Signals and Systems
Mekelweg 4,
2628 CD Delft
The Netherlands
https://sps.ewi.tudelft.nl/

SPS-2023-08

M.Sc. Thesis

Using Tensor Decompositions To Obtain Biomarkers From Auditory Event-Related Potentials

Kenneth Stunnenberg B.Sc.

Faculty of Electrical Engineering, Mathematics and Computer Science **Delft University of Technology**

Using Tensor Decompositions To Obtain Biomarkers From Auditory Event-Related Potentials

Thesis

submitted in partial fulfillment of the
requirements for the degree of

Master of Science

in

Electrical Engineering

by

Kenneth Stunnenberg B.Sc.
born in Amsterdam, The Netherlands

This work was performed in:

Signals and Systems Group
Department of Microelectronics
Faculty of Electrical Engineering, Mathematics and Computer Science
Delft University of Technology

Delft University of Technology

Delft University of Technology
Department of
Microelectronics

The undersigned hereby certify that they have read and recommend to the Faculty of Electrical Engineering, Mathematics and Computer Science for acceptance a thesis entitled **"Using Tensor Decompositions To Obtain Biomarkers From Auditory Event-Related Potentials"** by **Kenneth Stunnenberg B.Sc.** in partial fulfillment of the requirements for the degree of **Master of Science**.

Dated: 31-08-2023

Chairman: _____
Dr. B. Hunyadi

Advisor: _____
Dr.ir. R.C. Hendriks

Committee Members: _____
Dr.ir. K. Batselier

_____
Ir. M.L. Adank

_____
Dr.ir. J. Vroegop

Abstract

Brain disorders in children pose significant challenges to their development, impacting cognition, speech, movement, and behavior. The uncertainty surrounding prognostic information at the time of diagnosis leaves families with numerous questions about the future. The Child Brain Lab at Erasmus MC Sophia Children's Hospital conducts IQ, electroencephalogram (EEG), speech, and movement tests in playful environments, enhancing scientific research and healthcare practices for a better understanding of disease progression.

The Otolaryngology department at the Child Brain Lab focuses on auditory-related potentials (ERPs) obtained from EEG measurements to predict the future development of children with brain disorders. Analyzing ERP data from experiments like Mismatch Negativity (MMN) and Acoustic Change Complex (ACC) yields insights into developmental trajectories and connections between hearing, language, and brain development.

This thesis aims to explore alternative methodologies for extracting comprehensive information from ERPs, overcoming limitations of the commonly used peak amplitude and latency analysis. Tensor decompositions are employed to exploit structural information present in the data, using data fusion methods to combine multiple datasets for improved classification and deeper insights into group differences.

Simulations on artificial ERP data demonstrate that data fusion methods perform better on two ERP tensors compared to single tensor decomposition when group differences are shared between datasets. On a real dataset, tensor decompositions show promise for classifying subjects based on auditory event-related potentials while giving more insights into the neurological sources.

This report proposes an alternative method for analyzing ERP data, highlighting the potential of tensor decompositions and data fusion techniques.

Acknowledgments

I would like to start by expressing my deepest appreciation and give a big thanks to my mentors, Richard Hendriks and Borbala Hunyadi, for their invaluable guidance and interesting discussions. Additionally, I am honored to have collaborated with Jantien Vroegop and Marloes Adank, whose unwavering support, enthusiasm, and expertise in EEG and ERP signals have proven instrumental in my journey. The past year has exceeded my expectations in terms of personal growth and intellectual challenges. I am really thankful for the chance to contribute to this research field in collaboration with all four of you.

I also want to thank my family, for their help and support during challenging times. A special thanks to my girlfriend Lexi for helping me throughout the whole journey and being there for me in stressful moments. And to all my fellow students from the Signals Processing Systems group, it was a fun two years taking courses together, working on projects, and most importantly doing fun stuff besides studying. Without you guys, my Master here would not have been such a great time. Finally, I am also grateful to all my friends for being part of this successful and enjoyable journey.

Kenneth Stunnenberg B.Sc.
Delft, The Netherlands
31-08-2023

Contents

List of Figures

List of Tables

1 Introduction

Brain disorders in children pose significant challenges to their development and can affect various aspects such as cognition, speech, movement, and behavior [1]. Many brain disorders, particularly rare ones, lack clear prognostic information at the time of diagnosis. This uncertainty leaves parents with numerous questions about the future, such as "Will my child ever walk?" or "Can my child live independently?". The Child Brain Lab testing facility in Erasmus MC Sophia Children's Hospital aims to contribute to this research field. By conducting IQ, electroencephalogram (EEG), speech, and movement tests in rooms designed as playful environments, valuable data on these disorders is obtained. This multidimensional approach allows Erasmus MC to collect more information about brain disorders in children, which in turn facilitates scientific research and enhances healthcare practices. Through this systematic collection of data, a clearer understanding of disease progression and the consequences of different brain disorders can be obtained. Consequently, they hope to provide more accurate answers to the aforementioned questions, alleviating uncertainties and offering guidance to affected families.

Among the research groups that utilize the data from the Child Brain Lab, the otolaryngology department focuses on auditory event-related potentials (ERPs) obtained from EEG measurements. Specifically, their research aims to predict the future development of children based on these ERPs. Within a 20-minute time frame, two experiments are conducted to measure the ERPs. These two experiments are the Mismatch Negativity (MMN) [2] and the Acoustic Change Complex (ACC) [3] [4]. The MMN can be excited by occasional deviant stimuli embedded in a train of frequently presented standard stimuli. The ACC can be obtained by a stimulus that contains multiple time-varying acoustic changes. This will result in multiple and often overlapping ERP responses. By focusing on these two experiments and analyzing the corresponding ERPs, the otolaryngology department aims to unravel vital insights into the developmental trajectories of children with brain disorders. This research holds promise for predicting future outcomes and understanding the intricate connections between hearing, language, and brain development.

1.1 Problem Statement

The analysis of EEG data obtained from children with altered neurological development holds great potential in identifying biomarkers that can distinguish different groups based on their neurological profiles [5]. In particular, the MMN and ACC paradigms provide valuable insights into neurodevelopmental variations. However, the conventional approach of examining peak amplitudes and latencies of event-related potentials as biomarkers restricts the utilization of the complete information stored within ERPs.

The primary objective of this thesis is to find a method that can identify potential biomarkers capable of distinguishing between different groups based on their neurological profiles. While the examination of peak amplitudes and latencies of ERPs is a common clinical practice, it inherently limits the exploitation of the comprehensive information contained within ERPs. Consequently, the central challenge lies in exploring alternative methodologies to extract additional information from ERPs, enabling the discovery of more accurate and robust biomarkers. For example, the multichannel data can be exploited to do more with spatial information present in the ERP measurements. Next to that, the ERP data from both experiments can be used together in the form of data fusion. This thesis aims to answer the following questions:

1. How can alternative methods be employed to extract more comprehensive information from ERPs beyond peak amplitudes and latencies?

2. What novel techniques can facilitate the identification of more accurate and informative biomarkers for distinguishing between groups with different neurological development?

3. How can these advanced methods be effectively applied to the EEG data collected within the research environment of Erasmus MC, ensuring their practicality and reproducibility?

By addressing these research questions, this study aims to contribute to the development of an innovative approach for extracting comprehensive information from ERPs, ultimately leading to enhanced differentiation of neurodevelopmental profiles in children with altered neurological development.

1.2 Outline

In order to fully comprehend the methods and terminology used in this thesis, a better understanding of event-related potentials is needed. In the upcoming section, an introduction about ERPs and the two paradigms used to record the EEG data is given. The section after that will give a detailed explanation of the measurement steps and is followed by a literature study of some of the most common methods to analyze ERP data. This is useful for determining a technique that can be used on the dataset from Erasmus MC. Chapter 2 will contain a detailed explanation of tensor decompositions and data fusion methods using tensors. By employing data fusion techniques with tensors, the integration of information from diverse sources becomes possible, leading to a more comprehensive and accurate understanding of the underlying phenomena. Chapter 3 and 4 will contain simulations to see how the different methods perform. This will be done on generated simulation data and a dataset of real ERP measurements from another research investigating differences between schizophrenia patients and healthy subjects [6].

1.3 Event Related Potentials

ERPs, or event-related potentials, refer to the measured brain response that is the result of a specific stimulus (sensory, cognitive, or motor event) [7]. These responses are often extracted from the EEG signal through a process of signal averaging. By combining multiple trials of the EEG signal, background noise can be effectively eliminated. The analysis of a single brain response (trial) to a stimulus, in terms of signal-to-noise ratio (SNR), is often hindered by substantial noise, rendering proper analysis difficult [8]. Trials are small timeframes of the measured continuous EEG signal with a time-lock on the stimulus onset. Usually, a trial ranges from around -100 to 1000 milliseconds but this can vary for different experiments and research questions. ERPs can be characterized by their amplitude, latency, and waveform and the different peaks have specific names. Notably, the P1 and N1 components are frequently observed in ERPs, representing the early stages of sensory input processing [7]. Another example is the P300 component, which is often associated with attention and memory processes.

The analysis of ERPs is widely employed in neuroscience research, enabling insights into various cognitive processes such as language, memory, and attention. This approach offers a non-invasive means of investigation and provides high temporal resolution, facilitating the examination of neural responses over brief time intervals.

The focus of this thesis pertains specifically to auditory event-related potentials [9]. A commonly studied auditory ERP, known as the P1-N1-P2 complex, is derived by averaging multiple brain responses to auditory stimuli. This type of ERP starts with a P1 component and is immediately followed by an N1 component, and ended with a P2 component. These different components are the positive or negative peaks that together form the ERP. The P1-N1-P2 complex reflects the processing of auditory information across various levels, ranging from physical features such as frequency and intensity to higher-level features like phonetic or semantic attributes.

1.3.1 ACC

The approach described by Ostroff et al. [3] and Martin et al. [4] tries to evoke multiple, and often overlapping, P1-N1-P2 responses, collectively referred to as the Acoustic Change Complex (ACC). Figure 1.1 provides an illustration of an ACC response, wherein the stimulus is the phoneme /ui/. Notably, two distinct P1-N1-P2 responses can be seen. The second one is elicited by the acoustic change from /u/ to /i/ at the midpoint of the stimulus and is the classic ACC example. Acoustic changes can be pitch, loudness, or frequency. The source of the ACC waveform is located in the auditory cortex.

The ACC provides valuable insights into the brain's capacity to discriminate and detect changes in acoustic features present in speech signals. By analyzing the ACC waveform, researchers can gain an understanding of how the brain perceives and processes auditory information. This can contribute to the knowledge of speech perception and cognitive processes related to auditory discrimination. There are also a few advantages to using this technique, which are:

- Can be obtained in the absence of attention.

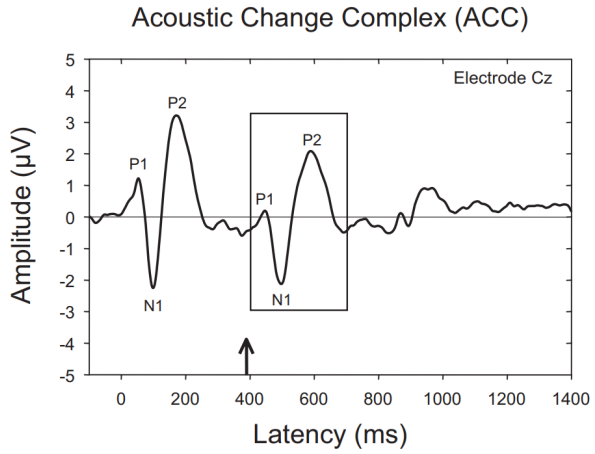


Figure 1.1: ACC response to phoneme /ui/ [7]

- Requires relatively few stimuli to get a recording with good SNR.
- High in amplitude which contributes to a better SNR.
- Is very consistent in individual subjects on a test-to-test experiment.

1.3.2 MMN

The Mismatch Negativity (MMN) is a neurophysiological response that can be elicited by presenting occasional deviant stimuli within a sequence of frequently presented standard stimuli [2]. This allows for developing a sensory memory for the standards, which is violated by the deviants. For instance, in a simulated MMN experiment, a standard tone of 1000Hz and a deviant tone of 1200Hz. The MMN can be seen as an increased negative waveform in the deviant stimuli as opposed to the stimuli presented as standards. This distinction is typically obtained by subtracting the average waveforms of the deviant and standard stimuli. This operation emphasizes the difference and provides insights into the auditory system's ability to discriminate between the two stimuli [7]. The MMN signal is predominantly localized in the auditory cortex and the frontal lobe [10].

The observed difference in the waveform is due to the amplitude drop-off in response to the frequent presentation of standard stimuli in contrast to the less frequent occurrence of deviant stimuli. After first hearing a tone, the amplitude of the corresponding ERP is relatively high and quickly drops if this same tone is repeated.

A very important reason that MMN is popular is that the response can be obtained when the patient is not paying attention to the sounds. However, MMN also has a few disadvantages and those are:

- MMN has a small amplitude which creates a poor SNR ratio. This is especially visible after subtraction.
- MMN is not reliable for individual test-to-test experiments because of the poor SNR. So it is often only used in group comparisons.

- Less reliable in children compared to adults.
- Tests are long in comparison with other methods such as the ACC.

There has been quite a significant amount of research performed on MMNs. This establishes it as a valuable test to perform when starting new experiments. These results can be easily compared with the broad literature that is available on this topic. An illustration of the difference between measured MMNs with subjects that pay attention and subjects that do not do this can be seen in figure 1.2.

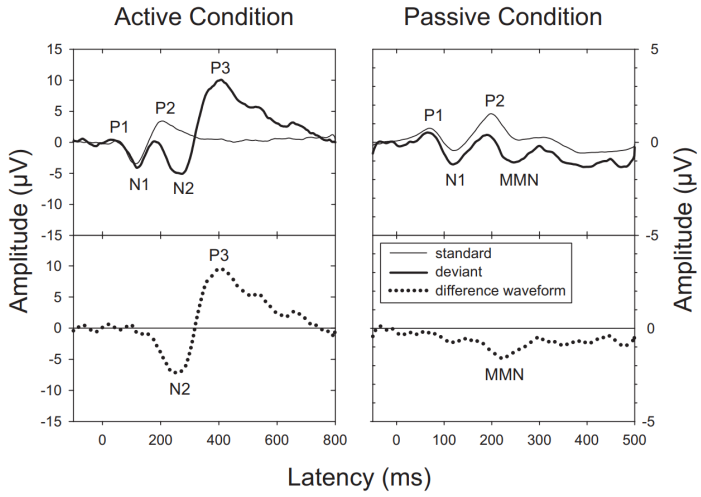


Figure 1.2: MMN waveform from an active and passive condition [7]

1.3.3 Measurements

The measurement procedures are important due to the high sensitivity of EEG data, which can be influenced by numerous factors. Within the context of the Child Brain Lab experiments conducted at Erasmus MC, the subject is situated within a dedicated room, where a speaker is positioned directly in front of them. The type of EEG cap used is a hydrogel GSN with 128 electrodes. When the sounds of the ACC and MMN experiments are being played, the subject is watching a silent movie to minimize any potential movement artifacts. The EEG data gets measured as a continuous signal that contains labels at the time point a sound was played. With the help of these labels, the response of the brain to a specific sound can be collected and averaged per electrode.

However, EEG often contains artifacts that can heavily influence the analysis of these signals. These artifacts can be from the body such as eye movements and muscle contractions but also external devices such as line noise [11]. This is problematic for EEG analysis because the true brain activity is masked by these artifact signals making the data difficult to analyze. After EEG measurements, a common practice is to remove the time frames which contain artifacts by visual inspection or a threshold. For event-related potentials (ERPs), artifacts are problematic as well because a lot of ERPs have a low SNR and are easily distorted [12][13]. Instead of throwing away data segments, it would be better if these artifacts could be removed so that the data can be kept for

analysis. This is why some more preprocessing steps are performed on the EEG dataset in order to obtain a clean ERP waveform. An overview of the different steps can be seen below [8]:

1. Load EEG data in Matlab and assign electrodes and a location map.
2. Bandpass filter between 0.1 and 30 Hz and apply a Notch filter at 60 Hz. This is to remove Line noise at 60 Hz and focus on the important neurological information between 0.1 and 30 Hz.
3. Downsample to 256 Hz for a faster computational time.
4. Make a second copy of the dataset for ICA artifact removal. This will only be used to remove eye blinks.
5. Use a sharper bandpass filter on this second copy to have more focus on the eye blink artifacts and remove unimportant sections of the data which contain a lot of noise. Which is usually at the beginning and end of the measurement.
6. Perform ICA on the second copy and do automatic eye blink removal by analyzing the components with a machine learning algorithm and transfer the weights of this ICA decomposition to the original dataset.
7. Reference data to an average of the two mastoid electrodes and divide the data into different epochs.
8. Remove epochs that have data that go above a certain threshold and average all the remaining epochs to obtain an ERP waveform for each electrode.

1.3.4 Event Related Potential Analysis

The most common approach of analyzing ERPs that is often used in hospitals is by only looking at the amplitudes and delays of certain peaks. For example, to see if there is a difference in the N1 peak between healthy patients and patients with a neurological disorder. Zhao et. al. [14] uses delays and amplitudes of the N170 and N270 ERP peak to see a difference between patients with ADHD and a healthy control group. They found that the N270 peak is delayed in the ADHD group. White et. al. [15] use the amplitude and delay of the P300 component to be able to detect mild cognitive impairment. Cardon et. al. [16] measure the delay and amplitude of the P3 component to look for differences in patients with tinnitus. Tenssay et. al. [17] try to find differences in the amplitude and latencies of ERPs but also combines this with the spectral power of the alpha, theta, and beta bands. There are many more papers that make use of the amplitude and delays of ERPs [18]–[21].

