University of Reading

Listening in a noisy world: the impact of acoustic cues and background music on speech perception in autism

Article

Published Version

Creative Commons: Attribution 4.0 (CC-BY)

Open Access

Li, J., Sujawal, M., Bernotaite, Z., Cunnings, I. ORCID: https://orcid.org/0000-0002-5318-0186 and Liu, F. ORCID: https://orcid.org/0000-0002-7776-0222 (2026) Listening in a noisy world: the impact of acoustic cues and background music on speech perception in autism. Autism, 30 (1). pp. 134-149. ISSN 1461-7005 doi: 10.1177/13623613251376484 Available at https://centaur.reading.ac.uk/124001/

It is advisable to refer to the publisher's version if you intend to cite from the work. See Guidance on citing.

To link to this article DOI: http://dx.doi.org/10.1177/13623613251376484

Publisher: SAGE Publications

All outputs in CentAUR are protected by Intellectual Property Rights law, including copyright law. Copyright and IPR is retained by the creators or other copyright holders. Terms and conditions for use of this material are defined in the End User Agreement.

www.reading.ac.uk/centaur

University of Reading

CentAUR

Central Archive at the University of Reading

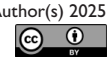
Reading's research outputs online

Check for updates

Original Article

autism

Listening in a noisy world: The impact of acoustic cues and background music on speech perception in autism

Autism
1–16
© The Author(s) 2025

Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/13623613251376484
journals.sagepub.com/home/aut
Sage

Jiayin Li, Maleeha Sujawal, Zivile Bernotaite, Ian Cunnings and Fang Liu

Abstract

Recognising speech in noise involves focusing on a target speaker while filtering out competing voices and sounds. Acoustic cues, such as 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, recruited in the United Kingdom, 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.

Lay abstract

This study examined how autistic and non-autistic adults understand speech when other voices or music were playing in the background. Participants focused on one main speaker while another voice played simultaneously. Sometimes, the second voice differed from the main one in gender or where the sound was coming from. These differences made it easier to tell the voices apart and understand what the main speaker was saying. Both autistic and non-autistic participants did better when these differences were present. But autistic individuals struggled more when the two voices were the same gender and came from the same location. Background music also made it harder to understand speech for everyone, but it especially affected autistic participants who tended to focus more on small details. These findings help us understand how autistic individuals process speech in noisy environments and could lead to better ways to support communication.

Keywords

acoustic cue, autism, background music, speech-in-noise processing

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,

University of Reading, UK

Corresponding author:
Fang Liu, School of Psychology & Clinical Language Sciences, University of Reading, Earley Gate, Reading RG6 6AL, UK.
Email: f.liu@reading.ac.uk

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 autistic individuals' 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 day-to-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-speaker scenario was associated with lower intelligence quotient (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 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.

This 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 SNR 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 during 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 et al., 2008; Molnar-Szakacs & Heaton, 2012; Ouimet et al., 2012).

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

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	0.942	–0.01
Autistic traits (AQ)	37.89 (7.20)	16.08 (7.53)	1258	< 0.001	0.94
Musical training years	5.01 (6.67)	5.67 (7.67)	608.5	0.647	–0.06
Non-verbal reasoning (RSPM raw score)	54.00 (3.68)	54.08 (3.26)	669.5	0.810	–0.03
Non-verbal reasoning (RSPM percentile)	51.94 (25.62)	51.5 (26.43)	661.5	0.879	0.02
Receptive vocabulary (ROWPVT-4 raw score)	166.56 (10.77)	167.50 (10.77)	633.0	0.870	–0.02
Receptive vocabulary (ROWPVT-4 standard score)	111.25 (16.27)	112.47 (14.15)	622.0	0.774	–0.04
Digit span	7.03 (1.48)	7.25 (1.18)	562.0	0.322	–0.13
Pitch threshold	0.22 (0.10)	0.33 (0.47)	644.5	0.973	–0.01
Global advantage	107.87 (103.21)	102.88 (43.99)	467.0	0.850	0.03
Local-to-global interference	20.41 (26.88)	15.21 (76.89)	652.0	0.016	0.36

AQ: Autism Spectrum Quotient; RSPM: Raven's Standard Progressive Matrices; ROWPVT-4: Receptive One-Word Picture Vocabulary Test, 4th edition.

