NATIONAL TECHNICAL UNIVERSITY OF ATHENS
SCHOOL OF ELECTRICAL AND COMPUTER ENGINEERING
SCHOOL OF MECHANICAL ENGINEERING

INTERDISCIPLINARY POSTGRADUATE PROGRAMME
"Translational Engineering in Health and Medicine"

Evaluating Multisensory Integration in Individuals with Autism and Schizophrenia utilizing EEG Data

Postgraduate Diploma Thesis

Theodoros Balougias

Supervisor:

Konstantina Nikita, MD, PhD
Professor, School of Electrical and Computer Engineering

Athens, October 2024

NATIONAL TECHNICAL UNIVERSITY OF ATHENS
SCHOOL OF ELECTRICAL AND COMPUTER ENGINEERING
SCHOOL OF MECHANICAL ENGINEERING

INTERDISCIPLINARY POSTGRADUATE PROGRAMME
"Translational Engineering in Health and Medicine"

Evaluating Multisensory Integration in Individuals with Autism and Schizophrenia utilizing EEG Data

Postgraduate Diploma Thesis

Theodoros Balougias

Supervisor:

Konstantina Nikita, MD, PhD
Professor, School of Electrical and Computer Engineering

The postgraduate diploma thesis has been approved by the examination committee on 30th October 2024

1st member	*2nd member*	*3rd member*
Nikita Konstantina, Professor	***Delis Ioannis, Lecturer***	***Golemati Spyretta, Associate Professor***

Athens, October 2024

…………………………………………………………………………………………………..

Theodoros Balougias
Graduate of the Interdisciplinary Postgraduate Programme,
“Translational Engineering in Health and Medicine”,
Master of Science,
School of Electrical and Computer Engineering,
National Technical University of Athens

Abstract

Multisensory integration (MSI) is the brain's ability to combine inputs from multiple sensory modalities, such as auditory and visual stimuli, to form a cohesive perceptual experience. Rapid and accurate interactions with the environment depend on effective MSI, but people with neurodevelopmental disorders such as schizophrenia (SCZ) and autism spectrum disorder (ASD) frequently struggle with this process. This thesis performs a electroencephalography (EEG) data analysis to examine the neural mechanisms of MSI in these populations relative to healthy controls (CN). Through the use of a reaction-time task, the study examines how various groups process multisensory and unisensory information. Participants are exposed to auditory (A), visual (V), and audiovisual (AV) stimuli and are asked to push a button as quickly as possible. Utilizing sophisticated decomposition techniques like Slice Tensor Component Analysis (SliceTCA) and Tucker decomposition, important neural components that differentiate people who perform better with multisensory stimuli ("integrators") from people who do not ("non-integrators") were extracted from the EEG data. These methods maintain the crucial brain dynamics associated with MSI while simplifying the high-dimensional EEG data. In addition to classifying integrators and non-integrators, a continuous analysis was performed using SliceTCA to examine the relationship between brain activity and response gain times—a measure of the speed advantage gained from multisensory stimuli compared to the faster unisensory condition. Beyond simple binary classification, this continuous analysis made it possible to understand MSI performance in more detail. The study found neural-behavior relationships by correlating neural components with response time gains. This finer-grained analysis highlighted how MSI impairments in these clinical populations manifest at both behavioral and neural levels. A key result of this study was the identification of P-300 playing a role in correlating neural activity with behavior. Moreover, EEGNet, a compact deep convolutional neural network, was utilized to distinguish between the ASD, SCZ, and CN groups since it is well-suited for extracting spatiotemporal features from EEG signals. Through utilization of the complete trial-by-trial EEG dataset, the model yielded further insights into the neural variations that underlie MSI. The general knowledge of how MSI is disrupted in people with ASD and SCZ has been aided by EEGNet's capacity to identify spatial and temporal variations in EEG data. All things considered, this work provides an extensive investigation of MSI through the integration of deep learning models and conventional decomposition methods. By highlighting notable variations in brain activity and behavior among ASD, SCZ, and CN groups, the study sheds light on the neural underpinnings of deficits in multisensory integration. Future diagnostic techniques and treatments targeted at enhancing sensory processing in neurodevelopmental disorders may benefit from these findings.

Keywords: Multisensory integration (MSI)**,** Electroencephalography (EEG)**,** Schizophrenia (SCZ)**,** Autism Spectrum Disorder (ASD)**,** Tucker decomposition, Slice Tensor Component Analysis (SliceTCA), EEGNet, Convolutional Neural Networks (CNN), P-300 Event Related Potential (P-300)

Περίληψη

Οι γρήγορες και ακριβείς αλληλεπιδράσεις με το περιβάλλον εξαρτώνται από την αποτελεσματική ταυτόχρονη επεξεργασία ηχητικών και οπτικών αισθητηριακών σημάτων (MSI), αλλά άτομα με νευροαναπτυξιακές διαταραχές όπως η σχιζοφρένεια (SCZ) και η διαταραχή του φάσματος του αυτισμού (ASD) έχει παρατηρηθεί ότι δυσκολεύονται με αυτή τη διαδικασία. Αυτή η διατριβή χρησιμοποιεί ανάλυση δεδομένων ηλεκτροεγκεφαλογραφίας (EEG) για να εξετάσει τους νευρικούς μηχανισμούς της ταυτόχρονης επεξεργασίας ηχητικών και οπτικών αισθητηριακών σημάτων (MSI) σε αυτούς τους πληθυσμούς σε σύγκριση με υγιείς συμμετέχοντες (CN). Μέσω ενός πειράματος χρόνου αντίδρασης, η μελέτη εξετάζει πώς διάφορες ομάδες επεξεργάζονται πολυαισθητηριακές και μονοαισθητηριακές πληροφορίες. Οι συμμετέχοντες εκτίθενται σε ακουστικά (A), οπτικά (V) και οπτικοακουστικά (AV) ερεθίσματα και πρέπει να πατήσουν το γρηγορότερο δυνατό ένα κουμπί. Χρησιμοποιώντας προηγμένες τεχνικές αποσύνθεσης όπως η Slice Tensor Component Analysis (SliceTCA) και η αποσύνθεση Tucker, που απλοποιούν τα δεδομένα EEG υψηλών διαστάσεων. σημαντικά νευρικά συστατικά που διαφοροποιούν άτομα που αποδίδουν καλύτερα με πολυαισθητηριακά ερεθίσματα ("ενσωματωτές") από άτομα που δεν το κάνουν ("μη ενσωματωτές") εξήχθησαν από τα δεδομένα EEG. Εκτός από την ταξινόμηση των "ενσωματωτών" και "μη ενσωματωτών", πραγματοποιήθηκε μια συνεχής ανάλυση χρησιμοποιώντας το SliceTCA για να εξεταστεί η σχέση μεταξύ της εγκεφαλικής δραστηριότητας και των χρόνων κέρδους απόκρισης—ένα μέτρο του πλεονεκτήματος ταχύτητας που αποκτάται από πολυαισθητηριακά ερεθίσματα σε σύγκριση με την ταχύτερη μονοαισθητηριακή κατάσταση. Πέρα από την απλή δυαδική ταξινόμηση, αυτή η συνεχής ανάλυση επέτρεψε την κατανόηση της απόδοσης στην ταυτόχρονη επεξεργασία ηχητικών και οπτικών αισθητηριακών σημάτων με περισσότερες λεπτομέρειες. Η μελέτη ανακάλυψε σχέσεις μεταξύ νευρικών και συμπεριφορικών δεδομένων, συσχετίζοντας νευρικά συστατικά με τα κέρδη χρόνου απόκρισης σε άτομα διαγνωσμένα με ASD και SCZ. Ένα βασικό αποτέλεσμα αυτής της μελέτης ήταν η ταυτοποίηση του P-300, το οποίο διαδραματίζει ρόλο στη συσχέτιση της νευρικής δραστηριότητας με τη συμπεριφορά. Παράλληλα, το EEGNet, ένα μοντέλο βαθειάς μάθησης που αξιοποιεί συνελικτικά επίπεδα, χρησιμοποιήθηκε για να διακρίνει μεταξύ των ομάδων ASD, SCZ και CN, καθώς είναι κατάλληλο για την εξαγωγή χωροχρονικών χαρακτηριστικών από σήματα EEG. Συνολικά, αυτή η εργασία παρέχει μια εκτενή διερεύνηση της ταυτόχρονης επεξεργασίας ηχητικών και οπτικών αισθητηριακών σημάτων μέσω της ενσωμάτωσης μοντέλων βαθιάς μάθησης και παραδοσιακών μεθόδων αποσύνθεσης. Επισημαίνοντας αξιοσημείωτες διαφορές στην εγκεφαλική δραστηριότητα και τη συμπεριφορά μεταξύ των ομάδων ASD, SCZ και CN, η μελέτη ρίχνει φως στα νευρικά θεμέλια των ελλειμμάτων στην ταυτόχρονη επεξεργασία ηχητικών και οπτικών αισθητηριακών σημάτων. Αυτά τα ευρήματα μπορεί να ωφελήσουν μελλοντικές διαγνωστικές τεχνικές και θεραπείες που στοχεύουν στη βελτίωση της αισθητηριακής επεξεργασίας σε νευροαναπτυξιακές διαταραχές.

Acknowledgments

I would like to express my deep gratitude to Dr. Ioannis Delis for providing me with both the research topic and dataset, and most importantly, for his unwavering support and interest in this project. His guidance through our weekly meetings and insightful discussions was invaluable, and his mentorship was both inspiring and encouraging throughout the entire process.

Additionally, I would like to thank Amir and Atyaf, PhD students of Dr. Delis, for their collaboration. Working alongside them was a truly rewarding experience, marked by a spirit of cooperation, mutual support, and a shared commitment to the project's success.

Table of contents

List of Figures

List of Tables

1. Introduction

1.1 Motivation - Multisensory Integration in individuals with Neurodevelopmental Differences

The brain's capacity to integrate data from several sensory modalities, such as auditory (A) and visual (V) inputs, to create a cohesive perceptual experience is known as multisensory integration, or MSI. Effective engagement with the environment depends on this process, which enables people to react to stimuli more rapidly and precisely. When people are exposed to multisensory stimuli (such as audio-visual or AV) as opposed to the same stimuli when they are presented individually as unisensory inputs (such as visual alone or auditory alone), they usually "gain" in performance. The brain processes sensory inputs optimally when they are congruent in time, which is why this advantage is particularly noticeable in activities where the brain must integrate simultaneous visual and auditory stimuli (Giard & Peronnet, 1999, Molholm et al., 2002).

However, not everyone responds to MSI equally, especially those with neurodevelopmental conditions like schizophrenia (SCZ) and autism spectrum disorder (ASD). ASD and SCZ are similar in that they both have abnormalities in social-emotional processing (e.g., difficulties with theory of mind and alexithymia) and cognitive functioning (e.g., executive function impairments). These conditions also frequently co-occur at higher-than-normal rates (Chandrasekhar et al., 2019). The connection between ASD and SCZ, which has been recognized for a long time (Rapoport et al., 2009), has been recently revisited through the neurodevelopmental continuum model proposed by Owen and O'Donovan (2017). This approach, which is based on genetic evidence, implies that overlapping pathogenic processes may be involved, and that ASD and SCZ may result from distinct consequences of disturbed brain development. Studies have repeatedly demonstrated that people with ASD frequently display notable impairments in MSI, which results in a persistently reduced MSI gain. When compared to neurotypical people, this is manifested in delayed and less accurate reactions to multisensory stimuli (Brandwein et al., 2015; Collignon et al., 2013). In contrast, the findings in SCZ are more variable. Some studies have reported similar MSI deficits in SCZ, with reduced MSI gain during AV tasks (Williams et al., 2010), while others have found no significant differences in MSI between SCZ patients and healthy controls (CN) (Wynn et al., 2014). These inconsistencies suggest that MSI in SCZ may be influenced by other factors, such as the specific tasks used or the heterogeneity within the SCZ population. Notably, the simultaneous comparison of MSI across CN, ASD, and SCZ individuals using the same task is a novel approach, as detailed in (Toumaian et al., 2024) making this investigation the first of its kind to provide direct comparative insights across these groups.

The importance of studying MSI in individuals with ASD and SCZ stems from its fundamental role in perception and cognitive function. MSI allows the brain to combine information from different sensory systems, such as sight and sound, to improve accuracy and response speed to stimuli. This process is crucial for tasks ranging from simple perception to complex cognitive decisions (Talsma et al., 2010). However, research has revealed that MSI is commonly compromised in persons with neurodevelopmental disorders, leading to impairments in sensory processing and behavior. Examining these deficits may yield important information about the brain differences between neurotypical people and those with ASD and SCZ. It also draws attention to potential sharing processes between the two diseases. The objective of this research is to improve diagnosis techniques and therapeutic alternatives by investigating the behavioral impacts and brain activity linked to disruptions in MSI. Technological developments such as machine learning and neuroimaging offer novel approaches to quantify and simulate MSI, which in turn improves our comprehension of sensory processing problems.

Research on the neurological and behavioral components of MSI frequently employs the AV synchronous target detection paradigm (Brandwein et al., 2011, 2013, 2015; Ostrolenk et al., 2019. In this paradigm, participants press a button to reply as fast as they can to A, V and AV stimuli. In order to prevent anticipated responses, catch trials—in which no stimulus is presented—are also included. The AV condition, which has been shown to improve response times (RT) via MSI, is an important measure of sensory integration efficiency for evaluating individuals in a variety of categories, such as neurotypical individuals, people with ASD, and people with SCZ (Stefanou et al., 2020, Collignon et al., 2013). Research using the AV synchronous target detection paradigm has consistently reported reduced MSI gain in reaction times among individuals with ASD, indicating that these individuals benefit less from the concurrent presentation of multiple sensory stimuli. Studies by Brandwein and colleagues (2011, 2013, 2015) have shown these reductions in MSI gain across both children and adolescents with ASD, suggesting a pervasive impairment in multisensory processing. Results in schizophrenia, on the other hand, vary significantly. While some studies, like Wynn et al. (2014), revealed no significant changes, others, like Williams et al. (2011), observed significantly lower MSI gain in SCZ patients compared to healthy controls, suggesting that MSI deficits in SCZ may vary depending on the particular tasks or stimuli employed. Despite these contradictory results, until recently there had not been a thorough simultaneous assessment of the behavioral and electrophysiological MSI effects in patients with ASD and SCZ, along with healthy controls (Toumaian et al., 2024). In order to close this gap, the last study evaluates the early sensory and later cognitive components of MSI using the AV synchronous target detection paradigm. This approach offers new information about the particulars of MSI deficiencies in these groups.

1.2 Previous studies

Previous studies have utilized electrophysiological data, accumulated through EEG recordings, to analyse various ERPs, offering insights into the temporal dynamics of sensory processing (Russo N., 2010, Brandwein et al., 2013). Research has focused on early perceptual components, such as the N1 (auditory) and P1 (visual), and mid-latency components like P2 (auditory) and

N2 (visual) (Wynn et al., 2014). Additionally, the cognitive P3 component has been investigated to understand its role in modulating attention and response variability during MSI (Osaka, 2012, Marucci et al., 2021, Saville et al., 2011, Valakos et al., 2020). Although MSI deficits have been consistently observed in the behavioral data for ASD and SCZ groups, these deficits are not always reflected in the ERP data.

Regarding the results of event-related potentials (ERPs) related to multisensory MSI, the findings from (Toumaian et al., 2024) indicate that MSI effects were observed across several ERP components, including increased peak amplitudes under multisensory conditions compared to unisensory conditions. Specifically, the N1 amplitude was significantly larger in the AV condition compared to the A condition, with a group-by-condition interaction showing a smaller N1 in SCZ patients compared to CN individuals. There were no significant differences in the P2 amplitude between SCZ and CN groups. Interestingly, no significant group differences were found in the P3 amplitude across conditions, although MSI deficits were still evident at the behavioral level in reaction times. However, no correlation could be established even between the ERPs different across groups and the reaction times. The study concludes that this discrepancy suggests that the behavioral deficits observed in MSI may arise at later stages of processing, potentially during response selection, rather than during early sensory processing stages (Toumaian et al., 2024). In a systematic review on MSI and SCZ (Gröhn, C., 2022), it was further highlighted that while MSI-related behavioral deficits were robust in SCZ, differences in early ERP components such as P1 and N2 were less consistent. SCZ patients showed some delayed or reduced ERP components, but these were not consistently different across all studies, indicating that while SCZ patients may have impaired MSI at a behavioral level, this was not always reflected in early electrophysiological responses.

EEG research on schizophrenia has shown characteristic oscillatory activity abnormalities in a variety of frequency bands, providing information about the neurophysiological basis of the disorder. The shift from studying ERPs to oscillatory frequencies has brought to light several facets of brain dysfunction in schizophrenia. Analyzing in the frequency domain provides an alternative viewpoint. Patients with schizophrenia have aberrant oscillatory behavior in a number of frequency bands, such as the alpha/gamma (White et al., 2010) and theta/delta (Boutros et al., 2008) bands. An inability to maintain neuronal synchrony during sensory processing has been connected to deficits in these bands. For example, research has shown that when individuals with schizophrenia are exposed to various stimuli, their gamma and beta power decreases (Uhlhaas & Singer, 2013). Several studies examining neural oscillations in individuals with schizophrenia exposed to auditory stimuli have identified deficits in steady-state evoked potentials, particularly in the gamma frequency range, with some impairments also observed in lower frequency bands (e.g., Krishnan et al., 2009; Kwon et al., 1999).

Another hypothesis regarding MSI is the concept of simple additive effects, as proposed by Miller (1982) in his “race model,” which suggests that the brain selects the fastest response from the available sensory inputs. To assess whether true MSI occurs, statistical methods have been developed that compare the observed reaction time distributions during multisensory stimulation to those expected from additive effects of unisensory stimulation. This approach, pioneered by Diederich and Colonius (1987) and further refined by Ulrich et al. (2007), has been widely used to evaluate MSI gain at the group level. More recently, innovative approaches like the bootstrap method have been developed to test MSI gain at the individual level (Toumaian et al., 2024). This method involves creating bootstrapped RT distributions by randomly pairing RTs from auditory and visual conditions, then comparing the mean RT of the audio-visual condition to these distributions to determine if it significantly exceeds expectations, thus confirming true MSI gain. Using our identical dataset, the previous study's racial model inequality analysis identified variations in the incidence of MSI gain among the three data groups (Control (CN), ASD, and SCZ). The CN group exhibited considerable departures from the race model, indicating that MSI gain was present. On the other hand, SCZ patients showed significant deviations for a narrower range of bootstrapped distribution quantiles, indicating a less prominent MSI growth in comparison to the control group. Finally, there were no notable departures from the race model among the ASD patients, suggesting that MSI increase did not occur in this group.

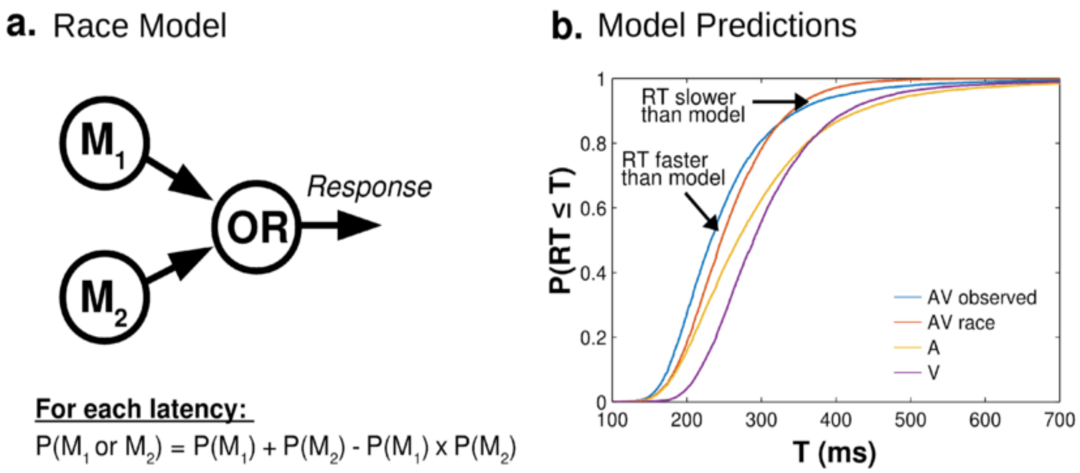


Figure 1: The race model of multisensory reaction times. (A) Diagram displaying the framework of the race model. Each sensory modality's independent processing channels (M1 and M2) feed into an OR operator, which triggers a reaction. The reaction time for a particular trial in this arrangement is determined by which of the two channels is faster. Based on the statistical function of the OR operator (as indicated by the equation), the race model forecasts the chance of a response happening at or before a given time T, assuming no correlation exists between the channels. (B) A comparison of the actual reaction time (RT) distribution (blue) for AV trials in the current dataset with the predictions of the race model (red). The race model prediction for each time bin is calculated using the cumulative RT distributions for the unisensory (A, yellow) and (V, purple) trials. The model underestimates the probability of faster trials, showing that the quickest responses (left tail) are faster than predicted by the race model. Conversely, it overestimates the likelihood of slower trials, indicating that the slowest responses (right tail) are slower than the race model anticipates (Plass, J., & Brang, D., 2021).

1.3 Current study

This research attempts to investigate the brain mechanisms behind MSI in three different groups: 32 CN individuals, 35 SCZ persons, and 23 ASD individuals' data were used. Every subject finished a straightforward RT task with four different trial types: AV, V, A stimuli and catch trials. Participants in the visual condition saw a red disc, while those in the auditory condition heard a tone at 1000 Hz. The tone and the disc were shown simultaneously during the AV condition, which allowed us to compare the way individuals processed multisensory and unisensory signals. Using this data provided by Nikolaos Smyrnis and the Neurosciences and Precision Medicine Research Institute "COSTAS STEFANIS," we look at how these groups differ in MSI performance, particularly in relation to brain activity and behavioral outcomes, such as response time gains.

Tucker decomposition, a potent method for lowering the complexity of high-dimensional data, is what we start with. Tucker decomposition breaks the EEG data into a core structure and factor matrices that represent participants, time points, and EEG channels. We are able to record significant interactions between different dimensions through this technique, providing a comprehensive image of the brain activity that takes place during MSI tasks. We use methods such as AdaBoost and K-Nearest Neighbors (KNN) to develop a classification pipeline after the data has been deconstructed. This pipeline uses the most pertinent components found by Tucker decomposition to classify participants as "integrators" or "non-integrators" depending on their MSI performance. The aim is to identify the brain components that differentiate people who gain from multisensory signals from those who do not.

