École Polytechnique Fédérale de Lausanne

MASTER PROJECT IN BIOENGINEERING

Carried out in Gabriel Kreiman Laboratory at Boston, Harvard Medical School.

---

24 hours in the human brain

Authors:
Eleonora Iaselli

Supervisors:
Prof. Gabriel Kreiman
Associate Professor Boston Children's Hospital,
Harvard Medical school

Prof. Michael Herzog
Laboratory of psychophysics (LPSY), EPFL

Date:
March 2, 2018

Acknowledgements

I would like to thank, first among all, Prof.Gabriel Kreiman from Boston Children's Hospital, Harvard Medical School, who granted me the incredible opportunity to work in his laboratory and collaborate with all his colleagues and students. Taking part to lab meetings and journal clubs allowed me to understand the dynamics of the laboratories in the world of research and all the fascinating questions that can be asked about the functioning of our brain. Moreover, weekly discussions with him allowed me to acquire interesting and helpful knowledge to apply to continue the analysis of my project.

I would also like to thank my supervisor at the EPFL, Prof.Michael Herzog to accept to evaluate my work of these incredible six months.

Finally, I would like to thank the collaborators who helped me in the extensive and tedious work of annotating the subject's behaviors and my family, which has always been by my side during not only my stay in Boston, but through the entire adventure of University,the friends back home and the new friends found here that never made me miss home.

To all of you, thank you!
Eleonora Iaselli

Contents

1 Introduction

The most illuminating way to obtain information about the operation of the brain is through neural recordings. In fact, the brain is composed by cells, whose way of action is mostly dependent on electrical currents and signals that interconnect the neurons creating large and extensive networks. This circuitry allows to spread information in different parts of the brain which are dedicated and involved in various behaviors. Although the general structure of the brain is fixed for every species with well defined regions and specific cell types, the exponential possibilities of interconnected systems allows us to develop our own personality and life experiences.

In order to relate the external world and the brain, in 1860 Gustav Theodor Fechner coined the term psychophysics to identify the analysis of physical stimuli and perceptions[1].
From 1860, a wide number of experiences were conducted to analyze the relationship between the external phenomenon (provided by the world) and the subject reaction (e.g. movement of the eyes, perception of an object, neural recordings). The main aim of all these experiences was to obtain a deeper knowledge about the pattern of activation of the neurons in a specific situation, in order to visualize the similarities and differences among people.

Moreover, the advances in electronics, computer science and functional imaging provided the tools to analyze the brain machinery in details. As an example, electroencephalograms (EEG) and magnetoencephalograms (MEG) have been extensively used in the last century to characterize the electric and magnetic fields emerging from neural currents within the brain. These techniques were followed by more sophisticated and invasive procedures as the electrocorticography (ECoG) and Deep Brain Stimulation (DBS)[2].

The information that can be collected through these techniques can be used to improve living conditions of patients, establish innovative forms of rehabilitation and for the pure love of knowledge. Brain machine interfaces (BCI) permit to create an intermediate system that associates the patient's intentions to a specific activity. This engineering approach used to solve every-day issues is of fundamental importance, as an example in the case of spinal injuries or neurodegenerative diseases [3] [4].
The aim of this project is to investigate the neural circuits through computer modeling, neurophysiological recordings and psychophysical measurements to better understand the phenomenons involved in perception and cognition.

1.1 ECoG and epilepsy

As already mentioned, neuronal communication is based on electrical currents formed in the brain. All active cellular processes in a given volume of the brain superimpose at a given location in the extracellular medium and generate a potential, V_e, with respect to a reference potential [5] . The electric field produced by the difference between V_e in two different locations can be recorded by extracellularly placed electrodes. EEG, ECoG and LFP are three techniques that record these signals, although they differ according to the invasiveness of the electrodes. In fact, EEG is recorded from the scalp, ECoG is obtained using subdural grid electrodes on the cortical surface and LFPs are accessed through intracranial recordings. Since EEG is separated by several layers of tissue (the meninges, blood, bone, and skin) from the signal, its bandwidth is limited to 50 Hz and suffers spatial blurring in a bigger extent compared to ECoG [6]. Figure 1 shows the positions of ECoG and EEG electrodes relative to the brain layers.

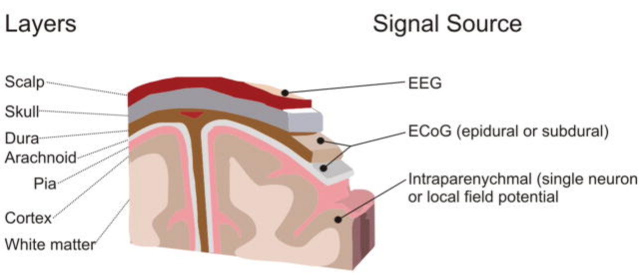


Figure 1: Brain layers and localization of the electrodes for EEG and ECoG. From [?]
.

All the techniques have the objective of recording extracellular signals in the brain.

ECoG has been and it is still commonly used in epileptology to define the localization of the epileptogenic zone

(EZ). Epilepsy is a neurologic chronic condition characterized by episodes of abnormal excessive or synchronous neuronal activity in the brain[7]. Intractable epilepsy is usually treated by surgery when seizures arise from a single resectable epileptogenic zone[8]. When the candidates for this type of surgery are individuated, they undergo a surgery to implant invasive electrodes in order to define the location of the seizure onset (EZ). Once the zone is localized, the surgeon proceeds removing it.

1.2 Machine learning

The rise of artificial intelligence (AI) that took place in the 1950s, was possible because of three generations of researches who realized that computers were not only capable of solving basic mathematics problems, but they had a higher and more sophisticated potential [9].
Machine learning is a fundamental concept of AI research since it is the study of computer algorithms that improve automatically through experience [10]. In fact, the field of machine learning consists in the ability to create automated systems capable of extracting meaningful information from a large dataset and learn from it [11]. The knowledge acquired could be consequently used to discern a class from another in a set of new instances. The two processes of learning and application are called data classification [12].

In the aim of this project, the field of machine learning investigated involves that the real classification output is known, therefore the learning process is called supervised [13].
On the other hand when instances are not labeled, the techniques investigated are called unsupervised. Compared to the unsupervised methodologies, the supervised techniques provide a powerful method to evaluate the performance of a classifier since it is possible to calculate the percentage of the correct predictions. Indeed, the accuracy prediction of a classifier is essential to predict its future reliability and to choose the best model from a given set [14]. One of the most used methods to evaluate classifier accuracies is called k fold cross validation. It is a statistical method which use part of the data to learn the model and another part to test them. The dataset is split into k mutually exclusive subsets of approximately the same size. The classifier is therefore trained on k-1 folds and tested on the remaining subset. This procedure is repeated until when all the folds have been used as the testing sample. The classifier accuracy is estimated averaging the various performances. A schematic of the procedure of cross-validation is shown in figure 2.

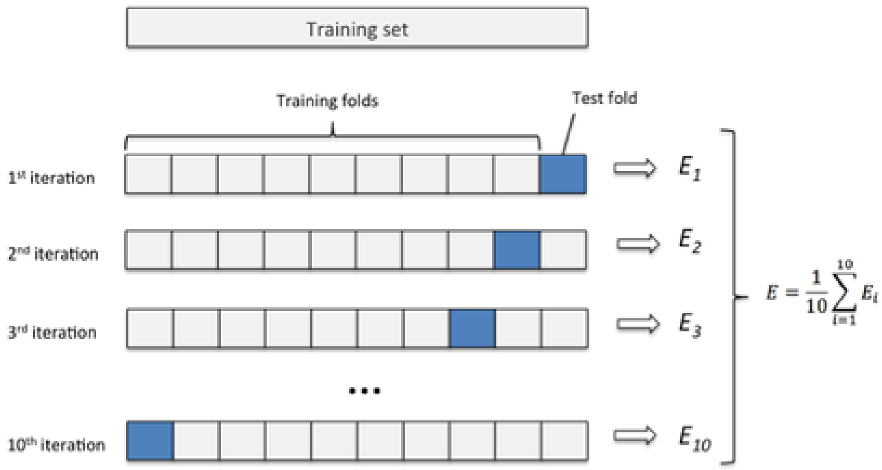
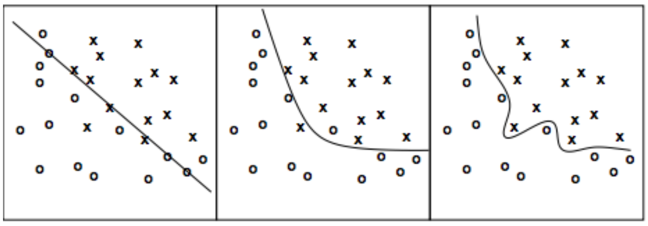


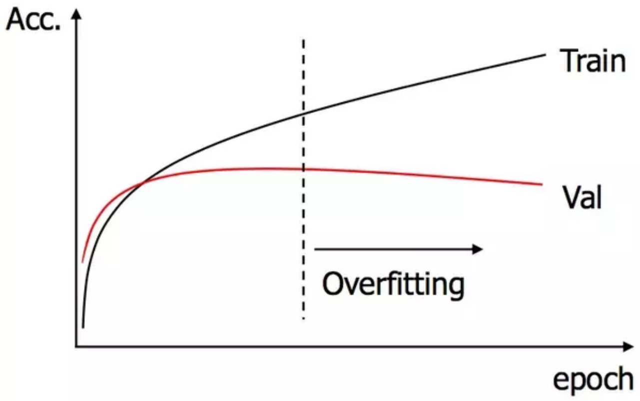
Figure 2: Schematic showing the process of k fold cross-validation, in this image k =10. The entire dataset is split into 10 subsets of data. The classifier is trained iteratively and independently with k-1 folds and tested on the remaining validation fold. The final error is computed as the average over the ten iterations.

The main reason why cross-validation is essential for a stable prediction of the reliability of the classification method is that using all the available data to train a model could cause a phenomenon called overfitting. The outcome of this situation is that the classifier will perform very well on the data it was trained on, but it will fail in the prediction of future unseen data. Overfitting can be explained as a result of including more terms than necessary in the prediction model, creating therefore a model more flexible than what it is required to be. This will add complexity compared to a simple linear classifier without entailing benefits in the performance or even causing a decrease in the accuracy[15].

A visualization of overfitting can be observed in figure 3. In figure 3a they are shown the various ways of classification for a linear, a quadratic and an overfitted classifier. The linear discriminant analysis (LDA) is a fundamental method in data analysis which tries to reduce the dimensionality of the original data points by solving a generalized eigenvalue system[16].
Essentially, data samples of different classes are projected in a lower dimension space and separability is defined according to statistical measures of mean value and variance. Fisher's intuition, when he proposed the model, was to find an hyperplane that could maximize the distance between the two means ans at the same time to minimize the variance in each class [?].LDA can also be extended to non-LDA, which implies that the disjoint regions of the measurement space corresponding to each class are assigned by quadratic boundaries [17].



(a) Visualization of the process of overfitting and comparison with linear and quadratic classifiers.



(b) Effect of overfitting on classifier accuracies for training and testing subsets of data. When the model is overfitted, it will seek to represent very specifically the training data and will have a lower performance in discriminating new instances.

Figure 3: Schematics and visualization of the phenomenon of overfitting and its consequences.

1.3 State of the art on neuronal engineering

The brain is the most mysterious and complex organ of humans and primates. In the last century the scientific community increased its interest in the analysis of the operation of this structure, since it could provide interesting and extremely useful tools for the development of interfaces between the brain and the computer (BCI), but also in the treatment of neurological disorders as strokes and epilepsy.
In fact, deepening the knowledge about how the neurons interact and their space-time patterns of activation or as well as the portion of the brain involved in a particular lesion or disease could provide essential information in how the brain works. Knowing the functioning of the brain is one of the most exciting and challenging researches that could potentially revolutionize our daily life. Indeed, modeling the brain and its components would impact the entire scientific community and theories about specific mechanisms could be pre-tested in this digital brain.

1.3.1 BCI

Brain computer interfaces are an example of how brain signal recordings and analysis can have a direct impact on human life. In fact, signal from brain activity can be used to communicate the user's intent to other people or to a device capable of performing this intent. Over the past decades, many laboratories have started to explore this technology whose main aim is to convert electrophysiological inputs from the user into outputs that can control external devices.
This is particularly interesting when damages in the peripheral nerves and muscles are present. EEG is usually the most used method to construct BCIs, because it is the less invasive machinery, although ECoG and single-unit activity recorded from within cortex would provide higher signal-to-noise ratios. EcoG seems to be the good compromise to obtain recordings less prone to artifacts and with higher spatio-temporal resolution, but that does not require the implantation of cortical electrodes.

In particular, ECoG signals have been successfully used to provide information about arm movement direction, individual finger movements and even continuous arm movement trajectories[18].
In 2012, human ECoG signals were used to distinguish between two types of grasp movements using single-trial ECoG recordings from the human motor cortex. Decoding the various possibility of grasping would require the introduction of a very high number of degrees of freedom and very complex patterns of inter-finger coordinations. Therefore, T. Pistohl et al. took into account the synergies which were present during hand movements to avoid to control all the joints separately. As a result, the authors discovered that low (2-6 Hz) and high (54-114 Hz) frequencies bands resulted more rigorous in the detection accuracy (oscillated between 70 % and 80%), compared to intermediate frequency bands (14-46 Hz).

Brain computer interfaces have been incrementally acquiring attraction from both the academic and the industrial researches because of the incredible opportunities they could generate. Nevertheless, the constrained experimental paradigms of the performance and decoding of controlled movements under strict conditions limits the application of these methodologies to natural behaviors and consequently to real life applications[19].

In consequence it seems extremely important to develop robust decoding algorithms that can face these challenges. One of the first investigations in this area was performed in 2006 with the aim of discovering the relationship between motor cortical activity and movements in monkeys not during repetitive and trained tasks.[20]A simple neuroprosthetic device was conceived to record neural brain activity of single neurons in primary motor cortex for several weeks at a time, with the intent of converting neural brain activity in one area to electrical stimuli to deliver in another region. This device was used to investigate the correlation between cortical activity and simultaneously-recorded electromyogram (EMG) activity of the arm during extended periods of free behavior in two monkeys. The final outcome of the study confirmed robust correlations between single motor cortex neurons and individual arm muscles also during unstrained behavior.

These kind of analysis are even more rare when considering humans. In fact, it is difficult and more problematic to obtain neural recordings from humans rather than with monkeys and the constraints that could be imposed to patients are obviously subordinated to human necessities. For all these reasons, it would be very convenient to obtain a decoding algorithm based on non-experimental tasks, but simultaneously it is very challenging to obtain the data to conceive them.
Unfortunately, the easiest way to have information about the real life of the subjects is to investigate the behaviors performed using manual labeling based on video and audio [18] [21] [22].

Only one study was performed not using manual labeled annotations and it is described in details in the following paragraph.

1.3.2 Wang and Olson

In this study of Wang et al. [23], long-term electrocorticography (ECoG) data were recorded over many days with simultaneous audio and video recordings. The same procedure was also used in the current research, the main difference residing in how the labels of the behaviors were achieved. In the current research, as already investigated from other studies [18] [21] [22], the ground truth was obtained by manual observations and subsequent annotations using a Graphical User Interface (GUI). On the other hand, Wang et al. used computer vision, speech processing, and machine learning techniques to automatically determine the ground truth labels for the subjects' activities. The main behavioral categories investigated were movement, speech and rest. The main limitation of this method is therefore the impossibility to examine more specific activities, as different kind of movements or actuated by distinct parts of the body. Moreover, the same authors underline the lack of specificity in the detection algorithm when other people enter the frame of the camera and most of all the impossibility to distinguish who is the speaker detected by the speech recognition algorithm.

Another difference with the current research is that in this article the main aim of the authors was to correlate unsupervised decoding of neural recordings performed by hierarchical clustering of power spectral features to the automatic algorithm of computer vision capable of distinguish the three classes of activities. In fact, ECoG activity was first analyzed by unsupervised methods to identify hierarchical k-means clustering on the first day of recordings and they were tested on 3 subsequent days whose videos and audios were not available. The obtained F1 scores for the detection of the behaviors were promising and encourage other researches in this field.

Considering all these stimulating studies which started only recently to consider the importance of natural behaviors compared to perfectly controlled experimental situations, we tried to add our contribution by analyzing 24 hours in the life of a human focusing on a few actions that were repetitively executed.

2 Materials and methods

2.1 Patient characteristics

The patient investigated was a 10 years old child diagnosed with untreatable epilepsy. Figure 4 shows the localization of the electrodes implanted prior to the surgery aimed to locate the seizure foci. The implanted micro depth electrodes were provided with platinum contacts and their diameter size was 1.3 mm. The total number of electrodes implanted was 104, which recorded at a sampling rate of 500 Hz.

The procedure and the implantation of the electrodes required the patient to remain in controlled conditions for one week before undergoing the surgery, which allowed to obtain 24 hours video recordings of his life. In fact, during this period, the subject was attached to an electric generator and consequently his range of movements was constrained. Therefore, subject's life and behavior were documented on both the neurophysiology side, investigating the neural recordings from his brain, and the action side, analyzing the activities performed. This information combined could provide relevant knowledge about how the brain works and contribute to correlate every day activities with neurons' activation.

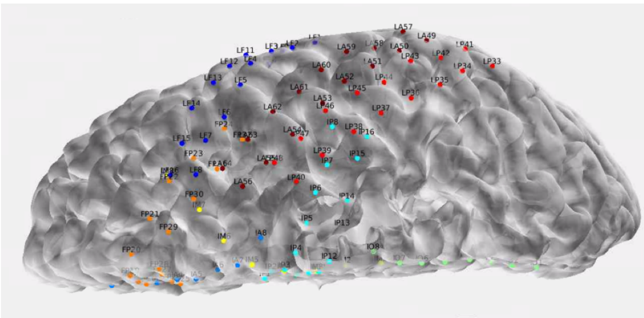


Figure 4: Position of the electrodes implanted on the surface of patient m00043's brain.

2.2 Annotations

As already mentioned in the previous paragraph, videos of the subject's life were accessible. The first task was to determine a set of behaviors which were not only interesting to investigate, but also sufficiently repeated during the day. The following behaviors were considered:

1. **Body movement**: the subject is performing any kind of movement that involves his body, but not his head or his face;
2. **Head movement**: the subject is performing movements of the neck and consequently of his head. Facial movements or chewing were not considered in this class;
3. **Eating**: the subject is eating, which means he is actively chewing or he is drinking;
4. **Video games**: the subject is involved in activities with the use of a computer. Sometimes these activities include psychophysics tasks for further experiments;
5. **Watching TV**: the patient is watching television and paying attention to it.
6. **Someone is talking**: this behavior does not require active participation of the subject, it only requires people talking into the same room;
7. **Patient is talking**;
8. **Physical contact**: any kind of physical contact with the subject;
9. **Sleep** : the patient is sleeping.

Once the behaviors to annotate were selected, a group of 4 people was charged to define the beginning and the end of the above-mentioned activities using a GUI interface as shown in figure 5. The neural recordings and the video were both visible using the Natus® NeuroWorks® software, allowing to annotate using the GUI on the right. NeuroWorks ® collects, retrieve and review neurological data while managing the studies with an intuitive user interface. Every behavior was associated with a button ON and a button OFF that were clicked respectively at the beginning and at the end of the activity. The elapsed time (reference of time starting from the beginning of the recording) and the relative behavior dynamics were stored in a text file.