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 weak central coherence (WCC) theory, autistic individuals show a cognitive bias towards 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 compromise 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 non-verbal 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.

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 Supplementary Material for details). 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 1 for participant demographics). Cognitive measures included non-verbal IQ (Raven's Standard Progressive Matrices; Raven & Court, 1998), receptive vocabulary (Receptive One-Word Picture Vocabulary Test–Fourth Edition (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

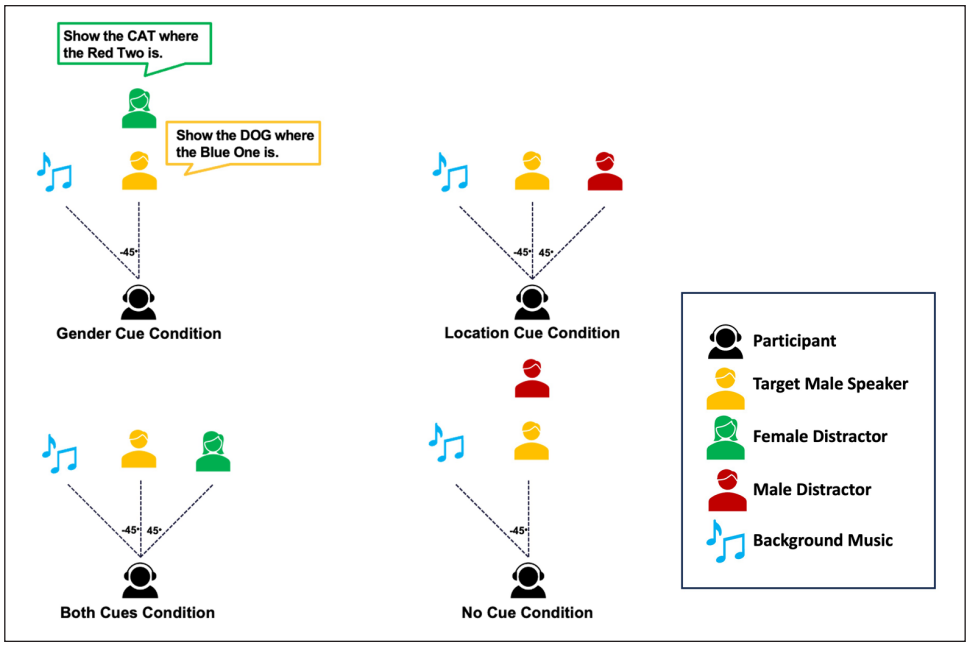


Figure 1. Schematic representation of design. The figure illustrates a single configuration of distractor and music locations; in the actual experiment, these locations vary dynamically across trials, with the music positioned symmetrically opposite the distractor.

small Ss) configurations. Two metrics were derived: global advantage (reaction time (RT) difference for global vs local judgements on congruent trials, indicating a bias towards 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.

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 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 positions; (4) both-cues: mismatched gender, separated positions. 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 towards 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 (see Supplementary Material for details). Each excerpt was taken from the middle section of the original track, avoiding the beginning and end where dynamic or structural changes are more likely to occur. This ensured consistent volume and texture throughout the listening period.

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 Supplementary Material for details). 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.

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 Supplementary Material) 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.

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).

Linear mixed-effects models. 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 LMMs (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 cue): 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.

Generalised additive mixed model. 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 Supplementary Material).

Pearson correlation. A Pearson correlation analysis was conducted to examine the relationship between individual cognitive factors and task performance. 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 non-verbal 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).

Community involvement

There was no community involvement in this study.

