Personalized Educational Framework for Autism Using EEG and ERP Biomarkers with Deep Learning Tools

Master's thesis in Biomedical Engineering

SIYI CHENG and YIZHEN WEN

DEPARTMENT OF ELECTRICAL ENGINEERING

CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden 2026
www.chalmers.se

Master's thesis 2026

Personalized Educational Framework for Autism Using EEG and ERP Biomarkers with Deep Learning Tools

SIYI CHENG and YIZHEN WEN

CHALMERS
UNIVERSITY OF TECHNOLOGY

Department of Electrical Engineering
Division of Signal Processing and Biomedical Engineering
Chalmers University of Technology
Gothenburg, Sweden 2026

Personalized Educational Framework for Autism
Using EEG and ERP Biomarkers with Deep Learning Tools
SIYI CHENG and YIZHEN WEN

Supervisor: Oana Geman, Computer Science and Engineering
Examiner: XueZhi Zeng, Electrical Engineering

Master's Thesis 2026
Department of Electrical Engineering
Division of Signal Processing and Biomedical Engineering
Chalmers University of Technology
SE-412 96 Gothenburg
Telephone +46 31 772 1000

Typeset in LaTeX
Printed by Chalmers Reproservice
Gothenburg, Sweden 2026

Personalized Educational Framework for Autism
Using EEG and ERP Biomarkers with Deep Learning Tools
SIYI CHENG and YIZHEN WEN
Department of Electrical Engineering
Chalmers University of Technology

Abstract

Children with Autism Spectrum Disorder (ASD) exhibit heterogeneous auditory processing profiles that are not accommodated by standardised acoustic environments. This thesis presents a two-stage framework for objective and personalized acoustic assessment using electroencephalography (EEG). The first stage employs a Convolutional Neural Network–Bidirectional Long Short-Term Memory (CNN-BiLSTM) classifier with trial-level voting to identify children with ASD from single-trial EEG. The second stage introduces the Neural Processing Efficiency Score (NPES), a composite index derived from four Event-Related Potential (ERP) biomarkers using within-subject normalisation, to recommend the optimal acoustic configuration for each child across five stimulus dimensions. The within-subject approach was adopted after group-level ERP comparisons yielded no significant differences between ASD and typically developing children. Evaluation on 22 ASD participants confirmed substantial heterogeneity in the recommended configurations, supporting the need for individualised rather than group-level acoustic assessment. The framework provides a methodological basis for neurophysiologically grounded acoustic personalisation in educational and therapeutic contexts.

Keywords: Autism Spectrum Disorder, EEG, Event-Related Potentials, Deep Learning, Personalized Assessment, Neural Processing Efficiency, Auditory Processing

Acknowledgements

We would like to express our sincere gratitude to our supervisor, Oana Geman, for her guidance and support throughout this project. Her feedback and encouragement kept us on track during the more challenging stages of the work. We also thank our examiner, Xuezhi Zeng, for her valuable comments and careful review of this thesis. We are also grateful to the research team at Ştefan cel Mare University in Suceava, Romania, for providing the EEG dataset that made this study possible.
Finally, we would like to thank each other for the collaboration and mutual support over the course of this thesis. Writing a master's thesis is not easy, and having a reliable partner made the process much more manageable.

Yizhen Wen and Siyi Cheng, Gothenburg, May 2026

List of Acronyms

Below is the list of acronyms that have been used throughout this thesis listed in alphabetical order:

ADI-R	Autism Diagnostic Interview–Revised
ADOS	Autism Diagnostic Observation Schedule
ASD	Autism Spectrum Disorder
BCI	Brain-Computer Interface
BiLSTM	Bidirectional Long Short-Term Memory
CAR	Common Average Reference
CNN	Convolutional Neural Network
DL	Deep Learning
DSM-5	Diagnostic and Statistical Manual of Mental Disorders, 5th Edition
EEG	Electroencephalography
ERP	Event-Related Potential
GFP	Global Field Power
ICA	Independent Component Analysis
LSTM	Long Short-Term Memory
NPES	Neural Processing Efficiency Score
ROI	Region of Interest
SHAP	SHapley Additive exPlanations
TD	Typically Developing
TEACCH	Treatment and Education of Autistic and related Communication-handicapped Children

Contents

List of Figures

List of Tables

1

Introduction

1.1 Motivation and Problem Statement

Autism Spectrum Disorder (ASD) is a heterogeneous neurodevelopmental condition characterised by persistent differences in social communication and restricted, repetitive patterns of behaviour [1]. Beyond the core diagnostic criteria, atypical sensory processing, particularly in the auditory domain, is reported in a majority of individuals with ASD and has direct consequences for daily functioning, including communication, learning, and social participation [2]. In educational and therapeutic settings, auditory stimuli are typically delivered at a standardised pace and intensity that does not account for the altered temporal integration, frequency sensitivity, or noise tolerance observed in many children with ASD. As a result, these children may experience reduced comprehension, heightened distress, or attentional disengagement under acoustic conditions that are adequate for their typically developing (TD) peers.

A critical yet frequently overlooked aspect of this challenge is the substantial inter-individual heterogeneity within the ASD population. One child may struggle to process rapid speech, while another may show heightened sensitivity to background noise or respond differently to variations in speaker pitch. Such heterogeneity implies that a single, group-averaged intervention strategy is unlikely to be effective across the spectrum. Despite this recognition, current clinical assessment relies predominantly on behavioural observation instruments such as the Autism Diagnostic Observation Schedule (ADOS) and the Autism Diagnostic Interview–Revised (ADI-R) [3]. These tools are inherently subjective and time-consuming, and they capture behavioural phenotypes rather than the underlying neural mechanisms that would be required to guide individualised accommodations.

Objective neurophysiological measurement offers a route toward addressing this gap. Electroencephalography (EEG) and the derived Event-Related Potentials (ERPs) provide the temporal resolution necessary to capture rapid neural dynamics underlying auditory processing. Specific ERP components, such as the N100 reflecting early auditory attention and the P300 reflecting cognitive resource allocation, have demonstrated sensitivity to ASD-related processing differences [4]. Large-scale multi-site validation efforts, including the Autism Biomarkers Consortium for Clinical Trials, have further confirmed that EEG-derived markers possess moderate-to-high psychometric stability [5], supporting their potential use as objective indices of auditory processing efficiency.

However, the majority of existing Deep Learning (DL) applications to ASD EEG

have focused on binary classification, distinguishing ASD from TD at the group level [6]. While such classifiers can contribute to screening, a binary diagnostic label provides no information about the specific acoustic conditions under which a given child processes auditory input most efficiently. This thesis is motivated by the need to move beyond categorical classification toward individualised functional profiling: not only identifying children who may benefit from adapted acoustic environments, but also determining what adaptation each child requires.

1.2 Aims and Objectives

The primary aim of this thesis is to develop a framework that translates EEG and ERP data into personalized acoustic recommendations for children with ASD. The framework integrates automated screening with individualised neural profiling, enabling a two-stage assessment process. First, the system determines which children may require personalized intervention. Second, it identifies the acoustic configuration that best supports each child's auditory processing.
To achieve this aim, the thesis is organised around three objectives:

1. **Develop an EEG-based deep learning model for ASD screening.**
   A hybrid architecture combining a Convolutional Neural Network with a Bidirectional Long Short-Term Memory network (CNN-BiLSTM) is trained on single-trial EEG data. Trial-level predictions are then aggregated to the subject level through trial voting, allowing the model to serve as the screening front-end of the proposed system.
2. **Characterize individual neural processing profiles using ERP biomarkers.**
   Conventional group-level ERP comparisons between ASD and TD children are first examined to assess whether statistically stable biomarkers exist at the population level. The findings of this analysis motivate a shift toward a within-subject approach, in which a composite Neural Processing Efficiency Score (NPES) is constructed from four established biomarkers: N100 latency, P200 latency, P300 latency, and Global Field Power (GFP).
3. **Build and evaluate a personalized acoustic recommendation system.**
   The NPES is used to identify the optimal acoustic configuration for each ASD child across five stimulus dimensions: speech rate, background noise, speaker distance, ear of presentation, and pitch. The system is evaluated through heterogeneity analysis, biomarker contribution analysis, and split-half stability analysis. The intended application is to provide educators, clinicians, and acoustic environment designers with an objective, data-driven tool that can inform the configuration of learning environments, assistive listening devices, and individualised pedagogical practice for children with ASD.

1.3 Thesis Structure

The remainder of this thesis is organised as follows. Chapter 2 reviews the theoretical background and related work, covering ASD auditory processing, the ERP

biomarkers employed in this study, deep learning approaches to EEG-based classification, and existing personalized assessment frameworks. Chapter 3 details the methodology, including the data acquisition protocol, signal preprocessing and ERP extraction pipeline, the deep learning classification framework, and the design of the NPES-based recommendation system. Chapter 4 presents the experimental results, organised into data quality and cohort characterisation, group-level ERP analysis, trial-level and subject-level classification, model interpretation, and personalized recommendation evaluation. Chapter 5 synthesises the findings, discusses their implications and limitations, and outlines directions for future work. Chapter 6 concludes the thesis.

2
Background

This chapter reviews the theoretical foundations and prior work relevant to the proposed framework. Section 2.1 discusses ASD and the auditory processing differences that motivate the present study. Section 2.2 introduces the ERP biomarkers used for neural profiling. Section 2.3 surveys deep learning approaches to EEG-based ASD classification. Section 2.4 reviews personalized approaches for ASD in educational contexts. Section 2.5 identifies the research gap addressed by this thesis.

2.1 ASD and Auditory Processing

2.1.1 Clinical Profile and Heterogeneity

ASD is defined by the Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition (DSM-5) as involving persistent deficits in social communication and interaction alongside restricted, repetitive patterns of behaviour [1]. The spectrum encompasses a wide range of functional profiles, from individuals with intellectual disability and minimal verbal language to those with average or above-average cognitive ability. This heterogeneity extends to sensory processing: while atypical sensory responsivity is recognised as a diagnostic feature under the DSM-5, the specific modalities affected and the direction of atypicality (hyper- or hypo-reactivity) vary substantially across individuals [3].

Clinical diagnosis currently relies on standardised behavioural observation instruments, primarily ADOS and ADI-R. Although these tools have well-established reliability, they assess behavioural phenotypes rather than the underlying neural mechanisms. Diagnosis typically occurs between 18 and 24 months of age when behavioural symptoms become distinguishable, yet neurophysiological deviations have been detected considerably earlier using EEG-based methods [4]. This temporal discrepancy highlights the potential of objective neuroimaging to complement behavioural assessment.

2.1.2 Auditory Processing Differences in ASD

Auditory perception is foundational to language acquisition, social interaction, and classroom learning. ERP research has consistently demonstrated that individuals with ASD exhibit atypical neural responses during both early sensory encoding and later cognitive processing stages [7].

At the early sensory encoding stage, children with ASD frequently show reduced N100 amplitudes in response to speech stimuli, suggesting a deficit in initial cortical

registration of auditory input [8]. This attenuation has been observed across variations in pitch, intensity, and spatial location, indicating a pervasive rather than condition-specific impairment. Additionally, elevated inter-trial variability in the EEG response to identical stimuli has been reported in ASD, suggesting that neural instability compounds the encoding deficit [9].
At later cognitive stages involving attention allocation, the deficits become more context-dependent. Reduced P300 amplitudes have been observed specifically during monaural left-ear presentation, consistent with a spectro-temporal processing imbalance between hemispheres [10]. Atypical P300 latency patterns have also been reported as a function of speaker distance, with ASD children showing shortened latencies for proximal stimuli that may reflect a hyper-reactive cognitive response to high-intensity social input [11].

2.1.3 Sensitivity to Acoustic Parameters

The auditory processing differences summarised above are not uniform across acoustic conditions, making it important to consider how specific stimulus parameters modulate neural responses in ASD. Five acoustic dimensions are particularly relevant to the present work:

1. **Speech rate.** Speech rate determines the temporal density of linguistic information. Faster speech compresses phonemic transitions, placing greater demands on temporal integration. Children with ASD, who often exhibit widened temporal integration windows, may be disproportionately affected by rapid speech [2].
2. **Background noise.** Background noise degrades the signal-to-noise ratio of the speech signal and requires the listener to segregate the target voice from competing sounds. Neurophysiological evidence indicates that noise attenuates early cortical components and impairs neural tracking of the speech envelope in ASD [12]. Behavioural and EEG findings further confirm that ASD listeners show delayed semantic processing under noisy conditions [13].
3. **Speaker distance.** Speaker distance affects the perceived intensity and spatial characteristics of the auditory signal. Proximal stimuli elicit stronger neural responses and may trigger hyper-reactive attentional orienting in ASD [11].
4. **Ear of presentation.** Ear of presentation probes hemispheric lateralisation of auditory processing. The right-ear advantage for speech, which reflects left-hemisphere dominance for language, may be reduced or absent in ASD. Monaural presentation to the left ear has been associated with diminished P300 amplitudes in ASD relative to TD children [10].
5. **Pitch variation.** Pitch variation alters the fundamental frequency of the speech signal and is closely linked to prosodic processing. Children with ASD have been reported to show both enhanced and diminished sensitivity to pitch changes depending on stimulus complexity, suggesting that pitch processing in ASD is atypical rather than uniformly impaired [7].

The diversity of these findings underscores two points relevant to this thesis. First, no single acoustic parameter captures the full extent of auditory processing differ-

ences in ASD. Second, the direction and magnitude of these differences vary across individuals, motivating an individualised rather than group-level approach to acoustic assessment.

2.2 EEG Biomarkers for Auditory Processing

This section introduces the six ERP components and one global metric extracted in the present study. The components are divided into onset components, time-locked to stimulus onset and reflecting early sensory and attentional processing, and offset components, time-locked to stimulus offset and reflecting higher-order language processing. Figure 2.1 provides a simulated schematic overview of these components and their approximate temporal positions within a typical auditory ERP waveform.

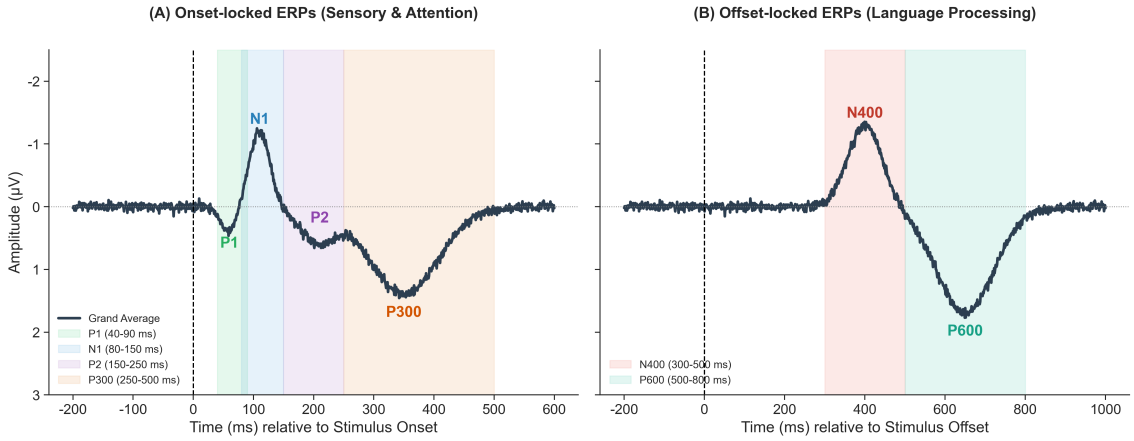

Figure 2.1: Schematic diagram of the extracted Auditory ERP components, showing (A) onset-locked components reflecting early sensory gating and attentional orienting, and (B) offset-locked linguistic components capturing semantic integration and syntactic repair.

2.2.1 Early Sensory Components: N100 and P200

The N100 is a negative-going component peaking approximately between 80 and 150 ms after stimulus onset, maximal over fronto-central scalp regions, and is commonly interpreted as reflecting auditory attention orienting and early cortical sound registration [14]. In ASD, atypical N100 responses, including reduced amplitudes and delayed latencies, have been reported, although findings vary across paradigms and participant groups [8, 15].

The P200 is a positive-going component peaking approximately between 150 and 250 ms after stimulus onset, also maximal over fronto-central sites. It is associated with early stimulus evaluation and is sensitive to stimulus salience, intensity, and spectral content [14].

A still earlier component, the P100, or P1, peaks approximately between 40 and 90 ms after stimulus onset and reflects the initial cortical registration of auditory

input [14]. Although it is not a primary biomarker in the present study, it is included in the extraction pipeline for completeness.

2.2.2 Attention and Cognitive Component: P300

The P300 is a positive-going component typically observed between 250 and 500 ms after stimulus onset, with a maximal centro-parietal scalp distribution [14, 16]. It is commonly associated with the allocation of attentional resources to salient or task-relevant stimuli, with amplitude reflecting attentional engagement and latency reflecting the time required for stimulus evaluation and categorisation [16]. In ASD research, atypical P300 responses have been linked to differences in attentional engagement and cognitive evaluation, although findings vary across paradigms, age groups, and stimulus types [15, 17]. In the present study, P300 latency is included in the NPES as an indicator of context-dependent auditory processing efficiency.

