Reduced Context Effect on Lexical Tone Normalization in Children with Autism Spectrum Disorder: A Speech-Specific Mechanism

Chen Kuang [1], Fei Chen [1]*, Jinting Yan [1, 2], Gang Peng [1, 3]*

[1] *School of Foreign Languages, Hunan University*

[2] *School of Education, Cangzhou Normal University*

[3] *Research Centre for Language, Cognition, and Neuroscience, Department of Chinese and Bilingual Studies, The Hong Kong Polytechnic University*

Running head: Reduced Context Effect on Lexical Tone Normalization in ASD

First author:

Chen Kuang. Address: School of Foreign Languages, Hunan University, Lushannan Road No. 2, Yuelu District, Changsha City, Hunan Province, China. Tel.: (+86) 15274656369.

Email address: kuangchen@hnu.edu.cn (Chen Kuang).

***Corresponding authors:**

Fei Chen. Address: School of Foreign Languages, Hunan University, Lushannan Road No. 2, Yuelu District, Changsha City, Hunan Province, China. Tel.: (+86) 17788926788.

Email address: chenfeianthony@gmail.com (Fei Chen)

Gang Peng. Address: Research Centre for Language, Cognition, and Neuroscience, Department of Chinese and Bilingual Studies, The Hong Kong Polytechnic University, 11 Yuk Choi Road, Kowloon, Hong Kong SAR, China.

Email address: gpengjack@gmail.com (Gang Peng)

ORCID

Chen Kuang https://orcid.org/0000-0002-3025-9932

Fei Chen https://orcid.org/0000-0002-4500-2590

Gang Peng https://orcid.org/0000-0002-1465-1301

Funding:

This work was supported by the Humanities and Social Sciences Youth Foundation, Ministry of Education of the People's Republic of China [grant number 22YJC740008].

Declarations:

Conflict of interest: The authors do not have any conflict of interest to declare.

Reduced Context Effect on Lexical Tone Normalization in Children with Autism Spectrum Disorder: A Speech-Specific Mechanism

Abstract

Purpose: Existing literature has demonstrated that individuals with autism spectrum disorder (ASD) exhibit atypical use of contextual information in their surroundings. However, there is limited understanding regarding their integration of contextual cues in speech processing. This study aims to explore how Mandarin-speaking children with and without ASD identify lexical tones in speech and nonspeech contexts, and to determine whether the size of context effect would be modulated by children's cognitive abilities.

Methods: Twenty-five children with ASD and 25 typically developing (TD) children were asked to identify Mandarin lexical tones preceded by three types of contexts (speech, nonspeech, and nonspeech-flattened contexts). We also tested child participants' verbal intelligence, nonverbal intelligence, and working memory capacity.

Results: Results revealed that the context effect was only observed in the speech contexts, where Mandarin-speaking children with ASD exhibited a reduced context effect compared to TD children. Moreover, TD children with higher verbal intelligence demonstrated a diminished context effect. However, nonverbal intelligence and working memory capacity were not significantly associated with the size of context effect in either group.

Conclusion: These findings revealed a subtle yet important difference between ASD and TD children's utilization of speech contexts in lexical tone identification, and validated a speech-specific mechanism underpinning children's lexical tone normalization.

Keywords Autism spectrum disorder · Mandarin tone · Context effect · Speech-specific mechanism

Introduction

Autism spectrum disorder (ASD) is a neural-developmental disorder. People diagnosed with ASD are clinically characterized by impaired social interaction, narrow interests, and repetitive behaviors (American Psychiatric Association, 2013, 2022). Beyond these core symptoms, autistic individuals' detail-oriented perceptual style has been documented since the first delineation of the autistic syndrome (Kanner, 1943). Their local processing bias has been extensively observed across both visual and auditory modalities. In particular, individuals with ASD focused more on the composing features of visual stimuli (Booth & Happé, 2018; Gauthier et al., 2009; Happé, 1999), exhibited intact or superior local processing abilities of pitch in pure tones (Bonnel et al., 2010; Eigsti & Fein, 2013; Heaton et al., 2008; Mottron et al., 2006), and showed overly focused auditory processing of speech (Heaton et al., 2008; Järvinen-Pasley, Pasley, et al., 2008; Järvinen-Pasley, Wallace, et al., 2008). Yet, the specific mechanisms that underlie the atypical processing style in ASD remain poorly understood.

Atypical Perceptual Processing in ASD

Two long-standing cognitive theories have been proposed to account for the atypical perceptual processing in individuals with ASD. The Weak Central Coherence (WCC) theory posits that individuals with ASD tend to prioritize the processing of local information over global integration, and exhibit a diminished or impaired ability to extract information from the bigger picture or global context (Frith, 1989; Happé & Frith, 2006). Individuals with ASD are often absorbed in details and find it challenging to integrate information within the surrounding context to the same extent as their typically developing (TD) counterparts (D'Souza et al., 2016; O'Connor, 2012). For example, autistic people exhibited difficulty in integrating global visual information (Booth & Happé, 2018; Happé, 1999) or using contextual information to derive the meaning of words and narratives (Happé, 1999; Nuske & Bavin, 2011). Contrastively, the Enhanced Perceptual Functioning (EPF) theory advocates that individuals with

ASD can exhibit heightened perceptual abilities in processing elementary physical properties without global perception deficits (Mottron et al., 2006; Mottron & Burack, 2001). Consistent with the EPF theory, prior studies have indicated that participants with ASD demonstrated a strong observation bias towards local shapes while maintaining intact spatial integration (Hadad & Ziv, 2015), or they may show enhanced local processing of musical pitch alongside intact holistic processing of musical chords (Heaton, 2003).

The above two models concur that autistic individuals exhibit significant differences in sensory input processing compared to TD individuals. However, they diverge in their explanations concerning the atypical nature of perceptual processing in ASD (Pellicano & Burr, 2012). The WCC theory posits that deficiencies in top-down modulation can result in hypersensitivity to sensory stimuli in ASD, whereas the EPF theory proposes that autistic perception is characterized by enhancements in bottom-up, feed-forward perceptual operations.

Most recently, researchers (Johnson, 2017; Johnson et al., 2015, 2021; Klin et al., 2020) have proposed a neuroconstructivist model that explains perceptual processing in autism from a developmental perspective. They argue that the processing style characterized by an overly focal orientation in autism can be understood as an adaptive response of an atypical brain to the pace and quantity of incoming information. A poor signal-to-noise ratio within each sensory modality may impede effective cross-modal integration. This suggests that the more featural and detail-oriented style of processing in individuals with autism may reflect inherent limits in parallel processing capabilities, leading to a strategic advantage in restricting sensory input to a single channel or a specific area of external space.

According to the neuroconstructivist model (Johnson et al., 2021), a neurocognitive modifier system can influence the relationship between atypical early-stage processing (e.g., early sensory disruptions) and phenotypic outcomes such that robust neurocognitive modifier system can push trajectories towards typical outcomes. A range of evidence supported the important role of the neurocognitive modifier system, especially in early ASD. While overt social behavioral differences are limited in the first year for infants who later develop ASD (Ozonoff et al., 2010), subtler differences have been extensively reported, including a decline in eye gaze between 2 and 6 months (W. Jones & Klin, 2013), reduced face looking from 6 to 24 months (Ozonoff et al., 2010), diminished temporal lobe sensitivity to social stimuli at 5 months (Lloyd-Fox et al., 2018), and altered neural processing of faces from 6 to 10 months (Elsabbagh et al., 2012; E. J. H. Jones et al., 2016).

Despite the abundance of research on perceptual processing abilities in ASD (DiStefano et al., 2019; Henderson et al., 2011; Järvinen-Pasley, Peppé, et al., 2008; Lau et al., 2021; Zhao et al., 2024), few investigations have actually explored the effects of preceding acoustic or semantic contexts on the perception of speech sounds among individuals with ASD. Notably, Gabay et al. (2024) investigated the impact of preceding contextual information, including acoustic (spectral and temporal) and semantic (lexical) cues, on phoneme categorization in Hebrew-speaking adults with or without autism. They discovered that adults with ASD could utilize all categories of contextual information to disambiguate speech sounds to the same extent as their IQ-matched TD peers, indicating that adults with ASD were able to use prior contextual information in perceiving segments (consonants and vowels). Yet, it is essential to know whether similar patterns would emerge in the categorical perception of suprasegmental components (e.g., lexical tones).

Context Effects in Lexical Tone Normalization

The meaning of speech sounds in tonal languages bears heavily on the accurate recognition of lexical tones (W. S.-Y. Wang, 1972). In real-life situations, listeners need to recognize the same tone produced by different talkers or the same talker in various conditions. The process of mapping different tonal variants onto the same tonal category is referred to as tone normalization (Francis et al., 2006). Successful tone normalization depends on the utilization of both intrinsic and extrinsic cues (Chen & Peng, 2016; Francis et al., 2006; C. Zhang et al., 2012). The intrinsic cues are the acoustic correlates of the target words, such as fundamental frequency (F0), duration, intensity, and voice quality. Extrinsic cues (i.e., contextual cues) refer specifically to the acoustic features of the surrounding contexts, including F0 range and average F0 height. Listeners' use of extrinsic contexts to identify lexical tones is known as extrinsic tone normalization.

The integration of contextual F0 cues and target lexical tones constitutes a form of global processing. Previous results unequivocally demonstrated a significant impact of a talker's speaking F0 in contexts on adults' normalization of Cantonese level tones and widely reported a contrastive effect elicited by F0 shifts (Francis et al., 2006; Wong & Diehl, 2003; C. Zhang et al., 2012). Specifically, a Cantonese mid-level tone is more likely to be identified as a high-level tone when preceded by a context with a low F0, and as a low-level tone after a context with a high F0. Unlike the Cantonese level tones that are differentiated by F0 heights, Mandarin tones are primarily defined by F0 contours (Peng et al., 2012; Zhu et al., 2023). As such, the contrastive context effect due to changes in F0 heights was merely found in adults' perception of ambiguous Mandarin tones close to the perceptual boundary, but was scarcely noted in adults' perception of prototypical Mandarin tones in studies using the paradigm of categorical perception (Chen & Peng, 2016; Huang & Holt, 2009, 2011). A series of studies explicitly examined the context effect by using a tone continuum ranging from Mandarin high-level tone (Tone 1, T1) to Mandarin mid-rising tone (Tone 2, T2) (Chen & Peng, 2016; Huang & Holt, 2009, 2011). Their findings revealed that when the ambiguous tones (i.e., stimuli positioned in the middle of the continuum) were preceded by a context with a lower mean F0, they were perceptually categorized as the Mandarin T1 with higher pitch onset. Conversely, when the precursor context had a higher mean F0, these same ambiguous tones were perceived as the Mandarin T2 with lower pitch onset. Yet, no context effect was observed in the ending stimuli (typical T1 and T2). In a similar vein, prior researchers also probed into the relative contribution of F0 heights to children's perception of Mandarin T1 and T2. Chen et al. (2023)'s recent work reported prominent contrastive context effects in Mandarin-speaking TD children aged from 6 to 8, suggesting that 6- to 8-year-olds can be proficient in utilizing the acoustic cues embedded in speech contexts for tone perception. Nevertheless, it remains unclear whether Mandarin-speaking children with ASD might exhibit TD-like behavior when processing contextual cues in lexical tone identification.

