University of Reading

Speech-in-Noise Processing in Autism Spectrum Disorder: Electrophysiological and Behavioural Evidence

A thesis submitted in fulfilment of the requirements for the degree of Doctor of Philosophy

School of Psychology and Clinical Language Sciences

Jiayin Li

January 2026

Abstract

Recognising speech in the presence of competing sounds requires listeners to effectively process both acoustic and semantic information. Autistic individuals often experience greater difficulties with speech-in-noise (SiN) processing, but the underlying mechanisms remain unclear. Existing electrophysiological (EEG) research has primarily focused on auditory-level processing in autism and frequently relies on short, discrete stimuli, such as tones or single words. As a result, higher-level semantic processing and its interaction with auditory mechanisms during continuous speech remains largely unexplored. This thesis addresses these gaps through three studies combining EEG and behavioural measures to investigate SiN processing in autistic and non-autistic adults. Study 1 evaluated listeners' use of acoustic cues (i.e., difference in speaker gender and spatial location) to attend to a speaker amid competing voices. While both groups benefited from these cues, autistic participants showed lower accuracy and smaller improvements over time. Study 2 examined the impact of background music on semantic processing using a sentence acceptability task. Autistic participants showed lower accuracy and attenuated N400 responses to semantic incongruities, reflecting difficulties with semantic integration. Unlike non-autistic participants, whose behavioural and neural responses varied with lyric intelligibility, autistic participants showed minimal variation across conditions. Study 3 explored auditory and semantic processing in the presence of intelligible or unintelligible background speech using the same task. Autistic participants showed reduced auditory encoding, as captured by EEG-derived temporal response functions, and delayed semantic processing reflected by N400 latency. While non-autistic participants adjusted their auditory and semantic processing according to the intelligibility of the background speech, autistic participants showed no such modulation. Overall, these findings offer new insight into SiN processing in autism, highlighting inefficient auditory and semantic processing, and a consistent lack of modulation in response to different noise types. This reduced flexibility in adapting to complex listening conditions indicates broader differences in processing strategies.

Acknowledgements

I would like to express my deepest gratitude to my supervisors, Fang Liu and Ian Cunnings. I feel incredibly fortunate to have had Fang as my primary supervisor. She has supported me in every aspect of my PhD, from shaping the academic direction of my work to offering practical advice and detailed feedback on manuscripts. Her commitment, generosity with her time, and guidance in both research and professional development have been invaluable. Beyond her academic mentorship, she also provided emotional support, showing patience and encouragement whenever I faced challenges. My sincere thanks also go to Ian, whose expertise and thoughtful feedback enriched this thesis. He was always willing to answer questions, however many I had, and I could rely on him for constructive guidance that improved both my research and writing. His perspectives often helped me see issues in a new light, and I greatly valued his generosity with time and support throughout this journey.

I would also like to thank Maleeha Sujawal, Zivile Bernotaite, and Jess Akhurst, whose support with participant recruitment and data collection was invaluable. Their dedication and efficiency made the process much smoother. I am grateful as well to Stuart Rosen, Paul Iverson, and Susanne Brouwer for generously providing materials and guidance on their use, which greatly supported the design and implementation of this research. I am particularly thankful to Paul for sharing insights into data analysis that helped shape my approach. My thanks also go to Anna Petrova and Jia Hoong Ong, who helped me settle in and offered advice that eased my transition into doctoral research.

I am profoundly grateful to the South East Network for Social Sciences (SeNSS) studentship, which funded my tuition fees and provided generous support for training and conference attendance. My data collection was supported jointly by SeNSS and Fang's European Research Council grant (ERC 678733, CAASD). I am also sincerely thankful to all the participants who generously gave their time to my studies, as their involvement made this research possible.

Finally, I'd like to express heartfelt thanks to my husband, whose love, encouragement, and companionship sustained me throughout this journey. As a researcher himself, he has been an inspiration, and I have learned a lot from him. I am also deeply grateful to my family for their patience, love, and belief in me, which carried me through the most difficult times.

Declaration: I confirm that this is my own work and the use of all material from other sources has been properly and fully acknowledged.

Jiayin Li

Declaration of Authorship

I declare that the three studies presented in this thesis are my original work, conducted in collaboration with my supervisors and co-authors. These studies have been published or submitted, with myself as the first author on all papers.

Study 1

Li, J., Sujawal, M., Bernotaite, Z., Cunnings, I., & Liu, F. (2025). Listening in a noisy world: The impact of acoustic cues and background music on speech perception in autism. *Autism*, 13623613251376484. DOI: https://doi.org/10.1177/13623613251376484

Study 2

Li, J., Sujawal, M., Bernotaite, Z., Cunnings, I., & Liu, F. (2026). Semantic Processing in Autism During Speech-in-Music Listening: Insights From Congruency and Surprisal-Based N 400 Analyses. *Psychophysiology*, *63*(1), e70232. DOI: https://doi.org/10.1111/psyp.70232

Study 3

Li, J., Sujawal, M., Bernotaite, Z., Cunnings, I., & Liu, F. (2025). Auditory and Semantic Processing of Speech-in-Noise in Autism: A Behavioral and EEG Study. *Autism Research*. DOI: https://doi.org/10.1002/aur.70097

Table of Contents

List of Tables

Chapter 2

Chapter 3

Chapter 4

List of Figures

Chapter 2

Chapter 3

Chapter 4

Chapter 1

General Introduction

Autism spectrum disorder (ASD) is a neurodevelopmental condition characterised by early-emerging difficulties in social communication and interaction, alongside restricted and repetitive patterns of behaviour, interests, and atypical sensory sensitivities (American Psychiatric Association, 2013). Although not part of the core diagnostic criteria, many autistic children experience significant delays in expressive and receptive language, and in some cases, these challenges persist into their adulthood (Brignell et al., 2018). Even among those with age-appropriate language abilities, atypical processing has been observed at both auditory and semantic levels (Gonçalves & Monteiro, 2023; Key & D'Ambrose Slaboch, 2021). These individuals often show heightened sensitivity to specific acoustic features, yet reduced efficiency in integrating broader contextual information, suggesting differences in higher-level cognitive mechanisms involved in speech processing (see Section 1.1).

Recent discussions in autism science have highlighted the importance of how these processing differences are conceptualised and interpreted. Pellicano and den Houting (2022) argue that autism research has traditionally been guided by a medical model, in which autistic characteristics are often described as deficits relative to neurotypical norms. They propose a shift towards a neurodiversity paradigm, which treats autism as a form of natural variation in human neurodevelopment. Importantly, this perspective does not deny the presence of cognitive or neurodevelopmental differences, nor does it reject experimental or mechanistic approaches. Instead, it cautions against interpreting departures from neurotypical patterns as inherently undesirable and encourages a more neutral interpretation of empirical findings. From this perspective, greater emphasis is placed on the role of environmental context in shaping everyday experiences. Drawing on the social model of disability, neurodiversity-informed accounts suggest that many challenges associated with autism emerge through interactions between individuals and their social and physical environments, rather than arising solely from individual characteristics (Pellicano & den Houting, 2022).

Accordingly, this thesis takes a neurodiversity-informed perspective. It examines differences in speech processing without assuming that atypical neural or behavioural patterns are

inherently suboptimal, and it focuses on how autistic individuals process speech in everyday listening environments, where multiple sound sources compete for attention. Situations such as workplace meetings, busy cafés, or social gatherings often involve overlapping voices and background noise, making it more difficult to focus on and understand the target speech (Bronkhorst, 2000; Cherry, 1953). Successfully navigating these situations requires the ability to separate relevant speech from competing sounds (Bregman, 1990; Griffiths & Warren, 2004; Woods & Colburn, 1992). This process, known as speech-in-noise (SiN) recognition, draws on both bottom-up processing of acoustic features and top-down cognitive and linguistic mechanisms that support attention, prediction, and comprehension (see Section 1.2).

While neurotypical individuals with normal hearing can generally integrate bottom-up and top-down processes flexibly, autistic individuals often exhibit measurable behavioural difficulties in noisy environments, including slower response times and reduced accuracy (Ruiz Callejo & Boets, 2023). Qualitative research further indicates that such challenges are common in daily life. Many autistic individuals describe feeling overwhelmed, disconnected, or excluded in complex auditory settings, which can contribute to heightened stress, reduced social engagement, and a diminished sense of inclusion (Sturrock et al., 2022; Bendo et al., 2024). These accounts underscore the need for experimental paradigms that reflect the complexity of real-world communication and clarify the cognitive and neural mechanisms underlying listening challenges in autism.

Although neuropsychological research on SiN processing remains limited, existing studies have identified differences in both bottom-up auditory encoding and top-down cognitive mechanisms (see Section 1.3). However, much of this work has relied on traditional event-related potential (ERP) paradigms using repeated presentations of discrete tones or syllables, which do not reflect the dynamic and context-dependent nature of speech. In addition, previous studies largely employed controlled, laboratory-based forms of background noise (e.g., amplitude-modulated noise or multi-speaker babble) (Sturrock et al., 2022), while other ecologically valid sound sources, such as music, have received little attention. As a result, it remains unclear how autistic individuals process continuous speech in realistic auditory environments and how they integrate information across linguistic and cognitive levels.

This thesis aims to address these gaps by investigating auditory and semantic processing of speech in the presence of various background sounds, using both behavioural and EEG

methods. Three empirical studies are conducted, each targeting different aspects of speech processing across listening scenarios that vary in acoustic, linguistic, and cognitive complexity.

Study 1 investigates whether listeners can use differences in speaker characteristics (i.e., voice pitch and spatial location) to identify and follow a target speaker when two voices are presented simultaneously. While previous studies have examined similar questions by analysing mean accuracy across trials (Emmons et al., 2022; Lau et al., 2022), the present study also models trial-by-trial performance using Generalised Additive Mixed Models, enabling a more fine-grained analysis of behavioural fluctuations over time. To further mimic real-world listening conditions, the study introduces instrumental music into the listening scenario to examine its effect on speech recognition in autistic and non-autistic participants, representing the first investigation of background music in competing speech scenarios.

Study 2 extends the investigation of background music by examining whether vocal features influence semantic processing during speech comprehension. Specifically, it focuses on the presence and intelligibility of linguistic content by comparing three types of background music: instrumental music without lyrics, music with intelligible English lyrics, and music with unintelligible lyrics. Participants hear target sentences presented concurrently with background music and judge their semantic congruency while EEG is recorded. Semantic integration is assessed using both behavioural accuracy and the N400 response to semantic incongruency, a well-established neural marker of meaning processing. This study offers new insight into the neural dynamics of semantic comprehension in the presence of background music in autistic and non-autistic individuals.

Study 3 further examines the interplay between auditory encoding and semantic integration during competing speech recognition. Semantic processing is indexed using the N400, while an advanced neural tracking method is employed to measure cortical tracking of acoustic features over time. Unlike traditional ERP approaches used by previous studies, neural tracking enables the analysis of continuous speech and captures how acoustic information unfolds dynamically in the brain (Crosse et al., 2016). Previous evidence suggests that non-autistic listeners adapt their listening strategies depending on the type of background noise, relying more on acoustic decoding in less structured conditions and more on the linguistic context of the target speech when the interference is intelligible (e.g., Song et al., 2020). Based on this, the study investigates whether autistic listeners show similar flexibility. To test this, masker

complexity is manipulated to introduce varying levels of acoustic and linguistic interference. This design allows for the assessment of how participants coordinate multiple levels of processing under different listening demands.

Collectively, these studies offer a systematic and ecologically grounded investigation of how autistic and non-autistic individuals process speech under diverse and realistic auditory conditions. By integrating behavioural and neural measures across auditory and semantic processing, this thesis advances understanding of the perceptual and cognitive mechanisms that support SiN comprehension in autism. To contextualise these research aims, the rest of this chapter reviews relevant empirical work. Section 1.1 reviews findings on auditory and semantic processing of speech in the absence of background noise, highlighting baseline differences in autism and interpreting these patterns in light of existing theoretical accounts of cognitive and perceptual processing. Section 1.2 provides a broader overview of key concepts in SiN processing, including types of masking, available cues, and the cognitive mechanisms that support effective cue use. Section 1.3 introduces current evidence on SiN processing in autism, focusing on behavioural and neural findings. Finally, Section 1.4 discusses limitations in the existing literature and explains how the present thesis addresses these gaps by outlining the design and main findings of the three empirical studies.

1.1 Speech processing in autism

Speech perception is a hierarchically organised and dynamically adaptive process, shaped by the continuous interplay between bottom-up sensory input and top-down linguistic and cognitive mechanisms (Hickok & Poeppel, 2007). Early auditory processing involves the extraction of temporal and spectral features from the acoustic signal, which are mapped onto phonetic and prosodic representations (Giraud & Poeppel, 2012; Poeppel et al., 2008). These features support access to word meaning by activating possible word candidates based on phonemic cues, primarily through feedforward processing routes (Norris, 1994, 1999; Norris & McQueen, 2008). However, successful comprehension often requires more than bottom-up decoding. Interactive models propose that lexical knowledge can feed back to influence phoneme-level processing, enabling top-down information to resolve ambiguity in real time (McClelland & Elman, 1986). Predictive coding frameworks further emphasise the role of dynamic top-down predictions in shaping perception according to incoming input (Blank & Davis, 2016; Sohoglu et al., 2012). These bidirectional processes support flexible and context-

sensitive comprehension, especially under adverse listening conditions such as background noise or signal degradation (Mattys et al., 2012; Shannon et al., 1995).

Several theories have been proposed to account for atypical speech processing in autism. The Enhanced Perceptual Functioning model (Mottron et al., 2006; Mottron & Burack, 2001) posits that autistic individuals show superior access to low-level perceptual information, such as pitch, reflecting enhanced bottom-up processing. Although global integration is not necessarily impaired, it may be less spontaneously activated, which can affect the mapping of perceptual inputs onto higher-order linguistic representations. In contrast, the Weak Central Coherence framework was originally proposed as a reduced tendency to extract global meaning or gestalt from incoming information (Frith, 1989). This account was subsequently extended to language, emphasising reduced reliance on contextual information during meaning resolution, such as the use of sentence context to disambiguate words or interpret discourse (Frith & Happé, 1994). Building on a large body of empirical work, Happé and Frith (2006) argued that WCC is best understood as a detail-focused cognitive processing bias rather than a core difference in global integration. Specifically, they proposed that individuals on the autism spectrum show a robust tendency to prioritise local information by default, while global or contextual processing remains available and can be effectively engaged when task demands explicitly cue or require integration. Applied to speech processing, this theory predicts reduced spontaneous weighting of higher-order contextual cues, such as prosodic and semantic information.

Predictive coding has also been applied to autism, offering a computational account of differences in perceptual and integrative processing. In typical perception, the brain continuously updates internal models to minimise prediction error (Clark, 2013; Feldman & Friston, 2010; Friston, 2009). Autism-specific accounts propose that this process is disrupted by atypical precision weighting, leading to an altered balance between sensory input and prior expectations. This imbalance may result in reduced sensitivity to linguistic predictability and diminished neural adaptation, whereby predictable stimuli fail to elicit the usual attenuation (Lawson et al., 2014; Van Boxtel & Lu, 2013; Van De Cruys et al., 2014). The Neural Complexity Hypothesis (e.g., Belmonte et al., 2004; Just et al., 2012) provides a structural account. It proposes that perceptual and cognitive differences in autism arise from atypical brain organisation, particularly reduced long-range connectivity and inefficient integration across distributed networks. These connectivity patterns are thought to constrain the

coordination of lower-level perceptual and higher-level linguistic processes during speech comprehension.

Additionally, the Social Motivation Theory (Chevallier et al., 2012) highlights differences in attentional engagement. It suggests that autistic individuals may assign lower intrinsic reward value to social stimuli, leading to reduced spontaneous orientation to communicative input from early development. This diminished attention may limit opportunities to form robust predictive models for speech, particularly for socially salient features such as prosody, speaker identity, and pragmatic intent.

Overall, these frameworks converge on the view that speech processing in autism involves an imbalance between heightened local encoding and reduced top-down modulation. On one side, the Enhanced Perceptual Functioning model emphasises heightened sensitivity to low-level acoustic detail, reflecting stronger local processing. On the other side, theories such as Weak Central Coherence, Predictive Coding, the Social Motivation Theory, and the Neural Complexity Hypothesis propose that higher-level influences on interpretation are less spontaneously engaged, whether this reflects a detail-focused processing bias, altered use of prior expectations, reduced attention to socially informative cues, or less efficient integration across neural systems. Despite their different starting points, these accounts converge in predicting that context-dependent speech comprehension may be less flexible. To examine how these mechanisms shape speech processing, the following sections review behavioural and neural evidence across auditory and semantic domains under quiet listening conditions. This structure distinguishes differences arising during early auditory encoding from those emerging in higher-level lexical and semantic integration. By linking findings in each domain to the theoretical models outlined above, the review evaluates how well these accounts explain processing patterns in autism and considers whether distinct mechanisms contribute at different levels of the speech processing hierarchy.

1.1.1 Auditory processing

1.1.1.1 Spectral processing

Spectral processing is essential for auditory perception, enabling the discrimination of speech sounds, identification of voices, and organisation of complex acoustic scenes. In autism, differences in the encoding and use of spectral cues may influence both low-level auditory

resolution and higher-order speech processing. Pitch, or fundamental frequency (F0), is one of the most widely studied spectral features in autism, with atypical perception often reported and thought to reflect broader auditory differences (O'Connor, 2012). In speech, pitch conveys prosody, lexical tone, and pragmatic intent (Belyk & Brown, 2014; Chen et al., 2022), and plays a key role in auditory scene analysis by supporting stream segregation and speaker identification in multi-speaker environments (see Section 1.2). As pitch plays a central role in auditory processing and is closely linked to speech-in-noise perception, this section focuses on pitch-related processing in autism, while considering other features such as formant when relevant.

Behavioural studies consistently show that at least a subset of autistic individuals demonstrate intact or enhanced pitch perception, particularly in non-speech contexts (Ouimet et al., 2012). These strengths are evident in tasks involving pure tones and melody-based stimuli, where autistic individuals often outperform non-autistic controls in detecting subtle pitch differences, identifying pitch direction, and tracking melodic changes (Bonnel et al., 2003, 2010; Chowdhury et al., 2017; Heaton et al., 1998; Heaton, Hudry, et al., 2008; Järvinen-Pasley et al., 2008; O'Riordan & Passetti, 2006). However, recent evidence suggests that these advantages are not uniform and may be shaped by individual cognitive abilities. Ong et al. (2024) found that intelligence and memory capacity significantly modulated performance across pitch perception tasks, indicating that cognitive factors partly account for variability in autistic individuals' pitch processing. Beyond these low-level perceptual tasks, strengths may also extend to higher-level auditory processing, including recognising melodic contours, distinguishing notes within chords, and retaining musical pitch in memory (Heaton et al., 1999; Heaton, 2003; Heaton, Williams, et al., 2008; Jamey et al., 2019; Stanutz et al., 2012, 2014).
Neurophysiological evidence supports these behavioural findings, indicating that autistic individuals often show enhanced neural sensitivity to frequency changes. This is most consistently reflected in the mismatch negativity (MMN) or mismatch field (MMF), a frontocentral response occurring 100–300 ms after a deviant sound, which indexes pre-attentive acoustic change detection. These responses are typically elicited using an oddball paradigm, where a repetitive standard stimulus is occasionally replaced by a spectrally deviant sound, allowing researchers to assess neural sensitivity to unexpected changes (Haesen et al., 2011). Studies have reported comparable or enhanced MMN/MMF responses to frequency deviations in autistic individuals relative to non-autistic controls, suggesting intact or heightened early auditory discrimination (Čeponienė et al., 2003; Kujala et al., 2007; Lepistö

et al., 2005, 2007; Yu et al., 2015). These effects are observed across age groups, particularly among autistic individuals with higher verbal or intellectual abilities, and align with the Enhanced Perceptual Functioning account, which posits superior local auditory processing in autism (Mottron et al., 2006). However, while early responses may be enhanced, recent evidence suggests that the temporal dynamics of auditory processing are less efficient in autism. A large-scale study by Hudac et al. (2018) found that autistic children not only exhibited heightened P3a responses to novel sounds but also showed delayed habituation of both N1 and P3a amplitudes over time, indicating prolonged processing of auditory deviance. Habituation, defined as the typical reduction in neural response to repeated stimuli, reflects an adaptive mechanism that filters out redundant information (Rankin et al., 2009). The observed delay in this process suggests difficulties in sensory adaptation, which may contribute to sensory overload or reduced flexibility in autistic participants.

Evidence is more mixed when frequency-based spectral cues, including pitch, are examined in linguistic contexts, with performance varying across tasks and stimuli (Key & D'Ambrose Slaboch, 2021; L. Wang et al., 2023). In speech, pitch signals prosodic features such as intonation, lexical stress, and emotional tone, and carries lexical meaning in tonal languages (Crystal, 1969; Xu, 2005; Yip, 2002). Some behavioural studies report typical or even enhanced pitch perception in autistic individuals, including accurate identification of pitch contours and lexical tone contrasts (Cheng et al., 2017; Heaton, Hudry, et al., 2008; Järvinen-Pasley et al., 2008). Other studies, however, indicate difficulties when pitch signals sentence-level meaning, such as contrastive stress or pragmatic intent (Jiang et al., 2015; Lyons et al., 2014; McCann et al., 2007). These findings suggest that enhanced basic pitch sensitivity does not always support effective integration in complex linguistic contexts.

Neurophysiological studies of speech processing commonly use vowel or syllable stimuli in oddball paradigms, in which standard and deviant sounds differ in their spectral properties, including pitch and formant-related structure. Depending on the nature of the stimuli contrast, these paradigms can index either early detection of acoustic change or, when acoustic distance is controlled across phoneme boundaries, the efficiency with which spectral information is mapped onto phonemic categories (Key & D'Ambrose Slaboch, 2021). Using a passive oddball paradigm, Čeponienė et al. (2003) examined neural responses to spectral changes in speech and non-speech in autistic children. Deviant vowels were created by shifting all formant frequencies upward by 10%. Autistic children showed a clear MMN to vowel deviants that did

not differ from controls, indicating preserved early discrimination of formant-based speech changes. In contrast, the later P3a response was absent for speech deviants in the autistic group, despite being present for tone deviants. This pattern suggests intact early sensory detection of spectral change, alongside reduced automatic attentional orienting to speech. A similar pattern was reported by Lepistö et al. (2005), who examined neural discrimination of pitch changes and vowel identity contrasts (e.g., /a/ versus /o/) using speech vowels and acoustically matched non-speech counterparts. Autistic children again showed robust MMN responses to both pitch and formant frequency changes. However, P3a responses were attenuated for speech stimuli while remaining relatively typical for non-speech sounds, indicating a selective reduction in automatic attentional engagement with speech-related spectral cues. Evidence for this speech-specific attenuation extends into adulthood. Lepistö et al. (2007) reported robust, and in some cases enhanced, MMN responses to spectral changes in adults with Asperger syndrome, alongside reduced P3a responses to speech changes but relatively stronger orienting to non-speech changes. Converging evidence for speech-specific differences in pitch processing comes from studies of tonal languages, where pitch variations carry lexical meaning. In these contexts, autistic children show reduced MMN amplitudes and poorer behavioural categorisation of tonal contrasts (Wang et al., 2017; Yu et al., 2015; Zhang et al., 2019). These effects persist even when autistic and non-autistic groups are matched on age and non-verbal IQ and appear to arise for speech stimuli but not for acoustically matched non-speech sounds, highlighting a speech-specific difficulty in autism (Zhang et al., 2019).

A particularly clear demonstration of this distinction is provided by Chen et al. (2021), who directly manipulated speech-relevant spectral cues while controlling lower-level acoustic properties. The authors exaggerated the first and second formants (F1 and F2) of a Mandarin vowel while holding fundamental frequency constant. Non-autistic children showed an enhanced early P1 response to the formant-exaggerated vowel, consistent with increased neural sensitivity to speech-relevant spectral structure. In contrast, autistic children did not show this enhancement, despite comparable overall P1 amplitudes to the vowel stimuli. Crucially, when the same formant-based manipulation was applied to acoustically matched non-speech analogues, both groups exhibited increased P1 responses, indicating preserved sensitivity to spectral exaggeration outside of a speech context. Together, these findings suggest that group differences are not driven by reduced sensitivity to spectral change per se, but by how formant-related spectral cues are neurally prioritised when embedded in speech. Evidence from earlier developmental stages further suggests that basic spectral encoding is preserved at lower levels

of the auditory system. Using auditory brainstem responses and speech-evoked frequency-following responses, Jones et al. (2020) examined subcortical auditory processing in toddlers with autism aged 2–3 years. The speech stimulus was a synthetic consonant–vowel syllable (/da/), enabling assessment of neural encoding of pitch (F0), the first formant (F1), and higher-frequency spectral components. The authors reported limited group differences, with no robust differences in the magnitude of F0 or F1 encoding between autistic and non-autistic toddlers. These findings indicate that early, subcortical encoding of speech spectral features may be largely intact in autism, and that differences observed in later childhood and adulthood are more likely to emerge at cortical or integrative stages of processing.

The neurophysiological findings reviewed above also suggest that differences in speech-related spectral processing may be shaped by attentional factors. In particular, the dissociation between preserved early change detection and reduced later orienting responses observed in MMN and P3a measures raises the possibility that autistic listeners detect acoustic changes in speech but do not automatically allocate attention to them. Direct evidence for this account comes from Whitehouse and Bishop (2008), who compared passive and active listening in an oddball paradigm. Reduced P3a responses in the autistic group were observed only during passive listening to speech vowels; when explicitly instructed to detect deviant sounds, their attentional responses were comparable to non-autistic peers. These findings suggest that autistic children are capable of attending to speech when task demands require it, but may show reduced spontaneous engagement with socially relevant auditory input in more naturalistic contexts. Supporting this view, Kuhl et al. (2005) found that individual differences in social attention modulated neural responses: autistic children who attended more to child-directed speech showed more typical MMN responses to consonant changes, whereas those who preferred non-speech sounds exhibited atypical MMN amplitudes and lateralisation patterns.

Beyond attention, the social relevance of specific acoustic cues may further shape how speech is processed. Pitch, in particular, plays a central role in conveying socially salient information such as speaker identity. There is strong evidence that vocal pitch cues support voice recognition (Abberton & Fourcin, 1978; Baumann & Belin, 2010; Kreitewolf et al., 2014; Lavner et al., 2000). In multi-speaker environments, the ability to use vocal pitch to distinguish speakers is essential for auditory scene analysis (see Section 1.2). Early studies suggested that autistic individuals may show reduced responses to vocal information. Klin (1991) found that autistic children did not show the typical preference for their mother's voice, interpreted as

reflecting reduced social motivation or atypical salience of vocal cues. Similarly, Boucher et al. (1998) reported poorer discrimination between familiar and unfamiliar voices in autistic children, relative to those with language delays. However, subsequent studies suggested that voice recognition difficulties may be task-dependent and would emerge only under high processing demands or when voices are acoustically similar (Boucher et al., 2000). A more comprehensive picture of vocal pitch processing in autism comes from a series of studies by Schelinski and colleagues. Behaviourally, Schelinski et al. (2017) found that autistic adults with typical cognitive abilities showed specific difficulties in discriminating vocal pitch especially when the voice is from an unfamiliar speaker, despite preserved sensitivity to musical pitch and vocal timbre, demonstrating their difficulty in learning novel voices. Neuroimaging evidence from the same group (Schelinski et al., 2016) revealed that while both autistic and non-autistic individuals activated the temporal voice areas (TVAs) in response to vocal sounds, only non-autistic participants showed increased TVA activation when actively discriminating speaker identity based on pitch cues. In autistic individuals, TVA responses remained unchanged in this context, suggesting intact low-level voice detection but reduced functional tuning for socially relevant auditory distinctions.

In summary, current evidence indicates that spectral processing of speech is often atypical in autism. While some autistic individuals show enhanced pitch sensitivity in non-linguistic contexts, behavioural and neural differences often emerge when spectral cues must be integrated with higher-level linguistic or socially meaningful information. This challenges the notion of a general enhancement in low-level auditory processing proposed by the Enhanced Perceptual Functioning model. Instead, the evidence points to difficulties in using spectral cues effectively when integration with linguistic or identity-related information is required. This pattern aligns with Predictive Coding accounts and the Weak Central Coherence hypothesis, both of which emphasise reduced top-down modulation. Evidence also indicates greater difficulty processing pitch in speech and voice contexts than in non-speech sounds. This pattern, consistent with the Social Motivation Theory, suggests reduced sensitivity to the communicative significance of spectral cues.

1.1.1.2 Temporal Processing

Speech is a temporally structured signal, requiring listeners to dynamically track how sounds unfold over time to extract linguistic information (Hickok, 2012; Luo & Poeppel, 2012). This ability, known as auditory temporal processing, enables listeners to segment the auditory

stream and map acoustic cues onto linguistic units that unfold over time (Ding et al., 2017; Ghitza, 2011; Giraud & Poeppel, 2012; Poeppel, 2003; Rosen, 1992).

Early behavioural evidence for atypical temporal processing in autism comes from speech-in-noise studies examining listeners' ability to take advantage of brief dips in background noise. These temporal dips create short periods of reduced masking and are thought to reflect sensitivity to fine-grained temporal structure (see Section 1.3.2 for further discussion). Behavioural findings suggest that autistic individuals derive less benefit from such modulations, resulting in poorer speech intelligibility in fluctuating maskers (Alcántara et al., 2004; Groen et al., 2009). Crucially, successful dip listening requires not only sufficient temporal resolution to detect the dips, but also the ability to use them effectively for stream segregation.

Difficulties in temporal processing in autism are consistently observed in behavioural tasks. Autistic individuals tend to show elevated thresholds for detecting brief silent gaps between stimuli (Bhatara et al., 2013; Boets et al., 2015; Foss-Feig et al., 2017) and for judging the order of temporally adjacent sounds (Kwakye et al., 2011). Specifically, in gap detection tasks, autistic individuals perform comparably to non-autistic controls when the gaps between sounds are long. However, as the interval shortens, their accuracy decreases and response times increase, indicating reduced sensitivity to subtle temporal differences. Alcántara et al. (2012) extended prior work on temporal resolution by examining auditory envelope processing in autistic children. Using a non-speech amplitude modulation detection task, the authors assessed sensitivity to changes in the temporal envelope of broadband noise across a wide range of modulation rates (2–512 Hz). This design allowed them to evaluate both temporal resolution (indexed by the highest modulation rate reliably detected) and processing efficiency (reflected in detection thresholds at slower modulation rates). While autistic children demonstrated similar temporal resolution to non-autistic peers, they showed consistently elevated detection thresholds across all rates, indicating reduced efficiency in tracking temporal-envelope cues. Crucially, performance on a control task assessing intensity discrimination did not differ between groups, ruling out general auditory or attentional deficits. Generally, behavioural results suggest that even when basic encoding of temporal information is preserved, autistic individuals may be less efficient in temporal integration processes critical for parsing dynamic auditory input.

Neurophysiological evidence from both passive and active paradigms indicates atypical processing of temporal duration cues in autism. ERP studies using passive oddball designs have consistently reported attenuated MMN responses to duration changes across both speech and non-speech stimuli in autistic children and adults, suggesting reduced pre-attentive sensitivity to temporal deviance (Lepistö et al., 2005, 2007). In Mandarin-speaking children, Huang et al. (2018) found preserved MMN responses to within-category speech duration contrasts but reduced P3a amplitudes, indicating diminished involuntary orienting to temporal changes in speech. Together, these findings point to a domain-general disruption in temporal encoding, particularly under passive listening conditions. Complementary evidence from active paradigms further supports this profile. In a gap detection task, autistic individuals demonstrated intact behavioural performance but reduced P2 amplitudes to near-threshold silent gaps, reflecting diminished engagement of attention or classification processes (Foss-Feig et al., 2018). Notably, P2 amplitude was positively associated with language ability, linking neural sensitivity to temporal features with communicative function. An fMRI study revealed increased activation in primary auditory cortex but reduced recruitment of higher-order auditory regions during temporally complex sound processing, suggesting an over-reliance on low-level encoding and limited integration of dynamic auditory input (Samson et al., 2011). In line with behavioural results, these findings highlight atypical temporal processing at both early and integrative stages.

Traditional ERP paradigms rely on averaging responses to discrete, repeated stimuli, which limits their ability to capture how the brain dynamically encodes temporal structure over time. As a result, they provide only a partial view of auditory processing, particularly when it comes to understanding how listeners track and integrate ongoing acoustic information in natural speech (Crosse et al., 2016). To address these limitations, recent research has increasingly adopted neural measures capable of capturing the brain's response to continuous, time-varying input (Palana et al., 2022). One widely used approach is neural tracking, which refers to the alignment between ongoing neural activity and the time-varying acoustic features of continuous speech (Obleser & Kayser, 2019). A central focus has been on tracking the speech envelope—the slow amplitude fluctuations that reflect the rhythmic and hierarchical structure of speech (Rosen, 1992). These modulations signal the timing of phonemes and syllables, supporting the extraction of both segmental and prosodic information. Neural responses in the delta (<4 Hz) and theta (4–8 Hz) bands synchronise with these modulations, corresponding to prosodic and syllabic timescales, respectively (Ahissar et al., 2001; Doelling et al., 2014;

Giraud & Poeppel, 2012; Luo & Poeppel, 2007). The temporal envelope plays a fundamental role in speech comprehension, as speech can remain intelligible based on envelope cues alone, even in the absence of fine spectral detail (Ghitza & Greenberg, 2009; Shannon et al., 1995). Although envelope tracking reflects low-level acoustic encoding, it is also modulated by higher-level processes (Ding & Simon, 2014). Stronger envelope tracking has been associated with greater speech intelligibility (e.g., Gross et al., 2013; Peelle et al., 2013), and attention has been shown to selectively enhance tracking of a target speaker in complex listening environments (Ding & Simon, 2012a; Kerlin et al., 2010; O'Sullivan et al., 2015; Zion Golumbic et al., 2013).

In early studies, neural tracking was measured by lagged cross-correlation and coherence analyses to examine phase synchronisation between neural activity and the speech signals (Ahissar et al., 2001; Luo & Poeppel, 2007). While effective in detecting overall alignment, these methods are limited by their sensitivity to autocorrelation and inability to separate overlapping acoustic features. To address these limitations, encoding and decoding models have been introduced (Crosse et al., 2016; Holdgraf et al., 2017; Mesgarani et al., 2009), using regularised linear regression to predict either the neural response from the speech stimulus (forward models) or reconstruct the speech envelope from neural activity (backward models). These models offer improved temporal precision, robustness to redundancy in the stimulus, and better generalisation across conditions (Crosse et al., 2021).

Recent work has begun to apply neural tracking methods to investigate auditory temporal processing in autism. Samoylov et al. (2024) examined neural tracking of non-speech rhythmic input by measuring phase synchronisation to 2 Hz amplitude-modulated tones, which mirrors prosodic timing in speech. Autistic children showed significantly reduced synchronisation in auditory cortex compared to non-autistic peers, and stronger entrainment predicted higher receptive language scores. These findings suggest atypical neural encoding of rhythmic temporal structure and its relevance for language development. Extending this work to naturalistic speech, X. Wang et al. (2023) used EEG to assess how young children's brain activity aligned with the speech envelope while they watched cartoons containing embedded speech. Neural responses were filtered into delta and theta bands and compared with the speech signal using linear modelling. Autistic children showed significantly reduced tracking in both bands, indicating weaker alignment with prosodic and syllabic structure. Importantly, these

effects were not accompanied by differences in visual attention, pointing to a specific neural difference in synchronising with the rhythmic structure of speech.

Temporal speech processing also requires the integration of information across timescales. One proposed mechanism is phase–amplitude coupling (PAC), in which the phase of low-frequency oscillations modulates the amplitude of higher-frequency activity. Using MEG, Jochaut et al. (2015) found reduced theta–gamma coupling in autistic adolescents and adults during speech listening, and this disruption predicted poorer sentence comprehension. X. Wang et al. (2023) reported atypical beta-gamma coupling in autistic children, with this non-canonical pattern emerging as the strongest predictor of individual speech reception scores, surpassing the influence of attention and envelope tracking.

Together, these findings suggest that while basic encoding of temporal cues may be preserved in some contexts, both behavioural and neural evidence consistently points to atypical temporal processing in autism. Observed difficulties include reduced sensitivity to brief temporal changes, diminished efficiency in tracking dynamic modulations, and weaker neural synchronisation with the rhythmic structure of speech. These difficulties become more apparent when processing demands increase, such as during fluctuating masking or when speech unfolds continuously over time. Neural tracking studies have highlighted reduced envelope alignment and atypical cross-frequency coupling as robust markers of disrupted temporal encoding and integration. This pattern supports theoretical accounts that emphasise attenuated top-down modulation and reduced prioritisation of temporally distributed cues (e.g., Weak Central Coherence; Predictive Coding).

1.1.2 Semantic processing

Semantic processing operates across multiple levels, from retrieving the meaning of individual words to integrating those meanings within larger sentence and discourse contexts (Hagoort, 2005). In autism, behavioural performance at the single-word level is often comparable to non-autistic peers, as shown in tasks such as word–picture matching and auditory semantic priming (Cantiani et al., 2016; Kamio et al., 2007). However, sentence-level semantic processing frequently reveals greater difficulty, with studies reporting reduced accuracy to sentence congruency judgement (Fishman et al., 2011; Manfredi et al., 2020). Differences in neural responses are more consistently observed across both levels of processing. ERP studies, particularly those examining the N400 component, often report attenuated or delayed responses

in autism (Hagoort, 2008; Mamashli et al., 2017; Pijnacker et al., 2010), while neuroimaging research points to atypical patterns of activation and connectivity within the language network (Cardinale et al., 2013; Gaffrey et al., 2007; Harris et al., 2006; Hwang et al., 2017; Just et al., 2012; Knaus et al., 2008). These findings suggest that while overt comprehension may appear intact, underlying neural mechanisms supporting semantic processing may be altered.

The N400 is a negative-going ERP component that peaks approximately 400 milliseconds after the onset of a potentially meaningful stimulus. Although it is widely recognised as a neural marker of semantic processing, its precise functional interpretation remains debated. One influential account links the N400 to semantic integration difficulty, with larger amplitudes reflecting increased effort to incorporate a word into its preceding context (Hagoort, 2008; Kutas & Hillyard, 1980). This view is primarily supported by semantic anomaly paradigms, in which incongruent sentence completions (e.g., "He spread the warm bread with socks") elicit larger N400 amplitudes than congruent ones. An alternative view posits that the N400 reflects earlier stages of processing, such as lexical access or the activation of semantic features (Federmeier, 2007; Lau et al., 2008). Supporting evidence comes from semantic priming paradigms, where semantically related word pairs (e.g., cat–mouse) reliably elicit smaller N400 amplitudes than unrelated pairs (e.g., cat–fork), even in the absence of syntactic or sentential context. According to this view, N400 amplitude reflects the degree to which lexical representations are pre-activated based on context or learned associations. More recent frameworks attempt to reconcile these accounts, proposing that the N400 reflects both probabilistic semantic prediction and the updating of contextual representations within a hierarchical predictive coding framework (Kuperberg, 2016).

Atypical N400 responses have been consistently reported in autism across both word-priming and sentence-level semantic paradigms. Attenuated or delayed N400 responses have been observed in studies employing word classification (Dunn et al., 1999; Dunn & Bates, 2005), semantic priming (DiStefano et al., 2019; Dunn et al., 1999; Dunn & Bates, 2005; McCleery et al., 2010) as well as semantic anomaly paradigms involving full sentence contexts (Braeutigam et al., 2008; Manfredi et al., 2020; Pijnacker et al., 2010; Ribeiro et al., 2013; Ring et al., 2007). In some cases, the typical N400 is absent, with a Late Positive Potential observed instead in response to semantic incongruities (Pijnacker et al., 2010; Ribeiro et al., 2013). This later positivity is thought to reflect delayed or compensatory integration processes that may support comprehension when early semantic processing is disrupted. These findings suggest

atypical semantic processing, which was interpreted as alterations in predictive semantic activation, lexical access, and/or context-based integration processes in autism. Importantly, reduced N400 responses to linguistic mismatches alongside intact responses to nonverbal mismatches indicate that semantic integration differences in autism may be specific to socially relevant speech input (McCleery et al., 2010).

Furthermore, individual differences in language ability appear to modulate neural responses to semantic information in autism. Reduced N400 amplitudes have been observed in autistic individuals with lower verbal skills (Coderre et al., 2017; DiStefano et al., 2019), whereas studies using the same paradigm but with better-matched language abilities report no significant group differences (O'Rourke & Coderre, 2021). However, atypical N400 effects have also been found in sentence-level tasks even when participants are matched for cognitive and verbal skills (Braeutigam et al., 2008). Taken together, these findings suggest that although language ability influences neural responses to meaning, atypical semantic processing in autism cannot be fully explained by general language or cognitive skills. Instead, it may reflect broader differences in how meaning is accessed and integrated.

Some studies also explored potential interactions between semantic and early auditory processing. Dunn et al. (1999) found that autistic children showed reduced P3b amplitudes in response to out-of-category words, suggesting decreased attentional allocation to semantic distinctions. Additionally, the autistic group exhibited prolonged latencies in early sensory components (N1, P2), which may reflect increased processing demands for word-level stimuli relative to simpler sounds like vowels or syllables. These findings were partially replicated in a larger sample (Dunn & Bates, 2005), who showed comparable N400 responses. However, delayed N1 responses were observed only in the younger autistic subgroup, suggesting that these delays may improve over time. Notably, there was no significant correlation between N1 and N400 latencies, indicating that early auditory delays were not directly linked to the timing of later semantic integration processes. These findings point to possible interactions between early auditory processing and later semantic integration, but the relationship appears to be complex and may vary with age, task demands, or individual differences.

Neuroimaging findings suggest that while the core semantic network remains functionally accessible in autism, its recruitment tends to be more variable, less lateralised, and often less efficient. Specifically, autistic individuals frequently exhibit more diffuse and less consistently

left-lateralised activation during language comprehension. Rather than focally engaging canonical left-hemisphere language areas, they often recruit right-hemisphere homologues and perceptual or associative regions, such as the fusiform gyrus and posterior visual cortices (Cardinale et al., 2013; Gaffrey et al., 2007; Harris et al., 2006; Knaus et al., 2008). These patterns suggest that semantic meaning may be constructed via alternative or compensatory neural pathways that are less specialised for language. Additionally, studies report reduced functional connectivity between key semantic regions, particularly between the left inferior frontal gyrus (IFG) and posterior temporal cortices (Hwang et al., 2017; Just et al., 2012; Kana et al., 2014). Together, atypical lateralisation, reduced functional integration, and compensatory activation of non-linguistic regions point to fundamental differences in the neurocognitive architecture underlying meaning construction in autism (Phan et al., 2021).

Overall, significant differences in semantic processing have been observed in autism, particularly at the sentence level, where successful comprehension depends on integrating meaning across words and over time. These difficulties are often interpreted within the framework of Weak Central Coherence, which proposes a reduced tendency to construct global, contextually coherent representations. Within a predictive coding framework, such difficulties are attributed to attenuated use of contextual priors during language comprehension, resulting in less top-down support for interpreting incoming input. Finally, atypical patterns of functional connectivity within the language network observed in neuroimaging studies lend support to the Neural Complexity Hypothesis, suggesting that reduced integration across distributed neural systems may further constrain meaning construction in autism.

1.1.3 Summary

Atypical auditory processing in autism has been documented across multiple domains—including spectral, temporal, and semantic levels—through converging behavioural and neural evidence. Autistic individuals often show intact or even heightened sensitivity to basic acoustic features, yet experience challenges when processing speech that unfolds over time or requires integration of linguistically relevant information, such as prosody, word meaning, or sentence-level context.

These findings have been interpreted through several theoretical frameworks, each capturing different aspects of the observed profile. The Enhanced Perceptual Functioning model accounts for strengths in local acoustic discrimination by proposing a bias toward detailed perceptual

analysis, though it does not explain broader integration difficulties. In contrast, the Weak Central Coherence theory characterises a detail-focused processing bias, marked by a default prioritisation of locally available information and reduced spontaneous use of broader contextual cues, which can result in less consistent recruitment of prosodic and semantic information in unconstrained listening situations. Predictive coding models attribute these integration difficulties to atypical weighting of prior expectations and reduced top-down predictions, which may limit adaptation to structured, time-varying input. Complementing these cognitive accounts, neuroimaging evidence of reduced functional connectivity within the language network supports the Neural Complexity Hypothesis, which posits inefficient coordination across distributed cortical systems.

Across different levels of auditory processing, a consistent dissociation has emerged between responses to speech and non-speech stimuli. While basic processing of non-speech sounds often appears intact, autistic individuals show attenuated neural responses to speech at both auditory and semantic levels (Chen et al., 2022; Galilee et al., 2017; McCleery et al., 2010; Piatti et al., 2021; Zhang et al., 2019). This domain-specific difficulty aligns with the Social Motivation Theory (Chevallier et al., 2012), which proposes that reduced early orientation to speech limits the development of specialised neural mechanisms for processing communicative signals. Over time, this diminished social-linguistic tuning may contribute to persistent differences in how speech is perceived and integrated, even in the presence of intact perceptual abilities for non-linguistic sounds.

While neurophysiological methods have revealed important processing differences in autism that may not be apparent through behavioural measures alone, several limitations remain. Most existing studies have examined auditory and semantic processing in isolation, providing limited insight into how these systems interact during naturalistic speech perception. Moreover, neural tracking methods remain underutilised in autism research, despite offering clear advantages over traditional ERP approaches. While ERP paradigms depend on averaging neural responses to isolated, time-locked stimuli, neural tracking analyses reflect how the brain continuously processes natural speech, yielding a richer and more ecologically valid picture of auditory encoding in real time.

To date, the few neural tracking studies in autism have primarily focused on neural entrainment, such as phase coherence with rhythmic input. While these approaches provide useful indices

of global alignment between neural activity and the speech signal, they lack the temporal specificity needed to resolve the fine-grained structure of neural responses. A more recently developed technique, the temporal response function (TRF), addresses these limitations by modelling how the brain responds to continuous speech over time, capturing both the timing and shape of neural activity associated with specific features of the input. The TRF approach offers enhanced temporal resolution and yields interpretable components—such as P1, N1, and P2—that correspond to established ERP markers while preserving the continuous dynamics of naturalistic speech. Despite these advantages, TRF has not yet been applied to investigate auditory encoding in autism, leaving a key gap in understanding how speech is processed in real time in this population (see Chapter 4 for further discussion).

1.2 Speech-in-noise processing

In everyday listening environments, individuals are often required to attend to a particular speaker while ignoring competing background sounds. This difficulty is commonly referred to as the "cocktail party problem" (Cherry, 1953). It arises because all ambient sounds reach the ears simultaneously and are combined into a single acoustic mixture (Shinn-Cunningham & Best, 2008). To interpret this complex input, the auditory system must extract the speech of interest (i.e., target), and separate it from irrelevant or distracting sounds (i.e., maskers). This process, known as auditory scene analysis, involves grouping and segregating acoustic elements into perceptually distinct auditory objects, such as the voice of the target speaker and other competing sources like speakers, music, or environmental noise (Bregman, 1990; Griffiths & Warren, 2004; Woods & Colburn, 1992). This section begins by introducing different types of masking that interfere with speech perception. It then reviews the acoustic and contextual cues that listeners use to manage speech-in-noise challenges. Finally, we discuss the cognitive factors that influence the efficiency of cue use in complex auditory environments.

1.2.1 Masking effects

Maskers disrupt auditory scene analysis not only by physically obscuring the target signal, but also by interfering with attention and linguistic processing, thereby reducing the intelligibility of the intended speech (Bronkhorst, 2000). These forms of interference are typically classified as energetic masking and informational masking.

Energetic masking occurs when background noise overlaps with the target speech in time and frequency, leading to competition at the auditory periphery. This acoustic interference disrupts the neural representation of speech and reduces its perceptual clarity. The degree to which this interference disrupts speech perception is strongly influenced by the signal-to-noise ratio (SNR), defined as the relative intensity of the target signal compared to the masker (Rhebergen & Versfeld, 2005). When the masker is substantially louder than the target, producing a low SNR, less of the speech signal is transmitted with sufficient fidelity to higher auditory processing stages. This results in reduced audibility and, in turn, impaired intelligibility (Brungart, 2001; Mattys et al., 2009). Although some evidence suggests that central mechanisms may play a role under certain conditions (Culling & Stone, 2017; Wang & Xu, 2021), energetic masking is primarily driven by interference at early, bottom-up stages of auditory processing.

Energetic masking can be partially mitigated when the masker fluctuates over time or frequency, a condition often referred to as modulated masking (Culling & Stone, 2017). In these conditions, listeners may exploit brief reductions in masker energy (i.e., temporal and spectral dips) to gain access to otherwise masked portions of the target speech. This process, known as dip listening or glimpsing, enables recovery of speech elements during moments when the local SNR becomes more favourable (Cooke, 2006; Miller & Licklider, 1950). Temporal dips refer to brief reductions in masker amplitude over time, whereas spectral dips involve frequency bands where the masker contains relatively little energy. A substantial body of psychoacoustic research has demonstrated that speech intelligibility is significantly improved in modulated noise compared to stationary noise (Festen & Plomp, 1990; Miller & Licklider, 1950). In these studies, participants were typically asked to identify words or sentences under varying masker conditions, and speech reception thresholds (SRTs) were measured to assess performance. SRTs, commonly defined as the SNR required for 50% correct identification, provide a reliable index of speech intelligibility. It is well established that neurotypical adults with normal hearing exhibit lower (i.e., better) SRTs in modulated compared to unmodulated noise, indicating their capacity to use tempo-spectral dips to reduce the impact of energetic masking (Howard-Jones & Rosen, 1993; Rhebergen & Versfeld, 2005).

However, in many real-world listening situations, speech intelligibility cannot be explained by audibility alone. Even when the target signal is clearly audible, comprehension may still be impaired due to higher-level interference from competing sounds. This phenomenon, known

as informational masking, reflects disruption at more central auditory and cognitive stages. It typically arises when the masker shares perceptual, structural, or semantic features with the target, making it difficult for listeners to segregate and attend to the relevant source (Brungart, 2001; Kidd et al., 2008). While informational masking can occur with complex non-speech sounds, it is most commonly examined in speech-on-speech contexts. In these scenarios, both the target and masker signals remain acoustically distinct and accessible, yet comprehension suffers due to the cognitive demands of suppressing irrelevant linguistic input and maintaining attention on the intended speaker (Cooke et al., 2008; Mattys et al., 2009; Rosen et al., 2013).

Several factors modulate the degree of informational masking. One well-established factor is intelligibility. Studies comparing forward (intelligible) and time-reversed (unintelligible) speech show that meaningful speech produces greater interference, indicating that access to linguistic content increases masking (Cherry, 1953; Freyman et al., 2001; Kellogg, 1939; Kidd et al., 2010; Ueda et al., 2017). A second factor is language familiarity. Maskers in a listener's native language typically cause more interference than unfamiliar ones, likely because they more readily engage lexical and semantic processing. For example, English speakers experience more masking from English than from Mandarin (Calandruccio et al., 2010; Cooke et al., 2008; Van Engen & Bradlow, 2007). However, this effect is less consistent in tasks with lower linguistic demands, suggesting that attentional load moderates the impact (Mattys et al., 2009). A third factor is linguistic similarity between the target and masker. When two languages are closely related and share features such as phonology, rhythm, or syntax, it becomes more difficult for listeners to perceptually separate them (Brouwer et al., 2012). For example, English speech is more disrupted by Dutch, a typologically similar language, than by a distant language like Mandarin (Calandruccio et al., 2013).

Together, these findings suggest that structured, meaningful, and familiar speech is particularly difficult to ignore and more likely to interfere with target speech processing. Such effects reflect not only surface-level intelligibility but also deeper cognitive engagement with linguistic content, underscoring the role of high-level language processing in informational masking. While informational masking is typically examined in the context of speech-based interference, it is important to recognise that it can also arise from non-speech sources. For example, music may also act as a potent informational masker, particularly when it includes lyrics, salient vocal lines, or rhythmic and prosodic structures that resemble speech (see

Chapter 3). These features may similarly capture attention and disrupt the listener's ability to focus on the target stream (Brouwer et al., 2022).

1.2.2 Cues to reduce masking effects

1.2.2.1 Speaker-related acoustic cues

In multi-talker environments, listeners rely on acoustic cues associated with the target speaker to distinguish their voice from competing speech streams (Bronkhorst, 2000). Two of the most effective cues are spatial location and vocal identity. Spatial cues distinguish the timing and intensity of sound arriving at each ear, allow listeners to localise the target speaker and to perceptually separate them from other sound sources (Arbogast & Kidd, 2000; Kidd, Mason, et al., 2005). Voice-related cues, such as timbre and gender, further enhance the distinctiveness of the target stream (Culling et al., 2003; Darwin et al., 2003). By making the target speech more acoustically and perceptually distinct, these cues reduce both energetic and informational masking and support selective attention.

Spatial location is a salient perceptual cue for segregating competing auditory streams in complex acoustic environments. Spatial release from masking (SRM) refers to the improvement in speech intelligibility when the target and masker are spatially separated. SRM operates through two primary mechanisms: improved signal-to-noise ratio at the ear with the more favourable input (better-ear listening) and binaural unmasking based on interaural time and level difference (Bronkhorst, 2000; Culling et al., 2004; Culling & Mansell, 2013; Culling & Stone, 2017; Edmonds & Culling, 2006; Kidd et al., 2010; Marrone et al., 2008). At the neural level, enhanced N1 responses to spatially separated versus co-located sources suggest early-stage perceptual sensitivity to spatial separation (Teder-Sälejärvi et al., 1999). Spatial cues are particularly effective in reducing informational masking when the target and masker share similar linguistic or vocal features, as in speech-on-speech interference (Best et al., 2012; Calandruccio et al., 2017; Freyman et al., 2001; Kidd et al., 2008, 2010, 2016; Rennies et al., 2019; Swaminathan et al., 2016; Viswanathan et al., 2016).

Similarly, voice-based cues, such as differences in pitch (fundamental frequency, F0), timbre, and vocal tract characteristics, facilitate stream segregation by distinguishing the target speaker from competing speakers. Behavioural studies show improved speech recognition when the target and competing speaker differ in pitch or vocal identity (Brown & Bacon, 2010; Brungart,

2001; Cullington & Zeng, 2008). In contrast, speech recognition declines when both voices share similar acoustic properties, such as gender or speaker identity (Allen et al., 2008; Brungart & Simpson, 2002; Noble & Perrett, 2002; Zhang et al., 2020). Neurophysiological findings support this distinction: ERP components such as N1 and P2 are modulated by speaker identity, with stronger responses observed when attention is consistently directed to a specific voice, indicating early-stage encoding of speaker-related features and their role in maintaining stream continuity (Pettigrew et al., 2004).

Together, these findings highlight the essential contribution of spatial and vocal cues to auditory stream segregation. They support early perceptual organisation by enhancing acoustic distinctiveness and enabling listeners to isolate the target signal from competing sounds, thereby forming the foundation for successful speech processing in complex environments. One widely used approach for experimentally examining the effects of both spatial and voice cues on SiN comprehension is the Coordinate Response Measure (CRM; Bolia et al., 2000) and its modified forms, such as the Children's CRM (CCRM; Messaoud-Galusi et al., 2011). These paradigms require listeners to identify keywords from a target sentence presented concurrently with one or more competing speech streams. Performance is significantly enhanced when the target and masker differ in spatial location or vocal characteristics, demonstrating how each cue facilitates stream segregation (Best et al., 2010; Ihlefeld & Shinn-Cunningham, 2008; Kidd, Arbogast, et al., 2005). Because the competing utterances are similar in syntactic structure and vocabulary, these tasks impose high demands on stream selection, making them particularly sensitive to differences in cue salience and individual processing strategies. The CRM paradigm has been widely used in studies of typical and atypical speech perception and is adopted in the present thesis to examine how spatial and voice cues support speech-in-noise comprehension in autistic and non-autistic adults (see Chapter 2).

1.2.2.2 Contextual cues

In challenging listening environments, speech comprehension depends not only on the acoustic feature of the target speaker but also on the listener's ability to apply linguistic knowledge. When low-level acoustic detail is degraded by noise, listeners increasingly rely on expectations about predictable words, sentence structure, and semantic meaning to support interpretation. These top-down contextual cues help compensate for incomplete sensory input by guiding perceptual inference and narrowing the range of plausible interpretations.

A foundational example is the phoneme restoration effect, in which listeners perceptually "fill in" speech segments masked by non-speech noise and report hearing both the missing phoneme and the masking sound (Warren, 1970). This effect demonstrates that lexical context can actively shape low-level auditory perception. Subsequent studies consistently show that listeners are better able to recognise degraded speech when it is embedded in syntactically and semantically coherent contexts (Boothroyd & Nittrouer, 1988; Miller et al., 1951). For example, sentence-final words are identified more accurately when preceded by strong semantic context (Kalikow et al., 1977). In addition, speech arranged in grammatical sentences is more intelligible than random word sequences, especially under multi-speaker conditions (Kidd et al., 2014). Together, these findings suggest that listeners use a range of contextual cues, including local phoneme-level expectations as well as broader syntactic and semantic structures, to compensate for reduced bottom-up input.

Neuroimaging studies extend behavioural and electrophysiological findings by demonstrating how the brain engages contextual linguistic information to support speech comprehension in noisy environments. Functional MRI research has shown that processing semantically coherent speech in noise is associated with increased activation in brain regions involved in semantic and syntactic integration, particularly the left inferior frontal gyrus (IFG) and bilateral superior temporal cortices (Scott et al., 2004; Scott & McGettigan, 2013). These regions are thought to support the prediction and integration of linguistic input, especially when bottom-up cues are ambiguous or degraded. Notably, IFG activation tends to peak when the acoustic signal is degraded but still partially accessible. This pattern suggests that top-down mechanisms are most effectively engaged when listeners can rely on both residual bottom-up information and higher-level linguistic predictions to interpret the speech signal (Zekveld et al., 2006).

Complementing these neuroimaging findings, ERP research has used the N400 component to examine how listeners draw on contextual cues to support semantic processing in noise. The N400 reflects the ease of integrating a word into its preceding context and is commonly interpreted as an index of how well semantic expectations are met during comprehension. In SiN paradigms, it provides a useful measure of listeners' ability to recruit top-down contextual information when the acoustic signal is degraded. Although findings on N400 modulation by noise are mixed, two broad patterns have emerged. Some studies report increased N400 amplitudes in the presence of distracting noise, particularly when the masker contains linguistic content, suggesting greater reliance on semantic context to support speech recognition under

these conditions (Devaraju et al., 2021; Kemp et al., 2019; Romei et al., 2011; Song et al., 2020). In contrast, others find reduced or delayed N400 responses, often interpreted as reflecting impaired lexical access or disrupted integration when the speech signal is severely masked (Aydelott et al., 2006; Hsin et al., 2023; Obleser & Kotz, 2011; Silcox & Payne, 2021; Strauß et al., 2013). These divergent findings may reflect differences in masker type, signal degradation, or the strength of contextual cues. Nevertheless, the N400 remains one of the most widely used neural markers of semantic processing effort in SiN research, particularly sensitive to variation in the acoustic and linguistic properties of the masker (e.g., Song et al., 2020). Accordingly, Studies 2 and 3 of this thesis employ the N400 to assess how varying masker types affect semantic processing under speech-in-noise conditions.

1.2.2.3 Top-down modulation of cue use

While speaker-related acoustic cues and contextual linguistic cues can support speech comprehension under masking, their effectiveness depends not only on properties of the signal but also on the listener's ability to detect, attend to, and integrate them. This process places high demands on cognition, requiring selective attention and executive control to focus on relevant input and suppress interference.

Neuroimaging studies show that the use of speaker-related cues such as spatial location and pitch recruits not only primary auditory regions but also higher-order attentional and control networks. For example, spatial stream segregation engages frontoparietal areas involved in attentional focus and cognitive control, particularly under high perceptual load or distraction (Best et al., 2012; Gutschalk & Dykstra, 2014; Lee et al., 2014). Voice-based cues similarly activate the right superior temporal sulcus and auditory cortex when listeners track a designated speaker in multi-speaker environments (Belin & Zatorre, 2003; Rosen et al., 2013). These findings suggest that the benefits of acoustic cues rely on both early perceptual mechanisms and top-down modulation that sustains attention on the target stream.

When such acoustic cues are absent or unreliable due to great energetic masking or high perceptual similarity between sources, listeners increasingly shift toward top-down strategies. McQueen & Huettig (2012) found that when speech was intermittently masked, listeners reduced reliance on word-onset competitors and instead drew more on rhyme-based lexical alternatives. This adaptive cue weighting becomes especially important under informational masking, where acoustic segregation is limited. Mattys et al. (2009) showed that under

cognitive load, recognition performance declined more sharply in the presence of competing speech than with energetic noise, indicating that successful comprehension depends on the ability to recruit contextual support.

However, this flexibility comes at a cognitive cost. When bottom-up information is degraded, understanding speech in noise demands greater use of general cognitive resources, such as working memory and attentional control (Mattys et al., 2012; Peelle, 2018; Pichora-Fuller et al., 2016). Neuroimaging studies demonstrate that effortful listening activates a broader frontoparietal network that supports speech processing and decision-making, including the IFG, dorsolateral prefrontal cortex, anterior cingulate cortex, and posterior parietal areas (McGettigan et al., 2012; Zekveld et al., 2012). These regions interact with attentional control systems, such as the frontal eye fields and intraparietal sulcus, which enable listeners to shift and maintain attention in dynamic auditory scenes (Corbetta et al., 2008; Lee et al., 2014).

Together, these findings underscore the importance of top-down modulation in managing acoustic and linguistic uncertainty during SiN processing. Effective comprehension under masking relies not only on the perceptual salience of individual cues but also on the listener's ability to flexibly allocate attention and exert executive control to prioritise meaningful input and suppress interference. As the following section will discuss, this reliance on adaptive, effortful processing can pose particular challenges for autistic individuals, who often experience difficulties with attentional control or cognitive flexibility in complex listening environments.

1.3 Speech-in-noise processing in autism

This section reviews empirical evidence on how autistic individuals perceive speech in the presence of background noise, with an emphasis on the auditory mechanisms that facilitate or interfere with processing under such conditions. As this thesis focuses specifically on auditory rather than audiovisual processing, studies involving multimodal paradigms are not considered.