2.2.3 Language-Related Components: N400 and P600

Two additional components are relevant to the sentence-level stimuli used in the present study. The N400 is a negative-going component typically observed approximately between 300 and 500 ms and is commonly associated with semantic processing and meaning integration [18]. Larger N400 amplitudes are generally interpreted as reflecting greater difficulty or processing effort in integrating a word or meaningful stimulus into its preceding context [18].
The P600 is a positive-going component typically observed approximately between 500 and 800 ms and is commonly associated with syntactic anomaly detection, repair, and reanalysis processes [19, 20]. Together, the N400 and P600 provide complementary indices of semantic and syntactic aspects of sentence-level language processing.
Although the N400 and P600 are included in the ERP extraction pipeline and contribute to the group-level waveform analysis, they are not incorporated into the NPES due to the additional variability introduced by offset-locked extraction across speed conditions. This decision is discussed further in Section 3.3.1.

2.2.4 Global Field Power

GFP is computed as the spatial standard deviation, or root mean square, of the electrical potential across recording electrodes at each time point [21, 22]. It provides a reference-independent measure of the overall strength of the scalp potential field and is commonly used in multichannel ERP analysis to summarise global response strength across the electrode montage [21, 23].
Unlike single-channel ERP amplitudes, GFP summarises the spatial dispersion of the scalp potential field rather than the voltage at a specific electrode. Higher GFP values therefore indicate stronger global field strength, but they do not by themselves specify whether the underlying processing is more efficient or less efficient. In the present study, lower GFP under a given acoustic condition is operationally interpreted as reflecting more spatially focused and potentially more efficient processing only when considered together with latency-based ERP measures.

This property is particularly relevant for low-density EEG recordings, where single-channel amplitudes can be more sensitive to electrode selection and reference-related polarity ambiguities. For this reason, GFP is used in the present study as a complementary biomarker to latency-based ERP measures, rather than as a direct substitute for component-specific amplitude analysis.

2.2.5 Within-Subject Normalisation

When using ERP biomarkers for individualised assessment rather than group-level comparison, a methodological challenge arises: absolute biomarker values differ across individuals due to factors unrelated to the experimental manipulation, including skull thickness, electrode impedance, and baseline neural excitability. Within-subject z-scoring addresses this by expressing each trial-level biomarker value relative to the individual's own distribution across all conditions [14]. The resulting standardised values can be combined across biomarkers without one component dominating due to its absolute scale. This normalisation strategy forms the basis of the composite NPES developed in Chapter 3.

2.3 Deep Learning for EEG-Based ASD Classification

2.3.1 Architectures

The high dimensionality and non-stationary nature of EEG data have motivated the application of DL architectures that learn hierarchical representations directly from raw or minimally preprocessed signals. Three architectural families are relevant to ASD EEG classification.

CNNs apply learnable filters along the temporal dimension of the EEG signal, extracting local features such as waveform shapes and frequency-band power. One-dimensional CNNs operating on multi-channel time series have been applied to ERP-based classification [24].

Recurrent architectures, particularly Long Short-Term Memory (LSTM) networks and their bidirectional variants (BiLSTM), model temporal dependencies across the full epoch. By maintaining a hidden state that integrates information from earlier and later time points, BiLSTMs can capture relationships between ERP components distributed across the epoch.

Hybrid CNN-BiLSTM architectures combine the strengths of both families: the CNN front-end extracts local spatiotemporal features, and the BiLSTM back-end models their sequential dependencies. Sharghilavan et al. [25] demonstrated that a CNN-BiLSTM architecture with explainability analysis achieved high classification accuracy on a small ASD cohort ($N = 26$), suggesting that the hybrid design captures richer spatiotemporal signatures than either component alone.

2.3.2 Subject-Independent Evaluation

A critical methodological consideration in EEG classification is the evaluation protocol. Because EEG signals exhibit strong inter-individual variability, a model trained and tested on trials from the same participant can achieve high accuracy by memorising individual neural signatures rather than learning generalisable features. Subject-independent evaluation, in which no participant appears in both training and test sets, provides a more realistic estimate of clinical utility but typically yields lower accuracy, particularly with small sample sizes [26].
Studies on smaller, focused cohorts have reported exceptionally high accuracy: Peck et al. [27] achieved 100% using support vector machines on 54 participants, while Tawhid et al. [28] reported accuracies up to 99% on 16 subjects. However, scaling to larger and more heterogeneous datasets has proven challenging. Xu et al. [29] reported that accuracies dropped to approximately 81% on 189 participants, highlighting the difficulty of generalisation and the importance of rigorous subject-independent validation.

2.3.3 Explainability

The clinical application of DL models requires that their decisions be interpretable. SHapley Additive exPlanations (SHAP) is a game-theoretic framework that assigns each input feature an importance value representing its contribution to the model output [30]. Applied to EEG classification, SHAP values can reveal which channels and time points drive the prediction, enabling comparison with known neurophysiological patterns. This transparency is essential for building trust in automated tools intended for educational or clinical use.

2.3.4 Limitations of Existing Approaches

Despite promising results, several limitations characterise existing DL approaches to ASD EEG classification. First, many studies report results on small, single-site datasets with fewer than 50 participants [26]. Second, the majority of studies employ within-subject or trial-level evaluation protocols that inflate accuracy estimates. Third, the focus on binary classification provides no information about the functional profile of the individual, limiting the translational value for intervention planning. This thesis addresses the third limitation directly by extending the classification framework with a personalized recommendation component.

2.4 Personalized Approaches for ASD

2.4.1 Adaptive Learning for ASD

Adaptive educational technologies modify instructional content or delivery in response to the learner's performance or state. For learners with ASD, the potential benefits are amplified by the heterogeneity of the population: a system that adapts

to individual needs can accommodate the diverse sensory and cognitive profiles that characterise the spectrum.

2.4.2 The TEACCH Framework

The Treatment and Education of Autistic and related Communication-handicapped Children (TEACCH) programme is one of the most widely implemented structured teaching approaches for ASD [31, 32, 33]. TEACCH emphasises environmental predictability, visual structure, and individualised task organisation to reduce sensory overload and support independent functioning. While TEACCH structures the physical and visual environment, it does not explicitly address the acoustic environment. The present thesis extends this principle of environmental structuring to the auditory domain, using neurophysiological data to identify the acoustic parameters that best support each child's processing efficiency.

2.4.3 The Diagnosis-to-Intervention Gap

A persistent challenge in ASD research and practice is the gap between diagnosis and individualised intervention. Current DL models applied to EEG predominantly function as binary classifiers, providing a categorical label but no guidance on how to adapt the environment for the identified individual [6]. A diagnostic output cannot indicate whether a child processes speech more efficiently in quiet or in noise, at close or far distance, or with slow or fast delivery. Bridging this gap requires a framework that not only classifies the individual but also profiles their functional characteristics across relevant stimulus dimensions.

2.5 Research Gap and Contributions

The literature reviewed above identifies three converging gaps. First, while ERP biomarkers have been shown to differ between ASD and TD groups, the high inter-individual heterogeneity within ASD limits the practical utility of group-level statistical contrasts for guiding individualised intervention. Second, existing DL models for ASD EEG classification do not extend beyond binary labelling to functional profiling. Third, current educational frameworks for ASD do not incorporate objective neurophysiological measurement to inform acoustic environment design.
This thesis contributes to addressing these gaps by developing a two-stage pipeline that combines DL classification with biomarker-based personalized recommendation, grounded in a composite neural efficiency index that captures individual auditory processing profiles across multiple acoustic dimensions.

3
Methods

This chapter describes the complete methodological pipeline, from raw EEG acquisition through preprocessing, feature extraction, and statistical analysis, to the deep learning classification framework and the design of the personalized acoustic recommendation system.

3.1 Data Acquisition and Preprocessing

3.1.1 Dataset and Recording Setup

This study utilised a secondary EEG dataset acquired by collaborating researchers at Ştefan cel Mare University in Suceava, Romania. The dataset comprised recordings from 89 Romanian children: 43 with a clinical diagnosis of ASD (27 male, 16 female; mean age = 9.88 years, SD = 1.37; age range: 7–12 years) and 46 TD peers (23 male, 23 female; mean age = 11.15 years, SD = 1.09; age range: 8–12 years). ASD diagnoses were confirmed through comprehensive clinical evaluation, including the ADI-R, ADOS, the Childhood Autism Rating Scale, and the Vineland Adaptive Behaviour Scales, and verified via psychiatric records. All participants had normal hearing and no history of neurological, psychiatric, or other medical conditions beyond ASD where applicable. None were taking medication at the time of testing. All participant data were anonymised prior to analysis; no personally identifiable information was used in any processing step or reported in this thesis. Details of the ethical approvals and data governance procedures are provided in Section 5.4.

EEG data were recorded using the Ultracortex Mark IV headset equipped with 16 dry electrodes, positioned according to the international 10-20 system, as illustrated in Figure 3.1. Data were acquired using the OpenBCI CytonDaisy biosensing system, which combines a Cyton board governed by a 32-bit PIC32 microcontroller with a Daisy module to facilitate 16-channel recording. The analogue EEG signals were digitised using 24-bit analogue-to-digital converters at a sampling rate of 125 Hz. The data were exported via the BrainFlow API, which internally applies the hardware scaling factor to convert raw digital values into microvolts. For compatibility with the MNE-Python analysis framework, these values were subsequently converted to volts by applying a 10^{-6} multiplier.

The system transmitted data wirelessly to a computer via an RFduino Bluetooth module. The wireless and child-friendly design was particularly advantageous for paediatric participants, as it minimised mechanical artifacts, cable-induced noise, and involuntary movements. During the preparation phase, participants were seated

Figure 3.1: EEG recording setup showing the Ultracortex Mark IV headset with 16 dry electrodes and the OpenBCI CytonDaisy acquisition system.

facing an empty wall without visual stimulation. After placing the electrocap and verifying that electrode impedances were below 10 kΩ, recordings were initiated with participants' eyes closed.

A total of 98 auditory stimuli were presented to each participant. Each stimulus was marked in the continuous EEG data using an event marker (trigger). The hardware marker pulse fired at the stimulus offset (end of the audio). The stimulus onset was retrospectively recovered during preprocessing by subtracting the known duration of the audio clip from the offset marker.

The stimuli varied across six dimensions: syntax type, pitch, speed, background noise, direction of presentation, and speaker distance. The specific conditions are detailed in Table 3.1.

Table 3.1: Experimental stimuli conditions and parameters.

Variable	**Conditions**
Syntax Type	Declarative, Imperative, Interrogative
Pitch	Normal, High, Low
Speed	Normal, Slow, Fast
Noise	No noise, White noise
Direction	Binaural, Monaural Left, Monaural Right
Distance	0.35 m, 0.5 m, 1 m, 2 m

Based on presentation speed, the stimuli fall into three duration categories. The marker indices and durations for each category are summarised in Table 3.2.

Table 3.2: Classification of stimuli based on presentation speed and duration.

Stimulus Type	Marker Indices	Duration (s)
Slow	69–75, 77–78	4.5
Fast	79–86, 88	1.5
Normal	All other markers	3.0

The exhaustive list of all 98 stimuli, detailing the specific combinations of acoustic manipulations and sentence types for each trigger, is provided in Appendix A.

3.1.2 Signal Preprocessing

All preprocessing was performed using MNE-Python (version 1.12.1) [34]. The complete pipeline is illustrated in Figure 3.2 and described in detail below. The same procedure was applied identically to every participant.

Spectral filtering. A notch filter centred at 50 Hz was first applied to attenuate power line interference from the European mains supply. A zero-phase bandpass filter was then applied with a passband of 0.1–40 Hz. The 0.1 Hz high-pass cutoff removes slow baseline drifts unrelated to neural activity while preserving slow cortical potentials relevant to ERP analysis. The 40 Hz low-pass cutoff suppresses high-frequency electromyographic contamination while retaining all ERP components examined in this study [14].

Common average reference. Following spectral filtering, a Common Average Reference (CAR) was computed across all 16 channels. By subtracting the mean signal of all electrodes from each individual channel, CAR reduces the contribution of spatially diffuse noise sources while preserving topographically localised neural activity. This referencing strategy is standard practice for low-density EEG montages where dedicated reference electrodes are not available [35].

Ocular artifact removal via ICA. Independent Component Analysis (ICA) was applied to identify and remove ocular and muscular artifacts. Because ICA decomposition is sensitive to slow drifts, a temporary copy of the EEG data was first high-pass filtered at 1.0 Hz. The FastICA algorithm was then fitted to extract up to 15 independent components. The frontal electrodes Fp1 and Fp2 were employed as proxy channels to automatically identify blink and eye-movement components via correlation. The spatial weights of the identified artifactual components were excluded, and the correction was applied back to the primary 0.1–40 Hz filtered signal. In cases where ICA decomposition failed due to very short or excessively noisy recordings, preprocessing continued without ICA correction.

Epoch extraction and artifact rejection. Epochs were extracted starting 200 ms prior to stimulus onset (baseline period) and extending to 1000 ms after stimulus

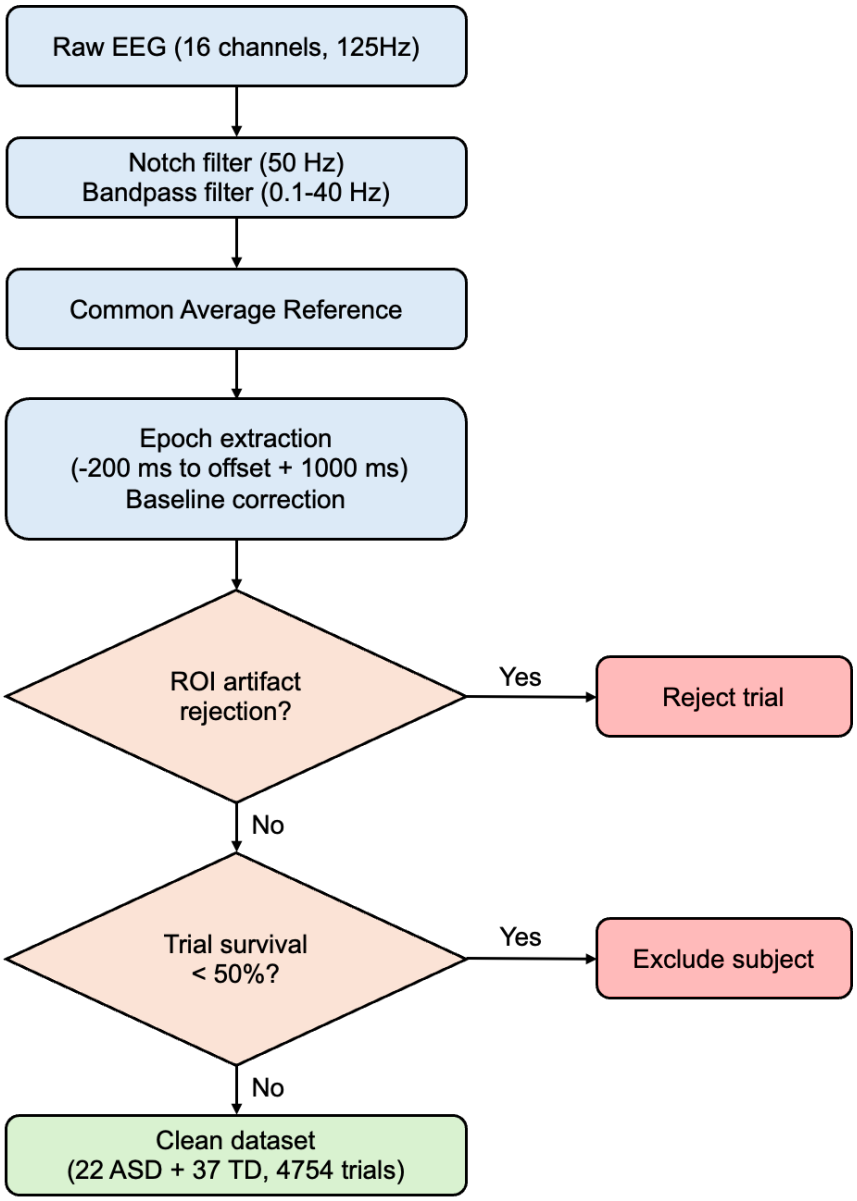

Figure 3.2: EEG preprocessing pipeline. Each recording passed through spectral filtering, common average re-referencing, and ICA-based artifact removal before epoch extraction. Trials exceeding the 200 μV amplitude threshold on any ROI channel were rejected. Participants retaining fewer than 50% of trials were excluded.

offset, capturing all onset ERP components as well as the offset-locked N400 and P600. Baseline correction was applied using the −200 to 0 ms pre-stimulus window. Artifact rejection was targeted solely at the ERP region of interest (ROI: F3, F4, C3, C4, P3, P4). A trial was rejected only if the absolute peak amplitude in any of these six ROI channels exceeded 200 μV. Noise occurring on peripheral channels did not trigger trial rejection, preserving trials that contained usable ERP data on the ROI electrodes despite transient artifacts elsewhere. This targeted strategy accommodates the elevated rate of movement artifacts typical in paediatric ASD populations while rejecting gross motor contamination of the analysis-relevant channels.

Subject exclusion. Participants retaining fewer than 50% of the original 98 trials (i.e., fewer than 49 trials) after artifact rejection were excluded from all subsequent analyses. This threshold balances the need for sufficient trial counts to produce stable ERP estimates against the risk of excessive data loss in a paediatric clinical population.

3.1.3 ERP Feature Extraction

Following preprocessing, ERP features were extracted from the trial-level EEG matrices. Six ERP components were examined, selected on the basis of their established relevance to auditory processing, attentional allocation, and language comprehension [14, 16, 18, 19, 20].
The component latency windows and regions of interest (ROIs) were defined according to standard ERP component descriptions and adapted to the low-density electrode montage used in the present study [14, 22, 35].

Table 3.3: ERP component definitions. Onset components are measured relative to stimulus onset. Offset components are measured relative to stimulus offset and use a local baseline correction.

