Linguistic Tone and Non-Linguistic Pitch Imitation in Children with Autism Spectrum Disorders: A Cross-Linguistic Investigation

Fei Chen[1], Candice Chi-Hang Cheung[2], and Gang Peng[2]

[1]*School of Foreign Languages, Hunan University, Changsha, China;*

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

Abstract

The conclusions on prosodic pitch features in autism spectrum disorders (ASD) have primarily been derived from studies in non-tonal language speakers. This cross-linguistic study evaluated the performance of imitating Cantonese lexical tones and their non-linguistic (nonspeech) counterparts by Cantonese- and Mandarin-speaking children with and without ASD. Acoustic analyses showed that, compared with typically developing peers, children with ASD exhibited increased pitch variations when imitating lexical tones, while performed similarly when imitating the nonspeech counterparts. Furthermore, Mandarin-speaking children with ASD failed to exploit the phonological knowledge of segments to improve the imitation accuracy of non-native lexical tones. These findings help clarify the speech-specific pitch processing atypicality and phonological processing deficit in tone-language-speaking children with ASD.

Keywords: lexical tone; non-linguistic pitch; imitation; Mandarin; Cantonese; ASD

Linguistic Tone and Non-Linguistic Pitch Imitation in Children with Autism Spectrum Disorders: A Cross-Linguistic Investigation

Introduction

Prosody is a broad term including suprasegmental properties such as intonation, tone, rhythm, and stress, which is used to convey various linguistic, attitudinal, emotional, pragmatic, and idiosyncratic functions (Bolinger, 1972; Cutler & Isard, 1980). The acoustic correlates of speech prosody involve pitch (fundamental frequency, F0), duration, intensity, and their mutual interaction. The two gold-standard clinical assessments of autism spectrum disorders (ASD), namely, Autism Diagnostic Observation Schedule, second version (ADOS-2; Lord et al., 2012) and Autism Diagnostic Interview-Revised (ADI-R; Rutter et al., 2003), have indexed prosodic atypicality as one of the diagnostic characteristics. Starting from the earliest delineation of the autistic syndrome with peculiar use of the tone of voice (Kanner, 1943), unusual prosody has been frequently identified as a central feature of speech and communication in autism (Baltaxe & D'Angiola, 1992; Diehl & Paul, 2012; Loveall et al., 2021; McCann & Peppé, 2003; Paul et al., 2005a; Peppé et al., 2011; Shriberg et al., 2001; Tager-Flusberg, 1981). Furthermore, prosodic differences have been found to be tightly associated with the general ratings of social and communicative competence in autism (Paul et al., 2005b).

To investigate the prosodic pitch features in ASD, several studies focused on the production of intonation, which is expressed as the variation in voice pitch at the sentence level. Some earlier reports based on observation or subjective ratings revealed dull, monotonic, or machine-like intonation in speech produced by some children with autism (Fay & Schuler, 1980; Kanner, 1943, 1971; Tager-Flusberg, 1981). However, contrary to the common impression of monotonic speech in autism, most of the studies adopting acoustic analyses consistently showed a wider pitch range and/or greater pitch standard deviation (SD), indicating an increased pitch variability of intonation produced by individuals with ASD (Bonneh et al., 2011; Diehl et al., 2009; Filipe et al., 2014; Fosnot & Jun, 1999; Green & Tobin, 2009; Hubbard & Trauner, 2007; Nadig & Shaw, 2012; Sharda et al., 2010). The only exception was reported by Nakai et al. (2014), which showed a smaller pitch range of intonation in Japanese-speaking children with ASD compared to typically developing (TD) peers at school age. In brief, the subjective impression of "flat" intonation

in ASD has not been confirmed by accumulating evidence from the acoustic analyses in most of previous studies in non-tonal language speakers with ASD, including English-speaking children (Diehl et al., 2009; Fosnot & Jun, 1999; Hubbard & Trauner, 2007; Nadig & Shaw, 2012), Hebrew-speaking children (Bonneh et al., 2011; Green & Tobin, 2009), Portuguese-speaking children (Filipe et al., 2014), and English-Hindi bilinguals (Sharda et al., 2010).

World's languages exhibit a natural diversity, among which the tonal languages (such as Mandarin and Cantonese) make use of F0 changes at the syllable level to mark phonological contrasts (Wang, 1973). Thus, in tonal languages, F0-based prosodic changes could be realized not only in the larger prosodic unit of a whole sentence (i.e., intonation), but also in the smaller unit of a syllable (i.e., lexical tone). The production of intonation in tone-language-speaking individuals with ASD was investigated by Chan and To (2016). The results showed that, in consistent with previous findings in non-tonal language speakers, Cantonese-speaking adults with high-functioning autism also demonstrated significantly higher SD of F0 than neurotypical adults, suggesting that the atypical sentence-level intonation may be a universal characteristic of individuals with ASD. However, to the best of our knowledge, no studies so far have depicted the syllable-level prosodic profile in tone-language-speaking children with ASD. There has been an analogical relationship between lexical tone and intonation as "small ripples riding on large waves" (Chao, 1968:39), implying that the dynamic changes in intonation at the sentence level might not transform or modify the lexical tones at the syllable level. However, some other scholars proposed a close interaction between lexical tone and intonation (Liu & Xu, 2005; Ma et al., 2006; Yuan, 2011). The first aim of this study is therefore to address this issue by performing prosodic pitch analyses of lexical tone imitations in tone-language-speaking children, to explore whether the overall prosodic pitch pattern at the syllable level is more variable in the clinical population of ASD.

More importantly, beyond the basic prosodic features, lexical tones constitute a speech element distinguishing word meanings in a tonal language, with the same linguistic effect as changing a vowel or a consonant. For instance, /ji/ spoken with different lexical tones in Cantonese can respectively mean /ji55/ *doctor*, /ji25/ *chair*, /ji33/ *meaning*, /ji21/ *son*, /ji23/ *ear*, and /ji22/ *two.* The digits here refer to lexical tone transcriptions in Chao's five-scale tone letters (Chao, 1930), with 5 being the highest and 1 being the lowest "relative" pitch level of a speaker's normalized pitch change. The mispronounced pitch contours of lexical tones might lead to comprehension and communication barriers.

Cantonese and Mandarin are two widely-spoken and well-studied tonal languages, with Cantonese tonal system being more complex. In Cantonese tone (CT) inventory, there are six citation tones in open syllables contrasting in both pitch height and slope (Gandour, 1981): Three level tones (high-level CT55 vs. mid-level CT33 vs. low-level CT22), two rising tones (low-rising CT23 vs. high-rising CT25), and one falling tone (low-falling CT21). The three level tones in CT (i.e., CT55–CT33–CT22) differ mainly in pitch height, while the other three contour tones (i.e., CT23–CT25–CT21) in both pitch height and direction. In Mandarin tone (MT) inventory, there are only four citation tones, each of which carries a distinct pitch contour (Wang, 1973; Yip, 2002): high-level MT55, high-rising MT35, high-falling MT51, and dipping or low-falling-rising MT214 (being realized as high-rising *MT35 at the non-final position when occurred before another dipping tone, and as low-falling *MT21 when the following tone is not a dipping tone). According to a corpus-based study of Mandarin and Cantonese (Peng, 2006), the acoustic distribution of CT25 on the tone chart is very similar to that of MT35, CT55 is similar to MT55, and CT21 is similar to *MT21, while there are no direct MT counterparts for two level CTs (CT33 and CT22) and the low-rising CT23.

In the current cross-linguistic study, the production data of imitating CTs in both Cantonese- and Mandarin-speaking children with ASD were analyzed and compared. Imitation skills play a potential role in the emergence and evolution of phonological systems (Lewis, 1957; Nguyen & Delvaux, 2015). During the process of speech sound acquisition, it is generally believed that imitation from adult models generates the most natural forms for its underlying mechanisms (Ingersoll, 2008; Kim & Clayards, 2019; Messum, 2008). Some researchers proposed that children with ASD could not accurately imitate the prosodic patterns of adult models (Diehl & Paul, 2012; Fosnot & Jun, 1999; Hubbard & Trauner, 2007; Peppé et al., 2011). On the other hand, others claimed that children with ASD could imitate the tone of voice and rhythm of other speakers as well as TD children (Fan et al., 2010; Frankel et al., 1987). Since the complex tonal system of Cantonese contains fine-grained pitch differences regarding both pitch height and direction, imitation of CT in both Cantonese- and Mandarin-speaking children with ASD offers a valuable chance to evaluate the imitation abilities in children with ASD, and to further illustrate how such performance changes as a function of language experience.

As suggested by cross-linguistic processing models such as the Speech Learning Model (SLM; Flege, 1995, 2007), the Perceptual Assimilation Model for suprasegmentals (PAM-S; So & Best, 2010, 2014), and the Similarity Differential Rate Hypothesis (SDRH; Major & Kim, 1996), the outcome of non-native phonetic processing can be

related to the "cross-linguistic similarity" between a non-native item and its closest native counterpart. Accordingly, when imitating the Cantonese tones, Mandarin-speaking children might assimilate the acoustically similar non-native tones of CT25, CT55, and CT21 into their native lexical tones of MT35, MT55, and *MT21 respectively. In contrast, when there is no similarity between a non-native sound and its native sound, the formation of mental representation for a novel and unfamiliar non-native category will gradually occur (Flege, 2007). That's to say, for Mandarin-speaking individuals, the three non-native lexical tones of CT55, CT25, and CT21 are familiar tonal stimuli, while the other three L2 stimuli of CT33, CT22, and CT23 are presumed to be unfamiliar since they have no direct acoustic counterparts in Mandarin (Peng, 2006; So & Best, 2010). Besides, the suprasegmental lexical tones are superimposed on the segmental components of each syllable, which might in turn exert an influence on the processing of non-native lexical tones. For instance, for healthy Mandarin-speaking adults, the native and familiar segments /fu/, /ji/ (existing in both Mandarin and Cantonese) helped improve their discrimination and/or identification accuracy of CT compared with the non-native and unfamiliar segments /si/ and /sɛ/ (only existing in Cantonese; note that Cantonese syllable /si/ is different from Mandarin /sɿ/) (Wang & Peng, 2014), indicating a top-down influence of phonological processing. Therefore, in the current study, the tonal familiarity and segmental familiarity of the speech models were manipulated to investigate the influence of phonological knowledge on the performance of lexical tone imitation in children with ASD.

In a nutshell, this cross-linguistic study evaluated the capacity of lexical tone and non-linguistic (i.e., nonspeech) pitch imitation in Cantonese- and Mandarin-speaking children with and without ASD. In speech condition, CTs were adopted as the models due to a richer inventory of tonal types than MTs (Gandour, 1983; Peng, 2006). Besides, in this study of pitch imitation, nonspeech analogues were also generated sharing exactly the same pitch trajectories with the three level tones (CT55, CT33, CT22) and three contour tones (CT23, CT25, CT21) in Cantonese, in an effort to test whether the atypical imitation performance in ASD was speech-specific or domain-general. Three general research questions are advanced. First, would the prosodic pitch pattern in tone-language-speaking children with ASD show an increased pitch variation compared to TD children during lexical tone imitations? Second, when imitating native and non-native lexical tones, would children with ASD be able to imitate normal-like lexical tones that are acoustically comparable to those produced by TD peers? Third, how would top-down phonological knowledge (segmental familiarity: familiar vs. unfamiliar; tonal familiarity: familiar vs. unfamiliar) influence the imitation accuracy in children with and without ASD? Two experiments were conducted in this study. Experiment 1 performed

acoustic analyses of prosodic pitch pattern as well as lexical tone imitations in order to answer the first and the second research questions respectively. The perceptual identification study in Experiment 2 was conducted to answer the third research question.

