Production of Mandarin consonant aspiration and monophthong in children with Autism Spectrum Disorder

Yan FENG[a,b#], Fei CHEN[c#], Junzhou MA[d], Lan WANG[e], and Gang PENG[b, e*]

[a]*School of Foreign Studies, Nanjing University of Science and Technology, Nanjing, China*

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

[c]*School of Foreign Languages, Hunan University, Hunan, China*

[d]*School of Foreign Languages, Taizhou University, Zhejiang, China*

[e]*Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Guangdong, China*

[#]The first two authors contributed equally to this study.

[*]Corresponding author:

E-mail: gpengjack@gmail.com

Tel: (+852) 3400 8462; Fax: (+852) 2334 0185

Address: *Department of Chinese and Bilingual Studies, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong SAR, China*

Production of Mandarin consonant aspiration and monophthong in children with Autism Spectrum Disorder

Abstract

Impaired speech sound production adds difficulties to social communication in children with Autism Spectrum Disorder (ASD), while a limited attempt has been made to figure out the speech sound production among Mandarin-speaking children with ASD. The current study conducted both auditory-perceptual scoring and quantitative acoustic analysis of speech sound imitated by 27 Mandarin-speaking children with ASD (3.33-7.00 years) and 30 chronological-age-matched typically developing (TD) children. Auditory-perceptual scoring showed significantly lower scores for aspirated/unaspirated consonants and monophthongs in children with ASD. Moreover, the correlation between the developmental age of language and production accuracy in children with ASD emphasised the importance of language assessment. The quantitative acoustic analysis further indicated that the ASD group produced a much shorter voice onset time for aspirated consonants and showed a reduced vowel space than the TD group. Early interventions focusing on these production patterns should be introduced to improve the speech sound production in Mandarin-speaking children with ASD.

Keywords: Autism Spectrum Disorder; Mandarin-speaking children; Impaired aspiration; Reduced vowel space.

Introduction

Autism Spectrum Disorder (ASD) is a life-long developmental disorder affecting social communication and interaction (American Psychiatric Association, 2013). Among children with ASD, some fail to develop any functional language capabilities. Even those who are relatively verbal almost always have language disorders (Klinger et al., 2014). Previous literature has well-documented that children with ASD are impaired in pragmatic, syntactic, and lexical development because of social-cognitive impairment (Cheung et al., 2020; Kjelgaard & Tager-Flusberg, 2001; Naigles, 2016; Tager-Flusberg et al., 2005), while there are some controversial findings on their impairment of speech sound production. Empirical studies have indicated a strong correlation between the difficulty of speech sound production and the severity of language impairment in individuals with ASD (Wolk & Brennan, 2013). Defective speech sound production has a detrimental impact on daily communication. Thus, there is an urgent need to improve our understanding of speech sound production in young children with ASD to provide a potential reference for clinical diagnosis and early intervention.

Mandarin speech sound development in typically developing children

There are around 1.4 billion people speaking Chinese in the world, including people from China, Singapore, Malaysia, etc. In China, the prevalence of children suffering from ASD has been as high as 0.7% in recent years (Zhou et al., 2020). There are many dialects differing in phonology in China. Among these, Mandarin (standard

Chinese) is a mainstream language with reference to Northern China dialects, which children need to learn and speak in kindergartens and schools in China (Li & To, 2017). Moreover, Mandarin is well-known for its variety of voiceless consonants with subtle phonemic differences in the place and manner of articulation, such as retroflex versus alveolar and aspirated versus unaspirated consonants. These were reported to be challenging for native young children and also for non-native speakers (Wang & Shangguan, 2004; Xie, 2009). The unique features of Mandarin phonology and articulation offer us a valuable opportunity to explore the generalisability of the findings about speech sound production from English-speaking individuals with ASD in a cross-linguistic framework.

To better explore the impairment of speech sound production in children with ASD, we firstly need to understand the development of speech sound production in typically developing (TD) children. Li and To (2017), and Peng and Chen (2020) reviewed the speech sound development in Mandarin-speaking children, and summarised the general acquisition order of Mandarin tones, vowels, and consonants in TD children based on previous research (e.g., Chen & Kent, 2010; Li & Thompson, 1977; Shi & Wen, 2007; Xie, 2009; Zhu & Dodd, 2000; Zhu, 2002). Zhu and Dodd (2000) is one of the early studies that comprehensively explored Mandarin speech sound acquisition in 129 native monolingual TD children (1;6-4;6) via picture naming and description tasks. They found that Mandarin lexical tones were the earliest acquired, followed by the acquisition of vowels and syllable-final consonants. The acquisition of syllable-initial consonant system was the latest completed.

Besides the general developmental order, previous studies have reported the age of phonological stabilisation for specific phonemes in TD children. Researchers have found that TD children acquire [t] before the age of two, [p] and [tʰ] before the age of three, [k], [kʰ] and [pʰ] at the age of three, [tɕ] and [tɕʰ] at the age of four, and [ts] and [tsʰ] at the age of five (Li & To, 2017; Si, 2006; Zhu, 2002; Zhu & Dodd, 2000). Li and To (2017) reported that the two retroflex affricates [tʂ] and [tʂʰ] are more challenging, and some TD children acquire these two sounds even after six or seven years old. With a larger sample size, Xie (2009) explored the developmental order of Mandarin syllable-initial consonants in 149 TD children (2;4-6;0) via picture naming and description tasks, and found that Mandarin-speaking TD children acquired unaspirated consonants earlier than the corresponding aspirated consonants. Such finding indicates that children may need more time to acquire the production of aspirated consonants. In terms of monophthong acquisition, Shi and Wen (2007) investigated the developmental order of Mandarin monophthongs among 40 TD children (1;0-6;0) via naming tasks and imitation tasks. The authors found that TD children acquire [a], [i], and [ɤ] at the age of one, [u] at the age of three, [ɻ̩] at the age of four, [ɹ̩] at the age of five, and [y] at the age of six. Via acoustic analysis, they also observed that the formants of [y], [ɻ̩], and [ɹ̩] were unstable in the spectrogram, reflecting the instability of tongue position during monophthong production in Mandarin-speaking children.