Results

Linear mixed-effects models

Accuracy. Figure 2 displays the mean accuracy across cue conditions and groups, while Table 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. In addition, 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 (see Supplementary Material for full results). The only significant comparison emerged in the non-autistic group, where accuracy in the both-cues condition was significantly lower with background music than without it, $\chi^2(1) = 23.64$, $p < 0.001$, OR = 11.78, 95% CI = [4.08, 34.00]. No other comparisons yielded significant effects (all p-values > 0.09).

Reaction times. Figure 3 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. In addition, no significant interactions between these factors were observed (see Table 3).

Generalised additive mixed models

To investigate how SiN 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

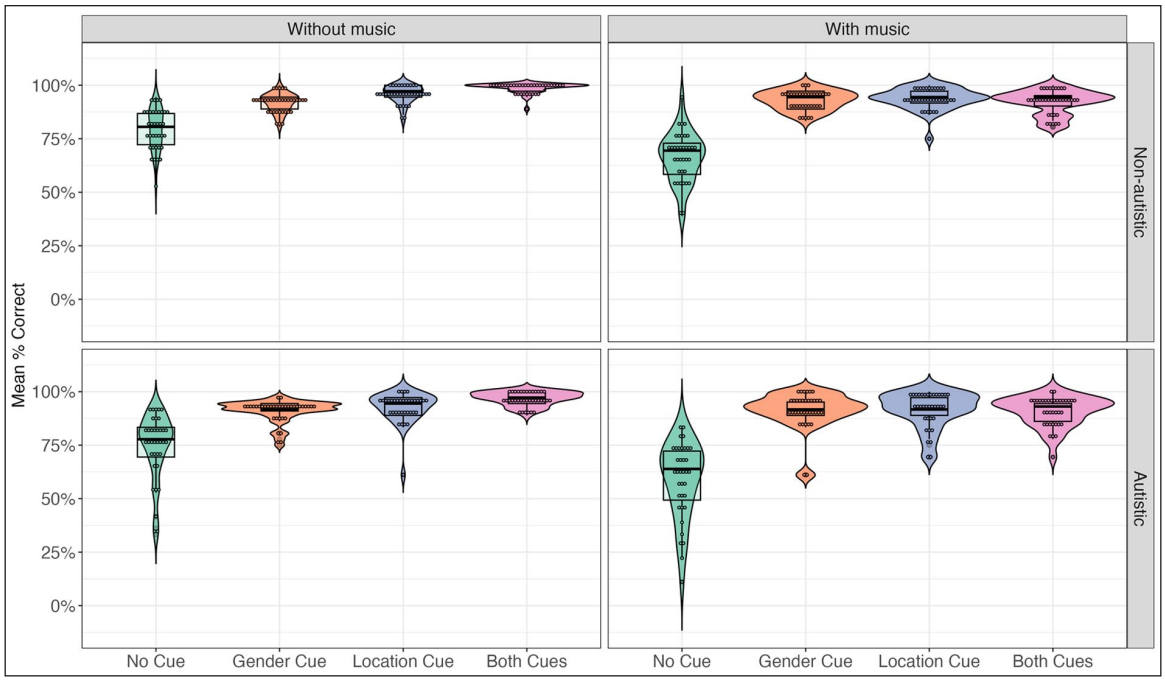


Figure 2. Mean accuracy rate across groups and conditions.

Table 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	**0.027**
Music	0.94	0.18	5.20	2.56	[1.80, 3.64]	25.78	**<0.001**
Cue1	–2.56	0.20	–12.77	0.08	[0.05, 0.11]	135.31	**<0.001**
Cue2	–0.75	0.22	–3.41	0.47	[0.31, 0.73]	11.44	**<0.001**
Group × Cue1	0.05	0.16	0.34	1.05	[0.77, 1.44]	0.11	0.737
Group × Cue2	0.27	0.19	1.42	1.31	[0.90, 1.89]	1.86	0.172
Music × Cue1	–0.01	0.38	–0.04	0.99	[0.47, 2.09]	0.00	1.000
Music × Cue2	–1.48	0.44	–3.36	0.23	[0.10, 0.54]	10.99	**<0.001**
Group × Music	–0.24	0.14	–1.81	0.78	[0.60, 1.02]	3.06	0.080
Group × Music × Cue1	0.64	0.24	2.69	1.90	[1.19, 3.05]	6.87	**0.008**
Group × Music × Cue2	0.79	0.38	2.09	2.20	[1.05, 4.62]	4.06	**0.044**

