ASHA American Speech-Language-Hearing Association

JSLHR

Research Article

Sung Speech Training Improves Prosodic Focus Marking in a Nondominant Language in Children With Autism Spectrum Disorder

Yixin Zhang,[a] Si Chen,[a,b,c,d] Meixuan Li,[a] Bin Li,[e] Shuang Lu,[f] Angel Chan,[a,b,c,g] Haoyan Ge,[h] Tempo Tang,[i] and Zhuoming Chen[j]

[a] Department of Chinese and Bilingual Studies, The Hong Kong Polytechnic University, China [b] Research Centre for Language, Cognition, and Neuroscience, The Hong Kong Polytechnic University, China [c] Peking University Research Centre on Chinese Linguistics, The Hong Kong Polytechnic University, China [d] Research Institute for Smart Ageing, The Hong Kong Polytechnic University, China [e] Department of Linguistics and Translation, City University of Hong Kong, China [f] School of Foreign Languages, Renmin University of China, Beijing [g] Speech Therapy Unit, The Hong Kong Polytechnic University, China [h] School of Education and Languages, Hong Kong Metropolitan University, China [i] The Hong Kong Child and Youth Services, China [j] The First Hospital of Jinan University, Guangdong, China

ARTICLE INFO

Article History:
Received February 15, 2025
Revision received July 4, 2025
Accepted July 22, 2025

Editor-in-Chief: Jessica E. Huber
Editor: Leah Catherine Fabiano

https://doi.org/10.1044/2025_JSLHR-25-00104

ABSTRACT

Introduction: Music and speech prosody share notable parallels, and music-based interventions have shown promise in fostering language development and social responsiveness. Song-based training, leveraging acoustic similarities between song and speech, is especially effective. This study examined whether short-term song-based training could enhance prosodic focus-marking in nondominant languages for autistic children. Specifically, it explored improvements in focus-marking strategies, such as on-focus expansion (OFE) and post-focus compression (PFC), and the number of prosodic correlates used.
Method: A short-term sung speech training intervention was designed, aligning melodic patterns with Mandarin's prosodic focus marking. Eighteen native Cantonese-speaking children with autism spectrum disorder underwent short-term sung speech training, and their pre- and posttraining performance was compared with two control groups: 18 Cantonese-speaking and 20 Mandarin-speaking typically developing children. Comparisons were made across participant groups as well as within the autistic group before and after the training.
Results: Sung speech training improved OFE use, particularly in fundamental frequency range, for noncontrastive focus marking in autistic children. Effects on PFC were less evident, and the training primarily enhanced OFE rather than increasing the number of prosodic correlates used. Control Cantonese-speaking participants showed no comparable improvements.
Conclusion: These findings highlight the potential of short-term, perception-based sung speech training as a supplementary intervention for improving prosodic focus marking in trilingual autistic children's nondominant languages, indicating positive cross-domain effects on speech-processing abilities.
Supplemental Material: https://doi.org/10.23641/asha.30347731

Autism spectrum disorder (ASD) is a neurodevelopmental condition characterized by challenges in social communication and interaction, as well as restricted interests and repetitive behaviors (American Psychiatric Association Division of Research, 2013). These social and communication difficulties often extend to various aspects of verbal and nonverbal expression, including speech prosody. Speech prosody, which encompasses rhythm, melody, and emphasis, plays a vital role in conveying meaning, emotion, and sentence structure (Gussenhoven & Chen, 2021). Research has shown that autistic children often show delayed, deviant development and deficits in their expressive use of speech prosody, for example, producing utterance of higher

Correspondence to Si Chen: sarah.chen@polyu.edu.hk. ***Disclosure:*** *The authors have declared that no competing financial or nonfinancial interests existed at the time of publication.*

mean fundamental frequency ($F0$), larger $F0$ range, and greater within-group variation (Asghari et al., 2021; Hubbard et al., 2017; Lau et al., 2023). Since speech prosody plays a critical role in children's sociopragmatic development (Hübscher & Prieto, 2019), such deficits can hinder their ability to effectively navigate social interactions. Consequently, speech prosody is an important target for interventions aimed at enhancing both communication and social skills in autistic children.

In addition, there is also an increase in the multilingual exposure among autistic children as multilingualism has become a global trend (Stein-Smith, 2016). Although recent studies reported no adverse influence of multilingualism on the language development of children with ASD (for a review, see Gilhuber et al., 2023), multilingual autistic children face no fewer difficulties when communicating in their second and third languages, just like in their first language (Chen et al., 2025; Ge et al., 2024; Wang et al., 2024). Therefore, effective interventions need to be designed to improve their prosodic abilities in nondominant languages. As a response to these problems, the present study explores the potential of using sung speech as an intervention method to improve prosodic focus marking in a third language of children with ASD.

Sung Speech Interventions for Autistic Children

Music has been identified as a pivotal element in human life since the early stage (Malloch & Trevarthen, 2009) and, more importantly, a strength in autistic individuals when compared to speech (for a review, see O'Connor, 2012). Studies have reported not only advanced music processing abilities among the autistic population but also a positive correlation between autistic traits and musical skills (Altgassen et al., 2005; Jones et al., 2009). As a result, the application of music-based interventions in language development for children with ASD has received growing attention in recent years. An increasing amount of studies suggests that social responsiveness and language development are improved by the use of musical interventions; namely, both language development and social skills can be targeted at when autistic children are engaged in various musical activities (Lim, 2011; A. Paul et al., 2015; Vaiouli et al., 2015; Williams et al., 2024). There is evidence suggesting that people with ASD are able to identify emotions in music (e.g., Quintin, 2019), and musical therapies contribute to the improvement in behavior, neural connectivity, social communication, and caregiver–child interaction among autistic children (for a review, see Gassner et al., 2022).

Growing evidence suggests that musical training enhances neural processing of speech, suggesting that music-based interventions can improve linguistic processing, including prosody (for a review, see Kraus & Chandrasekaran, 2010; Neves et al., 2022). Functional magnetic resonance imaging and electroencephalogram (EEG) studies show that music training can shape several aspects of cortical auditory processing, including those related to instrumental and pure-tone perception, and melody and rhythm perception (Hyde et al., 2011; Li et al., 2018; Moreno et al., 2009; Zendel et al., 2019). To be specific, the EEG study by Moreno et al. (2009) demonstrated that musical training was linked to a greater amplitude of a long-lasting positivity in response to little changes in sentential intonation, as well as a larger N300 amplitude during the perception of small pitch variations in nonspeech melodies. Zendel et al. (2019), similarly, found that music training increased N1 amplitude during speech perception in noisy context and enhanced a positive-going activity during word repetition. These findings provide positive evidence for the notion that music and speech share neurocognitive pathways (Peretz et al., 2015). This phenomenon, often referred to as cross-domain effects, highlights how training in one cognitive domain (e.g., music) can transfer and enhance performance in another (e.g., speech). In this case, engaging with music, which demands high precision in pitch, rhythm, and timing, can lead to neuroplastic changes that benefit auditory and linguistic processing, particularly in areas such as prosody, where pitch and rhythm are also critical (Patel, 2011, 2014). This explanation has been further expanded into the OPERA hypothesis (Patel, 2011, 2014); namely, in addition to the shared neural networks between music and speech processing (Overlap) and the higher precision required by music perception (Precision), music is also emotionally charged as speech (Emotion) while involving repetitive practice (Repetition) and requiring focused attention (Attention).

Sung speech, which blends music and speech, is particularly promising for improving prosodic abilities including focus marking among autistic children. First, speech and songs, compared to instrumental music, share even larger acoustic similarities as both of them are produced through human vocal tract (Sundberg, 1977). In fact, the natural combination of words and melodies in songs contributes to child language learning in various societies. Cross-linguistically, child-directed speech (i.e., "parentese") often features the blend of music and language elements including exaggerated use of pitch and special rhythms, which helps the early acquisition of both speech and music (Papoušek, 2007). Children's engagement in nursery rhymes, songy speech, and self-invented songs also contributes to the context for socialization and communication with other peers and adults (Campbell, 2010). Lying in the middle of the music-to-speech continuum, songs allow various combinations of music and speech (der Nederlanden et al., 2023). In this way, when applied to speech interventions, the music melodies and rhythms in songs can be designed to mimic the use of prosodic correlates in focus marking (e.g., *F*0 level and range, duration and intensity). As proposed in OPERA, the higher precision required in the processing of music melodies in songs may also heighten the trainees' speech-processing precision (Patel, 2014). In this way, the acquisition of prosodic patterns in speech may be reinforced by the music in songs, and the sung speech interventions may generate promising results.

Second, the preference to music over speech and the enhanced musical skills among autistic children may add to the effectiveness of sung speech interventions (for a review, see Simpson & Keen, 2011). On the one hand, when listening to sung speech, their preference for musical stimuli may increase their orientation and attention to sung speech. For instance, compared to their typically developing (TD) peers, autistic children have larger connectivity and activation in the left inferior frontal gyrus when listening to songs than to speech (Lai et al., 2012). Sharda et al. (2015) also found increased frontotemporal connectivity among the autistic children when they are listening to sung words rather than spoken words. Since it is reported that the language and sociocommunicative deficits found in autistic children are related to atypical left–right lateralization in frontotemporal regions, interrupted frontotemporal circuits, and reduced number of streamlines in the anterior and long segments of the accurate fasciculus in the left hemisphere (for children, see Conti et al., 2017; for adults, see Catani et al., 2016), the aforementioned positive changes brought by sung speech may facilitate language learning in children with ASD. On the other hand, behavioral studies also report particularly advanced pitch-processing skills in music among the autistic children, which may further add to the potential efficacy of sung speech intervention targeting speech prosody. Individuals with ASD show better recognition, discrimination, and memory of musical chords and melodic contours (Heaton, 2003; Heaton et al., 2008; Jiang et al., 2015; Mottron et al., 2000). L. Wang et al. (2022) found that autistic children performed better at discriminating between the rising and falling musical glides than on speech utterances expressing statements (i.e., a falling intonation) and questions (i.e., a rising intonation), but these children were still able to discriminate and imitate the statement–question intonations and showed developmental changes with ages. Such findings further suggest that the pitch sensitivity autistic children have in music can facilitate their speech prosody performance, and the combination of speech and music in songs may further add to such facilitation.

The review so far pointed to the great potential of sung speech interventions in improving children's speech prosody performance, but as relatively new methods, sung speech has not yet been fully integrated into the established interventions targeting at speech prosody, such as vivo modeling, video modeling, parent-implemented instruction, peer-mediated instruction and intervention, pivotal response training, reinforcement, and social narratives (for a review, see Holbrook & Israelsen, 2020). Instead, the existing literature has focused on either the neural mechanism behind song processing or the general effects of music therapies on autistic children's sociocommunicative development (for reviews, see James et al., 2015; Sharda et al., 2018). Moreover, despite the huge increase in the number of multilingual populations, the existing intervention focused on native rather than nondominant languages. To fill in the gap, the present study employed sung speech training and tested its effects on the improvement of Cantonese-speaking autistic children's expressive use of speech prosody in nondominant speech. We narrowed our targets to the use of prosodic correlates in Mandarin focus marking by Cantonese-speaking autistic children.

Prosodic Focus Marking by Autistic Children

Cross-linguistically, speech prosody can be used to mark focus, that is, directing the listeners' sentence to the new or important information (Lambrecht, 1994). Based on the size of the focal domain, there are broad and narrow focus, corresponding to the focus falling on the whole utterance and the focus on individual words (Halliday, 1967; Ladd, 2008). Furthermore, (narrow) focus can be used to indicate direct rejection of alternatives, which is known as contrastive or corrective focus, and noncontrastive narrow focus is used to answer *wh*-questions (Sahkai et al., 2013). Prosodically, narrow focus is often marked by on-focus expansion (OFE), that is, the exaggeration of various prosodic correlates of the on-focus constituents. For instance, these constituents are often realized with raised *F*0 level, increased *F*0 range, prolonged duration, and amplified intensity. In languages such as English and Mandarin, focus is also marked by post-focus compression (PFC), which is a prosodic focus-marking strategy that reduces *F*0 range and intensity of the post-focus syllables (Xu, 1999; Xu et al., 2012). In other words, many languages reduce the prosodic prominence of the post-focus components so as to increase the prominence of the on-focus ones. However, the more complicated tone system in Cantonese may restrict the usage of *F*0 in prosodic marking in this language due to the need to preserve tonal contours. Several studies have suggested that durational expansion rather than *F*0 exaggeration is the primary cue to signal focus in Cantonese, and PFC is not an established focus-marking strategy in Cantonese either (Fung & Mok, 2018).

