Project Title	**:**	Developing a Social Robot for Cantonese and Mandarin Speech Prosody Training in Children With Autism Spectrum Disorder
Grantee	**:**	The Hong Kong Polytechnic University
Principal Investigator	**:**	CHEN Si The Hong Kong Polytechnic University
Co-Investigator(s)	**:**	CHAN Wing-shan The Hong Kong Polytechnic University
		LI Bin City University of Hong Kong
		TANG Po-yi The University of Hong Kong
		CHEN Zhuo-ming First Affiliated Hospital of Jinan University
		CHEUNG Chung-wai The Hong Kong Polytechnic University
		WEN Chun-yi The Hong Kong Polytechnic University
		LV Shuang Renmin University of China
		LIU Yan The Hong Kong Polytechnic University

Final Report

by

Principal Investigator

Report

(a) Title: Developing a Social Robot for Cantonese and Mandarin Speech Prosody Training in Children with Autism Spectrum Disorder

(b) Abstract *(write a brief description of the report that summarises the objective(s) and significance of the project, its methodology, findings, conclusions and recommendations in a single paragraph and in no more than 250 words for the report written in English or 160 characters for that in Chinese)*

Abnormal speech prosody has been widely reported in individuals with autism. Many studies on children and adults with autism spectrum disorder speaking a non-tonal language showed deficits in using prosodic cues to mark focus. However, focus marking by autistic children speaking a tonal language is rarely examined. Cantonese children may face additional difficulties because tonal languages require them to use prosodic cues to achieve multiple functions simultaneously such as lexical contrasting and focus marking. Also, the acquisition of speech prosody in a non-native language is rarely examined. This study bridges this research gap by acoustically evaluating the use of Cantonese and Mandarin speech prosody to mark information structure by Cantonese-speaking children with and without autism spectrum disorder. We designed speech production tasks to elicit natural broad and narrow focus production among these children in sentences with different tonal combinations. Acoustic correlates of prosodic focus marking like f0, duration and intensity of each syllable were analyzed to examine the effect of participant group, focus condition and lexical tones. Our results showed differences in focus marking patterns between Cantonese-speaking children with and without autism spectrum disorder in Cantonese and Mandarin. In addition, we have provided sung speech training and developed a social robot that can implement the training in Cantonese and Mandarin. Sung speech training has significantly improved the use of speech prosody in Cantonese and Mandarin.

(c) Keyword(s) (provide a list of up to seven alphabetised words or short phrases that *are central and specific to the project)*

autism spectrum disorder, focus marking, human-robot interaction, speech prosody

(d) Introduction

Autism spectrum disorder (ASD) is a neurodevelopmental disorder marked by persistent deficiencies in social communication and interaction as well as limited and repetitive behavior, interests and activities (APA, 2013). Language deficits, compounded by challenges in social interactions, may remain a persistent and life-long challenge for many autistic individuals, and hence are regarded as important targets of early intervention for children with ASD.

Peculiar tones of voice and disturbances of prosody have been identified as the earliest characteristics of ASD. Children with ASD tend to show atypical patterns of speech prosody. The research on prosody production among individuals with ASD is important because speech prosody is a key component in communication. It is also reported that prosodic impairments and social communication are strongly correlated (Paul et al., 2005) and impairments in speech prosody can negatively affect friends making and job seeking (Eigsti et al., 2012). However, the existing research on prosody production in ASD, has been focusing on speakers of non-tonal languages, leaving the interaction between lexical tones and intonation in tone languages under-investigated (for a review see (Fusaroli et al., 2016). Tonal languages may offer a more challenging situation for individuals with ASD in using discourse functions such as focus marking because the acoustic cues such as fundamental frequency (f0) are used to achieve both lexical contrasts and focus marking. Acquisition of non-native speech prosody by children with ASD is also less examined. In addition, as an emerging intervention for language development of the children with special needs, the acoustic similarities between songs and speech have not been largely considered in previous interventions, and the potential benefits that song-based interventions can bring on certain aspects of speech, e.g., speech prosody have yet to be examined. The present study aims to fill in this research gap by analyzing the acoustic features of focus-marking by Cantonese-speaking children with ASD in comparison with their typically developing (TD) peers. The results may improve our understanding of prosodic production deficits in the population with ASD and may have clinical implications. We also aim to test the effect of sung-speech training on the improvement of speech prosody in both Cantonese and Mandarin and create a robot-assisted program based on the findings.

(e) Review of literature of the project

Speech prosody is the vocal modulation accompanying speech, which comprises variations in f0, duration, intensity and voice quality and serves a wide range of communication functions, such as signaling information structure and expressing the speakers' emotions and attitudes (Cutler & Pearson, 2018). A typical example of information structure categories is focus, which marks new information to the receiver(s) in a sentence, (Lambrecht, 1996; Gundel, 1999). There are two main focus types: broad focus (i.e., focus falling on the entire utterance) and narrow focus (i.e., focus falling on a selective part of an utterance). Narrow focus can be further categorized into non-contrastive and contrastive narrow focus, with the latter providing an explicit contrast to alternatives (Gundel, 1999). Focus can be marked by morpho-syntactic and prosodic means. Acoustic correlates of focus on and beyond the components on focus have been reported. Despite language-specific differences, components on focus are often realized with longer duration, higher f0 values or larger f0 range, and/or increased intensity than the components carrying no focus (for English see (Eady & Cooper, 1986; Xu & Xu, 2005), for German see (Féry & Kügler, 2008), for Mandarin see (Xu, 1999), for Japanese see (Ishihara, 2011), and components following on-focus syllables are also realized with reduced f0 range and intensity (i.e., post-focus compression, PFC) in languages like English, Greek, Dutch, Korean, and Mandarin (for review, see (Xu et al., 2012)).

