CLASSIFICATION OF SEMANTIC MEMORIES USING MULTITAPER SPECTRAL ESTIMATION

SEBASTIAN DALIN-VOLSING

Bachelor's thesis
2015:K7

LUND UNIVERSITY

Faculty of Science
Centre for Mathematical Sciences
Mathematical Statistics

CENTRUM SCIENTIARUM MATHEMATICARUM

Abstract

Classification of semantic memories using multitaper spectral estimation

by Sebastian Dalin-Volsing

The research on classification of semantic memories is still very young. Several methods have been tested ranging from magnetic resonance imaging (MRI) to electrocorticography (ECoG). This report describes an alternative way of classifying signals collected from an electroencephalogram (EEG) into categories using the Thomson multitaper method of spectral estimation, as well as a logistic regression model. The aim for this report is to expand the research field with an approach that complements the current options of classification. Data was distributed from the department of Psychology at Lund University, and the experimental paradigm was to classify three types of semantic memories (faces, landmarks and objects) based on their neural patterns. Based on the cross-validation from the mentioned methods, a classifier could successfully be trained for the "faces" and "landmarks" categories with an average success rate of 55% and 51% respectively. The classifier accurately responded to the onset of the stimuli ($p < 0.001$ for faces, $p = 0.015$ for landmarks). No classifier for the "objects" category could be trained using this method. These results indicate that the multitaper method of spectral estimation can be useful in detecting neural patterns. Several ways to refine these methods are discussed.

Contents

Contents ii

List of Figures iv

List of Tables v

1 Introduction 1
- 1.1 Background . . . 1
- 1.2 Question at issue . . . 2

2 Theory 3
- 2.1 Some basics . . . 3
- 2.2 Stochastic processes . . . 3
- 2.3 Fourier transform . . . 4
- 2.4 Spectral analysis . . . 4
  - 2.4.1 Periodogram . . . 5
  - 2.4.2 Modified periodogram . . . 5
  - 2.4.3 Multitaper . . . 6
  - 2.4.4 Spectrogram . . . 6
  - 2.4.5 Example of spectral analysis . . . 7
- 2.5 Logistic Regression . . . 9

3 Method 11
- 3.1 Experimental paradigm . . . 11
  - 3.1.1 Material . . . 12
  - 3.1.2 Familiarity . . . 12
  - 3.1.3 Test . . . 13
- 3.2 Classification . . . 14
- 3.3 Analysis . . . 14
  - 3.3.1 Model . . . 15
  - 3.3.2 Feature Selection . . . 15

4 Results 17
- 4.1 Results . . . 17
  - 4.1.1 Multitaper vs. Modified periodogram . . . 17
  - 4.1.2 Grid comparison . . . 18
  - 4.1.3 Bandwidth comparison . . . 20
  - 4.1.4 Perception vs. imagery . . . 21

5 Discussion **23**
5.1 Discussion 23
5.1.1 Analysis 23
5.1.2 Possible errors 25
5.1.3 Future studies 25
5.1.4 Conclusion 26

Bibliography **27**

List of Figures

1.1 Classifier performance as a function of time after stimulus (Morton et al, 2012). 2

2.1 An example signal 3
2.2 The signal from EEG of channel C3 (top) and O2 (bottom) 7
2.3 The periodogram estimate for C3 (top) and O2 (bottom) 8
2.4 The multitaper estimate for C3 (top) and O2 (bottom) 8
2.5 Spectrogram of C2 8

3.1 The different electrodes that were used. 11
3.2 A scheme describing the experimental paradigm. 14
3.3 The channels included in the feature selection 16
3.4 The different sized grids, big (left) and small (right) 16

4.1 Classifiers when using multitaper method (N=64, K=8, small grid) 18
4.2 Classifiers when using multitaper method (N=64, K=8, large grid) 19
4.3 Confusion matrix for small grid. 19
4.4 Confusion matrix for large grid. 20
4.5 Confusion matrix for small grid, K=4. 21
4.6 Classifiers when using multitaper method (N=64, K=4, small grid) 21
4.7 Classifiers when using multitaper method (N=64, K=4, small grid) with imagery features. 22

List of Tables

4.1 Amount of features selected in multitaper and periodogram. 17
4.2 Trials of different N and K using the small grid. 20

Chapter 1

Introduction

1.1 Background

We don't know that much about the brain, and even though many groups of scientists around the world have spent a lot of time trying to find neurological markers to cognitive processes, it's still hard to come to conclusions.

Research on the human memory is one of the research areas where progress has been made. With help of magnetic resonance imaging (MRI) it has been found that the brain structure Hippocampus is involved in many of the processes of the creation of memories (Eichenbaum, 2000). Today it is commonly believed that memory is created through Hippocampus, and that this structure keeps indexed representations of patterns in the cortex[1]. This index is then used to recollect memories by recreating the same pattern. The activation of patterns in cortex is what's considered as the actual memorization.

Exactly what happens during the learning process is still partly unknown. In this report we aimed to find out if it was possible to classify different types of semantic memories with use of an electroencephalogram (EEG). There are different approaches to finding neural correlates of cognitive functions, a few being PET (positron emission topography) or f-MRI scans. Although f-MRI scans are a really good way of finding neural correlates to cognitive functions, it still has some well-known flaws, one being the rather low temporal resolution.

With f-MRI studies we can pinpoint where neurons are activated a lot better than what's possible with EEG, but with EEG it's instead possible to locate the time-span of when a certain activation of neurons is present. Since f-MRI studies on classification already

[1]The cerebral cortex is the brain's outer layer of folded neural tissue, where most of our cognitive processes just as attention, thought and conciousness

has been made, and has been shown to separate patterns between groups of semantic memories, it can be argued that the same would be possible for EEG, although with a worse spatial resolution. With that in mind, we can instead focus on the temporal resolution, and how the frequency power changes of time in the different areas of the brain.

An experiment concerning EEG and classifier training was conducted by Morton et al, 2012. In their study, they managed to make classifications with help of EEG.