Effect of Context Type on Lexical Tone Normalization

There is an ongoing debate about whether lexical tone normalization in neurotypical populations recruits a general perceptual mechanism or a speech-specific mechanism. The general perceptual account holds that speech and nonspeech sounds are processed in a comparable manner within the human cognitive system, which involves common auditory and cognitive mechanisms to encode acoustic information (Diehl et al., 2004; Holt, 2006). Some studies (Huang & Holt, 2009; H. Zhang et al., 2022) found that both speech and nonspeech contexts can elicit significant contrastive context effects, thereby supporting the engagement of a general perceptual mechanism. In contrast, the speech-specific account believes that speech signals, different from nonspeech ones, are processed via a specialized brain network (Liberman et al., 1967; Liberman & Mattingly, 1985). A large number of studies suggested that only speech contexts can significantly facilitate the normalization of lexical tones, whereas the effect of other types of contexts was almost negligible (Chen & Peng, 2016; Francis et al., 2006; C. Zhang et al., 2012; K. Zhang et al., 2017). Importantly, a recent work conducted by Chen et al. (2023) evidenced a bigger role of speech contexts, as opposed to nonspeech contexts, in Mandarin-speaking TD children's normalization of Mandarin tones. A question that follows is: Will the larger context effect generated by speech carriers in TD children's lexical tone normalization also be evident in their ASD counterparts with potential speech-specific pitch processing atypicality?

Abundant behavioral and neuroimaging evidence suggested that people with ASD manifested intact or even enhanced pitch processing abilities in nonspeech stimuli (Bonnel et al., 2010; Foxton et al., 2003; Heaton, 2005; C. R. G. Jones et al., 2009; O'Riordan & Passetti, 2006) but showed impaired performance when processing pitch information in lexical tones and linguistic prosody (Chen et al., 2016, 2021; Jiang et al., 2015; X. Wang et al., 2017; Yu et al., 2015). Drawing from these observations, researchers proposed that individuals with ASD may struggle to develop or utilize specialized networks for vocal processing and phonetic learning in speech sounds (Kujala et al., 2013; Lindell & Hudry, 2013; O'Connor, 2012). To the best of our knowledge, no previous studies have attempted to compare the context effects in speech and nonspeech contexts among autistic children, leaving uncertainty about the influence of speech-specific pitch processing difficulties on their utilization of both speech and nonspeech contexts.

Of note, previous studies have suggested that local versus global auditory processing is often intertwined with stimulus complexity (Haesen et al., 2011; Samson et al., 2006), making it essential to consider stimulus complexity in our study. The neural complexity hypothesis predicts that individuals with ASD will exhibit enhanced perception of simple low-level stimuli and local features, while their ability to process more complex and global information may be spared or impaired (Samson et al., 2006). Consistent with the complexity hypothesis, which posits that individuals with ASD struggle to process spectrally and temporally complex auditory information (Samson et al., 2006), auditory temporal processing impairments have been well-documented in the ASD literature (Alcántara et al., 2004; Foss-Feig et al., 2017; Gabay et al., 2024; Groen et al., 2009; Meilleur et al., 2020). Given this, we created nonspeech-flattened contexts that are simpler than the nonspeech stimuli with temporal dynamics. Consequently, there exists a gradient of stimulus difficulty, with speech contexts representing the most complex, followed by nonspeech contexts, and finally nonspeech-flattened contexts.

In addition, although researchers have reached a consensus that individuals use F0 cues in the contexts to normalize lexical tones, little is known about which type of F0 information plays the critical role. Francis et al. (2006) introduced a mean F0 condition where the original F0 of a speech sentence was replaced with a flattened mean F0, effectively eliminating cues of a speaker's full F0 range. Compared with the original F0 condition, the mean F0 condition was found to be more helpful in adults' normalization of Cantonese level tones. However, C. Zhang et al. (2012) discovered that the flattened mean F0 condition, when compared to the original F0 contour condition, was less effective in facilitating adults' Cantonese tone normalization in speech contexts, which highlighted the dominant role of the F0 range over the mean F0. It is worth mentioning that earlier research mainly concentrated on adults' utilization of F0 cues in lexical tone normalization, while the extent to which children with and without ASD use these cues for the same purpose remains unknown. Thus, the current study intends to address this gap by comparing the perceptual performance in original nonspeech and nonspeech-flattened contexts among ASD and TD children.

Roles of Cognitive Abilities in Pitch Processing

It is well-documented that substantial individual differences exist in cognitive functions within the ASD population (Johnson et al, 2015; Klin et al., 2020; Mottron & Bzdok, 2020). For instance, children with ASD may exhibit a highly variable profile of verbal/nonverbal abilities (Hudry et al., 2010; Mecca et al., 2014) and working memory (WM) capacity (Ozonoff & Strayer, 2001; Williams et al., 2005) when compared with TD children. Numerous published studies on ASD and TD individuals (e.g., Chowdhury et al., 2017; Mayer et al., 2016; McHaney et al., 2021; Ong et al., 2024; Sota et al., 2018; L. Wang et al., 2023) have described the link between their cognitive abilities and their performance in both speech and nonspeech pitch processing tasks.

Individuals' verbal ability or verbal intelligence (VIQ) reflects their receptive and/or expressive spoken language skills (Volkmar, 2021). In TD, prior studies showed that better performance in pitch processing was associated with improved verbal abilities (Chung & Bidelman, 2021; Mayer et al., 2016; L. Wang et al., 2023). In ASD, verbal skills could effectively predict their ability to process pitch in lexical tones or intonations (Chen et al., 2016; Rong et al., 2023; L. Wang et al., 2023). In terms of nonverbal intelligence (NVIQ), it refers to the ability to analyze, reason, and solve problems using visual and spatial information, without relying on verbal production and comprehension (Volkmar, 2021). Previous research indicated a positive correlation between NVIQ and auditory pitch processing abilities in individuals with and without ASD (Chowdhury et al., 2017; Stanutz et al., 2014). Furthermore, as a cognitive system responsible for the temporary storage and manipulation of information for complex cognitive tasks, WM is widely acknowledged for its critical role in facilitating language acquisition and linguistic processing (Baddeley, 1992, 2000; Gathercole & Baddeley, 2014). In TD individuals, higher WM facilitated pitch perception in lexical tones (McHaney et al., 2021) and supported absolute pitch learning (Van Hedger et al., 2015). Similarly, WM in children or adolescents with ASD was proven to be a reliable predictor of their pitch processing abilities in speech and nonspeech carriers (Chen & Peng, 2021; Ong et al., 2024; Sota et al., 2018; L. Wang et al., 2023).

Speech normalization, including lexical tone normalization, represents a regulated process in human information processing that demands sufficient cognitive resources (Chen et al., 2023). Enhanced cognitive abilities might help children decipher speech and nonspeech pitch in contexts, bolster their retention of pitch information, and allow them to efficiently compare contexts and targets, ultimately leading to a more salient contrastive context effect (Chen et al., 2023). In this regard, it is worthwhile to investigate whether and how cognitive abilities, such as VIQ, NVIQ, and WM, influence lexical tone normalization by children with and without ASD, a topic that has been scarcely researched in this field.

The Current Study

In order to gain more insight into auditory processing mechanisms in ASD, this study seeks to explore how Mandarin-speaking children with and without ASD perceive lexical tones in speech and nonspeech contexts. The current study firstly attempts to examine whether children with ASD can utilize extrinsic contexts to normalize lexical tones in a TD-like fashion. Based on the WCC model, we hypothesize that children with ASD may exhibit a reduced context effect compared to their TD peers. In contrast, an alternative perspective grounded on the EPF model suggests that autistic children may show comparable integration of contexts with targets as TD children. Furthermore, drawing on the neuroconstructivist model, we anticipate that children with ASD may face challenges in utilizing preceding acoustic contexts to perceive lexical tones; considering the role of neurocognitive modifiers, the difference between the ASD and TD groups might be subtle.

Secondly, the present study intends to investigate how context type modulates ASD and TD children's Mandarin tone normalization. Based on previous findings in Mandarin-speaking TD children (Chen et al., 2023), we expect a more pronounced context effect in TD children within speech contexts compared to nonspeech and nonspeech-flattened contexts, indicating the involvement of a speech-specific mechanism. For children with ASD, it is possible that they might exhibit a speech-specific mechanism like TD children. However, given their potential perceptual bias towards nonspeech stimuli (Bonnel et al., 2010; Foxton et al., 2003; Heaton, 2005; C. R. G. Jones et al., 2009; O'Riordan & Passetti, 2006) and the higher stimulus complexity of speech contexts, it is also possible that the general perceptual mechanism might play a dominant role in autistic children's perception.

The third purpose of this study is to evaluate whether the size of the context effect would be indexed by VIQ, NVIQ, as well as WM in children with and without ASD. According to previous research highlighting these factors as robust predictors of ASD and TD individuals' pitch processing (e.g., Chowdhury et al., 2017; Mayer et al., 2016; McHaney et al., 2021; Ong et al., 2024; Sota et al., 2018; L. Wang et al., 2023), it is postulated that both ASD and TD children's VIQ, NVIQ, WM might effectively predict the size of the context effect. Higher VIQ, NVIQ, and WM might be associated with a larger size of the context effect.

Method

Power Analysis

To estimate the required sample size, we conducted a simulation-based power analysis using the SIMR package (Green & Macleod, 2016) in R (R Core Team, 2023). The expected effect size and the dataset for the simulation were derived from a pilot study conducted in our laboratory that employed a similar experimental design. Since the current study aimed to determine whether the context effect could be observed in both the ASD and TD groups, the power analysis was specifically designed to assess the ability to detect the main effect of contextual F0 height in the two groups. Based on the statistical analyses from 1,000 simulations, we determined that the experimental design of this study would have sufficient power (90.2%) to detect the main effect of contextual F0 height with a sample size of 20 subjects at an alpha level of 0.05. Therefore, we should include a minimum of 20 subjects in each group for this study.