Children with ASD tend to show delayed, deviant development and deficits in speech prosody. Meta-analyses of acoustic studies on prosodic features of vocal productions suggest that speech prosody of the autistic population is characterized by significantly higher mean f0, larger f0 range, longer voice duration and greater f0 variability (Fusaroli et al., 2016; Asghari et al., 2021). There is a paucity in research focusing on the production of prosodic prominence by autistic children.

In terms of prosodic focus marking, Diehl and Paul (Diehl & Paul, 2009; Diehl & Paul, 2011) also found that the differences between syllables carrying or not carrying focus in the autistic speech were less prominent than those in the TD speech. It is worth mentioning that in Diehl and Paul's studies, children with ASD tended to over-lengthen the syllables carrying no focus, unlike those in Paul et al.'s study, who did not lengthen the stressed syllables enough. The differences may arise from the different tasks and stimuli used in these two studies. Paul et al. elicited speech via imitation using the Tennessee Test of Rhythm and Intonation Patterns (T-TRIP, [32]) which involved 25 pre-recorded nonsense syllable /ma/ varying in rhythm and intonation. Diehl and Paul, however, used Profiling Elements of Prosodic Systems (PEPS-C), which assesses children's abilities to discriminate and articulate the prosodic forms in four areas of communication where prosody plays a critical role, namely, interaction, affect, boundary and focus (Peppé & McCann, 2003). Studies using PEPS-C have generally reported a significantly worse performance of the autistic children than their TD peers in both perceptual and production tasks (Diehl & Paul, 2011； DePape et al., 2012).

Meanwhile, there are also studies reporting comparable performance between the autistic and TD children. For instance, Nadig & Shaw (Nadig & Shaw, 2012) acoustically analyzed on- and post-focus syllables produced by English-speaking children with and without ASD and found that both groups produced significantly longer and louder on-focus syllables than post-focus ones, but neither of them used mean f0 in focus marking. The existing research has reported complex results in the use of f0 in focus marking by the autistic children. DePape et al. (Peppé et al., 2006) found that it were the autistic children with moderate rather than high language skills that used f0 range to mark information structure, although children with moderate skills did not necessarily master the correct usage of f0 range, and their performance may be influenced by the intervention they previously received.

From the studies reviewed so far, it seems that the use of f0 cues by autistic children in focus marking, in particular, seems to be more problematic. This makes prosodic focus marking in tone-language speaking children with ASD an interesting topic, as they do not only need to make the components on focus acoustically more prominent but also to keep the shape of lexical tones so as to convey the core meanings of words, which remains to be explored.

Cantonese is a typical tone language that uses f0 to contrast meanings of words. There are six full tones (i.e. carried by open syllables) and three checked tones (i.e. carried by syllables ending with /p/, /t/ or /k/) in Cantonese. An example of all full tones on the [fu] syllables is given as follows: [fu] with Tone 1 (55/53) ‘to call’; Tone 2 (25) ‘bitter; Tone 3 (33) ‘rich’; Tone 4 (21) ‘to hold’; Tone 5 (23) ‘woman’; and Tone 6 (22) ‘rotten’ (the numbers in bracket are Chao Tone Numeral, which marks the lowest pitch point with 1 and the highest with 5) (Chen et al., 2019).

As mentioned earlier, prosodic marking of focus is usually manifested in acoustic cues such as f0, intensity and duration (Xu & Xu, 2005). In addition to the adjustment of acoustic cues of on-focus words (e.g. higher f0 values, larger f0 range, longer duration and larger intensity), post-focus compression (i.e. reduced f0 range and intensity of words after the on-focus words (Xu, 2011), has also been found in many languages. However, the acoustic correlates of focus marking in Cantonese remain controversial. Some studies report on-focus f0 expansion and post-focus f0 compression in Cantonese (Gu,2007; Man, 2002) but others suggest that prosodic prominence in Cantonese is primarily signalled by on-focus lengthening (Fung & Mok, 2018; Mok et al., 2014). For instance, Mann (Man, 2002) examined the f0 changes of Cantonese monosyllabic words in broad and narrow focus conditions and found an expansion of f0 range for narrow focus, and yet the expansion may be affected by tone-focus interaction. However, using six sentences with the same tones on each syllable (from all Tone 1, all Tone 2 up to all Tone 6), Wu and Xu (Wu & Xu, 2010) found an increment of f0 excursion size in the dynamic tones but no increment in the static tones, and they reported no post-focus compression for Cantonese. In a more recent study, Fung and Mok (Fung & Mok, 2018) found no significant on-focus f0 changes, arguing that corrective focus in Cantonese is marked solely by durational expansion.

By contrast, prosodic-marking in Mandarin features both OFE and PFC, with f0 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 f0 ranges of contour tones. Mandarin also marks focus with OFE in duration and intensity as well as PFC in f0 range, f0 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 f0 changes.

The similarities between speech and music prosody has long been recognized in the existing literature, as pitch contour and rhythmic grouping are critical dimensions to both of them. In music, pitch and rhythmic relations define musical tunes while in speech, pitch and rhythm are important source of lexical, pragmatic and paralinguistic information (Thompson et al., 2004).

Despite of its great potential, the effect of song-based training or therapies to improve speech prosody remains understudied. 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 non-native speech.

