Multi-Parameter Modelling of Surface Electromyography Data

Open-Source Model for Surface EMG Simulations

Ahmad Alosta
and
Josef Djärf

2024

Master's Thesis in
Biomedical Engineering

Supervisor: Robin Rohlén
Assistant Supervisor: Jonathan Lundsberg

LUND
UNIVERSITY

Faculty of Engineering LTH
Department of Biomedical Engineering

Abstract

Surface electromyography (sEMG) measures skeletal muscle function by recording muscle activity from the surface of the skin. The technique can be used to diagnose neuromuscular diseases, and as an aid in rehabilitation, biomedical research, and human-computer interaction. A simulation model for sEMG data can assess decomposition algorithms and help develop new diagnostic tools. Such simulation models have previously not been available. We have written open-source code in Python to generate synthetic sEMG data. The code is publicly accessible via GitHub, an online platform for software development. The implemented model has multiple parameters that influence the artificially generated signal. The model was implemented with a bottom-up design, starting at a single muscle fibre and ending with the sEMG signal generated from up to hundreds of active motor units. The simulated signal can be recorded in potentially dozens of selectively positioned surface electrodes. The model's foundation is mathematical equations found throughout the scientific literature surrounding motor control and biological signalling, e.g., action potential propagation, membrane current distribution, and motor unit recruitment. We assert that the model incorporates the most significant features for generating sEMG data. The synthetically generated data was decomposed to study the simulated motor unit action potentials. The presented model can be used as ground truth to assess the performance of decomposition algorithms for sEMG. The analysis of sEMG signals can provide valuable insights into muscle activity, contributing to our understanding of motor control and aiding the development of prosthetics and assistive technologies.

Sammanfattning

Ytelektromyografi (sEMG) mäter skelettmuskelfunktionen genom att registrera muskelaktivitet från hudens yta. Tekniken kan användas för att diagnostisera neuromuskulära sjukdomar och som ett hjälpmedel vid rehabilitering, biomedicinsk forskning och för interaktion mellan människa och dator. En simuleringsmodell för sEMG-data kan bedöma avkodningsalgoritmer och hjälpa till att utveckla nya diagnostiska verktyg. Sådana simuleringsmodeller har tidigare inte varit tillgängliga. Vi har skrivit öppen källkod i Python för att generera syntetisk sEMG-data. Koden är tillgänglig via GitHub, en onlineplattform för mjukvaruutveckling. Den implementerade modellen har flera parametrar som påverkar den artificiellt genererade signalen. Modellen implementerades med en bottom-up-design, som börjar med en enda muskelfiber och slutar med sEMG-signalen genererad från upp till hundratals aktiva motoriska enheter. Den simulerade signalen kan registreras i potentiellt dussintals selektivt placerade ytelektroder. Modellens grund är matematiska ekvationer som finns i den vetenskapliga litteraturen kring motorisk kontroll och biologisk signalering, t.ex. aktionspotentialutbredning, membranströmfördelning och rekrytering av motoriska enheter. Vi hävdar att modellen innehåller de viktigaste funktionerna för att generera sEMG-data. Den syntetiskt genererade datan avkodades för att studera de simulerade motorenheternas aktionspotentialer. Den presenterade modellen kan användas som grundsanning för att bedöma prestandan av andra avkodningsalgoritmer för sEMG. Analysen av sEMG-signaler kan ge värdefulla insikter om muskelaktivitet, vilket bidrar till vår förståelse av motorisk kontroll och bidrar till utvecklingen av proteser och hjälpmedelsteknologier.

Preface

This thesis is the product of a completed Master's thesis, conducted at the Department of Biomedical Engineering at the Faculty of Engineering, Lund University. The project was completed in association with the Neuroengineering Group, under the examination of Dr. Christian Antfolk and the supervision of Dr. Robin Rohlén and Jonathan Lundsberg. The project was completed over the academic semester Aug 2023 − Jan 2024.

The project's main goal was to develop an open-source Python program for sEMG data simulation. The purpose of the project was to develop our programming ability and understanding of bio-signal processing. The model's foundation is based on mathematical equations and relevant theory found in the scientific literature. Two previous MATLAB programs, by (McDonald-Bowyer, Jager, Vanhoestenberghe, & Lancashire, 2020) and (Pah, 2003), influenced the implementation and design of early versions of the code. Additional MATLAB code snippets were provided by Dr Rohlén and Lundsberg.

We are thankful to Dr. Antfolk and the entire Department of Biomedical Engineering for allowing us the opportunity to complete our thesis in close connection to your ongoing research. We sincerely appreciate Dr. Rohlén and Lundsberg for mentoring us through the project. Thank you for investing your time in our project, your advice has been very helpful.

Table of Contents

List of Abbreviations

AP	Action potential
EMG	Electromyography
MCD	Membrane current distribution
MU	Motor unit
MUAP	Motor unit action potential
NMJ	Neuromuscular junction
sEMG	Surface electromyography
STA	Spike-triggered average

Mathematical Notation

A_f	Average area of a single fibre
A_k	Area encompassed by each motor unit territory
A_m	Muscle cross-sectional area
E_{max}	Maximum excitation
$FR_k(t)$	Firing rate response function
I_m	Membrane current distribution
K_a	Electrical conductivity ratio along and across the fibres
PFR_1	Peak firing rate first motor unit
P_i	Tripole amplitude of the pole i
P_k	Peak twitch force for the motor unit k
P_{total}	number of fibres required to exert one unit of force
V_m	Membrane action potential
g_e	Excitatory gain
nf_k	Number of fibres innervated by each motor unit
nf_{total}	Total number of fibres in the muscle
σ_r	Radial electrical conductivity

Φ_j	Potential on a plane skin surface at position (x, z) for the fibre j
A	Constant to fit the action potential amplitude
B	Resting membrane potential
C	Proportionality constant to fit the current distribution amplitude
$E(t)$	Excitatory drive function
MFR	Minimum firing rate
PFR	Peak firing rate of the motor unit
$PFRD$	Peak firing rate difference
RP	Range of the peak twitch forces
RR	Range of recruitment thresholds
RTE	Recruitment threshold excitation
a	Coefficient of recruitment threshold excitation assignment
b	Coefficient for twitch force
i	Pole index
j	Muscle fibre index
k	Current motor unit index
n	Number of motor units
x	Distance across the fibre
y	Depth of the fibre
z	Distance along the fibre
λ	Scaling factor
∂	Motor unit fibre density

Introduction

All voluntary movement of the body entails neuronal signals transmitted from the motor cortex in the central nervous system. The signal is transmitted via motor neurons to the peripheral nervous system and the muscle fibres in skeletal muscle. A collection of muscle fibres and the innervating motor neuron is called a motor unit (MU). The neuronal signal sent to a MU is called the action potential (AP), which constitutes the basis of electrophysiological recordings, such as electromyography (EMG) (Moritani, Stegeman, & Merletti, 2004). EMG is used extensively as a diagnostic tool for neuromuscular disease and as a research tool for studying locomotion and motor control. When using sEMG, electrodes are placed on the skin surface, registering the cumulative electrical potential from up to thousands of fibres and many MUs. The electrical signal can be decoded into neural information using digital processing tools, called decomposition algorithms. The sEMG signal can be used to determine MU activation, control prosthetic devices, or as a basis for human-computer interaction (Holobar, Farina, & Zazula, 2016).

This thesis presents a multiple-parameter model for synthetic sEMG data. Synthetically generated sEMG data can be used to develop diagnostic tools and progress the field of neuromuscular research (Farina, Mesin, Martina, & Merletti, 2004). The modelled data may also be used as ground truth when evaluating the performance of decomposition algorithms for EMG signal processing. Similar models have been described in the scientific literature but are not publicly available (Merletti, Lo Conte, Avignone, & Guglielminotti, 1999). We aim to contribute to the advancement of sEMG research, increase the availability of research-related code, and promote open-science, with a publicly available simulation model for sEMG data.

The purpose of the project was to develop our Python programming ability and deepen our understanding of bio-signal processing and modelling. The goal of the project was to provide the source code to the research field. The code is intended to be further developed by its users and has a modular design. The modules are based on relevant sEMG elements.

This thesis includes background theory, project methodology, code implementation, results, a self-critical discussion, and literature references.

Theory

Motor Control

The AP is an electrical phenomenon in the nervous system that enables information transmission. It is the fundamental unit of communication between neurons (electrically excitable cells). The AP represents the rapid and temporary changes in electrical potential across a neuron's cell membrane. Signalling in the nervous system utilizes changes in frequency to code information. Muscle contractions are enabled by electrical and chemical activity transmitted from the central nervous system (cortex), via the spinal cord to the peripheral nervous system, and finally, to the skeletal muscles. The signal originates from the motor cortex and is transmitted via a motor nerve, toward the muscle tissue. At the distal end, the nerve branches out to individual motor neurons, innervating individual muscle fibres. The point of connection between a motor neuron and a muscle fibre is called the neuromuscular junction (NMJ) (Moritani, Stegeman, & Merletti, 2004). *Figure 1* illustrates the connection between the cortex in the central nervous system, the spinal cord in the peripheral nervous system, and the skeletal muscle biceps brachii. Additionally, a schematic of the MU is displayed with the organization of muscle fibres, the innervating motor neuron, and the connection point known as the NMJ.