Participants

To ensure a sufficient statistical power, the initial sample for this study consisted of 54 Mandarin-speaking children, with 29 children with ASD and 25 TD controls. To ensure that all child participants could accurately perceive synthetic sounds, a minimum accuracy of 80% in identifying two ending stimuli (i.e., prototypical T1 and T2) was required in the formal experimental task (see Chen et al., 2023). Subsequently, the final analysis included 25 TD children (three girls, M_{age} = 8.10 years, SD = 1.78 years) and 25 children with ASD (three girls,

M_{age} = 8.49 years, *SD* = 2.24 years). Based on the DSM-5 (American Psychiatric Association, 2013) and the ADOS-2 (Lord et al., 2012), the diagnosis of ASD was conducted by pediatricians and child psychiatrists at local hospitals prior to enrollment. All children were reported to have normal or corrected-to-normal hearing and vision by their parents. Written informed consent from the parents and verbal assent from the child participants were obtained before participation. This research protocol was granted by the Research Ethics Committee of <institution redacted for peer review>.

Children's VIQ and NVIQ were assessed using the Wechsler Intelligence Scale for Children, Fourth Edition (WISC-IV) (Wechsler, 2003). A forward digital span task was used to measure the children's short-term phonological WM (Chen & Peng, 2021; Millman & Mattys, 2017; Yan et al., 2021). Table 1 provides demographic details of the child participants. The two testing groups were matched in terms of chronological age, NVIQ, and WM (all *p*s > .05), but the VIQ in children with ASD was lower than that in TD children (t = -2.16, p = .036, Cohen's d = -0.61).

Table 1 Demographic characteristics of participants in ASD and TD groups

	ASD group (*n* = 25)		TD group (*n* = 25)				
	M (*SD*)	Range	*M* (*SD*)	Range	*t*	*p*	Cohen's *d* [95% CI]
Age (y)	8.49 (2.24)	6;0–12;4	8.10 (1.78)	6;5–12;1	0.67	0.506	0.19 [-0.38, 0.76]
VIQ	99.00 (18.98)	60–136	109.00 (13.22)	82–133	-2.16	**0.036***	-0.61 [-1.18, -0.04]
NVIQ	106.32 (15.11)	73–130	105.96 (12.70)	82–129	0.09	0.928	0.03 [-0.54, 0.59]
WM	12.92 (2.61)	7–18	12.12 (2.74)	7–18	1.06	0.296	0.30 [-0.27, 0.87]

Note. Bold values indicate *p* values smaller than .05. VIQ = verbal IQ, NVIQ = non-verbal IQ, WM = working memory. *p < .05

Materials

As illustrated in Fig. 1, a trial in the lexical tone identification task was composed of two parts: context and target. There were three types of contexts, speech contexts (see Fig. 1a), nonspeech contexts (see Fig. 1b), and nonspeech-flattened contexts (see Fig. 1c), and each contains a high-F0 version (the blue lines in Figs. 1a, 1b, and 1c) and a low-F0 version (the red lines in Figs. 1a, 1b, and 1c).

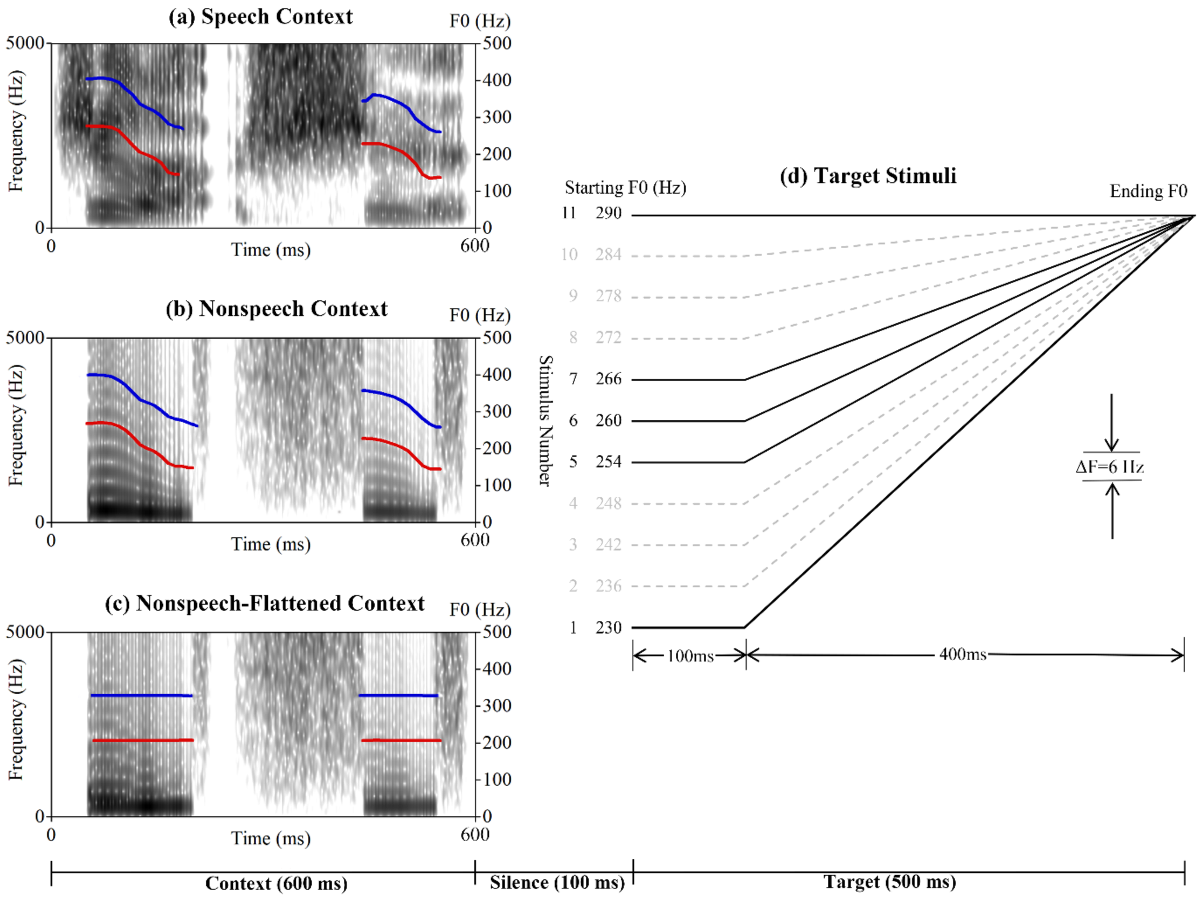

Fig. 1 The speech contexts (a), nonspeech contexts (b), nonspeech-flattened contexts (c), and target stimuli (d) in the context-dependent tone identification task. The blue lines are F0 contours of high contexts, and the red lines are F0 contours of low contexts

The preceding speech contexts consisted of a short Mandarin phrase /zhe4 shi4/ (transliterated in *Pinyin*; "This is" 这是). The reasons for choosing the phrase /zhe4 shi4/ as context were three-fold. First, it is a sentence that remains neutral in both meaning and emotion. Second, both words are familiar to 4-year-olds based on the Tong corpus (Deng & Yip, 2018) within the CHILDES database (MacWhinney, 2000), making them suitable materials for the present study involving children aged 6 to 12. Third, both syllables in this context carry Mandarin T4 (a falling tone), which is distinct from the target level and rising tones (T1 and T2), mitigating a potential priming effect. Speech contexts were generated based on recordings of /zhe4 shi4/ by a female Mandarin speaker (see Fig. 1a). First, the recording was normalized to a duration of 600 ms and an intensity of 70 dB. Subsequently, the entire F0 trajectory (mean F0 = 260 Hz) was raised and lowered by four semitones, respectively, simulating a high-F0 speaker context (mean F0 = 328 Hz, four semitones higher) and a low-F0 speaker context (mean F0 = 206 Hz, four semitones lower).

Nonspeech contexts were created using a triangle wave and white noise. The voiced portion in nonspeech contexts was simulated with the triangle wave, while the unvoiced part was synthesized using the white noise. These nonspeech contexts were designed to replicate the F0 trajectories of the speech contexts, generating both high and low nonspeech contexts (see Fig. 1b). To ensure that the nonspeech contexts had a similar perceived loudness level as the speech contexts, their intensity was set at 80 dB, which is 10 dB higher than the intensity of the speech context (Chen & Peng, 2016).

Finally, the nonspeech-flattened contexts were generated based on nonspeech contexts (see Fig. 1c). The original F0 contour of the nonspeech contexts was replaced with the flattened F0 trajectory while preserving the mean F0 values at 328 Hz in the high context and 206 Hz in the low context.

The contexts and target stimuli were created using recordings from the same female speaker. Each trial included a 100-ms silence interval between the context and the target (see Fig. 1). The target syllable selected for this study was the monosyllabic word /yi/. Both T1 and T2 syllables containing the sound /yi/ were monosyllabic words in Mandarin (/yi1/ denoting 'clothes' and /yi2/ signifying 'aunt'). The original recording of /yi1/ was first normalized to a duration of 500 ms and an intensity of 70 dB. To create the target tone continuum, we manipulated the F0 contours of the original /yi1/ recording. Using Praat's built-in overlap-add synthesis method (Boersma & Weenink, 2021), we replaced the original F0 trajectory with 11 different F0 trajectories, as shown in Fig. 1d.

The 11 F0 trajectories were synthesized in Praat using interpolation. We defined three key points at 0 ms (the starting position), 100 ms, and 500 ms (the ending position). By manipulating the positions of these three points, an 11-step /yi1/-/yi2/ continuum with a step size of 6 Hz was obtained. As depicted in Fig. 1c, the onset F0 for typical /yi1/ (Stimulus 11) was set at 290 Hz, while the onset F0 for typical /yi2/ (Stimulus 1) was set at 230 Hz. The offset F0 for both /yi1/ and /yi2/ remained at 290 Hz. The onset F0 value of /yi2/ (230 Hz) was determined based on the natural recording of /yi2/ from the same female speaker. Considering that typical Mandarin T2 does not rise immediately from the beginning but instead remains relatively stable during the first 20% of its F0 trajectory, we maintained a flat F0 for the first 100 ms. This design aligns with previous studies that have taken into account the acoustic realizations of typical Mandarin T2 (e.g., Chen et al., 2017, 2023; Peng et al., 2010).

