MASTERARBEIT | MASTER'S THESIS

Titel | Title

The Sound of Emotion: Behavioural and Electrophysiological Findings on Autism and Auditory Affective Processing

verfasst von | submitted by

Katharina Egger BSc

angestrebter akademischer Grad | in partial fulfilment of the requirements for the degree of

Master of Science (MSc)

Wien | Vienna, 2026

Studienkennzahl lt. Studienblatt | Degree programme code as it appears on the student record sheet:	UA 066 840
Studienrichtung lt. Studienblatt | Degree programme as it appears on the student record sheet:	Masterstudium Psychologie
Betreut von | Supervisor:	Assoz. Prof. Matthew Pelowski PhD

Acknowledgements

I would like to sincerely thank my supervisor at the University of Vienna, Associate Professor Matthew Pelowski, for his guidance and support throughout the writing of this thesis.

I am also very grateful to Professor Jonna Vuoskoski and the AEPA project team at the University of Oslo for giving me the opportunity to contribute to their research and for the formative research experience I gained. In particular, I would like to thank Liv Merve Akca Abrahamsson for her mentorship during my stay in Oslo and her continued support throughout the writing process.

My deepest thanks go to my family for their constant support, love, and encouragement - and for always being there for me, every step of the way.

Finally, I would like to thank Paul and Anja, my friends and thesis-writing companions, for being by my side throughout this process. Writing this thesis would not have been nearly as rewarding and funny without them.

Note on the Use of AI-Supported Tools

AI-supported tools were used in a supportive capacity during the preparation of this thesis. ChatGPT (OpenAI; versions 4o and 5), NotebookLM (Google), and elicit.com were used to assist with literature exploration, text organization, language refinement, and technical support in programming. All interpretations, analyses, and final text were independently developed and critically evaluated by the author.

1. Introduction

Emotions are a fundamental component of everyday social interaction. The ability to perceive and interpret emotional cues is central to effective communication. Such cues are conveyed not only visually through facial expressions but also auditorily through vocalisations, including emotional prosody and nonverbal affective sounds such as laughter or sobbing. The accurate perception of emotional information in the auditory domain therefore represents a key element of successful social communication (Wallbott & Scherer, 1986).

In autism spectrum disorder (ASD), difficulties in emotion perception have been widely reported (e.g. Leung et al., 2022; Uljarevic & Hamilton, 2013). A substantial body of research has demonstrated impairments in the recognition of emotional facial expressions, indicating that visual emotion processing is frequently affected. In contrast, findings regarding auditory emotion perception are far less consistent. Individuals with ASD often show atypical sensory processing, and some studies even suggest enhanced auditory perceptual abilities, with certain sounds being perceived or discriminated more precisely than in neurotypical individuals (e.g. Ouimet et al., 2012). This raises the question of whether emotional information conveyed through sound is processed differently in ASD.

When it comes to affective auditory perception specifically, research findings are mixed. While some studies report intact recognition and evaluation of emotional sounds (e.g. (Grossman et al., 2010; Jones et al., 2011; Järvinen et al., 2016), others indicate reduced accuracy (e.g. Globerson et al., 2015), altered intensity ratings (e.g. Gebauer et al., 2014), or atypical neural responses (Korpilahti et al., 2007; M. D. Lerner et al., 2013). Much of the existing literature relies predominantly on behavioural measures, such as emotion identification or rating tasks. Consequently, it remains unclear whether observed similarities or differences at the behavioural level reflect comparable neural processing mechanisms or instead are related to underlying differences at certain neural stages of auditory or affective processing.

The present Master's thesis addresses this gap by combining behavioural and electrophysiological (EEG) approaches to investigate auditory emotion processing in ASD. Participants were presented with emotional vocalisations while neural responses were recorded using EEG. In a subsequent task, the same vocalisations were re-presented and participants identified the expressed emotion and rated its perceived intensity. Behavioural

measures of recognition accuracy and subjective intensity were analyzed alongside event-related potentials (ERPs) in the form of the P200 component.

The P200 is an ERP component that has been linked to the processing of salient auditory and affective information. It is thought to reflect early stages of sensory-affective integration and attentional allocation to emotionally relevant acoustic features. Examining the P200 therefore allows for the investigation of whether potential group differences emerge already at early stages of auditory emotion processing, even in the absence of overt behavioural differences.

By integrating behavioural performance with neural indices, this thesis examines whether individuals with ASD and neurotypical controls differ in the processing of affective auditory stimuli at the behavioural level, the neural level, or both. In particular, it investigates whether potential differences emerge at early stages of sensory-affective encoding and how neural responses relate to subjective behavioural judgments. Focusing on brief, nonverbal emotional vocalisations that are largely independent of linguistic and contextual cues, the study provides new evidence on auditory emotion processing in ASD. Such insights are important for refining models of sensory-affective processing in autism and may help explain inconsistencies in previous findings that relied solely on behavioural measures.

2. Theoretical Background

The following chapter provides an overview of the theoretical and empirical foundations relevant to the present thesis, focusing on vocal emotion processing and autism spectrum disorder. It is structured into three main sections.

The first section addresses auditory emotion perception, with a specific focus on vocal emotion processing. It introduces key theoretical concepts and summarizes findings from cognitive and neuropsychological research on emotion perception and its underlying neural mechanisms. This section further discusses event-related potentials (ERPs), with particular emphasis on the P200 component and its relevance for early stages of auditory and affective processing.

The second section provides an overview of ASD, including its diagnostic criteria, core characteristics, prevalence, and etiological factors, establishing the clinical framework necessary for understanding emotion processing differences in ASD.

The third section integrates the two topics by reviewing the current state of research on vocal emotion processing in individuals with ASD. Both behavioural and neurophysiological findings are considered, and existing inconsistencies in the literature and key gaps that motivate the aims of the present thesis are discussed.

2.1. Auditory Emotion Processing

Emotions are fundamental components of our everyday lives – they guide our behaviour and shape our experiences, motivate us to seek situations that are beneficial to us, help us avoid those that may be harmful or unpleasant (Baumeister et al., 2007). They can be a crucial influence when we make decisions (Lerner et al., 2015) and guide us in social interactions (Van Kleef, 2009).

Many theoretical models have been proposed to explain the nature and function of emotions. One influential psychological theory of emotion was developed by Ekman (1992), who identified a set of “basic” emotions, anger, fear, happiness, sadness, surprise and disgust, that are proposed to be universal across cultures and biologically innate. According to this view, basic emotions are processed rapidly and automatically and are associated with distinct patterns of facial expression. These emotions are assumed to serve adaptive functions by enabling fast and efficient responses to biologically relevant events in the environment.

In contrast to this categorical perspective, which conceptualizes emotions as discrete affective states that exist without reference to each other, Russel (1980) proposed a

dimensional account of emotion in the form of the circumplex model. This model represents emotional experiences as positions within a continuous space defined by the dimensions of valence (ranging from pleasant to unpleasant) and arousal (ranging from low to high activation).

Most researchers distinguish three main components that constitute an emotional response: 1) a physiological component (such as blood pressure, heart rate and respiration), 2) a behavioural component (such as facial expressions, posture, gestures, vocal changes, and 3) a cognitive component (such as subjective feelings, thoughts, and language, Scherer, 1984). There are several brain regions that are sensitive to and selective for emotion perception. Different emotional states are represented in distinct patterns of brain activation (Kragel & LaBar, 2015). The posterior superior temporal sulcus is a core region in the "social brain" and plays an important role in the perception of social information and also crucial for developing theory of mind (Yang et al., 2015). This region is directly connected with primary auditory and visual brain regions and is selectively activated to social stimuli vs. non-social cues (Watson et al., 2014) and responsible for the temporal integration of multisensory (e.g. visual and auditory) emotional cues (Belin et al., 2000; Kreifelts et al., 2009).

Another important brain region in the perception of emotion is the amygdala, as it codes for the salience of emotional information. When lesioned, it can result in the inability to recognize emotions from facial expressions (Adolphs et al., 1994). In general, the amygdala seems to be especially relevant for unpredictable or unambiguous emotional stimuli (Adolphs, 2009).

The somatosensory cortex plays a crucial role in discriminating visual (Adolphs et al., 2000) and auditory emotions (Banissy et al., 2010) and certain emotional states correlate with certain voxel activations as shown with functional magnetic resonance imaging (fMRI) studies (Kragel & LaBar, 2015). Also, activation in the somatosensory cortex is similar both when experiencing a certain emotion oneself and perceiving the same emotion in others (Winston et al., 2003). Suppression of activity in the somatosensory cortex leads to impaired discrimination of auditory emotions (Banissy et al., 2010).

Although much of the research on emotion perception has focused on the visual domain, particularly facial expressions, emotional information is also conveyed through auditory signals. Vocal expressions constitute a primary channel through which affective states are communicated in everyday interaction. Because emotional meaning in the voice relies on modality-specific acoustic and neural mechanisms, vocal emotion processing

represents a distinct domain of affective perception. The following sections therefore examine the mechanisms and characteristics of vocal emotion processing in greater detail.

2.1.1. Vocal Emotion Processing

Vocal information is a central aspect of daily life. Beyond carrying linguistic content, voices convey cues about the speaker's identity, intentions, and emotional state (Latinus & Belin, 2011). Among the various types of information transmitted by the human voice, emotional cues are particularly important for social interaction. The ability to perceive and interpret vocal expressions of emotion enables listeners to infer others' affective states even without visual or linguistic input (Bachorowski, 1999).

Often, it only takes a brief moment to detect whether a speaker sounds annoyed, anxious, or surprised. Emotional states are accompanied by physiological changes such as alterations in heart rate, muscle tension, and blood flow, which systematically influence acoustic properties of the voice (Scherer, 1986). Together with communicative intentions, these physiological processes shape characteristic vocal patterns associated with specific emotions. As a result, emotional states tend to become audible in the voice—whether intentionally expressed or not (Banse & Scherer, 1996; Patel et al., 2011).

Emotional meaning in the voice is conveyed through a set of psychoacoustic features, such as loudness (amplitude), the fundamental frequency (F0), which reflects the vibration rate of vocal folds and is perceived as pitch, as well as tempo or speech rate (Banse & Scherer, 1996; Coutinho & Dibben, 2013). Additionally, voice quality is an important cue: it refers to the perceptual impression created by an acoustic pattern made up of the distribution of a speaker's frequency pattern (e.g. a shrill or soft-sounding voice, Banse & Scherer, 1996).

Not all emotions are recognized equally well in vocal expression. Emotions such as anger and sadness tend to be detected more reliably than fear or happiness (Chronaki et al., 2018). Beyond differences across emotion categories, substantial individual variability exists in the ability to recognize emotions. Research suggests that emotion recognition ability is positively associated with emotional understanding, empathy, and openness, and negatively related to alexithymia (e.g., Laukka et al., 2021). Furthermore, gender differences have been reported, with female adults and children typically outperforming males in identifying nonverbal emotional expressions (Allgood & Heaton, 2015; Laukka et al., 2021).

Vocal expression of emotion occurs in two main forms: speech prosody (intonation patterns accompanying spoken language) and nonverbal vocalisations (brief affective utterances such as laughter, sobbing, or screaming (Mauchand & Zhang, 2023). Producing

prosody requires highly precise and rapid articulatory movements, whereas nonverbal affective sounds rely on less fine-grained supraglottal movements (Schirmer & Kotz, 2006; Scott et al., 2010). Affective vocalisations appear to be partly universal across cultures, as shown for emotional speech (Scherer et al., 2001) and nonverbal expressions (Sauter et al., 2009).

Taken together, vocal emotion perception relies on the rapid extraction of structured acoustic cues that reflect underlying physiological states and communicative intentions. Although certain emotional expressions appear to be broadly recognizable across cultures, recognition accuracy varies across emotion categories and individuals. Vocal affect perception therefore represents a complex and multifaceted process, integrating low-level acoustic analysis with higher-level social interpretation.

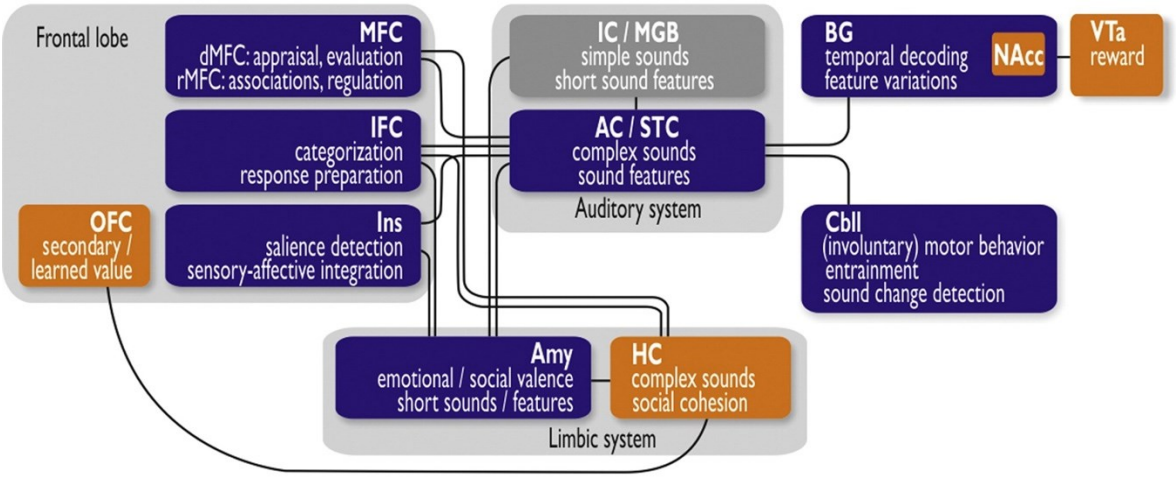
2.1.2. Neurocognitive Models of Auditory Affect Processing

Research on auditory emotion processing has proposed cognitive and neural network models describing a set of cortical and subcortical regions consistently activated by emotional sounds, ranging from short affective vocalisations to emotional music.

The amygdala plays a central role, as it responds particularly strongly to short and salient acoustic events and may act as a detector for novel and emotionally relevant sounds (Blackford et al., 2010). It is strongly interconnected with other regions that play a role in emotional sound processing (Frühholz et al., 2016). By contrast, the auditory cortex is more engaged in decoding complex sounds and in integrating emotional information over time (Frühholz et al., 2016). Another important region is the frontal cortex, which supports higher-order processes such as evaluation, appraisal, and the categorical representation of affective sounds. Subregions within the FC have distinct functions: the medial frontal cortex is associated with memory-related aspects, the inferior frontal cortex with emotional classification, and the insula with salience detection and the linking of acoustic input to internal emotional states. Additional subcortical structures are also important. The basal ganglia is connected to the AC and anticipates and decodes the temporal information of sounds, while the cerebellum is crucial for behavioural reactions to emotional sounds, such as involuntary motor reactions (Frühholz et al., 2016).

Figure 1

Neural model for the processing of affective sounds

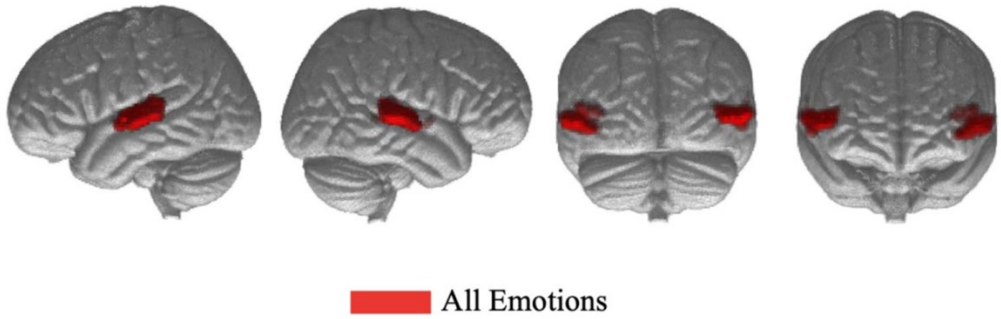


Note. Figure from Frühholz et al. 2016.

Neuroimaging research has identified voice-selective areas in the human brain–so-called Temporal Voice Areas in the superior temporal sulcus (STS), which show a robust response to vocal sounds including both speech and nonverbal expressions (Belin et al., 2000). The STS is not only sensitive to vocal information in general but appears to play a particularly important role in decoding emotional cues from nonverbal vocal expressions. Activity in this region is significantly enhanced for emotional compared to neutral voices, underscoring its role in the perceptual analysis of affective vocal signals (Grandjean et al., 2005; Mauchand & Zhang, 2023, Fig. 2). These regions in the STS are considered central hubs for the extraction of socially relevant information from voices and form the basis for higher-level vocal and affective processing.

Figure 2

Cluster Activation in the STS for Emotional Vocalisations



Note. Taken from Mauchand and Zhang (2023)

With a similar approach, fMRI studies have identified a network of "Emotional Voice Areas" (EVA) in the STC and primary auditory regions that are specifically sensitive to affective vocal information (Ethofer et al., 2012). Also, differential activation patterns have been observed depending on stimulus type: affective vocalisations elicit stronger responses in the left Heschl's gyrus, whereas emotional speech prosody is associated with greater activation in the right posterior superior temporal gyrus (Mauchand & Zhang, 2023).

Regarding the temporal aspects of affective vocal processing, neurocognitive models propose a multi-stage framework (Schirmer & Kotz, 2006; Bestelmeyer et al., 2014), suggesting a hierarchical flow of information from early auditory regions (e.g. the ventral auditory pathway) to areas associated with emotional and cognitive evaluation. According to this model, this begins with (1) low-level analyses of acoustic features in the primary and secondary auditory cortex, followed by (2) the integration of emotionally relevant cues into a coherent perceptual representation within the superior temporal sulci and gyri. These representations are subsequently projected to frontal and limbic regions, where (3) higher-order evaluative and cognitive processes take place, and emotional significance is applied (Schirmer & Kotz, 2006).

EEG Findings

Event-related potentials (ERPs) are voltage changes in the electroencephalogram (EEG) that are time-locked to sensory, cognitive, or motor events and reflect synchronized neural activity in response to external stimuli. ERPs are extracted from the ongoing EEG signal using filtering and signal-averaging techniques, allowing for the isolation of stimulus-related brain responses with high temporal resolution (Picton et al., 2000). In the auditory

domain, cortical auditory evoked potentials provide insight into successive stages of auditory processing, with distinct waveform components reflecting activity in different regions and processing levels of the auditory system (Schirmer & Kotz, 2006).

ERP components are conventionally labeled according to their polarity (positive or negative) and their typical latency following stimulus onset (e.g., P100, N100). Based on their temporal characteristics, these components are often classified into early and late responses. Early components, typically occurring within the first 50–200 milliseconds (ms) post-stimulus, are thought to reflect pre-attentive sensory encoding and early perceptual analysis in the cerebral cortex. Later components emerging after approximately 200 ms are associated with higher-order cognitive, attentional, and evaluative processes (Picton et al., 2000).

Regarding the temporal characteristics of vocal emotion perception, behavioural and electrophysiological findings indicate that the extraction of affective information from auditory signals occurs rapidly. Emotional content in both speech prosody and music can be recognized following very brief stimulus exposures, in some cases within the first 100 ms after onset (Nordström & Laukka, 2019). While this suggests that affective cues are available early in the auditory processing stream, recognition accuracy continues to improve as more acoustic information becomes available. For emotional speech prosody, accuracy has been shown to stabilize within a time window of approximately 500–600 ms, although there seems to be substantial variability across emotions—for example, stabilization occurring earlier for fearful prosody and considerably later for emotions such as disgust (Pell & Kotz, 2011).

Electrophysiological studies indicate that affective vocal processing follows a temporally structured sequence of sub-processes, which can be mapped onto distinct ERP components. Consistent with neurocognitive models of vocal emotion processing, these findings support a three-stage framework comprising early sensory encoding, intermediate emotional integration, and later cognitive evaluation. These stages are commonly indexed by a central N100, a frontal P200, and late positive potential (LPP; Nussbaum et al., 2022; Schirmer & Kotz, 2006; Schirmer & Gunter, 2017).

The P200

The P200, typically observed around 180–250 ms post-stimulus with a fronto-central scalp distribution, has been consistently implicated in the early integration of emotional information in auditory stimuli (Nussbaum et al., 2022; Schirmer & Kotz, 2006; Paulmann & Kotz, 2008). For nonverbal vocalisations, multiple studies have demonstrated modulation of

P200 amplitudes by emotional valence or arousal, with emotional sounds generally eliciting larger P200 amplitudes compared to neutral stimuli (Gerdes et al., 2013; Liu et al., 2012; Sauter & Eimer, 2010; Schirmer & Gunter, 2017). These findings suggest that the P200 indexes a critical processing stage at which emotional relevance is extracted from auditory input.

2.2. Autism Spectrum Disorder

Autism spectrum disorder is a neurodevelopmental condition characterized by atypical social communication and social interaction which are often associated with difficulties in the expression and understanding of emotions.

2.2.1. Diagnostic Criteria and Prevalence

According to the current version of the *Diagnostic and Statistical Manual of Mental Disorders* (DSM-5), deficits in social–emotional reciprocity and communicative behaviours, in conjunction with restricted and repetitive patterns of behaviour, are required for a diagnosis of ASD. Additional diagnostic criteria include the presence of symptoms in early childhood and clinically significant impairment in everyday functioning (see Table 1; APA, 2022). While earlier editions of the DSM and the *International Statistical Classification of Diseases and Related Health Problems* (ICD) classified autism under the category of pervasive developmental disorders, distinguishing between subtypes such as childhood autism, atypical autism, and Asperger syndrome, the current editions (DSM-5 and ICD-11) have adopted a single diagnostic category of ASD. This revision was motivated by the lack of sufficient scientific evidence for reliable and distinct criteria to separate the subtypes, as well as by the considerable symptom overlap between them (Noterdaeme et al., 2017). The current DSM-5 diagnostic criteria for ASD are summarized in Table 1.

