

# **Automatic Biosignal Quality Analysis for Surface Electromyography**

By

Graham D. Fraser

A thesis submitted to the Faculty of Graduate Studies and Research in partial fulfillment of the requirements for the degree of

**Masters of Applied Science**

**in Biomedical Engineering**

Ottawa-Carleton Institute for Biomedical Engineering

Department of Systems and Computer Engineering

Carleton University

Ottawa, Ontario, Canada

July 2012

Copyright © Graham D. Fraser, 2012



Library and Archives  
Canada

Published Heritage  
Branch

395 Wellington Street  
Ottawa ON K1A 0N4  
Canada

Bibliothèque et  
Archives Canada

Direction du  
Patrimoine de l'édition

395, rue Wellington  
Ottawa ON K1A 0N4  
Canada

*Your file Votre référence*

*ISBN: 978-0-494-93479-1*

*Our file Notre référence*

*ISBN: 978-0-494-93479-1*

#### NOTICE:

The author has granted a non-exclusive license allowing Library and Archives Canada to reproduce, publish, archive, preserve, conserve, communicate to the public by telecommunication or on the Internet, loan, distribute and sell theses worldwide, for commercial or non-commercial purposes, in microform, paper, electronic and/or any other formats.

The author retains copyright ownership and moral rights in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

#### AVIS:

L'auteur a accordé une licence non exclusive permettant à la Bibliothèque et Archives Canada de reproduire, publier, archiver, sauvegarder, conserver, transmettre au public par télécommunication ou par l'Internet, prêter, distribuer et vendre des thèses partout dans le monde, à des fins commerciales ou autres, sur support microforme, papier, électronique et/ou autres formats.

L'auteur conserve la propriété du droit d'auteur et des droits moraux qui protègent cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

---

In compliance with the Canadian Privacy Act some supporting forms may have been removed from this thesis.

While these forms may be included in the document page count, their removal does not represent any loss of content from the thesis.

Conformément à la loi canadienne sur la protection de la vie privée, quelques formulaires secondaires ont été enlevés de cette thèse.

Bien que ces formulaires aient inclus dans la pagination, il n'y aura aucun contenu manquant.

Canada

# Abstract

Surface electromyography (sEMG) is a useful tool for non-invasively measuring the electrical activity associated with contracting skeletal muscles. Applications of sEMG include ergonomics, controlling powered prostheses, and clinical neuromuscular assessment. SEMG signals can become contaminated with various forms of noise which can invalidate conclusions drawn from the data. This thesis presents methods to detect, identify, quantify, and mitigate various types of contamination in sEMG.

A least squares adaptive algorithm is evaluated for estimating power line interference. Analog-to-digital converter clipping can be detected by searching for consecutive extrema in the signal. Quantization noise can be estimated from the smallest observed difference between any two values. Amplifier saturation can be detected if the sEMG amplitude fails a test for normality. Electrocardiogram interference and motion artifact can be quantified using a combination of moving averages. A one-class support vector machine is also trained to differentiate clean from contaminated sEMG. These methods were evaluated on simulated and real sEMG artificially contaminated at controlled signal-to-noise ratios (SNR). Performance is expressed as a function of SNR. The objective is to contribute towards the development of an open-source user-friendly software tool to automatically assess sEMG quality. Future work should investigate other types of contaminants, the differentiation between similar contaminants, and the isolation of different contaminants in the same signal.

# **Acknowledgements**

I would like to thank my supervisors, Dr. Adrian Chan and Dr. James Green for their excellent guidance, my family and friends for their continued support, and my colleagues, Varsha Chaugai, Andy Huang, and Shabana Baig for their intellectual support as well as motivational support during those long hours in the lab.

I would also like to thank the Natural Sciences and Engineering Research Council of Canada for funding this research.

# Statement of Originality

This thesis presents the results of the work conducted by the author at Carleton University for the degree of M.A.Sc. in Biomedical Engineering. Many of these results have been presented and appear in conference publications. A summary of these publications and the results of the research is provided below.

**G. D. Fraser**, A. D. C. Chan, J. R. Green, D. MacIsaac. "Removal of Electrocardiogram Artifacts in Surface Electromyography using a Moving Average Method." 2012 IEEE Symposium on Medical Measurements and Applications, Budapest, Hungary, pp. 128-131, May 18-19, 2012.

The results of this conference paper appear in Chapter 6. The author collected all surface electromyography (sEMG) and electrocardiography (ECG) data used in the analysis, performed a thorough investigation of the contaminating effects of the ECG on the sEMG, and developed an innovative method to remove this contamination. The manuscript was written, revised with feedback from co-authors, and finalized for publication by the author. This paper was presented as an oral presentation by the author at the 2012 IEEE Symposium on Medical Measurements and Applications (MeMeA 2012) in Budapest, Hungary.

**G. D. Fraser**, A. D. C. Chan, J. R. Green, D. MacIsaac. "Detection of ADC Clipping, Quantization Noise, and Amplifier Saturation in Surface Electromyography." 2012 IEEE Symposium on Medical Measurements and Applications, Budapest, Hungary, pp. 162-166, May 18-19, 2012.

The results of this conference paper appear in Chapter 5. The author collected the sEMG used for this analysis, wrote the code to simulate the effects of analog-to-

digital converter clipping, quantization noise, and amplifier saturation, wrote the code to detect each of these three contaminants in sEMG, and ran computer simulations on simulated and real sEMG to evaluate the performance of the three methods of contaminant detection. The manuscript was written, revised with feedback from co-authors, and finalized for publication by the author. This paper was presented as a poster by the author at the 2012 IEEE Symposium on Medical Measurements and Applications (MeMeA 2012) in Budapest, Hungary.

**G. D. Fraser, A. D. C. Chan, J. R. Green, N. Abser, D. MacIsaac.** "CleanEMG - Power line interference estimation in sEMG using an adaptive least squares algorithm." 33rd Annual International Conference of the IEEE EMBS, Boston, Massachusetts USA, August 30 - September 3, 2011, pp. 7941- 7944.

The results of this conference paper appear in Chapter 4. The author wrote and tested code to quantify and mitigate power line interference in sEMG. The author then evaluated this code on simulated and real sEMG (collected by the author). The manuscript was written, revised with feedback from co-authors, and finalized for publication by the author. This paper was presented as a poster by the author at 33rd Annual International Conference of the IEEE EMBS (EMBC 2012) in Boston, Massachusetts, USA.

In Chapter 7, all sEMG data was collected by the author and all code to interface and train the LIBSVM tool was written by the author. The code to run the simulations and the evaluation of the classifier performance was done by the author. These results will be disseminated in a journal paper and submitted for publication in IEEE Transactions on Instrumentation and Measurement.

# Table of Contents

<b>ABSTRACT .....</b>	<b>III</b>
<b>ACKNOWLEDGEMENTS.....</b>	<b>IV</b>
<b>STATEMENT OF ORIGINALITY .....</b>	<b>V</b>
<b>TABLE OF CONTENTS .....</b>	<b>VII</b>
<b>LIST OF TABLES.....</b>	<b>X</b>
<b>LIST OF FIGURES .....</b>	<b>XI</b>
<b>LIST OF ABBREVIATIONS.....</b>	<b>XV</b>
<b>1 INTRODUCTION.....</b>	<b>1</b>
1.1 INTRODUCTION .....	1
1.2 MOTIVATION .....	1
1.3 STATEMENT OF THE PROBLEM.....	3
1.4 CONTRIBUTIONS.....	3
1.5 ORGANIZATION OF THESIS.....	6
<b>2 LITERATURE REVIEW.....</b>	<b>8</b>
2.1 INTRODUCTION .....	8
2.2 PRINCIPLES OF ELECTROMYOGRAPHY .....	8
2.3 APPLICATIONS OF ELECTROMYOGRAPHY .....	9
2.4 ELECTROMYOGRAPHY DATA ACQUISITION .....	11
2.4.1 <i>Needle vs. Surface Electromyography</i> .....	11
2.4.2 <i>Surface Electrodes</i> .....	12
2.4.3 <i>Electrode configurations</i> .....	13
2.4.4 <i>Amplifier</i> .....	14
2.4.5 <i>Analog-to-Digital Converter</i> .....	15
2.4.6 <i>Skin Preparation</i> .....	15
2.4.7 <i>Positioning of Electrodes</i> .....	16
2.5 SOURCES OF CONTAMINATION IN SURFACE ELECTROMYOGRAPHY .....	17
2.6 BIOSIGNAL QUALITY ANALYSIS.....	20
2.6.1 <i>Categories of Biosignal Quality Analysis</i> .....	20
2.6.2 <i>Biosignal Quality Analysis in Surface Electromyography</i> .....	22
2.7 SUMMARY.....	25
<b>3 SURFACE ELECTROMYOGRAPHY DATA .....</b>	<b>26</b>
3.1 INTRODUCTION .....	26
3.2 SURFACE ELECTROMYOGRAPHY SIMULATION .....	26
3.3 DATA COLLECTION EQUIPMENT.....	27
3.4 SKIN PREPARATION .....	28
3.5 MUSCLE CONTRACTIONS AND LANDMARKS .....	29
3.6 DATA COLLECTION PROCEDURE .....	34
3.7 REMOVAL OF CONTAMINATED DATA .....	38
3.8 SUMMARY.....	39
<b>4 POWER LINE INTERFERENCE ESTIMATION AND MITIGATION .....</b>	<b>40</b>

4.1	INTRODUCTION .....	40
4.2	LEAST SQUARES ADAPTIVE POWER LINE ESTIMATION ALGORITHM .....	41
4.3	SEMG SIMULATION AND INSERTION OF NOISE .....	43
4.4	SIMULATION RESULTS .....	43
4.4.1	<i>sEMG Spectrum Shape</i> .....	44
4.4.2	<i>Frequency and SNR</i> .....	45
4.4.3	<i>Signal Length and SNR</i> .....	48
4.4.4	<i>Comparison with Notch Filtering</i> .....	50
4.5	RESULTS WITH REAL SEMG .....	54
4.6	DISCUSSION .....	55
4.7	CONCLUSION.....	57
<b>5</b>	<b>DETECTION OF ADC CLIPPING, QUANTIZATION NOISE, AND AMPLIFIER SATURATION .....</b>	<b>58</b>
5.1	INTRODUCTION .....	58
5.2	DETECTION OF ADC CLIPPING .....	61
5.2.1	<i>Method</i> .....	61
5.2.2	<i>Simulated sEMG Results</i> .....	62
5.2.3	<i>Validation of ADC Clipping Detection with Real sEMG</i> .....	65
5.3	DETECTION OF QUANTIZATION NOISE.....	68
5.4	DETECTION OF AMPLIFIER SATURATION.....	71
5.4.1	<i>Method</i> .....	71
5.4.2	<i>Simulated sEMG Results</i> .....	72
5.4.3	<i>Validation of Amplifier Saturation Detection with Real sEMG</i> .....	73
5.5	DISCUSSION .....	75
5.6	CONCLUSION.....	77
<b>6</b>	<b>DETECTION AND MITIGATION OF ELECTROCARDIOGRAM ARTIFACT AND QUANTIFICATION OF MOTION ARTIFACT .....</b>	<b>78</b>
6.1	INTRODUCTION .....	78
6.2	MOVING AVERAGE METHOD .....	80
6.3	TEMPLATE SUBTRACTION METHOD .....	83
6.4	DATA ACQUISITION .....	84
6.5	COMPARISON OF METHODS .....	85
6.6	DISCUSSION .....	87
6.7	QUANTIFICATION OF MOTION ARTIFACT .....	88
6.7.1	<i>Introduction</i> .....	88
6.7.2	<i>Data</i> .....	89
6.7.3	<i>Methodology</i> .....	89
6.7.4	<i>Results</i> .....	91
6.8	CONCLUSION.....	93
<b>7</b>	<b>CLASSIFICATION OF SEMG USING SUPPORT VECTOR MACHINES .....</b>	<b>95</b>
7.1	INTRODUCTION .....	95
7.2	ONE-CLASS SVM .....	95
7.3	FEATURE EXTRACTION .....	98
7.3.1	<i>Autoregressive Model Coefficients</i> .....	98
7.3.2	<i>sEMG Histogram</i> .....	99
7.3.3	<i>Mean Absolute Value</i> .....	99
7.3.4	<i>Modified Median Frequency</i> .....	99
7.3.5	<i>Modified Mean Frequency</i> .....	100
7.3.6	<i>Power Spectral Density</i> .....	100
7.3.7	<i>Slope Sign Changes</i> .....	100

7.3.8	<i>Willison Amplitude</i>	101
7.3.9	<i>Waveform Length</i>	101
7.3.10	<i>Zero Crossings</i>	101
7.4	FEATURE OPTIMIZATION	102
7.5	SVM PERFORMANCE EVALUATION	104
7.6	RESULTS FOR CLASSIFICATION OF SEMG WITH A SINGLE CONTAMINANT	106
7.6.1	<i>Classification of sEMG with Power Line Interference, ECG Interference, or Motion Artifact</i>	106
7.6.2	<i>Classification of sEMG with Quantization Noise</i>	110
7.6.3	<i>Classification of sEMG with ADC Clipping</i>	113
7.6.4	<i>Classification of sEMG with Amplifier Saturation</i>	115
7.6.5	<i>Using the SVM Classification Approach to Quantify Signal Contamination</i>	118
7.7	CLASSIFICATION OF SEMG WITH TWO SIMULTANEOUS CONTAMINANTS	124
7.8	CONCLUSION	128
<b>8</b>	<b>THISIS SUMMARY AND FUTURE RECOMMENDATIONS</b>	<b>129</b>
8.1	SUMMARY	129
8.2	RECOMMENDATIONS FOR FUTURE WORK	130
	<b>APPENDIX A: LEAST SQUARES ADAPTIVE POWER LINE ESTIMATION ALGORITHM DERIVATION</b>	<b>132</b>
	<b>APPENDIX B: MOVING AVERAGE FREQUENCY RESPONSE DERIVATION</b>	<b>136</b>
	<b>APPENDIX C: INFORMED CONSENT FORM</b>	<b>137</b>
	<b>REFERENCES</b>	<b>145</b>

# List of Tables

TABLE 2-1: SEMG CONTAMINANTS CLASSIFIED AS INTERFERENCE, MEASUREMENT ARTIFACT, OR INSTRUMENTATION PROBLEM .....	17
TABLE 3-1: LANDMARKS AND CONTRACTION DESCRIPTION FOR BICEPS BRACHII .....	30
TABLE 3-2: LANDMARKS AND CONTRACTION DESCRIPTION FOR RECTUS FEMORIS.....	31
TABLE 3-3: LANDMARKS AND CONTRACTION DESCRIPTION FOR TIBIALIS ANTERIOR .....	32
TABLE 3-4: AMPLIFIER SETTINGS .....	36
TABLE 3-5: DATA ACQUISITION SETTINGS .....	36
TABLE 4-1: SEMG SPECTRUM SHAPE INDEX VALUES AND CORRESPONDING PARAMETER VALUES. THESE 4 PARAMETER PAIRS WERE CHOSEN AS THEY WERE THE SAME AS USED IN [5].....	44
TABLE 5-1: SHAPING PARAMETER PAIRS FOR SEMG SIMULATION.....	61
TABLE 7-1: CONTAMINANT LEVELS USED FOR FEATURE OPTIMIZATION .....	103
TABLE 7-2: OPTIMAL FEATURE SET FOR SVM CLASSIFICATION OF SEMG .....	104
TABLE 7-3: TRANSITION POINTS IN SVM CLASSIFICATION ACCURACY FOR EACH CONTAMINANT FOR SIMULATED AND REAL SEMG .....	118
TABLE 7-4: AMOUNTS OF CONTAMINATION FOR EACH NUMERIC CONTAMINANT LEVEL.....	122
TABLE 7-5: NOISE COMBINATIONS USED IN FIGURE 7-19 WITH CORRESPONDING NOISE LEVELS BELOW AND ABOVE THE TRANSITION POINT FOR EACH CONTAMINANT.....	126

# List of Figures

FIGURE 1-1: A TYPICAL SEMG SIGNAL .....	2
FIGURE 2-1: INNERVATION OF MUSCLE FIBRES BY A MOTOR NEURON TO FORM A MOTOR UNIT.....	9
FIGURE 2-2: TYPICAL EMG DATA ACQUISITION SETUP .....	11
FIGURE 2-3: MVAP-II AG-AGCL ELECTRODES. ....	13
FIGURE 3-1: EFFECT OF PARAMETERS ON SIMULATED SEMG POWER SPECTRAL DENSITY AS BASED ON [6].....	27
FIGURE 3-2: CONTRACTION STARTING POSITION AND ELECTRODE PLACEMENT FOR THE BICEPS BRACHII AS BASED ON [11].....	30
FIGURE 3-3: CONTRACTION STARTING POSITION AND ELECTRODE PLACEMENT FOR THE RECTUS FEMORIS AS BASED ON [11] .....	31
FIGURE 3-4: CONTRACTION STARTING POSITION AND ELECTRODE PLACEMENT FOR THE TIBIALIS ANTERIOR AS BASED ON [11].....	32
FIGURE 3-5: MVC FEEDBACK BAR GRAPH .....	34
FIGURE 3-6: HIERARCHY OF TERMINOLOGY FOR SEMG MVC DATASET (A) AND CONTRACTION DATASET (B).....	35
FIGURE 4-1: BOX PLOT OF RMS ERROR FOR 4 SEMG SPECTRUM SHAPES BETWEEN ORIGINAL CLEAN SIGNAL AND NOISY SIGNAL WITH NOISE ESTIMATE SUBTRACTED. ....	45
FIGURE 4-2: AVERAGE RMS ERROR FOR 5 DIFFERENT POWER LINE FREQUENCIES AT VARIOUS SNRS. ERROR BARS ARE AT PLUS OR MINUS ONE STANDARD DEVIATION. FREQUENCIES ARE OFFSET ON THE SNR-AXIS SUCH THAT ERROR BARS DO NOT OVERLAP.....	46
FIGURE 4-3: MEAN FREQUENCY ERROR FOR 5 DIFFERENT POWER LINE FREQUENCIES AT VARIOUS SNRS. ERROR BARS ARE AT PLUS OR MINUS ONE STANDARD DEVIATION. FREQUENCIES ARE OFFSET ON THE SNR-AXIS SUCH THAT ERROR BARS DO NOT OVERLAP. ....	47
FIGURE 4-4: DISTRIBUTION OF POWER LINE FREQUENCY ESTIMATES FOR 60 HZ NOISE AT (A) SNR=-5 dB AND (B) SNR=40 dB. NOTE THE X-AXIS SCALES ARE DIFFERENT. ....	48
FIGURE 4-5: AVERAGE RMS ERROR FOR 7 DIFFERENT SIGNAL LENGTHS AT VARIOUS SNRS. ERROR BARS ARE AT PLUS OR MINUS ONE STANDARD DEVIATION. SIGNAL LENGTHS ARE OFFSET ON THE SNR-AXIS SUCH THAT ERROR BARS DO NOT OVERLAP. SIGNAL SAMPLED AT 1000 HZ.....	49
FIGURE 4-6: MEAN ESTIMATED SNR FOR EACH ACTUAL SNR FOR THE NOISY SEMG SIGNAL AT VARIOUS SIGNAL LENGTHS. IDEAL CASE IS SHOWN FOR REFERENCE. ERROR BARS ARE AT PLUS OR MINUS ONE STANDARD DEVIATION. SIGNAL SAMPLED AT 1000 HZ.....	50
FIGURE 4-7: RMS ERROR FOR NOTCH FILTERED SEMG AT 60 HZ AT DIFFERENT SNRS AND Q-FACTORS. ....	51
FIGURE 4-8: RMS ERROR FOR NOTCH FILTERED SEMG AT 60 HZ FOR DIFFERENT SNR AND POWER LINE FREQUENCIES.....	52
FIGURE 4-9: RMS ERROR FOR NOTCH FILTERED SEMG (Q-FACTOR = 30) AT 60 HZ FOR DIFFERENT SNR AND SIGNAL LENGTHS.....	53
FIGURE 4-10: COMPARISON OF NOTCH FILTERING (Q-FACTOR = 30) AND ADAPTIVE FILTERING FOR SHORT SIGNALS (64 SAMPLES) AND LONG SIGNALS (4096 SAMPLES AT 1000 HZ) AT VARIOUS SNR .....	53
FIGURE 4-11: AVERAGE RMS ERROR FOR 5 DIFFERENT POWER LINE FREQUENCIES AT VARIOUS SNRS FOR ENTIRE REAL SEMG DATASET. ERROR BARS ARE AT PLUS OR MINUS ONE STANDARD DEVIATION. FREQUENCIES ARE OFFSET ON THE SNR-AXIS SUCH THAT ERROR BARS DO NOT OVERLAP.....	54
FIGURE 5-1: TYPICAL SEMG COLLECTION SETUP.....	58
FIGURE 5-2: SEMG CLIPPED TO $\pm 1$ V, SAMPLED AT 1 KHZ.....	59
FIGURE 5-3: SEMG SIGNAL QUANTIZED TO 16-BITS AND 6 BITS SHOWING THE DISTORTION EFFECT OF QUANTIZATION NOISE .....	60
FIGURE 5-4: AMPLIFIED SEMG SIGNAL WITH AND WITHOUT AMPLIFIER SATURATION WHERE THE AMPLIFIER MAXIMUM IS AT 10 V .....	61

FIGURE 5-5: CLIPPING DETECTION SENSITIVITY AS A FUNCTION OF THE NUMBER OF CLIPPED SAMPLES IN THE SIGNAL (PLOT FOR ADC MAX OF 2.5 NOT SHOWN AS SENSITIVITY WAS ALWAYS 1)...	63
FIGURE 5-6: HISTOGRAM FOR NUMBER OF SAMPLES CLIPPED OVER EACH CLIPPING LEVEL. NOTE THAT THE DISTRIBUTION FOR ADC MAX OF 2.5 AND PART OF THAT FOR 2.75 IS CUT OFF. ....	64
FIGURE 5-7: CLIPPING DETECTION SENSITIVITY AS A FUNCTION OF THE NUMBER OF CLIPPED SAMPLES IN THE SIGNAL FOR REAL SEMG AT A SAMPLING RATE OF (A) 3000.3 Hz, (B) 1500.15 Hz, AND (C) 1000.1 Hz. NUMBER OF SAMPLES ARE NORMALIZED BY DIVIDING BY THE SIGNAL LENGTH AND MULTIPLYING BY 10000.....	66
FIGURE 5-8: HISTOGRAMS FOR NUMBER OF SAMPLES CLIPPED OVER EACH CLIPPING LEVEL FOR REAL SEMG AT A SAMPLING RATE OF (A) 3000.3 Hz, (B) 1500.15 Hz, AND (C) 1000.1 Hz. NUMBER OF SAMPLES ARE NORMALIZED BY DIVIDING BY THE SIGNAL LENGTH AND MULTIPLYING BY 10000. ....	67
FIGURE 5-9: CLIPPING DETECTION SENSITIVITY AS A FUNCTION OF THE NUMBER OF CLIPPED SAMPLES IN THE SIGNAL FOR REAL SEMG (3000.3 Hz) BY SEARCHING FOR (A) TWO CONSECUTIVE EXTREMA AND (B) THREE CONSECUTIVE EXTREMA. NUMBER OF SAMPLES ARE NORMALIZED BY DIVIDING BY THE SIGNAL LENGTH AND MULTIPLYING BY 10000.....	68
FIGURE 5-10: ESTIMATED AVERAGE SQNR AT DIFFERENT BIT RESOLUTIONS AND SIGNAL LENGTHS WITH ONE STANDARD DEVIATION ERROR BARS. AVERAGED ACTUAL SQNR ARE SHOWN AS DOTTED LINES. SIGNALS SAMPLED AT 1000 Hz. ....	70
FIGURE 5-11: TRANSFER FUNCTION OF A TYPICAL AMPLIFIER WITH LINEAR AND NON-LINEAR RANGES OF OPERATION AND A GAIN FACTOR OF 1. LINEAR RANGE CORRESPONDS TO AN INPUT BETWEEN -0.4 AND 0.4. ....	72
FIGURE 5-12: PEARSON CORRELATION COEFFICIENT CALCULATED WHEN DIFFERENT FRACTIONS OF THE SIGNAL SAMPLES WERE IN THE NONLINEAR ZONE OF THE AMPLIFIER RANGE FOR SIMULATED SEMG. ....	73
FIGURE 5-13: PEARSON CORRELATION COEFFICIENT CALCULATED WHEN DIFFERENT FRACTIONS OF THE SIGNAL SAMPLES WERE IN THE NONLINEAR ZONE OF THE AMPLIFIER RANGE FOR REAL SEMG. ERROR BARS ARE AT $\pm 1$ STANDARD DEVIATION.....	74
FIGURE 5-14: CORRELATION COEFFICIENT DISTRIBUTION FOR REAL SEMG. MEAN IS INDICATED BY THE RED VERTICAL LINE. ....	75
FIGURE 6-1: MAGNITUDE AND PHASE OF THE FREQUENCY RESPONSE FOR A MOVING AVERAGE FILTER WITH $L=15$ .....	81
FIGURE 6-2: (A) SEMG SIGNAL CONTAMINATED WITH ECG ARTIFACT (SNR = -8 dB). THE CONTAMINATED SEMG FILTERED USING A MOVING AVERAGE WINDOW SIZE OF (B) 50 MS AND (C) 10 MS. ....	82
FIGURE 6-3: BLOCK DIAGRAM OF THE MOVING AVERAGE METHOD. ....	83
FIGURE 6-4: (A) SEMG CONTAMINATED WITH ECG (SNR = 2 dB). (B) SEMG FILTERED WITH MOVING AVERAGE METHOD. (C) SEMG ESTIMATE ERROR WITH MOVING AVERAGE (RMS ERROR = 0.1883). (D) SEMG FILTERED WITH TEMPLATE SUBTRACTION METHOD. (E) SEMG ESTIMATE ERROR WITH TEMPLATE SUBTRACTION (RMS ERROR = 0.2229).....	86
FIGURE 6-5: AVERAGE RMS ERROR BETWEEN FILTERED AND ORIGINAL CLEAN SEMG FOR ECG ARTIFACT MITIGATION WITH THE MOVING AVERAGE AND TEMPLATE SUBTRACTION METHODS. .	87
FIGURE 6-6: ESTIMATED VS. ACTUAL SMR FOR SIMULATED SEMG SIGNALS USING THREE METHODS OF QUANTIFICATION. ERROR BARS ARE AT PLUS OR MINUS ONE STANDARD DEVIATION. ....	91
FIGURE 6-7: RMS ERROR SIGNALS FOR SIMULATED SEMG SMR USING THREE METHODS OF ..... QUANTIFICATION. ....	92
FIGURE 6-8: ESTIMATED VS. ACTUAL SMR FOR REAL SEMG SIGNALS USING THREE METHODS OF QUANTIFICATION. ERROR BARS ARE AT PLUS OR MINUS ONE STANDARD DEVIATION. ....	92

FIGURE 6-9: RMS ERROR SIGNALS FOR REAL SEMG SMR USING THREE METHODS OF ..... QUANTIFICATION. .....	93
FIGURE 7-1: CLASSIFICATION ACCURACY OF ONE-CLASS SVM TRAINED WITH SIMULATED SEMG AND TESTED ON SEMG CONTAMINATED WITH POWER LINE INTERFERENCE, ECG INTERFERENCE, OR MOTION ARTIFACT WITH VARYING SNR. SNR VALUES ARE OFFSET FOR VISIBILITY. ERROR BARS ARE AT $\pm 1$ STANDARD DEVIATION. ....	107
FIGURE 7-2: THE EFFECT OF POWER LINE INTERFERENCE AT 10 dB ON A SIMULATED SEMG SIGNAL IN THE (A) TIME AND (B) FREQUENCY DOMAIN. ....	108
FIGURE 7-3: THE EFFECT OF ECG INTERFERENCE AT 3 dB ON A SIMULATED SEMG SIGNAL IN (A) TIME AND (B) FREQUENCY DOMAIN. ....	108
FIGURE 7-4: THE EFFECT OF MOTION ARTIFACT AT 5 dB ON A SIMULATED SEMG SIGNAL IN (A) TIME AND (B) FREQUENCY DOMAIN. ....	109
FIGURE 7-5: CLASSIFICATION ACCURACY OF ONE-CLASS SVM TRAINED WITH REAL SEMG AND TESTED ON SEMG CONTAMINATED WITH POWER LINE INTERFERENCE, ECG INTERFERENCE, OR MOTION ARTIFACT WITH VARYING SNR. SNR VALUES ARE OFFSET FOR VISIBILITY. ERROR BARS ARE AT $\pm 1$ STANDARD DEVIATION. ....	110
FIGURE 7-6: CLASSIFICATION ACCURACY OF SIMULATED SEMG WITH VARYING DEGREES OF QUANTIZATION NOISE. ERROR BARS ARE AT $\pm 1$ STANDARD DEVIATION. ....	111
FIGURE 7-7: THE EFFECT OF QUANTIZATION NOISE AT AN ADC STEP SIZE OF $2^{-4.5}$ ON SIMULATED SEMG IN THE TIME DOMAIN. ....	112
FIGURE 7-8: CLASSIFICATION OF REAL SEMG CONTAMINATED WITH QUANTIZATION NOISE AT DIFFERENT ADC STEP SIZES. ERROR BARS ARE AT $\pm 1$ STANDARD DEVIATION. ....	113
FIGURE 7-9: CLASSIFICATION OF SIMULATED SEMG CONTAMINATED WITH ADC CLIPPING AT DIFFERENT MEAN NUMBER OF SAMPLES CLIPPED (NORMALIZED BY DIVIDING BY THE SIGNAL LENGTH OF 4096 AND MULTIPLYING BY 10,000). ERROR BARS ARE AT $\pm 1$ STANDARD DEVIATION. ....	114
FIGURE 7-10: THE EFFECT OF CLIPPING WITH A NORMALIZED NUMBER OF SAMPLES CLIPPED OF 49 ON SIMULATED SEMG IN THE (A) TIME DOMAIN AND (B) FREQUENCY DOMAIN. ....	114
FIGURE 7-11: CLASSIFICATION OF REAL SEMG CONTAMINATED WITH ADC CLIPPING AT DIFFERENT MEAN NUMBER OF SAMPLES CLIPPED (NORMALIZED BY DIVIDING BY THE SIGNAL LENGTH OF 4096 AND MULTIPLYING BY 10,000). ERROR BARS ARE AT $\pm 1$ STANDARD DEVIATION. ....	115
FIGURE 7-12: CLASSIFICATION OF SIMULATED SEMG CONTAMINATED WITH AMPLIFIER SATURATION AT DIFFERENT MEAN FRACTIONS OF THE SIGNAL IN THE AMPLIFIER NONLINEAR ZONE. ERROR BARS ARE AT $\pm 1$ STANDARD DEVIATION. ....	116
FIGURE 7-13: THE EFFECT OF AMPLIFIER SATURATION WHEN 16% OF THE SIGNAL FALLS IN THE AMPLIFIER NONLINEAR ZONE ON SIMULATED SEMG IN THE (A) TIME DOMAIN AND (B) FREQUENCY DOMAIN. ....	117
FIGURE 7-14: CLASSIFICATION OF REAL SEMG CONTAMINATED WITH AMPLIFIER SATURATION AT DIFFERENT FRACTIONS OF THE SIGNAL IN THE AMPLIFIER NONLINEAR ZONE. ERROR BARS ARE AT $\pm 1$ STANDARD DEVIATION. ....	117
FIGURE 7-15: DECISION VALUES FOR CLEAN TEST SET SEMG NORMALIZED BY TRAINING DECISION VALUES FOR EACH SVM TESTING SET PARTITION. THE DECISION BOUNDARY INDICATES THE CHANGE IN CLASSIFICATION VALUE. ....	120
FIGURE 7-16: MEAN DECISION VALUE FOR EACH CONTAMINANT OVER ALL TEST SET PARTITIONS AT DIFFERENT CONTAMINANT LEVELS AT (A) FULL SCALE AND (B) ZOOMED IN. THE CONTAMINANT LEVELS ARE DEFINED FOR EACH CONTAMINANT TYPE IN TABLE 7-3. ....	121
FIGURE 7-17: MEAN DECISION VALUES OVER ALL TESTING SET PARTITIONS FOR EACH CONTAMINANT BELOW AND ABOVE ITS TRANSITION POINT: 1) POWER LINE INTERFERENCE AT 5 dB AND 10 dB, 2) ECG INTERFERENCE AT -5 dB AND 0 dB, 3) MOTION ARTIFACT AT 5 dB AND 10 dB, 4) QUANTIZATION NOISE AT ADC STEP SIZE OF $2^{-3.5}$ AND $2^{-4}$ , 5) ADC CLIPPING AT AN ADC MAX OF 3 V AND 3.5 V, AND 6) AMPLIFIER SATURATION AT AN AMPLIFIER GAIN OF 4 AND 3.5. .....	123

FIGURE 7-18: CLASSIFICATION ACCURACY FOR REAL SEMG BOTH WITH AND WITHOUT ECG INTERFERENCE AT 0 DB AND QUANTIZATION NOISE AT VARIOUS ADC STEP SIZES. ERROR BARS ARE AT  $\pm 1$  STANDARD DEVIATION. .... 125

FIGURE 7-19: AVERAGE RECTIFIED DECISION VALUES FOR EACH CONTAMINANT COMBINATION FROM TABLE 7-5 BELOW AND ABOVE THE TRANSITION POINT. .... 127

# List of Abbreviations

ADC	Analog to Digital Converter
ANOVA	Analysis of Variance
ARM	Autoregressive Model
CMRR	Common Mode Rejection Ratio
ECG	Electrocardiogram
EMG	Electromyogram
FFT	Fast Fourier Transform
IIR	Infinite Impulse Response
MES	Myoelectric Signal
MU	Motor Unit
MUAPT	Motor Unit Action Potential Train
MVC	Maximum Voluntary Contraction
RBF	Radial Basis Function
RMS	Root Mean Square
sEMG	Surface Electromyogram
SFAPT	Single-Fibre Action Potential Train
SMR	Signal-to-motion Artifact Ratio
SNR	Signal-to-noise Ratio
SVM	Support Vector Machine

# 1 Introduction

## 1.1 Introduction

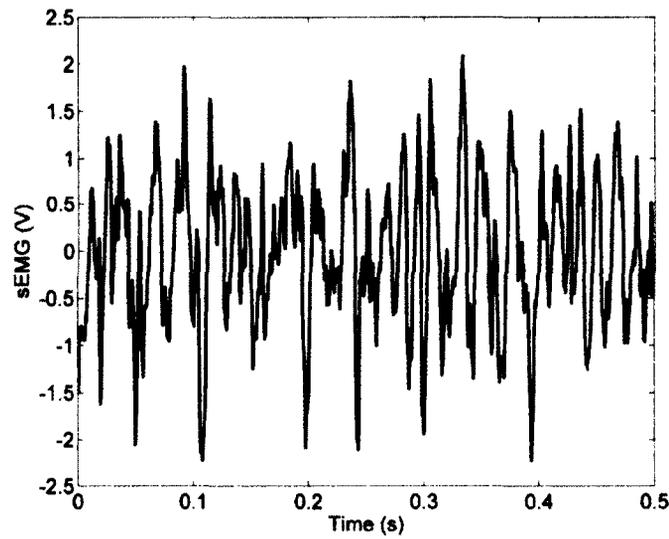
The focus of the work in this thesis is to develop and evaluate methods to detect, identify, quantify, or mitigate contamination in surface electromyography (sEMG). sEMG is used in a variety of applications but the presence of contamination can cause uncertainty or errors in the interpretation of the data. This chapter introduces surface electromyography (sEMG) along with its current applications and limitations. The scope of the work in this thesis is defined, as well as how this work fits into a larger ongoing research project, *CleanEMG* [1]. The aim of *CleanEMG* is to develop a user-friendly open-source software tool to perform automatic signal quality assessment in sEMG. Such a tool would remove the requirement for a trained sEMG technician and would enable reliable acquisition of clean sEMG without additional costs or complexities. This thesis contributes methods to detect six individual signal contaminants in sEMG, as well as a pattern classification approach to differentiate clean from contaminated sEMG. The need for this work in the field of sEMG acquisition and signal processing is demonstrated and the overall structure, contributions, and results of this thesis are outlined.

## 1.2 Motivation

Electromyography (EMG) is the measuring of the electrical activity associated with the contraction of skeletal muscles in the body. sEMG is the non-invasive measurement of EMG from the surface of the skin. Currently, sEMG is used in ergonomic studies, exercise physiology, movement and gait analysis, rehabilitation, biofeedback, powered control of prostheses, and clinical neuromuscular assessment. Acquired sEMG signals are susceptible to various forms of contamination which can invalidate conclusions drawn from the data. Many sources of contamination can be avoided through a proper data acquisition system

setup; however, this can be complex and requires considerable training and expertise from an sEMG technician. Furthermore, since there currently exists no universally accepted method to determine sEMG signal quality, sEMG cannot be reliably collected without the presence of such a technician which can incur considerable cost, add a significant time requirement, and severely limit the feasibility of sEMG collection. The result is limited clinical use of sEMG despite its promising potential [2].

Uncontaminated sEMG closely resembles random noise (Figure 1-1) and even with proper training it can be difficult to know if collected sEMG is clean or contaminated.