Experiment 1. Acoustic Analyses of Lexical Tone and Non-Linguistic Pitch Imitation

Methods

Participants

In total, 104 participants took part in this study and completed all the tests. There were 26 Cantonese-speaking children with ASD (CASD, two girls, M_{age} = 7.44 yr), 26 Cantonese-speaking TD children (CTD, one girl, M_{age} = 7.48 yr), 26 Mandarin-speaking children with ASD (MASD, one girl, M_{age} = 7.69 yr), and 26 Mandarin-speaking TD children (MTD, one girl, M_{age} = 7.65 yr). The Cantonese-speaking children with and without ASD were recruited from Hong Kong, and all spoke Cantonese as their first language at home and school with little exposure to Mandarin. The Mandarin-speaking participants were recruited from Shenzhen and used Mandarin as their first language with little exposure to Cantonese. All the child participants had neither hearing impairment nor comorbidities such as developmental motor speech disorder. Permission to conduct this study was obtained from the Hong Kong Polytechnic University, ensuring appropriate adherence to informed consent procedures.

The clinical diagnosis of ASD was established according to the DSM-5 (American Psychiatric Association, 2013), and the ADOS-2 (Lord et al., 2012) by pediatricians and child psychiatrists with expertise in diagnosing ASD in local hospitals before enrollment. The participants were 6- to 11-year-old high-functioning autism without major cognitive delays [nonverbal intelligence quotient (IQ) $\geqslant$70; mean = 105.23] and without severe language delays (i.e., able to use full, complex sentences). The school-age children were chosen since neurotypical preschoolers are still fine-tuning their control over coordinating pitch range, slope, and curvature in the production of native tonal categories (Mok et al., 2019, 2020; Rattanasone et al., 2018; Wong, 2013). The obtained language score in Cantonese-speaking participants was averaged across three subtests (Textual Comprehension Test, Expressive Nominal Vocabulary Test, and Test of Hong Kong Cantonese Grammar) in the *Hong Kong Cantonese Oral Language Assessment Scale* (T'sou et al., 2006). The nonverbal IQ in Cantonese-speaking participants was evaluated with *The Primary Test of Nonverbal Intelligence* (Ehrler & McGhee, 2008). The nonverbal and verbal IQs in Mandarin-speaking participants were assessed

with the *Wechsler Intelligence Scale for Children* (WISC-IV, Mandarin Chinese version) (Wechsler, 2003). As shown in Table 1, the Mandarin- or Cantonese-speaking ASD group did not differ from the corresponding TD group in terms of chronological age and nonverbal IQ, while slightly lagged behind TD children in the general language functioning ($p < .05$).

[Table 1 about here]

Materials

The speech models contained 24 Cantonese syllables with three level tones (CT55, CT33, CT22) and three contour tones (CT23, CT25, CT21) superimposed on four segments in Cantonese (/fu/, /ji/, /sɛ/, /si/). The Cantonese-speaking children were familiar with all of the native speech models. However, the Mandarin-speaking children exhibited familiar and unfamiliar distinctions of lexical tones (familiar tones: CT55, CT25, CT21; unfamiliar tones: CT33, CT22, CT23) as well as segments (familiar segments: /fu/, /ji/; unfamiliar segments: /sɛ/, /si/) in the non-native Cantonese syllables. These 24 Cantonese syllables were recorded 10 times in a natural way from 10 Cantonese adult speakers (five females; five males) who were born and raised in Hong Kong. We picked out the speech samples spoken by two representative speakers (one female voice and one male voice) whose tonal production was closer to the median of pitch height and slope among the 10 native speakers. Then, one speech sample (with the best voice quality) for each syllable was chosen from 10 repetitions by a phonetically trained native speaker based on the clarity and stability. Altogether, 48 speech stimuli (6 lexical tones × 4 segments × 2 voice genders) recorded from one male voice and one female voice were selected as the speech models. The six tonal categories of speech models deviated from each other in the two dimensions of pitch height and pitch slope without acoustic overlapping. Moreover, the speech models were double-checked by a Cantonese-speaking linguist to ensure that they showed no perceived tone merging. The mean F0 (*SD*) for the speech models of female voice and male voice were 252 (*52*) Hz and 121 (*23*) Hz respectively. Finally, to generate the non-linguistic/nonspeech pitch models, F0 trajectories of the zero-onset syllable /ji/ (6 lexical tones × 2 voice genders) were extracted to synthesize nonspeech models using equal-amplitude triangle waves, which have a different harmonic structure from that of speech sounds (Chen & Peng, 2016). The mean F0 (*SD*) for the nonspeech models of female voice and male voice were 250 (*45*) Hz and 119 (*24*) Hz respectively. The duration of both speech and nonspeech models was not normalized to make them sound more natural. Since the nonspeech stimuli sounded lower perceptually when compared to the speech stimuli of the same intensity, for the purpose of matching the

loudness level, the average intensity level of nonspeech models was set to 80 dB SPL, 15 dB higher than that of speech models.

Procedure

The experimenters were native speakers in each language background. The participants were tested for their nonverbal IQ, verbal IQ/language ability at first (Table 1). During these cognitive and language tests, all the child participants showed no difficulties in understanding the verbal instructions, indicating that they were not deficient in perceiving L1 speech sounds in connected and natural speech. Then, in the imitation tasks, stimulus presentation was implemented with E-Prime 2.0 program (Psychology Software Tools Inc., USA). The speech/nonspeech models were played in sound-attenuated rooms via bilateral loudspeakers (JBL CM220) located at 45 degrees to the left and right of the participants at a distance of approximately 1 meter. Before the formal experiment, a training session was provided by asking participants to imitate both the speech stimuli (another Cantonese syllable /fan/ with six CTs) and nonspeech counterparts as accurately as possible to familiarize children with the procedure and requirement. The imitations in the training session were not recorded, while all the imitations in the formal test were recorded using an external microphone (SHURE MV51) around 10 cm away from the mouth of the participant. The microphone was connected to a laptop computer through a USB audio interface with a sampling rate of 44,000 Hz. In the formal test, there were two testing blocks (speech block and nonspeech block), which were presented in a random order among participants. Within each testing block, the speech or nonspeech model stimuli were repeated three times and played randomly. The whole imitation task, including training and testing parts, lasted about 30-40 minutes for each participant.

Coding and Measurements

The recorded imitations were coded offline in Praat (Boersma & Weenink, 2016). Acoustic measurement of F0 was derived by using ProsodyPro (Xu, 2013) through the automatic F0 tracking on 20 equidistant points along the pitch contour. These F0 points were further checked and manually corrected for any “pitch-halving” or “pitch-doubling” errors. Then, 10% of both leftmost and rightmost F0 points were discarded, and only the middle 80% (i.e., 16 points) for each pitch trajectory were analyzed to decrease the tone-irrelevant variation (Peng, 2006). Besides, the intensity and duration values of each imitation were also measured and entered as covariates in the statistical analyses.

For the analyses of prosodic pitch pattern, three acoustic measures – pitch mean, pitch range, and pitch SD – were analyzed in the current study. The raw F0 values (in Hz) were transformed into semitones, with a reference frequency of 100 Hz (Rattanasone et al., 2018). In particular, the pitch range was calculated as the minimum F0 subtracted from maximum F0. These three pitch measures (in semitone) were calculated for each participant regardless of different tonal categories. The pitch mean was used to provide a general characterization of prosodic pitch (high vs. low); both pitch range and pitch SD were used to depict pitch variations. The expanded pitch range and/or larger pitch SD indicated more pitch variations, and vice versa (Diehl et al., 2009).

For the analyses of lexical tone production, the pitch of each tonal category was transformed from Hz to log-scale 5-level value (Peng & Wang, 2005), consistent with the time-honored selection of number of levels for linguists to transcribe lexical tones (Chao, 1930). The log-scale 5-level value was adopted since it calculated each speaker's normalized linguistic pitch distribution and eliminated inter-speaker variations in absolute pitch differences. Furthermore, to better analyze the fine-grained and dynamic pitch changes of lexical tones which were nonlinear in nature, the second-order orthogonal polynomial models were adopted which is a multilevel regression technique designed for analyzing time course data (Mirman, 2014; Rattanasone et al., 2018; Tang et al., 2019). According to Mirman (2014), the polynomial function generates three "time terms": the intercept term (i.e., pitch height), the first-order linear term (i.e., pitch slope), and the second-order quadratic term (i.e., pitch curvature). These three terms capture not only the height and slope of pitch contours, but also the steepness of the quadratic curvature. More specifically, the positive linear trend means a rising pitch contour, whereas negative means a falling contour; a larger absolute value of the linear trend represents a steeper slope. The positive quadratic trend indicates a concave F0 contour and negative indicates a convex contour, with a larger absolute value of quadratic trend suggesting more curvy contours. In addition, for the acoustic comparison of the rising tonal pair in Cantonese (CT23 vs. CT25), they additionally show a covert contrast at the temporal distinction of "inflection point". The minimum F0 value appears slightly earlier in the high-rising CT25 compared to that in low-rising CT23 along the pitch contour (Mok et al., 2020). Thus, the positions of the inflection point were obtained by locating the lowest F0 point in the first two thirds of the rising pitch contour.

Statistical Analysis

First, linear mixed-effect models (LMMs) in R (R Core Team, 2014) were used to analyze the three acoustic measures of prosodic pitch pattern. The package of lme4 (Bates et al., 2014) was used to fit the LMMs. An advantage of LMM is that it is possible to fit models with large, unbalanced data, such as the production data by children with and without ASD. The visual inspection of Q-Q plots and plots of residuals revealed no obvious deviations from homoskedasticity after exclusion of extreme data by a model-based trimming. In each LMM of prosodic pitch analyses, the pitch mean/range/SD (in semitone) in speech/nonspeech condition was entered as the dependent variable, with *group* (ASD, TD), *voice gender* (female voice, male voice), *language* (Cantonese, Mandarin), and all possible interactions acting as fixed effects. When fitting the LMMs, factors of *intensity* and *duration* were involved as controlled covariates, which were centered to reduce multicollinearity; *participant* and *item* were included as random effects.

Second, the growth curve analysis (Mirman, 2014) in R was adopted to analyze the lexical tone and non-linguistic pitch production. The pitch contours (in log-scale 5-level value) measured over 16 normalized time points were modeled with a second-order orthogonal polynomial, with fixed effects of *group* (ASD, TD), *language* (Cantonese, Mandarin), and their interaction on all time terms. The model also included *participant* random effects on all time terms (intercept, linear, and quadratic terms). Besides, the centered *intensity* and *duration* were included as covariates. In speech condition, the second-order polynomial models were conducted for each tonal category (6 lexical tones) and each type of segment (familiar and unfamiliar) separately (12 models in total: 6 for the familiar segment and 6 for the unfamiliar segment). In nonspeech condition, the second-order polynomial models were conducted for each tonal category (6 models in total).

Third, for the analysis of inflection point, the generalized Poisson regression models (Consul & Famoye, 1992) were constructed in R, with *group* (ASD, TD), *language* (Cantonese, Mandarin), *tonal pair* (CT23, CT25), and all the possible interactions acting as fixed effects. The generalized Poisson regression model has been found useful in fitting the dependent variables of integer data. When fitting the regression models in speech or nonspeech condition, *participant* and *item* were included as random effects.

For all the generated LMMs, polynomial models, and Poisson regression models mentioned above, the random slopes and their intercepts for all the relevant fixed effects were included in the initial model to make it maximally generalizable across the data (Barr et al., 2013). The *p*-values of main effects and interaction effects were obtained using Satterthwaite's approximations in R package lmerTest (Kuznetsova et al., 2017). When a significant

main effect of a multilevel factor or a significant interaction effect was detected, post-hoc pairwise comparisons were performed using the lsmeans package (Lenth, 2016) with Tukey adjustment.