(f) Theoretical and/or conceptual framework of the project

Theoretical frameworks have emerged to account for the relationship between musical training and speech processing. One such framework is the OPERA hypothesis, proposed by Patel (2011, 2012 , 2014), which posits that musical training can improve the neural representation of speech under five specific conditions. These conditions include: Overlap – the brain network responsible for processing the target music signals and speech cues should overlap; Precision – music requires higher precision in processing the target signals, which has been expanded to encompass higher demands in sensory or cognitive process (Patel, 2014); Emotion, Repetition and Attention – musical training activities should evoke strong positive emotions, involve frequent repetition, and require attention allocation, all of which are typical in musical training. These conditions explain why musical training has the potential to induce adaptive changes in auditory processing circuits, enhancing their precision beyond what is typically needed for speech processing (Patel, 2012). Supporting this hypothesis, Tierney & Kraus (2013) found a correlation between synchronization with a beat and auditory brainstem response, suggesting shared perception of timing details.

Building upon the OPERA hypothesis, they further proposed that musicians may have advantages in phonological skills due to the entrainment practice involved in musical training (Tierney & Kraus, 2014).

For the acquisition of non-native speech prosody, we used two models Cumulative-enhancement model (CEM, Flynn et al., 2004) and Scalpel model (SM, Slabakova, 2017). They 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.

(g) Methodology

To examine the acquisition of Cantonese speech prosody, native Cantonese-speaking children with ASD and Cantonese TD children participated in the experiment. All of the ASD participants in the experiment were formally diagnosed with ASD by professionals in established institutions based on ADOS-2 and other assessments. No participants were diagnosed of or suspected to have any other disorders. No TD participants had any speech or language disorders or suspected to have any disorders. Participants were invited to the speech laboratory at the Hong Kong Polytechnic University accompanied by parents. All child participants and parents were well-informed and agreed to participate in the experiment. Written consent was obtained from parents of child participants and verbal consent was obtained from child participants. The parents 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 child participants, and they also filled in questionnaires on the demographic and clinic conditions (if applied) of the children. All protocols were carried out in accordance with relevant guidelines and regulations. All participants were compensated for participating in the experiment.

ASD and TD participants with and without ASD were matched in age, gender, linguistic background and musical training background. All participants spoke Cantonese as their first and dominant language at home and school.

All participants were formally tested using the verbal language tests (expressive naming and narration) in Hong Kong Cantonese Oral Language Assessment Scale (HKCOLAS) (T'sou, 2006) and the non-verbal analytical intelligence with the Raven's Progressive Matrices (IQ) (Raven, 1989). The standard scores and age equivalent were obtained. HKCOLAS is a standardized speech and language assessment tool for Cantonese-speaking children. Two subtests (Narrative Test and Expressive Nominal Vocabulary Test) from HKCOLAS were used to assess the participants’ language ability in the current study. Raven’s Progressive Matrices test is a non-verbal intelligence test to assess abstract reasoning. There are sixty multiple choice questions on pattern matching. All questions were grouped into five sets, and within each set the questions were presented in an order where the difficulty of each set increased.

In total, 15 target sentences were used as stimuli in the experiment. Each sentence contains five monosyllabic words. They all depict an action and have a subject, a verb and an object. The prosodic complexity of stimuli is controlled by using two types of sentences: sentences with all words bearing the same tone (one from the six tones: Tone 1, Tone 2, Tone 3, Tone 4, Tone 5 and Tone 6), and sentences with a mixture of tones in which subjects carried one tone while the verbs and objects carried a different tone.

Fifteen corresponding pictures depicting the content of the target sentences were used to elicit natural answers from participants. Target sentences were grouped into five blocks and each block contains three target sentences. All the stimuli were presented randomly to each participant and the order of blocks was also randomized. For each sentence, a series of questions were designed to elicit the desired types of focus (i.e. broad, narrow and contrastive focus) in initial (subject), middle (verb), or final (object) positions.

The experimental session was made up of five blocks and each block contained 42 randomized trials [3 out of 15 target sentences * (1 broad focus + 1 non-contrastive narrow focus * 3 positions + 1 contrastive narrow focus * 3 positions) * 2 repetitions]. In total, 210 target sentences (42 trials * 5 blocks) were collected for each participant. The experiment was programmed in E-prime 2.0 [46].

Experiments were conducted in a sound-proof booth at the speech lab of the Hong Kong Polytechnic University. Audio Technica ATone 2035 condenser microphone and Steinberg UR22mkII USB Audio Interface were used to record participants' speech production with the sample rate of 44100 Hz in Audacity (Schneider, 2002).

Every block consisted of a practice session and a test session. During the practice session, the participants were instructed to familiarize themselves with the pictures of people and animals performing different actions so that they could consistently label people, animals, and the actions depicted in order to successfully play the game. Then they repeated each sentence recorded by a native Cantonese-speaking female speech therapist. The practice helped to reduce production errors in the later experiment. We reduced the memory load by using three stimulus sentences in each block so that children were able to remember 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.

During the experimental session, we followed the design of the game "under the shape" (Schneider, 2002). In each trial, the 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 (Fig 1).

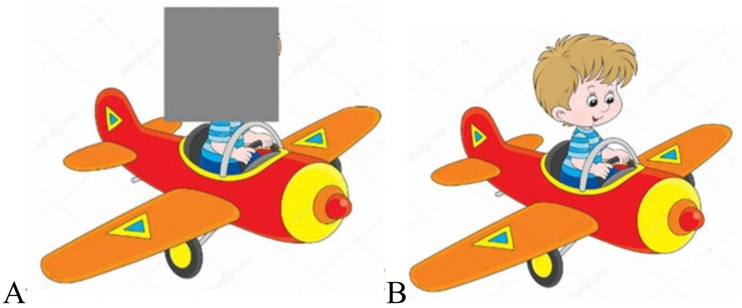

Fig 1. Illustration of the game “under the shape”. The sentence describes here is 張生揸飛機 "Mr. Cheung is operating an airplane", where all the words have Cantonese Tone 1.

