The following publication 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 is available at https://dx.doi.org/10.1044/2021_JSLHR-20-00725. The Journal of Speech, Language, and Hearing Research is available at https://pubs.asha.org/journal/jslhr.

Categorical Perception of Pitch Contours and Voice Onset Time in Mandarin-Speaking Adolescents with Autism Spectrum Disorders

Fei Chen [a, *], Gang Peng [b, *]

[a] School of Foreign Languages, Hunan University, Changsha, China

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

[*] **Corresponding Authors:**

Gang Peng: gpengjack@gmail.com

Tel: (+852) 3400 8462; Fax: (+852) 2334 0185

Fei Chen: chenfeianthony@gmail.com

Tel: (+86) 17788926788

Conflict of Interest: The authors have declared that no competing interests existed at the time of publication.

Funding Statement: This study was supported by grants from the Major Program of National Social Science Foundation of China (grant number 18ZDA293), the Social Science Foundation of Hunan Province (grant number 20ZDB003), and the Fundamental Research Funds for the Central Universities, Hunan University (grant number 53111801066).

Abstract

Purpose: Previous studies have shown enhanced pitch and impaired time perception in individuals with autism spectrum disorders (ASD). However, it remains unclear whether such deviated patterns of auditory processing depending on acoustic dimensions would transfer to the higher-level linguistic pitch and time processing. In this study, we compared the categorical perception (CP) of lexical tones and voice onset time (VOT) in Mandarin Chinese, which utilize pitch and time changes respectively to convey phonemic contrasts.

Method: The data were collected from 22 Mandarin-speaking adolescents with ASD and 20 age-matched neurotypical controls. In addition to the identification and discrimination tasks to test CP performance, all the participants were evaluated with their language ability and phonological working memory. Linear mixed-effects models were constructed to evaluate the identification and discrimination scores across different groups and conditions.

Results: The basic CP pattern of cross-boundary benefit when perceiving both native lexical tones and VOT was largely preserved in high-functioning adolescents with ASD. The degree of CP of lexical tones in ASD was similar to that in typical controls, whereas the degree of CP of VOT in ASD was greatly reduced. Furthermore, the degree of CP of lexical tones correlated with language ability and digit span in ASD participants.

Conclusions: These findings suggest that the unbalanced acoustic processing capacities for pitch and time can be generalized to the higher-level linguistic processing in ASD. Furthermore, the higher degree of CP of lexical tones correlated with better language ability in Mandarin-speaking individuals with ASD.

Keywords: lexical tone, VOT, categorical perception, pitch contours, Mandarin, ASD

Categorical Perception of Pitch Contours and Voice Onset Time in Mandarin-Speaking Adolescents with Autism Spectrum Disorders

1 Introduction

The speech and language difficulties are central to autism spectrum disorders (ASD) and are one of the critical signs and symptoms for establishing a diagnosis with ASD (Landa, 2008; You et al., 2017). However, the mechanisms that underlie speech and language delay/impairment in ASD remain poorly understood. At the basic acoustic processing level, individuals with ASD tended to show unbalanced auditory perceptual skills towards different aspects of spectral vs. temporal cues, as represented by hypersensitive processing of pitch and hyposensitive discrimination of sound duration (see Haesen et al., 2011; O'connor, 2012 for reviews).

Specifically, clinical observations reported that a small portion of low-functioning autistic children owned 'islands of genius', such as case descriptions of musical savants with autism owning absolute or 'perfect' pitch (Heaton et al., 1998; Kanner, 1943). In addition, more and more research suggested that compared to typically developing (TD) controls, both low- and high-functioning children with ASD (Heaton, 2003, 2005; Heaton et al., 2001, 2008; O'Riordan & Passetti, 2006) and adolescents/adults with ASD (Bonnel et al., 2003; Foxton et al., 2003; Mottron et al., 2000) had a better pitch memory and was generally more accurate at the processing of melodic pitch contours, and better at identification and discrimination of pitch changes in pure-tone stimuli (a nonspeech material). In stark contrast, both low- and high-functioning children with ASD (Brodeur et al., 2014; Maister & Plaisted-Grant, 2011; Szelag et al., 2004) and adolescents/adults with ASD (Falter et al., 2012; Martin et al., 2010) tended to show poorer performance in the basic auditory processing of sound duration from the evidence of both behavioral and neuroimaging studies, indicating that timing impairments may underpin core features of ASD. As suggested by the functional hypothesis (Lancker, 1980), the accurate and complete perception of speech sounds in native language speakers involves the processing of acoustic information as well as phonological information. It would be meaningful to investigate whether and how the differential auditory and acoustic sensitivity to spectral and temporal cues in ASD extended to the higher-level phonological processing.

Mandarin provides us an ideal opportunity to compare the perception of spectral and temporal information, due to the phonemic status of pitch and duration variations over syllables or morphemes in Mandarin. Besides the supra-segmental lexical tones, Mandarin phonology is also well-known for its varieties of aspirated vs. unaspirated consonants. The lexical tone uses the spectral cue of fundamental frequency (F0) to differentiate lexical meanings. For example, the Mandarin syllable *ba*/pa/ (*Pinyin*, an alphabetic phonological coding system used in Mainland China, and the corresponding International Phonetic Alphabet enclosed by backslashes) with high-level pitch (Tone 1) means "eight", and the same syllable means "to pull" when it is pronounced with high-rising pitch contour (Tone 2). Furthermore, the distinctive feature of unaspirated vs. aspirated contrast was acoustically realized as the temporal cue of voice onset time (VOT, defined as the time interval between the beginning of release burst and the onset of glottal pulsing) in Mandarin stops (Abramson & Whalen, 2017; Lisker & Abramson, 1964). For instance, the aspirated Mandarin stop *p*/p^h/ carries a much longer VOT compared to the corresponding unaspirated *b*/p/. A series of behavioral and neurophysiological studies have proved that Mandarin-speaking TD adults and children perceived both the linguistic pitch (lexical tone: Chen et al., 2017; Wang, 1976; Xi et al., 2010; Xu et al., 2006) and linguistic time (VOT; Cheung et al., 2009; Feng, 2018; Xi et al., 2009) in a highly categorical perception (CP) mode, with a sharp identification boundary and enhanced sensitivity to between-category contrasts relative to within-category ones (Liberman et al., 1957; Massaro, 1987). Moreover, developmental studies showed that Mandarin-speaking TD children from six-year-olds started to show an adult-like competence in the CP of lexical tones (Chen et al., 2017; Xi et al., 2009), and 10-year-old children generally reached an adult-like CP of VOT (Feng, 2018). Two studies investigated the CP of lexical tones in Mandarin-speaking children with ASD (Chen et al., 2016; Wang et al., 2017), but none of the previous studies have simultaneously explored the processing of lexical tones and VOT in ASD within one single study. To fill the research gap, in the current study, we compared the competence of CP of lexical tones and VOT in Mandarin-speaking individuals with ASD. This study aimed to investigate how different auditory weighting systems towards spectral and temporal cues in ASD influenced their linguistic processing of native phonological categories of lexical tones (Tone 1 vs. Tone 2) and VOT (*b*/p/ vs. *p*/p^h/) as indexed by CP measures. Answers to this question would deepen our understanding of the influence of lower-level acoustic processing on the higher-level phonological processing during speech perception with evidence from clinical populations.

Furthermore, previous studies on the development of speech perception in TD infants showed a language-dependent perceptual reorganization between 6 and 12 months, by perceptual narrowing for the native phonological contrasts and by perceptually ‘tuning out’ irrelevant acoustic information (Best & McRoberts, 2003; Kuhl, 2000, 2004). Such Perceptual Magnet Effect (Kuhl, 1991; Kuhl et al., 2008) around prototypes laid a foundation for the CP mode of native phonological categories in TD children. However, individuals with ASD tended to exhibit a lack of early interest in social interaction and a reduced preference to speech sounds (Constantino et al., 2004, 2007; Dawson et al., 1998; Klin, 1991; Kuhl et al., 2005), as well as atypical processing bias towards local details and low-level acoustic features (Frith, 1989; Happé & Frith, 2006; Mottron & Burack, 2001). Thus, they might show difficulties in extracting relevant invariant phonetic features and forming proper phonological categories, as reflected by a reduced degree of CP of speech sounds in ASD.

In the existing literature, one mismatch negativity (MMN) study (Wang et al., 2017) revealed a deficit in CP of lexical tones in Mandarin-speaking children with autism (mean age = 10.4 years, age range = 9–13), with similar mismatch response (MMR) amplitudes elicited from between-category tonal deviants relative to within-category tonal deviants. Another behavioral study (Chen et al., 2016) also suggested that the low-verbal and low-functioning children with ASD (age range = 6–8, developmental age = 3.7) exhibited no enhanced discrimination accuracy for the between-category tonal pairs (Tone 1 vs. Tone 2), and showed a much wider perceptual width around the boundary relative to age-matched TD children, both pointing to an impaired CP pattern of lexical tones. Also, a huge within-group heterogeneity in the degree of CP was observed based on the behavioral performance (Chen et al., 2016). These corroborating findings seemed to imply an impaired CP of lexical tones in tone language speakers with ASD. However, the notion of an impaired CP pattern (i.e., continuous/non-categorical perception pattern) among all autistic individuals should be interpreted with caution. Firstly, the MMR component (Wang et al., 2017) is elicited without behavioral requirements and in the absence of focal attention, thus it is highly likely that the attention of ASD perceivers per se might exert an influence on phonological processing of speech sounds (Whitehouse & Bishop, 2008). Secondly, the autistic spectrum showed huge variability in terms of speech and language development. The high-functioning ASD with better language and cognitive capacity may not necessarily show an impaired CP pattern. Thirdly, even TD young children often displayed a less precise categorization of speech sounds compared to healthy adults (Hoonhorst et al., 2011; Medina et al., 2010), which was related to the less sufficient experience to native speech sounds compared to adults. Similarly, children with ASD might merely exhibit "weaker or delayed" category formation compared to age-matched TD controls (Soulières et al., 2007), but rather an "impaired" CP pattern throughout their lifespan. To test these hypotheses, by using a behavioral CP paradigm, we investigated the performance of CP of native speech sounds (lexical tones and VOT) in high-functioning adolescents with ASD without severe language/cognitive delays.

It remained unclear whether the CP of lexical tones was universally impaired among all the native individuals of the autistic spectrum regardless of age and language/cognitive capacity. If this were the case, it would imply the possibility of developing the CP index as one of the biomarkers for early diagnosis of ASD from speech perception. Alternatively, if the degree of CP was altered among different subgroups of ASD and different ages, it would be necessary to uncover the possible influential factors. The degree of CP of native phonemes in TD children increased with age, due to an accumulation of perceptual development from the tonal information of ambient sound input (Chen et al., 2017). Also, some sub-tests of phonological working memory, such as digit span and nonword repetition, were considered to contribute to the behavioral performance of speech perception (Millman & Mattys, 2017). In this study, we also aimed to investigate whether the high-functioning adolescents with ASD who had longer native language experience could perceive native speech sounds in a preserved CP manner, and whether the performance in CP of speech would be indexed by chronological age, language ability as well as phonological working memory in ASD.

