



**POLITECNICO DI TORINO**

**DIPARTIMENTO DI INGEGNERIA MECCANICA E  
AEROSPAZIALE (DIMEAS)**

Master's Degree in Biomedical Engineering

Master Degree Thesis

**Machine learning algorithms for  
facial gesture recognition: a first  
analysis based on event-driven  
sEMG acquisition**

**Supervisors**

Prof. Danilo DEMARCHI  
Ph.D. Paolo MOTTO ROS  
M.Sc. Fabio ROSSI  
M.Sc. Andrea MONGARDI

**Candidate**

Angela CARNEVALE

Torino, March 2021



*Ai mie nipoti, Remo ed Antonio.*

*"To strive, to seek, to find, and not to  
yield."*

*– Alfred Lord Tennyson*



# Abstract

Facial gesture recognition has wide application in Human-Machine Interaction (HMI), which, in the medical area, can be identified with behavioral and emotional analyses, as well as rehabilitative procedures. Although historical approaches for facial expression recognition rely on videos and images data, in recent years, with the progress of the sensors technology and Machine Learning (ML) algorithms, recognition is also being achieved using biological signals, as surface ElectroMyoGraphic (sEMG) signal.

The thesis focuses on recognizing and classifying jaw movements and facial expressions from sEMG signals recorded by face muscles during the execution of such actions. The innovative event-driven technique, named Average Threshold Crossing (ATC), is applied to the amplified and filtered sEMG signal to extract the ATC feature. This feature is computed by averaging the events generated when an sEMG signal exceeds a voltage threshold on a predefined time window. Past works demonstrated the benefits of the ATC technique in terms of reduction of data processing, transmission and related power consumption, allowing it to be an optimal solution in the development of wearable, miniaturized and energy-efficient data acquisition system. With the aim to develop an ATC-based facial network, the thesis' goal is to understand whether the ATC approach is suitable for the recognition of facial gestures.

A first step towards this direction was to define which were the movements to be recognized, the corresponding musculature and, consequently, the electrodes position for proper signals detection. A preliminary test, beside confirming the feasibility of this idea, was needed to organized the sensors location of the facial network better. At this point, with the goal to train facial expression classifiers, a data collection was launched, involving 21 subjects. Each subject performed a list of eight gestures for different session in order to obtain a robust dataset. The raw sEMG signals have been recorded using the g.HIamp-Research amplifier; then, data were processed to extract the ATC parameter, used as input of a classifier.

In fact, several Machine Learning algorithms have been implemented to recognize

facial movements: Random Forest (RF), Support Vector Machine (SVM), k-Nearest Neighbour (k-NN), and Artificial Neural Networks (ANN). All the four classifiers perform the recognition obtaining similar accuracies. In particular, they reached an overall percentage of success greater than 60% when recognizing eight expressions. In comparison, they improve their average recognition rate up to 75% when two not well-defined expressions are removed from the dataset. These percentages will pave the way to the application of the ATC technique to facial gesture recognition.

# Acknowledgements

Ringrazio il Prof. Danilo Demarchi e il Dott. Paolo Motto Ros per avermi dato l'opportunità di lavorare a questo progetto. Un ringraziamento particolare va a Fabio Rossi e Andrea Mongardi per la disponibilità, il sostegno, l'aiuto, i preziosi consigli ricevuti e per avermi guidato durante tutto il periodo di tesi.

Ringrazio i compagni di tesi e tutto il VLSI per essere stati parte attiva del mio lavoro, sottoponendosi ai miei, a volte, non molto piacevoli test.

Ora, mi tocca esprimere almeno a parole, non essendo molto espansiva e amorevole nei gesti, la mia gratitudine ai miei affetti più cari, che mi hanno sostenuto e sopportato in questi anni.

Inizio con il ringraziare i miei genitori, che mi hanno permesso con il loro aiuto di arrivare fino a qui. Grazie per essere riusciti sempre a capirmi e ad appoggiare tutte le mie scelte.

Grazie a mia sorella e a mio fratello, per avermi detto sempre la giusta parola nel momento opportuno e avermi spronato ad inseguire tutto ciò che potesse rendermi felice, nonostante le difficoltà che dovevano essere affrontate. Spero che ora che la mia vita da studentessa è finita riusciate a chiamarmi con il mio vero nome e mettere da parte il nomignolo che mi avete affibbiato sin dalla nascita.

Grazie Beatrice, presenza costante in qualsiasi ricordo, per essermi stata accanto nei momenti più tristi e difficili da affrontare, nonostante il mio silenzio, e per aver condiviso con me gli innumerevoli aspetti felici della vita. Semplicemente grazie per ciò che hai fatto ieri e che continui a fare oggi malgrado la distanza.

Un ringraziamento speciale va ai miei compagni universitari: Chiara, protagonista indiscussa dei miei innumerevoli scherzi, Massimo, compagno di soprannomi e Lucia, complice di sbadataggine. Senza di voi le giornate e gli esami universitari sarebbero stati troppo pesanti da affrontare. Porterò sempre con me una parte di voi.

Grazie Alessia, per i sorrisi e i momenti di allegria soprattutto nei periodi in cui l'Angela "felice" veniva meno. In questi anni a Torino, mi sono mancati la sfogliatella sul lungomare di Mergellina dopo un esame, le serate alla Cumana, le cene a casa Monaco e il pollo al radicchio che costringevi tutti a mangiare per una settimana. Sei sempre stata nei miei pensieri.

Infine, ringrazio tutta la mia grande famiglia e gli amici di infanzia, per avermi incoraggiato sempre e accolto allegramente ad ogni mio ritorno a casa.

*Angela*

# Contents

<b>Abstract</b>	VI
<b>Acknowledgements</b>	VII
<b>List of Figures</b>	XII
<b>List of Tables</b>	XIV
<b>1 Introduction</b>	1
1.1 Muscular System . . . . .	1
1.1.1 The Skeletal Muscle . . . . .	2
1.1.2 Skeletal Muscle Contraction . . . . .	5
1.1.3 Facial muscles . . . . .	5
1.2 ElectroMyoGraphic (EMG) Signal . . . . .	9
1.2.1 The surface EMG signal . . . . .	11
1.2.2 sEMG Feature Extraction . . . . .	12
1.2.3 Facial sEMG . . . . .	14
1.3 The Average Threshold Crossing Technique (ATC) . . . . .	15
1.4 Classification Algorithms . . . . .	16
1.4.1 Artificial Neural Networks (ANN) . . . . .	16
1.4.2 Support Vector Machine (SVM) . . . . .	17
1.4.3 k-Nearest Neighbour (k-NN) . . . . .	19
1.4.4 Decision Tree (DT) . . . . .	20
1.4.5 Random Forest (RF) . . . . .	21
1.5 Machine Learning Metrics . . . . .	22
<b>2 State of art</b>	24
2.1 ATC technique applied to sEMG . . . . .	24
2.2 Facial sEMG Applications . . . . .	26
2.3 Facial gesture recognition . . . . .	27
2.3.1 Video and images Applications . . . . .	28
2.3.2 sEMG Applications . . . . .	30

<b>3</b>	<b>Preliminary studies</b>	<b>33</b>
3.1	Equipment specifications . . . . .	33
3.1.1	gHIamp-Research . . . . .	33
3.1.2	g.Recorder Software . . . . .	35
3.1.3	sEMG Quality Improvement . . . . .	36
3.1.4	Post-processing for ATC Extraction . . . . .	37
3.2	Movements Invenstingations . . . . .	37
3.2.1	Jaw movements . . . . .	37
3.2.2	Singing and talking . . . . .	40
3.2.3	Bite Analysis . . . . .	42
3.2.4	Facial expressions . . . . .	45
3.2.5	Cought gesture . . . . .	47
<b>4</b>	<b>Data Acquisition</b>	<b>50</b>
4.1	Performed Movements . . . . .	50
4.2	Electrodes Placement . . . . .	51
4.3	Acquisition Protocol . . . . .	52
4.3.1	Experimental Protocol: Graphical User Interface (GUI) . . . . .	53
4.3.2	Experimental Protocol: Phases . . . . .	56
<b>5</b>	<b>Processing Data</b>	<b>58</b>
<b>6</b>	<b>Machine Learning Implementation</b>	<b>64</b>
6.1	Parameters Algorithms . . . . .	65
6.1.1	Random Forest (RF) . . . . .	65
6.1.2	k-Nearest Neighbour (k-NN) . . . . .	66
6.1.3	Support Vector Machine (SVM) . . . . .	67
6.1.4	Artificial Neural Networks (ANN) . . . . .	67
6.2	Total Gestures Recognition . . . . .	68
6.2.1	Random Forest (RF) . . . . .	68
6.2.2	k-Nearest Neighbour (k-NN) . . . . .	69
6.2.3	Artificial Neural Networks (ANN) . . . . .	70
6.2.4	Support Vector Machine (SVM) . . . . .	71
6.3	Reduced Gestures Recognition . . . . .	72
6.3.1	Random Forest (RF) . . . . .	72
6.3.2	k-Nearest Neighbour (k-NN) . . . . .	73
6.3.3	Artificial Neural Networks (ANN) . . . . .	74
6.3.4	Support Vector Machine (SVM) . . . . .	74
6.4	Jaw Movements Recognition . . . . .	75
6.4.1	Random Forest (RF) . . . . .	75
6.4.2	k-Nearest Neighbour (k-NN) . . . . .	76
6.4.3	Artificial Neural Networks (ANN) . . . . .	77
6.4.4	Support Vector Machine (SVM) . . . . .	77

6.5	Recognition Reinforcement Model . . . . .	78
6.5.1	Random Forest (RF) . . . . .	78
6.5.2	k-Nearest Neighbour (k-NN) . . . . .	79
6.5.3	Artificial Neural Networks (ANN) . . . . .	80
6.5.4	Support Vector Machine (SVM) . . . . .	81
<b>7</b>	<b>Conclusion and Future Perspective</b>	<b>83</b>
	<b>Bibliography</b>	<b>85</b>

# List of Figures

1.1	Different types of muscular tissue . . . . .	2
1.2	Skeletal muscle overview . . . . .	2
1.3	Inner view of a muscle fiber . . . . .	3
1.4	Sarcomere . . . . .	4
1.5	Motor Unit anatomy . . . . .	5
1.6	Mimics muscles . . . . .	6
1.7	Jaw-closing and masticatory muscles . . . . .	8
1.8	Jaw-opening muscles . . . . .	8
1.9	Skeletal muscle action potential . . . . .	9
1.10	Intramuscular ElectroMyoGraphy (iEMG) . . . . .	10
1.11	Surface ElectroMyoGraphy (sEMG) . . . . .	11
1.12	Example of pre-gelled electrode . . . . .	12
1.13	Example of dry electrode . . . . .	12
1.14	sEMG sampling vs ATC sampling. . . . .	15
1.15	Example of artificial neural network architecture . . . . .	17
1.16	Optimal hyperplane using the SVM algorithm . . . . .	18
1.17	Representation of a non-linear model and its equivalent linear . . . . .	18
1.18	Different types of kernel . . . . .	19
1.19	Example of a k-NN classification. . . . .	19
1.20	Example of a Decision Tree . . . . .	20
1.21	Example of Random Forest algorithm . . . . .	22
2.1	ATC, Force, ARV signals in time domain . . . . .	24
2.2	Architecture of a gesture recognition system . . . . .	28
2.3	Emotracker system . . . . .	29
2.4	Headband . . . . .	31
3.1	g.HIamp-Research multi-channel system . . . . .	34
3.2	Markers definition . . . . .	36
3.3	Electrodes used for sEMG acquisition. . . . .	36
3.4	Preliminary electrodes placement . . . . .	38
3.5	sEMG signals during jaw movements. . . . .	39

3.6	ATC envelope during jaw movements. . . . .	39
3.7	Electrodes placement on <i>superior orbicularis oris</i> . . . . .	40
3.8	The performing of singing movement - sEMG signals. . . . .	41
3.9	The performing of talking act - sEMG signals. . . . .	41
3.10	The performing of talking act - ATC signals envelope. . . . .	42
3.11	Main features of the masticatory muscles. . . . .	43
3.12	sEMG masticatory signals . . . . .	44
3.13	ATC masticatory envelopes . . . . .	45
3.14	Electrodes placement on the mimics muscles . . . . .	46
3.15	sEMG signals during facial expressions protocol. . . . .	46
3.16	ATC envelopes during facial expressions protocol. . . . .	47
3.17	sEMG signals during cough. . . . .	48
3.18	ATC signals envelope during cough. . . . .	48
4.1	Performed movements . . . . .	51
4.2	Electrode positioning setup on the right hemiface. . . . .	52
4.3	Start screen . . . . .	54
4.4	Language choice . . . . .	54
4.5	Experimental protocol view screen . . . . .	55
4.6	Pop-up request for protocol continuation . . . . .	56
4.7	Exit pop-up . . . . .	56
4.8	Recording of the sEMG signal seen in real-time on g.Recording. . . . .	57
5.1	Settling time g.tec system. . . . .	58
5.2	Frequency response Butterworth filter. . . . .	59
5.3	sEMG signals pre- and post- digital filtering. . . . .	59
5.4	g.Recorder screen. . . . .	60
5.5	sEMG signal characterized by a high environmental noise. . . . .	61
5.6	TC events superimposed on the sEMG signal. . . . .	62
5.7	ATC signals envelope of facial protocol. . . . .	62

# List of Tables

3.1	g.HIamp-Research amplifier techical specification. . . . .	35
6.1	RF hyperparameters setting. . . . .	68
6.2	RF results. . . . .	69
6.3	k-NN hyperparameters setting. . . . .	69
6.4	k-NN results. . . . .	70
6.5	ANN hyperparameters setting. . . . .	70
6.6	ANN results. . . . .	71
6.7	SVM hyperparameters setting. . . . .	71
6.8	SVM results. . . . .	72
6.9	Better compromise of parameters for Random Forest. . . . .	72
6.10	Statistical results Random Forest. . . . .	73
6.11	Better compromise of parameters for k-NN. . . . .	73
6.12	Statistical results k-NN. . . . .	73
6.13	Better compromise of parameters for ANN. . . . .	74
6.14	Statistical results ANN. . . . .	74
6.15	Better compromise of parameters for SVM. . . . .	75
6.16	Statistical results SVM. . . . .	75
6.17	Random Forest best parameters setup. . . . .	76
6.18	Performances obtained with Random Forest. . . . .	76
6.19	k-NN best parameters setup. . . . .	76
6.20	Performances obtained with K-NN. . . . .	77
6.21	ANN best parameters setup. . . . .	77
6.22	Performances obtained with ANN. . . . .	77
6.23	SVM best parameters setup. . . . .	78
6.24	Performances obtained with SVM. . . . .	78
6.25	RF hyperparameters. . . . .	79
6.26	RF prediction results. . . . .	79
6.27	k-NN hyperparameters. . . . .	80
6.28	k-NN prediction results. . . . .	80
6.29	ANN hyperparameters. . . . .	80
6.30	ANN prediction results. . . . .	81

6.31 SVM hyperparameters. . . . . 81  
6.32 SVM prediction results. . . . . 82





# Chapter 1

## Introduction

### 1.1 Muscular System

The muscular system allows the movement of the body. It ensures the proper operation of the internal organs to perform some vital functions, such as eating and swallowing food, breathing, grasping objects, etc. According to the different functional and anatomical characteristics, muscle tissue can be of three types [1]:

- **Skeletal muscle** is the active component of the locomotor system. The skeletal musculature can wiggle the body by moving the bones and maintain posture in various positions through its connection with the skeleton. Its contraction is under the control of the Central Nervous System (CNS), it means depending on the individual's will. Muscle fibers are single, tubular, multinucleated, and striated.
- **Cardiac muscle** is only in the heart. Its involuntary contraction allows the blood to pump through the heart and into and through arteries and veins of the circulatory system and also the maintenance of blood pressure. It is composed of branched chains of uni- or binucleate cells with obvious striations.
- **Smooth muscle** is in the hollow organs (e.g., intestines, blood vessels, respiratory tubes), except in the heart. It is innervated by the Autonomic Nervous System (ANS), and it is responsible for the mobility of the organs that perform vegetative functions that are not under voluntary control. Unlike the other types of tissue, the fibers are nonstriated, grouped, and fusiform.

For completeness, it is useful to have a more in-depth view of the peculiar characteristics of skeletal muscle, which are related to movement, and more importantly, to know the sEMG signal associated with it.

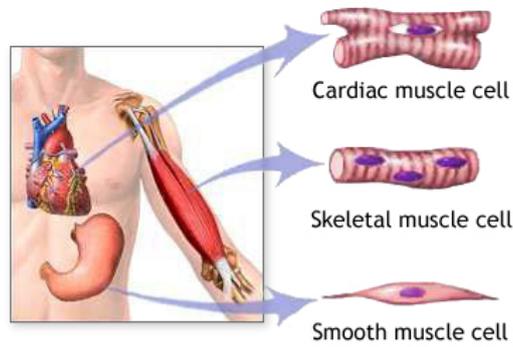


Figure 1.1: Different types of muscular tissue [2].

### 1.1.1 The Skeletal Muscle

Skeletal muscles under the Nervous System's direction generate the force, known as *muscular force*, needed to stabilize the skeleton in a wide range of conditions. This force's intensity depends on the muscle's architecture involved in a specific movement, particularly on its cross-sectional area and its pennation angle [3].

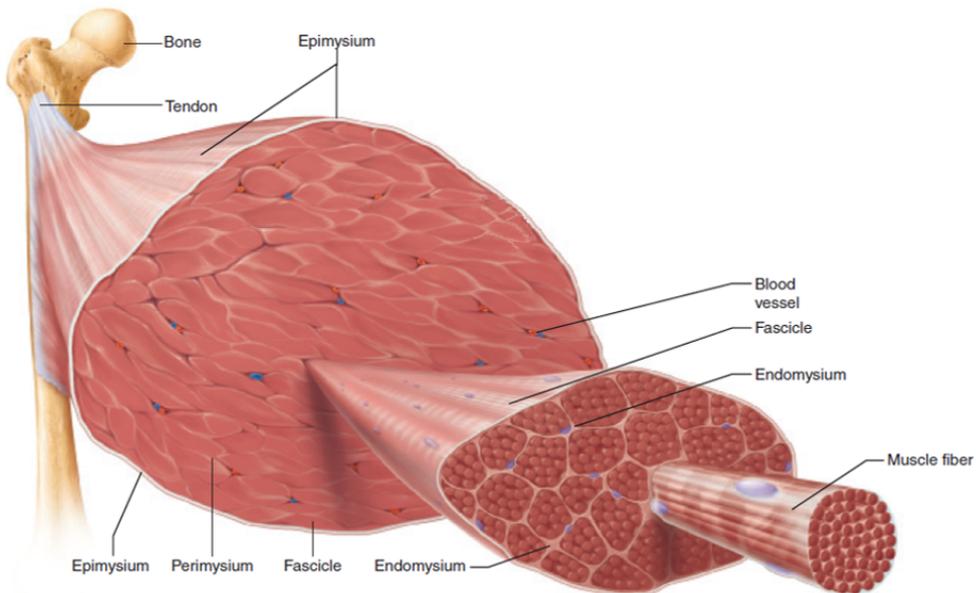


Figure 1.2: Skeletal muscle overview. Adapted from [1].

The skeletal muscle is externally surrounded by *epimysium* and consists of connective tissue wrappings, blood vessels, and individual muscle fibers bundled into fascicles. Each fiber group is encircled by a fibrous connective tissue layer, called *perimysium*, which separates individual fascicles from one another. Each fiber is protected by the *endomysium*, the innermost membrane of the muscle, and the extracellular fluid and the nutrients needed for the fibers' survival within the fascicle. All three sheaths may extend beyond the muscle fibers to form a *tendon*, which attaches the muscle to bones (shown in Figure 1.2).

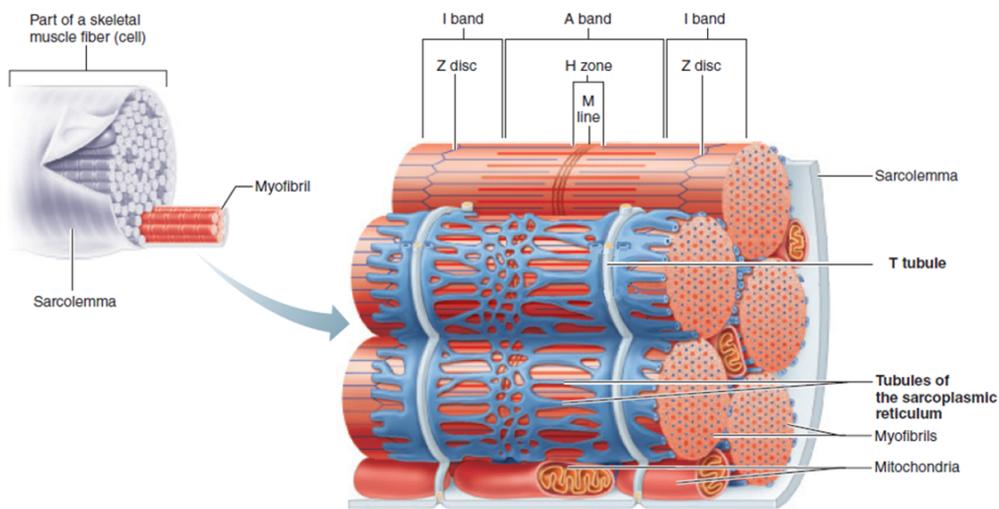


Figure 1.3: Inner view of a muscle fiber. Modified from [1].

In the inner part of the muscle, multinuclear cells are located just beneath the *sarcolemma*, the cell membrane of muscle fiber. The sarcolemma propagates the nerve impulse in the fiber and is connected to a system of *transverse (T) tubules*, filled with extracellular fluid. The T-tubule system acts as a channel for ions flow and, along with the sarcolemma, regulates muscle contraction, ensuring that action potential reaches all muscle fiber parts simultaneously. Within the sarcolemma is the *sarcoplasm*, the cytoplasm of a muscle fiber. Sarcoplasm contains a substantial amount of glycogen, myoglobin, mitochondria, and a large number *myofibrils*.

A myofibril, surrounded by *the sarcoplasmic reticulum*, is the functional unit of a muscle cell composed of repeating segments called *sarcomeres*. By observing the longitudinal section of the muscle, as shown in Figure 1.4, a sarcomere is made up of parallel overlapping matrices of myosin and actin, which are proteins shaped in thick and thin filaments, respectively.

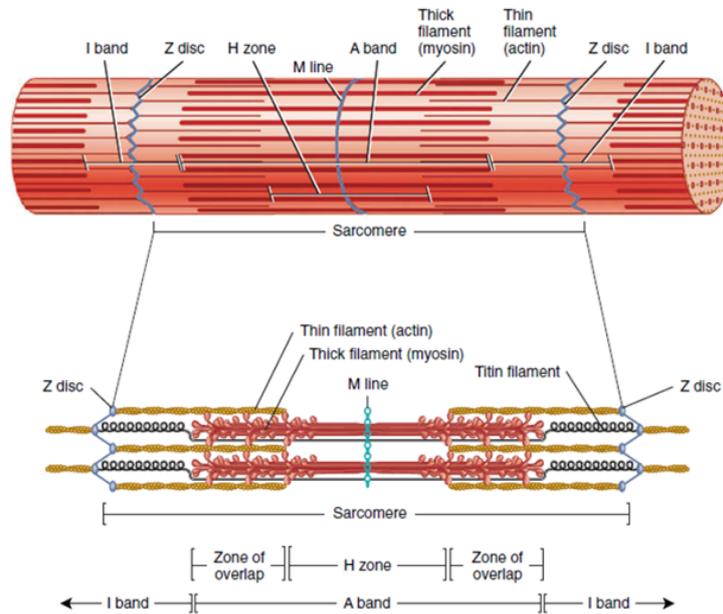


Figure 1.4: Sarcomere [4].

It is separated from other contractile units by dense zig-zagging protein-based structures called **Z discs**.

Other components of the sarcomere, which can be identified observing it through a microscope, are:

- **A band:** is the central darker part of the sarcomere consisting of thick filaments and also includes regions where there is an overlapping of actin and myosin filaments.
- **I band:** is the lighter area due to contains only thin filament. At the center of each I band is a Z disc.
- **H band:** is the region in the center of A band and is composed of thick filaments.
- **M line:** is a thickened area in the H zone with proteins that hold the filaments together in place.

The striped appearance of the fiber results from the alternation of light and dark bands. Besides, the overlap of the various thin and thick filaments derives from the state of the muscle, whether it is contracted, relaxed, or stretched [4].

### 1.1.2 Skeletal Muscle Contraction

The Central Nervous System (CNS) activates muscle fibers by sending signals to the nerves of specific skeletal muscles. Muscle contraction begins when the nervous system generates a signal, known as Action Potential, which spreads via a motor neuron and through its axons reaches muscle fibers at the neuromuscular junction. In the junction, a neurotransmitter called *acetylcholine* is released in response to changes in the properties of the sarcolemma, which triggers the contraction of the fibers following the interaction of calcium ions with the actin filaments.

