

This version of the article has been accepted for publication, after peer review (when applicable) and is subject to Springer Nature's AM terms of use (<https://www.springernature.com/gp/open-research/policies/accepted-manuscript-terms>), but is not the Version of Record and does not reflect post-acceptance improvements, or any corrections. The Version of Record is available online at: <http://dx.doi.org/10.1007/s10803-024-06698-4>.

Title: Acquisition of speech prosody in a non-native language Mandarin by Cantonese-speaking Children with and without Autism Spectrum Disorder

Authors: Si Chen^{1,2,3,4}, Yixin Zhang^{1*}, Meixuan Li¹, Bin Li⁶, Shuang Lu⁷, Angel Chan^{1,2,3,5}, Haoyan Ge⁸, Tempo Tang⁹, Zhuoming Chen¹⁰

Affiliations:

1. Department of Chinese and Bilingual Studies, The Hong Kong Polytechnic University
2. Research Centre for Language, Cognition, and Neuroscience, The Hong Kong Polytechnic University
3. The Hong Kong Polytechnic University – Peking University Research Centre on Chinese Linguistics
4. Research Institute for Smart Ageing, The Hong Kong Polytechnic University.
5. Speech Therapy Unit, The Hong Kong Polytechnic University, Hong Kong, Hong Kong SAR, China
6. Department of Linguistics and Translation, City University of Hong Kong
7. School of Foreign languages, Renmin University of China
8. School of Education and Languages, Hong Kong Metropolitan University
9. The Hong Kong Child and Youth Services
10. The First Hospital of Jinan University

***Corresponding author: yixin-yz.zhang@polyu.edu.hk**

Acknowledgement:

This work was supported by the Hong Kong Polytechnic University (1-ZVRT; 1-ZE0D; 1-W08C; 1-WZ02) and partly supported by Sin Wai Kin Foundation Limited (R-ZH5Z), the Standing Committee on Language Education and Research, Education Bureau, HKSAR (K-ZB2P) and RGC (A-PB1B). All grants were awarded to the first author. We would also like to thank Jiaqi Ma, Baobao Yang and Luwei Ye for their help in data analysis.

Abstract

Children with autism spectrum disorder (ASD) often show abnormal speech prosody in their native language. Multilingual exposure is increasing among children with ASD, but the acquisition of speech prosody of non-native languages, especially tonal languages has rarely been investigated. Tonal languages can pose more difficulties as speakers need to use acoustic cues to make lexical contrasts as well as encode the focal function. This study aims to fill in this gap by studying prosodic focus-marking in Mandarin by autistic Cantonese-speaking children who also speak English and Mandarin ($n=25$), in comparison with their typically developing peers ($n=20$) and native Mandarin-speaking children ($n=20$). Natural prosodic marking of different types of focus was elicited by picture-based prompt questions. Four acoustic cues, f_0 range, mean f_0 , duration and intensity were used for linear mixed effects model fitting in relation to groups and focus conditions. The autistic children made use of fewer acoustic cues and produced less evident on-focus expansion in these cues than the typically developing children, especially the native-Mandarin speaking peers. They also demonstrated a clear preference to on-focus expansion than to post-focus compression. Moreover, these children, together with their native Cantonese-speaking peers, hyper-performed in tone realization, prioritizing lexical prosody over focus marking. Such hyper-performance may further limit their use of prosodic cues in focus marking in Mandarin. However, the difficulties the autistic children faced in the acquisition of speech prosody in a non-native tone language, though found, are not more than those they face in their mother tongue. Instead, they have some knowledge of using on-focus expansion in non-native languages. Hence, multilingual exposure may help them master the use of some focus marking strategies though they still need interventions to help them to implement their focus-marking knowledge more sufficiently in both native and non-native languages.

1. Introduction

Multilingualism has become a global trend and multilingual exposure is also increasing among children with autism spectrum disorder (ASD) (Smith, 2016). ASD is a neurodevelopmental disorder characterized by different degrees of deficits and deviation in social communication and interaction, the development of cognition and emotions and sensory processing (Lord et al., 2020). Although language skills are no longer part of the autism diagnosis (cf. DSM-5, APA, 2013), existing studies have reported delayed and divergent language development in autistic children compared to common milestones (Gernsbacher et al., 2016; Gagnon et al., 2021; Prévost & Tuller, 2022). In the late decades of the last century, professionals and parents have therefore questioned the effects of multilingual exposure on the language development of children with ASD. Some of them suggest monolingual exposure so as to avoid additional challenges and delays in the language development of this special population, but more and more recent studies report no adverse effects or positive influences of bilingualism on child language development (for review, see Gilhuber, 2023). Hence, it becomes essential to investigate the acquisition of multiple languages by autistic children. Compared to segments, there is a paucity in the research on speech prosody performance in non-native language by the autistic children, despite the importance of speech prosody and the ubiquitous abnormality in the use and development of prosody among autistic individuals (Paul et al., 2005; Lloyd-Fox et al., 2013). Speech prosody, the suprasegmental feature of speech that serves a wide range of pragmatic and paralinguistic functions by varying acoustic cues, is of high importance to people's daily speech communication (Gussenhoven & Chen, 2021).

