This is the accepted version of the following article: Wang Bruce, 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 which has been published in https://doi.org/10.1044/2023_JSLHR-23-00508.

English prosodic focus marking by Cantonese trilingual children with and without autism spectrum disorder

Bruce Xiao Wang[1], Si Chen[2], Fang Zhou[2], Jiang Liu[3], Cheng Xiao[3], Angel Chan[2], Tempo Tang[2]

[1]Department of English and Communication, Hong Kong Polytechnic University; [2]Department of Chinese and Bilingual Studies, Hong Kong Polytechnic University; [3]Department of Languages, Literatures and Cultures, University of South Carolina

Correspondence to Si Chen: sarah.chen@polyu.edu.hk

Abstract

Purpose:

The current study investigated English prosodic focus marking by autistic and typically-developing (TD) Cantonese trilingual children, and examined the potential differences in this regard compared to native English-speaking children.

Method:

Forty-eight participants were recruited with 16 speakers for each of the three groups (CASD: Cantonese-speaking autistic group; CTD: Cantonese-speaking TD group; ETD: English-speaking TD group), and prompt questions were designed to elicit desired focus type (i.e., broad, narrow and contrastive focus). Mean duration, mean f0, f0 range, mean intensity and f0 curves were used as the acoustic correlates for linear mixed effects model fitting and functional data analyses in relation to groups and focus conditions (i.e., broad, narrow and contrastive pre-, on- and post-focus).

Results:

The CTD group had post-focus compression (PFC) patterns via reducing mean duration, narrowing f0 range, and lowering mean f0, f0 curve and mean intensity for words under both narrow and contrastive post-focus conditions, while the CASD group only had shortened mean duration and lowered f0 curves. However, neither the CTD nor CASD group showed much of on-focus expansion (OFE) patterns. The ETD group marked OFE by increasing mean duration, mean f0, mean intensity and higher f0 curve for words under on-focus conditions.

Conclusions:

The CTD group utilised more acoustic cues than the CASD group when it comes to PFC. The ETD group differed from the CASD and CTD groups in the use of OFE. Further, both the CASD and CTD groups showed positive L1 transfer in the use of duration and intensity and, potentially, successful acquisition in the use of f0 for prosodic focus marking. Meanwhile, the differences in the use of OFE between the Cantonese-speaking and English-speaking groups, not PFC, might indicate that Cantonese-speaking children acquire PFC prior to OFE.

Keywords: prosodic focus marking; autism spectrum disorder; trilingual children

1. Introduction

Autism spectrum disorder (ASD) is a neurodevelopmental disorder that affects various aspects of life (Asghari et al., 2021; Lord et al., 2018), with deficits in social communication and interaction (be it verbal or non-verbal) being one of the major symptoms (American Psychiatric Association, 2013). Prosody portrays the suprasegmental features of speech and serves an important role in communicative functions (Peppé et al., 2007). Prosodic cues used in focus marking indicate the relative prominence of certain segments and signal out the new or corrective information in an utterance (Bolinger, 1972; Cruttenden, 1997; Ladd, 2008; Xu & Xu, 2005). For example, a lengthened syllable in the sentence medial position could mark the centre of informativeness in a discourse. While the encoding and decoding of prosodic focus marking in a discourse seems to be straightforward for typically developing (TD) individuals, many studies (McCann & Peppé, 2003; Peppé et al., 2007) have reported impaired performance among autistic individuals. The main difficulties in understanding social norms in communication for autistic individuals often results from the lack of the ability to perceive and produce reciprocal prosodic cues (American Psychiatric Association, 2013; Thye et al., 2018). Many studies have investigated the prosodic focus marking patterns of autistic speech among monolingual speakers. However, there is a relative scarcity of research looking into the multilingual community as well as the acquisition of focus marking in a second language, despite significant prevalence of the multilingual community in the modern society. In the context of Hong Kong, for example, over 50% of the population (appx. 4 million people; Population By-census, 2016) use at least two languages in their daily communications. It is then essential to investigate the effect of multilingual exposure on the autistic population, especially when studies have shown potential benefits to bilingual exposure for individuals with autism. For example, Ratto et al. (2020) investigated the effects of bilingual exposure on executive functioning and social communication skills among monolingual and bilingual autistic children. All of the bilingual children had English as L2, but different L1 backgrounds such as Amharic, Arabic, and Mandarin etc. They found that the bilingual group had fewer executive functioning problems and better social communication skills when compared to the monolingual group. Further, a recent study reported that bilingual exposure might enhance the development of L1 focus marking (Ge et al., 2023). Specifically, Ge et al. compared the focus production of autistic L1 Cantonese L2 English bilingual children to their monolingual autistic peers and monolingual and bilingual typically developing children, showing that bilingual autistic children performed equally well to typically developing children and outperformed autistic monolingual autistic children. Recognising the potential benefits of a bilingual or/and multilingual context on the development of focus in autistic children, it is crucial to gain a deeper understanding of focus marking production in trilingual autistic children. The current study examines the production of English prosodic focus marking of trilingual (i.e., Cantonese, English, Mandarin) Hong Kong Cantonese-speaking autistic children (hereafter the CASD group) and compares their production with that of Hong Kong Cantonese-speaking TD children (hereafter the CTD group) and native English-speaking TD children (hereafter the ETD group).

1.1 Focus and its acoustic correlates

Halliday (1967) first introduced the concept of information structure, which describes the focal point of information within a discourse. The contrast between the focal and non-focal point can be marked by prosodic cues. Truckenbrodt (1995) further proposed the focus prominence rule. Namely, more prominence is given to the words or syllables that correspond to the focused element in a discourse, which normally contains information not given in the preceding conversation, i.e., accompanied by new information to the receiver (Lambrecht, 1994). According to Halliday (1976), new information is marked by three criteria: 1) new information needs to be the most important part in the current discourse; 2) new information is contrastive to what occurred in the preceding discourse; 3) new information answers the WH-question. Generally speaking, focus can be categorised into broad, narrow and contrastive conditions depending on the preceding questions (Cruttenden, 1997). Table 1 presents an example of focus types as a means of the preceding questions. In broad focus (Table 1 a.), the entire sentence is under focus (Lambrecht, 1994, 2000), while in narrow focus (b) and contrastive focus (c), the on-focus words (in bold and italics) have higher prominence than pre-focus (i.e., "Amy") and post-focus (i.e., "doll") words (Halliday, 1967; Xu & Xu, 2005). In narrow and contrastive focus, although the on-focus word could be the same (i.e., in sentences b. and c.), the difference lies in that the contrastive focus is accompanied by a limited number of alternatives (i.e., potential answers to the questions in c.) in the discourse (Chen, 2018) while narrow focus is not.

Table 1 Focus type as a function of the prompt questions. Words on focus are marked by bold italics (see Supplemental material A for the full list of questions).

Many studies have explored the acoustic correlates of English prosody and focus marking. Early studies such as Fry (1955) examined the role of duration and intensity in speech prosody and showed that duration plays a more salient cue than intensity for judgments of words on focus. Lieberman (1960) later investigated patterns of f0, intensity, and duration in relation to stressed and unstressed syllables of American English. Although Lieberman found similar f0 patterns to Fry, i.e., f0 is an important cue for signalling focus, he concluded that intensity played a more salient cue than duration. Further, Cooper et al. (1985) reported that native English adult speakers consistently produced on-focus words with longer duration than off-focus words in both short and long sentences; meanwhile, a sharp drop was observed in the f0 of post-focus words in short sentences. More recently, Xu & Xu (2005) concluded that on-focus words showed on-focus expansion (OFE), exhibiting a higher maximum f0, larger rise size and shorter rise time in comparison to those at pre-focus and post-focus locations. Meanwhile, post-focus words are often accompanied by a lowered f0 (i.e., a pattern of post-focus compression; PFC). Similar findings were reported by Breen et al. (2010), where they found that words at on-focus locations in general have longer duration, higher mean f0 and intensity than those at off-focus locations. Although previous studies showed no consensus on which acoustic feature contributes most to the prosodic focus marking in a discourse (Guo, 2022), it can be concluded that f0, duration and intensity are the three most studied acoustic correlates of prosodic focus marking in not only English, but other languages such as German (Mücke & Grice, 2014), Cantonese (Wu & Xu, 2010), Finnish (Arnhold & Kyröläinen, 2017), Mandarin (Yang & Chen, 2014) .

