Danmarks
Tekniske
Universitet

DTU

Time-frequency analysis of EEG

Author

Giovanni Grego - s202287

June 31, 2021

Contents

Abstract

The aim of this special course was to study the basis of the EEG analysis, through a hands-on approach supported by readings of papers and tutorials. For this reason the report focuses more on the framework, on the pipeline and the process rather than the final results.

The report starts with an introduction over the theory of the EEG signals and the McGurk illusion. Secondly it presents the experiment and the preprocessing. The third chapter focuses on the methods and the framework used for analysing the data, especially from a theoretical point of view. Finally, the results are shown and commented. At the very end, some personal comments on the course have been added.

The report doesn't have the goal to cover systematically and extensively all topics treated but to have a general and easy to read reference for highlight the most interesting parts. Some resources are cited for each part for future references. All the codes are stored in a github repository at this link: https://github.com/GiovanniGr/Time-frequency_analysis.git

Chapter 1

Introduction

1.1 EEG

The EEG measures the electrical activity of the cerebral cortex. The electrical activity of the brain is mainly caused by the post-synaptic potentials (PSPs), in particular for a change in membrane conductance and transmembrane potential provoked by the neurotransmitter released by the travel of the action potential to the nerve terminal.

The signal can lead to an excitatory post-synaptic potential (EPSP) on the dentrites or to an inhibitory post-synaptic potentials (IPSPs) on the cell body of the neuron. The combination of EPSPs and IPSPs induces currents that flow within and around the neuron. The EEG is essentially measuring these voltage changes in the extracellular matrix. The measurement is taken on the scalp by a system of electrodes and amplifiers.

The interaction between cortex and thalamus as well as the functional properties of large neuronal networks in the cortex that have an intrinsic capacity for rhythmicity create recognizable EEG patterns, varying in different areas of neocortex that allow us to make sense of the complex world of brain waves.

A longer and more accurate explanation, from which this summary has been extracted, can be found in the book [4].

1.2 Evoked and induced signals

This section has been written on the basis of the paper [1].

Evoked and induced oscillations are two types of cortical oscillatory activity. These oscillations differ in their phase-relationships to the stimulus. Evoked oscillations are phase locked to the stimulus, whereas induced oscillations are not.

Evoked oscillations are extracted from the EEG signal by first averaging over trials (i.e. getting rid of the not phase-locked components where the sum averages zero). Then the time-frequency analysis is applied to give an event-related response.

Induced oscillations are estimated by applying a time-frequency decomposition to each trial and hence averaging among trials' power. The power of evoked and background components are subtracted from this total power to reveal induced power.

In other words, evoked responses can be seen as the power of the average of the trials, whereas induces responses are the average power that cannot be explained by the power of the average.

A common conception is that evoked oscillations reflect a stimulus-locked event-related response in time-frequency space. On the other hand, induced oscillations are generated by some distinct high-order process.

1.3 McGurk illusion

McGurk effect is a multisensory illusion occurring with audiovisual speech.

It consists in a conflict of information coming from different senses, namely sight and hearing. In practice the illusion can be observed when one is asked to watch a video of lip movements alongside listening to sounds uttered, apparently by the same person whose lip movements one is watching.

If the lip movements and the sounds do not match, for example if the lip movements indicate a /ba/ sound whereas the auditory information is that of /ga/, one typically experiences an illusory third sound, for example /da/.

The reason for the great impact is that this is a striking demonstration of multisensory integration. It shows that auditory and visual information is merged into a unified, integrated percept. It is a very useful research tool since the strength of the McGurk effect can be taken to reflect the strength of audiovisual integration. [15]

Chapter 2

Experiment

The details of the design and conduction of the experiment, as well as the preprocessing of the raw data can be read in [5]. Below, a brief description of the experiment is reported.

2.1 Experiment:

The scope of the experiment is to analyse the McGurk fusion and the congruent conditions. Therefore visual and auditory stimulus have to be prepared and to be presented to the individuals. It was decided to record video clips of a speaker uttering the syllables /ba/ (then B) and /ga/ (then G) at 25 fps for the video and 44100 Hz for the audio. Then, frequency-matched noise for both audio signal and fast Fourier transform (FFT) signal of the video has been added to the original recordings with three different levels of signal-to-noise ratios (SNRs) (high, mid or low).

The different signals available are therefore visual G and B and auditory G and B, all of them with the three levels of SNRs. Visual and auditory stimulus are hence combined together in order to generate the McGurk fusion (auditory B, visual G) and congruent (auditory B, visual B and auditory G, visual G) conditions, in addition to only visual (G or B) and only auditory (G or B) conditions.

Twenty-four individuals (15 female) participated in the experiment. Each trial started with the first frame of the video (a still image of a face) lasting between 800ms and 1200ms. Then, for audio-visual conditions, the video playback started along with the onset of auditory noise. The syllable started 580 ms into the videos, which lasted for 1320 ms in total. For auditory-only presentation, the still image lasted for the whole presentation of the stimulus and visual-only stimuli were presented without sound. A total of 1980 trials were delivered, distributed in a balanced way among the different conditions. At the end of each trial, the participant was asked to press a button of a Cedrus response box labelled "B", "D" or "G", with respect to the syllable perceived. In Figure 2.1 the timeline of a general trial is shown.

Electroencephalograms (EEG) were recorded from a 128-channel cap. Eye movements

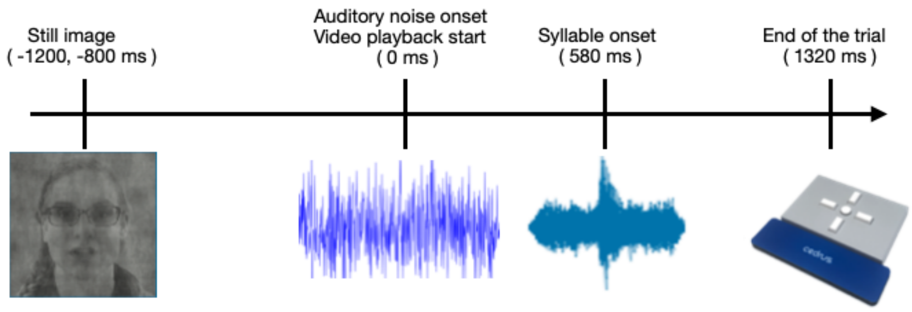

Figure 2.1: Timeline of each trial/epoch

were recorded with two electrooculography electrodes. The signal was recorded at 1000 Hz in reference to an electrode positioned at the tip of the nose, using an online high-pass filter of 0.1 Hz and a low-pass filter of 250 Hz. Three subjects were removed from analyses. Two had severe artefacts in the EEG and one was not able to distinguish between the visual stimuli. Thus, 21 subjects were used for the further analyses.