The present study aims to fill in this gap by investigating the expressive use of prosody in Mandarin focus marking by native Cantonese-speaking trilingual children with and without ASD who also speak English and Mandarin, in comparison with their typically developing (TD) peers who speak Mandarin as their first language. In particular, we aim to contribute to the following two under-explored areas: 1) the expressive use of

speech prosody by autistic children in a non-native tone language (i.e., Mandarin) in comparison to their TD peers and 2) the expressive use of speech prosody of tone language speaking children (i.e., native Cantonese-speaking children) in a non-native tone language (i.e., Mandarin) in comparison to their native counterparts. The results will deepen our understandings of prosodic development in multilingual children with and without ASD.

1.1 Focus marking and Speech Prosody in Mandarin and Cantonese

Focus marking plays a critical role in information packaging by directing listener's attention to new information (Lambrecht, 1994). Depending on the size of on-focal elements, focus can be categorized into broad focus, where the focus falls on the whole utterance, and narrow focus where the focus falls on individual words (Halliday, 1967; Ladd, 1996). Narrow focus used to mark contrastive or corrective information is referred to as contrastive or corrective focus, while narrow focus used to answer wh-questions is referred to as on-contrastive narrow focus; the key difference is that contrastive focus indicates a direct rejection of the alternatives of limited number while non-contrastive focus marks information with non-limited alternatives (Halliday, 1967; Gussenhoven, 2004). How different linguistic cues is used by children to comprehend and mark focus remains a critical issue in the field of language acquisition for years (cf., Crain et al., 1994; Paterson et al., 2006; Zhou et al., 2012; Szendrői et al., 2018).

Cross-linguistically, prosodic means are used to raise perceived prominence of the constituents on focus, including the fundamental frequency (f_0), duration and intensity (Gussenhoven, 2004; Baumann et al., 2006; Gussenhoven and Chen, 2021). It is worth mentioning that (narrow) focus can be marked not only by increasing the prominence of the on-focus components such as giving them higher f_0 level, larger f_0 excursion, longer duration and larger intensity (i.e., on-focus expansion, OFE), but also by decreasing the prominence of the components not carrying focus. For instance, post-focus compression (PFC), a prosodic focus-marking strategy that reduces f_0 range and

intensity of the syllables following syllables carrying narrow focus, is found in many languages (for English see Eady & Cooper, 1987; Xu & Xu, 2005; for German see Féry & Kügler, 2008; for Japanese see Ishihara, 2011; for Mandarin see Xu, 1999; Xu et al., 2012). Nevertheless, PFC is not as natural and hence not widely used as OFE since many languages and dialects lack PFC, and it is also notoriously difficult to learn PFC in both native and non-native languages (Wu & Xu, 2010; Xu & Chen, 2016; Zhang & Wang, 2016; Liu et al, 2019).

In tone languages like Cantonese and Mandarin, f_0 is also used to distinguish the core meanings of words, in addition to express utterance-level prosody. Cantonese has six full tones (Tone 1: High-Level, Tone 2: High-Rising, Tone 3: Mid-Level, Tone 4: Low-Level/Falling, Tone 5: Low-Rising, and Tone 6: Low-mid Level), while Mandarin has four full tones (T1: High-level, T2: Mid-rising, T3: Low-convex and T4: High-falling and a neutral tone). The more complex tone system in Cantonese limits the use of f_0 in prosodic marking when compared to Mandarin and English, and the existing studies reports complex results. In terms of OFE, several studies report durational expansion as the primary cue to signal focus in Cantonese rather than f_0 exaggeration (Wu & Xu, 2010; Mok et al., 2014; Fung & Mok, 2018); while Man (2000) and Wu and Xu (2010) reported tone-specific OFE of f_0 excursion, Fung & Mok (2019) reported no significant on-focus f_0 changes. OFE of intensity in Cantonese focus-marking is also controversial, Wu and Xu (2010) did find an on-focus increase in intensity, but Hsu et al., (2018) did not. With regard to PFC, although studies find a compressing tendency in f_0 range or velocity at post-focus positions in Cantonese (e.g., Hsu et al., 2018; Lee & Barnes, 2023), PFC is far from an established feature in this language and no PFC in intensity has been reported in Cantonese focus-marking to the best of the authors' knowledge.