1.2 Focus marking by autistic individuals

Many studies have attempted to explore distinctive prosodic features in autistic speech using an acoustic approach. For example, early studies have investigated the duration and frequency of pauses in speech produced by individuals with autism (e.g, Feldstein et al., 1982; Thurber & Tager-Flusberg, 1993), and some later ones have investigated the differences in duration, f0 and intensity between autistic individuals and their TD counterparts (e.g., Bone et al., 2016; Diehl & Paul, 2012; Filipe et al., 2014; Hubbard & Trauner, 2007; Paul et al., 2008; Sharda et al., 2010). Although atypical prosody in autistic speech has been widely reported (McCann & Peppé, 2003; Peppé et al., 2007), most of the previous studies have reported differences in prosody production in general (see Asghari et al., 2021; Fusaroli et al., 2017 for a review), e.g., the autistic group has overall higher mean f0 and wider f0 range than the TD group. Only a few studies have explored the use of prosodic focus marking among autistic individuals and how they differed from their TD peers. For example, Paul et al. (2008) compared word duration between autistic and TD speech in relation to stressed and unstressed conditions. They used an imitation task showing that the duration of stressed words was significantly longer than that of unstressed words for both groups. In addition, Nadig & Shaw (2015) investigated more acoustic correlates (i.e., duration, mean f0 and intensity) of prosodic focus marking in relation to focus positions (i.e., on-focus vs. post-focus). Their findings on word duration resembled those reported by Paul et al (2008) where both the autistic and TD group produced significantly longer duration for on-focus words than that for post-focus words. They also reported that both autistic and TD groups had significantly higher intensity for on-focus syllables than post-focus syllables; however, neither the autistic nor TD children had such patterns in terms of mean f0. The reported acoustic patterns suggest that autistic individuals might have similar knowledge to their TD counterparts in using acoustic cues to mark focus and signal sentence prominence. However, DePape et al. (2012) showed that the ability of autistic individuals using focus marking can vary in relation to language skills, where they found that autistic individuals with moderate language skills tend to use f0 range to mark information structure, whereas those with high language functioning did not. The findings from DePape et al seem to be counter-intuitive; however, they explained that although autistic children with high language skills had some language advantages over those with moderate language skills, they did not necessarily have better knowledge of where to and how to vary f0 range for marking information prominence in an utterance. DePape et al. also mentioned that the difference in marking information structure between autistic children with high and moderate language skills might also be attributed to prior speech intervention received (see DePape et al. 2012 for detailed discussion).

1.3 Prosodic focus marking by bilingual speakers

Some studies have investigated L2 prosodic focus marking patterns among adult bilingual speakers (e.g., Fung & Mok, 2014; Guo, 2022; Lee & Xu, 2020; Liu et al., 2019; Wu & Chung, 2011; Wu & Xu, 2010), suggesting that the acquisition of prosodic focus marking in L2 does not just depend on prosodic similarities between languages, but, more importantly, the background of bilingual speakers, e.g., early (simultaneous, sequential) and late bilinguals and daily contact of L2. Wu & Chung (2011) and Fung & Mok (2014) investigated the f0 variability of L1 Cantonese L2 English speakers as a function of whether the target word was on- or off-focus showing that prosodic focus marking had a low transferability across languages. Specifically, their findings suggested that the Cantonese speakers did not show similar f0 patterns in their English production, i.e., OFE and PFC, to native English speakers. Since duration and intensity are the main acoustic cues for prosodic focus marking in Cantonese (Wu & Xu, 2010), the discrepancy in f0 patterns between Cantonese and English speakers in their English production is likely to be a negative L1 transfer of prosodic focus marking. In a different study, Liu et al. (2019) explored the production of prosodic focus marking of L1 Korean L2 English bilinguals as a function of English proficiency (i.e., groups with advanced, intermediate and low English proficiency). Among the three acoustic correlates (duration, f0 range and mean intensity) they investigated, they showed that the advanced L2 English speakers used all three acoustic correlates to indicate prosodic focus marking and had a similar OFE pattern (but no PFC) to the native English speakers; meanwhile, groups with intermediate and low English proficiency did not use any acoustic cues to indicate OFE nor PFC in their English production. Since both English and Korean have OFE and PFC, the lack of positive L1 transfer of prosodic focus marking indicates that language proficiency has an important role in the acquisition of L2 prosody; meanwhile, studies also reported that PFC is likely to be acquired after

OFE (Liu et al., 2019) and has low transferability between languages (Chen et al., 2012; Liu et al., 2019; Wu & Chung, 2011).

Many other studies have also investigated prosodic focus marking among other adult bilingual communities, e.g., Mandarin - English (Guo, 2022; Visceglia et al., 2012), Japanese - English (A. Lee & Xu, 2020); however, only a few studies have investigated the use of prosodic focus marking among bilingual children. For example, Liu et al. (2016) and Liu (2017) investigated the use of prosodic focus marking among Bai-Mandarin sequential bilingual children, showing that the Bai-Mandarin bilingual group had partially acquired acoustic cues to mark focus in their L2 Mandarin production. Specifically, Bai-Mandarin sequential bilinguals had longer duration for on-focus words than pre-focus and post-focus words, and wider f0 range for on-focus words than post-focus words. Since the Bai language only utilises duration (not f0) as an acoustic cue for marking focus, the use of duration can be attributed to positive L1 transfer, whereas the use of f0 range can only be explained by successful acquisition (Liu et al., 2016). In general, findings from previous studies indicate that prosodic focus marking can be acquired by L2 speakers; however, the age of acquisition seems to play a more important role than the prosodic similarity between languages.

Previous studies examined participants with various background i.e., early (simultaneous, sequential) and late bilinguals, it is evident that bilingual speakers exhibit different prosodic patterns in their L2 production. Depending on the age of acquisition and daily L2 input, L2 prosodic focus marking can be, at least partially, acquired. However, to the best of our knowledge, no study has investigated the production of prosodic focus marking in L2 by autistic children. Although atypical prosodic focus marking has been observed in the native language of autistic children, the questions arise pertaining to the use of prosodic focus marking in the L2 or L3 production by multilingual autistic children and whether their acquisition exhibits variations compared to their TD peers. We address these questions by investigating the production of English prosodic focus marking of CASD and CTD trilingual children, and examining the potential differences in this regard compared to native English-speaking children.

1.4 The current study

This is the first study to investigate the production of English prosodic focus marking by Cantonese-speaking autistic children and how they differ from their TD peers, i.e., CTD group, and ETD group. Based on findings from studies comparing focus marking between autistic and TD speech (section 1.2) among monolingual English children, the CASD group is likely to show some degree of OFE or /and PFC patterns that are similar to those of the ETD group. However, the specific acoustic cues that CASD groups employ for focus marking seems intricate and remains empirical. While past studies show that duration and intensity serve as the primary acoustic correlates of focus marking for L1 Cantonese L2 English adult speakers (section 1.3), this pattern may differ for autistic children.

Following previous studies mentioned in sections 1.1 to 1.3, we used the most widely studied acoustic correlates, i.e., mean duration, mean f0, f0 range, mean intensity, as well as f0 curves to investigate the prosodic focus marking patterns of three groups. To the best of the authors' knowledge, no studies have compared f0 curves between the autistic and TD population. However, in general acoustic phonetic research involving TD population, studies (e.g., Chen et al., 2023) have demonstrated that f0 curves potentially provide more nuanced insights into prosodic focus marking. The analysis of f0 curves may offer some new insights on the prosodic focus marking strategies used by the three groups. Specific research questions are (1) Do the Cantonese-speaking autistic children show OFE and PFC patterns in their English production and what acoustic cues do they employ, (2) how does the CASD group differ from the CTD group in terms of OFE and PFC patterns in their English production, (3) how do the trilingual Cantonese-speaking (both the autistic and TD group) children differ from the ETD group in terms of OFE and PFC patterns?

