

PROJECT

02466 PROJECT WORK - BACHELOR OF ARTIFICIAL INTELLIGENCE AND DATA

ANALYSIS OF EEG SIGNALS

Cecilie Dahl Hvilsted (s214605), Ida Lund Raagart (s204010),

Katharina Strauss Søgaard (s214634)



June 19, 2023

1 Abstract

To understand how human perception differs under different circumstances, an increasingly big study is needed. This project considers a small piece of the investigation of perception, namely how information processing is affected by auditory and visual stimuli. Especially 2 peaks in electroencephalography (EEG signals), respectively named N1 and P2, are considered to be relevant for the audiovisual processing of information. The present method for analyzing EEG signals (by which the data is recorded) only takes 1 electrode, namely 'Cz' (located on top of the head), into account. For this analysis a reproduction of the results from the research paper "*Electrophysiological evidence for speech-specific audiovisual integration*" [12] is carried out. The data is gathered given 4 different stimuli, respectively auditory, visual, and both congruent and incongruent audiovisual stimuli. The comparison of EEG signals given the different stimuli is made by statistical t-test. This project examines if a 'synthetic' electrode will produce better results in an analysis of EEG signals. The 'synthetic' electrode is considered a linear combination of all 36 electrodes present in the study. It appears as a result of performing Group Independent Component Analysis (group ICA), such that the 'synthetic' electrode account for data from all test subjects. Prior to performing group ICA both individual Principal Component analysis (PCA) and group-level PCA is implemented. These steps center and whitens the data before it is inputted into the ICA algorithm. Throughout the analysis, several reality checks are implemented in order to make sure that the data analysis is done correctly. Lastly, a comparison of the t-test results of the 'Cz' electrode and the t-test results of the 'synthetic' electrode is carried out. The t-test for 'Cz' reveals that despite the test subject being conscious about the auditory stimuli being speech, and whether the auditory and visual stimuli are congruent or not, a significant difference in information processing is found at the N1 peak. T-test results for the 'synthetic' electrode (meaning that the test is based on back projected ICA data) show that on a significance level of 5 % only a significant difference at the N1 peak for the 'non-speech' group is found. This suggests that only when the test subjects are not aware that the sound is speech, the auditory-only data is different in the information processing compared to the auditory stimuli from audiovisual data. When comparing these t-test results to the 'Cz' channel t-test, they differ a bit. The 'Cz' channel t-test found several significant differences between stimuli for both speech groups and both peaks indicating a multisensory integration effect, while the 'synthetic' channel t-test only find an integration effect for the non-speech group at the N1 peak. Lastly to compare the two t-tests the effect size of all combinations of auditory stimuli across each group and each peak. This showed that in general, the effect size is larger for 'Cz' channel. However, the P2 peak for the speech mode has a much greater effect-size with the 'synthetic' electrode compared to the 'Cz' channel. This exact P2 peak for the speech group is relevant since the peak shows that a difference occurs when one knows whether a sound is speech or not. Therefore the 'synthetic' electrode is preferable compared to the 'Cz' channel.

2 Preface

During the last couple of years, the study of brain functioning and brain-related disorders has been found to play an important role in EEG signals. The most relevant feature of recording EEG signals to understand brain activity is that it is non-invasive as well as enables investigating brain response in real time. Nevertheless, several electrodes can capture the same brain activity. Thereby it is hard to determine where the actual activity occurs. Until now the analysis of the signals only uses 1 electrode, namely 'Cz', which is located on top of the head. Our analysis seeks to examine another method, namely using a linear combination of all electrodes in the analysis. It is thereby interesting to see if these 2 methods find the same results or if one is better than the other. However, the methods have caused us some trouble in determining the dimensions in the matrix calculations as well as visualizing our results as scalp maps (that reveal the areas of brain activity as well as the amount of activation). It means that our greatest problem has been the back projection of components, that are found from both individual and group-level PCA and group-level ICA. The back projection is done as only components in data space are possible to visualize. This is the main implementation of the analysis together with statistical tests of differences in brain activity occurring with different types of stimuli.

Contents

1 Abstract	1
2 Preface	2
3 Introduction	4
4 Data	5
4.1 Pre-processing	7
4.2 Ethical aspects	7
5 Methods	8
5.1 ERP and Grand Averages	8
5.2 Statistical test on Grand Averages	9
5.3 Factor-analysis	10
5.3.1 FastICA	10
5.4 Data shape	12
5.5 Whitening	12
5.5.1 Individual PCA	13
5.5.2 Group PCA	14
5.6 Group ICA	16
5.6.1 Reality check of individual ICA algorithm	17
5.6.2 Reality check of group ICA algorithm	18
5.6.3 Back projection of ICA components	19
6 Results	21
6.1 Grand Averages and statistical test on 'Cz' channel	21
6.2 PCA and dimension reduction	22
6.3 Scalp maps from individual PCA, group PCA, and group ICA	22
6.4 Grand averages and statistical test on back-projected data	27
7 Discussion	28
8 Conclusion	31
9 Appendix	34
9.1 Appendix A	34
9.2 Appendix B	35

3 Introduction

The complexity of the brain is difficult to address, and even more difficult when the relationship between the brain and its understanding of the world is taken into context [7]. Nevertheless one of the deeply researched areas of the brain is the areas processing sound and speech. This means that brain activity occurs in different areas according to how the auditory stimuli are presented. One tendency is that the perception of speech is dependent on the visual stimuli (i.e. what the eye sees) due to humans automatically starting to lip-read if possible. An example of this is the McGurk effect which is a phenomenon in audiovisual speech integration. Explained with an experiment, it compares the results of presenting test subjects with auditory-only stimuli to congruent audiovisual stimuli or McGurk stimuli. The McGurk stimuli is an incongruent visual syllable meaning that the sound coming from the auditory stimuli is different from what the face is saying (without sound), which corresponds to the visual stimuli [14]. Results of the experiment show that when incongruent audiovisual stimuli are presented, the test subject picks up a mix of the stimuli. Comparing congruent audiovisual stimuli to auditory-only stimuli results in a change in brain activity, indicating that the visual assists in decoding the sound.

An example of a study that studies the McGurk effect is the research paper "*Electrophysiological evidence for speech-specific audiovisual integration*" by Martijn Baart, Jeroen J. Stekelenburg and Jean Vroomen [12]. Here it was discovered that when humans are able to lipread while listening to speech, the human brain combines what we see with what we hear at different levels. The study is able to make this conclusion based on an investigation of different peaks in the electroencephalography signals (EEG signals), that are recorded from electrodes located on the scalp. More specifically, two peaks, namely N1 and P2, are found to be relevant for audiovisual speech processing [12]. The study reveals that the N1 peak occurred faster when people saw and heard speech together (audiovisual stimuli), compared to when they only heard speech (auditive stimuli). The time of a peak is measured on a time scale, that is chosen to be 1100 ms. This reveals faster processing of the sound when lips can be seen. It was also discovered that the P2 peak only responded differently between auditive and audiovisual data (both congruent and in-congruent) when people were aware that the sound was speech. Congruent audiovisual data occurs when the lips are saying the same as the sound. All in all, it showed that our brain tends to combine what we see and hear if possible in order to obtain a faster understanding of the sound. This is exactly what we would like to reproduce in this project.

Another relevant study is "*Group Independent Component Analysis for Event-Related EEG Data*" which uses the application of Independent Component Analysis (ICA) to EEG data at group level [8]. Their analysis shows that group-level ICA is effective in decomposing signals into a number of components as it entails a reduction of noise. It also results in an accurate reconstruction of event-related source signals (ERPs). This approach holds promise for enhancing our understanding of brain activity during specific events or tasks in a group context. In this case, group context is used as an expression for a situation that can account for several people (not necessarily at the same time).

These studies express several methods that are relevant for us to use in this project, as the objective of this project is to improve the research of EEG signals, that determine the McGurk effect.

As stated above, this project seeks to examine exactly how human perception changes when being aware

that the sound is speech, as well as being presented with both congruent or in-congruent audiovisual stimuli. It is based on the expectation of a change in the area of the brain that processes sound. This arises from the fact that the area processing speech is activated only when we are conscious of the stimuli being speech and if not, then the area processing random sounds is used [3]. The presentation of stimuli is throughout the report also referred to as events.

EEG has many advantages. One of them is that it is a non-invasive method of recording electrical activity in the brain using electrodes placed on the scalp. The EEG signals are generated by the electrical activity of neurons in the brain and can be used to study various aspects of brain function such as cognition, perception, and emotion [4]. This project focuses on the investigation of perception and the brain activity is captured as EEG signals for each individual test subject, by which the EEG signals then can be considered as event-related potentials (ERPs). They are referred to as event-related because the activity occurs as a result of a presentation of an event, namely a stimulus. Each trial is performed with 36 electrodes across the scalp of the subject. Despite having data for all 36 electrodes, in the typical analysis, only data from 1 electrode is considered, namely 'Cz' which is located right on top of the subject's head.

The method described above, accordingly only considering the 'Cz' electrode in the analysis, may not be the best way to analyze the data. A relevant method to consider is a combination of electrodes which might be more fitting. But due to the large variability across subjects, it is not likely that choosing the same electrodes/combination of electrodes would be ideal for all individuals. It means that these 2 approaches to analyzing ERPs only differ in the number of electrodes being considered. Another approach is using a combination of electrodes to analyze the data, which can be done by performing ICA. Particularly for this matter, we choose group-level ICA to produce a more generalized result accounting for several subjects. Also, it is stated in other studies that this method has been conducted to be adequate for EEG analysis. For each method, a 2-paired t-test is calculated to determine whether it found a significant difference between auditory data and auditory stimuli from audiovisual data (both congruent and in-congruent). In order to come to a conclusion about which of these methods is most suitable for similar analysis of event-related data, the methods are analyzed with coherent statistical analysis. More specifically, the effect size is calculated for each combination of stimuli across each group and for each peak. Lastly, this can be considered in order to evaluate the most suitable method for analyzing event-related brain activity. This will provide an understanding of how outer stimuli affects brain activity in humans.

This project thereby seeks to answer the question of whether a method of considering all electrodes when analyzing the data provides a more optimal way of analyzing event-related data, than only considering 1 electrode, namely 'Cz'.

4 Data

The data is extracted from a study made by Martijn Baart, Jeroen J. Stekelenburg and Jean Vroomen [12]. Besides that we are trying to reproduce the results from this research study, it has also been a large source for state-of-the-art on the topic of multi-sensory integration of information. The data has been an inspiration to a variety of different aspects such as detecting regularities from a sound sequence in

2-year-old infants [11], and it has also formed the basis of an investigation of whether autistic traits are associated with multi-sensory integration during information processing [16]. These 2 research studies mentioned above were both able to conclude the importance of multi-sensory integration when processing information in the brain. More specifically it was found that integration is developed already at the first months of life and also that is connected to autistic symptoms. Since the state-of-the-art covers a broad spectrum of research questions, it must also hold various different experimental tasks. The data that is examined in this project is made using visual enhancement of noise-embedded speech.

In Baart, Stekelenburg, and Vroomen's study titled "*Electrophysiological evidence for speech-specific audiovisual integration*" [12] they refer to 2 specific peaks, namely the N1 (found in interval 50-150 ms) and P2 (found in interval 150-250 ms) peaks. These were chosen to analyze since they are the first negative and positive deflections in EEG responses [15]. Also, in audiovisual speech processing, N1 (negative) and P2 (positive) show differential responses. N1 is suppressed by spatial alignment and predictability, while P2 is influenced by emotional content and phonetic congruence. This indicates their roles in integrating and processing multi-sensory information [12]. The peaks are what we seek to find in the time series of the test subjects as they reveal how fast the brain responds to stimuli.

The data set includes a total of 28 test subjects. To ensure randomization and balanced representation, these subjects are split randomly into two groups: a speech mode consisting of $\frac{N}{2}$ subjects, and a non-speech mode group with an equal number of test subjects. The stimuli used in the experiments involve recordings of a speaker saying the words "/tabi/" and "/tagi/" respectively. The subjects are exposed to both congruent and in-congruent audiovisual stimuli, where they listen to the audio while simultaneously observing a speaker's mouth articulating either the same word or the other word. This design allows for an examination of how the integration of auditory and visual affects perception. A third experiment is carried out with only auditive stimuli present for both groups. This is done in order to determine how humans' perception is affected when visual stimuli are added (both when being consistent with the auditive stimuli and not). To create a complete analysis of this a fourth experiment is carried out with only visual stimuli. In the analysis, the data from the visual stimuli is subtracted from the data from both types of audiovisual stimuli, since this will produce the brain activity caused by the auditive stimuli in the audiovisual data.