2.2 Preprocessing:

The preprocessing for the data used was already done in the paper [6]. The EEG signal was first high-passed at 1 Hz (Hamming windowed sinc FIR filter, order 3301) and then notch filtered at 49-51 Hz (Hamming windowed sinc FIR filter, order 3301) and down-sampled to 250 Hz. Subsequently, it was time-locked to syllable onset and cut into epochs starting 1100 ms before syllable onset and ending 1100 ms after syllable onset.

Then, data was inspected by eye and excessively noisy channels and highly noisy epochs were removed. This lead to removal of 7.1 (3 - 11) of 128 channels and 5.46% (1.92% - 11.11%) of epochs. The signal was re-referenced to the average reference (excluding the EOG channels) and an Independent Component Analysis (ICA) decomposition was computed using the 'binica' algorithm.

Independent components (ICs) were inspected and components representing eye movements, cardiac artefacts, cap movements and high-amplitude muscular activity were projected out from the data. On average 8.11% (5.60- 13.11%) of ICs were projected out.

Finally, the removed channels were linearly interpolated and epochs with an absolute amplitude exceeding 100 μV within the interval [-800, 800] ms compared to syllable onset were removed. This resulted in the removal of an additional 1.22% (0-5.37%) of trials.

Chapter 3

Methods

In this project, both Matlab and Python have been used. The original data was stored in matlab fieldtrip structures [3] and has to be converted in Python-readable types. This in order to be able to use and apply the functions of the MNE python library [8]. MNE is the most famous library used in Python to analyse and work with EEG signals. It contains tons of built-in functions that allow the user to perform easily a time-frequency analysis as well as a cluster based test, the two main tools used in the project.

3.1 MNE overview

There are many tutorials in the MNE website in order to start using the library. The following overview is extracted by the tutorial [10].

The basic blocks on which the library works are the raw data, the epochs data and the evoked data.

The raw data consists of the continuous EEG signal, in which each data point is the amplitude value of the signal (in volt) for a specific channel and time. The raw data, in the project, was useful to have a coarse insight on the signals and to initially build the epochs. In the second version of the code, the raw data has just not taken into account, preferring to create the epochs directly from the Matlab trials.

The epochs data is a discontinuous version of the raw data. In particular, the epochs are segments of the raw data, related and synchronized to an event registered during the data acquisition. Epochs are heavily used in the whole project as the aim is to compare signals arising from specific events.

Finally, another important data structure in MNE is the evoked data. The evoked, in general, represents the average of signals. In particular, in the project, evoked are used to average all the epochs (trials in Matlab) from the same event or the same person, in order to have a better estimate and average out unintentional factors.

3.2 Preprocessing

The data I worked with was already preprocessed. For this reason there has not been a real section of preprocessing. On the other hand the data used was stored in Matlab Fieldtrip structures, hence the work has been to transfer the information into Python MNE structures.

The idea is to convert each elementary piece into Numpy arrays and then building up the MNE structures from those arrays. The final idea is the result of many iterations of trial and error. This gave me the opportunity to face and try different ideas, methodologies and functions. The path towards the solution has been mainly guided by the functions of the library, following the requested input parameters and building them up during the conversions. One thing that arose in the process is the lack of compatibility between the two libraries, Fieldtrip and MNE. It would be desirable, in the future, a better compatibility in order to exploit the advantages of both the solutions.

It is valuable to highlight the problem with the time synchronization among trials and persons. In particular, for some of the files, the starting point of the trials were not synchronized and this led to a time-shift of some trials and to a wrong analysis. All the details can be found in appendix A.

Finally, in Figure 3.1 the ROI used in the entire project is shown. The region of interest covers the central-right region of the cerebral cortex.

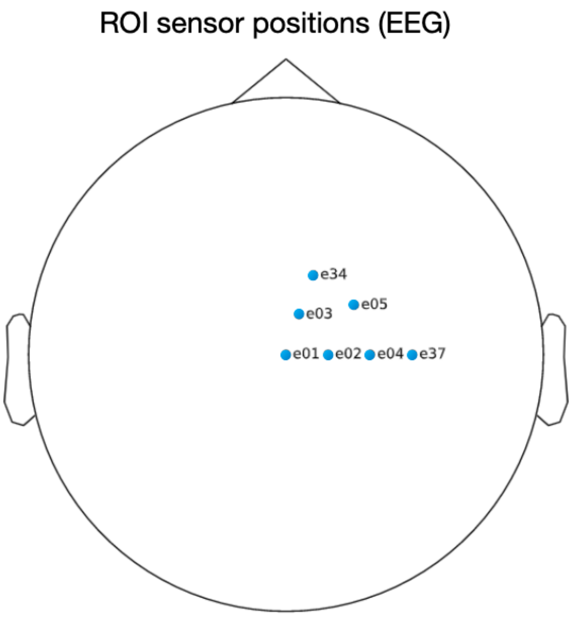

Figure 3.1: ROI used in the project

A premise to be made before starting with the data analysis is that, as shown in Appendix B, already in the computation of powers there is something wrong. It is deemed that the problems are in the parameters of the functions used than not on the framework. For this reason the results below are not correct but it has been decided to comment and analyse the different steps of the pipeline as if the results were correct, focusing more on the theory behind the methods and on the general framework.

3.3 Event related potential

First of all, in order to have an insight on the data, the event related potential is studied. In particular, the signal we are most interested in is the auditory response in the channels at the centre of the scalp.

Firstly the epochs have been created for each user, binding each epoch with the corresponding event. Subsequently the average of the epochs (evoked object) for each event has been computed, applying a baseline in the 100ms before the syllable onset (i.e. [480ms 580ms]. Finally, for visualisation purposes, the mean over the users for each event has been computed in order to have a better estimation of the signal.