2. Methods

2.1 Participants

Forty-eight participants were recruited in the current study with 16 speakers for each of the three groups, i.e., groups of CASD (mean age: 9.6 years old), CTD (mean age: 9.7 years old) and ETD (mean age: 9.9 years old). Participants of the Cantonese-speaking groups were recruited from local Hong Kong primary schools and the Speech Therapy Unit (STU) at the Department of Chinese and Bilingual Studies at the Hong Kong Polytechnic University. The CASD group received formal diagnosis of ASD from Developmental-Behavioural Paediatric Centres or hospitals in Hong Kong. The English-speaking children who participated as the control group of native speakers were from South Carolina in the United States. The CASD group started learning English and Mandarin at around 3 and 4.4 years old, and the CTD group started at around 3.1 and 3.3 years old. Table 2 shows the mean and standard deviation of the age of acquisition and self-reported language proficiency scores (i.e., reading, writing, listening and speaking) of Mandarin and English. The language proficiency scores ranged from 1 to 5 with 5 indicating the highest proficiency level and 1 indicating the lowest. In addition, the non-verbal Intelligence Quotient and English proficiency of all participants were tested using *Raven's Progressive Matrices* (Raven, 1989) and *Comprehensive Assessment of Spoken Language* (CASL; Carrow-Woolfolk, 1999) respectively. Table 3 gives test scores, age and gender along with the Autism Spectrum Quotient (AQ, Children's Version; Auyeung et al., 2008) scores. Participants were matched in terms of linguistic background, age and gender; however, the IQ and language scores[1] were not matched as it might led to biased sample collection for the autistic group (Dennis et al., 2009).

Table 2. Mean and standard deviation of age of acquisition (year) and self-reported language proficiency scores (i.e., reading, writing, listening and speaking) of the CASD and CTD groups. The language proficiency scores range from 1 to 5 (5 indicates the highest proficiency and 1 indicates the lowest).

Table 3. Mean and standard deviation of age, IQ and CASL scores of participants.

2.2 Stimuli

Fifteen target sentences paired with 15 pictures were designed for the production test and each of them described an on-going action involving a subject, a verb and an object. A set of prompt questions were designed to elicit desired focus type (i.e., broad, narrow and contrastive focus) in subject, verb and object position from the participants resulting in 7 conditions (Table 4). Under narrow and contrastive

[1] A two-tailed t-test was conducted, revealing no significant difference in English reading ($p = .25$), writing ($p = .12$), listening ($p = .09$) and speaking ($p = .15$) scores between the CASD and CTD groups.

focus conditions, words before and after the on-focus words were categorised into pre- and post-focus respectively. Target sentences were grouped into five blocks and each block contained three target sentences. All the stimuli were presented randomly to each participant with randomised blocks. In total, 210 stimuli (15 target sentences * 7 conditions * 2 repetitions) were collected from each participant, and the experiment was programmed in E-prime 2.0 (Schneider et al., 2002).

Table 4. Examples of prompt questions and target sentences in relation to different focus conditions.

2.3 Procedure

Cantonese participants were invited to the Speech and Language Sciences Laboratory in the Hong Kong Polytechnic University for the production test in a sound-proof booth. Audio Technica AT2035 condenser microphone and Steinberg UR22mkII USB Audio Interface were used to record participants' speech production with a sampling rate of 44100 Hz in Audacity (Audacity Team, 2021). English participants carried out the production test in a quiet room, and the whole procedure was audio-recorded locally with the same settings in Audacity (Audacity Team, 2021). All participants were compensated for their participation, and signed consent forms in compliance with a protocol approved by the Human Subjects Ethics Sub-committee at The Hong Kong Polytechnic University.

There were practice and test sessions before the real experiment for each block. During the practice participants were instructed to familiarise themselves with all the questions and target sentences as well as character names and actions associated with the pictures. During the experiment, prompt questions were asked by professional speech therapists (ST) playing a game called 'Under the shape' (Chen, 2011) with the participants. In each trial, participants were required to answer questions based on pairs of pictures presented on a computer screen. For example, Figure 1 (left panel) was first given to a participant with ST asking "*Who is kissing the doll*?"; then, the complete picture (Figure 1, right panel) was shown to the participants, and the participant was expected to answer "***Amy*** is kissing the doll." with narrow focus on the initial subject "***Amy***". The same stimuli were used for Cantonese- and English-speaking children. All target sentences were arranged into five blocks and three sentences were tested in each block. All the stimuli were shown randomly to each participant following a randomized block order. See Appendix A for a full list of stimuli.

Figure 1. An example of a pair of pictures used in the production test. In a trial eliciting narrow focus, the picture on the left with the grey area blocking the missing part was first shown to a participant while the ST asked “Who is kissing the doll?”; then, the picture on the right removing the grey area was shown to the participant, and he or she was expected to answer “***Amy*** is kissing the doll”.

2.4 Segmentation and feature extraction

Speech elicitation tasks were recorded and target sentences were manually segmented into words and syllables in Praat (Boersma & Weenink, 2023) following the segmentation criteria in Turk et al. (2012). All the words were segmented apart from the auxiliary verb “*is*” and the article “*the*”. Acoustic measures, i.e., word duration, mean f0, f0 range and mean intensity, were extracted using a Praat script (ProsodyPro; Xu, 2013) where the f0 was extracted using the autocorrelation method (Boersma, 1993). In addition, the maximum and minimum f0 values of each target word were extracted for the calculation of f0 range. Furthermore, f0 measurements were taken from 20 time points over the voiced portion of the target words with equal intervals for curve analyses.

2.5 Statistical analyses

Linear Mixed-Effects Models (LMM) were used for word duration, mean f0, f0 range and mean intensity analyses, and functional data analysis (FDA; Ramsay et al., 2009) was used for f0 curve analyses. In LMM analyses, model building started using word duration, mean f0, f0 range or mean intensity as the response variables with two random effects (i.e., participants and words). The explanatory variables, i.e., groups (CASD group, CTD group, ETD group), focus condition (broad, narrow and contrastive pre-, on-, post-focus) as well as the interaction term between group and focus condition were added to the model step by step followed by Likelihood Ratio (LR) tests to evaluate the significance of each fixed effect. The optimal model was selected based on the lowest Akaike information criterion (Bates et al., 2015). A *post-hoc* comparison was conducted if any significant effect was found. The implementation of LMM fitting, and *post-hoc* comparison were carried out using `lmer4` (Bates et al., 2015) and `emmeans` (Lenth et al., 2023) in R (R Core Team, 2023).

FDA (Ramsay et al., 2009) was carried out to investigate the difference between f0 curves in relation to focus condition, and a similar procedure in Chen et al. (2017, 2023) was adapted to obtain specific regions where the two curves showed significant differences across the voiced portion of the target syllables. The f0 curves were fitted using the following equation,

$$y_i(t_j) = f_i(t_j) + \epsilon_{ij}$$

where $y_i(t_j)$ denotes the f0 values at t_j in the curve i, $i = 1, \dots, n$ and $j = 1, \dots, m$.; ϵ_{ij} denotes the error term following a normal distribution N (0, σ^2).

A model was fit for each pair of f0 curves in relation to groups (i.e., CASD, CTD, ETD) and focus conditions (i.e., broad, narrow and contrastive pre-, on- and post-focus). Then functional t-tests were carried out to test regions where two f0 curves show significant differences. For each group, the f0 curve under broad focus condition was compared to those under narrow and contrastive pre-, on- and post-focus conditions. For each comparison, 200 random sampling was conducted, and the observed t-statistic was compared with the maximum 0.05 critical value. Statistical significance was reached when the observed t-statistics were larger than the maximum critical value. We adjusted the significance level based on the Bonferroni correction to reduce the potential risk of significance inflation due to multiple comparisons (Chen et al., 2017; Xu & Xu, 2005), and the maximum critical value was calculated based on the corrected alpha value. The R package `fda` (verision 6.0.5) was used for the FDA of f0 curves (Ramsay et al., 2009, 2022).

3. Results

Figures 2 to 5 give the results of LMM analyses of mean duration, mean f0, f0 range and mean intensity respectively. The black dots are the predicted mean value under each focus conditions and the horizontal whiskers indicate the 95% confidence intervals. Red dashed lines represent the estimated mean value under broad focus condition.