Component	Window (ms)	Lock	ROI electrodes	Functional role
P1	40–90	Onset	F3, F4, C3, C4	Early sensory registration
N1	80–150	Onset	F3, F4, C3, C4	Auditory attention orienting
P2	150–250	Onset	F3, F4, C3, C4	Sensory gating
P300	250–500	Onset	P3, P4, C3, C4	Attentional resource allocation
N400	300–500	Offset	P3, P4, C3, C4	Semantic integration
P600	500–800	Offset	P3, P4, C3, C4	Syntactic repair

ROI electrode selection. Two ROIs were defined based on the established topographic distributions of the target ERP components and adapted to the 16-channel montage used in the present study. As illustrated in Figure 3.3, fronto-central electrodes (F3, F4, C3, C4) were used for the early sensory components (P100, N100, P200), which are commonly observed over fronto-central scalp regions for auditory stimuli [14]. Centro-parietal electrodes (P3, P4, C3, C4) were used for the later

attention- and language-related components (P300, N400, P600), reflecting the typical centro-parietal distribution of the P300 and the posterior or centro-parietal involvement often reported for language-related ERP components [16, 18, 19, 20]. Because the 16-channel montage does not include midline electrodes such as Fz, Cz, or Pz, the selected lateral electrode pairs were used as the closest available approximation to conventional midline ROIs. This adaptation is consistent with the practical constraints of low-density EEG recordings and with the use of fronto-central channels in comparable auditory ERP studies involving ASD [35, 8]. The central electrodes (C3, C4) were included in both ROIs to provide spatial continuity between the fronto-central and centro-parietal regions.

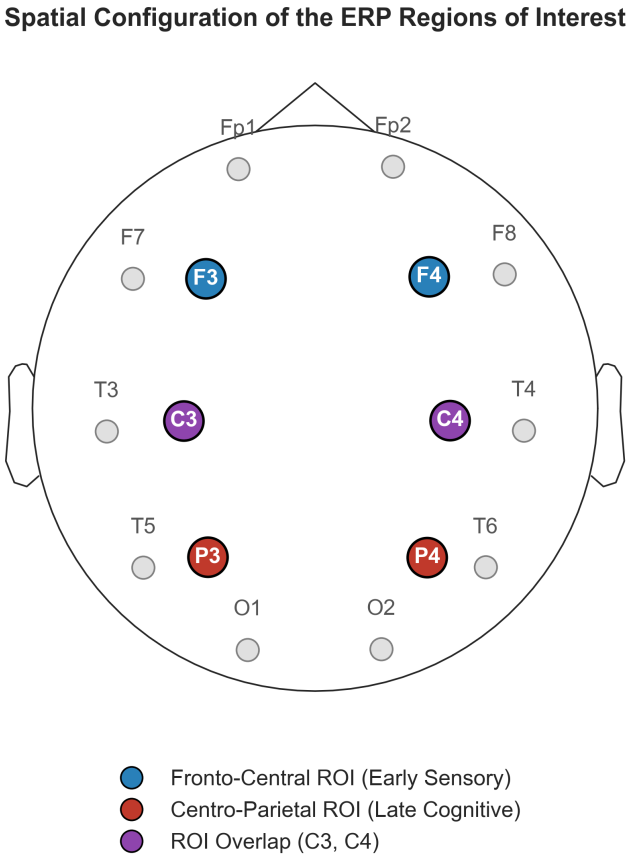

Figure 3.3: Topographic map illustrating the overlapping ROIs mapped on the 16-channel OpenBCI montage.

Onset ERP extraction. For each trial, the signal within each component time window was averaged across the four ROI channels to produce a single waveform segment. The mean voltage across this segment was taken as the component amplitude, and the time point of the peak voltage (minimum for negative components, maximum for positive) was taken as the component latency. Mean amplitude was preferred over peak amplitude because it is more robust to trial-level noise, particularly at the low sampling rate of 125 Hz [14].

Offset ERP extraction. Offset components (N400, P600) were measured relative to stimulus offset rather than onset, as they reflect post-stimulus cognitive processing that begins after the sentence has ended. Because the three speed conditions produce different stimulus durations (Fast: 1.5 s, Normal: 3.0 s, Slow: 4.5 s), the offset time point differs across trials. For each trial, a local baseline correction was applied by subtracting the mean voltage in the 200 ms immediately preceding stimulus offset from the post-offset signal. Amplitude and latency were then extracted using the same procedure as for onset components.

Global Field Power. GFP was computed as the spatial standard deviation across all 16 channels at each time point [21]. A net GFP value was derived for each trial by subtracting the mean GFP during the 200 ms pre-stimulus baseline from the mean GFP during the stimulus period. This measure captures the overall level of neural engagement regardless of topographic distribution and is not affected by the choice of reference electrode, making it particularly suitable for low-density montages where re-referencing can introduce polarity ambiguities in individual channel amplitudes.

Grand average waveforms. For visualisation, trial-level EEG matrices were averaged in two stages. First, within each participant, all trials belonging to a given Group $\times$ Condition cell were averaged to produce a subject-level waveform. Second, subject-level waveforms were averaged across participants within each group to produce the grand average. Shaded regions in all waveform figures represent ± 1 standard error of the mean across subjects. Onset plots display the interval from -200 to $+600$ ms relative to stimulus onset. Offset plots display the interval from -200 to $+900$ ms relative to stimulus offset, with individual trials aligned to their respective offset times before averaging.

3.1.4 Preliminary Group-Level Analysis

To assess whether stable group-level differences exist between ASD and TD children in auditory ERP biomarkers, a systematic statistical comparison was conducted across all six experimental dimensions. This analysis serves a dual purpose: it characterises the dataset at the population level and, critically, evaluates whether group-level contrasts are of sufficient magnitude and consistency to support personalized intervention design. As presented in Section 4.2, the outcome of this analysis motivates the shift from between-group statistics to the within-subject approach adopted in the personalized recommendation system.

All statistical tests were performed on subject-level mean values. For each ERP feature, trial-level values were first averaged within each participant for each condition, producing one value per subject per condition. Group differences were assessed using independent-samples Welch's t-tests (ASD versus TD) for each condition. Effect sizes were quantified using Cohen's d. Mixed-design analyses of variance were conducted with Group (ASD, TD) as the between-subjects factor and the experimental variable (e.g., Speed, Noise, Distance) as the within-subjects factor. Only participants contributing data to all levels of the within-subjects factor were included to ensure balanced designs. Statistical analyses were performed using Pingouin [36]

and SciPy [37].

Because the experimental design varies each acoustic dimension independently while holding others at baseline, each dimension was analysed separately. False discovery rate correction was applied to the full set of t-test p-values using the Benjamini-Hochberg procedure [38].

3.2 Deep Learning Framework for ASD/TD Classification

This section describes the trial-level binary classification framework used to distinguish children with ASD from typically developing peers from raw EEG signals. Three model architectures are compared under identical experimental conditions: a support vector machine serving as the classical machine learning baseline, and two deep learning architectures of increasing complexity. All models receive the same raw spatiotemporal EEG representation as input, ensuring that performance differences reflect architectural capacity rather than differences in feature engineering.

The framework was designed to investigate whether ASD-related neural signatures are reliably detectable at the single-trial level and whether subject-level contextualisation improves classification robustness. In addition to comparing architectures of increasing complexity, a subject-baseline difference representation was introduced to examine whether ASD-related neural alterations are more effectively characterised as deviations from an individual's typical neural response pattern rather than absolute waveform morphology. Figure 3.4 summarises the input preparation procedure, including temporal cropping, subject-independent partitioning, training-set normalisation, and construction of the subject-baseline difference representation.

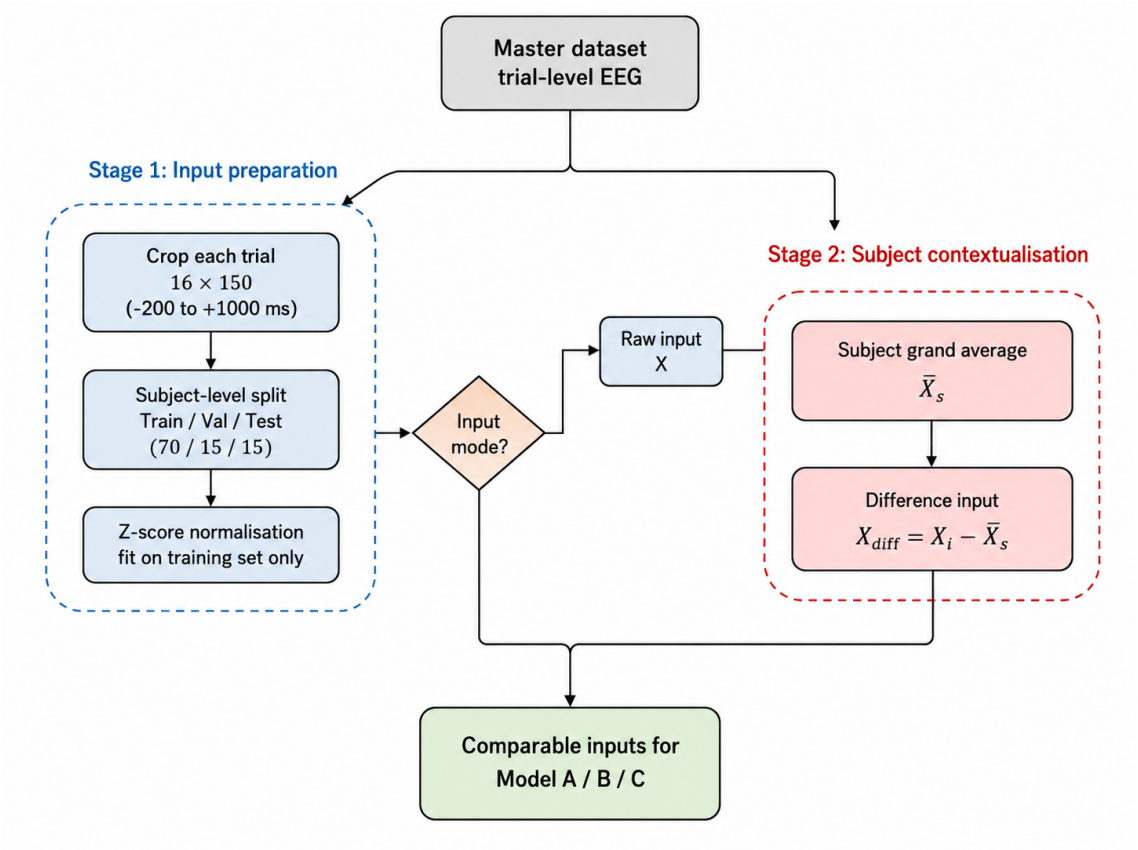

Figure 3.4: Input preparation and subject-baseline difference representation for EEG classification.

3.2.1 Data Preparation

Temporal cropping. Each trial in the master dataset was stored as a matrix of shape $(16 \times T)$, where T varies across speed conditions (Fast: 338 samples; Normal: 526 samples; Slow: 688 samples). To obtain a uniform input representation, each matrix was cropped to the first 150 samples, corresponding to the interval from -200 ms (baseline onset) to $+1000$ ms relative to stimulus onset. This window was selected because it is shorter than the minimum epoch length across all speed conditions, requiring no zero-padding, and fully encompasses all onset ERP components of interest including the P300 at up to 500 ms. Each trial was thus represented as a tensor of shape $(16, 150)$, labelled with the binary class of its corresponding participant (ASD $= 1$, TD $= 0$).

Subject-baseline difference representation. Single-trial EEG signals contain substantial inter-subject variability arising from stable individual neural characteristics unrelated to ASD-specific processing differences. To reduce the influence of such baseline activity, a subject-baseline difference representation was constructed. For each participant, a grand-average EEG template was first computed by averaging

all retained trials belonging to that participant:

$$\bar{X}_s = \frac{1}{N_s} \sum_{i=1}^{N_s} X_i \tag{3.1}$$

where X_i denotes the EEG matrix of trial i, and N_s is the number of retained trials for subject s. Each trial was then transformed into the difference representation:

$$X_{\text{diff}} = X_i - \bar{X}_s \tag{3.2}$$

This operation emphasises deviations from the participant's typical neural response pattern while suppressing stable subject-specific baseline activity unrelated to the classification target. Unlike simple channel concatenation, the difference representation preserves the original input dimensionality $(16, 150)$ while explicitly encoding relative neural dynamics at the single-trial level. Because the dataset was partitioned independently at the subject level (see below), each participant's grand-average template was computed only from trials belonging to that participant, introducing no cross-subset information leakage.

Subject-independent partitioning. A critical methodological requirement for EEG-based classification is that the model must generalise to unseen individuals rather than memorising the idiosyncratic neural signatures of participants seen during training. To enforce this, the dataset was partitioned at the subject level: all trials from a given participant were assigned exclusively to one of three subsets (training, validation, or test), with no subject appearing in more than one subset. This strategy eliminates the information leakage that would arise from a trial-level split. Partitioning was performed using subject-level stratified shuffling via StratifiedShuffleSplit from scikit-learn [39], with an approximate 70/15/15 allocation by subject count while preserving the ASD/TD ratio across subsets. The mild class imbalance (overall ASD:TD trial ratio of 0.600) was addressed during training through a positive class weight equal to the ratio of TD to ASD trials in the training set (see Section 3.2.3).

Per-channel normalisation. Per-channel z-score normalisation was applied to stabilise gradient-based optimisation and ensure that no single channel dominated the input representation by virtue of its absolute amplitude. The channel-wise mean and standard deviation were computed over all training trials and all time points, then applied identically to training, validation, and test data. Normalisation statistics were derived exclusively from the training set to prevent distributional information from validation or test subjects from influencing the training process. For the difference representation, the transformed EEG signals were normalised using the same training-set statistics derived from the raw training data.

3.2.2 Model Architectures

Three model architectures were evaluated, each receiving a $(16, 150)$ EEG tensor as input. The models are described in order of increasing architectural complexity and are summarised in Figure 3.5.

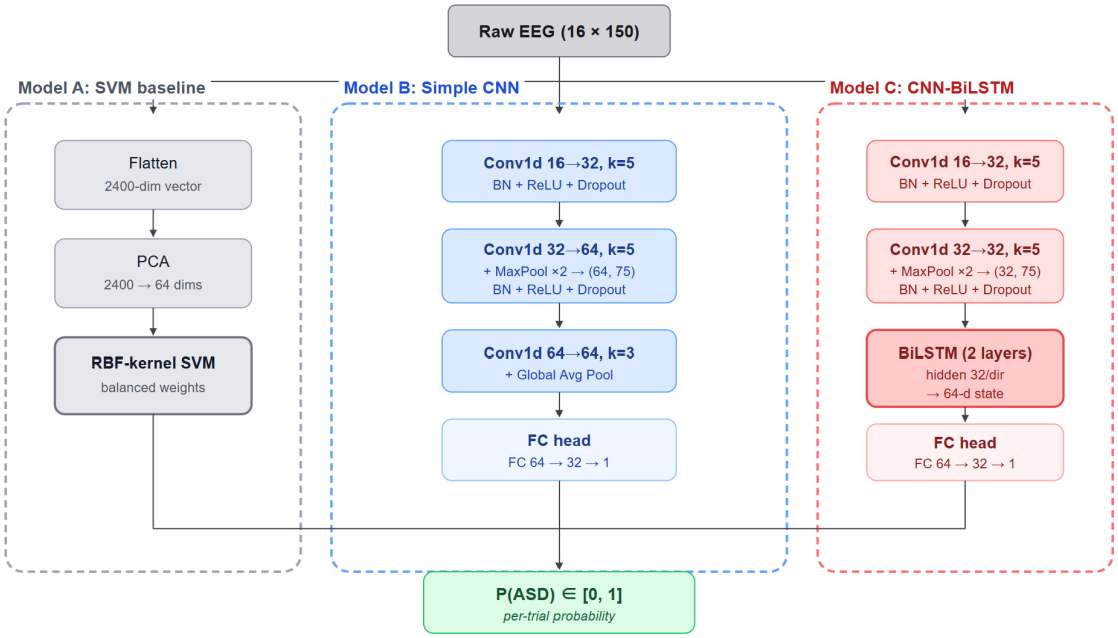

Figure 3.5: Comparison of the three EEG-based classification architectures evaluated in this study, shown side by side over the same (16×150) raw EEG input.

Model A: Support Vector Machine. A support vector machine (SVM) with a radial basis function kernel served as the classical machine learning baseline. Each trial was first flattened into a vector of $16 \times 150 = 2{,}400$ dimensions. Principal component analysis [40] was then applied to reduce this representation to 64 components, fitted exclusively on the training set, to regularise the high-dimensional input and reduce computational cost. The SVM was trained with balanced class weighting, regularisation parameter $C = 1.0$, and the scale heuristic for the kernel bandwidth ($\gamma = 1/(d \cdot \mathrm{Var}(X))$, where d is the number of features). Probability estimates were calibrated via Platt scaling to enable AUC-ROC computation. The SVM was selected as the baseline because it is well-established in EEG classification literature and provides a meaningful reference point for assessing the representation-learning advantage of deep architectures on the same raw input.

Model B: Convolutional Neural Network. A one-dimensional convolutional neural network was implemented as the deep learning baseline. The architecture comprised two convolutional blocks followed by a classification head. Each block consisted of a `Conv1d` layer with kernel size 5 and same-padding, batch normalisation, ReLU activation, and dropout ($p = 0.30$). The first block produced 32 feature maps; the second produced 64 feature maps and was followed by max-pooling with stride 2, halving the temporal resolution to 75 time steps. A third convolutional layer with 64 filters and kernel size 3 extracted higher-order features, followed by global average pooling to collapse the temporal dimension. The classification head comprised two dropout layers ($p = 0.60$ and $p = 0.40$) interleaved with a fully connected layer of 32 units with ReLU activation, and a final linear projection to a scalar logit.