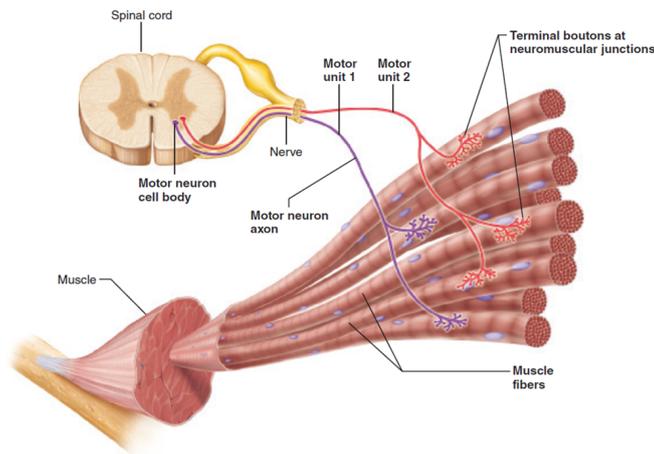


Figure 1.5: Motor Unit anatomy. Adapted from [1].

The functional unit, composed of a single motor neuron and a group of fibers, is called Motor Unit (MU). All the skeletal muscle fibers in the MU work together. When the muscle fiber is activated, a small amount of force is generated since a muscle contraction (also called twitch) has occurred. A more significant amount of force is generated within the whole muscle if more motor neurons are recruited, which activate more muscle fibers. Generally, the number of fibers in a MU may vary from a few units up to several hundred fibers depending on the movement. The control of muscle precision and strength depends on the number of fibers that respond to the stimulus. Precise control is required by the muscles with small dimensions, such as extraocular muscles. As a result, fewer fibers are involved. The opposite occurs in powerful movements (e.g., hip movement) performed by weight-bearing and bulky muscles [1] [5].

### 1.1.3 Facial muscles

The purpose of this study is the analysis of sEMG signals acquired from the muscles of the face and neck while performing certain movements. Good knowledge of

these muscles allows to better examine the results obtained and to correctly place the electrodes during the recording sessions. The focus is mainly on the muscles involved in facial expressions, jaw movements and, chewing. In the following pages, they are listed and briefly described.

### Mimics muscles

The muscles responsible for expressing emotions have a thin and variable shape so that the adjacent muscles often merge. They differ from other skeletal muscles due to they are inserted on the skin and not on the skeleton.

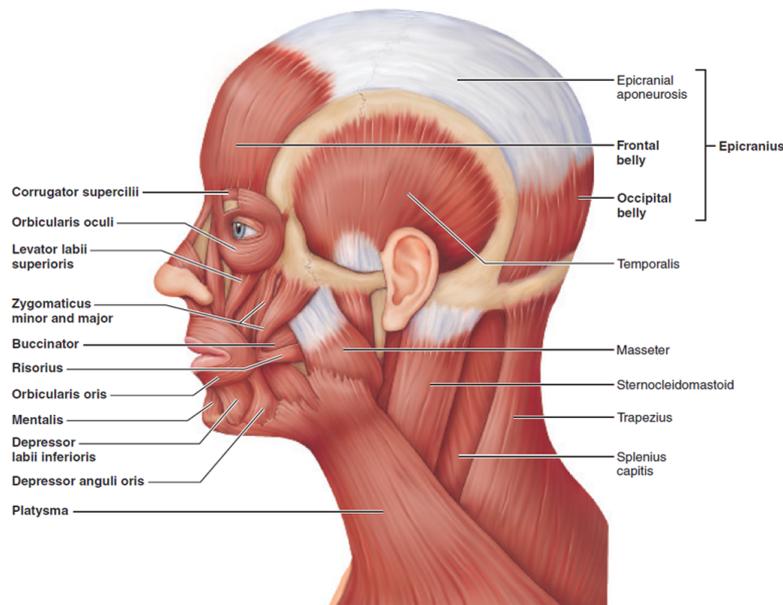


Figure 1.6: Mimics muscles [1].

- **Orbicularis oculi** is a thin sphincter muscle that is arranged around the eyelids and surrounds the orbit's rim. Its main function is closing eyes such as blinking and squinting.
- **Orbicularis oris** is a multilayered muscle which encircles the mouth and is arranged in concentric fibers bands running in different directions, mostly circularly. It allows closing and puckering lips in the act of whistling, sucking, and kissing.
- **Corrugator supercilii** is a small, pyramidal-shaped facial muscle located at the medial end of the eyebrow. It moves, together with the Orbicularis Oris, the eyebrows medially and inferiorly, towards the nose and the inside of the eye, creating wrinkles skin vertically on the forehead.

- **Zygomaticus** is a muscle pair that stretches diagonally from the cheekbone to the corner of the mouth. It raises the labial commissure superiorly and laterally.
- **Epicranius (Occipitofrontalis)** is a bipartite muscle that includes frontal and occipital bellies attached by the epicranial aponeurosis. The alternation of their actions pulls the scalp anteriorly and posteriorly.
- **Mentalis** is a muscle located on the chin's tip and represents the primary muscle of the lower lip movement. It is activated when the lower lip is pushed forward, causing some wrinkles on the chin.
- **Platysma** is an unpaired and superficial anterior neck muscle. It is involved in the lowering of the jaw and the corner of the mouth, tightening the lower face's skin and the front of the neck.
- **Risorius** is located in the lower part of the face in a lateral position below the Zygomaticus. Its action is to retract the corners of the mouth during a smile, laughter, and grimace. Unlike the smile resulting from the action of the Zygomaticus, it, not involving the Orbicularis oculi, does not produce the wrinkles around the eyes, known as "crow's feet".
- **Levator labii superioris** is a slender muscle between the orbicularis oris and the inferior eye's margin. It is responsible for opening the lips by elevating and furrowing the upper lip.
- **Depressor labii inferioris** is a small muscle that runs from the jaw to the lower lip. It is activated when the lower lip is lowered by pouting.  

item textbf Depressor anguli oris is a muscle located lateral to the Depressor Labii Inferioris. It is a zygomatic antagonist muscle that curves the corners of the mouth when the forehead is raised and the eyebrows are wrinkled.
- **Buccinator** is a thin, horizontal cheek muscle that pushes the cheek towards molars. It works with tongue to keep food between occlusal surfaces and out of the oral cavity during the chewing (it is also included in the masticatory muscles). By contracting bilaterally, it also allows blowing.

### Muscles of jaw movements

The muscles involved in the jaw movements can be divided into two groups: one related to the closing of the jaw and the mastication function, the other one to the opening movements.

- **Jaw-closing and masticatory muscles**

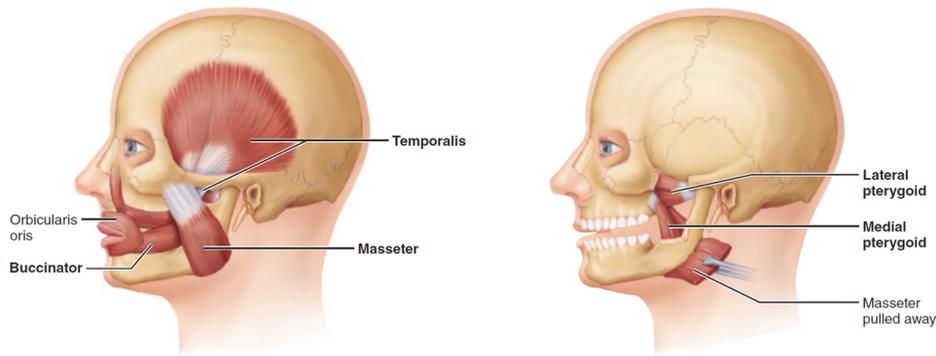


Figure 1.7: Jaw-closing and masticatory muscles [1].

- **Temporal** is fan-shaped with horizontal, vertical, and oblique fibers. It is involved in the retraction and elevation of the mandible, causing the jaw-closing.
  - **Masseter** is square in shape and placed laterally to the ramus of mandible. It allows the elevation of the jaw and it is also responsible for the biting force, involved in the gnashing of teeth both in the static form (clenching) and in the dynamic one (bruxism).
  - **Lateral (or External) pterygoid** consists of two bundles (upper and lower), with two separate origins and a single insertion site. It aids to chew and grind food with lower teeth, protracting mandible and moving it side-to-side.
  - **Medial (or Internal) pterygoid** is a thick two-headed muscle and a quadrilateral in shape, which is widely concealed by the mandible since it runs along the internal surface of that bone. It is involved with the lateral pterygoid in the jaw's elevation, and its action balances the action of the masseter.
- **Jaw-opening muscles**

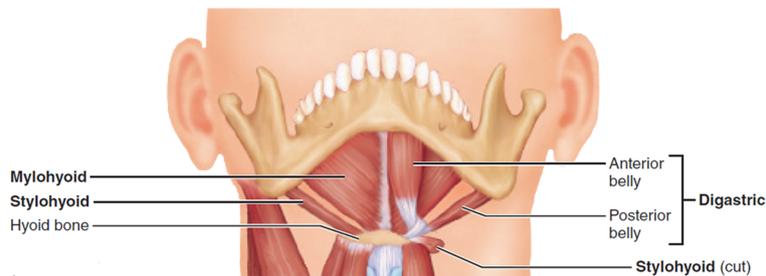


Figure 1.8: Jaw-opening muscles. Adopted from [1].

- **Digastric** includes two bellies connected by a tendon that slides in a fibrous ring attached to the bone hyoid. It is involved in the lowering and the retrusion of the mandible, leading to the elevation of the hyoid bone and its stabilization during phonation and swallowing.
- **Stylohyoid** is located above the posterior belly of the digastric muscle. It retracts the hyoid bone posteriorly and superiorly during a swallowing action.
- **Mylohyoid** is a triangular-shaped muscle between the mandible and the hyoid bone which forms the floor of the mouth. It lowers the mouth floor and, therefore, of the jaw during chewing, raises the hyoid bone when the jaw is fixed (swallowing), and steady during phonation.
- **Geniohyoid** is a long, thin, cylindrical muscle that extends from the chin to the hyoid bone.

## 1.2 ElectroMyoGraphic (EMG) Signal

The study of muscle electrical activity, caused in the MU by the CNS (discussed in Sec.1.1.2), can be performed through the *ElectroMyoGraphic (EMG)* technique which provides information about the control and execution of the movements. The set of several MUs, activated simultaneously by a moderate force contraction, create a Motor Unit Action Potential (MUAP). The waveform of a MUAP, which is the space-time summation of the individual AP produced by the muscle fibers' depolarization, provides information on how the MUs are organized within it. The single AP is structured as shown in Fig. 1.9: the transmembrane differential voltage is  $-70\text{mV}$  (resting state), without any conduction.

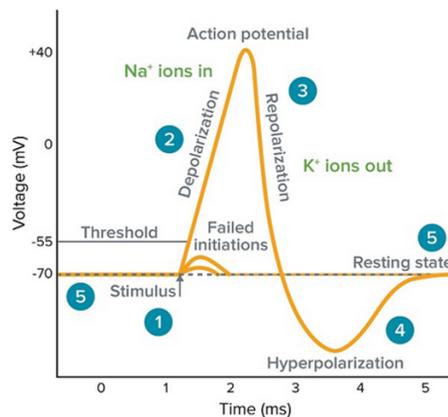


Figure 1.9: Skeletal muscle action potential [6].

Then, it increases until a threshold of  $-55\text{mV}$  is achieved which triggers the AP to

reach a maximum of 30 mV. At this point, there is a depolarization phase, followed by a repolarization phase until the original potential is restored.

A single UM contraction undergoes a spatial-temporal dispersion since it is not transmitted simultaneously to all the fibers present within it. It can also involve the fibres belonging to neighbouring UMs. The shape, phase, and duration of MUAPs can be influenced by some aspects such as the distance of the electrodes from the motor unit, the type of electrodes, the equipment used [7].

There are two different types of EMG signal:

- **Intramuscular ElectroMyoGraphy (iEMG)**: is an invasive technique since the electrodes used to record the signal are small and inserted directly into the muscle. The adoption of needle electrodes allows detecting the single MUAP obtaining a good muscle selectivity and a high signal-to-noise ratio. iEMG is used for clinical diagnosis, recording signals from deep or slender muscle (e.g., diagnosis of myopathies, diseases of the neuromuscular junction). This technique's disadvantages are related to the too-small size of the registration region, making it more difficult to reposition the needle to detect the same motor units. Furthermore, the extraction of parameters representative of the fiber membrane properties is made difficult [8].

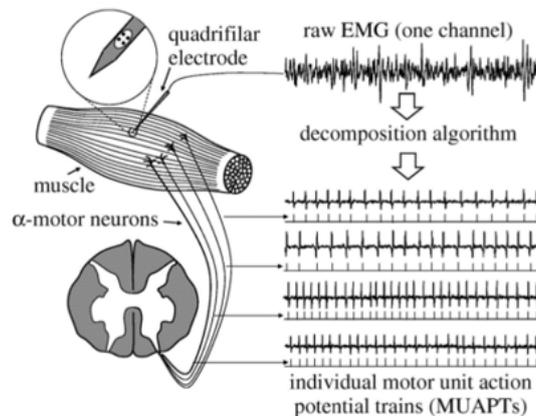


Figure 1.10: Intramuscular ElectroMyoGraphy (iEMG) [8].

- **Surface ElectroMyoGraphy (sEMG)**: is the non-invasive study of muscle functioning through electrodes placed on the skin overlying the muscle. It is used to evaluate the duration of muscle contraction, the action of different muscles during body movements, and the process of muscle fatigue. Signal detection is performed on a relatively large area of the muscle. The bio-signal, therefore, can contain the crosstalk originating from the surrounding muscles since the recorded signal is the sum of all the MUs (as shown in Fig. 1.11),

below the electrode recruited in the movement. In contrast to iEMG, sEMG is safer and easier to use but less accurate.

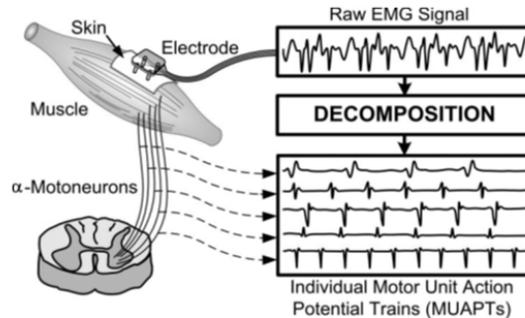


Figure 1.11: Surface ElectroMyoGraphy (sEMG) [9].

### 1.2.1 The surface EMG signal

The surface ElectroMyoGraphic (sEMG) signal has an amplitude that can vary between 0 mV and 10 mV and a frequency spectrum ranges between a few Hz to 500 Hz, with a significant contribution in the 20Hz - 150Hz range. Unfortunately, this signal can be easily subject to different noises:

- **motion artifact** derives from a relative motion in the skin-electrode interface, caused by muscle contraction, which changes the electrode charge layer, and by the deformation of skin below the electrode. It is a low-frequency noise ( $< 20\text{Hz}$ ) and can be damped by using a separating conductive gel and easy filtering in the pre-processing, being out of the sEMG spectrum;
- **cross-talking** is due to the electrical activity of the muscles close to the one being evaluated. It cannot be filtered, but the likelihood of detecting it can be reduced by carefully placing the electrode, usually in the midline of the abdominal muscle;
- **external interference** occurs due to the setup of the laboratory environment and electronics equipment used to detect, amplify, and record the signals. In particular, it mainly derives from the irradiation of the power line (50Hz) and is often maintained in the circuit because it is in the frequency spectrum;
- **muscular fatigue** occurs when muscle activity decreases and causes the signal to have a lower frequency and amplitude.

Currently, *wet* gel electrodes and *dry* metal electrodes are used to record the sEMG signal, depending on the type of exam which needs to be performed.

Wet electrodes use an electrolytic gel as an interface between the skin and the electrode's metal part. Particular chemical redox reactions occurs in metal-gel interface. For electrode metal part, the most used materials are silver/silver chloride (Ag/AgCl), silver chloride (AgCl) but also silver (Ag) or gold (Au). They can be disposable or reusable. The disposable ones are the most common as they are very light and in a wide range of shapes and sizes. Also, those with a built-in gel layer (Fig.1.12) are available to reduce application times.



Figure 1.12: Example of pre-gelled electrode [10].

The dry electrodes (Fig.1.13) are in direct contact with the skin without gel, providing an ideal condition for long-lasting recordings. They are heavy, inflexible, and suffer from high movement artifacts because they do not have a good adaption to the skin's contours. The absence of the gel prevents the subject from discomfort due to skin irritation but causes an increase in skin-electrode impedance. A signal conditioning circuit is needed to keep the impedance under control.



Figure 1.13: Example of dry electrode [11].

### 1.2.2 sEMG Feature Extraction

Feature extraction is a method for extracting relevant and useful information from sEMG signal. Three categories of features can be distinguished: time domain, frequency domain, and time-frequency domain. The mixed time-frequency domain (such as the wavelet transform) require a long computation time, making their use

difficult in real-time applications, for therapeutic devices, and classifier training. And as a result, the latter are not described in this introduction [12].

### Time domain features

The temporal features have as main characteristics, the straightforward implementation and the processing speed, the need for any transformation. The main disadvantages are the non-stationary characteristics of the signal and muscle crosstalk's influence due to their calculation essentially based on signal amplitude values.

- **Integrated EMG (IEMG)**: is the integral of the full-wave rectified sEMG signal, representing the area under the curve of the rectified signal. It is employed for muscle activity as a pre-activation index.

$$IEG = \sum_{i=1}^N |x_i|$$

Where N indicates the length of a signal segment, i is the segment increment and  $x_i$  denotes the value of the signal amplitude in that segment.

- **Mean Absolute Value (MAV)**: is defined as the moving average of the absolute values of the sEMG signal. It is used to detect and gauge different levels of muscle contraction.

$$MAV = \frac{1}{N} \sum_{i=1}^N |x_i|$$

Where the terms have the same meaning as above.

- **Mean Absolute Value Slope(MAVS)**: estimates the difference between the MAVs of the neighboring segments of the sEMG signal.

$$MAVS_i = MAV_{i+1} - MAV_i$$

Where i is the index of MAV taken into account.

- **Simple Square Integral (SSI)**: is the sum of the square values of the sEMG signal amplitude and correlated to the energy of the signal itself.

$$SSI = \sum_{i=1}^N |x_i|^2$$

- **Variance of EMG (MAVS)**: is useful for measuring signal power. It can be mathematically represents as:

$$MAV = \frac{1}{N-1} \sum_{i=1}^N x_i^2$$

- **Root Mean Square (RMS)**: is the square root of the average power of the EMG signal for a fixed time window.

$$RMS = \sqrt{\frac{1}{N} \sum_{i=1}^N |x_i|^2}$$

- **Waveform Length (WL)**: is the measure of EMG signal complexity in terms of waveform amplitude, frequency, and duration. It is defined as the cumulative length of the signal waveform.

$$WL = \sum_{i=1}^{N-1} |x_{i+1} - x_i|$$

### Frequency domain features

Features in the spectral or frequency domain are mainly used for muscle fatigue studies and motor unit recruitment analysis [13]. Compared to the features of the time domain, they require higher computation time and resources, considering that they are based on the estimate of the power spectral density (PSD).

- **Frequency Median (MDF)**: is a frequency value at which the EMG power spectrum is divided into two equal parts.

$$\sum_{i=1}^{FMD} P_j = \sum_{i=FMD}^M P_j = \frac{1}{2} \sum_{i=1}^M P_j$$

Where  $P_j$  is the signal power spectrum value at the frequency  $j$  and  $M$  is the maximal spectrum frequency.

- **Frequency Mean (MNF)**: is the sum of the product of the EMG power spectrum and the frequency divided by the total sum of the power spectrum.

$$MNF = \frac{\sum_{i=FMD}^M P_j f_j}{\sum_{i=1}^M P_j}$$

Where the terms have the same meaning as above.

### 1.2.3 Facial sEMG

Facial EMG is an electromyographic technique that measures muscle activity to reveal the very small electrical impulses generated by the facial muscles when they contract. It can be considered a valid method to evaluate masticatory function which depends on the facial muscles and to diagnose the oral system's disorder. It is also a tool to study human speech production and to help in speech recognition. Considering that facial expressions act as a non-verbal language through which it is possible to communicate the mood, feelings, and physical states, facial sEMG is further used because it allows the revelation of some facial expressions and sensations difficult visibly. [14].

### 1.3 The Average Threshold Crossing Technique (ATC)

The Average Threshold Crossing (ATC) is an event-driven technique in which the number of events increases whenever a myoelectric signal exceeds a fixed threshold, producing the quasi-digital Threshold Crossing (TC) signal. The figure 1.14 highlights the difference between the classic approach of transmitting data and the just mentioned technique.

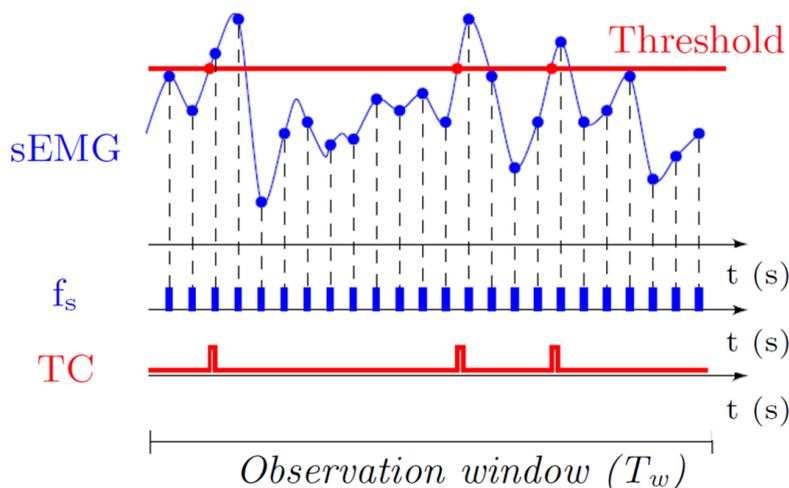


Figure 1.14: Classic sEMG sampling vs Average Threshold Crossing (ATC) sampling technique [15].

The ATC parameter is then the number of times the sEMG signal passes the threshold, in a defined time, divided by the period's duration. This innovative method allows reducing the amount of data to be digitized, saved, and transmitted, which also results in lower power consumption [16]. Moreover, unlike the classic transmission of the sEMG signal, which requires the use of an analog-digital converter (ADC), the TC signal can be sent directly as input to a microcontroller or a wireless transmission module. This implies that fewer electronic components are required, resulting in a decrease in the silicon circuit area's size and complexity [17].

In summary, the advantages of the event-driven technique are:

- low power consumption;
- reduced informations transmitted or stored;
- simpler acquisition hardware;

The disadvantages are related to the difficulty in choosing the correct threshold but above all to the loss of the signals' morphological properties and the impossibility of their possible reconstruction. Consequently, the ATC technique can not be used for diagnostic purposes but only for applications that do not require complete signal analysis. For the threshold problem, instead, in a previous study [18] the use of a dynamic threshold is proposed, involving an increase in the robustness of the ATC technique compared to the application of a fixed threshold. However, regarding to the correlation between the ATC parameter and the relative strength of a movement performed, as demonstrated by an early study [17], this technique can get ahead in that research field.

## 1.4 Classification Algorithms

In this chapter several supervised Machine Learning (ML) algorithms, used for the system validation, are presented and briefly described.

### 1.4.1 Artificial Neural Networks (ANN)

Artificial Neural Networks (ANN) have wide applications in many fields such as pattern recognition, prediction, optimization, associative memory, and control [19]. Just like the brain is made up of billions of neurons highly linked through connections, called synapses, an elementary operating unit in a neural network is a neuron-like node. Each neuron is a perceptron and receives a series of inputs, appropriately weighted, from the other neurons. The unit calculates the weighted sum of these inputs, followed by the application of a non-linear function. Therefore, the output of the neuron is obtained as:

$$y_j = f\left(\sum_i w_i x_i + b\right)$$

where  $x_i$ ,  $w_i$ ,  $b$ , and  $f$  are referred to input, weights, bias, and activation function, respectively. The changes in weights associated with synapses lead to learning by the brain, so the set of weights and biases is the neuron's information in the training phase. The value associated with the bias is variable, as it is for weights, and is corrected in this phase. Bias, put merely, determines whether and to what extent the neuron should fire up. During the training phase, the weights are repeatedly adjusted until the system output becomes consistent with the training data outputs or with the expected results. We thus speak of the learning rate. The activation function causes generating output and propagating the information from a neuron to another only if the inputs exceed a certain threshold. This is usually a non-linear function, such as a step function, a sigmoid, or a logistic function [20].

Many times a single neuron model is not enough. Thus, we speak of Multilayer Artificial Neural Network. The ANN architectures are organized into levels :

- **Input layer:** contains the input features vector;
- **Hidden layer:** represents the internal variables that encode the correlations between both the input and output to be generated. Sometimes it may not be present, but there may also be multiple hidden layers.
- **Output layer:** are the output values, that is the response of the neural network;

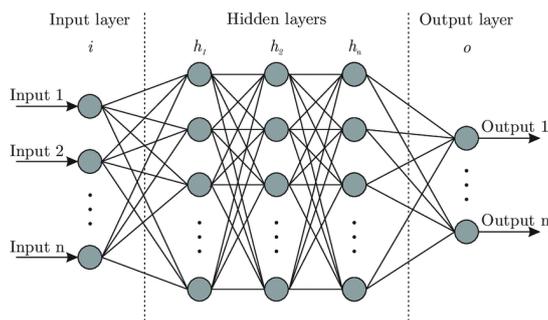


Figure 1.15: Example of artificial neural network architecture [21].