By contrast, prosodic-marking in Mandarin features both OFE and PFC, with f_0 used as the primary cue. Unlike English speakers who can mark focus by distinctive pitch patterns in addition to an increase in duration and intensity (Ouyang & Kaiser 2015), Mandarin speakers still need to maintain lexical tone contrasts when marking focus,

and hence to conform to its tonal system. To be specific, in syllables on narrow focus, the high pitch target in the embedded Tone 1, 2, or 4 is raised but the low target in Tone 3 and 4 is lowered (cf. Xu, 1999; Lee et al., 2016), and such strategies often result into an expansion in f_0 ranges of contour tones. Mandarin also marks focus with OFE in duration and intensity as well as PFC in f_0 range, f_0 height and intensity (Cao 2004; Liu & Xu, 2005; Wang et al., 2024), though Cao (2004) suggests that duration and intensity are not as important as f_0 changes.

1.2 Production of Prosodic Focus-marking by Children with and without Autism

Abnormal prosodic production has been widely found among the speakers with ASD (for a review, see Fusaroli et al., 2017). Overall, children with ASD often show delayed, deviant development and deficits in their use of speech prosody, e.g., producing utterance with higher mean f_0 , larger f_0 range and greater within-group variation (Hubbard et al., 2017; Asghari et al., 2021; Lau et al., 2022). With regard to focus-marking in particular, the existing studies mainly focused on English, and many of them used Profiling Elements of Prosodic Systems (PEPS-C), an assessment of children's abilities to discriminate and articulate the prosodic forms targeting at focus marking, boundary marking, interaction and affective expression (cf. Peppé & McCann, 2003). While autistic children were able to mark focus with longer duration and larger intensity for on-focus syllables like their TD counterparts, the contrast between the on-focus and non-focus syllables produced by the autistic children was not as prominent as the TD (Diehl & Paul, 2009, 2013; Nadig & Shaw, 2015). The expressive use of f_0 cues by autistic children is also less proficient. Nadig and Shaw (2015) reported that neither ASD nor TD group produced on-focus syllables with higher mean f_0 , while Diehl and Paul (2013) found significantly larger f_0 range and within-group f_0 variation in focus-marking among autistic children. Furthermore, Wang et al. (2024) found that the native Cantonese-speaking autistic children who speak English as their second language did

not show as proficient use of PFC as their TD peers when marking focus in English, though neither autistic or TD children in their study showed much OFE patterns.

Recent study investigated focus-marking in native tonal languages by autistic children, but non-native languages, especially tonal languages have rarely been examined. Ge et al. (2023) found that being Cantonese-English bilingual facilitates autistic children's prosodic marking in their L1 compared to the monolingual peers, but their performance is still not comparable to the TD group. Chen et al. (accepted) found that Cantonese-speaking children with ASD produced less sufficient OFE and showed a less sophisticated use of f_0 cues on both lexical and utterance levels. Little research has been done on the non-native use of utterance-level prosody. Research on focus-marking by TD children in tone language report tone-specific patterns of acquisition. Studies find that native Mandarin-speaking children at 4-5 years old are able to use pitch-related cues in focus marking like adults for Tone 1 and 4; however, their use of pitch-related cues for Tone 2 and 3 does not reach adult-like proficiency even at 11 years old (Chen et al., 2019). The adult-like use of non-pitch cues like duration, however, is developed as young as 4-5 years old (Yang & Chen, 2018). Liu et al. (2016) found that native Bai-speaking children do use prosodic cues to mark focus in Mandarin at 6-7 years old, but they are more proficient in using duration. However, little is known about focus-marking in Mandarin produced by children who speak Cantonese or English as their mother tongue and how the clinical condition of ASD may affect focus-marking in a non-native tone language.

1.3 The Current Study

Overall, there is a large paucity in the literature on children's acquisition of speech prosody in non-native languages. For both autistic children and their TD peers who are native speakers of a tone language, it remains unknown whether and how multilingualism influences their use of f_0 to encode both lexical and focal functions in the non-native languages. More importantly, whether and how ASD may add to the

difficulty of such a complex task. To fill in these gaps, we investigated prosodic focus-marking in Mandarin produced by native Cantonese-speaking children with and without ASD, as well as Mandarin-speaking children with matched backgrounds. We are interested in the following two questions:

RQ1. What is the effect of the clinical condition ASD on native Cantonese-speaking children's acquisition of prosody in Mandarin, namely, whether and how do Cantonese-speaking children with ASD differ from their TD peers in the expressive use of prosody in focus-marking in Mandarin?

RQ2. What is the effect of linguistic background on children's acquisition of utterance prosody in Mandarin, namely, whether and how do Cantonese-speaking children with and without ASD differ from the native Mandarin-speaking TD peers in the expressive use of prosody in focus-marking in Mandarin?

RQ3: How does focus marking interact with the realization of tones, and whether the interaction differs among the three participant groups?