On the one hand, the studies mentioned above establish a basis for the current study to observe the speech sound development in children with ASD. On the other hand, these studies have proved that both auditory-perceptual analysis and acoustic analysis are necessary and important for investigating the development of speech sound production. The classical auditory-perceptual analysis could subjectively evaluate the general speech sound production from listeners' perceptual perspective which is important in speech communication, and the acoustic analysis could objectively uncover fine-grained production patterns in children with ASD and provide implications for clinical intervention.

Speech sound production in children with ASD

Many studies on children with ASD focused on their impaired abilities in semantics, syntax, and pragmatics (Cheung et al., 2020; Naigles, 2016), while relatively limited attempts have been made to figure out their performance of speech sound production, and there are some controversial findings on the development of speech sound production in this population.

One of the controversial arguments is whether children with ASD show impaired speech sound production. Many studies have reported the impairments of speech sound production in individuals with ASD (Rapin et al., 2009; Shriberg et al., 2001; Wolk & Brennan, 2013; Wolk et al., 2016; Wolk & Giesen, 2000). Rapin et al. (2009) conducted a cluster analysis in 62 children with ASD aged 7-9 years. The authors demonstrated that around 24% of school-age children with ASD showed significant speech sound disorders. However, some research demonstrated that speech sound production was relatively intact in children with ASD, compared with other language behaviours (Bartak et al., 1975; Bartolucci et al., 1976; Kjelgaard & Tager-Flusberg, 2001). For instance, Kjelgaard and Tager-Flusberg (2001) tested phonological, lexical, semantic, and grammatical skills in 89 children with ASD. They found that children with ASD had relatively intact speech sound production, whereas other abilities showed a large heterogeneity. However, this conclusion might underestimate the extent to which children with ASD experienced difficulties in speech sound production as single word articulation was only subjectively judged as correct or incorrect in this study. Although the auditory-perceptual binary classification is a classical method to analyse speech sound development in children, this may overlook many articulatory problems made by individuals with ASD (e.g., producing aspirated consonants with inappropriate aspiration in Yang (2018)) and thus the binary perceptual judgement of correctness does not reflect the possible deficits in speech sound production. A quantitative measure, such as an acoustic measure, is necessary to understand the speech production skills in children.

Another controversial argument is whether children with ASD show typical or atypical production patterns. McCleery et al. (2006) compared the consonant production in 14 children with ASD (2;1-6;11) and 10 language-matched TD children (1;1-1;2) via both spontaneous production and imitation tasks. They observed similar production patterns in TD children and children with ASD in regard to developmental difficulty and voicing. Both TD children and children with ASD produced more early developing sounds (e.g., /d/, /b/, /h/, /m/, /n/) than later developing sounds (e.g., /dʒ/, /l/, /ɹ/, /s/, /t/), and both groups produced more voiced consonants than voiceless consonants. Schoen et al. (2011) also found that speech-like vocalisation (e.g., syllables) in children with ASD was similar to language-matched TD children, while children with ASD showed more atypical nonspeech vocalisation (e.g., a greater number of high-pitched squeal) than both age-matched and language-matched TD children. However, Wolk and Giesen (2000) observed a ‘chronological mismatch’ of speech sound development in children with ASD such that early developing sounds (e.g., /m/) were absent whereas later developing sounds (e.g., /θ/) were present. Moreover, Wolk and Brennan (2013) analysed typical/atypical speech sound error patterns using object-naming tasks in eight children with ASD aged 5-15 years. Results showed that all participants exhibited some typical error patterns (e.g., pre-vocalic voicing, post-vocalic devoicing, fronting, stopping, and gliding) reflecting phonological delays, whereas some of them also showed atypical error patterns (e.g., pre-vocalic devoicing, post-vocalic voicing, deaffrication, migration, and backing). Chenausky et al. (2021) also compared speech sound production in English-speaking TD children and minimally verbal children with ASD via acoustic analysis, and found a narrower vowel space produced by children with ASD. Indeed, some research has reported motor planning problems for the temporal sequences of the articulator movements in children with ASD, which may partly account for the impaired speech sound production and articulation errors (Mody et al., 2017; Pang et al., 2016; Wong et al., 2020). Taken together, the controversial conclusions in the previous literature encouraged us to further investigate the speech sound production in Mandarin-speaking children with ASD.

Previous studies focus more on high-functioning children with ASD, especially those from English-speaking countries (Naigles, 2016). Recently, an increasing number of studies have set out to address this shortcoming. Wu et al. (2020) explored speech sound development in 16 Mandarin-speaking children with ASD aged three to six years via Mandarin picture naming tasks and imitation tasks. They found some typical developmental patterns (e.g., aspirated consonants were acquired later than unaspirated ones, which was also reported in TD children in Xie (2009)) in children with ASD. Besides, they also found that children with ASD acquired vowels earlier than tones, whereas chronological-age-matched TD children acquired tones earlier than vowels, indicating an atypical developmental sequence of Mandarin phonology among individuals with ASD. However, only an auditory-perceptual assessment was conducted by Wu et al. (2020), which was inadequate to objectively reflect production patterns in Mandarin-speaking children with ASD.