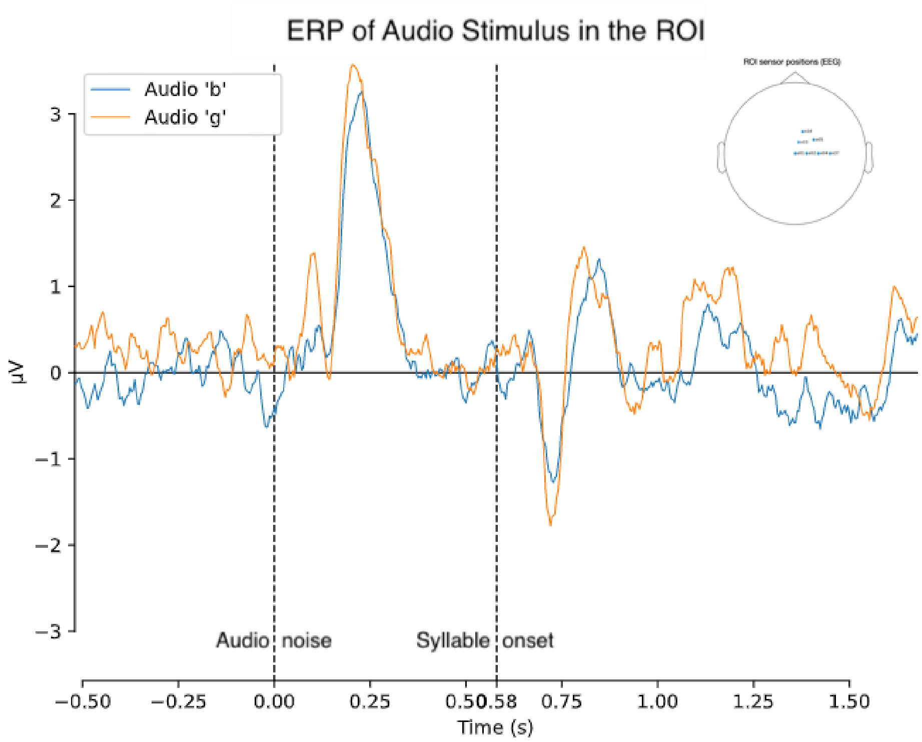

Figure 3.2: Event related potential of auditory response for syllable /ga/ and /ba/

In Figure 3.2 both the event related potentials of the syllable /ga/ and /ba/ are shown. The vertical line at 0s corresponds to the starting of the video and the beginning of the audio noise. The vertical line at $0.58s$ corresponds to the syllable onset. As it can be seen from the figure, there is a peak soon after the starting of the noise and then one after the syllable ongoing. Both of them are interesting and predicted, as we expect a perturbation in response to the stimulus. In particular, the most interesting part is soon after the second dashed line, that corresponds to the syllable onset for both events. It can be seen that there is a negative peak after around 150ms from the syllable onset followed by a positive peak. This is a behaviour found also in the thesis [5].

3.4 Time-frequency analysis

The following step in the analysis consists of computing and inspecting the powers related to each event condition [7]. To compute the power, the Morlet wavelet transformation [16] has been applied to all the epochs for each condition. Then the powers have been averaged for condition. Finally, just for visualisation purposes, the powers per condition have been averaged also among the users. In the figures below, some examples are shown.

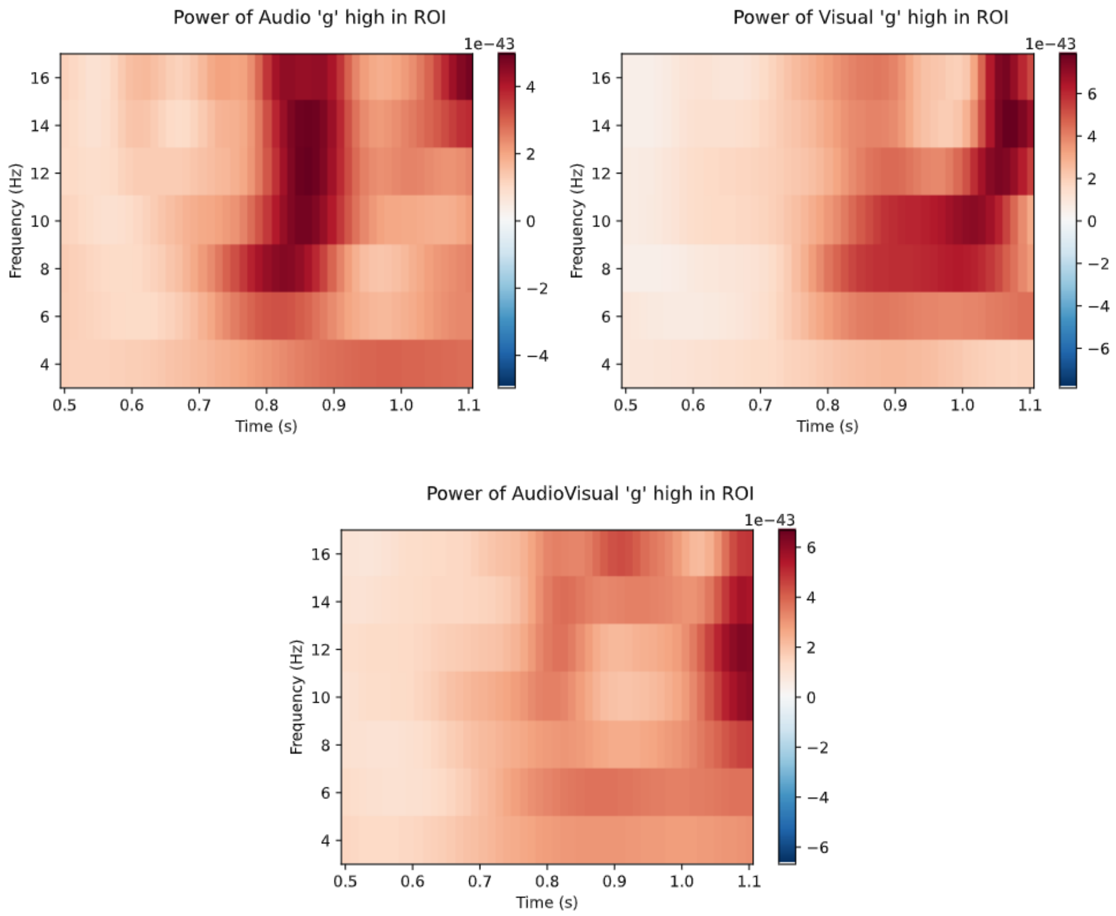

Figure 3.3: Examples of time-frequency powers (dB) per condition

In Figure 3.3 some examples of time-frequency plots of the induced signals are shown. In the left top corner, the audio /ga/ condition is shown. It can be seen that there is a positive peak at around 850ms. The power is almost constant for the different frequencies. In the visual /ga/ high (top right) there is a positive peak after 1s. For audiovisual the behaviour is similar to the visual, in which the peak is around the end of the experiment. More or less all of them have similar trends across frequencies.

3.5 Permutation cluster test

The last step aims to compute and to visualize whether there are statistical differences among powers of different conditions. In this section, the theory behind the permutation cluster test has been sketched, having as references different tutorials and papers [11], [13], [2], [12]. The implementation part is left in the code and involves only one function of the mne library [9].

It is of particular interest to analyse the difference among certain events, for example between "audiovisual" with respect to "audio" and "visual" or between the same event but with different levels of SNR.

The test used is the cluster-level statistical permutation test based on the F-statistic. The problem that led to the conception of this method is the multiple comparisons problem: the more a test is repeated, the higher is the probability that by chance an effect is detected.