2. Methodology

2.1 Participants

Twenty-five native Cantonese-speaking children with ASD (18 Males, 7 Females), twenty TD native Cantonese-speaking children (14 Males, 6 Females) and twenty TD native Mandarin-speaking children (14 Males, 6 Females) participated in this experiment (for demographic information see Table 1). All of the autistic participants were formally diagnosed with ASD and none of them was diagnosed of or was suspected to have any other disorders. No TD participants had any speech or language disorders or suspected to have any disorders. According to their self-reported data, all the native Cantonese-speaking participants also spoke English and Mandarin, but they learned English earlier or at the same time as Mandarin, but used Cantonese and English

as their primary languages to communicate. ANOVA tests found no significant age differences among the three participant groups [$F(2, 63) = 0.27, p = 0.76$], and the age of acquisition of Mandarin [$F(1, 41.57) = 1.82, p = 0.08$] and English [$F(1, 37.35) = 0.59, p = 0.56$] were not significantly different between the two native Cantonese-speaking groups.

Child participants came to the speech laboratory at the Hong Kong Polytechnic University accompanied by parents, who signed the consent forms of a protocol approved by the Human Subjects Ethics Sub-committee at the Hong Kong Polytechnic University on behalf of the children. Both the parents and the children were well-informed and agreed to participate in the experiment, and written consent was obtained from parents on behalf of their children. They were also compensated for participating in the experiment.

2.2 Tests

Participants completed Raven's Progressive Matrices (Raven, 2003) and Wechsler Intelligence Scale for Children (WISC) (Wechsler & Kodama, 1994) respectively. The autistic participants were also assessed with Autism Diagnostic Observation Schedule (ADOS-2) (Lord et al., 2012). Tests results are presented in Table 1. ANOVA tests indicated significant differences in the Raven's Progressive Matrices [$F(2, 62) = 4.85, p < 0.05$] and the overall WISC scores [$F(2,62) = 16.21, p < 0.0001$]. Post-hoc comparisons showed that the Mandarin group was of significantly higher Raven score than the Cantonese-speaking children with ASD ($p < 0.01$), as well as significantly higher WISC score than the Cantonese-speaking TD ($p < 0.05$) and ASD groups ($p < 0.0001$) and other groups did not exhibit significant differences.

Table 1 Demographic information and test scores of the participants

	Cantonese ASD	Cantonese TD	Mandarin TD

Age		9.50 ± 1.51	9.77 ± 1.11	9.44 ± 1.49
English Learning Age		3.06 ± 0.82	2.90 ± 0.97	5.55 ± 1.16
Mandarin Learning Age		4.04 ± 1.65	3.30 ± 1.08	4.33 ± 1.67
Musical Training Age		5.96 ± 1.42	5.91 ± 1.30	5.67 ± 1.80
Musical Training Duration		2.43 ± 1.95	3.64 ± 2.16	2.56 ± 1.40
IQ score		103.64 ± 14.47	108.95 ± 17.77	117.95 ± 13.78
WISC-V-Chinese score	Overall	29.12 ± 10.39	37.00 ± 9.67	45.05 ± 7.39
	Similarities	20.72 ± 7.39	21.95 ± 9.16	28.70 ± 5.29
	Vocabulary	23.52 ± 11.06	33.85 ± 8.31	36.80 ± 6.57
	Comprehension	14.76 ± 7.88	22.00 ± 5.65	23.45 ± 6.11
ADOS score	Overall	12.54 ± 5.99		
	Communication	4.13 ± 2.16		
	Reciprocal Social Interaction	8.42 ± 4.06		
				-

2.3 Materials

The game, “under the shape”, was used to elicit natural responses from the participants. We used 12 pictures and 12 corresponding sentences depicting the content of the pictures as the materials. All sentences were five-syllable sentences, made up of a two-syllable subject, a one-syllable verb and a two-syllable object. For each sentence, a series of questions were designed to elicit natural production of the desired types of focus on subject, verb or object (for an example, see Figure 1; for the full list see Appendix 1). Target sentences were grouped into 4 blocks and each block contains three target sentences so as to reduce the memory load. In this way, all the child participants were able to memorize the sentences describing the pictures. The experimental session

was made up of 4 blocks, each containing 42 randomized trials [3 out of 12 target sentences * (1 broad focus + 1 non-contrastive narrow focus * 3 positions + 1 contrastive narrow focus * 3 positions) * 2 repetitions]. In total, 168 target sentences (42 trials * 4 blocks) and 840 (168 sentences * 5 syllables) syllables were collected from each participant.

2.4 Procedure

The participants' speech production was recorded in a sound-proof booth using the Steinberg UR22mkII USB Audio Interface and the Audio Technica ATone 2035 condenser microphone at a sampling rate of 44100 Hz in Audacity.