Each test subject is monitored with 36 electrodes by which the continuous EEG signals are recorded. The placement of electrodes on the scalp is placed by the extended International 10-20 system. For this experiment, they are sampled at a frequency of 512 Hz over a number of time series, each lasting 1100 ms. This is repeated to create approximately 570 epochs for each test subject for all 4 types of stimuli. The term 'epoch' is used to represent each recording of 1100 ms. It means that for each test person around 570 recordings are made.[12]

Since the data of all test subjects are stacked into 1 data matrix later in the analysis, the number of epochs for all test subjects must agree. Therefore a specific number of epochs is chosen to be 97 within each type of stimuli as this still is very representative of the recordings from each test subject.

4.1 Pre-processing

The data has been pre-processed by several different methods such that the data only holds informative and reliable data signals. Unreliable signals occur very much in this setting, since the electrodes pick up all present signals, meaning signals that occur due to irrelevant stimuli, eye blinks, and so on.

First, band-pass filtering in the interval 0.5 - 40 Hz is applied such that only frequencies in this interval are kept. This is due to the fact that task-related information is present in this interval and thereby reveals the brain's response to specific stimuli. Nevertheless, it also helps with noise reduction in the sense of not capturing source signals at higher frequencies such as environmental interference and muscle artifacts [21].

Another noise reduction method that is used is discarding values of activation, that is the potential differences, above 150 μ V. These don't relate to the presentation of the specific stimuli as they in fact are created otherwise for instance by eye movements. This exactly shows that EEG signals are sensitive to non-brain activity, such as muscle movements and eye blinks, which can contaminate the signals recorded from the brain. This causes a prepossessing of the data to be relevant to remove non-relevant frequencies in order to decrease the noise in the signals [4]. When plots of grand averages (an example is figure 11) across all test subjects are made, the noise is averaged out by averaging all individual epochs first and afterward combining this, to find the average signal of all test subjects. This exactly decreases the random noise that occurs from stimuli not relevant to the experiment.

The data is also pre-processed by down-sampling to 128 Hz instead of 512 Hz, which is the true sampling rate of the data. Down-sampling is performed in relation to Nyquist's Theorem, which states that the sampling rate should be twice as fast as the greatest frequency [5]. Since the data is filtered such that the largest frequency is 40 Hz, a sampling rate of 512 Hz is way higher than necessary. This means that we still make sure to access all frequencies in the spectrum. Down-sampling also affects the number of time steps in each epoch as the recording frequency is decreased. Each signal now only holds 141 time steps during the time period of 1100 ms (which is each epoch).

The signal is recorded before the actual stimuli are presented in order to obtain a baseline that holds the random noise in the signal. The baseline signal is recorded from -0.1 to 0 milliseconds (i.e. the first 100 ms of the recorded signal). The random noise of the signal between all test subjects has a large variability as is it dependent on each of the individual subjects. This is the reason why a lot of trials are performed on each subject. When averaging all the trials the true signal will still show as the stimuli are time-locked to begin at 0 sec.

4.2 Ethical aspects

Considering the ethical aspects of the data set which is found in the experiment analyzed in [12] both bias and privacy of the test subjects are relevant. Firstly, all participants gave informed consent prior to the experiment. Nevertheless, medical data is in general more personally sensitive meaning that the data must be gathered with great care. But it is not data that can identify an individual, but one should be aware of keeping it secure and anonymized. Apart from this is the bias of the data set. All participants are from the same country (the Netherlands) and 20 out of 28 participants are female, all of

similar age and with normal hearing as well as normal/corrected vision. This could cause problems for reproducibility as the test persons are not representative of the larger population. The sample size is also quite small with only 28 test persons, which again could cause problems in terms of being representative.

5 Methods

For the project, we used various packages, including the `mne` package, which specializes in visualizing and analyzing human neurophysiological data captured through EEG signals. Additionally, we used other packages `scipy`, `sklearn`, and `numpy`. The complete code for this project is in the GitHub directory https://github.com/cdhvilsted/Fagpakkeprojekt2023_EEG.git.

When using our code it is important to first run the `ICA_data_Import.py` script to create the datafiles used throughout the project. This was done this way because when running the other scripts many times it makes it faster to have the data treated to arrays before.

5.1 ERP and Grand Averages

A method of visualizing physiological data from EEG signals is by plotting the time series of the signals. This reveals how the brain activity, which is as activation, looks in a time sequence. For the data considered in this project, it is the brain activity over 1100 ms. These individual time sequences are referred to as epochs. The method is especially relevant when working with event-related data due to the fact that it reveals the event-related potentials (ERPs). The ERPs are visualizations of the time series for each test subject during the presentation of an event, in this case, a stimulus. The data work with the events of being represented with either auditory, visual, or either congruent or in-congruent audiovisual stimuli.

Since the data set is divided into a speech group and a non-speech group, we are interested in how the ERPs of these 2 groups differ in auditory data across all test subjects in each group. Thereby we would like to compare the auditory-only data (A) with the auditory stimuli from the audiovisual data (AV). This is done by subtracting the visual data (V) from the audiovisual data and obtaining AV - V for both congruent and incongruent audiovisual data. Then it is possible to subtract the effect of the visual stimuli as we isolate the auditory stimuli. Both a visual representation of the comparison as well as statistical tests is performed to show the interaction effect of auditory and visual stimuli.

The visual representation of the interaction effect seeks initially to reduce the random noise in the signals. The noise is averaged out by averaging the epochs of each individual person for the specific stimuli. This will reduce the noise due to the assumption that the noise is Gaussian distributed as $\epsilon \sim \mathcal{N}(0, \sigma^2)$. This is shown in figure 1 at which 2 random epochs for auditory data (A) and the auditory stimuli from congruent audiovisual stimuli (AVc - V), both for subject 1 in the non-speech group, are plotted. It states how both N1 and P2 peaks become clear when averaging across all epochs of the individual test subjects while they are not very present for each individual epoch. Afterward, to produce a grand average across all test subjects in the respective groups is calculated as the average signal across all test subjects' individually averaged data. The data of each group is thereby averaged twice (first individually, then combined) and is a grand average. The grand average ERPs for both groups are shown in figure 11 in the Results section.

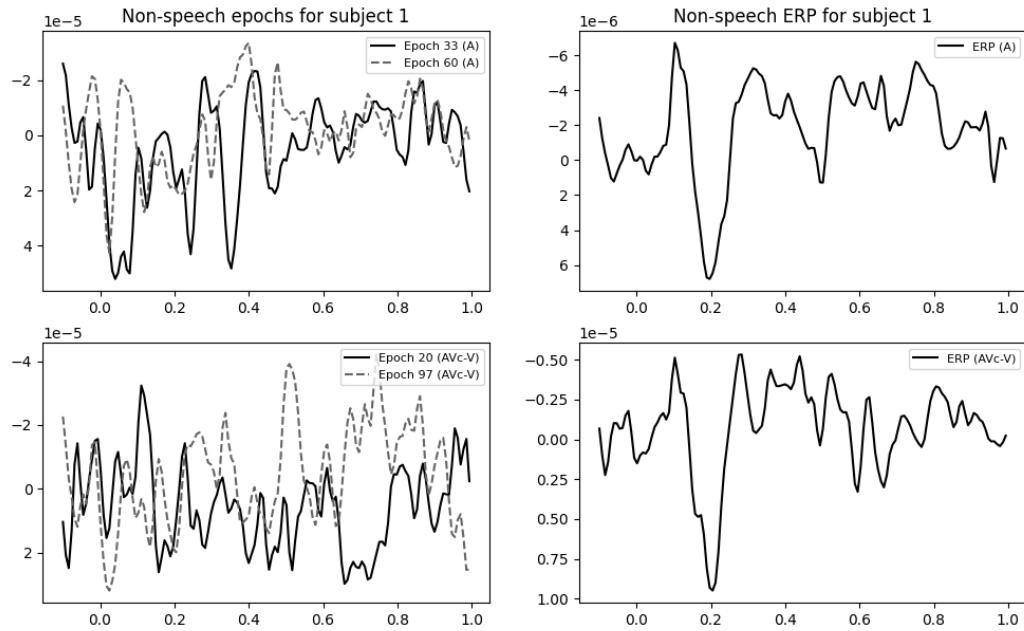


Figure 1: Visualization of individual epochs and the corresponding ERPs (averaged across all epochs for A and AVc-V stimuli) for subject 1 in the non-speech group. The figure has an inverted y-axis since it is the standard procedure for EEG data. Nevertheless, not everyone fulfills this and therefore one should be aware of the orientation. The y-axis holds the average value of activation. Also, the data is set to begin at time -0.100 milliseconds, which is due to the fact that the first 100 milliseconds are used as a baseline as this is before the stimuli are shown. Thereby the general brain activity from each subject is subtracted from the subsequent signal. This ensures that before the stimuli representation, the brain activity for all test subjects is close to 0. It means that the x-axis holds the times series of 1100 ms. It shows how both N1 and P2 peaks are clear when noise is averaged out and only the signals caused by the presentation of stimuli are shown (right). The noise is clear to see in the individual epochs where no clear peaks or tendencies across different epochs are seen (left). This is shown for auditory data (A) and auditory stimuli from congruent audiovisual data (AVc-V).

5.2 Statistical test on Grand Averages

The statistical test for the audiovisual interaction effect corresponds to the difference in amplitudes for auditory data (A) and audiovisual data minus the visual stimuli (AV - V). The specific statistical test is a paired 2-sample t-test. All test subjects are presented with the same stimuli, hence we start by averaging all epochs of each test subject. Then their minimum amplitudes for N1 are joined in a sample, as well as their maximum amplitudes for P2, to be used in 4 t-tests for the N1 component and 4 t-tests for the P2 component (2 for speech and 2 for non-speech mode). The t-test is paired since the data is obtained from the same test subjects, meaning that the data between the different stimuli can be related due to it being recorded from the same individual. Specifically, for both speech mode and non-speech mode we tested the difference in amplitudes between A and $AV_{Ic} - V$ (in-congruent audiovisual data with subtracted visual stimuli), and between A and $AV_c - V$ (congruent audiovisual data minus visual stimuli) as explained further in figure 2. As seen it corresponds to 8 paired t-tests in total which exactly compares how the auditory response differs when visual stimuli are integrated.

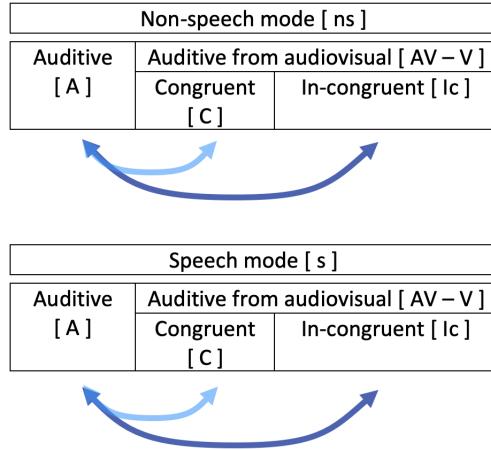


Figure 2: Method of comparing data in both speech and non-speech. The arrows show which comparisons are made between auditive data and auditive stimuli from the audiovisual data. All 4 tests are repeated for both peaks (N1 and P2) as well as for both groups (speech and non-speech). The two tests are a paired t-test and an effect-size test.

5.3 Factor-analysis

To analyze the data across test subjects, it is relevant to perform a factor-analysis method, in this case, group-level ICA [10]. We choose to perform the method on a group level instead of individual analysis as we seek to find a universal component that counts for all test subjects. The universal component will then hold the N1/P2 peak for all subjects, while if it were performed on an individual level, then the peaks would occur in different components for each subject. The universal component can also be considered as a 'synthetic' electrode that is a linear combination of all individual components from the two rounds of PCA, that is individual PCA and group-level PCA.

A reason why group-level ICA is appropriate to use is that EEG signals can't be assumed to be Gaussian distributed due to the fact, that the components are dependent on each other and similar components share variance. As PCA produces orthogonal and independent components, it is not a suitable method when it is a group-level analysis. Despite this, ICA produces non-orthogonal components and is therefore chosen to analyze the combined (and dimension-reduced) data.