Table 1

Diagnostic criteria for Autism Spectrum Disorder (DSM-5)

Criterion	Domain	Description
A	Social communication and interaction	Persistent deficits across contexts, as manifested by all of the following: • Social-emotional reciprocity • Nonverbal communicative behaviours used for social interaction • Developing, maintaining, and understanding relationships
B	Restricted and repetitive behaviours	As manifested by at least two of the following: • Stereotyped or repetitive motor movements, use of objects, or speech • Insistence on sameness or inflexible adherence to routines • Highly restricted, fixated interests (abnormal in intensity or focus) • Hyper- or hyporeactivity to sensory input or unusual interest in sensory aspects of the environment
C	Developmental onset	Symptoms present in the early developmental period
D	Functional impairment	Symptoms cause clinically significant impairment in social, occupational, or other important areas of functioning
E	Differential diagnosis	Disturbances not better explained by intellectual disability or global developmental delay

Note. Adapted from APA (2013)

In addition to the diagnostic criteria, the DSM-5 specifies severity levels for ASD based on the degree of support required, ranging from Level 1 (requiring support), through Level 2 (requiring substantial support), to Level 3 (requiring very substantial support).

Despite the standardized diagnostic criteria, diagnosing ASD remains challenging. The diagnostic process is complex and requires specialized, interdisciplinary clinical expertise, and symptoms can vary widely between individuals and overlap with other developmental disorders. Additionally, commonly used diagnostic tools are time-consuming and not well suited for large-scale screening. Nevertheless, early and accurate diagnosis is crucial, as early intervention has been shown to improve long-term outcomes for individuals with ASD (Lord et al., 2022; Wang & Wang, 2024).

Prevalence numbers for ASD vary considerably across different studies, which can be attributed to differing diagnostic methods, sample ages or sample sizes. From the latest systematic reviews, a median prevalence of roughly 100 cases of ASD in 10.000 individuals can be estimated, with a range between 1.09 and 436.0 (Wang & Wang, 2024; Zeidan et al., 2022). Regarding prevalence in Norway, a systematic review by (Jensen de López & Thirup

Møller, 2024) estimates a prevalence of 0.6 in Norwegian children. There has been a steady rise in the prevalence of diagnosed autism over time, especially in Asia, Europe and North America and in diagnoses of high-functioning autism (Salari et al., 2022). This can be partly explained by a growing body of research on autistic traits and their diverse manifestation, alongside changing diagnostic criteria and the improvement of screening instruments. Another possible explanation is enhanced awareness of autism in the general population and possible also due to an increase in risk factors (Baxter et al., 2015; Lai et al., 2014).

Autism is diagnosed approximately four times more frequently in males than in females (Giarelli et al., 2010; Zeidan et al., 2022). However, accumulating evidence suggests that this pronounced gender difference may partly reflect underrecognition of autism in females rather than true prevalence differences. Women with ASD have been shown to more frequently employ compensatory strategies, often referred to as masking or camouflaging, in order to adapt their behaviour to social expectations (Milner et al., 2019). While such strategies may facilitate social integration, they can also contribute to delayed or missed diagnoses, with many women receiving a diagnosis later in life than would be optimal (Atherton et al., 2022). Moreover, diagnostic criteria and assessment practices have historically been shaped by male-typical symptom presentations, which may further obscure autistic traits in females (Baron-Cohen et al., 2011; Loomes et al., 2017). These considerations are important when interpreting findings from clinical samples and highlight that observed gender distributions in research cohorts may not directly reflect underlying prevalence.

It is estimated that more than 70 % of people with ASD have comorbid conditions, the most common of which include (≈ 45%), attention-deficit/hyperactivity disorder (28–44%), anxiety (42–56%), depression (12–70%), and sleep disorders (50–80%; Lai et al., 2014)). Another common neurological comorbidity is epilspy, which has been reported to co-occur in 2,4% - 26% of ASD (El Achkar & Spence, 2015). Regarding the etiology of autism spectrum disorder, it cannot be attributed to a single universal biomarker or risk factor. Instead, ASD is more likely the result of a complex interplay between multiple genetic and environmental influences (Hertz-Picciotto et al., 2018; Newschaffer et al., 2007).

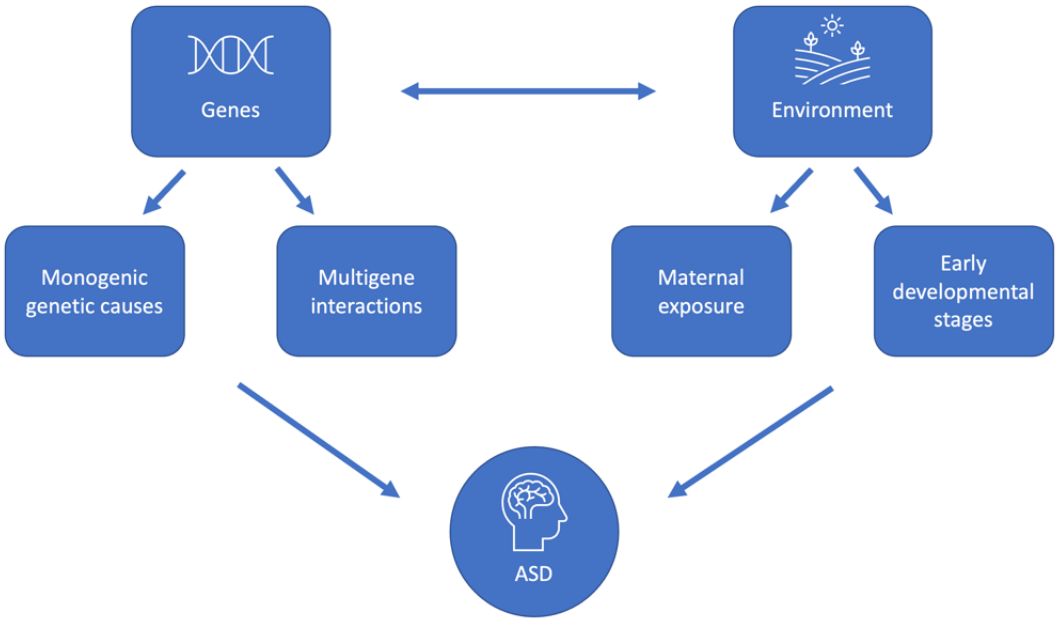
A small proportion of ASD cases involve identifiable genetic factors. For instance, chromosomal rearrangements or single-gene disorders are found in an estimated 5–15% of individuals with ASD, and an additional 5–10% of cases are linked to rare de novo or inherited copy-number variations (CNVs, Devlin & Scherer, 2012). However, the precise

genetic etiology remains unknown for at least 70% of ASD cases, indicating that non-genetic factors also play a substantial role in the development of the disorder.

Studies investigating pre-, peri-, and postnatal risk factors have yielded inconsistent and, in some cases, contradictory findings. Frequently discussed environmental risk factors include advanced maternal age (over 35 years), maternal infections such as rubella or cytomegalovirus, exposure to toxins during pregnancy, and elevated concentrations of adrenal cortical hormones in the amniotic fluid (Gardener et al., 2009; Ornoy et al., 2016). In sum, current evidence suggests that ASD most likely arises from an interaction between an individual's genetic makeup and environmental factors encountered during development (see Figure 3). In particular, certain genetic variants may increase susceptibility to environmental influences, such that exposure to specific risk factors has a stronger impact on neurodevelopment in genetically vulnerable individuals. When these influences occur during sensitive periods of brain development, they may contribute to an increased likelihood of developing ASD (Masi et al., 2017; Qin et al., 2024).

Figure 3

Illustration of Gene–Environment Interactions



Note. Adapted from Qin et al. (2024)

2.2.2. Theories of Autism

A variety of hypotheses and conceptual frameworks to explain autism have been proposed, emphasizing different aspects of the autistic phenotype, including cognitive,

neurobiological, genetic, and social dimensions. An early approach in theorizing about autism was the attempt to identify a specific "core deficit" that would be both unique to ASD (it must be exclusive to the group) and at the same time universally applicable (It must apply to every person in the group). However, no theory has been able to meet this claim to date, and given the heterogeneity of the condition, it is questionable whether such a framework could ever be achieved. In recent years, more complex and multidimensional theoretical models have been developed, aiming to describe autism across multiple levels of analysis.

While some theories see impaired social communication and related processes as the core root for autism, other theories have tried to explain ASD with more general perceptual differences that affects both social and non-social information:

Theory of Mind

One early and very influential account of autism is The Theory of Mind (ToM) hypothesis, which proposes that individuals with autism have difficulties inferring the mental states of others (= mentalising), which leads to impairments in social interaction (Fletcher-Watson & Happé, 2019; Long et al., 2025; Rajendran & Mitchell, 2007).

Early studies (e.g. Baron-Cohen et al., 1985) found that most children with ASD had difficulties with tasks designed to measure the Theory of Mind, such as the false belief task. In this task, children are asked to predict what a character will do based on the character's mistaken belief about a situation. To succeed, the child must infer that another person can have a belief that differs from reality and also from the child's own knowledge – which shows whether they understand that others have their own mental states (Wimmer & Perner, 1983).

There is a lot of evidence suggesting that individuals with ASD show difficulties with ToM, not only in the aspect of knowledge, but also in understanding other's wishes, intentions, and the ability to deceive (Gao et al., 2023; Steele et al., 2003; Talwar et al., 2012; Williams & Happé, 2009). However, it is questioned whether ToM is able to pose as an unifying theory to explain autism. On the one hand, ToM deficits are also observed in individuals with other psychiatric conditions , and on the other hand, it only accounts for certain aspects of ASD (Bruning et al., 2005).

Local vs. Global Processing

The term 'central coherence' was first introduced by Frith & Happé (1994) and refers to the process of integrating individual pieces of information into a larger whole, binding

information together and seeing ‘the bigger picture’. While this habit is more often seen in neurotypical individuals, individuals with ASD have been shown to display a reduced tendency for central coherence, which lead to the formulation of the "Theory of Weak Central Coherence" (WCC; Frith & Happé, 1994).

The WCC model proposes a continuum of cognitive styles, with hyperfocus on details (weak coherence, local processing style) on one end, and hyperfocus on the global meaning (strong coherence, global processing style) on the other end (Happé & Frith, 2006). Individuals with ASD are generally found toward the detail-focused end of this continuum, and WCC has therefore been suggested as a fundamental mechanism underlying characteristic features of autism, such as cognitive rigidity, insistence on sameness, and heightened attention to detail (Happé & Frith, 2006; Scher Lisa & Shyman, 2019).

This reduced tendency for global processing is not only seen to be disadvantageous, but has also been linked to enhanced local processing, including superior attention to and memory for details in which individuals with ASD frequently outperform neurotypical individuals (Fletcher-Watson & Happé, 2019; Frith & Happé, 1994). A commonly used measure to assess central coherence is the Embedded Figures Task (EFT), in which participants are asked to detect simple shapes hidden within a larger, more complex figure. Performance is typically evaluated by the speed and accuracy of identifying the target shape. Individuals with ASD often outperform neurotypical individuals in this task, showing faster detection times (Scher Lisa & Shyman, 2019).

Neuroscientific investigations of the WCC framework have reported altered patterns of neural activation associated with local processing biases in autism. For example, Manjaly et al., (2007) observed increased activation in right-hemispheric regions, including extrastriate visual areas and the cerebellum, in autistic compared to neurotypical participants during visual processing tasks. Other studies have suggested that enhanced local processing in autism may be related to hyperexcitability of the extrastriate cortex, potentially facilitating superior perception of fine visual details (Schwarzkopf et al., 2014). Taken together, contemporary accounts of WCC increasingly emphasize that autistic cognition is better characterized by a superiority in local processing rather than a fundamental deficit in global processing (Scher Lisa & Shyman, 2019).

2.3. Auditory Processing in ASD

Research on sensory processing in ASD has consistently identified atypical responses to auditory stimulation. Individuals with ASD frequently show both heightened sensitivity

(hyperacusis) and reduced reactivity (hyporeactivity) to sound, as well as difficulties filtering background noise (Rosenhall et al., 1999; Teder-Sälejärvi et al., 2005). Studies report atypical behavioural and neural responses to a wide range of auditory inputs, indicating that the processing of acoustic information may differ in certain ways from that of neurotypical individuals (O'Connor, 2012).

When it comes to simple auditory features, such as pitch discrimination or the detection of pure tones, several studies have reported enhanced abilities in individuals with ASD compared to neurotypical controls (Heaton et al., 1998; O'Riordan & Passetti, 2006; Ouimet et al., 2012) Enhanced pitch memory (Heaton, 2003) and superior detection of changes in melodies have also been observed, despite comparable global music processing between groups (Mottron et al., 2000). This pattern has been interpreted within the Enhanced Perceptual Functioning model (Mottron et al., 2006), which proposes a locally oriented perceptual style characterized by heightened low-level processing in ASD. Closely related to this account, the Neural Complexity Hypothesis (Samson et al., 2006) suggests that individuals with ASD show heightened sensitivity to simple auditory stimuli processed in primary auditory cortex, but reduced performance for acoustically complex sounds that require integration across temporal and spectral dimensions in higher-order auditory regions. Support for these models comes from a more recent electrophysiological study that demonstrated atypicalities in early perceptual processing, that may ultimately influence later cognitive mechanisms (Kaplan-Kahn et al., 2021).

At the neural level, there is also evidence for structural brain differences connected with auditory processing in ASD individuals, with studies reporting higher cortical thickness in the auditory cortex (Foster et al., 2015; Hyde et al., 2010). A meta-analysis by Williams and colleagues (2021) reported irregular neural processing of simple auditory stimuli, with prolonged ERP latencies and reduced amplitudes. When it comes to more complex sounds, studies report delayed or reduced responses as well as atypical lateralization of auditory stream processing in ASD individuals (Jorgensen et al., 2021).

A systematic review on auditory processing in ASD by Ouimet and colleagues (2012) came to the conclusion that there is strong support for enhanced processing of simple auditory stimuli, but the research findings on complex auditory stimuli are mixed. These findings raise the question of how such auditory processing characteristics affect the perception of socially relevant and emotionally meaningful sounds, such as human vocalisations.

Vocal Sound Processing and ASD

Human vocalisations constitute a particularly important class of auditory stimuli, as they convey not only acoustic information but also socially and emotionally relevant cues. Several studies have therefore specifically examined whether the processing of vocal sounds differs from that of non-vocal auditory stimuli in ASD.

Early neuroimaging evidence suggests atypical neural responses to vocal sounds in individuals with ASD. Gervais et al. (2004) reported reduced activation in voice-selective regions of the STS in participants with ASD when listening to human vocalisations, accompanied by less pronounced activation in the primary auditory cortex compared to neurotypical controls. Importantly, these differences were specific to vocal stimuli, as responses to non-vocal sounds did not differ between groups. Additionally, participants with ASD showed poorer memory for vocal sounds, while recall for non-vocal sounds was intact, suggesting reduced sensitivity or attentional efforts for socially relevant auditory information.

Electrophysiological studies provide similar evidence for atypical processing of vocal speech sounds in ASD. Reduced ERP amplitudes in response to vocal stimuli have been reported, indicating alterations in early perceptual and cognitive stages of vocal processing (Nelson & McCleery, 2008). Similarly, Whitehouse and Bishop (2008) found attenuated N2, P3, and N4 components for vocal speech sounds in participants with ASD, whereas responses to complex non-vocal tones were comparable to those of neurotypical individuals. Notably, the group differences were no longer observed when participants were instructed to attend to the auditory stream, suggesting that attentional engagement can modulate these processing deficits.

At a meta-analytic level, Jorgensen and colleagues (2021) compared studies that used electrophysiological measures to investigate vocal auditory processing in ASD, and found delayed P100 ERP components bilaterally and delayed N100 ERP components in the right hemisphere only. These findings suggest atypicalities in the temporal dynamics of early auditory processing for vocal sounds, although significant heterogeneity across studies was observed. A study by Schelinski and von Kriegstein (2019) also highlighted that auditory processing differences in ASD tend to be more pronounced for vocal than for non-vocal stimuli, indicating that difficulties might emerge primarily when acoustic information carries social meaning. These results have been interpreted as reflecting a reduced sensitivity to socially relevant, vocal information accompanied by a bias toward non-vocal sounds.

2.3.1. Emotion Processing in ASD

Research on emotion processing in autism spectrum disorder has predominantly focused on the visual domain, particularly facial expression recognition. Across a wide range of paradigms, including images or videos, and interactive tasks, individuals with ASD have frequently been shown to exhibit reduced accuracy and slower processing of facial emotions compared to neurotypical controls. Meta-analytic evidence supports the presence of a general emotion recognition impairment in ASD within the visual domain, encompassing facial expressions and body gestures (Uljarevic & Hamilton, 2013).

However, fewer studies have investigated emotion processing in non-visual modalities, and findings across domains are less consistent. Cross-domain meta-analyses suggest that emotion processing difficulties in ASD are not uniform, but depend on stimulus modality, emotion category, and age (Leung et al., 2022). Importantly, impairments appear more robust for socially salient human stimuli (e.g., faces, speech) than for non-social or abstract stimuli (e.g., cartoons, tones), which highlights the importance of domain-specific investigations (Leung et al., 2022).

2.3.2. Vocal Emotion Processing in ASD

While research on emotion processing in the visual domain provides relatively consistent evidence for impairments in individuals with ASD, findings in the auditory domain are considerably more heterogeneous. A substantial number of studies report deficits in the recognition of emotional vocal stimuli in individuals with ASD across different age groups and task paradigms, suggesting reduced accuracy or altered sensitivity to affective cues in vocalisations (Day et al., 2024; Globerson et al., 2015b; Golan et al., 2007; Heaton et al., 2012; Korpilahti et al., 2007; Lindner & Rosén, 2006; Peppé et al., 2007; Philip et al., 2010; Schelinski & von Kriegstein, 2019).

In contrast, several other studies have reported comparable performance between autistic and typically developing (TD) individuals on vocal emotion recognition tasks (Baker et al., 2010; Boucher et al., 2000; Brennand et al., 2011; Grossman et al., 2010; Jones et al., 2011). Such findings challenge the notion of a global deficit in affective auditory processing and instead point toward substantial inter-individual variability within the autism spectrum. In line with this interpretation, Lerner et al. (2013) propose that difficulties in vocal emotion recognition may not characterize the ASD population as a whole, but rather a specific subgroup of individuals. This perspective underscores the importance of considering

heterogeneity within ASD, as well as task characteristics, stimulus types when evaluating affective auditory processing.

Globerson et al. (2015) reported reduced recognition of basic emotional prosody in individuals with ASD and demonstrated that vocal emotion recognition was closely linked to psychoacoustic abilities, or specifically, pitch discrimination ability, particularly strongly in the ASD group. Apart from the impaired emotion recognition in the ASD group, the pitch discrimination ability were intact. Also, they found the association between psychoacoustic abilities and emotion recognition not to be moderated by group. The authors therefore argued that group differences in affective vocal processing may reflect impairments in higher-order emotional processing mechanisms rather than deficits in basic auditory perception.

In contrast, Schelinski and von Kriegstein (2019) reported both impaired vocal emotion recognition and reduced vocal pitch discrimination ability in individuals with ASD, alongside intact non-vocal pitch processing. Notably, correlations between vocal pitch discrimination and vocal emotion recognition were present only in the neurotypical group, but not in the ASD group. The authors concluded that individuals with ASD may be less able to utilize vocal pitch information when interpreting emotional prosody, suggesting that vocal emotion recognition deficits in ASD may, at least partly, be rooted in atypical sensory-level processing rather than exclusively higher-order social-cognitive mechanisms.

Meta-analytic evidence further shows the complexity of affective auditory processing in ASD. Leung (2022) conducted a cross-domain meta-analysis on emotion recognition in ASD and reported overall impairments across visual and auditory modalities. Moderator analyses revealed age as a significant factor, with larger group differences observed in older participants. Importantly, emotion-specific analyses within the auditory domain showed mixed results: reduced recognition accuracy in the ASD group was found for angry, happy, and disgusted vocal stimuli, whereas no group differences were observed for fearful or sad stimuli. Similarly, Zhang et al. (2022) conducted a systematic review and meta-analysis focusing specifically on vocal emotion recognition in autism. While an initial moderate-to-large group difference between ASD and neurotypical individuals was observed, this effect was no longer significant after correcting for publication bias. Moreover, group differences varied across emotion categories, with the largest effect observed for surprise and a significant difference found only for happiness among the remaining basic emotions. No reliable group differences emerged for fearful, sad, or angry vocal stimuli. The authors did not identify significant moderating effects of age or gender; however, they cautioned that the strong male bias in the available data limits firm conclusions regarding gender-related effects.

Beyond performance differences, there is also evidence that individuals with ASD may rely on different cues when interpreting vocal emotions. For example, autistic listeners have been shown to rely more strongly on verbal content or contextual information, rather than emotional prosody itself, to infer a speaker's emotional state (Leung et al., 2022; Lindner & Rosén, 2006; Le Sourn-Bissaoui et al., 2013; Stewart et al., 2013). This finding is particularly relevant for studies employing brief, non-verbal emotional vocalisations, in which contextual and linguistic cues are absent and reliance on acoustic affective information is required.