Model C: CNN-BiLSTM (Proposed Architecture). The proposed model augments the convolutional front-end with a bidirectional LSTM network to jointly capture spatial and temporal dependencies in the EEG signal. The convolutional front-end comprised two blocks (32 filters each, kernel size 5, batch normalisation, ReLU, and dropout at $p = 0.30$), followed by max-pooling, producing a representation of shape $(32, 75)$. The temporal sequence was then processed by a two-layer BiLSTM with hidden size 32 per direction, inter-layer dropout ($p = 0.40$), and bidirectional outputs concatenated to yield a 64-dimensional state vector at the final time step. The classification head was identical to Model B: two dropout layers ($p = 0.60$ and $p = 0.40$) bracketing a fully connected layer of 32 units, followed by a linear output projection. The bidirectional design allows the network to integrate information from both earlier and later time points within the epoch, which is advantageous for ERP components distributed across a temporal window rather than concentrated at a single latency.

3.2.3 Training Procedure

Models B and C were trained under identical conditions. Binary cross-entropy with logits was used as the loss function, with a positive class weight of $N_{\text{TD}}/N_{\text{ASD}}$ computed from the training set. Parameters were optimised using AdamW [41] with an initial learning rate of 5×10^{-4} and weight decay 10^{-3}. Gradient norms were clipped at 1.0 to prevent exploding gradients during BiLSTM training. Training proceeded for a maximum of 80 epochs with a mini-batch size of 64. A `ReduceLROnPlateau` scheduler halved the learning rate when validation loss failed to decrease for seven consecutive epochs, while early stopping with patience 15 terminated training when validation AUC-ROC did not improve. The two criteria were deliberately mixed: validation loss reacts faster to local optimisation plateaus and is therefore suitable for adaptive learning-rate adjustment, whereas AUC-ROC reflects discriminative performance directly and is therefore the appropriate criterion for model selection. The checkpoint with the highest validation AUC-ROC was retained for test-set evaluation and subsequent explainability analysis. Table 3.4 summarises all hyperparameters.

3.2.4 Evaluation and Explainability

Classification metrics. All three models were evaluated on the held-out test set, which contained no subjects seen during training or validation. Three primary metrics were computed: area under the receiver operating characteristic curve (AUC-ROC), F1-score, and overall accuracy. AUC-ROC was treated as the primary metric because it is threshold-independent and robust to class imbalance. Confusion matrices and ROC curves were generated for all models to facilitate direct visual comparison of classification behaviour.

Subject-level aggregation via trial voting. Subject-level classification was performed by aggregating trial-level predictions from the best-performing model (CNN-BiLSTM with difference input) via probability averaging. For each partici-

Table 3.4: Hyperparameters for deep learning model training (Models B and C).

Hyperparameter	Value
Optimiser	AdamW
Learning rate	5×10^{-4}
Weight decay	1×10^{-3}
Batch size	64
Maximum epochs	80
LR scheduler	ReduceLROnPlateau (patience = 7, factor = 0.5)
Early stopping patience	15 epochs
Intra-block dropout	0.30
Head dropout	0.60 / 0.40
BiLSTM inter-layer dropout	0.40 (Model C only)
Gradient clipping	1.0 (global norm)
Loss function	Binary cross-entropy with logits
Class weighting	$N_{\mathrm{TD}}/N_{\mathrm{ASD}}$

pant in the test set, the model was applied to every retained trial, and the resulting ASD probabilities were averaged across trials to yield a single subject-level estimate. The procedure is illustrated in Figure 3.6 and formalised in Equation 3.3.

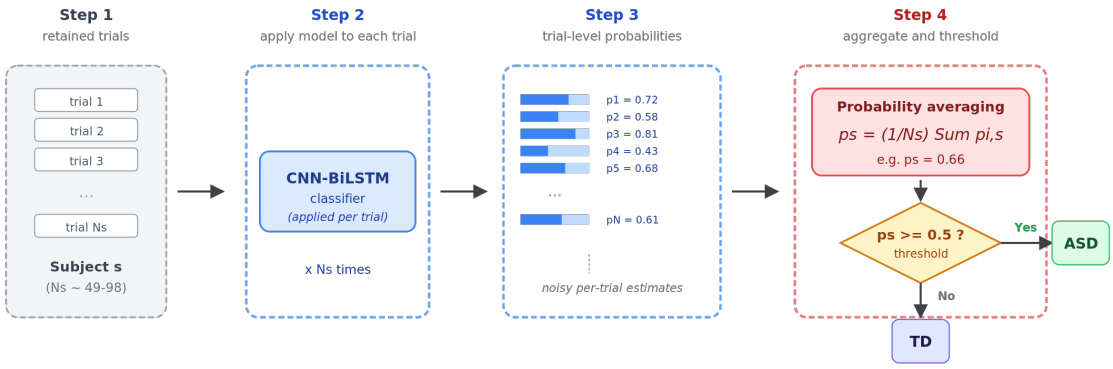

Figure 3.6: Subject-level prediction via probability averaging. Trial-level CNN-BiLSTM outputs are averaged across all retained trials of a subject and thresholded at 0.5.

$$\hat{p}_s = \frac{1}{N_s} \sum_{i=1}^{N_s} \hat{p}_{i,s} \tag{3.3}$$

where $\hat{p}_{i,s}$ denotes the predicted ASD probability for trial i of subject s, N_s is the number of retained trials for that subject, and $\hat{p}_s$ is the resulting subject-level ASD probability. This estimate was then thresholded at 0.5 to obtain a binary group prediction. The aggregation procedure is analogous to the grand-average approach in classical ERP analysis, where averaging across trials suppresses trial-level noise to reveal stable neural signatures. If ASD-related neural characteristics are genuinely stable individual-level traits rather than transient single-trial phenomena, aggregating predictions across trials should produce a more reliable group estimate than any

single epoch alone. The trial-level and subject-level AUC-ROC values were compared directly to quantify the benefit of aggregation, using the same underlying model and test set throughout.

SHAP-based channel attribution. To investigate which EEG channels and time points most influenced the classification decisions of the proposed CNN-BiLSTM model, SHAP values were computed using a gradient-based explainer [30]. For each test trial, SHAP assigned a scalar importance value to each element of the $(16, 150)$ input tensor, representing its marginal contribution to the predicted ASD probability. Channel importance was obtained by averaging absolute SHAP values over all test trials and all time points for each of the 16 channels. Temporal importance was similarly obtained by averaging over trials and channels, yielding a time-resolved importance profile aligned to the epoch time axis. The resulting importance maps were compared with the ERP regions of interest defined in Section 3.1.3 to assess whether the model identified neurophysiologically plausible spatial and temporal features as decision-relevant.

3.3 Personalized Recommendation System

Following the automated ASD screening described in Section 3.2, children identified as ASD proceed to an individualised acoustic assessment. This section describes the biomarker-based recommendation framework that determines the optimal acoustic configuration for each child. The complete two-stage pipeline is illustrated in Figure 3.7.

3.3.1 Neural Processing Efficiency Score

The design of the NPES draws on two established principles in cognitive neuroscience. First, composite indices that combine multiple z-scored EEG or ERP features into a single trial-level score have been successfully applied in domains ranging from infant cognition [42] to cognitive workload assessment [43], demonstrating that within-subject z-scoring and feature aggregation can yield meaningful trial-level predictions even in populations with high noise and limited trial counts. Second, the neural efficiency hypothesis holds that more efficient neural processing is characterised by faster information transmission (shorter ERP latencies) and more spatially focused cortical activation (lower GFP), a relationship that has been validated both behaviourally [44] and metabolically through the demonstrated linear relationship between GFP and respiratory energy expenditure [45]. The NPES integrates these two dimensions, latency-based processing speed and GFP-based processing effort, into a single composite score.

Component selection. Four biomarkers were selected from the six ERP components extracted in Section 3.1.3. The selected components and the rationale for their inclusion are as follows.
N1 latency (80–150 ms) reflects the speed of auditory attention orienting. Shorter latency indicates faster allocation of attention to the incoming stimulus. P2 latency

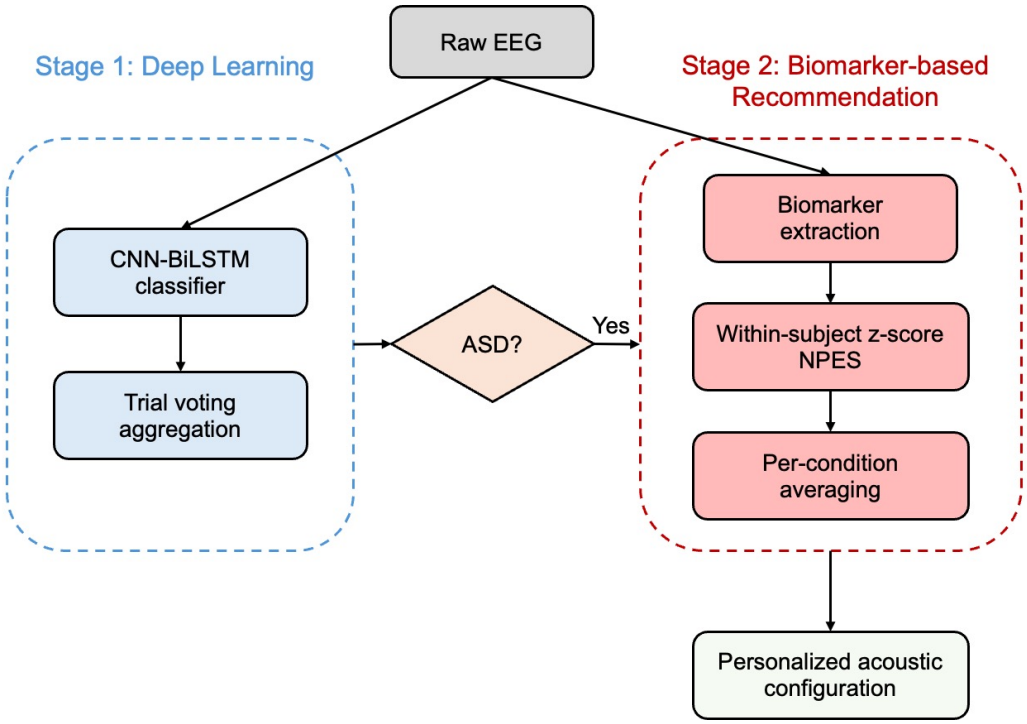

Figure 3.7: Two-stage personalized recommendation pipeline. Stage 1 employs a CNN-BiLSTM classifier with trial-level voting to screen for ASD. Stage 2 computes a composite Neural Processing Efficiency Score from four ERP biomarkers and recommends the acoustic condition maximising neural efficiency for each dimension.

(150–250 ms) reflects the speed of early cortical processing and sensory gating. P300 latency (250–500 ms) reflects the speed of higher-order cognitive resource allocation. For all three latency measures, lower values indicate faster processing.
GFP, computed as the spatial standard deviation across all 16 channels within the 0–1000 ms post-onset window, captures the overall spatial dispersion of neural activation. Lower GFP indicates more spatially concentrated processing, which under the neural efficiency hypothesis corresponds to reduced processing effort [21].
All four indicators follow a consistent "lower is better" convention, enabling their combination through a uniform sign inversion.

Excluded components. Two components were excluded from the NPES. The N400 and P600 are measured relative to stimulus offset rather than onset (Section 3.1.3). Because stimulus duration varies across the three speed conditions (1.5 s, 3.0 s, 4.5 s), the offset time point differs across trials, and each trial requires an individual local baseline correction. This offset-locked extraction introduces additional measurement variability that would increase noise in the composite score. The exclusion is further supported by the grand average analysis (Section 4.2.1), which showed that identifiable N400 and P600 components did not reliably emerge at the group level for minority speed conditions with limited trial counts.
Amplitude-based measures (e.g., N1 amplitude, P300 amplitude) were also excluded. As discussed in Section 4.2.2, the common average reference computed from 16 channels can alter the apparent polarity and magnitude of components at individual

electrodes. While this effect cancels out in within-subject comparisons when the reference is constant, it introduces an additional source of noise that latency and GFP measures are not subject to. The decision to use latency and GFP rather than amplitude therefore reduces the number of noise sources in the composite score.

Within-subject normalisation. For each participant, each of the four biomarkers was standardised to zero mean and unit variance across all of that participant's retained trials. This within-subject z-scoring, a standard approach in ERP analysis for removing inter-individual baseline differences [14], ensures that the NPES reflects relative efficiency across conditions within the same child rather than absolute differences between children. The NPES for trial i of participant j was computed as

$$\mathrm{NPES}_{i,j} = \frac{1}{4}\sum_{k=1}^{4}\Big(-z_{i,j,k}\Big), \tag{3.4}$$

where $z_{i,j,k}$ denotes the within-subject z-score of the k-th biomarker for trial i. The sign inversion converts the "lower is better" convention to "higher is better," so that higher NPES indicates more efficient neural processing. Equal weighting was adopted to avoid introducing assumptions about the relative importance of each biomarker. The validity of this choice is evaluated empirically through the biomarker contribution analysis in Section 4.4.3.

3.3.2 Acoustic Parameter Optimisation

For each ASD participant, the NPES was averaged across all trials sharing the same condition level within each acoustic dimension, yielding a condition-level efficiency profile. The recommended condition for a given dimension was the one with the highest mean NPES:

$$c_d^* = \arg\max_{c \in \mathcal{C}_d} \frac{1}{|\mathcal{T}_{j,c}|} \sum_{i \in \mathcal{T}_{j,c}} \mathrm{NPES}_{i,j}, \tag{3.5}$$

where d indexes the acoustic dimension, $\mathcal{C}_d$ is the set of conditions for dimension d (e.g., $\{\mathrm{Slow}, \mathrm{Normal}, \mathrm{Fast}\}$ for Speed), and $\mathcal{T}_{j,c}$ is the set of trial indices for participant j under condition c. The five dimensions were optimised independently, producing a five-parameter acoustic configuration per child.

This within-subject optimisation is motivated by the neural efficiency hypothesis: the condition under which the brain allocates the fewest and most spatially concentrated resources is the one under which the auditory input is most compatible with the individual's processing capacity. The resulting recommendation is specific to the child and not transferable to other individuals, as the NPES profile depends on the individual's own neural response pattern.

3.3.3 Recommendation Evaluation

The recommendation framework was evaluated along three dimensions.

Heterogeneity analysis To assess whether personalisation is necessary, the distribution of recommended conditions across all 22 ASD participants was examined for each acoustic dimension. If a single condition dominated the recommendations (e.g., more than 80% of participants), a universal recommendation would suffice and personalisation would offer little additional benefit. Diverse recommendations across participants would confirm that ASD children exhibit heterogeneous acoustic processing profiles.

Biomarker contribution analysis To determine whether the composite NPES provides a more comprehensive recommendation than any single biomarker, each of the four constituent biomarkers was used independently to generate single-biomarker recommendations by selecting the condition with the lowest value of that biomarker. The agreement rate between each single-biomarker recommendation and the full NPES recommendation was computed per dimension. If all four biomarkers agreed with the NPES recommendation, the composite would be redundant. If agreement varied across biomarkers and dimensions, the composite captures complementary information not available from any single indicator.

Split-half stability analysis To assess the reliability of the recommendations given the limited number of trials per minority condition, each ASD participant's trials were randomly divided into two equal halves. Recommendations were computed independently from each half, and the agreement rate between the two halves was recorded. This procedure was repeated 50 times with different random splits to obtain a stable estimate. The resulting agreement rate was compared against the chance baseline for each dimension, defined as the reciprocal of the number of conditions (33% for three-condition dimensions, 50% for two-condition dimensions, and 25% for four-condition dimensions).

4

Results

This chapter presents the experimental findings. Section 4.1 describes the data quality and final cohort characteristics. Section 4.2 reports the group-level ERP analysis and its implications for the analytical approach. Section 4.3 presents the trial-level and subject-level classification results. Section 4.4 evaluates the personalized recommendation system.

4.1 Data Preprocessing and Quality Control

Following the preprocessing and quality control pipeline described in Section 3.1.2, a total of 4,754 trials were retained across 59 participants. Table 4.1 summarises the retention statistics by group. The ASD group showed substantially lower and more variable trial survival rates than the TD group, consistent with the elevated movement and ocular artifact rates typical of paediatric ASD populations. Participants with fewer than 50% of trials surviving the ROI amplitude threshold were excluded from further analysis, resulting in 22 retained ASD participants and 37 retained TD participants.

Table 4.1: Participant retention and trial survival statistics after preprocessing and quality control. Values represent the number of participants and mean trial counts per group. SD: standard deviation.