It is noteworthy that all studies we mentioned above adopted different methods for eliciting speech sound production, which may cause variations in evaluating children’s performance (James et al., 2016; McLeod & Masso, 2019). One commonly used method is to collect spontaneous speech via picture/object naming tasks or by recording daily communication between children and parents or clinicians (e.g., James, et al., 2016; Wolk & Brennan, 2013). However, it is difficult to use this method to elicit spontaneous speech in minimally verbal children with ASD, especially younger ones, since the lack of social motivation could curb speech output (McCleery et al., 2006). Speech sound production assessed by picture/object naming tasks may also be affected by the selection of testing items (James, et al., 2016) and the mental vocabulary size of children with ASD. The other method is imitation, which is more suitable for children with ASD who lack the desire to communicate or minimally verbal ones with small vocabulary sizes. Some research reported that spontaneous production and imitation did not cause different results in children (Johnson et al., 2004; McCleery et al., 2006; McLeod & Masso, 2019). These studies also claimed that the imitation task required much less time to complete and could be regarded as a valid alternative to the spontaneous conversation task (Johnson et al., 2004; McLeod & Masso, 2019). However, Wolk et al. (2016) argued that imitation may overestimate children's actual production abilities since speech models are provided. Considering the factors mentioned above, Wu et al. (2020) combined the picture-naming task and imitation task for assessing speech sound production. However, that still could not solve the issue caused by the differing natures of tasks. Even in the imitation task, different procedures may also make some differences. Edwards (2014) clearly distinguished imitation and emulation. Strictly speaking, imitation is defined as a process that people reproduce the form and result of an action (Edwards, 2014). In Wu et al. (2020), the experimenter read the target word and children could reproduce the word (i.e., result) via imitating the movement of articulators (i.e., form). Different from imitation, emulation is a process that people reproduce the result of an action using their own execution. In this case, children's speech sound production may not be overestimated too severely, and researchers can also focus on the difficulty of speech sound production in children and avoid the possible effect of vocabulary size and social motivation. More investigations are needed to uncover and compare the role of different tasks in the assessment of speech sound production.

The present study

Mandarin consonant aspiration is unique in Mandarin phonology compared to English (the most widely investigated language in ASD research). Thus, the mainstream of speech sound production research focusing on English-speaking children with ASD did not include the production pattern of consonant aspiration. As we have introduced above, Wolk and Brennan (2013) found production problems of voicing in children with ASD, and voicing was reported similar to aspiration (e.g., Lisker & Abramson, 1964; Whitehill, 2010). Moreover, the main acoustic difference between aspirated and unaspirated consonants in Mandarin is the temporal cue of voice onset time (VOT). The temporal processing deficits in individuals with ASD have been commonly observed with evidence from several behavioural and neuroimaging studies (Brodeur et al., 2014; Falter et al., 2012; Martin et al., 2010). Taken together, it is expected that Mandarin-speaking children with ASD may also show difficulties in producing consonant aspiration. And these are the reasons why we aim to investigate consonant aspiration in this study. In regard to monophthongs, Chenausky et al. (2021) reported a narrower vowel space in children with ASD. Previous research has proposed that the narrow vowel space could influence speech clarity, and the size of vowel space might predict language development in young children (Liu et al., 2003). Thus, we also expect to investigate monophthongs in the current study.

Therefore, the first objective of this study is to figure out whether Mandarin-speaking children with ASD show difficulties in producing aspirated/unaspirated consonants and monophthongs using auditory-perceptual measurements. The second objective of this study is to investigate the production patterns of aspirated/unaspirated consonants and monophthongs in Mandarin-speaking children with ASD by performing quantitative acoustic analyses. The third objective of the current study is to explore the possible correlations between production performance and chronological age/developmental age of language/social impairment in children with ASD.

Methods

Participants

We recruited 27 Mandarin-speaking monolingual children with ASD (26 boys) aged 3.33-7.00 years (mean age = 4.76, SD = 0.97) and 30 chronological-age-matched TD children (23 boys, mean age = 4.54, SD = 0.98) as the control group. Appendixes A and B show the chronological age of each child. They came from different families speaking Mandarin and had no history of hearing impairment according to the parents' report. A consent form was obtained from each child's parent with a protocol approved by the Human and Animal Experiment Ethics Committee of Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences. TD children were recruited from mainstream kindergartens and primary schools in Mainland China, without explicit language impairments according to their parents' and teachers' reports. Children with ASD were diagnosed based on the diagnostic and statistical manual of mental disorders criteria (American Psychiatric Association, 2013) and the Childhood Autism Rating Scale (Ozonoff et al., 2005) by pediatricians and child psychiatrists with expertise in diagnosing ASD. All children with ASD completed the translated version of Psychoeducational Profile-Third Edition (PEP-3) (Schopler et al., 2005; Yu et al., 2019). PEP-3 is a standardised assessment with a referenced norm for children with ASD between the ages of 2 and 7.5 years. There are ten subtests in PEP-3, including cognitive verbal/preverbal, expressive language (EL), receptive language (RL), fine motor, gross motor, visual-motor imitation, affective expression (AE), social reciprocity (SR), characteristic motor behaviour (CMB), and characteristic verbal behaviour (CVB). The rating scale of each test item is from 0 to 2: that is, "passing" (2 points), "emerging" (1 point), and "failing" (0 points). PEP-3 also provides a score transformation system, in which the scores for all test items are converted into developmental ages based on a TD norm. The scores of PEP-3 subtests in children with ASD are shown in Appendix A. Their developmental age of language ability ranging from 1.13 to 4.58 years (mean age = 2.21, SD = 0.76) was obtained from the mean of the subtests of EL and RL in the PEP-3. The score of maladaptive behaviour, obtained from the subtests of AE, SR, CMB, and CVB in PEP-3, was regarded as an indicator of social impairment in children with ASD. Table 1 lists the demographic information of ASD and TD groups.

Table 1. Demographic information of children with ASD and TD children.

	n	Chronological age (year)		Developmental age of language (year)	
		Mean (SD)	Range	Mean (SD)	Range
ASD	27	4.76 (1.01)	3.33-7.00	2.21 (0.78)	1.13-4.58
TD	30	4.52 (0.90)	3.33-7.00	N/A	

Note. n = number of participants, ASD = Autism Spectrum Disorder, TD = typically developing, N/A = not applicable.

Materials and procedures

Sixteen Mandarin consonant-vowel syllables carrying high-level tone were chosen and shown here in Pinyin and the corresponding International Phonetic Alphabet: bū [pu], pū [pʰu], bō [po], pō [pʰo], dē [tɤ], tē [tʰɤ], gā [ka], kā [kʰa], jī [tɕi], qī [tɕʰi], jū [tɕy], qū [tɕʰy], zī [tsɹ̩], cī [tsʰɹ̩], zhī [tʂɻ̩], chī [tʂʰɻ̩]. These syllables included six pairs of voiceless aspirated and unaspirated consonants and eight monophthongs with reference to Zhu and Dodd (2000). As previous studies reported that the high-level tone was the earliest acquired tone for native Mandarin-speaking children (Li &

Thompson, 1977; Li & To, 2017; Zhu, 2002), we chose syllables with the high-level tone to reduce the difficulty of tasks. Also, controlling the tone of each syllable was to reduce the impact of the tone-context effect. To avoid causing fatigue in children with ASD, we had to shorten the duration of the whole experiment and used these 16 Mandarin syllables as materials in this experiment.