For each sentence, a series of questions were designed to elicit each desired types of focus, namely, broad focus, non-contrastive narrow focus, and contrastive narrow focus. The positions of focus are initial, middle, or final positions. One picture covered by a grey shape was presented to participants in each trial. The experimenter will proceed to ask a question about the presented pictures. For example, in Fig 1, the participants were presented with the picture with a grey shape covering the person flying an airplane, and the experimenter asked in Cantonese, "Who is operating an airplane?" Then, the experimenter pressed a button and the grey shape on the picture was removed. The participant was then expected to answer the experimenter's question by saying "Mr. Cheung is operating an airplane" with a focus on the subject. If a participant made a mistake in answering the question, namely, did not use the five-syllable answer required, the experimenter would ask the question again rather than simply ask for a correction so as to elicit a natural response. The maximum number of attempts was three, and none of the participants failed to correct themselves in this experiment.

To examine the acquisition of Mandarin speech prosody, Cantonese-speaking children with ASD, TD native Cantonese-speaking children and TD native Mandarin-speaking children participated in this experiment. 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 communication languages. The intelligence and language ability of the participants were evaluated with part of 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).

Similarly, 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. 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.

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.

For the training, each session consisted three phases. In the first phase, the children were presented with congruous pairs only (i.e., the answers contained the correct types of focus on the correct constituents as the response to the questions). After listening to each pair of question and answer sung, they were required to identify which part of the answer was on what kind of focus by pressing on the buttons corresponding to subject, verb and object different times: zero press for broad focus, one press for narrow focus, and two presses for contrastive focus. In the second phase, children were presented with congruous and incongruous pairs alternatively, and they were asked to press the corresponding buttons to indicate whether the pairs they heard was congruous or not. The third phase involves the presentation of congruous and incongruous pairs randomly with feedback.

(h) Data collection and analysis

For data analysis of pre- and post-training speech production in Cantonese, 9660 target sentences (15 sentences * 7 conditions * 2 repetitions * 23 participants * 2 groups) were acoustically analyzed for f0, duration and intensity. The five syllables of each sentences were manually segmented using Praat (Boersma, 2001), following the procedure of segmentation written by Jangjamras (2011). Obstruents were not included into the segmentation and we focused only on the sonorant parts of the syllables. The data were extracted using ProsodyPro (Xu, 2013), and abnormal data were mannually checked by the first and second authors. In total, 5285 syllables were removed from the 48300 syllables due to creakiness and other abnormality. None of the participants had data loss larger than 20 percent.

The f0 range (i.e. the difference between maximum and minimum f0), the mean f0, the duration and mean intensity of the sonorant part were calculated for each syllable in each sentence. These four acoustic parameters were treated as the dependent (i.e., outcome) variables as they are widely used in prosodic marking cross-linguistically. The two f0 parameters can also index children's performance of tone realization.

For independent (i.e., explanatory) variables, we were interested in the influence of Participant Group (i.e. ASD vs. TD), Focus Condition of the syllables, Tone Shape, Prosodic Complexity of the sentence and their interaction. Focus Condition was defined as the relative position to focus of a syllable, that is, 1) carrying broad focus (i.e. On-broad-focus), 2) preceding a syllable carrying contrastive or non-contrastive narrow focus (i.e. Pre-narrow-focus), 3) carrying narrow focus (i.e. On-narrow-focus), and 4) following a syllable carrying contrastive or non-contrastive narrow focus (i.e. Post-narrow-focus). Tone Shape refers to the shape of tones carried by each syllable, which was grouped into 1) Non-low Level (Tone 1 and 3), 2) Rising (Tone 2 and 5) and 3) Low (Tone 4 and 6) tones. Prosodic Complexity was defined based on the tonal combination of the answers, which was grouped into 1) Single-tone (i.e. the five syllables in an answer carries the same tone) and 2) Mixed-tone (i.e. the two subject syllables carries a different tone from the verb and object syllables in an answer).

Linear mixed effects (LME) models were fitted to evaluate the fixed effects and their interactions on the four outcome variables using lmer4 package (Bates et al., 2015) in R (Team R, 2024). The optimal fixed structure of each model was selected by stepwise comparisons from the simplest structure to the most complex, and Likelihood Ratio (LR) tests were used to determine whether including factors from the analysis led to a better fit. Tukey post-hoc tests were used for post-hoc comparisons of the interactions of interests using emmeans (Kuznetsova, 2017).

For data analysis of pre- and post-training speech production in Mandarin, 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 f0 range (= maximum f0 value - minimum f0 value), the mean f0, 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.

(i) Results and Discussion

For the acquisition of Cantonese, Cantonese-speaking children with ASD employed the same acoustic cues to mark focus as their TD peers, but used them in different ways. Both the ASD and TD groups expanded f_0 range and duration of the on-focus syllables while compressed the intensity of the post-focus syllables; nevertheless, the degree of on-focus expansion in the ASD group was smaller, and the two groups' use of these acoustic cues show tone-specific patterns. Since the ASD and TD groups in the present study did not significantly differ from each other in IQ scores and language abilities, the clinical condition may be the primary factor that led to the results observed here.