In the research field of auditory and speech processing, several recent studies have reported a speech-specific pitch processing atypicality in tone-language-speaking children with ASD (Lau et al., 2020; Jiang et al., 2015; Wang et al., 2017; Yu et al., 2015). In terms of syllable-level pitch processing, Mandarin-speaking children with ASD showed an atypical or impaired processing of lexical tones (Wang et al., 2017; Yu et al., 2015), whereas they showed normal or even enhanced processing of the same pitch information in the nonspeech materials (pure tone or harmonic sound). Such domain specificity of pitch processing was inconsistent with the relevant findings based on non-tonal language speakers. For instance, one study explored the pitch discrimination of real word, nonword, and nonspeech stimuli in English-speaking autistic children (mean age = 126 months) without significant language impairment (Heaton et al., 2008). The perceptual results showed that English-speaking children with ASD were generally more proficient at discriminating pitch contours from both speech (real word and nonword) and nonspeech conditions relative to TD children, pointing to a domain-general account of pitch processing superiority in non-tonal language speakers with ASD (Haesen et al., 2011; Heaton et al., 2008; Järvinen-Pasley & Heaton, 2007). The discrepancy between different language backgrounds suggested that the speech-specific lexical tone processing difficulties in autism were likely to be related to the unique phonological role of lexical tones. Given that lexical tones are superimposed on the syllabic segments, the semantic status of the pitch carriers in speech (real word vs. nonword) might contribute to the performance in CP of lexical tones. Furthermore, as suggested by the Neural Complexity Hypothesis (Samson et al., 2006) in depicting auditory processing in ASD, which proposes that individuals with ASD have difficulties in processing spectrally and temporally complex auditory information. The speech stimuli are more complex regarding the spectro-temporal components, compared with different types of nonspeech stimuli (e.g., pure tones, harmonics, or filtered sounds) adopted in previous studies. Thus, in order to test whether the acoustic complexity also plays a role, it is necessary to introduce another type of nonspeech material in this study, such as iterated rippled noise (IRN) which is comparable to the speech materials in terms of the spectro-temporal complexity (see *Methods* for more details). All in all, this study tried to uncover the nature of speech-specific lexical tone processing difficulties in tone-language-speaking individuals with ASD, by comparing the CP of pitch contours embedded in various types of pitch carriers with varying levels of spectro-temporal complexity or phonemic/semantic relevance.

In a nutshell, we focus on two prominent phonological features in Mandarin Chinese, lexical tones and voice onset time (VOT), which utilize pitch and time changes respectively to convey phonological contrasts, aiming to address three main questions: (1) Whether Mandarin-speaking high-functioning adolescents with ASD could perceive two speech continua varying in lexical tone and VOT in a similar categorical manner as neuro-typical peers, (2) Whether the performance in the CP of speech was related to chronological age, language ability as well as phonological working memory in individuals with ASD, and (3) Especially, whether different types of speech and nonspeech pitch carriers with varying levels of spectro-temporal complexity or phonemic/semantic relevance could exert an impact on the perception of pitch contours.

2 Methods

2.1 Participants

We have initially recruited 22 high-functioning Mandarin-speaking adolescents with ASD and 20 age-matched TD controls to participate in this study. Assessed by the local administrant hospitals, all the participants with ASD had nonverbal IQ above 70 using the Raven's Standard Progressive Matrices Test (Raven & Court, 1998) and without moderate to severe language delays (i.e., capable of using full, complex sentences). They were screened for hearing loss using pure tone audiometry and met the criteria for normal hearing. The clinical diagnosis of ASD was established according to the DSM-5 criteria for ASD (American Psychiatric Association, 2013), and further confirmed using the Autism Diagnostic Observation Schedule-

2 (ADOS-2; Lord et al., 2012), or Gilliam Autism Rating Scale–Second Edition (GARS-2; Gilliam, 2006) by pediatricians and child psychiatrists with expertise in diagnosing ASD in local hospitals. Approval of the research was granted by the local institutional review board of the Hong Kong Polytechnic University, and a written consent form was obtained from each participant.

To avoid the inclusion of individuals who had problems in perceiving synthetic sounds, a minimum accuracy score of 80% in the identification of two ending stimuli (i.e., prototypical tonal/aspiration category) in each continuum was required for the analyses of CP data. All the 20 TD controls (males = 14) met the accuracy criterion for the identification accuracy of both pitch contours and VOT. Of the 22 adolescents with ASD, 20 (males = 16) and 15 (males = 11) subjects with ASD met the criterion for pitch contour and VOT condition, respectively. The overview of participant characteristics is presented in Table 1. By using the Package 'Partiallyoverlapping' in R for the comparison of two partially overlapping samples (Derrick et al., 2017), the two ASD subgroups did not differ from each other in terms of chronological age (t = -0.23, p = .821), language ability (t = -0.11, p = .916), forward digit span (t = -0.03, p = .974), as well as nonword repetition (t = -0.13, p = .899). Moreover, both ASD subgroups had similar chronological age and forward digit span as the TD controls (ps > .05), but significantly lagged behind the TD controls in terms of language ability (both ps < .01) as well as nonword repetition (both ps < .001).

[Insert Table 1 around here]

2.2 Stimuli

The pitch contours ranging from Mandarin Tone 2 (high-rising tone) to Tone 1 (high-level tone) were embedded in four types of sound materials: real word (speech), nonword (speech), IRN (nonspeech), and pure tone (nonspeech). The stimulus features are shown in Table 2 in terms of pitch contrast, spectro-temporal complexity, phonemic contrast, and semantic contrast among four types of pitch carriers. The Mandarin monosyllabic words *ba*/pa/ with the Tone 1 and Tone 2 were recorded by a native female speaker (44100 Hz sampling rate, 16-bit resolution). Based on the natural pitch templates with Tone 2 (stimulus #1, meaning “to pull”) and Tone 1 (stimulus #7, meaning “eight”), the seven stimuli along the lexical tone continuum (Figure 1a) were synthesized using TANDEM-STRAIGHT software (Kawahara et al., 2009). Then, the seven pitch tiers were extracted and superimposed on the other three types of pitch carriers, including nonword (Figure 1b), IRN (Figure 1c), and pure tone (Figure 1d) using the Pitch-Synchronous Overlap Add implanted in Praat (Boersma & Weenink, 2016). Specially, the nonword *bü*/py/ was chosen since it does not exist in Mandarin, but its constituent units of consonant /p/ and vowel /y/ belong to native phonemes for Mandarin speakers. That is to say, the nonsense syllable *bü*/py/ contained the phonemic contrast but not semantic contrast in Mandarin. Additionally, two types of nonspeech materials with different levels of spectro-temporal complexity were adopted. The nonspeech material of pure tone is acoustically much simpler than the speech carriers, while the other nonspeech of iterated rippled noise (IRN) using 64 iteration steps (Swaminathan et al., 2008) has a comparable level of spectro-temporal complexity as the speech sounds. All the stimuli were normalized to be 300 ms. To further match the loudness level, the intensity level of pure tone was set to 85 dB, 15 dB higher than that of the other three types of pitch carriers (70 dB).

[Insert Table 2 around here]

[Insert Figure 1 around here]

A schematic diagram of the seven stimuli along the VOT continuum is shown in Figure 6(a). The VOT continuum was synthesized with the following procedures. First, the monosyllabic word *pa*/p^{h}a55/ (with Tone 1, meaning “lying down”) was produced from the same native female speaker, which was used as the basis for manipulation. Then, the syllable *pa*/p^{h}a55/ was normalized to be 300 ms, and it was divided into three parts: the burst release (~5 ms), aspiration (~48 ms), and vowel /a55/ (~247ms). The burst release was referred to as the abrupt burst in the waveform caused by the sudden release of the oral closure when producing stops; the aspiration part contained the frication noise along with the expiratory airflow right after the release of closure and before the vowel portion. During manipulation, the burst release was kept constant, while the aspiration part and vowel part were shortened and lengthened respectively in seven steps

(Δ = 8 ms). Lastly, the three parts were concatenated in Praat, generating a continuum of seven equally distanced stimuli ranging from the unaspirated *ba*/pa55/ (stimulus #1, meaning “eight”) to the aspirated *pa*/p^ha55/ (stimulus #7, meaning “lying down”). All the seven stimuli along the VOT continuum were set to be 300 ms in duration, and 70 dB in mean intensity.

2.3 Tasks and Procedure

Two classical tests for the CP of speech sounds, the identification test and the discrimination test, were both conducted via E-Prime 2.0. The identification test was performed firstly for all the participants, and then the discrimination test. Before performing experimental CP tests, three additional tasks evaluating language ability, digit span, and nonword repetition were performed respectively for each subject.

Language ability: The overall language ability (Chen et al., 2017; Ning, 2013) was evaluated for each Mandarin-speaking participant, which consists of five subtests (including *Test of Mandarin Grammar*, *Word Definition Test*, *Rapid Automatized Naming*, *Narrative Test*, and *Sentence Comprehension Test*). These subtests evaluated both language comprehension and language expression, and aimed to assess different aspects of language abilities such as phonology, lexicons, grammar, and semantics. The administration time is around 30 minutes.

Digit span: In order to evaluate the short-term phonological working memory, we administered a digit span task, which included both the forward digit span and backward digit span. However, during data collection, some of the individuals with ASD could not fully understand and follow the instructions of backward digit span with a requirement on both storage information and manipulation by the executive control (Hamann, 2017). Consequently, only the results from the simpler task of forward digit span were reported to ensure the reliability of performance. In the forward digit span task, a series of numbers were played to participants auditorily and they were asked to repeat them immediately. For each digit length (two to nine digits), there were two separate items. The response for each item was regarded as correct and awarded 0.5 points only when the participants could correctly repeat every digit in the right order. The full score for the test of forward digit span is 8.

Nonword repetition: The nonword repetition task was comprised of 60 items divided equally into 20 monosyllabic (e.g., rai4), 20 disyllabic (e.g., bong1nua2), and 20 trisyllabic (e.g., sua3piong4buai1) nonwords, which is compatible with the number of syllables in previous studies (Gathercole et al., 1991). Some of the syllables carried a nasal coda, and each syllable carried diphthongs or triphthongs together with one of the four Mandarin lexical tones, in an effort to increase the task demand and phonological complexity. All the onsets, finals, and lexical tones existed in Mandarin phonology, but the combined syllables within each nonword did not exist in Mandarin. The items, prerecorded by a female native speaker of Mandarin, were played back to participants using E-Prime 2.0 (Psychology Software Tools Inc., USA) on a Windows-based laptop. Monosyllabic, disyllabic, and trisyllabic nonwords were presented in three separate blocks with a pause between the blocks; the items within each block were randomized. Three practice trials were presented within each block before the experimental trials to familiarize subjects with the task. All the TD and ASD participants were asked to repeat the nonwords as accurately as possible. The repeated productions from each subject were recorded, and the recordings were transcribed and scored by a native phonetician afterward. For each item, the average percentage of phonemes (including consonant, vowel, and lexical tone) correctly repeated per nonword was calculated.