To sum up, the literature suggests that vocal emotion recognition in ASD cannot be straightforwardly characterized as uniformly impaired. Instead, the mixed findings likely reflect a complex interaction between individual differences within the autism spectrum, task demands, stimulus characteristics, and the relative contribution of basic perceptual versus higher-order cognitive mechanisms. This heterogeneity underscores the need for further systematic investigations of affective auditory processing in ASD, particularly studies that carefully control stimulus properties and isolate specific acoustic cues underlying emotional perception.

Neurophysiological Measures

While most research on autism and vocal emotion perception has been conducted using behavioural measures, typically in form of vocal emotion recognition tasks, a smaller body of work has investigated the underlying neural mechanisms. Neurophysiological methods such as EEG as well as fMRI, offer tools for identifying and quantifying the temporal and functional characteristics of perceptual and cognitive processes involved in affective auditory processing.

Evidence from fMRI studies suggests atypical neural processing of vocal emotional stimuli in individuals with ASD. Specifically, reduced activation has been reported in brain regions implicated in vocal emotion processing, including the superior temporal sulcus, as well as the insula and amygdala (Gervais et al., 2004; Rosenblau et al., 2016). These findings point toward altered neural responses to affective vocal information at the level of distributed emotion-processing networks.

EEG studies

EEG studies have provided more insights into the temporal dynamics of auditory affect processing in ASD, though results remain mixed. A recent study by Day et al. (2024)

reported impaired vocal emotion recognition performance in adolescents with ASD on behavioural rating tasks; however, event-related potential (ERP) responses, including the N100, P200, and late positive potential (LPP), did not differ significantly between participants with ASD and neurotypical individuals. The authors interpreted the absence of N100 differences as evidence for intact early auditory detection in individuals on the autism spectrum, given that the N100 is typically associated with low-level auditory perception and sensitivity to acoustic features such as pitch (Schirmer & Kotz, 2006) Moreover, the relationship between LPP amplitude and vocal emotion recognition performance, together with the finding that social cognition measures predicted emotion recognition ability, suggested that higher-order cognitive processes may play a central role in modulating vocal emotion recognition rather than early perceptual mechanisms.

In contrast, other EEG studies have reported alterations at earlier stages of auditory processing. Korpilahti et al. (2007) observed prolonged N100 latencies during vocal emotion recognition tasks in boys with ASD. Similarly, Lerner and colleagues (2013) found a negative correlation between N100 latency and vocal emotion recognition performance in adolescents with ASD, such that longer latencies were associated with poorer performance. No corresponding associations were observed for later amplitudes or latencies, that are usually connected with affect processing. Based on these findings, the authors argued that vocal emotion recognition difficulties in ASD may stem from atypicalities at very early stages of social perception, rather than exclusively from higher-order cognitive processes.

To sum up, neurophysiological findings on auditory affect processing in ASD present a heterogeneous pattern. While some studies suggest intact early auditory processing with impairments emerging at later, higher-order stages, others point toward atypicalities at early perceptual levels or compensatory attentional mechanisms. This inconsistency mirrors the mixed behavioural literature and underscores the need for further EEG studies that systematically disentangle early sensory processing, attentional modulation, and higher-order cognitive contributions to affective auditory processing in autism.

3. Research Gaps and Hypotheses

3.1. Research Gaps

What is based on current research known?

While research on visual emotion processing provides fairly robust evidence for atypicalities in ASD, findings in the auditory domain remain inconsistent.

Behavioural studies investigating vocal emotion processing in ASD predominantly report impairments in the recognition of affective vocalisations, whereas other studies describe intact performance comparable to neurotypical individuals. Meta-analytic evidence suggests small and emotion-specific effects alongside substantial heterogeneity across studies (Leung et al., 2022; Zhang et al., 2022).

Evidence from EEG studies is comparatively sparse and similarly inconclusive. Some studies report intact early auditory processing, suggesting that differences in affective processing may arise at later, higher-order stages (e.g., Day, 2024). Other studies indicate early-stage atypicalities, such as altered N100 or P200 amplitudes or latencies (e.g., Lerner, 2013), while still others interpret early ERP differences as reflecting compensatory attentional mechanisms rather than primary perceptual deficits. These inconsistencies may be partly attributable to differences in task demands, stimulus characteristics, and the extent to which linguistic or contextual cues allow compensatory processing.

By using a combination of behavioural and neuropsychological methods, this study aims to contribute to the current state of research in vocal affect processing, and to help understand the following unresolved research gaps better:

RESEARCH GAP 1: Where exactly does auditory emotion processing differences arise?

Behavioural measures alone are not able to disentangle whether atypical emotion recognition performance in ASD arises from early sensory-perceptual mechanisms, attentional modulation, or later, higher-order interpretive processes. The P200 component, occurring approximately 180–240 ms after stimulus onset, is associated with early affective encoding of auditory signals (Schirmer & Kotz, 2006) as well as attentional allocation to emotionally salient sounds (Stefanou et al., 2024) and occurs before conscious interpretation, labeling of or reasoning about emotion. By linking P200 amplitude and latency to continuous intensity ratings, the present study aims to capture early perceptual–attentional dynamics underlying subjective emotion perception in ASD.

RESEARCH GAP 2: The role of stimulus type (nonverbal, brief vocalisations)

Many studies on vocal affect processing rely on speech prosody or relatively long auditory stimuli, which may allow participants to draw on top-down strategies such as semantic interpretation or contextual inference. Evidence suggests that individuals with ASD may rely more strongly on such contextual cues during auditory affect processing than neurotypical individuals (Sourn-Bissaoui et al., 2013; Stewart et al., 2013). Consequently, it remains unclear how emotional vocal signals are processed when linguistic and contextual information is minimized. Specifically, it is an open question whether early neural responses to emotional sounds remain intact under conditions that restrict top-down compensation.

By employing short, nonverbal emotional vocalisations, the present study aims to isolate early affective auditory processing and minimize higher-order interpretive strategies, thereby providing a more direct test of early-stage neural processing in ASD.

RESEARCH GAP 3: How are behavioural and neural measures related?

A further question concerns the relationship between neural and behavioural measures of auditory emotion processing in ASD. While a small number of EEG studies have examined associations between early ERP components and behavioural performance, the available evidence remains limited and inconsistent. Existing studies differ with respect to stimulus type, task demands, behavioural outcome measures (e.g., accuracy versus subjective ratings), and the ERP components examined, leading to mixed conclusions about the functional relevance of early neural responses for emotion perception.

As a result, it remains unclear whether variability in early electrophysiological responses, such as P200 amplitude and latency, systematically covaries with subjective emotional experience within individuals, and whether such brain–behaviour relationships differ between individuals with and without ASD. Investigating the coupling between ERP measures and behavioural intensity ratings allows for a investigation of the functional role of early auditory processing stages. Similar coupling patterns across groups would suggest that comparable neural mechanisms support emotion perception in individuals with and without ASD. In contrast, group-specific coupling patterns would point toward ASD specific atypicalities in how early neural responses are integrated into subjective emotional experience.

3.2. Hypotheses

Based on the existing literature and the research gaps outlined above, the following hypotheses of this study are as follows:

3.2.1. Behavioural Hypotheses

H1a – Accuracy

Participants with ASD will show lower accuracy in identifying the intended emotion of nonverbal vocal stimuli compared to neurotypical participants.

H1b – Intensity Ratings

Participants with ASD will rate emotional vocalisations as less intense than neurotypical participants.

H1c – Emotion-specific differences

Group differences in behavioural performance are expected to vary across emotion categories.

3.2.2. ERP Hypotheses

Due to inconsistent findings regarding early ERP components in auditory emotion processing in ASD, hypotheses concerning P200 amplitude and latency are formulated in a non-directional manner.

H2a – Amplitude

The amplitudes of P200 ERP components are expected to differ between the ASD and NT group.

H2b – Latency

The latencies of P200 ERP components are expected to differ between groups.

RQ2 – Emotion Differences

Do amplitudes and latencies of P200 differ between emotion categories?

3.2.3. Brain – Behaviour Relationship

As the existing literature does not provide clear or consistent evidence regarding the nature or direction of the relationship between early ERP components and auditory emotion recognition in autistic versus neurotypical individuals, this study formulates an exploratory

research question rather than a confirmatory hypothesis.

RQ3 – Brain-Behaviour Coupling

Does the relationship between early electrophysiological markers of affective auditory processing (P200 amplitude and latency) and perceived emotional intensity ratings differ between participants with ASD and neurotypical participants?

4. Methods

This thesis was part of a larger research project, *Auditory Emotion Processing in Autism* (AEPA), conducted at the University of Oslo. This larger project investigates how individuals with ASD and neurotypical individuals process emotional information in vocal and musical stimuli using behavioural and electrophysiological measures. While the broader study included various tasks, stimuli, and questionnaires, only a subset relevant to vocal emotional processing is analyzed in this thesis. In the following section (Study Procedure), the complete procedure and material will be briefly described for the sake of completeness.

4.1. Study Procedure

The Guidelines for Conducting Research Studies with the Autistic Community (Gowen et al., 2019) were used as a reference for the planning of design and procedure of the experimental setup. These include, for example, sending participants a comprehensive overview of the study ahead of time, including both written and visual explanations of each step, allowing them to prepare and feel more at ease during the experiment. The schedule also included several opportunities for breaks and the availability of snacks and drinks between experimental blocks.

Upon arrival, participants were welcomed and guided through the first block of the experiment, which involved reading and signing the informed consent form. This form included a detailed explanation of the study's goals, procedures, and the participant's right to withdraw at any point without any negative consequences. Following consent, participants completed a brief cognitive assessment using the Matrix Reasoning subtest of the Wechsler Adult Intelligence Scale (WAIS-IV; Wechsler, 2008).

The next block consisted of the EEG session. During this phase, participants listened to randomized sequences of emotional vocal and musical stimuli while performing a simple auditory 1-back task. In this task, they were instructed to press a button whenever the same sound was presented consecutively. This was done to ensure that participants maintained attentional focus during the passive listening task The recording was conducted in a sound attenuated, dimly lit laboratory to minimize distractions. For this thesis, only the vocal stimuli data were analyzed.

Following the EEG session, participants completed an emotional rating task in which they rated each stimulus on four separate 10-point Likert scales reflecting the perceived emotional intensity for the emotion categories happy, sad, angry and fearful. The stimuli used

in the rating task were identical to those presented during EEG acquisition but shown in randomized order.

Finally, participants filled out five questionnaires: the Autism Spectrum Quotient (AQ, Baron-Cohen et al., 2001), the Toronto Alexithymia Scale (TAS-20, Bagby et al., 2020), the State-Trait Anxiety Inventory (STAI, Spielberger, 2012), the Goldsmiths Musical Sophistication Index (Gold-MSI, Müllensiefen et al., 2014), and the Sensory Perception Quotient (SPQ-35, Tavassoli et al., 2014).

The entire experimental session lasted approximately 2 to 3 hours, depending on the length of the breaks taken by each participant. The study was ethically approved and was conducted in accordance with the declaration of Helsinki.

4.2. Stimuli

The vocal stimuli used in this study were selected from the VENEC corpus**,** a validated database of non-linguistic emotional vocal expressions (Laukka et al., 2010) and are categorized into four emotional expressions: happy, fearful, angry, and sad. For the purposes of this study, all stimuli were standardized to a duration of exactly 1 second. The final set of stimuli consisted of 20 sounds, with five stimuli per emotion category.

4.3. EEG Recording

Stimuli were presented via two loudspeakers placed to the left and right of a monitor, at a comfortable listening distance. Ahead of the task, participants were instructed to sit still and to try to minimize movements and eye blinks during the EEG recording in order to reduce muscular and ocular artefacts.

The EEG signal was continuously recorded at a 2048 Hz sampling rate using a BioSemi ActiveTwo system (BioSemi, n.d.)with 64 active electrodes mounted on an electrode cap following the 10-20 international system.

Additionally, six external electrodes placed on the participant's ears, front and under their eyes were used to monitor electromyographic and electrooculographic activity, to allow for later identification of artefacts such as muscle activity or eye movements.

4.4. Behavioural Task

Following the EEG session, participants completed a behavioural task on a computer where they listened to the same 20 vocal stimuli that had been presented during the EEG recording. The sounds were played in a randomized order to minimize sequence effects. For

each sound, participants were asked to rate the extent to which it expressed four emotions using four separate slider scales. Each scale ranged from 0 ("Not at all") to 9 ("Very much so") and was phrased as: ″This person sounds … (“sad”, angry, happy, and fearful").

4.5. Questionnaire

AQ-50

To measure the self-reported degree of autism, we implemented the adult autism spectrum quotient (AQ-50, Baron-Cohen et al., 2001). The questionnaire consists of three subscales (social interaction and spontaneity, imagination and creativity, communication and reciprocity). For the purposes of this thesis, only the total score was used for analysis, since the goal was to use the scale for an overall estimate of the participant’s level autistic traits. Although the AQ has been subject to criticism regarding its factor structure and sensitivity, it remains one of the most widely used and validated self-report measures of autistic traits in adults (English et al., 2020). A score above the threshold of 26 has been suggested to indicate possible autism (Woodbury-Smith et al., 2005).

Group assignment in the present thesis was based on prior clinical diagnosis (ASD) or absence thereof (NT) and was not determined by AQ scores. The AQ was used descriptively to characterize autistic trait levels within each group. Participants were therefore not excluded based on AQ cutoff values, consistent with the dimensional conceptualization of autistic traits and the use of the AQ as a screening instrument rather than a diagnostic tool (Baron-Cohen et al., 2001).

4.6. Participants

Recruitment

The recruitment process involved the collaboration with several ASD associations in and around Oslo, Norway. Also, recruitment took place in several social media groups, some of them specifically made by and for neurodivergent individuals. We also worked with posters hung up at public places, mostly university buildings.

Inclusion / Exclusion Criteria

Inclusion criteria for all participants were normal hearing, normal or corrected-to-normal vision, and the ability to speak and understand English or one of the Scandinavian languages. For the ASD group, a formal diagnosis of ASD, according to the DSM-5, made by a licensed clinician or psychologist, was required. Given the high comorbidity rates

associated with ASD, certain additional psychiatric conditions were permitted, specifically Attention-Deficit/Hyperactivity Disorder (ADHD) and/or Anxiety Disorder. For the neurotypical group, inclusion required a generally healthy condition, with no history of neurological or psychiatric disorders.

Participants

The final sample consisted of N = 42 participants, including 21 individuals diagnosed with ASD and 21 neurotypical controls (NT). Participants in the NT group were matched to the ASD group with respect to age and gender. The autism group comprised 17 females and the comparison group comprised 16 females. The predominance of female participants was not intentional but occurred by chance during recruitment. The mean age of the total sample was 28.48 years. Participants completed the study in either English or Norwegian, depending on their language preference. All participants received a gift voucher as financial compensation (500 NOK, approximately 40 EUR).

5. Data Processing and Analysis Plan

5.1. Behavioural Data

The behavioural data were prepared and analyzed using RStudio (Posit team, 2025). From the results of the emotion rating task two types of scores were computed, each capturing a distinct aspect of emotional perception: 1) Accuracy Rate – indicating how accurately participants identified the intended emotion in each stimulus and 2) Intensity Rating – reflecting the perceived strength of the expressed emotion.

Accuracy Rate

Accuracy Rate was defined as the proportion of emotion items for which a participant correctly identified the target emotion. A response was classified as accurate if the participant gave a higher rating on the correct emotional scale than on any of the three incorrect scales for that item. The accuracy rate was computed as the number of correctly identified items divided by the total number of items (20), yielding a score from 0 (no items accurately recognized) to 1 (all items accurately recognized). In addition to the overall accuracy rate, accuracy rates were also computed separately for each emotion category. To avoid bounded zero and one values, accuracy rates were first adjusted using the method described by Smithson and Verkuilen (2006):

$$p_{adj} = \frac{\mathrm{y(n-1)} + 0.5}{n}$$

where y is the accuracy rate and n the number of trials per participant. The adjusted proportions were then logit-transformed to stabilize variances and better approximate normality:

$$logit(p_{adj}) = \ln\left(\frac{p_{adj}}{1 - p_{adj}}\right)$$

Intensity Ratings

The Intensity Rating score was calculated as the mean rating on the correct emotion scale for each item. This measure reflects the perceived emotional strength of the target emotion in the stimulus—regardless of whether it was accurately identified. Both an overall mean intensity score across all stimuli and separate mean scores per emotion category were computed, resulting in a score of minimum 0 and maximum 9.

5.2. EEG Data

Preprocessing of the EEG data was carried out using MATLAB (The MathWorks Inc., 2024) and the toolboxes EEGLab (Version 2022.1; Delorme & Makeig, 2004) and FieldTrip (Oostenveld et al., 2011). Initially, the data was downsampled to 256 Hz and re-referenced to the average of all scalp electrodes**.** Then a notch filter (47–53 Hz) was applied to remove line noise, followed by a bandpass filter (1-30 Hz) to retain frequency ranges relevant to cognitive ERP components. Next, bad channels were identified and interpolated using spherical interpolation, based on visual inspection. The datasets were then segmented into epochs from -300 ms to 1000 ms relative to stimulus onset and baseline-corrected using the 300 ms pre-stimulus interval.

To remove artifacts, an Independent Component Analysis (ICA) was then performed, followed by manual rejection of components associated with eye blinks, eye movements, and muscle artifacts. After ICA cleaning, data were again bandpass filtered (1–30 Hz).

Finally, the data was epoched separately for each emotion condition (happy, angry, fearful, sad) based on stimulus markers. For further analysis, single-electrode time courses were exported as .csv files.

Extraction of ERP Amplitudes and Latencies, Outlier Detection and Imputation

Event related potentials (ERPs) were extracted from the single electrode data for each participant and condition. The analysis focused on the P200 component as prior literature has shown this ERP component to be a relevant marker for (vocal) auditory emotion processing (e.g. Schirmer & Kotz, 2006; Pell et al., 2015). Consistent with previous research indicating that the auditory P200 is maximal at fronto-central scalp sites (e.g. Liu et al., 2012; Pell et al., 2015), a cluster of fronto-central electrodes (F1, Fz, F2, FC1, FCz, FC2, C1, Cz, and C2) was used. For each participant and condition, signals from these electrodes were averaged to derive a representative waveform for peak detection. To extract the peak amplitudes and latencies of the P200 component, a semi-automated peak detection approach was applied. Within each ERP waveform, local maximas were identified within a predefined time window of 180–240 ms post-stimulus onset, corresponding to the typical latency of the auditory P200. Detected peaks were then verified by visual inspection.

To reduce the influence of extreme values, outlier detection was conducted using Cook's distance (see next subchapter). Outlier values were then replaced via mean imputation, using the mean of the corresponding group × emotion cell.

5.3. Outlier Cleaning and Imputation

For both behavioural and electrophysiological data, the same systematic approach was used to screen for influential observations to ensure that the results were not biased by extreme data points. First, an ordinary least squares (OLS) proxy model was fitted for each dependent variable using the formula:

$$variable \sim group \times emotion$$

Subsequently, Cook's distance was computed for each observation (Cook, 1977) to assess its influence on the fitted model. Observations with values exceeding the threshold of $D_i > 4/N$ (with N representing the number of rows in the respective long-format dataset) were flagged as potential outliers and inspected. To keep a balanced design for the factorial analyses, missing values due to removed outliers were then imputed by the mean value of the corresponding group × emotion cell. The resulting imputed dataset was used for all inferential statistical analyses. Detailed tables listing flagged outliers as well as comparisons of raw versus cleaned datasets are presented in the Appendix C.

5.4. Statistical Analysis

Statistical Assumptions

Prior to conducting the factorial analyses, relevant statistical assumptions were assessed. The normality of residuals was evaluated through visual inspection of histograms and Q–Q plots, and then tested using the Shapiro–Wilk test (Shapiro & Wilk, 1965). Homogeneity of variances was assessed using Levene's test (Levene, 1960). For within-subject factors, in case of violations of the sphericity assumption, Greenhouse–Geisser corrections (Greenhouse & Geisser, 1959) were applied to adjust degrees of freedom and control for inflated Type I error. Where assumptions were substantially violated, robust alternatives were employed.

Statistical Testing

To assess possible group differences (NT vs. ASD) and emotion-related effects (happy, angry, fearful, sad) in the behavioural measures (accuracy scores and intensity ratings) and ERP measures (amplitude and latency), two-way mixed design ANOVAs were conducted, with group as a between-subjects factor and emotion as a within-subjects factor.

This approach allows the simultaneous testing of main effect of group, main effect of emotion, as well as emotion x group interactions, thus allowing for examining whether

emotional processing differs between groups and whether groups differ consistently across all emotions. Where significant main or interaction effects involving the Emotion factor were found, Holm-adjusted post hoc pairwise comparisons were performed to control for multiple testing while preserving statistical power.

Correlation Analysis for behavioural ratings and ERP data

To assess possible associations between the behavioural ratings and the electrophysiological measures (P200 amplitude and latency) the following two approaches were conducted:

First, repeated-measures correlations (rmcorr; (Bakdash & Marusich, 2017) were calculated for both groups to quantify the relationship between ratings and ERP measures across all four emotions within individuals. This approach takes into account the non-independence of measurements within an individual and is therefore suitable for within-subject designs.

Next, emotion specific Pearson correlations were assessed, where for each of the four emotions (happy, angry, fearful, sad), the association between ratings and P200 measures was calculated. To account for multiple testing in the Pearson correlations (four emotions × two ERP parameters = eight tests), p-values were corrected using the Benjamini–Hochberg false discovery rate (FDR) procedure (Benjamini & Hochberg, 1995). The FDR approach controls the expected proportion of false positives among all significant findings and is therefore well suited for exploratory analyses involving several correlated tests.