**Figure 1-1: A typical sEMG signal**

The development of a set of methods to detect, identify, quantify, or mitigate contamination in sEMG could reduce the cost associated with acquiring sEMG, remove uncertainty surrounding sEMG in a variety of its current applications, thus, improving their quality and encouraging clinical use, and potentially lead to new applications that were not previously feasible.

### **1.3 Statement of the Problem**

In this thesis, a set of methods are proposed and evaluated to detect, identify, quantify, or mitigate specific types of contaminants in sEMG. Each contaminant (e.g., power line interference, motion artifact, or ECG interference) will manifest itself in a different manner within the sEMG signal. Contaminant detection is specific to each individual form of contamination and so a separate method is required for each type. To assess the performance of the proposed methods, each method is evaluated by artificially contaminating a clean sEMG signal, allowing for a known ground truth. Both simulated and real sEMG are used for performance evaluation. Contamination levels are evaluated on the raw signals and not on extracted features which are typically used in a clinical setting. The acceptable level of contamination is expected to be application specific. Findings from this research can be adapted appropriately to suit each application (e.g., by setting appropriate thresholds of contamination).

A second objective of this thesis is to determine if clean and contaminated sEMG can be automatically differentiated regardless of contaminant type. This is done by investigating the classification of clean versus contaminated sEMG via a one-class support vector machine (SVM). In practice, it is unlikely that the type of contamination is known beforehand so this approach may be more appropriate for most applications. Contaminants are also unlikely to occur in isolation. It is also important to be able detect the presence of contamination in general even if the type or types cannot be identified.

### **1.4 Contributions**

This thesis provides six main contributions in the area of biosignal quality assessment for surface electromyography (sEMG). These contributions are as follows:

**1. Development and evaluation of a method using a least squares adaptive algorithm to identify, quantify, and mitigate power line interference in sEMG.**

Our results indicate that power line interference can be effectively detected using a least squares adaptive algorithm. The proposed method can accurately quantify and mitigate the interference for SNR below 20 dB and for typical length sEMG signals (i.e., 5 seconds or more). This performance is superior to that of notch filtering and offers much less signal distortion.

**2. A method to identify and quantify analog-to-digital converter clipping for typical length sEMG signals.**

Analog-to-digital converter (ADC) clipping can be reliably detected when the ADC range is unknown by searching for consecutive minimum or maximum values in the signal. Two consecutive extrema are sufficient for 1000-2000 Hz sampling rates. Above this, it is beneficial to search for 3 consecutive extrema to reduce the false positive rate.

**3. A method to identify, and quantify the quantization noise in an sEMG signal.**

Quantization noise can be quantified successfully for typical length sEMG signals (i.e., 5 seconds or more) with up to 16 bits of resolution by measuring the smallest difference between any two values in the signal.

**4. A method to identify the potential for amplifier saturation in sEMG.**

Amplifier saturation can be detected by measuring the correlation between the Gaussian probability density function and the sEMG time-domain amplitude distribution.

**5. A method using moving averages to quantify and mitigate electrocardiogram interference in sEMG. A similar method can also be used to quantify motion artifact.**

Electrocardiogram (ECG) interference can be estimated using a combination of moving averages with different window lengths. The accuracy of the noise estimate is superior to a popular existing method (i.e., template subtraction) at lower SNR and exhibits smaller variance. The proposed method is also better suited to automation. Performance degrades as SNR increases, however, at high SNR, the interference becomes insignificant. A moving average can also be used to quantify motion artifact in sEMG up to 10 dB. A modified implementation of Sinderby's method [3] can be used to extend the range of accuracy up to 20 dB.

**6. A method using a one-class support vector machine to differentiate clean from contaminated sEMG.**

A one-class SVM can be trained to differentiate clean and contaminated sEMG with one or more contaminants. The normalized decision function value from the SVM can be used as a confidence indicator in the classification decision. Accuracy will depend on the intensity of the contamination, as well as the quality of the clean dataset. Motion artifact, power line interference, and ECG interference can be detected with high accuracy, although detection is also possible in many cases by visually examining the signal or frequency spectrum. Detection of quantization noise, clipping, and amplifier saturation is possible even when no contamination is evident by visually examining the signal.

Methods developed here were directed at sEMG, however many are likely applicable to other biosignals such as electroencephalography (EEG) and ECG.

Some of these contributions were disseminated in the following publications:

- **G. D. Fraser**, A. D. C. Chan, J. R. Green, D. MacIsaac. "Removal of Electrocardiogram Artifacts in Surface Electromyography using a Moving Average Method." 2012 IEEE Symposium on Medical Measurements and Applications, Budapest, Hungary, pp. 128-131, May 18-19, 2012.
- **G. D. Fraser**, A. D. C. Chan, J. R. Green, D. MacIsaac. "Detection of ADC Clipping, Quantization Noise, and Amplifier Saturation in Surface Electromyography." 2012 IEEE Symposium on Medical Measurements and Applications, Budapest, Hungary, pp. 162-166, May 18-19, 2012.
- **G. D. Fraser**, A. D. C. Chan, J. R. Green, N. Abser, D. MacIsaac. "CleanEMG - Power line interference estimation in sEMG using an adaptive least squares algorithm." 33rd Annual International Conference of the IEEE EMBS, Boston, Massachusetts USA, August 30 - September 3, 2011, pp. 7941- 7944.

Related publications that the author has contributed to are:

- Nurul Abser, Dawn MacIsaac, Adrian D. C. Chan, **Graham Fraser**, James R. Green. "Clean EMG: Comparing Interpolation Strategies for Power Line Interference Quantification in Surface EMG Signals." Canadian Medical and Biomedical Engineering Conference, Halifax, Nova Scotia, Canada, June 18-22, 2012.
- Nurul Abser, Dawn MacIsaac, **Graham Fraser**, Adrian D. C. Chan, James R Green. "CleanEMG: Quantifying power line interference in surface EMG signals." 34th Conference of the Canadian Medical & Biological Engineering Society and Festival of International Conferences on Caregiving, Disability, Aging and Technology, Toronto, Canada, 69825, pp. 1-4, June 4- 9, 2011.

## ***1.5 Organization of Thesis***

This thesis consists of eight chapters.

Chapter 2 serves as a background in sEMG and a review of relevant literature in sEMG biosignal quality assessment.

Chapter 3 presents a method to simulate sEMG signals as well as a protocol for collection of clean sEMG for evaluation purposes.

Chapter 4 examines power line interference and evaluates an adaptive algorithm in terms of its ability to detect, identify, quantify, and mitigate power line interference under certain conditions. This algorithm is compared with traditional notch filtering.

Chapter 5 presents and evaluates methods for detecting three types of contaminants related to a poor data acquisition system setup, namely clipping, quantization noise, and amplifier saturation.

Chapter 6 examines contamination due to the electrocardiogram (ECG) and motion artifact. A method for quantifying and mitigating ECG interference is evaluated along with a comparison of methods for quantifying motion artifact.

In Chapter 7, the biosignal quality assessment problem is approached from the top down. Signals are processed by a support vector machine (SVM) which is trained to detect any contamination in the sEMG. The performance of this classification is evaluated for the six types of noise examined in this thesis, as well as with varying signal-to-noise ratios (SNR).

Chapter 8 presents a summary of contributions and provides recommendations for future work in sEMG signal quality assessment.

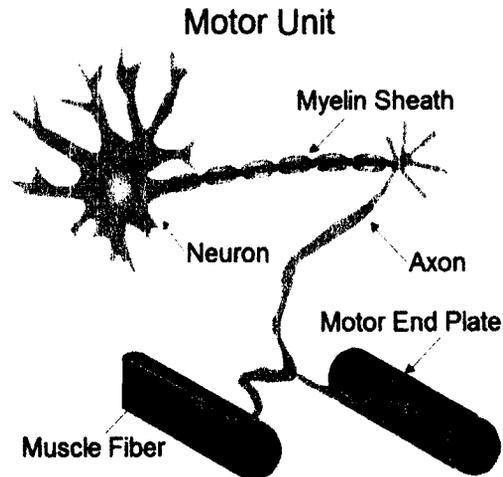
## **2 Literature Review**

### ***2.1 Introduction***

This chapter presents a background of relevant research in biosignal quality analysis using digital signal processing techniques, with particular emphasis on surface electromyography (sEMG). It serves to provide information regarding sEMG, summarizes previous research on biosignal quality analysis, and identifies gaps that will be addressed by the work presented in this thesis.

### ***2.2 Principles of Electromyography***

Electromyography (EMG) refers to the measuring of the electrical activity associated with skeletal muscle contractions. All muscles in the body are composed of cells organized to form muscle fibres, each of which are innervated by the axon of a motor neuron (and in rare cases, multiple motor neurons [4]) from the peripheral nervous system (Figure 2-1). The innervation zone of a muscle fibre is a given region (typically a few millimetres in length) where the fibre is innervated by axons from a given motor unit [5]. These motor neurons are stimulated by the motor system of the central nervous system, resulting in action potentials that travel along the axon and subsequently initiate action potentials in muscle fibres. These action potentials travel along the muscle fibres and cause the distance between cells that form the muscle to shorten, resulting in a muscle contraction.



**Figure 2-1: Innervation of muscle fibres by a motor neuron to form a motor unit**

Each motor neuron will produce a train of action potentials referred to as a single-fibre action potential train (SFAPT). All the muscle fibres innervated by a single motor neuron along with the motor neuron itself are referred to collectively as a motor unit (MU). The summation of all the SFAPTs in a MU is called a motor unit action potential train (MUAPT). The more MUs that fire, and the more often they fire, the higher the contractile force of the muscle will be [6]. The observed interference pattern, resulting from the superposition of MUAPTs under the electrode pickup area, is the EMG signal. The EMG signal is also referred to as the myoelectric signal (MES).

### ***2.3 Applications of Electromyography***

EMG was initially used to monitor abnormal neuromuscular conditions in patients; however, its applications have widened since the discovery of its ability to quantify muscle activity levels [6]. EMG applications now include ergonomic studies, exercise physiology, movement and gait analysis, rehabilitation, biofeedback, and control of powered prostheses [6].

Ergonomic studies have been used to identify problems with physical loading on the body when performing strenuous tasks such as heavy material handling in the workplace [6]. Due to the inability to observe force loadings and internal force distributions on the body, EMG

lends itself well to providing this additional information. Of particular note is the use for postural investigation as these problems can arise when a person is at rest under normal musculoskeletal loading conditions [6]. Load estimation and fatigue studies are often employed as part of ergonomics. The moving average of the rectified value of an EMG signal is related to the muscular force of an individual, although the force-EMG relationship can differ between patients [6]. Muscular fatigue can also be identified via EMG either by observing a decrease in the mean/median frequency of the power spectrum [7], or by a decrease in the average rectified value of the raw signal (although controversy exists with this approach [8]). Fatigue can be a useful indicator of problems related to ergonomics.

Psychological stress has also been shown to manifest itself in increased EMG activity, particularly in the trapezius muscle [6]. This suggests it may be possible to use EMG as an indicator for elevated stress while performing a task.

EMG has been applied to determine sequences of muscle activation during different exercises such as walking, speed walking, and running [6]. Analysis has identified agonist and antagonist muscles during different activities as well as correlating muscle-fibre conduction velocity in the vastus lateralis of the quadriceps with sprinting capability [6]. EMG has also shed light on the science of weight-training, showing that MU firing rate and recruitment are the determinants of muscle force. A variance in the firing threshold of MUs is what differentiates highly-conditioned strength athletes and amateurs with similar levels of muscle hypertrophy [6].

With respect to weight training-related injuries, EMG has been applied to determine musculoskeletal loadings during different exercises prone to injury. This information can then be used to determine safe and proper ways to perform these exercises while minimizing risk to the body [6].

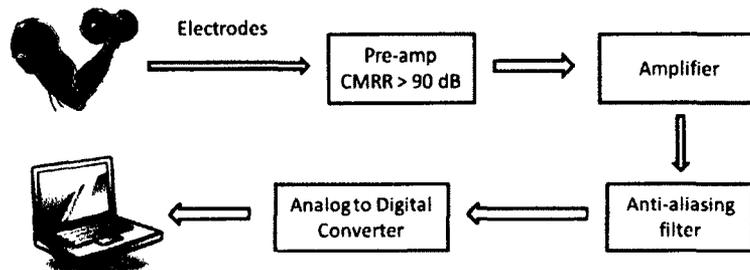
EMG has been used successfully in gait analysis by determining the sequence of muscle activation in normal and abnormal gaits [6].

The EMG of a muscle or muscle group during a rehabilitation phase allows the monitoring of the motor control and, if required, the retraining of those muscles via biofeedback [6]. Muscle retraining can be used to both increase or decrease muscle activation if required.

Perhaps one of the most significant applications of EMG is to the control of powered prostheses. If a limb is amputated, a prosthetic limb can be fabricated and attached in place of the amputated limb. The EMG from either residual muscles, in the case of a partial amputation, or alternative muscles, in the case of a complete amputation, is used as a control signal to the prosthesis. EMG information representing the signal intensity and MU firing rate are extracted from the signal and used to make control decisions [6]. Different strategies exist for interpreting control parameters and making control decisions.

## 2.4 Electromyography Data Acquisition

### 2.4.1 Needle vs. Surface Electromyography



**Figure 2-2: Typical EMG data acquisition setup**

EMG data acquisition consists of a number of stages (Figure 2-2). The EMG signal is measured from the patient's muscles via electrodes. An EMG signal can be recorded from a

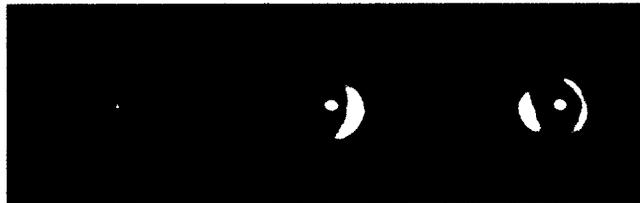
subject invasively using needle electrodes, or non-invasively using surface electrodes [9]. Needle electrodes will record from fewer muscle fibres than surface electrodes, offering greater spatial selectivity than surface electrodes. Needle electrodes may be more susceptible to motion artifact even when the subject appears stationary due to small movements within the muscle tissue [9]. On the other hand, surface electrodes are easier to apply than needles, more suitable to patient comfort, and carry a much smaller risk of infection or laceration. EMG collected from surface electrodes is susceptible to the low-pass filtering properties of the skin and fat tissue layer between the muscle fibres and surface electrodes [10]. EMG recorded with surface electrodes is often referred to as surface electromyography (sEMG). Due to its non-invasive nature, sEMG is often the preferred method of signal acquisition unless increased spatial selectivity is required, in which case needles should be used [9]. This thesis is focused solely on sEMG.

Factors such as electrode placement, orientation, electrode configuration, and electrode type can influence the sEMG being measured. The SENIAM project (Surface Electromyography for the Non-Invasive Assessment of Muscles) is a compilation of research to serve as a guide in best practices for reliable, repeatable sEMG acquisition [11].

#### **2.4.2 Surface Electrodes**

SEMG electrodes will either be dry or gelled. Dry electrodes (e.g., stainless steel, silver, gold, brass, or anodized aluminum [12]) are in direct contact with the skin (i.e., no gel in between) and are used when gelled electrodes are not feasible due to size or geometry constraints such as with electrode arrays [13]. Dry electrodes are also used for long-term recording as the electrolytic gel can dry up in a matter of hours [14]. Dry electrodes will have higher impedances than gelled electrodes [13]. Hence, dry electrodes generally provide poorer signal quality and are more susceptible to motion artifact.

Gelled electrodes are coated on the surface with an electrolytic gel. The purpose of the gel is to facilitate transmission of current from the muscle fibre to the electrode via the skin which ensures minimal noise. The most common type of gelled electrode is silver-silver-chloride (Ag-AgCl, Figure 2-3) [13]. Gelled electrodes can be disposable or reusable; however disposable electrodes are common as they are light and are fitted with a highly-adhesive backing to minimize noise associated with electrode movement [13].



**Figure 2-3: MVAP-II Ag-AgCl Electrodes.**

Electrode shape and size can vary, however circular electrodes 8 to 10 mm in diameter are the most common [6] [11].

### **2.4.3 Electrode configurations**

Three main types of electrode configurations for sEMG measurement are unipolar, bipolar, and double differential. Unipolar measurements are taken from a single electrode with respect to a reference (ground) electrode placed on an electrically inactive site on the body, usually above a bony area. A bipolar configuration uses two recording electrodes and one reference electrode. It measures the difference between the two recording electrodes. The double differential configuration measures the difference between two adjacent bipolar electrode configurations where the centre electrode is common. SEMG is usually recorded using a bipolar configuration aligned with the muscle fibres and away from the innervation zone and tendons with an inter-electrode distance of the smaller of 20 mm and  $\frac{1}{4}$  the length of the muscle fibre [11]. Other electrode configurations offer advantages in terms of ease of recording and muscle selectivity [15–17]. The electrode configuration will also apply

a transfer function such that sEMG recorded under different configurations are generally not directly comparable [2].

#### **2.4.4 Amplifier**

Amplification is essential in sEMG signal acquisition to improve the signal-to-noise ratio (SNR) [13]. A significant advantage to using a differential (bipolar) or double differential electrode configuration is to leverage the common mode rejection capability of the differential amplifier. Common signals to both electrodes such as those due to noise or deep muscle fibres are rejected [13]. The common mode rejection ratio (CMRR) for reliable sEMG acquisition should be at least 95 dB [9].

An amplifier is often separated into two components: a pre-amplifier and an amplifier (or power amplifier). The pre-amplifier is often as close to the electrodes as possible (and in some cases, contained within the electrodes) to ensure minimal contamination. The pre-amplifier will provide a small amount of voltage amplification and pass the signal to the amplifier which provides more substantial amplification.

sEMG is typically in the range of 6 mV peak to peak [13]; therefore, a gain of at least 1000 is recommended [9], although a higher gain may be needed for low amplitude signals [9]. The majority of the sEMG bandwidth occurs between 10 Hz and 500 Hz [9]. Other frequencies can be filtered out using amplifier settings although, notch filtering (e.g., for 60 Hz power line interference) is not recommended as it can cause loss of useful sEMG signal information [9].

The input impedance of the amplifier needs to be considerably higher than the skin impedance to avoid distortion due to input loading [13]. For Ag-AgCl electrodes, an input of

at least 10 M $\Omega$  is recommended [13]. Dry electrodes will result in a much larger skin impedance and can require an amplifier input impedance as high as 1 G $\Omega$  [13].

#### **2.4.5 Analog-to-Digital Converter**

An analog-to-digital converter (ADC) is required to sample and quantize the analog input from the amplifier. The ADC should have sufficient bit resolution to accurately represent the signal, and it should have a sufficient sampling rate to represent the entire signal bandwidth. A resolution of 12 bits is adequate for most applications [9].

The sampling rate should be set to at least twice the bandwidth of the signal as given by the Nyquist sampling theorem [18]. For sEMG this is at least 1000 Hz but often higher sampling rates are used to account for the finite roll-off of the anti-aliasing filter and avoid signal loss [9].

#### **2.4.6 Skin Preparation**

Proper skin preparation in sEMG acquisition is essential to minimize electrode movement and to reduce skin impedance [9]. Skin preparation is more significant if dynamic contractions are to be used which can increase the risk of motion artifact [9]. As discussed below, skin preparation consists of removing the hair, cleaning and abrading of the skin, and securely applying the electrodes as specified in [11].

Moderate to excessive amounts of hair can compromise the adhesive abilities of the electrodes, increase the skin-electrode impedance, and increase the potential for motion artifact. It is recommended to shave the area of the skin which will be in contact with the electrodes [11].

The skin should be cleaned with alcohol and then allowed to dry as a wet surface could inhibit electrode adhesion [11]. The skin should then be abraded by a fine sand paper or a gauze pad with abrasive gel [9]. A moderate amount of force should be used here to remove the outermost layer of the skin (i.e., stratum corneum) as it contributes most to the skin impedance [19]. Abrading the skin has also been found to reduce the skin stretch reflex which is a significant contributor to motion artifact [20]. After skin abrasion, the skin should have a light red appearance [9].

The skin impedance should be below 30 k $\Omega$  and ideally below 10 k $\Omega$  for best results [9]. If skin impedance is too high, this may warrant another round of skin preparation and the application of new electrodes [9]. It should be noted that skin impedance will decrease with time after application of electrodes and can take 2-3 minutes to stabilize [21]. Note that skin impedance is frequency dependent and tends to be highest for frequencies below 10 Hz and slightly lower for frequencies up to 100 Hz [22]. Impedance measurements can vary depending on the frequency of the impedance testing device. Proper skin preparation will drastically decrease the low frequency impedance of the skin [22].

#### **2.4.7 Positioning of Electrodes**

The precise positioning of the electrodes will vary depending on the muscle being recorded. In general, electrodes should be placed on top of the muscle being measured and in the direction of its muscle fibres [11]. For differential recordings, the muscle innervation zone should be avoided [6]. The action potentials generated by this motor unit will travel along the muscle fibre in either direction and therefore, differential electrodes placed on top of the innervation zone will likely not measure the true superposition of action potentials constituting the MUAPTs. It has been shown that sEMG amplitude is considerably smaller when measured from the innervation zone [5]. To avoid measuring from the innervation zone, guidelines in [11] should be followed.

## 2.5 Sources of Contamination in Surface Electromyography

A number of sources of contamination exist which can compromise the signal quality in sEMG. Contamination occurs when electrical potentials arising from sources other than that being measured are captured by the bioinstrumentation equipment, or if there is a problem with the equipment itself. Sources of contaminants include power line interference [9], baseline wander [9], electrocardiogram (ECG) artifact [9], electrode-induced motion artifact [3], skin-stretch-induced motion artifact [23], motion artifact due to cable-movement [24], crosstalk from adjacent muscles [3], ADC-related clipping and quantization noise, and amplifier saturation. These contaminants can be characterized as either interference from an unwanted signal, measurement-induced artifact due to a problem with how the signal is being measured, or problems with the instrumentation setup as shown in Table 2-1.

**Table 2-1: SEMG contaminants classified as interference, measurement artifact, or instrumentation problem**

<b>Contaminant</b>	<b>Type</b>
Power line	Interference
ECG	Interference
Crosstalk	Interference
Radio Frequency	Interference
Baseline wander	Measurement artifact
Motion artifact	Measurement artifact
Electrode Lift	Measurement artifact
ADC clipping	Instrumentation problem
Quantization noise	Instrumentation problem
Amplifier saturation	Instrumentation problem

Power line interference is caused by capacitive coupling of the acquisition equipment to the power line [25]. Ampere's Law states that a time-changing magnetic field can induce

current flow and vice versa. A stray electric field can also directly induce current flow. Intense electric and magnetic fields from the power line (50 Hz in Europe or 60 Hz in North America) can induce current flow in EMG recording equipment which will manifest as unwanted changes in potential.

ECG artifact occurs when EMG is measured from the torso, in close proximity to the heart [26]. The PQRST waves commonly seen in the ECG recording will contaminate the EMG signal.

Crosstalk in sEMG occurs from sEMG signals originating from muscles other than that (or those) of interest. As much as 30% of the sEMG detected directly over the muscle fibres can also be detected at nearby, but otherwise inactive, sites on the body [27].

Radio frequency interference is due to electromagnetic radiation, often due to telecommunications equipment, which can be picked up by the recording equipment.

Baseline wander occurs when the signal baseline changes over time. This can occur as a result of poor electrode contact either due to a malfunctioning electrode or excessive sweat or hair preventing proper adhesion. This improper contact causes a time-varying skin impedance which appears as a low frequency change in potential (less than 1 Hz) [28].

Motion artifact occurs when there is motion of either the electrode or cable, or stretching of the skin. Electrode motion will cause an artifact similar to baseline wander. This occurs as the half-cell potential of the electrode-electrolyte interface is disturbed due to motion [20]. This disturbance causes a charge redistribution at the electrode-electrolyte interface; the artifact occurs as a result of the charge attempting to re-establish an equilibrium. This charge redistribution is more significant for polarizable electrodes. As Ag-AgCl electrodes

are non-polarizable, they tend to be robust to electrode-induced motion artifact [20]. A more significant cause of motion artifact for non-polarizable electrodes is due to the skin-stretch [20]. When an electrode is either pulled or pressed upon, the skin beneath will stretch based on location and amount of adipose tissue. Larger amounts of stretch will induce more significant artifacts [20]. Skin abrasion should be performed as it reduces the skin potential between the outermost layers which, in turn, will reduce the amount of skin-stretch-induced motion artifact that can occur [20].

Motion of the electrode cables or nearby movement from other sources can also induce motion artifact in an sEMG signal. Such movement can cause a triboelectric effect where static electricity discharge induces current flow in the recording equipment [29]. This effect can be reduced by proper cable design and the use of active electrodes with pre-amplifiers located on the electrode.

Electrode lift is when the electrode does not make a good contact and begins to lift off the skin either due to a failure of the electrode to adhere to the skin, or due to excessive tugging on the electrode. This results in a poor conduction of current between the skin and electrode.

Clipping and quantization noise are both results of digitizing an analog signal with an ADC. The ADC can only represent a limited number of discrete values within its given range. Values outside this range are clipped and values within this range are quantized where the round-off error is known as the quantization error or quantization noise.

Amplifier saturation occurs when the amplifier is operating outside its linear range [30]. This range will be dictated by the amplifier gain and the maximum output voltage, or saturation voltage.

## **2.6 Biosignal Quality Analysis**

### **2.6.1 Categories of Biosignal Quality Analysis**

Quality analysis of sEMG, or any biosignal can be broken down into the following four categories: 1) detection, 2) identification, 3) quantification, and 4) mitigation. Detection is the simplest stage as it consists of determining only if contamination is present in the signal. This is useful in biosignal quality analysis as it can serve as an indicator that an sEMG is contaminated and should not be used.

Identification often follows from detection and serves to isolate what type (or types) of contamination are present in the signal. This information is more useful than detection alone as an sEMG operator may be able to correct the problem if the specific source is known.

Quantification may follow from either detection or identification. Here, the amount of contamination is measured and expressed numerically. This provides the operator with the ability to make a decision to accept or reject a signal based on the amount of contamination present. In many cases, contamination is always present (e.g., power line interference, motion artifact, quantization noise) but can be safely ignored at low levels. In some cases, quantification is possible but identification is not. The amount of contamination can be estimated but the specific source of said contamination is not identifiable. This is adequate if one intends to quantify contamination as a whole and is not interested in what individual types are present.

The last category of sEMG signal quality analysis is mitigation and it is often the most difficult to perform. To perform mitigation, some or all of the contaminant is removed while preserving the clean sEMG signal as much as possible. If mitigation is possible, a moderately or severely contaminated signal can be rectified without having to acquire new

data. Mitigation usually follows from quantification, but not necessarily from identification. Mitigation may be possible but not warranted if it results in more signal degradation than not attempting mitigation at all. This can be the case when the level of signal contamination is low.

Although quantification and mitigation are useful, it is often sufficient if detection and potentially identification are performed. This will indicate a problem in the data acquisition setup that the operator should be able to easily rectify and then the data collection can be repeated.

The biosignal quality assessment can be approached from two perspectives. A bottom-up approach involves examining the signal for individual contaminants in isolation using individually tailored identification (or quantification or mitigation) methods. This information is then combined to make a more informed analysis. For instance, detection of ECG interference could also be identified as motion artifact as both can appear as lower frequency components. Both types of contamination identified together may warrant further, more sophisticated methods to differentiate the two. Quantification of each contaminant would then be performed and, if possible, mitigation. The end result would be an overall assessment of the signal where each contaminant is detected, identified, quantified, or mitigated in isolation. Alternatively, a top-down approach can be employed which involves examining the signal as a whole for any deviation from a typical clean sEMG signal using statistical analysis or pattern classification techniques. This approach can facilitate the detection stage (and can provide a rough estimate of quantification), but does not need to be tailored to a single contaminant.

## 2.6.2 Biosignal Quality Analysis in Surface Electromyography

Signal quality analysis in sEMG is difficult as a typical sEMG signal will appear as random noise (and is often modeled as such [31]). It is difficult to visually discern contamination in sEMG except when the level of contamination is extreme. There is no typical time-domain waveform shape in sEMG such as the PQRST waveform in the electrocardiogram (ECG). The sEMG power spectrum does have a general shape [31]; however, significant variability can exist between different subjects, muscles (types, sizes, fibre types), contraction levels, and contraction types (static or dynamic). More sophisticated methods of quality assessment are thus required.

Although related work has been done on the various categories of biosignal quality analysis, no other works have looked at biosignal quality analysis as a whole in the context of sEMG. In [32], the authors evaluate a method for detecting contaminated segments of esophageal EMG. This method consists of extracting four features from the EMG, namely the maximum area between two consecutive zero crossings, the skewness of the signal amplitude distribution, the signal-to-noise ratio (SNR), and the spectral deformation. Measuring the deviation in feature values can lead to the detection of signal contamination with good confidence. This study was performed only on diaphragmatic EMG and the types of contamination in the signals were unlabelled (i.e., identification was not performed). Furthermore, the gold standard was the visual classification performed by analysts which can be subjective. The results from this paper suggest a potential method of contamination *detection* only, albeit when detection via visual inspection is also possible.

In [3], the authors address the detection and identification, and quantification categories by developing computer algorithms to objectively quantify contamination in diaphragmatic EMG. They evaluate four indices: 1) a signal-to-motion artifact ratio (SMR) to quantify

motion artifact; 2) a ratio describing the peak of the power spectrum to ensure EMG is being recorded (instead of noise); 3) a signal-to-noise ratio where noise, in this case, is assumed to have a flat power spectrum and is contributed by a variety of unknown sources; 4) an index of spectral deformation to detect crosstalk or non-QRS ECG interference. These methods were evaluated on simulated EMG power spectra contaminated with simulated ECG and motion artifact power spectra, as well as on real signals. The evaluation of simulated signals was to determine relationships between the various indices and the levels of contamination in the signal. The authentic signals were contaminated with motion artifact, EMG interference from the esophagus, and ECG interference. Correlations between the amount of noise in the signal and the computed indices were calculated. A number of assumptions are made to this end (e.g., the EMG power spectrum is linear for low frequencies, noise has uniform power and is unaffected by EMG at high frequencies). The paper concludes by presenting acceptable ranges for the four indices and by claiming their values can quantify the amount of contamination in diaphragmatic EMG.

Detection is also examined in [33] within the context of monopolar multichannel sEMG. Each signal is characterized by two non-overlapping-windowed standard deviations computed with different window lengths. Thresholds are chosen which classify signals as very poor, slightly poor, or good, in accordance with the visual classification of two expert operators. As no dataset partitioning (training/testing) was used in determining the threshold, this method is likely biased and prone to overfitting the sample data. Nevertheless, the results suggest a possible metric for detecting contamination in monopolar multichannel sEMG.

The work presented in [34] also looked at detecting noisy channels in multichannel sEMG. SEMG was separated into training and testing databases to avoid bias. Datasets consisted of SEMG from a variety of subjects, contraction types, muscle groups, and force levels. Three features were extracted from the data and a multivariate outlier detection method was

applied. Performance was evaluated against a gold standard derived from the classifications of three human operators. High accuracy was reported, although the method assumes that at least 75% of the recorded channels are uncontaminated. No attempt was made to quantify the degree of sEMG contamination. As such, it is unknown whether the contaminated channels were severely contaminated or mildly contaminated. If operators were able to visually discern the contamination, likely the contamination was moderate to severe and the question can be raised as to if this method is able to detect contamination that is not visually discernible. The authors also note that this method may not be applicable for smaller muscle groups, or sEMG collected without a high-density electrode array.

The works presented above have certain limitations which are addressed in this thesis. In [32] and [34], only contamination detection is performed and results are based on visual classification which is prone to inter-expert variability and is insensitive to low levels of contamination. By artificially inserting known quantities of contamination, this thesis may include signals with measurably varying degrees of contamination and is not limited to severely contaminated signals.

In [32], the issue of detection when visual inspection is inconclusive (i.e., smaller amounts of interference) is not addressed. This issue is addressed further in this thesis, along with the identification, quantification, and mitigation categories. In addition, sEMG recorded from different sites of the body are considered.

In [3], identification or quantification of individual types of contamination (i.e., motion artifact or ECG, but not both) is not possible as they both impact the SMR. Furthermore, other EMG from other muscles are not considered in the study. These limitations are

addressed in this thesis by including sEMG from a variety of muscles and attempting to quantify and, in some cases, mitigate individual contaminants.

Both [33] and [34] are limited to electrode array data and present no method of contaminant quantification. Multichannel recordings are used to search for outliers which would not be applicable to single channel recordings or low density arrays. This thesis will use bipolar recordings collected from a single channel which is less limiting as the results could easily be extended to arrays.

## ***2.7 Summary***

Signal analysis and interpretation of results are still obstacles in sEMG [6]. There currently do not exist any universally accepted methods or criteria to quantify contamination in sEMG [3]. Many types of signal contaminants exist and each one manifests in a different manner. Currently, a proper background in sEMG signal acquisition is critical to ensure an acquisition setup is satisfactory, and that signal contaminants are at acceptable levels. SEMG setup and collection is normally performed only by a trained sEMG technician due to the aforementioned complexities with its acquisition. Untrained individuals attempting to acquire and analyze sEMG will likely encounter signal contamination due to an improper setup and be unaware of its presence. As such, development of an automatic method of sEMG quality assessment is warranted [1].

## 3 Surface Electromyography Data

### 3.1 Introduction

Both simulated and real sEMG are used to develop and evaluate the artifact detection methods in this thesis. This chapter first presents a method for simulating sEMG. This method is used to generate large datasets consisting of signals which are guaranteed to be contaminant free. Second, a protocol is defined to support the repeatable reliable acquisition of real sEMG signals which are sufficiently free of contamination. Data acquired with this protocol can serve as examples of real clean sEMG and the signals collected were manually verified to ensure their quality. The simulated and real clean sEMG can then be artificially contaminated and the SNR will be known for evaluation purposes. Research in this thesis was approved by the Carleton University Research Ethics Board. Five volunteers with no known neuromuscular disorders were selected and informed consent was provided (Appendix C:). Data collection procedures are based on the guidelines from [11].

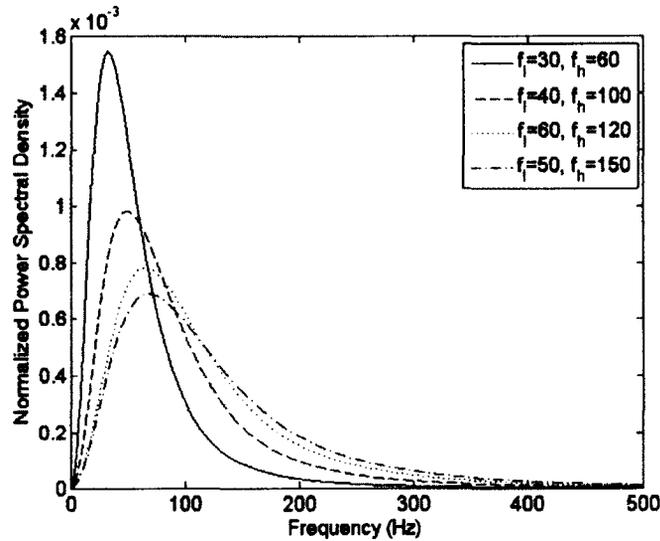
### 3.2 Surface Electromyography Simulation

Each biosignal quality assessment method presented in this thesis is first evaluated with simulated sEMG as this provides a much larger dataset of signals guaranteed to be clean with which the method can be tested. SEMG data are simulated by passing white Gaussian noise through a shaping filter [31] at a sampling rate of 1 kHz although the sampling rate can vary depending on the method being tested. The shaping filter transfer function is given in (3-1).

$$H_{EMG}(f) = \frac{jKf_h^2 f}{(f_l + jf)(f_h + jf)^2} \quad (3-1)$$

In (3-1),  $f_l$  and  $f_h$  are parameters which adjust the shape of the EMG spectrum and  $K$  is the gain factor. Four shaping parameter ordered pairs were chosen as shown in Figure 3-1

based on [6]. Typically, these shaping parameters are varied in time to support the nonstationarity of the EMG signal; however, we use short signal durations for our simulations (less than 5 seconds) so that stationarity can be reasonably assumed, especially for isometric, isotonic contractions. The gain factor  $K$  was adjusted to normalize the power of the sEMG signal to 1.



**Figure 3-1: Effect of parameters on simulated sEMG power spectral density as based on [6]** Other models of simulated sEMG were not considered in this work. More sophisticated models for simulating sEMG exist, however they would still not represent the same variability as in real sEMG. We use simulated sEMG to produce large datasets which would not be feasible if only real sEMG were used. We then ensure results are consistent with real sEMG and a much smaller dataset. It is unlikely that using other methods of simulated sEMG would have any significant impact on results.

### 3.3 Data Collection Equipment

SEMG data were collected using pre-gelled Ag-AgCl surface electrodes (MVAP-II, MVAP Medical Supplies Inc., Newbury Park, CA, USA). The electrodes were circular and 10 mm in

diameter. The electrodes were arranged in a bipolar configuration with an inter-electrode distance of 20 mm in adherence with the guidelines from [11].

The signals obtained via the bipolar electrodes were amplified (Model 15A54 Grass amplifier, Grass Telefactor, West Warwick, RI, USA) and then sampled (12-bit, PCI-6071E, National Instruments, Austin, TX, USA). The digitized signals were processed offline with Matlab (Mathworks, Natick, MA, USA).

### **3.4 Skin Preparation**

The following skin preparation was performed prior to applying electrodes to each active recording site for each volunteer.