We choose to make our own code for performing group-level ICA, as it is a quite complicated process to understand and keep track of achieving the correct dimensions of the resulting matrices. The process of ICA is inspired by the multiview ICA on the GitHub repository created by Hugorichard [18]. This group-level ICA algorithm uses both PCA and ICA to analyze the data. In this case, PCA is used to whiten the data before it is provided to the ICA algorithm. The different ICA algorithms in general do not vary much, so this was chosen as it was very accessible namely FastICA [1].

5.3.1 FastICA

The type of ICA, we use in this project, is sklearn's FastICA [9]. In FastICA, the goal is to estimate statistically independent components from our whitened group-level PCA data. For measuring the statistical independence, the Central Limit Theorem is central as independent variables seem to be more non-Gaussian distributed compared to dependent variables. The way this non-Gaussianity is measured

in FastICA is by the approximation of negentropy. Entropy (referred to as H) is a known concept from information theory, and for continuous random variables, it is called differential entropy. The calculation of differential entropy is stated in equation (1).

$$H(\mathbf{y}) = - \int f(\mathbf{y}) \log f(\mathbf{y}) d\mathbf{y} \quad (1)$$

In equation 1 the term $f(\mathbf{y})$ is the probability density function for the random variable \mathbf{y} . It is known that Gaussian distributed variables have the largest entropy among all random variables of equal variance.

The role of entropy in negentropy is used in FastICA, as it calculates the amount of information gain obtained when observing a variable. In our case the unobserved variables are components. This is used in the formula for negentropy $J(\mathbf{y})$, which is defined as the difference in entropy between the component \mathbf{y} and the random Gaussian variable $\mathbf{y}_{\text{gauss}}$ in the equation below:

$$J(\mathbf{y}) = H(\mathbf{y}_{\text{gauss}}) - H(\mathbf{y}) \quad (2)$$

It means that negentropy is a measure of non-Gaussianity calculated as the deviation from a Gaussian-distributed random variable. Since it measures the contrast between the 2 variables, it is often referred to as a contrast function. This formula states, that for more non-Gaussianity in a component, the larger value of negentropy, since the entropy of a non-Gaussian variable is small (since random Gaussian holds maximum entropy).

Some problems might occur when calculating the negentropy since all density distributions of the components must be known. To solve this an approximation of the negentropy is often used. Specifically, the sklearn's FastICA uses the approximation '*logcosh*', namely the logarithm of the hyperbolic cosine function. In this case, a non-quadratic function G is chosen for each component to be defined by equation (3).

$$G_n(\mathbf{y}) = \frac{1}{a_n} \log \cosh a_n \cdot \mathbf{y} \quad (3)$$

Using a_n as a constant in interval $[1, 2]$. Then the negentropy can be approximated to:

$$J(\mathbf{y}) \approx \sum_{i=1}^p k_i \cdot [\mathbb{E}[G_i(\mathbf{y})] - \mathbb{E}[G_i(\mathbf{y}_{\text{gauss}})]] \quad (4)$$

Initially, an unmixing matrix W is defined randomly holding random independent components. Then the optimal update for each entry in the unmixing matrix W is found by maximizing the non-Gaussianity of the components in the reduced data. The reduced data is the outcome of the first individual PCA and second group-level PCA. In this formula, k_i is a positive constant, and the random Gaussian variable is given by $\mathbf{y}_{\text{gauss}} \sim \mathcal{N}(0, 1)$. It is also noticeable that \mathbb{E} corresponds to the expectation, such that the difference is measured as the difference in the expected value of the non-quadratic function values of each variable.

The above steps are repeated until no further maximum non-Gaussianity can be found, meaning that the components in the unmixing matrix have converged into holding the most independent components [2].

5.4 Data shape

The methods of group-level PCA and group-level ICA presuppose that the data for each test subject is stacked to 1 long time series, such that a matrix can be arranged with the data series of all test subjects. As the data is divided into 2 groups 'non-speech' and 'speech' (each of 14 participants) the stacked 3-dimensional matrix holds the shape (14, 36, 54708). The data is recorded from 36 electrodes (referred to as channels) at 141 time steps for each of the 97 epochs. The number of epochs varies between the individuals in the range from 97 to 160 in each represented stimulus. It is necessary for the analysis that participants have the same amount of epochs, therefore the epochs are cropped to be 97 for all participants. Then the last dimension in the matrix is found by stacking the data from all time steps in all epochs giving $141 \cdot 97 = 13677$. Since we do this for 2 different types of stimuli, namely auditory and visual for both groups (non-speech and speech), then the last dimension becomes $13677 \cdot 4 = 54708$. This is illustrated in Figure 3.

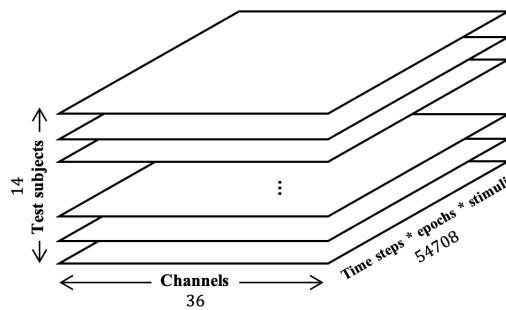


Figure 3: Sketch of stacked data X which for all 14 test subjects holds the data from 36 channels in rows and columns consisting of all time series for all epochs for the 2 types of stimuli for each group. Thereby the last dimension becomes $\text{epochs} * \text{timesteps} * \text{stimuliTypes} * \text{speechTypes} = 97 \cdot 141 \cdot 2 \cdot 2 = 54708$.

5.5 Whitening

Before using the ICA algorithm the data has been whitened and centered. Centering the data ensures that the scale of the data points is equal for all. The centering is done as in equation (5) for n being each row in data matrix X and the vector of ones being of length n , meaning that each row is centered by subtracting the value 1 from each value and multiplying by the mean of the row.

$$\tilde{\mathbf{X}}_n = \mathbf{X}_n - [1, 1, \dots, 1] \cdot \text{mean}(\mathbf{X}_n) \quad (5)$$

When whitening the data orthogonal components of the data are found. To do this we use the method Singular Value Decomposition (SVD) as it is a method for eigendecomposition [17]. This is exactly a method of PCA, which we seek to perform on both an individual and group level. By this, we obtain:

$$\tilde{\mathbf{X}} = \mathbf{U} \cdot \Sigma \cdot \mathbf{V}^T \quad (6)$$

The matrix Σ is a diagonal matrix of the eigenvalues, while both \mathbf{U} and \mathbf{V} are orthogonal matrices holding orthonormal vectors. Specifically, the matrix \mathbf{V} holds the eigenvectors, while matrix \mathbf{U} holds orthonormal vectors made from multiplying the data $\tilde{\mathbf{X}}$ with each eigenvector from matrix \mathbf{V} .

This method is repeated twice for the data to be the optimal input to the ICA algorithm. When performing PCA, dimension reduction is chosen in order to keep 95% of the variance in all data. The first round of PCA is on an individual level. Here the dimensions are reduced to ease the amount of computation needed. When reducing dimensions it means that we select a specific number of components less than 36. The second round of PCA is on a group level. The group level treats the subjects as one group, which is considered as some components explaining the same variance. We thereby seek to find the components that explain as much variance as possible across test subjects.

5.5.1 Individual PCA

The stacked data is used as input to the individual PCA, but at this point, we are able to access the data of each test subject in the stacked data matrix. For dimension reduction of the data matrix, PCA is used on the individual subjects. The outcome of PCA is a projection matrix \mathbf{R} as well as a matrix $\mathbf{X}_{reduced}$ of the reduced data. The matrix \mathbf{R} has the shape $(n1_{components}, 36, 14)$ and is visualized in figure 4. This is also referred to in step (1) in figure 8. The variable $n1_{components}$ is determined as finding the number of components holding at least 95% of the total variance in the data from each participant. This is possible since the components obtained by PCA each holds a relative amount of variance from the data (called the explained variance). The components are then sorted according to which holds most of the variance in the data.

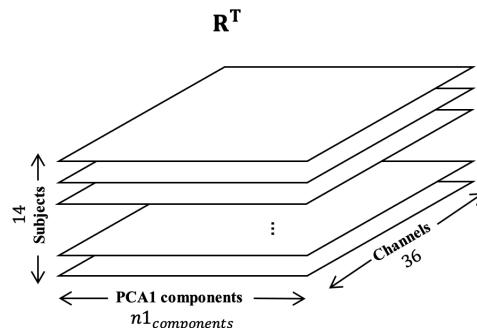


Figure 4: Matrix R^T that is 3 dimensional. It holds stacked 2-dimensional matrices for each test subject. The individual matrices have dimensions $(14, 36)$ as $n1_{components} = 14$ and data is recorded from 36 electrodes.

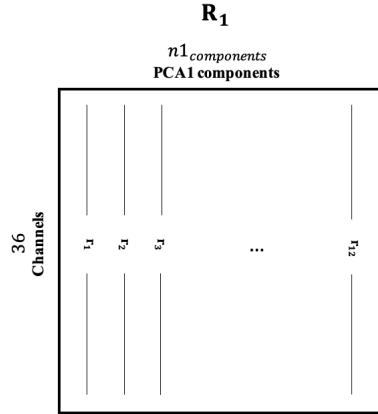


Figure 5: For each test subject a matrix as the one visualized here is constructed. As also expressed in figure 4, it holds individual principal components in columns and data from all 36 channels in rows. This corresponds to the matrix created for each individual test subject when performing step (1) in figure 8.

In order to find the optimal number of components in PCA, an investigation of the explained variance of each component is made. Explained variance is the fraction of how much of the total variance is held in each component. The concept is described in equation 7 [19] for n components of all M components, and σ^2 is found as the eigenvalues of matrix \mathbf{R} :

$$\text{Variance explained} = \frac{\sum_{i=1}^n \sigma_i^2}{\sum_{i=1}^M \sigma_i^2} \quad (7)$$

From this, the number of principal components covering a minimum of 95% of the variance is chosen. The specific value of components for individual PCA $n1_{components}$ is found to be the maximum number of components needed for the subjects (since the subjects, in this case, are individual). It means that we obtain a vector of explained variances for each subject. The largest number of components among all test subjects needed to cover 95% of all variance in data is 14.

5.5.2 Group PCA

PCA is now performed on the entire group to find the group-specific components, meaning that we do not distinguish between the individual subjects since it now is considered as a whole. For each subject $n1_{components}$ principal components are found from the stacked reduced data \mathbf{X} . It means that in step (2) there is not done any extra dimension reduction and thereby $n2_{components} = 14 \cdot n1_{components}$. This is explained further in the Discussion section. The output is then a projection matrix \mathbf{G} together with a matrix of reduced data \mathbf{Y} . The projection matrix \mathbf{G} has dimensions (196, 196). This is explained by $(n2_{components}, n1_{components} \cdot subjects)$, such that the first dimension is the group-level PCA components, while the second dimension comes from the individual PCA components for each test subject, as $14 \cdot 14 = 196$. An illustration of the matrix \mathbf{G} is shown in figure 6. The reduced matrix \mathbf{Y} is then used as input to the group-level ICA algorithm. Nevertheless, this step is done in the FastICA algorithm by setting the variable 'whiten'. It means that matrix \mathbf{G} is a FastICA attribute. The mixing matrix \mathbf{A} is found as a FastICA attribute as well, and when whitening is used it is already dotted with matrix \mathbf{G}^T . It means that the ICA components in \mathbf{A} are in individual PCA space. Thereby the reduced data matrix $\mathbf{X}_{reduced}$ from individual PCA is the actual input used for the ICA algorithm.