In the lexical tone identification task, only five stimuli were selected from the 11-step tone continuum, including two typical tonal stimuli at either end of the continuum (Stimuli 1 and 11) and three ambiguous stimuli (Stimuli 5, 6, and 7). Two typical stimuli were included to ensure that child participants understood the task instructions and took the task seriously. Our data analysis excluded the data from children who failed to achieve a minimum accuracy score of 80% in identifying two specific stimuli (Stimuli 1 and 11) during the formal task (see Chen et al., 2023). Besides, the rationale behind selecting the three ambiguous stimuli (Stimuli 5, 6, and 7) was two-fold. First, recent studies showed that context-dependent shifts of Mandarin T1 and T2 identification were found in ambiguous stimuli (Chen & Peng, 2016; K. Zhang et al., 2017); therefore, three ambiguous targets were incorporated to examine the context effect on lexical tone identification. Second, Chen et al. (2017), who adopted the same /yi1/-/yi2/ continuum, found that the middle stimulus (Stimulus 6) was the most ambiguous for both Mandarin-speaking children and adults, with its identification rate approaching chance level (50% in a two-alternative forced choice task).

Procedure

Before the formal experiment, a training session was conducted to make sure that children had a comprehensive understanding of the experimental procedure. During the training session, the experimenter began by instructing the child participants to point at a picture depicting a car driving on a level road upon hearing the typical /yi1/ (Stimulus 11), and they were taught to point at a picture showing a car driving up a hill upon hearing the typical /yi2/ (Stimulus 1). Once they understood which picture to select after hearing /yi1/ or /yi2/, the experimenter presented two unambiguous targets preceded by high and low contexts to the participants. This allowed the children to become familiar with the contextual stimuli they would encounter in the formal experiment. Only participants who successfully matched the unambiguous target stimulus with the correct picture were allowed to proceed to the formal experiment.

In the formal task, three blocks were administered in a counterbalanced order: one block with speech contexts, one with nonspeech contexts, and one with nonspeech-flattened contexts. The stimulus including typical targets (Stimulus 1 or 11) was repeated five times, while the stimulus containing ambiguous targets (Stimulus 5, 6, or 7) was repeated four times, resulting in a total of 132 trials (3 context types [speech, nonspeech, and nonspeech-flattened] × 2 context heights [high vs. low] × 2 unambiguous target stimuli [Stimuli 1 and 11] × 5 repetitions + 3 context types [speech, nonspeech, and nonspeech-flattened] × 2 context heights [high vs. low] × 3 ambiguous target stimuli [Stimuli 5, 6, and 7] × 4 repetitions). The tone identification task was conducted on a laptop using

E-prime 2.0 software (Psychology Software Tools, 2012) in a sound-attenuated room.

Children were instructed to indicate their identification of the target syllable by pointing at the corresponding picture. No feedback was given to the participants during the formal task. The experimenter recorded the children's responses by pressing "1" on the keyboard when children pointed at the picture of a car driving on a level road or "2" when children pointed at the picture of a car driving up a hill. The entire tone identification task, including both training and the formal task, took approximately one hour to complete. If children exhibited the normalization process, their perception of the same ambiguous target would change in a contrastive manner according to the F0 heights of the preceding contexts. Specifically, they would give more T2 responses (a tone with a relatively low F0) in high-F0 contexts and more T1 responses (a tone with a relatively high F0) in low-F0 contexts.

Data Analysis

In analyzing participants' responses in three types of contexts, the statistical analyses were conducted using the generalized linear mixed-effects models (GLMMs) with the *lme4* package (Bates et al., 2015) in R (version 4.3.1) (R Core Team, 2023). "Response" was the dependent variable. "Context Height" (high and low contexts), "Group" (ASD and TD groups), and their interaction were fixed factors. We anticipated that the relations among test variables would differ for each participant or for each stimulus; therefore, "Participant" and "Trial" were included as random factors. In order to control for potential confounding factors, the centered values of VIQ, NVIQ, and WM were incorporated as covariates. Post-hoc pairwise comparisons were conducted using the *emmeans* package (Lenth, 2018) with Tukey adjustment. Furthermore, hierarchical multiple regressions were conducted to delve into the roles of VIQ, NVIQ, and WM in predicting the size of context effect in lexical tone normalization. Notably, only the ambiguous targets (Stimuli 5, 6, and 7) were included in our data analyses because the identification of typical Mandarin tones was largely unaffected by contextual cues (Chen et al., 2023; Chen & Peng, 2016).

Results

Context-Dependent Identification of Mandarin Lexical Tones

In order to examine whether ASD and TD children exhibited a significant lexical tone normalization process, we compared the participants' responses in high-F0 vs. low-F0 contexts within each condition. Fig. 2 illustrates the perception results of both ASD and TD groups in the speech, nonspeech, and nonspeech-flattened contexts.

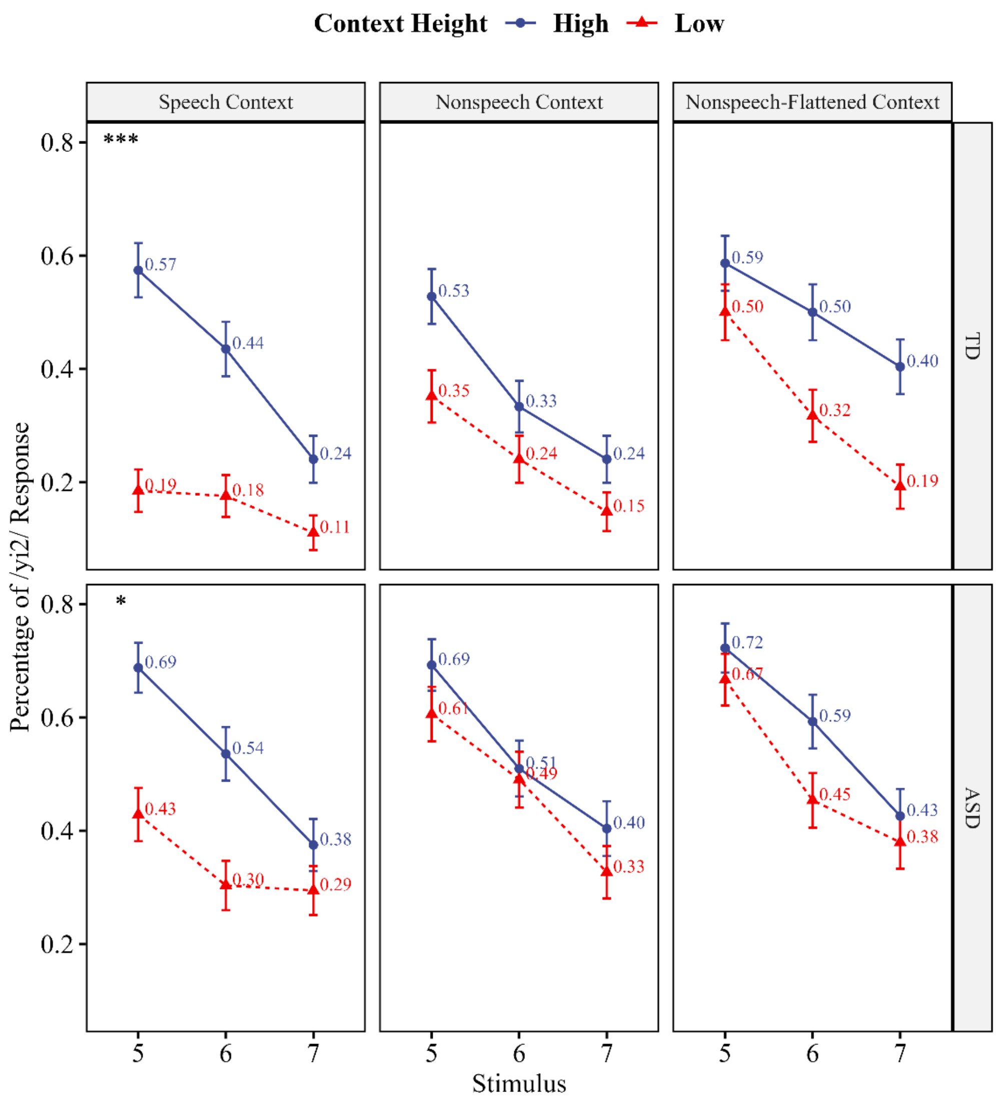

Fig. 2 The percentage of /yi2/ responses in speech, nonspeech, and nonspeech-flattened contexts for each group.

The error bars represent the standard error of the mean (± 1). The asterisks mean that the participants provide significantly more /yi2/ responses in high-F0 contexts than in low-F0 contexts, indicating a context-dependent tone identification. *$p < .05$, ***$p < .001$

Three GLMMs were constructed in speech, nonspeech, and nonspeech-flattened contexts, respectively. Table 2 presents the statistical results of the three GLMMs. In the speech contexts, there was a significant main effect of "Context Height" [χ^2 (1) = 5.32, p = .021] and a statistically significant interaction between "Context Height" and "Group" [χ^2 (1) = 5.10, p = .024]. To further analyze the interaction, post-hoc pairwise comparisons showed higher proportions of /yi2/ responses in high-F0 context in both ASD and TD children, but a larger difference between high- and low-F0 contexts was found in the TD group (β = 1.72, SE = 0.51, z = 3.41, $p < .001$) compared with the ASD group (β = 1.06, SE = 0.49, z = 2.17, p = .030). In the nonspeech and nonspeech-flattened contexts, no significant main effect or interaction was observed (all *p*s > .05), suggesting that no context effect was present in either ASD or TD children. Taken together, these results revealed that the context effect was only elicited when the preceding context was speech type, and that the TD group displayed a more salient context effect than the ASD group in speech contexts.