1. Determine the location of the electrode contact area using a tape measure and the guidelines from [11] for the particular muscle being recorded (see Section 3.5). Mark the electrode centers on the skin using a marker.
2. Clean the contact area at the recording site on the skin with alcohol swabs and let dry completely (approximately 20 seconds). Moisten a gauze pad with a moderate amount of Nuprep Skin Prep Gel (D.O. Weaver and Co., Aurora, Colorado, USA) and spread the gel over the recording site. Using a circular rubbing motion, abrade the skin using moderate force using a dry portion of the gauze pad for approximately 10 seconds or until skin is dry and slightly red/pink.
3. If the marks on the skin representing the electrode locations were removed in the process, mark them again.
4. Remove electrodes from adhesive backing and apply EC2 electrode cream (Grass Telefactor West Warwick, RI, USA) to the contact area on the electrode. Place the electrodes on the skin on the marked locations and attach the electrode leads. The use of EC2 electrode cream serves a similar purpose as the electrolyte gel on the

electrodes. It is used here to mitigate any issues that may occur if the electrolyte gel has dried out.

5. Wrap the electrode wires to the skin to prevent tugging using Life sports tape (Shoppers Drug Mart, Toronto, ON, Canada).
6. Perform an electrode impedance test using the Grass amplifier electrode impedance test function [35]. If the impedance is greater than 10 k $\Omega$  then remove electrodes and perform the skin preparation again. Note that a threshold of 10 k $\Omega$  was chosen from [9], although 5 k $\Omega$  or less was observed to be optimal.

### ***3.5 Muscle Contractions and Landmarks***

Three muscles (biceps, quadriceps, and tibialis anterior) were chosen for recording such that a range of muscle sizes (in terms of number of fibres) were covered. The electrode placements and a description of the contractions are listed in Table 3-1, Table 3-2, and Table 3-3. These placements are based on the guidelines from [11].

**Table 3-1: Landmarks and Contraction Description for Biceps Brachii**

<b>Muscle</b>	Biceps Brachii
<b>Starting Position</b>	Sitting on a chair with the elbow flexed at a right angle and the dorsal side of the forearm in a horizontal downwards position (supination). The palm should be pressed against the bottom of a table or desk.
<b>Electrode Location</b>	On the line between medial acromion (highest part of shoulder bone) and cubital fossa (frontal elbow joint) at 1/3 the distance from the cubital fossa (see Figure 3-2).
<b>Electrode Orientation</b>	Parallel to the line between medial acromion and cubital fossa.
<b>Reference Electrode</b>	On wrist bone.
<b>Contraction</b>	From starting position, press against supinated forearm in direction of extension.



**Figure 3-2: Contraction starting position and electrode placement for the biceps brachii as based on [11]**

**Table 3-2: Landmarks and Contraction Description for Rectus Femoris**

<b>Muscle</b>	Quadriceps femoris (rectus femoris)
<b>Starting Position</b>	Sitting on a table with legs hanging over the edge. One leg should be hanging right in front of the table leg and is fastened to the table by a strap around the ankle. The knee joint is at 90 degrees and the foot is above the ground (see Figure 3-3).
<b>Electrode Location</b>	On the line 50% of the distance between the anterior superior iliac spine (front of hip bone) and the superior part of the patella (see Figure 3-3).
<b>Electrode Orientation</b>	In the direction of the line between the anterior superior iliac spine (front of hip bone) and the superior part of the patella.
<b>Reference Electrode</b>	On the malleolus (ankle).
<b>Contraction</b>	Attempt to extend the leg away from the table without moving the thigh or the hips. The edge of the table can be held with the hands for stability.



**Figure 3-3: Contraction starting position and electrode placement for the rectus femoris as based on [11]**

**Table 3-3: Landmarks and Contraction Description for Tibialis Anterior**

<b>Muscle</b>	Tibialis anterior
<b>Starting Position</b>	Sitting on a table with legs hanging over the edge. One knee should be bent such that the bottom of the foot is pressed against the table leg. The foot is fastened to the table with a strap around the ball of the foot. The foot does not make contact with the ground (see Figure 3-4).
<b>Electrode Location</b>	1/3 the distance between the proximal tip of the fibula and medial malleolus (see Figure 3-4).
<b>Electrode Orientation</b>	In the direction of the line between the tip of the fibula and medial malleolus.
<b>Reference Electrode</b>	On ankle bone.
<b>Contraction</b>	Start with the foot at full eversion and plantarflexed. Attempt to invert and dorsiflex the foot against resistance.



**Figure 3-4: Contraction starting position and electrode placement for the tibialis anterior as based on [11]**

Muscle contractions were chosen to be isometric (constant length) and isotonic (constant force) to minimize motion artifact and ensure consistency between recordings. Three levels

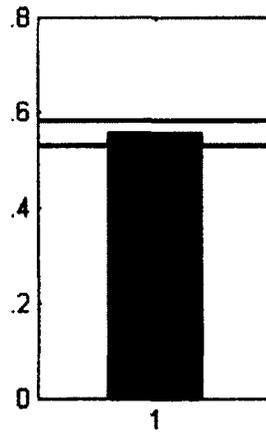
of force (20%, 40%, and 60% of the maximum voluntary contraction (MVC)) were chosen for each volunteer. For each subject and muscle, three MVC calibration recordings are performed where the subject contracts at his/her maximum for three seconds. Counter-resistance was provided using the bottom of a desk for the biceps, and a Velcro strap for the rectus femoris and tibialis anterior. For each of the three MVC recordings, the MVC is computed as the maximum of the root mean square (RMS) of the signal using a sliding window with a length of 1 second. If  $L$  is the number of samples corresponding to a window length of 1 second ( $L = \frac{f_s}{1s}$  for sampling rate  $f_s$ ) and  $x[n]$  is the sEMG signal then this computation is realized as shown in (3-2) where  $*$  is the convolution operator.

$$MVC = \max(\sqrt{x^2[n] * w[n]}) \quad (3-2)$$

$$w[n] = \begin{cases} \frac{1}{L}, & 0 \leq n \leq L-1 \\ 0, & n < 0 \text{ or } n > L-1 \end{cases}$$

The MVC value that is used thereafter is the maximum of the three MVC calibration recordings.

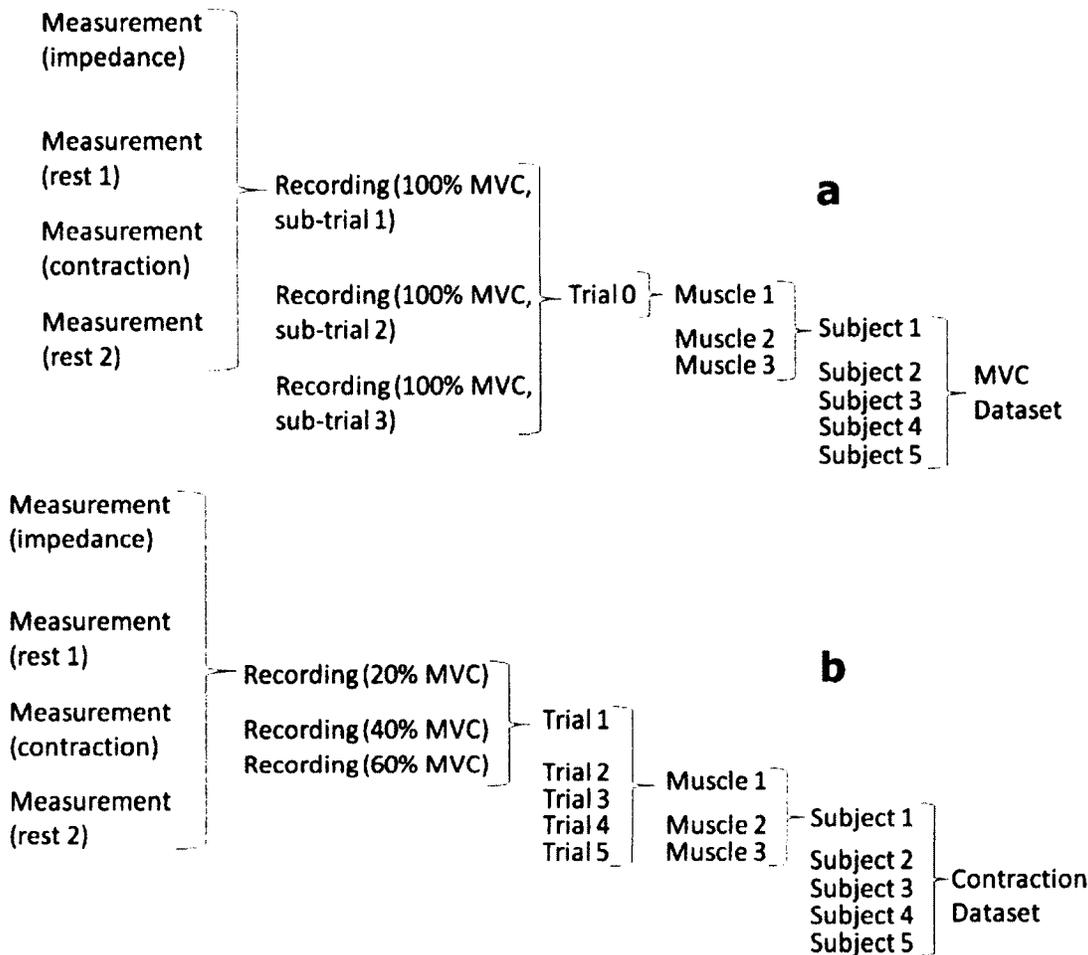
For contractions at fractions of the MVC, a bar graph is displayed on the screen (Figure 3-5) to provide feedback to the subject regarding their contraction level. Two horizontal lines are drawn, indicating the region of 5% tolerance on that MVC value. The subject is instructed to try to keep the bar in that range. The colour of the bar changes from red to green to blue if the bar is above, within, or below the target range, respectively. The bar graph is updated approximately every 10 ms and computes the instantaneous contraction level as the RMS from the most recent 1 s of the signal.



**Figure 3-5: MVC feedback bar graph**

### **3.6 Data Collection Procedure**

For each volunteer, the procedure consisted of recording five *trials* for three different muscles at three different fractions of their maximum voluntary contraction (MVC) (i.e., 20%, 40%, and 60% MVC). Initially, the MVC level must be determined for each of the three muscles (this *MVC calibration trial* is referred to as *trial 0*) as specified in the previous section (Section 3.5). This is then followed by 5 *contraction trials*. Each trial, including the MVC calibration trial (*trial 0*) involved three *recordings*. Each *recording* is made at a different fraction of the MVC for the *contraction trials*. In the case of *trial 0*, the three recordings are all at 100% MVC and the maximum over all three recordings is taken as the overall MVC. Each *recording* consisted of four recording parts (referred to here as *measurements*) consisting of: 1) an impedance measurement, 2) a rest measurement, 3) a contraction measurement, and 4) a second rest measurement. This classification of measurements, recordings, and trials is summarized in Figure 3-6.



**Figure 3-6: Hierarchy of terminology for sEMG MVC dataset (a) and contraction dataset (b).**

Each measurement was examined for validity by ensuring the spectrum shape was typical of sEMG with no low frequency or power line frequency spikes, that the signal amplitude fell within the range of the ADC (no clipping or quantization noise), and that the rest recording was very low amplitude with no baseline wander. If any of these conditions were not met, the measurement was repeated. The file format for saving the files was "s<subject\_num>m<muscle\_num> t<trial\_num>l<mvc\_level\_num>(i | r1 | r2 | c).daq" where angle brackets indicate numbered identifiers. For example, an impedance recording for subject #2, muscle #1 (biceps), trial #4, and MVC level #1, the file would be saved as s2m1t4l1i.daq. Note that trial 0 has no level (always 100% MVC) but three sub-trial

recordings are performed from which to get the maximum value to use as the MVC. The level number indicates the sub-trial number for trial 0 only.

Data collection proceeds as follows:

1. Run the amplifier configuration software. Program channels 1 through 3 to the settings indicated in Table 3-4.

**Table 3-4: Amplifier settings**

Setting	Value
Low Freq (HPF)	0.3 Hz
High Freq (LPF)	1000 Hz
Gain	2000*
Line Filter	Disabled

\* Note that this value was increased if the signal amplitude was observed to be small (approximately less than 20% of the ADC range)

2. Run the data acquisition software and set the data acquisition settings as given in Table 3-5.

**Table 3-5: Data acquisition settings**

Setting	Value	
	MVC Calibration	% MVC Contraction
Number of channels	3	3
Sampling rate	3000.3 Hz	3000.3 Hz
Sampling duration	3 s	10 s
ADC range	±10 V	±5 V
MVC level	N/A	0.2 for level 1, 0.4 for level 2, 0.6 for level 3

3. Apply the skin preparation and attach the electrodes based on the guidelines in Section 3.4. Attach all electrodes and ensure each muscle is connected to a separate recording channel. Channels 1 through 3 should be connected to one of the biceps, quadriceps, and tibialis anterior, respectively. Wrap the electrode wires to the body close to the electrode leads to minimize tugging on the electrode leads.

4. **MVC calibration:** Note that the first impedance-rest-contraction trial should be numbered as trial 0 and this indicates the MVC calibration trial. The recording level number indicates the MVC sub-trial in this case. The ADC range will need to be increased to 10 V for this and the sampling duration reduced to 3 seconds. Once MVC calibration is complete, reset the settings as given in Table 3-5. Three sub-trials are performed and the maximum is taken as the MVC for that muscle. For example, the second MVC calibration sub-trial for a contraction measurement for subject #1 and muscle #3 would be saved as "s1m3t0l2c.daq".

5. **Impedance Test:** Run the Grass Telefactor amplifier electrode impedance test function (subject is relaxed) and ensure the electrode impedance is less than 10 k $\Omega$ . Save the recording as

"s<subject\_num>m<muscle\_num>t<trial\_num>l<mvc\_level\_num>i.daq" where <subject\_num>, <muscle\_num>, <trial\_num>, and <mvc\_level\_num> are numbers indicating the subject, muscle, trial, and MVC level respectively.

If the impedance is too high, wait a minute and perform another repetition. Mark the current data file as erroneous by appending "\_error#" to the filename (where # is incremented for each time an error occurs for that filename). If the impedance is not below 10 k $\Omega$  after 5 repetitions, remove and discard electrodes and perform the skin prep again. Note that this did not occur for the data collected in this thesis, indicating well-performed skin prep.

6. **Rest1:** Specify the save file as

"s<subject\_num>m<muscle\_num>t<trial\_num>l<mvc\_level\_num>r1.daq" and instruct the subject to relax the muscle being recorded. Once the subject's muscle is relaxed, start recording. If a problem is identified with the given recording, perform another repetition and mark the current file as erroneous by appending "\_error#" to the filename.

7. **Contraction:** Specify the save file as

"s<subject\_num>m<muscle\_num>t<trial\_num>l<mvc\_level\_num>c.daq" and instruct the subject to perform the appropriate contraction for the first muscle. If this is the MVC calibration trial (number 0), the subject is instructed to contract as hard as they can for 3 seconds. Once the subject is performing the contraction, start recording. For subsequent trials (numbers 1-5), the subject should try to maintain the contraction level of corresponding to the current MVC level (20%, 40%, or 60% MVC). The subject's predicted MVC level should be relayed visually to them in real time.

Observe the signal as it is recorded and when recording is complete, plot the power spectrum and ensure everything is as expected. If a problem is identified with the given recording, perform another repetition and mark the current file as erroneous by appending "\_error#" to the filename.

8. **Rest2:** Specify the save file as

"s<subject\_num>m<muscle\_num>t<trial\_num>l<mvc\_level\_num>r2.daq" and instruct the subject to relax the muscle being recorded once again as in step 7.

9. Repeat steps 5-9 for each of the 3 muscles being recorded, for 5 trials per muscle (and one MVC calibration trial), and for each MVC level recording (20%, 40%, and 60% MVC). Note that each MVC level should be performed in sequence for a given trial and muscle. Once 5 trials are complete, the next muscle can be recorded in the same manner.

10. When all recordings are complete, carefully remove all electrodes from the subject's skin and clean the contact area with alcohol wipes and clean gauze pads.

### ***3.7 Removal of Contaminated Data***

To ensure all data recorded was sufficiently free of contamination, an SNR was estimated for each signal by measuring the power of the sEMG signal and dividing it by the power of the

rest signal taken immediately afterwards (rest2). If the SNR was below 18 dB, the signal was deemed contaminated and was removed from the dataset. Out of the 225 signals recorded, 58 signals were removed in this manner.

### ***3.8 Summary***

This chapter described the methodology in simulating sEMG and collecting real sEMG for use in evaluation of the methods in this thesis. The purpose of the simulated sEMG is to facilitate the generation of a very large sEMG dataset. The real sEMG provides the variability likely to be encountered in practice on a much smaller dataset. The procedure for collection of real sEMG is to ensure contamination in the collected sEMG is minimal.

## **4 Power Line Interference Estimation and Mitigation**

### ***4.1 Introduction***

Power line interference is a major source of noise in sEMG measurements due to capacitive and magnetic coupling between the subject or sEMG data acquisition apparatus, and the AC power source (60 Hz/120 V in North America) [25]. It has been demonstrated that power line interference can compromise the analysis of EMG signals [36–38].

There exist methods of mitigating power line interference including removing sources of noise and shielding, however the focus here is on digital signal processing techniques to remove power line interference, without requiring modifications or additions to the data acquisition equipment. Traditionally, power line interference is removed from sEMG signals using a notch filter centered on the contaminated frequency components. A number of disadvantages exist with this approach. Firstly, the notch filter removes all frequencies within the stopband, which includes frequency components of the sEMG signal. Secondly, practical notch filters have finite roll-off on each side of the stopband, which causes distortion in those sEMG frequency components [39]. Lastly, there exists variance in the power line interference frequency; a notch filter could miss the power line interference entirely [40].

Adaptive algorithms can be used to filter a reference input which is correlated with the power line interference to produce an accurate estimate of the noise [41]. This reference input must either be a recorded signal, which is expensive in terms of memory, or a synthetic signal which presupposes a known power line frequency.

In this chapter, a least squares adaptive power line estimation algorithm is evaluated which estimates the unknown amplitude, phase, and frequency of the power line interference, with no need for a reference input. The noise estimate can be used to quantify the signal-to-noise ratio (SNR). The power line interference can subsequently be mitigated in the sEMG by subtracting this noise estimate. The algorithm is evaluated in terms of its ability to accurately estimate the power line interference at different frequencies in simulated sEMG with varying spectral content, SNR, and signal length. The results are then validated by running the algorithm on real sEMG contaminated with artificial power line interference.

## 4.2 Least Squares Adaptive Power Line Estimation Algorithm

An iterative least squares steepest descent algorithm adapted from [42] is used to estimate the parameters of the power line interference. A full derivation of the algorithm can be found in Appendix A. The power line interference,  $\hat{m}[n]$ , is assumed to consist of a single frequency component; that is, a sinusoid with an unknown amplitude  $\hat{A}$ , angular frequency  $\hat{\omega}$ , and phase  $\hat{\phi}$  as given in (4-1).

$$\hat{m}[n] = \hat{A} \cos(\hat{\omega} \cdot n + \hat{\phi}) \quad (4-1)$$

Such a sinusoid can be expressed as a linear combination of a cosine and sine function as:

$$\hat{m}[n] = \hat{a} \cos(\hat{\omega} \cdot n) + \hat{b} \sin(\hat{\omega} \cdot n) \quad (4-2)$$

where  $\hat{a}$  and  $\hat{b}$  are the amplitudes of the in-phase and quadrature components,  $\hat{A} = \sqrt{\hat{a}^2 + \hat{b}^2}$ , and  $\hat{\phi} = -\tan^{-1} \frac{\hat{b}}{\hat{a}}$ .

At iteration  $k$  of the least squares algorithm, the mean squared error function is defined as:

$$E_k = \frac{1}{N} \sum_{n=0}^{N-1} (x[n] - \hat{m}_k[n])^2 \quad (4-3)$$

In (4-3),  $x$  is the noisy sEMG signal,  $\hat{m}_k$  is the power line noise estimate, and  $N$  is the signal length. It can be shown that minimizing the mean squared error ( $E_k$ ) between  $x$  and  $\hat{m}_k$  is equivalent to minimizing the mean squared error between  $\hat{m}_k$  (the noise estimate) and the

actual noise, given that the clean sEMG signal (without power line interference) is uncorrelated with the noise. A proof of this statement can be found in Appendix A.

The mean squared error function in (4-3) is minimized using an iterative steepest descent algorithm in the frequency parameter space. The frequency is updated as specified by (4-4), (4-5), and (4-6).

$$\hat{\omega}_{k+1} = \hat{\omega}_k - 0.5\mu \frac{\partial E_k}{\partial \hat{\omega}_k} \quad (4-4)$$

$$\frac{\partial E_k}{\partial \hat{\omega}_k} = -2[x - Qh_{opt}]^T Qh_{opt} \quad (4-5)$$

$$h_{opt} = \begin{bmatrix} \hat{a}_{opt} \\ \hat{b}_{opt} \end{bmatrix} \quad (4-6)$$

In (4-5),  $h_{opt}$  can be determined using the linear least squares estimator since  $\hat{m}$  is linear in terms of  $\hat{a}$  and  $\hat{b}$  as given in (4-7). The parameter  $\mu$  is the learning rate.

$$h_{opt} = (Q^T Q)^{-1} Q^T x \quad (4-7)$$

$$Q = \begin{bmatrix} \cos(\hat{\omega}_k \cdot n) \\ \sin(\hat{\omega}_k \cdot n) \end{bmatrix}^T \quad (4-8)$$

A disadvantage of the iterative steepest descent approach is that it can converge on a local minimum. To mitigate this, we confine the search space by assuming that the frequency is in the range of 59.5 Hz to 60.5 Hz. The initial frequency for the search was determined by minimizing the mean squared error defined by (4-3) using the least squares estimator in terms of  $\hat{a}$  and  $\hat{b}$ . For each frequency in the desired range at intervals of 0.001 Hz, (4-7) was used to calculate  $h_{opt}$ ,  $\hat{m}_k$  was then calculated from (4-1) and  $E_k$  from (4-3). Once the mean squared error,  $E_k$ , is calculated for each frequency in the search space, the minimum can be located and that frequency is chosen as the starting point for the steepest descent algorithm.

The algorithm proceeds by iteratively stepping in the opposite direction of the gradient vector and terminates when the frequency step size is less than  $10^{-7}$  Hz, or if this condition is not met after 200 iterations. It was determined that if the algorithm does not converge after 200 iterations, it is unlikely that it ever will because it is actually diverging from the power line frequency. The learning rate,  $\mu$ , represents the fraction of the gradient to use as the step size for each iteration. It is normally chosen as a small number to ensure accurate convergence. The smaller the learning rate, the more accuracy can be achieved, however more iterations will be required. The learning rate was initialized to  $10^{-8}$ . If the frequency step changes direction, this indicates a local minimum was overshoot due to the step size being too large. To prevent continually overshooting the minimum, the learning rate was reduced by 10% whenever the frequency step changed direction to decrease the step size.

### ***4.3 sEMG Simulation and Insertion of Noise***

To evaluate the performance of the least squares adaptive power line estimation algorithm, sEMG data was simulated using the method described in Chapter 3. Power line interference was added to the simulated sEMG signal by generating a sinusoid with a given amplitude and frequency; the phase was randomized using a uniform distribution in the range  $[-\pi, \pi]$  radians. The amplitude of the sinusoid was calculated based on the desired SNR. For the simulations, SNR, power line frequency, signal length, and sEMG spectrum shape were varied to examine their impact on algorithm performance.

### ***4.4 Simulation Results***

The purpose of the simulation was to estimate the sinusoidal noise components in terms of their frequencies and amplitudes where amplitude is related to SNR. That is to say, the goal was to evaluate the accuracy of the least squares adaptive power line estimation algorithm's estimated frequency and estimated SNR with respect to each control variable in the

simulation, namely sEMG spectrum shape, power line frequency, SNR, and signal length. Refer to section 3.2 for the sEMG simulation method.

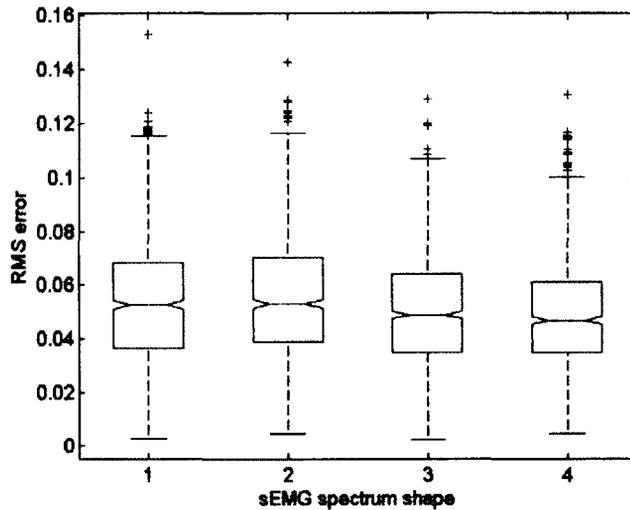
#### 4.4.1 sEMG Spectrum Shape

The effect of shaping parameters  $f_l$  and  $f_h$  was investigated for a constant signal length of 4096 (sampled at 1000 Hz), SNR of 10 dB, and power line frequency of 60 Hz. Simulations were replicated 1000 times. SEMG shapes were defined as indices for simplicity as given in Table 4-1.

**Table 4-1: sEMG spectrum shape index values and corresponding parameter values. These 4 parameter pairs were chosen as they were the same as used in [6].**

Index	Parameters ( $f_l, f_h$ )
1	(30,60)
2	(40,100)
3	(60,120)
4	(50,150)

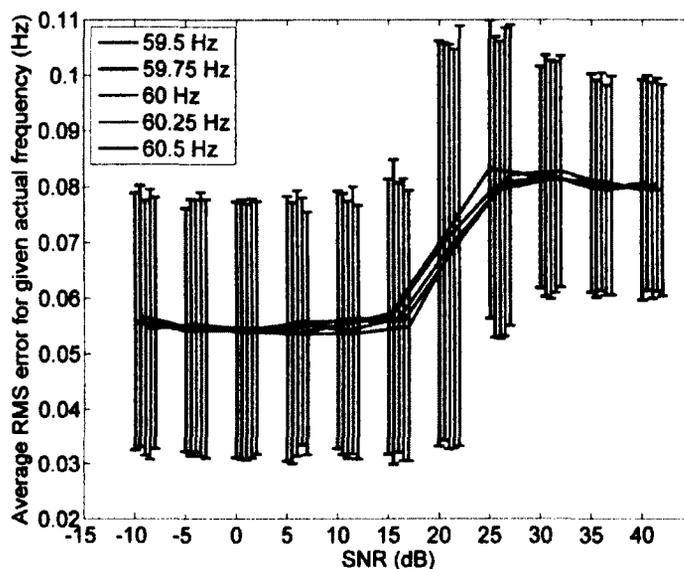
In Figure 4-1, a box plot of the root mean square (RMS) error between the original sEMG signal before noise was added, and the final noisy signal with the noise estimate subtracted is shown for each sEMG shape. Analysis of variance (ANOVA) resulted in a p-value very close to zero. This indicates that there is strong evidence that the null hypothesis (that each of the distributions for different sEMG spectrum shapes have the same mean) is false and can be rejected ( $p < 0.05$ ). That is to say, that sEMG spectrum shape does have a significant effect on RMS error. However, the largest difference between any two mean RMS errors was 0.0063 and the largest difference in medians was 0.0060. These small differences suggest that sEMG shape, although statistically significant, does not have a practical significance on the RMS error.



**Figure 4-1: Box plot of RMS error for 4 sEMG spectrum shapes between original clean signal and noisy signal with noise estimate subtracted.**

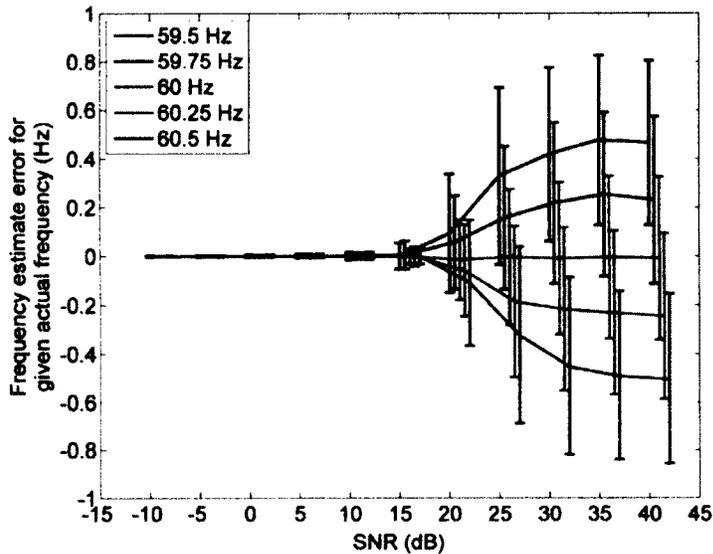
#### 4.4.2 Frequency and SNR

The average RMS error was calculated over all 1000 iterations for 5 different power line frequencies (59.5, 59.75, 60, 60.25, 60.5) Hz. These frequencies were chosen as they represent a range much larger than what would typically be observed in North America. A constant signal length of 4096 samples (sampled at 1000 Hz) was used, along with a constant sEMG shape of  $(f_l, f_h) = (30, 60)$ . It was observed that power line frequency had no effect on the RMS error. This can be seen in Figure 4-2 where the RMS error for each power line frequency is shown at each SNR. We can observe that each frequency follows the same trend, and the variances are all similar.



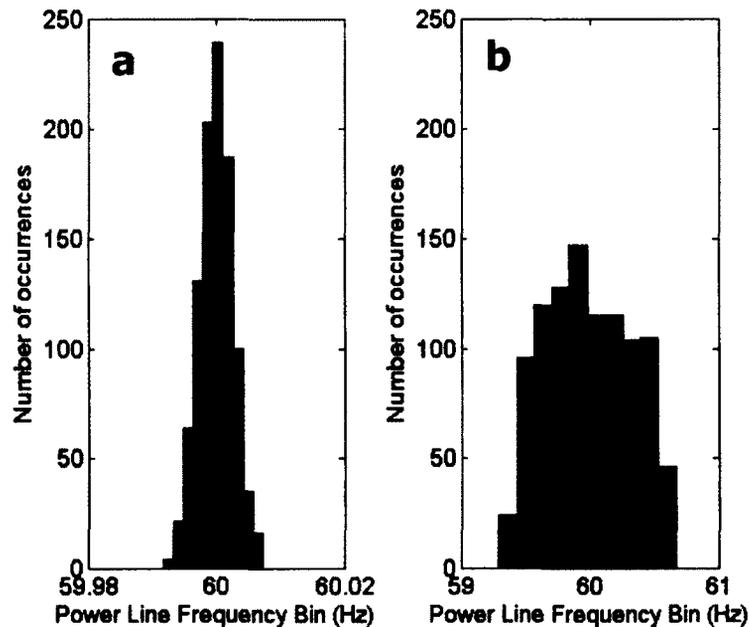
**Figure 4-2: Average RMS error for 5 different power line frequencies at various SNRs. Error bars are at plus or minus one standard deviation. Frequencies are offset on the SNR-axis such that error bars do not overlap.**

We notice that for large noise power (SNR between -10 dB and 15 dB), the RMS error is smaller. For SNR above 15 dB, the RMS error becomes larger, with the variance in the error also noticeably larger. For high SNR sEMG signals (greater than 15 dB), the power line interference is small and there may exist larger sinusoidal components in the sEMG signal that the least squares adaptive algorithm can mistakenly identify as power line interference. This can also be seen in Figure 4-3 as a plot of mean frequency error versus actual SNR.



**Figure 4-3: Mean frequency error for 5 different power line frequencies at various SNRs. Error bars are at plus or minus one standard deviation. Frequencies are offset on the SNR-axis such that error bars do not overlap.**

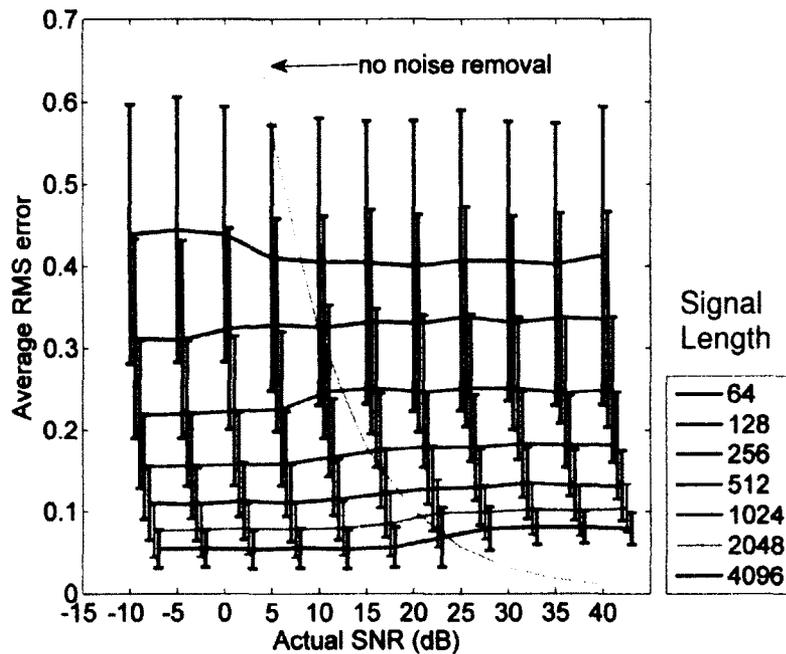
We see a divergence of errors for high SNR (past 15 dB) when the algorithm performed poorly. This divergence is explained by the constraining of the frequency parameter space to 59.5 Hz to 60.5 Hz. When the algorithm starts to perform badly, its estimate becomes essentially useless and will reflect the expected value of the frequency parameter space (in this case, the mean was 60 Hz). This produces an error of virtually zero when the actual frequency was 60 Hz, but produces larger errors as the actual frequency deviates from the mean (e.g., 59.5 Hz or 60.5 Hz). This is supported by Figure 4-4 which shows histograms for two SNRs run for 1000 iterations. At a low SNR, the distribution has a mean of 60 Hz (the actual power line frequency) and a small variance. For high SNR the variance is much larger and the distribution begins to resemble a uniform distribution. This indicates that for high SNR, the algorithm is tracking a random frequency component in the sEMG which resembles a sinusoid, and that these frequency components are essentially uniformly distributed within the parameter space.



**Figure 4-4: Distribution of power line frequency estimates for 60 Hz noise at (a) SNR=-5 dB and (b) SNR=40 dB. Note the x-axis scales are different.**

#### 4.4.3 Signal Length and SNR

The average RMS errors were calculated over 1000 iterations for each SNR and signal length  $N$  where  $N \in \{64,128,256,512,1024,2048,4096\}$  as shown in Figure 4-5. We observe that signal length has a large effect on the RMS error between the original clean signal and the final signal after the noise estimate has been removed. Increasing the SNR has a negligible impact on RMS error since, even though the noise estimate worsens, the magnitude of the noise is small enough that the difference has minimal impact on the overall RMS error. The dependency of RMS error on signal length can be explained by recognizing that shorter signals have much poorer temporal and frequency resolution than longer signals. This lack of resolution results in identifying sEMG frequency components together with noise components as interference and ultimately would cause a loss of useful information when the signal is cleaned.

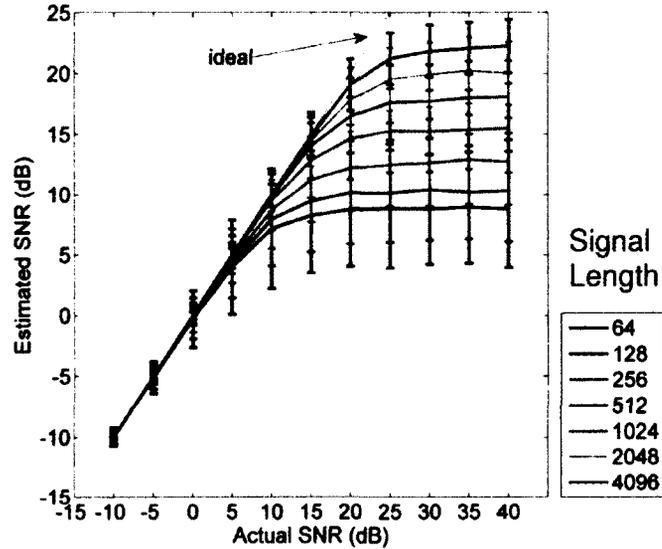


**Figure 4-5: Average RMS error for 7 different signal lengths at various SNRs. Error bars are at plus or minus one standard deviation. Signal lengths are offset on the SNR-axis such that error bars do not overlap. Signal sampled at 1000 Hz**

In Figure 4-5, a curve representing the RMS error between the noisy sEMG without noise cancellation, and the original clean sEMG is shown for reference. This curve effectively represents the square root of the noise signal power. When the average RMS error at a given signal length and SNR is larger than the square root of the noise power, this is an indication that noise cancellation results in a larger RMS error than if nothing was done and the original noisy signal used as is. It can be seen that at an SNR past 25 dB, it is better to not perform noise cancellation, regardless of signal length. However, for shorter signal lengths this limit shifts to a lower SNR.