Children with ASD also face difficulties in prosodic focus marking. In nontonal languages such as English, autistic children can mark on-focus syllables with greater intensity and duration but use $F0$ cues less proficiently (Diehl & Paul, 2012; Nadig & Shaw, 2015; R. Paul et al., 2008). In tone languages such as Cantonese, autistic children also have difficulty producing sufficient OFE (S. Chen et al., 2024). Such prosody deficits are observed not only in their native language production but also in non-dominant languages. According to a recent study by B. X. Wang et al. (2024) on native Cantonese-speaking children who also speak English and Mandarin, children with ASD failed to use PFC in English focus marking in the same way as their TD peers. S. Chen et al. (2025) also found that compared to the native Mandarin-speaking TD children, the native Cantonese-speaking autistic children who learn Mandarin as their third language also preferred OFE over PFC in Mandarin focus marking but did not produce as evident OFE or used as many prosodic correlates; that is, autistic children did not make use of intensity as TD children did.

Research Questions

The present training was targeted at the native Cantonese-speaking autistic children who speak Mandarin as their third language, aiming to test "whether and how the sung speech training improves the prosody focus marking in Mandarin by the Cantonese-speaking children with ASD." In particular, we would like to investigate the following questions:

Research Question 1: Does the sung speech training have a positive cross-domain effect, that is, helping autistic children improve their focus marking strategies including OFE and PFC in speech production?

Research Question 2: Does sung speech training increase the number of prosodic correlates autistic children use in both strategies OFE and PFC (i.e., $F0$ level and range, duration, and intensity)?

Method

Participants

Eighteen native Cantonese-speaking children with ASD (Cantonese-speaking ASD; 12 boys, six girls), 18 native Cantonese-speaking children with typical development (Cantonese TD; 13 boys, five girls), and 20 native Mandarin-speaking children with typical development (Mandarin TD; 14 boys, six girls) participated in the present experiment. Table 1 presents demographic information of the participants. All of the autistic participants in the experiment were formally diagnosed with ASD, and none was diagnosed of or suspected to have any other disorders such as attention-deficit/hyperactivity disorder. No TD participants had any or suspected of having any speech or language disorders. According to their self-reported data, all the native Cantonese-speaking participants learned English earlier or at the same time as Mandarin but used Cantonese and English

Table 1. Demographic information and test scores of the participants.

Demographic variable		Cantonese ASD	Cantonese TD	Mandarin TD
Age		9.46 ± 1.52	9.85 ± 1.15	9.44 ± 1.49
English learning age		2.86 ± 0.59	2.89 ± 1.02	5.55 ± 1.16
Mandarin learning age		3.95 ± 1.63	3.33 ± 1.14	4.33 ± 1.67
Musical training age		5.77 ± 1.40	5.70 ± 1.16	5.67 ± 1.80
Musical training duration		2.62 ± 2.09	3.90 ± 2.08	2.56 ± 1.40
IQ score		104.44 ± 14.47	108.11 ± 17.60	117.95 ± 13.78
WISC-V Chinese score	Overall	28.33 ± 11.49	36.50 ± 10.09	45.05 ± 7.39
	Similarities	20.11 ± 7.96	22.00 ± 9.60	28.70 ± 5.29
	Vocabulary	22.56 ± 12.34	34.00 ± 8.77	36.80 ± 6.57
	Comprehension	14.22 ± 8.06	21.72 ± 5.78	23.45 ± 6.11
ADOS-2 score	Overall	14.35 ± 5.33		
	Communication	5.00 ± 1.97		
	Reciprocal social interaction	9.35 ± 3.67		

Note. ADOS-2 does not have a specific cutoff score that definitively determines whether a participant has autism, but higher scores indicate more severe autism symptoms. According to Lord et al. (2012), scores of 1–4 indicate no worse than mild autism symptoms, scores of 5–7 indicate moderate autism symptoms, and scores of 8 or higher indicate high levels of autism symptom. ASD = participants with autism spectrum disorder; TD = typically developing participants; WISC-V = Wechsler Intelligence Scale for Children–Fifth Edition; ADOS-2 = Autism Diagnostic Observation Schedule–Second Edition.

as their primary communication languages. The analysis of variance (ANOVA) tests indicated no significant age differences among the three participant groups, $F(2, 53) = 0.485$, $p = .618$, or in the age of acquisition of Mandarin, $F(1, 30.39) = 1.3153$, $p = .1982$, between the two native Cantonese-speaking groups.

Participants and their parents were well informed and agreed to participate in the experiment. The parents signed the consent forms of a protocol approved by the Human Subjects Ethics Subcommittee at The Hong Kong Polytechnic University on behalf of the child participants. All participants were compensated for participating in the experiment.

Tests

All participants were tested using part of Raven's Progressive Matrices (Raven & Raven, 2003) and Wechsler Intelligence Scale for Children (WISC; Wechsler & Kodama, 1949) for intelligence and language evaluations, respectively. Autistic participants were also assessed with the Autism Diagnostic Observation Schedule–Second Edition (Lord et al., 2012). Tests results are also summarized in Table 1. The ANOVA test indicated a significant effect of participant group on IQ, $F(2, 53) = 4.334$, $p < .05$, and the overall WISC score, $F(2, 53) = 14.01$, $p < .001$. Post hoc comparisons showed that the Mandarin group was of significantly higher IQ than the native Cantonese-speaking children with ASD ($p < .05$), as well as significantly higher WISC scores than the Cantonese-speaking group with ASD ($p < .001$) and typical development ($p < .05$). Post hoc comparisons were conducted using Tukey's honestly significant difference test.

Materials

The experiment consisted of pre- and posttraining testing sessions and sung speech training sessions. In total, 12 five-syllable subject–verb–object sentences, together with 12 pictures depicting the content of these sentences, were used in both the pre- and posttraining sessions. A series of questions were designed to elicit the desired types of focus on subject, verb, object, or the entire sentences.

Six sentences were selected from the same 12 sentence stimuli as the training stimuli. They were evenly divided into two sets and used in two separate training sessions, respectively. A professional composer was invited to compose the music melody according to the pitch and rhythmic features of each types and position of focus (i.e., broad focus, as well as contrastive and noncontrastive narrow tone on subject, verb, and objects). The details of the music are presented in Supplemental Materials S1–S3. Two professional female signers (a répétiteur in Beijing Contemporary Music Academy and a master student in Central Conservatory of Music) were invited to record the question–answer pairs in a soundproof booth at the speech lab of The Hong Kong Polytechnic University. The singer singing the answers was instructed not only to sing according to the melodies and rhythmic patterns but also to sing the constituents on focus with higher intensity as illustrated on the score. Audio Technica ATone 2035 condenser microphone and Steinberg UR22mkII USB Audio Interface were used to record the sung speech with a sample rate of 44100 Hz in Audacity.

Procedure

All the experimental sessions were conducted and recorded in the same soundproof booth using the same equipment. Stimuli in both training and testing sessions were presented through loudspeakers as some autistic children found headphones uncomfortable or distressing. All participants were allowed to adjust the loudness to a level they found comfortable during the practice session so as to create a supportive and accessible environment, especially for autistic children with high sensory sensitivity.

In the testing session(s) attended by all groups, a practice session was conducted before the formal testing to familiarize the participants with the stimuli and reduce production errors in the testing trials. We also made the 12 target sentence stimuli (mentioned in the material session) into four blocks to reduce the memory load so that all child participants were able to memorize the sentences describing the pictures with no errors. 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).

The game, "Under the Shape" (A. Chen, 2011), was used to elicit natural responses from the participants. Participants were presented with a sequence of pictures on the computer screen, and they needed to answer the question asked by the experimenter according to the picture (see Figure 1). If a participant made a mistake in answering the question, the experimenter would ask the question again rather than simply ask for a correction so as to keep the naturalness of focus marking in the responses. For trials that contained a mistake, all the participants were able to correct themselves within three times of repetitions. Each question–picture combination was repeated twice in a random order, resulting into 168 sentences produced by each individual (= 12 Stimuli × [Broad Focus + 3 Position (S-V-O) × 2 Narrow Focus] × 2).

All the participants attended the testing session. Both the Cantonese-speaking ASD training group and the

Figure 1. Illustration of the game "Under the Shape." Take the noncontrastive focus trial as an example. The above picture with a gray square covering the person's hand was first shown to the participant. The experimenter asked in Mandarin "/wɑŋ35 uən35 tʰi35 ʂən35 mɤ3/" (What is Wang Wen carrying?). After the question was asked, the gray area was removed. The participant was expected to answer the target Mandarin sentence with narrow focus on the object, "/wɑŋ35 uən35 tʰi35 [pʰaɪ35 tɕʰjoʊ35]$_{focus}$/" (Wang Wen is carrying [(the) volleyball]$_{focus}$).

Cantonese-speaking TD control group were tested twice, with a 1-week interval between sessions. The ASD training group received two training sessions during this interval, each held 1 week apart. The training program consisted of three structured phases, as outlined in Table 2. The Cantonese-speaking TD control group did not receive any training but was also tested twice, with the same 1-week interval, to control for potential practice effects between pre- and posttests. This design allowed us to distinguish improvements due to training from those resulting from repeated exposure to the test materials. The native Mandarin-speaking TD control group, however, participated in a single testing session. Given their native proficiency, their performance was near ceiling, and repeated testing was not deemed necessary or informative in this group.

Data Analysis

The data of the pre- and posttraining testing sessions were analyzed using Praat (Boersma & Weenink, 2023). The five-syllable sentences produced were manually segmented following the procedure of segmentation written by Jangjamras (2011), with prosodic information extracted using ProsodyPro (Xu, 2013). The extracted values were also manually checked. The *F*0 range (i.e., the difference between the maximum and minimum *F*0 value), the mean *F*0, the duration, and the mean intensity of the sonorant part were calculated for each syllable so as to measure how different types of focus were marked by participants in each group. Syllables with over 50% data loss (i.e., no meaningful *F*0 values can be extracted, mainly due to creakiness or large environmental noises caused by child's movements) were excluded as no meaningful mean *F*0 or *F*0 range could be extracted from such syllables (except for one Mandarin-speaking child who had 12% data loss, all participants had data loss less than 4%, resulting in

Table 2. The procedure of the training sessions.

Phase	Stimuli (question–answer pairs presented as sung speech)		Response required from the participants	
	Matchedness	**Randomization**	**Task**	**Action**
Phase I (self-paced learning)	Matched pairs only	—	Identify the focus type and position while listening to the pairs	When hearing a question, the participants need to get ready to press the buttons that represent different focus types (times to press the buttons: no pressing for broad focus, once for narrow focus, twice for contrastive focus) and position on either subject, verb, object, or the entire utterance
Phase II (supervised learning)	Matched and mismatched pairs	A matched pair followed by a mismatched pair	Listen to both matched and mismatched pairs for a comparison	Pressing on the buttons with corresponding labels ("matched" and "mismatched")
Phase III (pseudotesting)	Matched and mismatched pairs	Randomized	Judge whether the question and answer in each trial match with each other	Pressing on the buttons with corresponding labels ("matched" and "mismatched")

Note. The participants will be given feedback on their responses. Matchedness describes whether the prosody of the answers matched the question, for example, for question "/uən55 ɪŋ55 mo55 ʂən35 mə0/" (What is Wen Ying touching?), "/uən55 ɪŋ55 mo55 [ɕjɛn55 xwa55]$_{focus}$/" (Wen Ying is touching [(the) flower]$_{focus}$) is a matched answer, whereas "/[uən55 ɪŋ55]$_{focus}$ mo55 ɕjɛn55 xwa55/" ([Wen Ying]$_{focus}$ is touching (the) flower) is a mismatched answer.

approximately 840 analyzable syllables per participant and session). This high observation-to-subject ratio notably increases the statistical power of the linear mixed-effects (LME) models introduced below, even with a relatively small sample size.