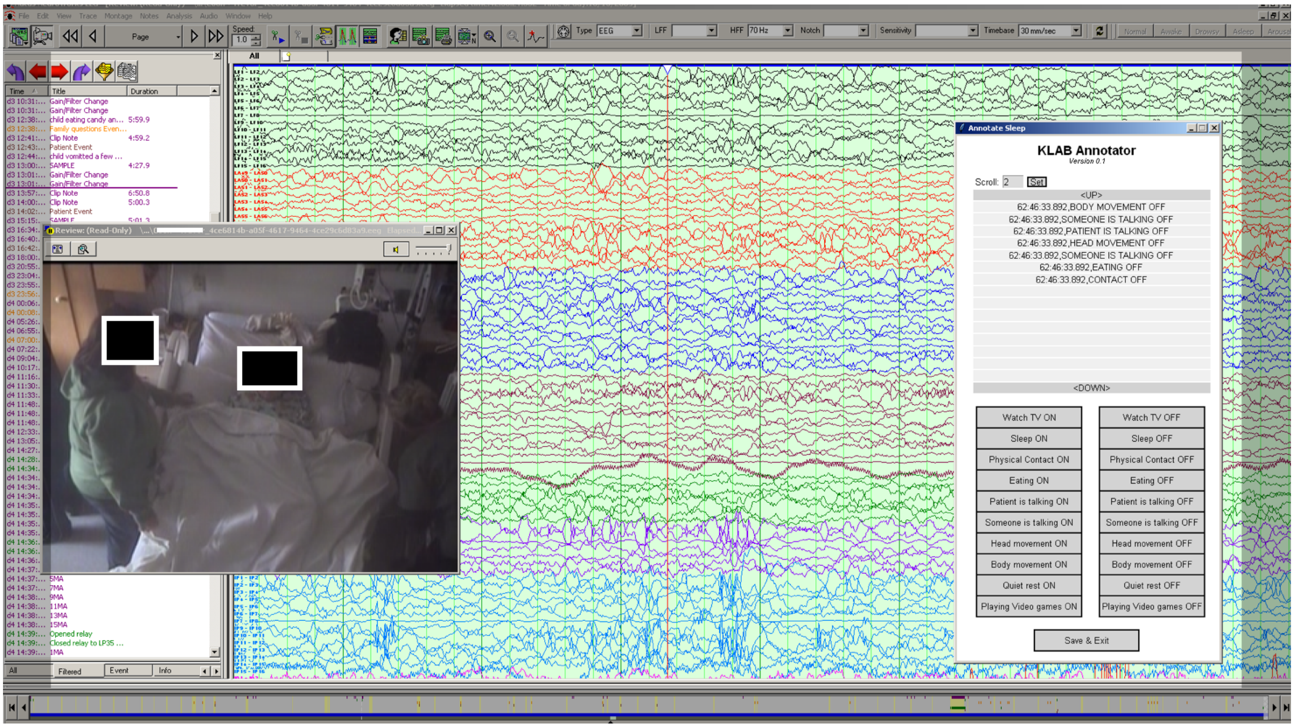


Figure 5: Figure showing the GUI interface (on the right) and the software Natus ® NeuroWorks ® used to annotate the time dynamics of every behavior. The main window in green displays the signals recorded by the electrodes and the video on the left allows to annotate the subject's real life. The beginning and the end of the behaviors discussed were monitored using the GUI interface, which allowed to store information about the activity performed and its relative time fluctuations.

The resolution of the elapsed time was at the ms level (format elapsed time was "hh:mm:ss.fff"). The text file was used to align the neuronal recordings and the behavioral annotations. In fact, the software Natus ® NeuroWorks ® associated every value of the elapsed time to a number, the ensemble of these numbers not being linear but containing some breaks. These breaks represented when the neuronal signals were not recorded or available. All this information was used to correctly align the behaviors and the electrodes' recordings on the time-scale.

2.3 Data extraction and labeling

Neuronal recordings were organized in a matlab® structure which contained the signals (voltage over time in μV x 104 channels) and reference values necessary to associate the index of the dataset and the elapsed time. The signals were Notch filtered at 60 Hz and harmonics (removal of frequencies multiple of 60 Hz) to eliminate the noise derived from the oscillations of alternating current. Moreover every channel was normalized subtracting the respective mean and dividing by the associated standard deviation. The labels for every behavior were extracted in a binary form (1 when the behavior was ON and 0 when the behavior was OFF) and aligned to the electrode signals. Nevertheless, there were a few breaks in the recording that accounted for approximately six minutes of the entire day that had to be taken into account for a correct alignment. In fact, the above-mentioned indexes were not linear and presented discontinuities when the neural data were not available. On the other hand, the GUI interface does not have information about the neural recordings availability, therefore annotations in those intervals had to be discarded. Moreover, an additional 2% of the final data were not available.

These data were further used for classification purposes. Although the labeling was executed at the ms resolution because of the elapsed-time time scale, it was necessary to take into account the human factor to investigate accuracy. In fact, in spite of the incredible speed of processing of our brain, we could not retain possible to be able to recognize the beginning of the behavior and the subsequent activity of annotating it at the same resolution of the sampling frequency. For this reason, we chose a round window of one second to extract feature from the data. The following four different features were considered:

- Root mean square (RMS)
$$RMS = \sqrt{\frac{1}{n}(x_1^2 + x_2^2 + x_3^2 + ...x_n^2)} \tag{1}$$
- Mean absolute value (MAV)
$$MAV = \frac{\sum_{i=1}^{n} |x_i|}{n} \tag{2}$$
- MinAmp: minimal value of the signal
$$MinAmp = min(x_1, x_2, x_3, ...x_n) \tag{3}$$
- MaxAmp: maximal value of the signal
$$MaxAmp = max(x_1, x_2, x_3, ...x_n) \tag{4}$$

where $x_1, x_2, ..x_n$ were the data sample of the signal recorded from the electrodes in one second.

As a result, we obtained one value of each of the features for every second and every channel. One matrix for every feature was therefore computed whose dimensions were driven by the number of seconds and the number of electrodes.
Not all the calculated values were considered to train the classifier. In fact, the resolution of the annotations was at the ms level and the feature calculation at the second level. Consequently, a new vector containing the labels was created to respect the new time scale. This implied that some feature points represented intervals overlapping the zones of the beginning and the end of the behavior. Moreover, we already mentioned that we decided to take into account the human factor and its undetermined loss of accuracy. For all these reasons, the data calculated in these poorly defined periods were not used. Only full seconds of activity ON or activity OFF were investigated and provided to the classifier. Additionally, in order to avoid non-uniform distributions for on and off classes, the feature matrix used to train the classifiers included the same number of cases for the active and inactive behavior.

2.4 Division in three different modalities

For all the behaviors, the dataset was unbalanced because the data informing about the behavior were less frequent. For this reason, the data belonging to class "0" (implying that the behavior was OFF and not performed) were undersampled to produce the same number of instances for both classes. Three different modalities were considered for classification. The main difference in these procedures involved the determination of the negatively labeled examples, which were chosen according to different time scales. Figure 6 shows the schematics of how these operations were constructed.

In the first modality, referred in the following as methodology A, the negative cases were randomly chosen into the time interval of the entire 24 hours, which involved that they could also all be located at a specific timing when artifacts were present. This would lead to an imprecise classification caused by the specific pattern of the negative labeled examples. Therefore, in the case of non-stationarities in the data, this methodology could arise doubts in trusting the performance accuracy of the classifier.
For all these reasons, another methodology was investigated, that we defined methodology B. In this case, the negative samples of the behavior were chosen randomly in a defined interval close to the behavior itself. More clearly, for every sample "ON", a sample "OFF" was chosen, if localized in an interval of 1 minute before or after the sample itself. If it was not found, the sample ON was discarded. Furthermore, every "OFF" sample had to be unique and consequently represented only once in the dataset.

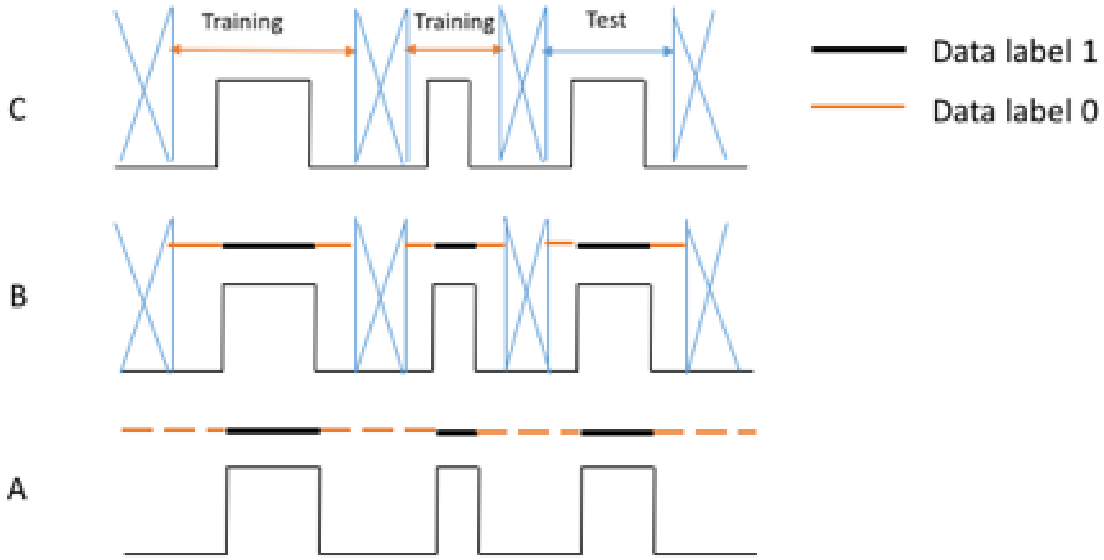


Figure 6: Schematics of the data used for methodologies A, B and C. The binary signal described when the behavior was performed (label "1")and when not (label "0"). For all the methodologies, all the data labeled as "1" were selected and provided to the classifier. For methodology A the negative samples were randomly chosen in the 24 hours recorded; for methodology B only class "0" data which were close in time to the positive events were selected; for methodology C, the data were the same considered for methodology B, but the training and testing set were forced to be at different time periods of the day.

Moreover, this methodology was partition in two different modalities, according to whether the choice of the new dataset was performed before or after the calculation of the features. We will define the two flow processes as methodology B1 and B2.
The data for methodology B1 were extracted according to the following algorithm. "Data" was the original matrix containing the neural recordings; "Labels" was a cell vector consisting of the identification of when the behavior was performed or not performed and "behavior" was the action investigated.
The output of the function were "DataOff" and "DataOn", which included the new Data matrices respectively when the behavior was not performed and when it was performed. Additionally "indexYes" and "indexNo" were computed and investigated for debugging purposes. They represent respectively the indexes of the initial matrix of data points used to build the "DataOff" and "DataOn" datasets.

```

function [indexYes, indexNo, lb, DataOff,DataOn]= MethodologyB(Data, Labels, behavior)

lb = Labels{1,behavior}(1:size(Data,1));
uno = find(lb); % find the indexes of the events


% samples contained in 1  minutes round window
rw = 60*500;

% initialize the variables indexOn, DataOff, DataOn
a=1;
  for i=1:length(uno)
    indexOn =uno(i);
    if indexOn > rw
```

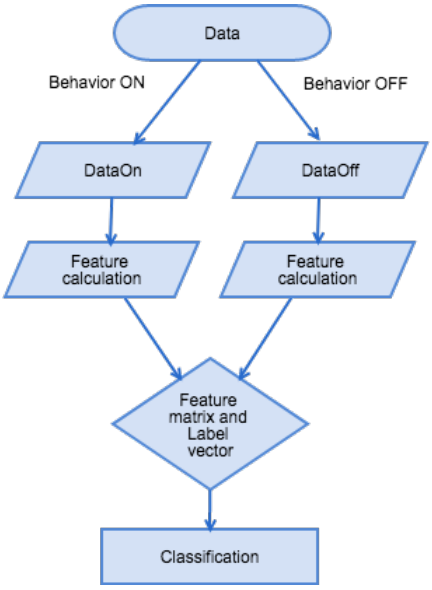
```
        interval_before = (indexOn -rw):(indexOn-1);
        zeros_before = find(lb(interval_before) ==0);
        if ~isempty(zeros_before)
            x= zeros_before(randi([1,length(zeros_before)]));

            indexNo(a) = interval_before(x);
            DataOn(a,:) = Data(indexOn, :);
            indexYes(a)= indexOn;
            DataOff(a,:) = Data(indexNo(a),: );

            lb(indexNo(a))= 3;
             a=a+1;

        elseif indexOn+rw  < length(lb)
            interval_after = (indexOn+1):(indexOn+rw);
            zeros_after = find(lb(interval_after)==0);

            if ~isempty(zeros_after)
                        x=zeros_after(randi([1,length(zeros_after)]));
                indexNo(a) = interval_after(x);
                indexYes(a)= indexOn;
                DataOn(a,:) = Data(indexOn, :);
                DataOff(a,:) = Data(indexNo(a),: );

                lb(indexNo(a))= 3;
                a=a+1;
            end
        end
    end
end
 DataOff= DataOff(~isnan(DataOff(:,1)),:);
 DataOn = DataOn(~isnan(DataOn(:,1)),:);

end
```

Unfortunately, methodology B1 could suffer of an argument already discussed in the previous sections. In fact, according to the considerations already exposed about the human factor, this approach would suggest that humans could be able to annotate very precisely following a "ms" time-scale. In fact, the original annotations were obtained considering a very precise beginning and end of the behavior and the labels were computed according to these annotations. In methodology B1, we select the ON and OFF samples based on these labels, which inform about the behavior at every "ms". Unfortunately, we already discussed about the possible noise derived from treating the data samples without using the round window considered for the feature calculation. Consequently, another approach was investigated.

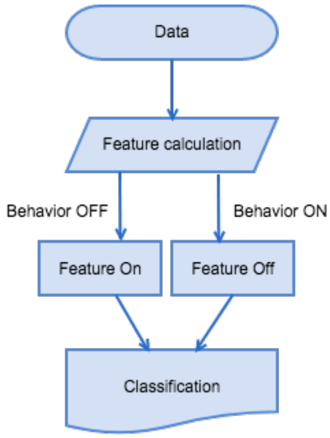
In methodology B2, we tried to overcome the problem by selecting the data after the feature using datapoints of one second was calculated. This means that the one-second features were calculated for all the data and the selection of the events close in time was performed directly on the features. In this way, transitory seconds (when the label was not constant over the second) were discarded and only features calculated over entire seconds of OFF behavior close to the ON event were considered. Also in this case, the round window examined to define the "0" labeled sample was one minute.

In summary, the main difference between methodology B1 and methodology B2 consisted of when the selection of the data was performed. Figure 7 shows the two processes of the second methodology and at what moment of the algorithm the negative examples were selected.

Finally, another modality was examined, referred as C. In this case,the data of the 24 hours were divided into clusters of three hours each, creating eight different sections of data. According to the number of events presented for every behavior in the three-hours sections, the classifier was trained and tested on different clusters. As an



(a) DataOn and DataOff were selected according to the initial labeling vector provided after the annotations. DataOff were selected randomly in the one-minute window previous or following the performance of the behavior. After the data were selected, the feature was calculated in the one-second window. This algorithm suggests that humans are able to characterize the behaviors at high resolution.



(b) In this second method, the selection of the Data to use for classification is performed after the feature over the 1-second window was calculated. Therefore, unclear seconds which contained both active and inactive behavior were discarded.

Figure 7: The two flow processes to select the Data for classification using methodology B. On figure 7a it is shown the flow for methodology B1 and in 7b for methodology B2.

example, in figure 6, the first two clusters were used to train the classifier and the third one was considered to test it.

For the three methodologies, different analysis were performed. The single channel analysis and the Fisher analysis were performed independently for the three methodologies. In addition, for the first methodology, another investigation that implied the shifting of the labels of different amounts of time was executed.

2.4.1 Classifiers

The main aim of this project was to build a model able to address the problem of identifying to which of a set of categories a new observation belongs. This argument is called classification in the field of machine learning. In this particular case, since the real information about the instances was provided, the problem falls under the process of supervised learning [16].
Supervised techniques are defined as machine learning tasks of inferring functions that maps an input to an output based on example input-output pairs [?]. The classifiers built in this analysis were characterized by statistical approaches rather than Artificial Neural Networks or decision trees. Therefore, these procedures were described by the probability of an object to belong to a given class and not only on the classification performance [24].

Particularly, in this project we focused on the binary classification, which involves only two classes. One of the most basic approaches to understand the problem is the Bayes's Rule Classifier.
In fact, this classification strategy is based on the investigation of the posterior probabilities of a given sample to belong to one of the two classes. Considering that a randomly selected observation X = x belongs either to $\prod_1$ or to $\prod_2$ respectively with probabilities π_1 and π_2, we could suppose that the conditional multivariate probability density of X for the i^{th} class is

$$P(X = x \mid \ X \in \prod_i) = f_i(x) \tag{5}$$

The function f_i does not require to be continuous. Subsequently, the posterior probabilities could be estimated from Bayes' theorem according to the following formula:

$$p(\prod_i \mid x) = P(X = \prod_i \mid X = x) = \frac{f_i(x)\pi_i}{f_1(x)\pi_1 + f_2(x)\pi_2} \tag{6}$$

Finally, the observation x would be assigned to the class with the higher posterior probability.
When the multivariate probability densities of the two classes are assumed to be multivariate Gaussian having arbitrary mean vectors and a common covariance matrix, the Bayes' rule classifier is defined as a Gaussian linear classifier. This hypothesis infers that f_1 follows $N((\vec{\mu_0}, \Sigma_0))$, f_2 follows $(\vec{\mu_1}, \Sigma_1)$ and $\Sigma_0 = \Sigma_1 = \Sigma$. The classifier obtained will be referred as linear in the following.
Afterwards, we also investigated the data assuming that the two classes followed different covariance matrices, which implies $\Sigma_0 \neq \Sigma_1$. In this case the function obtained is called a quadratic discriminant function and the boundary between the two classes is a quadratic function of the data.

Moreover, we also investigated these same type of classifiers but assuming that the electrodes' recorded signals were independent from each other, This would produce diagonal covariance matrices, Σ_d for the diaglinear classifier and Σ_{d1} and Σ_{d2} for the diagquadratic.

2.4.2 Single Channel

For the single channel analysis, every single channel was considered individually for training and testing. Therefore, the feature RMS in one-second round window was calculated for every electrode, but they were individually provided to train a linear classifier. The classification error estimation was implemented through a 10 fold cross-validation repeated four times. This provided ten values of the classification error for each repetition of the partition, which grants to calculate the average and the standard deviation over 40 values.

2.4.3 Fisher analysis

Fisher score is one of the most used filter methods for supervised feature selection. Filter methods aim to rank each feature according to some univariate metric and select the highest ranking features. Fisher score discrimination is obtained optimizing the separation of the projected samples by having a large difference between the means with respect to a measure of the standard deviation for each feature in each class, according to equation 7. μ_- and σ_- are the mean and standard deviation of the signal when the behavior is not performed and on the other hand μ_+ and σ_+ are the mean and standard deviation of the same electrode when the subject is executing the behavior. W is the projection to optimize.

$$J(w) = \frac{|\mu_- - \mu_+|}{\sigma_-^2 + \sigma_+^2} \tag{7}$$

The outcome of this algorithm is a ranking of the channels according to their capability to discriminate the two classes finding a projection of the data where the difference between the classes is maximized. The goal of Fisher discriminant analysis is to give a large separation of the class means while also keeping the in-class variance small.
One of the main limitations of this algorithm is that it investigates features independently and does not consider a possible correlation between variables, therefore causing a selection of redundant features. On the contrary, in the current study, this limitation appears to be useful since we are interested in knowing the performance of all the channels, in order to locate the active portions of the brain. Fisher score of individual electrodes was evaluated for every training partition, which involves that every electrode would be described by ten values of Fisher score. Moreover, the process was repeated 5 times, which led to the final number of 50 scores for every channel. The mean over these 50 values was calculated and the signals were ranked accordingly to this parameter to perform the evaluation of the best channels to compare to the results of the single channel analysis of the distinct methodologies.
After the evaluation of the Fisher scores for every training partition, different types of classifiers were trained and tested increasing the number of electrodes considered. More specifically, 104 classifiers were trained with the first one using only the signal recorded by the best electrode and the last one using all the channels. This was repeated for all the types of classifiers investigated which are the following: linear, diaglinear, quadratic and diagquadratic.
Moreover, each of these processes was embedded in a 10 fold cross validation.

2.4.4 Labels shifting analysis

An additional control to verify the robustness of classifier's performance was to shift the labels of various round windows either forward or backward on time with probability 0.5. The round windows chosen were: 1 second, 30 minutes and 1 hour. The Fisher score analysis was then repeated with the new labeling vector. The shifting of the labels was created using the following algorithm:

```
function [LabelsShifted] = ShiftLabels(tau, Labels)

indexes = find(diff(Labels));

initial = indexes(1:2:end)+1;
finals = indexes(2:2:end);

for i= 1:length(finals)
    if initial(i) < tau
         r(i) = 0.7;
    else
    r(i)= rand(1,1);
    end
    if r(i) > 0.5 % move to the right
       LabelsShuffled(initial(i):initial(i)+tau-1)= 0;
       LabelsShuffled(initial(i)+tau:tau + finals(i)) = 1;
    else
  % move to the left
       LabelsShuffled(initial(i)-tau:finals(i)-tau -1) = 1;
       LabelsShuffled(finals(i)- tau:finals(i) ) = 0;
    end

end
```

where “tau“ is the amount of shifting required, “Labels“ the initial vector containing the binary information about the class investigated and “LabelsShifted“ is the resulting vector after the shifting factor was applied.

3 Results

3.1 Annotations

A schematic illustration of the time-dynamics of the behaviors is shown in figure 8. The 24 hours recording starts at midnight and ends at midnight of the following day. The subject appears to have a very irregular sleep, above all during the night period, furthermore he also exhibits frequent and short movements of his head and his body. It is also possible to remark the questionable eating habits of the patient, who had small snacks during the entire day, but not a proper and long meal. Consequently, the eating events were short, but frequent during the day. The behavior "Someone is talking" is mostly constituted by conversations between the subject's family members as well as physical contact, which is particularly performed with parents and occasionally with nurses and doctors.