In terms of f_0 range, the autistic children in our study did not produce on-focus syllables with an expansion of f_0 range compared to their TD peers. Autistic children did not only produce contour tones with significantly smaller f_0 range than TD children at the post-narrow-focus position, but also low tones regardless of focus condition. In other conditions, the f_0 range produced by the TD group was also slightly larger, though the difference did not reach statistical significance. At the first glance, this finding seems to be in line with early studies that reported prosodic production among the autistic population to be monotonic and machine-like (for review see (Peppé, 2003)). However, since more recent studies suggest that the population with ASD tends to produce sing-songy prosody, we attribute these results to the autistic children's failure to implement lexical and utterance prosody simultaneously, that is, to produce lexical tone accurately while marking information structure clearly.

With regard to duration, while both the autistic and TD children produced long post-narrow-focus syllables, such lengthening may be due to the final lengthening (see (Wong, 2002) for instance). This is because two-thirds of the post-narrow-focus syllables fell on objects, namely, the last words of the sentences. The present finding is more in line with the findings by Paul et al. and Grossman et al. that English speakers with ASD did not lengthen the stressed syllables enough. However, unlike in Diehl & Paul's study, the autistic individuals in our study did not over-lengthen the syllables carrying no focus as pre-narrow-focus syllables produced by our autistic participants were the shortest. The differences between the present finding and Diehl & Paul's study may be due to the differences in language background, namely, their participants were English speakers while ours were Cantonese speakers. Unlike English which used f_0 patterns to mark utterance focus (Gussenhoven, 1994), the major cue used for focus marking in Cantonese is the on-focus expansion of duration. Therefore, our participants with ASD still showed a tendency of on-focus lengthening, though not as sufficient as the TD peers.

In addition, we found an overall influence of lexical tones on the use of acoustic cues in both the ASD and TD groups, indicating that children face extra difficulties in marking prosodic focus in a tonal language. On the one hand, children need to vary f_0 (and other acoustic cues) so as to produce accurate lexical tones. Previous studies have found that autistic children have speech-related deficits in tone production. Autistic children showed more f_0 variations in imitating Mandarin lexical tones, but not in imitating non-speech stimuli (Chen et al., 2022). On the other hand, they need to mark focus using acoustic cues involved in tone production. The difficulties in encoding both the lexical and focal function may have led to the smaller f_0 range produced by the autistic children than the TD peers in general. The difficulties observed in focus marking especially for low tones in the present study may be due to the extra difficulty involved in low tone acquisition and production (Wong & Strange, 2017). Moreover, for the ASD group, only on low tones were the on-narrow-focus syllables longer than on-broad-focus. Our results thus showed that the ASD group could mark focus using on focus expansion of duration only on the low tone. The low tone is reported to be among the shortest of Cantonese tones in its citation form, the lengthening in on-narrow-focus syllables may thus be more dramatic than other tones in focus marking due to its original short duration (Kong, 1987) .Also, it seems that final lengthening is more prominent on non-low level tones for both groups. It may be due to the fact that non-low level tones tend to have longer duration in the citation form and thus the final lengthening effect may be more prominent.

Based on these findings, we propose that Cantonese-speaking children with ASD did not use on-focus expansion in f_0 range and intensity to mark focus, but showed some post-focus compression in these two cues. It is worth mentioning, however, unlike Mandarin and English, Cantonese is not a language with typical post-focus compression (Wu, 2021). The seemingly smaller f_0 range in post-focus syllables may alternatively be explained by the lack of f_0 range expansion in the on-focus syllables, since in the ASD group no significance was found in f_0 range between pre-focus and on-focus syllables when the embedded tones were level and rising tones and syllables on broad focus had the smallest f_0 range when carrying low tones.

According to the neuro-imaging study conducted by Eigsti et al. (2012), more generalized neural regions were activated in the ASD group compared to the TD group.

Echoing Eigsti et al, Yu et al. (2022) also found that different from the TD children, children with ASD did not show left-lateralized late negative response distinction when processing native lexical prosody. The reduced neural specialization involved in linguistic prosody processing may lead to the fact that the autistic population need cognitive control and resources in processing prosody, which is intrinsically challenging because it involves integration from multiple levels of language. As a result, the ASD group in the present study had some difficulties in marking focus and failed to keep as distinctive shapes of lexical tones as the TD peers while marking focus at the same time. ASD children were also reported to have difficulties in mapping acoustic cues and information structure (Chen, 2021). Although they may use syntactic cues in comprehending focus, the ability to use prosodic cues to comprehend focus was significantly worse compared to their TD peers (Ge et al., 2022). It has been reported that prosodic cues may help identify alternatives and affects implicature computation. The deficits in the mapping thus may lead to weaker identification of alternatives and implicature computation (Gotzner, 2019). In turn, the deficit may lead to difficulties in using acoustic cues to mark information structure in speech production.

For the acquisition of Mandarin, 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.

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 ASD adds to such difficulties.

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), 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. 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, 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).

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. As significant differences between tones in the two Cantonese-speaking groups than in the native Mandarin-speaking group, and the tone production in the latter was influenced more by focus conditions. It may be argued that these two non-native groups tend to prioritize the acoustic realization of lexical prosody, that is, the lexical tones over that of utterance prosody, the focus, 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.

Regarding the effects of sung-speech training on the acquisition of Cantonese speech prosody, we identified the following two conditions as indicators of improvement after musical training: 1) Post-Training Correct Adjustment: In their pre-training production, they did not show significant adjustment of the target prosodic cues (i.e., they did not adjust the prosodic cues or adjusted them incorrectly, such as pre-focus/post-focus increase but on-focus compression). In their post-training production, they were able to adjust the target prosodic cues correctly. 2) Post-Training Error Reduction: In their pre-training production, they showed significant but incorrect adjustment of the target prosodic cues. After musical training, the significant adjustment disappeared, indicating that they overcame the incorrect adjustment, although they still could not adjust the prosodic cues correctly.