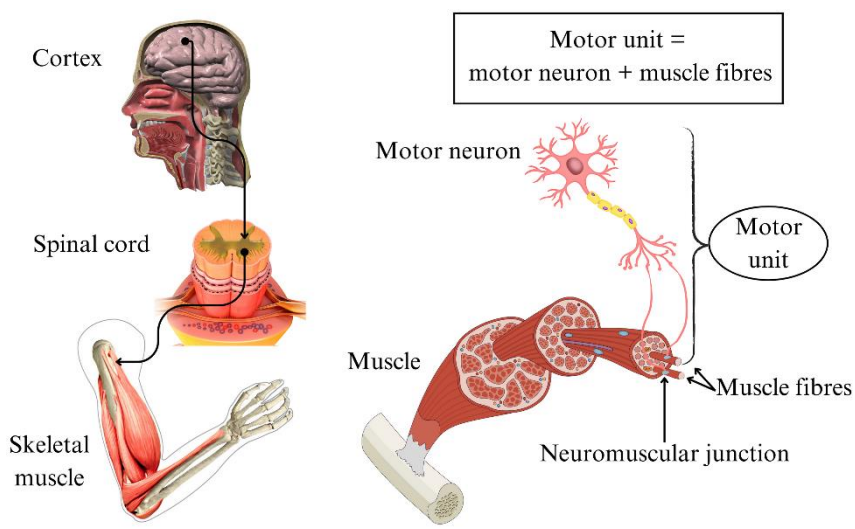

Figure 1. Schematic drawing over motor control pathways from the cortex (brain), via a motor nerve in the spinal cord, to the individual fibres within skeletal muscle.

Current Distribution and Action Potential

When an AP reaches the NMJ, the neurotransmitter acetylcholine (a signalling molecule) is released. The neurotransmitter binds to ligand-gated receptors on the muscle fibre, triggering cell membrane depolarization. The depolarized muscle fibre contracts as the AP propagates along the fibre. The signal often propagates in two opposing directions since the NMJ sits somewhere along the length of the muscle fibre (Moritani, Stegeman, & Merletti, 2004). *Figure 2* illustrates how the signal propagates inside a fibre. The muscle fibre is displayed as a grey cylinder, with the AP illustrated as a black wave moving in opposing directions along the z-axis (depolarization front). The AP is surrounded by regions where the fibre is at rest. At the centre of the fibre sits the NMJ, which is innervated by a motor neuron originating in the central nervous system. The depolarization of the fibre membrane starts at the NMJ and moves outward.

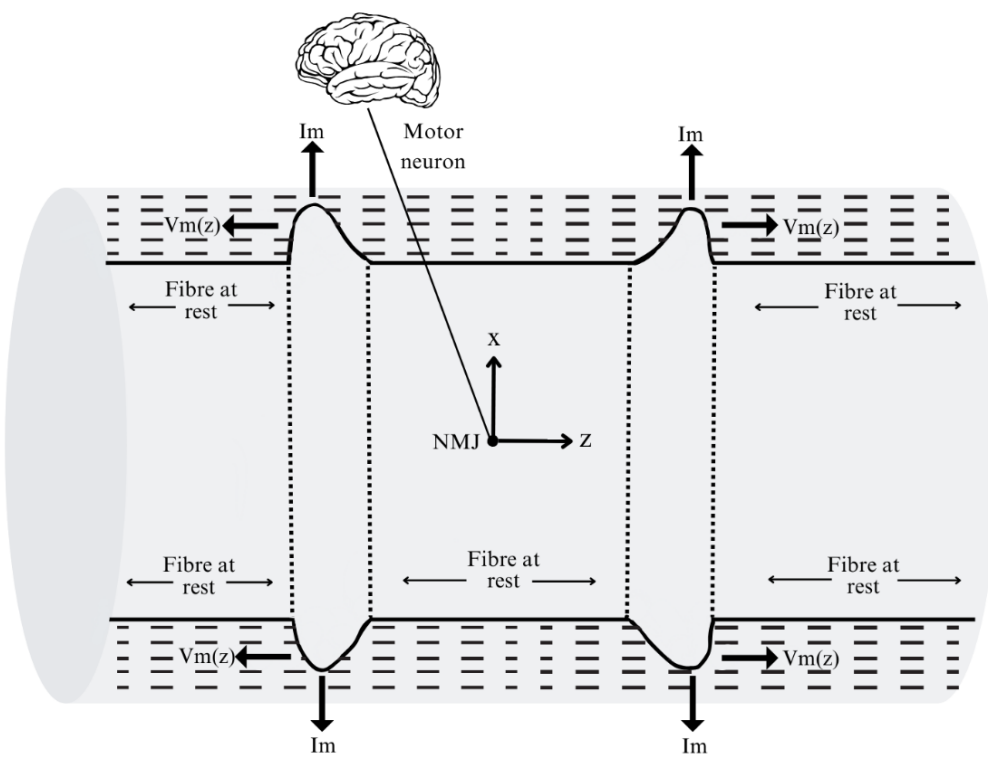

Figure 2. Schematic drawing of the action potential, Vm(z), propagating along a fibre, surrounded by regions with the fibre at rest. The direction of the membrane current distribution, Im, is indicated by arrows, and the neuromuscular junction is positioned in the centre of the fibre.

The membrane current distribution (MCD), I_m, and the AP, V_m, of skeletal muscle have been mathematically modelled by (Merletti, Lo Conte, Avignone, & Guglielminotti, 1999) according to the equations:

$$I_m = C\frac{d^2V_m(z)}{dz^2} = CA\lambda^2(\lambda z)[6 - 6\lambda z + (\lambda z)^2]e^{-\lambda z} \quad (1)$$

$$V_m(z) = A(\lambda z)^3 e^{-\lambda z} - B \quad (2)$$

where A denotes a constant to fit the amplitude, B denotes the resting membrane potential, λ is a scaling factor expressed in mm^{-1}, C is a proportionality constant, and z denotes the distance along the fibre. The MCD given by (1) has the shape of a tripole, indicated by the vertical lines P1, P2, and P3 in *Figure 3*. The figure displays the normalized MCD with the associated tripole lines and the membrane AP of the muscle fibre.

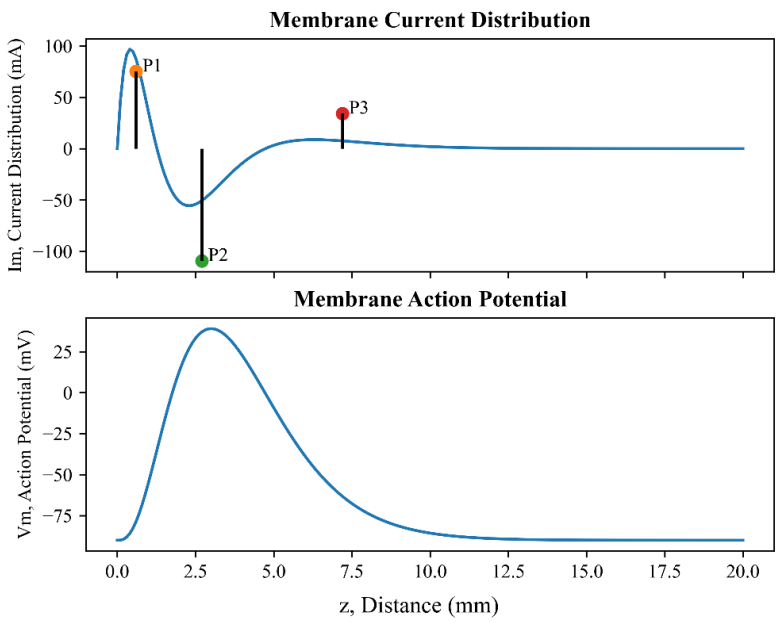

Figure 3. Change in membrane current distribution (top) and action potential amplitude (bottom) along the distance of a muscle fibre. Poles P1, P2, and P3 in the membrane current distribution indicate the tripole shape.

The tripole shape is the result of sequential opening and closing of voltage-sensitive ion channels in the muscle fibre membrane. The channels give rise to the mechanism by which the AP propagates (Huxley, 1974). The duration of a muscle AP is approximately 5 ms and the conduction velocity along the muscle fibre is roughly $4 - 5\ m/s$ (Merletti, Lo Conte, Avignone, & Guglielminotti, 1999).

Signal Propagation

Once the AP has been generated, at the NMJ, it propagates along the length of the fibre, until it reaches the end of the muscle (tendon). The electrical potential is also transmitted through the surrounding tissue, eventually reaching the surface of the skin. At the skin surface, the electrical signal may be recorded with sEMG. The potential, (Φ), measured on a plane skin surface at position (x, z), has been described by (Merletti, Lo Conte, Avignone, & Guglielminotti, 1999) according to the equation:

$$\Phi_j(x,z) = \frac{1}{2\pi\sigma_r}\sum_{i=1}^{6}\frac{P_i}{\sqrt{((x-x_i)^2+y_i^2)K_a+(z-z_i)^2}} \qquad (3)$$

where j is the muscle fibre index, $P_i(x_i, y_i, z_i)$ is the amplitude of the tripole, σ_r denotes the radial electrical conductivity, K_a is the electrical conductivity ratio along the fibres and across the fibres, and y_i denotes the depth of the fibre. *Equation 3* was combined with the cylindrical shape of the volume conductor, described by (Farina, Mesin, Martina, & Merletti, 2004). This means the plane skin surface instead has a cylindrical shape and positions on the skin surface (x, z), are determined by applying the law of cosines.

