!

Last, but definitely not the least, I would also like to thank my family, especially my parents for providing me with the financial and moral support to pursue a Master's degree at UNB. A special thanks to all my friends here at UNB who have kept my spirits up throughout the journey!

Table of Contents

ABSTRACT ..... **ii**

ACKNOWLEDGEMENTS ..... **iii**

Table of Contents ..... **iv**

List of Tables ..... **vi**

List of Figures ..... **vii**

List of Symbols and Abbreviations ..... **viii**

Chapter 1: INTRODUCTION **1**

Chapter 2: SIMULATING SEMG SIGNALS **4**

2.1 Modelling SEMG signals 4

2.1-1 Signal Generation 4

2.1-2 Data Capture 14

2.2 Myosim 3.0 – A User Friendly SEMG Simulation Tool 18

2.3 A closer look at PANDA-based SEMG quality analysis 22

Chapter 3: ADDING INSTRUMENTATION EFFECTS TO MYOSIM **24**

3.1 Baseline Noise 24

3.1-1 Methods 24

3.1-2 Results 28

3.1-3 Conclusions 30

3.2 Filtering and Quantisation 31

3.3 Inspecting the effects of Instrumentation Noise on Signal Features 33

3.4 Summary 35

Chapter 4: IMPROVING MYOSIM MODEL PARAMETER SETTINGS **37**

4.1 Setting Embedded Model Parameters 37

4.2 Choosing User-defined Parameters Using a Genetic Algorithm 39

4.2-1 Setting up the Genetic Algorithm 46

4.2-2 Evaluating the Genetic Algorithm 50

4.3 Summary 60

Chapter 5: Evaluating the impact of MYOSIM upgrades on PANDA-BASED SEMG SIGNAL QUALITY ANALYSIS **61**

5.1 Methods 61

5.2 Results and Discussion 63

5.2-1 PANDA performance without noise 63
5.2-2 PANDA performance in the presence of noise 65
Chapter 6: Conclusion 68
6.1 Recommendations 69
REFERENCES 71
APPENDIX A: Description of features 75
Curriculum Vitae

List of Tables

Table 2.1: Description of SFAP parameters in MUAP modelling 12

Table 3.1: Baseline noise parameters 29

Table 3.2: Simulation Parameter values used in the Inspection of the effects of Instrumentation Noise on SEMG features 33

Table 4.1: Updated model parameters 38

Table 4.2: Tool parameters and GA search space 47

Table 4.3: Simulation Matching 55

Table 4.4: Comparison between features from Subject 4 records and all other well-behaved records. 58

Table 5.1: Cross Validation Results 64

Table 6.1: Feature description (Th=0.005) 78

List of Figures

Figure 2.1: Summary of the EMG modelling process 6
Figure 2.2: Characteristics of a single muscle fibre 7
Figure 2.3: Transmembrane potential, source, tissue filter and SFAP 8
Figure 2.4: Motor unit model parameters 11
Figure 2.5: Motor unit geometry 11
Figure 2.6: Example MUAP, MUAPT and SEMG 14
Figure 2.7: Typical SEMG signal collection set-up 15
Figure 2.8: Myosim 3.0 20
Figure 2.9: 'Myosim Parameters' window 21
Figure 2.10: PANDA with different test signals 23
Figure 3.1: Expected PSD of baseline noise 26
Figure 3.2: PSD of EMG signal and baseline noise 29
Figure 3.3: Simulated baseline noise 30
Figure 3.4: Frequency response of the band-pass filter 31
Figure 3.5: An example of Simulated Unfiltered vs Filtered SEMG 32
Figure 3.6: Effects of instrumentation on signal features 35
Figure 3.7: Myosim GUI for setting instrumentation effects 36
Figure 4.1: Example initial population 41
Figure 4.2: Stages in a typical GA 42
Figure 4.3: Roulette Wheel Selection 44
Figure 4.4: Crossover operator 45
Figure 4.5: Feature comparison between the two bands 53
Figure 4.6: Matched signal spectrum. 55
Figure 4.7: Typical distribution of simulated signal features compared to the record 57
Figure 4.8: Atypical distribution of simulated signal features compared to the record 58
Figure 4.9: Effect of channel on features 59
Figure 5.1: Overview of the process used to investigate impacts on PANDA. 63
Figure 5.2: Cross-validation results – Subject 1 65
Figure 5.3: PANDA performance at varying SNR 66

List of Symbols and Abbreviations

SEMG	Surface Electromyography
PANDA	Pairwise Attribute Noise Detection Algorithm
SFAP	Single Fibre Action Potential
MU	Motor Unit
MUAP	Motor Unit Action Potential
MUAPT/MUAPt	Motor Unit Action Potential Train
IPI	Inter-Pulse Interval
MAV	Mean Absolute Value
SSC	Slope Sign Change
ZC	Zero Crossings
WL	Wavelength/Waveform Length
EN	Entropy
SMR	Signal to Motion Artefact Ratio
MVC	Maximum Voluntary Contraction
FDR	Full Dynamic Range
AA	Anti-Aliasing
PSD	Power Spectral Density
SNR	Signal to Noise Ratio
MAE	Mean Absolute Error
FR	Frequency Ratio
MA	Motion Artefact
PLI	Power-line Interference
SAT	Amplifier Saturation

CHAPTER 1: INTRODUCTION

Surface Electromyography signals (SEMG) are signals that originate from muscles. These signals are stochastic in nature with amplitude values of about $0 - 10\,mV$ P-P [1]. Typically, the signals have a total bandwidth of about $500\,Hz$ with most of the energy contained in the $50\,Hz - 150\,Hz$ band. They can be represented by a Gaussian distribution function with zero mean [1]. SEMG signals are acquired using instrumentation that is specifically designed to detect, amplify and record these signals.

In the process of acquiring SEMG signals, noise inevitably gets introduced into the signals. Noise is any unwanted signal that contaminates the data and can make proper analysis difficult. It is not possible to obtain a noise-free signal, but its effects can be mitigated with high quality instrumentation and prudent instrumentation tuning and measurement procedures. Noise that is still present even after proper techniques have been employed can be considered as baseline noise, and is generally present to some degree in all SEMG signals. Despite the measurement challenges associated with SEMG, many SEMG-based diagnostic and assistive devices exist, for example, portable workplace SEMG-based fatigue monitors and SEMG controlled powered prosthetics [2], [3]. The presence of noise in the SEMG signals poses a challenge to varying degrees depending on the application. Several powerful techniques have been proposed to deal with noise, including Wavelet De-noising, Independent Component Analysis, Adaptive filtering, etc. [4].

To better enable SEMG-based technologies, Chan and MacIsaac initiated a project called the cleanEMG project in 2011 [5]. According to the researchers, the aim of the project is

to provide open source, user friendly methods for automated quality analysis of SEMG signals. They acknowledge that sophisticated acquisition instrumentation is available to ensure high quality data when set up properly, but also note that complicated measurement systems may not be fully or optimally implemented in portable devices, and emphasise that remote monitoring often takes place in the absence of an expert who can properly set up or intermittently tune the instrumentation. To accommodate these situations, SEMG-based devices could include automated quality analysis components.

One of the techniques developed as part of the cleanEMG project, involves the use of a Pairwise Attribute Noise Detection Algorithm (PANDA) to identify noisy records [6]. The algorithm basically assigns a noise factor to a record by comparing characteristic attributes across a set of records. The value of the noise factor can be used to differentiate between clean and noisy records. One of the challenges for this technique is that it requires a control set of clean signals for comparison. Acquiring sufficiently clean signals through measurement is difficult to guarantee, so instead, the cleanEMG researchers opted to try a simulated data set, using a computerised SEMG modelling tool called Myosim [7].

Preliminary studies which aimed to identify simulated noisy signals planted among simulated clean signals yielded encouraging results [8]. In these studies, a set of signals was simulated and then some of them were modified to simulate amplifier saturation, motion artefact and power line interference. However, in unreported studies where

testing was extended to using in vivo SEMG records in comparison to the simulated control signals, all records, clean or otherwise, were identified as suspect.

The researchers acknowledged several limitations in the simulated signals that they used as clean control signals and recommended follow-up work to explore the impact of improvements to the simulation tool for better use with PANDA [6]. In particular, they noted that the simulator they used did not emulate any instrumentation effects which would surely be present in SEMG records, including baseline noise, quantisation, and filtering. They also suggested developing an approach to better match model parameters with SEMG records. The purpose of this work was to conduct this follow-up research.

Details of the model and simulation tool used in the original PANDA studies [6] are described in Chapter 2. Work completed to add instrumentation effects to the simulator is described in Chapter 3, and Chapter 4 describes work completed to refine fixed model parameters, and to facilitate better choice of user-defined model parameters. Chapter 5 reports on an investigation conducted to elucidate effects of these amendments on PANDA-based SEMG quality analysis.

CHAPTER 2: SIMULATING SEMG SIGNALS

The purpose of this work was to explore ways to improve a tool designed to simulate SEMG. The tool (Myosim) and the model which forms its foundation are described below. Details are provided based on the version of the tool used by the researchers who originally applied PANDA to SEMG [6]. While opportunities for improvement are identified here, details about implementing those improvements are described in Chapter 2 and Chapter 4.

2.1 Modelling SEMG signals

An SEMG signal is the result of measuring the electrical activity of muscles at the surface of the skin. An accurate representation of this signal requires a model of the generation of this activity as observed by the sensor at the skin surface but it must also consider effects that the sensor and its associated measuring instrumentation may induce. The simulation tool used in the original PANDA research included a comprehensive generating model, but ignored most of the instrumentation effects which would necessarily be induced during data capture. Both aspects of the complete model are described below.

2.1-1 Signal Generation

The fundamental unit of skeletal muscle is the motor unit (MU) which encompasses a single motor neuron and all the muscle fibres it innervates [9]. Whenever a motor neuron is activated, all the fibres it innervates are also activated. A muscle contraction is the result of simultaneous activation of one or more motor units. The number of motor units per muscle (M), and the number of fibres per motor unit (N) varies from person to person

and from muscle to muscle. A study conducted by Bromberg in [10] reported the number of Motor Units in Bicep Brachii across 10 subjects to vary from $235 \leq M \leq 1580$ with a mean of $M = 553$. This is in loose agreement with the average values reported by Buchthal and Schmalbruch [11] who reported an average of $M = 774$ MUs. The authors in [11] have also reported an average of $N = 750$ fibres/MU in the Biceps Brachii.

The electrical activity of an activated motor unit can be observed and recorded by placing electrodes near the motor unit territory. The pulse that is recorded from an activated motor unit is called the Motor Unit Action Potential (MUAP). The train of pulses formed by several consecutive MUAPs is called a Motor Unit Action Potential Train (MUAPT). The time between consecutive pulses is called the Inter-Pulse Interval (IPI). An SEMG signal can be considered as the summation of several MUAPTs.

An MUAPT is a train of MUAPs and each MUAP in turn is the summation of the action potentials which propagate along fibres that belong to the unit. Each fibre's action potential originates at the site where the motor nerve innervates the fibre which is called the innervation point (IP). Action potentials propagate in both directions away from the innervation point and the speed at which they propagate is called the conduction velocity (v). The activity observed from a single active fibre is called the Single Fibre Action Potential (SFAP).

Figure 2.1 highlights the complete process for modelling the generation of SEMG based on the finite muscle length model originally proposed by Plonsey in [12] and further

developed by Cueto and Parker in [13]. It models observation of the electrical activity of all active fibres at the skin surface with point electrodes.

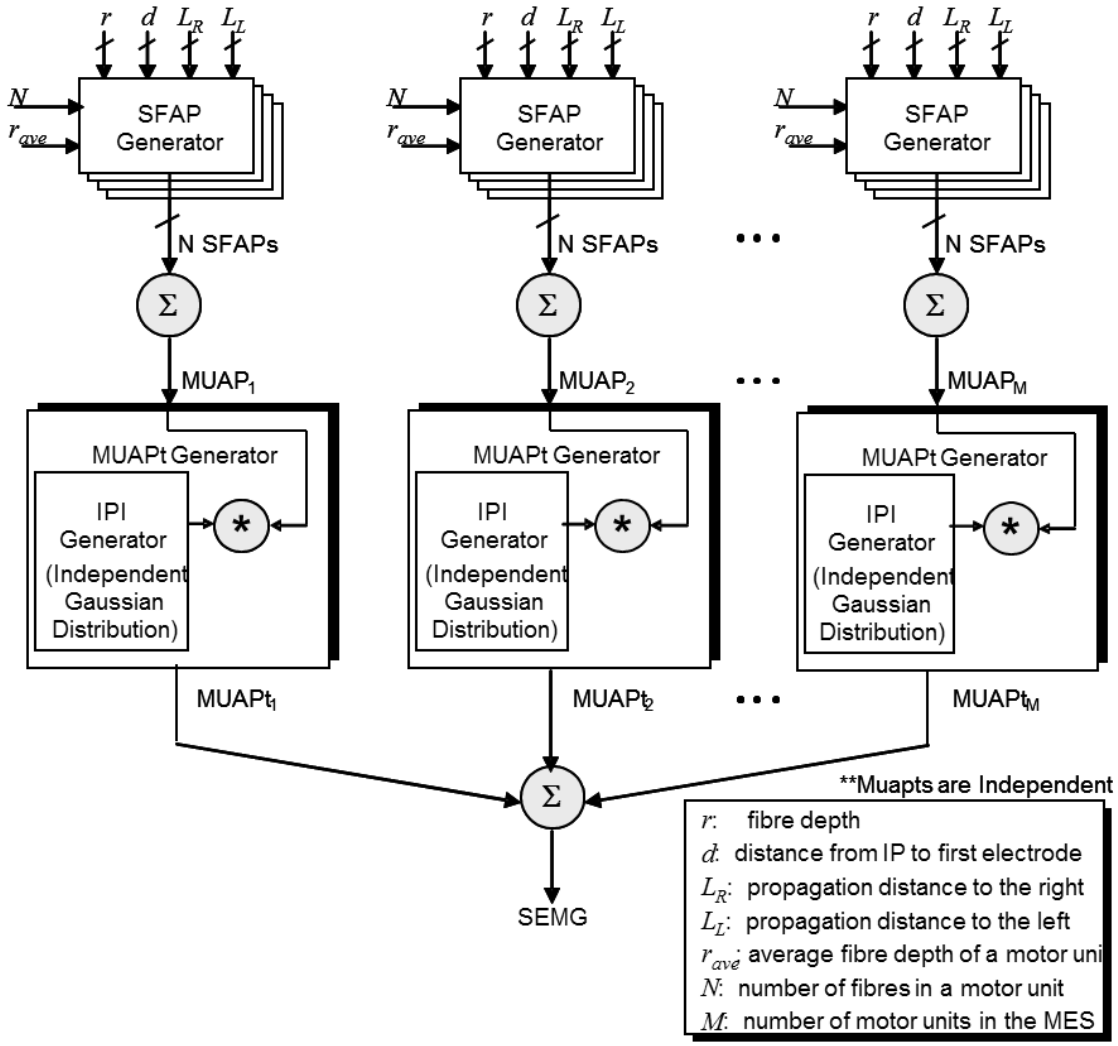
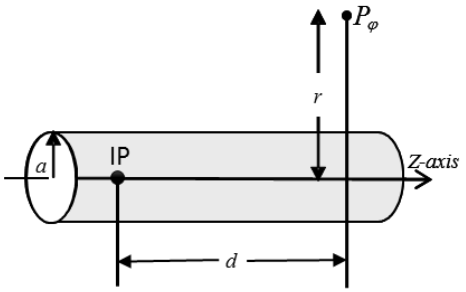


Figure 2.1: Summary of the EMG modelling process [14]

2.1-1a Single Fibre Action Potentials

In the model, muscle fibres are assumed to be cylindrical with radius 'a' and longitudinal Z-axis located at a depth 'r' as observed from a point 'P_φ', representing electrode location. Figure 2.2 illustrates these geometrical parameters.



Typical Fibre Parameter Values [15]:

radius: $a = 25\ \mu m$
conductivities: $\sigma_i = 0.45\ s.m^{-1}$,
$\sigma_e = 2.5\ s.m^{-1}, \sigma_l/\sigma_t = 5$
conduction velocity: $v = 4 - 6\ m.s^{-1}$

Figure 2.2: Characteristics of a single muscle fibre [14]

The extracellular potential as observed from the point P_φ represents the SFAP and is calculated by convolving a double layer differential source with a tissue filter that considers various geometric and physiological parameters of the fibre. Figure 2.3 shows an example of the source, tissue Filter and resulting SFAP. The following equation taken from Cueto-Gonzalez and Parker [13] describes the convolution process:

$$SFAP(t) \cong K1 \times s(t) * [h_L(t, d, r, v, L_L) + h_R(t, d, r, v, L_R)\ ] \quad \textbf{2.1}$$

In equation 2.1, $K1 = {\sigma_i \theta^2}/{8\sigma_e}$, where $\theta = {a}/{v}$, and σ_i and σ_e are the intracellular and extracellular conductivities which constrain the fibre's transmembrane potential $V_m(t)$. The value of $K1$ was nominally set to 1 in Myosim 1.0 because prior work involving Myosim only focused on the shape of the resultant signal rather than the absolute value

of the amplitude and thus, $K1$ could be ignored. Nevertheless, values provided in Figure 2.2 are representative of what is found in the literature.