6. Results

6.1. AQ Questionnaire

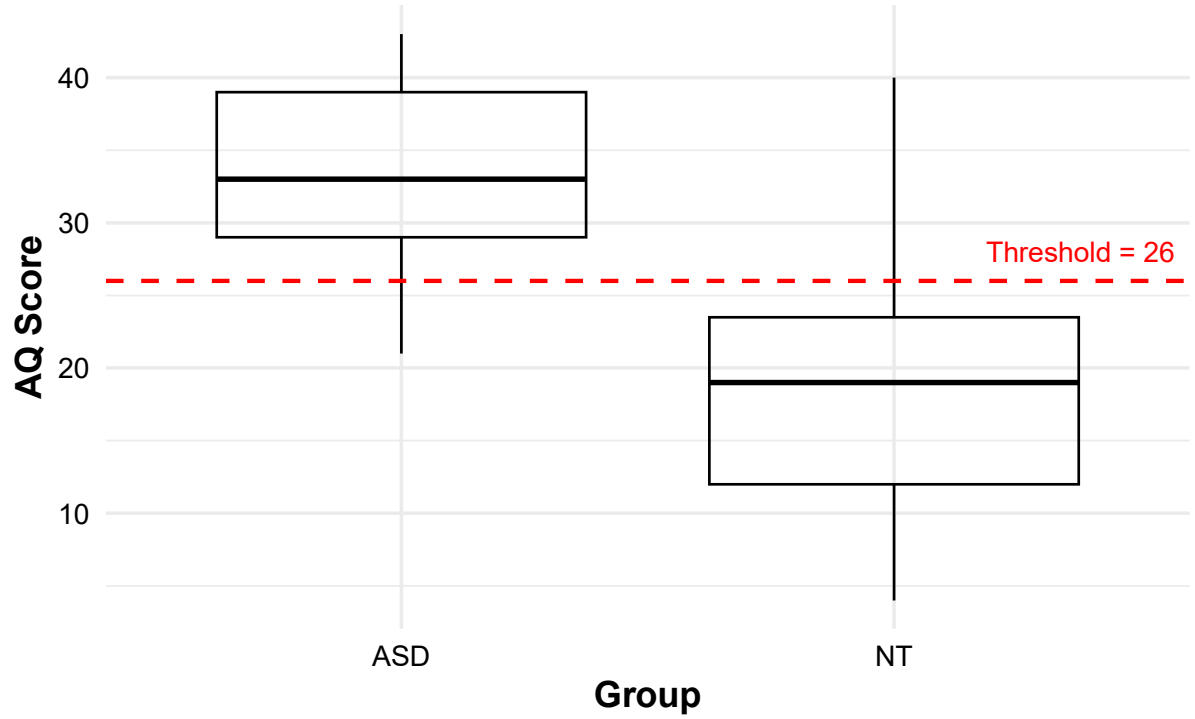
The Autism Spectrum Quotient (AQ) was administered to validate group differences in autistic traits. Descriptive statistics indicated substantially higher AQ scores in the ASD group (M = 33.30, SD = 6.45) compared to the neurotypical (NT) group (M = 19.12, SD = 9.06; see Figure 1). In the ASD group, 18 of 21 participants (85.7%) scored at or above the commonly used AQ threshold of 26 (Woodbury-Smith et al., 2005), whereas this was the case for 5 of 21 participants (23.8%) in the NT group.

An independent-samples t-test confirmed that AQ scores were significantly higher in participants with ASD than in neurotypical participants, $t(36.26) = 5.83$, $p < .001$, with a large effect size ($d = 1.80$, 95% CI [1.07, 2.51]). Assumptions of normality and homogeneity of variances were met (see Appendix B, Table 8).

As the AQ measure was administered to characterize autistic traits descriptively within each group, no participants were excluded based on AQ scores.

Figure 4

Boxplot of AQ Scores by Group



Note. Boxplots display AQ scores by group (NT = neurotypical participants, ASD = participants with ASD) and include the median, 25th and 75th percentiles, and 1.5 × the interquartile range. Scores of 26 or above indicate possible autism.

6.2. Behavioural Data

Accuracy Rates

Following data cleaning, 6 of 168 observations (3.6%) exceeded the Cook's distance threshold, were flagged as influential and replaced using mean imputation within the corresponding group × emotion cell. Detailed information on flagged outliers and comparisons between raw and imputed data is provided in Appendix C (Table 9 Table 10).

Descriptive statistics indicated lower overall accuracy in the ASD group (M = 1.38, SD = 0.83) compared to the neurotypical group (M = 1.65, SD = 0.74). Descriptive statistics by group and emotion are presented in Table 2, and Figure 1 shows boxplots of accuracy rates by group.

Table 2

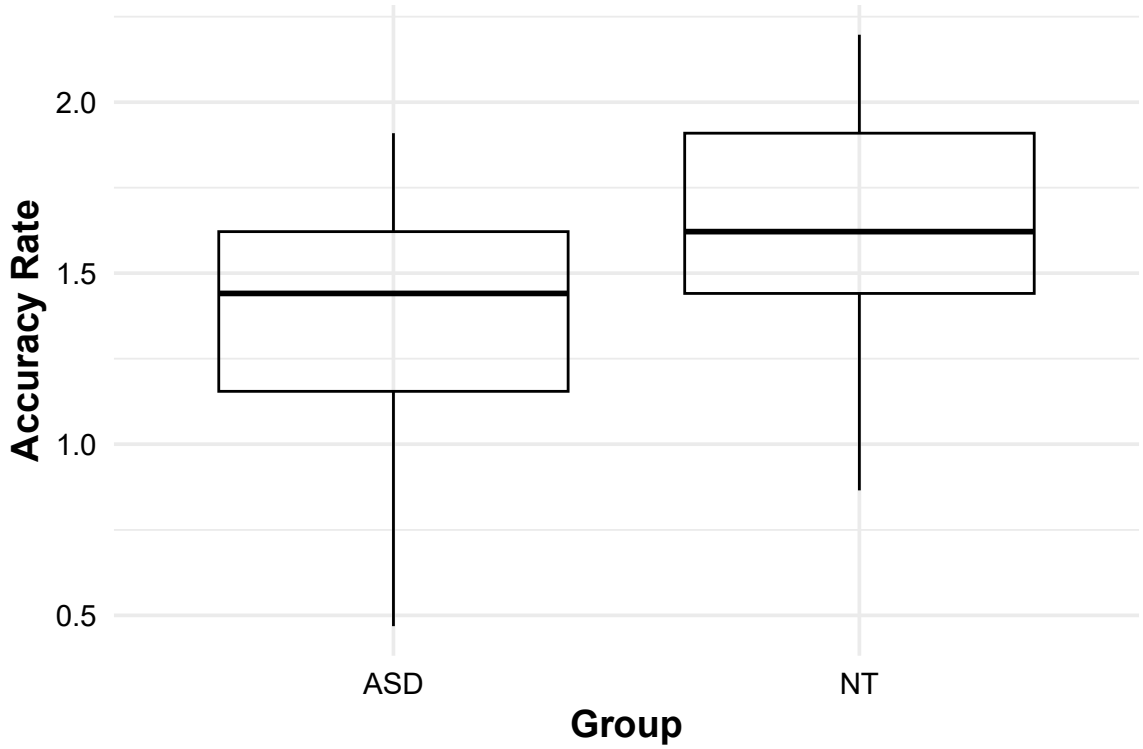
Descriptives of Accuracy Rates by Group and Emotion

Emotion	Group	M	SD	95% CI
Angry	ASD	1.05	0.79	[0.69, 1.41]
	NT	1.25	0.87	[0.85, 1.65]
Fearful	ASD	0.95	0.86	[0.56, 1.34]
	NT	1.72	0.68	[1.41, 2.03]
Happy	ASD	1.84	0.65	[1.54, 2.14]
	NT	1.78	0.66	[1.48, 2.08]
Sad	ASD	1.66	0.68	[1.35, 1.97]
	NT	1.83	0.61	[1.55, 2.11]

Note. Accuracy rates are logit-transformed. M = mean, SD = standard deviation, CI = confidence interval

Figure 5

Boxplot of Accuracy Rates by Group



Note. Boxplots display accuracy rates by group and include the median, 25th and 75th percentiles, and 1.5 × the interquartile range.

Because accuracy rates violated the assumption of normality, an aligned rank transform (ART) ANOVA was conducted with Group (ASD, NT) as a between-subjects factor and Emotion (angry, fearful, happy, sad) as a within-subjects factor. The analysis revealed a significant main effect of Group, a significant main effect of Emotion, and a significant Group × Emotion interaction (see Table 3).

Table 3

ANOVA Results for Accuracy Rates

Effect	F	df	Df.res	*p*
group	10.33	1	40	.0026 **
emotion	8.83	3	120	<.0001 ***
group × emotion	3.62	3	120	.0152 *

Note. Aligned rank transform (ART) ANOVA was used. Accuracy rates are logit-transformed.

Follow-up aligned-rank contrast tests with Holm-adjusted p-values indicated that participants with ASD showed significantly lower accuracy than neurotypical participants for fearful vocalisations only (see Table 4). No significant group differences were observed for angry, happy, or sad stimuli. Pairwise emotion contrasts within each group are reported in Appendix D, Table 17. Figure 6 illustrates accuracy rates across emotion categories for both groups.

Table 4

Contrasts of Accuracy Rates Between Groups by Emotion

contrast	emotion	estimate	SE	p.value	95% Confidence Interval for Difference		Cohen's d
					Lower Bound	Upper Bound	
ASD - NT	a	-9,714	12,781	1,000	-34,96	15,53	-0,120
	f	-40,810	12,781	0,039	-66,05	-15,57	-0,506
	h	3,524	12,781	1,000	-21,72	28,77	0.044
	s	-10,810	12,781	1,000	-36,05	14,43	0.134

Note. Estimated marginal mean differences are based on aligned-rank contrasts (ART-C). Negative estimate values indicate lower accuracy in the ASD group.

Figure 6

Line Plot of Accuracy Rates by Group and Emotion

Note. The figure shows mean accuracy rates with standard deviation error bars for each emotion category (a=angry, f=fearful, h=happy, s=sad), separately for ASD and NT participants.

Intensity Ratings

Outlier screening identified 12 of 168 observations (7.1%) that exceeded the Cook's distance threshold and were flagged as influential. These observations were replaced via group × emotion mean imputation to preserve a balanced factorial design. Detailed information on flagged outliers and comparisons between raw and imputed data is provided in Appendix C (Table 11&Table 12).

Descriptive statistics indicated that neurotypical participants rated emotional vocalisations as more intense overall (M = 6.29, SD = 1.56) than participants with ASD (M = 5.35, SD = 1.48). Descriptive statistics by group and emotion are provided in Table 5, and Figure 7 illustrates boxplots of intensity ratings by group.

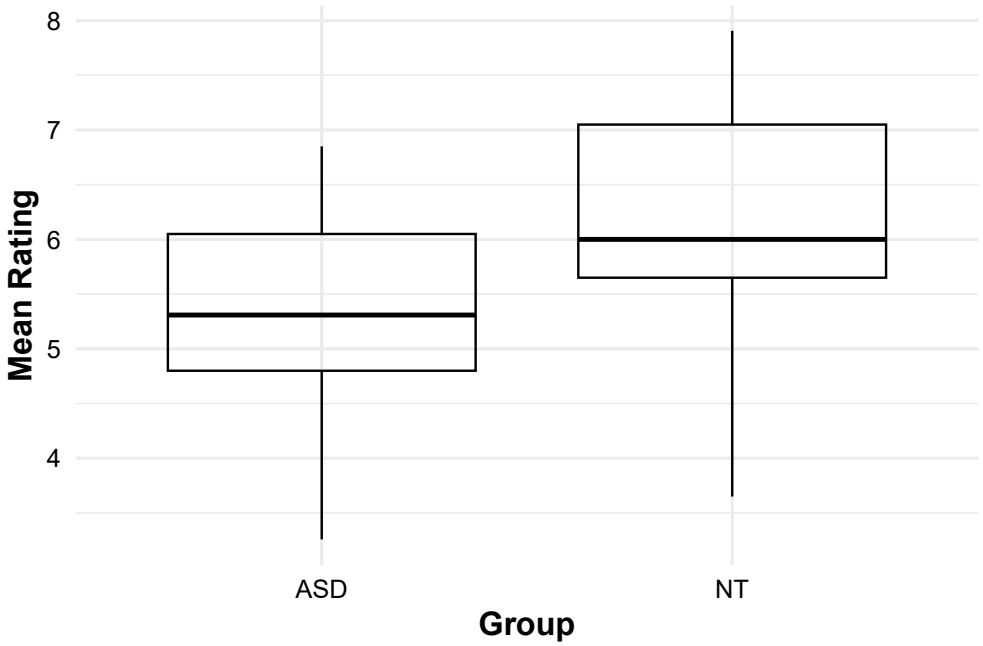
Table 5

Descriptive Statistics of Intensity Ratings by Group and Emotion Category

Emotion	Group	M	SD	Min	Max	95% CI
Angry	ASD	4.40	1.39	0.8	6.4	[3.77, 5.03]
	NT	5.30	1.66	2.2	7.8	[4.54, 6.06]
Fearful	ASD	5.44	1.16	2.4	6.8	[4.91, 5.97]
	NT	6.53	1.41	3	9	[5.89, 7.17]
Happy	ASD	6.02	1.09	3.6	8	[5.52, 6.52]
	NT	6.62	1.29	4.4	8.6	[6.03, 7.21]
Sad	ASD	5.53	1.78	2.4	9	[4.72, 6.34]
	NT	6.71	1.53	3.8	9	[6.01, 7.41]

Figure 7

Boxplot of Mean Intensity Ratings by Group



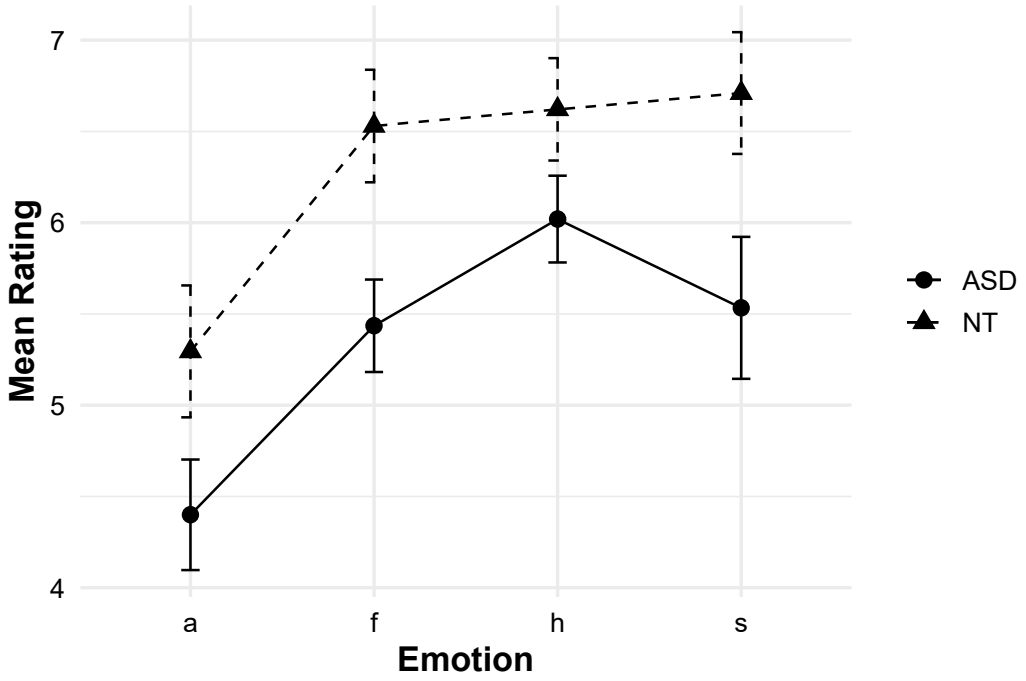
Note. Boxplots display accuracy rates by group and include the median, 25th and 75th percentiles, and 1.5 × the interquartile range.

Assumptions of normality and homogeneity of variances were met (see Appendix B, Table 8). A two-way mixed ANOVA with Group (ASD, NT) as a between-subjects factor and Emotion (angry, fearful, happy, sad) as a within-subjects factor revealed a significant main effect of Group, $F(1, 40) = 9.24$, $p = .004$, $\eta^2_p = .188$, indicating higher intensity ratings in the NT group. A significant main effect of Emotion was also observed, $F(2.91, 116.21) = 13.29$, $p < .001$, $\eta^2_p = .249$, reflecting differences in perceived intensity across emotions. The Group × Emotion interaction was not significant, $F(2.91, 116.21) = 0.50$, $p = .678$.

Exploratory Holm-adjusted post hoc comparisons indicated that group differences were significant for fearful and sad vocalisations, with participants with ASD providing lower intensity ratings than neurotypical participants. No significant group differences were observed for happy or angry stimuli. Detailed post hoc comparisons are reported in Appendix D, Table 18 - Table 20. Figure 8 illustrates intensity ratings across emotion categories for both groups.

Figure 8

Line Plot of Intensity Ratings by Group and Emotion



Note. The figure shows mean subjective intensity ratings with standard deviation error bars for each emotion category, separately for ASD and NT participants.

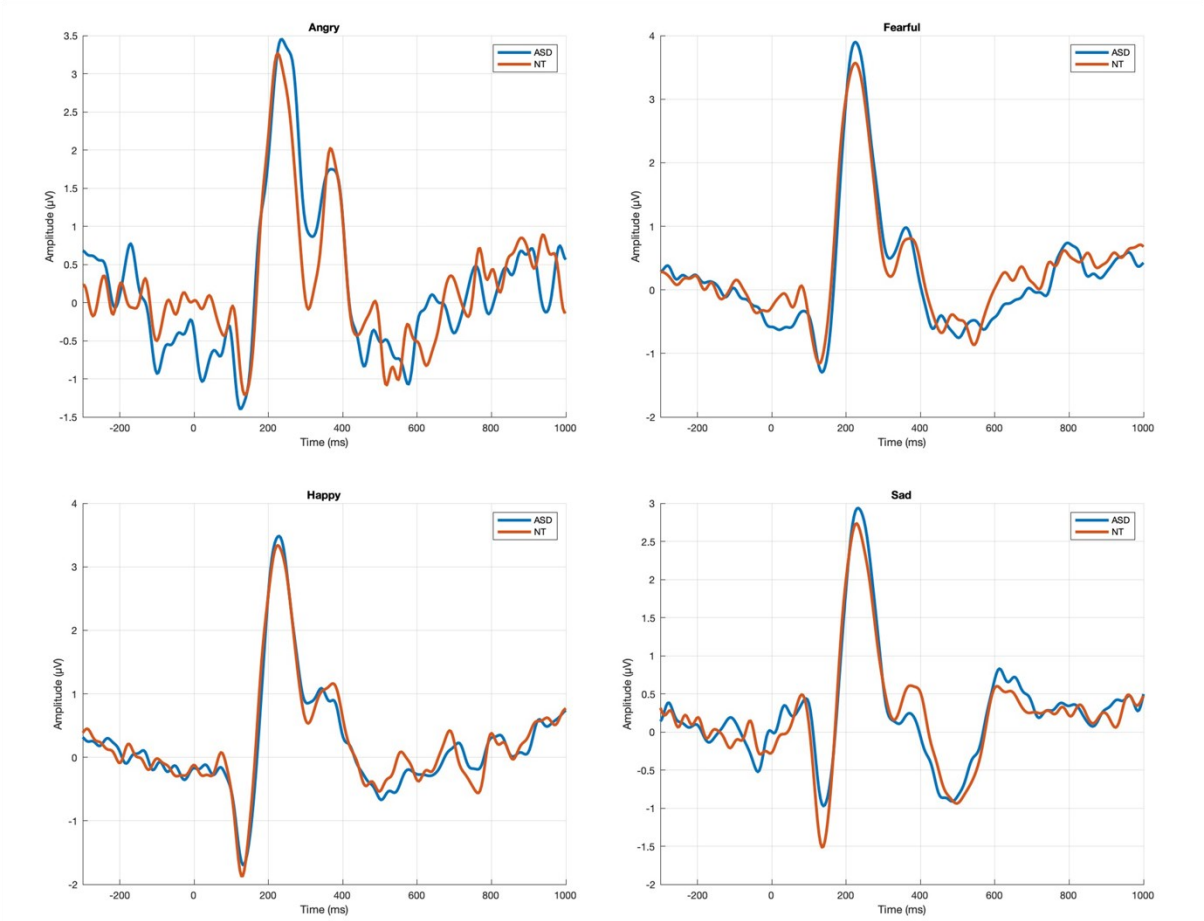
6.3. EEG Results

This section reports event-related potential (ERP) results focusing on the P200 component elicited by emotional vocalisations. P200 amplitude and latency were analyzed as a function of Group (ASD, NT) and Emotion (happy, angry, fearful, sad). EEG data from three participants (two NT, one ASD) were excluded due to excessive noise.

Figure 9 shows group-level averaged ERP waveforms across a frontocentral electrode cluster (F1, Fz, F2, FC1, FCz, FC2, C1, Cz, C2), for each emotion and group. A clear P200 component was observed around 200 ms in all conditions, with no pronounced group differences apparent at the waveform level.

Figure 9

Group-level Averaged ERP Waveforms by Emotion Category



Note. Waveforms were averaged across frontocentral electrodes (F1, Fz, F2, FC1, FCz, FC2, C1, Cz, C2)

6.3.1. Descriptive Statistics

Following outlier detection, a small number of amplitude and latency values were replaced using group × emotion mean imputation (see Appendix C, Table 13-

Table 16).