GLMM: generalised linear mixed-effects model; 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. Significant *p*-values are presented in bold.

each group–condition combination. Participant-specific smooth terms were also included to account for individual variability (see Table 4). Figure 4 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

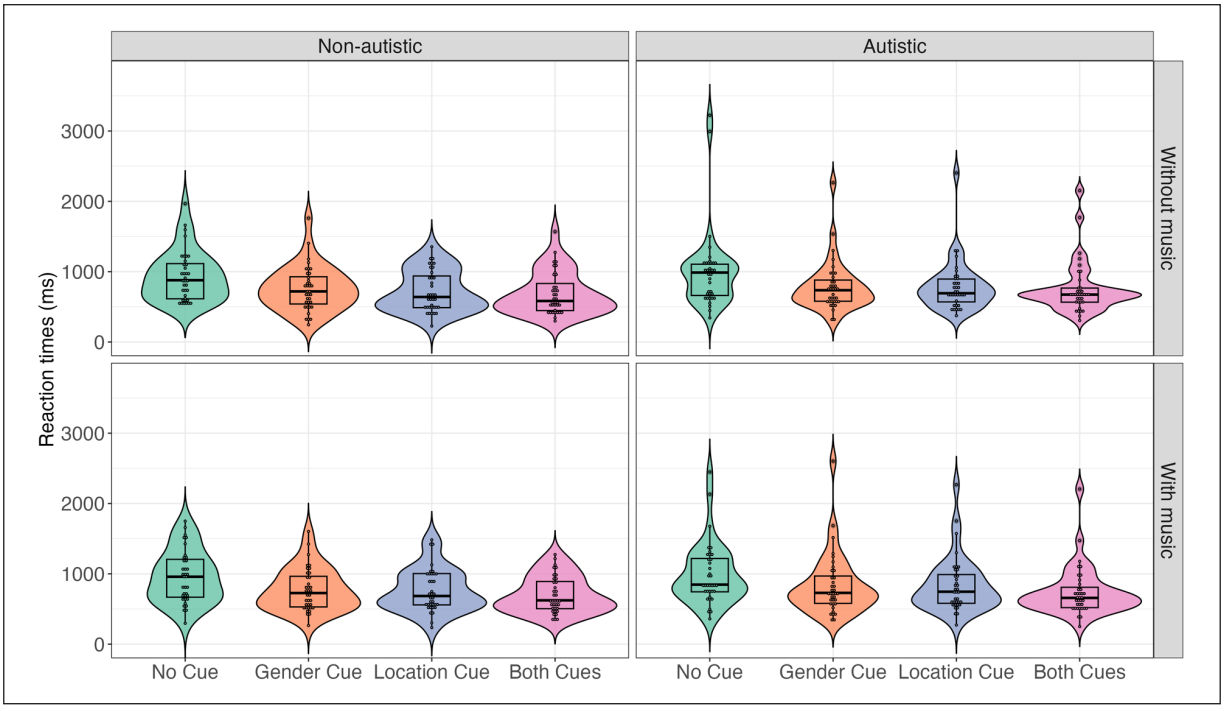


Figure 3. Mean RTs across groups and conditions.
RT: reaction time.