In this case the problem stems from the large number of points compared. The solution adopted by the cluster-level statistical permutation test is, as stated in the name, to perform the test not on each data point but on an overall value that explains the cluster.

In particular, the procedure can be summarised in two steps: the cluster-forming, which converts one high-dimensional observation into a quantifiable summary regarding its cluster structure, and the creation of a surrogate null distribution, against which the observed data is compared to obtain p values.

The case under consideration has data of shape (time x frequency x space) and the inference regards a binary condition contrast across subjects.

In the first step, in the two-condition case, a first-order test statistic (e.g. t test) is calculated comparing the values of the two conditions for each coordinate. The H0 of this test is that, in an unobserved population of subjects, exposed to the same experimental manipulation, the difference between the two conditions would equal precisely 0.

Repeating this procedure for the entire data set results in a t score map. Comparing these t values to the t distribution yields p values.

But due to the number of scores, many are expected to exceed a critical value even if the true effect is zero everywhere. The idea is to exploit clusters of data in order to reduce the dimensionality. For this, voxels are thresholded according to an a priori defined criterion, and adjacent voxels with t scores exceeding this value are grouped together.

Finally, groups are summarized into a single number by, for example, taking the sum of the t values-yielding the cluster size(s).

In the second step, a second-order inference stage is employed. The null hypothesis of this stage is that the cluster structure of the data identified in the first stage is exchangeable between conditions (i.e., clusters simply reflect the inherent correlation of data in time, space, and frequency).

The probability of the data under a null hypothesis of exchangeability can be established with permutation tests. The test simply realizes this null hypothesis by enumerating all possible assignments of data points to the conditions.

Although the full permutation problem is computationally intractable, there exists some approximation algorithms that yields satisfactory results. In each iteration, the data is randomly assigned to conditions, the structure of the cluster is established and the value is stored. This is repeated until a large number of samples under the null hypothesis of exchangeability has been obtained. Finally, the empirical cumulative density function of these surrogate-null values is computed as an approximation of the distribution of the test statistic under the null.

The percentage of surrogate-null values that the actual observed data exceed corresponds to the p value under the null of exchangeability. Importantly, while the original data is high dimensional and thus prone to multiple comparison issues, the first stage reduces it to a single number, and it is this number whose probability under the null is established.

Chapter 4

Results and Conclusions

The combinations of conditions for which the cluster test has been applied can be grouped in two areas. One test has been computed for each combination of two different level of SNR for the same stimulus (e.g. "visual /ga/ high" and "visual /ga/ medium"). This group involves two signals and will be called from now on "Group SNRs comparisons". Another set of comparisons has been applied between a multi-sensory event (e.g. "audiovisual /ga/ low") and the sum of the respective single-sensory events (e.g. "audio /ga/" plus "visual /ga/ low"). This group involves three signals and will be called "Group multi-stimulus". Finally, other comparisons have been made among the same stimulus but for different syllables (e.g. "visual /ga/ low" and "visual /ba/ low"). This has been made just for testing reasons, and this group will be named "Group test same stimulus". It is easier to group the different comparisons as it will be seen that the results within the same group are similar.

4.1 Analysis of the threshold

An important part in the permutation cluster test is played by the threshold parameter. In general the lower the threshold, the broader the inspection and the fewer the clusters [14], [13]. All the details of the analysis can be found in Appendix C. The final result (i.e. significant difference or not) doesn't change with respect to the thresholds used, so for convenience it was decided to take $t = 0.5$. The results reported in the sections below are computed with that threshold.

4.2 Group test same stimulus

In this group, the test has been performed between "visual /ga/ low" and "visual /ba/ low" (i.e. visual stimulus with SNR low). This comparisons have been made to test the method as we don't expect too much difference between these two conditions, as the stimulus is very disturbed. In the figure below it can be seen the result for the visual events.

	0	1	2	3
p-value of Cluster	0.5087	1	1	1
nr. points in Cluster	84991	1	2	1

Table 4.1: P-value and number of points for each cluster in the "visual /ga/ vs /ba/ low" conditions.

As expected, in Table 4.1 it can be seen that for each cluster, its p-value is greater both than 0.01 and 0.05 (i.e. two standard thresholds for p-value). This means that the two signals are not statistically different as there is neither one cluster with a p-value lower than one of the two thresholds. This result was expected as the two events are similar because the two images have a low SNR, so it's difficult to see them and the stimulus doesn't differ significantly.

From the number of points for each cluster it can be seen that the great majority of the points are put in the same cluster, the most significant. This aspect has been discussed in more detail in Appendix C.

Another interesting aspect to analyze is the cluster-formation.

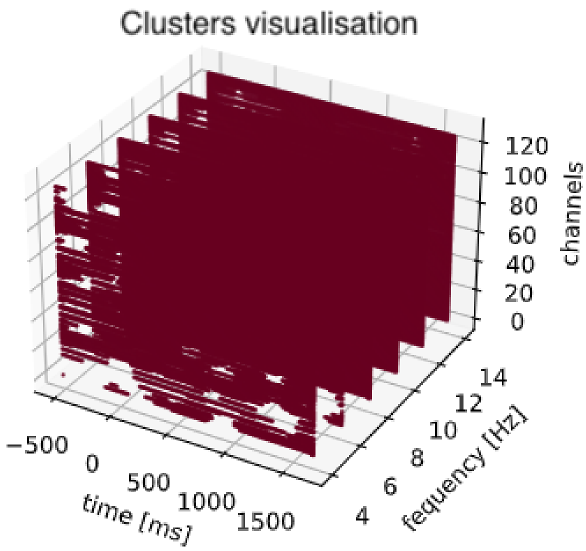

Figure 4.1: Visualization of the clusters for the "visual low" test

In Figure 4.1 each cluster of the data points is coloured differently. From the image it can be seen that some points are removed from the variable space because the F-value was lower of a certain threshold in the first step of the algorithm. The data points left are then grouped into clusters, based on adjacencies in time, frequency and space (channels).

Adjacency in time and frequency is trivial (each point is adjacent to the previous and to the following) whereas for the channels, the montage is needed and it was retrieved by the original fieldtrip structure.

In Figure 4.1 is not easy to see the cluster formation, as there is only one big cluster and other few clusters with one or two elements. For this reason, the visualisation of the clusters for the same condition but with an higher threshold is shown in the figure below.