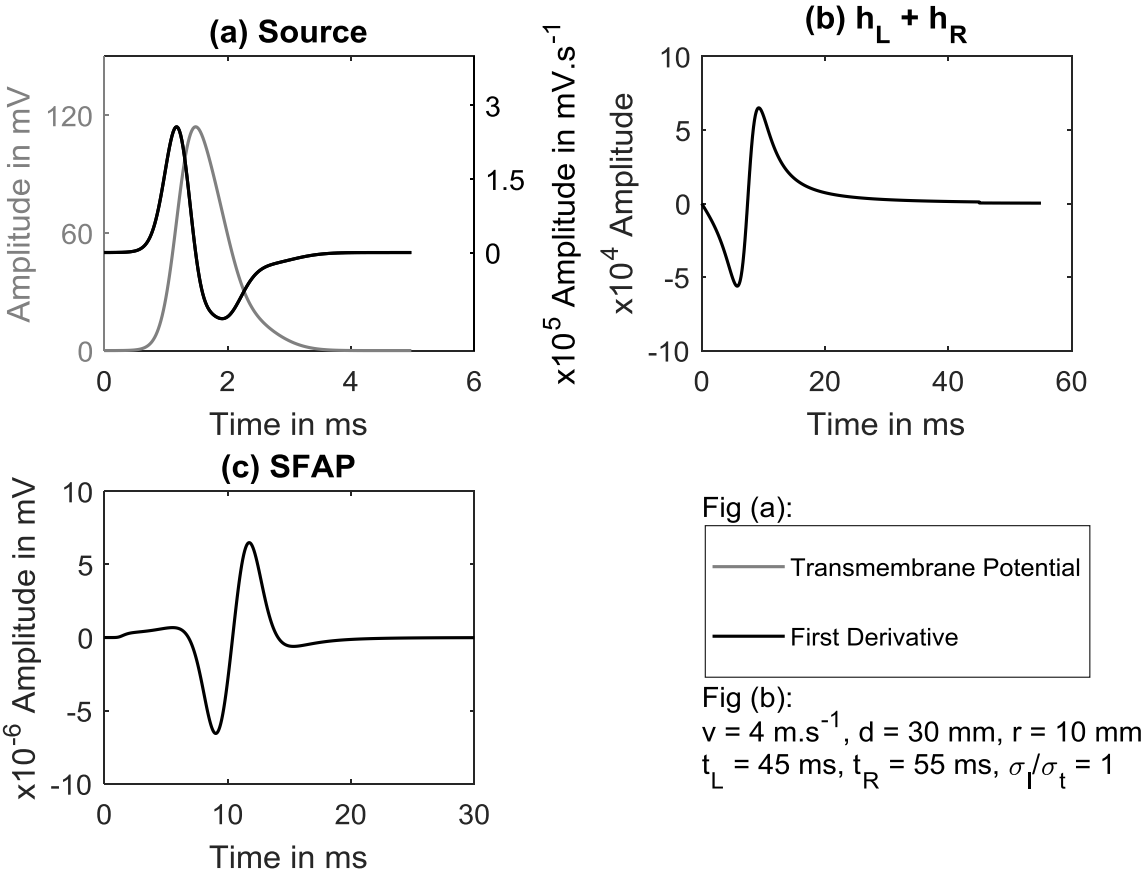


Figure 2.3: Transmembrane potential, source, tissue filter and SFAP

The source, $s(t)$ is modelled as the first derivative of the fibre transmembrane potential (i.e. $s(t) = \partial V_m(t)/\partial t$). Plonsey's model [12], empirically derived, is used in Myosim to model the transmembrane potential:

$$V_m(Z) = \sum_{a=1}^{3} C_a \times e^{F_a(Z)} \qquad \textbf{2.2}$$

where $C_1 = 51, C_2 = 72, C_3 = 18, F_1(Z) = -64(Z - 0.54)^2, F_2(Z) = -28.41(Z - 0.66)^2$ and $F_3(Z) = -11.09(Z - 0.86)^2$.

Plonsey provided the model as a function of Z, so a transformation assuming $Z = v_s t$, was necessary to derive the model in time. Plonsey did not provide the value of v_s for the particular fibre from which he was taking measurements, so Cueto-Gonzalez and Parker assumed a nominal value of $v_s = 5\ m.s^{-1}$ to complete the transformation. Taking the derivative of the above equation with respect to Z and applying the transformation yields the first derivative of the transmembrane potential with respect to time:

$$\frac{\partial V_m(t)}{\partial t} = v_s \times \sum_{a=1}^{3} C_a \times \frac{dF_a(v_s t)}{dt} \times e^{F_a(v_s t)} \qquad \textbf{2.3}$$

A plot of the transmembrane potential and its first derivative are depicted in Figure 2.3(a) with $v_s = 5\ m.s^{-1}$.

The tissue filter in Cueto-Gonzalez and Parker's work was modelled as delineated in equation 2.4 where σ_l and σ_t are longitudinal and transversal conductivities respectively and all of the other parameters are as defined previously.

$$h(t) = \frac{t}{\left[\frac{\sigma_l}{\sigma_t}\left(\frac{r}{v}\right)^2 + t^2\right]^{3/2}} \qquad \textbf{2.4}$$

Cueto-Gonzalez and Parker considered propagation down the fibre in both directions when applying the tissue filter yielding the sum of the left h_L and right h_R impulse responses given by:

$$h_L(t,d,r,v,t_L) = \begin{cases} h\left(t + {}^{d}/_{v}\right), & 0 \le t \le t_L \\ 0, & otherwise \end{cases} \quad \textbf{2.5}$$

$$h_R(t,d,r,v,t_R) = \begin{cases} h\left(t - {}^{d}/_{v}\right), & 0 \le t \le t_R \\ 0, & otherwise \end{cases} \quad \textbf{2.6}$$

where, t_L and t_R are the time taken to propagate to each end of the fibre. With the same justification provided for $K1$, the value of ${}^{\sigma_l}/_{\sigma_t}$ was nominally set to 1 in Myosim 1.0, but can more accurately be set as ${}^{\sigma_l}/_{\sigma_t} = 5$ according to [15].

2.1-1b Motor Unit Action Potential

A motor unit pulse detected by surface electrodes is the result of the summation of potentials of the N fibres belonging to that motor unit. Fibres in the MU will have different depths, lengths, innervation points, and conduction velocities. Figure 2.4 and Figure 2.5 help to visualise a motor unit's geometry, and Table 2.1 describes the parameters shown on the figures:

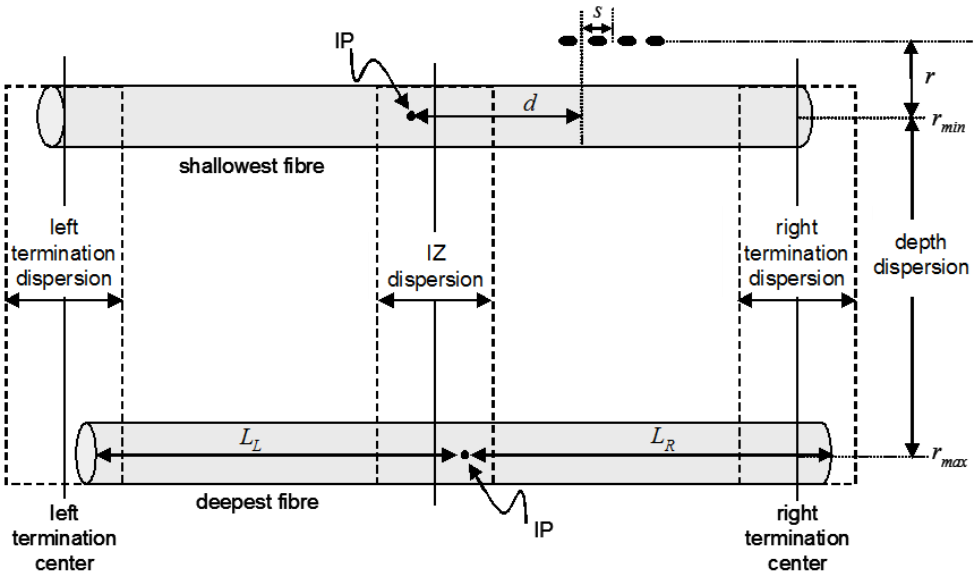


Figure 2.4: Motor unit model parameters

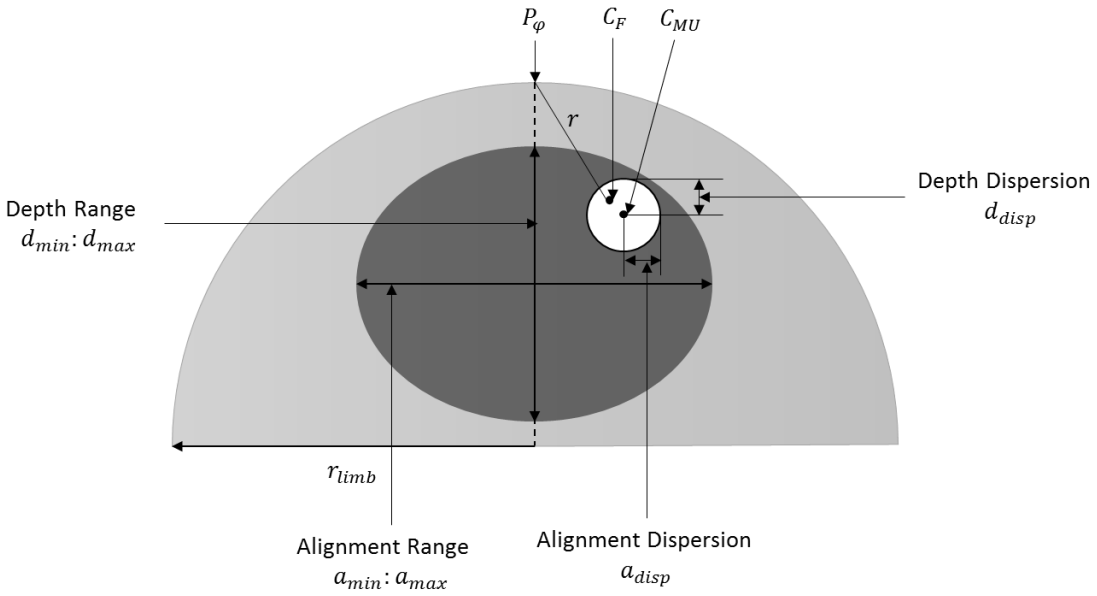


Figure 2.5: Motor unit geometry

Parameter	Description
s	Electrode spacing
C_{MU}	Centre of a motor unit, chosen from a uniform random distribution from the ellipse described by the depth and alignment ranges.
C_F	Centre of a fibre, chosen from a uniform random distribution from the ellipse described by the depth and alignment dispersions.
r	Depth of a given fibre, Euclidean distance between the centre of fibre C_F and the recording point P_φ.
d	Distance between IP of a given fibre and the closest electrode. IP location chosen relative to IZ centre from a uniform distribution within the IZ dispersion limits.
L_R	Distance between IP of a given fibre and its right termination. Fibre termination location chosen relative to right termination centre from a uniform distribution within the right termination dispersion limits.
L_L	Distance between IP of a given fibre and its left termination. Fibre termination location chosen relative to left termination centre from a uniform distribution within the left termination dispersion limits.

Table 2.1: Description of SFAP parameters in MUAP modelling

For a motor unit consisting of N fibres, the MUAP is modelled using the simple equation:

$$MUAP(t) = \sum_{n=1}^{N} SFAP_n(t) \qquad \textbf{2.7}$$

2.1-1c Motor Unit Action Potential Trains and the SEMG Signal

An MUAPT is a train of MUAPs with the pulses separated by some IPI. The values of IPI in the train are random, arguably Gaussian in nature [16]. The term 'Firing Rate' (FR) is also

used in the literature to describe the separation between the pulses and is equal to the inverse of the mean IPI [16].

An MUAPT can be modelled by convolving an MUAP with an impulse train consisting of randomly located impulses. Each location is set by incrementing by the random IPI. Each IPI is set by sampling from a Gaussian distribution with a specified mean and standard deviation. The firing rate depends on the muscle and the force being exerted. Typical values for the Bicep Brachii range from $15\ Hz - 30\ Hz$ [40]. Assuming there are P pulses randomly positioned in the train, the MUAPT is modelled as:

$$MUAPT(t) = MUAP(t) * \sum_{p=1}^{P} \delta(t - t_p) \qquad \textbf{2.8}$$

where t_p is the location of the Pth impulse.

Figure 2.6(a) shows an example MUAP consisting of $N = 200$ fibres which are uniformly distributed in a circular MU whose centre is located $C_{MU} = 10\ mm$ below the surface with a diameter of $8\ mm$ and the corresponding MUAPT with an average firing rate of $7\ Hz$ and a coefficient of variation of 0.25 is shown in Figure 2.6(b).

The SEMG signal is obtained by summing several MUAPTs. For M MUAPTs, the resultant SEMG is obtained using equation 2.9. An example SEMG generated using $M = 50$ Motor Units containing $N = 200$ fibres each is shown in Figure 2.6(c).

$$SEMG(t) = \sum_{m=1}^{M} MUAPT_m(t) \qquad 2.9$$

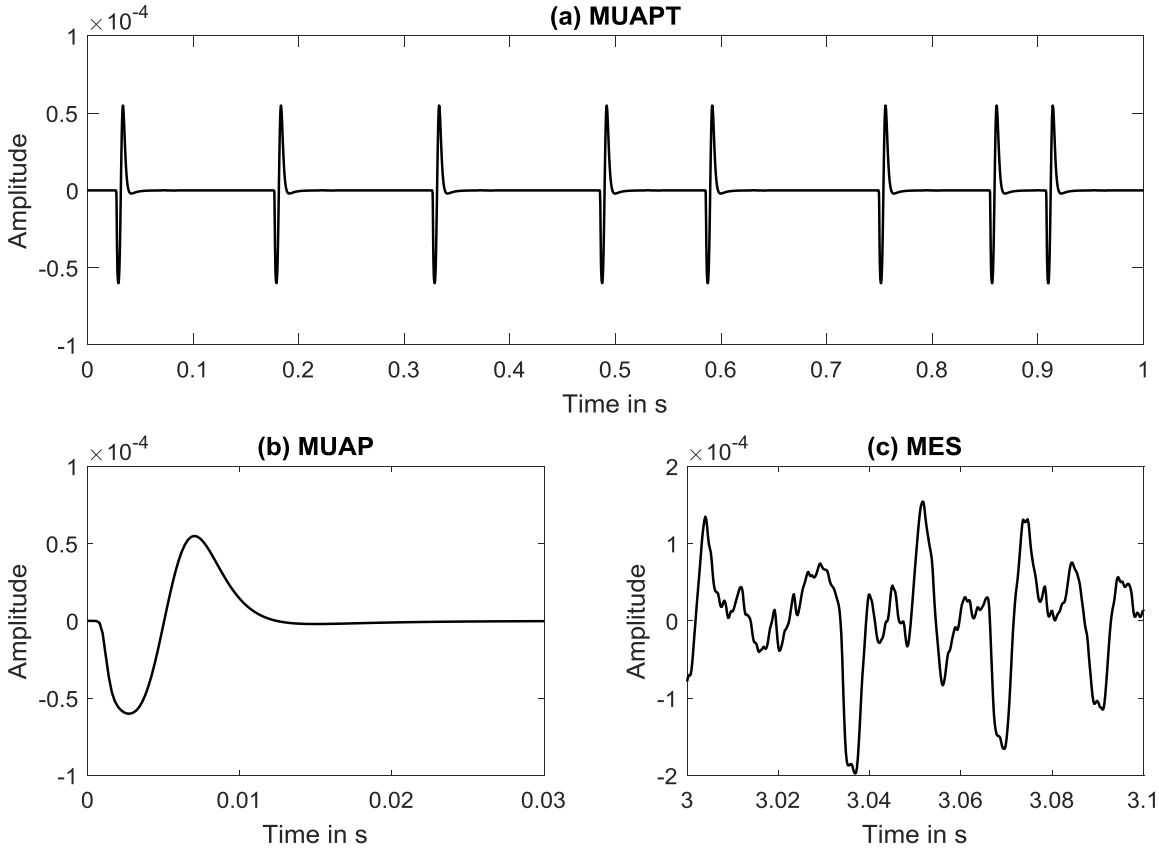


Figure 2.6: Example MUAP, MUAPT and SEMG

2.1-2 Data Capture

A typical instrumentation set up to capture signals in vivo is depicted in Figure 2.7:

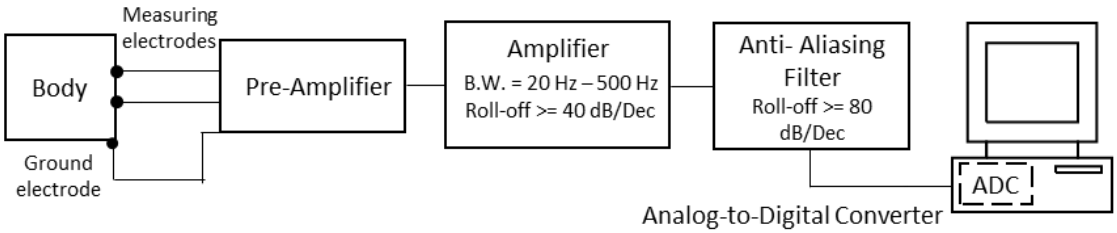


Figure 2.7: Typical SEMG signal collection set-up

The instrumentation consists of several components including sensors, amplifiers, filters, and an Analogue-to-Digital Converter. These components are made from electronic circuit elements such as electrodes, capacitors, resistors, transistors, operational amplifiers, etc. Thus, even 'clean' records are band-pass filtered and quantised, and contain some unavoidable instrumentation noise. The simulation tool used in the original PANDA research ignored these 'Instrumentation Effects'. Though none of these effects are drastic, they may still have contributed to PANDA's bias for assigning high error factors to the records. A description of each is provided in the sections that follow.

It is also worth noting that electrode configuration and shape can influence SEMG records. Typically, electrodes are used in one of two configurations: monopolar or bipolar [17]. In a monopolar configuration, a single surface electrode is placed on the muscle of interest and the detected signal is amplified and recorded. In a bipolar configuration, two electrodes are placed on the muscle within a few centimetres of each other and a differential amplifier is employed to record and amplify the SEMG signals. The main advantage of using a bipolar configuration is its ability to supress interference from other biological sources, thus yielding a cleaner signal. Myosim provides the option to generate both monopolar and bipolar signals. The authors in [17] also point out that the shape of

the electrodes does not matter, only the surface area does. Myosim currently does not currently incorporate the electrode shape or area in the model.

2.1-2a Baseline Noise

Baseline noise originating from instrumentation is a result of the fundamental properties of device components and the charge carriers, electrons. While generally small compared to amplitudes of signals, this noise is ever present. Specific types of instrumentation noise are well established in the literature and include white noise (thermal, shot), pink noise (flicker), and red/brown noise (popcorn (burst), avalanche) [18] - [20]. Some sources of noise are more dominant than the others and should be taken into consideration, while weaker sources can be ignored. Literature suggests that the most dominant types of instrumentation noise in SEMG are [21] - [24]:

- Thermal Noise – dominant at higher frequencies and stays relatively constant (white).
- Flicker Noise – dominant at lower frequencies rolling off according to $K \times 1/f$ (pink).

Instrumentation noise is not the only source of baseline noise in SEMG. Power line interference can be present for instance, because our bodies are capacitively coupled to the powered devices in our surroundings. This noise is cyclic in accordance with the power mains. While well designed amplifiers can attenuate power line interference through common mode rejection, it is rarely fully mitigated in biosignal records and often presents substantially in SEMG measurements [25]. Low amplitude power line interference can therefore be considered along with thermal and flicker noise as possible components of baseline noise when modeling instrumentation effects.

With well setup instrumentation, baseline noise can be expected to have minimal effects on SEMG measurements. Nevertheless, adding options to add baseline noise to simulated SEMG signals allows for the analysis of the effects on signal features, not only when they are minimal, but also when they may become more profound due to poorly designed or set instrumentation.

2.1-2b Band-pass Filtering

It is common practice to filter an SEMG signal as it is being measured. Amplifiers are set up to band-pass the input to avoid noise components which fall outside the bandwidth of the bulk of SEMG. Well-designed amplifiers have filtering capabilities with a reasonably flat passband and roll-offs typically around $40\ dB/decade$ [26], [27]. While the exact specifications of the filters might vary, typical standards are [26], [27]:

- Low- frequency cut-off: $10\ Hz - 20\ Hz$, roll-off rate: $40\ dB/decade$ (or 2^{nd} order IIR filter)
- High-frequency cut-off: $500\ Hz$ with a roll-off rate: $40\ dB/decade$ (or 2^{nd} order IIR filter)

Also, often anti-aliasing filters are used as an additional low pass filter because they have a sharper roll-off, typically greater than $80\ dB/decade$ [28]. Pass band filtering can therefore be taken into consideration when modelling instrumentation effects.