Experimental CP task 1: Identification test. First, in the identification test, participants were asked to perform a two-alternative forced-choice (2AFC) paradigm. In the identification training of pitch contours among the four continua, participants were trained to point to one picture depicting a car driving on a level road when hearing a ‘level tone (Tone 1)’, and point to the other picture depicting a car driving on a rising road when hearing a ‘rising tone (Tone 2)’. In the identification training of the VOT continuum, participants were trained to point to one picture with a blue circle when hearing the sound of ‘*ba*/pa55/’, and point to the other picture with a blue square when hearing ‘*pa*/p^ha55/’. After participants have acquired the matching between the sounds and their corresponding pictures, a practice block was offered before formal testing. In the practice block, the two ending stimuli of each continuum (#1 and #7) were repeated four times, and minimum accuracy of 80% was required before moving on to the formal block. During the practice blocks, the feedback was offered to the participant, but not in the formal blocks. In the formal block, each stimulus was repeated five times and played randomly. There were totally five different blocks of five continua, including four continua for the pitch condition and one continuum for the VOT condition. Subjects were asked to identify 175 sounds in total (7 stimuli × 5 repetitions × 5 continua) among five formal blocks. The four identification blocks containing four types of pitch contours were randomly presented, and the order of pitch and VOT conditions was also presented in a random order among participants. The participants' responses were logged by the experimenter via pressing the corresponding keys on the keyboard. Both TD and ASD adolescents were free to have a rest whenever they wanted. The whole identification test lasted around 25 minutes for each participant.

Experimental CP task 2: Discrimination test. Then, for the discrimination test, the AX paradigm was adopted to instruct subjects to discriminate the two sounds of each pitch/VOT pair as the 'same' or 'different'. Also, during the training stage, participants were trained to point to one picture (a happy face with two identical eyes) representing the same pairs, and point to the other picture (a sad face with two different eyes) representing the different pairs. Each practice block contained four pairs along the pitch/VOT continuum (i.e., 1-1, 7-7, 1-7, and 7-1), with each practice pair repeating twice. The minimum discrimination accuracy of 80% was required before moving to the formal block. Feedback was provided to the participants in the practice blocks, but not in the formal blocks. A modified 2-step discrimination test was adopted with no overlapping discrimination units (1-3, 3-5, and 5-7), which were used in some of the ERP studies (e.g., Chen & Peng, 2020). There were 10 testing pairs for each pitch/VOT continuum in the 2-step discrimination task, including six pairs (different pairs) consisting of two different stimuli separated by 2 steps in either forward (1-3, 3-5, 5-7) or reverse order (3-1, 5-3, 7-5), as well as four pairs (same pairs) each paired with itself (1-1, 3-3, 5-5, 7-7). Each pair was repeated four times randomly within one formal block, with a 500 ms inter-stimulus interval. There were 200 pairs (10 pairs × 4 repetitions × 5 continua) distributed among five formal blocks (four pitch blocks and one VOT block). The four blocks of pitch discrimination were randomly presented, and the order of pitch and VOT conditions was presented in a random order as well. All participants were free to have a rest whenever they wanted, and the whole discrimination test lasted around 40 minutes for each subject.

2.4 Scoring and Data Analyses

For both identification and discrimination tests, only the data in formal blocks were involved in further analyses. First, the identification curve was analyzed in term of two key parameters using probit analyses (Finney, 1971): boundary position, which is defined as the corresponding 50% crossover point in a continuum, and boundary width, defined as the linear distance along the stimulus step between the 25th and 75th percentiles as determined by the mean and standard deviation obtained from probit analysis (Hallé, Chang, & Best, 2004). We have replaced 0% with 0.1%, and 100% with 99.9% for individual identification curves at both ends, in order to fit the asymptotic property of probit function (Peng et al., 2010). The boundary position refers to the identification midpoint dividing the two tonal/aspiration categories, and the boundary width indicates the steepness of the response shift around the categorical boundary. Importantly, the boundary width was used to measure the degree of CP in the identification test (Chen et al., 2017). The narrower the boundary width, the steeper the boundary shift, and vice versa.

Second, the discrimination pairs were divided into three comparison units (units 1-3, 3-5, and 5-7), each containing four types of discrimination pairs: the same pairs (AA and BB) and different pairs (AB and BA). Adjacent comparison units contained overlapping AA or BB pairs (Peng et al., 2010; Xu et al., 2006). Then, the discrimination accuracy (%) was transferred into the sensitivity index d' for each comparison unit (Macmillan & Creelman, 2005), which takes response bias into consideration. Specifically, for each comparison unit, the d-prime (d') score was computed as the difference between standard normal deviate (*z*-score) of hit rate (“different” responses to different pairs: AB and BA) and that of false alarm rate (“different” responses to the same pairs: AA and BB). In reference to the boundary position, the comparison units were further classified as between-category type and within-category type for each subject. For instance, if one participant showed a boundary position of 3.94, then for this subject, the between-category sensitivity was referred to the *d′* of the comparison unit of 3-5, while the within-category sensitivity was calculated as the averaged *d′* for the comparison units of 1-3 and 5-7. Finally, the *d′* score of between-category type minus that of within-category type was referred to as the "peakedness score" (Jiang et al., 2012), which represents the benefit magnitude in the discrimination test.

Statistical analyses were conducted using linear mixed-effect models (LMMs) in R (R Core Team, 2014), by using the package of lme4 (Bates et al., 2014) to create the LMMs. Data points with standardized residuals over 2.5 standard deviations (*SD*) were removed due to the violation of normal distribution. For the condition of pitch perception, the models were built with *group* (ASD vs. TD), *pitch carrier* (real word, nonword, IRN, and pure tone), and their two-way interaction acting as fixed factors to analyze the boundary width and boundary position for identification analysis, as well as the peakedness score for discrimination analysis. Another LMM was constructed for the pitch discrimination performance using *category type* (within-category vs. between-category), *group* (ASD vs. TD), *pitch carrier* (real word, nonword, IRN, and pure tone), and all possible interactions acting as fixed factors. Furthermore, for each LMM in the VOT condition, the boundary width/boundary position/peakedness score was entered as the dependent measure, with *group* (ASD vs. TD) acting as the fixed effect. In addition, to compare the within- vs. between-category sensitivities to VOT changes, the LMM was built with *category type* (within-category vs. between-category), *group* (ASD vs. TD), and their interaction acting as fixed factors. When fitting all the LMMs in the analyses of identification and discrimination data, the factors of *language ability*, *digit span*, and *nonword repetition* were regarded as controlled covariates, which were centered to reduce multicollinearity; *participant* was included as a random effect. By-participant random intercepts and slopes for all possible fixed factors were included in the initial model (Barr et al., 2013), which was compared with a simplified model that excluded a specific fixed factor using the ANOVA function in lmerTest package (Kuznetsova et al., 2017). Post-hoc pairwise comparisons were performed using the lsmeans package (Lenth, 2016) with Tukey adjustment.

Furthermore, linear regression models were constructed in R to examine the potential variables contributing to the ASD participants' CP performance. The approach of linear regression models is considered superior to traditional methods of correlation analyses such as Pearson/Spearman's correlation (Koerner & Zhang, 2017), since the linear regression models consider the mutual influence of different predictors. Hypothesized predictors for the CP of pitch/VOT included chronological age, language ability, forward digit span, and nonword repetition. Evaluation of the degree of CP in the current report included (1) boundary width across four types of pitch carriers, (2) peakedness score across four types of pitch carriers, (3) boundary width in the VOT condition, (4) peakedness score in the VOT condition. Separate models were created for each estimate of CP performance, with all the four predictors added as fixed effects. Parameter estimates, standard errors, *t* values, and *p* values for the fixed effects were assessed and reported.

3 Results

3.1 Categorical Perception of Pitch Contours

3.1.1 Identification Result

Figure 2(a) shows the overall identification curves, and Figure 2(b) displays the boundary positions for the ASD and the TD groups among four types of pitch carriers. The LMM on boundary position showed a significant two-way interaction of *group* × *pitch carrier* [χ^2 (3) = 13.07, $p < .01$], which was further analyzed under different types of pitch carriers respectively. In the real word condition, compared to the TD group ($M = 4.42$, $SD = 0.43$), the ASD group ($M = 4.97$, $SD = 0.41$) exhibited a much larger boundary position which was closer to the level end (β = 0.52, SE = 0.19, $t = 2.73$, $p < .01$). However, as shown in Figure 2(b), ASD group and TD group showed a similar boundary position in nonword condition (β = 0.25, SE = 0.19, $t = 1.29$, $p = .202$), in IRN condition (β = 0.04, SE = 0.19, $t = 0.20$, $p = .839$), as well as in pure tone condition (β = -0.01, SE = 0.19, $t = -0.06$, $p = .954$).

[Insert Figure 2 around here]

Moreover, the obtained boundary widths for the ASD and the TD groups among four types of pitch carriers are shown in Figure 3. The mean boundary widths (*SD*) for adolescents with ASD and TD controls were 0.84 (*0.55*) and 0.63 (*0.33*) respectively in the real word condition, 0.87 (*0.68*) and 0.78 (*0.40*) in nonword condition, 0.84 (*0.43*) and 0.66 (*0.20*) in IRN condition, 0.93 (*0.57*) and 0.71 (*0.29*) in pure tone condition. The LMM on boundary width revealed neither significant main effects of *group* [χ^2 (1) = 0.97, p = .325], *pitch carrier* [χ^2 (3) = 2.01, p = .569], nor a significant interaction of *group* × *pitch carrier* [χ^2 (3) = 0.82, p = .845]. The obtained boundary width in the identification of pitch contours did not differ between ASD and TD groups, and among different types of pitch carriers (Figure 3).

[Insert Figure 3 around here]

3.1.2 Discrimination Result

The d' values of the between-category and within-category comparison units in ASD and TD groups are displayed in Figure 4 across different pitch carriers. Statistical analysis revealed a significant three-way interaction of *group* × *category type* × *pitch carrier* [χ^2 (3) = 8.47, $p < .05$], which was further analyzed under different pitch carriers respectively. First, in the real word condition, LMM on d' values revealed significant main effects of *category type* [χ^2 (1) = 19.16, $p < .001$], and *group* [χ^2 (1) = 5.28, $p < .05$], while the interaction of *group* × *category type* did not reach significance [χ^2 (1) = 0.31, p = .575]. Second, in the nonword condition, there were significant main effects of *category type* [χ^2 (1) = 19.81, $p < .001$], and *group* [χ^2 (1) = 6.16, $p < .05$] on the d' values, while there was no significant interaction effect of *group* × *category type* [χ^2 (1) = 2.32, p = .127]. Third, the LMM on d' values in the IRN condition showed significant main effect of *category type* [χ^2 (1) = 23.46, $p < .001$], while the interaction effect of *group* × *category type* was not significant [χ^2 (1) = 0.06, p = .813]. These findings above indicated that both ASD and TD groups showed much higher d' values in response to between-category unit compared to the within-category unit across pitch carriers of real word, nonword, and IRN. Moreover, as shown in Figure 4, relative to TD controls, the ASD group showed relatively smaller *d′* values in discriminating both within-category and between-category comparison units in pitch carriers of real word and nonword. Fourth, in the pure tone condition, LMM on *d′* values exhibited significant main effect of *category type* [χ^2 (1) = 35.30, $p < .001$], as well as interaction effect of *group* × *category type* [χ^2 (1) = 10.47, $p = .001$]. Post-hoc pairwise comparisons showed that relative to TD group, the ASD group had a lower *d′* in response to between-category unit (β = -0.89, SE = 0.33, $t = -2.70$, $p < .01$), while the two groups had similar *d′* values in response to within-category unit (β = 0.21, SE = 0.33, $t = 0.62$, $p = .535$). Moreover, the between-category unit generated a much higher *d′* value than the within-category type in pure tone condition (Figure 4), for both ASD group (β = 0.64, SE = 0.23, $t = 2.80$, $p < .01$), and TD group (β = 1.74, SE = 0.23, $t = 7.57$, $p < .001$).