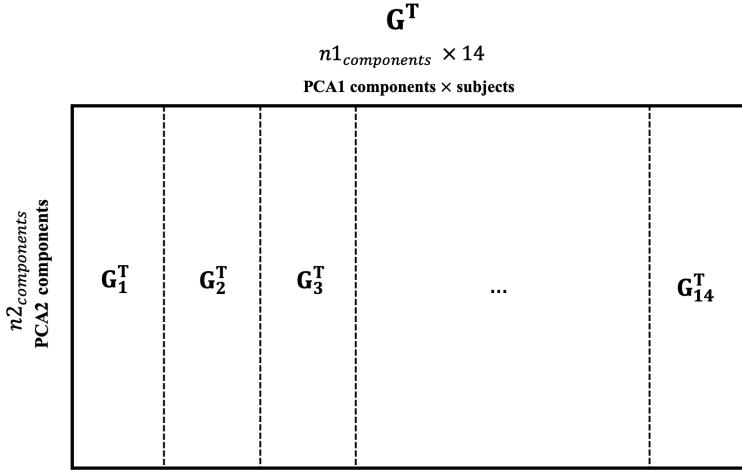


Figure 6: Visual representation of matrix G^T . The rows hold components from step (2) in 8 meaning the group-level PCA components. The columns consist of individual principal components from step (1) for each test subject.

For each individual test subject the matrix \mathbf{G}_n^T then has dimensions (196, 14) illustrated in figure 7.

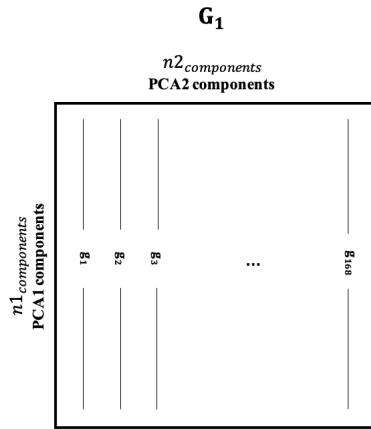


Figure 7: Each test subject has a corresponding matrix like this. Here it is illustrated for subject 1. The rows hold the group principal components while the columns are the individual principal components from step (1) in figure 8.

For group PCA in step (2) in Figure 8, the group-level PCA components in data space is found by back projecting the group PCA components onto the individual PCA component space. Then it is possible to visualize the principal components as well. The reconstruction is done as in equation 8 by indexing in both \mathbf{G} and \mathbf{R} to obtain the data corresponding to the same test subject n :

$$PCA2_{components} = \mathbf{R}_n \cdot \mathbf{G}_n \quad (8)$$

Matrix \mathbf{G} holds the components in group PCA space, while matrix \mathbf{R} holds the components from individual PCA space, which is used as the change of basis matrix. Since the matrices are orthogonal, the inverse matrix is the same as the transposed matrix, namely $\mathbf{G}^{-1} = \mathbf{G}^T$, which also holds for matrix \mathbf{R} . This is why we use the transposed change of basis matrix. In this case, all subjects are treated as one gathered group.

The input data for the ICA algorithm is now found. It takes the input of the reduced data found by individual PCA that whitens the data. As in step (2) there is no dimension reduction performed in step (3) in figure 8 since we are only interested in finding the sorted components holding the most variance. It means that $n_3 \text{components} = n_2 \text{components}$. The ICA components are sorted according to the percentage of variance explained, which is given in equation (19). The terms in the equation are explained beneath the equation.

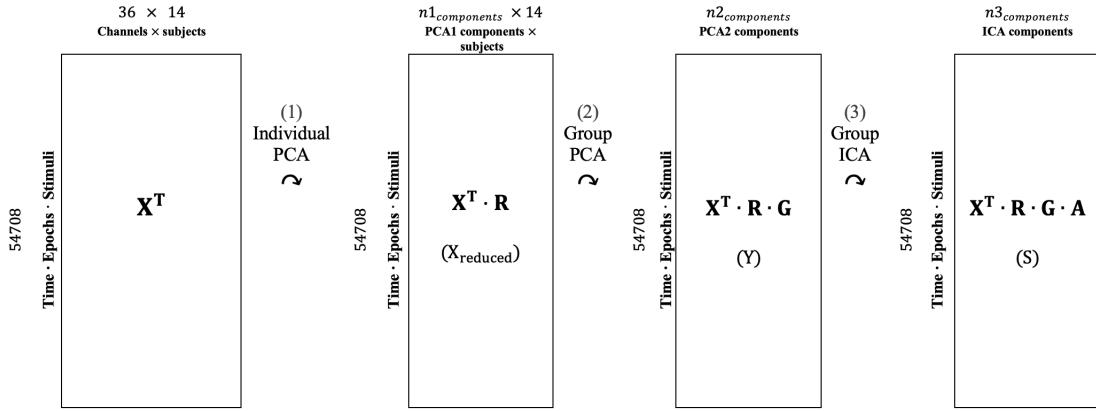


Figure 8: This figure shows the overall process of how we start with data \mathbf{X} , performing 2 times PCA and lastly ICA. The number of components used to reduce dimension in step (1) is $n_1 \text{components} = 14$. The number of components in step (2) is then $n_2 \text{components} = 14 \cdot 14 = 196$. Lastly, the number of components used when performing ICA is also $n_3 \text{components} = 196$.

5.6 Group ICA

The group-level ICA algorithm returns 3 matrices dividing the reduced data matrix \mathbf{Y} into a source matrix \mathbf{S} and a mixing matrix \mathbf{A} and the projection matrix \mathbf{W} . The projection matrix is the matrix that takes the data from PCA space into ICA space, meaning that it holds the independent components. The relationship between the mixing matrix \mathbf{A} and the unmixing matrix \mathbf{W} is $\mathbf{W}^{-1} = \mathbf{A}$. The inverse matrix \mathbf{W}^{-1} holds dimensions (196, 196) and is visualized in Figure 9.

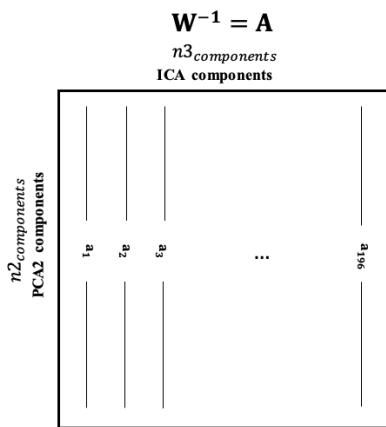


Figure 9: Inverse matrix \mathbf{W}^{-1} holding rows of group principal components and columns consists of the group independent components. The independent components are referred to as a_n in the figure. This is due to the relation $\mathbf{W}^{-1} = \mathbf{A}$.

The source matrix \mathbf{S} holds the time series (meaning the signals) for each independent component and the mixing matrix \mathbf{A} is used to hold the mixing of the true source signals, such that the input data can be reconstructed by $\mathbf{Y} = \mathbf{S} \cdot \mathbf{A}^T$. This corresponds to step (3) in figure 8.

5.6.1 Reality check of individual ICA algorithm

To make sure we are able to reconstruct our original data, we perform a reality check. We simulated some data from trigonometric functions. There are 2000 steps in *time* from 0 to 10 and the source is defined as the 3 signals:

$$S_1 = \sin(2 \cdot \pi \cdot time) \cdot 90 \quad (9)$$

$$S_2 = \cos(5 \cdot \pi \cdot time) \cdot 100 \quad (10)$$

$$S_3 = \sin(11 \cdot \pi \cdot time) \cdot 30 \quad (11)$$

From here, the source matrix is arranged as in equation 12.

$$\mathbf{S} = \begin{bmatrix} S_1 \\ S_2 \\ S_3 \end{bmatrix} \quad (12)$$

Here \mathbf{S} is referred to as the true source. A random mixing matrix with size (3×3) is chosen and the data is created by:

$$\mathbf{X} = \mathbf{S} \mathbf{A}^T \quad (13)$$

Then ICA and PCA are done separately and the following plots are the results shown in figure 15. Both methods are on an individual level since only 1 subject is considered with 3 channels.

We also checked that ICA was able to find an estimate of a mixing matrix, by back projecting to normal space $X_{back} = S_{ICA} \cdot A^T$.

The outcome of this reality check shows that the estimated sources from ICA and PCA differ relatively much (seen from plots with 'Simulated Data' in both ICA space and PCA space in figure 15). Comparing both plots to the plot of the true source reveals that ICA is much better at finding the correct frequencies. At the same time, ICA estimates the amplitude of all signals to be equal. This comes from the ICA mixing matrix. The estimated mixing matrices from ICA and PCA are scaled and compared to the true mixing matrix (that is randomly initialized and also scaled). The scaling is done as normalization by dividing with the norm of the specific column. Comparing the scaled ICA mixing matrix and the scaled true mixing matrix reveals that they are equal. In this case, the equality is accepted despite the components being in different rows such that the order of the columns is different. Also, the values of the components can hold different signs.

To get back from data in ICA space to the simulated data in normal channel space we use the back projection formula:

$$\mathbf{X} = S_{ICA} A_{ICA}^T \quad (14)$$

From this, we achieve the same plot as the top plot called 'Simulated Data in Normal Space' in figure 15. This holds for both PCA and ICA.

As we are now sure that ICA finds the true mixing matrix, we seek to perform the same analysis on our EEG data. Initially, we attempt to only use auditory data for subject 1 in the non-speech group. The individual data is inputted to PCA with no dimension reduction. When inputted to the PCA algorithm the data is transposed such that it fulfills dimensions $(n_{samples}, n_{features})$. Then the new data \mathbf{X}_{PCA} with size $(13677, 36)$. The ICA algorithm prefers the same input shape as PCA. Thereby the data can be given directly to ICA at this time. Visualizing the scalp maps obtained from the mixing matrix \mathbf{A} , the results are shown in figure 16.

Figure 16 also shows the time series corresponding to each individual component. These time series, which are held in the estimated source matrix \mathbf{S}_{ICA} are back-projected as stated in equation 15.

$$\mathbf{S} = \mathbf{S}_{ICA} \cdot \mathbf{A}^T \quad (15)$$

5.6.2 Reality check of group ICA algorithm

To validate our ICA method we use our method on simulated data. To do this we need to know the true sources beforehand. We expect that the estimated ICA sources will be able to identify this common source in terms of its wavelength as it will be statistically independent of the others.

We applied the same treatment, as for the EEG data set, to simulated data where the true sources are known. We constructed the simulation with two subjects, 1000 samples, and three sources for each subject. In our project data set, the sources can be considered equivalent to electrodes. To simulate the data, we created the sources as follows:

$$\begin{aligned} s_1 &= [\sin(13 \cdot \pi \cdot t), \quad \sin(3.3 \cdot \pi \cdot t) \cdot 3, \quad \cos(0.5 \cdot \pi \cdot t)] \\ s_2 &= [\cos(0.5 \cdot \pi \cdot t), \quad \sin(11 \cdot \pi \cdot t \cdot 3), \quad \sin(19 \cdot \pi \cdot t) \cdot 3] \end{aligned} \quad (16)$$

As can be seen, the true sources were generated as sinusoidal and cosine waves, with one source set to the same for the subjects (referred to as the common source).

$$\begin{aligned} \hat{\mathbf{S}} \text{ has dimensions: } & ((2, 3, 1000)) \\ \hat{s}_1 &= \mathbf{s}_1^T \\ \hat{s}_2 &= \mathbf{s}_2^T \\ \mathbf{S} &= \hat{\mathbf{S}}^T \end{aligned} \quad (17)$$

In the equation above t is the time variable (also sample index). The source number 2 for both subjects is the same, therefore we refer to this as our common source. In figure 17 we have plotted the true sources. The common source is marked with a bolder line and we expect the ICA, to estimate it well as one common source.

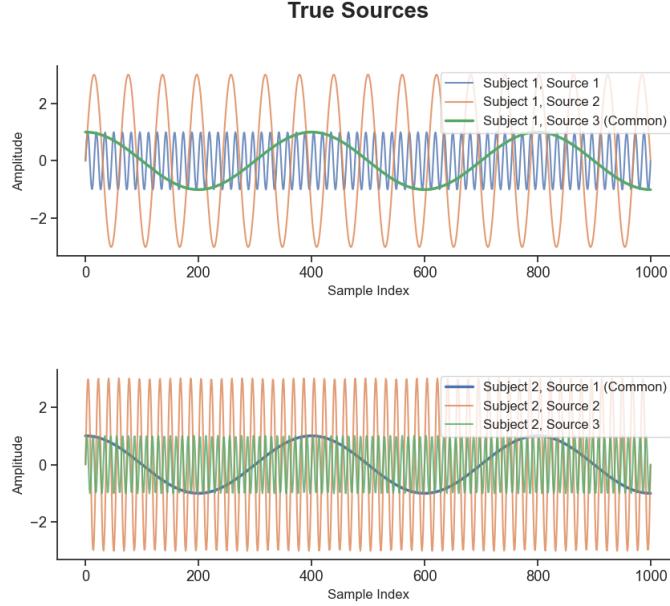


Figure 10: Each plot shows the 3 simulated sources for each test subject. The common source is highlighted with extra bold on both plots. It is also clear that the 2 other sources differ in frequency between the test subjects.