Machine Learning

Only looking at peaks and amplitudes can limit the amount of information that is present in ERP data. In the literature, this analysis is often expanded with more advanced methods to maximize the information use that an ERP contains. A widely

used method for ERP analysis is machine learning and can for example help with the following points [22]:

- Automatically extract complex and subtle features from ERP data. This can result in features that are not easy to identify even by human experts and can reveal hidden patterns or relationships in the data.
- Can be used to classify ERP components by training models on a labeled dataset. The machine learning algorithm can learn to distinguish between different conditions, tasks, or groups.
- Machine learning algorithms can automate the analysis process making this more efficient and reducing human bias.
- Can exploit the multivariate nature of ERP signals by capturing spatial-temporal patterns between the features of different electrodes and subjects.
- Can integrate ERP data with other modalities, such as MRI, to get better insights into the cognitive processes underlying ERP responses.

Machine learning methods also come with some disadvantages and one of them is data interpretability. The complex nature of some algorithms, such as deep learning models, can make it really difficult to understand the learned features or patterns. When the dataset is small or noisy, overfitting can result in poor model performance. Therefore, machine learning algorithms often require a large amount of labeled data for training which is not always present in the available dataset. Another non-trivial part is choosing the right machine learning algorithm and its hyperparameters. This requires consideration of various factors and can lead to suboptimal results if not done right.

The most simple approach is using the time domain features, such as amplitudes and delay, in combination with a machine learning algorithm. Kim et. al. [23] compare ERPs from emergency events in cars with different mental distractions for the driver. The analysis is done with regularized linear discriminant analysis (LDA). Lee et. al. [24] use LDA to decode ERPs for a spellchecker system. LDA is a widespread approach for machine learning tasks in a lot of papers [25]–[27]. Shimigasa et. al. [28] measure the components of the P300 and N100 ERPs to classify EEG signals with artificial neural networks. Cecotti et. al. [29] try to detect P300 in EEG data and use a convolutional neural network to see which ERP corresponds to which character in a spelling paradigm. Khatun et. al [30] combine an ERP's temporal and spectral features with SVM to detect mild cognitive impairment. In a later publication [31] they use SVM in combination with 590 ERP features like amplitudes, latencies, and power in frequency bands. Williams et. al. [32] use an ensemble sparse classifier to discriminate mild cognitive impairment patients from healthy controls.

Matrix Decompositions

Machine learning methods might not perform well if the dataset is small and noisy or the wrong algorithm got selected. A way to rely less on the amount of data and get more

information about the underlying sources in the EEG signal is to use decompositions. A very common decomposition is called independent component analysis (ICA) and a few of the benefits of utilizing ICA are [33]:

- The aim of ICA is to decompose the ERP data into independent sources and can separate underlying neural sources from mixed signals. This enables the identification of brain processes or artifacts and enhances the interpretability of the ERP data.
- ICA is very effective in identifying artifacts such as eye blinks. These artifact components can be isolated and removed resulting in a cleaner ERP signal and are especially helpful with noisy data.
- Provides temporal and spatial information about the identified components which enables researchers to examine the spatial patterns and timings of specific neural sources and artifacts.
- ICA is a data-driven approach and does not require a priori assumptions other than that the data have statistically independent sources.
- Can reduce the dimensionality of ERP data and make it easier to focus on only relevant sources. This can enhance computational efficiency.
- Can integrate ERP with other neuroimaging modalities such as MRI and can provide a more comprehensive understanding of the underlying neural processes.
- It is a well-established technique in the field of ERP analysis which makes it a standardized and reproducible method. This makes it easier to compare with different studies and build upon existing knowledge.

ICA is a blind source separation method and in the case of ERP data, the problem can be written as the ERP data matrix $\mathbf{X}$ equal to an unknown mixing matrix $\mathbf{A}$ multiplied by the source signals $\mathbf{S}$:

$$\mathbf{X} = \mathbf{AS}, \tag{1.1}$$

where $\mathbf{X} \in \mathcal{R}^{N \times M}$ is the ERP data matrix with N channels and M data points, $\mathbf{A} \in \mathcal{R}^{N \times N}$ is the unknown mixing matrix, and $\mathbf{S} \in \mathcal{R}^{N \times M}$ is the matrix which contains N different sources with M data points. The reversed equation of this formula is:

$$\hat{\mathbf{S}} = \mathbf{WX}, \tag{1.2}$$

where $\hat{\mathbf{S}} \in \mathcal{R}^{N \times M}$ is the estimation of the N sources, $\mathbf{W} \in \mathcal{R}^{N \times N}$ is the inverse mixing matrix of $\mathbf{X}$. Blind source separation techniques basically try to estimate this $\mathbf{W}$ matrix. ICA is a linear decomposition technique that tries to reveal the underlying statistical sources of mixed signals. This decomposition has become popular in the application of EEG signals because it provides two important features. These are artifact removal [34][35] and the ability to disentangle the mixed brain signals [36]. ICA decomposes the ERP data into a set of independent components or ICs. In figure

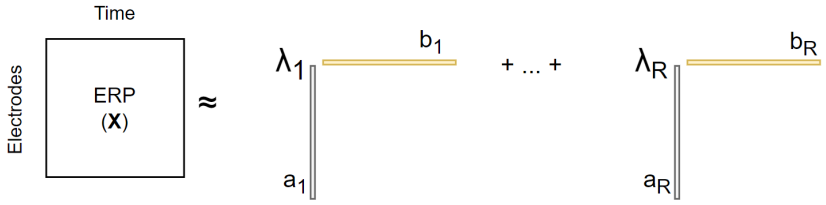


Figure 1.3: Example of matrix decomposition

1.3, a sum of rank 1 matrices can be seen which represents an ICA decomposition. The $a_1 ... a_R$ vectors contain the spatial information of the sources $b_1 ... b_R$.

An example of a study that uses ICA for ERP analysis is by Kalyakin et. al. [37]. They use ICA to extract components from an MMN waveform and divide them into MMN-like and non-MMN-like components. A cleaner MMN waveform can be obtained with this method. Sugi et. al. [38] make use of ICA for averaging ERPs together to get a cleaner end result. Cong et. al. [39] combine wavelet transform and ICA in order to find out how many single trials of MMN recordings are necessary to obtain a good response.

Sometimes, the assumption of ICA that the underlying sources are statistically independent is too heavy and does not hold for EEG data [40]. Another type of decomposition that does not have this assumption and can be used for analysis, is called Empirical Mode Decomposition (EMD). EMD is developed by Huang et. al. [41] to be used for decomposing nonstationary signals into a set of simpler components. These components are called Intrinsic Mode Functions (IMFs). The idea is that EMD decomposes a signal into a finite number of oscillatory components with varying scales and it tries to adaptively get these components directly from the data. The advantages of EMD are that it is an adaptive technique. Unlike ICA, which assumes statistical independence, EMD can adapt to the specific signal characteristics and decomposes them into components based on the local properties. This is especially helpful when dealing with ERP data that contains complex and time-varying oscillatory components. A second advantage is that EMD can capture the local oscillatory behavior and temporal dynamics without requiring a priori knowledge. Another benefit is that EMD is a multiscale decomposition which allows for the characterization of ERP data at different temporal resolutions. This can provide information on the transients and long-lasting oscillatory patterns. This way, EMD can uncover hidden temporal dynamics and provides information about ERP phenomena that occur at different time scales. However, a main disadvantage is that this decomposition can quickly become sensitive to noise and nonstationarities in the data. The algorithm consists of the following steps:

1. Get all local minima and maxima of the signal

2. Fit an envelope to the local minima and to the local maxima (Usually done with cubic spline interpolation)

3. Calculate the mean of the upper and lower envelopes in order to obtain the mean envelope

4. The first IMF is calculated by taking the difference between the original signal and the mean envelope
5. Repeat the first four steps to obtain more IMFs
6. Continue these steps until a stopping criterion is met (usually a number of maxima and minima or properties of the calculated IMF)

Different use cases can be artifact separation, reducing the dimensionality of the data, and providing interpretable components corresponding to specific oscillatory patterns. Chang et. al. [42] use EMD to extract beta-related oscillations from EEG signals of subjects that perform self-paced right and left index-finger lifting tasks. Siuly et. al. [43] perform EMD on ERP signals to get statistical features for a machine-learning algorithm. They want to be able to distinguish ERPs from schizophrenia patients and healthy controls.

EEG signals are also commonly analyzed by their frequency characteristics. Often, the information present in the different frequency bands such as delta, theta, alpha, beta, and gamma is used. EMD and ICA do not provide explicit time-frequency information. If frequency analysis is required, a popular technique is called wavelet decompositions and is very frequently used in biomedical signal processing [44]. This is a technique that decomposes a signal into a set of wavelet functions and allows for frequency analysis at different scales. It is especially useful for nonstationary signals because it provides time and frequency information. Unlike EMD, which decomposes the signal based on local oscillatory behavior, wavelet transform captures frequency information in a structured and systematic manner. This can help with the analysis of specific frequency bands or scales of interest. Wavelet transform could be a good choice because of several reasons:

- Allows for multiresolution analysis of the ERP because it decomposes the data into different frequency components at varying scales. This allows for information on ERP phenomena across different temporal resolutions.
- Wavelet transform provides excellent time-frequency localization. This can help with the precise identification of when specific frequency components are present in the ERP signal. EMD and ICA do not provide explicit time-frequency information.
- Is a flexible method because it offers a wide range of wavelet functions that can be tailored toward the specific problem and data type. Although, this can also be seen as a disadvantage.
- Provides really clear and interpretable ERP data in the time-frequency domain.

Quiroga et. al. [45] make use of wavelet transform by decomposing an ERP signal in different frequency bands and taking a closer look at the alpha and gamma band. Begum et. al. [46] show that with the help of wavelet transform, they find increased alpha and delta waves in a group that has a higher education level compared to the second group. Wavelet transform is also used a lot to obtain features for machine learning algorithms [44], [47]–[51].

Tensors

A growing field in the literature is using tensor decompositions for biomedical signal processing. These types of decompositions are performed on higher dimensional matrices and preserve structural information compared to the matrix decompositions shown before. Tensor structures are very natural in a lot of data sets, especially in biomedical analysis because this often contains group comparison. In the field of EEG/ERP analysis, a matrix is usually not enough to fully explain all the different dimensions obtained from the data. For example, data could have three dimensions such as subjects x electrodes x time. Or for a single patient using wavelet transform, electrodes x frequency x time. For the matrix decompositions, these higher-order tensors often get changed into two-order matrices with the help of unfolding. This can result in a wrong data interpretation and a loss of information among these modes [52]. An illustration of a tensor decomposition is shown in figure 1.4.

Cong et. al. [53] uses nonnegative tensor decompositions to extract multi-domain features of the MMN for cognitive research. Lee et. al. [54] perform nonnegative tensor factorization on a continuous EEG signal in order to classify multiple mental tasks. Wang et. al. [55] build a 5th-order tensor with dimensions channel x frequency x time x subject x condition and perform a nonnegative tensor decomposition to extract multi-mode ERP features. Aghabeig et. al. [56] conduct an oddball ERP experiment and try to characterize differences in the processing of stimuli using multi-domain features. Pouryazdian et. al. [57] propose a technique to select multi-domain features of an MMN among all extracted features by a template-matching approach. The template consists of information on the MMN temporal and spectral signatures. They use statistical tests to determine which feature significantly discriminates two groups of subjects. Cong et. al. [58] use a nonnegative tensor factorization to get multi-domain MMN features that can discriminate two groups of children which cannot be done by only looking at delays and amplitudes of the MMN.

During EEG measurements, often, multiple experiments/tests are done with the subject. Sometimes this includes psychological tests, MRI data, or just another EEG experiment. This extra data can provide more relevant information for the decomposition and a common method to combine this data is with data fusion methods [59]. By jointly analyzing these different data modalities, more detailed information could be extracted. For example, Acar et. al. [60] use functional MRI, structural MRI, and EEG data to look for biomarkers that say something about the difference between patients with schizophrenia and healthy control subjects. An advanced coupled matrix tensor factorization (ACMTF) is used to fuse these two data types and use a single decomposition. An example of using MRI data in combination with an EEG tensor can be seen in figure 1.5.

1.3.5 Discussion

In short, the most common approach to analyzing ERP signals is done by looking at the delays and amplitudes of the different ERP peaks. In the literature, a very popular approach is by doing this analysis with machine learning or deep learning. Matrix decompositions such as ICA, EMD, and wavelet transform can also be used to extract

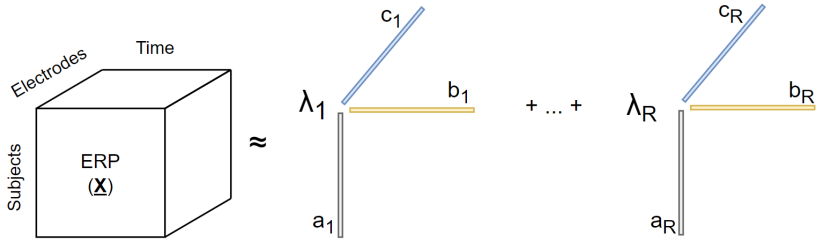


Figure 1.4: Example of tensor decomposition

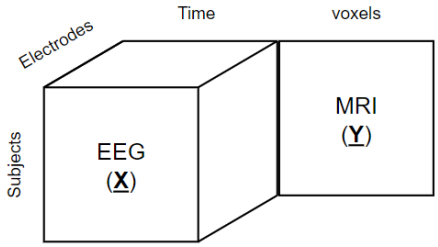


Figure 1.5: Data structure of joint tensor and matrix analysis

more information from the ERP signals and are often combined with a machine-learning method. The data can usually be naturally written in a tensor structure and changing this to a matrix can result in wrong data interpretation and information loss. Therefore, tensor decomposition can be used to keep the structural information. These types of decompositions are gaining popularity in the biomedical signal-processing literature.

However, a choice has to be made for a method to use in this thesis on the data from Erasmus MC. Since the measurements just started, not a lot of data will be present. Machine learning and deep learning approaches usually require a lot of data and therefore might not be the best method to try on the available dataset (at the end of the thesis there was only a dataset of six patients which is not included in this report since it did not add anything to the results and conclusion of the project). A matrix decomposition method can also be used and might require less data compared to machine learning approaches. However, the data can be naturally written in a tensor with dimensions subjects x electrodes x time, and changing this data to a matrix can result in a loss of information and wrong data interpretation. Therefore, the rest of this thesis contains a more detailed look into tensors and how to use these decompositions to look at the effect of neurodevelopment on auditory event-related potentials.

2 Methods

As discussed in section 1.3.4, tensor decompositions have become more prevalent in EEG-related data processing and cognitive neuroscience. A lot of data can already naturally be represented with tensors. For example, ERP data can be written as a tensor with dimensions electrode x time x trial or electrode x time x patient. The data can also be transformed to the frequency domain in order to get a tensor with dimensions electrode x frequency x time. Processing high-order tensors is often difficult and time-consuming and sometimes has an impact on the stability and convergence of the method used. It is common to reshape a high-order tensor into a low-order tensor by unfolding it into a two-order matrix or combining modes together. However, this can result in a wrong data interpretation and loss of information among these modes [52]. If the higher-order tensor is kept as it is, a tensor decomposition can be performed. For the data from Erasmus MC, certain components extracted from the decomposition can be used to tell something about group differences present in the data. In figure 2.1, the data that will be obtained by the measurements from Erasmus MC can be seen. The two tensors contain the ERP data over time for each electrode and each subject for the ACC and MMN experiment.

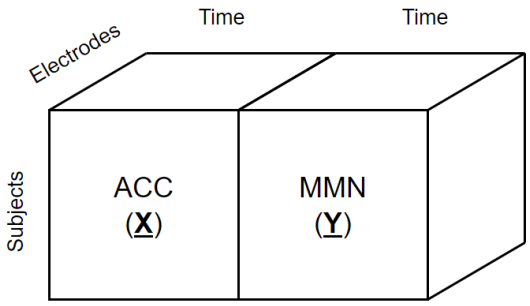


Figure 2.1: Data from Erasmus MC

Further in this chapter, a detailed explanation will be given of what tensor decompositions are, how they can be used, and how this can be applied to the data used in this thesis.

2.1 Decompositions

Many signal and data analysis techniques rely on matrix decompositions, where the data/signal matrix is decomposed into a set of factor matrices. These factor matrices can carry important information about the data. Similarly, if the data/signal is in the

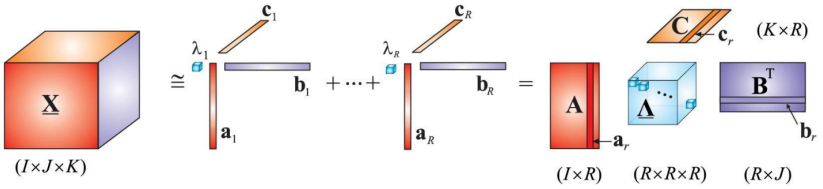


Figure 2.2: Schematic of CP decomposition [61]

form of a tensor, a tensor decomposition can be used. Tensor decompositions exploit the interactions among multiple modes of the tensor. Two very common decompositions are Canonical Polyadic decomposition (CPD) and Tucker decomposition (TD). CPD is also known as Parallel Factor Analysis (PARAFAC) and canonical decomposition (CANDECOMP) [61].