While properly set up filters preserve the bulk of SEMG information, it is challenging to choose a cut-off frequency which precisely attenuates noise without attenuating any SEMG signal. So even in a simulated signal with no noise, filtering may affect signal features, especially those which are sensitive to low and high frequency components.

These will become more pronounced as baseline noise is added to the signal. These effects are expected to be small, but may be worth examining.

2.1-2c Quantisation

Since SEMG signals are generally recorded digitally, the effects of digital conversion must be considered. An analogue-to-digital Converter quantises analogue inputs into digital bins. The number of bins (L) is defined by the number of bits in the storage register ($L = 2^{N_{bits}}$, where N_{bits} is the number of bits). Resolution of the quantised result is therefore dependent on the bit size of the register, and the step-size in volts (ΔV) is determined by dividing the full dynamic range (V_{FDR}) of the converter by L.

Quantisation noise is often characterised as a random signal with 0-mean, uniformly distributed between $\pm {}^{1}/_{2}\,\Delta V$ [29]. As such, power in this noise source is typically estimated as $V_{RMS}^2 = {}^{\Delta V^2}/_{12}$. Typically, SEMG measurement systems use converters with $N_{bits} = 8, 12$ or $16\ bits$ to represent a signal yielding $L = 256, 4096$ or 65536 respectively. As such, quantisation error is small and its effects are usually negligible on the SEMG signal amplitude. However, some signal features, such as those derived from the amplitude histogram of a signal, may be affected. Last bin count and entropy are two examples. Thus, it is worth adding the quantisation process to Myosim.

2.2 Myosim 3.0 – A User Friendly SEMG Simulation Tool

Myosim (MYOelectric SIMulator) is a computerised simulation tool developed in MATLAB by MacIsaac et al. [7] to simulate surface EMG. The tool uses a finite fibre length model described by Cueto-Gonzalez and Parker [12] as detailed in the previous section. SFAPs

are generated by convolving a transmembrane potential source with a tissue filter, based on specified model parameters. SFAPs are grouped and summed to produce a set of motor unit action potentials (MUAPs) and when these are convolved with a set of pulse trains, the result is summed to produce an SEMG signal.

Figure 2.8 depicts the graphical user interface available in Myosim Version 3.0. While the tool offers a wide array of functionality gradually added as it migrated from Version 1 to Version 3, only the aspects of interest here will be highlighted. They are mostly found in the 'Parameters' dialog displayed in Figure 2.9. The tool sets model parameter values based on user inputs and random variations about the inputs, the limits of which are also user specified. Model parameters include the geometry of each fibre in each motor unit relative to electrode location (depth, alignment, fibre terminations, and innervation point) as well as the number of motor units, the number of fibres per motor unit, conduction velocity, and firing statistics. The tool also allows a user to specify more than one channel location, the electrode configuration (monopolar, bipolar, double differential), and the electrode spacing. Finally, sampling rate and signal duration are also user specified.

Practically, it is problematic to presume the value of model parameters since they cannot be directly measured. The tool is designed to constrain parameters to within physiological limits, but true values can still vary considerably from person-to-person, muscle-to-muscle, activity-to-activity, etc. This presents a problem when trying to simulate clean control signals for use with PANDA. If the parameter values chosen for the control signals

do not yield sample signals which sufficiently reflect in vivo recorded SEMG signals under test, PANDA will be biased to assign high error factors to the records [6].

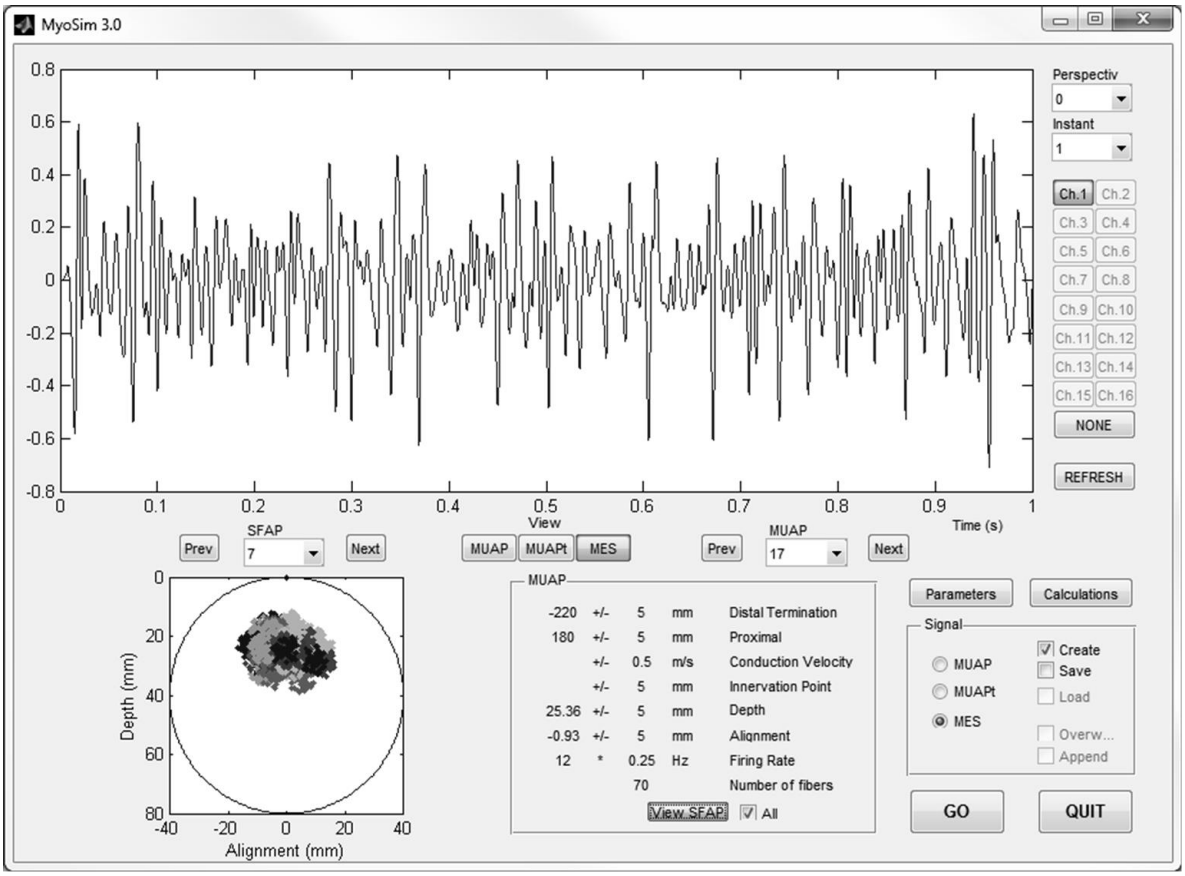


Figure 2.8: Myosim 3.0

Figure 2.9: 'Myosim Parameters' window

2.3 A closer look at PANDA-based SEMG quality analysis

Since the principal motivation of this work was to improve the simulator for use with PANDA-based SEMG quality analysis, details regarding PANDA are described here. First proposed by researchers in [30], PANDA is a ranking algorithm based on a 'noise factor' it assigns to the signals, calculated from pairwise comparisons of attributes across the signals. The algorithm was adapted by researchers in [6] to classify SEMG signals as clean or noisy.

In the context of SEMG, the data set used in PANDA-based analysis typically consists of several clean 'control' signals and a few test signals which must be flagged as clean or noisy. The attributes used for comparison in PANDA can be represented by features of the SEMG signal. The original work developing PANDA for use with SEMG used the following nine features [6]: Mean Absolute Value (MAV), Slope Sign Change (SSC), Zero Crossing (ZC), Wavelength (WL), Entropy (EN), Histogram (HIST), Signal to Motion Artefact Ratio (SMR), $60\ Hz$ power (PLI60) and $180\ Hz$ power (PLI180).

For each feature, PANDA divides the clean control signals into a specified number of bins based on similarity of that feature. A signal under test is then put into the bin whose mean is closest to its own feature value. The deviation of the test signal for the remaining features with respect to the means of the bin are calculated (and normalised by the standard deviation) and the sum of normalised deviations is used to calculate the noise factor for the signal under test. The higher the deviation of the features from the average values for each bin, the higher the noise factor, and consequently the more likely that the

signal under consideration is 'suspect'. For a more in-depth explanation of PANDA, readers are directed to [6] and [30].

The original PANDA work reported encouraging results when first investigating PANDA for use with SEMG signals [6]. The researchers simulated a set of control signals, and then simulated a set of test signals. They maintained a clean copy of each test signal and then corrupted a copy of each by simulating power line noise, motion artefact, and saturation. PANDA could easily distinguish between the simulated clean and noisy test signals, but limitations in the simulation tool prevented it from producing simulated control signals which represented in vivo SEMG records. When used to test the records, test Noise Factors were always well above the noise threshold, even for clean test records [6]. Figure 2.10 shows examples of PANDA in use with (a) simulated test signals and (b) in vivo test records. Power line interference and Motion artefact were included in the noisy test records at an SNR of $5\ dB$ in both examples.

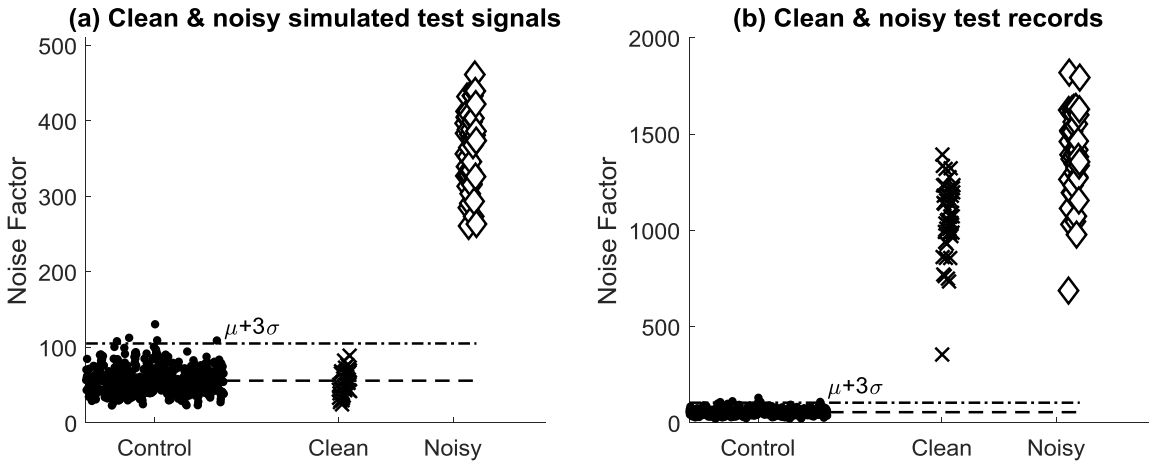


Figure 2.10: PANDA with different test signals

Chapter 3: ADDING INSTRUMENTATION EFFECTS TO MYOSIM

To more realistically represent in vivo recorded SEMG signals (herein termed 'records'), instrumentation effects were built into Myosim V4.0. Once the pure signal is generated, simulated baseline noise is added, and the result is filtered, then quantised. This more accurately reflects the capture process typical for SEMG signals. While the effects of these additions are expected to be small, they may help to better simulate 'clean' signals for use in PANDA-based SEMG quality analysis. Details about how each of these additions was made and how their addition influences the signal are described below.

3.1 Baseline Noise

An empirical study was conducted to verify the components of baseline noise as reported in Chapter 2 and to establish a model (and default parameter values) for implementing the noise as part of Myosim V4.0. Thermal and Flicker noise were expected to be the significant contributors along with some residual power line interference.

3.1-1 Methods

3.1-1a Data Collection

Data were captured from two subjects participating in another SEMG study [31] with contractions from the biceps brachii under investigation. As such, the measurement protocol reflected one that is typical for SEMG signal acquisition. Prior to collection, the skin where electrodes were place was cleansed with alcohol. Four channels from a flexible eight channel bar electrode array were used in bipolar configuration. Bars were fixed at $10\ mm$ apart. Fastened with adhesive tape, the array was centred over the short

head of the muscle and placed in parallel to the muscle fibres. Conductive gel was applied through holes fitted in the array to provide better conductivity. The reference electrode, a Red Dot electrode, was placed on the neck.

All channels were processed identically first through a pre-amplifier configuration described previously by Lovely [32] and then through a Tektronix AM502 differential amplifier. The pre-amplifier gain was set to 20 while the main amplifier gain was set to 500. Since baseline noise was the focus for analysis here, the pass-band region was extended beyond what is typical for SEMG - the main amplifier was set from $0.1\ Hz - \ 3\ KHz$. Each channel was further processed through an anti-aliasing filter with a cut-off frequency of $2\ KHz$ (the maximum setting available). The resulting signal was sampled at a rate of $5\ KHz$ using a 16 bit MCC PCIM-DAS 1602/16 A/D board to be recorded on a computer. To acquire only noise, rest segments were collected. Each subject was asked to sit with their arm resting on a table while the noise was recorded for 60 seconds.

3.1-1b Data Analysis

A 15 second segment from an arbitrary channel for each subject was selected for analysis. The gain was removed from each segment and the power spectral density (PSD) was computed using Welch's method with 1 second non-overlapping windows. The resulting spectrum was visually inspected to identity characteristic features of the noise.

Based on a visual inspection of the spectrum, the power spectral model depicted in Figure 3.1 was chosen to parameterise the baseline noise. This is a commonly used model for

instrumentation noise [23] and combines pink and white noise as a piecewise linear log-log plot. Since the power of pink noise is inversely proportional to frequency, it is much more pronounced at lower frequencies, which is depicted as the 'pink region' in the figure. The spectrum rolls off proportionally to the frequency from some value P_0 for $f > 0$, and after a certain frequency it falls flat to a value P_W. The corner frequency f_{cp}, represents the frequency after which the white noise begins to dominate, and thus the spectrum appears flat, which is depicted as the 'white region' in the figure.

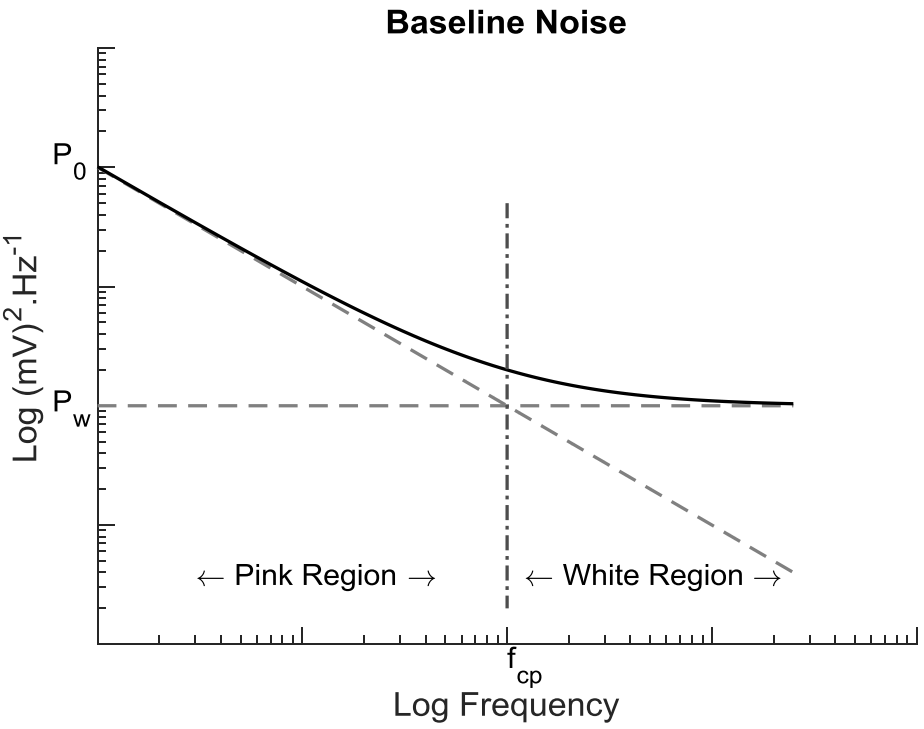


Figure 3.1: Expected PSD of baseline noise

The values of the parameters depend on several factors, such as the device, temperature, components used, and were therefore user defined in Myosim V4.0. However, default values were established using the data collected for this work. The corner frequency f_{cp} can be obtained by visually inspecting the PSD of the recorded noise segments to find the

frequency at which the spectrum beings to fall flat. Since more than one record was available, the average across both records can be used. In the pink region, the power ϕ_i at a frequency bin f_i can be approximated as $\phi_i \cong {P_0}/{f_i}$. Thus, an estimate of P_0 can be obtained by taking the mean across frequency bins according to $\sum_{i=1}^{n} {\phi_i \times f_i}/{n}$, $0 < f_n \leq f_{cp}$. Again, this process can be averaged across segments. With P_0 and f_{cp} set, the constant white-band power P_W can be approximated as the power at the corner frequency, i.e., $P_W = {P_0}/{f_{cp}}$

Once values for the parameters were obtained, a random signal with a Gaussian distribution was shaped to simulate the noise. The mean was set to 0 and the variance was set proportional to P_0. The magnitude component of its Fourier transform corresponding to frequency bin f_i was scaled by $\sqrt{{1}/{f_i}}$, $0 < f_i \leq f_{cp}$ or by $\sqrt{{1}/{f_{cp}}}$, $f_i > f_{cp}$. The inverse Fourier transform was then used as the simulated white and pink noise.

The final component of baseline noise is the power-line components present at $60\ Hz$ and its harmonics. The power components in the bands $[\ k \times 60 - 1, k \times 60 + 1\ ]$, $k = 1,2 \ldots 5$ was summed to obtain an estimate of the power of each of the harmonics, $P_{k\times 60}$ for each record and the values were averaged across the records. Power-line interference was then simulated by generating sinusoids with random phase at corresponding frequencies with amplitudes proportional to the observed powers. For this work, only the first five harmonics were considered.

3.1-2 Results

The PSD of one of the segments with the gains removed is shown in Figure 3.2(a) on a linear plot. This plot is representative of what was observed in both segments. For comparison, the spectrum of an SEMG signal is also shown on a different scale. It was already established that the baseline noise level is typically small compared to the SEMG signal, and this is evident in the plot (the SEMG scale is 30X smaller than the noise scale). It is worth noting however that the instrumentation amplifier used in this experiment was well designed with a common mode rejection ratio rated as $CMMR > 100\ dB$ [32] yet remnant power line interference is still evident. It is also worth noting that the noise captured in this experiment exemplifies what would be collected in a pass band much wider than what is typical during an SEMG study ($0.1\ Hz - 2\ KHz$ vs $20\ Hz - 500\ Hz$), and still the white-band region was not captured. This is evident in Figure 3.2(b), which shows the PSD of the noise on a log-log plot. The pink noise region is clearly evident with each spectrum showing a $1/f$ decline, but there is no evidence of any leveling off. Thus, for this instrumentation, the value of f_{cp} was nominally chosen to be equal to the Nyquist frequency of $2.5\ KHz$.