When making group ICA across subjects we expect that this source will be estimated. The mixing matrix \mathbf{A} is created randomly, and mixed with the source as $S_i \cdot A_i^T$ for i being the number of subjects.

Now we follow our group-level ICA method. We do PCA on each individual and concatenate all the data in PCA space to what we end up calling X_{PCA} . This is the data that we put into FastICA. The fast ICA was trained with a min 1000 iteration and a tolerance at $1 \cdot 10^{-7}$. From this, we get S_{ICA} which is the estimated source matrix. which is plotted in Figure 17.

We see in the figure that component 2 represents the wavelength of the common component. This, therefore, illustrates that the ICA performs as intended. To see component 2 plotted on its own it can be found in appendix 9.2.

5.6.3 Back projection of ICA components

For plotting the independent components in ICA space, it is required to project the components back into the initial channel space. To perform the back projection all 3 component matrices are used, namely \mathbf{R} , \mathbf{G} , and \mathbf{A} . Each holds components for individual PCA, group PCA, and group ICA. Then the back-projected ICA components in channel space occur as in equation 18. Once again, in matrices \mathbf{G} and \mathbf{R} we index n to capture the data of the individual test subjects.

$$ICA_{components} = \mathbf{R}_n \cdot \mathbf{G}_n \cdot \mathbf{A} \quad (18)$$

For sorting the independent components $n3_{components}$ that is outputted from the ICA algorithm, the percentage of variance accounted for is calculated for each component that came out of step (2) in figure 8. From this, it is then possible to choose the components accounting for the most variance of the data, despite the fact that several independent components can cover the same variance. The formula is given from [13] in equation 19:

$$\text{pvaf}(\text{component}) = 100 - 100 \cdot \frac{\text{mean}(\text{var}(\text{data} - \text{backProjection}))}{\text{mean}(\text{var}(\text{data}))} \quad (19)$$

The *backProjection* term in equation 20 is the reconstruction of the data sources \mathbf{S} after the reduction has been performed in step (3). It is dotted with a mixing matrix (inverse of the unmixing matrix \mathbf{W}), that holds the independent components obtained from performing ICA. The percentage of variance accounted for is thus calculated for the reduced data set $\mathbf{X}_{\text{reduced}}$ without each back projected component. However, it is important to keep in mind that the matrix \mathbf{Y} is a hidden intermediate link between $\mathbf{X}_{\text{reduced}}$ and \mathbf{S} that is incorporated in the FastICA algorithm.

$$\text{backProjection} = \mathbf{S} \cdot \mathbf{A}^T \quad (20)$$

It is now relevant to make ERPs for the source components in the ICA space and determine with statistical testing once again, whether the auditive data is significantly different from the auditive stimuli from the audiovisual data. To do this, the previous method of statistical testing and visualizing grand average ERPs is repeated.

Since the source components in the ICA space only hold 1-time series for all test subjects, a back projection of the data must be performed such that 1-time series for each test subject occurs. The original data for each test subject must thereby be projected onto the corresponding independent component in channel space (above this is mentioned as *ICA_{components}*). The most relevant independent component to consider is component 4 (time series and scalp maps are visualized in figure 18). As \mathbf{A}^{-1} , the unmixing matrix is already in *PCA1space* we are able to extract each person's individual component by indexing in the component, $A_{4,n}^{-1}$ meaning component 4 person n. The data is back projected onto this component. For n being the individual test subjects:

$$\mathbf{S}_n = \mathbf{X}_n \cdot \mathbf{R}_n \cdot A_{4,n}^{-1} \quad (21)$$

The component is normalized before indexing to ensure it doesn't influence the scale of the signals. As the S_n can be oriented incorrectly, such that the N1 peak is positive and the P2 peak is negative, each of the S_n signals are compared against the first S_0 by calculating the covariance and if it was negative the signal is flipped, $S_n = S_n \cdot -1$. Then it was checked if the orientation of all the signals pointed the correct way, having negative N1 and positive P2.

Then a t-test can be carried out, as we did for only the 'Cz'-component earlier. In doing so, we average the epochs across each test subject to obtaining an ERP for each test subject and identify both the minimum N1 peak as well as the maximum P2 peak within the average ERPs. We then take these peaks to the same paired t-test as illustrated in figure 2 to see if there are any differences. Also, a visual representation, namely the grand averages plot, is shown.

To determine the most suitable method for this type of data, given either the 'Cz' channel or the 'synthetic' channel, the effect size for each combination of stimuli across each group for both N1 and P2 peak is calculated. It means that we end up with 8 effect sizes, as for the t-tests. The effect size is calculated by fitting a Gaussian distribution to data given by $A - (AV - V)$. This is done for both congruency and in-congruency for both groups and for both peaks. For each fit the effect is given as:

$$\eta^2 = \frac{\mu}{\sigma} \quad (22)$$

For μ and σ being the mean value and the standard deviation from the Gaussian distribution. The value of η^2 then reveals the number of standard deviations, that the distribution of data points in $A - (AV - V)$ deviates from 0.

6 Results

6.1 Grand Averages and statistical test on 'Cz' channel

The grand average ERPs using only the 'Cz' channel, which is located on top of the head are shown in figure 11.

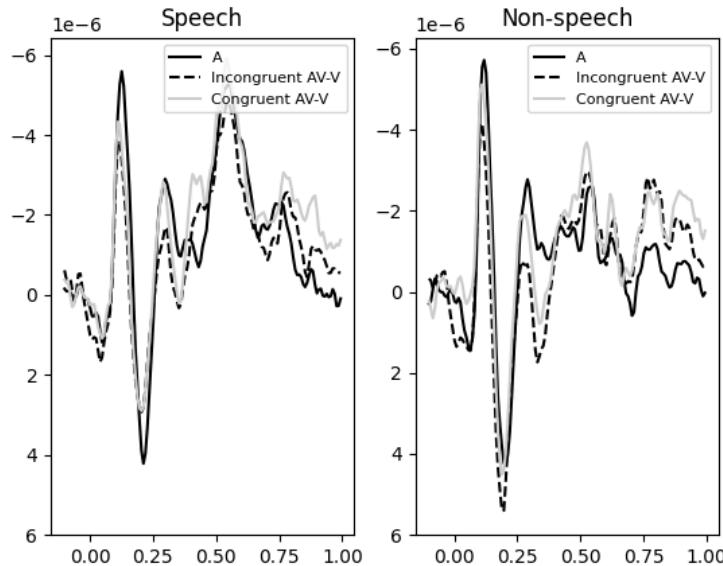


Figure 11: ERP's of grand averages across all test subjects on 'Cz' channel. The y-axis holds the average value of activation. The x-axis holds the times series of 1100 ms. The N1 and P2 peaks can be seen in both the speech and non-speech group, with greater peaks for the non-speech group.

Both an N1 and P2 peaks are present in figure 11 as the negative and positive peak occurring in the interval [0.0, 0.2], which is compatible with the grand average plot found in state-of-the-art article [12].

The results from the 8 paired t-tests are stated in Table 1.

Peak	N1		P2	
p-value	$p(A, AV_C - V)$	$p(A, AV_{Ic} - V)$	$p(A, AV_C - V)$	$p(A, AV_{Ic} - V)$
Speech	0.004	0.007	0.006	0.01
Non-speech	0.04	0.003	0.01	0.3

Table 1: Results from the t-tests on 'Cz' channel

The results of the statistical t-test above (for which only the 'Cz' electrode is considered) reveal, that when using a significance level of 5%, several significant differences between auditory data and the auditory

stimuli from the audiovisual data are found at every N1 peak despite of which group it is, as well as for P2 in speech group. This indicates that we found multisensory integration effects between auditory and visual stimuli for the speech group, and also at the N1 peak in the non-speech group. At the same time it also shows that for the non-speech group at P2 peak, congruency has an impact on the perception of the information since data from auditory only and congruent audiovisual stimuli without the visual stimuli is significantly different, while it is not significantly different from congruent auditory stimuli from audiovisual data.

Coherent with our results is, that in research paper [12] they found integration effects for both groups at the N1 peak, and only in the speech group at P2 from the statistical test.

6.2 PCA and dimension reduction

When individual PCA is performed, the explained variance of each component is calculated in order to determine the number of principal components to consider. For each 14 test subjects, the optimal number of components is stated in Table 2:

Subject number	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Number of components	6	6	5	14	11	11	3	2	12	4	6	3	6	6

Table 2: This table states how many principal components are optimal to consider when doing dimension reduction. The final optimal number that is used is the maximal of these 14 numbers.

The results from this showed that the maximal number of components needed across all test subjects is 14 principal components. To make sure that the data of all subjects hold at least 95% of the variance in the data. The number of individual principal components is thereby chosen to be 14, namely variable $n_{1,components} = 14$.

6.3 Scalp maps from individual PCA, group PCA, and group ICA

From the individual PCA in Figure 8 at step (1), the first 4 principal components were obtained and their corresponding scalp maps can be observed in Figure 12. Notably, subjects 1, 2, 6, 7, 9, 10, and 11 exhibit N1 and P2 peaks for the second component, while subjects 3 and 5 display these peaks in the third component. Furthermore, the principal component denoted 'component 0' is interpreted as the mean value of the entire data set holding most of the variance in the data, that is noise.

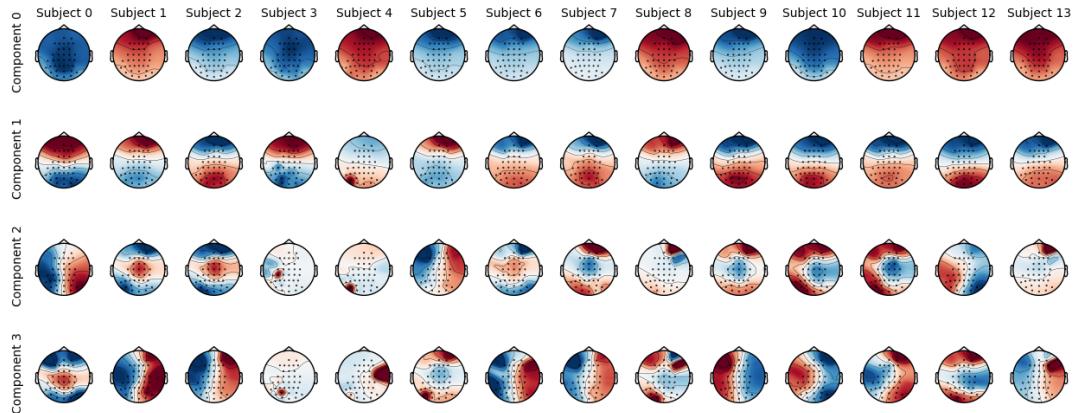


Figure 12: Across the 14 subjects, the principal components 0-3 are shown. The amount of activation is clear from the saturation of the respective color. The focus lies on identifying the activated component and the amount of activation. To determine whether it corresponds to an N1 or P2 peak, time series plotting is required.

We plotted the components in Figure 8 to validate if our process in step (1) seems correct, which it does. We also validated this with an example plot in Figure 13.

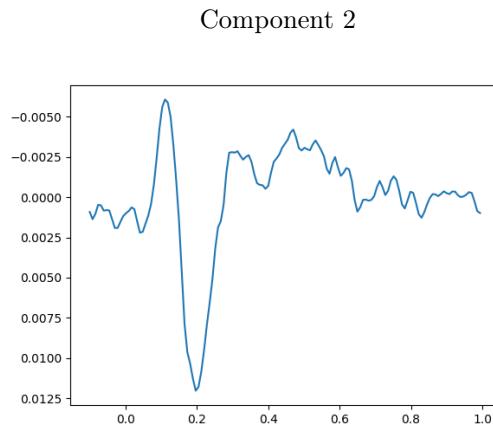


Figure 13: Here is an example of the time series plot for component 2, subject 1. The P2 peak is really prominent at time (horizontal axis) 0.2, also a small N1 peak can be observed. The y-axis holds the activation, not in μV but in an arbitrary unit. The x-axis holds the time from -0.1 to 1.