For duration, we observed post-training correct adjustment primarily in pre-focus syllables, indicating that after musical training, the children improved their ability to compress pre-focus syllables to signal the upcoming focus. In terms of intensity, we noted both post-training correct adjustment and post-training error reduction in pre-focus and post-focus syllables, as well as post-training error reduction in on-focus syllables for T6.

For mean f0, the improvement was mainly in T1, T5, and T6. Children showed post-training correct adjustment and error reduction in pre-focus and post-focus syllables. Conversely, the improvement in f0 range was mainly seen as post-training error reduction. We observed error reduction in pre-focus and post-focus syllables of all the tones except for T1, and in the on-focus syllable of T3.

Our findings demonstrate a close correlation between music and speech, supporting theoretical frameworks such as OPERA hypothesis (Patel, 2011, 2012,_ 2014). When the processing materials and cognitive mechanisms are shared, the beneficial effect of musical training on speech processing is established. Specifically, this close correlation benefits autistic children's prosodic focus production. Autistic individuals are found to have deficits in processing socio-communication auditory information (Paul et al., 2007) and are more sensitive to non-speech sounds than speech sounds (e.g., Chen et al., 2022). In this study, we transfer the prosodic patterns in the sounds they are less sensitive to (i.e., speech) to sounds they are more sensitive to (music). When processing music, their cognitive load is lower, allowing them to attend to and learn the prosodic patterns embedded in it, significantly improving their prosodic focus production.

Additionally, autistic children are known to have disorders in multisensory coordination (Baum et al., 2015). It is likely that they know what to produce but have difficulties coordinating and articulating prosodic cues. Previous studies have demonstrated that musical training enhances individuals' ability to coordinate between different sensory modalities. Music can engage multisensory mechanisms, thereby refining vocal production (Stegemöller et al., 2008). Our findings support this correlation, showing that musical training improves autistic children’s coordination of their vocal articulation to more closely align with their intended prosodic patterns.

Furthermore, we found that it is more challenging to train autistic children to use f0 range to indicate focus. However, through musical training, we observed error reduction, indicating that musical training can adjust children’s atypical prosody production and make their sentences sound less confusing.

Regarding the effects of sung-speech training on the acquisition of Mandarin speech prosody, positive training effects have been found in different acoustic parameters, focus conditions and tones, though the major improvement was seen on their use of OFE, especially in non-contrastive focus marking. After training, the autistic children showed significant OFE in the f0 range of T2 and T4 when marking contrastive and non-contrastive focus, and they also learned to mark non-contrastive focus with OFE in the mean f0 of T1 and T2. At the same time, their use of OFE in intensity also became evident when marking broad and contrastive focus after the training session. Although PFC was notoriously difficult to learn, the autistic group showed a tendency of PFC in the f0 range of T2.

The positive training effects add to the evidence that song-based training facilitates speech prosody acquisition in the autistic population, even in non-native languages. Using similar acoustic cues in music and speech, the sung speech is designed in a way to boost the acoustic cues such as f0, duration and intensity in on-focus words reflected in the melody line and rhythmic pattern, which reinforces the acoustic cues generally used in speech and in turn may have improved their mapping of prosodic cues to information(structure (Lima and Castro, 2011) and help them to transfer what they have learned through the music to speech production. The facilitating effects of the current training may have been further enhanced by our design, namely, we did not conform the melodies to lexical tones but to focus marking prosody only. In this way, the trainees may be able to mainly pay attention to the pitch patterns on the utterance-level than the lexical tones. Their post-training performance, therefore, may be less constrained by accurate tone realization and hence they showed more evident OFE than their TD peers. The exaggerated realization of acoustic cues on constituents on non-contrastive focus in the songs may help them realize that that they also need to use prosodic cues to make the newly given information more prominent in answering wh-questions.

(j) Conclusions and Recommendations

To conclude, this project has found that Cantonese-speaking children with ASD did not use as sufficient on-focus expansion to mark focus in Cantonese as their TD peers. The children with ASD also produced less distinctive f_0 range for different tone shapes and focus conditions than TD children, but their focus-marking was not influenced by the prosodic complexity of the sentences. The findings of the present study have clinical implications. Our findings suggest that Cantonese-speaking children with ASD are not as sophisticated in prosodic focus marking as their TD peers, and therefore requires specific training, especially on how to retain distinctive f_0 range for different tone shapes while marking focus more evidently.

In addition, this project 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. 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, when compared with the 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 the knowledge of OFE as their TD peers in their third language. Therefore, multilingual exposure to children with ASD may not necessarily bring negative effect. It is not recommended that autistic children are restricted from multilingual exposure.

Sung speech training has been shown to be effective in improving the use of prosodic cues in both Cantonese and Mandarin. It is recommended that more song-based interventions are used to improve speech prosody.

(k) Bibliography *(use APA Editorial Style throughout the report)*

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). https://doi.org/10.1038/s41598-021-02487-6

Bao, M. (2008). *Phonetic realization and perception of prominence among lexical tones in Mandarin Chinese*. University of Florida.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, *67*, 1–48. https://doi.org/10.18637/jss.v067.i01

Baum, S. H., Stevenson, R. A., & Wallace, M. T. (2015). Behavioral, perceptual, and neural alterations in sensory and multisensory function in autism spectrum disorder. *Progress in neurobiology*, 134, 140-160.