Tucker decomposition reduces dimensionality well, but it's not always easy to figure out how many components is the right amount and how to break down the intricate relationships between them. To address this, we also applied Slice Tensor Component Analysis (SliceTCA), a more specialized method that enabled us to maintain shared topographic maps across participants while allowing individual differences in the timing of neural responses. SliceTCA is particularly well-suited for EEG data analysis because it is able to preserve a common spatial structure, which is essential when comparing brain activity between groups. This method enables us to analyze how temporal dynamics vary between participants while keeping the spatial organization intact, making it easier to identify group-specific and stimulus-specific neural responses related to MSI.

Following SliceTCA decomposition, we investigated the connection between the neuronal constituents and behavioral results using a correlation analysis. Our correlation analysis operates on a continuous scale, in contrast to the conventional binary classification method, which only classifies participants as integrators or non-integrators. This method has multiple benefits: by looking at the degree of MSI gain as a continuous variable, it enables a more accurate assessment of the relationship between brain activity and MSI performance.

This allows us to examine how different degrees of MSI gain correspond to brain activity, providing us with a much more thorough picture of these processes than simply classifying someone as an integrator or not.

To identify particular time periods where brain activity differed significantly between audiovisual and unisensory stimuli, we performed a sliding window bootstrap analysis. During these periods, we computed important EEG metrics (mean, maximum, and minimum differences) and compared them to the response gains of the participants, or the speed at which they responded to multisensory stimuli as opposed to unisensory ones. This continuous-level approach enables us to find finer and more complicated correlations between brain activity and behavioral performance that may be missed with binary classification. By combining SliceTCA for neural decomposition with a continuous correlation analysis, this study provides a detailed look at MSI in SCZ and ASD populations. This approach, which balances spatial consistency with the flexibility to analyze temporal variations, allows us to better understand the neural mechanisms driving MSI and how these processes differ across various groups.

In addition to the previously mentioned methods, we also employ EEGNet, a compact convolutional neural network architecture specifically designed for EEG data, to classify the different groups in our dataset: ASD versus CN, CN versus SCZ and SCZ versus CN. Across a range of BCI scenarios, EEGNet's depthwise and separable convolutions enable effective and understandable feature extraction from EEG data (Lawhern et al., 2018). Even with a small amount of training data, this network has demonstrated resilience in identifying EEG signals from many paradigms, including movement-related cortical potentials, P300 visual-evoked potentials, error-related negative responses, and sensory motor rhythms. It has also achieved good performance in this regard.

We employ EEGNet to classify the different groups by analyzing their EEG responses during the MSI task. The architecture of EEGNet is highly suitable for this purpose, as it effectively captures the spatiotemporal characteristics of EEG signals. We evaluate the characteristics that contribute to the categorization between groups using feature visualization approaches after the model has been trained. With this method, we may identify the precise brain structures and patterns that set apart people with ASD and SCZ from those in the CN group. We want to improve upon the results of SliceTCA and Tucker decomposition by integrating EEGNet into our analytical framework, providing a more comprehensive understanding of the brain mechanisms producing MSI and their differences among distinct clinical groups.

.

2. Dataset

The pre-processed dataset was kindly provided to us by Nikolaos Smyrnis and the Neurosciences and Precision Medicine Research Institute “COSTAS STEFANIS”.

2.1 Participants

This study involved a total of 90 participants, categorized into three groups: 32 healthy controls (CN) (14 males, 18 females; mean age = 26.62 years; age range = 17-38 years), 35 individuals with schizophrenia (SZ) (26 males, 9 females; mean age = 28.29 years; age range = 17-44 years), and 23 individuals diagnosed with autism spectrum disorder (ASD) (20 males, 3 females; mean age = 25.70 years; age range = 18-41 years). All SZ and ASD participants were diagnosed by a psychiatrist according to the DSM-IV-TR criteria and were in a stable phase of their disorder. Healthy controls reported no significant medical history or systemic illnesses. Participants in the ASD group were recruited through day-care centers in Athens, and SZ participants were recruited through the Psychiatry Department of Eginition Hospital. Participants with SZ had to be currently receiving benzodiazepine medication in order to be excluded, while participants with ASD had to be excluded if they were also diagnosed with attention deficit/hyperactivity disorder (ADHD). The Eginition University Hospital's ethical committee approved the study protocol. All participant data were anonymized using special codes, and data confidentiality was scrupulously upheld. Authorized workers only had access to safe storage and restricted storage. The examination of multisensory integration differences between the ASD and SZ groups was made possible by using the CN group as a baseline for comparing the responses and neurophysiological parameters of these groups.

2.2 Stimuli and Experimental Procedure

Three stimulus conditions—auditory (A), visual (V), and audiovisual (AV)—were randomly given with equal probability to participants in a basic reaction time (RT) detection test. A 1000 Hz tone (duration: 60 ms; 75 dB SPL; rise/fall time: 5 ms) was provided in the auditory condition. In the visual condition, an LCD panel showed a red disc with a diameter of 3.2 cm against a black background for 60 milliseconds. The auditory and visual stimuli were presented simultaneously in the AV condition. The participants' task methods and response delaying capacity were evaluated through the use of randomly inserted blank catch trials, which lacked any stimuli. Participants were instructed to press a button with their right index finger on a specially designed response keyboard (Empirisoft) upon detecting any stimulus (A, V, or AV) within a 1400 ms window. The inter-trial interval varied between 2000 and 3000 ms, and the stimuli were presented in blocks of 100 trials each, with 25 repetitions per stimulus condition, for a total of 6 blocks.

2.3 EEG Acquisition and Pre-processing

EEG data were recorded from 61 scalp electrodes using the ISO-1064CE and CONTROL-1164 Braintronics AC system (Almere, The Netherlands) following the 10-20 system. Impedance was maintained below 10 kΩ using an electrolyte gel (Neurgel, Spes Medica s.r.l., Genova, Italy). The reference electrode was placed between CPz and Pz, while the ground electrode was located between Fz and AFz. EEG signals were recorded with a time constant of 10 seconds and a low-pass filter of 100 Hz, with an additional notch filter at 50 Hz to eliminate power line interference. Data were sampled at 1024 Hz and stored for offline processing.

Matlab 2015a (MathWorks, USA)'s Fieldtrip toolbox (version 11.11.15) was used for EEG pre-processing. 200 ms of pre-stimulus and 1000 ms of post-stimulus data were included in each trial. Pre-processing was done in four steps: (1) visually inspecting trials and channels to remove noisy ones; (2) removing trials and channels using inter-correlated variance with a variance threshold of two; (3) using independent component analysis (ICA) to remove artifacts (electrooculographic, electrocardiographic, and electromyographic signals); and (4) re-referencing to the average of all electrodes. A 12 Hz low-pass filter was used to filter the EEG data, and event-related potentials (ERPs) were computed for every electrode, participant, and stimulus condition (A, V, AV, and catch).

2.4 Grouping participants into “Integrators” vs “Non Integrators”

The multisensory integration (MSI) study's participants were categorized as either "integrators" or "non-integrators" according to how well they performed in an audiovisual (AV) target detection test. A bootstrap approach in conjunction with a race model test was used to classify the data. When processing auditory and visual information separately, the race model predicts an individual's reaction time to AV stimuli. In the event that the individual's actual reaction time (RT) to auditory stimuli was noticeably quicker than the predicted RT, multisensory integration (MSI) was being advantageous, and the participant was labeled as a "integrator."

To create a distribution of reaction times from the unisensory (visual and auditory) conditions, the bootstrap approach was utilized. A person was considered to have MSI gain if their AV reaction time was substantially faster than this bootstrapped distribution. However, the subject was labeled as a "non-integrator" if there was no discernible increase in RT for AV stimuli when compared to the unisensory circumstances. The results of the study showed a progressive drop in MSI capacity from healthy individuals to those with schizophrenia and autism, with 78% of healthy controls (CN), 49% of people with schizophrenia, and 26% of those with autism demonstrating MSI gain. The following table provides a summary of the dataset.

Table 1: Summary table of dataset

Group	Number of Participants	Male	Female	Mean Age (years)	Age Range (years)	MSI Integrators (%)	Recruitment Source	Exclusion Criteria
Healthy Controls (CN)	32	14	18	26.62	17-38	78	General population	No significant medical history or systemic illnesses
Schizophrenia (SZ)	35	26	9	28.29	17-44	49	Psychiatry Department of Eginition Hospital	No current benzodiazepine treatment
Autism Spectrum Disorder (ASD)	23	20	3	25.70	18-41	26	Day-care centers in Athens	No concurrent diagnosis of attention deficit/hyperactivity disorder (ADHD)

3. Methods

3.1 Tucker Decomposition and Classification Analysis

In this methodology, we utilize Tucker decomposition to reduce the dimensionality of EEG data collected from autistic, schizophrenic, and control participants. This decomposition allows us to identify the core underlying factors that capture essential neural dynamics across participants, time, channels and stimuli. Subsequently, we apply classification algorithms using the most informative components from the decomposition to distinguish between integrators and non-integrators based on the participants' behavioral response times.

The Tucker decomposition is a form of higher-order principal component analysis (PCA) that breaks down a multi-dimensional tensor into a core tensor and factor matrices. Specifically, in a three-way case where the tensor has dimensions I×J×K it is approximated by a core tensor that is transformed by factor matrices along each mode.

$$X \approx G \times_1 A \times_2 B \times_3 C$$

where:

- X is the original tensor,
- G is the core tensor that captures the interactions between the different components,
- A, B and C are factor matrices associated with each mode of the tensor,
- $\times_n$ denotes the mode-n product.

These factor matrices, typically orthogonal, represent the principal components in each dimension. The core tensor reveals the interactions between the components across modes, allowing for a compressed version of the original data when the factor matrices have fewer components than the original dimensions (Tucker L.R.,1966). This approach is flexible and can be tailored to specific applications, depending on the dimensions and structure of the data. It serves as a foundational tool in tensor decomposition, offering insights into the multi-dimensional structure of datasets. (Kolda, T. G., 2009)

In this study we implemented both Tucker and non-negative Tucker decomposition, while also using the algorithm FISTA in Numpy for faster convergence. NN-Tucker decomposition is a constrained form of Tucker decomposition where the core tensor G and the factor matrices A, B, and C are constrained to have non-negative elements. This non-negativity constraint is particularly useful in applications where the data or its underlying components are non-negative. This is not the case for our data, but we tried it out and obtained some meaningful and interpretable results. FISTA is an optimization algorithm that accelerates the convergence of the basic Iterative Shrinkage-Thresholding Algorithm (ISTA) for solving large-scale convex optimization problems, especially those involving sparsity-inducing penalties (e.g., Lasso).

The Basic ISTA

The ISTA is designed to solve problems of the form:

$$\min_x \{f(x)+g(x)\}$$

where f(x) is a smooth, differentiable convex function and g(x) is a non-smooth convex function. In the context of NN-Tucker, f(x) could represent the reconstruction error and g(x) the non-negativity constraint.

The update rule for ISTA is:

$$x^{(k+1)} = \mathrm{prox}_{\gamma g}(x^{(k)} - \gamma \nabla f(x^{(k)})$$

where $\mathrm{prox}_{\gamma g}$ is the proximal operator of g with step size γ.

FISTA improves upon ISTA by introducing an additional momentum term, which significantly speeds up convergence. The update rules are:

$$y^{(k)} = x^{(k)} + \frac{t^{(k-1)} - 1}{t^k} (x^k - x^{(k-1)})$$

$$x^{(k+1)} = \mathrm{prox}_{\gamma g}(y^{(k)} - \gamma \nabla f\big(y^{\{k\}}\big))$$

$$t^{\{k+1\}} = \frac{1 + \sqrt{1 + 4(t^{\{k\}})^2}}{2}$$

Here, $t^{(k)}$ is a sequence that controls the momentum term, and the update uses both the current and previous iterates to accelerate convergence. (Beck, A., & Teboulle, M., 2009)

Tucker decomposition offers a distinct advantage over CP decomposition due to its flexibility. Unlike CP, which uses the same rank across all modes, Tucker allows for different ranks in each mode, enabling it to capture more complex and diverse interactions within the data. This flexibility makes Tucker decomposition particularly effective when the relationships between dimensions are not uniform, providing a richer and more detailed representation of the data through its core tensor, which highlights these varying interactions. In our case, this flexibility enables us to explore a range of interactions between participants, time points and channels, allowing us to identify those interactions that yield the most effective discrimination in the subsequent classification analysis. However, this added flexibility also introduces a key drawback—the need to select different ranks for each mode can make the implementation and interpretation of Tucker decomposition more challenging. To address the issue of interpretability, we visualized heatmaps of the core tensors, as shown in the results section, to illustrate the interaction strengths between the factor matrices. Additionally, the computational complexity of Tucker decomposition requires more sophisticated and time-consuming algorithms to process, particularly as the number of modes and ranks increases. This higher complexity makes it more resource-intensive and less efficient than CP decomposition, especially when working with large or high-dimensional datasets

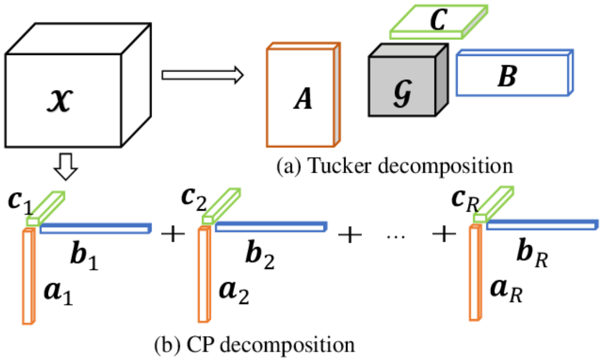


Figure 2: Illustration of the two decompositions (Jiang, T. X., 2017)

Tensor formation

Our original EEG dataset is structured as a four-way tensor of shape (90, 1025, 61, 600), where 90 represents the participants across all groups (32 control, 23 ASD, and 35 CN), 1025 denotes the time points, 61 refers to the EEG channels, and 600 corresponds to the number of trials. The trials consist of four types: 150 audio trials, 150 visual trials, 150 audiovisual trials, and 150 catch trials. The catch trials were excluded from further analysis, reducing the number of trials in the tensor, but we still work with a four-way tensor at this stage. To capture more variance and simplify the data, we averaged the EEG signals over trials within each condition. Specifically, we calculated the average EEG signals for A, V and AV conditions separately for each participant. The resulting tensor has a shape of (90, 1025, 61, 3), where the first dimension represents participants, the second and third dimensions correspond to time points and electrode channels, respectively, and the fourth dimension includes the three conditions (A, V, and AV). The fact that we stack the participants of all groups in the first axis leads to a group-agnostic approach, which is desirable for capturing the neural correlates for integration across these groups. The Tucker decomposition is then applied to these reduced tensors, breaking them down into a core tensor and factor matrices. The core tensor provides a lower-dimensional representation of the original data, capturing interactions between the participants, time points, and channels. The factor matrices, corresponding to participants, time points, channels and stimuli, reveal latent patterns within each dimension, helping to uncover the structure of the EEG signals and the underlying neural mechanisms involved in multisensory integration.

Classification Using Best Components

The classification process aims to differentiate integrators from non-integrators by leveraging the most meaningful components identified through Tucker decomposition. This dataset is well-balanced, with integrators representing 53% and non-integrators 47%. The classification model uses participant-specific factors, which capture individual variations in neural activity, as key features in distinguishing between the two groups. By employing cross-validation techniques, the model assesses different components to identify those that most effectively separate integrators from non-integrators. Since this analysis spans across all groups, identifying a component with high predictive accuracy could reveal a common neural mechanism that is prominent in integrators but less pronounced in non-integrators. Various algorithms were explored for the classification task, with the best results achieved using AdaBoost and KNN, highlighting their effectiveness in this context. This approach not only helps link neural activity patterns to behavioral outcomes, such as response times, but also enhances our understanding of the underlying neural processes associated with multisensory integration. The final classification accuracy is averaged over the five folds used in cross-validation, ensuring the robustness and generalizability of the model's predictions.

AdaBoost, short for Adaptive Boosting, is an ensemble learning algorithm designed to improve the performance of weak classifiers by combining them into a strong classifier. The core idea behind AdaBoost is to iteratively train weak classifiers, typically decision trees, and adjust their weights based on the classification errors. During each iteration, AdaBoost assigns higher weights to the misclassified instances, forcing the next classifier to focus more on the difficult cases. By combining the outputs of these weak learners through a weighted majority vote, AdaBoost creates a strong classifier capable of handling more complex classification tasks (Freund & Schapire, 1997). One of the key advantages of AdaBoost is its ability to adaptively modify the weights of weak classifiers, allowing it to improve accuracy without overfitting. It is particularly effective in scenarios where the data contains noise or complex patterns. AdaBoost has been successfully applied in a variety of fields, including image recognition, text classification, and, as in our case, neural data classification.

K-Nearest Neighbors (KNN) is a simple yet effective classification algorithm widely used for pattern recognition and classification tasks. KNN operates on the principle of similarity by classifying a data point based on the majority label of its closest neighbors in the feature space. The "k" in KNN refers to the number of neighbors considered during classification, which is typically an odd number to avoid ties. The algorithm calculates the distance between data points using metrics such as Euclidean distance, and assigns the new data point to the class most common among its nearest neighbors (Cover & Hart, 1967). One of the strengths of KNN is its simplicity and ease of implementation, as it does not require an explicit training phase like other algorithms. However, its performance heavily depends on the choice of k, which in our case was 3 after tuning, and the distance metric, which in our case is the Euclidean distance. The Euclidean distance between two points $p=(p_1,p_2,\ldots,p_n)$ and $q=(q_1,q_2,\ldots,q_n)$ an n-dimensional space is given by the following equation:

$$d(p,q) = \sqrt{(p1 - q1)^2 + (p2 - q2)^2 + \cdots + (pn - qn)^2}$$

KNN is particularly well-suited for multi-class problems and is effective in classification tasks where the boundaries between classes are not linear.

3.2 SliceTCA decomposition

Next, we tested another decomposition method called Slice Tensor Component Analysis or SliceTCA. The key difference between SliceTCA and Tucker decomposition lies in their approach to factorization. While Tucker decomposition yields a core tensor and factor matrices for each dimension (participants, time, channels and stimulus conditions), SliceTCA decomposes such a tensor by preserving one dimension, such as channels, consistently across all data entries, while simultaneously identifying distinct patterns in the remaining dimensions, such as participants and time. This method enables the separation of complex co-variability by maintaining a fixed dimension (channel) and disentangling the contributions of other dimensions, thereby capturing how different temporal patterns or participant-specific dynamics interact with shared neural activity across the preserved dimension. This makes SliceTCA particularly effective for identifying shared spatial patterns (common topographic maps) across participants, while still allowing for variability in temporal profiles or participant-specific responses (Pellegrino A., 2024). Conversely, without this explicit focus on maintaining spatial consistency across groups, Tucker decomposition applies a more generic factorization, which makes it less intuitive for directly comparing spatial activity across all participants and time points at the same time.

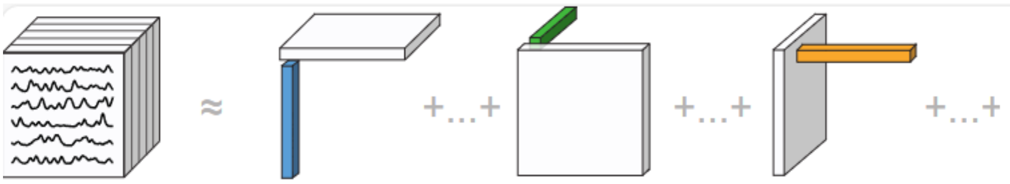


Figure 3: SliceTCA decomposition(Pellegrino A., 2024)

The most commonly known Tensor Component Analysis (TCA) decomposes a data tensor into components represented as the outer product of vectors from each dimension, enforcing that each component captures shared variability across all dimensions simultaneously. On the other hand, by enabling components to be defined along distinct dimensions or "slices" of the tensor, such as participants, time, or channel dimensions, SliceTCA provides more flexibility. This implies that structured variability can be captured by SliceTCA along particular dimensions without requiring a uniform structure along all of them. SliceTCA is more flexible when dealing with complicated data structures since it isolates components within certain dimensions, unlike TCA, which is limited to lie at the intersection of all variable classes.

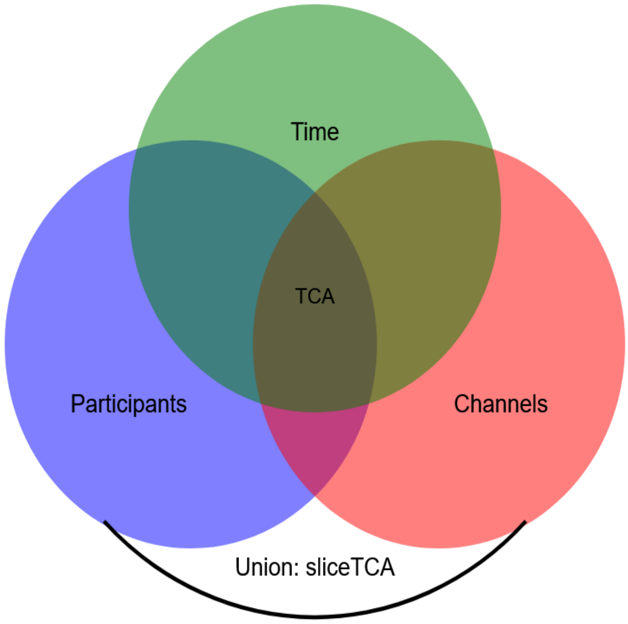


Figure 4: Variability captured by SliceTCA compared to TCA

In terms of comparing it to Principal Component Analysis (PCA), single slicing SliceTCA and PCA are viewed as equivalent in the sense that both methods reduce data dimensionality by capturing the most important components that explain the variance. In single slicing SliceTCA, the data tensor is unfolded along a chosen dimension, such as electrode channels, resulting in a lower-dimensional representation akin to the principal components identified by PCA in a matrix. The main advantage of single slicing SliceTCA lies in its ability to confine the principal components to a specific dimension, such as channels or time in EEG studies. This feature is particularly valuable for EEG data analysis, where it is crucial to focus on neural activity across trials or time while maintaining the interpretability of the results. By isolating significant components within a designated dimension, SliceTCA facilitates targeted analysis of neural signals, something traditional PCA cannot achieve as it considers all dimensions of the data. This approach makes SliceTCA more effective for extracting meaningful patterns from complex EEG data, where emphasis on particular dimensions is necessary.

SliceTCA has been used to separate complex forms of variability in a number of neural data studies. Compared to conventional methods, it has revealed task-relevant structures with fewer components when used to analyze motor cortical dynamics during reaching tasks. SliceTCA revealed different covariability classes in the midbrain, thalamus, hippocampus, and visual cortex among other brain regions in multi-region recordings. By taking trial-based variability into account, it has also been used to identify task-specific manifolds in cortical and cerebellar data collected during cued motor tasks. These uses demonstrate how well SliceTCA captures complex neural patterns in various experimental settings (Pellegrino A., 2024). To our knowledge, it is the first time that SliceTCA has been tested on EEG data.