3.1 Word duration

The optimal LMM model indicated that group alone was not a significant predictor for duration, but the inclusion of focus condition ($\chi^2 = 403.54$; df = 6; $p < .0001$) and the two-way interaction between group and focus condition ($\chi^2 = 474.91$; df = 12; $p < .0001$) significantly improved the model fitting. Overall, Figure 2 shows that the CASD group consistently has longer mean word duration than their TD peers across focus conditions. A *post-hoc* comparison showed that CASD group produced narrow on-focus ($p = .05$) and post-focus ($p = .01$) words as well as contrastive post-focus ($p = .001$) words significantly longer than those produced by CTD group.

Within each group, the CASD and CTD groups showed some patterns of PFC (i.e., mean duration of narrow and contrastive post-focus words is shorter than that of broad focus words; black dots on the far left of the red dashed line), while the ETD group yielded patterns of OFE (i.e., mean duration of on-focus words is longer than that of the broad focus words; black dots on the far right of the red dashed line). In terms of statistical significance, the CASD group had narrow post-focus words significantly shorter than broad focus words ($p < .0001$), and the CTD group produced both narrow and contrastive post-focus words significantly shorter than broad focus words ($ps < .0001$). For the ETD group, contrastive on-focus words had significantly longer duration than broad focus words ($p < .0001$).

Figure 2 Predicted mean duration (black dots) with 95% confidence interval across focus conditions of three groups.

Table 5 *Post-hoc* comparisons of mean word duration of three groups in relation to focus conditions. Narrow and contrastive on- and post- focus words are compared with the baseline (i.e., broad focus). Positive estimates indicate a higher value for broad focus condition.

3.2 Mean f0

For the mean f0, the optimal LMM showed that that the inclusion of group ($\chi^2 = 7.24$; df = 2; $p = .027$), focus condition ($\chi^2 = 6889.90$; df = 6; $p < .0001$), and the interaction between group and focus condition ($\chi^2 = 111.59$; df = 12; $p < .0001$) significantly improved the model fit. Figure 3 shows that the CASD group has higher mean f0 than their TD peers across matching focus conditions; however, a significant difference was only reached between the CASD and ETD groups ($p < .05$). For OFE and PFC patterns, all groups showed some degree of PFC where the narrow post and contrastive post words had significantly lower mean f0 than that of the broad focus words (i.e., black dots of post-focus condition are on the left of the red dashed lines). However, only the ETD group showed some degree of OFE by producing significant and marginally significant higher mean f0 for narrow and contrastive on-focus words (i.e., black dots are on the right of the red dashed line) than those of broad focus words (Table 6).

Figure 3. Predicted mean f0 (black dots) with 95% confidence interval across focus conditions of three groups.

Table 6 *Post-hoc* comparisons of mean f0 of three groups in relation to focus conditions. Narrow and contrastive on- and post- focus words are compared with the baseline (i.e., broad focus). Positive estimates indicate a higher value for broad focus condition.

3.3 F0 range

For f0 range, the optimal LMM indicated that group alone did not significantly improve the model fit, but the inclusion of focus condition ($\chi^2 = 78.52$; df = 6; $p < .0001$) as well as the interaction between group and focus condition ($\chi^2 = 192.16$; df = 12; $p < .0001$) significantly improved the model fit. Figure 4 indicates that the CASD group generally has wider f0 range than the CTD and ETD groups across matching focus conditions. However, significant differences were only observed in post-focus locations, but not in on-focus locations. Specifically, the CASD group exhibited significantly wider f0 range than the CTD group for words under narrow post-focus ($p = .04$) and contrastive post-focus ($p = .04$) conditions.

Within each group, both the CASD and ETD groups produced on-focus words with wider f0 range than broad focus words (i.e., patterns similar to OFE; black dots are on the right of the red dashed lines); however, those differences did not reach statistical significance. Surprisingly, the ETD group exhibited a wider f0 range for post-focus words (i.e., black dots are on the right of the red dashed line), particularly for contrastive post-focus words where the range was significantly wider than broad focus words. The

CTD group did not have patterns that are similar to OFE. However, the CTD group showed patterns of PFC where the narrow and contrastive post-focus words had significantly narrower f0 range than those of broad focus words (*ps* $< .0001$, Table 7).

Figure 4. Predicted mean f0 range (black dots) with 95% confidence interval across focus conditions of three groups.

Table 7. *Post-hoc* comparisons of f0 range of three groups in relation to focus conditions. Narrow and contrastive on- and post- focus words are compared with the baseline (i.e., broad focus). Positive estimates indicate a higher value for broad focus condition.

3.4 Intensity

For mean intensity, the optimal LMM model indicated that the inclusion of group ($\chi^2 = 8.51$; df = 2; $p = .014$), focus condition ($\chi^2 = 328.29$; df = 6; $p < .0001$) and the interaction between group and focus condition ($\chi^2 = 239.20$; df = 12; $p < .0001$) significantly improved the model fit. Figure 5 suggests that the CASD group generally has lower mean intensity than their TD peers across matching focus conditions; however, a significant difference was only observed between the CASD group and the ETD group (*ps* $< .05$), while there was no significant difference between the CASD group and the CTD group.

Within each group, the CASD children had some patterns of OFE (i.e., black dots of on-focus words are on the right of the broad focus words) where they produced contrastive on-focus words with significantly higher mean intensity than the broad focus words ($p = .024$). Meanwhile, the CTD group exhibited similar patterns to PFC where narrow ($p < .0001$) and contrastive ($p < .001$) post-focus words had significantly lower mean intensity than the broad focus words (i.e., black dots of post-focus words are on the left of the red dashed line). For the ETD group, both OFE and PFC patterns were observed. Specifically, the ETD children produced contrastive on-focus words with significantly higher mean intensity than broad focus words, and narrow and contrastive post-focus words with significantly lower mean intensity (Figure 5 leftmost panel; Table. 8).

Figure 5. Predicted mean intensity (black dots) with 95% confidence interval across focus conditions of three groups.

Table 8. *Post-hoc* comparisons of mean intensity of three groups in relation to focus conditions. Narrow and contrastive on- and post- focus words are compared with the baseline (i.e., broad focus). Positive estimates indicate a higher value for broad focus condition.

3.5 F0 curves

Figure 6 shows the mean smoothed f0 curves of each group in relation to focus conditions. The grey solid lines indicate the f0 curves of words under broad focus and the coloured dash lines represent those under narrow and contrastive on- and post-focus conditions. Figure 6 shows that the CASD group had consistently higher f0 values than their TD peers across the whole syllable with matching focus conditions. This is similar to the pattern in mean f0 (Figure 3). Within each group, there is not much difference in f0 curves between narrow and contrastive focus conditions. In terms of OFE and PFC patterns, a visual inspection indicates that the f0 curves under narrow and contrastive post-focus conditions are much lower than those under broad focus conditions across the whole syllable for all three groups, suggesting some degree of PFC. However, it seems that there is not much OFE patterns for the three groups, i.e., the difference in f0 curves between on-focus and broad focus conditions is not as substantial as that between post-focus and broad focus conditions.

Figure 6. Smoothed f0 curve plots in relation to groups and focus conditions. The strip texts on top indicate three groups and different colours account for focus conditions.

Figure 7 presents the results of FDA analyses, and the orange colour denotes the regions where two f0 curves reached significant differences. The x-axis indicates the percentage throughout the vocalic portion of the target words and y-axis gives the focus condition that is being compared with the broad focus. Results from FDA analyses reassert the PFC patterns in Figure 6, showing that the f0 curves under narrow and contrastive post-focus conditions are significantly different than those under broad focus words across the whole syllable for all three groups. Meanwhile, only the ETD group showed partially OFE in on-focus *vs.* broad focus comparisons, i.e., the f0 curves of words under contrastive on-focus condition and broad focus conditions were significantly different from 26% to 48% of the syllable (Figure 7).

Figure 7. Significance regions based on results from functional t-tests when broad focus is compared to narrow and contrastive on-focus and post-focus words. Orange fill denotes regions where a significant difference is reached (see Appendix B for statistics).