[Insert Figure 4 around here]

The peakedness scores (the *d′* score of between-category unit minus that of within-category unit) in TD and ASD groups are shown in Figure 5 across different pitch carriers. The mean peakedness scores (*SD*) for the ASD group and TD group were 1.26 (*1.50*) and 1.00 (*1.59*) respectively in the real word condition, 0.75 (*1.54*) and 1.40 (*1.19*) in nonword condition, 1.35 (*2.00*) and 1.48 (*1.76*) in IRN condition, 0.67 (*1.09*) and 1.74 (*1.00*) in pure tone condition. For the LMM on peakedness score in the discrimination of pitch contours, neither main effects of *group* [χ^2 (1) = 0.88, $p = .348$], *pitch carrier* [χ^2 (3) = 1.47, $p = .688$], nor interaction effect of *group* × *pitch carrier* [χ^2 (3) = 5.63, $p = .131$] reached significance. The peakedness scores were comparable between ASD and TD groups, and among different types of pitch carriers (Figure 5).

[Insert Figure 5 around here]

3.1.3 Linear Regression Result

Table 3 shows the regression coefficients indicating the relationships between predictors (chronological age, language ability, digit span, nonword repetition) and the degree of CP of pitch contours in individuals with ASD. In the real word condition, *language ability* (β = -0.03, SE = 0.01, $t = -3.01$, *p*

< .01) and *digit span* (β = -0.11, SE = 0.05, t = -2.17, p < .05) were significant predictors for the boundary width (Table 3). The negative regression coefficients (β) indicated that the better language ability or digit span in ASD led to a narrower boundary width (i.e., a steeper identification slope) in the identification of lexical tones in real word. In the nonword condition, only nonword repetition in ASD was significantly associated with the boundary width in the identification of pitch contours embedded in nonword material (β = -5.30, SE = 1.66, t = -3.20, p < .01), with higher accuracy of nonword repetition contributing to a steeper slope. In the nonspeech pitch carriers of IRN and pure tone, however, none of the significant correlations between predictors and the degree of CP were detected as shown in Table 3.

[Insert Table 3 around here]

3.2 Categorical Perception of VOT

Figure 6(b) shows the identification curves of VOT perception in ASD and TD groups. The analysis on boundary position of VOT continuum showed that two groups had a similar boundary position (β = 0.33, SE = 0.38, t = 0.89, p = .383). However, compared to ASD group, the TD controls showed a much narrower boundary width in the identification of VOT as shown in Figure 6c (β = -0.88, SE = 0.30, t = -2.92, p < .01), as well as a higher peakedness score in the discrimination of VOT as shown in Figure 6e (β = 1.65, SE = 0.62, t = 2.68, p < .05). Furthermore, the LMM on d' values of different category types in VOT condition exhibited a significant interaction effect of *group* × *category type* [χ^2 (1) = 8.60, p < .01]. Post-hoc analyses demonstrated that both ASD group (β = 1.23, SE = 0.43, t = 2.89, p < .01) and TD group (β = 2.88, SE = 0.37, t = 7.82, p < .001) showed a higher between-category d' value relative to the within-category one (Figure 6d). In addition, the ASD participants showed a much lower d' in response to between-category unit (β = -2.07, SE = 0.40, t = -5.20, p < .001), while the two groups had similar d' values in response to within-category unit during VOT perception (β = -0.42, SE = 0.40, t = -1.05, p = .297).

[Insert Figure 6 around here]

The regression coefficients (estimates, standard errors, *t* values, and *p* values) are presented in Table 4, indicating the relationships between predictors and the degree of CP of VOT among ASD participants. None of the four predictors in ASD participants revealed a significant relationship with the boundary width in the VOT identification test (all *p*s > .05). Similarly, there were no significant correlations between the ASD participants' performance on chronological age/language ability/digit span/nonword repetition and the peakedness score in the VOT discrimination test (all *p*s > .05).

[Insert Table 4 around here]

4 Discussion

4.1 Preserved CP Pattern in High-functioning Adolescents with ASD and its Influential Factors

By using a behavioral CP paradigm, we investigated the performance of CP of lexical tones and VOT in high-functioning adolescents with ASD without severe language/cognitive delays. As seen in Figure 4 and Figure 6(d), the high-functioning adolescents with ASD in the current study did perceive the lexical tones and VOT in a preserved CP pattern, as indicated by a much higher *d′* for between-category pairs than that for within-category pairs in both types of continua. Furthermore, the preserved CP pattern of cross-boundary benefit in the speech context was transferred to nonspeech counterparts for both ASD and TD groups (Figure 4), reflecting a carry-over influence of long-term phonological processing from the speech to the nonspeech domain. The preserved CP pattern in ASD was also detected in the perception of other types of speech sounds, such as the CP of vowels (/i-y/ continuum) and consonants by place of articulation (/d-b/ continuum) in both high-functioning children with autism and Asperger syndrome (You et al., 2017), as well as the CP of VOT (/g-k/ continuum) in high-functioning and cognitively able adults with ASD (Stewart et al., 2018). Thus, following the above evidence from the perception of various types of speech sounds, we would be confident to infer that the impaired CP pattern might not apply to all the autistic individuals, but rather tend to be part of a shared vulnerability of language or cognitive delay/impairment in a subgroup of ASD.

Thus, the autistic individuals showed large variabilities regarding the degree of CP of speech sounds, with some individuals showing a profoundly impaired CP pattern (Chen et al., 2016; Wang et al., 2017) while others exhibiting a preserved CP pattern albeit with varying levels of competence (this study; Stewart et al., 2018; You et al., 2017). In the current study, we further investigated whether and how chronological age, language ability, and phonological working memory (digit span and nonword repetition) were related to the level of CP competence among individuals with ASD. As shown in Table 3, the overall language ability in ASD was a significant predictor for the boundary width ($p < .01$) in real word condition. The autistic participants with better language ability tended to elicit a narrower boundary width (i.e., a steeper slope) in the identification of lexical tones. The higher degree of CP of lexical tones correlated with better language ability in Mandarin-speaking high-functioning adolescents with ASD. Such correlation was observed as well in the low-functioning younger autistic children (Chen et al., 2016). Furthermore, the degree of CP was correlated with the verbal ability of reading, lexical decision, and verbal IQ in adults with ASD (Stewart et al., 2018). Collectively, the close relationship between phonological processing capacities and language functions was consistently observed in individuals with ASD of various cognitive abilities and different age ranges (Bishop et al., 2004; Chen et al., 2016; Constantino et al., 2007; Stewart et al., 2018). The implication is that some aspects of language difficulties found in individuals with ASD may be related to the reduced CP competence of speech sounds. Our current findings suggested the necessity of further examining the potential links among social competence, speech processing, and language functioning among autistic individuals in future prospective longitudinal work. If the CP competence of native speech turned out to be a reliable predictor of certain language-related abilities in ASD, it would call for an inclusion of CP-related testing and training in the evaluation and intervention of ASD at an early stage.

Furthermore, the regression analyses showed that the capacity of digit span in ASD could be a contributing factor for the identification acuity of lexical tones; the nonword repetition in the autism group was a significant predictor for the identification acuity of pitch contours in nonword condition (Table 3). There findings were not surprising given that in the behavioral CP tests, three forms of memory—sensory memory and the short- and long-term forms of categorical memory—are involved (Xu et al., 2006). Besides, the discrimination task in the AX pattern required the recruitment of short-term working memory to store one stimulus and then to compare it with the subsequent one (Mitterer & Mattys, 2017). Both digit span and nonword repetition were used to evaluate short-term phonological working memory (Hamann, 2017; Rispens & Baker, 2012); the nonword repetition task further draws on sub-lexical knowledge to access and maintain new phonological codes, which is thought to measure the representations of "chunks' of phonemes (Shao et al., 2020; Szewczyk et al., 2018). The close correlation between digit span/nonword repetition and the degree of CP found in this study called for the controlling of such confounding factors of the cognitive capacities such as phonological working memory of ASD (Boets et al., 2015). Contrary to our prediction, this study failed to reveal a relationship between chronological age and CP competence across all the stimulus conditions. However, the lack of age effect must be interpreted with caution as it may be attributed to the relatively matured perceptual development in our samples of high-functioning adolescents with ASD, and the lack of power.

4.2 The Influence of Low-level Acoustic Processing on the High-level Phonological Processing

In the research field of general auditory processing, a plethora of studies have implied that individuals with ASD were reported to show atypical and unbalanced auditory processing depending on the acoustic dimensions (spectral vs. temporal) (Alcántara et al., 2012; Groen et al., 2009). As CP of speech sounds reflects the higher-level phonological processing mode, by comparing the CP competence of linguistic pitch (lexical tone) and linguistic time (VOT) in native speakers with ASD at the same time, it can help uncover whether and how lower-level acoustic processing could influence higher-level phonological processing.

Based on the current findings, although both the perception of lexical tones and VOT showed a typical CP pattern in Mandarin-speaking adolescents with ASD, the degree of CP varied greatly. As for the behavioral indexes of CP competence (Chen et al., 2019), the boundary width was used to measure the degree of CP in the identification function, with a narrower boundary width indicating a steeper slope, and vice versa; the peakedness score reflected the magnitude of cross-boundary benefit in a discrimination test. During the perception of the lexical tone continuum, the ASD participants showed similar boundary width (Figure 3) and peakedness score (Figure 5) relative to the neuro-typical peers. In stark contrast, during the perception of the VOT continuum, the ASD group showed a much wider boundary width (Figure 6c) and lower peakedness score (Figure 6e). Taken together, for the high-functioning Mandarin-speaking adolescents with ASD, the CP of native lexical tones was largely intact, meanwhile the degree of CP of VOT was greatly reduced. These findings suggest that the unbalanced acoustic processing capacities for pitch and time can be generalized to higher-level linguistic processing from the evidence in ASD. There is a concern that the inferior performance on the CP of VOT in ASD could also be attributed to the relative difficulty levels since the aspirated vs. unaspirated contrast tended to be acquired later relative to lexical tones in Mandarin-speaking TD children (Hua & Dodd, 2000). However, even for the high-functioning adults with ASD, they also showed a less categorical fashion in the perception of the VOT continuum when compared with IQ-matched typically developed adults (Stewart et al., 2018). Thus, the auditory processing difficulties of sound duration in autism are manifested profoundly and further persist into the higher-level phonological processing that involves the basic CP competence of VOT, and the processing of vowel length contrast phonemically to mark semantic distinction such as in Finnish-speaking (Lepistö et al., 2005) and Japanese-speaking (Kasai et al., 2005) individuals with ASD. To conclude, the current findings provide direct evidence that lower-level acoustics underlie higher-level phonological processing in speech perception, since the unbalanced acoustic processing skill (pitch vs. time) in ASD extends to the CP of speech sounds (lexical tones vs. VOT) in native perceivers from the clinical population.

4.3 Lexical Tone Perception Difficulties in ASD and the Underlying Mechanisms

For tone language speakers with ASD, several studies (Chen et al., 2016; Cheng et al., 2017; Lau et al., 2020; Wang et al., 2017; Wu et al., 2020; Yu et al., 2015) pointed to the native lexical tone perception difficulties at both behavioral and neural levels. Yet, our full understanding of lexical tone perception difficulties and their underlying mechanisms in tone language speakers with ASD are still far from complete.