We aimed to compare the pre- and posttraining focus-marking performance of the Cantonese-speaking ASD training group with the performance of the other two TD groups. Therefore, we included testing group (Cantonese ASD pretraining vs. Cantonese ASD post-training vs. Cantonese TD Test 1 vs. Cantonese TD Test 2 vs. Mandarin TD), focus condition, and lexical tone embedded in the syllable (high-level T1 vs. rising T2 vs. falling T4) as the fixed effects. Here, focus condition was defined as the relative position to focus of a syllable, that is, carrying broad focus (*Broad Focus*), preceding a syllable carrying contrastive or noncontrastive narrow focus (*Pre-Focus*), carrying noncontrastive narrow focus (*Noncontrastive Focus*), carrying contrastive narrow focus (*Contrastive Focus*), and following a syllable carrying contrastive or noncontrastive narrow focus (*Post-Focus*).

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). To further mitigate the risk of Type I errors associated with multiple comparisons—especially given the number of conditions and interactions tested—we employed Tukey-adjusted post hoc comparisons using the emmeans package (Lenth et al., 2023) in R. The Tukey correction is a conservative method that controls the familywise error rate, ensuring that any reported pairwise differences among factor levels are statistically reliable. All specific p values of the post hoc comparisons were reported in Appendix, and the following Results section only reported significance levels. The full details of the LME models are presented alongside the results. Since none of testing group, focus condition, and lexical tone were of significant effects on duration, only results on the $F0$ range, mean $F0$, and intensity were reported.

Results

F0 Range

The optimal LME model showed that the effects of testing group, $\chi^2(4) = 15.5220$, $p < .001$; focus condition, $\chi^2(4) = 41.60$, $p < .001$; and lexical tone, $\chi^2(2) = 51.37$, $p < .001$, as well as their two-way (Testing Group × Focus Condition, $\chi^2(16) = 5.79$, $p < .001$; Focus Condition × Lexical Tone, $\chi^2(8) = 6.69$, $p < .001$; Testing Group × Lexical Tone, $\chi^2(8) = 117.31$, $p < .001$) and three-way, $\chi^2(32) = 3.13$, $p < .001$, interactions were significant.

Differences Between Testing Groups

Across Tones 1 and 4, all Cantonese-speaking groups (ASD and TD) consistently produced a significantly or marginally larger $F0$ range than the Mandarin TD group (ps < .01), regardless of focus condition. For Tone 2 (rising tone), more nuanced differences were observed. In Test 2, the Cantonese TD group produced T2 with smaller $F0$ range overall—significantly smaller than that of the Cantonese ASD group under both broad and narrow focus conditions (ps < .05), but not significantly different from that of the Mandarin TD group. This suggests a potential de-exaggeration of focus marking in the Cantonese-speaking TD group in the second testing session, possibly due to increased familiarity. Notably, no significant differences were found between pre- and posttest within either the Cantonese TD or ASD group overall. However, when examining focus-marking strategies, consistent patterns emerged that suggest meaningful training effects in the ASD group.

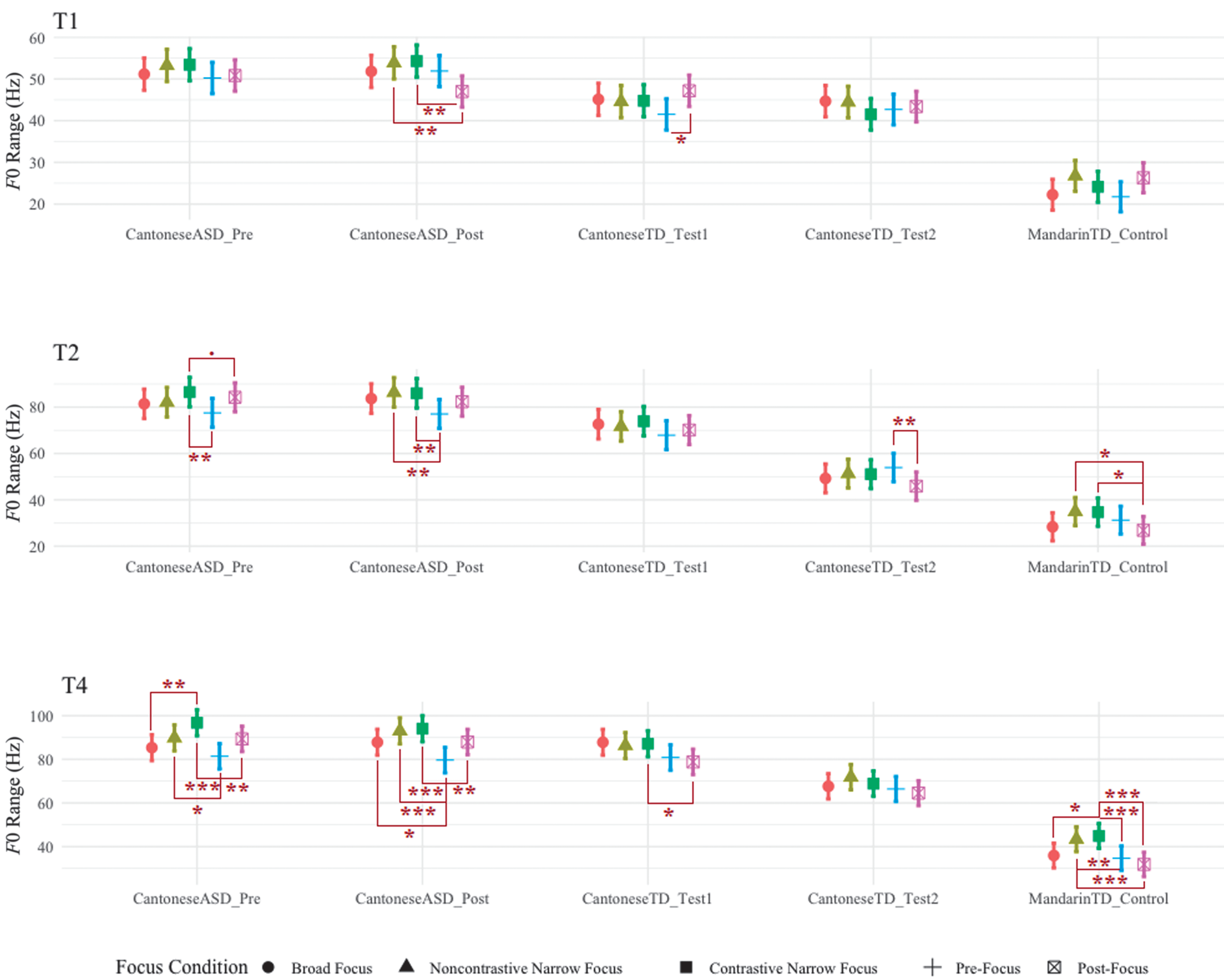
Group-Specific Strategies and Training Effects

The most prominent training-related change was an improvement in OFE in the Cantonese ASD group (see Figure 2). After training, this group showed significantly larger $F0$ range in narrow focus conditions (both contrastive and noncontrastive) compared to the pre-focus or post-focus conditions. Since such improvement was not seen between the two tests of the Cantonese TD groups, it can be attributed to the effects of sung speech training, rather than the repetition of the test itself.

Neither the Mandarin TD nor the Cantonese TD showed clear OFE or PFC in the $F0$ range of the high-level T1, probably because the $F0$ range of the level tone may not be as an effective index of focus-marking strategies as that of the contour tones. However, the Cantonese ASD group, after training, demonstrated some OFE by producing T1 with significantly larger $F0$ range in the two narrow focus conditions than in the post-focus condition.

With regard to contour tones, for the rising T2, the Cantonese ASD before training only marked contrastive focus with significantly larger $F0$ range than *Pre-Focus* and marginally larger $F0$ range than *Post-Focus*; the $F0$ range of T2 carrying noncontrastive focus was even smaller than post-focus T2. After training, the $F0$ range of T2 in both contrastive and noncontrastive focus conditions was significantly larger than in the pre-focus condition and also higher than in the post-focus condition, although the latter difference did not reach statistical significance. For the falling T4, prior to training, the ASD group marked contrastive focus with a significantly larger $F0$ range than both *Pre-Focus* and *Broad Focus*; after training, T4 produced under both narrow and broad focus

Figure 2. Estimated marginal means of *F*0 range in each test group. Significance: *** represents $p < .001$, ** represents $p < .01$, * represents $p < .05$, and • represents $p < .055$—marginal significance. *F*0 = fundamental frequency; ASD = autism spectrum disorder; TD = typically developing.



conditions had significantly larger *F*0 range than *Pre-Focus*. It seemed that the training slightly reduced the prominence of contrastive focus marking but boosted the ability of noncontrastive marking abilities in the ASD group. By contrast, the Cantonese TD groups, in either Test 1 or Test 2, still showed some tendency of OFE; that is, T4 on narrow focus showed larger *F*0 range than other conditions, and PFC (i.e., post-focus T4) showed smaller *F*0 range than other conditions, but only the difference between *Contrastive Focus* and *Post-Focus* in Test 1 reached statistical significance. By comparing the performance of Cantonese-speaking ASD with Cantonese TD, it can be seen that sung speech training led to improved OFE of *F*0 range.

By contrast, the use of PFC was not reliably observed in Cantonese-speaking groups. For the contour tones T2 and T4, the ASD group consistently produced the smallest *F*0 range in the pre-focus rather than the post-focus condition, both before and after training. The Cantonese-speaking TD group did not produce T2 or T4 with *F*0 range significantly smaller in the post-focus condition compared to the other conditions.

Nevertheless, it was the Mandarin TD group that showed the clearest OFE and PFC strategies when producing the two contour tones, as the *F*0 range of T2 and T4 was the smallest in the post-focus condition but the largest in the two narrow focus conditions. To be specific, for T2, the *F*0 range in the post-focus condition was significantly smaller than *Contrastive Focus* and *Noncontrastive Focus.* For T4, *Contrastive Focus* and *Noncontrastive Focus* was marked with significantly larger *F*0 range than *Pre-Focus* and *Post-Focus*, and contrastive focus was made even more evident as the difference between contrastive and broad focus was also significant.

Mean F0

The optimal LME model showed that the effects of testing group, $\chi^2(4) = 17.21$, $p < .001$; focus condition, $\chi^2(4) = 250.36$, $p < .001$; and lexical tone, $\chi^2(2) = 33.69$, $p < .001$, as well as their two-way (Testing Group × Focus Condition, $\chi^2(12) = 60.89$, $p < .001$; Focus Condition × Lexical Tone, $\chi^2(8) = 14.95$, $p < .001$; Testing Group × Lexical Tone, $\chi^2(6) = 99.39$, $p < .001$) and three-way, $\chi^2(24) = 6.93$, $p < .001$, interactions were significant.