Difficulties with speech-in-noise (SiN) perception in autism are often discussed in relation to atypical sensory experiences, particularly heightened sensitivity to sound. Section 1.3.1 therefore begins by examining how auditory hypersensitivity and reduced sound tolerance may contribute to broader difficulties with sensory adaptation and noise filtering. The subsequent

subsections explore how autistic individuals respond to two types of interference: energetic masking, which results from overlapping sound energy that obscures the target signal, and informational masking, which involves interference at higher cognitive levels such as linguistic or attentional processing. Section 1.3.2 focuses on responses to energetic masking, while Section 1.3.3 turns to informational masking and considers how top-down mechanisms may influence SiN perception in autism. Section 1.3.4 concludes with a summary of key findings and offers an interpretation of these results in relation to broader theoretical frameworks.

1.3.1 Hypersensitivity

Auditory hypersensitivity is commonly reported in autism and is reflected in the DSM-5 as "hyper-reactivity to sensory input" (American Psychiatric Association, 2013). Many autistic individuals describe heightened sensitivity to loud, high-pitched, or unexpected sounds, which can feel overwhelming and trigger behaviours such as covering their ears, pulling away, or showing visible signs of discomfort (Gomes et al., 2008; Rosenhall et al., 1999; Williams et al., 2021). When auditory input becomes too intense or unpredictable, it can lead to sensory overload, characterised by distress, anxiety, and, in some cases, shutdowns or meltdowns (Belek, 2019; Charlton et al., 2021).

Neuroimaging research suggests that auditory hypersensitivity in autism may reflect a combination of heightened bottom-up reactivity and diminished top-down regulatory control. Increased activation in central auditory regions in response to moderately intense sounds points to atypical sensory gain mechanisms that may amplify incoming input (Foss-Feig et al., 2017; Rubenstein & Merzenich, 2003; Zimmerman et al., 2020). At the same time, reduced functional connectivity between auditory areas and regions involved in emotional and cognitive regulation suggests that top-down modulation of sensory input may be less effective (Linke et al., 2018; Matsuzaki et al., 2017; Vissers et al., 2012).

These auditory sensitivities may make it particularly difficult for autistic individuals to process speech in noisy environments, where even moderate background noise can disrupt their attention and participation (Anderson et al., 2018; Gelbar et al., 2014; Robertson & Simmons, 2015). Hyperacusis, defined as reduced tolerance to the loudness of everyday sounds, is a well-documented issue in autism and is often reported despite normal hearing thresholds (Danesh et al., 2021; Gomes et al., 2008; Khalfa et al., 2004). In such cases, noise that is perceived as harmless by non-autistic individuals may be experienced as intrusive or overwhelming, making

it harder to focus on speech. In addition to hyperacusis, autistic individuals often report increased sensitivity to background sounds that are unpredictable, complex, or beyond their control. These types of noise can be particularly disruptive and are associated with slower adaptation to noise and greater emotional discomfort (Stansfeld, 1992; Stansfeld & Clark, 2019). Collectively, these sensory differences can create additional perceptual and affective demands, making speech comprehension in noisy contexts more effortful and less effective.

1.3.2 Processing speech in energetic masking

Behavioural research on SiN processing in autism has largely focused on speech reception thresholds (SRTs) under energetic masking, particularly on listeners' sensitivity to dips in the masker. These dips are brief reductions in masker energy over time (temporal dips) or across frequency (spectral dips), which create short windows in which parts of the speech signal become more audible (see Section 1.2.1). Across studies, autistic individuals have shown reduced benefit from modulated maskers containing such dips (Alcántara et al., 2004; Dunlop et al., 2016; Groen et al., 2009; Ruiz Callejo et al., 2023; Schelinski & Von Kriegstein, 2020).

For example, Alcántara et al. (2004) found that non-autistic young adults improved their SRTs when temporal or spectro-temporal dips were present, whereas autistic participants did not. A similar pattern was observed in children, with Groen et al. (2009) reporting reduced benefit from temporal dips in the autistic group compared to controls. Building on these findings, Ruiz Callejo et al. (2023) showed in a larger sample that autistic individuals continued to show reduced benefit from temporally modulated noise, even after controlling for cognitive and language abilities. This suggests a core difference in how autistic individuals process the temporal structure of speech, consistent with evidence from studies in quiet settings that have identified atypical temporal processing in autism (see Section 1.1.1.2). Task demands may influence whether group differences emerge. While most studies compute SRTs using a 50% accuracy criterion, Schelinski and von Kriegstein (2020) employed a more stringent 75% threshold. Under this more challenging condition, autistic participants showed greater difficulty recognising speech in both temporally modulated and spectrally dipped noise. Additionally, they found that in the non-autistic group, better pitch perception was associated with better performance, whereas no such relationship was observed in the autistic group. This pattern suggests that pitch perception ability contributes less to SiN performance in autistic individuals compared to non-autistic individuals.

Neuroimaging studies have begun to elucidate the neural basis of these behavioural differences. Fadeev et al. (2024) used MEG to assess phoneme-level encoding in autistic and non-autistic children as they passively listened to synthetic vowels with preserved or disrupted phoneme-relevant spectral structure. A sustained processing negativity (SPN), maximal in non-primary regions of the left temporal cortex, was elicited in both groups by synthetic vowels with preserved phoneme-relevant spectral structure. However, SPN amplitude was significantly reduced in the autism group, particularly in the left hemisphere, suggesting diminished integration of spectral detail into phoneme-level units. Crucially, lower SPN amplitudes in the left hemisphere predicted poorer performance on a word recognition task in amplitude-modulated noise, linking atypical phonemic encoding to reduced benefit from temporal fluctuations in complex listening environments.

Extending this work to sentence-level processing in adults, two fMRI studies have reported reduced engagement of subcortical and cortical auditory systems in autistic adults during sentence recognition in stationary pink noise, despite comparable behavioural performance. At the subcortical level, only the non-autistic group showed enhanced activation in both left and right inferior colliculi (IC), whereas autistic participants exhibited this response only in the left IC, suggesting altered encoding of masked speech (Schelinski et al., 2022). At the cortical level, the autistic group demonstrated significantly weaker activation in the left inferior frontal gyrus (IFG) compared to the non-autistic group during SiN processing (Schelinski & Von Kriegstein, 2023). This reduced IFG activation in autism was interpreted as diminished engagement of top-down mechanisms such as articulatory simulation and predictive processing.

Evidence from MEG further implicates the IFG and disrupted top-down dynamics in autism. Mamashli et al. (2017) compared mismatch field (MMF) responses in autistic and non-autistic adolescents under quiet and noise conditions. Both groups showed typical MMF activity in the right superior temporal gyrus (STG) and IFG in quiet. However, under noise, the IFG MMF response was preserved in the non-autistic group but significantly reduced in the autistic group. This reduced recruitment of IFG in noise suggests atypical top-down regulation of auditory processing. Supporting this interpretation, autistic participants also showed reduced beta-band coherence between frontal and temporal regions, pointing to disrupted feedback connectivity.

A more socially oriented interpretation of these difficulties comes from Hernandez et al. (2020), who used fMRI to investigate how autistic and non-autistic youth process socially relevant speech in ecologically valid noisy environments. Although both groups showed similar activation in auditory and language regions, autistic participants performed significantly worse on recognition of questions in the SiN condition. Notably, successful speech recognition was associated with increased activation in the left angular gyrus, a region involved in both language comprehension and theory of mind. This activation correlated positively with social motivation and cognitive functioning, suggesting that some autistic individuals may engage this region as a compensatory mechanism to support the processing of socially relevant speech in complex auditory scenes. These findings reinforce that instead of low-level acoustic processing impairments, SiN difficulties in autism reflect differences in higher-level attentional and social-cognitive processes.

Recent research using eye-tracking techniques has also provided physiological evidence of the enhanced cognitive demands experienced by autistic individuals during SiN processing. Xu et al. (2024) assessed listening effort in Mandarin-speaking autistic and non-autistic school-age children using a sentence recognition task in steady-state, speech-shaped noise. Behavioural accuracy was comparable across groups; however, pupillometry revealed increased peak pupil dilation in autistic children, indicating heightened arousal, alongside reduced mean dilation, suggesting diminished sustained attentional engagement. These findings imply that autistic listeners may experience elevated cognitive load and inconsistent attentional resource allocation during speech perception, even under relatively simple masking conditions.

In summary, autistic individuals tend to gain less benefit from temporal and spectral fluctuations that typically support SiN perception. Even when behavioural performance appears intact, energetic masking may impose greater cognitive and self-regulatory demands. These challenges likely reflect both early-stage differences in encoding acoustic detail and reduced flexibility in engaging top-down attentional mechanisms.

1.3.3 Processing speech in informational masking

Informational masking poses a particular challenge for autistic individuals in everyday environments where competing speech and cognitive demands are high. In a large-scale self-report study, Bendo et al. (2024) found that many autistic adults struggled to understand speech in settings such as cafés, group conversations and classrooms, which are characterised by

overlapping speech and high levels of informational masking. These situations were frequently described as overwhelming or mentally exhausting.

These real-world listening difficulties under informational masking are mirrored in experimental research. Behavioural studies consistently show that autistic individuals perform worse than non-autistic peers when the masker contains linguistic content. For example, DePape et al. (2012) found that autistic adults showed significantly lower accuracy in identifying sentences masked by multi-speaker babble, despite comparable performance in stationary noise. Similarly, Bhatara et al. (2013) reported reduced accuracy in autistic adolescents under two-speaker babble, while performance in speech-shaped noise remained largely intact. Ruiz Callejo et al. (2023) replicated these findings using a controlled set of masker conditions. Participants were tested on sentence recognition tasks in stationary noise, eight-speaker babble, and a single competing speaker. Although both groups performed similarly in stationary noise, autistic participants were disproportionately affected by maskers containing speech. Notably, those with a history of early language delay performed particularly poorly in the single-speaker condition, suggesting that increased linguistic demands may exacerbate difficulties with auditory stream segregation. Dunlop et al. (2016) reported similar behavioural patterns, with autistic individuals showing poorer sentence recognition in multi-speaker babble but not in stationary noise. Importantly, their study also examined potential mechanisms underlying these difficulties, focusing on attentional filtering and sensory discomfort. Using the Auditory Attention and Discomfort Questionnaire, they found that autistic participants reported significantly greater attentional difficulties in situations involving background speech. These attentional ratings were closely aligned with performance, supporting the idea that central filtering difficulties contribute to reduced SiN recognition under informational masking. In contrast, auditory hypersensitivity scores did not differ between groups and were unrelated to task performance, suggesting that sensory discomfort alone cannot explain the observed challenges. These studies suggest that autistic individuals may experience greater difficulty with informational masking, and that their performance is shaped by individual differences in attentional control and language development history.

Neural studies investigating responses to socially salient speech cues under informational masking offer additional insight into how attention and language ability shape speech processing in autism. One such cue is one's own name (OON), which typically captures attention in multi-speaker environments. However, prior research has shown that autistic

individuals often fail to detect their name in these settings (e.g., Newman, 2005), and may show reduced orienting even in quiet conditions (Dawson et al., 2004; Miller et al., 2017), reflecting differences in selective attention, salience detection, and self-referential processing. Building on this work, Schwartz et al. (2020) used EEG to assess neural responses to OON under informational masking. They measured both early MMR and later positive potentials (LPP) while participants passively listened to multi-speaker scenes containing either their own name or a stranger's name. Non-autistic individuals and verbally fluent autistic participants showed stronger neural responses to their own name, whereas minimally and low-verbal participants did not reliably differentiate between the two. Crucially, larger LPP amplitudes in the multi-speaker condition were positively correlated with caregiver-rated auditory filtering abilities, suggesting that individual differences in attention modulate neural sensitivity to socially salient speech cues. These findings highlight the dynamic interplay between language ability and attentional control in shaping SiN processing in autism under informational masking.

Although socially salient cues such as OON can help guide attention, listeners more commonly rely on speaker-specific acoustic features (e.g., spatial location and vocal pitch) when navigating multi-speaker environments. This process, known as auditory scene analysis, requires listeners to perceptually organise complex auditory input into distinct streams and selectively attend to a target speaker (see Section 1.2.2). However, research suggests that autistic individuals may derive less consistent benefit from such acoustic cues. For example, Schafer et al. (2020) assessed speech recognition in multi-speaker babble. Although both groups showed improved performance when acoustic cues (i.e., spatial separation or gender differences between voices) were available, autistic individuals derived significantly smaller benefits from these cues than their non-autistic peers. Similarly, Emmons et al. (2022) used a selective attention task to test cue use in young autistic adults. While both groups performed best when spatial and voice gender cues were combined, autistic participants demonstrated lower overall accuracy and failed to show the typical attentional "switch cost" observed in non-autistic participants. These results suggest broader differences in how auditory attention is deployed and updated in dynamic listening contexts. Notably, this inefficient cue use persists even when groups are matched on verbal and nonverbal IQ (Lau et al., 2022).

Neural evidence further supports the idea that autistic individuals may process speaker-specific acoustic cues differently during stream segregation. Lepistö et al. (2009) used EEG to examine how children with Asperger syndrome passively processed tone sequences designed to either

form a single stream or segregate into two based on pitch differences. While MMN responses were typical in the integrated condition, children with AS showed significantly reduced MMN amplitudes when pitch cues were needed to segregate the streams, suggesting diminished pre-attentive discrimination of concurrent auditory input. Complementary findings come from Teder-Sälejärvi et al. (2005), who used a spatial selective attention task with competing speech streams presented from different locations. Autistic adults showed reduced ability to attend to a target location, indicating less efficient use of spatial cues to separate sound sources. Together, these studies suggest that atypical neural responses to pitch and spatial location cues may contribute to the reduced benefit autistic individuals derive from acoustic features in complex listening environments.

1.3.4 Summary

Auditory processing differences in autism span both sensory reactivity and challenges with speech perception. Hypersensitivity to everyday sounds is commonly reported and may contribute to sensory overload or emotional distress in noisy environments, affecting attention and engagement. However, hypersensitivity alone does not reliably predict performance in SiN tasks (Dunlop et al., 2016). Behavioural difficulties are evident under both energetic and informational masking, particularly in tasks requiring flexible attention and the use of acoustic cues such as temporal dips (Alcántara et al., 2004; Groen et al., 2009) and speaker-related features (Emmons et al., 2022; Lau et al., 2022; Schafer et al., 2020). Notably, even when accuracy is preserved, studies report elevated listening effort (Xu et al., 2024) and altered neural responses (Schelinski et al., 2022; Schwartz et al., 2020).

Neural findings converge on disruptions across both bottom-up and top-down stages of auditory processing. Autistic individuals show atypical encoding of speech-relevant features, including reduced sustained responses to phoneme-level spectral cues (Fadeev et al., 2024), altered mismatch responses to vocal pitch (Lepistö et al., 2009), and early-stage subcortical atypicalities during SiN processing (Schelinski et al., 2022). At higher levels, cortical regions involved in language processing, articulatory simulation, and predictive coding show reduced and less flexible activation in noisy conditions (Schelinski & von Kriegstein, 2023; Mamashli et al., 2017). Additional evidence for disrupted top-down engagement includes reduced beta-band coherence between frontal and temporal areas, indicating impaired coordination between control and sensory systems (Mamashli et al., 2017), and the absence of attentional enhancement for spatial cues during localisation (Teder-Sälejärvi et al., 2005).

These findings can be interpreted through several theoretical frameworks proposed to account for sensory and cognitive processing in autism. The reduced benefit from acoustic cues, along with diminished phoneme integration and subcortical differences, challenges accounts of Enhanced Perceptual Functioning, which posit heightened low-level perceptual sensitivity. Instead, these results align more closely with the Weak Central Coherence theory, which suggests a bias toward local over global processing. In the context of SiN, this may result in reduced integration of temporally distributed speech features into coherent linguistic units. At the top-down level, inflexible modulation of cortical speech-related regions, reduced coordination across brain systems, and attenuated attentional enhancement of spatial cues support Predictive Coding models, which propose that autistic individuals generate imprecise or overly rigid internal predictions about incoming sensory input. Disrupted functional connectivity may also reflect reduced neural integration and flexibility, consistent with the Neural Complexity Hypothesis. Finally, compensatory engagement of language and social-cognitive regions during successful SiN recognition, particularly among individuals with higher social motivation and cognitive ability (Hernandez et al., 2020), alongside reduced orienting to salient social stimuli (e.g., one's own name; Schwartz et al., 2020), is in line with the Social Motivation Theory, highlighting the role of motivational and social-cognitive factors in shaping auditory attention.

Notably, while these group-level findings point to specific neural and behavioural mechanisms underlying SiN difficulties in autism, considerable heterogeneity remains. The extent and nature of these processing differences vary substantially across individuals, likely shaped by factors such as early language delay (Callejo et al., 2023), cognitive ability (Lau et al., 2022), and attentional filtering capacity (Dunlop et al., 2016). These differences are observed across developmental stages, from childhood (Groen et al., 2009; Xu et al., 2024) to adulthood (Schelinski & von Kriegstein, 2023). In more complex cases, children with co-occurring language difficulties exhibit broader auditory processing differences that affect performance even in quiet listening conditions (Russo et al., 2009).

1.4 The current thesis

The previous sections highlighted that successful speech-in-noise (SiN) perception depends on the dynamic coordination of bottom-up auditory encoding and top-down mechanisms,

including attentional control and semantic integration. These demands intensify in complex acoustic environments, where listeners must track and interpret speech amid competing sound streams. In autism, difficulties in such contexts are associated with both reduced encoding of acoustic features and atypical higher-order modulation involving language and attention.

Despite valuable insights from existing research, several key limitations constrain our understanding of SiN processing in autism. First, relatively few studies have used EEG to investigate how autistic individuals process speech in complex auditory environments, leaving limited insight into their real-time processing. Even for speech processing in the absence of noise, most neurophysiological research has focused on early-stage auditory encoding using highly controlled, decontextualised stimuli such as isolated syllables or words (see Section 1.1). While such designs capture low-level responses, they provide limited insight into how autistic individuals construct meaning over time or how semantic interpretation interacts with acoustic encoding in real-world listening.

Second, the range of masking conditions used in previous research has been relatively narrow. Most studies rely on energetic maskers, such as modulated noise, which limits our understanding of how autistic individuals respond to ecologically valid challenges. Although a few studies have used informational maskers, such as competing speech, they often overlook how the increased social and linguistic demands of these maskers influence processing. Other naturalistic non-speech maskers, such as background music, have been almost entirely neglected, despite their frequent presence in everyday listening environments. This gap is especially relevant given evidence that autistic individuals often show reduced attention to speech and less efficient use of linguistic and semantic context, alongside relatively intact early acoustic processing (see Sections 1.1 and 1.2). At the same time, enhanced sensitivity to music-related acoustic features, particularly pitch and melodic structure, is a salient trait observed in many autistic individuals (see Section 1.1.1.1). These co-occurring features constitute a pattern that implies an asymmetric competition between speech and music for attentional and linguistic processing resources in autism. Background music during speech processing may therefore exert a stronger influence on autistic individuals by drawing on resources that would otherwise be allocated to the target speech. Accordingly, examining background music as a masker is a theoretically grounded extension of existing work on speech-in-noise processing in autism.

Finally, many existing studies have not adequately accounted for variability in language and cognitive abilities, raising the possibility that observed group differences in SiN processing may partly reflect sample heterogeneity rather than fundamental processing differences.

Overall, these gaps highlight the need for more ecologically valid and temporally sensitive approaches to investigate how acoustic and semantic processing unfold in realistic auditory scenes. The present thesis addresses these challenges by combining behavioural and EEG methods to examine how autistic and non-autistic individuals process continuous speech under background conditions that vary in acoustic structure and intelligibility. It focuses on two central research questions:

(1) How do different types of background sound affect speech comprehension in autistic and non-autistic individuals?

(2) What processing strategies are employed under varying levels of auditory and semantic demand?

This work aims to provide a more comprehensive account of the mechanisms underlying SiN processing in autism. Across three empirical studies, the thesis employs time-sensitive neural measures and advanced analytical techniques, each tailored to the demands of distinct listening scenarios involving different sources of auditory and semantic interference. Study 1 (Chapter 2) investigates whether autistic individuals benefit from bottom-up acoustic cues related to the target speaker (i.e., voice gender and spatial location) when attempting to follow speech in complex auditory scenes. Participants completed a speech recognition task in the presence of competing speech and instrumental background music. Importantly, no explicit cue was provided regarding the presence or features of the maskers, simulating real-world demands for internally guided auditory selection. In addition to linear mixed-effects models, generalised additive mixed models were used to examine fine-grained performance changes over time. Results showed that both groups benefited from acoustic cues; however, autistic participants were overall less accurate and demonstrated smaller improvements over time in the absence of such cues, suggesting less efficient use of salient acoustic information. Instrumental music had a similar impact on both groups, but autistic individuals with a stronger local processing bias were more susceptible to interference, indicating that musical masking may interact with individual perceptual styles rather than diagnostic group alone.

Study 2 (Chapter 3) examines how different types of background music influence semantic processing, focusing on the degree of linguistic content in the masker. Participants listened to sentences presented with one of three types of background music: instrumental music (no lyrics), music with unintelligible Simlish lyrics, or music with intelligible English lyrics. This design allowed for a graded investigation of linguistic interference. Semantic processing was assessed using both behavioural accuracy and the N400 effect. Non-autistic participants showed clear sensitivity to the type of masker: they exhibited larger N400 responses in the instrumental condition, with progressively reduced N400 amplitudes as the linguistic content of the masker increased, paralleling a decline in comprehension accuracy. Autistic participants, by contrast, demonstrated overall lower accuracy and reduced N400 amplitudes, with minimal differentiation across conditions, suggesting less modulation of semantic processing in response to the linguistic characteristics of the masker.

Study 3 (Chapter 4) investigates both acoustic and semantic processing using continuous speech presented in quiet, babble noise, and competing speech conditions. This study combines neural tracking of the speech envelope using forward-modelled temporal response functions (TRFs) with N400 responses to semantic processing. This dual approach allows examination of how bottom-up and top-down mechanisms co-occur and potentially trade off under varying levels of listening difficulty. Autistic participants showed significantly reduced TRF amplitudes and delayed N400 onset, indicating less efficient auditory encoding and slower semantic processing. Despite similar N400 amplitudes and behavioural performance between groups, distinct group-level differences emerged in how neural resources were allocated. In the competing speech condition, non-autistic participants exhibited a shift in strategy: increased semantic integration (larger N400s) was accompanied by reduced envelope tracking (smaller TRFs), consistent with a trade-off between encoding and interpretation under linguistic competition. Autistic participants showed no such modulation across conditions, suggesting reduced flexibility in adapting to changing processing demands.

Collectively, the three studies offer a layered account of SiN processing in autism. Study 1 showed that although both groups benefited from bottom-up acoustic cues, autistic individuals exhibited reduced adaptation in their absence and were more susceptible to interference from music, particularly depending on individual perceptual styles. Studies 2 and 3 revealed that, even with comparable behavioural accuracy, autistic individuals demonstrated reduced neural

encoding of both acoustic and semantic information, alongside differences in neurocognitive strategy. These included less flexible adjustment to contextual demands and weaker coordination between acoustic and semantic processing. Overall, the findings indicate that SiN difficulties in autism involve both auditory and semantic processes and are characterised by reduced processing flexibility across varying task demands. Chapter 2–4 present each study in detail. Chapter 5 synthesises the findings across studies and situated them within key theoretical accounts of autism, highlighting how atypical auditory and semantic processing reflect broader differences in processing mechanism.

Chapter 2

Study 1: The Impact of Acoustic Cues and Background Music on Speech Perception in Autism

Abstract

Recognising speech in noise involves focusing on a target speaker while filtering out competing voices and sounds. Acoustic cues, like vocal characteristics and spatial location, help differentiate between speakers. However, autistic individuals may process these cues differently, making it more challenging for them to perceive speech in such conditions. This study investigated how autistic individuals use acoustic cues to follow a target speaker and whether background music increases processing demands. Thirty-six autistic and 36 non-autistic participants identified information from a target speaker while ignoring a competing speaker and background music. The competing speaker's gender and location either matched or differed from the target. The autistic group exhibited lower mean accuracy across cue conditions, indicating general challenges in recognising speech in noise. Trial-level analyses revealed that while both groups showed accuracy improvements over time without acoustic cues, the autistic group demonstrated smaller gains, suggesting greater difficulty in tracking the target speaker without distinct acoustic features. Background music did not disproportionately affect autistic participants but had a greater impact on those with stronger local processing tendencies. Using a naturalistic paradigm mimicking real-life scenarios, this study provides insights into speech-in-noise processing in autism, informing strategies to support speech perception in complex environments.

Keywords: acoustic cues, autism, background music, speech-in-noise processing

2.1 Introduction

Listening in a cocktail-party environment, where multiple speakers and background sounds overlap, places significant demands on both bottom-up and top-down processing mechanisms (Bronkhorst, 2000). Bottom-up processes enable the detection of acoustic differences between competing auditory streams, while top-down mechanisms, including selective attention and cognitive control, guide the selection of relevant auditory information and help suppress

irrelevant distractions. Together, these mechanisms integrate acoustic information into cohesive auditory objects, facilitating effective speech recognition in noise (Başkent & Gaudrain, 2016; Shinn-Cunningham & Best, 2008).

Autistic individuals often face challenges with speech-in-noise (SiN) recognition due to differences in auditory processing and cognition (O'Connor, 2012; Ruiz Callejo & Boets, 2023). These difficulties may exacerbate social communication difficulties (American Psychiatric Association, 2013), as social interaction often occurs in complex listening environments. Behaviourally, autistic individuals struggle to use brief reductions in noise intensity to recognise target speech (Alcántara et al., 2004; Groen et al., 2009; Schelinski & Von Kriegstein, 2020). Electrophysiological studies further report reduced spatial attention and diminished frequency discrimination in the presence of competing auditory streams (Lepistö et al., 2009; Teder-Sälejärvi et al., 2005). Moreover, attenuated neural encoding of vowels predicts difficulties in word-in-noise recognition, suggesting the effect of early acoustic disruptions on SiN comprehension (Fadeev et al., 2024). Collectively, these results indicate difficulties in extracting and integrating acoustic information during SiN processing.

In multi-speaker environments, differences in spatial location and vocal characteristics can be used to segregate overlapping speech and follow the target speaker (Culling et al., 2003; Darwin et al., 2003; Shinn-Cunningham et al., 2017). However, how autistic individuals use these cues remains unclear, despite frequent reports of every-day difficulties in managing competing voices (Bendo et al., 2024). Emmons et al. (2022) addressed this by examining whether participants could use gender and location cues to direct attention between competing speakers. Non-autistic participants performed at ceiling with either cue alone, whereas autistic participants showed better performance only when both cues were available, indicating a greater reliance on multiple cues. Lau et al. (2023) found that lower SiN recognition in a three-speakers scenario was associate with lower IQ, but they did not directly examine how acoustic cues affected performance.

Another socially relevant yet often overlooked challenge is the interference of background music with speech recognition (Brown & Bidelman, 2022; Russo & Pichora-Fuller, 2008; Shi & Law, 2010). Autistic individuals often prefer music over speech, likely due to its structured, predictable, and emotionally resonant qualities, in contrast to the nuanced variability and social complexity of spoken language (Allen et al., 2009; Kuhl et al., 2005). Neural evidence suggests

that autistic children show stronger brain responses to music than to speech or environmental noise, indicating heightened sensitivity to musical input (Molnar-Szakacs & Heaton, 2012). Many autistic individuals also demonstrate enhanced musical abilities including superior pitch perception and melodic memory despite well-documented difficulties with speech and language processing (Heaton, Williams, et al., 2008; O'Connor, 2012; Ouimet et al., 2012). These findings suggest that music is particularly salient for autistic individuals and may draw greater perceptual and cognitive resources when presented concurrently with speech, potentially making it a more disruptive distractor during speech processing.

The current study addressed two key research questions:
1. Do autistic participants benefit from acoustic cues in resolving SiN challenges in a two-speaker scenario?

2. Does background music impose greater processing demands on autistic listeners compared to non-autistic listeners?

Participants were asked to identify speech from a target speaker presented simultaneously with a distractor speaker and instrumental background music. The spatial location and gender of the distractor were systematically manipulated, creating four conditions: no-cue, gender-cue, location-cue, and both-cues conditions. Importantly, the relative loudness of the speech compared to the background noise (i.e., signal-to-noise ratio, SNR) was kept fixed throughout the experiment. This contrasts with previous studies that adaptively varied SNRs to estimate speech detection thresholds (Alcántara et al., 2004; Groen et al., 2009). We maintained a fixed SNR to account for auditory hypersensitivity commonly reported in autism (Williams et al., 2021), as the gradual increase in noise used in adaptive procedures could cause sensory discomfort and, in turn, confound measures of speech perception (Danesh et al., 2021; Khalfa et al., 2004).

Beyond group comparisons of mean accuracy, we used Generalised Additive Mixed Models (GAMMs) (Wood, 2011, 2017) to analyse accuracy trajectories over trials across cue conditions and groups. By tracking changes in performance over time, this approach captured non-linear patterns that may reflect improvement, attentional shifts, or fatigue across the task. Building on prior evidence suggesting less efficient use of auditory cues as well as general SiN processing difficulties in autism, we expected autistic participants to show lower overall

accuracy across conditions and slower improvement over trials, particularly in conditions with fewer available cues. This would reflect greater difficulty in tracking the target speaker when salient acoustic distinctions are absent. We also hypothesised that background music would interfere more with speech recognition in the autistic group, based on previous findings that music is often more salient, emotionally engaging, and perceptually preferred over speech in autistic individuals (Heaton, Williams, et al., 2008; Molnar-Szakacs & Heaton, 2012; Ouimet et al., 2012).

Successful performance on this task requires participants to recognise and effectively utilise acoustic cues to segregate a target speaker from competing streams. This involves perceiving individual features and integrating them into a coherent auditory object over time, a process may be particularly demanding for autistic individuals. According to the Weak Central Coherence theory, autistic individuals show a cognitive bias toward local over global information, which may affect their ability to combine multiple auditory cues across time and sources (Happé & Frith, 2006). Predictive coding accounts offer a complementary explanation, suggesting reduced reliance on top-down predictions, which may limit their ability to anticipate and filter relevant speech in noisy or unpredictable contexts (Van De Cruys et al., 2014). Neurobiological accounts further propose that reduced functional connectivity and lower signal complexity may impair the integration of acoustic information into a coherent target stream (Belmonte et al., 2004; Just et al., 2012).

Accordingly, if these perceptual tendencies constrain acoustic integration in the current task, other cognitive mechanisms such as working memory and reasoning may compensate to support task performance. Conversely, if autistic individuals demonstrate heightened sensitivity to local, low-level features, as proposed by the Enhanced Perceptual Functioning account (Mottron et al., 2006), we might expect stronger associations between pitch discrimination ability and task accuracy, reflecting a reliance on fine-grained auditory detail. Therefore, to assess whether theoretically motivated individual differences contribute to SiN performance, we conducted a correlational analysis examining the associations between task accuracy and nonverbal IQ, working memory, pitch discrimination, and local-to-global processing style. This approach allowed us to evaluate whether these cognitive factors support cue-based listening and whether different mechanisms may be involved across groups.

2.2 Methods

Participants

A power analysis determined a target sample size of 70 participants (35 per group), providing nearly 80% power to detect most effects of interest (see Appendix A). Ultimately, we recruited 36 autistic and 36 non-autistic native English speakers aged 16 to 47. All participants had normal or corrected-to-normal vision, no colour blindness, and normal pure-tone hearing levels, at 0.5, 1, 2, and 4 kHz. Both groups had no current speech, language, or communication needs. Clinical diagnoses were confirmed for all autistic participants, while non-autistic participants had no personal or family history of autism and scored below 32 on the Autism Spectrum Quotient (AQ; Baron-Cohen et al., 2001)

To account for potential factors influencing SiN processing, we assessed cognitive and auditory abilities as well as musical training background (see Table 2-1). Cognitive measures included nonverbal IQ (Raven's Standard Progressive Matrices; Raven & Court, 1998), receptive vocabulary (ROWPVT-4; Martin & Brownell, 2011), and verbal short-term memory (digit span task; Wechsler et al., 2003). Cognitive processing style was assessed using Navon's paradigm (Navon, 1977) in which participants responded to composite letters in congruent (e.g., a large H composed of small Hs) or incongruent (e.g., a large H composed of small Ss) configurations. Two metrics were derived: global advantage (RT difference for global vs. local judgements on congruent trials, indicating a bias toward global processing) and local-to-global interference (RT difference for global judgements between congruent and incongruent trials, reflecting difficulty prioritising global over conflicting local information). Navon scores were based on 62 participants, as 10 (5 autistic, 5 non-autistic) did not complete the task. Auditory abilities were evaluated using a pitch direction discrimination task (Liu et al., 2010), where thresholds were determined through a "two down, one up" adaptive staircase method. Musical training background was measured as self-reported years of formal instrumental and vocal training, with total training years used as the primary metric (Pfordresher & Halpern, 2013).

Wilcoxon rank-sum tests revealed no significant differences between the autistic and non-autistic groups on key demographic and cognitive variables. Autistic participants scored significantly higher on the AQ. They also exhibited higher local-to-global interference scores on the Navon task, suggesting increased difficulty in prioritising global over local information.

This study received ethical approval from the University Research Ethics Committee. All participants provided written informed consent before participating and received financial compensation or course credits for their involvement.

Table 2-1. Characteristics of the autistic (*n* = 36) and non-autistic (*n* = 36) groups.

Variables	Autistic *Mean (SD)*	Non-autistic *Mean (SD)*	*W*	*p*	Rank-biserial Correlation
Gender (Female:Male:Others)	22:12:2	28:7:1			
Age	23.29 (5.60)	23.54 (5.99)	641.0	.94	−0.01
Musical training years	5.01 (6.67)	5.67 (7.67)	608.5	.65	−0.06
Nonverbal reasoning (RSPM raw score)	54.00 (3.68)	54.08 (3.26)	669.5	.81	0.03
Nonverbal reasoning (RSPM percentile)	51.94 (25.62)	51.5 (26.43)	661.5	.88	0.02
Receptive vocabulary (ROWPVT-4 raw score)	166.56 (10.77)	167.50 (10.77)	633.0	.87	−0.02
Receptive vocabulary (ROWPVT-4 standard score)	111.25 (16.27)	112.47 (14.15)	622.0	.77	−0.04
Digit Span	7.03 (1.48)	7.25 (1.18)	562.0	.32	−0.13
Pitch threshold	0.22 (0.10)	0.33 (0.47)	644.5	.97	−0.01
Global advantage	107.87 (103.21)	102.88 (43.99)	467.0	.85	0.03
Local-to-global interference	20.41 (26.88)	15.21 (76.89)	652.0	.02	0.36
Autistic traits (AQ)	37.89 (7.20)	16.08 (7.53)	1258.0	< .01	0.94

Stimuli and apparatus

The target and distractor speech stimuli were sourced from the Children's Coordinate Response Measure corpus (Messaoud-Galusi et al., 2011), recorded by three Southern Standard British English speakers. Each sentence followed the structure: *"Show the ANIMAL (dog/cat) where the COLOUR (black/blue/green/pink/red/white) NUMBER (1/2/3/4/5/6/8/9) is."* The number "7" was excluded due to its two-syllable pronunciation, making it easier to distinguish from the other numbers. To direct attention, the callsign "dog" was always used for the target speaker, while distractor speakers used the callsign "cat".

Figure 2-1 provides an overview of the experimental design. Acoustic cues were manipulated by varying the distractor's gender (matching or differing from the target) and spatial location (co-located or separated from the target), resulting in four conditions: 1) no-cue: matched gender, co-located position; 2) gender-cue: mismatched gender, co-located position; 3) location-cue: matched gender, separated position; 4) both-cues: mismatched gender, separated position. Spatial separation was achieved using binaural head-related transfer functions, which simulate realistic spatial positioning over headphones (Wenzel et al., 1993; Wightman & Kistler, 1989). In our setup, the target speaker was fixed at 0° azimuth, while distractors were either co-located or positioned at −45° (left) or +45° (right). This was experienced by participants as the distractor voice shifting toward the left or right ear, making it perceptually distinct from the centrally presented target.

To evaluate the effect of background music, half of the trials included peaceful instrumental music spatially mirrored to the distractor speaker's location. Music stimuli were derived from a validated set of film music excerpts developed to reflect distinct emotional qualities (Eerola & Vuoskoski, 2011). To minimise emotional or semantic interference while maintaining ecological validity, we chose excerpts characterised by high valence and low arousal. These pieces feature a slow tempo, smooth dynamics, and soft timbres. To avoid perceptual discontinuities and ensure smooth transitions, we selected each music excerpt from the middle section of the original track, avoiding the beginning and end where dynamic or structural changes are more likely to occur. This approach ensured that all clips maintained consistent volume and texture throughout without the need for artificial fade-ins or fade-outs.

The target speech was played at 55 dB SPL through headphones, with background music (when present) adjusted to a 0 dB SNR, ensuring equal intensity between the music and target speech. To balance task difficulty with performance feasibility and ecological validity, we included two SNR levels (−9 dB and −3 dB) determined through pilot testing (see Appendix B). Neurotypical participants achieved 60–70% accuracy at −9 dB across conditions, indicating substantial but manageable difficulty. The −3 dB level was included to ensure the task remained accessible to autistic individuals who may experience heightened sensitivity to sound intensity, potentially leading to discomfort or disengagement in more challenging conditions (Danesh et al., 2021; Khalfa et al., 2004). These effects may arise at a sensory processing level, independent of auditory segregation ability. Including only −9 dB trials could risk floor effects or excessive sensory load in the autistic group. The −3 dB level was therefore added to reduce

the overall sensory burden. Notably, SNR was not analysed as an experimental variable, and participants were not informed of the level changes during the experiment. Instead, the two SNR levels were randomly intermixed within blocks to introduce a naturalistic and unpredictable listening environment. This approach prioritised participant accessibility, ecological validity, and engagement, while avoiding potential confounds related to sensory sensitivity and ensuring that the task remained feasible and realistic across both groups.

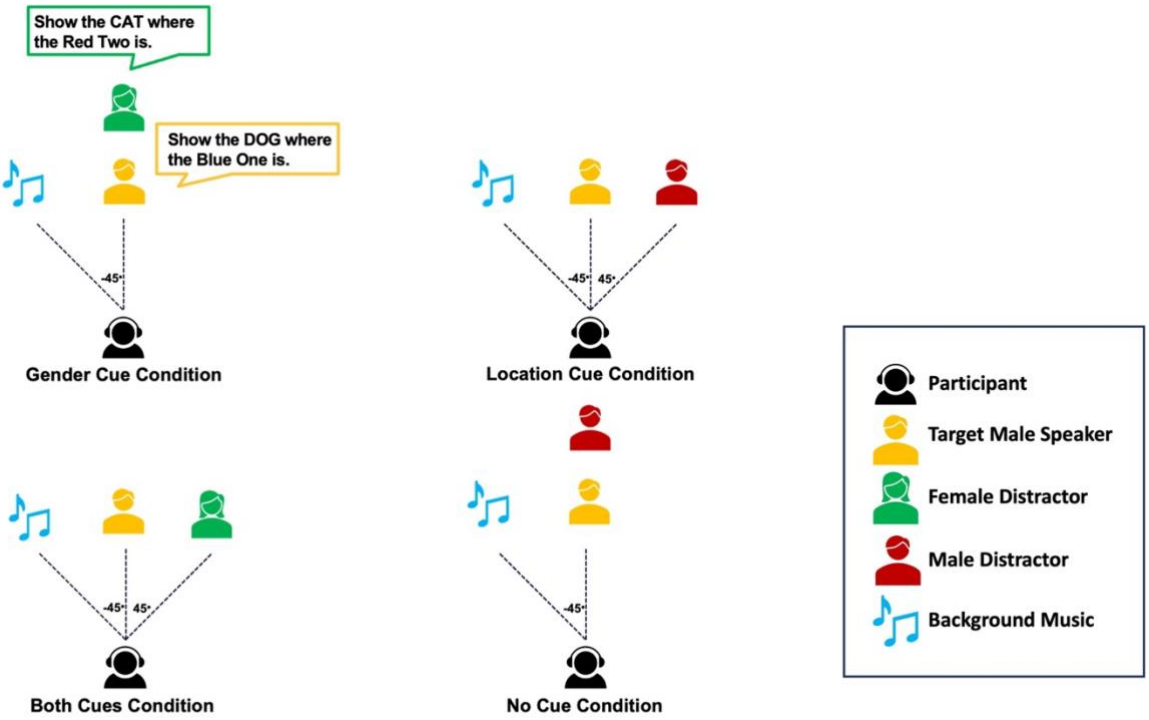

Figure 2-1. Schematic representation of design. The figure illustrates a single configuration of distractor and music locations; in the actual experiment, these locations vary dynamically across trials, with the music positioned symmetrically opposite the distractor.

Procedure

The experiment was implemented and presented using PsychoPy (version 2022.2.2; Peirce et al., 2019). On each trial, participants listened to either two or three simultaneous auditory streams delivered via headphones including a target speaker, a distractor speaker, and, in some trials, background music. Participants were instructed to focus on the target speaker, identified by the callsign "dog" and report the associated colour-number combination while ignoring the distractor speaker using the callsign "cat". Following the auditory stimulus, participants were shown an on-screen response grid containing all possible colour-number combinations (see Figure 2-2) and responded by clicking on the corresponding-coloured number box as quickly and accurately as possible using a computer mouse. A correct response required selecting both

the correct colour and number. The target and distractor speakers never shared the same callsign, colour, or number. Before the experiment, participants were informed that music might be present in some trials but were not given pre-trial cues about its presence.

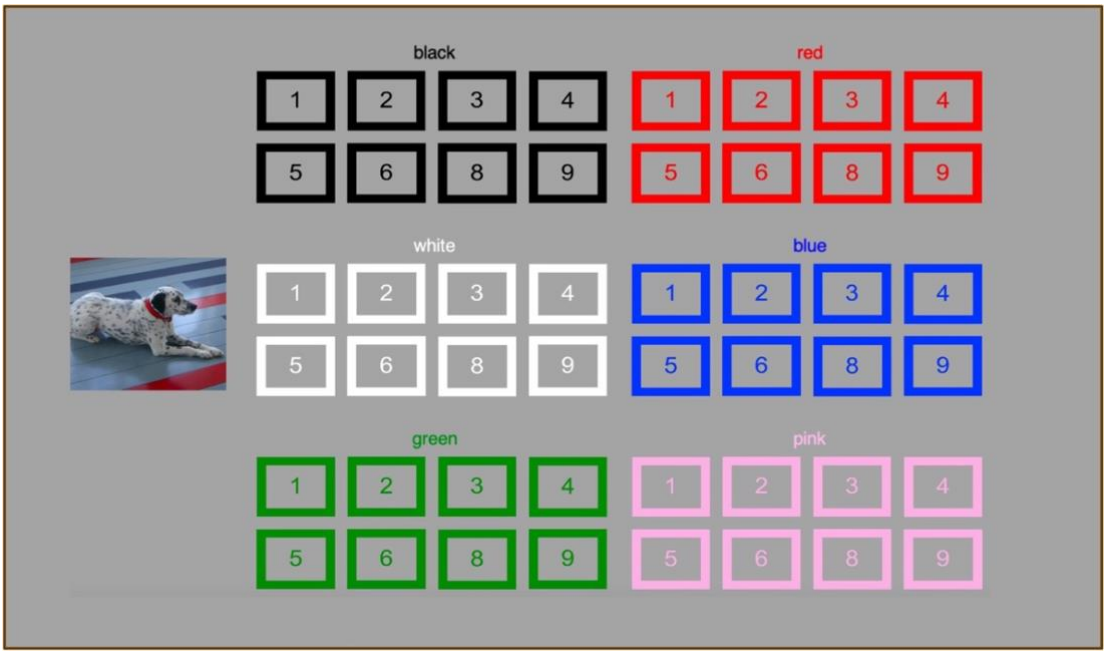

Figure 2-2. Response screen used in the sentence identification task. The image of the dog (left) represents the target callsign that participants were instructed to attend to. Six colour-coded response panels were displayed, each containing numbered response options. Participants responded by using the mouse to click on the square that matched the correct colour-number combination associated with the keyword spoken by the target speaker.

The experiment consisted of 288 trials, combining 48 unique colour-number pairs (six colours × eight numbers), two distractor genders (male, female), and three spatial locations (co-located, left, right). For analysis, performance in left and right distractor conditions was averaged to represent conditions with location cues. Trials were randomly presented across six blocks, with breaks between blocks to minimise fatigue. Conditions were mixed within each block to prevent participants from anticipating the presence of background music or specific acoustic cues. No prior information was given about the target speaker's gender or location. Before the main experiment, participants completed a training session consisting of eight trials with feedback to confirm their understanding of the task and the audibility of the target sentences.

Statistical analysis

All analyses were conducted in R (version 4.1.2, Posit Team, 2022). To examine the three-way interaction between acoustic cues, background music, and group, linear mixed-effects models

(LMMs) were constructed using the lme4 package (Bates et al., 2015). Accuracy was analysed using generalised linear mixed-effects models (GLMMs) with the BOBYQA optimiser. Reaction times (RTs) for correct responses were analysed using LMMs, with RTs log-transformed to correct for positive skewness. Outliers exceeding three standard deviations from each participant's mean RTs across conditions (< 2% of the data) were excluded.

Gender-cue and location-cue conditions were averaged into a single one-cue condition. Helmert coding was applied to compare cue conditions: 1) cue1 (no cue vs. any cues): No cue = 2/3; one cue = −1/3; both cues = −1/3; 2) cue2 (one cue vs. both cues): No cue = 0; one cue = 1/2; both cues = −1/2. "Any cues" refers to trials where at least one cue was present, encompassing both the one-cue and both-cues conditions. Fixed effects in the models included group (autistic = 1/2, non-autistic = −1/2), background music (without music = 1/2, with music = −1/2), acoustic cue (cue1, cue2), and their interactions.

Models were first fitted with maximal random effects structures, including by-participant and by-item random intercepts and slopes for all relevant fixed effects (Barr, 2013). Due to convergence issues, the structure was simplified in stages: first by removing correlations among random effects, then by removing random intercepts, and finally by adopting a forward selection approach. This involved starting with a model containing only random intercepts and incrementally adding random slopes, retaining only those that significantly improved model fit based on likelihood ratio tests. The final model reflected the most complex convergent structure. Fixed effects and interactions were assessed via likelihood ratio tests by comparing the final model to nested models with specific effects removed. Significant interactions were explored through simple effects analyses on subsetted data. All follow-up models used the most complex convergent structure shared across subsets. Bonferroni correction was applied.

To investigate the effects of group and cue condition on accuracy over time, we conducted a GAMM analysis using the mgcv (Wood, 2011, 2017) and itsadug packages (van Rij et al., 2015). Tensor function plots were generated to visualise interaction effects, identifying time windows of significant differences across group and condition (focusing on the no-cue and both-cues conditions). The SNR levels were randomly presented across conditions, which could potentially confound the Group × Condition interaction effect across trials. To address this, we constructed separate GAMMs for each SNR (see details in Appendix C).

A Pearson correlation analysis was conducted to examine the relationship between individual cognitive factors and task performance. A total of nine variables were included, resulting in 36 unique pairwise correlations. To maintain the integrity of our hypothesis-driven analysis, we did not apply multiple corrections, as this could obscure meaningful effects. This approach is consistent with recent methodological guidance suggesting that such corrections are not always necessary when testing a small number of *a priori* hypotheses and when no omnibus null hypothesis is being evaluated (García-Pérez, 2023). Factors included nonverbal IQ, working memory, musical training, and pitch processing ability, all previously linked to SiN processing (Gordon-Salant & Cole, 2016; Heinrich, 2021). Navon task scores were also examined to assess global–local processing style, given evidence of local processing bias in autism (Happé & Frith, 2006), which could influence the ability to integrate auditory information. Receptive vocabulary was not included, as our use of consistent sentence structures minimised lexical demands.

We examined three performance measures: overall accuracy, accuracy in the no-cue condition (the most difficult), and the background music effect (the accuracy difference between without- and with-music conditions). To normalise percentage accuracy scores, the rationalised arcsine transformation was applied before analysis (Studebaker, 1985).

2.3 Results

LMMs

Figure 2-3 displays the mean accuracy across cue conditions and groups, while Table 2-2 summarises the model results. A significant main effect of group revealed that autistic participants (M = 85.9%, SD = 34.8%) exhibited lower accuracy than their non-autistic counterparts (M = 88.9%, SD = 31.4%). Significant main effects were found for both acoustic cue contrasts. Accuracy was lower in the no-cue condition than in trials with at least one cue. Additionally, accuracy in the one-cue condition (gender: M = 91.7%, SD = 27.5%; location: M = 93.4%, SD = 24.9%) was lower than in the both-cues condition (M = 94.6%, SD = 22.5%).

There was a significant main effect of background music. Accuracy was lower in the with-music condition (M = 84.6%, SD = 36.1%) compared to the without-music condition (M = 90.1%, SD = 29.9%). We also observed significant three-way interactions between group, music and each cue contrast. To follow up, we conducted separate analyses for each interaction.

Interaction between group, music, and cue1 (no-cue vs. any cues). We first examined the group-by-cue1 interaction in trials with and without background music ($\alpha = .025$). There was no significant interaction between cue1 and group in both trials with music, $\chi^2(1) = 2.60$, $p = .107$, OR = 0.76, 95% CI = [0.55, 1.05]; and trials without background music, $\chi^2(1) = 0.83$, $p = .363$, OR = 1.18, 95% CI = [0.83, 1.68]. Next, we investigated the music-by-cue1 effect within each group. The interaction between music and cue1 was not significant for either the non-autistic group, $\chi^2(1) = 0.01$, $p = .933$, OR = 1.04, 95% CI = [0.45, 2.38], or the autistic group, $\chi^2(1) = 2.04$, $p = .153$, OR = 1.81, 95% CI = [0.82, 4.02].

Interaction between group, music, and cue2 (one-cue vs. both-cues). We then examined the group-by-cue2 interaction in trials with and without background music ($\alpha = .025$). There was no significant interaction between cue2 and group in trials with music, $\chi^2(1) = 0.03$, $p = .854$, OR = 0.97, 95% CI = [0.67, 1.39]; or without music, $\chi^2(1) = 1.00$, $p = .316$, OR = 1.30, 95% CI = [0.80, 2.11]. Similarly, we investigated the music-by-cue2 effect within each group. The interaction between music and cue2 was significant for the non-autistic group, $\chi^2(1) = 8.15$, $p = .004$, OR = 0.18, 95% CI = [0.05, 0.60], but not for the autistic group, $\chi^2(1) = 2.16$, $p = .142$, OR = 0.43, 95% CI = [0.14, 1.30]. Following this, we examined the simple effect of background music within the non-autistic group for each cue2 condition ($\alpha = .0125$). In the both-cues condition, accuracy was significantly lower in trials with background music compared to those without background music, $\chi^2(1) = 23.64$, $p < .001$. This effect was reflected in the odds ratio (OR = 11.78, 95% CI [4.08, 34.00]), indicating a markedly higher likelihood of correct responses in the absence of background music. However, in the one-cue condition, no significant music effect was observed, $\chi^2(1) = 0.24$, $p = .623$, OR = 1.14, 95% CI = [0.66, 2.01].

In summary, the analysis of the three-way interactions indicated a significant effect only in the non-autistic group: accuracy in the both-cues condition was lower with background music than without. No other comparisons were significant. In response to the examiner's comment, we conducted a supplementary analysis in which gender-cue and location-cue conditions were entered into the model as separate cue conditions, rather than being combined into a single one-cue condition. This analysis revealed no significant differences between the gender-cue and location-cue conditions. All other effects were consistent with those reported in the main analysis. Full details of the supplementary analysis, including models that incorporate all four cue conditions separately, are provided in Appendix D.

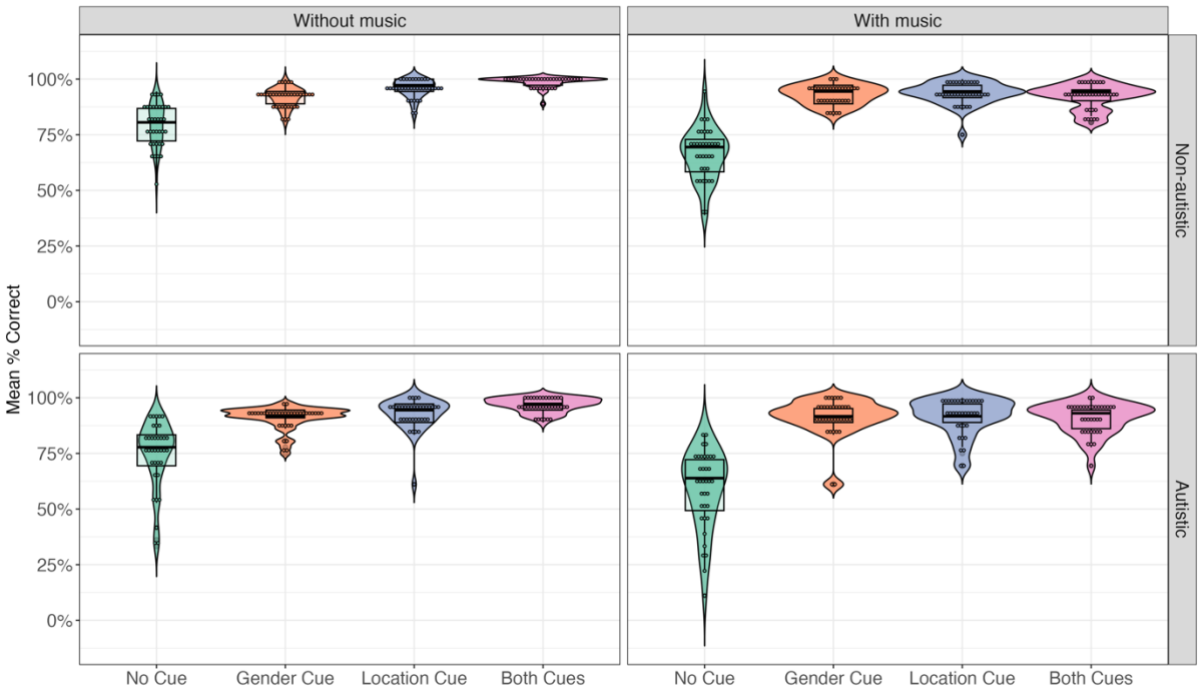

Figure 2-3. Mean accuracy rate across groups and conditions.

Table 2-2. Results of the GLMM for behavioural accuracy.

Fixed effects	***β***	**SE**	***z***	**OR**	**95% CI**	χ^2	***p***
(Intercept)	2.88	0.13	22.94	—	—	—	—
Group	−0.41	0.18	−2.26	0.66	[0.46, 0.95]	4.88	**.027**
Music	0.94	0.18	5.20	2.56	[1.80, 3.64]	25.78	**< .001**
Cue1	−2.56	0.20	−12.77	0.08	[0.05, 0.11]	135.31	**< .001**
Cue2	−0.75	0.22	−3.41	0.47	[0.31, 0.73]	11.44	**< .001**
Group × Cue1	0.05	0.16	0.34	1.05	[0.77, 1.44]	0.11	.737
Group × Cue2	0.27	0.19	1.42	1.31	[0.90, 1.90]	1.86	.172
Music × Cue1	−0.01	0.38	−0.04	0.99	[0.47, 2.09]	0.00	1.00
Music × Cue2	−1.48	0.44	−3.36	0.23	[0.10, 0.54]	10.99	**< .001**
Group × Music	−0.24	0.14	−1.81	0.78	[0.60, 1.02]	3.06	.080
Group × Music × Cue1	0.64	0.24	2.69	1.90	[1.19, 3.05]	6.87	**.008**
Group × Music × Cue2	0.79	0.38	2.09	2.20	[1.05, 4.62]	4.06	**.044**

Note. Significant *p*-values are presented in bold. OR = Odds ratio. Odds ratios are obtained by exponentiating the model's log-odds (β) coefficients. 95% confidence intervals (CIs) are similarly derived by exponentiating the CIs of the log-odds estimates.

Figure 2-4 presents the mean RTs across conditions and groups. The final model included fixed effects and by-item and by-participant random intercepts. Consistent with the accuracy results, a significant main effect of acoustic cues was observed. Longer RTs were required for accurate responses in the no-cue condition (M = 939.80, SD = 819.73) compared to the both-cues condition (M = 722.34, SD = 610.13) and the one-cue condition (gender-cue: M = 782.39, SD = 637.48; location-cue: M = 773.87, SD = 637.07). However, no significant main effects of music or group were found. Additionally, no significant interactions between these factors were observed (see Table 2-3).

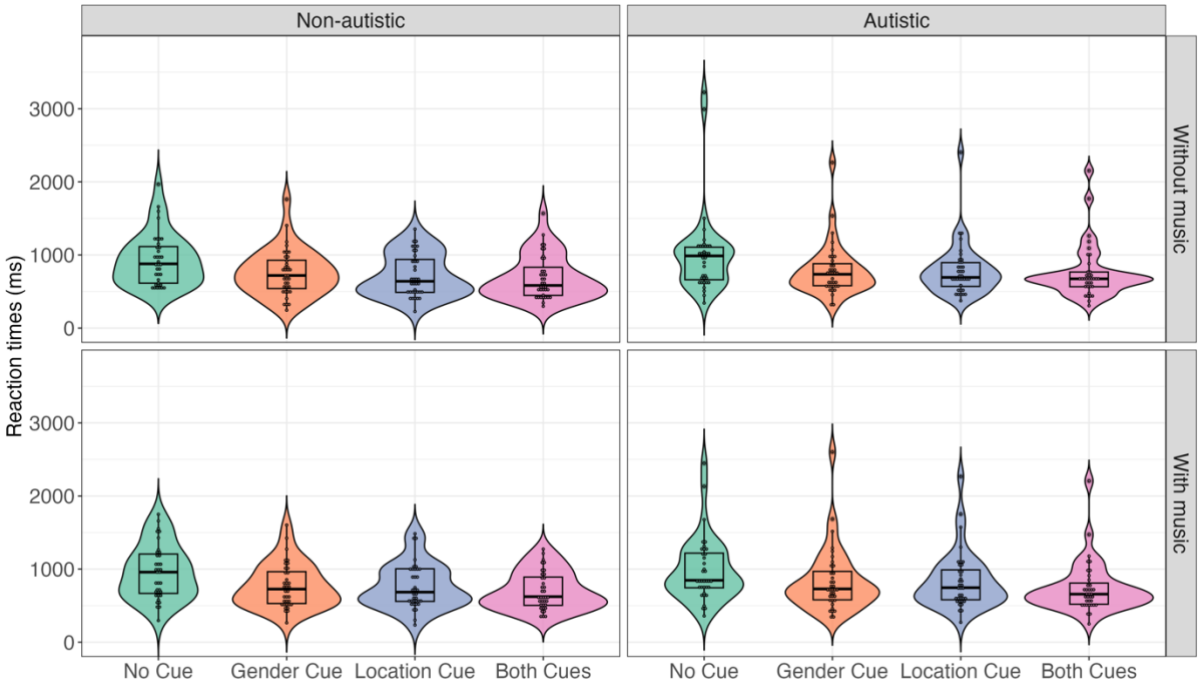

Figure 2-4. Mean RTs across groups and conditions.

Table 2-3. Results of the LMM for reaction times (RTs) of accurate responses.

Fixed effects	***β***	**SE**	**t**	**Exp(*β*)**	**95% CI**	**χ^2**	***p***
(Intercept)	6.41	0.05	126.76	—	—	—	—
Group	0.06	0.10	0.66	1.07	[0.88,1.30]	0.43	.512
Music	−0.03	0.02	−1.44	0.97	[0.92,1.01]	2.07	.150
Cue1	0.27	0.03	9.77	1.31	[1.24,1.38]	84.21	**< .001**
Cue2	0.10	0.03	3.48	1.10	[1.04,1.16]	11.88	**< .001**
Group × Cue1	0.00	0.03	0.17	1.00	[0.95,1.06]	0.03	.864
Group × Cue2	0.00	0.03	0.06	1.00	[0.95,1.05]	0.00	.950
Music × Cue1	0.00	0.05	0.01	1.00	[0.90,1.11]	0.00	1.00
Music × Cue2	−0.03	0.05	−0.60	0.97	[0.87,1.08]	0.36	.547

Group × Music	0.04	0.02	1.92	1.05	[1.00,1.09]	3.52	.061
Group × Music × Cue1	0.00	0.05	−0.02	1.00	[0.90,1.11]	0.00	.988
Group × Music × Cue2	−0.06	0.05	-1.15	0.94	[0.85,1.04]	1.31	.252

Note. Significant *p*-values are presented in bold. Exp(β) values are obtained by exponentiating the fixed-effect coefficients from the linear mixed-effects model predicting log-transformed response times. The resulting values reflect multiplicative effects on raw response times, where values greater than 1 indicate longer response times and values less than 1 indicate shorter response times relative to the reference level. The accompanying 95% confidence intervals (CIs) are derived by exponentiating the intervals for the log-scale estimates.

GAMMs

To investigate how speech-in-noise performance changed over time, we used GAMMs to model trial-level accuracy trajectories across cue conditions (no-cue vs. both-cues) and groups (autistic vs. non-autistic) for each SNR level. Each model included parametric effects for group and cue condition as well as smooth terms to capture time-varying trends within each group–condition combination. Participant-specific smooth terms were also included to account for individual variability (see Table 2-4). Figure 2-5 illustrates accuracy trends across trials (in bins of six) for each group and cue condition. As can be seen, both groups performed at ceiling in the both-cues condition with little change across trials. In contrast, in the no-cue condition, both groups showed improvements over time, particularly the non-autistic group, whose performance steadily increased across trials.

Parametric effects. In both models, the parametric coefficients revealed significant accuracy differences when comparing the baseline condition (both-cues in the autistic group) to the other group–condition combinations. Specifically, accuracy was significantly lower in the no-cue condition for both groups. No significant group differences were observed in the both-cues condition, indicating comparable accuracy.