The 16 Mandarin syllables were firstly pre-recorded by a female language teacher aged 27 years from Northern China using Praat (Boersma & Weenink, 2020) with a 44 100 Hz sampling rate and 16-bit resolution. The language teacher taught Mandarin, and had reached the category A of level 2 in the Putonghua Proficiency Test. Those who have reached this level generally have the standard pronunciation of Mandarin speech. Both ASD and TD children were asked to stay in a quiet aural rehabilitation room (a soundproof booth) with parents and listen to computer playbacks of pre-recorded audio files. Only auditory information was given to the participants to avoid giving visual information for imitation. This could avoid children imitating the movements of the speaker's articulators, such as the movement of lips. All children were instructed to imitate (similar to "emulate" in Edwards (2014)) the speech sound twice after hearing the pre-recorded sound. The sounds produced by children were recorded using an external microphone (SHURE MV51) connected to a laptop with Praat (44 100 Hz sampling rate, 16-bit resolution). The microphone was fixed around 10 cm away from the children's mouth.

Data analysis

Although the experiment was conducted in a quiet room, there was irrelevant noise from children and parents on-site during the recording. We excluded productions with irrelevant environmental noise that were impossible to extract acoustic information from speech sounds. If one of the two productions per syllable had irrelevant environmental noise, the other clear production without noise was used for further analysis. The exclusive criterion was only environmental noise. If both productions were clear, we chose the first production to conduct further analysis.

We recruited five Mandarin-speaking raters from Northern China (26-29 years old, three males), who majored in linguistics with a Master's degree for auditory-perceptual scoring. All raters had normal hearing according to self-reports and were familiar with the scoring and disordered speech. The raters independently assessed the children's production of monophthongs and aspirated/unaspirated consonants separately. As we mentioned in the introduction, a binary scale (correct or incorrect) may not be fine-grained enough and a 5-point scale has been supported in some previous studies (e.g., Chen et al., 2019; Strand et al., 2014). Therefore, we used a five-point scale from completely accurate (five points) to completely inaccurate (one point). Recordings of ASD and TD children were randomly presented to the raters, and they were blinded to the child's group to prevent bias. Inter-rater reliability was assessed by the intra-class correlation coefficient (ICC, Koo & Li, 2016). The two-way random-effects model, absolute-agreement, and the mean-rating ($k = 5$) were used to assess the rating reliability across the five raters. There was a good to excellent agreement for monophthong scoring (ICC = 0.884, $p < 0.001$) and consonant scoring (ICC = 0.908, $p < 0.001$). No intra-rater reliability was measured in this study. Pearson correlation analysis was conducted to explore the relationship between the production scores and chronological age/developmental age of language/social impairment in the ASD group.

For quantitative acoustic analysis of aspirated/unaspirated consonants, VOT in milliseconds from the release of sound (i.e., the onset of burst) to the onset of vocal fold vibration (i.e., the beginning point of periodicity) was extracted manually in Praat (Abramson & Whalen, 2017). VOT is a widely used acoustic indicator of aspirated and unaspirated consonants (Chao & Chen, 2008; Lisker & Abramson, 1964). Aspirated consonants tend to show longer VOTs than unaspirated ones in Mandarin. For acoustic analysis of monophthongs, the measurement of frequencies of the first and second formants (F1, F2) was defined as the articulatory-referenced locations based on the stability of formant patterns in Praat (Kent & Vorperian, 2018). F1 has been demonstrated to be inversely correlated with tongue height (i.e. the higher the F1 frequency, the lower the tongue height), and F2 is associated with tongue backness and lip rounding in previous research (i.e. the higher the F2 frequency, the more front the tongue position and the less rounded the lip) (Delattre, 1951; Lee et al., 2016). Formant values in Hertz were converted to the Bark scale using the formula in Zwicker and Terhardt (1980). The Bark scale is a psychoacoustical scale to map acoustic features to auditory perceptual representation. The purpose of this transformation is to reduce the influence of physiological and anatomical inter-speaker differences and maintain the separation of vowel categories (e.g., Flynn, 2011). This method was widely used in previous studies on the measurement of vowel space (e.g., Most et al., 2000; Neumeyer et al., 2010). Vowel space size was defined as a triangle area determined by three corner vowels (i.e. /i, a, u/) and calculated by the formula (1) proposed by Liu et al. (2009), where F1i represented the F1 of /i/, F2a represented the F2 of /a/ and so on. Because the age range of participants in this study was wide, which may lead to some individual differences in the size of vocal tracts, we further conducted 1:1 gender- and age-matching to compare the vowel space of TD children and those with ASD with 15 participants in each group.

$$\text{Vowel space size} = \text{Absolute value } \{[F1i \times (F2a - F2u) + F1a \times (F2u - F2i) + F1u \times (F2i - F2a)]/2\} \quad (1)$$

Given the potential context effect of different consonants and monophthongs and individual differences across subjects, linear mixed-effect models (LMMs) in R (R Core Team, 2019) were adopted. To examine whether Mandarin-speaking children with ASD showed difficulty in producing aspirated/unaspirated consonants and monophthongs, the production scores were compared in TD children and children with ASD. To figure out the production patterns of children with ASD, VOT of aspirated/unaspirated consonants and the vowel space of monophthongs were compared between the two groups. The packages of lme4 (Bates et al., 2015) and lmerTest (Kuznetsova et al., 2017) were used to create the LMMs. There were four

LMMs in total. Bonferroni correction was used to conduct multiple comparisons.