Some scholars have proposed a speech-specific mechanism to explain the pitch perception difficulties only in the speech context (Wang et al., 2017; Yu et al., 2015). Others tried to explain the deficits with the ‘allophonic perception’ theory for autism (Huang et al., 2018; M. O’Riordan & Passetti, 2006; You et al., 2017), due to the detail-oriented processing style and enhanced acoustic pitch discrimination skills in autism. By using a fine-grained CP approach, this study investigated the identification, as well as within-category and between-category discrimination of pitch contours embedded in various types of speech and nonspeech contexts. The four different types of pitch carriers (real word, nonword, IRN, pure tone) differ in the levels of spectro-temporal complexity or phonemic/semantic relevance.

The degree of CP, as assessed by both boundary width (Figure 3) and peakedness score (Figure 5), did not differ between ASD and TD groups among all the four types of pitch carriers, indicating the well-developed CP of lexical tones in high-functioning adolescents with ASD, and its carry-over influence of long-term phonological processing from the speech to nonspeech domain regardless of the word status and spectro-temporal complexity. Interestingly, the boundary position differed between ASD and TD groups only in the real word condition with semantic information. More specifically, as shown in Figure 2, Mandarin-speaking participants with ASD showed a much higher boundary position (i.e., closer to the level end) relative to TD controls in the real word condition, with a similar pattern called “psychophysical boundary” observed in the non-tonal language speakers who had no tonal language experience (Wang, 1976). In other words, the relative perceptual space for the level tone (Tone 1) in Mandarin-speaking individuals with ASD was compressed compared to the TD controls, with ASD participants displaying less tolerance for the ambiguous rising contours to be judged as Mandarin level tone. Compared to the other three types of pitch carriers, the real word condition additionally carried the semantic contrast with the level-ending stimuli (#7) meaning “eight” and the rising-ending stimuli (#1) meaning “to pull”. As suggested by the ‘Ganong effect’ (Ganong, 1980; Stewart & Ota, 2008), which proposed that the boundary of phonetic categorization shifted as a function of lexical-semantic influence from real words, the enlarged perceptual space for the high-level Tone 1 in TD group might be attributed to the stronger influence from semantic effect in the high-frequency numeric word "eight". Therefore, it was possible that the autistic individuals were less susceptible to higher-level semantic capture when performing a pitch identification task, in line with the 'Weak Central Coherence' theory (Frith, 1989; Happé & Frith, 2006). These findings implied that one of the potential reasons responsible for the speech-specific lexical tone perception difficulties in ASD might be caused by a weaker feedback loop from the lexicon to phonemic activation (McClelland & Elman, 1986). This hypothesis should be tested in future studies with the paradigm of 'Ganong effect' (Ganong, 1980; Stewart & Ota, 2008), which directly investigates whether the extent to which tone categorization biases the judgment toward a known word is weakened in the ASD group relative to neurotypicals.

In the discrimination test, the sensitivity to within-category pitch discrimination was not elevated for the ASD group compared to the TD group in the speech conditions, which seemed to contradict the 'allophonic perception' theory for autism (Huang et al., 2018; M. O'Riordan & Passetti, 2006; You et al., 2017). But this phenomenon should be explained with caution. On the one hand, the within-category discrimination does not merely reflect the acoustic pitch processing for the native speakers, since the "dulled" within-category sensitivity was gradually formed with native language experience by perceptually 'tuning out' irrelevant acoustic information. Following this line, the high-functioning adolescents with ASD in this study who had an intact CP pattern might not show the 'allophonic perception' feature, which was corroborated with the findings in the high-functioning adults with ASD who did not show more accurate within-category discrimination in comparison with TD adults (Stewart et al., 2018). We would speculate the 'allophonic perception' pattern to emerge in low-functioning young children with ASD who lack inhibitory mechanisms for suppressing the detection of irrelevant within-category pitch differences, and thus to cause an impaired CP pattern. On the other hand, the behavioral AX discrimination task taps into attentional, and working memory processes, and is not solely assessing discrimination, which has been noted to be unsuitable for the ASD population since a large proportion of them were accompanied with attention and working memory deficits (Heaton et al., 2008). This might be one of the reasons to explain why the autistic subjects performed inferiorly compared to controls across the board regardless of within- and between-category pitch discriminations in the conditions of real word, nonword, as well as IRN. Although with such profound attentional and memory disadvantage, the ASD group nevertheless showed comparable d' values (even higher values but not statistically significant) in response to within-category pairs relative to TD controls in the pitch carrier of pure tone. Compared with other carriers of speech sounds and the nonspeech IRN, the pure tone is much simpler in terms of spectro-temporal complexity (Figure 1). Our current observations on pure tone were consistent with the previous behavioral and MMN findings (Bonnel et al., 2003; Ferri et al., 2003; Gomot et al., 2002; M. O'Riordan & Passetti, 2006), which highly supports the Neural Complexity Hypothesis (Samson et al., 2006). That is, the individuals with autism may display enhancement in pitch discrimination where spectro-temporally simple but not complex stimuli yield superior performances.

4.4 Limitations and Future Directions

This study has several limitations. First, it is important to note that the current conclusions were limited to the high-functioning adolescents with ASD, but not necessarily extended to younger children or low-functioning individuals with ASD. Given the huge heterogeneity within the autistic spectrum, in order to obtain a more robust statistical power, a larger sample size with a broader range of demographic characteristics is needed. Second, one of the big challenges for performing behavioral studies in autism is the reliability and validity of the data, given the serious attention deficits in most individuals with ASD. The current study adopted a modified discrimination test with no overlapping discrimination units, and future studies should try to include all the possible discrimination pairs to increase statistical power. Besides, before formal CP tests we have performed the training stage and practice blocks to familiarize the subjects with the experimental procedures, and a minimum accuracy of 80% was required before moving to the formal blocks. However, in the formal testing blocks, no control items were included to monitor the performance of the participants in this study. For instance, in the discrimination task, only the two-step discrimination pairs were incorporated as the testing stimuli, which was hard for us to judge the reliability of participants' responses. In future studies, we could also involve the easily discriminable pairs (such as the discrimination of the two ending stimuli) with a much larger acoustic distance in the testing blocks in an effort to monitor the behavioral responses during the testing stage. Furthermore, only behavioral research of one representative lexical tone or VOT pair was adopted in the current study. Further investigations with more pairs as well as electrophysiological approaches are warranted to comprehensively uncover the neural mechanisms of the CP of lexical tones and VOT in tone language speakers with ASD, and their correlations with behavioral measurements. Finally, future longitudinal research with younger children with ASD is necessary to chart and compare the developmental trajectory of CP of lexical tones and VOT, to further investigate the age effect on the perception of linguistic pitch and linguistic time in ASD.

5 Conclusions

Despite large individual variability, findings of the current study revealed a preserved CP pattern when perceiving the native lexical tones and VOT in Mandarin-speaking high-functioning adolescents with ASD, with a much higher sensitivity to the between-category pairs compared to the within-category pairs. The degree of CP of lexical tones in ASD was similar to TD controls, whereas the degree of CP of VOT was greatly reduced, reflecting the influence from lower-level acoustic processing of pitch and time. The language ability, digit span, and nonword repetition of ASD participants were found to be significant predictors for the levels of CP competence in some speech conditions of pitch perception. Furthermore, individuals with ASD showed a "psychophysical boundary" similar to the non-tonal language speakers, potentially due to the reduced access to the semantic information of real word. These findings deepened our understanding of phonological processing of different speech elements in the subgroup of high-functioning ASD without severe language/cognitive delay.

6 Acknowledgments

We would like to thank Yan Feng and Yumeng Yang from The Hong Kong Polytechnic University for their assistance in implementing this study.

References

Abramson, A. S., & Whalen, D. H. (2017). Voice Onset Time (VOT) at 50: Theoretical and practical issues in measuring voicing distinctions. *Journal of Phonetics*, *63*, 75–86. https://doi.org/10.1016/j.wocn.2017.05.002

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

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders: DSM-5* (5th ed.). Arlington, VA : American Psychiatric Publishing.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255–278. https://doi.org/10.1016/j.jml.2012.11.001

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2014). Fitting Linear Mixed-Effects Models using lme4. *ArXiv:1406.5823 [Stat]*. http://arxiv.org/abs/1406.5823

Best, C. C., & McRoberts, G. W. (2003). Infant perception of non-native consonant contrasts that adults assimilate in different ways. *Language and Speech*, *46*(2–3), 183–216. https://doi.org/10.1177/00238309030460020701

Bishop, D. V. M., Maybery, M., Wong, D., Maley, A., Hill, W., & Hallmayer, J. (2004). Are phonological processing deficits part of the broad autism phenotype? *American Journal of Medical Genetics Part B: Neuropsychiatric Genetics*, *128B*(1), 54–60. https://doi.org/10.1002/ajmg.b.30039

Boersma, P., & Weenink, D. (2016). *Praat: Doing phonetics by computer (Version 6.0. 14) [Computer program]*. http://www.praat.org/

Boets, B., Verhoeven, J., Wouters, J., & Steyaert, J. (2015). Fragile spectral and temporal auditory processing in adolescents with autism spectrum disorder and early language delay. *Journal of Autism and Developmental Disorders*, *45*(6), 1845–1857. https://doi.org/10.1007/s10803-014-2341-1

Bonnel, A., Mottron, L., Peretz, I., Trudel, M., Gallun, E., & Bonnel, A.-M. (2003). Enhanced pitch sensitivity in individuals with autism: A signal detection analysis. *Journal of Cognitive Neuroscience*, *15*(2), 226–235. https://doi.org/10.1162/089892903321208169

Brodeur, D. A., Green, C. G., Flores, H., & Burack, J. A. (2014). Time estimation among low-functioning individuals with autism spectrum disorders: Evidence of poor sensitivity to variability of short durations. *Autism Research*, *7*(2), 237–244. https://doi.org/10.1002/aur.1364

Chen, F., & Peng, G. (2020). Reduced sensitivity to between-category information but preserved categorical perception of lexical tones in tone language speakers with congenital amusia. *Frontiers in Psychology*, 11:581410. https://doi.org/10.3389/fpsyg.2020.581410

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 2016*, 233–237. https://doi.org/10.21437/Interspeech.2016-1133

Chen, F., Zhang, H., Wang, S. Y., & Peng, G. (2019). Intrinsic cues and vowel categorical perception. *Linguistic Sciences*, *18*(4), 410–425. https://doi.org/10.7509/j.linsci.201808.032304

Cheng, S. T. T., Lam, G. Y. H., & To, C. K. S. (2017). Pitch perception in tone language-speaking adults with and without autism spectrum disorders. *I-Perception*, *8*(3), 2041669517711200. https://doi.org/10.1177/2041669517711200

Cheung, H., Chung, K. K. H., Wong, S. W. L., McBride-Chang, C., Penney, T. B., & Ho, C. S. H. (2009). Perception of tone and aspiration contrasts in Chinese children with dyslexia. *Journal of Child Psychology and Psychiatry*, *50*(6), 726–733. https://doi.org/10.1111/j.1469-7610.2008.02001.x

Constantino, J. N., Gruber, C. P., Davis, S., Hayes, S., Passanante, N., & Przybeck, T. (2004). The factor structure of autistic traits. *Journal of Child Psychology and Psychiatry*, *45*(4), 719–726. https://doi.org/10.1111/j.1469-7610.2004.00266.x

Constantino, J. N., Yang, D., Gray, T. L., Gross, M. M., Abbacchi, A. M., Smith, S. C., Kohn, C. E., & Kuhl, P. K. (2007). Clarifying the associations between language and social development in autism: A study of non-native phoneme recognition. *Journal of Autism and Developmental Disorders*, *37*(7), 1256–1263. https://doi.org/10.1007/s10803-006-0269-9

Dawson, G., Meltzoff, A. N., Osterling, J., Rinaldi, J., & Brown, E. (1998). Children with autism fail to orient to naturally occurring social stimuli. *Journal of Autism and Developmental Disorders*, *28*(6), 479–485. https://doi.org/10.1023/A:1026043926488

Derrick, B., Russ, B., Toher, D., & White, P. (2017). Test statistics for the comparison of means for two samples that include both paired and independent observations. *Journal of Modern Applied Statistical Methods*, 16(1), 137-157. doi: 10.22237/jmasm/1493597280

Falter, C. M., Noreika, V., Wearden, J. H., & Bailey, A. J. (2012). More consistent, yet less sensitive: Interval timing in autism spectrum disorders. *Quarterly Journal of Experimental Psychology*, *65*(11), 2093–2107. https://doi.org/10.1080/17470218.2012.690770

Feng, Y. (2018). *Development of categorical perception in Mandarin-speaking children* [Master Dissertation]. Central China Normal University.