Time-varying effects (smooth terms). Smooth terms of the models revealed significant non-linear changes in performance across trials, but only in the no-cue condition. Significant increase of accuracy was observed in both groups across both SNR levels, indicating improved performance over trials. In contrast, no significant trial effects were found in the both-cues condition for either group, reflecting stable ceiling-level performance from the beginning of the task.

Group and condition contrasts over time (difference plots). To visualise when and where group and condition differences emerged during the task, we examined pairwise comparisons using difference plots (Figure 2-6). These plots highlight time windows where significant contrasts appeared and are interpreted in light of the accuracy trends shown in Figure 2-5. In the no-cue condition, significant group differences emerged during the later trials (Figure 2-6, A2 and B2), with non-autistic participants outperforming autistic participants from trials 29-36 at −3 dB and from trials 14–30 at −9 dB. As shown in Figure 2-5, this difference reflects the fact that accuracy in the no-cue condition continued to increase for the non-autistic group, while the autistic group's performance remained more stable or variable. This widening gap suggests that the non-autistic group continued to improve with exposure, whereas the autistic group showed less consistent change.

Cue-related differences within each group were illustrated in Figure 2-6 in panels A3-A4 and B3–B4. For the non-autistic group (A3, B3), the difference between both-cues and no-cue conditions decreased over time, mirroring the upward trend in no-cue accuracy seen in Figure 2-5. This suggests improved performance over trials. For the autistic group (A4, B4), the size of the cue-related difference remained relatively stable, especially at −3 dB. Figure 2-5 supports this and shows that while both-cues accuracy stayed high throughout, performance in the no-cue condition fluctuated and showed less overall improvement.

In summary, both groups used the acoustic cues effectively when available. However, in the no-cue condition, only the non-autistic group showed steady gains over time. The autistic group also improved, but their performance was more variable, and they did not fully catch up in the later trials.

Table 2-4. Summary of GAMMs for accuracy by Group and Cue at each SNR level.

	−3 dB SNR				−9 dB SNR			
Parametric coefficients	***β***	**SE**	***z***	***p***	***β***	**SE**	***z***	***p***
(Intercept)	3.10	0.16	19.53	**< .001**	2.65	0.15	18.00	**< .001**
NAS.Both cues	0.36	0.24	1.48	.138	0.21	0.21	0.99	.322
AS.No cue	−2.32	0.15	−15.81	**< .001**	−2.00	0.13	−15.93	**< .001**
NAS.No cue	−2.08	0.19	−10.87	**< .001**	-1.65	0.19	−8.79	**< .001**
Smooth terms	**edf**	**Ref.df**	***χ²***	***P***	**edf**	**Ref.df**	***χ²***	***p***
AS.Both cues	2.06	2.57	4.48	.205	1.00	1.00	2.56	.110
NAS.Both cues	1.00	1.00	0.44	.506	1.00	1.00	0.50	.480
AS.No cue	2.09	2.61	9.78	**.020**	1.00	1.00	7.99	**.005**
NAS.No cue	1.00	1.00	12.66	**< .001**	1.00	1.00	10.85	**< .001**
Participants	47.68	646.00	153.93	**< .001**	68.49	646.00	256.92	**< .001**
	R^2(adjusted) = 0.174				R^2(adjusted) = 0.170			
	Deviance explained = 19.5%				Deviance explained = 18.2%			

Note. Significant p-values are presented in bold. Abbreviations: SE=standard error; AS=autistic group; NAS=non-autistic group, EDF=effective degrees of freedom. Formula: Accuracy ~ Group × Cue + s(Trial, by = Group × Cue, k = 8) + s(Trial, bs = "fs", m = 1).

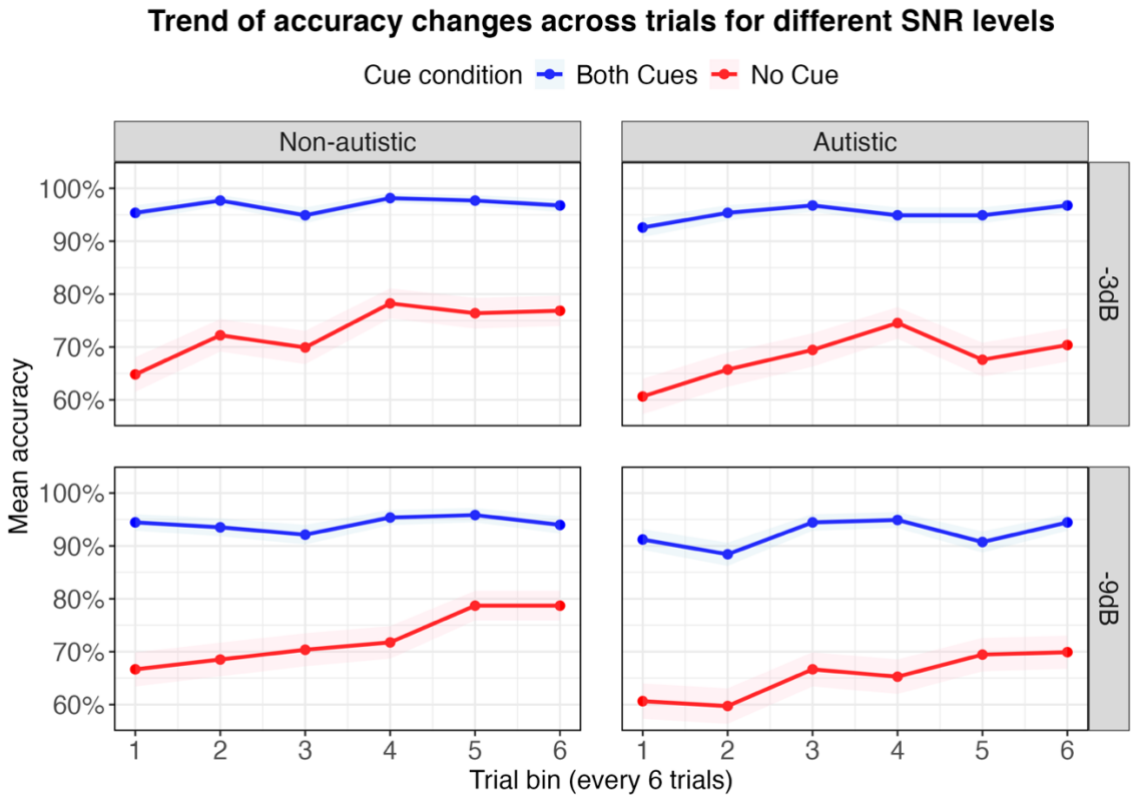

Figure 2-5. The trend of mean accuracy changes across trial bins (every 6 trials) for different SNR levels across group and condition. The shaded area indicating the 95% confidence interval.

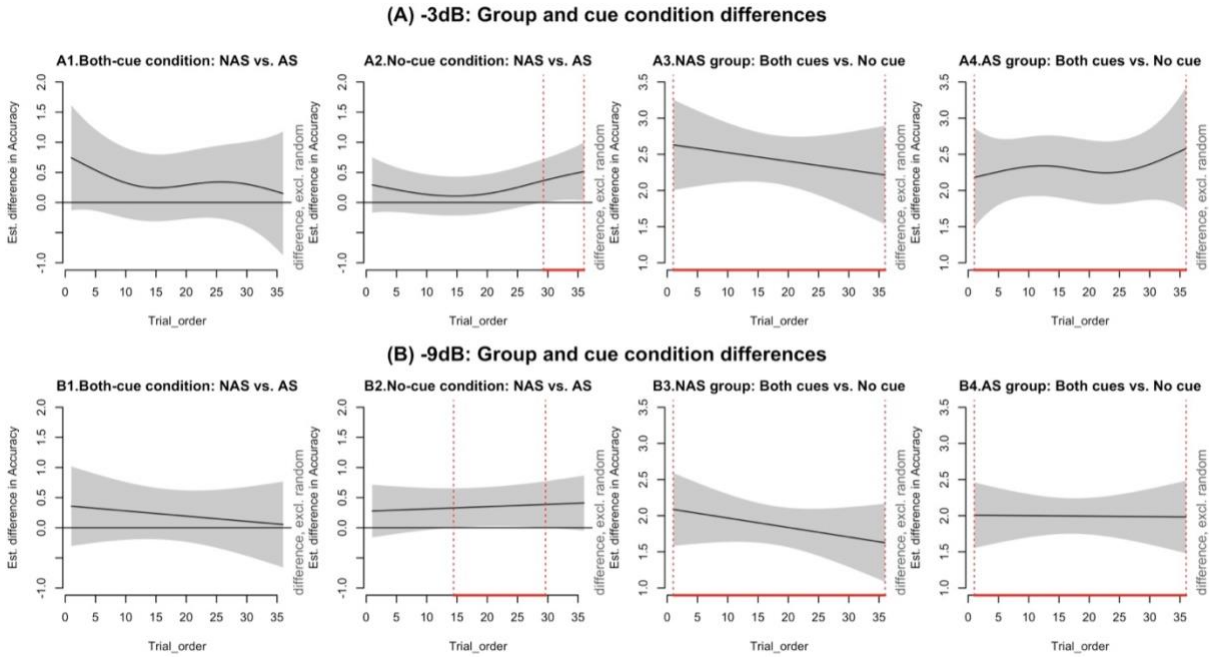

Figure 2-6. Estimated differences in accuracy over trials. The black line represents the estimated difference, with the grey shaded area indicating the 95% confidence interval. Red segments highlight trial ranges where the difference is statistically significant ($p < .05$).

Correlations

Figure 2-7 presents significant correlations for both groups. In the non-autistic group, lower pitch discrimination thresholds (indicating better pitch processing) were associated with higher mean accuracy and better performance in the no-cue condition. A follow-up analysis excluding an outlier (>3 SD from the mean) yielded consistent results (see Figure 2-8). In the autistic group, higher digit span scores were linked to better overall and no-cue accuracy. Additionally, participants who showed stronger local-to-global interference, indicating weaker global processing, also showed larger accuracy declines in the presence of background music. Appendix E reports a supplementary analysis in which false discovery rate (FDR) correction was applied across all pairwise correlations to ensure transparency with respect to multiple comparisons. Most of the reported associations remained significant after correction, with the exception of the relationship between pitch discrimination ability and mean accuracy in the non-autistic group. All correlations observed in the autistic group remained significant following correction. However, when the outlier in the pitch discrimination task was excluded, no pitch-related correlations in the non-autistic group survived FDR correction.

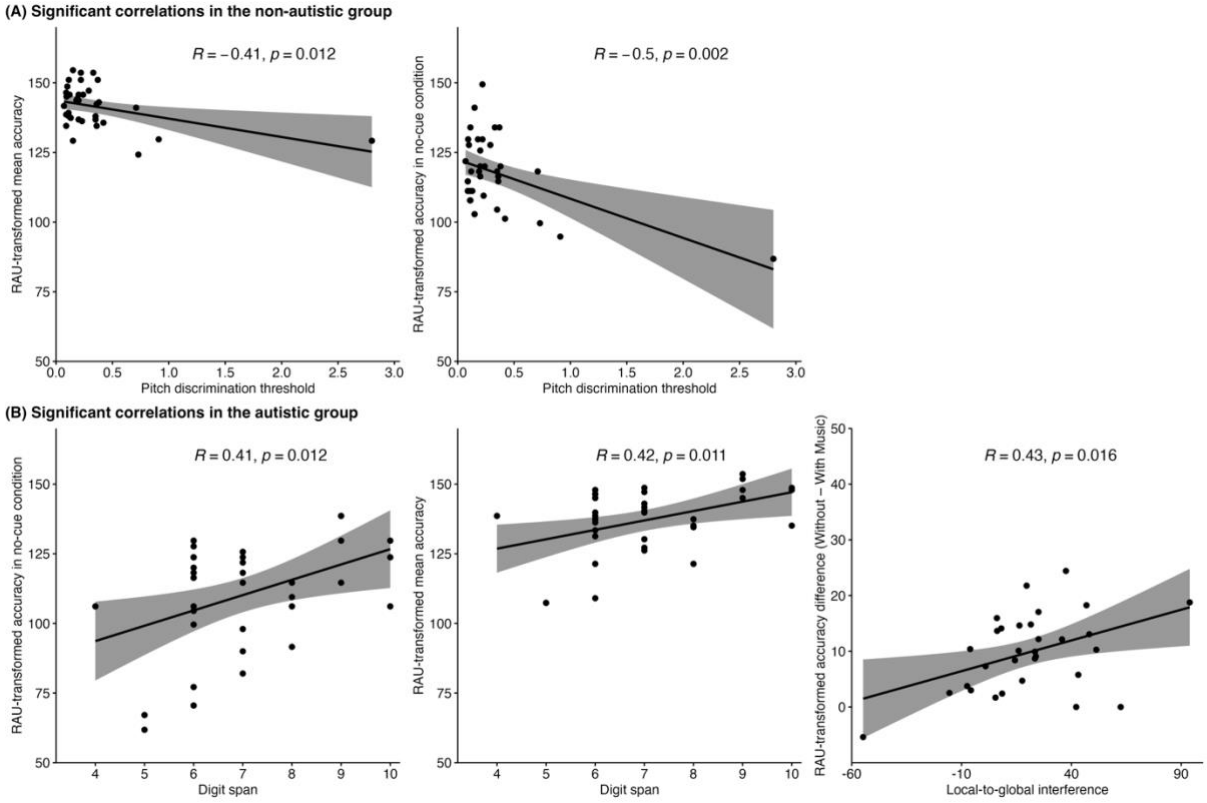

Figure 2-7. Significant correlations between accuracy and cognitive factors in the non-autistic group (A) and the autistic group (B). The grey shaded area indicates the 95% confidence interval around the mean. RAU = Rationalised Arcsine Units.

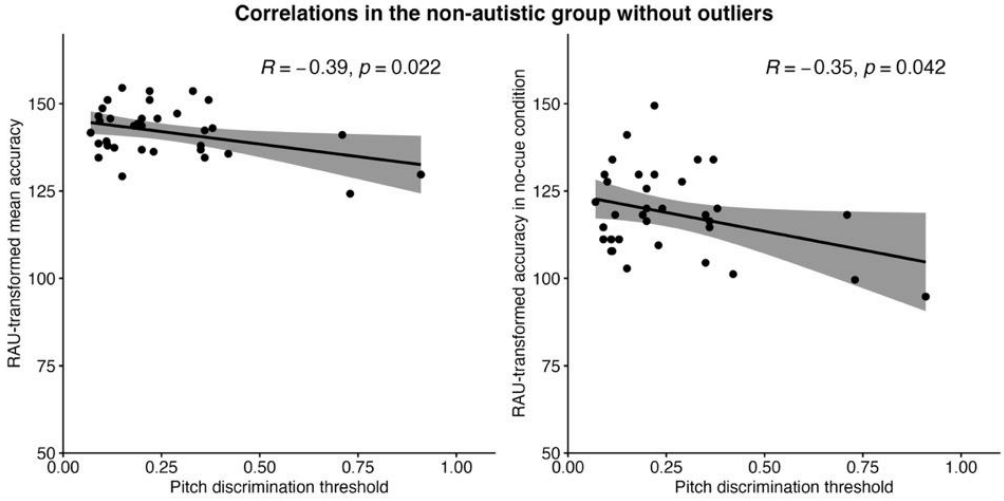

Figure 2-8. Correlations between pitch discrimination threshold and performance in the non-autistic group without the outlier. Accuracy values were transformed using the rationalised arcsine unit (RAU) transformation to normalise proportion data and reduce variance instability. The grey shaded area indicates the 95% confidence interval around the mean.

2.4 Discussion

This study examined speech processing in autistic and non-autistic adults in a competing-speaker scenario with background music, which required selective attention to the target speaker in a dynamic auditory environment. As expected, autistic participants exhibited lower accuracy than their non-autistic counterparts, reflecting greater challenges in recognising target speech in noisy environments.

2.4.1 Benefits of acoustic cues on mean accuracy

Both groups demonstrated higher accuracy and faster responses when at least one acoustic cue was present, highlighting the benefit of salient acoustic cues. This aligns with Emmons et al. (2022), who reported improved performance in autistic participants when both gender and location cues were available, compared to one-cue conditions. However, while they instructed participants to attend to specific acoustic features before each trial, our participants had to detect and use speaker-related cues independently based on the callsign within the speech stream. This closely mirrors real-life listening, where explicit instructions are rarely available. Also, our task involved identifying information from sentences, rather than recalling isolated words, further encouraging ongoing cue integration. Additionally, by incorporating a no-cue condition, our study extended the investigation to scenarios without salient cues, offering new insights into performance under more challenging conditions. Taken together, these features contribute to a more naturalistic assessment of speech-in-noise (SiN) processing and demonstrate that autistic listeners can benefit from speaker-related acoustic cues even in tasks that more closely approximate everyday communication demands.

2.4.2 Group differences in trial-level improvement

We expanded our analysis beyond mean accuracy to explore how performance changed over time across cue conditions using GAMMs. In the both-cues condition, both groups showed stable ceiling performance throughout the experiment. These findings suggest that autistic participants used acoustic cues as effectively as their non-autistic peers, indicating intact SiN processing in less demanding scenarios. However, in the more challenging no-cue condition, where two male speakers were collocated, both groups initially experienced processing difficulties but showed significant improvement over trials. Significant group differences emerged in later trials, with non-autistic participants achieving higher accuracy. Our results suggest that SiN performance is modulated by task complexity, and that autistic participants

may exhibit disproportionate difficulties as auditory scene complexity increased (Bendo et al., 2024).

Two factors may explain this group difference. The first concerns reduced implicit learning and sustained auditory attention in autism. Although no explicit instructions were given regarding the target speaker's identity or location, the same male speaker was consistently positioned at a fixed auditory location. Non-autistic participants may have implicitly detected this regularity (Reber, 1989), gradually becoming more familiar with the target voice and finding it easier to process over time (Holmes et al., 2021; Nygaard & Pisoni, 1998). In contrast, autistic participants may have struggled to form a stable auditory representation of the target speaker, potentially due to challenges with implicit learning (Lawson et al., 2018). This may relate to predictive coding accounts, which suggest reduced weighting of top-down predictions (Van de Cruys et al., 2014), potentially making it harder to develop expectations about the speaker's voice or location. Additionally, autistic individuals often exhibit atypical auditory attention, particularly under high-demand conditions (Čeponienė et al., 2003; Emmons et al., 2022; Keehn et al., 2013). Without salient cues to guide attention, they may have found it harder to consistently focus on the target stream. However, as implicit learning and attention were not directly measured, these interpretations require further investigation.

The second factor concerns group differences in processing strategies. Despite the lack of salient cues, non-autistic participants may have used vocal differences between speakers to segregate speech. Supporting this, better pitch discrimination was associated with higher accuracy in the non-autistic group, but not in the autistic group. While both groups showed similar non-vocal pitch discrimination ability, only non-autistic participants appeared to use this perceptual skill to support task performance. This indicates that autistic participants may have detected pitch differences but not spontaneously used them to guide stream segregation, possibly reflecting atypical top-down processing, as proposed by the predictive coding theory (Van de Cruys et al., 2014). However, It is also important to consider that although non-vocal pitch discrimination was comparable between groups, we did not assess vocal pitch perception, which tends to differ in autistic individuals and has been associated with variations in their SiN processing (Schelinski & Von Kriegstein, 2020). Autistic individuals may experience greater difficulty prioritising socially relevant acoustic cues, particularly when these cues are subtle or ambiguous (Hernandez et al., 2020; Schwartz et al., 2020).

Instead, autistic participants appeared to rely more on working memory to manage the increasing demands of the no-cue condition. Significant positive correlations between working memory and accuracy suggest that they engaged higher-level cognitive resources as a complementary strategy during SiN recognition. This aligns with research highlighting the role of working memory in mitigating SiN difficulties (Dryden et al., 2017). Such reliance may reflect broader differences in auditory processing. According to the Weak Central Coherence theory (WCC, Happé & Frith, 2006), autistic individuals may show a detail-focused processing bias, with a reduced tendency to prioritise global or contextual integration. A bias towards local acoustic detail may alter the balance between perceptual integration and controlled processing, increasing reliance on working memory when subtle cues (e.g., pitch) must be evaluated across time. The additional load of maintaining task-relevant information in memory may also have contributed to the group differences in performance (Lau et al., 2023).

2.4.3 The effect of background music

This study is the first to examine the effect of music on speech recognition in autism. While music reduced accuracy in both groups overall, a significant group difference emerged in the both-cues condition: the presence of music reduced accuracy in the non-autistic group, but not in the autistic group. With both cues available, non-autistic participants may have relied on automatic processing, which requires minimal attention effort (Schneider & Shiffrin, 1977). While efficient, such processing is more susceptible to unexpected distractions like background music. Thus, the decline in accuracy may not reflect the music's inherent distractibility, but the heightened sensitivity of ceiling-level performance to even subtle increases in task demands. In contrast, autistic participants showed no reduction in accuracy, which may suggest less reliance on automatic processing even when both cues were available. Instead, they may have sustained a more effortful, controlled focus on the speech signals, which made their performance less influenced by the presence of music (Xu et al., 2024). These results challenge our initial hypothesis that autistic participants would be more vulnerable to background music due to their heightened interest in music over speech. The structured and instrumental nature of the music used in this study may have lacked the personal or social relevance needed to elicit heightened distraction (Kiss & Linnell, 2021; Nadon et al., 2021).

Correlation analyses revealed that cognitive processing styles influenced autistic participants' susceptibility to background music. Autistic participants with stronger local biases exhibited greater performance declines in the presence of music. This was measured using the local-to-

global interference score, which reflects difficulty in focusing on global patterns when conflicting local details are present. These results provide support for the WCC theory (Happé & Frith, 2006). During SiN processing, a local bias may hinder the ability to group auditory elements into meaningful streams, making it more difficult to separate target speech from background music. As a result, background music may be processed as a distracting competing source, leading to greater interference and reduced performance. ERP studies support this interpretation, showing that while autistic individuals process individual acoustic elements accurately, they exhibit reduced neural responses when required to integrate multiple sound streams (e.g., Lepistö et al., 2009). This effect may have been especially pronounced in the current study due to the use of a 0 dB SNR, which increased listening demands and likely intensified the impact of local processing bias.

2.4.4 Limitations and directions for future research

One limitation of the current study relates to the nature of the stimuli, which may have reduced task demands and masked group differences. The use of predictable speech and emotionally neutral music likely made the task less challenging by minimising semantic and emotional interference. Future studies should use more naturalistic speech and vary music features (e.g., emotional tone, genre, lyrics) to better capture group differences under more realistic and cognitively demanding conditions (Brown & Bidelman, 2022; Russo & Pichora-Fuller, 2009; Shi & Law, 2010).

Additionally, our conclusions are based solely on behavioural measures. Emerging evidence suggests that, despite similar accuracy, autistic individuals may show increased listening effort, reflected in greater pupil dilation (Xu et al., 2024) and reduced MEG responses (Fadeev et al., 2024). Future research should incorporate neurophysiological measures to provide a more comprehensive understanding of SiN processing in autism.

Finally, our sample consisted of verbally and cognitively able adults, which helped control for confounds but limited the generalisability to broader autistic populations. Moreover, although the sample size was based on a power analysis, the relatively small pilot sample used for estimation may have reduced the precision of those calculations. Larger and more diverse samples are needed to explore potential subgroup differences within the autism spectrum.

Despite these limitations, our findings suggest that autistic individuals can achieve comparable speech recognition performance when listening conditions are structured and low in distraction. This points to the potential value of technologies such as remote microphone systems (Schafer et al., 2014) or sound-field amplification (Wilson et al., 2021), which enhance the salience of target speech when background noise cannot be fully controlled. Moreover, the observed improvement in autistic participants' performance over trials suggests that they may benefit from structured training. Since cue-based training has been shown to improve SiN perception in non-autistic individuals (Gohari et al., 2023), adapting similar interventions for autistic populations could enhance their ability to navigate multi-talker environments.

2.5 Conclusion

This study highlights the role of acoustic cues and background music in SiN processing in autism. While autistic listeners faced general difficulties, they effectively used acoustic cues to support speech recognition and showed improvement with repeated exposure in the absence of cues, though their progress was slower than that of non-autistic participants. Additionally, individual differences in sensitivity to background music highlight the heterogeneity of cognitive processing styles in autism, reinforcing the need for personalised support strategies.

Chapter 3

Study 2: Semantic Processing in Autism during Speech-in-Music Listening: Insights from Congruency and Surprisal-Based N400 Analyses

Abstract

Understanding speech in background music is a common real-world challenge, particularly when vocals compete for linguistic processing resources. This study examined how the presence and intelligibility of sung lyrics influence semantic processing in autistic and non-autistic adults. Twenty-nine participants per group performed a sentence acceptability judgement task while EEG was recorded. Sentences ended with either semantically congruent or incongruent words and were presented alongside instrumental, Simlish (phonologically English like but unintelligible), or English-lyric versions of the same songs. We analysed N400 responses using two complementary approaches: a categorical congruency contrast, indexing the neural cost of processing semantic anomalies, and a continuous lexical surprisal measure, capturing graded sensitivity to word predictability. In non-autistic participants, both analyses showed the largest N400 responses in the instrumental condition, attenuated responses in vocal conditions, and reduced behavioural accuracy as lyrics became more intelligible. Autistic participants showed lower overall accuracy, attenuated N400 effects in both analyses, and minimal variation across masking conditions, indicating inefficient semantic processing and reduced adaptability to changing listening demands. By combining ecologically valid speech-in-music masking with dual analytic approaches, this study provides the first neurophysiological evidence of these semantic processing differences in autism and demonstrates how integrating categorical and probabilistic measures can yield a richer and more nuanced account of speech processing in complex auditory environments.

Keywords: autism, background music, ERPs, N400, speech-in-noise

3.1 Introduction

Communication often takes place in acoustically complex environments where listeners must attend to a target speaker while ignoring competing sounds, a challenge known as the “cocktail party problem” (Cherry, 1953). Because all sound sources are blended into a single acoustic

waveform at the ears, segregating the target speech stream requires active perceptual and cognitive processing (Shinn-Cunningham & Best, 2008). Interference affecting speech intelligibility is generally grouped into two main categories. Energetic masking arises from spectral and temporal overlap between the target and masker, degrading the signal at the auditory periphery (Cooke, 2006; Festen & Plomp, 1990). In contrast, informational masking reflects higher-level interference, including difficulties in stream segregation, control of attention, or resolution of lexical and semantic competition (Brungart, 2001; Kidd et al., 2008). Such interference can arise from acoustic similarities, such as when the masker has speech like pitch, rhythm, or formant patterns, making it harder to distinguish from the target voice. It can also be linguistic, occurring when the masker activates competing lexical or semantic representations in the listener's mental lexicon (Summers & Roberts, 2020). Although acoustic and linguistic informational masking differ in their locus, they frequently co-occur in real-world listening. Successful comprehension under these conditions requires listeners to allocate resources flexibly between bottom-up perceptual analysis and top-down cognitive mechanisms such as selective attention, prediction, and semantic integration (Assmann & Summerfield, 2004; Mattys et al., 2012; Sohoglu et al., 2012). Evidence suggests that the relative contribution of these processes varies with the severity of masking: moderate masking increases reliance on top-down support, whereas severe masking can disrupt lexical access to such an extent that predictive mechanisms offer little benefit (Mattys et al., 2009, 2012).

Understanding how these processes unfold in real time can be approached through neurophysiological measures such as event-related potentials (ERPs). Among these, the N400 component is particularly relevant because it is closely linked to semantic processing and is sensitive to the interplay between bottom-up and top-down mechanisms during comprehension (Broderick et al., 2019). The N400 is a negative-going waveform that peaks around 400 ms and is typically largest over centroparietal scalp sites. It is widely used as an index of semantic processing, although its precise functional role remains debated (Kutas & Federmeier, 2011; Lau et al., 2008). Early accounts associated larger (more negative) N400 amplitudes with greater semantic integration difficulty, meaning increased effort to unify an incoming word with its preceding context (Hagoort, 2008; Kutas & Hillyard, 1980). This view is supported by semantic anomaly paradigms, in which incongruent sentence completions (e.g., *He spread the warm bread with socks*) elicit larger N400s than congruent ones (e.g., *He spread the warm bread with butter*). A second account, grounded in semantic priming paradigms, links the N400 to the ease of lexical-semantic access: predictable or related words (e.g., *doctor–nurse*) are

accessed more efficiently and evoke smaller amplitudes than unrelated ones (Federmeier, 2007; Lau et al., 2008). More recent predictive-processing models integrate these perspectives, proposing that the N400 reflects the degree of mismatch between context-based semantic predictions and the incoming input (Kuperberg, 2016). In this view, both semantically anomalous completions and highly unexpected words produce larger prediction errors, which manifest as greater processing effort and consequently larger N400 amplitudes. Semantic integration difficulty is thus seen as the downstream consequence of encountering unexpected information, while lexical access effects arise because related primes increase the predicted probability of certain semantic features, reducing the mismatch when the target appears. This predictive framework accommodates findings from both anomaly- and priming-based paradigms, providing a common basis for interpreting N400 effects in terms of listeners' sensitivity to semantic prediction.

Studies using the N400 to investigate speech perception in noise have yielded mixed findings that map onto the interplay between bottom-up and top-down processing. Under moderate masking when the signal remains largely intelligible, listeners appear to rely more on contextual cues, which is reflected in larger N400 amplitudes (Devaraju et al., 2021; Kemp et al., 2019; Romei et al., 2011; Song et al., 2020; Zheng et al., 2023). Under severe masking, however, phonological and lexical information can be degraded to the extent that contextual predictions are harder to deploy, leading to reduced or delayed N400 responses (Calma-Roddin & Drury, 2020; Hsin et al., 2023; Silcox & Payne, 2021). This mirrors findings from degraded-speech research, where manipulations such as noise-vocoding or time compression reduce intelligibility, weaken lexical activation, and attenuate N400 effects (Aydelott et al., 2006; Obleser & Kotz, 2011; Strauß et al., 2013). Jamison et al. (2016) reported non-linear effects of signal-to-noise ratio on the N400 in a passive listening paradigm, underscoring the complexity of acoustic degradation's impact on semantic processing. Such variability across studies likely reflects differences in signal-to-noise ratio, the type of semantic manipulation, and the characteristics of the masker, underscoring the importance of carefully controlling these factors when interpreting N400 effects in noisy listening. A major limitation of this literature, however, is its heavy reliance on unintelligible speech-shaped noise. Few studies have examined informational maskers such as competing speech or background music, which are more ecologically valid and cognitively demanding (Brown & Bidelman, 2022a; Calma-Roddin & Drury, 2020; Song et al., 2020), and thus better reflect speech-in-noise processing difficulties in real life (Bendo et al., 2024). Another gap is the limited attention to individual differences.

Few studies systematically consider factors such as musical expertise or attentional control (Jamison et al., 2016; Zheng et al., 2023), and even fewer examine neurodivergent populations.

This is particularly important in the context of autism spectrum disorder, a neurodevelopmental condition characterised by differences in social communication and sensory processing (American Psychiatric Association, 2013). Autistic individuals often experience greater difficulty understanding speech in noisy environments, especially under informational masking (Emmons et al., 2022; Lau et al., 2022; Li et al., 2025; Ruiz Callejo et al., 2023). These challenges have been linked to atypical auditory temporal processing and disruptions in top-down mechanisms, including reduced capacity to segregate concurrent streams, diminished context-based modulation of auditory input and inflexible cognitive resource allocation (DePape et al., 2012; Lepistö et al., 2009; Mamashli et al., 2017; Russo et al., 2009). Despite growing evidence for such auditory processing differences, N400 responses in autistic individuals during speech-in-noise remain largely unexplored. In our recent work, we compared autistic and non-autistic adults listening to sentences in quiet, babble noise, and competing speech, and found that the autistic group showed delayed and reduced N400 responses, indicating less efficient semantic processing under noisy conditions (Li et al., 2025). This result aligns with prior reports of atypical N400 responses in autism even without background noise (Braeutigam et al., 2008; Manfredi et al., 2020; Pijnacker et al., 2010; Ribeiro et al., 2013; Ring et al., 2007) and supports the Weak Central Coherence (WCC) account, which suggests a reduced tendency to form globally coherent representations in autism (Frith, 1989; Frith & Happé, 1994; Happé & Frith, 2006). Another framework relevant to these findings is the predictive coding theory, which views perception as the brain's ongoing effort to minimise prediction error by integrating sensory input with top-down expectations, weighted by their estimated precision (H. Feldman & Friston, 2010; Friston, 2009). In autism, atypical precision weighting may reduce the influence of prior expectations, leading to less consistent use of contextual information such as semantic predictability (Lawson et al., 2018; Van Boxtel & Lu, 2013; Van De Cruys et al., 2014). While the WCC focuses on a bias towards local over global processing outcomes, the predictive coding theory offers a mechanistic explanation for how this may arise, attributing it to altered integration of predictions with incoming sensory evidence. Together, these accounts suggest that autistic individuals may be less able to exploit reliable semantic context to guide comprehension in challenging listening environments.

One ecologically relevant source of informational masking that remains understudied especially in autistic populations is music, particularly vocal music. Because speech and music share overlapping acoustic and cognitive resources, background music can interfere with speech processing through competition for attentional and linguistic resources (Koelsch et al., 2005; Maess et al., 2001; Patel, 1998; Shi & Law, 2010). This interference is even stronger when the music contains intelligible lyrics, as they add a competing speech stream and introduce semantic competition (Brown & Bidelman, 2023; Crawford & Strapp, 1994; Russo & Pichora-Fuller, 2008; Scharenborg & Larson, 2018). However, isolating the role of intelligibility is challenging because manipulations often alter other acoustic properties. For example, speech-in-speech studies comparing familiar and unfamiliar languages (e.g., English vs. Mandarin) confound intelligibility with differences in rhythm and prosody (Van Engen, 2010). To address this, Brouwer et al. (2022) used Simlish, an artificial language created for *The Sims* video game. Simlish retains many of the acoustic characteristics of natural speech—such as syllable structure, phonotactics, and prosody—but does not correspond to any real language, making it linguistically unintelligible to listeners. The study focused on two targeted contrasts within the same musical material: (1) instrumental vs. vocal music, to test whether the presence of lyrics increases masking relative to instrumental versions; and (2) English vs. Simlish lyrics, to isolate the effect of intelligibility while controlling for other vocal-acoustic properties. These conditions represent different types and degrees of informational masking: instrumental music provides only acoustic masking; Simlish lyrics combine acoustic masking with non-meaningful linguistic masking; and English lyrics add acoustic masking plus meaningful linguistic masking. Participants listened to sentences with background music in these three forms. Recall accuracy was highest with instrumental music, lower with Simlish, and lowest with English lyrics. These findings suggest that vocal music disrupts speech processing more than instrumental music, and that intelligible lyrics introduce additional semantic interference, mirroring effects observed in speech-in-speech studies (Brungart, 2001; Simpson & Cooke, 2005).

Building on Brouwer et al. (2022), we examined how characteristics of background vocals, particularly their presence and intelligibility, influence speech processing in autistic and non-autistic adults. We adapted the paradigm by replacing the recall task with a sentence-acceptability judgement while recording EEG, allowing us to measure N400 responses as participants evaluated each sentence. Background music from the same songs was presented in the three masking conditions described above. The current study mainly addressed two

questions: (1) How does the presence and intelligibility of background vocals affect behavioural accuracy and N400 responses? (2) Do autistic and non-autistic adults differ in the extent to which their accuracy and N400 responses adapt to changes in masking condition?

Based on Brouwer et al. (2022), we predicted that both groups would show the highest accuracy with instrumental music, lower accuracy with Simlish lyrics, and the lowest accuracy with English lyrics. For the N400 responses, we considered two competing hypotheses. If more distracting conditions impose greater demands on semantic integration, they should elicit larger N400s (Devaraju et al., 2021; Kemp et al., 2019; Romei et al., 2011). Conversely, if increased distraction impairs phonological and lexical access, the result should be reduced or delayed N400s, reflecting less efficient semantic processing (Calma-Roddin & Drury, 2020; Silcox & Payne, 2021). Regarding group differences, we anticipated that autistic participants would show lower accuracy and attenuated or delayed N400s overall, with smaller condition-related changes, consistent with reports of reduced adaptability to noise conditions with varying listening demands (Li et al., 2025; Mamashli et al., 2017).

We examined the N400 using two complementary approaches. The congruency-based analysis compared ERPs elicited by semantically congruent versus incongruent sentence endings, a well-established method for probing categorical prediction violations. However, recent work has suggested that such contrasts capture only the most extreme mismatches and may overlook more graded variations in contextual predictability (Weissbart et al., 2020). To address this, we also applied a surprisal-based analysis, which uses a probabilistic language model to assign a continuous value reflecting the unexpectedness of the sentence-final word given its preceding context (Michaelov et al., 2023). This approach models the full distributional structure of contextual expectations, offering potentially greater sensitivity and a more mechanistic index of predictive processing. By incorporating both methods, we assess whether the effects of masking and group differences are consistent across categorical and continuous measures, and evaluate the robustness of N400 mechanisms under more complex listening conditions involving background music. Consistent with predictive-processing accounts (Kuperberg, 2016), we interpret larger N400 amplitudes in both analyses as reflecting greater prediction error when the incoming word diverges from context-based expectations.

3.2 Methods

Participants

29 autistic and 29 non-autistic participants (aged 17–47) were recruited. All were right-handed, native English speakers with normal or corrected-to-normal vision and passed a hearing screening at 25 dB HL for 0.5, 1, 2, and 4 kHz using an Amplivox manual audiometer. None reported speech, language, learning, neurological, or psychiatric conditions. Autism diagnoses were confirmed by clinicians with supporting documentation. Non-autistic participants had no personal or family history of autism and were screened using the Autism Spectrum Quotient (Baron-Cohen et al., 2001).

To account for cognitive factors influencing speech-in-music processing (Gordon-Salant & Cole, 2016; Heinrich, 2021), participants completed background assessments of nonverbal IQ (Raven's Progressive Matrices; Raven & Court, 1998), receptive vocabulary (Receptive One-Word Picture Vocabulary Test IV; Martin & Brownell, 2011), and verbal short-term memory (digit span; Wechsler et al., 2003). Given the potential impact of music familiarity, preference, and musical training on speech-in-music perception (Bidelman & Yoo, 2020; Brown & Bidelman, 2023; Russo & Pichora-Fuller, 2008; Strait & Kraus, 2011; Thompson et al., 2001), these factors were measured and matched across groups. Musical training was assessed using a questionnaire (Pfordresher & Halpern, 2013) that summed the total years of formal instruction across all instruments, including voice. Familiarity and liking were rated post-experiment on a 0–4 scale. Participants also completed a familiarity check by identifying the singer and song title. These responses were scored and combined with the ratings to yield a final familiarity score (0–6). Demographic and cognitive data are summarised in Table 3-1. Wilcoxon rank-sum tests revealed no significant group differences, except for higher AQ scores in the autistic group.

The study was approved by the University Research Ethics Committee. All participants provided written informed consent and received financial compensation and travel reimbursement. Psychology students recruited via the participant pool received course credit for their participation.

Table 3-1. Characteristics of the autistic (*n* = 29) and non-autistic (*n* = 29) groups.

Variables	Autistic *Mean (SD)*	Non-autistic *Mean (SD)*	*W*	*p*	Rank-biserial Correlation
Gender (Female:Male)	20:9	23:6			
Age	26.02 (8.09)	26.33 (7.59)	398.0	.73	−0.05
Nonverbal reasoning (RSPM raw score)	53.86(3.70)	54.45(3.57)	387.5	.61	−0.08
Nonverbal reasoning (RSPM percentile)	51.21(25.48)	52.55(28.50)	415.5	.94	−0.01
Receptive vocabulary (ROWPVT-4 raw score)	167.35(10.43)	170.24(8.06)	363.5	.38	−0.14
Receptive vocabulary (ROWPVT-4 standard score)	109.45(16.12)	113.21(13.67)	358.5	.34	−0.15
Digit span	7.10 (1.66)	7.07 (0.84)	407.0	.83	−0.03
Musical training	4.12 (5.76)	6.38 (6.67)	320.5	.11	−0.24
Music familiarity score	2.72 (0.59)	2.83 (0.38)	339.0	.19	−0.19
Music liking score	1.59 (0.98)	2.14 (0.83)	355.0	.30	−0.16
Autistic traits (AQ)	38.38 (6.58)	16.52 (7.54)	822.5	< .01	0.96

Stimuli and apparatus

The target stimuli consisted of 180 sentence pairs with highly constraining contexts (Stringer & Iverson, 2020). Each sentence contained 5–10 words and was recorded by a female native speaker of Standard Southern British English. The final word was either semantically congruent with the preceding context (e.g., *Children like pasta with tomato sauce*) or incongruent (e.g., *Children like pasta with tomato noon*). Congruent and incongruent final words were matched for word frequency, length (syllables, phonemes), phonological neighbourhood density, and stress pattern to ensure that differences in processing reflected semantic rather than lexical or phonological factors. Semantically incongruent sentences were expected to elicit larger N400 amplitudes than congruent sentences.

Two pop songs by Katy Perry, *Last Friday Night (T.G.I.F.)* and *Hot N Cold*, were presented in three versions: (1) a karaoke version (instrumental, non-vocal), (2) the original version with English lyrics, and (3) a version with Simlish lyrics. All versions were sourced from YouTube, normalised in Praat (Boersma, 2007), and RMS-equalised to control for energetic masking. To

ensure consistency across conditions, the English and Simlish versions began at the onset of Katy Perry's vocals, guaranteeing the presence of a voice. The instrumental version was aligned to start at the same time point as the other two. To preserve expectancy effects, background music was played continuously throughout each block, with each sentence paired with a unique music segment. Unlike Brouwer et al. (2022), who used a signal-to-noise ratio (SNR) of −15 dB, the present study employed an SNR of −6 dB (target at 60 dB SPL, background at 66 dB SPL). This adjustment is based on pilot testing (see Appendix A) to ensure that participants could reliably perceive and comprehend the target speech while avoiding ceiling effects.

The experiment was conducted in a soundproof booth using E-Prime 3.0. Stimuli were presented binaurally via Etymotic ER-1 earphones, delivered through a sound card. Participants were instructed to judge sentence acceptability while ignoring the background music. To ensure task comprehension, they completed three practice trials per condition prior to the main experiment. Each trial began with an audio file and a fixation cross, followed by a randomly jittered silent interval of 1.5–1.7 seconds. Participants judged sentence acceptability by pressing C (acceptable) or M (unacceptable). The experiment included six blocks (two per condition), each containing 60 trials and lasting 5–6 minutes. Congruent and incongruent sentences were randomly intermixed, and block order was randomised. To minimise context effects, condition assignments were rotated across three experimental lists, with each participant randomly assigned one. Self-paced breaks were provided between blocks to reduce fatigue.

EEG recording and pre-processing

Electroencephalography (EEG) data were recorded using a BioSemi ActiveTwo system with 64 Ag/AgCl scalp electrodes and six external electrodes (placed at the left and right mastoids, and at sites for vertical and horizontal electrooculography). Signals were sampled at 2048 Hz with the Common Mode Sense (CMS) and Driven Right Leg (DRL) electrodes serving as reference and ground, respectively. Electrode offsets were monitored in ActiView and kept below ±30 μV. Event triggers marking the onset of target words were generated in E-Prime and transmitted via the parallel port to ensure precise synchronisation with the EEG signal.

Pre-processing was conducted in EEGLAB (Delorme & Makeig, 2004) within Matlab R2022b (The MathWorks Inc, 2022). Continuous EEG data were bandpass filtered with a high-pass

filter at 0.1 Hz and a low-pass filter at 40 Hz using a zero-phase Butterworth filter. The data were then downsampled to 256 Hz and re-referenced to the average of the mastoid electrodes. Epochs were extracted from −200 ms to 800 ms relative to the onset of the target word and baseline-corrected using the pre-stimulus interval (−200 to 0 ms). Bad channels were manually identified through visual inspection and interpolated. Independent component analysis using the Infomax algorithm was applied to identify and remove components associated with ocular and muscle artefacts. Finally, trials with voltage fluctuations exceeding ±150 μV were excluded from further analysis.

EEG data analysis

To examine the neural correlates of semantic processing under masking, we used two complementary approaches: a categorical contrast of semantic congruency and a continuous measure of lexical surprisal derived from a probabilistic language model. These analyses capture different dimensions of context-based processing, enabling a more comprehensive account of N400 modulation.

Semantic congruency analysis. To identify the spatiotemporal distribution of the N400 congruency effect, we conducted cluster-based permutation testing using the FieldTrip toolbox (Oostenveld et al., 2011), following the method by Maris & Oostenveld (2007). Paired-samples t-tests were computed at each electrode and time point within the pre-specified 200–600 ms window to compare congruent and incongruent conditions. Samples with $p < .05$ were grouped into clusters based on spatial and temporal adjacency. For each cluster, the sum of t-values was used as the test statistic. A null distribution was generated using 1,000 random permutations of condition labels, and the maximum cluster statistic from each permutation was retained. Clusters were deemed significant if their statistic exceeded the 97.5th percentile or fell below the 2.5th percentile of the null distribution (two-tailed $\alpha = .05$).

Then, we used linear mixed-effects models (LMMs) to examine N400 amplitude differences between congruent and incongruent conditions. As N400 elicited by semantic incongruity is reported to be maximum over the centro-parietal region by prior studies (Kutas & Hillyard, 1984), we defined a centro-parietal region of interest (ROI) comprising electrodes C3, Cz, C4, CP3, CPz, CP4, P3, Pz, and P4, following the identical electrode selection reported by Koelsch et al. (2004). Within this ROI, we applied linear mixed-effects models to assess the effect of semantic congruency on N400 amplitude across group and condition. This congruency contrast

captures the neural cost of processing unexpected semantic information, operationalised as the amplitude difference between incongruent and congruent sentence completions.

Lexical surprisal analysis. To complement the traditional congruency analysis and capture more fine-grained variation in semantic predictability, we conducted a separate analysis using lexical surprisal as a continuous predictor. Surprisal was operationalised as the negative logarithm of a word's conditional probability given its preceding context (−log P(word|context)), following approaches that link word predictability to processing effort (Frank et al., 2015; Smith & Levy, 2013). Within a predictive-processing framework, higher surprisal values reflect greater prediction error, with more unexpected words eliciting larger (more negative) N400 amplitudes. Surprisal values were estimated using the GPT-2 language model. Each sentence was tokenised, and for each final word, the model generated a conditional probability distribution based on its preceding context. The surprisal value was extracted from the probability assigned to the final word. Although more recent studies have used larger models such as GPT−3 (e.g., Michaelov et al., 2024), we used GPT-2 due to its open accessibility, computational efficiency, and demonstrated ability to predict N400 amplitude variation reliably (Michaelov et al., 2023). As in the congruency analysis, N400 amplitude was modelled within the same centro-parietal ROI using LMMs, with surprisal included as a continuous fixed effect. This analysis enabled us to assess whether N400 amplitude varied systematically with lexical predictability, complementing the categorical contrast with a continuous measure of semantic expectation.

Statistical analysis

Analyses were conducted in R (version 4.1.2; Posit Team, 2022) using the lme4 package (Bates et al., 2015). Linear mixed-effects models (LMMs) were fitted to N400 amplitudes, and generalised linear mixed-effects models (GLMMs) were constructed for behavioural accuracy (binary outcome), with the BOBYQA optimiser applied to improve convergence. Two contrast-coded predictors captured condition effects: (1) non-vocal vs. vocal music (contrast coding: English = 1/3, Simlish = 1/3, non-vocal = −2/3), and (2) Simlish vs. English lyrics (English = 1/2, Simlish = −1/2, non-vocal = 0).

In the behavioural accuracy model, fixed effects included group (coded as 1/2 for autistic, −1/2 for non-autistic), the two condition contrasts, and their interactions. Two models were fitted for the N400: a congruency-based model, which included congruency (coded as 1/2 for

congruent, −1/2 for incongruent) and its interactions with group and condition; and a surprisal-based model, which included surprisal as a mean-centred continuous predictor, with interactions structured identically to those in the congruency model.

Model selection followed the recommendations of Barr (2013). Initial models were fitted with a maximal random-effects structure, including random intercepts and slopes for within-unit predictors for both participants and items, where possible. When maximal models failed to converge, the random-effects structure was simplified in a stepwise manner: (1) by removing correlations between random effects, and (2) by incrementally adding random slopes to an intercept-only model to identify the most parsimonious structure that captured meaningful variance. As models included group as a between-subject factor, random intercepts for participants were retained in all models to account for individual baseline differences. At each step, likelihood ratio tests were used to compare models and retain only random effects that significantly improved model fit. Fixed effects and interactions were tested using likelihood ratio tests by comparing the final model to nested models with specific fixed effects removed. Follow-up analyses on subsets of the data were conducted when significant interactions were found. For models with only categorical predictors (N400 amplitude and accuracy models), likelihood ratio tests were used. To follow up significant interactions involving the continuous surprisal predictor, we used the emtrends function from the emmeans package (Lenth, 2025) to estimate the slope of surprisal within each level of the interacting factor. Pairwise comparisons of these slopes were conducted to assess whether the effect of surprisal on N400 amplitude differed significantly across levels of the categorical variable.

3.3 Results

Behavioural accuracy

Figure 3-1 presents accuracy in judging sentence acceptability across different background music conditions. There was a marginally significant main effect of group ($\chi^2(1) = 3.85$, $p = .050$) with non-autistic participants ($M = 85.0\%$, $SD = 35.8\%$) showing higher accuracy than autistic participants ($M = 83.0\%$, $SD = 37.6\%$). Significant main effects were observed for both condition contrasts. Accuracy was higher in the non-vocal condition ($M = 94.0\%$, $SD = 23.9\%$) compared to the vocal conditions ($M = 79.0\%$, $SD = 40.8\%$; $\chi^2(1) = 126.10$, $p < .001$). Within the vocal conditions, participants were more accurate in the Simlish condition ($M = 80.3\%$, $SD = 39.7\%$) than in the English condition ($M = 77.6\%$, $SD = 41.7\%$; $\chi^2(1) = 13.94$, $p < .001$). The

interaction between group and condition was significant for contrast 1 ($\chi^2(1) = 4.75$, $p = .029$), and not significant for contrast 2 ($\chi^2(1) = 2.37$, $p = .124$).

Follow-up analyses were conducted to clarify the significant interaction between group and condition contrast 1 (non-vocal vs. vocal). In the non-vocal condition, a significant group effect was observed ($\chi^2(1) = 7.46$, $p = .006$), with non-autistic participants ($M = 95.2\%$, $SD = 21.4\%$) outperforming autistic participants ($M = 92.7\%$, $SD = 26.1\%$). In contrast, within the vocal conditions, the group difference was not significant ($\chi^2(1) = 1.21$, $p = .271$). Significant effects of condition were found in both the autistic ($\chi^2(1) = 56.70$, $p < .001$) and non-autistic group ($\chi^2(1) = 73.10$, $p < .001$). In both cases, participants performed better in the non-vocal condition than in the vocal conditions. Additionally, within the non-autistic group, a significant effect of condition contrast 2 (English vs. Simlish) was observed ($\chi^2(1) = 12.63$, $p < .001$), with higher accuracy in the Simlish condition ($M = 81.7\%$, $SD = 38.7\%$) than in the English condition ($M = 77.9\%$, $SD = 41.5\%$). This effect was not significant in the autistic group ($\chi^2(1) = 2.78$, $p = .096$).

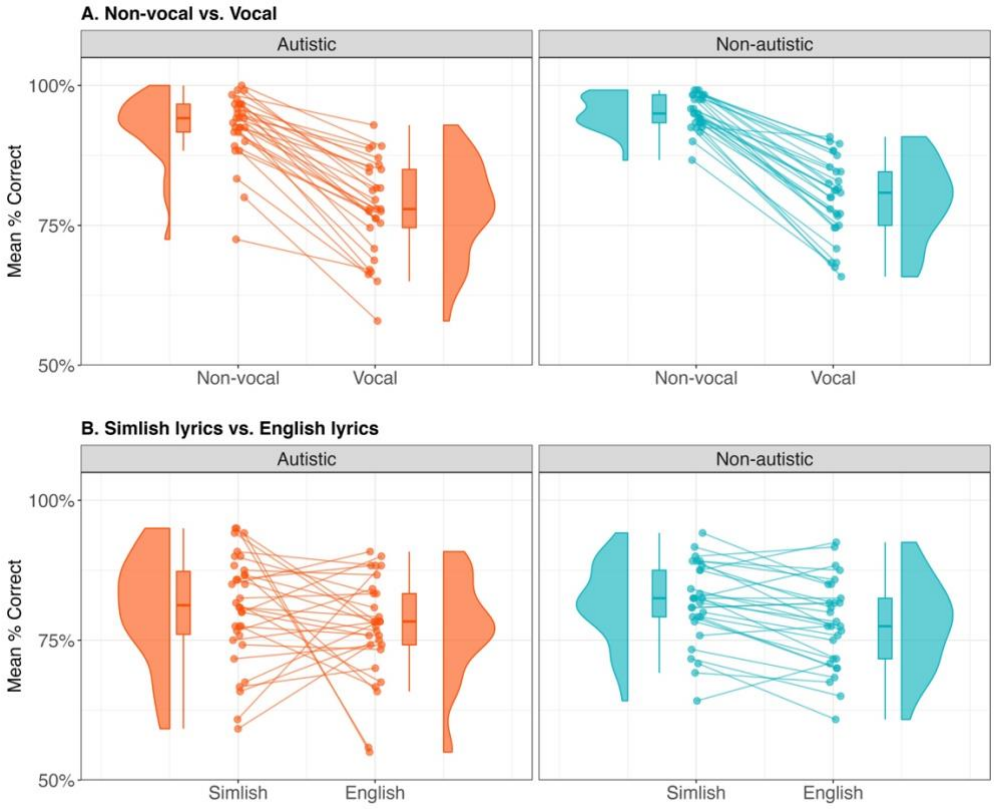

Figure 3-1. Performance accuracy across conditions for autistic and non-autistic groups. (A) Compares non-vocal condition to vocal conditions. (B) Compares Simlish to English conditions. Violin

plots with embedded box plots show the distribution of mean percentage accuracy, with individual data points connected to illustrate within-subject differences.

N400 amplitude (congruency-based analysis)

We first used cluster-based permutation tests to examine the distribution and timing of the N400 effect by comparing amplitudes elicited by incongruent versus congruent sentences across all conditions. Significant spatiotemporal clusters were identified in both groups (both *p*-values < .001), demonstrating a robust N400 effect for both autistic and non-autistic participants. In the non-autistic group, a significant cluster extended over central-parietal electrodes from 273 to 600 ms, with the strongest effects between approximately 350 and 500 ms. In the autistic group, the effect emerged later, with a significant cluster over a similar scalp distribution from 312 to 600 ms, reflecting a shorter cluster duration relative to the non-autistic group (see Figure 3-2A).

Having established the presence of the N400 effect, we next applied cluster-based permutation tests to assess whether its magnitude was modulated by background music condition within each group. In both groups, N400 amplitudes were more negative in the non-vocal than in the vocal condition, with significant central–parietal clusters spanning the 200–600 ms window (both *p*-values < .025). In the non-autistic group, this cluster began earlier and persisted longer than in the autistic group, whose cluster emerged later and was shorter in duration (see Figure 3-2B). No significant clusters were found when comparing the English and Simlish vocal conditions within either group.

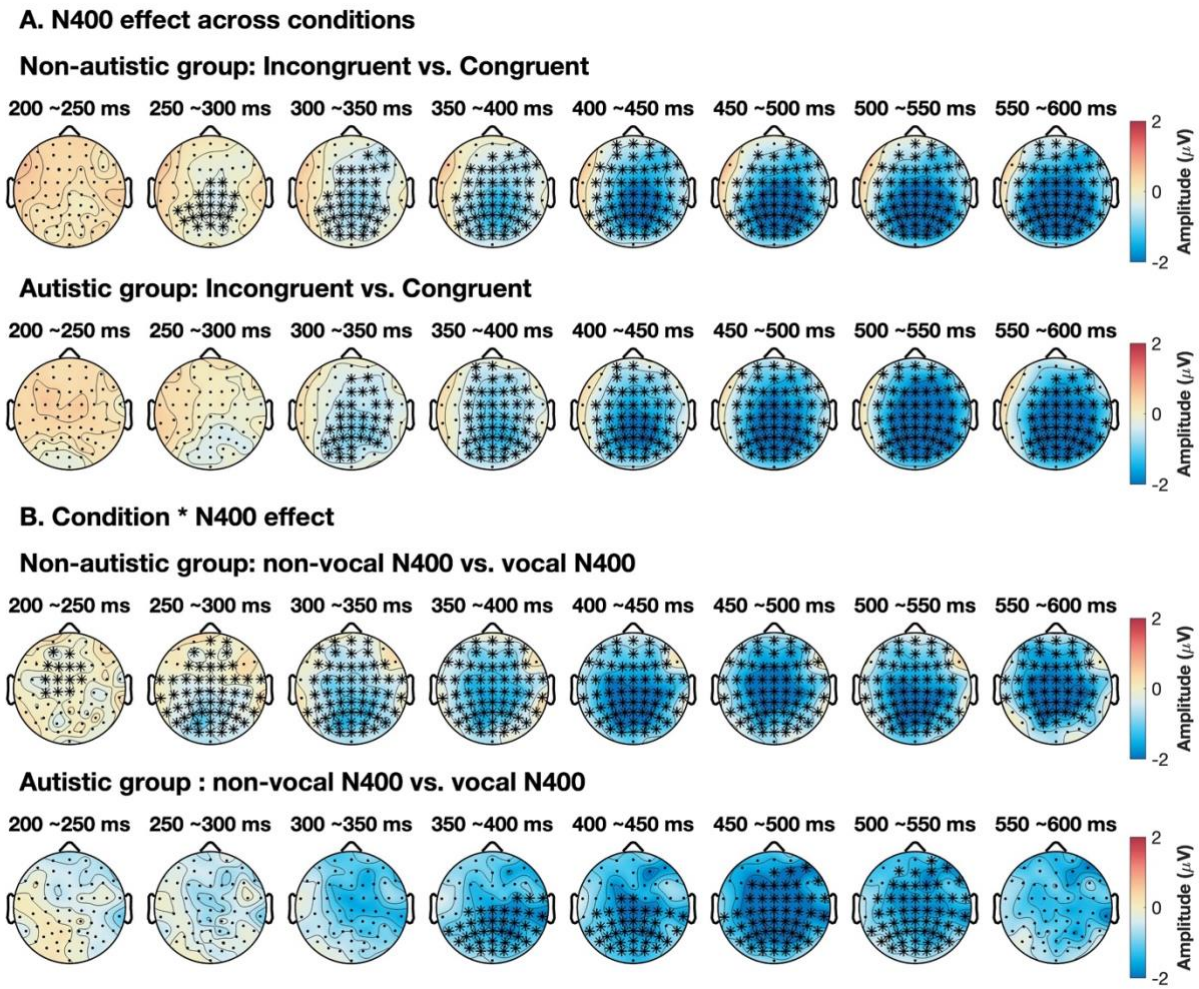

Figure 3-2. Cluster-based permutation tests on *t*-values across time and electrodes. Topographic maps show the spatial distributions, time windows, and strengths of significant clusters (marked with "*"). (A) Comparison of incongruent and congruent sentences across all conditions for each group. (B) Significant clusters for the comparison of N400 effect between non-vocal and vocal conditions.

To examine group differences in N400 amplitude, LMMs were constructed for the 300–600 ms time window (see Figure 3-3). As shown in Table 3-2, a significant three-way interaction between group, condition, and congruency was observed for the non-vocal vs. vocal contrast. No such interaction was found for the English vs. Simlish contrast. Follow-up analyses were conducted to further explore the interaction.

Group × Congruency. A significant interaction was observed in the non-vocal condition ($\chi^2(1) = 16.26$, $p < .001$). In contrast, the interaction in the vocal condition was not significant ($\chi^2(1) = 0.24$, $p = .624$). Follow-up analyses revealed a significant simple group effect in the non-vocal condition, with the non-autistic group showing larger N400 amplitudes in response to congruent sentences compared to the autistic group ($\chi^2(1) = 8.62$, $p = .003$). No group difference was observed for incongruent sentences ($\chi^2(1) = 2.26$, $p = .133$). These results

suggest that compared to the autistic group, the non-autistic group exhibited a stronger N400 effect in the non-vocal condition.

Condition1 × Congruency. Both groups showed a significant main effect of congruency, indicating robust N400 responses (NAS: $\chi^2(1) = 96.24$, $p < .001$; AS: $\chi^2(1) = 46.40$, $p < .001$). A significant interaction was observed in both non-autistic ($\chi^2(1) = 50.90$, $p < .001$) and autistic groups ($\chi^2(1) = 6.76$, $p = .009$). Follow-up analyses within the non-autistic group revealed that incongruent sentences elicited significantly more negative responses in the non-vocal condition than in the vocal condition ($\chi^2(1) = 54.26$, $p < .001$). A significant condition effect was also found for congruent sentences ($\chi^2(1) = 7.92$, $p = .005$), with smaller amplitudes observed in the vocal condition. Follow-up analyses for the autistic group revealed that incongruent sentences elicited significantly more negative responses in the non-vocal condition than in the vocal condition ($\chi^2(1) = 9.46$, $p = .002$), but no significant condition effect in congruent sentences was found ($\chi^2(1) = 0.35$, $p = .557$). In summary, these results suggest that both non-autistic and autistic participants exhibited stronger N400 effect in the non-vocal condition compared to the vocal condition.