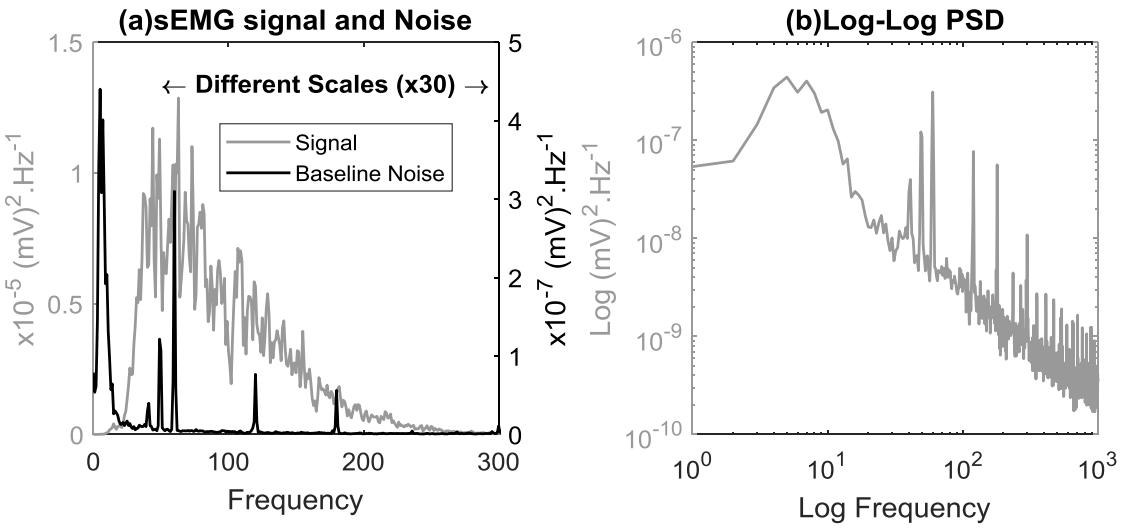


Figure 3.2: PSD of EMG signal and baseline noise

For the instrumentation set up in this experiment, the model parameter values obtained are delineated in Table 3.1:

Instrumentation Noise	Power line Interference
$P_0 \cong 5 \times 10^{-13} V^2.Hz^{-1}$	$P_{60} \cong 6 \times 10^{-13}\, V^2.Hz^{-1}$
$f_{cp} \geq 2.5\, KHz$	$P_{120} \cong\ 1 \times 10^{-13}\, V^2.Hz^{-1}$
	$P_{180} \cong\ 8 \times 10^{-14}\, V^2.Hz^{-1}$
	$P_{240} \cong\ 1 \times 10^{-14}\, V^2.Hz^{-1}$
	$P_{300} \cong\ 2.5 \times 10^{-14}\, V^2.Hz^{-1}$

Table 3.1: Baseline noise parameters

Using these values, baseline noise was simulated by using the model described previously. The time-domain and the PSD of the simulated noise are shown in Figure 3.3(a) and (b). One of the noise records obtained from the instrumentation setup is superimposed on each plot. Visual inspection of the figure indicates that the noise captured in this experiment was well modelled.

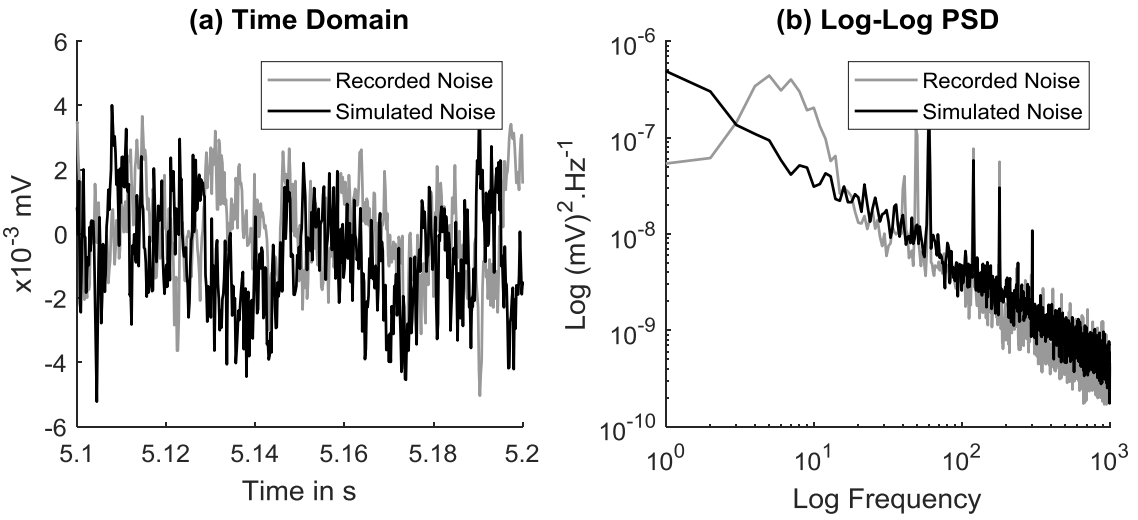


Figure 3.3: Simulated baseline noise

3.1-3 Conclusions

Experimental noise was consistent with what has been reported in the literature, and the piecewise pink-white model was sufficient to represent the baseline noise captured. Instrumentation used to record the signals was sufficiently well designed so that the white noise region was undetectable so P_W could not be established, but the pink noise parameter was established to be $P_0 = 5 \times 10^{-13}\ V^2.Hz^{-1}$. All that could be ascertained about the corner frequency was that it was greater than the pass-band set in this work, i.e. $f_{cp} \geq 2.5\ KHz$. Power line interference harmonics were also observed between $6 \times 10^{-13}\ V^2.Hz^{-1} \leq P_{k60} \leq 1 \times 10^{-14}\ V^2.Hz^{-1}$. The model was implemented in Myosim V4.0 with default parameters set as specified. Figure 3.7 depicts the GUI interface added to Myosim to allow users to specify baseline model parameters.

3.2 Filtering and Quantisation

To simulate the band-pass filtering process, a low-pass and a high-pass Butterworth filter were combined. This was implemented in Myosim V4.0 such that users can set the order and cut-off frequencies of each filter independently. Default values were set to 2nd order with a low-pass cut-off of $f_{lp} = 500\ Hz$, and a high-pass cut-off of $f_{hp} = 20\ Hz$.

Similarly, the anti-aliasing filter was implemented as a Butterworth low-pass filter with configurable order and cut-off frequencies. The defaults were set to 4^{th} order with a cut-off frequency of $f_{lp} = 500\ Hz$.

The magnitude responses of these filters are shown in Figure 3.4 within one octave of the region of interest.

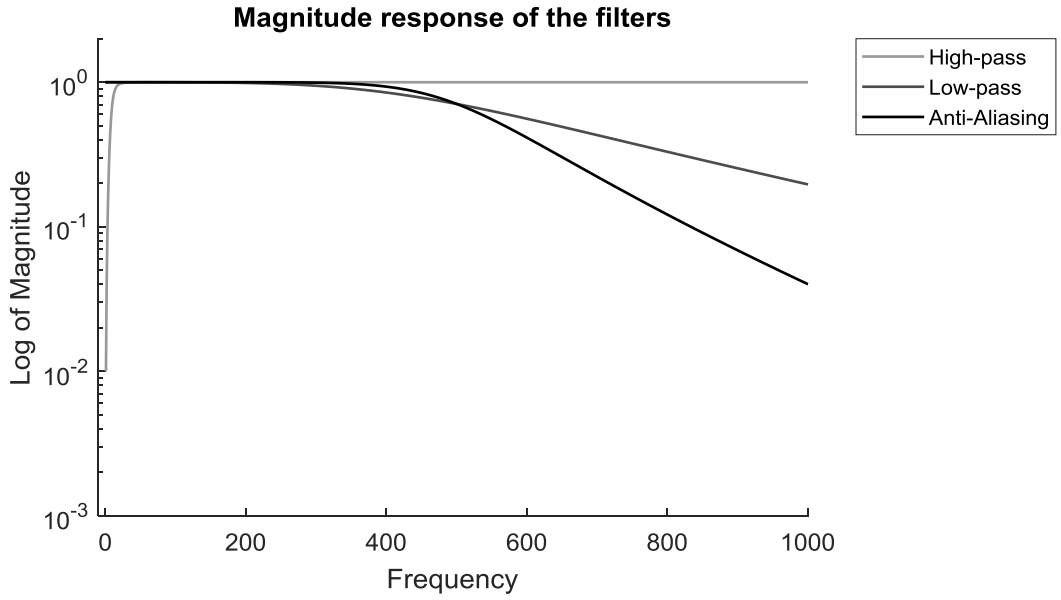


Figure 3.4: Frequency response of the band-pass filter

To demonstrate the effects of filtering on the SEMG signal, an SEMG signal was generated using the simulator and baseline noise was added to it. The baseline noise parameters

were chosen to reflect an instrumentation that was expected to be reasonable, but the power of the noise was scaled so that the Signal to Noise Ratio (SNR) met that which was measured in the SEMG record collected in the study described in the previous section. The average SNR obtained across the 2 subjects was found to be $25\ dB$ when an active segment was compared to the rest segment for each record. The filtered and unfiltered signals in time and frequency domain are depicted in Figure 3.5. The filter used was a 2^{nd} order band-pass filter with $f_{hp} = 10\ Hz$ and $f_{lp} = 500\ Hz$ cascaded with a 4^{th} order anti-aliasing filter with $f_{lp} = 500\ Hz$. No significant difference can be observed among the signals, except for the phase shift introduced by the filter that looks fairly linear within the band of interest. This was expected because parameters were set to filter beyond the band which contains the bulk of SEMG, in alignment with experimentation. This result highlights the merely marginal effect of filtering on SEMG records with little noise during acquisition.

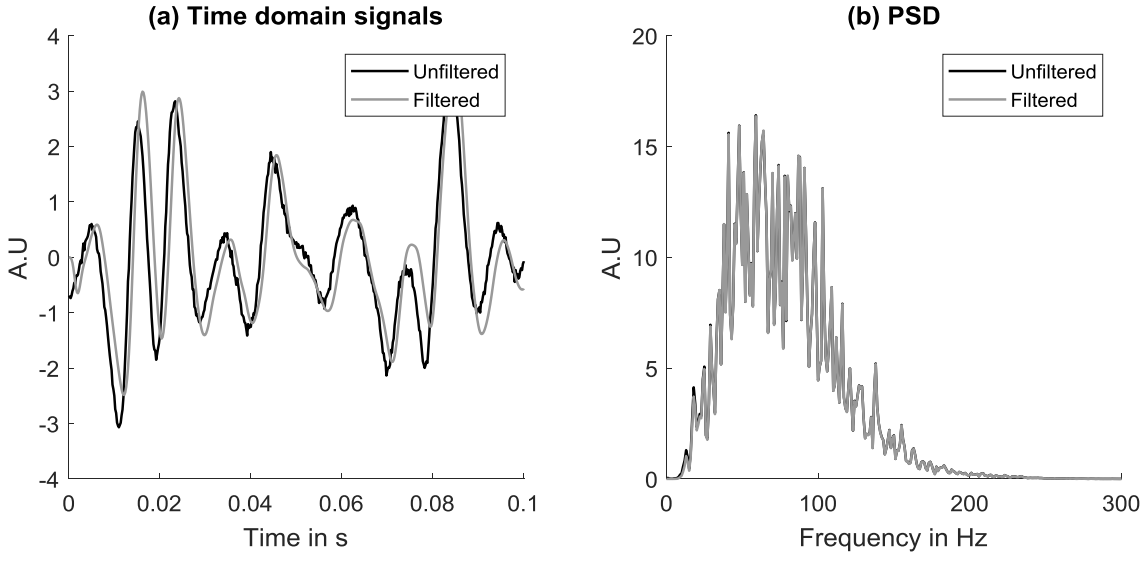


Figure 3.5: An example of Simulated Unfiltered vs Filtered SEMG

Finally, to model the quantisation process, a binning process was also implemented in Myosim V4.0 based on a user specified bit-size for the ADC register. Assuming a register with greater than 8 bits, the effects of quantisation should be almost negligible for SEMG, but some features of the signal particularly sensitive to histogram shape may show some measurable influence.

3.3 Inspecting the effects of Instrumentation Noise on Signal Features

The effects of baseline noise, filtering and quantisation on commonly extracted SEMG features are shown in Figure 3.6. For visual comparison between features, each feature was normalised with respect to the maximum value among all the five groups. The SEMG simulated to extract the features was generated using the following model parameters:

Parameters	Value
Number of fibres per MU	400
Number of MUs	100
Distal termination	$-220\ mm \pm 5\ mm$
Proximal termination	$180\ mm \pm 5\ mm$
Innervation Point dispersion	$\pm 5\ mm$
Vertical depth range of MUs	$10\ mm\ to\ 20\ mm$
Horizontal alignment of MUs	$-10\ mm\ to\ 10\ mm$
MU radius	$5\ mm$
Radius of limb	$40\ mm$
Conduction velocity	$4\ m.s^{-1} \pm 0.2\ m.s^{-1}$
Firing rate	$5\ to\ 20\ *\ 0.15\ Hz$
Electrode Locations	$30\ mm\ and\ 40\ mm$
Perspective	$0\ mm$
Sampling frequency	$5\ KHz$
Signal duration	$10\ s$

Table 3.2: Simulation Parameter values used in the Inspection of the effects of Instrumentation Noise on SEMG features

This clean signal and four altered versions of it are depicted:

- Clean filtered: clean signal passed through the filters (Clean Filtered)
- Noisy unfiltered: clean signal contaminated with baseline noise but not filtered
- Noisy filtered: clean signal contaminated with baseline noise and filtered
- Noisy, Filtered, quantised: clean signal contaminated with baseline noise, filtered and quantised.

The filter used was a 2^{nd} order band-pass filter with $f_{hp} = 10\ Hz$ and $f_{lp} = 500\ Hz$ cascaded with a 4^{th} order anti-aliasing filter with $f_{lp} = 500\ Hz$. The noise was generated with the parameters specified in Table 3.1 and the power was scaled such that the resultant SNR was $25\ dB$. A 16 bit quantiser was setup with appropriate full scale dynamic range.

Each of the signals was split into ten one second segments, and five features were calculated for each of the segments: Mean Absolute Value (MAV), Mean Frequency (MeanFreq), Slope Sign Change (SSC), Waveform Length (WL), and Zero Crossings (ZC). Each of the features was calculated via the methods provided by Chan [33]. All the features except MeanFreq were used as 'conventional features' in the original study and were reported to be widely used in SEMG studies [6]. MeanFreq was added because it is a popular metric used in SEMG studies, particularly relating to fatigue analysis [44].

Features such as MAV, and MeanFreq were observed to be largely unaffected by the instrumentation effects but ZC, WL and especially SSC are sensitive to high frequency

content, and thus noise can significantly affect these values. However, when the noisy signal was filtered, the values were observed to be indistinguishable from the clean signal.

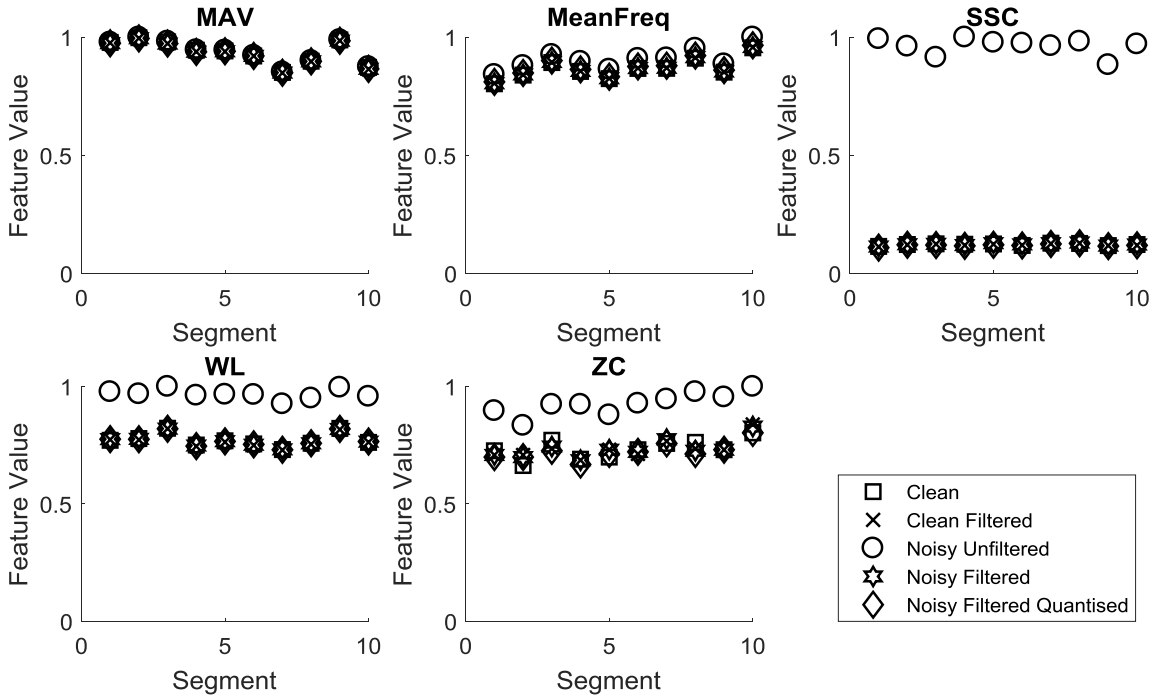


Figure 3.6: Effects of instrumentation on signal features

3.4 Summary

A model capturing baseline noise and the processes of filtering and quantisation were integrated into Myosim. Users can specify various parameters in a GUI that can be accessed via the Myosim Parameters window as shown in in Figure 3.7. Amplifier gain, band-pass and anti-aliasing filter and quantisation parameters can be set. The GUI also allows users to add baseline noise model parameters. Assuming instrumentation is setup properly, instrumentation effects were observed to be negligible. Future work to implement electrode geometry in Myosim may provide additional insight.