4. Discussion

Recall that the current study aims to explore three research questions. Regarding our first research question, whether the CASD group shows OFE and PFC patterns in their English productions and what acoustic cues they employ, our results show that OFE and PFC are realised via different acoustic cues among the CASD group. Specifically, children in the CASD group showed PFC patterns by reducing mean duration for words under narrow post-focus condition and lowering mean f0 and f0 curves for words under both narrow and contrastive post-focus conditions. Meanwhile, children in the CASD group also exhibited some patterns of OFE by increasing mean intensity for words under contrastive on-focus condition. Since duration and intensity are the main acoustic cues of prosodic focus marking in Cantonese (Wu & Xu, 2010), the use of duration and intensity for marking OFE and PFC seems to be a positive L1 transfer. Regarding the use of f0 as an indicator for OFE in Cantonese, Wu and Xu (2010) and Fung and Mok (2014) both reported the inconsistent use of mean f0 and f0 range to mark OFE, i.e., on-focus words had significantly higher mean f0 than broad focus words depending on the lexical tone and location of the on-focus words, e.g., T3 at sentence medial location, T4 at sentence final location. As for PFC, both studies reported that f0 was not an acoustic cue to mark PFC suggesting that PFC does not exist in Cantonese. Given these inconsistent patterns reported from previous studies, our findings could lead to a different interpretation, i.e., the use of mean f0 as an indicator of PFC among the CASD group could be the successful acquisition of a new feature in prosodic focus marking in their English production.

For the second research question, i.e., how does the CASD group differ from the CTD group in terms of OFE and PFC patterns in their English production, our results reveal that the CTD group exhibited a slightly more completed profile in utilising PFC. Specifically, the CTD group showed PFC patterns by reducing mean duration, narrowing f0 range, and lowering mean f0, f0 curves and mean intensity for words under both narrow and contrastive post-focus conditions, while the CASD group only shortened mean duration for words under narrow post-focus conditions and had no significant difference in f0 range across focus conditions. However, the CTD group did not show any patterns of OFE. Although there are some discrepancies in the production of English prosodic focus marking between the CASD and CTD groups, they both utilised word duration and mean f0 to mark PFC. This is partially similar to findings among monolingual English autistic and TD children (Nadig & Shaw, 2015; Paul et al., 2008), where both the autistic and TD children produced significantly longer duration for on-focus words than the post-focus words. Meanwhile, the CASD and CTD groups seem to share similar above-mentioned acquisition patterns, i.e., positive L1 transfer in the use of duration and intensity and potentially successful acquisition in the use of f0 for prosodic focus marking. The acquisition in using f0 to mark focus also supports the findings in Bai-Mandarin bilingual children (Liu et al., 2016) that the use of acoustic cues to mark focus in L2 can be, at least partially, acquired. Although the acoustic cues used for OFE and PFC among CASD and CTD children differ from L1 Cantonese L2 English adult speakers (Fung & Mok, 2014; Wu & Chung, 2011), the discrepancies are likely due to the different age of acquisition of the participants, sample size, methods used for speech elicitation as well as variability in individuals with autism.

With regard to the third research question, i.e., how do the CASD and CTD children differ from the native ETD children in terms of OFE and PFC patterns. Our findings show that the ETD children mostly differ from the CASD and CTD children in the use of OFE. Specifically, the ETD group has marked OFE by increasing mean duration, mean f0, mean intensity and f0 curve for words under on-focus conditions, whereas the CTD group did not show much of OFE patterns in terms of these acoustic cues and the CASD group only used intensity to mark OFE under contrastive on-focus condition. Among the ETD children, mean f0 was the only acoustic cue for signalling OFE under both contrastive and narrow on-focus conditions. For other acoustic cues, i.e., duration and intensity, OFE was only realised under contrastive on-focus conditions. Further, the f0 curve under contrastive on-focus conditions was significantly different from the one under broad focus condition from 26% to 48% of the syllable. This partial competence of utilising prosodic focus marking among the ETD children suggest some degree of immaturity in language processing, e.g., children start to exhibit brain responses (i.e., Focus Positive Shift) to narrow and contrastive focus that are similar to adults at twelve years old (Pannekamp et al., 2011).

It is worth noting that the Cantonese-speaking (i.e., CASD and CTD) and English-speaking (i.e., ETD) children mainly differ in the use of OFE, not PFC, which indicates that Cantonese-speaking children might acquire PFC prior to OFE in their L2 English production. This is especially true for patterns observed in f0 curves, i.e., all groups showed patterns of PFC while only the ETD children showed limited degree of OFE for words under contrastive on-focus conditions. This pattern regarding the acquisition of L2 English prosodic focus marking is different from findings reported in previous studies (e.g., Liu et al., 2019), where OFE is acquired prior to PFC. The observed differences in our study compared to Liu et al. (2019) could be attributed to the difference in age of L2 acquisition between our study (3 to 4 years old) and Liu et al (not reported). However, future research should explore how differing L2 acquisition ages influence the sequence of acquiring focus marking patterns, e.g., OFE and PFC.

It was noticed that all three groups of children showed asymmetric patterns in realising OFE and PFC in relation to narrow and contrastive conditions, i.e., OFE and PFC patterns were realised under both narrow and contrastive conditions for some acoustic cues, but not for the others[2]. This asymmetric pattern is likely due to the fact that the prosodic profile of these children is still under development. Previous studies (e.g., Hübscher et al., 2019a, 2020; Armstrong et al., 2018) reported that the use of prosody relating to various aspects (e.g., pragmatic) of verbal communication undergoes development during preschool years (3 to 5 years old), and the skills in using prosodic focus marking are language dependent (Chen, 2018). For example, Dutch-speaking children show adult-like focus marking patterns

[2] The raw acoustic differences between narrow and contrastive focus are subtle across all focus types (i.e., pre-focus, on-focus, post-focus) and three groups (i.e., CASD, CTD, ETD), and the only significant difference was observed in mean duration between narrow on-focus and contrastive on-focus conditions for the ETD group. Detailed statistics are given in Appendix C.

at around 10 to 11 years old (Romoren, 2016), while Korean-speaking children have adult-like focus marking patterns at around 4 to 5 years old (Yang, 2017).

Overall, our findings suggest that the CTD children utilised more acoustic cues than the CASD group when it comes to PFC, while the ETD group differed from both the CASD and CTD groups in the use of OFE. However, this does not suggest the degree of prosody maturity of each group given the multifaceted nature in the acquisition of prosody. When faced with perceiving information in a sentence, there are multiple layers of decoding involved. This encompasses not only the segmentation of the sound stream and its association with meaningful linguistic units within the sentence context, but also the interpretation of the relative pitch height across the sentence (also shown in previous behavioural and neural studies, e.g., Lee & Nusbaum, 1993; Tong et al., 2014; Wang et al., 2015). Taking one of the stimuli as an example, consider the response, 'John is flying the plane', to the questions 'Who is flying the plane?' (Q1) and 'What do you see in the picture?' (Q2). For both questions, children need to determine how to utilise different acoustic cues (e.g., duration, f0) for words under different focus conditions (e.g., on-focus words *vs.* post-focus words), and misaligning the acoustic cues with the on-focus and post-focus words is likely to lead to miscommunication. The current study mainly focused on the production of speech prosody, and future research is needed to incorporate production and perception within the context of language and prosody acquisition, encompassing both multilingual autistic and TD populations.

5. Conclusion

The current study investigated the production of English prosodic focus marking patterns (i.e., OFE and PFC) among trilingual CASD children, and compared their production with trilingual CTD children and native ETD children. This is the first study that attempts to reveal prosodic focus marking patterns among trilingual autistic children. By comparing the production between the CASD and CTD groups, our findings show similarities with previous studies using monolingual autistic English children, i.e., the autistic and TD children both used mean duration and mean f0 for marking focus. However, our findings also show some discrepancies from previous studies in terms of the acquisition order of prosodic focus marking. Previous studies showed that PFC is acquired after OFE, whereas the opposite order is found in our study. Apart from using static acoustic correlates (i.e., mean duration, f0, f0 range and intensity), we adapted a novel method using FDA comparing the f0 curves in relation to focus conditions, and the results further support our findings of the acquisition of prosodic focus marking, i.e., all groups showed PFC patterns under both contrastive and narrow post-focus conditions, while only the ETD group exhibited a limited degree of OFE for words under contrastive on-focus conditions.

