Comparison of Different Methods on EEG Signal Separating of Stuttering Adult and Child During the Pre-speech Auditory Modulation

Feiran Yang

A thesis
submitted in partial fulfillment of the
requirements for the degree of

Master of Science

University of Washington

2020

Reading Committee:

Ludo Max, Chair

Amy Orsborn

Program Authorized to Offer Degree:
Bioengineering

University of Washington

Abstract

Comparison of Different Methods on EEG Signal Separating of Stuttering Adult and Child During the Pre-speech Auditory Modulation

Feiran Yang

Chair of the Supervisory Committee:
Professor Ludo Max
Speech and Hearing Sciences

During the Event-related potential (ERP) study, ideally, the EEG recording only contains the event-related signal. However, there could exist irrelevant signals and noise. Unconscious activities, such as eye movement and muscle movement, and activities caused by the design of the experiment, could occur during the recording sessions. Meanwhile, due to the hyperactive nature of the child, there is more irrelevant signal inside child EEG signals. To solve this problem. there are three methods discussed in this paper, which are averaging, independent component analysis (ICA), and Autoencoder.

Averaging is the classical method applying to process data in ERP studies. Two advantages of this method are: 1) preserving the original information of the data most 2) eliminating non-activity-related Gaussian noise. There also are two pitfalls: 1) reducing the number of epoch in each group 2) failing to remove the irrelevant activity-related signals. This method is also unable to get useful information from the child data. And the signal to noise ratio (SNR) of this method is 30.21 for adult subjects.

ICA, a linear blind source separation method, is also a common method used by some of the studies. There are two advantages to this method: 1) preserving the number of epoch in each group. 2) removing the irrelevant eye movement and muscle movement signals. One pitfall is that bad rejection choice may cause losing information. This method improves some

of the result in child subjects. And the SNR of this method is 33.02 for adult subjects, which is higher than averaging.

Autoencoder is a nonlinear dimensionality reduction method. By creating proper loss function, a nonlinear independent feature learning method is applied to the EEG signals. The advantages are 1) nonlinearly learning the feature and linearly reconstructing the data at the same time 2) dimensionality reduction. One pitfall is currently no localization method to validate the features. And the SNR of this method is 22.94 for adult subjects, which is lower than averaging. And Autoencoder also can process part of the child data.

TABLE OF CONTENTS

LIST OF FIGURES

ACKNOWLEDGMENTS

First, I am very grateful to Dr. Max, for the opportunity to work in his lab and his guidance in my projects and this thesis. Then, I want to thank my lab mates for providing data to me and helping me when I have trouble. I really enjoy times in the lab and all those silly conversations.

Besides, I would like to thank my thesis committee member: Dr. Amy Orsborn, who takes time carefully reviewing my thesis.

Finally, I want to thank my two friends, Jing and Zhiqi, who spend almost every entire day zoom with me during this devastating period. It gives me humongous mental support to finish this thesis.

Chapter 1

INTRODUCTION

1.1 Stuttering

Stuttering is a speech disorder, characterized by prolongations and repetitions of certain sounds and syllables, and it affects speech fluency in roughly three million Americans. Considered as a physiological problem with important psychological consequences, People who stutter (PWS) often experience difficulties in communication, and many fear that it may affect their interpersonal relationships, education, and employment opportunities[2]. Different hypotheses have been put forward to explain stuttering, such as a failure to develop left-hemispheric dominance for speech[3] and malfunction in the auditory and speech system[4]. Recent advances in neuroscience have identified potential regions responsible for stuttering, such as the left inferior lateral cortex, pre-motor cortex, and basal ganglia[5]. While these regions have been identified, a singular network responsible for stuttering has not been fully described. The current standard of care for stuttering is limited to speech and behavioral therapy.

1.2 Experiment: Pre-speech Auditory Modulation

Previous studies observed abnormal prespeech auditory modulation in PWS comparing to nonstuttering individual[1][6].

The experiment is conducted on recruited subjects as showed in Figure 1.1. There are 3 different conditions, speaking (Figure 1.1 A), silent reading (Figure 1.1 B), and seeing (Figure 1.1 C). At time 0 ms, the screen will show the word on speaking and reading trials, and sign on seeing trials. At time 600 ms, there is an onset cue on the screen for each trial, and the subjects are requested to speak the word during the speaking trials. For $\frac{1}{3}$ of the trials in

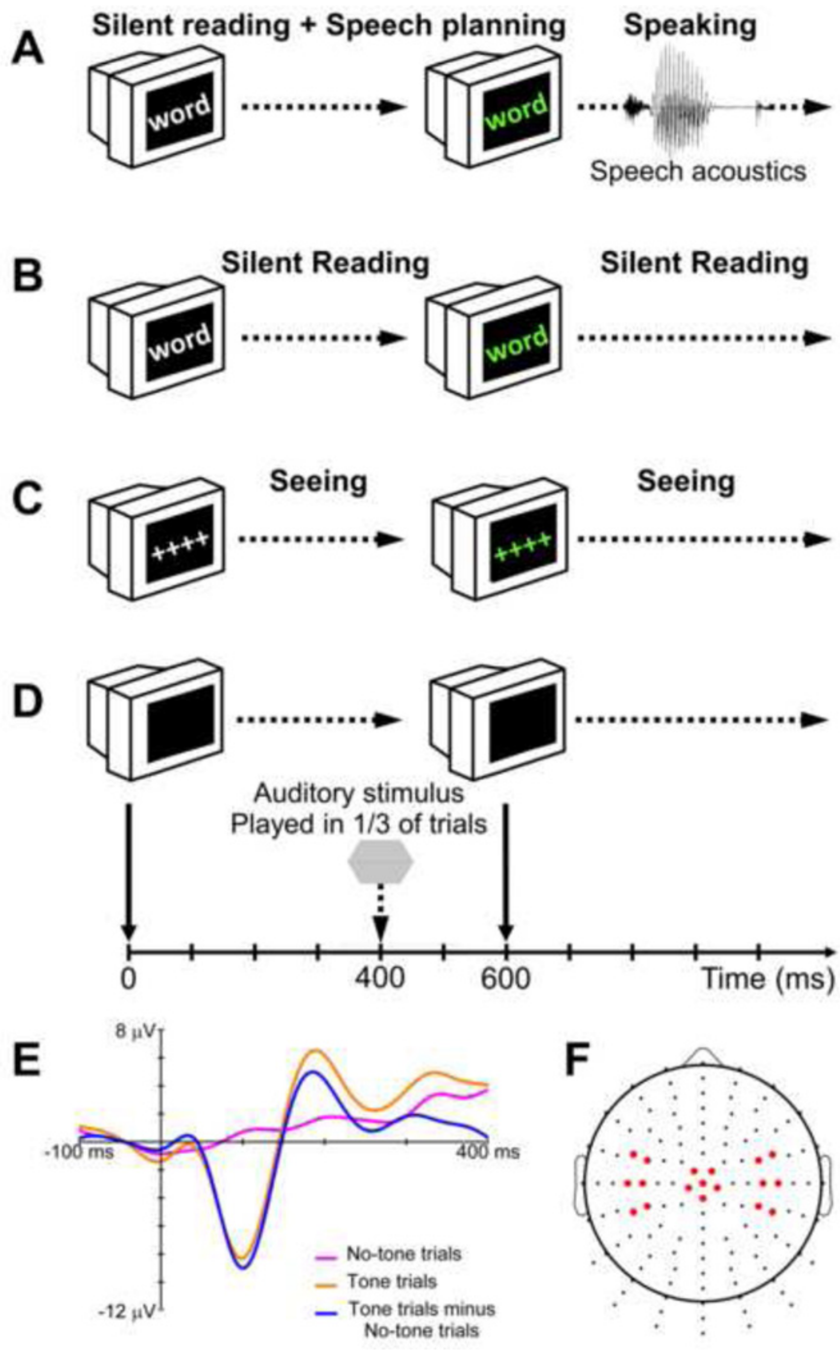

Figure 1.1: Experiment of modulation of auditory processing during speech movement planning[1]

each condition, there is an auditory stimulus (1 kHz, 40 ms duration, 75 dB SPL) onset at 400 ms. The words showed in this study are all monosyllabic consonant-vowel-consonant (CVC) words[1].

Figure 1.1 E is an example of the EEG recording the experiment. Figure 1.1 F is the region of interest (ROI) for auditory modulation.

1.3 Significance and Research Motivation

Event-related potential (ERP) is a technique to understand the brain. EEG recording has high temporal resolution and low spatial resolution. As a result, it is a powerful tool to understand the activities in the brain related to a specific event. Normally, the EEG recording contains event-related signals. There are also irrelevant signals, however, such as the noise caused by the resistance of scalp and power line noise, the signals generated by eye movement, and muscle movement, which were caused by speaking, irrelevant brain activities, and other artifacts[7].

There are some traditional methods to deal with those irrelevant signals. By using filters, some noise concentrates on the small range of frequencies that can be removed. The random noise can be canceled by each other just by averaging across each trial. For the nonevent-related signals, we can reject specific epochs, by setting an appropriate threshold on Electromyography (EMG) and Electrooculography (EOG) recording and also just by experiences.

However, sometimes the traditional methods can not solve all problems, such as on child, who is hyperactivity during the study. The rejection of bad epochs can't give us the expected waveform. Also by simply rejecting bad epochs, it'll also cause other problems, such as changing the number of trials in each group, and also the reduction of experiment trials number, which may also lead to the failure of the averaging due to the lack of epochs.

To solve this problem, I investigated and modify the ICA algorithms of artifacts removal of the EEG signals produced by stuttering children and adults during auditory modulation. The results on adult subjects were similar to the previously published results, but the algorithm was not as effective with children.

Looking for an alternative method, I contemplated the assumption that the EEG signals recorded from each channel were the linear combination of distinct brain activity inside the brain due to the principle of EEG recording, and concluded that there was a possibility I could find useful features using nonlinear methods. Meanwhile, the linear feature extraction methods, such as ICA and principal component analysis, generated the same number of components, which is far larger than the number of actual brain activities. Furthermore, my observations informed me that many components were high-frequency noise.

As such, I employed the Autoencoder that performed both nonlinear feature extraction and dimensional reduction as the basis of my new methods. The Autoencoder also enabled me to add constraints to the feature reduction method: prompting independence among each feature.