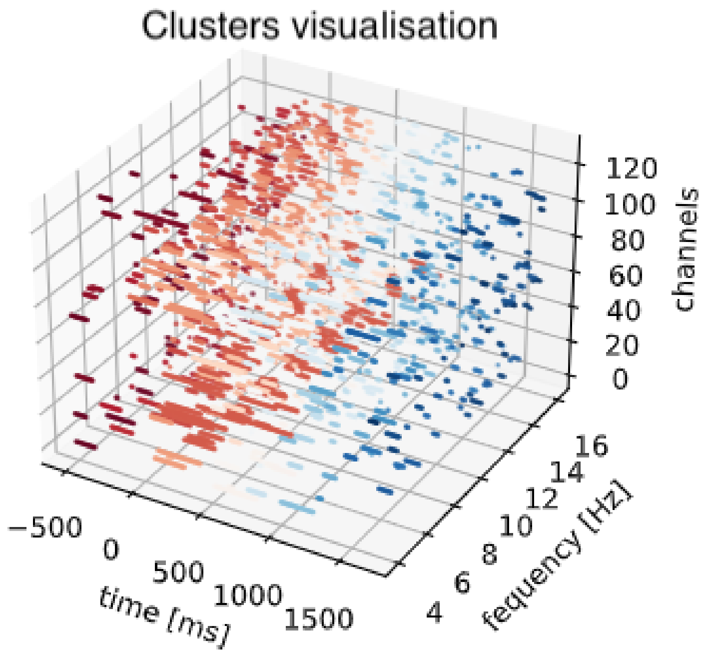

Figure 4.2: Visualization of the clusters for the "visual low" test, with t=5

In Figure 4.2 it's easy to see that there are many less data points than not in the figure above. This is because the threshold used is higher and more points are filtered out in the first step of the permutation cluster test. Moreover, also in this case each cluster is coloured differently. It can be seen that there are many clusters and that are oriented in the frequency and in the space directions whereas are split in the time direction.

Also in this case it can be seen that the higher the threshold the fewer the points and the more particular the analysis.

4.3 Group multi-stimulus

In this set of tests, the comparisons have been performed between a multi-sensory event and the sum of the respective single-sensory events. This has been done for congruent (e.g. same syllable for both "audiovisual", "audio" and "visual") signals. Moreover, each comparison has been made among the same level of SNR (e.g. "audiovisual medium" with "audio medium" and "visual medium") and for each SNR level. In total there are six tests: three levels for "audiovisual /ga/" and three for "audiovisual /ba/". Apart from "audiovisual /ga/ high" and "audiovisual /ga/ low", all the other tests have only one cluster of data and the p-value is below the 0.05 threshold. This means that there is statistical difference between the "audiovisual" power signal and the sum of the "audio" and "visual" powers.

For both "audiovisual /ga/ high" and "audiovisual /ga/ medium" more than one cluster have been found, of which one with p-value lower than 0.05 and the other higher. Hence also for these events there is significant difference. The results are summarized in the table below.

				Nr. clusters	Nr. clusters with p <0.05
Group m-s.	audiovisual vs audio+visual	/ba/	high	1	1
			medium	1	1
			low	1	1
		/ga/	high	4	1
			medium	1	1
			low	2	1

Table 4.2: Number of clusters and number of clusters with p<0.05 for each condition tested in Group multi-stimulus

4.4 Group SNRs comparisons

Lastly, in this set the same event has been compared for different levels of SNR. Also in this case, there are so three combinations of levels for "audiovisual /ga/" and three for "audiovisual /ba/", for a total of six tests.

				Nr. clusters	Nr. clusters with p <0.05
Group SNRs.	audiovisual	/ba/	high vs med	8	0
			med vs low	9	0
			high vs low	8	0
		/ga/	high vs med	15	0
			med vs low	4	1
			high vs low	3	1

Table 4.3: Number of clusters and number of clusters with p<0.05 for each condition tested in Group SNRs comparisons

In Table 4.3 an overview of the number of clusters and how many of them have a p value less than 0.05 is shown. It can be seen that all of them have different number of clusters and that for only two of them at least one cluster has $p < 0.05$.

Hence for all the different combinations of levels of SNR, the difference for the two signals is not statistically significant for the first four conditions and there is statistical difference for "audiovisual /ga/ medium" compared to "audiovisual /ga/ low" and "audiovisual /ga/ high" compared to "audiovisual /ga/ low".

4.5 Conclusions

From the results above it can be said that in general there is significantly difference for the tests where the power of the audiovisual stimulus is compared with the power of the sum of the audio and the visual stimuli.

On the other hand, there is not statistical difference among the tests of the second group, in which different SNR levels were tested together for the same syllable, apart for two of them where it was found differences.

As a general conclusion it's important to remember that the results depend also on the threshold used both for the first step of the cluster formation and for the threshold on the p-value. For these reasons, some differences can be found especially for Group SNRs comparisons. Nevertheless, as shown in Appendix C, the results are robust against the threshold parameter.

Finally, it has to be remembered that, as discussed in Appendix B, the results of the analysis are not correct and the important part is more what it should be computed and why than not the actual results.

Resources

[1] Olivier David. "Mechanisms of evoked and induced responses in MEG/EEG". In: (2006). DOI: https://doi.org/10.1016/j.neuroimage.2006.02.034.

[2] fieldtrip. *Cluster-based permutation tests on time-frequency data.* URL: https://www.fieldtriptoolbox.org/tutorial/cluster_permutation_freq/.

[3] fieldtrip. *Welcome to the FieldTrip website.* URL: https://www.fieldtriptoolbox.org.

[4] Jiyeoun Jenna Yoo MD Lara V. Marcuse MD Madeline C. Fields MD. *Rowan's Primer of EEG.* Elsevier, 2015. ISBN: 0323353878.

[5] Alma Lindborg. *Quantifying Audiovisual Integration in Speech Perception.* Technical University of Denmark, 2020. ISBN: 0909-3192.

[6] Alma Lindborg. "Speech-specific audiovisual integration modulates induced theta-band oscillations". In: (2019). DOI: https://doi.org/10.1371/journal.pone.0219744.

[7] mne. *Frequency and time-frequency sensor analysis.* URL: https://mne.tools/stable/auto_tutorials/time-freq/plot_sensors_time_frequency.html#sphx-glr-auto-tutorials-time-freq-plot-sensors-time-frequency-py.

[8] mne. *MNE.* URL: https://mne.tools/stable/index.html.

[9] mne. *mne.stats.permutation_ cluster_ test.* URL: https://mne.tools/stable/generated/mne.stats.permutation_cluster_test.html.

[10] mne. *Overview of MEG/EEG analysis with MNE-Python.* URL: https://mne.tools/stable/auto_tutorials/intro/plot_10_overview.html#sphx-glr-auto-tutorials-intro-plot-10-overview-py.

[11] mne. *Spatiotemporal permutation F-test on full sensor data.* URL: https://mne.tools/stable/auto_tutorials/stats-sensor-space/plot_stats_spatio_temporal_cluster_sensors.html#sphx-glr-auto-tutorials-stats-sensor-space-plot-stats-spatio-temporal-cluster-sensors-py.