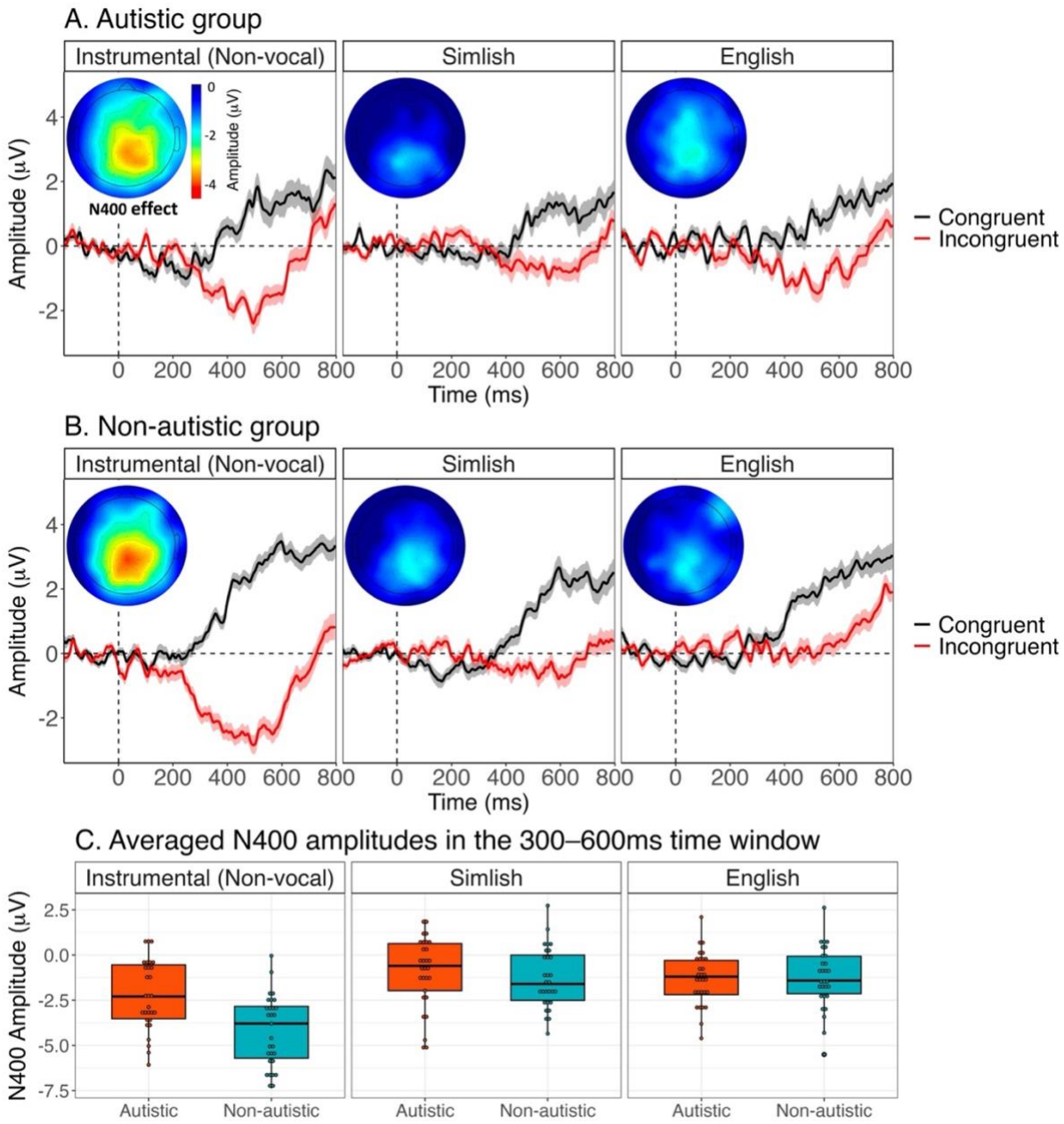

Figure 3-3. N400 analysis across groups and conditions. Topographic maps show the spatial distributions of mean N400 amplitudes in the 300–600 ms time window. (A, B) ERP waveforms for non-autistic (A) and autistic (B) groups, comparing congruent and incongruent sentences across conditions. (C) Averaged N400 amplitudes (300–600 ms) for each group and condition across electrodes of interest.

Table 3-2. LMM results for N400 amplitudes in the congruency-based analysis.

Fixed effects	β	SE	t	χ^2	p
(Intercept)	−0.01	0.17	−0.05	—	—
Group	−0.50	0.32	−1.57	2.42	.121
Congruency	1.92	0.19	10.15	65.54	**< .001**
Condition1	0.51	0.15	3.43	11.72	**< .001**
Condition2	0.18	0.17	1.06	1.13	.288
Group × Condition1	−0.22	0.30	−0.74	0.55	.458
Group × Condition2	−0.60	0.35	−1.73	3.01	.083
Congruency × Group	−0.75	0.33	−2.29	5.78	**.025**
Congruency × Condition1	−2.02	0.30	−6.77	45.84	**< .001**
Congruency × Condition2	0.16	0.34	0.46	0.21	.646
Congruency × Group × Condition1	1.76	0.60	2.93	8.59	**.003**
Congruency × Group × Condition2	0.35	0.69	0.51	0.26	.609

Note. The *p*-values of significant fixed effects are presented in bold.
Model: lmer(Amplitude ~ 1 + Group × Congruency × Condition1 + Group × Congruency × Condition2 + (1 + Congruency | Subject) + (1 | Item)).

N400 amplitude (surprisal-based analysis)

The results of surprisal-based analysis are presented in Table 3-3. First, a significant main effect of surprisal rate (SR) was observed: words with higher surprisal values elicited more negative ERP amplitudes in the N400 time window. This supports the traditional congruency-based findings, reinforcing that lexical predictability modulates N400 amplitude.

A significant interaction between group and SR was observed, indicating that the relationship between SR and N400 amplitude differed between autistic and non-autistic participants. Follow-up simple slopes analysis revealed that SR was negatively associated with N400 amplitude in both the non-autistic group (β = –0.17, 95% CI [–0.20, –0.14]) and the autistic group (β = –0.10, 95% CI [–0.13, –0.07]). The SR slope was significantly steeper in the non-autistic group than in the autistic group ($\Delta\beta$ = 0.07, p = .004), indicating stronger N400 amplitudes to lexical surprisal in the non-autistic group. There was also a significant interaction between Condition1 and SR, revealing different surprisal-N400 relationships for vocal and non-vocal conditions. In both conditions, higher SR was associated with reduced N400 amplitude, but the association was stronger in the non-vocal condition (β = –0.20, 95% CI [–0.24, –0.16]) compared to the vocal condition (β = –0.07, 95% CI [–0.10, –0.04]). The slope

difference indicated that the presence of vocals diminished neural responsiveness to lexical surprisal ($\Delta\beta = 0.11$, $p < .001$).

Together, these two-way interactions observed in the surprisal-based analysis closely mirror the significant two-way interactions found in the traditional N400 congruency analysis, reinforcing the robustness of group and condition effect in semantic processing across analytic approaches. However, unlike the congruency-based model, which showed a significant three-way interaction (Group × Condition1 × SR), this effect in the surprisal model was smaller and did not reach significance ($p = .064$). Visual inspection of the interaction plot (Figure 3-4) revealed a non-significant trend whereby the group difference in surprisal-related N400 modulation appeared greater in the non-vocal than in the vocal condition. This pattern parallels the congruency analysis and suggests convergence in the overall group and context effects across models.

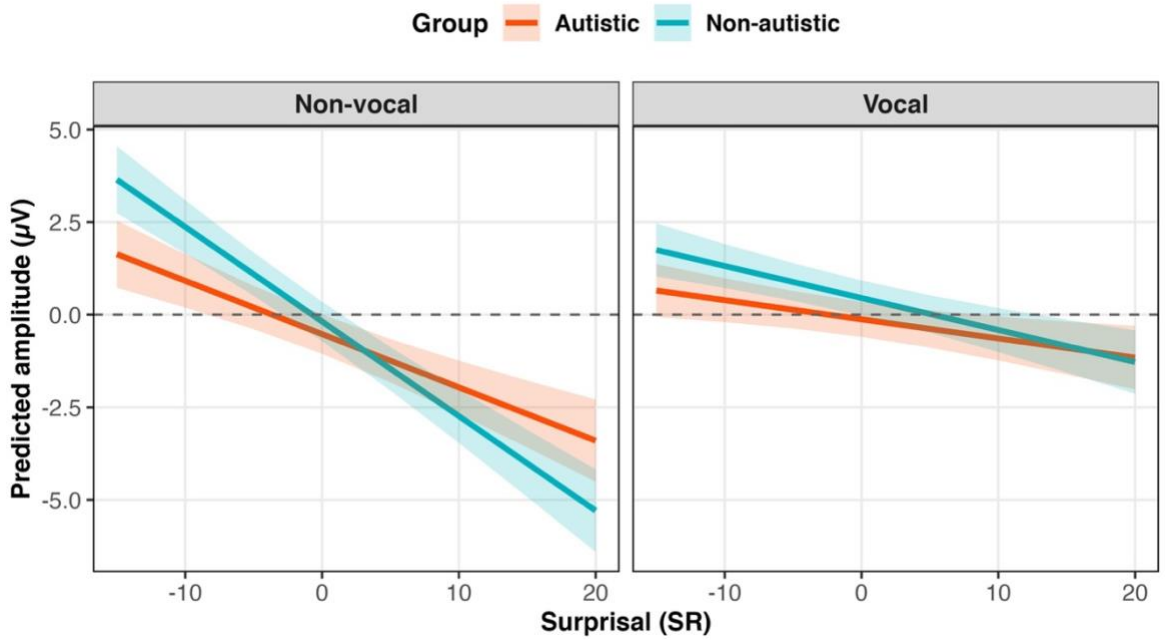

Figure 3-4. Predicted N400 amplitudes as a function of lexical surprisal (SR) for the non-vocal and vocal conditions. Shaded ribbons represent 95% confidence intervals around the model estimates. Lines are plotted separately for the autistic and non-autistic groups.

Table 3-3. LMM results for N400 amplitudes in the surprisal-based analysis.

Fixed effects	*β*	SE	t	χ^2	*p*
(Intercept)	−0.01	0.17	−0.05	—	—
Group	−0.50	0.32	−1.56	2.38	.123
SR	−0.11	0.01	−9.35	78.23	**< .001**
Condition1	0.52	0.15	3.46	11.97	**< .001**
Condition2	0.18	0.17	1.06	1.13	.288
Group × Condition1	−0.24	0.30	−0.78	0.61	.434
Group × Condition2	−0.60	0.35	−1.73	3.00	.083
SR × Group	0.06	0.02	3.11	9.68	**.002**
SR × Condition1	0.13	0.02	6.33	40.02	**< .001**
SR × Condition2	−0.02	0.02	−0.70	0.50	.481
SR × Group × Condition1	−0.08	0.04	−1.86	3.44	.064
SR × Group × Condition2	−0.05	0.05	−0.99	0.98	.322

Note. The *p*-values of significant fixed effects are presented in bold.
Model: lmer(Amplitude ~ 1 + Group × SR × Condition1 + Group × SR × Condition2 + (1 | Subject) + (1 | Item)).

Pearson correlation

Pearson correlation analyses were conducted separately for each group to examine relationships between individual difference measures, behavioural accuracy, and neural responses. Behavioural performance was indexed by mean accuracy, and neural responses were quantified as the mean N400 amplitude from 300–600 ms, averaged across the ROIs used in the N400 analysis, as an index of semantic processing. The individual difference measures comprised: (1) years of formal musical training; (2) Raven's Standard Progressive Matrices standard score (Nonverbal IQ); (3) ROWPVT percentile rank (receptive vocabulary); (4) digit span score (working memory); (5) Autism Spectrum Quotient score; and (6) self-reported ratings of liking and familiarity with the background music. P-values were adjusted for multiple comparisons using the False Discovery Rate correction procedure (Benjamini & Hochberg, 1995). No significant correlations remained after correction for multiple comparisons.

3.4 Discussion

3.4.1 Impact of background music on semantic processing

The present study examined how the presence and intelligibility of background vocals influence the comprehension of spoken sentences under ecologically valid masking conditions. All background music maskers were drawn from the same songs and sung by the same singer,

controlling for acoustic properties of the music and vocal characteristics. Simlish closely resembles English in phonology and prosody but lacks meaningful lexical content, allowing us to isolate the effects of intelligibility. Accuracy was significantly lower when target speech was accompanied by sung vocals than by instrumental music, and lower for intelligible (English) than for unintelligible (Simlish) lyrics. These patterns indicate that the presence of sung vocals impairs comprehension by increasing competition for perceptual and attentional resources, and that intelligible lyrics add a further source of interference through competition at lexical and semantic levels (Summers & Roberts, 2020). The replication of both effects mirrors findings from Brouwer et al. (2022), who used the same background music manipulations in a sentence repetition task. By using a less demanding sentence acceptability task, we were able to measure neural responses concurrently, demonstrating that these interference effects are robust across tasks and reflect general constraints on processing speech in the presence of competing vocal streams.

The N400 results showed significant differences in the vocal vs. non-vocal comparison. The attenuation of N400 amplitudes in vocal conditions suggests that listeners were less able to generate or use precise context-based predictions to guide semantic processing. This is consistent with studies reporting smaller N400 responses when speech is heavily degraded or masked, an effect attributed to reduced access to lexical representations or weakened top-down predictions (Aydelott et al., 2006; Obleser & Kotz, 2011; Silcox & Payne, 2021; Strauß et al., 2013). By contrast, other studies have observed larger N400s under moderate masking, interpreted as reflecting increased processing effort or greater reliance on contextual information to compensate for partial loss of bottom-up input (Devaraju et al., 2021; Kemp et al., 2019; Romei et al., 2011; Song et al., 2020). These divergent patterns likely reflect differences in the severity and nature of the masker. In the present study, the −6 dB SNR, selected based on pilot testing, likely introduced substantial energetic masking, reducing the fidelity of bottom-up input and thereby degrading prediction precision to the point where contextual information conferred little additional benefit. The use of sung vocals may have further increased cognitive load by simultaneously engaging musical and linguistic processing systems that share overlapping neural resources (Koelsch et al., 2005; Maess et al., 2001; Patel, 1998, 2003; Slevc et al., 2009). Under such conditions, prediction-based facilitation may be substantially reduced, resulting in attenuated N400 amplitudes.

In contrast to the behavioural results, the N400 did not differ significantly between the English and Simlish conditions. From a predictive processing perspective, this lack of differentiation suggests that under both types of vocal masking, the precision of context-based semantic predictions was already reduced to a similar extent. At −6 dB SNR, the combination of energetic masking from the overlapping speech-like acoustics and informational masking from competing vocal streams likely degraded the target signal enough that prediction error signals, as indexed by the N400, were attenuated to a comparable floor level in both conditions. Under such load, listeners may have relied more on controlled, effortful mechanisms to complete the task, drawing resources away from the N400's early semantic effects (Obleser & Kotz, 2011; Peelle, 2018; Strauß et al., 2013). This interpretation aligns with Load Theory (Lavie et al., 2004), which posits that high perceptual load limits the processing of competing information. Behavioural performance, however, remained sensitive to the intelligibility contrast, suggesting that comprehension under these challenging conditions was supported by later, controlled mechanisms outside the N400 time window (Blackford et al., 2012; Peelle, 2018; Wöstmann et al., 2015).

We also examined whether self-reported familiarity with and liking of the background music, as well as musical training background, modulated behavioural or neural outcomes. None of these factors showed significant correlations with accuracy or N400 measures. This finding contrasts with prior evidence that familiarity and musical expertise can influence speech perception in background music. Familiar music has been reported to either hinder performance, by capturing attention and evoking autobiographical memories (De Groot & Smedinga, 2014; Janata et al., 2007), or facilitate it, by increasing predictability and arousal, thereby aiding distractor suppression (Feng & Bidelman, 2015; Russo & Pichora-Fuller, 2008). Musical expertise has been associated with enhanced speech-in-noise perception and auditory stream segregation (Bidelman & Yoo, 2020; Strait & Kraus, 2011; Zendel et al., 2015), although evidence is less consistent when the masker is music rather than speech, as in speech-in-music paradigms (Brown & Bidelman, 2022b; Patston & Tippett, 2011).

Several methodological and sample-related factors may account for the absence of such effects in the present study. Our participants varied considerably in musical training, whereas familiarity and liking were relatively uniform, and all three variables were analysed as continuous predictors. Prior studies, by contrast, have typically categorised participants into high- and low-familiarity or high- and low-musicality groups (Brown & Bidelman, 2022b,

2023; Zheng et al., 2023), a strategy that maximises between-group contrasts and may increase sensitivity to effects that emerge only at higher levels of familiarity or skill. A second consideration is the nature of our paradigm, which combined both energetic and informational masking; this higher level of difficulty may have overridden any potential advantages of familiarity or expertise that might be more apparent in less acoustically demanding settings used by previous studies. Future research could address these possibilities by recruiting larger, more balanced samples that enable grouping-based analyses, systematically manipulating SNR to test effects under different masking loads and examining potential non-linear relationships between listener characteristics and speech-in-music processing. Furthermore, incorporating more naturalistic, continuous stimuli could increase sensitivity to familiarity and musical ability effects. For example, Brown & Bidelman (2022a) used continuous speech-in-music materials with cortical tracking analyses, providing a powerful approach for capturing subtle, listener-specific effects that may be missed with shorter, discrete stimuli.

3.4.2 Atypical semantic processing in autism

To our knowledge, this is the first study to examine speech-in-music processing in autism, combining behavioural measures with neural indices of semantic processing reflected by the N400. Music is a common and ecologically valid background sound in daily life, yet its impact on speech comprehension in autism remains largely unexplored. Previous research suggests that many autistic individuals show a strong interest in music and, in some cases, enhanced musical abilities such as superior pitch discrimination and melodic memory (Heaton, Williams, et al., 2008; O'Connor, 2012). This dissociation underscores the need to examine how background music influences speech comprehension in autistic listeners. Across conditions, autistic participants showed lower overall accuracy than non-autistic participants, indicating group-level differences in comprehension performance. These results align with and extend earlier behavioural evidence that autistic individuals experience greater difficulty than non-autistic peers in challenging listening situations involving informational masking, particularly when competing vocal information is present (Bendo et al., 2024; Emmons et al., 2022; Lau et al., 2023; Ruiz Callejo & Boets, 2023). While most previous work has focused on the effect of speech maskers, our findings show that background music can significantly disrupt comprehension, with a greater impact on autistic than non-autistic listeners, even among those with intact cognitive and language abilities.

N400 analyses also revealed clear group differences in semantic processing. In the congruency-based analysis, autistic participants showed a significant N400 cluster that emerged later, covered a shorter time range, and had reduced amplitudes compared with the non-autistic group. This pattern is consistent with prior reports of reduced amplitude and/or delayed latency in autism, both in quiet listening (Braeutigam et al., 2008; Manfredi et al., 2020; Pijnacker et al., 2010; Ribeiro et al., 2013; Ring et al., 2007) and under competing speech or babble maskers (Li et al., 2025). The present findings extend this evidence by showing that semantic processing differences in autism are not limited to speech-based maskers, but also arise when the competing signal is music, underscoring the generality of these effects across ecologically relevant listening conditions. Surprisal-based analyses provided converging evidence: while higher surprisal values elicited more negative-going N400 responses in both groups, the relationship was steeper in non-autistic listeners, indicating greater neural responsiveness to fine-grained variation in lexical predictability. The flatter slope in autistic listeners suggests weaker modulation of neural responses by probabilistic expectations, in line with the attenuated facilitation effects observed in the congruency-based analysis. Taken together, these results suggest less efficient prediction-based semantic processing in autism.

Group differences were most pronounced in the instrumental music condition, where autistic participants showed lower accuracy and attenuated N400 responses compared with non-autistic peers. This condition, combining a low SNR with minimal linguistic masking, was likely the most favourable listening environment overall. In this context, non-autistic listeners appeared able to capitalise on semantic cues to support comprehension, as reflected in their stronger N400 responses, suggesting more effective engagement of predictive processing when the target speech was relatively unmasked. In contrast, autistic listeners showed a markedly reduced N400 in this condition, with the largest group divergence occurring for congruent trials. Although N400 group differences are most often driven by responses to incongruent items, the present finding of a larger divergence for congruent trials is not unprecedented. Several studies have similarly reported group effects restricted to, or more pronounced for, congruent items (Ahtam et al., 2020; Li et al., 2025; O'Rourke & Coderre, 2021; Ring et al., 2007). This pattern is typically interpreted as reduced semantic facilitation, where supportive context does not fully pre-activate or ease the integration of predictable words. For example, in our previous speech-in-noise study (Li et al., 2025), autistic participants produced smaller N400 responses and lower accuracy for congruent sentences, particularly under competing speech masking, consistent with a reduced benefit from facilitative context. In the present study, the attenuation

to expected words, together with comparable responses to incongruent items, similarly points to a diminished capacity to capitalise on predictive context when it is available.

Autistic participants also showed similar accuracy when speech was embedded in music with either Simlish or English lyrics, despite these masker types differing in their degree of linguistic interference. While Simlish is unintelligible and should impose less linguistic masking than English, non-autistic listeners capitalised on this advantage, achieving higher accuracy in the Simlish condition. The absence of such a benefit in the autistic group suggests reduced sensitivity to changes in task demands, reflecting a more rigid processing strategy across conditions. A similar pattern emerged in our previous speech-in-noise study (Li et al., 2025), where autistic participants showed comparable N400 amplitudes across quiet, babble, and competing speech conditions, unlike non-autistic participants who modulated their responses with masker type. Converging evidence from Mamashli et al. (2017) further links such reduced adaptability to atypical large-scale connectivity patterns, which may limit the dynamic reallocation of processing resources needed to adjust semantic operations to varying acoustic and linguistic demands. Taken together, these findings suggest that autistic listeners may engage semantic processing in a relatively fixed manner, showing limited adjustment when masker characteristics change.

However, this group-level pattern should be interpreted with caution given the notable heterogeneity observed within the autistic group (Figure 3-1). While some participants showed greater difficulty with English lyrics compared to Simlish, others performed better in the English condition. Such variability suggests that not all autistic individuals adopt the same strategy, and that individual differences in processing style, attentional focus, or the capacity to adapt to specific masking conditions may play an important role (Hernandez et al., 2020; Kalas, 2012; Preis et al., 2016).

Taken together, the behavioural and neural findings indicate that autistic participants comprehended speech in background music less accurately and showed attenuated N400 responses, reflecting weaker engagement of predictive semantic context. These effects were accompanied by limited modulation across masking conditions, suggesting reduced flexibility in adapting semantic processing to the characteristics of the competing music maskers. This pattern is consistent with predictive coding accounts of autistic perception, which suggest that perceptual inference is shaped by atypical precision weighting, meaning differences in how

strongly the brain emphasises mismatches between incoming sensory input and prior expectations (Lawson et al., 2014; Van Boxtel & Lu, 2013; Van de Cruys et al., 2014). In typical listening, the influence of top-down predictions is dynamically adjusted according to the reliability of sensory input. Evidence from speech-in-noise studies shows that listeners increase their reliance on contextual predictions when the signal is clear and reduce it when the signal is degraded (Mattys et al., 2009, 2012). By contrast, in autism the precision assigned to prediction errors may be adjusted less flexibly, so the balance between sensory evidence and prior expectations does not adapt efficiently to changing listening conditions. Applied to our experiment, this suggests that under less challenging masking—particularly the non-vocal condition—non-autistic listeners could down-weight unpredictable acoustic details and rely more on semantic context, whereas autistic listeners may have continued to treat the signal as highly unpredictable. Such reduced flexibility could explain both the absence of accuracy gains and the attenuated N400 responses observed in autistic participants.

3.4.3 Complementary insights from congruency and surprisal analyses

Including both a categorical semantic anomaly contrast and a continuous surprisal measure allowed us to assess complementary aspects of context-based semantic processing. The congruency contrast captures the neural cost of processing sentence-final words that strongly violate contextual predictions, while surprisal provides a graded, trial-by-trial estimate of the unexpectedness of the sentence-final word given its preceding context, indexing more subtle variation in contextual prediction strength. Using these approaches in parallel enabled us to examine both robust, high-certainty violations and fine-grained probabilistic effects within the same dataset.

As anticipated, both analyses revealed robust N400 modulation by masking condition, with larger amplitudes in the non-vocal condition. This convergence supports the validity of our paradigm for eliciting semantic prediction effects under ecologically valid masking. However, the two measures differed in their sensitivity to higher-order interactions involving group and condition. The congruency-based model identified a significant Group × Congruency × Condition1 interaction, indicating that group differences in anomaly processing varied with masker type. In contrast, the surprisal model produced significant main effects of Group and Condition-1 but not the critical three-way interaction.

These differences may partly reflect the statistical properties of the models. The congruency variable was binary, allowing for a more straightforward random-slope structure, whereas surprisal was continuous and required a more constrained random-effects specification in its final form due to model convergence issues. This difference in modelling flexibility may have reduced sensitivity to complex higher-order interactions in the surprisal analysis. At the same time, the measures differ in scope: congruency effects are most likely to emerge when prediction precision is high, as in favourable non-vocal listening conditions, making group differences more detectable under clear violation contexts. Surprisal, by modelling incremental probability differences for each target word across all trials, captures more diffuse fluctuations in prediction error. While this broader coverage can be advantageous for detecting graded effects, it can also disperse the signal across contexts with varying levels of predictability, reducing statistical sensitivity to complex interactions in the present dataset.

From a methodological standpoint, surprisal-based analysis offers a powerful complement to categorical approaches by replacing binary stimulus labels with probability estimates that capture variation in contextual predictability even among trials classified within the same congruency category. Previous work shows that surprisal predicts N400 amplitude as well as, and in some cases better than, traditional cloze probability (Michaelov et al., 2023, 2024). This advantage has been demonstrated in recent continuous-speech EEG research, where surprisal is aligned to each word in a narrative and modelled alongside other linguistic and acoustic predictors. For example, Weissbart et al. (2020) computed word-by-word surprisal for an entire spoken story using a recurrent neural network language model and incorporated these values into a temporal response function model (Crosse et al., 2016), finding reliable cortical tracking of surprisal over time.

In the present study, surprisal was calculated only for the sentence-final target word, enabling a direct methodological match to the congruency-based analysis. While this approach provides a more limited estimate than word-by-word surprisal across entire sentences, it still captured variability in semantic processing across all trials, irrespective of categorical congruency labels. The generally consistent patterns observed across the two measures suggest that both categorical and graded perspectives are informative in ecologically valid listening contexts. Future studies could extend this approach by incorporating word-by-word surprisal measures, as in Weissbart et al. (2020), to provide a more detailed assessment of predictive processing under complex listening conditions.

3.5 Conclusion

This study is the first to examine speech-in-music processing in autism using combined behavioural and electrophysiological measures. We found that vocal music imposed greater impact on speech comprehension and N400 responses relative to instrumental music and intelligible lyrics exerting additional interference at the behavioural level. Autistic participants showed lower overall accuracy, attenuated N400 responses, and reduced sensitivity to changes in masker type, indicating differences in the use of contextual predictability.

Using both congruency- and surprisal-based analyses revealed generally convergent masking effects and complementary insights into semantic prediction, with the categorical measure highlighting robust violations of prediction and the continuous measure capturing graded probabilistic variation. Our results demonstrate the feasibility of applying surprisal in a discrete-trial ERP paradigm with background masking, providing a link to continuous-speech tracking approaches. These findings extend speech-in-noise research to the common but under-studied case of background music and highlight the utility of combining categorical and probabilistic methods to characterise semantic processing under ecologically valid conditions.

Chapter 4

Study 3: Auditory and Semantic Processing of Speech-in-Noise in Autism: A Behavioural and EEG Study

Abstract

Autistic individuals often struggle to recognise speech in noisy environments, but the neural mechanisms behind these challenges remain unclear. Effective speech-in-noise (SiN) processing relies on auditory processing, which tracks target sounds amidst noise, and semantic processing, which further integrates relevant acoustic information to derive meaning. This study examined these two processes in autism.

Thirty-one autistic and 31 non-autistic adults completed a sentence judgement task under three conditions: quiet, babble noise, and competing speech. Auditory processing was measured using EEG-derived temporal response functions (TRFs), which tracked how the brain follows speech sounds, while semantic processing was assessed via behavioural accuracy and the N400 component, a neural marker of semantic processing.

Autistic participants showed reduced TRF responses and delayed N400 onset, indicating less efficient auditory processing and slower semantic processing, despite similar N400 amplitude and behavioural performance. Moreover, non-autistic participants demonstrated a trade-off between auditory and semantic processing resources. In the competing speech condition, they showed enhanced semantic integration but reduced neural tracking of auditory information when managing linguistic competition introduced by intelligible speech noise. In contrast, the autistic group showed no modulation of neural responses, suggesting reduced flexibility in adjusting auditory and semantic demands.

These findings highlight distinct neural processing patterns in autistic individuals during SiN tasks, providing new insights into how atypical auditory and semantic processing shape SiN perception in autism.

Keywords: autism, N400, neural tracking, speech-in-noise, temporal response functions

4.1 Introduction

Recognising speech in noisy environments, a process known as speech-in-noise (SiN) processing, is a complex task influenced by both auditory and cognitive interference from competing sounds (Bronkhorst, 2000). Background noise can physically mask speech signals, obscuring key acoustic features and making perception more difficult. This challenge increases when the background contains intelligible speech with similar vocal characteristics, which introduces additional cognitive interference and makes it harder to focus on the target signal (Başkent & Gaudrain, 2016; Brungart, 2001).

For autistic individuals, these difficulties can be even more pronounced due to atypical auditory and cognitive profile (O'Connor, 2012; Ouimet et al., 2012). Previous research on SiN recognition in autism has predominantly focused on auditory processing difficulties, such as challenges in utilising temporal dips (Alcántara et al., 2004; Groen et al., 2009). Autistic participants are less able to use these brief reductions in noise intensity to enhance target speech recognition. These difficulties extend to continuous noise without temporal dips, particularly under stricter recognition criteria (Schelinski & Von Kriegstein, 2020). In multi-speaker scenarios, autistic listeners experience difficulties in using speaker-relevant cues, such as spatial or vocal features, to enhance speech separation (DePape et al., 2012; Schaeffer et al., 2023). Additionally, atypical auditory processing in autism is compounded by differences in higher-order cognitive functions, including verbal abilities (Ruiz Callejo et al., 2023; Russo et al., 2009), attentional control (Emmons et al., 2022), and the integration of auditory information (Lepistö et al., 2009). Neuroimaging studies provide further insights into the neural mechanisms underlying these auditory processing difficulties**.** Impairments in sensory control have been linked to reduced neural responses in the inferior frontal gyrus under noisy conditions, suggesting disrupted top-down modulation (Schelinski & von Kriegstein, 2023). Heightened activity in the speech-processing cortex during SiN tasks indicates compensatory mechanisms for managing auditory challenges (Hernandez et al., 2020). Additionally, increased recruitment of neural resources regardless of task difficulty points to inflexible resource allocation in autism (Mamashli et al., 2017).

Collectively, these findings highlight both auditory difficulties and cognitive challenges during SiN processing in autism. However, no studies have examined SiN recognition in autism with a combined focus on both auditory and semantic processing, despite the crucial role each plays

in successful comprehension. To test this, we used electroencephalography (EEG) to examine both auditory and semantic processing in autistic and non-autistic individuals. Our study builds on Song et al. (2020), who explored the effects of competing speech and babble noise on speech perception. Their findings revealed a significant trade-off between auditory and semantic processing in non-autistic listeners. Compared to the unintelligible babble masker, the intelligible speech masker resulted in amplified N400 amplitudes, indicating greater reliance on semantic processing. However, this increased semantic effort was accompanied by less accurate neural tracking of the target speech, suggesting reduced auditory processing. These results support the idea that cognitive resources are limited and dynamically allocated, with greater engagement in semantic processing diminishing resources available for auditory processing. When speech is degraded by noise, skilled listeners rely more on semantic context to compensate for lost acoustic information, thereby facilitating comprehension (Bilger et al., 1984; Kalikow et al., 1977).

The present study builds on this framework to examine whether autistic individuals adopt a similar compensatory strategy during SiN processing. Following Song et al. (2020), we employed a semantic congruency task across three listening conditions: quiet, single-talker speech noise, and babble noise. To investigate auditory processing, we measured neural tracking of speech envelopes, which capture continuous amplitude fluctuations in speech. Neural tracking reflects the brain's ability to synchronise with rhythmic external stimuli, such as speech (Brodbeck & Simon, 2020; Ding & Simon, 2012b). Neural tracking was estimated using a machine learning approach to predict neural responses from speech envelopes, known as forward modelling (Crosse et al., 2016, 2021). Compared to backward modelling, which reconstructs the stimulus from neural data (Song et al., 2020), forward modelling offers greater insight into the temporal dynamics of speech processing. This approach allows us to examine speech encoding over time (Holdgraf et al., 2017), making it particularly suited for examining the time-resolved neural processes involved in speech perception under noisy conditions (Ding & Simon, 2013; Gillis et al., 2022; Yasmin et al., 2023; Zhang et al., 2023). From this modelling, we obtained the temporal response function (TRF) and focused on P1, N1, and P2 responses. These components closely correspond to auditory evoked potentials (AEPs) and are thought to reflect different auditory processing stages. For example, P1 is associated with early acoustic encoding, while N1–P2 is linked to attention and speech intelligibility. Such temporally specific information is not accessible through backward modelling, which provides

a single global measure of decoding accuracy but lacks interpretable component-level resolution.

Although no previous studies have examined TRF components in autistic individuals during SiN tasks, the well-documented auditory processing difficulties in noise led us to hypothesise that autistic participants would exhibit reduced P1–N1–P2 responses across all conditions. This hypothesis is further supported by findings of atypical neural entrainment in autism in quiet environments (Jochaut et al., 2015), suggesting difficulties in synchronising brain activity with speech. AEP studies have also reported atypical P1–N1–P2 responses in autism, indicating reduced cortical responsiveness to acoustic input (O'Connor, 2012; Schwartz et al., 2023).

We evaluated semantic processing through both behavioural judgments of semantic violations within the semantic congruency task and corresponding neural responses. Autistic individuals often exhibit atypical cortical response to semantic information even without the presence of noise. This has been investigated using N400, an ERP component widely recognised as a neural marker of lexical-semantic processing (Kutas & Hillyard, 1980). Typically, N400 amplitudes are larger for less predictable or incongruent words, reflecting greater difficulty in resolving meaning (Hagoort, 2008; Osterhout & Holcomb, 1992). However, N400 responses are also influenced by individual differences in cognitive and language abilities, and considerable variability has been observed within the autistic population. Autistic individuals—particularly children with poor verbal abilities—often exhibit reduced, delayed, or atypically distributed N400 compared to their non-autistic peers, suggesting difficulties with semantic integration (Coderre et al., 2017; Fishman et al., 2011; Pijnacker et al., 2010). In contrast, studies focusing on autistic individuals with stronger verbal abilities have reported relatively typical patterns of semantic processing (DiStefano et al., 2019; Henderson et al., 2011; McCleery et al., 2010). Given previous findings of attenuated N400 responses in quiet conditions, we hypothesised that autistic participants would exhibit reduced and delayed N400 responses to SiN stimuli, indicating challenges in semantic integration in noisy environments.

Considering the effect of masker types, we also hypothesised that masker intelligibility would impact the trade-offs between auditory and semantic processing. For non-autistic participants, we expected stronger N400 and weaker TRF responses in the intelligible speech masker condition compared to the unintelligible babble condition, consistent with Song et al. (2020). In contrast, we predicted that autistic participants would show less differentiation between

conditions of varying intelligibility, reflecting reduced top-down modulation during SiN processing.

Finally, prior research has identified a range of cognitive factors that may contribute to variability in SiN perception among autistic individuals. For example, temporal processing difficulties have been found to correlate more closely with language ability than with autism diagnosis per se (DePape et al., 2012; Bhatara et al., 2013). Similarly, verbal IQ may influence performance at an individual level, even when group-level differences are not observed (Ruiz Callejo et al., 2023). Difficulties with selective auditory attention have also been reported in autism (Emmons et al., 2022; Lau et al., 2023). Taken together, these findings highlight the complex and multifactorial nature of SiN perception in autism. Based on this evidence, and in line with recent work showing that cognitive abilities can predict neural and behavioural responses to SiN (Ruiz Callejo & Boets, 2023), we conducted exploratory correlation analyses to examine potential associations among cognitive abilities, behavioural accuracy, and neural responses.

4.2 Methods

Participants

We recruited 31 autistic and 31 non-autistic participants, aged 17–47, all of whom were right-handed native English speakers. Participants passed a hearing screening using an Amplivox manual audiometer, confirming normal hearing in both ears at 25 dB for frequencies of 0.5, 1, 2, and 4 kHz. Both groups had no current speech, language, or communication needs. Autistic participants had diagnoses confirmed by professional clinicians and supported by clinical reports. Non-autistic participants reported no personal or family history of autism, and this was further supported by their scores on the Autism Spectrum Quotient (AQ) (Baron-Cohen et al., 2001), all of which were below the cut-off of 32.

We measured cognitive abilities that may influence SiN processing (Gordon-Salant & Cole, 2016; Heinrich, 2021). Nonverbal IQ was measured using Raven's Standard Progressive Matrices (Raven & Court, 1998), while receptive vocabulary, a proxy for verbal IQ, was assessed using the Receptive One-Word Picture Vocabulary Test (ROWPVT-4; Martin & Brownell, 2011). Verbal short-term memory was evaluated with the digit span task (Wechsler et al., 2003). Participants also completed a musical training questionnaire (Pfordresher &

Halpern, 2013), which recorded years of formal training across various instruments. Additionally, auditory-related traits were measured using the Auditory Attention and Discomfort Questionnaire (Dunlop et al., 2016), which assessed difficulties with auditory attention in noisy environments and sensitivity to auditory stimuli in daily life. Demographic and cognitive data are summarised in Table 4-1. There were no significant differences between autistic and non-autistic groups in chronological age, musical training background, receptive vocabulary, nonverbal reasoning ability, or verbal short-term memory. However, the autistic group scored significantly higher on the AQ, reflecting elevated autistic traits, and reported greater auditory attention difficulties and discomfort.

The study was approved by the University Research Ethics Committee, and all participants provided written informed consent. Participants received financial compensation. Student participants recruited from the psychology participant pool were awarded course credits.

Table 4-1. Characteristics of the autistic (*n* = 31) and non-autistic (*n* = 31) groups.

Variables	Autistic *Mean (SD)*	Non-autistic *Mean (SD)*	*W*	*p*	Rank-biserial Correlation
Gender (Female:Male)	22:9	26:5			
Age	25.73 (7.89)	25.78 (7.83)	484.0	.97	0.01
Musical training years	4.02 (5.61)	6.39 (7.02)	384.0	.16	−0.20
Nonverbal reasoning (RSPM raw score)	53.87 (3.59)	54.39 (3.61)	441.0	.58	−0.08
Nonverbal reasoning (RSPM percentile)	49.03 (23.96)	52.74 (29.32)	458.5	.75	−0.05
Receptive vocabulary (ROWPVT-4 raw score)	167.16 (10.40)	170.03 (8.35)	429.5	.48	−0.11
Receptive vocabulary (ROWPVT-4 standard score)	109.26 (15.92)	113.10 (14.69)	420.5	.40	−0.10
Digit Span	7.07 (1.61)	7.07 (1.03)	464.0	.82	−0.03
Auditory attention difficulty	38.58 (10.03)	24.74 (9.47)	811.5	< .01	0.69
Auditory discomfort	60.94 (9.76)	43.81 (10.44)	855.5	< .01	0.78
Autistic traits (AQ)	38.29 (6.62)	17.13 (8.49)	935.0	< .01	0.95

Stimuli and apparatus

The target stimuli consisted of 180 sentence pairs with highly constraining contexts (see Appendix B for the full list). The final word in each sentence was either semantically congruent (e.g., I passed my test and got my driving licence) or incongruent with the preceding context (e.g., I passed my test and got my driving discount).

Semantically incongruent sentences were expected to elicit larger N400 amplitudes than congruent sentences, reflecting the modulation of N400 responses during semantic integration. Sentences were drawn from a validated set developed by Stringer and Iverson (2020). Each sentence contained 5–10 words (5–13 syllables) and was recorded by a female native speaker of Southern British English.

The masker stimuli were adopted from Song et al. (2020). To prevent participants from relying on acoustic cues to differentiate between competing voices, maskers were created using recordings of the same English stories, read by the same speaker who recorded the target sentences. For the single-talker speech masker, pauses in the recordings were reduced to less than 25 milliseconds, and low-frequency amplitude modulations (<1 Hz) were attenuated through filtering. The babble masker was designed to eliminate intelligible linguistic structure and speaker-specific acoustic cues while preserving speech-like spectral and temporal properties. The stories were first segmented into short excerpts of 1.5 to 2.5 seconds, excluding segments containing more than 15% silence. These excerpts were then randomly reordered and concatenated, disrupting natural temporal continuity. Twelve of these randomised sequences were subsequently superimposed, producing a dense, speech-derived babble that was acoustically uniform and unintelligible. The signal-to-noise ratio was set to 0 dB, based on a pilot study (see Appendix A for details of the pilot study).

Participants completed the experiment using E-Prime 3.0 software in a soundproof booth. Audio stimuli were presented binaurally through Etymotic ER-1 earphones at 67 dB sound pressure level. Participants judged sentence acceptability while disregarding background noise. Prior to the experiment, participants completed three practice items per condition to ensure understanding of the task. During each trial, an audio file was played alongside a fixation cross displayed on the screen. After a silent interval (1.5–1.7 seconds), participants judged the sentence as acceptable or unacceptable.

The experiment comprised six blocks (two per condition), with 60 trials per block lasting 5−6 minutes. Sentences of varying congruency were randomly mixed, and block order was randomised. To minimise context effects, three experimental lists were created, with conditions counterbalanced across lists. Lists were randomly assigned to participants. Self-paced breaks between blocks were provided to reduce fatigue.

EEG recording and pre-processing

EEG data were recorded using a Biosemi Active Two system with 64 Ag/AgCl electrodes and six external electrodes (left/right mastoids and vertical/horizontal electrooculography). Signals were recorded at a sampling rate of 2048 Hz without referencing, and electrode impedances were kept below ±30 μV. Triggers marking the onset of target words were recorded with the EEG. Data pre-processing was performed in EEGLAB (Delorme & Makeig, 2004) within Matlab R2018b. For TRF analysis, EEG signals were band-pass filtered between 1−8 Hz using a zero-phase Butterworth filter to isolate low-frequency activity (Ahissar et al., 2001; Luo & Poeppel, 2007). The data were then downsampled to 64 Hz for computational efficiency. The speech envelope, used as the input acoustic feature for TRF modelling, was extracted via the Hilbert transform, downsampled to 64 Hz, and normalised with the EEG data (mean-subtracted and standardised). EEG trials were precisely aligned with stimulus segments to ensure matching data lengths.

For N400 analysis, signals were low-pass filtered at 40 Hz using a zero-phase Butterworth filter, downsampled to 256 Hz, and re-referenced to the average of the mastoids. Data were segmented into epochs ranging from −200 ms to 800 ms relative to the target word onset and baseline-corrected using the pre-stimulus interval (−200 ms to 0 ms). Bad channels were manually identified and interpolated. Independent Component Analysis was performed using the runica algorithm implemented in EEGLAB to decompose the continuous EEG data into independent components. Artefactual components were identified and rejected based on both automatic classification and manual inspection. Specifically, we used the ICLabel plugin (Pion-Tonachini et al., 2019) to estimate the probability that each component reflected neural activity, eye movements, muscle activity, or other sources of noise. Components classified as “eye” or “muscle” with a probability of at least 75% were considered candidates for removal. All flagged components were further examined manually, with particular attention to topography, time series, and power spectrum characteristics. On average, 3.87 trials per participant (approximately 1% of all trials) were excluded due to artefacts in the autistic group, and 2.00

trials per participant (approximately 0.6% of all trials) were excluded in the non-autistic group. A Wilcoxon rank-sum test showed no significant group difference in the number of excluded trials ($W = 577.5$, $p = .154$), with a small, non-significant effect size ($r = 0.20$, 95% CI [−0.08, 0.46]), indicating comparable trial rejection rates across groups.

EEG data analysis

TRF modelling. We conducted TRF modelling using the mTRF toolbox (Crosse et al., 2016). Models were fitted with a time-lag window of [−100, 400 ms] to capture neural responses at latencies between 0 and 300 ms (Di Liberto et al., 2018). Separate models were created for each condition and group, with ridge regression and regularisation (λ) employed to prevent overfitting. An individual, subject-specific approach was used to train and cross-validate the TRF models, following the procedures outlined in Crosse et al. (2021), to estimate the TRF that best fits each participant's neural responses. Optimal λ values were selected via 10-fold cross-validation, testing a range ($[10^{-6}, ..., 10^{4}]$) and selecting the λ yielding the highest average Pearson correlation between predicted and actual EEG signals (Zion Golumbic et al., 2013). The resulting TRF waveforms represent how the EEG signal at each electrode changes in response to a unit change in the speech stimulus envelope. Pearson correlation coefficient r between the predicted and recorded EEG signals was also calculated to evaluate the overall strength of neural tracking.

Cluster-based permutation tests. For both auditory (TRF) and semantic (N400) processing, we applied cluster-based permutation tests (CBPT, Maris & Oostenveld, 2007) using the FieldTrip toolbox (Oostenveld et al., 2011). Paired t-tests were conducted at each electrode and time point to assess differences between conditions or groups. To identify candidate clusters, a two-sided threshold of $p < .05$ was applied to the resulting sample-level t-tests and spatiotemporally adjacent significant data points were grouped into clusters. For each cluster, a cluster-level statistic was calculated as the sum of the *t*-values within the cluster. Statistical significance was assessed using a two-sided Monte Carlo permutation test with 1000 random permutations of condition labels. Clusters were considered significant if their cluster-level statistic fell within the top or bottom 2.5% of the permutation distribution, corresponding to an overall corrected alpha level of 0.05. For tests conducted separately across group and condition, Bonferroni correction was applied to control for multiple comparisons.

This non-parametric approach is especially useful for identifying spatiotemporally extended effects without imposing strong a priori constraints on when or where such effects might occur. Additionally, we complemented CBPTs with additional latency analyses and targeted statistical testing using linear mixed-effects models, allowing us to quantify amplitude and latency differences and assess interactions between group and condition effects with greater precision.

Latency analysis. Latency detection methods were chosen to match the temporal characteristics of each ERP/TRF component. For early TRF components (P1, N1, P2), which are characterised by sharp, time-locked peaks, we used traditional peak latency detection within predefined windows (Luck, 2005). In contrast, N400 latency was estimated using the fractional area latency (FAL) algorithm implemented in ERPLAB, a method recommended for broader and more variable components to provide robust and reliable estimates of onset latency (Lopez-Calderon & Luck, 2014). Latency windows for each TRF component were defined based on the mean and standard deviation (SD) of observed peaks across participants. Each window was set as mean ± 2 SD to capture approximately 95% of latency variability and was visually validated against grand-averaged waveforms to ensure alignment with observed peak distributions. Within these validated windows, peak latency and amplitude were identified for each participant and condition. The onset latency of the N400 effect was estimated using the fractional area latency (FAL) method, following the guidelines by Lopez-Calderon & Luck (2014). We computed the area under the N400 difference waveform (incongruent minus congruent) within the 200–500 ms time window at posterior midline electrodes (Cz, CPz, Pz), and identified the time point at which 20% of the total area was reached. This measure was calculated separately for each participant and condition.

Statistical analysis

Analyses were conducted in R (version 4.1.2, Posit Team, 2022). Linear mixed-effects models (LMMs) were fitted for TRF and N400 data including component amplitudes, latencies, and Pearson correlation (r). Generalised linear mixed-effects models (GLMMs) were constructed for behavioural accuracy (binary outcome), with the BOBYQA optimiser applied to improve convergence. All models were constructed using the lme4 package (Bates et al., 2015). Models compared performance across background conditions using two masker contrasts: (1) baseline (no maskers) vs. masker conditions (babble, speech) and (2) babble vs. speech maskers.

Fixed effects included group (autistic = 1/2, non-autistic = −1/2), masker (contrast1: babble = 1/3, speech = 1/3, baseline = −2/3; contrast2: babble = 1/2, speech = −1/2, baseline = 0), and their interactions. For ERP models, sentence type (congruent = 1/2, incongruent = −1/2) was included as an additional fixed effect. Model selection followed the recommendations of Barr (2013). Initial models were fitted with a maximal random-effects structure, including random intercepts for participants and by-participant random slopes for within-subject predictors. For the behavioural data, the maximal model also included by-item random effects (random intercepts and slopes). In contrast, TRF and N400 data were grand-averaged across trials for each condition and participant prior to statistical analysis to reduce trial-level noise; therefore, item-level variability was not modelled, and random effects for trials were not included.

When maximal models failed to converge, the random-effects structure was simplified in a stepwise manner: (1) by removing correlations between random effects, and (2) by incrementally adding random slopes to an intercept-only model to identify the most parsimonious structure that captured meaningful variance. As models included group effect as a between-subject factor, random intercepts for participants were retained in all models to account for individual baseline differences. At each step, likelihood ratio tests were used to compare models and retain only random effects that significantly improved model fit. Fixed effects and interactions were tested using likelihood ratio tests by comparing the final model to nested models with specific fixed effects removed. Significant interactions were followed up with simple effects analyses by subsetting the data and refitting the model. Bonferroni correction was applied to control for multiple comparisons, with the alpha level set at .025 for main and interaction effects, and .0125 for simple effects. For effect sizes, partial eta-squared (η_p^2) was computed for each fixed effect in LMMs with values ≥ .01, ≥ .09, and ≥ .25 interpreted as small, medium, and large effects, respectively (Cohen et al., 2013). For GLMMs with binary outcomes, odds ratios were calculated by exponentiating the model coefficients. An OR of 1 indicates no effect, while values farther from 1 (either above or below) reflect stronger effects.

4.3 Results

Behavioural results

Figure 4-1 shows accuracy for both masker contrasts across groups. Overall, both groups performed well on the task, particularly in the baseline condition, where ceiling performance was observed. The GLMM analysis (Table 4-2) revealed significant main effects of both

masker contrasts. Behavioural accuracy was higher in the baseline condition compared to the masker conditions (baseline: M_{NAS} = 97.7%, SD_{NAS} = 14.9%; M_{AS} = 96.9%, SD_{AS} = 17.2%. Maskers: M_{NAS} = 93.3%, SD_{NAS} = 25.1%; M_{AS} = 91.7%, SD_{AS} = 27.7%). Additionally, both groups performed better in the babble condition (M_{NAS} = 94.4%, SD_{NAS} = 23.1%; M_{AS} = 93.3%, SD_{AS} = 24.9%) than the speech condition (M_{NAS} = 92.2%, SD_{NAS} = 26.9%; M_{AS} = 90.0%, SD_{AS} = 30.0%). No significant group effects or interactions were found, indicating comparable accuracy rates across masker conditions.

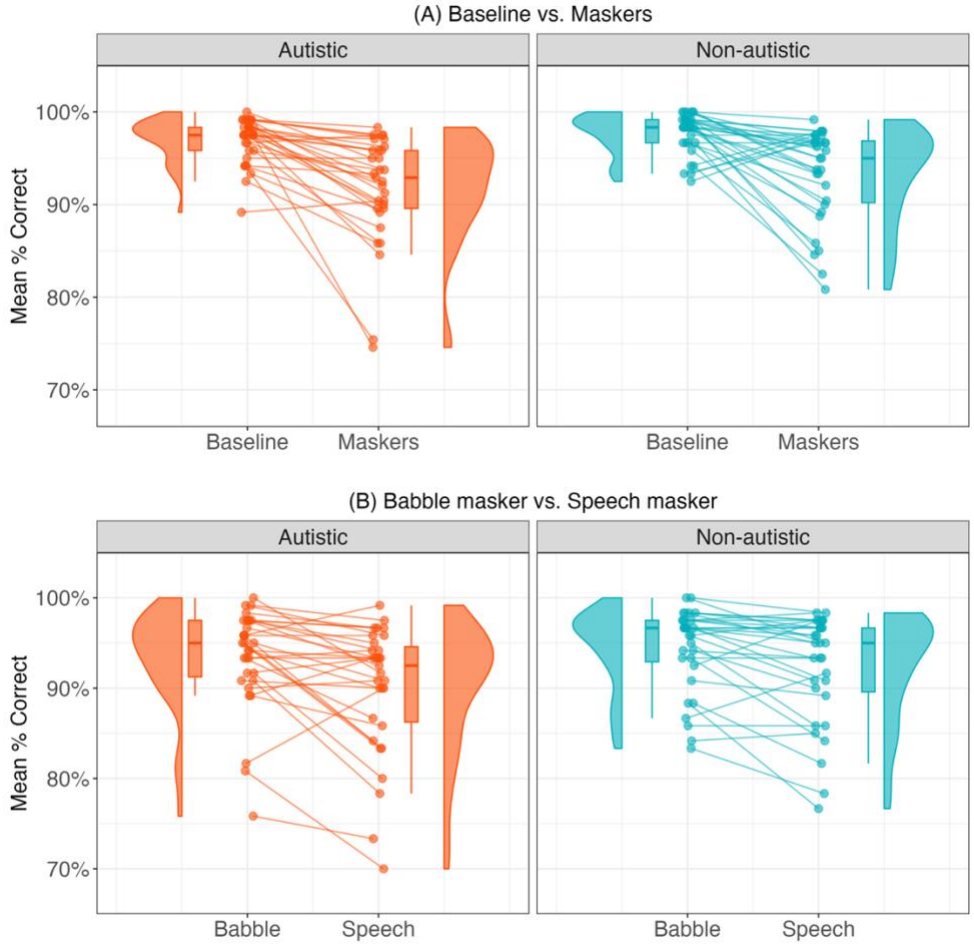

Figure 4-1. Performance accuracy across conditions for autistic and non-autistic groups. (A) Compares baseline to masker conditions (average of babble and speech maskers). (B) Compares babble to speech maskers. Violin plots with embedded box plots show the distribution of mean percentage accuracy, with individual data points connected to illustrate within-subject differences.

Table 4-2. Results of the GLMM for behavioural accuracy.

Fixed effects	*β*	SE	z	χ^2	*p*	OR
(Intercept)	3.51	0.11	31.22	—	—	—
Group	−0.31	0.19	−1.62	2.54	.112	0.73
Masker1	−1.10	0.12	−9.15	58.85	**< .001**	0.33
Masker2	0.46	0.09	5.10	21.56	**< .001**	1.58
Group × Masker1	0.07	0.22	0.32	0.10	.752	1.07
Group × Masker2	0.13	0.16	0.81	0.61	.434	1.14

Note. The *p*-values of significant fixed effects are presented in bold. Model structure: glmer(Accuracy ~ 1 + Group × Masker1 + Group × Masker2 + (1 + Masker1 + Masker2 | Subject) + (1 | Item)). OR: Odds ratios.

TRF results

We conducted cluster-based permutation tests (CBPTs) to identify statistically significant spatiotemporal clusters within a 0–300 ms time window. As CBPTs are limited to pairwise comparisons, we adopted a structured analysis plan to match our theoretical contrasts of interest and to remain as consistent as possible with our follow-up LMMs.

Initially, we explored main effects of group across condition, but no significant clusters emerged. We suspect this may be due to variability in the latency and polarity of the P1–N1–P2 complex across groups and conditions, which can dilute effects when aggregated. Therefore, we performed separate CBPTs within each group and condition and applied Bonferroni correction to account for multiple comparisons (McClannahan et al., 2019). This approach allowed us to better capture condition-specific or group-specific TRF effects without assuming consistent timing or morphology across all comparisons.

For the group effect, three tests were conducted (one per condition), resulting in a corrected alpha of 0.05/3 ≈ .017. For the condition effect, we examined two theoretically motivated contrasts within each group: (1) baseline vs. maskers and (2) babble vs. speech, resulting in four comparisons in total and a corrected alpha of 0.05/4 = .0125. These two contrasts were selected to remain consistent with our LMMs, which were designed to address the same comparisons, rather than testing each condition individually.

Figure 4-2A shows the results of cluster-based permutation tests examining group differences within each condition. The waveforms illustrate the latency, amplitude, and morphology of

TRF components. In the baseline condition, the P1, N1, and P2 peaks in the non-autistic group are clearly identifiable (as marked in the figure), closely resembling traditional auditory evoked potentials (AEPs) in both latency and polarity. In the speech masker condition, a significant cluster was observed between 109 and 172 ms (p = .010), as shown in the topographic map, with activity primarily distributed over fronto-central electrodes. This cluster falls within the expected N1 time window and reflects stronger neural tracking of the speech envelope in the non-autistic group compared to the autistic group.

Figure 4-2B presents the results of condition effects within each group. In the early P1 time window, both groups exhibited significant clusters when comparing baseline to masker conditions (both p-values < .001, 0–125 ms), indicating reduced TRF amplitudes in the presence of background noise. In addition, both groups showed significant clusters in the comparison between babble and speech maskers, with reduced TRF responses in the speech condition. For the non-autistic group, the cluster spanned 0–109 ms (p < .001), while for the autistic group, the cluster was observed from 31–94 ms (p < .001), both falling within the P1 response window.

In the later N1–P2 time range, a significant cluster was found in the non-autistic group for the baseline vs. masker contrast between 156–250 ms (p < .001), suggesting reduced auditory cortical responses in noisy compared to quiet conditions. This implies that neural tracking of the speech envelope was more robust in the absence of background noise for non-autistic participants. No corresponding effect was observed in the autistic group, indicating a lack of measurable differentiation between quiet and noisy conditions. For the babble vs. speech contrast, a significant cluster in the non-autistic group was observed between 172–234 ms (p = .002), reflecting stronger TRF responses in the babble condition. No significant differences were found in the autistic group, suggesting comparable auditory tracking responses across masker types.

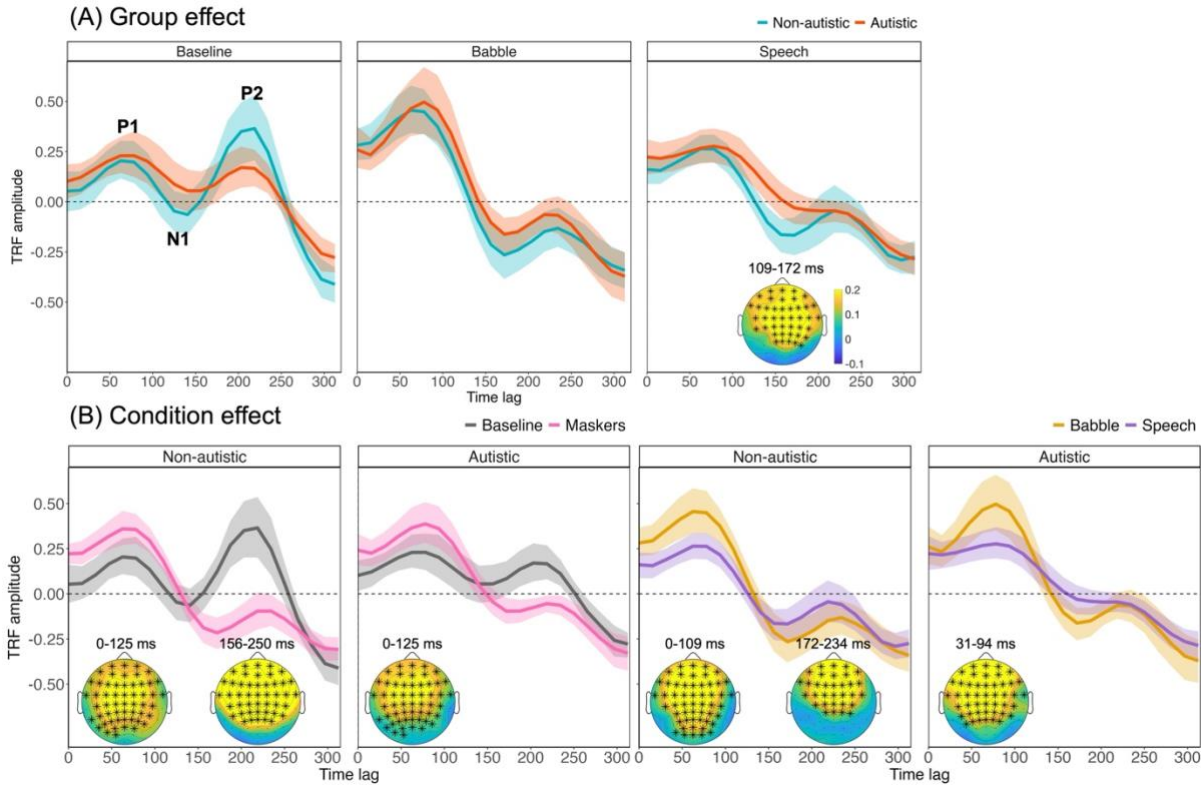

Figure 4-2. Results of the cluster-based permutation tests for TRF group and condition effects. Each panel includes line plots showing mean TRF waveforms, and topographic maps highlighting scalp regions and time windows where significant clusters were identified. Asterisks indicate the scalp locations of these clusters, with the corresponding time windows labelled next to each map. The maps reflect the absolute amplitude differences (in μV) between the compared groups or conditions, with the colour scale indicating the magnitude of the differences. (A) Group comparisons between autistic and non-autistic participants within each listening condition (baseline, babble, and speech). Approximate peaks of the P1, N1, and P2 components are labelled in the baseline waveform for reference. (B) Condition comparisons within each group. Left panels compare baseline to masker conditions (babble and speech combined); right panels compare babble to speech.

LMMs

Since CBPTs could not capture specific TRF components and latency variability, LMMs were conducted to examine P1, N1, and P2 responses separately, focusing on fronto-central electrodes (AFz, Fz, F1, F2, F3, F4, FCz, FC1, FC2, FC3, FC4, Cz, C1, C2, C3, C4) (Muncke et al., 2022). This approach allowed for a systematic interpretation of how individual TRF components drive the observed differences, providing more precise insights into auditory processing mechanisms. Peak amplitudes and latencies were examined for the P1 and N1 components. For the N1 component, a "larger" response indicates a more negative deflection, reflecting stronger neural activation. For the P2 component, only amplitude was analysed, as the peak was not reliably distinguishable across conditions and therefore unsuitable for latency analysis. Additionally, because P2 exhibited negative polarity in some conditions, we also examined the amplitude difference between P2 and N1 (P2 minus N1) as a more reliable index

of later auditory processing (e.g., Beauducel et al., 2000). Full results are reported in Table 4-3. Meanwhile, the Pearson correlation coefficient (r) between actual and predicted EEG signals was also included in the statistical analysis (see Table 4-4 for the results). Box plots for all measured variables are shown in Figure 4-3.

P1 amplitude. There was no significant effect of group or any group × condition interactions. However, both masker contrasts yielded significant main effects. P1 amplitude was reduced in the baseline condition (M = 0.40, SD = 0.47) relative to the masker conditions (M = 0.60, SD = 0.56). Within the masker conditions, babble noise (M = 0.78, SD = 0.65) elicited significantly greater P1 amplitudes than speech maskers (M = 0.43, SD = 0.37).