Ferri, R., Elia, M., Agarwal, N., Lanuzza, B., Musumeci, S. A., & Pennisi, G. (2003). The mismatch negativity and the P3a components of the auditory event-related potentials in autistic low-functioning subjects. *Clinical Neurophysiology*, *114*(9), 1671–1680. https://doi.org/10.1016/S1388-2457(03)00153-6

Finney, D. J. (1971). *Probit analysis* (3rd ed.). Cambridge: Cambridge University Press.

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

Frith, U. (1989). *Autism: Explaining the Enigma*. Blackwell, Oxford.

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., Willis, C., Emslie, H., & Baddeley, A. D. (1991). The influences of number of syllables and wordlikeness on children's repetition of nonwords. *Applied Psycholinguistics*, *12*(3), 349–367. https://doi.org/10.1017/S0142716400009267

Gilliam, J. E. (2006). *Gilliam Autism Rating Scale: GARS 2*. Austin, TX: PRO-ED.

Gomot, M., Giard, M.-H., Adrien, J.-L., Barthelemy, C., & Bruneau, N. (2002). Hypersensitivity to acoustic change in children with autism: Electrophysiological evidence of left frontal cortex dysfunctioning. *Psychophysiology*, *39*(5), 577–584. https://doi.org/10.1017/S0048577202394058

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

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

Hallé, P. A., Chang, Y., & Best, C. T. (2004). Identification and discrimination of Mandarin Chinese tones by Chinese vs. French listeners. *Journal of Phonetics*. 32, 395–421.

Hamann, C. (2017). Phonological working memory and language development: What are the measures and what do they measure? *Applied Psycholinguistics*, *38*(6), 1313–1321. https://doi.org/10.1017/S0142716417000261

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., Pring, L., & Hermelin, B. (2001). Musical processing in high functioning children with autism. *Annals of the New York Academy of Sciences*, *930*(1), 443–444. https://doi.org/10.1111/j.1749-6632.2001.tb05765.x

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

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

Hoonhorst, I., Medina, V., Colin, C., Markessis, E., Radeau, M., Deltenre, P., & Serniclaes, W. (2011). Categorical perception of voicing, colors and facial expressions: A developmental study. *Speech Communication*, *53*(3), 417–430. https://doi.org/10.1016/j.specom.2010.11.005

Hua, Z., & Dodd, B. (2000). The phonological acquisition of Putonghua (Modern Standard Chinese). *Journal of Child Language*, *27*(1), 3–42. https://doi.org/10.1017/S030500099900402X

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

Järvinen-Pasley, A., & Heaton, P. (2007). Evidence for reduced domain-specificity in auditory processing in autism. *Developmental Science*, *10*(6), 786–793. https://doi.org/10.1111/j.1467-7687.2007.00637.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

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

Kasai, K., Hashimoto, O., Kawakubo, Y., Yumoto, M., Kamio, S., Itoh, K., Koshida, I., Iwanami, A., Nakagome, K., Fukuda, M., Yamasue, H., Yamada, H., Abe, O., Aoki, S., & Kato, N. (2005). Delayed automatic detection of change in speech sounds in adults with autism: A magnetoencephalographic study. *Clinical Neurophysiology*, *116*(7), 1655–1664. https://doi.org/10.1016/j.clinph.2005.03.007

Kawahara, H., Takahashi, T., Morise, M., & Banno, H. (2009). Development of exploratory research tools based on TANDEM-STRAIGHT. *Proceedings of International Conference on Asia-Pacific Signal and Information Processing Association*, 111–121.

Klin, A. (1991). Young autistic children's listening preferences in regard to speech: A possible characterization of the symptom of social withdrawal. *Journal of Autism and Developmental Disorders*, *21*(1), 29–42. https://doi.org/10.1007/BF02206995

Koerner, T. K., & Zhang, Y. (2017). Application of linear mixed-effects models in human neuroscience research: A comparison with Pearson correlation in two auditory electrophysiology studies. *Brain Sciences*, *7*(3), 26. https://doi.org/10.3390/brainsci7030026

Kuhl, P. K. (1991). Human adults and human infants show a "perceptual magnet effect" for the prototypes of speech categories, monkeys do not. *Perception & Psychophysics*, *50*(2), 93–107. https://doi.org/10.3758/BF03212211

Kuhl, P. K. (2000). A new view of language acquisition. *Proceedings of the National Academy of Sciences*, *97*(22), 11850–11857. https://doi.org/10.1073/pnas.97.22.11850

Kuhl, P. K. (2004). Early language acquisition: Cracking the speech code. *Nature Reviews Neuroscience*, *5*(11), 831–843. https://doi.org/10.1038/nrn1533

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

Kuhl, P. K., Conboy, B. T., Coffey-Corina, S., Padden, D., Rivera-Gaxiola, M., & Nelson, T. (2008). Phonetic learning as a pathway to language: New data and native language magnet theory expanded (NLM-e). *Philosophical Transactions of the Royal Society B: Biological Sciences*, *363*(1493), 979–1000. https://doi.org/10.1098/rstb.2007.2154

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software*, *82*(1), 1–26. https://doi.org/10.18637/jss.v082.i13

Lancker, D. van. (1980). Cerebral lateralization of pitch cues in the linguistic signal. *Paper in Linguistics*, *13*(2), 201–277. https://doi.org/10.1080/08351818009370498

Landa, R. J. (2008). Diagnosis of autism spectrum disorders in the first 3 years of life. *Nature Clinical Practice Neurology*, *4*(3), 138–147. https://doi.org/10.1038/ncpneuro0731

Lau, J. C., To, C. K., Kwan, J. S., Kang, X., Losh, M., & Wong, P. C. (2020). Lifelong tone language experience does not eliminate deficits in neural encoding of pitch in autism spectrum disorder. *Journal of Autism and Developmental Disorders*, 1-20. https://doi.org/10.1007/s10803-020-04796-7

Lenth, R. V. (2016). Least-squares means: The R package lsmeans. *Journal of Statistical Software*, *69*(1), 1–33. https://doi.org/10.18637/jss.v069.i01

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

Liberman, A. M., Harris, K. S., Hoffman, H. S., & Griffith, B. C. (1957). The discrimination of speech sounds within and across phoneme boundaries. *Journal of Experimental Psychology*, *54*(5), 358–368. https://doi.org/10.1037/h0044417

Lisker, L., & Abramson, A. S. (1964). A cross-language study of voicing in initial stops: Acoustical measurements. *WORD*, *20*(3), 384–422. https://doi.org/10.1080/00437956.1964.11659830

Lord, C., Rutter, M., DiLavore, P. C., Risi, S., Gotham, K., & Bishop, S. (2012). *Autism Diagnostic Observation Schedule: ADOS–2*. Los Angeles, CA: Western Psychological Services.

Macmillan, N. A., & Creelman, C. D. (2005). *Detection Theory: A User's Guide* (2nd ed.). Mahwah, NJ: Erlbaum.

Maister, L., & Plaisted-Grant, K. C. (2011). Time perception and its relationship to memory in Autism Spectrum Conditions. *Developmental Science*, *14*(6), 1311–1322. https://doi.org/10.1111/j.1467-7687.2011.01077.x

Martin, J. S., Poirier, M., & Bowler, D. M. (2010). Brief report: Impaired temporal reproduction performance in adults with autism spectrum disorder. *Journal of Autism and Developmental Disorders*, *40*(5), 640–646. https://doi.org/10.1007/s10803-009-0904-3

Massaro, D. W. (1987). *Speech perception by ear and eye: A paradigm for psychological inquiry*. Hillsdale, NJ: Erlbaum.

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

Medina, V., Hoonhorst, I., Bogliotti, C., & Serniclaes, W. (2010). Development of voicing perception in French: Comparing adults, adolescents, and children. *Journal of Phonetics*, *38*(4), 493–503. https://doi.org/10.1016/j.wocn.2010.06.002

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

Mitterer, H., & Mattys, S. L. (2017). How does cognitive load influence speech perception? An encoding hypothesis. *Attention, Perception, & Psychophysics*, *79*(1), 344–351. https://doi.org/10.3758/s13414-016-1195-3

Mottron, K., & 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* (pp. 131–148). Mahwah, NJ: Lawrence Erlbaum Associates.

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

Ning, C. Y. (2013). *Test of Language Ability in Mandarin-speaking Preschoolers*. Institute of Linguistics, Tianjin Normal University: Tianjin University Press.

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

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

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

R Core Team. (2014). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing. http://www.R-project.org/

Raven, J., & Court, J. (1998). *Raven manual: Section 3. Standard progressive matrices*. Oxford: Oxford Psychologists Press.

Rispens, J., & Baker, A. (2012). Nonword repetition: The relative contributions of phonological short-term memory and phonological representations in children with language and reading impairment. *Journal of Speech, Language, and Hearing Research*, *55*(3), 683–694. https://doi.org/10.1044/1092-4388(2011/10-0263)

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

Shao, J., Wang, L., & Zhang, C. C. (2020). Talker processing in Mandarin-speaking congenital amusics. *Journal of Speech, Language, and Hearing Research*, *63*(5), 1361–1375. https://doi.org/10.1044/2020_JSLHR-19-00209

Soulières, I., Mottron, L., Saumier, D., & Larochelle, S. (2007). Atypical categorical perception in autism: Autonomy of discrimination? *Journal of Autism and Developmental Disorders*, *37*(3), 481–490. https://doi.org/10.1007/s10803-006-0172-4

Stewart, M. E., & Ota, M. (2008). Lexical effects on speech perception in individuals with “autistic” traits. *Cognition*, *109*(1), 157–162. https://doi.org/10.1016/j.cognition.2008.07.010

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

Swaminathan, J., Krishnan, A., Gandour, J. T., & Xu, Y. (2008). Applications of static and dynamic iterated rippled noise to evaluate pitch encoding in the human auditory brainstem. *IEEE Transactions on Biomedical Engineering*, *55*(1), 281–287. https://doi.org/10.1109/TBME.2007.896592

Szelag, E., Kowalska, J., Galkowski, T., & Pöppel, E. (2004). Temporal processing deficits in high-functioning children with autism. *British Journal of Psychology*, *95*(3), 269–282. https://doi.org/10.1348/0007126041528167

Szewczyk, J. M., Marecka, M., Chiat, S., & Wodniecka, Z. (2018). Nonword repetition depends on the frequency of sublexical representations at different grain sizes: Evidence from a multi-factorial analysis. *Cognition*, *179*, 23–36. https://doi.org/10.1016/j.cognition.2018.06.002

Wang, W. S.-Y. (1976). Language change. In S. R. Harnad, H. D. Steklis, & J. Lancaster (Eds.), *Origins and evolution of language and speech* (Vol. 280, pp. 61–72). New York: New York Academy of Sciences.

