



MASTER THESIS – DEPARTMENT OF BIOMEDICAL ENGINEERING

Multichannel surface EMG and machine learning for classification of facial expressions

Veerle Diederiks

Biomedical Signals and Systems (BSS)

GRADUATION COMMITTEE

Prof.dr. R.J.A. Van Wezel (Richard) – Chairman

Dr. S.U. Yazuz (Utku) – Daily supervisor

Dr.ir. K. Nizamis (Kostas) – External member

23 June 2021

UNIVERSITY OF TWENTE.

Preface

Dear reader,

I proudly present the results of the thesis that I have been working on for the past year. By carrying out this research, I am finalizing six years of studying at the University of Twente. During this period I have definitely learned a lot and have explored all that Enschede could bring me.

After finishing my Bachelors, I struggled with choosing a specialization for my Masters. Following a lot of contemplation I had chosen the track of Medical Device Design. However, after some time I realized that it was not what I hoped it would be. That is one of the reasons that I'm extremely grateful that my Graduation Committee, consisting of Utku Yavuz, Richard van Wezel and Kostas Nizamis, allowed me to perform my thesis in the field of Signal Analysis. A research area that I have found to be much more interesting. Thanks to my committee for guiding me through the process and enabling me to follow my interests, with a special thanks to Utku for being my daily supervisor and meeting up weekly.

I hope you enjoy reading my report,

Veerle

Abstract

Facial expressions are an important aspect of non-verbal communication, showing reactions and attention. In patients with Disorders of Consciousness (DOC) facial expressions are commonly less pronounced. Diagnosis of these patients is partly based on their response to external stimuli, measured by their facial expression. As these can be difficult to objectively measure, a high misdiagnosis rate exists. Development of a method to detect and identify expressions could support diagnosis and possibly improve communication between patients and caregivers or loved-ones. The main goal of this research is to evaluate to what extent facial surface electromyography (sEMG) signals can be used to classify four facial expressions (happiness, anger, sadness and fear) in healthy subjects. In addition, micro expressions are evoked and measured to mimic the diminished expressions of DOC patients. Lastly the predictive value of various channels is evaluated to determine the most efficient experimental set-up. An experimental protocol using a 32-channel unipolar micro-electrode set-up was designed to obtain EMG signals. Twenty-nine models were included in in this study. It was found that the model Subspace K-Nearest Neighbor (KNN) and feature Difference Absolute Mean Value (DAMV) performed best at classifying expressions of subjects it had not been trained on (with a test accuracy of 55.7%). Happiness was most often identified correctly. Additionally, this research has demonstrated that micro expressions, evoked by exposure to images of facial expressions, occur and can be measured with sEMG. The model Subspace KNN and feature Waveform Length (WL) succeeds in predicting these micro expressions for one of the subjects with a test accuracy of 47.1%. Evaluation of the predictive value of the 32 channels shows that a comparable test accuracy (53.4%) is obtained for a subset of only 15 channels with model Subspace KNN and feature WL. To develop a method applicable in clinical practice further research is needed, this research provides a good starting point.

Samenvatting

Gezichtsuitdrukkingen zijn een belangrijk aspect van non-verbale communicatie, en worden gebruikt voor het tonen van reacties en aandacht. Bij patiënten met een bewustzijnsstoornis zijn de gezichtsuitdrukkingen vaak minder uitgesproken. De diagnose van deze patiënten is onder andere gebaseerd op de reactie op prikkels van buitenaf, en wordt gemeten aan de hand van de gezichtsuitdrukking. Omdat dit moeilijk te meten is in deze patiëntengroep, is er een hoog percentage dat een verkeerde diagnose krijgt. De ontwikkeling van een methode om gezichtsuitdrukking te detecteren en identificeren zou artsen kunnen ondersteunen bij het stellen van een diagnose, en kan daarnaast mogelijk de communicatie tussen patiënten en zorgverleners of naasten verbeteren. Het belangrijkste doel van dit onderzoek is om te evalueren in hoeverre signalen van spieractiviteit (sEMG) gebruikt kunnen worden om vier gezichtsuitdrukkingen (blijdschap, woede, verdriet en angst) bij gezonde proefpersonen te classificeren. Daarnaast worden micro-expressies uitgelokt en gemeten om de verminderde gezichtsuitdrukking van patiënten met een bewustzijnsstoornis na te bootsen. Ten slotte wordt de voorspellende waarde van verschillende EMG-kanalen geëvalueerd om de meest efficiënte meetprocedure te bepalen. In dit onderzoek is een experimenteel protocol ontworpen dat met behulp van 32 micro-electroden de EMG signalen van verschillende spieren in het gezicht meet. Negenentwintig modellen werden geëvalueerd in deze studie. Hieruit bleek dat het model *Subspace K-Nearest Neighbor (KNN)* en feature *Difference Absolute Mean Value (DAMV)* het beste in staat was in het classificeren van uitdrukkingen van proefpersonen waarop het model niet was getraind (met een testnauwkeurigheid van 55,7%). Blijdschap was de uitdrukking die het vaakst correct geclassificeerd werd. Daarnaast heeft dit onderzoek aangetoond dat micro-expressies aanwezig zijn en gemeten kunnen worden door middel van sEMG. Het model *Subspace KNN* en feature *Waveform Length (WL)* slaagde erin om deze micro-expressies voor één van de proefpersonen te voorspellen met een testnauwkeurigheid van 47,1%. Evaluatie van de voorspellende waarde van de 32 kanalen laat zien dat een vergelijkbare testnauwkeurigheid (53,4%) wordt verkregen voor een subset van slechts 15 kanalen met model *Subspace KNN* en feature *WL*. Voor de ontwikkeling van een methode die daadwerkelijk in een klinische setting gebruikt kan worden is meer onderzoek nodig. Deze studie biedt een goed uitgangspunt.

Contents

Preface	2
Abstract	3
Samenvatting.....	4
1. Background.....	7
1.1. Introduction	7
1.1.1. Facial Expression Recognition	7
1.1.2. Clinical relevance	7
1.1.3. Motivation and objectives.....	8
1.2. EMG and Facial Expressions	9
1.2.1. Anatomy of the face	9
1.2.2. Facial expression of emotions	11
1.2.3. Electromyography (EMG)	11
1.2.4. State of the art	13
1.3. Machine Learning	14
1.3.1. Classification algorithms.....	15
2. Methodology	17
2.1. Experiments	17
2.1.1. Subjects	17
2.1.2. Expressions	17
2.1.3. Electrode configuration	18
2.1.4. Experimental procedure.....	19
2.2. Data analysis	22
2.2.1. EMG pre-processing	22
2.2.2. Feature extraction	23
2.2.3. Classification	24
2.2.4. Micro expressions.....	25
2.2.5. Channel selection	25
3. Results.....	26
3.1. Model performance.....	26
3.1.1. Validation accuracy	26
3.1.2. Test accuracy	27
3.2. Micro expressions.....	28
3.2.1. Manual identification	28
3.2.2. Model performance.....	30
3.3. Channel subsets.....	31

4. Discussion.....	34
4.1. Experimental procedure	34
4.2. Window lengths	34
4.3. Features	35
4.4. Classifier performance	35
4.5. Micro expressions.....	37
4.6. Channel subsets.....	37
5. Conclusion	38
Bibliography.....	39
Appendices	44
Appendix A: Protocol	44
Appendix B: Information Brochure and Informed Consent	53
Appendix C: FACES database	56
Appendix D: Results for all classifiers, features and window lengths	57
Appendix E: Optimization	60
Appendix F: Channel subsets.....	60
Appendix G: Speed measures.....	60

1. Background

In this chapter the motivation and objectives of this research are defined. Additionally, relevant concepts will be described, including facial anatomy, different types of expressions, electromyography and classification models.

1.1. Introduction

1.1.1. Facial Expression Recognition

Analysis of facial expression has been of great interest in several fields for quite some time. It has applications in marketing, surveillance, entertainment and healthcare, amongst others [1]. Obvious facial expressions that we all come across from time to time are disgust, fear, joy, surprise, sadness and anger (see Figure 1). Facial expressions are very important in non-verbal communication, showing reactions and attention. Although there are differences in communication between countries and cultures, these six facial expressions are universal [2].

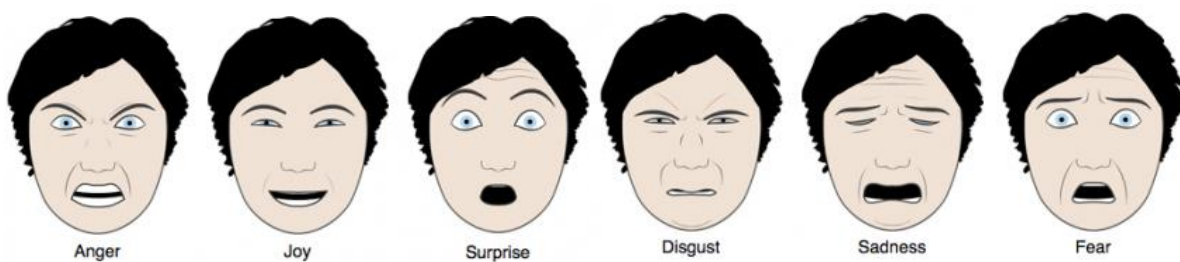


Figure 1: Facial characteristics of the six basic emotions: anger, joy, surprise, disgust, sadness and fear [3].

The faces in Figure 1 show exaggerated expressions. In real-life, much more often only subtle changes in expression take place as an expression of emotion. These subtle changes are more difficult to detect and distinguish. Some expressions are so subtle that they cannot be detected with the naked eye. Different techniques have been evaluated that could aid in detection of these expressions. Two methods with good prospects are video based facial expression detection and the use of EMG-signals.

The first method is used to automatically detect expression in videos of faces, as its name might suggest. Pitfalls are that spontaneous expressions are only recognized to a limited extent, real-time recognition remains difficult and rotated faces or faces that are off-center make detection harder [4]. Some expressions remain too subtle to be recognized by computer vision systems. In addition, the systems are not practical for wearable applications, as you would have to point a camera on the face at all times. EMG based facial expression detection can provide information about subtle changes. It is a non-invasive method to measure muscle activity [5]. As it is able to measure minimal changes in muscle activity, it can detect micro expressions [6]. Besides, wireless electrodes exist, making this set-up more suitable for wearable measurements. For both methods, video based and EMG based, classification learners (a type of machine learning) can be used to classify the facial expression based on the obtained signals.

1.1.2. Clinical relevance

People that would benefit from detection of their facial expressions are patients in whom these are less pronounced. In particular, patients with Disorders of Consciousness (DOC). This is an umbrella term for patients that awaken from a coma, but remain unaware. Their awareness might improve over time, and can depend on their response to external stimuli. One method to measure this response is by the patient's facial expression [7], [8]. DOC patients are known for their limited expressions. Some of them are so subtle that they are not visible to the naked eye [9]–[11]. Development of a technique that can detect these subtle expressions might aid in diagnosing DOC patients in the future.

In addition, this technique could assist other patients with diminished muscle activity as well. For example people who suffer from neurological disorders, like Parkinson's Disease. Their facial expressions are usually smaller, in some cases even absent, and take more time to take place [12]. Reduced facial expressions similarly occurs in other pathological cases. Damage in the facial nerve can result in weakness or inability to move the muscles, making it difficult to show emotions. Development of a method to detect and identify expressions could support diagnosis and possibly improve communication between patients and caregivers or loved-ones.

1.1.3. Motivation and objectives

This research aims at developing a method and experimental protocol for classification of facial expressions. Since EMG has a high potential to detect subtle expressions, this is the method selected in this study. Various classification models, features and window lengths will be explored to find an optimal combination to accurately predict facial expressions. The predictive power of a selection of channels will be evaluated to create a subset of required channels with a high information-density. Channels that show little predictive value can be disregarded in future research to simplify the protocol and make it more suitable in practice. In addition, this research will attempt to measure micro expressions to explore the ability of EMG to measure subtle expressions. If succeeded, these expression will be used for further examination of the classification accuracy of the models.

In summary, the main objectives are:

- Design of an experimental protocol to obtain facial EMG measurements;
- Development of one or multiple machine learning models to classify expressions based on the EMG data;
- Evaluation of the predictive power of all channels to create a subset of channels for simplification of data acquisition;
- Measurement of micro-expressions, to possibly further investigate the classification accuracy of the models.

Research question and hypothesis

To achieve the goal(s) the main research question to be answered is: To what extent can facial EMG signals be used to classify the facial expression of healthy subjects?

It is expected that EMG can be used to classify several facial expressions, as previous studies have proven (see section 1.2.4 *State of the art*). The expression of happiness will be most easily distinguished from the other expressions, due to characteristic activity in zygomatic major. Furthermore it is likely that electrodes located on muscles used in expressions explored in this research will yield the largest predictive value. Making predictions regarding micro expressions is difficult, as not every person will show them. Hopefully at least one of the subjects that participates will show them so they can be measured.

1.2. EMG and Facial Expressions

1.2.1. Anatomy of the face

The face has a complex anatomy, containing 42 muscles. These muscles can be divided into two groups: the muscles that control facial expressions (mimetic muscles) and the muscles that control movement of the jaw for chewing and grinding (muscles of mastication). For this research only the mimetic muscles will be discussed.

Mimetic muscles

The mimetic muscles originate from the skeleton and insert into skin. Contraction of these muscles creates folds in the skin, forming the basis of facial mimicry. In younger people this folding is reversible due to elasticity of the skin. However, in older people some folds may exist continuously. [13]

The mimetic muscles overlap each other, some muscles lie deep and other superficial. A sub-division of the facial muscles into four groups can be made: muscles in the area of the skull, eyes, nose and mouth [13]. The functions of relevant mimetic muscles can be seen in Table 1, location of the muscles is shown in Figure 2 on the next page.

Table 1: Relevant muscles and their function, divided into muscles of the skull, eyes, nose and mouth. Adapted from [13][14].

	Muscle	Function
Skull	Frontalis	Raising the eyebrows, creating wrinkles in the forehead
Eyes	Orbicularis oculi	Closing the eye
	Corrugator supercilii	Depressing the eyebrows, creating frown lines
Nose	Procerus	Pulling the forehead downwards, creating horizontal folds at the bridge of the nose
	Nasalis	Dilatation of nostrils
	Levator labii superioris (aleque nasi)	Pulling skin of the nasal openings and upper lip upwards, dilating the nostrils
Mouth	Orbicularis oris	Closing the lips
	Buccinator	Pressing cheeks against teeth, and pulling the corners of the mouth outwards
	Zygomaticus major	Pulling the corners of the mouth laterally upwards
	Risorius	Pulling the corners of the mouth laterally
	Levator anguli oris	Pulling the corners of the mouth upwards
	Depressor anguli oris	Pulling the corners of the mouth downwards
	Depressor labii inferioris	Lowering the bottom lip
	Mentalis	Creating a fold between the chin and lips
	Platysma	Depressing the lower lip, corners of the mouth, and mandible

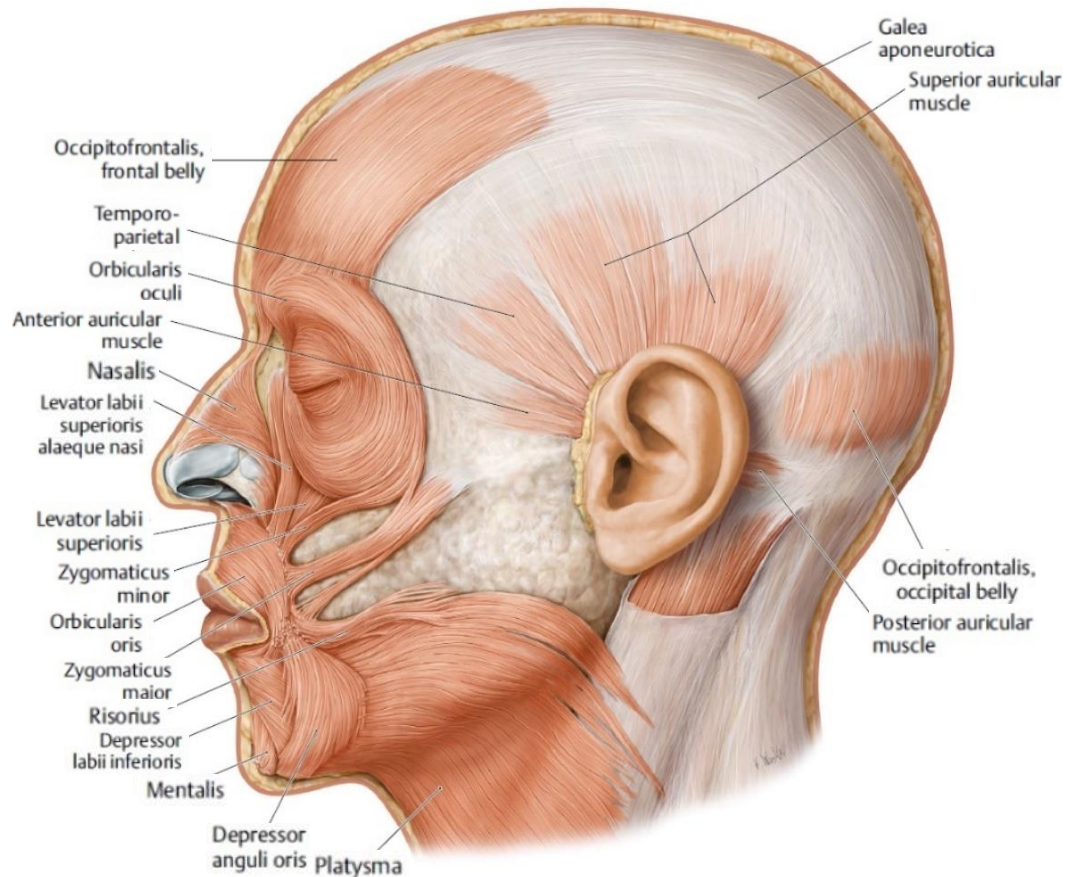
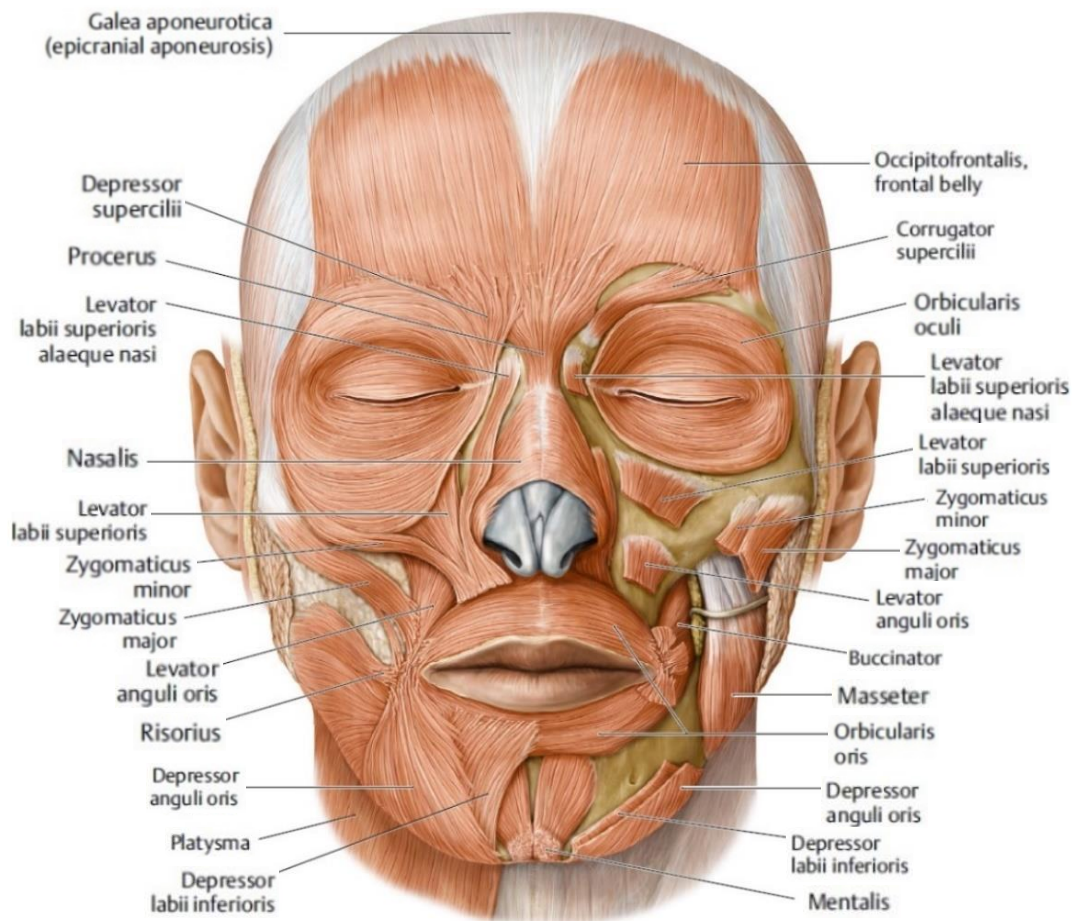


Figure 2: Anatomy of the facial muscles, front view (top) and side view (bottom) [15].

1.2.2. Facial expression of emotions

Emotions are expressed by contraction of certain mimetic muscles described in the previous section. The emotions and muscles can be linked by means of the EMFACS system, short for Emotional Facial Action Coding System [16]. Within this coding system, emotions are linked to Action Units (AU), which describe the muscle(s) needed to perform a certain action [17]. E.g. AU1 describes the inner brow raiser, performed by contraction of the frontalis. In section 2.1.2 *Expressions*, the expressions evaluated in this research with their corresponding AU's are presented.

Macro and micro expressions

Expressions are not always shown with the same intensity. On one hand there are the clear, significant expressions that can be observed with the naked eye. These are the 'normal' facial expressions, which typically last between 1 and 4 seconds [18]. Within this report these will be referred to as macro expressions. On the other hand, micro expressions exist. These are usually unconscious expressions with a duration of about half a second or less [18]. Micro expressions can be defined as a brief facial movement revealing an emotion that a person tries to conceal [19]. The two main factors that distinguish macro and micro expressions are the total duration and onset duration. In a study in 2013 [20] these two factors were found to be respectively 500 ms and 260 ms. However, the onset duration (time after exposure) varies in studies. Another study [21] found an onset duration of 300-400 ms and an even higher duration, of 500 ms [22], was found as well.

Multiple studies [22]–[24] have shown that micro expressions can be evoked by exposure to pictures of facial expressions. This means that when exposed to a happy face, people respond with higher activity in the Zygomaticus major (pulling the corners of the mouth laterally upwards) and when viewing an angry face, people respond with higher activity in the Corrugator supercilii (depressing the eyebrows). This corresponds with the findings of Wingenbach et al. [25], showing that covert facial mimicry (micro expressions) are emotion-specific. However, research has also shown that not all people are in possession of this quality. It is related to empathy [26]. Being highly empathic correlates with a higher reactivity to facial expressions. In addition, it may also vary based on the gender of the subject, females pronouncing a larger response to facial expressions [26]. It is thus uncertain that every person will show micro expressions at every occurrence.