SliceTCA and non-negative SliceTCA

For our dataset, we utilized both SliceTCA and non-negative SliceTCA methods, specifically focusing on channel slicing. To get a thorough representation of the data structure, we first used these methods on the complete dataset. We then separately applied non-negative SliceTCA to the original data's positive and negative subsets. For the negative data, we extracted the inherently negative values from the dataset and then inverted their sign before applying non-negative SliceTCA. This method allowed us to ensure that the non-negativity constraint was preserved while avoiding the cancellation of values that typically occurs when positive and negative components are combined. The purpose of this approach was to determine whether separating and analyzing the positive and negative components independently would facilitate the extraction of more meaningful and interpretable components. Non-negative SliceTCA seeks to improve interpretability by encouraging the extracted components to be represented as distinct, non-overlapping features of the data by placing a non-negativity constraint on the tensor factors. Therefore, we tested whether this could lead to a clearer and more intuitive understanding of the underlying structure.

Regarding SliceTCA, we employed an invariance methodology also from the original SliceTCA paper (Pellegrino A., 2024) to address the inherent ambiguities and ensure a more reliable decomposition of the neural data. Channel slicing involves capturing patterns of covariability across multiple channels over time and participants. However, because of the possible transformations within and between slice types, this can result in several equivalent solutions. If these invariances are not addressed, it becomes difficult to assign distinct and significant interpretations to the extracted components, since they might be a combination of overlapping contributions from various trials, time points, or channels.

A hierarchical optimization process is used in the methodology to reduce these uncertainties and produce a decomposition that is comprehensible and consistent. There are multiple stages to this process. The process starts by minimizing a global reconstruction loss to make sure that the overall tensor decomposition accurately matches the data that was observed. Next, in order to guarantee that the contributions of every slice type are independently optimized without compromising the global fit, the partial reconstruction losses unique to each slice type—such as channel, time, and trial slices—are minimized. In the final stage, additional constraints are introduced to promote component orthogonality and distinctiveness within and between slice types. This step ensures that the extracted components are unique and do not redundantly capture the same sources of variability. Through a hierarchical process of systematic minimization of these sub-loss functions, the methodology reaches a solution in which each element uniquely reflects a particular dimension of the variability in the data. By avoiding overlaps and guaranteeing a stable and comprehensible decomposition, this methodical approach facilitates a more useful comprehension of the temporal and spatial patterns of neural activity throughout various channels.

Data Preparation and Stacking

Initially divided into three participant groups (ASD, SCZ and CN), the data are first averaged over trials in order to eliminate noise specific to individual trials and concentrate on group-level patterns. The EEG responses to the three stimulus situations from each participant group are "stacked" into a single tensor. The stacking process transforms the data into a consistent format across the three groups, ensuring the tensor captures the overall neural dynamics while retaining distinctions between the participant groups and the stimulus types.

The stacking process ensures that the tensor accurately captures complex neural dynamics across participants, while retaining distinctions between the different groups (ASD, SCZ and CN) and stimulus types (A, V and AV). In this multi-dimensional format, the tensor has three primary axes: participants, time, and EEG channels. This structure allows SliceTCA to perform a decomposition that not only reveals patterns of brain activity across these dimensions but also disentangles the contributions of the different groups and stimuli. One of the strengths of SliceTCA is that it enables us to leverage the participant axis to disentangle the group-level and stimulus-specific neural responses. The recordings of all 90 participants from all three groups, exposed to the three stimuli, are stacked along the participant axis. Each component of the tensor decomposition produces a slice, and by focusing on the participant axis of each slice, we can separate and analyze the neural responses corresponding to different groups and stimuli.

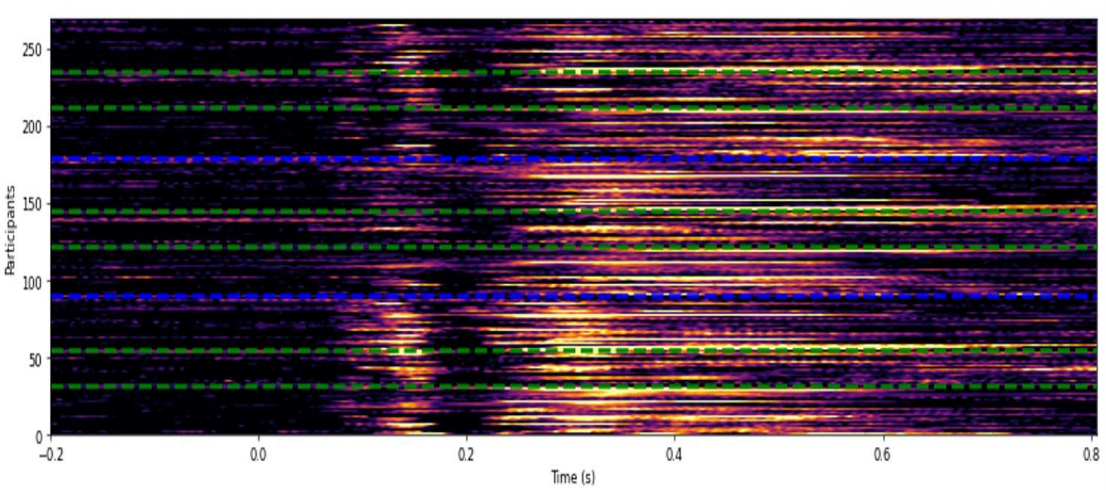


Figure 5: An example of a stacked participant x time slice. The blue dotted lines separate the 3 stimuli (A, V, AV from top to bottom) and the green dotted lines separate the 3 groups (CN, ASD, SCZ from top to bottom)

This disentanglement makes it possible to perform in-depth statistical comparisons and analysis at the stimulus and group levels. Slices that isolate participant reactions to a certain stimulus type (A, for instance) or concentrate on slices that reflect particular groups (ASD individuals, for instance) can be extracted. We may examine group-specific patterns or stimulus-specific effects by examining these slices, which enables a thorough comparison of the ways in which various groups react to each stimulus. The capacity to concentrate on particular tensor slices augments the analysis's accuracy and fineness, enabling the discovery of minute variations in brain activity among subjects and stimuli.

Grid Search for Optimal Number of Components

One of the key steps before running the SliceTCA algorithm is determining the optimal number of components to specify for the decomposition. This is done via a grid search process, which systematically tests different configurations of component numbers. The objective here is to minimize the loss function (typically reconstruction error) on the validation set. The methodology for splitting the tensor into training and validation sets using masks follows the approach outlined in the original SliceTCA paper (Pellegrino A., 2024). This method makes sure that distinct portions of the data are systematically distributed between training and testing, allowing for a more thorough assessment of the model's effectiveness. We guarantee that the decomposition process discovers significant patterns from the training set while the validation set is hidden by employing this mask-based partitioning, which enables an objective evaluation of the model's generalization skills. The goal of the grid search is to determine the optimal number of components to ensure that the model captures meaningful structure in the EEG data without overfitting, while also offering the ideal balance between model accuracy and complexity.

SliceTCA relies heavily on the grid search process because the number of components directly affects how easily the findings can be interpreted. A component count that is too low could oversimplify the data and miss important differences between groups, while a component count that is too high could introduce noise and obscure the findings. We can improve the model's robustness and generalizability by choosing the ideal number of components that result in a low loss function through the use of the grid search. In this study, we chose to rely on channel slicing and therefore we tested the optimal number of channel-dimension components while keeping zero components across the two remaining dimensions (time and participants). The usefulness of channel slicing will be made clear in the following section, however we also tested decompositions with more than one slice types. Although the validation loss kept decreasing, especially across the time axis which attributes to most variance of the data, the channel slicing components we want for our subsequent analysis were less interpretable in this mixed slicing scenario. This should be due to the fact that, when using only channel slices, the decomposition treats these channel slices as principal components and therefore captures more variance through them.

Channel Slicing Approach

Once the optimal number of components is determined, we proceed with the channel slicing approach. This approach makes effective use of the spatial structure of EEG channels to simplify interpretation and facilitate meaningful comparisons between individuals and groups, making it especially intuitive and potent for analyzing EEG data. With channel slicing, the algorithm breaks down the data so that each participant's unique temporal profile is preserved while we obtain a shared topographic map, or spatial pattern, across all participants and groups. This is a crucial feature because it maintains the spatial consistency offered by the EEG channels while enabling direct comparison of brain activity across various groups and stimuli conditions.

First, having a shared topographic map means that the spatial pattern of EEG data is the same across all groups and participants. This uniform spatial structure makes the analysis simpler by allowing us to focus on changes in the timing of the activity without altering the spatial patterns. It also helps in easily visualizing and understanding how brain regions process information differently, as represented by the EEG channels. Second, while the spatial map is fixed, the timing of neural activity can change from person to person. This approach provides the flexibility to examine how each individual's brain activity changes over time, which is useful for exploring differences in how individuals and groups respond to various stimuli. This temporal profile shows how EEG signals progress over time and allows for a more detailed look into cognitive processes related to specific tasks or stimuli. Lastly, after breaking down the data in this way, we can compare the temporal components for each participant to identify differences among ASD, SCZ and CN groups. These timing signals are important for detecting unique neural responses that may be linked to specific groups or stimuli, capturing details such as brainwave patterns, responses to events, or synchronization with stimuli.

3.3 Link to Behavioral Results: Correlation with Response Times

In this methodology, the main goal is to decompose EEG data from ASD, SCZ and CN participants to identify latent neural patterns that are common or differ across these groups. One of the key advantages of using the channel slicing approach is the ability to relate the temporal profiles extracted from the EEG data to behavioral outcomes—in this case, response times. Since each participant's temporal profile is distinct, we can perform correlation analyses between these profiles and the participants' behavioral data, such as response times during the experimental tasks. We can examine possible group differences and learn more about how neural dynamics affect task performance by analyzing each participant's temporal components and comparing them with response times. For example, in ASD versus CN participants, we may find that particular temporal patterns predict faster or slower response times. These correlations provide us important new insights into how various groups process sensory data and make decisions by connecting the neural data to cognitive or behavioral outcomes.

Bootstrap Analysis with Sliding Window: Temporal Signal Significance Testing

After the temporal signals are decomposed using SliceTCA, a crucial step is identifying the time windows where significant differences exist between brain activity in response to multisensory (AV) and unisensory (A or V) stimuli. This analysis helps reveal when and where the brain processes multisensory integration differently compared to the stronger unisensory stimuli (either A or

V). Our function compares the neural responses to AV stimuli against the stronger unisensory stimulus across time using a non-parametric bootstrap analysis. This is done on the averaged multisensory and unisensory temporal profiles. The goal is to detect statistically significant differences between these two conditions in localized temporal windows.

1. **Sliding Window Comparison**: The EEG data from both conditions (AV vs. stronger unisensory) is analyzed over a series of overlapping time windows. Each window captures a small segment of the time series (e.g., 50 time points), which is then shifted by a set number of points (e.g., 15). This approach enables continuous comparison across the entire time range of the signals while focusing on smaller, interpretable segments of neural activity.

2. **Bootstrap Sampling**: The function resamples the data with replacement across participants and time points within each time window. A distribution of potential differences within that window is produced for each bootstrap sample by computing a difference (such as the median or trimmed mean difference) between the multisensory and unisensory conditions.

3. **Estimating Differences**: A summary statistic (trimmed mean or median) is calculated for each time window to show the central tendency of the neural signals within that window. The computation of the difference between the two stimulus conditions yields a localized measure of the differences in brain activity between processing AV and the stronger unisensory stimuli.

4. **Confidence Intervals and P-Values**: The bootstrap-generated distribution of differences allows for the calculation of confidence intervals. These intervals estimate the likely range of the true difference between the two conditions. Additionally, p-values are computed to determine the significance of the observed differences, based on how much of the bootstrap distribution supports the null hypothesis (i.e., no difference between conditions).

5. **Significance Testing**: The function determines whether there is a statistically significant difference between the multisensory and unisensory stimuli for each time window. When the p-value falls below a predetermined threshold (e.g., 0.05), significant windows are identified, signifying the intervals of time during which there are discernible differences in brain activity between the two conditions.

Application to EEG Data

This method is particularly valuable for studying multisensory integration. By applying the sliding window approach, we can detect the specific temporal windows where brain activity in response to multisensory stimuli differs from that of the stronger unisensory stimulus. By shedding light on the specific circumstances under which the brain interprets these stimuli differently, this localized analysis advances our knowledge of the mechanisms underlying sensory processing.

Correlating these EEG differences with response gains for each participant comes next, after the significant temporal windows of EEG differences between AV and unisensory stimuli are identified. In this scenario, response gain is defined as the difference between the response time for the stronger unisensory stimulus (either audio or visual, depending on which one was faster) and the response time for the audio-visual stimulus. The calculation of response gain provides a measure of how much faster or slower a participant responded to multisensory stimuli compared to the unisensory condition. A higher response gain indicates that the participant responded more quickly to the AV stimulus, suggesting more efficient multisensory integration.

After identifying the significant windows in which the neural responses to AV and unisensory stimuli diverge, the focus shifts to quantifying the EEG differences within these windows for each participant. In each significant time window, the analysis looks at how the EEG signal differs between the two conditions, isolating the 50 consecutive time points where the average difference in neural activity is the highest. For each participant, a rolling average of EEG differences is computed within the significant windows, and the section of time that shows the highest average difference is selected. This ensures that the analysis is centered on the periods where the brain exhibits the most pronounced neural contrast between the two stimulus conditions.

Response Gain Calculation and Correlation Analysis

The trial-averaged minimum response time from the unisensory conditions is subtracted from the trial-averaged response time for the multisensory condition to determine the response gain for each participant. In comparison to the unisensory condition, this value indicates the speed advantage (or disadvantage) of responding to the multisensory condition. Next, these response gains are correlated to the differences in neural activity found in the EEG data during significant windows.

Correlating the EEG metrics with each participant's response gain is the next step. The Pearson correlation coefficient, a statistical indicator of the linear relationship between two variables, is used to calculate the correlation. In this instance, the degree to which response gain is correlated with neural differences between AV and unisensory conditions is ascertained using the Pearson correlation. This function computes the correlation matrix between two or more sets of variables, and in this analysis, it is used to compute the correlation between the EEG metrics (e.g., mean or max or min difference) and the response gain across all participants. The output is a matrix in which the off-diagonal elements represent the correlation coefficients between pairs of variables. A correlation coefficient closer to 1 or -1 indicates a strong relationship, while a value near 0 indicates little to no linear relationship.

A t-statistic is used to assess the statistical significance of the correlation in addition to the correlation coefficient. To ascertain whether the observed correlation deviates significantly from zero, the t-statistic is calculated. The formula for the t-statistic is:

$$t = r \cdot \sqrt{\frac{n-2}{1-r^2}}$$

where r is the Pearson correlation coefficient and n is the number of participants. The t-statistic helps assess whether the correlation is significant given the sample size. A high absolute value of the t-statistic indicates that the correlation is unlikely to be due to chance. Once the t-statistic is calculated, the corresponding p-value is derived, which determines the statistical significance of the correlation. A small p-value (typically less than 0.05) indicates that the correlation is statistically significant.

Each individual as well as specific subgroups, such as integrators versus non-integrators or clinical groups, such as participants with ASD, CN and SCZ, are subjected to the correlation analysis. This makes it easier to look into any differences that might exist between different groups' relationships between brain activity and behavioral performance. For each group, an independent correlation between the response gain and the EEG parameters is calculated. This allows for a more accurate determination of whether specific brain regions represent response patterns among the participant groups. Using the Pearson correlation and the t-statistic and p-value, the analysis ascertains which neural features from the significant time windows are most strongly associated with response gain. This sheds light on the ways in which brain activity during multisensory processing affects behavior.

3.4 EEGNet implementation

With this analysis, which is an ongoing work, we are currently not trying to link the neural dynamics to the behavioral results, but to capture distinct neural signatures, if any, for each group (CN, AS and SCZ) based on the AV trials of our task. To do that we implemented EEGNet is a compact convolutional neural network (CNN) architecture designed primary for EEG-based brain-computer interfaces (BCI). It is designed to work efficiently in a range of BCI paradigms, manage sparse data availability, and generate features that can be understood neurophysiologically (Lawhern et al., 2018). EEGNet is well-suited for use in clinical research and neuroscience because of its structure, which allows it to capture both the temporal and spatial aspects of EEG signals.

The architecture of EEGNet consists of a series of convolutional layers that are optimized for extracting relevant features from EEG data. Initially, the model applies temporal convolutions to the EEG signals to capture frequency information above a certain threshold. The network is able to isolate specific frequency bands from the EEG data thanks to this temporal filtering step, which is essential for comprehending the underlying neural dynamics. EEGNet uses depthwise convolution to learn spatial filters associate with the temporal ones. Because it allows the network to recognize spatial patterns linked to each temporal feature, this operation is especially helpful in EEG applications as it helps the network learn the relationships between the channels in the EEG data and temporal profiles. In comparison to conventional convolutional layers, depthwise convolution has fewer trainable parameters, which helps avoid overfitting, particularly when working with small datasets. After the depthwise convolution, separable convolution is used to further process the data. Separable convolution consists of a depthwise convolution followed by a pointwise convolution, which allows the network to first learn how to summarize each feature map individually and then optimally merge the outputs. This separation of feature extraction into two distinct operations helps EEGNet handle the complex, multi-dimensional nature of EEG signals more efficiently by decoupling the spatial and temporal relationships in the data.

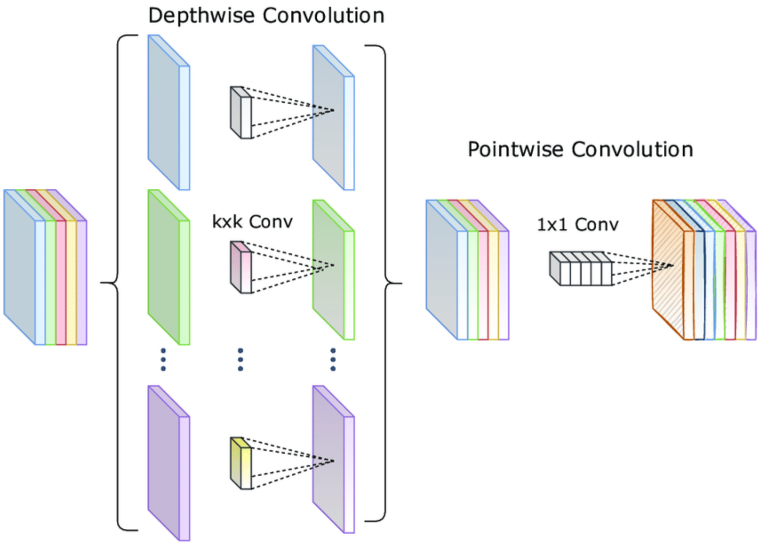


Figure 6: Depthwise separable convolutions (Sultonov, F., 2024)

To regularize the model and improve generalization, techniques such as Batch Normalization and Dropout are employed. While Dropout randomly omits a portion of the connections during training to lower the risk of overfitting, Batch Normalization standardizes the outputs of the convolutional layers to stabilize the learning process. A softmax classification layer, which outputs the probabilities of the various classes in the data, is the last stage in the EEGNet architecture. This classification layer is directly connected to the feature maps generated by the previous layers, which reduces the number of free parameters and helps the model remain lightweight and computationally efficient.

All things considered, EEGNet's architecture, which combines regularization methods, depthwise and separable convolutions, and temporal and spatial filtering, makes it an effective all-in-one tool for analyzing EEG data. Its high accuracy and interpretability in identifying pertinent patterns in brain signals makes it appropriate for a variety of BCI applications as well as exploratory EEG neural activity analysis (Lawhern et al., 2018). The figure below shows how the model is designed.

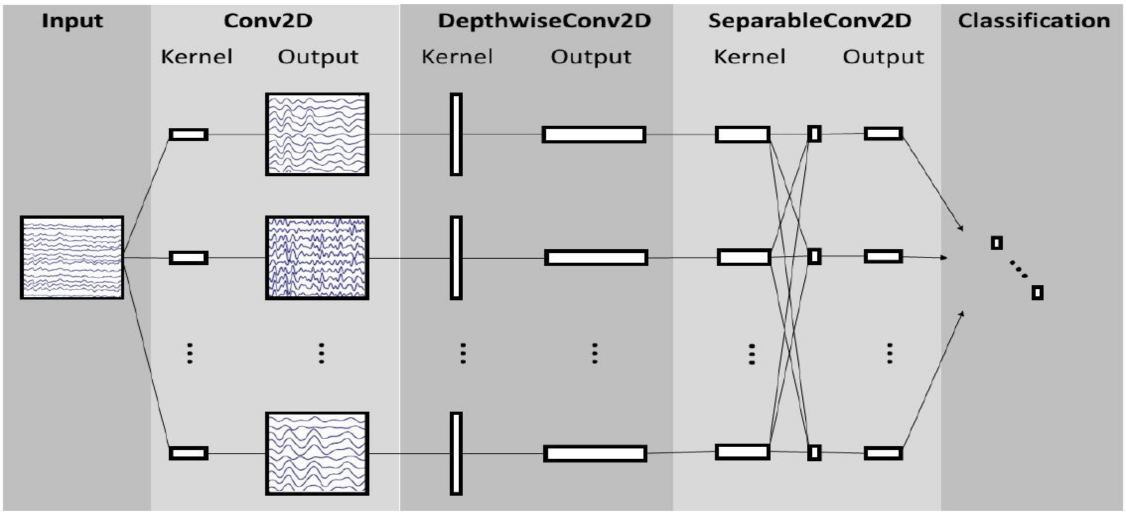


Figure 7: The EEGNet architecture. The convolutional kernel connection between inputs and outputs, sometimes referred to as feature maps, is shown by lines. The network first learns frequency filters using a temporal convolution (second column), and it then learns frequency-specific spatial filters using a depthwise convolution (middle column), which is connected to each feature map separately. The combination of a depthwise convolution, which learns a temporal summary for each feature map separately, and a pointwise convolution, which learns how to mix the feature maps together best, results in the separable convolution (fourth column) (Lawhern et al., 2018)

To delve deeper into explaining the function of the model we will discuss each layer's input and output separately. Each input is reshaped to a size of (1, C, T), which corresponds to one sample, C channels, and T time points. The model processes inputs in batches. The model's first layer is a 2D convolutional layer that uses filters of size F1×1×Fs/2, where Fs is the sampling rate, to extract temporal features. According to the Nyquist theorem, which states that the sampling frequency must be at least twice the highest frequency component in the signal to prevent information loss, the filter size is halved (or further reduced if necessary) to ensure anti-aliasing.