Results

Production score of aspirated/unaspirated consonants

The scores for consonants in ASD and TD groups are shown in Figure 1. We used an LMM to describe the scores of consonants averaged across the raters as a function of *group* (ASD versus TD). The model included a fixed factor of *group*. In addition, it included random factors for *subject* and *syllable*. The mean score of consonants was 3.425 in the ASD group and 4.726 in the TD group (β = 1.303, SE = 0.147). This difference was estimated to be reliable (χ^2 = 49.36, p < 0.0001), indicating an impaired production of aspirated/unaspirated consonants in children with ASD.

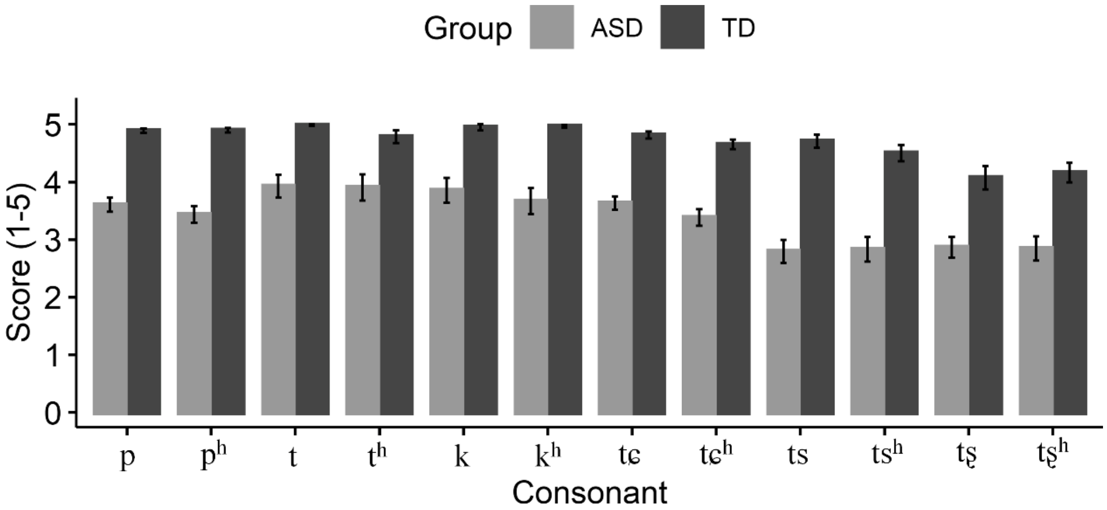

Figure 1. Scores for Mandarin aspirated/unaspirated consonants in ASD and TD groups. *Note.* Error bars = +/- 1 standard error, ASD = Autism Spectrum Disorder, TD = typically developing.***Production score of monophthongs***

The scores for eight Mandarin monophthongs in ASD and TD children are shown in Figure 2. An LMM was created to describe the scores of monophthongs averaged across the raters as a function of *group* (ASD versus TD). The model included a fixed factor of *group*, and random factors of s*ubject* and *syllable*. The mean score for monophthongs was 3.724 in the ASD group and 4.833 in the TD group ($\beta = 1.110$, $SE = 0.110$). This difference was significant ($\chi^2 = 58.34$, $p < 0.0001$), revealing an impaired monophthong production in children with ASD.

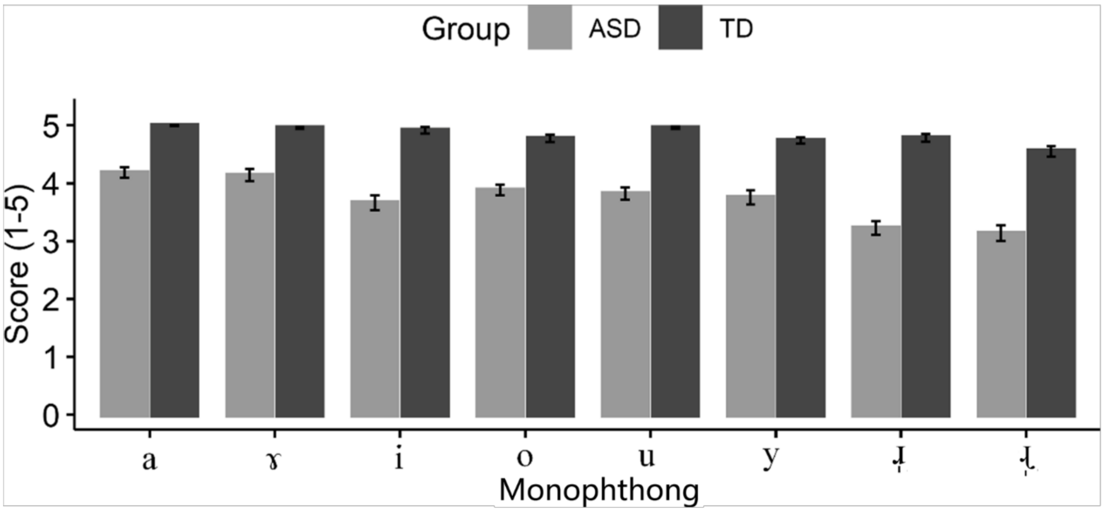

Figure 2. Scores for Mandarin monophthongs in ASD and TD groups. *Note.* Error bars = +/- 1 standard error, ASD = Autism Spectrum Disorder, TD = typically developing.

Correlation analysis

There was a significant positive correlation between developmental age of language and consonant score ($r = 0.453$, $p = 0.018$). This indicated that the production of aspirated/unaspirated consonants was better with an increased developmental age of language in children with ASD. However, no significant correlation was observed between chronological age and consonant score (*r* = -0.321, *p* = 0.102), suggesting that their production of aspirated/unaspirated consonants may not necessarily improve with chronological age. There was no significant correlation between social impairment (mean = 35.87, SD = 11.47) and consonant score (*r* = 0.179, *p* = 0.371) in children with ASD.