2.1.1 Canonical Polyadic Decomposition

The CP decomposition decomposes an Nth order tensor into a linear combination of terms, $\mathbf{u}_r^{(1)} \circ \mathbf{u}_r^{(2)} \circ \cdots \circ \mathbf{u}_r^{(N)}$. Each term in this formula is a rank-1 tensor and the full tensor can be written as:

$$\begin{aligned} \underline{\mathbf{X}} &= \sum_{r=1}^{R} \lambda_r \mathbf{u}_r^{(1)} \circ \mathbf{u}_r^{(2)} \circ \cdots \circ \mathbf{u}_r^{(N)} \\ &= \underline{\mathbf{\Lambda}} \times_1 \mathbf{U}^{(1)} \times_2 \mathbf{U}^{(2)} \cdots \times_N \mathbf{U}^{(N)}, \end{aligned} \tag{2.1}$$

where $\underline{\mathbf{X}} \in \mathcal{R}^{I_1 \times I_2 \times \cdots \times I_N}$ is an Nth-order tensor, λ_r are the diagonal entries of the core tensor $\underline{\mathbf{\Lambda}} \in \mathcal{R}^{R \times R \times \cdots \times R}$, and $\mathbf{U}^{(n)} = [\mathbf{u}_1^{(n)}, \mathbf{u}_2^{(n)}, \ldots, \mathbf{u}_R^{(n)}]$ are the factor matrices. Figure 2.2 shows a schematic of the CP decomposition.

The CPD is unique up to scaling and permutation with very mild conditions required and is usually computed using an Alternating Least Squares (ALS) algorithm [61]. Extra constraints can be added to ensure the data is unique or make the computation faster. For example, if the ERP data is transformed to the frequency domain, a nonnegative CP decomposition can be performed which can outperform the original CP decomposition [54].

$$\begin{aligned} &\min_{\mathbf{U}^{(1)},\ldots,\mathbf{U}^{(N)}} \frac{1}{2} ||\boldsymbol{\mathcal{X}} - [[\mathbf{U}^{(1)}, \ldots, \mathbf{U}^{(N)}]]||_F^2 \\ &s.t. \mathbf{U}^{(N)} \geq 0 \; for \; n = 1, \ldots, N, \end{aligned} \tag{2.2}$$

where $\mathbf{U}^{(N)} \in \mathcal{R}^{I_n \times R}$ are the estimated factor matrices. There are various optimization methods that can solve this NCP problem such as multiplicative updating (MU), alternating least squares (ALS), hierarchical alternating least squares (HALS), or the more recent alternating proximal gradient (APG) [55].

In figure 2.3, a graphical illustration of a tensor decomposition is shown with three factor matrices and a core tensor.

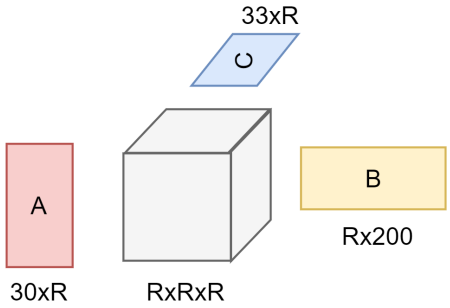


Figure 2.3: Graphic illustration of a tensor decomposition with 30 subjects, 200 data points, and 33 electrodes

2.1.2 Tucker Decomposition

The Tucker decomposition (TD) provides a more general factorization of an Nth-order tensor. It does this by decomposing the tensor into a core tensor and factor matrices and can be written as follow:

$$\begin{aligned} \underline{\mathbf{X}} &= \sum_{r_1=1}^{R_1} \cdots \sum_{r_N=1}^{R_N} g_{r_1 r_2 \cdots r_N} (\mathbf{u}_{r_1}^{(1)} \circ \mathbf{u}_{r_2}^{(2)} \circ \cdots \circ \mathbf{u}_{r_N}^{(N)}) \\ &= \underline{\mathbf{G}} \times_1 \mathbf{U}^{(1)} \times_2 \mathbf{U}^{(2)} \cdots \times_N \mathbf{U}^{(N)}, \end{aligned} \tag{2.3}$$

where $\underline{\mathbf{X}} \in \mathcal{R}^{I_1 \times I_2 \times \cdots \times I_N}$ is an Nth-order tensor, $\underline{\boldsymbol{G}} \in \mathcal{R}^{R_1 \times R_2 \times \cdots \times R_N}$ is the core tensor, and $\mathbf{U}^{(n)} = [\mathbf{u}_1^{(n)}, \mathbf{u}_2^{(n)}, \ldots, \mathbf{u}_R^{(n)}]$ are the mode-n factor matrices.

The difference between TD and CPD is that in TD the core tensor can be any tensor with compatible sizes but in CPD the core tensor has nonzero elements only in the diagonal. In addition to this, TD is not unique but extra constraints can be imposed on the factor matrices which can yield a unique core tensor and factor matrices [52]. In practice, the TD is often calculated with the Multilinear Singular Value Decomposition (MLSVD), which is also called the Higher-Order SVD (HOSVD). This can be seen as a constrained tucker decomposition where the factor matrices are orthogonal and the core tensor is all-orthogonal (slices in each mode are mutually orthogonal) [61]. Another difference between CPD and TD is that the number of components in the modes can be different from each other which is not the case in a CPD. Next to that, each component in each mode interacts with any other component in any other mode. That is why prior information on all modes is needed to get components of interest in TD because you cannot automatically say that the first columns of each factor matrix are solely related to each other [52]. In general, the tucker decomposition is a bit more flexible than CPD.

For the rest of this thesis, the CPD decomposition will be used. The main reason for this is the easier interpretability of the components. The TD also requires a choice of the number of components in each mode which is not trivial. In order to get more information about how different methods perform, CPD seems to be an easier choice to start with. For the next step in this project, it might be beneficial to also look at the performance of a TD or other decompositions.

2.2 Component Analysis

An important step after the decomposition is actually finding components of interest. If the EEG measurement setup is the same for all subjects you can assume that the underlying factors are the same across subjects, however, there is a subject-dependent strength and condition-dependent strength [55]. Which component is of interest depends on the research question, for example, the extracted components can be used to differentiate across groups, yield more insight into the data, offer a better representation of cognitive functions, or train a classifier for the diagnosis of patients. One way to choose relevant components can be with prior information from the performed test and expected results. For example, the MMN is a well-defined ERP and can easily be recognized from the components based on its properties in time, frequency, and spatial domain [57]. This can ensure that the components of interest actually give information about the MMN.

Another very common approach is by using a statistical test on the subject mode factor matrix columns to see if there is a clear group difference visible. Acar et. al. [60] use a two-sample t-test to find statistically significant components. This basically means the components that can say something about a group difference in the datasets. A two-sample t-test is performed on each column of the factor matrix from the subject mode. The vectors that have p-values < 0.05 are identified as statistically significant. After this, they use a simple k-means clustering algorithm to find potential groups in the data. The clustering is performed on all possible combinations of the columns of the factor matrix.

The method that will be used in this report is the statistical test in combination with a simple k-means clustering. If prior information is used about the ERP waveforms to select components of interest, maybe some variation in the ERP due to a specific neurological disorder is missed because of these pre-defined assumptions. In the future, k-means clustering can be changed into a more complex machine-learning algorithm.

2.3 Data Fusion

In order to perform analysis on the Erasmus MC data and look for group differences, the tensor decompositions discussed in the section 2.1 can be applied to the ACC and MMN tensor individually. This will result in a number of components for each tensor where the columns of factor matrix $\mathbf{A}$ say something about how strong a component is present in each patient, in factor matrix $\mathbf{B}$ about temporal information of the ERP waveform, and in factor matrix $\mathbf{C}$ about the spatial information. As discussed in the previous section, after the decompositions, the relevant components can be selected by performing a two-sample t-test. The classification can be performed with k-means clustering on values in the columns of factor matrix $\mathbf{A}$ and more information about this difference can be analyzed with factor matrix $\mathbf{B}$ and $\mathbf{C}$.

However, it is expected that the two tensors which contain the ERP information from two measurements have a correlation in the group differences. If a patient has trouble distinguishing sound, an effect is expected to be seen in the ACC and MMN. This is an assumption and is not certain but they are both tests where the subjects passively

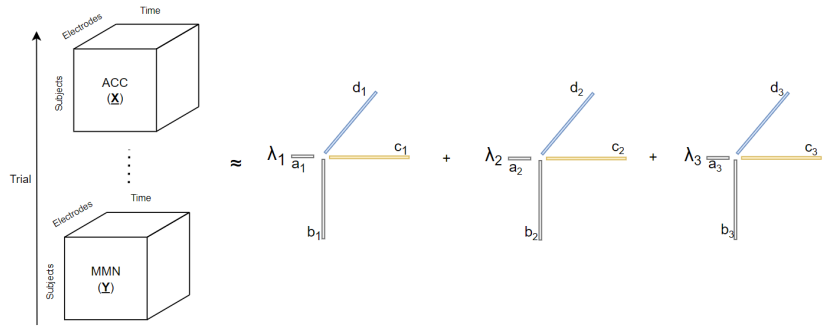


Figure 2.4: Example 4D tensor decomposition which combines the two paradigms

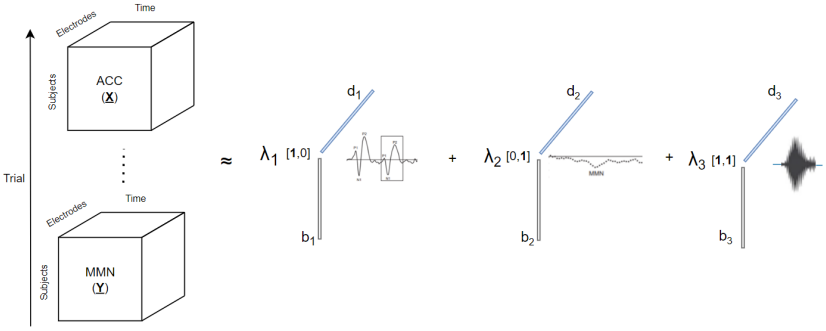


Figure 2.5: Example 4D tensor decomposition which has ideal interpretable components

listen to some tones. Instead of analyzing these two tensors separately, they can be simultaneously used for finding biomarkers. It might be better to use the information that both tensors contain about the group differences in a single decomposition and maybe this can give better results. One way of combining the data from the two experiments is by creating a 4D tensor with an extra domain called "Trial". This domain refers to the two measurement setups for the ACC and MMN. A schematic of the decomposition can be seen in figure 2.4. Instead of three vectors for each component, there are now four.

An ideal scenario can be seen in figure 2.5. The decomposition consists of three components, where one component is related to the ACC, one to the MMN, and the last one to a noise component. In the a_1, a_2, and a_3 vectors, it is clear that it is a component present in the ACC trial, MMN trial, or both. However, it does not work like this in practice because there are many degrees of freedom. The components will not be clearly related to the ACC or the MMN but will probably be a mixture of both. This makes it really hard to understand the components and get reliable and consistent results back.

In section 1.3.4, it was explained that Acar et. al. [60] use a joint decomposition to make use of the information contained in an fMRI matrix, sMRI matrix, and an EEG tensor. A common approach is by finding a shared dimension in both datasets which in the case of finding biomarkers is often the subject/patient dimension. After the joint decomposition, the components of the different datasets have the same subject/patient factor matrix. So if you find a biomarker in dataset "A" you can find a component

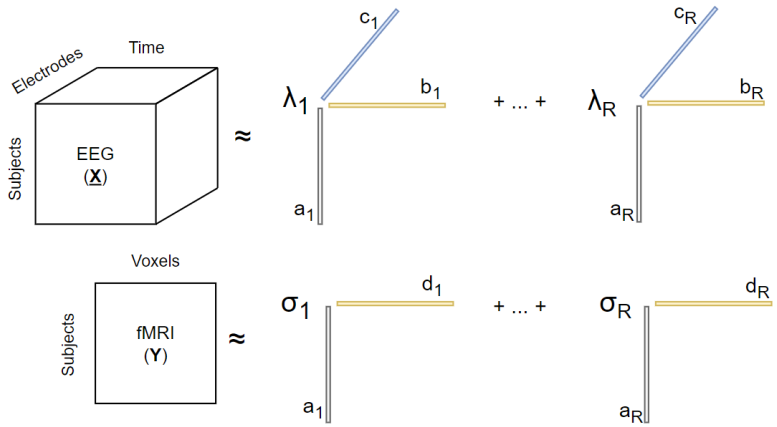


Figure 2.6: CMTF of EEG and fMRI

with a similar subject distribution in dataset "B". These MRI data matrices can also be replaced by an EEG tensor in order to use these joint decompositions on the dataset from Erasmus MC. The original approaches for data fusion are based on matrix factorizations for example joint independent component analysis (jICA) [62]. If EEG data is used, it has to be written as a matrix such as subject x time. However, as previously discussed, EEG data can often be naturally represented with a tensor and if it is transformed into a matrix, structural information is lost.

The rest of this section will explain how joint decompositions work, see different variations of joint decompositions [59], and rewrite the cost functions to work with the data from Erasmus MC.

2.3.1 Coupled Matrix Tensor Factorization

As seen before, a common method is to combine EEG with MRI data in order to gain more insight into the underlying neurological sources of the brain. Hunyadi et. al [62] use data fusion of an EEG data tensor together with a functional MRI matrix. This technique is called coupled matrix-tensor factorization (CMTF) and it works without the need of reorganizing the tensor so it can exploit the multidimensional structure. The cost function that needs to be minimized for an EEG tensor $\underline{\mathbf{X}} \in \mathcal{R}^{subjects \times time \times channel}$ and an fMRI matrix $\mathbf{Y} \in \mathcal{R}^{subjects \times voxels}$ looks as follows:

$$f(\lambda, \mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{S}^{fMRI}) = ||\underline{\mathbf{X}} - [[\lambda; \mathbf{A}, \mathbf{B}, \mathbf{C}]]||^2 + ||\mathbf{Y} - \mathbf{A}\mathbf{S}^{fMRI}||^2, \tag{2.4}$$

where $[[\lambda; \mathbf{A}, \mathbf{B}, \mathbf{C}]]$ corresponds to the CPD with factor matrices $\mathbf{A} \in \mathcal{R}^{subjects \times R}$, $\mathbf{B} \in \mathcal{R}^{time \times R}$, $\mathbf{C} \in \mathcal{R}^{channels \times R}$, and super diagonal of the core tensor $\lambda \in \mathcal{R}^{1 \times R}$. In figure 2.6 an illustration of the decomposition can be seen. What is clear from the figure and the equation is that both datasets share the $\mathbf{A}$ factor matrix. This means that if, for example, the first component in the decomposition tells something about a group difference. There is a corresponding fMRI and EEG component with that exact same group distribution.

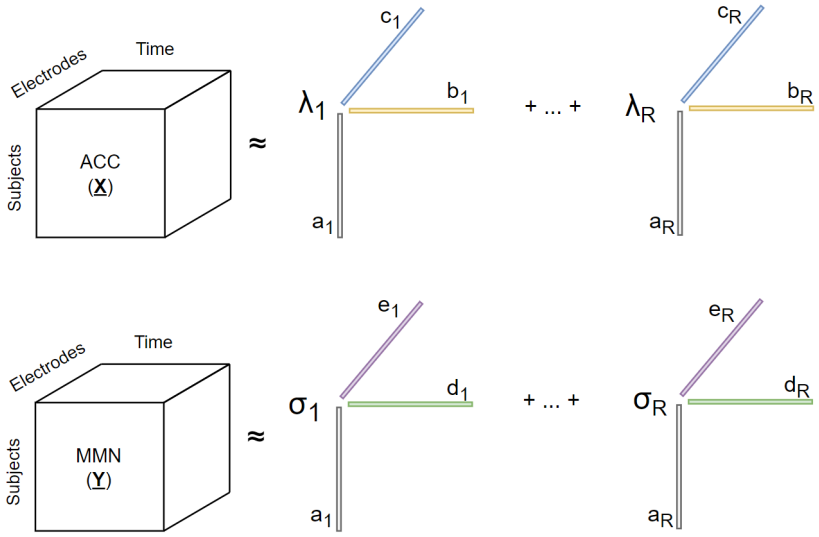


Figure 2.7: CMTF modified to fixed data fusion on Erasmus MC data

This CMTF method can also be extended to the ACC and MMN tensor from figure 2.1. It will not be a coupled matrix-tensor factorization anymore but will become a coupled tensor factorization. In the rest of this thesis, this method will be called fixed data fusion (FDF). To let this work with the ACC tensor $\underline{\mathbf{X}} \in \mathcal{R}^{subjects \times time \times electrodes}$ and MMN tensor $\underline{\mathbf{Y}} \in \mathcal{R}^{subjects \times time \times electrodes}$, the cost function can be modified as follows:

$$f(\lambda, \sigma, \mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D}, \mathbf{E}) = ||\underline{\mathbf{X}} - [[\lambda; \mathbf{A}, \mathbf{B}, \mathbf{C}]]||^2 + ||\underline{\mathbf{Y}} - [[\sigma; \mathbf{A}, \mathbf{D}, \mathbf{E}]]||^2, \tag{2.5}$$

where $\sigma \in \mathcal{R}^{1 \times R}$ and $\lambda \in \mathcal{R}^{1 \times R}$ are the superdiagonal of the two core tensors and $\mathbf{A} \in \mathcal{R}^{subjects \times R}, \mathbf{B} \in \mathcal{R}^{time \times R}, \mathbf{C} \in \mathcal{R}^{electrodes \times R}, \mathbf{D} \in \mathcal{R}^{time \times R}$, and $\mathbf{E} \in \mathcal{R}^{electrodes \times R}$ are the corresponding factor matrices.