Every block consisted of a practice session and a test session. The practice helped to reduce production errors in the test session. The order of blocks was counterbalanced across participants within each group and all the trials in each session were presented randomly by the software E-prime 2.0 (Schneider et al., 2002).

Every trial involved a series of pictures displayed on a computer screen, and the participants' task was to respond to the researcher's question based on the picture provided in each trial. The experimenter would ask the question again instead of just asking for a correction if a participant answered the question incorrectly in order to preserve the naturalness of the response. Every participant in the experiment was able to correct themselves up to a maximum repetition of three times.

3.3 Data Analyses

The five syllables of each sentences were manually segmented using Praat (Boersma & Weenink, 2023), following the procedure of segmentation written by Jangjamras (2011). The data were extracted using ProsodyPro (Xu, 2013), and abnormal data were manually checked by the first and second authors. In total, 223 syllables (0.41%) were removed from the 48300 syllables due to creakiness and other abnormality. The f_0 range

(= maximum f_0 value - minimum f_0 value), the mean f_0 , and the mean intensity of the sonorant part as well as the duration of these syllables were calculated. These four acoustic parameters were used as the outcome variables of the statistical analyses. *Participant Group* (ASD vs. Cantonese TD vs. Mandarin TD), *Focus Condition* (i.e., the relative position of a syllable to a certain type of focus, Broad focus vs. Non-contrastive focus vs. Contrastive focus vs. Post-focus vs. Pre-focus), the embedded *Lexical Tone* (High-level T1 vs. Rising T2 vs. Falling T4) were used as explanatory variables.

Linear mixed effects (LME) models were fitted to evaluate the fixed effects and their interactions on the four outcome variables using the package *lme4* (Bates et al., 2015) in R (R Core Team, 2024). Tukey post-hoc tests were used for post-hoc comparisons of the interactions of interests, using the package *emmeans* (Lenth et al., 2023) in R.

3. Results

Figure 2 shows the predicted values of f_0 range, mean f_0 , duration and mean intensity generated by the corresponding LME models.

Figure 2 Predicted values of f_0 range, mean f_0 , duration and mean intensity generated by the corresponding LME models/Boxplots

3.1 f_0 range

Evaluation of the LME model showed that the inclusion of *Participant Group* [$\chi^2(2) = 26.95, p < .0001$], *Focus condition* [$\chi^2(4) = 39.96, p < .0001$], *Lexical Tone* [$\chi^2(2) = 6.91, p < .0005$], as well as the two-way [*Participant Group***Focus condition*, $\chi^2(8) = 3.89, p < .0001$; *Focus condition***Lexical Tone*, $\chi^2(8) = 7.72, p < .0001$; *Participant Group***Lexical Tone*, $\chi^2(4) = 154.35, p = .0001$] and three-way interactions [$\chi^2(16) = 4.63, p < .0001$] among *Participant Group*, *Focus condition* and *Lexical Tone* significantly contributed to the model.

Group-wise differences. F_0 range of the two contour tones, T2 and T4 produced by the two native Cantonese-speaking groups was significantly larger than those produced by the native Mandarin-speaking group in all focus conditions ($ps < .0001$). In addition, the ASD group also produced T1 with significantly larger f_0 range than the native Mandarin speakers when syllables carry broad focus ($p < .05$) or are in pre-focus condition ($p < .005$).

Focus-marking strategies. In general, f_0 range was used in focus-marking in contour tones only by all three groups, especially in T4, the falling tone (Figure 2A). The Cantonese- and Mandarin-speaking TD group showed PFC in f_0 range while the ASD group did not.

When carrying T2, the syllables on non-contrastive and contrastive focus were produced with significantly larger f_0 range than the pre-focus syllables in the ASD group ($ps < .05$), while in the Mandarin TD group, f_0 range of the syllables on non-contrastive focus was marginally larger than the post-focus syllables ($p = .0537$); no significant difference between focus conditions was found in the Cantonese TD group.

When carrying T4, the pre-focus syllables were produced with the smallest f_0 range in the ASD group, significantly smaller than all the other focus conditions including the post-focus syllables ($ps < .001$), and the post-focus syllables were also produced with significantly smaller f_0 range than the syllables carrying contrastive focus ($p < .005$); in the two TD groups, both the pre- and post-focus syllables were produced with f_0 range significantly smaller than syllables on contrastive and non-contrastive focus (Cantonese: $ps < .05$; Mandarin: $ps < .001$). Furthermore, the range of T4-bearing syllables on broad focus was significantly larger than the post-focus syllables in the Cantonese TD group ($ps < .005$), while the range of T4-bearing syllables contrastive focus was also significantly larger than syllables on broad focus in the Mandarin TD group ($p < .05$).