Boersma, P. (2001). Praat, a system for doing phonetics by computer. *Glot. Int.*, *5*(9), 341-345.

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

Cao, W. (2004). A preliminary analysis of focus and ending in Chinese intonation. In *Speech Prosody 2004, International Conference*.

Chen, F., Cheung, C. C., & Peng, G. (2022). Linguistic Tone and Non-Linguistic Pitch Imitation in Children with Autism Spectrum Disorders: A Cross-Linguistic Investigation. *Journal of Autism and Developmental Disorders*, *52*(5), 2325–2343. https://doi.org/10.1007/s10803-021-05123-4

Chen, S., He, Y., Wayland, R., Yang, Y., Li, B., & Yuen, C. W. (2019). Mechanisms of tone sandhi rule application by tonal and non-tonal non-native speakers. *Speech Communication*, *115*, 67–77. https://doi.org/10.1016/j.specom.2019.10.008

Chen, S., Zhang, Y. X., Zhou, F., Chan, A., Li, B., Li, B., Tang, P.Y., Chun, E., Chen, Z. M. (accepted) Focus-marking in a tonal language: prosodic differences between Cantonese-speaking children with and without autism spectrum disorder. PloS One.

Cutler, A., & Pearson, M. (2018). On the analysis of prosodic Turn-Taking cues. In *Routledge eBooks* (pp. 139–156). https://doi.org/10.4324/9780429468650-8

DePape, A. R., Hall, G. B. C., Tillmann, B., & Trainor, L. J. (2012). Auditory Processing in High-Functioning Adolescents with Autism Spectrum Disorder. *PloS One*, *7*(9), e44084. https://doi.org/10.1371/journal.pone.0044084

Diehl, J. J., & Paul, R. (2009). The assessment and treatment of prosodic disorders and neurological theories of prosody. *International Journal of Speech-language Pathology*, *11*(4), 287–292. https://doi.org/10.1080/17549500902971887

Diehl, J. J., & Paul, R. (2011). Acoustic and perceptual measurements of prosody production on the profiling elements of prosodic systems in children by children with autism spectrum disorders. *Applied Psycholinguistics*, *34*(1), 135–161.

https://doi.org/10.1017/s0142716411000646

Eady, S. J., & Cooper, W. E. (1986). Speech intonation and focus location in matched statements and questions. *The Journal of the Acoustical Society of America, 80*(2), 402–415. https://doi.org/10.1121/1.394091

Eigsti, I., Schuh, J., Mencl, E., Schultz, R. T., & Paul, R. (2012). The neural underpinnings of prosody in autism. *Child Neuropsychology/Neuropsychology, Development, and Cognition. Section C, Child Neuropsychology*, *18*(6), 600–617. https://doi.org/10.1080/09297049.2011.639757

Féry, C., & Kügler, F. (2008). Pitch accent scaling on given, new and focused constituents in German. *Journal of Phonetics*, *36*(4), 680–703. https://doi.org/10.1016/j.wocn.2008.05.001

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

Fusaroli, R., Lambrechts, A., Bang, D., Bowler, D. M., & Gaigg, S. B. (2016). "Is voice a marker for Autism spectrum disorder? A systematic review and meta-analysis." *Autism Research*, *10*(3), 384–407. https://doi.org/10.1002/aur.1678

Ge, H., Liu, F., Yuen, H. K., Chen, A., & Yip, V. (2022). Comprehension of prosodically and syntactically marked focus in Cantonese-Speaking children with and without autism spectrum Disorder. *Journal of Autism and Developmental Disorders*, *53*(3), 1255–1268. https://doi.org/10.1007/s10803-022-05770-1

Gotzner, N. (2019). The role of focus intonation in implicature computation: a comparison with only and also. *Natural Language Semantics*, *27*(3), 189–226. https://doi.org/10.1007/s11050-019-09154-7

Gu, W., & Lee, T. (2007, August). Effects of tonal context and focus on Cantonese F0. In Proceedings of the 16th International Congress of Phonetic Sciences (ICPhS 2007) (pp. 1033-1036).

Gussenhoven, C. (1994). Focus and sentence accents in English. *Focus and natural language processing*, *3*, 83-92.

Hombert, J. M. (1976). Difficulty of producing different F in speech. *The Journal of the Acoustical Society of America*, *60*(S1), S44-S45.

Hombert, J. M. (1978). A model of tone systems. *Elements of Tone, Stress and Intonation*, 129-143.

Ishihara, S. (2011). Japanese focus prosody revisited: Freeing focus from prosodic phrasing. *Lingua*, *121*(13), 1870–1889. https://doi.org/10.1016/j.lingua.2011.06.008

Jangjamras, J. (2011). Perception and Production of English Lexical Stress by Thai Speakers. In *ProQuest LLC eBooks*. https://eric.ed.gov/?id=ED539316

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

Kong, Q. (1987). Influence of Tones upon Vowel Duration in Cantonese. *Language and Speech*, *30*(4), 387–399. https://doi.org/10.1177/002383098703000407

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: tests in linear mixed effects models. *Journal of statistical software*, *82*(13).

Lee, Y. C., Wang, T., & Liberman, M. (2016). Production and perception of tone 3 focus in Mandarin Chinese. *Frontiers in psychology*, *7*, 175546. https://doi.org/10.3389/fpsyg.2016.01058

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, C. N., & Thompson, S. A. (1978). The acquisition of tone. In *Tone* (pp. 271-284). Academic Press.