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*, 43254. https://doi.org/10.1038/srep43254

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

Wu, H., Lu, F., Yu, B., & Liu, Q. (2020). Phonological acquisition and development in Putonghua-speaking children with Autism Spectrum Disorders. *Clinical Linguistics & Phonetics*, *34*(9), 844–860.

Xi, J., Jiang, W., Zhang, L., & Shu, H. (2009). Categorical Perception of VOT and Lexical Tones in Chinese and the Developmental Course. *Acta Psychologica Sinica*, *41*(7), 572–579.

Xi, J., Zhang, L., Shu, H., Zhang, Y., & Li, P. (2010). Categorical perception of lexical tones in Chinese revealed by mismatch negativity. *Neuroscience*, *170*(1), 223–231. https://doi.org/10.1016/j.neuroscience.2010.06.077

Xu, Y., Gandour, J. T., & Francis, A. L. (2006). Effects of language experience and stimulus complexity on the categorical perception of pitch direction. *The Journal of the Acoustical Society of America*, *120*(2), 1063–1074. https://doi.org/10.1121/1.2213572

You, R. S., Serniclaes, W., Rider, D., & Chabane, N. (2017). On the nature of the speech perception deficits in children with autism spectrum disorders. *Research in Developmental Disabilities*, *61*, 158–171.

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

List of Abbreviations

ADOS-2: Autism Diagnostic Observation Schedule-2

ASD: Autism spectrum disorder

CP: Categorical perception

GARS-2: Gilliam Autism Rating Scale–Second Edition

IRN: Iterated rippled noise

LMM: linear mixed-effect model

MMN: Mismatch negativity

MMR: Mismatch Responses

SD: Standard deviation

TD: Typically developing

VOT: Voice onset time

Tables

Table 1. *Means (and standard deviations) of chronological age, language ability, forward digit span, and nonword repetition for two ASD subgroups in the perception of pitch contours and VOT respectively, and TD controls.*

Group	Number (male)	Chronological Age	Language Ability	Forward Digit Span	Nonword Repetition
ASD (Pitch)	20 (16)	13.87 (2.88)	87.40 (10.38)	6.45 (1.99)	75.03% (7.83%)
ASD (VOT)	15 (11)	14.06 (2.87)	87.73 (10.50)	6.47 (1.65)	75.42% (7.00%)
TD controls	20 (14)	13.46 (0.79)	98.60 (2.06)	7.40 (0.77)	84.96% (2.72%)

Table 2. *The stimulus features of four types of pitch carriers.*

Type	*Pitch contrast*	*Spectro-temporal complexity*	*Phonemic contrast*	*Semantic contrast*
Real word (speech)	+	+	+	+
Nonword (speech)	+	+	+	-
IRN (nonspeech)	+	+	-	-
Pure Tone (nonspeech)	+	-	-	-

Table 3. *The regression coefficients indicating the relationships between chronological age/language ability/digit span/nonword repetition and the degree of CP of pitch contours (boundary width/peakedness score) in individuals with ASD.*

Pitch Carriers	Predictors	Boundary Width				Peakedness Score			
		β	SE	*t*	*p* value	β	SE	*t*	*p* value
Real Word	Chronological Age	0.018	0.028	0.615	0.548	0.130	0.115	1.132	0.275
	Language Ability	**-0.032**	**0.011**	**-3.009**	**0.009****	-0.021	0.043	-0.482	0.637
	Digit Span	**-0.109**	**0.050**	**-2.170**	**0.046***	0.361	0.203	1.784	0.094
	Nonword Repetition	1.128	1.186	0.951	0.357	-7.081	4.774	-1.483	0.159
Nonword	Chronological Age	0.057	0.047	1.209	0.246	0.012	0.145	0.080	0.937
	Language Ability	0.013	0.016	0.801	0.435	0.046	0.049	0.918	0.373
	Digit Span	-0.084	0.077	-1.089	0.293	-0.296	0.236	-1.254	0.229
	Nonword Repetition	**-5.302**	**1.656**	**-3.201**	**0.005****	-3.043	5.079	-0.599	0.558
IRN	Chronological Age	0.005	0.037	0.129	0.899	-0.034	0.171	-0.197	0.846
	Language Ability	-0.020	0.013	-1.575	0.136	0.053	0.059	0.900	0.382
	Digit Span	-0.040	0.060	-0.675	0.510	0.376	0.279	1.346	0.198
	Nonword Repetition	1.234	1.286	0.959	0.353	-2.696	6.015	-0.448	0.660
Pure Tone	Chronological Age	0.056	0.045	1.232	0.237	0.012	0.105	0.113	0.911
	Language Ability	-0.022	0.016	-1.417	0.177	-0.033	0.036	-0.931	0.367
	Digit Span	-0.098	0.074	-1.333	0.202	0.026	0.171	0.149	0.883
	Nonword Repetition	0.542	1.589	0.341	0.738	3.545	3.678	0.964	0.350

** $p < .01$, * $p < .05$

Table 4. *The regression coefficients indicating the relationships between chronological age/language ability/digit span/nonword repetition and the degree of CP of VOT (boundary width/peakedness score) among participants with ASD.*

	Predictors	Boundary Width				Peakedness Score			
		β	SE	*t*	*p* value	β	SE	*t*	*p* value
VOT	Chronological Age	-0.001	0.141	-0.010	0.992	-0.134	0.175	-0.772	0.458
	Language Ability	0.068	0.078	0.874	0.403	-0.106	0.096	-1.094	0.299
	Digit Span	-0.354	0.490	-0.721	0.487	0.843	0.604	1.395	0.193
	Nonword Repetition	-5.537	5.480	-1.010	0.336	3.182	6.745	0.472	0.647

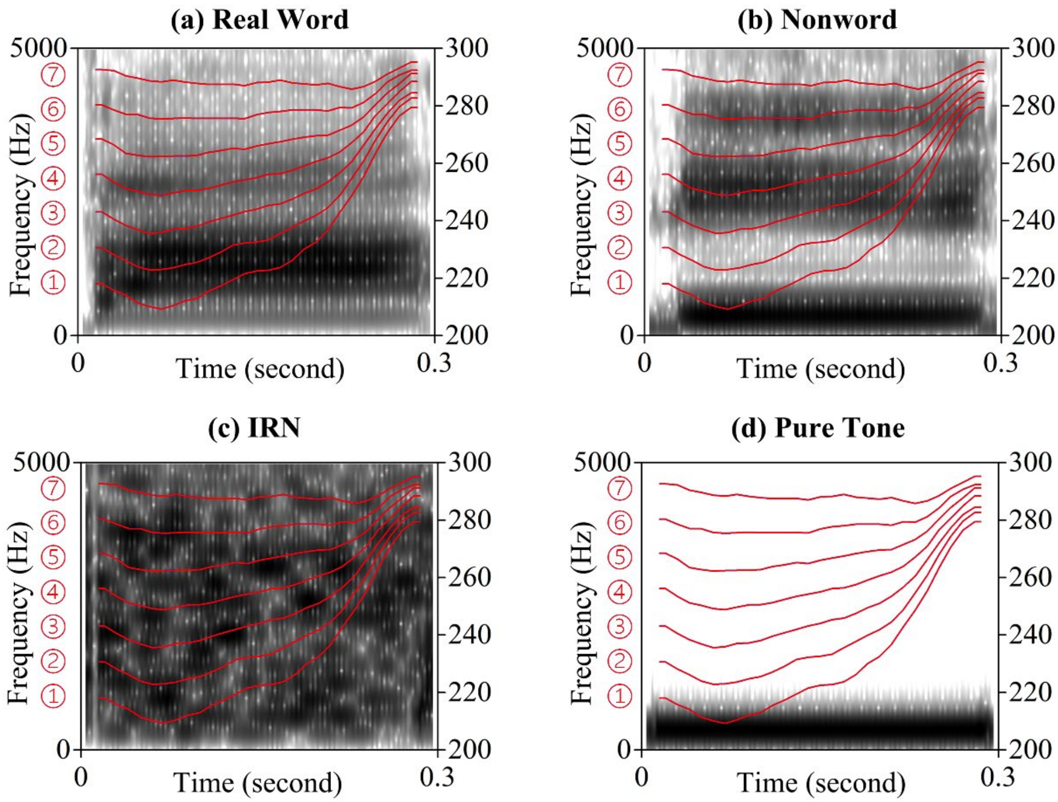

Figure 1. Schematic diagram of pitch contours embedded in (a) real word, (b) nonword, (c) IRN, and (d) pure tone. The right-side y-axis indicates the F0 in Hz. The red curves indicate the pitch contours along each continuum.