Mean SNR estimates were calculated for each SNR and signal length as shown in Figure 4-6. Ideally, the plot should have a linear one-to-one relationship, however, we see the SNR estimate fall below the actual SNR for values at some point which is a function of the signal length, that is, the SNR is underestimated for higher actual SNR values. For values above

this threshold, subtraction of this noise estimate will result in a distortion of the sEMG frequency components. The longer the signal length, the higher this threshold SNR will be which is consistent with the dependency on signal length observed in Figure 4-5. Values range between 0 dB and 15 dB where 15 dB is the breaking point for signals of length 4096.



**Figure 4-6: Mean estimated SNR for each actual SNR for the noisy sEMG signal at various signal lengths. Ideal case is shown for reference. Error bars are at plus or minus one standard deviation. Signal sampled at 1000 Hz.**

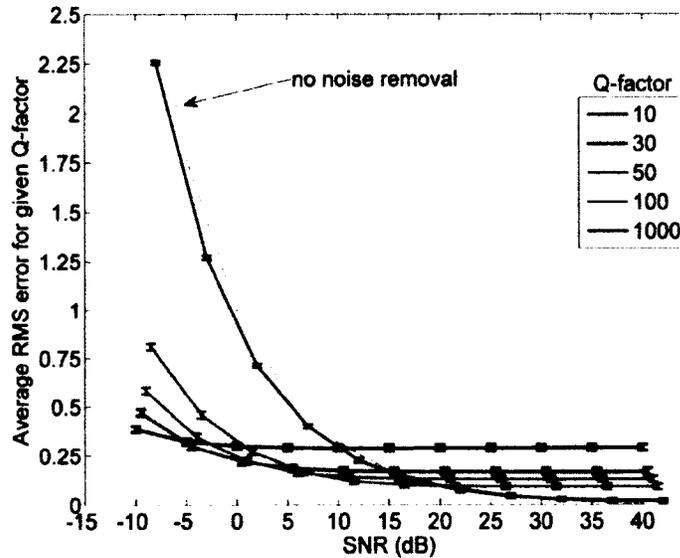
#### 4.4.4 Comparison with Notch Filtering

The same simulation for generating sEMG and power line interference was used for testing an infinite impulse response (IIR) notch filter with various bandwidths. The bandwidths were centered on 60 Hz and the same sampling rate of 1 kHz was used. The bandwidths ( $BW$ ) were varied by specifying the filter quality factor  $Q$  (or  $Q$ -factor).

$$Q = \frac{f_0}{BW} \quad (4-9)$$

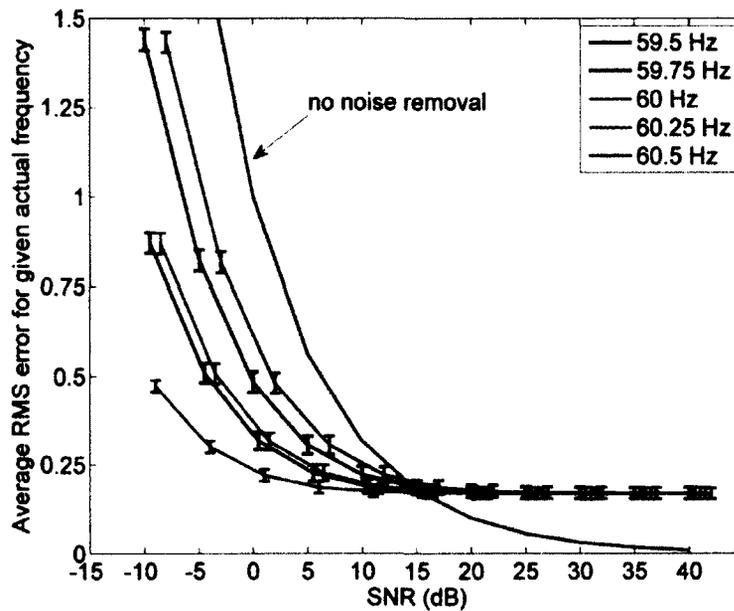
The RMS errors for the filtered sEMG signals are shown as functions of SNR at different  $Q$ -factors in Figure 4-7. Power line interference was generated at 60 Hz and all signal lengths were 4096 samples (sampled at 1000 Hz). It is shown that lower  $Q$ -factors produce lower RMS errors at lower SNRs. As SNR increases, higher  $Q$ -factors become slightly

advantageous. At around 10-15 dB, notch filtering becomes detrimental overall as the RMS error increases beyond that of no noise removal.



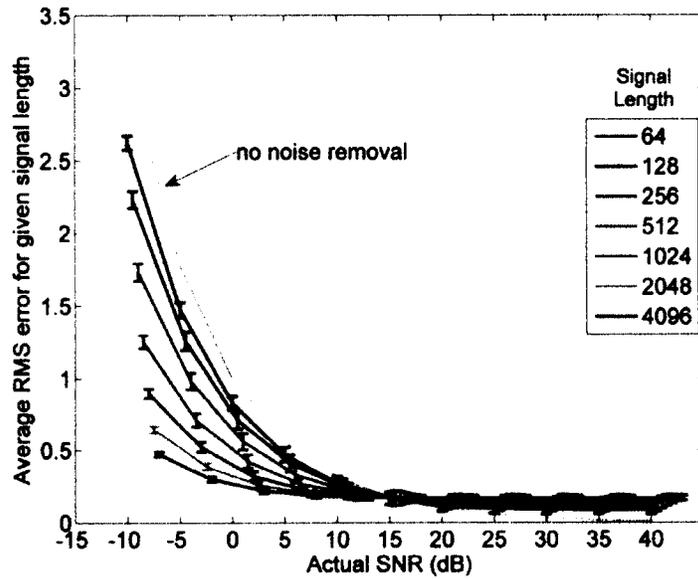
**Figure 4-7: RMS error for notch filtered sEMG at 60 Hz at different SNRs and Q-factors.**

From Figure 4-7, a Q-factor of about 30 is optimal over the range of SNRs at this frequency resolution. The effect of deviation in the power line frequency on the effectiveness of the notch filter can be seen from Figure 4-8. A Q-factor of 30 is used with a signal length of 4096 (sampled at 1000 Hz). The greater the deviation from the notch filter center frequency, the larger the RMS error. As expected, optimal performance is observed when the power line frequency is coincident with the notch filter center frequency.



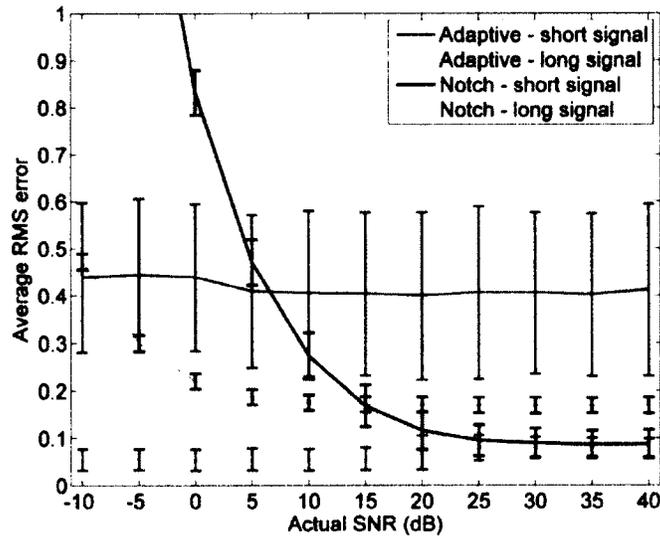
**Figure 4-8: RMS error for notch filtered sEMG at 60 Hz for different SNR and power line frequencies**

The signal length is varied in Figure 4-9 for a power line frequency of 60 Hz and notch filter Q-factor of 30. The notch filtering approach produces a larger RMS error for lower SNR and shorter signals. As SNR increases the RMS error decreases and eventually levels off. Notch filtering performance of high SNR signals is better for shorter signal lengths.



**Figure 4-9: RMS error for notch filtered sEMG (Q-factor = 30) at 60 Hz for different SNR and signal lengths**

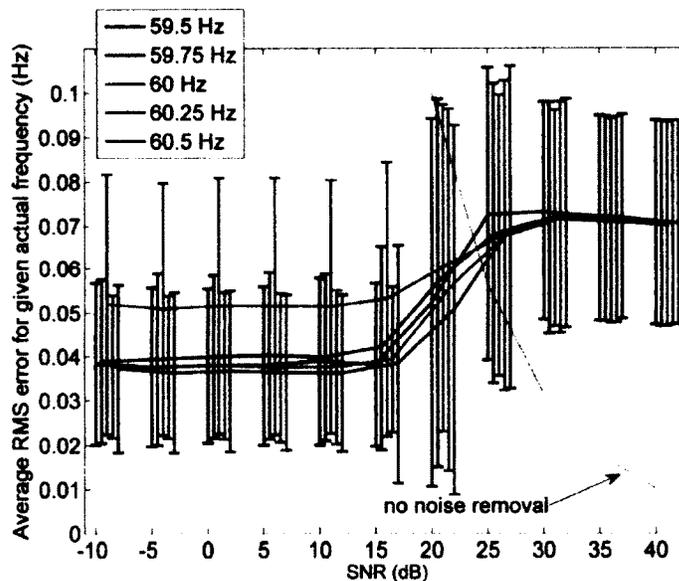
It is shown in Figure 4-10 that the adaptive algorithm results in a lower RMS error than notch filtering except for short signal lengths at high SNR. In other cases, it is advantageous to use the adaptive algorithm over notch filtering.



**Figure 4-10: Comparison of notch filtering (Q-factor = 30) and adaptive filtering for short signals (64 samples) and long signals (4096 samples at 1000 Hz) at various SNR**

## 4.5 Results with Real sEMG

Real sEMG (recorded as described in Chapter 3) was used to validate the results obtained with simulated data. Figures are not directly comparable due to different signal lengths (real sEMG was sampled at 3000.3 Hz for 10 seconds). The results for frequency estimate versus SNR were the same as Figure 4-3. Results for SNR estimate versus actual SNR showed a cut-off for the estimate around 15 to 20 dB in accordance with Figure 4-6 for long signals. Furthermore, we see the algorithm is applicable for SNR below approximately 20 to 25 dB as shown in Figure 4-11 in accordance with Figure 4-5.



**Figure 4-11: Average RMS error for 5 different power line frequencies at various SNRs for entire real sEMG dataset. Error bars are at plus or minus one standard deviation. Frequencies are offset on the SNR-axis such that error bars do not overlap.**

In Figure 4-11, we can observe larger variance and RMS error for 60 Hz interference at lower SNR. The explanation for this is that real sEMG will always have some true power line interference at 60 Hz unless all equipment is operating on battery power and is completely electrically shielded from outside interference. Since 60 Hz noise is present in the signal which is assumed to be clean, removal of this 60 Hz noise along with synthetic noise will

result in a larger RMS error even though no more clean sEMG is being sacrificed than for any of the other power line frequencies.

## **4.6 Discussion**

Power line frequency and sEMG spectrum shape were not significant factors for estimating power line interference. These results are not surprising as the sEMG itself is independent from the noise that is introduced. There is no reason that the two should influence each other. The adaptive power line estimation algorithm is designed to be invariant with respect to frequency within its expected range so any frequency chosen within this range should also have no impact on performance.

The adaptive power line estimation algorithm performed very well for lower SNR and higher signal lengths. As SNR increases (beyond 15 dB for a signal length of 4096 at 1000 Hz), the frequency estimate of the algorithm becomes inaccurate. This is because the power line interference is small and the least squares adaptive algorithm may be falsely tracking sinusoidal components that are part of the sEMG signal. This hypothesis is supported by examining the error in the estimate of the power line frequency, which also becomes inaccurate for SNR values above 15 dB. The range of power line frequency (i.e., [59.5 Hz, 60.5 Hz]) is larger than what would be anticipated in reality. Narrowing this range may help mitigate the error in power line frequency estimate.

Although SNR affected the accuracy of noise estimation, it was less significant when comparing the original signal before noise was introduced, and the final signal after the noise estimate was subtracted (i.e., the RMS error). Although the RMS error between the cleaned and original signals would increase in general with SNR, this increase was very small which is explained by the power line interference having small magnitude at high SNR and thus having little impact on the RMS error.

Signal length also was determined to affect the magnitude of the RMS error and accuracy of noise estimation. The explanation for this was that longer signals contained more sinusoidal periods of the power line interference contaminating the sEMG which allows for better estimation of the sinusoidal parameters. When the signal length is reduced, fewer periods of the sinusoidal power line interference exist and are still combined with sEMG so it is more difficult to extract information regarding to the sinusoid when a smaller amount of information is available to the algorithm.

Notch filtering results in a higher RMS error with an ideal quality factor, power line frequency and signal length. Changing these parameters will further worsen the performance of the notch filter. The only case where it was observed to perform better than the adaptive algorithm was for short signals at moderate to high SNR. It can be seen that notch filtering a short signal of high SNR is similar to (but slightly better than) performing no mitigation whatsoever. Since the signal length is small, the frequency resolution is poor. The 60 Hz noise is spread out over a range of frequencies and notch filtering only removes a small amount and leaves the rest unchanged. It is not worth applying a notch filter in this case due to the negligible improvement over doing nothing, even though the performance may be better than the adaptive algorithm (which should not be used in this circumstance either). For longer signals, it is always better to apply the adaptive algorithm over the notch filter, even when applying the adaptive algorithm is worse than doing nothing.

Whenever an attempt is made to mitigate the power line interference, there will likely be distortion of the sEMG signal. When the amount of interference is large (i.e., low SNR), the trade-off between reducing the effect of the interference and distorting the sEMG is justified. The trade-off may not be justified, or even required, for low levels of interference (i.e., high SNR). This algorithm provides an accurate estimate of the SNR up to 15 dB for a

signal length of 4096 when sampled at 1000 Hz. This SNR estimate can be used to decide whether or not any power line interference mitigation should be attempted or not. Using a shorter signal length will lower this SNR threshold and will increase the RMS error between the cleaned and original noiseless signal.

#### ***4.7 Conclusion***

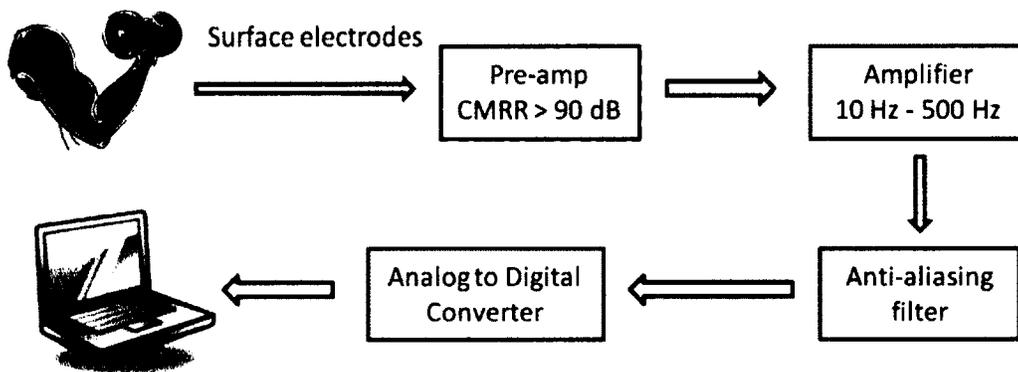
The application of the least squares adaptive power line estimation algorithm to quantify power line interference in sEMG produced reliable results for SNR below 15 dB for a signal length of 4096 when sampled at 1000 Hz. For sufficiently short sEMG signals such that the signal can be assumed to be stationary, this method facilitates the mitigation of power line interference with minimal distortion to the original signal and can do so in a timely manner. The method is not appropriate for signals with too few samples or signals with minimal power line interference. If the SNR is high (above 25 dB), it is unlikely that the power line interference will have much of an impact on the signal and filtering may not be necessary. Otherwise, an adaptive least squares method can significantly improve sEMG signal quality when contaminated with power line interference and results in less signal distortion than when applying a notch filter. While the power line interference estimation method was presented here in the context of sEMG, this method would be just as applicable to other data collection setups (e.g., electrocardiograms and electroencephalograms).

# 5 Detection of ADC Clipping, Quantization Noise, and Amplifier Saturation

## Amplifier Saturation

### 5.1 Introduction

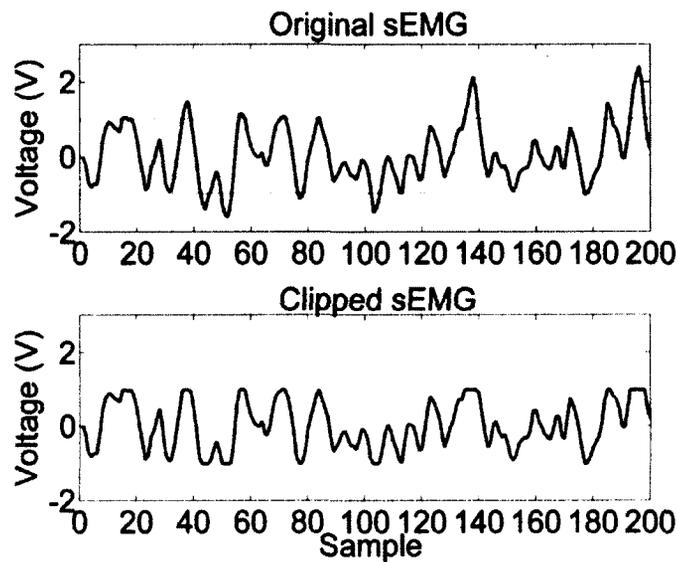
A typical sEMG collection setup is shown in Figure 5-1. In Chapter 4, the focus was on power line interference in the first stage of the collection setup. In this chapter, issues with the amplifier and analog-to-digital converter (ADC) stages are examined; specifically, ADC clipping, quantization noise, and amplifier saturation. When acquiring sEMG, the expected signal amplitude needs to be considered when setting the gain for the amplifier and range for the ADC. This process may be trivial for a trained sEMG technician, but can be a significant source of error for others, and since the overarching goal of this thesis is to enable non-expert users to incorporate sEMG analysis into clinical workflows, automating this process would be a valuable contribution.



**Figure 5-1: Typical sEMG collection setup**

An ADC is a device that will take a continuous input signal, sample it at discrete time intervals, and then quantize the signal amplitudes to discrete values. If the input dynamic range of the ADC is too small, *ADC clipping* will occur when the input signal exceeds this range. An ADC that has saturated will clip the output to either its maximum or minimum

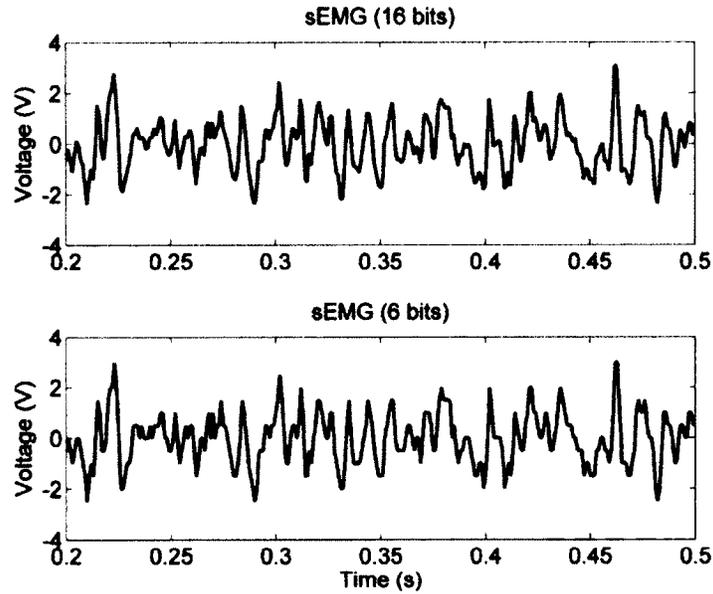
value (Figure 5-2). This is easily detected if the dynamic range of the ADC is known; however, this information may not be readily available. The motivation behind this work is to enable sEMG acquisition without the requirement for trained bioengineers or experts otherwise familiar with the acquisition equipment and setup. Furthermore, in certain cases, sEMG acquisition and analysis are performed separately (and by different people). The acquisition information may not be readily available to all parties.



**Figure 5-2: sEMG clipped to  $\pm 1$  V, sampled at 1 kHz**

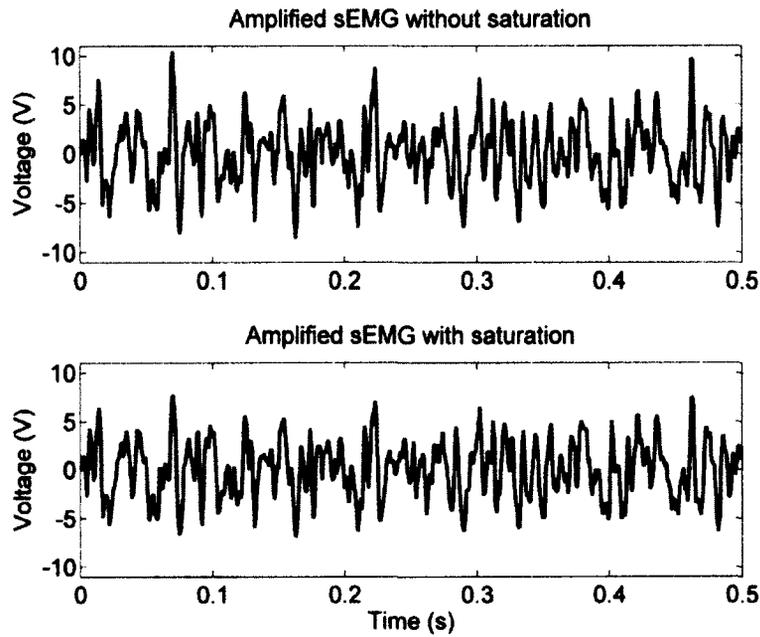
Conversely, if the ADC dynamic range is too large with respect to the signal amplitude, the *quantization noise* in the signal can become significant. Quantization noise is a result of rounding the analog amplitude value to a discrete value (note that truncation may be used instead, however, only rounding associated with mid-tread quantizers will be considered here). Quantization noise will always be present in a sampled signal [43]. The amount of quantization noise can be expressed as a signal-to-quantization-noise ratio (SQNR). The SQNR of a signal increases by 6 dB per additional bit of resolution [44]. An  $n$ -bit ADC that is not utilizing its full range (i.e. when the ADC input range is set too high) will behave like a  $k$ -bit ADC where  $k < n$ . As such, knowing the bit resolution of the ADC is insufficient information to determine the effective SQNR. The step-size of the ADC can be used to

determine the effective SQNR, however this information may not be readily available, as was the case for the ADC range. Quantization noise can result in a signal that is less smoothed and more step-like (Figure 5-3).



**Figure 5-3: SEMG signal quantized to 16-bits and 6 bits showing the distortion effect of quantization noise**

*Amplifier saturation* occurs when the signal amplitude, after applying the amplifier gain, is outside the operating range of the amplifier, which is limited by its electrical supply. The amplifier will no longer provide a linear response when saturated, resulting in signal distortion (Figure 5-4).



**Figure 5-4: Amplified sEMG signal with and without amplifier saturation where the amplifier maximum is at 10 V**

In this chapter, a set of methods are proposed and evaluated for detecting and quantifying ADC clipping, quantization noise, and amplifier saturation in sEMG signals. Performance of these methods is evaluated using simulated sEMG [31] with random shaping filter parameter pairs (Table 5-1).

**Table 5-1: Shaping parameter pairs for sEMG simulation**

$f_l$ (Hz)	30	40	60	50
$f_h$ (Hz)	60	100	120	150

Performance is then validated using real sEMG collected using the protocol in Chapter 3.

## **5.2 Detection of ADC Clipping**

### **5.2.1 Method**

The detection of clipping is trivial if the range of the ADC is known. However, if the range is not known, clipping can be detected by searching for two consecutive maximum or

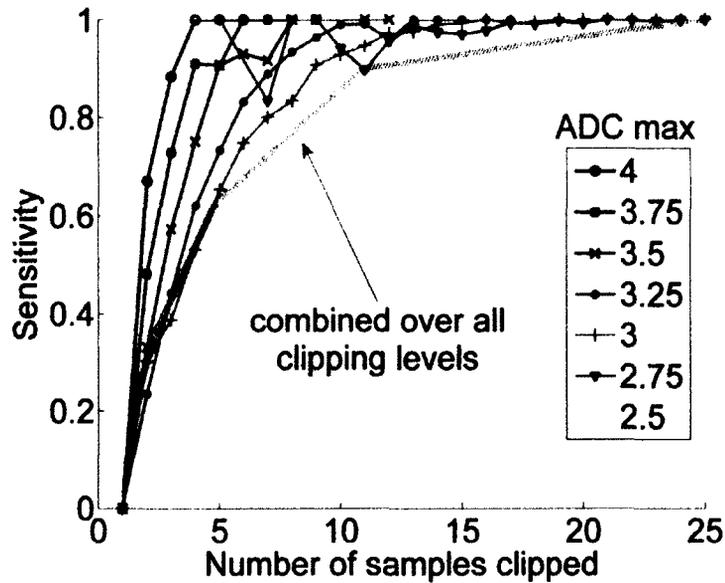
minimum values in the signal. This is accomplished by examining the signal in two passes. The first pass is to determine both the maximum and minimum values in the signal. The second pass is then performed to identify all occurrences where either of these values appears consecutively, two or more times. Due to random noise and signal variability, observing two consecutive minimum or maximum values is an unlikely occurrence in acquired biological signals unless the ADC has truly saturated. An ADC will quantize an analog signal and, when the ADC output saturates, clip the signal at ADC's maximum or minimum value. This is shown in (5-1) where  $s(t)$  is the analog signal,  $s[n]$  is the digitized signal,  $T$  is the sampling period,  $\Delta$  is the ADC step size, and  $A_{min}$  and  $A_{max}$  are the ADC minimum and maximum values, respectively.

$$s[n] = \begin{cases} \text{Around}\left(\frac{s(nT)}{\Delta}\right), & A_{min} \leq s(nT) \leq A_{max} \\ A_{min}, & s(nT) < A_{min} \\ A_{max}, & s(nT) > A_{max} \end{cases} \quad (5-1)$$

### 5.2.2 Simulated sEMG Results

To evaluate the proposed ADC clipping detection method, simulated sEMG (normalized to a power of 1) with one of four random shaping parameter pairs as specified in Table 5-1, were manually quantized to a fixed-point 16 bits at seven different voltage ranges (4, 3.75, 3.5, 3.25, 3, 2.75, and 2.5) leading to increasing degrees of signal clipping. The signals were then clipped at those voltage levels using (5-1). For each level of clipping, 10,000 iterations were performed. In each iteration, a quantized sEMG signal was simulated (4096 samples in length and sampled at 1 kHz). If the signal was clipped, the number of samples clipped was recorded. If clipping did not occur then the signal was skipped. If a clipping event was detected in the signal, the iteration was considered a true positive (TP); if no clipping was detected, the iteration was considered a false negative (FN). Note that a TP does not necessitate that all clipped samples in the signal were detected. A sensitivity (TP/(TP+FN))

can then be derived for the method. If clipping was detected in the original signal (before manual clipping was introduced), the iteration was considered a false positive (FP).

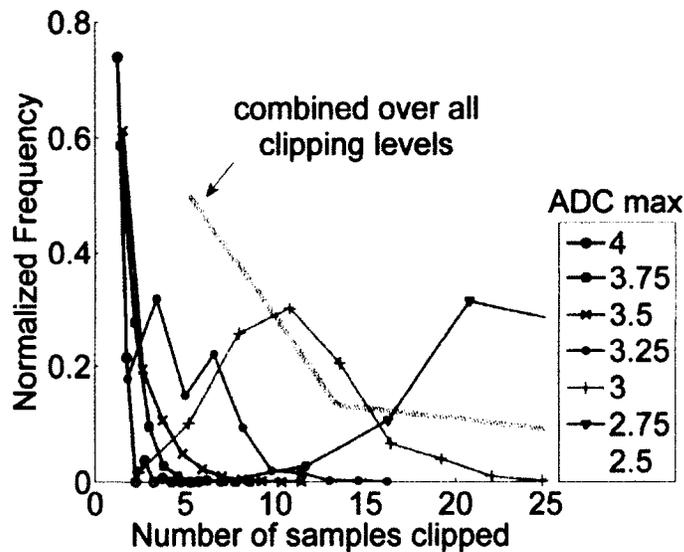


**Figure 5-5: Clipping detection sensitivity as a function of the number of clipped samples in the signal (plot for ADC max of 2.5 not shown as sensitivity was always 1).**

Figure 5-5 is a plot of the sensitivity of the clipping detection algorithm as an increasing number of samples in the sEMG were clipped. As the ADC range is lowered, it is more likely that a larger number of samples will be clipped which means the plot of the sensitivity curve for that ADC range will be pulled further to the right. Generally speaking, the lower the ADC range, the more clipping should occur and the easier this clipping will be to detect. However, occurrences where only a few samples were clipped at a small ADC range are actually harder to detect than if the ADC range was higher since these occurrences are very seldom, and when they do occur, they are unlikely to occur as consecutive minima or maxima. This is why sensitivity is actually lower for a lower ADC range when the number of samples clipped is also low. As expected, the sensitivity for each range tends to increase with the number of samples clipped and reaches 1 after about 20 samples clipped out of 4096. Note that 5 false positives occurred (out of 10,000 iterations) as a result of detecting

clipping on an sEMG signal before the clipping had been simulated. This results in a specificity of 99.95%.

In Figure 5-6, the distribution of the number of samples clipped for each clipping level is shown. The distributions will be approximately Gaussian; however for lower levels of clipping, the distribution is skewed as the number of samples clipped cannot be less than 1. The means of the distribution increase with the level of clipping, confirming that higher levels of clipping will result in more samples clipped on average.



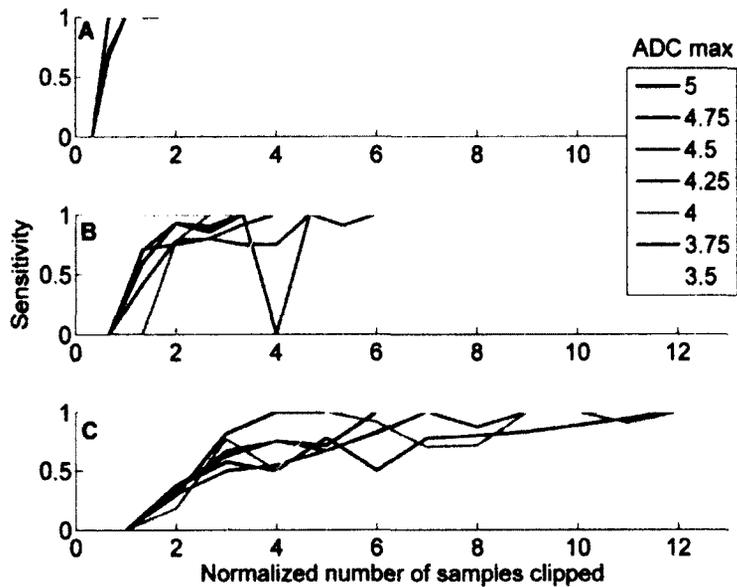
**Figure 5-6: Histogram for number of samples clipped over each clipping level. Note that the distribution for ADC max of 2.5 and part of that for 2.75 is cut off.**

The method was also tested using real data obtained from a microphone connected to a computer, and using real sEMG in section 5.2.3. Different signal amplitudes were recorded over a range of frequencies by speaking and whistling into the microphone, as well as playing music through speakers and then feeding the music back into the microphone. Clipping was correctly identified in the majority of instances. Some signals whose amplitudes were near the threshold of the ADC would only clip at one or two (non-consecutive) samples. This would go undetected using this method; however it also

represents a very small loss of information so the false negative is both not surprising and not critical. It is conceivable to have a signal which repeatedly clips at only a single sample over and over but never manifests multiple consecutive clipped samples. Such a signal would also not be flagged as clipped by this method, however it was determined that producing such a signal is difficult (unless done artificially) and likely not to occur in practice. False positive identifications were seldom observed, however they could occur for very low amplitude signals where the bit resolution was poor, thus making it more likely to observe consecutive minimum or maximums due to the smaller number of possible values available to the ADC at that range. Such an occurrence would also be detected as a source of quantization noise and, when corrected, would no longer be a problem. Quantization noise is discussed in section 5.3.

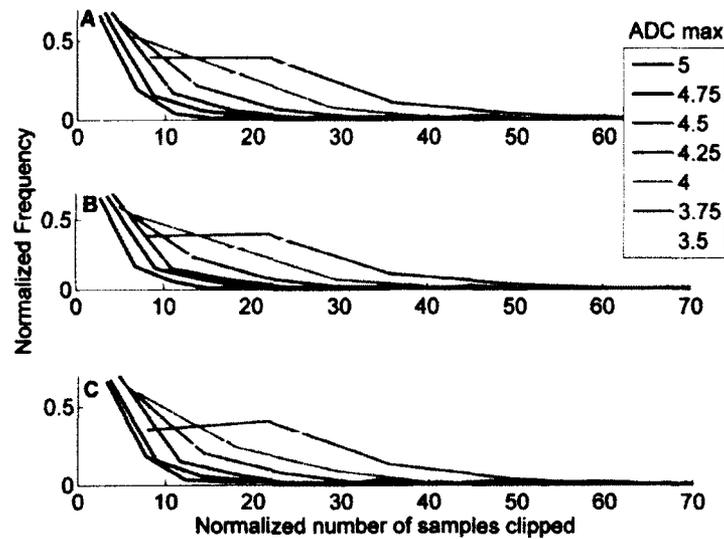
### **5.2.3 Validation of ADC Clipping Detection with Real sEMG**

Real sEMG was normalized and manually clipped at seven voltage levels (5, 4.75, 4.5, 4.25, 4, 3.75, and 3.5). Each signal was then processed using the same clipping detection method as for simulated sEMG. The real sEMG was sampled at 3000.3 Hz for 10 seconds. Two additional iterations were performed using the real data downsampled to 1500.15 Hz, and 1000.1 Hz by keeping every second and third sample, respectively, to investigate the effect of sampling rate on accuracy. No low-pass filtering was performed in conjunction with downsampling. Results are shown in Figure 5-7 and Figure 5-8.



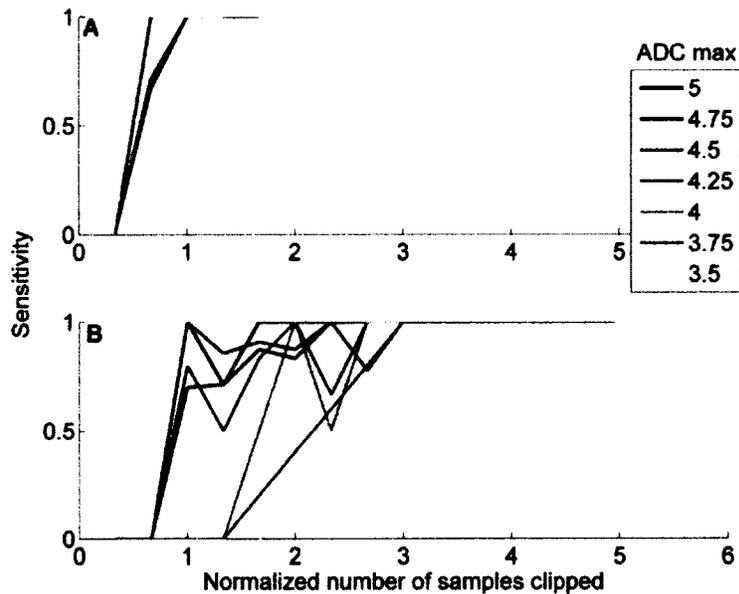
**Figure 5-7: Clipping detection sensitivity as a function of the number of clipped samples in the signal for real sEMG at a sampling rate of (a) 3000.3 Hz, (b) 1500.15 Hz, and (c) 1000.1 Hz. Number of samples are normalized by dividing by the signal length and multiplying by 10000.**

We can observe that as the sampling rate is decreased, the sensitivity for a fixed number of samples clipped appears to decrease while the distribution of normalized samples clipped remains constant (Figure 5-8). This is expected as using a high sampling rate will increase the probability of observing consecutive maxima or minima. However, this also increases the chances of a false positive. The total number of false positives that occurred out of the 167 signals before clipping was induced was 11, 5, and 1, for sampling rates of 3000.3 Hz, 1500.15 Hz, and 1000.1 Hz, respectively.



**Figure 5-8: Histograms for number of samples clipped over each clipping level for real sEMG at a sampling rate of (a) 3000.3 Hz, (b) 1500.15 Hz, and (c) 1000.1 Hz. Number of samples are normalized by dividing by the signal length and multiplying by 10000.**

We can mitigate the 11 false positives that occur at 3000.3 Hz by adjusting our method based on the sampling rate. For higher sampling rates (e.g., 2000 Hz and above) we can search for 3 consecutive extrema as opposed to 2. These results are compared to the original (using 2 consecutive extrema) in Figure 5-9. In the 3 consecutive extrema case, the false positive rate was reduced to 0, however it took a slightly larger number of clipped samples before the detection sensitivity reached 1.



**Figure 5-9: Clipping detection sensitivity as a function of the number of clipped samples in the signal for real sEMG (3000.3 Hz) by searching for (a) two consecutive extrema and (b) three consecutive extrema. Number of samples are normalized by dividing by the signal length and multiplying by 10000.**

### **5.3 Detection of Quantization Noise**

The signal-quantization-noise ratio (SQNR) is a useful indicator of the amount of quantization noise present in the signal. Such an indicator can be used to determine if the level of quantization noise can compromise the signal quality in the current setup.