Table 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	0.512
Music	–0.03	0.02	–1.44	0.97	[0.92,1.01]	2.07	0.150
Cue1	0.27	0.03	9.77	1.31	[1.24,1.38]	84.21	**< 0.001**
Cue2	0.10	0.03	3.48	1.10	[1.04,1.16]	11.88	**< 0.001**
Group × Cue1	0.00	0.03	0.17	1.00	[0.95,1.06]	0.03	0.864
Group × Cue2	0.00	0.03	0.06	1.00	[0.95,1.05]	0.00	0.950
Music × Cue1	0.00	0.05	0.01	1.00	[0.90,1.11]	0.00	1.000
Music × Cue2	–0.03	0.05	–0.60	0.97	[0.87,1.08]	0.36	0.547
Group × Music	0.04	0.02	1.92	1.05	[1.00,1.09]	3.52	0.061
Group × Music × Cue1	0.00	0.05	–0.02	1.00	[0.90,1.11]	0.00	0.988
Group × Music × Cue2	-0.06	0.05	–1.15	0.94	[0.85,1.04]	1.31	0.252

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. Significant *p*-values are presented in bold.

over trials. In contrast, no significant trial effects were found in the both-cues condition for either group, reflecting their 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 (see Figure 5). These plots highlight time windows where significant contrasts appeared and are interpreted in light of the accuracy trends shown in Figure 4. In the no-cue condition, significant group differences emerged during the later trials (see Figure 5, 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 4, 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

Table 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	**< 0.001**	2.65	0.15	18.00	**< 0.001**
NAS.Both cues	0.36	0.24	1.48	0.138	0.21	0.21	0.99	0.322
AS.No cue	–2.32	0.15	–15.81	**< 0.001**	–2.00	0.13	–15.93	**< 0.001**
NAS.No cue	–2.08	0.19	–10.87	**< 0.001**	–1.65	0.19	–8.79	**< 0.001**
Smooth terms	edf	Ref.df	χ^2	*p*	edf	Ref.df	χ^2	*p*
AS.Both cues	2.06	2.57	4.48	0.205	1.00	1.00	2.56	0.110
NAS.Both cues	1.00	1.00	0.44	0.506	1.00	1.00	0.50	0.480
AS.No cue	2.09	2.61	9.78	**0.020**	1.00	1.00	7.99	**0.005**
NAS.No cue	1.00	1.00	12.66	**< 0.001**	1.00	1.00	10.85	**< 0.001**
Participants	47.68	646.00	153.93	**< 0.001**	68.49	646.00	256.92	**< 0.001**
	R^2(adjusted) = 0.174				R^2(adjusted) = 0.170			
	Deviance explained = 19.5%				Deviance explained = 18.2%			

GAMMs: generalised additive mixed models; SNR: signal-to-noise ratio; *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). Significant *p*-values are presented in bold.

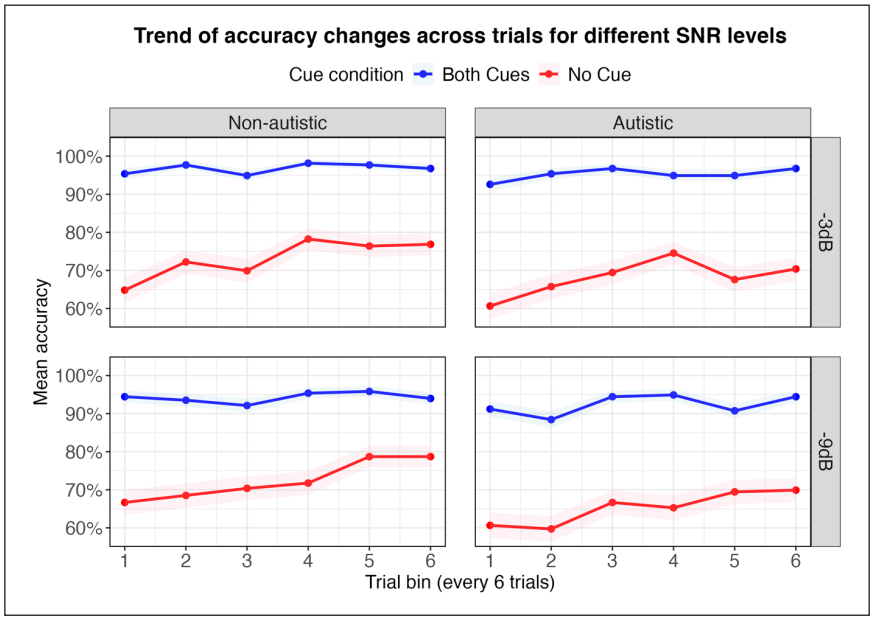


Figure 4. The trend of mean accuracy changes across trial bins (every six trials) for different SNR levels across group and condition with the shaded area indicating the 95% confidence interval.
SNR: signal-to-noise ratio.