Differences Between Testing Groups

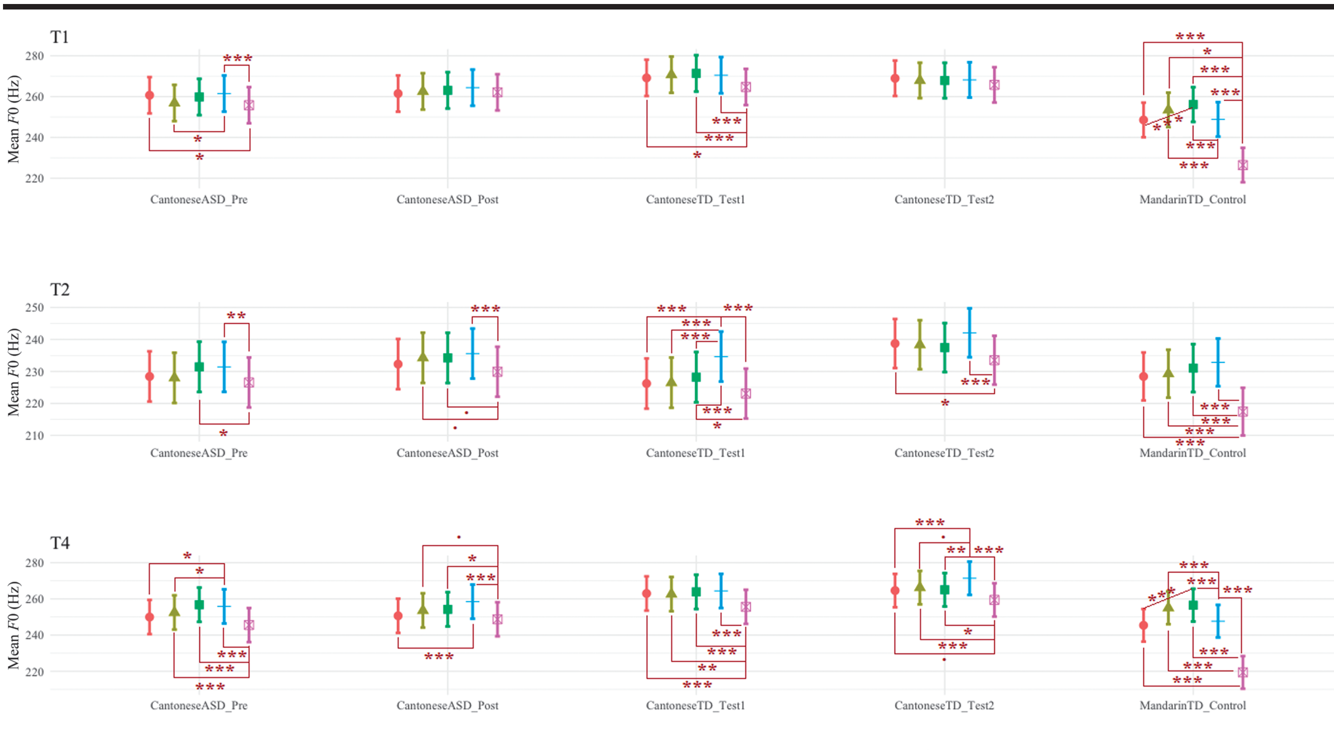
Significance was only found between the pre- and posttraining tests attended by Cantonese ASD, indicating evident training effects in the marking of noncontrastive focus. To be specific, they produced significantly higher T1 in the posttraining test in the noncontrastive focus ($p < .01$) and post-focus ($p < .001$) conditions, as well as significantly higher T2 in the posttraining test in the noncontrastive focus and pre-focus conditions ($ps < .05$).

Group-Specific Strategies and Training Effects

All testing groups showed either significant or trending post-focus lowering across all three tones, with the most notable training effect observed in OFE under the noncontrastive focus condition (see Figure 3). In terms of T1, before training, *Pre-Focus* was the highest, being significantly higher than *Pre-Focus* as well as *Noncontrastive Focus* and *Broad Focus.* In the posttraining test, *Contrastive Focus* and *Noncontrastive Focus* were of similar mean *F*0 to *Pre-Focus*, and all were higher than *Broad Focus* and *Post-Focus*, though none of these differences reached statistical significance. The Cantonese TD controls, by contrast, only showed significant post-focus lowering in Test 1, while the Mandarin TD showed significant OFE in *Contrastive Focus* and *Noncontrastive Focus*, being significantly higher than *Pre-Focus* and *Post-Focus*, as well as post-focus lowering with *Post-Focus* being significantly lower than the other focus conditions. Furthermore, contrastive focus was again marked more evidently as the difference between contrastive and broad focus was also significant.

For T2, the training also helped the Cantonese ASD to mark noncontrastive with marginally higher mean *F*0 than *Post-Focus* in the posttraining test, while *Post-Focus* mean *F*0 was significantly or marginally lower than that of pre-focus and contrastive focus in both pre- and posttraining tests. The Cantonese TD controls, in Test 1, produced significantly higher *Pre-Focus* than all the other conditions but significantly lower *Post-Focus* than *Pre-Focus* and *Contrastive Focus.* In Test 2, however, the significance was only found between *Pre-Focus*, *Broad Focus,* and the lowest *Post-Focus.* The Mandarin TD also produced an evident pre-focus raising of T2, and the significance was

Figure 3. Estimated marginal means of mean *F*0 in each test group. Significance: *** represents $p < .001$, ** represents $p < .01$, * represents $p < .05$, and • represents $p < .055$—marginal significance. *F*0 = fundamental frequency; ASD = autism spectrum disorder; TD = typically developing.



seen between the other focus conditions and the lowest *Post-Focus.*

Not much training effects were seen in the mean $F0$ of T4, since before the training, contrastive and noncontrastive focus was already marked with significantly higher mean $F0$ than *Post-Focus* in the ASD group. *Pre-Focus* raising was seen in the Cantonese ASD and TD groups, but not the Mandarin TD controls; Mandarin TD adopted a focus-marking strategy similar to T1, and the pattern of significance in T4 was also the same as that in T1 in this group.

Intensity

The optimal LME model showed that the effects of testing group, $\chi^2(4) = 3.02$, $p = .022$; focus condition, $\chi^2(4) = 79.66$, $p < .001$; and their interaction, $\chi^2(16) = 47.95$, $p < .001$, were significant. Under the same focus condition, post hoc comparisons found no significant differences between testing groups.

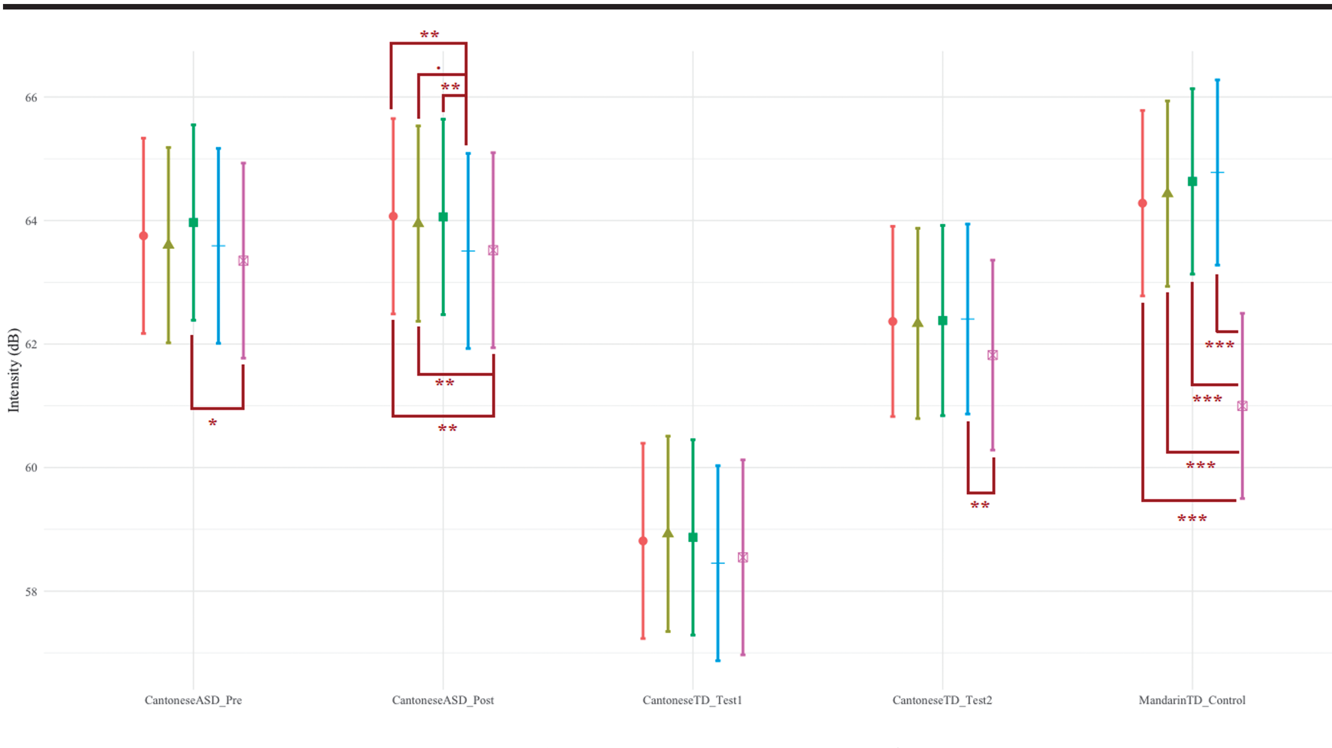
Group-Specific Strategies and Training Effects

Training effects were still more evidently found in OFE (see Figure 4). Before the training, the Cantonese ASD group only produced significantly louder *Contrastive Focus* than *Post-Focus*; after training, *Broad Focus* and *Contrastive Focus* were both significantly louder than *Pre-Focus* and *Post-Focus*, and *Noncontrastive Focus* was also marginally louder than *Pre-Focus.* Cantonese TD group produced significantly louder *Pre-Focus* than *Post-Focus*, showing a PFC tendency. The Mandarin TD, however, showed significant PFC but not OFE, because while the intensity in the post-focus condition was significantly lower than in other conditions, the highest mean intensity was found in the pre-focus condition.

Discussion

The present study investigated the effects of a short-term song-based training targeting prosodic focus-marking performance in Mandarin by native Cantonese-speaking children with ASD. The sung speech was designed in a way to boost the prosodic correlates of words carrying contrastive and noncontrastive focus using corresponding melody lines and rhythmic patterns, which reinforces prosodic patterns used in focus marking in speech. Positive effects of the speech–speech training used in this study were also observed. With regard to the two focus-marking strategies (OFE and PFC in Research Question 1), the

Figure 4. Estimated marginal means of mean intensity in each test group. Significance: *** represents $p < .001$, ** represents $p < .01$, * represents $p < .05$, and • represents $p < .055$—marginal significance. ASD = autism spectrum disorder; TD = typically developing.



training strengthened their use of OFE, especially in the noncontrastive focus marking. With regard to the prosodic correlates used in focus marking (Research Question 2), before training, the native Cantonese-speaking autistic children were already able to use *F*0 range, mean *F*0, and mean intensity to mark focus in Mandarin just like Cantonese and Mandarin TDs. This ceiling finding makes the improvements in the number of prosodic correlates used less evident than the improvement in focus-marking strategies.

Since similar improvements were not seen in the second test performance by the Cantonese TD controls, the improvements seen in the Cantonese ASD group can be largely attributed to the training effects rather than the familiarity of the test itself raised by repetition. In fact, Cantonese TD's performance in the second test did not show much improvement—the OFE of major *F*0 cues related to focus marking in Mandarin; that is, the *F*0 range of the contour tones T2 and T4 and in the mean *F*0 of the high-level T1, in fact, became less evident in the second test. The Cantonese TD controls might have felt bored or become less careful in the second test as they became more familiar with the test, leading to the aforementioned changes. The following discussion will focus more on the performance of the autistic group.

Noncontrastive Focus-Marking Improvement

Before training, the ASD group displayed a notable asymmetry in their use of OFE between contrastive and noncontrastive focus marking. While they successfully marked contrastive focus with significantly larger *F*0 range, higher mean *F*0, and larger intensity, their marking of noncontrastive focus was insufficient. The prominence of noncontrastive focus bearing syllables, measured in any of these three cues, was often less prominent than those bearing broad focus or no focus. Although the native Mandarin-speaking TD group, in many cases, did mark contrastive focus with more evident OFE significant differences between contrastive and broad focus, the syllables bearing noncontrastive focus were still produced with larger *F*0 range, higher mean *F*0, and larger intensity than those bearing broad focus. This asymmetry may stem from overall prosodic impairments often observed in autistic individuals (e.g., Diehl & Paul, 2012; McCann & Peppé, 2003; R. Paul et al., 2008). However, the failure to apply OFE to noncontrastive focus, despite using it effectively for contrastive focus, may also reflect autistic children's deficits in sociocommunication skills. Pragmatic language impairments are profound in children with ASD in assessment-based studies, which further influence their expressive use of verbal cues such as prosody (e.g., Reindal et al., 2023).

The sung speech training in the study significantly increased the use of OFE in noncontrastive focus marking in Mandarin by the Cantonese ASD group. The sung speech training's exaggerated realization of prosodic correlates in noncontrastive focus may help the autistic children to acquire such prosodic patterns and apply them to the test later. Additionally, the training may have helped them to transfer their prosodic skills to noncontrastive focus marking. The Cantonese ASD group already demonstrated knowledge of OFE in contrastive focus marking before training, so that sung speech training may not only contribute to the acquisition of the prosodic patterns but also, more importantly, enable them to recognize the pragmatic similarities between the contrastive and noncontrastive narrow focus; that is, they need to raise the prominence of the newly given information in the answers to *wh*-questions just as they raise the prominence of the contrastive/correct information. Then, they may apply similar prosodic strategies in this scenario as when they make corrections. This pattern is in line with studies such as Finnigan and Starr (2010) and Lim and Draper (2011), who report significant improvement in autistic children's pragmatic skills such as turn-taking and echoic production when using song-based stimuli. Overall, our findings suggest that song-based training contributes to the development of prosodic focus-marking abilities in autistic children, potentially aligning with previous research on the sociocommunicative benefits of music-based interventions (see Marquez-Garcia et al., 2022; Vaiouli & Andreou, 2018).