1.4 EEG Data Structure

Channel	Adult	Child
EEG	1-128	1-32
Reference	129,130	33,34
EOG	131,132	35,36
EMG	133-136	37-40
Mic	137	41
Total	137	41

Table 1.1: Data Structure

This paper contains both adult and child EEG data during the auditory modulation experiment. The data set included 28 adult right-handed subjects, and 14 of them are non-stuttering subjects. Three of the subjects in each group are female, and the rest 11 are male. The Age of stuttering subjects range from 19 to 43, and the nonstuttering subject is paired to each stuttering subject based on age (± 3) and sex. We also did 4 child test experiments, including 1 nonstuttering male child, 1 stuttering male child, and 2 nonstuttering female children. The EEG signals are recorded by the Biosemi active-electrode EEG system.

For the adult EEG data, we have 137 channels. And for the child EEG data, we have 41 channels. Detail of all channels is shown in Table 1.1.

Chapter 2
METHOD 1: AVERAGING AND REJECTION

2.1 Background and Motivation

EEG signal contains both neural signals and non-neural noise signals, including eye movement, muscle movement, noise from the environment, and etc. Some of the noise is small and constant, which can be eliminated by averaging across trials. And some noise are in the specific frequency, like line noise, can be eliminated by appropriate filter. Others are big and transient, like eye movement and muscle movement, and they are hard to be removed directly[7]. In this chapter, EOG recording and EMG recording were used as the indicators of the quality of each epoch. Two thresholds were set to reject bad epochs.

2.2 Methods

2.2.1 Data Pre-processing

Data was processed base on the EEGLAB toolbox[8]. First, all data were re-referenced by two reference channels, which are two electrodes located at the left and right mastoids. Then all channels were cleaned individually by filters. EEG signals were filtered by a band-pass filter (two-way least-squares FIR filtering) with cutoff frequency at 0.5 Hz and 15 Hz. EOG signals were filtered by band-pass filter with cutoff frequency at 1 Hz and 10 Hz. EMG signals were filtered by band-pass filter with cutoff frequency at 10 Hz and 300 Hz.

The data was separated into 4 different events, speaking tone trials, speaking no-tone trials, reading tone trials, and reading no-tone trials. The data was cut 100 ms before the onset of its corresponding event, and 400 ms after the onset of its corresponding event. Then the data was baselined by subtracting the average of data from -100 ms to 0 ms.

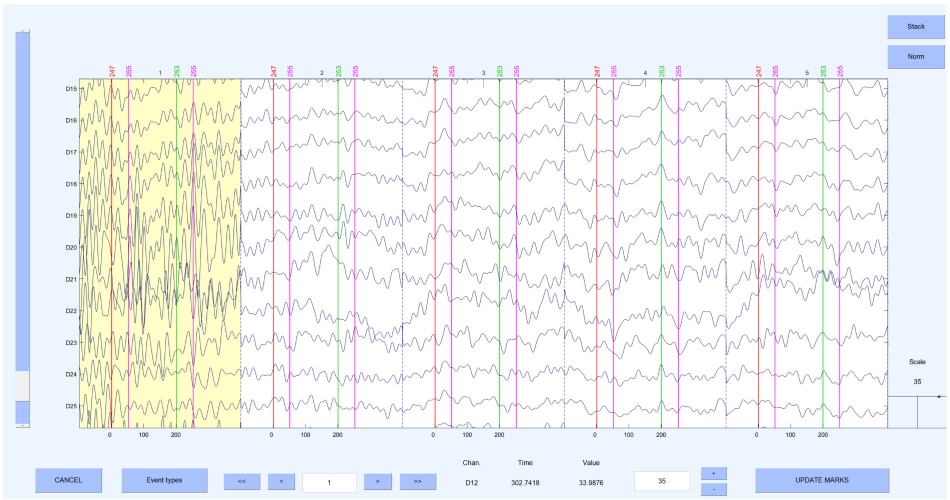

Figure 2.1: Example of bad epochs removing by threshold (EEG)

2.2.2 Rejection

Then the threshold for EEG channels ([-100 μV, 100 μV], Figure 2.1)was set as the indicator of the bad epoch, which was rejected. Meanwhile, the rest bad epochs were rejected manually by scrolling over the recording.

2.2.3 Averaging

The cleaned data first was averaged through different trials in the same groups. The averaged data was averaged across the channels within each ROI (Right ROI: 49, 50, 53, 54, 62, 63; Center ROI: 1, 2, 33, 65, 97, 111; Left ROI: 107, 108, 115, 116, 123, 124).

This study is focused on the pre-speech auditory modulation, thus the most important measure is the difference between the average of tone trials and the average of no-tone trials, which eliminates all the non-task-related neural signal[9]. First, we subtract the signal for no-tone trials from the signal for tone trials separately for the speaking task and the silent reading task. Then we compare the amplitude of N1 of the difference signal between both speaking and reading tasks.

2.3 Results

2.3.1 Adult

Difference Signal

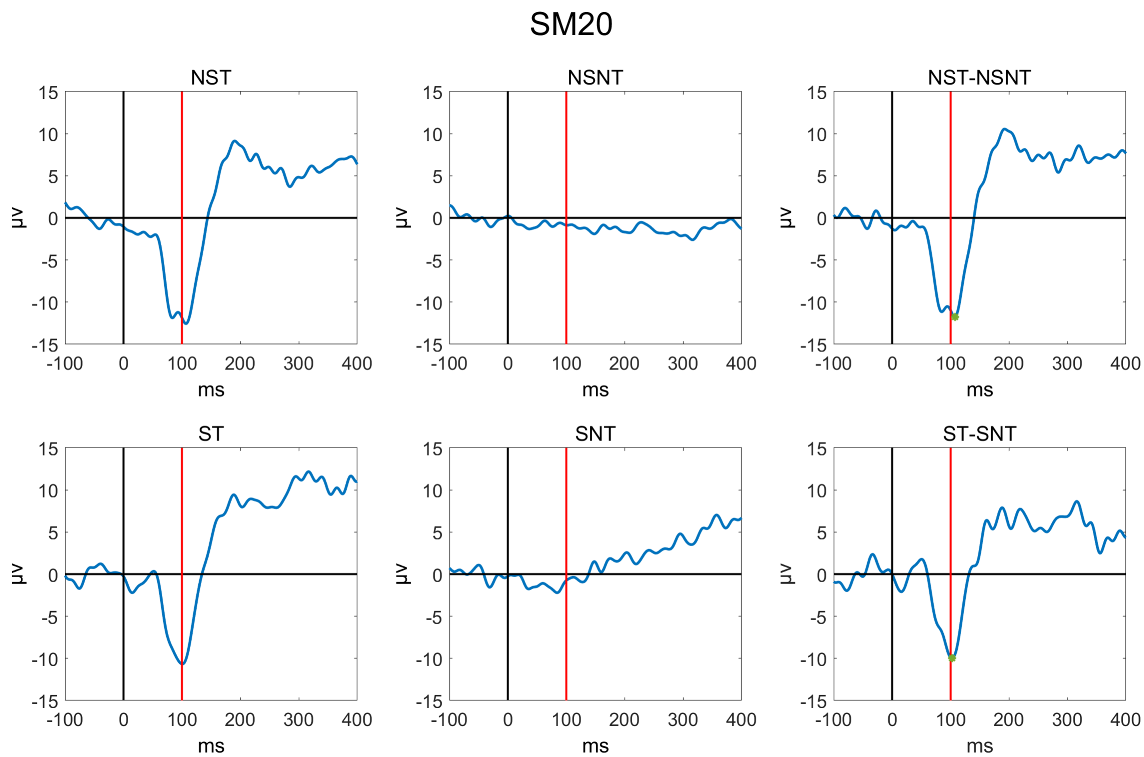

Figure 2.2: Example of adult difference signal and its corresponding N1 (subject SM20). NonSpeaking Tone(NST); NonSpeaking No Tone(NSNT); Speaking Tone(ST); Speaking no Tone(SNT);

Figure 2.2 shows an example of the averaged signals in 4 different conditions and the difference signal. The pattern of the difference signal is with an N1 onset at around 100 ms, and with a P2 onset at around 200 ms. After averaging and rejection, most of the subjects show the onset of N1 and P2 clearly as Figure 2.2. There are two subjects in the stuttering group and two subjects in the nonstuttering group not showing a significant negative peak.

N1

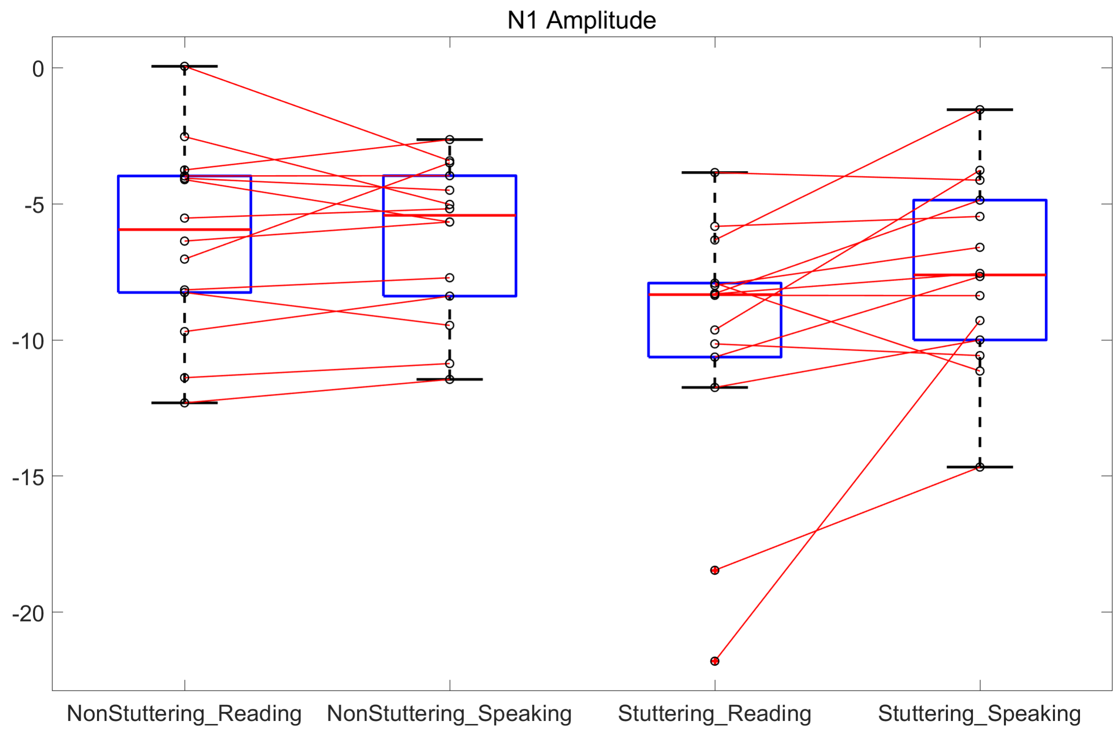

Figure 2.3: N1 amplitude in stuttering group (Right) and nonstuttering group (Left). The red line connects the same subject in different tasks.