Group	Total	Retained	Dropped	Avg. trials	Survival (%)	SD (%)
ASD	43	22	21	49.0	50.0	37.3
TD	46	37	9	68.6	70.0	29.1

Figure 4.1 illustrates the effect of spectral filtering on the power spectral density (PSD) of the EEG signal, averaged across all 16 channels for a representative participant. Prior to filtering, the signal is dominated by low-frequency power concentrated below 5 Hz, arising from slow cortical drifts and electrode settling. Following application of the 50 Hz notch filter and the 0.1–40 Hz zero-phase bandpass filter, the DC component and high-frequency noise are attenuated, yielding a flatter spectrum within the analysis band. The residual peak visible near 25 Hz in the filtered spectrum reflects a minor ringing artefact at the subharmonic of the notch frequency introduced by the FIR filter design; this does not affect any ERP components of interest, which are distributed below 15 Hz.
Figure 4.2 shows a representative accepted and rejected trial, illustrating the ROI-based amplitude threshold criterion. In the accepted trial (upper panel), all six ROI

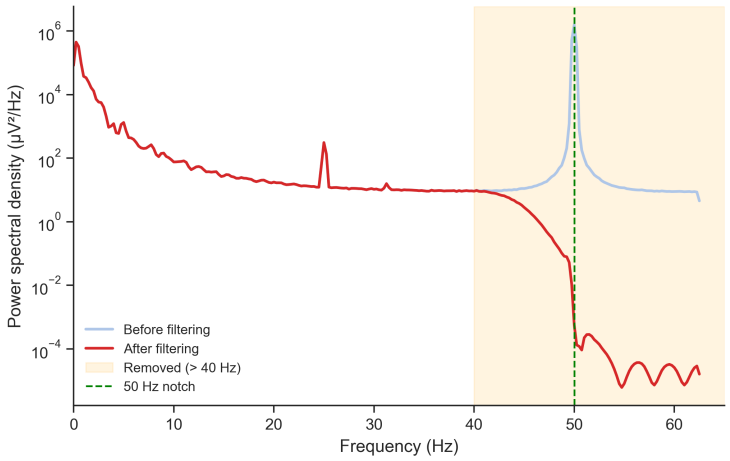

Figure 4.1: Mean power spectral density across 16 channels before and after preprocessing, displayed on a logarithmic scale. Shaded regions indicate frequency bands removed by the 0.1 Hz high-pass and 40 Hz low-pass filters. The green dashed line marks the 50 Hz notch filter.

channels remain within the ±200 μV threshold throughout the epoch, indicating a clean recording suitable for ERP analysis. In the rejected trial (lower panel), at least one ROI channel exceeds this threshold, reflecting a transient movement or ocular artefact. Grey traces correspond to non-ROI channels, which did not contribute to the rejection decision.

Table 4.2 reports the mean (SD) number of valid trials per condition per participant for each experimental dimension. The distribution reflects the design of the stimulus set, in which a small number of conditions within each factor were presented with fewer repetitions than the predominant baseline condition. Notably, the No-noise, Normal-speed, Both-ear, Normal-pitch, and 0.35 m distance conditions each account for the large majority of trials per participant, whereas the contrasting conditions (e.g., Yes-noise: 6–8 trials; Fast and Slow speed: 7–9 trials) were presented less frequently. This asymmetry should be considered when interpreting statistical comparisons across conditions, as the higher-repetition conditions yield more stable ERP averages. The balance between ASD and TD groups was consistent across all conditions, indicating no systematic condition-level attrition bias.

4.2 Group-Level ERP Analysis

4.2.1 Grand Average Waveforms

Grand average ERP waveforms were computed for each acoustic dimension to verify that the passive auditory paradigm produced measurable auditory responses and to examine any visible differences between the ASD and TD groups. Figure 4.3 presents the onset-locked waveforms for the Noise condition. Under the noise-absent condition, both groups showed early deflections in the expected N1 and P2 time windows, though the components were not sharply defined at the grand average level. Under the noise-present condition, waveform variability increased in both

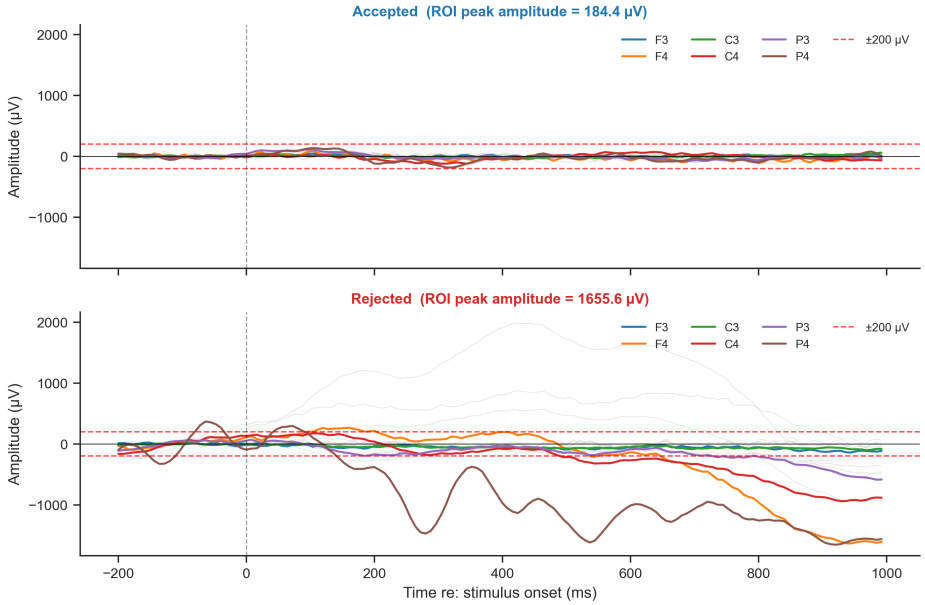

Figure 4.2: Representative accepted (upper) and rejected (lower) trial epochs shown for the six ROI channels (coloured traces: F3, F4, C3, C4, P3, P4) and the ten non-ROI channels (grey traces). The red dashed lines indicate the ±200 μV ROI rejection threshold. Trials were retained only if all ROI channels remained within this threshold.

Table 4.2: Mean (SD) number of valid trials per participant per condition after quality control, for the ASD and TD groups separately.

Factor	Condition	ASD ($n = 22$) Mean (SD)	TD ($n = 37$) Mean (SD)
Speed	Fast	8.5 (1.0)	8.2 (1.3)
	Normal	65.0 (13.4)	65.2 (12.2)
	Slow	7.6 (2.0)	6.8 (2.6)
Noise	No	74.3 (13.3)	72.7 (13.4)
	Yes	6.8 (3.0)	7.6 (2.0)
Distance	0.35 m	59.0 (10.5)	58.8 (10.9)
	0.5 m	7.3 (1.8)	7.2 (1.8)
	1 m	9.7 (2.3)	9.4 (2.3)
	2 m	5.0 (1.0)	4.9 (1.2)
Ear	Both	66.8 (12.3)	66.2 (12.2)
	Left	6.5 (2.4)	6.6 (1.8)
	Right	7.7 (1.4)	7.5 (1.7)
Pitch	High	7.6 (2.1)	8.0 (1.6)
	Low	7.9 (1.4)	7.5 (1.9)
	Normal	65.5 (12.6)	64.8 (13.1)
Sentence type	Declarative	26.8 (4.8)	26.6 (5.3)
	Imperative	26.8 (5.9)	27.1 (5.3)
	Interrogative	27.5 (5.3)	26.5 (5.3)

groups, as reflected in the wider standard error bands, and the early components became less distinguishable.

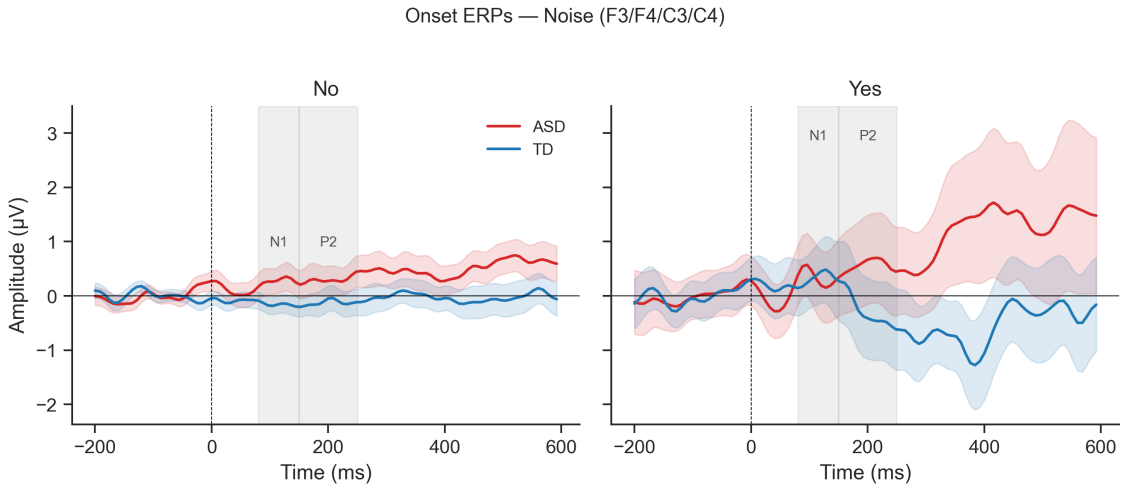

Figure 4.3: Grand average onset ERPs for the Noise dimension (ROI: F3, F4, C3, C4). Shaded regions indicate ±1 SEM across subjects. Grey bands mark the N1 and P2 time windows.

Figure 4.4 shows the onset-locked waveforms for monaural left, monaural right, and binaural presentation. The waveform morphology varied across presentation modes, with the monaural conditions producing larger amplitude responses than the binaural condition. In the Left and Right conditions, the ASD and TD waveforms showed substantial overlap throughout the epoch. In the Both condition, the two groups showed a visible separation in the post-stimulus period, with the ASD group trending toward more positive values, although this difference did not reach statistical significance in the formal analysis (Section 4.2.2).

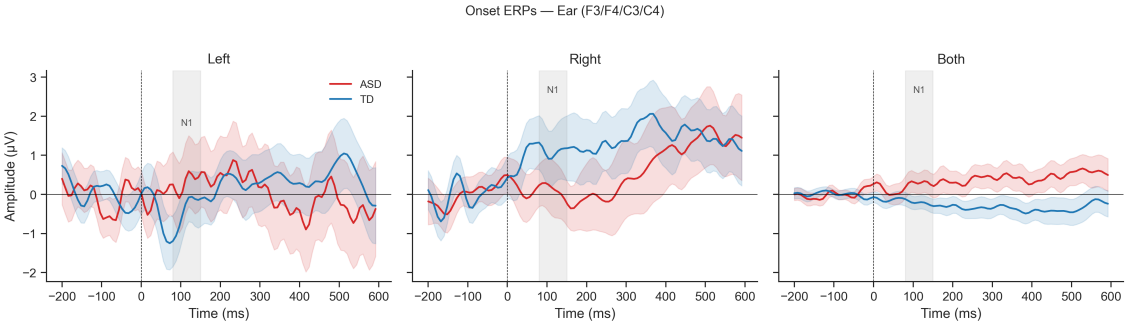

Figure 4.4: Grand average onset ERPs for the Ear dimension (ROI: F3, F4, C3, C4). Left, Right, and Both (binaural) presentation conditions are shown in separate panels.

To illustrate the effect of trial count on waveform quality, Figure 4.5 presents the offset-locked grand averages for the Pitch dimension. The Normal pitch condition, which comprises approximately 65 trials per subject, produced stable waveforms with narrow standard error bands. By contrast, the Low and High pitch conditions,

which comprise only 7–8 trials per subject, produced waveforms with wide standard error bands spanning several microvolts. No identifiable P600 component emerged at the grand average level for these minority conditions. This pattern was observed across all minority conditions in the dataset (Speed Fast/Slow, Noise Yes, Distance 0.5/2 m, Ear Left/Right), indicating that the unbalanced trial distribution limits the reliability of condition-specific grand averages when trial counts are small.

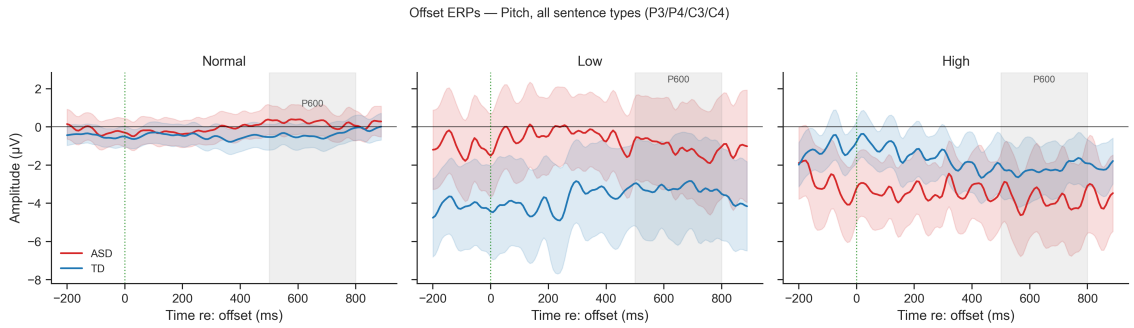

Figure 4.5: Grand average offset ERPs for the Pitch dimension (ROI: P3, P4, C3, C4). The Normal condition (approximately 65 trials per subject) produces stable waveforms, while the Low and High conditions (approximately 7–8 trials per subject) show large inter-subject variability.

4.2.2 Statistical Comparison: ASD versus TD

A total of 38 Welch's t-tests were conducted, comparing ASD and TD groups on ERP amplitude, latency, and GFP measures across all acoustic conditions. After FDR correction, none of the 38 comparisons reached statistical significance ($p_{\text{FDR}} < 0.05$). Effect sizes were generally small, with a median $|d| < 0.3$ across all comparisons. Mixed ANOVAs similarly revealed no significant Group $\times$ Condition interactions for any acoustic dimension.

Two factors likely contribute to the absence of significant group effects. First, the small sample size (22 ASD, 37 TD) provides limited statistical power to detect the small to moderate effect sizes typically reported in paediatric ASD ERP studies. Second, as illustrated in Figure 4.5, minority conditions with fewer than 10 trials per subject produce unstable subject-level averages, further reducing the sensitivity of group-level contrasts for these conditions.

Among the biomarkers examined, GFP showed the most consistent behaviour across conditions. As GFP is computed as the spatial standard deviation across all channels, it is not dependent on the polarity at any individual electrode. However, even GFP did not yield significant group differences after correction. While GFP does not differentiate ASD from TD at the group level in this dataset, it remains a stable measure of neural processing effort that is suitable for within-subject comparisons across conditions.

4.2.3 Implications for the Analytical Approach

The absence of reliable group-level ERP differences has a direct consequence for the design of the personalized recommendation system. A recommendation framework based on population-level biomarker contrasts would require stable and significant group differences as its foundation. Since such differences were not observed in the present dataset, the recommendation system instead adopts a within-subject approach in which each child's neural processing efficiency is characterised relative to their own baseline across conditions. This within-subject strategy, implemented through the NPES described in Section 3.3.1, does not require group-level contrasts and instead relies on the relative ordering of conditions within each individual.

4.3 Deep Learning Classification

4.3.1 Test Set Composition Across Splits

Because the dataset was partitioned at the subject level, the number of subjects in each test set was fixed by the 70/15/15 split ratio. However, the number of trials varied across splits, as different participants contributed different numbers of retained trials. Figure 4.6 illustrates this variation across 100 random splits, and Table 4.3 summarises the key statistics.

Table 4.3: Summary statistics of test set composition across 100 random splits.

Variable	Mean	SD	Min	Max
ASD subjects per split	3.0	0.0	3.0	3.0
TD subjects per split	6.0	0.0	6.0	6.0
Total subjects per split	9.0	0.0	9.0	9.0
ASD trials per split	242.0	24.1	188.0	285.0
TD trials per split	479.7	30.7	415.0	552.0
Total trials per split	721.7	40.1	627.0	821.0
ASD trial fraction	0.335	0.026	0.280	0.398

The number of subjects per split was constant across all 100 partitions (3 ASD, 6 TD), as expected from the stratified split procedure. However, the total number of trials varied between 627 and 821, with ASD trials ranging from 188 to 285. This variation reflects differences in trial survival rates across participants: some subjects retained more trials after artifact rejection than others. As a result, the class balance in the test set also varied, with the ASD trial fraction ranging from 0.28 to 0.40. This variation in test set composition is a likely source of the performance variance observed across splits in the classification results reported below.

4.3.2 Trial-Level Classification

Table 4.4 shows the trial-level classification results for all six model configurations. For the deep learning models, the difference input gave equal or better results than the raw input, while for the SVM the two inputs performed similarly.