The maximum quantization error for a uniform mid-tread quantizer is half the quantization step size ( $\Delta$ ). The quantization error can be considered additive and the signal can be represented as in (5-2), where  $x[n]$  is the quantized signal,  $s[n]$  is the signal before quantization, and  $q[n]$  is the quantization noise (round off error).

$$x[n] = s[n] + q[n] \quad (5-2)$$

We can define the SQNR as in (5-3), where  $P_s$  and  $P_q$  are the signal power of  $s[n]$  and  $q[n]$ , respectively.

$$SQNR = 10 \log \left( \frac{P_s}{P_q} \right) \quad (5-3)$$

We can make a reasonable assumption that the quantization error is uniformly distributed between  $-\Delta/2$  and  $\Delta/2$ . Thus,  $P_q$  is equal to the variance of this uniform distribution [45]:

$$P_q = \frac{\Delta^2}{12} \quad (5-4)$$

The signal prior to quantization ( $s[n]$ ) is unknown but its power can be estimated from the recorded signal  $x[n]$ . Since  $s[n]$  and  $q[n]$  are uncorrelated and  $q[n]$  is zero mean,

$$P_s = P_x - P_q \quad (5-5)$$

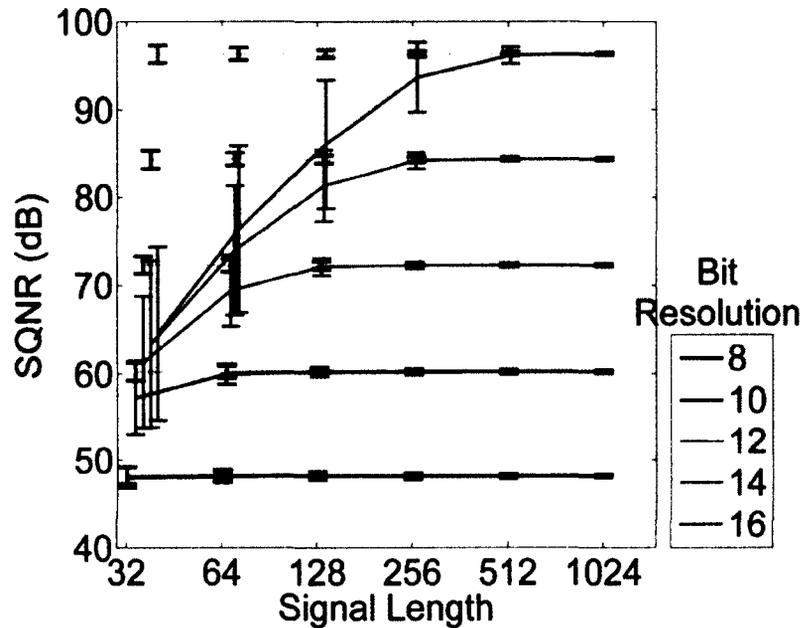
$$SQNR = 10 \log \left( \frac{P_x}{P_q} - 1 \right) \quad (5-6)$$

$$SQNR = 10 \log \left( 12 \frac{P_x}{\Delta^2} - 1 \right) \quad (5-7)$$

If the step size  $\Delta$  is not known, it can be estimated from the signal by searching for the smallest absolute difference between any two values in the signal, provided that the signal is long enough to yield a high probability of observing two values that are one step size apart.

Accuracy of estimating SQNR using (5-7) and the estimated step size were investigated by generating signals consisting of random numbers with predefined bit resolutions. Bit resolutions from 8 to 16 bits were used by varying the step size and appropriately rounding off the sEMG values. Signal length was also varied from 32 to 1024 samples, increasing by factors of 2. The SQNR was estimated from the quantized signal (computed using (5-7)) and compared to the actual SQNR (computed using (5-3)). Each bit resolution and signal length combination was evaluated on 1000 iterations.

The SQNR estimates were averaged over all iterations for each signal length and bit resolution and are shown in Figure 5-10 along with the actual SQNR. It can be seen that longer signal lengths will produce more accurate SQNR estimates. The higher the bit resolution, the longer the signal length should be in order to accurately estimate the SQNR.



**Figure 5-10: Estimated average SQNR at different bit resolutions and signal lengths with one standard deviation error bars. Averaged actual SQNR are shown as dotted lines. Signals sampled at 1000 Hz.**

From Figure 5-10, it is shown that a signal length greater than 256 is more than sufficient to produce an accurate estimate at 12-bit resolution which is a typical resolution for an ADC for sEMG acquisition. This simulation assumed that all values within the signal range are equally likely to occur. In sEMG signals, the time-domain voltage is approximately normal [31] so values close to the baseline are more likely to occur than peak values. However, the purpose of this simulation was to show how the estimated SQNR will approach the actual SQNR given by the bit resolution in the simulation. The actual minimum and maximum values of the signal are not important. What is important is that given certain minimum and

maximum values, the SQNR can be estimated by ultimately estimating the signal step size, and the accuracy of this step size estimate is determined by signal length.

## 5.4 Detection of Amplifier Saturation

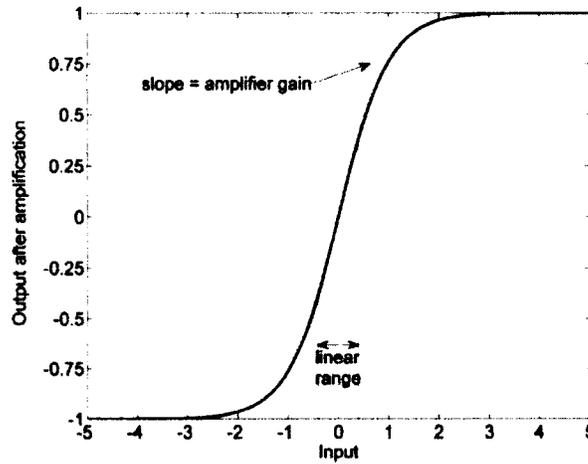
Real amplifiers are not ideal and will have both linear and nonlinear ranges of operation. As the expected output approaches the amplifier rails, the amplifier tends to provide a diminishing gain factor. This nonlinearity introduces contamination in the sEMG signal. The approximate shape of this curve is shown in Figure 5-11; it is a sigmoid function scaled to the desired amplifier range and shifted vertically to yield an odd function about the origin. The equation for the curve is given in (5-8) where  $s_{input}(t)$  is the original signal,  $s_{output}(t)$  is the signal after amplification,  $G$  is the gain factor, and  $Amp_{max}$  is the amplifier maximum value (minimum value is  $-Amp_{max}$ ).

$$s_{output}(t) = Amp_{max} \left( \frac{2}{1 + e^{\frac{-2G}{Amp_{max}} s_{input}(t)}} - 1 \right) \quad (5-8)$$

### 5.4.1 Method

The amplitude of an sEMG signal can be modeled as a Gaussian process [31]. The histogram of a given sEMG signal with a given mean and variance should exhibit a high correlation with a normal distribution with equal mean and variance. A similar normality test was evaluated in [46]. As the level of amplifier saturation increases, the value of the Pearson correlation coefficient [47] between the sEMG histogram and normal probability density function should decrease. This measure of correlation is a potential method of automatically detecting amplifier saturation in sEMG. The level of amplifier saturation was measured by calculating the fraction of samples in the nonlinear range. The nonlinear range was defined to be when the percent difference between the actual and expected amplified values was greater than 5%. This occurs when the condition in (5-9) is met.

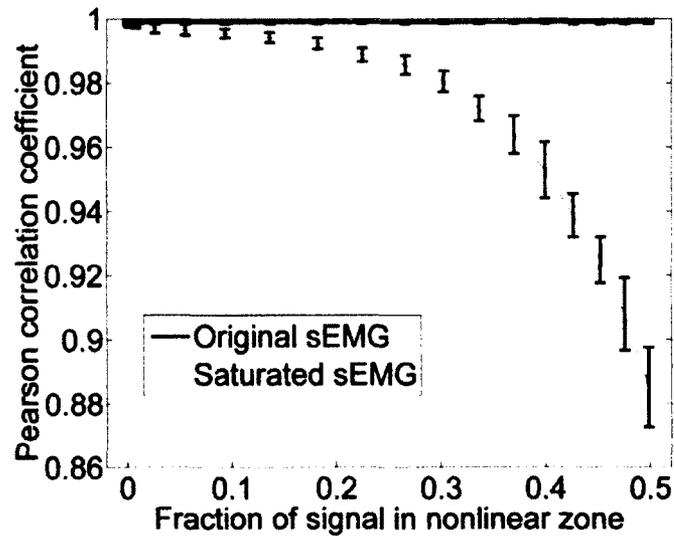
$$\frac{|s_{input}(t)|G}{Amp_{max}} > 0.4 \quad (5-9)$$



**Figure 5-11: Transfer function of a typical amplifier with linear and non-linear ranges of operation and a gain factor of 1. Linear range corresponds to an input between -0.4 and 0.4.**

#### 5.4.2 Simulated sEMG Results

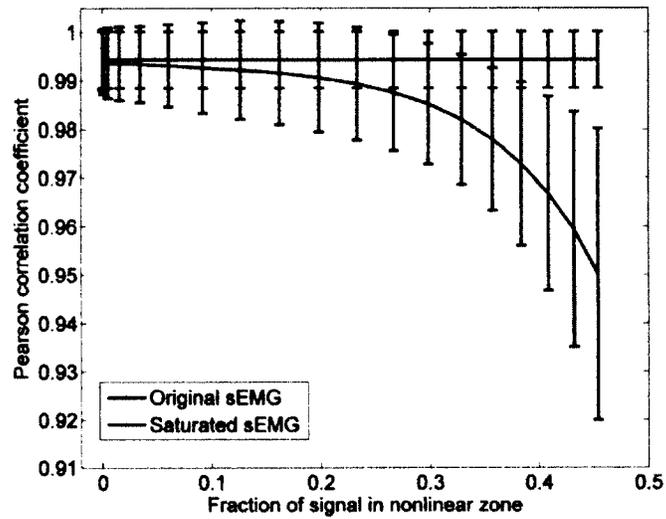
Simulated sEMG with a length of 4096 samples (sampled at 1000 Hz) were scaled to have an initial signal power of 0.1. Using (5-8) to simulate the predicted amplifier output, the gain  $G$  was varied from 1 to 20, with  $Amp_{max}=10$ . An amplifier maximum of 10 V is the case for our lab and gain values were chosen to include a range from little to no amplifier saturation to severe amplifier saturation. These gain values were applied in software and will differ from values used on a real amplifier in hardware. Each simulation was repeated over 1000 iterations and the average Pearson correlation coefficients between the sEMG amplitude histogram and normal probability density function were computed for both the original signal and each amplified signal. The trend is shown in Figure 5-12. A decrease in correlation can be observed as the fraction of the signal amplitude in the nonlinear zone increases.



**Figure 5-12: Pearson correlation coefficient calculated when different fractions of the signal samples were in the nonlinear zone of the amplifier range for simulated sEMG.**

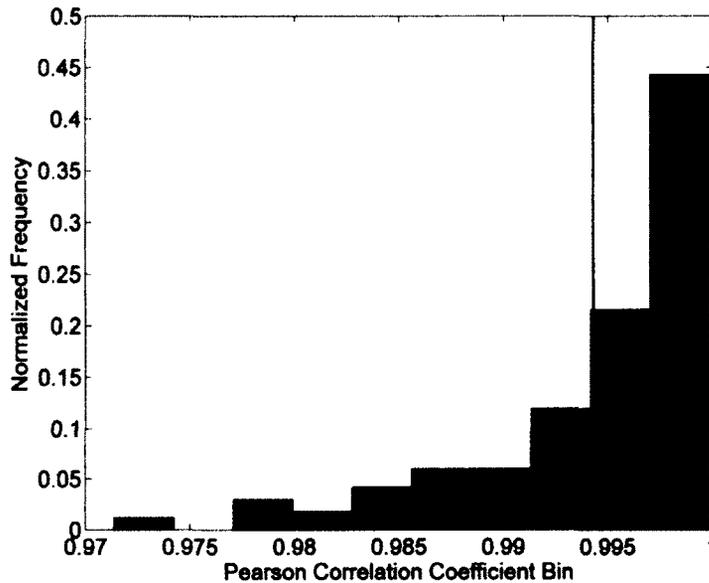
### 5.4.3 Validation of Amplifier Saturation Detection with Real sEMG

Real sEMG were also used here to ensure the hypothesis of normality was reasonable in practice (i.e., real sEMG without significant amplifier saturation should pass the normality test). Simulating amplifier saturation in the same method using real sEMG after normalization to a signal power of 0.1 yields the results in Figure 5-13. The plot is similar to Figure 5-12. The decrease in correlation coefficient is more gradual. Variance is larger for real sEMG than simulated sEMG under controlled conditions which is expected.



**Figure 5-13: Pearson correlation coefficient calculated when different fractions of the signal samples were in the nonlinear zone of the amplifier range for real sEMG. Error bars are at  $\pm 1$  standard deviation.**

In Figure 5-14 we see a histogram of correlation coefficients for the real sEMG without amplifier saturation with mean 0.9943 and standard deviation 0.0058. It can be seen that the majority of the distribution falls very close to 1, indicating a strong correlation with the normal probability density function.



**Figure 5-14: Correlation coefficient distribution for real sEMG. Mean is indicated by the red vertical line.**

## **5.5 Discussion**

As ADC clipping becomes more significant, the sensitivity increases quickly, reaching 90% with 10 of the 4096 samples clipped (Figure 5-5). Such a small amount of clipping would be difficult to observe visually, yet can be detected with the proposed method. ADCs operate with fixed ranges so it is more likely that clipping will be either significant or nonexistent as opposed to being present but very minimal. In the case that only a few samples are clipped and this goes undetected, the resulting signal distortion could be considered insignificant.

SQNR can be reliably estimated given a signal of sufficient length. For lower bit resolutions, a shorter signal length is sufficient for an accurate SQNR estimate. The smallest difference between any two values will become a reasonable estimate of the true ADC step size as the signal length increases. A larger pool of samples will result in a greater probability that the difference between two of them will represent the true step size. As sEMG is usually sampled at a minimum of 1000 Hz and often for at least a few seconds, signal lengths of

4096 and greater are typical which, from Figure 5-10, would translate into a highly accurate estimate of the SQNR and a reliable indicator of quantization noise.

The assumption that sEMG amplitude is approximately normal is reasonable for both simulated and real sEMG. The normality decreases as the degree of amplifier saturation increases (Figure 5-12). It is possible to determine a threshold value for the Pearson correlation coefficient to distinguish clean signals from those affected by amplifier saturation. Other factors can influence the normality of the signal (e.g., ADC clipping). Also, the large variance for real data when there is a large amount of saturation present makes choosing a correlation coefficient threshold more difficult (i.e., this method could result in more false positives for real sEMG data). As such, while this test can detect amplifier saturation, additional tests would be required to definitively identify the source of the sEMG contaminant. In addition, motion artifact or baseline wander can increase the kurtosis of the distribution and, although it may still be strongly correlated with the normal probability density function, the variance in the correlation may increase and warrant a less restrictive threshold. Real sEMG signals will always have a small amount of motion artifact and will produce distributions that are more peaked (i.e., exhibit higher kurtosis). As these signals become saturated, the tails of their distributions become truncated which usually results in a decrease in kurtosis. Since the initial kurtosis of the real sEMG was above normal, the correlation to the normal distribution will actually increase with a small amount of amplifier saturation (this can be seen in Figure 5-13 but it is extremely slight). As the saturation becomes more significant, the correlation will decrease and will follow a similar trend to that of simulated sEMG as shown in Figure 5-12. More investigation is required to determine a robust correlation coefficient threshold, but our current results suggest that the normality of the sEMG amplitude can be used as an indicator for amplifier saturation.

## **5.6 Conclusion**

Methods of detecting ADC clipping, quantization noise, and amplifier saturation in sEMG signals were investigated. ADC clipping can be detected in sEMG by searching for consecutive minimum or maximum values. Quantization noise can be measured by estimating the SQNR of the signal by determining the smallest step size. Amplifier saturation can be detected by examining the correlation between the sEMG amplitude histogram and the Gaussian probability density function. More investigation is required to determine the robustness of this method, however, especially with real signals and hardware. These methods can be employed automatically in software in any sEMG acquisition setup.

## **6 Detection and Mitigation of Electrocardiogram Artifact and Quantification of Motion Artifact**

### ***6.1 Introduction***

When acquiring sEMG signals from the abdomen, back, or chest, significant electrocardiogram (ECG) artifact can contaminate the sEMG signal [26]. This is of particular concern for esophageal and diaphragmatic sEMG, as well as recording of any muscle in the torso (e.g., myoelectrically controlled prostheses with targeted muscle reinnervation [48]). The high signal power of the QRS complex from the ECG can greatly increase the power of the recorded signal and alter extracted features such as mean/median frequency [49]. Although a variety of methods exist to mitigate this interference, there is currently no universally accepted method for mitigation of ECG artifact [50]. Alternatives include spike clipping [26], gating [49], high-pass filtering [26], adaptive filtering [51], and template subtraction [49].

Spike clipping and high-pass filtering were used in [26]; however, only sEMG between QRS complexes were used to compute an average rectified value for performance evaluation. This presupposes that the P and T waves are not significant sources of interference. Also, the average rectified value may not be the best performance indicator for all applications. Spike clipping clamps the QRS complex at a predetermined (or adaptive) threshold and results in replacing the complex with a wave that is almost square and discontinuous which compromises the signal coincident with each clamped QRS complex. Furthermore, if the P and T wave amplitude are significant but below the clipping threshold, this contamination will remain. High-pass filtering can remove most ECG contamination but may also remove low frequency sEMG components due to their overlapping spectra [50].

Gating is similar to clipping. Segments of the signal that exceed a threshold (either positively or negatively) are clamped, but to zero, instead of the threshold value. This results in a loss of information. If the heart rate is particularly high, a large portion of the sEMG can be removed along with the ECG [49].

Adaptive filtering has been successfully applied to ECG artifact mitigation in [51]; however, it requires a separate reference channel which may not be suitable for all applications. This additional reference channel increases system complexity and reduces usability.

Template subtraction involves creating an ECG beat morphology template based on portions of the signal consisting of little sEMG, aligning this template to the original signal through cross-correlation, and then subtracting it from the original [49]. This method requires that a silent period (i.e., a period free from the sEMG signal content) exist within the signal (such as expiration for recording from the diaphragm), or requires a parameter optimization procedure to fit an averaged waveform to each ECG beat [52], which can be computationally expensive.

In this chapter, a new subtraction method is presented to mitigate ECG artifacts in sEMG recordings. This method does not use a beat-averaged template, but rather an estimate of the ECG waveform derived from a moving average of the ECG contaminated sEMG recording. As such, this method does not require a silent period from which to obtain an ECG template. Performance of the moving average method is evaluated on real sEMG artificially contaminated with ECG and is compared to the template subtraction method [49]. The template subtraction method is chosen for comparison as it has been shown to offer good, reliable performance when compared with other methods [50]. The use of a

moving average is later applied to quantifying motion artifact which is similar in frequency content to the P and T waves of the ECG.

## 6.2 Moving Average Method

The moving average method for ECG mitigation uses two moving averages to estimate the ECG signal, which can then be subtracted from the contaminated sEMG recording. A moving average  $y$  can be calculated by the linear convolution of the original signal  $x$ , and the impulse response of the moving average filter  $h$  where  $N$  is the signal length:

$$y[n] = \sum_{m=0}^{N-1} h[m]x[n-m] \quad (6-1)$$

In this work a rectangular window is used. Other windows were investigated (Hamming and Hann) but the rectangular window was found to produce the best performance, empirically. The filter impulse function is therefore:

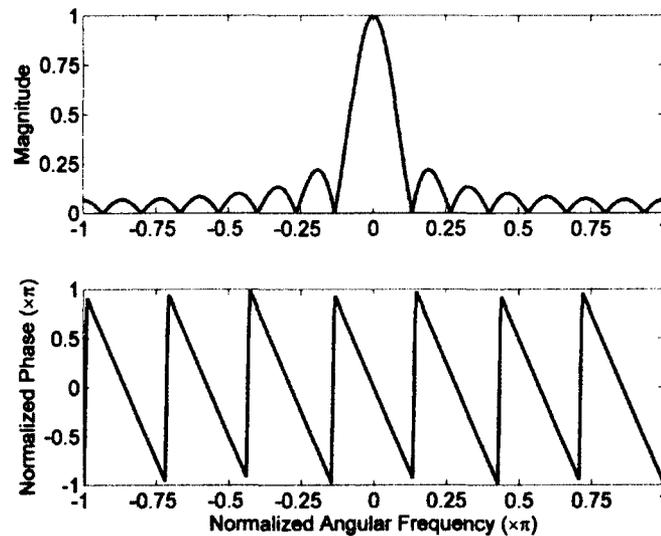
$$h[n] = \frac{1}{L}, \quad -\frac{L-1}{2} \leq n \leq \frac{L-1}{2} \quad (6-2)$$

where  $L$  is the length of the filter, which must be an odd valued integer. This filter impulse function is centered around  $n=0$ , so the filter is non-causal; however, this is not an issue for offline data processing of pre-recorded data and real-time operation is still possible with appropriate data buffering.

The frequency response of this filter is given by (6-3). A full derivation can be found in Appendix B.

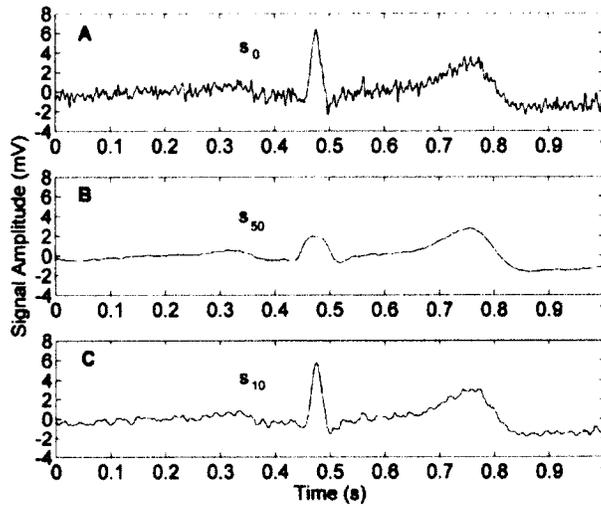
$$H(e^{j\omega}) = \begin{cases} 1, & \omega = 2\pi k, k \in \mathbb{Z} \\ e^{-j\omega \frac{L-1}{2}} \left( \frac{\sin(\frac{\omega L}{2})}{\sin(\frac{\omega}{2})} \right), & \omega \neq 2\pi k, k \in \mathbb{Z} \end{cases} \quad (6-3)$$

Figure 6-1 shows a plot of the frequency response for a window length of 15. It is shown that the moving average filter acts as a type of low-pass filter and has nulls at  $\pm 2\pi k/L$  for any nonzero integer  $k$ .



**Figure 6-1: Magnitude and phase of the frequency response for a moving average filter with  $L=15$ .**

Two moving averages are combined to obtain an accurate estimate of the ECG morphology; one to estimate the low frequency parts of the waveform (P and T wave) and another to estimate the high frequency parts (QRS complex). A larger window size smoothes out more of the sEMG but may also smooth out the ECG waveform. An appropriate window size to obtain the shape for the P and T waves was determined empirically to be 50 ms. Due to its higher frequency components, the QRS complex is attenuated using such a large window size (Figure 6-2b). A smaller window size of 10 ms is used to detect the QRS complex even though significant high-frequency sEMG will remain (Figure 6-2c). As the moving average filter can be thought of as a low-pass filter, the window length is directly related to the frequency cutoff. A longer window length will offer a greater smoothing effect on the signal and translates to a lower frequency cutoff (lower bandwidth low-pass filter). This is needed to estimate the P and T waves as they consist of low frequencies. The QRS complex contains higher frequency components and requires less aggressive low-pass filtering.



**Figure 6-2: (a) SEMG signal contaminated with ECG artifact (SNR = -8 dB). The contaminated sEMG filtered using a moving average window size of (b) 50 ms and (c) 10 ms.**

Let the moving average computed with a 50 ms window be denoted  $s_{50}[n]$ , and the moving average computed with a 10 ms window to represent the QRS complexes be denoted  $s_{10}[n]$ . The estimate of the ECG waveform is  $s_{50}[n]$ , except during the QRS complexes where it is  $s_{10}[n]$ . If we let  $T$  be the set of sample indices identified as part of a QRS complex, the ECG estimate can be expressed as:

$$\hat{s}_{ECG}[n] = \begin{cases} s_{10}[n], & n \in T \\ s_{50}[n], & n \notin T \end{cases} \quad (6-4)$$

QRS detection must be performed to identify the set of samples  $T$ . QRS detection was performed on  $s_{50}[n]$ , which was first filtered in the forward and backward direction, ensuring a zero-phase response, using a 2<sup>nd</sup> order Butterworth low-pass filter with a cutoff frequency of 15 Hz. A derivative threshold method was then applied to locate the QRS complexes. The Q, R, and S waves were then labelled by finding where the derivative was zero. A minimum R-R interval (300 ms, corresponding to a maximum pulse rate of 200 beats per minute) was imposed to reduce false positive QRS detection.

The ECG estimate  $\hat{s}_{ECG}[n]$  is subtracted from the original contaminated signal  $s_o[n]$  and the result  $\hat{s}_{EMG}[n]$  is an estimate of the sEMG. The moving average method is summarized in Figure 6-3.

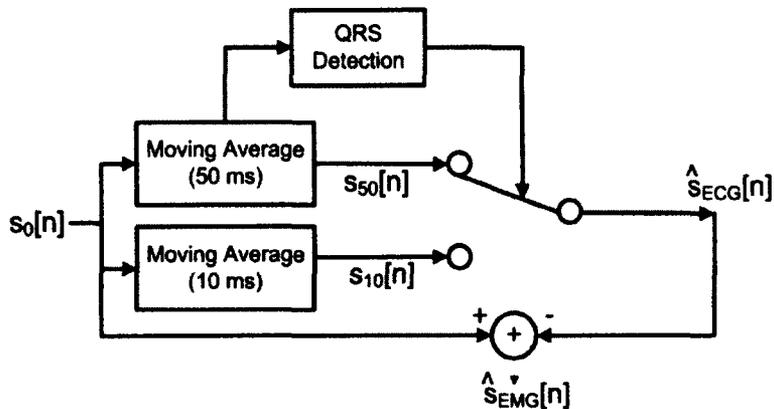


Figure 6-3: Block diagram of the moving average method.

### 6.3 Template Subtraction Method

The template subtraction method [49] requires a reference signal of the contaminating ECG, which can be obtained by recording while the muscle is relaxed. QRS detection is performed on the sEMG-free signal (i.e., ECG-only signal) using a thresholding technique. This is a simpler technique than that used for the moving average method. In this case, the QRS complexes are identified from an ECG signal free of sEMG and the Q and S wave locations are unimportant. The first and last beats are discarded to ensure no partial beats. The remaining beats are segmented (divided at the midpoints of the R-R intervals) and aligned to each other using cross-correlation. Zero-padding is used at the start and end of shorter beat segments to ensure consistent signal lengths, while keeping the beat segments centered on their R waves. A subtraction template is formed by averaging the beat segments.

Cross-correlation is performed between the contaminated signal and the subtraction template. In this work, PQRST complexes were identified at locations where the cross-

correlation value was above the 99<sup>th</sup> percentile of the cross-correlation function computed over the entire signal. At these points, the aligned template was subtracted from the contaminated signal. The resulting signal was the sEMG with ECG artifact removed.

## **6.4 Data Acquisition**

This research was approved by the Carleton University Research Ethics Board. ECG and sEMG were recorded from one male subject with no known cardiovascular or neuromuscular disorders. An amplifier system with programmable gain and bandwidth (Model 15A54, Grass Telefactor, West Warwick, RI, USA) and a 12-bit analog-to-digital converter (PCI-6071E, National Instruments, Austin, TX, USA) were used to record the signals, which were processed offline using Matlab. SEMG and ECG were sampled at 3000.3 Hz. Each recording was 10 seconds in duration.

ECG recordings were made using Ag/AgCl electrodes (Blue Sensor T, Ambu, Ballerup, Denmark) with four standard configurations: lead I, lead II, V4 and V5. These leads were chosen to ensure a positive QRS deflection. The skin was cleaned with alcohol before applying electrodes. Signals were band-pass filtered in hardware from 0.3 Hz to 100 Hz, with a gain of 2000. Two separate recordings were performed for a total of eight ECG waveforms. The first recording was used as a contaminant, and the second served as the ECG-only reference signal required for template subtraction. The subject was relaxed and breathed normally to ensure minimal sEMG contamination.

SEMG were collected from the *biceps brachii* of the right arm using Ag/AgCl electrodes (MVAP-II, Newbury Park, CA, USA) in bipolar configuration with an inter-electrode distance of 20 mm. Electrodes were placed according to SENIAM guidelines [11]. Skin preparation involved cleaning with alcohol, abrading the skin with Nuprep gel, and applying EC2 electrode cream to the electrodes. Impedance between the electrodes, measured by the

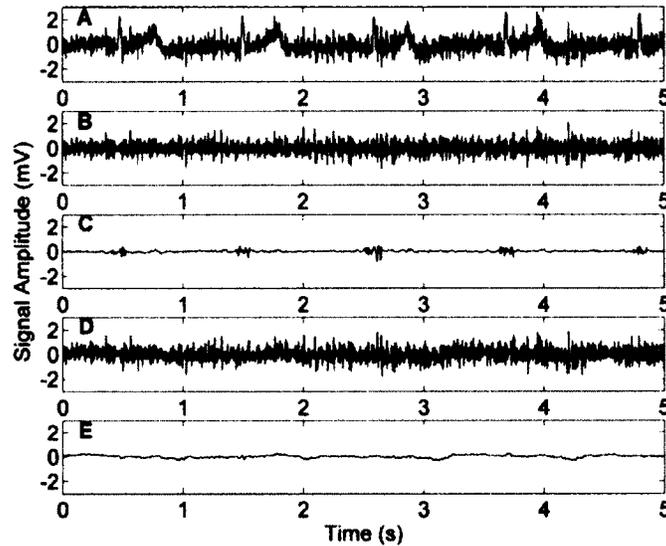
amplifier system between recordings, was 8-12 k $\Omega$  throughout the recording process. Ten sEMG recordings were made consisting of five isometric, isotonic contractions, and five rest trials when no muscle contraction occurred. Biceps contractions were made holding a 15 lb load with the forearm parallel to the floor and upper arm perpendicular to the floor. SEMG recordings were band-pass filtered from 0.3 Hz and 1000 Hz, with a gain of 2000.

## **6.5 Comparison of Methods**

Each of the five sEMG signals was added to each of the four ECG contaminant signals yielding a total of 20 sEMG/ECG combinations. SEMG was normalized to a signal power of 1 and ECG was scaled to obtain an SNR between -8 dB and 8 dB in 2 dB increments. A wide SNR range was used as the degree of ECG contamination can vary substantially in practice [26], [53]. Each of the ECG-only signals was paired with a corresponding ECG contaminant signal. Both signals were scaled using the same factor to preserve their magnitudes relative to one another. SEMG and rest recordings were also paired and scaled using the same factors.

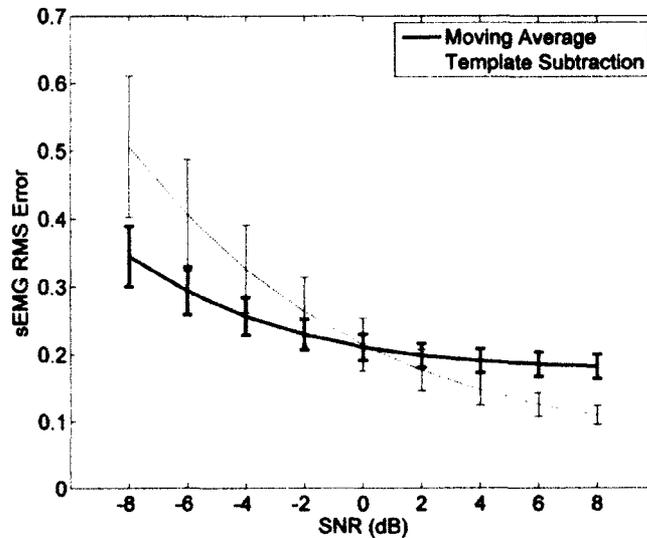
For each signal combination, both the moving average method and the template subtraction method were applied to mitigate the ECG contamination. For the moving average method,  $s_{50}[n]$  and  $s_{10}[n]$  were computed with moving average filter lengths of  $L=151$  and  $L=31$ , respectively. The root mean square (RMS) error between the cleaned signal and original sEMG signal was computed.

A sample contaminated sEMG signal is shown in Figure 6-4 at a signal to noise ratio of 2 dB along with the filtered signal using the moving average method and the error between the filtered and original sEMG signals.



**Figure 6-4: (a) sEMG contaminated with ECG (SNR = 2 dB). (b) sEMG filtered with moving average method. (c) sEMG estimate error with moving average (RMS error = 0.1883). (d) sEMG filtered with template subtraction method. (e) sEMG estimate error with template subtraction (RMS error = 0.2229).**

QRS detection was successful for both methods except for one signal combination where one QRS complex was missed for  $\text{SNR} \geq 6 \text{ dB}$  for the moving average method. The result of missing the QRS complex was that the moving average estimate  $s_{50}[n]$  was used instead of  $s_{10}[n]$ . The missed QRS complex could be detected by a more robust QRS detection algorithm; however, the consequence of the missed QRS complex was minimal given that the SNR was already high. The RMS error, as a function of SNR, is shown in Figure 6-5 for the moving average and template subtraction methods.



**Figure 6-5: Average RMS error between filtered and original clean sEMG for ECG artifact mitigation with the moving average and template subtraction methods.**

## 6.6 Discussion

The moving average method was found to perform better than the template subtraction method at lower SNR ( $< 0$  dB) in terms of RMS error. The RMS error of both methods decreased with increasing SNR; however this decrease was more gradual for moving average. This is explained by ECG artifact dominating the signal at low SNR. Template subtraction works by forming an average template over a silent period and small differences between each PQRST waveform are significant at low SNR. This translates into a poor estimate with the average template and produces a large RMS error. With a moving average, the ECG shape can still be obtained with relative accuracy.

At higher SNR, the ECG artifact may not be visible. The artifact will become clearer when moving averages are taken; however, these averages can still be dominated by the high signal power from the sEMG and the performance of the moving average method will saturate. The template subtraction method does not suffer from this drawback as it uses an sEMG-free signal to form the template. This allows for accurate ECG estimates even when

ECG power is relatively low. Furthermore, small differences in the ECG template will be negligible at high SNR. As such, the template subtraction method outperforms the moving average method at high SNR (when the signal power of the sEMG is dominant).

The moving average method has the advantage of being less sensitive to SNR and does not require an EMG-free segment in the recording. It also offers an advantage in terms of execution delay and automation. Both methods are comparable in terms of execution time; however, the template subtraction requires an entire signal consisting of the contaminated and silent (ECG only) portions. The more heart beats present, the better the subtraction template estimate. It is also required that the silent portion be manually identified by a technician which prevents real-time execution, although it is possible to potentially automate this process as well. Conversely, the performance of the moving average method does not depend on the number of heart beats. It can be applied automatically as ECG is acquired and filtered in real-time with no requirement for user interaction or any additional recordings.

## ***6.7 Quantification of Motion Artifact***

### **6.7.1 Introduction**

Motion artifact occurs when motion of the electrode cables, electrode leads, or stretching of the skin disturbs the half-cell potential of the electrode-electrolyte interface [20]. This artifact will appear as low frequency components in the signal, typically less than 20 Hz [3]. Motion artifact is a serious concern for dynamic muscle contractions where motion of electrode and cable are unavoidable. High pass filtering can be performed in hardware to eliminate this interference; however useful sEMG will be sacrificed due to overlapping spectra, as was the case for ECG artifact.

## 6.7.2 Data

Motion artifact was artificially added to simulated sEMG and real sEMG. The motion artifact was obtained from [54], [55] which consists of motion artifact, typical in ambulatory ECG recordings, recorded by placing electrodes on the limbs such that the ECG was not visible. The motion artifact was sampled at 360 Hz. SNR was varied from -40 dB to 40 dB in 5 dB increments. Simulated sEMG (refer to section 3.2) was sampled at 1080 Hz and the motion artifact was upsampled to the same sampling rate (using low-pass interpolation). Real sEMG was sampled at 3000.3 Hz and, in that case, the motion artifact was first upsampled to 9000 Hz and then downsampled to 3000 Hz. For each signal, a random window of motion artifact (with the same length as the sEMG signal) was selected from the pre-recorded sample in [54], [55] to contaminate the sEMG.