Figure 2.3 shows the distribution of N1 amplitude among 2 groups. The previous studies shows that the difference of N1 amplitude between two conditions are consistent in nonstuttering group. The absolute value of N1 amplitude during silent reading condition is higher than it during speaking condition. However, in stuttering groups, the relationship varies. By averaging, for most of the nonstuttering subjects, the N1 amplitude during the reading task is slightly higher than it during the speaking task. While for the stuttering group, the relationship between N1 during the reading task and it during speaking task vary, Which is similar to the previous publication. The difference of the result from the previous paper might caused by the inclusion of data set. During previous study, some of the data are left out due to the improper behavior during the experiment, such as falling asleep during the experiment.

Signal-to-Noise Ratio (SNR)

The average SNR is 30.21. This number is set as the baseline to evaluate the performance of the following methods.

2.3.2 Child

By using averaging and rejection, we can not observe the difference signal in child subject with the same pattern as it in the adult subject. Here is an example (Figure 2.4).

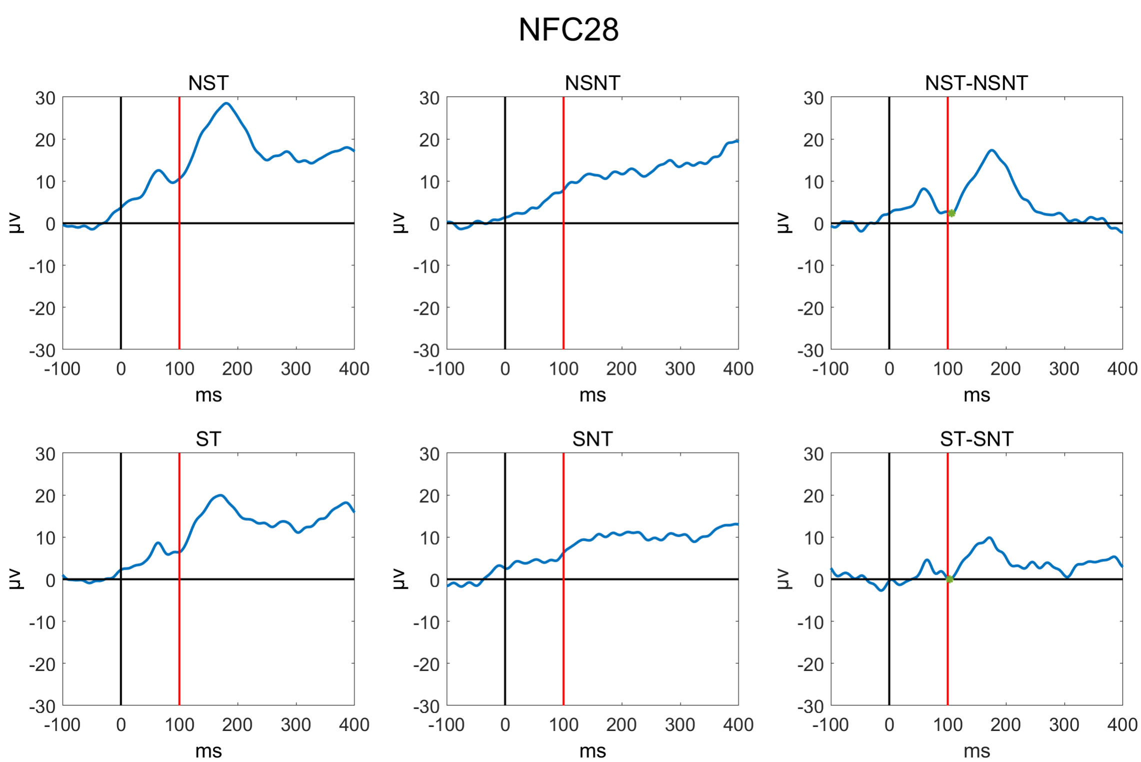

Figure 2.4: Example of child difference signal and its corresponding N1 (subject NFC28)

2.4 Discussion

This method is able to process most of the adult subjects and gets both N1 at 100 ms and P1 at 200 ms. However, on child subjects, it fails to produce the same kind of result. Meanwhile, as shown in Figure 2.5, there are some bad epochs being removed. Some subjects even reject half of the epochs. And the mean of the number of rejection is around 20% for the nonstuttering group and 40% for the stuttering group. Although some of the epochs are rejected, this method still gains a good result, and it mostly preserves the information of the original data.

This result led to seek other methods to solve both problems above, which is a more adaptive method with less trial reduction.

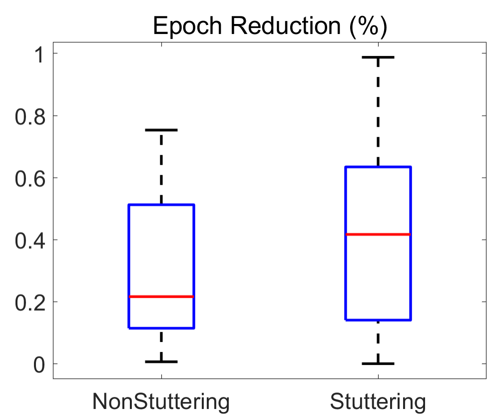

Figure 2.5: The proportion of epochs reduction

Chapter 3

METHOD 2: INDEPENDENT COMPONENT ANALYSIS (ICA)

3.1 Background and Motivation

In order to remove noise and artifacts with a method that aims to identify the individual sources that gave rise to the recorded overall EEG signal, Independent component analysis is introduced. Due to the principle of EEG, the EEG recording is the linear combination of the sources generated by brain activities. Thus, ICA, a linear blind source separation method, fits this concept.

Assume there are n source: $S = (s_1, s_2, \ldots, s_n)^T$, and there are n detector, which detects n signal: $X = (x_1, x_2, \ldots, x_n)^T$. Each x_i is the linear combination of the source: $x_i = a_{i1}s_1 + a_{i2}s_2 + \ldots + a_{in}s_n$.

Thus:

$$\begin{aligned} X &= AS \\ S &= A^{-1}X \end{aligned}$$

If we can find the matrix A, we can reconstruct the source signals with the detected signals. However, with the information we have, we can not find the matrix A by solving the equation. Thus, we need other methods to find the matrix A.

Since the source signals are not related to each other, The goal of the separation is to make the predicted source signal as independent as possible. To achieve this goal, we could promote nongaussianity. By calculating kurtosis or negentropy, we can calculate nongaussianity. Also we could minimize mutual information among the source or use maximum likelihood estimation. We used Infomax ICA in this method, which is based on maximizing entropy by using neural network[10][11].

3.2 Pre-processing

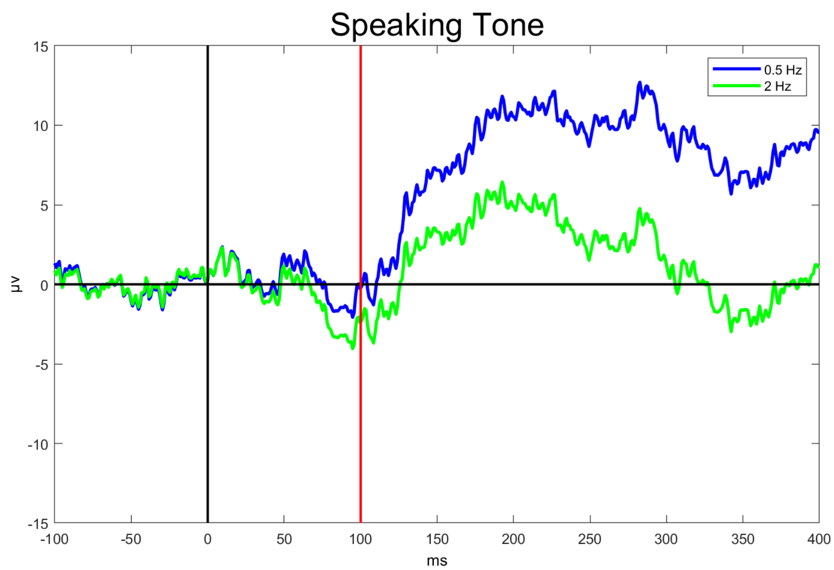

Figure 3.1: Example of difference between using high-pass filter at 0.5 Hz and 2 Hz

ICA is focused on information with frequency at 3 to 15 Hz. Previous work shows that by using a high-pass filter at 1 to 2 Hz can get signal with the highest SNR[12]. However, there still is useful information in low-frequency EEG signals. By using a 1 to 2 Hz high-pass filter can cause the amplitude reduction at the later part of each trial (Figure 3.1).

Since ICA is a linear method, a filter should not change the relationship among each channel. Thus, a band-pass filter with cutoff frequency at 2 Hz and 15Hz was applied first on the original signals. To reduce the running time and learning the features better, the filtered signals were downsampled to 100Hz. Both EEG and EOG signals were feed into the infomax ICA algorithm to obtain the ICA transformation matrix A^{-1}_{2Hz}. Then the ICA transformation matrix A^{-1}_{2Hz} was applied to signals filtered by band-pass filter with cutoff frequency at 0.5 Hz and 15Hz (Figure 3.6).

The other steps are the same as in chapter 2. EOG signals were filtered by band-pass filter with cutoff frequency at 1 Hz and 10 Hz.