Descriptively, the ASD group showed a mean P200 amplitude of M = 3.92 microvolt (μV, SD = 1.32) and a mean latency of M = 232.13 ms (SD = 16.56). In the NT group, mean amplitude was M = 3.72 μV (SD = 1.16) and mean latency was M = 229.16 ms (SD = 15.02). Detailed descriptive statistics by group and emotion are provided in Appendix D, Table 21.

Assumptions of normality and homogeneity of variances were met for both amplitude and latency data (see Appendix B, Table 8).

6.3.2. P200 Amplitude

A mixed ANOVA with Group as a between-subjects factor and Emotion as a within-subjects factor revealed a significant main effect of Emotion, $F(2.54, 94.10) = 10.04$, $p < .001$, $\eta^2_p = .21$. Neither the main effect of Group, $F(1, 37) = 0.40$, $p = .530$, nor the Group × Emotion interaction, $F(2.54, 94.10) = 0.69$, $p = .536$, reached significance. Post hoc comparisons indicated that P200 amplitudes for sad vocalisations were significantly lower than for angry, fearful, and happy stimuli. No group-specific differences in amplitude were observed.

6.3.3. P200 Latency

For P200 latency, the mixed ANOVA revealed a significant main effect of Emotion, $F(2.91, 116.21) = 4.36$, $p = .009$, $\eta^2_p = .11$. The main effect of Group was not significant, $F(1, 40) = 0.84$, $p = .364$. Importantly, the Group × Emotion interaction was significant, $F(2.91, 116.21) = 5.22$, $p = .003$, $\eta^2_p = .12$. Follow-up comparisons indicated that participants with ASD showed significantly longer P200 latencies than neurotypical participants for angry vocalisations only, whereas no group differences emerged for fearful, happy, or sad stimuli (see Table 6). Detailed post hoc results are reported in Appendix D (Table 24Table 25).

Table 6

Contrasts of P200 Latency Between Groups by Emotion Category

contrast	emotion	estimate	SE	p.value	95% Confidence Interval for Difference		Cohen's d
					Lower Bound	Upper Bound	
ASD - NT	a	16,896	5,343	0,003	6,069	27,723	1.040
	f	0,642	4,583	0,889	-8,645	9,928	0.046
	h	-1,708	4,696	0,718	-11,223	7,808	-0.120
	s	-3,938	4,675	0,405	-13,41	5,534	-0.277

Note. Pairwise comparisons are based on estimated marginal means (EMMs) from the mixed ANOVA model. *p*-values are Holm-adjusted.

6.4. Relationship Between Behavioural Ratings and ERP Data

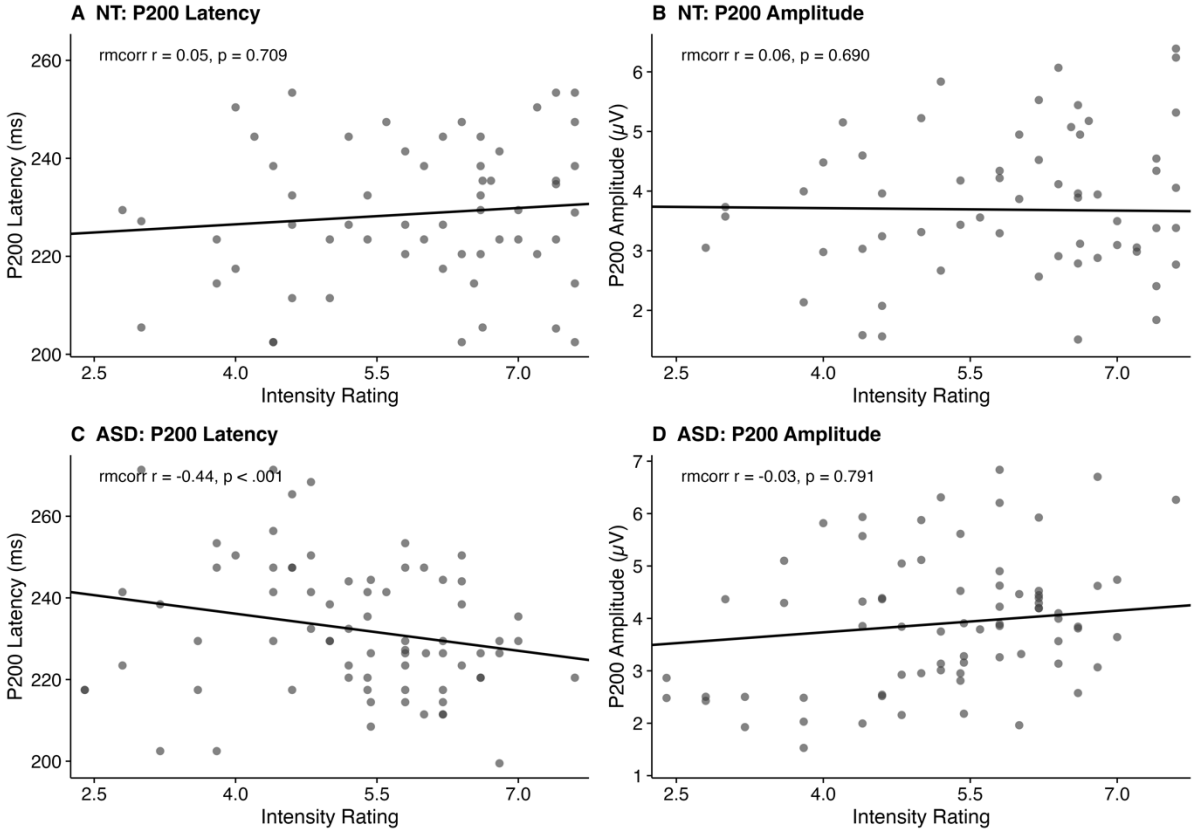
Repeated-measure correlations were conducted to examine the association between subjective intensity ratings and P200 measures (amplitude and latency) within participants (see Figure 10). In the neurotypical (NT) group, no significant correlations emerged, neither

between ratings and P200 latency (r = .052, p = .709) nor between ratings and P200 amplitude (r = .055, p = .690).

In the ASD group, a moderate and statistically significant negative correlation was found between ratings and P200 latency (r = –.443, p < .001, 95% CI [–.63, –.22]), indicating that sounds rated as more intense were processed more quickly. No significant association was observed between ratings and P200 amplitude in the ASD group.

Figure 10

RMCORR Between Intensity Ratings and P200 ERPs



Note. Scatterplots illustrate repeated-measures correlations between intensity ratings and P200 measures across emotion conditions within participants, separately for ASD and NT groups and P200 amplitudes and latencies.

6.4.1. Emotion Specific Correlations Between Ratings and ERP Data

To explore emotion-specific relationships between behavioural and electrophysiological responses, Pearson correlations between ratings and P200 parameters were computed separately for each emotion category (see Table 7). No significant correlations were found between ratings and P200 amplitude for any emotion (all |r| ≤ .17, all

$p \geq .311$). For P200 latency, only angry stimuli showed a significant negative correlation ($r = –.39$, $p = .016$), indicating faster processing of higher rated angry sounds. However, this effect did not remain significant after FDR correction ($p_{FDR} = .131$). No significant latency–rating correlations were observed for happy, sad, or fearful stimuli (all $|r| \leq .19$, all $p \geq .258$). Overall, the emotion-specific analyses did not show any robust correlations after correction for multiple testing.

Table 7

Correlations Between Ratings and ERP Data by Emotion Category

Emotion	ERP measure	r	p.value (uncorrected)	p.value (FDR)
Happy	Amplitude	0.17	.311	.674
Happy	Latency	–0.04	.818	.818
Angry	Amplitude	0.05	.763	.818
Angry	Latency	–0.39	.016	.131
Sad	Amplitude	0.16	.348	.674
Sad	Latency	0.19	.258	.674
Fearful	Amplitude	0.11	.514	.686
Fearful	Latency	0.13	.422	.674

Note. Pearson Correlations. p-values were corrected using Benjamini–Hochberg FDR (k = 8 tests).

7. Discussion

7.1. Summary of Key Findings

The aim of this thesis was to investigate auditory emotion processing in individuals with ASD and neurotypical individuals using a combination of behavioural and electrophysiological measures. Specifically, the study examined emotion recognition accuracy, subjective intensity ratings, and early neural responses to brief, nonverbal emotional vocalisations. Given the mixed and inconclusive findings in the auditory domain of emotion processing in autism, this work sought to contribute to a better understanding by assessing both behavioural performance and early ERP markers.

Overall, the findings indicate both shared and distinct aspects of auditory emotion processing in individuals with ASD and neurotypical persons. Behaviourally, participants with ASD showed lower overall accuracy and reduced perceived emotional intensity compared to neurotypical participants, with group differences in accuracy being particularly pronounced for fearful vocalisations. In contrast, both groups demonstrated similar differentiation between emotion categories in their intensity ratings.

At the neural level, P200 amplitudes were modulated by emotion but did not differ between groups, whereas P200 latencies showed an emotion-specific group difference, with participants with ASD exhibiting delayed responses to angry vocalisations. Brain–behaviour analyses revealed a negative association between P200 latency and perceived emotional intensity in the ASD group only, suggesting a closer coupling between early neural processing speed and subjective emotion perception in individuals with ASD.

7.1.1. Behavioural Data

Regarding behavioural performance, participants with ASD demonstrated lower overall accuracy and reduced subjective intensity ratings compared to neurotypical participants. These findings are broadly consistent with previous studies reporting impairments in auditory emotion recognition in ASD (e.g., Day et al., 2024; Globerson et al., 2015; Schelinski & von Kriegstein, 2019). The present results also indicate that these differences are not similar across emotional categories. Group differences in recognition accuracy were particularly pronounced for fearful vocalisations, whereas no significant group differences emerged for angry, happy, or sad stimuli. This pattern contrasts with meta-analytic findings suggesting stronger group differences for happy and angry vocal expressions (Leung, 2022; Zhang et al., 2023).

Although participants with ASD rated emotional vocalisations as less intense overall, both groups showed comparable differentiation between emotion categories. This suggests that while emotional salience may be less strong in ASD, the relative structure of emotion perception remains largely preserved. However, it should be noted that the present study was primarily designed to assess overall group differences rather than fine-grained emotion-specific effects. Given the relatively modest sample size and the statistical tests used, the emotion-specific findings should therefore be interpreted cautiously and considered exploratory. Nevertheless, the observed pattern supports the notion that auditory emotion recognition in ASD is not uniformly impaired but varies across emotional categories, underscoring the importance of examining emotion-specific mechanisms in future research.

7.1.2. EEG data

With respect to the ERP component data, no main effect of group was found for either P200 amplitudes or overall P200 latency, although emotion-specific group differences emerged at the latency level, indicating that early sensory-affective encoding of emotional vocalisations is largely comparable between individuals with ASD and neurotypical individuals. This finding aligns with previous EEG studies reporting intact early auditory ERP responses in ASD at the level of P200 amplitude (e.g., Day et al., 2024). The absence of group differences in P200 amplitude contrasts with models proposing a fundamental deficit in early auditory emotion encoding in autism.

7.1.3. Brain – Behaviour Coupling

Although early ERP responses to emotional vocalisations did not differ robustly between groups, analyses of brain–behaviour relationships revealed group-specific patterns. Specifically, in participants with ASD, longer P200 latencies were associated with lower perceived emotional intensity, whereas no such association was observed in the neurotypical group. This finding suggests that early neural responses and subjective emotion perception are more tightly coupled in autism.

More specifically, this pattern suggests that early auditory processing may play a more functionally relevant role in supporting emotion perception in individuals with ASD. Importantly, the results of this study do not support a primary deficit in early sensory encoding, as P200 amplitudes were largely comparable across groups. Instead, they point toward differences in how early perceptual and attentional processes are integrated into subjective emotional experience. Individuals with ASD may rely more strongly on early-

stage auditory processing to construct emotional meaning from vocal signals, particularly when contextual and linguistic cues are minimized (Leung et al., 2022).

An alternative interpretation of this observed brain–behaviour coupling is that the relationship reflects interindividual differences within the autistic group rather than a uniform processing strategy. It is possible that only a subgroup of participants with, namely those who generally perceived the stimuli as less emotionally intense, exhibited prolonged P200 latencies. This interpretation would be consistent with previous findings suggesting that early auditory processing differences in ASD may be driven by individual variability rather than group-wide effects. Specifically, Lerner et al. (2013) showed that N100 latencies predicted whether an individual with ASD belonged to a subgroup with intact or impaired emotion recognition ability, suggesting that early ERP components may serve as a marker of individual variability in social-emotional functioning.

However, the present thesis was not designed to disentangle subgroup-specific effects, and therefore cannot determine which interpretation is more likely. Taken together, these findings underscore the importance of future studies that look closer into this relationship between neural responses and behavioural performance by taking into account interindividual differences within diagnostic groups and not only rely on group-average comparisons. Future studies employing larger samples and/or person-centered or clustering approaches may help clarify whether distinct subgroups or processing profiles underlie the observed coupling between early neural responses and subjective emotional perception.

7.2. Implications for the Temporal Characteristics of Auditory Emotion Processing Differences

Taken together, these behavioural and neuropsychological results allow implications for the question where auditory emotion processing differences in autism are most likely to arise. The absence of robust group differences in P200 amplitude and latency suggests that early sensory-affective encoding of emotional vocalisations is largely intact in individuals with ASD. At the same time, differences in behavioural performance and in brain–behaviour coupling exist, suggesting that differences appear during the integration of early auditory signals into subjective emotional experience. Thus, atypicalities in autism seem more closely related to how early perceptual and attentional processes contribute to emotion evaluation than to basic sensory deficits.

7.2.1. Role of nonverbal vocalisations

This study used short, nonverbal vocalisations to elicit emotional responses. Previous research suggests that individuals with ASD may rely more strongly contextual cues when processing emotional sounds (e.g., Le Sourn-Bissaoui et al., 2013). An open question therefore is whether early neural responses to emotional vocalisations in autism remain intact when such cues are minimized. The absence of robust group differences in P200 amplitude, together with largely comparable P200 latencies across groups, indicates that early sensory-affective encoding of emotional vocal sounds is largely preserved in individuals with ASD, even under conditions that reduce higher-order interpretive support. The presence of emotion-specific latency differences further suggests that early processing is modulated by emotional content rather than globally altered. To better understand the role of nonverbal vocalisations relative to speech prosody, future studies should directly compare verbal and nonverbal emotional stimuli within the same experimental design.

7.2.2. Female sample

A notable deviation from other study samples concerns the high proportion of female participants in the autistic group (approximately 80%), with a gender-matched neurotypical comparison group. This contrasts with much of the existing autism literature, which is strongly male-biased, both in clinical samples and in experimental research (e.g. Loomes et al., 2017). Importantly, the uneven gender distribution may have influenced the present findings. Previous research suggests that gender can moderate auditory and emotional processing, with women often showing heightened sensitivity to emotional auditory cues compared to men (e.g., Allgood & Heaton, 2015). Moreover, there is also evidence that autistic women may differ from autistic men in social perception, compensatory strategies, and sensory processing profiles (e.g. Kothari et al., 2013), which could have affected both the behavioural and neural responses to emotional stimuli in this study.

At the same time, the predominance of female participants can also be viewed as a strength of the present study. Women with ASD are known to be underrepresented in autism research, which goes in line with a limited understanding of gender-specific processes and potentially biased theoretical models of autism. Future research should explicitly investigate gender as a moderating factor, ideally with designs that include balanced gender samples or that directly compare women and men on the autism spectrum in order to better be able to distinguish gender- and diagnosis-related effects.

7.2.3. Group-based comparisons vs. dimensional approaches

Although the groups differed significantly in their AQ scores as expected, some individuals in the ASD group scored below the clinical threshold, as well as some participants in the neurotypical group scores above the threshold. This pattern underscores that autism is not a uniform construct, but rather a broad and dimensional spectrum characterised by intra- and interindividual variability. Moreover, it underscores that strict categorical grouping (ASD vs. NT) cannot always reflect all relevant differences or similarities in neurocognitive processing. These findings are consistent with current theoretical approaches that view autism as a continuous dimension of social-cognitive, sensory and emotional characteristics rather than a dichotomous clinical category (Fletcher-Watson & Happé, 2019). For future studies, a dimensional approach (e.g. using continuous measures such as AQ or sensory profiles as predictors) could therefore enable a more differentiated analysis and this could possibly respond more sensitively to individual differences than group-based comparisons.

7.3. Caveats and Targets for Future Research

Several limitations should be considered when interpreting the present findings. First, the study focused specifically on the P200 component as an index of early sensory-affective processing. Although this allowed for targeted investigation of early auditory encoding, emotion perception unfolds across multiple temporal stages. Future research integrating early and later ERP components within the same experimental framework may provide a more comprehensive understanding of the temporal dynamics underlying auditory emotion processing in autism.

Second, the exclusive use of brief, nonverbal vocalisations limited the scope of auditory stimuli examined. While this design minimized contextual influences, it remains unclear whether similar neural–behavioural patterns would emerge for other types of vocalisations, such as prosody. Direct comparisons between verbal and nonverbal emotional stimuli would help clarify modality-specific and context-dependent processing mechanisms.

Third, the high proportion of female participants may limit generalizability to predominantly male samples in the existing literature on ASD. Given evidence for gender-related differences in emotional and auditory processing, future studies should explicitly examine gender as a moderating factor.

Finally, although group-based comparisons revealed certain patterns, autism represents a heterogeneous and dimensional condition. The observed variability in brain–behaviour coupling underscores the importance of approaches that move beyond categorical contrasts and incorporate continuous trait measures or person-centered analyses. Larger samples will be necessary for disentangling subgroup-specific processing profiles.

7.4. Conclusion

To conclude, this Master's thesis contributes to the mixed literature on auditory emotion processing in autism by demonstrating a dissociation between behavioural performance and early neural responses to affective vocalisations. At the behavioural level, autistic and neurotypical groups differed in both emotion recognition accuracy and subjective intensity ratings. At the neural level, early auditory–affective processing indexed by the P200 ERP component was largely comparable between groups. However, the relationship between P200 latency and behavioural measures differed, indicating altered brain–behaviour coupling in autism.

By integrating behavioural and electrophysiological measures within one experimental framework and employing brief, nonverbal vocal stimuli, the present study provides further insights into auditory emotion processing in autism. Rather than pointing to uniform sensory deficits, the results emphasize heterogeneity within ASD and suggest that differences in the integration of neural signals into conscious emotional experience may play a central role in auditory emotion perception.