Results

Prosodic Pitch Pattern

Pitch Mean. In speech condition, the LMM on the pitch mean (in semitone) showed significant main effects of *group* [χ^2 (1) = 9.77, $p < .01$], and *voice gender* [χ^2 (1) = 27.90, $p < .001$], while the main effect of *language* [χ^2 (1) = 0.00, $p = .981$] and all interaction effects did not reach significance (all *p*s > .05). As shown in the left column of Figure 1a, the ASD group regardless of language backgrounds generally demonstrated a higher average pitch ($M = 16.5$) when imitating the lexical tones compared to the TD children ($M = 15.8$). Moreover, as expected, both ASD and TD children enhanced their mean pitch when imitating the speech models of female voice ($M = 16.5$) relative to male voice ($M = 15.9$). Similarly, the LMM on the pitch mean (in semitone) in nonspeech condition also revealed significant main effects of *group* [χ^2 (1) = 4.87, $p < .05$], and *voice gender* [χ^2 (1) = 56.37, $p < .001$]. That is to say, as presented in the left column of Figure 1b, children with ASD ($M = 16.7$) also tended to exhibit a higher pitch mean at non-linguistic pitch imitation than their neurotypical peers ($M = 16.1$). Moreover, child participants produced a relatively higher pitch mean when imitating the nonspeech stimuli containing the female pitch contours ($M = 17.0$) than the male pitch contours ($M = 15.8$).

Pitch Range. For the pitch range in speech condition, only a main effect of *group* [χ^2 (1) = 8.76, $p < .01$] was found. Neither the main effects of *voice gender* [χ^2 (1) = 0.22, $p = .637$], *language* [χ^2 (1) = 0.39, $p = .531$], nor any two-way or three-way interactions were significant (all *p*s > .05). The obtained pitch range (in semitone) of lexical tone imitations was 14.5 in the ASD group and 13.0 in the TD group (in the middle column of Figure 1a). The main effect of *group* suggested that both Mandarin- and Cantonese-speaking participants with ASD generally produced a wider pitch range in the imitation of lexical tones, which might reveal an exaggerated pitch in ASD. For the pitch range in nonspeech condition, all the main effects and interaction effects fell short of significance (all *p*s > .05). As displayed in the middle column of Figure 1b, children with ASD ($M = 11.0$) showed comparable pitch range (in semitone) as the TD controls ($M = 10.6$) when imitating the non-linguistic pitch contours in nonspeech condition.

Pitch SD. The LMM was performed on pitch SD in speech condition, and the statistical results showed significant main effects of *group* [χ^2 (1) = 4.42, $p < .05$], and *voice gender* [χ^2 (1) = 8.65, $p < .01$], while the main effect of *language* [χ^2 (1) = 0.01, $p = .908$] and all interaction effects did not reach statistical significance (all *p*s > .05). As illustrated in the right column of Figure 1a, there was a significant difference in pitch SD (in semitone) of lexical tone imitations between the two groups. Overall, children with ASD ($M = 2.68$) showed a significantly greater SD across F0 samples in speech condition than TD children ($M = 2.49$). Moreover, participants showed a larger pitch SD when imitating the CT models of female voice ($M = 2.66$) than male voice ($M = 2.50$). Next, in the nonspeech condition, only the main effect of *voice gender* was found to be marginally significant [χ^2 (1) = 2.97, $p = .085$]. The non-significant main effect of *group* [χ^2 (1) = 0.01, $p = .940$] suggested that, different from imitating lexical tones, children with ASD ($M = 2.58$) generated a pretty similar pitch SD compared to the TD children ($M = 2.57$) when imitating the non-linguistic pitch contours (in the right column of Figure 1b).

[Figure 1 about here]

Both Cantonese- and Mandarin-speaking children with ASD showed greater pitch variations when imitating lexical tones in speech condition, as indicated by an expanded pitch range and a larger pitch SD (Figure 1a). In order to further examine whether the pitch variations of lexical tone imitations were correlated with the language/verbal abilities in children with ASD per se, we conducted Spearman's correlation in Cantonese- and Mandarin-speaking with ASD respectively. For CASD (Figure 2a), a very strong positive correlation was found between pitch range and pitch SD ($r = .92$, $p < .001$) as expected. However, the language score of CASD was not correlated with pitch range ($r = .20$, $p = .330$), or pitch SD ($r = .09$, $p = .655$). In a similar manner, there was a positive correlation between pitch range and pitch SD in MASD ($r = .77$, $p < .001$). However, neither the correlation between verbal IQ and pitch range ($r = .25$, $p = .223$), nor that between verbal IQ and pitch SD ($r = .33$, $p = .104$) reached significance in MASD when imitating lexical tones (Figure 2b).

[Figure 2 about here]

Lexical Tone and Non-Linguistic Pitch Imitation

Figure 3 displays the pitch contours along 16 time points produced by four subgroups of child participants (CASD, CTD, MASD, MTD) when imitating six CTs and the non-linguistic pitch models. The seemingly overlapping pitch

contours across the four subgroups implied that all the child participants could generally produce the global pitch contours (Figure 3), consistent with high-level (CT55), mid-level (CT33), low-level (CT22), low-rising (CT23), high-rising (CT25), and low-falling (CT21) descriptions. However, if we zoomed in on the fine-grained pitch differences, all the pitch trajectories showed dynamic changes in terms of pitch height, pitch slope, and pitch curvature. Second-order orthogonal polynomial models were built for each tonal category. In the polynomial models, the intercept term, linear term (ot1), and quadratic term (ot2) capture the F0 contour's pitch height, pitch slope, and pitch curvature, respectively (Tables 3 & 4).

[Figure 3 about here]

Level Tones. Table 2 shows the statistical results of fixed effects on the pitch height, slope, and curvature when imitating three level CTs. First, for the imitation of high-level CT55 (Table 2a), there was only a significant effect of *language* on the linear term (pitch slope) in both speech and nonspeech conditions. Compared to Mandarin-speaking children, the Cantonese-speaking children tended to produce a relatively more falling F0 slope (Figure 3) when imitating the high-level CT55 in the familiar segment (β = -0.13, SE = 0.05, t = -2.62, p = .01) and unfamiliar segment (β = -0.21, SE = 0.05, t = -4.12, p < .001), as well as in nonspeech condition (β = -0.19, SE = 0.09, t = -2.14, p < .05). Then, for the imitation of both mid-level CT33 and low-level CT22, the results merely revealed significant main effect of *language* on the linear term (pitch slope) in the speech condition. Specifically, when imitating the mid-level CT33 (Table 2b), Cantonese-speaking children showed more falling F0 slope, with significant negative estimates relative to Mandarin-speaking children in the familiar segment (β = -0.24, SE = 0.05, t = -4.56, p < .001), as well as unfamiliar segment (β = -0.20, SE = 0.05, t = -4.09, p < .001). Also, the pitch trajectories of low-level CT22 imitations (Table 2c) tended to be more falling for Cantonese-speaking children in familiar segment (β = -0.21, SE = 0.06, t = -3.37, p = .001), as well as unfamiliar segment (β = -0.28, SE = 0.06, t = -5.04, p < .001). It should be noted that, when imitating the three level tones in both speech and nonspeech conditions, neither the main effect of *group* nor the interaction of *group* × *language* on the pitch height, slope, or curvature was found to be significant (Table 2).

[Table 2 about here]

Contour Tones. Table 3 shows the statistical results of fixed effects on the pitch height, slope, and curvature in the production of three contour tones. When imitating the low-rising CT23 (Table 3a) or low-falling CT21 (Table 3c), none of the fixed effects on the time terms reached significance in both speech and nonspeech conditions. The

results for high-rising CT25 showed a significant effect of *group* on the linear term (pitch slope) but only in speech condition (Table 3b). That was, when imitating the high-rising CT25, both Mandarin- and Cantonese-speaking children with ASD produced rising contours with shallower slopes than age-matched TD children in familiar segment (β = -0.29, SE = 0.12, t = -2.42, $p < .05$) and unfamiliar segment (β = -0.24, SE = 0.12, t = -2.01, $p < .05$). In addition, there was a significant negative effect of *group* on the quadratic term (pitch curvature) for children with ASD, suggesting that they produced a flatter F0 curve than TD children when imitating the high-rising CT25 in familiar segment (β = -0.13, SE = 0.07, t = -2.02, $p < .05$). All other fixed effects were not significant (see Table 3 for full results).

[Table 3 about here]

Inflection Point of CT23 vs. CT25. Additionally, we compared the inflection points of the rising minimal pair (low-rising CT23 vs. high-rising CT25) using the generalized Poisson regression model. In speech condition, the regression model on inflection point showed a significant main effect of *tonal pair* [χ^2 (1) = 52.49, $p < .001$], while the other main effects and interaction effects did not reach significance (all ps > .05). All the child participants, regardless of language backgrounds and clinical condition, produced an earlier inflection position when imitating high-rising CT25 (M = 4.52) than the low-rising CT23 (M = 5.41) in speech condition. Similarly, in nonspeech condition, only the significant main effect of *tonal pair* [χ^2 (1) = 12.41, p < .001] was found, with an earlier inflection position when imitating the nonspeech CT25 (M = 4.23) than CT23 (M = 5.10) for all the child participants. It should be noted that when imitating the two rising CTs in both speech and nonspeech conditions, neither the main effect of *group* nor its interaction effect on the inflection position was found to be significant (all ps > .05).

Experiment 2. Identification of Low-Rising CT23 vs. High-Rising CT25 Imitations

As shown in the acoustic of analyses of lexical tone and non-linguistic pitch imitation in Experiment 1, the group difference (ASD vs. TD) was only detected during the imitation of high-rising CT25 in speech condition (Table 3b). Specifically, both Mandarin- and Cantonese-speaking children with ASD produced a shallower pitch slope and a flatter F0 curve in the imitation of high-rising CT25 relative to TD children. The minimal pair of two rising tones in Cantonese phonology mainly differed in terms of pitch slope, with high-rising CT25 showing a much steeper slope

compared to the low-rising CT23. Based on the acoustic analyses, it is likely that Native Cantonese listeners might show a higher identification accuracy of high-rising CT25 imitations produced by TD children relative to those by children with ASD. However, given the nature of categorical perception of native speech sounds (Liberman et al., 1957; Xu et al., 2006), the shallower pitch slope of high-rising CT25 imitation in children with ASD did not necessarily entail identification difficulties for native speakers. To shed light on this issue, we further conducted an identification test (Experiment 2), by asking the native Cantonese-speaking adults to perceive and identify the minimal-pair imitations of two rising tones (CT23 vs. CT25). The perceptual analysis in Experiment 2 was performed to complement acoustic measurements in Experiment 1.

Methods

Participants

In total, 16 neurotypical undergraduate and graduate students in college (8 males; M_{age} = 24.6 yr, SD = 2.9) whose first language is Cantonese participated in the identification test. They were not majoring in linguistics or psychology, and had no reported speech, language, or hearing disorders. None of the participants had received formal musical training over one year. All participants gave informed consent in compliance with the protocols approved by the Research Ethics Committee of Hong Kong Polytechnic University, and they were paid for their travel and time.

Stimuli and Procedure

Totally there were 1,664 syllables produced by all the child participants (CASD, CTD, MASD, and MTD) through imitating the speech models of CT25 and CT23, and these lexical tone imitations were included as the perceptual stimuli. The stimuli were not normalized in terms of intensity and duration, in an effort to keep these perceived sounds unmodified. Instead, the duration and intensity values were included as covariates in the statistical analysis to control for confounding factors. The stimuli were presented using E-prime 2.0, and were divided into four testing blocks based on four different carrying segments (fu, ji, se, si). The four testing blocks were counterbalanced across participants. The perceptual stimuli were played in a random order within each testing block. Before the formal test, there was a practice block with the adult speech models of CT25 and CT23 included as the practice stimuli to familiarize participants with the identification procedure. The participants were asked to conduct a two-alternative forced choice (2AFC) identification task. After the presentation of each syllable, they would be asked to identify the target syllables

as Cantonese characters “婦” (CT23) or “苦” (CT25) in the block of “fu”; as “耳” (CT23) or “倚” (CT25) in the block of “ji”; as “社” (CT23) or “寫” (CT25) in the block of “se”; as “市” (CT23) or “史” (CT25) in the block of “si” by pressing corresponding keyboard buttons. The participants were allowed to play the target syllable repeatedly until they were confident to make a judgement. The whole identification test, including the practice block, lasted approximately 1.5 h for each participant.