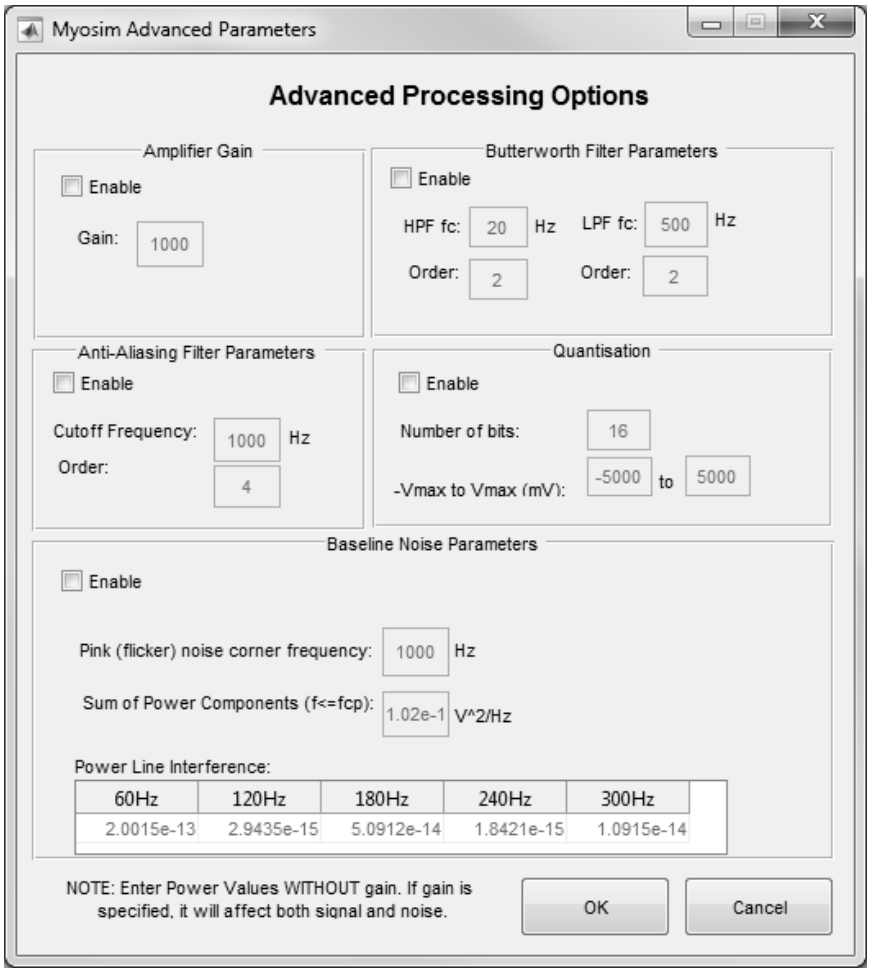


Figure 3.7: Myosim GUI for setting instrumentation effects

Chapter 4: IMPROVING MYOSIM MODEL PARAMETER SETTINGS

Simulated control signals offers an advantage over recorded control signals for use in PANDA in that it can be guaranteed to be clean, but it must sufficiently represent the characteristics of SEMG records to be useful. The model used in the Myosim tool is well supported in literature [12, 13, 15, 46, 47, 48] providing a rationale for its use. However, even given validity of the model, several model parameters must be set before the simulator can produce intended output signals. Some of these parameters are invariant and embedded into the tool's implementation of the model (e.g. intracellular and extracellular conductivities, transverse and longitudinal conductivities, etc.). Others may vary with the contraction being modelled and are user defined (e.g. number of fibres, number of motor units, conduction velocity, and parameters related to muscle geometry). This work aimed to improve on the selection of both, where possible.

4.1 Setting Embedded Model Parameters

The model used to simulate SEMG was thoroughly described in Chapter 2. Using that model, SFAPs are generated according to equation 2.1, repeated here for convenience:

$$SFAP(t) \cong K1 \times s(t) * [h_L(t, d, r, v, L_L) + h_R(t, d, r, v, L_R)\,]$$

$$\text{where } K1 = {\sigma_i \theta^2}\Big/{8\sigma_e}, \text{ and } \theta = {a}/{v}$$

$$h_L(t, d, r, v, t_L) = \begin{cases} h\left(t + {d}/{v}\right), & 0 \le t \le t_L \\ 0, & otherwise \end{cases} \qquad \textbf{4.1}$$

$$h_R(t, d, r, v, t_R) = \begin{cases} h\left(t - {}^{d}/_{v}\right), & 0 \le t \le t_R \\ 0, & otherwise \end{cases}$$

$$h(t) = \frac{t}{\left[\frac{\sigma_l}{\sigma_t}\left(\frac{r}{v}\right)^2 + t^2\right]^{3/2}}$$

Because previous work using Myosim considered only normalised signals, the embedded value for the scaling factor in equation 4.1 was set to $K1 = 1$, along with the conductivity ratio (ie. ${}^{\sigma_l}/_{\sigma_l} = 1$). However, this work required non-normalised amplitude values, so these parameters had to be fully specified. Literature [15] detailing what is known about the values was also highlighted in Chapter 2. Based on those recommendations, the following adjustments were made to the embedded model parameter values:

	Original Model Values	**Updated Model values**
σ_i	not defined	$0.45\ s.m^{-1}$
σ_e	not defined	$2.5\ s.m^{-1}$
a	not defined	$25\ \mu m$
v	not defined	$4\ m.s^{-1}$
σ_l/σ_t	1	5
$K1$	1	$8.79 \times 10^{-13}\ s^2$

Table 4.1: Updated model parameters (note: a and v are known to vary but in proportion to each other such that the ration a/v remains relatively constant)

To confirm that the chosen values yielded expected amplitudes, a single MU was generated close to the surface and the resultant amplitude was compared to values reported in the literature. A literature review was conducted to find Intramuscular EMG (IEMG) studies that reported values for the amplitude of a single MU from human muscle. However, most studies involving MU observation focus on other characteristics, typically firing rates, and thus ignore amplitudes altogether, or report normalised amplitudes. A

single IEMG study [36], clearly reported experimentally obtained values of the amplitudes of a human MU and this was used as reference for this work. The values obtained from the simulation was found to be smaller than reported in the study by two orders of magnitude even when $N = 700$ fibres were included in the motor unit. To account for this discrepancy, the embedded value for $K1$ was scaled by a factor of $K_{scale} = 500$.

Differences between observed measures and model outputs are likely the results of limitations on the assumptions made in the model. One assumption is that each fibre is in an infinite volume conductor. Practically, the muscle is finite in volume and the boundary conditions affect the signal. Another assumption is that the volume conductor is filled with only one type of conducting material. The effects due to neighbouring fibres and other tissues such as fat, tendons and other organic material such as bones are not considered in the model and could have a significant effect on the output signal.

Using a model which simplifies the volume conductor offers a reduction in computational complexity, but at the cost of ignoring some variability across amplitude outputs. Nevertheless, preliminary analysis with embedded features set to the parameter values specified for the simplified model indicated that the model was sufficient to produce simulated SEMG which matched recorded SEMG, as described in the rest of this chapter.

4.2 Choosing User-defined Parameters Using a Genetic Algorithm

While physiological limits can provide a general range for acceptable user-defined parameter values, the researchers who first proposed PANDA for use with SEMG found it problematic to isolate a set of parameters that could sufficiently represent SEMG records.

Given the general range, a brute-force approach to specifying parameter values would likely yield combinations of parameters that produce acceptable control signals, but the computational overhead associated with generating signals with all possible combinations would be excessive. The work detailed below describes an alternative approach investigated, employing a genetic algorithm to isolate combinations of parameters fit for use.

Genetic Algorithms (GAs), first proposed by John Holland in 1975, are a subset of Evolutionary Algorithms (EAs) which were inspired by biological evolution and use the concepts of chromosome propagation and random mutation to find optimal solutions(s) for a given problem [34]. GAs are good at optimising a wide range of problems and generally find acceptable solutions in a relatively short time [35]. They are often a preferred strategy for optimisation because they avoid getting caught searching about local minima/optima [41].

The basic goal of the GA is to optimise some function $f(x_1, x_2, \dots, x_N)$, commonly called the 'objective function'. A common scenario involves finding some combination of values for the parameters $x_1, x_2, \dots, x_N$ in a given search space that minimises (or maximises) that function. There are several approaches to implementing GAs depending on the nature of the problem being optimised and resource constrains. For instance, if the problem being optimised has a simple objective function, then a canonical GA (Holland's original algorithm) can be used. If there are multiple objective functions that need to be simultaneously optimised, common strategies include: combining the objective functions

to form a single function to be optimised, or optimising subpopulations to specific objective functions [37]. Also, if the population is very large and several sets of optimal values are to be found, then a parallel GA implementation such as an Island GA can be used, which runs the GA on several subpopulations in parallel [38]. Though the design of a GA depends on the application and implementation, the GA is described generally below, in terms of the canonical algorithm used in this work.

To optimise the objective function, the parameters, also referred to as 'genes', are combined to form an array like structure called a 'chromosome'. The algorithm starts by initialising a random value for each gene in the chromosome to create an 'individual', also commonly referred to as a 'member'. Initial values are selected from within a specified range of allowable values for each gene. Several individuals are generated to create an initial 'population'. An example population is shown in Figure 4.1 which contains $i = 1{:}\, I$ individuals, each of which contain $n = 1{:}\, N$ genes. Each value X_{in} is a random value for gene n of individual i.

X_{11}	X_{12}	...	X_{1M}
X_{21}	X_{22}	...	X_{2M}
⋮	⋮	...	⋮
X_{N1}	X_{N2}	...	X_{NM}

Figure 4.1: Example initial population

Individuals in the population are evaluated using the objective function $f(\cdot)$ to produce outputs $O_1, O_2, \dots, O_I$. The algorithm then determines the 'fitness' of each individual using

the output. To do this, each output O_i is mapped to a score F_i that provides an indication of the individual's strength as a solution. This mapping is done using a 'fitness function' $F(\cdot)$, i.e., $F(O_i) = F_i$. A higher fitness score indicates a better chance that the individual is close to an optimal solution. Fitness values are typically non-negative. Commonly used fitness functions include $F(Y) = Y$ for the case of maximisations and $F(Y) = \frac{1}{1+Y}$ for the case of minimisations [41]. If one or more individuals has a fitness score exceeding a predetermined threshold, then the process is stopped and the individuals producing the required scores are said to be the solutions. Otherwise, the algorithm proceeds to the next stage. The flowchart shown in Figure 4.2 outlines the typical GA process.

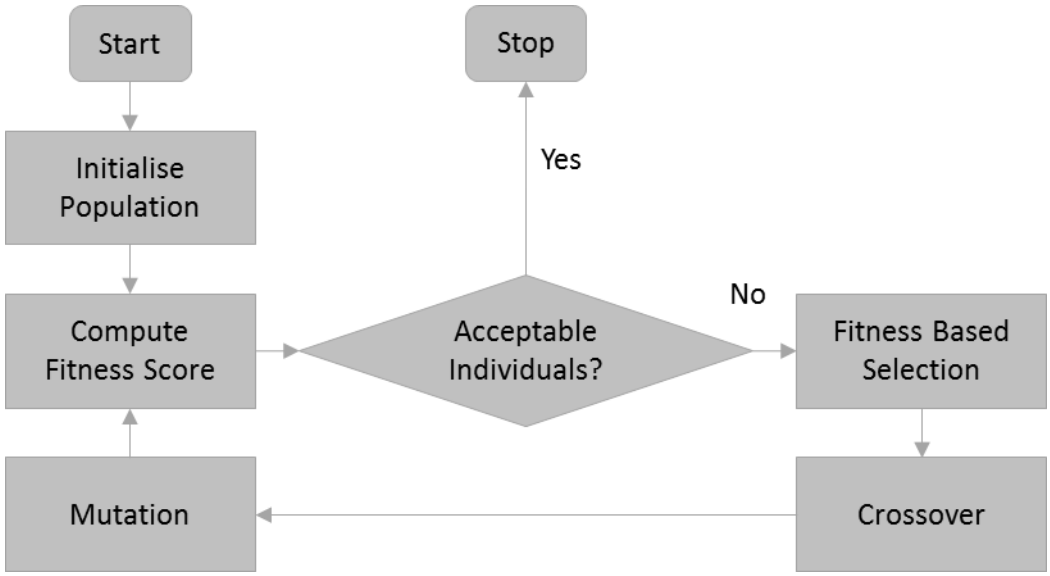


Figure 4.2: Stages in a typical GA

If no individuals are fit, the GA mutates a selection of the population to try and generate fitter individuals. To do this, the GA selects individuals to form an 'intermediate population', based on their fitness scores. Typically, a fitness proportional selection

operator called a 'roulette wheel selector' is used to select individuals based on their fitness score. This is a preferred selection operator since it is faster to compute and also has a better chance of preserving diversity than other algorithms such as tournament selection or rank based selection [39].

Using the roulette wheel selector, An individual i with fitness score F_i has a 1 in $F_i/\bar{F}$ chance of being selected, where $\bar{F} = \sum_{i=1}^{I} F_i$ is the total fitness of the population. The spinning of the wheel is simulated by sampling a number from a uniform random process such that the number lies between 0 and 1. The first individual whose cumulative fitness yields a value greater than or equal to the sampled number is selected for the intermediate population. The cumulative fitness for an individual is the sum of fitness scores up to and including the score of the individual.

Figure 4.3 shows an example roulette wheel with three individuals in which Individual 1 contributes 0.56 (or $56\ \%$) of the total fitness, Individual 2 contributes 0.33 (or $33\ \%$) while Individual 3 contributes 0.11 (or $11\ \%$). The cumulative fitness scores are 0.56 for individual 1, $0.56 + 0.33 = 0.89$ for individual 2, and $0.56 + 0.33 + 0.11 = 1$ for individual 3. Given these scores, if the sampled random number is 0.3, then Individual 1 is selected for the intermediate population since its cumulative fitness is $0.56 > 0.3$. Similarly, if the sampled random number is 0.7, then Individual 2 is selected since it is the first Individual to have a cumulative fitness score greater than 0.7 (i.e. $0.89 > 0.7$).

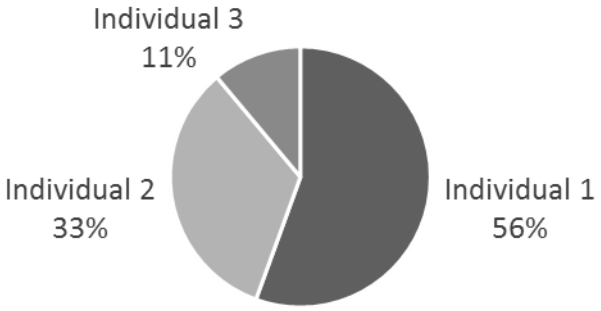


Figure 4.3: Roulette Wheel Selection

The higher the contribution to the overall fitness, the more likely that an individual is selected to place a copy of itself in the intermediate population. This process is repeated as many times as there are individuals in the population. Note that the intermediate population can contain multiple copies of individuals since the random selection can result in the same individual being selected multiple times.

In order to find better solutions from the given population, a crossover operation is performed on the intermediate population. Crossover involves swapping genes of random individuals in the population to possibly produce individuals that are fitter than the individuals that produced them. Just as an offspring inherits genes from both parents, the crossover operator swaps genes between two random pairs of chromosomes to produce two new individuals. The number of individuals that undergo crossover is determined by the crossover rate, p_c. Two individuals are chosen, then 'split' at one or more crossover points, after which alternate segments are exchanged. An example crossover with 2 crossover points is shown in Figure 4.4 where the two individuals contain 6 genes each.

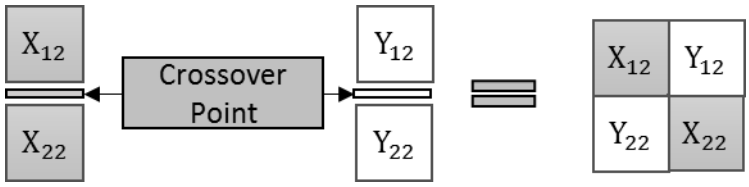


Figure 4.4: Crossover operator

The number of crossover points is typically 1 or 2, and the locations of the crossover points are usually randomly sampled from a uniform random distribution [41], [42]. Note that the individuals that were not selected for the crossover remain as members of the intermediate population while the chosen individuals are replaced by the offspring.

The final stage of the genetic algorithm is the mutation operation, where genes in the population are selected at random and the values are reinitialised from the given search space. A mutation rate is used to set the probability p_m that a gene of an individual will be selected for mutation. The premise behind mutation is to add diversity to the existing population and allow the algorithm to escape any local minima/optima. This can especially be problematic if the algorithm has undergone several iterations, potentially causing the population to become too homogeneous.

The configuration parameters of the GA such as population size, p_c, p_m, etc., influence the performance of the algorithm and it has been suggested that the GA can perform relatively well with a range of settings [43].

After mutation, the intermediate population becomes the population for the next iteration. Then the process of computing fitness, crossover, and mutation is repeated until either the algorithm achieves convergence or a set number of iterations is complete.

Thus, the algorithm iteratively selects the fittest individuals from a population, swaps genes between a few of these individuals and randomly mutates others to try and reach an optimal solution.

4.2-1 Setting up the Genetic Algorithm

4.2-1a Configuring the Genetic Algorithm

In the context of this work, a GA was used to identify a set of Myosim user-defined parameters which would yield an output signal that matches an SEMG record. Matching aims to ensure that the control signals used with PANDA are realistic. As delineated in Chapter 2, the resultant signal is determined by the summation of several MUAPTs, and the characteristics of the MUAPTs are determined by the SFAPs associated with the MU. The tissue filter described in equation 4.1 determines the characteristics of the SFAP and is dependent on the depth and CV of the fibre. Thus, four Myosim parameters were chosen to be optimised in this work (i.e. included in the chromosome as genes): Number of fibres (N), Number of MUs (M), Depth offset (D_{OFF}), and CV (v). Other user-defined parameters such as signal duration (T) and sampling rate (f_s) were set to match what was known about the SEMG record.

The parameters listed above do not fully specify all of the physiological parameters included in the model. Besides the parameters embedded in the implementation, Myosim also allows users to assign values to other parameters which influence outputs, including fibre length (termination points), electrode locations (relative to innervation points),

muscle radius, and firing rate. These parameters were pre-set to typical values for the contracting muscle being matched as delineated in Table 4.2.

As also indicated in the table, Myosim allows users to specify limits for including random variation about many specified values. This models typical variation across fibres. All of the random variation was pre-set and mostly applied to the pre-set parameters, but was also applied to CV. For instance, if the GA suggested a conduction velocity $v = 4\ m.s^{-1}$, v was specified as $v = 4 \pm 0.2\ m.s^{-1}$, and the simulated value of v for each fibre was randomly determined (from a uniform distribution), to take on any value between $3.8\ m.s^{-1} \leq v \leq 4.2\ m.s^{-1}$. As a result of all of this variation, each solution identified by the GA actually yields a set of output signals, only one of which represents the output evaluated for fitness by the GA. Whether or not all signals in the set provided reasonable representations of SEMG records, was left to be determined through analysis.