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 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 upwards trend in no-cue accuracy seen in Figure 4. 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 4 supports this and shows that while both-cues accuracy stayed high throughout, performance in the no-cue condition fluctuated and showed less overall improvement.

Taken together, 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.

Correlations

Figure 6 presents significant correlations for both groups. In the non-autistic group, lower pitch discrimination

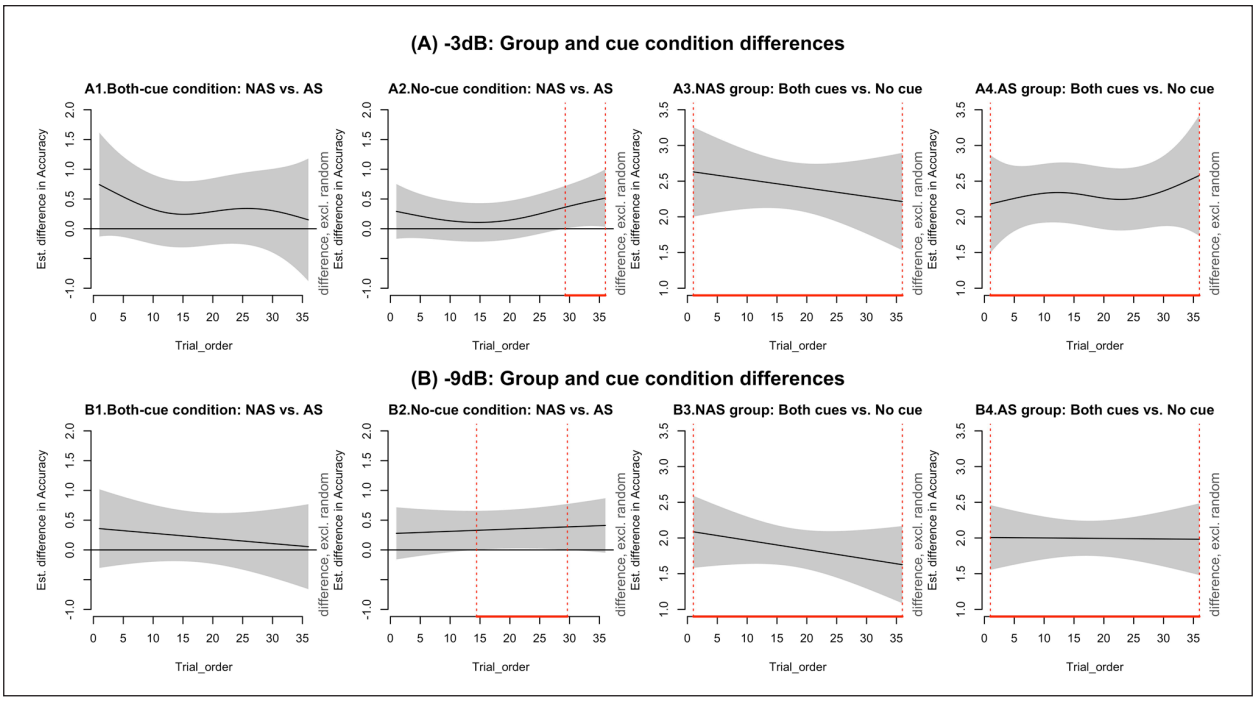


Figure 5. Estimated differences in accuracy over trials.
The black line represents the estimated difference, with the grey shaded area indicating the 95% confidence interval. Red segments highlight trial ranges where the difference is statistically significant ($p < 0.05$).