While the present study did not directly assess sociocommunicative outcomes, prior research has reported that song-based interventions can enhance therapist–child interaction, reduce negative behavior, and improve task engagement (Dieringer et al., 2017; Gee et al., 2014). These factors may have played a supportive role in the observed improvements in prosodic performance in the present study. Although further research is needed to determine whether these prosodic gains translate to broader communicative contexts, the present findings contribute to the growing body of literature on the potential of song-based training for children with ASD.

OFE Improvement

The improvement in noncontrastive focus-marking posttraining was primarily observed in OFE rather than PFC. Notably, the Cantonese-speaking autistic children in this study did not show substantial deficits in using OFE for contrastive focus marking in *F*0 cues and intensity, differing from some previous studies that reported atypical prosody use in autistic individuals (e.g., Asghari et al., 2021; McCann & Peppé, 2003; R. Paul et al., 2008). This discrepancy may stem from the focus on nondominant rather than native languages. However, B. X. Wang et al. (2024) also reported poorer prosodic focus-marking performance in English among Cantonese–English–Mandarin trilingual children with ASD than their TD peers, with TD peers consistently using more cues (duration, *F*0, intensity) to mark PFC than the ASD group, though neither group demonstrated significant OFE.

The most striking difference between our study and that of B. X. Wang et al. (2024) is that OFE was an effectively used and quickly improved focus-marking strategy by our autistic participants, whereas PFC was limited to post-focus lowering in Cantonese-speaking ASD and TD groups. The improvement in OFE and the lack of improvement in PFC may be explained by an effective first-language transfer, according to the cumulative enhancement model (CEM; Flynn et al., 2004) and scalpel model (SM; Slabakova et al., 2024). Both models argue that there is property-by-property transfer of linguistic features between first, second, and third languages, influenced more by the perceived structural similarities between languages than acquisition order. CEM emphasizes on the perceived facilitativeness of the transfer, while SM incorporates additional factors such as feature complexity and frequency. In this case, the *F*0 patterns in the sung speech melodies in training may highlight similarities between the OFE used in Cantonese and Mandarin focus marking, raising the perceived facilitativeness of the OFE transfer. In addition, OFE is more common cross-linguistically than PFC (Xu et al., 2012), which further contributes to its transfer and hence the improvement observed.

By contrast, PFC, a more complex and less frequent strategy, is not an established feature in Cantonese despite some compressing tendencies observed in post-focus syllables (e.g., Hsu et al., 2018; Lee & Barnes, 2023). Hence, in the present study, limited use was seen in PFC in neither the Cantonese ASD nor TD group. The null findings for PFC may also indicate that the training, while effective in highlighting pitch expansion in focused syllables, was less effective in cueing the suppression of post-focus prosody. This asymmetry might reflect a natural bias toward enhancing new information (OFE) rather than deemphasizing given information (PFC), especially in children with ASD, who may have difficulty with pragmatic inferences and context tracking.

Further Discussion

The observed effectiveness of the sung speech training may be explained by the potential facilitation of prosodic learning through musical components. Previous studies have suggested that some autistic individuals demonstrate strengths in identifying and memorizing melodic contours (Heaton, 2003; Jiang et al., 2015), which may support their perception and retention of pitch-based patterns. Although our study did not directly assess the music processing abilities of participants, it is possible that the melodic structure of the sung stimuli, which was aligned with focus-marking *F*0 contours, enhanced the salience of prosodic cues and contributed to participants' performance during posttraining assessment.

In addition, as previously introduced, neural activation in frontotemporal regions, that is, areas associated with sociocommunicative and prosodic processing, has been observed during sung speech listening (Lai et al., 2012; Sharda et al., 2015). Repeated exposure to sung speech stimuli like those in this study may have reinforced activation in these regions, supporting the encoding of prosodic patterns. These observations align with broader evidence suggesting partial overlap in the neural mechanisms underlying pitch processing in music and speech in both neurotypical and autistic individuals (Cheng et al., 2017; L. Wang et al., 2022). Although we did not test these neural mechanisms directly, the theoretical possibility of shared processing pathways, especially for pitch processing, may help explain how sung speech training can support prosodic learning. However, further research is needed to determine whether and how such mechanisms relate to broader communicative functions and long-term changes in autistic children.

The design of the training may have further enhanced its effectiveness by focusing solely on prosodic focus marking rather than lexical tones, minimizing the interference from tones. Lexical tones in tone languages, such as Mandarin, often complicate intonation realization (Jiang et al., 2015; Liu & Xu, 2005), resulting in the better perpetual performance of intonation in nontone languages than in tone languages by autistic children (L. Wang et al., 2022). Our participants were able to mainly pay attention to the pitch patterns on the utterance level than the lexical tones due to our design. Their posttraining performance, therefore, may be less constrained by accurate tone realization, and hence, they showed more evident use of the OFE strategy than their Cantonese-speaking TD peers in some cases.

In addition, instead of passive song listening, we also required responsive actions from the autistic children in all three phases in each training session. As reviewed by Marquez-Garcia et al. (2022), interactive responses add to the effectiveness of song-based therapy for ASD. The multimodal activities involved simultaneous processing of visual, auditory, sensory, and motoric information in the present training and, if made long term, may further contribute to the brain plasticity (Reybrouck et al., 2018), enhancing the therapeutic effects of sung speech training.

Implications and Limitations

The positive outcomes observed in this song-based training suggest potential clinical applications. Such perception-based training using song speech could improve autistic children's prosody performance in nondominant language and hence has the potential of being used as a supplement to production training widely adopted in clinical practice targeting at autistic individuals' prosody deficits (Hargrove et al., 2009). In particular, the significant improvement in the OFE of *F*0 cues indicates that the increased neural connectivity and activation during song listening among the autistic children may be retained and benefit their speech performance in long term.

We also acknowledge several limitations. This was a small-scale preliminary study with a limited number of participants and training sessions, which constrains the generalizability of the results. Additionally, some effects did not approach significance, and generalization of prosodic gains to spontaneous or naturalistic contexts was not assessed. Importantly, there was a potential IQ mismatch between the Cantonese-speaking ASD group and the Mandarin-speaking TD controls. Although this IQ difference does not directly influence the observed training-related improvements in the Cantonese ASD group, the particularly strong performance of the Mandarin TD group may, at least in part, reflect their generally higher cognitive abilities rather than native advantages. In other words, it is possible the ASD group's posttraining performance might have more closely approximated that of the native Mandarin-speaking TD group with similar IQ. Future studies should control more rigorously for cognitive factors, include generalization measures, and assess long-term outcomes. Expanding this line of research to language pairs with greater typological distance than Cantonese and Mandarin may also provide further insights into cross-linguistic and cross-modal transfer in prosody learning.

Author Contributions

Yixin Zhang: Formal analysis, Visualization, Writing – original draft, Writing – review & editing. **Si Chen:** Conceptualization, Data curation, Writing – review & editing. **Meixuan Li:** Data curation, Formal analysis. **Bin Li:** Data curation, Writing – review & editing. **Shuang Lu:** Data curation, Writing – review & editing. **Angel Chan:** Data curation, Writing – review & editing. **Haoyan Ge:** Data curation, Writing – review & editing. **Tempo Tang:** Data curation. **Zhuoming Chen:** Data curation, Writing – review & editing.

Ethics Statement

This study has been approved by the PolyU Institutional Review Board (Reference No. HSEARS20210920005).

Data Availability Statement

The data that support the findings of this study are provided as supplemental materials.

Acknowledgments

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 corresponding author. We would like to thank our composer, Michael Andreas Häringer, and our two singers, Yan Meng and Wennan Cheng. We would also like to thank Jiaqi Ma, Baobao Yang, and Luwei Ye for their help in data analysis.

References

Altgassen, M., Kliegel, M., & Williams, T. I. (2005). Pitch perception in children with autistic spectrum disorders. *British Journal of Developmental Psychology, 23*(4), 543–558. https://doi.org/10.1348/026151005X26840

American Psychiatric Association Division of Research. (2013). Highlights of changes from DSM-IV to DSM-5: Somatic symptom and related disorders. *Focus, 11*(4), 525–527. https://doi.org/10.1176/appi.focus.11.4.525

Asghari, S. Z., Farashi, S., Bashirian, S., & Jenabi, E. (2021). Distinctive prosodic features of people with autism spectrum disorder: A systematic review and meta-analysis study. *Scientific Reports, 11*(1), Article 23093. https://doi.org/10.1038/s41598-021-02487-6

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

Boersma, P., & Weenink, D. (2023). *Praat: Doing phonetics by computer* (6.3.08) [Computer software]. http://www.praat.org/

Campbell, P. S. (2010). *Songs in their heads: Music and its meaning in children's lives*. Oxford University Press.

Catani, M., Dell'Acqua, F., Budisavljevic, S., Howells, H., Thiebaut de Schotten, M., Froudist-Walsh, S., D'Anna, L., Thompson, A., Sandrone, S., Bullmore, E. T., Suckling, J., Baron-Cohen, S., Lombardo, M. V., Wheelwright, S. J., Chakrabarti, B., Lai, M. C., Ruigrok, A. N., Leemans, A., Ecker, C., . . . Murphy, D. G. (2016). Frontal networks in adults with autism spectrum disorder. *Brain, 139*(2), 616–630. https://doi.org/10.1093/brain/awv351

Chen, A. (2011). Tuning information packaging: Intonational realization of topic and focus in child Dutch. *Journal of Child Language, 38*(5), 1055–1083. https://doi.org/10.1017/S0305000910000541

Chen, S., Zhang, Y., Li, M., Li, B., Lu, S., Chan, A., Ge, H., Tang, T., & Chen, Z. (2025). Acquisition of speech prosody in a non-native tone language by children with and without autism spectrum disorder. *Journal of Autism and Developmental Disorders*. Advance online publication. https://doi.org/10.1007/s10803-024-06698-4

Chen, S., Zhang, Y., Zhou, F., Chan, A., Li, B., Li, B., Tang, T., Chun, E., & Chen, Z. (2024). Focus-marking in a tonal language: Prosodic differences between Cantonese-speaking children with and without autism spectrum disorder. *PLOS ONE, 19*(7), Article e0306272. https://doi.org/10.1371/journal.pone.0306272

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). https://doi.org/10.1177/2041669517711200

Conti, E., Mitra, J., Calderoni, S., Pannek, K., Shen, K. K., Pagnozzi, A., Rose, S., Mazzotti, S., Scelfo, D., Tosetti, M., Muratori, F., Cioni, G., & Guzzetta, A. (2017). Network overconnectivity differentiates autism spectrum disorder from other developmental disorders in toddlers: A diffusion MRI study. *Human Brain Mapping, 38*(5), 2333–2344. https://doi.org/10.1002/hbm.23520

der Nederlanden, C. M. V. B., Qi, X., Sequeira, S., Seth, P., Grahn, J. A., Joanisse, M. F., & Hannon, E. E. (2023). Developmental changes in the categorization of speech and song. *Developmental Science, 26*(5), Article e13346. https://doi.org/10.1111/desc.13346

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

Dieringer, S. T., Porretta, D. L., & Sainato, D. (2017). Music and on-task behaviors in preschool children with autism spectrum disorder. *Adapted Physical Activity Quarterly, 34*(3), 217–234. https://doi.org/10.1123/apaq.2015-0033

Finnigan, E., & Starr, E. (2010). Increasing social responsiveness in a child with autism: A comparison of music and non-music interventions. *Autism, 14*(4), 321–348. https://doi.org/10.1177/1362361309357747

Flynn, S., Foley, C., & Vinnitskaya, I. (2004). The cumulative-enhancement model for language acquisition: Comparing adults' and children's patterns of development in first, second and third language acquisition of relative clauses. *International Journal of Multilingualism, 1*(1), 3–16. https://doi.org/10.1080/14790710408668175

Fung, H. S. H., & Mok, P. P. K. (2018). Temporal coordination between focus prosody and pointing gestures in Cantonese. *Journal of Phonetics, 71,* 113–125. https://doi.org/10.1016/j.wocn.2018.07.006

Gassner, L., Geretsegger, M., & Mayer-Ferbas, J. (2022). Effectiveness of music therapy for autism spectrum disorder, dementia, depression, insomnia and schizophrenia: Update of systematic reviews. *European Journal of Public Health, 32*(1), 27–34. https://doi.org/10.1093/eurpub/ckab042

Ge, H., Lee, A. K. L., Yuen, H. K., Liu, F., & Yip, V. (2024). Bilingual exposure might enhance L1 development in Cantonese–English bilingual autistic children: Evidence from the production of focus. *Autism, 28*(7), 1795–1808. https://doi.org/10.1177/13623613231207449

Gee, B. M., Thompson, K., & St John, H. (2014). Efficacy of a sound-based intervention with a child with an autism spectrum disorder and auditory sensory over-responsivity. *Occupational Therapy International, 21*(1), 12–20. https://doi.org/10.1002/oti.1359

Gilhuber, C. S., Raulston, T. J., & Galley, K. (2023). Language and communication skills in multilingual children on the autism spectrum: A systematic review. *Autism, 27*(6), 1516–1531. https://doi.org/10.1177/13623613221147780

Gussenhoven, C., & Chen, A. (Eds.). (2021). *The Oxford handbook of language prosody*. Oxford University Press.