The output of this first layer has dimensions (F1, C, T), where F1 is the number of temporal filters, chosen as a hyperparameter. The second layer is a depthwise separable convolutional layer that learns D spatial filters with size C×1, producing an output size of (D×F1, 1, T). This depthwise convolution layer offers several benefits. Firstly, because its neurons are not fully connected, it significantly reduces the number of learned parameters, thereby decreasing the risk of overfitting—particularly advantageous when dealing with relatively small EEG datasets. Additionally, the layer's ability to learn specific spatial filters tied to corresponding temporal filters from the previous layer enhances interpretability of the model. Each spatial filter is tailored to a particular temporal pattern, making the visualization of learned features more intuitive. The final layer is a separable convolutional layer that aggregates the information from the previous layers, combining spatial and temporal filters. In this layer, F2=D×F1 pointwise filters of size (1,1) are used to merge the learned spatial and temporal features. This results in an output of size (F2, 1, T//4), where average pooling is applied to reduce the number of time points. Further down the network, additional pooling is applied to average over time, leading to a flattened output size of (F2, T//32) which is then passed into a softmax layer with N neurons, where N is the number of classes in the classification task. This final layer generates the model's predictions.

Block	Layer	# filters	size	# params	Output	Activation	Options
1	Input				(C, T)		
	Reshape				(1, C, T)		
	Conv2D	F_1	(1, 64)	$64 * F_1$	(F_1, C, T)	Linear	mode = same
	BatchNorm			$2 * F_1$	(F_1, C, T)		
	DepthwiseConv2D	D * F_1	(C, 1)	$C * D * F_1$	(D * F_1, 1, T)	Linear	mode = valid, depth = D, max norm = 1
	BatchNorm			$2 * D * F_1$	(D * F_1, 1, T)		
	Activation				(D * F_1, 1, T)	ELU	
	AveragePool2D		(1, 4)		(D * F_1, 1, T // 4)		
	Dropout*				(D * F_1, 1, T // 4)		$p = 0.25$ or $p = 0.5$
2	SeparableConv2D	F_2	(1, 16)	$16 * D * F_1 + F_2 * (D * F_1)$	(F_2, 1, T // 4)	Linear	mode = same
	BatchNorm			$2 * F_2$	(F_2, 1, T // 4)		
	Activation				(F_2, 1, T // 4)	ELU	
	AveragePool2D		(1, 8)		(F_2, 1, T // 32)		
	Dropout*				(F_2, 1, T // 32)		$p = 0.25$ or $p = 0.5$
	Flatten				(F_2 * (T // 32))		
Classifier	Dense	N * (F_2 * T // 32)			N	Softmax	max norm = 0.25

Figure 8: Detailed overview of each layer of the EEGNet architecture (Lawhern et al., 2018).

The table below illustrates the different ways that EEGNet has been applied to different datasets, showing that it can achieve strong performance that frequently equals or exceeds that of larger and more complex models (Lawhern et al., 2018). This demonstrates how adaptable and effective EEGNet is at managing a range of EEG data types

Table 2: Summary of EEGNet applications (Lawhern et al., 2018)

Dataset	Description	EEGNet Handling
P300 Event-Related Potential (ERP)	Stereotyped neural response to novel visual stimuli, commonly elicited using the visual oddball paradigm. Participants respond to infrequent target stimuli, producing a positive deflection 300ms post-stimulus.	EEGNet applies depthwise and separable convolutions to capture spatial and temporal features from the EEG signals, allowing for detection and classification of P300 responses across multiple participants with high accuracy in BCI tasks.
Feedback Error-Related Negativity (ERN)	Characterized by negative and positive EEG components following incorrect feedback, usually 350-500ms post-stimulus. Used in tasks like forced-choice paradigms and BCI spellers.	EEGNet processes time-series EEG data to extract class-specific discriminant features and evaluates performance on a trial-by-trial basis. The architecture is optimized for detecting ERN patterns across trials for two-class classification of feedback outcomes.
	Neural responses related to voluntary movements of hands and feet, observable along central and midline	EEGNet generalizes across different BCI paradigms such as MRCP, using convolutional layers to

Movement-Related Cortical Potential (MRCP)	electrodes. The components include readiness potential before movement and motor potential at movement onset.	detect spatial and temporal EEG features related to motor movements, facilitating the classification of movement intentions from EEG data.
Sensory Motor Rhythm (SMR)	Involves mu (8-12Hz) and beta (18-26Hz) band desynchronization over the sensorimotor cortex during actual or imagined movements. Used in motor imagery BCIs.	EEGNet captures oscillatory features (such as SMR) by applying spatial filters and convolutional operations to extract frequency-specific information. This allows robust classification of imagined movements in tasks like motor imagery.

Implementation on our Dataset

Our implementation of the EEGNet was to differentiate between AV EEG trials of participants between two groups each time: ASD versus SCZ, ASD versus CN and CN versus SCZ. This approach allows us to utilize all individual trials rather than averaging them, as was done in the decomposition analyses. Deep convolutional neural networks, such as EEGNet, are well-suited to processing large datasets, and their performance improves with an increase in the volume of data. By leveraging every trial, the model can capture the nuanced temporal and spatial variations in the EEG signals that might be lost through averaging, ultimately enhancing the network's ability to classify the data more effectively. This was a key motivation for incorporating a deep learning architecture alongside decomposition methods. By doing so, we aimed to complement the decomposition approach, allowing the deep learning model to fully exploit the richness of the data by utilizing all individual trials and capturing more complex patterns that may not be accessible through traditional averaging techniques. With this approach, our aim was to capture differences between groups by focusing on the AV trials, as these are the key trials for assessing multisensory integration across participants. Unlike the hypothesis in the SliceTCA analysis, where we compared multisensory and unisensory conditions to correlate neural differences with behavioral data, here we sought to identify significant spatial and temporal components that distinguish the groups during the multisensory task. Our goal was to uncover potential neural differences between groups specifically within the context of multisensory processing.

The input data to the neural network is structured as (time points, channels) for each individual trial, where each trial consists of EEG signals recorded across multiple electrode channels over time. For each trial, the number of time points corresponds to 1025 samples, representing the duration of the AV stimulus, while the number of channels is 61, corresponding to the electrodes used during recording. The model processes batches of 32 trials at a time, meaning that during training, the network ingests a batch of data with the shape (32, time points, channels). This batch-based input structure allows the network to efficiently handle multiple trials simultaneously, optimizing learning by processing groups of trials in parallel, while still preserving the detailed temporal and spatial information crucial for distinguishing between participant groups. In terms of hyperparameters, we selected 8 temporal filters with a kernel length of 64, to be extracted from the first convolutional layer, along with 2 spatial filters for each temporal filter from the depthwise layer, allowing the model to capture both temporal and spatial features in the EEG data.

The data was first divided into training and test sets in an 80-20 ratio to assess the performance of the model. This ensured that 80% of the data was used for training and 20% for testing. The training data was further split into training and validation sets in the same 80-20 ratio for each fold in order to conduct cross-validation within the training set using a 10-fold split. By ensuring that the model was trained and validated on various subsets of the training data, this technique reduced the possibility of overfitting, allowed for robust evaluation, and guaranteed generalization to previously unseen data.

During the 1000 epochs of training, the model's performance was monitored, and early stopping was applied to prevent overtraining. This means the training process was halted once the model's performance on a validation set stopped improving, reducing the risk of overfitting. The model's parameters, such as the learning rate and the dropout rate, were fine-tuned to optimize the classification accuracy. Finally, the trained model was used to classify EEG trials for both the ASD and SCZ groups, as well as ASD versus CN and SCZ versus CN. The performance of the classification was evaluated using several metrics, including the cross-entropy loss, accuracy, F1 scores and Area Under the Curve (AUC). After achieving satisfying performance our goal was to determine the important features that discriminate between the groups by applying the following feature visualization pipeline.

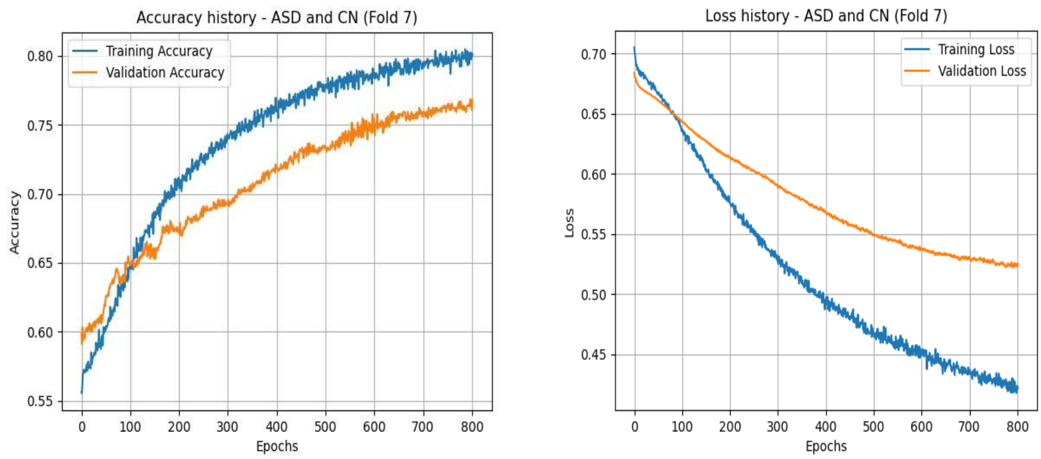


Figure 9: Example of training on a fold in the classification task ASD vs CN. Utilizing early stopping, the model stopped training after epoch 800 preventing overfitting.

Feature Visualization Using Grad-CAM

Gradient-weighted Class Activation Mapping, or Grad-CAM, is a popular visualization method for convolutional neural network predictions. It provides insights into which regions of the input contribute the most to the model's decision, thus offering interpretability to otherwise black-box deep learning models (Selvaraju et al., 2017). This is accomplished by generating heatmaps that indicate the most pertinent input regions for a given prediction by calculating the gradients of the output class score with respect to the convolutional feature maps of a certain layer. Grad-CAM is very useful for EEG analysis applications since it can aid in comprehending the spatio-temporal patterns that are essential for making classification decisions. Applying this to EEGNet, both the temporal features retrieved from the first layers and the spatial features detected by the depthwise convolutional layers that follow can be interpreted using Grad-CAM. In our case, we apply Grad-CAM to the two-group classification problem in order to interpret the temporal and spatial filters from the first two convolutional layers of EEGNet. The primary goal is to comprehend how various EEG data components affect the model's predictions.

The Temporal Convolutional Layer (First Convolutional Layer) in EEGNet is designed to extract both temporal and spatial features by learning temporal filters for each EEG channel independently. This approach provides a low-level representation that captures frequency-specific patterns, essential for obtaining interpretable features directly from the EEG signals. To enhance interpretability, we apply Grad-CAM to this layer, using the correctly predicted trials and their associated labels from the test set as input. This process allows us to focus on the patterns that contributed most significantly to correct classifications. By generating Grad-CAM heatmaps for these correctly classified trials, we identify the temporal segments that have the highest relevance across channels, distinguishing between classes. These heatmaps are visualized as time series across channels, highlighting the time points and regions of increased importance for the classification task. This method not only reveals the spatial and temporal features critical for model predictions but also supports interpretability by showcasing the segments where the model detected meaningful activity associated with specific class labels.

4. Results

4.1 Significant Tucker components for integrator classification

The results of the Tucker decomposition analysis, which was carried out to find components that can tell integrators from non-integrators independent of the groups they belong to, making it a group-agnostic analysis, are described in detail in this section. Our goal was to identify recurring themes in the data that are strongly associated with these two categories by utilizing participant factor matrices and predefined integrator labels. This approach was intended to extract components that capture variations connected to integrator status, thereby shedding light on the neural and behavioral traits that set integrators apart from non-integrators.

By methodically evaluating every possible combination of rankings up to a maximum of 5 for participants, 5 for time, 5 for channels, and 2 for stimuli, we were able to do the Tucker and non-negative Tucker decomposition analysis. We were able to determine which components were most pertinent for identifying patterns in the data after conducting a thorough investigation. We utilized the participant factor matrices from the components as inputs for additional classification analysis for each decomposition. Since greater values would result in more intricate interactions within the core tensor and make it more difficult to understand the distinct contributions of each dimension, we decided against going over these rank restrictions. By limiting our analysis to these ranks, we ensured a good balance between extracting meaningful components and maintaining the interpretability of the results, thereby facilitating the differentiation between participant groups. The variance explained by some lower and higher rank decompositions for both Tucker and non-negative Tucker are presented in the table below.

Table 3: The variance explained by the reconstructed tensor for lower and higher rank Tucker decompositions

Rank Configuration	Variance Explained
(2, 1, 1, 1)	0.084
(2, 1, 2, 1)	0.108
(2, 1, 3, 1)	0.122
(5, 5, 3, 2)	0.169
(5, 5, 4, 2)	0.178
(5, 5, 5, 2)	0.188

Table 4: The variance explained by the reconstructed tensor for lower and higher rank Tucker decompositions

Rank Configuration	Variance Explained
(2, 1, 1, 1)	0.0175
(2, 1, 1, 2)	0.0186
(2, 1, 2, 1)	0.0195
(5, 5, 5, 1)	0.0718
(5, 5, 4, 2)	0.0699
(5, 5, 5, 2)	0.0757

The process of finding decompositions and components within participant factor matrices that effectively classify integrators involved a detailed classification analysis. A number of machine learning models were employed, each with customized settings to improve performance, including LDA, Nearest Neighbors, Random Forest, and AdaBoost. Using the integrator and non-integrator labels as the goal classification variable, techniques such as leave-one-out cross-validation and K-Fold cross-validation with five splits were used to validate the efficacy of the models. For the final analysis, 5-fold cross-validation was ultimately selected. The analysis systematically explored all available decompositions, extracting the participant factor matrices from each one. For each component of the decomposition each participant vector was then used as inputs to the classification models. For each combination of component and classifier, the average classification accuracy was determined using cross-validation.

After calculating the accuracy scores for all classifiers and decompositions, the results were analyzed to find the decompositions that yielded the highest performance for each model. The most effective decompositions—those achieving the best classification accuracy—were identified for each classifier. This step made it possible to choose the decompositions that were most successful in distinguishing integrators from non-integrators. This thorough process allowed the analysis to uncover specific decompositions and participant factor components that demonstrated the highest classification accuracy, enabling us to visualize the remaining time and channel components associated with these participant ones, as demonstrated in the subsequent sections. The decompositions with the highest accuracy scores across the two best performing models (KNN and AdaBoost) are presented in the next tables.

Table 5: Accuracy score for integrators vs non-integrators of the best decompositions using the KNN model

Table: KNN Model (K=3)			
Decomposition Type	Rank Configuration	Variance Explained	Accuracy Score
Tucker	(3, 2, 2, 2)	0.148	0.689
Tucker	(5, 2, 4, 1)	0.174	0.689
Tucker	(5, 4, 4, 2)	0.189	0.667
Non-negative Tucker	(3, 1, 5, 1)	0.047	0.656
Non-negative Tucker	(5, 4, 5, 2)	0.075	0.656

Table 6: Accuracy score for integrators vs non-integrators of the best decompositions using the AdaBoost model

Table: AdaBoost Model			
Decomposition Type	Rank Configuration	Variance Explained	Accuracy Score
Tucker	(3, 4, 4, 2)	0.163	0.689
Tucker	(3, 2, 5, 2)	0.158	0.667
Tucker	(4, 1, 5, 2)	0.138	0.667
Tucker	(5, 3, 2, 2)	0.167	0.667
Tucker	(5, 2, 4, 1)	0.174	0.667

Interaction Between Components

Understanding the relationships between the participants, time, and channels—the different factors of the decomposition—is one of the main objectives of this analysis. This interaction is complex and necessitates a thorough analysis of the core tensor, which represents the combination of these elements. To see how certain components interact across the time and channel dimensions, we extract slices from the core tensor after averaging over the stimulus dimension. These interactions are visually represented through the usage of heatmaps. For instance, we can see the relationship between temporal dynamics and the activity in particular EEG channels by looking at particular core slices. With the use of these heatmaps, one can gain a better understanding of the intricate correlations that exist between spatial and temporal patterns and the activation throughout time of different brain areas, as indicated by EEG channels.

We identify the two optimal combinations of participant, time, and channel components that exhibit the strongest interactions, as shown in the heatmaps.

These interacting components play the crucial role in driving the highest classification performance.

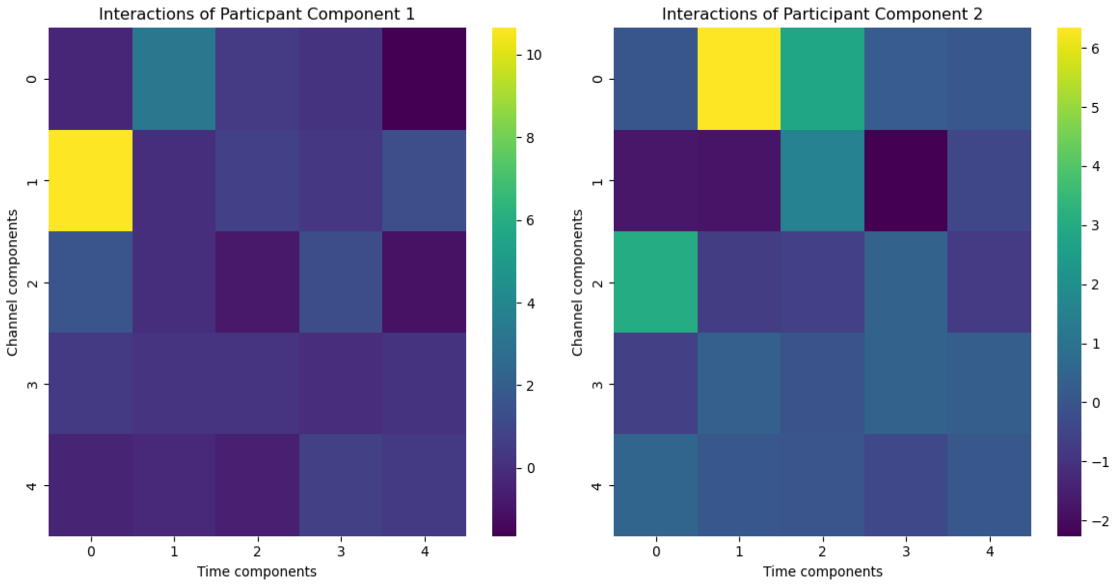


Figure 10: Heatmaps of core tensors, which depict the intensity of interactions between a given participant and the time and channel components.

Visualization and Interpretation of Components

The method ends with a visual representation of the important components found during classification. The most important factors in separating integrators from non-integrators were these elements. We can learn more about the brain dynamics that separate the two groups by looking at the temporal and spatial patterns of these components.

KNN

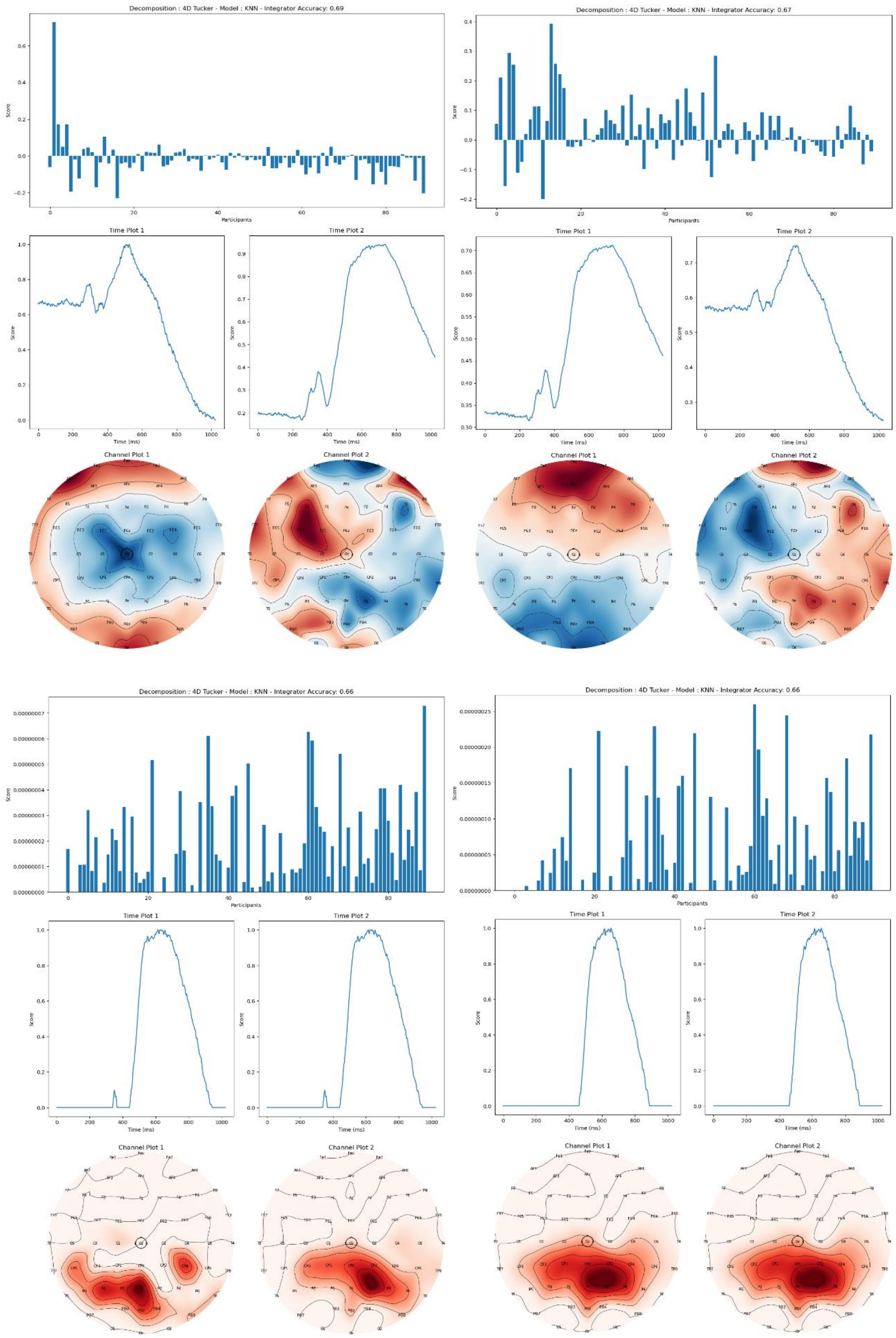


Figure 11: Significant Tucker and NN-Tucker components using KNN with K = 3 for classification

AdaBoost

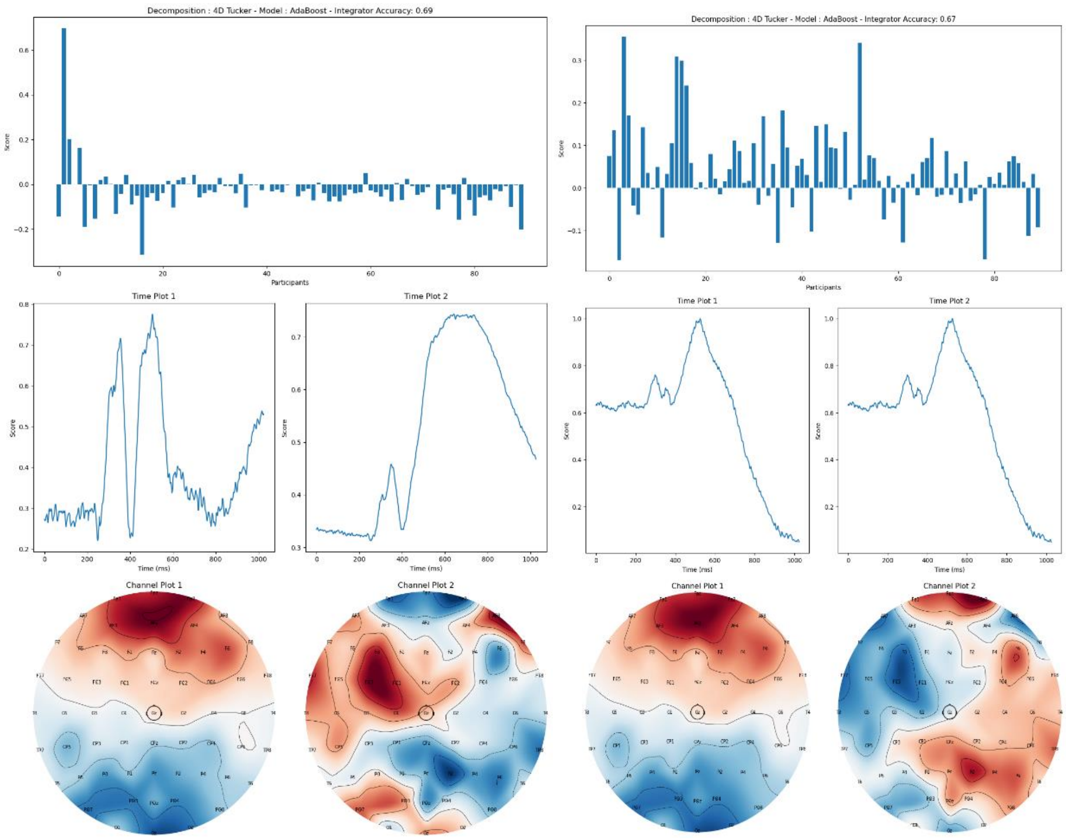


Figure 12: Significant Tucker components using AdaBoost classification

A comprehensive understanding of the neural mechanisms that may underlie the "integrator" status can be obtained by analyzing the temporal and spatial filters of the components above. The temporal profiles show a general peak around 300 ms after stimulus onset (here the stimulus is shown on time point 200) that is suggestive of stimulus evaluation and decision-making. These traits align with the neural mechanisms connected to the P300 component, which are frequently connected to activities involving sustained attention and the cognitive processing of multisensory data. These temporal patterns imply that integrators are capable of efficiently processing and assessing sensory inputs, which allows them to react to stimuli that call for integration of various sensory modalities.