3.3 Methods

3.3.1 ICA on epoch Data VS Continuous Data

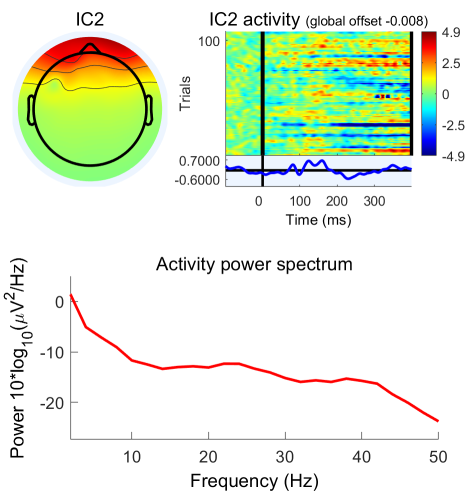

Figure 3.2: Example of component represent eye movement

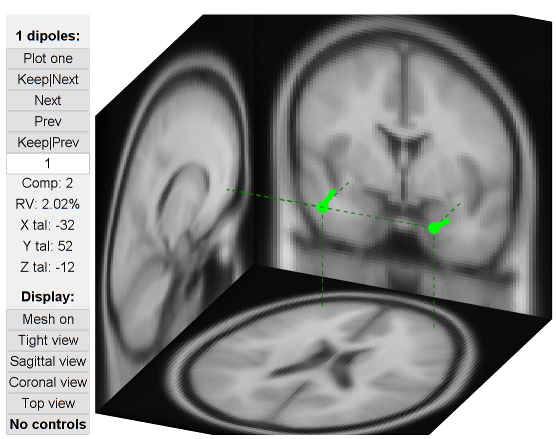

Figure 3.3: Example of dipole location represent eye movement

There are discussions on whether to use epoch data or continuous data for ICA. Epoch data has a smaller data size and more proportion of information related to the event. While continuous data could provide more information in general. Thus both epoch data and continuous data were used, and continuous data were found to be more suitable for this experiment.

3.3.2 Artifact Rejection

In this algorithm, two kinds of artifacts, eye movement, and muscle movement were rejected.

Eye Movement

The eye movement signal usually looks like a neural signal from the component spectra. For both of them, most of their signals lie under 20 Hz. However, eye movement only happens at the front of the head[13]. It's easy to differentiate eye movement from other brain activities by scalp map (Figure 3.2) and component localization (Figure 3.3). Meanwhile, EOG channels were included in the artifact rejection. By examining the changes in EOG channels, the validity of rejection can be reassured.

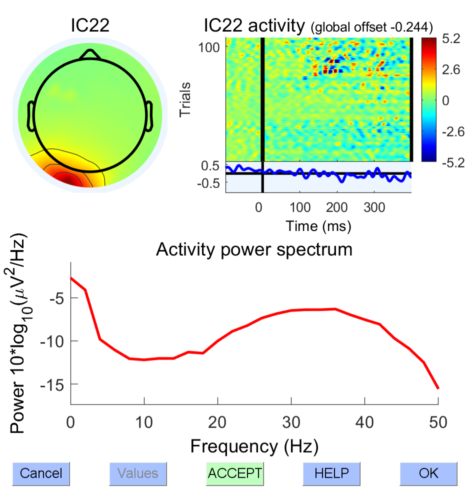

Figure 3.4: Example of component represent muscle movement

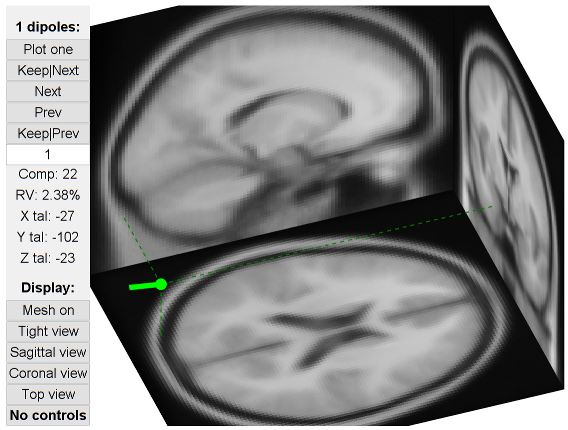

Figure 3.5: Example of dipole location represent muscle movement

Muscle Movement

Muscle movement signal looks like high-frequency noise[13]. By using component spectra, they can be discriminated from other signals (Figure 3.4). Meanwhile, since it's movement generated by muscles, the origin of it is out of the brain (Figure 3.5). Thus, component localization can be another criterion.

3.3.3 Back Projection

ICA originally aims to find the resource from the detected data. However, the category of some components are uncertain, and it's hard to choose which single component could represent the real source signal. Thus, in this algorithm, I decided to first reject those two kinds of components above and back-projected the rest of the components to the original channel space.

Figure 3.6 shows the procedures of this ICA algorithm. The cleaned data was applied to the averaging algorithm in chapter 2. The difference signal and its corresponding N1 are still the important measurements.

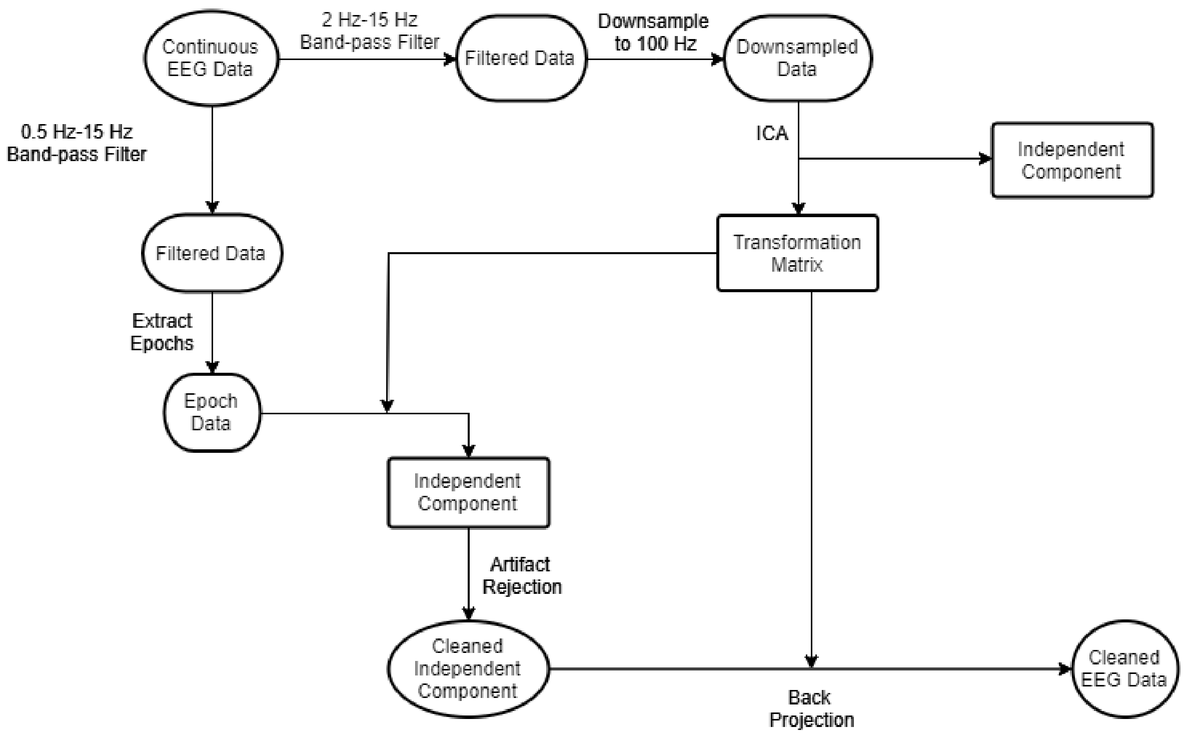

Figure 3.6: Procedures of ICA algorithm

3.4 Results

3.4.1 Adult

Difference Signal: Averaging VS ICA

Figure 3.7 shows an example of the comparison between averaging and ICA. CF61 is one of the subjects that can not be obtained the normal two-peak pattern trajectory by averaging and rejection. For this subject, almost 90% of epochs were rejected by the threshold. By using ICA, the trajectory of the speaking condition was cleaned, and can clearly show both the onset of N1 and P2. For the 4 abnormal data that can not be processed by the averaging algorithm, three of them change into a better shape trajectory as subject CF61 to some extent. Only one of them in stuttering subjects still remains the same.

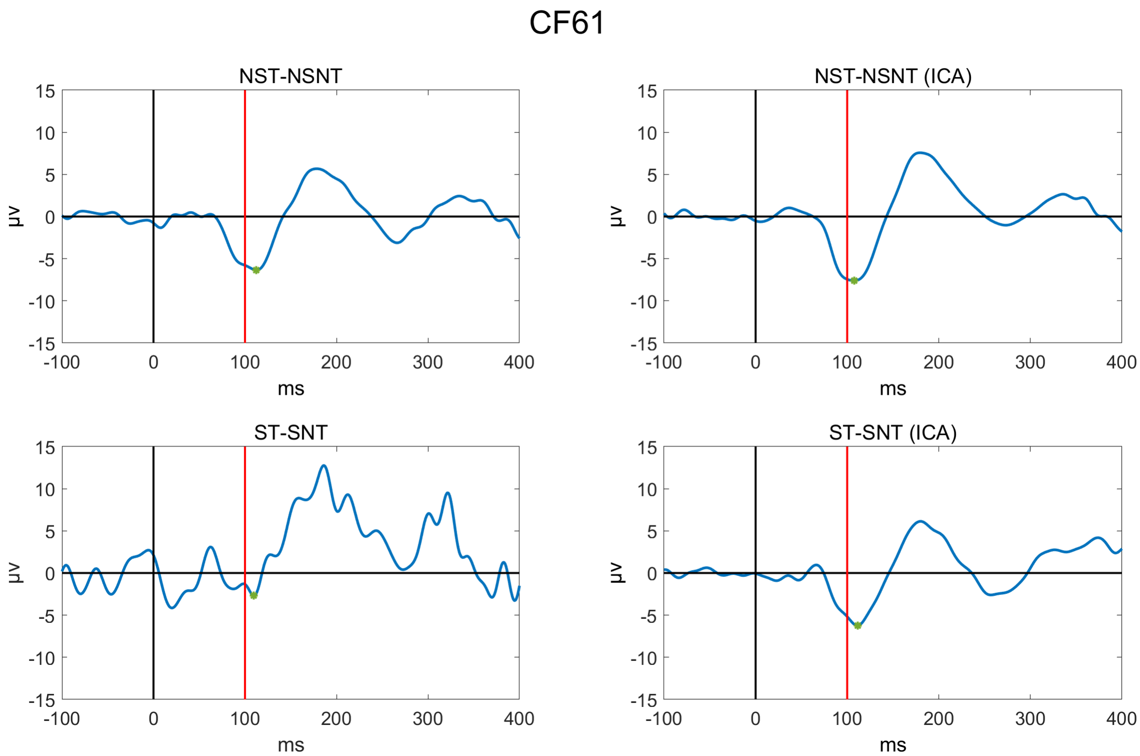

Figure 3.7: Example of the result by averaging and rejection (left), and the result by ICA (right). NST-NSNT: the difference between tone and no tone trial during the silent reading task.ST-SNT: the difference between tone and no tone trial during the speaking task.