As the same plots were made in the second whitening PCA, the plots can be seen in Figure 14. Here we did not see as clear peaks as in Figure 12, which also makes sense to us, as the data does not fulfill the assumptions of PCA, namely that the data should be Gaussian distributed. As the method was done correctly, we decided to continue with the ICA as the next step.

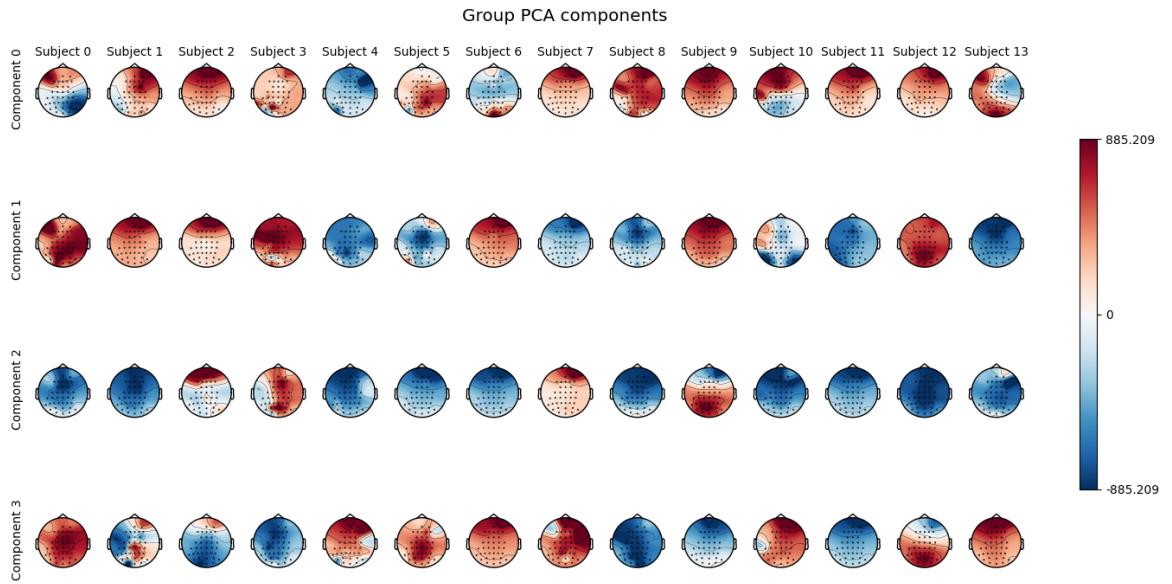


Figure 14: Across the 14 subjects, the principal components 0-3 are shown. While the exact interpretation of the blue and red colors as positive or negative values cannot be determined, it is established that one color represents negativity while the other represents positive. Nevertheless, the amount of activation is clear from the saturation of the respective color. The focus lies on identifying the activated component and the amount of activation. To determine whether it corresponds to an N1 or P2 peak, time series plotting is required.

For the individual ICA, the result of the simulated data is shown in Figure 15. The data is projected into both PCA space and ICA space.

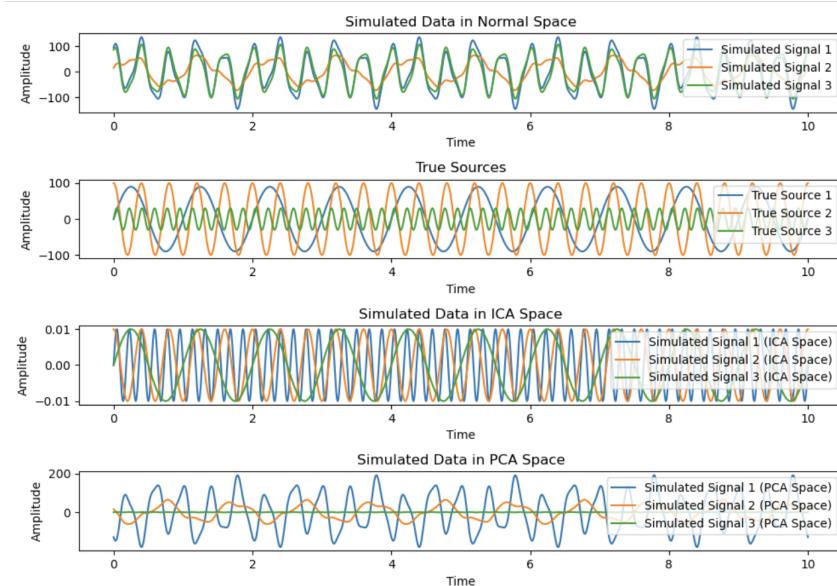


Figure 15: Sources have been simulated and mixed by both PCA and ICA. The signals are mixed using an initially random mixing matrix. From these plots, it is possible to investigate how ICA and PCA estimate the true source from the mixed signals. Afterward, an estimated mixing matrix is found for both methods as well. It is clear that ICA is better at finding the right frequencies of the sources than PCA. When data is reconstructed from the estimated source and mixing matrices the simulated data in data space occurs. We also find that ICA is able to find the true mixing matrix.

We see that the 'simulated data in ICA space' plot estimates the true sources in some degree. This suggests that ICA is better at finding the amplitudes of the true sources, while PCA does not manage

to unmix the simulated data into the true sources since it actually looks more similar to the plot of simulated data.

As a part of our reality check, we performed individual ICA before proceeding to group ICA. The results of individual ICA, when subject 1 from the non-speech group is considered are shown in figure 16.

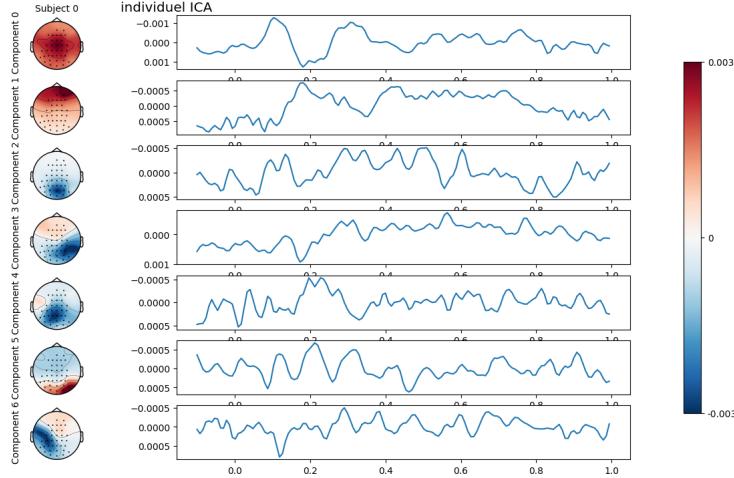


Figure 16: For subject 1 the auditory and visual data for both non-speech and speech is selected. Components 0-6 are plotted as scalp maps, as well as the corresponding time series for each component. It can be seen that component 0 has a scalp map with an active area surrounding the 'Cz' channel and a timeseries with a clear N1 and P2 peak. The components are sorted according to PVAF (equation 19). The timeseries plots have the activation on the y-axis, in an arbitrary unit, and the time on the x-axis.

Figure 16 shows smooth scalp maps with several components that reveal a centered distribution of activation, which corresponds to the N1 and P2 peaks. Especially components 1 and component 3 hold clear scalp maps. For their time series, component 1 seems to have the opposite peaks of N1/P2, since the earliest peak holds positive activation and the second peak holds negative activation. Despite this, the time series for component 3 looks very similar to having an N1 and P2 peak. This would therefore be the optimal component to work further with.

Since individual ICA is completed, the next step was group ICA. For Group ICA on simulated data, we got these estimated sources in Figure 17.

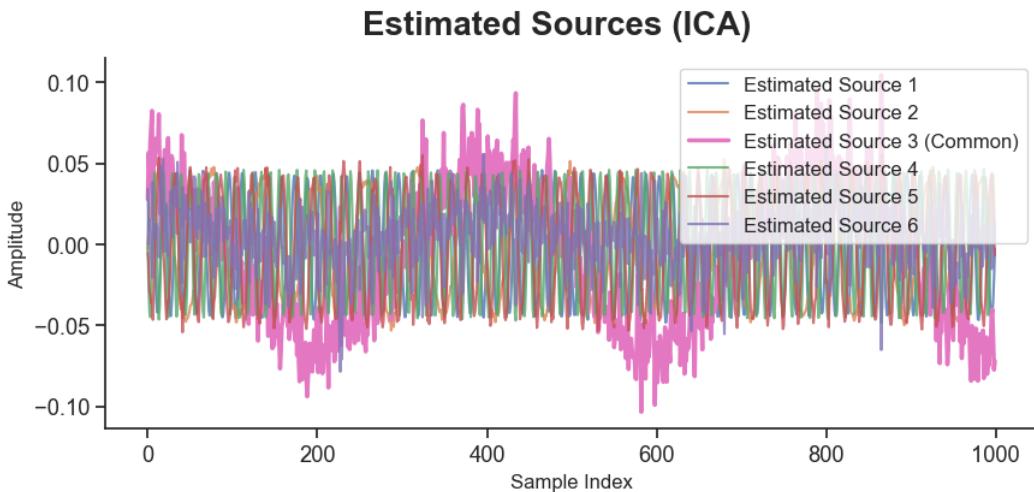


Figure 17: Estimated Sources for group ICA.

We see in Figure 17 that component 2 represents the wavelength of the common component. This, therefore, illustrates that the ICA performs as intended. To see component 2 plotted on its own it can be found in Appendix 9.2.

Group ICA on EEG data is exactly what the project has been about, namely making an analysis of different stimuli across all test subjects and finding an optimal component from this. The scalp maps and corresponding time series are visualized below:

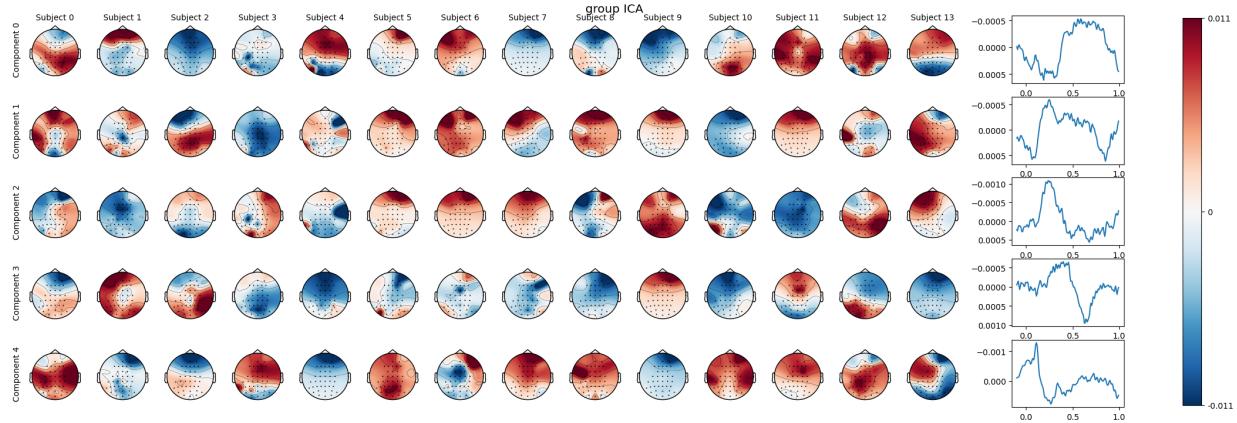


Figure 18: The 0-4 components and coherent time series are plotted for group ICA. The components are sorted using PVAF (equation 19). The most interesting component, meaning it reveals a centered scalp map around 'Cz' electrode, is component 4. Therefore this component is chosen to use for further work. The time series for component 4 is visualized in figure 19.

In Figure 18, we see the ICA components. It can be seen that the Group ICA was better than the PCA, at determining N1 and P2 peaks.

The optimal component holding the smoothest scalp maps for several subjects as well as the time series with the most present N1/P2 peaks is component 4. To investigate the time series for this component it is plotted in Figure 19.