Halliday, M. A. K. (1967). Notes on transitivity and theme in English: Part 2. *Journal of Linguistics, 3*(2), 199–244. https://doi.org/10.1017/S0022226700016613

Hargrove, P., Anderson, A., & Jones, J. (2009). A critical review of interventions targeting prosody. *International Journal of Speech-Language Pathology, 11*(4), 298–304. https://doi.org/10.1080/17549500902969477

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

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

Holbrook, S., & Israelsen, M. (2020). Speech prosody interventions for persons with autism spectrum disorders: A systematic review. *American Journal of Speech-Language Pathology, 29*(4), 2189–2205. https://doi.org/10.1044/2020_AJSLP-19-00127

Hsu, Y. Y., & German, J. S.. (2018, November). Prosodic organization and focus realization in taiwan mandarin. In *The 32nd Pacific Asia Conference on Language, Information and Computation (PACLIC)*.

Hubbard, D. J., Faso, D. J., Assmann, P. F., & Sasson, N. J. (2017). Production and perception of emotional prosody by adults with autism spectrum disorder. *Autism Research, 10*(12), 1991–2001. https://doi.org/10.1002/aur.1847

Hübscher, I., & Prieto, P. (2019). Gestural and prosodic development act as sister systems and jointly pave the way for children's sociopragmatic development. *Frontiers in Psychology, 10,* Article 1259. https://doi.org/10.3389/fpsyg.2019.01259

Hyde, K. L., Zatorre, R. J., & Peretz, I. (2011). Functional MRI evidence of an abnormal neural network for pitch processing in congenital amusia. *Cerebral Cortex, 21*(2), 292–299. https://doi.org/10.1093/cercor/bhq094

James, R., Sigafoos, J., Green, V. A., Lancioni, G. E., O'Reilly, M. F., Lang, R., Davis, T., Carnett, A., Achmadi, D., Gevarter, C., & Marschik, P. B. (2015). Music therapy for individuals with autism spectrum disorder: A systematic review. *Review Journal of Autism and Developmental Disorders, 2*(1), 39–54. https://doi.org/10.1007/s40489-014-0035-4

Jangjamras, J. (2011). *Perception and production of English lexical stress by Thai speakers*. University of Florida.

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

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

Kraus, N., & Chandrasekaran, B. (2010). Music training for the development of auditory skills. *Nature Reviews Neuroscience, 11*(8), 599–605. https://doi.org/10.1038/nrn2882

Ladd, D. R. (2008). *Intonational phonology*. (2nd ed.). Cambridge University Press. https://doi.org/10.1017/CBO9780511808814

Lai, G., Pantazatos, S. P., Schneider, H., & Hirsch, J. (2012). Neural systems for speech and song in autism. *Brain, 135*(3), 961–975. https://doi.org/10.1093/brain/awr335

Lambrecht, K. (1994). *Information structure and sentence form: Topics, focus, and the mental representations of discourse referents*. Cambridge University Press. https://doi.org/10.1017/CBO9780511620607

Lau, J. C., Losh, M., & Speights, M. (2023). Differences in speech articulatory timing and associations with pragmatic language ability in autism. *Research in Autism Spectrum Disorders, 102,* Article 102118. https://doi.org/10.1016/j.rasd.2023.102118

Lee, C. K., & Barnes, J. (2023). Contour shape and prosodic focus marking: The case of Hong Kong Cantonese. In *Proceedings of the 20th International Congress of Phonetic Sciences* (pp. 1300–1304).

Lenth, R. V., Buerkner, P., Giné-Vázquez, I., Herve, M., Jung, M., Love, J., Miguez, F., Riebl, H., & Singmann, H. (2023). *emmeans: Estimated marginal means, aka least-squares means* (1.8.4-1) [Computer software]. https://CRAN.R-project.org/package=emmeans

Li, Q., Wang, X., Wang, S., Xie, Y., Li, X., Xie, Y., & Li, S. (2018). Musical training induces functional and structural auditory-motor network plasticity in young adults. *Human Brain Mapping, 39*(5), 2098–2110. https://doi.org/10.1002/hbm.23989

Lim, H. A. (2011). *Developmental speech-language training through music for children with autism spectrum disorders: Theory and clinical application*. Jessica Kingsley Publishers.

Lim, H. A., & Draper, E. (2011). The effects of music therapy incorporated with applied behavior analysis verbal behavior approach for children with autism spectrum disorders. *Journal of Music Therapy, 48*(4), 532–550. https://doi.org/10.1093/jmt/48.4.532

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., Goode, S., Heemsbergen, J., Jordan, H., Mawhood, L., & Schopler, E. (2012). Autism Diagnostic Observation Schedule: A standardized observation of communicative and social behavior. *Journal of Autism and Developmental Disorders, 19*(2), 185–212.

Malloch, S., & Trevarthen, C. (2009). *Communicative musicality*. Oxford University Press.

Marquez-Garcia, A. V., Magnuson, J., Morris, J., Iarocci, G., Doesburg, S., & Moreno, S. (2022). Music therapy in autism spectrum disorder: A systematic review. *Review Journal of Autism and Developmental Disorders, 9*(1), 91–107. https://doi.org/10.1007/s40489-021-00246-x

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

Moreno, S., Marques, C., Santos, A., Santos, M., Castro, S. L., & Besson, M. (2009). Musical training influences linguistic abilities in 8-year-old children: More evidence for brain plasticity. *Cerebral Cortex, 19*(3), 712–723. https://doi.org/10.1093/cercor/bhn120

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

Nadig, A., & Shaw, H. (2015). Acoustic marking of prominence: How do preadolescent speakers with and without high-functioning autism mark contrast in an interactive task? *Language, Cognition and Neuroscience, 30*(1–2), 32–47. https://doi.org/10.1080/01690965.2012.753150

Neves, L., Correia, A. I., Castro, S. L., Martins, D., & Lima, C. F. (2022). Does music training enhance auditory and linguistic processing? A systematic review and meta-analysis of behavioral and brain evidence. *Neuroscience & Biobehavioral Reviews, 140,* Article 104777. https://doi.org/10.1016/j.neubiorev.2022.104777

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

Papoušek, M. (2007). Communication in early infancy: An arena of intersubjective learning. *Infant Behavior & Development, 30*(2), 258–266. https://doi.org/10.1016/j.infbeh.2007.02.003

Patel, A. D. (2011). Why would musical training benefit the neural encoding of speech? The OPERA hypothesis. *Frontiers in Psychology, 2,* Article 142. https://doi.org/10.3389/fpsyg.2011.00142

Patel, A. D. (2014). Can nonlinguistic musical training change the way the brain processes speech? The expanded OPERA hypothesis. *Hearing Research, 308,* 98–108. https://doi.org/10.1016/j.heares.2013.08.011

Paul, A., Sharda, M., Menon, S., Arora, I., Kansal, N., Arora, K., & Singh, N. C. (2015). The effect of sung speech on socio-communicative responsiveness in children with autism spectrum disorders. *Frontiers in Human Neuroscience, 9,* Article 555. https://doi.org/10.3389/fnhum.2015.00555

Paul, R., Chawarska, K., Cicchetti, D., & Volkmar, F. (2008). Language outcomes of toddlers with autism spectrum disorders: A two-year follow-up. *Autism Research, 1*(2), 97–107. https://doi.org/10.1002/aur.12

Peretz, I., Vuvan, D., Lagrois, M. É., & Armony, J. L. (2015). Neural overlap in processing music and speech. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences, 370*(1664), Article 20140090. https://doi.org/10.1098/rstb.2014.0090

Quintin, E. M. (2019). Music-evoked reward and emotion: Relative strengths and response to intervention of people with ASD. *Frontiers in Neural Circuits, 13,* Article 49. https://doi.org/10.3389/fncir.2019.00049

Raven, J., & Raven, J. (2003). Raven Progressive Matrices. In R. S. McCallum (Ed.), *Handbook of nonverbal assessment*. Springer. https://doi.org/10.1007/978-1-4615-0153-4_11

R Core Team. (2024). *R: A language and environment for statistical computing* [Computer software]. https://www.R-project.org

Reindal, L., Nærland, T., Weidle, B., Lydersen, S., Andreassen, O. A., & Sund, A. M. (2023). Structural and pragmatic language impairments in children evaluated for autism spectrum disorder (ASD). *Journal of Autism and Developmental Disorders, 53*(2), 701–719. https://doi.org/10.1007/s10803-020-04853-1

Reybrouck, M., Vuust, P., & Brattico, E. (2018). Brain connectivity networks and the aesthetic experience of music. *Brain Sciences, 8*(6), Article 107. https://doi.org/10.3390/brainsci8060107

Sahkai, H., Kalvik, M.-L., & Mihkla, M. (2013). Prosodic effects of information structure in Estonian. In E. L. Asu & P. Lippus (Eds.), *Proceedings of International Conference: Nordic Prosody XI, Tartu, 15.-17.08.2012* (pp. 323–332). Peter Lang.

Schneider, W., Eschman, A., & Zuccolotto, A. (2002). *E-Prime* (Version 2.0) [Computer software and manual].

Sharda, M., Midha, R., Malik, S., Mukerji, S., & Singh, N. C. (2015). Fronto-temporal connectivity is preserved during sung but not spoken word listening, across the autism spectrum. *Autism Research, 8*(2), 174–186. https://doi.org/10.1002/aur.1437

Sharda, M., Tuerk, C., Chowdhury, R., Jamey, K., Foster, N., Custo-Blanch, M., Tan M., Nadig A., & Hyde, K. (2018). Music improves social communication and auditory–motor connectivity in children with autism. *Translational Psychiatry, 8*(1), Article 231. https://doi.org/10.1038/s41398-018-0287-3

Simpson, K., & Keen, D. (2011). Music interventions for children with autism: Narrative review of the literature. *Journal of Autism and Developmental Disorders, 41*(11), 1507–1514. https://doi.org/10.1007/s10803-010-1172-y

Slabakova, R., Leal, T., & Domínguez, L. (2024). *Research methods in generative second language acquisition*. Routledge. https://doi.org/10.4324/9781003160762

Stein-Smith, K. (2016). The role of multilingualism in effectively addressing global issues: The sustainable development goals and beyond. *Theory and Practice in Language Studies, 6*(12), Article 2254. https://doi.org/10.17507/tpls.0612.03

Sundberg, J. (1977). Singing and timbre. *Music Room Acoustics, 17,* 57–81.