Tone realization. Although different tones have different f_0 contour, but the distinctiveness of tonal contour indexed by f_0 range was only significant in the

Cantonese TD group, namely, the f_0 range of T4 was significantly larger than T1 in syllables on broad, contrastive and non-contrastive focus ($ps < .05$), whereas in the ASD group, the f_0 range of T4 was marginally larger than T1 in post-focus syllables ($p = .0512$).

3.2 Mean f_0

Evaluation of the LME model showed that not only the inclusion of *Participant Group* [$\chi^2(2) = 1.67, p < .0001$], *Focus condition* [$\chi^2(4) = 316.99, p < .0001$], *Lexical Tone* [$\chi^2(2) = 38.29, p < .0001$], but also the two-way [*Participant Group*Focus condition*, $\chi^2(8) = 121.53, p < .0001$; *Focus condition*Lexical Tone*, $\chi^2(8) = 21.82, p < .0001$; *Participant Group*Lexical Tone*, $\chi^2(4) = 205.87, p < .0001$] and three way interactions [$\chi^2(16) = 11.35, p < .0001$] between them significantly contributed to the model.

Group-wise differences. Compared to f_0 range, not much significance was found between groups in mean f_0 with focus condition and lexical tones controlled. Only in the post-focus condition was the mean f_0 of T1 and T4 produced by the Mandarin TD significantly smaller than those produced by the two native Cantonese-speaking groups (T1: ASD vs. Mandarin TD: $p < .005$; Others: $ps < .05$), indicating the significance of post-focus lowering in the Mandarin group.

Focus-marking strategies. All three groups used a post-lowering strategy in focus-marking; however, the two Cantonese-speaking groups, with and without ASD, also showed pre-focus raising in the two contour tones, while the Mandarin TD group showed OFE in the mean f_0 of T1 and T4 (Figure 2B).

For T1, no statistical significance was found in the ASD group, but in the Cantonese TD group, post-focus T1 was significantly lower than T1 in other focus conditions except for T1 on broad focus ($ps < .0005$). In the Mandarin TD group, not only post-focus T1 was significantly lower than T1 in other focus conditions ($ps < .0001$), but the T1 on contrastive was also significantly higher than T1 on broad focus and pre-focus

T1 ($p < .001$).

For T2, in the ASD group, although the post-focus T2 was the lowest, but only its difference from the pre-focus T2 was significant ($p < .0001$). In the Cantonese TD group, while post-focus T2 was significantly lower than T2 on contrastive focus ($p < .05$), pre-focus T2 was significantly higher than T2 in all other focus conditions ($ps < .0005$). In the Mandarin TD group, however, it was the post-focus T2 that was significantly lower than T2 in all the other conditions, including pre-focus ($ps < .0001$).

For T4, in the two Cantonese-speaking groups, post-focus T4 was significantly lower than the T4 on contrastive and non-contrastive focus (ASD: $ps < .05$; TD: $ps < .0001$), while the pre-focus T4 was significantly higher than post-focus T4 and T4 on broad focus (ASD: $ps < .001$; TD: $ps < .0001$). By contrast, all the differences between focus conditions in the Mandarin TD group were statistically significant ($ps < .0005$), except those between broad focus and pre-focus and between the two types of narrow focus.

Tone realization. In all three groups, the mean f_0 of T1 and T4 was significantly higher than T2 in all focus conditions, except that no statistical significance was seen in the post-focus condition in the Mandarin TD group.

3.3 Duration

Evaluation of the LME model showed that the inclusion of *Participant Group* [$\chi^2(2) = 0.13, p = .88$], *Focus condition* [$\chi^2(4) = 1.89, p = .11$] and their interaction [$\chi^2(8) = 1.68, p = .52$] contributed to the model.

With focus condition controlled, no significant difference was found between groups, and only in the Mandarin-speaking TD group existed significant differences between focus conditions, namely, syllables on contrastive focus were significantly longer than pre- and post- syllables ($ps < .01$) and marginally longer than syllables on narrow ($p = .052$) and broad focus ($p = .051$).

3.4 Intensity

Evaluation of the LME model showed that the inclusion of *Focus condition* [$\chi^2(4) = 186.97, p < .0001$], *Lexical Tone* [$\chi^2(2) = 6.23, p < .005$], as well as the two-way and three-way interactions between *Participant Group*, *Focus condition* and *Lexical Tone* significantly contributed to the model [*Participant Group*Focus condition*, $\chi^2(8) = 178.02, p < .0001$; *Focus condition*Lexical Tone*, $\chi^2(8) = 5.434, p < .0001$; *Participant Group*Lexical Tone*, $\chi^2(4) = 42.50, p < .0001$; *Participant Group*Focus condition*Lexical Tone*, $\chi^2(4) = 2.00, p < .05$].

Group-wise differences and Tone realization. With focus condition and lexical tone controlled, no significant difference was observed between groups. Also, no significant difference between different tones was found in intensity with participant group and focus condition controlled.