Lima, C. F., & Castro, S. L. (2011). Speaking to the trained ear: musical expertise enhances the recognition of emotions in speech prosody. Emotion, 11(5), 1021.

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. *Journal of Autism and Developmental Disorders*.

Man, V. C. H. (2002, April 11). *Focus Effects on Cantonese Tones: An Acoustic study*. https://doi.org/10.21437/speechprosody.2002-102

Mok, P. P., Fung, H. S., & Li, J. (2014, May 20). *A preliminary study on the prosody of broadcast news in Hong Kong Cantonese*. https://doi.org/10.21437/speechprosody.2014-204

Nadig, A., & Shaw, H. (2012). 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

Ouyang, I. C., & Kaiser, E. (2015). Prosody and information structure in a tone language: an investigation of Mandarin Chinese. *Language, Cognition and Neuroscience*, *30*(1-2), 57-72. https://doi.org/10.1080/01690965.2013.805795

Paul, R., Augustyn, A., Klin, A., & Volkmar, F. R. (2005). Perception and Production of Prosody by Speakers with Autism Spectrum Disorders. *Journal of Autism and Developmental Disorders*, *35*(2), 205–220. https://doi.org/10.1007/s10803-004-1999-1

Paul, R., Chawarska, K., Fowler, C., Cicchetti, D., Volkmar, F (2007) Listen my children and you shall hear: auditory preferences in toddlers with autism spectrum disorders. Journal of Speech Language and Hearing Research, 50 (5) (2007), p. 1350, 10.1044/1092-4388(2007/094)

Peppé, S., & McCann, J. (2003). Assessing intonation and prosody in children with atypical language development: the PEPS-C test and the revised version. *Clinical Linguistics & Phonetics*, *17*(4–5), 345–354. https://doi.org/10.1080/0269920031000079994

Peppé, S., McCann, J., Gibbon, F., O'Hare, A., & Rutherford, M. (2006). Assessing prosodic and pragmatic ability in children with high-functioning autism. *Journal of Pragmatics*, *38*(10), 1776–1791. https://doi.org/10.1016/j.pragma.2005.07.004

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

Raven, J. (1989). The Raven Progressive Matrices: A review of national norming studies and ethnic and socioeconomic variation within the United States. *Journal of Educational Measurement*, *26*(1), 1–16. https://doi.org/10.1111/j.1745-3984.1989.tb00314.x

Raven, J. (2003). Raven progressive matrices. In *Handbook of nonverbal assessment* (pp. 223-237). Boston, MA: Springer US.

Schneider, W., Eschman, A., & Zuccolotto, A. (2002). *E-Prime (Version 2.0)*. [Computer software and manual]. Pittsburgh, PA: Psychology Software Tools Inc.

Schneider, W., Eschman, A., & Zuccolotto, A. (2002). *E-prime (version 2.0). Pittsburgh, PA: Psychology Software Tools Inc*. Retrieved 18/04/2019 from https://pstnet. com/products/e-prime.

Slabakova, R. (2017). The scalpel model of third language acquisition. *International Journal of Bilingualism*, *21*(6), 651-665. https://doi.org/10.1177/1367006916655413

T'sou, B., Lee, T., Tung, P., Man, Y., Chan, A., To, C. K. S., & Chan, Y. (2006). Hong Kong Cantonese oral language assessment scale. *Hong Kong: City University of Hong Kong*.

Team, R. (2024). RStudio: integrated development for R. RStudio, PBC, Boston, MA. 2020.

Thompson, W. F., Schellenberg, E. G., & Husain, G. (2004). Decoding speech prosody: Do music lessons help?. *Emotion*, 4(1), 46.

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*, 1-20. https://doi.org/10.1044/2023_JSLHR-23-00508

Wechsler, D., & Kodama, H. (1949). *Wechsler intelligence scale for children* (Vol. 1). New York: Psychological corporation.

Wong, P., & Strange, W. (2017). Phonetic complexity affects children's Mandarin tone production accuracy in disyllabic words: A perceptual study. *PloS One*, *12*(8), e0182337. https://doi.org/10.1371/journal.pone.0182337

Wong, W. Y. P., Brew, C., Beckman, M. E., & Chan, S. D. (2002). Using the segmentation corpus to define an inventory of concatenative units for Cantonese speech synthesis. In *COLING-02: The First SIGHAN Workshop on Chinese Language Processing*.

Wu, W. L., & Xu, Y. (2010, May 10). *Prosodic focus in Hong Kong Cantonese without post-focus compression*. https://doi.org/10.21437/speechprosody.2010-85

Xu, Y. (1999). Effects of tone and focus on the formation and alignment of f0contours. *Journal of Phonetics*, *27*(1), 55–105. https://doi.org/10.1006/jpho.1999.0086

Xu, Y. (2011, August). Post-focus Compression: Cross-linguistic Distribution and Historical Origin. In *ICPhS* (pp. 152-155).

Xu, Y. (2013). ProsodyPro—A tool for large-scale systematic prosody analysis. Laboratoire Parole et Langage, France.

Xu, Y., & Xu, C. X. (2005). Phonetic realization of focus in English declarative intonation. *Journal of Phonetics*, *33*(2), 159–197. https://doi.org/10.1016/j.wocn.2004.11.001

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.

Yu, L., Huang, D., Wang, S., & Zhang, Y. (2022). Reduced Neural Specialization for Word-level Linguistic Prosody in Children with Autism. *Journal of Autism and Developmental Disorders*, *53*(11), 4351–4367. https://doi.org/10.1007/s10803-022-05720-x