N1

Figure 3.8 shows the distribution of N1 amplitude among 2 groups and their 2 conditions. For most of the nonstuttering subjects, the N1 amplitude during the reading task is higher than it during the speaking task. While for the stuttering group, the relationship between N1 during the reading task and it during speaking task vary, Which is similar to the previous publication.

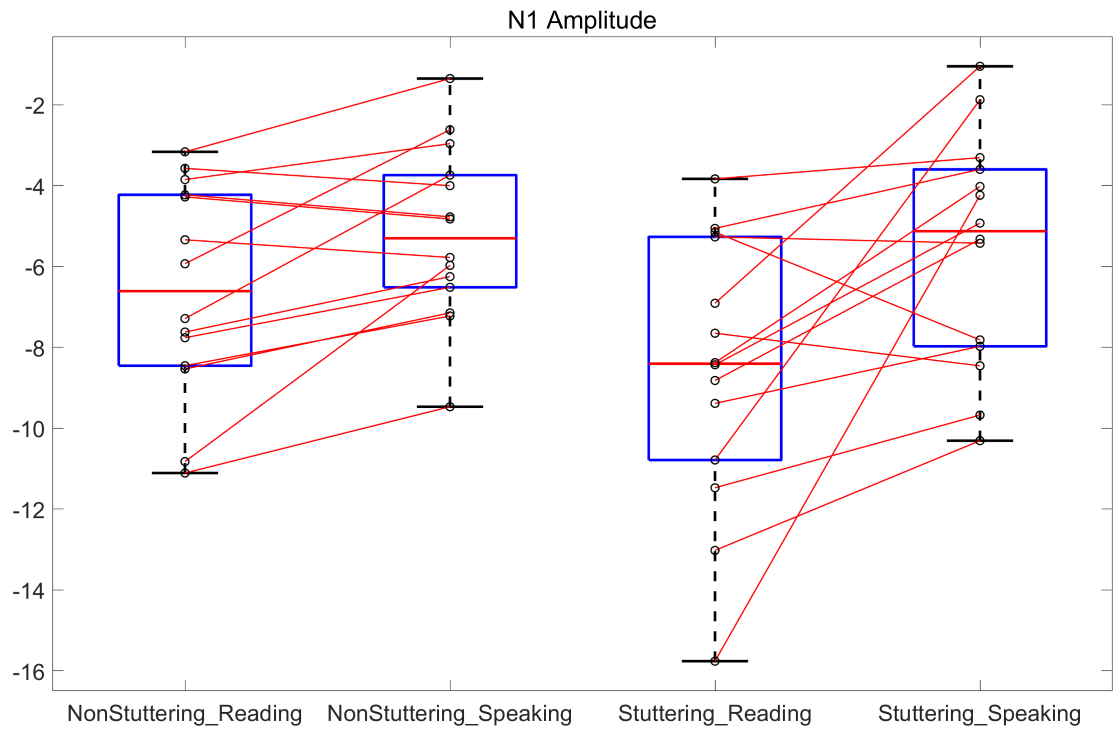

Figure 3.8: (ICA) N1 amplitude in stuttering group (Right) and nonstuttering group (Left). The red line connects the same subject in different tasks.

Signal-to-Noise Ratio (SNR)

The average SNR for this method is 33.02, which is higher than the baseline 30.21 for averaging algorithm.

3.4.2 Child

By using ICA, the result (Figure 3.9) is improved slightly, since speaking groups showing a negative peak of around 100 ms. However, it does not clean all the irrelevant signals in the child EEG data.

For some child subject, such as NFC29 (Figure 3.10), the algorithm successfully removes the peak on set at 50 ms. And both of them show a negative peak on set at 100 ms. It

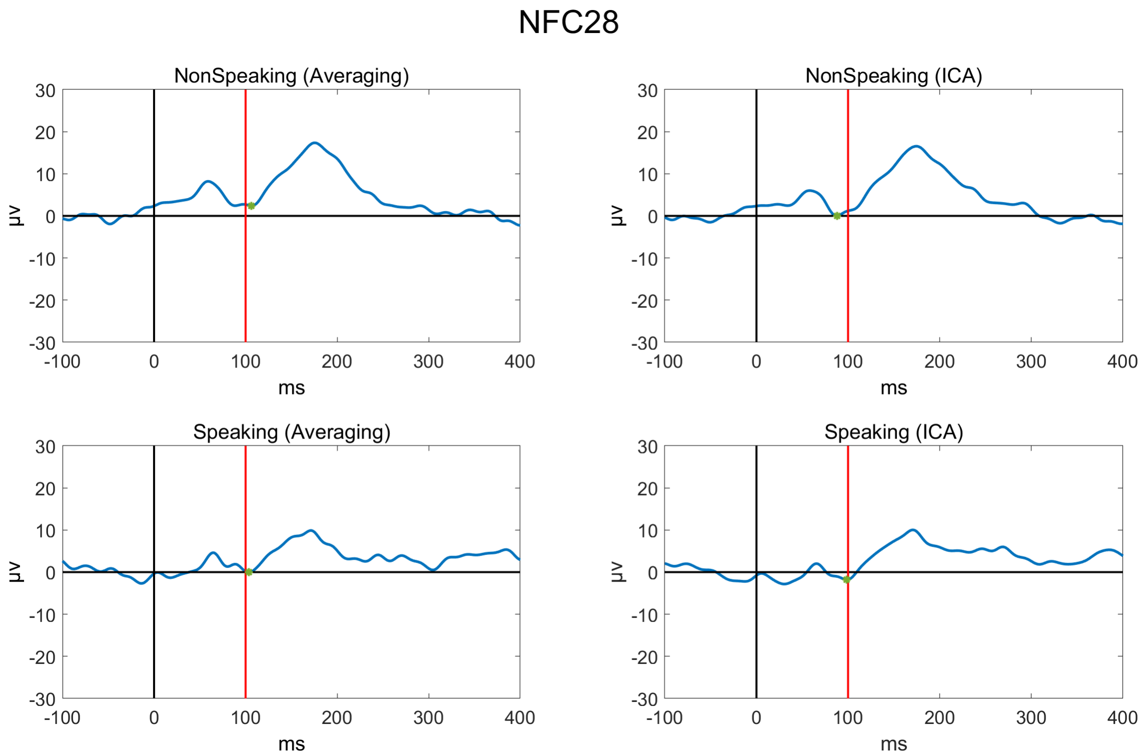

Figure 3.9: Example of child difference signal and its corresponding N1 after applying ICA vs Averaging (subject NFC28)

indicates that ICA can process part of the child data.

3.5 *Discussion*

This method has a better performance than averaging on both adult and child subjects. It cleaned some of the data that averaging can not process, and it gets a more steady result without outlier, than the averaging algorithm. Meanwhile, by the principle of the ICA, there is no epoch reduction, which enables all data to remain the same number as it in the experiment. Meanwhile, ICA increases the SNR compared to the averaging algorithm, which indicates that this method has a better performance on denoising.

However, there are still some pitfalls in this method. For example, one of the subjects

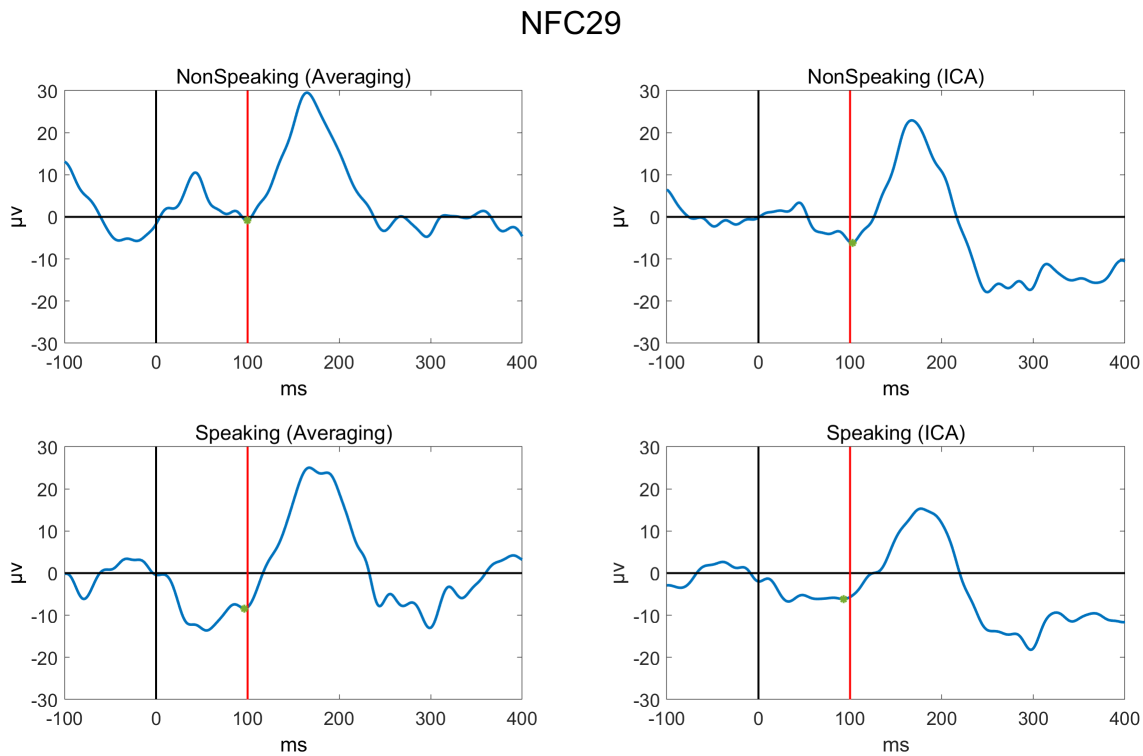

Figure 3.10: Example of child difference signal and its corresponding N1 after applying ICA vs Averaging (subject NFC29)

in the adult group can not be successfully cleaned by ICA, and the improvement in child subjects are only in some subjects. Thus we need a better solution.

The principle of EEG is assuming all the channels that are recorded are the linear combination of the source signals that generated inside the brain. However, if we neglect this principle, and use other methods to find the pattern of each epoch, we might also find useful information. The idea of nonlinear methods led to use Autoencoder as another method to solve this problem.