Motor Unit Recruitment

The sum of all simultaneous APs, from each fibre in one MU, is called the motor unit action potential (MUAP). The sum of all MUAPs from a specific unit is known as a MU signal. The sum of all MU signals is the entire electrical activity of a muscle. The number of recruited MUs is dependent on muscle contraction level. A stronger voluntary muscle contraction will result in greater MU recruitment and increased firing frequencies. The early recruited MUs innervate fewer muscle fibres and thus generate less contraction force. The later recruited MUs innervate more muscle fibres and have larger cell bodies. Thus, they can generate a larger contraction force. The relationship between MU size and recruitment is known as Henneman's size principle (Henneman, Somjen, & Carpenter, 1965). The early recruited MUs are typically firing at the maximum frequency, while the last recruits have the lowest firing frequency. The characteristics of MUs are, in this work, dependent on several parameters, such as peak twitch force (P_k), the number of fibres innervated by each MU (nf_k), and the area encompassed by each MU territory (A_k). The following equations have been described by (Fuglevand, Winter, & Palta, 1993) and are used to determine the number of innervating muscle fibres and the cross-sectional area of a MU:

$$P_k = e^{b \cdot k} \tag{4}$$

$$b = \frac{\ln(RP)}{n} \tag{5}$$

$$nf_{total} = \frac{A_m}{A_f} \tag{6}$$

$$P_{total} = \sum_{i=1}^{n} P_k \tag{7}$$

$$nf_k = (nf_{total}/P_{total}) \cdot P_k \tag{8}$$

$$A_k = \frac{nf_k}{\partial} \tag{9}$$

where b denotes a coefficient for twitch force, RP is the range of twitch forces, nf_{total} denotes the total number of fibres in the muscle, A_m is the cross-sectional area of the muscle, A_f represents the average area of a single fibre, P_{total} represents the number of fibres required to exert one unit of force (1 unit force ≈ twitch force of smallest MU), ∂ denotes the MU fibre density, n is the number of MUs, and k denotes the current MU index. The same authors (Fuglevand, Winter, & Palta, 1993) have also described the MU recruitment model in great detail. They outline several MU features which influence the recruitment order: recruitment threshold excitation (RTE), peak firing rate (PFR), maximum excitation (E_{max}), and the firing rate response ($FR_k(t)$). The following equations describe the intricate mathematical synergy of these features:

$$RTE_k = e^{a \cdot k} \tag{10}$$

$$a = \frac{\ln(RR)}{n} \tag{11}$$

$$PFR_k = PFR_1 - PFRD \cdot \frac{RTE_k}{RTE_n} \tag{12}$$

$$E_{max} = RTE_n + \frac{PFR_n - MFR}{g_e} \tag{13}$$

$$FR_k(t) = g_e \cdot (E(t) - RTE_k) + MFR \quad E(t) \geq RTE_k \tag{14}$$

where $PFRD$ denotes the peak firing rate difference, MFR is the minimum firing rate, g_e is the excitatory gain, a represents a coefficient for the assignment of recruitment threshold excitation, RR is the range of recruitment thresholds, n is the total number of MUs, t denotes time, $E(t)$ is the excitatory drive function, and k denotes the current MU index.

The muscle, MUs, and individual muscle fibres are all modelled as cylinders. This relates to the model described by (Farina, Mesin, Martina, & Merletti, 2004), and it resembles anatomical structure. However, MUs are functional units, distinct from the anatomical structure of the muscle itself. Every MU's cross-sectional area A_i is modelled as a circle, with MU radius r_i. *Figure 4* illustrates a schematic drawing of fibre organisation inside the MUs and MU organisation inside the muscle. Also, it displays a two-dimensional cross-section of the muscle and the neural firing patterns of three recruited MUs.

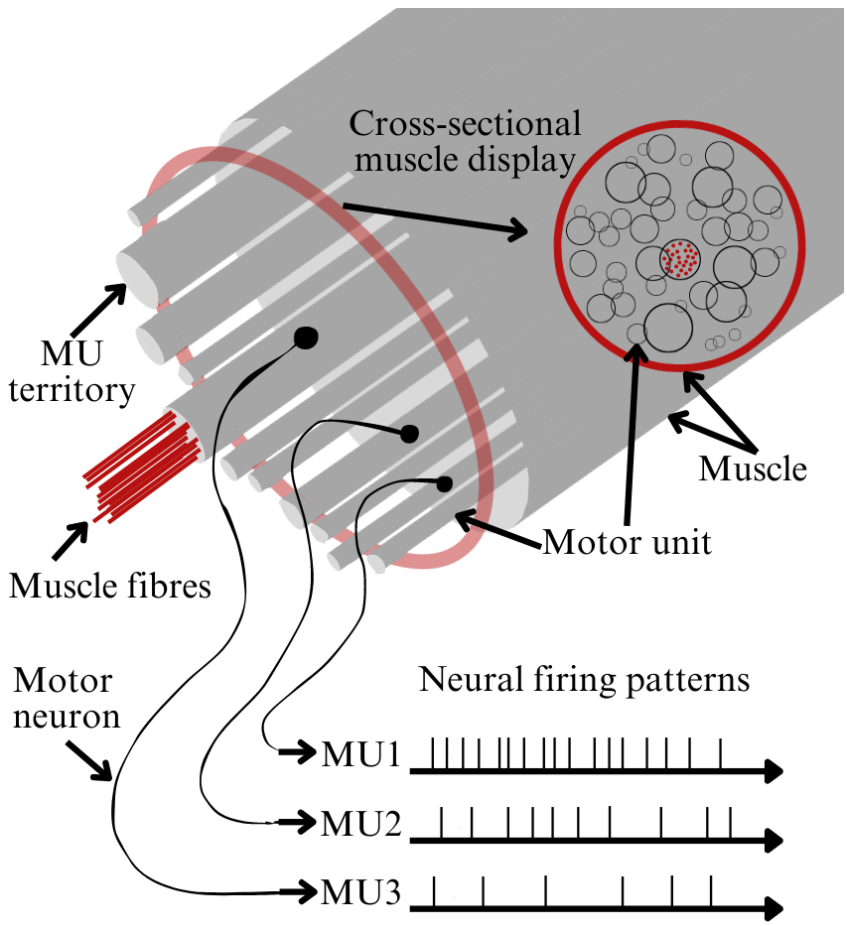

Figure 4. Schematic of muscle fibre and MU organisation inside the muscle, a cross-sectional display of significantly overlapping motor unit territories, and the neural firing patterns of the first three recruited motor units.

Electromyography

EMG is a recording technique to measure skeletal muscle response from central nervous system stimulation. It registers the electrical activity generated in the muscle tissue. Intramuscular EMG requires a needle to penetrate the skin, measuring the electrical potential inside the muscle. These recordings are precise and very local, only recording the space around the tip of the needle. Contrary, sEMG measures the electrical potential after being transmitted through the muscle, fat, and skin layers on the surface of the skin.

Isolating and sorting the APs originating from the same MU is highly relevant for EMG research, and it is accomplished by applying decomposition algorithms to the recorded signal. Clinically, EMG can be used to detect muscle abnormalities, diagnose muscular disorders, measure muscle activation, and determine MU recruitment. Post-processed EMG data can enable prosthetic control or be used as the basis for human-computer interaction (Farina, Roberto, & Stegeman, Biophysics of the Generation of EMG Signals, 2004).

Surface Electromyography

sEMG is a non-invasive recording from the skin surface, registering the muscle activity from below the electrodes. The recorded signal is a summation of electrical activity generated by up to hundreds of MUs and thousands of fibres, firing simultaneously within a muscle, superimposing on one another. Clinical sEMG is typically performed with a recording duration of seconds, up to several minutes, and may be applied to muscles throughout the body. Comprehensive sEMG utilizes two-dimensional electrode arrays with varying shapes, sizes, and number of electrodes. These types of recordings are called high-density sEMG, referring to the closely packed electrodes in the array (Moritani, Stegeman, & Merletti, 2004).

Analysing sEMG data may aid a physician when diagnosing muscle abnormalities, since recordings from healthy and diseased muscles differ. The elicited signal is also unique to a specific muscle and bodily movement. The recorded signal can be imagined as an activity map of the muscle. The level of contraction greatly affects the signal amplitude, as it will change the number of active MUs. The MU is considered the electrically generating entity in the context of sEMG, even though signals originate in individual fibres inside the MU (Farina, Roberto, & Stegeman, Biophysics of the Generation of EMG Signals, 2004). *Figure 5* illustrates how sEMG may be used on the wrist.

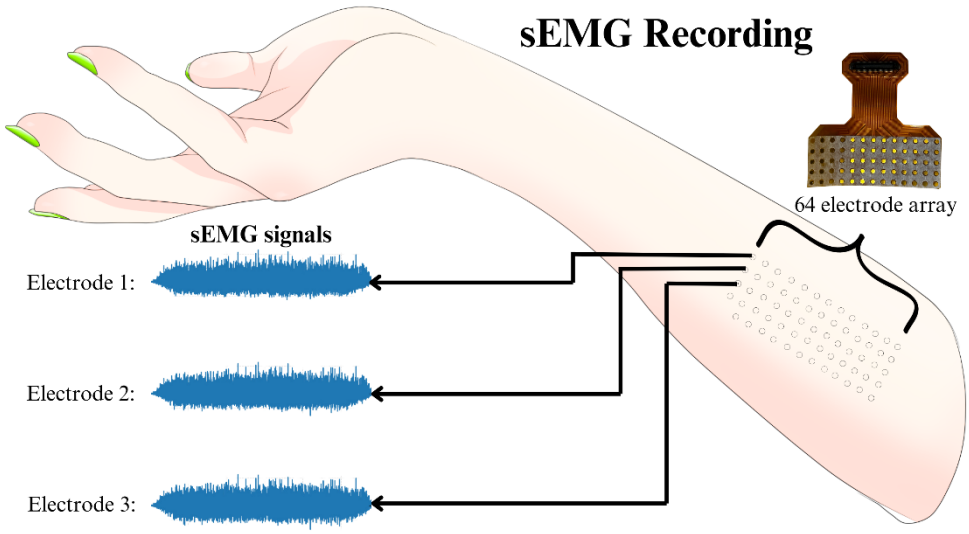

Figure 5. Illustration of how a surface electromyogram may be placed on the wrist to record muscle activation. The signal is recorded with a 64-electrode array. Example signals are displayed for three electrodes.