Acknowledgements

This work was supported by Department of Chinese and Bilingual Studies at Hong Kong Polytechnic University [1-ZVRT; 1-ZE0D; 1-W08C], and partly supported by the SCOLAR, Education Bureau, HKSAR [K-ZB2P] and RGC [A-PB1B].

Conflict of Interest

There are no conflicts of interest.

Data Availability Statement

The datasets analysed during the current study are available from the first author on request.

References

American Psychiatric Association. (2013). *Diagnostic and Statistical Manual of Mental Disorders (DSM-5) 5th ed*.

Arnhold, A., & Kyröläinen, A.-J. (2017). Modelling the interplay of multiple cues in prosodic focus marking. *Laboratory Phonology*, *8*(1), Article 1. https://doi.org/10.5334/labphon.78

Armstrong, M., Esteve Gibert, N., Hübscher, I., Igualada, A., & Prieto, P. (2018). Developmental and cognitive aspects of children’s disbelief comprehension through intonation and facial gesture. *First Language*, *38*(6), 596–616. https://doi.org/10.1177/0142723718789278

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

Audacity Team. (2021). *Audacity* [Computer software]. https://audacityteam.org/

Auyeung, B., Baron-Cohen, S., Wheelwright, S., & Allison, C. (2008). The Autism Spectrum Quotient: Children's Version (AQ-Child). *Journal of Autism and Developmental Disorders*, *38*(7), 1230–1240. https://doi.org/10.1007/s10803-007-0504-z

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

Boersma, P. (1993). Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound. *Proceedings of the Institute of Phonetic Sciences*, *17*, 97–110.

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

Bolinger, D. (1972). Accent Is Predictable (If You're a Mind-Reader). *Language*, *48*(3), 633–644. https://doi.org/10.2307/412039

Bone, D., Bishop, S., Gupta, R., Lee, S., & Narayanan, S. S. (2016). Acoustic-Prosodic and Turn-Taking Features in Interactions with Children with Neurodevelopmental Disorders. *Interspeech 2016*, 1185–1189. https://doi.org/10.21437/Interspeech.2016-1073

Breen, M., Fedorenko, E., Wagner, M., & Gibson, E. (2010). Acoustic correlates of information structure. *Language and Cognitive Processes*, *25*(7–9), 1044–1098. https://doi.org/10.1080/01690965.2010.504378

Carrow-Woolfolk, E. (1999). Comprehensive assessment of spoken language (CASL). *American Guidance Service*.

Chen, A. (2009). The phonetics of sentence-initial topic and focus in adult and child Dutch. In M. Vigário, S. Frota, & M. J. Freitas (Eds.), *Phonetics and Phonology: Interactions* *and interrelations* (pp. 91–106). John Benjamins Publishing Company. https://doi.org/10.1075/cilt.306.05che

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, A. (2018). Get the focus right across languages: Acquisition of prosodic focus-marking in production. *The Development of Prosody in First Language Acquisition*, 295–314.

Chen, S., Hong, Y., Li, B., & Chun, E. (2023). The f0 perturbation effects in focus marking: Evidence from Korean and Japanese. *PLOS ONE*, *18*(3), e0283139. https://doi.org/10.1371/journal.pone.0283139

Chen, S., Zhang, C., McCollum, A. G., & Wayland, R. (2017). Statistical modelling of phonetic and phonologised perturbation effects in tonal and non-tonal languages. *Speech Communication*, *88*, 17–38. https://doi.org/10.1016/j.specom.2017.01.006

Chen, Y., Guion-Anderson, S., & Xu, Y. (2012). Post-Focus Compression in Second Language Mandarin. *Speech Prosody 2012*, 410–413.

Cooper, W. E., Eady, S. J., & Mueller, P. R. (1985). Acoustical aspects of contrastive stress in question-answer contexts. *The Journal of the Acoustical Society of America*, *77*(6), 2142–2156. https://doi.org/10.1121/1.392372

Cruttenden, A. (1997). *Intonation*. Cambridge University Press.

Dennis, M., Francis, D. J., Cirino, P. T., Schachar, R., Barnes, M. A., & Fletcher, J. M. (2009). Why IQ is not a covariate in cognitive studies of neurodevelopmental disorders. *Journal of the International Neuropsychological Society*, *15*(3), 331–343. https://doi.org/10.1017/S1355617709090481

DePape, A.-M., Chen, A., Hall, G., & Trainor, L. (2012). Use of Prosody and Information Structure in High Functioning Adults with Autism in Relation to Language Ability.

Frontiers in Psychology, *3*. https://www.frontiersin.org/articles/10.3389/fpsyg.2012.00072

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

Feldstein, S., Konstantareas, M., Oxman, J., & Webster, C. D. (1982). The chronography of interactions with autistic speakers: An initial report. *Journal of Communication Disorders*, *15*(6), 451–460. https://doi.org/10.1016/0021-9924(82)90018-1

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

Frith, U. (1989). *Autism: Explaining the Enigma.* (1st ed.). Blackwell Publishing.

Frith, U. (2008). *Autism: A Very Short Introduction*. OUP Oxford.

Fry, D. B. (1955). Duration and Intensity as Physical Correlates of Linguistic Stress. *The Journal of the Acoustical Society of America*, *27*(4), 765–768. https://doi.org/10.1121/1.1908022

Fung, H. S. H., & Mok, P. P. K. (2014). Realization of Narrow Focus in Hong Kong English declaratives—A Pilot Study. *Speech Prosody 2014*, 964–968. https://doi.org/10.21437/SpeechProsody.2014-182

Fusaroli, R., Lambrechts, A., Bang, D., Bowler, D. M., & Gaigg, S. B. (2017). “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., Lee, A. K. L., Yuen, H. K., Liu, F. and Yip, V. (2023) Bilingual exposure might enhance L1 development in Cantonese-English bilingual autistic children: evidence from the production of focus. *Autism*.

Guo, X. (2022). Acoustic Correlates of English Lexical Stress Produced by Chinese Dialect Speakers Compared to Native English Speakers. *Frontiers in Psychology*, *13*. https://www.frontiersin.org/articles/10.3389/fpsyg.2022.796252

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

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

Hübscher, I., Garufi, M., & Prieto, P. (2019). The development of polite stance in preschoolers: How prosody, gesture, and body cues pave the way. *Journal of Child Language*, *46*(5), 825–862. https://doi.org/10.1017/S0305000919000126

Hübscher, I., Wagner, L., & Prieto, P. (2020). Three-year-olds infer polite stance from intonation and facial cues. *Journal of Politeness Research*, *16*(1), 85–110. https://doi.org/10.1515/pr-2017-0047

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

Lambrecht, K. (1994). *Information structure and sentence form: Topics, focus, and the mental representations of discourse referents* (pp. xvi, 388). Cambridge University Press. https://doi.org/10.1017/CBO9780511620607

Lambrecht, K. (2000). When subjects behave like objects: An analysis of the merging of S and O in Sentence-Focus Constructions across languages. *Studies in Language. International Journal Sponsored by the Foundation "Foundations of Language"*, *24*(3), 611–682. https://doi.org/10.1075/sl.24.3.06lam

Lee, A., & Xu, Y. (2020). Focus prosody in Japanese-English early bilinguals: A pilot study. *Speech Prosody 2020*, 843–847. https://doi.org/10.21437/SpeechProsody.2020-172

Lee, L., & Nusbaum, H. C. (1993). Processing interactions between segmental and suprasegmental information in native speakers of English and Mandarin Chinese. *Perception & Psychophysics*, *53*(2), 157–165. https://doi.org/10.3758/BF03211726

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

Lieberman, P. (1960). Some acoustic correlates of word stress in American English. *Journal of the Acoustical Society of America*, *32*, 451–454. https://doi.org/10.1121/1.1908095

Liu, J., Xu, Y., & Lee, Y. (2019). Post-focus compression is not automatically transferred from Korean to L2 English*. *Phonetics and Speech Sciences*, *11*(2), 15–21. https://doi.org/10.13064/KSSS.2019.11.2.015

Liu, Z., Chen, A., & Van De Velde, H. (2016). Prosodic focus marking in Bai-Mandarin sequential bilinguals' Mandarin. *Speech Prosody 2016*, 951–955. https://doi.org/10.21437/SpeechProsody.2016-195

Liu, Z. H. (2017). *The Development of Prosodic Focus-marking in Early Bilinguals' L2: A Study of Bai-Mandarin Early Bilinguals' Mandarin* [Doctoral thesis]. Utrecht University.