## 6.7.3 Methodology

Three methods were compared for quantifying motion artifact. The first method is an implementation of Sinderby's method for computing a signal-to-motion artifact ratio (SMR) [3]. Sinderby's method assumes that the power spectrum of a clean sEMG signal is roughly linear between 0 and 20 Hz and that respiratory-induced motion artifact will manifest itself as an increase in low frequency sEMG components in this range. Sinderby's method first uses a sliding window on the power spectrum between 35 Hz and 600 Hz (or the maximum frequency) to find the largest mean power density. A straight line is then drawn from the zero frequency to this largest mean power density. The signal-to-motion artifact ratio (SMR) is computed as the ratio of total signal power to the signal power below 20 Hz that exceeds that straight line. A number of issues exist with this approach. First, the method uses a basic FFT to compute the power spectrum. This results in high frequency resolution but very high variance in power between adjacent frequencies. This will produce a largest mean power density which is lower than what would be expected when the shape of the spectrum is examined visually. Second, motion artifact often has a frequency component at the zero

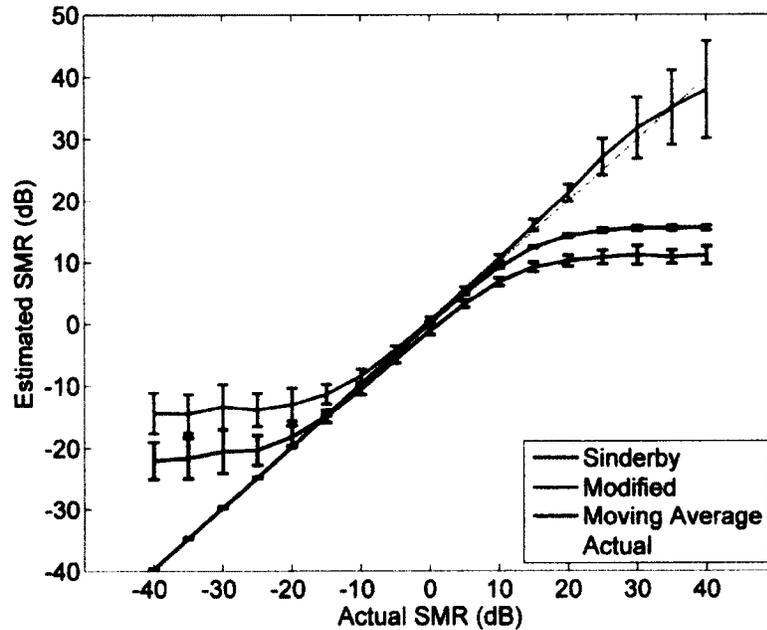
frequency (not necessarily DC due to limited frequency resolution). Drawing a line from the observed signal power at zero frequency may underestimate the motion artifact if the signal power at this point is not close to zero. In fact, it is unclear whether the method was actually intended to draw the line from the origin or from the true signal power at the zero frequency. Third, the method does not properly compute an SMR as would be done intuitively. If noise below 20 Hz and above the line is considered motion artifact, one would assume that clean signal power should be everything else not including this portion of the spectrum. Sinderby's method uses the entire signal power including motion artifact and then divides by the estimated motion artifact power. The result is a ratio that is always strictly greater than 1 or greater than 0 dB. For the purposes of this comparison, Sinderby's method is implemented by drawing the straight line from the origin, and the SMR is computed as the clean signal power (excluding the power above the line) divided by the motion artifact (the power above the line) as the results are more accurate.

Sinderby's method was modified slightly and used as the second method to detect motion artifact. Welch's averaged modified periodogram with 8 Hamming windows and 50% overlap was used to compute the power spectrum. The maximum signal power between 10 Hz and 20 Hz was then determined along with its corresponding frequency. A straight line from the origin to this maximum power point was drawn and power below 20 Hz and above this line was considered noise while the remainder of the spectrum was considered to be clean sEMG.

The third method used a moving average with 50 ms window length as was used for ECG artifact detection. The P and T wave portions of the ECG waveform are similar in frequency content to motion artifact. The resulting signal from the moving average was used as the ECG artifact estimate with which the SMR could be computed.

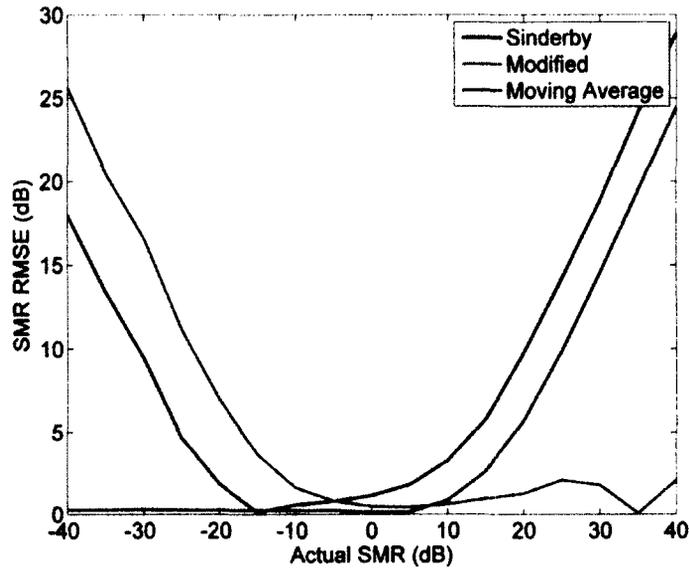
Each of the three methods was applied to quantify motion artifact. SMR was computed over all iterations and then averaged. Results for simulated data are shown in Figure 6-6.

#### 6.7.4 Results



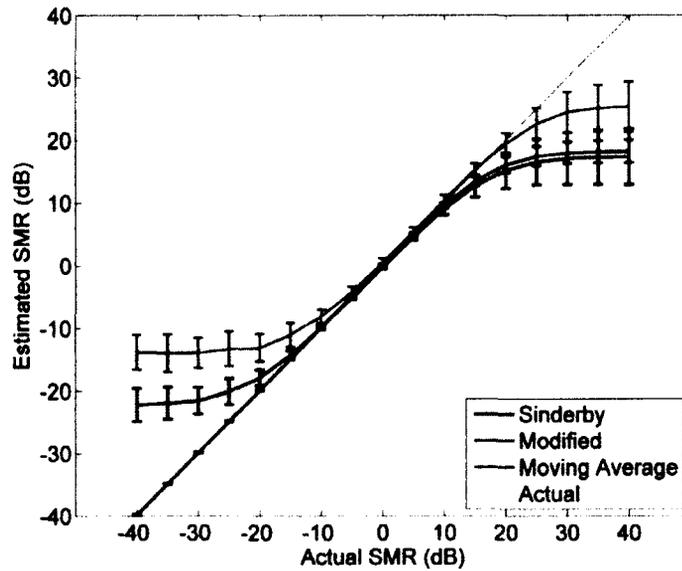
**Figure 6-6: Estimated vs. actual SMR for simulated sEMG signals using three methods of quantification. Error bars are at plus or minus one standard deviation.**

For simulated sEMG, the modified version of Sinderby's algorithm is superior than the original above -5 dB. However, the moving average is superior to both methods for SNR of approximately 8 dB and below, above which the modified Sinderby algorithm is better. This can be seen more accurately by examining the SNR estimate errors in Figure 6-7. No method is accurate over the entire range of SNR, but a combination of moving average and modified Sinderby's method could be used to expand the range of accuracy beyond that of any of the lone methods.



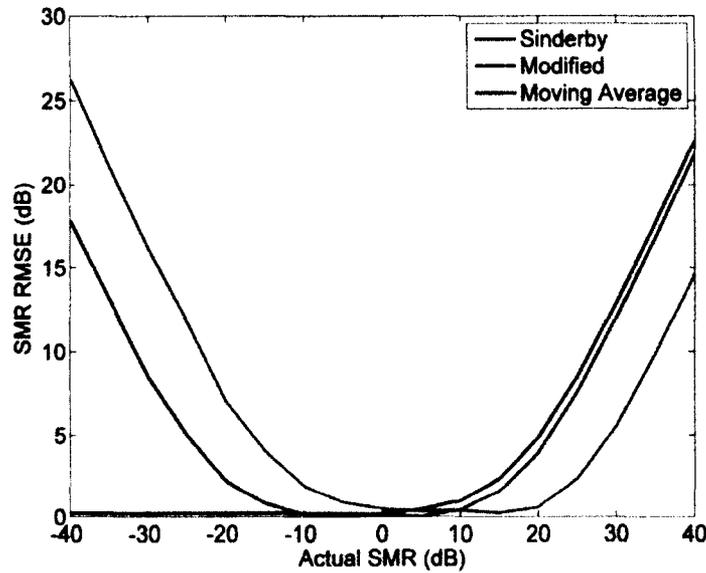
**Figure 6-7: RMS error signals for simulated sEMG SMR using three methods of quantification.**

Results for real sEMG are similar (Figure 6-8), albeit less accurate for the modified Sinderby method at higher SNR. This is likely due to the presence of real motion artifact in the signal, causing the SMR to be underestimated.



**Figure 6-8: Estimated vs. actual SMR for real sEMG signals using three methods of quantification. Error bars are at plus or minus one standard deviation.**

In this case, Sinderby's method is superior to the modified method below approximately 5 dB (instead of -5 dB for the simulated case). The moving average is superior to both methods below 10 dB, above which the modified Sinderby method is superior. However, even the best method is only accurate to 20 dB (Figure 6-9).



**Figure 6-9: RMS error signals for real sEMG SMR using three methods of quantification.**

These results indicate that a moving average is the best method to quantify motion artifact below 10 dB. If desired, this accuracy can be extended to 20 dB by using the modified implementation of Sinderby's algorithm in conjunction with the moving average. The moving average can be used as a first step and if the predicted SMR is above 10 dB, the modified Sinderby method can be used.

## **6.8 Conclusion**

The comparison of a moving average method and the template subtraction method for ECG artifact mitigation in sEMG showed comparable results. The moving average method is superior for lower SNR (< 0 dB). It is also more robust to SNR, better suited for real-time

application, and does not require user interaction or a relaxed sEMG recording. However as SNR increases, better performance can be achieved with template subtraction.

Future work will investigate a smoother interpolation between moving averages as given in (6-4) to avoid discontinuities. Further investigation is also needed for dynamic contractions where motion artifact and baseline wander can be present.

A moving average can also be used to accurately quantify motion artifact below 10 dB. If used in conjunction with a modified implementation of Sinderby's method, this range can be extended to 20 dB. Baseline wander could also be detected in this manner as it appears as low frequency components similar to motion artifact but typically less than 1 Hz.

## **7 Classification of sEMG using Support Vector Machines**

### ***7.1 Introduction***

Noise detection in sEMG is traditionally performed by testing for a particular form of contamination in the given signal. If an sEMG signal is presented which is potentially contaminated with an unknown form of noise, it would need to be tested systematically using each individual method before it could be declared clean. Also, methods presented previously assumed the presence or absence of particular contaminants. Having multiple contaminants present could cause issues with some of the methods. Alternatively, a method to differentiate between clean and contaminated sEMG could perform this function in a single step.

Such differentiation could be achieved using a two-class pattern classification method. However, contamination in sEMG comes in many forms and is highly variable. Two-class classification is only accurate when sufficient training examples exist for both classes, and when the training examples are sufficiently representative of the full range of variability of each class [56]. It would be nearly impossible to simulate or collect sufficient sEMG with every possible type of contamination in order to represent the entire class of contaminated sEMG. For this reason, classifying clean versus contaminated sEMG is better suited to one-class classification in which only clean sEMG training data are used. For this analysis, a support vector machine (SVM) is used.

### ***7.2 One-class SVM***

Support vector machines are tools in pattern classification which are used to discriminate between two classes. Linearly inseparable datasets are mapped to a higher dimensional space using a kernel function such that they become linearly separable. An optimization

problem is then solved to find the hyperplane that maximizes the margin between the two classes. SVM training algorithms have been extended to support one-class classification as implemented in [57]. The optimization goal for one-class SVM is to find a hyperplane that best separates the data from the origin within kernel (i.e., higher-dimensional) space. This is equivalent to finding a tightly contained volume for the data in the original feature space.

A set of  $l$  input feature vectors are given  $(x_1, x_2, \dots, x_l)$  and  $x \in X$  where  $X$  is called the feature space. We further assume that a feature map  $\Phi$  exists which maps  $X \rightarrow F$  where  $F$  is a higher-dimensional space called inner product space or kernel space.

The primal optimization problem is expressed as follows [58].

$$\begin{aligned} \min \frac{1}{2} \|w\|^2 + \frac{1}{\nu l} \sum_i \xi_i - b & \quad (7-1) \\ \text{s. t. } w \cdot \Phi(x_i) \geq b - \xi_i, \quad \xi_i \geq 0 & \end{aligned}$$

In (7-1),  $w$  is the normal vector of the hyperplane in kernel space and  $b$  is the hyperplane bias. The constraint in (7-1) is that the inner product of  $w$  and  $\Phi(x_i)$  (i.e., the mapping of  $x_i$  in  $F$ ) should be greater than or equal to the bias  $b$  to classify  $x_i$  in the positive (+1) class since, in a one-class classification problem, all training data is assumed to be from the positive class. The variable  $\xi_i$  is a slack variable to allow some error in the classification margin where  $\xi_i \geq 0$ . A classifier which allows a degree of slack is referred to as a soft margin classifier.

The parameter  $\nu \in (0,1]$  in (7-1) is used to adjust the penalty of permitting slack in the objective function and is optimized as part of the SVM training. Optimizing this function is accomplished using the method of Lagrange multipliers [58]. The Lagrange function  $L$  is given for multipliers  $\alpha_i, \beta_i \geq 0$ :

$$L(w, \xi, b, \alpha, \beta) = \frac{1}{2} \|w\|^2 + \frac{1}{\nu l} \sum_i \xi_i - b - \sum_i \alpha_i ((w \cdot \Phi(x_i)) - b + \xi_i) - \sum_i \beta_i \xi_i \quad (7-2)$$

Differentiating (7-2) with respect to  $w, \xi, b$  and setting this equal to zero yields the following solutions, respectively.

$$w = \sum_i \alpha_i \Phi(x_i) \quad (7-3)$$

$$\alpha_i + \beta_i = \frac{1}{\nu l} \quad (7-4)$$

$$\sum_i \alpha_i = 1 \quad (7-5)$$

The final decision function is shown in (7-6). Note that it is assumed here that  $\text{sgn}(0) = 1$ .

$$f(x) = \text{sgn}(w \cdot \Phi(x) - b) \quad (7-6)$$

Substituting (7-3) into (7-6), we have:

$$f(x) = \text{sgn}(\sum_i \alpha_i \Phi(x_i) \cdot \Phi(x) - b) \quad (7-7)$$

We further introduce a kernel function  $k(x, y)$  (such as the Gaussian kernel [58]) which represents the dot product of  $\Phi(x)$  and  $\Phi(y)$ , i.e., the mapping of feature vectors  $x$  and  $y$  in kernel space). Note that the function  $\Phi(x)$  need not be known explicitly so long as its dot product is known.

$$k(x, y) = \Phi(x) \cdot \Phi(y) \quad (7-8)$$

Substituting (7-8) into (7-7):

$$f(x) = \text{sgn}(\sum_i \alpha_i k(x_i, x) - b) \quad (7-9)$$

We can transform the primal optimization problem in (7-1) into the dual optimization problem by using (7-3), (7-4), (7-5), and (7-8):

$$\min \frac{1}{2} \sum_{i,j} \alpha_i \alpha_j k(x_i, x_j) \quad (7-10)$$

$$\text{s.t. } 0 \leq \alpha_i \leq \frac{1}{v_i}, \quad \sum_i \alpha_i = 1$$

where for all feature vectors  $x_i$  for which the constraint in (7-1) is an equality:

$$b = \sum_j \alpha_j k(x_j, x_i) \quad (7-11)$$

The feature vectors  $x_i$  which satisfy (7-11) are the *support vectors*, consisting of the hardest points to classify in the dataset. Equation (7-10) can be solved using quadratic optimization techniques and an optimum is guaranteed to exist. The two difficulties are 1) reaching the optimum efficiently, which depends on the algorithm efficiency, and 2) generalizing well to unseen data, which depends on the training data.

### **7.3 Feature Extraction**

To avoid the curse of dimensionality [59], pattern classification of biosignals such as sEMG is better performed on descriptive features extracted from each signal to be classified rather than operating on the raw signal values themselves. The better the discriminability of these features, the more accurate the classification will be. Discriminability refers to the degree to which each feature is able to discriminate between signals of each class.

The following features [60] were extracted from each sEMG signal and were used for classification. The discriminability of each feature was later examined.

#### **7.3.1 Autoregressive Model Coefficients**

An autoregressive model (ARM) computes the  $n^{\text{th}}$  term of a sequence,  $x$ , as a linear combination of the previous  $M$  terms. The aim is to find coefficients,  $a_1, a_2, \dots, a_M$ , which minimize the squared error in (7-12). Note that  $w_n$  is a white noise error term.

$$x_n = \sum_{i=1}^M a_i x_{n-i} + w_n \quad (7-12)$$

These coefficients are found using the linear least squares estimator. Once the coefficients are determined from an sEMG training sample, the coefficients may be used as descriptive features as input to the SVM classifier. For this analysis,  $M=4$  is used.

### 7.3.2 sEMG Histogram

The range of the sEMG signal is computed and is split into  $n$  bins of equal size. The number of signal samples whose value falls in each bin is counted. The number of occurrences for each bin is used as a separate feature. For this analysis,  $n=10$  is used.

### 7.3.3 Mean Absolute Value

The mean absolute value (*MAV*) is computed for a signal,  $x$ , of length  $N$  as shown in (7-13). Note it is also referred to as the average rectified value. The MAV can be used directly as a feature.

$$MAV = \frac{1}{N} \sum_{i=1}^N |x_n| \quad (7-13)$$

### 7.3.4 Modified Median Frequency

The median frequency of a signal is the frequency where the power spectrum is divided into two partitions where the total power of one partition is equal to that of the other. The modified median frequency is similar, only it is computed from the amplitude spectrum rather than the power spectrum. If the amplitude spectrum is denoted by vector  $A$  of length  $N$  then (7-14) should be satisfied for index  $M$  corresponding to the modified median frequency. Values from the amplitude spectrum are used here as they have less variability (since the amplitude values are not being squared).

$$\sum_{i=1}^M A_i = \sum_{i=M}^N A_i \quad (7-14)$$

### 7.3.5 Modified Mean Frequency

The modified mean frequency (MMNF) is the weighted average frequency computed over the amplitude spectrum. It is computed for amplitude spectrum  $A$  with length  $N$  as shown in (6-15).

$$MMNF = \frac{\sum_{i=1}^N f_i A_i}{\sum_{i=1}^N A_i} \quad (7-15)$$

### 7.3.6 Power Spectral Density

The power spectral density is computed using the Fast Fourier Transform (FFT) and then squaring the result. For a signal of  $N$  samples, the length of the spectrum for positive angular frequencies (0 to  $\pi$ ) is  $N/2$ . This power spectrum can be shortened by dividing it into  $k$  frequency bins (where  $k < N$ ) and then computing the mean of each bin's spectrum values. These averaged values are used as features for classification. For this analysis,  $k=10$  was used.

### 7.3.7 Slope Sign Changes

The number of slope sign changes (SSC) is computed as shown in (7-16). A threshold value  $T \leq 0$  can be used to control the false positive identification rate due to random noise, however this may also increase the false negative rate. A value of 0 should be used if very small slope values are expected. For this analysis, a value of  $T=0$  was used.

$$SSC = \sum_{i=2}^{N-1} f((x_i - x_{i-1})(x_{i+1} - x_i)) \quad (7-16)$$

$$f(x) = \begin{cases} 1, & x \leq T \\ 0, & x > T \end{cases}$$

### 7.3.8 Willison Amplitude

The Willison Amplitude (*WAMP*) is defined as the number of times the absolute difference between adjacent samples exceeds a predetermined threshold  $T$  as shown in (7-17). An accepted threshold (depending on amplifier gain) is usually between 10 mV and 100 mV [60]. For this analysis, a value of 10 mV ( $T=0.01$ ) was used.

$$WAMP = \sum_{i=1}^{N-1} f(|x_{i+1} - x_i|) \quad (7-17)$$

$$f(x) = \begin{cases} 1, & x \geq T \\ 0, & x < T \end{cases}$$

### 7.3.9 Waveform Length

The waveform length (*WL*) is calculated as the cumulative length of the entire signal, or the sum of the consecutive absolute differences of the signal as given in (7-18). The computation is identical to that from (7-17) only without applying the threshold function  $f$ .

$$WL = \sum_{i=1}^{N-1} |x_{i+1} - x_i| \quad (7-18)$$

### 7.3.10 Zero Crossings

The number of zero crossings (*ZC*) is defined as the number of times the sEMG signal amplitude crosses the x-axis (either positive amplitude to negative, or negative amplitude to positive). Once again, a threshold value  $T < 0$  can be used here to avoid low amplitude fluctuations about the baseline. The number of zeros crossings is found using (7-19). For this analysis, a value of  $T=0$  was used.

$$ZC = \sum_{i=1}^{N-1} f(x_n x_{n+1}) + \sum_{i=1}^{N-1} g\left(\frac{x_{n+1}}{x_n}\right) \quad (7-19)$$

$$f(x) = \begin{cases} 1, & x < T \\ 0, & x \geq T \end{cases}$$

$$g(x) = \begin{cases} 1, & x = 0 \\ 0, & x \neq 0 \end{cases}$$

## **7.4 Feature Optimization**

Not all features in Section 7.3 were useful for differentiating clean and contaminated sEMG. Some features were found to vary too much for clean sEMG samples (i.e., these features exhibited large intraclass variability) and may not be sensitive enough to the contamination to provide good discriminability. The goal of feature optimization is to identify the smallest subset of features that provides good discriminability between clean and contaminated sEMG samples. By reducing the total number of features used in classification, we also reduce the complexity of the classifier and help to avoid the curse of dimensionality.

To evaluate the discriminability of each feature, sEMG was simulated with varying shaping parameters and a signal length of 4096 samples (sampled at 1080 Hz). Referring back to section 3.2,  $f_i$  was randomly generated in the range [30 60] and  $f_h$  in the range [60 160]. The simulated dataset consisted of 204 clean signals and 204 contaminated signals. The contamination was artificially generated and added to the clean signals. Six contaminants (power line interference, ECG interference, motion artifact, quantization noise, ADC clipping, and amplifier saturation) were inserted on equal-sized subsets of the dataset (i.e., 34 signals per contaminant type). The levels of contamination inserted were chosen such that on an initial feature set (consisting of the first AR coefficient, the first and last 3 bins of the 10-bin amplitude histogram, the first 2 bins of the 10-bin power spectrum, and all remaining features), classification accuracy ranged from 70% to 80%. This initial feature set was chosen arbitrarily based on a preliminary, non-comprehensive manual (trial and error) feature optimization. These contamination levels are shown in Table 7-1.

**Table 7-1: Contaminant levels used for feature optimization**

<b>Contaminant</b>	<b>Amount</b>
Motion Artifact	SNR at 3 dB
Power line interference	SNR at 5 dB
ECG interference	SNR at 3 dB
Quantization noise	Step size at $2^{-4.25}$ V
Clipping	ADC max at 2.75 V
Amplifier saturation	Amplifier gain at 3.5*

\* Note that this is an artificial software gain factor applied after the hardware gain factor and after the signal has been normalized (multiplied by a constant).

Each of the 31 features was removed from the dataset, one at a time. For each feature, 10 different datasets were simulated and used to train the SVM. The SVM was trained on the clean sEMG only and then tested on the contaminated sEMG. An average and standard deviation classification accuracy was computed for the 10 SVM training iterations for each feature removal. Once all 31 features had been removed in isolation, the feature removal which resulted in the highest classification accuracy (deemed the worst feature) was permanently removed. Following this greedy feature selection approach, the procedure continued removing each remaining feature and measuring the accuracy while permanently removing the worst feature at each iteration. The procedure automatically terminated when 3 features were remaining, regardless of whether or not they were optimal. When determining the highest accuracy at the end of each round of feature removals, an effective accuracy ( $acc'$ ) was computed as given in (7-20).

$$acc' = \mu_{acc} - 0.25\sigma_{acc} \quad (7-20)$$

In (7-20), the effective accuracy,  $acc'$ , is the weighted difference of the mean accuracy ( $\mu_{acc}$ ) and the standard deviation ( $\sigma_{acc}$ ). This was done to ensure that a marginally higher classification accuracy with a much higher variance was not chosen to be preferable.

Two independent executions of the feature optimization were performed. Classification accuracy was recorded before and after each feature removal iteration. For both execution runs, optimal accuracy with low variance was observed after 24 feature removals (i.e., when 7 features remained in the feature set). However, feature sets were identical for both execution runs after one more feature removal (i.e., 6 features remaining in the feature set). Therefore, the 6 features in common to both separate optimizations were chosen as the optimal feature set. The optimal feature set consisted of the first and last bins of the 10-bin amplitude histogram, the mean absolute value, the Willison amplitude, and the first 2 bins in the 10-bin power spectrum. This is summarized in Table 7-2.

**Table 7-2: Optimal feature set for SVM classification of sEMG**

	<b>Feature</b>
<b>1</b>	10-bin amplitude histogram (first bin)
<b>2</b>	10-bin amplitude histogram (last bin)
<b>3</b>	Mean absolute value
<b>4</b>	Willison amplitude
<b>5</b>	10-bin power spectrum (first bin)
<b>6</b>	10-bin power spectrum (second bin)

## **7.5 SVM Performance Evaluation**

LIBSVM [57] was used to train SVMs on both simulated data [31] and real data collected using the experimental protocol described in Chapter 3. For one-class classification, the SVM is trained on the positive examples (clean sEMG) only and then tested on both clean and contaminated (negative) sEMG. The classification accuracy of the SVM was evaluated by varying the SNR of the signals being contaminated. Different contaminants were used and evaluated separately first, and then in combination.

Two datasets (i.e., one consisting of simulated sEMG and the other consisting of real sEMG) were used to evaluate the SVM classification accuracy. A dataset of 400 sEMG signals was simulated [31] and partitioned into a 50/50 training/testing split. All simulated signals were quantized to 16 bits using equation (5-1), shifted to zero mean, and normalized to a signal power of 1. One hundred iterations were repeated, simulating new signals at each iteration, and results were averaged. A dataset consisting of 167 real sEMG signals (shifted to zero mean and normalized to a signal power of 1) collected using the experimental protocol in Chapter 3 was also used. Fourteen iterations were performed where each iteration consisted of leaving out one entire muscle group for a given subject (5 subjects and 3 muscles per subject) to be used as the testing set. One entire muscle group was removed from the dataset due to poor signal quality which resulted in only 14 iterations instead of 15. Each iteration consisted of a test set size between 5 and 15 signals. The variation in test set size was due to the removal of signals with low SNR from the entire group. This SNR was estimated using the signal power from the rest signal immediately following contraction. The testing dataset was then manually contaminated with one of power line interference, ECG interference, motion artifact, quantization noise, clipping, or amplifier saturation. The degree of contamination for each contaminant type is defined in section 7.6.

The 6 features determined by feature optimization in Section 7.4 were extracted from each signal. All features for each signal were combined into a feature vector and written to either a training file or testing file in LIBSVM format. The training data was linearly scaled to  $[-1 +1]$  and the same scaling factor was then applied to the testing data. A two-dimensional grid search was then performed to locate the optimal values of  $\nu, \gamma$  for the SVM classification. The parameter  $\nu$  is used as given in (7-1) and  $\gamma$  is the spread parameter in

the Gaussian radial basis function (RBF) kernel as shown in (7-21). Note that  $x, y$  are feature vectors.

$$k(x, y) = e^{-\gamma \|x-y\|^2} \quad (7-21)$$

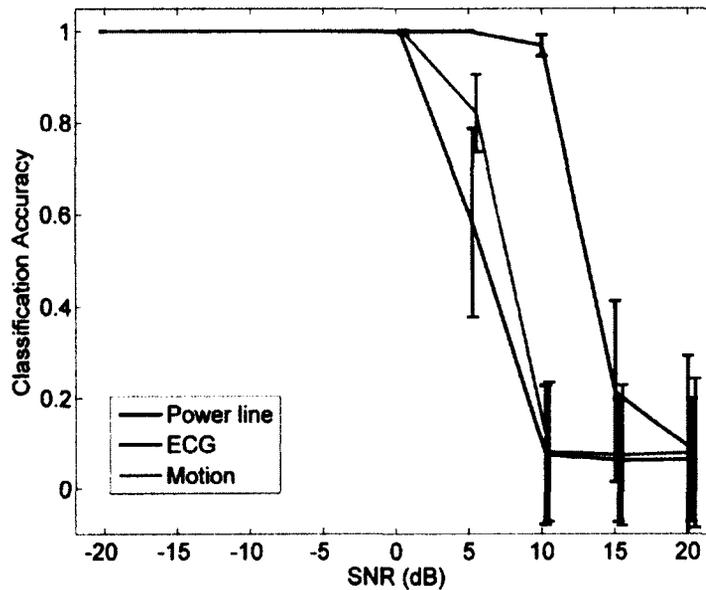
Once the optimal values for parameters  $\nu, \gamma$  were found, the SVM was trained using those parameters on the training dataset and the classification performance evaluated on the testing dataset.

## **7.6 Results for Classification of sEMG with a Single Contaminant**

The entire test set was contaminated with a single type of contamination (power line interference, ECG interference, motion artifact, quantization noise, clipping, or amplifier saturation). Classification accuracy of the SVM was first evaluated on a clean test set (i.e., before any contamination was introduced), and second, accuracy was evaluated as a function of the amount of contamination inserted for each of the six contaminants. Accuracy is computed as the fraction of correct classifications.

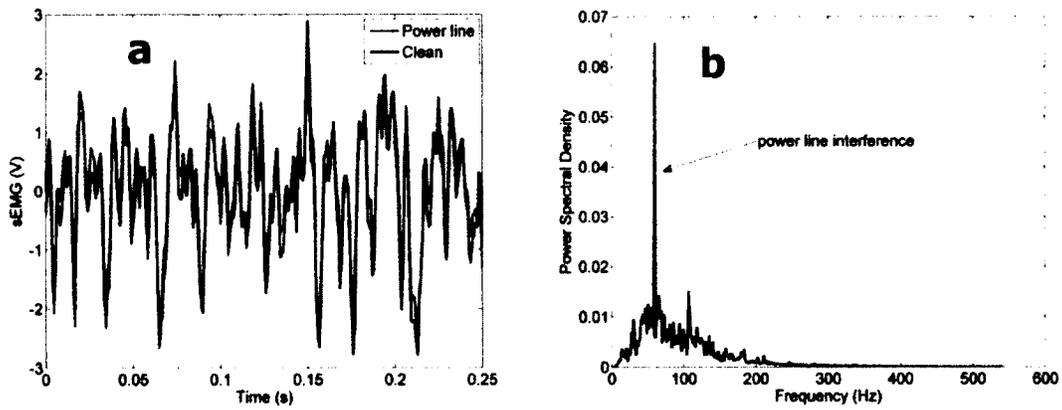
### **7.6.1 Classification of sEMG with Power Line Interference, ECG Interference, or Motion Artifact**

The testing dataset was manually contaminated with one of power line interference, ECG interference, or motion artifact. The power line interference was created artificially by generating a sinusoid with a random frequency between -59.5 Hz and 60.5 Hz, a random phase, and amplitude corresponding to SNR between -20 dB and 20 dB in increments of 5 dB. ECG interference was inserted from real signals used in Chapter 6, and motion artifact from [54], [55]. The performance with simulated sEMG is shown in Figure 7-1.



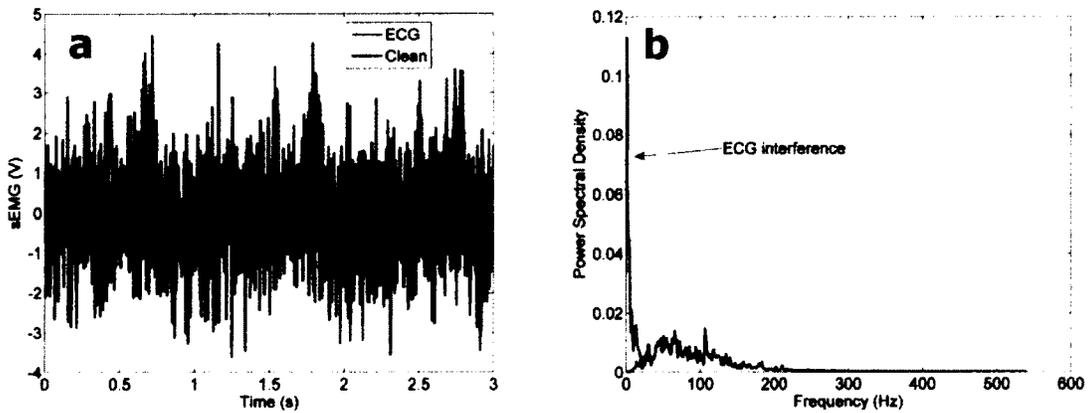
**Figure 7-1: Classification accuracy of one-class SVM trained with simulated sEMG and tested on sEMG contaminated with power line interference, ECG interference, or motion artifact with varying SNR. SNR values are offset for visibility. Error bars are at  $\pm 1$  standard deviation.**

Classification accuracy is 100% for low SNR (large amount of contamination). A transition point exists for each contaminant where accuracy drops substantially (approximately 10 dB for power line interference, 3 dB for ECG interference, and 5 dB for motion artifact). Past this transition point, the contamination levels are too low to be detected in the signal and the accuracy falls to 0. In this section, the transition point is defined as the smallest amount of contamination where the classification accuracy is 80% or greater. Transition point values are approximate and will vary with the dataset. We see the effect of power line interference at the transition point of 10 dB on simulated sEMG in the time and frequency domain (Figure 7-2). Note that classifying high SNR signals as clean is an acceptable error (and a reasonable decision). In these cases, the noise is minute and is not distorting any of the features which is the reason for the SVM being unable to differentiate them from clean signals.



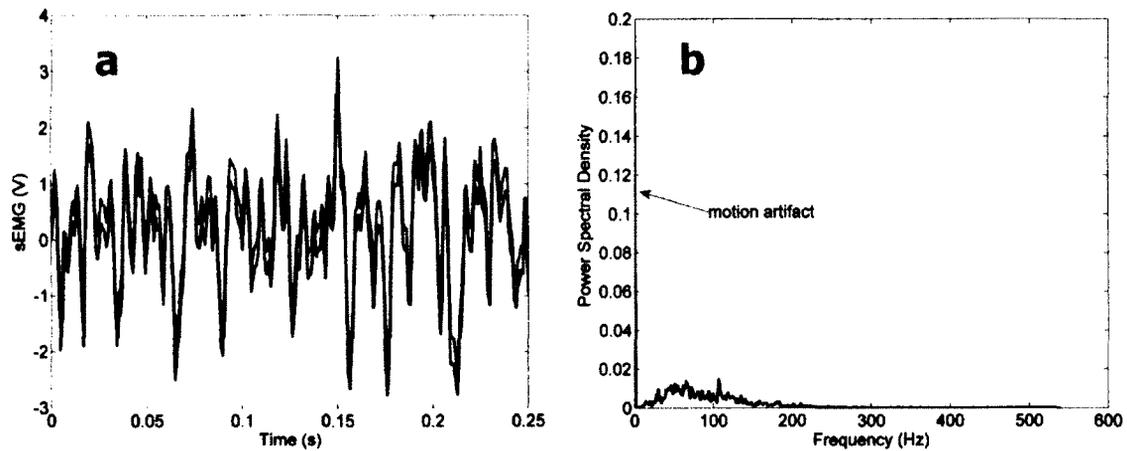
**Figure 7-2: The effect of power line interference at 10 dB on a simulated sEMG signal in the (a) time and (b) frequency domain.**

In Figure 7-3, we see time and frequency domain plots showing the effect of ECG interference at 3 dB.



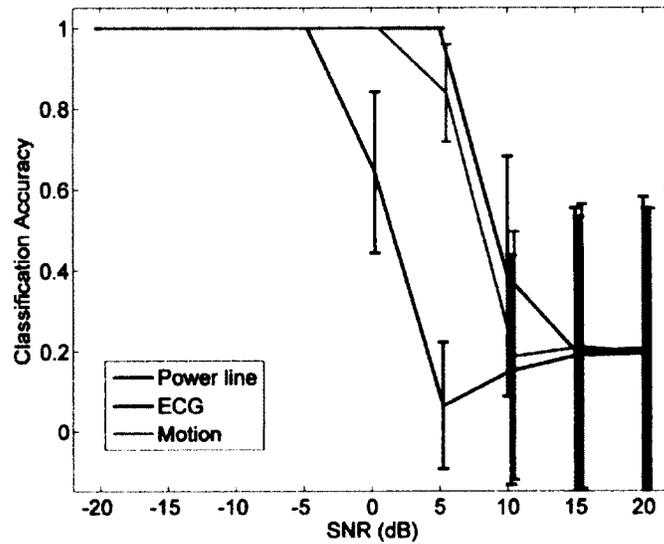
**Figure 7-3: The effect of ECG interference at 3 dB on a simulated sEMG signal in (a) time and (b) frequency domain.**

In Figure 7-4, we see time and frequency domain plots showing the effect of motion artifact at 5 dB.