There are two types of multilayer ANN: *feedforward* (see Figure 1.15), where each node in a layer is connected only to the nodes of the next layer, and *recurrent*, where each node of a layer can be connected to nodes of its own layer or even to nodes of the previous layers, as well as the subsequent layers. ANN learning algorithm aims to determine the values of the weight connections to minimize the error. The gradient descent method can help solve this optimization problem. For intermediate nodes, understanding the error rate is problematic since a middle layer's output does not coincide with the desired outcome. The backpropagation technique is applied and consists of comparing the system output value with the desired one. Based on the calculated difference (i.e., error), the algorithm modifies the network weights, trying to converge the output results with the desired ones. The implementation of neural networks is suitable if there is a lot of data for use in the training phase. It is impossible to identify a model a priori, and a robust method is required despite the noisy data. The main disadvantage of artificial neural networks is the high computing cost. The choice of the hidden layers' size is also an issue: an underestimation of the number of neurons can result in poor generalization skills, but an overestimation can lead to overfitting [22] [23].

## 1.4.2 Support Vector Machine (SVM)

Support Vector Machine (SVM) is widely applied in computational biology [24], in face recognition [25], and text categorization [26]. It can solve linear and non-linear

problems. The SVM model's idea is to create a line or hyperplane that separates the data into two classes in the simplest case. This boundary is called the *decision boundary*. Points closest to the line, *support vectors*, are more prone to adding new data. A small variation of the hyperplane can change their classification. Farther points, on the other hand, are more likely to be correctly classified by the algorithm. The distance between the support vectors and the decision boundary is defined the *margin* (see Figure 1.16). Narrow margin models are more flexible and can adapt to many training sets. A large margin ensures a good generalization of the algorithm in classifying new data.

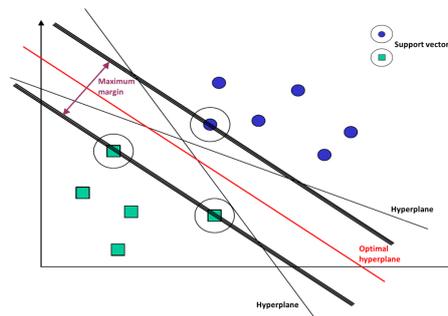


Figure 1.16: Optimal hyperplane using the SVM algorithm in a linear problem [23].

An SVM is also capable of handling non-linear problems, that is, when one line is not enough, as in the previous case, to be able to divide the two sets with precision. This is done by using a non-linear transformation function, *feature function*, that maps a non-linear model into a new space, defined *feature space*. It is, therefore, a question of raising a given model to a higher dimension to make it linearly separable [27]. The use of the kernel functions (see Figure 1.17) in the input space allows determining the optimal separation hyperplane without any calculation in the space of the higher dimensional characteristics. [28].

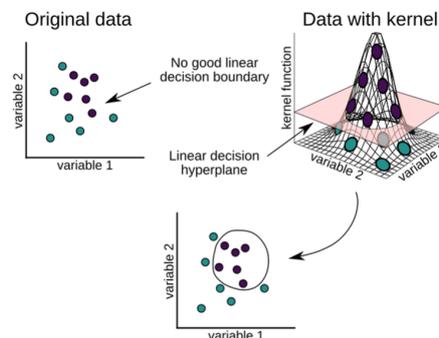


Figure 1.17: Representation of a non-linear model and its equivalent linear [29].

There are many kernels, the most popular ones are: *linear*, *polynomial*, *radial basis function (RBF)* and *sigmoid*.

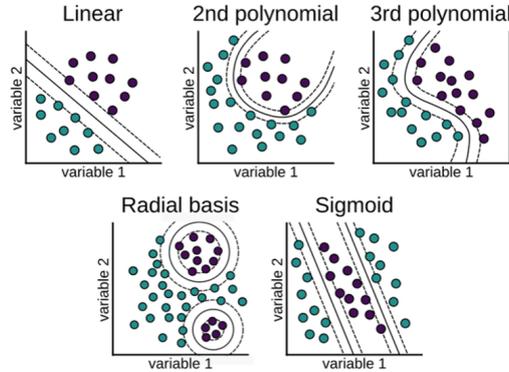


Figure 1.18: Different types of Kernel [29].

This algorithm in its basic form can be used for binary problem while in multiclass classification the optimization approach or the decomposing into a series of binary class models can be exploited. In the former case kernel function is used and in the latter One-Versus-One (OVO), One-Versus-All (OVA), or Directed Acyclic Graphs (DAGs) techniques are adopted [30].

### 1.4.3 k-Nearest Neighbour (k-NN)

The k-Nearest Neighbour classifier is successful in pattern recognition, in-text categorization, ranking models, object recognition, and event recognition applications [31]. k-NN is a supervised learning and lazy algorithm in which learning consists of storing the characteristics and classes of the learning set data and then classifying the new samples based on their similarity to the examples of the training set, without learning model creation. The k-NN is also a non-parametric method since it does not make assumptions about the distribution of the analyzed data [32].

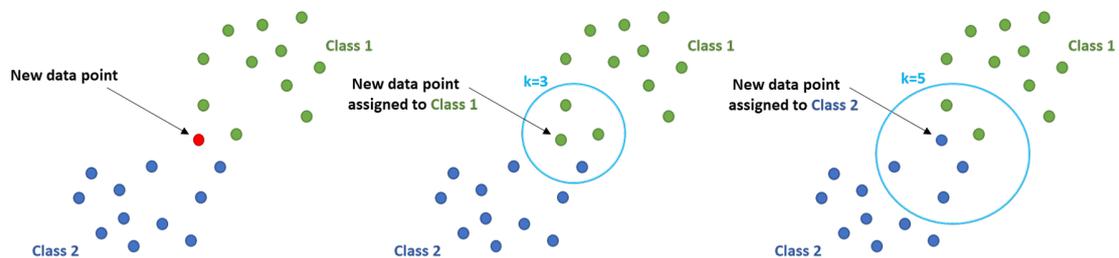


Figure 1.19: Example of a k-NN classification. A new data point could be classified in class 1 (green circles) or in class 2 (blue circles), according to the neighborhood selected. If  $k = 3$  it would be assigned to class 1 while if  $k = 5$  to class 2.

The idea behind the method is straightforward: data are classified according to the class of their neighbors. Given an unlabeled sample, the algorithm searches for the  $k$  examples of the training data set that are most similar to this point. It assigns the label that occurs most frequently among the  $k$  closest labels, as shown in Figure 1.19. The  $k$ -NN algorithm uses a similarity measure of the distances between the new and identified neighborhoods. Typically, the most used distance is the Euclidean distance but other distance functions, such as *Minkowski correlation*, and *Chi-square*, are available [33]. The parameter  $k$  is a positive integer and depends on the characteristics and type of data. High values of  $k$  reduce the effect of noise in the classification, but the boundaries between them are less defined. The main advantage of this memory-based approach is that the classifier adapts immediately as new data is added. On the other hand, a problem is a computational complexity that grows almost linearly with the number of learning vectors. Therefore, storage space and the number of distances to be calculated can become a nodal issue when working with large data sets [31].

#### 1.4.4 Decision Tree (DT)

Decision Tree (DT) is a non-parametric supervised machine learning algorithm. The principle on which the method is based in the construction of a tree. From the formal point of view, a tree is a finite set of elements called nodes (see Fig.1.20):

- *Root node*: represents a decision that will bring about the subdivision of dataset into two or more totally unrelated subsets.
- *Internal node*: is one of the potential decisions accessible at any point of the tree structure. This node is associated with its parent nodes on the upper end, on the lower one with its child nodes (or leaf node).
- *Leaf nodes*: is the results of decisions.

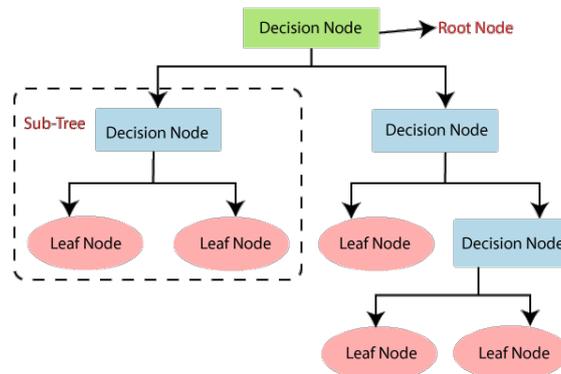


Figure 1.20: Example of a Decision Tree [34].

The set of nodes, except the root node, can be divided into  $n$  distinct sets, called *sub-trees*.

The method labels the nodes with the attributes' names while the arcs (the branches of the tree) with the possible values that the feature can assume. Each leaf represents a class label, while the other nodes are the conditions under which attributes are compared and separated. Classifying an instance is therefore reduced to applying, starting from the root node, the test conditions of each node to the example itself and following the branch composed of the comparisons' results. The leaf node will determine the class of the starting instance. So, the algorithm learns from the data through a decision pathway based on "if-then" rules [35].

The tree is built with a top-down approach using a strategy known as "divide et conquer" [36]. The most important phases of DT learning are:

- Selecting a split rule for each node determining the variables and the corresponding threshold value, which will be used to partition the dataset in each node;
- Determining which nodes are intended as terminals since it is necessary to decide when to stop splits for each node. Without an adequate rule, there is a risk of building trees that are too large with little generalization ability or are small in size and approximate the data incorrectly;
- Assigning the class to each terminal node.

Control criteria are used to limit the growth of trees. This area includes the *pruning phase*, which consists of obtaining the smallest subtree from a tree that does not compromise the classification accuracy. A branch or subtree that are irrelevant because represent noises in the data or outliers, can be removed [35] [37].

One of the advantages of decision trees is the production of clear classification rules and their ease of interpretation. However, they often have poor predictive performance. To overcome this possible issue, model ensemble technologies have been developed. One of the techniques used is the one called Random Forest (RF).

### 1.4.5 Random Forest (RF)

The Random Forest algorithm consists of using, for a given classification problem, several decision trees. This technique offers different points of view of the same problem managing to guarantee better accuracy results. It is a type of ensemble method that combines predictions from different decision trees, each of which is made from training set data taken randomly according to a specific probability distribution. The final result derives from the aggregation of the single tree's outputs (see Fig:1.21): the most voted class if it is a label problem or the average if the trees produce a numerical value.

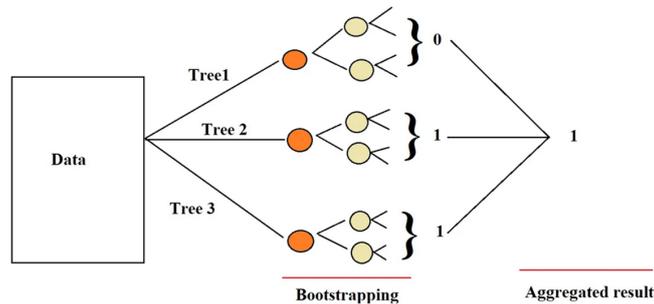


Figure 1.21: Example of Random Forest algorithm which gives the most voted class as output [38].

The random forest algorithm does not require scaling features or handling outliers because the data's variance does not affect how the underlying decision trees find the optimal split point. This makes it fast enough to implement. Additionally, random forests are far less likely to have overfitting problems than decision trees as they are trained on a more diverse dataset than individual trees. One major drawback of the Random forest algorithm is that it is more difficult to interpret. In a DT, predictions are derived from various combinations of features. In a RF, however, the inferences span numerous trees. It will also take longer to train than simple decision trees and can run slowly with large data sets. [39].

## 1.5 Machine Learning Metrics

To evaluate a given classification model's performance, there are several and various metrics to analyze the results. Functions require real ground truth labels and those deriving from model inference as input. They return a numeric value indicating the quality of the results. On the next few pages, we will look at some of the most popular classification metrics: *accuracy*, *precision*, *recall*, and *F1-Score*. Before seeing them in detail, it may be useful to define the following concepts:

- **Positive (TP)**: consists of the samples labelled and classified in the same way;
- **True Negative (TN)**: number of true negatives (classified as false are false).
- **False Positive (FP)**: is a set consisting of the samples classified by the system as belonging to one class but really belonging to another;
- **False Negative (FN)**: consists of the samples classified by the system as not belonging to a specific class which, in reality, are part of it.

The *Confusion Matrix* is a useful and effective evaluation tool. For a classification problem with  $N$  classes, the confusion matrix  $M$  is an  $N \times N$  matrix where the element  $M_{ij}$  corresponds to the number of samples belonging to the class as  $i$  but classified by the network as  $j$ . The elements of the diagonal correspond to the correctly classified samples. Therefore, the confusion matrix allows us to understand the model's performance in case the classes involved are not excessive. In classification cases in which the number of classes is high, it becomes challenging to extrapolate useful information for a first evaluation. For this reason, other summary metrics are used, such as those explained below.

- **Accuracy:** is the proportion of samples that are correctly classified to the total number of samples.

$$accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

- **Precision:** is the ratio of the number of correct predictions of class to the total number of times the model predicts it.

$$precision = \frac{TP}{TP + FP}$$

- **Recall:** measures the sensitivity of the model. It is the ratio of the correct predictions for a class to the total number of cases in which it actually occurs.

$$recall = \frac{TP}{TP + FN}$$

- **F1 score:** is a weighted harmonic average of the Precision and Recall metrics.

$$F1 - score = \frac{2 \times precision \times recall}{precision + recall}$$

All parameters range from 0% (worst) to 100% (best).

# Chapter 2

## State of art

### 2.1 ATC technique applied to sEMG

In recent years, the ATC technique has been combined with electromyographic signals with the possibility of obtaining the basic requirements for the realization of a wearable, portable, low energy consumption, wireless and multichannel acquisition system.

In [16], ATC approach has been merged with IR-UWB wireless technology in the biomedical applications field for the development of a low-complexity radio system. The use of IR-UWB technology allows the system to communicate muscle strength information with a receiver in two-way communication, without needing to digitize the data with an ADC. The work aimed was to demonstrate the correlation of performance system in terms of TC events (digital pulses) and evaluate the ARV values calculated on the raw sEMG signal. Furthermore, this information was combined with a force signal acquired by a dynamometer during the execution of the Maximum Voluntary Contraction (MVC) up to 70%. It has been possible to find a broad correlation between the above-mentioned signals (see Fig.2.1), obtaining an ATC-force correlation of  $0.95 \pm 0.02$  and ARV-force one of  $0.97 \pm 0.02$ , and show the effectiveness of the wireless-ATC based system for an estimate muscle force.

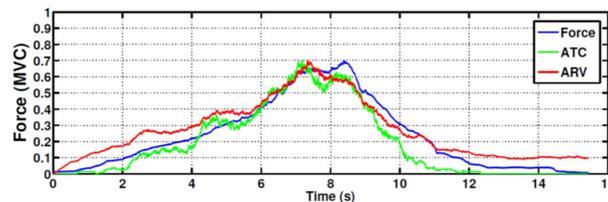


Figure 2.1: ATC, Force, ARV signals in time domain [16].

Another study [40] has proposed a more complex multi-channel acquisition system that relied on the previous ATC-based wireless version. The pattern of this system is based on the previous one: from the acquired raw signal the quasi-digital signal (TC signal) is extracted, serialized and sent to an encoder which, with an AER (Address-Event Representation) approach, forms a data packet detecting the input channel and triggering the IR-UVB transmitter for data transmission. This work allowed on one side to demonstrate the possibility of reducing the size and complexity of the circuit by removing ADC and on the other to evaluate the robustness of the ATC approach by varying SNR, amplifier distortion, saturation, and the number of lost events. Regarding this latter point, 5 - 6dB of SNR and 70% of lost events can be tolerated.

In [41], the system described above was tested with in-vivo experiments. The dependence of ATC impulses increase on the variation of force applied both in isometric and isotonic contractions was demonstrated. During the experimental protocol, a set of different weights (i.e., 2 kg, 4 kg, 6 kg, 8 kg and 10 kg) was used to prove this aspect. By adopting the ATC technique, the discrimination of the resting condition and contractions, performed using weights of at least 6 kg of difference, were possible without errors.

In [42], an sEMG wearable acquisition board was developed and validated with experimental tests. In particular, the muscle force correlation with the ATC was evaluated through isometric contractions adopting different weights. In this work, the ATC events were computed and compared with sEMG standard parameters (i.e. MNF, MDF, RMS, peak to peak amplitude (Vpp), and total power (Ptot)). The comparison showed ATC approach presents the highest median values of the correlation coefficients, resulting in a good parameter to measure muscle force.

ATC is also considered a valid technique for machine learning applications, such as hand gesture recognition, as shown in [43] and [44].

In [43], ATC data from four wrist movements, acquired with three input channels, was used to train an SVM classifier. An average accuracy of 92.87 % was achieved with this model, resulting in only 5.34% lower than that obtained with the training based on features extracted by raw sEMG signal. In addition, by using the ATC approach, a significant reduction of data rate was achieved from 6.14 kB/s, resulting from the sEMG signal-based model, to 30 B/s. It has also been demonstrated an energy consumption decrease of about 14% with the transmission of the ATC data acquired from 4 channels compared to that measured using one sampled sEMG signal.

In [44], a low-power embedded system including three acquisition channels and the

Apollo2 microcontroller from Ambiq Micro was presented in order to recognize six hand gestures. The Neural Network (NN) method was used to perform the movements classification, reaching an accuracy of 96.34%, with a latency of 268.5 ms and a power consumption of 2.9 mW.

In [15], ATC was also used in another application to control Functional Electrical Stimulation (FES), a rehabilitation treatment that has the function of strengthening or loosening a muscle applying low energy electrical pulses to increase its strength. The acquisition board used is described in [42] and mentioned above with the exception that the IR-UWB was replaced with Bluetooth Low Energy (BLE). A time window of 130 ms was chosen to compute ATC values from 4 acquisition channels, managing to get a data throughput of about 30.7 B/s. For the power consumption aspect, 5.126 mW was measured for the ATC acquisition phase alone, and 20.23 mW was reached also including the BLE transmission.

This system was tested in [45] by evaluating the correlation between the voluntary movements from which the ATC is acquired and the motions induced by the FES protocol. The cross-correlation value obtained was overall above 0.9.

## 2.2 Facial sEMG Applications

Facial sEMG is a topic widely treated in the researches's works. Several examples are presented below.

Jun-Wen Tanhave et al. [46] have done a study on analysis intending to discriminate three different emotional states, negative, neutral and positive by monitoring two facial muscles activity (Zygomaticus major and Corrugator Supercilii) in response to a visual stimuli. The two muscles can be used as indicators to test negative and positive sensations. This can represent a starting point for the future development of automated methods for emotional states identification such as Human-computer interaction in situations such as monitoring of elderly people.

In [47], facial sEMG is also used as an index of palatability by exploring muscles like the levator labii superioris/alaque nasi region. The present study aims to find the correlation among the sensory hedonic muscular responses elicited from different tastes in the region of levator labii muscle.

G. S. Meltzner et al. [48] have presented a speech recognition system using 11 recording channels placed on the face and the neck to classify three different modes of speaking conditions (i.e., voiced, mouthed, and mentally rehearsed). A standard Hidden Markov Models (HMMs) was implemented for the recognition phase, reaching in vocalized and mouthed speech modes mean rates of 92.1% and 86.7%, respectively. Good recognition is not possible for mentally rehearsed mode, due to

poor muscle activity. This system is useful for the speaker-dependent case since a small amount of data (65 words) is used to train the classifier.

C. Huang et al. [49] have designed a speech recognition system to help speechless people. Variance (VAR), IEMG and WL were extracted from sEMG signal and were used as inputs to a ANN algorithm. ANN was a binary classifier and was trained to recognize two words, "yes" or "no". Several conditions concerning vector input, training data, and size of hidden layers were tested. The overall accuracy obtained for each condition is not been optimal. Despite this, the prototype system could be used for speech pattern recognition by introducing some improvements such as database and extracted features expansion.

E. Lopez-Larraz et al. [50] have proposed a prototype silent-speech recognition system to overcome limitations dued to ambient noise and to be useful for people with disorders as laryngectomy. Different syllabes, which are organized in five group in line with involved articulation (i.e. labials, dentals, palatals, velars, and alveolars), have to be recognized. The classifier adopted was based on boosting and has obtained a global mean accuracy of 69%.

In [51], facial sEMG signals were used to assess the response in masticatory muscles between healthy people and patients with Cerebral Pulsy (CP) at different degree of oral motor function impairment. It was demonstrated that the two groups' muscular activity differs statistically in each simulated situations (at rest, opening and closing the mouth).

In [52], the impact of head positions on muscular activity among adult people with brainstem stroke and control group was evaluated. Head position does not influence masticatory muscles' performance but the activity is widely different during Maximal Voluntary Contraction (MVC) in the two groups. The sEMG could be used as a tool to analyze impairments in stroke patients.

## 2.3 Facial gesture recognition

Humans express their emotions not only verbally but also through their actions. It has been shown that 93% of communication is non-verbal and facial expressions and body gestures play a fundamental role [53].

That is exactly why gesture recognition is recently a popular topic in researchers' work, with applications in various fields.

The capture of images or videos of body movements and the recording of the sEMG signal has allowed the implementation of gesture recognition systems by using different approaches.

However, they share several steps before the gesture recognition phase: data pre-processing and features extraction.

The architecture for these typical systems consist of four main building blocks, as shown in Figure 2.2:

- **Input:** handles directly the system input which can be a raw image, a video or a sEMG signal;
- **Pre-processing data:** includes noise reduction, face detection, alignment, normalization, augmentation, and enhancement phases for images and a filtering step for the sEMG signal;
- **Features extraction:** extracts useful data in the form of appearance and geometric aspects, temporal and spatio-temporal information from the image/video and in time or frequency indices (described in Sec.1.2.2) or ATC parameter (discussed in Sec.1.3) from sEMG signal;
- **Classification:** related to classification algorithm, some of which are described in Sec.1.4.

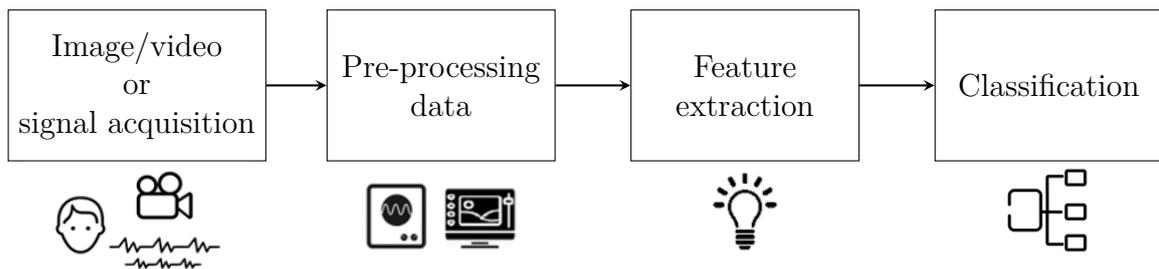


Figure 2.2: Architecture of a gesture recognition system.

### 2.3.1 Video and images Applications

The technology that uses mathematical algorithms to analyze facial muscles activation in images and videos goes by the name of Facial Expression Recognition (FER). Since the emotional state is involved in activating the muscles, facial expressions can be detected and distinguished via FER systems [54] [55]. It has a wide application in Human-Computer Interaction (HCI) (g.e. interactive gaming and virtual reality), in the medical field for behavioral and emotional analysis, such as Autism, and in the field of surveillance [56]. However, not only the face image but also physiological signals (i.e.EMG or EEG signal) can be used as input data in FER applications [57]. But despite this, the camera for its easy availability and simple usage is the most widely used sensor.

Camera-based FER systems can rely on the analyzing of a face static and single image or focusing on videos in dynamic scenarios. In addition, a video-based system can be of two types: frame-based and sequence-based. The former refers to a system that considers a single frame to distinguish different facial emotions. The latter, instead, extracts temporal information in the frames needed to detect expressions [58].

The most relevant problems affecting the performance of these FER systems are:

- The variation in the background illumination of the image or video can worsen the accuracy of the features extracted from the face. This issue can make images appear differently due to shadows appearing or part of the face image being darkened.
- The difficulty in recognizing generic facial expressions but only those of pre-trained human faces (subject-dependence), since human faces differ according to their nationalities, skin texture, people’s age, and even modes of expression.
- The changing of the head pose makes challenging to detect expressions since, in most existing approaches, the face is in the frontal position. For example, head rotations can result in the loss of some portions of the face and complicate image processing techniques and consequently make inaccurate expression classification algorithms.

Due to these challenges, it is difficult to analyse facial deep information with 2-D static images and video sequences. To improve FER system camera data can be merged with information from other sensors, obtaining a multimodal sensors. In [58] three categories of sensors were identified:

- Detailed-Face: focuses on other parts of the face such as the eyes directly related to mental states’ expression such as attention or sleepiness. I. Hupont et al. [59] have proposed the combination of an eye tracking sensor, which analyzes where a person is looking, and FER technology (see Fig.2.3).



Figure 2.3: Emotracker system: a 17-inch TFT monitor with integrated IR diodes to track user’s gaze and a webcam to detect facial expressions [59].

This system monitors not only gaze but also has several intuitive visualization options for human emotions. It could have a good following in the field of psychological research.

- Non-Visual: includes audio, biosignals sensors and depth camera.
- Target-Focused: acquires infrared thermal images, which calculate the temperatures distributed in different parts of the image. It can solve lighting variation in the image.

3D facial imaging can provide the possibility to tackle the issues mentioned earlier by including geometric, time series data and depth information. For example, it manages to bypass head pose problem by extracting geometrically invariant features and having a more extensive consideration of rigid head motions [60]. Although the 3D approach also allows for superficial face and ear-to-ear measurements, it requires higher storage and computation costs and a large time consumption, proving to be not always suitable for real-time applications [54]. Deep learning technique can also be used to overcome conventional approaches, since it reduces the image pre-processing phase, ensure a certain reliance on physics-based models, and are more robust for uncontrolled environments with various types of elements and different lighting. In addition, it provides also improvements in terms of efficiency i.e. time, computational and space complexity.