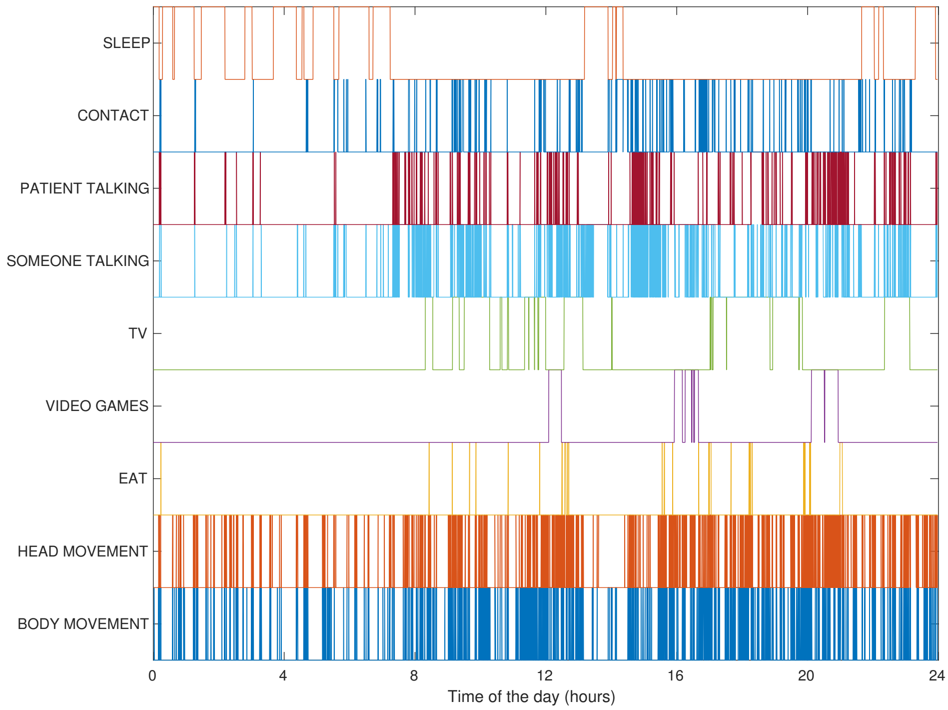


Figure 8: Time-dynamics of every behavior in the 24 hours analyzed for the specific subject TC starting from midnight. Body movement and head movement were the most frequent and lasted for a few seconds, for this reason their distribution in the figure resulted very dense. "VIDEO GAMES" and "EATING" resulted the least recurrent.

3.2 Methodology A

For this methodology, all the "OFF" data during the 24 hours were eligible to be selected before training the classifier. At first, every single electrode was used independently and uniquely, to determine the relevance of the single channel in the classification performance. Subsequently, the Fisher score for every channel and every behavior was computed, followed by a set of classifiers trained adding the signals of the electrodes according to the ranking provided by the Fisher algorithm.

3.2.1 Single Channel

The results obtained from the single channel analysis using RMS as feature are shown in Table 1 to 6 and Figures 8 to 13. The behaviors "PHYSICAL CONTACT" and "SOMEONE IS SPEAKING" did not present single channel classification errors lower than 0.4 and for this reason results were not shown.

Electrodes identified by numbers 4,11,12,13, 14 and 102 were selected in all the behaviors involving a movement either of the body, of the head or of the mouth when eating. Since channels 13 and 4 can be also found in the best performing electrodes of the behavior "PATIENT IS TALKING", we could advance the hypothesis that this portion of the brain is involved in the movement of part of the face. In fact, facial movements and expressions were not clearly distinguished when annotating "BODY MOVEMENT", which involves that the two could be simultaneous. Moreover, it seems reasonable that the movement of the body is correlated to facial movements. On the other hand, these signals were not significantly relevant for the other behaviors.
Electrode 38 seemed to be interesting for all the behaviors, which suggest an artifact or a non-stationarity in the data.

Channel	Classification error (± std)
14	0.3619 (± 0.0070)
13	0.3650(± 0.0082)
4	0.3762 (± 0.0090)
102	0.3802 (± 0.0098)
38	0.3861 (± 0.0094)
11	0.3882 (± 0.0089)
12	0.3937 (± 0.0104)
39	0.3975 (± 0.0082)

Table 1: **Methodology A.** BODY MOVEMENT: Electrodes that showed the lowest classification error when used individually to train a linear classifier with RMS as feature calculated over 1 second round windows.

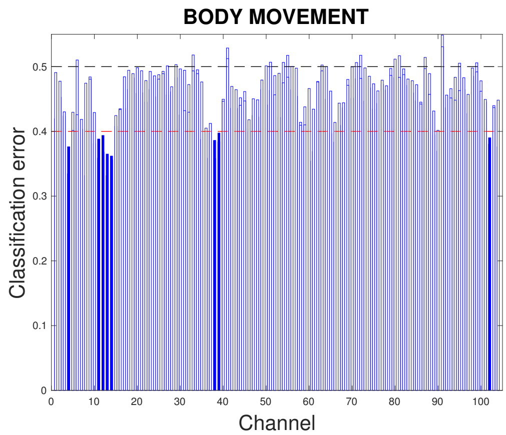


Figure 9: **Methodology A.** Classification errors obtained using a linear classifier trained with the RMS of the data computed over 1 second round window. Every single electrode was used independently. The dotted black line individuates chance, which is 50% and the red dotted line represents the threshold chosen of 40 % to determine the most significant signals.

Channel	Classification error (± std)
11	0.3530 (± 0.0128)
13	0.3629 (± 0.0128)
14	0.3643 (± 0.0108)
12	0.3729 (± 0.0114)
4	0.3747 (± 0.0113)
5	0.3786 (± 0.0124)
102	0.3822 (± 0.0110)
38	0.3889 (± 0.0113)
59	0.3903 (± 0.0091)
32	0.3963 (± 0.0088)

Table 2: **Methodology A.** HEAD MOVEMENT: Electrodes that showed the lowest classification error when used individually to train a linear classifier with RMS as feature calculated over 1 second round windows.

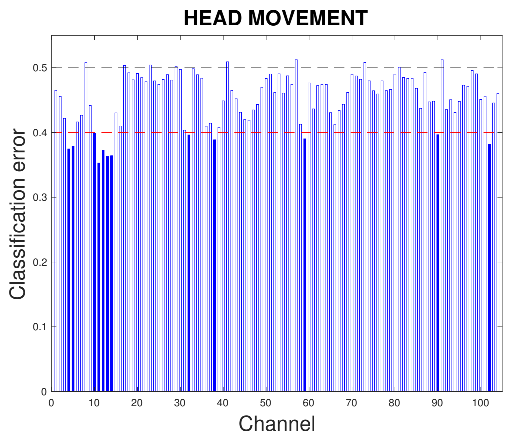


Figure 10: **Methodology A.** Classification errors obtained using a linear classifier trained with the RMS of the data computed over 1 second round window. Every single electrode was used independently. The dotted black line individuates chance, which is 50% and the red dotted line represents the threshold chosen of 40 % to determine the most significant signals.

Channel	Classification error	Channel	Classification error
13	0.3316 (± 0.0250)	90	0.3695 (± 0.0317)
8	0.3317 (± 0.0296)	104	0.3761(± 0.0308)
4	0.3346(± 0.0276)	10	0.3809(± 0.0280)
60	0.3540 (± 0.0281)	66	0.3833(± 0.0276)
14	0.3559 (± 0.0231)	39	0.3837 (± 0.0281)
38	0.3582 (± 0.0286)	35	0.3877 (± 0.0231)
36	0.3587 (± 0.0213)	102	0.3901 (± 0.0286)
11	0.3589 (± 0.0281)	89	0.3943 (± 0.213)
12	0.3680 (± 0.0259)	65	0.3966 (± 0.0281)
61	0.3689 (± 0.02017)	37	0.3979 (± 0.0259)

Table 3: **Methodology A.** EATING: Electrodes that showed the lowest classification error when used individually to train a linear classifier with RMS as feature.

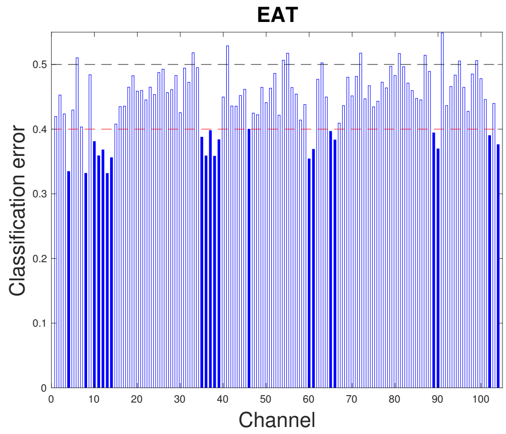


Figure 11: **Methodology A.** Classification errors obtained using a linear classifier trained with the RMS of the data computed over 1 second round window. Every single electrode was used independently. The dotted black line individuates chance, which is 50% and the red dotted line represents the threshold chosen of 40 % to determine the most significant signals.

Channel	Classification error (± std)
80	0.3176 (± 0.0212)
8	0.3325 (± 0.0189)
74	0.3491 (± 0.0174)
38	0.3727 (± 0.0193)
13	0.3740 (± 0.0205)
92	0.3786 (± 0.0204)
72	0.3775 (± 0.0190)
29	0.3788 (± 0.0164)
4	0.3887 (± 0.0195)

Table 4: **Methodology A.** PATIENT IS TALKING: Electrodes that showed the lowest classification error when used individually to train a linear classifier with RMS as feature.

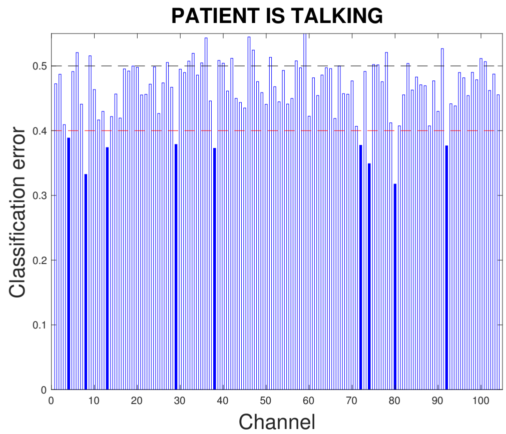


Figure 12: **Methodology A.** Classification errors obtained using a linear classifier trained with the RMS of the data computed over 1 second round window. Every single electrode was used independently. The dotted black line individuates chance, which is 50% and the red dotted line represents the threshold chosen of 40 % to determine the most significant signals.

Channel	Classification error	Channel	Classification error
38	0.3176 (± 0.0127)	53	0.3887 (± 0.0131)
37	0.3325 (± 0.0111)	59	0.3887 (± 0.132)
4	0.3491 (± 0.0101)	92	0.3887 (± 0.0129)
60	0.3727 (± 0.0158)	27	0.3887 (± 0.0128)
36	0.3740 (± 0.0106)	51	0.3887 (± 0.0114)
13	0.37867 (± 0.0099)	47	0.3887 (± 0.0139)
104	0.3775 (± 0.0107)	22	0.3887 (± 0.0119)
66	0.3788 (± 0.0133)		

Table 5: **Methodology A.** VIDEO GAMES: Electrodes that showed the lowest classification error when used individually to train a linear classifier with RMS as feature.

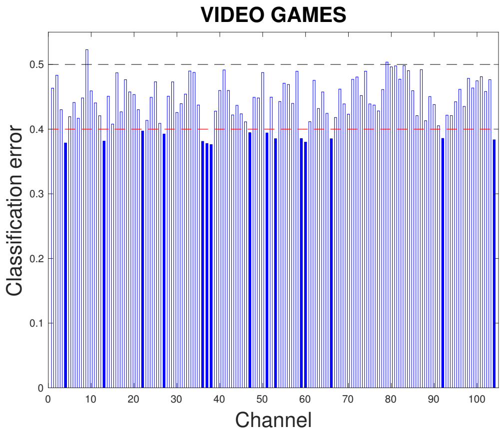


Figure 13: **Methodology A.** Classification errors obtained using a linear classifier trained with the RMS of the data computed over 1 second round window. Every single electrode was used independently. The dotted black line individuates chance, which is 50% and the red dotted line represents the threshold chosen of 40 % to determine the most significant signals.

Channel	Classification error (± std)
38	0.3316 (± 0.0067)
13	0.3452 (± 0.0069)
7	0.3613 (± 0.0078)
66	0.3647 (± 0.0068)
46	0.3670 (± 0.0071)
14	0.3778 (± 0.0078)
4	0.3827 (± 0.0057)
39	0.3857 (± 0.0065)
15	0.3960 (± 0.0078)
3	0.3968 (± 0.0082)
88	0.3986 (± 0.0063)
60	0.3989 (± 0.0073)

Table 6: **Methodology A.** WATCH TV: Electrodes that showed the lowest classification error when used individually to train a linear classifier with RMS as feature.

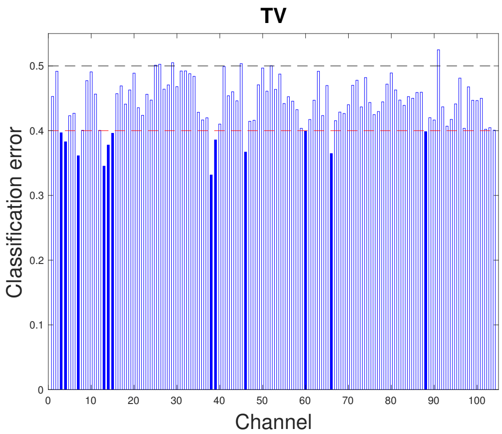
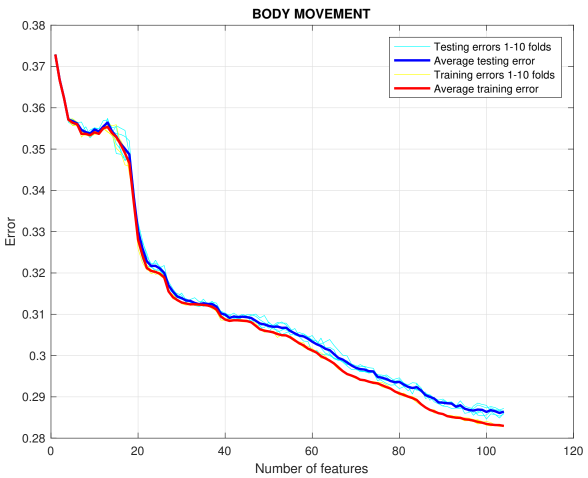


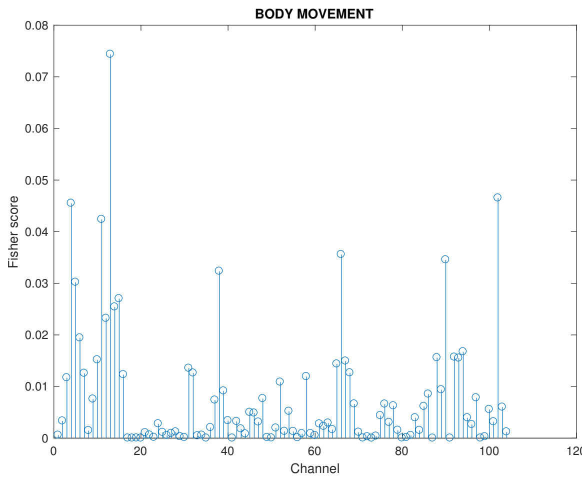
Figure 14: **Methodology A.** Classification errors obtained using a linear classifier trained with the RMS of the data computed over 1 second round window. Every single electrode was used independently. The dotted black line individuates chance, which is 50% and the red dotted line represents the threshold chosen of 40 % to determine the most significant signals.

3.2.2 Fisher

The results of Fisher algorithm include 104 classifiers built with the following reasoning: the first one trained only with the first best signal, the second trained with the two best signals and continuously until the last classifier which is trained with all the electrodes' information. Figure 15a shows the results of this algorithm for one of the behaviors, specifically when the patient was moving his body. As already remarked in the single channel analysis, a linear classifier trained with RMS of the best channel will have a performance of approximately 63 % (respective error of 37 %), subsequently the error continued to decrease adding the signals provided to the classifier.
The dark red and blue line in figure 15a represent the average of the training and testing error over 10 repetitions, whereas the results of the individual simulations were shown with light colored curves. Interestingly, for this behavior there was a temporary increase in the classification error when using between 10 and 15 electrodes, but there was no evidence of overfitting, since there was not an increase in the training error when more signals were considered by the classifier. The trend of these curves was similar in all the behaviors analyzed (refer to the Appendix for the behaviors not shown).
In figure 15b the Fisher scores obtained by the individual electrodes were shown for behavior "BODY MOVEMENT" . The first 8 electrodes with the highest Fisher score were: 13,102,4,11,66,90,38,5. Five values out of the eight confirmed the single channel analysis, electrodes 14 and 12 present respectively the 10th and the 11th best fisher score, whereas electrode 39 was at the 29th position of the ranking. Standard deviations (not shown) for the calculation of Fisher Power were lower than 0.002 for all the signals.



(a) **Methodology A.** Classification error for training and testing set according to the number of electrodes provided to the classifier following the ranking obtained by Fisher algorithm. The light curves represent the individual simulations and the dark lines describe the average. The worst classification error was approximately 0.37 and the best was in the range of 28 %- 29% for both the training and the testing set.



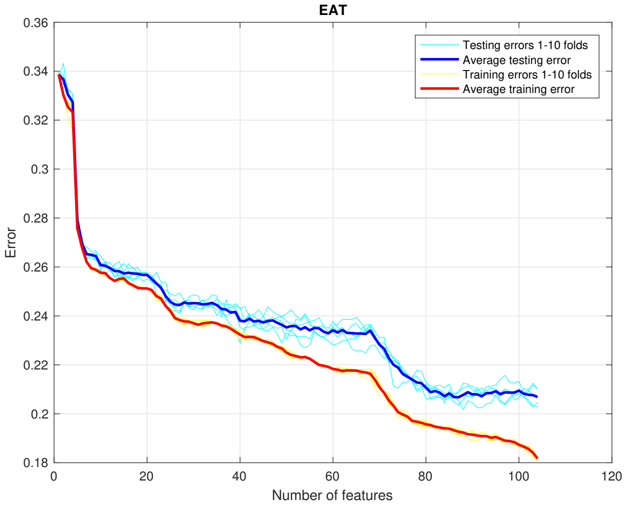
(b) **Methodology A.** Fisher scores for individual channels obtained using Fisher algorithm.These results were used to create a ranking of the electrodes (referred as features in fig 15a).

Figure 15: Results of Fisher analysis for BODY MOVEMENT

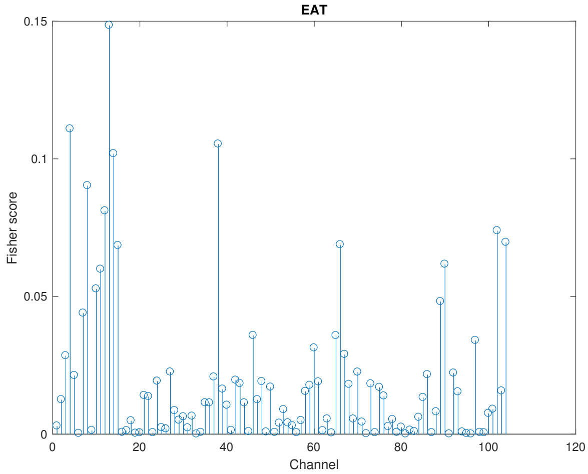
The only activity that slightly presented overfitting was "EATING", as shown in figure 16a. In fact, when approximately more than 90 channels were used to train the classifier, there was a noticeable decrease in the training error and a small increase in the testing error. As already noticed from the results of the single channel classification, the main relevant electrodes for "EATING" significantly overlapped with the ones involved in "BODY MOVEMENT". In fact, channels 4,8,12,13,14 and 38 presented the best Fisher scores.

Regarding the behavior "HEAD MOVEMENT", all the electrodes individuated by the single channel classification had a relative Fisher score among the twenty highest values except for signals recorded from electrode 59 and 14, which showed not relevant Fisher scores.

Interestingly, the most frequent channels selected for the behaviors which involved a movement (body, head, eating) were not relevant to decode when the patient was having a conversation. According to both single channel performances and Fisher scores, electrodes 29, 92, 72,74, 8 and 80 could provide information about when this behavior was performed. Therefore, these results suggest that channels 4, 13 and 38, which are



(a) **Methodology A.**Classification error for training and testing set according to the number of electrodes provided to the classifier following the ranking obtained by Fisher algorithm. The light curves represent the individual simulations and the dark lines describe the average. The worst classification error approximates around 0.34 and the best was in the range of 18 %- 22% respectively for the training and the testing set.



(b) **Methodology A.**Fisher scores for individual channels obtained using Fisher algorithm.These results were used to create a ranking of the electrodes (referred as features in fig 15a).

Figure 16: Results of Fisher analysis for EATING

frequently considered for the behaviors involving movements, could consequently represent specifically facial expression and activation of muscles of the face, as already suggested from the single channel analysis.

In order to understand whether channels 11,12,13,14,3,8 and 4 presented artifacts, we repeated the Fisher analysis for the "BODY MOVEMENT" behavior using RMS as feature removing the signals recorded from these electrodes. The results of the classification are shown in figure 17. The initial and final performances were lower compared to the analysis including all the signals, but the trend of the curve and the final outcome of the classification don't differ significantly. Consequently, we can conclude that the high performance of the classifier was not due by artifacts found in the most frequently chosen electrodes.