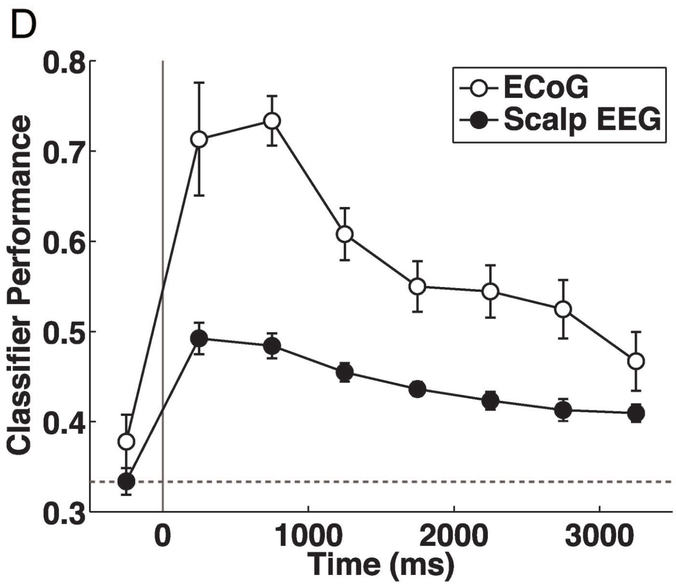

Figure 1.1: Classifier performance as a function of time after stimulus (Morton et al, 2012).

It was shown that the accuracy of the classifier became stronger when the stimulus material was presented. This is of course a very intuitive (but nevertheless important) finding. This means that EEG is able to pick up changes in neural patterns which corresponds to the memory of a certain stimulus.

The standard when making statistical analysis of these kinds of experiments is to use a mathematical model called MVPA (multi-voxel pattern analysis). The aim for this report is to complement the results from the MVPA with results from other mathematical methods explained in the next chapter.

1.2 Question at issue

The question at issue for this report is finally: "Is it possible to classify different semantic memories from EEG signals using the multitaper method of spectral estimation?"

Chapter 2

Theory

2.1 Some basics

Some few things that needs to be said before is that a signal can be defined as "a function that conveys information about the behaviour or attributes of some phenomenon" (Priemer, 1991). In this case, the phenomenon is the activity of the neurons in the brain. What the spectrum analysis aims for is to convert the signals to the frequency domain, in which the underlying information about the power of the different frequencies in the signal is revealed.

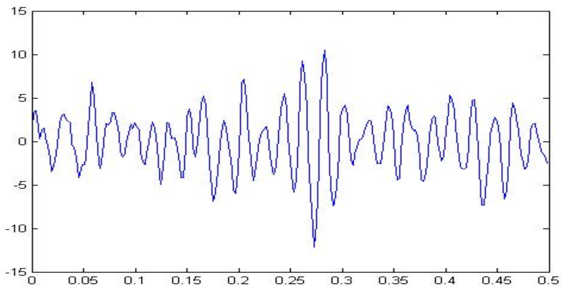

FIGURE 2.1: An example signal

2.2 Stochastic processes

In this report it is assumed that the signal behaves in the form of a stochastic process. In order to analyse the signal we will need to look at a realisation of the signal that has finite length.

Let X_n be a stochastic process with mean zero and a continuous spectrum $R_x(f)$ where f denotes frequency. If $\{x(t),\, t = 0, 1, 2, 3, ..., n-1\}$ is a sequence of sample data from X_n, then $x(t)$ is considered to be a realisation of X_n.

2.3 Fourier transform

If $x(t)$ is a integrable function $x : \mathbb{R} \to \mathbb{C}$ then the Fourier transform of x is defined as:

$$\mathcal{X}(f) = \int_{-\infty}^{\infty} \mathrm{x(t)} e^{-i2\pi f t} \mathrm{d}t, \tag{2.1}$$

In context, $\mathcal{X}(f)$ is the frequency-domain representation of the time-domain represented $x(t)$.

2.4 Spectral analysis

In order to convert the signal to the frequency domain we first need to define the power spectral density (PSD) of a signal. The meaning of PSD is that it resembles the distribution of variance of the process over the different frequencies.

If the covariance function $r(\tau)$ of a stationary process[1] $X(t)$, $t \in \mathbb{R}$ is continuous, there exists a positive, symmetric and integrable function $R(f)$ such that:

$$r(\tau) = \int_{-\infty}^{\infty} e^{-i2\pi f \tau} R(f) df \tag{2.2}$$

The expression (2.2) is called the spectral representation of $r(\tau)$ and $R(f)$ is the spectral density of the covariance function $r(\tau)$ of the process $X(t)$.

If the spectrum is continuous, then the spectral density is given by the Fourier inversion formula:

[1]A stationary stochastic process, or stationary process, is a process whose joint probability distribution does not change when shifted in time. All processes in this report are considered to be stationary.

$$R(f) = \int_{-\infty}^{\infty} e^{i2\pi f\tau} r(\tau) d\tau \tag{2.3}$$

Combining equation (2.2) and (2.3) implies that $R(f)$ and $r(\tau)$ is a Fourier transform pair.

2.4.1 Periodogram

Let $\{x(t),\, t = 0, 1, 2, ..., n-1\}$ be a sequence of real-valued data. Assume this data is a realisation of the stationary process $\{X_n, n \in \mathbb{Z}\}$.

Then the estimate of the spectral density $R_x(f)$ of the realisation $x(t)$ is defined as:

$$\hat{R}_x(f) = \frac{1}{n} |\mathcal{X}(f)|^2 \tag{2.4}$$

where

$$\mathcal{X}(f) = \sum_{t=0}^{n-1} x(t) e^{-i2\pi f t} \tag{2.5}$$

is the Fourier transform of the data vector $x(t)$.

2.4.2 Modified periodogram

Because of the properties of the periodogram (not being a good estimator of the PSD due to its high spectral leakage), we need to modify it. If we introduce a window function $w(t)$ of length N which has certain properties:

1. $w(t)$ is even (i.e., $w(-t) = w(t)$)
2. $w(t) = 0$ for $|t| \geq M$, where $M < N$
3. $w(0) = 1$
4. $w(t)$ decays smoothly to zero with t

we can define the modified periodogram as in (2.2) but with

$$\mathcal{X}(f) = \sum_{t=0}^{n-1} w(t)x(t)e^{-i2\pi ft} \tag{2.6}$$

being the modified periodogram Fourier transform of the data vector $x(t)$.

2.4.3 Multitaper

When we use periodogram method to estimate the spectral density of a signal we often assume that the coefficients we get from the Fourier transform will be a good estimation of the amplitude of the corresponding frequency. This is not always a valid assumption, since the representation $x(t)$ might include a lot of noise.

By averaging over many independent calculations of the spectral density for the same sample we can reduce the variance.