### 2.3.2 sEMG Applications

The sEMG technique, particularly, has wide application in kinesiology and in the identification of neuromuscular disease and myoelectric control systems. Similarly, gesture and facial expressions via the sEMG approach are focused on the study of emotions, nursing, HCI, affective computation [61] and also in design and control of a virtual crane training system [62] or in rehabilitation devices like wheelchair [63]. Compared to the system based on image or video, the sEMG-based one has a high temporal resolution, can detect some facial muscle movements invisible to the naked eye, is efficient in data processing, not influenced by head pose and measurement environment and suitable to be embedded in wearable devices [61].

In the following pages the literature works related to facial gesture recognition are listed, with particular emphasis on accuracy score.

G. Gibert et al. [64] have implemented a classifier, the Gaussian model, which able to discriminate six facial expressions using eight electrodes placed on the specific muscles. To obtain a result, they have used, as sEMG feature, the envelope of absolute values. During the testing phase, they have demonstrated that the classifier worked fine with a final average accuracy of 92.19 % on all facial expressions, while making some mistakes in the beginning and the end of each movement performance.

L. Ang et al. [65] have focused only on three muscles, which have a pretty blatant response in terms of contraction for emotions that need to be recognized. The Mean, the Standard Deviation, the Root Mean Square (RMS), and the Power density spectrum (PDS) of the sEMG signal were computed for each muscle. In addition, they have reduced the collected data dimensions with the Feature Differentiation technique to implement the Minimum-distance classifier and have reached an overall performance of 92.78%.

UM. Hamedi et al. [66] group have presented a method for recognizing eight facial gestures through three recording channels in a bipolar configuration. The feature vector's size was reduced, using only RMS, directly related to the contraction force of a muscle. Unlike the other works, not only facial expressions related above all to emotions have taken into consideration but also gesturing as opening the mouth as if to say "a". Two classification techniques were applied, Support Vector Machine (SVM) and Fuzzy C-means clustering (FCM), succeeding to obtain 91.8% and 80.4% recognition ratio, respectively.

Y.Chenhey et al. [67] have focused on motor expressions related to the movement of the eyebrows and designed a headband (shown in Figure 2.4) to allow easy placement of electrodes and also to obtain a standardized signal recording. Three-time domain indices of the sEMG signals were extracted for each muscle,  $RMS_{max}$ ,  $RMS_{mean}$ ,  $RMS_{var}$ , and then used as inputs for the classification algorithms. Two types of neural networks were proposed (Ensemble Neural Networks (ENN) and Back Propagation Neural Networks (BPNN)) and their results have been compared. The average accuracy achieved in both cases was nearly perfect. The ENN(Training: 97.12% Test: 96.12%) had slightly better performance than the BPNN(Training: 95.24% and Test: 95.56%), by evaluating all two models with the training and the test set.



Figure 2.4: Headband [67].

Y. Cai et al. [68] have studied the recognition of eight facial gestures, distinguishing them into transient expressions (i.e., sadness, happiness, pout, angry and surprise) and periodic expressions (i.e., chewing, speaking, gargling) via sEMG signals recording with six channels. Many features were extracted from each signal, both in frequency and time domain to then develop three different classifiers, cubic SVM, Gaussian SVM, and cubic K-Nearest Neighbors (KNN). All algorithms have achieved good results, but the Cubic SVM classifier had the best performance with an accuracy of up to 99.52

In particular, the work of S. Orguc et al. [69] is interesting because, in contrast to the above works, jaw movements are investigated . The group has used the Discrete Wavelet Transform (DWT) as a unique feature, more suitable for non-stationary signals such as sEMG but computationally expensive for the feature's high dimensionality. An SVM algorithm was implemented to distinguish in real-time three jaw movements (i.e. clenching, chewing, and jaw opening), obtaining an accuracy greater than 85%.

# Chapter 3

## Preliminary studies

The aim of the thesis work concerns facial movement recognition, and in this direction, a preliminary analysis of facial expression and jaw movements has been carried out. All the studies reported in this chapter have been conducted to understand which muscles could be properly monitored in order to recognize related gestures. This analysis has been performed firstly, recording sEMG signals with g.HIamp-Research multi-channel amplifier [70], during the execution of test movements, and then post-processing them in order to figure out muscle-actions associations. The ATC technique is applied to post-processed sEMG signal in order to assess if an event-driven processing is suitable for recognizing the facial gestures.

### 3.1 Equipment specifications

Through the sEMG technique, insights have been gained into understanding certain facial muscle behavior when performing particular movements. The recording of the sEMG signals, acquired in the bipolar configuration with pre-gelled sensors, has been made possible using a multichannel biomedical amplifier, produced by *g.tec*, integrated with user-friendly support software. The signals have been then post-processed in order to extract the ATC feature via software and understand if the event-driven approach could be valid in identifying face gestures.

#### 3.1.1 gHIamp-Research

g.HIamp-Research is a multi-channel biosignal amplifier. It is used to measure brain functions in an invasive or non - invasive way, and it is intended for research purposes and not for medical applications. The amplifier consists of 80, 144, or 256 channels per unit, and each block of 64 channels is connected with a multi-pole connector to the electrode headbox, which can be active or passive. It is connected via USB to PC, and each channel has 24-bit ADC with an internal

sampling frequency until 38.4 kHz. It allows to make also the measurement, without any saturation, of:

- Electrocoorticulogram (ECoG);
- Electrooculogram (EOG);
- Electrocardiogram (ECG);
- Electromyogram (EMG).



Figure 3.1: g.HIamp-Research multi-channel system: (a) g.HIamp amplifier; (b) 64-channel passive electrode connector box [70].

g.HIamp includes the following components:

- g.HIamp USB biosignal amplifier;
- 64 channel electrode connector box;
- connection cable from 64 channel electrode connector box to g.HIamp;
- GlobTek GTM21097-3005 – medical power supply unit;
- power line cord;
- USB cable;
- instruction for use;

In this thesis, the 144-channel g.HIamp model was used for the recording of sEMG signals collected from the subject's face. The amplifier's technical specifications are described below.

Table 3.1: g.HIamp-Research amplifier technical specification.

<b>Model</b>	g.HIamp-Research
<b>Type</b>	144 - Channel Research Amplifier
<b>Rated power consumption</b>	20 VA
<b>Rated DC voltage</b>	5 V
<b>Rated current of fuse</b>	4 A
<b>Rated voltage of fuse</b>	250 V
<b>Sensitivity</b>	$\pm 250mV$
<b>Highpass</b>	0 Hz
<b>Lowpass</b>	19.2 kHz
<b>Input Impedance</b>	$>100 M\Omega$
<b>ADC resolution</b>	24 bit
<b>Sampling frequency</b>	38.4 kHz per channel
<b>Number of ADCs</b>	144

### 3.1.2 g.Recorder Software

The g.tec support program in Windows is *g.Recorder*, a novel, user-friendly software package for recording, displaying, and storing biosignals. It is also possible to capture trigger information, video sequences and inspect data after recording through an easy-to-use *Replay mode*. Some extensions of software allow a few signal parameters, useful for clinical research, to be extracted and visualized online: Compressed Spectral Array (CSA), Heart Rate (HR), and Heart Rate Variability (HRV). The data is stored in hdf5 format. In order to perform these preliminary tests, g.Recorder software has been used to provide a direct control on acquisition hardware (i.e., g.HIamp-Research). It is possible to configure the amplifier by setting the sampling frequency, the common reference, the active channels by defining the type of configuration (bipolar or unipolar), and the filters to be applied. Furthermore, the acquisition unit also provides configurable filter setups in order to design low-, high- or bandpass- filter: for the high-pass filter 0.1 Hz, 1 Hz, 2 Hz, 5 Hz, and for the low-pass one 30 Hz, 60 Hz, 100Hz, 200 Hz, 250 Hz, 500 Hz are the selectable frequencies, respectively. The g.Recorder also offers the user the possibility to define and identify markers (see Fig. 3.2) indicating a particular time point and area markers specifying a time window to highlight specific events during measurement. These markers are useful for the labeling of the signals being classified. Once the channels have been configured and the tags defined, the setup can be saved and used when needed.

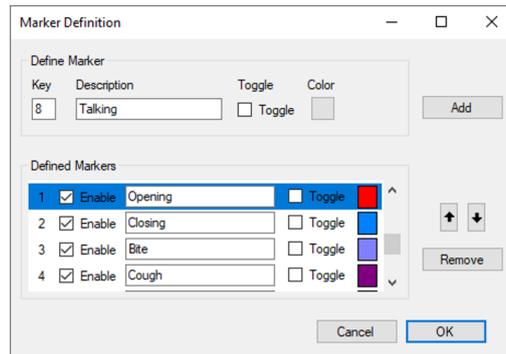


Figure 3.2: Markers definition.

### 3.1.3 sEMG Quality Improvement

The quality of the sEMG signal can be improved with good skin preparation and a proper electrode application. Skin cleaning with alcohol and drying before electrode placements can improve their adhesion. The epidermis is a poor conductor of electricity and may create an artefact that can distort the signal. The electrodes used for sEMG acquisition are the Kendall™ H124SG model produced by Covidien [71]. The sensor is pre-gelled, disposable, and with circular surface Ag/AgCl coated and a 24 mm diameter.



Figure 3.3: Electrodes used for sEMG acquisition.

The face electrode placement has been carried out based on the literature that focuses on the facial muscles activity [51, 52, 72] and according to the guidelines for sEMG recording [73]. In the following pages, the movements that have been analyzed and the muscles involved will be introduced and explained in detail the electrodes' positioning. Some actions will be discarded and some muscles until the electrodes' final configuration and the movements to be carried out are defined. As the gestures involve muscles located in different facial areas, more than one pair of electrodes will be used.

### 3.1.4 Post-processing for ATC Extraction

The recorded sEMG signals have been saved in order to be processed in the Matlab environment to improve signal quality (whenever necessary) and extract the ATC feature. The steps taken are briefly reported below:

- Signal filtering: pass-band filter between 30-400 Hz.
- Setting of parameters useful for calculating the ATC feature extraction:
  - Definition of a threshold value for each channel considered. The threshold is computed as:

$$V_{th} = baseline + mean\_noise + 3 * std$$

*baseline* represents the noise baseline evaluated considering the first six seconds of the recording; *mean\_noise* and *std* are respectively the mean and the standard deviation of the noise calculated after rectifying the part of the signal considered.

- An hysteresis of 15 mV has been considered around the computed threshold value in order to take into account spurious muscle signal activations.
  - A time window of 130 ms for the ATC feature computation.
- Processing of the ATC characteristic by evaluating how many times the sEMG signal exceeds the voltage value set at  $V_{th} \pm 15$  mV.
- The ATC parameter is defined by summing all TC events in the chosen observation window.
- Plotting the ATC envelope superimposed on the sEMG signal, both signals are normalized concerning their maximum value.

## 3.2 Movements Investigations

### 3.2.1 Jaw movements

A first analysis is based on investigation of jaw movements: *opening*, *closing*, *bite*, and *clenching*. As described in Section 1.1.3, different muscles are involved in jaw actions; therefore, just the most superficial and palpable ones have been selected. In particular, to examine jaw opening, the *mylohyoid* and *digastric* muscular activities have been recorded. Instead, to analyze jaw closing, clenching, or biting, *temporal anterior* and *masseter* have been investigated.

In Figure 3.4 it is possible to observe a preliminary positioning of the electrodes on selected muscles according to easily palpable and specific anatomical landmarks.

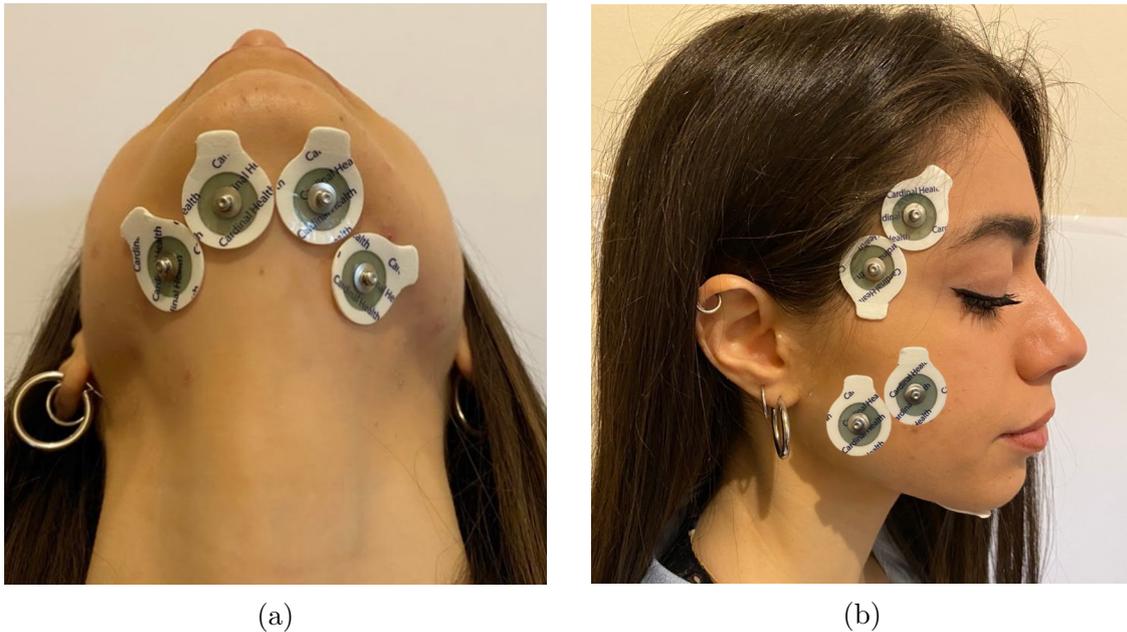


Figure 3.4: Preliminary electrodes placement. Muscles considered are: (a) *digastric* and *mylohyoid*; (b) *masseter* and *anterior temporal*.

- By clenching the teeth, it is possible to palpate the *anterior temporal* and *masseter* to place the electrodes on the first muscle, vertically along the anterior margin and on the second, on the lower third of the line between the lateral corner of the eye and the gonial angle, the more lateral and palpable point of jaw angle.
- Lifting the head allowed electrodes to be placed in the medial submental region at the base of the mandible to record the activity of the *digastric* muscle and on the *mylohyoid* by palpating the edge of the hyoid bone on which this latter muscle is located.

Signals were recorded in the following test conditions:

- **Jaw elevation:** mouth opening at maximum;
- **Rest or Closing:** subject kept lips relaxed;
- **Dental clenching:** subject was asked to clench their teeth with maximum force, named Maximum Voluntary Isometric Contraction (MVIC).

Each movement lasts 20 s without any rest between one and the other, and the series of three gestures was repeated twice with a pause of 30 s.

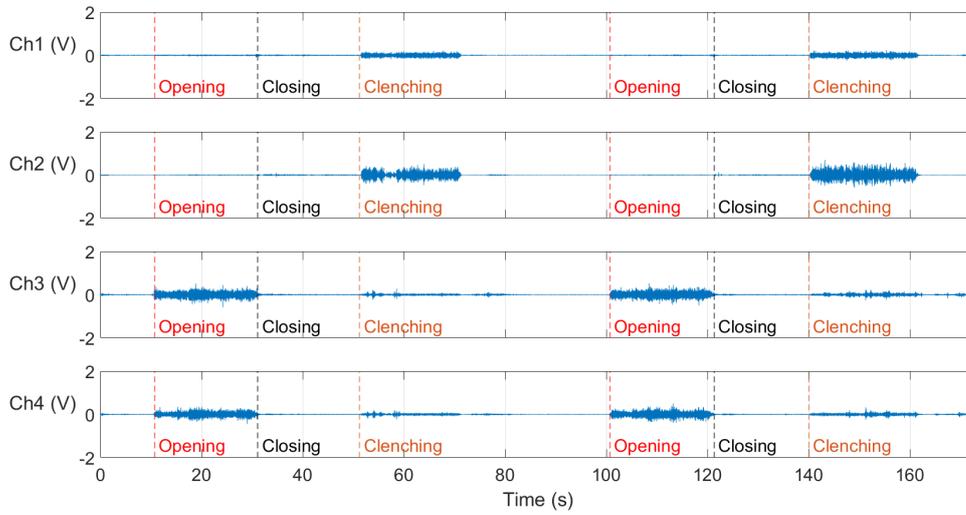


Figure 3.5: sEMG signals during jaw movements. Muscles activation investigated by each channel: Ch1 - *Anterior Temporal*, Ch2 - *Masseter*, Ch3 - *Digastric*, Ch4 - *Mylohyoid*.

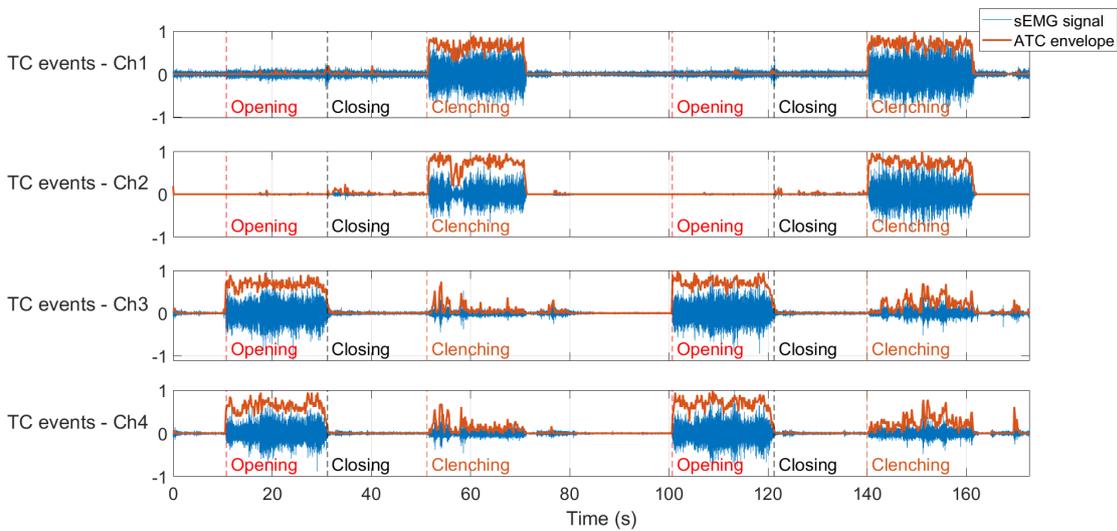


Figure 3.6: ATC envelope during jaw movements. Muscles activation investigated by each channel: Ch1 - *Anterior Temporal*, Ch2 - *Masseter*, Ch3 - *Digastric*, Ch4 - *Mylohyoid*.

The muscles involved in the opening movement (see Figure 3.5) still have a very similar muscle response. Consequently, in order to prevent misleading values from being acquired due to the electrodes' problematic positioning and minimize the

number of recording channels, it has been decided to discard the mylohyoid muscle. Furthermore, the masseter muscle response has been greater than one the temporal muscle as it is more affected by tooth clenching.

At this point, a comparison evaluation between the sEMG signal and its related ATC enveloped has been carried out. The Figure 3.6 shows that the single movements are well delineated by the ATC signal when there is muscle activity. The ATC envelope has null values during the closing since the sEMG signal does not exceed the set threshold, also proving the noise-robustness feature of the technique. At the end of the recording and the end of the first clenching movement, there should have been no TC events being a resting phase, but involuntary movements could appear due to muscle fatigue. On channel one, the baseline noise is more significant, probably because of unstable electrodes position over the muscle of interest. For these reasons, a different threshold has been set for each channel.

### 3.2.2 Singing and talking

Another action to be studied has been *singing*. In order to analyze it, electrodes were placed on the *superior orbicularis oris*, close to the angle of the mouth. With respect to the the previous acquisition protocol, described above, singing was introduced after clenching the teeth. The series of actions is repeated twice, separated by 20 s as resting phase. Furthermore, singing action can also be observed on the other muscles as Digastric.



Figure 3.7: Electrodes palcement on *superior orbicularis oris*.

Since singing may not make the performance of the subjects comfortable, it has been decided to analyze the act of talking at a later time, maintaining the same protocol used for singing. In the Figure 3.9 the muscle activity of the superior orbicularis oris during talking appears to be slightly more intense than the one previously observed with singing.

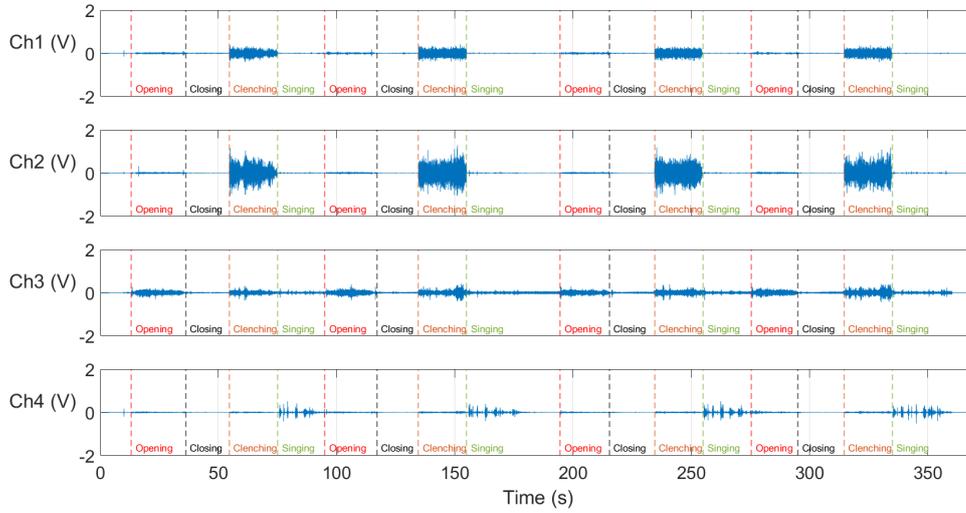


Figure 3.8: The performing of singing movement - sEMG signals. sEMG signals recorded by each channel: Ch1 - *Anterior Temporal*, Ch2 - *Masseter*, Ch3 - *Digastric*, Ch4 - *Superior Orbicularis Oris*.

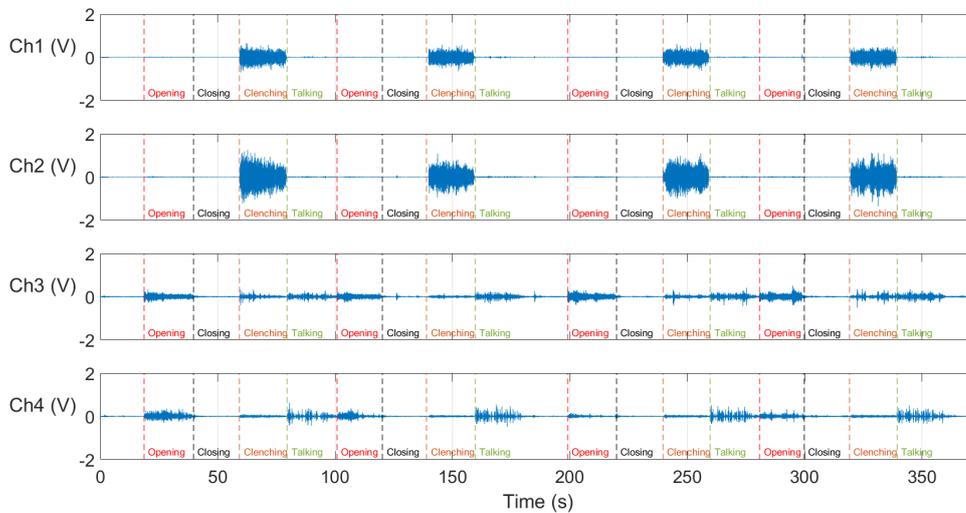


Figure 3.9: The performing of talking act - sEMG signals. Ch1 - *Anterior Temporal*, Ch2 - *Masseter*, Ch3 - *Digastric*, Ch4 - *Superior Orbicularis Oris*.

Talking seems to be a more continuous movement over time, while when singing, it is necessary to take breaks to breathe according to the type of intensity tone used (e.g., low, high, acute, or severe). On the digastric muscle, there can also be a muscular response.

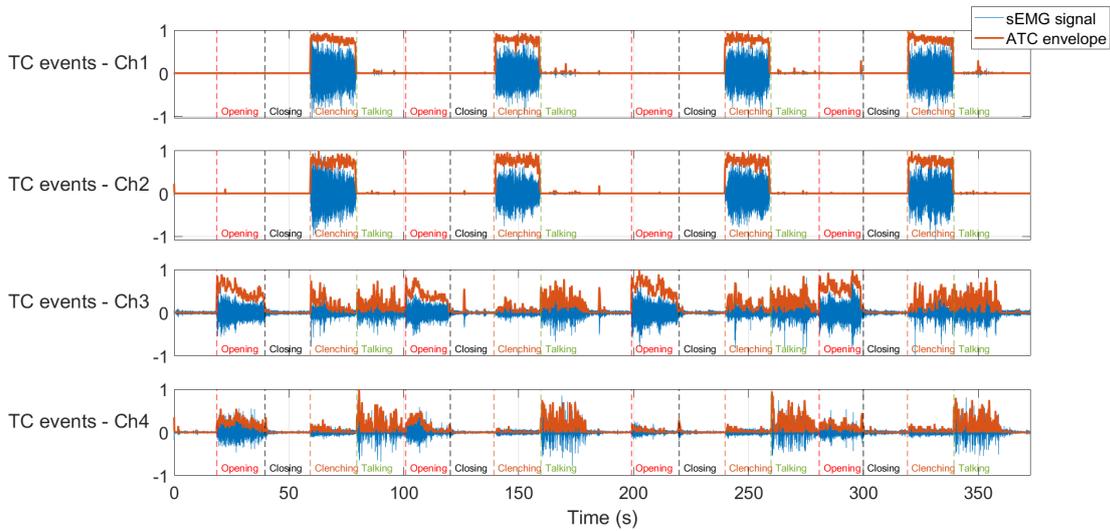


Figure 3.10: The performing of talking act - ATC signals envelope. Ch1 - *Anterior Temporal*, Ch2 - *Masseter*, Ch3 - *Digastric*, Ch4 - *Superior Orbicularis Oris*.