P1 Latency. No significant main effects or interactions emerged for P1 latency.

N1 Amplitude. A significant group effect was found, with non-autistic participants (M = −0.40, SD = 0.49) showing stronger (more negative) N1 responses than autistic participants (M = −0.24, SD = 0.43). Additionally, both masker contrasts showed significant main effects. N1 amplitude was stronger in the masker conditions (M = −0.39, SD = 0.46) compared to the baseline condition (M = −0.18, SD = 0.45). Meanwhile, more negative responses were observed in the babble condition (M = −0.50, SD = 0.50) relative to the speech condition (M = −0.28, SD = 0.40). No interactions reached significance.

N1 Latency. Group differences in latency were marginal (p = .078), with autistic participants (M = 174.62, SD = 27.24) showing delayed responses compared to non-autistic participants (M = 166.83, SD = 26.70). A significant main effect of condition was observed, with longer latencies in masker conditions (M = 177.21, SD = 23.73) than in the baseline (M = 157.78, SD = 29.14). No significant interactions were observed.

P2 Amplitude. Significant main effects of both masker contrasts were also detected. The baseline condition (M = 0.56, SD = 0.69) elicited larger amplitudes compared to masker conditions (M = 0.02, SD = 0.40). Between maskers, the speech condition (M = 0.07, SD = 0.40) showed slightly larger responses than babble (M = −0.03, SD = 0.40). However, given the variability in N1 across conditions, this result should be interpreted cautiously. A significant interaction between group and baseline-masker contrast was observed. Post hoc analyses revealed significant baseline-masker differences in both autistic ($\chi^2(1) = 17.73$, $p < .001$) and

non-autistic groups ($\chi^2(1) = 30.05$, $p < .001$). No group differences were found within either the baseline ($\chi^2(1) = 3.34$, $p = .067$) or masker conditions ($\chi^2(1) = 0.18$, $p = .671$).

N1–P2 Amplitude. A marginal group effect was observed ($p = .057$), with stronger N1–P2 responses in the non-autistic group ($M = 0.64$, $SD = 0.82$) compared to the autistic group ($M = 0.39$, $SD = 0.52$). A significant main effect of condition was also present: baseline responses ($M = 0.73$, $SD = 0.86$) were greater than those under masker conditions ($M = 0.41$, $SD = 0.57$). No significant interactions were observed.

Neural Tracking Strength (r). There was a significant difference between the baseline and maskers conditions, with greater *r*-values in the baseline condition ($M = 0.11$, $SD = 0.08$) compared to the masker conditions ($M = 0.09$, $SD = 0.07$). No group differences or interactions were found, indicating comparable tracking strength between groups.

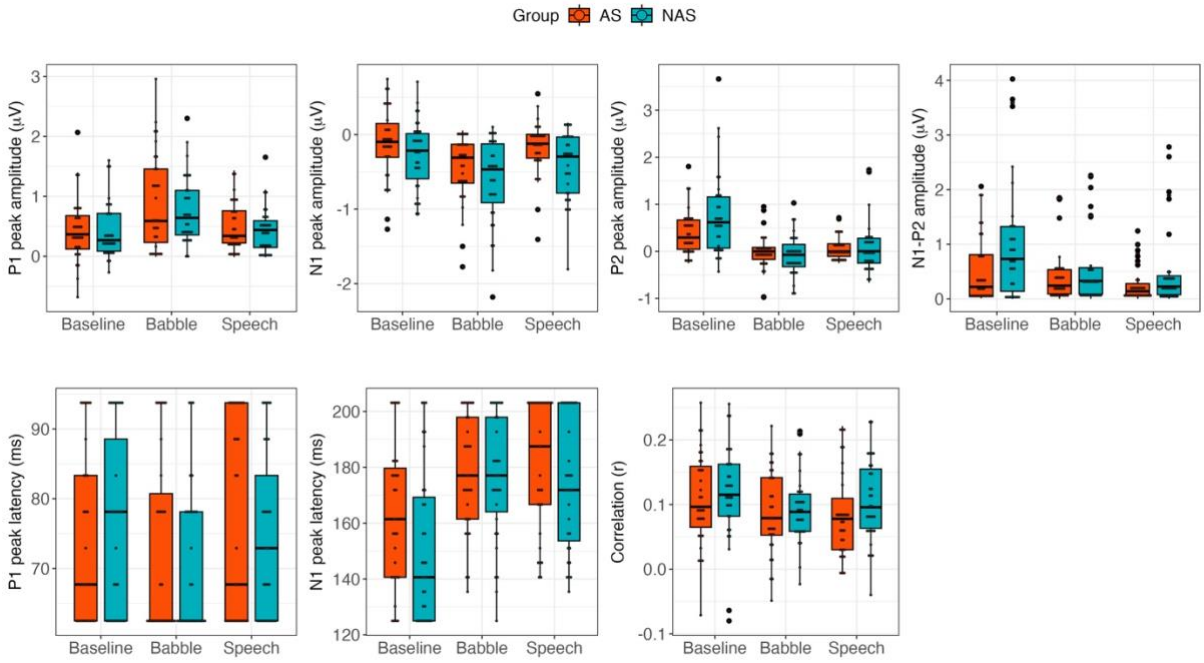

Figure 4-3. TRF component amplitudes, latencies, and model fit across conditions and groups. Boxplots show peak amplitudes and latencies of the TRF P1, N1 components, as well as the amplitude of P2 and N1–P2 (P2 minus N1), and Pearson correlation coefficients (r) between predicted and recorded EEG signals. Data are presented by condition (baseline, babble, speech) and group (AS: Autistic; NAS: Non-autistic).

Table 4-3. Results of the LMM for TRF component amplitudes and latency.

	Fixed effects	***β***	**SE**	**t**	χ^2	***p***	η_p^2
	(Intercept)	0.54	0.05	10.34	—	—	—
	Group	0.06	0.10	0.62	0.38	.537	0.01
P1	Masker1	0.20	0.05	3.86	13.36	**< .001**	0.19
amplitude	Masker2	0.35	0.06	5.80	26.84	**< .001**	0.35
	Group × Masker1	0.08	0.10	0.79	0.61	.433	0.01
	Group × Masker2	0.05	0.12	0.45	0.20	.656	0.00
	(Intercept)	74.17	0.98	75.47	—	—	—
	Group	1.79	1.97	0.91	0.82	.365	0.01
P1	Masker1	−1.95	1.69	−1.15	1.31	.252	0.02
latency	Masker2	−2.82	1.82	−1.55	2.35	.125	0.04
	Group × Masker1	2.72	3.38	0.81	0.65	.421	0.01
	Group × Masker2	-1.73	3.64	−0.48	0.23	.635	0.00
	(Intercept)	−0.32	0.04	−8.31	—	—	—
	Group	0.16	0.08	2.13	4.40	**.036**	0.07
N1	Masker1	−0.21	0.07	−3.26	9.78	**.002**	0.15
amplitude	Masker2	−0.22	0.05	−4.13	15.05	**< .001**	0.22
	Group × Masker1	0.04	0.13	0.31	0.09	.759	0.00
	Group × Masker2	−0.04	0.11	−0.42	0.17	.678	0.00
	(Intercept)	170.73	2.17	78.53	—	—	—
	Group	7.79	4.35	1.79	3.13	.077	0.05
N1	Masker1	19.43	2.98	6.52	32.33	**< .001**	0.41
latency	Masker2	−1.21	3.01	−0.40	0.16	.687	0.00
	Group × Masker1	-5.75	5.96	−0.96	0.92	.337	0.01
	Group × Masker2	-7.34	6.01	−1.22	1.47	.225	0.02
	(Intercept)	0.20	0.05	3.86	—	—	—
	Group	−0.08	0.10	−0.78	0.60	.439	0.01
P2	Masker1	−0.53	0.06	−8.63	48.95	**< .001**	0.55
amplitude	Masker2	−0.11	0.05	−2.18	4.59	**.032**	0.07
	Group × Masker1	0.35	0.12	2.79	7.35	**.007**	0.11
	Group × Masker2	0.06	0.10	0.66	0.43	.513	0.01
	(Intercept)	0.52	0.06	8.20	—	—	—
N1–P2	Group	−0.24	0.13	−1.93	3.61	.057	0.06
amplitude	Masker1	−0.32	0.09	−3.52	11.27	**< .001**	0.17
	Masker2	0.11	0.07	1.50	2.20	.138	0.03

Group × Masker1	0.31	0.18	1.67	2.73	.098	0.04
Group × Masker2	0.11	0.15	0.73	0.53	.465	0.01

Note. The *p*-values of significant effects are presented in bold. The same model was used for all analyses of amplitude and latency: lmer(Amplitude/Latency ~ 1 + Group × Masker1 + Group × Masker2 + (1 + Masker1 + Masker2 | Subject)).

Table 4-4. Results of the LMM for *r*-values of TRF modelling.

Fixed effects	β	**SE**	**t**	χ^2	p	η_p^2
(Intercept)	0.10	0.01	14.98	—	—	—
Group	−0.01	0.01	−0.85	0.72	.395	0.01
Masker1	−0.02	0.01	−2.51	5.99	**.014**	0.09
Masker2	0.00	0.01	−0.34	0.12	.733	0.00
Group × Masker1	0.00	0.02	−0.30	0.09	.761	0.00
Group × Masker2	0.01	0.02	0.86	0.73	.393	0.01

Note. The *p*-values of significant fixed effects are presented in bold. Model structure: lmer(r-value ~ 1 + Group × Masker1 + Group × Masker2 + (1 + Masker1 + Masker2 | Subject)).

ERP results

Two cluster-based permutation tests were conducted within the same 200–600 ms time window as used in Song et al. (2020), following the procedure described in that study. We first compared responses to incongruent vs. congruent sentences in each group to identify clusters reflecting N400 variation. Significant differences were observed in both groups across all masker conditions (both *p*-values < .001), indicating that both groups showed significant N400 effects. As shown in Figure 4-4, there was a significant cluster across the scalp between 200–600 ms for non-autistic listeners. In contrast, a significant cluster was found between 250 and 600 ms for autistic listeners, suggesting a delayed onset of the N400 response. This was further verified by a statistical analysis of N400 onset latency. The autistic group showed significantly longer latencies than the non-autistic group, indicating delayed semantic processing (see Table 4-5 for results). We then examined the effect of noise conditions on the N400 within each group by comparing N400 amplitudes across two masker contrasts: (1) baseline versus noise maskers, and (2) babble versus speech masker. After applying Bonferroni correction for multiple comparisons, no significant clusters were identified between conditions in either group.

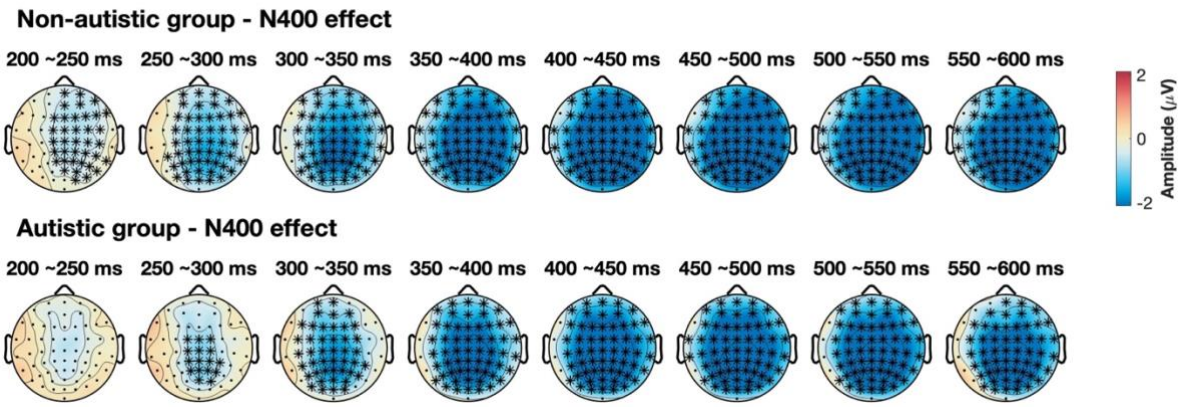

Figure 4-4. Results of the cluster-based permutation test of the N400 effect in each group. Topographic maps display the strength of ERP amplitude difference between incongruent and congruent sentences in 50 ms time bins from 200 to 600 ms for the non-autistic (top row) and autistic (bottom row) groups. Asterisks indicate the time windows and scalp regions where significant clusters were identified. The maps reflect the value of amplitude differences (in μV), with the colour scale indicating the polarity and magnitude of the effect.

Table 4-5. Results of the LMM for N400 onset latency.

Fixed effects	β	SE	z	χ^2	p	η_p^2
(Intercept)	208.16	1.48	141.01	—	—	—
Group	6.33	2.95	2.15	4.44	**.035**	0.07
Masker1	−0.65	2.47	−0.26	0.07	.793	0.00
Masker2	0.15	3.35	0.04	0.00	.965	0.00
Group × Masker1	1.79	4.93	0.36	0.13	.717	0.00
Group × Masker2	−6.87	6.69	−1.03	1.04	.307	0.02

Note. The *p*-values of significant fixed effects are presented in bold. Model structure: lmer (Latency ~ 1 + Group × Masker1 + Group × Masker2 + (1 + Masker1 + Masker2 | Subject)).

Then, LMMs were conducted on N400 amplitudes extracted from a predefined 300–500 ms time window and a predefined midline region of interest, both selected in accordance with Song et al. (2020), to assess between-group differences across conditions (see Figure 4-5 for the results). Mean amplitudes were averaged across five midline electrodes (Fz, FCz, Cz, CPz, and Pz) and entered as the dependent variable. As summarised in Table 4-6, significant three-way interactions were found between group, sentence type, and the two masker contrasts.

To better understand the three-way interactions, we conducted post-hoc analyses focusing on two key comparisons: (1) between-group differences in N400 effect within each masker condition and (2) within-group N400 effect across different masker conditions.

Group × Sentence Interaction. We first examined the group-by-sentence interaction in each condition. No significant interaction between group and sentence type ($\alpha = .025$) was observed in the baseline condition ($\chi^2(1) = 0.07$, $p = .792$), babble condition ($\chi^2(1) = 0.12$, $p = .730$) or speech condition ($\chi^2(1) = 4.51$, $p = .034$). These results indicate that there were no group differences in N400 amplitude in any conditions.

Masker × Sentence Interaction. Next, we examined masker-by-sentence interaction in each group. In the non-autistic group, a significant interaction between sentence type and masker contrast 1 (baseline vs. maskers) was observed ($\chi^2(1) = 770.25$, $p < .001$), while this interaction was marginally significant for autistic participants ($\chi^2(1) = 4.84$, $p = .027$). Simple effects analyses within the non-autistic group revealed significant differences between baseline and masker conditions for both congruent ($\chi^2(1) = 433.15$, $p < .001$) and incongruent sentences ($\chi^2(1) = 350.58$, $p < .001$). Specifically, the congruent condition showed larger amplitudes in the baseline compared to masker conditions, while the incongruent condition showed the opposite, with masker conditions eliciting larger amplitudes than the baseline. This pattern suggests that the presence of maskers amplified the N400 effect in the non-autistic group, a trend not observed in the autistic group. For masker contrast 2 (babble vs. speech), a significant interaction was observed in the non-autistic group ($\chi^2(1) = 582.56$, $p < .001$) but not in the autistic group ($\chi^2(1) = 1.42$, $p = .234$). Follow-up analyses in the non-autistic group showed no significant difference between the babble and speech conditions for congruent sentences ($\chi^2(1) = 5.42$, $p = .020$). However, a significant difference was found for incongruent sentences ($\chi^2(1) = 861.84$, $p < .001$), with the speech condition eliciting larger amplitudes than the babble condition. These findings suggest that while the non-autistic group exhibited a more pronounced N400 effect in the speech condition compared to the babble condition, no similar masker effect was detected in the autistic group.

Overall, there were no significant group differences in the N400 effect for any condition, indicating comparable N400 amplitudes between the autistic and non-autistic groups. However, condition effects were observed only within the non-autistic group. Specifically, they exhibited a significantly larger N400 response in the masker conditions compared to the baseline condition, as well as a significantly larger N400 in the speech condition relative to the babble condition. In contrast, no significant condition effects were observed in the autistic group.

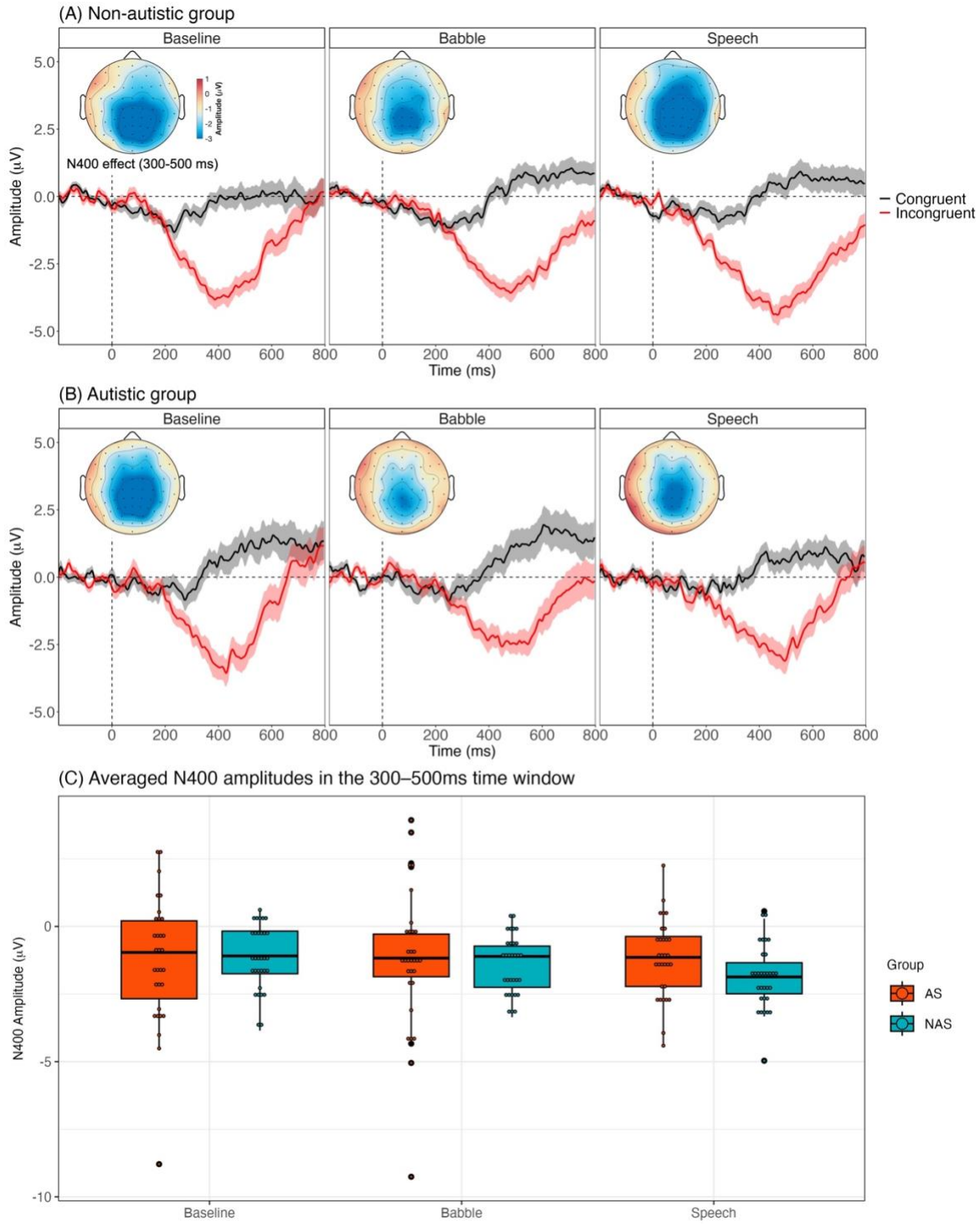

Figure 4-5. N400 amplitudes across groups and masker conditions. (A–B) ERP waveforms for the non-autistic (A) and autistic (B) groups, showing mean amplitudes for congruent (black) and incongruent (red) sentences across baseline, babble, and speech masker conditions. Shaded areas represent ±1 SEM. Topographic maps display the spatial distribution of the N400 effect (incongruent minus congruent) averaged across the 300–500 ms time window. (C) Averaged N400 amplitudes (300–500 ms window) for each condition (baseline, babble, speech) and group (AS: Autistic; NAS: Non-autistic).

Table 4-6. Results of the LMM for N400 amplitudes.

Fixed effects	β	SE	t	χ^2	p	η_p^2
(Intercept)	−0.56	0.11	−4.91	—	—	
Group	0.53	0.23	2.32	5.14	**.023**	0.08
Sentence	1.36	0.17	7.92	43.36	**<.001**	0.50
Masker1	−0.02	0.01	−2.48	6.13	**.013**	0.00
Masker2	0.21	0.01	25.09	628.90	**<.001**	0.00
Group × Masker1	−0.02	0.01	−1.54	2.36	.125	0.00
Group × Masker2	−0.03	0.02	−1.71	2.91	.088	0.00
Group × Sentence	−0.25	0.34	−0.74	0.54	.463	0.01
Masker1 × Sentence	0.23	0.01	16.20	206.28	**<.001**	0.00
Masker2 × Sentence	−0.24	0.02	−14.45	208.77	**<.001**	0.00
Group × Masker1 × Sentence	−0.56	0.03	−19.57	382.80	**<.001**	0.00
Group × Masker2 × Sentence	0.54	0.03	16.24	263.71	**<.001**	0.00

Note. The *p*-values of significant effects are presented in bold. Model structure: lmer(Amplitude ~ 1 + Group × Sentence × Masker1 + Group × Sentence × Masker2 + (1 + Sentence | Subject)).

Correlation

To investigate the relationships among cognitive abilities, neural measures of auditory and semantic processing, and task performance, we conducted Pearson correlation analyses for each group and condition. Each analysis included six individual difference measures, including 1) years of professional musical training; 2) Raven's standard score (nonverbal IQ); 3) ROWPVT percentile (receptive vocabulary); 4) digit span score (working memory); 5) AQ score; and 6) the summed score of auditory attention difficulty and discomfort. Meanwhile, three task-related measures were also examined including behavioural accuracy, TRF amplitude, and N400 amplitude. Behavioural performance was indexed by mean accuracy. Semantic processing was quantified using the mean N400 amplitude between 300–500 ms, averaged across five midline electrodes (Fz, FCz, Cz, CPz, and Pz). Auditory processing was measured by the difference between the P2 and N1 components of the TRF response across fronto-central electrodes. This single TRF index was used instead of separate components for two reasons: to reduce the number of variables in the correlation analysis, and because significant condition and group effects were observed within this time window. In total, nine variables were included in the correlation matrix for each group and condition. To control for multiple comparisons, *p*-values were adjusted using the False Discovery Rate procedure (Benjamini & Hochberg, 1995).

Across all conditions and groups, only two significant correlations emerged (see Figure 4-6). In the baseline condition, a significant negative correlation was found between auditory processing and behavioural performance in the autistic group ($R = –0.58$, $p = .025$). Specifically, autistic participants with larger N1–P2 amplitudes tended to show lower behavioural accuracy in response to semantic incongruency. This relationship was not observed in the non-autistic group. In the speech condition, autistic participants' self-reported auditory attention difficulty and discomfort score was also negatively correlated with behavioural accuracy ($R = –0.58$, $p = .021$), suggesting that those who experience greater auditory challenges in daily life performed more poorly under speech masking. But the relationship was not significant in the non-autistic group.

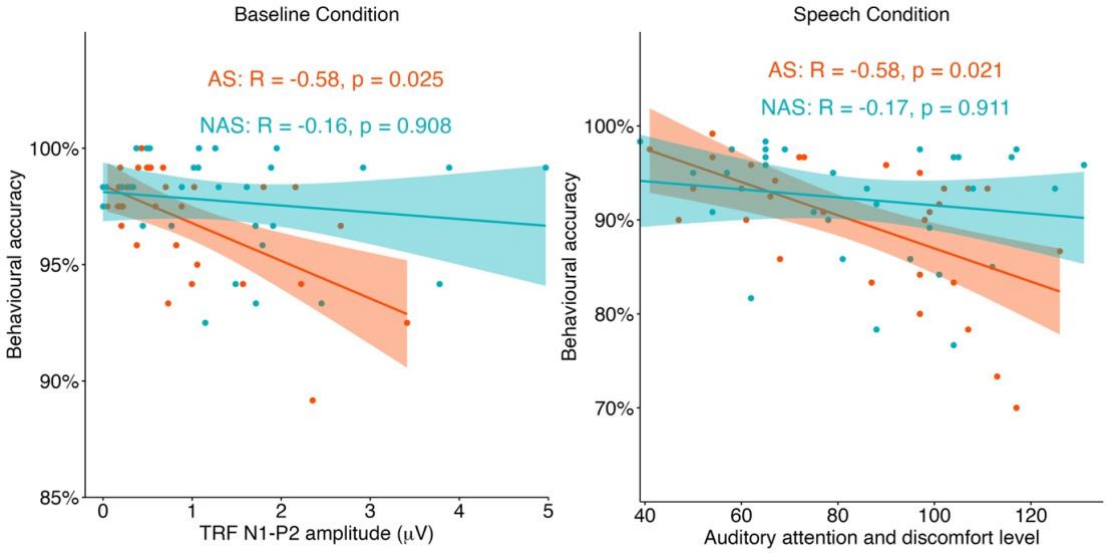

Figure 4-6. Scatter plots of significant correlations. Significant effects were found only in the autistic group (AS) in the baseline and speech conditions; non-autistic group (NAS) data are shown for comparison.

4.4 Discussion

This study examined SiN processing in autism by investigating auditory and semantic mechanisms. Although autistic participants showed similar behavioural accuracy and overall N400 amplitudes in response to semantic violations, they exhibited reduced TRF amplitudes, indicating less robust neural encoding of acoustic information and delayed N400 onset to semantic violations. Moreover, unlike the non-autistic group, whose neural responses reflected a trade-off between auditory and semantic processing based on masker type, autistic participants showed no such modulation.

4.4.1 Masker-modulated SiN processing in non-autistic individuals

To understand how background noise affects auditory processing, we examined TRF components (P1, N1, P2), which reflect stage-specific, time-locked neural responses contributing to speech envelope tracking, along with the *r*-value, which quantifies the overall fidelity of neural tracking by measuring how accurately and consistently the brain follows the speech envelope over time.

Compared to the baseline, masker conditions elicited significantly lower TRF *r*-values. In our forward modelling, this reflects weaker or less consistent neural tracking of speech envelope over time, likely due to interference from competing noise. This pattern is consistent with findings from backward modelling studies reporting lower reconstruction accuracy under noisy conditions (Song et al., 2020). Component-level results offered a more detailed view of processing stages. We found larger P1 amplitudes in masker conditions compared to baseline, consistent with prior studies linking larger P1 responses to degraded speech and louder sounds (Chen et al., 2023; Verschueren et al., 2022). This likely reflects increased demands on early-stage acoustic encoding due to the presence of interfering sounds. During the N1 time window, masker conditions elicited stronger responses (more negative in amplitude) and longer latencies compared to baseline, reflecting increased attentional engagement. This aligns with evidence showing heightened N1 responses in scenarios requiring greater attention, such as multi-speaker environments or vocal music processing (Brown & Bidelman, 2022a; Kong et al., 2014), as well as broader AEP studies linking enhanced N1 amplitude and delayed N1 latency to increased attentional or listening effort (Hillyard et al., 1973). In contrast, P2 amplitudes were reduced in masker conditions relative to baseline. P2 has been widely linked to auditory object formation and speech intelligibility, with larger P2 amplitudes typically associated with better stream segregation and more successful comprehension (Chen et al., 2023; Shinn-Cunningham et al., 2017). Supporting this, studies in complex auditory scenes have shown that robust P2 responses are linked to successful tracking of the attended speech stream, whereas competing speech often elicits alternative components such as N2 (Fiedler et al., 2019). Thus, the reduction in P2 under noise observed in the current study likely reflects increased difficulty in forming a coherent neural representation of the target speech. These findings align with AEP studies that highlight P2 as a crucial component for forming auditory objects in degraded listening conditions, where larger P2 amplitudes have been associated with more successful segregation of the target speech from noise (Näätänen & Picton, 1987; Strauß et al., 2013).

The comparison between babble and speech conditions revealed no significant differences in *r*-values, indicating similar overall neural tracking strength. However, component-level analyses showed attenuated P1 and N1 amplitudes in the speech masker condition relative to babble, suggesting decreased acoustic encoding and reduced attention orientation for speech maskers. Importantly, this does not necessarily indicate that the babble masker imposes greater auditory demands. As highlighted by Song et al. (2020), babble and speech maskers differ across multiple acoustic and linguistic dimensions, complicating direct comparisons. We therefore follow their approach, viewing these effects as the result of an interplay between semantic and auditory-level demands in non-autistic participants.

This interpretation is supported by non-autistic participants' behavioural and N400 results. Consistent with Song et al. (2020), we found a greater decline in behavioural accuracy in the speech masker condition compared to the babble condition. This was accompanied by larger N400 responses to incongruent words in the speech masker condition, indicating increased semantic processing effort due to greater linguistic interference. Taken together, these findings support the idea of a trade-off between auditory and semantic processing: when cognitive resources are increasingly allocated to resolving lexical competition under speech masking, fewer may remain available for early auditory encoding. This may account for the reduced early TRF responses (P1/N1) observed in the speech masker condition alongside enhanced semantic engagement (N400) in non-autistic participants. In conclusion, although different measures of auditory processing were used, our findings in the non-autistic group closely align with those of Song et al. (2020) and demonstrate the interplay between auditory and semantic processing during SiN listening modulated by masker types.

4.4.2 Atypical auditory-semantic processing in autistic individuals

This is the first study to examine neural tracking of acoustic information in autistic individuals during SiN listening. Our TRF analysis revealed a significant group difference in N1 amplitude, with autistic participants showing reduced responses (i.e., less negative amplitude), as well as a marginally reduced N1–P2 amplitude (i.e., P2 minus N1). These findings are consistent with an AEP study conducted by Teder-Sälejärvi et al. (2005), who reported flatter N1 spatial gradients in autistic adults during an auditory localisation task with competing distractors. These results were interpreted as evidence of a reduced ability to sustain auditory attention in noisy environments. Consistent with this interpretation, previous AEP studies in non-autistic listeners have shown that attending to speech in noise enhances N1 and P2 amplitudes and

shortens their latencies (Billings et al., 2011), whereas reduced motivation and increased listening fatigue are associated with attenuated N1 responses (Moore et al., 2017). Accordingly, the smaller N1 and N1–P2 magnitude we observed in autistic participants may similarly reflect diminished attentional engagement. This neural pattern coincides with behavioural differences: autistic participants reported greater difficulties with auditory attention and sensory sensitivity compared to their non-autistic peers (see Table 4-1). Moreover, only in the autistic group did we find a significant relationship between the score of auditory attention difficulty and discomfort (AAD) and behavioural accuracy in the most challenging speech condition. Autistic participants reporting greater everyday auditory challenges (higher AAD scores) performed more poorly when the target speech was presented with competing speech. Together, our findings suggest that background noise may have been more distracting for autistic participants at the acoustic level, making it harder for them to maintain focus and track the target speech stream especially in more challenging scenarios.

An alternative explanation for group differences in auditory responses comes from a study by Lepistö et al. (2009), which reported that autistic participants showed reduced AEP responses only when processing overlapping auditory streams, but not when the streams were presented separately. This suggests that neural differences may emerge specifically under noisy conditions that place high demands on auditory integration. In our study, cluster-based permutation tests across the full P1–N1–P2 time window revealed a group difference around the N1 time window in the speech condition. However, because this analysis is exploratory in nature, we followed up with a more targeted statistical approach using LMMs. The LMM analysis revealed a significant main effect of group on TRF N1 amplitude but did not identify any significant group differences within individual conditions. Thus, while the cluster-based results point to potential group-specific effects under challenging listening conditions, the statistical analysis does not provide strong evidence that these effects are condition-specific. This leaves open the question of whether the observed TRF differences are driven by masker complexity or reflect broader group-level auditory processing differences, even in noise-free conditions. Further research is needed to clarify how auditory stream integration contributes to SiN difficulties in autism.

Despite significant group differences in TRF amplitudes, *r*-values did not differ between groups, indicating similar overall encoding accuracy. At first glance, this may seem inconsistent with findings from Jochaut et al. (2015), who reported reduced cortical tracking

of the speech envelope in autistic individuals under noise-free conditions. However, several key methodological differences likely account for this discrepancy. First, the two studies used different approaches to quantify speech tracking. Jochaut et al. linked fMRI responses to the speech envelope to derive spatial tracking indices, which they then cross-correlated with EEG to assess theta-band dynamics. In contrast, our study employed forward modelling to estimate TRFs, with *r*-values reflecting how accurately the speech envelope predicts EEG responses in the time domain. Second, the experimental paradigms differed: Jochaut et al. used naturalistic, paragraph-length speech in a passive listening task, while our paradigm involved short, semantically manipulated sentences, likely placing lower demands on continuous tracking. Finally, the participant samples varied: the autistic group in Jochaut et al.'s study showed greater variability and generally lower IQ and language abilities, whereas our groups were more closely matched. Overall, our results suggested that while both groups track the speech envelope with comparable precision, they differ in the strength and temporal dynamics of neural encoding, which points to divergent auditory processing mechanisms.

We also offer new insight into SiN processing in autism by examining semantic processing with the N400 component, which has been largely overlooked in previous SiN research. Unexpectedly, unlike most prior studies that found significantly reduced N400 amplitudes and lower behavioural accuracy in noise-free conditions, our participants showed N400 amplitudes and accuracy comparable to non-autistic individuals across all the conditions. This is consistent with research demonstrating intact N400 responses to linguistic semantics in autistic adults, despite differences in experimental paradigms (Coderre et al., 2017; O'Rourke & Coderre, 2021). One possible explanation for the absence of group difference in N400 amplitudes is the close matching of verbal and cognitive abilities across groups in the current study, which helped control for potential confounding factors that may have influenced results in previous studies (DiStefano et al., 2019; McCleery et al., 2010). Our findings suggest that previously reported N400 differences in autism observed even under less challenging, noise-free conditions may be largely driven by individual differences in language ability, rather than reflecting a general deficit in semantic processing.

The absence of a group effect may also be attributable to the simplified task design, which reduced semantic demands by manipulating only the final word's congruency in each sentence. The task's predictability may have enabled autistic participants to rely on prior context to anticipate the incongruent word, rather than engaging in deeper semantic processing. This

likely contributed to the near-ceiling behavioural performance observed in both groups, particularly in the baseline condition. As a result, our task might not be sufficiently demanding to detect group differences in semantic processing. Future research could employ more challenging comprehension or decision-making tasks to better capture variability under noisy listening conditions. We also note that the distribution of N400 amplitudes, particularly in the autistic group (Figure 4-5), spanned a wider range and included several extreme values, reflecting greater variability across individuals. This variability could have influenced the observed group patterns. However, to reflect the heterogeneity of the autistic population and maintain transparency, we retained all data points, including outliers, in the analysis.

Importantly, although overall N400 amplitudes were comparable between groups, the cluster-based permutation test revealed a trend toward delayed N400 onset in the autistic group. This delay was accompanied by a more restricted, centrally focused distribution, compared to the broader activation observed in the non-autistic group during the 200–250 ms time window. Follow-up analyses of onset latency confirmed a significant group difference, consistent with patterns reported in previous studies under noise-free conditions (Braeutigam et al., 2008; DiStefano et al., 2019). In the present study, the delayed N400 onset in autistic participants, relative to non-autistic participants, occurred alongside preserved behavioural accuracy and comparable N400 amplitudes. This suggests that autistic individuals may have required slightly more time or cognitive effort to integrate semantic information to achieve similar outcomes. This interpretation is supported by an eye-tracking study, which found increased listening effort in autistic children during speech-in-noise recognition, despite similar accuracy to non-autistic peers (Xu et al., 2024).

In summary, we found significantly reduced TRF N1 responses and delayed N400 onset latency across conditions, yet overall similar N400 amplitudes in the autistic group compared with the non-autistic group. These findings suggest an atypical temporal profile in the auditory-to-semantic processing stream in autism. Specifically, the reduced N1 amplitude may reflect diminished attentional engagement or reduced efficiency in encoding acoustic features of speech, while the preserved N400 amplitude indicates that lexical-semantic integration was ultimately successful. One possibility is that the delayed N400 onset reflects a downstream consequence of atypical early auditory encoding, suggesting that semantic processing was preserved but required more time or effort to compensate for inefficient auditory processing (i.e., reduced TRF N1). Alternatively, as discussed above, the absence of group differences in

N400 amplitude may be partly due to the relatively low semantic complexity of the task. From this perspective, the delayed N400 onset may also reflect inefficient semantic processing that was not fully captured by the current paradigm. This interpretation is further supported by the absence of masker-related modulation effects in the autistic group (see Section 4.4.3), where N400 amplitudes remained similar across conditions despite varying levels of task difficulty. Future research could further clarify the interaction between auditory and semantic processing in autism by systematically varying both acoustic and semantic demands.

4.4.3 The absence of masker-modulation in autistic individuals

Both groups demonstrated higher accuracy in the babble than in the speech masker condition, indicating behavioural sensitivity to task difficulty. However, only the non-autistic group exhibited corresponding neural modulation. Specifically, they adjusted their auditory and semantic responses depending on masker type, suggesting a flexible, compensatory strategy that increased semantic processing in response to intelligible background speech. In contrast, the autistic group showed no such modulation at either the auditory or semantic level, suggesting no neural adjustment to listening difficulty. This was evident not only in comparisons across masker types, but also in their N400 responses between baseline and masker conditions. Even in the easier baseline condition—where behavioural performance was near ceiling and significantly better than in masker conditions—the autistic group showed significant N400 amplitudes similar to the masker conditions. This suggests that they engaged similar levels of semantic processing effort regardless of task difficulty. Such a pattern aligns with previous findings of heightened auditory effort under challenging listening demands in autism (Mamashli et al., 2017, Schelinski & von Kriegstein, 2023). One interpretation is that autistic participants may allocate more effort toward processing semantic congruency, potentially at the cost of reduced capacity for top-down modulation as well as reduced auditory processing. Supporting this, we observed a negative correlation between TRF amplitudes and behavioural accuracy in the baseline condition for the autistic group. Participants with larger auditory responses tended to perform worse behaviourally. Even without background noise, those who showed stronger auditory responses might have fewer cognitive resources available for efficient semantic processing, which resulted in lower behavioural accuracy. However, as we did not observe direct correlations between TRF and N400 amplitudes, the interaction between auditory and semantic processing remains speculative and should be explored further in future studies.

In conclusion, we found that while non-autistic participants flexibly reallocated cognitive resources between acoustic and semantic processing depending on masker type, no such modulation was observed in the autistic group. These findings suggest that autistic individuals process auditory and semantic information differently in noisy environments, likely due to a combination of differences in sensory encoding and reduced top-down control. This interpretation is consistent with that of Alcántara et al. (2004), who attributed difficulties in processing speech-in-noise with temporal dips to a combination of temporal processing impairments and reduced top-down modulation. Although no substantial speech recognition difficulties emerged in our controlled task, such atypical processing patterns may limit autistic individuals' ability to adapt in unpredictable or demanding environments, where effective communication often relies on flexible processing strategies and the integration of bottom-up and top-down information (Başkent & Gaudrain, 2016; Shinn-Cunningham & Best, 2008).

4.4.4 The effect of individual factors on SiN processing

Unlike many previous studies on SiN processing in autism, which often involved smaller samples and did not control for between-group differences in verbal and cognitive abilities (Ruiz Callejo & Boets, 2023), our study matched autistic and non-autistic participants on age, nonverbal IQ, vocabulary, working memory, and musical background. These factors have all been identified in prior research as potential contributors to performance in SiN tasks (Carroll et al., 2016; Gordon-Salant & Cole, 2016; Heinrich & Knight, 2016; Rönnberg et al., 2010). Although this group matching approach may limit the generalisability of our findings to the broader autistic population, it allowed us to minimise potential confounds and examine SiN processing within a more defined subgroup. Importantly, the presence of group differences even among autistic individuals with typical verbal and cognitive abilities suggests that their SiN difficulties are not solely due to general language or cognitive abilities but may instead reflect differences in listening strategies or processing patterns under varying conditions.

The only unmatched factor between groups was auditory attention and discomfort (AAD) scores, with the autistic group reporting significantly higher levels of attention difficulties and noise sensitivity. To assess whether this group difference in AAD scores influenced neural or behavioural responses, we conducted complementary (G)LMM analyses for behavioural accuracy, TRF measures and the N400 component. In these models, the AAD score was included as a covariate, while the fixed and random effects structures remained identical to those used in the main analyses. For both the TRF and N400 models, there were no significant

main effects of AAD, and the inclusion of AAD scores did not alter the observed group effects or group-by-condition interactions (see Appendix C for details). This indicates that group-level differences in AAD scores did not substantially influence neural responses. In contrast, the behavioural accuracy model revealed a significant main effect of AAD: participants with higher AAD scores showed lower accuracy. However, the group effect remained non-significant, consistent with the original model without the AAD covariate. This suggests that individual differences in auditory attention and discomfort may contribute to variability in behavioural performance, independent of diagnostic group. Additionally, within the autistic group, we found a significant negative correlation between AAD scores and behavioural accuracy in the speech masker condition, indicating that autistic participants with higher AAD scores tended to perform worse in the most challenging listening condition. These findings suggest that while AAD scores do not explain the group-level neural differences, they may contribute to individual variation in behavioural performance, particularly among autistic individuals in difficult listening conditions.

It should be noted that, although our sample size falls within—or even exceeds—the typical range reported in EEG research (see Clayson et al., 2019 for a discussion), it may still be underpowered to detect subtle group effects given the small effect sizes observed for group-related differences in both the TRF and N400 data. Future studies with larger and more diverse samples of autistic individuals will be necessary to better characterise the mechanisms underlying speech-in-noise processing in autism.

4.5 Conclusion

This is the first EEG study to examine both auditory- and semantic-level processing of SiN signals in autistic individuals combining neural tracking measures and N400. The findings highlight distinct auditory and semantic processing between autistic and non-autistic adults during SiN tasks. Despite similar behavioural accuracy and N400 amplitude, autistic participants showed reduced neural encoding of auditory information, delayed semantic processing, and a lack of modulation by masker type, suggesting differences in processing efficiency and flexibility across multiple levels. These findings contribute to a deeper understanding of SiN processing in autism. Future research could build on these insights to develop strategies that support autistic individuals in noisy social settings, enhancing communication and inclusion.

Chapter 5

General Discussion

Speech-in-noise (SiN) processing represents a persistent challenge for autistic individuals in everyday communication. While this difficulty has been consistently documented in both experimental tasks (Ruiz Callejo & Boets, 2023) and self-reports (Bendo et al., 2024; Dunlop et al., 2016), the underlying mechanisms remain insufficiently understood, partly because previous studies have relied primarily on behavioural measures of mean performance. Such designs obscure the dynamic adjustments listeners make over time and yield only limited insight into the neural mechanisms that support comprehension. Neurophysiological research has helped to address this gap, but its scope has also been restricted. Most ERP studies have focused on early auditory responses to simple, decontextualised stimuli such as tones or syllables (Lepistö et al., 2009; Russo et al., 2009; Teder-Sälejärvi et al., 2005), providing little understanding of how autistic listeners engage with continuous, naturalistic speech or higher-level semantic processing.

To address these limitations, this thesis investigated auditory and semantic mechanisms of speech processing using ecologically valid maskers, including competing speech and music, to better approximate real-world listening environments. To capture dynamic changes in performance over trials, we used Generalised Additive Mixed Models (GAMMs), which allowed us to model trial-level fluctuations and strategy use rather than relying solely on mean accuracy. For neural responses, auditory processing was assessed using temporal response function (TRF) modelling, which provides neurophysiologically interpretable components (P1–N1–P2) while preserving the dynamics of continuous speech processing (Crosse et al., 2016). Semantic processing was examined with N400-based measures to capture sensitivity to meaning under varying contexts (Kuperberg, 2016). Together, these methods offered a more comprehensive account of how autistic and non-autistic individuals process speech in noisy environments.

Study 1 used a behavioural paradigm to test whether speaker-related acoustic cues, namely voice pitch and spatial location, facilitate speech recognition in noise. Both groups benefited from these cues, but autistic participants showed lower overall accuracy and less improvement

across trials. Performance was further modulated by background music, particularly among autistic individuals with stronger local processing tendencies, suggesting that cognitive style influenced susceptibility to musical distraction. Studies 2 and 3 extended this investigation by combining behavioural and EEG measures to examine online speech processing under masking. Study 2 focused on background music and assessed semantic integration through the N400. Autistic participants demonstrated reduced accuracy alongside attenuated and delayed N400 responses, indicating weaker and slower integration of meaning in distracting musical contexts. Study 3 employed competing speech and babble as maskers, revealing that autistic participants not only showed slower semantic integration but also exhibited flatter TRF responses (N1–P2). Crucially, across both Studies 2 and 3, non-autistic participants flexibly adjusted their processing strategies depending on the masker, whereas autistic participants showed little evidence of such modulation.

Overall, this thesis provides novel evidence that autistic difficulties in noisy environments are not restricted to atypical sensory encoding but reflect broader differences in top-down mechanisms involving linguistic and cognitive processing. The following sections discuss these findings in depth, situating them within theoretical frameworks of autism and considering their implications for models of speech processing in naturalistic conditions.

5.1 Atypical auditory and semantic processing in autism

In Study 1, the target and distractor sentences were matched in structure, which minimised semantic interference and placed greater demands on acoustic segregation. Recognition therefore depended primarily on the ability to distinguish speakers based on pitch and spatial location. Autistic participants showed lower overall accuracy than their non-autistic peers, indicating reduced efficiency in functionally deploying acoustic features to segregate the target stream. Since pitch discrimination thresholds were directly measured and found to be matched across groups, these difficulties are unlikely to reflect impaired low-level sensitivity to basic acoustic cues. Rather, they suggest a challenge in applying such cues effectively for stream segregation (Schelinski et al., 2016, 2017; Teder-Sälejärvi et al., 2005).

Study 3 extended the investigation to neural measures of auditory processing using TRFs, which modelled how the brain continuously tracks speech envelope over time and yield interpretable components corresponding to classical ERP markers (Crosse et al., 2016, 2021;

Di Liberto et al., 2018). The early P1, indexing initial sensory registration was largely preserved in autistic participants, indicating intact low-level encoding (Chen et al., 2023; Verschueren et al., 2022). By contrast, the N1, linked to selective encoding of acoustic features and attentional engagement (Hillyard et al., 1973; Kong et al., 2014; Brown & Bidelman, 2022), was significantly reduced in autism, echoing prior findings of attenuated N1 responses under noisy conditions (Teder-Sälejärvi et al., 2005; Lepistö et al., 2009). The subsequent P2 associated with auditory object formation and successful stream segregation (Näätänen & Picton, 1987; Strauß et al., 2013; Shinn-Cunningham et al., 2017) was also slightly attenuated, indicating weaker integration of the speech signal. Together, these findings suggest that while early sensory registration is relatively intact, autistic listeners show diminished attention and integration at later stages of auditory processing, providing a neural counterpart to the poor behavioural performance observed in Study 1.

This thesis also provides novel evidence on semantic processing in autism. Using the N400 as the neural indicator, we were the first to examine how background music and competing speech varied in intelligibility influence semantic processing in autistic listeners. In both Study 2 and Study 3, autistic participants showed attenuated and delayed N400 responses, indicating less efficient use of semantic context to support comprehension. This pattern indicates less effective mapping of word meaning onto sentence context, consistent with prior reports of attenuated N400 effects in autism in the absence of background noise (Braeutigam et al., 2008; Fishman et al., 2011; Pijnacker et al., 2010; Ribeiro et al., 2013). Importantly, whereas many previous studies included autistic participants with lower cognitive or language abilities than their non-autistic counterparts, our groups were carefully matched. Yet, differences in N400 responses persisted, indicating that the observed semantic-processing differences cannot be explained solely by broader cognitive or language profiles. At the behavioural level, autistic participants showed lower accuracy in Study 2 but not in Study 3, underscoring that group differences emerged more consistently in neural responses than in overt performance. This suggests that SiN processing difficulties in autism may not always be detectable at the behavioural level.

In summary, across the three studies, the findings consistently indicate that autistic listeners process both auditory and semantic information less efficiently in noisy environments. These patterns point to difficulties in integrating information across processing domains. In the auditory domain, this may take the form of challenges in combining pitch, spatial, or temporal (i.e., envelope) cues to segregate target speech. In the semantic domain, it may involve a

reduced tendency to integrate word meanings into the broader sentence context. This interpretation closely aligns with the Weak Central Coherence (WCC) framework, which emphasises a reduced tendency to spontaneously integrate local cues into coherent global representations. Predictive Coding accounts offer a complementary perspective, suggesting that less precise top-down predictions could further limit the efficiency of auditory and semantic integration during comprehension (Van de Cruys et al., 2014). In contrast, these findings challenge the Enhanced Perceptual Functioning (EPF) account, which predicts heightened perceptual sensitivity (Mottron et al., 2006). Instead of showing enhanced low-level encoding, the evidence points to intact but not superior early acoustic encoding, together with reduced efficiency at later processing stages.

5.2 Atypical listening strategy in autism

Beyond reduced efficiency in auditory and semantic processing, the results also highlighted differences in how autistic participants adjusted their listening strategies. In this thesis, listening strategy refers to how listeners prioritise and weight available acoustic and semantic cues, how attentional and cognitive resources are engaged, and how these processes are adjusted as listening demands change over time or across contexts (Mattys et al., 2012; Peelle, 2018; Pichora-Fuller et al., 2016). Such adjustments reflect changes in cue weighting during ongoing processing, rather than consciously articulated plans. This distinction is often described in terms of implicit versus explicit processing. Implicit processing refers to relatively automatic, stimulus-driven adjustments that do not require deliberate control, whereas explicit processing involves more intentional regulation via executive mechanisms, including strategic control over attention and working memory (Pichora-Fuller et al., 2016; Peelle, 2018). Because the present studies did not include awareness measures or post-task strategy reports, they cannot determine whether listeners were consciously aware of these adjustments or deliberately selected particular approaches. Accordingly, the findings are interpreted in terms of differences in the degree of flexibility with which processing responded to task demands, rather than explicit strategy selection.

In Study 1, both groups performed well when salient acoustic cues were available, indicating effective target-speaker recognition under favourable listening conditions. When these cues were absent, both groups improved across trials, but autistic participants showed smaller gains in the later stages of the task. This divergence suggests reduced cumulative adaptation in the

most demanding condition, where successful performance relied on the spontaneous recruitment of subtle acoustic information and sustained attention to the target stream over time. Correlation analyses further clarify the resources associated with successful performance. In the non-autistic group, pitch discrimination ability predicted accuracy. Although this association does not imply conscious or explicit strategy use, it suggests that listeners with finer pitch resolution were better able to exploit voice-related information to segregate the target stream, particularly when the task lacked salient acoustic cues. This enhanced sensitivity may have supported more reliable use of subtle vocal differences and/or more robust extraction of voice characteristics under masking. In contrast, autistic participants' accuracy in the no-cue condition was associated with working memory capacity, indicating stronger dependence on controlled cognitive resources when acoustic cues were less accessible. This relationship does not imply atypical explicit strategy selection; rather, it suggests that maintaining and selecting task-relevant information placed greater demands on limited-capacity resources in autism. While controlled memory support may help sustain trial-by-trial performance, it may be less effective in supporting cumulative performance gains across trials. As a result, the autistic participants showed smaller late-stage improvements and a widening performance gap over time. Notably, pitch discrimination thresholds were comparable across groups, indicating that the group difference is unlikely to reflect reduced sensitivity to pitch itself. Instead, it may reflect differences in how pitch information was spontaneously recruited and integrated into stream segregation under higher listening demands.

The strategy differences observed in Study 1 were also evident in Studies 2 and 3, but here they were expressed as differences in how neural responses were modulated by masker characteristics across listening contexts. In Study 2, non-autistic participants showed clear masker-dependent modulation of semantic processing: N400 responses were larger when background music was non-vocal and therefore carried less linguistic interference. In Study 3, non-autistic participants showed coordinated masker-dependent changes across auditory and semantic processing: when the masker carried greater semantic load (intelligible speech), they reduced acoustic tracking while increasing semantic integration. Notably, however, the direction of N400 modulation differed across the two studies. In Study 2, music with intelligible lyrics elicited smaller N400s than instrumental music, whereas in Study 3, competing intelligible speech produced larger N400s than babble. At first glance, these effects appear contradictory, but they can be explained by the different SNR levels applied. As discussed in Chapter 3, severe masking (as in Study 2, −6 dB SNR) likely degraded target input

to the point where prediction mechanisms conferred little benefit, leading to attenuated N400 responses. In contrast, the less adverse SNRs in Study 3 preserved intelligibility, allowing listeners to engage predictive mechanisms more strongly; under these conditions, intelligible maskers increased contextual demands and produced larger N400s.

Across Studies 2 and 3, autistic participants showed comparatively little masker-dependent modulation of neural responses. Their TRF and N400 responses remained relatively stable across quiet, music, babble, and competing speech, indicating reduced sensitivity to changes in the acoustic and linguistic properties of the masker. Within cue-weighting and listening-effort frameworks, effective speech processing under adverse conditions often involves adjusting reliance on acoustic versus semantic information as their usefulness changes with listening demands (Mattys et al., 2012; Peelle, 2018; Pichora-Fuller et al., 2016). From this perspective, the reduced neural differentiation observed in the autistic group is best interpreted as reduced flexibility in online processing adjustment, rather than as evidence of consciously selected listening strategies. This reduced context sensitivity helps explain why autistic participants benefited less from conditions with lower interference, as reflected by their consistently lower accuracy across contexts.

Overall, these findings reveal two dimensions of reduced flexibility in autism. The first is reduced adaptation over time in the more demanding no-cue condition in Study 1. The second is reduced sensitivity to masker characteristics in Studies 2 and 3, reflected in limited modulation of acoustic and semantic neural responses across contexts. These effects point to a listening profile that is less responsive to changing demands, which may in turn explain greater self-reported listening effort and auditory discomfort in everyday noisy environments reported by the autistic participants. From a theoretical perspective, the WCC framework suggests that autistic listeners may focus on local features with less consistent spontaneous integration into broader patterns, which could contribute to the reduced functional use of pitch cues for stream segregation in Study 1 and limited integration of contextual changes in Studies 2 and 3. Predictive Coding accounts offer a complementary interpretation, proposing that perception depends on continuously balancing sensory evidence with top-down expectations, with the weight assigned to each adjusted according to their estimated reliability (Clark, 2013; Friston, 2009). In autism, this balance may be less flexibly regulated due to atypical precision weighting of prediction errors (Lawson et al., 2014; Van de Cruys et al., 2014). As a result, listeners may either rely too heavily on incoming detail or fail to adjust predictions effectively when

conditions change, limiting trial-by-trial updating (Study 1) and context-dependent modulation of processing across masker conditions (Study 2 and Study 3). In sum, WCC and Predictive Coding offer different explanatory angles on the same pattern of results, highlighting why autistic listeners showed reduced flexibility in adapting to changing listening demands over time and across contexts.

5.3 The impact of music

Another distinctive contribution of this thesis is the investigation of background music as a competing auditory source, a domain largely overlooked in autism research. Its complexity as a masker arises from being both acoustically rich and emotionally salient, particularly when it contains vocals or intelligible lyrics. Prior studies suggest that some autistic individuals may show heightened sensitivity or preference for musical features (Heaton, 2009; Bhatara et al., 2013), raising the possibility that music could carry particular salience in this group. Across the two studies that examined background music, its influence on autistic and non-autistic participants revealed both shared and distinct patterns.

In Study 1, instrumental background music was introduced as an additional auditory object. At the group level, performance differences between autistic and non-autistic participants were not significant. However, individual variability was evident: autistic participants with stronger local processing biases showed greater declines in performance in the presence of music. This finding is consistent with the WCC framework, which characterises a bias towards local detail alongside reduced prioritisation of global integration. For these listeners, instrumental music may have reduced the efficiency of forming and maintaining a coherent speech stream, particularly when successful performance depended on integrating cues across time and sources. The relatively neutral emotional content and absence of vocals, combined with a more moderate signal-to-noise ratio, likely reduced the disruptive potential of the music overall, which may explain why effects were confined to individual differences rather than observed at the group level.

Study 2 focused more directly on the role of music by manipulating its content and intelligibility. The background music, consisting of pop songs with rhythmic and emotional qualities, was more engaging and presented at a challenging signal-to-noise ratio of −6 dB. Under these conditions, significant group differences were observed. Autistic participants

showed reduced semantic processing, as reflected in attenuated N400 responses and lower accuracy in meaning judgements. Interestingly, while non-autistic participants were most disrupted by intelligible English lyrics compared to Simlish, autistic participants did not show a consistent difference between the two. One interpretation is that autistic listeners may be less sensitive to the intelligibility of lyrics as a speech-like distractor, treating music more uniformly as a competing auditory stream. At the same time, variability within the autistic group was striking: some participants reported heightened distraction from Simlish, possibly due to attempts to identify its linguistic properties. This underscores the importance of larger-scale studies to capture heterogeneity in autistic responses to vocal and non-vocal music.

The findings from Studies 1 and 2 indicate that the extent to which music disrupts speech processing in autism depends strongly on its characteristics. In Study 1, instrumental music presented at a moderate signal-to-noise ratio was relatively neutral and non-vocal, likely reducing its disruptive potential and explaining why no group-level effects were observed. By contrast, Study 2 employed well-known pop songs that were rhythmically engaging, emotionally salient, and presented at a more challenging −6 dB SNR. Under these conditions, autistic participants showed clear neural and behavioural disruptions, with attenuated N400 responses and reduced accuracy in acceptability judgements.

Reduced orientation to speech may also help explain why music had a stronger impact on autistic individuals in Study 2. Social Motivation Theory (Chevallier et al., 2012) proposes that communicative signals may carry lower intrinsic reward value in autism, reducing spontaneous attention to speech and the engagement of language-related networks (Schwartz et al., 2020; Hernandez et al., 2020). This attentional difference becomes particularly relevant when linguistic demands are high. Whereas Study 1 employed simple and repetitive sentences that required relatively little sustained processing, Study 2 used more naturalistic and less predictable sentences that placed greater demands on attention and semantic integration. Under these conditions, autistic participants may have found it more difficult to maintain focus on the speech signal, which in turn made competing auditory input—such as background music—more disruptive. From this perspective, the critical factor is not only the nature of the music itself but also the level of attentional engagement required by the speech, which together determine the extent to which music functions as a powerful masker.

5.4 Future directions

While this thesis advances understanding of auditory and semantic processing in speech-in-noise (SiN) among autistic and non-autistic individuals, several limitations should be acknowledged, highlighting important directions for future research.

First, methodological constraints in task design and stimulus selection limit the scope of interpretation. To ensure feasibility, task demands were deliberately kept low across the studies, both in the features of the maskers and the target speech, allowing autistic participants to complete the experiments reliably and without undue fatigue. However, this approach yielded generally high accuracy, particularly in the non-autistic group, which may have reduced sensitivity to more subtle group differences. For the maskers, fixed SNRs were established through pilot testing to avoid ceiling effects while maintaining target-speech intelligibility. This choice may have reduced variability, as more challenging or fluctuating SNRs could have amplified group differences and better approximated everyday listening conditions. The listening scenarios were also restricted, involving only two competing speakers and/or a narrow range of musical genres including neutral instrumental music in Study 1 and pop songs from a single artist in Study 2. Future work could build on this by incorporating multiple simultaneous talkers, a wider range of musical genres, or dynamically varying acoustic environments, thereby providing a closer match to the complexity of real-world listening scenes. For the target speech, the semantic paradigm was intentionally kept simple to ensure the task remained manageable for autistic participants while still eliciting reliable N400 responses. Semantic congruency was manipulated only at the sentence-final word, with each context presented twice (once congruent, once incongruent). While this design provided precise experimental control, it may have encouraged participants to focus narrowly on the final word rather than processing broader sentence context. Time constraints also prevented the inclusion of filler sentences that could have reduced predictability and discouraged deeper semantic engagement. To mitigate this limitation, Study 2 complemented the categorical manipulation with a surprisal-based analysis, which provided a more sensitive and context-driven measure of semantic processing. The two approaches converged: incongruent words in the categorical analysis, and less predictable (higher surprisal) words in the continuous analysis, both elicited stronger N400 responses that were modulated by the background music condition. This consistency suggests that participants were meaningfully engaging with sentence context despite the paradigm's simplicity. Nonetheless, future studies should increase demands on semantic processing by

incorporating more varied sentence structures and filler items. Another promising approach would be to use continuous narrative speech, which more closely approximates natural language comprehension.

A second avenue for future research is to deepen theoretical understanding of autistic SiN processing by directly testing competing accounts and extending them to mechanisms not addressed in this thesis. Among the frameworks considered here, WCC and predictive coding offer the most plausible accounts of the present findings, though they emphasise different processes. WCC proposes a detail-focused processing bias, in which local acoustic or linguistic information is prioritised by default and broader contextual integration is less spontaneously engaged, whereas predictive coding points to reduced flexibility in updating expectations and weighting of prior expectations under uncertainty. How to empirically distinguish these accounts remains an open question; this thesis therefore treats both as plausible and highlights their differentiation as an important direction for future research. Other theoretical perspectives may also shed light on the present findings. Social Motivation Theory proposes reduced orientation towards socially salient speech. It was referenced earlier to interpret the stronger impact of music in Study 2, yet it remains to be tested directly to assess its relevance more clearly. Although participants reported everyday listening difficulties, no measures explicitly captured selective attention. Future research could address this gap by incorporating direct indices of attentional allocation, such as comparing neural tracking of attended and unattended streams (Ding & Simon, 2012b; Orf et al., 2023; Power et al., 2012), measuring EEG alpha modulation (X. Wang et al., 2023), or using pupillometry and eye-tracking (Xu et al., 2024). Additionally, future research should test the Neural Complexity Hypothesis through targeted connectivity analyses to clarify whether reduced long-range network integration characterises autism.