Surface Electromyogram Decomposition

Decomposition of the sEMG signal is a process of decoding the signal into individual MU firing patterns. Decomposition can provide information about the firing times and APs of each MU, which in turn may give insight into muscle health or disease. Synthetic sEMG signals can be used as a tool for evaluating decomposition algorithms. By comparing the decomposed signal to the ground truth, the quality of the decomposition can be assessed. The MU firing patterns of synthetically generated data are known and can be compared to the decomposition output. Thus, decomposition may be used to evaluate the generated MUAPs (Holobar, Farina, & Zazula, 2016). The performance of the model and the quality of the generated data are outlined in the *Evaluation* section. The decomposition used for evaluation produces two sets of data for each MU: the decomposed neural firing pattern and the spike-triggered average (STA). Recall and precision of the decomposition may be calculated. The recall is determined by dividing the number of decomposed matched firings by the number of ground truth firings. The precision is determined by dividing the number of decomposed matched firings by the total number of decomposed firings. The STA provides a measure of the underlying MUAP by averaging by APs of the sEMG signal (Lundsberg, Björkman, Malešević, & Antfolk, 2022).

Method

Implementation

The project implementation was done using the Python programming language (version 3.11.5) (Foundation, 2023). All code was written in text editor Visual Studio Code (version 1.83) and the program has two dependencies: NumPy (version 1.26) and Matplotlib (version 3.8.2). The program follows object-oriented programming guidelines and consists of three classes: MotorUnit, SurfaceEMG, and SaveData. The first two classes are tied to the theoretical models of MUs and sEMG as described in the *Theory* section. The third class is used only for loading and saving the generated data. A table of all methods, for each class, is provided in the appendix. The simulation model was implemented modularly, with a bottom-up design. *Figure 6* displays the general implementation structure of the code.

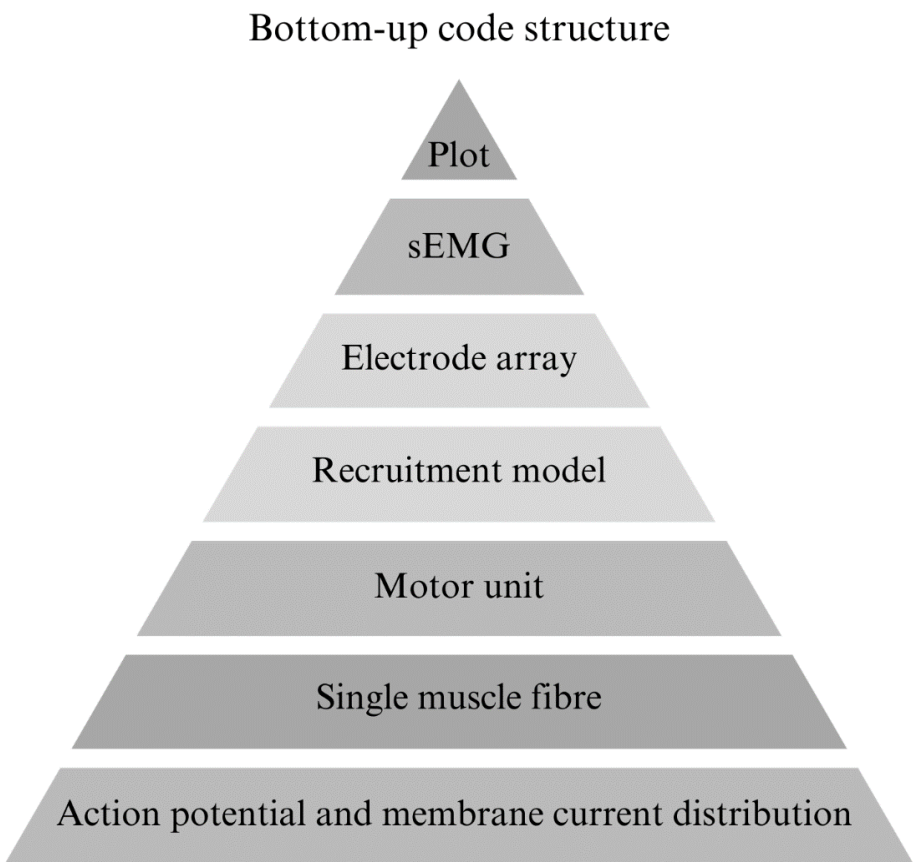

Figure 6. Order of code implementation with a bottom-up design.

Module 1 – Membrane Current Distribution and Action Potential

The first implemented module contained calculations of the MCD and AP. These two concepts established the bedrock of our program and were implemented in accordance with (Merletti, Lo Conte, Avignone, & Guglielminotti, 1999) as written in *Equations 1* and *2*. The tripole shape was determined by finding the centroid of each phase of the MCD. The phases are separated by a zero-crossing of the y-axis. The three poles (phases) are represented by the vertical lines P1, P2, and P3 in *Figure 3*. The output of the module was compared to the reference (Merletti, Lo Conte, Avignone, & Guglielminotti, 1999).

Module 2 – Single Muscle Fibre

The second implemented module contained code for a single muscle fibre. This component was implemented by defining a length vector on which the AP and MCD could propagate. To complete the theoretical model of the fibre, the NMJ, fibre ends, conduction velocity, and electrical conductivities were defined. Additionally, a fibre depth was established, i.e. the distance to the fibre from the surface of the skin. Then, the signal propagation inside the fibre could be modelled as described by (Merletti, Lo Conte, Avignone, & Guglielminotti, 1999) in *Equation 3*. Simultaneously, the shape of the skin surface was modified from planar to cylindrical as described by (Farina, Mesin, Martina, & Merletti, 2004), meaning the electrical potential in electrode positions (x, z), were determined by applying the law of cosines instead of the Pythagorean theorem. Additionally, a signal attenuation factor was added for signals propagating along the muscle fibres. This attenuation factor replaced the end of fibre phenomenon outlined by (Merletti, Lo Conte, Avignone, & Guglielminotti, 1999). The attenuating factor changed the denominator in the *Equation 3* to the following:

$$\sqrt{((x - x_i)^2 + y_i^2)K_a + e^{0.09 \cdot |z - z_i|}(z - z_i)^2} \qquad (15)$$

The attenuation factor was determined empirically by comparing the attenuation of the simulated signal to attenuation in real sEMG data.

Module 3 – Motor Unit

The third module was a type of iterator. The module generated repetitions of the single fibres from module 2 and summed them up to simulate a MU. Muscle fibre positions were spread randomly inside a cylinder representing the MU. However, the individual fibres were implemented without any cross-sectional area. Instead, fibres were simulated as two-dimensional parallel lines inside the MU cylinder. The third module included no variation in MU size, radius, or fibre innervation level, even though these properties are linked to the MU. Instead, these features were introduced together with the recruitment model in module 4. Completing module 3 allowed for the generation of MUAPs. The output from the module was compared to the reference (Merletti, Lo Conte, Avignone, & Guglielminotti, 1999).

Module 4 – Recruitment Model

The fourth module to be implemented was the recruitment model for MUs. First, *Equations 10-14* were implemented to simulate a population of MUs and calculate their respective firing times. Afterwards, *Equations 4-9* were added to determine the radius and fibre innervation level of each MU. *Figure 11*, in the *Result* section, shows the recruitment of MUs with their respective firing times. At this point, another cylinder was implemented to represent the entire muscle. The MUs were spread out randomly inside the muscle, in a similar fashion to fibres spread out inside the MUs. At this point, the volume conductor, i.e., the muscle itself, was completed. *Figure 4*, in the *Theory* section, illustrates the volume conductor as a muscle filled with MUs, and each MU filled up with muscle fibres. The output from the module was compared to the reference (Fuglevand, Winter, & Palta, 1993).

Module 5 – Electrode Array

The electrode array implementation was completed at this stage. A matrix of electrode positions was generated to simulate the recording sEMG array. The sum of all MUAPs, i.e., the original sEMG signal, was calculated in the defined electrode positions. The locations of the electrodes and the interelectrode distance are adjustable. The electrodes are simulated as a single point of detection. Afterward, simulated background Gaussian noise was added to the original signal. The sEMG signal with and without added noise is shown in *Figure 10*, in the *Result* section. The completion of this module enabled simulation of synthetic sEMG data.

Simulation and Evaluation of sEMG Data

To simulate relevant data without defining parameters, a user can call any method inside any of the classes, as seen in *Class and Method Structure* in the *Appendix*. The methods are tied to theoretical elements of sEMG instead of the modules. Pre-defined default values may be used to simulate data, which enables a new user to bring about figures and data directly. The method will return objects or data relevant to the selected method, for instance: a MU, an amplitude, an array of muscle fibre potentials, or various plots. However, the user can adjust any of the available parameters by manually overwriting the default value. Manipulating the input parameters will naturally affect the output accordingly. All the functional parameters in the program are listed under *Parameters* in the *Appendix*. The methods, as well as the objects they generate, are intended to be familiar to researchers in the field of neurophysiology and neuroengineering.

Decomposition of the sEMG signal decodes it into individual neural firing patterns, with their respective MUAPs. The quality of a new decomposition algorithm can be assessed by comparing the decomposed signal to the ground truth. However, by using an established decomposition algorithm, synthetically generated sEMG data may be evaluated. In the scenario of synthetic sEMG data, ground truth parameters are known, i.e., the number of simulated motor units, the simulated neural firing times, etc.

Result

We have successfully implemented a model for synthetic sEMG data following the description in the *Method* section. A significant part of the project's result is the source code itself, which is publicly available on GitHub (https://github.com/Josef-Djarf/sEMG-Sim).

Simulations

Figure 7 shows how a single MUAP propagates. Ten electrodes, located 20 mm apart, record the MUAP at different times of the simulation. The electrode array sits above the NMJs, located closest to electrode number 2, determined by the early detection of the MUAP. The figure replicates *Fig. 2* in *Modelling of Surface Myoelectric Signals−Part I: Model Implementation*, (Merletti, Lo Conte, Avignone, & Guglielminotti, 1999). Replication of their figure suggests accurate implementation of their model.