The Thomson multitaper method is defined as follows:

$$\hat{R}_x(f) = \frac{1}{K}\sum_{k=1}^{K} \hat{R}_{x,k}(f) = \frac{1}{K}\sum_{k=1}^{K} |\sum_{t=0}^{n-1} x(t)h_k(t)e^{-i2\pi ft}|^2, \tag{2.7}$$

where $\{h_k(t)\}$, $k = 1, .., K$ are the Slepian sequences. The reason to why we use this sequence of window functions is that they are all pairwise orthogonal to each other which in turn gives a variance reduction to the spectrum.

In this report we will talk about the values of N and K, where N is the length (in samples) of $h_k(t)$ and K is the number of orthogonal window functions used.

2.4.4 Spectrogram

Another way to see the behaviour of a signal is to see how the power of the frequencies change over time. The spectrogram lets us compute the spectrum for smaller segments of the signal with large overlap. In this report we use an overlap of $N - 4$ samples. This means that for each new segment, we will have only 4 changed samples from the previous segment.

The spectrogram S is defined as:

$$S(t, f) = |\mathcal{X}(t, f)|^2 = |\sum_{\tau=0}^{n-1} x(\tau)w(\tau - t)e^{-i2\pi f\tau}|^2. \qquad (2.8)$$

2.4.5 Example of spectral analysis

To illustrate how this can be applied to real-data we use EEG data from a study where a subject was placed in a chair with closed eyes in a silent dimmed light laboratory. Then a flickering light of 15 Hz, (Grass Photic stimulator Model PS22C), was introduced for a 5 s interval at a distance of 1 m.

In this example the channels C3 and O2 are well separated. First we look at the signals collected from this 5 second period where the light was introduced (Fig. 2.2).

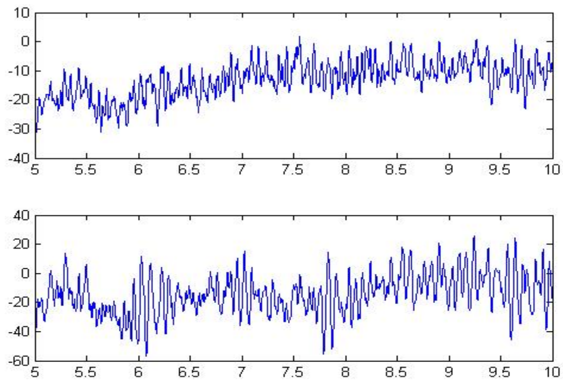

FIGURE 2.2: The signal from EEG of channel C3 (top) and O2 (bottom)

Fig. 2.3 shows the periodogram power spectral density estimate for these signals and Fig. 2.4 shows the power spectral density estimate using the multitaper method.

If we calculate the spectrogram for the signal we can in Fig. 2.5 see an increase of power in the frequency 15 Hz in the channel C2 during the time period 5-10 seconds, when the flickering light was introduced.

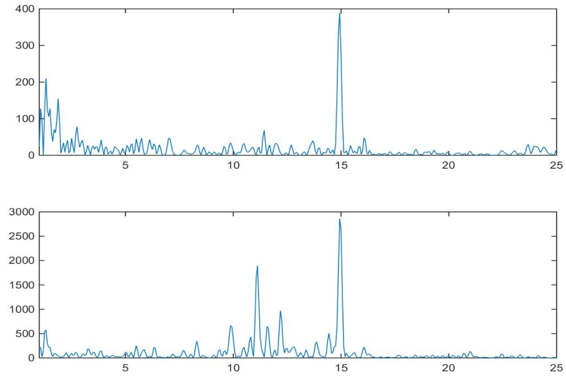

FIGURE 2.3: The periodogram estimate for C3 (top) and O2 (bottom)

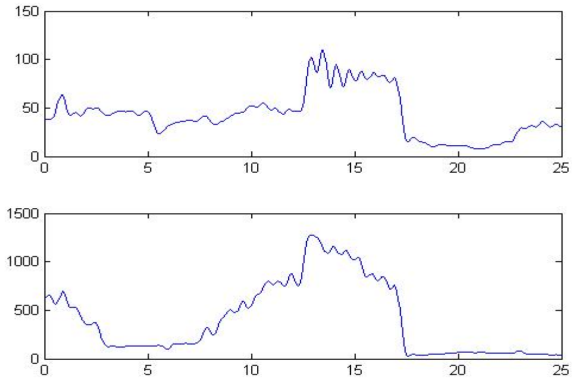

FIGURE 2.4: The multitaper estimate for C3 (top) and O2 (bottom)

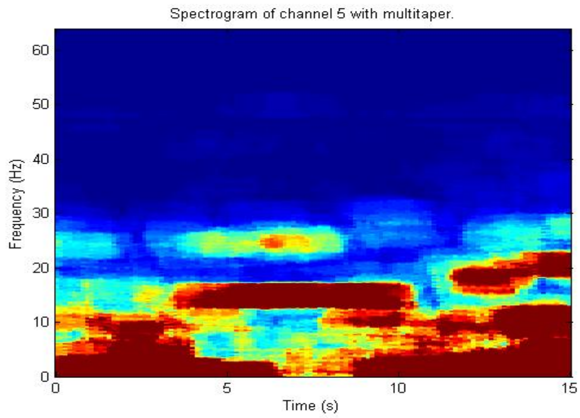

FIGURE 2.5: Spectrogram of C2

2.5 Logistic Regression

In order to be able to make a classifier for the different categories, we also need to introduce a model that can handle this kind of classification problem. A logistic regression model has a binary response variable that depends on some predictor variables (features). In this way categorical responses with possible values 0 or 1 can be modelled using a regression model (which can not be done with "normal" linear regression).

A multivariate linear regression model takes the form of:

$$Y = \beta_0 + \beta_1 X_1 + ...\beta_k X_k, \tag{2.9}$$

where β_j's $\{j = 0, ..., k\}$ are assumed to be non-random, and X_j's $\{j = 1, ..., k\}$ (our features) are assumed to have a linear relationship with our response variable Y (which in this case is real-valued).

However, when we have a categorical response variable like:

$$Y = \begin{cases} 0 & \text{if trial i doesn't belong to the given category} \\ 1 & \text{if trial i belongs to the given category,} \end{cases}$$

we need to make some transforms to able to work with this response variable in the same way as we do in the linear case.