8. Literature

Adolphs, R. (2009). The Social Brain: Neural basis of social knowledge. *Annual Review of Psychology*, *60*(Volume 60, 2009), 693–716. https://doi.org/10.1146/annurev.psych.60.110707.163514

Adolphs, R., Damasio, H., Tranel, D., Cooper, G., & Damasio, A. R. (2000). A role for somatosensory cortices in the visual recognition of emotion as revealed by three-dimensional lesion mapping. *Journal of Neuroscience*, *20*(7), 2683–2690. https://doi.org/10.1523/JNEUROSCI.20-07-02683.2000

Adolphs, R., Tranel, D., Damasio, H., & Damasio, A. (1994). Impaired recognition of emotion in facial expressions following bilateral damage to the human amygdala. *Nature*, *372*(6507), 669–672. https://doi.org/10.1038/372669a0

Allgood, R., & Heaton, P. (2015). Developmental change and cross-domain links in vocal and musical emotion recognition performance in childhood. *The British Journal of Developmental Psychology*, *33*(3), 398–403. https://doi.org/10.1111/bjdp.12097

Atherton, G., Edisbury, E., Piovesan, A., & Cross, L. (2022). Autism through the ages: A mixed methods approach to understanding how age and age of diagnosis affect quality of life. *Journal of Autism and Developmental Disorders*, *52*(8), 3639–3654. https://doi.org/10.1007/s10803-021-05235-x

Bachorowski, J.-A. (1999). Vocal Expression and Perception of Emotion. *Current Directions in Psychological Science*, *8*(2), 53–57. https://doi.org/10.1111/1467-8721.00013

Bagby, R. M., Parker, J. D. A., & Taylor, G. J. (2020). Twenty-five years with the 20-item Toronto Alexithymia Scale. *Journal of Psychosomatic Research*, *131*, 109940. https://doi.org/10.1016/j.jpsychores.2020.109940

Bakdash, J. Z., & Marusich, L. R. (2017). Repeated Measures Correlation. *Frontiers in Psychology*, *8*. https://doi.org/10.3389/fpsyg.2017.00456

Baker, K. F., Montgomery, A. A., & Abramson, R. (2010). Brief Report: Perception and lateralization of spoken emotion by youths with high-functioning forms of autism. *Journal of Autism and Developmental Disorders*, *40*(1), 123–129. https://doi.org/10.1007/s10803-009-0841-1

Banissy, M. J., Sauter, D. A., Ward, J., Warren, J. E., Walsh, V., & Scott, S. K. (2010). Suppressing sensorimotor activity modulates the discrimination of auditory emotions but not speaker identity. *Journal of Neuroscience*, *30*(41), 13552–13557. https://doi.org/10.1523/JNEUROSCI.0786-10.2010

Banse, R., & Scherer, K. R. (1996). Acoustic profiles in vocal emotion expression. *Journal of Personality and Social Psychology*, *70*(3), 614–636. https://doi.org/10.1037/0022-3514.70.3.614

Baron-Cohen, S., Leslie, A. M., & Frith, U. (1985). Does the autistic child have a "theory of mind" ? *Cognition*, *21*(1), 37–46. https://doi.org/10.1016/0010-0277(85)90022-8

Baron-Cohen, S., Lombardo, M. V., Auyeung, B., Ashwin, E., Chakrabarti, B., & Knickmeyer, R. (2011). Why are autism spectrum conditions more prevalent in males? *PLOS Biology*, *9*(6), e1001081. https://doi.org/10.1371/journal.pbio.1001081

Baron-Cohen, S., Wheelwright, S., Skinner, R., Martin, J., & Clubley, E. (2001). The autism-spectrum quotient (AQ): Evidence from Asperger syndrome/high-functioning autism, males and females, scientists and mathematicians. *Journal of Autism and Developmental Disorders*, *31*(1), 5–17. https://doi.org/10.1023/A:1005653411471

Baumeister, R. F., Vohs, K. D., Nathan DeWall, C., & Liqing Zhang. (2007). How Emotion Shapes Behavior: Feedback, Anticipation, and Reflection, Rather Than Direct Causation. *Personality and Social Psychology Review*, *11*(2), 167–203. https://doi.org/10.1177/1088868307301033

Baxter, A. J., Brugha, T. S., Erskine, H. E., Scheurer, R. W., Vos, T., & Scott, J. G. (2015). The epidemiology and global burden of autism spectrum disorders. *Psychological Medicine*, *45*(3), 601–613. https://doi.org/10.1017/S003329171400172X

Belin, P., Zatorre, R. J., Lafaille, P., Ahad, P., & Pike, B. (2000). Voice-selective areas in human auditory cortex. *Nature*, *403*(6767), 309–312. https://doi.org/10.1038/35002078

Benjamini, Y., & Hochberg, Y. (1995). Controlling the False Discovery Rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B (Methodological)*, *57*(1), 289–300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x

Bestelmeyer, P. E. G., Maurage, P., Rouger, J., Latinus, M., & Belin, P. (2014). Adaptation to vocal expressions reveals multistep perception of auditory emotion. *Journal of Neuroscience*, *34*(24), 8098–8105. https://doi.org/10.1523/JNEUROSCI.4820-13.2014

BioSemi. (n.d.). *Biosemi EEG amplifier electrodes*. Retrieved January 12, 2026, from https://www.biosemi.com/

Blackford, J. U., Buckholtz, J. W., Avery, S. N., & Zald, D. H. (2010). A unique role for the human amygdala in novelty detection. *NeuroImage*, *50*(3), 1188–1193. https://doi.org/10.1016/j.neuroimage.2009.12.083

Boucher, J., Lewis, V., & Collis, G. M. (2000). Voice processing abilities in children with autism, children with specific language impairments, and young typically developing children. *Journal of Child Psychology and Psychiatry*, *41*(7), 847–857. https://doi.org/10.1111/1469-7610.00672

Brennand, R., Schepman, A., & Rodway, P. (2011). Vocal emotion perception in pseudo-sentences by secondary-school children with Autism Spectrum Disorder. *Research in Autism Spectrum Disorders*, *5*(4), 1567–1573. https://doi.org/10.1016/j.rasd.2011.03.002

Bruning, N., Konrad, K., & Herpertz-Dahlmann, B. (2005). Bedeutung und Ergebnisse der Theory of Mind-Forschung für den Autismus und andere psychiatrische Erkrankungen. *Zeitschrift Für Kinder- Und Jugendpsychiatrie Und Psychotherapie*, *33*(2), 77–88. https://doi.org/10.1024/1422-4917.33.2.77

Chronaki, G., Wigelsworth, M., Pell, M. D., & Kotz, S. A. (2018). The development of cross-cultural recognition of vocal emotion during childhood and adolescence. *Scientific Reports*, *8*(1), 8659. https://doi.org/10.1038/s41598-018-26889-1

Cook, R. D. (1977). Detection of influential observation in linear regression. *Technometrics*, *19*(1), 15–18. https://doi.org/10.1080/00401706.1977.10489493

Coutinho, E., & Dibben, N. (2013). Psychoacoustic cues to emotion in speech prosody and music. *Cognition and Emotion*, *27*(4), 658–684. https://doi.org/10.1080/02699931.2012.732559

Day, T. C., Malik, I., Boateng, S., Hauschild, K. M., & Lerner, M. D. (2024). Vocal emotion recognition in autism: Behavioral performance and event-related potential (ERP) response. *Journal of Autism and Developmental Disorders*, *54*(4), 1235–1248. https://doi.org/10.1007/s10803-023-05898-8

Delorme, A., & Makeig, S. (2004). EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods*, *134*(1), 9–21. https://doi.org/10.1016/j.jneumeth.2003.10.009

Devlin, B., & Scherer, S. W. (2012). Genetic architecture in autism spectrum disorder. *Current Opinion in Genetics & Development, Molecular and Genetic Bases of Disease*, *22*(3), 229–237. https://doi.org/10.1016/j.gde.2012.03.002

Ekman, P. (1992). An argument for basic emotions. *Cognition and Emotion*, *6*(3–4), 169–200. https://doi.org/10.1080/02699939208411068

El Achkar, C. M., & Spence, S. J. (2015). Clinical characteristics of children and young adults with co-occurring autism spectrum disorder and epilepsy. *Epilepsy & Behavior: E&B*, *47*, 183–190. https://doi.org/10.1016/j.yebeh.2014.12.022

English, M. C. W., Gignac, G. E., Visser, T. A. W., Whitehouse, A. J. O., & Maybery, M. T. (2020). A comprehensive psychometric analysis of autism-spectrum quotient factor models using two large samples: Model recommendations and the influence of divergent traits on total-scale scores. *Autism Research*, *13*(1), 45–60. https://doi.org/10.1002/aur.2198

Ethofer, T., Bretscher, J., Gschwind, M., Kreifelts, B., Wildgruber, D., & Vuilleumier, P. (2012). Emotional Voice Areas: Anatomic location, functional properties, and structural connections revealed by combined fMRI/DTI. *Cerebral Cortex*, *22*(1), 191–200. https://doi.org/10.1093/cercor/bhr113

Fletcher-Watson, S., & Happé, F. (2019). *Autism: A New Introduction to Psychological Theory and Current Debate* (2nd ed.). Routledge. https://doi.org/10.4324/9781315101699

Foster, N., Tryfon, A., Ouimet, T., Doyle-Thomas, K., Anagnostou, E., Evans, A., Zwaigenbaum, L., Hyde, K., & Group, for N. A. imaging. (2015). Auditory cortical structure predicts superior pitch processing in children with autism. *International Journal of Developmental Neuroscience*, *47*(Part_A), 96–97. https://doi.org/10.1016/j.ijdevneu.2015.04.264

Frith, U., & Happé, F. (1994). Autism: Beyond "theory of mind." *Cognition*, *50*(1), 115–132. https://doi.org/10.1016/0010-0277(94)90024-8

Frühholz, S., Trost, W., & Kotz, S. A. (2016). The sound of emotions—Towards a unifying neural network perspective of affective sound processing. *Neuroscience & Biobehavioral Reviews*, *68*, 96–110. https://doi.org/10.1016/j.neubiorev.2016.05.002

Gao, S., Wang, X., & Su, Y. (2023). Examining whether adults with autism spectrum disorder encounter multiple problems in theory of mind: A study based on meta-analysis. *Psychonomic Bulletin & Review*, *30*(5), 1740–1758. https://doi.org/10.3758/s13423-023-02280-8

Gardener, H., Spiegelman, D., & Buka, S. L. (2009). Prenatal Risk Factors for Autism: A Comprehensive Meta-analysis. *The British Journal of Psychiatry : The Journal of Mental Science*, *195*(1), 7–14. https://doi.org/10.1192/bjp.bp.108.051672

Gebauer, L., Skewes, J., Westphael, G., Heaton, P., & Vuust, P. (2014). Intact brain processing of musical emotions in autism spectrum disorder, but more cognitive load and arousal in happy vs. Sad music. *Frontiers in Neuroscience*, *8*. https://doi.org/10.3389/fnins.2014.00192

Gerdes, A. B. M., Wieser, M. J., Bublatzky, F., Kusay, A., Plichta, M. M., & Alpers, G. W. (2013). Emotional sounds modulate early neural processing of emotional pictures. *Frontiers in Psychology*, *4*. https://doi.org/10.3389/fpsyg.2013.00741

Gervais, H., Belin, P., Boddaert, N., Leboyer, M., Coez, A., Sfaello, I., Barthélémy, C., Brunelle, F., Samson, Y., & Zilbovicius, M. (2004). Abnormal cortical voice processing in autism. *Nature Neuroscience*, *7*(8), 801–802. https://doi.org/10.1038/nn1291

Giarelli, E., Wiggins, L. D., Rice, C. E., Levy, S. E., Kirby, R. S., Pinto-Martin, J., & Mandell, D. (2010). Sex differences in the evaluation and diagnosis of autism spectrum disorders among children. *Disability and Health Journal*, *3*(2), 107–116. https://doi.org/10.1016/j.dhjo.2009.07.001

Globerson, E., Amir, N., Kishon-Rabin, L., & Golan, O. (2015a). Prosody Recognition in Adults With High-Functioning Autism Spectrum Disorders: From Psychoacoustics to Cognition. *Autism Research*, *8*(2), 153–163. https://doi.org/10.1002/aur.1432

Globerson, E., Amir, N., Kishon-Rabin, L., & Golan, O. (2015b). Prosody recognition in adults with high-functioning autism spectrum disorders: From psychoacoustics to cognition. *Autism Research: Official Journal of the International Society for Autism Research*, *8*(2), 153–163. https://doi.org/10.1002/aur.1432

Golan, O., Baron-Cohen, S., Hill, J. J., & Rutherford, M. D. (2007). The 'Reading the Mind in the Voice' Test-revised: A study of complex emotion recognition in adults with and without autism spectrum conditions. *Journal of Autism and Developmental Disorders*, *37*(6), 1096–1106. https://doi.org/10.1007/s10803-006-0252-5

Grandjean, D., Sander, D., Pourtois, G., Schwartz, S., Seghier, M. L., Scherer, K. R., & Vuilleumier, P. (2005). The voices of wrath: Brain responses to angry prosody in meaningless speech. *Nature Neuroscience*, *8*(2), 145–146. https://doi.org/10.1038/nn1392

Greenhouse, S. W., & Geisser, S. (1959). On methods in the analysis of profile data. *Psychometrika*, *24*, 95–112. https://doi.org/10.1007/BF02289823

Grossman, R. B., Bemis, R. H., Plesa Skwerer, D., & Tager-Flusberg, H. (2010). Lexical and affective prosody in children with high-functioning autism. *Journal of Speech,*

Language, and Hearing Research, *53*(3), 778–793. https://doi.org/10.1044/1092-4388(2009/08-0127)

Happé, F., & Frith, U. (2006). The Weak Coherence account: Detail-focused cognitive style in autism spectrum disorders. *Journal of Autism and Developmental Disorders*, *36*(1), 5–25. https://doi.org/10.1007/s10803-005-0039-0

Heaton, P. (2003). Pitch memory, labelling and disembedding in autism. *Journal of Child Psychology and Psychiatry*, *44*(4), 543–551. https://doi.org/10.1111/1469-7610.00143

Heaton, P., Hermelin, B., & Pring, L. (1998). Autism and pitch processing: A precursor for savant musical ability? *Music Perception*, *15*(3), 291–305. https://doi.org/10.2307/40285769

Heaton, P., Reichenbacher, L., Sauter, D., Allen, R., Scott, S., & Hill, E. (2012). Measuring the effects of alexithymia on perception of emotional vocalizations in autistic spectrum disorder and typical development. *Psychological Medicine*, *42*(11), 2453–2459. https://doi.org/10.1017/S0033291712000621

Hertz-Picciotto, I., Schmidt, R. J., & Krakowiak, P. (2018). Understanding environmental contributions to autism: Causal concepts and the state of science. *Autism Research*, *11*(4), 554–586. https://doi.org/10.1002/aur.1938

Hyde, K. L., Samson, F., Evans, A. C., & Mottron, L. (2010). Neuroanatomical differences in brain areas implicated in perceptual and other core features of autism revealed by cortical thickness analysis and voxel-based morphometry. *Human Brain Mapping*, *31*(4), 556–566. https://doi.org/10.1002/hbm.20887

Järvinen, A., Ng, R., Crivelli, D., Neumann, D., Arnold, A. J., Woo-VonHoogenstyn, N., Lai, P., Trauner, D., & Bellugi, U. (2016). Social functioning and autonomic nervous system sensitivity across vocal and musical emotion in Williams syndrome and autism spectrum disorder. *Developmental Psychobiology*, *58*(1), 17–26. https://doi.org/10.1002/dev.21335

Jensen de López, K. M., & Thirup Møller, H. (2024). Prevalence of Autism in Scandinavian Countries (Denmark, Norway, Sweden), and Nordic Countries (Finland, Iceland, the Faroe Islands, and Greenland). *Neuropsychiatric Disease and Treatment*, *20*, 1597–1612. https://doi.org/10.2147/NDT.S466081

Jones, C. R. G., Pickles, A., Falcaro, M., Marsden, A. J. S., Happé, F., Scott, S. K., Sauter, D., Tregay, J., Phillips, R. J., Baird, G., Simonoff, E., & Charman, T. (2011). A multimodal approach to emotion recognition ability in autism spectrum disorders.

Journal of Child Psychology and Psychiatry, and Allied Disciplines, *52*(3), 275–285. https://doi.org/10.1111/j.1469-7610.2010.02328.x

Jorgensen, A. R., Whitehouse, A. J. O., Fox, A. M., & Maybery, M. T. (2021). Delayed cortical processing of auditory stimuli in children with autism spectrum disorder: A meta-analysis of electrophysiological studies. *Brain and Cognition*, *150*, 105709. https://doi.org/10.1016/j.bandc.2021.105709

Kaplan-Kahn, E. A., Park, A., & Russo, N. (2021). Pathways of perceptual primacy: ERP evidence for relationships between autism traits and enhanced perceptual functioning. *Neuropsychologia*, *163*, 108065. https://doi.org/10.1016/j.neuropsychologia.2021.108065

Korpilahti, P., Jansson-Verkasalo, E., Mattila, M.-L., Kuusikko, S., Suominen, K., Rytky, S., Pauls, D. L., & Moilanen, I. (2007). Processing of affective speech prosody is impaired in asperger syndrome. *Journal of Autism and Developmental Disorders*, *37*(8), 1539–1549. https://doi.org/10.1007/s10803-006-0271-2

Kothari, R., Skuse, D., Wakefield, J., & Micali, N. (2013). Gender differences in the relationship between social communication and emotion recognition. *Journal of the American Academy of Child & Adolescent Psychiatry*, *52*(11), 1148-1157.e2. https://doi.org/10.1016/j.jaac.2013.08.006

Kragel, P. A., & LaBar, K. S. (2015). Multivariate neural biomarkers of emotional states are categorically distinct. *Social Cognitive and Affective Neuroscience*, *10*(11), 1437–1448. https://doi.org/10.1093/scan/nsv032

Kreifelts, B., Ethofer, T., Shiozawa, T., Grodd, W., & Wildgruber, D. (2009). Cerebral representation of non-verbal emotional perception: fMRI reveals audiovisual integration area between voice- and face-sensitive regions in the superior temporal sulcus. *Neuropsychologia*, *47*(14), 3059–3066. https://doi.org/10.1016/j.neuropsychologia.2009.07.001

Lai, M.-C., Lombardo, M. V., & Baron-Cohen, S. (2014). Autism. *The Lancet*, *383*(9920), 896–910. https://doi.org/10.1016/S0140-6736(13)61539-1

Latinus, M., & Belin, P. (2011). Human voice perception. *Current Biology*, *21*(4), R143–R145. https://doi.org/10.1016/j.cub.2010.12.033

Laukka, P., Bänziger, T., Israelsson, A., Cortes, D. S., Tornberg, C., Scherer, K. R., & Fischer, H. (2021). Investigating individual differences in emotion recognition ability using the ERAM test. *Acta Psychologica*, *220*, 103422. https://doi.org/10.1016/j.actpsy.2021.103422

Le Sourn-Bissaoui, S., Aguert, M., Girard, P., Chevreuil, C., & Laval, V. (2013). Emotional speech comprehension in children and adolescents with autism spectrum disorders. *Journal of Communication Disorders*, *46*(4), 309–320. https://doi.org/10.1016/j.jcomdis.2013.03.002

Lerner, J. S., Li, Y., Valdesolo, P., & Kassam, K. S. (2015). Emotion and Decision Making. *Annual Review of Psychology*, *66*(1), 799–823. https://doi.org/10.1146/annurev-psych-010213-115043

Lerner, M. D., McPartland, J. C., & Morris, J. P. (2013). Multimodal emotion processing in autism spectrum disorders: An event-related potential study. *Developmental Cognitive Neuroscience*, *3*, 11–21. https://doi.org/10.1016/j.dcn.2012.08.005

Leung, F. Y. N., Sin, J., Dawson, C., Ong, J. H., Zhao, C., Veić, A., & Liu, F. (2022). Emotion recognition across visual and auditory modalities in autism spectrum disorder: A systematic review and meta-analysis. *Developmental Review*, *63*, 101000. https://doi.org/10.1016/j.dr.2021.101000

Levene, H. (1960). Robust tests for equality of variances. In *Olkin, I., Ed., Contributions to Probability and Statistics, Stanford University Press, Palo Alto* (pp. 278–292).

Lindner, J. L., & Rosén, L. A. (2006). Decoding of emotion through facial expression, prosody and verbal content in children and adolescents with Asperger's syndrome. *Journal of Autism and Developmental Disorders*, *36*(6), 769–777. https://doi.org/10.1007/s10803-006-0105-2

Liu, T., Pinheiro, A. P., Deng, G., Nestor, P. G., McCarley, R. W., & Niznikiewicz, M. A. (2012). Electrophysiological insights into processing nonverbal emotional vocalizations. *NeuroReport*, *23*(2), 108. https://doi.org/10.1097/WNR.0b013e32834ea757

Liu, T., Pinheiro, A., Zhao, Z., Nestor, P. G., McCarley, R. W., & Niznikiewicz, M. A. (2012). Emotional Cues during Simultaneous Face and Voice Processing: Electrophysiological Insights. *PLoS ONE*, *7*(2), e31001. https://doi.org/10.1371/journal.pone.0031001

Long, E. L., Catmur, C., & Bird, G. (2025). The theory of mind hypothesis of autism: A critical evaluation of the status quo. *Psychological Review*. https://doi.org/10.1037/rev0000532

Loomes, R., Hull, L., & Mandy, W. P. L. (2017). What is the Male-to-Female ratio in autism spectrum disorder? A systematic review and meta-analysis. *Journal of the American*

Academy of Child & Adolescent Psychiatry, *56*(6), 466–474. https://doi.org/10.1016/j.jaac.2017.03.013

Lord, C., Charman, T., Havdahl, A., Carbone, P., Anagnostou, E., Boyd, B., Carr, T., Vries, P. J. de, Dissanayake, C., Divan, G., Freitag, C. M., Gotelli, M. M., Kasari, C., Knapp, M., Mundy, P., Plank, A., Scahill, L., Servili, C., Shattuck, P., … McCauley, J. B. (2022). The Lancet Commission on the future of care and clinical research in autism. *The Lancet*, *399*(10321), 271–334. https://doi.org/10.1016/S0140-6736(21)01541-5

Manjaly, Z. M., Bruning, N., Neufang, S., Stephan, K. E., Brieber, S., Marshall, J. C., Kamp-Becker, I., Remschmidt, H., Herpertz-Dahlmann, B., Konrad, K., & Fink, G. R. (2007). Neurophysiological correlates of relatively enhanced local visual search in autistic adolescents. *NeuroImage*, *35*(1), 283–291. https://doi.org/10.1016/j.neuroimage.2006.11.036

Masi, A., DeMayo, M. M., Glozier, N., & Guastella, A. J. (2017). An overview of autism spectrum disorder, heterogeneity and treatment options. *Neuroscience Bulletin*, *33*(2), 183–193. https://doi.org/10.1007/s12264-017-0100-y

Mauchand, M., & Zhang, S. (2023). Disentangling emotional signals in the brain: An ALE meta-analysis of vocal affect perception. *Cognitive, Affective, & Behavioral Neuroscience*, *23*(1), 17–29. https://doi.org/10.3758/s13415-022-01030-y

Milner, V., McIntosh, H., Colvert, E., & Happé, F. (2019). A qualitative exploration of the female experience of autism spectrum disorder (ASD). *Journal of Autism and Developmental Disorders*, *49*(6), 2389–2402. https://doi.org/10.1007/s10803-019-03906-4

Mottron, L., Dawson, M., Soulières, I., Hubert, B., & Burack, J. (2006). Enhanced perceptual functioning in autism: An update, and eight principles of autistic perception. *Journal of Autism and Developmental Disorders*, *36*(1), 27–43. https://doi.org/10.1007/s10803-005-0040-7

Mottron, L., Peretz, I., & Ménard, E. (2000). Local and global processing of music in high-functioning persons with autism: Beyond central coherence? *Journal of Child Psychology and Psychiatry*, *41*(8), 1057–1065. https://doi.org/10.1111/1469-7610.00693

Müllensiefen, D., Gingras, B., Musil, J., & Stewart, L. (2014). The musicality of non-musicians: An index for assessing musical sophistication in the general population. *PLOS ONE*, *9*(2), e89642. https://doi.org/10.1371/journal.pone.0089642

Nelson, C. A., & McCleery, J. P. (2008). Use of event-related potentials in the study of typical and atypical development. *Journal of the American Academy of Child & Adolescent Psychiatry*, *47*(11), 1252–1261. https://doi.org/10.1097/CHI.0b013e318185a6d8

Newschaffer, C. J., Croen, L. A., Daniels, J., Giarelli, E., Grether, J. K., Levy, S. E., Mandell, D. S., Miller, L. A., Pinto-Martin, J., Reaven, J., Reynolds, A. M., Rice, C. E., Schendel, D., & Windham, G. C. (2007). The epidemiology of autism spectrum disorders. *Annual Review of Public Health*, *28*(1), 235–258. https://doi.org/10.1146/annurev.publhealth.28.021406.144007

Nordström, H., & Laukka, P. (2019). The time course of emotion recognition in speech and music. *The Journal of the Acoustical Society of America*, *145*(5), 3058–3074. https://doi.org/10.1121/1.5108601

Noterdaeme, M., Ullrich, K., & Enders, A. (2017). *Autismus-Spektrum-Störungen (ASS)*. W. Kohlhammer GmbH. https://doi.org/10.17433/978-3-17-026849-4

Nussbaum, C., Schirmer, A., & Schweinberger, S. R. (2022). Contributions of fundamental frequency and timbre to vocal emotion perception and their electrophysiological correlates. *Social Cognitive and Affective Neuroscience*, *17*(12), 1145–1154. https://doi.org/10.1093/scan/nsac033

O'Connor, K. (2012). Auditory processing in autism spectrum disorder: A review. *Neuroscience & Biobehavioral Reviews*, *36*(2), 836–854. https://doi.org/10.1016/j.neubiorev.2011.11.008

Oostenveld, R., Fries, P., Maris, E., & Schoffelen, J.-M. (2011). FieldTrip: Open Source Software for Advanced Analysis of MEG, EEG, and Invasive Electrophysiological Data. *Computational Intelligence and Neuroscience*, *2011*, 1–9. https://doi.org/10.1155/2011/156869

O'Riordan, M., & Passetti, F. (2006). Discrimination in autism within different sensory modalities. *Journal of Autism and Developmental Disorders*, *36*(5), 665–675. https://doi.org/10.1007/s10803-006-0106-1

Ornoy, A., Weinstein- Fudim, L., & Ergaz, Z. (2016). Genetic syndromes, maternal diseases and antenatal factors associated with autism spectrum disorders (ASD). *Frontiers in Neuroscience*, *10*, 316. https://doi.org/10.3389/fnins.2016.00316

Ouimet, T., Foster, N. E. V., Tryfon, A., & Hyde, K. L. (2012). Auditory-musical processing in autism spectrum disorders: A review of behavioral and brain imaging studies.