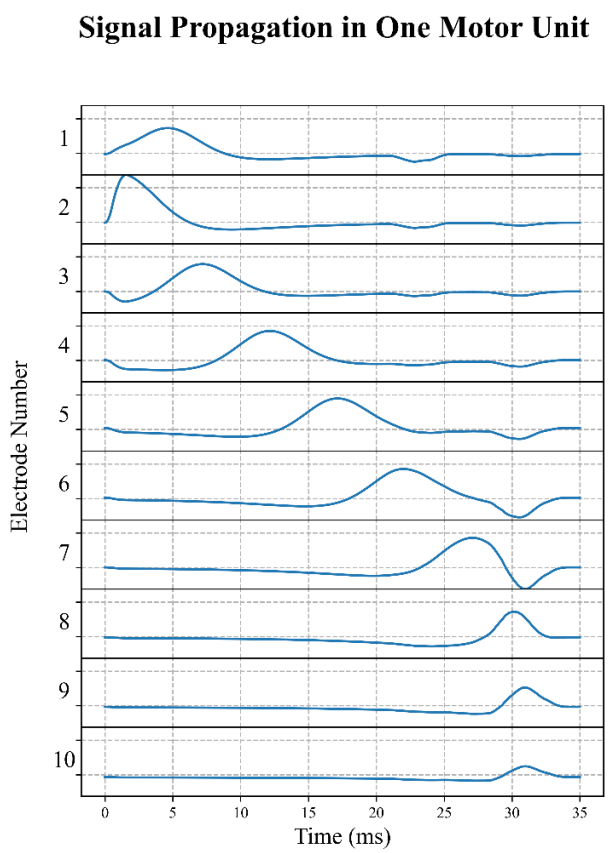

Figure 7. Simulation of Fig. 2 from (Merletti, Lo Conte, Avignone, & Guglielminotti, 1999), a motor unit action potential propagating and recorded in 10 electrode positions.

Figure 8 below displays the type of data a potential user may simulate. The plot contains the normalised sEMG signal generated from 100 MUs, as recorded by a 3x3-electrode array over 30 seconds. The parameter values used are presented in *Tables 1* and *2*.

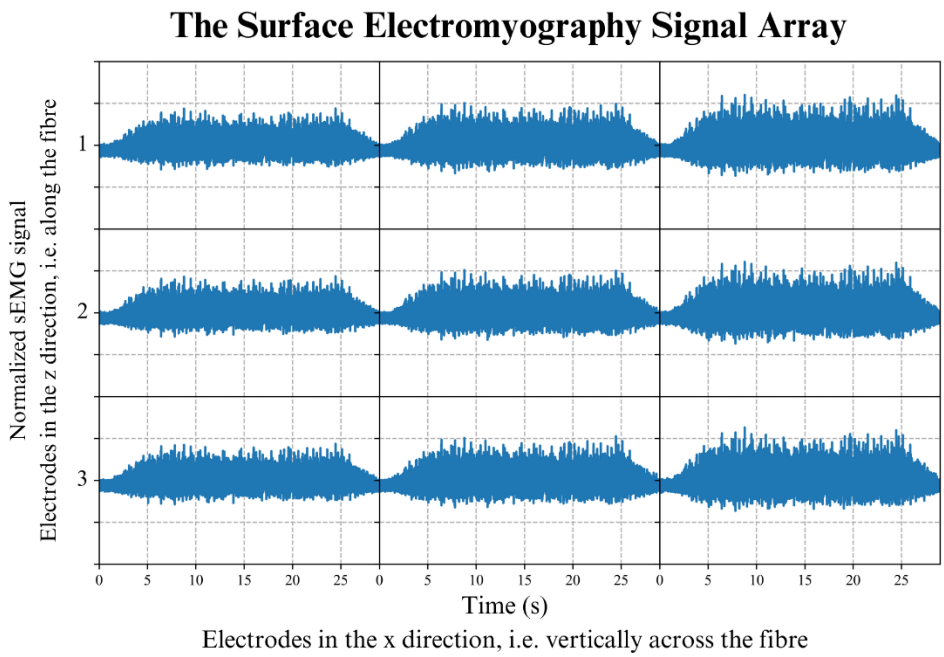

Figure 8. Normalised surface electromyogram data from 100 simulated MUs, recorded over 30 seconds, with a 3x3-electrode array.

Figure 9 was simulated using the same parameter values as the simulation of *Figure 8* but is recorded with only one electrode. The data in *Figure 9* is not normalised. Zooming in on the data, as illustrated by the arrow, resolves the signal amplitude changes occurring on a millisecond scale and the noise oscillations occurring at the microsecond scale.

Figure 10 contains the same simulation as *Figure 9*, but the original sEMG signal and the signal with added Gaussian noise have been colour-coded for illustration purposes. The added noise was generated with a signal-to-noise ratio of 3 dB.

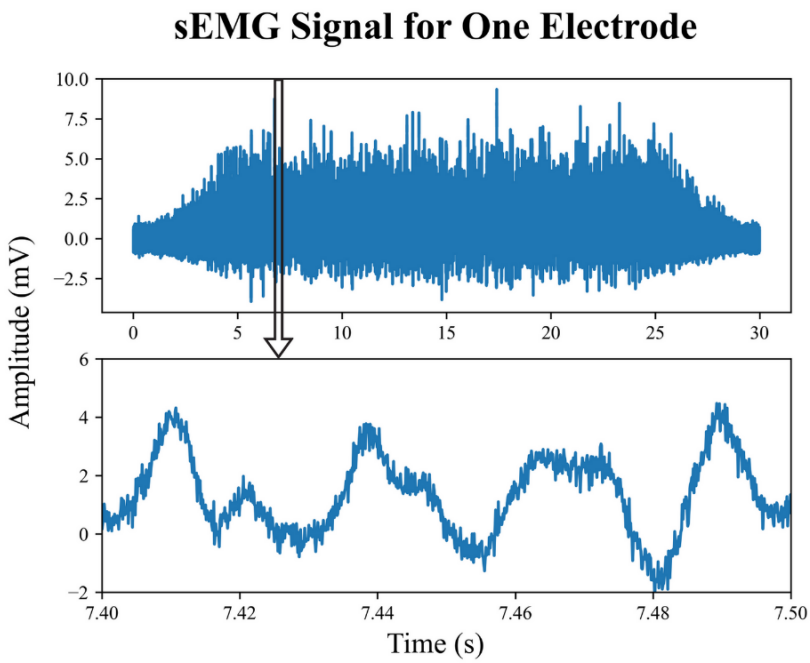

Figure 9. Surface electromyogram signal (top), zoomed in signal (bottom).

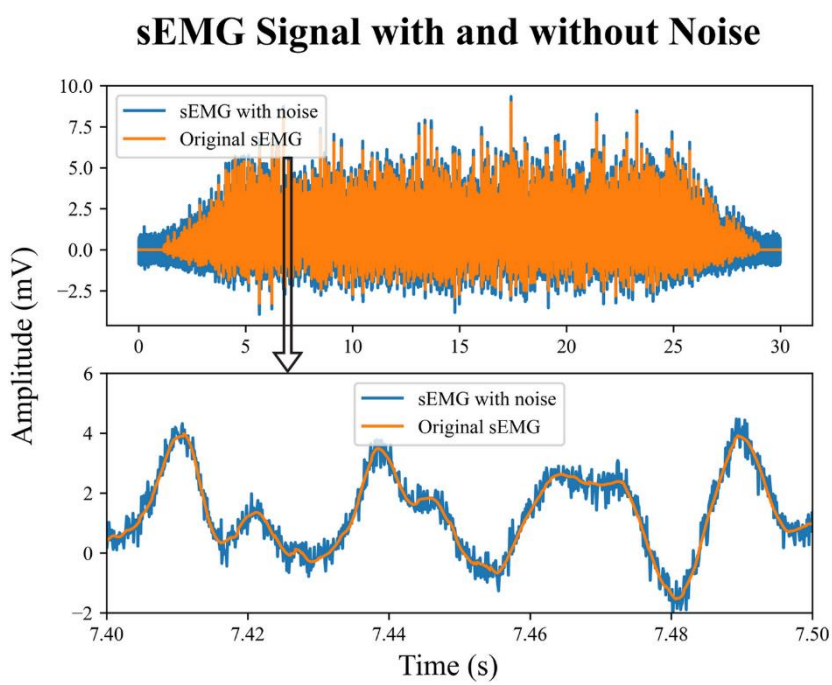

Figure 10. Color-coded original surface electromyogram signal and signal with noise, zoomed in signal (bottom).

The simulations in *Figures 8-10* were generated using the recruitment model. The recruitment of the MU pool followed a trapezoid excitatory drive function of 30 seconds. The function has three phases. First, it ramps up for 5 seconds, then it remains static for 20 seconds, and finally, it ramps down for 5 seconds. The excitatory drive function had a maximum excitation level of 20 %, and contained a total of 100 MUs in the pool. The MU recruitment and firing times are illustrated below in *Figure 11.* The remaining parameter values for the model are presented in *Tables 1* and *2*.