In figure 2.7 the cost function is illustrated. In figure 2.8, a matrix illustration of the joint decomposition with two core tensors can be seen. Both decompositions share the $\mathbf{A}$ factor matrix with information from the subject dimension. In this example, there are 30 subjects, 200 data points, 33 electrodes, and R components.

2.3.2 Advanced Coupled Matrix Tensor Factorization

The fixed data fusion method keeps the structural information from the tensors and can isolate shared components. However, this method assumes that the coupled datasets only have shared components and may fail to capture the underlying patterns in the presence of shared and unshared components. The hard constraint that the shared factors must be exactly the same in both decompositions is maybe a bit too strict [63]. For example, what if the components of interest are the ones that tell something about a group difference but this group difference is not exactly shared across the two datasets? The hard constraint of having exactly the same subject mode factor matrices is not great in that case and a relaxed version might be better. This is why another approach called structure revealing CMTF or advanced CMTF (ACMTF) is

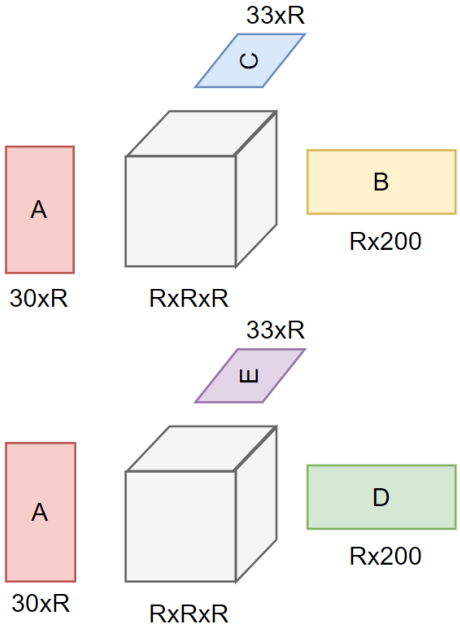


Figure 2.8: Graphical illustration of the fixed data fusion method

used. Acar et. al. [64] use this ACMTF to perform a joint decomposition on an fMRI matrix, sMRI matrix, and EEG tensor. This method allows for shared and unshared components to be present. The ACMTF model with R-components, EEG tensor $\underline{\mathbf{X}} \in \mathcal{R}^{subjects \times time \times electrodes}$, fMRI matrix $\mathbf{Y} \in \mathcal{R}^{subjects \times voxels}$, and sMRI matrix $\mathbf{Z} \in \mathcal{R}^{subjects \times voxels}$ minimizes the following objective function:

$$\begin{aligned} f(\lambda, \Sigma, \Gamma, \mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D}, \mathbf{E}) =& ||\underline{\mathbf{X}} - [[\lambda; \mathbf{A}, \mathbf{B}, \mathbf{C}]]||^2 + ||\mathbf{Y} - \mathbf{A}\Sigma\mathbf{D}^T||^2 \\ &+ ||\mathbf{Z} - \mathbf{A}\Gamma\mathbf{E}^T||^2 + \beta||\lambda||_1 + \beta||\sigma||_1 + \beta||\gamma||_1, \end{aligned} \tag{2.6}$$

where $\lambda, \sigma, \gamma \in \mathcal{R}^{1 \times R}$ contain the weights of the different components, Σ and $\Gamma \in \mathcal{R}^{R \times R}$ are diagonal matrices with σ and γ, $\beta > 0$ is a penalty parameter so that unshared components have weights close to zero, and $\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D}$, and $\mathbf{E}$ are the corresponding factor matrices. The method is illustrated in Figure 2.9.

The cost function can again be modified to fit the problem in this thesis with the ACC tensor $\underline{\mathbf{X}}$ and MMN tensor $\underline{\mathbf{Y}}$ as follows (ADF):

$$\begin{aligned} f(\lambda, \sigma, \mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D}, \mathbf{E}) =& ||\underline{\mathbf{X}} - [[\lambda; \mathbf{A}, \mathbf{B}, \mathbf{C}]]||^2 + ||\underline{\mathbf{Y}} - [[\sigma; \mathbf{A}, \mathbf{D}, \mathbf{E}]]||^2 \\ &+ \beta||\lambda||_1 + \beta||\sigma||_1. \end{aligned} \tag{2.7}$$

2.3.3 Soft Variants

The advanced data fusion method seems to be a more realistic option for real datasets since the group distribution might not be shared. The unshared components can still capture ERP components with a group difference that is heavily present in one dataset but not so strong in the other. However, a drawback is that it is sensitive to its parameters (number of components R and penalty parameter β) [63]. Advanced data

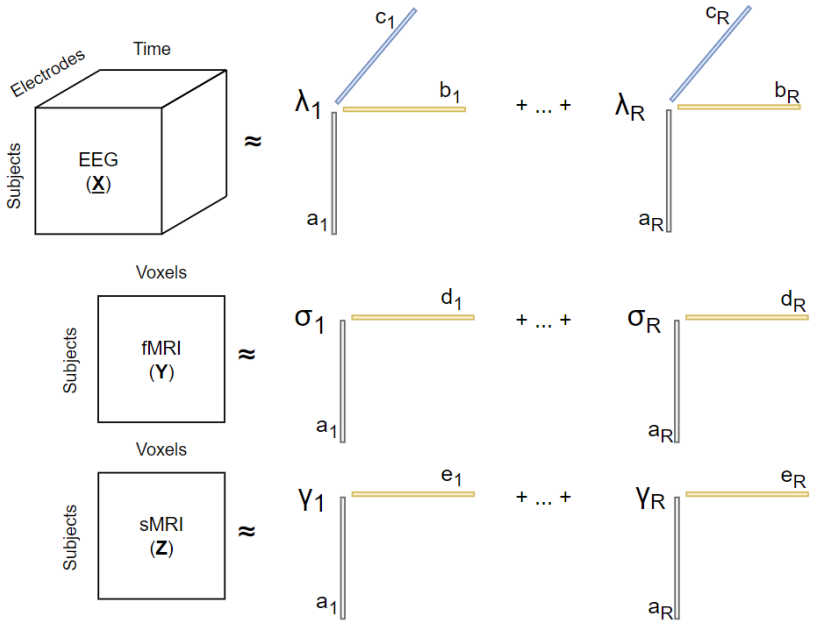


Figure 2.9: ACMTF data decomposition with EEG, fMRI, and sMRI data

fusion also relies on a hard coupling assumption, which means that the factor matrices are exactly the same across all datasets. However, it might be better to use a more relaxed assumption using a soft coupling approach [59]. For example, what if group differences are present in the MMN tensor but in the ACC this group difference is not exactly shared and is slightly different? Instead of the exact same factor matrix and allowing unshared components, a relaxation on the shared factor matrices can be added. The cost function for a soft data fusion can be written as follows (SDF):

$$f(\lambda, \sigma, \mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{U}, \mathbf{D}, \mathbf{E}) = ||\underline{\mathbf{X}} - [[\lambda; \mathbf{A}, \mathbf{B}, \mathbf{C}]]||^2 + ||\underline{\mathbf{Y}} - [[\sigma; \mathbf{U}, \mathbf{D}, \mathbf{E}]]||^2 + \alpha||\mathbf{A} - \mathbf{U}||^2. \tag{2.8}$$

Note that instead of the shared factor matrix $\mathbf{A}$, the second decomposition now uses a factor matrix $\mathbf{U}$. The similarity between these two factor matrices can be adjusted with the term $\alpha||\mathbf{A} - \mathbf{U}||_F^2$. For a big value of α, the factor matrices $\mathbf{A}$ and $\mathbf{U}$ will be really similar, and for smaller values of α, the factor matrices can be different from each other.

In figure 2.10, the factor matrices and core tensors of the decomposition can be seen. Instead of a shared factor matrix, there are two separate matrices of the subject dimension. The similarity can be changed with the α parameter.

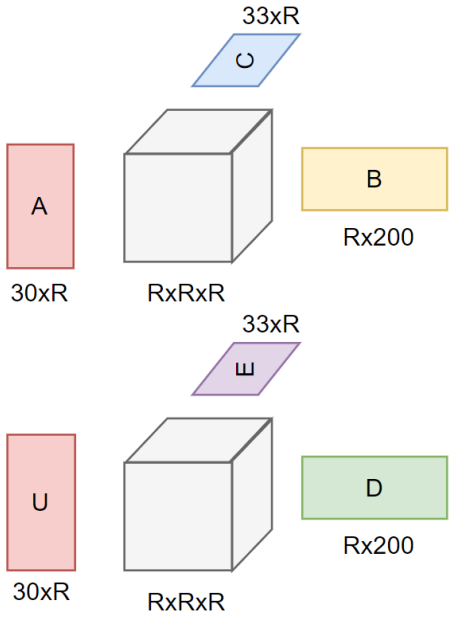


Figure 2.10: Graphic illustration of the soft data fusion method

3 Simulation

In this chapter, the tensor decomposition methods discussed in the previous chapter will be tested on simulated data. First, it will be explained how and what artificial data is generated. After this, there will be a section about how the tensor decompositions are used with this data which will be followed by the actual experiment results.

3.1 Generating Data

The data is generated through BESA simulator software, which allows producing continuous EEG data with event-related potentials. This software program lets you create an artificial ERP waveform at specific locations in the brain and automatically distributes this data across 33 electrodes (see figure 3.1). The user can set coordinates in the brain for a dipole source with a specific direction. Additionally, the waveform that is present at those locations can be created entirely from scratch. The user can add peak values and two consecutive points will automatically be connected. The output will be a continuous EEG signal with this artificial ERP present. The number of times the ERP occurs in the EEG signal and the time between two consecutive ERPs can be adjusted. Next to that, more parameters that can be adjusted are the amount of white noise, alpha noise, and the variation in the repeated ERP peaks.

To keep the simulation similar to what is measured at Erasmus MC, the simulated ERP data is based on the MMN paradigm, involving ERPs in response to standard and deviant sounds. Figure 3.1 shows the BESA application with an MMN-like waveform. Specifically, the focus area in the brain is the left and right auditory cortex, as they play

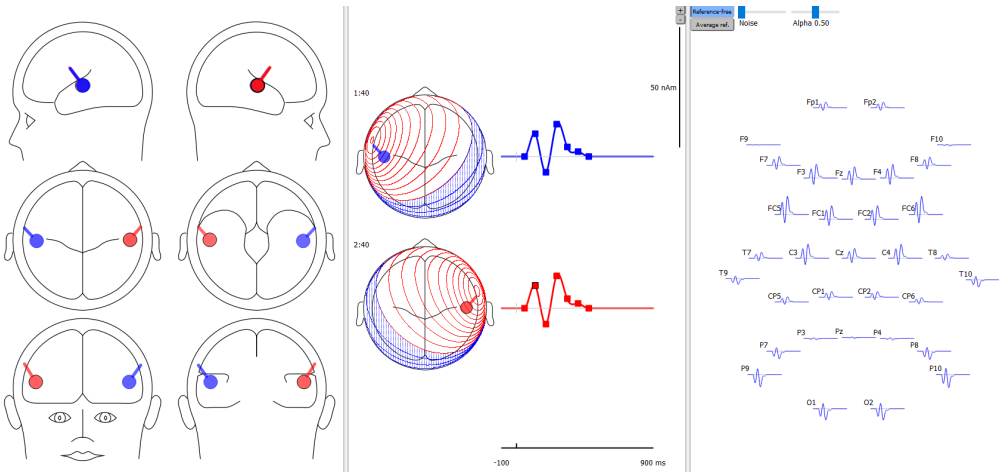


Figure 3.1: BESA simulator ERP design window

a crucial role in sound processing according to existing research [9]. The dipole direction is also chosen to replicate the MMN generator processes [65]. The middle part of figure 3.1 illustrates the actual ERP waveform that should simulate the real response of the brain to a sound. The height and shape of the different peaks are based on figure 1.2 in section 1.3.2. In order to resemble the MMN and ACC tests conducted at Erasmus MC, the ERP is repeated 200 times with an interval of 2 seconds. Additionally, to enhance the realism of the EEG data, white noise and alpha wave noise is added to the signal. With the output EEG data containing 200 ERPs from BESA simulator, Matlab is used to average the single ERPs together, resulting in a final averaged ERP waveform.

With the help of BESA simulator, two groups will be created that both have a deviant and standard waveform. The amount of difference between the deviant and standard in both groups can be adjusted to see how the methods perform. The difference will be present in the later peaks of the ERP as we have seen before in the MMN section 1.3.2. The waveforms of the two groups will be generated multiple times with different noise levels. These different noise levels can mask the difference that is present within a group or make it seem like there is a difference whilst there actually is not.

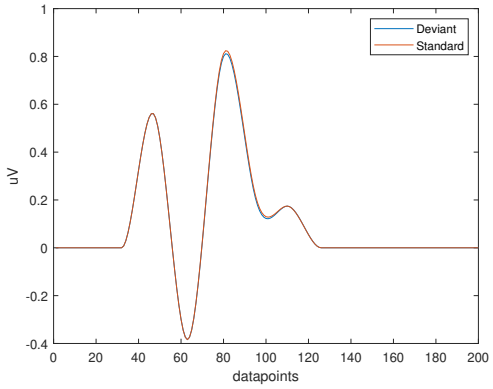
For example, a scenario where there is a difference between the standard and deviant in group 2 but not in group 1, can be seen in figure 3.2. In this case, group 1 could be subjects with a neurological disorder that has trouble with distinguishing sounds while group 2 are healthy control subjects. The clean waveforms are illustrated in figure 3.2a and 3.2b while the waveforms of three different noise levels can be seen in figure 3.2c, 3.2d, and 3.2e.

3.2 Using Tensor Decompositions To Reveal Group Differences

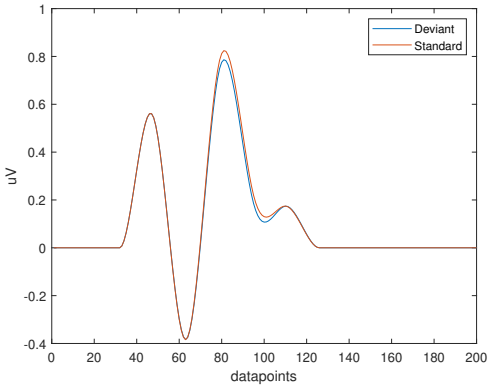
In order to see the effects of the different decompositions, the four methods discussed in chapter 2 will be compared to each other:

- CPD on single tensor
- Fixed joint decomposition (FDF) on two tensors
- Advanced joint decomposition (ADF) on two tensors
- Soft joint decomposition (SDF) on two tensors

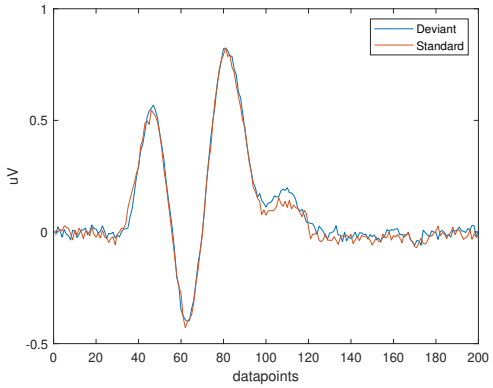
As explained in chapter 2, the most important reason for the choice of a CPD is the easy interpretability of the components. This makes it less difficult to understand what happens in the decomposition and helps with understanding the different data fusion methods. The classification of which subject belongs to which group will be done with k-means clustering on all the possible combinations of the columns of factor matrix $\mathbf{A}$. However, only columns corresponding to statistically significant components will be considered for analysis. The significant components will be detected with a two-sample t-test and p-values lower than 0.05 are identified as significant [60].



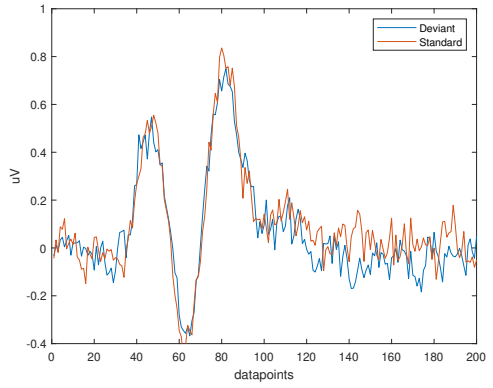
(a) The clean standard and deviant waveform of group 1



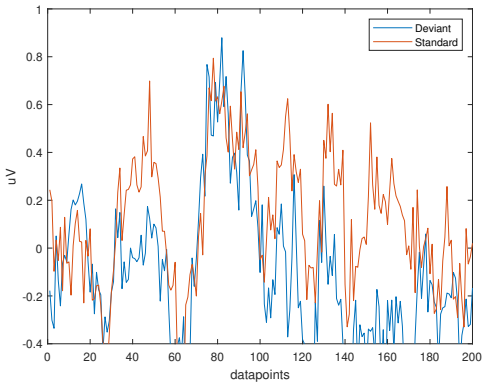
(b) The clean standard and deviant waveform of group 2



(c) The standard and deviant waveform of group 1 with an SNR of 21dB



(d) The standard and deviant waveform of group 1 with an SNR of 11dB



(e) The standard and deviant waveform of group 2 with an SNR of 4dB

Figure 3.2: Example of clean and noisy deviant/standard of both groups. Note that in group 1 there is no difference present between the standard and deviant. In group 2, the deviant is slightly different than the standard.