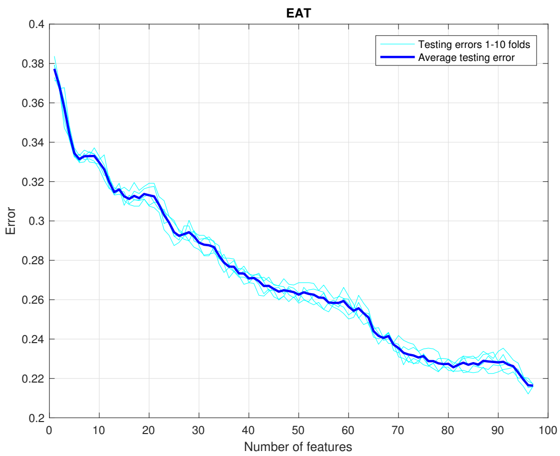
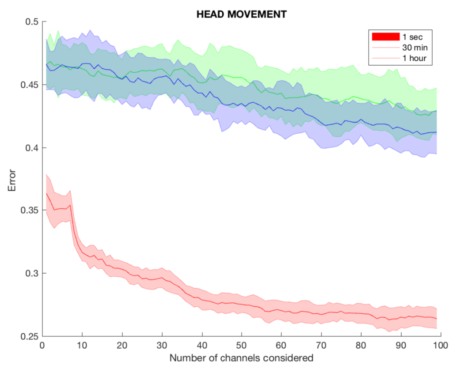


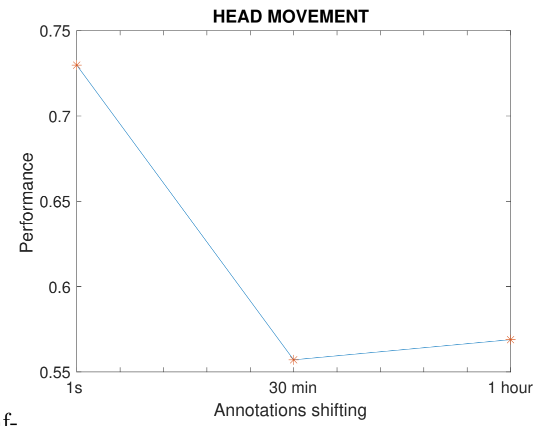
Figure 17: **Methodology A.** Classification error using Fisher algorithm after removing the channels 11,12,13,14,3,8 and 4 to investigate possible artifacts. The trend of the curve did not substantially differed from the one obtained using all the electrodes. As a result, artifacts in these signals were not considered relevant.

3.2.3 Shifting behavior analysis

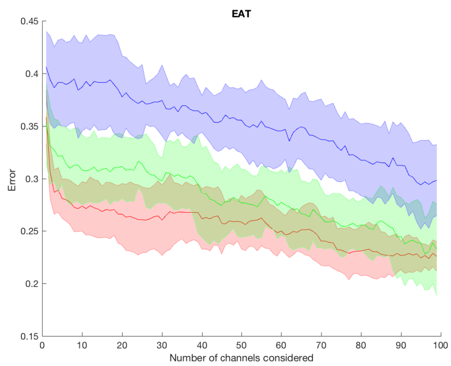
Another control to confirm the reliability of the classifier was implemented moving the labels on the time scale by 1 second, 30 minutes and 1 hour. Every single event for each behavior was shifted either backwards or forward with a probability of 0.5, as already explained in the Materials and Methods section. Figures 18a and 18b show the results obtained for one of the behaviors, specifically HEAD MOVEMENT. It is easily remarkable that the classification performance is lower when the shuffling is performed, but the classifier is not at chance level (0.5), meaning that it outperforms compared to reality. Moreover, there was an increase in the classification performance if the behaviors were shifted of one hour compared to the 30 minutes shift case. This characteristic can be explained by the time dynamics of this behavior. In fact, the subject moves his head frequently during the day and the shifting of one hour overlaps more than the 30 minutes shifting with the real labels. In fact, 30 minutes shift presented 3007 labels equal to the real classification and on the other hand 1 hour shift presented 3045 equal values. The same outcome can be achieved testing the EATING behavior (figures 18c and 18d). Unfortunately in this case there is an inconsistent difference between the performance before and after shuffling the labeling. Although for this activity there is a monotonic decrease in the performance from 0.75 to approximately 0.65, according to the overlaps within the labels. In fact, the amount of overlapping at 30 minutes shift was higher compared to the 1 hour shift.
It is important to note that although this proof did not achieve satisfying results and arises doubts on the functioning of the classifier, when shuffled labels were used for the same purpose they provided classification errors of 0.5 as expected. This suggests that the reasons of this result are more profound and deserve more accurate examinations.



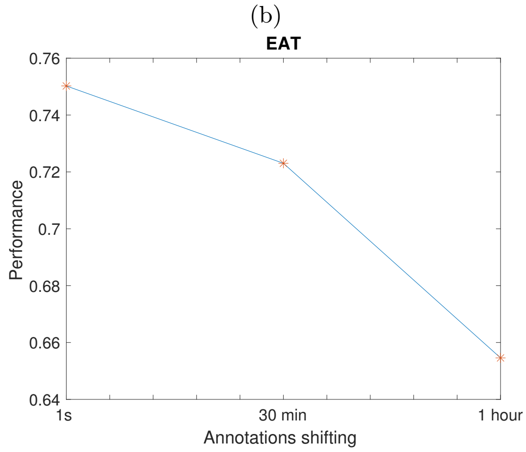
(a) Classification error using Fisher algorithm after removing the channels



(b)



(c) Classification error



(d)

Figure 18: The annotations for the behaviors HEAD MOVEMENT and EAT were shifted of 1 second, 30 minutes and 1 hour. The relative classification errors are shown in figure 18a and 18c. The shaded areas represent the standard deviation and the dark lines the average across trials. Figure 18b and 18d show the average performances for the three shifting factors considering only the classification errors obtained when all the electrodes were taken into account.

3.2.4 Classifier comparison

Various combinations of classifiers and features were used to evaluate the performance and the accuracy of the predictions. The features used were RMS, MAV, MaxAmp and MinAmp defined in equations 1, 2, 3 and 4 of the Materials and Methods section and the classifiers tested were: linear, diaglinear, quadratic and diagquadratic. In the following, in order to have a better understanding of the classifiers' achievements, we will investigate and compare them according to their performance in the classification accuracy, which is calculated as *1-classification error*.

The various performances for four different behaviors were shown in figure 19, this histogram was obtained considering only the classification results when all the 104 electrodes were used. The red horizontal line represents the chance level of 50%. The detailed results of the Fisher algorithm analysis for all the classifiers and all the features are shown in the Appendix. From figure 19 we could conclude that the linear classifiers always provided better and more stable performances compared to the other classifiers and the choice of the feature did not seem to impact substantially the accuracy of the classification. The quadratic and diagquadratic classifiers provided less stable accuracies across the features used and showed equal or slightly lower performances compared to the linear and diaglinear classifiers for all the behaviors in figure 19 except for "EATING". In fact in this case the best performances were provided by the quadratic classifier using MinAmp and MaxAmp.

Moreover, the complete Fisher analysis for the quadratic and diagquadratic classifiers (refer to Appendix figures 9-12) for all the features and all the behaviors presented ambiguous trends of the classification errors with high oscillations. An evident increase when most of the electrodes were considered probably caused by overfitting was also present, but it was often followed by a not very well explicable sudden decrease. For all these reasons, the linear and diaglinear classifiers should be considerate more confident.

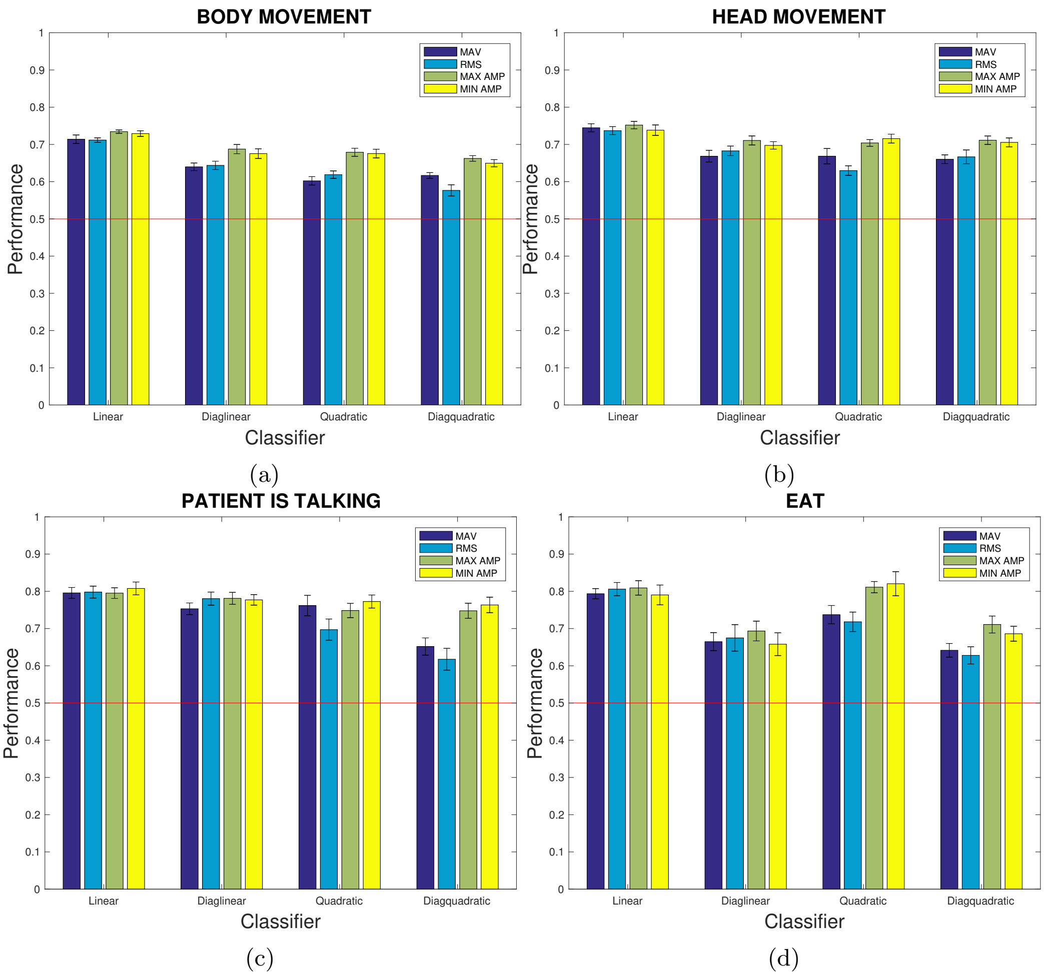


Figure 19: **Methodology A.** Comparison of the performances of the various combinations of classifiers and features. These results were based on the classification analysis when all the channels were taken into account. The red horizontal line represents chance at 50%.

3.3 Methodology B

The second methodology implied operations only with data labeled "0" (the subject is not performing the behavior) which were close in time to the data labeled "1". The aim of this methodology was to avoid that the classification was based on non-linearities in the data rather than real information about the patient's actions. Moreover, this methodology was also investigated using two different approaches to select the data for the classification. Th main difference between the two approaches was explained in the previous section and mainly concerns whether the data selection was performed before or after the feature calculation. The first approach is more sensitive to noise produced by inaccuracies of the annotations at the ms time-scale.

Methodology B1

3.3.1 Data selection

The selection of the data to use to fit the classifier caused the loss of some information. In fact, for some long behaviors as "Watching TV", "Video Games" and "Sleep", most of the initial data had to be discarded because it was not possible to obtain a reciprocal negative labeled sample in the 1 minute round window. This is shown in Table 7. This information also allowed to compare the number of elements provided to the classifiers when the feature was calculated before or after the data selection (methodologies B1 and B2).

Behavior	**Number of seconds**
Body movement	10827
Head movement	6061
Eating	870
Video games	583
TV	1337
Someone is talking	8131
Patient is talking	1934
Physical contact	3481
Sleep	1495

Table 7: Number of elements provided to the classifier for each behavior after the selection of the data and the feature calculation.

3.3.2 Single Channel

Interestingly, the results of the single channel analysis in this methodology differed substantially from the ones obtained in the previous procedure. The single channel analysis, as in the previous methodology, was based on the results of the classification using only one electrode and the feature considered was the root mean square (RMS). "SOMEONE IS TALKING" and "PHYSICAL CONTACT" did not present electrodes that with only theirs activity could produce a performance higher than 0.4 and for this reason they were not shown. This result was consistent with the analysis obtained using the methodology A for these behaviors. On the contrary, "WATCHING TV" and "EATING" presented respectively 48 and 40 channels whose signals could discriminate the behavior with an accuracy higher than 60%.

In Table 10 and 11 only the best 10 electrodes were shown.

Compared to the first methodology, the best electrodes for "BODY MOVEMENT", "HEAD MOVEMENT" and"EATING" differ completely, as it is shown correlating Tables 8,9, 10 and Tables 1,2 and 3. The only parameter which remained constant was the high classification accuracy produced using electrode 11 for the three behaviors. Moreover, this channel seemed to be relevant also for the talking activity of the patient.

Another electrode which appeared to be interesting to analyze for these behaviors was the electrode 10. The distributions of the feature RMS for these two channels and the relevant behaviors were exposed in figure 20. A two-Sample Kolmogorov-Smirnov test was executed to assess whether the distributions derived from the same continuous distribution. This hypothesis test evaluated the difference between the cumulative distribution functions of the RMS when the behavior was performed and when it was not active. In the three cases and for both electrodes 10 and 11 the hypothesis was rejected at the 5% significance level.

Channel	Classification error (± std)
11	0.3398 (± 0.0116)
10	0.3657 (± 0.0093)
58	0.3827 (± 0.0137)
90	0.3908 (± 0.0092)
14	0.3908 (± 0.0113)
67	0.4091 (± 0.0109)

Table 8: **Methodology B1.** BODY MOVEMENT: results of the classification process using the signals of only a single electrode. Only the performances obtained the best electrodes were shown.

Channel	Classification error	Channel	Classification error
49	0.3217 (± 0.0417)	61	0.3659 (± 0.0438)
24	0.3270 (± 0.0430)	36	0.3664 (± 0.0402)
11	0.3360 (± 0.0398)	89	0.3682 (± 0.0431)
84	0.3447 (± 0.0358)	27	0.3684 (± 0.0441)
75	0.3531 (± 0.0355)	35	0.3723 (± 0.0379)
41	0.3531 (± 0.0407)	97	0.3816 (± 0.0424)
2	0.3580 (± 0.0426)	21	0.3842 (± 0.0474)
10	0.3608 (± 0.0383)	12	0.3862 (± 0.0496)
73	0.3622 (± 0.0463)	99	0.3955 (± 0.0394)
57	0.3655 (± 0.0387)	58	0.3970 (± 0.0520)

Table 9: **Methodology B1.** HEAD MOVEMENT:results of the classification process using the signals of only a single electrode. Only the performances obtained the best electrodes were shown.

Channel	Classification error (± std)
11	0.2400 (± 0.0248)
10	0.2724 (± 0.0252)
12	0.2932 (± 0.0273)
65	0.2976 (± 0.0261)
102	0.3037 (± 0.0239)
55	0.3121 (± 0.0228)
50	0.3139 (± 0.0218)
36	0.3385 (± 0.0284)
4	0.3387 (± 0.0250)
104	0.3414 (± 0.0230)

Table 10: **Methodology B1.** EATING :results of the classification process using the signals of only a single electrode. Only the performances obtained the best electrodes were shown.

Channel	Classification error (± std)
75	0.3120 (± 0.0418)
40	0.3236 (± 0.0479)
35	0.3266 (± 0.0408)
33	0.3327 (± 0.0374)
42	0.3342 (± 0.0468)
101	0.3347 (± 0.0396)
41	0.3466 (± 0.0392)
88	0.3495 (± 0.0411)
21	0.3508 (± 0.0385)
63	0.3535 (± 0.0462)

Table 11: **Methodology B1.** WATCH TV:results of the classification process using the signals of only a single electrode. Only the performances obtained the best electrodes were shown.

Channel	Classification error
87	0.3551 (± 0.0602)
46	0.3554 (± 0.0484)
88	0.3645 (± 0.0633)
77	0.3757 (± 0.0657)
13	0.3771 (± 0.0669)
3	0.3783 (± 0.0713)
86	0.3798 (± 0.0786)
21	0.3805 (± 0.0585)
82	0.3890 (± 0.0742)
31	0.3926 (± 0.0576)

Table 12: **Methodology B1.** VIDEO GAMES:results of the classification process using the signals of only a single electrode. Only the performances obtained the best electrodes were shown.

Channel	Classification error	Channel	Classification error
11	0.2962 (± 0.0294)	76	0.3757 (± 0.0266)
86	0.3204 (± 0.0242)	24	0.3877 (± 0.0248)
87	0.3247 (± 0.0216)	3	0.3879 (± 0.0270)
88	0.3299 (± 0.0230)	99	0.3903 (± 0.0284)
10	0.3415 (± 0.0268)	4	0.3928 (± 0.0256)
2	0.3498 (± 0.0234)	94	0.3929 (± 0.0233)
1	0.3549 (± 0.0260)	33	0.3938 (± 0.0289)
93	0.3590 (± 0.0290)	78	0.3973 (± 0.0307)
92	0.3705 (± 0.0256)	25	0.3977 (± 0.0231)
98	0.3722 (± 0.0260)	96	0.3986 (± 0.0259)

Table 13: **Methodology B1.** PATIENT IS TALKING:results of the classification process using the signals of only a single electrode. Only the performances obtained the best electrodes were shown.

Furthermore, according to table 13, two clusters of significant electrodes appeared to be involved in the talking behavior of the subject: 86,87,88 and 92,93,98. In this case, a two-Sample Kolmogorov-Smirnov test was not only implemented to compare if the distributions differed for ON and OFF behaviors for single electrodes, but also between electrodes. The outcome of the analysis confirmed the difference of the calculated RMS when the subject was having a conversation and when he was not and additionally the signals recorded from the three electrodes seemed to be significantly different for both the clusters.

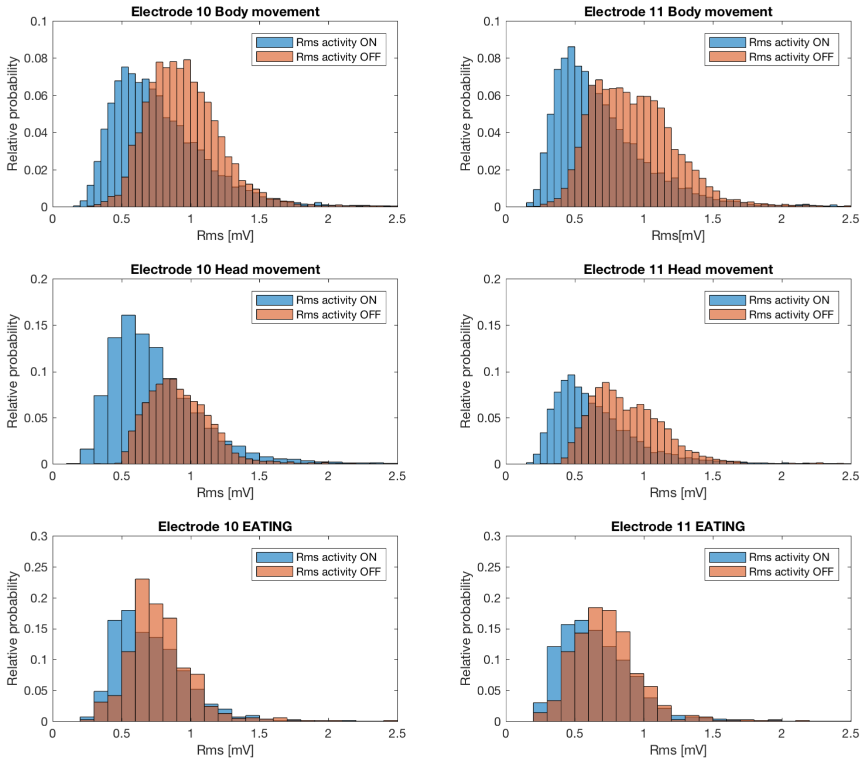
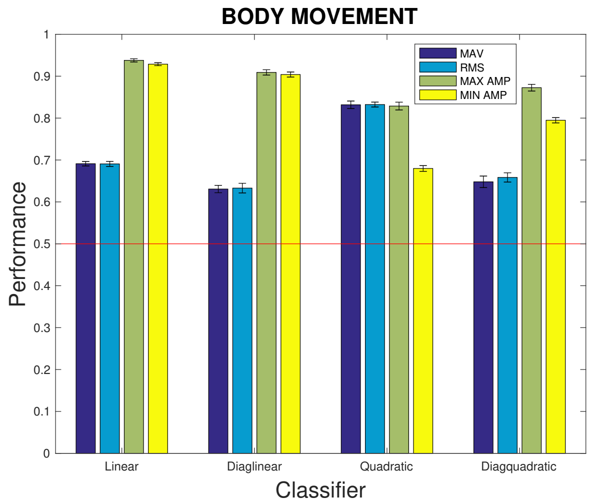


Figure 20: Distributions of the root mean square (RMS) calculated over 1 second round window for electrode 10 and 11 and three different behaviors. The red histogram represents the values calculated when the subject was executing the activity and in blue when he was at rest.