We are in some way interested in the probability p that our response takes the value 1, or in mathematical form $p = P(Y = 1)$. Since probabilities only takes values in [0,1] we introduce the odds function

$$Odds(p_i) = \frac{p_i}{1 - p_i}, \tag{2.10}$$

where i is the i'th observation of our predictor variables $X_{1,i}, ..., X_{k,i}$. The odds function can take values from [0,∞] and works better as a response variable than our previous binary one. However we would also like our response to be able to take negative values. This can be done by a logarithm transform. The final regression model is:

$$log(Odds(p_i)) = \beta_0 + \beta_1 X_{1,i} + ...\beta_k X_{k,i}, \tag{2.11}$$

with the same assumptions as for the linear case. From this, the probability p_i can be derived to be:

$$p_i = P(Y_i = 1) = \frac{e^{\beta_0 + \beta_1 X_1 + ... + \beta_k X_k}}{1 + e^{\beta_0 + \beta_1 X_1 + ... + \beta_k X_k}}. \tag{2.12}$$

The β_j's $\{j = 1, ...k\}$ in this case are interpreted as increase in log-odds for each increase by 1 for the corresponding X_j.

Chapter 3

Method

3.1 Experimental paradigm

The subjects were asked to do a three-step experiment where the different steps were familiarity, study and recall[1].

Before these tests, the cap of electrodes were attached to the head of the subjects, and they signed a form of consent. The distribution of electrodes in the cap can be seen in Fig. 3.1.

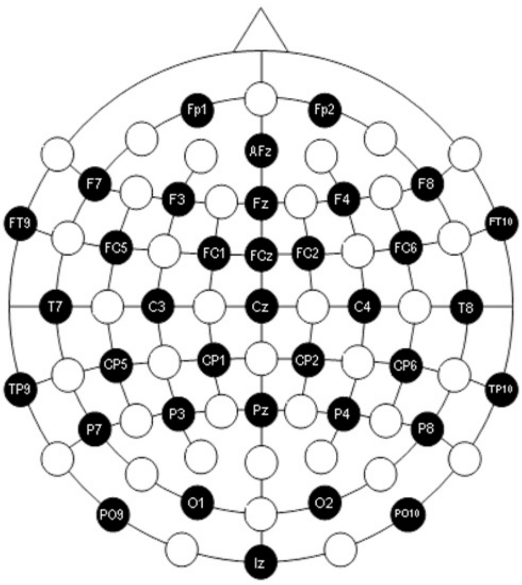

FIGURE 3.1: The different electrodes that were used.

The paradigm was to let the subjects encode abstract words together with one target stimuli of one of the specific semantic categories "faces", "landmarks" or "objects". The

[1]All data in this report is collected from the study phase.

same cue words was then going to be shown in the recall phase to see if the subjects remember what stimuli this specific cue was paired with in the study phase. The idea was to see if the classifier would serve as a pointer to if the subject would remember the correct stimuli or not.

The classifier was trained in the study phase and is considered to be unique for each person.

Everything was done in the program E-Prime.

3.1.1 Material

The material of the experiment itself was handed by the department of Psychology at Lund University. The experiment was conducted by students at the department of Psychology in Lund.

A total of 192 abstracts words were divided into three sets with similar length, frequency and concreteness. These 192 words were to be the associates to the target stimuli. The 192 target stimuli were divided into three categories (faces, landmarks and objects).

The chosen stimuli were faces of famous people (for the face category), well known buildings or sceneries (for the landmark category) and everyday objects (for the object category). When the paradigm was programmed into E-Prime, each of the 192 abstract words from before were associated with three different stimuli (one from each category) to prevent random associations. Each participant were presented with 192 different pair associates, 64 from each stimulus category.

3.1.2 Familiarity

In the first step of the experiment, the subjects were asked to rate their familiarity with the different targets. They were shown a picture of the target stimuli (face, landmark or object), together with the name of the target.

The scale was rated as follows:

1. I don't know this at all.
2. I know very well about this.
3. I recognise the picture but not the name.
4. I recognise the name but not the picture.

This was later used to exclude data from the collection.

3.1.3 Test

After the familiarity phase, the subjects entered the testing. This phase was divided into one study phase and one recall phase. In the study phase, the subjects were introduced to the encoding task of the experiment. The recordings of EEG started with this phase, and the subjects were put inside a Faradays cage to prevent disturbances of the measurements.

The encoding task proceeded as follows:

1. Inter-stimuli cross-hair (1 s)
2. Abstract word-cue (2 s)
3. Inter-stimuli cross-hair (1 s)
4. Abstract word-cue with target stimuli-context (2 s)
5. Inter-stimuli cross-hair (1 s)
6. Abstract word-cue (2 s)
7. Question, active rating 1-3 (6 s)

During part 3-4 and 5-6 in this encoding task the classifier was trained, by extracting the signals from the EEG and grouping it into the appropriate semantic category. The classification for part 3 and 4 is from the "perception" condition, where the subject could see the target, and the classification for part 5 and 6 is the "imagery" condition.

In part 6 the subject was asked to picture the target in their mind. This was done to separate between perception and imagery, to see which of the two would serve best as a classifier. This means that a total of 6 different kinds of classifiers was trained for each subject. One for each semantic category (face, landmark and object), and one for each of the conditions "perception" and "imagery".

After each pair-associate the subjects were asked to rate the difficulty of the association between word and target. The subjects got 6 seconds to answer. The question was phrased: "How simple was it to associate these two objects?" and the choices were:

1. Very hard (almost impossible)

2. It was OK

3. Very easy (they are normally already associated)

Figure 3.2 shows a time scheme of the experiment. The experiment was held in Swedish.

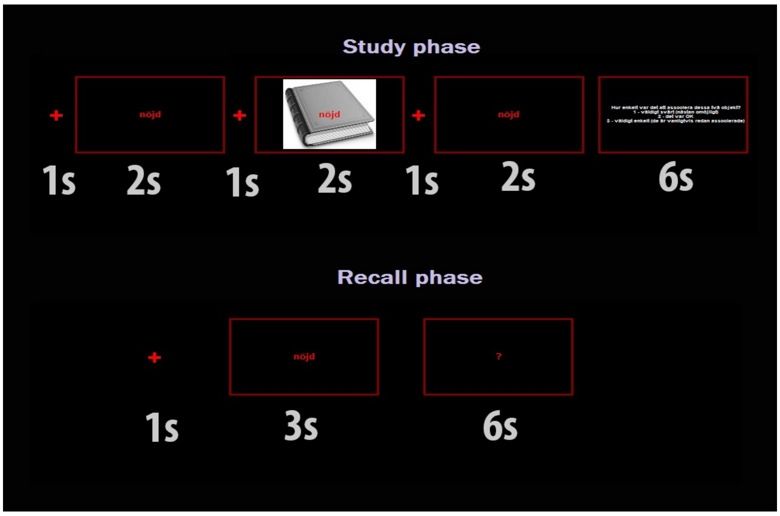

FIGURE 3.2: A scheme describing the experimental paradigm.