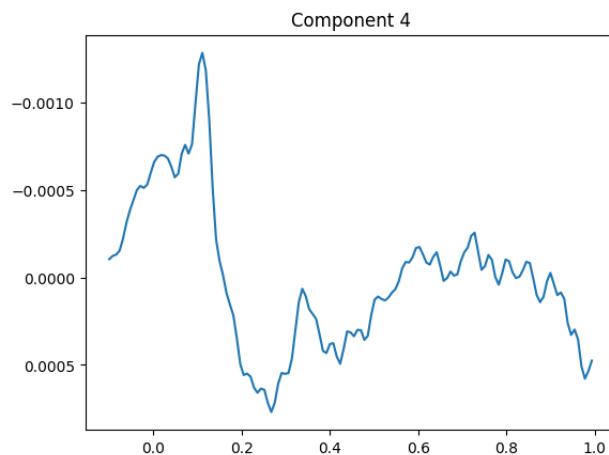


Figure 19: Time series of component 4 from group ICA in figure 18. This plot shows a clear N1 at around 100 ms and a clear P2 peak at 200 ms. The y-axis holds the activation in an arbitrary unit, and the x-axis holds the time.

This figure reveals a clear P2 peak at 200 ms. However, N1 is not that present for this component. Nev-

ertheless, both peaks do not need to be present in order for the data to tell something about audiovisual speech processing.

6.4 Grand averages and statistical test on back-projected data

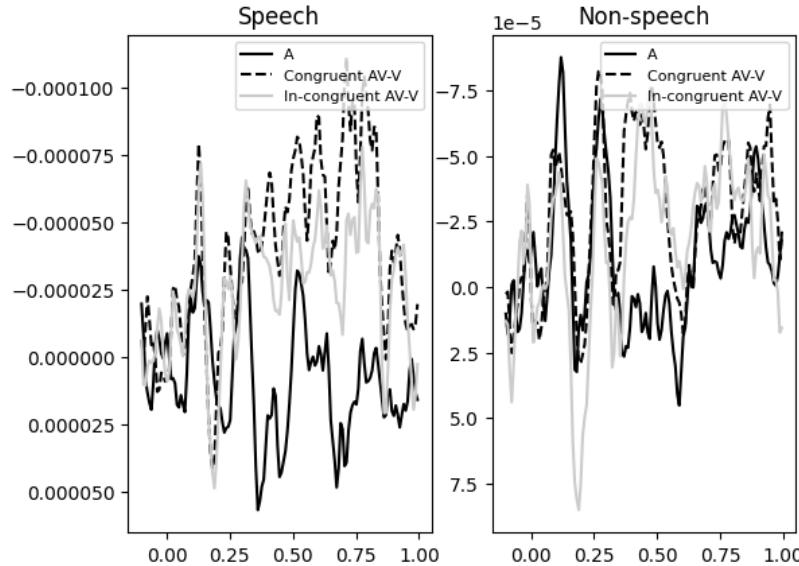


Figure 20: Grand averages for 'synthetic' channel from ICA data. For the group 'speech' both types of audiovisual both N1 and P2 peaks are present, while for auditory data the amplitudes are smaller. However, for the 'non-speech' group the peaks are more clear for all 3 types of stimuli. The y-axis holds the activation in an arbitrary unit, and the x-axis holds the time.

The statistical t-test on the ICA (back-projected) time series for all test subjects can be viewed as performed on a 'synthetic' channel, that is as a linear combination of all 36 channels that are included in the analysis. The results are stated in Table 3 below.

Peak	N1		P2	
p-value	$p(A, AV_C - V)$	$p(A, AV_{Ic} - V)$	$p(A, AV_C - V)$	$p(A, AV_{Ic} - V)$
Speech	0.27	0.37	0.064	0.056
Non-speech	0.0014	0.00011	0.15	0.11

Table 3: Results from the t-tests on a linear combination of all channels into one 'synthetic' channel

When comparing the 2 methods with different numbers of electrodes considered is given in Table 4 and Table 5 :

Peak	N1		P2	
	$p(A, AV_C - V)$	$p(A, AV_{Ic} - V)$	$p(A, AV_C - V)$	$p(A, AV_{Ic} - V)$
Effect for stimuli				
Speech	-0.97	-0.88	0.91	0.83
Non-speech	-0.64	-1.0	0.79	0.30

Table 4: Effect sizes of 'Cz' channel. Effect-size is also called Cohen's d. The effect-size is calculated as stated in formula 22. It is calculated for comparisons for between auditory and auditory part of audiovisual data.

Peak	N1		P2	
Effect for stimuli	$p(A, AV_C - V)$	$p(A, AV_{Ic} - V)$	$p(A, AV_C - V)$	$p(A, AV_{Ic} - V)$
Speech	-0.32	-0.26	-1.12	-1.51
Non-speech	0.56	0.58	0.43	0.47

Table 5: Effect sizes of 'synthetic' channel from back projected ICA data. The effect-size is calculated as stated in formula 22. It is calculated for comparisons for between auditive and auditive part of audiovisual data.

7 Discussion

Considering Figure 11, which we expected to be equal to Figure 3a (shown in appendix 9.1) in the research paper '*Electrophysiological evidence for speech-specific audiovisual integration*' by Martijn Baart, Jeroen J. Stecklenburg, and Jean Vroomen [12], then the 2 figures differs in such way that the auditive data is skewed positively in our figure. It means that the auditive data takes larger values for the speech group in our figure than in the referenced figure. At the same time, the incongruent audiovisual (without the visual) takes larger values for the non-speech group. There may be two reasons that can describe the differences. Before we handle the data preprocessing has been done, which discard signals that exceed one of several thresholds. The research article used band-pass filtering from 0.5-30 Hz and rejected absolute amplitudes exceeding 120 microvolts. Whilst the preprocessing done to the data used has a filter of 0.5-40 Hz and rejects absolute amplitudes exceeding 150 microvolts on non-frontal channels. This could explain why our data has greater amplitudes since the threshold is much lower. Considering discarding signals, we also use the same number of epochs for all test subjects in all stimuli. This means that we discard many epochs for some subjects since they on average hold approximately 570 epochs across the different stimuli. The discarding is done to make implementation easier as it then is equal for all subjects. Nevertheless, a lot of epochs are rejected and not used, which may cause our results from differing even more. But despite the different preprocessing and cutting down epochs, the figures look very similar otherwise. Specifically the two peaks N1 and P2 that we wish to study are very prominent in both figures.

When we concatenate the epochs, because of the given data set, the distribution of results from 'tabi' and 'tagi' are not necessarily equal, when used in the t-test. This might not be fair since it then will be biased towards one of them. A way of coming around this bias could be to manually select approximately $\frac{97}{2}$ from each of the events within the stimuli. This would make sure that the data is equally distributed among the bias, which might be skewed as our data is imported into this project. Another bias to consider is the small sample size of only 28 subjects. Despite each subject contributing to around 570 epochs in total, it might be relevant to capture EEG signals from more subjects in order to be able to generalize the results across different types of people. This also comes from the fact that almost all test subjects are women, and all test subjects are from the same geographic place and approximately all the same age, which is quite young. This causes the whole analysis to be biased towards young Dutch women, which could have been avoided if the EEG signals were recorded on a group of different ages, gender, and provenance.

We fit the ICA algorithm on only the auditive and visual data from both the speech and non-speech groups. This was done as we assumed they would have the clearest N1 and P2 peaks as a common tendency across all test subjects. However, this might not have been most ideal, as the in-congruent and

congruent audiovisual data also holds data of the tendency as seen in the grand averages across the 'Cz' channel 11. But as we expected the two audiovisual data types (for each group) to have smaller peaks we choose not to include them in the training of the ICA algorithm.

When we perform the individual PCA we reduce the dimensionality of the data from 36 to 12 components. This is done to reduce the computational power needed as well as get rid of redundant information, whilst still keeping a minimum of 95% of the variance within the data. However, in the group PCA we do not reduce the dimensionality. This is due to the fact that reducing already reduced data can cause the data to lose important information, as $0.95 \cdot 0.95 = 0.90$, therefore we choose not to reduce the data any further. When performing group-level ICA we did not reduce the data as we are only interested in choosing the important components using PVAF, meaning that we use PVAF to sort the components and pick the most relevant.

When looking at the group-level ICA scalp maps in figure 19, it can be seen that not all scalp maps from the chosen component are having activity in the 'Cz' region as we expected. Those having the 'Cz' activation distribution are for example subjects 6, 12, and 13. However, their 'Cz' activations are different since some are negative and some are positive. This influences the orientation of S , and is one of the reasons why we have to manually correct the orientation when creating the grand averages, as well as perform the t-test and calculate the effect-size. All these 3 things are done for both data from the 'Cz' channel and for the data projected onto the specific ICA component (chosen as component 4). The time series for component 4 in figure 19 has a clear N1 and P2 peak but does have some noise and fluctuation throughout it. This indicates it may be a good component, but it may not be ideal.

Throughout the project, we performed several reality checks to ensure we understood how the data was presented and implemented. This exactly has caused us some problems during the process. Throughout the analysis, the initial individual PCA was straightforward to both understand and implement. The scalp maps, shown in figure 12, look smooth and correctly sorted such that the first component holds the most variance, which in our case will be noise. Also, on some scalp maps, it is clear to observe the centering around the 'Cz' electrode on top of the head, which indicates coherent N1/P2 peaks. Nevertheless, when back projecting the data into data space, it became slightly more difficult to receive some satisfying visualizations. A way of making sure that the data fulfills all assumptions of PCA, an example is equation 6. Since it did not seem to be enough to debug the errors, a last attempt of solving the problem was made. It involved simulations of data in smaller dimensions, such that we manually were able to investigate if the correct matrices were outputted. These small extra implementations created security where the error did not occur and thereby narrowed down the debugging.

However, we should probably have started with simulated data before moving on to real data. This is something we will use in other projects moving forward. This helped us understand how a 'common source' occurs across several subjects, and how ICA is able to separate it from the rest of the sources. It also showed us many of the tricky dimensions and especially helped us keep track of the different dimensions. When using sklearn for both the ICA and PCA, some matrices were retrieved and transposed whilst others were not. Smaller reality checks within the simulated data script helped us understand the dimensions. We got a better understanding of the back-projection and also how different PCA algorithms handle the data when it finds orthogonal components. It was also a lot simpler having only

3 sources/channels instead of 36, and only 2 subjects. It thereby contributed to understanding the data on a smaller level in the beginning, which made the interpretation of data with increasing dimensions easier.

The t-test values from the test of the data projected onto the chosen ICA component were not what we expected 3. More of the p-values were above the significance level of 0.05. The effect-size test however showed the P2 for speech had a higher effect size. The scale goes from small to huge, where $d = 0.8 = \text{large}$, and the absolute values of the effect sizes from the 'Cz' test lie between 0.8 – 0.95. The ICA test however had absolute values between 1.10 – 1.5, where the scale has $d = 1.2 = \text{very large}$ and $d = 2 = \text{huge}$. [6]. Therefore the ICA had a greater effect size and is, therefore, more useful for finding a difference between different stimuli.

Research regarding EEG signals has a broad impact and helps fulfill many of the UN's Sustainable Development Goals. First and foremost achieving a better and deeper understanding of EEG data improves research which can improve the way different issues are treated in the healthcare system. Several neurological disorders are diagnosed using EEG data, therefore more people can be diagnosed and receive the correct treatment. This works towards the third goal of Good Health and Well-being. By diagnosing more people and understanding how to give the best possible treatment goal 10 will be fulfilled by reducing inequality.

Secondly, using these methods within research will improve education and the understanding of EEG signals, which will work towards the fourth goal of Quality Education.

Some brain-computer interface (BCI) uses EEG signals to enable a user to control the computer using brain signals. This allows people with disabilities to use computers, who otherwise would be unable to use computers. By improving the way the EEG signals are processed greater control can be achieved, which can improve the functionality of the BCI. This will work towards goal 9 (industry, innovation, and infrastructure), goal 4 (quality education) as usage of computers is widely spread throughout the education system, goal 3 (good health and well-being) through less isolation and more independence [20].

BCI and Brain Machine Interface (BMI) have many use cases with both rehabilitation after stroke, treatment of neurological diseases, and usage within virtual reality (VR), augmented reality (AR), extended reality (XR), and within each 'reality' the application can be within rehabilitation, entertainment, education, usages in military and professional settings [20]. EEG signals are used as a non-invasive and relatively cheap method of measuring brain signals. However as proven throughout the project using raw EEG signals might be less appropriate and instead, proper processing techniques such as Group ICA can reduce the noise and find the important components within the EEG signals.