Chapter 4

METHOD 3: AUTOENCODER

4.1 Background and Motivation

An Autoencoder is a kind of neural network, which aims to learn the low-dimensional representation of itself. The framework of this neural network contains three parts, input, latent features, and output, and they are connected by an encoder and a decoder. It is an unsupervised learning method, and both the input and output are the data itself[14].

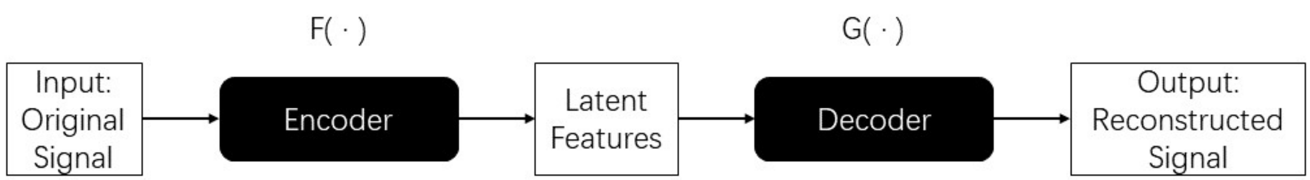

Figure 4.1: The framework of Autoencoder

The encoder ($F(\cdot)$) compresses the signal into low-dimensional features, and the purpose of the decoder ($G(\cdot)$) is trying to reconstruct the latent features into a signal that similar to the original signal. And the reconstructed signal can be written as $G(F(X))$. And the aim of this network is trying to make the difference between the original signal and the reconstructed signal as small as possible, which is $G(F(X)) = X$.

The general uses of Autoencoder include dimension reduction and signal denoising. By adding other constrains on the learning, we can obtain meaningful latent features. In this chapter, the independence among the latent features are promoted by a loss function, and the nonlinear learning structures are achieved by the nonlinear layers.

4.2 Methods

4.2.1 Pre-processing

Data were processed as methods above. EEG signals were filtered by band-pass filter with cutoff frequency at 0.5 Hz and 15Hz. EOG signals were filtered by band-pass filter with cutoff frequency at 1 Hz and 10 Hz. To simplify the process of independent feature learning, all data were normalized. The data was separated into 4 different events the same as the methods above.

4.2.2 Define the number of latent features: Hankel Matrix

Hankel matrix has a form shown as follow:

$$H = \begin{pmatrix} a_0 & a_1 & \cdots & a_{n-1} \\ a_1 & a_2 & \cdots & a_n \\ \vdots & \vdots & \ddots & \vdots \\ a_m & a_{m+1} & \cdots & a_{m+n-1} \end{pmatrix}$$

In the state-space system identification theory, The rank of the Hankel matrix often used as the order of the system[15]. In this methods, a_i represents the voltage of signal at time t_i. By doing singular value decomposition on this Hankel matrix, we can acquire the latent relationship of the signal across times. The first several large singular values can indicate at least how many latent signals controls the observed signals.

The time delay is set as 100 ms in this method. Figure 4.2 shows an example of the singular value of a Hankel matrix. From the figure, we can see after the 15^{th} singular value, they are approximately zero. It indicates that there are at least 15 latent signals can generate these observed signals. As a result, The number of latent features is set as 15, and the results are the same from both child and adult data.

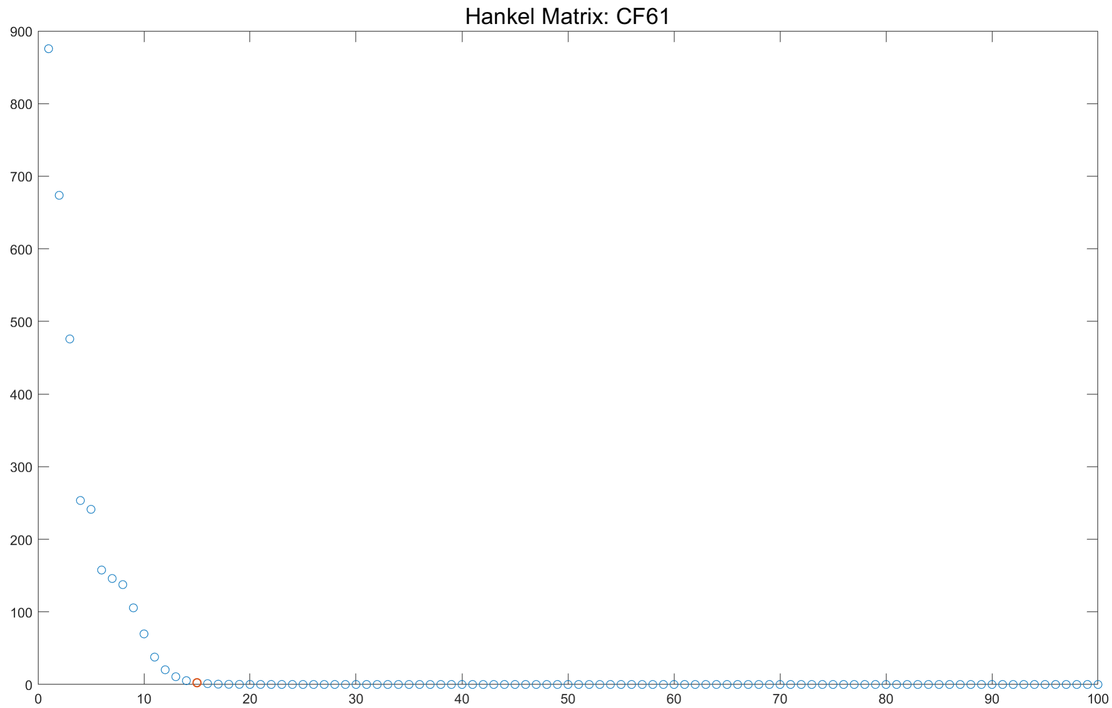

Figure 4.2: The singular value of a Hankel matrix with 100 ms time delay (Example: CF61)

4.2.3 Framework

The encoder of this neural network consist of two hidden layers. The first layer is a linear fully connected layer with 64 neurons. The second layer is a active function, which is the Softplus function with 15 neurons.

Due to the principle of EEG, it assumes that the observed signals are the linear combination of source signals. As a result, The decoder consists of only one linear fully connected layer to simulate the assumption. The Framework shows in Figure 4.3.

As for Child, the framework is the same with the fully connected layer consisting 32 of neurons and 15 latent features.

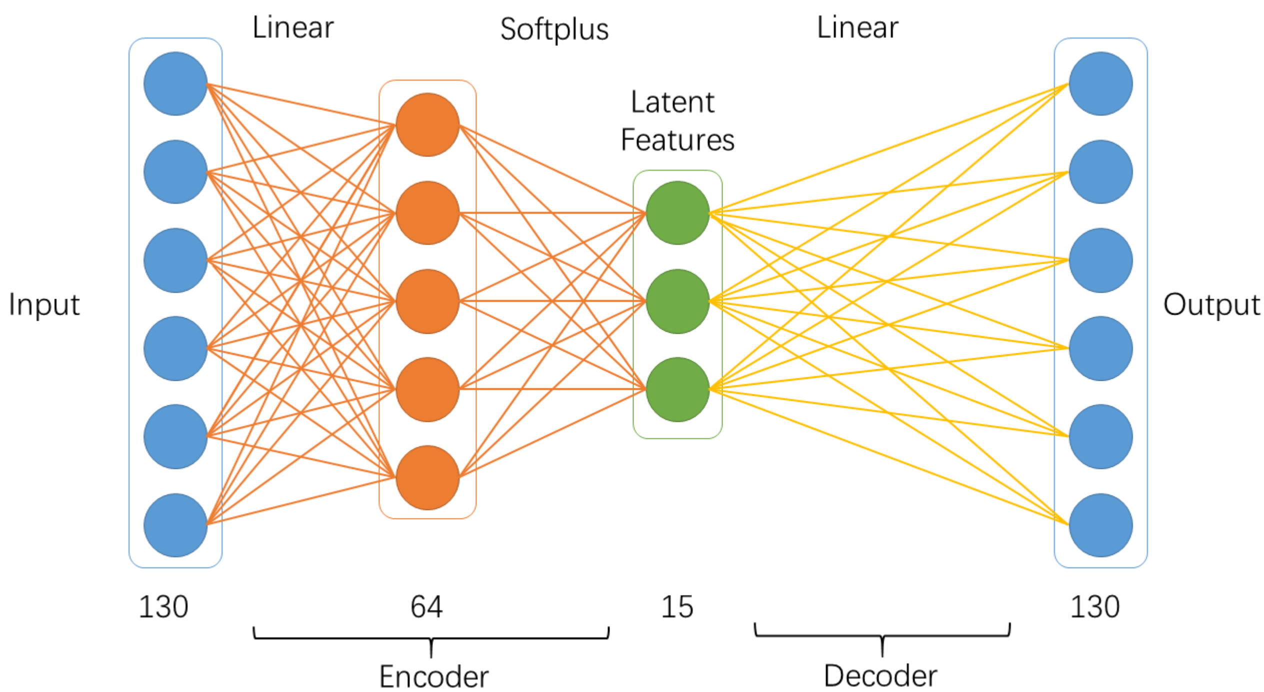

Figure 4.3: The Autoenoder Framework (Adult)

4.2.4 Input and Output

The last two channels of each data are the EOG channels. Signals represent eye movement should dominate those two channels. The linear combination of the latent features should be able to eliminate most of the eye movement signals in those channels. As a result, the output data for validation in this method is different from the input data. The last two channels are replaced by the average of the rest channels as a rough estimation of the brain signals.

4.2.5 Loss Function

To promote independence among features. Two kinds of methods are applied to creat the loss function.

Kurtosis

The low-dimensional features should not only be able to reconstruct the original signal, but it also should be independent. In order to promote independence, kurtosis was used in the loss function.

Kurtosis a classic way to measure the nongaussianity of the data[11]. It's defined by

$$Kurt(y) = E\{y^4\} - 3E\{y^2\}^2$$

If the data is independent to each other, the following equation holds.

$$Kurt(y_1 + y_2) = Kurt(y_1) + Kurt(y_2)$$

As a result, the kurtosis is used to promote independence among the latent features. The loss function uses the least square error to make the difference between the original signal and the reconstructed signal as small as possible. Here is the loss function.

$$Loss = \lambda_1 \|Y_{original} - Y_{predict}\| + \lambda_2 \sum_{i,j,i \neq j}^{n} |Kurt(F_i + F_j) - (Kurt(F_i) + Kurt(F_j))|$$

Where $Y_{original}$ is the original signal, $Y_{predict}$ is the predicted signal by the network, and F_i is the i^{th} feature.