The ATC signals envelope shows how in the case of static movements such as opening and clenching, the TC values in the various time windows are primarily stable. During speaking, instead in moments of pause between one word and another, there is less muscle activation and therefore fewer events.

The action of talking is not necessarily observable on the orbicularis oris, but a muscular response is also visible on the digastric. Consequently, in order to reduce the active channel number and eliminate the problem that the electrodes above the orbicularis oris could be detached during other facial movements (e.g., bite action), this muscle was put aside.

### 3.2.3 Bite Analysis

Subsequently, a second analysis has been performed acquiring sEM signals during bite and chewing executions. The masticatory muscles' typical signals during an apple bite are shown in Figure 3.11 in which the main features are indicated. The first bite and subsequent chews are spaced relatively evenly so that the whole signal appears periodic. The initial bite is quite broad and distinguishable from the next one that makes up the chewing. The end of chewing corresponds to swallowing, which, however, is not always visible. Looking at the figure shows that the moment of swallowing is barely visible on channel 1. It can also be seen that there is a correspondence in the activity of the two muscles and a similar involvement, different only in terms of signal amplitude.

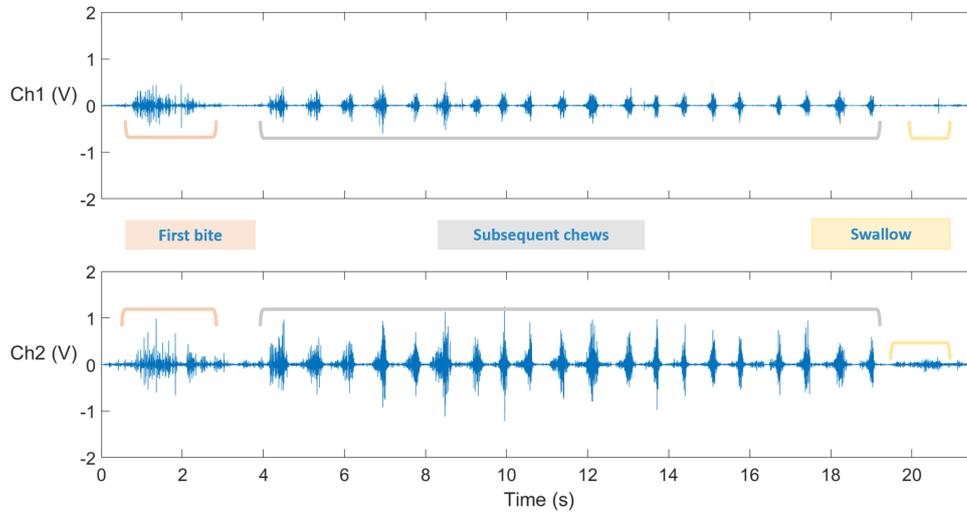


Figure 3.11: Main features of the masticatory muscles. Ch1 - *Anterior Temporal*, Ch2 - *Masseter*.

Food texture can modulate several aspects of these actions: bite force, masticatory force, jaw movements, length of chewing cycles, and the cycle number before the swallowing [74]. When eating food, the individual has her chewing way, which depends on the behavioral and anatomical aspects. In particular, chewing differs from one person to another for the breakdown of the food bolus, influenced by the salivary flow, the ability to reform and move food in the mouth, and different muscle activation in terms of muscle recruitment [75]. The force of the bite (in the first chewing cycle) can vary due to subjects' experience of food texture. Therefore, it has been decided to test different types of food to distinguish the different types of bite and evaluate their different intensity. The foodstuff chosen were the following:

- cracker: soft and friable aliment;
- chewing gum: relatively consistent texture, soft and unbreakable, despite repeated chewing cycles;
- nougat: crunchy and hard food;
- apple: floury texture.

A chewing gum has not been considered in food tested since chewing force and duration do not change during the movement. Unlike other foods, chewing gum did not disintegrate during the masticatory cycles. For this reason, chewing gum has not been tested while apple, cracker, and nougat have been chosen for their assorted characteristics. The sEMG signals reported in the Figure 3.13 highlight the difference among the food types:

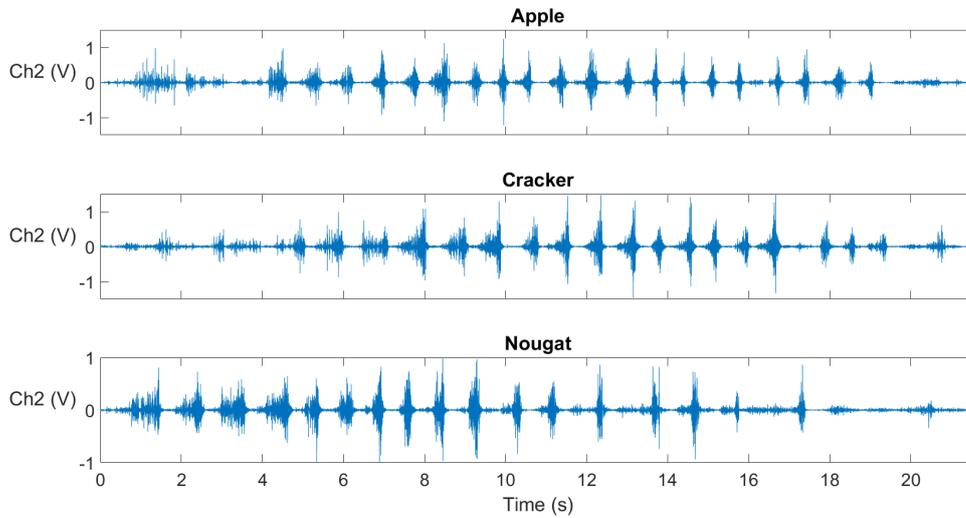


Figure 3.12: sEMG masticatory signals testing different foods. Ch2 -*Masseter*.

- The bite of the apple has longer duration since depending on the quantity of injected food. The chewing phase does not begin immediately after the bite but is preceded by a further bite much less intense than the first and one, which is necessary to make food consumption easier.
- The bite of the cracker, due to its friability, is not very intense, unlike the chewing cycles that reach higher intensity peaks in amplitude, even compared to those of other foods.
- Last, regarding the nougat bite, it lasts less than the one of the apple. Furthermore, the first chewing cycles are very similar to the first bite, given the nougat's rubbery consistency and the difficulty of disintegrating it immediately. They become much more distant from each other because there is probably little food left in the mouth.

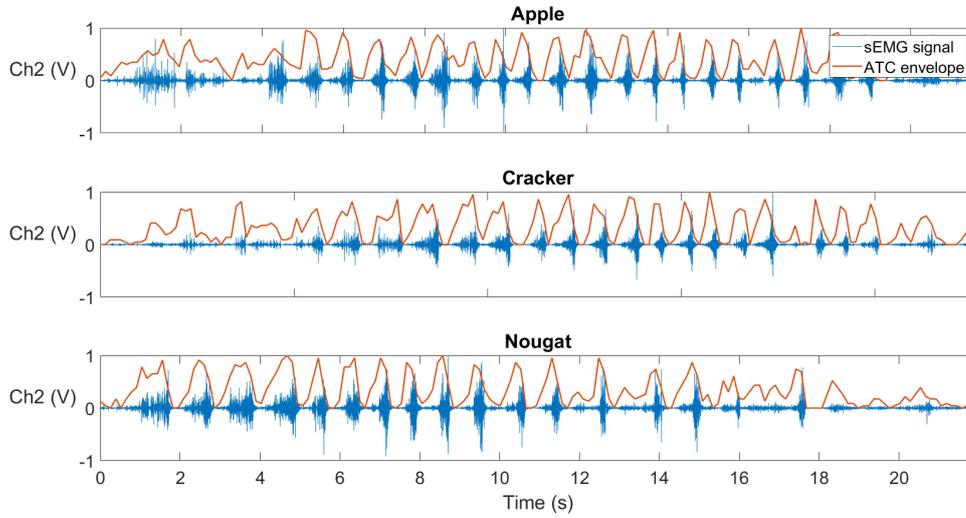


Figure 3.13: ATC masticatory envelopes testing different foods. Ch2 -*Masseter*.

The ATC signals follows well the chewing cycles. The bites of the apple and the nougat are represented by a higher number of TC events being more consistent, unlike the cracker one which is less intense.

### 3.2.4 Facial expressions

Facial expressions are known for their role in emotions communications, and they often tell us how people feel. It is essential to correctly identify and recognize human emotional reactions in order to improve the interactions among digital devices and their users. Some of them express basic emotions such as happiness or anger. In order to investigate them, two mimics muscles (discussed in 1.1.3) have been introduced in the electrodes placement setup: *zygomaticus major* and *corrugator supercilii*. *Zygomaticus major* pulls mouth corners upward and laterally for smiling while *corrugator supercilii* lowers eyebrows for angering [76]. Electrodes have been placed on muscle regions (see Fig. 3.14), according to the guidelines for fEMG positioning [73]. In order to investigate and increase other movements to be recognized, eye gestures have been considered like *winking*. Winking can be an extra control element in systems like the video games sector and in situations where the users have busy hands [77, 78]. The muscles involved in winking can be *zygomaticus major*, *corrugator supercilii*, and anterior temporal.



Figure 3.14: Electrodes placement on the mimics muscles.

An acquisition protocol (shown in Figure 3.15) has been defined to observe emotional and winking expressions. The gestures to be performed have been in the following order: wink, smile, and scowl. The three movements have been repeated twice within the first session with 5 sof rest between expressions. Then, after a 20 spause, a second session has been performed.

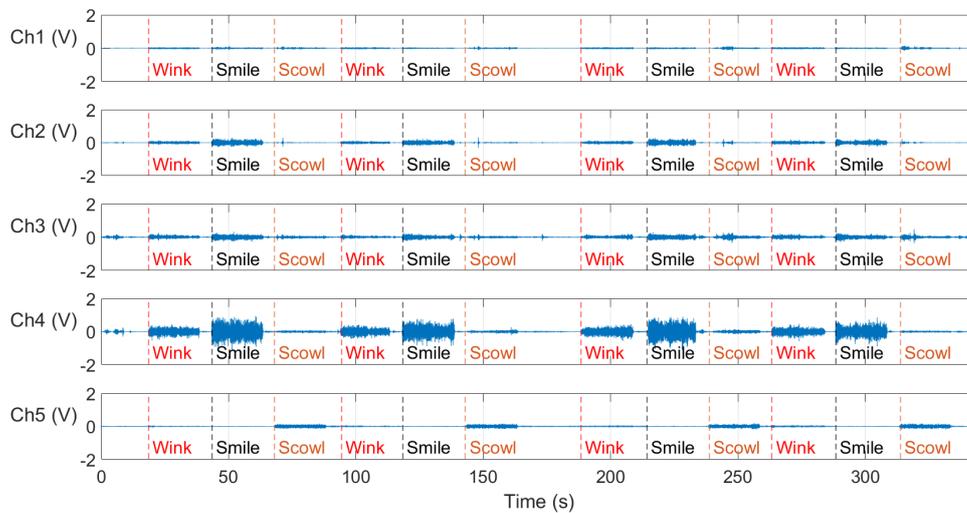


Figure 3.15: sEMG signals during facial expressions protocol. sEMG signals recorderd by each channel: Ch1 - *Anterior Temporal*, Ch2 - *Masseter*, Ch3 - *Digastric*, Ch4 - *Zygomatikus Major*, Ch5 - *Corrugator Supercilii*.

In the Figure 3.16, it is possible to observe how there is muscle activation on all channels except the last one during the wink. As expected, a more intense muscular response is on the zygomatic and moderate on the digastric and masseter. On the

other hand, the smile has a wide activity on the zygomatic, but evident effects are also visible on the masseter and digastric. Finally, as regards the scowl, it activates a muscular reaction only on the corrugator.

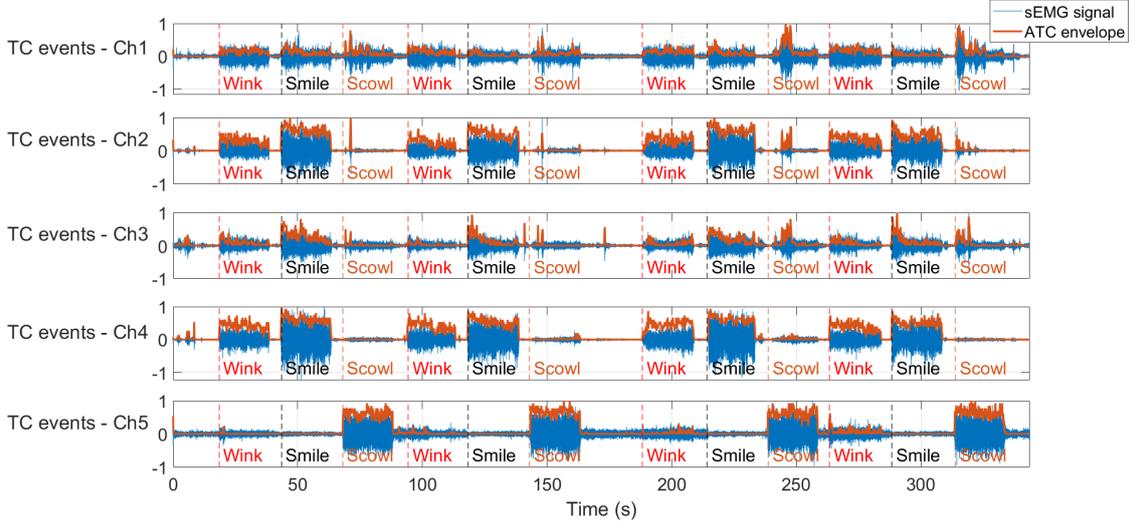


Figure 3.16: ATC envelopes during facial expressions protocol. Channel: Ch1 - *Anterior Temporal*, Ch2 - *Masseter*, Ch3 - *Digastric*, Ch4 - *Zygomaticus Major*, Ch5 - *Corrugator Supercilii*.

Regarding the above, it is possible to be successful with the event-driven approach by analyzing the ATC signal's envelope to monitor facial expressions.

### 3.2.5 Cough gesture

Finally, cough has been also analyzed as a gesture to be recognized for a potential future application to the detection and diagnosis of voice dysfunctions [79]. The muscular response could be observed on the digastric muscle but also the zygomaticus. Muscle activity during cough varies greatly from person to person since it is not a standard movement but also due to the difficulty of simulating it. It has been decided to acquire the muscles already tested with the other gestures during the cough simulation. Registration has been made under these conditions:

- simulation of cough for 5 times with 10s of pause between one repetition and another;
- resting of 20 s;
- cough for 5 times with 10-seconds rest-time between the different repetitions.

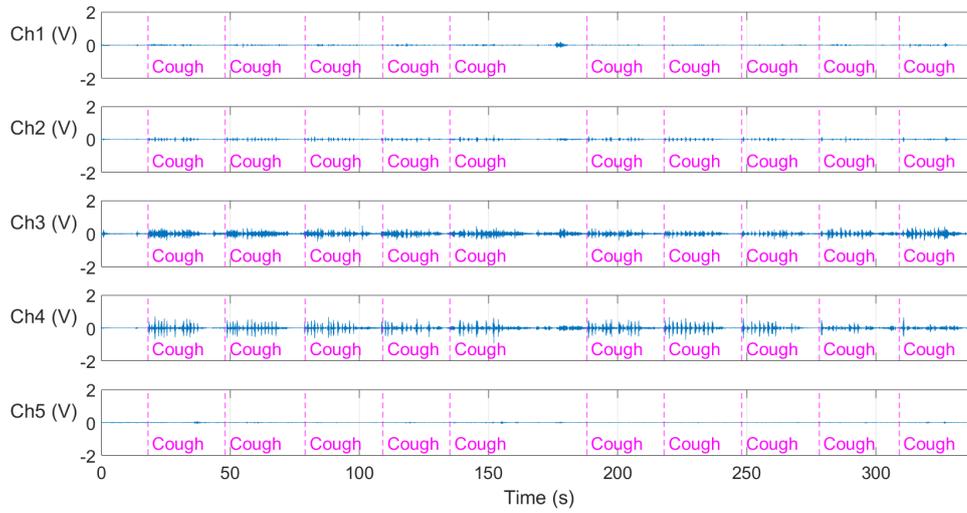


Figure 3.17: sEMG signal during cough. Channel used for recording: Ch1 - *Anterior Temporal*, Ch2 - *Masseter*, Ch3 - *Digastric*, Ch4 - *Zygomaticus Major*, Ch5 - *Corrugator Supercilii*.

The cough, as shown in the Figure 3.17, is a movement that is neither static like a smile nor periodic like a bite. It varies depending on how it is simulated. An evident muscular activity is on the digastric and the zygomatic, but also the masseter appears to be involved during the cough.

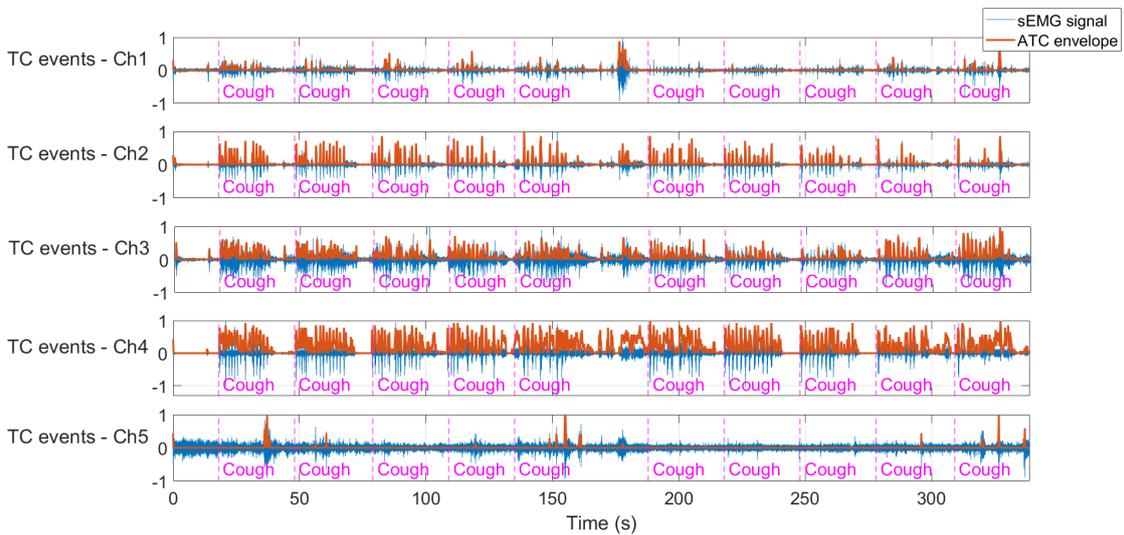


Figure 3.18: ATC signals envelope during cough. Channel: Ch1 - *Anterior Temporal*, Ch2 - *Masseter*, Ch3 - *Digastric*, Ch4 - *Zygomaticus Major*, Ch5 - *Corrugator Supercilii*.

Furthermore, it can also be noted that after the first session during the resting phase, the affected muscles are not immediately inactive, probably due to the effort involved in the execution of the movement.

Finally, observing the ATC envelope (Figure 3.18), it can be concluded that the activation is not present on the temporal and corrugator since these muscles are far from the area of the face most involved in carrying out the movement.

After the muscles to be investigated and the facial gestures to be recognized have been decided, it has been necessary to establish an experimental protocol to be performed by a group of volunteers in order to train machine learning algorithms to recognize facial gestures.

# Chapter 4

## Data Acquisition

In the first preliminary phase, the movements to be performed and the facial muscles involved have been studied. A Data collection of the sEMG signal from different people, during the execution of the movements listed in Chapter 3, is fundamental to cover a proper sample of the population. The movements chosen for the definition of an experimental protocol can be divided into:

- *static gestures*: opening, closing, wink, and smile;
- *transient gestures*: bite, cough, and talking.

The electrode positioning setup has been defined to begin the data acquisition phase in order to standardize signal acquisition during the data collection.

### 4.1 Performed Movements

The list of gestures to be performed by the volunteers was selected on the basis of some literature, presented in the Section 2.2 and the analysis carried out in Chapter 3. The definitive list of movements is reported in the following:

- **Mouth opening**: it consists in the lowering of the mandible (Figure 4.2a). The mainly used muscle is *digastric*.
- **Mouth closing**: it is the elevation of the jaw without the clenching of the teeth (Figure 4.2b). The involved muscles are *anterior temporal* and *masseter*.
- **Bite**: it consists in the sinking of the teeth in a food, to detach a piece and then to chew it (Figure 4.2c). *Anterior temporal* and *masseter* are the most used.
- **Cough**: it is a modified breath, consisting of a short inhalation phase followed by a sudden and violent exhalation (Figure 4.1d). The mainly used muscles is *digastric* and *zygomaticus major*.

- **Wink:** it is an act of closing one eye quickly (Figure 4.1e). The useful muscles are *zygomaticus major*, *anterior temporale* and *corrugator supercillii*.
- **Smile:** it consists in the pulling the mouth corners upward and laterally (Figure 4.1f). The muscle involved is *zygomaticus major*.
- **Scowl:** it is the act of frowning. The most used muscle is *corrugator supercillii*.
- **Talking:** it entails lip movement while reading a text (Figure 4.1h). Involved muscle is *digastric*.

The mouth closing and the resting position are considered a single class.

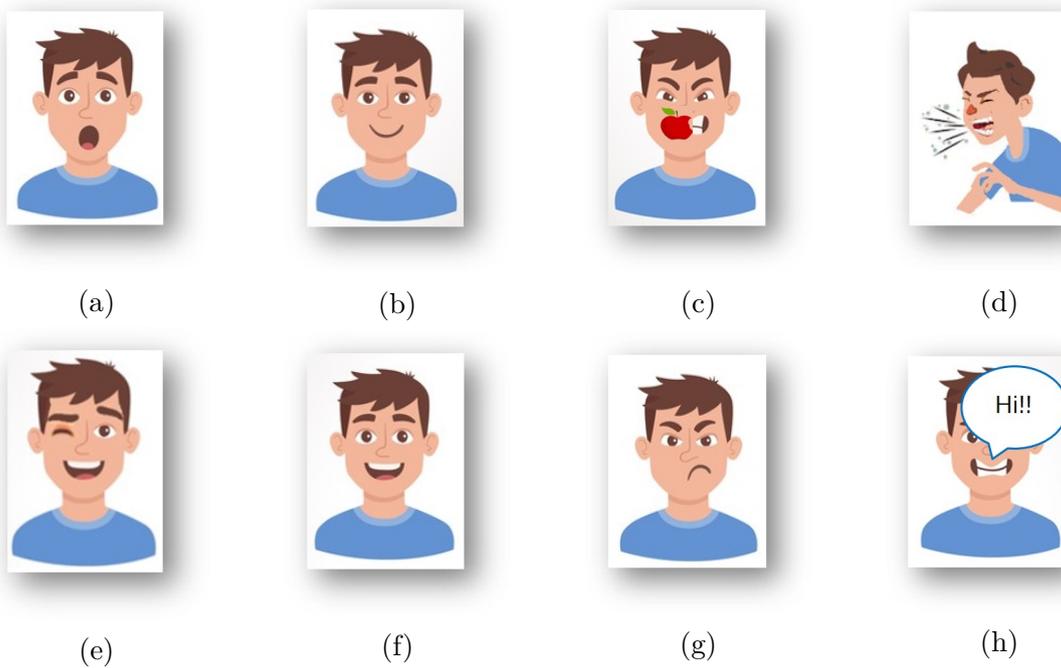


Figure 4.1: Examples of all performed movements: (a) mouth opening, (b) mouth closing, (c) bite, (d) cough, (e) wink, (f) smile, (g) scowl, (h) talking.

## 4.2 Electrodes Placement

The electrodes' placement is undoubtedly a critical aspect since the facial muscles are tiny and close to each other. Consequently, considering the problematic of muscular cross-talk, a careful application is needed. Other problems may be encountered due to the shape of the face, the possible presence of a beard or double chin, which can make it challenging to identify the muscle and place the electrodes. Moreover, depending by skin condition or length of beard, electrodes attachment could vary

during the entire protocol, affecting the quality of acquired signals. sEMG electrodes are the same used in the preliminary analysis (i.e., H124SG [71]). The final electrodes placement on right face side is the following (see Figure 4.2):

- First couple of electrodes is placed on the *anterior temporal*. The main contribution is to the *bite* and *wink*, but some useful value are recorded also during *smile*.
- Second pair is placed on the *masseter*. These electrodes are mainly used in *bite*.
- Third couple is placed on the *digastric*. Main effects are obviously on the *mouth opening*, but they are also necessary for *talking* and *cough*.
- Fourth pair is placed on the *zygomaticus major*. These electrodes are mainly used in *smile*, but could have also a relevant effect on *wink* and *cough* (depending by how the subject perform such expression).
- Fifth couple is on the *corrugator supercilii* and it has evident effects on *scowl* but also on *wink*.
- Last, the reference electrode has been placed on the forehead in an electrical neutral area.