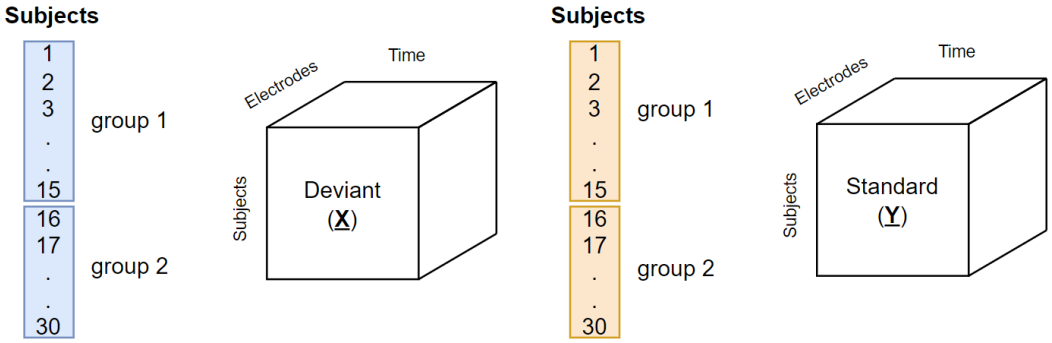


Figure 3.3: Two tensors build from the deviants and standards of both groups.

In order to perform a tensor decomposition, two tensors are created from the deviant and standard of each subject. Figure 3.3 illustrates the two tensors containing the deviant and standard waveforms. The Signal-to-Noise Ratio can be adjusted in BESA simulator to get a more realistic EEG waveform. Although the clean standard and deviant are the same for all subjects within each group, the noisy ERP is different for each subject because this gets added individually. The difference between the standard and deviant of both groups is less visible and can be masked by the added noise. This makes it more difficult for any decomposition to detect differences between both groups. The number of subjects will be 30 because this seems to be a good representation of what is also commonly used in the literature [60].

The classification accuracy is likely to depend on two factors which are the amount of SNR and the actual difference between the two groups. The amount of difference there is in a group can be represented by the similarity percentage between the two waveforms. In this case, we calculate the fit between the two groups in a tensor with the following formula [60]:

$$Fit = 100 \times (1 - \frac{||\underline{\mathbf{X}}_{group1} - \underline{\mathbf{X}}_{group2}||^2}{||\underline{\mathbf{X}}_{group1}||^2}), \tag{3.1}$$

where $\underline{\mathbf{X}}_{group1}$ is a tensor representation of the first group present in $\underline{\mathbf{X}}$ and $\underline{\mathbf{X}}_{group2}$ is the tensor representation of the second group present in $\underline{\mathbf{X}}$. If this formula is used to check the similarity between the deviant waveforms (figure 3.2a and figure 3.2b) of both groups in tensor $\underline{\mathbf{X}}$, a value of 99% is the result. If the value was 100%, the waveform of the two groups would be completely identical which is the case in tensor $\underline{\mathbf{Y}}$ from figure 3.3. Obviously, if noise is added individually to each patient, this value goes down but the actual difference of the ERPs is harder to detect.

3.3 Performed Experiments

During the simulations, two cases will be tested, namely, if there is only a difference present in one tensor and if there is a difference present in both tensors. These two cases will hopefully show the benefit of using the joint decompositions but also see how the decompositions perform when the assumption that both tensors contain the same group

information is not actually true. The most extreme case is if the second tensor contains no group difference at all. Different SNR values will be tested in both simulations to see how the methods perform in certain SNR scenarios. Three SNR values will be selected to show in this report ranging from "bad" results to "perfect" results in classification accuracy. This is to see what the relationship is between changing the SNR value and thus changing the difficulty for the methods to be able to detect the difference between the two groups.

As mentioned before, the difference that is present in the later ERP peaks is based on the MMN. There is no specific reason for the exact amount of difference and this is just a value chosen in combination with the SNR levels to get results ranging from "bad" to "perfect". If the difference was chosen to be smaller, the SNR values would have been higher to get the same results.

3.4 No Shared Difference

The first test case is that the standard waveform is the same in both groups but there is a difference in the deviant. The clean waveform data is the same as in figure 3.2a and figure 3.2b. If all algorithms are tested and k-means clustering is performed on the significant components, the results are as in figure 3.4. This figure shows the classification accuracy for 10 iterations that can be achieved with the four different methods for three SNR values. The first method, CPD on a single tensor, is performed on the deviant and standard tensors.

The optimal number of components for each decomposition is chosen by comparing the classification accuracy of different values and choosing the one that gives the best result. Table 3.1 gives an overview of the effect of different component numbers for each algorithm. The component numbers that give the highest accuracy are used for figure 3.4. The number of components needed is generally higher for the data fusion methods compared to the CPD on a single tensor. This is especially the case for ADF and SDF because unshared components are also allowed. In general, that results in more important components that can be included in the decomposition. The simulated data in this chapter is really simple so the extra components needed are usually just one or two. However, for more complex data, this number can be in the range of 10 components [60]. The advanced data fusion and soft data fusion also have some extra α and β variables. These values are chosen in the same way as the components, which is by comparing the classification accuracy for different values and choosing the one that results in the best performance. The differences are not that big for different values of α and β, however, in the extremes you do see an effect. For example, if β is approaching zero, the advanced data fusion seems to behave more as the fixed data fusion variant. This is the same for the soft data fusion if α gets large.

3.4.1 Classification Results

The blue bar in figure 3.4 represents the result of using the information in the components of a CPD on the deviant tensor $\underline{\mathbf{X}}$. The deviant tensor contains a difference between the waveforms of each group and therefore it is expected to be able to detect

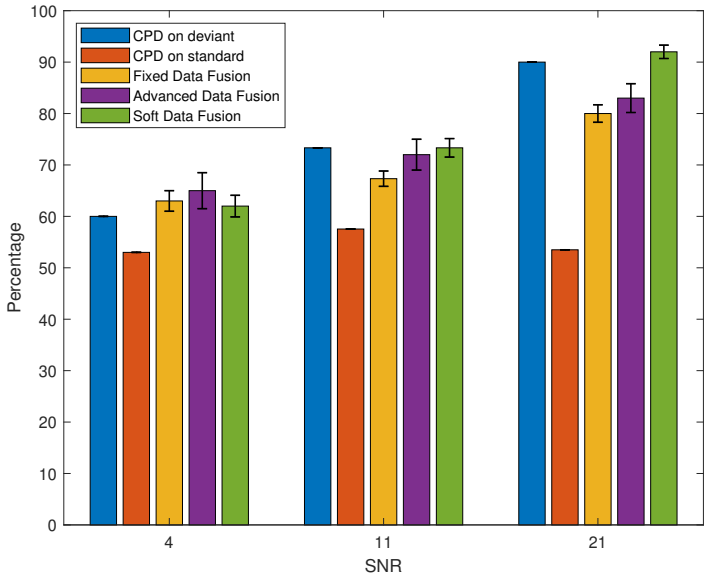
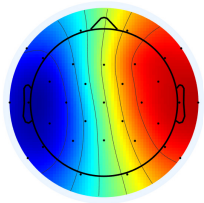


Figure 3.4: Classification accuracy for different SNR values if standard tensor has no group differences

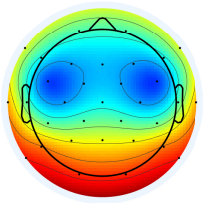
this depending on the SNR values. As can be seen in figure 3.4, the classification accuracy keeps going up when the SNR increases while the actual difference that is present is really small and easily masked by noise.

In contrast, the orange bar shows the result of the CPD on the standard tensor $\underline{\mathbf{Y}}$. The standard waveforms are the same in both groups so the accuracy is expected to be close to 50% due to the difference not being present. In figure 3.4 this behavior can be seen for all three SNR values. Both CPD results on the single tensors do not show any variation in the end result for all 10 iterations. As explained in chapter 2, this is because the CPD is generally unique up to scaling and permutation.

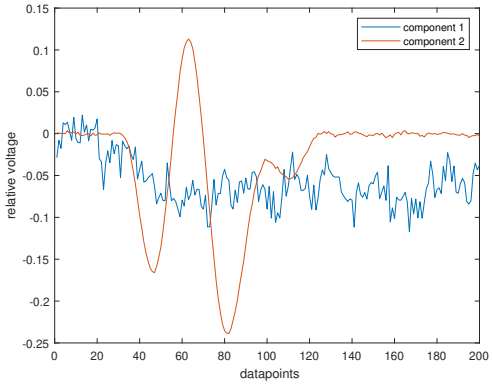
The results of the data fusion methods are represented by the yellow, purple, and green bars. For all three methods, there is a deviation in the classification accuracy present. This is because the data fusion methods are more complex and have more degrees of freedom. Especially advanced data fusion and soft data fusion because they also include unshared components. For a low SNR value, the data fusion results are similar to each other and are basically on the level of the CPD on the deviant tensor. When the SNR is increased, the difference between the data fusion methods gets more visible. For an SNR of 11dB and 21dB, the advanced data fusion and the soft data fusion methods perform better compared to the fixed data fusion. This is expected because there is no group difference present in the standard tensor. The fixed data fusion has trouble including the component that tells something about the group difference in the deviant tensor because it is hard to find a component with the same group distribution in the standard tensor. For ADF, this is easier because unshared components are allowed. Although the unshared component can be given a lower weight, ADF still needs to find a component with the same group difference in the standard tensor. The



(a) Scalp map from component 1



(b) Scalp map from component 2



(c) Waveform of components 1 and 2

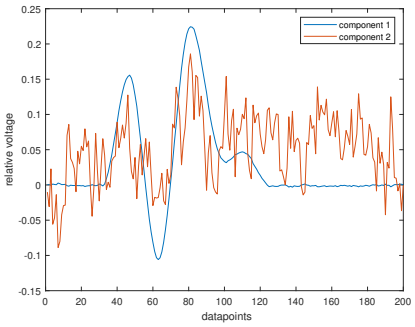
Figure 3.5: ERP information from all components of the CPD on the deviant tensor

SDF is more flexible since it has two different factor matrices instead of a fixed one.

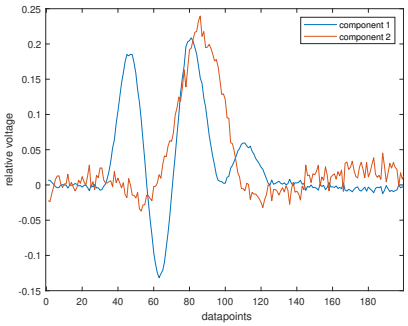
3.4.2 Component Interpretation

To show what the components of a decomposition look like and what information they contain. The components of the CPD on the deviant tensor with an SNR of 11dB are shown in figure 3.5. From table 3.1, it can be seen that the optimal number of components is two. In figure 3.5a and figure 3.5b, the scalp maps from the columns of factor matrix $\mathbf{C}$ can be seen. The first component tells something about noise that is present in all the subjects. However, the second component gives a source that is present in the exact brain regions that are selected for the ERP in BESA simulator. The highest classification accuracy is also achieved if k-means clustering is performed on the column of the second component in factor matrix $\mathbf{A}$. The two waveforms of the components can be seen in figure 3.5c. Component one is indeed related to some noise, while the second component gives the clean ERP back.

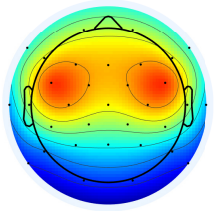
Another interesting decomposition result to look at is advanced data fusion. In table 3.1, the optimal number of components seems to be 4 with a β value of 0.01. This decomposition results in one statistically significant component that is illustrated in figure 3.6 together with a component that is not statistically significant (equally present in all subjects). The weights of the first component are similar while the weights of the second significant component are higher for the deviant tensor component compared to the standard tensor component. This means that the first component is



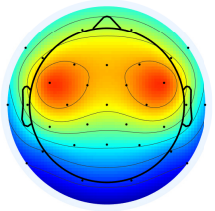
(a) Waveform of not statistically significant component and statistically significant component in standard tensor



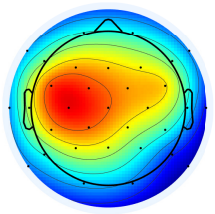
(b) Waveform of not statistically significant component and statistically significant component in deviant tensor



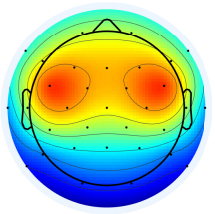
(c) Scalp map not statistically significant component standard tensor



(d) Scalp map not statistically significant component deviant tensor



(e) Scalp map statistically significant component standard tensor



(f) Scalp map statistically significant component deviant tensor

Figure 3.6: ERP information from a not statistically significant component and a statistically significant component. In other words, one component is present in all subjects while the other component tells something about the group differences.

shared amongst both tensors while the second significant component is unshared in the standard tensor (i.e. the group distribution in the A factor matrix column is present in the deviant tensor but not in the standard tensor). If the scalp maps are analyzed of the statistically significant component, the auditory cortex regions can be seen in the scalp map from the deviant tensor (figure 3.6f) while this is not the case for the scalp map in the standard tensor (figure 3.6e). The reason is that there is no ERP component present in the standard tensor that shares the same group difference as the deviant tensor. To still include this component, the decomposition tries to find some noise that shares the same group difference and gives that component a low weight.

If the waveform is analyzed of the statistically significant component, it can be seen that it is related to some noise in figure 3.6a. However, in figure 3.6b, there is a positive peak. This positive peak is related to the difference in the deviant waveform from both groups. If one deviant from each group would be subtracted, the result would be that peak which is exactly at the location the difference is present. This is actually similar to the MMN waveform discussed in chapter 1.

Performance for 10 iterations			
Algorithm	R	Accuracy (%)	Deviation (%)
CPD on $\underline{\mathbf{X}}$	1	66.67	0
	2	**73.33**	0
	3	73.33	0
	4	70.0	0
	5	66	5.48
CPD on $\underline{\mathbf{Y}}$	1	**57.54**	0
	2	56.67	0
	3	57.12	0
	4	56	1.49
	5	56.14	4.47
FDF	1	65.54	0
	2	**67.33**	1.49
	3	66.67	4.71
	4	65.33	3.80
	5	67.20	1.7
	6	64.67	2.98
ADF $\beta = 0.001$	2	68	2.98
	4	**69.33**	3.80
	8	68.67	1.49
ADF $\beta = 0.01$	1	66	1.49
	2	68	2.98
	3	69.45	2.35
	4	**72**	3.80
	5	70.33	4.7
	6	68.12	4.35
	8	68.67	1.49
ADF $\beta = 0.1$	2	**69.33**	1.49
	4	68.67	2.98
	8	64	4.35
SDF $\alpha = 0.001$	2	70.13	1.43
	4	**71.54**	2.23
	8	68.98	2.87
SDF $\alpha = 0.01$	1	66.67	0.6
	2	**73.33**	1.6
	3	73.33	1.78
	4	72	1.83
	5	70	2.36
	6	68.67	1.83
	8	69.33	2.79
SDF $\alpha = 0.1$	2	69.21	1.56
	4	**70.76**	2.21
	8	68.47	3.56

Table 3.1: Effect of different number of components on deviant and standard tensors with an SNR of 11dB for 10 iterations

3.5 Shared Difference

The second test case is that the group difference is also present in the standard tensor. If this is true, the data fusion methods are expected to work better compared to the single decompositions because there is extra information available during the decomposition. The components and variables of the decompositions are chosen the same way as in the previous section and the classification results can be seen in figure 3.7.

The blue and orange bars in figure 3.7 represent the CPD on the single tensors again. It can clearly be seen that the results of the CPD on the standard tensor are now on an equal level to the CPD on the deviant tensor. As the SNR increases, the classification accuracy gets higher because the difference is less masked by the noise and thus easier detectable.

The data fusion methods perform similarly for all SNR values since there are no more unshared differences in both tensors. By taking advantage of the extra information, the data fusion methods do perform better compared to the CPD on a single tensor. This is especially the case for higher SNR values. Although, there is still more deviation in the results of the data fusion methods due to the algorithms being more complex.

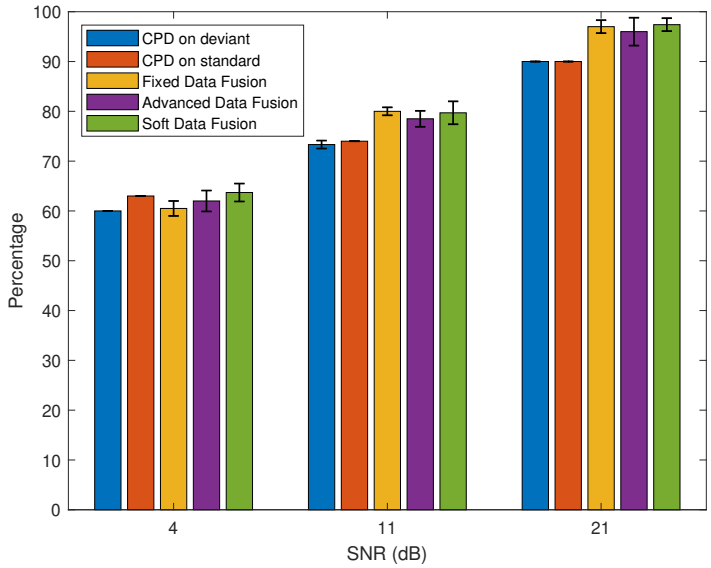


Figure 3.7: Classification accuracy for different SNR values if standard tensor has a group difference as well

3.6 Discussion

With the results of the simulations performed in the previous sections, a few questions can be answered. The most important one is that the data fusion methods perform better than the single decompositions when there is a shared difference in both tensors. This is to be expected since there is simply more information in the decomposition. However, the data fusion methods appear not to be as good if there is no shared

difference at all. The constraints are too hard to really capture the same statistically significant components as the single tensor decomposition on the tensor that has a group difference. Although, the ADF and SDF do come close if the SNR value is higher.