Third, participant characteristics also constrain generalisability, underscoring the need for larger and more diverse samples in future work. Our autistic participants were adults with cognitive and language abilities comparable to those of the non-autistic group, which minimised potential confounds but also narrowed the scope of generalisation. This is important because autistic individuals with different cognitive or language profiles may show distinct patterns of speech-in-noise processing (Ruiz Callejo & Boets, 2023). In addition, the sample size, though typical for EEG research, limited the ability to capture the heterogeneity that characterises the autistic population. This was particularly evident in Study 2, where responses

to Simlish versus English lyrics varied substantially across participants. Larger-scale studies could better examine such variability and explore predictors such as sensory sensitivity, cognitive style, or musical experience, thereby clarifying the sources of individual differences.

Finally, future research should also extend to applied contexts by examining how targeted interventions and assistive technologies can support autistic individuals in everyday listening environments. The trial-by-trial gains observed in the most demanding no-cue condition of Study 1 suggest a capacity for learning and strategy refinement, but the reduced flexibility observed across studies indicates that this potential may not be fully realised without structured support. Testing approaches such as cue-focused training and explicit coaching in strategy use (Gohari et al., 2023; Mathews et al., 2024; Schafer et al., 2024) represents one promising direction. Future studies could also evaluate the effectiveness of technologies that enhance the salience of target speech. Remote-microphone systems, for example, have been shown to improve listening in noise, reduce physiological stress, and support broader outcomes in autistic youth (Feldman et al., 2022; Rance et al., 2017; Schafer et al., 2014). Classroom sound-field amplification represents another promising approach, offering an environmental adjustment in educational contexts (Wilson et al., 2021).

5.5 Conclusion

This thesis has provided the first comprehensive investigation of speech-in-noise processing in autism that integrates behavioural and neural measures with ecologically valid stimuli. Across three studies, advanced analyses (GAMMs, temporal response function modelling, and N400 indices) were used to capture how autistic and non-autistic listeners process continuous speech in the presence of competing speech and/or music.

The results show that autistic individuals' difficulties with speech in noise are multifaceted, involving both bottom-up auditory encoding and top-down semantic processes. These difficulties varied systematically with task demands and masker type. In the most challenging no-cue condition of Study 1, autistic participants showed smaller late-trial improvements than their non-autistic peers, indicating reduced adaptation when segregation relied on subtle acoustic cues. In Studies 2 and 3, when masker characteristics shifted the balance between acoustic and semantic demands, autistic participants demonstrated reduced flexibility in reallocating processing resources. Taken together, the findings suggest a limitation in top-down

mechanisms that may hinder the flexible coordination of auditory and semantic processes, thereby constraining speech comprehension under demanding listening conditions.

In conclusion, this thesis advances understanding of speech-in-noise processing in autism by situating listening difficulties within the dynamic interplay of auditory encoding and semantic integration. It illuminates the mechanisms that contribute to reduced comprehension in noisy environments and provides a foundation for future strategies to support autistic individuals' communication in everyday life.

References

Abberton, E., & Fourcin, A. J. (1978). Intonation and Speaker Identification. *Language and Speech*, *21*(4), 305–318. https://doi.org/10.1177/002383097802100405

Ahissar, E., Nagarajan, S., Ahissar, M., Protopapas, A., Mahncke, H., & Merzenich, M. M. (2001). Speech comprehension is correlated with temporal response patterns recorded from auditory cortex. *Proceedings of the National Academy of Sciences*, *98*(23), 13367–13372. https://doi.org/10.1073/pnas.201400998

Alcántara, J. I., Cope, T. E., Cope, W., & Weisblatt, E. J. (2012). Auditory temporal-envelope processing in high-functioning children with Autism Spectrum Disorder. *Neuropsychologia*, *50*(7), 1235–1251. https://doi.org/10.1016/j.neuropsychologia.2012.01.034

Alcántara, J. I., Weisblatt, E. J. L., Moore, B. C. J., & Bolton, P. F. (2004). Speech-in-noise perception in high-functioning individuals with autism or Asperger's syndrome. *Journal of Child Psychology and Psychiatry*, *45*(6), 1107–1114. https://doi.org/10.1111/j.1469-7610.2004.t01-1-00303.x

Allen, K., Carlile, S., & Alais, D. (2008). Contributions of talker characteristics and spatial location to auditory streaming. *The Journal of the Acoustical Society of America*, *123*(3), 1562–1570. https://doi.org/10.1121/1.2831774

Allen, R., Hill, E., & Heaton, P. (2009). `Hath charms to soothe . . .': An exploratory study of how high-functioning adults with ASD experience music. *Autism*, *13*(1), 21–41. https://doi.org/10.1177/1362361307098511

American Psychiatric Association. (2013). *Diagnostic and Statistical Manual of Mental Disorders (DSM-5)*. https://psychiatryonline.org/doi/book/10.1176/appi.books.9780890425596

Anderson, A. H., Carter, M., & Stephenson, J. (2018). Perspectives of University Students with Autism Spectrum Disorder. *Journal of Autism and Developmental Disorders*, *48*(3), 651–665. https://doi.org/10.1007/s10803-017-3257-3

Arbogast, T. L., & Kidd Jr, G. (2000). Evidence for spatial tuning in informational masking using the probe-signal method. *The Journal of the Acoustical Society of America*, *108*(4), 1803–1810. https://pubs.aip.org/asa/jasa/article-abstract/108/4/1803/553528

Assmann, P., & Summerfield, Q. (2004). The Perception of Speech Under Adverse Conditions. In S. Greenberg, W. A. Ainsworth, A. N. Popper, & R. R. Fay (Eds), *Speech Processing in the Auditory System* (pp. 231–308). Springer. https://doi.org/10.1007/0-387-21575-1_5

Aydelott, J., Dick, F., & Mills, D. L. (2006). Effects of acoustic distortion and semantic context on event-related potentials to spoken words. *Psychophysiology*, *43*(5), 454–464. https://doi.org/10.1111/j.1469-8986.2006.00448.x

Baron-Cohen, S., Wheelwright, S., Skinner, R., Martin, J., & Clubley, E. (n.d.). *The Autism-Spectrum Quotient (AQ): Evidence from Asperger Syndrome/High-Functioning Autism, Malesand Females, Scientists and Mathematicians*.

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

Barr, D. J. (2013). Random effects structure for testing interactions in linear mixed-effects models. *Frontiers in Psychology*, *4*. https://doi.org/10.3389/fpsyg.2013.00328

Başkent, D., & Gaudrain, E. (2016). Musician advantage for speech-on-speech perception. *The Journal of the Acoustical Society of America*, *139*(3), EL51–EL56. https://doi.org/10.1121/1.4942628

Bates, D., Maechler, M., Bolker, B., Walker, S., Christensen, R. H. B., Singmann, H., Dai, B., Grothendieck, G., Green, P., & Bolker, M. B. (2015). Package 'lme4'. *Convergence*, *12*(1), 2. http://dk.archive.ubuntu.com/pub/pub/cran/web/packages/lme4/lme4.pdf

Baumann, O., & Belin, P. (2010). Perceptual scaling of voice identity: Common dimensions for different vowels and speakers. *Psychological Research Psychologische Forschung*, *74*(1), 110–120. https://doi.org/10.1007/s00426-008-0185-z

Beauducel, A., Debener, S., Brocke, B., & Kayser, J. (2000). On the Reliability of Augmenting/Reducing. *Journal of Psychophysiology*, *14*(4), 226–240. https://doi.org/10.1027//0269-8803.14.4.226

Belek, B. (2019). Articulating Sensory Sensitivity: From Bodies with Autism to Autistic Bodies. *Medical Anthropology*, *38*(1), 30–43. https://doi.org/10.1080/01459740.2018.1460750

Belin, P., & Zatorre, R. J. (2003). Adaptation to speaker's voice in right anterior temporal lobe. *Neuroreport*, *14*(16), 2105–2109. https://journals.lww.com/neuroreport/fulltext/2003/11140/adaptation_to_speaker_s_voice_in_right_anterior.19.aspx

Belmonte, M. K., Allen, G., Beckel-Mitchener, A., Boulanger, L. M., Carper, R. A., & Webb, S. J. (2004). Autism and Abnormal Development of Brain Connectivity. *The Journal of Neuroscience*, *24*(42), 9228–9231. https://doi.org/10.1523/JNEUROSCI.3340-04.2004

Belyk, M., & Brown, S. (2014). Perception of affective and linguistic prosody: An ALE meta-analysis of neuroimaging studies. *Social Cognitive and Affective Neuroscience*, *9*(9), 1395–1403. https://academic.oup.com/scan/article-abstract/9/9/1395/1679321

Bendo, G. J., Sturrock, A., Hanks, G., Plack, C. J., Gowen, E., & Guest, H. (2024). The diversity of speech-perception difficulties among autistic individuals. *Autism & Developmental Language Impairments*, *9*, 23969415241227074. https://doi.org/10.1177/23969415241227074

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

Best, V., Marrone, N., Mason, C. R., & Kidd, G. (2012). The influence of non-spatial factors on measures of spatial release from masking. *The Journal of the Acoustical Society of America*, *131*(4), 3103–3110. https://doi.org/10.1121/1.3693656

Best, V., Shinn-Cunningham, B. G., Ozmeral, E. J., & Kopčo, N. (2010). Exploring the benefit of auditory spatial continuity. *The Journal of the Acoustical Society of America*, *127*(6), EL258–EL264. https://pubs.aip.org/asa/jasa/article-abstract/127/6/EL258/939992

Bhatara, A., Babikian, T., Laugeson, E., Tachdjian, R., & Sininger, Y. S. (2013). Impaired Timing and Frequency Discrimination in High-functioning Autism Spectrum Disorders. *Journal of Autism and Developmental Disorders*, *43*(10), 2312–2328. https://doi.org/10.1007/s10803-013-1778-y

Bidelman, G. M., & Yoo, J. (2020). Musicians Show Improved Speech Segregation in Competitive, Multi-Talker Cocktail Party Scenarios. *Frontiers in Psychology*, *11*, 1927. https://doi.org/10.3389/fpsyg.2020.01927

Billings, C. J., Bennett, K. O., Molis, M. R., & Leek, M. R. (2011). CORTICAL ENCODING OF SIGNALS IN NOISE: EFFECTS OF STIMULUS TYPE AND RECORDING PARADIGM. *Ear and Hearing*, *32*(1), 53–60. https://doi.org/10.1097/AUD.0b013e3181ec5c46

Blackford, T., Holcomb, P. J., Grainger, J., & Kuperberg, G. R. (2012). A funny thing happened on the way to articulation: N400 attenuation despite behavioral interference in picture naming. *Cognition*, *123*(1), 84–99. https://doi.org/10.1016/j.cognition.2011.12.007

Blank, H., & Davis, M. H. (2016). Prediction errors but not sharpened signals simulate multivoxel fMRI patterns during speech perception. *PLoS Biology*, *14*(11), e1002577. https://journals.plos.org/plosbiology/article?id=10.1371/journal.pbio.1002577

Boersma, P. (2007). Praat: Doing phonetics by computer. *Http://Www.Praat.Org/*. https://cir.nii.ac.jp/crid/1571417125760875008

Boets, B., Verhoeven, J., Wouters, J., & Steyaert, J. (2015). Fragile Spectral and Temporal Auditory Processing in Adolescents with Autism Spectrum Disorder and Early Language Delay. *Journal of Autism and Developmental Disorders*, *45*(6), 1845–1857. https://doi.org/10.1007/s10803-014-2341-1

Bolia, R., Nelson, W., Ericson, M., & Simpson, B. (2000). A speech corpus for multitalker communications research. *The Journal of the Acoustical Society of America*, *107*, 1065–1066. https://doi.org/10.1121/1.428288

Bonnel, A., McAdams, S., Smith, B., Berthiaume, C., Bertone, A., Ciocca, V., Burack, J. A., & Mottron, L. (2010). Enhanced pure-tone pitch discrimination among persons with autism but not Asperger syndrome. *Neuropsychologia*, *48*(9), 2465–2475. https://doi.org/10.1016/j.neuropsychologia.2010.04.020

Bonnel, A., Mottron, L., Peretz, I., Trudel, M., Gallun, E., & Bonnel, A.-M. (2003). Enhanced Pitch Sensitivity in Individuals with Autism: A Signal Detection Analysis. *Journal of Cognitive Neuroscience*, *15*(2), 226–235. https://doi.org/10.1162/089892903321208169

Boothroyd, A., & Nittrouer, S. (1988). Mathematical treatment of context effects in phoneme and word recognition. *The Journal of the Acoustical Society of America*, *84*(1), 101–114. https://doi.org/10.1121/1.396976

Boucher, J., Lewis, V., & Collis, G. (1998). Familiar face and voice matching and recognition in children with autism. *The Journal of Child Psychology and Psychiatry and Allied Disciplines*, *39*(2), 171–181. https://www.cambridge.org/core/journals/journal-of-child-psychology-and-psychiatry-and-allied-disciplines/article/familiar-face-and-voice-matching-and-recognition-in-children-with-autism/C288964B8741ABD34014BDD1FD05E886

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

Braeutigam, S., Swithenby, S. J., & Bailey, A. J. (2008). Contextual integration the unusual way: A magnetoencephalographic study of responses to semantic violation in individuals with autism spectrum disorders. *European Journal of Neuroscience*, *27*(4), 1026–1036. https://doi.org/10.1111/j.1460-9568.2008.06064.x

Bregman, A. (1990). Auditory Scene Analysis: The Perceptual Organization of Sound. In *Journal of The Acoustical Society of America—J ACOUST SOC AMER* (Vol. 95). https://doi.org/10.1121/1.408434

Brignell, A., Morgan, A. T., Woolfenden, S., Klopper, F., May, T., Sarkozy, V., & Williams, K. (2018). A systematic review and meta-analysis of the prognosis of language outcomes for individuals with autism spectrum disorder. *Autism & Developmental Language Impairments*, *3*, 2396941518767610. https://doi.org/10.1177/2396941518767610

Brodbeck, C., & Simon, J. Z. (2020). Continuous speech processing. *Current Opinion in Physiology*, *18*, 25–31. https://doi.org/10.1016/j.cophys.2020.07.014

Broderick, M. P., Anderson, A. J., & Lalor, E. C. (2019). Semantic Context Enhances the Early Auditory Encoding of Natural Speech. *The Journal of Neuroscience*, *39*(38), 7564–7575. https://doi.org/10.1523/JNEUROSCI.0584-19.2019

Bronkhorst, A. W. (2000). The cocktail party phenomenon: A review of research on speech intelligibility in multiple-talker conditions. *Acta Acustica United with Acustica*, *86*(1), 117–128. https://www.ingentaconnect.com/contentone/dav/aaua/2000/00000086/00000001/art00016

Brouwer, S., Akkermans, N., Hendriks, L., Van Uden, H., & Wilms, V. (2022). "Lass frooby noo!" the interference of song lyrics and meaning on speech intelligibility. *Journal of Experimental Psychology: Applied*, *28*(3), 576–588. https://doi.org/10.1037/xap0000368

Brouwer, S., Van Engen, K. J., Calandruccio, L., & Bradlow, A. R. (2012). Linguistic contributions to speech-on-speech masking for native and non-native listeners: Language familiarity and semantic content. *The Journal of the Acoustical Society of America*, *131*(2), 1449–1464. https://doi.org/10.1121/1.3675943

Brown, C. A., & Bacon, S. P. (2010). Fundamental frequency and speech intelligibility in background noise. *Hearing Research*, *266*(1), 52–59. https://doi.org/10.1016/j.heares.2009.08.011

Brown, J. A., & Bidelman, G. M. (2022a). Familiarity of Background Music Modulates the Cortical Tracking of Target Speech at the "Cocktail Party". *Brain Sciences*, *12*(10), 1320. https://doi.org/10.3390/brainsci12101320

Brown, J. A., & Bidelman, G. M. (2022b). Song properties and familiarity affect speech recognition in musical noise. *Psychomusicology: Music, Mind, and Brain*, *32*(1–2), 1–6. https://doi.org/10.1037/pmu0000284

Brown, J. A., & Bidelman, G. M. (2023). Attention, Musicality, and Familiarity Shape Cortical Speech Tracking at the Musical Cocktail Party. *bioRxiv*, 2023.10.28.562773. https://doi.org/10.1101/2023.10.28.562773

Brungart, D. S. (2001). Informational and energetic masking effects in the perception of two simultaneous talkers. *The Journal of the Acoustical Society of America*, *109*(3), 1101–1109. https://doi.org/10.1121/1.1345696

Brungart, D. S., & Simpson, B. D. (2002). The effects of spatial separation in distance on the informational and energetic masking of a nearby speech signal. *The Journal of the Acoustical Society of America*, *112*(2), 664–676. https://doi.org/10.1121/1.1490592

Calandruccio, L., Brouwer, S., Van Engen, K. J., Dhar, S., & Bradlow, A. R. (2013). Masking release due to linguistic and phonetic dissimilarity between the target and masker speech. *American Journal of Audiology*, *22*(1), 157–164. https://doi.org/10.1044/1059-0889(2013/12-0072)

Calandruccio, L., Buss, E., & Bowdrie, K. (2017). Effectiveness of Two-Talker Maskers That Differ in Talker Congruity and Perceptual Similarity to the Target Speech. *Trends in Hearing*, *21*, 2331216517709385. https://doi.org/10.1177/2331216517709385

Calandruccio, L., Dhar, S., & Bradlow, A. R. (2010). Speech-on-speech masking with variable access to the linguistic content of the masker speech. *The Journal of the Acoustical Society of America*, *128*(2), 860–869. https://doi.org/10.1121/1.3458857

Calma-Roddin, N., & Drury, J. E. (2020). Music, Language, and The N400: ERP Interference Patterns Across Cognitive Domains. *Scientific Reports*, *10*(1), 11222. https://doi.org/10.1038/s41598-020-66732-0

Cantiani, C., Choudhury, N. A., Yu, Y. H., Shafer, V. L., Schwartz, R. G., & Benasich, A. A. (2016). From Sensory Perception to Lexical-Semantic Processing: An ERP Study in Non-Verbal Children with Autism. *PLOS ONE*, *11*(8), e0161637. https://doi.org/10.1371/journal.pone.0161637

Cardinale, R. C., Shih, P., Fishman, I., Ford, L. M., & Müller, R.-A. (2013). Pervasive Rightward Asymmetry Shifts of Functional Networks in Autism Spectrum Disorder. *JAMA Psychiatry*, *70*(9), 975–982. https://doi.org/10.1001/jamapsychiatry.2013.382

Čeponienė, R., Lepistö, T., Shestakova, A., Vanhala, R., Alku, P., Näätänen, R., & Yaguchi, K. (2003). Speech–sound-selective auditory impairment in children with autism: They can perceive but do not attend. *Proceedings of the National Academy of Sciences*, *100*(9), 5567–5572. https://doi.org/10.1073/pnas.0835631100

Charlton, R. A., Entecott, T., Belova, E., & Nwaordu, G. (2021). "It feels like holding back something you need to say": Autistic and Non-Autistic Adults accounts of sensory experiences and stimming. *Research in Autism Spectrum Disorders*, *89*, 101864. https://doi.org/10.1016/j.rasd.2021.101864

Chen, F., Zhang, H., Ding, H., Wang, S., Peng, G., & Zhang, Y. (2021). Neural coding of formant-exaggerated speech and nonspeech in children with and without autism spectrum disorders. *Autism Research*, *14*(7), 1357–1374. https://doi.org/10.1002/aur.2509

Chen, Y., Schmidt, F., Keitel, A., Rösch, S., Hauswald, A., & Weisz, N. (2023). Speech intelligibility changes the temporal evolution of neural speech tracking. *NeuroImage*, *268*, 119894. https://doi.org/10.1016/j.neuroimage.2023.119894

Chen, Y., Tang, E., Ding, H., & Zhang, Y. (2022). Auditory Pitch Perception in Autism Spectrum Disorder: A Systematic Review and Meta-Analysis. *Journal of Speech, Language, and Hearing Research*, *65*(12), 4866–4886. https://doi.org/10.1044/2022_jslhr-22-00254

Cheng, S. T. T., Lam, G. Y. H., & To, C. K. S. (2017). Pitch Perception in Tone Language-Speaking Adults With and Without Autism Spectrum Disorders. *I-Perception*, *8*(3), 2041669517711200. https://doi.org/10.1177/2041669517711200

Cherry, E. C. (1953). Some experiments on the recognition of speech, with one and with two ears. *Journal of the Acoustical Society of America*, *25*, 975–979. https://pure.mpg.de/rest/items/item_2309493_5/component/file_2309492/content

Chevallier, C., Kohls, G., Troiani, V., Brodkin, E. S., & Schultz, R. T. (2012). The Social Motivation Theory of Autism. *Trends in Cognitive Sciences*, *16*(4), 231–239. https://doi.org/10.1016/j.tics.2012.02.007

Chowdhury, R., Sharda, M., Foster, N. E. V., Germain, E., Tryfon, A., Doyle-Thomas, K., Anagnostou, E., & Hyde, K. L. (2017). Auditory Pitch Perception in Autism Spectrum Disorder Is Associated With Nonverbal Abilities. *Perception*, *46*(11), 1298–1320. https://doi.org/10.1177/0301006617718715

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences*, *36*(3), 181–204. https://www.cambridge.org/core/journals/behavioral-and-brain-sciences/article/whatever-nextpredictive-brains-situated-agents-and-the-future-of-cognitivescience/33542C736E17E3D1D44E8D03BE5F4CD9

Clayson, P. E., Carbine, K. A., Baldwin, S. A., & Larson, M. J. (2019). Methodological reporting behavior, sample sizes, and statistical power in studies of event-related potentials: Barriers to reproducibility and replicability. *Psychophysiology*, *56*(11), e13437. https://doi.org/10.1111/psyp.13437

Coderre, E. L., Chernenok, M., Gordon, B., & Ledoux, K. (2017). Linguistic and Non-Linguistic Semantic Processing in Individuals with Autism Spectrum Disorders: An ERP Study. *Journal of Autism and Developmental Disorders*, *47*(3), 795–812. https://doi.org/10.1007/s10803-016-2985-0

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2013). *Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences* (3rd edn). Routledge. https://doi.org/10.4324/9780203774441

Cooke, M. (2006). A glimpsing model of speech perception in noise. *The Journal of the Acoustical Society of America*, *119*(3), 1562–1573. https://pubs.aip.org/asa/jasa/article/119/3/1562/849084

Cooke, M., Garcia Lecumberri, M. L., & Barker, J. (2008). The foreign language cocktail party problem: Energetic and informational masking effects in non-native speech perception. *The Journal of the Acoustical Society of America*, *123*(1), 414–427. https://doi.org/10.1121/1.2804952

Corbetta, M., Patel, G., & Shulman, G. L. (2008). The Reorienting System of the Human Brain: From Environment to Theory of Mind. *Neuron*, *58*(3), 306–324. https://doi.org/10.1016/j.neuron.2008.04.017

Crawford, H. J., & Strapp, C. M. (1994). Effects of vocal and instrumental music on visuospatial and verbal performance as moderated by studying preference and personality. *Personality and Individual Differences*, *16*(2), 237–245. https://doi.org/10.1016/0191-8869(94)90162-7

Crosse, M. J., Di Liberto, G. M., Bednar, A., & Lalor, E. C. (2016). The Multivariate Temporal Response Function (mTRF) Toolbox: A MATLAB Toolbox for Relating Neural Signals to Continuous Stimuli. *Frontiers in Human Neuroscience*, *10*. https://doi.org/10.3389/fnhum.2016.00604

Crosse, M. J., Zuk, N. J., Di Liberto, G. M., Nidiffer, A. R., Molholm, S., & Lalor, E. C. (2021). Linear Modeling of Neurophysiological Responses to Speech and Other Continuous Stimuli: Methodological Considerations for Applied Research. *Frontiers in Neuroscience*, *15*, 705621. https://doi.org/10.3389/fnins.2021.705621

Crystal, D. (1969). *Prosodic systems and intonation in English*. https://www.cambridge.org/bt/universitypress/subjects/languages-linguistics/phonetics-and-phonology/prosodic-systems-and-intonation-english

Culling, J. F., Hawley, M. L., & Litovsky, R. Y. (2004). The role of head-induced interaural time and level differences in the speech reception threshold for multiple interfering sound sources. *The Journal of the Acoustical Society of America*, *116*(2), 1057–1065. https://doi.org/10.1121/1.1772396

Culling, J. F., Hodder, K. I., & Toh, C. Y. (2003). Effects of reverberation on perceptual segregation of competing voices. *The Journal of the Acoustical Society of America*, *114*(5), 2871–2876. https://doi.org/10.1121/1.1616922

Culling, J. F., & Mansell, E. R. (2013). Speech intelligibility among modulated and spatially distributed noise sources. *The Journal of the Acoustical Society of America*, *133*(4), 2254–2261. https://doi.org/10.1121/1.4794384

Culling, J. F., & Stone, M. A. (2017). Energetic Masking and Masking Release. In J. C. Middlebrooks, J. Z. Simon, A. N. Popper, & R. R. Fay (Eds), *The Auditory System at the Cocktail Party* (pp. 41–73). Springer International Publishing. https://doi.org/10.1007/978-3-319-51662-2_3

Cullington, H. E., & Zeng, F.-G. (2008). Speech recognition with varying numbers and types of competing talkers by normal-hearing, cochlear-implant, and implant simulation subjects. *The Journal of the Acoustical Society of America*, *123*(1), 450–461. https://doi.org/10.1121/1.2805617

Danesh, A. A., Howery, S., Aazh, H., Kaf, W., & Eshraghi, A. A. (2021). Hyperacusis in Autism Spectrum Disorders. *Audiology Research*, *11*(4), Article 4. https://doi.org/10.3390/audiolres11040049

Darwin, C. J., Brungart, D. S., & Simpson, B. D. (2003). Effects of fundamental frequency and vocal-tract length changes on attention to one of two simultaneous talkers. *The Journal of the Acoustical Society of America*, *114*(5), 2913–2922. https://doi.org/10.1121/1.1616924

Dawson, G., Toth, K., Abbott, R., Osterling, J., Munson, J., Estes, A., & Liaw, J. (2004). Early Social Attention Impairments in Autism: Social Orienting, Joint Attention, and Attention to Distress. *Developmental Psychology*, *40*(2), 271–283. https://doi.org/10.1037/0012-1649.40.2.271

De Groot, A. M., & Smedinga, H. E. (2014). Let the music play!: A short-term but no long-term detrimental effect of vocal background music with familiar language lyrics on foreign language vocabulary learning. *Studies in Second Language Acquisition*, *36*(4), 681–707. https://www.cambridge.org/core/journals/studies-in-second-language-acquisition/article/let-the-music-play/C28D533232A39482804924C26DF247DF

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

DePape, A.-M. R., Hall, G. B. C., Tillmann, B., & Trainor, L. J. (2012). Auditory Processing in High-Functioning Adolescents with Autism Spectrum Disorder. *PLOS ONE*, *7*(9), e44084. https://doi.org/10.1371/journal.pone.0044084

Devaraju, D. S., Kemp, A., Eddins, D. A., Shrivastav, R., Chandrasekaran, B., & Hampton Wray, A. (2021). Effects of Task Demands on Neural Correlates of Acoustic and Semantic Processing in Challenging Listening Conditions. *Journal of Speech, Language, and Hearing Research*, *64*(9), 3697–3706. https://doi.org/10.1044/2021_JSLHR-21-00006

Di Liberto, G. M., Peter, V., Kalashnikova, M., Goswami, U., Burnham, D., & Lalor, E. C. (2018). Atypical cortical entrainment to speech in the right hemisphere underpins phonemic deficits in dyslexia. *NeuroImage*, *175*, 70–79. https://doi.org/10.1016/j.neuroimage.2018.03.072

Di Liberto, G. M., O'Sullivan, J. A., & Lalor, E. C. (2015). Low-Frequency Cortical Entrainment to Speech Reflects Phoneme-Level Processing. *Current Biology*, *25*(19), 2457–2465. https://doi.org/10.1016/j.cub.2015.08.030

Ding, N., Patel, A. D., Chen, L., Butler, H., Luo, C., & Poeppel, D. (2017). Temporal modulations in speech and music. *Neuroscience & Biobehavioral Reviews*, *81*, 181–187. https://doi.org/10.1016/j.neubiorev.2017.02.011

Ding, N., & Simon, J. Z. (2012a). Emergence of neural encoding of auditory objects while listening to competing speakers. *Proceedings of the National Academy of Sciences*, *109*(29), 11854–11859. https://doi.org/10.1073/pnas.1205381109

Ding, N., & Simon, J. Z. (2012b). Neural coding of continuous speech in auditory cortex during monaural and dichotic listening. *Journal of Neurophysiology*, *107*(1), 78–89. https://doi.org/10.1152/jn.00297.2011

Ding, N., & Simon, J. Z. (2013). Adaptive Temporal Encoding Leads to a Background-Insensitive Cortical Representation of Speech. *The Journal of Neuroscience*, *33*(13), 5728–5735. https://doi.org/10.1523/JNEUROSCI.5297-12.2013

Ding, N., & Simon, J. Z. (2014). Cortical entrainment to continuous speech: Functional roles and interpretations. *Frontiers in Human Neuroscience*, *8*. https://doi.org/10.3389/fnhum.2014.00311

DiStefano, C., Senturk, D., & Jeste, S. S. (2019). ERP evidence of semantic processing in children with ASD. *Developmental Cognitive Neuroscience*, *36*, 100640. https://doi.org/10.1016/j.dcn.2019.100640

Doelling, K. B., Arnal, L. H., Ghitza, O., & Poeppel, D. (2014). Acoustic landmarks drive delta–theta oscillations to enable speech comprehension by facilitating perceptual parsing. *NeuroImage*, *85*, 761–768. https://doi.org/10.1016/j.neuroimage.2013.06.035

Dryden, A., Allen, H. A., Henshaw, H., & Heinrich, A. (2017). The Association Between Cognitive Performance and Speech-in-Noise Perception for Adult Listeners: A Systematic Literature Review and Meta-Analysis. *Trends in Hearing*, *21*, 2331216517744675. https://doi.org/10.1177/2331216517744675

Dunlop, W. A., Enticott, P. G., & Rajan, R. (2016). Speech Discrimination Difficulties in High-Functioning Autism Spectrum Disorder Are Likely Independent of Auditory Hypersensitivity. *Frontiers in Human Neuroscience*, *10*. https://doi.org/10.3389/fnhum.2016.00401

Dunn, M. A., & Bates, J. C. (2005). Developmental Change in Neutral Processing of Words by Children with Autism. *Journal of Autism and Developmental Disorders*, *35*(3), 361–376. https://doi.org/10.1007/s10803-005-3304-3

Dunn, M., Vaughan Jr., H., Kreuzer, J., & Kurtzberg, D. (1999). Electrophysiologic Correlates of Semantic Classification in Autistic and Normal Children.

Developmental Neuropsychology, *16*(1), 79–99. https://doi.org/10.1207/s15326942dn160105

Edmonds, B. A., & Culling, J. F. (2006). The spatial unmasking of speech: Evidence for better-ear listening. *The Journal of the Acoustical Society of America*, *120*(3), 1539–1545. https://doi.org/10.1121/1.2228573

Eerola, T., & Vuoskoski, J. K. (2011). A comparison of the discrete and dimensional models of emotion in music. *Psychology of Music*, *39*(1), 18–49. https://doi.org/10.1177/0305735610362821

Emmons, K. A., KC Lee, A., Estes, A., Dager, S., Larson, E., McCloy, D. R., St. John, T., & Lau, B. K. (2022). Auditory Attention Deployment in Young Adults with Autism Spectrum Disorder. *Journal of Autism and Developmental Disorders*, *52*(4), 1752–1761. https://doi.org/10.1007/s10803-021-05076-8

Fadeev, K. A., Romero Reyes, I. V., Goiaeva, D. E., Obukhova, T. S., Ovsiannikova, T. M., Prokofyev, A. O., Rytikova, A. M., Novikov, A. Y., Kozunov, V. V., Stroganova, T. A., & Orekhova, E. V. (2024). Attenuated processing of vowels in the left temporal cortex predicts speech-in-noise perception deficit in children with autism. *Journal of Neurodevelopmental Disorders*, *16*(1), 67. https://doi.org/10.1186/s11689-024-09585-2

Federmeier, K. D. (2007). Thinking ahead: The role and roots of prediction in language comprehension. *Psychophysiology*, *44*(4), 491–505. https://doi.org/10.1111/j.1469-8986.2007.00531.x

Feldman, H., & Friston, K. J. (2010). Attention, uncertainty, and free-energy. *Frontiers in Human Neuroscience*, *4*, 215. https://www.frontiersin.org/articles/10.3389/fnhum.2010.00215/full

Feldman, J. I., Thompson, E., Davis, H., Keceli-Kaysili, B., Dunham, K., Woynaroski, T., Tharpe, A. M., & Picou, E. (2022). Remote Microphone Systems Can Improve Listening-in-Noise Accuracy and Listening Effort for Youth with Autism. *Ear and Hearing*, *43*(2), 436–447. https://doi.org/10.1097/AUD.0000000000001058

Feng, S.-Y., & Bidelman, G. M. (2015). Music familiarity modulates mind wandering during lexical processing. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *37*. https://escholarship.org/uc/item/5hv3n49c

Festen, J. M., & Plomp, R. (1990). Effects of fluctuating noise and interfering speech on the speech-reception threshold for impaired and normal hearing. *The Journal of the Acoustical Society of America*, *88*(4), 1725–1736. https://doi.org/10.1121/1.400247

Fiedler, L., Wöstmann, M., Herbst, S. K., & Obleser, J. (2019). Late cortical tracking of ignored speech facilitates neural selectivity in acoustically challenging conditions. *NeuroImage*, *186*, 33–42. https://doi.org/10.1016/j.neuroimage.2018.10.057

Fishman, I., Yam, A., Bellugi, U., Lincoln, A., & Mills, D. (2011). Contrasting patterns of language-associated brain activity in autism and Williams syndrome. *Social Cognitive and Affective Neuroscience*, *6*(5), 630–638. https://doi.org/10.1093/scan/nsq075

Foss-Feig, J. H., Schauder, K. B., Key, A. P., Wallace, M. T., & Stone, W. L. (2017). Audition-Specific Temporal Processing Deficits Associated with Language Function in Children with Autism Spectrum Disorder. *Autism Research : Official Journal of the International Society for Autism Research*, *10*(11), 1845–1856. https://doi.org/10.1002/aur.1820

Foss-Feig, J. H., Stavropoulos, K. K. M., McPartland, J. C., Wallace, M. T., Stone, W. L., & Key, A. P. (2018). Electrophysiological response during auditory gap detection: Biomarker for sensory and communication alterations in autism spectrum disorder?

Developmental Neuropsychology, *43*(2), 109–122. https://doi.org/10.1080/87565641.2017.1365869

Frank, S. L., Otten, L. J., Galli, G., & Vigliocco, G. (2015). The ERP response to the amount of information conveyed by words in sentences. *Brain and Language*, *140*, 1–11. https://doi.org/10.1016/j.bandl.2014.10.006

Freyman, R. L., Balakrishnan, U., & Helfer, K. S. (2001). Spatial release from informational masking in speech recognition. *The Journal of the Acoustical Society of America*, *109*(5), 2112–2122. https://doi.org/10.1121/1.1354984

Friston, K. (2009). The free-energy principle: A rough guide to the brain? *Trends in Cognitive Sciences*, *13*(7), 293–301. https://www.cell.com/trends/cognitive-sciences/abstract/S1364-6613(09)00117-X

Frith, U. (1989). Autism and "Theory of Mind". In C. Gillberg (Ed.), *Diagnosis and Treatment of Autism* (pp. 33–52). Springer US. https://doi.org/10.1007/978-1-4899-0882-7_4

Frith, U., & Happé, F. (1994). Autism: Beyond "theory of mind". *Cognition*, *50*(1–3), 115–132. https://doi.org/10.1016/0010-0277(94)90024-8

Gaffrey, M. S., Kleinhans, N. M., Haist, F., Akshoomoff, N., Campbell, A., Courchesne, E., & Müller, R.-A. (2007). A typical participation of visual cortex during word processing in autism: An fMRI study of semantic decision. *Neuropsychologia*, *45*(8), 1672–1684. https://doi.org/10.1016/j.neuropsychologia.2007.01.008

Galilee, A., Stefanidou, C., & McCleery, J. P. (2017). Atypical speech versus non-speech detection and discrimination in 4- to 6- yr old children with autism spectrum disorder: An ERP study. *PLOS ONE*, *12*(7), e0181354. https://doi.org/10.1371/journal.pone.0181354

García-Pérez, M. A. (2023). Use and misuse of corrections for multiple testing. *Methods in Psychology*, *8*, 100120. https://doi.org/10.1016/j.metip.2023.100120

Gelbar, N. W., Smith, I., & Reichow, B. (2014). Systematic Review of Articles Describing Experience and Supports of Individuals with Autism Enrolled in College and University Programs. *Journal of Autism and Developmental Disorders*, *44*(10), 2593–2601. https://doi.org/10.1007/s10803-014-2135-5

Ghitza, O. (2011). Linking speech perception and neurophysiology: Speech decoding guided by cascaded oscillators locked to the input rhythm. *Frontiers in Psychology*, *2*, 130. https://www.frontiersin.org/articles/10.3389/fpsyg.2011.00130/full

Ghitza, O., & Greenberg, S. (2009). On the Possible Role of Brain Rhythms in Speech Perception: Intelligibility of Time-Compressed Speech with Periodic and Aperiodic Insertions of Silence. *Phonetica*, *66*(1–2), 113–126. https://doi.org/10.1159/000208934

Gillis, M., Decruy, L., Vanthornhout, J., & Francart, T. (2022). Hearing loss is associated with delayed neural responses to continuous speech. *European Journal of Neuroscience*, *55*(6), 1671–1690. https://doi.org/10.1111/ejn.15644

Giraud, A.-L., & Poeppel, D. (2012). Cortical oscillations and speech processing: Emerging computational principles and operations. *Nature Neuroscience*, *15*(4), 511–517. https://doi.org/10.1038/nn.3063

Gohari, N., Dastgerdi, Z. H., Rouhbakhsh, N., Afshar, S., & Mobini, R. (2023). Training Programs for Improving Speech Perception in Noise: A Review. *Journal of Audiology & Otology*, *27*(1), 1–9. https://doi.org/10.7874/jao.2022.00283

Gomes, E., Pedroso, F. S., & Wagner, M. B. (2008). Hipersensibilidade auditiva no transtorno do espectro autístico. *Pró-Fono Revista de Atualização Científica*, *20*(4), 279–284. https://doi.org/10.1590/s0104-56872008000400013

Gonçalves, A. M., & Monteiro, P. (2023). Autism Spectrum Disorder and auditory sensory alterations: A systematic review on the integrity of cognitive and neuronal functions related to auditory processing. *Journal of Neural Transmission*, *130*(3), 325–408. https://doi.org/10.1007/s00702-023-02595-9

Gordon-Salant, S., & Cole, S. S. (2016). Effects of Age and Working Memory Capacity on Speech Recognition Performance in Noise Among Listeners With Normal Hearing. *Ear & Hearing*, *37*(5), 593–602. https://doi.org/10.1097/AUD.0000000000000316

Griffiths, T. D., & Warren, J. D. (2004). What is an auditory object? *Nature Reviews Neuroscience*, *5*(11), 887–892. https://doi.org/10.1038/nrn1538

Groen, W. B., Van Orsouw, L., Huurne, N. T., Swinkels, S., Van Der Gaag, R.-J., Buitelaar, J. K., & Zwiers, M. P. (2009). Intact Spectral but Abnormal Temporal Processing of Auditory Stimuli in Autism. *Journal of Autism and Developmental Disorders*, *39*(5), 742–750. https://doi.org/10.1007/s10803-008-0682-3

Gross, J., Hoogenboom, N., Thut, G., Schyns, P., Panzeri, S., Belin, P., & Garrod, S. (2013). Speech Rhythms and Multiplexed Oscillatory Sensory Coding in the Human Brain. *PLoS Biology*, *11*(12), e1001752. https://doi.org/10.1371/journal.pbio.1001752

Gutschalk, A., & Dykstra, A. R. (2014). Functional imaging of auditory scene analysis. *Hearing Research*, *307*, 98–110. https://doi.org/10.1016/j.heares.2013.08.003

Hagoort, P. (2005). On Broca, brain, and binding: A new framework. *Trends in Cognitive Sciences*, *9*(9), 416–423. https://doi.org/10.1016/j.tics.2005.07.004

Hagoort, P. (2008). The fractionation of spoken language understanding by measuring electrical and magnetic brain signals. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *363*(1493), 1055–1069. https://doi.org/10.1098/rstb.2007.2159

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

Harris, G. J., Chabris, C. F., Clark, J., Urban, T., Aharon, I., Steele, S., McGrath, L., Condouris, K., & Tager-Flusberg, H. (2006). Brain activation during semantic processing in autism spectrum disorders via functional magnetic resonance imaging. *Brain and Cognition*, *61*(1), 54–68. https://doi.org/10.1016/j.bandc.2005.12.015

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., Hudry, K., Ludlow, A., & Hill, E. (2008). Superior discrimination of speech pitch and its relationship to verbal ability in autism spectrum disorders. *Cognitive Neuropsychology*, *25*(6), 771–782. https://doi.org/10.1080/02643290802336277

Heaton, P., Pring, L., & Hermelin, B. (1999). A pseudo-savant: A case of exceptional musical splinter skills. *Neurocase*, *5*(6), 503–509. https://doi.org/10.1080/13554799908402745

Heaton, P., Williams, K., Cummins, O., & Happé, F. (2008). Autism and pitch processing splinter skills: A group and subgroup analysis. *Autism*, *12*(2), 203–219. https://doi.org/10.1177/1362361307085270

Heinrich, A. (2021). The role of cognition for speech-in-noise perception: Considering individual listening strategies related to aging and hearing loss. *International Journal of Behavioral Development*, *45*(5), 382–388. https://doi.org/10.1177/0165025420914984

Henderson, L. M., Baseler, H. A., Clarke, P. J., Watson, S., & Snowling, M. J. (2011). The N400 effect in children: Relationships with comprehension, vocabulary and decoding. *Brain and Language*, *117*(2), 88–99. https://doi.org/10.1016/j.bandl.2010.12.003

Hernandez, L. M., Green, S. A., Lawrence, K. E., Inada, M., Liu, J., Bookheimer, S. Y., & Dapretto, M. (2020). Social Attention in Autism: Neural Sensitivity to Speech Over Background Noise Predicts Encoding of Social Information. *Frontiers in Psychiatry*, *11*, 343. https://doi.org/10.3389/fpsyt.2020.00343

Hickok, G. (2012). The cortical organization of speech processing: Feedback control and predictive coding the context of a dual-stream model. *Journal of Communication Disorders*, *45*(6), 393–402. https://doi.org/10.1016/j.jcomdis.2012.06.004

Hickok, G., & Poeppel, D. (2007). The cortical organization of speech processing. *Nature Reviews Neuroscience*, *8*(5), 393–402. https://doi.org/10.1038/nrn2113

Hillyard, S. A., Hink, R. F., Schwent, V. L., & Picton, T. W. (1973). Electrical Signs of Selective Attention in the Human Brain. *Science*, *182*(4108), 177–180. https://doi.org/10.1126/science.182.4108.177

Holdgraf, C. R., Rieger, J. W., Micheli, C., Martin, S., Knight, R. T., & Theunissen, F. E. (2017). Encoding and Decoding Models in Cognitive Electrophysiology. *Frontiers in Systems Neuroscience*, *11*, 61. https://doi.org/10.3389/fnsys.2017.00061

Holmes, E., To, G., & Johnsrude, I. S. (2021). How Long Does It Take for a Voice to Become Familiar? Speech Intelligibility and Voice Recognition Are Differentially Sensitive to Voice Training. *Psychological Science*, *32*(6), 903–915. https://doi.org/10.1177/0956797621991137

Howard-Jones, P. A., & Rosen, S. (1993). Uncomodulated glimpsing in '"checkerboard"'noise. *The Journal of the Acoustical Society of America*, *93*(5), 2915–2922. https://pubs.aip.org/asa/jasa/article-abstract/93/5/2915/965049

Hsin, C.-H., Chao, P.-C., & Lee, C.-Y. (2023). Speech comprehension in noisy environments: Evidence from the predictability effects on the N400 and LPC. *Frontiers in Psychology*, *14*. https://doi.org/10.3389/fpsyg.2023.1105346

Huang, D., Yu, L., Wang, X., Fan, Y., Wang, S., & Zhang, Y. (2018). Distinct patterns of discrimination and orienting for temporal processing of speech and nonspeech in Chinese children with autism: An event-related potential study. *European Journal of Neuroscience*, *47*(6), 662–668. https://doi.org/10.1111/ejn.13657

Hudac, C. M., DesChamps, T. D., Arnett, A. B., Cairney, B. E., Ma, R., Webb, S. J., & Bernier, R. A. (2018). Early enhanced processing and delayed habituation to deviance sounds in autism spectrum disorder. *Brain and Cognition*, *123*, 110–119. https://doi.org/10.1016/j.bandc.2018.03.004

Hwang, B. J., Mohamed, M. A., & Brašić, J. R. (2017). Molecular imaging of autism spectrum disorder. *International Review of Psychiatry*, *29*(6), 530–554. https://doi.org/10.1080/09540261.2017.1397606

Ihlefeld, A., & Shinn-Cunningham, B. (2008). Spatial release from energetic and informational masking in a selective speech identification task. *The Journal of the Acoustical Society of America*, *123*(6), 4369–4379. https://doi.org/10.1121/1.2904826

Jamey, K., Foster, N. E. V., Sharda, M., Tuerk, C., Nadig, A., & Hyde, K. L. (2019). Evidence for intact melodic and rhythmic perception in children with Autism Spectrum Disorder. *Research in Autism Spectrum Disorders*, *64*, 1–12. https://doi.org/10.1016/j.rasd.2018.11.013

Jamison, C., Aiken, S. J., Kiefte, M., Newman, A. J., Bance, M., & Sculthorpe-Petley, L. (2016). Preliminary Investigation of the Passively Evoked N400 as a Tool for Estimating Speech-in-Noise Thresholds. *American Journal of Audiology*, *25*(4), 344–358. https://doi.org/10.1044/2016_AJA-15-0080

Janata, P., Tomic, S. T., & Rakowski, S. K. (2007). Characterisation of music-evoked autobiographical memories. *Memory*, *15*(8), 845–860. https://doi.org/10.1080/09658210701734593

Järvinen-Pasley, A., Pasley, J., & Heaton, P. (2008). Is the Linguistic Content of Speech Less Salient than its Perceptual Features in Autism? *Journal of Autism and Developmental Disorders*, *38*(2), 239–248. https://doi.org/10.1007/s10803-007-0386-0

Jiang, J., Liu, F., Wan, X., & Jiang, C. (2015). Perception of Melodic Contour and Intonation in Autism Spectrum Disorder: Evidence From Mandarin Speakers. *Journal of Autism and Developmental Disorders*, *45*(7), 2067–2075. https://doi.org/10.1007/s10803-015-2370-4

Jochaut, D., Lehongre, K., Saitovitch, A., Devauchelle, A.-D., Olasagasti, I., Chabane, N., Zilbovicius, M., & Giraud, A.-L. (2015). Atypical coordination of cortical oscillations in response to speech in autism. *Frontiers in Human Neuroscience*, *9*. https://doi.org/10.3389/fnhum.2015.00171

Jones, M. K., Kraus, N., Bonacina, S., Nicol, T., Otto-Meyer, S., & Roberts, M. Y. (2020). Auditory Processing Differences in Toddlers With Autism Spectrum Disorder. *Journal of Speech, Language, and Hearing Research*, *63*(5), 1608–1617. https://doi.org/10.1044/2020_JSLHR-19-00061

Just, M. A., Keller, T. A., Malave, V. L., Kana, R. K., & Varma, S. (2012). Autism as a neural systems disorder: A theory of frontal-posterior underconnectivity. *Neuroscience and Biobehavioral Reviews*, *36*(4), 1292–1313. https://doi.org/10.1016/j.neubiorev.2012.02.007

Kalas, A. (2012). Joint Attention Responses of Children with Autism Spectrum Disorder to Simple versus Complex Music. *Journal of Music Therapy*, *49*(4), 430–452. https://doi.org/10.1093/jmt/49.4.430

Kalikow, D. N., Stevens, K. N., & Elliott, L. L. (1977). Development of a test of speech intelligibility in noise using sentence materials with controlled word predictability. *The Journal of the Acoustical Society of America*, *61*(5), 1337–1351. https://doi.org/10.1121/1.381436

Kamio, Y., Robins, D., Kelley, E., Swainson, B., & Fein, D. (2007). Atypical Lexical/Semantic Processing in High-Functioning Autism Spectrum Disorders without Early Language Delay. *Journal of Autism and Developmental Disorders*, *37*(6), 1116–1122. https://doi.org/10.1007/s10803-006-0254-3

Kana, R. K., Uddin, L. Q., Kenet, T., Chugani, D., & MÃ¼ller, R.-A. (2014). Brain connectivity in autism. *Frontiers in Human Neuroscience*, *8*. https://doi.org/10.3389/fnhum.2014.00349

Keehn, B., Müller, R.-A., & Townsend, J. (2013). Atypical attentional networks and the emergence of autism. *Neuroscience & Biobehavioral Reviews*, *37*(2), 164–183. https://doi.org/10.1016/j.neubiorev.2012.11.014

Kellogg, E. W. (1939). Reversed Speech. *The Journal of the Acoustical Society of America*, *10*(4), 324–326. https://doi.org/10.1121/1.1915995

Kemp, A., Eddins, D., Shrivastav, R., & Hampton Wray, A. (2019). Effects of Task Difficulty on Neural Processes Underlying Semantics: An Event-Related Potentials Study. *Journal of Speech, Language, and Hearing Research*, *62*(2), 367–386. https://doi.org/10.1044/2018_JSLHR-H-17-0396

Kerlin, J. R., Shahin, A. J., & Miller, L. M. (2010). Attentional Gain Control of Ongoing Cortical Speech Representations in a “Cocktail Party”. *The Journal of Neuroscience*, *30*(2), 620–628. https://doi.org/10.1523/JNEUROSCI.3631-09.2010

Key, A. P., & D’Ambrose Slaboch, K. (2021). Speech Processing in Autism Spectrum Disorder: An Integrative Review of Auditory Neurophysiology Findings. *Journal of*

Speech, Language, and Hearing Research, *64*(11), 4192–4212. https://doi.org/10.1044/2021_JSLHR-20-00738

Khalfa, S., Bruneau, N., Rogé, B., Georgieff, N., Veuillet, E., Adrien, J.-L., Barthélémy, C., & Collet, L. (2004). Increased perception of loudness in autism. *Hearing Research*, *198*(1), 87–92. https://doi.org/10.1016/j.heares.2004.07.006

Kidd, G., Jr., Arbogast, T. L., Mason, C. R., & Gallun, F. J. (2005). The advantage of knowing where to listena). *The Journal of the Acoustical Society of America*, *118*(6), 3804–3815. https://doi.org/10.1121/1.2109187

Kidd, G., Jr., Mason, C. R., & Best, V. (2014). The role of syntax in maintaining the integrity of streams of speech. *The Journal of the Acoustical Society of America*, *135*(2), 766–777. https://doi.org/10.1121/1.4861354

Kidd, G., Jr., Mason, C. R., Swaminathan, J., Roverud, E., Clayton, K. K., & Best, V. (2016). Determining the energetic and informational components of speech-on-speech masking. *The Journal of the Acoustical Society of America*, *140*(1), 132–144. https://doi.org/10.1121/1.4954748

Kidd, G., Mason, C. R., Best, V., & Marrone, N. (2010). Stimulus factors influencing spatial release from speech-on-speech masking. *The Journal of the Acoustical Society of America*, *128*(4), 1965–1978. https://doi.org/10.1121/1.3478781

Kidd, G., Mason, C. R., Brughera, A., & Hartmann, W. M. (2005). The role of reverberation in release from masking due to spatial separation of sources for speech identification. *Acta Acustica United with Acustica*, *91*(3), 526–536. https://www.ingentaconnect.com/contentone/dav/aaua/2005/00000091/00000003/art00014?crawler=true

Kidd, G., Mason, C., Richards, V., & Durlach, N. (2008). Informational Masking. In *Auditory Perception of Sound Sources* (Vol. 29, pp. 143–189). https://doi.org/10.1007/978-0-387-71305-2_6

Kiss, L., & Linnell, K. J. (2021). The effect of preferred background music on task-focus in sustained attention. *Psychological Research*, *85*(6), 2313–2325. https://doi.org/10.1007/s00426-020-01400-6

Klin, A. (1991). Young autistic children's listening preferences in regard to speech: A possible characterization of the symptom of social withdrawal. *Journal of Autism and Developmental Disorders*, *21*(1), 29–42. https://doi.org/10.1007/bf02206995

Knaus, T. A., Silver, A. M., Lindgren, K. A., Hadjikhani, N., & Tager-Flusberg, H. (2008). fMRI activation during a language task in adolescents with ASD. *Journal of the International Neuropsychological Society*, *14*(6), 967–979. https://doi.org/10.1017/s1355617708081216

Koelsch, S., Gunter, T. C., Wittfoth, M., & Sammler, D. (2005). Interaction between Syntax Processing in Language and in Music: An ERP Study. *Journal of Cognitive Neuroscience*, *17*(10), 1565–1577. https://doi.org/10.1162/089892905774597290

Koelsch, S., Kasper, E., Sammler, D., Schulze, K., Gunter, T., & Friederici, A. D. (2004). Music, language and meaning: Brain signatures of semantic processing. *Nature Neuroscience*, *7*(3), 302–307. https://doi.org/10.1038/nn1197

Kong, Y.-Y., Mullangi, A., & Ding, N. (2014). Differential modulation of auditory responses to attended and unattended speech in different listening conditions. *Hearing Research*, *316*, 73–81. https://doi.org/10.1016/j.heares.2014.07.009

Kreitewolf, J., Gaudrain, E., & Von Kriegstein, K. (2014). A neural mechanism for recognizing speech spoken by different speakers. *NeuroImage*, *91*, 375–385. https://doi.org/10.1016/j.neuroimage.2014.01.005

Kuhl, P. K., Coffey-Corina, S., Padden, D., & Dawson, G. (2005). Links between social and linguistic processing of speech in preschool children with autism: Behavioral and electrophysiological measures. *Developmental Science*, *8*(1), F1–F12. https://doi.org/10.1111/j.1467-7687.2004.00384.x

Kujala, T., Aho, E., Lepistö, T., Jansson-Verkasalo, E., Nieminen-von Wendt, T., Von Wendt, L., & Näätänen, R. (2007). Atypical pattern of discriminating sound features in adults with Asperger syndrome as reflected by the mismatch negativity. *Biological Psychology*, *75*(1), 109–114. https://doi.org/10.1016/j.biopsycho.2006.12.007

Kumle, L., Võ, M. L.-H., & Draschkow, D. (2021). Estimating power in (generalized) linear mixed models: An open introduction and tutorial in R. *Behavior Research Methods*, *53*(6), 2528–2543. https://doi.org/10.3758/s13428-021-01546-0

Kuperberg, G. R. (2016). Separate streams or probabilistic inference? What the N400 can tell us about the comprehension of events. *Language, Cognition and Neuroscience*, *31*(5), 602–616. https://doi.org/10.1080/23273798.2015.1130233

Kutas, M., & Federmeier, K. D. (2011). Thirty Years and Counting: Finding Meaning in the N400 Component of the Event-Related Brain Potential (ERP). *Annual Review of Psychology*, *62*(1), 621–647. https://doi.org/10.1146/annurev.psych.093008.131123

Kutas, M., & Hillyard, S. A. (1980). Reading Senseless Sentences: Brain Potentials Reflect Semantic Incongruity. *Science*, *207*(4427), 203–205. https://doi.org/10.1126/science.7350657

Kutas, M., & Hillyard, S. A. (1984). Brain potentials during reading reflect word expectancy and semantic association. *Nature*, *307*(5947), 161–163. https://doi.org/10.1038/307161a0

Kwakye, L. D., Foss-Feig, J. H., Cascio, C. J., Stone, W. L., & Wallace, M. T. (2011). Altered auditory and multisensory temporal processing in autism spectrum disorders.

Frontiers in Integrative Neuroscience, *4*, 129. https://doi.org/10.3389/fnint.2010.00129

Lau, B., Emmons, K. A., Lee, A. K. C., Munson, J., Dager, S. R., & Estes, A. M. (2023). The prevalence and developmental course of auditory processing differences in autistic children. *Autism Research*, *16*(7), 1413–1424. https://doi.org/10.1002/aur.2961

Lau, B. K., Emmons, K., Maddox, R. K., Estes, A., Dager, S., (Astley) Hemingway, S. J., & Lee, A. K. (2022). *The impact of cognitive ability on multitalker speech perception in neurodivergent individuals*. Psychiatry and Clinical Psychology. https://doi.org/10.1101/2022.09.19.22280007

Lau, E. F., Phillips, C., & Poeppel, D. (2008). A cortical network for semantics: (De)constructing the N400. *Nature Reviews Neuroscience*, *9*(12), 920–933. https://doi.org/10.1038/nrn2532

Lavner, Y., Gath, I., & Rosenhouse, J. (2000). The effects of acoustic modifications on the identification of familiar voices speaking isolated vowels. *Speech Communication*, *30*(1), 9–26. https://doi.org/10.1016/S0167-6393(99)00028-X

Lawson, R. P., Aylward, J., Roiser, J. P., & Rees, G. (2018). Adaptation of social and non-social cues to direction in adults with autism spectrum disorder and neurotypical adults with autistic traits. *Developmental Cognitive Neuroscience*, *29*, 108–116. https://doi.org/10.1016/j.dcn.2017.05.001

Lawson, R. P., Rees, G., & Friston, K. J. (2014). An aberrant precision account of autism. *Frontiers in Human Neuroscience*, *8*. https://doi.org/10.3389/fnhum.2014.00302

Lee, A. K. C., Larson, E., Maddox, R. K., & Shinn-Cunningham, B. G. (2014). Using neuroimaging to understand the cortical mechanisms of auditory selective attention. *Hearing Research*, *307*, 111–120. https://doi.org/10.1016/j.heares.2013.06.010

Lenth, R. (2025). *emmeans: Estimated Marginal Means, aka Least-Squares Means* (Version R package version 1.11.2-8) [Computer software]. https://rvlenth.github.io/emmeans/

Lepistö, T., Kuitunen, A., Sussman, E., Saalasti, S., Jansson-Verkasalo, E., Nieminen-von Wendt, T., & Kujala, T. (2009). Auditory stream segregation in children with Asperger syndrome. *Biological Psychology*, *82*(3), 301–307. https://doi.org/10.1016/j.biopsycho.2009.09.004

Lepistö, T., Kujala, T., Vanhala, R., Alku, P., Huotilainen, M., & Näätänen, R. (2005). The discrimination of and orienting to speech and non-speech sounds in children with autism. *Brain Research*, *1066*(1), 147–157. https://doi.org/10.1016/j.brainres.2005.10.052

Lepistö, T., Nieminen-von Wendt, T., von Wendt, L., Näätänen, R., & Kujala, T. (2007). Auditory cortical change detection in adults with Asperger syndrome. *Neuroscience Letters*, *414*(2), 136–140. https://doi.org/10.1016/j.neulet.2006.12.009

Li, J., Sujawal, M., Bernotaite, Z., Cunnings, I., & Liu, F. (2025). Auditory and Semantic Processing of Speech-in-Noise in Autism: A Behavioral and EEG Study. *Autism Research*, *n/a*(n/a). https://doi.org/10.1002/aur.70097

Linke, A. C., Jao Keehn, R. J., Pueschel, E. B., Fishman, I., & Müller, R.-A. (2018). Children with ASD show links between aberrant sound processing, social symptoms, and atypical auditory interhemispheric and thalamocortical functional connectivity. *Developmental Cognitive Neuroscience*, *29*, 117–126. https://doi.org/10.1016/j.dcn.2017.01.007

Liu, F., Patel, A. D., Fourcin, A., & Stewart, L. (2010). Intonation processing in congenital amusia: Discrimination, identification and imitation. *Brain*, *133*(6), 1682–1693. https://doi.org/10.1093/brain/awq089

Lopez-Calderon, J., & Luck, S. J. (2014). ERPLAB: An open-source toolbox for the analysis of event-related potentials. *Frontiers in Human Neuroscience*, *8*. https://doi.org/10.3389/fnhum.2014.00213

Luck, S. J. (n.d.). *Ten Simple Rules for Designing and Interpreting ERP Experiments*.

Luo, H., & Poeppel, D. (2007). Phase Patterns of Neuronal Responses Reliably Discriminate Speech in Human Auditory Cortex. *Neuron*, *54*(6), 1001–1010. https://doi.org/10.1016/j.neuron.2007.06.004

Luo, H., & Poeppel, D. (2012). Cortical Oscillations in Auditory Perception and Speech: Evidence for Two Temporal Windows in Human Auditory Cortex. *Frontiers in Psychology*, *3*. https://doi.org/10.3389/fpsyg.2012.00170

Lyons, M., Schoen Simmons, E., & Paul, R. (2014). Prosodic Development in Middle Childhood and Adolescence in High-Functioning Autism. *Autism Research*, *7*(2), 181–196. https://doi.org/10.1002/aur.1355

Maess, B., Koelsch, S., Gunter, T. C., & Friederici, A. D. (2001). Musical syntax is processed in Broca's area: An MEG study. *Nature Neuroscience*, *4*(5), 540–545. https://doi.org/10.1038/87502

Mamashli, F., Khan, S., Bharadwaj, H., Michmizos, K., Ganesan, S., Garel, K. A., Ali Hashmi, J., Herbert, M. R., Hämäläinen, M., & Kenet, T. (2017). Auditory processing in noise is associated with complex patterns of disrupted functional connectivity in autism spectrum disorder. *Autism Research*, *10*(4), 631–647. https://doi.org/10.1002/aur.1714

Manfredi, M., Cohn, N., Sanchez Mello, P., Fernandez, E., & Boggio, P. S. (2020). Visual and Verbal Narrative Comprehension in Children and Adolescents with Autism Spectrum Disorders: An ERP Study. *Journal of Autism and Developmental Disorders*, *50*(8), 2658–2672. https://doi.org/10.1007/s10803-020-04374-x

Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods*, *164*(1), 177–190. https://doi.org/10.1016/j.jneumeth.2007.03.024

Marrone, N., Mason, C. R., & Kidd, G. (2008). Tuning in the spatial dimension: Evidence from a masked speech identification task. *The Journal of the Acoustical Society of America*, *124*(2), 1146–1158. https://doi.org/10.1121/1.2945710

Martin, N. A., & Brownell, R. (2011). *Expressive one-word picture vocabulary test-4 (EOWPVT-4)*. Academic Therapy Publications.