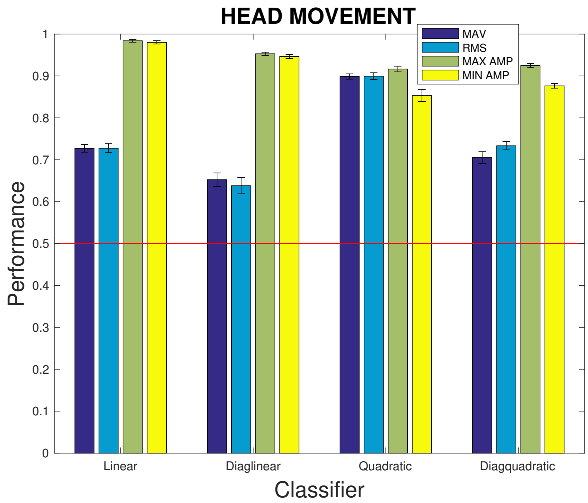
3.3.3 Discriminant analysis

The discriminant analysis using Fisher algorithm was also investigated with the results of this methodology. We tested all the possible combinations of type of classifier (linear, diaglinear, quadratic and diagquadratic) and feature (RMS, MAV, MaxAmp and MinAmp). Fisher algorithm provided performances according to the increasing number of electrodes furnished to the classifier; the following analysis focused on the results obtained when all the signals were considered. As for methodology A, these results were better understandable when characterized by the performances and not by the classification errors.
Mostly all the behaviors showed the same type of results according to the combination used. The smallest classification accuracies were obtained using the linear and diaglinear classifiers trained with MAV and RMS. For these two classifiers the minimal and maximal value (MinAmp and MaxAmp) showed unusual and curious very high accuracies very close to one. The quadratic classifier produced high and unexpected performances for all the features in all the behaviors analyzed. Also for the diagquadratic classifier the MAV and RMS produced accuracies in the range of approximately 60%-80% , meanwhile the performances obtained using MinAmp and MaxAmp as features did not present a prevalent attitude, but they differed for every behavior. Also in this case, accuracies were very high. Figure 21 shows the typical results of this analysis for the four behaviors already presented in the first methodology. It is therefore possible to directly compare the results of the first and second methodology. RMS and MAV in all the classifiers and in all the behaviors in methodology B1 show lower or equal performances than methodology A. But when using the MinAmp and MaxAmp as features, the

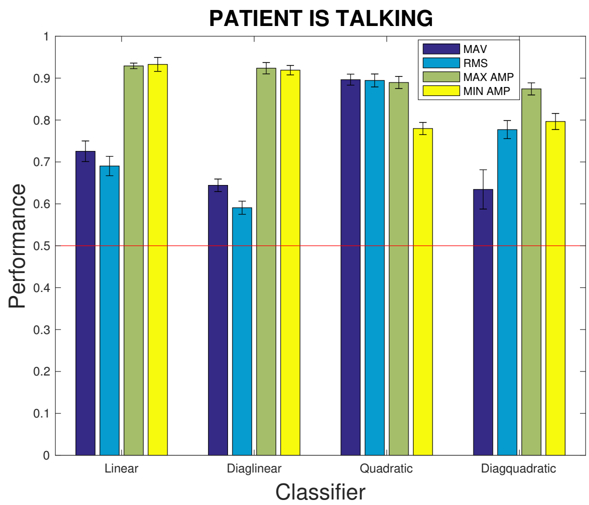
performances increase for all the behaviors and all the classifiers compared to methodology A.



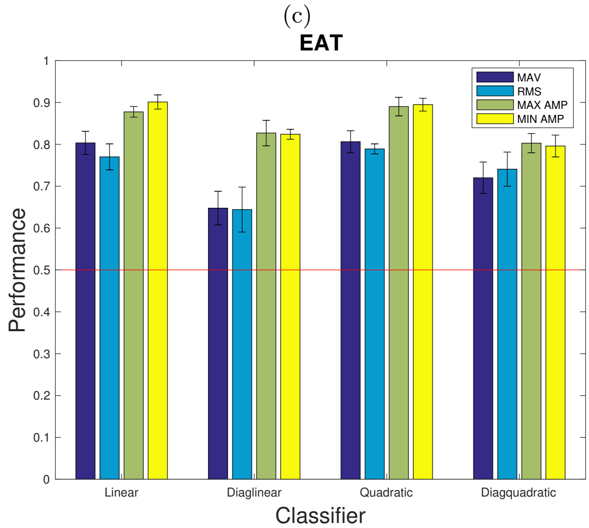
(a)



(b)



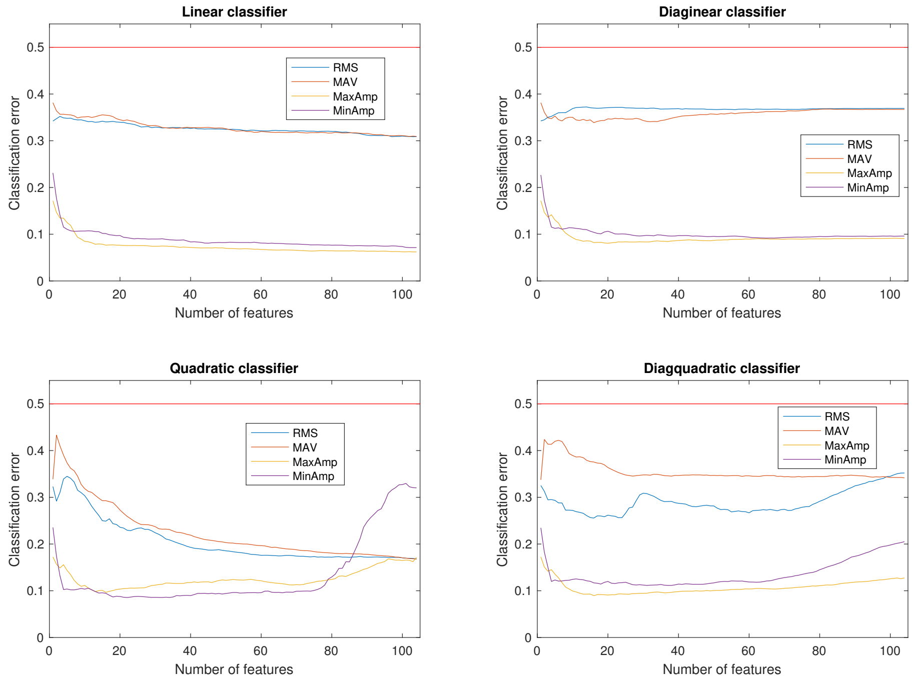
(c)



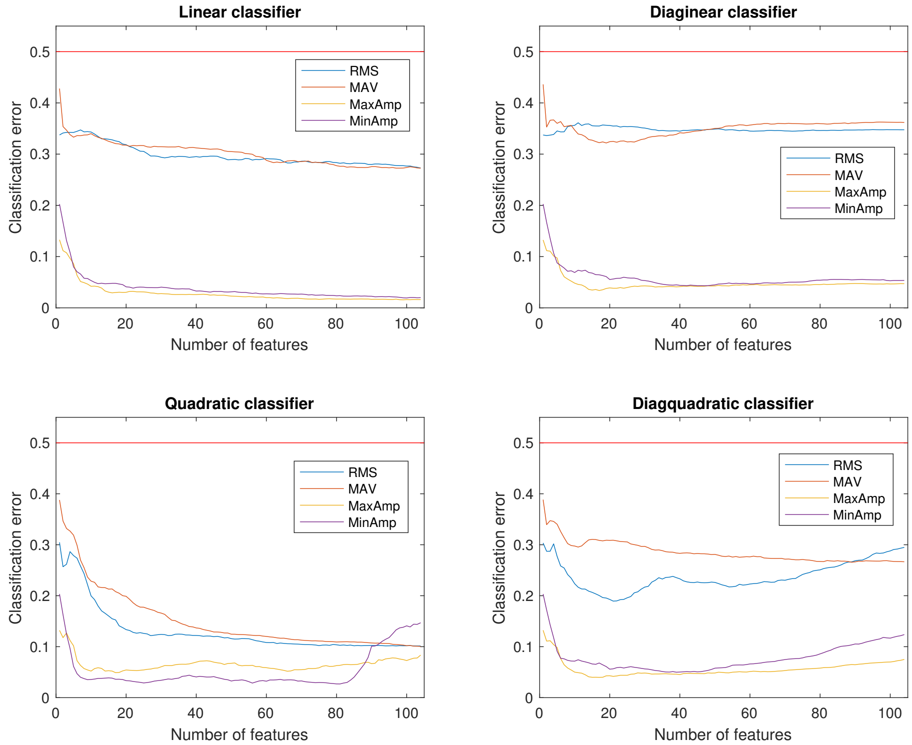
(d)

Figure 21: **Methodology B1.** Comparison of the performances of the various combinations of type of classifiers and features using negative samples of the behaviors collocated in the 1 minute round window prior or subsequent to the positive event. These results were based on the classification analysis when all the channels were taken into account. The red horizontal line represents chance at 50%.

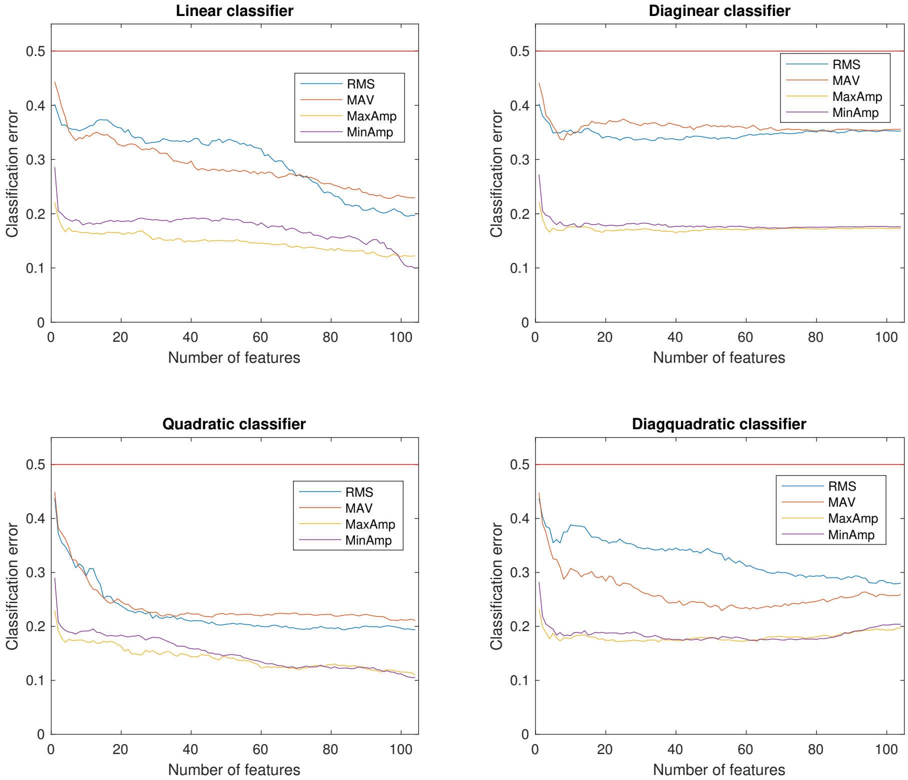
Figure 22 shows the results of the Fisher analysis for the same behaviors of figure 21 conforming to the classifier and the feature used. The minimal and maximal value present very high accuracies close to 90% for all the classifiers and all the behaviors. Moreover, the error computed using the minimal amplitude constantly increased when more than 80 channels were provided to the quadratic and diagquadratic classifier. This conduct could also be observed in less extent when the maximal amplitude was used. On the other hand, the classification errors for RMS and MAV oscillated around 30% and 40% using the linear and diaglinear classifiers for all the behaviors except for "EATING", where the error achieved 20 % when all the channels were considered (figure 39c upper left). These two features also seemed to have the same conduct for all the behaviors in the case of the quadratic classifier. The error decreases when more signals are used and the range observes varies between 0.4 and 0.2. The same argument can not be proposed for the diagquadratic classifier, where errors were less stable.



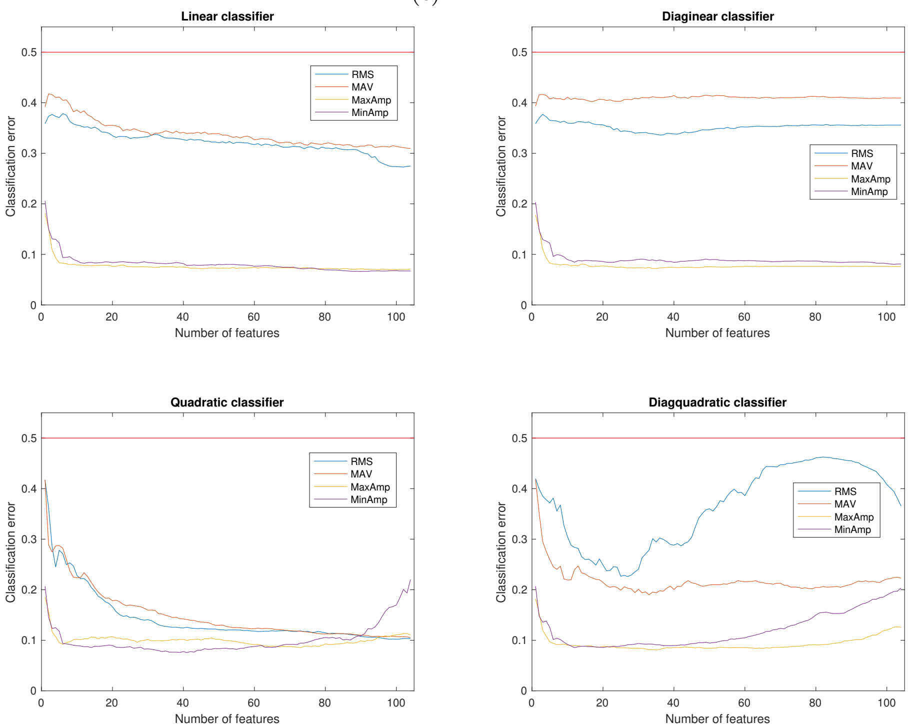
(a) BODY MOVEMENT



(b) HEAD MOVEMENT



(c) EAT



(d) PATIENT IS TALKING

Figure 22: **Methodology B1.** Comparison of the classification errors using Fisher algorithm for the types of classifiers analyzed (linear, diaglinear, quadratic and diagquadratic) and the features calculated with a round window of one second. The horizontal red line represents chance level at 50%. The number of features indicated in the x-axis characterizes the number of signals provided to the classifier according to the Fisher scores.

The trend of the classification error computed with the features MinAmp and MaxAmp require additional investigations to understand this uncommon and very accurate classification.

Because of all these questions about the reliability of the results obtained with methodology B1, we proceeded analyzing the performance of the classifier according to the data selection executed with methodology B2.

Methodology B2

3.4 Feature analysis

In this second approach for methodology B, the features were calculated before the selection of the data. As a result, all the features' values that were obtained from seconds where the behavior was not completely defined were discarded. For this reason, the total number of elements considered for classification decreased compared to methodology B1, as it can be observed from Table 14. The activities that were mostly affected were those that involved the subject active movement and when other people were speaking.

Behavior	Number of seconds
Body movement	7979
Head movement	5024
Eating	840
Video games	580
TV	1313
Someone is talking	5584
Patient is talking	1280
Physical contact	3241
Sleep	1473

Table 14: Number of seconds used for the analysis of the classifier for each behavior for the methodology B2, which involved the data selection after the feature calculation.

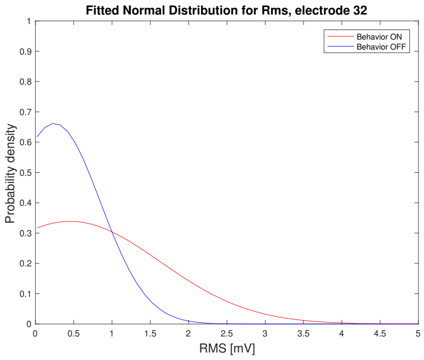
3.4.1 Single Channel

The classification accuracies of the methodology B2 for the single channel analysis using RMS as feature were lower than the ones obtained using methodology B1.
The smallest errors computed and the relative electrode for each behavior were the following:

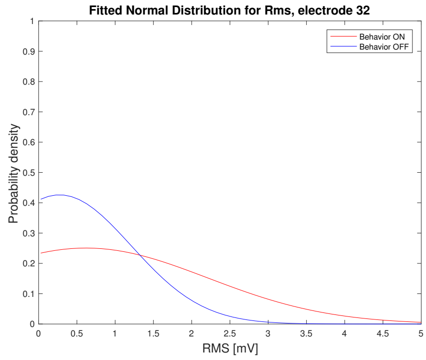
BODY MOVEMENT:	**0.4308**($\pm$0.008)	*electrode* **32**
HEAD MOVEMENT:	**0.3975**($\pm$0.001)	*electrode* **32**
EAT:	**0.4222**($\pm$0.0467)	*electrode* **88**
VIDEO GAMES:	**0.43**($\pm$0.0387)	*electrode* **101**
TV:	**0.4479**($\pm$0.0275)	*electrode* **45**
SOMEONE IS TALKING:	**0.46**($\pm$0.0133)	*electrode* **16**
PATIENT IS TALKING:	**0.42**($\pm$0.0275)	*electrode* **11**
PHYSICAL CONTACT:	**0.46**($\pm$0.0191)	*electrode* **102**
SLEEP:	**0.44**($\pm$0.0277)	*electrode* **65**

A normal distribution was fitted to feature RMS calculated on the data for electrode 32 in case of "BODY MOVEMENT" and "HEAD MOVEMENT" when these behaviors were executed and when they were not active. Figure 23 shows the obtained Gaussian curves. Only the positive values of the fitted distribution were shown because of the definition of RMS.
Remarkably, for both the behaviors the shape of the RMS ON was similar, which suggests that it would be very difficult to distinguish between BODY MOVEMENT and HEAD MOVEMENT. On the other hand the distribution of RMS OFF was very different between the behaviors and substantially differed also from the RMS ON. We could then propose that the signal recorded from this electrode was normally very variable, but acquired more specific values in case a movement was performed. Therefore, electrode 32 could be relevant to understand the movements of the subject. To test this hypothesis, we also investigated the fitted Gaussian curves for electrode 32 for a behavior which did not involve movements, more specifically "PHYSICAL CONTACT".
Interestingly, in this case, which can be observed in figure 23c, the distributions of RMS ON and RMS OFF were very similar and not easily distinguishable and they are both comparable to the curves of RMS ON for the already analyzed behaviors. Additionally, it is important to remark that in figure 23c we discarded the information about the other behaviors. This implies that the physical contact could be highly correlated to the behaviors involving movements. For this reason we added the same analysis, but when the patient was actively

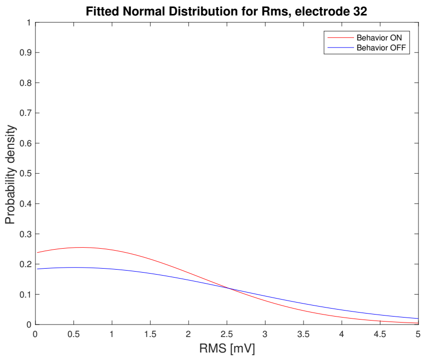
playing video games, which does not seem to be highly correlated to movements according to the people who made the annotations since the patient has to pay attention to the psychophysic tasks shown on the screen of the computer. From figure 23d we could observe that the distribution of RMS ON completely differed from the three other behaviors, but more importantly, the distribution of the feature when the behavior was ON and when it was OFF did not differ. This suggests that the distribution of RMS during movements for electrode 32 acquired a very-well defined shape.