1.2.3. Electromyography (EMG)

When muscles contract, electrical potentials originate from the motor units. These electrical potentials can be measured via electrodes, either directly in the muscle or on the surface of the skin [27]. This latter method, sEMG, is nowadays more widely used as it is non-invasive. The EMG signal is usually represented as μV over time, see Figure 3 below. The signals are typically relatively noisy, due to surrounding electrical circuits, movement artifacts and cross-talk between muscles [27], [28]. Thorough filtering needs to be performed to make the signals useful for further investigation, usually consisting of a bandpass filter to remove low-frequent movement artifacts and high-frequent noise, followed by a notch filter to remove powerline noise (50 or 60 Hz) [28].

EMG signals contain a substantial amount of information. Researchers usually look at a selection of properties of the signal, also called features, to make it more convenient to work with. These EMG features are generally in the time or frequency domain. One of the most commonly used time domain features for EMG analysis is the Root Mean Square (RMS) of the signal [29], see bottom graph in Figure 3. This feature is so popular due to its quick calculation and easy implementation, whilst preserving significative information. Many other features exist as well, and new ones are still being discovered, making selection of a compact, non-redundant feature set challenging.

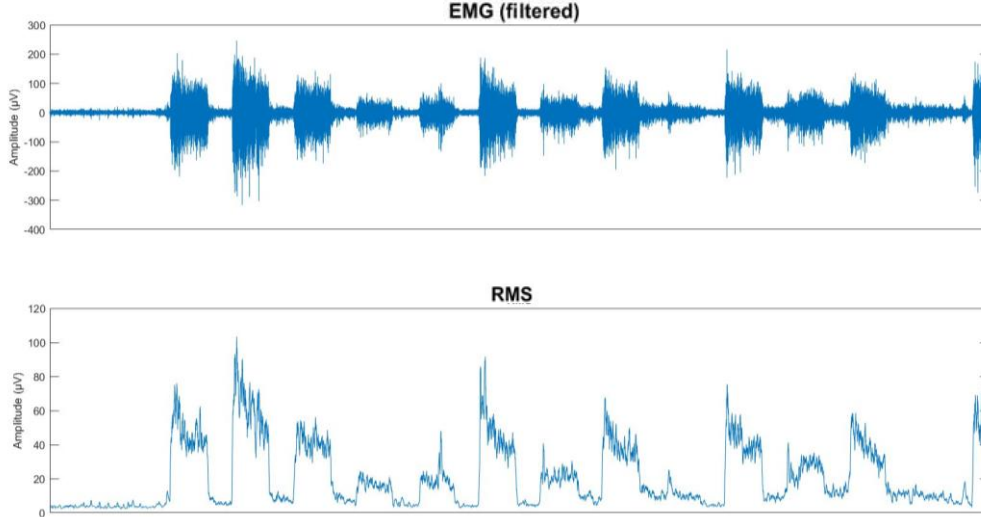


Figure 3: Filtered EMG signal (top) and corresponding Root Mean Square (RMS) values (bottom).

Two extensive studies [30], [31] have examined respectively 26 and 44 features for EMG-based algorithms. The first study [30] evaluated the features for myoelectric control based on wearable EMG sensors located on the wrist. They found that the features L-scale (LS), Integrated Absolute Value (IAV), Mean Absolute Value (MAV), Root Mean Square (RMS), Waveform Length (WL) and Difference Absolute Mean Value (DAMV), Difference Absolute Standard Deviation Value (DASDV) and Mean Value of the Square Root (MSR) have the best prospects in terms of classification rates. Most of these features are in agreement with findings of the second study [31], who evaluated the features for decoding of hand movements as well. They recommended MAV, Standard Deviation (STD), WL, DAMV and IAV to obtain high recognition accuracy and low processing time. As a result, the following selection of features will be used in this research: MAV, RMS, WL, SD, DAMV and IAV. The features are calculated with the formulas presented in Table 2 below, with the EMG data as x_i and the number of samples in each time window as N .

Table 2: The six features used in this research and their formula's, with EMG data as x_i and the number of samples in each time window as N . Adapted from [31], [32].

Feature	Formula
Mean Absolute Value (MAV)	$MAV = \frac{1}{N} \sum_{i=1}^N x_i $
Root Mean Square (RMS)	$RMS = \sqrt{\frac{\sum_{i=1}^N x_i ^2}{N}}$
Waveform Length (WL)	$WL = \sum_{i=1}^N (x_i - x_{i-1})$
Standard Deviation (SD)	$SD = \sqrt{\frac{1}{N-1} \sum_{i=1}^N (x_i - \bar{x})^2}$
Difference Absolute Mean Value (DAMV)	$DAMV = \frac{1}{N} \sum_{i=1}^{N-1} x_{i+1} - x_i $
Integrated Absolute Mean Value (IAV)	$IAV = \sum_{i=1}^N x_i $

1.2.4. State of the art

Researchers have already paved the way for classification of emotions based on EMG signals. Numerous papers regarding this subject can be found, varying in the expressions examined, number of channels used, channel placement, and type of features and classifiers evaluated. Accuracies vary in the range of 60% to almost 99%. The vast number of variables between studies makes it difficult to directly compare performance in terms of prediction accuracy. Nonetheless, some works related to this research are summarized in Table 3 and discussed below.

Table 3: Related research with some important properties. (FCM = fuzzy c-means, LDA = linear discriminant analysis, NARX = nonlinear autoregressive exogenous network).

Ref.	Expressions	n subjects	n channels	Features	Classifiers	Accuracy (%)
[33]	5 Happiness, anger, rage, frowning, neutral	4	2 bipolar sets	Root Mean Square (RMS)	FCM	90.8%
[34]	11 Anger, happiness, fear, sadness, surprise, neutral, clenching, half smile (left), half smile (right), frown, kiss	42	10	Root Mean Square (RMS) Waveform Length (WL) Sample Entropy (SE) Cepstral Coefficient (CC)	LDA	74.9%
[35]	3 Happiness, anger, disgust	12	2 bipolar sets	Wavelet Packet Transform (WPT): Mean Standard Deviation (STD) Energy	LDA	91.7%
[36]	4 Joy, anger, sadness, pleasure	Not reported	Not reported	Root Mean Square (RMS) Variance (VAR) Mean Absolute Value (MAV) Integrated EMG (IEMG)	NARX	98.8%

Hamed et al. [33] obtained an accuracy of 90.8% on five expressions with only 2 bipolar electrode sets. However, clear explanation on which data was used as training data and which as test data is lacking. Therefore it is uncertain whether these results are for a personalized model or for a general model, a distinction that is very important for interpretation of these results. In [34] the researchers aimed at training the classification model with only 1 trial per subject. They succeeded with an accuracy of 74.9% for detection of 11 expressions. Kehri et al. [35] obtained an accuracy of 91.7% with only 4 bipolar electrodes, for detection of happiness, anger and disgust. They used Linear Discriminant Analysis (LDA) and Wavelet Packet Transformed (WPT) features. Some of the best results in the field of facial expression recognition are obtained with deep learning. In [36] an accuracy of 98.8% was achieved for the classification of four expressions: joy, anger, sadness and pleasure. Unfortunately, deep learning is not always applicable due to the requirement of a large dataset.

An important aspect for application of a classification model on several patient groups is that a pre-trained model should be able to make predictions on new subjects, as it is fairly unfeasible to train the model on patients with limited facial movements. To evaluate the ability of a model to do this, the accuracy of the model on a completely new subject should be calculated. To our knowledge, this measure has not been reported before.

1.3. Machine Learning

Interest in machine learning has been rising in the last years. There are varying definitions going around, but the essence is that machine learning (ML) enables computers to think and learn independently, without explicit programming [37]. It is a type of artificial intelligence that can be applied in a widespread of domains: logistics, gaming, health-care, and so on. Some people confuse machine learning with deep learning and use those definitions interchangeably. However, that is incorrect. Deep learning is a type of machine learning, in which artificial neural networks adapt and learn from vast amounts of data, see Figure 6. A vast amount of data is one of the main requirements for using a deep learning approach, as it will not function properly on a smaller dataset.

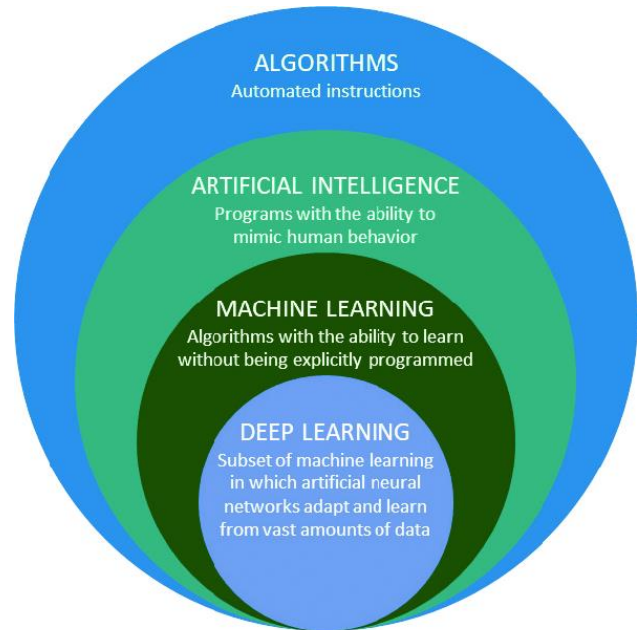


Figure 6: Schematic overview of the field of algorithms, comprising artificial intelligence, machine learning and deep learning [56]

Machine learning can be divided into supervised learning and unsupervised learning. The latter uses only input data to group and interpret the dataset, whereas the first makes predictions based on both input and output data. A problem that can be tackled using unsupervised learning is the clustering problem. The algorithm creates clusters based on similarities in the dataset, and new data is placed in the appropriate cluster. This type is also referred to as a statistic based approach. Supervised learning, where the outputs are known beforehand, can tackle either a classification or a regression problem. When the output is continuous, e.g. temperature in degrees Celsius, a regression model needs to be used. A categorical output, e.g. gender (being either male or female), calls for a classification model. This example is binary, but these type of models can also handle multi-class problems. [37]

The problem addressed in this research is a multi-class classification problem, several expressions (classes) need to be recognized. Besides supervised and unsupervised learning there are more categories. However, these two are the most commonly used. The general workflow of machine learning is as follows: features and labels are extracted from raw or pre-processed data. The data is divided into a training/validation dataset and a testing dataset. The model is trained and validated using the first dataset, see Figure 7 for a schematic overview. Once a proper model is obtained, the test accuracy can be determined by evaluating the predictive performance on a completely new dataset, not seen by the model before (the testing dataset).

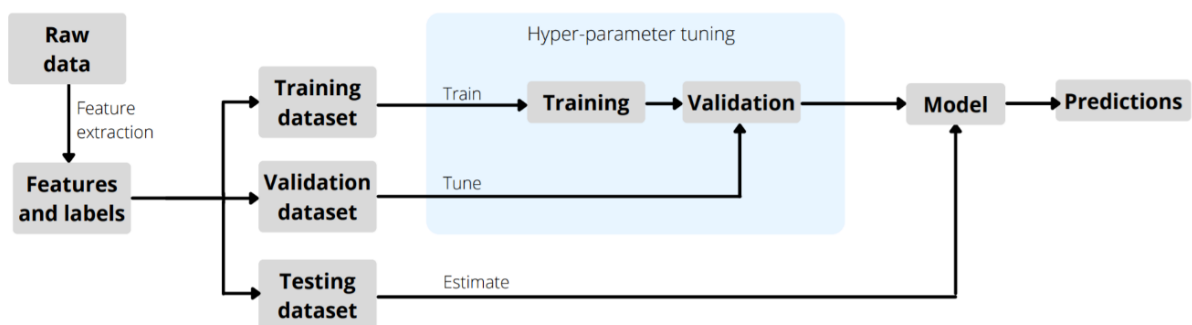


Figure 7: Workflow for machine learning. Features and labels are extracted from the raw data, which are then divided into a training, validation and testing dataset, used to create and test the model. Adapted from [57]

1.3.1. Classification algorithms

EMG pattern recognition typically calls for a classification algorithm. Commonly used classifiers include Support Vector Machines (SVM), Linear Discriminant Analysis (LDA), K-Nearest Neighbors (KNN), Random Forests and Naïve Bayes [38]. A comprehensive tool for evaluating the performance of multiple classifiers at once is MATLABs Classification Learner [39]. Roughly thirty classifiers can be trained in one go, see Table 4. Depending on data size this process can be time-consuming. Nonetheless, in this research all the classifiers are trained to explore as many options as possible. The most promising two are described below.

Table 4: The 29 classifiers trained in this research, listed by type.

Type	Sub-type	Type	Sub-type
Tree	<i>Fine tree</i>	Discriminant Analysis	<i>Linear discriminant</i>
	<i>Medium tree</i>		<i>Quadratic discriminant</i>
	<i>Coarse tree</i>	Naïve Bayes	<i>Gaussian Naïve Bayes</i>
K-nearest neighbors (KNN)	<i>Fine KNN</i>		<i>Kernel Naïve Bayes</i>
	<i>Medium KNN</i>	Support Vector Machine (SVM)	<i>Linear SVM</i>
	<i>Coarse KNN</i>		<i>Quadratic SVM</i>
	<i>Cosine KNN</i>		<i>Cubic SVM</i>
	<i>Cubic KNN</i>		<i>Fine Gaussian SVM</i>
	<i>Weighted KNN</i>		<i>Medium Gaussian SVM</i>
Ensemble	<i>Boosted Trees</i>		<i>Coarse Gaussian SVM</i>
	<i>Bagged Trees</i>	Neural Network (NN)	<i>Narrow NN</i>
	<i>Subspace Discriminant</i>		<i>Medium NN</i>
	<i>Subspace KNN</i>		<i>Wide NN</i>
	<i>RUSBoosted Trees</i>		<i>Bilayered NN</i>
			<i>Trilayered NN</i>

Ensemble classifiers combine results from multiple learners into one model to improve predictive performance. By fusing various algorithms the weaknesses of single learners diminish while their strengths add up to improve the outcome [40]. The learners can be combined in several ways (e.g. bagged, boosted, subspace), creating different subtypes of classifiers. Bagged Trees and subspace KNN ensembles will be reviewed below.

Ensemble classifier: Bagged Trees

The ensemble of Bagged Trees combines different decision tree learners to improve accuracy. Decision trees essentially break down problems into smaller decisions. It starts with a root node, being the first decision to be made. From this node several paths can follow to new nodes, forming a new decision. The final node, where no other nodes extend from, is called the terminal node or leaf, see Figure 8. The number of layers a tree has is referred to as the depth of that tree. With deep trees one needs to be cautious of overfitting. [41]

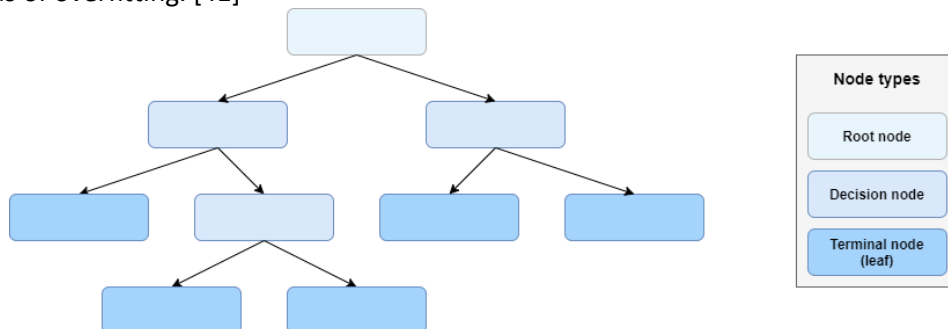


Figure 8: Schematic overview of a decision tree learner, with the three different node types listed: root node (start), decision node (middle) and terminal node (end).

Bagging is a combination of **bootstrap** and **aggregating**. It generates bootstrap replicates from the training data (see Figure 11), trains numerous learners on these sets, and averages the results (majority vote) to create one prediction, see Figure 9. In this case the learners are all decision trees. Another term for bootstrap replicating is sampling with replacement. This means that datapoints used in one replicate set, go back to the original dataset, after which they can be chosen again for a new replicate set. In other words, datapoints can occur in multiple replicate sets and not all datapoints are required to be used in the replicate sets. [42], [43]

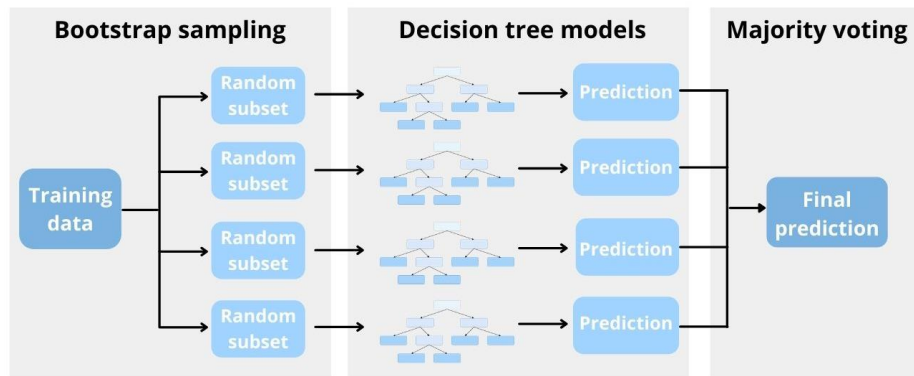


Figure 9: Overview of the Bootstrap Aggregating (bagging) method for decision tree learners. The training data is divided into several subsets on which multiple decision trees are trained. A majority vote results in one final prediction.

Ensemble classifier: Subspace KNN

The subspace KNN ensemble combines various K-nearest neighbor classifiers to improve accuracy. With KNN, data is classified according to the class of the k nearest neighbors of a datapoint [44]. E.g. the new datapoint in Figure 10 will be placed in either class A, B or C based on the smallest distance to a number of k datapoints in those classes.

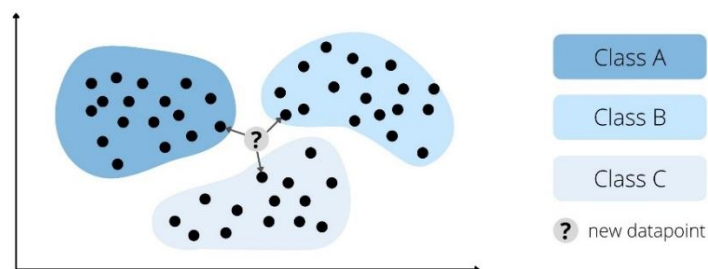


Figure 10: K-nearest neighbor (KNN) learner. The class of the new point will be determined based on the smallest distance to surrounding datapoints in either class A, B or C.

Those various KNN classifiers can be combined using the random subspace method. Random subsets of features are created, each training a weak learner. When new data is applied, the average class (majority vote) of all the weak learners is selected as prediction [45]. It is similar to bagging, however, the subsets are created across features instead of across training data, see Figure 11.

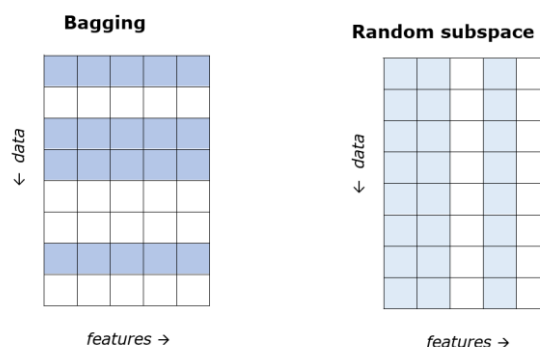


Figure 11: Two methods to create ensemble classifiers: bagging (left) and random subspace (right). Subsets are created across the data or across the features respectively.

2. Methodology

In this section the methods for the experimental procedure and data analysis are presented, together with substantiation and relevant literature.

2.1. Experiments

The experiments were conducted at a facility of the University of Twente: Zuid-Horst 285. This lab was equipped with a monitor, comfortable chair for the subjects and the TMSi Refa multichannel amplifier. For an elaborate list of used materials, see Appendix A: Protocol. Software used in the experiment include Matlab (version 2021b) with Psychtoolbox [46], TMSi Polybench toolbox [47], Stimulus Presenter toolbox [48] and Feature Extraction Toolbox [49]. This study received ethical approval from the ethical department of the University of Twente. Informed consent forms were obtained from all participants.

2.1.1. Subjects

The experiment was conducted on five healthy participants, no neuromuscular disorders and/or facial lesions were present. Each of the subjects had good vision, at least within 2 meters. Four of the subjects were female, one was male. Mean age of these subjects was 23.4 years with a standard deviation of 1.02 years, all subjects were Caucasian and were students at the University of Twente. See Table 5 for an overview of the subjects. Dutch was the native language of the participants, but all spoke sufficiently English to take part in this study.

Table 5: Information of the 5 participating subjects.

Number of participants	5
Gender	
- Male	1
- Female	4
Age	
- 22	1
- 23	2
- 24	1
- 25	1
Ethnicity	
- Caucasian	5

2.1.2. Expressions

This study focusses on four expressions: happiness, anger, sadness and fear, see Table 6. The first three are chosen due to their variation in action units. Anger and sadness share AU 4, which causes the eyebrows to lower. Other than that, these three expressions use different action units. In anger, there is more action around the eyes than only lowering of the eyebrows: the upper eyelids rise and tighten to get that intense look. The area around the mouth contracts as well, tightening the lips. In contrast to anger, the inner corners of the eyebrows can rise when sad. In addition sadness can usually be detected by depressed corners of the mouth: a characteristic feature of sadness, which is a main identification mark of the sad emoticon as well. Happiness is one of the main expressions where there is high activity in the Zygomatic Major, raising the corners of the mouth to create that distinctive smile. In addition, the muscles around the eye (Orbicularis Oculi) contract to raise the cheeks. The fourth expression is chosen for its similarities with sadness and anger, this is the expression of fear. Raising of the inner brow can occur when in fear, but it also appears when sad. Lowering of the brow and raising and tightening of the eyelids happens in both anger and fear. Action units that distinguish fear from the other three expressions are number 20 and 26: stretching the lips and dropping the jaw. The similarities and differences between these four expressions allow for proper investigation of the distinctive character of the expression classification algorithm(s).

Table 6: Emotions with corresponding Action Units (AU), muscles and resulting actions. Adapted from [50].