Mathews, L., Schafer, E. C., Gopal, K. V., Lam, B., & Miller, S. (2024). Speech-in-Noise and Dichotic Auditory Training Students With Autism Spectrum Disorder. *Language, Speech, and Hearing Services in Schools*, *55*(4), 1054–1067. https://doi.org/10.1044/2024_LSHSS-23-00168

Matsuzaki, J., Kagitani-Shimono, K., Sugata, H., Hanaie, R., Nagatani, F., Yamamoto, T., Tachibana, M., Tominaga, K., Hirata, M., Mohri, I., & Taniike, M. (2017). Delayed Mismatch Field Latencies in Autism Spectrum Disorder with Abnormal Auditory Sensitivity: A Magnetoencephalographic Study. *Frontiers in Human Neuroscience*, *11*. https://doi.org/10.3389/fnhum.2017.00446

Mattys, S. L., Brooks, J., & Cooke, M. (2009). Recognizing speech under a processing load: Dissociating energetic from informational factors. *Cognitive Psychology*, *59*(3), 203–243. https://doi.org/10.1016/j.cogpsych.2009.04.001

Mattys, S. L., Davis, M. H., Bradlow, A. R., & Scott, S. K. (2012). Speech recognition in adverse conditions: A review. *Language and Cognitive Processes*, *27*(7–8), 953–978. https://doi.org/10.1080/01690965.2012.705006

McCann, J., Peppé, S., Gibbon, F. E., O'Hare, A., & Rutherford, M. (2007). Prosody and its relationship to language in school-aged children with high-functioning autism.

International Journal of Language & Communication Disorders, *42*(6), 682–702. https://doi.org/10.1080/13682820601170102

McClannahan, K. S., Backer, K. C., & Tremblay, K. L. (2019). Auditory Evoked Responses in Older Adults with Normal Hearing, Untreated, and Treated Age-Related Hearing Loss. *Ear and Hearing*, *40*(5), 1106–1116. https://doi.org/10.1097/AUD.0000000000000698

McCleery, J. P., Ceponiene, R., Burner, K. M., Townsend, J., Kinnear, M., & Schreibman, L. (2010). Neural correlates of verbal and nonverbal semantic integration in children with autism spectrum disorders. *Journal of Child Psychology and Psychiatry*, *51*(3), 277–286. https://doi.org/10.1111/j.1469-7610.2009.02157.x

McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology*, *18*(1), 1–86. https://doi.org/10.1016/0010-0285(86)90015-0

McGettigan, C., Evans, S., Rosen, S., Agnew, Z., Shah, P., & Scott, S. (2012). An application of univariate and multivariate approaches in fMRI to quantifying the hemispheric lateralization of acoustic and linguistic processes. *Journal of Cognitive Neuroscience*, *24*(3), 636–652. https://doi.org/10.1162/jocn_a_00161

McQueen, J. M., & Huettig, F. (2012). Changing only the probability that spoken words will be distorted changes how they are recognized. *The Journal of the Acoustical Society of America*, *131*(1), 509–517. https://doi.org/10.1121/1.3664087

Mesgarani, N., David, S. V., Fritz, J. B., & Shamma, S. A. (2009). Influence of Context and Behavior on Stimulus Reconstruction From Neural Activity in Primary Auditory Cortex. *Journal of Neurophysiology*, *102*(6), 3329–3339. https://doi.org/10.1152/jn.91128.2008

Messaoud-Galusi, S., Hazan, V., & Rosen, S. (2011). Investigating speech perception in children with dyslexia: Is there evidence of a consistent deficit in individuals? *Journal*

of Speech, Language, and Hearing Research : JSLHR, *54*(6), 1682–1701. https://doi.org/10.1044/1092-4388(2011/09-0261)

Michaelov, J. A., Bardolph, M. D., Van Petten, C. K., Bergen, B. K., & Coulson, S. (2024). Strong Prediction: Language Model Surprisal Explains Multiple N400 Effects. *Neurobiology of Language*, *5*(1), 107–135. https://doi.org/10.1162/nol_a_00105

Michaelov, J. A., Coulson, S., & Bergen, B. K. (2023). So Cloze Yet So Far: N400 Amplitude Is Better Predicted by Distributional Information Than Human Predictability Judgements. *IEEE Transactions on Cognitive and Developmental Systems*, *15*(3), 1033–1042. IEEE Transactions on Cognitive and Developmental Systems. https://doi.org/10.1109/TCDS.2022.3176783

Miller, G. A., Heise, G. A., & Lichten, W. (1951). The intelligibility of speech as a function of the context of the test materials. *Journal of Experimental Psychology*, *41*(5), 329–335. https://doi.org/10.1037/h0062491

Miller, G. A., & Licklider, J. C. R. (1950). The intelligibility of interrupted speech. *Journal of the Acoustical Society of America*, *22*, 167–173. https://doi.org/10.1121/1.1906584

Miller, M., Iosif, A.-M., Hill, M., Young, G. S., Schwichtenberg, A. J., & Ozonoff, S. (2017). Response to Name in Infants Developing Autism Spectrum Disorder: A Prospective Study. *The Journal of Pediatrics*, *183*, 141-146.e1. https://doi.org/10.1016/j.jpeds.2016.12.071

Molnar-Szakacs, I., & Heaton, P. (2012). Music: A unique window into the world of autism. *Annals of the New York Academy of Sciences*, *1252*(1), 318–324. https://doi.org/10.1111/j.1749-6632.2012.06465.x

Moore, T. M., Key, A. P., Thelen, A., & Hornsby, B. W. Y. (2017). Neural mechanisms of mental fatigue elicited by sustained auditory processing. *Neuropsychologia*, *106*, 371–382. https://doi.org/10.1016/j.neuropsychologia.2017.10.025

Mottron, L., & Burack, J. A. (2001). Enhanced perceptual functioning in the development of autism. In *The development of autism: Perspectives from theory and research* (pp. 131–148). Lawrence Erlbaum Associates Publishers.

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

Muncke, J., Kuruvila, I., & Hoppe, U. (2022). Prediction of Speech Intelligibility by Means of EEG Responses to Sentences in Noise. *Frontiers in Neuroscience*, *16*, 876421. https://doi.org/10.3389/fnins.2022.876421

Näätänen, R., & Picton, T. (1987). The N1 Wave of the Human Electric and Magnetic Response to Sound: A Review and an Analysis of the Component Structure. *Psychophysiology*, *24*(4), 375–425. https://doi.org/10.1111/j.1469-8986.1987.tb00311.x

Nadon, É., Tillmann, B., Saj, A., & Gosselin, N. (2021). The Emotional Effect of Background Music on Selective Attention of Adults. *Frontiers in Psychology*, *12*. https://doi.org/10.3389/fpsyg.2021.729037

Navon, D. (1977). Forest before trees: The precedence of global features in visual perception. *Cognitive Psychology*, *9*(3), 353–383. https://doi.org/10.1016/0010-0285(77)90012-3

Newman, R. S. (2005). The Cocktail Party Effect in Infants Revisited: Listening to One's Name in Noise. *Developmental Psychology*, *41*(2), 352–362. https://doi.org/10.1037/0012-1649.41.2.352

Noble, W., & Perrett, S. (2002). Hearing speech against spatially separate competing speech versus competing noise. *Perception & Psychophysics*, *64*(8), 1325–1336. https://doi.org/10.3758/BF03194775

Norris, D. (1994). Shortlist: A connectionist model of continuous speech recognition. *Cognition*, *52*(3), 189–234. https://doi.org/10.1016/0010-0277(94)90043-4

Norris, D. (1999). The merge model: Speech perception is bottom-up. *The Journal of the Acoustical Society of America*, *106*(4_Supplement), 2295–2295. https://doi.org/10.1121/1.427854

Norris, D., & McQueen, J. M. (2008). Shortlist B: A Bayesian model of continuous speech recognition. *Psychological Review*, *115*(2), 357–395. https://doi.org/10.1037/0033-295X.115.2.357

Nygaard, L. C., & Pisoni, D. B. (1998). Talker-specific learning in speech perception. *Perception & Psychophysics*, *60*(3), 355–376. https://doi.org/10.3758/BF03206860

Obleser, J., & Kayser, C. (2019). Neural Entrainment and Attentional Selection in the Listening Brain. *Trends in Cognitive Sciences*, *23*(11), 913–926. https://doi.org/10.1016/j.tics.2019.08.004

Obleser, J., & Kotz, S. A. (2011). Multiple brain signatures of integration in the comprehension of degraded speech. *NeuroImage*, *55*(2), 713–723. https://doi.org/10.1016/j.neuroimage.2010.12.020

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

Ong, J. H., Zhao, C., Bacon, A., Leung, F. Y. N., Veic, A., Wang, L., Jiang, C., & Liu, F. (2024). The Relationship Between Autism and Pitch Perception is Modulated by Cognitive Abilities. *Journal of Autism and Developmental Disorders*, *54*(9), 3400–3411. https://doi.org/10.1007/s10803-023-06075-7

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), 156869. https://doi.org/10.1155/2011/156869

Orf, M., Wöstmann, M., Hannemann, R., & Obleser, J. (2023). Target enhancement but not distractor suppression in auditory neural tracking during continuous speech. *iScience*, *26*(6). https://doi.org/10.1016/j.isci.2023.106849

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

O'Rourke, E., & Coderre, E. L. (2021). Implicit Semantic Processing of Linguistic and Non-linguistic Stimuli in Adults with Autism Spectrum Disorder. *Journal of Autism and Developmental Disorders*, *51*(8), 2611–2630. https://doi.org/10.1007/s10803-020-04736-5

Osterhout, L., & Holcomb, P. J. (1992). Event-related brain potentials elicited by syntactic anomaly. *Journal of Memory and Language*, *31*(6), 785–806. https://doi.org/10.1016/0749-596X(92)90039-Z

O'Sullivan, J. A., Power, A. J., Mesgarani, N., Rajaram, S., Foxe, J. J., Shinn-Cunningham, B. G., Slaney, M., Shamma, S. A., & Lalor, E. C. (2015). Attentional Selection in a Cocktail Party Environment Can Be Decoded from Single-Trial EEG. *Cerebral Cortex*, *25*(7), 1697–1706. https://doi.org/10.1093/cercor/bht355

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

Palana, J., Schwartz, S., & Tager-Flusberg, H. (2022). Evaluating the use of cortical entrainment to measure atypical speech processing: A systematic review. *Neuroscience & Biobehavioral Reviews*, *133*, 104506. https://doi.org/10.1016/j.neubiorev.2021.12.029

Patel, A. D. (1998). Syntactic Processing in Language and Music: Different Cognitive Operations, Similar Neural Resources? *Music Perception*, *16*(1), 27–42. https://doi.org/10.2307/40285775

Patston, L. L. M., & Tippett, L. J. (2011). The Effect of Background Music on Cognitive Performance in Musicians and Nonmusicians. *Music Perception*, *29*(2), 173–183. https://doi.org/10.1525/mp.2011.29.2.173

Peelle, J. E. (2018). Listening Effort: How the Cognitive Consequences of Acoustic Challenge Are Reflected in Brain and Behavior. *Ear and Hearing*, *39*(2), 204. https://doi.org/10.1097/AUD.0000000000000494

Peelle, J. E., Gross, J., & Davis, M. H. (2013). Phase-Locked Responses to Speech in Human Auditory Cortex are Enhanced During Comprehension. *Cerebral Cortex*, *23*(6), 1378–1387. https://doi.org/10.1093/cercor/bhs118

Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., Kastman, E., & Lindeløv, J. K. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods*, *51*(1), 195–203. https://doi.org/10.3758/s13428-018-01193-y

Pellicano, E., & den Houting, J. (2022). Annual Research Review: Shifting from 'normal science' to neurodiversity in autism science. *Journal of Child Psychology and Psychiatry*, *63*(4), 381–396. https://doi.org/10.1111/jcpp.13534

Pettigrew, C. M., Murdoch, B. M., Kei, J., Chenery, H. J., Sockalingam, R., Ponton, C. W., Finnigan, S., & Alku, P. (2004). Processing of English Words with Fine Acoustic Contrasts and Simple Tones: A Mismatch Negativity Study. *Journal of the American Academy of Audiology*, *15*(1), 47–66. https://doi.org/10.3766/jaaa.15.1.6

Pfordresher, P. Q., & Halpern, A. R. (2013). Auditory imagery and the poor-pitch singer. *Psychonomic Bulletin & Review*, *20*(4), 747–753. https://doi.org/10.3758/s13423-013-0401-8

Phan, L., Tariq, A., Lam, G., Pang, E. W., & Alain, C. (2021). The Neurobiology of Semantic Processing in Autism Spectrum Disorder: An Activation Likelihood Estimation Analysis. *Journal of Autism and Developmental Disorders*, *51*(9), 3266–3279. https://doi.org/10.1007/s10803-020-04794-9

Piatti, A., Van der Paelt, S., Warreyn, P., & Roeyers, H. (2021). Atypical attention to voice in toddlers and pre-schoolers with autism spectrum disorder is related to unimpaired cognitive abilities. An ERP study. *Research in Autism Spectrum Disorders*, *86*, 101805. https://doi.org/10.1016/j.rasd.2021.101805

Pichora-Fuller, M. K., Kramer, S. E., Eckert, M. A., Edwards, B., Hornsby, B. W. Y., Humes, L. E., Lemke, U., Lunner, T., Matthen, M., Mackersie, C. L., Naylor, G., Phillips, N. A., Richter, M., Rudner, M., Sommers, M. S., Tremblay, K. L., & Wingfield, A. (2016). Hearing Impairment and Cognitive Energy: The Framework for Understanding Effortful Listening (FUEL). *Ear & Hearing*, *37*(1), 5S-27S. https://doi.org/10.1097/AUD.0000000000000312

Pijnacker, J., Geurts, B., Van Lambalgen, M., Buitelaar, J., & Hagoort, P. (2010). Exceptions and anomalies: An ERP study on context sensitivity in autism. *Neuropsychologia*, *48*(10), 2940–2951. https://doi.org/10.1016/j.neuropsychologia.2010.06.003

Pion-Tonachini, L., Kreutz-Delgado, K., & Makeig, S. (2019). ICLabel: An automated electroencephalographic independent component classifier, dataset, and website. *NeuroImage*, *198*, 181–197. https://doi.org/10.1016/j.neuroimage.2019.05.026

Poeppel, D. (2003). The analysis of speech in different temporal integration windows: Cerebral lateralization as 'asymmetric sampling in time'. *Speech Communication*, *41*(1), 245–255. https://doi.org/10.1016/s0167-6393(02)00107-3

Poeppel, D., Idsardi, W. J., & van Wassenhove, V. (2008). Speech perception at the interface of neurobiology and linguistics. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *363*(1493), 1071–1086. https://doi.org/10.1098/rstb.2007.2160

Posit team. (2022). *RStudio: Integrated Development Environment for R*. Posit Software, PBC. http://www.posit.co/

Poulsen, R., Williams, Z., Dwyer, P., Pellicano, E., Sowman, P. F., & McAlpine, D. (2024). How auditory processing influences the autistic profile: A review. *Autism Research*, *17*(12), 2452–2470. https://doi.org/10.1002/aur.3259

Power, A. J., Foxe, J. J., Forde, E., Reilly, R. B., & Lalor, E. C. (2012). At what time is the cocktail party? A late locus of selective attention to natural speech. *European Journal of Neuroscience*, *35*(9), 1497–1503. https://doi.org/10.1111/j.1460-9568.2012.08060.x

Preis, J., Amon, R., Silbert Robinette, D., & Rozegar, A. (2016). Does Music Matter? The Effects of Background Music on Verbal Expression and Engagement in Children with Autism Spectrum Disorders. *Music Therapy Perspectives*, *34*(1), 106–115. https://doi.org/10.1093/mtp/miu044

Rance, G., Chisari, D., Saunders, K., & Rault, J.-L. (2017). Reducing Listening-Related Stress in School-Aged Children with Autism Spectrum Disorder. *Journal of Autism and Developmental Disorders*, *47*(7), 2010–2022. https://doi.org/10.1007/s10803-017-3114-4

Rankin, C. H., Abrams, T., Barry, R. J., Bhatnagar, S., Clayton, D., Colombo, J., Coppola, G., Geyer, M. A., Glanzman, D. L., Marsland, S., McSweeney, F., Wilson, D. A., Wu, C.-F., & Thompson, R. F. (2009). Habituation Revisited: An Updated and Revised Description of the Behavioral Characteristics of Habituation. *Neurobiology of Learning and Memory*, *92*(2), 135–138. https://doi.org/10.1016/j.nlm.2008.09.012

Raven, J. C., & Court, J. H. (1998). *Raven's progressive matrices and vocabulary scales*. Oxford Psychologists Press Oxford. http://www.v-psyche.com/doc/IQ/Raven-Vocabulary.doc

Reber, A. S. (1989). Implicit learning and tacit knowledge. *Journal of Experimental Psychology: General*, *118*(3), 219. https://psycnet.apa.org/record/1989-38920-001

Rennies, J., Best, V., Roverud, E., & KiddJr., G. (2019). Energetic and Informational Components of Speech-on-Speech Masking in Binaural Speech Intelligibility and Perceived Listening Effort. *Trends in Hearing*, *23*, 2331216519854597. https://doi.org/10.1177/2331216519854597

Rhebergen, K. S., & Versfeld, N. J. (2005). A Speech Intelligibility Index-based approach to predict the speech reception threshold for sentences in fluctuating noise for normal-hearing listeners. *The Journal of the Acoustical Society of America*, *117*(4), 2181–2192. https://doi.org/10.1121/1.1861713

Ribeiro, T. C., Valasek, C. A., Minati, L., & Boggio, P. S. (2013). Altered semantic integration in autism beyond language: A cross-modal event-related potentials study. *NeuroReport*, *24*(8), 414–418. https://doi.org/10.1097/WNR.0b013e328361315e

Ring, H., Sharma, S., Wheelwright, S., & Barrett, G. (2007). An Electrophysiological Investigation of Semantic Incongruity Processing by People with Asperger's Syndrome. *Journal of Autism and Developmental Disorders*, *37*(2), 281–290. https://doi.org/10.1007/s10803-006-0167-1

Robertson, A. E., & Simmons, D. R. (2015). The Sensory Experiences of Adults with Autism Spectrum Disorder: A Qualitative Analysis. *Perception*, *44*(5), 569–586. https://doi.org/10.1068/p7833

Romei, L., Wambacq, I. J. A., Besing, J., Koehnke, J., & Jerger, J. (2011). Neural indices of spoken word processing in background multi-talker babble. *International Journal of Audiology*, *50*(5), 321–333. https://doi.org/10.3109/14992027.2010.547875

Rosen, S. (1992). Temporal information in speech: Acoustic, auditory and linguistic aspects. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, *336*(1278), 367–373. https://doi.org/10.1098/rstb.1992.0070

Rosen, S., Souza, P., Ekelund, C., & Majeed, A. A. (2013). Listening to speech in a background of other talkers: Effects of talker number and noise vocoding. *The Journal of the Acoustical Society of America*, *133*(4), 2431–2443. https://doi.org/10.1121/1.4794379

Rosenhall, U., Nordin, V., Sandström, M., Ahlsen, G., & Gillberg, C. (1999). Autism and hearing loss. *Journal of Autism and Developmental Disorders*, *29*(5), 349–357. https://idp.springer.com/authorize/casa?redirect_uri=https://link.springer.com/article/10.1023/A:1023022709710&casa_token=MIWKz8inNPYAAAAA:aj3VX2mQwN4-x9XMUDxxaRNkAZ2Xqg2rjFNDZHL57VllD5_RDRQoyKB7ZwvN5nSTflXpyjXbHZl4KfEJWLU

Rubenstein, J. L. R., & Merzenich, M. M. (2003). Model of autism: Increased ratio of excitation/inhibition in key neural systems. *Genes, Brain and Behavior*, *2*(5), 255–267. https://doi.org/10.1034/j.1601-183x.2003.00037.x

Ruiz Callejo, D., & Boets, B. (2023). A systematic review on speech-in-noise perception in autism. *Neuroscience & Biobehavioral Reviews*, *154*, 105406. https://doi.org/10.1016/j.neubiorev.2023.105406

Ruiz Callejo, D., Wouters, J., & Boets, B. (2023). Speech-in-noise perception in autistic adolescents with and without early language delay. *Autism Research*, *16*(9), 1719–1727. https://doi.org/10.1002/aur.2966

Russo, F. A., & Pichora-Fuller, M. K. (n.d.). *Tune In or Tune Out: Age-Related Differences in Listening to Speech in Music*.

Russo, F. A., & Pichora-Fuller, M. K. (2008). Tune in or tune out: Age-related differences in listening to speech in music. *Ear and Hearing*, *29*(5), 746–760. https://journals.lww.com/ear-hearing/fulltext/2008/10000/Perilymph_Modiolar_Communication_Routes_in_the.8.aspx

Russo, N., Zecker, S., Trommer, B., Chen, J., & Kraus, N. (2009). Effects of Background Noise on Cortical Encoding of Speech in Autism Spectrum Disorders. *Journal of Autism and Developmental Disorders*, *39*(8), 1185–1196. https://doi.org/10.1007/s10803-009-0737-0

Samoylov, I., Arcara, G., Buyanova, I., Davydova, E., Pereverzeva, D., Sorokin, A., Tyushkevich, S., Mamokhina, U., Danilina, K., Dragoy, O., & Arutiunian, V. (2024). Altered neural synchronization in response to 2 Hz amplitude-modulated tones in the auditory cortex of children with Autism Spectrum Disorder: An MEG study.

International Journal of Psychophysiology, *203*, 112405. https://doi.org/10.1016/j.ijpsycho.2024.112405

Samson, F., Hyde, K. L., Bertone, A., Soulières, I., Mendrek, A., Ahad, P., Mottron, L., & Zeffiro, T. A. (2011). Atypical processing of auditory temporal complexity in autistics. *Neuropsychologia*, *49*(3), 546–555. https://doi.org/10.1016/j.neuropsychologia.2010.12.033

Schaeffer, J., Abd El-Raziq, M., Castroviejo, E., Durrleman, S., Ferré, S., Grama, I., Hendriks, P., Kissine, M., Manenti, M., Marinis, T., Meir, N., Novogrodsky, R., Perovic, A., Panzeri, F., Silleresi, S., Sukenik, N., Vicente, A., Zebib, R., Prévost, P., & Tuller, L. (2023). Language in autism: Domains, profiles and co-occurring conditions. *Journal of Neural Transmission*, *130*(3), 433–457. https://doi.org/10.1007/s00702-023-02592-y

Schafer, E. C., Gopal, K. V., Mathews, L., Miller, S., & Lam, B. P. W. (2024). Impact of an Auditory Processing Training Program on Individuals With Autism Spectrum Disorder. *American Journal of Audiology*, *33*(4), 1221–1236. https://doi.org/10.1044/2024_AJA-24-00134

Schafer, E. C., Mathews, L., Gopal, K., Canale, E., Creech, A., Manning, J., & Kaiser, K. (2020). Behavioral Auditory Processing in Children and Young Adults with Autism Spectrum Disorder. *Journal of the American Academy of Audiology*, *31*(9), 680–689. https://doi.org/10.1055/s-0040-1717138

Schafer, E., Traber, J., Layden, P., Amin, A., Sanders, K., Bryant, D., & Baldus, N. (2014). Use of Wireless Technology for Children with Auditory Processing Disorders, Attention-Deficit Hyperactivity Disorder, and Language Disorders. *Seminars in Hearing*, *35*(03), 193–205. https://doi.org/10.1055/s-0034-1383504

Scharenborg, O., & Larson, M. (2018). The Conversation Continues: The Effect of Lyrics and Music Complexity of Background Music on Spoken-Word Recognition. *Interspeech 2018*, 2280–2284. https://doi.org/10.21437/Interspeech.2018-1088

Schelinski, S., Borowiak, K., & Von Kriegstein, K. (2016). Temporal voice areas exist in autism spectrum disorder but are dysfunctional for voice identity recognition. *Social Cognitive and Affective Neuroscience*, *11*(11), 1812–1822. https://doi.org/10.1093/scan/nsw089

Schelinski, S., Roswandowitz, C., & von Kriegstein, K. (2017). Voice identity processing in autism spectrum disorder. *Autism Research*, *10*(1), 155–168. https://doi.org/10.1002/aur.1639

Schelinski, S., Tabas, A., & von Kriegstein, K. (2022). Altered processing of communication signals in the subcortical auditory sensory pathway in autism. *Human Brain Mapping*, *43*(6), 1955–1972. https://doi.org/10.1002/hbm.25766

Schelinski, S., & Von Kriegstein, K. (2020). Brief Report: Speech-in-Noise Recognition and the Relation to Vocal Pitch Perception in Adults with Autism Spectrum Disorder and Typical Development. *Journal of Autism and Developmental Disorders*, *50*(1), 356–363. https://doi.org/10.1007/s10803-019-04244-1

Schelinski, S., & Von Kriegstein, K. (2023). Responses in left inferior frontal gyrus are altered for speech-in-noise processing, but not for clear speech in autism. *Brain and Behavior*, *13*(2), e2848. https://doi.org/10.1002/brb3.2848

Schneider, W., & Shiffrin, R. M. (1977). Controlled and automatic human information processing: I. Detection, search, and attention. *Psychological Review*, *84*(1), 1. https://psycnet.apa.org/journals/rev/84/1/1/

Schwartz, S., Wang, L., Shinn-Cunningham, B. G., & Tager-Flusberg, H. (2020). Neural Evidence for Speech Processing Deficits During a Cocktail Party Scenario in

Minimally and Low Verbal Adolescents and Young Adults with Autism. *Autism Research*, *13*(11), 1828–1842. https://doi.org/10.1002/aur.2356

Schwartz, S., Wang, L., Uribe, S., Shinn-Cunningham, B. G., & Tager-Flusberg, H. (2023). Auditory evoked potentials in adolescents with autism: An investigation of brain development, intellectual impairment, and neural encoding. *Autism Research*, *16*(10), 1859–1876. https://doi.org/10.1002/aur.3003

Shannon, R. V., Zeng, F.-G., Kamath, V., Wygonski, J., & Ekelid, M. (1995). Speech Recognition with Primarily Temporal Cues. *Science*, *270*(5234), 303–304. https://doi.org/10.1126/science.270.5234.303

Shi, L.-F., & Law, Y. (2010). Masking effects of speech and music: Does the masker's hierarchical structure matter? *International Journal of Audiology*, *49*(4), 296–308. https://doi.org/10.3109/14992020903350188

Shinn-Cunningham, B., Best, V., & Lee, A. K. C. (2017). Auditory Object Formation and Selection. In J. C. Middlebrooks, J. Z. Simon, A. N. Popper, & R. R. Fay (Eds), *The Auditory System at the Cocktail Party* (pp. 7–40). Springer International Publishing. https://doi.org/10.1007/978-3-319-51662-2_2

Shinn-Cunningham, B. G., & Best, V. (2008). Selective Attention in Normal and Impaired Hearing. *Trends in Amplification*, *12*(4), 283–299. https://doi.org/10.1177/1084713808325306

Silcox, J. W., & Payne, B. R. (2021). The costs (and benefits) of effortful listening on context processing: A simultaneous electrophysiology, pupillometry, and behavioral study. *Cortex*, *142*, 296–316. https://doi.org/10.1016/j.cortex.2021.06.007

Simpson, S. A., & Cooke, M. (2005). Consonant identification in N-talker babble is a nonmonotonic function of N. *The Journal of the Acoustical Society of America*, *118*(5), 2775–2778. https://doi.org/10.1121/1.2062650

Smith, N. J., & Levy, R. (2013). The effect of word predictability on reading time is logarithmic. *Cognition*, *128*(3), 302–319. https://doi.org/10.1016/j.cognition.2013.02.013

Sohoglu, E., Peelle, J. E., Carlyon, R. P., & Davis, M. H. (2012). Predictive Top-Down Integration of Prior Knowledge during Speech Perception. *Journal of Neuroscience*, *32*(25), 8443–8453. https://doi.org/10.1523/JNEUROSCI.5069-11.2012

Song, J., Martin, L., & Iverson, P. (2020). Auditory neural tracking and lexical processing of speech in noise: Masker type, spatial location, and language experience. *The Journal of the Acoustical Society of America*, *148*(1), 253–264. https://doi.org/10.1121/10.0001477

Stanutz, S., Wapnick, J., & Burack, J. (2012). Pitch discrimination and melodic memory in children with autism spectrum disorders. *Autism*, *18*, 137–147. https://doi.org/10.1177/1362361312462905

Strait, D. L., & Kraus, N. (2011). Can You Hear Me Now? Musical Training Shapes Functional Brain Networks for Selective Auditory Attention and Hearing Speech in Noise. *Frontiers in Psychology*, *2*. https://doi.org/10.3389/fpsyg.2011.00113

Strauß, A., Kotz, S. A., & Obleser, J. (2013). Narrowed Expectancies under Degraded Speech: Revisiting the N400. *Journal of Cognitive Neuroscience*, *25*(8), 1383–1395. https://doi.org/10.1162/jocn_a_00389

Stringer, L., & Iverson, P. (2020). Non-native speech recognition sentences: A new materials set for non-native speech perception research. *Behavior Research Methods*, *52*(2), 561–571. https://doi.org/10.3758/s13428-019-01251-z

Studebaker, G. A. (1985). A 'Rationalized' Arcsine Transform. *Journal of Speech, Language, and Hearing Research*, *28*(3), 455–462. https://doi.org/10.1044/jshr.2803.455

Sturrock, A., Guest, H., Hanks, G., Bendo, G., Plack, C. J., & Gowen, E. (2022). Chasing the conversation: Autistic experiences of speech perception. *Autism & Developmental Language Impairments*, *7*, 23969415221077532. https://doi.org/10.1177/23969415221077532

Summers, R. J., & Roberts, B. (2020). Informational masking of speech by acoustically similar intelligible and unintelligible interferers. *The Journal of the Acoustical Society of America*, *147*(2), 1113–1125. https://doi.org/10.1121/10.0000688

Swaminathan, J., Mason, C. R., Streeter, T. M., Best, V., Roverud, E., & Kidd, G. (2016). Role of Binaural Temporal Fine Structure and Envelope Cues in Cocktail-Party Listening. *Journal of Neuroscience*, *36*(31), 8250–8257. https://doi.org/10.1523/jneurosci.4421-15.2016

Teder-Sälejärvi, W. A., Hillyard, S. A., Röder, B., & Neville, H. J. (1999). Spatial attention to central and peripheral auditory stimuli as indexed by event-related potentials. *Cognitive Brain Research*, *8*(3), 213–227. https://doi.org/10.1016/S0926-6410(99)00023-3

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*(2–3), 221–234. https://doi.org/10.1016/j.cogbrainres.2004.10.021

The MathWorks Inc. (2022). *MATLAB version: 9.13.0 (R2022b), Natick, Massachusetts: The MathWorks Inc.* [Computer software]. https://www.mathworks.com

Thompson, W. F., Schellenberg, E. G., & Husain, G. (2001). Arousal, Mood, and The Mozart Effect. *Psychological Science*, *12*(3), 248–251. https://doi.org/10.1111/1467-9280.00345

Ueda, K., Nakajima, Y., Ellermeier, W., & Kattner, F. (2017). Intelligibility of locally time-reversed speech: A multilingual comparison. *Scientific Reports*, *7*(1). https://doi.org/10.1038/s41598-017-01831-z

Van Boxtel, J. J. A., & Lu, H. (2013). A predictive coding perspective on autism spectrum disorders. *Frontiers in Psychology*, *4*. https://doi.org/10.3389/fpsyg.2013.00019

Van De Cruys, S., Evers, K., Van Der Hallen, R., Van Eylen, L., Boets, B., de-Wit, L., & Wagemans, J. (2014). Precise minds in uncertain worlds: Predictive coding in autism. *Psychological Review*, *121*(4), 649–675. https://doi.org/10.1037/a0037665

Van Engen, K. J. (2010). Similarity and familiarity: Second language sentence recognition in first- and second-language multi-talker babble. *Speech Communication*, *52*(11), 943–953. https://doi.org/10.1016/j.specom.2010.05.002

Van Engen, K. J., & Bradlow, A. R. (2007). Sentence recognition in native- and foreign-language multi-talker background noise. *The Journal of the Acoustical Society of America*, *121*(1), 519–526. https://doi.org/10.1121/1.2400666

Van Rij, J., Hendriks, P., Van Rijn, H., Baayen, R. H., & Wood, S. N. (2019). Analyzing the Time Course of Pupillometric Data. *Trends in Hearing*, *23*, 2331216519832483. https://doi.org/10.1177/2331216519832483

Van Rij, J., Wieling, M., & Baayen, R. H. (2015). *itsadug: Interpreting Time Series and Autocorrelated Data Using GAMMs* (p. 2.4.1) [Data set]. https://doi.org/10.32614/CRAN.package.itsadug

Verschueren, E., Gillis, M., Decruy, L., Vanthornhout, J., & Francart, T. (2022). Speech Understanding Oppositely Affects Acoustic and Linguistic Neural Tracking in a Speech Rate Manipulation Paradigm. *The Journal of Neuroscience*, *42*(39), 7442–7453. https://doi.org/10.1523/JNEUROSCI.0259-22.2022

Vissers, M. E., X Cohen, M., & Geurts, H. M. (2012). Brain connectivity and high functioning autism: A promising path of research that needs refined models, methodological convergence, and stronger behavioral links. *Neuroscience & Biobehavioral Reviews*, *36*(1), 604–625. https://doi.org/10.1016/j.neubiorev.2011.09.003

Viswanathan, N., Kokkinakis, K., & Williams, B. T. (2016). Spatially separating language masker from target results in spatial and linguistic masking release. *The Journal of the Acoustical Society of America*, *140*(6), EL465–EL470. https://doi.org/10.1121/1.4968034

Wang, L., Ong, J. H., Ponsot, E., Hou, Q., Jiang, C., & Liu, F. (2023). Mental representations of speech and musical pitch contours reveal a diversity of profiles in autism spectrum disorder. *Autism*, *27*(3), 629–646. https://doi.org/10.1177/13623613221111207

Wang, X., Delgado, J., Marchesotti, S., Kojovic, N., Sperdin, H. F., Rihs, T. A., Schaer, M., & Giraud, A.-L. (2023). Speech Reception in Young Children with Autism Is Selectively Indexed by a Neural Oscillation Coupling Anomaly. *The Journal of Neuroscience*, *43*(40), 6779–6795. https://doi.org/10.1523/JNEUROSCI.0112-22.2023

Wang, X., Wang, S., Fan, Y., Huang, D., & Zhang, Y. (2017). Speech-specific categorical perception deficit in autism: An Event-Related Potential study of lexical tone processing in Mandarin-speaking children. *Scientific Reports*, *7*(1), 43254. https://doi.org/10.1038/srep43254

Wang, X., & Xu, L. (2021). Speech perception in noise: Masking and unmasking. *Journal of Otology*, *16*(2), 109–119. https://doi.org/10.1016/j.joto.2020.12.001

Warren, R. M. (1970). Perceptual Restoration of Missing Speech Sounds. *Science*, *167*(3917), 392–393. https://doi.org/10.1126/science.167.3917.392

Wechsler, H., Nelson, T. E., Lee, J. E., Seibring, M., Lewis, C., & Keeling, R. P. (2003). Perception and reality: A national evaluation of social norms marketing interventions to reduce college students' heavy alcohol use. *Journal of Studies on Alcohol*, *64*(4), 484–494. https://doi.org/10.15288/jsa.2003.64.484

Weissbart, H., Kandylaki, K. D., & Reichenbach, T. (2020). Cortical Tracking of Surprisal during Continuous Speech Comprehension. *Journal of Cognitive Neuroscience*, *32*(1), 155–166. https://doi.org/10.1162/jocn_a_01467

Wenzel, E. M., Arruda, M., Kistler, D. J., & Wightman, F. L. (1993). Localization using nonindividualized head-related transfer functions. *The Journal of the Acoustical Society of America*, *94*(1), 111–123. https://doi.org/10.1121/1.407089

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

Wightman, F. L., & Kistler, D. J. (1989). Headphone simulation of free-field listening. II: Psychophysical validation. *The Journal of the Acoustical Society of America*, *85*(2), 868–878. https://doi.org/10.1121/1.397558

Williams, Z. J., He, J. L., Cascio, C. J., & Woynaroski, T. G. (2021). A review of decreased sound tolerance in autism: Definitions, phenomenology, and potential mechanisms. *Neuroscience & Biobehavioral Reviews*, *121*, 1–17. https://doi.org/10.1016/j.neubiorev.2020.11.030

Wilson, W. J., Harper-Hill, K., Armstrong, R., Downing, C., Perrykkad, K., Rafter, M., & Ashburner, J. (2021). A preliminary investigation of sound-field amplification as an inclusive classroom adjustment for children with and without autism spectrum disorder. *Journal of Communication Disorders*, *93*, 106142. https://www.sciencedirect.com/science/article/pii/S0021992421000654?casa_token=

wwh_6HHABtkAAAAA:42N6hb7a3cpM2LL9NshKUrf2y0Wj17mkkTRacbbu_Tx-wRTvwYQdSC8whAX82Yre-zB602bG-CVf

Wong, P. C. M., Skoe, E., Russo, N. M., Dees, T., & Kraus, N. (2007). Musical experience shapes human brainstem encoding of linguistic pitch patterns. *Nature Neuroscience*, *10*(4), 420–422. https://doi.org/10.1038/nn1872

Wood, S. N. (2011). Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, *73*(1), 3–36. https://academic.oup.com/jrsssb/article-abstract/73/1/3/7034726

Wood, S. N. (2017). *Generalized additive models: An introduction with R*. chapman and hall/CRC. https://www.taylorfrancis.com/books/mono/10.1201/9781315370279/generalized-additive-models-simon-wood

Woods, W. S., & Colburn, H. S. (1992). Test of a model of auditory object formation using intensity and interaural time difference discrimination. *The Journal of the Acoustical Society of America*, *91*(5), 2894–2902. https://doi.org/10.1121/1.402926

Wöstmann, M., Herrmann, B., Wilsch, A., & Obleser, J. (2015). Neural Alpha Dynamics in Younger and Older Listeners Reflect Acoustic Challenges and Predictive Benefits. *The Journal of Neuroscience*, *35*(4), 1458–1467. https://doi.org/10.1523/JNEUROSCI.3250-14.2015

Xu, S., Zhang, H., Fan, J., Jiang, X., Zhang, M., Guan, J., Ding, H., & Zhang, Y. (2023). *Auditory Challenges and Listening Effort in School-Age Children with Autism: Insights from Pupillary Dynamics during Speech in Noise Perception*. Social Sciences. https://doi.org/10.20944/preprints202309.1636.v1

Xu, S., Zhang, H., Fan, J., Jiang, X., Zhang, M., Guan, J., Ding, H., & Zhang, Y. (2024). Auditory Challenges and Listening Effort in School-Age Children With Autism: Insights From Pupillary Dynamics During Speech-in-Noise Perception. *Journal of Speech, Language, and Hearing Research*, *67*(7), 2410–2453. https://doi.org/10.1044/2024_JSLHR-23-00553

Xu, Y. (2005). Speech melody as articulatorily implemented communicative functions. *Speech Communication*, *46*(3), 220–251. https://doi.org/10.1016/j.specom.2005.02.014

Yasmin, S., Irsik, V. C., Johnsrude, I. S., & Herrmann, B. (2023). The effects of speech masking on neural tracking of acoustic and semantic features of natural speech. *Neuropsychologia*, *186*, 108584. https://doi.org/10.1016/j.neuropsychologia.2023.108584

Yip, M. J. W. (2002). *Tone*. Cambridge University Press. https://books.google.com/books?hl=en&lr=&id=KFv2lojXjpwC&oi=fnd&pg=PA6&dq=Yip,+M.+(2002).+Tone.+Cambridge+University+Press.&ots=mdHEFDzMGG&sig=v8zzESu5n0LXo_w68kcUNjVKoCo

Yu, L., Fan, Y., Deng, Z., Huang, D., Wang, S., & Zhang, Y. (2015). Pitch Processing in Tonal-Language-Speaking Children with Autism: An Event-Related Potential Study. *Journal of Autism and Developmental Disorders*, *45*(11), 3656–3667. https://doi.org/10.1007/s10803-015-2510-x

Zekveld, A. A., Heslenfeld, D. J., Festen, J. M., & Schoonhoven, R. (2006). Top–down and bottom–up processes in speech comprehension. *Neuroimage*, *32*(4), 1826–1836. https://www.sciencedirect.com/science/article/pii/S1053811906005076?casa_token=xJWs-cnOJHEAAAAA:92SYY9yYGB4d3HRsXdsmlPeyro0bYyDMQJj9X-R0H3dGG6ZSZUQfXzp5FKjgts0oVKu_PqBdFbBS

Zekveld, A. A., Rudner, M., Johnsrude, I. S., Heslenfeld, D. J., & Rönnberg, J. (2012). Behavioral and fMRI evidence that cognitive ability modulates the effect of semantic context on speech intelligibility. *Brain and Language*, *122*(2), 103–113. https://doi.org/10.1016/j.bandl.2012.05.006

Zendel, B. R., Tremblay, C.-D., Belleville, S., & Peretz, I. (2015). The Impact of Musicianship on the Cortical Mechanisms Related to Separating Speech from Background Noise. *Journal of Cognitive Neuroscience*, *27*(5), 1044–1059. https://doi.org/10.1162/jocn_a_00758

Zhang, J., Meng, Y., Wu, C., Xiang, Y.-T., & Yuan, Z. (2019). Non-speech and speech pitch perception among Cantonese-speaking children with autism spectrum disorder: An ERP study. *Neuroscience Letters*, *703*, 205–212. https://doi.org/10.1016/j.neulet.2019.03.021

Zhang, J., Wang, X., Wang, N., Fu, X., Gan, T., Galvin, J. J., Willis, S., Xu, K., Thomas, M., & Fu, Q.-J. (2020). Tonal Language Speakers Are Better Able to Segregate Competing Speech According to Talker Sex Differences. *Journal of Speech, Language, and Hearing Research : JSLHR*, *63*(8), 2801–2810. https://doi.org/10.1044/2020_JSLHR-19-00421

Zhang, X., Li, J., Li, Z., Hong, B., Diao, T., Ma, X., Nolte, G., Engel, A. K., & Zhang, D. (2023). Leading and following: Noise differently affects semantic and acoustic processing during naturalistic speech comprehension. *NeuroImage*, *282*, 120404. https://doi.org/10.1016/j.neuroimage.2023.120404

Zheng, Y., Gao, P., & Li, X. (2023). The modulating effect of musical expertise on lexical-semantic prediction in speech-in-noise comprehension: Evidence from an EEG study. *Psychophysiology*, *60*(11). https://doi.org/10.1111/psyp.14371

Zimmerman, R., Smith, A., Fech, T., Mansour, Y., & Kulesza, R. J. (2020). In utero exposure to valproic acid disrupts ascending projections to the central nucleus of the inferior colliculus from the auditory brainstem. *Experimental Brain Research*, *238*(3), 551–563. https://doi.org/10.1007/s00221-020-05729-7

Zion Golumbic, E. M., Ding, N., Bickel, S., Lakatos, P., Schevon, C. A., McKhann, G. M., Goodman, R. R., Emerson, R., Mehta, A. D., Simon, J. Z., Poeppel, D., & Schroeder, C. E. (2013). Mechanisms Underlying Selective Neuronal Tracking of Attended Speech at a "Cocktail Party". *Neuron*, *77*(5), 980–991. https://doi.org/10.1016/j.neuron.2012.12.037

Appendices for Study 1

Appendix A. Power Analysis

To determine the sample size, we conducted a power analysis using preliminary accuracy data (10 participants per group). Simulations were run using the mixedpower package (Kumle et al., 2021) with 1000 iterations, incorporating group, cue condition, background music, and their interactions as fixed effects.

To enhance statistical power, we grouped the location-cue and gender-cue conditions into a one-cue condition and applied Helmert coding to focus on two key contrasts: 1) no-cue vs. any cue conditions (the average of one-cue and both-cues); 2) one-cue vs. both-cues. This allowed us to prioritise the most relevant contrasts and enhance statistical power.

A generalised linear model (GLM) was fitted to pilot data, including fixed effects and random intercepts for participants and items, from which we obtained beta coefficients to define the smallest effect size of interest (SESOI). To account for uncertainty in effect size estimates, beta coefficients were reduced by 15% (Kumle et al., 2021).

As shown in Figure A1, a sample size of 70 participants (n = 35 per group) would provide approximately 80% power to detect most effects of interest, particularly three-way interactions. Based on these results, we set our target sample size at 70 participants.

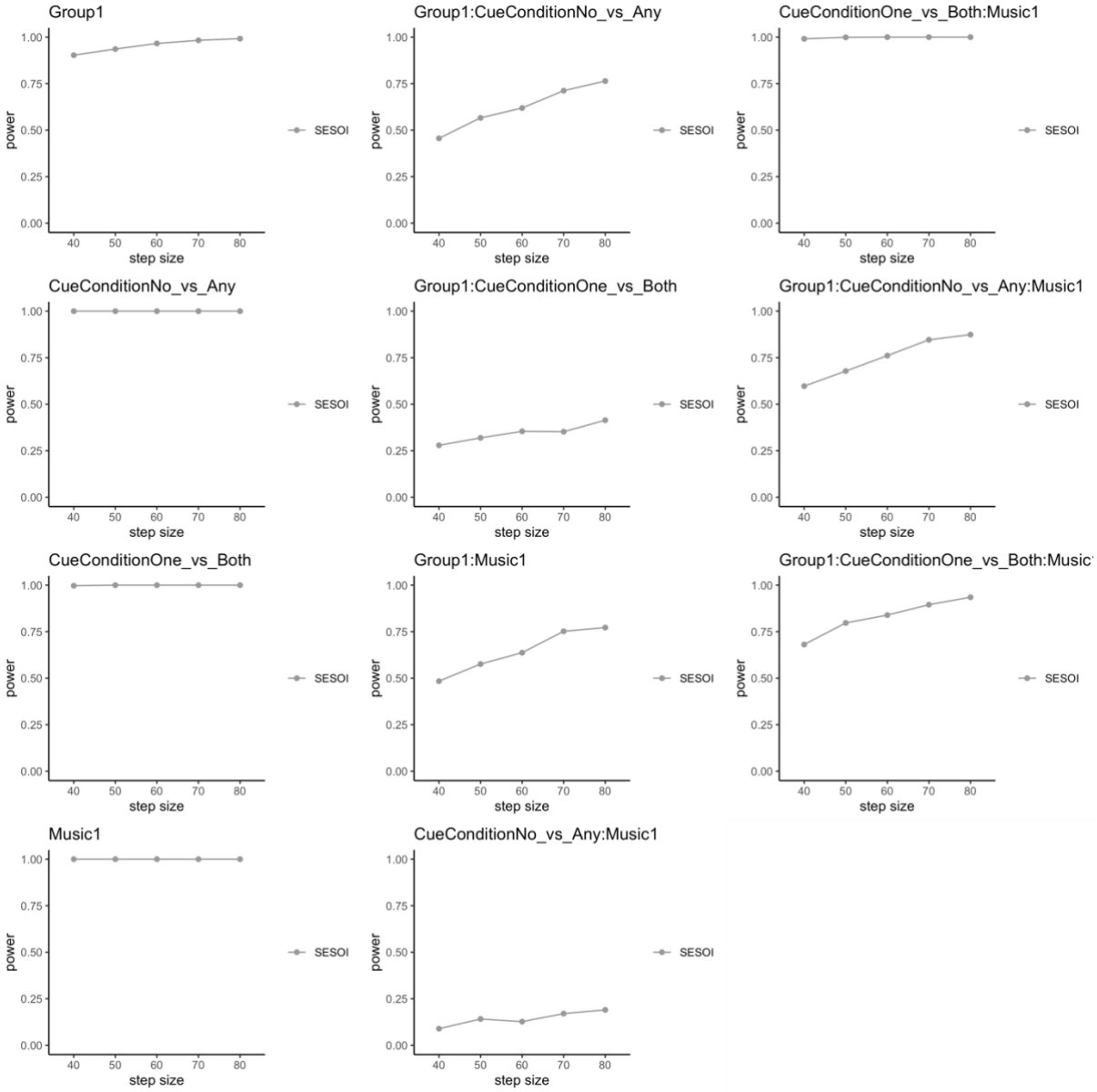

Figure A1. **Power analysis results.** The plots show power estimates (y-axis) across different step sizes (x-axis) for main effects and interactions in the model. Each panel represents a specific effect or interaction, with error bars indicating uncertainty based on the smallest effect size of interest.

Appendix B. Pilot Study

To determine the appropriate signal-to-noise ratio (SNR) for the distractor speech, a pilot study was conducted with three neurotypical, English-speaking participants with normal hearing. Participants listened to target speech while distractor speech was presented at four SNR levels: −3 dB, −6 dB, −9 dB, and −12 dB. An SNR of −3 dB means the distractor speech is 3dB louder than the target speech, with progressively lower SNR levels indicating more challenging listening conditions.

The pilot study comprised 192 trials (48 per SNR level) across four blocks, with all four cue conditions randomly presented within each block. Results (see Table B1) showed high accuracy at −3 dB and −6 dB (~80% overall, with near-ceiling performance in the both-cues condition). At −9 dB, accuracy declined to 60-70%, balancing task difficulty with reasonable performance. At −12 dB, accuracy dropped to ~50%, indicating substantial difficulty and possible reliance on guesswork.

Based on these findings, −9 dB was initially selected as it provided a balance between task difficulty and performance. However, given potential sensitivity to high sound levels in autistic participants (Danesh et al., 2021; Khalfa et al., 2004), the final experiment included a mix of −3 dB and −9 dB trials to reduce difficulty while maintaining ecological validity.

Notably, the effect of SNR level was not a focus of the study and was not analysed as an experimental variable. Instead, the inclusion of both levels served to introduce natural variation in task difficulty, support sustained participant engagement, and ensure the task was suitable for individuals with differing sensory profiles. During the experiment, the two SNR levels were randomly intermixed within blocks rather than presented in separate blocks. This design choice was intended to minimise learning or strategy effects that could arise if participants became aware of predictable shifts in difficulty. Randomisation across trials helped to maintain a more naturalistic and unpredictable listening environment, reflecting real-world situations in which background noise levels vary unexpectedly. Participants were not informed of the SNR changes in advance, further encouraging sustained attention and adaptive listening strategies. In summary, our approach prioritised participant accessibility, ecological validity, and engagement, while avoiding potential confounds related to sensory sensitivity and ensuring that the task remained feasible and realistic across both groups.

Table B1. Mean accuracy rate of SNR levels included in the pilot study.

SNR Level	−3 dB	−6 dB	−9 dB	−12 dB
	Mean (SD)	***Mean (SD)***	***Mean (SD)***	***Mean (SD)***
No Cue	58% (50%)	72% (45%)	58% (50%)	47% (51%)
Gender Cue	86% (35%)	78% (42%)	72% (45%)	50% (51%)
Location Cue	85% (37%)	86% (35%)	72% (45%)	58% (50%)
Both Cues	96% (17%)	97% (17%)	66.7% (48%)	72% (45%)
Grand average	81% (39%)	83% (37%)	67% (47%)	57% (50%)

Appendix C. Generalised additive mixed models (GAMMs)

C1. Accuracy GAMM reported in main text (no-cue vs. both-cues)

C1.1 Rationale for model construction

The effect of background music was excluded from GAMM analyses because it was presented randomly across trials, focusing on its incidental influence on overall performance rather than trial-level dynamics. To balance analytical focus with statistical reliability, our main GAMM analysis focused on the contrast between the both-cues and no-cue conditions. This decision was guided by both empirical and methodological considerations. First, the gender and location cue conditions showed accuracy patterns that were broadly similar to the both-cues condition, with performance typically ranging from 85–100% across trials for both groups (see Figure C1). This similarity indicates that the one-cue conditions did not produce meaningfully different behavioural outcomes from the both-cues condition. Second, including all four cue conditions would have significantly increased the number of group–cue comparisons and model complexity, reducing statistical power and increasing the risk of overfitting. Focusing on the most theoretically and behaviourally distinct contrast (i.e., both cues vs. no cue) allowed us to preserve statistical precision and maximise interpretability.

Although the analysis of reaction times (RTs) for accurate responses was included in our linear mixed-effects models (LMMs), it was not employed in the main GAMM analysis due to limitations in interpretability. Unlike LMMs, which focus on estimating mean effects across conditions, GAMMs model changes over time and rely on consistent data density within each condition to estimate smooth terms reliably. In our dataset, reduced accuracy especially in the no-cue condition resulted in fewer correct trials and thus sparser RT data. This sparsity compromised the stability of the smooth estimates. Moreover, because we examined RTs only for correct responses, more difficult conditions yielded RT data from a restricted and potentially non-representative subset of participants, introducing selection bias. Together, these limitations made RTs unsuitable for reliable GAMM modelling in the main analysis. In contrast, accuracy was recorded on every trial, providing a complete and more representative basis for modelling trial-level dynamics.

In response to reviewer suggestions, we included exploratory GAMM analyses using RT data as well as a four-condition accuracy model in the Appendix. While these analyses provide a

broader view of cue-specific and time-varying effects, their results should be interpreted with caution due to the limitations outlined above.

To avoid conflating the effects of cue condition and signal-to-noise ratio (SNR), we fitted separate models for each SNR level (−3 dB and −9 dB). This allowed for clearer interpretation of condition and group effects without confounding influences from SNR differences.

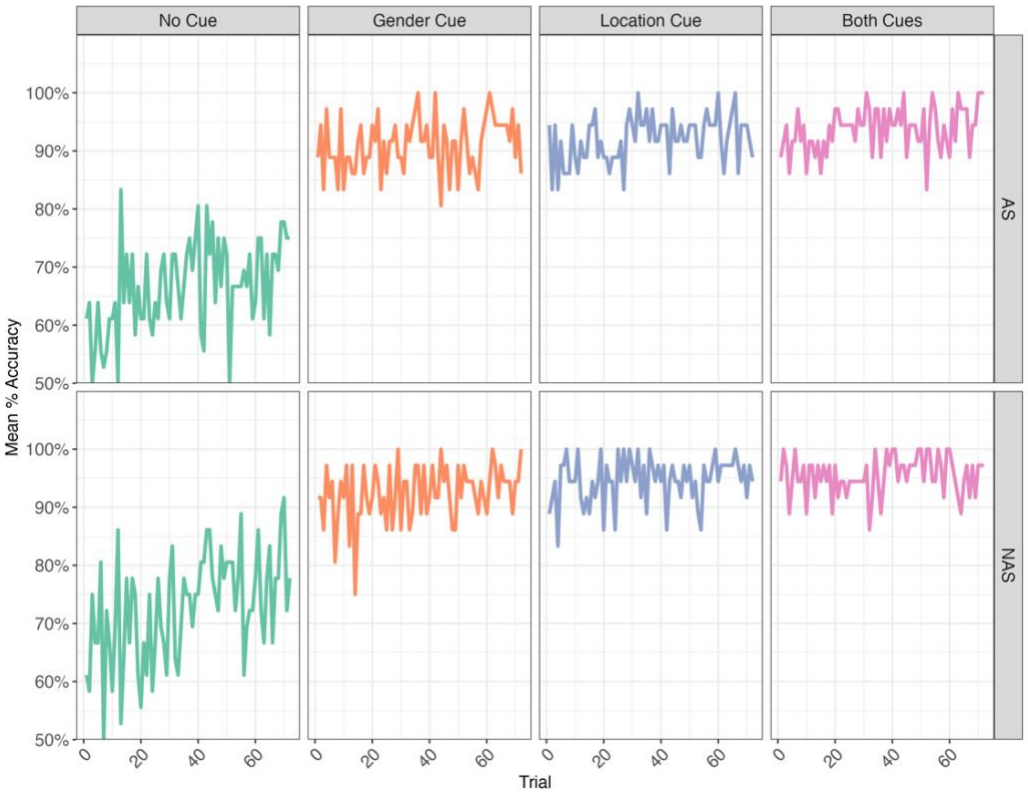

Figure C1. Trial-level mean accuracy in each condition for autistic (AS) and non-autistic participants (NAS).

C1.2 Procedure of model fitting

We constructed a series of nested models with increasing complexity and compared them using the compareML function from the *itsadug* package (van Rij et al., 2019). Three models were fitted:

- **Model 1 (m1)** included cue condition as a fixed parametric effect, a global smooth term for trial order to capture non-linear learning or adaptation effects, and a by-participant smooth to account for individual variability across trials.
- **Model 2 (m2)** extended m1 by allowing separate smooths for each cue condition, enabling trial-level effects to vary non-linearly across conditions.

- **Model 3 (m3)** replaced the fixed cue condition term with a Group × Cue interaction (two groups × two cue conditions). In m3, the autistic group in the both-cues condition served as the reference level, and separate smooths were estimated for each group–condition combination to capture potentially distinct learning curves.

Model comparisons were conducted sequentially (m1 vs. m2, then m2 vs. m3). Because compareML reports half of the likelihood-ratio statistic under "Difference," the χ^2 values shown here correspond to those printed by the function, while the reported *p*-values are based on the doubled statistic (the conventional likelihood-ratio test). As likelihood-ratio tests for smooth terms provide approximate *p*-values, we interpreted them cautiously and placed emphasis on retaining theoretically motivated predictors and interactions. This approach ensured that models remained aligned with our research questions, even when purely statistical criteria (e.g., AIC) suggested a simpler structure.

For the −9 dB data, model comparisons revealed that m2 significantly improved fit over m1, $\chi^2(2) = 6.10$, $p = .002$, supporting the inclusion of condition-specific trial order variability. Furthermore, m3 significantly improved fit over m2, $\chi^2(6) = 6.95$, $p = .031$, indicating that the Group × Cue interaction contributed significantly to explaining performance. Because testing this interaction was central to our research questions, m3 was selected as the best-fitting model for the −9 dB data.

For the −3 dB data, m2 provided a non-significant improvement over m1, $\chi^2(2) = 0.69$, $p = .504$, and m3 similarly yielded a non-significant improvement over m2, $\chi^2(6) = 3.72$, $p = .283$. Nonetheless, we retained m3 as the final model to ensure a consistent model structure across SNR levels and to allow estimation of the theoretically important Group × Cue interaction.

To ensure that model complexity did not result in overfitting, we conducted extensive model criticism. Diagnostic plots were used to ensure key assumptions were met (Figure C2). The Q-Q plots indicated that residuals closely followed a normal distribution, with only minor deviations in the tails, generally supporting the assumption of normally distributed residuals. The residuals vs. fitted values plots showed a scattered pattern without discernible structure, suggesting homoscedasticity and the absence of systematic trends in residuals. Temporal dependencies were assessed using the autocorrelation function (ACF) of residuals (Van Rij et al., 2019). As shown in Figure C2, ACF values at Lag 1 and beyond remained close to zero,

with no significant deviations across all lags, indicating that the model adequately accounted for temporal dependencies. Therefore, no further corrections for autocorrelation were necessary.

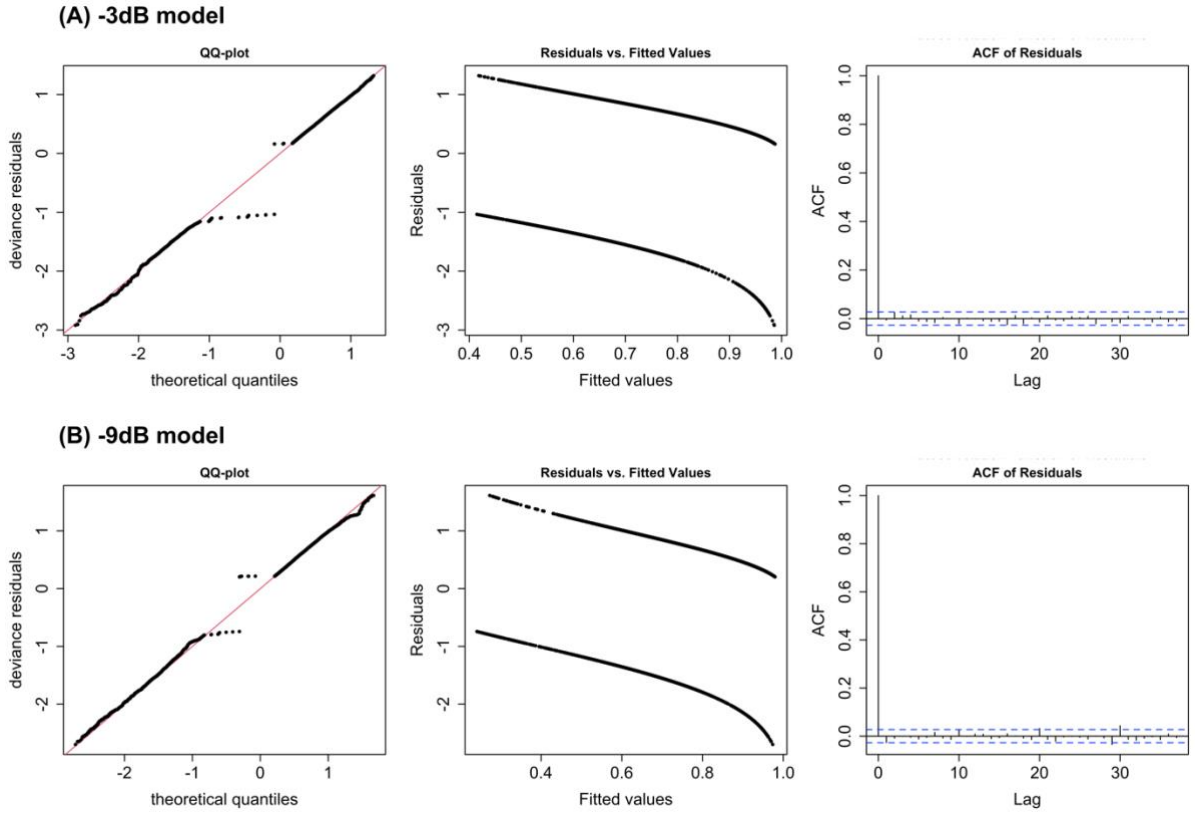

Figure C2. Accuracy model (No-cue vs. Both-cues) diagnostics for the selected GAMM for each SNR level.