The activations display both more dispersed lateralized patterns and concentrated central-parietal patterns in terms of location. The parietal cortex, which is essential for working memory, attentional control, and sensory integration, may be involved in the central-parietal activity, as it is consistent with sources of the P300 that are known to be involved. Multisensory stimuli most likely cause this region to become active, and effective integration improves cognitive performance. In addition, the lateralized and diffuse activations indicate broader recruitment of neural networks, potentially involved in sensory filtering and flexible attentional allocation.

4.2 SliceTCA Components and Correlations

SliceTCA is utilized with channel slicing to effectively disentangle the neural data into latent components that reflect distinct classes of covariability—across participants, time, and channels. This decomposition approach is particularly advantageous for analyzing neural datasets such as EEG, where capturing the structure of task-relevant neural dynamics is essential. By employing SliceTCA with a focus on channel slicing, we can reveal the shared variances specific to individual EEG channels, which in turn enhances the interpretation of neural responses and allows for a more detailed understanding of which brain regions interact during specific tasks.

The link between channel slicing components and response time gains (RT gains) is critical, as it refers to the phenomenon where responses to multisensory stimuli, like AV inputs, are faster compared to unimodal inputs, a concept central to multisensory integration studies. RT gains are observed as a decrease in reaction time when stimuli from multiple modalities are processed concurrently. By correlating SliceTCA components with RT gains, it becomes possible to discern how specific EEG channel activities relate to enhanced response speeds. This correlation not only reflects the neural synchrony but also indicates the efficiency of the brain's processing mechanisms when dealing with integrated multisensory inputs, which can provide insights into neural efficiency and adaptability.

Grid search results for channel slicing

For grid search, we followed the methodology used in the original SliceTCA paper. Using the 3D EEG tensor data, the grid search involved methodically evaluating various configurations while making sure the training to validation data ratio was maintained at or around 80%. This was accomplished by controlling the tensor's selection of blocks for the training and validation sets by adding the proper masks. The objective was to maximize the model's performance while guaranteeing that the training and validation data were proportionately represented. We performed grid search on the channel slicing decomposition for both SliceTCA and NN-SliceTCA. For both types, the optimal number of channel slices turned out to be 4. More components led to overfitting the data and therefore higher validation set loss. The loss function used is the MSE loss between the original tensor and the reconstructed one from the channel components.

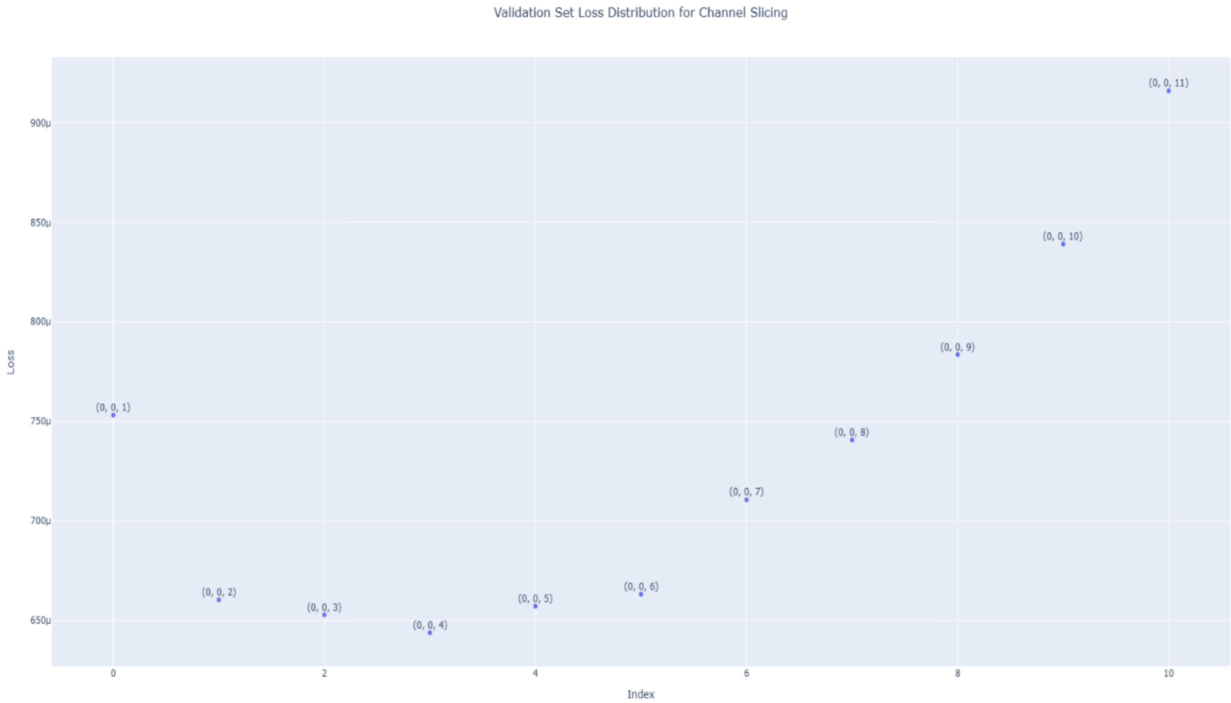


Figure 13: MSE loss for SliceTCA decomposition using 1-11 channel ranks

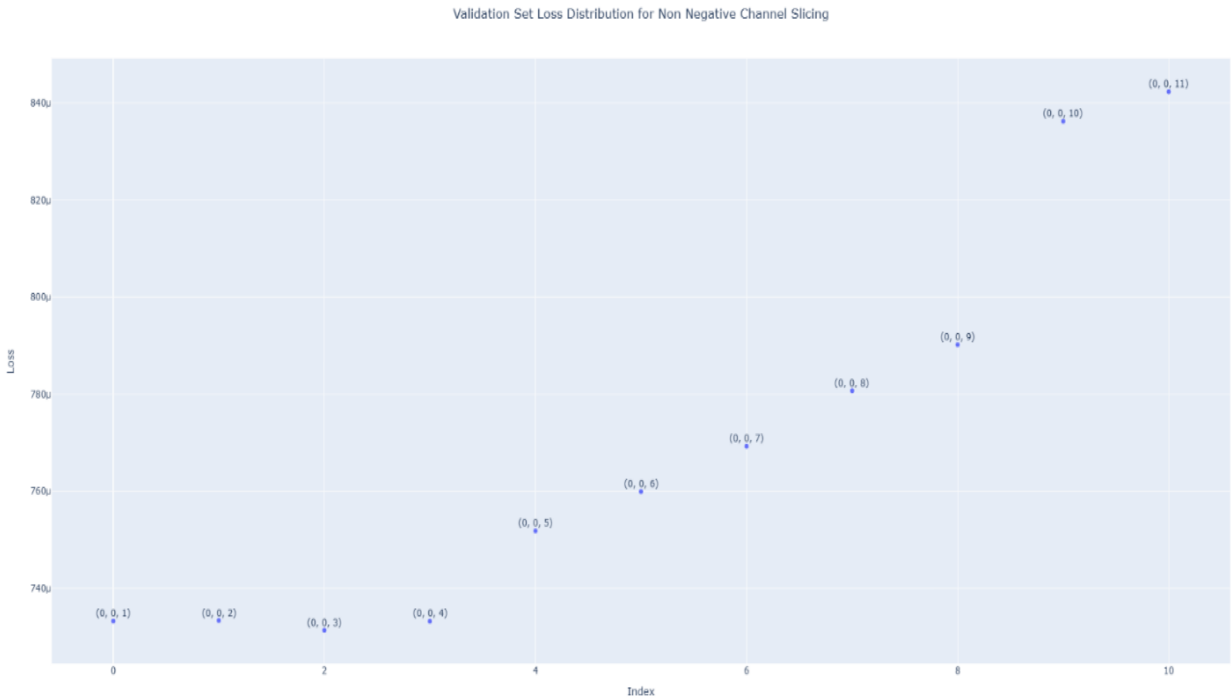


Figure 14: MSE loss for NN- SliceTCA decomposition using 1-11 channel ranks

Component comparison of different decomposition types

Here, we examine the differences between SliceTCA and NN-SliceTCA using our EEG data, with an emphasis on channel slicing for intuitive results. SliceTCA breaks down neural data into discrete patterns that span participants, channels and time. A non-negativity constraint is added by NN-SliceTCA, which is helpful for interpreting purely additive neural activations by emphasizing only positive contributions. To evaluate how well each method handles various data characteristics, we applied these methods to the entire dataset and NN-SliceTCA subsets that only included positive or negative data points.

Several important insights are revealed when the components of various decompositions are compared. First, the components look very similar when using NN-SliceTCA with all data as opposed to just positive data points. The difference is mostly in the intensity of the signal. When using all data points the intensity is slightly higher, which can be explained by the inverted negative values adding to it. This suggests that there is no discernible cancellation effect between positive and negative values, and that NN-SliceTCA accurately captures real variance within the data. On the other hand, when negative data points are inverted and then subjected to NN-SliceTCA, new, interpretable components emerge that are not seen when the full dataset is used. Furthermore, when using SliceTCA without non-negativity constraints but within the invariance pipeline discussed, distinct components emerge, capturing additional structures. Certain components, for example, appear to represent auditory processing, which is not present in other decompositions. These results highlight how crucial it is to use various SliceTCA variations in order to guarantee the robustness of the outcomes and to fully capture the diversity of data structures.

SliceTCA components

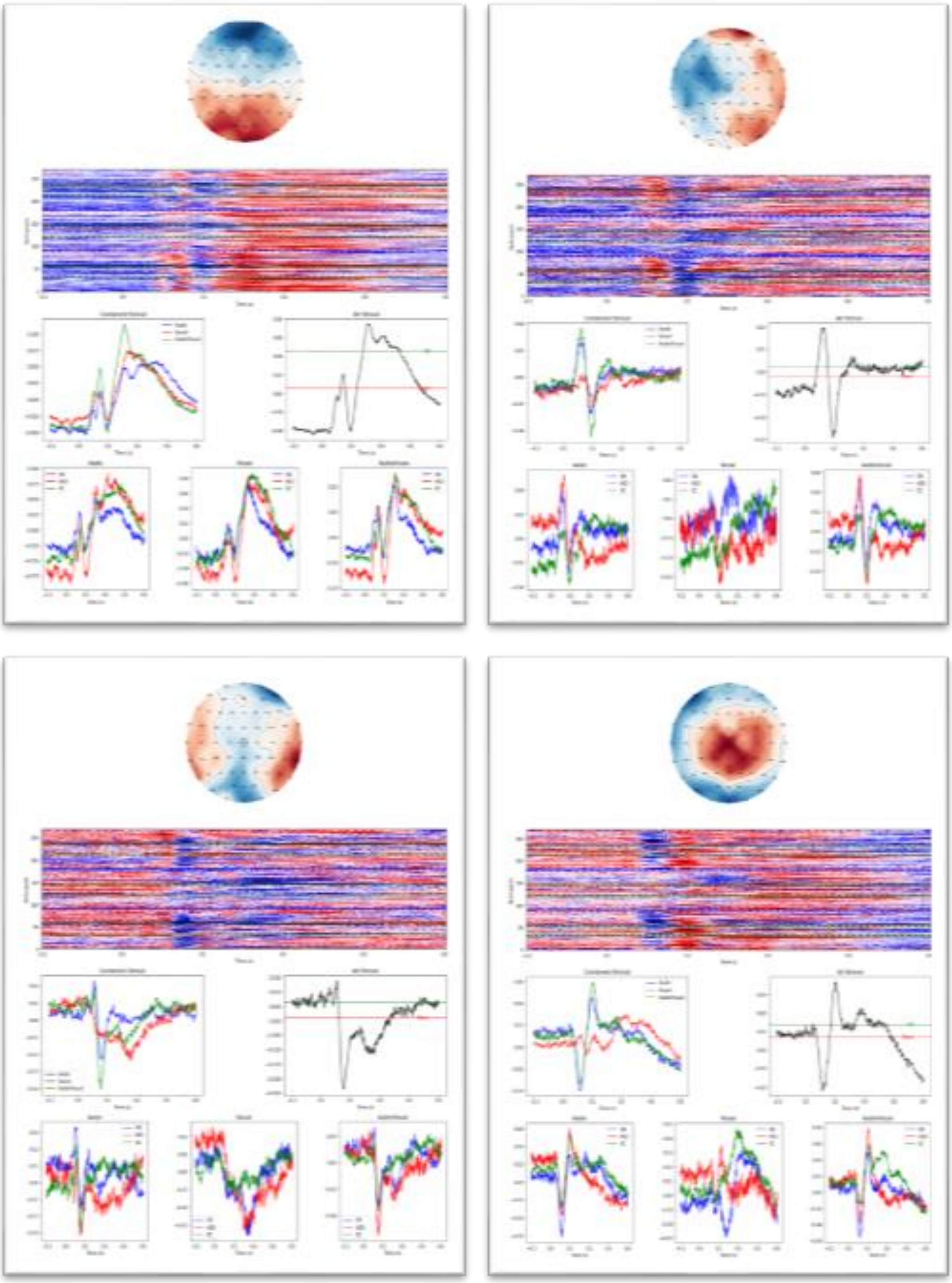


Figure 15: The 4 components for SliceTCA. The spatial topographies reveal varied activations across frontal, central, and parietal regions, with distinct lateralized patterns. Frontal and central areas show strong engagement, suggesting involvement in attentional control and cognitive processing. In images 2, 3, and 4, there is prominent parietal lobe activation, which aligns with auditory sensory processing, indicating that these components may reflect integration of auditory stimuli.

NN-SliceTCA components for positive data points

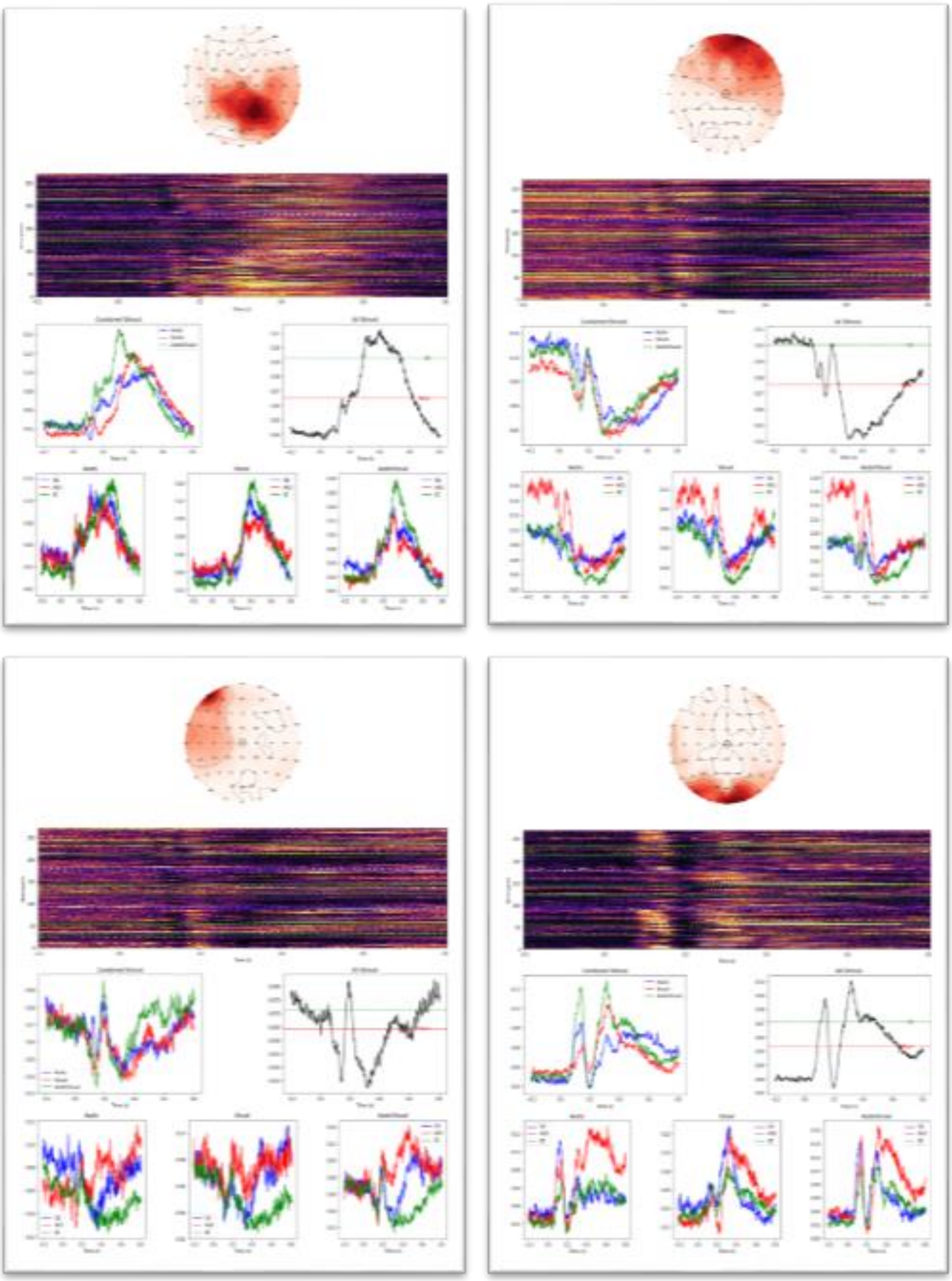


Figure 16: The 4 components for NN-SliceTCA using only the positive EEG data. The spatial topographies across the images show varying activation patterns primarily in frontal, central, and parietal regions, with each component displaying unique distributions. The first image exhibits strong central-parietal activation, suggesting involvement in attention and sensory processing. Subsequent images display more lateralized or diffuse patterns, indicating engagement of diverse brain areas across different tasks or conditions.

NN-SliceTCA components for inverted negative data points

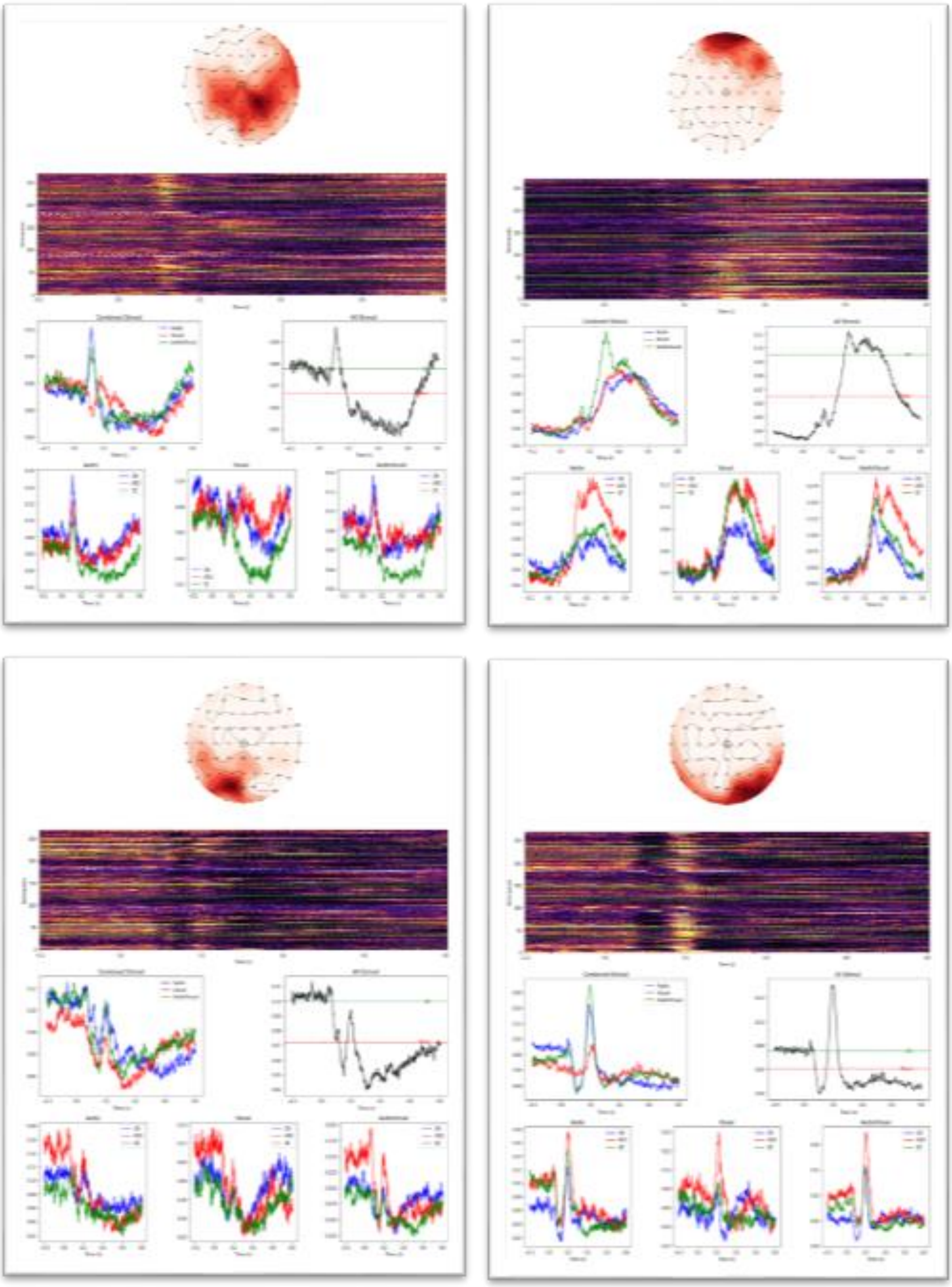


Figure 17: The 4 components for NN-SliceTCA using only the negative EEG data after inverting their sign. The spatial topographies show varying activation patterns across central-parietal, frontal, and lateralized regions, indicating diverse neural processes. Central-parietal activations suggest sensory integration and attention, while frontal engagement reflects executive functions like decision-making. Lateralized and diffuse parietal patterns indicate specialized sensory processing (A and V).