Vaiouli, P., & Andreou, G. (2018). Communication and language development of young children with autism: A review of research in music. *Communication Disorders Quarterly, 39*(2), 323–329. https://doi.org/10.1177/1525740117705117

Vaiouli, P., Grimmet, K., & Ruich, L. J. (2015). "Bill is now singing": Joint engagement and the emergence of social communication of three young children with autism. *Autism, 19*(1), 73–83. https://doi.org/10.1177/1362361313511709

Wang, B. X., Chen, S., Zhou, F., Liu, J., Xiao, C., Chan, A., & Tang, T. (2024). English prosodic focus marking by Cantonese trilingual children with and without autism spectrum disorder. *Journal of Speech, Language, and Hearing Research, 67*(3), 782–801. https://doi.org/10.1044/2023_JSLHR-23-00508

Wang, L., Beaman, C. P., Jiang, C., & Liu, F. (2022). Perception and production of statement–question intonation in autism spectrum disorder: A developmental investigation. *Journal of Autism and Developmental Disorders, 52*(8), 3456–3472. https://doi.org/10.1007/s10803-021-05220-4

Wechsler, D., & Kodama, H. (1949). *Wechsler Intelligence Scale for Children*. The Psychological Corporation.

Williams, T. I., Loucas, T., Sin, J., Jeremic, M., Meyer, S., Boseley, S., Fincham-Majumdar, S., Aslett, G., Renshaw, R., & Liu, F. (2024). Using music to assist language learning in autistic children with minimal verbal language: The MAP feasibility RCT. *Autism, 28*(10), 2515–2533. https://doi.org/10.1177/13623613241233804

Xu, Y. (1999). Effects of tone and focus on the formation and alignment of *F*0 contours. *Journal of Phonetics, 27*(1), 55–105. https://doi.org/10.1006/jpho.1999.0086

Xu, Y. (2013). *ProsodyPro—A tool for large-scale systematic prosody analysis*. https://www.semanticscholar.org/paper/Prosody-Pro-—-A-Tool-for-Large-scale-Systematic-Xu/453480b67a15cac4ec47b52c3a2eea241c18db64

Xu, Y., Chen, S. W., & Wang, B. (2012). Prosodic focus with and without post-focus compression: A typological divide within the same language family? *The Linguistic Review, 29*(1), 131–147. https://doi.org/10.1515/tlr-2012-0006

Zendel, B. R., West, G. L., Belleville, S., & Peretz, I. (2019). Musical training improves the ability to understand speech-in-noise in older adults. *Neurobiology of Aging, 81,* 102–115. https://doi.org/10.1016/j.neurobiolaging.2019.05.015

Appendix (p. 1 of 5)

Post Hoc Comparisons

Table A1. Post hoc comparisons of *F*0 range and mean *F*0 between different focus conditions in the same group.

		*F*0 range									
		CantoneseASD_Pre		CantoneseASD_Post		CantoneseTD_Test1		CantoneseTD_Test2		Mandarin TD	
		t ratio	*p* value	*t* ratio	*p* value	*t* ratio	*p* value	*t* ratio	*p* value	*t* ratio	*p* value
T1	Broad vs. Non-contrastive_Narrow	−1.1	0.999996	−1.07	0.999997	−1.07	0.999997	0.11	1	−2.49	0.7016
	Broad vs. Contrastive_Narrow	−1.19	0.999982	−1.28	0.999927	−1.28	0.999927	1.69	0.9943	−1.05	0.999998
	Broad vs. Pre-Focus	0.55	1	−0.05	1	−0.05	1	1.22	0.99997	0.31	1
	Broad vs. Post-Focus	0.2	1	2.88	0.3964	2.88	0.3964	0.8	0.99999999	−2.54	0.6672
	Non-contrastive vs. Contrastive	−0.09	1	−0.22	1	−0.22	1	1.58	0.9978	1.44	0.9995
	Non-contrastive vs. Pre-Focus	1.8	0.9864	1.18	0.99998	1.18	0.99998	1.09	0.99999	3.15	0.218
	Non-contrastive vs. Post-Focus	1.45	0.9994	4.13	0.0086	4.13	0.0086	0.67	1	0.29	1
	Contrastive vs. Pre-Focus	1.9	0.9742	1.41	0.9996	1.41	0.9996	−0.7	1	1.51	0.9989
	Contrastive vs. Post-Focus	1.55	0.9984	4.34	0.0036	4.34	0.0036	−1.13	0.99999	−1.35	0.9998
	Pre vs. Post-Focus	−0.4	1	3.39	0.1163	3.39	0.1163			−3.27	0.1629
T2	Broad vs. Non-contrastive_Narrow	−0.2884	1	−1.0674	0.999998	0.3884	1	−0.8649	1	−2.8054	0.4519
	Broad vs. Contrastive_Narrow	−2.0604	0.938	−0.9211	1	−0.5168	1	−0.7497	1	−2.7087	0.5295
	Broad vs. Pre-Focus	1.7985	0.9869	3.0579	0.2722	2.2044	0.8823	−2.2231	0.8732	−1.4125	0.99962
	Broad vs. Post-Focus	−1.3094	0.9999	0.6272	1	1.1736	0.99999	1.5988	0.9974	0.7289	1
	Non-contrastive vs. Contrastive	−1.7713	0.9893	0.1493	1	−0.9052	1	0.1168	1	0.0949	1
	Non-contrastive vs. Pre-Focus	2.1273	0.9151	4.3013	0.0043	1.7602	0.9901	−1.2339	0.99996	1.7914	0.9876
	Non-contrastive vs. Post-Focus	−0.9792	1	1.8556	0.9807	0.7296	1	2.5866	0.6285	3.9222	0.01959
	Contrastive vs. Pre-Focus	4.1535	0.0079	4.1418	0.0083	2.7954	0.4598	−1.3705	0.99977	1.6818	0.9946
	Contrastive vs. Post-Focus	1.0459	1	1.6914	0.9942	1.7643	0.9898	2.4599	0.7261	3.8104	0.02949
	Pre vs. Post-Focus	−3.6328	0.0543	−2.8611	0.4087	−1.205	0.99998	4.4608	0.00214	2.4845	0.7079
T4	Broad vs. Non-contrastive_Narrow	−1.776	0.9889	−2.0418	0.9435	0.5894	1	−1.7014	0.9937	−3.1114	0.2405
	Broad vs. Contrastive_Narrow	−4.498	0.0018	−2.451	0.7326	0.26	1	−0.5009	1	−3.7376	0.0381
	Broad vs. Pre-Focus	1.785	0.9881	3.6851	0.0456	3.1518	0.2181	0.5721	1	0.5674	1
	Broad vs. Post-Focus	−1.8078	0.9861	−0.0409	1	4.055	0.0118	1.4538	0.9994	1.9239	0.9705
	Non-contrastive vs. Contrastive	−2.751	0.4952	−0.4092	1	−0.3297	1	1.2008	0.99998	−0.626	1
	Non-contrastive vs. Pre-Focus	3.8325	0.0272	6.0196	5.20E−07	2.4757	0.7145	2.5139	0.6856	4.1154	0.0093
	Non-contrastive vs. Post-Focus	0.2133	1	2.2997	0.8321	3.3775	0.1196	3.3913	0.1149	5.4604	1.39E−05
	Contrastive vs. Pre-Focus	6.9171	1.38E−09	6.4878	2.61E−08	2.8562	0.4125	1.144	0.99999	4.8296	0.0004
	Contrastive vs. Post-Focus	3.3383	0.1337	2.7686	0.4811	3.7594	0.0353	2.0247	0.9482	6.1724	2.01E−07
	Pre vs. Post-Focus	−4.1928	0.0068	−4.3658	0.0033	1.0525	0.9999982	1.0267	0.9999989	1.5728	0.99796

(table continues)

Appendix (p. 2 of 5)

Post Hoc Comparisons

Table A1. *(Continued).*

	mean *F*0									
	CantoneseASD_Pre		CantoneseASD_Post		CantoneseTD_Test1		CantoneseTD_Test2		Mandarin TD	
	t ratio	*p* value	*t* ratio	*p* value	*t* ratio	*p* value	*t* ratio	*p* value	*t* ratio	*p* value
T1	2.77	0.4835	–1.38	0.9998	–1.14	0.9999	0.78	1	–3.58	0.0641
	0.62	1	–1.73	0.9918	–1.61	0.997	0.81	1	–5.53	9.24E–06
	–0.7	1	–1.18	0.9999	–1.09	0.9999	0.67	1	–0.23	1
	4.06	0.0116	–0.5	1	3.7	0.0435	2.75	0.4939	18.35	0
	–2.14	0.9092	–0.37	1	–0.48	1	0.02	1	–1.95	0.9656
	–3.85	0.0256	–1.53	0.9987	0.2	1	–0.22	1	3.86	0.0251
	0.91	0.9999	0.38	1	4.99	0.0002	1.86	0.9801	22.43	0
	–1.41	0.9997	–1.1	0.9999	0.74	1	–0.25	1	6.08	3.56E–07
	3.35	0.1299	0.79	0.9999	5.54	9.02E–06	1.84	0.9829	24.66	0
	5.47	1.33E–05	2.18	0.8934	5.51	1.05E–05	2.39	0.776	21.36	0
T2	0.27	1	–1.25	0.99995	–0.17	1	0.25	1	–0.56	1
	–1.92	0.9717	–1.24	0.99996	–1.26	0.99994	0.83	1	–1.68	0.9948
	–2.19	0.8905	–2.38	0.781	–6.16	2.20E–07	–2.51	0.6859	–3.24	0.1767
	1.37	0.9998	1.72	0.9925	2.3	0.8302	3.91	0.0205	8.03	6.05E–13
	–2.19	0.8893	0.02	1	–1.09	0.999996	0.59	1	–1.12	0.99999
	–2.5	0.698	–0.96	0.9999997	–5.96	7.55E–07	–2.79	0.4607	–2.6	0.6171
	1.06	0.999998	3.18	0.2055	2.5	0.6947	3.63	0.0555	8.67	2.12E–13
	0	1	–0.97	0.9999996	–4.71	0.00067	–3.47	0.0905	–1.32	0.99989
	3.56	0.0687	3.16	0.2164	3.75	0.0368	2.97	0.333	9.94	2.18E–13
	4.15	0.008	4.83	0.00039	9.88	2.31E–13	7.49	2.15E–11	13.11	0
T4	–1.52	0.9987	–1.76	0.9902	0.21	1	–1.01	0.9999	–5.86	1.39E–06
	–4.08	0.0105	–2.12	0.9161	–0.54	1	–0.27	1	–6.74	4.73E–09
	–4.03	0.013	–5.33	2.82E–05	–0.98	0.9999	–4.82	0.0004	–1.58	0.9978
	3.03	0.2873	1.31	0.9999	5.08	0.0001	3.65	0.0519	17.92	0
	–2.58	0.6347	–0.37	1	–0.75	1	0.75	1	–0.88	0.9999
	–2.31	0.8247	–3.34	0.1348	–1.21	0.99997	–3.65	0.0512	5.11	9.38E–05
	4.79	0.0005	3.34	0.1346	4.84	0.0004	4.8	0.0005	24.58	0
	0.64	1	–1.54	0.9986	–0.36	1	–4.51	0.0017	6.11	2.91E–07
	7.69	4.53E–12	3.75	0.036	5.7	3.64E–06	3.95	0.0175	25.58	0
	8.23	3.41E–13	7.77	2.47E–12	7.05	5.45E–10	9.82	2.51E–13	22.63	0

Post Hoc Comparisons

Table A2. Post hoc comparisons of *F*0 range and mean *F*0 between different groups under the same condition.