Annals of the New York Academy of Sciences, *1252*(1), 325–331. https://doi.org/10.1111/j.1749-6632.2012.06453.x

Patel, S., Scherer, K. R., Björkner, E., & Sundberg, J. (2011). Mapping emotions into acoustic space: The role of voice production. *Biological Psychology*, *87*(1), 93–98. https://doi.org/10.1016/j.biopsycho.2011.02.010

Paulmann, S., & Kotz, S. A. (2008). Early emotional prosody perception based on different speaker voices. *NeuroReport*, *19*(2), 209. https://doi.org/10.1097/WNR.0b013e3282f454db

Pell, M. D., & Kotz, S. A. (2011). On the time course of vocal emotion recognition. *PLoS ONE*, *6*(11), e27256. https://doi.org/10.1371/journal.pone.0027256

Pell, M. D., Rothermich, K., Liu, P., Paulmann, S., Sethi, S., & Rigoulot, S. (2015). Preferential decoding of emotion from human non-linguistic vocalizations versus speech prosody. *Biological Psychology*, *111*, 14–25. https://doi.org/10.1016/j.biopsycho.2015.08.008

Peppé, S., McCann, J., Gibbon, F., O'Hare, A., & Rutherford, M. (2007). Receptive and expressive prosodic ability in children with high-functioning autism. *Journal of Speech, Language, and Hearing Research*, *50*(4), 1015–1028. https://doi.org/10.1044/1092-4388(2007/071)

Philip, R. C. M., Whalley, H. C., Stanfield, A. C., Sprengelmeyer, R., Santos, I. M., Young, A. W., Atkinson, A. P., Calder, A. J., Johnstone, E. C., Lawrie, S. M., & Hall, J. (2010). Deficits in facial, body movement and vocal emotional processing in autism spectrum disorders. *Psychological Medicine*, *40*(11), 1919–1929. https://doi.org/10.1017/S0033291709992364

Picton, T. w., Bentin, S., Berg, P., Donchin, E., Hillyard, S. a., Johnson JR., R., Miller, G. a., Ritter, W., Ruchkin, D. s., Rugg, M. d., & Taylor, M. j. (2000). Guidelines for using human event-related potentials to study cognition: Recording standards and publication criteria. *Psychophysiology*, *37*(2), 127–152. https://doi.org/10.1111/1469-8986.3720127

Posit team. (2025). *RStudio: Integrated Development Environment for R* [Computer software]. Posit software, PBC, Bosten, MA. http://www.posit.co/

Qin, L., Wang, H., Ning, W., Cui, M., & Wang, Q. (2024). New advances in the diagnosis and treatment of autism spectrum disorders. *European Journal of Medical Research*, *29*(1), 322. https://doi.org/10.1186/s40001-024-01916-2

Rajendran, G., & Mitchell, P. (2007). Cognitive theories of autism. *Developmental Review*, *27*(2), 224–260. https://doi.org/10.1016/j.dr.2007.02.001

Rosenblau, G., Kliemann, D., Dziobek, I., & Heekeren, H. R. (2016). Emotional prosody processing in Autism Spectrum Disorder. *Social Cognitive and Affective Neuroscience*, nsw118. https://doi.org/10.1093/scan/nsw118

Rosenhall, U., Nordin, V., Sandström, M., Ahlsén, G., & Gillberg, C. (1999). Autism and hearing loss. *Journal of Autism and Developmental Disorders*, *29*(5), 349–357. https://doi.org/10.1023/A:1023022709710

Russell, J. A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology*, *39*(6), 1161–1178. https://doi.org/10.1037/h0077714

Salari, N., Rasoulpoor, S., Rasoulpoor, S., Shohaimi, S., Jafarpour, S., Abdoli, N., Khaledi-Paveh, B., & Mohammadi, M. (2022). The global prevalence of autism spectrum disorder: A comprehensive systematic review and meta-analysis. *Italian Journal of Pediatrics*, *48*(1), 112. https://doi.org/10.1186/s13052-022-01310-w

Samson, F., Mottron, L., Jemel, B., Belin, P., & Ciocca, V. (2006). Can spectro-temporal complexity explain the autistic pattern of performance on auditory tasks? *Journal of Autism and Developmental Disorders*, *36*(1), 65–76. https://doi.org/10.1007/s10803-005-0043-4

Sauter, D. A., & Eimer, M. (2010). Rapid detection of emotion from human vocalizations. *Journal of Cognitive Neuroscience*, *22*(3), 474–481. https://doi.org/10.1162/jocn.2009.21215

Sauter, D. A., Eisner, F., Ekman, P., & Scott, S. K. (2009). *Universal vocal signals of emotion*. 31st Annual Meeting of the Cognitive Science Society.

Schelinski, S., & von Kriegstein, K. (2019). The relation between vocal pitch and vocal emotion recognition abilities in people with autism spectrum disorder and typical development. *Journal of Autism and Developmental Disorders*, *49*(1), 68–82. https://doi.org/10.1007/s10803-018-3681-z

Scher Lisa, J., & Shyman, E. (2019). Challenging Weak Central Coherence: A brief exploration of neurological evidence from visual processing and linguistic studies in autism spectrum disorder. *Annals of Behavioral Neuroscience*, 136–143. https://doi.org/10.18314/abne.v2i1.1606

Scherer. (1986). Vocal affect expression: A review and a model for future research. *Psychological Bulletin*. https://doi.org/10.1037/0033-2909.99.2.143

Scherer, K. R. (1984). Emotion as a multicomponent process: A model and some cross-cultural data. *Review of Personality & Social Psychology*, *5*, 37–63.

Scherer, K. R., Banse, R., & Wallbott, H. G. (2001). Emotion inferences from vocal expression correlate across languages and cultures. *Journal of Cross-Cultural Psychology*, *32*(1), 76–92. https://doi.org/10.1177/0022022101032001009

Schirmer, A., & Gunter, T. C. (2017). Temporal signatures of processing voiceness and emotion in sound. *Social Cognitive and Affective Neuroscience*, *12*(6), 902–909. https://doi.org/10.1093/scan/nsx020

Schirmer, A., & Kotz, S. A. (2006). Beyond the right hemisphere: Brain mechanisms mediating vocal emotional processing. *Trends in Cognitive Sciences*, *10*(1), 24–30. https://doi.org/10.1016/j.tics.2005.11.009

Schwarzkopf, D. S., Anderson, E. J., Haas, B. de, White, S. J., & Rees, G. (2014). Larger extrastriate population receptive fields in autism spectrum disorders. *Journal of Neuroscience*, *34*(7), 2713–2724. https://doi.org/10.1523/JNEUROSCI.4416-13.2014

Scott, S. K., Sauter, D., & McGettigan, C. (2010). Brain mechanisms for processing perceived emotional vocalizations in humans. In S. M. Brudzynski (Ed.), *Handbook of Behavioral Neuroscience* (Vol. 19, pp. 187–197). Elsevier. https://doi.org/10.1016/B978-0-12-374593-4.00019-X

Shapiro, S. S., & Wilk, M. B. (1965). An analysis of variance test for normality (complete samples). *Biometrika*, *52*(3–4), 591–611. https://doi.org/10.1093/biomet/52.3-4.591

Smithson, M., & Verkuilen, J. (2006). A better lemon squeezer? Maximum-likelihood regression with beta-distributed dependent variables. *Psychological Methods*, *11*(1), 54–71. https://doi.org/10.1037/1082-989X.11.1.54

Spielberger, C. D. (2012). *State-Trait Anxiety Inventory for Adults* [Dataset]. https://doi.org/10.1037/t06496-000

Steele, S., Joseph, R. M., & Tager-Flusberg, H. (2003). Brief Report: Developmental Change in Theory of Mind Abilities in Children with Autism. *Journal of Autism and Developmental Disorders*, *33*(4), 461–467. https://doi.org/10.1023/A:1025075115100

Stewart, M. E., McAdam, C., Ota, M., Peppé, S., & Cleland, J. (2013). Emotional recognition in autism spectrum conditions from voices and faces. *Autism: The International Journal of Research and Practice*, *17*(1), 6–14. https://doi.org/10.1177/1362361311424572

Talwar, V., Zwaigenbaum, L., Goulden, K. J., Manji, S., Loomes, C., & Rasmussen, C. (2012). Lie-telling behavior in children with autism and its relation to False-Belief

understanding. *Focus on Autism and Other Developmental Disabilities*, *27*(2), 122–129. https://doi.org/10.1177/1088357612441828

Tavassoli, T., Hoekstra, R. A., & Baron-Cohen, S. (2014). The Sensory Perception Quotient (SPQ): Development and validation of a new sensory questionnaire for adults with and without autism. *Molecular Autism*, *5*(1), 29. https://doi.org/10.1186/2040-2392-5-29

Teder-Sälejärvi, W. A., Pierce, K. L., Courchesne, E., & Hillyard, S. A. (2005). Auditory spatial localization and attention deficits in autistic adults. *Cognitive Brain Research*, *23*(2), 221–234. https://doi.org/10.1016/j.cogbrainres.2004.10.021

The MathWorks Inc. (2024). *MATLAB version: 24.2 (R2024b)* [Computer software]. https://www.mathworks.com

Uljarevic, M., & Hamilton, A. (2013). Recognition of emotions in autism: A formal meta-analysis. *Journal of Autism and Developmental Disorders*, *43*(7), 1517–1526. https://doi.org/10.1007/s10803-012-1695-5

Van Kleef, G. A. (2009). How Emotions Regulate Social Life: The Emotions as Social Information (EASI) Model. *Current Directions in Psychological Science*, *18*(3), 184–188. https://doi.org/10.1111/j.1467-8721.2009.01633.x

Wallbott, & Scherer. (1986). Cues and channels in emotion recognition. *Journal of Personality and Social Psychology*. https://doi.org/10.1037/0022-3514.51.4.690

Wang, C., & Wang, H. (2024). The growing challenge of autism spectrum disorder: A comprehensive review of etiology, diagnosis, and therapy in children. *All Life*, *17*(1), 2415057. https://doi.org/10.1080/26895293.2024.2415057

Watson, R., Latinus, M., Charest, I., Crabbe, F., & Belin, P. (2014). People-selectivity, audiovisual integration and heteromodality in the superior temporal sulcus. *Cortex*, *50*, 125–136. https://doi.org/10.1016/j.cortex.2013.07.011

Whitehouse, A. J. O., & Bishop, D. V. M. (2008). Do children with autism 'switch off' to speech sounds? An investigation using event-related potentials. *Developmental Science*, *11*(4), 516–524. https://doi.org/10.1111/j.1467-7687.2008.00697.x

Williams, D. M., & Happé, F. (2009). What Did I Say? Versus What Did I Think? Attributing false beliefs to self amongst children with and without autism. *Journal of Autism and Developmental Disorders*, *39*(6), 865–873. https://doi.org/10.1007/s10803-009-0695-6

Williams, Z. J., Abdelmessih, P. G., Key, A. P., & Woynaroski, T. G. (2021). Cortical auditory processing of simple stimuli is altered in autism: A meta-analysis of auditory

evoked responses. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, *6*(8), 767–781. https://doi.org/10.1016/j.bpsc.2020.09.011

Wimmer, H., & Perner, J. (1983). Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition*, *13*(1), 103–128. https://doi.org/10.1016/0010-0277(83)90004-5

Winston, J. S., O'Doherty, J., & Dolan, R. J. (2003). Common and distinct neural responses during direct and incidental processing of multiple facial emotions. *NeuroImage*, *20*(1), 84–97. https://doi.org/10.1016/S1053-8119(03)00303-3

Woodbury-Smith, M. R., Robinson, J., Wheelwright, S., & Baron-Cohen, S. (2005). Screening Adults for Asperger Syndrome Using the AQ:A Preliminary Study of its Diagnostic Validity in Clinical Practice. *Journal of Autism and Developmental Disorders*, *35*(3), 331–335. https://doi.org/10.1007/s10803-005-3300-7

Yang, D. Y.-J., Rosenblau, G., Keifer, C., & Pelphrey, K. A. (2015). An integrative neural model of social perception, action observation, and theory of mind. *Neuroscience & Biobehavioral Reviews*, *51*, 263–275. https://doi.org/10.1016/j.neubiorev.2015.01.020

Zeidan, J., Fombonne, E., Scorah, J., Ibrahim, A., Durkin, M. S., Saxena, S., Yusuf, A., Shih, A., & Elsabbagh, M. (2022). Global prevalence of autism: A systematic review update. *Autism Research*, *15*(5), 778–790. https://doi.org/10.1002/aur.2696

Zhang, M., Xu, S., Chen, Y., Lin, Y., Ding, H., & Zhang, Y. (2022). Recognition of affective prosody in autism spectrum conditions: A systematic review and meta-analysis. *Autism*, *26*(4), 798–813. https://doi.org/10.1177/1362361321995725

9. List of Abbreviations

Abbreviatioin	**Definition**
AC	Auditory Cortex
ASD	Autism Spectrum Disorder
DSM-5	Diagnostic and Statistical Manual of Mental Disorders (5th ed.)
EEG	Electroencephalography
ERP	Event-Related Potential
fMRI	Functional magnetic resonance imaging
F0	Fundamental frequency
NT	Neurotypical
pAC	Primary Auditory Cortex
STS	Superior Temporal Sulcus
ToM	Theory of Mind
WCC	Theory of Weak Central Coherence

10. Table of Tables

11. Table of Figures

12. Appendix

12.1. Appendix A: Abstract

Abstract - English

Auditory emotion processing in autism spectrum disorder (ASD) has received less attention than visual emotion processing, and existing findings are mixed regarding both behavioural performance and neural mechanisms. This study investigated auditory emotion processing in autistic and neurotypical individuals by combining behavioural measures with electrophysiological indices of early auditory processing. Participants first listened to brief, nonverbal emotional vocalisations (happy, angry, fearful, sad) while EEG was recorded. In a subsequent task, they identified the expressed emotion and rated its perceived intensity. Analyses focused on the P200 event-related potential (ERP) component as an index of early sensory–affective processing.

Behaviourally, the autistic group showed lower overall recognition accuracy and reduced perceived emotional intensity compared to neurotypical participants, with group differences in accuracy particularly pronounced for fearful vocalisations. Both groups, however, demonstrated comparable differentiation between emotion categories in their intensity ratings. At the neural level, P200 amplitudes were modulated by emotion but did not differ between groups, suggesting largely intact early sensory–affective encoding in ASD. P200 latencies revealed an emotion-specific group difference, with delayed responses to angry vocalisations in the ASD group. Moreover, P200 latency was negatively associated with perceived emotional intensity in the ASD group only, indicating altered brain–behaviour coupling.

These findings suggest that auditory emotion processing differences in autism are less likely to reflect fundamental early sensory deficits and may instead arise during the integration of early auditory signals into subjective emotional experience.

Abstract – German

Die auditive Emotionsverarbeitung bei Autismus-Spektrum-Störungen (ASS) hat bislang weniger Aufmerksamkeit erhalten als die visuelle Emotionsverarbeitung, und die bisherigen Befunde sind sowohl hinsichtlich behavioraler Ergebnisse als auch neuronaler Mechanismen uneinheitlich. Die vorliegende Masterarbeit untersuchte die auditive Emotionsverarbeitung bei autistischen und neurotypischen Personen durch die Kombination behavioraler und elektrophysiologischer Maße. Die Teilnehmenden hörten zunächst kurze,

nonverbale emotionale Vokalisierungen (freudig, wütend, ängstlich, traurig), während ein Elektroenzephalogramm (EEG) aufgezeichnet wurde. In einer anschließenden Aufgabe identifizierten sie die dargestellte Emotion und bewerteten deren wahrgenommene Intensität. Die elektrophysiologischen Analysen fokussierten auf die P200-Komponente ereigniskorrelierter Potenziale (EKPs) als Index früher sensorisch-affektiver Verarbeitung.

Behavioral zeigte die ASS-Gruppe eine geringere Genauigkeit in der Emotionserkennung sowie niedrigere Intensitätsbewertungen im Vergleich zur neurotypischen Gruppe, wobei die Unterschiede in der Genauigkeit besonders bei ängstlichen Lautäußerungen ausgeprägt waren. Beide Gruppen differenzierten jedoch vergleichbar zwischen den Emotionskategorien. Auf neuronaler Ebene wurden die P200-Amplituden durch Emotion moduliert, unterschieden sich jedoch nicht zwischen den Gruppen, was auf eine weitgehend intakte frühe sensorisch-affektive Verarbeitung bei ASS hindeutet. Die P200-Latenzen zeigten hingegen emotionsspezifische Gruppenunterschiede, mit verzögerten Reaktionen auf wütende Vokalisierungen in der ASS-Gruppe. Zudem bestand nur in der ASS-Gruppe ein signifikanter Zusammenhang zwischen P200-Latenz und wahrgenommener Intensität.

Insgesamt sprechen die Ergebnisse dafür, dass Unterschiede in der auditiven Emotionsverarbeitung bei ASS weniger auf grundlegende frühe sensorische Defizite zurückzuführen sind, sondern vielmehr auf Unterschiede in der Integration früher auditiver Signale in die subjektive Emotionswahrnehmung.

12.2. Appendix B: Assumption Testing

Table 8

Assumption Checks for Behavioural and Electrophysiological Measures

Measure	Assumption Tested	Test	Statistic	*p*
AQ score	Normality	Shapiro–Wilk	$W = 0.985$	.849
AQ score	Homogeneity	Levene's	$F = 1.26$	.268
Accuracy rate	Normality	Shapiro–Wilk	$W = 0.938$	< .001
Accuracy rate	Homogeneity	Levene's	$F = 0.37$	.545
Intensity rating	Normality	Shapiro–Wilk	$W = 0.985$	.064
Intensity rating	Homogeneity	Levene's	$F = 0.10$	.760
P200 amplitude	Normality	Shapiro–Wilk	$W = 0.987$	.163
P200 amplitude	Homogeneity	Levene's	$F = 1.40$	.238
P200 latency	Normality	Shapiro–Wilk	$W = 0.984$	.073
P200 latency	Homogeneity	Levene's	$F = 0.59$	.442

Note. Shapiro–Wilk tests were used to assess normality of residuals, and Levene's tests were used to assess homogeneity of variances between groups. Visual inspections of histograms and Q–Q plots were additionally conducted for all measures (see Appendix Figures X–Y). Accuracy rates violated the assumption of normality and were therefore analyzed using an aligned rank transform procedure. All other measures met assumptions for parametric analyses.