Emotion	AU	Related muscles	Action
Happiness	6	Orbicularis oculi, pars orbitalis	Cheek raiser
	12	Zygomatic Major	Lip corner puller
Anger	4	Depressor Glabellae (procerus) Depressor Supercilli Corrugator	Brow lowerer
	5	Levator palpebrae superioris	Upper lid raiser
	7	Orbicularis oculi, pars palpebralis	Lid tightener
	23	Orbicularis Oris	Lip tightener
Sadness	1	Frontalis, pars medialis	Inner brow raiser
	4	Depressor Glabellae (procerus) Depressor Supercilli Corrugator	Brow lowerer
	15	Depressor anguli oris (triangularis)	Lip corner depressor
Fear	1	Frontalis, pars medialis	Inner brow raiser
	2	Frontalis, pars lateralis	Outer brow raiser
	4	Depressor Glabellae Depressor Supercilli Corrugator	Brow lowerer
	5	Levator palpebrae superioris	Upper lid raiser
	7	Orbicularis oculi, pars palpebralis	Lid tightener
	20	Risorius	Lip stretcher
	26	Masseter; temporal and internal pterygoid relaxed	Jaw drop

2.1.3. Electrode configuration

For this experiment, 32 unipolar microelectrodes were used. The placement of these electrodes is determined based on the action units described above, together with guidelines from Fridlund and Cacioppo [27], which are commonly used in facial EMG research. However, these guidelines were created for a bipolar configuration, and as unipolar electrodes are used in this research the positioning might vary slightly.

First, muscles involved in the four expressions were covered with at least one electrode. These electrodes are represented in red in Figure 12. Additional electrodes, presented in blue, were added to cover a larger area of the face. All electrodes were located on muscle bellies as much as possible. Elaborate guidelines for placement of the electrodes can be seen in *Appendix A: Protocol*. Most positions are determined with relative distance to facial marks, others are placed at a fixed distance to certain points. Due to anatomical variety within subjects, electrodes can be placed slightly different to this setup. Especially when expressing emotions, the skin may fold or wrinkle on some places making it difficult for electrodes to properly adhere to the skin.

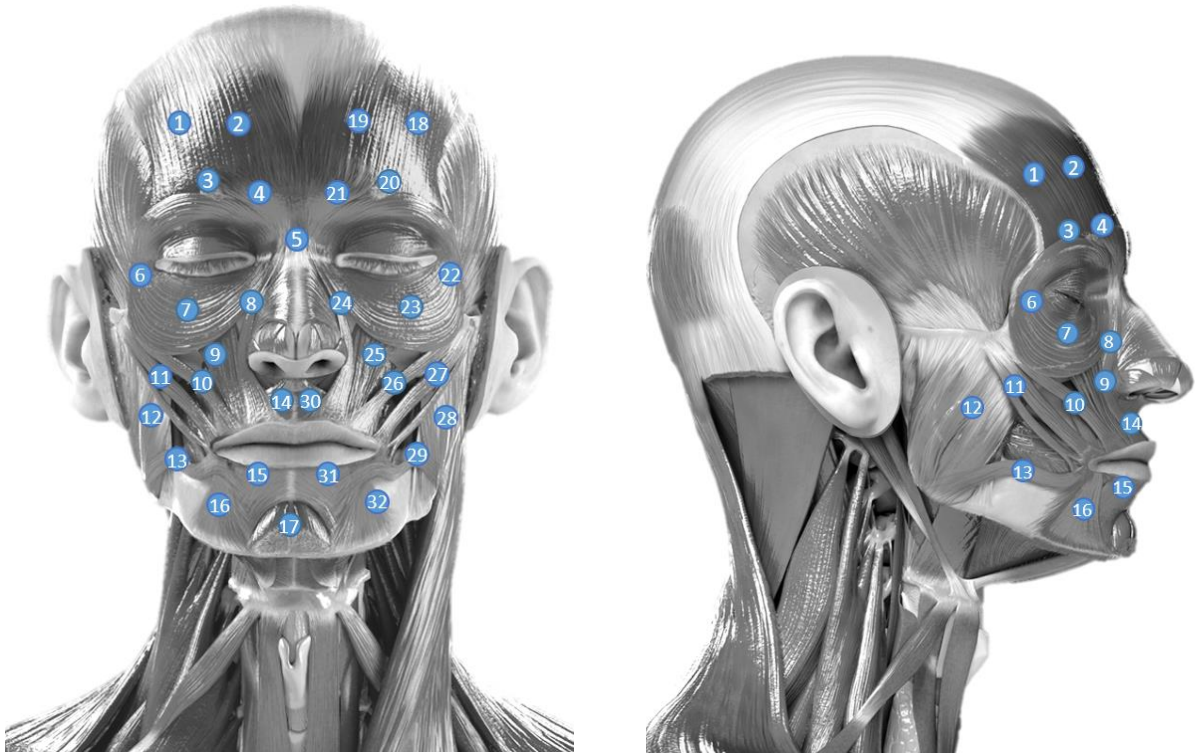


Figure 12: Configuration of the 32 unipolar microelectrodes, front view (left) and side view (right). Adapted from [58].

2.1.4. Experimental procedure

An overview of the setup used in the experiments can be seen in Figure 13 and Figure 14. The subject was seated in a comfortable chair, looking at the stimulus monitor. The stimulus monitor was controlled via the researchers laptop. The 32 microelectrode channels, plus the ground electrode, were connected to the researchers monitor via the EMG amplifier (TMSi Refa, with a 2048 Hz sampling frequency). A digital trigger device, connected to both the researchers laptop and the EMG amplifier, was used to send a trigger at the beginning of the experiment from the laptop to the amplifier for synchronization purposes.

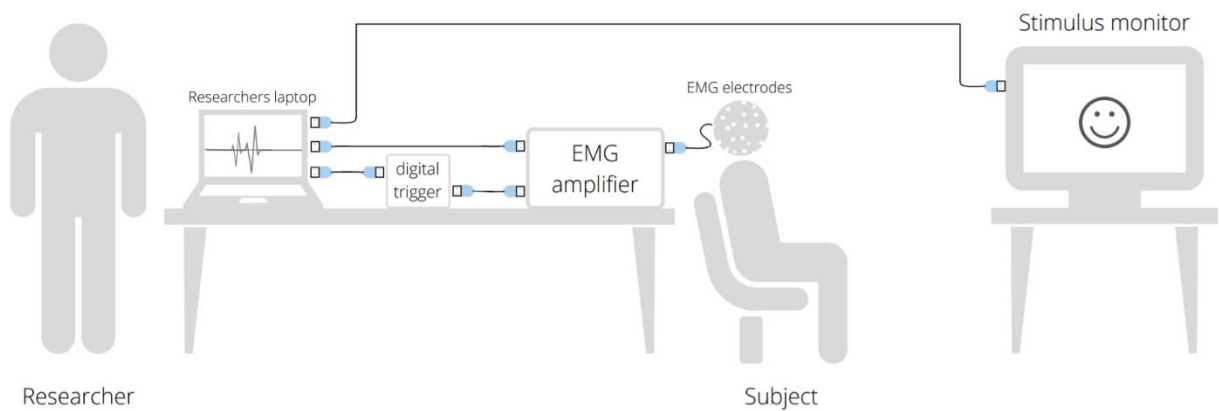


Figure 13: Schematic overview of the set-up for the experimental procedure.



Figure 14: Pictures taken during the experiment. The subject is seated in a chair in front of a screen on which stimuli are presented (left), the used electrode configuration (right).

The experiment consisted of three parts, all set up following the same construction (see Figure 15). There were four blocks in which 24 stimuli (either images of facial expressions or written expressions) were shown for a duration of 5 seconds, followed by a 3 second break (black screen). Once a block was finished, there was a longer break with a duration of 30 seconds. The stimuli was shown in a random order to vary within and between subjects.

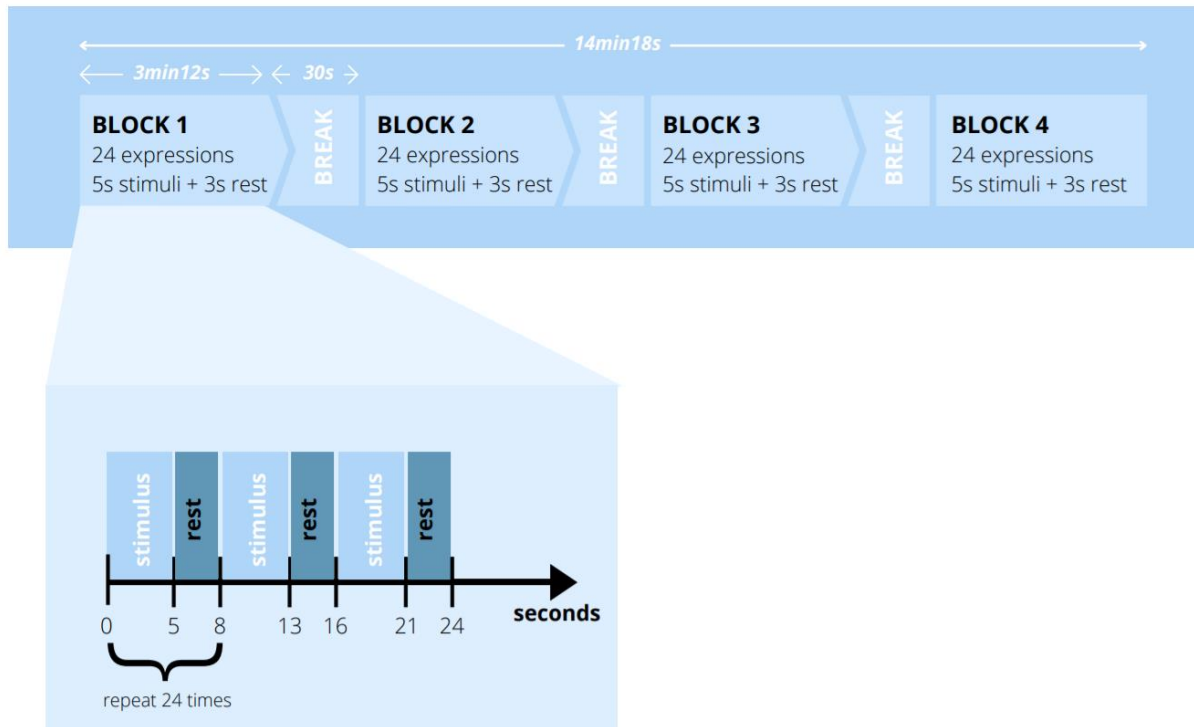


Figure 15: Schematic overview of the experimental procedure.

This format was performed three times with varying stimuli and instructions (see Figure 16). The stimuli could either be an image of one of the four facial expressions, or this emotion shown as a word. The images were acquired from the FACES database [51]. These consisted of staged expressions, performed by subjects (male and female) of three different age groups: young, middle-aged and old. For all images used, see *Appendix C: FACES database*.

In part I, subjects were instructed to look at the images without doing anything. They had to sit as still as possible to minimize movement artifacts. The goal of this part of the experiment was to measure micro-expressions, which have been shown to occur when looking at images of facial expressions. The theory behind this is to mimic the lower amplitude responses that can occur in DOC patients [52].

The stimuli used for part II were the same as used for part I: the images of facial expressions were showed again in random order. This time, the instructions were different. The subjects were told to mimic the images for as long as the images appeared on the screen (5 seconds). They did not have to worry or think about which expressions were showed, they simply had to mirror the image. The goal of this setup was to measure high-amplitude expressions. The used images all showed distinct emotions, which ensured activations of the main muscles involved is the specific emotions when mirroring.

Part III was added to the experiment to allow the subjects to express the emotions more naturally. The four emotions were randomly shown on screen in text (e.g. “Angry”, “Happy”). Subjects were instructed to express the emotion shown on screen for 5 seconds. They did not have to show exactly the same expression for each emotion, they were allowed to variate. The main objective was to naturally show these emotions, without giving it too much thought.

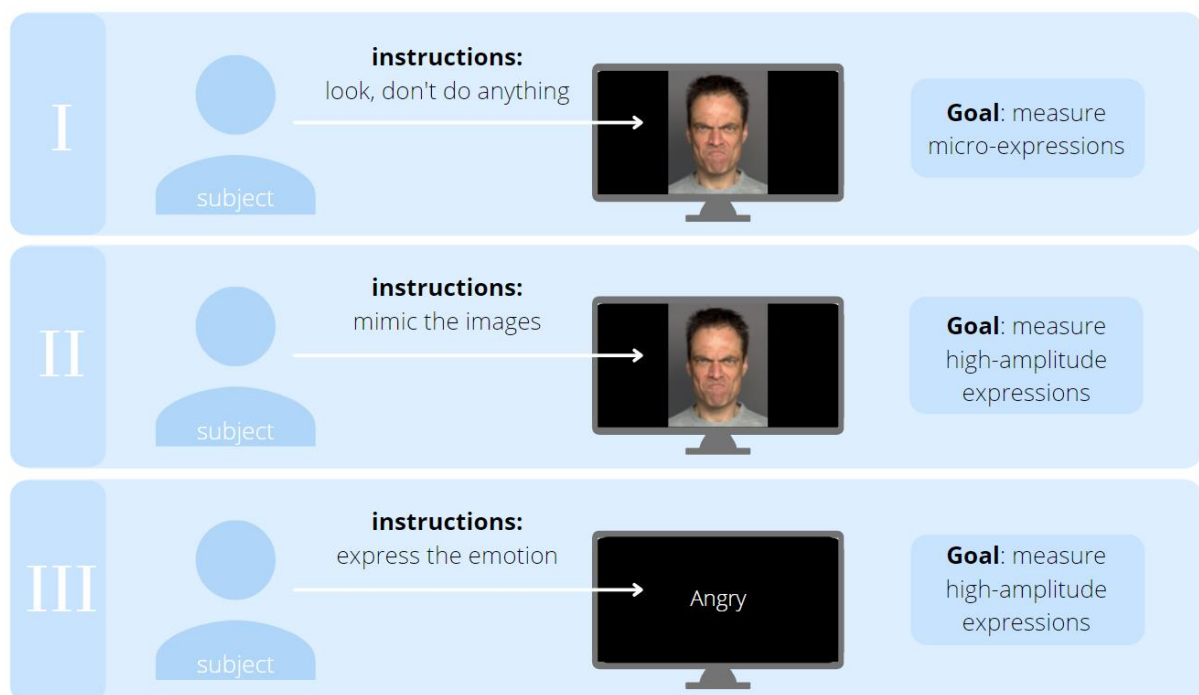


Figure 16: Overview of the three parts of the experiment.

2.2. Data analysis

Facial expression recognition algorithms based on EMG data usually consist of the following components: pre-processing, feature extraction and classification [38]. These three components will be discussed in the following sections.

2.2.1. EMG pre-processing

The raw EMG signals were processed offline, in MATLAB R2021a. Fast Fourier Transform (FFT) plots were created to inspect the frequency spectrum of the data and select appropriate filter values (see Figure 17). This resulted in filtering the signal at 40-500 Hz with a 4th order Butterworth bandpass filter, followed by a 50 Hz Notch filter to remove powerline noise. The signal was rectified and normalized as % of the baseline value, as done in previous research [53]. This baseline value was determined in the first 3 seconds of the experiments, where the participants looked at a cross positioned in the middle of the screen while maintaining a neutral face.

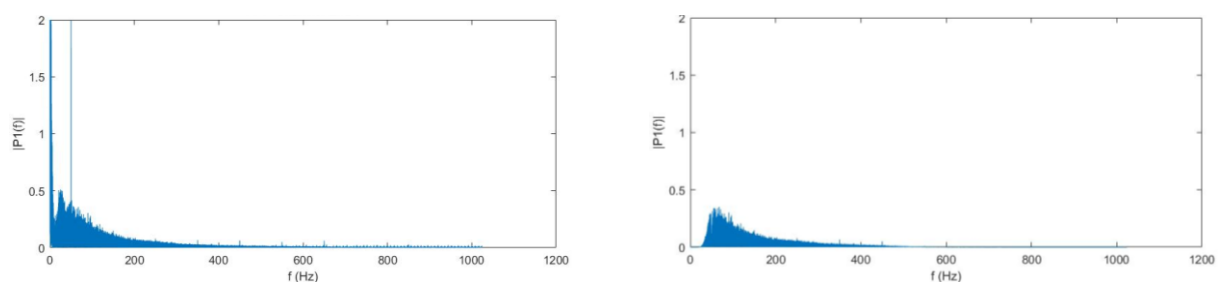


Figure 17: FFT plots of the raw signal (left) and filtered signal (right). Two filters are applied: a 4th order Butterworth bandpass filter at 40-500 Hz and a 50 Hz Notch filter.

Stimuli and EMG signals are synchronized by sending a digital signal to the EMG amplifier at the beginning of the experiment, at a known timepoint. The time between that trigger and all subsequent stimuli is known, therefore the corresponding EMG values can be detected and transformed to match the stimuli (see Figure 18 for a schematic overview).

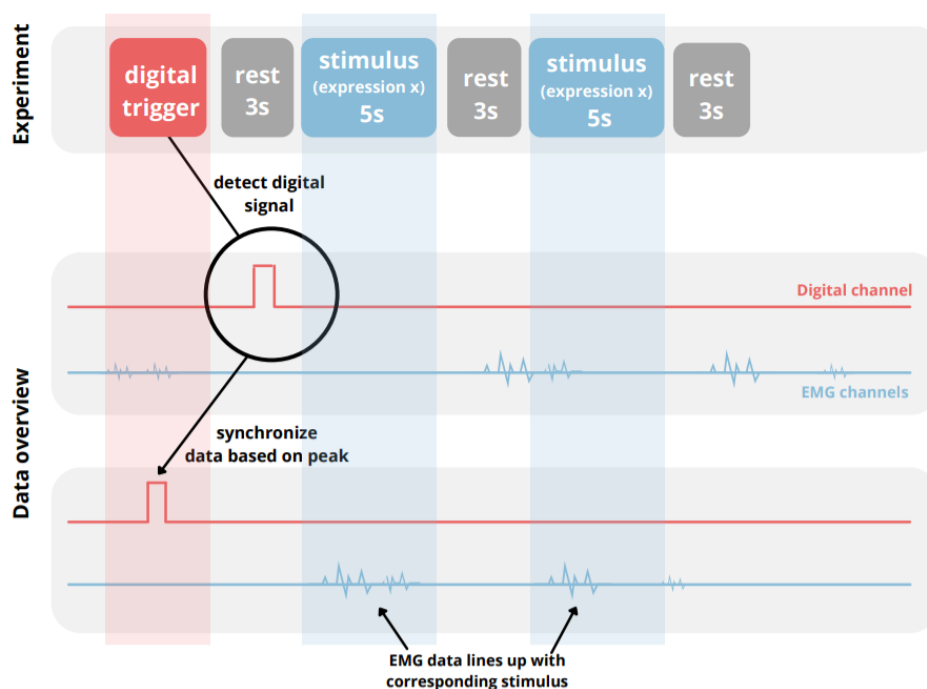


Figure 18: The process of synchronization. A digital trigger, sent at a known timepoint, is used to transform the data and match presented stimuli and EMG data.

2.2.2. Feature extraction

As discussed in section 1.2.3 *Electromyography (EMG)*, the six features used in this research were: MAV, RMS, WL, SD, DAMV and IAV. The features were extracted from the data with the Feature Extraction Toolbox [49]. All features were evaluated separately, after which the most promising were combined to possibly gain higher accuracy.

Window length

Features are not calculated over the whole signal at once but over epochs (time windows) of the signal. The length of these time windows varies between studies and substantiation for a certain length is often lacking. A comprehensive study [34] evaluated window lengths between 50 and 1500 ms, with a step size of 50 ms. Based on recognition accuracy an optimal window length was determined to be 1100 ms. In general, increasing the window length increased predictive accuracy. However, between 50 and 200 ms there was evident improvement in recognition accuracy, but after this point the recognition accuracy did not improve significantly. One also needs to take into account that a larger window contains more information. This means that processing takes more time, resulting in a possible delay when making real-time predictions. This relation between window length and amount of information has been evaluated, and a window length of 200 ms for static contractions and 300 ms for dynamic contractions was found to contain the maximum information [54]. In addition, it was found that the relation between the amount of information and window length varied between features. This shows that choosing the right window length is a meticulous decision, depending on required processing speed, selected features and type of contractions, among other things.

To determine the most proper window length for this specific research, three different non-overlapping window lengths were evaluated: 250, 500 and 1000 ms. After selection of the most promising window length, the remaining two lengths were disregarded from further analysis.

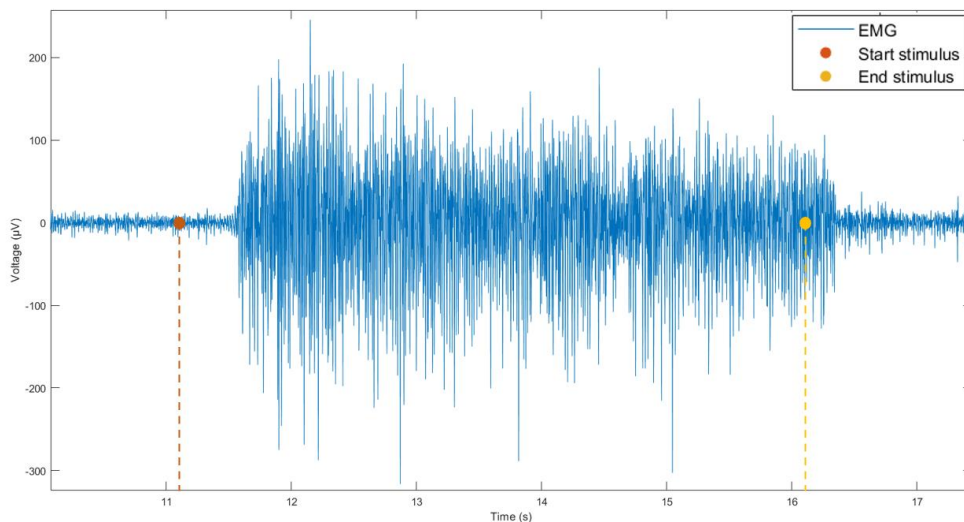


Figure 19: The delay (± 500 ms) between stimulus presentation (red dotted line) and the EMG signal (blue graph).

Delay

Each stimulus was shown for a duration of 5 seconds. However, there is a delay in the subjects between viewing the stimulus and expressing it. This is displayed in Figure 19: the stimulus appears on screen at the red dot (± 11.1 sec) and about 0.5 seconds later (± 11.6 sec) the facial muscles starts to contract. The length of this delay might vary between subjects. To ensure that only the period of muscle activation is selected for analysis, the first 1 second of every expression was excluded.