By adding the relaxation terms to the data fusion methods, more information on the actual component waveforms and scalp maps can be revealed. ADF and SDF allow for unshared components to be present to capture components that tell something about a group difference even though this is not present in the second tensor. In ADF, the weights give information about which tensor contains the group distribution that is present in the column of factor matrix $\mathbf{A}$. The deviation of the results is higher with the data fusion methods because there is more freedom in the number of components it can use. This might cause problems with data interpretability because the solution is not unique and has to be taken into account if these methods are used for biomarker extraction. Another difficulty with the ADF and SDF is the extra variable, however, the results do not change that much if a slightly different than optimal value is chosen. In the literature, the α and β values are often around 0.01 and this seems to behave pretty stable and optimal on the simulation data as well.

This chapter only showed two extreme cases, however, more scenarios have been tested:

- Added different levels of variation in the peaks for every patient in the same group
- Tested with different levels of shared differences in both tensors

The results of both tests were as expected and therefore not added to this report. The conclusion is that if a variation is added to the individual peaks of subjects in a group, the tensor methods have difficulties unraveling the group differences depending on the level of variation and the SNR. If the variation is low, the results are the same as shown in this chapter. However, if the variation is high, the actual difference between the two groups is masked and cannot be captured by the decompositions.

The conclusion of the second test is that by changing the percentage of shared group difference, the results move in between the two extreme cases shown in this chapter. If the two tensors have a shared difference but this difference is not exactly for the same patients, the results are more similar to the plot in figure 3.7. However, if the two tensors have only a few shared subjects with a group difference, the results are more comparable to figure 3.4.

4 Schizophrenia Detection

In the previous chapter, the different decompositions were tested on simulated ERP data. The results show that if a joint decomposition is used and the two datasets contain similar group information, the classification accuracy will be higher. However, the simulated data is really basic and simple. The question is if these methods still perform on a real dataset. Therefore, in this chapter, the proposed methods will be evaluated on an openly available dataset. In a study from Ford et. al [6], it was investigated whether schizophrenia can be diagnosed based on ERP data. This was done with ERP experiments on patients with schizophrenia (26) and healthy controls (22). After the publication, they released additional measurements for a total of 81 subjects. In this chapter, the bigger dataset of 81 subjects will be used for the evaluation of the tensor decompositions. The experiments performed on the subjects are as follows:

- Subjects press a button and a 1000Hz, 80dB sound is played
- Subjects passively listen to a 1000Hz, 80dB sound
- Subjects press a button and no sound is played

An example of the ERP waveform of experiment one can be seen in figure 4.1 where the ERP of subject 1 at electrode Fz is shown. At timepoint 0 the button is pressed and the sound is immediately played. After the measurements, Ford et. al. [6] analyzed a few features to see if there are differences in both groups. These features are the N1 peak, P2 peak, and readiness potential (RP). The RP is the average value of the EEG data in a timeframe before the actual stimulus is played. The conclusion is that on average there is a small difference between the two groups, however, the ERP measurements and the features cannot be used for any individual classification of whether a person has schizophrenia or not. If this is still tried on the bigger dataset of 81 subjects and with the features that are used in the paper, the classification results are not great and reflect the conclusion of the paper. A simple k-means clustering algorithm is used on the N1 peaks and the results can be seen in figure 4.2. The accuracy stays below 70 percent for the three main electrodes that are also used in the analysis from Ford et. al. [6].

The next section will explain how the proposed methods in this report will be used on the available dataset. After this, another section will show the results and evaluate the performance.

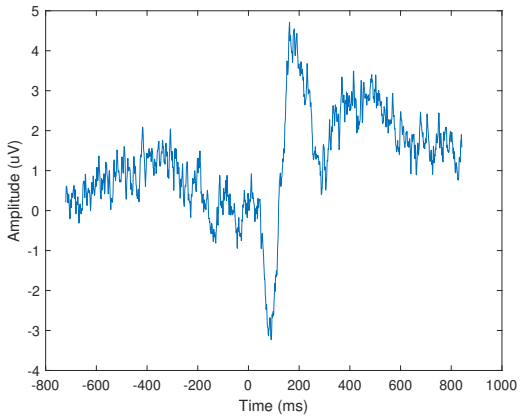


Figure 4.1: Example of ERP of subject 1 at electrode Fz

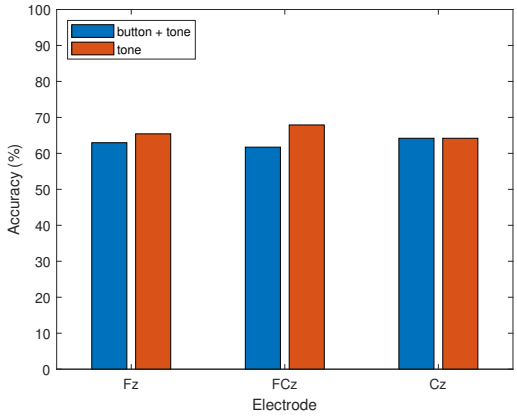


Figure 4.2: Classification accuracy of N1 peaks of two experiments at three main electrodes

4.1 Using Tensor Decompositions To Reveal Group Differences

To be able to use the data for the methods in this thesis, the data has to be rewritten in a tensor format (see figure 4.3) with dimensions subjects x time x electrodes. There are in total 81 subjects and 9 electrodes. Originally, the EEG was measured with 64 electrodes, however, the published dataset only includes 9 electrodes of the most important locations. The formulas in chapter 2 are for two tensors but can be rewritten easily to include a third tensor. However, for the sake of simplicity, consistency, and resemblance to the Erasmus data, tensor $\underline{\mathbf{Z}} \in \mathcal{R}^{81\times1700\times9}$ is not included in the analysis. Additionally, only in the first two measurements did the subjects actually listen to a tone and this is more related to the experiments of Erasmus MC as well.

The two tensors will be from experiment 1 $\underline{\mathbf{X}} \in \mathcal{R}^{81\times1700\times9}$ and experiment 2 $\underline{\mathbf{Y}} \in \mathcal{R}^{81\times1700\times9}$. The bigger dataset of 81 subjects will be used to evaluate the tensor methods and was released by the author of the original research after the initial publication. The data is present in an Excel file as an averaged ERP for each patient. This

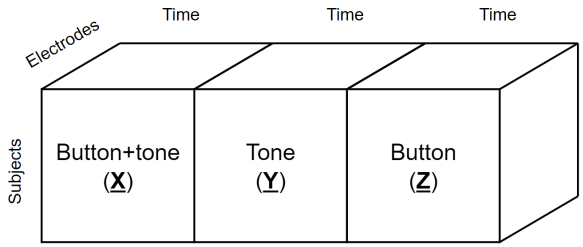


Figure 4.3: Tensors representations of the three experiments

means that pre-processing is already performed and includes the following main steps:

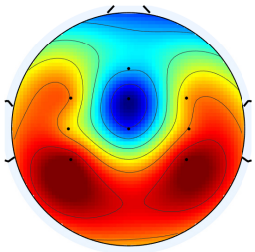
- 0.1Hz high-pass filter
- Interpolation of outlier channels in the continuous EEG data
- Get 3s epochs from continuous EEG data
- Baseline correction with time frame -100 to 0 ms
- Canonical Correlation Analysis to remove muscle artifacts and high-frequency white noise
- Rejection of outlier single trials and removal of outlier components from an ICA decomposition
- Interpolation of outlier channels within single trials
- Averaging remaining trials to get an averaged ERP for each patient at each electrode

The averaged ERP for each patient at every electrode is put into a tensor dataset for each experiment. The tensor is further processed by normalizing the horizontal slices of the tensor and standardizing both tensors by dividing it with their Frobenius norm.

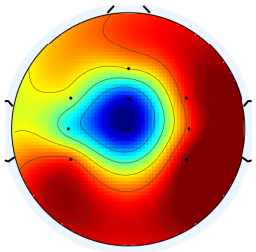
4.2 Tensor Decompositions On All Subjects

The next step is to see how a tensor decomposition performs in trying to see if there is a group difference present. In order to do that, firstly, the single tensors $\underline{\mathbf{X}}$ and $\underline{\mathbf{Y}}$ are decomposed with a CPD. The number of components is chosen by computing the decomposition for different component numbers R and checking when it performs the best. The results for 10 iterations can be seen in table 4.1 and this is the same procedure as in chapter 3. A low number of components seems to perform the best for classification accuracy and if a higher number is chosen, the variance also gets higher.

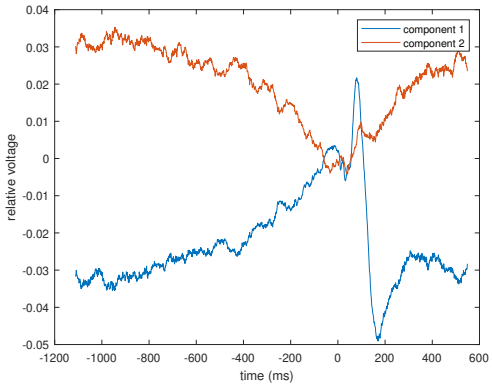
The components of the CPD of tensor $\underline{\mathbf{X}}$ with $R = 2$ can be seen in figure 4.4. Figure 4.4a and figure 4.4b contain the scalp maps of the two components. The scalp maps show some activity in the central and left/right motor regions in the brain which are also in the results of Ford et. al. [6]. Figure 4.4c contains the data over time from



(a) Scalp map from component 1



(b) Scalp map from component 2



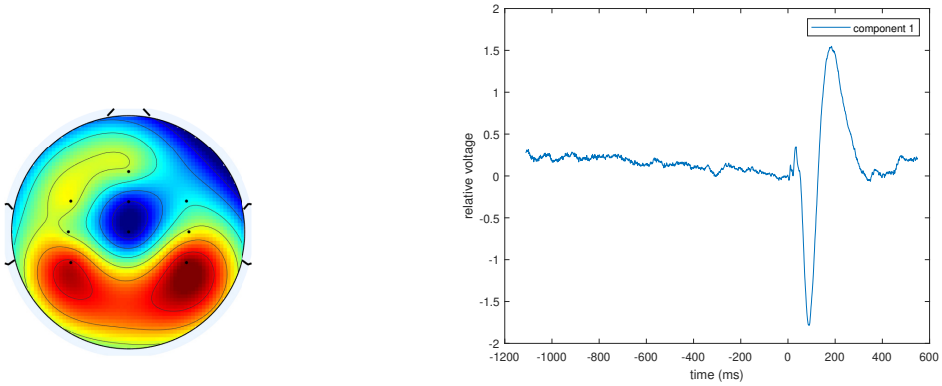
(c) Waveform of components 1 and 2

Figure 4.4: ERP information from all components of CPD on tensor $\underline{\mathbf{X}}$

the two components. The first component clearly contains the ERP response, while the second component also seems to show a really small ERP. All the possible combinations of the columns of factor matrix A are used for the K-mean clustering algorithm. The best classification result is given by using k-means clustering only on the first column. This means that the first component says more about the group difference than the second component does.

The CPD components of tensor $\underline{\mathbf{Y}}$ can be seen in figure 4.5. The best classification accuracy is achieved if only one component is used (see table 4.1). The data over time of the component is illustrated in figure 4.5b and is a cleaner version of the ERP itself. This might suggest that there is a group difference present in one of the ERP peaks between 0-400 ms and not necessarily in the data related to the readiness potential. The scalp map shows some activity in the central area of the brain but also in the left and right motor regions. This is similar to the scalp maps of the CPD on tensor $\underline{\mathbf{X}}$. The clustering result on the features in the column of the significant component of factor matrix $\mathbf{A}$ can be seen in figure 4.6. The two identified groups are illustrated in red and blue and the features are ordered from low to high.

In figure 4.7, the classification accuracies of using all the different decompositions for 10 iterations can be seen. The optimal number of components determined by the classification accuracy is shown in table 4.1. The single tensor decompositions get a 65 and 75 percent accuracy, however, when the joint decompositions are used, the values



(a) Scalp map of component 1

(b) Waveform of component 1

Figure 4.5: ERP information from all components of CPD on tensor $\underline{\mathbf{Y}}$

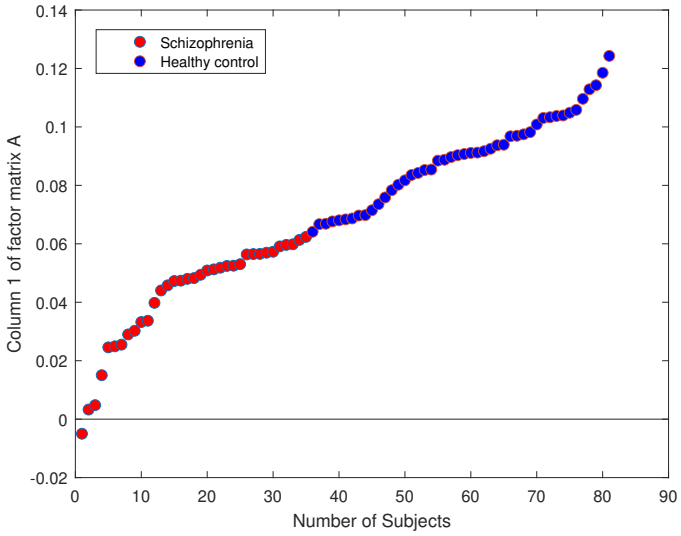


Figure 4.6: Results of clustering on features from component 1

stay in between the results of the single tensor decompositions at around 70 percent.

The bad performance of the data fusion methods could be due to the big difference between the two experiments. The subjects are actively doing something in the first experiment while they are only passively listening in the second. This can cause too many neurological differences because other parts of the brain are used. The group difference might not be shared in these two tensors due to the experiments being too different. In chapter 3, the data fusion methods were also in between the two CPD results on a single tensor when the group difference was not shared in both tensors. Another reason for the low accuracy could be that age has too much variety and it is well-known that ERPs also change due to age effects [66]. The decomposition then captures these age differences instead of the actual differences in the clinical symptoms.

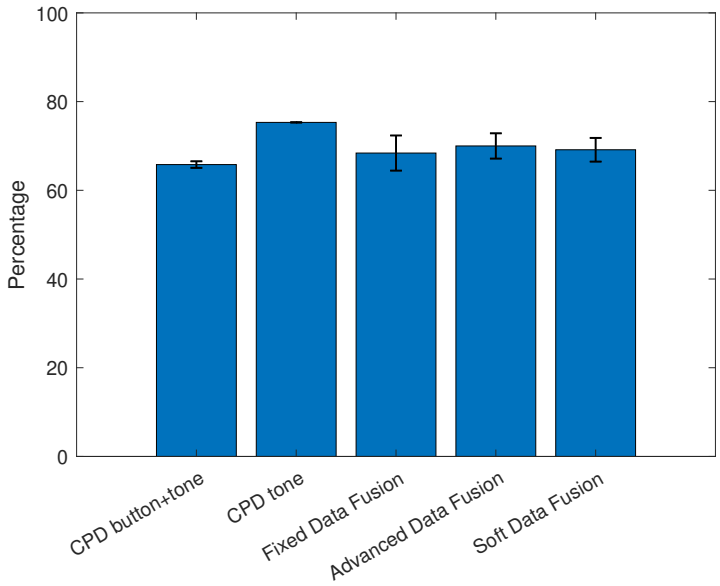


Figure 4.7: Classification accuracy of all decomposition methods

Performance for 10 iterations			
Algorithm	R	Accuracy (%)	Deviation (%)
CPD on $\underline{\mathbf{X}}$	1	59.26	0
	2	**65.80**	0.74
	3	64.12	1.42
	4	63.46	2.12
	5	61.23	1.67
CPD on $\underline{\mathbf{Y}}$	1	**75.31**	0
	2	74.81	0.64
	3	72.72	0.39
	4	68.40	3.41
	5	75.06	2.59
FDF	1	58.77	3.69
	2	63.70	3.20
	3	67.16	4.40
	4	**68.40**	3.96
	5	66.67	3.13
	6	65.80	3.14
ADF $\beta = 0.001$	2	65.92	1.86
	3	65.56	3.84
	4	66.54	2.99
	5	67.16	2.80
	6	68.53	3.03
	8	**70.43**	3.8
	10	69.24	3.85
ADF $\beta = 0.01$	3	67.65	3.32
	6	66.54	1.75
	8	**68.27**	3.24
ADF $\beta = 0.1$	3	66.17	3.20
	6	**69.81**	2.76
	8	67.14	3.63
SDF $\alpha = 0.001$	3	67.41	3.74
	6	**69.17**	3.42
	8	68.78	2.95
SDF $\alpha = 0.01$	2	67.53	2.85
	3	66.54	2.21
	4	67.16	1.95
	5	68.91	3.01
	6	**70.96**	2.93
	8	67.30	4.87
	10	67.64	2.67
SDF $\alpha = 0.1$	3	65.93	3.50
	6	**68.83**	3.46
	8	66.72	2.98

Table 4.1: Effect of different number of components on decompositions of all subjects for 10 iterations

4.3 Tensor Decompositions On Subjects Above 30

As shown in the previous section, the tensor decompositions do find some components that tell something about a possible group difference. However, the results of the data fusion methods were not as expected. This might be due to the big age difference between the subjects which causes too many differences in the ERP. To see if the big variation in age between the subjects really matters, the average ERP response of all age groups can be in figure 4.8. The ERP waveform does indeed change a lot for each age group and could mask the difference between the clinical symptoms. This scenario is very similar to adding a variation in the peaks of the individual subjects which was discussed in section 3.6.