Lord, C., Elsabbagh, M., Baird, G., & Veenstra-Vanderweele, J. (2018). Autism spectrum disorder. *The Lancet*, *392*(10146), 508–520. https://doi.org/10.1016/S0140-6736(18)31129-2

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

Mücke, D., & Grice, M. (2014). The effect of focus marking on supralaryngeal articulation – Is it mediated by accentuation? *Journal of Phonetics*, *44*, 47–61. https://doi.org/10.1016/j.wocn.2014.02.003

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

Pannekamp, A., Van Der Meer, E., & Toepel, U. (2011). Context- and Prosody-Driven ERP Markers for Dialog Focus Perception in Children. *Brain Topography, 24*(3–4), 229–242. https://doi.org/10.1007/s10548-011-0194-x

Paul, R., Bianchi, N., Augustyn, A., Klin, A., & Volkmar, F. (2008). Production of Syllable Stress in Speakers with Autism Spectrum Disorders. *Research in Autism Spectrum Disorders*, *2*(1), 110–124. https://doi.org/10.1016/j.rasd.2007.04.001

Peppé, S., McCann, J., Gibbon, F., O'Hare, A., & Rutherford, M. (2007). Receptive and Expressive Prosodic Ability in Children With High-Functioning Autism. *Journal of Speech, Language, and Hearing Research*, *50*(4), 1015–1028. https://doi.org/10.1044/1092-4388(2007/071)

Population By-census (2016). Population by language. https://www.bycensus2016.gov.hk

R Core Team. (2023). *R: A language and environment for statistical computing* [Computer software]. https://www.R-project.org/

Ratto, A.B., Potvin, D., Pallathra, A.A., Saldana, L. and Kenworthy, L., 2020. Parents report fewer executive functioning problems and repetitive behaviors in young duallanguage speakers with autism. *Child Neuropsychology*, *26*(7), pp.917-933.

Ramsay, J. O., Graves, S., & Hooker, G. (2022). *fda: Functional Data Analysis* (6.0.5) [Computer software]. https://CRAN.R-project.org/package=fda

Ramsay, J. O., Hooker, G., & Graves, S. (2009). *Functional data analysis with R and MATLAB*. Springer.

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

Romøren, A. S. H. (2016). *Hunting highs and lows: The acquisition of prosodic focus marking in Swedish and Dutch* [Unpublished PhD dissertation, Utrecht University.]. https://dspace.library.uu.nl/handle/1874/338040

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

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

Thurber, C., & Tager-Flusberg, H. (1993). Pauses in the narratives produced by autistic, mentally retarded, and normal children as an index of cognitive demand. *Journal of Autism and Developmental Disorders*, *23*(2), 309–322. https://doi.org/10.1007/BF01046222

Thye, M. D., Bednarz, H. M., Herringshaw, A. J., Sartin, E. B., & Kana, R. K. (2018). The impact of atypical sensory processing on social impairments in autism spectrum disorder. *Developmental cognitive neuroscience*, *29*, 151-167.

Tong, X., McBride, C., Lee, C.-Y., Zhang, J., Shuai, L., Maurer, U., & Chung, K. K. H. (2014). Segmental and suprasegmental features in speech perception in Cantonese-speaking second graders: An ERP study. *Psychophysiology*, *51*(11), 1158–1168. https://doi.org/10.1111/psyp.12257

Truckenbrodt, H. (1995). *Phonological Phrases: Their Relation to Syntax, Focus, and Prominence* [Doctoral thesis, Universität Tübingen]. chrome-extension://efaidnbmnnnibpcajpcglclefindmkaj/http://www.ai.mit.edu/projects/dm/theses/truckenbrodt95.pdf.

Turk, A., Nakai, S., & Sugahara, M. (2012). Acoustic Segment Durations in Prosodic Research: A Practical Guide. In *Acoustic Segment Durations in Prosodic Research: A Practical Guide* (pp. 1–28). De Gruyter. https://doi.org/10.1515/9783110914641.1

Visceglia, T., Su, C., & Tseng, C. (2012). Comparison of English narrow focus production by L1 English, Beijing and Taiwan Mandarin speakers. *2012 International Conference on Speech Database and Assessments*, 47–51. https://doi.org/10.1109/ICSDA.2012.6422454

Wang, M., Li, C., & Lin, C. Y. (2015). The Contributions of Segmental and Suprasegmental Information in Reading Chinese Characters Aloud. *PLoS ONE*, *10*(11), e0142060. https://doi.org/10.1371/journal.pone.0142060

Wonnacott, E., & Watson, D. G. (2008). Acoustic emphasis in four year olds. *Cognition*, *107*(3), 1093–1101. https://doi.org/10.1016/j.cognition.2007.10.005

Wu, W., & Chung, L. (2011). Post-focus compression in English-Cantonese bilingual speakers. *In 17th International Congress of Phonetic Sciences, ICPhS 2011*, 148–151.

Wu, W., & Xu, Y. (2010). Prosodic Focus in Hong Kong Cantonese without Post-focus Compression. *Proceeding of Speech Prosody*. Speech Prosody, Chicago.

https://www.semanticscholar.org/paper/Prosodic-Focus-in-Hong-Kong-Cantonese-without-Wu-Xu/25fe9e429e3ead0dd6e101144dca80ece1cb5890

Xu, Y. (2013). ProsodyPro—A Tool for Large-scale Systematic Prosody Analysis. *Tools and Resources for the Analysis of Speech Prosody*, 7–10. https://www.semanticscholar.org/paper/ProsodyPro-%E2%80%94-A-Tool-for-Large-scale-Systematic-Xu/453480b67a15cac4ec47b52c3a2eea241c18db64

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

Yang, A. (2017). *The acquisition of prosodic focus-marking in Mandarin Chinese- and Seoul Korean-speaking children proefschrift*. LOT.

Yang, A., & Chen, A. (2014). Prosodic Focus Marking in Child and Adult Mandarin Chinese. *14th International Symposium on Tonal Aspects of Language*, 54–58.

Tables and Figures

Figure 1_1, Figure 1_2

Figure 2.

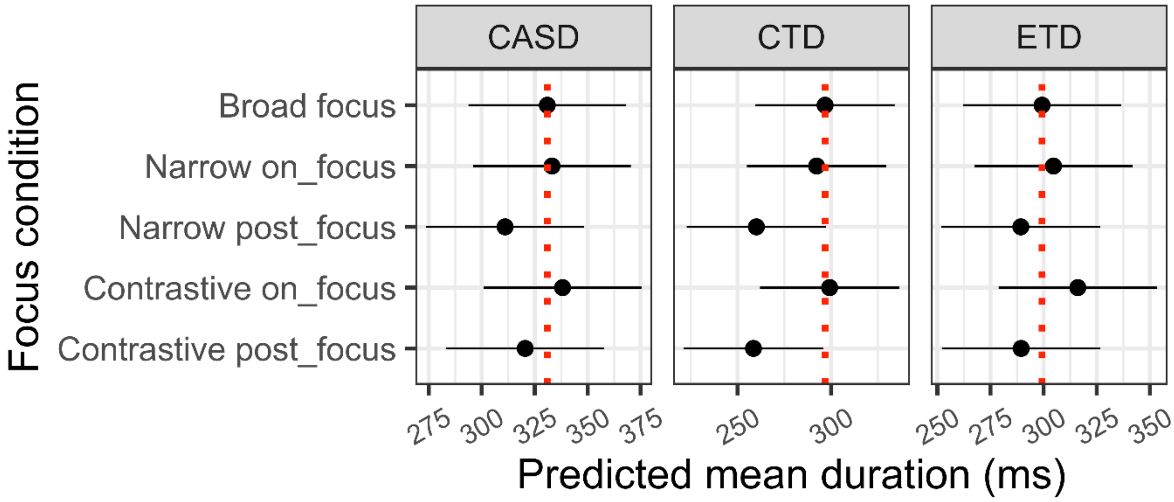

Figure 3.