Statistical Analysis

To analyze the identification accuracy, a generalized linear mixed-effects model (GLMM) was created in R using the lme4 package (Bates et al., 2014). For the construction of GLMM, the dichotomous response to each stimulus (“1” meaning correct response or “0” indicating incorrect response) was entered as the dependent measure, with *group* (ASD, TD), *language* (Cantonese, Mandarin)*, segment* (familiar, unfamiliar), *tonal pair* (CT23, CT25), and all their possible interactions acting as fixed effects. When fitting GLMM, *participant* and *item* were included as random effects. Moreover, the centered duration and intensity values for each stimulus were included as the controlled covariates. The other methods for GLMM calculation were consistent with LMM as shown in Experiment 1.

Results

Figure 4 shows box plots of the identification accuracy (%) across different conditions. GLMM, on identification accuracy, revealed a significant three-way interaction of *group* × *language* × *tonal pair* [χ^2 (1) = 13.40, $p < .001$], as well as a significant three-way interaction of *group* × *segment* × *tonal pair* [χ^2 (1) = 8.32, $p < .01$]. First, post-hoc pairwise comparisons for the interaction of *group* × *language* × *tonal pair* showed that Cantonese speakers’ identification accuracy was similar in the perception of high-rising CT25 imitations produced by CASD and by CTD (β = -0.12, SE = 0.07, t = -1.65, p = .099), as well as those produced by MASD and by MTD (β = -0.01, SE = 0.07, t = -0.12, p = .904). Moreover, the identification accuracy was similar in the perception of low-rising CT23 imitations produced by CASD and by CTD (β = -0.02, SE = 0.07, t = -0.25, p = .804), whereas the accuracy was much higher in the perception of CT23 imitations produced by the MTD than those by MASD (β = -0.28, SE = 0.07, t = -3.89, $p < .001$). Then, post-hoc pairwise comparisons for the three-way interaction of *group* × *segment* × *tonal pair* showed that native speakers’ identification accuracy of CT23 imitations produced by TD children was significantly higher compared to those produced by children with ASD when the carrying segment was familiar (β = -0.19, SE = 0.08, t =

-2.38, $p < .05$). When the segment was unfamiliar, the identification accuracy was similar ($\beta = -0.10$, SE = 0.08, $t = -1.22$, $p = .224$).

[Figure 4 about here]

Discussion

The abnormalities of prosodic pitch production have been noted since the earliest report of ASD (Kanner, 1943), but our full understandings of the language-specific features and the underlying mechanisms are currently inconclusive. The previous conclusions on prosodic and suprasegmental features in ASD have primarily been derived from non-tonal language speakers. Importantly, the changes in pitch also play a crucial role in distinguishing phonological contrasts and word meanings at the syllable level for tonal language speakers. This study adopted an imitation task to investigate the prosodic pitch pattern and lexical tone production in tone-language-speaking children with ASD. The major findings and relevant discussions were shown in the following parts.

Atypical Prosodic Pitch Pattern in Tone-Language-Speaking Children with ASD

The prosodic pitch pattern was investigated with pitch mean/range/SD in the imitation of speech syllables and nonspeech sounds. When imitating speech models, both Cantonese- and Mandarin-speaking children with ASD showed a higher pitch mean, a larger pitch range, as well as a greater pitch SD than peers with TD (Figure 1a). When imitating nonspeech models, children with ASD only produced a higher pitch compared to the TD participants (Figure 1b). The group differences of the increased pitch of intonation in speakers with ASD have been found in some studies (Chan & To, 2016; Edelson et al., 2007; Sharda et al., 2010) but not others (Diehl et al., 2009; Nadig & Shaw, 2012). Most of the previous studies that employed acoustic measurements, focused on the pitch range and/or pitch SD of intonation at the sentence level. Contrary to the traditional stereotype of monotonic intonation in autism, children with ASD generally showed a significantly larger pitch range and/or pitch SD compared to TD children, indicating increased pitch variations of intonation in the ASD group (Bonneh et al., 2011; Chan & To, 2016; Diehl et al., 2009; Filipe et al., 2014; Fosnot & Jun, 1999; Green & Tobin, 2009; Hubbard & Trauner, 2007; Nadig & Shaw, 2012; Sharda et al., 2010). These previous studies only compared the F0 differences between ASD and TD groups, while neglected the influence from other prosodic features such as intensity and duration. Actually, the pitch-related parameters almost

always involve concomitant variations in other prosodic features (Xu & Prom-on, 2019). After controlling for intensity and duration, the current findings corroborated the notion of increased pitch variations in ASD with the empirical evidence from a smaller prosodic unit at the syllable level. Especially, for individuals with autism who speak a tonal language, the atypical prosodic feature of increased F0 variations emerged broadly, not only on the larger prosodic unit of intonation (Chan & To, 2016), but also on the smaller unit of lexical tone (this study).

The conclusion of increased F0 variation in ASD as a prominent feature of prosody could be reached with high reliability and generalizability, since the same pattern was found consistently across various studies in low-functioning (Baltaxe, 1984; Fosnot & Jun, 1999) and high-functioning (this study; Chan & To, 2016; Diehl et al., 2009; Filipe et al., 2014; Green & Tobin, 2009; Nadig & Shaw, 2012) children with ASD; from tonal (this study; Chan & To, 2016) and non-tonal (Bonneh et al., 2011; Diehl et al., 2009; Filipe et al., 2014; Fosnot & Jun, 1999; Green & Tobin, 2009; Hubbard & Trauner, 2007; Nadig & Shaw, 2012; Sharda et al., 2010) language backgrounds; from wide age ranges in children (this study; Bonneh et al., 2011; Filipe et al., 2014; Fosnot & Jun, 1999; Green & Tobin, 2009; Nadig & Shaw, 2012; Sharda et al., 2010), adolescents (Diehl et al., 2009), and even adults (Chan & To, 2016); from analyses in both spontaneous (Bonneh et al., 2011; Chan & To, 2016; Diehl et al., 2009; Filipe et al., 2014; Green & Tobin, 2009; Nadig & Shaw, 2012; Sharda et al., 2010) and imitation data (this study; Fosnot & Jun, 1999; Hubbard & Trauner, 2007). There was a concern about the influence of overall language/cognitive functioning on the prosodic abnormalities, since children with specific language impairment (Goffman, 1999; Marshall et al., 2009) or major cognitive delays (Shriberg & Widder, 1990) also revealed prosodic deficits. In this study, the ASD and TD groups were matched in terms of nonverbal IQ, although the general language functioning in ASD slightly lagged behind TD children. We additionally performed correlation analysis between the general language functioning and pitch range/SD in CASD and MASD respectively (Figure 2), but no significant correlations were found from our study samples. Furthermore, even in studies with matched comparison groups on variables of both IQ and general language functioning (Diehl et al., 2009; Nadig & Shaw, 2012), the pattern of increased pitch variation in speakers with ASD was observed as well. Thus, the prosodic pitch differences produced by ASD and TD children tended to be specific to prosody, rather than an artifact of more general language and/or intellectual functioning. Such findings highlight the presence of prosodic pitch atypicalities even in very high-functioning and linguistically developed individuals with ASD, which could be a stigmatizing barrier to communication competence and social acceptance for speakers with ASD who evidence prosodic oddities.

What are the underlying mechanisms responsible for the atypical prosodic pitch pattern in individuals with ASD confirmed in the current and multiple other studies? One study proposed that the phenomenon of more variable prosody might be caused by a delay in developmental trajectory of speech in ASD (Sharda et al., 2010). The observations of increased pitch range/SD and pitch mean could also be discovered in speech directed to infants commonly known as "motherese", which had distinct prosodic patterns characterized by heightened pitch and exaggerated pitch contours (Segal & Newman, 2015). Early intonation features of younger TD children under 2 years also mimicked motherese-like features, but diminished gradually after 2–3 years of age (Eguchi, 1969). Thus, the increase in pitch variability in speakers with ASD might reflect prolonged mimicry of the prosodic pitch patterns of child-directed speech in this group, relative to TD children (Sharda et al., 2010). Others have labeled the atypical prosodic pitch pattern in ASD as aberrant rather than delayed development of speech prosody in ASD (Rapin & Dunn, 2003). This perspective could be supported by observing such an atypical pattern in adolescents and even in adults with ASD (Chan & To, 2016; Diehl et al., 2009). A possible explanation for the increased pitch variability in the ASD group was a disruption in the basic pitch-controlling speech production mechanisms, which could stem from a deficit at the production level (Bonneh et al., 2011), or reflected speech compensation for auditory feedback perturbations to overcome a noisy channel supposed to transmit "efference copy" information (Houde et al., 2007). More studies are needed to uncover the nature of abnormal suprasegmental aspects of speech production, or prosody in speakers with ASD.

Preserved Lexical Tone Imitation Skill in Cognitively Able Children with ASD

In this study, the complex CTs with changes in both pitch height and slope (Gandour, 1981) were imitated by both Cantonese- and Mandarin-speaking children with and without ASD. As shown in Figure 3, it was found that all the child participants at school age could generally imitate the global tone contours of six CTs, which was important for maintaining tonal category distinctions. The growth curve analysis indicated that the ASD group only produced a shallower slope and/or a flatter F0 curve in the production of high-rising CT25 relative to TD children. However, such fine-grained and within-category acoustic differences did not cause difficulties in native speakers' perceptual judgement, as evidenced by a comparable accuracy of identifying the CT25 imitation produced by ASD and TD children (Figure 4). Moreover, even for the imitation of the covert contrast of the inflection points of CT23 vs. CT25, which was a reliable acoustic difference not perceivable by naïve speakers (Edwards & Beckman, 2008), all the child

participants, including the MASD, correctly produced an earlier inflection point for CT25 than CT23. Our current findings offered strong supports to the notion that the echoed speech by children with ASD could imitate complex tonal contours from the adult models accurately in a preserved manner (Fan et al., 2010; Frankel et al., 1987). Such imitation skills seemed to be unaffected by the linguistic status of children with ASD, as a lack of interaction effect of *group* × *language* in all the acoustic analyses. That is, even the MASD could largely imitate pitch contours of unfamiliar/familiar lexical tones as well as CASD. It appeared to be the case that children with ASD adopted a bottom-up mechanism when imitating the pitch contours at the syllable level, and they were largely intact at the processing of local pitch information, as suggested by the enhanced perceptual functioning in autism (Mottron & Burack, 2001).

However, we should be prudent in generalizing the current finding of preserved lexical tone imitation skill to each individual on the autistic spectrum especially those with intellectual disabilities or severe language delays. On the one hand, this study adopted a relatively simple task, which asked participants to imitate each lexical tone in isolation at the syllable level. Such a task might obscure group differences that would be present in more difficult tasks such as lexical tone imitation in connected speech. On the other hand, the participants with ASD in this study belonged to the subgroup without major cognitive delay/severe language delay. Since there could be strong relationship between immediate imitation skill and language ability (Rogers et al., 2008; Toth et al., 2006), it remains unclear whether low-functioning children with ASD with severe language delays would be able to imitate pitch contours of lexical tones that were acoustically comparable to those imitations in TD children.