2.2.3. Classification

MATLAB's Classification Learner was used to train all 29 classifiers for all features and all three window lengths. This resulted in $29 \times 6 \times 3 = 522$ model performances. Out of these results, the best two combinations of window length, feature(s) and model were optimized. Data used for this originates from part II and III from the experiment, where macro expressions were measured. The micro expressions were excluded from analysis for this part, their analysis can be found in *section 2.2.4. Micro expressions*.

Model performance

The model's performance can be evaluated in terms of accuracy. There are two types of accuracy: validation accuracy and test accuracy. The first one is calculated based on the datasets used to train the model, see Figure 20. The test accuracy is calculated over a separate dataset, which the model has not seen before. All accuracies are calculated according to a 5-fold cross-validation method (see Figure 21). Using this method, the data is divided into 5 equal parts. Subsequently, all parts are excluded from the dataset and used as testing dataset in an iterative process. The final model accuracy is calculated as the average of these 5 iterations.

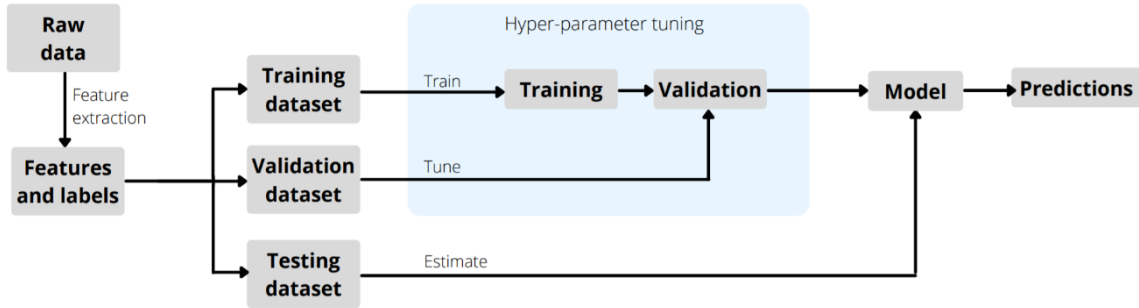


Figure 20: Workflow for machine learning. Features and labels are extracted from the raw data, which are then divided into a training, validation and testing dataset, used to create and test the model. Adapted from [57].

When developing a general model, as done in this research, it is important that a pre-trained model should be able to make predictions on new subjects. That is because it is not always possible to train a model on patients that are incapable of performing expressions voluntarily. To evaluate the ability of a model to predict well on new data, the testing accuracy is of utmost importance. Usually this is calculated using a subset of all the data, see the left image in Figure 21. This subset thus contains data

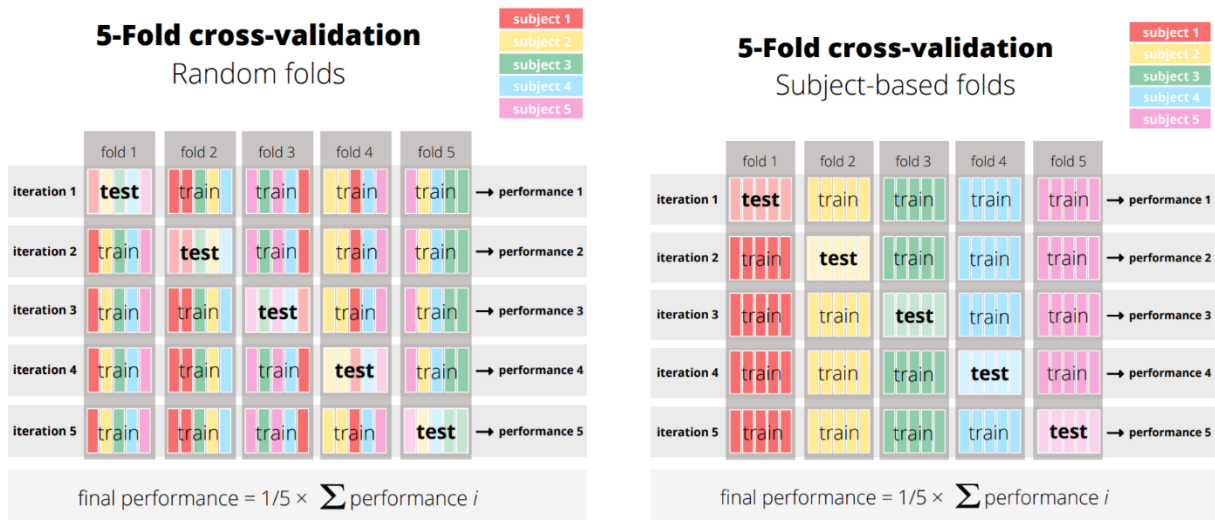


Figure 21: Two methods for 5-fold cross-validation: random folds (left) and subject-based folds (right). Data is divided into 5 equal parts. Subsequently all parts are excluded from the dataset and used as testing data. The final performance is calculated as the average of the 5 iterations. The colors represent the different subjects.

from all subjects used to train and validate the data as well. However, in this research we are interested in using a completely new subject as testing data. A proper way to evaluate the ability of the model to work with new data is by excluding subjects from the training dataset and using them as separate test datasets, see the right image in Figure 21. By doing this, the training data does not contain any data from that excluded subject. Consecutively this will be done for every subject, after which the average test accuracy can be calculated. When test accuracy is mentioned in the rest of this report, it refers to the latter method.

2.2.4. Micro expressions

The preprocessing steps for the micro expression data were similar to the macro expression data. The only difference is that features were extracted at other timepoints. As discussed in section 1.2.2. *Facial expression of emotions*, micro expression can occur within the first second after exposure to visual stimuli. Therefore features were calculated in this first second, which was excluded in analysis of the macro expressions.

For evaluation of the models performances on micro expressions, first instances where these micro expressions occur had to be identified. This is because the true class needs to be known in order to calculate the accuracy of the model. To identify the true classes, i.e. the micro expressions, activity maps were created. These activity maps were generated for every subject and every possible micro expression, over the following time points: 0-250, 250-500, 500-750 and 750-1000 ms, which is the duration after stimulus presentation. These activity maps were manually compared to the mean activity maps of every subject for the respective expression, see Figure 22. These mean activity maps were calculated over the EMG signals obtained from the macro expressions. Occurrences from all subjects where a micro expression seemed to evolve were combined to create a test dataset, after which the accuracy of the model for identification of micro expressions was evaluated, following the same steps as described in the previous section.

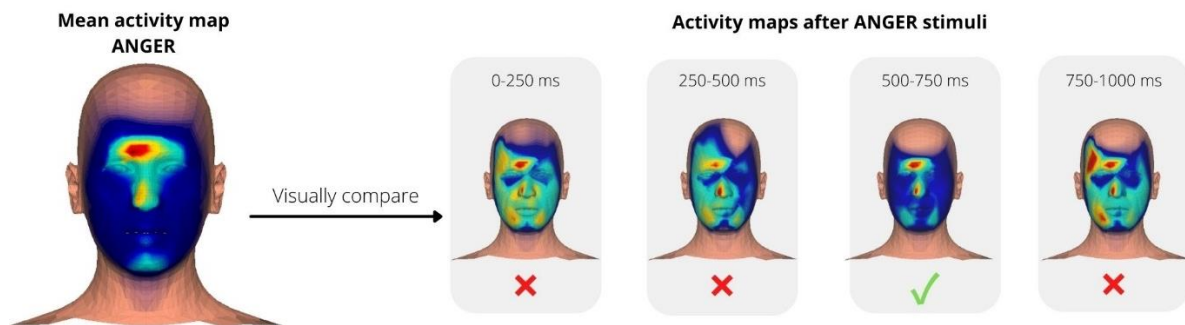


Figure 22: Overview of identification of micro-expressions, based on facial activity maps. The image at the third timepoint (500-750 ms) is a visual match for the mean activity map shown on the left.

2.2.5. Channel selection

In this research data was acquired with a 32-channel EMG set-up. Not all of these channels will add to the model's performance to the same extent. To simplify the experimental setup in the future, subsets of channels were evaluated. The predictive power of all channels was determined and interpreted via a predictor importance plot in MATLAB. Next, the models validation and test accuracies with the most important channels were determined. In an iterative process, channels were added consecutively to assess improvement of performance. During this iterative process facial symmetry was preserved, meaning that if a channel on one side of the face had a high predictive importance, the mirrored channel on the other side of the face was added as well. Based on these results, channels that show little predictive value can be disregarded in future research and an optimal channel subset can be established to accurately identify the four emotions included in this research.

3. Results

In the following sections the results of the experiments and data analysis will be presented. This contains the model performance (in terms of validation accuracy and test accuracy), micro expressions and channel subsets.

3.1. Model performance

3.1.1. Validation accuracy

For this research the validation accuracy of 29 classifiers, over 6 features and 3 window lengths has been evaluated. Results for all classifiers, features and window lengths can be found in *Appendix D: Results for all classifiers, features and window lengths*. Optimization was performed, but did not yield better results (see *Appendix E: Optimization*). The best accuracies were obtained with a window length of 250 ms and the ensemble models:

- Subspace KNN with feature DAMV
- Subspace KNN with feature WL
- Bagged Trees with feature DAMV
- Bagged Trees with feature WL

Confusion matrices for each of these four models are shown in Figure 23. The numbers represent the percentages of the expressions horizontally classified as the expressions vertically. The expressions classified correctly most often are happiness and anger. Sadness is the expression most often misclassified, and was for the majority misidentified as anger. Overall performance is good, indicated by the deep blue color on the diagonal and a white color off diagonal.

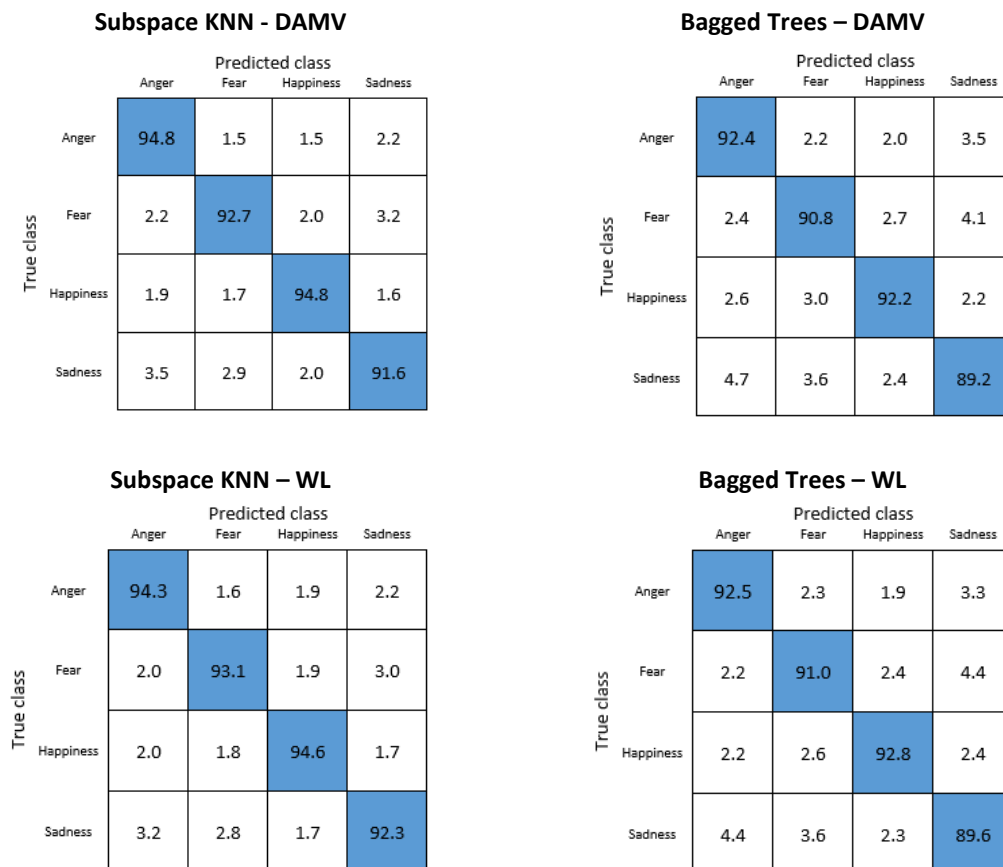


Figure 23: Validation confusion matrices for all four models, the numbers represent the percentage (%) of each expression shown horizontally that was classified as the expressions vertically. (DAMV=Difference Absolute Mean Value; WL=Waveform Length).

The largest overall validation accuracy was 93.6% for the subspace KNN model with features WL and DAMV separately, see Figure 24. Bagged Trees performed slightly worse, with a validation accuracy of 91.1% with feature WL and 91.2% with feature DAMV.

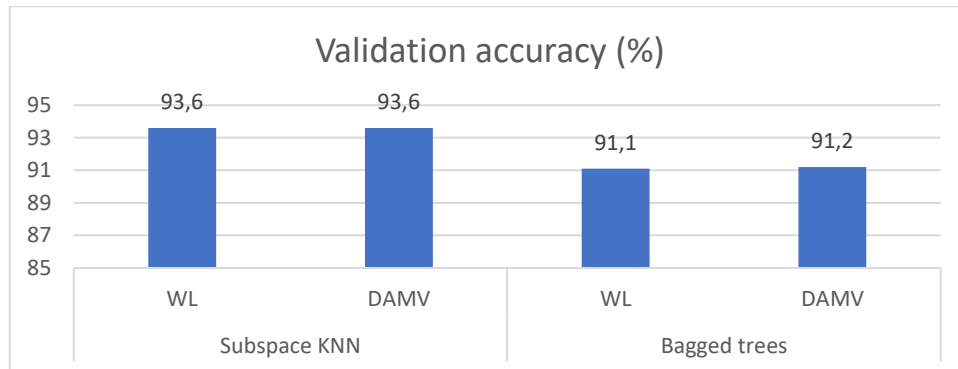


Figure 24: Validation accuracies for the four models. From left to right: Subspace KNN with feature Waveform Length (WL), Subspace KNN with feature Difference Absolute Mean Value (DAMV), Bagged Trees with feature WL and Bagged Trees with feature DAMV.

3.1.2. Test accuracy

The confusion matrices for the test sets are presented in Figure 25. The subspace KNN models classified the expression of happiness most often correctly, with a true positive rate of 81.8% for feature DAMV and 80.3% for feature WL. Sadness was most often misclassified, in most cases as anger. In the Bagged Trees model with feature DAMV, happiness was also most often classified correctly, with a true positive rate of 68.8%. For the Bagged Trees model with feature WL, fear was most often correctly predicted, with a true positive rate of 66.0%. Sadness was most often misclassified, the models misidentified it as fear most often.

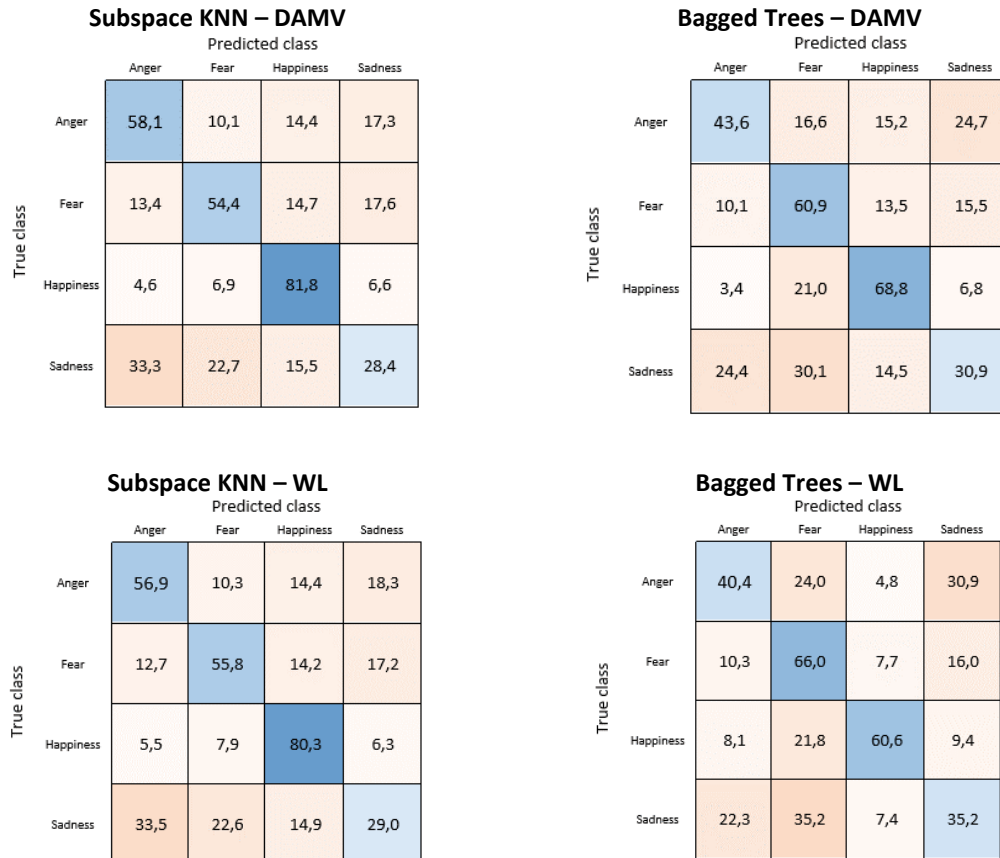


Figure 25: Test confusion matrices for all four models. The numbers represent the percentage (%) of each expression shown horizontally that was classified as the expressions vertically. (DAMV=Difference Absolute Mean Value; WL=Waveform Length).

Test accuracies can be seen in Figure 26 below. Subspace KNN with feature DAMV performed best, with an accuracy of 55.7%, followed by Subspace KNN with feature WL (55.5%). Bagged Trees performed worse, with test accuracies of 51.0% and 50.6% for respectively feature DAMV and WL.

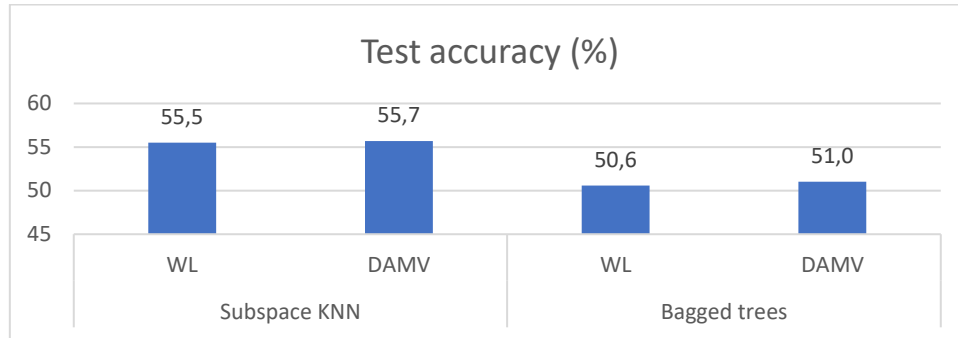


Figure 26: Test accuracies for the four models. From left to right: Subspace KNN with feature Waveform Length (WL), Subspace KNN with feature Difference Absolute Mean Value (DAMV), Bagged Trees with feature WL and Bagged Trees with feature DAMV.

3.2. Micro expressions

3.2.1. Manual identification

Mean activity maps per expression for all five subjects are shown in Table 8 on the next page. The number of micro expressions manually identified, based on these mean activity maps, is shown in Table 7. For subject 1 the most micro expressions were detected, with a total of 33. For the other subjects there were significantly less micro expressions identified, with no expressions at all for subject 4.

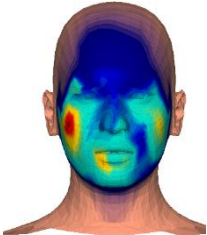
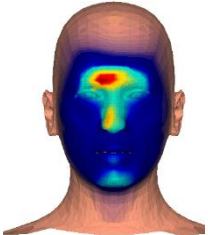
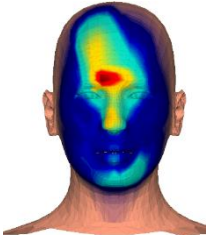

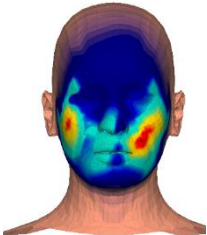


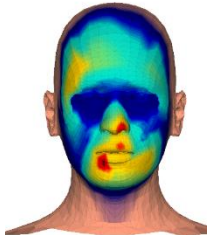
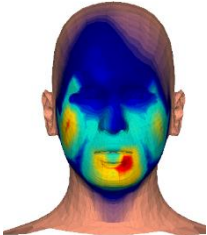
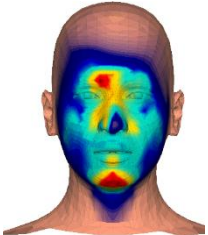
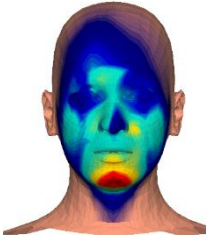

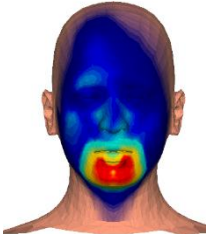
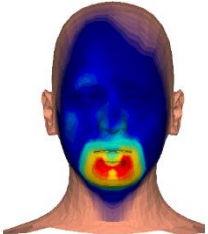
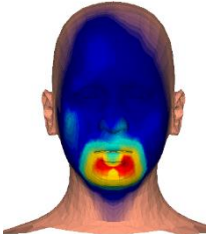
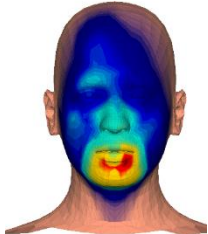
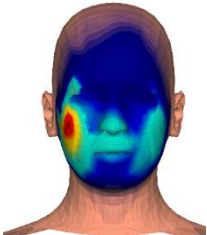
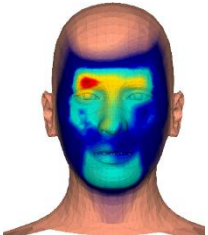

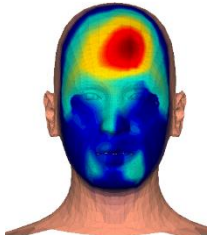
Table 7: Overview of the number of manually identified micro expressions per subject and in total.

Subject ↓	Expression				Total
	Happiness	Anger	Sadness	Fear	
1	8	11	9	5	33
2	3	1	2	3	9
3	4	0	0	3	7
4	0	0	0	0	0
5	0	0	4	1	5
Total	15	12	15	12	54

The mean activity maps of subject 4 stand out, being almost identical for each of the four expressions, with high activity at the chin. Subject 1 shows overlap between anger, sadness and fear with high activity at the right eyebrow and forehead. For subject 2 anger and sadness overlap as well, with high activity around both eyebrows. Happiness and fear are dissimilar to each other and the rest. In subject 3, both happiness and fear show activity at the left corner nearby the bottom lip, and both anger and sadness show high activity on the middle of the chin. Anger differs in that additionally there is high activity at the right eyebrow. For subject 5, both sadness and fear show increased activity at the left forehead. The activity map of anger shows distinctive activity at the right eyebrow, and happiness is characterized by high activity in the right cheek.