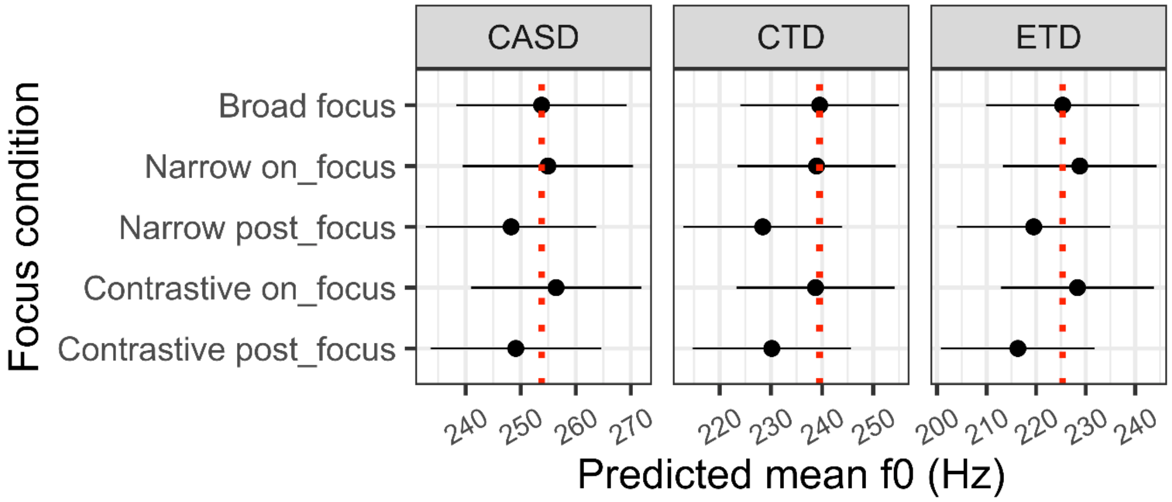

Figure 4.

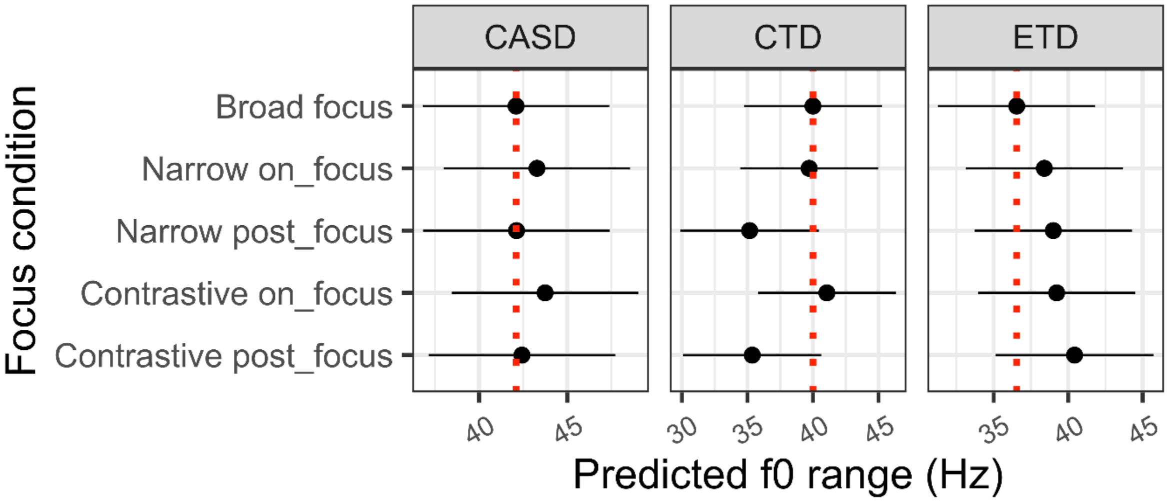

Figure 5.

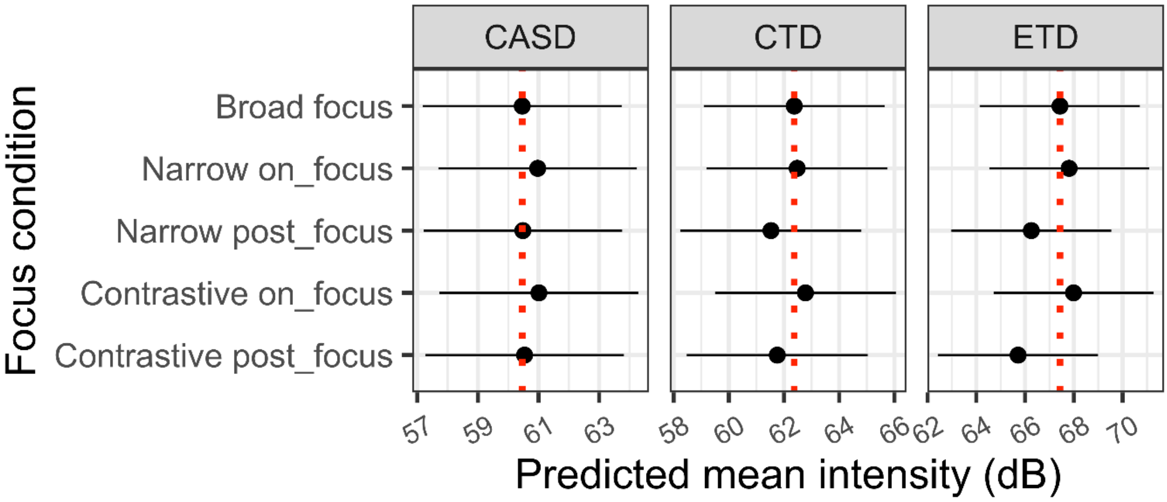

Figure 6.

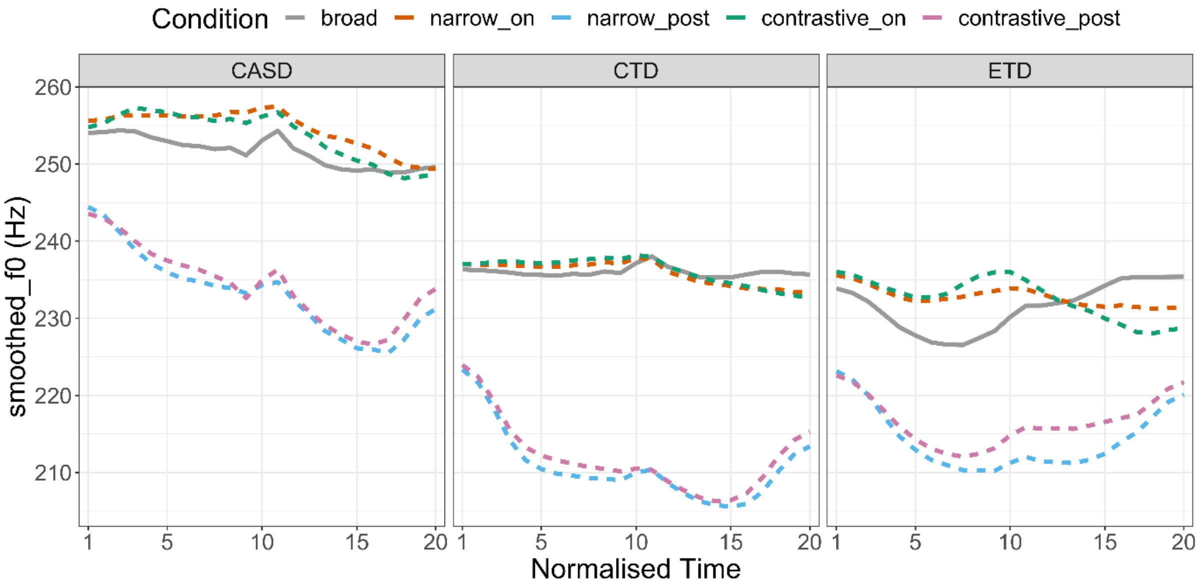

Figure 7.