NN-SliceTCA components

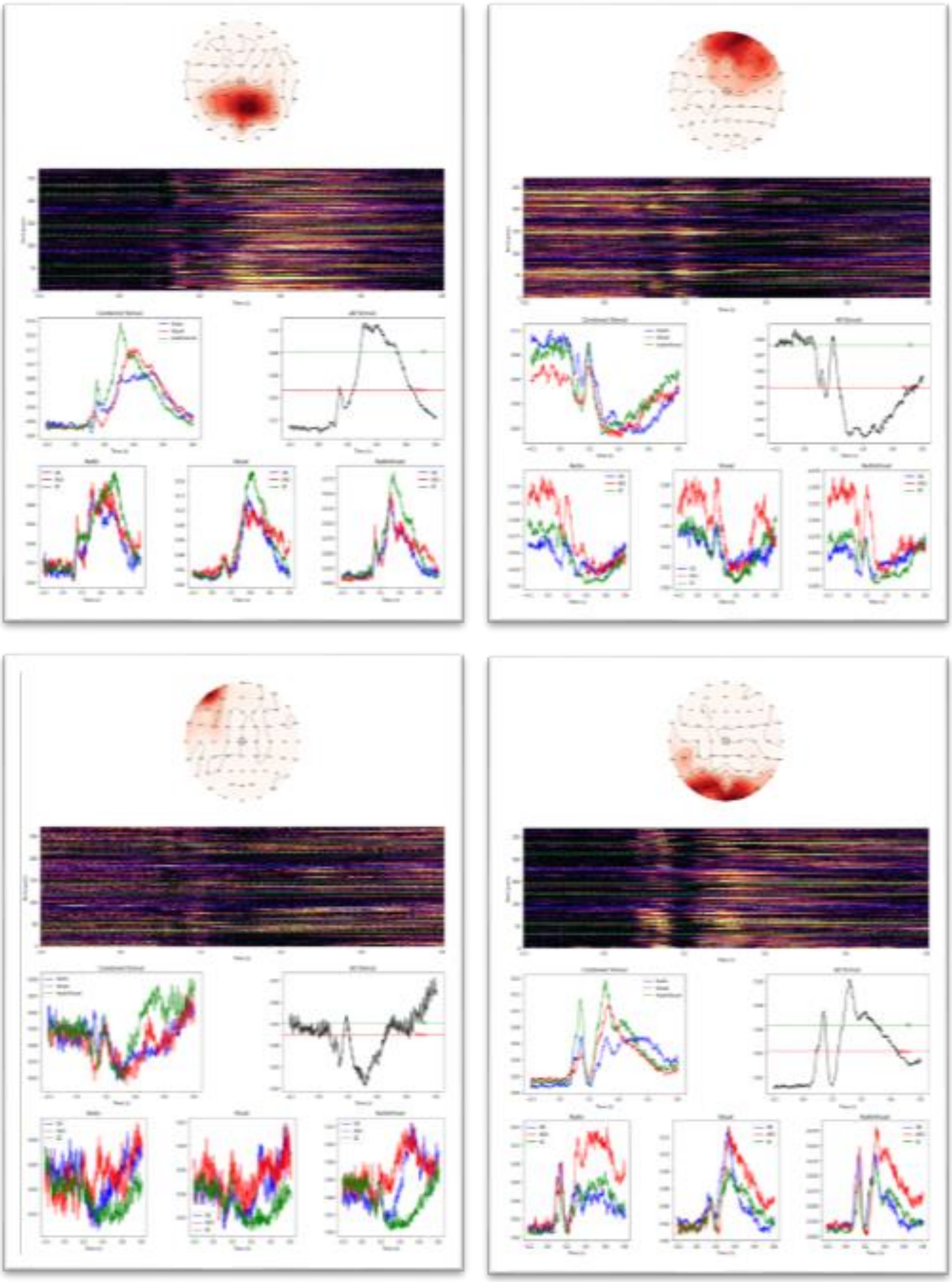


Figure 18: The 4 components for NN-SliceTCA. The components are almost the same as those from Figure 16.

The four components of NN-SliceTCA and SliceTCA show different but meaningful spatial topographies, demonstrating the structured neural processes and diversity that each method captures. Different frontal, central, and parietal regions show different levels of activation across SliceTCA components. Notable lateralized patterns indicate roles in attentional control, cognitive processing, and auditory sensory integration. Similar to this, although with different distributions, the NN-SliceTCA components for both positive and negative EEG data highlight distinct activation patterns across related brain regions. The NN-SliceTCA positive components highlight central-parietal regions for attention and sensory processing, while more diffuse and lateralized patterns point to broader engagement across tasks. Conversely, the negative EEG data in NN-SliceTCA exhibits central-parietal and frontal activations that align with sensory integration, attention, and executive functions. Together, these findings underscore the diversity in neural components identified by SliceTCA and NN-SliceTCA, while also revealing consistent spatial structures that reflect underlying cognitive processes, sensory integration, and attention mechanisms.

Bootstrap methodology for identifying significant windows

This section describes a bootstrap-based methodology that we use to find time windows where there are notable differences in brain activity between stimuli that are unisensory (A or V) and multisensory (AV). We extract time-specific neural responses by using SliceTCA to decompose the temporal EEG signals, and we compare these responses across conditions using a sliding window method. To create a distribution of possible differences for each window, we use non-parametric bootstrap sampling, which makes it possible to estimate p-values and confidence intervals. We can identify critical periods of enhanced or diminished multisensory integration using this method, which allows us to pinpoint precise time intervals where neural processing of multisensory integration differs significantly from the stronger unisensory responses. The selection of significant time intervals becomes clear in the following figures.

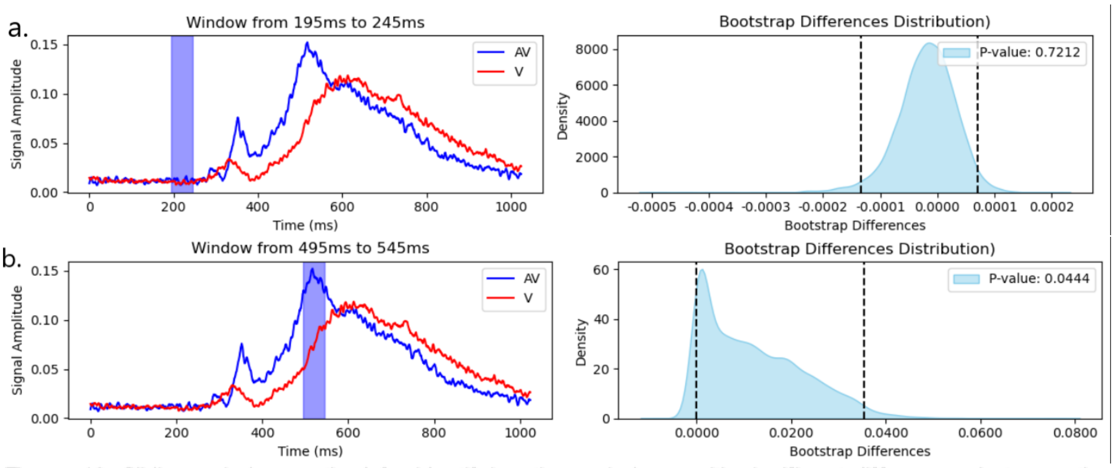


Figure 19: Sliding window method for identifying time windows with significant differences between the multisensory and the stronger unisensory (in this case V) stimulus. (a) Insignificant time interval based on the bootstrapped difference distribution. (b) Significant time interval

The following plots showcase all the components from SliceTCA and NN-SliceTCA where significant windows were found. We find significant windows in the following components that correspond to early ERPs like the N100 and P100, which are important for first stimulus detection in response to visual and auditory inputs. Compared to the auditory or visual stimuli alone, the audio-visual condition exhibits a notably stronger P300 component. The P300's importance in the AV case draws attention to its possible use in combining multiple senses at once. Additionally, some significant windows appear later in the signal, potentially indicating more complex cognitive processes such as response selection or decision-making that extend beyond primary sensory processing. These subsequent elements might be the result of ongoing brain activity connected to the processing and interpretation of multisensory information.

SliceTCA components

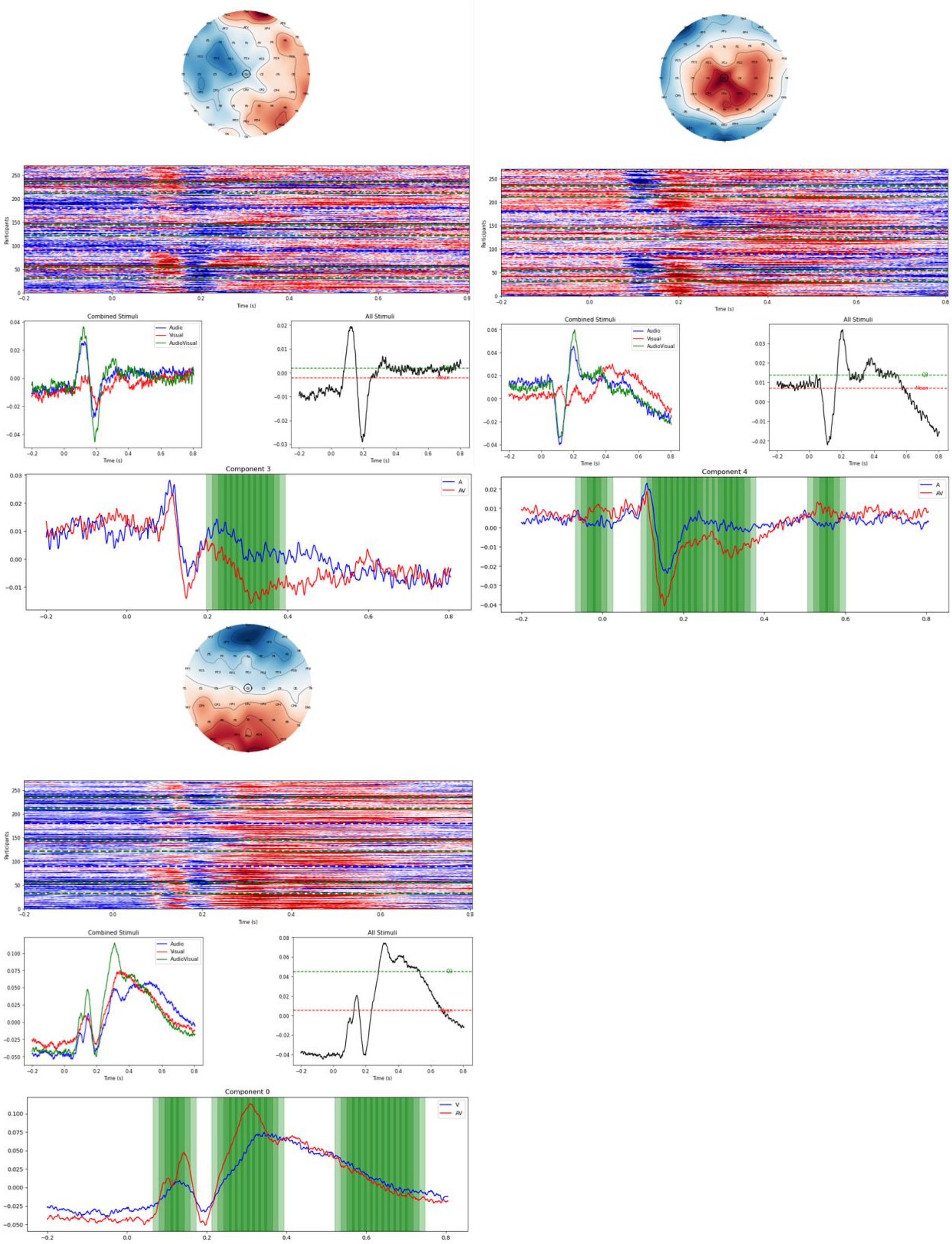


Figure 20: Significant windows for SliceTCA components. In the second image, we can observe a N100 auditory component, which is typically seen around 100ms post-stimulus. This component is often associated with early auditory sensory processing and is visible here, aligning with the timing of auditory stimuli. In the third image, the green windows seem to capture both audio and visual P100 components — an early component in response to sensory stimuli, particularly visual and auditory, which occurs around 100ms. Additionally, a P300 component is present, typically appearing later (around 300ms or more) and linked to cognitive processes like attention and stimulus evaluation. This suggests a more complex processing stage, where both sensory and cognitive processing are active.

NN-SliceTCA components

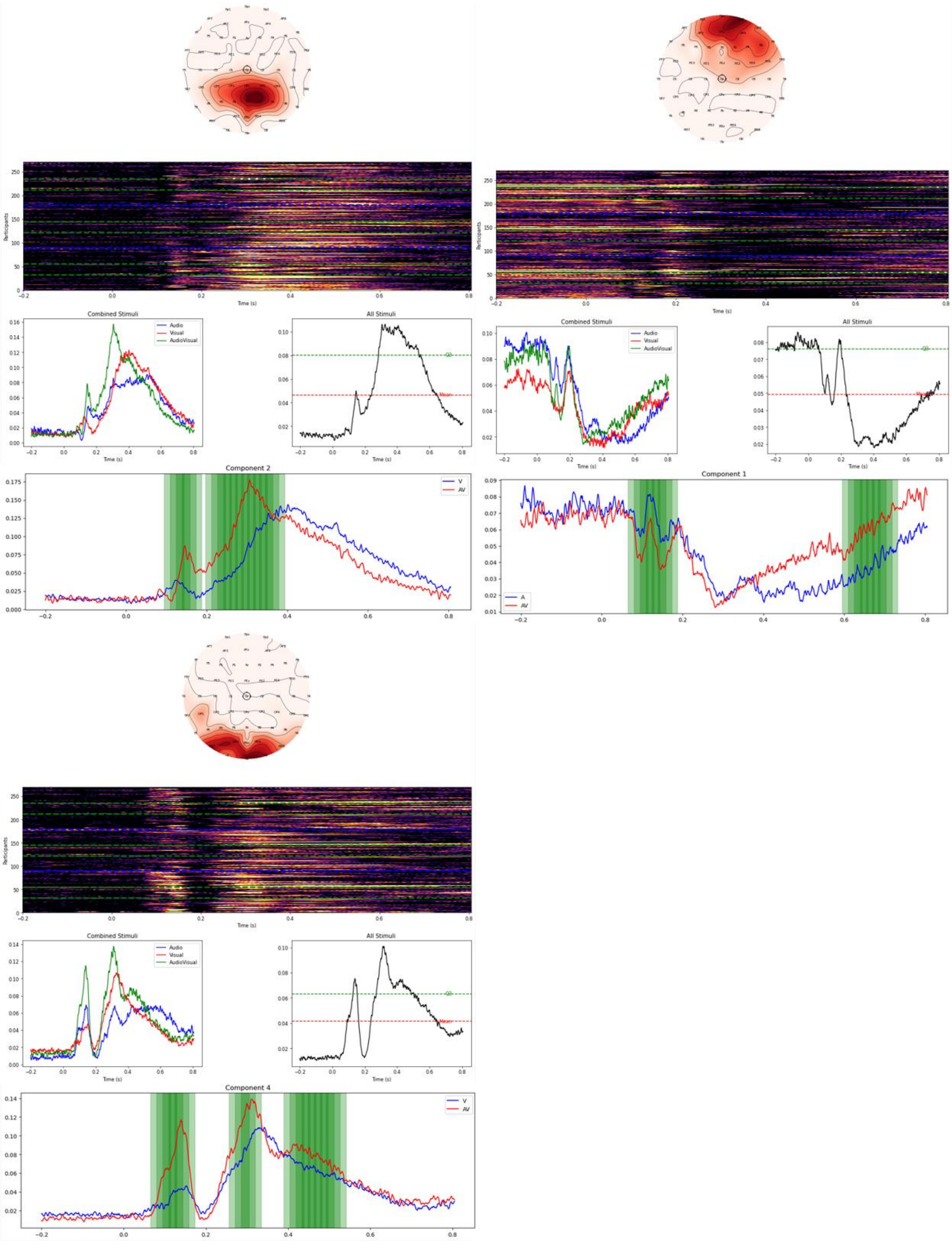


Figure 21: Significant windows for NN-SliceTCA components. We are able to discern windows that correspond with stimulus detection ERPs once more using this decomposition, in addition to the P300 ERP. Processing at a later stage is also captured.

The significant windows in Figures 20 and 21, identified between the AV and stronger unisensory conditions, suggest enhanced neural engagement for the AV stimuli. The presence of early components, such as the N100 and P100, alongside the later P300 component, indicates a combination of rapid sensory and later cognitive processing. This enhanced processing during AV stimuli potentially supports quicker responses, a hypothesis we aim to test further in our subsequent correlation analysis.

Correlation between EEG differences and response gain

This section examines the relationship between behavioral response gain and EEG differences in the three participant groups separately and collectively. The hypothesis is that the EEG intensity difference between responses to multisensory and unisensory stimuli in certain time ranges could explain the speed gain in button pressing. We assess the relationship between these neural differences and participants' responses to multisensory stimuli using the EEG temporal profiles obtained from SliceTCA decomposition. Response gain is computed by comparing participants' trial-averaged reaction times to multisensory stimuli (AV) with their reaction times to the stronger unisensory condition (either A or V alone). The response gain quantifies the improvement in reaction speed when both sensory modalities are presented together, indicating the efficiency of multisensory integration.

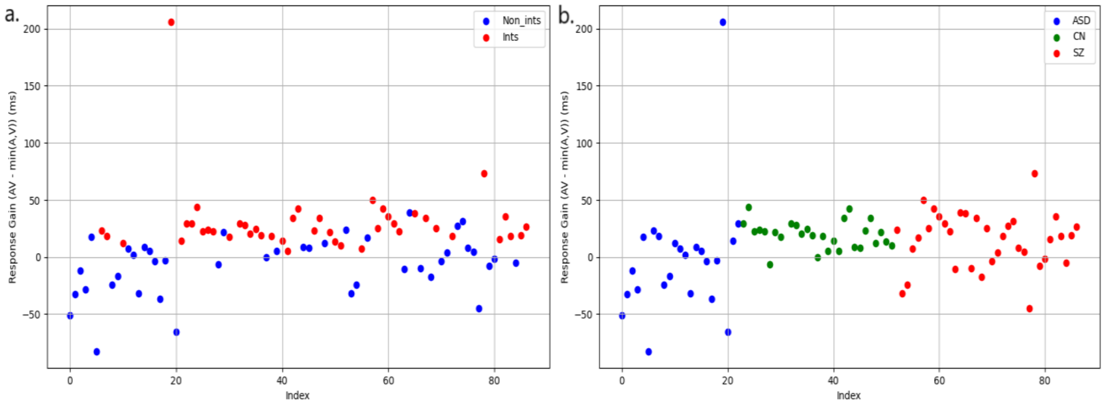
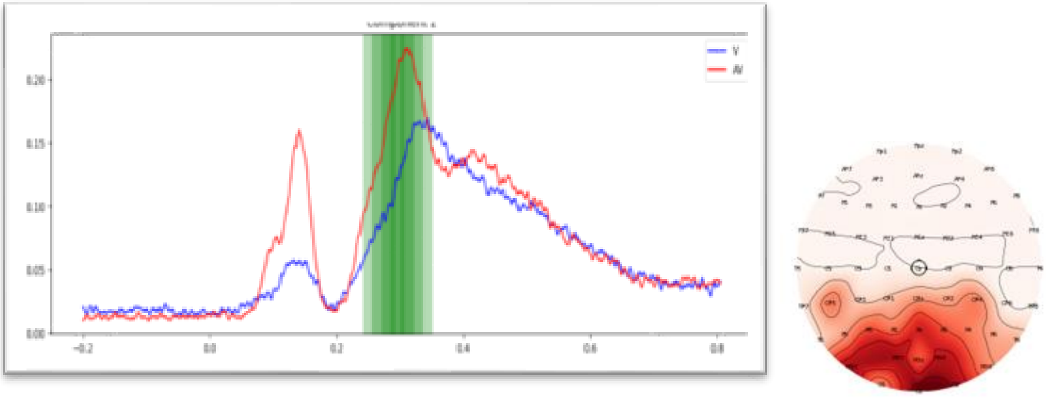


Figure 22: Response gain after averaging over trials for all 90 participants. (a) Labels based on "integrator status". We can see that the general trend shows that the integrators have bigger response gains, meaning they get faster during the AV stimulus. (b) Labels based on groups. The ASD group shows smaller or negative response gains which agrees with the small rates of integrators in the group. CN groups shows the smallest variability and 79% is classified as an integrator. SCZ group lies in between with 49% being classified as integrators

We apply a t-statistic method to evaluate the statistical significance of the correlation between EEG differences and response gain. The correlation coefficient and the number of participants are used to calculate the t-statistic for each correlation, which gives an indication of the likelihood that the observed correlation deviates from zero. The significance of these correlations is then ascertained by computing p-values, which allows us to identify the critical brain regions that are substantially linked to improved multisensory processing. We can find significant patterns in neural dynamics across various groups and conditions thanks to this thorough approach.