The activity maps of all subjects for happiness show high activity in the right cheek, with exception of subject 4. When expressing anger, there is activity around the eyebrows for all subjects (excl. Subject 4). For subjects 2, 3 and 5 this is mainly on the right side of the face, subject 2 shows higher activity on both sides. Sadness manifests in different ways. For subjects 3 and 4 the main activity is around the chin and for subject 1 the main activity is around the eyebrows. Subjects 2 and 5 show high activity both at the chin and eyebrows. The expression of fear also varies between subjects. Subject 2 and 3 show mainly activity around the mouth region whereas subject 1 shows activity around the right eyebrow and forehead and subject 5 around the left forehead.

Table 8: Mean facial muscle activity maps for the four emotions (happiness, anger, sadness and fear) shown for all five subjects. Red indicates a high activity, blue indicates a low activity.

		Expression			
		Happiness	Anger	Sadness	Fear
Subject	1	Subject 1. Happiness (mean) 	Subject 1. Anger (mean) 	Subject 1. Sadness (mean) 	Subject 1. Fear (mean) 
	2	Subject 2. Happiness (mean) 	Subject 2. Anger (mean) 	Subject 2. Sadness (mean) 	Subject 2. Fear (mean) 
	3	Subject 3. Happiness (mean) 	Subject 3. Anger (mean) 	Subject 3. Sadness (mean) 	Subject 3. Fear (mean) 
	4	Subject 4. Happiness (mean) 	Subject 4. Anger (mean) 	Subject 4. Sadness (mean) 	Subject 4. Fear (mean) 
	5	Subject 5. Happiness (mean) 	Subject 5. Anger (mean) 	Subject 5. Sadness (mean) 	Subject 5. Fear (mean) 

An example viewing the progress of muscle activity over time is shown below in Figure 27. Normalized RMS values of all 32 channels are plotted as a function of time after exposure to a happy stimulus. The large peak at 750-1000 ms represents high activity in the left Zygomaticus major (cheek), as can be seen in the corresponding activity map (4). After 1 second the values rapidly decrease again.

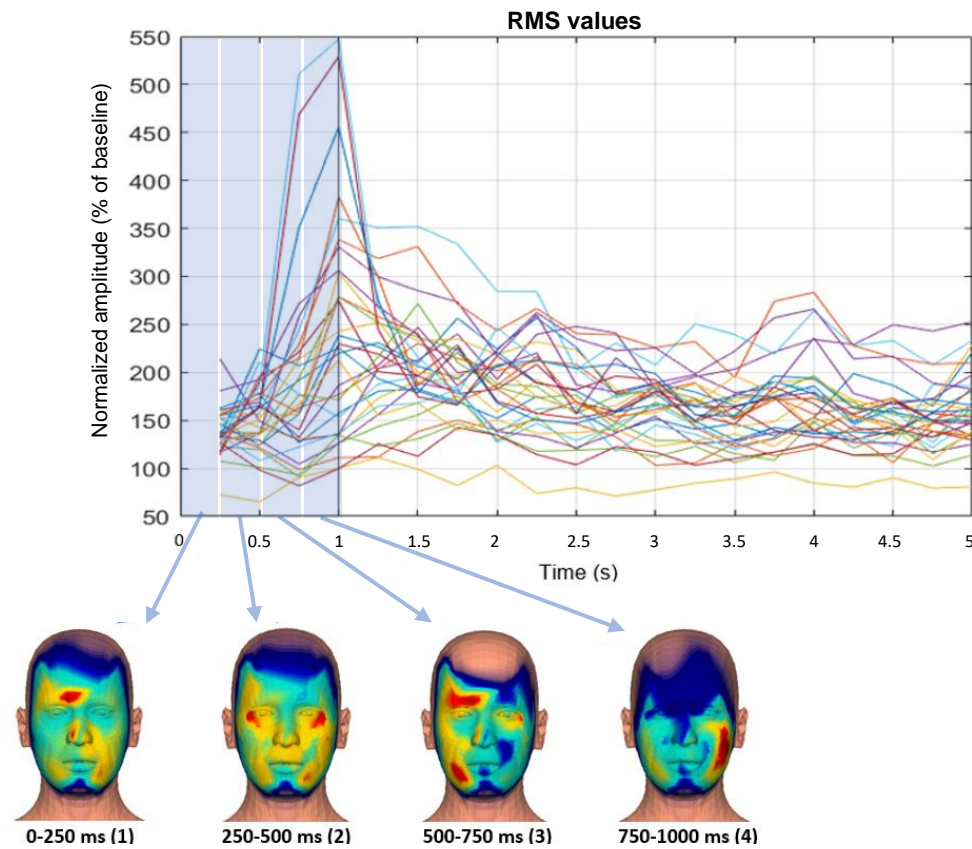


Figure 27: RMS of all 32 channels during exposure to happy stimulus (top), time on x-axis represents time after exposure. The corresponding facial activity maps are shown on the bottom.

3.2.2. Model performance

Only for subject 1 a sufficient amount of micro expressions were manually identified (Table 7). The maximum test accuracy (47.1%) for this subject was obtained with model subspace KNN and feature WL. The confusion matrix is presented in Figure 28. The model succeeds in predicting happiness and sadness, indicated by the dark blue color on the diagonal. Anger is mistaken as sadness in 79.2% of the cases, shown by the dark red square.

		Predicted class			
		Anger	Fear	Happiness	Sadness
True class	Anger	4,2	4,2	12,5	79,2
	Fear	33,3	16,7	33,3	16,7
	Happiness	0,0	5,0	75,0	20,0
	Sadness	0,0	0,0	20,0	80,0

Figure 28: Test confusion matrix for the micro expressions of subject 1. The numbers represent the percentage (%) of each expression shown horizontally that was classified as the expressions vertically.

3.3. Channel subsets

The importance of all channels is shown in Figure 29. The 32 channels correspond to the values on the x-axis. It can be seen that the most important channel is number 4, followed by channel 2, 1, and 11. After an iterative process of adding and removing channels (see *Appendix F: Channel subsets*) two promising subsets have been found. The first set (set A) consists of 8 channels, represented by the blue dots in Figure 30. The second subset consists of 7 additional channels (red dots in Figure 30) leading up to a total of 15 channels.

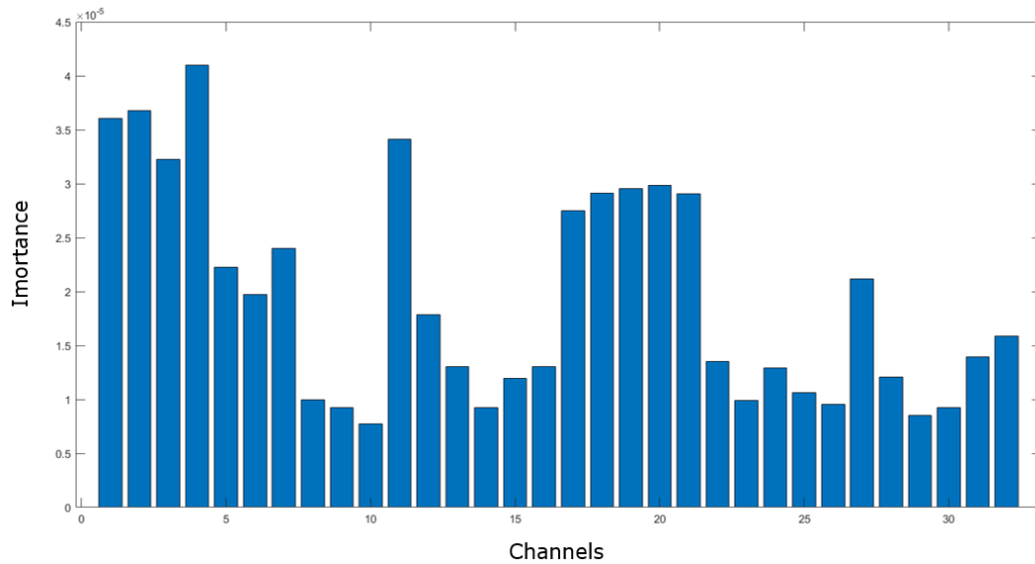


Figure 29: Plot of the channel importance for bagged tree classification algorithms. The 32 channels correspond to the numbers on the x-axis.

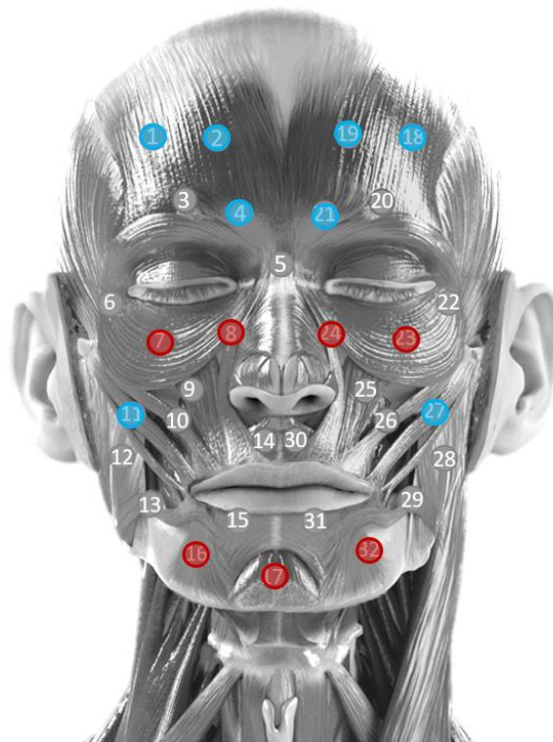


Figure 30: The electrode configuration of the two subsets of channels that have a high predictive value. Subset A is shown in blue ($n=8$), and subset B is shown in blue+red ($n=15$).

The validation and test accuracies for all channels (n=32) and the subsets (n=15 and n=8) are displayed respectively in Figure 31 and Figure 32. Largest validation accuracy for subset A (n=8) is obtained with Bagged Trees and feature WL (86.9%), and for subset B (n=15) with Subspace KNN and feature WL (92.1%). Highest test accuracies are achieved with Subspace KNN, feature WL (53.4%) and Bagged Trees, feature DAMV (45.9%) for respectively subset B (n=15) and subset A (n=8).

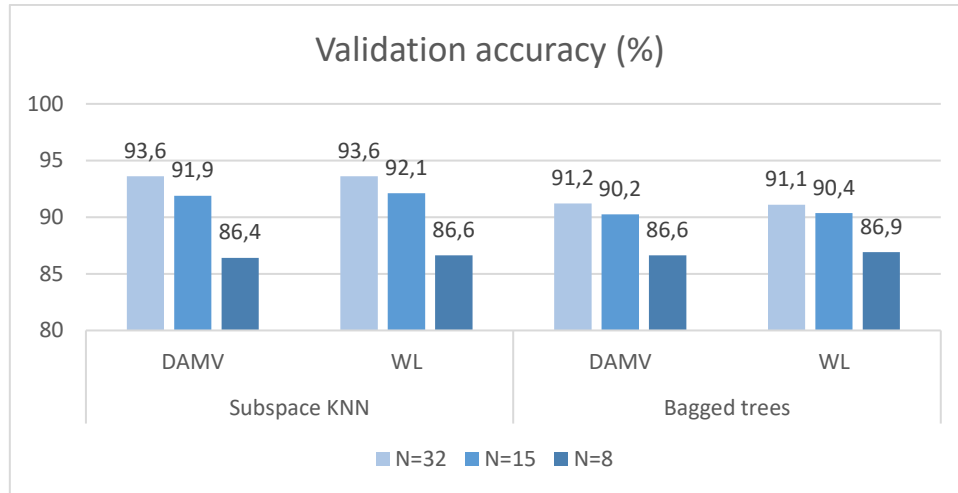


Figure 31: Validation accuracies for electrode configurations with all channels (n=32) and subset A (n=8) and subset B (n=15). From left to right: Subspace KNN with feature Difference Absolute Mean Value (DAMV), Subspace KNN with feature Waveform Length (WL), Bagged Trees with feature DAMV and Bagged Trees with feature WL.

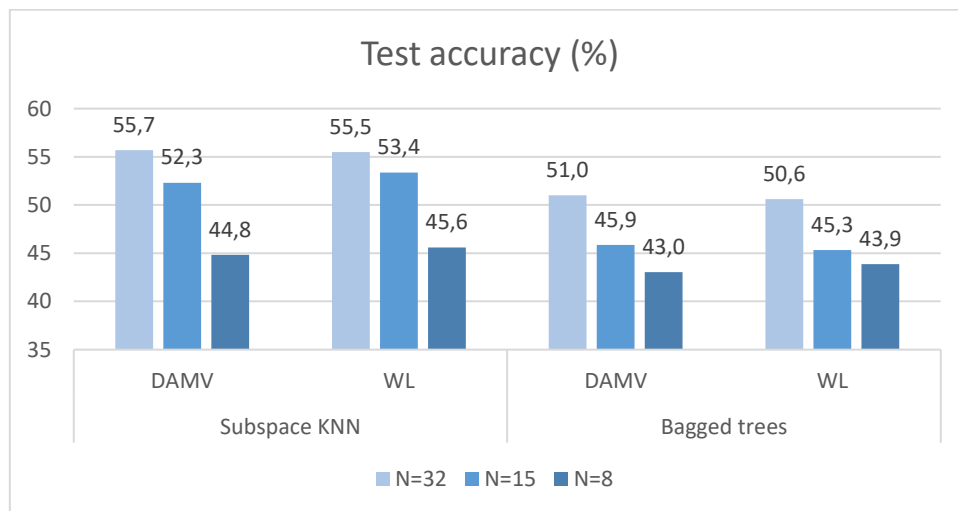


Figure 32: Test accuracies for electrode configurations with all channels (n=32) and subset A (n=8) and subset B (n=15). From left to right: Subspace KNN with feature Difference Absolute Mean Value (DAMV), Subspace KNN with feature Waveform Length (WL), Bagged Trees with feature DAMV and Bagged Trees with feature WL.

Confusion matrices for both subsets are shown in Figure 33 on the next page. It can be seen that for both subsets, the Subspace KNN models correctly classify the expression of happiness most often. For Bagged Tree models, fear is classified most often correctly, with exception of subset B (n=15) with feature DAMV, which correctly predicted happiness most often. In all models, sadness is misclassified most often. It is misidentified as fear in most of the cases.

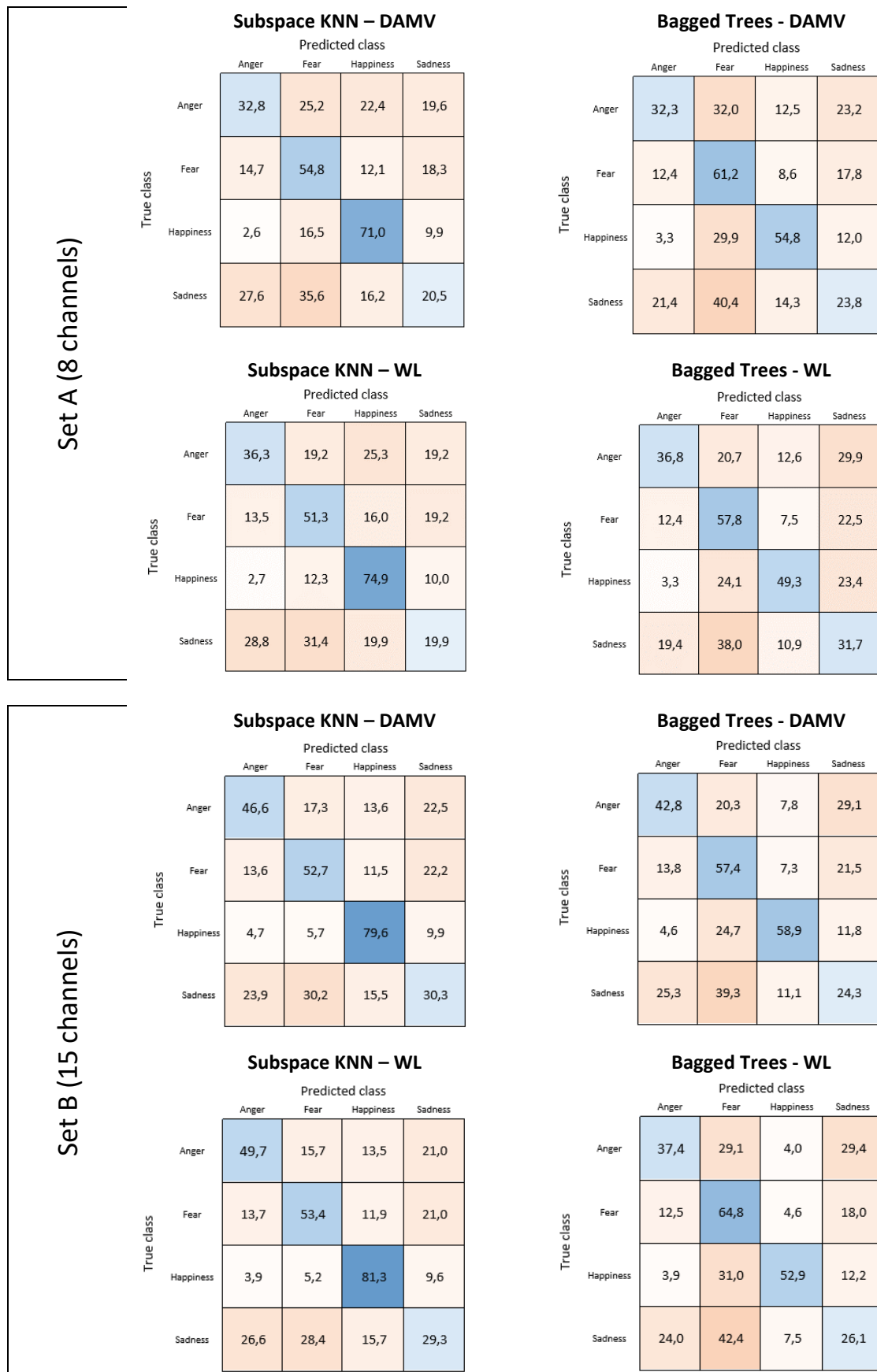


Figure 33: Test confusion matrices for all four models, for channel set A (top) and set B (bottom). The numbers represent the percentage (%) of each expression shown horizontally that was classified as the expressions vertically. (DAMV=Difference Absolute Mean Value; WL=Waveform Length).

4. Discussion

The main goal of this research was to evaluate to what extent facial sEMG signals can be used to classify the facial expression in healthy subjects. Twenty-nine models were included in this study, after which the most promising four models had been chosen for further development and validation. The model Subspace KNN and feature DAMV performed best at classifying expressions of subjects it had not been trained on (with a test accuracy of 55.7%). Evaluation of the predictive value of the 32 channels showed that a comparable test accuracy (53.4%) was obtained for a subset of only 15 channels with model Subspace KNN and feature WL. With 8 channels the maximum test accuracy reduced to a value of 45.6% for that same model. Additionally, this research has demonstrated that micro expressions, evoked by exposure to images of facial expressions, occur and can be measured with sEMG. The model Subspace KNN and feature WL succeeded in predicting these micro expressions for one of the subjects with a test accuracy of 47.1%.

4.1. Experimental procedure

To obtain the facial sEMG measurements an experimental protocol was designed. This protocol proved successful in attaining all required measures, including both the macro and micro expressions. One of the practical weaknesses of the protocol is the time required for positioning the electrodes on the face of the subjects. This meticulous process of preparation and placement of the 32 unipolar electrodes took up about one hour per subject. However, this 32-channel set-up was never intended for use in practice. The reason for investigating this large amount of channels was to identify the locations of the face that are most important when identifying the four expressions included in this study. When used in a practical setting, fewer electrodes can be used, as the evaluation of the channel subsets in this research have shown. This will decrease the preparation time.

In addition to being time consuming, the placement of the electrodes is rather subjective since the positioning is performed manually. Despite its subjectiveness, the placement instructions are based on the guidelines described by Fridlund and Cacioppo [27] which is the most commonly used method in facial EMG research. Nonetheless, variation in electrode positions between subjects could possibly lead to discrepancies in the signals measured, resulting in inaccurate predictions. Therefore the repeatability of this method should be thoroughly evaluated in future research. A possible method to position the electrodes more objectively would be to use a fixed template on which the electrodes are mounted. This is done by Cha et al. [34], who investigate the configuration of electrodes in a head-mounted display. However, a drawback of such a fixed template is that it has to be suitable for all subjects. The probability that this is the case is small since there is a lot of variety in the structure of people's faces. In future research, possibilities to make the protocol more suitable for application in clinical practice should be evaluated.

4.2. Window lengths

The optimal window length (out of 250, 500 and 1000 ms) was found to be 250 ms. The longer the window length, the worse the validation accuracy became. An optimal length of 250 ms is consistent with a previous study [31] that states that EMG segments with a length between 100 and 300 ms contain the maximum information. However, there could be another explanation. With a shorter window length, the number of datapoints increases. E.g. one second of data will be divided into 1 datapoint for a window length of 1000 ms, 2 datapoints for a window length of 500 ms and 4 datapoints for a window length of 250 ms. This thus results in a larger dataset to train the model on which contains datapoints similar to each other. In this way, it increases the chance that datapoints in the validation sets are comparable to datapoints in the training sets. To avoid this overlap in datapoints, the test accuracies should be calculated as done afterwards (by excluding each subject from the dataset subsequently). Though, as that was a laborious process, it was not feasible to calculate it for all window lengths and features on all 29 models included in the research. Priorities had to be set in order

to complete this thesis in time. Additionally, the optimal window length of 250 ms is in the same order of magnitude as a length of 256 ms, frequently used in similar research [33][36][55]. Therefore it is not expected that the window lengths of 500 and 1000 ms would yield better results.

4.3. Features

The most promising features were found to be the Waveform Length (WL) and Difference Mean Absolute Value (DAMV) separately. Combinations of various features did not improve accuracy. This is not in accordance with previous studies that typically use multiple features [34], [36]. In our research the best accuracies were achieved with only one separate feature for roughly all 29 types of classifiers and all three window lengths, from which it can be concluded that it does not depend on these two variables. Possibly all six features included in this research behaved similar. Adding multiple features together would not add much predictive value in that case. The features have in common that they are all time domain features. This decision was made because of recommended features from Abbaspour et al. [31] and Phinyomark et al. [32]. However, as both of these studies evaluated muscles around the wrist, it could be that their optimal feature sets are not optimal for use on facial EMG. As a result it is possible that a combination of other features would improve accuracy when used in combination instead of separate, but this hypothesis will have to be examined to accept with certainty.