8 Conclusion

Initially, in our analysis, we tried to reproduce the results of a research paper that is very relevant for the state-of-the-art on this topic. The results of the statistical t-test for the 'Cz' channel produced results similar to the results in the research paper. More specifically the outcome revealed significant differences suggesting that multi-sensory integration causes a difference in brain activity when information processing. After both performing individual PCA, whitening and group-level ICA data have been reduced the dimensionality of components as well as back projecting the components from group-level PCA space into data space.

Throughout the analysis several reality checks are implemented to secure the data analysis proceeds correctly. An example hereof is testing whether the ICA algorithm finds the correct mixing matrix, and estimates true signals that are common among test subjects. This contributes to the optimal performance of the ICA algorithm. Nevertheless, a great difference between the methods cannot be found in this project.

The 2-paired t-test carried out on the back-projected data shows a significant difference in auditive data when the test subjects are not aware of the sound being speech at the first negative peak in the ERPs (called N1). The 2 methods are compared by calculating the effect size. This allows us to compare how close the distributions of data given different stimuli are to being zero-centered. This showed that in general, the effect size is larger for 'Cz' channel. However, the P2 peak for the speech mode has a much greater effect-size with the 'synthetic' electrode compared to the 'Cz' channel. This is exactly the peak we wanted to find a greater effect size for as the t-test for the 'Cz' channel found a significant difference in both of the speech P2 t-tests but not for both non-speech P2 t-tests. Therefore the 'synthetic' electrode is preferable compared to the 'Cz' channel.

Concludingly, Group ICA is a great resource for studying event-related potentials to avoid the manual selection of components, as it instead finds one relevant component across all subjects. This method can be used within many life-changing studies and innovations, as EEG data is used in both diagnostic and professional settings. By individualizing the selection of 'synthetic' electrodes but still having a group view, the outcome has a greater effect-size.

References

- [1] Aapo Hyvärinen. “Fast and robust fixed-point algorithms for independent component analysis”. eng. In: *Ieee Transactions on Neural Networks* 10.3 (1999), pp. 626–634. ISSN: 19410093, 10459227. DOI: 10.1109/72.761722.
- [2] A Hyvärinen and E Oja. “Independent component analysis: Algorithms and applications”. eng. In: *Neural Networks* 13.4-5 (2000), pp. 411–430. ISSN: 18792782, 08936080. DOI: 10.1016/S0893-6080(00)00026-5.
- [3] Dale Purves et al. *Neuroscience. 2nd edition.* eng. Sinauer Associates, 2001. ISBN: 0878937420, 9780878937424.
- [4] Scott Makeig et al. “Mining event-related brain dynamics”. In: *Trends in Cognitive Sciences* 8.5 (2004), pp. 204–210. DOI: 10.1016/j.tics.2004.03.008.
- [5] Walt Kester. “Sampling thoery”. eng. In: *Data Conversion Handbook*. Elsevier Inc., 2005, pp. 76–76. ISBN: 0080477011, 0750678410, 1281009997, 1417565985, 9780080477015, 9780750678414, 9781281009999, 9781417565986. DOI: 10.1016/B978-075067841-4/50011-7.
- [6] Shlomo S. Sawilowsky. “New Effect Size Rules of Thumb”. In: *Journal of Modern Applied Statistical Methods* 8.2 (Nov. 2009), pp. 597–599. ISSN: 1538-9472. DOI: 10.22237/jmasm/1257035100.
- [7] Danielle S Bassett and Michael S Gazzaniga. “Understanding complexity in the human brain”. eng. In: *Trends in Cognitive Sciences* 15.5 (2011), pp. 200–209. ISSN: 1879307x, 13646613. DOI: 10.1016/j.tics.2011.03.006.
- [8] Tom Eichele et al. “EEGIFT: Group Independent Component Analysis for Event-Related EEG Data”. In: *Computational Intelligence and Neuroscience* 2011 (2011), pp. 1–9. ISSN: 1687-5265. DOI: 10.1155/2011/129365.
- [9] Fabian Pedregosa et al. “Scikit-learn: Machine Learning in Python”. In: *Journal of Machine Learning Research* 12.85 (2011), pp. 2825–2830. URL: <http://jmlr.org/papers/v12/pedregosa11a.html>.
- [10] Seungjin Choi. “Independent component analysis”. eng. In: *Handbook of Natural Computing*. Vol. 1-4. Springer Berlin Heidelberg, 2012, pp. 435–459. ISBN: 1784026999, 3540929096, 354092910X, 354092910x, 3540929118, 9781784026998, 9783540929093, 9783540929109, 9783540929116. DOI: 10.1007/978-3-540-92910-9__13.
- [11] R A Otte et al. “Detecting violations of temporal regularities in waking and sleeping two-month-old infants”. In: *Biological Psychology* 92.2 (2013), pp. 315–322. ISSN: 0301-0511. DOI: <https://doi.org/10.1016/j.biopsych.2012.09.009>. URL: <https://www.sciencedirect.com/science/article/pii/S0301051112002013>.
- [12] Martijn Baart, Jeroen J Stekelenburg, and Jean Vroomen. “Electrophysiological evidence for speech-specific audiovisual integration”. eng. In: *Neuropsychologia* 53.1 (2014), pp. 115–121. ISSN: 00283932, 18733514. DOI: 10.1016/j.neuropsychologia.2013.11.011.
- [13] Makoto Miyakoshi. *ICA & percent variance accounted for*. Dec. 2014. URL: <https://sccn.ucsd.edu/pipermail/eeglablist/2014/009134.html>.

- [14] Michael S Beauchamp. “Audiovisual Speech Integration: Neural Substrates and Behavior”. eng. In: *Neurobiology of Language*. Elsevier Inc., 2015, pp. 515–526. ISBN: 0124077943, 0124078621, 9780124077942, 9780124078628. DOI: 10.1016/B978-0-12-407794-2.00042-0.
- [15] Michael Bruyns-Haylett et al. “The neurogenesis of P1 and N1: A concurrent EEG/LFP study”. eng. In: *Neuroimage* 146 (2017), pp. 575–588. ISSN: 10959572, 10538119. DOI: 10.1016/j.neuroimage.2016.09.034.
- [16] Thijs van Laarhoven, Jeroen J Stekelenburg, and Jean Vroomen. “Increased sub-clinical levels of autistic traits are associated with reduced multisensory integration of audiovisual speech”. eng. In: *Scientific Reports* 9.1 (2019). ISSN: 20452322. DOI: 10.1038/s41598-019-46084-0.
- [17] Reza Bagheri. *Understanding Singular Value Decomposition and its Application in Data Science*. Jan. 2020. URL: <https://towardsdatascience.com/understanding-singular-value-decomposition-and-its-application-in-data-science-388a54be95d>.
- [18] Hugo Richard et al. “Modeling Shared Responses in Neuroimaging Studies through MultiView ICA”. und. In: (2020). URL: <https://github.com/hugorichard/multiviewica/tree/master/multiviewica>.
- [19] Tue Herlau, Mikkel N. Schmidt, and Morten Mørup. *Introduction to Machine Learning and Data Mining*. 1.0. Technical University of Denmark, Apr. 2022.
- [20] Varun Kohli et al. “A review on Virtual Reality and Augmented Reality use-cases of Brain Computer Interface based applications for smart cities”. eng. In: *Microprocessors and Microsystems* 88 (2022). ISSN: 18729436, 01419331. DOI: 10.1016/j.micpro.2021.104392.
- [21] Arayamparambil C. Anilkumar Chetan S. Nayak. “EEG Normal Waveforms”. In: () .

9 Appendix

9.1 Appendix A

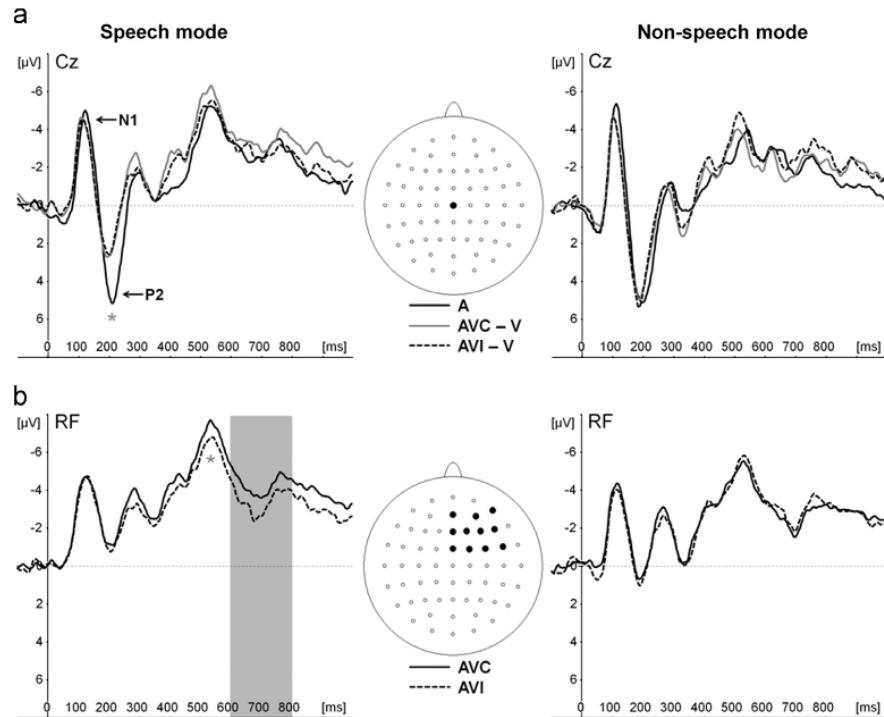
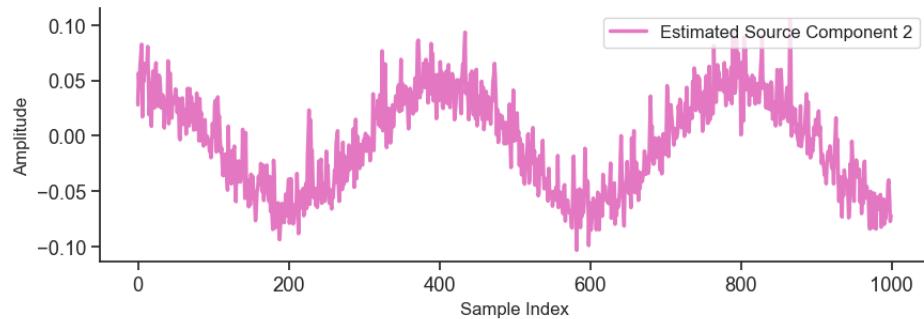


Fig. 3. (a) Averaged event-related potentials (ERPs) at Cz for auditory-only (A) and the audiovisual congruent – visual (AVC – V) and incongruent – visual (AVI – V) conditions for the speech mode (left) and non-speech mode (right). Lower panels (b) display the ERPs averaged across right-frontal electrodes for AV congruent- (AVC) and incongruent (AVI) stimuli. The asterisks and shaded area indicate significant peak- and mean amplitude differences between speech- and non-speech mode conditions.

Figure 21: Figure 3 from research paper 'Electrophysiological evidence for speech-specific audiovisual integration' by Martijn Baart, Jeroen J. Stecklenburg and Jean Vroomen [12]

9.2 Appendix B

Estimated Source Component 2 (Common)



Moving Average of Estimated Source Component 2

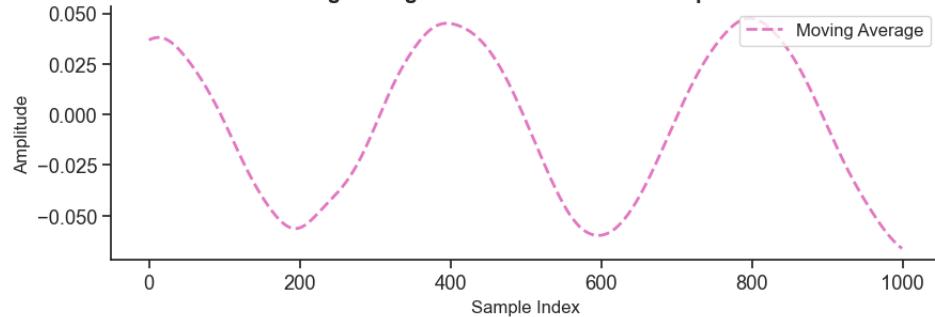


Figure 22: In this figure the estimated source 2 is plotted. The bottom plot is an moving average with window size 15