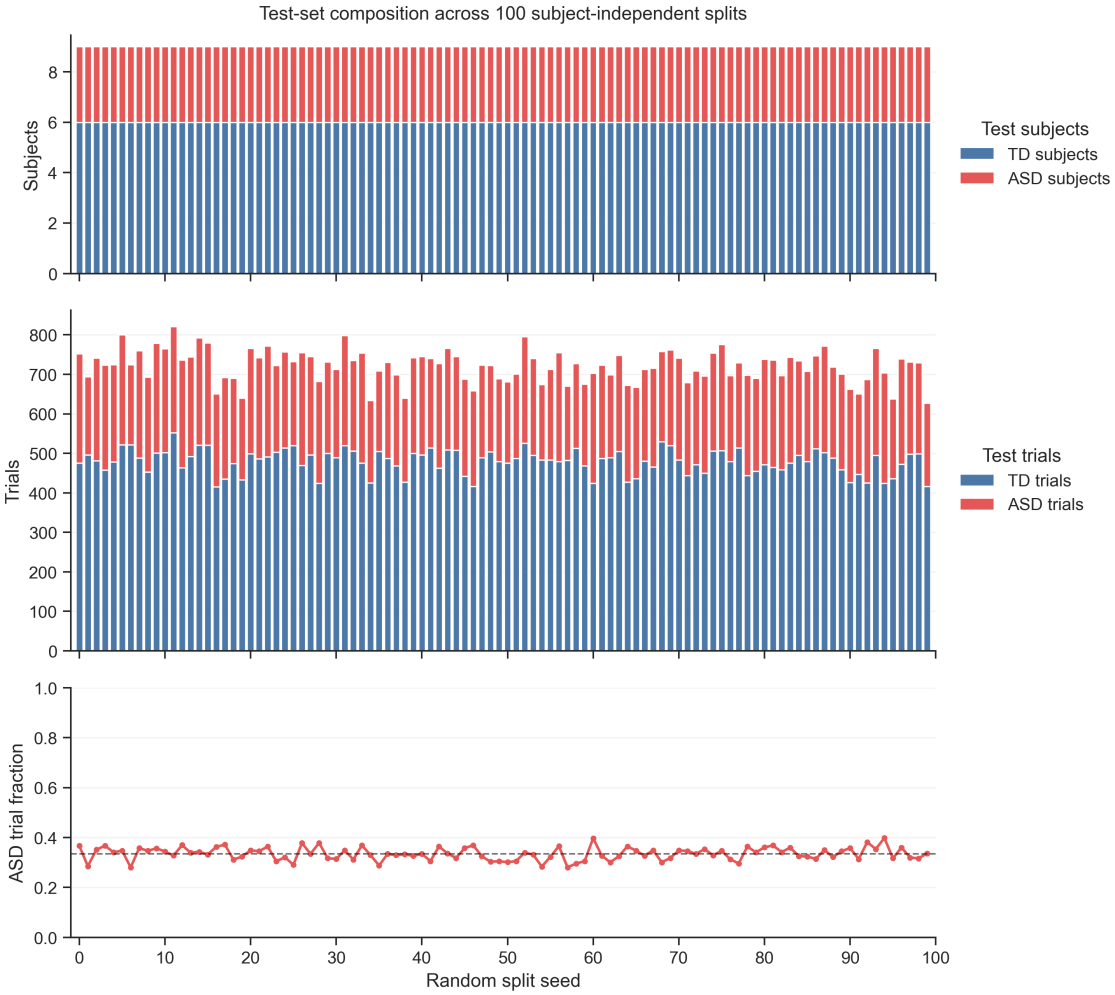

Figure 4.6: Test set composition across 100 random subject-independent splits. Top: number of ASD and TD subjects per split. Middle: number of ASD and TD trials per split. Bottom: ASD trial fraction per split. The dashed line indicates the overall mean.

Table 4.4: Trial-level classification performance for all six model configurations. Difference input refers to the subject-baseline difference representation, where each trial is expressed relative to the participant's own average EEG response.

Model	Input	Acc.	AUC	F1	Sens.	Spec.
SVM	Raw	0.64 ± 0.09	0.61 ± 0.16	0.43 ± 0.16	0.40 ± 0.15	0.76 ± 0.07
SVM	Difference	0.65 ± 0.10	0.61 ± 0.19	0.42 ± 0.18	0.39 ± 0.16	0.78 ± 0.09
Simple CNN	Raw	0.72 ± 0.25	0.77 ± 0.25	0.70 ± 0.16	0.83 ± 0.12	0.66 ± 0.38
Simple CNN	Difference	0.72 ± 0.17	0.80 ± 0.20	0.61 ± 0.22	0.68 ± 0.29	0.76 ± 0.23
CNN-BiLSTM	Raw	0.71 ± 0.12	0.72 ± 0.18	0.56 ± 0.18	0.59 ± 0.25	0.78 ± 0.12
CNN-BiLSTM	Difference	0.75 ± 0.14	0.80 ± 0.07	0.68 ± 0.10	0.75 ± 0.07	0.75 ± 0.22

The SVM performed close to chance level with both inputs (AUC = 0.61), suggesting that flattened EEG with PCA reduction does not capture enough useful information for ASD classification. Both deep learning models performed notably better. The CNN-BiLSTM with difference input reached the highest trial-level AUC (0.80) and had the smallest variation across partitions (SD = 0.07), while the Simple CNN reached the same AUC but with much larger variation (SD = 0.20). Using the difference input also improved CNN-BiLSTM performance compared to raw input (AUC: 0.80 vs 0.72), suggesting that encoding each trial as a deviation from the participant's own average is more useful than using the raw signal directly.
That said, trial-level AUC values were only moderate across all models. Single EEG trials are noisy, and classifying individual epochs in a small paediatric sample is inherently difficult. This motivated the subject-level aggregation analysis below.

4.3.3 Subject-Level Classification via Trial Voting

Table 4.5 and Figure 4.7 show the results when trial-level predicted probabilities were averaged within each participant and thresholded at 0.5 to produce a single subject-level prediction.

Table 4.5: Subject-level classification performance obtained by averaging trial-level predicted probabilities within each participant. All models use the difference input.

Model	Accuracy	AUC	F1	Sensitivity	Specificity
SVM	0.76 ± 0.09	0.67 ± 0.36	0.36 ± 0.35	0.27 ± 0.28	1.00 ± 0.00
Simple CNN	0.78 ± 0.22	0.86 ± 0.23	0.60 ± 0.43	0.67 ± 0.47	0.83 ± 0.20
CNN-BiLSTM	0.78 ± 0.14	0.86 ± 0.15	0.72 ± 0.09	0.80 ± 0.18	0.77 ± 0.28

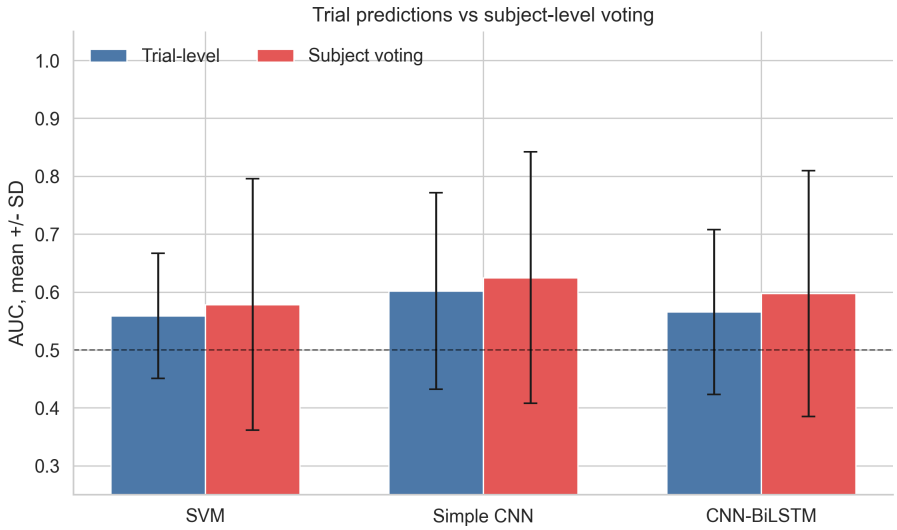

Figure 4.7: Mean AUC (± SD) at trial level and after subject-level voting, for all three models with difference input. The dashed line marks chance level (AUC = 0.5).

Moving to subject-level predictions improved results for both deep learning models. The CNN-BiLSTM AUC increased from 0.80 to 0.86, F1 from 0.68 to 0.72, and Sensitivity from 0.75 to 0.80. The Simple CNN showed a similar AUC gain but with much higher variance in F1 (SD = 0.43 vs 0.09 for CNN-BiLSTM), meaning its performance was less consistent across different data splits.

The SVM behaved differently at the subject level. Its Specificity reached 1.00 with no variance, while Sensitivity fell to 0.27. This means the SVM predicted almost every participant as TD, regardless of their true group. This happened because the SVM already had low Sensitivity at the trial level (0.39), and averaging those predictions across trials pushed the subject-level score below 0.5 for nearly all participants. The SVM therefore did not learn meaningful ASD-related patterns and is included only as a baseline comparison.

Based on these results, the CNN-BiLSTM with difference input was chosen as the final model. Its subject-level AUC of 0.86 and Sensitivity of 0.80 mean it correctly identified around four out of five ASD children on average, with the most stable performance across partitions among all tested configurations.

4.3.4 Exploratory SHAP Channel Attribution

An exploratory SHAP analysis was performed to inspect which EEG channels the CNN-BiLSTM relied on most for classification. SHAP values were computed for five randomly selected partitions and averaged across partitions and test trials. Figure 4.8 shows the mean absolute SHAP attribution per channel, with ERP ROI channels (F3, F4, C3, C4, P3, P4) highlighted in red.

P4 showed the highest mean attribution, followed by F3, Fp2, T3, and C3. Among the top five channels, three (P4, F3, C3) belong to the ERP regions of interest defined in Section 3.1.3, while Fp2 and T3 are outside the ROI. Overall, four of the six ROI channels (P4, F3, C3, F4) ranked among the top eight most attributed channels, suggesting that the model drew information from regions associated with auditory and cognitive processing. The remaining ROI channels (P3 and C4) ranked lower, and non-ROI channels such as Fp2 and T3 also contributed notably. The large error bars on several channels reflect variability across the five partitions, which is expected given the small test set size. This result is therefore interpreted as an exploratory check rather than a definitive characterisation of the model's decision process.

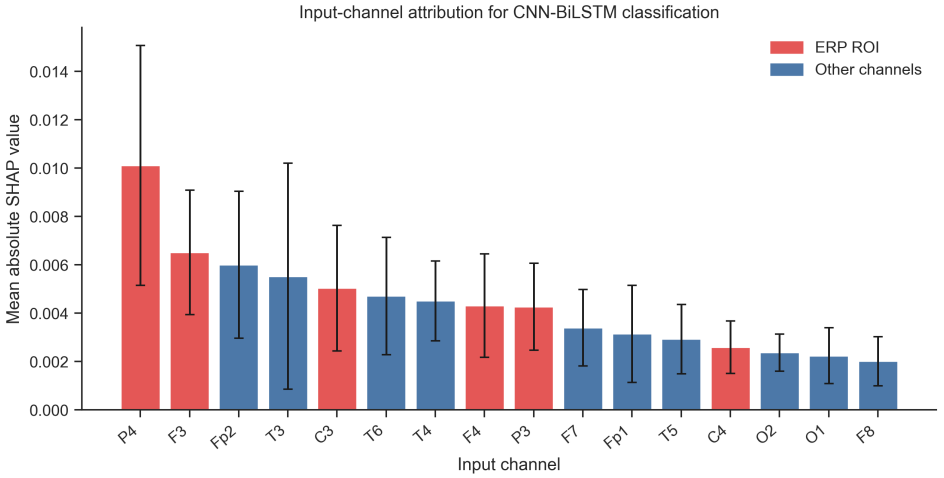

Figure 4.8: Mean absolute SHAP values per input channel, averaged across five randomly selected partitions and all test trials. Red bars indicate ERP ROI channels (F3, F4, C3, C4, P3, P4); blue bars indicate non-ROI channels. Error bars show standard deviation across the five partitions.

4.4 Personalized Recommendations

4.4.1 Recommendation Heterogeneity

Figure 4.9 shows the distribution of recommended conditions across the 22 ASD participants for each acoustic dimension. No single condition received a recommendation from more than 65% of participants in any dimension. The Ear dimension exhibited the most even distribution (Both: 36%, Left: 32%, Right: 32%), while Noise showed the strongest majority (Yes: 65%, No: 35%). Speed, Distance, and Pitch all had at least three distinct conditions represented in the recommendations. These results confirm that the optimal acoustic configuration varies substantially across ASD children, providing empirical justification for individualised rather than group-level recommendations.

4.4.2 Individual NPES Profiles

Figure 4.10 presents the NPES profiles of three representative ASD participants to illustrate the diversity of individual acoustic processing patterns. ASD Subject 1 showed highest efficiency for slow speech, no background noise, and a 1 m speaker distance, whereas ASD Subject 2 processed slow speech and high pitch most efficiently but performed best with background noise present. ASD Subject 3 exhibited yet a different pattern, favouring fast speech, close speaker distance, and left-ear presentation. These contrasting profiles demonstrate that the optimal acoustic environment cannot be inferred from group-level averages and must be determined individually.

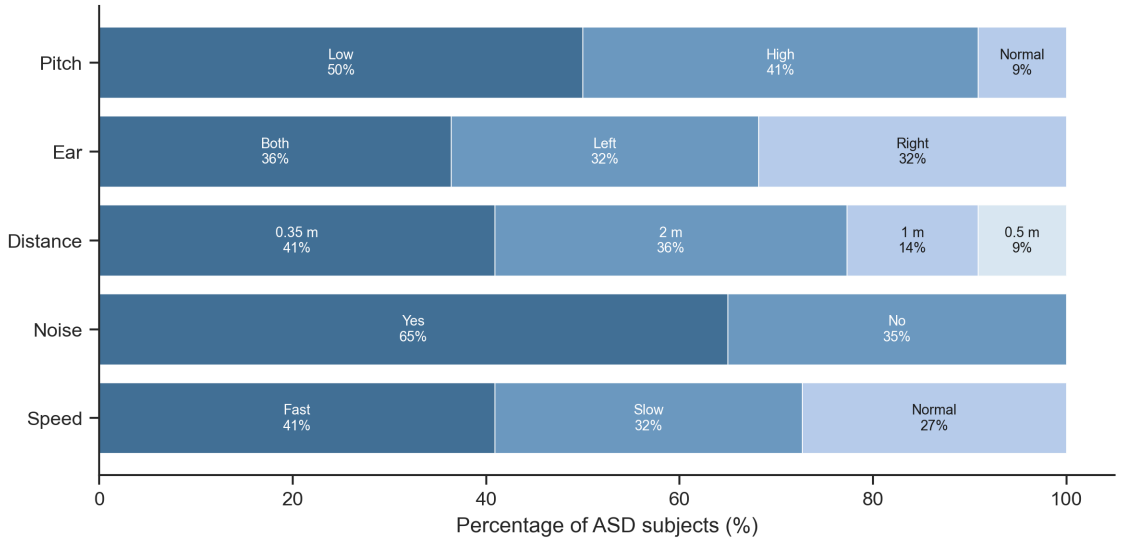

Figure 4.9: Distribution of NPES-based recommendations across 22 ASD participants for each acoustic dimension. Each colour represents a different recommended condition. No single condition dominates any dimension, confirming the need for individualised assessment.

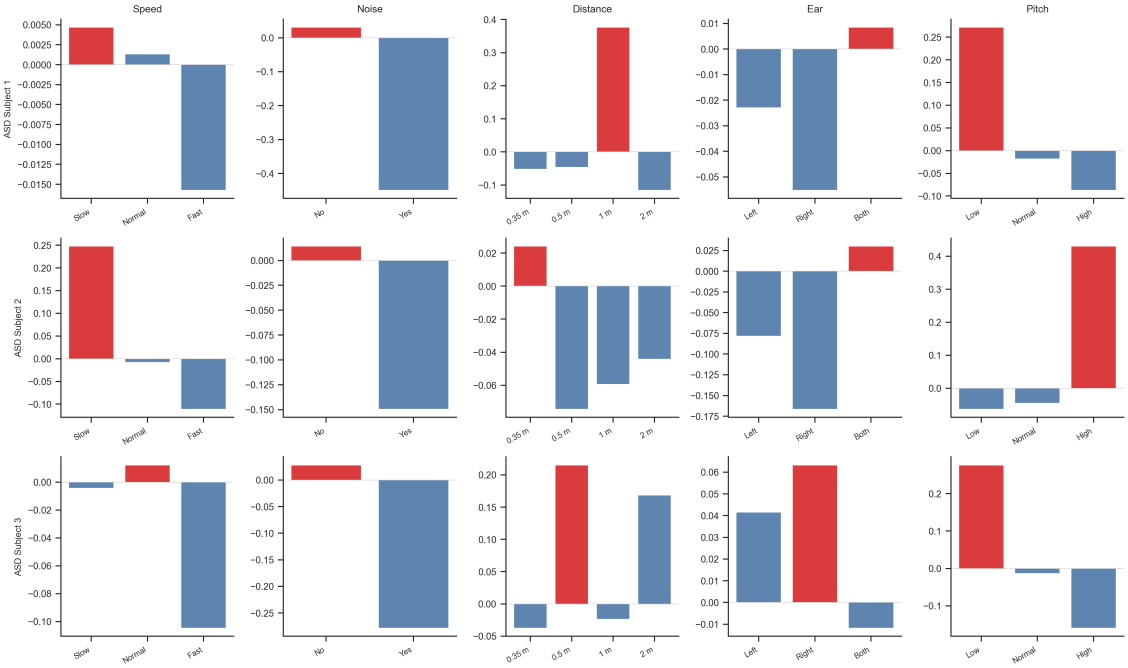

Figure 4.10: NPES profiles for three representative ASD participants across the five acoustic dimensions. Red bars indicate the recommended condition (highest NPES) for each dimension. The zero line represents the participant's own mean efficiency across all conditions.

4.4.3 Biomarker Contribution

Figure 4.11 shows the agreement rate between each single-biomarker recommendation and the full NPES recommendation for each acoustic dimension. GFP exhibited the highest single-biomarker agreement in the Noise dimension (80%) and Pitch dimension (68%), while N1 latency showed the strongest contribution in the Noise dimension (75%) and Speed dimension (59%). P2 and P300 latencies contributed most to the Ear dimension (both 64%). No single biomarker consistently dominated across all five dimensions, with agreement rates ranging from 32% to 80%. This result validates the composite NPES design: each biomarker captures a distinct aspect of neural processing efficiency, and the combination provides a more comprehensive characterisation than any individual component.