**Figure 7-4: The effect of motion artifact at 5 dB on a simulated sEMG signal in (a) time and (b) frequency domain.**

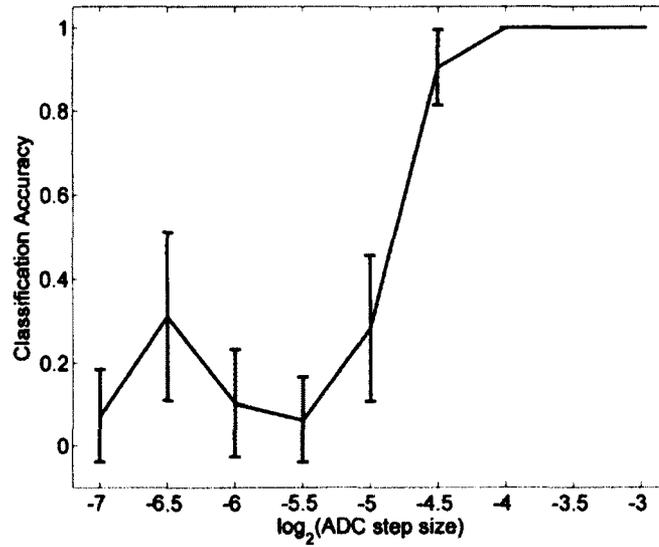
The performance on the real sEMG data is shown in Figure 7-5. The trend is similar to that of the simulated data (see Figure 7-1). With real data, the transition point where the accuracy begins to drop occurs at a lower SNR. This is likely due to a combination of two factors. The first factor is that real sEMG will be more variable than the idealized simulated sEMG used previously. The second factor is that there is real unintended motion artifact and power line interference in the real data collected. Usually the amounts are small, however, they are still measurable and when the SVM is trained on slightly contaminated data, it will become less sensitive to those types of contaminants in the test data.



**Figure 7-5: Classification accuracy of one-class SVM trained with real sEMG and tested on sEMG contaminated with power line interference, ECG interference, or motion artifact with varying SNR. SNR values are offset for visibility. Error bars are at  $\pm 1$  standard deviation.**

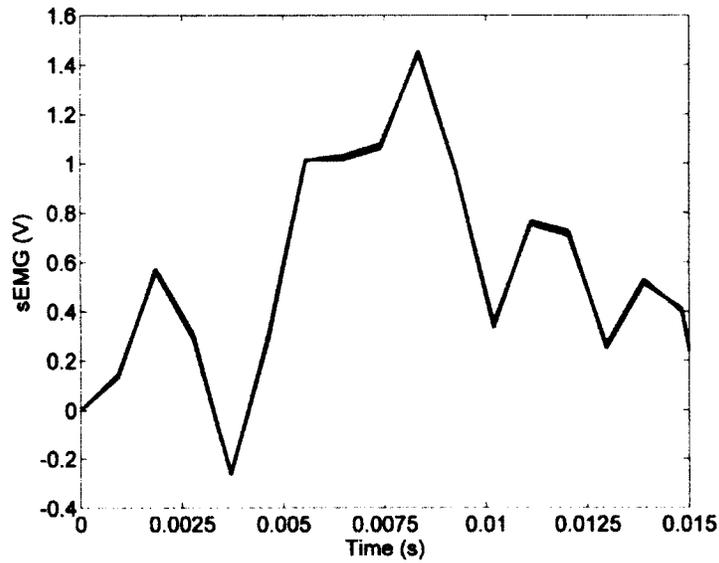
### 7.6.2 Classification of sEMG with Quantization Noise

The testing dataset was manually contaminated with quantization noise using (5-1) and varying the ADC step size to each the following:  $2^{-3}$ ,  $2^{-3.5}$ ,  $2^{-4}$ ,  $2^{-4.5}$ ,  $2^{-5}$ ,  $2^{-5.5}$ ,  $2^{-6}$ ,  $2^{-6.5}$ ,  $2^{-7}$ . Classification accuracy as a function of step size is shown in Figure 7-6.



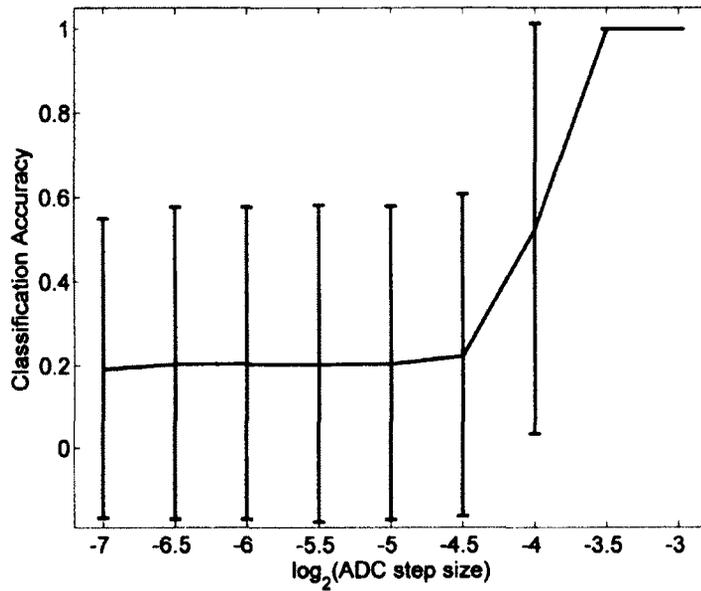
**Figure 7-6: Classification accuracy of simulated sEMG with varying degrees of quantization noise. Error bars are at  $\pm 1$  standard deviation.**

The transition point for SVM classification with quantization noise occurs at an ADC step size of approximately  $2^{-4.5}$ . The effect of quantization noise at this step size on simulated sEMG in the time domain is shown in Figure 7-7. Note that the small deviations in the time domain due to the rounding (quantization) error have no effect on the frequency domain. Although these deviations can be seen here, they would not be apparent if the clean signal were not available.



**Figure 7-7: The effect of quantization noise at an ADC step size of  $2^{-4.5}$  on simulated sEMG in the time domain**

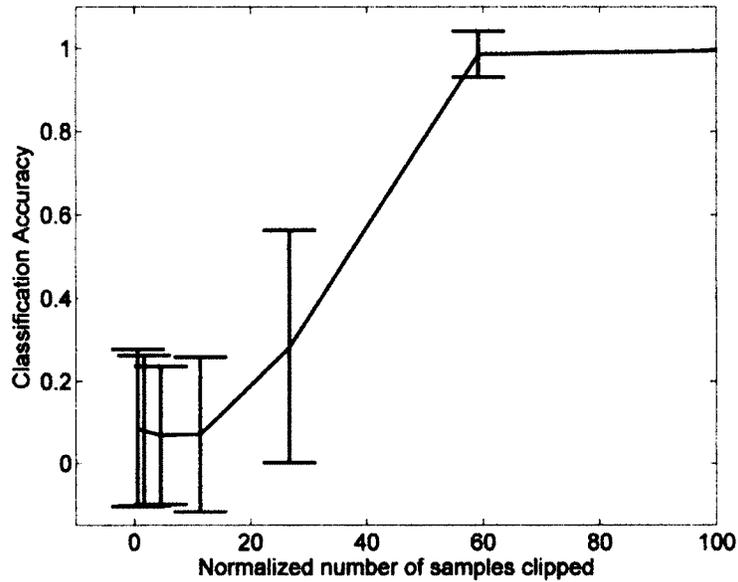
The performance on the dataset of real sEMG is shown in Figure 7-8. With real data, the accuracy is low (the variance is due to clean signals being incorrectly identified as contaminated in the dataset, causing unexpected performance for certain test set partitions and not others) until a step size of about  $2^{-3.5}$  when accuracy reaches 100%. Thus, quantization noise appears to be more difficult to detect in real, as opposed to simulated signals. This could partly be due to the fact that the amplitude of the real signals tended to be larger than that of the simulated signals, despite the fact that both were normalized to a power of 1. Higher amplitude signals will be less impacted by quantization noise at a given resolution when compared to signals with lower amplitude.



**Figure 7-8: Classification of real sEMG contaminated with quantization noise at different ADC step sizes. Error bars are at  $\pm 1$  standard deviation.**

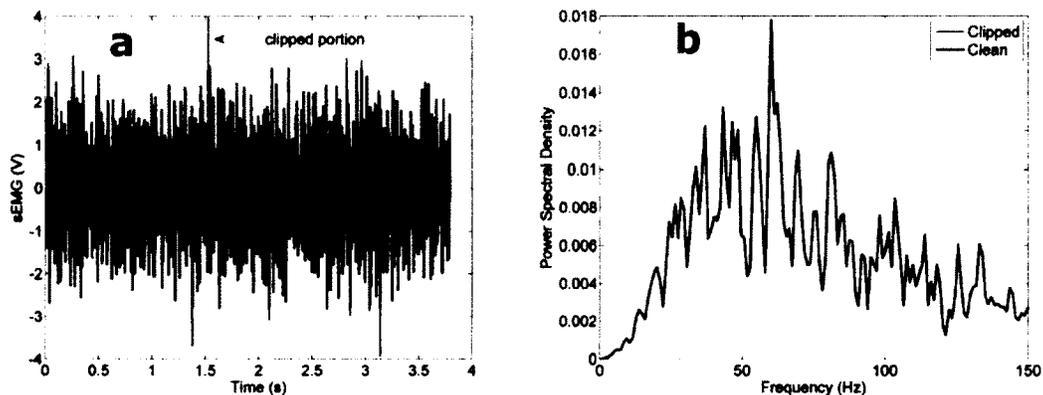
### 7.6.3 Classification of sEMG with ADC Clipping

The testing dataset was manually contaminated with ADC clipping using (5-1) with a variable ADC max (2-4 V in 0.25 V increments). A plot of classification accuracy as a function of the normalized number of signal samples clipped is shown in Figure 7-9. The normalized number of clipped samples is computed by counting the number of clipped samples, dividing by the signal length and multiplying by 10000.



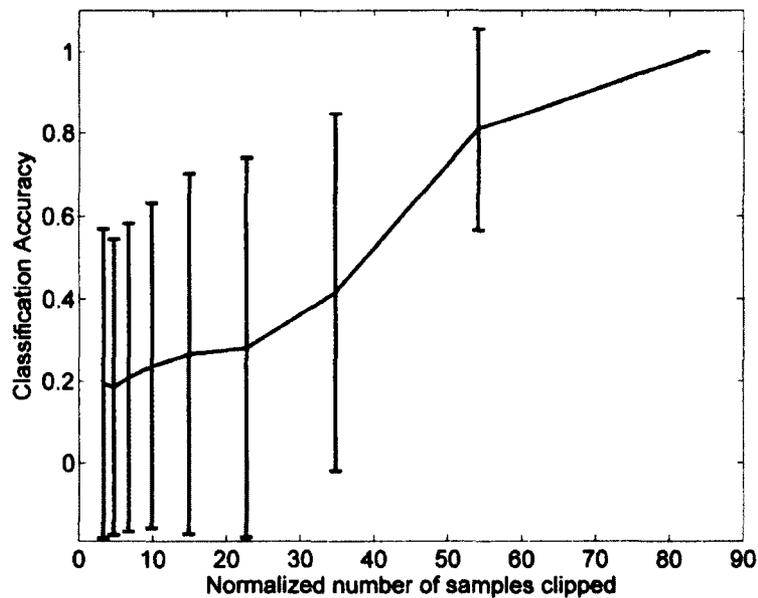
**Figure 7-9: Classification of simulated sEMG contaminated with ADC clipping at different mean number of samples clipped (normalized by dividing by the signal length of 4096 and multiplying by 10,000). Error bars are at  $\pm 1$  standard deviation.**

For ADC clipping, the transition point for the classification accuracy occurs at a normalized number of samples clipped of approximately 50 for simulated sEMG. The effect of clipping in the time and frequency domain with a normalized number of samples clipped of 49 is shown in Figure 7-10. Portions of the signal where the red (clipped) series is visible indicate regions where the signal was clipped.



**Figure 7-10: The effect of clipping with a normalized number of samples clipped of 49 on simulated sEMG in the (a) time domain and (b) frequency domain.**

The classification accuracy for the real sEMG dataset is shown in Figure 7-11. The accuracy increases gradually with the number of samples clipped. The accuracy is reasonable beyond a normalized number of samples clipped of approximately 55, just marginally higher than the 50 for simulated data. However, considering the relatively high variance here (standard deviation of 12.8 for real data, and 21.7 for simulated) there does not appear to be a significant difference between the transition points for simulated versus real data in this case. It should be noted that the simulated and real data were at different amplitudes despite having the same signal power. As such, different ADC ranges were used (for real data, the range values were increased from 2-4 V to 3-5 V) in order to achieve a similar range of number of samples clipped.

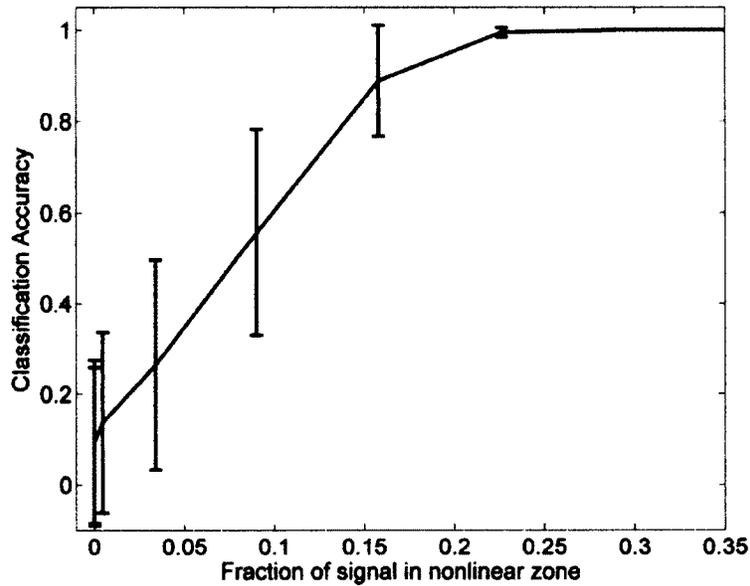


**Figure 7-11: Classification of real sEMG contaminated with ADC clipping at different mean number of samples clipped (normalized by dividing by the signal length of 30,003 and multiplying by 10,000). Error bars are at  $\pm 1$  standard deviation.**

#### 7.6.4 Classification of sEMG with Amplifier Saturation

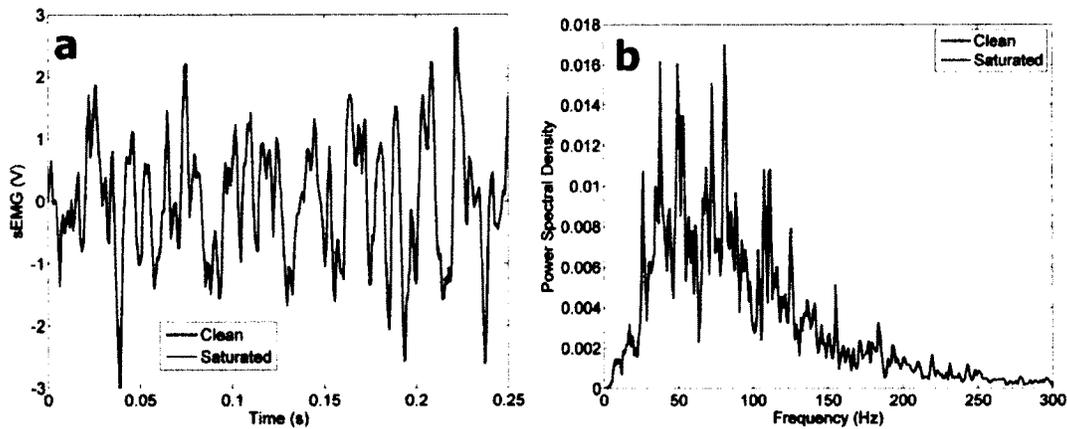
The testing dataset was manually contaminated with amplifier saturation using (5-8) with an amplifier max set to 10 V and variable amplifier gain (0.5-4.5 in increments of 0.5). A

plot of classification accuracy as a function of the fraction of the signal distorted is shown in Figure 7-12. The fraction of the signal distorted is defined as the fraction of signal samples outside the linear range of the amplifier (i.e., regions where equation (5-9) is satisfied).



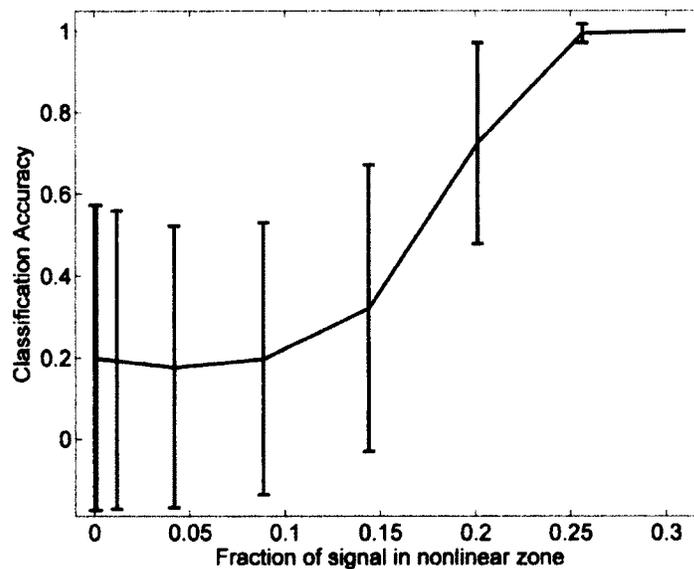
**Figure 7-12: Classification of simulated sEMG contaminated with amplifier saturation at different mean fractions of the signal in the amplifier nonlinear zone. Error bars are at  $\pm 1$  standard deviation.**

The transition point for amplifier saturation occurs at approximately 16-17% of the signal in the nonlinear zone for simulated sEMG. The effect of amplifier saturation in the time and frequency domain at this degree of amplifier saturation is shown in Figure 7-13.



**Figure 7-13: The effect of amplifier saturation when 16% of the signal falls in the amplifier nonlinear zone on simulated sEMG in the (a) time domain and (b) frequency domain.**

The classification accuracy for the real sEMG dataset contaminated with amplifier saturation is shown in Figure 7-14. The transition point for the real data seems to be higher at around 23% of the signal saturated and the real data exhibits higher variance. The variance in the fraction of the signal in the nonlinear zone figures was small for both real and real simulated (approximately 0.7% for real data and 0.5% for simulated data).



**Figure 7-14: Classification of real sEMG contaminated with amplifier saturation at different fractions of the signal in the amplifier nonlinear zone. Error bars are at  $\pm 1$  standard deviation.**

Table 7-3 summarizes the transition points for each contaminant for simulated and real data.

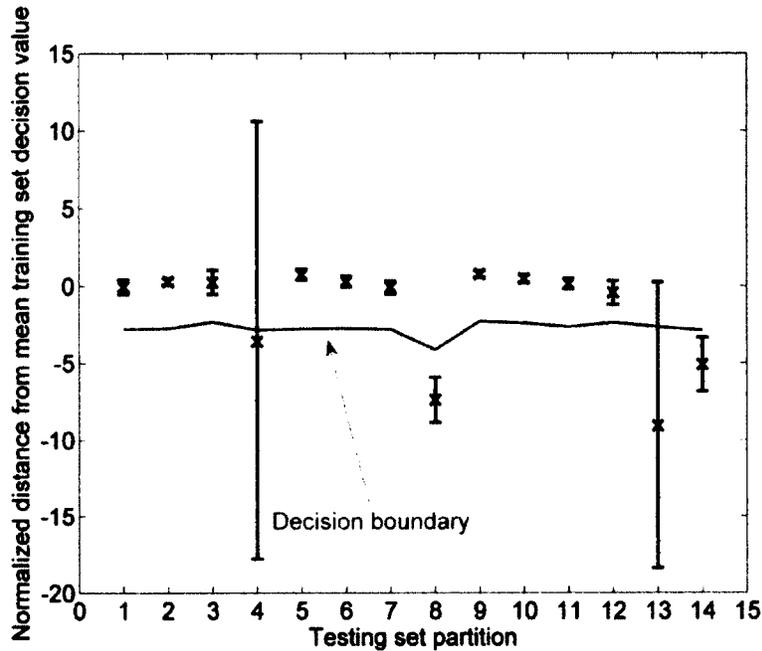
**Table 7-3: Transition points in SVM classification accuracy for each contaminant for simulated and real sEMG**

<b>Contaminant</b>	<b>Transition point for simulated data</b>	<b>Transition point for real data</b>
Power line interference	10 dB	SNR: 5 dB
ECG interference	3 dB	SNR: -3 dB
Motion artifact	5 dB	SNR: 5 dB
Quantization noise	ADC step size: $2^{-4.5}$	ADC step size: $2^{-3.5}$
ADC clipping	Normalized number of samples clipped: 50	Normalized number of samples clipped: 55
Amplifier saturation	16% of signal in amplifier nonlinear zone	23% of signal in amplifier nonlinear zone

### 7.6.5 Using the SVM Classification Approach to Quantify Signal Contamination

The SVM can perform a classification for clean versus contaminated sEMG with a degree of accuracy which is SNR-dependent. However, it is not designed to express the degree of cleanliness or contamination in the sEMG. This is not an issue for signals which are severely contaminated (low SNR); however it would be desirable to have a measure of contaminant quantification for signals with moderate to low amounts of contamination. In such circumstances, the magnitude of the decision function in (7-9) (i.e., the raw decision value before the *sgn* function is applied) may be very close to 0, indicating that the signal may fall very close to the decision boundary. It was hypothesized that the raw decision value of (7-9) can be useful in determining how much confidence to put in a given SVM classification, and may also serve as a means to estimate the degree of contamination in the signal. A human operator can set their own threshold of acceptability (i.e., set their own decision boundary). For example, a higher decision boundary value would make the system more conservative, classifying signals that were borderline cases as contaminated.

For this portion of the analysis, a clean uncontaminated test set was used. This was done to observe how the decision values vary between and within test set partitions. For each of the 14 SVM test partitions for the real dataset, the mean and standard deviation of the raw decision values for the training set were calculated. These values were used to normalize (subtract mean and divide by standard deviation) the decision values obtained from classifying the test set with that SVM. By normalizing the testing decision values in this manner with respect to the training decision values, we are able to compare how the test vector decision values cluster for a given SVM relative to others. Furthermore, considering that the test partitions correspond to different muscle groups for different subjects, this analysis examines inter-subject and inter-muscle variability. This is shown in Figure 7-15. The position of each marker is the mean decision value for that testing dataset partition after normalization relative to the training set. A position above the red line indicates a clean (+1) classification. A position below the red line indicates a contaminated (-1) classification, which is incorrect in this case since these results are on the clean portion of the test set only. Tighter error bars indicate better clustering of the test set data and a value close to zero indicates that that test set partition clustered close to the training set partition.

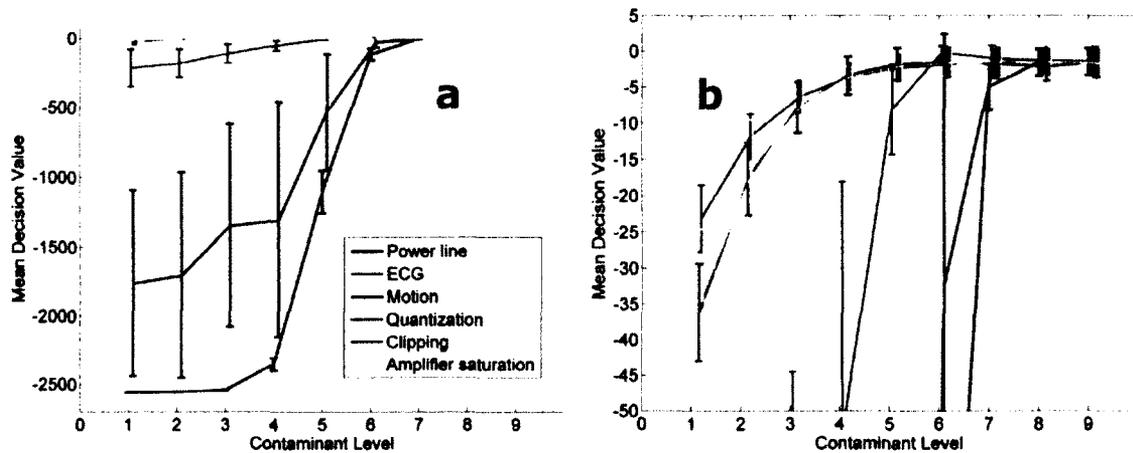


**Figure 7-15: Decision values for clean test set sEMG normalized by training decision values for each SVM testing set partition. The decision boundary indicates the change in classification value**

The plot in Figure 7-15 also serves as an indicator for outliers. Testing set partitions which exhibit abnormally large variance may contain certain sEMG signals that are not, in fact, clean. Partitions which fall entirely below the decision boundary may contain sufficient contamination to be completely rejected by the classifier. We can observe that partitions 4 and 13 exhibit high variance, and partitions 8 and 14 fall entirely below the decision boundary. It was noted that for partition 4, the subject did not always begin the contraction before the recording began, thus, creating a silent portion at the beginning of certain recordings. This was observed to cause a deviation in the feature values from the mean of the dataset. For partition 13, the power spectrum was observed to show motion artifact in some circumstances. This would likely account for the large variance. The power spectrum for some signals in partition 8 contained power line interference and harmonics, despite having a high SNR estimate when compared with the rest recording. Lastly, no problems were observed with the signals from partition 14; however these signals were from the subject with the largest number of rejected signals (due to low SNR) and could be a result

of improper contraction technique or an improper setup for that contraction. These four partitions were still included in the dataset which could have skewed the results if these partitions really did represent contaminated sEMG. It should be noted that training on a reduced dataset free of these outliers will not necessarily increase the transition point for the SVM classification. Rather, the amount of variance in the normalized decision values observed for low levels of contamination will be reduced. SVMs can operate in the presence of noisy data so long as the majority of the training set is representative of the class of data in question.

The remaining analysis will use a contaminated test set where the contamination level is measurably varied. The decision values for the contaminated test set were computed for each of the 6 contaminants and the 14 testing set partitions at each of the 9 levels of contamination. These values were observed to be correlated with the amount of contamination in the signal. The decision values were averaged over all testing set partitions for each contaminant. The mean decision value curve as a function of the contaminant level is shown in Figure 7-16. The contaminant levels for each contaminant type are defined in Table 7-4.



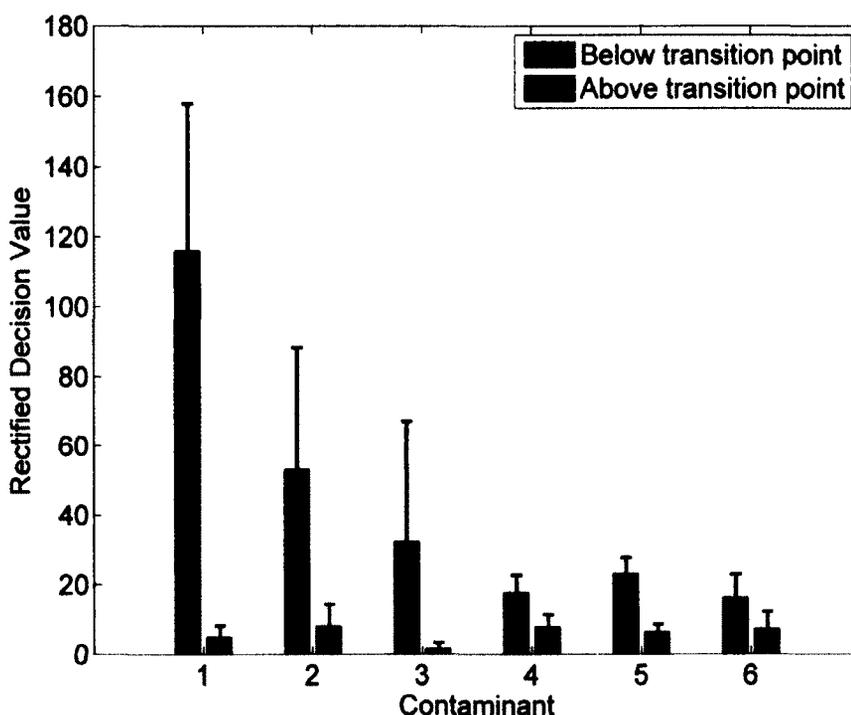
**Figure 7-16: Mean decision value for each contaminant over all test set partitions at different contaminant levels at (a) full scale and (b) zoomed in. The contaminant levels are defined for each contaminant type in Table 7-4.**

**Table 7-4: Amounts of contamination for each numeric contaminant level**

<b>Contaminant Level</b>	<b>Power Line, ECG, or Motion Artifact (SNR in dB)</b>	<b>Quantization (ADC step size)</b>	<b>Clipping (ADC max)</b>	<b>Amplifier Saturation (Gain)</b>
1	-20	$2^{-3}$	3	4.5
2	-15	$2^{-3.5}$	3.25	4
3	-10	$2^{-4}$	3.5	3.5
4	-5	$2^{-4.5}$	3.75	3
5	0	$2^{-5}$	4	2.5
6	5	$2^{-5.5}$	4.25	2
7	10	$2^{-6}$	4.5	1.5
8	15	$2^{-6.5}$	4.75	1
9	20	$2^{-7}$	5	0.5

We can see that the curves in Figure 7-16 follow an S shape. High levels of contamination cause the decision values to saturate at very low values. Low levels of contamination cause the decision values to approach zero. There is also a transition region in between containing an inflection point. This is where the ability of the SVM is to detect the contamination will change. Detection of quantization noise, clipping, and amplifier saturation was more difficult and the ranges of contaminant levels chosen were skewed towards the right end of the S curve so only half of this trend is seen. It is likely that a wider range would show the same relationship as is seen for power line interference, ECG interference, and motion artifact.

Instead of looking at the whole curve, Figure 7-17 is a plot of the rectified decision values (as they are negative values) for contamination levels during the transition region (i.e., below and above the transition point) for each contaminant.



**Figure 7-17: Mean decision values over all testing set partitions for each contaminant below and above its transition point: 1) Power line interference at 5 dB and 10 dB, 2) ECG interference at -5 dB and 0 dB, 3) motion artifact at 5 dB and 10 dB, 4) quantization noise at ADC step size of  $2^{-3.5}$  and  $2^{-4}$ , 5) ADC clipping at an ADC max of 3 V and 3.5 V, and 6) Amplifier saturation at an amplifier gain of 4 and 3.5.**

The decision values above the transition point in Figure 7-17 are similar for all contaminants. The decision values below the transition point can vary, but are usually sufficiently greater than the values above the transition point to allow one to develop a confidence threshold in the classification result based on the decision value. This implies that a generalized decision threshold can be used to discern contaminated signals (e.g., from Figure 7-17, a decision threshold of -10 could be used). The large variance in the values below the transition point is likely due to outliers. A larger dataset with a strict method of outlier rejection should reduce this variance, however, such a restricted dataset may not be suited for all applications. This variance in decision values below the transition point may be acceptable, in which case the threshold value could be adjusted based on an acceptable false negative rate.

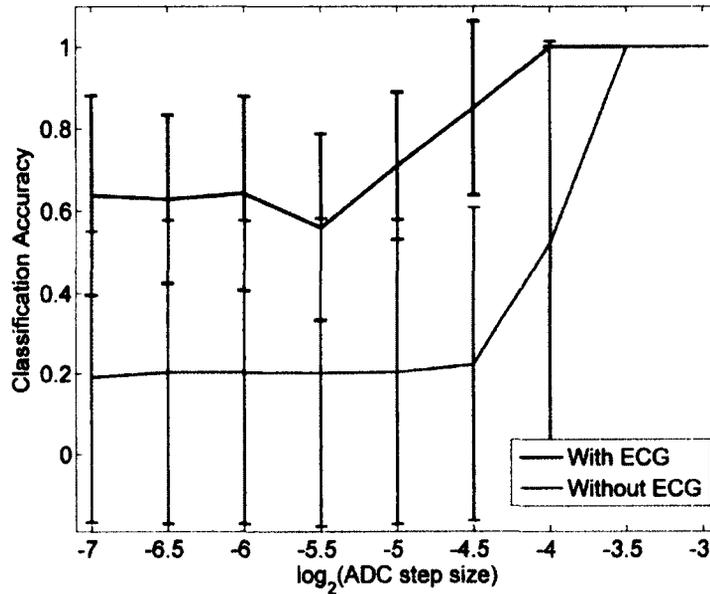
## **7.7 Classification of sEMG with Two Simultaneous Contaminants**

The real sEMG dataset was combined with two of the six contaminants (motion artifact, power line interference, ECG interference, quantization noise, ADC clipping, and amplifier saturation). All 15 combinations of the contaminants were used. The order that motion artifact, power line interference, and ECG interference are applied does not affect the result. However, amplifier saturation, ADC clipping, and quantization noise must be applied after the former three contaminants. When a subset of the latter three contaminants were applied together, clipping was always applied after amplifier saturation and quantization noise was always applied very last. When a subset of motion artifact, power line interference, and ECG interference was used, SNR for the second contaminant applied was computed relative to the original uncontaminated signal, and not to the entire signal containing the first contaminant.

Each of the fifteen combinations of two of the six contaminants was applied at 9 noise levels per contaminants (for a total of 81 noise level combinations). The contaminated dataset was then used to test each of the 14 SVMs.

The results of the simulations reflected similar transition points for each contaminant pair as to when the contaminants were applied in isolation. In some cases, the transition point for a given contaminant would occur with a smaller amount of noise due to the combined impact of that contaminant and the second contaminant, even though the noise level of the second contaminant may have been after its own transition point (i.e., the second contaminant would not have been detected on its own with good accuracy). For instance, ECG artifact is not detected on its own with reliable accuracy at 0 dB (approximately 60% from Figure 7-5), however when combined with quantization noise the accuracy reaches 80% at a step

size of  $2^{-4.5}$  and almost 100% at  $2^{-4}$ , indicating a wider range of detection than when ECG interference was not present (Figure 7-18).

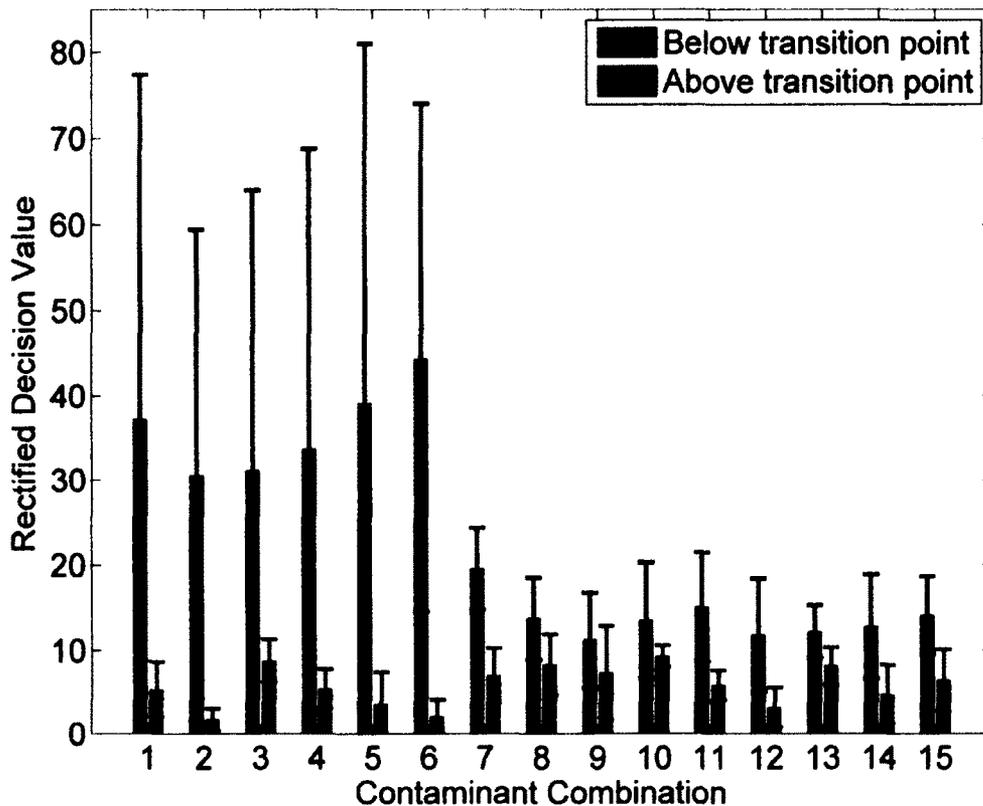


**Figure 7-18: Classification accuracy for real sEMG both with and without ECG interference at 0 dB and quantization noise at various ADC step sizes. Error bars are at  $\pm 1$  standard deviation.**

Figure 7-19 is a plot of the decision values for each of the 15 combinations of contaminants on either side of the transition point. Note that if multiple points were below the transition point and still at high accuracy, the closest point to 60% accuracy was chosen. The contaminants are listed below in Table 7-5 along with their respective noise levels both below and above the transition point.