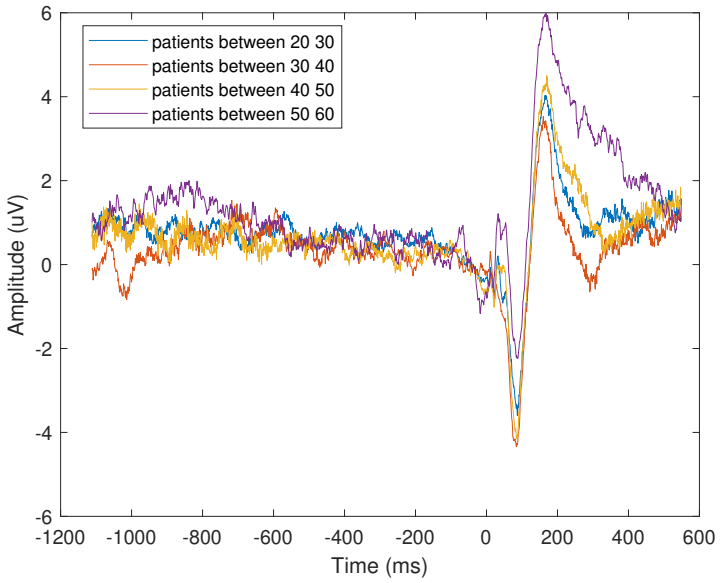


Figure 4.8: ERP differences for each age group

If an age group is removed, the classification accuracy is expected to go up. This can be tested by only including subjects above 30 and throwing away the ERP waveforms of subjects below 30. The dataset will now be containing 52 subjects instead of 81. The results of all the methods used on this 52-subject dataset can be seen in figure 4.9. The components and variables are chosen in the same way as in the previous section. The figure shows that the classification accuracy gets higher. The results of the CPD on the single experiments are now a value of around 70 and 81 percent. However, the data fusion methods stay between the two values of the single decomposition, which again suggests that the group differences are not shared. As discussed in the previous section as well, the problem is most likely that actively and passively listening to sounds gives too much neurological variation. Additionally, there are also different levels of schizophrenia and the dataset only contains binary labels. Although an assumption, it might be that subjects with a mild form of schizophrenia have an ERP waveform that is similar to the healthy control group when actively listening to sounds as opposed to passively listening.

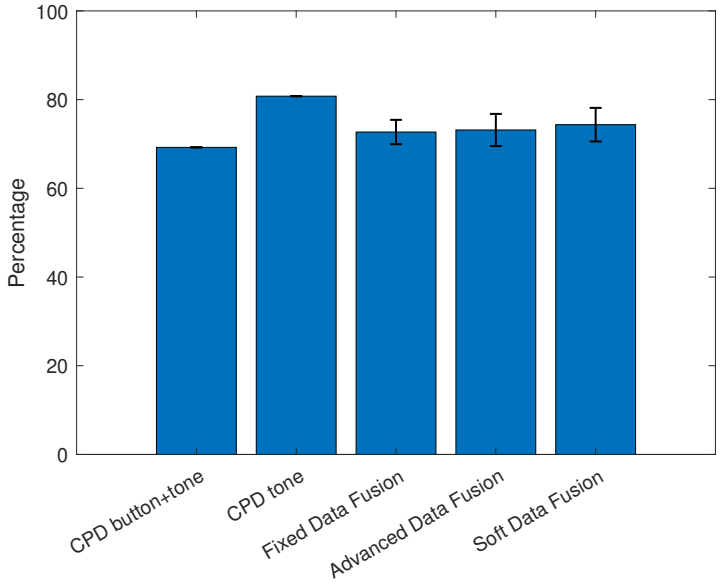
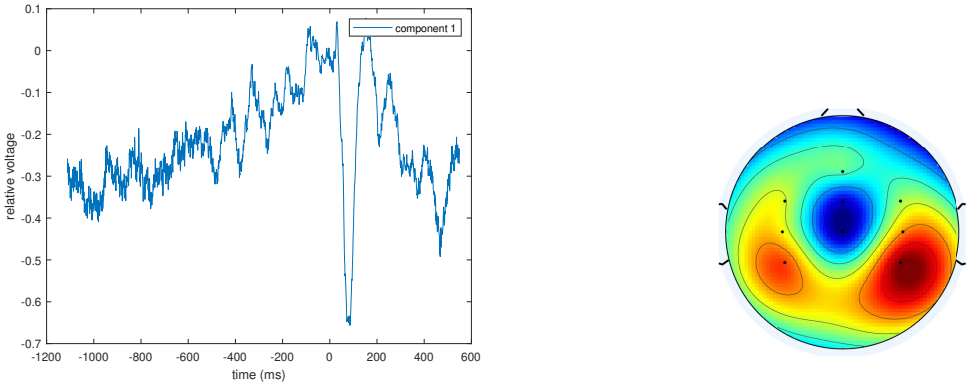


Figure 4.9: Classification accuracy of all decomposition methods on subjects above 30



(a) Waveform of component 1

(b) Scalp map of component 1

Figure 4.10: ERP information from all components of CPD on tensor $\underline{\mathbf{Y}}$

Because the accuracy increases, the quality of the biomarker also improves because there is a component that more accurately says something about the group difference present. The statistically significant component in the single decomposition of experiment two, or tensor $\underline{\mathbf{Y}}$, can be seen in Figure 4.10. It is interesting to see that from the entire ERP, only one peak is left, namely, the N1 peak. This could mean that the difference between the two groups is mostly present and related to this N1 peak and again shows the benefit of using the tensor decompositions. Not only is the classification accuracy relatively high, but it also contains information about the source of the actual difference. In the research of Ford et. al [6], they also found that on average, there is a difference related to N1 peak between schizophrenia subjects and healthy controls. The clustering result on the features in the column of the significant component of factor

matrix **A** can be seen in figure 4.11. The two identified groups are illustrated in red and blue and the features are ordered from low to high.

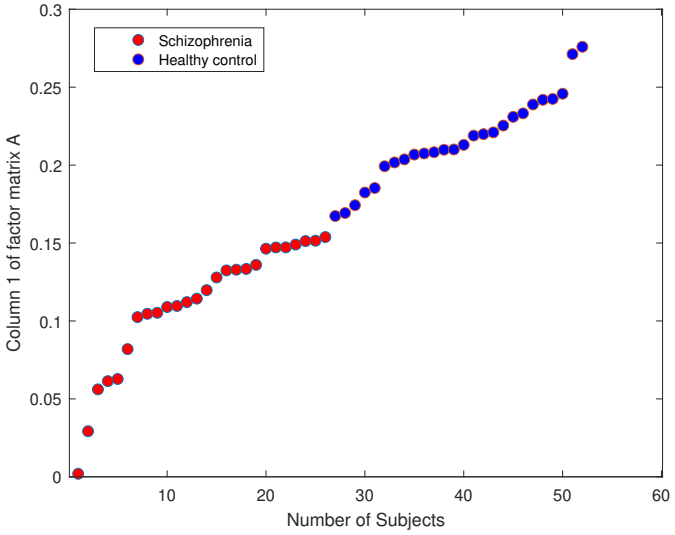


Figure 4.11: Results of clustering on features from component 1

4.4 Discussion

The data fusion methods do not seem to outperform the single tensor decompositions for this dataset. As discussed in the previous sections, this could be due to the neurological differences between the two experiments. Additionally, the various levels of schizophrenia might respond differently to one experiment compared to the other. However, this simulation does show that tensor decompositions can indeed be used for getting a higher classification accuracy compared to the methods used in the original paper. When the age group of 20-30 is removed, the results improve because there is less age-related variety present in the ERP. The tensor decomposition can now focus better on the actual difference between the clinical symptoms. This results in a component that tells something about the N1 peak which is also a possible difference between the groups that was found in the paper of Ford et. al. [6]. The simulation in chapter 3 also covered a scenario where there is a variation in the peaks for every subject which masks the actual clinical difference. If this variation is made smaller, by removing an age group, the decompositions can better unravel the difference between the two groups.

There are some things that could make the results better than they are now. Currently, the decomposition only uses the data over time for every electrode and subject, however, as discussed in chapter 2, a common way to analyze EEG/ERP data is by looking at the frequency domain. The tensor can be expanded or replaced by performing a wavelet transform on the ERP data to see if there are differences in frequency components between the subjects. Empirical Mode Decomposition can also be used to build a tensor and see if it adds an increase in classification accuracy.

These scenarios were actually tested to see if this would give better results, however, after creating these tensors, the results were not higher than shown in this chapter. This can still be looked into and optimized since the quick tests performed were not very thorough. Another possible improvement that was tested was instead of doing a joint tensor decomposition, a joint matrix-tensor decomposition was performed. The created tensor was still from experiment 2 but the matrix was created from some demographic data including age, gender, and education level. This problem is basically a CMTF that was discussed in chapter 2 but did not give better results than the one obtained with the joint tensor decompositions.

A combination with machine learning could also be a good solution. The features from the tensor decomposition can be combined with other features such as demographic information, frequency band power, or statistical information to get a higher classification accuracy. There are already a few papers that have done some machine learning on this dataset and the highest they get is around a 90% accuracy [43], [67]–[72]. This is higher than the tensor decomposition, however, it is often not a fair comparison due to the number of features that are used for those machine learning algorithms. Some of these machine learning methods use features from wavelet transform, empirical mode decomposition, or a lot of statistical features. The downside of most of these methods is that the focus is mostly on getting a high classification accuracy and not on identifying what the source of the difference actually is. They also often lack any explanation of why certain features are used.

Another improvement could be by checking if the performance increases with using more flexible tensor decompositions, such as tucker decomposition, block term decomposition, or higher order singular value decomposition. The main reason for the use of CPD was the simplicity and interpretability of the components. However, the assumption that the data can be perfectly represented as a sum of rank-one-tensors might not hold. Although, there is no direct proof from the components that this might be the case. The flexible decompositions can be tested to see if they can capture the group differences better in terms of classification accuracy and interpretability of the components.

5 Discussion and Future Work

This chapter will contain a discussion to talk in more detail about the results from chapter 3 and chapter 4. After the discussion, some possible directions for future work will be explained.

Discussion

Tensor decompositions seem a promising method to extract biomarkers from ERP data. In chapter 3, it was shown that by using tensor decompositions, small differences between two groups can be detected. The decomposition results in components that can be used for the classification of each subject. Additionally, the components give more information about what the cause of this difference might be.

The data fusion methods are a way of combining more information to achieve better results and get extra information about the underlying neurological sources that cause the differences. In the simulations from chapter 3, it was shown that the data fusion methods can outperform the single decompositions if the additional dataset has similar differences between the two groups. If this is not the case, the data fusion methods are more of a compromise between the single decomposition results. For biomarker extraction, however, the data fusion methods are less reliable due to the complexity of the methods and more deviation in the results.

The initial simulations were to show the benefit that data fusion methods can have in comparison to single decompositions. It was also a way to test the methods on a similar dataset from Erasmus MC, namely, two tensors with ERP data. In the literature, data fusion methods have of course been tested on all kinds of data, however, the combination of two ERP tensors is not common at all. To the authors' knowledge, these data fusion methods have never even been performed on two tensors that contain ERP data.

Using data fusion methods opens up more possibilities and is not limited to only combining two ERP tensors. As explained in chapter 2, data fusion can also be done with MRI data or even some demographic data, psychological score data, or movement score data. The psychological score and movement score data is also measured in the Child Brain Lab and although currently the dataset is really small, this could definitely be an option in the future.

The simulations on the Schizophrenia dataset in chapter 4 were to show if tensor decomposition could work on a real dataset instead of only the simple artificial data from chapter 3. Additionally, this experiment was to see how the data fusion of two ERP tensors would work on actual data. Was it better than the single decompositions and comparable to some scenarios from chapter 3? Would the tensor decompositions even get better results than in the original paper? The results show that by using tensor decompositions, the classification results are significantly higher than only using

amplitudes and delays. Sadly, the data fusion methods do not perform better than the single decomposition of experiment two. However, due to the simulations performed in chapter 3, the reason why this happens can be explained. The data fusion classification results end up in between the results of the single decompositions if the shared group difference is not present. This most likely means that the variation of the response between passively and actively listening to sound is too much. Age-related ERP differences also affect the result and can mask the actual difference in the clinical symptoms. By removing an age group, the classification results are higher and give more detailed information about what the difference between the two groups is. The statistically significant component is related to the N1 peak which was also found as a potential biomarker in the original paper. However, only using this amplitude does not give high classification accuracy.

The tensor decomposition methods allow for the extraction of more comprehensive information from ERPs and can facilitate the identification of biomarkers for distinguishing different groups. Next to that, it is shown how to apply these techniques to the data from Erasmus MC theoretically in chapter 2 and experimentally during the two simulations.

Future Work

This section will contain possible ways to improve the methods used in this report, Throughout the different chapters, some explanation about possible future work is already given but this section just summarizes all of this to give a better overview.

More information in tensor decompositions

In the simulations, a 3D tensor is created with dimensions subjects x time x electrodes. However, in the introduction about ERP analysis methods, different techniques were discussed. Some of these techniques transformed the data to the frequency domain or used empirical mode decomposition to get more information on oscillations present in the data. By adding dimensions to the existing tensors or creating new tensors this extra information can be exploited, and perhaps, the resulting components can describe the biomarkers more accurately compared to only using spatial and temporal information.

In addition to that, the data fusion methods can be changed from having two ERP tensors to adding different data types. In Erasmus MC, the MRI of children with a neurological disorder is often already measured. The data fusion methods can be expanded with an extra MRI matrix as we have seen before [60]. During the Child Brain Lab measurements, the subjects also get psychological tests and movement tests. From both of these results, score matrices can be built to include in the data fusion. An example of this can be seen from Kinnet et. al. [73], who use psychological tests to create a patient x score matrix and combine this with an EEG tensor. Tensor decompositions and data fusion are used to estimate development scores for 22 children under 11. So there are lots of opportunities to exploit the data fusion algorithms compared to only using the ERP tensors.

Different tensor decompositions

As mentioned earlier in this report as well, CPD is mainly chosen due to the complexity and component interpretability. However, the assumption that the data can be perfectly estimated with a sum of rank-1 tensors might not hold for some more complex data. During the simulations, no direct evidence has been found that the CPD underperforms. Although, more flexible methods might perform better in isolating biomarkers related to group differences. Examples of these methods are tucker decomposition, block-term decomposition, and higher-order singular value decomposition. The flexible decompositions can be tested to see if they can capture the group differences better in terms of classification accuracy and interpretability of the components.

Better machine learning algorithm

For the classification, a k-means clustering algorithm has been used so far which is a really basic clustering algorithm. If the Erasmus MC dataset grows, the features from the tensor decomposition can potentially be combined with other features commonly used in ERP analysis to exploit more information. This might result in better classification results.

6 Conclusion

Brain disorders in children pose significant challenges to their development and can affect various aspects such as cognition, speech, movement, and behavior. The lack of clear prognostic information at the time of diagnosis leaves parents and children with numerous uncertainties about the future. The Child Barin Lab at Erasmus MC Sophia Children's Hospital is contributing to this research field by conducting IQ, EEG, speech, and movement tests in playful environments to obtain valuable data on these disorders. This multidimensional approach enhances scientific research and healthcare practices, leading to a better understanding of disease progression and its consequences.

The Otolaryngology department focuses on auditory-related potentials (ERPs) obtained from the EEG measurement in the Child Brain Lab. Their aim is to predict the future development of children with brain disorders using ERPs, particularly focusing on the Mismatch Negativity (MMN) and the Acoustic Change Complex (ACC) paradigms. By analyzing the ERP data from these experiments, they seek to unravel crucial insights into development trajectories and connections between hearing, language, and brain development.

However, the conventional approach of examining peak amplitudes and latencies of ERPs as biomarkers limits the utilization of the complete information stored within ERPs. To overcome this limitation, the primary objective of the thesis is to explore alternative methodologies that can extract more comprehensive information from ERPs and identify accurate biomarkers for distinguishing between different neurological profiles.

In this report, tensor decompositions have been utilized to exploit the structural information that is naturally present in the data which cannot be achieved with matrix decompositions. Specifically, tensor data fusion methods have been tested to combine multiple datasets for one tensor decomposition. First of all, it is concluded that, by using tensor decompositions, ERP data can be decomposed into different components that can be used for classifying patients and biomarker information. By using data fusion methods, data with a shared subject dimension can give higher classification results and give more insights into the underlying phenomena that cause group differences. By performing simulations on artificial ERP data, it has been shown that the different data fusion methods perform better on two ERP tensors compared to a single tensor decomposition. However, this is only the case if the group difference is shared among both datasets. It also has been shown on a real dataset that using tensor decompositions has the potential to classify subjects by using auditory event-related potentials. Additionally, it provides a better understanding of which neurological sources cause the difference.

In short, tensor decompositions and data fusion methods show great potential for analyzing ERP data. With the help of this report, an alternative method, as opposed to the conventional analysis methods, has been proposed.

Bibliography

[1] M. Aghabeig, B. Bałaj, J. Dreszer, *et al.*, "Perception of non-native phoneme contrasts in 8-13 months infants: Tensor-based analysis of eeg signals," in *2019 27th European Signal Processing Conference (EUSIPCO)*, 2019, pp. 1–5. DOI: 10.23919/EUSIPCO.2019.8903174.

[2] R. Naatanen, *Attention and brain function*. Hillsdale, NJ: Erlbaum, 1992.

[3] J. M. Ostroff, B. A. Martin, and A. Boothroyd, "Cortical evoked response to acoustic change within a syllable," *Ear and hearing*, vol. 19, no. 4, pp. 290–297, 1998.