Parameters	Value	GA Parameters	Search Space
Distal termination	$-220 \pm 5\ mm$	N	$100{:}\ 1{:}\ 400$
Proximal termination	$180 \pm 5\ mm$	M	$50{:}\ 1{:}\ 200$
IP dispersion	$\pm 5\ mm$	D_{OFF}	$1{:}\ 0.1{:}\ 15\ mm$
Radius of limb	$40\ mm$	v	$3{:}\ 0.1{:}\ 6\ m.s^{-1}$
Firing Rate	$10\ to\ 30\ *\ 0.15\ Hz$		
Electrode Locations	$50\ mm\ and\ 60\ mm$		
Sampling frequency	$5\ KHz$		
Signal duration	$5\ s$		
v dispersion	$\pm 0.2\ m.s^{-1}$		
MU Depth range (D)	$5\ mm\ to\ 21mm + D_{OFF}$		
Alignment	$-10\ mm\ to\ 10\ mm$		
MU dispersion	$\pm 5\ mm$		

Table 4.2: Tool parameters and GA search space

Given this problem set up, the canonical GA was an ideal candidate for generating solutions since it is relatively fast, easy to implement, and is designed to generate single

(1 chromosome) short (a few genes) solutions. A nominal population size of 100 was initially generated from the given search space. A roulette wheel was used to select individuals for inclusion in the intermediate population. The crossover rate was set to $p_c = 10\ \%$, with the number of crossover points set to 1. The mutation rate was set to $p_m = 5\ \%$. Since more than one solution can be produced for this problem, the GA ran until 5 solutions were identified. The function used to evaluate fitness is described in the next section.

4.2-1b Selecting the Fitness Function

Given the GA configuration, an individual chromosome represents a combination of four Myosim parameters which can be adapted to generate a simulated signal which matches an SEMG record. The simulated signal is the output which is mapped to a fitness score used by the GA to evaluate the individual's strength as a solution during the selection stage. For this work, the mean absolute error (MAE) between the power spectral densities of the record and simulated signals was used as a metric to compute the fitness score. The metric was normalised to the total power in the record being matched as delineated in equation 4.2, where $\phi(f)$ is the power spectral density of the signal. The normalisation was included so that a common threshold for fitness could be established across different records. Power spectral densities were obtained using Welch's averaging method using half second, non-overlapping windows.

$$MAE = \frac{|\phi_{Rec}(f) - \phi_{Sim}(f)|}{\sum \phi_{Rec}(f)} \qquad \textbf{4.2}$$

Resultant MAE values were used to assign a fitness score according to:

$$F(E_i) = \frac{A}{A + E_i} \tag{4.3}$$

where E_i is the MAE for individual i. The scaling factor A was set to the threshold for acceptance of a solution so that the following conditions were met:

$$F(E_i) \approx \begin{cases} 0, & E_i \ll A \\ \frac{1}{2}, & E_i = A \\ 1, & E_i \gg A \end{cases} \tag{4.4}$$

Thus, any fitness score greater than or equal to $^1/_2$ was accepted as a solution.

The threshold was set by establishing the minimum expected variation between two SEMG assumed to come from the same contraction. This variation represents sampling variation and is unavoidable when comparing SEMG records. To establish this value, a long signal was generated in Myosim with $T = 500$ seconds. This signal was split into 100 5-sec segments. One of these signals was arbitrarily assigned as the reference signal, and the rest were used as comparison signals. For each segment, a power spectral density was calculated. These spectra were used to calculate MAE between the reference and comparison signals. The mean (μ_{MAE}) and standard deviation (σ_{MAE}) of the 99 MAE values were used to estimate the minimum expected variation. Results yielded an expected MAE of $E_{MAE}(\mu, \sigma) \cong (0.0054, 0.00035)$. To accommodate for additional variation which can exist as a result of physiological parameter changes during a contraction (e.g.

switching active fibres) the threshold was broadened to $A = 1.15 \times (\mu_{MAE} + 3 \times \sigma_{MAE})$, yielding $A \cong 0.0075$.

4.2-1c Establishing Performance Metrics

Comparing the power spectral densities of the records and simulated output signals provides a good indication of fitness, but it compares only a partial representation of the signals. To more completely evaluate how well each solution's simulated output signals matched the records, several other output signal features were also inspected. These included the same five features (MAV, MeanFreq, SSC, WL, and ZC) used to inspect the effects of instrumentation noise as defined in Chapter 3. As in the work described in Chapter 3, each was calculated via the methods provided by Chan as delineated in Appendix 1 [33].

To establish a range of expected values for any particular record, the features were computed for five 1-second segments extracted from the record and the mean and standard deviation for each feature was calculated. The expected range was then set to $\mu \pm 3\sigma$. Then features extracted from simulated outputs were inspected against these ranges. If the features of the simulated signals fell within the ranges specified, then the simulated signals were considered to be a good match to the records.

4.2-2 Evaluating the Genetic Algorithm

4.2-2a Refining the Fitness Function

Before evaluating the GA for its ability to match SEMG records, a preliminary experiment was conducted to refine the fitness function. Since most of the SEMG power resides in

the frequency band less than $150\ Hz$, it made sense for this application to band-limit MAE calculations for fitness evaluation to frequencies ranging from $0 \leq f \leq 150\ Hz$. However, to avoid the influence of any noise that may be present in the SEMG record which is being matched, a further reduction in frequency bands was investigated. Since the goal was to avoid noise prone regions, the reduced range avoided low bands susceptible to the effects of motion artefact, and the mid band susceptible to power line interference. Thus, the reduced range included only the frequency bands from $20\ Hz \leq f \leq 58\ Hz\ \&\ 62\ Hz \leq f \leq 150\ Hz$. That is, the bands from $0\ Hz - 20\ Hz$ and $58\ Hz - 62\ Hz$ were ignored in the calculation of MAE.

An analysis was conducted comparing output signals produced from solutions identified by the GA using full-range and reduced-range fitness functions. The records used for matching in this comparison were judiciously collected to avoid noise according to the procedures described below. Since there was no significant noise in the records being matched, the comparison between solutions was useful in identifying any differences in solutions caused by the reduced-range fitness function. If no differences were noted, the reduced-range function could be used in fitness evaluation with no loss of veracity, but with the benefit of preventing noise-prone regions from influencing solution generation. That is, even if the GA were to match a record with noise, the noise would have no influence on the solution.

A total of 24 records (12 subjects $\times 2$ contraction levels) were matched in this comparison analysis (levels were $15\ \%$ and $25\ \%$ of each subject's maximum voluntary

contraction). These signals were collected as part of another SEMG study aimed at investigation variability in SEMG features (the same one as mentioned in Chapter 3) [31]. It was important for that study to limit noise in the records as much as possible, so the records made excellent candidates for this comparison analysis. The instrumentation was set up to filter signals below $20\ Hz$ and above $500\ Hz$ thus minimising the effects of low and high frequency noise. The gain was adjusted to utilise the full dynamic range of the ADC while making sure the amplifier was not saturated. An instrumentation amplifier with CMRR $> 100\ dB$ was used to mitigate power line interference. Multiple channels were used to record the signals so that a channel could be selected for processing which minimised proximity to the innervation zone and tendon regions.

To further ensure that the records was noise-free, they were also pre-processed through the cleanEMG noise detection tool introduced in Chapter 1 [5]. This tool checks for power line interference, quantisation noise, saturation, motion artefact, and any unusual spectral deformations. Any records that was flagged as noisy by the tool was discarded. Two of the 24 records were flagged with a spectral deformation which was visually confirmed. These records, both from the same subject, were discarded from the set. A five second segment from the middle of each of the remaining 22 records from 11 subjects was selected for further analysis.

Figure 4.5 provides a visual comparison of the 5 performance metrics calculated from output signals produced from solutions identified for each record with the full-range and reduced-range fitness functions.

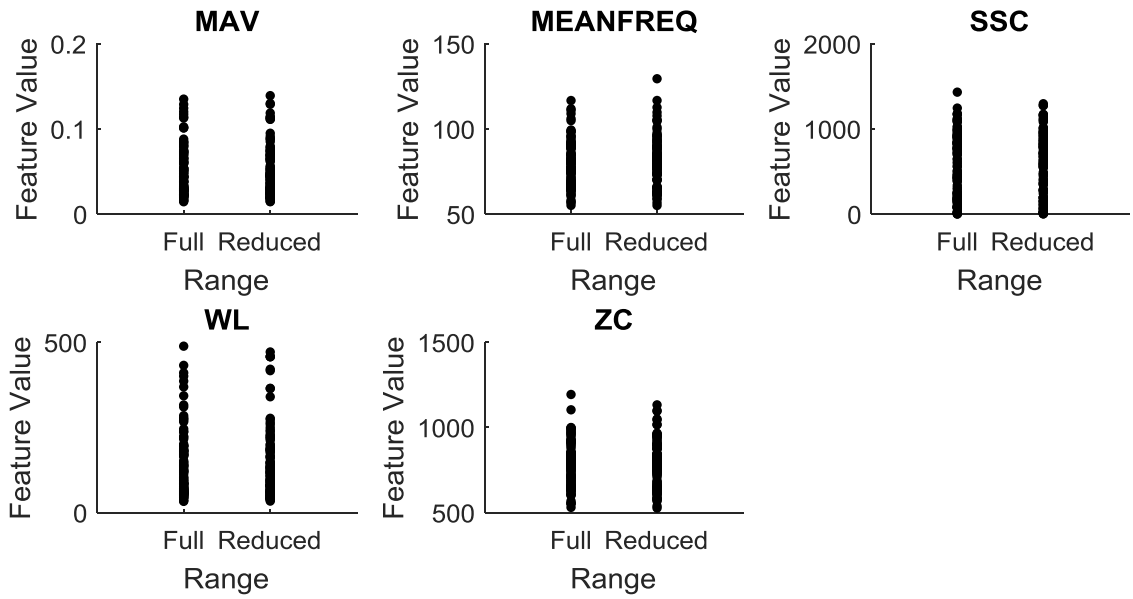


Figure 4.5: Feature comparison between the two bands

A two-way ANOVA was computed with range as a 2-level factor, and records ($N = 22$) as a block. No significant differences were revealed between any of the five features for signals produced from the full and reduced ranges ($\alpha = 0.05, 0.2 < p < 0.9$). Thus, the reduced-range fitness function was adopted for use in this work.

This analysis compared output signals simulated based on solutions produced with two slightly different fitness functions, but made no comparison between the simulated outputs and the records they were simulated to match, or evaluation of the veracity of the solutions. To validate the solutions, such a comparison was conducted, as delineated in the rest of this chapter.

4.2-2b Validating the Fitness Function with simulated signals

To first test the validity of the fitness function, a simulated signal was generated in Myosim and the GA was setup to match this reference signal. All the pre-set values were

set to match. Then, the values of the simulation parameters used to generate the reference signal were compared to the GA solution parameter values.

Results from this investigation were expected be difficult to interpret because of the way Myosim includes random fibre-to-fibre variation when generating signals with a given set of solution parameters. This predictably leads to deviations from the reference beyond what would be expected simply as a result of the non-zero fitness threshold. Since Myosim allows for random variation, each solution yields multiple output signals, each with a different set of fibre-to-fibre parameter values. There is a chance that two or more solutions can give rise to a similar set of fibre parameter values and thus generate a similar output signal (similar enough for the chromosome to be considered fit). The variation, in essence, allows for a solution with different parameters from the reference to give rise to a matching output signal.

Table 4.3 lists an example set of reference signal parameters together with the average value of 100 solutions obtained from the GA trying to match it. Also listed for comparison are the expected values of the initial population which represent random values selected from the search space. For visual comparison, Figure 4.6 shows the spectrum of one of the matched signals superimposed on the spectrum of the reference signal. These results were typical of the results obtained when conducting this comparison with other reference signals.

GA Parameters	Reference Signal	GA Results	Random Solutions
N	150	188	250
M	60	112	110
D_{OFF}	11	9.9	8
v	5.5	4.7	4.5

Table 4.3: Simulation Matching

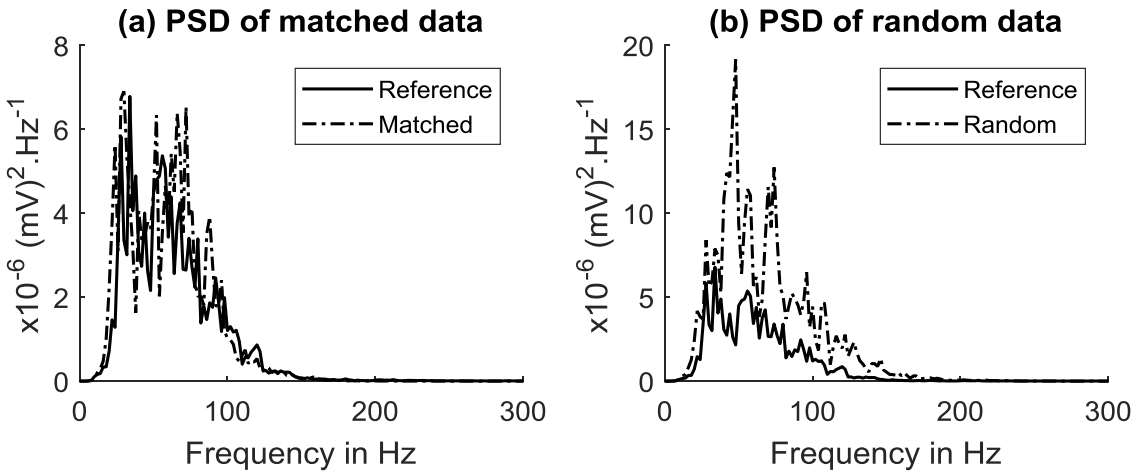


Figure 4.6: Matched signal spectrum. Parameters for (a): $N = 139, M = 193, D_{OFF} = 8.2, v = 4.3$, (b): $N = 247, M = 107, D_{OFF} = 1.7, v = 4.4$

As indicated in the example provided, there was some suggestion that the algorithm does converge towards the reference configuration when compared to the expected values from the initial population - three of the four solution parameter averages are closer to the reference than the initial population's parameter averages.

In trying to interpret the results, another important observation surfaced. There was indication that a trade-off in focus among solution parameters was also contributing to the variation in the results. For instance, to increase amplitude of the simulated SEMG to improve fitness, the GA can either increase the number of MUs, decrease their depths, or choose some balanced combination between these changes. Different solutions reflect

a different focus on which of these paths the GA followed. This kind of trade-off yielded another source of deviation in solution parameter values from the reference.

The different sources of variation resulted in a solution pool rather than a single solution (that is, different combinations of values could be equally fit), but there is indication that the pool of solutions converged toward the reference. The validity of the solutions with respect to producing output signals that match SEMG records is evaluated in the next section.

4.2-2c Evaluating GA solutions

The records used to refine the fitness function was used again to evaluate solutions produced by the GA. This data set consisted of 22 5-sec segments extracted from the records captured from 11 subjects at 2 contraction levels. The GA was used to identify 5 solutions for each record and two 5-sec simulated signals were generated from each solution, yielding a total of $5 \times 2 = 10$ simulated signals for analysis for each record (or $10 \times 2 = 20$ output signals for each subject contracting at 2 levels). The five performance metrics were computed for all record segments and simulated signal segments, both in one-second increments. The features were then visually inspected to see if the simulated signal features fell within the region defined by the mean and standard deviation of the features from the records. Values which fell within three standard deviations from the mean were considered to be well-behaved.

Figure 4.7 displays a comparison between the features of the record and the features of the simulated signals for the stronger contraction level captured from Subject 5.

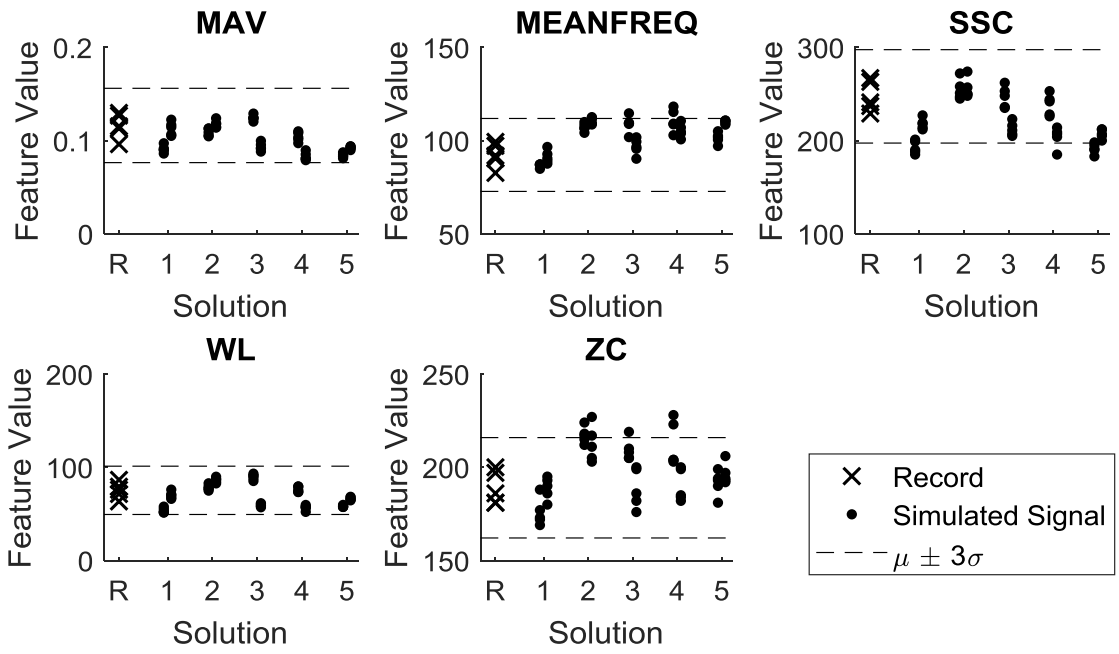


Figure 4.7: Typical distribution of simulated signal features compared to the record

These results are representative of results observed across subjects and levels of contraction. As expected, all MAV values calculated from the simulated signals (dots) fell within the expected range defined by the MAV values of the record (Xs). This was expected because the fitness function was based on the power spectrum, which has a direct relationship to MAV. While a few output signal segments fell outside expected ranges for the other features, all were within or close to the boundary and were considered acceptable since only 5 data points were used to establish the range.

Some variation was expected in the features of the simulated signals as compared to the record, but Figure 4.8 depicts an atypical result which was not expected, observed for the weaker contraction level from Subject 4. Most of the simulated signal feature values fell well outside expected bounds and this was representative of both contractions for that subject. These atypical results were most pronounced in the SSC feature.