**Table 7-5: Noise combinations used in Figure 7-19 with corresponding noise levels below and above the transition point for each contaminant.**

Combination	First Contaminant	Noise level		Second Contaminant	Noise level	
		Below	Above		Below	Above
1	Motion artifact	5 dB	10 dB	Power line interference	20 dB	10 dB
2	Motion artifact	5 dB	10 dB	ECG interference	15 dB	5 dB
3	Motion artifact	5 dB	10 dB	Quantization noise	ADC step: $2^{-6}$	ADC step: $2^{-4}$
4	Motion artifact	5 dB	10 dB	Clipping	ADC max: 4.75	ADC max: 3.5
5	Motion artifact	5 dB	10 dB	Amplifier saturation	Gain: 1.5	Gain: 3
6	Power line interference	10 dB	10 dB	ECG interference	-5 dB	5 dB
7	Power line interference	10 dB	10 dB	Quantization noise	ADC step: $2^{-3.5}$	ADC step: $2^{-4.5}$
8	Power line interference	10 dB	10 dB	Clipping	ADC max: 3.25	ADC max: 3.5
9	Power line interference	10 dB	10 dB	Amplifier saturation	Gain: 3.5	Gain: 3
10	ECG interference	0 dB	5 dB	Quantization noise	ADC step: $2^{-4.5}$	ADC step: $2^{-4}$
11	ECG interference	0 dB	5 dB	Clipping	ADC max: 4	ADC max: 3.5
12	ECG interference	0 dB	5 dB	Amplifier saturation	Gain: 2.5	Gain: 3
13	Clipping	ADC max: 3.25	ADC max: 3.5	Quantization noise	ADC step: $2^{-6.5}$	ADC step: $2^{-4.5}$
14	Amplifier saturation	Gain: 3.5	Gain: 3	Quantization noise	ADC step: $2^{-4}$	ADC step: $2^{-4.5}$
15	Amplifier saturation	Gain: 3	Gain: 3	Clipping	ADC max: 3	ADC max: 3.25



**Figure 7-19: Average rectified decision values for each contaminant combination from Table 7-5 below and above the transition point.**

The results from Figure 7-19 show that combinations which involve entire additive signals (such as power line interference, motion artifact, or ECG interference, i.e., combinations 1-6) are much better discriminated by the SVM. Significant distortion is apparent after visual examination of the signal in these cases. For quantization noise, ADC clipping, and amplifier saturation, the decision values are more clustered below and above the transition point. However, contamination is not evident from the visual examination of the time domain or frequency domain signal. Thus, the SVM classifier may offer a method for automatically detecting these types of contaminants, although it cannot identify the contaminants.

The decision values above the transition point are similar for Figure 7-17 and Figure 7-17 Figure 7-19. A candidate threshold decision value could be -10. Values below this could be identified as contaminated. The large error bars in some cases present the possibility of type II errors (false negatives). False positives could occur but they are more dependent on how well the SVM conforms to the class of clean sEMG. The confidence threshold could be adjusted to achieve a desired sensitivity and specificity, although more investigation is required here along with a much larger dataset.

## **7.8 Conclusion**

In this chapter, the use of a one-class SVM to differentiate clean and contaminated sEMG was evaluated. Results indicate that this classification is feasible and accuracy will depend on the intensity of the contamination, as well as the quality of the clean dataset. Motion artifact, power line interference, and ECG interference can be detected with high accuracy, although detection is also possible in many cases by visually examining the signal or frequency spectrum. Detection of quantization noise, clipping, and amplifier saturation is possible and the SVM decision function value can be used as a confidence indicator and compared to a pre-calibrated threshold. This provides an advantage in terms of automation and when visual examination of the signal is insufficient to detect contamination. The presence of multiple contaminants in the signal will often increase the classification accuracy at a given noise level. More investigation with a larger training dataset could lead to a robust SVM classifier with a reliable confidence threshold.

## 8 Thesis Summary and Future Recommendations

### 8.1 Summary

In this thesis, methods were presented and evaluated in terms of their accuracy in signal quality assessment in surface electromyography (sEMG). This thesis contributed methods to identify, quantify, and mitigate power line interference; identify, and quantify analog-to-digital converter clipping and quantization noise; identify amplifier saturation; quantify and mitigate electrocardiogram artifact; and quantify motion artifact. These methods were tested on real and simulated sEMG artificially contaminated with either synthetic noise, or noise obtained from [54], [55] at different signal-to-noise ratios. Results indicate that power line interference could be mitigated at up to 20 dB SNR for signals 5 seconds in length or more. Clipping could be detected at levels well below that when signal quality becomes compromised. Quantization noise could be quantified for signals 5 seconds in length or more at 16-bit resolution. Amplifier saturation could be detected when a significant amount ( $> 40\%$ ) of the signal is outside the linear range of the amplifier. ECG interference can be mitigated using a moving average method. A moving average can also be used to quantify motion artifact up to 10 dB.

A procedure to train a one-class support vector machine (SVM) to differentiate clean and contaminated sEMG (signal quality assessment from the top-down) was also investigated. The accuracy of the SVM was evaluated for different levels of contamination. The SVM could reject power line interference, motion artifact, ECG interference, quantization noise, ADC clipping, and amplifier saturation depending on the level of contamination and the SVM decision function value could be used as a confidence indicator. Detection of quantization noise, clipping, and amplifier saturation is possible even when no contamination is evident by visually examining the signal.

The contributions in this thesis are towards a reliable open-source user-friendly software tool for automatic signal quality assessment in sEMG.

## ***8.2 Recommendations for future work***

Future work should address the following items.

### **1. Investigation of baseline wander and crosstalk**

Identification, quantification, and mitigation of baseline wander and crosstalk were introduced but not a focus in this thesis. Baseline wander is a form of low frequency contamination that could be mitigated in a similar manner as motion artifact (using a moving average or a related high-pass filtering method). Crosstalk is a difficult contaminant to mitigate without placing restrictions on the sEMG acquisition setup, however an investigation could utilize the increased selectivity of a double-differential electrode configuration to collect sEMG with and without crosstalk. The electrode transfer function would need to be taken into account. A procedure would need to be developed where an individual muscle could be recorded without crosstalk as well as its impact on a neighbouring muscle. Some correlation should be observed with an increasing contraction level of the interfering muscle.

### **2. Contaminant differentiation**

Further investigation should be performed to differentiate one contaminant from another. A procedure should be in place to identify each contaminant unambiguously and rule out the rest. All contaminant identification methods should be tested on all forms of contamination to ensure they correctly identify the contaminant of interest. Further testing could then be done by combining contaminants and ensuring they can still be differentiated. In some cases, this may not be possible due to significant overlap (e.g., motion artifact and electrocardiogram interference).

### **3. SVM classification**

More investigation is required in training the one-class SVM classifier with a larger dataset of clean sEMG. Such an investigation could facilitate the selection of a confidence decision value threshold to achieve a range of desired sensitivities and specificities depending on the application. A larger, more comprehensive dataset could help to identify signals which are contaminated (outliers) with greater certainty.

### **4. Impact of contamination on sEMG features**

This thesis examines the effect of signal contamination on the raw signal only and not on features such (e.g., mean absolute value and median frequency) which are more useful for many sEMG applications. An investigation should be conducted where the impact of signal contamination on different sEMG features is examined.

### **5. Larger dataset**

Each method in this thesis should be validated using a larger dataset of real sEMG. Furthermore, methods evaluated using additive pre-recorded contaminants (i.e., ECG and motion artifact) should be evaluated with a larger database of those contaminants. Additional ECG artifact should be collected from more subjects under different conditions, as well as additional recordings of motion artifact.

# Appendix A: Least Squares Adaptive Power Line Estimation Algorithm Derivation

Appendix A includes a derivation of the least squares adaptive power line estimation algorithm presented in section 4.2 and adapted from [42]. The derivation shows how to compute the derivative of the mean squared error objective function between an sEMG signal contaminated with power line interference and an estimate of that power line interference. The minimum of the MSE between the noise and the noise estimate can be located iteratively by moving in small steps in the opposite direction (steepest descent) to the derivative. First, as the power line interference in the signal cannot be measured directly, we need a method for estimating the error between the noise and the noise estimate.

## A.1 Computing the Error Function for the Noise Estimate

The noisy sEMG signal  $x[n]$  is the sum of the clean signal  $s[n]$  and the noise  $m[n]$

$$x[n] = s[n] + m[n] \quad (\text{A-1})$$

The noise is assumed to be purely due to power line interference and of the form,

$$\hat{m}[n] = \hat{a} \cos(\hat{\omega}n) + \hat{b} \sin(\hat{\omega}n) \quad (\text{A-2})$$

where  $\hat{m}[n]$  is the noise estimate,  $\hat{a}$  and  $\hat{b}$  are the amplitude estimates of the in-phase and quadrature components, and  $\hat{\omega}$  is the power line frequency.

The squared error between the signal and the noise estimate is

$$e^2 = (x[n] - \hat{m}[n])^2.$$

The expectation of the squared error (the MSE) is as follows.

$$E[e^2] = E[(x[n] - \hat{m}[n])^2]$$
$$E[e^2] = E[(s[n] + m[n] - \hat{m}[n])^2]$$

$$E[e^2] = E[s^2[n]] + E[(m[n] - \hat{m}[n])^2] + 2E[s[n](m[n] - \hat{m}[n])]$$

We will assume that the clean signal  $s[n]$  is uncorrelated with the difference in noise and noise estimate  $m[n] - \hat{m}[n]$ .

$$E[e^2] = E[s^2[n]] + E[(m[n] - \hat{m}[n])^2] + 2E[s[n]]E[(m[n] - \hat{m}[n])]$$

We will further assume that  $\hat{m}[n]$  is an unbiased estimator for  $m[n]$  so  $E[(m[n] - \hat{m}[n])] = 0$ .

$$E[e^2] = E[s^2[n]] + E[(m[n] - \hat{m}[n])^2]$$

The minimum of the MSE is then defined as follows.

$$\min(E[e^2]) = \min(E[s^2[n]]) + \min(E[(m[n] - \hat{m}[n])^2])$$

Since the signal  $s[n]$  is being measured, the minimum MSE between noise and the noise estimate  $\hat{m}[n]$  is achieved by minimizing the MSE between the signal  $x[n]$  and the noise estimate. This tells us that the best estimate we have for  $x[n]$  is  $\hat{m}[n]$ .

$$\hat{x}[n] = \hat{m}[n] \tag{A-3}$$

The error function to minimize using steepest descent is:

$$E = \frac{1}{N} \sum_{n=0}^{N-1} (x[n] - \hat{x}[n])^2. \tag{A-4}$$

## A.2 Computing the Derivative of the Error Function

Let  $x = x[n]$  where  $x = [x[0] \ x[1] \ \dots \ x[N-1]]^T$ .

From  $\hat{m}[n] = \hat{a} \cos(\hat{\omega}n) + \hat{b} \sin(\hat{\omega}n)$  (A-2) and  $\hat{x}[n] = \hat{m}[n]$  (A-3),

$$\hat{x}[n] = \hat{a} \cos(\hat{\omega}n) + \hat{b} \sin(\hat{\omega}n)$$

Let  $\hat{x} = \hat{x}[n]$  where  $\hat{x} = [\hat{x}[0] \ \hat{x}[1] \ \dots \ \hat{x}[N-1]]^T$ .

We can rewrite (A-4) as follows. Note that the  $\frac{1}{N}$  is a constant so it will be dropped since minimizing the total squared error is equivalent to minimizing the MSE.

$$E = \|x - \hat{x}\|_2^2$$

$$E = x^T x + \hat{x}^T \hat{x} - 2\hat{x}^T x \tag{A-5}$$

Let  $h = [\hat{a} \ \hat{b}]^T$  and  $Q = [\cos(\hat{\omega}n) \ \sin(\hat{\omega}n)]$ , then

$$\hat{x} = Qh \tag{A-6}$$

Combining (A-5) and A-6),

$$E = x^T x + h^T Q^T Q h - 2h^T Q^T x \quad (\text{A-7})$$

Since (A-7) is linear in terms of  $h$ , we can find the optimal  $h$  as follows.

$$\frac{\partial E}{\partial h} = 2Q^T Q h - 2Q^T x$$

Let  $h = h_{opt}$  when  $\frac{\partial E}{\partial h} = 0$ .

$$2Q^T Q h_{opt} - 2Q^T x = 0$$

$$h_{opt} = (Q^T Q)^{-1} Q^T x \quad (\text{A-8})$$

The minimum error can then be found from (A-7) and (A-8).

$$E_{min} = x^T x + h_{opt}^T Q^T Q h_{opt} - 2h_{opt}^T Q^T x \quad (\text{A-9})$$

Combining (A-8) and (A-9) and noting that,  $h_{opt}^T = x^T Q (Q^T Q)^{-1}$ ,

$$E_{min} = x^T x + x^T Q (Q^T Q)^{-1} Q^T Q (Q^T Q)^{-1} Q^T x - 2x^T Q (Q^T Q)^{-1} Q^T x$$

$$E_{min} = x^T x + x^T Q (Q^T Q)^{-1} Q^T x - 2x^T Q (Q^T Q)^{-1} Q^T x$$

$$E_{min} = x^T x - x^T Q (Q^T Q)^{-1} Q^T x \quad (\text{A-10})$$

Taking the derivative of (A-9) with respect to  $\hat{\omega}$  where,

$$Q' = \frac{d}{d\hat{\omega}} Q$$

$$Q' = [-n \sin \hat{\omega} n \quad n \cos \hat{\omega} n]$$

$$\frac{\partial}{\partial \hat{\omega}} E_{min} = h_{opt}^T Q'^T Q h_{opt} + h_{opt}^T Q^T Q' h_{opt} - 2h_{opt}^T Q'^T x \quad (\text{A-11})$$

We notice that,

$$h_{opt}^T Q'^T Q h_{opt} = [a_{opt} \quad b_{opt}] [-n \sin \hat{\omega} n \quad n \cos \hat{\omega} n]^T [\cos(\hat{\omega} n) \quad \sin(\hat{\omega} n)] [a_{opt} \quad b_{opt}]^T$$

$$h_{opt}^T Q^T Q' h_{opt} = (-n a_{opt} \sin \hat{\omega} n + n b_{opt} \cos \hat{\omega} n) (a_{opt} \cos(\hat{\omega} n) + b_{opt} \sin(\hat{\omega} n))$$

The value  $h_{opt}^T Q'^T Q h_{opt}$  is a scalar. This makes sense since  $\frac{\partial}{\partial \hat{\omega}} E_{min}$  must also be a scalar. This means that

$$h_{opt}^T Q'^T Q h_{opt} = [h_{opt}^T Q'^T Q h_{opt}]^T$$

$$h_{opt}^T Q'^T Q h_{opt} = h_{opt}^T Q^T Q' h_{opt} \quad (\text{A-12})$$

Substituting (A-12) into (A-11),

$$\frac{\partial}{\partial \hat{\omega}} E_{\min} = 2h_{opt}^T Q^T Q h_{opt} - 2h_{opt}^T Q^T x$$

$$\frac{\partial}{\partial \hat{\omega}} E_{\min} = 2h_{opt}^T Q^T (Q h_{opt} - x)$$

We can apply the transpose once again here since  $\frac{\partial}{\partial \hat{\omega}} E_{\min}$  is a scalar.

$$\frac{\partial}{\partial \hat{\omega}} E_{\min} = -2(x - Q h_{opt})^T Q^T h_{opt} \quad (\text{A-13})$$

(A-13) can then be applied using the iterative descent update rule for a given learning rate,  $\mu$ .

$$\hat{\omega}_{next} = \hat{\omega} - \frac{1}{2} \mu \frac{\partial}{\partial \hat{\omega}} E_{\min} \quad (\text{A-14})$$

## Appendix B: Moving Average Frequency Response Derivation

Appendix B lists the steps in the derivation of the frequency response presented in (6-3) in Section 6.2. A moving average can be implemented as a finite impulse response filter with  $L$  coefficients each equal to  $1/L$ . The impulse response of this filter is:

$$h(n) = \frac{1}{L} \sum_{m=0}^{L-1} \delta(n-m) \quad (\text{B-1})$$

The transfer function can then be written as:

$$H(z) = \frac{1}{L} (1 + z^{-1} + z^{-2} + \dots + z^{-(L-1)}) \quad (\text{B-2})$$

And the frequency response:

$$H(e^{j\omega}) = \frac{1}{L} (1 + e^{-j\omega} + e^{-j2\omega} + \dots + e^{-j(L-1)\omega}) \quad (\text{B-3})$$

(B-3) is the sum of a geometric series with a ratio of  $e^{-j\omega}$ . This sum is equivalent to:

$$H(e^{j\omega}) = \left(\frac{1}{L}\right) \frac{1-e^{-j\omega L}}{1-e^{-j\omega}}, \quad \omega \neq 2\pi k, \quad k \in \mathbb{Z} \quad (\text{B-4})$$

$$H(e^{j\omega}) = \left(\frac{1}{L}\right) \frac{e^{-j\omega(\frac{L}{2})} - e^{-j\omega(\frac{L}{2})}}{e^{j\frac{\omega}{2}} - e^{-j\frac{\omega}{2}}} \left(\frac{e^{-j\omega(\frac{L}{2})}}{e^{-j\frac{\omega}{2}}}\right), \quad \omega \neq 2\pi k, \quad k \in \mathbb{Z}$$

$$H(e^{j\omega}) = \left(\frac{1}{L}\right) e^{-j\omega\frac{L-1}{2}} \left(\frac{\sin(\frac{\omega L}{2})}{\sin(\frac{\omega}{2})}\right), \quad \omega \neq 2\pi k, \quad k \in \mathbb{Z} \quad (\text{B-5})$$

In (B-3), the sum evaluates to 1 if  $\omega$  is 0 or any multiple of  $2\pi$ . The result is (B-6) which is the same equation shown in (6-3).

$$H(e^{j\omega}) = \begin{cases} 1, & \omega = 2\pi k, \quad k \in \mathbb{Z} \\ e^{-j\omega\frac{L-1}{2}} \left(\frac{\sin(\frac{\omega L}{2})}{\sin(\frac{\omega}{2})}\right), & \omega \neq 2\pi k, \quad k \in \mathbb{Z} \end{cases} \quad (\text{B-6})$$

# Appendix C: Informed Consent Form

Appendix C contains a copy of the informed consent form that each participant was asked to read and sign. This form was approved by the Carleton University Research Ethics Board. The date of ethics approval was June 16, 2003 and this clearance expires on August 31, 2012.

## Biological Signals Research Informed Consent



Dr. Adrian D.C. Chan  
Associate Professor  
Dept Systems & Computer Engineering  
1125 Colonel By Drive  
Ottawa ON K1S 5B6

I, \_\_\_\_\_ have been invited by Dr. Adrian Chan, and his research associates, of the Department of Systems and Computer Engineering at Carleton University, Ottawa, ON to participate in a study on biological signals.

The purpose of the study is to examine various biological signals including:

1. electrocardiogram (ECG) – heart signals
2. myoelectric signal (MES) – muscle signals
3. electroneurogram (ENG) – nerve signals
4. electroencephalogram (EEG) – brain signals
5. acoustic sound – speech for example
6. phonocardiogram (PCG) – heart sounds
7. RadioVibrometer – anatomical vibrations

These signals can be monitored non-invasively using electrodes on the skin surface. While being non-invasive, these signals contain valuable information pertaining to the status of internal organs (e.g. heart, muscles, nerves). These signals are used in a variety of areas including: disease diagnosis, myoelectrically controlled prosthesis, monitoring muscle fatigue, and expanding physiological knowledge of our bodies.

Research undertaken as part of a graduate/undergraduate course requirement:

YES  NO

If YES specify course name and course number

Course Name	
Course Number	

Research undertaken as part of a graduate or undergraduate thesis research:

YES  NO

## Measurement Equipment and Risks

### **Bioelectric signals**

During the data collection stage, surface electrodes with conductive gel will be placed on the subject to acquire the necessary biological signals. Depending on the study, the number and exact placement of the electrodes will vary. A maximum of 16 simultaneous channels will be used.

**Risks:** The application of surface electrodes causes an associated risk that is similar to that of obtaining an electrocardiogram (cardiac) record (i.e. slight irritation of the skin under the electrode site).

Bioelectric signals will be used in this study: YES  NO

If YES, I acknowledge this requires the application of surface electrodes

## **Phonocardiogram**

If the phonocardiogram (PCG) is required, this will be obtained using an electronic stethoscope. This stethoscope is the same as you would find used by a medical doctor, except that it provides an electronic audio output connection that enables recording.

**Risks:** None.

Phonocardiogram will be used in this study: YES  NO

If YES, I acknowledge this requires the use of a stethoscope

## **Acoustic signals**

The acoustic signal may also be required in some speech related studies. In these circumstances, a microphone will be used to record the participant's voice.

**Risks:** None.

Acoustic signals will be used in this study: YES  NO

If YES, I acknowledge this requires the use of a microphone and may record my voice

## **RadioVibrometer**

The RadioVibrometer (Aliph, San Francisco CA) is used to detect anatomical vibrations, typically used in research associated with the production of speech (e.g. detecting vibrations near the trachea). The RadioVibrometer operates in a similar fashion to a radar, emitting a very low power (less than a milliwatt of power or 1/1000 that of many mobile phones) and detecting the return signal. It is safe for continuous use under current FDA and FCC RF regulations. Use of the RadioVibrometer requires the application of electrodes on the skin surface.

**Risks:** There is always a risk associated with the emission of electromagnetic waves; however, the RadioVibrometer uses a very low emission power (less than a milliwatt of power or 1/1000 that of many mobile phones). It is safe for continuous use under current FDA and FCC RF regulations.

RadioVibrometer will be used in this study: YES  NO

If YES, I acknowledge this requires the use of the RadioVibrometer

### **Electrical stimulation**

In some research, it may be required to elicit a nerve or muscle response. This will be accomplished using applied electrical stimulation. The electrical stimulus will initially be applied at the minimum level and slowly increased, in order to ensure the subject's safety and comfort. Stimulation will require the application of a pair of surface electrodes with conductive gel.

**Risks:** With electrical stimuli applied to the subject, there is a chance of electrocution. This risk has been minimized by using an electrical isolation unit, and by increasing the electrical stimuli slowly from the minimum level. The subject may feel slight discomfort if when a high stimulus level is used. The sensation is similar to some striking the stimulation site with a flick of a finger.

Electrical stimulation will be used in this study: YES  NO

If YES, I acknowledge this requires the application of surface electrodes and electrical stimuli.

### **Procedures**

I may be asked some questions to ascertain some additional data, which may include:

1. age
2. sex
3. height
4. weight

5. native language
6. birth place
7. whether they have any known neuromuscular disorders

I have been fully informed of the study that I am participating in: YES

I have been informed in advance that the study session for the data collections will be:

approximately one hour

up to three hours

I may be asked to return for additional sessions, and will be informed of this at the end of the study session. I am in no way obligated to participate in these additional sessions.

**Should the experimenter note any unusual readings during the course of the experiment the study will be stopped immediately. The experimenter is not a physician and cannot make a medical diagnosis. I will be asked to contact my family physician. The researcher will contact my physician in writing explaining why the experiment was stopped. I may not return to the study or undertake any further experiments without the written consent of my physician.**

### Benefits

There are no direct benefits or remuneration for my participation in this study.

### Confidentiality

My identity will be kept strictly confidential unless otherwise discussed with the researcher. Any scientific report, presentation, or publication of the data will refer to me using a subject number. Information on my gender, age, and the presence of known neuromuscular disorders may be used.

I understand that the data from this research will be used in scientific reports, presentations, and publications and my identity will remain confidential. Data will be kept electronically and my consent forms and information will be kept as a hardcopy. Access to the data will be restricted to the researcher investigators. Data may be shared with research associates at other academic institutions. Data may be kept for an indefinite period of time.

### Ethical review

This study has been reviewed and received ethics clearance through the Carleton University Research Ethics Committee in accordance to the Tri-Council Policy Statement for Ethical Conduct for Research Involving Humans. Should I have any concerns or questions about my involvement in this study I may contact the committee chair:

Professor Antonio Gualtieri, Chair  
Carleton University Research Ethics Committee  
Office of Research Services  
Carleton University  
1125 Colonel By Drive  
Ottawa, Ontario K1S 5B6  
Tel: 613-520-2517  
E-mail: ethics@carleton.ca

### Withdrawal

Participation in this study is strictly voluntary. I am free to withdraw from the experiment at any time and without any consequences. I will also declare if the researcher can/cannot use the data I have provided should I decide to withdraw from the study.

### Consent for Participation in the Study

I hereby agree to participate in this study and consent to the use of this research data in scientific reports, presentations, and publications with the understanding that my identity will remain confidential. I have read and understand the above explanation of the research

procedure and all my questions have been answered to my satisfaction. I understand that I am free to withdraw from this research at any time and without any consequence.

Participant		Signature	
Experimenter		Signature	
Date			

## Biological Signals Research Consent for Photographs



Dr. Adrian D.C. Chan  
Associate Professor  
Dept Systems & Computer Engineering  
1125 Colonel By Drive  
Ottawa ON K1S 5B6

I understand that photographs (conventional/digital) may be required to document portions of the study. I have the option to consent or decline the usage of photographs during the study. Usage of these photographs will be restricted to scientific reports, presentations, or publications. I understand that my confidentiality will be maintained as best as possible; however, the usage of photographs may compromise my confidentiality.

Photographs will be required in this study: YES  NO

If YES photographs will be: CONVENTIONAL  DIGITAL

I, \_\_\_\_\_ hereby consent to having photographs taken.

Participant		Signature	
Experimenter		Signature	
Date			

# References

- [1] A. D. C. Chan and D. MacIsaac, "CleanEMG: Assessing the quality of EMG signals," in *Conf. Med. Biol. Eng. Soc. and FICCDAT*, 2011, pp. 1-4.
- [2] J.-Y. Hogrel, "Clinical applications of surface electromyography in neuromuscular disorders," *Clinical Neurophysiology*, vol. 35, pp. 59-71, 2005.
- [3] C. Sinderby, L. Lindström, and A. E. Grassino, "Automatic assessment of electromyogram quality," *Journal of Applied Physiology*, vol. 79, no. 5, pp. 1803-15, Nov. 1995.
- [4] E. V. S. J. V. Trontelj, "Multiple Innervation of Muscle Fibers in Myasthenia Gravis," *Muscle & Nerve*, vol. 18, no. 2, pp. 224-228, 1995.
- [5] T. Masuda, H. Miyano, and T. Sadoyama, "The Position of Innervation Zones in the Biceps Brachii Investigated by Surface Electromyography," *Biomedical Engineering, IEEE Transactions on*, vol. BME-32, no. 1, pp. 36-42, 1985.
- [6] R. Merletti and P. Parker, *Electromyography - Physiology, Engineering, and Noninvasive Applications*. 1999.
- [7] V. T. Inman, H. J. Ralston, J. B. D. C. M. Saunders, M. B. B. Feinstein, and E. W. W. Jr., "Relation of human electromyogram to muscular tension," *Electroencephalography and Clinical Neurophysiology*, vol. 4, no. 2, pp. 187-194, 1952.
- [8] M.-I. Lin, H.-W. Liang, K.-H. Lin, and Y.-H. Hwang, "Electromyographical assessment on muscular fatigue—an elaboration upon repetitive typing activity," *Journal of Electromyography and Kinesiology*, vol. 14, pp. 661-669, 2004.
- [9] P. Konrad, *The ABC of EMG - A Practical Introduction to Kinesiological Electromyography*. 2005.
- [10] D. Farina and A. Rainoldi, "Compensation of the effect of sub-cutaneous tissue layers on surface EMG: a simulation study," *Medical Engineering & Physics*, vol. 21, no. 6-7, pp. 487-497, 1999.
- [11] R. Merletti, G. Rau, C. Disselhorst-Klug, D. F. Stegerman, and G. M. Hägg, "SENIAM." [Online]. Available: <http://www.seniam.org>. [Accessed: 05-Dec-2011].
- [12] A. Searle and L. Kirkup, "A direct comparison of wet, dry and insulating bioelectric recording electrodes," *Physiological Measurement*, vol. 21, no. 2, p. 271, 2000.
- [13] S. Day, *Important Factors in Surface EMG Measurement*. Bortec Biomedical Ltd., 2002.
- [14] R. Betts and B. Brown, "Method for recording electrocardiograms with dry electrodes applied to unprepared skin," *Medical and Biological Engineering and Computing*, vol. 14, no. 3, pp. 313-315, 1976.

- [15] P. Zipp, "Effect of electrode parameters on the bandwidth of the surface e.m.g. power-density spectrum," *Medical and Biological Engineering and Computing*, vol. 16, no. 5, pp. 537-541, 1978.
- [16] G. E. Loeb and C. Gans, *Electromyography for Experimentalists*. University of Chicago Press, 1986, pp. 66-70.
- [17] T. J. Koh and M. D. Grabiner, "Cross talk in surface electromyograms of human hamstring muscles," *Journal of Orthopaedic Research*, vol. 10, no. 5, pp. 701-709, 1992.
- [18] H. J. Landau, "Sampling, data transmission, and the Nyquist rate," *Proceedings of the IEEE*, vol. 55, no. 10, pp. 1701-1706, 1967.
- [19] T. Yamamoto and Y. Yamamoto, "Electrical properties of the epidermal stratum corneum," *Medical and Biological Engineering and Computing*, vol. 14, no. 2, pp. 151-158, 1976.
- [20] H. Tam and J. G. Webster, "Minimizing Electrode Motion Artifact by Skin Abrasion," *Biomedical Engineering, IEEE Transactions on*, vol. BME-24, no. 2, pp. 134-139, Mar. 1977.
- [21] T. Yamamoto and Y. Yamamoto, "Analysis for the change of skin impedance," *Medical and Biological Engineering and Computing*, vol. 15, no. 3, pp. 219-227, 1977.
- [22] J. Rosell, J. Colominas, P. Riu, R. Pallas-Areny, and J. G. Webster, "Skin impedance from 1 Hz to 1 MHz.," *IEEE transactions on bio-medical engineering*, vol. 35, no. 8, pp. 649-51, Aug. 1988.
- [23] J. G. Webster, "Reducing Motion Artifacts and Interference in Biopotential Recording," *Biomedical Engineering, IEEE Transactions on*, vol. BME-31, no. 12, pp. 823-826, 1984.
- [24] A. B. Simakov and J. G. Webster, "Motion Artifact from Electrodes and Cables," *Iranian Journal of Electrical and Computer Engineering*, vol. 9, no. 2, pp. 139-143, 2010.
- [25] J. C. Huhta and J. G. Webster, "60-HZ interference in electrocardiography.," *IEEE transactions on bio-medical engineering*, vol. 20, no. 2, pp. 91-101, Mar. 1973.
- [26] P. Zhou, B. Lock, and T. A. Kuiken, "Real time ECG artifact removal for myoelectric prosthesis control," *Physiol. Meas.*, vol. 28, no. 4, 2007.
- [27] J. P. P. van Vugt and J. G. van Dijk, "A convenient method to reduce crosstalk in surface EMG," *Clinical Neurophysiology*, vol. 112, no. 4, pp. 583-592, 2001.
- [28] J. M. Łęski and N. Henzel, "ECG baseline wander and powerline interference reduction using nonlinear filter bank," *Signal Processing*, vol. 85, no. 4, pp. 781-793, 2005.
- [29] D. H. Gordon, "Triboelectric Interference in the ECG," *Biomedical Engineering, IEEE Transactions on*, vol. BME-22, no. 3, pp. 252-255, May 1975.

- [30] A. A. Eltholth, A. R. Mekhail, A. Elshirbini, M. I. Dessouki, and A. I. Abdelfattah, "Modeling the Effect of Clipping and Power Amplifier Non-Linearities on OFDM Systems," *Ubiquitous Computing and Communication Journal*, vol. 3, no. 1, pp. 54-59.
- [31] E. Shwedyk, R. Balasubramanian, and R. N. Scott, "A nonstationary model for the electromyogram.," *IEEE transactions on bio-medical engineering*, vol. 24, no. 5, pp. 417-24, Sep. 1977.
- [32] A. Arvidsson, A. Grassino, and L. Lindstrom, "Automatic selection of uncontaminated electromyogram as applied to respiratory muscle fatigue," *Journal of Applied Physiology*, vol. 56, no. 3, pp. 568-575, 1984.
- [33] C. Grönlund, K. Roeleveld, A. Holtermann, and J. Karlsson, "On-line signal quality estimation of multichannel surface electromyograms," *Medical and Biological Engineering and Computing*, vol. 43, no. 3, pp. 357-364, 2005.
- [34] H. Marateb, M. Rojas-Martínez, M. Mansourian, R. Merletti, and M. Mañanas Villanueva, "Outlier detection in high-density surface electromyographic signals," *Medical and Biological Engineering and Computing*, vol. 50, no. 1, pp. 79-89, 2012.
- [35] *Grass Technologies 15 & 15LT Amplifier Systems User & Service Manual Release C.* 2012, pp. 4-56.
- [36] D. T. Mewett, K. J. Reynolds, and H. Nazeran, "Reducing power line interference in digitised electromyogram recordings by spectrum interpolation.," *Medical & biological engineering & computing*, vol. 42, no. 4, pp. 524-31, Jul. 2004.
- [37] R. V. Baratta, M. Solomonow, B. H. Zhou, and M. Zhu, "Methods to reduce the variability of EMG power spectrum estimates.," *Journal of Electromyography and Kinesiology*, vol. 8, no. 5, pp. 279-85, Oct. 1998.
- [38] G. Filligoi and F. Felici, "Detection of hidden rhythms in surface EMG signals with a non-linear time-series tool.," *Medical engineering & physics*, vol. 21, no. 6-7, pp. 439-48, 1999.
- [39] Y. U. Gang, L; Ling, L; Qilian, Y; Xuemin, "A new adaptive coherent model algorithm for removal of power-line interference," *Journal of Clinical Engineering*, vol. 20, pp. 147-150, 1995.
- [40] Y.-der Lin, H.-hsun Huang, and F.-ching Chong, "Adaptive cancellation of power-line interference from biopotential measurement," *Proceedings of the 20th Annual International Conference of the IEEE Engineering in Medicine and Biology Society. Vol.20 Biomedical Engineering Towards the Year 2000 and Beyond (Cat. No.98CH36286)*, vol. 20, no. 3, pp. 1643-1644, 1998.
- [41] B. Widrow et al., "Adaptive noise cancelling: Principles and applications," *Proceedings of the IEEE*, vol. 63, no. 12, pp. 1692-1716, 1975.

- [42] Y. A. Mahgoub and R. M. Dansereau, "Time Domain Method for Precise Estimation of Sinusoidal Model Parameters of Co-Channel Speech," *Research Letters in Communications*, vol. 2008, pp. 1-5, 2008.
- [43] E. A. Clancy, E. L. Morin, and R. Merletti, "Sampling, noise-reduction and amplitude estimation issues in surface electromyography," *Journal of Electromyography and Kinesiology*, vol. 12, no. 1, pp. 1-16, 2002.
- [44] B. P. Lathi, *Modern Digital and Analog Communication Systems (Third Edition)*, 3rd ed. Oxford University Press, 1998.
- [45] J. O. Smith, "Round-Off Error Variance," in *Mathematics of the Discrete Fourier Transform (DFT) with Audio Applications, Second Edition*, 2007.
- [46] J. Filliben, "The Probability Plot Correlation Coefficient Test for Normality," *Technometrics*, vol. 17, no. 1, pp. 111-117, 1975.
- [47] I. Cohen, Y. Huang, J. Chen, and J. Benesty, "Pearson Correlation Coefficient," in *Noise Reduction in Speech Processing*, vol. 2, Springer Berlin Heidelberg, 2009, pp. 1-4.
- [48] P. Zhou, M. Lowery, R. Weir, and T. Kuiken, "Elimination of ECG Artifacts from Myoelectric Prosthesis Control Signals Developed by Targeted Muscle Reinnervation," in *27th Annual International Conference of the IEEE-EMBS*, 2005.
- [49] A. Bartolo, R. R. Dzwonczyk, C. Roberts, and E. Goldman, "Description and validation of a technique for the removal of ECG contamination from diaphragmatic EMG signal," *Med. Biol. Eng. Comput.*, vol. 34, no. 1, pp. 76-81, 1996.
- [50] J. D. M. Drake and J. P. Callaghan, "Elimination of electrocardiogram contamination from electromyogram signals: an evaluation of currently used removal techniques," *J. Electromyogr. Kinesiol.*, vol. 16, no. 2, pp. 175-187, 2006.
- [51] P. Akkiraju and D. C. Reddy, "Adaptive cancellation technique in processing myoelectric activity of respiratory muscles," *Transactions on Biomedical Engineering*, vol. 39, no. 6, 1992.
- [52] R. Bloch, "Subtraction of electrocardiographic signal from respiratory electromyogram," *Journal of applied physiology*, vol. 55, no. 2, 1983.
- [53] A. Bartolo, C. Roberts, R. R. Dzwonczyk, and E. Goldman, "Analysis of diaphragm EMG signals: comparison of gating vs. subtraction for removal of ECG contamination," *J. Appl. Physiol.*, vol. 80, pp. 1898-1902, 1996.
- [54] G. B. Moody, W. E. Muldrow, and R. G. Mark, "A noise stress test for arrhythmia detectors," *Computers in Cardiology*, vol. 11, pp. 381-384, 1984.
- [55] A. L. Goldberger et al., "{PhysioBank, PhysioToolkit, and PhysioNet}: Components of a New Research Resource for Complex Physiologic Signals," *Circulation*, vol. 101, no. 23, pp. 215-220.

- [56] M. Yousef, S. Jung, L. C. Showe, and M. K. Showe, "Learning from positive examples when the negative class is undetermined--microRNA gene identification.," *Algorithms for molecular biology : AMB*, vol. 3, p. 2, Jan. 2008.
- [57] C.-C. Chang and C.-J. Lin, "{LIBSVM}: A library for support vector machines," *ACM Transactions on Intelligent Systems and Technology*, vol. 2, no. 3, pp. 1-27, 2011.
- [58] B. Schölkopf, J. C. Platt, J. Shawe-Taylor, A. J. Smola, and R. C. Williamson, "Estimating the Support of a High-Dimensional Distribution," *Neural Computation*, vol. 13, no. 7, pp. 1443-1471, Jul. 2001.
- [59] R. E. Bellman, *Adaptive control processes: a guided tour*. 1961.
- [60] A. Phinyomark, C. Limsakul, and P. Phukpattaranont, "A Novel Feature Extraction for Robust EMG Pattern Recognition," pp. 71-80, Dec. 2009.