[12] mne. *Statistical inference.* URL: https://mne.tools/dev/auto_tutorials/stats-sensor-space/10_background_stats.html.

[13] Jona Sassenhagen. "Clusterâbased permutation tests of MEG/EEG data do not establish significance of effect latency or location". In: *Psychophysiology* (2018). DOI: 10.1111/psyp.13335.

[14] Stephen M. Smith. "Threshold-free cluster enhancement: Addressing problems of smoothing, threshold dependence and localisation in cluster inference". In: (2009). DOI: https://doi.org/10.1016/j.neuroimage.2008.03.061.

[15] Kaisa Tiippana. *What is the McGurk effect?* URL: https://www.frontiersin.org/articles/10.3389/fpsyg.2014.00725/full.

[16] wikipedia. *Morlet wavelet.* URL: https://en.wikipedia.org/wiki/Morlet_wavelet.

Appendix A

Starting time distribution analysis

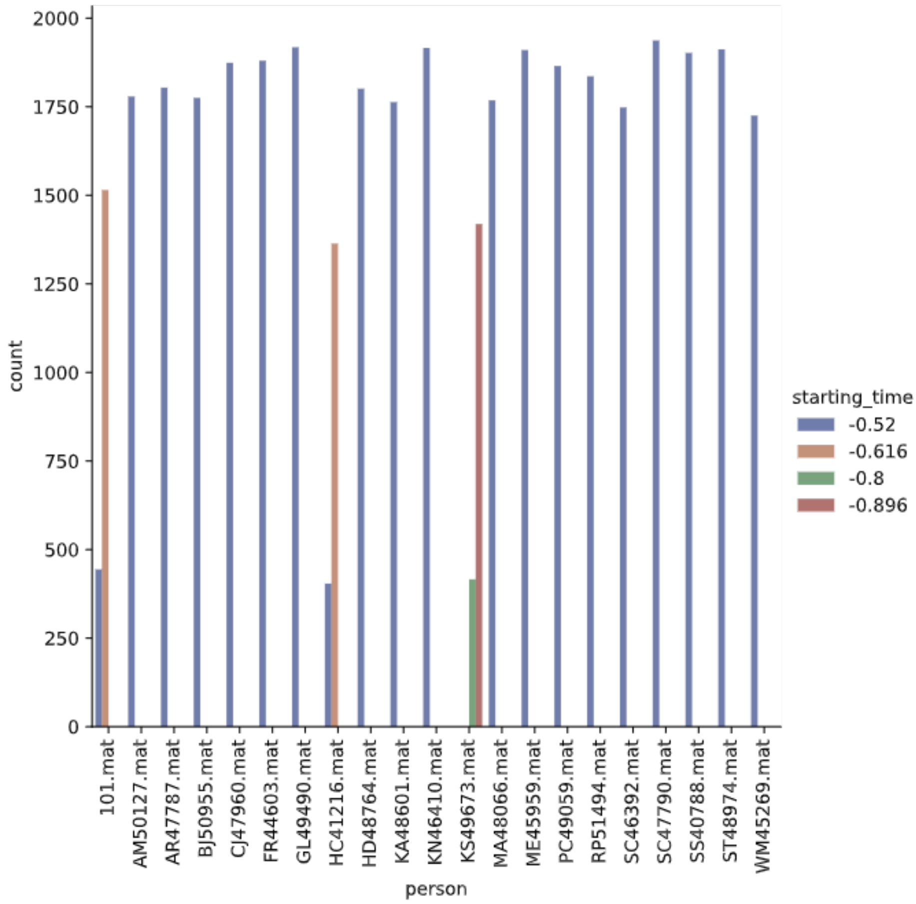

Figure A.1: nr. of epochs per person and time

In Figure A.1 the distribution per person and time is shown. It can be seen that only for three people there are different starting times than not -0.52. It turned out that the time information in the Matlab data was wrong and the solution was simply to not take care of the time in the data, but assuming that all the trials start at -0.52s.

Appendix B

Pipeline validation on Evoked

The results of the analysis for the evoked version of the signal are shown in the thesis [5]. For this reason it is possible to check the pipeline used for the induced version. Below, the results for the evoked signals are shown.

B.1 Power of the evoked signals

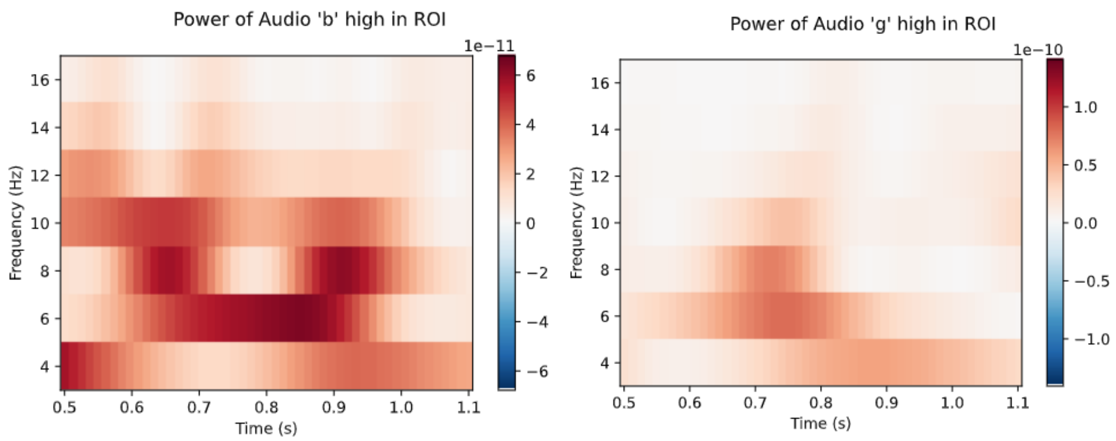

Figure B.1: Time-frequency powers (dB) of the stimulus audio /ba/ (left) and audio /ga/ (right)

As it can be seen in Figure B.1 the same event (i.e. audio) but for two different syllables is shown. These two signals should be similar but in the images above it can be seen that the powers are considerably different. In fact, in the image on the left, the signal has a behaviour different from the one on the left. First of all the maximum power (in dB) differs for a factor two between the two signals. Secondly also the peaks are different, with the audio /ga/ with a peak around 750ms (150ms after the syllable onset) between 6 and 8 Hz whereas for audio /ba/ there is not a clear peak and the higher levels of power is spread over more frequencies and for a longer time.