Focus-marking strategies. The two Cantonese-speaking groups showed some significant OFE in T1-bearing syllables, i.e., the pre-focus T1 was of significantly lower intensity than the T1 on contrastive focus in both groups (ASD: $p < .05$; TD: $p < .005$, and also the T1 on non-contrastive focus in the Cantonese TD group ($p < .05$). By contrast, the Mandarin TD group showed evident PFC across all three tones in intensity, with the post-focus syllables significantly lower than syllables in other focus conditions ($ps < .0001$).

4. Discussion

This acoustic study analyzed the expressive use of prosody in focus-marking in Mandarin by native Cantonese-speaking children with and without ASD. By comparing their performance to the native Mandarin-speaking children with matched backgrounds, we found evident influence of the clinical condition, nativeness, and their interaction on children's acquisition of speech prosody in a non-native tone language. Acoustic

cues serving different focus-marking strategies (i.e., OFE and PFC) used by each participant group are summarized in Table 2.

Table 2 Focus-marking Strategies (Acoustic cues listed here are all of statistical significance).

	OFE	PFC
ASD	f_0 range of T2 (Contrastive & Non-contrastive); intensity of T1 (Contrastive)	mean f_0 of T4
Cantonese TD	f_0 range of T4 (Contrastive & Non-contrastive); intensity of T1 (Contrastive & Non-contrastive)	f_0 range of T4; mean f_0 of T2 and T4
Mandarin TD	f_0 range of T4 (Contrastive & Non-contrastive); mean f_0 of T1 (Contrastive); mean f_0 of T4 (Contrastive & Non-contrastive); duration (Contrastive)	f_0 range of T2; f_0 range of T4; mean f_0 of T1, T2 and T4; intensity of T1, T2 and T4

With regard to our first research question, the Cantonese-speaking children with ASD mainly differ from their native Cantonese- and Mandarin-speaking TD peers in the focus-marking strategies. Compared to the two TD groups, especially the Mandarin-speaking TDs, the autistic children demonstrated a less complete acoustic profile in utilising both OFE and PFC. In fact, the significant lowering of post-focus tones found in the present study has not been widely reported in the existing literature on PFC, as PFC typically referred to the compression in pitch excursion rather than height (e.g., Xu, 1999; Xu and Xu, 2005; Xu et al., 2012). From this point of view, while the Cantonese-speaking TD group was able to compress f_0 range of the post-focus T4 to increase the prominence of the preceding on-focus constituents, no typical PFC was found regarding f_0 range in the ASD group. Our finding confirms again that PFC is a feature difficult to acquire, and the clinical condition of ASD adds to such difficulties. Furthermore, it is worth pointing out that the ASD group actually showed some tendency of OFE in the f_0 range of T4, the mean f_0 of T2 and T4 as well as the intensity of T2 and T4 (Figure 2), but none of these reached statistical significance. Such

tendency, together with the significant OFE found in the f_0 range of T2 and the intensity of T1, indicates that the Cantonese-speaking ASD children may have some knowledge of OFE as their TD peers, but their expressive performance was less sufficient. This is in line with the existing studies, which report that the contrasts between the on-focus and non-focus constituents produced by the autistic children are less prominent (Diehl & Paul, 2009, 2013; Nadig & Shaw, 2015).

At the same time, the performance of the ASD group in the present study also differs from the existing literature in several aspects. Firstly, while the studies on English focus-marking suggest that duration and intensity are the common cues used by the autistic children (e.g., Diehl & Paul, 2009, 2013), the ASD group in our study used f_0 cues, intensity but not duration. We argue that the use f_0 cues may be due to the fact that our participants were native tone-language speakers, and the testing language is also a tone language. The native Cantonese-speaking autistic children also used lowered f_0 curve in post-focus position to mark focus in their L2 English (Wang et al., 2024). The lack of durational change, however, is slightly surprising because duration is considered the primary cue of focus-marking in Cantonese, but neither the Cantonese-speaking ASD or TD group here transferred their L1 knowledge to Mandarin, that is, to mark focus with OFE or PFC in duration. Wang et al. (2024), by contrast, reported that native Cantonese-speaking children with ASD shortened the duration of post-focus syllables. In fact, they reported that Cantonese-speaking children, whether having ASD or not, used PFC but not OFE, but the ASD group in this study did not show much PFC except in mean f_0 but a clear preference to OFE. It is reported that autistic children did not have sufficient OFE in f_0 range and duration when marking focus in their native language Cantonese (Chen et al., accepted). However, this clinical condition did not make autistic children have more difficulties in acquiring OFE in their third language compared to their TD peers.