Figure 11

Q-Q Plot of Residuals for AQ scores

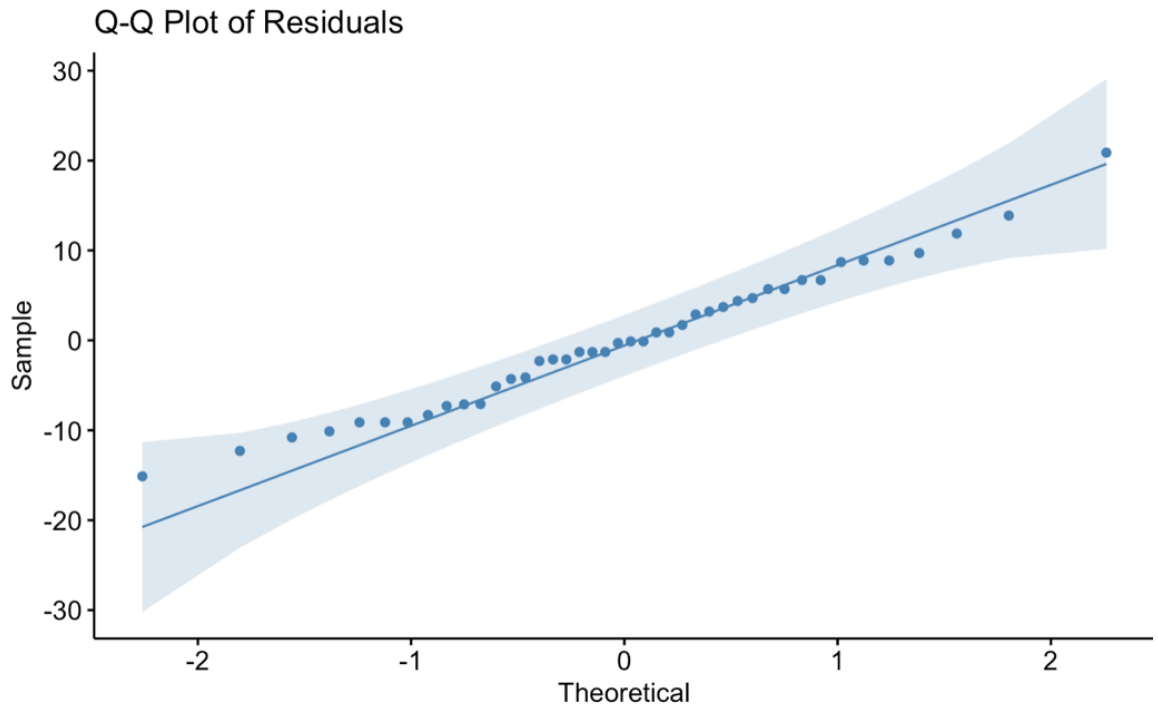


Figure 12

Q-Q Plot of Residuals for Accuracy Rates

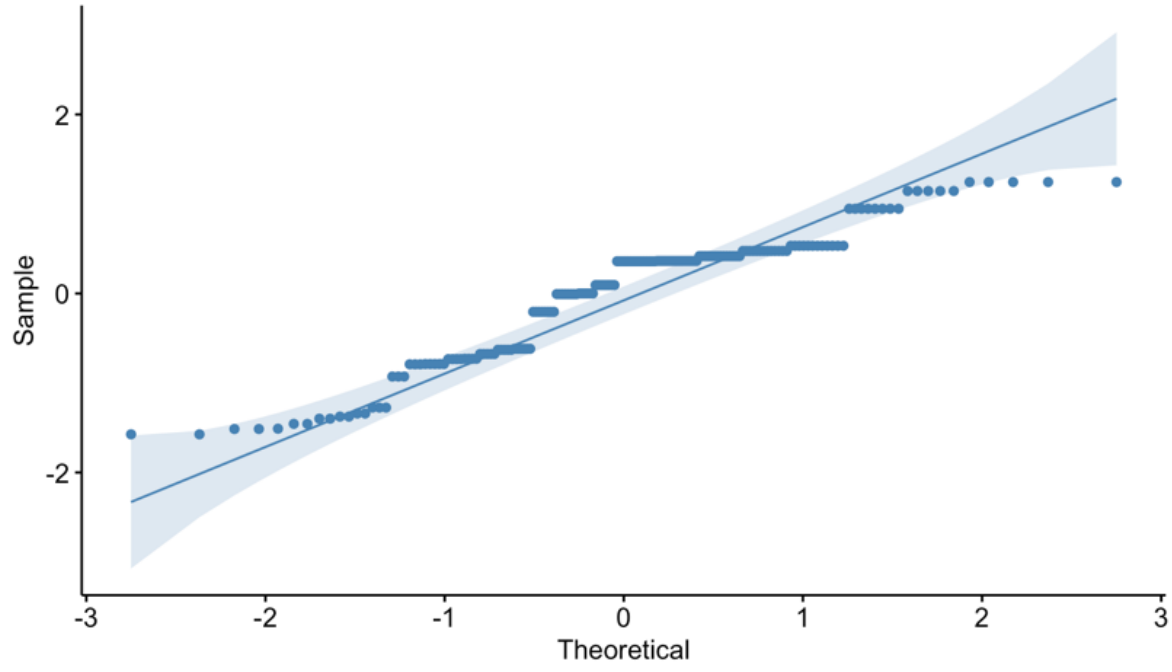


Figure 13

Q-Q Plot of Residuals for P200 Amplitudes

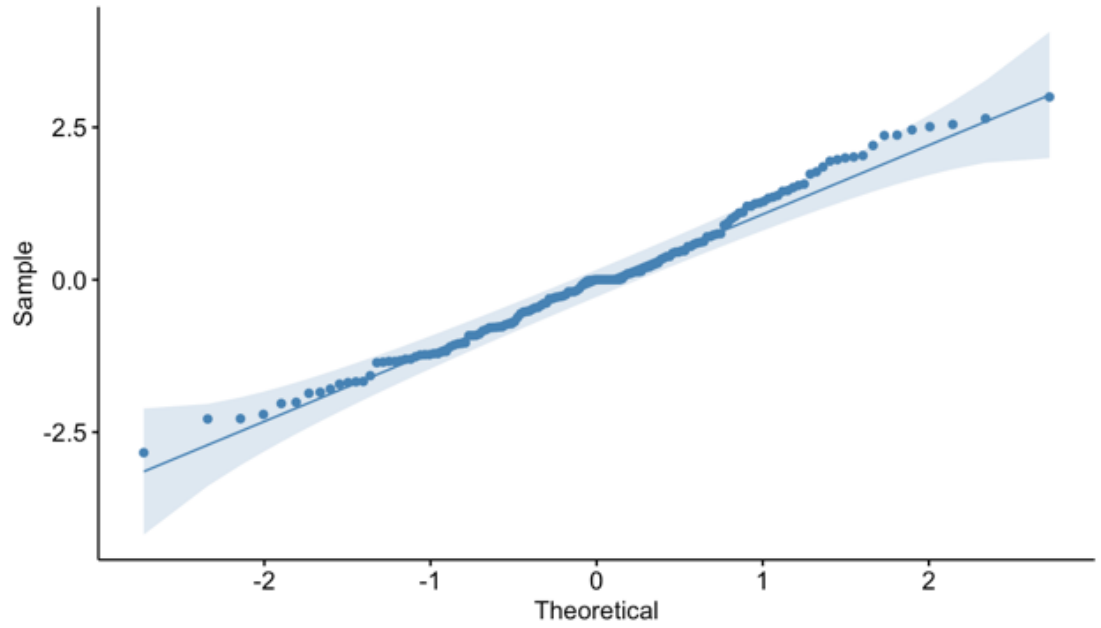
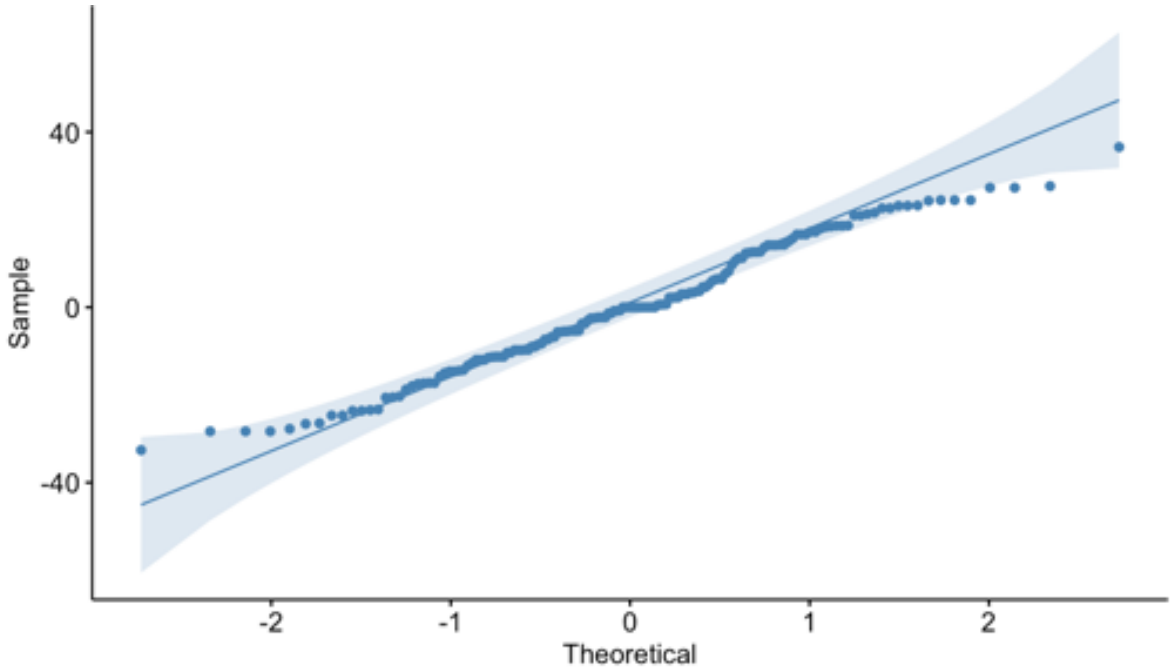


Figure 14

Q-Q Plot of Residuals for P200 Latencies



12.3. Appendix C: Outlier Cleaning and Imputation

Accuracy Rates

Table 9

Flagged Outlier for Accuracy Rates

Group	Emotion	N flagged	IDs	Cook's D cutoff
ASD	a	2	28, 30	0.0238
ASD	h	1	20	0.0238
ASD	s	1	12	0.0238
NT	f	1	27	0.0238
NT	h	1	9	0.0238

Note. Rates are Logit – Transformed

Table 10

Descriptive Accuracy Rates Before vs. After Cleaning

Group	Emotion	Raw Data: M (SD)	Cleaned Data: M (SD)
ASD	a	0.85 (1.01)	1.05 (0.79)
ASD	f	0.95 (0.86)	0.95 (0.86)
ASD	h	1.7 (0.9)	1.84 (0.65)
ASD	s	1.54 (0.9)	1.66 (0.68)
NT	a	1.25 (0.87)	1.25 (0.87)
NT	f	1.62 (0.81)	1.72 (0.68)
NT	h	1.64 (0.91)	1.78 (0.66)
NT	s	1.83 (0.61)	1.83 (0.61)

Note. Rates are Logit-Transformed

Intensity Ratings

Table 11

Flagged Outlier for Intensity Ratings

Group	Emotion	N flagged	IDs	Cook's D cutoff
ASD	a	3	20, 28, 36	0.0238
ASD	f	4	12, 20, 24, 26	0.0238
ASD	h	1	20	0.0238
NT	f	1	27	0.0238
NT	h	2	9, 27	0.0238

Table 12

Descriptive Intensity Ratings Before vs. After Cleaning

Group	Emotion	Raw Data: M (SD)	Cleaned Data: M (SD)
ASD	a	4.23 (2.11)	4.4 (1.39)
ASD	f	5.01 (2.1)	5.44 (1.16)
ASD	h	5.78 (1.55)	6.02 (1.09)
ASD	s	5.53 (1.78)	5.53 (1.78)
NT	a	5.3 (1.66)	5.3 (1.66)
NT	f	6.26 (1.89)	6.53 (1.41)
NT	h	6.19 (1.87)	6.62 (1.29)
NT	s	6.49 (1.84)	6.71 (1.53)

P200 ERP Component

Table 13

Flagged Outlier for P200 Amplitudes

Group	Emotion	N flagged	IDs	Cook's D cutoff
ASD	a	2	18, 22	0.02564
ASD	h	2	18, 32	0.02564
ASD	f	1	32	0.02564
NT	a	3	27, 31, 39	0.02564
NT	h	1	39	0.02564
NT	s	1	39	0.02564

Table 14

Descriptive P200 Amplitudes Before vs. After Cleaning

Group	Emotion	Raw Data: M (SD)	Cleaned Data M (SD)
ASD	a	4.33 (2.02)	4.37 (1.48)
ASD	f	4.36 (1.56)	4.19 (1.38)
ASD	h	4.23 (1.48)	3.84 (0.89)
ASD	s	3.27 (1.25)	3.27 (1.25)
NT	a	4.03 (1.97)	3.87 (1.31)
NT	f	3.98 (1.13)	3.98 (1.13)
NT	h	4.11 (1.59)	3.84 (1.06)
NT	s	3.35 (1.27)	3.18 (1.04)

Table 15

Flagged Outlier for P200 Latencies

Group	Emotion	N flagged	IDs	Cook's D cutoff
ASD	a	3	10, 24, 42	0.02564
ASD	h	1	4	0.02564
ASD	f	3	4, 10, 24	0.02564
NT	a	1	25	0.02564
NT	h	1	37	0.02564
NT	s	2	13, 37	0.02564

Table 16

Descriptive P200 Latencies Before vs. After Cleaning

Group	Emotion	Mean_raw (SD)	Mean_cleaned(SD)
ASD	a	247.56 (26.3)	244.06 (18.26)
ASD	f	229.88 (22.12)	226.44 (11.78)
ASD	h	224.79 (17.45)	227.23 (13.63)
ASD	s	230.78 (16.48)	230.78 (16.48)
NT	a	230.28 (20.11)	227.16 (14.83)
NT	f	225.8 (16.56)	225.8 (16.56)
NT	h	226.76 (18.33)	228.94 (15.67)
NT	s	235.11 (18.65)	234.72 (12.29)

12.4. Appendix D: Additional Descriptives & Statistical Testing

Accuracy Rates

Table 17

Pairwise Emotion Contrasts Within Groups for Accuracy Rates

Group	contrast	estimate	SE	p.value	95% Confidence Interval for Difference		Cohen's d
					Lower Bound	Upper Bound	
ASD	a - f	6,238	12,563	1,000	-18,63	31,11	0,045
	a - h	-41,619	12,563	0,031	-66,49	-16,74	-0,302
	a - s	-31,048	12,563	0,253	-55,92	-6,17	-0,226
	f - h	-47,857	12,563	0,006	-72,73	-22,98	-0,348
	f - s	-37,286	12,563	0,076	-62,16	-12,41	-0,271
	h - s	10,571	12,563	1,000	-14,30	35,45	0,077
NT	a - f	-24,857	12,563	0,752	-49,73	0,02	-0,181
	a - h	-28,381	12,563	0,411	-53,26	-3,51	-0,206
	a - s	-32,143	12,563	0,223	-57,02	-7,27	-0,234
	f - h	-3,524	12,563	1,000	-28,40	21,35	-0,026
	f - s	-7,286	12,563	1,000	-32,16	17,59	-0,053
	h - s	-3,762	12,563	1,000	-28,64	21,11	-0,027

Note. Estimated marginal mean differences reflect differences in accuracy between emotion conditions, tested separately within each group using aligned-rank contrast tests (ART-C). Negative values indicate lower accuracy in the ASD group.

Intensity Ratings

Table 18

Pairwise Comparisons of Intensity Rates for Emotion (Main Effect)

contrast	estimate	SE	p.value	95% Confidence Interval for Difference		Cohen's d
				Lower Bound	Upper Bound	
A - F	-1,135	0,264	0,001	-1,843	-0,428	-1.360
A - H	-1,473	0,277	< 0.001	-2,216	-0,73	-1.680
A - S	-1,274	0,251	< 0.001	-1,946	-0,602	-1.607
F - H	-0,338	0,234	0,482	-0,966	0,29	-0.456
F - S	-0,139	0,251	0,945	-0,813	0,535	-0.175
H - S	0,199	0,26	0,87	-0,499	0,896	0.242

Note. Pairwise comparisons are based on estimated marginal means (EMMs) from the mixed ANOVA model. *p*-values are Holm-adjusted.

Table 19

Pairwise Comparisons of Intensity Ratings Between Groups by Emotion

contrast	emotion	estimate	SE	p.value	95% Confidence Interval for Difference		Cohen's d
					Lower Bound	Upper Bound	
ASD - NT	a	-0,895	0,471	0,065	-1,848	0,057	-0.601
	f	-1,095	0,399	0,009	-1,901	-0,288	-0.867
	h	-0,601	0,368	0,11	-1,344	0,142	-0.517
	s	-1,177	0,512	0,027	-2,212	-0,141	-0.726

Note. Pairwise comparisons are based on estimated marginal means (EMMs) from the mixed ANOVA model. *p*-values are Holm-adjusted. Negative estimate values indicate lower intensity ratings in the ASD group.

Table 20

Pairwise Emotion Contrasts of Intensity Ratings Within Groups

Group	contrast	estimate	SE	p.value	95% Confidence Interval for Difference		Cohen's d
					Lower Bound	Upper Bound	
ASD	a - f	-1,035	0,373	0,04	-2,036	-0,035	-0.877
	a - h	-1,62	0,392	0,001	-2,671	-0,569	-1.306
	a - s	-1,133	0,354	0,014	-2,083	-0,183	-1.0111
	f - h	-0,585	0,331	0,305	-1,473	0,304	-0.558
	f - s	-0,098	0,356	0,993	-1,051	0,855	-0.087
	h - s	0,487	0,368	0,554	-0,5	1,473	0.418
NT	a - f	-1,235	0,373	0,01	-2,235	-0,234	-1.046
	a - h	-1,326	0,392	0,008	-2,377	-0,275	-1.069
	a - s	-1,415	0,354	0,001	-2,365	-0,465	-1.262
	f - h	-0,091	0,331	0,993	-0,98	0,797	-0.087
	f - s	-0,18	0,356	0,957	-1,133	0,773	-0.160
	h - s	-0,089	0,368	0,995	-1,075	0,897	-0.076

Note. Estimated marginal mean differences reflect differences in intensity ratings between emotion conditions, tested separately within each group.

P200 ERPs

Table 21

Descriptives of P200 Amplitudes and Latencies by Group and Emotion

	Emotion	Group	M	SD	Min	Max	95% CI
P200 Amplitudes (μV)	Angry	ASD	4.37	1.48	1.53	6.73	[3.68, 5.06]
		NT	3.87	1.31	1.58	6.24	[3.24, 4.50]
	Fearful	ASD	4.19	1.38	2.18	6.84	[3.54, 4.84]
		NT	3.98	1.13	2.41	6.44	[3.44, 4.52]
	Happy	ASD	3.84	0.89	2.00	5.88	[3.42, 4.26]
		NT	3.84	1.06	1.56	6.39	[3.33, 4.35]
	Sad	ASD	3.27	1.25	1.24	6.26	[2.68, 3.86]
		NT	3.18	1.04	1.47	5.18	[2.68, 3.68]
P200 Latencies (ms)	Angry	ASD	244.06	18.26	211.46	271.37	[235.51, 252.61]
		NT	227.16	14.83	202.48	250.4	[220.01, 234.31]
	Fearful	ASD	226.44	11.78	208.47	247.41	[220.93, 231.95]
		NT	225.8	16.56	202.48	262.38	[217.82, 233.78]
	Happy	ASD	227.23	13.63	199.48	250.4	[220.85, 233.61]
		NT	228.94	15.67	202.48	253.4	[221.39, 236.49]
	Sad	ASD	230.78	16.48	202.48	253.4	[223.07, 238.49]
		NT	234.72	12.29	217.45	262.38	[228.80, 240.64]

Table 22

ANOVA Results on P200 Amplitudes

Effect	F	df	Df.res	*p*
group	0,403	1	37	.530
emotion	10,038	2,543	94,1	<.001
group × emotion	0,692	2,543	94,1	.536

Table 23

Pairwise Contrasts for P200 Amplitude by Emotion (Main Effect)

contrast	estimate	SE	p.value	95% Confidence Interval for Difference		Cohen's d
				Lower Bound	Upper Bound	
A - F	0,031	0,229	0,999	-0,585	0,647	0.044
A - H	0,275	0,188	0,471	-0,232	0,782	0.480
A - S	0,894	0,198	0.0003	0,362	1,426	1.487
F - H	0,244	0,161	0,436	-0,188	0,677	0.500
F - S	0,864	0,162	<.0001	0,428	1,3	1.752
H - S	0,619	0,163	0,003	0,18	1,058	1.248

Note. Pairwise comparisons are based on estimated marginal means (EMMs) from the mixed ANOVA model. *p*-values are Holm-adjusted.

Table 24

ANOVA Results on P200 Latencies

Effect	F	df	Df.res	*p*
group	0,843	1	37	0,364
emotion	4,359	2,654	98,21	0,009
group × emotion	5,219	2,654	98,21	0,003

Table 25

Pairwise Contrasts for P200 Latency by Emotion (Main Effect)

contrast	estimate	SE	p.value	95% Confidence Interval for Difference		Cohen's d
				Lower Bound	Upper Bound	
A - F	9,493	3,435	0,042	0.226	1.580	0.909
A - H	7,53	3,221	0,108	0.096	1.431	0.769
A - S	2,86	2,877	0,754	-0.324	0.973	0.327
F - H	-1,963	2,581	0,871	-0.895	0.398	-0.250
F - S	-6,633	2,923	0,124	-1.408	-0.075	-0.746
H - S	-4,67	2,425	0,235	-1.289	0.031	-0.633

Note. Pairwise comparisons are based on estimated marginal means (EMMs) from the mixed ANOVA model. *p*-values are Holm-adjusted.