		*F*0 range									
		Broad_Focus		Non-contrastive_Narrow_Focus		Contrastive_Narrow_Focus		Pre-Focus		Post-Focus	
		t ratio	*p* value	*t* ratio	*p* value	*t* ratio	*p* value	*t* ratio	*p* value	*t* ratio	*p* value
T1	CantoneseASD_Pre vs. CantoneseASD_Post	−0.3478	1	−0.3262	1	−0.4556	1	−1.2366	0.99996	2.8472	0.41938
	CantoneseASD_Pre vs. CantoneseTD_Test1	1.2284	0.99997	1.7622	0.98996	1.7575	0.9903	1.8453	0.982	0.7709	1
	CantoneseASD_Pre vs. CantoneseTD_Test2	1.3359	0.99985	1.8088	0.98596	2.449	0.73412	1.6188	0.99688	1.5966	0.99745
	CantoneseASD_Pre vs. MandarinTD_Control	6.0345	0.000000475	5.5264	0.00000959	6.1071	0.000000302	6.1846	0.000000186	5.3192	0.0000303
	CantoneseASD_Post vs. CantoneseTD_Test1	1.3626	0.99979	1.8891	0.97614	1.9337	0.96871	2.1978	0.88536	−0.0409	1
	CantoneseASD_Post vs. CantoneseTD_Test2	1.4718	0.99927	1.9373	0.96804	2.627	0.59591	1.976	0.96016	0.774	1
	CantoneseASD_Post vs. MandarinTD_Control	6.1682	0.000000206	5.6581	0.00000451	6.2849	0.000000098	6.5447	1.78E−08	4.4872	0.0019
	CantoneseTD_Test1 vs. CantoneseTD_Test2	0.0911	1	0.0231	1	0.6678	1	−0.251	1	0.8154	1
	CantoneseTD_Test1 vs. MandarinTD_Control	4.7732	0.0005	3.7181	0.04076	4.304	0.00425	4.2906	0.00449	4.5286	0.00158
	CantoneseTD_Test2 vs. MandarinTD_Control	4.7467	0.00057	3.7464		3.679		4.6083		3.7556	
T2	CantoneseASD_Pre vs. CantoneseASD_Post	−0.9196	0.999999878	−1.7076	0.993400523	0.2109	1	0.2762	1	1.092	0.999996252
	CantoneseASD_Pre vs. CantoneseTD_Test1	1.0677	0.999997575	1.2709	0.999937857	1.5313	0.99863708	1.2014	0.999977463	1.7602	0.990110899
	CantoneseASD_Pre vs. CantoneseTD_Test2	3.9677	0.01649626	3.799	0.030722999	4.3738	0.003141615	2.9802	0.322710612	4.8412	0.00035985
	CantoneseASD_Pre vs. MandarinTD_Control	6.63	1.01E−08	5.8975	1.10E−06	6.4712	2.91E−08	5.9179	9.68E−07	7.3359	6.62E−11
	CantoneseASD_Post vs. CantoneseTD_Test1	1.3447	0.999834165	1.7846	0.988176745	1.4682	0.99929382	1.1413	0.999991311	1.5223	0.998754612
	CantoneseASD_Post vs. CantoneseTD_Test2	4.247	0.005396613	4.3192	0.003975277	4.31	0.004134687	2.9193	0.365490277	4.6001	0.001136947
	CantoneseASD_Post vs. MandarinTD_Control	6.9114	1.44E−09	6.424	3.97E−08	6.4069	4.44E−08	5.8564	1.40E−06	7.0918	3.97E−10
	CantoneseTD_Test1 vs. CantoneseTD_Test2	2.8857	0.39017581	2.5111	0.687768494	2.8217	0.439121383	1.7627	0.989920161	3.0573	0.272592708
	CantoneseTD_Test1 vs. MandarinTD_Control	5.5347	9.16E−06	4.5938	0.001170453	4.9003	0.000268681	4.6856	0.000762019	5.5304	9.38E−06
	CantoneseTD_Test2 vs. MandarinTD_Control	2.6503	0.577033701	2.0806	0.93159018	2.0736	0.933872153	2.942	0.349221478	2.4706	0.718259881
T4	CantoneseASD_Pre vs. CantoneseASD_Post	−0.9769	0.999999582	−1.264	0.999943618	1.0687	0.999997531	0.9682	0.999999651	0.8054	0.999999992
	CantoneseASD_Pre vs. CantoneseTD_Test1	−0.3145	1	0.449	1	1.2296	0.999965594	0.0728	1	1.3826	0.999735066
	CantoneseASD_Pre vs. CantoneseTD_Test2	2.2924	0.836319253	2.3308	0.813549142	3.6121	0.058160058	1.9907	0.956794602	3.3053	0.146438008
	CantoneseASD_Pre vs. MandarinTD_Control	6.4791	2.76E−08	6.0901	3.36E−07	6.7987	3.16E−09	6.2814	1.00E−07	7.7416	3.09E−12
	CantoneseASD_Post vs. CantoneseTD_Test1	0.0025	1	0.8563	0.999999972	0.8826	0.999999948	−0.1556	1	1.1939	0.999979909
	CantoneseASD_Post vs. CantoneseTD_Test2	2.6135	0.606874246	2.7422	0.50236512	3.2605	0.165268868	1.7593	0.990170427	3.1144	0.238790363
	CantoneseASD_Post vs. MandarinTD_Control	6.8038	3.05E−09	6.5038	2.34E−08	6.4429	3.51E−08	6.0477	4.37E−07	7.549	1.33E−11
	CantoneseTD_Test1 vs. CantoneseTD_Test2	2.6114	0.608520262	1.8746	0.978229208	2.3666	0.790892997	1.9172	0.971642102	1.9042	0.973793209
	CantoneseTD_Test1 vs. MandarinTD_Control	6.8026	3.08E−09	5.6259	5.44E−06	5.5387	8.95E−06	6.2078	1.60E−07	6.3237	7.63E−08
	CantoneseTD_Test2 vs. MandarinTD_Control	4.2183	0.006083123	3.7808	0.032771813	3.1876	0.199556111	4.3279	0.003830515	4.4585	0.002162802

(table continues)

Appendix (p. 4 of 5)

Post Hoc Comparisons

Table A2. *(Continued).*

	mean *F0*									
	Broad_Focus		Non-contrastive_Narrow_Focus		Contrastive_Narrow_Focus		Pre-Focus		Post-Focus	
	t ratio	*p* value	*t* ratio	*p* value	*t* ratio	*p* value	*t* ratio	*p* value	*t* ratio	*p* value
T1	−4.14	0.0084	−4.14	0.0084	−2.35	0.8019	−2.96	0.3337	−6.48	2.72E−08
	−1.15	0.9999	−1.15	0.9999	−0.96	1	−0.75	1	−0.74	1
	−0.93	1	−0.93	1	−0.68	1	−0.56	1	−0.84	1
	0.29	1	0.29	1	0.31	1	1.08	0.9999	2.5	0.694
	−0.68	1	−0.68	1	−0.69	1	−0.51	1	−0.22	1
	−0.45	1	−0.45	1	−0.41	1	−0.32	1	−0.31	1
	0.77	1	0.77	1	0.59	1	1.33	0.9999	3.04	0.2829
	0.24	1	0.24	1	0.29	1	0.19	1	−0.09	1
	1.47	0.9993	1.47	0.9993	1.3	0.9999	1.85	0.9819	3.26	0.1638
	1.25	0.99996	1.25	0.99996	1.01	0.9999	1.67	0.995	3.4	0.1128
T2	−2.45	0.735	−4	0.0147	−1.79	0.9873	−3.75	0.0363	−3.01	0.3055
	0.21	1	0.14	1	0.31	1	−0.31	1	0.33	1
	−0.99	0.9999994	−1	0.9999993	−0.58	1	−1.03	0.9999987	−0.68	1
	0	1	−0.13	1	0.04	1	−0.14	1	0.9	0.9999999
	0.58	1	0.74	1	0.58	1	0.09	1	0.65	1
	−0.62	1	−0.39	1	−0.31	1	−0.63	1	−0.35	1
	0.38	1	0.49	1	0.31	1	0.27	1	1.22	0.9999681
	−1.21	0.9999755	−1.14	0.9999909	−0.89	0.9999999	−0.72	1	−1.01	0.9999991
	−0.22	1	−0.27	1	−0.28	1	0.18	1	0.55	1
	1.02	0.999999	0.9	0.9999999	0.64	1	0.92	0.9999999	1.6	0.9973
T4	−5.86	1.39E−06	−0.669	1	1.535	0.998589	−2.224	0.8726	−2.71	0.5283
	−6.74	4.73E−09	−0.803	0.99999999	−0.562	1	−0.68	1	−0.799	0.99999999
	−1.58	0.9978	−1.102	0.99999554	−0.662	1	−1.253	0.9999519	−1.118	0.9999941
	17.92	0	−0.21	1	0.02	1	0.664	1	2.131	0.9135
	−0.88	0.9999	−0.715	0.99999999	−0.765	0.99999999	−0.472	1	−0.546	1
	5.11	9.38E−05	−1.013	0.99999914	−0.867	0.99999996	−1.041	0.9999985	−0.862	0.99999997
	24.58	0	−0.12	1	−0.188	1	0.878	0.99999995	2.391	0.7748
	6.11	2.91E−07	−0.288	1	−0.092	1	−0.563	1	−0.309	1
	25.58	0	0.614	1	0.596	1	1.362	0.9997942	2.951	0.343
	22.63	0	0.918	0.99999988	0.699	1	1.959	0.9637	3.309	0.1449

Post Hoc Comparisons

Table A3. Post hoc comparisons of intensity.

	CantoneseASD_Pre		CantoneseASD_Post		CantoneseTD_Test1		CantoneseTD_Test2		Mandarin TD	
	t ratio	*p* value	*t* ratio	*p* value	*t* ratio	*p* value	*t* ratio	*p* value	*t* ratio	*p* value
Broad_Focus vs. Non-contrastive_Narrow_Focus	0.8	0.99999999	0.62	1	−0.6	1	0.17	1	−0.86	0.9999
Broad_FocusContrastive_Narrow_Focus	−1.14	0.99999162	0.05	1	−0.3	1	−0.08	1	−1.97	0.9621
Broad_FocusPre-Focus	0.98	0.99999952	3.38	0.12	2.18	0.8915	−0.24	1	−3.16	0.2129
Broad_FocusPost-Focus	2.43	0.7474	3.31	0.1455	1.61	0.997	3.38	0.1186	20.78	0
Non-contrastive_Narrow_FocusContrastive_Narrow_Focus	−1.94	0.9674	−0.57	1	0.3	1	−0.25	1	−1.11	0.999995
Non-contrastive_Narrow_FocusPre-Focus	0.07	1	2.68	0.5542	2.87	0.3987	−0.44	1	−2.18	0.8934
Non-contrastive_Narrow_FocusPost-Focus	1.52	0.9988	2.61	0.6116	2.3	0.8301	3.18	0.2045	21.76	0
Contrastive_Narrow_FocusPre-Focus	2.28	0.8417	3.32	0.1391	2.53	0.6722	−0.15	1	−0.92	0.999999
Contrastive_Narrow_FocusPost-Focus	3.73	0.0394	3.26	0.1672	1.96	0.9636	3.47	0.0898	23.01	0
Pre-FocusPost-Focus	1.68	0.9947	−0.09	1	−0.66	1	4.2	0.0065	27.66	0
	Broad_Focus		**Non-contrastive_Narrow_Focus**		**Contrastive_Narrow_Focus**		**Pre-Focus**		**Post-Focus**	
	t ratio	*p* value	*t* ratio	*p* value	*t* ratio	*p* value	*t* ratio	*p* value	*t* ratio	*p* value
CantoneseASD_Pre vs. CantoneseASD_Post	−1.67	0.9952	−1.85	0.9814	−0.48	1	0.62	1	−1.26	0.9999
CantoneseASD_Pre vs. CantoneseTD_Test1	2.23	0.8712	2.11	0.9223	2.3	0.8329	2.32	0.8195	2.17	0.8977
CantoneseASD_Pre vs. CantoneseTD_Test2	0.63	1	0.58	1	0.73	1	0.54	1	0.7	1
CantoneseASD_Pre vs. MandarinTD_Control	−0.24	1	−0.39	1	−0.31	1	−0.55	1	1.09	0.999996
CantoneseASD_Post vs. CantoneseTD_Test1	2.37	0.7889	2.26	0.8517	2.34	0.8083	2.28	0.8413	2.25	0.8614
CantoneseASD_Post vs. CantoneseTD_Test2	0.78	1	0.74	1	0.77	1	0.5	1	0.78	1
CantoneseASD_Post vs. MandarinTD_Control	−0.1	1	−0.22	1	−0.27	1	−0.59	1	1.17	0.999987
CantoneseTD_Test1 vs. CantoneseTD_Test2	−1.62	0.9967	−1.56	0.9983	−1.6	0.9973	−1.81	0.9858	−1.5	0.999
CantoneseTD_Test1 vs. MandarinTD_Control	−2.53	0.6738	−2.55	0.6593	−2.67	0.5643	−2.93	0.3565	−1.14	0.999992
CantoneseTD_Test2 vs. MandarinTD_Control	−0.9	0.9999999	−0.99	0.9999995	−1.06	0.999998	−1.11	0.9999944	0.39	1