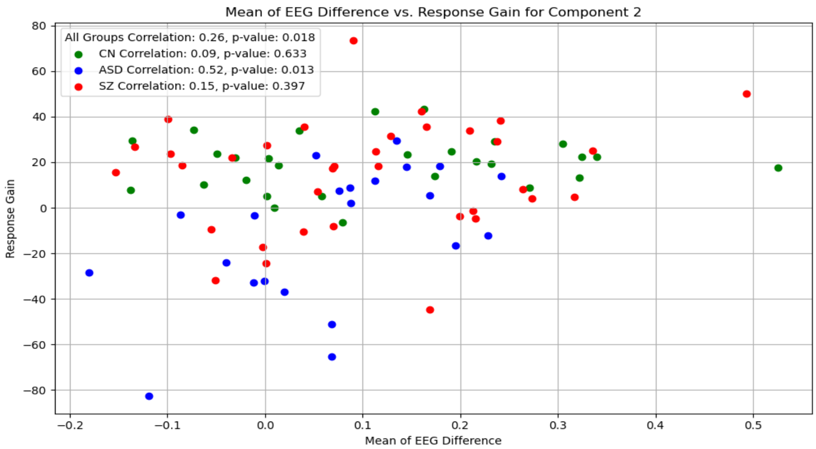


Figure 23: Significant window that correlates across all participant groups in the non-negative SliceTCA decomposition . Taking groups separately it correlates with the ASD group. The significant temporal window corresponds to the P300 cognitive ERP.

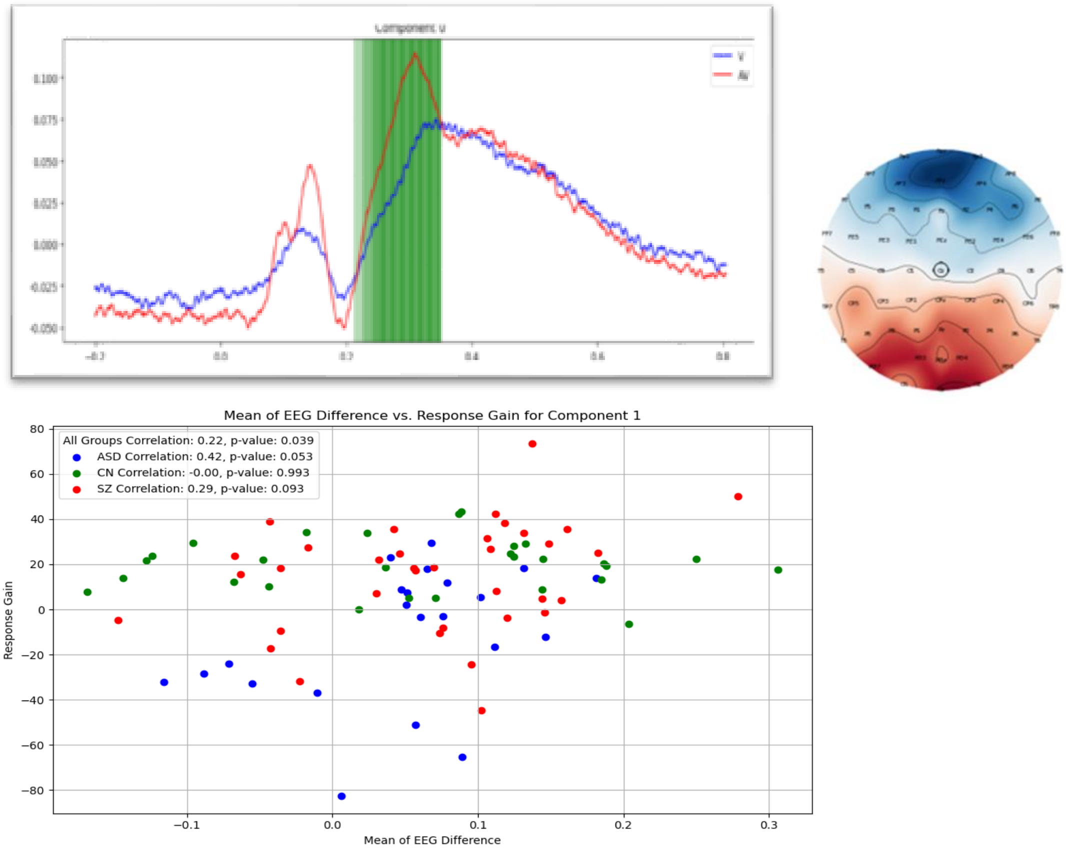


Figure 24: The same component from the SliceTCA decomposition. This time the SCZ group is also more correlated

We identified the windows that displayed significance for all three population groups, allowing us to suggest that this time interval could represent a neural mechanism related to response gain. For the window in Figure 23 and 24, which corresponds to the P300 cognitive ERP, we observe that the ASD participants influence the correlation the most followed by the SCZ group. Our positive correlation finding suggests that this ERP component may be associated with response gain in our multisensory task The P300 may indicate improved processing efficiency in multisensory environments, such as those with visual and auditory stimuli, as the brain may combine data from various sensory channels to create a cohesive perceptual experience.

On the other hand, as the correlation was prominent mostly in the ASD group, which exhibited the highest degree of variability in response gains, it may also suggest that the increased P300 response could be attributed to compensatory mechanisms for this specific group that facilitate faster response times when visual and auditory stimuli are presented simultaneously.

4.3 EEGNet Results

In the context of distinguishing CN, ASD, and SCZ groups, EEGNet will be applied to AV stimulus trial data to capture neural dynamics specific to each group. These patterns could be neural processing strategies during multisensory integration. By leveraging EEGNet's capacity for classification and feature extraction, this method seeks to provide a nuanced understanding of the neural correlates associated with these disorders, contributing to a more comprehensive model of multisensory processing across neurodivergent populations.

Model performance
To begin with, we present accuracy across each fold for each of the three models (ASD vs CN, ASD vs SCZ and ASD vs SCZ) the. We also made use of the average ROC curve over folds plotted for the test set, which is particularly useful for interpreting imbalanced dataset results. By focusing on the true positive rate versus the false positive rate across different thresholds, the ROC curve provides a more comprehensive evaluation of the model's performance, ensuring that the classification accuracy is not skewed by the class imbalance present in the data.

<u>ASD vs CN</u>

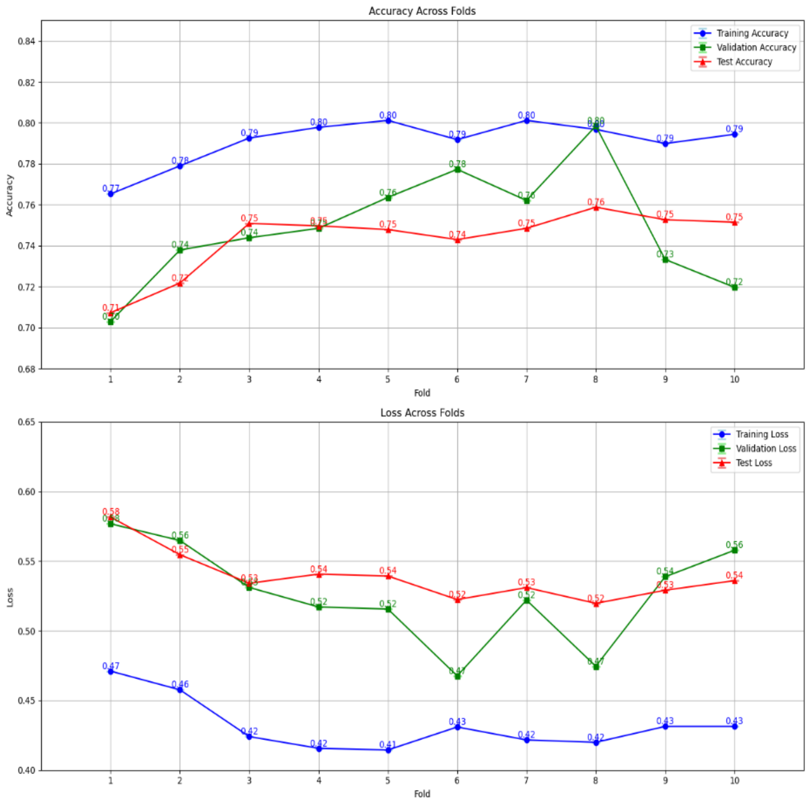


Figure 25: Accuracy and training loss across the 10 folds for ASD vs CN

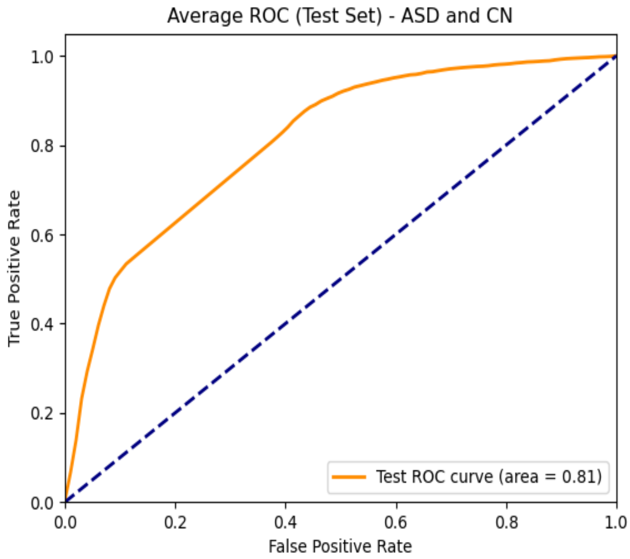


Figure 26: Average ROC for the test set of ASD vs CN

ASD vs SCZ

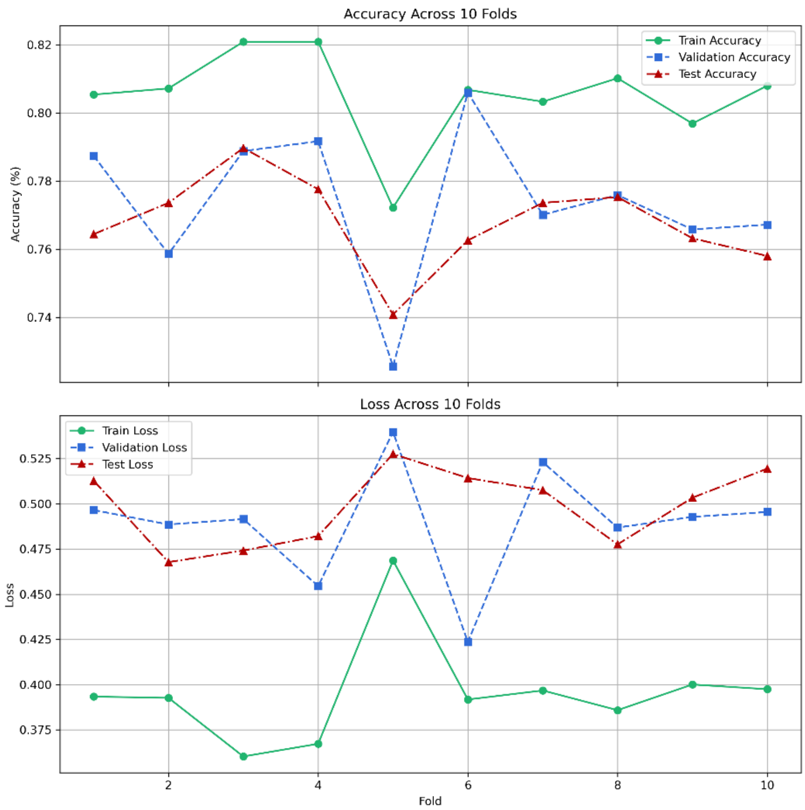


Figure 27: Accuracy and training loss across the 10 folds for ASD vs SCZ

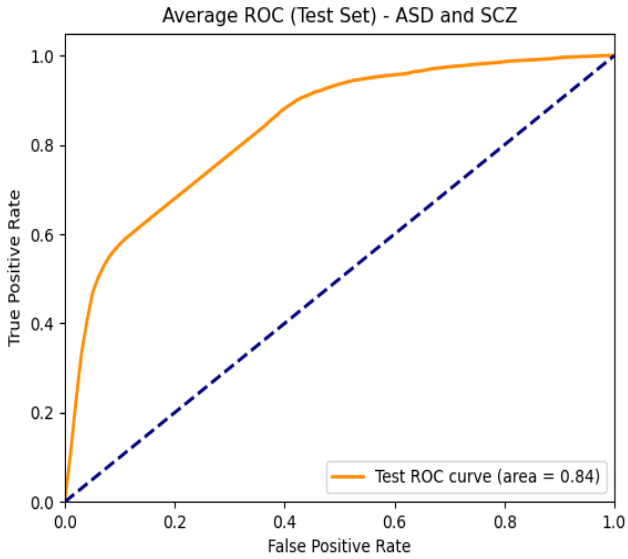


Figure 28: Average ROC for the test set of ASD vs SCZ

CN vs SCZ

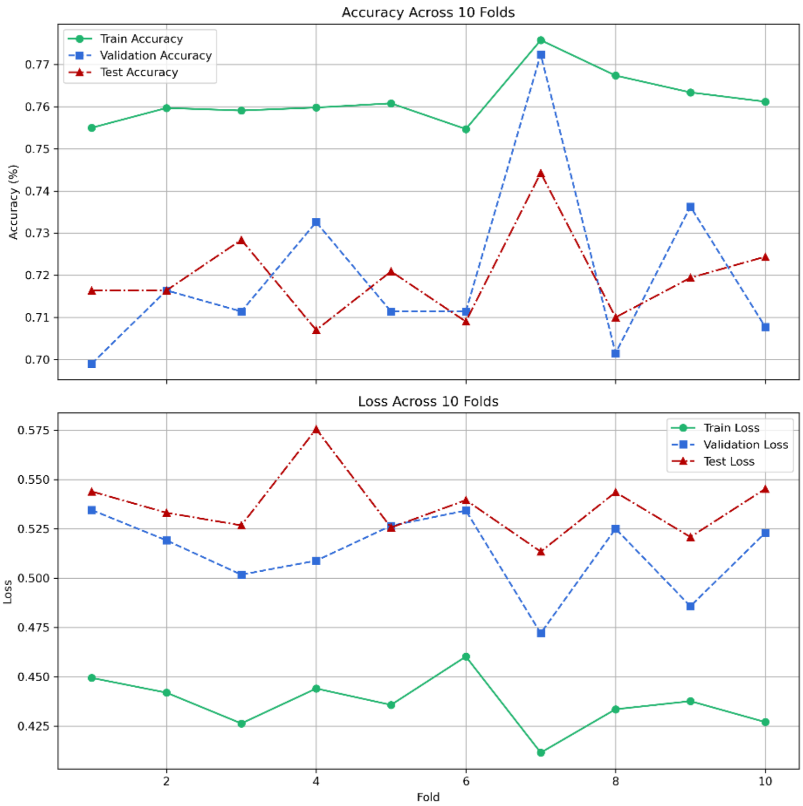


Figure 29: Accuracy and training loss across the 10 folds for CN vs SCZ

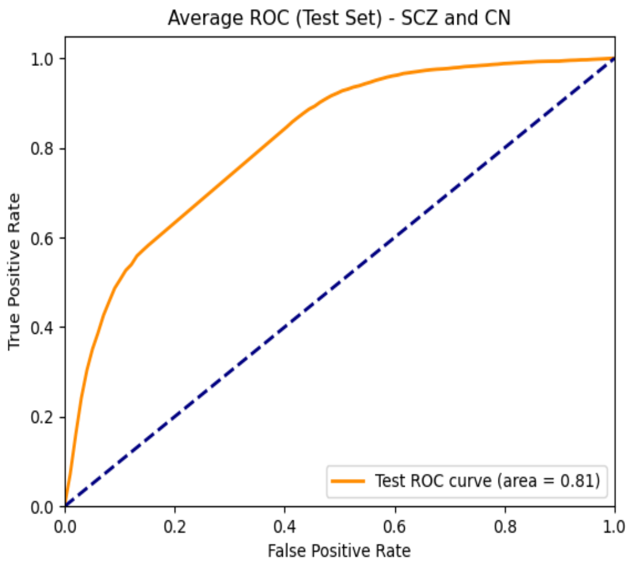


Figure 30: Average ROC for the test set of CN vs SCZ

Table 7: Summary performance table for the 3 classification models

Model	Average Accuracy	Average AUC
ASD vs CN	0.75	0.81
ASD vs SCZ	0.77	0.84
CN vs SCZ	0.72	0.81

The model demonstrates strong classification capabilities in distinguishing between ASD, CN, and SCZ groups, particularly in differentiating ASD from both CN and SCZ in AV trials. Though the CN and SCZ groups have the most data available, the model performs worse when attempting to distinguish between the two in terms of accuracy. This could suggest that during AV trials, CN and SCZ share more similar neural features, or that the neural signatures linked to SCZ are more nuanced and difficult to distinguish from those of neurotypical controls. This decreased classification accuracy may also be attributed to the variability within the SCZ population or possible shared cognitive processing patterns with CN individuals during AV tasks. For the subsequent feature visualization pipeline, the weights of the best performing fold of each models are used to extract significant discriminating features.

Feature Visualization

The GradCAM heatmaps, generated from the first convolutional layer of the model on correctly predicted trials from the test set across all three classifications (ASD vs. CN, ASD vs. SCZ, and CN vs. SCZ), reveal key differences in the model's learning patterns and highlight interpretable ERP-like features. This analysis shows that the model captures meaningful neural dynamics in these correctly classified cases, with distinct spatial and temporal activation patterns that align with known electrophysiological markers. Conversely, in incorrectly predicted trials, GradCAM heatmaps do not exhibit these structured, ERP-like patterns. Instead, the activation is more diffuse and lacks clear spatial or temporal focus, suggesting that the model fails to capture coherent neural features in these cases. An example is shown in the heatmaps below extracted from the ASD vs SCZ model and depict the activation of the outputs of the first convolutional layer in the case of correctly and incorrectly predicted ASD trials.

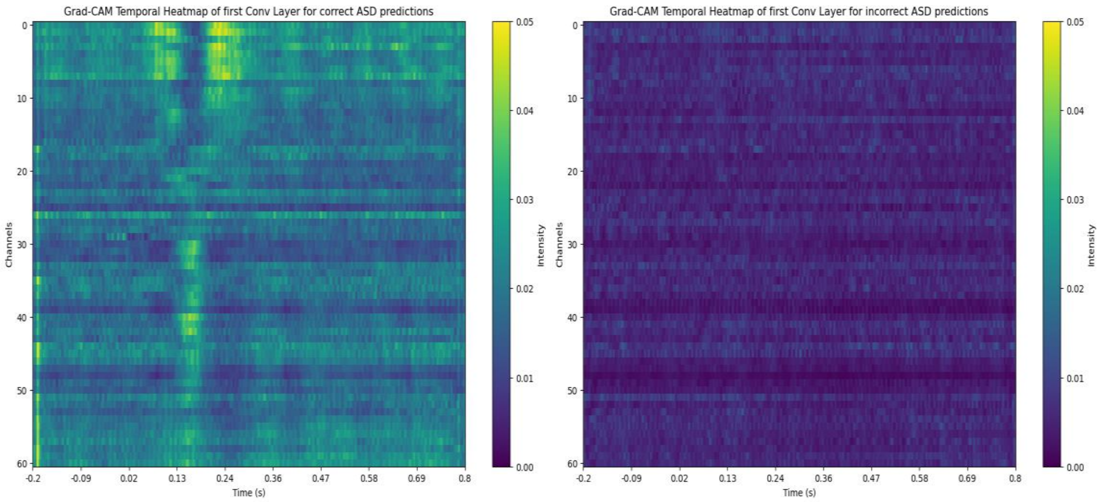


Figure 31: Heatmaps of first convolutional layer in the case of correctly and incorrectly predicted ASD trials

For each classification model (ASD vs. CN, ASD vs. SCZ, and CN vs. SCZ), we present the GradCAM heatmaps, as well as detailed spatial plots from the moment of stimulus occurrence up to 500 ms post-stimulus. This temporal window is chosen because it captures the period with the highest activation intensity in the heatmaps, aligning with the occurrence of key ERP components and encompassing the significant windows identified by the SliceTCA components.

<u>ASD vs CN</u>

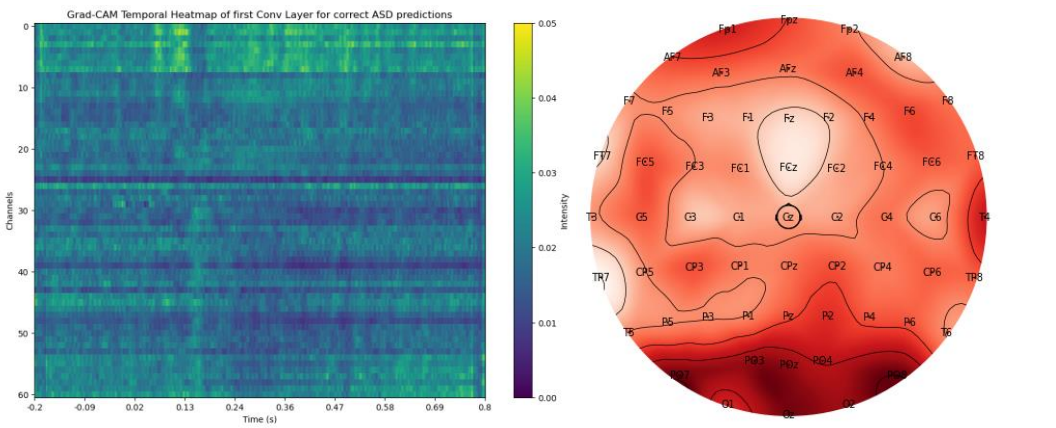


Figure 32: Heatmap and averaged topoplot for 0-0.5s for correctly predicted ASD trials (ASD vs CN)

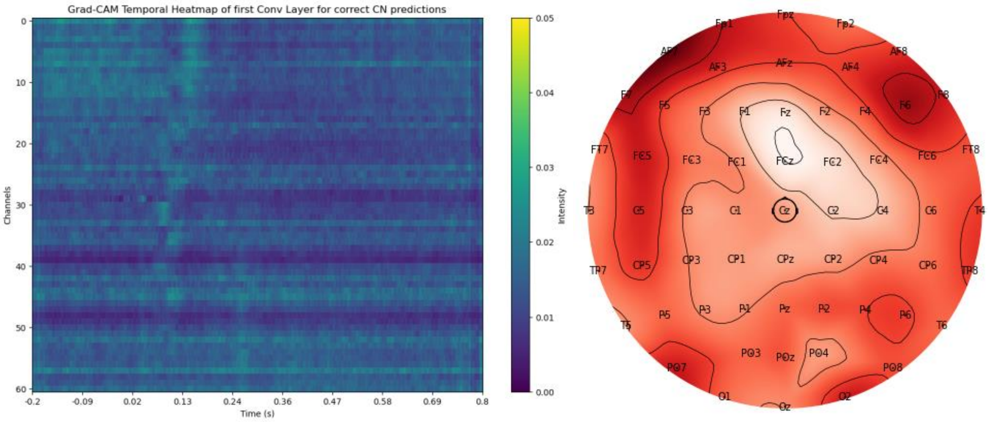


Figure 33: Heatmap and averaged topoplot for 0-0.5s for correctly predicted CN trials (ASD vs CN)

The GradCAM-based heatmaps for correctly predicted ASD and CN trials reveal notable differences in activation intensity and spatial distribution. Higher overall intensity activations are seen in the ASD trials (first figure), with a concentration of activations in the brain's posterior regions. This strong activation may be a sign of the higher cognitive load or increased neural effort that people with ASD need to process multisensory stimuli. This is corroborated by research indicating that people with ASD frequently exhibit abnormal and enhanced neural responses when performing tasks involving sensory integration.