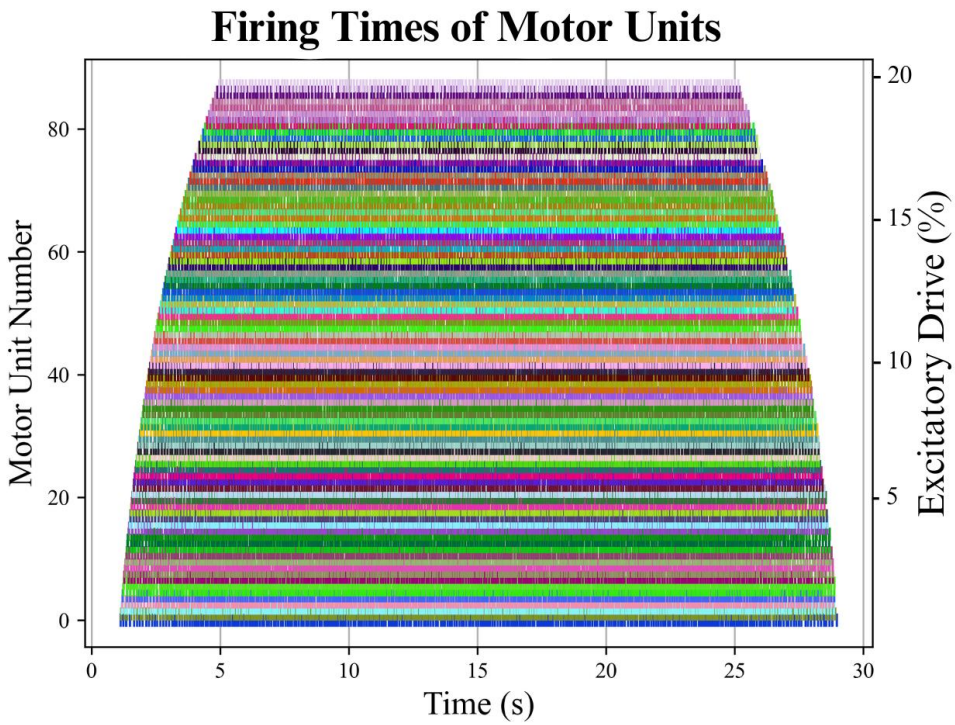

Figure 11. Recruitment and firing patterns of motor units following a 30 second trapezoid excitatory drive function, with maximum excitation level 20 %.

The early recruited MUs are firing at the maximum frequency and contain the least muscle fibres, while the last recruits have the lowest firing frequency and contain the most muscle fibres. The model of recruitment is based on *Equations 10-14.* The first MUs also encompass the smallest territory, and the last recruits encompass the largest territory, following Henneman's size principle. The relationship between MU number, size, generated peak twitch force, and the number of innervating fibres, is described by *Equations 4-9.* In *Figure 11,* the firing times of each MU is denoted by a coloured vertical line along the x-axis. The y-axis represents the MU number and excitatory drive level.

The parameters in *Table 1* lists the values used to create the simulations above for the SurfaceEMG class. All parameters are set to their default value with the exception of parameters coloured in grey. Parameters come with the appropriate unit, but some parameters are unitless.

Table 1. List of parameter values for SurfaceEMG class.

SurfaceEMG	
Simulation time (s)	30
Sampling rate (Hz)	10 000
Excitatory drive function (s)	[5 20 5]
Maximum excitation level (%)	20
Signal to noise ratio (dB)	3
Signal amplitude offset (mV)	0
Number of motor units in the motor pool	100
Recruitment range (n.u.)	30
Excitatory gain	1
Minimum firing rate (Hz)	8
Peak firing rate first MU (Hz)	35
Peak firing rate difference of motor units (Hz)	10
Inter-spike-interval coefficient of variation	0.15
Twitch force range (n.u.)	100
Motor unit density (fibres/mm^2)	20
Number of fibres in smallest MU	25
Number of fibres in largest MU	2 725
Muscle fibre diameter (mm)	0.046
Muscle diameter (mm)	15
Muscle depth (mm)	10
Number of electrodes in z-axis	3
Number of electrodes in x-axis	3
Y-axis minimum plot limit	-1
Y-axis maximum plot limit	1

The parameters in *Table 2* lists the values used to create the simulations above for the MotorUnit class. All parameters are set to their default value with the exception of parameters coloured in grey. Parameters come with the appropriate unit, but some parameters are unitless.

Table 2. List of parameter values for MotorUnit class.

MotorUnit	
AP amplitude constant (V)	0.096
Membrane resting potential (V)	-0.090
AP scaling factor (mm^{-1})	1
AP proportionality constant	1
AP plot length (mm)	20
Muscle fibre length (mm)	210
Conduction velocity (m/s)	4
Ratio axial/radial conductivity	6
Radial conductivity (S/m)	0.303
Interelectrode spacing (mm)	10
Number electrodes z-axis	3
Number electrodes x-axis	3
Electrode shift z-axis (mm)	165
Neuromuscular junction position (mm)	90
Fibre depth (y-axis) (mm)	10
Fibre x-position (mm)	10
Extermination zone width (mm)	10
Innervation zone width (mm)	5
Time fibre simulation (ms)	35
Motor unit radius (mm)	1
Number of fibres	100
Motor unit depth (y-axis) (mm)	10
Motor unit x-position (mm)	0
Y-axis minimum plot limit	-1
Y-axis maximum plot limit	1

Evaluation

New data was simulated for the evaluation process. This time parameter values were adjusted to align with the decomposition algorithm, empirically testing different values to resemble previous experimentally collected sEMG data. The following parameter values were utilised: an 8x8-electrode array, 50 MUs in the unit pool, $55\ mm$ muscle diameter, $30\ mm$ muscle depth, 0.001 in the ratio between axial and radial conductivity, electrode shift z-axis $75\ mm$, and conduction velocity for each MU varying between $4 - 6\ m/s$. The sampling rate was set to $10\,000\ Hz$, but the data was downsampled to $2000\ Hz$ before decomposition. The remaining parameter values were set to default, as shown in *Tables 1* and *2*. The simulated sEMG data was decomposed using algorithms implemented in MATLAB (Inc., 2023), since decomposition algorithms in Python are not available. The simulated data was exported to a MAT-file format. The decomposition algorithm employed is comprehensively described by (Lundsberg, Björkman, Malešević, & Antfolk, 2022). The decomposition produces two sets of data for each MU: the decomposed neural firing pattern and the signal STA.

The number of individual firings was determined for both the decomposed signal and the ground truth. The second decomposed MU was selected for illustration since its firing pattern most accurately matched with ground truth MU firing pattern. The ground truth of the second MU contained 286 firings, while the decomposed correlate contained 849 firings. The decomposition overestimated the number of firings in the MU by $\approx 300\ \%$. The recall of this estimation was $\approx 34\ \%$, using a variation threshold $0.9/1$ and a detection window of $5\ ms$, i.e., the duration of a muscle AP. The precision was calculated to $\approx 11\%$. 32/50 MUs in the pool were activated by the recruitment model. The decomposition algorithm estimated 4 out of those 31 MUs. The estimated MUs contained on average 820 firings, while ground truth MUs contained 275. *Figure 12* illustrates the firing pattern agreement of the second simulated MU to ground truth.

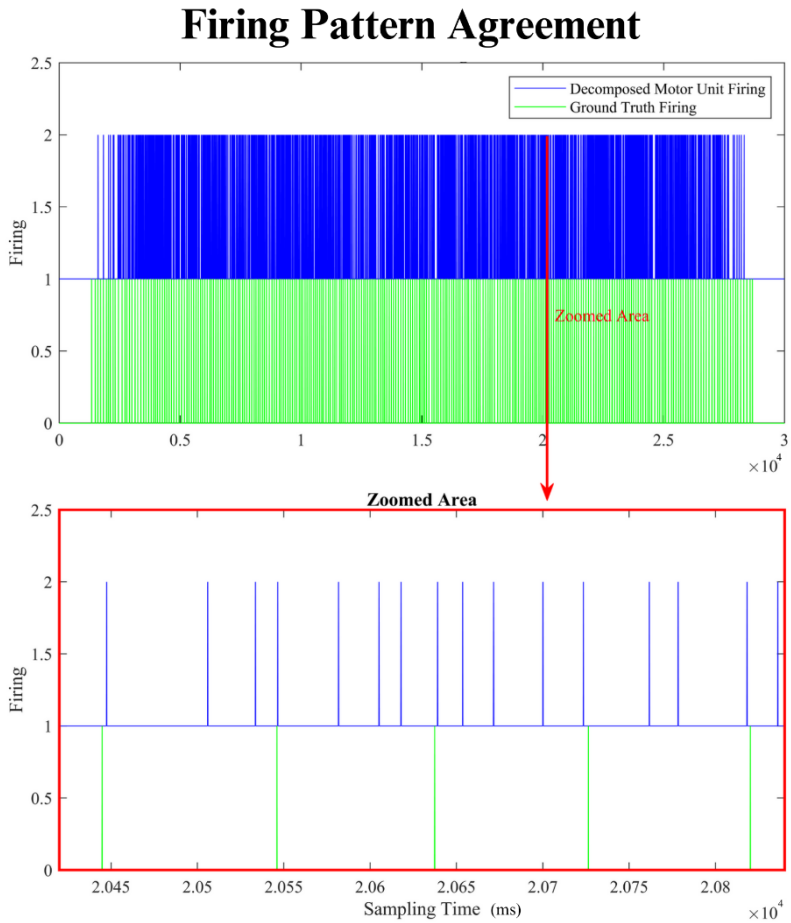

Figure 12. Decomposed firing pattern (blue) and ground truth firing pattern (green) of the second simulated motor unit. The red arrows indicate the zoomed in area (bottom).

Decomposed MU firings and the ground truth firings occurring simultaneously ($< 5\ ms$) are assumed to be correlated, indicating they originate from the same MU. The decomposition algorithm overestimated the number of firings for each MU and did not identify all simulated MUs. Thus, the decomposed firing pattern contain firings from multiple simulated MUs. The overestimation is visible in the zoomed-in area in *Figure 12*, where there are 3 times as many decomposed firings as ground truth firings, visible as blue lines without a matching green line.