Nonetheless, with features WL and DAMV separately, a maximum validation accuracy of 93.6% was achieved, comparable to related research using multiple features [33], [35]. When applying the classification algorithm in real-time, computation of only one feature could mean faster classification. However, this measure has not been elaborately explored in this research. The prediction speed of all four models was determined using MATLAB's classification learner (see *Appendix G: Speed measures*), but this did not include time for pre-processing steps (consisting of filtering, segmentation and extraction of the features). Therefore conclusions regarding the influence of the number of features on the prediction speed cannot be drawn yet. For future research into real-time applications, this is a measure that needs to be explored.

4.4. Classifier performance

Out of the 29 models included in this research, the ensemble models Subspace KNN and Bagged Trees outperformed all others. The acquired validation performances (with a maximum of 93.6%) are in comparison with results obtained by previous researchers. As cited in *section 1.2.4 State of the art* Kehri et al. [35] showed a classification accuracy of 91.66% on 12 subjects. The classifier used was a SVM and three facial emotions were classified: happiness, anger and disgust. Results obtained from SVM in our research are much lower, with a maximum validation accuracy of 73.4% (Fine Gaussian SVM – WL). These were non-optimized results, which may explain the poorer performance. This model was not optimized in this research because the ensemble classifiers outperformed all others. Another explanation for this large difference in validation accuracy of the SVM models could be the large variety between the studies. Kehri et al. use different expressions (disgust instead of fear and sadness), different features, use the Wavelet Packet Transform (WPT) and have a higher number of participants. This amount of variety between studies makes it difficult to pinpoint exactly what causes the differences. Researchers that have obtained higher accuracies often use deep learning [36], which was not possible in this research due to the limited number of subjects. For future research, the assessment of deep learning algorithms is recommended, to possibly improve accuracy even more.

The test performances (calculated by excluding each subject from the training set subsequently and testing the algorithm on that subject) were found to be 55.7%, 55.5%, 51.0% and 50.6% for respectively Subspace KNN (DAMV), Subspace KNN (WL), Bagged Trees (DAMV) and Bagged Trees (WL). There is room for improvement. A plausible explanation for these average results is the limited number of subjects used in this research. We attempted to create a generalized model. However, there is much more diversity in a population than displayed by these 5 people. In addition, possible deviations are

not smoothed out as much as would be when more people were used to train the model. The decision for such a limited number of subjects was made as this was a proof of principle study, to explore the practical potential. In this research the results from one of the subjects stood out, being much different than data from the other subjects. The mean test performance, without the results from testing that subject, show a maximum accuracy of 63.6% for Subspace KNN (DAMV). This is an absolute improvement of 7.9%, highlighting the influence of one subject on such a small sample group. It is expected that a larger subject group would yield better test performances.

The confusion matrices provide insight into the extent to which the models were able to distinguish between specific expressions. Happiness was correctly identified most often. This is in accordance with expectations, as activity in the Zygomaticus major only occurs in expression of happiness, being a good predictor. Sadness was the expression most often misclassified. It was misidentified mostly as anger (for Subspace KNN) and fear (for Bagged Trees). This is also in accordance with expectations since these expressions use similar muscles.

In practice, this would mean that sadness could be mistaken as anger or fear. To determine whether this would impose problems we need to take the desired application into account. For some applications it could be possible that discriminating between positive and negative expressions is more important than discriminating between the subcategories. Russell's circumplex model of affect (see Figure 34) displays affective states in terms of arousal (vertical axis) and valence (horizontal axis). Expressions on the left side of the graph are generally unpleasant, and expressions on the right side of the graph are experienced as pleasant. From this it could be concluded that distinguishing between happiness versus fear, anger and sadness is most important. The classification models developed in this research, especially the Subspace KNN models, appear to succeed at this. The four expressions explored in this research were chosen from an engineering point of view. To determine which divisions are essential for application in certain patient groups, medical specialists should be consulted.

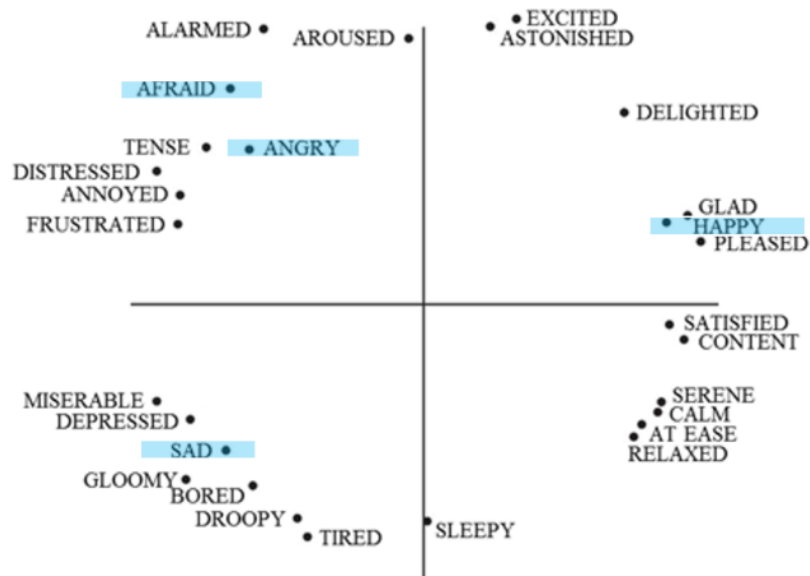


Figure 34: Russell's circumplex model of affect. Emotions are displayed in term of arousal or activation (vertical axis) and valence or pleasantness (horizontal axis). [59].

4.5. Micro expressions

One of the sub-goals of this research was to explore the possibility to provoke and measure micro expressions. The experimental procedure to elicit the micro expressions was based on previous research of Dimberg et al. [22]–[24], showing that images of facial expressions can induce micro expressions in certain subjects. To determine whether micro expressions occurred, they looked at graphs of the EMG response after exposure to stimuli for the evaluated electrodes ($n=2$). As in our research a much larger amount of electrodes was evaluated ($n=32$), this exact method was found to be unsuitable. A similar approach of visualization of the electromyographic response over time was chosen by use of the facial activity maps.

To identify micro expressions, mean facial activity maps for each expression were compared to the activity maps during the first second of exposure to the visual stimuli. Possible micro expressions were manually labelled, making this a subjective method. Nevertheless, micro expressions were identified in four out of five subjects. Subject 1 showed the most events with a total of 33 micro expressions. Following the method used by Dimberg et al. [22]–[24], Figure 27 confirms existence of micro expressions, and the ability of sEMG to measure them. This is in accordance with the expectations determined at the beginning of this research. Using the identified micro expressions as input for the machine learning models, a maximum test accuracy of 47.1% was obtained for subspace KNN with feature WL for subject 1. This shows a potential for classifiers trained on healthy subjects to function on patients with minimal facial activity as well. However, one side note that should be taken is that these micro expressions were studied to mimic the diminished facial expressions that occur in DOC patients. In this research only healthy subjects have been measured, and evoking micro expressions seemed like the most suitable method to simulate this behavior. In future research measurements should be performed on real DOC patients to determine whether the developed models properly function on that patient group.

While evaluating the micro expressions, results from one of the subjects stood out. The mean activity maps from subject 4 appear to be similar for all expressions (see Table 8). This indicates that either subject 4 does not show a clear difference between the four evaluated expressions, or a more probable cause, that normalization of the data did not work correctly. It is expected that this subject was not sitting still during the measurement of the baseline value. This only became apparent after all models had been developed. Would it be detected earlier, an attempt would have been made to establish the baseline value at another timepoint. It is expected that this subject caused a slight decrease in the model performances in this research.

4.6. Channel subsets

Evaluation of the subsets of channels show that with only 8 and 15 channels, maximum test accuracies of respectively 45.6% and 53.4% were obtained for the ensemble model of Subspace KNN and feature WL. The test accuracy for this model with all 32 channels was 55.5%. These results are relatively close to each other. This shows that, for these expressions, the classifiers perform similar with data of fewer electrodes. However, this does not mean that the specific subsets work well for detection of other expressions as well, as other muscles might be used in that case. This study provides evidence that an increase in electrodes does not yield a linear increase in predictor performance. In future research, once desired expressions to be detected have been determined, the subsets should be evaluated again.

5. Conclusion

From the results of this research it can be concluded that facial sEMG signals can be used to detect four facial expressions (happiness, sadness, anger and fear). The ensemble classification algorithms Subspace KNN and Bagged Trees show the most promising results, with features WL and DAMV separately. Out of three window lengths, 250 milliseconds resulted in the largest classification accuracy. Based on the required expressions to be detected, a subset of channels could provide sufficient information for classification. In addition, this research has proven that micro expressions can be provoked in certain subjects by presenting images of facial expressions, and that these micro expressions can be measured with sEMG.

Recommendations for future research include increasing the number of subjects to draw more significant conclusions. If a larger dataset is acquired, deep learning algorithms could be explored as well, avoiding the need for manual feature selection. In addition, medical experts should be consulted to determine which expressions need to be distinguished for the application on DOC patients, and possibly for other patient groups with weakened facial expressions as well. This will help to create an algorithm that is suitable for clinical application.

Bibliography

- [1] D. Canedo and A. J. R. Neves, "Facial expression recognition using computer vision: A systematic review," *Applied Sciences (Switzerland)*, vol. 9, no. 21. MDPI AG, Nov. 01, 2019. doi: 10.3390/app9214678.
- [2] I. M. Revina and W. R. S. Emmanuel, "A Survey on Human Face Expression Recognition Techniques," *Journal of King Saud University - Computer and Information Sciences*, Sep. 2018, doi: 10.1016/j.jksuci.2018.09.002.
- [3] T. Fadrus and O. Spindler, "Grimace." <http://www.grimace-project.net/> (accessed Oct. 21, 2020).
- [4] J. L. Bernat, "Chronic disorders of consciousness," *Lancet*, vol. 367, no. 9517, pp. 1181–1192, Apr. 2006, doi: 10.1016/S0140-6736(06)68508-5.
- [5] B. Jin, ; Yue Qu, ; Liang Zhang, Z. Gao, and L. Zhang, "Diagnosing Parkinson Disease Through Facial Expression Recognition: Video Analysis", doi: 10.2196/18697.
- [6] J. Perdiz, G. Pires, and U. J. Nunes, "Emotional state detection based on EMG and EOG biosignals: A short survey," Mar. 2017. doi: 10.1109/ENBENG.2017.7889451.
- [7] J. T. Giacino, Y. G. Bodien, and C. Chatelle, "CRS-R COMA RECOVERY SCALE-REVISED Administration and Scoring Guidelines."
- [8] J. Stender *et al.*, "Diagnostic precision of PET imaging and functional MRI in disorders of consciousness: A clinical validation study," *The Lancet*, vol. 384, no. 9942, pp. 514–522, Aug. 2014, doi: 10.1016/S0140-6736(14)60042-8.
- [9] F. Plum and J. B. Posner, *The Diagnosis of Stupor and Coma*, 3rd ed. Philadelphia, 1980. Accessed: Sep. 29, 2020. [Online]. Available: https://books.google.nl/books?hl=nl&lr=&id=Pbl4CH4NIQsC&oi=fnd&pg=PA1&ots=GwA5a-B78g&sig=xS4SNI7K8SAItT49Y3MghDpuqMk&redir_esc=y#v=onepage&q&f=false
- [10] J. M. Zadoks and R. Gijzen, "Naar meer bewustzijn Passende zorg voor mensen met langdurige bewustzijnsstoornissen," May 2018.
- [11] A. Rietveld, W. S. van Erp, P. E. Vos, B. Post, E. G. J. Zandbergen, and J. C. M. Lavrijssen, "Langdurige bewustzijnsstoornissen: diagnose, prognose en behandeling," *Tijdschrift voor Neurologie & Neurochirurgie*, vol. 117, no. 3, pp. 97–102, Sep. 2016, Accessed: Sep. 30, 2020. [Online]. Available: <http://www.coma.ulg.ac.be/nl/>
- [12] O. Gosseries *et al.*, "Disorders of Consciousness: Coma, Vegetative and Minimally Conscious States," in *States of Consciousness*, Springer, Berlin, Heidelberg, 2011, pp. 29–55. doi: 10.1007/978-3-642-18047-7_2.
- [13] W. Platzer, *Sesam Atlas van de anatomie. Deel 1: Bewegingsapparaat*, 1st ed., vol. 3. Amersfoort: ThiemeMeulenhoff, 20AD.
- [14] T. Marur, Y. Tuna, and S. Demirci, "Facial anatomy," *Clinics in Dermatology*, vol. 32, no. 1. pp. 14–23, Jan. 2014. doi: 10.1016/j.clindermatol.2013.05.022.
- [15] M. Schuenke, E. Schulte, and U. Schumacher, *Thieme Atlas of Anatomy. Head, Neck and Neuroanatomy*, 2nd ed., vol. 1. Stuttgart: Thieme Medical Publishers Inc, 2016.

- [16] L. Yao, Y. Wan, H. Ni, and B. Xu, "Action unit classification for facial expression recognition using active learning and SVM," *Multimedia Tools and Applications*, pp. 1–15, Apr. 2021, doi: 10.1007/s11042-021-10836-w.
- [17] B. C. Ko, "A brief review of facial emotion recognition based on visual information," *Sensors (Switzerland)*, vol. 18, no. 2, Feb. 2018, doi: 10.3390/s18020401.
- [18] "Micro Expressions | Facial Expressions | Paul Ekman Group."
<https://www.paulekman.com/resources/micro-expressions/> (accessed Jun. 14, 2021).
- [19] P. Ekman and W. v. Friesen, "The repertoire of nonverbal behavior: Categories, origins, usage, and coding," in *Mouton Classics: From Syntax to Cognition. From Phonology to Text*, vol. 1, no. 1, De Gruyter Mouton, 2013, pp. 819–868. doi: 10.1515/semi.1969.1.1.49.
- [20] W. J. Yan, Q. Wu, J. Liang, Y. H. Chen, and X. Fu, "How Fast are the Leaked Facial Expressions: The Duration of Micro-Expressions," *Journal of Nonverbal Behavior*, vol. 37, no. 4, pp. 217–230, Dec. 2013, doi: 10.1007/s10919-013-0159-8.
- [21] U. Dimberg and M. Thunberg, "Rapid facial reactions to emotional facial expressions," *Scandinavian Journal of Psychology*, vol. 39, no. 1, pp. 39–45, Mar. 1998, doi: 10.1111/1467-9450.00054.
- [22] U. Dimberg and M. Thunberg, "Empathy, emotional contagion, and rapid facial reactions to angry and happy facial expressions," *PsyCh Journal*, vol. 1, no. 2, pp. 118–127, Dec. 2012, doi: 10.1002/pchj.4.
- [23] U. Dimberg, M. Thunberg, and K. Elmehed, "Unconscious facial reactions to emotional facial expressions," *Psychological Science*, vol. 11, no. 1, pp. 86–89, May 2000, doi: 10.1111/1467-9280.00221.
- [24] U. Dimberg, P. Andréasson, and M. Thunberg, "Emotional empathy and facial reactions to facial expressions," *Journal of Psychophysiology*, vol. 25, no. 1, pp. 26–31, 2011, doi: 10.1027/0269-8803/a000029.
- [25] T. S. H. Wingenbach, M. Brosnan, M. C. Pfaltz, P. Peyk, and C. Ashwin, "Perception of Discrete Emotions in Others: Evidence for Distinct Facial Mimicry Patterns," *Scientific Reports*, vol. 10, no. 1, Dec. 2020, doi: 10.1038/s41598-020-61563-5.
- [26] U. Dimberg and L. O. Lundquist, "Gender differences in facial reactions to facial expressions," *Biological Psychology*, vol. 30, no. 2, pp. 151–159, Apr. 1990, doi: 10.1016/0301-0511(90)90024-Q.
- [27] A. J. Fridlund and J. T. Cacioppo, "Guidelines for Human Electromyographic Research," *Psychophysiology*, vol. 23, no. 5, pp. 567–589, 1986, doi: 10.1111/j.1469-8986.1986.tb00676.x.
- [28] A. v H, "A Review on Noises in EMG Signal and its Removal," *International Journal of Scientific and Research Publications*, vol. 7, no. 5, p. 23, 2017, Accessed: May 18, 2021. [Online]. Available: www.ijsrp.org
- [29] A. Phinyomark, P. Phukpattaranont, and C. Limsakul, "Feature reduction and selection for EMG signal classification," *Expert Systems with Applications*, vol. 39, no. 8, pp. 7420–7431, Jun. 2012, doi: 10.1016/j.eswa.2012.01.102.

- [30] A. Phinyomark, R. N. Khushaba, and E. Scheme, "Feature extraction and selection for myoelectric control based on wearable EMG sensors," *Sensors (Switzerland)*, vol. 18, no. 5, May 2018, doi: 10.3390/s18051615.
- [31] S. Abbaspour, M. Lindén, H. Gholamhosseini, A. Naber, and M. Ortiz-Catalan, "Evaluation of surface EMG-based recognition algorithms for decoding hand movements," *Medical and Biological Engineering and Computing*, vol. 58, no. 1, pp. 83–100, Jan. 2020, doi: 10.1007/s11517-019-02073-z.
- [32] A. Phinyomark, P. Phukpattaranont, and C. Limsakul, "Feature reduction and selection for EMG signal classification," *Expert Systems with Applications*, vol. 39, no. 8, pp. 7420–7431, Jun. 2012, doi: 10.1016/j.eswa.2012.01.102.
- [33] M. Hamed, I. Mohammad Rezazadeh, and M. Firoozabadi, "Facial gesture recognition using two-channel bio-sensors configuration and fuzzy classifier: A pilot study," in *INECC 2011 - International Conference on Electrical, Control and Computer Engineering*, 2011, pp. 338–343. doi: 10.1109/INECC.2011.5953903.
- [34] H. S. Cha, S. J. Choi, and C. H. Im, "Real-time recognition of facial expressions using facial electromyograms recorded around the eyes for social virtual reality applications," *IEEE Access*, vol. 8, pp. 62065–62075, 2020, doi: 10.1109/ACCESS.2020.2983608.
- [35] V. Kehri, R. Ingle, S. Patil, and R. N. Awale, "Analysis of facial EMG signal for emotion recognition using wavelet packet transform and SVM," in *Advances in Intelligent Systems and Computing*, 2019, vol. 748, pp. 247–257. doi: 10.1007/978-981-13-0923-6_21.
- [36] S. A. Mithbavkar and M. S. Shah, "Recognition of Emotion Through Facial Expressions Using EMG Signal," Jan. 2019. doi: 10.1109/ICNTE44896.2019.8945843.
- [37] J. Alzubi, A. Nayyar, and A. Kumar, "Machine Learning from Theory to Algorithms: An Overview," in *Journal of Physics: Conference Series*, Nov. 2018, vol. 1142, no. 1. doi: 10.1088/1742-6596/1142/1/012012.
- [38] A. Phinyomark and E. Scheme, "EMG pattern recognition in the era of big data and deep learning," *Big Data and Cognitive Computing*, vol. 2, no. 3. MDPI AG, pp. 1–27, Sep. 01, 2018. doi: 10.3390/bdcc2030021.
- [39] MATLAB & Simulink - MathWorks United Kingdom, "Classification Learner." <https://uk.mathworks.com/help/stats/classificationlearner-app.html> (accessed Jun. 14, 2021).
- [40] MATLAB & Simulink - MathWorks United Kingdom, "Classification Ensembles." <https://uk.mathworks.com/help/stats/classification-ensembles.html> (accessed Jun. 14, 2021).
- [41] S. R. Safavian and D. Landgrebe, "A Survey of Decision Tree Classifier Methodology," *IEEE Transactions on Systems, Man and Cybernetics*, vol. 21, no. 3, pp. 660–674, 1991, doi: 10.1109/21.97458.
- [42] L. Breiman, "Bagging predictors," *Machine Learning*, vol. 24, no. 2, pp. 123–140, 1996, doi: 10.1007/bf00058655.
- [43] J. Barugahare, A. N. Amirkhanian, F. Xiao, and S. N. Amirkhanian, "Predicting the dynamic modulus of hot mix asphalt mixtures using bagged trees ensemble," *Construction and Building Materials*, vol. 260, Nov. 2020, doi: 10.1016/j.conbuildmat.2020.120468.

- [44] A. Gul *et al.*, “Ensemble of a subset of kNN classifiers,” *Advances in Data Analysis and Classification*, vol. 12, no. 4, pp. 827–840, Jan. 2018, doi: 10.1007/s11634-015-0227-5.
- [45] J. Shin, “Random Subspace Ensemble Learning for Functional Near-Infrared Spectroscopy Brain-Computer Interfaces,” *Frontiers in Human Neuroscience*, vol. 14, p. 236, Jul. 2020, doi: 10.3389/fnhum.2020.00236.
- [46] “Psychtoolbox.” <https://www.psychtoolbox.net/> (accessed Jun. 14, 2021).
- [47] TMSi, “Software: TMSi Polybench.” <https://info.tmsi.com/software-tmsi-polybench> (accessed Jun. 14, 2021).
- [48] S. ter Haar *et al.*, “Matlab Stimulus Presenter.” http://eegget-it.nl/presenter_manual.html#msp_manual_tips (accessed Jun. 14, 2021).
- [49] J. Too, “EMG Feature Extraction Toolbox.” <https://uk.mathworks.com/matlabcentral/fileexchange/71514-emg-feature-extraction-toolbox> (accessed Jun. 14, 2021).
- [50] I. Tautkute and T. Trzciński, “Classifying and Visualizing Emotions with Emotional DAN,” *Fundamenta Informaticae*, vol. 168, no. 2–4, pp. 269–285, 2019, doi: 10.3233/FI-2019-1832.
- [51] N. Ebner, M. Riediger, and U. Lindenberger, “FACES: A database of facial expressions in young, middle-aged, and older women and men. Max Planck Society.” <https://faces.mpdl.mpg.de/imeji/collection/IXTdG721TwZwyZ8e?q=> (accessed Jun. 14, 2021).
- [52] C. M. Fiacconi and A. M. Owen, “Using facial electromyography to detect preserved emotional processing in disorders of consciousness: A proof-of-principle study,” *Clinical Neurophysiology*, vol. 127, no. 9, pp. 3000–3006, Sep. 2016, doi: 10.1016/j.clinph.2016.06.006.
- [53] A. van Boxtel, “Facial EMG as a Tool for Inferring Affective States,” 2010.
- [54] F. D. Farfán, J. C. Politti, and C. J. Felice, “Evaluation of EMG processing techniques using Information Theory,” *BioMedical Engineering Online*, vol. 9, no. 1, pp. 1–18, Nov. 2010, doi: 10.1186/1475-925X-9-72.
- [55] M. Revina and S. Emmanuel, “Face expression recognition using electromyography,” *International Journal of Latest Trends in Engineering and Technology*, pp. 53–55, 2017.
- [56] J. Vrana and R. Singh, “The NDE 4.0: Key Challenges, Use Cases, and Adaption,” 2020, Accessed: Jun. 14, 2021. [Online]. Available: https://www.researchgate.net/publication/339997962_The_NDE_40_Key_Challenges_Use_Cases_and_Adaption
- [57] A. Basavaraju, J. Du, F. Zhou, and J. Ji, “A Machine Learning Approach to Road Surface Anomaly Assessment Using Smartphone Sensors,” *IEEE Sensors Journal*, vol. 20, no. 5, pp. 2635–2647, Mar. 2020, doi: 10.1109/JSEN.2019.2952857.
- [58] “Anatomy Next.” <https://anatomy.net/> (accessed Jun. 22, 2021).
- [59] J. A. Russell, “A circumplex model of affect,” *Journal of Personality and Social Psychology*, vol. 39, no. 6, pp. 1161–1178, Dec. 1980, doi: 10.1037/h0077714.