[4] B. A. Martin and A. Boothroyd, "Cortical, auditory, event-related potentials in response to periodic and aperiodic stimuli with the same spectral envelope," *Ear and hearing*, vol. 20, no. 1, pp. 33–44, 1999.

[5] S. J. Luck, D. H. Mathalon, B. F. O'Donnell, *et al.*, "A roadmap for the development and validation of event-related potential biomarkers in schizophrenia research," *Biological psychiatry*, vol. 70, no. 1, pp. 28–34, 2011.

[6] J. M. Ford, V. A. Palzes, B. J. Roach, and D. H. Mathalon, "Did i do that? abnormal predictive processes in schizophrenia when button pressing to deliver a tone," *Schizophrenia bulletin*, vol. 40, no. 4, pp. 804–812, 2014.

[7] K. P. Martin BA Tremblay KL, "Speech evoked potentials: From the laboratory to the clinic," Jun. 2008, pp. 285–313. DOI: 10.1097/aud.0b013e3181662c0e.

[8] S. J. Luck, *An introduction to the event-related potential technique*. MIT press, 2014.

[9] K. Joos, A. Gilles, P. Van de Heyning, D. De Ridder, and S. Vanneste, "From sensation to percept: The neural signature of auditory event-related potentials," *Neuroscience & Biobehavioral Reviews*, vol. 42, pp. 148–156, 2014.

[10] M. I. Garrido, J. M. Kilner, K. E. Stephan, and K. J. Friston, "The mismatch negativity: A review of underlying mechanisms," *Clinical neurophysiology*, vol. 120, no. 3, pp. 453–463, 2009.

[11] X. Jiang, G.-B. Bian, and Z. Tian, "Removal of artifacts from eeg signals: A review," *Sensors*, vol. 19, no. 5, p. 987, 2019.

[12] M. Mennes, H. Wouters, B. Vanrumste, L. Lagae, and P. Stiers, "Validation of ica as a tool to remove eye movement artifacts from eeg/erp," *Psychophysiology*, vol. 47, no. 6, pp. 1142–1150, 2010.

[13] I. Winkler, S. Debener, K.-R. Müller, and M. Tangermann, "On the influence of high-pass filtering on ica-based artifact reduction in eeg-erp," in *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, IEEE, 2015, pp. 4101–4105.

[14] F. Zhao, C. Yang, and Y. Zheng, "The n270 in facial s1-s2 paradigm as a biomarker for children with attention-deficit/hyperactivity disorder," *IEEE Access*, vol. 8, pp. 88 969–88 976, 2020.

[15] A. T. White, R. B. Merino, S. Hardin, and S. Kim, "Non-invasive, cost-effective, early diagnosis of mild cognitive impairment in an outpatient setting: Pilot

study," in *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, IEEE, 2018, pp. 13–16.

[16] E. Cardon, I. Joossen, H. Vermeersch, *et al.*, "Systematic review and meta-analysis of late auditory evoked potentials as a candidate biomarker in the assessment of tinnitus," *PloS one*, vol. 15, no. 12, e0243785, 2020.

[17] F. Tenssay and H. Wang, "Analysis of eeg signals during visual processing: An erp study," in *2019 IEEE International Conference on Signal Processing, Communications and Computing (ICSPCC)*, IEEE, 2019, pp. 1–5.

[18] R. Nawaz, H. Nisar, and Y. V. Voon, "Analyses of stroop eeg using event related potentials and time-frequency decomposition," in *2020 IEEE-EMBS Conference on Biomedical Engineering and Sciences (IECBES)*, IEEE, 2021, pp. 401–405.

[19] J. Hine and S. Debener, "Late auditory evoked potentials asymmetry revisited," *Clinical Neurophysiology*, vol. 118, no. 6, pp. 1274–1285, 2007.

[20] S. Waninger, C. Berka, A. Meghdadi, *et al.*, "Event-related potentials during sustained attention and memory tasks: Utility as biomarkers for mild cognitive impairment," *Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring*, vol. 10, pp. 452–460, 2018.

[21] D. De Venuto, V. F. Annese, and G. Mezzina, "Remote neuro-cognitive impairment sensing based on p300 spatio-temporal monitoring," *IEEE Sensors Journal*, vol. 16, no. 23, pp. 8348–8356, 2016.

[22] M.-P. Hosseini, A. Hosseini, and K. Ahi, "A review on machine learning for eeg signal processing in bioengineering," *IEEE reviews in biomedical engineering*, vol. 14, pp. 204–218, 2020.

[23] J.-P. Kim and S.-W. Lee, "Time domain eeg analysis for evaluating the effects of driver's mental work load during simulated driving," in *2017 5th International Winter Conference on Brain-Computer Interface (BCI)*, IEEE, 2017, pp. 79–80.

[24] M.-H. Lee, J. Williamson, D.-O. Won, S. Fazli, and S.-W. Lee, "A high performance spelling system based on eeg-eog signals with visual feedback," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 26, no. 7, pp. 1443–1459, 2018.

[25] B. Blankertz, S. Lemm, M. Treder, S. Haufe, and K.-R. Müller, "Single-trial analysis and classification of erp components—a tutorial," *NeuroImage*, vol. 56, no. 2, pp. 814–825, 2011.

[26] Y. Zhang, G. Zhou, Q. Zhao, J. Jin, X. Wang, and A. Cichocki, "Spatial-temporal discriminant analysis for erp-based brain-computer interface," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 21, no. 2, pp. 233–243, 2013.

[27] J. Ma, Y. Zhang, A. Cichocki, and F. Matsuno, "A novel eog/eeg hybrid human–machine interface adopting eye movements and erps: Application to robot control," *IEEE Transactions on Biomedical Engineering*, vol. 62, no. 3, pp. 876–889, 2014.

[28] T. Shimogasa and H. Higa, "Toward a control of a mouse pointer: Eeg analysis for feature extraction," in *2014 14th International Conference on Intelligent Systems Design and Applications*, IEEE, 2014, pp. 151–155.

[29] H. Cecotti and A. Graser, "Convolutional neural networks for p300 detection with application to brain-computer interfaces," *IEEE transactions on pattern analysis and machine intelligence*, vol. 33, no. 3, pp. 433–445, 2010.

[30] S. Khatun, B. I. Morshed, and G. M. Bidelman, "Single channel eeg time-frequency features to detect mild cognitive impairment," in *2017 IEEE International Symposium on Medical Measurements and Applications (MeMeA)*, IEEE, 2017, pp. 437–442.

[31] S. Khatun, B. I. Morshed, and G. M. Bidelman, "A single-channel eeg-based approach to detect mild cognitive impairment via speech-evoked brain responses," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 27, no. 5, pp. 1063–1070, 2019.

[32] P. Williams, A. White, R. B. Merino, S. Hardin, J. C. Mizelle, and S. Kim, "Facial recognition task for the classification of mild cognitive impairment with ensemble sparse classifier," in *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, IEEE, 2019, pp. 2242–2245.

[33] J. Onton, M. Westerfield, J. Townsend, and S. Makeig, "Imaging human eeg dynamics using independent component analysis," *Neuroscience & biobehavioral reviews*, vol. 30, no. 6, pp. 808–822, 2006.

[34] T.-P. Jung, S. Makeig, M. Westerfield, J. Townsend, E. Courchesne, and T. J. Sejnowski, "Removal of eye activity artifacts from visual event-related potentials in normal and clinical subjects," *Clinical Neurophysiology*, vol. 111, no. 10, pp. 1745–1758, 2000.

[35] T.-P. Jung, S. Makeig, C. Humphries, *et al.*, "Removing electroencephalographic artifacts by blind source separation," *Psychophysiology*, vol. 37, no. 2, pp. 163–178, 2000.

[36] S. Makeig, M. Westerfield, T.-P. Jung, *et al.*, "Dynamic brain sources of visual evoked responses," *Science*, vol. 295, no. 5555, pp. 690–694, 2002.

[37] I. Kalyakin, N. González, and H. Lyytinen, "Extraction of the mismatch negativity on two paradigms using independent component analysis," in *2008 21st IEEE International Symposium on Computer-Based Medical Systems*, IEEE, 2008, pp. 59–64.

[38] T. Sugi, K. Kimura, S. Nishida, *et al.*, "Feature extraction for single trial record of visual mismatch negativity by use of independent component analysis," in *2007 IEEE/ICME International Conference on Complex Medical Engineering*, IEEE, 2007, pp. 1458–1462.

[39] F. Cong, I. Kalyakin, T. Ristaniemi, and H. Lyytinen, "How many single trials' eeg recordings are enough to extract children's brain activity of mismatch negativity by independent component analysis?" In *2011 7th International Symposium on Image and Signal Processing and Analysis (ISPA)*, IEEE, 2011, pp. 546–551.

[40] N. Zhuang, Y. Zeng, L. Tong, C. Zhang, H. Zhang, B. Yan, *et al.*, "Emotion recognition from eeg signals using multidimensional information in emd domain," *BioMed research international*, vol. 2017, 2017.

[41] N. E. Huang, Z. Shen, S. R. Long, *et al.*, "The empirical mode decomposition and the hilbert spectrum for nonlinear and non-stationary time series analysis,"

Proceedings of the Royal Society of London. Series A: mathematical, physical and engineering sciences, vol. 454, no. 1971, pp. 903–995, 1998.

[42] H.-C. Chang, P.-L. Lee, M.-T. Lo, Y.-T. Wu, K.-W. Wang, and G.-Y. Lan, "Inter-trial analysis of post-movement beta activities in eeg signals using multivariate empirical mode decomposition," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 21, no. 4, pp. 607–615, 2013.

[43] S. Siuly, S. K. Khare, V. Bajaj, H. Wang, and Y. Zhang, "A computerized method for automatic detection of schizophrenia using eeg signals," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 28, no. 11, pp. 2390–2400, 2020.

[44] H. Gandhi, D. Green, J. Kounios, C. M. Clark, and R. Polikar, "Stacked generalization for early diagnosis of alzheimer's disease," in *2006 International Conference of the IEEE Engineering in Medicine and Biology Society*, IEEE, 2006, pp. 5350–5353.

[45] R. Q. Quiroga, O. Sakowitz, E. Basar, and M. Schürmann, "Wavelet transform in the analysis of the frequency composition of evoked potentials," *Brain Research Protocols*, vol. 8, no. 1, pp. 16–24, 2001.

[46] T. Begum, F. Reza, A. Ahmed, S. Elaina, and J. Abdullah, "Delta signal in high educational level in auditory oddball paradigm-a wavelet study," in *2011 4th International Congress on Image and Signal Processing*, IEEE, vol. 5, 2011, pp. 2725–2728.

[47] G. Jacques, J. L. Frymiare, J. Kounios, C. Clark, and R. Polikar, "Multiresolution analysis for early diagnosis of alzheimer's disease," in *The 26th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, IEEE, vol. 1, 2004, pp. 251–254.

[48] C. Cecere, C. Corrado, and R. Polikar, "Diagnostic utility of eeg based biomarkers for alzheimer's disease," in *2014 40th Annual Northeast Bioengineering Conference (NEBEC)*, IEEE, 2014, pp. 1–2.

[49] N. Stepenosky, D. Green, J. Kounios, C. M. Clark, and R. Polikar, "Majority vote and decision template based ensemble classifiers trained on event related potentials for early diagnosis of alzheimer's disease," in *2006 IEEE International Conference on Acoustics Speech and Signal Processing Proceedings*, IEEE, vol. 5, 2006, pp. V–V.

[50] M. Ahiskali, D. Green, J. Kounios, C. M. Clark, and R. Polikar, "Erp based decision fusion for ad diagnosis across cohorts," in *2009 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, IEEE, 2009, pp. 2494–2497.

[51] B. A. Balut, M. T. Karnick, D. Green, J. Kounios, C. M. Clark, and R. Polikar, "Ensemble based data fusion from parietal region event related potentials for early diagnosis of alzheimer's disease," in *2007 International Joint Conference on Neural Networks*, IEEE, 2007, pp. 2409–2414.

[52] F. Cong, Q.-H. Lin, L.-D. Kuang, X.-F. Gong, P. Astikainen, and T. Ristaniemi, "Tensor decomposition of eeg signals: A brief review," *Journal of neuroscience methods*, vol. 248, pp. 59–69, 2015.

[53] F. Cong, A.-H. Phan, P. Astikainen, *et al.*, "Multi-domain feature extraction for small event-related potentials through nonnegative multi-way array decomposition from low dense array eeg," *International journal of neural systems*, vol. 23, no. 02, p. 1 350 006, 2013.

[54] H. Lee, Y.-D. Kim, A. Cichocki, and S. Choi, "Nonnegative tensor factorization for continuous eeg classification," *International journal of neural systems*, vol. 17, no. 04, pp. 305–317, 2007.

[55] D. Wang, Y. Zhu, T. Ristaniemi, and F. Cong, "Extracting multi-mode erp features using fifth-order nonnegative tensor decomposition," *Journal of neuroscience methods*, vol. 308, pp. 240–247, 2018.

[56] M. Aghabeig, B. Bałaj, J. Dreszer, *et al.*, "Perception of non-native phoneme contrasts in 8-13 months infants: Tensor-based analysis of eeg signals," in *2019 27th European Signal Processing Conference (EUSIPCO)*, IEEE, 2019, pp. 1–5.

[57] S. Pouryazdian, A. Chang, D. J. Bosnyak, L. J. Trainor, S. Beheshti, and S. Krishnan, "Multi-domain feature selection in auditory mismatch negativity via parafac-based template matching approach," in *2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, IEEE, 2016, pp. 1603–1607.

[58] F. Cong, A. H. Phan, Q. Zhao, *et al.*, "Benefits of multi-domain feature of mismatch negativity extracted by non-negative tensor factorization from eeg collected by low-density array," *International journal of neural systems*, vol. 22, no. 06, p. 1 250 025, 2012.

[59] D. Lahat, T. Adali, and C. Jutten, "Multimodal data fusion: An overview of methods, challenges, and prospects," *Proceedings of the IEEE*, vol. 103, no. 9, pp. 1449–1477, 2015.

[60] E. Acar, C. Schenker, Y. Levin-Schwartz, V. D. Calhoun, and T. Adali, "Unraveling diagnostic biomarkers of schizophrenia through structure-revealing fusion of multi-modal neuroimaging data," *Frontiers in neuroscience*, vol. 13, p. 416, 2019.

[61] A. Cichocki, N. Lee, I. Oseledets, A.-H. Phan, Q. Zhao, D. P. Mandic, *et al.*, "Tensor networks for dimensionality reduction and large-scale optimization: Part 1 low-rank tensor decompositions," *Foundations and Trends® in Machine Learning*, vol. 9, no. 4-5, pp. 249–429, 2016.

[62] B. Hunyadi, W. Van Paesschen, M. De Vos, and S. Van Huffel, "Fusion of electroencephalography and functional magnetic resonance imaging to explore epileptic network activity," in *2016 24th European Signal Processing Conference (EUSIPCO)*, IEEE, 2016, pp. 240–244.

[63] C. Chatzichristos, S. Van Eyndhoven, E. Kofidis, and S. Van Huffel, "Coupled tensor decompositions for data fusion," in *Tensors for data processing*, Elsevier, 2022, pp. 341–370.

[64] E. Acar, E. E. Papalexakis, G. Gürdeniz, *et al.*, "Structure-revealing data fusion," *BMC bioinformatics*, vol. 15, no. 1, pp. 1–17, 2014.

[65] T. Frodl-Bauch, N. Kathmann, H.-J. Möller, and U. Hegerl, "Dipole localization and test-retest reliability of frequency and duration mismatch negativity generator processes," *Brain Topography*, vol. 10, pp. 3–8, 1997.

[66] R. E. Dustman, R. Y. Emmerson, R. Ruhling, *et al.*, "Age and fitness effects on eeg, erps, visual sensitivity, and cognition," *Neurobiology of aging*, vol. 11, no. 3, pp. 193–200, 1990.

[67] S. K. Khare, V. Bajaj, and U. R. Acharya, "Spwvd-cnn for automated detection of schizophrenia patients using eeg signals," *IEEE Transactions on Instrumentation and Measurement*, vol. 70, pp. 1–9, 2021.

[68] M. Baygin, O. Yaman, T. Tuncer, S. Dogan, P. D. Barua, and U. R. Acharya, "Automated accurate schizophrenia detection system using collatz pattern technique with eeg signals," *Biomedical Signal Processing and Control*, vol. 70, p. 102 936, 2021.

[69] L. Zhang, "Eeg signals classification using machine learning for the identification and diagnosis of schizophrenia," in *2019 41st annual international conference of the ieee engineering in medicine and biology society (EMBC)*, IEEE, 2019, pp. 4521–4524.

[70] S. K. Khare, V. Bajaj, S. Siuly, and G. Sinha, "Classification of schizophrenia patients through empirical wavelet transformation using electroencephalogram signals," in *Modelling and Analysis of Active Biopotential Signals in Healthcare, Volume 1*, IOP Publishing, 2020.

[71] S. K. Khare and V. Bajaj, "A hybrid decision support system for automatic detection of schizophrenia using eeg signals," *Computers in Biology and Medicine*, vol. 141, p. 105 028, 2022.

[72] S. K. Khare and V. Bajaj, "A self-learned decomposition and classification model for schizophrenia diagnosis," *Computer Methods and Programs in Biomedicine*, vol. 211, p. 106 450, 2021.

[73] E. Kinney-Lang, A. Ebied, B. Auyeung, R. F. Chin, and J. Escudero, "Introducing the joint eeg-development inference (jedi) model: A multi-way, data fusion approach for estimating paediatric developmental scores via eeg," *IEEE transactions on neural systems and rehabilitation engineering*, vol. 27, no. 3, pp. 348–357, 2019.