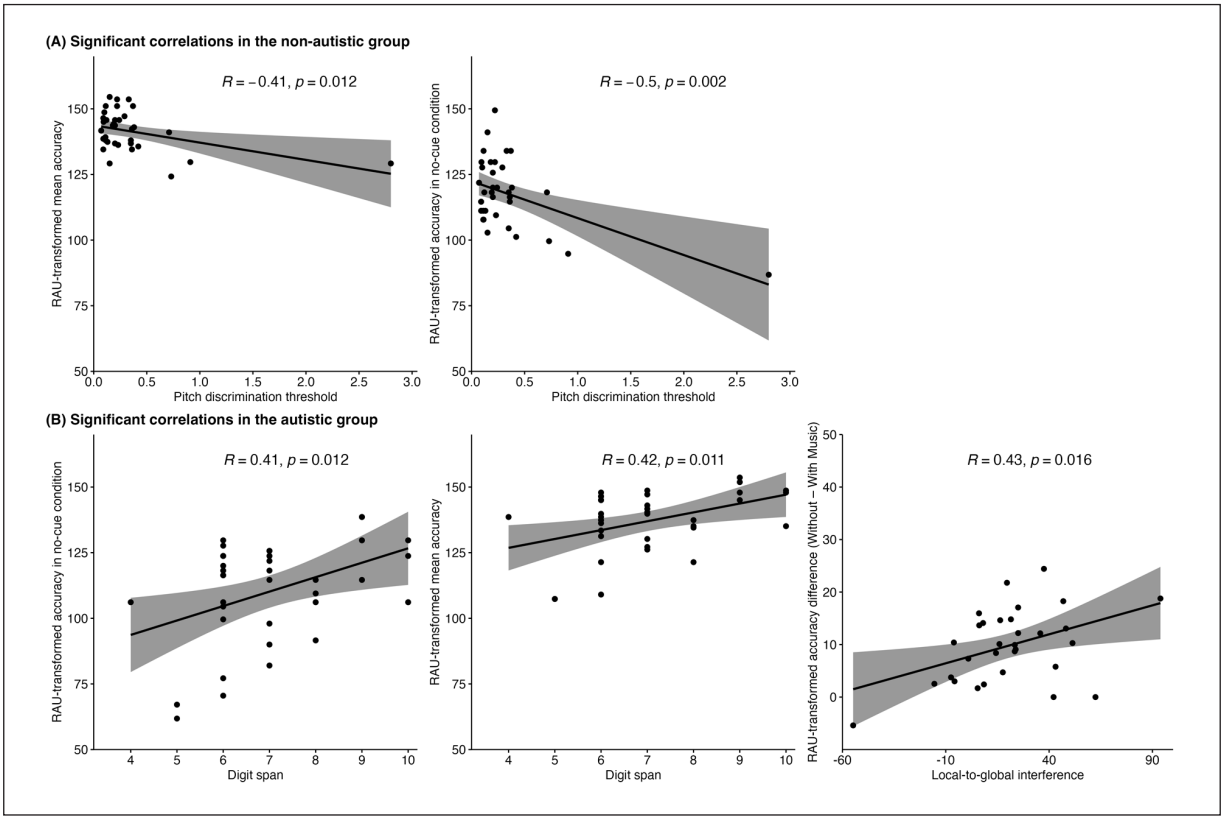


Figure 6. Significant correlations between accuracy and cognitive factors in the non-autistic group (A) and the autistic group (B).
The grey shaded area indicated the 95% confidence interval across the mean. RAU: rationalised arcsine units.

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 Supplementary Material). In the autistic group, higher digit span scores were linked to better overall and no-cue accuracy. In addition, participants who showed stronger local-to-global interference, indicating weaker global processing, also showed larger accuracy declines in the presence of background music.

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.