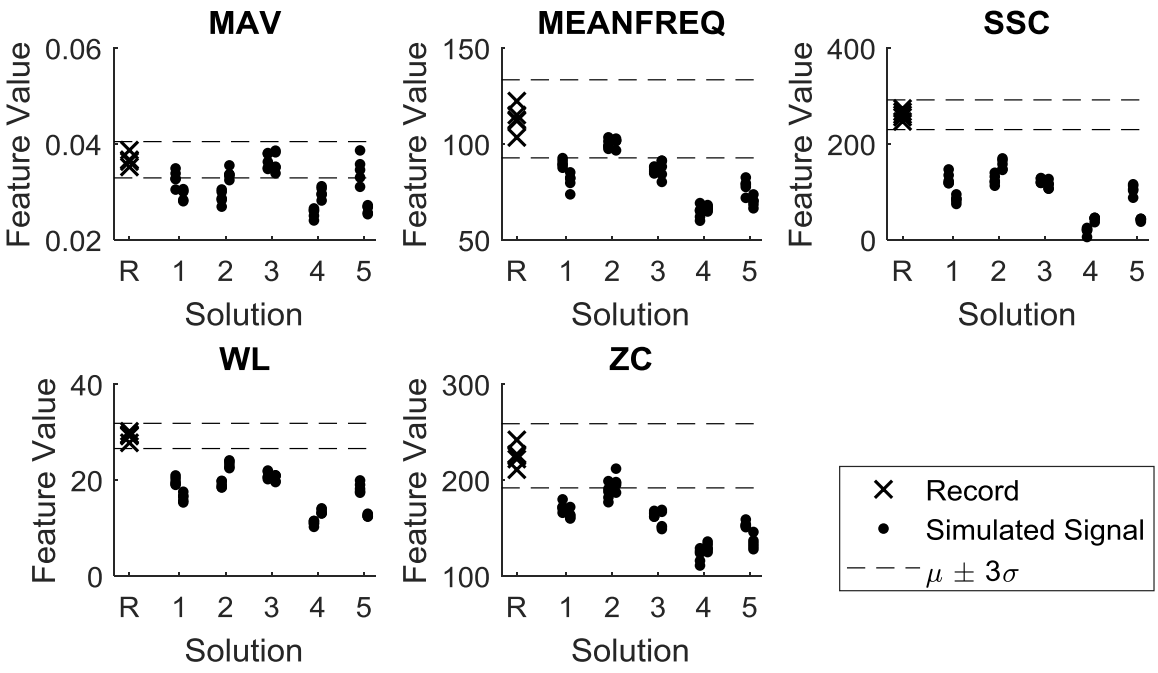


Figure 4.8: Atypical distribution of simulated signal features compared to the record

To better understand these results, feature values from the records for this subject were compared to the feature values across all well-behaved records using a two-sample Kolmogorov-Smirnov nonparametric test.

Table 4.4 delineates the results obtained from the comparison and includes results from both contraction levels. The test suggests that the features of the records corresponding to subject 4 were statistically different as compared to the well-behaved records.

Feature	Weak Contraction		Strong Contraction	
	Well-behaved	**Subject 4**	**Well-behaved**	**Subject 4**
MAV	0.038 ± 0.021	0.037 ± 0.001*	0.74 ± 0.033	0.61 ± 0.002
MEANFREQ	80.25 ± 7.04	113.11 ± 6.7**	79.28 ± 6.83	108.36 ± 4.41**
SSC	116.08 ± 53	261 ± 10.32**	177.68 ± 43.1	287 ± 20.31**
WL	21.16 ± 11.84	29.22 ± 0.88*	41.32 ± 19.7	45.7 ± 2.14
ZC	148.98 ± 10.9	225.4 ± 11.1**	157.5 ± 15.67	221.6 ± 14.22**

Table 4.4: Comparison between features from Subject 4 records and all other well-behaved records. Asterisks indicate statistically different results: ** Indicates $p < 0.0001$ and * indicates $p < 0.012$.

Ad hoc examination of channel location with simulated signals indicated that this could have been a cause for the discrepancy. Figure 4.9 shows the features of an example signal as observed from different channels. The figure clearly indicates channel effect on features, which is well supported in the literature [45]. MAV was observed to be the least affected by channel selection, while SSC was observed to be most affected. The investigation suggests that the record being matched in the atypical result may have been selected from a channel close to the innervation zone. However, literature also indicates that features like MAV and SSC can be affected in similar ways by other factors such as the number of deep fibres contributing to the measurement. More work needs to be completed to more fully understand this anomalous result.

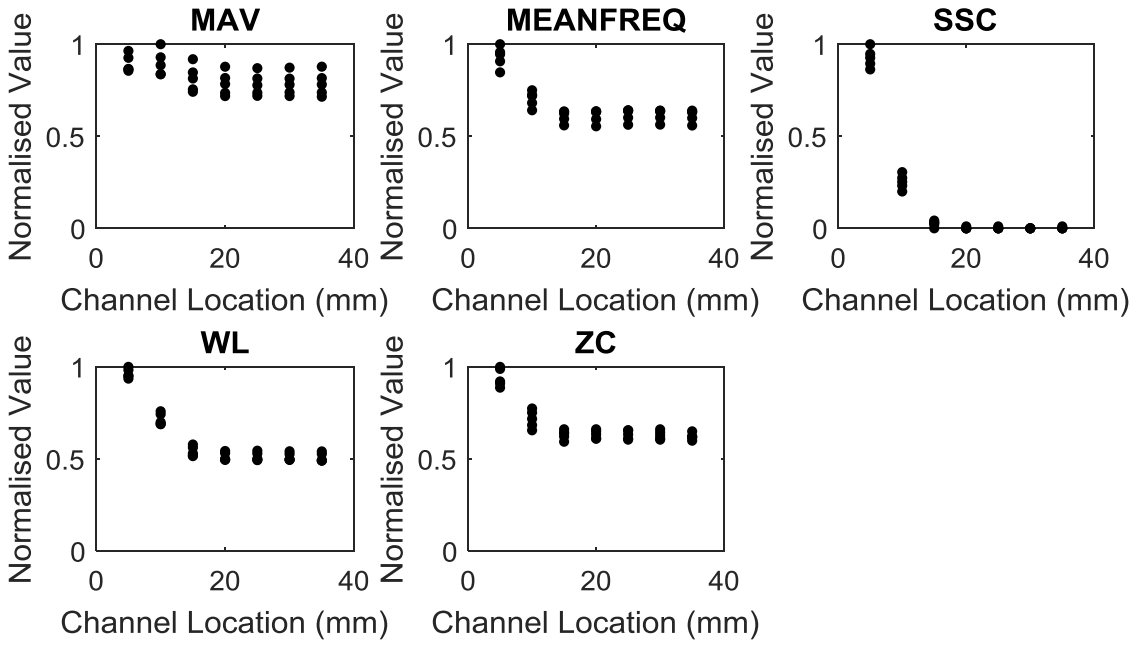


Figure 4.9: Effect of channel on features

4.3 Summary

A GA was setup to find certain Myosim parameters that produce simulated signals that matched SEMG records. The parameters were used to generate simulated signals whose features were compared to the features of the records.

Future work is suggested to explore the possibility of optimising all the physiological parameters associated with each fibre and MU in the muscle that matches a record. Such an implementation allows the possibility of predicting the physiological parameters associated with an SEMG signal and decompose the signal into MUAPs and SFAPs. Although the total number of parameters being optimised using this approach is relatively higher, it could be mitigated to a certain extent using a parallel GA implementation. Exploration of a different fitness function is also suggested here that can better capture other features such as SSC. This could be accomplished by the use of a more complex fitness function that can combine spectrum with other signal features.

CHAPTER 5: EVALUATING THE IMPACT OF MYOSIM UPGRADES ON PANDA-BASED SEMG SIGNAL QUALITY ANALYSIS

In Chapter 2, the original model used in Myosim was fully specified, and values for model parameters originally ignored in the model were quantified, based on what is reported in the literature. In Chapter 3, instrumentation effects were investigated and modelled. In Chapter 4, a genetic algorithm was used to better determine user defined Myosim parameters for improved matching with SEMG records. This chapter describes an investigation of the effects of these upgrades on the performance of PANDA. Since instrumentation effects were found throughout the work reported thus far to be negligible compared to effects of the other upgrades, they were ignored in the work which follows.

5.1 Methods

To conduct this study, the simulated signals generated in Chapter 4 were used to construct the clean control signals for use with PANDA, while the 22 records were used as test signals. The clean records were segmented into 1 second segments giving rise to 330 segments (22 records $\times$ 15 seconds per record). Each segment was run through cleanEMG which marked 16 as noisy so those segments were discarded, yielding 314 test segments. Similarly, the simulated signals were segmented into 1 second segments resulting in 1100 segments (22 records $\times$ 5 solutions per record $\times$ 2 signals per solution $\times$ 5 segments per signal).

The aim of this work was to evaluate the simulated control signals in terms of their ability to help PANDA interpret clean records as clean, and noisy records as noisy. Thus, noisy test segments were also required. To control the amount of noise present in each test segment, simulated noise was added to the clean test segments. The test segments were contaminated to simulate the presence of 3 different types of noise to align with the original PANDA work [6]: Motion Artefact (MA), Power-line Interference (PLI) and Amplifier Saturation (SAT). For MA and PLI, the noise was added at SNR values starting at $-20\ dB$ and increasing to $20\ dB$ with $1\ dB$ increments. Then, the two sources of noise were combined and the process was repeated. Finally, for SAT, the percentage values saturated was increased from $1\ \%$ to $10\ \%$ with $1\ \%$ increments. PANDA was set up to analyse the test segments based on the configuration recommended in the original PANDA work [6]. Accordingly, all 9 features as specified in Chapter 2 were used as attributes, including: MAV, SSC, WL, ZC, SMR, ENT, HIST, PLI60, and PLI180. Appendix 1 details how each of these were calculated. The PANDA binning process was set to include 5 bins of equal widths. Given these settings, PANDA calculated a noise factor for each of the simulated control segments and the mean plus three standard deviations of the noise factors was used as the threshold to distinguish between noise factors from clean and noisy test segments.

Again, to align with the original PANDA work [6], a 11-fold cross-validation was conducted to evaluate PANDA's performance. The cross validation was necessary to ensure that the signals simulated to match a particular subject's records were left out of the control signals when PANDA was assessing that subject's records. Thus, evaluation was

conducted by withholding the simulated control signals corresponding to a particular subject's matches and evaluating the performance of PANDA on that subject's records with the remaining simulated control signals in action. Figure 5.1 depicts an overview of the entire process.

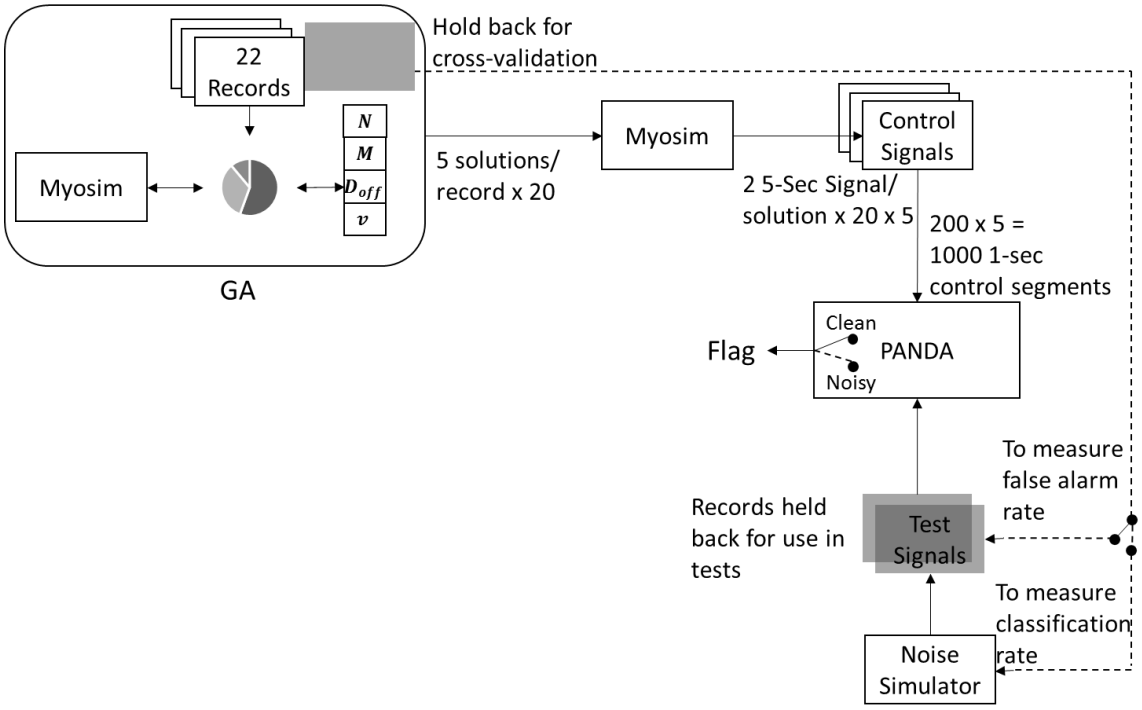


Figure 5.1: Overview of the process used to investigate impacts on PANDA.

5.2 Results and Discussion

5.2-1 PANDA performance without noise

Table 5.1 lists the results obtained from the cross validation on test segments without noise. These results delineate the capacity for simulated control signals to help PANDA identify clean records as clean. The average probability of false alarm (i.e. flagging clean records as noisy) was about $20\ \%$ when the threshold was set to $\mu + 3\sigma$ where μ and σ

are the mean and standard deviation of the noise factors of the control signals respectively. This was a marked improvement over the original PANDA work [6], which reported a false alarm rate of $96.2\ \%$ with simulated control signals.

Subject	Segments Flagged/Total Segments
1	$0/29\ (0\ \%)$
2	$13/29\ (44.8\ \%)$
3	$6/28\ (21.4\ \%)$
4	$25/30\ (83.3\ \%)$
5	$21/30\ (70\ \%)$
6	$0/30\ (0\ \%)$
7	$0/24\ (0\ \%)$
8	$0/30\ (0\ \%)$
9	$1/28\ (3.5\ \%)$
10	$1/29\ (3.5\ \%)$
11	$0/27\ (0\ \%)$
Mean False Alarm	$20.6\ \%$

Table 5.1: Cross Validation Results

Obviously, the original PANDA work [6] concluded that simulated control signals could not be used, yet the results from this work show substantial improvement. In fact, the rate reported here is more comparable to the rate reported for the in vivo control records in the original work ($9.6\ \%$). Only four subjects contributed to higher false alarm rates than that reported metric (subjects $2\ -\ 5$), and in fact the remaining subjects (subjects $1, 6\ -\ 11$) all had rates close to $0\ \%$.

Figure 5.2 (a) and (b) show plots of noise factors obtained from the cross-validation tests for subjects 1 and 2. These exemplify the strong and weak performers respectively.

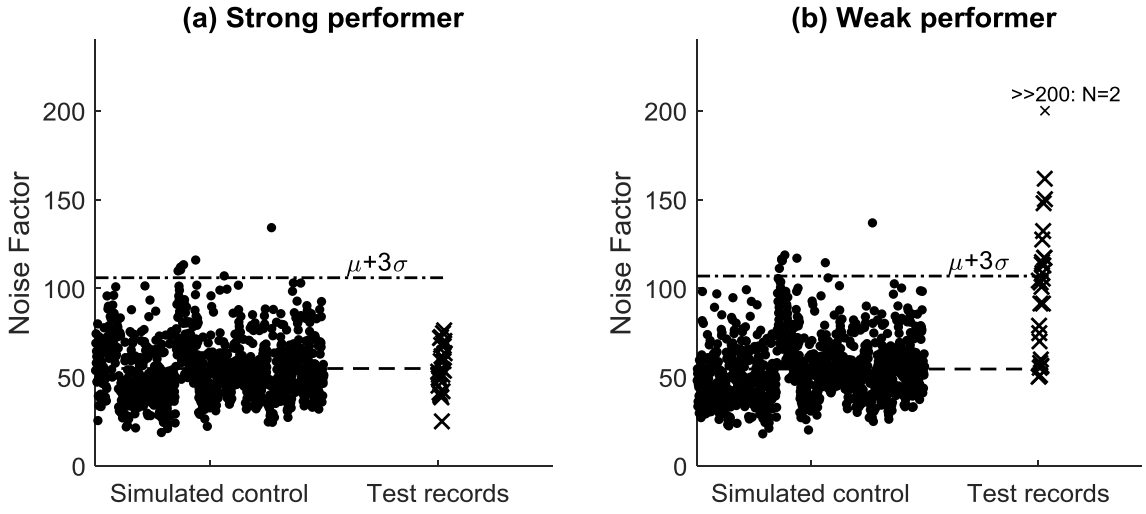


Figure 5.2: Cross-validation results – Subject 1 (False Alarm Rate = 0 %)

5.2-2 PANDA performance in the presence of noise

To test PANDA's ability to detect noisy records, the cross-validation process was repeated with the noisy test segments (MA, PLI, SAT, MA+PLI). Figure 5.3 depicts classification rates based on the average across cross-validation trials, for each level of noise. These results delineate the capacity for simulated control signals to help PANDA identify noisy records as noisy. For comparison purposes, results from the original work with PANDA [6] (using in vivo control records) were reproduced and are also superimposed on the figure. Also, to accommodate discrepancies in false alarm rates between the studies, classification results based on a threshold of σ (which reduces false alarm rates to 10 %) are also included.

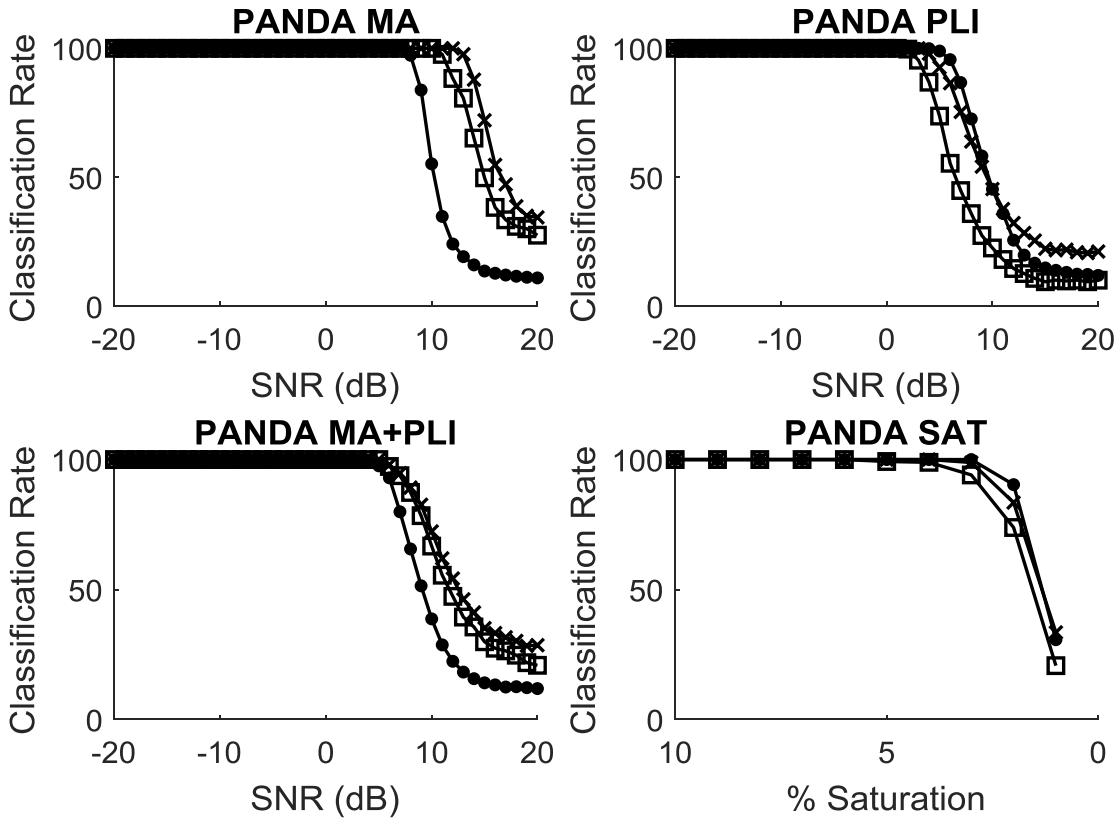


Figure 5.3: PANDA performance at varying SNR (—●— = Recorded control signals $Th = 3\sigma$, —×— = Simulated control signals $Th = 3\sigma$, —□— = Simulated Control signals $Th = 5\sigma$).