Speech-Specific and Contour-Biased Lexical Tone Processing Atypicalities in ASD

We have observed two speech-specific phenomena from the current imitation data. First, compared to TD peers, children with ASD showed increased pitch range/SD of lexical tone imitations, while they exhibited similar pitch range/SD when imitating the nonspeech sounds. Second, children with ASD showed some deviations from the TD children in coordinating pitch slope and curvature when imitating the high-rising CT25 superimposed on speech segments, but the two groups did not differ from each other when imitating the same pitch contour of CT25 embedded in nonspeech materials. The speech-specific imitation atypicality in ASD lent further support to the notion that children with ASD showed domain-specific pitch processing difficulties. Specifically, Mandarin-speaking individuals with ASD showed atypical or impaired processing of lexical tone (Lau et al., 2020; Wang et al., 2017; Yu et al., 2015) and intonation (statements vs. questions, Jiang et al., 2015), whereas they showed normal or even enhanced processing of

the same pitch information in the domains of music and nonspeech. In line with extant findings in pitch processing, more and more research proposed a speech-specific viewpoint that speech and language learners with autism failed to engage or develop specialized networks for vocal processing and phonetic learning in speech sounds (Haesen et al., 2011; Kujala et al., 2013; Lindell & Hudry, 2013; O'connor, 2012).

In addition, there are two types of lexical tones in general: contour tones and level tones. Contour tones change their pitch height and direction apparently over time, whereas level tones remain at approximately a steady pitch (Yip, 2002). One recent study by Cheng et al. (2017) investigated the ability to discriminate "level tones" embedded in real syllable, pseudo-syllable, and non-speech in Cantonese-speaking individuals with ASD. However, no group differences (ASD vs. TD) were found across all three conditions. It seemed that the speech-specific lexical tone processing difficulties in ASD tended to be biased towards the processing of contour tones (high-rising tone vs. high-falling tone, Wang et al., 2017; Yu et al., 2015), while less apparent in the processing of level tones (Cheng et al., 2017). However, the conclusion was far from clear in literature as none of the previous studies incorporated both level and contour lexical tones in one single study. In this study, children with and without ASD were asked to imitate both Cantonese level and contour tones. When imitating three level tones (CT55, CT33, CT22), both Cantonese- and Mandarin-speaking children with ASD produced a comparable pitch height, slope, and curvature relative to TD children (Table 2). Only a main effect of *language* on the linear term (pitch slope) of level tones was found, with Cantonese-speaking children eliciting a more falling F0 slope compared to Mandarin-speaking ones. It was reported that Cantonese speakers tended to produce a slightly falling contour in their actual realization of three level tones, especially for the low-level and mid-level tones (Mok et al., 2020; Zhang et al., 2018). Thus, the cross-linguistic differences in the acoustic realization of level tones could be attributed to the influence of long-term native language experience. Then, when imitating three contour tones (CT23, CT25, CT21), the two groups did not differ in the acoustic realizations of low-falling CT21, and low-rising CT23, whereas differed in terms of pitch slope and curvature of high-rising CT25. Although subsequent identification test proved that such fine-grained acoustic differences did not lead to perceptual ambiguity, children with ASD nevertheless exhibited some difficulties in coordinating exactly the same acoustic pitch trajectory of the more dynamic and fast-changing contour tones with a steeper slope (i.e., CT25 in this study). To conclude, the current findings pointed to a speech-specific and contour-biased lexical tone processing atypicalities in individuals with ASD.

Top-Down Phonological Processing Deficits in Children with ASD

It was proposed that pitch processing capacity was intact or even superior at the bottom-up acoustic processing level but impaired due to a top-down phonological processing deficit in individuals with ASD (Jiang et al., 2015; Wang et al., 2017; Yu, 2018). The hypothesis was that in tonal models tapping the phonological processing abilities of the child participants, comparatively inferior imitation performance could arise from either the lack or the impairment of relevant phonological representations (Kuhl, 2011). As mentioned earlier, the Mandarin-speaking children with ASD in our study could imitate the global tone contours of both familiar and unfamiliar tonal categories with similar acoustic performances to TD children. It seemed that acoustic pitch realizations of lexical tone imitation persisted independently of speech familiarity, and were not influenced by the phonological status of the tonal categories or the linguistic status of the carrying syllables in children with ASD. However, as shown in Figure 4, Cantonese-speaking adult listeners showed higher identification accuracy of CT23 imitations with familiar segment produced by MTD than imitations produced by MASD. Actually, the MTD and MASD produced similar F0 realizations of CT23 with familiar segment, which meant that different perceptual accuracy was affected by some other factors beyond F0 (the primary correlate of lexical tones). This is not surprising since several secondary cues, such as intensity profile, duration, and voice quality, also contribute to lexical tone perception (Zhang et al., 2012). In our statistical model, the duration and intensity have been entered as covariates. Thus, when imitating the non-native and unfamiliar CT23, MTD could utilize the phonological knowledge of native and familiar segment (/fu/ or /ji/) to produce a better voice quality of that syllable. However, in contrast, MASD failed to exploit such top-down phonological knowledge to compensate for the imitation of syllables with non-native tonal category. That is, MASD demonstrated compromised performance relative to MTD, when imitating unfamiliar Cantonese tonal stimuli superimposed on familiar segments (/fu/ and /ji/), but comparatively normal performance when on unfamiliar segments (/si/ and /sɛ/). These findings implied that the lexical tone processing difficulties in speech condition (Wang et al., 2017; Wu et al., 2020; Yu et al., 2015) reported in some children with ASD were caused by a phonological processing deficit rather than the acoustic pitch processing deficit.

Clinical and Theoretical Implications

Languages of the world exhibit a natural diversity, and around 60–70% of the world's languages are tonal (Yip, 2002), which is not reflected in the mainstream autism research on the pitch and general prosodic processing skills. The

current study firstly investigated the prosodic pitch pattern and lexical tone imitation skill in tone-language-speaking children with ASD in a cross-linguistic context. The findings have important implications for ASD assessment and remediation to examine the mechanisms of prosodic atypicalities in tone-language-speaking children with ASD. On the one hand, we found increased prosodic pitch variations in tone-language-speaking children with ASD, which was consistent with previous findings in children with ASD from non-tonal language backgrounds. The accumulating evidence of atypical prosodic pitch pattern contributes to the possibility of developing pitch-based measures as one of the behavioral biomarkers for ASD, which are both quantitative and objective (Bonneh et al., 2011). On the other hand, there is a trend of distinct patterns across different language speakers with ASD (tonal language vs. non-tonal language), pointing to a language-specific pitch processing pattern in ASD. Specifically, previous research that documented superior pitch processing skills among individuals with ASD (such as autistic musical savants) was disproportionately drawn from non-tonal language speakers with ASD. However, lexical tone processing is a complex process that involves an interface between pitch and meaning and incorporates both acoustic and phonological processing. Several recently conducted studies (Lau et al., 2020; Wang et al., 2017; Yu et al., 2015) pointed to lexical tone perception difficulties in tone-language-speaking individuals with ASD. The related findings in this study further lent support to the notion of lexical tone processing difficulties in ASD, with extended evidence from two different tonal language systems and from an imitation task. Furthermore, our current observations firstly revealed the nature of such processing difficulties, which occurred at the top-down phonological processing level, rather than at the bottom-up acoustic processing level. To conclude, the current findings help clarify speech-specific and language-specific auditory pitch processing atypicalities in children with ASD from different tonal language backgrounds, which are not only theoretically interesting, but also clinically relevant.

Limitations and Future Research

Our study has several limitations that we must note. First, in this cross-linguistic study, the nonverbal IQ and general language functioning among Cantonese- and Mandarin-speaking participants were evaluated with different testing materials. Our initial concern was to ensure the ASD group to be matched with TD group in terms of nonverbal IQ in children from each language background. But without unified measurement among all the child participants, we could not yet include these factors as covariates in the statistical models. Unfortunately, there was no standard oral language assessment scale thus far applicable to both Cantonese- and Mandarin-speaking children. Second, perceptual ability

in child participants might be tested in future studies to investigate whether the perceptual ability and vocal imitation performance in ASD are closely related, or to some extent distinct, since there might be distinct representations used to support speech imitation and perception tasks (Hutchins & Peretz, 2012). Third, the Cantonese native speakers' identification accuracy of CT23 and CT25 imitations in Experiment 2 was surprisingly low. On the one hand, we did not control for the possibility that some native speakers in Experiment 2 might merge the two rising tones in Cantonese (Fung & Lee, 2019; Mok et al., 2013). On the other hand, in the current identification study, we adopted a blocked-segment design that contained the imitation stimuli produced from different talkers in one single block. Listeners might have struggled to estimate the upper and lower F0 boundaries of a particular voice within a block, thus were unable to map each rising pitch stimulus to the corresponding tone category with reference to its relative position in that talker's F0 range. Future identification studies could, for example, present stimuli through a blocked-talker design and exclude the native adult participants who merge the two rising tones. Furthermore, given that the imitation task adopted in this study was simple, which could be reliably performed in younger and low-functioning children with ASD, future study could test the imitation abilities in low-functioning preschoolers with ASD who show intellectual disability/severe language delay. It would be meaningful to see how the prosodic pitch pattern and lexical tone imitation skill change among different subgroups of the autistic spectrum and among different age groups in future studies. Finally, in contrast with the imitation task adopted in the current study, investigations into lexical tone processing skills in a more natural setting, such as spontaneous speech samples in daily life, would be an important next step.

References

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

Baltaxe, C. A. (1984). Use of contrastive stress in normal, aphasic, and autistic children. *Journal of Speech and Hearing Research*, *27*(1), 97–105. https://doi.org/10.1044/jshr.2701.97

Baltaxe, C. A. M., & D'Angiola, N. (1992). Cohesion in the discourse interaction of autistic, specifically language-impaired, and normal children. *Journal of Autism and Developmental Disorders*, *22*(1), 1–21. https://doi.org/10.1007/BF01046399

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

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

Bolinger, D. L. (1972). *Intonation*. Harmondsworth: Penguin.

Bonneh, Y. S., Levanon, Y., Dean-Pardo, O., Lossos, L., & Adini, Y. (2011). Abnormal speech spectrum and increased pitch variability in young autistic children. *Frontiers in Human Neuroscience*, *4*, 237. https://doi.org/10.3389/fnhum.2010.00237

Chan, K. K. L., & To, C. K. S. (2016). Do individuals with high-functioning autism who speak a tone language show intonation deficits? *Journal of Autism and Developmental Disorders*, *46*(5), 1784–1792. https://doi.org/10.1007/s10803-016-2709-5

Chao, Y. R. (1930). A system of tone letters. *Le Maiître Phonétique*, *45*, 24–27.

Chao, Y. R. (1968). *A grammar of spoken Chinese*. Berkeley, CA: University of California 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

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

Consul, P. C., & Famoye, F. (1992). Generalized poisson regression model. *Communications in Statistics - Theory and Methods*, *21*(1), 89–109. https://doi.org/10.1080/03610929208830766

Cutler, A., & Isard, S. D. (1980). The production of prosody. In *Language production* (Vol. 1, pp. 245–269). London: Academic Press.

Diehl, J. J., & Paul, R. (2012). Acoustic differences in the imitation of prosodic patterns in children with autism spectrum disorders. *Research in Autism Spectrum Disorders*, 6(1), 123–134. https://doi.org/10.1016/j.rasd.2011.03.012

Diehl, J. J., Watson, D., Bennetto, L., Mcdonough, J., & Gunlogson, C. (2009). An acoustic analysis of prosody in high-functioning autism. *Applied Psycholinguistics*, *30*(3), 385–404. https://doi.org/10.1017/S0142716409090201

Edelson, L., Grossman, R., & Tager-Flusberg, H. (2007). Emotional prosody in children and adolescents with autism. *Poster Session Presented at the Annual International Meeting for Autism Research*. Poster session presented at the annual international meeting for Autism Research, Seattle, WA.