Appendices

Appendix A: Protocol

Location: ZH285. **Duration:** 2 hours (+/- 1 hour subject preparation and 1 hour experiment)

Materials

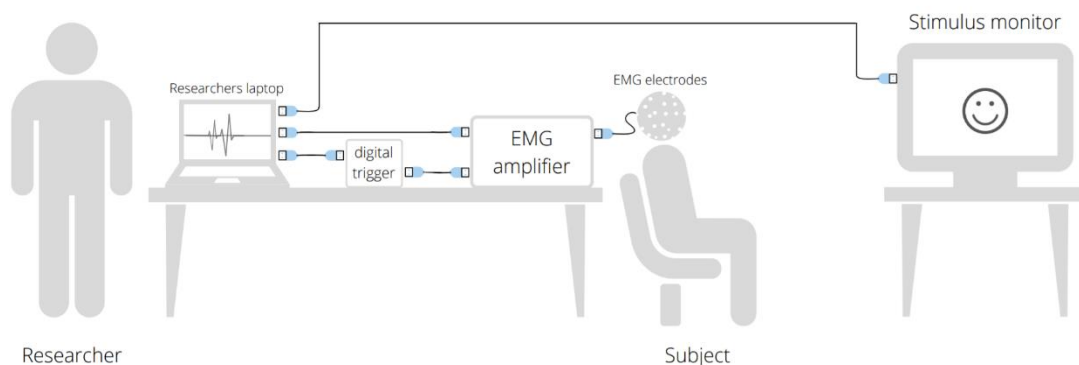
- Microelectrodes unipolar – 35 pcs.
- Microelectrodes bipolar – 2 pcs.
- 'Regular' electrodes – 1 pcs.
- Stickers (to adhere the electrodes to the skin) – 37 pcs.
- Electrode paste (Ten20)
- Spatula/knife (to add gel)
- Fine liner (to mark electrode locations before placement)
- Measurement tape (to determine electrode locations)
- Alcohol wipes
- Cleansing peeling (Nuprep)
- Face shield (Covid measures)
- Face mask (Covid measures)
- Protective gloves (Covid measures)

Hardware

- Laptop
- Monitor
- TMSi Refa
- Required cables to connect abovementioned

Software

- MATLAB
- Psychtoolbox
- TMSi Polybench toolbox
- Stimuli presenter GUI

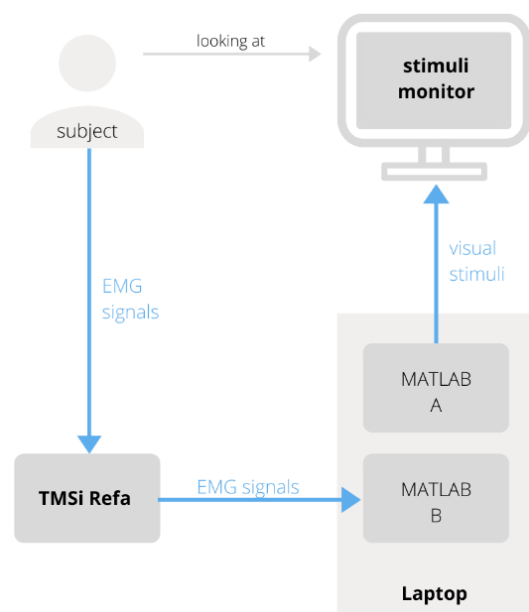


Experimental protocol

1. Preparation

The following steps should be performed before subject arrival.

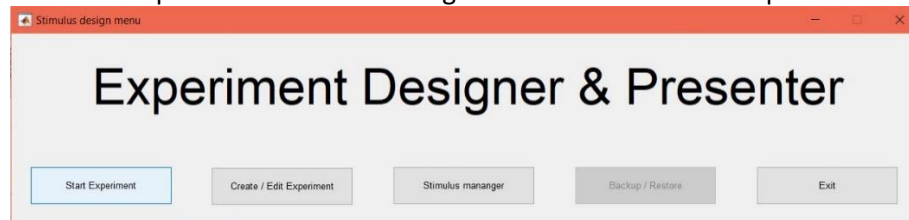
- 1) Start the operator laptop.
- 2) Connect the stimuli monitor to the laptop with a HDMI cable and ensure that the monitor is connected to the power supply.
- 3) Set the connected monitor to be the main screen. To do this, go to 'settings' on the laptop, 'monitor' and select that screen 2 should be the main screen. (note: if you skip this step, the stimuli will show on the operator laptop instead of on the separate monitor).
- 4) Open two individual MATLAB windows.



1.1 MATLAB window A

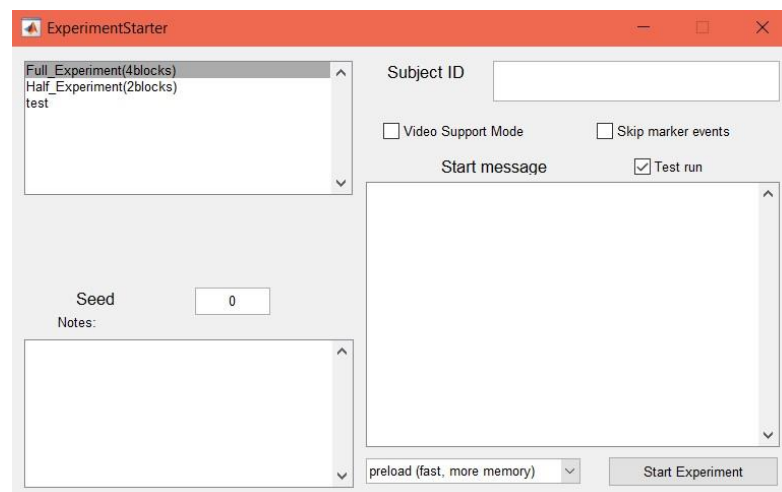
5) In window A, open the Matlab Stimulus Presenter. To do so, ensure the folder is added to the path, type 'start' in the command window and press enter.

6) Now a window will open titled 'Stimulus design menu'. Click on 'Start Experiment'.



7) The following screen will open. There are three experiments available:

1. Full_Experiment(4blocks) one full part of the experiment
2. Half_Experiment(2blocks) half part of the experiment
3. Test one image, to test the setup



8) In the same window there is a line with 'Seed'. This number determines the order of the stimuli presented. The stimuli will be presented in a random order, but with the same seed value, the order will be identical. Therefore, for each subject and each part of the experiment, this number should be changed. Use the table below.

9)

Subject ID	Part I	Part II
1	Seed = 0	Seed = 1
2	Seed = 2	Seed = 3
3	Seed = 4	Seed = 5
4	Seed = 6	Seed = 7
5	Seed = 8	Seed = 9

10) On the top right a subject ID can be filled in. This ID will show in the results of the stimuli presentation script, that is automatically saved as excel file. In this file the order and timing of presented stimuli is saved.

1.2 MATLAB window B

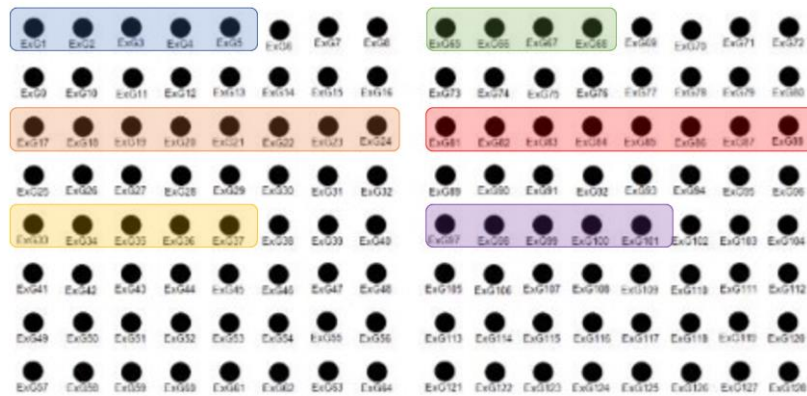
- 11) In the other MATLAB window, the scripts for impedance measurements and data acquisition can be opened. These script are called 'ImpedanceTest' and 'PlotAndSave_Poly5'.

1.3 Electrode preparation

- 12) For easier application the electrodes can be grouped beforehand. A division can be made as seen in the table below.

Area on face (left/right)	Specific area	Electrode numbers
Left	Forehead	1-5
	Cheeks	17-24
	Mouth	33-37
Right	Forehead	65-68
	Cheeks	81-88
	Mouth	97-101

The electrodes should be grouped in the Refa as well, according to the figure below. The colors from the last column of the table correspond with the colors in the figure.

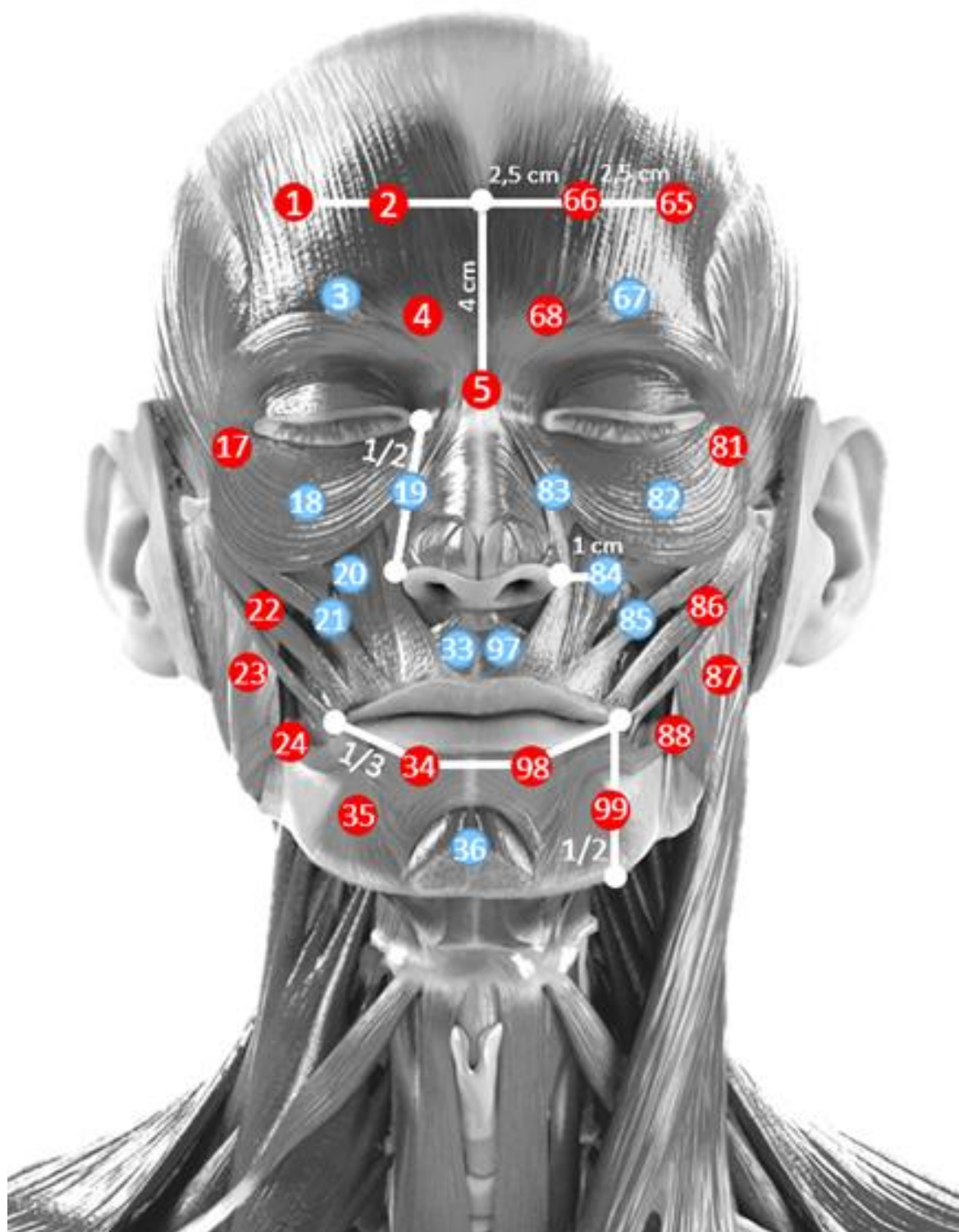


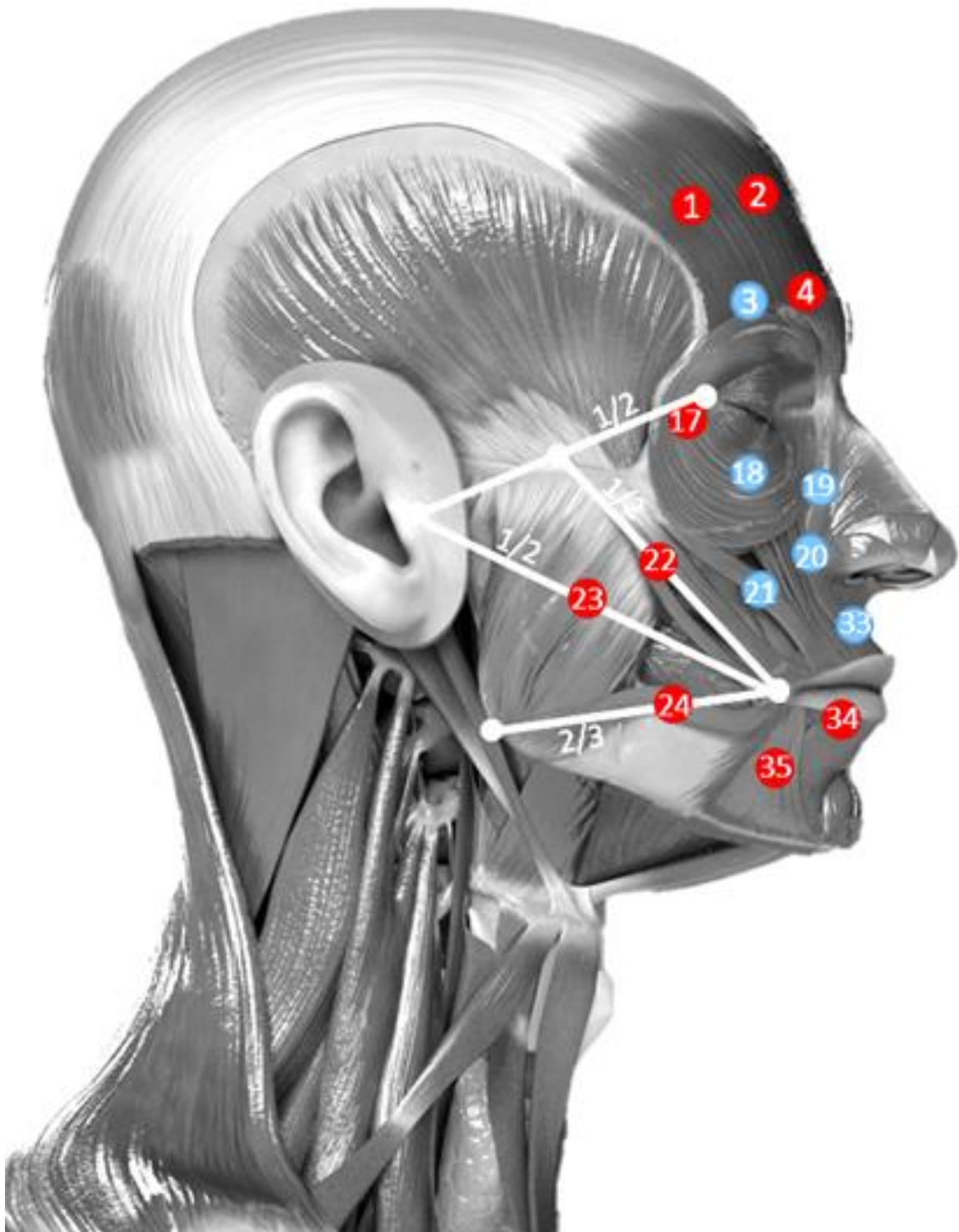
- 13) Besides grouping, the double-sided stickers can already be placed on the electrodes.



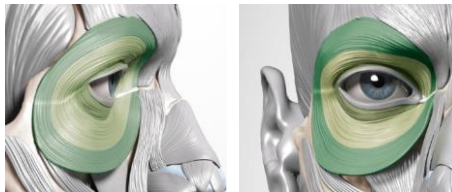



2. Electrode placement

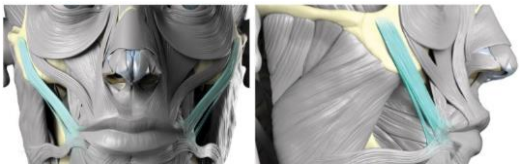


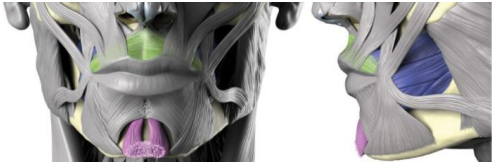
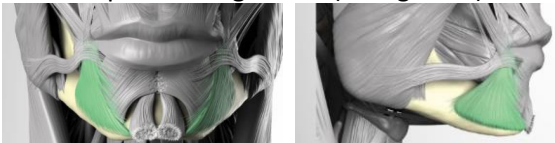
Once the subject arrives, they should be informed about the course of the experiment. They will be asked if they understand everything and they must sign an informed consent form at the beginning of the experiment.

- 1) Prepare the face of the subject by rubbing the cleansing peeling on the skin with a cotton swab. cleaning it with an alcohol wipe.
- 2) Then, clean the face with an alcohol wipe. This might sting a bit.
- 3) Determine electrodes placement based on the information in the figures and tables below. Mark the positions with a fine liner. NOTE: place the mark 0,5 cm above where the center of the electrode should go.
- 4) Once all positions are marked, place the electrodes (35 pcs.) on the skin. Use the knife/spatula to put a small drop of electrode gel in the hole of the electrode, peel of the sticker and stick to the skin. Don't forget to attach the ground and electrodes on the biceps.
- 5) Once all electrodes are in place, check the impedance. To do this, run the script 'ImpedanceTest'. The impedance values for all channels are calculated and saved. The impedance should be <10 kΩ, ideally 5 kΩ. If this is not the case, review your electrodes.



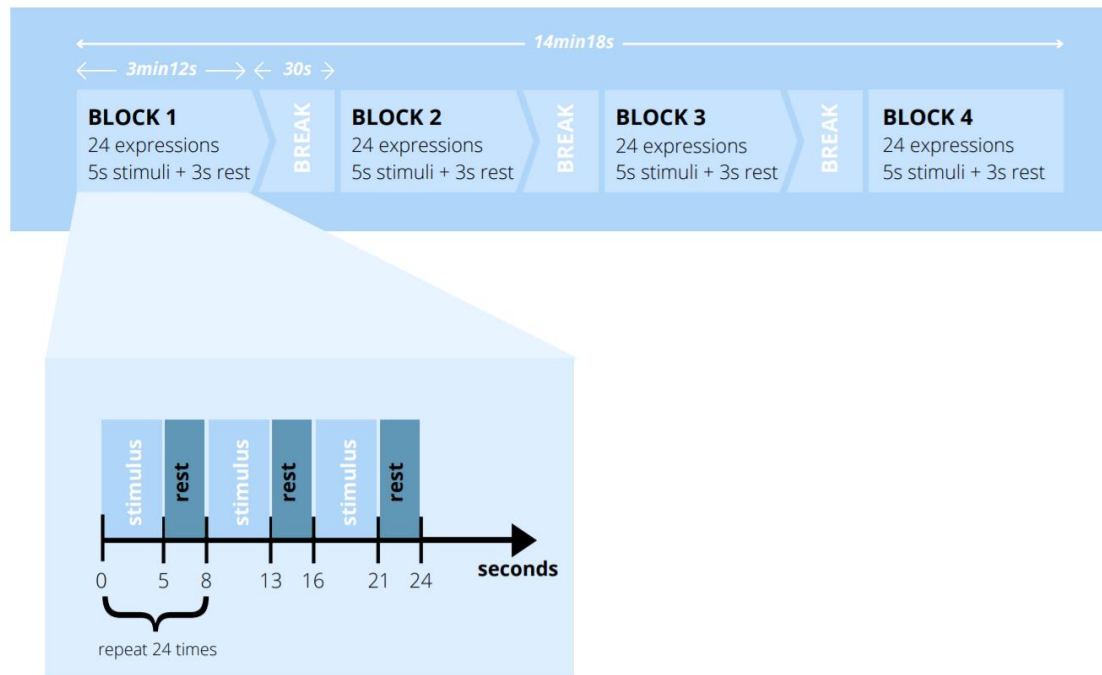


Electrode number	Placement instructions	Muscle(s) involved
FOREHEAD		
1,65	Measure from deepest/lowest point on nasal bone 4 cm upwards onto the forehead. From this point on, go 2,5 cm outwards (both left and right) and place the electrodes.	Frontalis, pars medialis (blue) and pars lateralis (green) 
2,66	Measure 2,5 cm outwards from electrode 1,65, and place the electrode there on both sides of the forehead.	
3,67	Place the electrodes right above the eyebrow, in the middle of electrodes 1-2 and 65-66.	
4,68	Place the electrodes just above the beginning of the eyebrow.	Procerus (purple), Corrugator Supercilii (teal) and Depressor Supercilii (red) 
5	Place the electrode at the deepest point of the nasal bridge, between the inner eye corner and beginning of eyebrow. Ask subject to frown to determine place where electrodes can properly adhere.	
CHEEKS		
17,81	Place the electrodes in the outer corner of both eyes. Locate a bit out and under to avoid wrinkling of the skin when smiling.	Orbicularis oculi 
18,82	Following the curve of the eye socket, place this electrode 2 cm from electrodes 17,81.	
19,83	Measure from the inner eye corner to outer point of nostril, place electrode at 1/2.	Levator labii superioris alaeque nasi 
20,84	Place the electrodes 1 cm outwards from the outer point of the nostrils. Instruct the subject to smile, and ensure electrodes aren't placed in skin crease.	Levator labii superioris 
21,85	From 20,84, go 1 cm outwards, in direction of the angle of the mandible (palpate).	Zygomaticus minor 

22,86	Measure from the outer eye corners to the tragus, mark half. Then, from that mark to the corner of the mouth, mark half and place the electrodes.	<p>Zygomaticus Major</p> 
23,87	Measure from the corner of the mouth to the tragus, place the electrode at 1/2.	<p>Masseter</p> 
24,88	Measure from the corner of the mouth to the corner of the mandible (palpate). Place the electrodes at 1/3 from the corner of the mouth.	<p>Risorius</p> 
MOUTH		
33,97	Place the electrodes between the nose and upperlip, as close to the crease as possible. Instruct subject to smile to determine exact placement.	<p>Orbicularis Oris (green) and Mentalis (purple)</p> 
34,98	Follow the bottom lip and locate the electrodes at 1/3 distance from the corners of the mouth.	
36	Place the electrodes 2-2,5 cm below the electrodes 34,98, in the middle of the chin	
35,99	Locate between the corners of the mouth and the mandible, at 1/2.	<p>Depressor anguli oris (triangularis)</p> 
GROUND ELECTRODE		
ground	Palpate the left collarbone and place a 'regular' electrode (no microelectrode)	

3. Experiment

Now the real experiment can start. The experiment consists of three parts, each taking ± 15 minutes. Four blocks of expressions will be showed during each part, all followed by a 20 second break. See the figure below for a complete time overview.