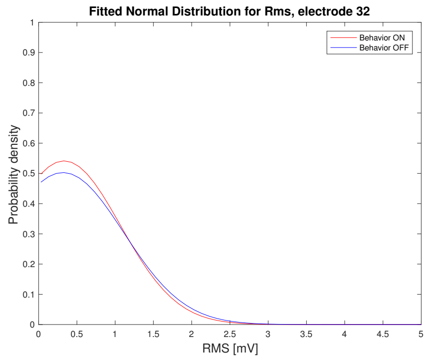
(a) BODY MOVEMENT



(b) HEAD MOVEMENT



(c) PHYSICAL CONTACT

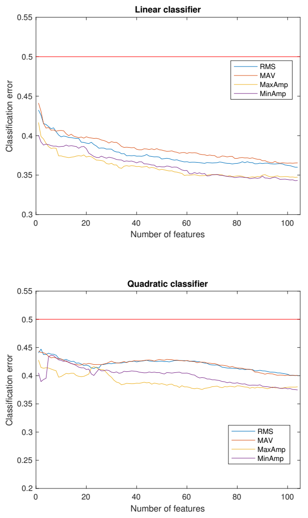


(d) PLAYING VIDEO GAMES

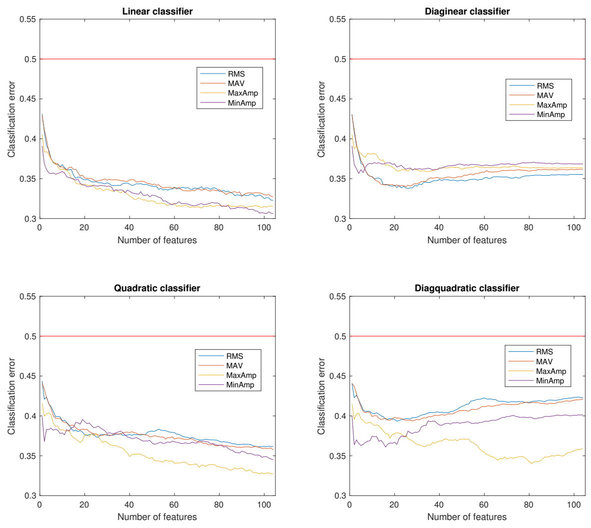
Figure 23: Normal distributions fitted to the data of Rms for electrode 32 for three different behaviors. The red line is obtained using the values calculated when the patient was moving and the blue line where it was resting. Negative values were not considered because RMS can only be positive.

3.4.2 Discriminant analysis

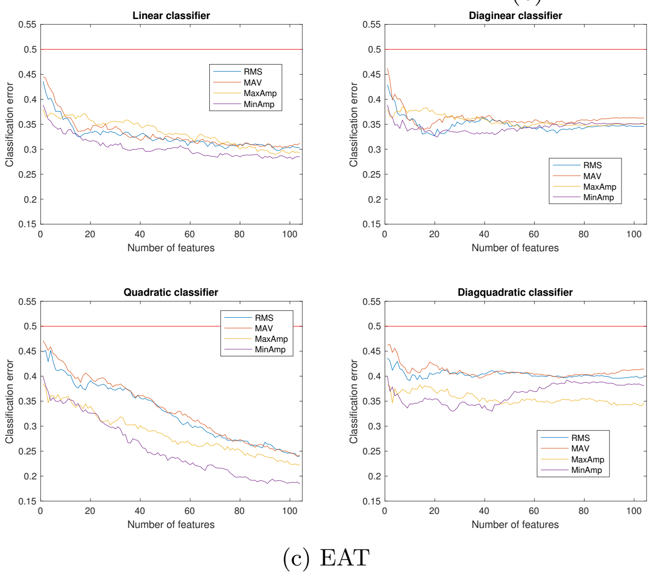
The Fisher analysis with all the possible combinations of classifiers and features was finally investigated also in this methodology. In order to allow an easier interpretation of the results the classification errors' trends were shown in figure 24 for the same behaviors.
Differently from methodology B2, in this case the MinAmp and the MaxAmp features had the same performances of RMS and MAV. The error oscillated around 35% and 45% and the only evidence of overfitting was observed when the diagquadratic classifier was used, except for when the patient was talking. Moreover, after the first twenty electrodes were provided, the errors seemed to stabilize, implying that most of the electrodes were not relevant.



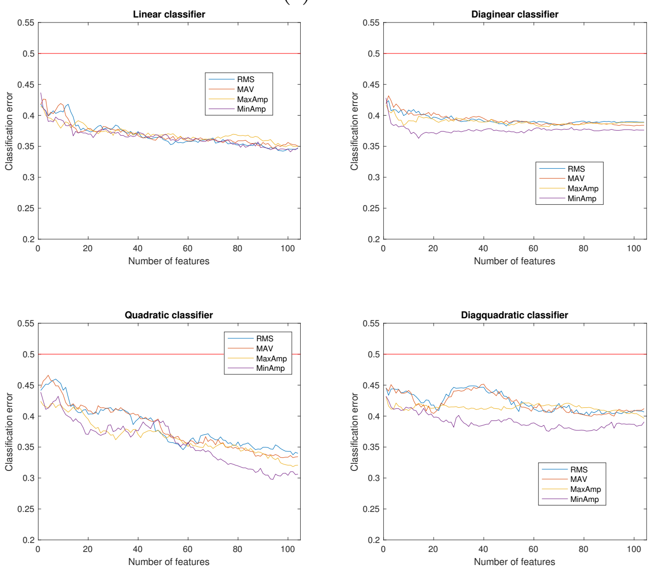
(a) BODY MOVEMENT



(b) HEAD MOVEMENT



(c) EAT



(d) PATIENT IS TALKING

Figure 24: **Methodology B2.** Comparison of the classification errors using Fisher algorithm for the types of classifiers analyzed (linear, diaglinear, quadratic and diagquadratic) and the features calculated with a round window of one second. The horizontal red line represents chance level at 50%. The number of features indicated in the x-axis characterizes the number of signals provided to the classifier according to the Fisher scores.

3.5 Methodology C

In methodology C, we tried to differentiate the behaviors according to the time at which they were performed. The aim of this analysis was to investigate whether the signals recorded by the electrodes differed over time and if the behavior physiology characteristics also differed during the day. More specifically, the question addressed was: *"Could it be possible that breakfast is different from lunch? Is it possible that a classifier trained using only the data of the morning is not capable of detecting the same behavior in the afternoon?"* In case of a negative reply to these questions we could conclude that the classifiers could be considered efficient and reliable.
In order to answer these questions, we separated the 24 hours in 8 clusters, each of them containing the data of 3 hours of the day. This means that the first cluster would contain the information from midnight to 3 am, the second cluster from 3 am to 6 am and so on until the end of the day. The classifiers were subsequently trained and tested using defined clusters, which also represent different time points during the day.
Unfortunately for this analysis, only a limited number of iterations was possible. In fact, the distributions of the behavior and consequently of the data was not uniform during the day and therefore over all the clusters. For example, the number of the observations labeled "1" (when the behavior is ON) in each cluster for the behaviors "EATING", "BODY MOVEMENT", "HEAD MOVEMENT" and "PATIENT IS TALKING" were the following:

Number of elements	BODY MOVEMENT	HEAD MOVEMENT
Cluster 1	436003	314924
Cluster 2	553332	256248
Cluster 3	534427	392771
Cluster 4	1107228	552353
Cluster 5	684388	346024
Cluster 6	1439985	737001
Cluster 7	1064741	583967
Cluster 8	617311	825543

Number of elements	EAT	PATIENT IS TALKING
Cluster 1	3923	27830
Cluster 2	0	14772
Cluster 3	14226	51557
Cluster 4	21936	81565
Cluster 5	161260	128213
Cluster 6	245008	648172
Cluster 7	191477	430329
Cluster 8	0	152697

Table 15: Data labeled "1" in each cluster and in each behavior. This information was used to determine which clusters to use to train and test the classifier. In fact, we wanted to keep an approximate proportion of 90% of observation for training and 10% for testing.

We can observe that the patient during the night was sometimes awake to drink water (the patient was drinking in the cluster from midnight to 3 am) and he also spoke a few words when waking up to his parents or nurses. This is the reason why the number of elements in PATIENT IS TALKING is never zero in any cluster.
In order to keep the proportion of data of approximately 90 % for the training and 10 % for the testing of the classifier, not all the possible combinations of the clusters were used. In fact, we could not randomly decide which cluster to use to train the classifier and which one to test it, because of the non-uniformity of the data. For this reason only a few possibilities were investigated, the features used were MaxAmp and RMS.
For the behavior BODY MOVEMENT the classifier was trained on clusters from 1 to 6 and tested on the remaining clusters 7 and 8. The testing error were respectively 0.3104 when using MaxAmp and 0.3882 when using RMS.
For the behavior HEAD MOVEMENT two classifiers were investigated, the first one used the first seven clusters for training and was tested on the eighth and the second one used for testing cluster 6 and was trained on the remaining ones. The obtained testing errors were 0.3682 for the RMS and 0.3592 for MaxAmp for the first separation and 0.4423 for RMS and 0.4055 for MaxAmp for the second division.

For the behavior EAT the clusters five and six were used to train the classifier and the first four to test it. The resulting testing errors were approximately 0.44 for both the features.

Finally, also for "PATIENT IS TALKING" two separations of the data were considered. When only the last cluster was used for testing the errors were 0.3652 for MaxAmp and 0.4064 for RMS. On the contrary, when only cluster five was not provided to train the classifier, the errors were 0.3947 and 0.3816 for RMS and MaxAmp.

3.6 Locations of the electrodes in the three classifiers

In this section we provide the description of the location of the electrodes in the brain of the patient. In the previous sections the electrodes were identified by a given number in order to have a simple and straightforward way to compare the different methodologies. In fact, the mnemonics given by numbers were easy to recognize and recall.
Nevertheless, it is interesting and significant to investigate the positions of the electrodes and the portions of the brain related to them. Every number is associated to an acronym as reported in the following.

1	**IP1**	11	**IP11**	21	**FP5**	31	**FP15**	41	**L9**	51	**LA3**	61	**LA13**	71	**LF7**
2	**IP2**	12	**IP12**	22	**FP6**	32	**FP16**	42	**LP10**	52	**LA4**	62	**LA14**	72	**LF8**
3	**IP3**	13	**IP13**	23	**FP7**	33	**LP1**	43	**LP11**	53	**LA5**	63	**LA615**	73	**LF9**
4	**IP4**	14	**IP14**	24	**FP8**	34	**LP2**	44	**LP12**	54	**LA6**	64	**LA16**	74	**LF10**
5	**IP5**	15	**IP15**	25	**FP9**	35	**LP3**	45	**LP13**	55	**LA7**	65	**LF1**	75	**LF11**
6	**IP6**	16	**IP16**	26	**FP10**	36	**LP4**	46	**LP14**	56	**LA8**	66	**LF2**	76	**LF12**
7	**IP7**	17	**FP1**	27	**FP11**	37	**LP5**	47	**LP15**	57	**LA9**	67	**LF3**	77	**LF13**
8	**IP8**	18	**FP2**	28	**FP12**	38	**LP6**	48	**LP16**	58	**LA10**	68	**LF4**	78	**LF14**
9	**IP9**	19	**FP3**	29	**FP13**	39	**LP7**	49	**LA1**	59	**LA11**	69	**LF5**	79	**LF15**
10	**IP10**	20	**FP4**	30	**FP14**	40	**LP8**	50	**LA2**	60	**LA12**	70	**LF6**	80	**LF16**

81	**IA1**	91	**IM3**	101	**IO5**
82	**IA2**	92	**IM4**	102	**IO6**
83	**IA3**	93	**IM5**	103	**IO7**
84	**IA4**	94	**IM6**	104	**IO8**
85	**IA5**	95	**IM7**		
86	**IA6**	96	**IM8**		
87	**IA7**	97	**IO1**		
88	**IA8**	98	**IO2**		
89	**IM1**	99	**IO3**		
90	**IM2**	100	**IO4**		

In turn, every acronym was associated to a portion of the brain according to the parcellation of Destrieux et al.[25].A schematic of this parcellation is shown in figure 25 (from [25]). The area of the brain associated to each electrode is defined in the Appendix.
All the electrodes were located in the right hemisphere of the patient's brain, the m00043_destrieux.fig figure attached to this report shows the locations of the electrodes and their identifying acronyms on the brain. The numerical indexes in figure 25 refer to the anatomical regions defined in Table 1 of the above-mentioned article [25]. In fact, this ATLAS is considered both anatomically valid and reliable for subdividing the human cerebral cortex into standard gyral-based neuroanatomical regions.
It is important to know that also other ATLAS and parcellations of human cortical gyri and sulci exist [26] [27]. The electrodes 10 and 11, which seemed to be significant for the decoding of the movements in both methodologies A and B1, are referred as IP10 and IP11 and they were located respectively at the middle-posterior part of the cingulate gyrus and sulcus (pMCC) and in the superior frontal gyrus (F1) according to Destrieux et al. Interestingly, they were not located in an area in proximity to the motor cortex. We also investigated the positions of the electrodes IA6, IA7, IA8 (86,87,88)and IM4, IM5 and IO2 (92,93,98) because they seemed to create clusters in the determination of the behavior "PATIENT IS TALKING" according to the classification accuracies obtained in methodology B1.
All the electrodes were positioned in the superior frontal gyrus (F1) except for IO2, which was located in the Precuneus. Therefore, this allocation suggests to be very relevant for the communicative activity of the subject. In fact, other studies already proposed that F1 could remarkably contribute to higher cognitive functions and in working memory [28].
Finally, the positions of the electrodes individuated by the single channel analysis of methodology B2 were determined. The electrode FP16 (32), which we found to be relevant for the determination of the movements,

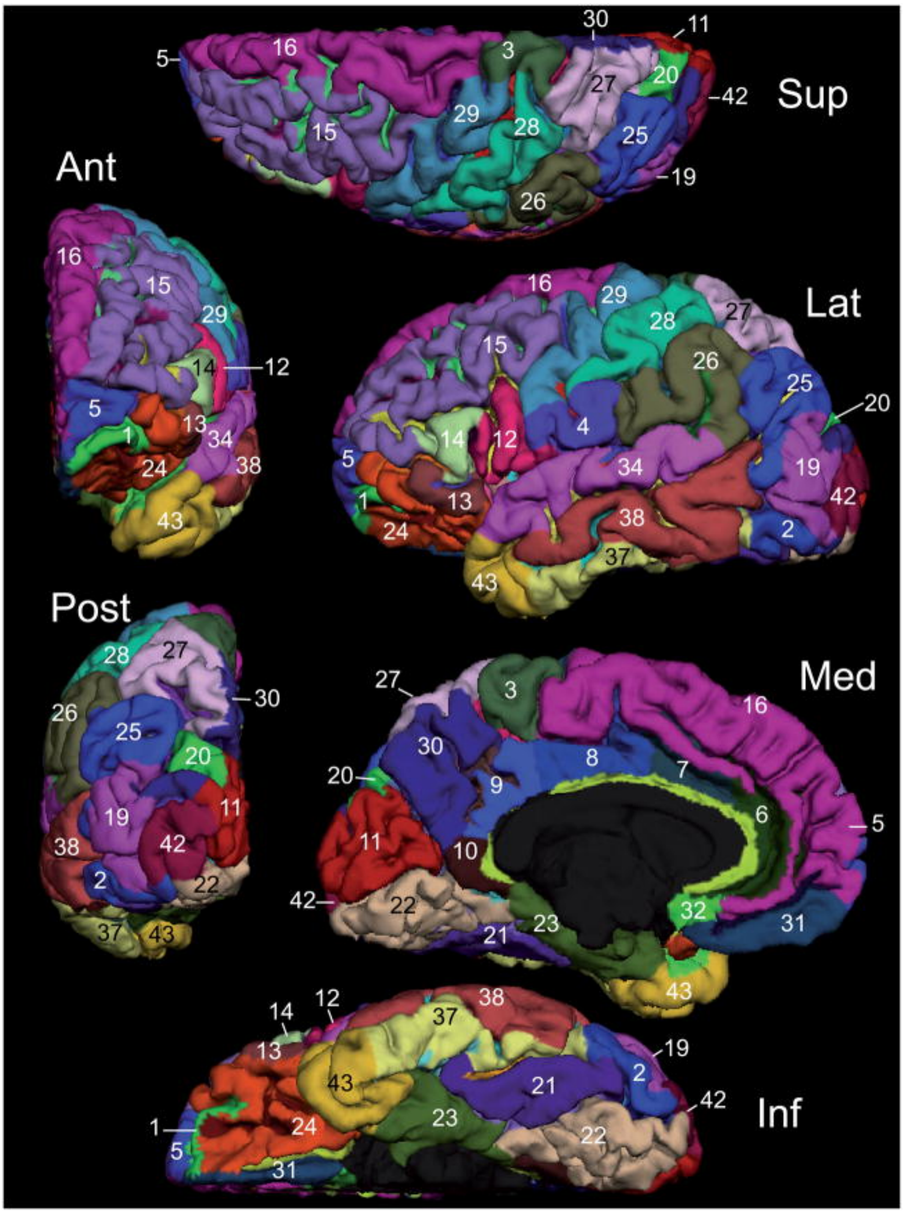


Figure 25: Parcellation of human cortical gyri and sulci according to Destrieux et al. Superior (Sup), anterior (Ant), lateral (Lat), posterior (Post), medial (Med), and inferior views are provided. To note: Precuneus 30, pMCC 8, F1 16, F2 15, T2 38, postcentral gyrus 28, precentral gyrus 29, paracentral lobule and sulcus 3.

was located in the middle frontal gyrus (F2). This association has never been found before. In fact, it has already been associated to selective attention by serving as a circuit-breaker to interrupt ongoing endogenous attentional processes in the dorsal network and reorient attention to an exogenous stimulus, but never to motor activities [29].

Electrode LF1 (65), involved when the patient was sleeping, was located in the middle temporal gyrus (T2), whereas electrodes LP45 (45) and IP16 (16) were recording respectively from the postcentral and and precentral gyrus. On the contrary, electrodes IO5 (101) and IO6 (102) were positioned in the Precuneus and paracentral lobule and sulcus. Lastly, electrode IA8 (88), which was involved in the subject's eating and drinking activities, was located in the Precuneus.

4 Discussion and further scientific questions

The main aim of this project was to investigate whether it was possible to apply machine learning techniques to decode behaviors not only in very well experimental controlled experiments, but in real life situations. In fact, in the future, the importance of detecting natural behaviors will certainly increase for the necessity of using this information in real life applications.
In this study, we found that very simple algorithms could already be suitable for this purpose, leading to accuracies much higher than chance.
Three different modalities were investigated to improve the reliability of the classifier, starting from the broadest methodology which considers all the data in the 24 hours and ending with a classifier whose performance does not change over time (methodology C).

This project allows room for a high number of subsequent analysis.
In fact, at first it would be a necessary and interesting survey to investigate more deeply the reason why the MinAmp and MaxAmp features produce very high classification accuracies in methodology B1 and if this underlines a specific physiological phenomenon or it is the result of an artifact. Additionally, when using the methodology B2, which only slightly differ from B1, the mentioned features are again comparable with RMS and MAV and produce more stable classification errors.

Another necessary analysis would be to apply the shifting of the labels also to the methodologies B1 and B2, in order to confirm the highest reliability provided by these two methods to select the data compared to methodology A.

Evaluation of the behaviors and between-behaviors autocorrelations and correlations could also be very interesting and useful. As an example, in figure 23 it would be very significant to associate the distributions of the features between the behaviors with an analysis of the correlations. Indeed, the investigation would be more precise and significant. Given a specific electrode which is relevant for one behavior and not discriminative for another action, the distribution of the features calculated from the signal recorded in OFF and ON cases could be compared knowing the information about the correlation of the behaviors. Consequently, these distributions could be investigated using both the data when the actions are performed at the same time and when only one is executed. This would produce more specific investigations.

Another important point is constituted by the annotations themselves. A narrow examination of the annotations was already performed in this project (not shown). The same time frame was annotated from two different people and the the overlapping varied between 2 % and 5 % difference. Although, it is important to note that this percentage was calculated based on the first labels, which are at the ms time-resolution. A more accurate investigation should consider the features' values over the 1 second round window according to the various "annotators", which should increase in number.

Moreover, other possible combinations are possible to train and test the classifier according to methodology C, which were not considered in the present article.

Finally, all the process could be repeated using the data from another subject. Unfortunately, this investigation has some limitations since the positions of the electrodes in the subjects' brains are obviously subordinated to the clinical purposes. For this reason, it would be more difficult to compare the final outcomes. On the contrary, the positive aspect is that more areas could be investigated [30].

References

[1] G. A. Gescheider, *Psychophysics*. 1997.

[2] G. Buzsáki, C. A. Anastassiou, and C. Koch, “The origin of extracellular fields and currents — EEG , ECoG , LFP and spikes,” vol. 13, no. June, pp. 407–420, 2012.

[3] A. I. Sburlea, L. Montesano, and J. Minguez, “Advantages of eeg phase patterns for the detection of gait intention in healthy and stroke subjects,”

[4] I. Transactions, O. N. Neural, and R. Engineering, “Guest Editorial Brain – Computer Interface Technology : A Review of the Second International Meeting,” vol. 11, no. 2, pp. 94–109, 2003.

[5] G. J. Ortega, R. G. Sola, and J. Pastor, “Complex network analysis of human ECoG data,” *Neuroscience Letters*, vol. 447, no. 2, pp. 129–133, 2008.

[6] J. A. Wilson, E. A. Felton, P. C. Garell, G. Schalk, and J. C. Williams, “ECoG Factors Underlying Multimodal Control of a Brain–Computer Interface,” vol. 14, no. 2, pp. 246–250, 2006.