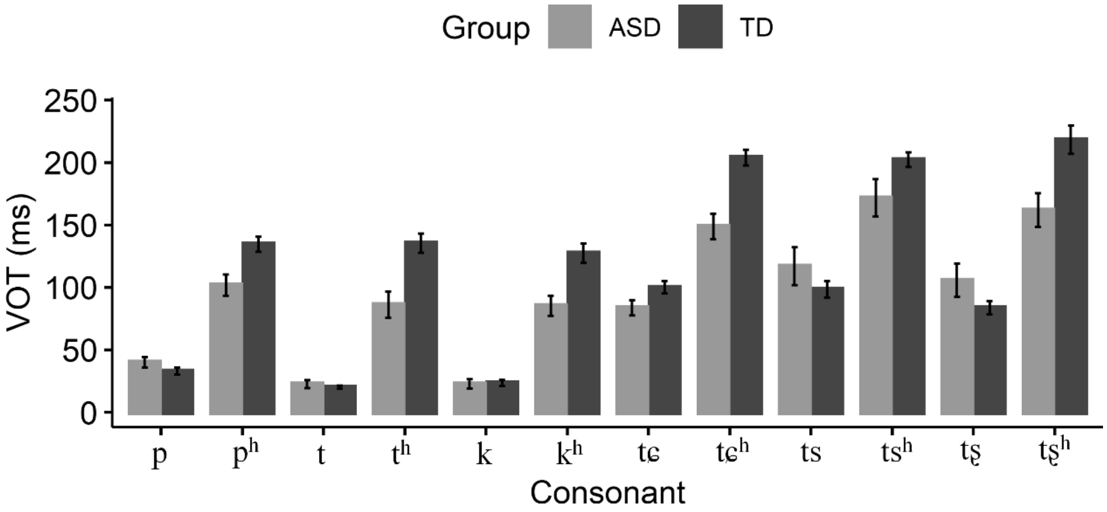

Figure 3. VOT of Mandarin aspirated/unaspirated consonants in ASD and TD groups. *Note.* Error bars = +/- 1 standard error, VOT = voice onset time, ASD = Autism Spectrum Disorder, TD = typically developing.

Similarly, a significant positive correlation between monophthong score and developmental age of language (*r* = 0.568, *p* = 0.002) was found in children with ASD. This demonstrated that the production of monophthong improved with increased developmental age. However, there was no significant correlation between chronological age and monophthong score (*r* = -0.363, *p* = 0.063) in the ASD group. This revealed that the production of monophthongs in children with ASD may not necessarily improve with increased chronological age. No significant correlation was detected between social impairment and monophthong score ($r = 0.295$, $p = 0.135$) in children with ASD.

Production patterns of aspirated/unaspirated consonants

Figure 3 shows the VOT of aspirated/unaspirated consonants in ASD and TD groups. We created an LMM to describe the VOT of consonants as a function of *group* (ASD versus TD), and *aspiration* (aspirated versus unaspirated). It consisted of fixed factors of *group*, *aspiration*, and the *group* × *aspiration* interaction, and random factors of *subject* and *syllable*. The mean VOT for aspirated consonants in the ASD group was 125.559 ms, and that in the TD group was 170.443 ms. The mean VOT for unaspirated consonants in the ASD group was 65.753 ms, and that in the TD group was 61.911 ms. We observed a significant interaction effect between *group* and *aspiration* ($\chi^2 = 68.01$, $p < 0.0001$). For both groups, VOT for unaspirated consonants was significantly shorter than that for aspirated ones (for the ASD group: $\beta = 60.85$, $SE = 18.97$, $t = 3.208$, $p = 0.028$; for the TD group: $\beta = 108.60$, $SE = 18.90$, $t = 5.747$, $p = 0.0001$). The difference in VOT of unaspirated consonants was not significant between the two groups ($\beta = 2.62$, $SE = 7.26$, $t = 0.361$, $p = 1.000$). However, children with ASD produced shorter VOT for aspirated consonants than TD children ($\beta = -45.13$, $SE = 7.25$, $t = -6.229$, $p < 0.0001$).

Production patterns of monophthongs

Figure 4 (A) shows acoustic vowel space measured by F1 and F2 in the Bark scale across ASD (solid line) and TD children (dashed line). Figure 4 (B) presents the vowel space of 1:1 gender- and age-matched TD children and those with ASD with 15 participants in each group. The vowel space was described in the LMM as a function of *group* (ASD versus TD). The model consisted of a fixed factor of *group,* and a random factor of *subject*. The mean vowel space was 7.18 in the ASD group and 15.76 in the TD group (β = 8.581, *SE* = 1.620). The difference was estimated to be significant (F = 28.07, $p < 0.0001$). The results showed a significantly reduced vowel space in children with ASD. After 1:1 gender- and age-matching, we still found the pattern that each child with ASD (mean = 6.31) consistently showed a narrower vowel space than the age-matched TD child (mean = 14.14). The difference was also significant between the two groups (F = 15.26, p = 0.001). The similar patterns between the whole group comparison (i.e., 27 ASD vs. 30 TD) and the 1:1 gender- and age-matching comparison (i.e., 15 ASD vs. 15 TD) (also see Figures 4 (A) and (B)) indicated the robustness of such finding.