There was also a recall part where the person was tested on these pair-associates, but this report focuses only on the training of classifier.

3.2 Classification

The aim for the classification of different semantic memories was to try to find differences in neural patterns that can be seen in the signals from the EEG. These signals are obtained in the study-phase of the experiment and analysed and compared with the mathematical methods described in Chapter 2.

3.3 Analysis

The data was pre-processed at the department of Psychology before it was ready for analysis. The data was run through a band-pass filter, to remove unwanted frequencies[2]. Then all the trials with artifacts (blinking, yawning, moving of the head etc.) were removed from the data.

[2]The wanted frequencies in this report ranged between $\approx 3 - 45$ Hz.

In the analysis we agreed on focusing on the spectrograms to see what differences there were between the different categories. We started by computing a mean spectrogram for each channel. This was made by averaging over all trials in the same category. Then we divided the mean spectrograms into smaller blocks. T-tests were made on smaller time-frequency-blocks of the mean spectrograms comparing the same blocks for the different categories. The blocks that were significantly different from each other was then added to the model of that specific classifier.

3.3.1 Model

The model that was used for the classifiers was a logistic regression model. Only the significant previously computed time-frequency blocks was used as the parameters in the model. In the case of having a total of k significant parameters, a classifier for the category "Faces" would look like this:

$$P(Y = Face) = \frac{e^{\beta_0+\beta_1 X_1+\ldots+\beta_k X_k}}{1+e^{\beta_0+\beta_1 X_1+\ldots+\beta_k X_k}} \tag{3.1}$$

The same kind of model was made for the categories "Landmarks" and "Objects". R was used for computation of the estimates of $\hat{\beta}_j$, $j = \{1, \ldots, k\}$. Here X_j, $j = \{1, \ldots, k\}$ corresponds to the mean power of the corresponding j:th selected time-frequency block.

3.3.2 Feature Selection

In order to add only the important features X_j into the model in equation 3.1, a few selections were made.

Firstly, a logistic model like the one above require the features to be independent from each other, so the first step was to reduce the number of channels used. The channels were chosen so that they would be far from each other, but covering the entire scalp (see Fig. 3.3).[3].

In each one of these channels, a spectrogram was made for each of the 64 trials[4] in each classifier. Then the spectrogram was averaged over each of the trials to make one

[3]Initially, we wanted to use the channels F7 and F8 instead of FC5 and FC6, but due to noise problems with channel F8, we decided to exclude that one from the selection of features.

[4]In our case, five of the trials in the categories faces and landmarks, and three for the object category was rejected due to artifacts (like blinking, moving, yawning etc). In order to balance between the categories, the last two trials of the object category was removed.

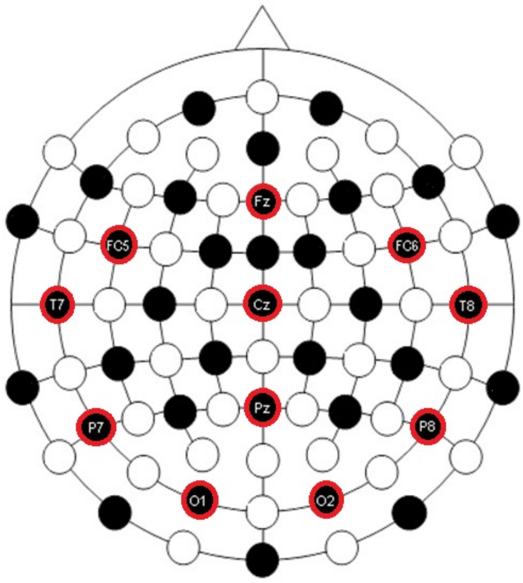

FIGURE 3.3: The channels included in the feature selection

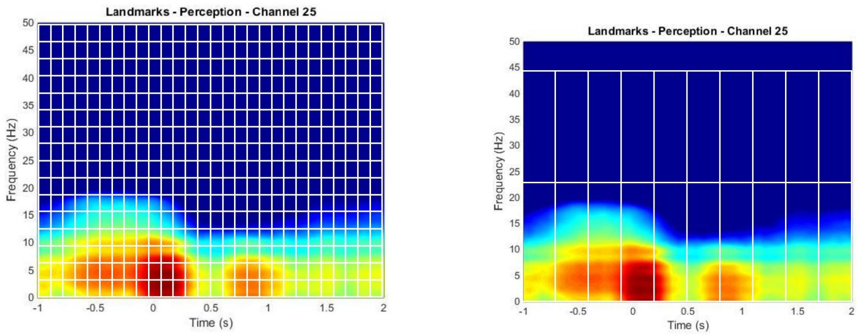

FIGURE 3.4: The different sized grids, big (left) and small (right)

average spectrogram for the categories "faces", "landmarks" and "objects" in a specific channel.

Then, for each channel, the mean spectrogram was divided into a grid. Different grids were tested, with different sizes (see Fig. 3.4). These sizes was chosen based on the bandwidth $B = \frac{K+3}{N}$. It was argued that our choice of B would limit the choices of grid.

Now T-tests were made between the different classifiers for "faces", "landmarks" and "objects" in the same grid to see if the values in there have the same mean. From there, only the significantly different features were selected for our model. In that way, we reduce the dependency further.

Chapter 4

Results

4.1 Results

The results in this section are based on the cross-validation that was made to compare the models. The cross-validation technique that was used was "Repeated random sub-sampling validation", where 5 random trials were left out of the feature selection and then tested on the models, to see the accuracy of the classifiers. This was repeated 10 times.

All data used in the result section is from one participant only.

4.1.1 Multitaper vs. Modified periodogram

When the modified periodogram method was used, much less of the data was collected as features. From this, it was decided to only use the multitaper method (N=64 and K=8) as spectral estimate, using a p-value of 0.001 for the t-tests between categories.

Table 4.1: Amount of features selected in multitaper and periodogram.

Spectral Estimate	p-value	% features selected
Multitaper	0.001	$\approx 70\%$
Multitaper	0.05	$\approx 83\%$
Mod. Periodogram	0.001	$\approx 25\%$
Mod. Periodogram	0.05	$\approx 35\%$

4.1.2 Grid comparison

Firstly, the results of the perception category, where the targets were presented as an image, using the smaller grid.