Table 2 Statistical results of GLMMs for the context-dependent identification of Mandarin tones in speech, nonspeech, and nonspeech-flattened contexts, respectively

Context Types	Effects	*df*	χ2	*p*
Speech	Group	1	2.92	.088
	Context Height	1	5.32	**.021***
	VIQ	1	0.12	.724
	NVIQ	1	2.52	.112
	WM	1	0.03	.863
	Group × Context Height	1	5.10	**.024***
Nonspeech	Group	1	3.26	.071

	Context Height	1	0.95	.330
	VIQ	1	0.89	.346
	NVIQ	1	1.82	.178
	WM	1	0.83	.363
	Group × Context Height	1	1.05	.305
Nonspeech-Flattened	Group	1	1.21	.272
	Context Height	1	1.50	.221
	VIQ	1	0.24	.625
	NVIQ	1	3.33	.068
	WM	1	0.19	.659
	Group × Context Height	1	1.21	.271

Note. Bold values indicate *p* values smaller than .05. GLMM = generalized linear mixed-effects model; VIQ = verbal IQ, NVIQ = non-verbal IQ, WM = working memory; *df* = degree of freedom. **p* < .05

Predicting the Size of Context Effect Based on Children's Cognitive Abilities

Based on the fact that we found significant contrastive context effects only in speech contexts, hierarchical multiple regressions were performed to investigate the effects of VIQ, NVIQ, and WM on the magnitude of context effect in speech contexts while controlling for age. The context effect size, which quantified the impact of contextual F0 cues on lexical tone identification, was calculated as the difference between the percentage of /yi2/ responses in the high-F0 context and the percentage of /yi2/ responses in the low-F0 context (Chen et al., 2023; Chen & Peng, 2016; Huang & Holt, 2011; K. Zhang et al., 2021). Considering the wide age range in the ASD and TD children, age was entered as a predictor in the first step, followed by VIQ, NVIQ, and WM (see Table 3).

Table 3 Statistical results of the hierarchical regression analyses

Group	Variables	*B*	*SE*	R^2 change	*F* change	*p*
TD ($n = 25$)	Step 1: Age	< 0.01	< 0.01	< .01	0.01	.911
	Step 2: VIQ	-0.01	< 0.01	.06	4.58	**.036***
	Step 3:					

	NVIQ	< 0.01	< 0.01	< .01	0.16	.694
	Step 4: WM	0.01	0.02	< .01	0.29	.595
ASD (n = 25)	Step 1: Age	< 0.01	< 0.01	.02	1.62	.207
	Step 2: VIQ	< 0.01	< 0.01	.01	0.58	.449
	Step 3: NVIQ	-0.01	0.01	.01	0.74	.394
	Step 4: WM	0.02	0.02	.01	0.67	.415

Note. Bold values indicate p values smaller than .05. VIQ = verbal IQ, NVIQ = non-verbal IQ, WM = working memory; B (raw beta) and SE (standard error of beta) are the final statistics of the full model, whereas R^2 change, F change, and p value are change statistics for the respective step as the last-entered step, representing its additional contribution. *$p < .05$

The regression analyses on TD children's context effect size revealed that after controlling for age, introducing VIQ explained an additional 6% of the variance, and that change in R^2 was statistically significant [F (1,72) = 4.58, p = .036]. The inclusion of TD children's NVIQ [F (1,71) = 0.16, p = .694] and WM [F (1,70) = 0.29, p = .595] accounted for less than 1% of the variance each, and the changes in R^2 values were not statistically significant. As for children with ASD, the current results indicated that after controlling for age, there was no significant contribution of VIQ, NVIQ, or WM (all ps > .05).

Discussion

The present study aimed to explore Mandarin-speaking ASD and TD children's context-dependent tone identification. In agreement with the neuroconstructivist model, children with ASD showed a reduced context effect compared to their TD peers, but the difference between the two groups was relatively subtle. Meanwhile, the context effect was only evident in speech contexts, suggesting a speech-specific mechanism. With respect to the correlation between cognitive abilities and the size of context effect, a higher VIQ was associated with a smaller context effect size in TD children. In contrast, the cognitive factors, including VIQ, NVIQ, and WM, failed to predict the magnitude of context effect in children with ASD.

Potential Explanations for the Reduced Context Effect in Children with ASD

In comparison with precedent findings that autistic individuals had difficulty in using contexts to derive the meaning of words and narratives (Happé, 1999; Nuske & Bavin, 2011), our study revealed a subtle reduction in sensitivity to acoustic contexts in speech among children with ASD. Our findings partially support the WCC theory from the auditory modality. Within the framework of the WCC model, the relatively reduced context effect in children with ASD could be roughly attributed to their impaired global processing, which requires a top-down, meaning-based approach. Of importance, while we identified a reduced context effect in the ASD group, the difference between the ASD and TD groups was relatively subtle. This finding suggests that the diminished context effect in ASD cannot be fully explained by the WCC theory alone.

In line with the neuroconstructivist model (Johnson, 2017; Johnson et al., 2015, 2021; Klin et al., 2020), our findings suggest that the performance of children with ASD might not simply be attributed to ongoing early sensory disruptions; rather, it may reflect early brain adaptation. The profiles of early perturbations and later neurocognitive modifiers interact to produce phenotypic outcomes, and strong neurocognitive modifier system activity can steer developmental trajectories toward typical outcomes (Johnson et al., 2021). In the current study, the subtle yet important differences in the utilization of acoustic contexts might indicate the potential impact of the underlying neurocognitive modifier system in ASD. It is crucial to closely monitor the nuanced atypicality in children with ASD, as these may provide valuable insights into the disorder's early developmental mechanisms. Indeed, evidence from studies involving infant siblings has shown early disruptions in attention to voices and faces in infants later diagnosed with ASD (Dawson et al., 2023). These previous findings, along with our results, underscore the importance of adopting a developmental perspective to enhance our understanding of ASD.

In our study, children with ASD could recognize the target tone based on the preceding speech contexts, yet the context effect was weaker than that seen in their TD counterparts. Thus, in contrast to the WCC theory, the neuroconstructivist model may represent a more powerful framework for understanding the complex processes involved in the acquisition and development of speech and language. However, prospective longitudinal studies with larger sample sizes and narrow age groups should build on the present evidence to further investigate developmental perceptual trajectories in ASD and to examine the application of the neuroconstructivist model from a developmental perspective.

Furthermore, their atypical performance might be related to their attenuated perceptual priors. As suggested by the Bayesian theory of autism (Pellicano & Burr, 2012), perception involves an inferential process of sensory stimuli that is influenced by the likelihood (i.e., one's internal response to the presented stimulus) and the priors (i.e., one's historical experiences that also inform predictions about the probability of stimuli in the environment). Weaker Bayesian priors may be responsible for the perceptual experiences observed in individuals with ASD. In other words, their perception is less influenced by prior experiences or preceding contexts. In our study, the inadequate use of contextual information in children with ASD could be a key factor behind the diminished impact of prior contexts when compared to their TD peers.

However, our findings regarding the varying susceptibility to speech contexts for ASD and TD children in tone perception contradict previous research on segment perception. Gabay et al. (2024) reported that autistic adults could felicitously use contexts to categorize speech sounds to a degree similar to their TD counterparts. In our results, although significant speech-specific context effects were detected in both ASD and TD children, the ASD group demonstrated a weaker contrastive context effect. The discrepant outcomes in our study and Gabay et al. (2024)'s research might be accounted for by several methodological factors. Unlike Gabay et al. (2024) who recruited ASD and TD adults, our study focused on ASD and TD children who may still be developing in cognitive and language domains. In terms of the spectral contexts and targets, we specifically examined spectral contexts with varying F0 levels, using real words as target sounds at both ends of the continuum. On the contrary, Gabay et al. (2024) concentrated on the influence of spectral center of gravity (COG) in their contexts, with the target forming a real word at one end of the continuum. Moreover, while we aimed to investigate low-level acoustic contexts in terms of spectral information (F0 information in particular), Gabay et al. (2024) investigated not only the low-level acoustic contexts but also high-level lexical contexts generated by long-term lexical knowledge (Ganong, 1980). Thus, further research is essential to elucidate the influence of different contexts on the perception of segments and suprasegments in the ASD group.

Speech-Specific Context Effect on Children's Lexical Tone Normalization

Consistent with previous findings of speech-specific context effects on Cantonese and Mandarin tone normalization in adults (Chen & Peng, 2016; Francis et al., 2006; C. Zhang et al., 2012; K. Zhang et al., 2017), the current experimental results reported that only speech contexts reliably facilitated tone normalization among Mandarin-speaking children with and without ASD. There are several possible explanations for the unequal effects of speech and nonspeech contexts. Firstly, the nonlinguistic nature of the nonspeech contexts might be a contributing factor (Chen & Peng, 2016; C. Zhang et al., 2012; K. Zhang et al., 2017). As these nonspeech contexts (nonspeech and nonspeech-flattened contexts) are acoustically different from the target stimuli and carry no linguistic meaning (H. Zhang et al., 2022), children may perceive them as unwanted noises (C. Zhang et al., 2012).

Secondly, children's restricted exposure to artificially-synthesized nonspeech sounds might be connected to the imbalanced effects of speech and nonspeech contexts (Chen & Peng, 2016; K. Zhang et al., 2017). Lee et al.

(1996)'s study found that Cantonese speakers were more accustomed to lexical tones compared to relatively unfamiliar non-lexical pitches, and they posited that this disparity in familiarity might explain why Cantonese speakers could effectively differentiate between lexical tones but struggled with closely-matched pitches present in nonwords. This assumption could also be extended to lexical tone normalization in Mandarin-speaking children with and without ASD, implying that their limited experience in processing nonspeech might result in a decreased impact of nonspeech contexts in contrast to speech contexts.

More importantly, our results provided empirical support for the speech-specific mechanism and reinforced the assertion that speech sounds are represented within a distinct phonetic module as the intended vocal gestures that produce them (Liberman et al., 1967; Liberman & Mattingly, 1985). The context-sensitive categorization of target sounds is a result of the phonetic module's utilization of learned information about the overlapping vocal gestures and their acoustic consequence (H. Zhang et al., 2022). Since nonspeech sounds are unlikely to overlap with speech sounds in vocal utterances, the perception of target speech sounds is expected to be unaffected by nonspeech precursors. In light of the speech-specific theory, different neural networks in human brains have evolved to process speech and nonspeech sounds separately (Liberman et al., 1967; Liberman & Mattingly, 1985). After being processed in the primary auditory cortex, speech and nonspeech signals are directed to distinct cortical regions (Bregman, 1990; Diehl & Kluender, 1989). Speech signals are transmitted to the auditory association cortex for comparison with the speech templates established through long-term linguistic exposure. Conversely, nonspeech signals, devoid of corresponding mental representations, are not subjected to the same processing pattern.