.

On the other hand, the CN trials (second figure) show a more uniform activation throughout the scalp, especially in the frontal areas and the parietal lobes. This spatial distribution is consistent with normal EEG responses, in which processing and responding to AV stimuli is occurring in the frontal cortex of the brain. Comparing the concentrated, high-intensity activations observed in the ASD trials with the wider distribution of activations observed in the CN trials suggests a more distributed and potentially more efficient neural processing mechanism. Additionally, the concentrated posterior activation in ASD suggests reliance on different neural circuits or strategies compared to the CN group, which might relate to differences in sensory processing pathways in ASD.

<u>ASD vs SCZ</u>

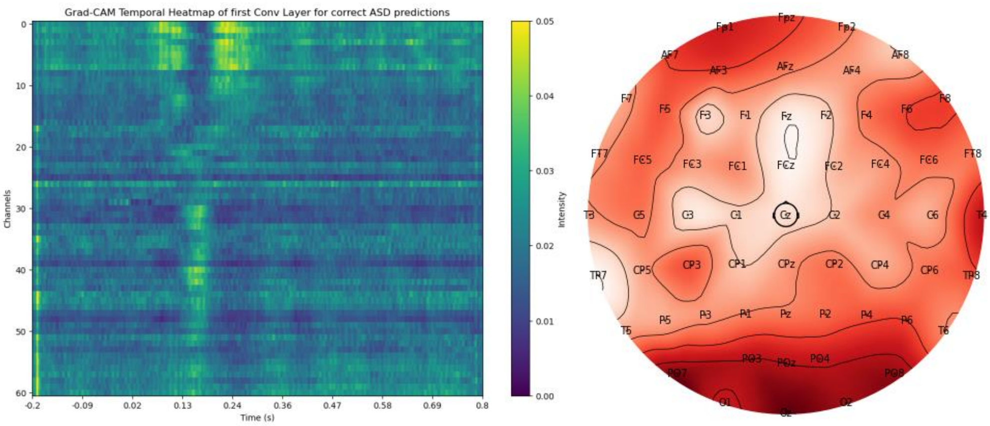


Figure 34: Heatmap and averaged topoplot for 0-0.5s for correctly predicted ASD trials (ASD vs SCZ)

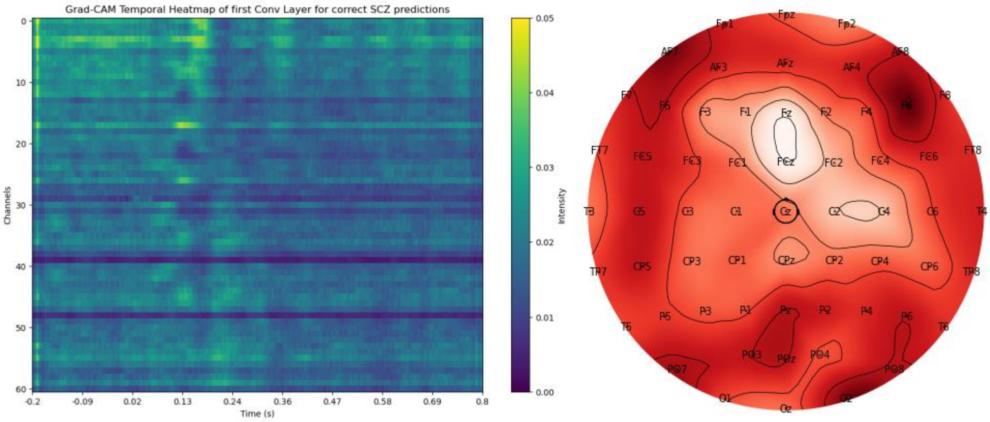


Figure 35: Heatmap and averaged topoplot for 0-0.5s for correctly predicted SCZ trials (ASD vs SCZ)

In this comparison, we observe more distinct ERP-like activations across the groups. The first peak in the ASD heatmaps appears approximately 100 ms after the stimulus, and it most likely corresponds to the P100 component, which is linked to early visual processing and usually originates in the occipital regions. After that, there is a second prominent peak at approximately 300 ms. This corresponds to the P300 component, which is associated with cognitive processing and attentional allocation and is frequently observed in parietal-occipital regions.

The spatial activation pattern in the SCZ trials, on the other hand, is more varied, with a notable peak appearing approximately 200 ms after the stimulus onset. This broader activation pattern could indicate a more distributed and potentially more efficient strategy for multisensory integration, similar to the patterns observed in the CN group. In contrast, the ASD trials exhibit more concentrated activation, which may reflect a less flexible or more narrowly focused neural response strategy. As in the previous comparison, the intensities are higher for the ASD trials.

<u>CN vs SCZ</u>

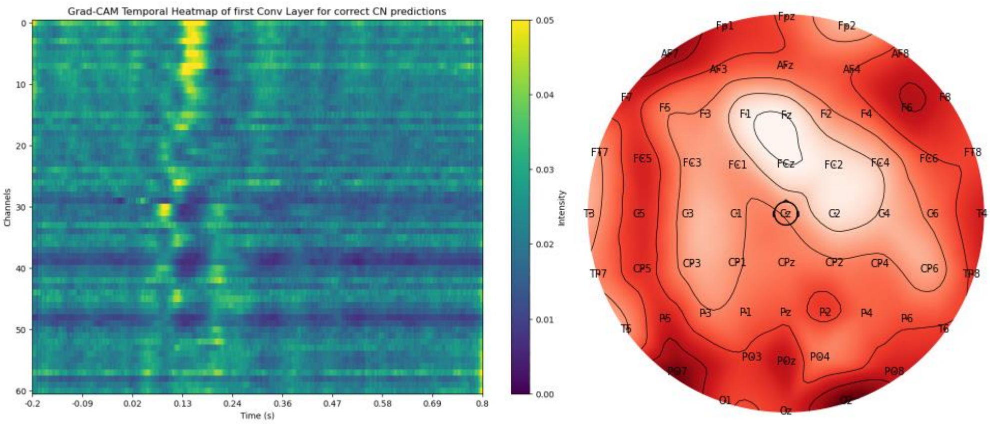


Figure 36: Heatmap and averaged topoplot for 0-0.5s for correctly predicted CN trials (CN vs SCZ)

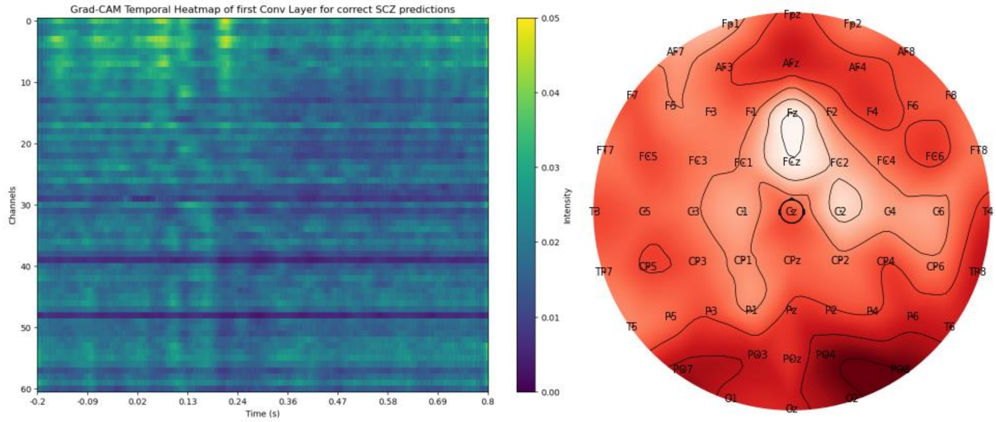


Figure 37: Heatmap and averaged topoplot for 0-0.5s for correctly predicted SCZ trials (CN vs SCZ)

We note that the activation intensity in the CN trials stays relatively high, similar to the SCZ group, in contrast to the ASD vs. CN comparison. Strong early visual processing is indicated by a prominent peak in the occipital region around 200 ms in the CN trials. However, the CN group exhibits a more uniform and evenly distributed activation pattern across all channels in contrast to the more focalized areas of activation shown by the SCZ group.

Probably, the key takeaway from these observations is that the channel activation distribution becomes progressively more uniform across groups, with CN displaying the most even distribution, followed by SCZ, and then ASD. his implies that CN interacts with a wider range of brain areas, potentially indicating a consistent and effective approach to multimodal integration. On the other hand, ASD trials exhibit more intense and localized activations, suggesting a reliance on particular regions, perhaps as a result of higher processing effort in the brain. Interestingly, the SliceTCA analysis also identifies the P300 component as significant, and the ASD trials show a greater prominence for this component. Notably, the P300 component is more prominent in the ASD trials, aligning with findings from the SliceTCA analysis, which also highlights this component as significant. This pronounced P300 in ASD suggests heightened attentional processing and may reflect compensatory mechanisms as the brain allocates additional resources to interpret and integrate sensory information. Distinct neural processing strategies for multisensory integration are highlighted by the variations in both spatial distribution and intensity among the groups.

5. Discussion

This study examined the neural mechanisms underlying multisensory integration by leveraging Tucker decomposition, SliceTCA and EEGNet. This allowed for a thorough investigation of the temporal and structural patterns found in EEG data of neurotypical individuals and individuals diagnosed with ASD and SCZ recorded during an AV synchronous target detection paradigm. This study aimed to provide light on potential underlying neural mechanisms linked to multisensory processing deficits by identifying the temporal and spatial characteristics of neural responses to AV stimuli as well as how these responses may differ or align amongst the three groups.

Tucker decomposition was initially employed for this study but yielded limited interpretability regarding the behaviorally relevant components. By contrast, through SliceTCA's unique channel slicing approach, distinct spatial patterns emerged, capturing localized covariability across time and participant groups, which provided insights into neural correlates of behavioral responses during AV tasks. The decomposition revealed significant ERP components, including early sensory components like N100 and P100, as well as the P300, associated with cognitive processing and attention. Using a bootstrap method, we identified significant time windows that differentiated neural responses between the stronger unisensory and multisensory stimuli, allowing us to isolate intervals where multisensory integration was most prominent. Within these windows, the P300 component was particularly notable for its positive correlation with response gain, especially among the ASD group and secondly in the SCZ group. The ASD group displayed the greatest variability in response gains, indicating that the heightened P300 response may also reflect compensatory mechanisms that contribute to faster response times when both auditory and visual stimuli are presented together.

In comparison, EEGNet provided an alternative yet complementary perspective. As a convolutional neural network specifically designed for EEG analysis, EEGNet identified interpretable spatial-temporal patterns in the data, particularly in response to AV stimuli. The model achieved high classification accuracy in distinguishing between the groups, underscoring its utility for detecting neural patterns associated with multisensory integration in ASD, SCZ, and control groups. Consistent with SliceTCA findings, EEGNet identified heightened activity in parietal and occipital regions, further supporting the role of these areas in multisensory integration. The GradCAM visualizations revealed that both ASD and SCZ groups exhibited altered activation patterns relative to controls, with the ASD group showing intensified activation indicative of heightened cognitive effort. Notably, both SliceTCA and EEGNet captured consistent neural signatures, particularly in the P100-N100 and P300 range, associated with stimulus recognition and enhanced cognitive processing of the sensory inputs, respectively.

Next Steps

The next steps for this research involve improving the analysis and looking into neural and behavioral correlations more thoroughly, building on the current findings. First, within the AV condition, we want to find significant EEG differences in the SliceTCA components that distinguish integrators from non-integrators and perform a similar correlation analysis with behavioral response data by isolating these windows. This will allow us to ascertain whether response gains are correlated with the same ERP components, like the P300, or if other components emerge as significant indicators.

In relation to the EEGNet model, our ongoing work includes testing all participant groups together within a three-class classification framework (ASD, SCZ, and neurotypical). Furthermore, we are investigating stimulus-specific classification by differentiating between stimuli that are AV, A, and V within each group as well as between groups. This method will assist in identifying distinct neural responses to various sensory situations. Both the SliceTCA and EEGNet methodology pipeline could also be applied to response-locked EEG data instead of the stimulus-locked in this study. This analysis may shed light to different structure, as response-locked analysis accounts for variability in individual reaction times, allowing for a more precise examination of neural activity related to response generation.

Lastly, to enhance the interpretability of our results, we intend to utilize GradCAM visualizations on an individual level Our goal is to reduce the influence of inter-subject variability and provide a more individualised perspective on multisensory performance by visualising temporal and spatial features for every participant. Through repeated trials, adaptive neural changes may be revealed by this individual-level analysis, which could provide insight into how each participant's multisensory processing changes over the course of the experiment.

Bibliography

Beck, A., & Teboulle, M. (2009). A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM journal on imaging sciences*, *2*(1), 183-202.

Boutros, N. N., Arfken, C., Galderisi, S., Warrick, J., Pratt, G., & Iacono, W. (2008). The status of spectral EEG abnormality as a diagnostic test for schizophrenia. *Schizophrenia research*, *99*(1-3), 225-237.

Brandwein, A. B., Foxe, J. J., Butler, J. S., Frey, H. P., Bates, J. C., Shulman, L. H., & Molholm, S. (2015). Neurophysiological indices of atypical auditory processing and multisensory integration are associated with symptom severity in autism. *Journal of autism and developmental disorders*, *45*, 230-244.

Brandwein, A. B., Foxe, J. J., Butler, J. S., Russo, N. N., Altschuler, T. S., Gomes, H., & Molholm, S. (2013). The development of multisensory integration in high-functioning autism: high-density electrical mapping and psychophysical measures reveal impairments in the processing of audiovisual inputs. *Cerebral cortex*, *23*(6), 1329-1341.

Brandwein, A. B., Foxe, J. J., Russo, N. N., Altschuler, T. S., Gomes, H., & Molholm, S. (2011). The development of audiovisual multisensory integration across childhood and early adolescence: a high-density electrical mapping study. *Cerebral cortex*, *21*(5), 1042-1055.

Chandrasekhar, T., Copeland, J. N., Spanos, M., & Sikich, L. (2020). Autism, psychosis, or both? Unraveling complex patient presentations. *Child and Adolescent Psychiatric Clinics*, *29*(1), 103-113.

Collignon, O., Charbonneau, G., Peters, F., Nassim, M., Lassonde, M., Lepore, F., ... & Bertone, A. (2013). Reduced multisensory facilitation in persons with autism. *cortex*, *49*(6), 1704-1710.

Diederich, A., & Colonius, H. (1987). Intersensory facilitation in the motor component? A reaction time analysis. *Psychological Research*, *49*(1), 23-29.

Freund, Y., & Schapire, R. E. (1997). A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of computer and system sciences*, *55*(1), 119-139.

Cover, T., & Hart, P. (1967). Nearest neighbor pattern classification. *IEEE transactions on information theory*, *13*(1), 21-27.

Giard, M. H., & Peronnet, F. (1999). Auditory-visual integration during multimodal object recognition in humans: a behavioral and electrophysiological study. *Journal of cognitive neuroscience*, *11*(5), 473-490.

Gröhn, C., Norgren, E., & Eriksson, L. (2022). A systematic review of the neural correlates of multisensory integration in schizophrenia. *Schizophrenia Research: Cognition*, *27*, 100219.

Jiang, T. X., Huang, T. Z., Zhao, X. L., & Deng, L. J. (2017). A novel nonconvex approach to recover the low-tubal-rank tensor data: When t-SVD meets PSSV. *arXiv preprint arXiv:1712.05870*.

Kolda, T. G., & Bader, B. W. (2009). Tensor decompositions and applications. *SIAM review*, *51*(3), 455-500.

Krishnan, G. P., Hetrick, W. P., Brenner, C. A., Shekhar, A., Steffen, A. N., & O'Donnell, B. F. (2009). Steady state and induced auditory gamma deficits in schizophrenia. *Neuroimage*, *47*(4), 1711-1719.

Kwon, J. S., O'Donnell, B. F., Wallenstein, G. V., Greene, R. W., Hirayasu, Y., Nestor, P. G., ... & McCarley, R. W. (1999). Gamma frequency–range abnormalities to auditory stimulation in schizophrenia. *Archives of general psychiatry*, *56*(11), 1001-1005.

Lawhern, V. J., Solon, A. J., Waytowich, N. R., Gordon, S. M., Hung, C. P., & Lance, B. J. (2018). EEGNet: a compact convolutional neural network for EEG-based brain–computer interfaces. *Journal of neural engineering*, *15*(5), 056013.

Marucci, M., Di Flumeri, G., Borghini, G., Sciaraffa, N., Scandola, M., Pavone, E. F., ... & Aricò, P. (2021). The impact of multisensory integration and perceptual load in virtual reality settings on performance, workload and presence. *Scientific Reports*, *11*(1), 4831.

Miller, J. (1982). Divided attention: Evidence for coactivation with redundant signals. *Cognitive psychology*, *14*(2), 247-279.

Molholm, S., Ritter, W., Murray, M. M., Javitt, D. C., Schroeder, C. E., & Foxe, J. J. (2002). Multisensory auditory–visual interactions during early sensory processing in humans: a high-density electrical mapping study. *Cognitive brain research*, *14*(1), 115-128.

Osaka, M., Fukami, T., Shimada, T., Ishikawa, F., Arikura, S., Kurihara, M., & Suzuki, T. (2012, July). Analysis of P300 evoked by multisensory stimulus consisting of visual and auditory stimuli. In *2012 ICME International Conference on Complex Medical Engineering (CME)* (pp. 717-721). IEEE.

Ostrolenk, A., Bao, V. A., Mottron, L., Collignon, O., & Bertone, A. (2019). Reduced multisensory facilitation in adolescents and adults on the autism spectrum. *Scientific Reports*, *9*(1), 11965.

Owen, M. J., & O'Donovan, M. C. (2017). Schizophrenia and the neurodevelopmental continuum: evidence from genomics. *World Psychiatry*, *16*(3), 227-235.

Pellegrino, A., Stein, H., & Cayco-Gajic, N. A. (2024). Dimensionality reduction beyond neural subspaces with slice tensor component analysis. *Nature Neuroscience*, 1-12.

Plass, J., & Brang, D. (2021). Multisensory stimuli shift perceptual priors to facilitate rapid behavior. *Scientific Reports*, *11*(1), 23052.

Rapoport, J., Chavez, A., Greenstein, D., Addington, A., & Gogtay, N. (2009). Autism spectrum disorders and childhood-onset schizophrenia: clinical and biological contributions to a relation revisited. *Journal of the American Academy of Child & Adolescent Psychiatry*, *48*(1), 10-18.

Russo, N., Foxe, J. J., Brandwein, A. B., Altschuler, T., Gomes, H., & Molholm, S. (2010). Multisensory processing in children with autism: high-density electrical mapping of auditory–somatosensory integration. *Autism Research*, *3*(5), 253-267.

Saville, C. W., Shikhare, S., Iyengar, S., Daley, D., Intriligator, J., Boehm, S. G., ... & Klein, C. (2012). Is reaction time variability consistent across sensory modalities? Insights from latent variable analysis of single-trial P3b latencies. *Biological psychology*, *91*(2), 275-282.

Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., & Batra, D. (2017). Grad-cam: Visual explanations from deep networks via gradient-based localization. In *Proceedings of the IEEE international conference on computer vision* (pp. 618-626).

Stefanou, Maria Elena, et al. "Late attentional processes potentially compensate for early perceptual multisensory integration deficits in children with autism: evidence from evoked potentials." *Scientific Reports* 10.1 (2020): 16157.

Sultonov, F., Park, J. H., Yun, S., Lim, D. W., & Kang, J. M. (2022). Mixer U-Net: An improved automatic road extraction from UAV imagery. *Applied Sciences*, *12*(4), 1953.

Talsma, D., Senkowski, D., Soto-Faraco, S., & Woldorff, M. G. (2010). The multifaceted interplay between attention and multisensory integration. *Trends in cognitive sciences*, *14*(9), 400-410.

Toumaian, M., Covanis, P., Mantas, A., Karantinos, T., Kayas, S., Kentikeleni, A., ... & Smyrnis, N. (2024). Multisensory integration deficits in Schizophrenia and Autism evidenced in behaviour but not event related potentials. *Psychiatry Research*, *332*, 115727.

Tucker, L. R. (1966). Some mathematical notes on three-mode factor analysis. *Psychometrika*, *31*(3), 279-311.

Uhlhaas, P. J., & Singer, W. (2013). High-frequency oscillations and the neurobiology of schizophrenia. *Dialogues in clinical neuroscience*, *15*(3), 301-313.

Ulrich, R., Miller, J., & Schröter, H. (2007). Testing the race model inequality: An algorithm and computer programs. *Behavior Research Methods*, *39*(2), 291-302.

Valakos, D., d'Avossa, G., Mylonas, D., Butler, J., Klein, C., & Smyrnis, N. (2020). P300 response modulation reflects breaches of non-probabilistic expectations. *Scientific reports*, *10*(1), 10254.

White, T. P., Joseph, V., O'Regan, E., Head, K. E., Francis, S. T., & Liddle, P. F. (2010). Alpha–gamma interactions are disturbed in schizophrenia: a fusion of electroencephalography and functional magnetic resonance imaging. *Clinical Neurophysiology*, *121*(9), 1427-1437.

Williams, L. E., Light, G. A., Braff, D. L., & Ramachandran, V. S. (2010). Reduced multisensory integration in patients with schizophrenia on a target detection task. *Neuropsychologia*, *48*(10), 3128-3136.

Wynn, J. K., Jahshan, C., & Green, M. F. (2014). Multisensory integration in schizophrenia: a behavioural and event-related potential study. *Cognitive neuropsychiatry*, *19*(4), 319-336.