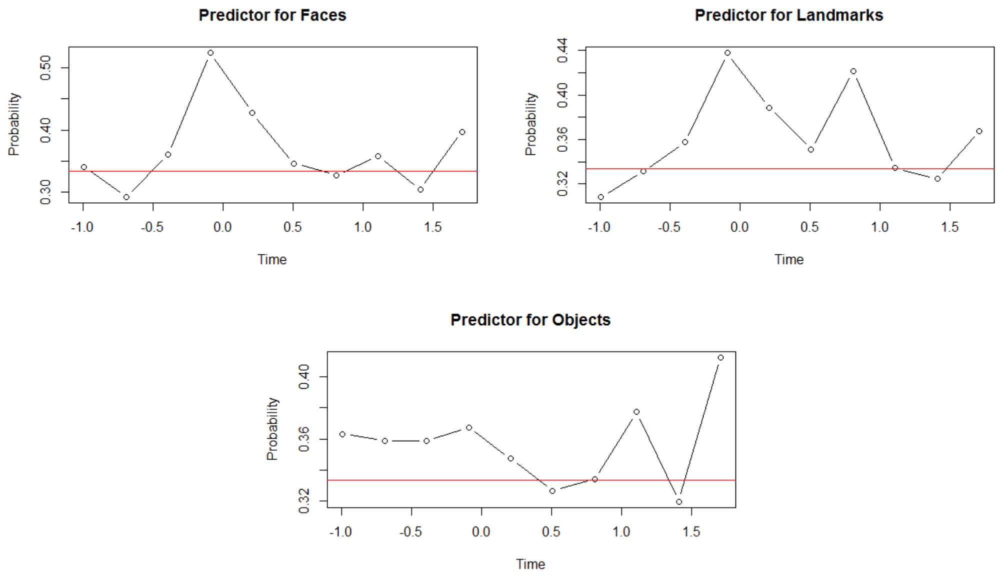

FIGURE 4.1: Classifiers when using multitaper method (N=64, K=8, small grid)

The red line in the accuracy plots is the probability of $\frac{1}{3}$, which indicates chance. For the predictors of "face" and "landmark" we have a significant difference of accuracy for the classifier when the stimuli is shown at $T = (-0.1s, 0.2s)$ ($p < 0.001$ for faces, $p = 0.015$ for landmarks). However for objects, that classifier is not significantly different when the stimuli is shown ($p = 0.684$).

The same tests for the large grid gave a significant difference at stimulus onset $T = (-25\text{ms}, 50\text{ms})$ for the faces category only ($p = 0.011$). The p-values for the landmark category and the object category in this classifier was $p = 0.316$ and $p = 0.171$ respectively. This is shown in Fig. 4.2.

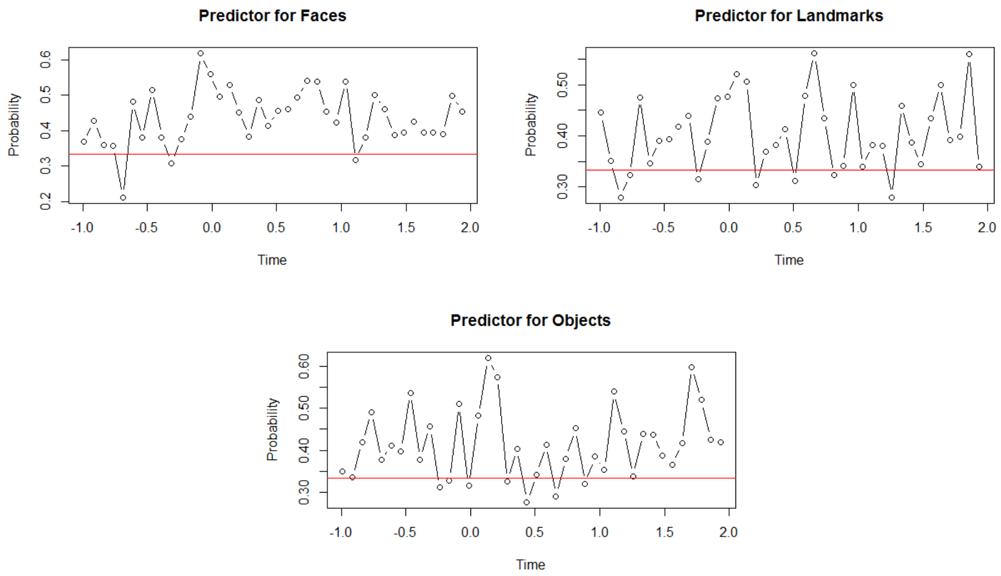

FIGURE 4.2: Classifiers when using multitaper method (N=64, K=8, large grid)

When the cross-validation was made, we also let the classifier predict what type of signal it was. This was made by comparing the probabilities for the different categories, and picking the one that was the largest. The results are as follows:

True class

Predicted class	Face	Landmark	Object
Face	28 (55%)	14 (27%)	15 (28%)
Landmark	10 (20%)	26 (51%)	18 (34%)
Object	13 (25%)	11 (22%)	20 (38%)
Total	51 (100%)	51 (100%)	53 (100%)

FIGURE 4.3: Confusion matrix for small grid.

Large grid:

True class

Predicted class	Face	Landmark	Object
Face	25 (50%)	12 (24%)	14 (28%)
Landmark	17 (34%)	13 (26%)	21 (42%)
Object	8 (16%)	25 (50%)	15 (30%)
Total	50 (100%)	50 (100%)	50 (100%)

FIGURE 4.4: Confusion matrix for large grid.

What we want to see in these confusion matrices are large numbers on the top-left to bottom-right diagonal. They are the correct guesses of the classifiers.

4.1.3 Bandwidth comparison

The next step was to try the smaller grid for different values of N and K, to see if the choice of bandwidth $B = \frac{K+3}{N}$ has any impact on the selection of features. The results are shown in table 4.2.

TABLE 4.2: Trials of different N and K using the small grid.

Trial	N	K	B (Hz)	Total Accuracy
1	64	8	22	47.7%
2	64	4	14	48%
3	128	8	11	39.3%
4	32	8	44	44.6%

A window length of $N = 64$ in our case is 0.5 seconds long. The total accuracy is the calculated observed probability of having the classifier correctly predicting the category. This observed probability is:

$$\frac{\text{Number of trials tested}}{\text{Number of correct predictions}} \tag{4.1}$$

The confusion matrix from trial 2 (using N=64 and K=4) was:

		True class		
		Face	**Landmark**	**Object**
Predicted class	**Face**	26 (52%)	13 (26%)	17 (34%)
	Landmark	16 (32%)	20 (40%)	12 (24%)
	Object	8 (16%)	17 (34%)	21 (42%)
	Total	50 (100%)	50 (100%)	50 (100%)

FIGURE 4.5: Confusion matrix for small grid, K=4.