Mutual Information

There is another method that can promote the independence among latent features, mutual information between n variables, $y_1, y_2, \cdots, y_n$, which is defined by:

$$I(y_1, y_2, \cdots, y_n) = \sum_{i=1}^{n} H(y_i) - H(y_{gaussian})$$

Where H(y) is the Entropy of y, and $y_{gaussian}$ is a random variable with Gaussian distribution of the same covariance matrix as y. It can also write in this form.

$$I(y_1, y_2, \cdots, y_n) = C - \sum_{i=1}^{n} J(y_i)$$

Where C is a constant, and J(y) is negentropy of y. However J(y) is very hard to be calculated. As a result, it can be approximated by

$$J(y) \approx \frac{1}{12}E\{y^3\}^2 + \frac{1}{48}Kurt(y)^2$$

By minimize the mutual information among the features, we can obtain the independent features. Here is the loss function:

$$Loss = \lambda_1 \|Y_{original} - Y_{predict}\| + \lambda_2 \setminus \sum_i^n |\frac{1}{12}E\{F_i^3\}^2 + \frac{1}{48}Kurt(F_i)^2|$$

Where $Y_{original}$ is the original signal, $Y_{predict}$ is the predicted signal by the network, and F_i is the i^{th} feature.

Figure 4.4: The difference signal: original signal vs predict signal (Example: CF60)

4.2.6 Code

The Autoencoder build based on TensorFlow framwork[16]. 5 fold cross-validation was used to validate the model. The optimizer used in this training is Adam optimizer[17].

4.3 Results

4.3.1 Adult

Kurtosis

By using kurtosis in the loss function, the latent features tend to be more Gaussian or sub-Gaussian rather than independent. It causes the predicted signal to be sub-Gaussian as well. As shown in Figure 4.4, the predicted signal contains less information than the original signal. As a result, only mutual information was used in the loss function.

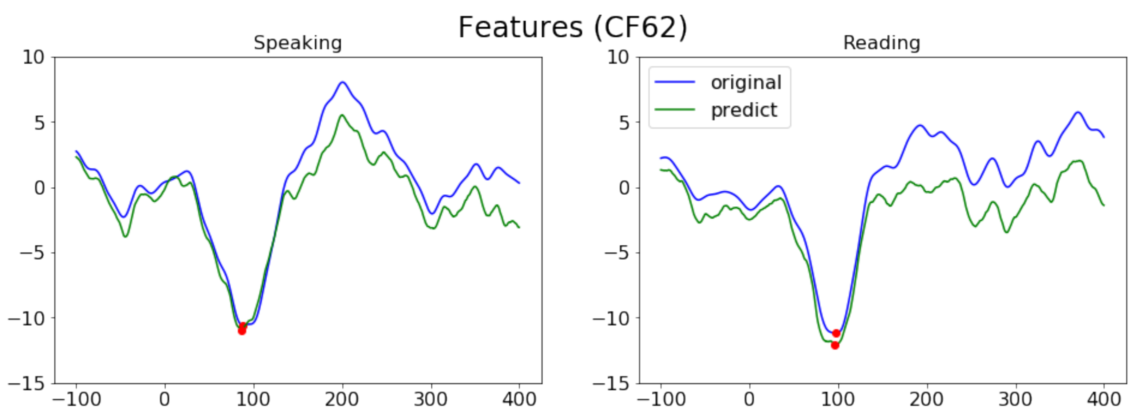

Figure 4.5: The difference signal: original signal vs predict signal (Example: CF62)

Validation

The average difference between the original signal and the predicted signal in the training set is 14.8. And the average difference between the original signal and the predicted signal in the testing set is 23.1. Both of them are similar to each other, which means that the Autoencoder is adaptive.

Difference Signal

Figure 4.5 shows an example of the difference signal. The predicted signal has the similar trajectory as the original signal with some changes as the assumption.

N1

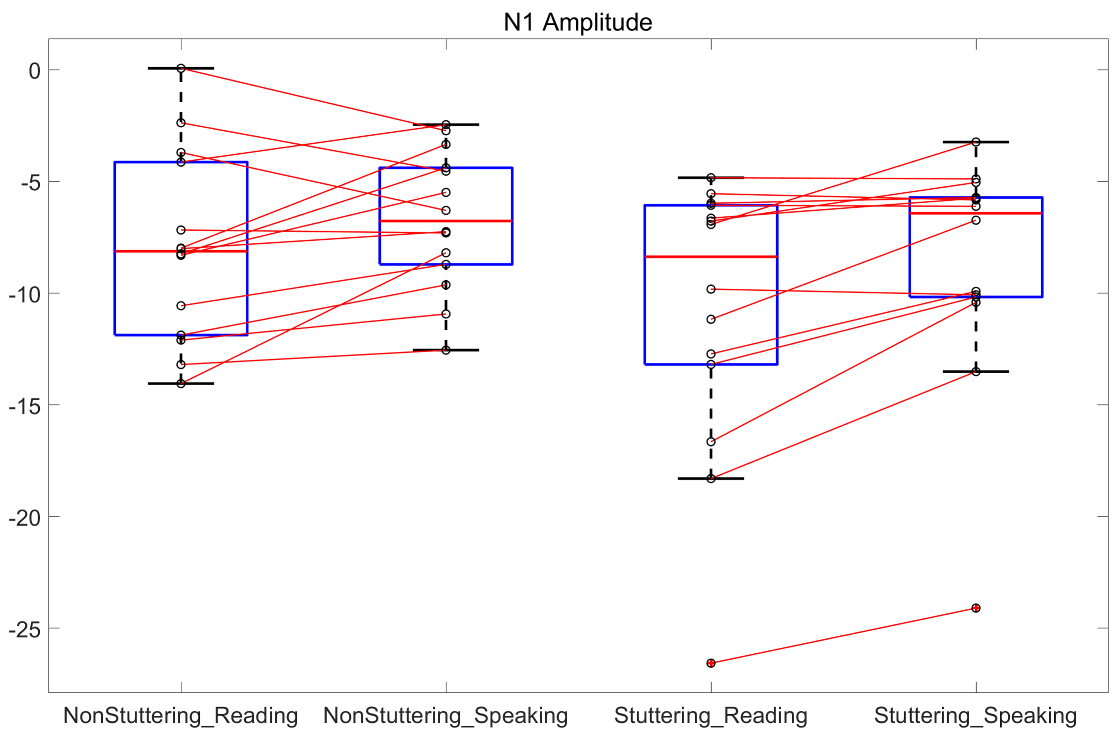

Figure 4.6: (Autoencoder) N1 amplitude in stuttering group (Right) and nonstuttering group (Left). The red line connects the same subject in different tasks.

Figure 4.6 shows the distribution of N1 amplitude among 2 groups and their 2 conditions. For both nonstuttering group and stuttering group, most of the subjects have a higher the

N1 amplitude during the reading task than it during the speaking task, which is not similar to the previous publication.

SNR

The average SNR for this method is 22.94, which lower than the baseline 30.21 for the averaging algorithm. By comparing three different methods (Figure 4.7), There is a significant difference between Averaging and Autoencoder (p=0.0018), a significant difference between ICA and Autoencoder (p=1.4×10^{-4}), and no significant difference between ICA and Averaging (p=0.24).

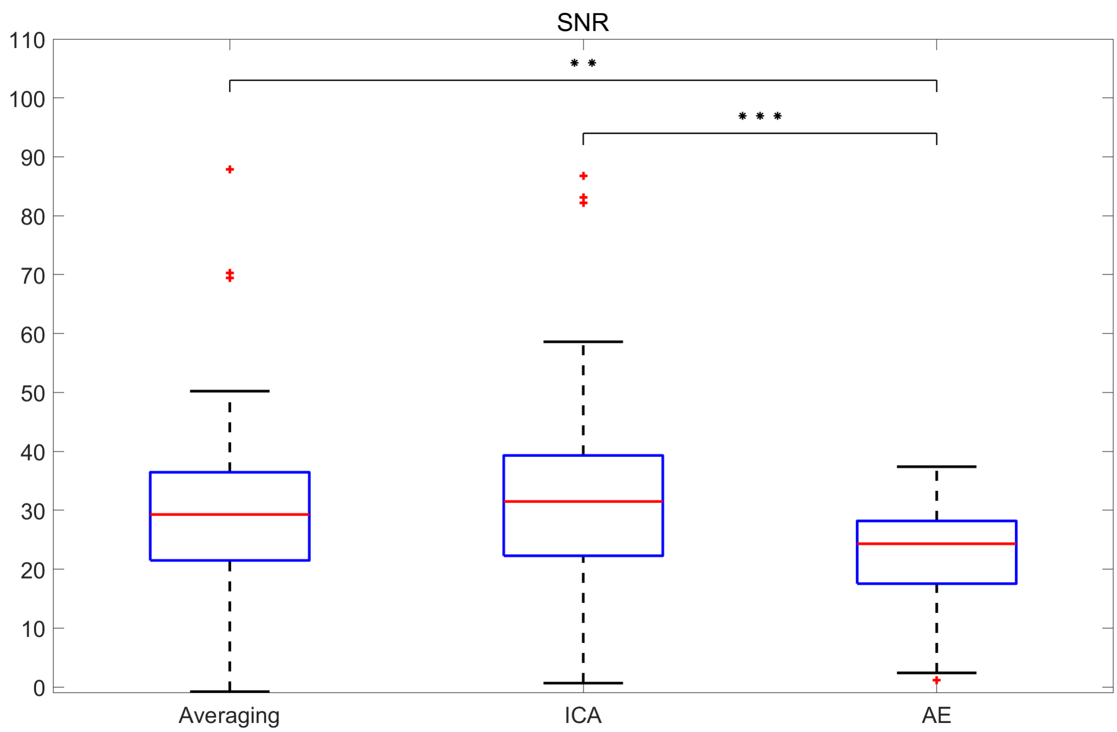

Figure 4.7: The distribution of SNR among different methods

4.3.2 Child

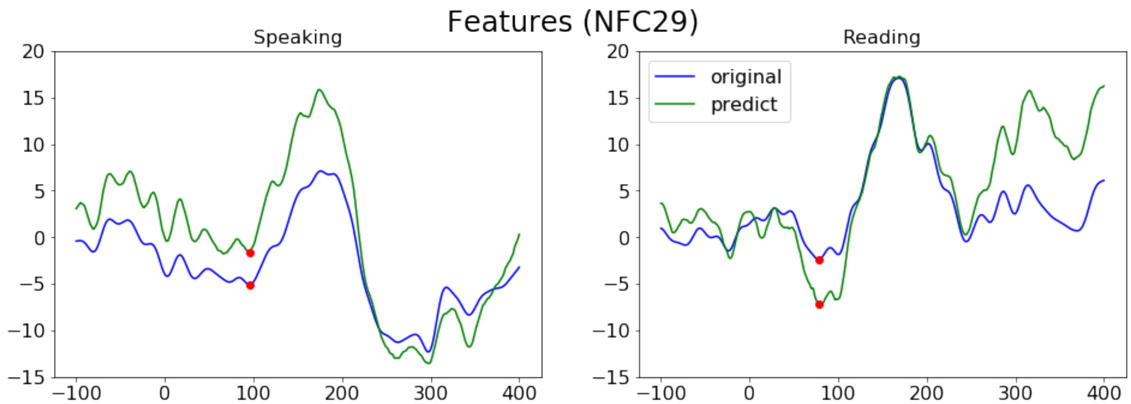

Figure 4.8: Example of child difference signal original VS Predicted (subject NFC29)