In addition, neither a context effect was found in the nonspeech contexts nor in the nonspeech-flattened contexts. On the one hand, the findings in the ASD group were at odds with the neural complexity hypothesis

(Samson et al., 2006). Based on the complexity hypothesis, both the speech and the original nonspeech contexts, characterized by dynamic changes in F0 contours, are expected to impose greater demands on auditory processing in individuals with ASD compared to nonspeech-flattened contexts. The current study, however, found that participants with ASD were able to integrate the complex auditory cues in speech contexts. It is possible that the speech-specific mechanism may play a larger role than the stimulus complexity. Still, there is a need to revisit the understanding of how auditory processing operates in autistic individuals' lexical tone normalization. On the other hand, the current findings did not identify which type of F0 information was more important in lexical tone normalization across the two nonspeech contexts. As discussed earlier, the lack of context effects in both nonspeech and nonspeech-flattened contexts might be attributed to the nonlinguistic nature of the nonspeech stimuli (Chen & Peng, 2016; C. Zhang et al., 2012; K. Zhang et al., 2017), children's limited exposure to artificially generated nonspeech sounds (Chen & Peng, 2016; K. Zhang et al., 2017), and the distinct neural pathways involved in processing speech versus nonspeech auditory stimuli (Liberman et al., 1967; Liberman & Mattingly, 1985). Consequently, both ASD and TD children in our study may lack the inclination to extract a talker's full F0 range from the F0 contour in nonspeech contexts, or to derive a talker's mean speaking F0 in nonspeech-flattened contexts for the purpose of normalizing Mandarin tones, which suggests that any preference for F0 range or mean F0 is unlikely to emerge. Nonetheless, given the sample size and the relatively high functioning of the recruited autistic children, caution is advised in interpreting these statistical results.

One novel finding in the present study is that Mandarin-speaking children with ASD, ranging from 6 to 12 years of age and without severe cognitive deficits, may develop distinct neural networks for processing speech and nonspeech stimuli in their brains. However, their ability to effectively utilize these neural systems could be constrained, as evidenced by their reduced context effect in comparison with the TD group. Future investigations employing electrophysiological techniques are crucial to uncover the underlying mechanisms behind these findings.

Relationship between Cognitive Factors and the Size of Context Effect

Our study found that higher VIQ was associated with a smaller size of contrastive context effect in speech contexts among TD children. A great bulk of literature has confirmed a correlation between high VIQ and enhanced categorical perception competence (e.g., greater accuracy in distinguishing between categories and narrower boundary width) (Chen et al., 2016; Chen & Peng, 2021; Rong et al., 2023; Stewart et al., 2018). In our study, we assessed the size of context effect in the identification of ambiguous tones from the middle of the continuum. It is likely that TD children with higher VIQ rely less on contextual information when judging these ambiguous tones, possibly due to their better categorical perception skills associated with higher VIQ. Alternatively, TD controls with higher VIQ might adopt a more flexible strategy in our experimental task. Since child participants were not explicitly instructed to utilize the contexts when identifying target tones and our two-alternative forced-choice task can be less demanding for TD children with higher VIQ, their relatively richer linguistic experience might allow them to focus on the target sound at the end and identify the target tone based on the intrinsic cues only. Yet, this finding should be interpreted with caution, and future studies with larger sample sizes are still needed to further investigate the role of VIQ.

Inconsistent with previous findings (Chen & Peng, 2021; Rong et al., 2023; L. Wang et al., 2023), this research failed to uncover a significant link between the VIQ of children with ASD and their performance in normalizing lexical tones. Our research examined the role of VIQ in a high-level pitch processing task requiring the integration of pitch information in the speech contexts and the target lexical tone. In contrast, earlier investigations predominantly adopted a low-level pitch processing task to examine pitch processing in lexical tones or intonations alone (Chen & Peng, 2021; Rong et al., 2023; L. Wang et al., 2023). Therefore, autistic children's VIQ may exert a greater impact on local pitch processing than global pitch processing. Future investigations should delve deeper into the potential connections between verbal abilities and local/global pitch processing in individuals with ASD. Moreover, the ASD participants we recruited were children with relatively higher verbal abilities (see Table 1), with a mean VIQ of 99. However, precursor studies like Rong et al. (2023), which revealed a notable positive association between their verbal abilities and lexical tone perception, enrolled autistic children with severe language delays. It is important to note that the VIQ difference between ASD and TD groups in Rong et al. (2023)'s research was larger than that in our study. Hence, it is conceivable that the predictive role of VIQ in lexical tone normalization may be more prominent in autistic children with significant language delays.

The results of the study, however, do not support an association between NVIQ and context effect size in both groups. It is important to consider that the lexical tone normalization task in this study entails a more top-down approach, whereas the nonverbal reasoning tasks in the WISC-IV require more bottom-up processing (Chowdhury et al., 2017). Interestingly, previous research using the Raven's Progressive Matrices (Raven, 1938) as NVIQ assessment suggested that NVIQ was a reliable predictor of autistic individuals' pitch perception (L. Wang et al., 2023), and responses to the Raven matrices are commonly attributed to the utilization of top-down analytical reasoning processes (de Winter et al., 2023). Therefore, additional work is needed to compare the efficacy of cognitive measurements when investigating the potential association between NVIQ and auditory pitch processing.

We hypothesized that WM might play an active role in the process of lexical tone normalization. Listeners would need to maintain contextual cues within their WM systems to facilitate comparison with the target stimuli.

Unexpectedly, our current results revealed that WM did not serve as a significant predictor of the context effect magnitude in either ASD or TD children. This finding was in agreement with the research by Chen et al. (2023), which showed that the performance of Mandarin-speaking TD children in Mandarin tone normalization was unaffected by their WM capacity. These outcomes indicate that executive functions except WM, such as attention, inhibitory control, and mental flexibility, may be more crucial in supporting lexical tone normalization in both ASD and TD children. The present study merely investigated the role of WM in children's normalization of lexical tones. To provide a clearer picture of the cognitive strategies employed by autistic children in lexical tone normalization, future studies should examine the impacts of these additional executive functions, especially given the well-documented executive dysfunction in the ASD population (Demetriou et al., 2018).

According to the neuroconstructivist model (Johnson et al., 2015), autism exhibits significant genetic heterogeneity, resulting in extensive and varied differences across multiple brain regions and systems. It also demonstrates considerable phenotypic diversity, encompassing a broad spectrum of IQ and language abilities with varying profiles of strengths and weaknesses, and comorbidities with common psychiatric conditions such as anxiety, mood disorders, and attention disorders (Klin et al., 2020; Tordjman et al., 2017). While this study examined the impact of phenotypic heterogeneity on IQ and memory functions within children with ASD, it is important to note that we only included participants with relatively higher functioning. Therefore, our sample may not represent the entire autism spectrum, emphasizing the need for future studies to include a broader range of participants across the autism spectrum.

Conclusion

In conclusion, the present study delved into the normalization of lexical tones in Mandarin-speaking children with and without ASD. Our results uncovered a nuanced but important difference between ASD and TD children in the utilization of contextual F0 cues to identify lexical tones. Meanwhile, we found that ASD and TD children's lexical tone normalization was influenced only by the speech contexts, thereby lending support to the speech-specific account. Furthermore, our results also reveal that TD children with higher VIQ exhibit a smaller context effect size, indicating that they are less susceptible to extrinsic contexts when identifying the target tonal categories. Overall, our findings provide valuable insights into how Mandarin-speaking children with and without ASD normalize lexical tones in speech and nonspeech contexts. These insights may inform clinicians in their development of speech training strategies, such as incorporating more speech contexts, that could potentially enhance lexical tone processing and acquisition in children with ASD, though further research and clinical validation are warranted before implementation. Nevertheless, the limited sample size and wide age range in this study might potentially undermine the robustness and generalizability of our conclusions. Moving forward, further investigations with a larger sample size and more narrowly defined age groups would be advantageous in better understanding the atypical perceptual processing in children with ASD.

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

American Psychiatric Association. (2013). *Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition*. American Psychiatric Publishing.

American Psychiatric Association. (2022). *Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition, Text Revision*. American Psychiatric Association.

Baddeley, A. (1992). Working Memory. *Science*, *255*(5044), 556–559. https://doi.org/10.1126/science.1736359

Baddeley, A. (2000). The episodic buffer: A new component of working memory? *Trends in Cognitive Sciences*, *4*(11), 417–423. https://doi.org/10.1016/S1364-6613(00)01538-2

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Boersma, P., & Weenink, D. (2021). *Praat: Doing phonetics by computer (Computer program, version 6.2.03)*. http://www.praat.org/

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

Booth, R. D. L., & Happé, F. G. E. (2018). Evidence of reduced global processing in autism spectrum disorder. *Journal of Autism and Developmental Disorders*, *48*(4), 1397–1408. https://doi.org/10.1007/s10803-016-2724-6

Bregman, A. S. (1990). *Auditory scene analysis*. MIT Press.

Chen, F., & Peng, G. (2016). Context effect in the categorical perception of Mandarin tones. *Journal of Signal Processing Systems*, *82*(2), 253–261. https://doi.org/10.1007/s11265-015-1008-2

Chen, F., & Peng, G. (2021). Categorical perception of pitch contours and voice onset time in Mandarin-speaking adolescents with autism spectrum disorders. *Journal of Speech, Language, and Hearing Research*, *64*(11), 4468–4484. https://doi.org/10.1044/2021_JSLHR-20-00725

Chen, F., Peng, G., Yan, N., & Wang, L. (2017). The development of categorical perception of Mandarin tones in four- to seven-year-old children. *Journal of Child Language*, *44*(6), 1413–1434. https://doi.org/10.1017/S0305000916000581

Chen, F., Yan, N., Pan, X., Yang, F., Ji, Z., Wang, L., & Peng, G. (2016). Impaired categorical perception of Mandarin tones and its relationship to language ability in autism spectrum disorders. *Proceedings of Interspeech*, 233–237. https://doi.org/10.21437/Interspeech.2016-1133

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, F., Zhang, K., Guo, Q., & Lv, J. (2023). Development of achieving constancy in lexical tone identification with contextual cues. *Journal of Speech, Language, and Hearing Research*, *66*(4), 1148–1164. https://doi.org/10.1044/2022_JSLHR-22-00257

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

Chung, W.-L., & Bidelman, G. M. (2021). Mandarin-speaking preschoolers' pitch discrimination, prosodic and phonological awareness, and their relation to receptive vocabulary and reading abilities. *Reading and Writing*, *34*(2), 337–353. https://doi.org/10.1007/s11145-020-10075-9

D'Souza, D., Booth, R., Connolly, M., Happé, F., & Karmiloff-Smith, A. (2016). Rethinking the concepts of 'local or global processors': Evidence from Williams syndrome, Down syndrome, and autism spectrum disorders. *Developmental Science*, *19*(3), 452–468. https://doi.org/10.1111/desc.12312