Part I

- 1) Instruct the subject that they have to view the screen and stay still. In between stimuli they should remain neutral.
- 2) Ensure all scripts are set correct:
 - a. Stimulus presenter: seed and subject ID on the correct number
 - b. Data acquisition: subject ID and part of experiment in line 6.
- 3) First start data acquisition in MATLAB window B. To do this, run the script. (All ranges can be set by 'r', a dialog allows to put in a range. Press 'a' for autoscale).
- 4) Next, start the stimulus presenter by clicking on the 'start' button.
- 5) Once the stimulus presenter is ended, stop the data acquisition by pressing 'q' or the cross.
- 6) IMPORTANT: Go to the results folder of the stimulus presenter and manually set the name of the file to: 'subjectX_partX'. (If you don't change the name, the next file won't be saved).
- 7) Now the subject can have a 10 minute break to relax or stretch their face. The electrodes must remain in place.

Part II

- 8) Instruct the subject that they have to view the screen and mimic the expressions, for as long as the images appear on screen (5s). In between stimuli, they should remain neutral.
- 9) Repeat steps 2-6 from Part I. IMPORTANT: set seed +1.

Part III

- 10) Instruct the subject that they have to view the screen and express the emotions written on screen, for as long as the words appear on screen (5s). In between they should remain neutral.
- 11) Repeat steps 2-6 from Part I.

Once the experiment is finished, all electrodes can be taken off, the face can be cleaned and the subject is free to go.

4. After the experiment

- 1) Remove all stickers from the electrodes and clean all cables and the setup with an alcohol wipe.
- 2) Check if you have six files for this subject:
 1. stimuli overview part I
 2. stimuli overview part II
 3. stimuli overview part III
 4. EMG data part I
 5. EMG data part II
 6. EMG data part III

Appendix B: Information Brochure and Informed Consent

Information brochure

High density EMG to detect and identify facial expression

Dear reader,

In this letter we would like to inform you about the research “High-density EMG to detect and identify facial expression” that you are asked to participate in. Your decision to accept or deny participation should be based on proper information. This letter consists of the aim of the research, details about the procedure, and possible risks. If you decide to participate, you are free to withdraw from the study at any time, without stating a reason. Within 24 hours after the experiment, you can decide that your data may not be used for the research after all, again, without statement of a reason.

At the end of the entire research, you can be informed about the results obtained by means of a debriefing, if desired. Signing this document, you declare that you are a volunteer and will not receive remuneration. Details about the date and location of the experiment will be shared with you separately.

Purpose of the research

Analysis of facial expression has been of great interest in several fields for quite some time. It has applications in marketing, surveillance, entertainment, and healthcare, amongst others. Obvious facial expressions that we all come across from time to time are disgust, fear, joy, surprise, sadness and anger. Facial expressions are very important in non-verbal communication, showing reactions and attention. And even though there are differences in communication between countries/cultures, these six facial expressions are universal. In real-life, often only subtle changes in expression take place. These subtle changes can be difficult to detect and distinguish. Some expressions are so subtle that they cannot be detected with the naked eye. Different techniques have been evaluated that could aid in detection of these expressions. One method with good prospects is EMG based facial expression detection.

In the proposed research, electromyography (EMG) measurements of the face will be recorded as response to presented pictures of facial expressions. These EMG measurements show the muscle activity corresponding with different emotions. As this research is in its initial state, the main aim is to verify the experimental setup and to check the feasibility of EMG to identify facial expressions. The results of this study (the established, tested and validated protocol) can be used to detect the facial expressions in patients in whom this is less evident (e.g. patients with facial paralysis, facial lesion or disorders of consciousness). It is expected to detect and identify the facial expressions in these patients with limited functionality, improving communication.

Procedure

Duration

The experiment will take about two consecutive hours including preparation, familiarization, and performing the experimental task.

Preparation

In order to measure muscle activity, 32 micro electrodes will be placed on your face (see figure 1). To do this properly, the skin must be cleaned, and if needed, shaved. We will clean your face using alcohol and a cleansing peeling. The alcohol might dry out your skin. Be aware of this and, if desired, bring some moisturizing cream and/or make-up to apply afterwards.

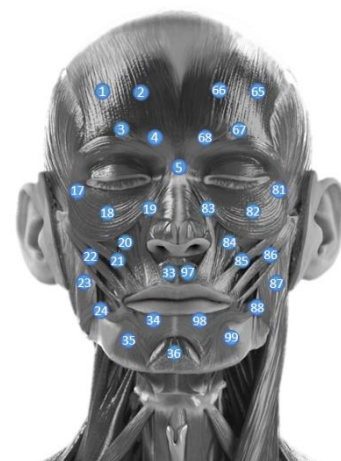


Figure 1: Electrode configuration

Experiment

During the measurements you will be seated in front of a screen. On this screen, several images of faces that display a specific emotion, and names of emotions, will be shown. You are informed to observe and/or mimic the emotions, during which your muscle activity will be measured. Several rounds of visual stimuli and measurements will take place. In between rounds there is time to rest and relax your muscles.

Inclusion/Exclusion criteria

- The participant should be 18 years or older;
- The participant should not have any neuromuscular disorder and/or facial lesion.

Benefits and risks

There are no expected risks. Alcohol is used to clean the face which might dry the skin a bit. Furthermore, all equipment used in this experiment is completely safe. We will be cautious, no physical or mental harm is expected. You will be confronted with a number of pictures to induce a particular emotional state. Be aware that this might provoke unwanted feelings.

There will be no direct benefits for you from participating in this research. On the long run, we hope to be able to detect facial expressions in persons with limited muscle activity, improving communication.

Data collection

Your data will be handled confidentially and will be anonymized right after the experiment. Your name will be in no manner connected to the data. We will never disclose the data to third parties without your permission. Disclosure of data include, but is not limited to, publication in a journal, presentation at a conference, presentation in a colloquium, or discussion in internal consultation. In publication of this research, as MSc thesis and possibly scientific publications, the personal information (e.g. gender, age) will be given on group level, no individual data will be disclosed.

Yours sincerely,

Contact information

Researcher

Veerle Diederiks
University of Twente
+31657752130
[v.l.diederiks@student.
utwente.nl](mailto:v.l.diederiks@student.utwente.nl)

Research leader

Utku Yavuz
University of Twente
Faculty of EEMCS
Horst -Zuidhorst De Horst 2
7522LW Enschede
The Netherlands
Tel: +31(0)534898158
s.u.yavuz@utwente.nl

Ethics Committee member

Dr.ir. J.R. Buitenweg
University of Twente
Faculty of EEMCS
Horst -Zuidhorst De Horst 2
7522LW Enschede
The Netherlands
Tel: +31534892705
j.r.buitenweg@utwente.nl

Informed consent

High density EMG to detect and identify facial expression

- I declare that I've read and understand the information brochure about the research titled: "High density EMG to detect and identify facial expression". I have been able to ask questions and, if applicable, been answered to my satisfaction;
- I declare that I voluntarily participate in the abovementioned experiment. I'm aware that I can withdraw from this study at any time, up to 24 hours after the experiment, without stating any reason;
- I give permission to process my data in abovementioned manner. Data will be anonymized and not shared with any third parties without my consent.

☐ I would like to be informed in the future about the results of this study. I give permission to reach out to me at the following email address:

Participant information

Age: _____

Gender:






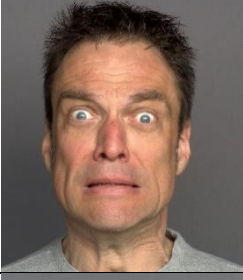






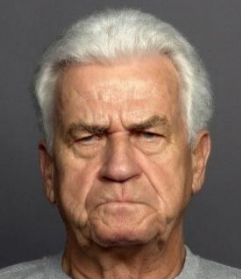

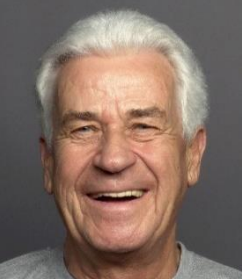
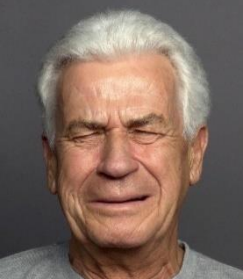

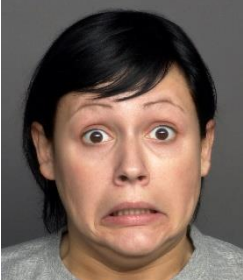






- ☐ Male
- ☐ Female
- ☐ Other: _____

_____	_____	____/____/____
Name participant	Signature	Date

The researcher declares to have properly informed the participant about the research and, to the best his/her ability, ensured that the participant understands to what they are freely consenting.

_____	_____	____/____/____
Name researcher	Signature	Date

Appendix C: FACES database

	Anger	Fear	Happiness	Sadness
Middle-aged: female				
Middle-aged: male				
Old: female				
Old: male				
Young: female				
Young: male				

Appendix D: Results for all classifiers, features and window lengths

In the tables on the following three pages the results for all classifiers (n=29) and features (n=6) are shown for window lengths of 250, 500 and 1000 ms respectively. Results shown is validation accuracy (%).

Window length = 250 ms		Feature →						
Model ↓		MAV	RMS	WL	SD	DAMV	IAV	all
Tree	<i>Fine tree</i>	76.5	75.8	76.0	75.5	76.2	76.0	77.3
	<i>Medium tree</i>	69.4	68.6	68.0	68.2	68.8	69.5	68.9
	<i>Coarse tree</i>	55.5	55.6	56.9	56.1	56.7	55.5	58.4
Discriminant Analysis	<i>Linear discriminant</i>	54.7	54.8	55.8	54.5	56.0	54.6	61.1
	<i>Quadratic discriminant</i>	64.7	64.7	64.3	64.0	64.1	64.3	-*
Naïve Bayes	<i>Gaussian Naïve Bayes</i>	47.5	47.2	47.7	49.0	46.5	46.6	48.2
	<i>Kernel Naïve Bayes</i>	65.3	65.7	62.6	66.0	62.6	65.5	65.5
Support Vector Machine	<i>Linear SVM</i>	64.1	64.1	65.0	63.9	64.8	63.9	65.5
	<i>Quadratic SVM</i>	70.4	70.5	70.7	69.8	70.9	70.4	71.0
	<i>Cubic SVM</i>	72.4	72.1	72.8	71.4	72.6	72.1	72.3
	<i>Fine Gaussian SVM</i>	72.5	72.5	73.4	71.6	73.1	72.5	72.8
	<i>Medium Gaussian SVM</i>	69.3	69.0	69.5	68.9	69.4	69.1	69.5
	<i>Coarse Gaussian SVM</i>	63.7	63.3	63.8	63.5	63.9	63.8	64.1
K-nearest neighbors	<i>Fine KNN</i>	78.1	77.9	79.1	77.2	79.1	78.2	78.4
	<i>Medium KNN</i>	74.3	74.1	74.7	73.6	74.4	74.4	74.6
	<i>Coarse KNN</i>	66.6	66.4	66.7	66.5	66.5	66.3	66.4
	<i>Cosine KNN</i>	74.4	74.2	75.2	73.5	74.7	74.6	74.5
	<i>Cubic KNN</i>	73.9	73.8	74.4	72.9	74.1	73.9	74.1
	<i>Weighted KNN</i>	77.1	76.7	77.8	75.9	77.6	77.0	77.2
Ensemble	<i>Boosted Trees</i>	74.1	74.3	74.4	74.5	74.2	74.3	75.5
	<i>Bagged Trees</i>	90.2	90.1	91.5	89.5	91.0	90.1	90.9
	<i>Subspace Discriminant</i>	69.0	68.4	69.7	68.0	69.4	68.2	63.5
	<i>Subspace KNN</i>	92.8	92.4	93.6	91.7	93.6	92.8	87.2
	<i>RUSBoosted Trees</i>	69.4	68.6	68.0	68.2	68.8	69.5	68.9
Neural Network	<i>Narrow NN</i>	68.4	68.2	69.6	67.8	69.2	68.8	69.7
	<i>Medium NN</i>	71.2	70.8	71.6	70.5	71.8	71.2	71.5
	<i>Wide NN</i>	73.8	73.5	73.9	72.8	74.2	73.9	73.7
	<i>Bilayered NN</i>	68.8	68.9	69.7	68.2	69.4	68.7	70.1
	<i>Trilayered NN</i>	69.0	69.3	69.8	68.7	69.2	68.7	69.6

* failed → error: one or more of the classes have singular covariance matrices for their predictor values. This makes training fail. Try the "Covariance structure: Diagonal" advanced option.

Window length = 500 ms		Feature →						
Model ↓		MAV	RMS	WL	SD	DAMV	IAV	all
Tree	<i>Fine tree</i>	75.1	75.6	76.2	75.5	76.4	75.6	76.7
	<i>Medium tree</i>	68.6	68.3	67.6	69.2	68.1	69.1	69.4
	<i>Coarse tree</i>	54.8	55.5	57.3	55.7	57.6	55.4	58.2
Discriminant Analysis	<i>Linear discriminant</i>	55.6	55.7	56.7	55.3	56.9	55.7	62.5
	<i>Quadratic discriminant</i>	65.4	65.4	65.0	64.8	65.2	65.5	-*
Naïve Bayes	<i>Gaussian Naïve Bayes</i>	47.7	46.2	48.0	48.8	46.1	47.3	47.4
	<i>Kernel Naïve Bayes</i>	65.7	66.1	61.4	66.8	62.0	65.6	65.9
Support Vector Machine	<i>Linear SVM</i>	64.4	64.5	65.2	64.4	64.7	64.5	65.6
	<i>Quadratic SVM</i>	69.5	69.5	70.0	68.9	70.2	69.2	69.9
	<i>Cubic SVM</i>	70.8	70.7	71.1	70.3	71.0	70.8	70.8
	<i>Fine Gaussian SVM</i>	70.8	70.7	71.2	70.1	71.1	70.6	71.0
	<i>Medium Gaussian SVM</i>	68.0	68.1	68.0	68.2	68.5	67.8	68.2
	<i>Coarse Gaussian SVM</i>	61.8	61.9	61.7	61.6	61.9	61.4	62.0
K-nearest neighbors	<i>Fine KNN</i>	75.9	75.5	76.4	75.2	76.8	76.1	76.5
	<i>Medium KNN</i>	70.2	70.0	70.4	69.6	70.5	70.1	70.3
	<i>Coarse KNN</i>	63.8	63.9	64.1	63.5	64.1	64.0	64.1
	<i>Cosine KNN</i>	69.9	69.9	70.2	69.8	70.4	69.9	70.2
	<i>Cubic KNN</i>	69.6	69.9	70.0	69.1	70.1	69.7	70.0
	<i>Weighted KNN</i>	73.9	74.1	74.5	72.9	74.9	73.6	74.1
Ensemble	<i>Boosted Trees</i>	75.3	75.0	74.1	74.6	73.9	75.3	75.8
	<i>Bagged Trees</i>	87.5	87.9	88.1	86.8	87.9	87.6	87.8
	<i>Subspace Discriminant</i>	69.1	69.1	70.3	69.1	70.1	69.5	64.7
	<i>Subspace KNN</i>	89.9	89.4	90.1	89.1	90.4	89.7	86.9
	<i>RUSBoosted Trees</i>	68.6	68.3	67.6	69.2	68.0	69.1	69.4
Neural Network	<i>Narrow NN</i>	68.1	68.8	68.2	68.0	69.0	68.8	68.8
	<i>Medium NN</i>	70.4	70.1	71.2	69.6	71.0	70.3	70.6
	<i>Wide NN</i>	72.5	72.2	72.6	71.8	72.9	71.5	72.8
	<i>Bilayered NN</i>	68.2	68.6	68.8	68.1	68.6	68.1	69.5
	<i>Trilayered NN</i>	68.5	68.5	68.8	67.9	68.0	68.2	69.1

Window length = 1000 ms		Feature →						
Model ↓		MAV	RMS	WL	SD	DAMV	IAV	all
Tree	<i>Fine tree</i>	74.2	74.9	75.3	74.8	74.7	75.3	75.8
	<i>Medium tree</i>	69.8	69.4	68.3	69.5	67.7	70.6	70.1
	<i>Coarse tree</i>	55.2	56.7	55.7	56.6	56.1	55.2	58.8
Discriminant Analysis	<i>Linear discriminant</i>	56.3	56.6	57.4	56.0	57.7	55.9	64.0
	<i>Quadratic discriminant</i>	66.3	66.4	66.1	66.5	66.1	65.7	~*
Naïve Bayes	<i>Gaussian Naïve Bayes</i>	48.4	46.7	44.7	47.3	47.3	45.4	45.9
	<i>Kernel Naïve Bayes</i>	65.1	66.2	61.1	66.6	60.8	65.5	64.8
Support Vector Machine	<i>Linear SVM</i>	64.4	64.5	64.7	64.7	65.0	64.5	65.3
	<i>Quadratic SVM</i>	68.3	68.3	68.6	68.0	69.1	69.2	68.2
	<i>Cubic SVM</i>	68.2	68.0	69.1	67.9	69.3	68.9	68.8
	<i>Fine Gaussian SVM</i>	68.1	68.2	68.0	67.3	68.3	68.1	67.9
	<i>Medium Gaussian SVM</i>	66.8	67.0	67.0	67.1	67.6	67.3	67.0
	<i>Coarse Gaussian SVM</i>	58.9	59.4	59.0	59.2	59.1	59.2	59.2
K-nearest neighbors	<i>Fine KNN</i>	70.4	70.3	70.8	70.1	71.3	70.4	70.3
	<i>Medium KNN</i>	66.9	67.6	67.9	67.1	67.8	67.6	67.8
	<i>Coarse KNN</i>	61.3	61.4	61.0	61.7	60.9	61.1	61.5
	<i>Cosine KNN</i>	66.9	66.8	67.3	66.5	67.6	67.3	67.2
	<i>Cubic KNN</i>	67.4	67.4	67.7	66.6	67.7	67.8	68.0
	<i>Weighted KNN</i>	69.3	70.0	69.9	69.7	69.9	70.0	69.0
Ensemble	<i>Boosted Trees</i>	73.8	75.1	73.6	75.0	73.6	74.3	75.4
	<i>Bagged Trees</i>	82.5	83.6	82.6	82.3	82.8	83.6	82.8
	<i>Subspace Discriminant</i>	69.7	69.7	70.8	69.9	70.6	69.7	66.2
	<i>Subspace KNN</i>	84.0	84.0	84.1	84.4	84.8	84.5	72.3
	<i>RUSBoosted Trees</i>	69.8	69.3	68.3	69.5	67.7	70.6	69.8
Neural Network	<i>Narrow NN</i>	67.7	68.0	68.7	66.0	68.1	66.9	67.6
	<i>Medium NN</i>	68.7	68.5	68.3	67.8	68.4	68.4	68.4
	<i>Wide NN</i>	69.4	69.1	69.1	69.2	69.5	69.6	69.1
	<i>Bilayered NN</i>	66.7	68.3	68.0	67.2	67.7	66.8	67.7
	<i>Trilayered NN</i>	67.5	68.0	67.3	67.0	68.5	67.1	68.3

Appendix E: Optimization

For possible increase of performance, combinations of the features have been evaluated. Results show that accuracy did not improve significantly.

		WL+DAMV	WL+DAMV+MAV	WL+MAV	MAV+RMS
Ensemble	<i>Bagged Trees</i>	91.3%	91.2%	91.0%	90.1%
	<i>Subspace KNN</i>	93.4%	93.3%	93.5%	92.7%

Appendix F: Channel subsets

Channels a(b) with a and b being symmetrical channels	Ensemble: subspace KNN		Ensemble: bagged Trees	
	DAMV	WL	DAMV	WL
+4(21), 11(27)	60.0%	60.3%	71.8%	71.8%
+1(18)	69.0%	69.1%	84.2%	84.3%
+2(19)	86.6%	86.2%	86.2%	86.8%
+3(20)	86.8%	86.4%	87.4%	87.4%
-1(18)	84.7%	85.1%	85.2%	84.9%
+17, +1(18)	90.2%	91.1%	88.9%	88.6%
+5	90.2%	91.2%	88.8%	89.2%
+7(23), -5	90.8%	91.3%	89.4%	89.2%
+16(32)	91.4%	91.4%	89.9%	89.8%
+8(24)	92.2%	91.8%	90.1%	89.9%
-3(20)	91.8%	90.7%	90.3%	90.0%

Appendix G: Speed measures

The table below shows the prediction speed (in classified objects per second) and training time (in seconds) for the four models: Subspace KNN with feature WL, Subspace KNN with feature DAMV, Bagged Trees with feature WL and Bagged Trees with feature DAMV. Bagged Trees operated significantly faster.

	Subspace KNN		Bagged Trees	
	WL	DAMV	WL	DAMV
Prediction speed	686 obs/sec	838 obs/sec	29600 obs/sec	31600 obs/sec
Training time	96.691 sec	77.18 sec	15.6129 sec	14.400 sec