Edwards, J., & Beckman, M. E. (2008). Methodological questions in studying consonant acquisition. *Clinical Linguistics & Phonetics*, *22*(12), 937–956. https://doi.org/10.1080/02699200802330223

Eguchi, S. (1969). Development of speech sounds in children. *Acta Otolaryngol*, *257*, 1–51.

Ehrler, D. J., & McGhee, R. L. (2008). *PTONI: Primary test of nonverbal intelligence*. Austin, TX: Pro-Ed.

Fan, Y. T., Decety, J., Yang, C. Y., Liu, J. L., & Cheng, Y. (2010). Unbroken mirror neurons in autism spectrum disorders. *Journal of Child Psychology and Psychiatry*, 51(9), 981-988.

Fay, W., & Schuler, A. L. (1980). *Emerging language in autistic children*. Baltimore, MD: University Park Press.

Filipe, M. G., Frota, S., Castro, S. L., & Vicente, S. G. (2014). Atypical prosody in Asperger Syndrome: Perceptual and acoustic measurements. *Journal of Autism and Developmental Disorders*, *44*(8), 1972–1981. https://doi.org/10.1007/s10803-014-2073-2

Flege, J. E. (1995). Second language speech learning: Theory, findings, and problems. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 233–277). Timonium, MD: York Press.

Flege, J. E. (2007). Language contact in bilingualism: Phonetic system interactions. In J. Cole & J. I. Hualde (Eds.), *Laboratory Phonology 9* (pp. 353–382). Berlin: Walter de Gruyter.

Fosnot, S. M., & Jun, S. (1999). *Prosodic characteristics in children with stuttering or autism during reading and imitation*. Paper presented at the 14th international congress of phonetic sciences.

Frankel, F., Simmons, J. Q., & Richey, V. E. (1987). Reward value of prosodic features of language for autistic, mentally retarded, and normal children. *Journal of Autism and Developmental Disorders*, *17*(1), 103–113. https://doi.org/10.1007/BF01487263

Fung, R. S. Y., & Lee, C. K. C. (2019). Tone mergers in Hong Kong Cantonese: An asymmetry of production and perception. *The Journal of the Acoustical Society of America*, *146*(5), EL424–EL430. https://doi.org/10.1121/1.5133661

Gandour, J. (1981). Perceptual dimensions of tone: Evidence from Cantonese. *Journal of Chinese Linguistics*, *9*(1), 20–36. JSTOR.

Gandour, J. (1983). Tone perception in Far Eastern languages. *Journal of Phonetics*, *11*(2), 149–175. https://doi.org/10.1016/S0095-4470(19)30813-7

Goffman, L. (1999). Prosodic influences on speech production in children with specific language impairment and speech deficits: Kinematic, acoustic, and transcription evidence. *Journal of Speech, Language, and Hearing Research: JSLHR*, *42*(6), 1499–1517. https://doi.org/10.1044/jslhr.4206.1499

Green, H., & Tobin, Y. (2009). Prosodic analysis is difficult … but worth it: A study in high functioning autism. *International Journal of Speech-Language Pathology*, *11*(4), 308–315. https://doi.org/10.1080/17549500903003060

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

Houde, J., Nagarajan, S., & Heinks‐Maldonado, T. (2007). Dynamic cortical imaging of speech compensation for auditory feedback perturbations. *The Journal of the Acoustical Society of America*, *121*(5), 3045–3045. https://doi.org/10.1121/1.4781744

Hubbard, K., & Trauner, D. A. (2007). Intonation and emotion in autistic spectrum disorders. *Journal of Psycholinguistic Research*, *36*(2), 159–173. https://doi.org/10.1007/s10936-006-9037-4

Hutchins, S., & Peretz, I. (2012). Amusics can imitate what they cannot discriminate. *Brain and Language*, *123*(3), 234–239. https://doi.org/10.1016/j.bandl.2012.09.011

Ingersoll, B. (2008). The social role of imitation in autism: Implications for the treatment of imitation deficits. *Infants & Young Children*, *21*(2), 107–119. https://doi.org/10.1097/01.IYC.0000314482.24087.14

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.

Kanner, Leo. (1971). Follow-up study of eleven autistic children originally reported in 1943. *Journal of Autism and Childhood Schizophrenia*, *1*(2), 119–145. https://doi.org/10.1007/BF01537953

Kim, D., & Clayards, M. (2019). Individual differences in the link between perception and production and the mechanisms of phonetic imitation. *Language, Cognition and Neuroscience*, 34(6), 769-786.

Kuhl, P. K. (2011). Who's Talking? *Science*, *333*(6042), 529–530. https://doi.org/10.1126/science.1210277

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

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

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

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

Lewis, M. M. (1957). *How Children Learn to Speak*. London: Harrap.

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

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

Liu, F., & Xu, Y. (2005). Parallel encoding of focus and interrogative meaning in Mandarin intonation. *Phonetica*, *62*(2–4), 70–87. https://doi.org/10.1159/000090090

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.

Loveall, S. J., Hawthorne, K., & Gaines, M. (2021). A meta-analysis of prosody in autism, Williams syndrome, and Down syndrome. *Journal of Communication Disorders*, 89, 106055. https://doi.org/10.1016/j.jcomdis.2020.106055

Ma, J. K.-Y., Ciocca, V., & Whitehill, T. L. (2006). Effect of intonation on Cantonese lexical tones. *The Journal of the Acoustical Society of America*, *120*(6), 3978–3987. https://doi.org/10.1121/1.2363927

Major, R. C., & Kim, E. (1996). The similarity differential rate hypothesis. *Language Learning*, *46*(3), 465–496. https://doi.org/10.1111/j.1467-1770.1996.tb01244.x

Marshall, C. R., Harcourt-Brown, S., Ramus, F., & van der Lely, H. K. J. (2009). The link between prosody and language skills in children with specific language impairment (SLI) and/or dyslexia. *International Journal of Language & Communication Disorders*, *44*(4), 466–488. https://doi.org/10.1080/13682820802591643

McCann, J., & Peppé, S. (2003). Prosody in autism spectrum disorders: A critical review. *International Journal of Language & Communication Disorders*, *38*(4), 325–350. https://doi.org/10.1080/1368282031000154204

Messum, P. R. (2008). *The role of imitation in learning to pronounce* [Doctoral thesis]. University of London.

Mirman, D. (2014). *Growth curve analysis and visualization using R*. FL: Chapman & Hall/CRC.

Mok, P. P. K., Fung, H. S. H., & Li, V. G. (2019). Assessing the link between perception and production in Cantonese tone acquisition. *Journal of Speech, Language, and Hearing Research*, *62*(5), 1243–1257. https://doi.org/10.1044/2018_JSLHR-S-17-0430

Mok, P. P. K., Li, V. G., & Fung, H. S. H. (2020). Development of phonetic contrasts in Cantonese tone acquisition. *Journal of Speech, Language, and Hearing Research*, *63*(1), 95–108. https://doi.org/10.1044/2019_JSLHR-19-00152

Mok, P. P. K., Zuo, D., & Wong, P. W. Y. (2013). Production and perception of a sound change in progress: Tone merging in Hong Kong Cantonese. *Language Variation and Change*, *25*(3), 341–370. https://doi.org/10.1017/S0954394513000161

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.

Nadig, A., & Shaw, H. (2012). Acoustic and perceptual measurement of expressive prosody in high-functioning autism: Increased pitch range and what it means to listeners. *Journal of Autism and Developmental Disorders*, *42*(4), 499–511. https://doi.org/10.1007/s10803-011-1264-3

Nakai, Y., Takashima, R., Takiguchi, T., & Takada, S. (2014). Speech intonation in children with autism spectrum disorder. *Brain and Development*, *36*(6), 516–522. https://doi.org/10.1016/j.braindev.2013.07.006

Nguyen, N., & Delvaux, V. (2015). Role of imitation in the emergence of phonological systems. *Journal of Phonetics*, 53, 46-54.

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

Paul, R., Augustyn, A., Klin, A., & Volkmar, F. R. (2005a). Perception and production of prosody by speakers with autism spectrum disorders. *Journal of Autism and Developmental Disorders*, 35(2), 205–220. https://doi.org/10.1007/s10803-004-1999-1

Paul, R., Shriberg, L. D., McSweeny, J., Cicchetti, D., Klin, A., & Volkmar, F. (2005b). Brief report: Relations between prosodic performance and communication and socialization ratings in high functioning speakers with autism spectrum disorders. *Journal of Autism and Developmental Disorders*, *35*(6), 861–869. https://doi.org/10.1007/s10803-005-0031-8

Peng, G. (2006). Temporal and tonal aspects of Chinese syllables: A corpus-based comparative study of Mandarin and Cantonese. *Journal of Chinese Linguistics*, *34*(1), 134–154.

Peng, G., & Wang, W. S. Y. (2005). Tone recognition of continuous Cantonese speech based on support vector machines. *Speech Communication*, *45*(1), 49–62. https://doi.org/10.1016/j.specom.2004.09.004

Peppé, S., Cleland, J., Gibbon, F., O'Hare, A., & Castilla, P. M. (2011). Expressive prosody in children with autism spectrum conditions. *Journal of Neurolinguistics*, 24(1), 41–53. https://doi.org/10.1016/j.jneuroling.2010.07.005

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

Rapin, I., & Dunn, M. (2003). Update on the language disorders of individuals on the autistic spectrum. *Brain and Development*, *25*(3), 166–172. https://doi.org/10.1016/S0387-7604(02)00191-2

Rattanasone, N. X., Tang, P., Yuen, I., Gao, L., & Demuth, K. (2018). Five-year-olds' acoustic realization of mandarin tone sandhi and lexical tones in context are not yet fully adult-like. *Frontiers in Psychology*, *9*, 1–10.

Rogers, S. J., Young, G. S., Cook, I., Giolzetti, A., & Ozonoff, S. (2008). Deferred and immediate imitation in regressive and early onset autism. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, *49*(4), 449–457. https://doi.org/10.1111/j.1469-7610.2007.01866.x

Rutter, M., Le Couteur, A., & Lord, C. (2003). *Autism Diagnostic Interview—Revised*. Los Angeles: Western Psychological Services.

Segal, J., & Newman, R. S. (2015). Infant preferences for structural and prosodic properties of infant-directed speech in the second year of life. *Infancy*, *20*(3), 339–351. https://doi.org/10.1111/infa.12077

Sharda, M., Subhadra, T. P., Sahay, S., Nagaraja, C., Singh, L., Mishra, R., Sen, A., Singhal, N., Erickson, D., & Singh, N. C. (2010). Sounds of melody—Pitch patterns of speech in autism. *Neuroscience Letters*, *478*(1), 42–45. https://doi.org/10.1016/j.neulet.2010.04.066

Shriberg, L. D., Paul, R., McSweeny, J. L., Klin, A., Cohen, D. J., & Volkmar, F. R. (2001). Speech and prosody characteristics of adolescents and adults with high-functioning autism and Asperger syndrome. *Journal of Speech, Language, and Hearing Research*, 44(5), 1097–1115.