Dawson, G., Rieder, A. D., & Johnson, M. H. (2023). Prediction of autism in infants: progress and challenges. *The Lancet Neurology*, *22*(3), 244–254. https://doi.org/10.1016/S1474-4422(22)00407-0

de Winter, J. C. F., Dodou, D., & Eisma, Y. B. (2023). Responses to Raven matrices: Governed by visual complexity and centrality. *Perception*, *52*(9), 645–661. https://doi.org/10.1177/03010066231178149

Demetriou, E. A., Lampit, A., Quintana, D. S., Naismith, S. L., Song, Y. J. C., Pye, J. E., Hickie, I., & Guastella, A. J. (2018). Autism spectrum disorders: A meta-analysis of executive function. *Molecular Psychiatry*, *23*(5), 1198–1204. https://doi.org/10.1038/mp.2017.75

Deng, X., & Yip, V. (2018). A Multimedia Corpus of Child Mandarin: The Tong Corpus. *Journal of Chinese Linguistics*, *46*(1), 69–92. https://doi.org/10.1353/jcl.2018.0002

Diehl, R. L., & Kluender, K. R. (1989). On the objects of speech perception. *Ecological Psychology*, *1*(2), 121–144. https://doi.org/10.1207/s15326969eco0102_2

Diehl, R. L., Lotto, A. J., & Holt, L. L. (2004). Speech perception. *Annual Review of Psychology*, *55*, 149–179. https://doi.org/10.1146/annurev.psych.55.090902.142028

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

Eigsti, I.-M., & Fein, D. A. (2013). More is less: Pitch discrimination and language delays in children with optimal outcomes from autism. *Autism Research*, *6*(6), 605–613. https://doi.org/10.1002/aur.1324

Elsabbagh, M., Mercure, E., Hudry, K., Chandler, S., Pasco, G., Charman, T., Pickles, A., Baron-Cohen, S., Bolton, P., & Johnson, M. H. (2012). Infant neural sensitivity to dynamic eye gaze is associated with later emerging autism. *Current Biology*, *22*(4), 338–342. https://doi.org/10.1016/j.cub.2011.12.056

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*, *10*(11), 1845–1856. https://doi.org/10.1002/aur.1820

Foxton, J. M., Stewart, M. E., Barnard, L., Rodgers, J., Young, A. H., O'Brien, G., & Griffiths, T. D. (2003).

Absence of auditory “global interference” in autism. *Brain*, *126*(12), 2703–2709. https://doi.org/10.1093/brain/awg274

Francis, A. L., Ciocca, V., Wong, N. K. Y., Leung, W. H. Y., & Chu, P. C. Y. (2006). Extrinsic context affects perceptual normalization of lexical tone. *The Journal of the Acoustical Society of America*, *119*(3), 1712–1726. https://doi.org/10.1121/1.2149768

Frith, U. (1989). *Autism: Explaining the enigma.* Basil Blackwell.

Gabay, Y., Reinisch, E., Even, D., Binur, N., & Hadad, B.-S. (2024). Intact utilization of contextual information in speech categorization in autism. *Journal of Autism and Developmental Disorders*. *54*(10), 3837–3853. https://doi.org/10.1007/s10803-023-06106-3

Ganong, W. F. (1980). Phonetic categorization in auditory word perception. *Journal of Experimental Psychology: Human Perception and Performance*, *6*(1), 110–125. https://doi.org/10.1037/0096-1523.6.1.110

Gathercole, S. E., & Baddeley, A. D. (2014). *Working memory and language*. Psychology Press. https://doi.org/10.4324/9781315804682

Gauthier, I., Klaiman, C., & Schultz, R. T. (2009). Face composite effects reveal abnormal face processing in autism spectrum disorders. *Vision Research*, *49*(4), 470–478. https://doi.org/10.1016/j.visres.2008.12.007

Green, P., & MacLeod, C. J. (2016). SIMR: An R package for power analysis of generalized linear mixed models by simulation. *Methods in Ecology and Evolution*, *7*(4), 493–498. https://doi.org/10.1111/2041-210X.12504

Groen, W. B., van Orsouw, L., Huurne, N. Ter, 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

Hadad, B.-S., & Ziv, Y. (2015). Strong bias towards analytic perception in ASD does not necessarily come at the price of impaired integration skills. *Journal of Autism and Developmental Disorders*, *45*(6), 1499–1512. https://doi.org/10.1007/s10803-014-2293-5

Haesen, B., Boets, B., & Wagemans, J. (2011). A review of behavioural and electrophysiological studies on auditory processing and speech perception in autism spectrum disorders. *Research in Autism Spectrum Disorders*, *5*(2), 701–714. https://doi.org/10.1016/j.rasd.2010.11.006

Happé, F. (1999). Autism: Cognitive deficit or cognitive style? *Trends in Cognitive Sciences*, *3*(6), 216–222. https://doi.org/10.1016/S1364-6613(99)01318-2

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

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

Heaton, P. (2005). Interval and contour processing in autism. *Journal of Autism and Developmental Disorders*, *35*(6), 787–793. https://doi.org/10.1007/s10803-005-0024-7

Heaton, P., Davis, R. E., & Happé, F. G. E. (2008). Research note: Exceptional absolute pitch perception for spoken words in an able adult with autism. *Neuropsychologia*, *46*(7), 2095–2098. https://doi.org/10.1016/j.neuropsychologia.2008.02.006

Henderson, L. M., Clarke, P. J., & Snowling, M. J. (2011). Accessing and selecting word meaning in autism spectrum disorder. *Journal of Child Psychology and Psychiatry*, *52*(9), 964–973. https://doi.org/10.1111/j.1469-7610.2011.02393.x

Holt, L. L. (2006). Speech categorization in context: Joint effects of nonspeech and speech precursors. *The Journal* *of the Acoustical Society of America*, *119*(6), 4016–4026. https://doi.org/10.1121/1.2195119

Huang, J., & Holt, L. L. (2009). General perceptual contributions to lexical tone normalization. *The Journal of the Acoustical Society of America*, *125*(6), 3983–3994. https://doi.org/10.1121/1.3125342

Huang, J., & Holt, L. L. (2011). Evidence for the central origin of lexical tone normalization (L). *The Journal of the Acoustical Society of America*, *129*(3), 1145–1148. https://doi.org/10.1121/1.3543994

Hudry, K., Leadbitter, K., Temple, K., Slonims, V., McConachie, H., Aldred, C., Howlin, P., & Charman, T., & PACT Consortium. (2010). Preschoolers with autism show greater impairment in receptive compared with expressive language abilities. *International Journal of Language & Communication Disorders*, *45*(6), 681–690. https://doi.org/10.3109/13682820903461493

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

Järvinen-Pasley, A., Peppé, S., King-Smith, G., & Heaton, P. (2008). The relationship between form and function level receptive prosodic abilities in autism. *Journal of Autism and Developmental Disorders*, *38*(7), 1328–1340. https://doi.org/10.1007/s10803-007-0520-z

Järvinen-Pasley, A., Wallace, G. L., Ramus, F., Happé, F., & Heaton, P. (2008). Enhanced perceptual processing of speech in autism. *Developmental Science*, *11*(1), 109–121. https://doi.org/10.1111/j.1467-7687.2007.00644.x

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

Johnson, M. H. (2017). Autism as an adaptive common variant pathway for human brain development. *Developmental Cognitive Neuroscience*, *25*, 5–11. https://doi.org/10.1016/j.dcn.2017.02.004

Johnson, M. H., Charman, T., Pickles, A., & Jones, E. J. H. (2021). Annual research review: Anterior Modifiers in the Emergence of Neurodevelopmental Disorders (AMEND)—a systems neuroscience approach to common developmental disorders. *Journal of Child Psychology and Psychiatry*, *62*(5), 610–630. https://doi.org/10.1111/jcpp.13372

Johnson, M. H., Jones, E. J. H., & Gliga, T. (2015). Brain adaptation and alternative developmental trajectories. *Development and Psychopathology*, *27*(2), 425–442. https://doi.org/10.1017/S0954579415000073

Jones, C. R. G., Happé, F., Baird, G., Simonoff, E., Marsden, A. J. S., Tregay, J., Phillips, R. J., Goswami, U., Thomson, J. M., & Charman, T. (2009). Auditory discrimination and auditory sensory behaviours in autism spectrum disorders. *Neuropsychologia*, *47*(13), 2850–2858. https://doi.org/10.1016/j.neuropsychologia.2009.06.015

Jones, E. J. H., Venema, K., Earl, R., Lowy, R., Barnes, K., Estes, A., Dawson, G., & Webb, S. J. (2016). Reduced engagement with social stimuli in 6-month-old infants with later autism spectrum disorder: a longitudinal prospective study of infants at high familial risk. *Journal of Neurodevelopmental Disorders*, *8*, 7. https://doi.org/10.1186/s11689-016-9139-8

Jones, W., & Klin, A. (2013). Attention to eyes is present but in decline in 2–6-month-old infants later diagnosed with autism. *Nature*, *504*(7480), 427–431. https://doi.org/10.1038/nature12715

Kanner, L. (1943). Autistic disturbances of affective contact. *Nervous Child*, *2*, 217–250.

Klin, A., Micheletti, M., Klaiman, C., Shultz, S., Constantino, J. N., & Jones, W. (2020). Affording autism an early brain development re-definition. *Development and Psychopathology*, *32*(4), 1175–1189.

https://doi.org/10.1017/S0954579420000802

Kujala, T., Lepistö, T., & Näätänen, R. (2013). The neural basis of aberrant speech and audition in autism spectrum disorders. *Neuroscience and Biobehavioral Reviews*, *37*(4), 697–704. https://doi.org/10.1016/j.neubiorev.2013.01.006

Lau, J. C. Y., To, C. K. S., Kwan, J. S. K., Kang, X., Losh, M., & Wong, P. C. M. (2021). Lifelong tone language experience does not eliminate deficits in neural encoding of pitch in autism spectrum disorder. *Journal of Autism and Developmental Disorders*, *51*(9), 3291–3310. https://doi.org/10.1007/s10803-020-04796-7

Lee, Y.-S., Vakoch, D. A., & Wurm, L. H. (1996). Tone perception in Cantonese and Mandarin: A cross-linguistic comparison. *Journal of Psycholinguistic Research*, *25*(5), 527–542. https://doi.org/10.1007/BF01758181

Lenth, R. (2018). *Estimated marginal Means, aka Least-Squares means.* https://github.com/rvlenth/emmeans

Liberman, A. M., Cooper, F. S., Shankweiler, D. P., & Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychological Review*, *74*(6), 431–461. https://doi.org/10.1037/h0020279

Liberman, A. M., & Mattingly, I. G. (1985). The motor theory of speech perception revised. *Cognition*, *21*(1), 1–36. https://doi.org/10.1016/0010-0277(85)90021-6

Lindell, A. K., & Hudry, K. (2013). Atypicalities in cortical structure, handedness, and functional lateralization for language in autism spectrum disorders. *Neuropsychology Review*, *23*(3), 257–270. https://doi.org/10.1007/s11065-013-9234-5

Lloyd‐Fox, S., Blasi, A., Pasco, G., Gliga, T., Jones, E. J. H., Murphy, D. G. M., Elwell, C. E., Charman, T., & Johnson, M. H. (2018). Cortical responses before 6 months of life associate with later autism. *European Journal of Neuroscience*, *47*(6), 736–749. https://doi.org/10.1111/ejn.13757

Lord, C., Rutter, M., DiLavore, P. C., Risi, S., Gotham, K., & Bishop, S. (2012). *Autism diagnostic observation* *schedule: ADOS–2*. Western Psychological Services.