The aforementioned differences between our studies and the existing literature light on the mechanisms that determine transfer or cross-linguistic influence on L2 and L3

acquisition (a citation on review will be added). Cumulative-enhancement model (CEM, Flynn et al., 2004) and Scalpel model (SM, Slabakova, 2017) assume that the linguistic features of people's acquired language, L1 or L2, are transferred to L3 property-by-property, and it is the perceived similarity between the linguistic structures rather than the order of acquisition or general language distance that influences the transfer. CEM is mainly different from SM as it argues that transfer is determined by whether it is perceived as facilitative or not, while SM identifies more factors that can affect the transfer other than facilitativeness such as complexity and structure frequency. In our case, the transfer of OFE from Cantonese can be explained by either model. Firstly, the transfer facilitates focus marking in Mandarin. Moreover, compared to PFC, OFE is a focus marking strategy more frequently used in world languages (Xu et al., 2012). Therefore, even autistic children can easily acquire this strategy in learning Mandarin, whereas the development of PFC may be more cognitively challenging and hence was only observed in TD children. The "failure" in the transfer of OFE in duration may be explained by the low perceived facilitativeness, as duration is not the primary cue for focus marking in Mandarin but may influence tonal realization (e.g., Bao, 2008).

The overall deficits autistic children demonstrate in focus-marking may also be due to the difficulties they have in using prosodic features to perceive focus. For instance, Ge et al. (2022) reported significantly lower accuracy and longer reaction time in Cantonese-speaking children with ASD in the perception of focus-marking even in their mother tongue. The lack of perceptual sensitivity to the prosodic cues signaling prominence may contribute to the failure in the acquisition of the expressive use of these cues among the autistic population.

With regard to our second research question, in addition to the differences in focus marking strategies, we also found some direct between-group differences in the two f_0 parameters, especially the f_0 range. This also helps answer our third research question, as significant differences in tones were found between the two Cantonese-speaking groups and the native Mandarin-speaking group, and the tone production in the native

Mandarin-speaking group was influenced more by focus conditions. It may be argued that these two non-native groups tend to prioritize the acoustic realization of the lexical function, that is, the lexical tones over that of the focal function, which limits their use of acoustic cues in focus marking. In other words, the non-native speakers hyper-produced the lexical tones with larger f_0 excursions and levels, and hence make it harder to further vary these cues to mark focus, resulting into their less complete acoustic profile of focus-marking observed in the present study. The hyper-articulation of tones was even more evidently seen in the ASD group, as more and larger statistically significant differences were found between them and the Mandarin-speaking TDs.

We offer two explanations to such hyper-articulation. Firstly, according to ‘hyper-articulation and hypo-articulation’ (H&H) theory (Lindblom, 1990), economy of effort can cause speech production to be hypo-articulated, and communicative contrasts from the listeners’ perspective may lead to it becoming hyper-articulated. Non-native speakers might go to the “hyper-articulation end”, allowing longer duration and larger f_0 movement so as to enhance intelligibility by native listeners (Lindblom, 1990). As for the autistic group, in addition to enhancing intelligibility, the hyper-articulation in prosody observed in the autistic group was in fact in line with the sing-songy characteristics reported by previous studies (e.g. Wehrle et al. 2002). Secondly, Cantonese is a tone language with more complex tone system than Mandarin, which may therefore lead to higher requirements on the accuracy of tone realization in various focus conditions. Even in producing Mandarin with a less complex tone system, these non-native children made more efforts in encoding lexical contrasts, which may be due to the influence from their L1.

The differences observed in the tone realization in fact echo with the hyper-articulation found in the two Cantonese-speaking groups as well as the different focus-marking strategies the three groups used. To be specific, the significance and marginal significance found between tones in the two f_0 parameters suggest that the two Cantonese-speaking groups, especially the TDs, made an extra effort in keeping the

distinctiveness of the tone shapes; by contrast, the lack of significant differences between T2 and the other tones in the post-focus condition in the native Mandarin-speaking TD is likely due to PFC, that is, all tones are too compressed to be significantly different from each other in f_0 range and level.

5. Conclusion

This study is the first to examine the expressive use of prosodic cues in focus marking in Mandarin, a tone language, by native Cantonese-speaking children with and without ASD. In this way, we did not only contribute to the less investigated area of prosodic focus-marking production in a non-native language by autistic children, but also the less investigated population, i.e., tone-language-speaking autistic children. We identified that their use of prosodic means is less proficient than the native and non-native peers in Mandarin focus-marking. However, they still showed some use of prosodic cues in focus-marking as their TD peers. However, considering the results from previous studies, our findings do not suggest that they face even more difficulties in acquiring speech prosody in non-native languages than in their mother tongue. Instead, their performance indicated that they have some knowledge of OFE in their third language. Therefore, multilingual exposure to children with ASD may not necessarily bring negative effect in their acquisition of speech prosody. More interventions are needed to help them better use prosody in focus-marking in the future.