The accuracy plots are seen in Fig. 4.6.

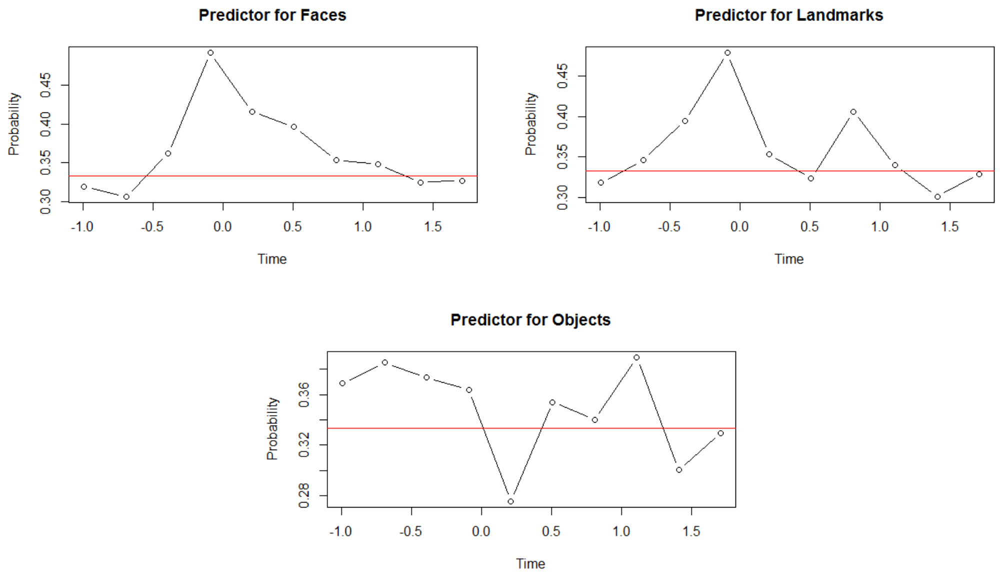

FIGURE 4.6: Classifiers when using multitaper method (N=64, K=4, small grid)

4.1.4 Perception vs. imagery

A similar cross-validation was made using the small grid on the "imagery"-trials using $N = 64$ and $K = 8$. Here the target stimuli was not presented but instead it was asked for the participants to keep the image in mind.

The accuracy plots from this trial can be seen in Fig. 4.7.

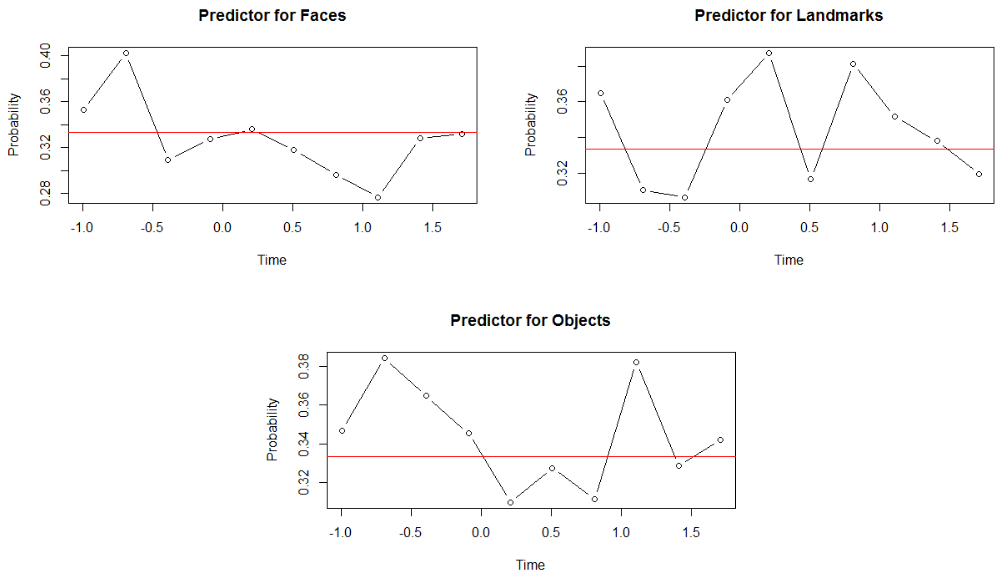

FIGURE 4.7: Classifiers when using multitaper method (N=64, K=4, small grid) with imagery features.

The total accuracy for these set of classifiers was 38%.

Chapter 5

Discussion

5.1 Discussion

5.1.1 Analysis

According to the results previously given, it can be argued that the Thomson multitaper method is successful in producing a classifier for semantic memories. However, there are some issues with modeling in this way. It is natural to want to include as many features as possible, because we want to be able to explain as much as possible. But in signal processing, we're generally going to have a lot of correlation. This is partly fixed by using the multitaper method, but since we are trying to find differences within the same brain, the channels will most likely correlate with each other, since the electrodes have very poor spatial resolution.

In the feature selection, an average of $\approx 70\%$ was kept to be included in the classifier. This is one of the reasons to why we chose multitaper over the modified periodogram. It is well known that the multitaper method reduces the variance compared to the modified periodogram, and this is probably why we got such a low value of features collected if we use the modified periodogram. When we compute the T-tests for the different regions of the grids, this takes the variance into account, and does not let as many features through the selection due to the high variance.

However, using the multitaper method with N=64 and K=8 (or 4) with a small grid lets us use a fair amount of the data we have, while at the same time dealing with the variance. The 55% prediction accuracy for faces, and 51% for landmarks suggests that this method is useful. The results from the accuracy plots in Fig. 4.1 and Fig. 4.6 is comparable with the result of Morton et al, 2012 (see Fig. 1.1). For faces and landmarks we can see a big increase in accuracy when the stimuli is shown.

When a bigger grid was used, we only had a significant difference in the very beginning of the stimulus onset in the faces category. It could be that we get a lot of redundancy, since we already average over time when we compute the spectrum, and so the time resolution is already limited by our choice of bandwidth. The results in this case are not significant due to the high variance of the accuracy. Fig. 4.2 illustrates this variance. Since so many features were included in the model, the probability of the different features being correlated is high. This makes the model very unstable, and that's probably why the variance is so high.