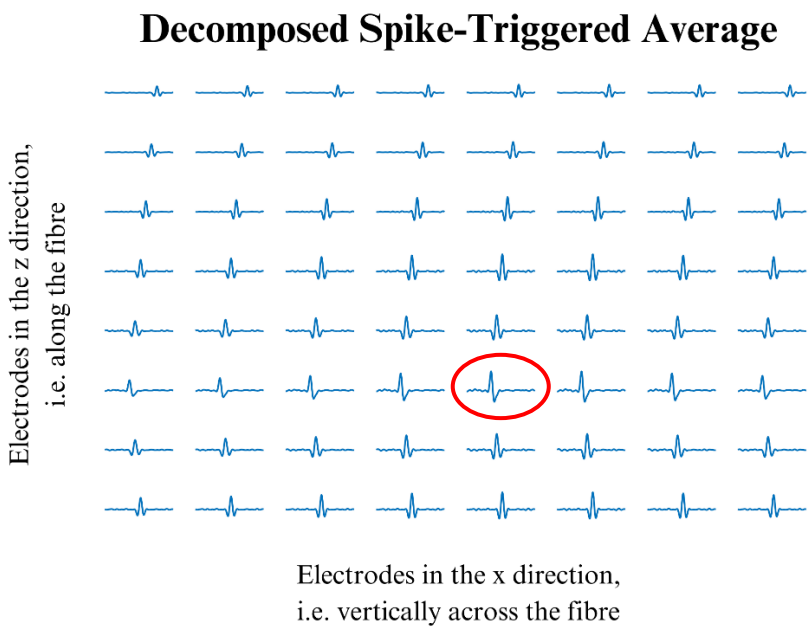

Figure 13. Spike-triggered average in 8x8-electrode array of the decomposed motor unit. The red circle indicates the spike-triggered average motor unit action potential shown in figure 15.

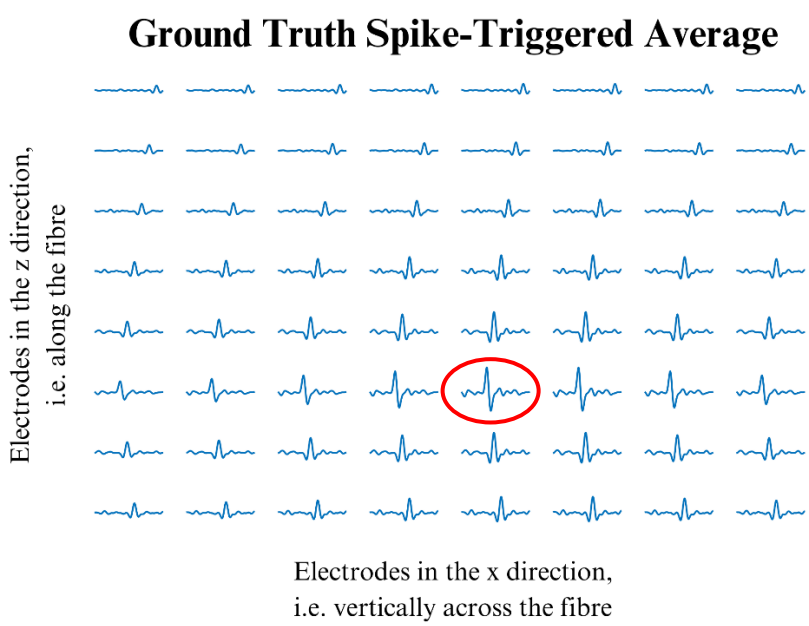

Figure 14. Ground truth spike-triggered average in 8x8-electrode array of the second simulated motor unit. The red circle indicates the spike-triggered average motor unit action potential shown in figure 15.

The STA estimates the average MUAP shape by using the time stamps of the neural firing pattern. In *Figure 13*, the STA from the second simulated MU is displayed in the 8x8 electrode positions. The red circle in the figure indicates where the signal amplitude is the highest. The highest amplitude corresponds to the proximity between this electrode and the signal source (the NMJ). Conversely, the electrodes further away from the source exhibit lower amplitudes, since the signal attenuates as it propagates. *Figure 14*, illustrates the STA in the 8x8 electrode positions from the second simulated MU. This MU likely corresponds to the best estimated MU.

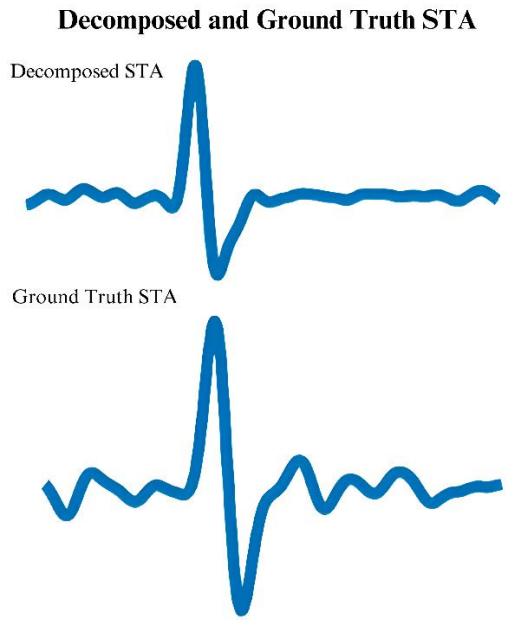

Figure 15. Decomposed spike-triggered average, with the largest amplitude, from the decomposed motor unit compared to the corresponding ground truth spike-triggered average, i.e., the second simulated motor unit.

The STA provides insight into how neurons respond to specific events or triggers. In this context, the STA represents the average response to neural firings at each electrode position. The red circles in *Figure 13* and *14* indicate the signal enlarged and displayed below. *Figure 15* shows a comparison between the decomposed STA and the ground truth STA. The signal, i.e., the MUAP, has the typical tripole shape, as expected of a real muscle contraction, reflecting the sequential opening and closing of voltage-sensitive ion channels in the muscle fibre membrane (Vecchio, et al., 2020). The STA may help to identify and characterise the MUAPs. Displaying the STA in the 8x8-electrode array configuration sheds light on the temporal dynamics and attenuation of the AP within the muscle.

Discussion

In this thesis, we implemented a sEMG simulation model based on mathematical equations described in the scientific literature. The implementation consisted of five modules. First, a mathematical description of the MCD and the AP. Second, a length vector resembling the single muscle fibre on which the signal could propagate. Third, an iterator to generate MUs by repeated simulations of fibres. Fourth, the MU recruitment model. Fifth, the electrode array. Completing the five modules allowed for sEMG data to be simulated. Afterwards, the generated data was evaluated using a decomposition algorithm. The five implemented modules match the outline of project objectives, written before the project started:

(1) Implement a single muscle fibre AP
(2) Implement a MUAP
(3) Implement the signal volume conductor (the tissue layers)
(4) Implement the surface electrode array to generate sEMG data
(5) Evaluate the generated data

Method descriptions and parameter explanations have been added to the code. Abbreviations were avoided and parameter names are closely tied to their theoretical underpinning to promote intelligibility. Additional comments were added throughout the code to help the reader follow the sequence of calculations, line by line. The code was written according to object-oriented guidelines. The project's main goal, providing open-source code to the research field, has been reached. Our Python programming ability and understanding of bio-signal processing have improved, which was the purpose of this project.

Adherence to the theoretical sources was validated several times throughout the project. After each module, the code was reviewed and the output was compared to the relevant literature source. A major contention in the project was the replication of *Fig. 2* from *Modeling of Surface Myoelectric Signals−Part I: Model Implementation,* (Merletti, Lo Conte, Avignone, & Guglielminotti, 1999), resulting in *Figure 7*. The cited source does not provide any code or quantitative measure of their results, but the visual resemblance between the source figure and the modelled figure argues for correct interpretation and implementation. We accept the implementation as equivalent but lack conclusive evidence of identical reproduction.

Supported by the theoretical adherence, we argue the data in Figures 8-10 visually resembles experimental sEMG data. In terms of overall appearance, the synthetic data is indistinguishable from experimental sEMG data. Visual

appearance may not reflect accurate physiological modelling, but combining the qualitative (visual) and quantitative (decomposition) assessment strengthens the overall validation of the synthetically generated sEMG. Additionally, the distinctive tripole shape of the extracted MUAP, displayed in Figure 15, aligns well with the expected MUAP from a muscle contraction (Vecchio, et al., 2020) adding credibility to the physiological resemblance of the synthetic data.

As described in the *Evaluation* section, the simulated data was assessed using a decomposition algorithm. However, the decomposition algorithm is limited in its accuracy and has an inherent variance threshold for adjusting sensitivity. The variance threshold determines the sEMG signal amplitude variation permitted for the identification of MUs. A larger variance threshold will allow greater fluctuation in the AP amplitudes from estimated MUs, and vice versa. Thus, the decomposition algorithm is not a perfect quantitative measure, instead it only indicates to what extent relevant sEMG features can be identified. With a lower variance threshold fewer motor units will be detected, but will be more accurately matched with ground truth. Also, adjusting the detection window for firing correspondence will greatly affect the recall and precision of the decomposition. We are critical to having multiple sensitivity thresholds when analysing data, since they may overlap on or even invalidate each other. We decomposed the simulated data mainly to highlight successful extraction of the neural firing pattern and an STA. It demonstrates that the simulated data indeed contains relevant sEMG features. The decomposition heavily overestimated the number of firings in each MU, from around 275 to more than 800 firings. This drastic overestimation could be explained by the fact that all simulated MUs are generated with the same MCD. The MCDs of real MUs are unique. *Figure 12* displays several decomposed estimated firings (blue lines) without origin in the ground truth firing pattern (green lines). The overestimated firings likely belong to one of the other 31 recruited MUs, but the algorithm has not identified these as separate. The algorithm could not distinguish all the MUs overlapping in the sEMG signal. We believe the synthetic data is missing intricate variations found in a real sEMG signals. The simulated data is constructed bottom-up from key components of sEMG, outlined by each module in the *Method* section, but is nonetheless fabricated. It is generated as a discrete signal, based on strictly specified parameter values. Experimentally recorded sEMG signals, on the other hand, are by nature continuous signals with infinitesimal resolution and variation, recorded at a certain sampling frequency.