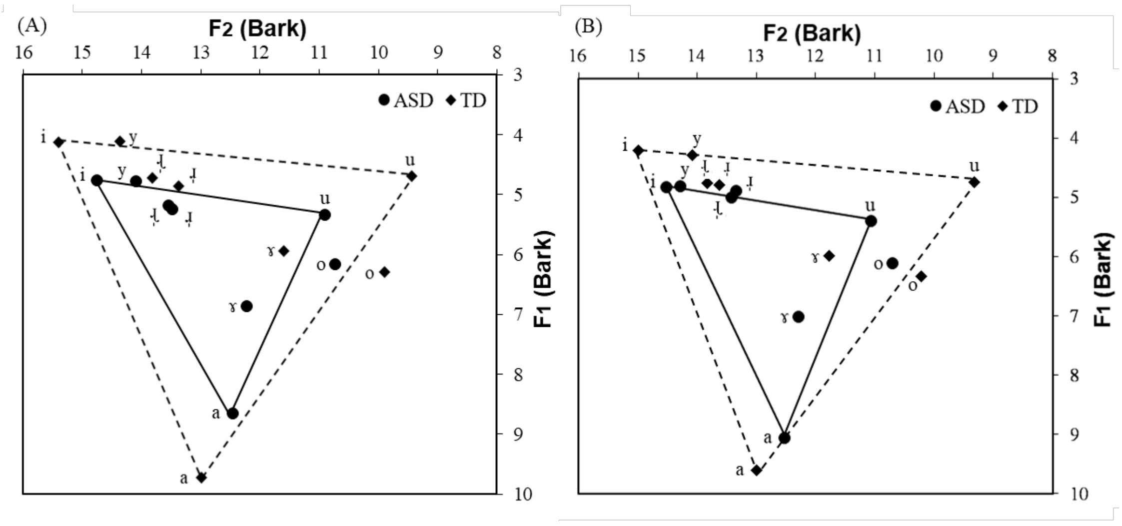

Figure 4. (A) F1 and F2 in the Bark scale of Mandarin monophthongs among ASD and TD groups. (B) Vowel space of 1:1 gender- and age-matched TD children and those with ASD with 15 participants in each group. *Note.* F1 = the first formant, F2 = the second formant, ASD = Autism Spectrum Disorder, TD = typically developing.

Discussion

This study intended to investigate the production of Mandarin aspirated/unaspirated consonants and monophthongs among native young children with ASD. The auditory-perceptual scoring results suggested that Mandarin-speaking young children with ASD exhibited a significantly impaired production of aspirated/unaspirated consonants and monophthongs. Besides, the production of monophthongs and aspirated/unaspirated consonants in children with ASD was found to be better with increased developmental age. However, no evidence showed that their production performance improved with increased chronological age or different degrees of social impairment. Furthermore, the quantitative acoustic analysis indicated that children with ASD produced a shorter VOT for aspirated consonants and a reduced vowel space for monophthongs, in comparison to age-matched TD children.

Impaired production of aspirated/unaspirated consonants and monophthongs in children with ASD

In our study, TD children have obtained auditory-perceptual scores close to the ceiling (five points), with the exception of the two retroflex affricates [tʂ] and [tʂ^h] (close to four points) which were reported to be challenging even for TD children at six or seven years of age (Li & To, 2017). Compared to TD children, the chronological-age-matched children with ASD showed significantly lower scores for Mandarin aspirated/unaspirated consonants and monophthongs. This indicated an impaired production of aspirated/unaspirated consonants and monophthongs among Mandarin-speaking children with ASD, which is consistent with the findings in previous literature (Wolk & Brennan, 2013; Experiment 1 in Wu et al., 2020).

Additionally, the production performance was better with increased developmental age of language in children with ASD, consistent with the results in Wolk and Brennan (2013) and Wu et al. (2020). And we did not observe any significant correlations between chronological age and production scores. The two findings implied that their speech sound production may not necessarily improve with chronological age. Thus, the language assessment is necessary, and early interventions should be taken to improve their speech sound production.

Production patterns of consonant aspiration and monophthongs in children with ASD

We observed several production patterns in Mandarin-speaking children with ASD via the quantitative acoustic analysis. VOT for aspirated consonants in children with ASD was shorter than TD ones. Besides, there was a reduced vowel space in children with ASD, which was consistent with Chenausky et al. (2021).

For the unaspirated consonants, no significant difference in VOT was found between the ASD and TD groups in the acoustic analysis, while the auditory-perceptual analysis showed a significant difference in production scores between the two groups. It is noteworthy that the formant transition (reflecting the rapid changes of the vocal tract after the release of consonants) plays an important role in consonant perception (e.g., Walley & Carrell, 1983). In both auditory-perceptual and acoustic analyses, children with ASD underperformed in the production of monophthongs. Although raters were asked to assess the consonants without considering the following monophthongs, the formant transition accompanied by the following monophthongs might still affect raters' auditory-perceptual scoring. However, the acoustic analysis of VOT was not affected by the formant transition at all. Therefore, it is reasonable that the two analyses showed asymmetrical results.

Although the correlation results between production performance and social impairment in this study showed no evidence for the mechanisms underlying the difficulties of aspiration production in children with ASD, previous research indeed proposed several possible reasons. Firstly, previous studies reported that the impairment of social interaction in children with ASD impedes their development of speech production. It is well-documented that social interaction plays an important role in children's language development (Kuhl, 2000). Via social interaction, children could learn how to communicate with others in an intelligible and appropriate way. They may realise their speech errors and correct these errors to produce typical-like speech sounds approaching their social partners during social interactions (McKeever et al., 2019). The speech attunement framework proposes that children need to engage in the ambient environment and to develop speech production to support intelligible and socially appropriate communication (Shriberg et al., 2011). However, the impairment of social interaction is an evident characteristic of children with ASD. Their development of speech production cannot be efficiently benefited from social interactions like TD ones, and their impaired speech production further hinders social interactions. The speech attunement framework posits that the challenges of social interaction may affect their abilities to monitor and correct speech errors, and thus they cannot develop appropriate speech production. Since we used the imitation task in this study, the data we collected may not fully reflect children's speech communication in real life. Therefore, no significant correlation between production score and social impairment was observed. Secondly, Pang et al. (2016) using magnetoencephalography found an abnormal latency and activation in the primary motor cortex, motor planning and executive control areas, and temporal sequencing and sensorimotor integration areas in children with ASD. Consonant aspiration requires speakers to coordinate the articulator movements to realise a burst of airflow and an immediate onset of phonation (Wong et al., 2020). Thus, the deficits of speech motor planning skills in children with ASD may degrade their performances on sequencing the articulatory movements that are necessary for the aspirated sounds. Thirdly, the perceptual deficit in children with ASD (Huang et al., 2018, Wang et al., 2017; You et al., 2017; Yu et al., 2015) may also limit their ability to perceive others' speech and adjust their speech production to ensure an intelligible and appropriate speech production.