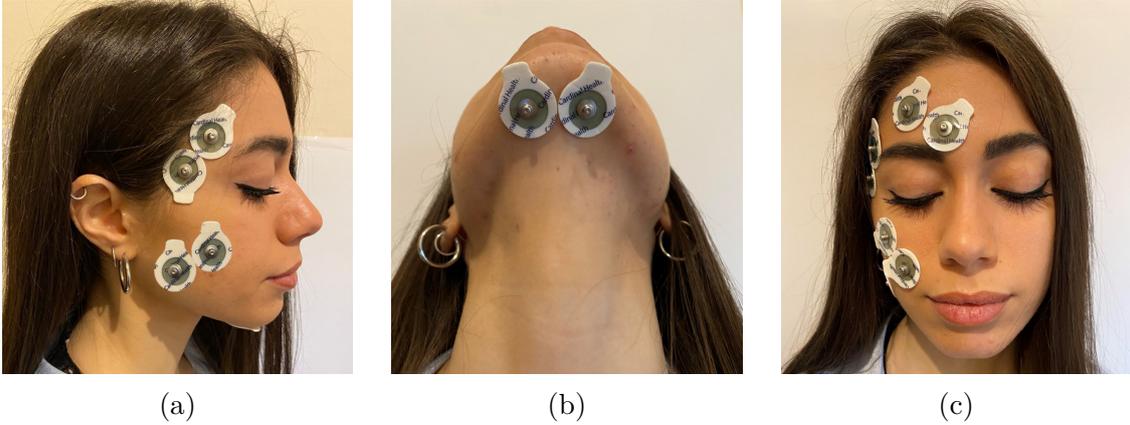


Figure 4.2: Electrode positioning setup on the right hemiface: (a) *Anterior temporal* and *masseter*, (b) *digastric*, (c) *zygomaticus major* and *corrugator supercilii*.

### 4.3 Acquisition Protocol

Following the preliminary analysis, an *in vivo* experimentation has been launched in order to obtain enough data from various individuals to train and test the classifiers. The experimental protocol, abiding with the current regulations for scientific

research in healthy human volunteers experimentation promoted by University of Study of Torino, has received approval from the corresponding bioethics committee [80]. Twenty-one healthy subjects, 16 males and 5 females (with age between 24 and 39 years old), have been recruited from Polytechnic of Turin's Department of Electronics and Telecommunications (DET). They have all been thoroughly briefed on the experiment and its potential risks. They have been received informed consent for the study and agreed to participate in the signing of the form, drafted according to the local bio-ethical committee's guidelines. The subjects' privacy rights have been observed since they have been assigned a progressive code, and the latter has been the only identifier used within the study.

During the experiment, each subject seated comfortably, straight on a chair with their hands put on the knees, kept at  $90^\circ$ . After explaining the study, the facial skin has been cleaned with 70% alcohol and cotton to reduce skin-electrode impedance before electrodes have been placed on the right side of the subject's face. This placement is critical due to the presence of a double chin or beard in males. The electrodes have been positioned as described above. In particular, for the electrodes' locations on the anterior temporal muscle and the masseter muscle, subjects were requested to clench their teeth. Regarding the digastric and zygomaticus major muscles, on the other hand, they have been asked to raise their head and smile to adjust the electrode's position on the skin overlying the submental region and close as possible to the mouth corner, respectively.

In order to be compliant with the COVID-19 regulations, each subject has been wearing the protective mask until the electrodes have been placed on the face. Therefore, during the texts' execution, the supervisor used a face shield as additional protection and a protective mask.

A short calibration phase, in which the subject has been asked to quickly perform all the gestures, has been conducted before the experiment execution in order to verify muscle activations on all the channels with prepared electrode configuration.

### **4.3.1 Experimental Protocol: Graphical User Interface (GUI)**

The subjects during the protocol have been supported by a Graphical User Interface (GUI) developed in the Matlab<sup>®</sup> environment through the Toolbox App Designer. Below, the features of the application are briefly explained.

- Once the application is launched, the start screen represented in Fig.4.3 shows up.

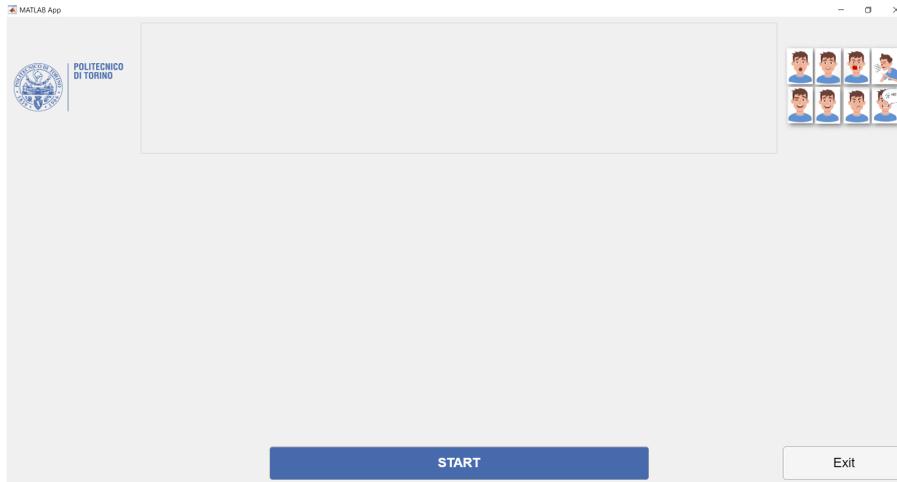


Figure 4.3: Start screen: *Start* and *Exit* are button objects.

- When the user presses the *Start* button at the beginning of the session, two new buttons, *Italian* and *English*, appear on the interface (see Figure 4.4) for choosing the language to be used to give indications regarding the movements to be performed.

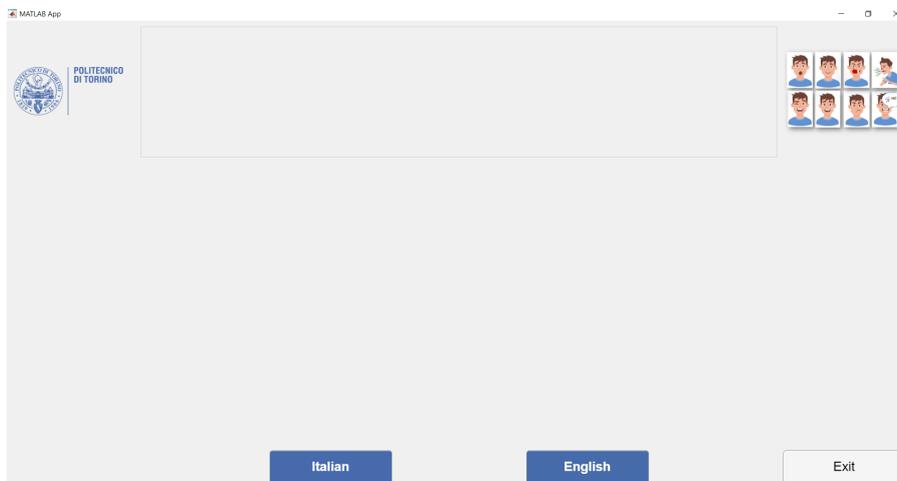
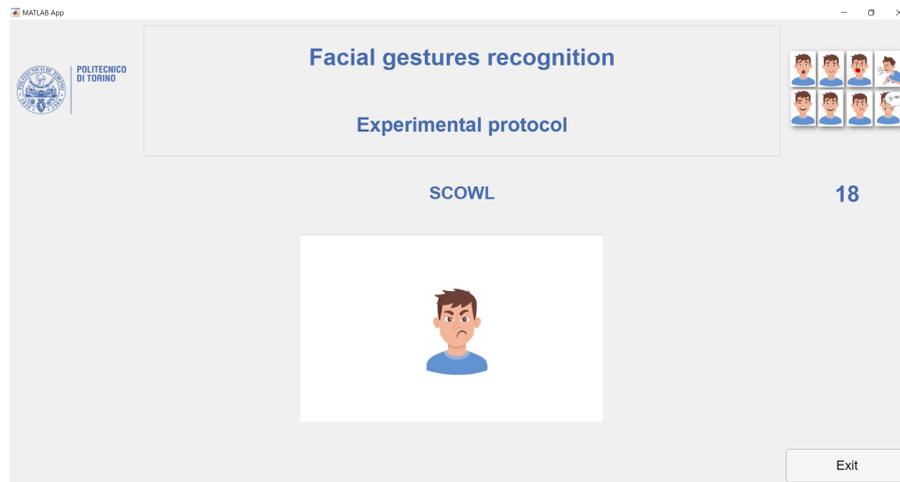


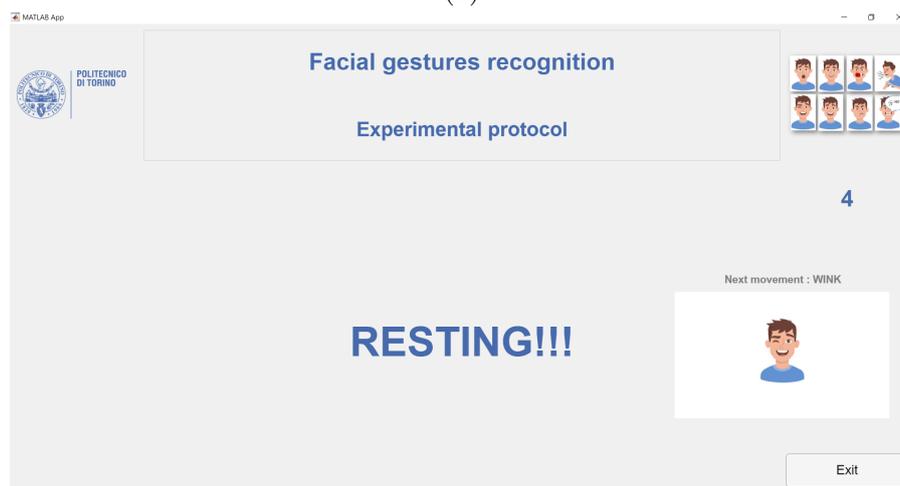
Figure 4.4: Language choice: *English* and *Italian* are button objects enable.

- Then, the session starts and the supervisor simultaneously launches the acquisition on the g.Recorder program (described in Sec. 3.1.2).
- In the view of the experimental protocol screen, the movement to be performed is shown with a cartoon image, and a down counter is inserted to

mark a rhythm. Furthermore, during the resting phase, a preview of the next movement appears on the screen to facilitate the subject in its execution.



(a)



(b)

Figure 4.5: Experimental protocol view screen. (a) Movement to be performed. (b) Preview of the movement to be executed.

- After the end of the first recording, a pop-up requesting to continue appears (shown in Figure 4.6a). If the continue button is pressed, the pause time in min is asked for before the start of the new recording (see Figure 4.6b).
- The exit button is always enabled to be able to exit the application at any time after responding to a confirmation pop-up (see Fig. 4.7).



Figure 4.6: Pop-up request for protocol continuation: (a) request to continue with the sEMG acquisition; (b) dialogue box to insert pause time before the next recording.

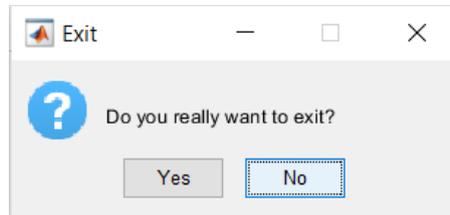


Figure 4.7: Exit pop-up.

### 4.3.2 Experimental Protocol: Phases

During the protocol, the subjects have to perform each gesture sequentially for 20 s, with 5 s of rest between them. The sequence of movements within the same session was performed three times, with 20 s of rest between each repetition. The protocol is repeated three times with a 2 min pause between the sessions, and in each session, different foods have been tested. The order of the foods to be examined during the experiment was entirely up to each subject in order to increase the variability of data collection. The subjects have been guided in the execution of the gestures by a GUI, which help them in performing the required action with the proper timing. The protocol can be summarized in the following steps:

1. The GUI is launched to mark a rhythm (see Figure 4.3) by the subject and the acquisition of the signal is started.
2. The GUI shows the preview of the movement to be performed (Figure 4.5b), according to the list of movements detailed in Sec. 4.1;
3. The gesture is maintained for 20 s.
4. The rest of 5 s is observed. If there are any remaining movements for the actual repetition, the flow returns to point 2.

5. After completing all series of the movements, a 20-seconds rest time is followed.
6. New repetition restarts from 2 unless three repetitions have already been done.
7. Data is stored on the computer. At the end of the session, a pop-up request (shown in Figure 4.6a) allows the supervisors/user to decide if go through another session. If a further session is needed, a dialogue box (Figure 4.6b) is displayed to insert the pause minutes before it begins. Flow restarts from 2.

During the acquisition, it has been possible to observe the signal collected on the g.Recorder software in real-time (see Figure 4.8). Accordingly, if there have been problems with the electrode's displacement, the electrodes have been repositioned in the best possible way at the end of the session.

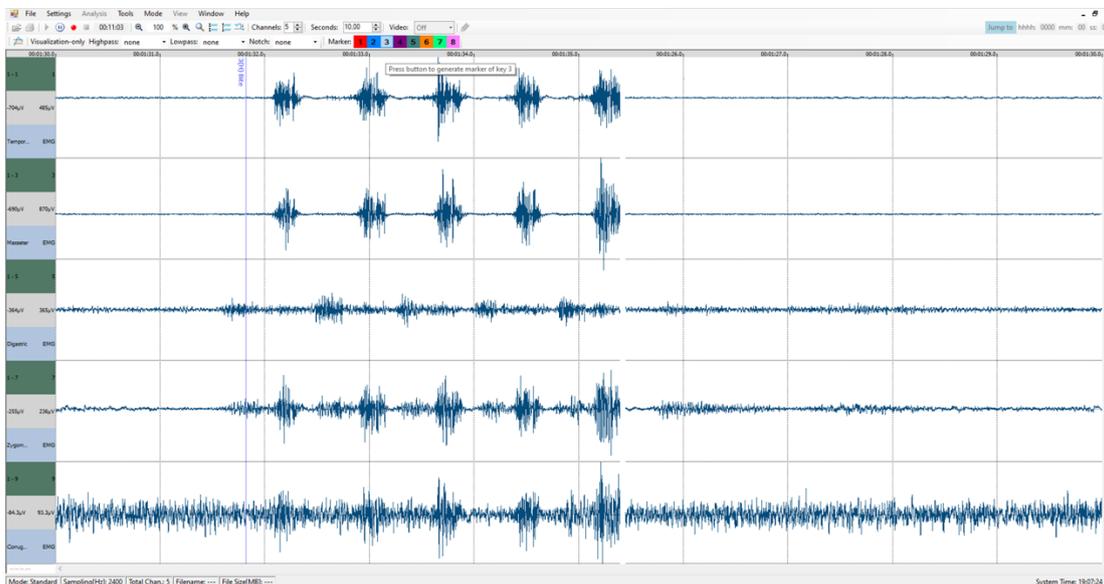


Figure 4.8: Recording of the sEMG signal seen in real-time on g.Recording.

## Chapter 5

# Processing Data

The data acquired during the acquisition phase have been processed in order to extract the useful feature (i.e., ATC parameter) for the classification process. Once the multi-channel sEMG data have been saved on the computer, they have been uploaded to Matlab<sup>®</sup> to be elaborated. A Matlab script has been implemented to process sEMG signal and extract the ATC feature, which will corresponds to the classifier inputs. The code flow is the following:

- All the muscular sEMG signals of each subject, corresponding to the movement execution of the acquisition protocol described in 4.3, have been loaded separately.
- The first 4 seconds of the recording have been removed for the presence of a spike due to the settling of the g.tec system.

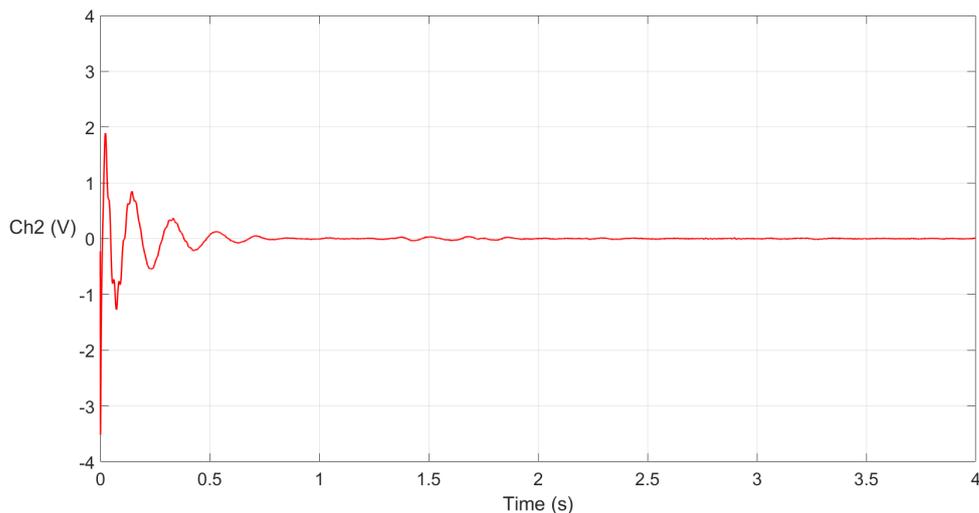


Figure 5.1: Settling time g.tec system. Ch2: *Masseter*.

In these instants of time (shown in Figure 5.1) the system stabilizes so that the acquired signal can reach a certain precision and remain within the specified accuracy range.

- The signals have been pre-filtered hardware by g.HIamp with a high pass filter with a cutoff frequency of 5 Hz and a stopband filter (i.e., Notch filter) at 48-52 Hz. As a result, in order to eliminate low fluctuations signal caused by movement artefacts, and considering a typical sEMG frequency spectrum, the sEMG signals have been further filtered in post-processing using a 10th order Butterworth bandpass filter with a passband between 30-400 Hz. In the Figure 5.2 the module of the frequency response of the implemented filter can be observed.

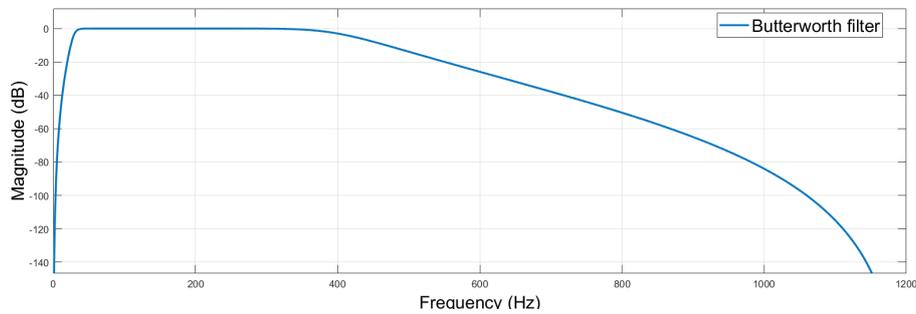


Figure 5.2: Frequency response Butterworth filter.

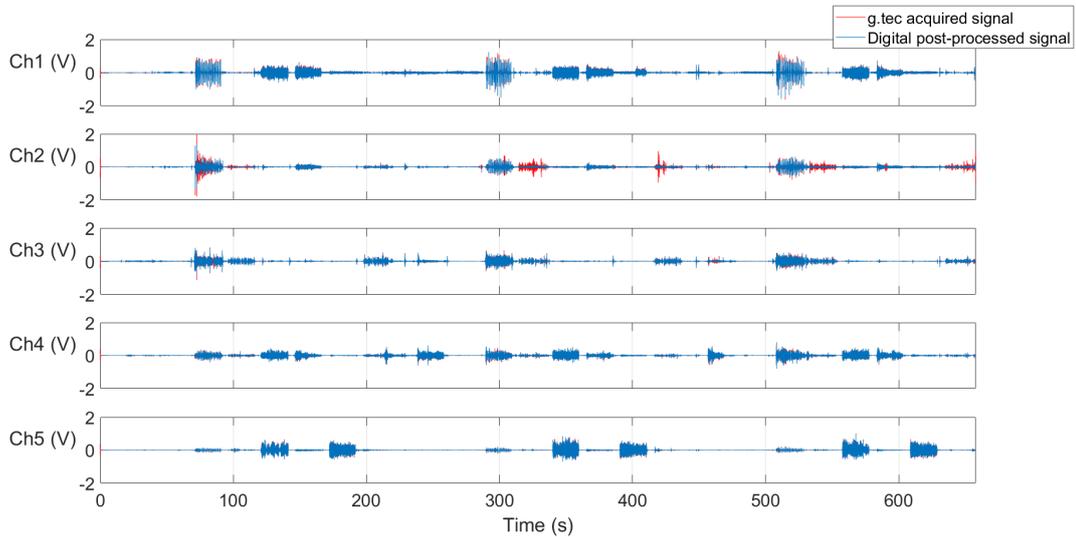


Figure 5.3: sEMG signals pre- and post-digital filtering. Muscles activation investigated by each channel: Ch1 - *Anterior Temporal*, Ch2 - *Masseter*, Ch3 - *Digastric*, Ch4 - *Zygomaticus major*, Ch5 - *Corrugator supercillii*.



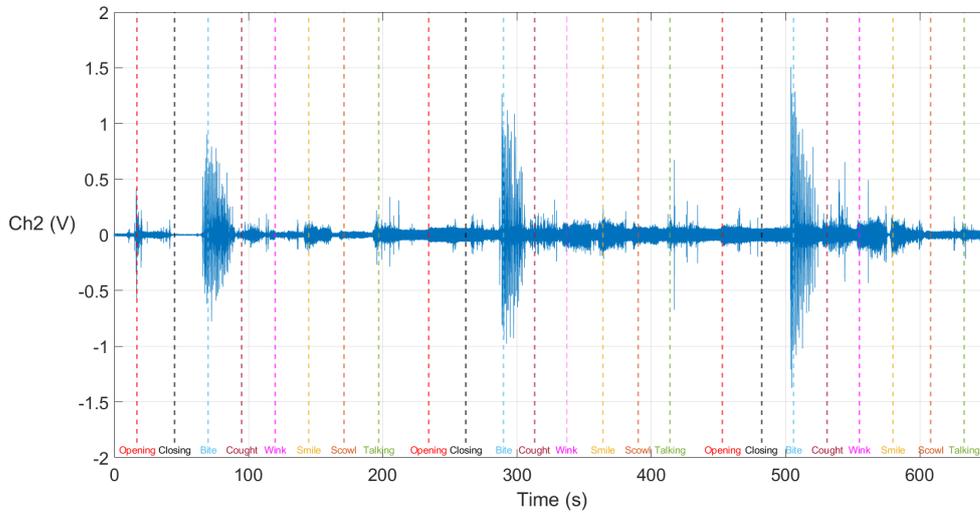


Figure 5.5: sEMG signal characterized by a high environmental noise due to a possible displacement electrode. Muscle activation investigated by Ch2 -*Masseter*.

The threshold value changes according to the channel electrical features and varies among subject because of different skins and bodies conditions. In particular, the baseline noise can be influenced by: electrode placement and state; involuntary facial movements or saliva swallowing; skin condition such as possible beard presence. Figure 5.5 shows masseter muscular activity and highlights some of the problems mentioned above. It can be observed that the signal baseline after first bite movement probably suffered a slight electrode detachment on the masseter, caused by effort caused by biting the nougat. On the other hand, during the acquisition, the background noise gradually increased because the subject had a beard which reduced the electrode-skin contact.

For these typical issues, the threshold has been evaluated, taking into account the baseline noise for the entire duration of the recording.

In particular, in the case examined, not doing this type of threshold assessment would have led to an inadequate evaluation of the ATC activation related to each movement performed in terms of events.

- An hysteresis of 15 mV has been considered around the computed threshold value in order to take into account spurious muscle signal activation ATC. The algorithm evaluates how many times the sEMG signal has gone above the  $V_{th} \pm 15$  mV threshold. The ATC feature has got by summing up all TC events in a time window of 130 ms.

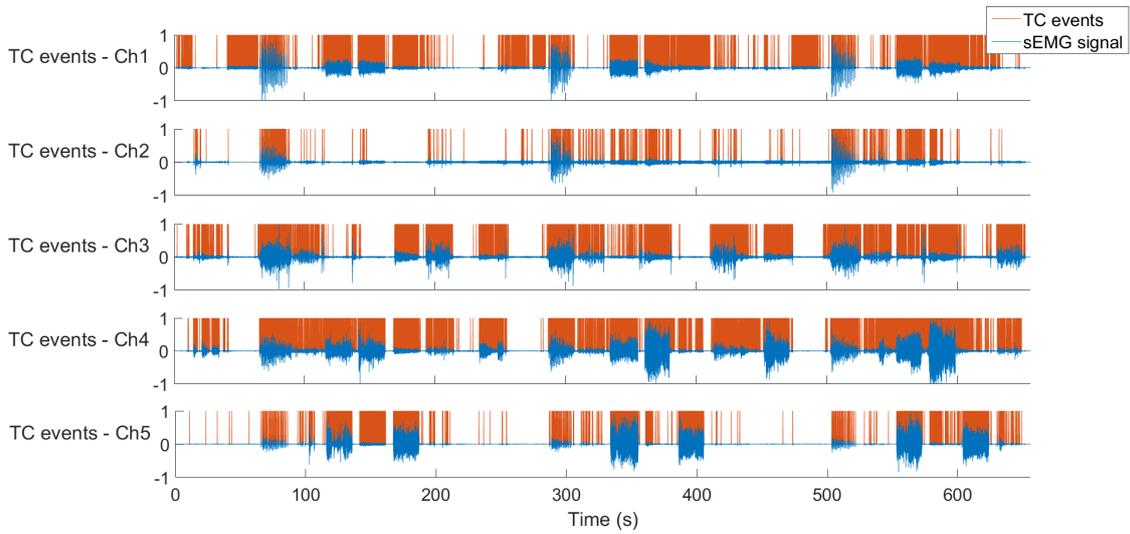


Figure 5.6: TC events superimposed on the sEMG signal. Channel: Ch1 - *Anterior Temporal*, Ch2 - *Masseter*, Ch3 - *Digastric*, Ch4 - *Zygomaticus major*, Ch5 - *Corrugator supercili*.