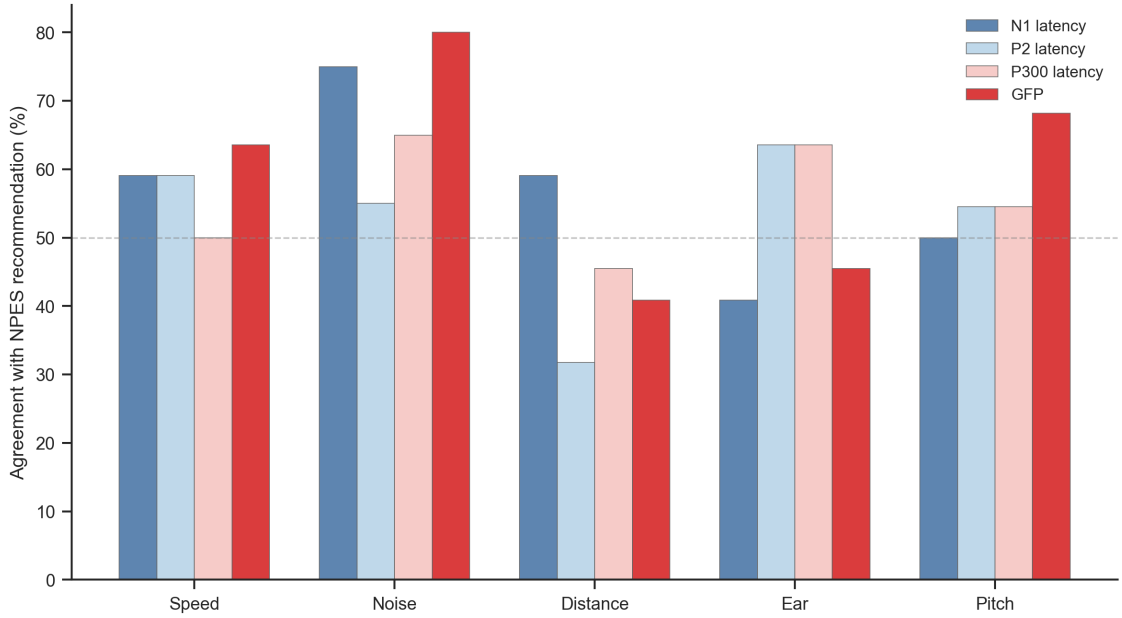

Figure 4.11: Agreement between single-biomarker recommendations and the full NPES recommendation for each acoustic dimension. The dashed line indicates the 50% reference level. No single biomarker dominates all dimensions, supporting the composite score design.

4.4.4 Recommendation Stability

Table 4.6 presents the split-half agreement rates for each acoustic dimension alongside the corresponding chance baselines. Speed and Pitch showed agreement rates above their respective chance levels, indicating that the NPES-based recommendations for these dimensions contain signal beyond random assignment. Noise, with only two conditions, showed agreement close to its 50% chance baseline. Distance and Ear exhibited the lowest agreement rates, consistent with these dimensions having the smallest number of trials per minority condition (Distance 2 m: mean 5.0 trials per subject; Ear Left: mean 6.5 trials per subject). When each split-half contains only 2–3 trials per condition, the NPES estimate is dominated by single-trial

noise.

Table 4.6: Split-half recommendation agreement per acoustic dimension (50 random repetitions per subject). Chance baseline is defined as $1/k$ where k is the number of conditions.

Dimension	Conditions (k)	Chance (%)	Agreement (%)
Speed	3	33	45
Noise	2	50	48
Distance	4	25	24
Ear	3	33	29
Pitch	3	33	40

These results indicate that recommendations derived from a full assessment session are more reliable than those from abbreviated protocols. The limited stability observed for Distance and Ear reflects a data limitation (insufficient trials per minority condition) rather than a flaw in the NPES framework itself, and would be expected to improve with experimental designs that allocate more trials to each condition.

5

Discussion

This chapter discusses the findings in the context of prior work, examines their implications for educational practice, identifies the limitations of the present study, and outlines directions for future research.

5.1 Summary of Findings

This thesis developed a two-stage framework for personalized acoustic assessment of children with ASD, combining deep learning-based screening with biomarker-driven recommendation.

The first stage addressed automated ASD identification. A CNN-BiLSTM classifier trained on single-trial raw EEG achieved moderate trial-level classification performance, outperforming both the SVM baseline and the Simple CNN across multiple data partitions. Aggregating trial-level predictions to the subject level via probability averaging improved classification accuracy, demonstrating that the model captures ASD-related neural signatures that become more reliable when averaged over multiple observations.

The second stage addressed the characterisation of individual auditory processing profiles. A preliminary group-level analysis of ERP biomarkers revealed no statistically significant differences between ASD and TD children after multiple comparison correction, despite the use of six ERP components across five acoustic dimensions. This finding, rather than representing a negative outcome, provided the empirical motivation for shifting from between-group statistics to within-subject profiling. The Neural Processing Efficiency Score was constructed from four biomarkers (N1 latency, P2 latency, P300 latency, and GFP) using within-subject z-scoring and equal-weight aggregation.

The NPES-based recommendation system produced heterogeneous acoustic configurations across the 22 ASD participants, with no single condition dominating any of the five acoustic dimensions. The biomarker contribution analysis confirmed that no single component could replicate the full NPES recommendation across all dimensions, supporting the value of the composite design. Split-half stability analysis indicated that recommendation reliability is constrained by the number of available trials per condition, with dimensions containing fewer minority-condition trials showing lower agreement rates.

5.2 Implications for Educational Practice

The proposed framework is designed to function as a single-session assessment tool. Following a standardised EEG recording session of approximately 30 minutes, the system outputs an ASD screening decision and a five-dimensional acoustic configuration tailored to the individual child.

The acoustic recommendations have direct relevance to several aspects of educational practice. A recommendation for reduced speech rate suggests that the child's temporal integration processes benefit from slower phonemic transitions, which can inform how teachers pace verbal instructions. A recommendation for reduced background noise indicates that the child's auditory system is less able to segregate speech from competing sounds, supporting the use of sound-absorbing materials in the classroom. Recommendations regarding speaker distance and pitch can guide seating arrangements and vocal delivery adjustments.

This approach extends the environmental structuring principles of established frameworks such as TEACCH [31], which emphasise visual and spatial predictability for ASD learners, to the acoustic domain. While TEACCH structures the physical environment based on clinical observation, the present system structures the auditory environment based on objective neurophysiological measurement. The two approaches are complementary: TEACCH can inform the visual and spatial layout of the learning environment, while the NPES-based assessment can inform its acoustic properties.

It is important to note that the current system provides static assessment rather than real-time adaptation in physical classrooms; the recommendations are intended to inform environment design decisions made by educators, not to control live acoustic parameters dynamically. However, extending beyond traditional physical settings, this framework represents a potential application direction for seamless integration within AI-driven, voice-controlled online learning systems. In such digital environments, the NPES-generated acoustic configuration can serve as a direct initialization profile. For instance, an interactive virtual tutor could automatically adjust its Text-to-Speech engine to match the optimal speech rate and pitch identified for a specific ASD child. Furthermore, the system could adaptively calibrate its response latency to accommodate the child's auditory processing speed and implement active software-level noise cancellation based on their sensory gating profile. This transition from a static assessment report to a closed-loop, dynamically adaptive digital learning environment represents a highly promising frontier for inclusive neuro-educational technologies.

5.3 Limitations

A general limitation of this thesis is that it was based on a secondary dataset. Therefore, the experimental protocol, stimulus distribution, participant recruitment, and behavioural assessment design could not be modified retrospectively. The limitations discussed below should be understood within this context that the present work aimed to develop and evaluate a methodological framework using the available

data, rather than to design a new data collection protocol from the beginning.

Dataset Size and Data Quality The first limitation is the relatively small sample size. The final dataset included 59 participants, with 22 ASD children and 37 TD children. This is small for machine learning research and limits the stability of performance estimates.
The original dataset included 102 participants. However, technical limitations in the synchronised communication mechanism during data acquisition reduced the number of samples with complete and accurate stimulus markers. Further preprocessing and data quality checks reduced the final sample size. The reduction was more severe in the ASD group than in the TD group.
This data loss reflects the practical difficulty of collecting EEG data from children with ASD. Children may show lower task compliance, more movement, and higher sensitivity to the experimental environment. These factors can reduce signal quality and increase the exclusion rate. Although subject-independent evaluation was used to reduce overfitting, the small number of test subjects in each split still leads to high variability across random partitions.

Experimental Paradigm The study used a passive auditory paradigm with a fixed stimulus sequence. A major limitation of this design is the absence of subjective behavioural assessments. Without behavioural scores, it is not possible to directly validate whether the neural responses correspond to the child's actual auditory experience or learning performance.
As a result, the NPES should be interpreted as a neurophysiological estimate of processing efficiency, rather than a fully validated measure of educational benefit. The framework identifies the condition under which neural processing appears more efficient, but it cannot yet confirm whether this condition improves comprehension, comfort, or learning outcomes.
Another limitation is the imbalance of stimulus conditions. Most trials belonged to the baseline condition, while minority conditions contained only 5 to 8 trials per subject. This limited the reliability of condition-specific NPES estimates. It also affected the split-half stability analysis, especially for dimensions with few available trials.

EEG Montage and Biomarker Selection The EEG data were recorded using a 16-channel montage. This provides limited spatial resolution compared with high-density EEG systems. The common average reference computed from a low-density montage may also introduce polarity distortions, especially for amplitude-related measures.
For this reason, amplitude biomarkers were excluded from the final NPES design. The NPES instead used latency-based measures and GFP. A higher-density montage would allow more reliable artifact rejection, better source localisation, and more stable amplitude-based biomarkers.

Generalisability The dataset was collected from Romanian-speaking children using Romanian-language stimuli. Neural responses to speech rate, pitch, and syn-

tactic structure may vary across languages. Therefore, the findings should not be assumed to generalise directly to other linguistic or cultural contexts.
This limitation is especially relevant for the speed and pitch dimensions, because these acoustic features are closely related to the phonological and prosodic properties of each language. Cross-linguistic validation is needed before the framework can be applied more broadly.

Validation of Recommendations The NPES-based recommendations have not yet been validated against behavioural or educational outcomes. The system can identify the condition under which a child's neural response appears most efficient, but it remains unclear whether this leads to better attention, reduced distress, improved comprehension, or stronger learning performance in real educational settings. The deep learning component is also limited to the screening stage. During the development of this thesis, several deep learning approaches for the recommendation stage were explored, including the use of classifier probability as a GFP proxy and trial-level optimality prediction. These approaches did not produce sufficiently reliable results on the present dataset. Therefore, the final recommendation stage used the biomarker-based NPES framework instead.

5.4 Societal and Ethical Considerations

Societal relevance. This work addresses the intersection of two United Nations Sustainable Development Goals: SDG 4 (Quality Education) and SDG 3 (Good Health and Well-being) [46]. Current educational and diagnostic approaches for ASD rely predominantly on subjective behavioural observation, which may inadvertently disadvantage non-verbal or minimally verbal children whose internal processing difficulties are not reflected in observable behaviour. The proposed framework contributes toward more equitable assessment by providing an objective, neurophysiologically grounded tool that characterises each child's auditory processing profile independently of their verbal or behavioural presentation. If validated at scale, such a tool could reduce the dependence on specialist clinical judgement for acoustic environment design, potentially improving access to individualised support in resource-limited educational settings.

Ethical framework and data governance. This study utilised a secondary EEG dataset collected from children, including minors with a clinical diagnosis of ASD, which constitutes sensitive personal health data under the General Data Protection Regulation (GDPR) [47]. The original data collection was approved by the Research Ethics Committee of Ştefan cel Mare University of Suceava, Romania, and written informed consent was obtained from the parents or legal guardians of all participants in accordance with the Declaration of Helsinki [48]. Secondary use of the data for the present thesis was approved by the Ethics Committee at Chalmers University of Technology. All data were stored and processed on local machines in compliance with institutional data handling policies. No personally identifiable information was used in the analysis or reported in this thesis; participant identifiers

were anonymised prior to all processing steps.

Responsible use of AI in educational contexts. The application of deep learning to clinical and educational decision-making raises concerns about transparency and accountability. To address this, the classification model developed in this thesis was evaluated using subject-independent protocols that provide realistic estimates of generalisation performance, and SHAP-based explainability analysis was applied to identify which EEG features contribute to the model's decisions. The recommendation component of the framework is based on interpretable ERP biomarkers rather than opaque model outputs, ensuring that the reasoning behind each acoustic recommendation can be understood and verified by clinicians and educators. It should be noted that the proposed system is intended as a decision-support tool to inform professional judgement, not as a standalone diagnostic or prescriptive device.

5.5 Future Work

The limitations identified in this thesis suggest several directions for future research. Since the present work was based on a secondary dataset, future work should focus on validating and extending the proposed framework using datasets that are specifically designed for personalized acoustic assessment.

Framework Validation on Purpose-Designed Datasets Future validation should be conducted on larger and more balanced datasets designed for personalized acoustic assessment. Such datasets should include more balanced ASD and TD groups, more evenly distributed trials across acoustic conditions, and behavioural measures of auditory comfort, comprehension, attention, and learning outcomes.
The purpose of such validation would not only be to improve classification performance or recommendation stability, but also to test whether the NPES-based recommendations correspond to meaningful behavioural and educational benefits. This would provide stronger external validation than is possible with the current offline secondary dataset.

Analysis of Condition-Change Responses Another future direction is to analyse neural responses to changes between acoustic conditions. In the present dataset, the normal condition was the dominant condition, while the other acoustic conditions occurred less frequently. Instead of treating each condition only as an independent category, future work could examine whether transitions from the normal condition to a modified acoustic condition elicit mismatch-like responses.
Such responses could be interpreted as MMN-like responses, although the current paradigm was not specifically designed as a classical oddball MMN experiment. This analysis could provide additional information about how ASD and TD children respond to acoustic changes over continuous trials. For example, changes in speech rate, background noise, speaker distance, or pitch may produce transient neural responses that are not fully captured by condition-wise ERP averaging.

This direction is particularly relevant because the current dataset did not show robust group-level ERP differences after multiple comparison correction. A transition-based analysis may therefore offer a complementary way to investigate whether ASD and TD children differ in their neural sensitivity to acoustic changes, rather than only in their average responses to static acoustic conditions.

Adaptive Experimental Paradigms Future studies could also investigate adaptive experimental paradigms. In the present dataset, the stimulus sequence was fixed and the number of trials was uneven across acoustic conditions. An adaptive paradigm could select stimulus conditions dynamically based on the child's neural responses.
This would have two main advantages. First, it could help balance the number of trials across acoustic conditions, improving the reliability of condition-specific NPES estimates. Second, it could allow the system to progressively identify the most suitable acoustic configuration during the assessment session. Methods such as Bayesian optimisation or multi-armed bandit algorithms could be used to explore the acoustic parameter space efficiently.

Behavioural and Educational Validation A key next step is to validate whether NPES-based recommendations lead to measurable benefits in real educational or therapeutic settings. The current system identifies the acoustic condition under which a child's neural response appears most efficient, but it does not yet show whether this condition improves real-world outcomes.
Future intervention studies could compare children receiving acoustically personalized instruction with children receiving standard instruction. Relevant outcome measures may include speech comprehension, attention, task engagement, distress, and learning performance. Such studies would provide the behavioural evidence needed to evaluate the practical value of the proposed framework.

Improved EEG Recording and Biomarker Design Future studies could benefit from higher-density EEG recordings, such as 64-channel systems. Higher-density EEG would improve spatial resolution, support more reliable ICA-based artifact rejection, and enable more stable source-level analysis.
Improved EEG recording quality may also allow amplitude-based ERP biomarkers to be reconsidered. In the present thesis, amplitude measures were excluded from the final NPES design due to concerns related to low-density common average referencing. With higher-quality recordings, future versions of the NPES could include a broader set of biomarkers and provide a more complete description of auditory processing efficiency.

Deep Learning for Recommendation With larger and more balanced datasets, it may become possible to incorporate deep learning into the recommendation stage. A future model could learn the mapping from raw EEG features to optimal acoustic conditions directly, rather than relying only on explicit ERP biomarker extraction. In this setting, the NPES framework developed in this thesis could serve as an initial labelling strategy. Biomarker-based recommendations could provide supervised

targets for a neural network trained on raw EEG input. This would connect the screening and recommendation stages more closely and may support faster personalized assessment in the future.

6
Conclusion

This thesis presented a framework for objective and personalized acoustic assessment of children with ASD. By combining EEG-based classification with ERP-based recommendation, the work explored how neural responses can be used not only for screening, but also for characterising individual auditory processing profiles.
The results show that subject-level aggregation of single-trial EEG predictions can support automated ASD screening. They also show that personalized acoustic recommendations can be derived from within-subject ERP profiles, even when stable group-level ERP differences are not observed. This shift from group-level comparison to individual profiling is a central contribution of the thesis.
The proposed NPES framework revealed substantial variability among ASD participants, supporting the need for individualised acoustic assessment rather than a universal acoustic configuration. Although the current system requires further validation with behavioural outcomes and purpose-designed datasets, it provides an initial methodological basis for neurophysiologically grounded acoustic personalisation.
Overall, this thesis demonstrates the feasibility of using EEG and ERP biomarkers to support personalized acoustic assessment for children with ASD. The framework provides a foundation for future adaptive systems that can better support learning and therapeutic environments.

Bibliography

[1] C. Lord, M. Elsabbagh, G. Baird, and J. Veenstra-Vanderweele, "Autism spectrum disorder," *The Lancet*, vol. 392, no. 10146, pp. 508–520, 2018.

[2] S. Sharghilavan, O. Geman, and R. Toderean, "Speech perception and speech attention: a case study based on event-related potential using the openbci system," in *2024 E-Health and Bioengineering Conference (EHB)*, pp. 1–4, IEEE, 2024.