B.2 Permutation cluster test on evoked signals

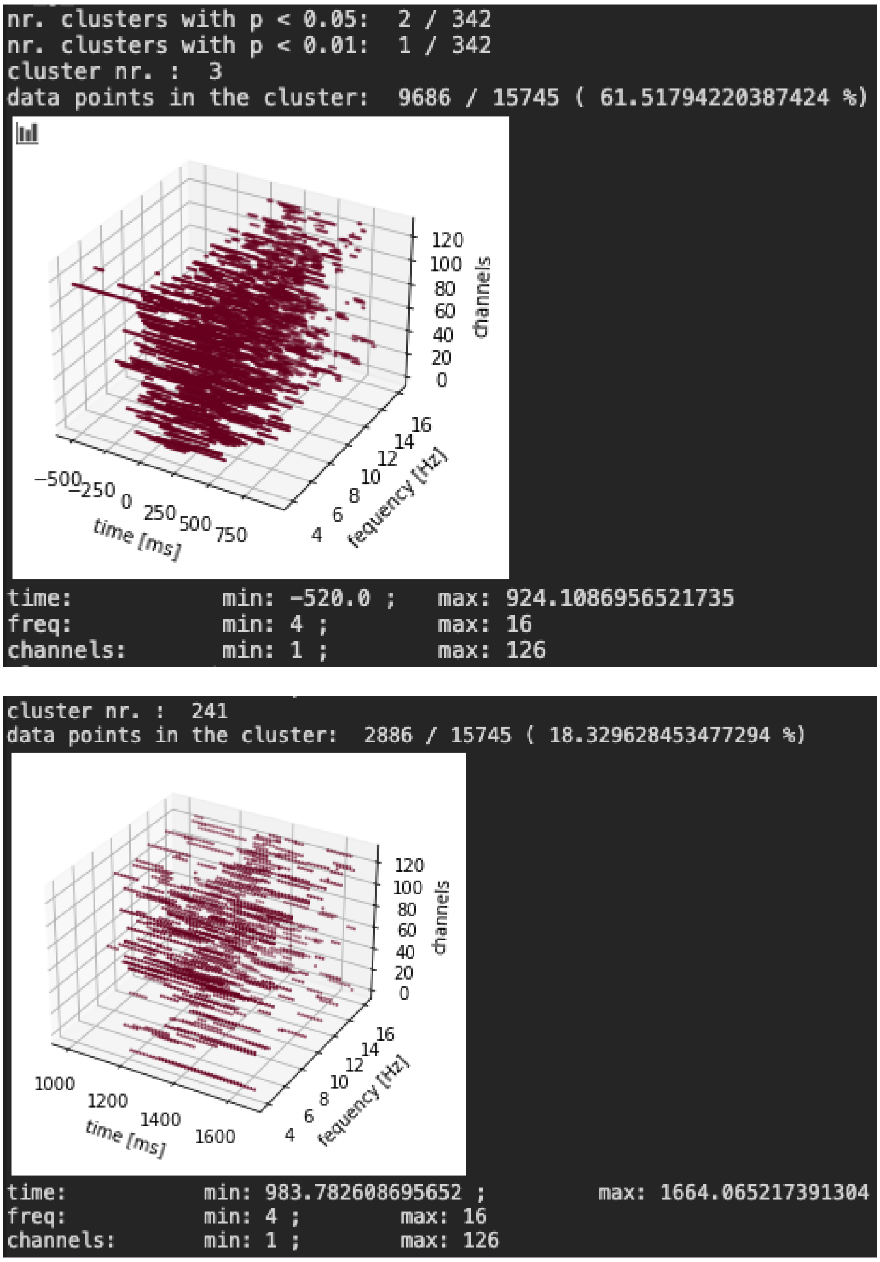

Figure B.2: Visualisation and information of the significant clusters found for the test between "Audiovisual /ga/ high" and "Audiovisual /ba/ high"

In the figure above it can be seen that the clusters span completely the frequencies and the channels whereas they are limited in time even if the cluster n. 3 (top figure) and the n. 241 (in the bottom) together span almost all of it.

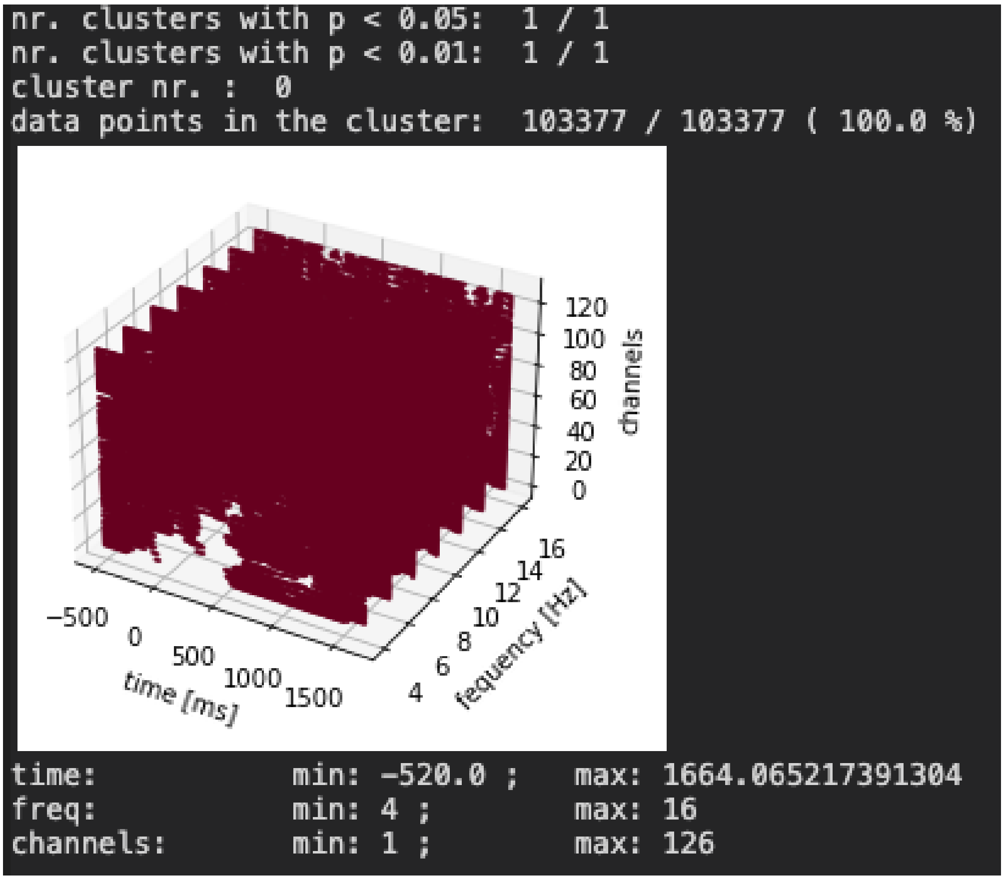

Figure B.3: Visualisation and information of the significant cluster found for the test between "Audiovisual /ba/ low" and "Audio /ba/ low" plus "Visual /ba/ low"

In the example above the results of the permutation cluster test for the "Audiovisual" against "Audio" plus "Visual" is shown. In this case, there is only one big cluster that spans the whole variable space. The data points removed in the first step of the permutation cluster test is centred around zero but, as the baseline is taken from 480ms to 580ms, the F-value of the points should be lower in that range than not around zero.