- The relative ATC envelopes have been superimposed on the sEMG signals (see Figure 5.7) in order to observe whether the threshold imposed was correct in assessing the muscle activation of each gesture performed.

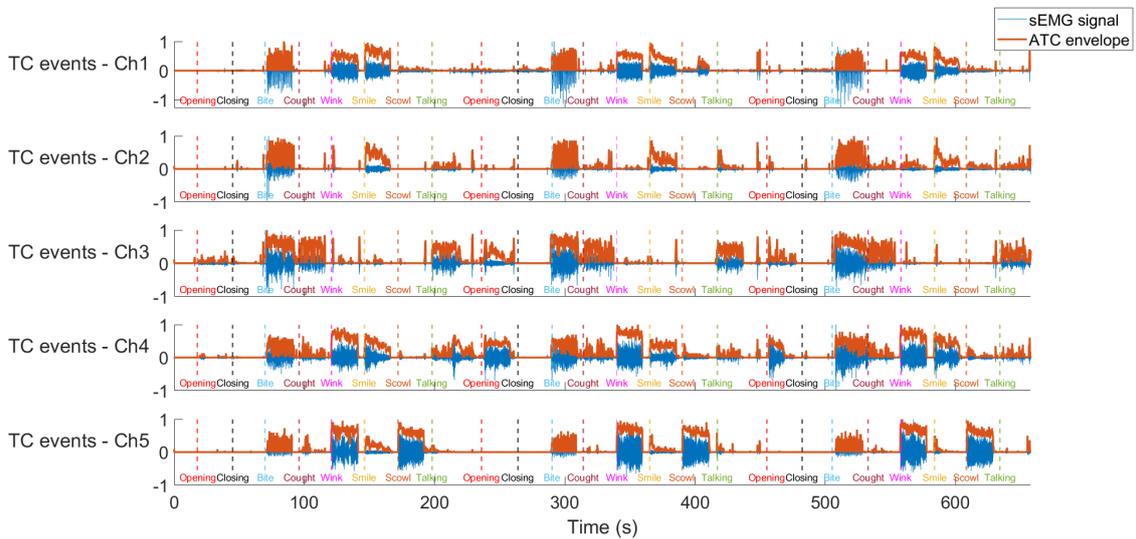


Figure 5.7: ATC signals envelope of facial protocol. Muscles activation investigated by each channel: Ch1 - *Anterior Temporal*, Ch2 - *Masseter*, Ch3 - *Digastric*, Ch4 - *Zygomaticus major*, Ch5 - *Corrugator supercili*.

- The ATC signal is saved in .csv format, useful for the classification phase.

After analyzing all of the signals, a structure has been created for each subject's data relating to the three protocol repetitions in order to make them easily accessible for future analysis.

## Chapter 6

# Machine Learning Implementation

Considering the number of performed gestures, and their nested information related to the acquisition channels, the use of Artificial Intelligence (AI) algorithm has been identified as the proper way to achieve facial recognition reaching suitable performance. In particular, the aim has been to understand whether it was possible to recognize gestures not only by observing muscle activation alone but by implementing ML models. This chapter presents the results regarding the accuracy of the classifiers' performance in gesture recognition. All models have been developed using ML libraries developed for the Python<sup>®</sup> environment. For all ML algorithms implemented, the following steps were applied:

- The available data set has been divided into the *training set* and *test set* with the *Stratified k-fold* algorithm, choosing  $k = 5$ . This algorithm splits the data set on  $k$  fold so that distributes approximately in each fold the same percentage of samples for each class of the available dataset. Moreover, before the splitting, data are also shuffled to enhance the performance of the classifiers.
- A hyperparameter optimization has been implemented with *GridSearchCV*, a library function included in sklearn's model selection package [81]. It is used to train a machine learning model with multiple combinations of hyperparameter training and find the best variety of parameters that optimize the evaluation metric. It creates an exhaustive set of hyperparameter and training models for each combination.
- For each ML algorithm, a statistical evaluation has been carried out in terms of accuracy, recall, precision, and F1-score to evaluate the goodness of the classification model found.

## 6.1 Parameters Algorithms

This section describes the parameters considered for each ML algorithm and the ones used during the optimization phase before the final choice of the most accurate model.

### 6.1.1 Random Forest (RF)

Random Forest has been implemented using *scikit-learn* library [81]. The following parameters have been varied [81]:

- **n\_estimators**: is the number of trees in the forest. On but one hand, many trees allow obtaining a better performance of the model, but on other hand, it could slow down its training by increasing the computational costs.
- **max\_features**: is the maximum number of features to be considered while seeking out the satisfactory split in trees. The higher this number, the better the algorithm's behavior in terms of recognition accuracy, but at the same time, the less fast the model processing time will be. There are several ways to search for them:
  - *Auto*: it consists of taking all the characteristics that make sense for a given tree. No restrictions are placed on the individual decision tree.
  - *Sqrt*: for each run, the square root of the available variables are taken as a max features number.
  - *Log2*: the log2 of the features is taken into account.
- **min\_samples\_leaf**: is the sample number required to be at a leaf node. It is a useful parameter for the overfitting problem: it ensures that the tree cannot overfit the training data set by creating branches with a number of samples smaller than the one chosen at the leaf node.
- **min\_samples\_split**: is the minimum number of samples required to split the node. High values can lead the model to be too generalized with the risk of losing some specific learning relationships for the sample selected for a tree.
- **max\_depth**: is the maximum depth of a single tree. The deeper the tree grows, the greater the model's complexity will be because it acquires more information on the data by adapting too much on the latter, losing the ability to generalize on the data set.

### 6.1.2 k-Nearest Neighbour (k-NN)

*scikit-learn* library [81] is used to develop k-NN model. The following parameters have been taken into account [81]:

- **n\_neighbors**: is the number of nearest neighbors. If it is high, the field cost increases, while if it is reduced, the final prediction of the results will be more influenced by noise. The choice of k depends on the available data set. It is recommended to perform preliminary tests to highlight the best number. Consequently, for each collection of data taken into consideration, the graph of k value and the dataset's corresponding error rate have been plotted. Then, it is possible to choose a range of k values to be tested during the tuning phase with a low error rate.
- **weights**: is the weight function that defines each point's weight within the model. There are two types of weighted functions:
  - *Auto*: in each neighbourhood, all samples are weighted similarly;
  - *Distance*: the points are weighted based on the inverse of their distance: closer elements will be more relevant.
- **algorithm**: indicates the type of algorithm used to determine the neighborhood. There are several metrics:
  - *brute*: it is the simplest and most intuitive as it provides for calculating the distances among all the pairs of points in the data set. It is the least performing approach that can only be used with datasets that are not too large; otherwise, the calculation times increase exponentially.
  - *kd\_tree*: it is a methodology based on the use of data structures to make the algorithm more efficient. By making aggregate considerations, it aims to decrease the number of necessary calculations: an example could be that of having a generic point A far from a point B, which however, is close to a third point C, then A is also far from C.
  - *ball\_tree*: it was introduced to improve the previous algorithm for large data sets; differs from the *kd\_tree* in data aggregation methodologies. This method attempts to reduce the required number of distance calculations by efficiently encoding the sample's aggregated distance information.
  - *auto*: adapts the algorithm based on the dataset to be managed.
- **p**: is an integer that identifies the type of formula for calculating distances.
  - $p = 1$  is equivalent to using the Manhattan distance.
  - $p = 2$  Euclidean distance is used.

### 6.1.3 Support Vector Machine (SVM)

SVM have been implemented using *scikit-learn* library [81]. The following parameters have been utilized [81] :

- **C**: is a regularization parameter. When C is a low value, a more significant number of misclassifications occurs due to the choice of a decision boundary with a large margin. SVM with a large C tries to reduce erroneous predictions adopting a decision boundary with a lower margin, resulting in a model to more generalized.
- **kernel**: indicates the type of kernel used by the model. It can be *poly*, *rbf* or *sigmoid* among those available.
- **gamma**: is used with non-linear problems. It is a parameter that considers the influence of the distance of a training set sample. Low gamma values result in multiple points clustered together, outlining a wide-ranging similarity among points themselves. For high gamma values, the samples must be very close together to be considered in the same group (or class). Therefore, models with enormous gamma values tend to be poorly generalized. There are two types of settings for this parameter:
  - scale: is equal to  $1 / (n\_features * variance)$ ;
  - auto: uses  $1/n\_features$  as gamma value.

### 6.1.4 Artificial Neural Networks (ANN)

The neural network model was created using a combination of *keras* [82] and *tensorflow* [83] libraries. In order to find the right model, many different combinations have been tried by varying these parameters:

- **architecture**: is the structure of the neural network, which is composed as follows:
  - input layer: k neurons as the number of channels;
  - hidden layer: dense layer with n neurons and Rectified Linear Unit (ReLU) as the activation function, which sends the result to the output if positive; otherwise, it will output zero;
  - output layer: m neurons as the number of gestures. Softmax is the activation function and creates a vector of probabilities that sum to one, representing the potential results' probability distributions.
- **learning rate (LR)**: controls the learning speed. Adam optimizer have been used during the training phase with a learning rate that have been varied. It is

a stochastic gradient-based optimizer that relied on the adaptive estimation of the mean (first moment) and the uncentered variance (second-order moments). The aim is to update the net weights during the epochs of the training phase.

- **epochs**: indicates how many times each data point will be used.

## 6.2 Total Gestures Recognition

All gestures performed during the acquisition protocol has been classified. Both biting and chewing have been included in one class, named bite, no distinction having been made based on the type of food tested. Below, the results reached with to be considered ML algorithms.

### 6.2.1 Random Forest (RF)

After the optimization, the chosen hyperparameters were shown in Table 6.1:

Table 6.1: RF hyperparameters setting.

Hyperparameter	Value
n_estimators	900
max_features	'log2'
min_samples_leaf	2
min_samples_split	8
max_depth	10

The model has a low ability to predict gestures such as cough and talking, as shown in Table 6.2, since the poor repeatability of these actions over time and in different situations, even by the same subject. In fact, the average accuracy achieved is 61.23%.

Table 6.2: RF results.

	<b>Accuracy(%)</b>	<b>Precision(%)</b>	<b>Recall(%)</b>	<b>F1-Score(%)</b>
Opening	59.26	49.13	47.44	48.27
Closing	80.32	67.85	80.30	73.55
Bite	62.36	54.12	62.38	57.96
Cought	23.13	37.96	6.18	10.63
Wink	67.27	76.91	56.87	65.39
Smile	78.63	59.24	78.66	67.58
Scowl	85.19	78.37	85.21	81.64
Talking	33.66	34.30	32.42	33.33
<b>Avg</b>	<b>61.23</b>	<b>57.23</b>	<b>56.18</b>	<b>54.79</b>

## 6.2.2 k-Nearest Neighbour (k-NN)

Table 6.3 report the algorithm model obtained after optimization.

Table 6.3: k-NN hyperparameters setting.

<b>Hyperparameter</b>	<b>Value</b>
n_neighbors	100
weights	'uniform'
p	2
algorithm	'brute'

With this model, a very low average accuracy (60.66%) has been obtained, even moderately lower than the RF one. It can be observed in Table 6.4 that the opening movement accuracy slightly exceeds 50%, probably due to the displacement of the electrodes over the digastric muscle in some bearded male subjects, resulting in the acquisition of noised signals. On the other hand, TC events are distributed during the opening in a fragmented way comparable to ones obtained during talking or coughing. Therefore, the three movements turn out to be incorrectly classified.

Table 6.4: k-NN results.

	<b>Accuracy(%)</b>	<b>Precision(%)</b>	<b>Recall(%)</b>	<b>F1-Score(%)</b>
Opening	50.73	50.82	42.28	46.16
Closing	80.30	67.59	80.30	73.40
Bite	62.83	54.68	60.11	57.27
Cought	19.94	26.56	8.41	12.77
Wink	75.40	77.15	56.28	65.08
Smile	78.23	58.84	78.23	67.16
Scowl	85.38	77.81	85.38	81.42
Talking	32.46	33.76	32.46	33.10
<b>Avg</b>	<b>60.66</b>	<b>55.90</b>	<b>55.43</b>	<b>54.55</b>

### 6.2.3 Artificial Neural Networks (ANN)

After the optimization, the best hyperparameters are presented in Table 6.5:

Table 6.5: ANN hyperparameters setting.

<b>Hyperparameter</b>	<b>Value</b>
architecture	5-32-16-8
LR	0.001
epochs	50

The recognition ratio obtained is 61.05%, with a performance very similar to the RF model and slightly higher than the k-NN one. The Table 6.6 shows that the closing movement achieves a higher accuracy than the other classes. The same thing can be observed in the previous models. The availability of more data and the similarity of the closing sEMG signal among the different subjects allows this outcome to be achieved. Furthermore, during the wink, the accuracy of the model's performance is over 60% since muscle recruitment is not the same for all subjects. Muscular activation varies according to how the movement is performed.

Table 6.6: ANN results.

	<b>Accuracy(%)</b>	<b>Precision(%)</b>	<b>Recall(%)</b>	<b>F1-Score(%)</b>
Opening	57.85	47.40	48.83	48.10
Closing	81.92	65.92	81.92	73.05
Bite	64.14	52.39	64.14	57.67
Cought	24.92	33.16	7.24	11.89
Wink	63.67	76.91	56.77	65.32
Smile	75.30	61.47	78.66	67.58
Scowl	84.86	77.01	84.86	80.74
Talking	35.72	37.48	25.68	30.48
<b>Avg</b>	<b>61.05</b>	<b>56.47</b>	<b>55.59</b>	<b>54.37</b>

## 6.2.4 Support Vector Machine (SVM)

The best prediction SVM model is characterized by parameters described in Table 6.7.

Table 6.7: SVM hyperparameters setting.

<b>Hyperparameter</b>	<b>Value</b>
C	1
kernel	'rbf'
gamma	auto

A correct prediction is obtained with gestures like smile and scowl, as can be seen in Table 6.8, because they activate highly the zygomatic and corrugator supercillii. They are, therefore, more defined movements, and their execution appears to be performed similarly by all the subjects.

Table 6.8: SVM results.

	<b>Accuracy(%)</b>	<b>Precision(%)</b>	<b>Recall(%)</b>	<b>F1-Score(%)</b>
Opening	55.70	43.02	45.84	
Closing	81.79	65.75	81.67	72.85
Bite	67.67	55.31	61.90	58.42
Cought	21.85	44.76	3.19	5.96
Wink	60.32	79.23	52.87	63.42
Smile	78.61	58.50	78.61	67.08
Scowl	86.62	75.79	86.30	80.70
Talking	32.15	33.13	31.93	32.52
<b>Avg</b>	<b>60.59</b>	<b>57.69</b>	<b>54.94</b>	<b>53.35</b>

The algorithms' accuracy has been not good since movements, like cough and talking, have not been well performed or appeared less defined. This last aspect had a significant impact on recognition performance.

## 6.3 Reduced Gestures Recognition

The success rate achieved by recognizing all eight gestures has not been optimal. As a result, gestures that have been negatively impacted the models' predictive ability and exhibited very low accuracy have been removed from the target. Also in this analysis, no distinction has been made on the bite based on the food tested. The setting of the best parameters has been sought for each model, and it has been possible to achieve better performance in recognition.

### 6.3.1 Random Forest (RF)

The best setting for a RF model obtained was shown in Table 6.9:

Table 6.9: Better compromise of parameters for Random Forest.

<b>Hyperparameter</b>	<b>Value</b>
n_estimators	100
max_features	'auto'
min_samples_leaf	1
min_samples_split	5
max_depth	10

Compared to the previous case with eight classes, it manages to achieve an average accuracy greater than 10%. It can be noted that the least performing gestures are opening and biting since no discrimination has been made on the different food bite and in some bearded subjects, the opening was not well defined.

Table 6.10: Statistical results Random Forest.

	Accuracy(%)	Precision(%)	Recall(%)	F1-Score(%)
Opening	59.50	63.97	53.88	58.50
Closing	84.43	79.60	84.61	82.03
Bite	67.55	62.55	65.91	64.19
Wink	73.71	78.99	56.83	66.10
Smile	77.92	68.41	78.58	73.14
Scowl	84.64	83.03	84.64	83.83
Avg	<b>74.63</b>	<b>72.76</b>	<b>70.74</b>	<b>71.30</b>

### 6.3.2 k-Nearest Neighbour (k-NN)

The parameters that guarantee the better performance of the k-NN model are listed in Table 6.11:

Table 6.11: Better compromise of parameters for k-NN.

Hyperparameter	Value
n_neighbors	100
weights	'uniform'
p	2
algorithm	'brute'

Table 6.12: Statistical results k-NN.

	Accuracy(%)	Precision(%)	Recall(%)	F1-Score(%)
Opening	58.68	65.95	55.47	58.27
Closing	84.46	79.50	84.46	81.90
Bite	62.83	63.93	62.66	63.29
Wink	75.39	80.37	55.62	65.74
Smile	79.40	67.13	78.89	72.53
Scowl	85.24	82.34	85.24	83.77
Avg	<b>74.33</b>	<b>72.44</b>	<b>70.39</b>	<b>70.92</b>

The Table 6.12 shows the statistical results obtained with this newest k-NN version, which results in an improved statistical outcomes with respect to previous architecture. In particular, the low bite accuracy may be due to its almost periodic nature. Moreover, the chewing cycles can have different durations depending on the subjects and the food eaten. As a result, they have a strong impact on the model's prediction performance.

### 6.3.3 Artificial Neural Networks (ANN)

The results for the ANN algorithm are presented in Table 6.14 and they have been obtained with the setting, described in Table 6.13.

Table 6.13: Better compromise of parameters for ANN.

Hyperparameter	Value
Architecture	5-32-16-8-8
LR	0.001
epoch	70

Table 6.14: Statistical results ANN.

	Accuracy(%)	Precision(%)	Recall(%)	F1-Score(%)
Opening	62.81	66.48	45.04	53.70
Closing	87.91	76.54	87.91	81.83
Bite	62.57	64.97	62.57	63.75
Wink	69.11	80.51	54.82	65.22
Smile	76.41	68.81	76.41	72.41
Scowl	86.28	80.85	86.28	83.48
Avg	<b>74.18</b>	<b>73.03</b>	<b>68.84</b>	<b>70.07</b>

This model has a predictive capacity very similar to that of the previous ones, slightly moderately lower.

### 6.3.4 Support Vector Machine (SVM)

The better parameters setup for SVM model is shown in Table 6.15.

Table 6.15: Better compromise of parameters for SVM.

Hyperparameter	Value
C	1
kernel	'rbf'
gamma	'auto'

This algorithm among the one tested allows obtaining the highest overall accuracy. The worst results shown in the Table 6.16 in terms of class accuracy are related to the opening, bite, and wink movements for the reasons already described above. In particular, for the wink gesture, this outcome is due to the variability among subjects in recruiting the muscles examined during its execution.

Table 6.16: Statistical results SVM.

	Accuracy(%)	Precision(%)	Recall(%)	F1-Score(%)
Opening	62.83	62.73	55.06	58.64
Closing	84.99	70.36	84.97	82.07
Bite	67.70	62.63	65.64	63.93
Wink	70.36	79.63	56.71	66.24
Smile	77.59	70.12	77.15	73.47
Scowl	84.42	83.13	84.28	83.70
Avg	<b>74.65</b>	<b>72.88</b>	<b>70.63</b>	<b>71.34</b>

## 6.4 Jaw Movements Recognition

In order to move towards a future clinical application, it has been decided to focus only on the jaw movements, removing three acquisition channels. The channels placed on the corrugator supercillii and on the zygomaticus major have not been considered since they have been introduced for facial expressions. Within the muscles involved in chewing and biting, only the masseter has been chosen for its most evident response regarding the sEMG signal's amplitude.

### 6.4.1 Random Forest (RF)

After the optimization, the best RF performances have been achieved with the parameters shown in Table 6.17.

Table 6.17: Random Forest best parameters setup.

Hyperparameter	Value
n_estimators	100
max_features	'log2'
min_samples_leaf	2
min_samples_split	5
max_depth	10

Observing the statistical results (Table 6.20), the opening movement remains below 70% in terms of accuracy, probably always due to possible problems due to the electrode positioning with some subjects. It is also possible to analyze that the bite performance precision improves because only one of the masticatory muscles has been considered. On the other hand, considering only the mandibular movements' muscles, the risk of having unusual muscular activity on other muscles (i.e., chewing on the corrugator supercillii) has been reduced. However, it is possible to obtain an overall accuracy of close to 80%.

Table 6.18: Performances obtained with Random Forest.

	Accuracy(%)	Precision(%)	Recall(%)	F1-Score(%)
Opening	67.60	61.56	52.05	56.41
Closing	87.17	85.97	87.26	86.61
Bite	80.55	68.42	75.38	71.73
Avg	<b>78.44</b>	<b>72.76</b>	<b>70.74</b>	<b>71.30</b>

## 6.4.2 k-Nearest Neighbour (k-NN)

The selected k-NN model has been characterized by the parameters listed in Table 6.19.

Table 6.19: k-NN best parameters setup.

Hyperparameter	Value
n_neighbors	20
weights	'uniform'
p	1
algorithm	'ball_tree'

The performances obtained with this model have been the lowest compared to the other algorithms ones.

Table 6.20: Performances obtained with K-NN.

	Accuracy(%)	Precision(%)	Recall(%)	F1-Score(%)
Opening	66.12	62.60	37.70	47.06
Closing	89.63	82.59	89.63	85.96
Bite	75.40	66.98	74.29	70.45
Avg	<b>77.05</b>	<b>70.72</b>	<b>67.21</b>	<b>67.82</b>

### 6.4.3 Artificial Neural Networks (ANN)

The results of Table 6.22 have been obtained with hyperparameters shown in Table 6.21.

Table 6.21: ANN best parameters setup.

Hyperparameter	Value
architecture	2-9-9-3
LR	0.001
epoch	70

Table 6.22: Performances obtained with ANN.

	Accuracy(%)	Precision(%)	Recall(%)	F1-Score(%)
Opening	66.13	58.11	58.17	58.64
Closing	88.77	87.67	85.09	86.36
Bite	80.05	68.21	75.85	71.83
Avg	<b>78.32</b>	<b>71.63</b>	<b>73.03</b>	<b>72.27</b>

The best results are obtained with the closing and biting movements, unlike the opening, which always suffers from challenges caused by the skin condition (beard) and double chin.

### 6.4.4 Support Vector Machine (SVM)

The SVM model that allows achieving more satisfactory results is characterized by the parameters listed in the Table 6.23.

Table 6.23: SVM best parameters setup.

Hyperparameter	Value
C	10
kernel	'rbf'
gamma	'scale'

Also, in this case, the best results are obtained with the svm algorithm; in the case of opening, it is possible to have a success rate that is close to 70% (see Table 6.24).

Table 6.24: Performances obtained with SVM.

	Accuracy(%)	Precision(%)	Recall(%)	F1-Score(%)
Opening	69.02	61.72	51.82	56.34
Closing	87.25	86.01	87.21	86.61
Bite	79.19	68.28	75.85	71.87
Avg	<b>78.46</b>	<b>71.63</b>	<b>71.60</b>	<b>71.60</b>

## 6.5 Recognition Reinforcement Model

In all previous models, the bite and chewing cycles have been classified as a single class. The bite lasts a few seconds and it has been challenging to classify with little data available. Consequently, an attempt has been made to classify the three types of bite (apple, cracker and nougat), training the classifiers on a single subject, which has been asked to undergo a further acquisition phase. The subject has repeated the experimental protocol by performing only the bite movement, without the other gestures. From the acquired signals, only the part relating to the bite has been extracted without considering chewing. The data set has been strengthened by adding more bite information and classifying the set of reduced gestures (discussed in Section 6.3), including bite classes and chewing movement.

### 6.5.1 Random Forest (RF)

The results for the RF algorithm are presented in Table 6.26 and they have been obtained with the setting, described in Table 6.25.

Table 6.25: RF hyperparameters.

Hyperparameter	Value
n_estimators	200
max_features	'log2'
min_samples_leaf	2
min_samples_split	8
max_depth	10

Table 6.26: RF prediction results.

	Accuracy(%)	Precision(%)	Recall(%)	F1-Score(%)
Opening	86.06	72.14	85.71	78.34
Closing	89.04	86.09	88.67	87.36
Chewing	85.83	64.63	62.85	63.73
Wink	83.39	84.64	87.32	85.96
Smile	90.14	87.41	88.34	87.87
Scowl	87.03	94.10	87.03	90.43
Apple	70.64	74.67	53.33	62.22
Cracker	70.77	71.83	71.83	71.83
Nougat	70.59	73.53	46.73	57.14
<b>Avg</b>	<b>81.50</b>	<b>78.78</b>	<b>74.65</b>	<b>76.10</b>

The bite is the part of the signal that differed significantly according to the food eaten. Compared to the other classes, those relating to different bites have lower accuracy. This aspect depends on the different durations and intensities of the bites. The recognition of chewing alone (85.83 %) is instead increased in terms of accuracy.

### 6.5.2 k-Nearest Neighbour (k-NN)

The results of Table 6.28 have been obtained with hyperparameters shown in Table 6.27.