[3] J. Shan, Y. Gu, J. Zhang, X. Hu, H. Wu, T. Yuan, and D. Zhao, "A scoping review of physiological biomarkers in autism," *Frontiers in Neuroscience*, vol. 17, p. 1269880, 2023.

[4] W. J. Bosl, H. Tager-Flusberg, and C. A. Nelson, "Eeg analytics for early detection of autism spectrum disorder: a data-driven approach," *Scientific Reports*, vol. 8, no. 1, p. 6828, 2018.

[5] S. J. Webb, A. J. Naples, A. R. Levin, *et al.*, "The autism biomarkers consortium for clinical trials: initial evaluation of a battery of candidate eeg biomarkers," *American Journal of Psychiatry*, vol. 180, no. 1, pp. 41–49, 2023.

[6] E. Gkintoni, M. Panagioti, S. P. Vassilopoulos, G. Nikolaou, B. Boutsinas, and A. Vantarakis, "Leveraging ai-driven neuroimaging biomarkers for early detection and social function prediction in autism spectrum disorders: a systematic review," *Healthcare*, vol. 13, no. 2, p. 177, 2025.

[7] S. Wang, C. Yang, Y. Liu, Z. Shao, and T. Jackson, "Early and late stage processing abnormalities in autism spectrum disorders: An ERP study," *PLoS ONE*, vol. 12, no. 5, p. e0178542, 2017.

[8] S. Sharghilavan, L. Mehdizadeh Fanid, O. Geman, H. Shahrokhi, and H. Seyedarabi, "N100 as a neural marker of atypical early auditory encoding in autism: sensitivity to pitch, distance-based intensity, and spatial location," *bioRxiv*, pp. 2025–08, 2025.

[9] M. Latinus, Y. Mofid, K. Kovarski, J. Charpentier, M. Batty, and F. Bonnet-Brilhault, "Atypical sound perception in ASD explained by inter-trial (in)consistency in EEG," *Frontiers in Psychology*, vol. 10, p. 1177, 2019.

[10] S. Sharghilavan, L. Mehdizadeh Fanid, O. Geman, H. Shahrokhi, and H. Seyedarabi, "Exploring hemispheric contributions in the processing of social speech in autism," *Journal of Pediatric Perspectives*, vol. 13, no. 9, pp. 19650–19660, 2025.

[11] S. Sharghilavan, L. Mehdizadeh Fanid, O. Geman, H. Shahrokhi, and H. Seyedarabi, "Too close to focus? neural evidence of altered auditory spatial attention in autism," *Journal of Pediatric Perspectives*, vol. 13, no. 8, pp. 19618–19627, 2025.

[12] K. Fadeev, I. Romero Reyes, *et al.*, "Attenuated processing of vowels in the left temporal cortex predicts speech-in-noise perception deficit in children with autism," *Journal of Neurodevelopmental Disorders*, vol. 16, no. 1, p. 67, 2024.

[13] J. Li, M. Sujawal, Z. Bernotaite, I. Cunnings, and F. Liu, "Auditory and semantic processing of speech-in-noise in autism: a behavioral and EEG study," *Autism Research*, vol. 18, no. 10, pp. 2011–2030, 2025.

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

[15] S. S. Jeste and C. A. Nelson III, "Event related potentials in the understanding of autism spectrum disorders: an analytical review," *Journal of autism and developmental disorders*, vol. 39, no. 3, pp. 495–510, 2009.

[16] T. W. Picton, "The p300 wave of the human event-related potential," *Journal of clinical neurophysiology*, vol. 9, no. 4, pp. 456–479, 1992.

[17] T. Cui, P. P. Wang, S. Liu, and X. Zhang, "P300 amplitude and latency in autism spectrum disorder: a meta-analysis," *European child & adolescent psychiatry*, vol. 26, no. 2, pp. 177–190, 2017.

[18] M. Kutas and K. D. Federmeier, "Thirty years and counting: finding meaning in the n400 component of the event-related brain potential (erp)," *Annual review of psychology*, vol. 62, no. 1, pp. 621–647, 2011.

[19] L. Osterhout and P. J. Holcomb, "Event-related brain potentials elicited by syntactic anomaly," *Journal of memory and language*, vol. 31, no. 6, pp. 785–806, 1992.

[20] A. D. Friederici, "Towards a neural basis of auditory sentence processing," *Trends in cognitive sciences*, vol. 6, no. 2, pp. 78–84, 2002.

[21] D. Lehmann and W. Skrandies, "Reference-free identification of components of checkerboard-evoked multichannel potential fields," *Electroencephalography and clinical neurophysiology*, vol. 48, no. 6, pp. 609–621, 1980.

[22] M. M. Murray, D. Brunet, and C. M. Michel, "Topographic erp analyses: a step-by-step tutorial review," *Brain Topography*, vol. 20, no. 4, pp. 249–264, 2008.

[23] H. L. Hamburger and M. A. vd Burgt, "Global field power measurement versus classical method in the determination of the latency of evoked potential components," *Brain topography*, vol. 3, no. 3, pp. 391–396, 1991.

[24] H. Ardakani, M. Taghizadeh, and F. Shayegh, "Diagnosis of autism disorder based on deep network trained by augmented EEG signals," *International Journal of Neural Systems*, vol. 32, no. 11, p. 2250046, 2022.

[25] S. Sharghilavan, O. Geman, *et al.*, "Explainable ai in eeg waves based classification for early identification in autism," in *2025 IEEE Medical Measurements Applications (MeMeA)*, 2025.

[26] H. Hatim, Z. Alyasseri, and N. Jamil, "A recent advances on autism spectrum disorders in diagnosing based on machine learning and deep learning," *Artificial Intelligence Review*, vol. 58, no. 10, p. 313, 2025.

[27] F. C. Peck *et al.*, "Prediction of autism spectrum disorder diagnosis using nonlinear measures of language-related eeg at 6 and 12 months," *Journal of Neurodevelopmental Disorders*, vol. 13, no. 1, p. 57, 2021.

[28] M. N. A. Tawhid *et al.*, "A spectrogram image based intelligent technique for automatic detection of autism spectrum disorder from eeg," *PLOS ONE*, vol. 16, no. 6, p. e0253094, 2021.

[29] Y. Xu *et al.*, "Autism spectrum disorder diagnosis with eeg signals using time series maps of brain functional connectivity and a combined cnn-lstm model," *Computer Methods and Programs in Biomedicine*, vol. 250, p. 108196, 2024.

[30] S. M. Lundberg and S. I. Lee, "A unified approach to interpreting model predictions," in *Advances in Neural Information Processing Systems (NeurIPS)*, vol. 30, pp. 4765–4774, 2017.

[31] G. B. Mesibov, V. Shea, E. Schopler, L. W. Adams, E. Merkler, S. Burgess, M. Mosconi, S. M. Chapman, C. Tanner, and M. E. Bourgondien, *The TEACCH approach to autism spectrum disorders*. Springer, 2004.

[32] P. Sanz-Cervera *et al.*, "The effectiveness of TEACCH intervention in autism spectrum disorder: a review study," *Papeles Del Psicólogo*, vol. 39, no. 1, pp. 40–50, 2018.

[33] J. Virues-Ortega, F. M. Julio, and R. Pastor-Barriuso, "The teacch program for children and adults with autism: A meta-analysis of intervention studies," *Clinical psychology review*, vol. 33, no. 8, pp. 940–953, 2013.

[34] A. Gramfort, M. Luessi, E. Larson, *et al.*, "Meg and eeg data analysis with mne-python," *Frontiers in Neuroscience*, vol. 7, p. 267, 2013.

[35] T. W. Picton, S. Bentin, P. Berg, E. Donchin, S. A. Hillyard, R. Johnson, G. A. Miller, W. Ritter, D. S. Ruchkin, M. D. Rugg, *et al.*, "Guidelines for using human event-related potentials to study cognition: recording standards and publication criteria," *Psychophysiology*, vol. 37, no. 2, pp. 127–152, 2000.

[36] R. Vallat, "Pingouin: statistics in python.," *J. Open Source Softw.*, vol. 3, no. 31, p. 1026, 2018.

[37] P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, *et al.*, "Scipy 1.0: fundamental algorithms for scientific computing in python," *Nature methods*, vol. 17, no. 3, pp. 261–272, 2020.

[38] Y. Benjamini and Y. Hochberg, "Controlling the false discovery rate: a practical and powerful approach to multiple testing," *Journal of the Royal Statistical Society: Series B*, vol. 57, no. 1, pp. 289–300, 1995.

[39] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, *et al.*, "Scikit-learn: Machine learning in python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.

[40] I. T. Jolliffe, *Principal Component Analysis*. Springer, 2 ed., 2002.

[41] I. Loshchilov and F. Hutter, "Decoupled weight decay regularization," in *International Conference on Learning Representations (ICLR)*, 2019.

[42] F. Leroy, S. Naik, R. Gulbinaite, M. Palu, D. Battaglia, and G. Dehaene-Lambertz, "Dynamics of the attentional blink in preverbal infants," *Proceedings of the National Academy of Sciences*, vol. 123, no. 10, p. e2526752123, 2026.

[43] F. Si, S. Miao, Y. Yang, S. Li, G. Liu, W. Cao, R. Zhang, Z. Yang, H. Yang, K. Wang, *et al.*, "The effects of polarization characteristics on visual fatigue: an empirical study based on subjective and objective indicators," *Frontiers in Neuroscience*, vol. 19, p. 1715236, 2025.

[44] B. Dunst, M. Benedek, E. Jauk, S. Bergner, K. Koschutnig, M. Sommer, A. Ischebeck, B. Spinath, M. Arendasy, M. Bühner, *et al.*, "Neural efficiency as a function of task demands," *Intelligence*, vol. 42, pp. 22–30, 2014.

[45] M. J. Wenger, J. T. Townsend, and S. F. Newbolds, "The neural efficiency score: Validation and application," *bioRxiv*, pp. 2024–07, 2024.

[46] United Nations, "Transforming our world: The 2030 agenda for sustainable development." Department of Economic and Social Affairs, 2015.

[47] European Parliament and Council of the European Union, "Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation)." Official Journal of the European Union, L 119, 4 May 2016, pp. 1–88, 2016.

[48] W. M. Association *et al.*, "World medical association declaration of helsinki: ethical principles for medical research involving human subjects," *Jama*, vol. 310, no. 20, pp. 2191–2194, 2013.

A

Appendix 1

Table A.1: Exhaustive list of all 98 stimuli and their acoustic manipulations.

Order	Trigger ID	Ear	Distance	Pitch	Speed	Noise	Sentence Type
1	1	Left	0.35 m	Normal	Normal	No	Interrogative
2	2	Both	1 m	Normal	Normal	No	Interrogative
3	3	Right	0.35 m	Normal	Normal	No	Interrogative
4	4	Both	0.35 m	Normal	Normal	No	Imperative
5	5	Both	0.5 m	Normal	Normal	No	Imperative
6	6	Both	1 m	Normal	Normal	No	Imperative
7	7	Left	0.35 m	Normal	Normal	No	Imperative
8	8	Both	2 m	Normal	Normal	No	Imperative
9	9	Right	0.35 m	Normal	Normal	No	Imperative
10	10	Both	0.35 m	Normal	Normal	No	Declarative
11	11	Both	0.5 m	Normal	Normal	No	Declarative
12	12	Both	1 m	Normal	Normal	No	Declarative
13	13	Left	0.35 m	Normal	Normal	No	Declarative
14	14	Both	2 m	Normal	Normal	No	Declarative
15	15	Right	0.35 m	Normal	Normal	No	Declarative
16	16	Both	0.35 m	Normal	Normal	No	Interrogative
17	17	Both	0.5 m	Normal	Normal	No	Interrogative
18	18	Both	1 m	Normal	Normal	No	Interrogative
19	1	Left	0.35 m	Normal	Normal	No	Interrogative
20	2	Both	1 m	Normal	Normal	No	Interrogative
21	11	Both	0.5 m	Normal	Normal	No	Declarative
22	3	Right	0.35 m	Normal	Normal	No	Interrogative
23	4	Both	0.35 m	Normal	Normal	No	Imperative
24	5	Both	0.5 m	Normal	Normal	No	Imperative
25	6	Both	1 m	Normal	Normal	No	Imperative
26	7	Left	0.35 m	Normal	Normal	No	Imperative
27	8	Both	2 m	Normal	Normal	No	Imperative
28	9	Right	0.35 m	Normal	Normal	No	Imperative
29	10	Both	0.35 m	Normal	Normal	No	Declarative
30	11	Both	0.5 m	Normal	Normal	No	Declarative
31	12	Both	1 m	Normal	Normal	No	Declarative
32	12	Both	1 m	Normal	Normal	No	Declarative
33	13	Left	0.35 m	Normal	Normal	No	Declarative

Continued on next page

Table A.1 – *Continued from previous page*

Order	Trigger ID	Ear	Distance	Pitch	Speed	Noise	Sentence Type
34	14	Both	2 m	Normal	Normal	No	Declarative
35	15	Right	0.35 m	Normal	Normal	No	Declarative
36	16	Both	0.35 m	Normal	Normal	No	Interrogative
37	17	Both	0.5 m	Normal	Normal	No	Interrogative
38	18	Both	1 m	Normal	Normal	No	Interrogative
39	1	Left	0.35 m	Normal	Normal	No	Interrogative
40	2	Both	1 m	Normal	Normal	No	Interrogative
41	3	Right	0.35 m	Normal	Normal	No	Interrogative
42	4	Both	0.35 m	Normal	Normal	No	Imperative
43	13	Left	0.35 m	Normal	Normal	No	Declarative
44	5	Both	0.5 m	Normal	Normal	No	Imperative
45	6	Both	1 m	Normal	Normal	No	Imperative
46	7	Left	0.35 m	Normal	Normal	No	Imperative
47	8	Both	2 m	Normal	Normal	No	Imperative
48	9	Right	0.35 m	Normal	Normal	No	Imperative
49	19	Both	0.35 m	High	Normal	No	Declarative
50	20	Both	0.35 m	High	Normal	No	Interrogative
51	21	Both	0.35 m	High	Normal	No	Imperative
52	19	Both	0.35 m	High	Normal	No	Declarative
53	20	Both	0.35 m	High	Normal	No	Interrogative
54	14	Both	2 m	Normal	Normal	No	Declarative
55	21	Both	0.35 m	High	Normal	No	Imperative
56	19	Both	0.35 m	High	Normal	No	Declarative
57	20	Both	0.35 m	High	Normal	No	Interrogative
58	21	Both	0.35 m	High	Normal	No	Imperative
59	22	Both	0.35 m	Low	Normal	No	Declarative
60	23	Both	0.35 m	Low	Normal	No	Interrogative
61	24	Both	0.35 m	Low	Normal	No	Imperative
62	22	Both	0.35 m	Low	Normal	No	Declarative
63	23	Both	0.35 m	Low	Normal	No	Interrogative
64	24	Both	0.35 m	Low	Normal	No	Imperative
65	15	Right	0.35 m	Normal	Normal	No	Declarative
66	22	Both	0.35 m	Low	Normal	No	Declarative
67	23	Both	0.35 m	Low	Normal	No	Interrogative
68	24	Both	0.35 m	Low	Normal	No	Imperative
69	25	Both	0.35 m	Normal	Slow	No	Declarative
70	26	Both	0.35 m	Normal	Slow	No	Interrogative
71	27	Both	0.35 m	Normal	Slow	No	Imperative
72	25	Both	0.35 m	Normal	Slow	No	Declarative
73	26	Both	0.35 m	Normal	Slow	No	Interrogative
74	27	Both	0.35 m	Normal	Slow	No	Imperative
75	25	Both	0.35 m	Normal	Slow	No	Declarative
76	16	Both	0.35 m	Normal	Normal	No	Interrogative
77	26	Both	0.35 m	Normal	Slow	No	Interrogative
78	27	Both	0.35 m	Normal	Slow	No	Imperative

Continued on next page

Table A.1 – *Continued from previous page*

Order	Trigger ID	Ear	Distance	Pitch	Speed	Noise	Sentence Type
79	28	Both	0.35 m	Normal	Fast	No	Declarative
80	29	Both	0.35 m	Normal	Fast	No	Interrogative
81	30	Both	0.35 m	Normal	Fast	No	Imperative
82	28	Both	0.35 m	Normal	Fast	No	Declarative
83	29	Both	0.35 m	Normal	Fast	No	Interrogative
84	30	Both	0.35 m	Normal	Fast	No	Imperative
85	28	Both	0.35 m	Normal	Fast	No	Declarative
86	29	Both	0.35 m	Normal	Fast	No	Interrogative
87	17	Both	0.5 m	Normal	Normal	No	Interrogative
88	30	Both	0.35 m	Normal	Fast	No	Imperative
89	31	Both	0.35 m	Normal	Normal	Yes	Declarative
90	32	Both	0.35 m	Normal	Normal	Yes	Interrogative
91	33	Both	0.35 m	Normal	Normal	Yes	Imperative
92	31	Both	0.35 m	Normal	Normal	Yes	Declarative
93	32	Both	0.35 m	Normal	Normal	Yes	Interrogative
94	33	Both	0.35 m	Normal	Normal	Yes	Imperative
95	31	Both	0.35 m	Normal	Normal	Yes	Declarative
96	32	Both	0.35 m	Normal	Normal	Yes	Interrogative
97	33	Both	0.35 m	Normal	Normal	Yes	Imperative
98	18	Both	1 m	Normal	Normal	No	Interrogative

DEPARTMENT OF SOME SUBJECT OR TECHNOLOGY
CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden
www.chalmers.se

CHALMERS
UNIVERSITY OF TECHNOLOGY