C2. Reaction times GAMMs (exploratory)

C2.1 Procedure of model fitting

In response to reviewer feedback, we conducted an exploratory GAMM analysis of reaction times (RTs) for correct responses, using a model-fitting procedure consistent with that applied to the accuracy data. Likelihood ratio tests were not applicable for the RTs' GAMMs due to the way degrees of freedom are estimated in models with smooth terms. Specifically, generalised additive models use penalised smoothing, which can result in non-integer and even negative differences in effective degrees of freedom when comparing models. This occurs when more complex models include smooth terms that are heavily penalised so that they contribute little explanatory value. As a result, the assumption of nested models with valid, positive degrees of freedom required for likelihood ratio testing is violated. Accordingly, we based our model selection for RTs on AIC differences and theoretical interpretability.

For the −9 dB data, model 2 (m2), which included cue-specific smooths, showed a modest improvement in AIC ($\Delta AIC = 1.66$) over the simpler global-smooth model (m1). Model 3 (m3), which included Group × Cue-specific smooths, significantly improved model fit relative to m2, $\chi^2(6) = 8.48$, $p = .009$, and further reduced AIC ($\Delta AIC = 4.39$). Based on these results, we selected m3 as the best-fitting and theoretically most appropriate model for the −9 dB reaction time data.

For the −3 dB data, m2 showed slightly worse fit than m1, with a higher ML score and AIC ($\Delta AIC = -0.26$), indicating that cue-specific smooths did not improve model performance. Model 3 (m3) further worsened fit compared to m2 ($\Delta AIC = -2.23$). Accordingly, we selected m1 as the most parsimonious and best-fitting model for the −3 dB data. Notably, this model did not include the Group × Cue interaction of interest, as the data did not support the inclusion of more complex terms. We therefore report the −3 dB RT results for completeness but interpret them with caution, in contrast to the −9 dB analysis where model fit supported the inclusion of the interaction.

Model check was conducted for the RT-based GAMMs at both SNR levels (Figure C3). The Q-Q plots showed that residuals were approximately normally distributed, with only mild deviations in the tails, consistent with expectations for reaction time data. Residuals vs. fitted values plots showed no discernible structure, suggesting no major violations of homoscedasticity. ACF plots revealed no notable autocorrelation, indicating that the models

adequately accounted for temporal structure. These diagnostics support the appropriateness of the models for exploratory purposes.

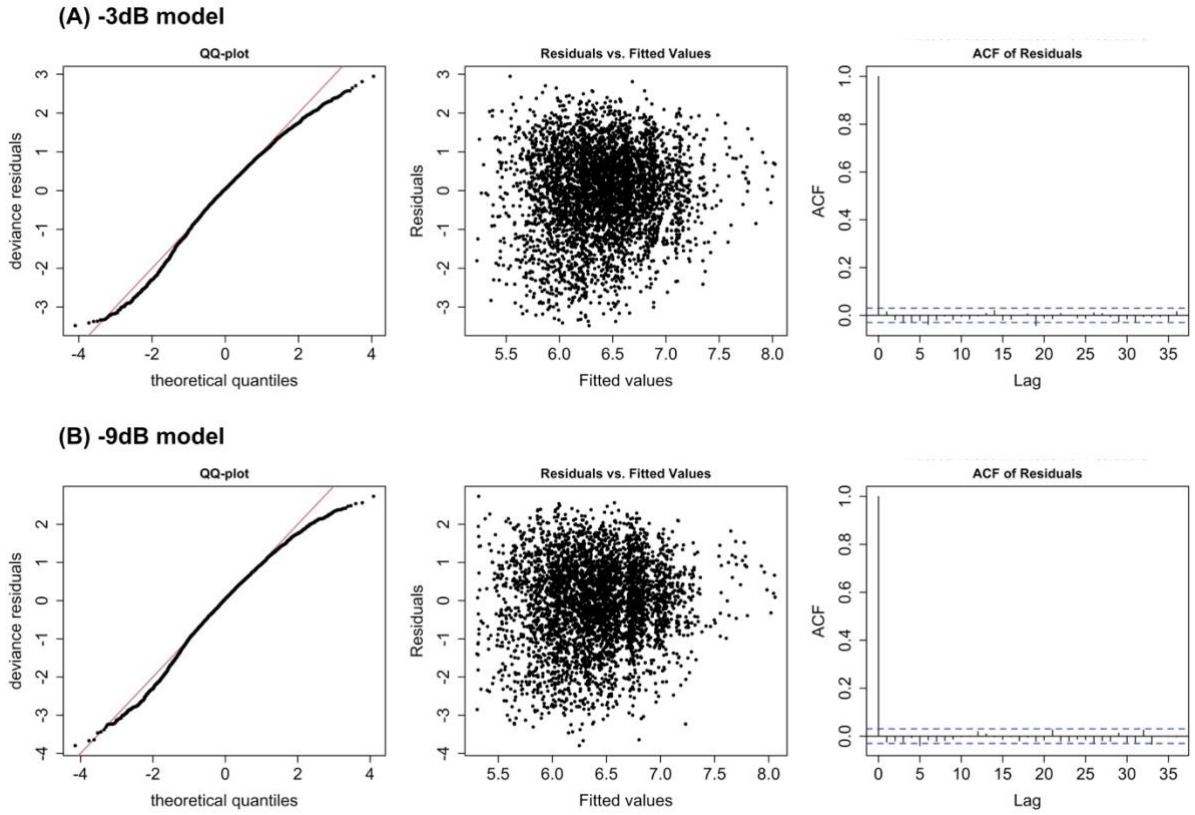

Figure C3. RTs model diagnostics for the selected GAMM for each SNR level.

C2.2 Results

For the −3 dB model, the parametric coefficients revealed a significant difference in RTs between the two cue conditions. RTs were significantly longer in the no-cue condition compared to the both-cues condition ($\beta = 0.25$, $p < .001$), suggesting that listeners responded more slowly when acoustic cues were absent. The smooth term for trial order was not statistically significant (*edf* = 1.00, *Ref.df* = 1.00, $F = 2.45$, $p = .118$), indicating no consistent change in RTs across trials when averaged across all participants and conditions. This likely reflects the limitations of the simpler model structure selected for the −3 dB condition, which did not include cue- or group-specific smooths due to concerns about overfitting. Consequently, the model could not capture potential trial-wise differences across conditions or groups. The final model explained approximately 24.2% of the deviance (*adjusted* $R^2 = 0.263$).

For the −9 dB model, the parametric coefficients revealed significant differences in RTs relative to the baseline (autistic group, both-cues condition). RTs were significantly longer in the no-cue condition for both the autistic group ($\beta = 0.35$, $p < .001$) and the non-autistic group ($\beta = 0.20$, $p = .042$). These findings align with the −3 dB results and further suggest that both groups responded more slowly when acoustic cues were absent. There was no significant RT difference between groups in the both-cues condition ($\beta = –0.05$, $p = .635$), indicating comparable response latencies when both cues were available. Analysis of the smooth terms revealed a significant decrease in RTs over trials only in the autistic group's both-cues condition (*edf* = 1.00, *Ref.df* = 1.00, $F = 3.90$, $p = .048$). The final model explained approximately 26% of the deviance (*adjusted* $R^2 = 0.282$). Pairwise comparisons were examined using difference plots to evaluate trial-level effects of cue condition and group (Figure C4). No significant group differences were observed in either the both-cues or no-cue condition, indicating comparable RTs between autistic and non-autistic participants within each cue condition. As expected, RTs were consistently faster in the both-cues condition than in the no-cue condition for both groups. This difference remained stable over trials for the non-autistic group, but the autistic group showed a gradual reduction in cue-related RT differences over time. Descriptive trend plots (Figure C5) suggest this may reflect a decrease in RTs across trials in the no-cue condition for autistic participants, potentially indicating adaptation or improved tracking. However, due to reduced and uneven trial counts, these results should be interpreted with caution.

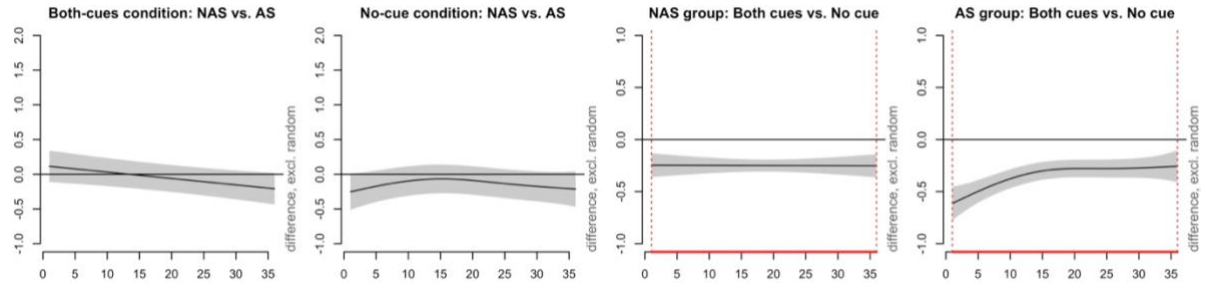

Figure C4. Estimated differences in RTs over trials. The black line represents the estimated difference, with the grey shaded area indicating the 95% confidence interval. Red segments highlight trial ranges where the difference is statistically significant ($p < .05$).

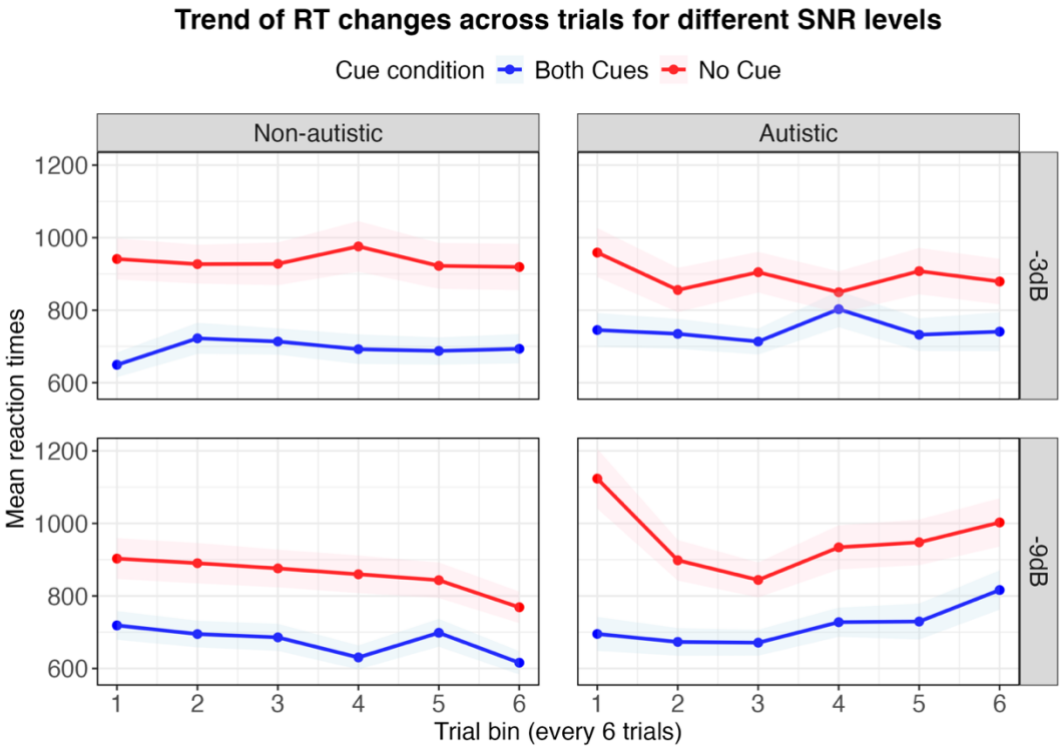

Figure C5. The trend of mean RT changes across trial bins (every 6 trials) for different SNR levels across group and condition with the shaded area indicating the 95% confidence interval.

C3. Full accuracy-based GAMMs: all cue conditions (exploratory)

As outlined earlier, our main analysis focused on the contrast between the no-cue and both-cues conditions due to its direct relevance to our research aims. In response to reviewer feedback, we conducted an exploratory GAMM analysis incorporating all four cue conditions (no cue, gender cue, location cue, both cues) to investigate potential cue-specific effects. We did not include pairwise difference plots in the four-cue model to avoid inflating Type I error. Given the large number of potential comparisons and the absence of inherent correction for multiplicity, such plots could lead to false positives. We therefore focused on interpreting parametric coefficients and smooth terms from the model summary.

C3.1 Procedure of model fitting

The same model-fitting procedure described earlier was used, with the smoothing parameter k increased from 8 to 10 to better capture trial-level variability.

For the −3 dB data, model comparisons showed that m2 fit better than m1 ($\chi^2(6) = 5.25$, $p = .105$), and m3 further improved fit over m2 ($\chi^2(12) = 4.15$, $p = .529$), though neither comparison reached conventional significance. Nonetheless, we retained m3 for consistency across SNR levels and to enable estimation of the Group × Cue interaction.

For the −9 dB data, m2 did not improve model fit over m1 ($\chi^2(6) = 0.04$, $p = 1.00$), but m3 significantly outperformed m2 ($\chi^2(12) = 14.60$, $p = .004$), justifying inclusion of the interaction terms. Thus, m3 was selected as the best-fitting model for the −9 dB data.

Model diagnostics indicated a good fit for both final models (see Figure C6). Deviance residuals were approximately normally distributed with only minor deviations in the tails. Residuals vs. fitted plots showed curved patterns expected for bounded data, and autocorrelation checks revealed no temporal dependency, supporting model adequacy.

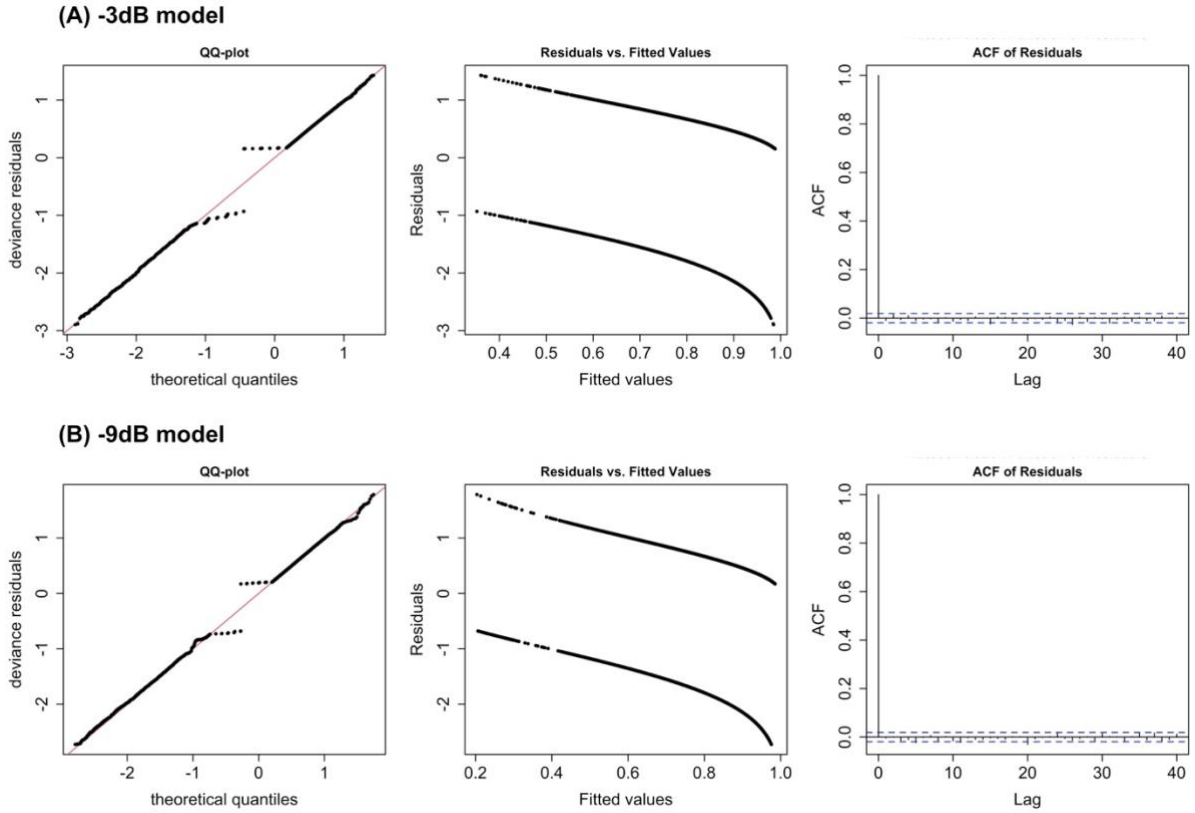

Figure C6. Accuracy model (4 cues) diagnostics for the selected GAMM for each SNR level.

C3.2 Results

At −3 dB, parametric results showed that the only group–cue combinations not significantly different from the baseline (autistic group, both-cues condition) were the non-autistic group in the both-cues and location-cue conditions. All other combinations showed significantly lower accuracy. Specifically, gender cue trials showed reduced accuracy for both the autistic group (β = −0.84, p < .001) and the non-autistic group (β = −0.70, p < .001). The location cue condition yielded a significant accuracy drop for the autistic group (β = −0.35, p = .047). No-cue trials showed the most pronounced reduction for both the autistic (β = −2.31, p < .001) and non-autistic (β = −2.07, p < .001) groups. Trial-level effects showed significant accuracy improvements in the no-cue condition for both groups (autistic: *edf* =2.10, *Ref.df* = 2.61, χ^2= 9.73, p = .020; non-autistic: *edf* =1.00, *Ref.df* = 1.00, χ^2 = 12.61, p < .001), consistent with findings from the main analysis. The model explained 16.3% of the deviance (*adjusted* R^2 = 0.14). These findings reinforce our main analysis by showing that trial-wise improvements are primarily present in the no-cue condition.

At −9 dB, parametric effects indicated that accuracy was significantly lower only in the no-cue condition for both groups (autistic: $\beta = -2.04$, $p < .001$; non-autistic: $\beta = -1.70$, $p < .001$). All other group–cue combinations showed no significant difference from the baseline. For time-varying effects, significant improvements in accuracy over trials were found in the no-cue condition for both groups (autistic: *edf* =1.00, *Ref.df* = 1.00, $\chi^2 = 8.03$, $p = .005$; non-autistic: *edf* =1.00, *Ref.df* = 1.00, $\chi^2 = 10.81$, $p = .001$). Additionally, the autistic group showed significant improvement in the gender cue condition (*edf* = 1.69, *Ref.df* = 2.10, χ^2 =7.54, $p = .026$), and a marginal effect was observed for the non-autistic group (*edf* = 1.26, *Ref.df* = 1.48, $\chi^2 = 5.54$, $p = .054$). The model accounted for 18.2% of the deviance (*adjusted* $R^2 = 0.16$).

Together, these exploratory GAMMs broadly aligned with the main analysis, confirming that trial-level accuracy improvements were most consistent in the no-cue condition. At −9 dB, however, an additional improvement was observed in the gender cue condition for the autistic group (and marginally in the non-autistic group), suggesting some adaptation to gender cue information. Descriptive trend plots are shown in Figure C7.

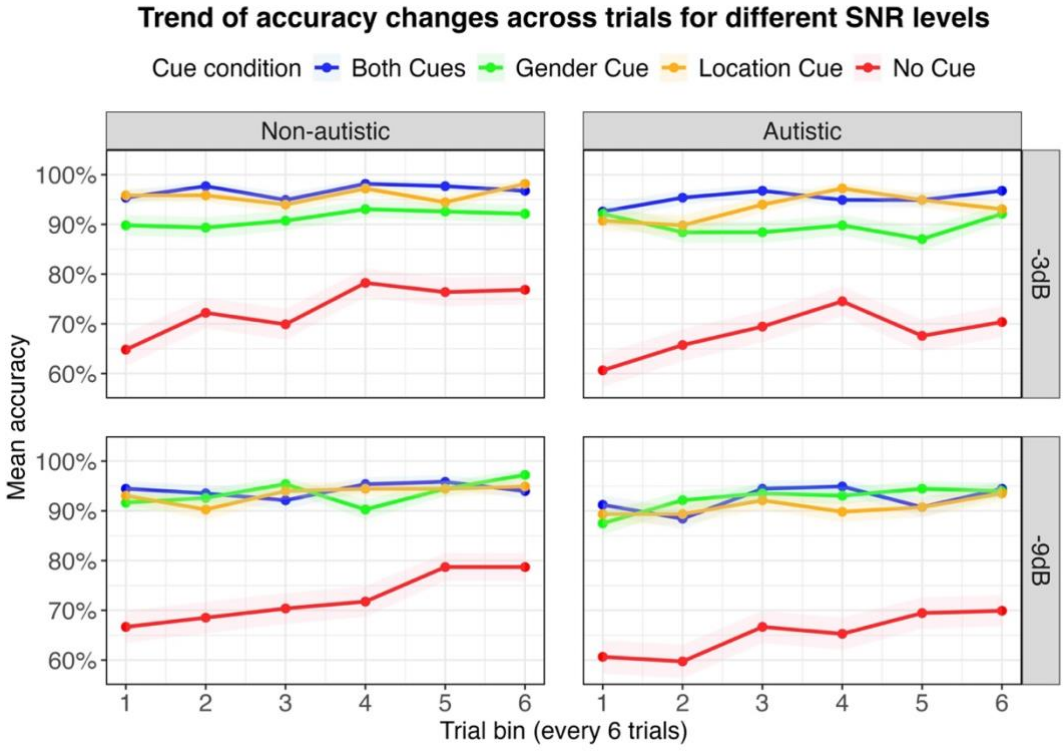

Figure C7. The trend of mean accuracy changes across trial bins (every 6 trials) for different SNR levels across group and condition (4 cues) with the shaded area indicating the 95% confidence interval.

Appendix D. (G)LMM analysis with four cue conditions

In response to the examiner's comment, a supplementary analysis was conducted in which the gender-cue and location-cue conditions were entered into the model as separate factors, rather than being combined into a single one-cue condition. Contrast coding followed the same principles as in the main analysis, with the exception of the cue condition. Three Helmert contrasts were specified to compare the four cue conditions. The first two contrasts were identical to those used in the main analysis and were included to allow direct comparability of results. Specifically, the first contrast compared the no-cue condition with all other cue conditions (Cue contrast 1: no cue = 3/4; gender cue = −1/4; location cue = −1/4; both cues = −1/4), and the second contrast compared the both-cues condition with the single-cue conditions (Cue contrast 2: no cue = 0; gender cue = −1/3; location cue = −1/3; both cues = 2/3). An additional third contrast was included to directly compare the location-cue and gender-cue conditions (Cue contrast 3: no cue = 0; gender cue = 1/2; location cue = −1/2; both cues = 0). The final models for both accuracy and reaction time included all fixed effects and their interactions. Random effects comprised random intercepts for participants and items. For the accuracy model, an additional random slope for Cue contrast 1 (no cue vs. any cue) was included. As shown in Tables D1 and D2, no significant main effects or interactions involving the direct comparison between gender-cue and location-cue conditions were observed. All other significant fixed effects and interactions were consistent with those reported in the main analysis, indicating that separating the gender-cue and location-cue conditions did not alter the pattern of results. Overall, these findings indicate that there was no significant difference between gender-cue and location-cue conditions for either accuracy or reaction time.

Table D1. Results of the GLMM for behavioural accuracy.

Fixed effects	*B*	SE	*z*	OR	95% CI	χ^2	*p*
(Intercept)	3.00	0.12	24.67	—	—	—	—
Group	−0.39	0.18	−2.15	0.68	[0.48, 0.97]	4.44	**.035**
Music	0.76	0.17	4.46	2.14	[1.53, 2.98]	19.07	**< .001**
Cue1	−2.44	0.19	−12.51	0.09	[0.06, 0.13]	129.16	**< .001**
Cue2	0.75	0.22	3.39	2.11	[1.37, 3.26]	11.29	**< .001**
Cue3	−0.19	0.24	−0.82	0.82	[0.52, 1.31]	0.65	.420
Group × Cue1	0.01	0.15	0.10	1.01	[0.76, 1.36]	0.01	.920
Group × Cue2	−0.26	0.19	−1.38	0.77	[0.53, 1.12]	1.75	.186
Group × Cue3	0.31	0.17	1.78	1.36	[0.97, 1.90]	2.93	.087

Music × Cue1	0.23	0.37	0.61	1.26	[0.61, 2.60]	0.34	.558
Music × Cue2	1.47	0.44	3.34	4.36	[1.84, 10.34]	10.87	**< .001**
Music × Cue3	−0.22	0.48	−0.47	0.80	[0.32, 2.04]	0.21	.647
Group × Music	−0.21	0.13	−1.65	0.81	[0.64, 1.04]	2.51	.113
Group × Music × Cue1	0.52	0.22	2.40	1.69	[1.10, 2.58]	5.46	**.019**
Group × Music × Cue2	−0.77	0.38	−2.05	0.46	[0.22, 0.97]	3.90	**.048**
Group × Music × Cue3	0.46	0.34	1.34	1.58	[0.81, 3.10]	1.65	.199

Note. Significant *p*-values are presented in bold. OR = Odds ratio. Odds ratios are obtained by exponentiating the model's log-odds (β) coefficients. 95% confidence intervals (CIs) are similarly derived by exponentiating the CIs of the log-odds estimates.

Table D2. Results of the LMM for times (RTs) of accurate responses.

Fixed effects	***B***	**SE**	***t***	**Exp(β)**	**95% CI**	χ^2	***p***
(Intercept)	6.39	0.05	126.98	—	—	—	—
Group	0.06	0.10	0.66	1.07	[0.88, 1.29]	0.43	.513
Music	−0.04	0.02	−1.71	0.96	[0.92, 1.01]	2.91	.088
Cue1	0.25	0.03	9.32	1.29	[1.22, 1.35]	77.46	**< .001**
Cue2	−0.10	0.03	−3.49	0.91	[0.86, 0.96]	11.91	**< .001**
Cue3	0.01	0.03	0.36	1.01	[0.95, 1.08]	0.13	.719
Group × Cue1	0.00	0.03	0.16	1.00	[0.95, 1.06]	0.03	.870
Group × Cue2	0.00	0.03	−0.06	1.00	[0.95, 1.05]	0.00	.951
Group × Cue3	−0.02	0.03	−0.65	0.98	[0.93, 1.04]	0.42	.516
Music × Cue1	0.01	0.05	0.11	1.01	[0.91, 1.12]	0.01	.916
Music × Cue2	0.03	0.05	0.60	1.03	[0.93, 1.15]	0.36	.549
Music × Cue3	0.04	0.06	0.56	1.04	[0.91, 1.17]	0.31	.578
Group × Music	0.04	0.02	1.69	1.04	[0.99, 1.08]	2.85	.091
Group × Music × Cue1	0.01	0.05	0.17	1.01	[0.91, 1.12]	0.03	.868
Group × Music × Cue2	0.06	0.05	1.15	1.06	[0.96, 1.17]	1.32	.251
Group × Music × Cue3	0.02	0.06	0.38	1.02	[0.91, 1.15]	0.15	.701

Note. Significant *p*-values are presented in bold. Exp(β) values are obtained by exponentiating the fixed-effect coefficients from the linear mixed-effects model predicting log-transformed response times. The resulting values reflect multiplicative effects on raw response times, where values greater than 1 indicate longer response times and values less than 1 indicate shorter response times relative to the reference level. The accompanying 95% confidence intervals (CIs) are derived by exponentiating the intervals for the log-scale estimates.

Appendix E. Correlation analysis with FDR correction

In response to the examiner's suggestion, a supplementary correlation analysis with false discovery rate (FDR) correction was conducted to provide a conservative assessment of the reported associations. As discussed in Chapter 2 (Study 1), the correlation analyses were theory driven rather than exploratory in nature. Nevertheless, FDR correction was applied here to ensure transparency with respect to multiple comparisons.

For the non-autistic group, the association between accuracy in the no-cue condition and pitch discrimination ability remained significant when all participants were included (FDR-corrected $p = .033$). However, this association was no longer significant after excluding an outlier defined as greater than 3 SDs from the mean (FDR-corrected $p = .458$). No other pitch-related correlations in the non-autistic group survived FDR correction.

For the autistic group, the association between digit span and accuracy survived FDR correction for both mean accuracy and accuracy in the no-cue condition (both FDR-corrected $ps = .048$). In addition, the relationship between local-to-global interference and accuracy decline in the presence of background music also remained significant after correction (FDR-corrected $p = .048$).

Appendices for Study 2

Appendix A. Pilot Study

Six native English speakers participated in this pilot study. Target sentences were presented at 60 dB SPL, accompanied by background music (either English or Simlish lyrics) under four signal-to-noise ratio (SNR) conditions: 0 dB, −3 dB, −6 dB, and −9 dB. Participants were asked to listen to all the stimuli planned for the main experiment and to judge the acceptability of each sentence, with their accuracy recorded. As shown in Table A1, accuracy rates at 0 dB and −3 dB approached or exceeded 90%. At −6 dB, accuracy dropped to 82.8% in the English lyrics condition and 80.0% in the Simlish condition. At −9 dB, performance declined further to 68.3% (English) and 72.2% (Simlish), with participants reporting considerable difficulty hearing the target sentences. Based on these results, −6 dB was selected as the SNR level for the main experiment to avoid ceiling effects while still allowing participants to process linguistic information.

Table A1. Mean accuracy rate in the pilot study across conditions and SNR levels.

Condition	**SNR level**	***Mean***	***SD***
English lyrics	0 dB	86.1%	34.7%
	−3 dB	88.9%	31.5%
	−6 dB	82.8%	37.9%
	−9 dB	68.3%	46.6%
Simlish lyrics	0 dB	90.0%	30.1%
	−3 dB	91.1%	28.5%
	−6 dB	80.0%	40.1%
	−9 dB	72.2%	44.9%

Appendix B. Stimuli for Target Speech

The following presents the full set of target sentences used in the experiment. Each sentence ended with a critical word designed to elicit an N400 response. Each sentence ended with a critical word that was either congruent (semantically appropriate within the sentence context) or incongruent (semantically anomalous within the sentence context). The congruent examples are indicated by the expected completion in uppercase (e.g., UNIVERSITY), while the incongruent counterparts are shown in parentheses (e.g., (POSSIBILITY)).

1. She wants to get a degree from a famous UNIVERSITY (POSSIBILITY).
2. The scientist is in the lab doing the EXPERIMENT (ABILITY).
3. The chef used a lot of salt and PEPPER (NOVELS).
4. A large church is called a CATHEDRAL (DIPLOMA).
5. Last night we saw the stars and the MOON (HOLE).
6. To earn money you need a JOB (TALK).
7. My children enjoy singing simple SONGS (BOOKS).
8. Beef and chicken are types of MEAT (CREW).
9. The clothes are cheap because they are on SALE (DIRT).
10. Camels usually live in the DESERT (PROJECT).
11. The light hangs from the CEILING (LADDER).
12. Remote controls can change the TV CHANNEL (QUARTER).
13. Beef and milk come from COWS (BAYS).
14. He parks his cars in his GARAGE (MEMBER).
15. She usually wakes up early in the MORNING (LADY).
16. Cars and factories can cause air POLLUTION (GYMNASTICS).
17. The shop assistant served all the CUSTOMERS (BENEFITS).
18. The north is colder than the SOUTH (PANTS).
19. The passengers thanked the bus DRIVER (SOLDIER).
20. My hair was too long so I got a HAIRCUT (TECHNIQUE).
21. I went to the post office to buy a STAMP (QUIZ).
22. After dinner we asked the waiter for the BILL (COLD).
23. Football and running are types of SPORT (RANGE).
24. The sun can burn your SKIN (PAINT).
25. There are three pictures hanging on the WALL (PAIN).

26. The opposite of midday is MIDNIGHT (KNOWLEDGE).
27. Sick people should see a DOCTOR (BUSINESS).
28. A T rex was a big DINOSAUR (COCONUT).
29. Spring and summer are two of the four SEASONS (WARNINGS).
30. He opened the lock with a KEY (POP).
31. Zebras have many black and white STRIPES (FLUTES).
32. We crossed the river by walking over the BRIDGE (THROAT).
33. Children like pasta with tomato SAUCE (NOON).
34. Every country is run by the GOVERNMENT (MEMORY).
35. We knocked on the front DOOR (CHECK).
36. There are sixty seconds in a MINUTE (HUMAN).
37. In the day we get light from the SUN (FAIR).
38. Bosses should be kind to their EMPLOYEES (ADVENTURES).
39. February is always the shortest MONTH (GIFT).
40. Giraffes have spots and a long NECK (FORM).
41. After his shower he got dried with a TOWEL (CHAT).
42. Trousers and skirts are types of CLOTHES (STEPS).
43. Eggs come from a duck or a CHICKEN (BOYFRIEND).
44. Doctors try to cure dangerous DISEASES (PIANOS).
45. The bride is wearing a white DRESS (CLUB).
46. My shoes are made of brown LEATHER (HOCKEY).
47. Athletes get instructions from their COACH (BLOCK).
48. Friday is my favourite day of the WEEK (GUESS).
49. I get my hair cut by my favourite HAIRDRESSER (PINEAPPLE).
50. In winter there can be very cold WEATHER (CANDY).
51. I keep my wallet in my trouser POCKET (LESSON).
52. She smelled the flowers using her NOSE (CASH).
53. Your sister’s son is your NEPHEW (BLANKET).
54. The boss of a ship is called the CAPTAIN (OFFICE).
55. You wear shoes on your FEET (TRUCKS).
56. Apples and bananas are types of FRUIT (YARD).
57. The popular girl has lots of FRIENDS (THOUGHTS).
58. They gave a prize to the competition WINNER (MODEL).
59. There are eleven players on a football TEAM (FIRE).

60. The baseball player hit the ball with his BAT (ROW).
61. Magicians know a lot of card TRICKS (SCENES).
62. The queen is married to the KING (NEWS).
63. My phone doesn't work because it's run out of BATTERY (COMEDY).
64. These clothes were made by the fashion DESIGNER (RELATION).
65. Villages are smaller than cities and TOWNS (DRINKS).
66. A big sea is called an OCEAN (APPLE).
67. Rain falls from big black CLOUDS (SNACKS).
68. Famous people are also called stars or CELEBRITIES (VARIETIES).
69. Every morning he washes in the sink in the BATHROOM (FINAL).
70. Please don't tell anyone my SECRET (COLLEGE).
71. The south is warmer than the NORTH (CHOICE).
72. The student makes a lot of spelling MISTAKES (PARTNERS).
73. Honest people always tell the TRUTH (LUCK).
74. On her birthday she ate chocolate CAKE (BELLS).
75. The girl likes toast with strawberry JAM (CORN).
76. My favourite flowers are red ROSES (GASES).
77. In rush hour there is a lot of TRAFFIC (WINTER).
78. Trees grow lots of green LEAVES (GAPS).
79. Footballers are happy when they score a GOAL (TUNE).
80. I can't read your terrible HANDWRITING (LUXURY).
81. The policeman shot the thief with his GUN (PART).
82. The opposite of war is PEACE (SNOW).
83. Someone who makes bread is called a BAKER (SINGER).
84. He cuts vegetables with a sharp KNIFE (SQUARE).
85. The photographer took pictures with a CAMERA (PRISON).
86. Your mum and dad are your PARENTS (CLINICS).
87. The people who live near you are your NEIGHBOURS (MELONS).
88. The husband bought flowers for his WIFE (STUFF).
89. The class went on a history trip to the MUSEUM (SOLUTION).
90. Every night I read my children a STORY (COUPLE).
91. We went sailing on the lake in our new BOAT (MESS).
92. He typed using the computer's KEYBOARD (BROCHURE).
93. I write my homework sitting at my DESK (CROWD).

94. We waited an hour in the long QUEUE (BLOG).
95. The orchestra played some classical MUSIC (SERVICE).
96. When we go camping we sleep in a TENT (GUM).
97. Many people died in the second world WAR (TOP).
98. The sports team built a big new STADIUM (GRANDDAUGHTER).
99. Clothes for sleeping in are called PYJAMAS (RECYCLING).
100. The little girl made a new dress for her DOLL (MEAL).
101. A sandwich has two pieces of BREAD (SNAKE).
102. I grow beautiful flowers in my GARDEN (PAINTING).
103. The prince's parents are the king and QUEEN (SUIT).
104. The king lives in an old stone CASTLE (PENNY).
105. Get out of the ocean if you see a SHARK (POUND).
106. He often buys his wife a bunch of FLOWERS (TURKEYS).
107. Workers get instructions from their BOSS (FOOD).
108. Sick animals are cared for by a VET (DISK).
109. Jungles have a hot and wet CLIMATE (BACKPACK).
110. Students go to university to get a DEGREE (POEM).
111. The day is light but the night is DARK (GROUP).
112. We heard rain falling on the house's ROOF (TRADE).
113. Women sometimes wear nice smelling PERFUME (SUNSHINE).
114. After the main course we ordered DESSERT (REPAIRS).
115. The wife cooked dinner for her HUSBAND (PROBLEM).
116. Birds fly by using their WINGS (CAPS).
117. I'll call you if you give me your telephone NUMBER (PERSON).
118. Go to the dentist if you have TOOTHACHE (SURNAMES).
119. A very small town is called a VILLAGE (PLANET).
120. The thick book had five hundred PAGES (GLASSES).
121. There are billions of websites on the INTERNET (EMBASSY).
122. The model wore a top and a short SKIRT (BREEZE).
123. She ate her food with a knife and FORK (CHEEK).
124. Zoos have a lot of dangerous ANIMALS (ENEMIES).
125. The competition winner received a PRIZE (CHAIN).
126. Astronauts use rockets to go to SPACE (HEAT).
127. Penguins and ducks are types of BIRD (VAN).

128. Carrots and potatoes are types of VEGETABLE (LEMONADE).
129. Children should never talk to STRANGERS (TOILETS).
130. The girl brushed her long blonde HAIR (BAR).
131. Asia is not a country, it's a CONTINENT (PHARMACY).
132. Really scary dreams are called NIGHTMARES (BASEBALLS).
133. Children love playing with noisy TOYS (PORTS).
134. Tablets and pills are types of MEDICINE (VIDEO).
135. A jacket isn't as warm as a long COAT (LAKE).
136. The bride and groom had a traditional WEDDING (ARMY).
137. Your eyes and mouth are part of your FACE (DAD).
138. In Asia people eat a lot of RICE (CAVES).
139. You can get fit by working out at the GYM (FLAG).
140. The largest animal in Africa is the ELEPHANT (ORANGE).
141. The tourists are visiting the capital CITY (POWER).
142. Keep your neck warm with a long SCARF (BULB).
143. After dinner we left the waiter a small TIP (BONE).
144. Cars and buses are types of VEHICLE (PRISONER).
145. I often borrow books from the LIBRARY (CHOCOLATE).
146. You can't control the beating of your HEART (SIDE).
147. Managers often earn a high SALARY (FESTIVAL).
148. In the past teachers wrote on the BLACKBOARD (CHECKOUT).
149. People with toothache should visit the DENTIST (SUNSET).
150. The rock musician plays the GUITAR (REWARD).
151. Carrying a heavy bag can hurt your BACK (SET).
152. To visit some countries you need a VISA (TOPIC).
153. I prefer typing to writing with a PEN (DUCK).
154. The traditional furniture is made of WOOD (GOLF).
155. A holiday after your wedding is called a HONEYMOON (CHAMPION).
156. Keep your feet warm with wool SOCKS (TINS).
157. The funniest people at the circus are the CLOWNS (DUST).
158. Forests have many tall green TREES (KICKS).
159. The teacher helps the children in her CLASS (FAULT).
160. My favourite fish is grilled pink SALMON (COLA).
161. In the summer we go to the swimming POOL (CREAM).

162. At the gym I put my things in the LOCKER (TIGER).
163. When it is hot you should drink lots of WATER (TROUBLE).
164. Before we ordered we looked at the restaurant MENU (BUCKET).
165. She called the doctor to make an APPOINTMENT (EXAMPLE).
166. The postman delivered the important LETTER (KISSES).
167. Sweets and biscuits have a lot of SUGAR (MONKEYS).
168. Children are punished for their bad BEHAVIOUR (DEPARTURE).
169. People enjoy reading their birthday CARDS (BRAINS).
170. Reading and photography are common HOBBIES (CANDLES).
171. He carefully filled out the job APPLICATION (ENTERTAINMENT).
172. Russia is the world's largest COUNTRY (MOMENT).
173. Clothes fit best if they are the right SIZE (FEAR).
174. Doctors and nurses work in a HOSPITAL (DETECTIVE).
175. The criminal was caught by the POLICE (WATCHES).
176. There are books and CDs on the SHELF (DRUM).
177. Your heart moves blood around your BODY (REASON).
178. He called the restaurant to make a RESERVATION (GENERATION).
179. She speaks with a strong Scottish ACCENT (ENTRY).
180. The nervous fans watched the football MATCH (FILE).

Appendices for Study 3

Appendix A. Pilot Study

The signal-to-noise ratio (SNR) level was set to 0 dB based on a pilot study. In this study, six native English speakers participated in this pilot study. Participants were presented with target sentences under four SNR conditions (6 dB, 3 dB, 0 dB, and −3 dB) and were asked to judge sentence acceptability, with accuracy recorded. Table A1 shows the results of the pilot study. At 6 dB and 3 dB, accuracy rates exceeded 90%. At 0 dB, accuracy dropped to 88.1% in the babble condition and 85.2% in the single-talker condition. At −3 dB, accuracy declined further to 75.6% (babble) and 78.9% (single speaker), with participants reporting significant difficulty hearing the target sentences. Based on these results, 0 dB was selected as the SNR level for the main experiment to prevent ceiling effects while ensuring participants could process linguistic information effectively.

Table A1. Accuracy rate in the pilot study across conditions and SNR levels.

Condition	SNR level	*Mean (SD)*
Babble	6 dB	94.4% (22.9%)
	3 dB	92.2% (26.8%)
	0 dB	88.1% (32.4%)
	−3 dB	75.6% (43.1%)
Speech	6 dB	94.4% (22.9%)
	3 dB	90.4% (29.6%)
	0 dB	85.2% (35.6%)
	−3 dB	78.9% (40.1%)

Appendix B. Stimuli for Target Speech

The following presents the full set of target sentences used in the experiment. Each sentence ended with a critical word designed to elicit an N400 response. Each sentence ended with a critical word that was either congruent (semantically appropriate within the sentence context) or incongruent (semantically anomalous within the sentence context). The congruent examples are indicated by the expected completion in uppercase (e.g., CALCULATOR), while the incongruent counterparts are shown in parentheses (e.g., (MILLIMETRE)).

1. In maths class we do sums using a CALCULATOR (MILLIMETRE).
2. Someone who doesn't eat meat is called a VEGETARIAN (DOCUMENTARY).
3. For breakfast children eat toast or CEREAL (LITERATURE).
4. Your aunt and uncle's children are your COUSINS (PROGRAMS).
5. Flats don't have gardens but they have BALCONIES (LOTTERIES).
6. You should put your rubbish in the BIN (RAIL).
7. The chef cooks in the hot KITCHEN (STATION).
8. When there is snow in the mountains we go SKIING (BANKING).
9. Eat breakfast in the morning and dinner in the EVENING (FIGURE).
10. The day after today is called TOMORROW (PROFESSOR).
11. I always look up new words in a DICTIONARY (BABYSITTER).
12. We went to visit our grandfather and GRANDMOTHER (PROPERTY).
13. He rides through the desert on a CAMEL (DISCO).
14. Turn it on using the remote CONTROL (REPORT).
15. In tennis you hit the ball with a RACKET (PUZZLE).
16. The mother and father have four CHILDREN (PIECES).
17. Every day I write my thoughts in my relative DIARY (RELATIVE).
18. Tourists read about the sights in their GUIDEBOOK (SNOWBOARD).
19. The footballer kicked the round shop BALL (SHOP).
20. There was lots of rain and lightning during the STORM (CLIFF).
21. The car has space for a driver and three PASSENGERS (SIGNATURES).
22. A book about someone's life is called a BIOGRAPHY (CURRICULUM).
23. They went to watch a play at the THEATRE (DOCUMENT).
24. Doctors choose medicine and then write a PRESCRIPTION (TOMATO).
25. We work during the week and relax at the WEEKEND (MAGIC).

26. There are hundreds of countries in the WORLD (THING).
27. The border guard put a stamp in my PASSPORT (SOFTWARE).
28. Grandfather has a moustache and a long BEARD (SHEET).
29. The customers queued in a straight LINE (CUT).
30. People sleep with their head on a PILLOW (FARMER).
31. They are drinking coffee in the CAFÉ (FERRY).
32. In some zoos animals live in small CAGES (PURSES).
33. One hundred years is called a CENTURY (BASKETBALL).
34. She made a special cake for her son's BIRTHDAY (MESSAGE).
35. Someone who owns a meat shop is called a BUTCHER (NECKLACE).
36. He keeps money in a leather WALLET (PHOTO).
37. The two boys are identical TWINS (DIALS).
38. I chose a recipe and bought all the INGREDIENTS (EXAMINERS).
39. Students have to write a lot of long ESSAYS (OLIVES).
40. There are three children and two parents in the FAMILY (RADIO).
41. The carpet is covering the FLOOR (TRIP).
42. We planned our journey using a gate MAP (GATE).
43. The mess is cleaned up by the CLEANER (SPEAKER).
44. The actors performed on the theatre's STAGE (COOK).
45. Children usually write with a pen or PENCIL (COMIC).
46. After school children must do their HOMEWORK (PLASTIC).
47. Clean your teeth with toothpaste and a TOOTHBRUSH (NIGHTCLUB).
48. Dollars and pounds are different types of CURRENCY (SCENERY).
49. There are many trees in the FOREST (RABBIT).
50. Scientists do experiments in a LABORATORY (CERTIFICATE).
51. Doctors take care of their PATIENTS (CREDITS).
52. In China the most famous drink is green TEA (SOUL).
53. Circles and squares are different SHAPES (FLATS).
54. Patients are cared for by doctors and NURSES (CROSSES).
55. The little girl loves her teddy BEAR (SEA).
56. People who design buildings are called ARCHITECTS (INVENTIONS).
57. Famous chefs usually work at expensive RESTAURANTS (BATTLES).
58. The new chemical was discovered by a SCIENTIST (CAPITAL).
59. The nasty cat caught the little MOUSE (JET).

60. I passed my test and got my driving LICENCE (DISCOUNT).
61. Footballers wear a t-shirt and SHORTS (BEANS).
62. Girls quickly grow up and become WOMEN (TODAY).
63. We checked in at the hotel RECEPTION (CONCLUSION).
64. A very bad cold is called the FLU (SKILL).
65. Cook the chicken at a high TEMPERATURE (QUALITY).
66. A zebra is similar to a HORSE (STAR).
67. Poor people don't have a lot of MONEY (PLACES).
68. Someone who writes for a newspaper is a JOURNALIST (STRAWBERRY).
69. Boxes are made of strong paper called CARDBOARD (LUNCHTIME).
70. You wash your hair using SHAMPOO (LETTUCE).
71. There are twenty-six letters in the English ALPHABET (WATERFALL).
72. Tomorrow there will be rain with thunder and LIGHTNING (TENNIS).
73. The walkers followed the forest PATH (BRIDE).
74. You have eight fingers and two THUMBS (LINKS).
75. Our house has two bathrooms and three BEDROOMS (CONTRACTS).
76. The pilot got the plane ready for the next FLIGHT (TRACK).
77. On each foot you have five TOES (PANS).
78. Letters are delivered by the POSTMAN (SPINACH).
79. History is lots of students' favourite SUBJECT (PURPOSE).
80. Buses and trains are types of public TRANSPORT (LUGGAGE).
81. The couple have two girls and a BOY (FINE).
82. The big university has thousands of clever STUDENTS (BOTTOMS).
83. The alphabet has five vowels and twenty-one CONSONANTS (MOTORWAYS).
84. Your mother's sister is your AUNT (ENGINE).
85. He holds up his trousers with a BELT (TONGUE).
86. I can't see because the TV has a small SCREEN (BRUSH).
87. Monkeys like to eat yellow BANANAS (POLICEMEN).
88. Stealing and killing people are types of CRIME (SPOT).
89. Biology and chemistry are types of SCIENCE (CAREER).
90. Most governments have a prime minister or a PRESIDENT (COMPANY).
91. Cats and dogs are popular PETS (MILES).
92. The journalist wrote a long ARTICLE (UNIVERSE).
93. The children are playing a fun GAME (RIDE).

94. Smoking is a very bad HABIT (REVIEW).
95. In the morning we drink tea or COFFEE (PRESENTS).
96. She has a gold ring on her FINGER (MARKET).
97. Football teams always have eleven PLAYERS (TURNINGS).
98. If you are lost, ask someone for DIRECTIONS (REPORTERS).
99. Next month the pregnant lady will have her BABY (MILLION).
100. Every day the chicken lays an EGG (INCH).
101. The desk has four wooden LEGS (BANDS).
102. People eat soup or cereal using a SPOON (FILM).
103. The singer has a beautiful VOICE (LIST).
104. An architect's job is to design BUILDINGS (GIRLFRIENDS).
105. For lunch I usually eat a cheese SANDWICH (JUNGLE).
106. In the morning people eat toast for BREAKFAST (MARRIAGE).
107. Children ask their teachers lots of QUESTIONS (BROTHERS).
108. Money you pay to the government is called TAX (RAP).
109. We showed our passports when we crossed the BORDER (FEVER).
110. A baby cat is called a KITTEN (POSTER).
111. In the morning she drinks orange JUICE (COAST).
112. Draw a straight line using the RULER (PARROT).
113. Lots of teachers work in that SCHOOL (CHANCE).
114. Sweet honey is made by BEES (HUTS).
115. Tourists often send their friends a POSTCARD (CLASSROOM).
116. In some countries schoolchildren wear a UNIFORM (LOCATION).
117. Children enjoy seeing clowns at the CIRCUS (DISTRICT).
118. Some children go to school on a yellow BUS (CLOCK).
119. The musician plays the piano and other INSTRUMENTS (ANNOUNCEMENTS).
120. Please put the flowers in a VASE (GRILL).
121. The painting is in a wooden FRAME (TUBE).
122. I take sandwiches to work to eat for LUNCH (FRONT).
123. Use the lift or walk up the STAIRS (TERM).
124. People you work with are your COLLEAGUES (PAINTERS).
125. Those bees make delicious sweet HONEY (MATTER).
126. When you eat, food goes down into your STOMACH (MACHINE).
127. The Italian restaurant sells slices of PIZZA (FOOTBALL).

128. Your hand is connected to your ARMS (EAST).
129. The tourists are staying in an expensive HOTEL (SYSTEM).
130. I drink coffee with sugar and MILK (STAFF).
131. Meat from a cow is called BEEF (ZONE).
132. There are sixty minutes in an HOUR (AIR).
133. He's drinking water out of the tall GLASS (BREATH).
134. My dentist looks after my TEETH (LAMPS).
135. Mice love to eat smelly CHEESE (VIEWS).
136. The tour guide is talking to a group of TOURISTS (BISCUITS).
137. We used the bridge to cross the RIVER (SUMMER).
138. She's cutting the paper using sharp SCISSORS (PIRATES).
139. People in England and China speak different LANGUAGES (CHARACTERS).
140. In the morning father always reads the NEWSPAPER (STUDIO).
141. A pilot's job is to fly a PLANE (FOOL).
142. The child loves his mother and FATHER (RUNNING).
143. The mother loves her son and DAUGHTER (SURPRISE).
144. Her wedding ring is made of GOLD (STORES).
145. We hear sound using our EARS (HALLS).
146. Children between thirteen and nineteen are TEENAGERS (CALENDARS).
147. He called the restaurant to book a TABLE (MIDDLE).
148. The doctor told me to quit SMOKING (DRAMA).
149. I buy all my food at the big SUPERMARKET (GRADUATION).
150. Cars and buses have four WHEELS (POTS).
151. Please help me to open the jam JAR (PHRASE).
152. The model looked at herself in the MIRROR (SECTION).
153. My car was fixed by a MECHANIC (TRANSLATION).
154. The student studied and got good GRADES (SHOCKS).
155. Penguins eat a lot of FISH (GUARDS).
156. On sunny days there are no clouds in the SKY (FIRM).
157. Silver, gold and iron are different types of METAL (CABLE).
158. Wine is usually made from GRAPES (WOOL).
159. The tea is in a small white CUP (ROLL).
160. Cars are made in a FACTORY (PERFORMANCE).
161. History is learning about what happened in the PAST (BREAK).

162. If you are hot you can open the WINDOW (FUTURE).
163. Your father's brother is your UNCLE (PICTURE).
164. I called the hotel to book a ROOM (CASE).
165. Keep your head warm by wearing a HAT (PARK).
166. Login using your username and PASSWORD (BRACELET).
167. I like driving instead of WALKING (DUTY).
168. Most vegetables are grown on a FARM (SPEECH).
169. Chefs are very good at COOKING (DOLLARS).
170. Some people wear blue trousers called JEANS (GROOMS).
171. I like jewellery made of gold more than SILVER (TALENT).
172. Children's films made of drawings are called CARTOONS (TRUMPETS).
173. Children quickly grow up and become ADULTS (OPTIONS).
174. Babies drink from a plastic BOTTLE (LEVEL).
175. Schoolchildren wear trousers and a white SHIRT (BUNCH).
176. Eating fruit and exercising are good for your HEALTH (PRICE).
177. The class listened to their TEACHER (DANGER).
178. You wear a hat on your HEAD (GIRL).
179. A small mountain is called a HILL (MATE).
180. The mother made her sick child some chicken SOUP (TEAR).

Appendix C. Complementary (G)LMM Analyses

Because the two groups differed significantly in their Auditory Attention and Discomfort (AAD) scores, complementary (G)LMM analyses were conducted to assess whether this group difference affected our results. Specifically, I re-ran the models for behavioural accuracy, TRF measures (the peak amplitude and latency for TRF components, and model fit r-value), and N400 responses (amplitude and onset latency), this time including AAD score as a covariate. The fixed and random effects structures remained identical to those used in the main analyses. Full results are provided in Tables C1-C5.

Table C1. Results of the GLMM for behavioural data including AAD score as a covariate.

Fixed effects	***β***	**SE**	**z**	χ^2	***p***	**OR**
(Intercept)	4.49	0.36	12.30	—	—	—
Group	−0.33	0.18	−1.80	3.17	.075	0.72
Masker1	−1.13	0.12	−9.34	58.85	**<. 001**	0.32
Masker2	0.47	0.09	5.37	21.56	**<. 001**	1.61
Group × Masker1	0.08	0.22	0.36	0.10	.752	1.07
Group × Masker2	0.12	0.16	0.76	0.61	.434	1.13
AAD score	−0.01	0.00	−2.82	7.56	**.006**	0.99

Note. The *p*-values of significant fixed effects are presented in bold. Model structure: glmer(Accuracy ~ 1 + Group × Masker1 + Group × Masker2 + (1 + Masker1 + Masker2 | Subject) + (1 | Item)). OR: Odds ratios.

Table C2. Results of the LMM for TRF amplitudes and latency including AAD score as a covariate.

	Fixed effects	***β***	**SE**	**t**	χ^2	***p***	η_p^2
	(Intercept)	0.66	0.15	4.27	—	—	—
	Group	0.06	0.10	0.59	0.38	.537	0.01
P1	Masker1	0.20	0.05	3.86	13.36	**< .001**	0.19
amplitude	Masker2	0.35	0.06	5.80	26.84	**< .001**	0.35
	Group × Masker1	0.08	0.10	0.79	0.61	.433	0.01
	Group × Masker2	0.05	0.12	0.45	0.20	.656	0.00
	AAD score	0.00	0.00	−0.83	0.67	.413	0.01
	(Intercept)	77.66	3.60	21.597	—	—	—
	Group	1.70	1.96	0.87	0.74	.389	0.01
P1	Masker1	−1.95	1.69	−1.15	1.31	.252	0.02
latency	Masker2	−2.82	1.82	−1.55	2.35	.125	0.04
	Group × Masker1	2.72	3.38	0.81	0.65	.421	0.01
	Group × Masker2	−1.73	3.64	−0.48	0.23	.635	0.00
	AAD score	−0.04	0.04	−1.01	0.95	.331	0.02

	(Intercept)	−0.21	0.14	−1.50	—	—	—
	Group	0.16	0.08	2.13	4.26	**.039**	0.07
N1	Masker1	−0.21	0.07	−3.26	9.78	**.002**	0.15
amplitude	Masker2	−0.22	0.05	−4.13	15.05	**< .001**	0.22
	Group × Masker1	0.04	0.13	0.31	0.09	.759	0.00
	Group × Masker2	−0.04	0.11	−0.42	0.17	.678	0.00
	AAD score	0.00	0.00	−0.79	0.59	.444	0.00
	(Intercept)	168.78	7.98	21.15	—	—	—
	Group	7.84	4.35	1.80	3.13	.077	0.05
N1	Masker1	19.43	2.98	6.52	32.33	**< .001**	0.41
latency	Masker2	−1.21	3.01	−0.40	0.16	.687	0.00
	Group × Masker1	−5.75	5.96	−0.96	0.92	.337	0.01
	Group × Masker2	−7.34	6.01	−1.22	1.47	.225	0.02
	AAD score	0.02	0.09	0.25	3.16	.075	0.00
	(Intercept)	0.13	0.15	0.88	—	—	—
	Group	−0.08	0.10	−0.76	0.57	.451	0.01
P2	Masker1	−0.53	0.06	−8.63	48.95	**< .001**	0.55
amplitude	Masker2	−0.11	0.05	−2.18	4.59	**.032**	0.07
	Group × Masker1	0.35	0.12	2.79	7.35	**.007**	0.11
	Group × Masker2	0.06	0.10	0.66	0.43	.513	0.01
	AAD score	0.00	0.00	0.45	0.19	.664	0.00
	(Intercept)	0.53	0.22	2.46	—	—	—
	Group	−0.24	0.13	−1.93	3.62	.057	0.06
N1-P2	Masker1	−0.32	0.09	−3.52	11.27	**< .001**	0.17
amplitude	Masker2	0.11	0.07	1.50	2.20	.138	0.03
	Group × Masker1	0.31	0.18	1.67	2.73	.098	0.04
	Group × Masker2	0.11	0.15	0.73	0.53	.465	0.01
	AAD score	0.00	0.00	−0.08	0.00	.935	0.00

Note. The *p*-values of significant effects are presented in bold. The same model was used for all analyses of amplitude and latency: lmer(Amplitude/Latency ~ 1 + Group × Masker1 + Group × Masker2 + (1 + Masker1 + Masker2 | Subject)).

Table C3. Results of the LMM for *r*-values of TRF modelling including AAD score as a covariate.

Fixed effects	***β***	**SE**	**t**	χ^2	***p***	η_p^2
(Intercept)	0.09	0.02	3.66	—	—	—
Group	−0.01	0.01	−0.83	0.69	.112	0.01
Masker1	−0.02	0.01	−2.51	5.99	**.014**	0.09
Masker2	0.00	0.01	−0.34	0.12	.733	0.00
Group × Masker1	0.00	0.02	−0.30	0.09	.761	0.00
Group × Masker2	0.01	0.02	0.86	0.73	.393	0.01
AAD score	0.00	0.00	0.49	0.23	.632	0.00

Note. The *p*-values of significant fixed effects are presented in bold. Model structure: lmer(r-value ~ 1 + Group × Masker1 + Group × Masker2 + (1 + Masker1 + Masker2 | Subject)).

Table C4. Results of the LMM for N400 onset latency including AAD score as a covariate.

Fixed effects	***β***	**SE**	***t***	χ^2	***p***	η_p^2
(Intercept)	208.20	5.31	39.23	—	—	—
Group	6.33	2.96	2.14	4.43	**.035**	0.07
Masker1	−0.65	2.47	−0.26	0.07	.793	0.00
Masker2	0.15	3.35	0.04	0.00	.965	0.00
Group × Masker1	1.79	4.93	0.36	0.13	.717	0.00
Group × Masker2	−6.87	6.69	−1.03	1.04	.307	0.02
AAD score	0.00	0.06	−0.01	0.00	.990	0.00

Note. The *p*-values of significant fixed effects are presented in bold. Model structure: lmer(Latency ~ 1 + Group × Masker1 + Group × Masker2 + (1 + Masker1 + Masker2 | Subject)).

Table C5. Results of the LMM for N400 amplitude including AAD score as a covariate.

Fixed effects	***β***	**SE**	**t**	χ^2	***p***	η_p^2
(Intercept)	−0.99	0.42	−2.38	—	—	—
Group	0.54	0.23	2.38	5.42	**.020**	0.08
Sentence	1.36	0.17	7.92	43.36	**< .001**	0.50
Masker1	−0.02	0.01	−2.48	6.13	**.013**	0.00
Masker2	0.21	0.01	25.09	628.9	**< .001**	0.00
Group × Masker1	−0.02	0.01	−1.54	2.36	.125	0.00
Group × Masker2	−0.03	0.02	−1.71	2.91	.088	0.00
Group × Sentence	−0.25	0.34	−0.74	0.54	.463	0.01
Masker1 × Sentence	0.23	0.01	16.20	206.28	**< .001**	0.00
Masker2 × Sentence	−0.24	0.02	−14.45	208.77	**< .001**	0.00
Group × Masker1 × Sentence	−0.56	0.03	−19.57	382.8	**< .001**	0.00
Group × Masker2 × Sentence	0.54	0.03	16.24	263.71	**< .001**	0.00

AAD score	0.01	0.00	1.08	1.14	.285	0.02

Note. The *p*-values of significant effects are presented in bold. Model structure: lmer(Amplitude ~ 1 + Group × Sentence × Masker1 + Group × Sentence × Masker2 + (1 + Sentence | Subject)).