Table 6.27: k-NN hyperparameters.

Hyperparameter	Value
n_neighbors	18
weights	'uniform'
p	2
algorithm	'brute'

Table 6.28: k-NN prediction results.

	Accuracy(%)	Precision(%)	Recall(%)	F1-Score(%)
Opening	91.99	65.56	91.99	74.47
Closing	86.70	86.49	86.33	86.41
Chewing	77.56	64.26	62.99	63.62
Wink	84.81	85.02	80.21	82.55
Smile	90.14	81.67	89.44	85.38
Scowl	88.05	90.32	86.01	88.11
Apple	65.14	89.36	38.53	53.85
Cracker	69.23	63.16	55.38	59.02
Nougat	65.55	78.13	42.02	54.64
Avg	<b>79.91</b>	<b>77.89</b>	<b>70.32</b>	<b>72</b>

The results are very similar to ones of the previous model except for the opening movement, which achieves a success rate greater than 90%. This improvement comes from having trained and randomly tested the model on a single person. Furthermore, being a woman, there were no problems with electrodes placement, obtaining a good signal recording.

### 6.5.3 Artificial Neural Networks (ANN)

The best ANN model, whose performance can be analyzed in the Table 6.30 is obtained with the setup of parameters listed below (Table 6.29).

Table 6.29: ANN hyperparameters.

Hyperparameter	Value
Architecture	5-9-9-9
LR	0.01
epoch	70

Table 6.30: ANN prediction results.

	<b>Accuracy(%)</b>	<b>Precision(%)</b>	<b>Recall(%)</b>	<b>F1-Score(%)</b>
Opening	87.46	69.34	87.46	77.35
Closing	89.98	85.93	89.79	87.82
Chewing	79.13	62.84	53.94	58.05
Wink	83.75	88.28	74.56	80.84
Smile	91.55	80.50	91.55	85.67
Scowl	86.69	91.37	86.69	88.97
Apple	64.22	56.14	58.72	57.40
Cracker	69.23	70.91	60.00	65.00
Nougat	71.43	78.95	37.82	51.14
<b>Avg</b>	<b>80.38</b>	<b>73.03</b>	<b>71.17</b>	<b>72.47</b>

It can be observed that among the three bites, the apple has the lowest accuracy, probably due to a similarity to the other two based on how the bite is performed by the subject. During the further submission to the experimental protocol, the subject had to perform several bites. After a series of trials, fatigue and disgust have arisen, the individual worsened her movement performance.

#### 6.5.4 Support Vector Machine (SVM)

A correct prediction of SVM model is obtained with the hyperparameters described in Table 6.31.

Table 6.31: SVM hyperparameters.

<b>Hyperparameter</b>	<b>Value</b>
C	1
kernel	'rbf'
gamma	'scale'

The results achieved (shown in Table 6.32) are very similar to the ones of the RF model both in the recognition of bites and in other gestures. These two models present the highest accuracy score.

Table 6.32: SVM prediction results.

	<b>Accuracy(%)</b>	<b>Precision(%)</b>	<b>Recall(%)</b>	<b>F1-Score(%)</b>
Opening	87.46	69.53	87.46	77.47
Closing	89.79	85.18	89.89	87.47
Chewing	83.46	63.36	65.35	64.34
Wink	83.04	84.21	79.15	81.60
Smile	88.38	83.11	88.38	85.67
Scowl	86.69	92.36	86.69	89.44
Apple	69.72	83.61	46.79	60.00
Cracker	69.23	67.24	60.00	63.41
Nougat	71.43	84.00	35.29	49.70
<b>Avg</b>	<b>81.02</b>	<b>79.18</b>	<b>71.00</b>	<b>73.23</b>

This analysis suggests the possibility of distinguishing the type of bite based on the kind of food tested. As expected, by training and testing data acquired by a single subject (albeit randomly) the different algorithms developed, encouraging results are achieved both in the analysis of jaw movements and facial expressions. Discrete outcomes, due to the lack of a more significant number of available data, also come out in the bite classification.

## Chapter 7

# Conclusion and Future Perspective

In this thesis work, a feasibility study on the combination of facial sEMG recording and ATC approach was proposed to recognize facial gestures.

The first part of the analysis focused on the movements to be performed, the corresponding musculature involved, and the sensors' correct arrangement to detect the facial sEMG signals.

As first step, jaw movements have been studied for clinical and rehabilitative interest, including coughing and the action of talking up to the inclusion of facial expressions. The ATC technique has been applied to post-processed sEMG signals in order to assess if an event-driven processing is suitable for identifying facial gestures.

Once the feasibility of an ATC-based facial network has been confirmed, it has been necessary to establish an experimental protocol in order to collect data from different individuals, thus obtaining a sufficiently large dataset useful to properly describe a sample of population. Therefore, after the approval of the experimental procedure by University of Study of Torino's bioethics committee, the acquisition campaign has been launched.

The test involved 21 subjects, 16 males and 5 females (with ages between 24 and 39 years old), which had to repeat a sequence of 8 gestures. In particular, movements to be performed by each subject were the following: *mouth opening*, *mouth closing*, *bite*, *cough*, *wink*, *smile*, *scowl*, and *talking*. sEMG signals were collected from the right emiface investigating some masticatory muscles and mimics ones: *anterior temporal*, *masseter*, *digastric*, *zygomaticus major*, and *corrugator supercillii*. After correctly preparing the skin, taking into account its condition or length of beard in

male subjects, sensors were positioned on the face of interest areas for the sEMG acquisition.

Considering the number of performed gestures, and their nested information related to the acquisition channels, the use of Artificial Intelligence (AI) algorithm has been identified as the proper way to achieve facial recognition reaching suitable performance. Offline ML analysis has been performed by implementing four algorithms: Random Forest, k-NN, Artificial Neural Networks, and Support Vector Machine.

In the first phase, all gestures performed during the acquisition protocol have been classified. Both biting and chewing have been included in one class, and no bite diversification was made based on the food tested. All algorithms have reached a similar success rate greater than 60%. Therefore, in order to improve average recognition accuracy by up to 75%, the not well recognized movements, like cough and talking, movements that had a strong influence on models predictive ability, have been eliminated from the available dataset.

With the aim of moving towards a future clinical application, only the jaw movements have been classified, eliminating the recording channels not interested in these gestures. As a result, an overall accuracy of close to 80 % has been achieved.

A further step regarded the possibility to distinguish the type of bite and the reduced gestures. It was impossible to classify the different bites using the entire dataset because the amount of data associated to this movement is quite limited (not being one of the main tasks during acquisition protocol). However, a preliminary version based and strengthened on a single subject has been proposed, achieving encouraging results with an accuracy score greater than 80%.

Through the experience gained with this thesis, and having examined the current challenges, the future developments are now introduced.

A first step will be to efficiently improve the ML algorithms predictive capacity through a new data acquisition campaign in order to obtain a more robust dataset and to achieve a more balanced data collection (in terms of individuals and gestures). As for the desire to distinguish bites based on their different intensity and characteristics, a proper guideline for next tests will be to record this movement as standalone.

Indeed, changes will be made to the defined experimental protocol, trying to optimize the acquisition features for the most critical recognized gestures, and increasing the resting time from one movement to another to avoid muscle fatigue

and make movements repetition more comfortable. Another important point, in order to ensure the acquisition of good quality signals, is to establish more specific criteria for inclusion in the experiment, such as requiring male subjects to shave their faces if necessary.

Finally, after optimizing the ML algorithms, an embedded implementation can be realized to design a wearable system capable of acquiring the sEMG signals from face muscles and classifying related gestures in real-time application.

In conclusion, advancing towards speech recognition by identifying keywords and establishing new muscles could open the door to new applications.

# Bibliography

- [1] E.N. Marieb, P.B. Wilhelm, and J.B. Mallatt. “Human Anatomy, Global Edition”. In: Pearson Education Limited, 2016. Chap. 10-11. ISBN: 9781292156873.
- [2] URL: <https://medlineplus.gov/ency/imagepages/19841.htm>.
- [3] D.A. Neumann. “Kinesiology of the Musculoskeletal System: Foundations for Physical Rehabilitation”. In: Mosby, 2002. Chap. 3, pp. 41–44. ISBN: 9780815163497.
- [4] G.J. Tortora and B.H. Derrickson. “Principles of Anatomy and Physiology, 14th Edition: 14th Edition”. In: Wiley, 2013. Chap. 10. ISBN: 9781118808856.
- [5] I. Peate and M. Nair. “Fundamentals of Anatomy and Physiology: For Nursing and Healthcare Students”. In: Fundamentals. Wiley, 2016. Chap. 6. ISBN: 9781119055525.
- [6] URL: <https://www.moleculardevices.com/applications/patch-clamp-electrophysiology/what-action-potential#gref>.
- [7] G. Robertson et al. *Research Methods in Biomechanics, 2E*. Human Kinetics, 2013. ISBN: 9780736093408.
- [8] R. Merletti and D. Farina. “Analysis of intramuscular electromyogram signals”. In: *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 367 (2008), pp. 357–368.
- [9] Carlo J De Luca et al. “Decomposition of surface EMG signals”. In: *Journal of neurophysiology* 96.3 (2006), pp. 1646–1657.
- [10] URL: <https://www.shimmersensing.com/products/emg-ecg-electrodes>.
- [11] Muhammad Jamal. “Signal Acquisition Using Surface EMG and Circuit Design Considerations for Robotic Prosthesis”. In: Oct. 2012. ISBN: 978-953-51-0805-4. DOI: 10.5772/52556.
- [12] Christopher Spiewak et al. “A Comprehensive Study on EMG Feature Extraction and Classifiers”. In: *Open Access Journal of Biomedical Engineering and its Applications* 1 (Feb. 2018). DOI: 10.32474/OAJBEB.2018.01.000104.

- [13] Angkoon Phinyomark et al. “The Usefulness of Mean and Median Frequencies in Electromyography Analysis”. In: Oct. 2012, pp. 195–220. ISBN: 978-953-51-0805-4. DOI: 10.5772/50639.
- [14] Cheng-Ning Huang, Chun-Han Chen, and Hung-Yuan Chung. “The review of applications and measurements in facial electromyography”. In: *Journal of Medical and Biological Engineering* 25 (Mar. 2005), pp. 15–20.
- [15] Fabio Rossi et al. “Wireless Low Energy System Architecture for Event-Driven Surface Electromyography”. In: May 2019, pp. 179–185. ISBN: 978-3-030-11972-0. DOI: 10.1007/978-3-030-11973-7\_21.
- [16] M. Crepaldi et al. “A quasi-digital radio system for muscle force transmission based on event-driven IR-UWB”. In: *2012 IEEE Biomedical Circuits and Systems Conference (BioCAS)*. 2012, pp. 116–119. DOI: 10.1109/BioCAS.2012.6418406.
- [17] S. Sapienza et al. “On Integration and Validation of a Very Low Complexity ATC UWB System for Muscle Force Transmission”. In: *IEEE Transactions on Biomedical Circuits and Systems* 10.2 (2016), pp. 497–506. DOI: 10.1109/TBCAS.2015.2416918.
- [18] A. Shahshahani et al. “An all-digital spike-based ultra-low-power IR-UWB dynamic average threshold crossing scheme for muscle force wireless transmission”. In: *2015 Design, Automation Test in Europe Conference Exhibition (DATE)*. 2015, pp. 1479–1484.
- [19] Anil K Jain, Jianchang Mao, and K Moidin Mohiuddin. “Artificial neural networks: A tutorial”. In: *Computer* 29.3 (1996), pp. 31–44.
- [20] V. Sze et al. “Efficient Processing of Deep Neural Networks: A Tutorial and Survey”. In: *Proceedings of the IEEE* 105.12 (2017), pp. 2295–2329. DOI: 10.1109/JPROC.2017.2761740.
- [21] Facundo Bre, Juan Gimenez, and Víctor Fachinotti. “Prediction of wind pressure coefficients on building surfaces using Artificial Neural Networks”. In: *Energy and Buildings* 158 (Nov. 2017). DOI: 10.1016/j.enbuild.2017.11.045.
- [22] B. Baharudin, Lam Hong Lee, and K. Khan. “A Review of Machine Learning Algorithms for Text-Documents Classification”. In: *Journal of Advances in Information Technology* 1 (2010), pp. 4–20.
- [23] Sotiris Kotsiantis. “Supervised Machine Learning: A Review of Classification Techniques”. In: *Informatika (Ljubljana)* 31 (Oct. 2007).
- [24] Konrad Rieck et al. “Support Vector Machines”. In: Jan. 2012, pp. 883–926. DOI: 10.1007/978-3-642-21551-3\_30.
- [25] Guodong Guo, Stan Li, and Kapluk Chan. “Face Recognition by Support Vector Machines”. In: *Proceedings of the Fourth IEEE International Conference on Automatic Face and Gesture Recognition* (Feb. 1970).

- [26] Thorsten Joachims. “Text Categorization with Support Vector Machines”. In: *Proc. European Conf. Machine Learning (ECML’98)* (Jan. 1998). DOI: 10.17877/DE290R-5097.
- [27] Richard Brereton and Gavin Lloyd. “Support Vector Machines for classification and regression”. In: *The Analyst* 135 (Feb. 2010), pp. 230–267. DOI: 10.1039/b918972f.
- [28] S. Arora et al. “Performance Comparison of SVM and ANN for Handwritten Devnagari Character Recognition”. In: *ArXiv* abs/1006.5902 (2010).
- [29] URL: <https://www.r-bloggers.com/2019/10/support-vector-machines-with-the-mlr-package/>.
- [30] Shirin Tavara. “Parallel Computing of Support Vector Machines: A Survey”. In: *ACM Computing Surveys* 51 (Jan. 2019), pp. 1–38. DOI: 10.1145/3280989.
- [31] Haneen Abu Alfeilat et al. “Effects of Distance Measure Choice on K-Nearest Neighbor Classifier Performance: A Review”. In: *Big Data* 7 (Aug. 2019). DOI: 10.1089/big.2018.0175.
- [32] S. Raschka and V. Mirjalili. *Python Machine Learning: Machine Learning and Deep Learning with Python, scikit-learn, and TensorFlow 2, 3rd Edition*. Packt Publishing, 2019, pp. 101–104. ISBN: 9781789958294.
- [33] Li-Yu Hu et al. “The distance function effect on k-nearest neighbor classification for medical datasets”. In: *SpringerPlus* 5 (Dec. 2016). DOI: 10.1186/s40064-016-2941-7.
- [34] URL: <https://www.javatpoint.com/machine-learning-decision-tree-classification-algorithm>.
- [35] Yan-Yan Song and Ying Lu. “Decision tree methods: applications for classification and prediction”. In: *Shanghai archives of psychiatry* 27 (Apr. 2015), pp. 130–5. DOI: 10.11919/j.issn.1002-0829.215044.
- [36] Hamid Alinejad-Rokny and S. Parvin. “2.16. Divide and Conquer Classification”. In: *International Journal of Sciences: Basic and Applied Research (IJSBAR)* 5 (Jan. 2011), pp. 2446–2452.
- [37] Mei-Hung Chiu et al. “The use of facial micro-expression state and tree-forest model for predicting conceptual-conflict based conceptual change”. In: Jan. 2016.
- [38] URL: <https://helloml.org/easy-introduction-to-random-forest-algorithm/>.
- [39] Jehad Ali et al. “Random Forests and Decision Trees”. In: *International Journal of Computer Science Issues(IJCSI)* 9 (Sept. 2012).

- [40] P. M. Ros et al. “A wireless address-event representation system for ATC-based multi-channel force wireless transmission”. In: *5th IEEE International Workshop on Advances in Sensors and Interfaces IWASI*. 2013, pp. 51–56. DOI: 10.1109/IWASI.2013.6576061.
- [41] S. Sapienza et al. “On Integration and Validation of a Very Low Complexity ATC UWB System for Muscle Force Transmission”. In: *IEEE Transactions on Biomedical Circuits and Systems* 10.2 (2016), pp. 497–506. DOI: 10.1109/TBCAS.2015.2416918.
- [42] D. A. F. Guzman et al. “Very low power event-based surface EMG acquisition system with off-the-shelf components”. In: *2017 IEEE Biomedical Circuits and Systems Conference (BioCAS)*. 2017, pp. 1–4. DOI: 10.1109/BIOCAS.2017.8325152.
- [43] S. Sapienza et al. “On-Line Event-Driven Hand Gesture Recognition Based on Surface Electromyographic Signals”. In: *2018 IEEE International Symposium on Circuits and Systems (ISCAS)*. 2018, pp. 1–5. DOI: 10.1109/ISCAS.2018.8351065.
- [44] A. Mongardi et al. “A Low-Power Embedded System for Real-Time sEMG based Event-Driven Gesture Recognition”. In: *2019 26th IEEE International Conference on Electronics, Circuits and Systems (ICECS)*. 2019, pp. 65–68. DOI: 10.1109/ICECS46596.2019.8964944.
- [45] F. Rossi et al. “An Event-Driven Closed-Loop System for Real-Time FES Control”. In: *2019 26th IEEE International Conference on Electronics, Circuits and Systems (ICECS)*. 2019, pp. 867–870. DOI: 10.1109/ICECS46596.2019.8965153.
- [46] Jun-Wen Tan et al. “Facial electromyography (fEMG) activities in response to affective visual stimulation”. In: (Apr. 2011). DOI: 10.1109/WACI.2011.5953144.
- [47] S. Hu et al. “Facial EMG as an indicator of palatability in humans”. In: *Physiology Behavior* 68 (1999), pp. 31–35.
- [48] Geoffrey Meltzner et al. “Speech recognition for vocalized and subvocal modes of production using surface EMG signals from the neck and face”. In: Jan. 2008, pp. 2667–2670.
- [49] Chun-Han Chen Chen-Ning Huang and Hung-Yuan Chung. “Speech pattern recognition with facial electromyography”. In: *Biomedical Engineering Society Annual Symposium*. 2003.
- [50] E. Lopez-Larraz et al. “Syllable-based speech recognition using EMG”. In: *2010 Annual International Conference of the IEEE Engineering in Medicine and Biology*. 2010, pp. 4699–4702. DOI: 10.1109/IEMBS.2010.5626426.

- [51] Miriam Matsui et al. “Differences between the activity of the masticatory muscles of adults with cerebral palsy and healthy individuals while at rest and in function”. In: 73 (Sept. 2016).
- [52] Chuyao Jian et al. “Multiparameter Electromyography Analysis of the Masticatory Muscle Activities in Patients with Brainstem Stroke at Different Head Positions”. In: *Frontiers in Neurology* 8 (May 2017). DOI: 10.3389/fneur.2017.00221.
- [53] H. Gunes, M. Piccardi, and T. Jan. “Face and Body Gesture Recognition for a Vision-Based Multimodal Analyzer”. In: *VIP*. 2003.
- [54] Francesca Nonis et al. “3D Approaches and Challenges in Facial Expression Recognition Algorithms—A Literature Review”. In: *Applied Sciences* 9 (Sept. 2019), p. 3904. DOI: 10.3390/app9183904.
- [55] Olufisayo Ekundayo and Serestina Viriri. “Facial Expression Recognition: A Review of Methods, Performances and Limitations”. In: Mar. 2019, pp. 1–6. DOI: 10.1109/ICTAS.2019.8703619.
- [56] Meriem Sari, Abdelouahab Moussaoui, and Abdenour Hadid. “Automated Facial Expression Recognition Using Deep Learning Techniques: An Overview”. In: *International Journal of Informatics and Applied Mathematics* 3 (2020), pp. 39–53.
- [57] Yunxin Huang et al. “Facial Expression Recognition: A Survey”. In: *Symmetry* 11 (Sept. 2019), p. 1189. DOI: 10.3390/sym11101189.
- [58] Najmeh Samadiani et al. “A review on automatic facial expression recognition systems assisted by multimodal sensor data”. In: *Sensors* 19.8 (2019), p. 1863.
- [59] I. Hupont et al. “The Emotracker: Visualizing Contents, Gaze and Emotions at a Glance”. In: *2013 Humaine Association Conference on Affective Computing and Intelligent Interaction*. 2013, pp. 751–756. DOI: 10.1109/ACII.2013.139.
- [60] T. Fang et al. “3D facial expression recognition: A perspective on promises and challenges”. In: *2011 IEEE International Conference on Automatic Face Gesture Recognition (FG)*. 2011, pp. 603–610. DOI: 10.1109/FG.2011.5771466.
- [61] M. Jiang et al. “Facial Expression Recognition with sEMG Method”. In: *2015 IEEE International Conference on Computer and Information Technology; Ubiquitous Computing and Communications; Dependable, Autonomic and Secure Computing; Pervasive Intelligence and Computing*. 2015, pp. 981–988. DOI: 10.1109/CIT/IUCC/DASC/PICOM.2015.148.
- [62] Iman Mohammad-Rezazadeh et al. “Using affective human–machine interface to increase the operation performance in virtual construction crane training system: A novel approach”. In: *Automation in Construction* 20 (May 2011), pp. 289–298. DOI: 10.1016/j.autcon.2010.10.005.

- [63] S. M. Firoozabadi, M. R. A. Oskoei, and H. Hu. “A Human – Computer Interface based on Forehead Multi-Channel Bio-signals to Control a Virtual Wheelchair”. In: 2002.
- [64] Guillaume Gibert et al. “Enhancement of Human Computer Interaction with Facial Electromyographic Sensors”. In: *Proceedings of the 21st Annual Conference of the Australian Computer-Human Interaction Special Interest Group: Design: Open 24/7*. OZCHI '09. Melbourne, Australia: Association for Computing Machinery, 2009, pp. 421–424. ISBN: 9781605588544. DOI: 10.1145/1738826.1738914. URL: <https://doi.org/10.1145/1738826.1738914>.
- [65] L. B. P. Ang et al. “Facial expression recognition through pattern analysis of facial muscle movements utilizing electromyogram sensors”. In: *2004 IEEE Region 10 Conference TENCON 2004*. Vol. C. 2004, 600–603 Vol. 3. DOI: 10.1109/TENCON.2004.1414843.
- [66] Mahyar Hamedi et al. “Surface Electromyography-Based Facial Expression Recognition in Bi-Polar Configuration”. In: *Journal of Computer Science 7* (2011), pp. 1407–1415.
- [67] Y. Chen, Zhongliang Yang, and J. Wang. “Eyebrow emotional expression recognition using surface EMG signals”. In: *Neurocomputing* 168 (2015), pp. 871–879.
- [68] Y. Cai et al. “Machine-learning approaches for recognizing muscle activities involved in facial expressions captured by multi-channels surface electromyogram”. In: *Smart Health* (2017), pp. 15–25.
- [69] S. Orguc et al. “EMG-based Real Time Facial Gesture Recognition for Stress Monitoring”. In: *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)* (2018), pp. 2651–2654.
- [70] URL: <https://www.gtec.at/product/ghiamp/>.
- [71] Bio-medical. *Covidien products*. <https://bio-medical.com/covidien-kendall-disposable-surface-emg-ecg-ekg-electrodes-1-24mm-50pkg.html>.
- [72] Hoo Lee et al. “Changes in Hyolaryngeal Movement and Swallowing Function After Neuromuscular Electrical Stimulation in Patients With Dysphagia”. In: *Annals of rehabilitation medicine* 39 (Apr. 2015), pp. 199–209. DOI: 10.5535/arm.2015.39.2.199.
- [73] A. Fridlund and J. Cacioppo. “Guidelines for human electromyographic research.” In: *Psychophysiology* 23 5 (1986), pp. 567–89.
- [74] Akiko Shimada et al. “Measurement of dynamic bite force during mastication”. In: *Journal of oral rehabilitation* 39 (Jan. 2012), pp. 349–56. DOI: 10.1111/j.1365-2842.2011.02278.x.

- [75] E. Kemsley and M. Defernez. “Relating Surface Electromyograms of the Facial Muscles During Mastication to the Mechanical and Sensory Properties of Foodstuffs”. In: 2012.
- [76] Anton Boxtel. “Facial EMG as a tool for inferring affective states”. In: Aug. 2010.
- [77] T. Bastos-Filho et al. “Multi-modal interface for communication operated by eye blinks, eye movements, head movements, blowing/sucking and brain waves”. In: *2013 ISSNIP Biosignals and Biorobotics Conference: Biosignals and Robotics for Better and Safer Living (BRC)*. 2013, pp. 1–6. DOI: 10.1109/BRC.2013.6487458.
- [78] Piotr Kowalczyk and Dariusz Sawicki. “Blink and wink detection as a control tool in multimodal interaction”. In: *Multimedia Tools and Applications* 78 (May 2019). DOI: 10.1007/s11042-018-6554-8.
- [79] Nicholas Smith et al. “Detection of Simulated Vocal Dysfunctions Using Complex sEMG Patterns”. In: *IEEE journal of biomedical and health informatics* 20 (Oct. 2015). DOI: 10.1109/JBHI.2015.2490087.
- [80] Università degli studi di Torino. *Comitato Bioetico dell’Ateneo*. <https://www.unito.it/ricerca/strutture-e-organi-la-ricerca/comitato-di-bioetica-dellateneo>.
- [81] F. Pedregosa et al. “Scikit-learn: Machine Learning in Python”. In: *Journal of Machine Learning Research* 12 (2011), pp. 2825–2830.
- [82] Antonio Gulli and Sujit Pal. *Deep learning with Keras*. Packt Publishing Ltd, 2017.
- [83] Martin Abadi et al. “Tensorflow: A system for large-scale machine learning”. In: *12th {USENIX} Symposium on Operating Systems Design and Implementation ({OSDI} 16)*. 2016, pp. 265–283.