Shriberg, L. D., & Widder, C. J. (1990). Speech and prosody characteristics of adults with mental retardation. *Journal of Speech, Language, and Hearing Research*, *33*(4), 627–653. https://doi.org/10.1044/jshr.3304.627

So, C. K., & Best, C. T. (2010). Cross-language perception of non-native tonal contrasts: Effects of native phonological and phonetic influences. *Language and Speech*, *53*(2), 273–293. https://doi.org/10.1177/0023830909357156

So, C. K., & Best, C. T. (2014). Phonetic influences on English and French listeners' assimilation of Mandarin tones to native prosodic categories. *Studies in Second Language Acquisition*, *36*(2), 195–221. https://doi.org/10.1017/S0272263114000047

Tager-Flusberg, H. (1981). On the nature of linguistic functioning in early infantile autism. *Journal of Autism and Developmental Disorders*, *11*(1), 45–56. https://doi.org/10.1007/BF01531340

Tang, P., Yuen, I., Rattanasone, N. X., Gao, L., & Demuth, K. (2019). The acquisition of Mandarin tonal processes by children with cochlear implants. *Journal of Speech, Language, and Hearing Research*, *62*(5), 1309–1325. https://doi.org/10.1044/2018_JSLHR-S-18-0304

Toth, K., Munson, J., N. Meltzoff, A., & Dawson, G. (2006). Early predictors of communication development in young children with autism spectrum disorder: Joint attention, imitation, and toy play. *Journal of Autism and Developmental Disorders*, *36*(8), 993–1005. https://doi.org/10.1007/s10803-006-0137-7

T'sou, B., Lee, T., Tung, P., Chan, A., Man, Y., & To, C. K. S. (2006). *Hong Kong Cantonese Oral Language Assessment Scale*. Hong Kong: Language Information Sciences Research Centre.

Wang, W. S. Y. (1973). The Chinese language. *Scientific American*, *228*(2), 50–60.

Wang, X., & Peng, G. (2014). Phonological processing in Mandarin speakers with congenital amusia. *The Journal of the Acoustical Society of America*, *136*(6), 3360–3370. https://doi.org/10.1121/1.4900559

Wang, Xiaoyue, 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

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

Wong, P. (2013). Perceptual evidence for protracted development in monosyllabic Mandarin lexical tone production in preschool children in Taiwan. *The Journal of the Acoustical Society of America*, *133*(1), 434–443. https://doi.org/10.1121/1.4768883

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.

Xu, Yi. (2013). ProsodyPro—A tool for large-scale systematic prosody analysis. *Proceedings of Tools and Resources for the Analysis of Speech Prosody (TRASP 2013)*, 7–10.

Xu, Yi, & Prom-on, S. (2019). Economy of effort or maximum rate of information? Exploring basic principles of articulatory dynamics. *Frontiers in Psychology*, *10*. https://doi.org/10.3389/fpsyg.2019.02469

Xu, Yisheng, 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

Yip, M. (2002). *Tone*. Cambridge: Cambridge University Press.

Yu, L. (2018). *An electrophysiological investigation of linguistic pitch processing in tonal-language-speaking children with autism (Unpublished doctoral dissertation)* [Ph.D. Thesis, Department of Speech, Language, and Hearing Sciences, University of Minnesota, USA]. https://conservancy.umn.edu/bitstream/handle/11299/201141/Yu_umn_0130E_19753.pdf?sequence=1

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

Yuan, J. (2011). Perception of intonation in Mandarin Chinese. *The Journal of the Acoustical Society of America*, *130*(6), 4063–4069. https://doi.org/10.1121/1.3651818

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, K., Peng, G., Li, Y., Minett, J. W., & Wang, W. S. Y. (2018). The effect of speech variability on tonal language speakers' second language lexical tone learning. *Frontiers in Psychology*, 9:1982. https://doi.org/10.3389/fpsyg.2018.01982

Table 1. Descriptive characteristics of study samples in Experiment 1

(a)	CASD ($n=26$)		CTD ($n=26$)		t	p
	M (*SD*)	Range	*M* (*SD*)	Range		
CA in years	7.44 (1.28)	6;0–10;4	7.48 (1.22)	6;0–10;0	−0.14	.891
Language score	37.53 (10.52)	14–54	43.10 (7.75)	25–52	−2.17	.035*
Nonverbal IQ	104.50 (22.81)	70–143	107.96 (19.21)	80–140	−0.59	.557
(b)	MASD ($n=26$)		MTD ($n=26$)		t	p
	M (*SD*)	Range	*M* (*SD*)	Range		
CA in years	7.69 (1.45)	6;0–11;7	7.65 (1.08)	6;6–10;0	0.09	.928
Verbal IQ	97.95 (17.06)	71–125	106.23 (10.99)	82–130	−2.07	.045*
Nonverbal IQ	105.96 (14.60)	79–129	108.62 (12.06)	87–128	−0.72	.478

CASD Cantonese-speaking children with ASD, *CTD* Cantonese-speaking TD children, *MASD* Mandarin-speaking children with ASD, *MTD* Mandarin-speaking TD children, *CA* Chronological age, *M* mean, *SD* standard deviation; *$p < .05$

Table 2. The results of fixed effects on the intercept term,linear term, and quadratic term for each level tone (df = 1)

Level tones	Time terms	Fixed effects	Speech condition				Nonspeech condition	
			Familiar segment		Unfamiliar segment			
			χ^2	*p*-value	χ^2	*p*-value	χ^2	*p*-value
(a) CT55	Intercept term	Group	3.32	.068	1.82	.177	2.42	.120
		Language	1.10	.294	1.84	.175	3.25	.072
		Group:Language	3.05	.081	2.28	.131	0.32	.570
	Linear term	ot1:Group	1.03	.310	0.41	.522	2.51	.113
		ot1:Language	**6.66**	**.010****	**15.71**	**< .001*****	**4.50**	**.034***
		ot1:Group:Language	0.76	.385	0.12	.726	0.96	.327
	Quadratic term	ot2:Group	0.69	.408	0.00	.985	0.16	.692
		ot2:Language	1.74	.188	0.00	.954	1.24	.266
		ot2:Group:Language	0.18	.675	0.31	.580	0.25	.616
(b) CT33	Intercept term	Group	0.16	.692	0.40	.525	1.11	.291
		Language	0.17	.679	0.00	.998	0.27	.604
		Group:Language	0.14	.708	0.00	.967	0.09	.767
	Linear term	ot1:Group	1.87	.171	2.65	.103	0.21	.647
		ot1:Language	**18.93**	**< .001*****	**15.55**	**< .001*****	0.04	.847
		ot1:Group:Language	0.77	.380	0.02	.894	0.04	.838
	Quadratic term	ot2:Group	0.43	.510	0.82	.366	0.00	.965
		ot2:Language	0.21	.650	0.25	.617	0.00	.990
		ot2:Group:Language	0.99	.319	0.24	.624	0.00	.944
(c) CT22	Intercept term	Group	1.74	.188	2.33	.127	1.92	.166
		Language	2.17	.141	2.77	.096	1.82	.177
		Group:Language	1.09	.297	0.08	.771	0.67	.413
	Linear term	ot1:Group	2.55	.110	0.55	.459	0.40	.526
		ot1:Language	**10.79**	**.001****	**22.77**	**< .001*****	2.15	.142
		ot1:Group:Language	0.04	.841	3.26	.071	0.16	.692
	Quadratic term	ot2:Group	0.27	.603	0.23	.633	0.00	.973
		ot2:Language	2.30	.129	0.05	.828	0.04	.837
		ot2:Group:Language	1.23	.267	0.53	.465	1.70	.192

The bold values indicate *p* values smaller than 0.05

***$p < .001$, **$p < .01$, *$p < .05$

Table 3. The results of fixed effects on the intercept term, linear term, and quadratic term for each contour tone (df = 1)

Contour tones	Time terms	Fixed effects	Speech condition				Nonspeech condition	
			Familiar segment		Unfamiliar segment			
			χ^2	*p*-value	χ^2	*p*-value	χ^2	*p*-value
(a) CT23	Intercept term	Group	1.56	.212	1.76	.185	2.74	.098
		Language	2.40	.122	0.21	.648	0.05	.826
		Group:Language	0.01	.936	0.50	.481	0.01	.929
	Linear term	ot1:Group	0.89	.345	0.14	.713	0.07	.794
		ot1:Language	2.38	.123	2.29	.130	3.43	.064
		ot1:Group:Language	0.06	.809	0.78	.376	0.26	.609
	Quadratic term	ot2:Group	1.17	.279	1.36	.244	1.69	.193
		ot2:Language	0.03	.857	0.19	.662	0.02	.890
		ot2:Group:Language	0.20	.654	0.11	.746	0.24	.624
(b) CT25	Intercept term	Group	2.28	.131	2.16	.141	2.87	.090
		Language	0.12	.732	0.66	.418	0.29	.588
		Group:Language	0.40	.526	0.23	.630	0.34	.558
	Linear term	ot1:Group	**5.68**	**.017***	**3.91**	**.048***	0.46	.497
		ot1:Language	0.60	.440	2.73	.098	1.19	.275
		ot1:Group:Language	0.11	.744	0.03	.872	0.27	.604
	Quadratic term	ot2:Group	**4.00**	**.046***	1.00	.317	0.21	.645
		ot2:Language	1.04	.308	0.12	.728	0.02	.902
		ot2:Group:Language	1.52	.218	1.73	.188	0.01	.911
(c) CT21	Intercept term	Group	1.84	.175	0.56	.453	2.66	.103
		Language	0.17	.682	0.37	.541	2.44	.118
		Group:Language	0.30	.584	0.32	.571	0.00	.957
	Linear term	ot1:Group	0.84	.359	1.82	.177	0.01	.912
		ot1:Language	0.41	.523	0.01	.919	1.34	.247
		ot1:Group:Language	0.23	.634	3.05	.081	2.62	.106
	Quadratic term	ot2:Group	2.23	.135	0.01	.905	1.00	.318
		ot2:Language	0.12	.734	0.04	.843	0.31	.581
		ot2:Group:Language	1.50	.221	0.17	.679	0.29	.593