[7] R. S. Fisher, W. V. E. Boas, W. Blume, C. Elger, P. Genton, P. Lee, and J. Engel, “Epileptic Seizures and Epilepsy : Definitions Proposed by the International League Against Epilepsy (ILAE) and the International Bureau for Epilepsy (IBE),” vol. 46, no. 4, pp. 470–472, 2005.

[8] E. Wyllie, Y. G. Comair, P. Kotagal, J. Bulacio, W. Bingaman, and P. Ruggieri, “Seizure Outcome after Epilepsy Surgery in Children and Adolescents,” pp. 740–748, 1998.

[9] M. Negnevitsky, *Artificial Intelligence*. 2011.

[10] P. N. Stuart J. Russell, *Artificial Intelligence*. 2010.

[11] K. R. Murphy, *Machine Learning*. 2012.

[12] I. G. Maglogiannis, *Emerging Artificial Intelligence Applications in Computer Engineering*. 2007.

[13] S. B. K. I. D. Z. P. E. Pintelas, “Machine learning : a review of classification and combining techniques,” no. 2006, pp. 159–190, 2007.

[14] R. Kohavi, “A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection,” vol. 5, 1995.

[15] D. M. Hawkins, “The Problem of Overfitting,” pp. 1–12, 2004.

[16] A. J. Izenman, *Linear Discriminant Analysis 8.1*.

[17] J. H. Friedman, “Regularized Discriminant Analysis,” vol. 84, no. 405, pp. 165–175, 2016.

[18] T. Pistohl, A. Schulze-bonhage, A. Aertsen, C. Mehring, and T. Ball, “NeuroImage Decoding natural grasp types from human ECoG,” *NeuroImage*, vol. 59, no. 1, pp. 248–260, 2012.

[19] A. Jackson, J. Mavoori, and E. E. Fetz, “Correlations Between the Same Motor Cortex Cells and Arm Muscles During a Trained Task , Free Behavior , and Natural Sleep in the Macaque Monkey,” pp. 360–374, 2018.

[20] A. Jackson, C. T. Moritz, J. Mavoori, T. H. Lucas, and E. E. Fetz, “The Neurochip BCI : Towards a Neural Prosthesis for Upper Limb Function,” vol. 14, no. 2, pp. 187–190, 2006.

[21] H. Neuroscience, “From speech to thought : the neuronal basis of cognitive units in non-experimental , real-life communication investigated using ECoG,” vol. 8, no. June, pp. 1–17, 2014.

[22] J. Ruescher, O. Iljina, D.-m. Altenmüller, A. Aertsen, A. Schulze-bonhage, and T. Ball, “NeuroImage Somatotopic mapping of natural upper- and lower-extremity movements and speech production with high gamma electrocorticography,” *NeuroImage*, vol. 81, pp. 164–177, 2013.

[23] N. X. R. Wang, J. D. Olson, J. G. Ojemann, and R. P. N. Rao, “Unsupervised Decoding of Long-Term , Naturalistic Human Neural Recordings with Automated Video and Audio Annotations,” vol. 10, no. April, pp. 1–13, 2016.

[24] V. N. Vapnik, “An Overview of Statistical Learning Theory,” vol. 10, no. 5, pp. 988–999, 1999.

[25] C. Destrieux, B. Fischl, A. Dale, and E. Halgren, “NeuroImage Automatic parcellation of human cortical gyri and sulci using standard anatomical nomenclature,” *NeuroImage*, vol. 53, no. 1, pp. 1–15, 2010.

[26] M. F. Glasser, T. S. Coalson, E. C. Robinson, C. D. Hacker, E. Yacoub, K. Ugurbil, J. Andersson, C. F. Beckmann, M. Jenkinson, S. M. Smith, and D. C. V. Essen, “Europe PMC Funders Group Europe PMC Funders Author Manuscripts A multi-modal parcellation of human cerebral cortex,” vol. 536, no. 7615,

pp. 171–178, 2017.

[27] R. S. Desikan, F. Se, B. Fischl, B. T. Quinn, B. C. Dickerson, D. Blacker, R. L. Buckner, A. M. Dale, R. P. Maguire, B. T. Hyman, M. S. Albert, and R. J. Killiany, “An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest,” vol. 31, pp. 968–980, 2006.

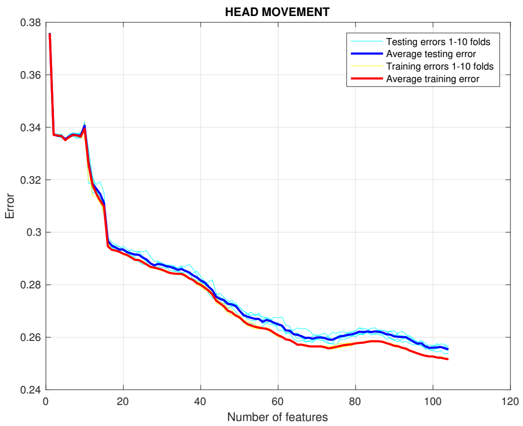
[28] F. Boisgueheneuc, R. Levy, E. Volle, M. Seassau, H. Duffau, S. Kinkingnehun, Y. Samson, S. Zhang, B. Dubois, and P. M. Curie-paris, “Functions of the left superior frontal gyrus in humans : a lesion study,” no. March, pp. 3315–3328, 2018.

[29] S. Japee, K. Holiday, M. D. Satyshur, I. Mukai, L. G. Ungerleider, C. J. Tinsley, N. Trent, T. Liu, and M. State, “A role of right middle frontal gyrus in reorienting of attention : a case study,” vol. 9, no. March, pp. 1–16, 2015.

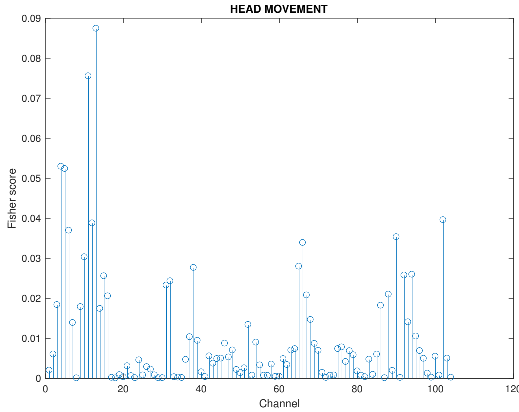
[30] S. American, N. America, and S. American, “new ways of making more powerful computers IT ’ S TIME TO CHANGE THE WAY WE STUDY THE BRAIN .,” vol. 306, no. 6, pp. 50–55, 2018.

5 Appendix

Methodology A

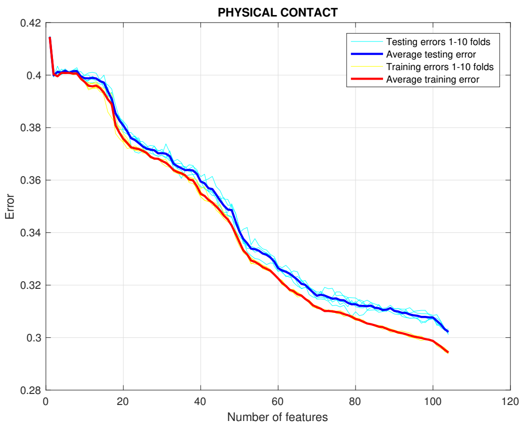


(a) Classification error obtained training a linear classifier with feature RMS and Fisher algorithm.

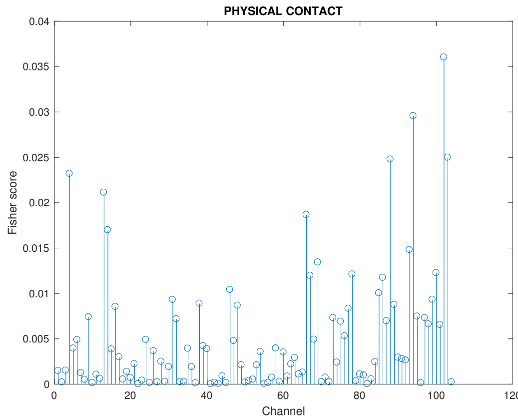


(b) Fisher scores for each electrode.

Figure 26: Results of Fisher analysis

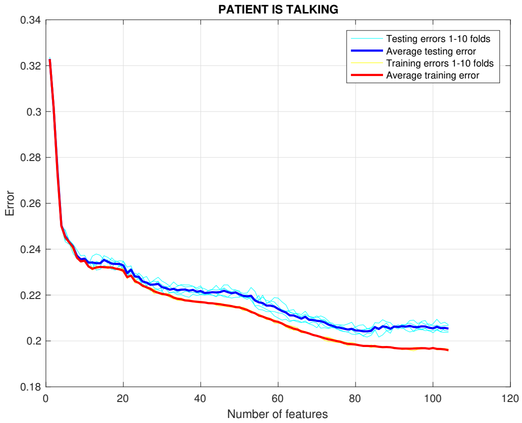


(a) Classification error obtained training a linear classifier with feature RMS and Fisher algorithm.

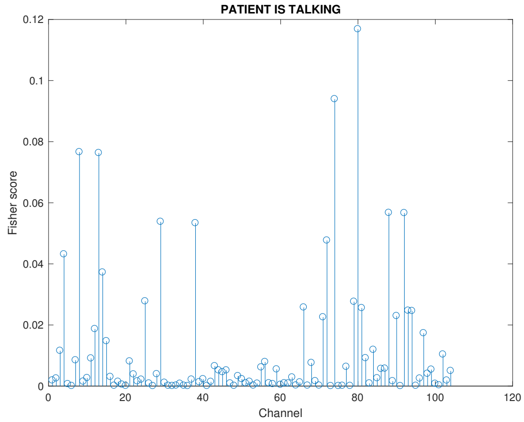


(b) Fisher scores for each electrode

Figure 27: Results of Fisher analysis

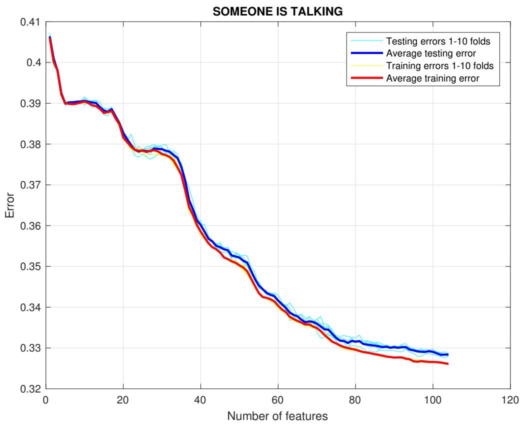


(a) Classification error obtained training a linear classifier with feature RMS and Fisher algorithm.

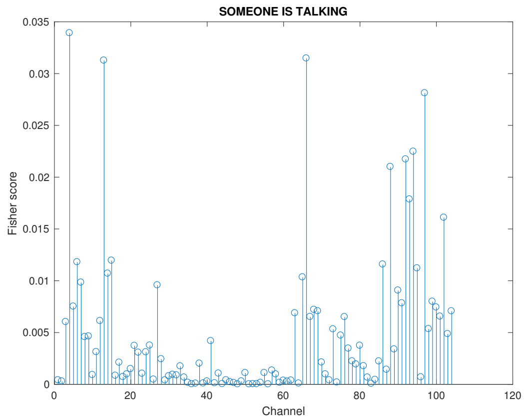


(b) Fisher scores for each electrode.

Figure 28: Results of Fisher analysis

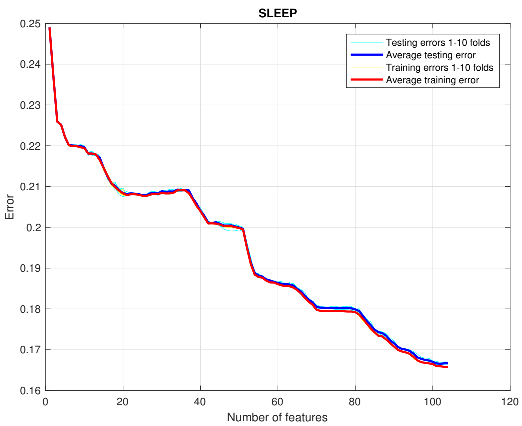


(a) Classification error obtained training a linear classifier with feature RMS and Fisher algorithm.

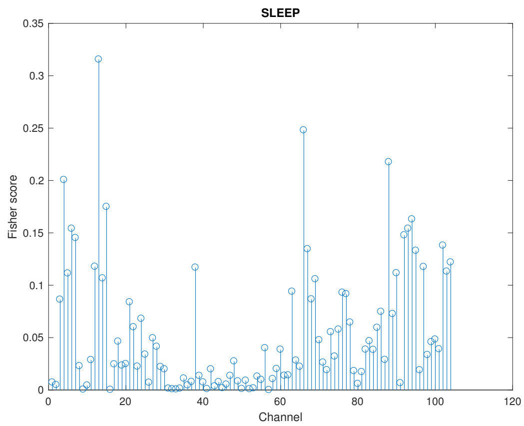


(b) Fisher scores for each electrode.

Figure 29: Results of Fisher analysis

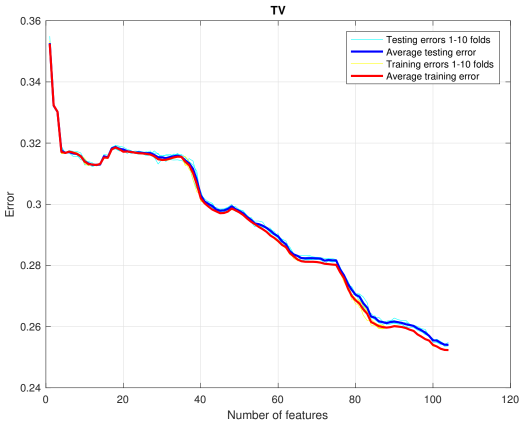


(a) Classification error obtained training a linear classifier with feature RMS and Fisher algorithm.

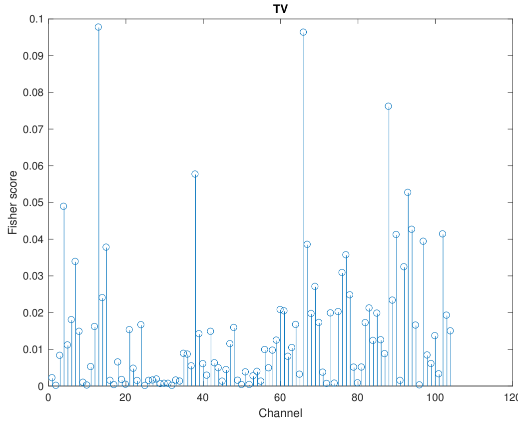


(b) Fisher scores for each electrode.

Figure 30: Results of Fisher analysis

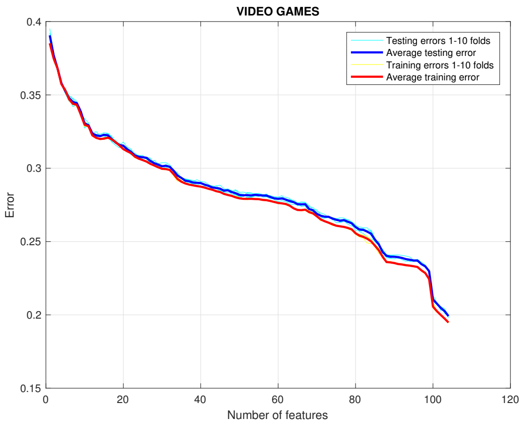


(a) Classification error obtained training a linear classifier with feature RMS and Fisher algorithm.

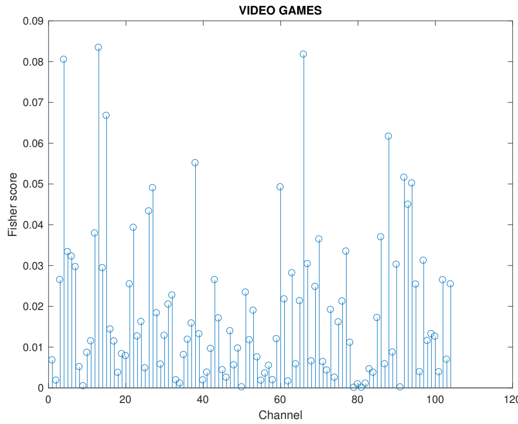


(b) Fisher scores for each electrode.

Figure 31: Results of Fisher analysis



(a) Classification error obtained training a linear classifier with feature RMS and Fisher algorithm.



(b) Fisher scores for each electrode.

Figure 32: Results of Fisher analysis

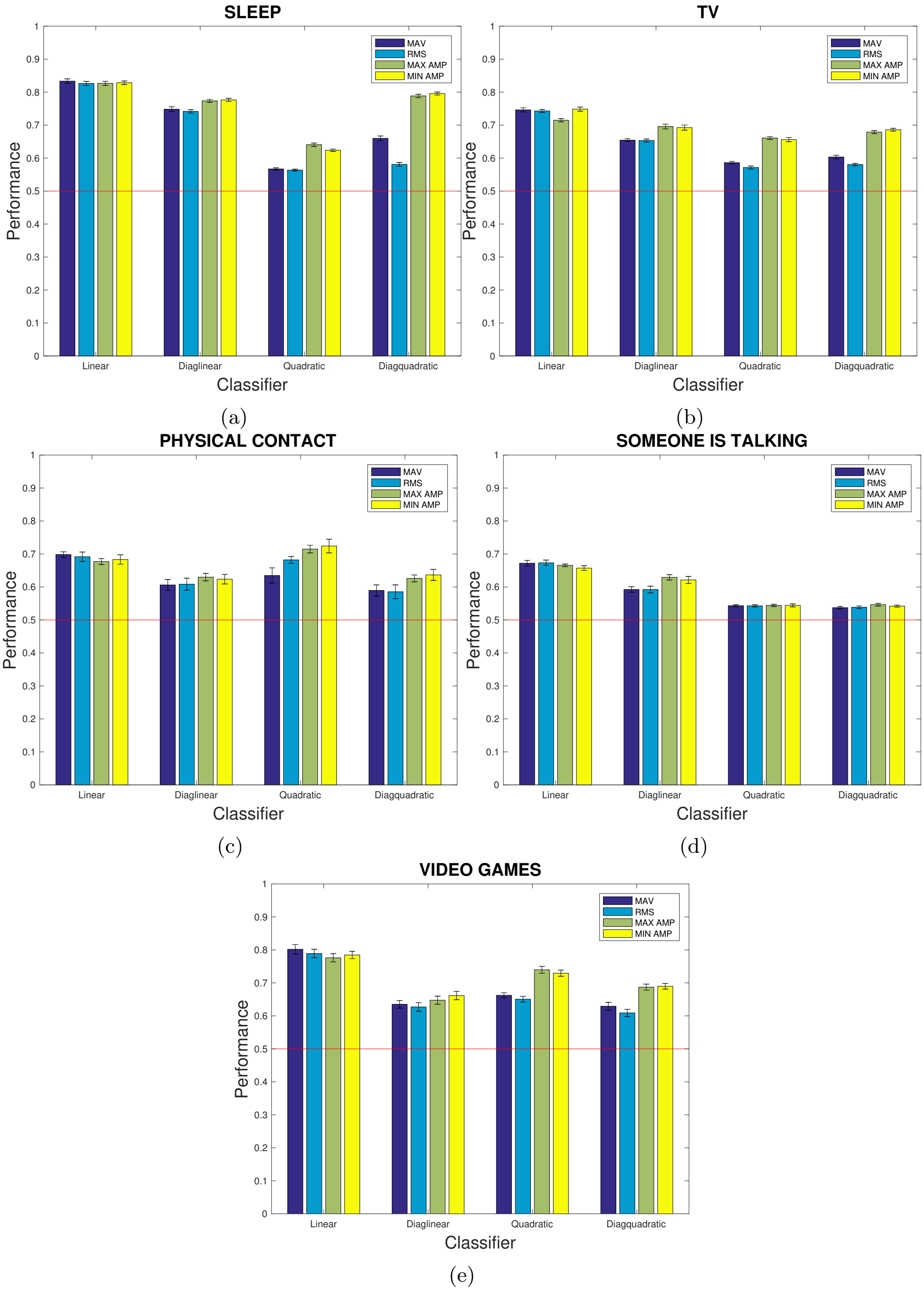


Figure 33: Comparison of the performances of the various type of classifiers and and features (RMS, MAV, Minimal value and maximal value) calculated using a round window of one second.