In Figure 4.8, after learning the latent features by Autoencoder, there is a negative peak onset during the reading task at around 100 ms, which does not exist in the original signal. While there is no improvement during the speaking task. Figure 4.9 shows the difference of subject NFC29 after processed by three different methods.

4.4 Discussion

Although the N1 amplitude is not similar to the publication, which indicates some pitfalls in the Autoencoder algorithm, Autoencoder is the most effective method that can learn features in child data. It successfully shows N1 in some of the child data, which can not be obtained by the other two methods. The SNR is lower than the averaging means the algorithm generates noise during the learning. Furthermore, this method only learns the latent features without any rejection. There could be artifacts among those latent features.

As a result, the algorithm needs to be improved to achieve three aims: 1) increasing SNR 2) N1 amplitude similar to the previous publication 3) adapting to all child data. Meanwhile, feature identification methods, such as localization, need to be developed to apply to this algorithm to reject artifacts.

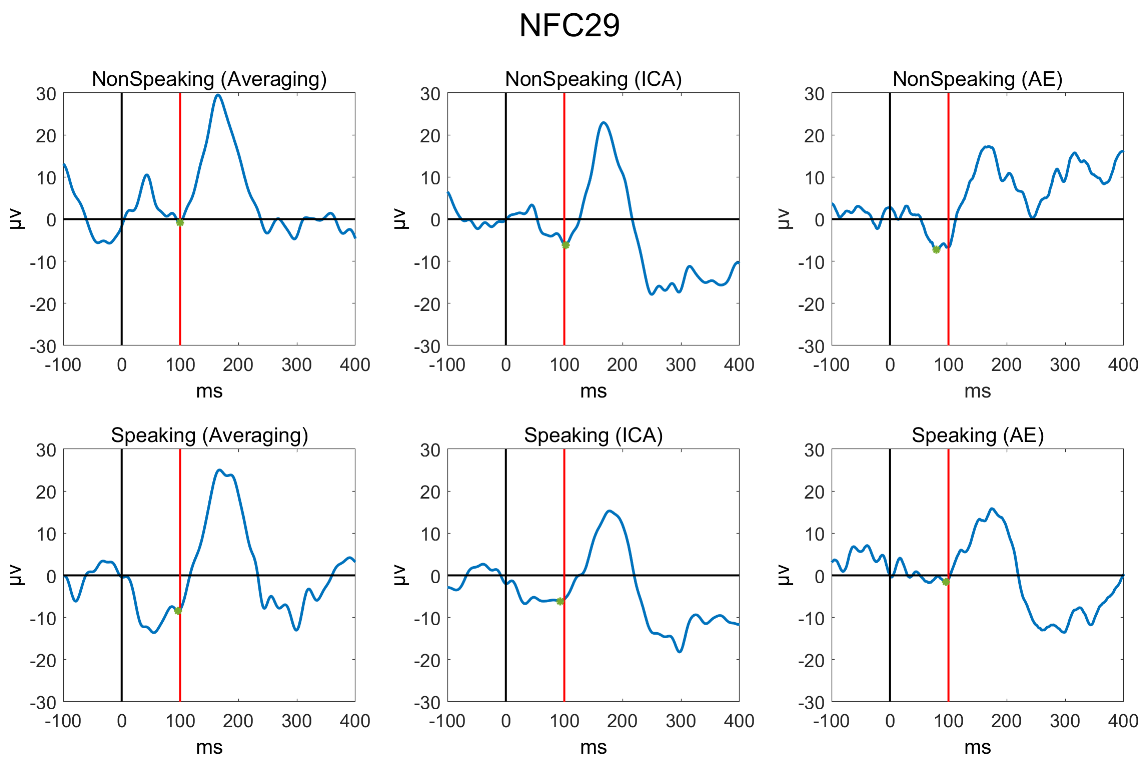

Figure 4.9: Example of child difference signal Averaging VS ICA VS Autoencoder (subject NFC29)

Chapter 5
CONCLUSION AND FUTURE WORK

5.1 Conclusion

In conclusion, this thesis compared three different methods, averaging, ICA, and Autoencoder, that could be used on EEG data processing and artifacts rejection. All three methods have their own advantages and pitfalls and can be chosen based on the data set.

Two advantages of averaging are preserving the original information of the data most and eliminating non-activity-related Gaussian noise. However, there are two pitfalls. It reduces the number of epoch in each group and fails to remove the irrelevant activity-related signals. ICA could solve part of both of the problems, and obtains a cleaner data with higher SNR than averaging. It successfully process some of the child data. To add nonlinearity to the algorithm, Autoencoder was introduced. It can generate low-dimensional independent features and combines both linearity and nonlinearity in the algorithm. And it can also process some of the child data. One pitfall is that currently there is no localization method to validate the trained features.

5.2 Future Work

For now, we only have 4 child data, and the stuttering female child is still missing. Ideally, We need to recruit more child subjects and paired them as the adult group to get a statistical result. The algorithm still needs to have further improvement to be adapted to child subjects.

By changing the structure of the Autoencoder and adding different constrains, we can develop a better algorithm. Besides, a localization algorithm can be developed for the Autoencoder method to better validate the latent features. Since the decoder is a linear map, which is the same as ICA, the localization algorithm should be similar to it for the ICA.

With the localization algorithm, it's easier to determine whether the further rejection of latent features is needed.

BIBLIOGRAPHY

[1] Ayoub Daliri and Ludo Max. Modulation of auditory processing during speech movement planning is limited in adults who stutter. *Brain and Language*, 143:59 – 68, 2015.

[2] Michael P. Boyle. Relationships between psychosocial factors and quality of life for adults who stutter. *American Journal of Speech-Language Pathology*, 24(1):1–12, 2015.

[3] O. Bloodstein and N.B. Ratner. *A Handbook on Stuttering*. Thomson Delmar Learning, 2008.

[4] C. Stromsta. *Elements of stuttering*. Atsmorts Publishing, 1986.

[5] Anna Craig-McQuaide, Harith Akram, Ludvic Zrinzo, and Elina Tripoliti. A review of brain circuitries involved in stuttering. *Frontiers in Human Neuroscience*, 8:884, 2014.

[6] Ayoub Daliri and Ludo Max. Stuttering adults' lack of pre-speech auditory modulation normalizes when speaking with delayed auditory feedback. *Cortex*, 99:55 – 68, 2018.

[7] S.J. Luck. *An Introduction to the Event-Related Potential Technique*. A Bradford Book. MIT Press, 2014.

[8] Arnaud Delorme and Scott Makeig. Eeglab: an open source toolbox for analysis of single-trial eeg dynamics including independent component analysis. *Journal of Neuroscience Methods*, 134(1):9 – 21, 2004.

[9] Pamela Bäß, Thomas Jacobsen, and Erich Schröger. Suppression of the auditory n1 event-related potential component with unpredictable self-initiated tones: Evidence for internal forward models with dynamic stimulation. *International Journal of Psychophysiology*, 70(2):137 – 143, 2008.

[10] A. Hyvärinen, J. Karhunen, and E. Oja. *Independent Component Analysis*. Adaptive and Cognitive Dynamic Systems: Signal Processing, Learning, Communications and Control. Wiley, 2004.

[11] A. Hyvärinen and E. Oja. Independent component analysis: algorithms and applications. *Neural Networks*, 13(4):411 – 430, 2000.

[12] Irene Dowding, Stefan Debener, Klaus-Robert Müller, and Michael Tangermann. On the influence of high-pass filtering on ica-based artifact reduction in eeg-erp. 08 2015.

[13] Tzyy-Ping Jung, Colin Humphries, Te-Won Lee, Scott Makeig, Martin J. McKeown, Vicente Iragui, and Terrence J Sejnowski. Extended ica removes artifacts from electroencephalographic recordings. In M. I. Jordan, M. J. Kearns, and S. A. Solla, editors, *Advances in Neural Information Processing Systems 10*, pages 894–900. MIT Press, 1998.

[14] Sebastian J. Wetzel. Unsupervised learning of phase transitions: From principal component analysis to variational autoencoders. *Phys. Rev. E*, 96:022140, Aug 2017.

[15] W. Recursive form of the eigensystem realization algorithm for system identification.

[16] Martín Abadi, Ashish Agarwal, Paul Barham, Eugene Brevdo, Zhifeng Chen, Craig Citro, Greg S. Corrado, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Ian Goodfellow, Andrew Harp, Geoffrey Irving, Michael Isard, Yangqing Jia, Rafal Jozefowicz, Lukasz Kaiser, Manjunath Kudlur, Josh Levenberg, Dan Mane, Rajat Monga, Sherry Moore, Derek Murray, Chris Olah, Mike Schuster, Jonathon Shlens, Benoit Steiner, Ilya Sutskever, Kunal Talwar, Paul Tucker, Vincent Vanhoucke, Vijay Vasudevan, Fernanda Viegas, Oriol Vinyals, Pete Warden, Martin Wattenberg, Martin Wicke, Yuan Yu, and Xiaoqiang Zheng. Tensorflow: Large-scale machine learning on heterogeneous distributed systems, 2016.

[17] Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization, 2014.