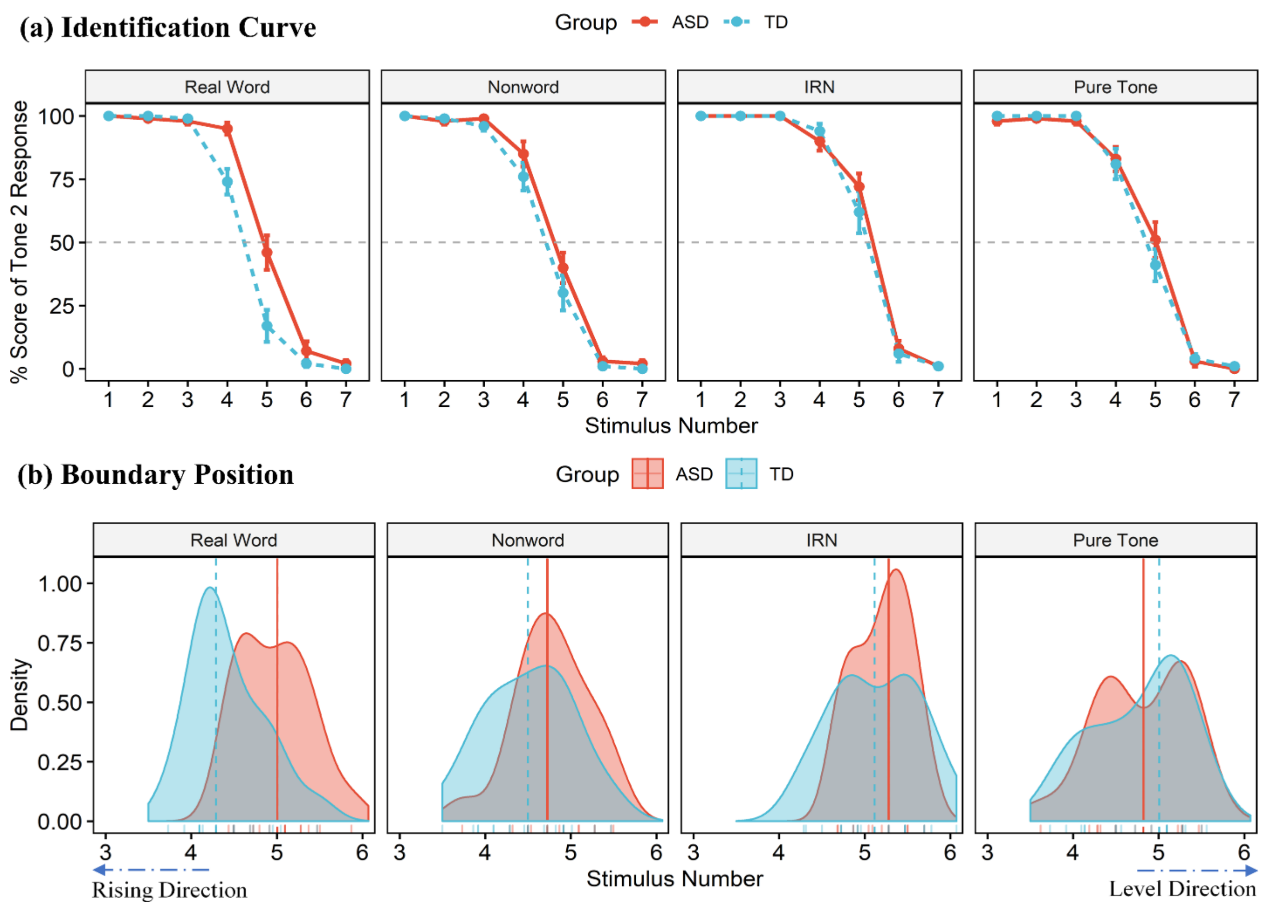

Figure 2. (a) The identification curves of Tone 2 responses, and (b) boundary positions in ASD and TD groups among four types of pitch contours (real word, nonword, IRN, and pure tone). The vertical lines indicate the median values of the boundary position.

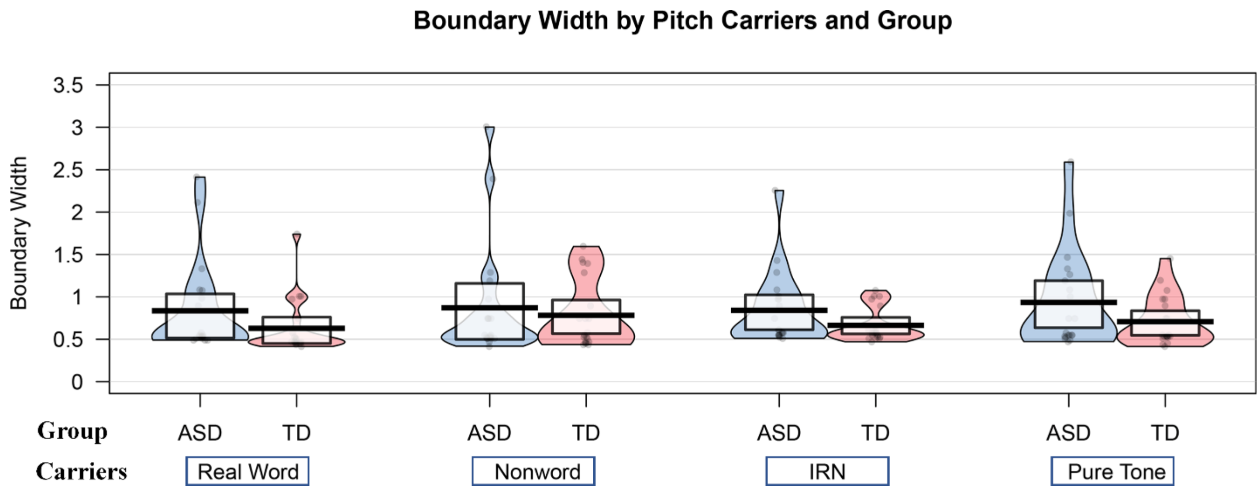

Figure 3. The boundary widths in ASD and TD groups among four types of pitch contours (real word, nonword, IRN, and pure tone).

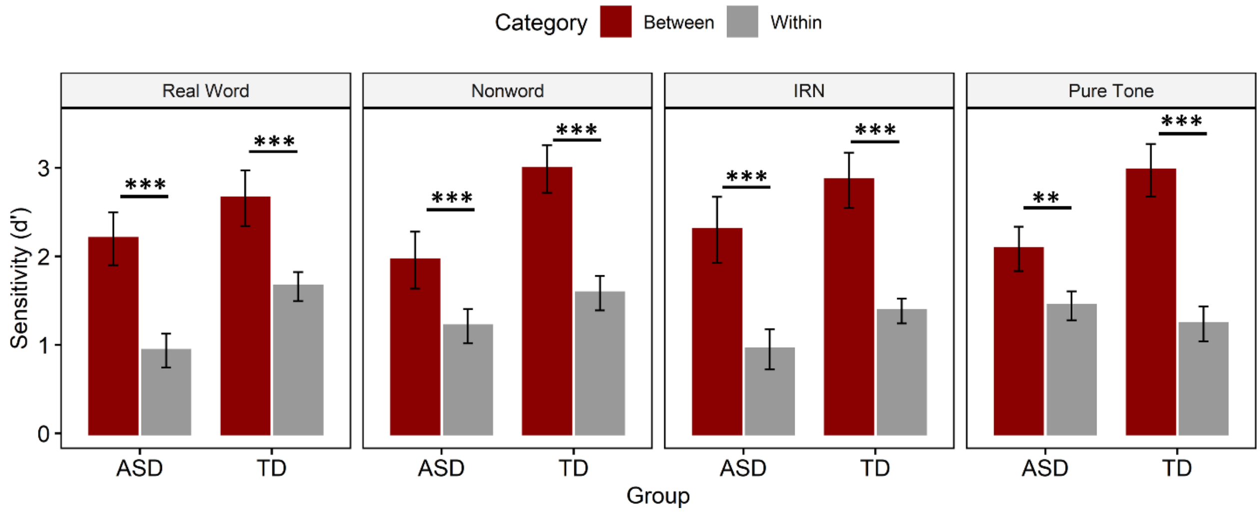

Figure 4. The *d′* of the between- and within-category units for the ASD and TD participants under different types of pitch carriers (real word, nonword, IRN, and pure tone). Error bars: +/- 1 standard error. ***p < .001, **p < .01

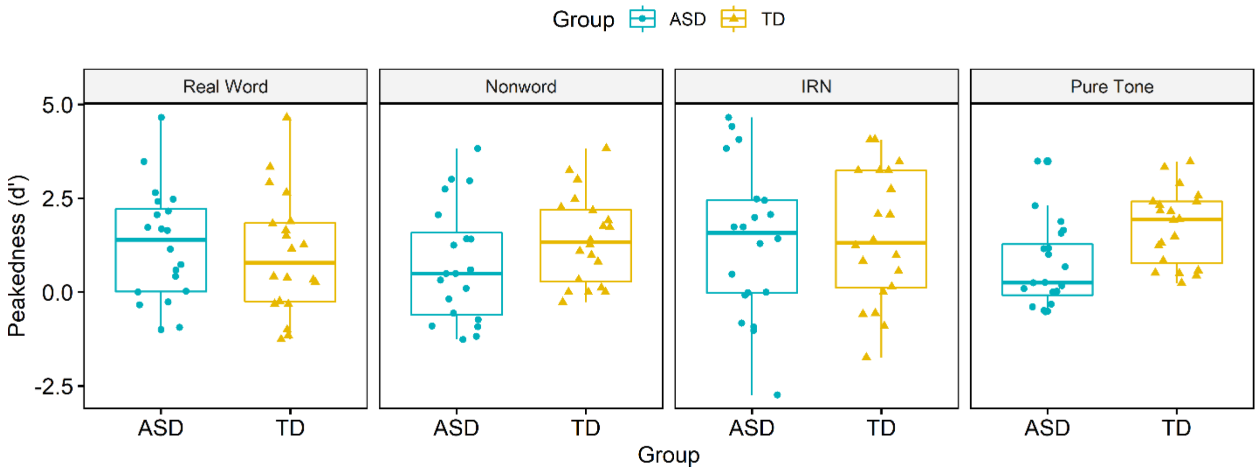

Figure 5. Box plots of peakedness scores (d') for the ASD and TD participants under different types of pitch carriers (real word, nonword, IRN, and pure tone).

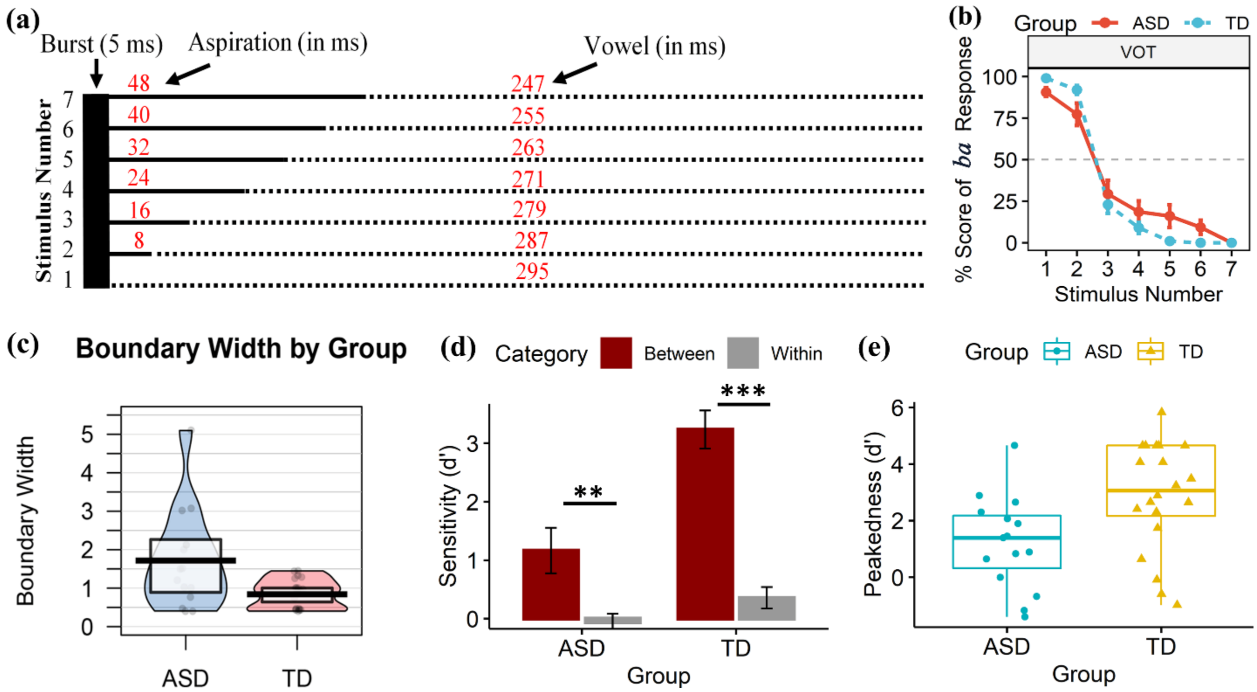

Figure 6. (a) Schematic diagram of VOT continuum; (b) The identification curves of unaspirated *ba*/pa/ responses; (c) The boundary widths for ASD and TD groups in the identification of VOT; (d) The *d′* values of the between-category and within-category units for the ASD and TD participants in the discrimination of VOT; (e) Box plots of peakedness scores (*d′*) for the ASD and TD participants in discrimination of VOT. ***$p < .001$, **$p < .01$