The results were encouraging. The simulated control signals appeared to perform as well as the in vivo control records from the previous study in terms of classification rate. The simulated control signals used in the current study were observed to have a slightly higher classification rate for records corrupted with MA (either solely, or in combination with PLI) compared to the in vivo control records obtained from the original study. Using the simulated control signals, the corner SNR (the maximum SNR which has a classification rate close to 100 %) for the test records corrupted by MA was observed to be $10\ dB$ and $12\ dB$, for thresholds of 3σ and 5σ respectively. This was higher than the case where PANDA was trained using the in vivo control records, where the corner SNR was observed

to be about $8\ dB$. Similarly, the corner SNR for test records corrupted by MA+PLI was observed to be $5\ dB$ for simulated control signals (both thresholds) and $3\ dB$ for in vivo control records. However, the corner SNR was slightly lower when the test records were corrupted only by PLI (for both the thresholds). The observed values were $3\ dB$ and $2\ dB$ for simulated control signals for thresholds of 3σ and 5σ respectively while the in vivo control records had a corner SNR of $4\ dB$. Similarly, the corner for test records corrupted with SAT (the minimum % saturation for $100\ \%$ classification rate) was observed to be $3\ \%$ for in vivo control records and simulated control signals with threshold of 3σ, and $4\ \%$ with the threshold set to 5σ.

The threshold can be changed depending on the application, at the expense of lower classification rate. That is, there is a trade-off between threshold and classification rate at a particular SNR value. If the application can tolerate some amount of noisy signals, then the threshold can be set higher. This way, the probability of false alarm will be reduced, but the classification rate will be reduced as well and thus some noisy signals may be flagged as clean. It is recommended that a lower threshold be set despite having a higher probability of false alarm, since this will ensure minimum chance of noisy signals being flagged as clean.

CHAPTER 6: CONCLUSION

This work aimed to investigate the poor performance of simulated control signals when used with PANDA to distinguish between clean and noisy in vivo SEMG records. The work which spawned this investigation suggested two issues which could have influenced their results: 1) Instrumentation effects which were not included in the model used to simulate the signals and 2) Insufficiently judicious selection of model parameter values. To investigate the first suggestion, a model for the instrumentation noise was developed and its effect on SEMG signal features was investigated. To investigate the second suggestion, a genetic algorithm was employed to find realistic tool parameters by matching SEMG records. The two investigations revealed that:

- Instrumentation effects were negligible when modeling signals collected with typical instrumentation setups.
- Better selection of embedded and user-defined tool parameters improved the performance of PANDA.

This work yielded 3 important contributions:

- A model of the instrumentation noise was proposed and implemented in Myosim.
- A better set of embedded model parameters were integrated into Myosim including conductivity values, σ_l/σ_t, σ_i, σ_e, and other amplitude scaling factors a/v, K_{scale}.
- A method to use a GA to find realistic Myosim tool parameters was proposed and implemented.

6.1 Recommendations

While the investigations detailed in this work yielded constructive results, they also uncovered questions for future consideration. Suggestions for further exploration include the following:

- Further improvements to the model used in Myosim: Myosim does not currently incorporate electrode geometry (specifically electrode surface area). Literature pertaining to SEMG signal measurement indicated that the SEMG records are influenced by the electrode surface area [17]. Thus, modelling the effects of electrode surface area and integrating it into Myosim may provide useful insight in SEMG studies. Another factor is to explore the observed difference in amplitudes between simulated SFAP and those observed experimentally. This work proposed a scaling factor (K_{scale}) to overcome the problem, but the cause of the difference is currently unclear. It is suspected that a combination of several factors such as presence of tendons, bones, fat tissues, etc., is the cause of the observed difference. Since Myosim does not incorporate these factors, a study to implement this and test the results can yield useful information.

- Exploring GA configuration parameters: In this work, the GA configuration parameters (i.e. crossover rate, mutation rate – also referred to as 'GA control parameters' in literature) were set heuristically. Literature suggests that a wide range of configuration parameters could improve performance and that a

dynamically changing set of configuration parameter may achieve the optimal performance [43]. These options could be explored to improve matching speed.

- Exploring alternative fitness functions to improve matching performance. This way, the obtained solution(s) would be precise, in the sense that a solution will always produce a single matching signal. The computational requirements for such a setup would be high, but can be offset by the fact that members of the population can be evaluated in parallel.

- Construction of a standard control set: The control signals used in this study may not generalise to all SEMG data. Exploration to produce a 'standard control data set' which represent all SEMG signals sufficiently well is suggested. This would require matching to various muscle groups and contraction levels initially, but once in place it could eliminate the need to match signals under test, making the process of noisy signal identification less time consuming.

REFERENCES

1. N. Nazmi et al., "A review of classification techniques of EMG signals during isotonic and isometric contractions," *Sensors*, vol. 16, issue 8, pp. 1304, 2016.

2. I. Conradsen et al., "Evaluation of novel algorithm embedded in a wearable sEMG device for seizure detection," *2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, San Diego, CA, 2012, pp. 2048-2051.

3. A. Herrera et al., "Design of an electrical prosthetic gripper using EMG and linear motion approach," *Proceedings from the 17th Florida Conference on the Recent Advances in Robotics (FCRAR),* Florida, USA, 2004.

4. R. H. Chowdhury et al., "Surface Electromyography Signal Processing and Classification Techniques," *Sensors*, vol. 13, issue 9, pp. 12431-12466, 2013.

5. A. D. Chan & D. MacIsaac, "CleanEMG: Assessing the quality of EMG signals," *34th Conference of the Canadian Medical & Biological Engineering Society and Festival of International Conferences on Caregiving, Disability, Aging and Technology,* Toronto, Canada, 2011, vol. 69826, pp. 1-4.

6. G. Phillips, "Pairwise Attribute Noise Detection Algorithm for Detecting Noise in Surface Electromyography Recordings," MScE Thesis, Dept. of Electrical and Computer Engineering, University of New Brunswick, Fredericton, Canada, 2016.

7. D. MacIsaac, "Simulating Myoelectric Signals with a Finite Length Model of Muscle," *Proceedings of the 29th Conference of the Canadian Medical and Biological Engineering Society*, 2006.

8. G. Phillips et al., "Pairwise Attribute Noise Detection Applied to Surface EMG," XX ISEK Congr, Rome, Italy, 2014, pp. 97.

9. R. J. Monti, R. R. Roland, & V. R. Edgerton, "Role of motor unit structure in defining function," *Muscle & nerve*, vol. 24, issue 7, pp. 848-866, 2001.

10. M. B. Bromberg, "Motor unit estimation: Reproducibility of the spike-triggered averaging technique in normal and ALS subjects," *Muscle & nerve*, vol. 16, issue 5, pp. 466-471, 1993.

11. F. Buchthal & H. Schmalbruch, "Motor unit of mammalian muscle," *Physiological reviews*, vol. 60, issue 1, pp.90-142, 1980.

12. R. Plonsey, "The active fiber in a volume conductor," *IEEE Transactions on Biomedical Engineering*, vol. BME-21, issue 5, pp. 371-381, 1974.

13. J. A. Gonzalez-Cueto & P. A. Parker, "Deconvolution estimation of motor unit conduction velocity distribution," *IEEE Transactions on Biomedical Engineering*, vol. 49, issue 9, pp. 955-962, 2002.

14. D. MacIsaac, “A Robust Index of Localized Muscle Fatigue,” PhD Dissertation, University of New Brunswick, Fredericton, Canada, 2004.

15. B. K. Van Veen et al., "Potential distribution and single-fibre action potentials in a radially bounded muscle model," *Medical and Biological Engineering and Computing*, vol. 30, issue 3, pp. 303-310, 1992.

16. H. P. Clamann, “Statistical analysis of motor unit firing patterns in a human skeletal muscle,” *Biophysical journal*, vol. 9, issue 10, pp.1233-1251, 1969.

17. H. J. Hermens et al., "Development of recommendations for SEMG sensors and sensor placement procedures," *Journal of electromyography and Kinesiology* vol. 10, issue 5, pp. 361-374, 2000.

18. D. T. Godin, P. A. Parker & R. N. Scott, "Noise characteristics of stainless-steel surface electrodes," *Medical and Biological Engineering and Computing*, vol. 29, issue 6, pp. 585-590, 1991.

19. M. S. Keshner, "1/f noise," *Proceedings of the IEEE*, vol. 70, issue 3, pp. 212-218, 1982.

20. R. Mancini, *Op amps for everyone: design reference*, Newnes, 2003.

21. R. Merletti & P. A. Parker, *Electromyography: physiology, engineering, and non-invasive applications*, Vol. 11, John Wiley & Sons, 2004.

22. M .B. Reaz, M. S. Hussain & F. Mohd-Yasin, "Techniques of EMG signal analysis: detection, processing, classification and applications," *Biological procedures online*, vol. 8, issue 1, pp. 11-35, 2006.

23. Analog Devices, "Op amp noise relationships: 1/f noise, rms noise, and equivalent noise bandwidth." MT-048 Tutorial, 2009.

24. M. H. Limaye, & M.V. Deshmukh, “ECG Noise Sources and Various Noise Removal Techniques: A Survey,” *International Journal of Application or Innovation in Engineering & Management (IJA1EM)*, vol. 5, issue 2, pp. 86-92, 2016.

25. D. T. Mewett, H. Nazeran & K. J. Reynolds, "Removing power line noise from recorded EMG," *2001 Conference Proceedings of the 23rd Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pp. 2190-2193 vol.3, 2001.

26. C. J. De Luca et al., "Filtering the surface EMG signal: Movement artifact and baseline noise contamination," *Journal of biomechanics*, vol. 43, issue 8, pp. 1573-1579, 2010.

27. D. F. Stegeman, & H. Hermens, “Standards for surface electromyography: The European project Surface EMG for non-invasive assessment of muscles (SENIAM),” *Enschede: Roessingh Research and Development*, pp.108-12, 2007.

28. M. T. Kyu, Z. M. Aung & Z. M. Naing, "Design and Implementation of Active Filter for Data Acquisition System," *2009 International Conference on Information Management and Engineering*, Kuala Lumpur, pp. 406-410, 2009.

29. W. Kester, "Taking the Mystery out of the Infamous Formula," SNR= 6.02 N+ 1.76 dB," and Why You Should Care,” MT-001 Tutorial, 2008.

30. J. D. Van Hulse, T. M. Khoshgoftaar & H. Huang, "The pairwise attribute noise detection algorithm," *Knowledge and Information Systems*, vol. 11, issue 2, pp. 171-190, 2007.

31. Y. Shi, “Within-trial and Between-trial variability in surface EMG features,” Dept. of Electrical and Computer Engineering, University of New Brunswick, Fredericton, Canada, 2017.

32. D. F. Lovely, "Low noise electrode amplifier for use in evoked potential studies," *Proceedings of the Canadian Medical and Biological Engineering Society Conference* (CMBES), Ottawa, Canada, 1993.

33. Adrian D. C. Chan, http://www.sce.carleton.ca/faculty/chan/index.php, last accessed: 2017-05-17.

34. D. Whitley, "A genetic algorithm tutorial," *Statistics and computing*, vol. 4, issue 2, pp. 65-85, 1994.

35. D. Beasley, D. R. Bull & R. R. Martin, "An overview of genetic algorithms: Part 1, fundamentals," *University computing*, vol. 15, issue 2, pp. 56-69, 1993.

36. F. Buchthal, C. Guld & P. Rosenfalck, "Volume Conduction of the Spike of the Motor Unit Potential Investigated with a New Type of Multielectrode," *Acta Physiologica*, vol. 38, issues 3-4, pp. 331-354, 1957.

37. T. Murata & H. Ishibuchi, "MOGA: multi-objective genetic algorithms," *Proceedings of 1995 IEEE International Conference on Evolutionary Computation*, Perth, WA, Australia, 1995, pp. 289.

38. D. Whitley, S. Rana, & R. B. Heckendorn, “The island model genetic algorithm: On separability, population size and convergence,” CIT. Journal of computing and information technology, vol. 7, issue 1, pp. 33-47, 1999.

39. N. M. Razali & J. Geraghty, “Genetic algorithm performance with different selection strategies in solving TSP,” *Proceedings of the world congress on engineering*, vol. 2, pp. 1134-1139, 2011.

40. K. Seki, & M. Narusawa, “Firing rate modulation of human motor units in different muscles during isometric contraction with various forces,” *Brain research*, vol. 719, issue 1, pp. 1-7, 1996.

41. T. V. Mathew, "Genetic algorithm," Report submitted at IIT Bombay, 2012.

42. K. A. De Jong, & W. M. Spears, “A formal analysis of the role of multi-point crossover in genetic algorithms,” *Annals of mathematics and Artificial intelligence*, vol. 5, issue 1, pp. 1-26, 1992.

43. J. J. Grefenstette, "Optimization of Control Parameters for Genetic Algorithms," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 16, issue 1, pp. 122-128, 1986.

44. A. Phinyomark et al, "The usefulness of mean and median frequencies in electromyography analysis," *Computational intelligence in electromyography analysis-A perspective on current applications and future challenges*, InTech, 2012.

45. L. Mesin, R. Merletti, & A. Rainoldi, “Surface EMG: the issue of electrode location,” *Journal of Electromyography and Kinesiology*, vol. 19, issue 5, pp. 719-726, 2009.

46. S. M. Fleisher, M. Studer, & G. S. Moschytz, “Mathematical model of the single-fibre action potential,” *Medical and Biological Engineering and Computing*, vol. 22, issue 5, pp. 433-439, 1984.

47. A. Miller-Larsson, “An analysis of extracellular single muscle fibre action potential field—Modelling results,” *Biological cybernetics*, vol. 51, issue 4, pp. 271-284, 1985.

48. J. Rodriguez-Falces, J. Navallas & A. Malanda, ”EMG modeling,” *Computational Intelligence in Electromyography Analysis-A Perspective on Current Applications and Future Challenges*, InTech, 2012 .

APPENDIX A: Description of features

Feature	Description	Formulae
Mean Absolute Value (MAV)	The mean of a typical EMG signal is approximately zero, therefore the mean absolute value is found to provide a reasonable value for comparisons. MAV is calculated by averaging the absolute values of each sample.	$$MAV = \frac{1}{N}\sum_{n=1}^{N}\lvert x(n\Delta t)\rvert$$
Mean Frequency (MeanFreq)	This metric estimates the centre of mass of the SEMG PSD ϕ.	$$MeanFreq = \sum_{n=0}^{N} f_n\phi_n / \sum_{n=0}^{N}\phi_n$$
Zero Crossings (ZC)	The zero crossings attribute indicates the number of sign changes that occur within the signal. The ZC count is augmented each time the signal crosses the zero line (x-axis). A threshold (Th) is set to omit any crossings that are negligible.	$$ZC = \sum_{n=1}^{N-1} f\Big(\big(x(n\Delta t)\cdot x((n+1)\Delta t)\big)\Big) + \sum_{n=1}^{N-1} g\left(\frac{x((n+1\Delta t))}{x(n\Delta t)}\right)$$ $$f(x) = \begin{cases}1, x < Th\\0, x \geq Th\end{cases}$$ $$g(x) = \begin{cases}1, x = 0\\0, x \neq 0\end{cases}$$
Slope Sign Change (SSC)	The slope sign change attribute denotes the number of slope sign changes incurred by the signal. Like the ZC	$$SSC = \sum_{n=2}^{N-1} f\Big(\big(x(n\Delta t) - x((n-1)\Delta t)\big)\cdot\big(x((n+1)\Delta t) - x(n\Delta t)\big)\Big)$$ $$f(x) = \begin{cases}1, x \leq Th\\0, x > Th\end{cases}$$

	attribute, a threshold (Th) is assigned to omit negligible slope sign changes.	
Wavelength (WL)	The wavelength of the signal is calculated by computing the shortest distance (the hypotenuse) between each adjacent pair of samples and summing the findings.	$WL = \sum_{n=2}^{N} \sqrt{(x(n\Delta t) - x((n-1)\Delta t))^2 + \Delta t^2}$
Entropy (EN)	Entropy measures the randomness of the amplitude of the input signal. Here, p_m represents the probability of occurrence of the m^{th} amplitude value (A_m). It is estimated by counting the number of times it occurs in the given signal (ie from a histogram of $x(n\Delta t)$).	$EN = -\sum_{m=1}^{M} (p_m \cdot \log_2(p_m))$ $p_m \approx \frac{count(A_m)}{N}$
Normality/Histogram (HIST)	The HIST feature is calculated by summing the values of the first and last bin of the 10-bin histogram H if the signal. This feature was found to be sensitive to saturation.	$HIST = H(1) + H(10)$

60Hz Interference	The $60\ Hz$ interference feature is typically computed by summing the spectral components in a narrow range around $60\ Hz$. A range was used to accommodate potential for $60\ Hz$ jitter described by Abser et al. In their work they suggest a range between $58 - 62\ Hz$ which was applied here. For this work, the difference between the expected values (using median filter method proposed by Abser et al.) and the measured value normalised by the total power of the signal is used as the feature.	$$Int(60) = \frac{\lvert \sum_{n=58}^{62} \Delta\phi_n \rvert}{\sum_{n=0}^{N} \phi_n}$$ $$\Delta\phi_n = \phi_{n(measured)} - \phi_{n(expected)}$$
180Hz Interference	The $180\ Hz$ interference feature is similar to the $60\ Hz$ interference feature, except the frequency range is $178 - 182\ Hz$.	$$Int(180) = \frac{\lvert \sum_{n=178}^{182} \Delta\phi_n \rvert}{\sum_{n=0}^{N} \phi_n}$$

Signal to Motion Artifact Ratio (SM)	The SM ratio is used to identify low-frequency noise in the signal. It is typically found by summing the power densities across the entire signal and dividing by the power densities at frequencies less than $20Hz$ which extend beyond a line drawn from the origin (zero frequency) to the highest mean power density. For this work, the inverse of this ratio is taken as the feature, i.e., the power above the line is divided by the total power.	$SM = \frac{\sum_{n=0}^{20} \phi_{peak_n}}{\sum_{n=0}^{N} \phi_n}$

Table 6.1: Feature description (Th=0.005)

Curriculum Vitae

Candidate's Full Name:	Shriram Tallam Puranam Raghu
Universities Attended	B. Tech, JNTU-H, India, 2014
Publications: None	
Conference Presentations: None	