All these results are different from the one expected from the thesis [5]. The reasons why the results are so different should be analysed deeper. In this case, as said during the report, the focus was more on the methods so this aspect has been disregarded.

Thanks to this evaluation it is known that the results are not correct and are not reliable. For this reason the steps for the induced signals, that were at least suspicious, are definitely not correct.

Appendix C

Analysis of the threshold parameter for the permutation cluster test

An important part in the permutation cluster test is played by the threshold parameter. In general the lower the threshold, the broader the inspection and the fewer the clusters [14], [13]. In order to inspect how the threshold influences the results and to find the best threshold, a process of trials and errors has been applied.

First of all, we inspected the number of clusters with respect to the threshold, for each set of conditions analysed.

				nr. clusters		
				t: 0.5	t: 1	t: 5
Group3	Audiovisual vs audio+visual	/ba/	high	1	6	172
			medium	1	4	159
			low	1	6	179
		/ga/	high	4	11	208
			medium	1	7	211
			low	2	13	200
Group2	Audiovisual	/ba/	high vs medium	8	49	457
			medium vs low	9	65	425
			high vs low	8	51	444
		/ga/	high vs medium	15	75	442
			medium vs low	4	36	364
			high vs low	3	18	342
GroupTest	Visual /ga/ low vs visual /ba/ low			6	55	430
	Audio /ga/ low vs audio /ba/ low			4	42	508

Table C.1: Nr. clusters with respect to the threshold used

In Table C.1 it can be seen that, as expected, lowering the threshold results in fewer clusters whereas increasing the threshold leads to a higher number of clusters. In other words, when the threshold decreases, there are more points left in the first step and the clusters are bigger and fewer, and the analysis is done broader, for a more general effect. On the other hand, when the threshold increases, there are less data points remaining and the clusters are smaller and more, the effect inspected is more in detail, as the clusters focus on a smaller and more particular portion of data points.

The most important aspect to have a look at is probably the number of clusters for which $p < 0.05$, always distinguishing among the threshold used. The results are shown in the table below.

				nr. clusters with p <0.05		
				t: 0.5	t: 1	t: 5
Group3	Audiovisual vs audio+visual	/ba/	high	1	1	1
			medium	1	1	1
			low	1	1	1
		/ga/	high	1	1	1
			medium	1	1	1
			low	1	1	1
Group2	Audiovisual	/ba/	high vs medium	0	0	0
			medium vs low	0	0	0
			high vs low	0	0	0
		/ga/	high vs medium	0	0	0
			medium vs low	1	1	1
			high vs low	1	1	1
GroupTest	Visual /ga/ low vs visual /ba/ low			0	0	0
	Audio /ga/ low vs audio /ba/ low			0	0	0

Table C.2: Nr. clusters with p<0.05 with respect to the threshold

The table above shows that even if the number of clusters changes heavily, the amount of them that are significantly different is always the same. This means that the conditions are significantly different or not independently from the number of clusters. Moreover, the number of clusters significantly different is always the same, even if the total number of clusters changes greatly.

The previous result appeared a bit suspicious and further analysis on the clusters have been made. In particular, the number of elements not filtered out in the first step of the cluster test has been reported in the table below. In addition, also the percentage of elements put by the algorithm in the biggest cluster has been shown.

				nr. element left (in thousands)			percentage of elements in the biggest cluster (%)		
				t: 0.5	t: 1	t: 5	t: 0.5	t: 1	t: 5
Group3	Audiovisual vs audio+visual	/ba/	high	121	102	33	100	99.98	96.08
			medium	118	99	32	100	99.98	96.57
			low	121	103	34	100	99.96	96.45
		/ga/	high	113	92	28	99.98	99.98	93.64
			medium	117	98	30	100	99.96	94.53
			low	112	92	27	99.98	99.95	94.29
Group2	Audiovisual	/ba/	high vs med	82	54	4	99.97	99.66	4.30
			med vs low	83	54	4	99.97	99.59	10.36
			high vs low	82	54	4	99.98	99.65	6.41
		/ga/	high vs med	76	47	3	99.94	99.22	1.76
			med vs low	103	78	13	99.99	99.88	59.99
			high vs low	108	84	15	99.99	99.84	61.51
GroupTest	Visual vs visual	/ga/ vs /ba/	low vs low	85	56	7	99.99	99.62	21.74
	Audio vs audio	/ga/ vs /ba/	low vs low	84	56	5	99.99	99.53	3.51

Table C.3: Nr. elements not filtered out and percentage of elements left in the biggest cluster

In Table C.3 can be seen that, as expected by the theory and intuitively, the higher the threshold the lower the amount of elements not filtered out.

The percentage of data in the biggest cluster shows that, eventually, almost the totality of points are clustered together in the biggest cluster and then most of the other clusters have only few elements. There are some exception, when $t = 5$, for which the percentage is not as high as in the other cases but, nevertheless, the biggest cluster is still considerably bigger than the others. Moreover, the cases for which the biggest cluster contains less than 50% of the points are also the ones for which no clusters are significantly different.

The distribution of the points in the clusters suggests that, even though the number of clusters increases heavily, at least when there is significant difference, there is always one big cluster and that cluster is the one for which the significance is arose. For this reason it has been decided to take $t = 0.5$, in this way there are few clusters and more data points left. In all cases, the results are consistent and the same results are obtained independently by the threshold used.

Appendix D

Personal comments

To conclude the report, I would like to add some comments relevant to the salient features of the course.

Throughout the duration of the course, I have learnt a number of theoretical topics such as EEG theory, differences between induced and evoked signals, calculation of signal powers and performing statistical tests.

The learning process was iterative, meaning that the literature reading was adapted with respect to the requirement at each step and recursively implementing the theory.

I also learnt how to look up and read through the most relevant research papers rather than readily available material from slides or books.

In addition to the same point, I also learnt how to work more independently in a way that I had the freedom to explore and implement my ideas with the guidance of the faculty.

Furthermore, I worked with new Matlab and Python libraries and new code implementations. Finally, I would like to say that even though the practical results are not in line with the theoretical expectation, I learnt a lot about researching theory and how to convert my literature reading into tangible ideas and code.