Limitations

The implemented model is based on equations and theoretical descriptions gathered from several independent sources. Therefore, the model is a synthesis of separate parts. Predictable and seamless synergy between the parts are assumptions about the model. The major assumptions about the code are:

(1) The model's parts synergise in a predictable manner
(2) All APs are generated using the same MCD
(3) The mechanism of tripole (MCD) generation is very simple
(4) Single fibres are two-dimensional and parallelly oriented in a muscle
(5) MU positions are randomly distributed in a muscle
(6) MUs are activated strictly following the recruitment model
(7) The volume conductor (muscle) is homogeneous.

These assumptions naturally outline possible future improvements. Each assumption could be reduced, and eventually removed, by adding new code to the model, to counteract the assumption. Some of the possible improvements have been outlined previously, by the authors in cited literature (Farina, Mesin, Martina, & Merletti, 2004) (Merletti, Lo Conte, Avignone, & Guglielminotti, 1999). As mentioned in the *Preface*, two other projects shaped the early program design, using their previous code as a template. Using their code was the biggest benefit but also the largest setback in the project. It enabled a jump-start in the initial phase of the project, small sets of data could be generated right away which got the project moving forward. However, later on, after several weeks of programming, flaws in the early code surfaced. Major stumbling blocks arose as a consequence of adapting code from other authors, and fragments of the early code are still part of the program to date. Here lies room for future improvement, as it would be possible to re-write the entire project from start to end without using any other source code. We believe the program structure and performance would benefit from such a re-writing. For instance, the processing speed of our program could be greatly improved by avoiding some for-loops and replacing every Python-list with a NumPy array (Harris, Millman, Van Der Walt, & et al., 2020). The current version is limited in terms of processing speed. Since the project is published on GitHub, new collaborators are encouraged to develop the code. The bonus objective, the implementation of a graphical user interface, could provide a smoother user experience.

Another important limitation of the model are the ambiguous parameter domains and codomains. It is possible to generate data which is not theoretically feasible, e.g., incredibly long fibre lengths, or huge radii muscles. Domain and codomain values, i.e., the accepted input and output values, are not specified for the list of parameters. However, most parameters have a theoretically limited domain. Additionally, some parameters are much more sensitive than others and runtime errors may occur with theoretically sound parameter values, due to limited processing power. For instance, manipulating the scaling factor λ, from *Equations 1* and *2*, has reasonable consequences on the MCD and AP, but changes the final sEMG data in incompatible ways. Adding a framework for parameter domains and codomains could benefit users, preventing ineffective and illogical use of the model. This would require an investigation of the limits and influence of each parameter.

Conclusion

We have successfully implemented sEMG simulation model for synthetic data. The model comprises five modules, each aligned with the outlined objectives. The project achieved its primary goal and purpose, providing open-source code to the research field and developing our Python programming ability. Theoretical adherence has been validated multiple times by replicating key aspects of the cited literature and the decomposition algorithm output indicates the generated data contain relevant sEMG features. The model's assumptions and limitations have been acknowledged, like synergy between parts and sensitivity to parameter changes. The presented model is intended to be further developed by its users and several examples for future improvement have been presented. We argue the project's work contributes a valuable sEMG simulation tool to the research community.

References

Farina, D., Mesin, L., Martina, S., & Merletti, R. (2004, March). A surface EMG generation model with multilayer cylindrical description of the volume conductor. *IEEE Transactions on Biomedical Engineering, 51*(3), 415-426. Retrieved from https://doi.org/10.1109/TBME.2003.820998

Farina, D., Roberto, M., & Stegeman, D. F. (2004). Biophysics of the Generation of EMG Signals. In P. P. Roberto Merletti (Ed.), *Electromyography: Physiology, Engineering, and Noninvasive Applications* (pp. 81-105). Institute of Electrical and Electronics Engineers. Retrieved from https://doi.org/10.1002/0471678384.ch4

Foundation, P. S. (2023). *Python documentation version 3.11.5*. Retrieved from Python Software Foundation: http://www.python.org

Fuglevand, A. J., Winter, D. A., & Palta, A. E. (1993, Dec). Models of Recruitment and Rate Coding Organization in Motor-Unit Pools. *Journal of Neurophysiology, 70*(6), 2470-2488. Retrieved from https://doi.org/10.1152/jn.1993.70.6.2470

Harris, C. R., Millman, K. J., Van Der Walt, S. J., & et al. (2020). Array programming with NumPy. *Nature, 3*, 357-362. Retrieved from https://doi.org/10.1038/s41586-020-2649-2

Henneman, E., Somjen, G., & Carpenter, D. O. (1965, May 01). EXCITABILITY AND INHIBITIBILITY OF MOTONEURONS OF DIFFERENT SIZES. *Journal of Neurophysiology, 28*(3), 599-620. Retrieved from https://doi.org/10.1152/jn.1965.28.3.599

Holobar, A., Farina, D., & Zazula, D. (2016). Surface EMG Decomposition. In D. F. Roberto Merletti (Ed.), *In Surface Electromyography : Physiology, Engineering, and Applications* (pp. 180-209). The Institute of Electricaland Electronics Engineers, Inc. Retrieved from https://doi.org/10.1002/9781119082934.ch07

Huxley, A. F. (1974). Muscle contraction. *Journal of Physiology, 243*(1), 1-43. Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1330687/

Inc., T. M. (2023). *MATLAB Online version (R2023b, Update 1)*. Retrieved from MATLAB: https://www.mathworks.com

Lundsberg, J., Björkman, A., Malešević, N., & Antfolk, C. (2022). Compressed spike-triggered averaging in iterative decomposition of surface EMG. *Computer Methods and Programs in Biomedicine, 228*. Retrieved from https://doi.org/10.1016/j.cmpb.2022.107250

McDonald-Bowyer, A., Jager, K. D., Vanhoestenberghe, A., & Lancashire, H. (2020, 08 19). Matlab Model of Surface EMG. Retrieved from https://doi.org/10.6084/m9.figshare.12827132.v1

Merletti, R., Lo Conte, L., Avignone, E., & Guglielminotti, P. (1999). Modeling of surface myoelectric signals. I. Model implementation. *IEEE Transactions on Biomedical Engineering, 46*, 810-820. Retrieved from https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=771190

Moritani, T., Stegeman, D. F., & Merletti, R. (2004). Basic Physiology and Biophysics of EMG Signal Generation. In P. P. Roberto Merletti (Ed.), *Electromyography: Physiology, Engineering, and Noninvasive Applications* (pp. 1-25). Institute of Electrical and Electronics Engineers. Retrieved from https://doi.org/10.1002/0471678384.ch1

Pah, N. D. (2003). MATLAB™ Implementation of Merletti's SEMG Model. *Ubaya Repository*, 1-13. Retrieved from http://repository.ubaya.ac.id/8172/1/Paper2new.pdf

Rosenfalck, P. (1969). Intra- and extracellular potential fields of active nerve and muscle fibres. A physico-mathematical analysis of different models. *Acta physiologica Scandinavica. Supplementum, 321*, 1-168. Retrieved from https://api.semanticscholar.org/CorpusID:42446465

Servier. (n.d.). NEUROMUSCULAR SYNAPSE. *Figure modified with discriptive texts and lines, adaptation of "NEUROMUSCULAR SYNAPSE" from Servier Medical Art by Servier, licensed under a Creative Commons Attribution 3.0 Unported License*. Retrieved from https://creativecommons.org/licenses/by/4.0/

Vecchio, A. D., Holobar, A., Falla, D. L., Felici, F., Enoka, R. M., & Farina, D. (2020). Tutorial: Analysis of motor unit discharge characteristics from high-density surface EMG signals. *Journal of Electromyography and Kinesiology, 53*. Retrieved from https://doi.org/10.1016/j.jelekin.2020.102426

Appendix

Class and Method Structure

SurfaceEMG	__init__
	simulate_recruitment_model
	plot_recruitment_model
	simulate_surface_emg
	add_noise
	plot_surface_emg_array_no_noise
	plot_surface_emg_array
	plot_one_electrode_surface_emg_no_noise
	plot_one_electrode_surface_emg
MotorUnit	__init__
	get_tripole_amplitude
	plot_current_distribution_action_potential
	get_tripole_distance
	simulate_fibre_action_potential
	plot_fibre_action_potential
	simulate_motor_unit
	plot_motor_unit
SaveData	__init__
	save_output_data
	open_and_load_saved_data
	plot_saved_motor_unit
	plot_saved_surface_emg_array_no_noise
	plot_saved_surface_emg_array
	plot_saved_one_electrode_surface_emg_no_noise
	plot_saved_one_electrode_surface_emg

Parameters

SurfaceEMG	MotorUnit
Simulation time	AP amplitude constant
Sampling rate	Membrane resting potential
Excitatory drive function	AP scaling factor
Maximum excitation level	AP proportionality constant
Signal to noise ratio	AP plot length
Signal amplitude offset	Muscle fibre length
Number of motor units	Conduction velocity
Recruitment range	Ratio axial-radial conductivity
Excitatory gain	Radial conductivity
Minimum firing rate	Inter electrode spacing
Peak firing rate first MU	Number of electrodes z-axis
Peak firing rate difference	Number of electrodes x-axis
Inter-spike-interval coefficient	Electrode shift z-axis
Twitch force range	Neuromuscular junction position
Motor unit density	Y-axis fibre position variation
Smallest MU number of fibres	X-axis fibre position variation
Largest MU number of fibres	Extermination zone width
Muscle fibre diameter	Innervation zone width
Muscle diameter	Time muscle fibre simulation
Muscle depth	Motor unit radius
Number of electrodes z-axis	Number of fibres
Number of electrodes x-axis	Motor unit y-position
Y-axis plotting limit	Motor unit x-position
X-axis plotting limit	Y-axis plotting limit
	X-axis plotting limit