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 the study by 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 call-sign 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. In addition, 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 SiN processing and demonstrate that autistic listeners can benefit from speaker-related acoustic cues even in tasks that more closely approximate everyday communication demands.

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 atypical 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. In addition, 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 WCC theory (Happé & Frith, 2006), autistic individuals may focus on local acoustic details at the expense of integrating information into a coherent global representation. Difficulties in binding subtle acoustic cues (e.g. pitch) into a unified auditory object may have increased the cognitive demands of the task, thereby prompting greater working memory involvement. The additional load of maintaining task-relevant information in memory may also have contributed to the group differences in performance (Lau et al., 2023).

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 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. Event-related potential (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 (Lepistö et al., 2009). This effect may have been especially pronounced in this study due to the use of a 0 dB SNR, which increased listening demands and likely intensified the impact of local processing bias.

Limitations and directions for future studies

One limitation of this 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 varying 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, 2008; Shi & Law, 2010).

In addition, 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 magnetoencephalography (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.

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. In addition, individual differences in sensitivity to background music highlight the heterogeneity of cognitive processing styles in autism, reinforcing the need for personalised support strategies.

ORCID iDs

Jiayin Li https://orcid.org/0000-0003-4847-6430
Ian Cunnings https://orcid.org/0000-0002-5318-0186
Fang Liu https://orcid.org/0000-0002-7776-0222

Author contributions

Jiayin Li: Data curation; Formal analysis; Funding acquisition; Investigation; Methodology; Project administration; Resources; Software; Writing – original draft; Writing – review & editing.
Maleeha Sujawal: Data curation; Investigation.
Zivile Bernotaite: Data curation; Investigation.
Ian Cunnings: Funding acquisition; Supervision; Writing – review & editing.
Fang Liu: Conceptualisation; Funding acquisition; Project administration; Supervision; Writing – review & editing.

Funding

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by South East Network for Social Sciences (SeNSS) Doctoral Training Partnership (DTP) (ES/P00072X/1); European Research Council (ERC) Starting Grant (ERC-StG-2015, CAASD, 678733).

Declaration of conflicting interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Data availability statement

Anonymised data and analytical code are available at https://osf.io/9gdmf/.

Supplemental material

Supplemental material for this article is available online.

References

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, 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://doi.org/10.1176/appi.books.9780890425596

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

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: Figure 1. *The Journal of Neuroscience*, *24*(42), 9228–9231. https://doi.org/10.1523/JNEUROSCI.3340-04.2004

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

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

Brown, J. A., & Bidelman, G. M. (2022). 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

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

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

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

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

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., Lee, K. C., 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

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

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

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

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

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

Heaton, P., 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

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*, Article 343. https://doi.org/10.3389/fpsyt.2020.00343

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

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 & Biobehavioral Reviews*, *36*(4), 1292–1313. https://doi.org/10.1016/j.neubiorev.2012.02.007

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

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

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

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

Lau, B. K., 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

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

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

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

Martin, N. A., & Brownell, R. (2011). *ROWPVT-4: Receptive one-word picture vocabulary Test-Fourth Edition*. Pearson Assessment.

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*, *54*(6), 1682–1701. https://doi.org/10.1044/1092-4388(2011/09-0261)

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

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

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

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

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

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

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

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

Posit Team. (2022). *RStudio: Integrated development environment for R*. Posit Software, PBC. http://www.posit.co/

Raven, J. C., & Court, J. H. (1998). *Raven's progressive matrices and vocabulary scales*. Oxford Psychologists Press. 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

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

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

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

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

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

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

Shiffrin, R. M., & Schneider, W. (1977). Controlled and automatic human information processing: II. Perceptual learning, automatic attending and a general theory. *Psychological Review*, *84*(2), 127–190. https://doi.org/10.1037/0033-295X.84.2.127

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

Shinn-Cunningham, B. G., 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

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

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

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 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

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

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

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

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 & Hall/CRC. https://www.taylorfrancis.com/books/mono/10.1201/9781315370279/generalized-additive-models-simon-wood

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