It's noticeable that the classifier for the big grid is not very predictive, and hereby we conclude that it's better to use the smaller grid. The "Face" predictor was not changed that much when we used the big grid in Fig. 4.4, but for the "Landmark" and "Object" category the predictor using the big grid was even worse than chance on average.

When we reduced the number of windows (K) to 4, we expected to see a better resolution in frequency, and then maybe we could pick up some larger differences between the categories. If we look at the confusion matrix for trial 2 (Fig. 4.5) it indicates that there might be a significance even in the case for the "Object" category, but when we look at the accuracy plots (Fig. 4.6) we see that the "Object" category is responding to something else than the shown stimuli, otherwise it would be more accurate when the stimuli is shown. From this we draw the conclusion that this method is unable to classify semantic memories from the group "Object". For the other two categories, we do see the same increase in accuracy as for $K = 8$, and see therefore no problem in continuing testing for both values of K.

The classifiers trained in this report was not able to accurately predict a signal corresponding to the category "Object". It could be that the neural patterns for a generic object is beyond this study in terms of precision. But it could also be that neural patterns for objects (like a pair of scissors or a pencil) are too different to be considered the same class. Maybe objects like these have a deeper connection with our memory (since we encounter them every day), and are therefore harder to trace as a neural pattern.

Finally, when we look at the accuracy plots of the imagery paradigm, we can see that it's probable that what we see in Fig. 4.7 is noise to a big extent. It does not seem to react at all to the fact that a stimulus is presented, and since the total accuracy was as low as 38% it does not seem like imagery is a stable enough hypothesis for building a classifier. We can also note that it seems like the peaks of the accuracy plots are in the baseline, which may indicate that the target is still lingering in the mind of the subject. For objects and faces this is the case.

All our numbers from the result section are based on the cross-validation that was made. This means we are yet to test these classifiers on actual recall. Accuracy of our classifiers only speak of accuracy in terms of predicting a signal of encoding, and not of recall. It can also be argued that another type of cross-validation might be better for such an analysis, based on the number of trials. This would've yielded another result.

5.1.2 Possible errors

Since we are dealing with brainwaves, it's important to realise that conclusions we draw are very unique for the specific brain that we are examining. Therefore it's important to not extrapolate out of the brain we are examining. The results we get are also based on the assumption that there are differences in these categories of memory, and that the grouping of stimuli actually resembles the categories that we are classifying for.

It could also be that the cross-validation technique of "Repeated random sub-sampling validation" is not suitable for this size of dataset. Since we had a total of 59 trials and 5 randomized trials are left out every iteration, then there is no way to make sure that all the trials are left out an equal amount of time. If there would be problems with the data, such as outliers or artifacts, then this would definitely cause some trouble in the classification. Also due to time constraints, only 10 repetitions of cross-validation was run for each trial. Ideally, this number should be a lot higher, if we want to make sure that our cross-validation is working.

5.1.3 Future studies

So what can be done to improve the classifiers?

A problem I see is that we would want to see what happens closer to the onset of the stimuli at T=0. Focusing on the signal around T=0 will probably not only increase the total accuracy, but it will also help us getting more sensitivity to when in time this increase happens for different categories. The whole point of using EEG in favor of other methods like fMRI or PET is that we have much better temporal resolution, and this can be used in a much bigger extent then what has been done in this specific study.

We could also focus more on the problem of redundancy. It's possible by help of step algorithms to remove unnecessary features from a model, which means that we would have more stable model in terms of computing the probabilities, but maybe not in terms of predictive power of the classifier. Also, instead of making a logistic regression and comparing the probabilities, for a classification problem like this, a multinomial logistic regression model would be more suitable.

The channels chosen in this task can also be changed to improve the classification. The choice of individual channels was based on a grid that would cover the entire scalp but at the same time remove redundant information. The reduction of channels also increase the possible errors (if the channels are not accurately measuring the neural activity for some reason). Maybe we should consider to average over a few channels close to each other when we form the grid next time. This will increase the redundancy a bit, but maybe reduce possible measuring errors.

5.1.4 Conclusion

The purpose of this report has been to find complementary methods to classification of semantic memories by using EEG. The Thomson multitaper method of spectral estimation has proved to be successful in detecting differences in neural patterns both over time and between the categories, based on the results from the cross-validation. The method has proved to be the most useful with a lower amount of features collected, which is intuitive due to the restrictions on resolution. There are many parameters that can be adjusted to make the classifiers better, and I personally think it's improvable with some more extensive research.

Bibliography

[1] Eichenbaum, H., (2000). *A cortical–hippocampal system for declarative memory.* Nature Reviews Neuroscience 1, 41-50; doi:10.1038/35036213

[2] Hansson-Sandsten, M., (2011, 22-27 May). *Cross-spectrum and coherence function estimation using time-delayed Thomson multitapers.* Paper presented at Acoustics, Speech and Signal Processing (ICASSP), 2011 IEEE International Conference, Prague. doi:10.1109/ICASSP.2011.5947289

[3] Ketz, N., O'Reilly, R. C. & Curran, T. (2014). *Classification aided analysis of oscillatory signatures in controlled retrieval.* NeuroImage, 85, 2: 749-760. doi:doi: 10.1016/j.neuroimage.2013.06.077

[4] Lindgren, G., Rootzén, H. & Sandsten, M., (2014). *Stationary stochastic processes for scientists and engineers.* Boca Raton: CRC Press.

[5] Morton, N. W., Kahana, M. J., Rosenberg, E. A., Baltuch, G. H., Litt, B., Sharan, A. D., Sperling, M. R. & Polyn, S. M., (2012). *Category-Specific Neural Oscillations Predict Recall Organization During Memory Search.* Cerebral Cortex October 2013, 23(1): 2407-2422. doi: 10.1093/cercor/bhs229

[6] Priemer, R., (1991). *Introductory Signal Processing.* World Scientific. p. 1. ISBN 9971509199. Retrieved 20 February, 2015 from `https://books.google.se/books?id=QBT7nP7zTLgC&pg=PA1&redir_esc=y#v=onepage&q&f=false`

[7] Stoica, P. & Moses, R., (1997). *Introduction to spectral analysis.* New Jersey: Prentice Hall

[8] Warner, R. M., (1998). *Spectral analysis of Time-Series Data.* New York: The Guildford Press.

Bachelor's Theses in Mathematical Sciences 2015:K7
ISSN 1654-6229

LUNFMS-4013-2015

Mathematical Statistics
Centre for Mathematical Sciences
Lund University
Box 118, SE-221 00 Lund, Sweden

http://www.maths.lth.se/