The bold values indicate *p* values smaller than 0.05

*$p < .05$

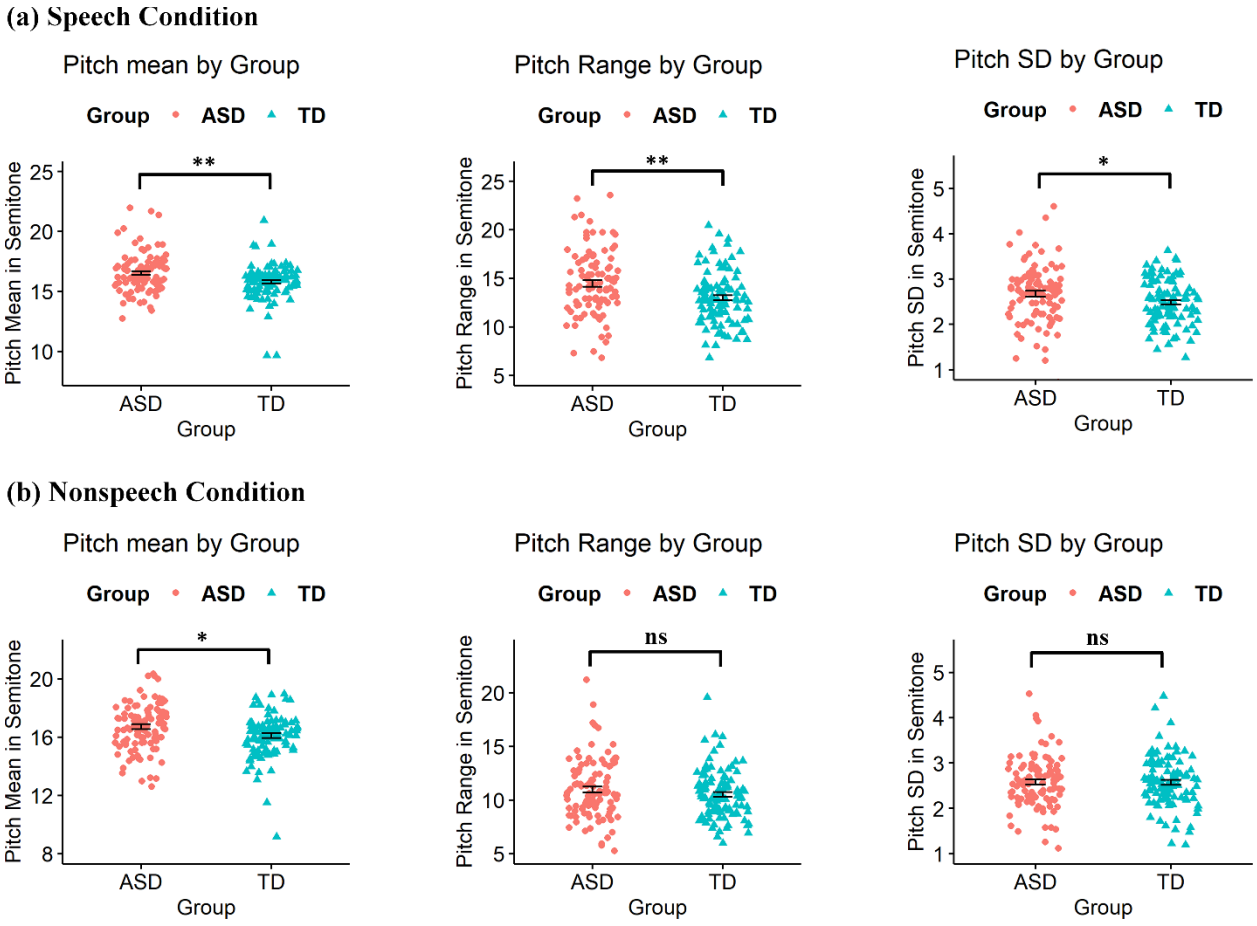

Figure 1. The pitch mean (left column), pitch range (middle column), and pitch SD (right column) produced by children with ASD and TD children when imitating models in **a** speech condition, and **b** nonspeech condition. The error bars were presented inside the jitters. $^{**}p < .01$; $^{*}p < .05$; *ns* not significant

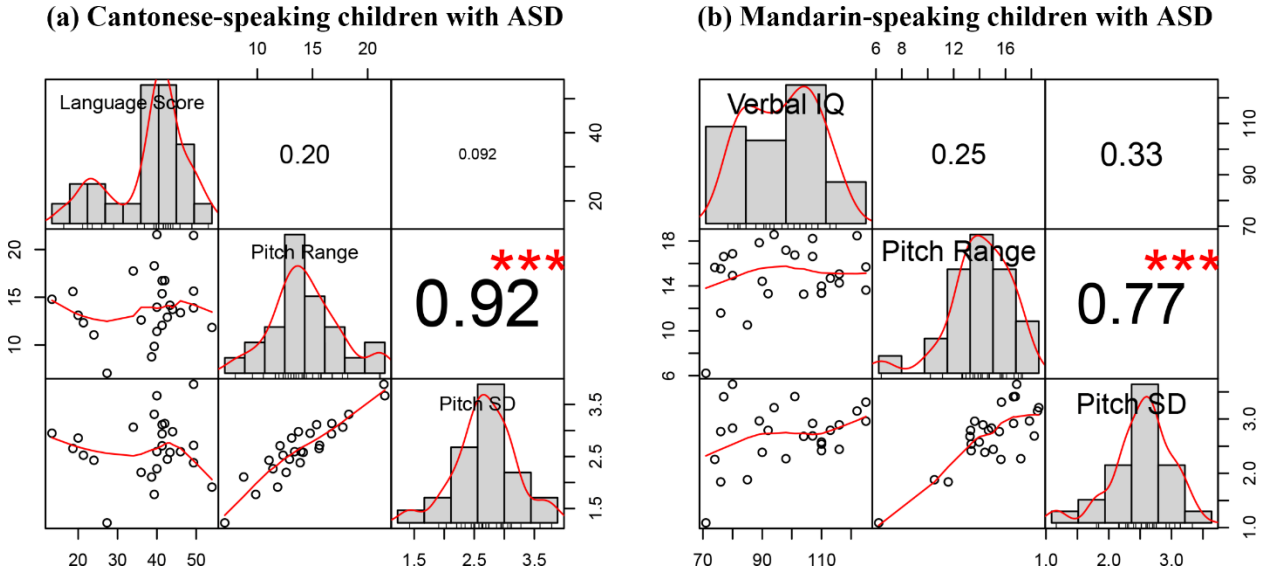

Figure 2. The correlations among language/verbal ability, pitch range, and pitch SD when imitating lexical tones in **a** Cantonese-speaking children with ASD, and **b** Mandarin-speaking children with ASD. The correlation coefficient *r* was displayed by numbers in the squares, with larger font indicating a larger *r* value

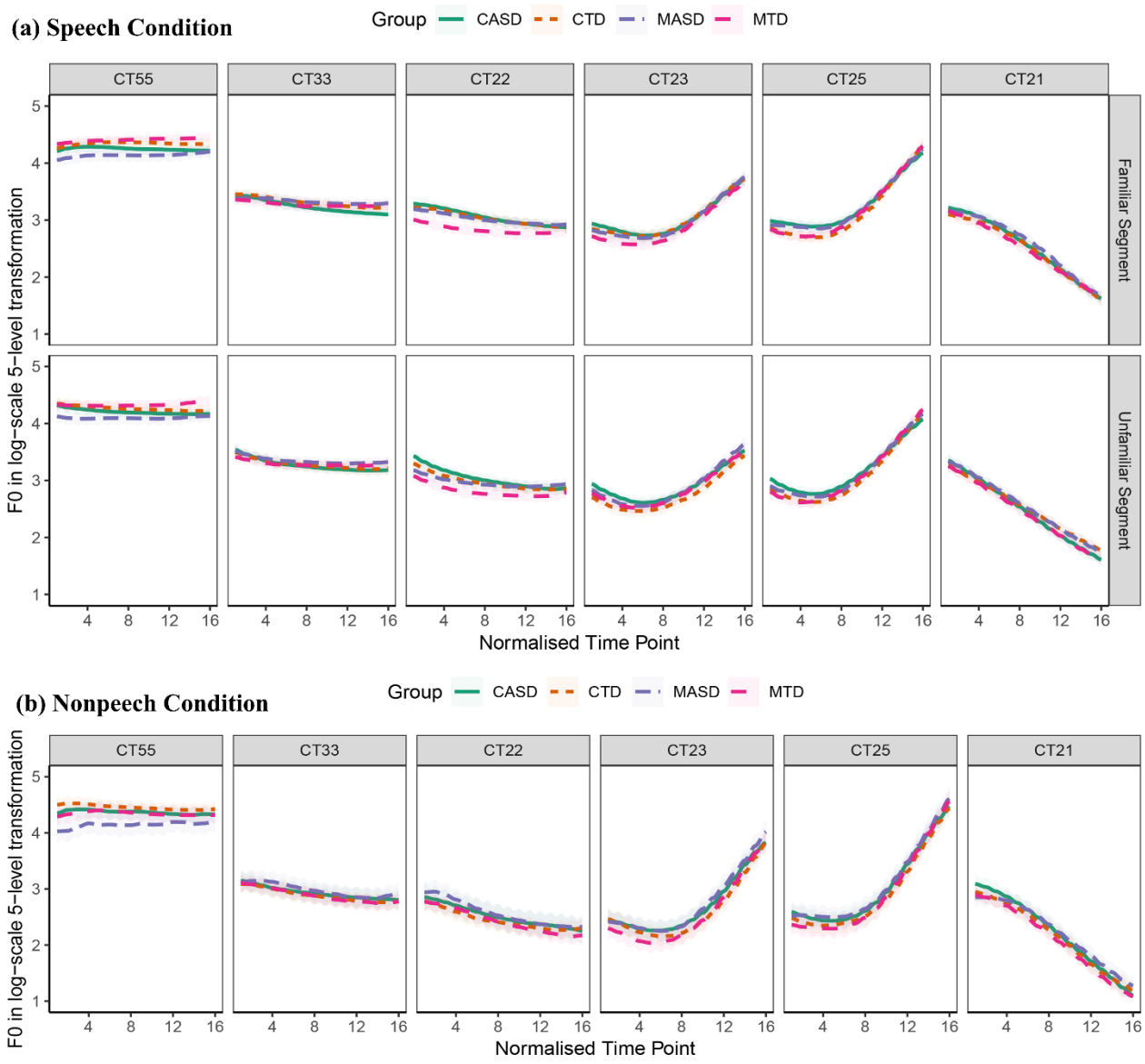

Figure 3. Pitch contours of **a** lexical tone and **b** non-linguistic pitch imitations produced by four subgroups (CASD, CTD, MASD, and MTD). The shades with light colors indicate standard error

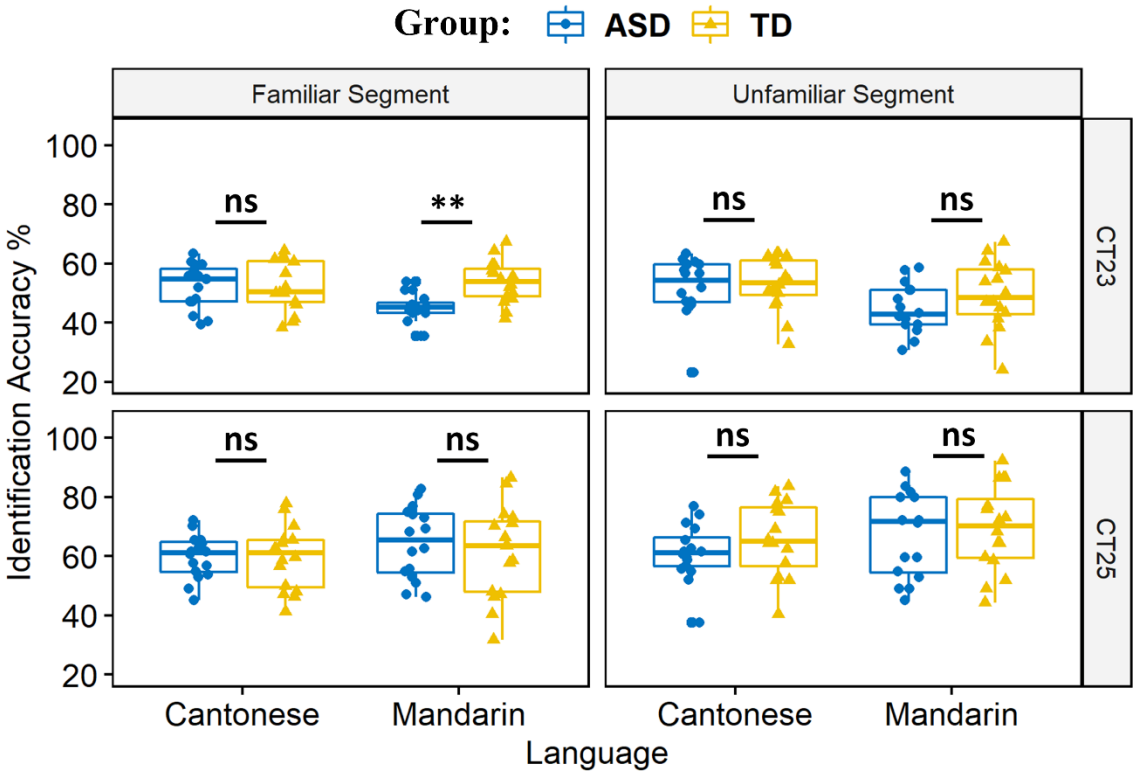

Figure 4. The Cantonese-speaking adults' identification accuracy of CT23 vs. CT25 imitations with the familiar and unfamiliar segments produced by CASD, CTD, MASD, and MTD. The bold line inside the boxes marks the median of identification accuracy, and the upper and lower boundaries of the boxes mark its upper and lower quartiles of accuracy. $^{**}p < .01$; *ns* not significant