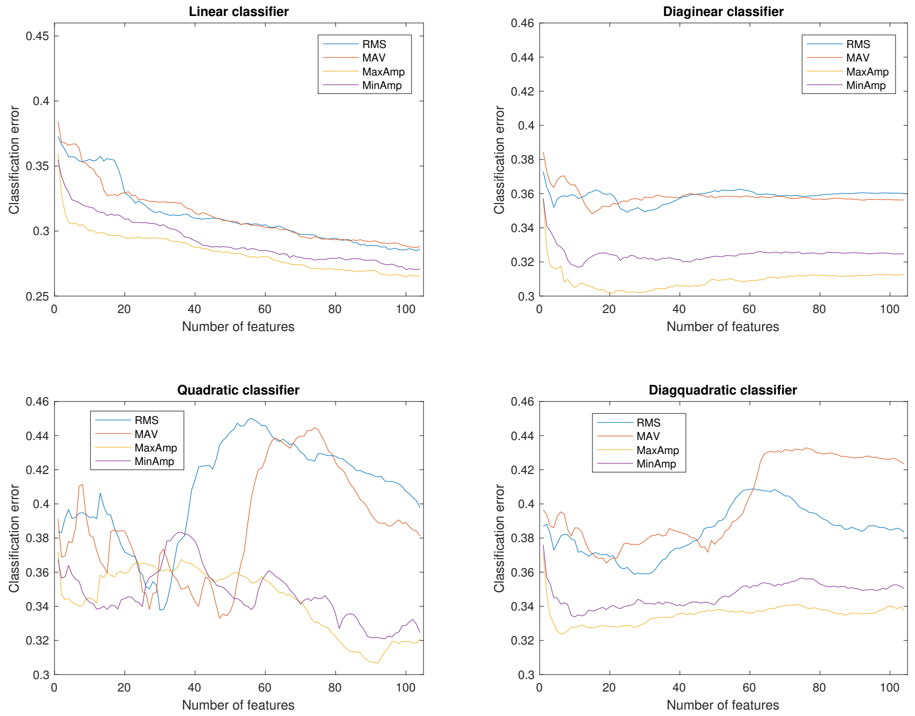


Figure 34: BODY MOVEMENT: Classification error for all the classifiers and all the features according to the number of signals provided to the classifier , which were selected following the ranking provided by the Fisher algorithm.

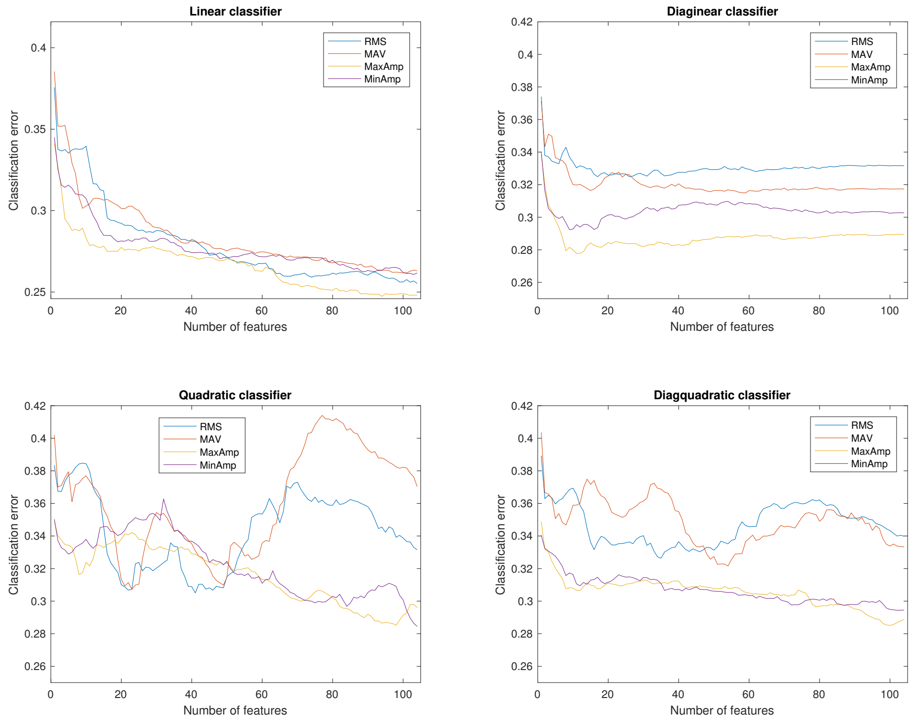


Figure 35: HEAD MOVEMENT: Classification error for all the classifiers and all the features according to the number of signals provided to the classifier , which were selected following the ranking provided by the Fisher algorithm.

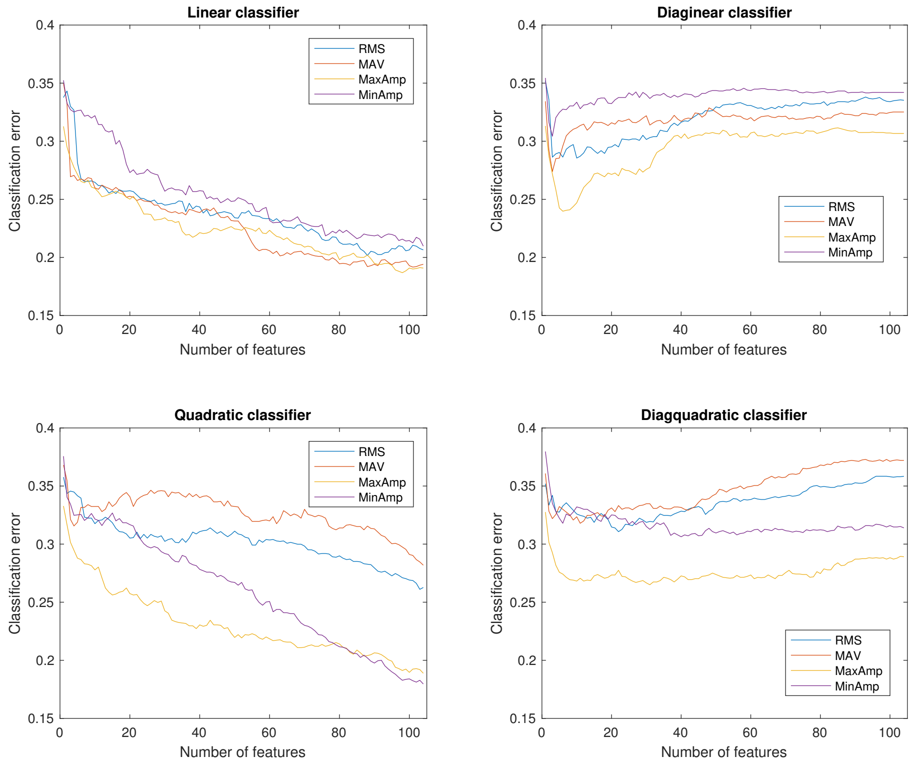


Figure 36: EAT: Classification error for all the classifiers and all the features according to the number of signals provided to the classifier , which were selected following the ranking provided by the Fisher algorithm.

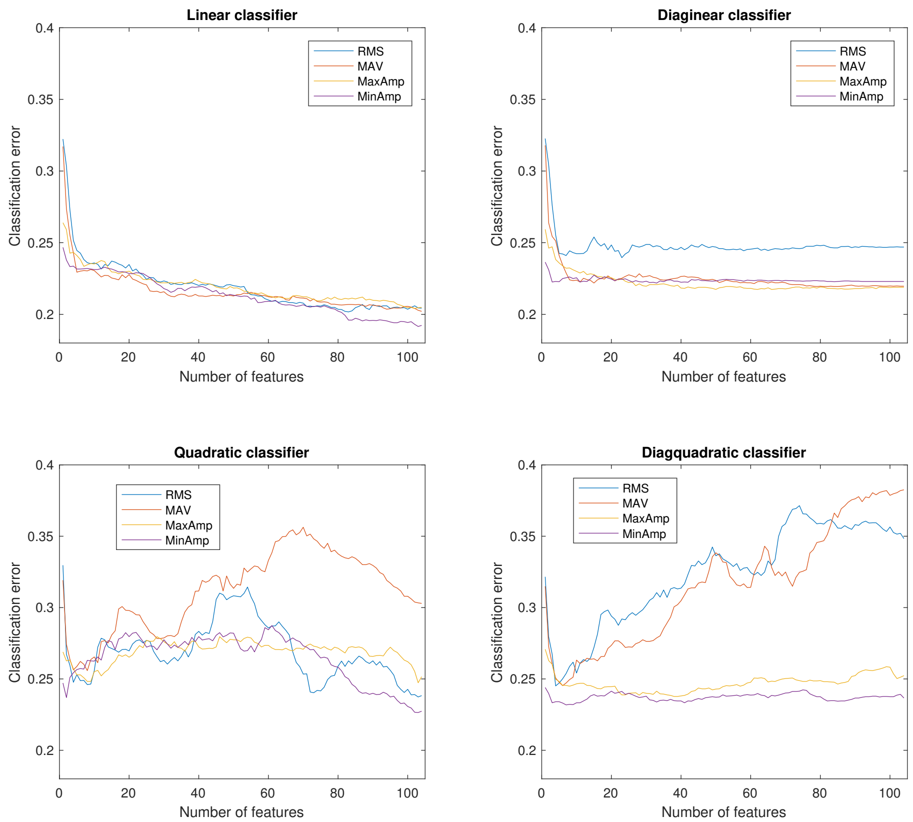
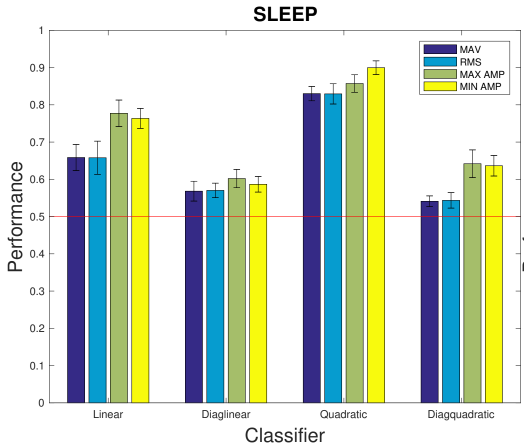
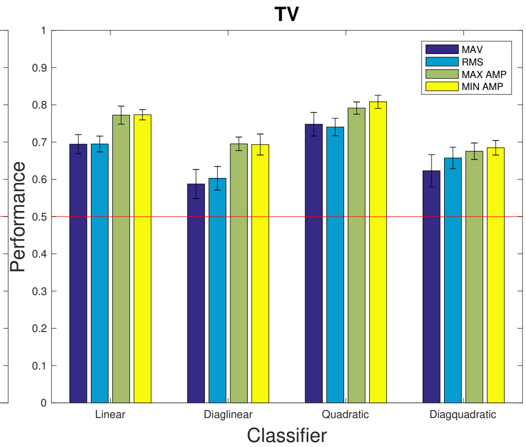


Figure 37: PATIENT IS TALKING: Classification error for all the classifiers and all the features according to the number of signals provided to the classifier , which were selected following the ranking provided by the Fisher algorithm.

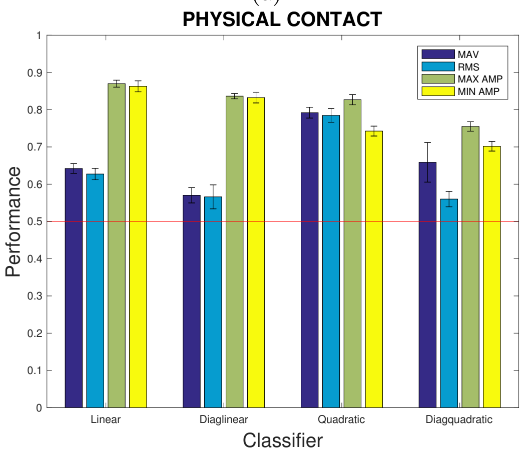
Methodology B1



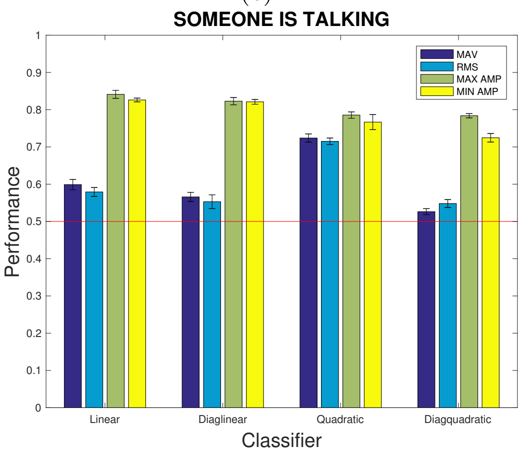
(a)



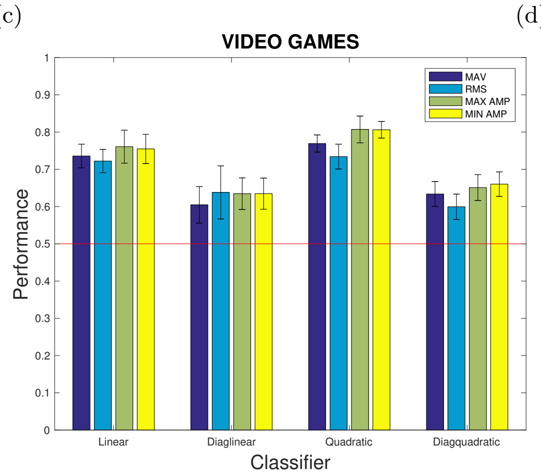
(b)



(c)



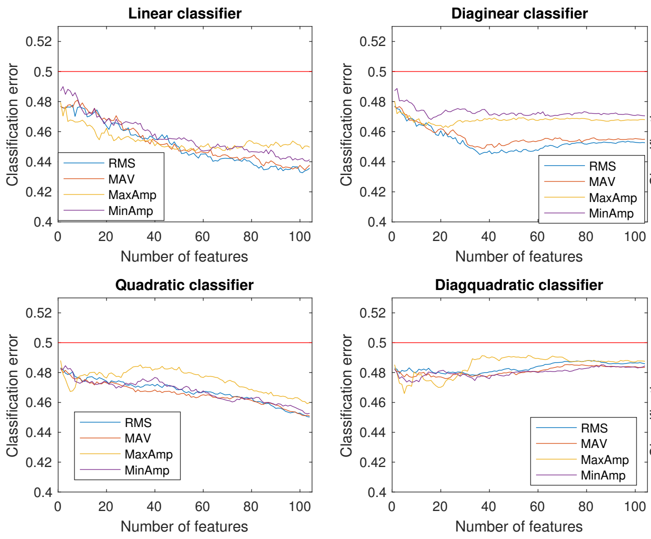
(d)



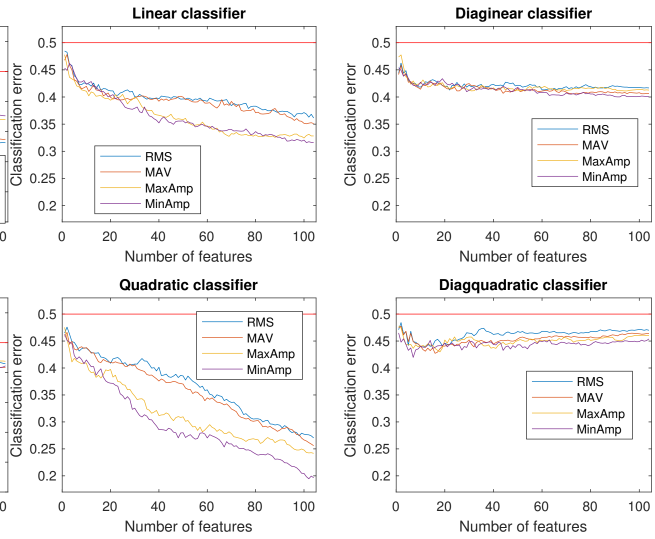
(e)

Figure 38: Comparison of the performances of the various type of classifiers and and features (RMS, MAV, Minimal value and maximal value) calculated using a round window of one second.

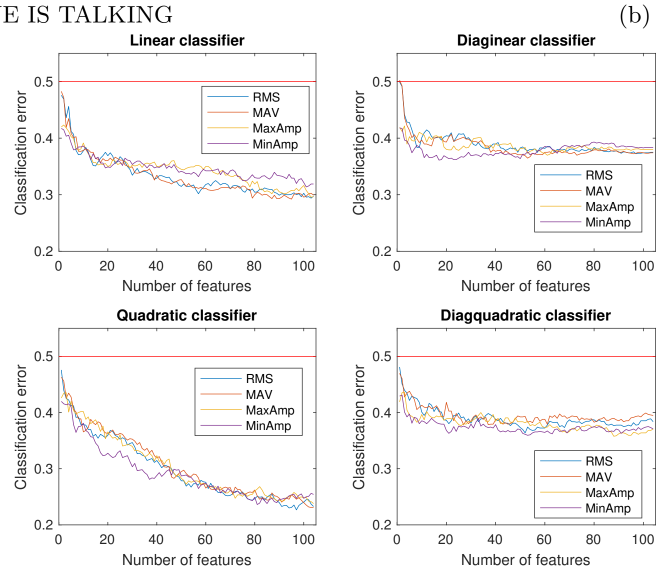
Methodology B2



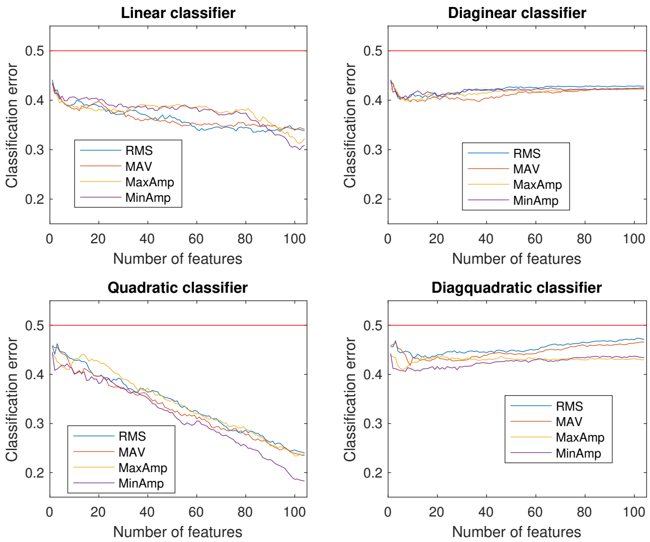
(a) SOMEONE IS TALKING



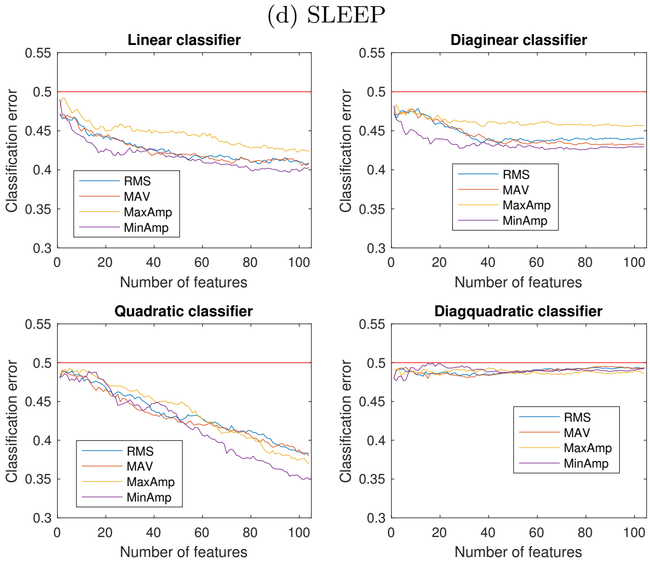
(b) TV



(c) VIDEO GAMES



(d) SLEEP



(e) PHYSICAL CONTACT

Figure 39: Comparison of the classification errors using Fisher algorithm for the types of classifiers analyzed (linear, diaglinear, quadratic and diagquadratic) and the features calculated with a round window of one second. The horizontal red line represents chance level at 50%. The number of features indicated in the x-axis characterizes the number of signals provided to the classifier according to the Fisher scores.

ACRONYM OF THE ELECTRODE AND RESPECTIVE AREA OF THE BRAIN ACCORDING TO DESTRIEUX ET AL.

IP 1 : G_and_S_cingul-Mid-Ant
IP 2-6,11,12,14: G_front_sup
IP 7-8,15,16: G_precentral
IP 9-10: G_and_S_cingul-Mid-Ant
IP 13: G_and_S_paracentral

FP 1,9: G_and_S_cingul-Ant
FP 2-6,10-15: G_front_sup
FP 7-8 16: G_front_middle

LP 1-3 G_pariet_inf-Angular
LP 4-5 12 -13 G_postcentral
LP 6-7 14 G_precentral
LP 8 G_front_sup
LP 9 G_pariet_inf-Angular
LP 10-11 G_pariet_inf-Supramar
LP 15-16 G_front_middle

LA 1,9: G_pariet_inf-Supramar
LA 3-4,10,11: G_postcentral
LA 5 12 13 G_precentral
LA 6-7,14,15: G_front_middle
LA 8,16: G_front_sup

LF 1: G_temporal_middle
LF 2: G_and_S_subcentral
LF 3-4: G_precentral
LF 5,10-12: G_front_inf-Opercular
LF 6-7,14-15: G_front_middle
LF 8,16: G_front_sup
LF 9: G_temp_sup-Lateral
LF 13: G_front_inf-Triangul

IA 1-2: G_and_S_cingul-Ant
IA 3-8: G_front_sup

IM1: G_and_S_cingul-Mid-Post
IM 2-8: G_front_sup

IO 1: G_cuneus
IO 2-5: G_precuneus
IO 6-8: G_and_S_paracentral