MacWhinney, B. (2000). *The CHILDES project: The database*. Psychology Press.

Mayer, J. L., Hannent, I., & Heaton, P. F. (2016). Mapping the developmental trajectory and correlates of enhanced pitch perception on speech processing in adults with ASD. *Journal of Autism and Developmental Disorders*, *46*(5), 1562–1573. https://doi.org/10.1007/s10803-014-2207-6

McHaney, J. R., Tessmer, R., Roark, C. L., & Chandrasekaran, B. (2021). Working memory relates to individual differences in speech category learning: Insights from computational modeling and pupillometry. *Brain and Language*, *222*, 105010. https://doi.org/10.1016/j.bandl.2021.105010

Mecca, T. P., Orsati, F. T., & de Macedo, E. C. (2014). Non-verbal cognitive profile of young children with autism spectrum disorders. *Psychology*, *05*(11), 1404–1417. https://doi.org/10.4236/psych.2014.511151

Meilleur, A., Foster, N. E. V., Coll, S.-M., Brambati, S. M., & Hyde, K. L. (2020). Unisensory and multisensory temporal processing in autism and dyslexia: A systematic review and meta-analysis. *Neuroscience and Biobehavioral Reviews*, *116*, 44–63. https://doi.org/10.1016/j.neubiorev.2020.06.013

Millman, R. E., & Mattys, S. L. (2017). Auditory verbal working memory as a predictor of speech perception in modulated maskers in listeners with normal hearing. *Journal of Speech, Language, and Hearing Research*, *60*(5), 1236–1245. https://doi.org/10.1044/2017_JSLHR-S-16-0105

Mottron, L., & Burack, J. A. (2001). Enhanced perceptual functioning in the development of autism. In: J. A. Burack, T. Charman, N. Yirmiya, & P. R. Zelazo (Eds.), *The development of autism: Perspectives from theory and research*. (pp. 131–148). Erlbaum.

Mottron, L., & Bzdok, D. (2020). Autism spectrum heterogeneity: fact or artifact? *Molecular Psychiatry*, *25*(12), 3178–3185. https://doi.org/10.1038/s41380-020-0748-y

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

Nuske, H. J., & Bavin, E. L. (2011). Narrative comprehension in 4-7-year-old children with autism: Testing the Weak Central Coherence account. *International Journal of Language and Communication Disorders*, *46*(1), 108–119. https://doi.org/10.3109/13682822.2010.484847

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

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

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

Ozonoff, S., Iosif, A.-M., Baguio, F., Cook, I. C., Hill, M. M., Hutman, T., Rogers, S. J., Rozga, A., Sangha, S., Sigman, M., Steinfeld, M. B., & Young, G. S. (2010). A Prospective Study of the Emergence of Early Behavioral Signs of Autism. *Journal of the American Academy of Child & Adolescent Psychiatry*, *49*(3), 256-266.e2. https://doi.org/10.1016/j.jaac.2009.11.009

Ozonoff, S., & Strayer, D. L. (2001). Further evidence of intact working memory in autism. *Journal of Autism and Developmental Disorders*, *31*(3), 257–263. https://doi.org/10.1023/A:1010794902139

Pellicano, E., & Burr, D. (2012). When the world becomes "too real": A Bayesian explanation of autistic perception. *Trends in Cognitive Sciences*, *16*(10), 504–510. https://doi.org/10.1016/j.tics.2012.08.009

Peng, G., Zhang, C., Zheng, H.-Y., Minett, J. W., & Wang, W. S.-Y. (2012). The effect of intertalker variations on acoustic-perceptual mapping in Cantonese and Mandarin tone systems. *Journal of Speech, Language, and Hearing Research*, *55*(2), 579–595. https://doi.org/10.1044/1092-4388(2011/11-0025)

Peng, G., Zheng, H.-Y., Gong, T., Yang, R.-X., Kong, J.-P., & Wang, W. S.-Y. (2010). The influence of language experience on categorical perception of pitch contours. *Journal of Phonetics*, *38*(4), 616–624. https://doi.org/10.1016/j.wocn.2010.09.003

Psychology Software Tools. (2012). *E-Prime 2.0*. Pittsburgh, USA. https://www.pstnet.com/

R Core Team. (2023). *R: The R Project for Statistical Computing*. https://www.r-project.org/

Raven, J. C. (1938). *Progressive matrices: A perceptual test of intelligence*. H. K. Lewis.

Rong, Y., Weng, Y., Chen, F., & Peng, G. (2023). Categorical perception of Mandarin lexical tones in language-delayed autistic children. *Autism*, *27*(5), 1426–1437. https://doi.org/10.1177/13623613221138687

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

Sota, S., Hatada, S., Honjyo, K., Takatsuka, T., Honer, W. G., Morinobu, S., & Sawada, K. (2018). Musical disability in children with autism spectrum disorder. *Psychiatry Research*, *267*, 354–359. https://doi.org/10.1016/j.psychres.2018.05.078

Stanutz, S., Wapnick, J., & Burack, J. A. (2014). Pitch discrimination and melodic memory in children with autism spectrum disorders. *Autism*, *18*(2), 137–147. https://doi.org/10.1177/1362361312462905

Stewart, M. E., Petrou, A. M., & Ota, M. (2018). Categorical speech perception in adults with autism spectrum conditions. *Journal of Autism and Developmental Disorders*, *48*(1), 72–82. https://doi.org/10.1007/s10803-

017-3284-0

Tordjman, S., Cohen, D., Anderson, G. M., Botbol, M., Canitano, R., Coulon, N., & Roubertoux, P. L. (2017). Reframing autism as a behavioral syndrome and not a specific mental disorder: Implications of genetic and phenotypic heterogeneity. *Neuroscience and Biobehavioral Reviews*, *80*, 210. https://doi.org/10.1016/j.neubiorev.2017.01.030

Van Hedger, S. C., Heald, S. L. M., Koch, R., & Nusbaum, H. C. (2015). Auditory working memory predicts individual differences in absolute pitch learning. *Cognition*, *140*, 95–110. https://doi.org/10.1016/j.cognition.2015.03.012

Volkmar, F. R. (2021). *Encyclopedia of Autism Spectrum Disorders*. Springer International Publishing.

Wang, L., Xiao, S., Jiang, C., Hou, Q., Chan, A. H. D., Wong, P. C. M., & Liu, F. (2023). The form and function processing of lexical tone and intonation in tone-language-speaking children with autism spectrum disorder. *The Journal of the Acoustical Society of America*, *154*(1), 467–481. https://doi.org/10.1121/10.0020271

Wang, W. S.-Y. (1972). The many uses of F0. In A. Valdman (Ed.), *Papers in Linguistics and Phonetics to the Memory of Pierre Delattre* (Vol. 76, Issue 1, pp. 487–503). Mouton. https://doi.org/10.1515/9783110803877-041

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

Wechsler, D. (2003). *Wechsler Intelligence Scale for Children–Fourth Edition*. The Psychological Corporation.

Williams, D. L., Goldstein, G., Carpenter, P. A., & Minshew, N. J. (2005). Verbal and spatial working memory in autism. *Journal of Autism and Developmental Disorders*, *35*(6), 747–756. https://doi.org/10.1007/s10803-

005-0021-x

Wong, P. C. M., & Diehl, R. L. (2003). Perceptual normalization for inter- and intratalker variation in Cantonese level tones. *Journal of Speech, Language, and Hearing Research*, *46*(2), 413–421. https://doi.org/10.1044/1092-4388(2003/034)

Yan, J., Chen, F., Gao, X., & Peng, G. (2021). Auditory-motor mapping training facilitates speech and word learning in tone language–speaking children with autism: An early efficacy study. *Journal of Speech, Language, and Hearing Research*, *64*(12), 4664–4681. https://doi.org/10.1044/2021_JSLHR-21-00029

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

Zhang, C., Peng, G., & Wang, W. S.-Y. (2012). Unequal effects of speech and nonspeech contexts on the perceptual normalization of Cantonese level tones. *The Journal of the Acoustical Society of America*, *132*(2), 1088–1099. https://doi.org/10.1121/1.4731470

Zhang, H., Ding, H., & Lee, W.-S. (2022). The influence of preceding speech and nonspeech contexts on Mandarin tone identification. *Journal of Phonetics*, *93*, 101154. https://doi.org/10.1016/j.wocn.2022.101154

Zhang, K., Sjerps, M. J., & Peng, G. (2021). Integral perception, but separate processing: The perceptual normalization of lexical tones and vowels. *Neuropsychologia*, *156*, 107839. https://doi.org/10.1016/j.neuropsychologia.2021.107839

Zhang, K., Wang, X., & Peng, G. (2017). Normalization of lexical tones and nonlinguistic pitch contours: Implications for speech-specific processing mechanism. *The Journal of the Acoustical Society of America*, *141*(1), 38–49. https://doi.org/10.1121/1.4973414

Zhao, C., Ong, J. H., Veic, A., Patel, A. D., Jiang, C., Fogel, A. R., Wang, L., Hou, Q., Das, D., Crasto, C., Chakrabarti, B., Williams, T. I., Loutrari, A., & Liu, F. (2024). Predictive processing of music and language in autism: Evidence from Mandarin and English speakers. *Autism Research*, *17*(6), 1230–1257. https://doi.org/10.1002/aur.3133

Zhu, M., Chen, F., Chen, X., & Yang, Y. (2023). The more the better? Effects of L1 tonal density and typology on the perception of non-native tones. *PLOS ONE*, *18*(9), e0291828. https://doi.org/10.1371/journal.pone.0291828