Clinical implications

We found that speech sound production in children with ASD may not necessarily improve with chronological age. Thus, we suggested that early intervention is necessary. Our study also uncovered some production patterns of children with ASD, providing direct references for the clinical intervention of speech sound production. Children with ASD were found to produce aspirated consonants with shorter VOT. Thus, intervention focusing on lengthening VOT of aspirated consonants may be helpful to children with ASD. Chen et al. (2019) designed a 3-D virtual pronunciation tutor with visible articulatory movements and airflow changes for Mandarin-speaking children with ASD. Aspirated consonants were accompanied by longer and larger airflow in the 3-D virtual tutor and the eye-tracking evidence showed that children with ASD exhibited more interests in the 3-D virtual tutor than real human face and indeed paid attention to the aspirated airflow of the 3-D virtual tutor. That was proved to be helpful to improve their consonant production ability.

Besides, we also found a reduced vowel space in children with ASD. Previous studies have provided evidence for the facilitative effect of larger vowel space on the development of speech discrimination skills and spoken word recognition in children (Liu et al., 2003; Song et al., 2010). Chen et al. (2021) also reported that children showed increased cortical response to formant-exaggerated stimuli. Therefore, input with formant exaggeration may facilitate monophthong production in children with ASD. The infant-directed speech is often produced with expanded vowel space (Chen et al., 2021). In clinical intervention, clinicians could use infant-directed speech so that children with ASD could receive the input with formant exaggeration. Besides, the vowels with formant exaggeration could be generated by a formant synthesizer (e.g., Klatt-type formant synthesizer in Hanson and Stevens (2002)), and presented to children with ASD by computer-assisted language training.

Limitations and future direction

This study presented impaired production of aspirated/unaspirated consonants and monophthongs and their production patterns in Mandarin-speaking children with ASD. However, the findings could not be directly generalised to complex syllables. Previous studies reported that the stimulus characteristics of syllables could influence children's performance (Hodges et al., 2017), and clinical judgements of children's speech production may also be benefited from tests with di- or polysyllabic words (James et al., 2016). Thus, more complex testing syllables should be included in future research. Considering the difficulty of collecting speech production in minimally verbal children, we used the imitation task which may overestimate their production abilities. Besides, to exclude the interruption of environmental noise in tokens to the quantitative acoustic analysis, we only analysed one token per syllable of each child. The data set may not be large enough to powerfully reflect their production abilities.

In addition, a further longitudinal study may provide more solid data for the correlation between chronological age and language development. Besides, TD children’s language abilities should also be assessed in the future to make a detailed comparison with children with ASD. Without cross-language investigation, our findings only focused on Mandarin-speaking children with ASD and could not be directly generalised to other native language speakers.

Furthermore, this study could not figure out the causal relationship between specific deficits (e.g., motor control deficit, or perceptual deficit) and the impairment of speech sound production. More attention should be paid to uncovering the specific deficits underlying the difficulty of speech sound production in children with ASD. Based on this, targeted intervention would be expected to improve their speech sound production.

Conclusions

The findings of this study indicated an impaired production of aspirated/unaspirated consonants and monophthongs in children with ASD, in comparison to chronological-age-matched TD children. Importantly, the quantitative acoustic analysis further demonstrated that Mandarin-speaking children with ASD showed shorter VOT for aspirated consonants and narrower vowel space. Specific interventions focusing on these production patterns need to be explored to improve their speech sound production in future research.

Acknowledgements

This study was supported by grants from National Natural Science Foundation of China (11974374), Hong Kong Research Grants Council – General Research Fund (15610321), Ministry of Education of Humanities and Social Science Project (20YJC740041), and the Fundamental Research Funds for the Central Universities (531118010660). We thank all reviewers for their valuable comments on this paper.

Disclosure of interest

The authors declare no conflict of interest.

References

Abramson, A. S., & Whalen, D. H. (2017). Voice Onset Time (VOT) at 50: Theoretical and practical issues in measuring voicing distinctions. *Journal of Phonetics, 63*, 75-86. https://doi.org/10.1016/j.wocn.2017.05.002

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders, 5th edition: DSM-5*. Washington, DC: American Psychiatric Pub.

Bartak, L., Rutter, M., & Cox, A. (1975). A comparative study of infantile autism and specific developmental receptive language disorder. *The British Journal of Psychiatry, 126*(2), 127–145. https://doi.org/10.1007/BF01540396

Bartolucci, G., Pierce, S., Streiner, D., & Eppel, P. T. (1976). Phonological investigation of verbal autistic and mentally retarded subjects. *Journal of Autism and Developmental Disorders, 6*(4), 303-316. https://doi.org/10.1007/BF01537908

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

Boersma, P., & Weenink, D. (2020). Praat: Doing phonetics by computer (Version 6.1.09) [Computer software]. http://www.praat.org

Brodeur, D. A., Green, C. G., Flores, H., & Burack, J. A. (2014). Time estimation among low-functioning individuals with autism spectrum disorders: Evidence of poor sensitivity to variability of short durations. *Autism Research, 7*(2), 237–244.

https://doi.org/10.1002/aur.1364

Chao, K. Y., & Chen, L. M. (2008). A cross-linguistic study of voice onset time in stop consonant productions. *International Journal of Computational Linguistics & Chinese Language Processing, 13*(2), 215-232.

Chen, L. M., & Kent, R. D. (2010). Segmental production in Mandarin-learning infants. *Journal of Child Language, 37*(2), 341–371. https://doi.org/10.1017/S0305000909009581

Chen, F., Wang, L., Peng, G., Yan, N., & Pan, X. (2019). Development and evaluation of a 3-D virtual pronunciation tutor for children with autism spectrum disorders. *PLoS ONE, 14*(1), e0210858. https://doi.org/10.1371/journal.pone.0210858

Chen, F., Zhang, H., Ding, H., Wang, S., Peng, G., & Zhang, Y. (2021). Neural coding of formant‐exaggerated speech and nonspeech in children with and without autism spectrum disorders. *Autism Research, 14*(7), 1357‐1374. https://doi.org/10.1002/aur.2509

Chenausky, K. V., Brignell, A., Morgan, A. T., Norton, A. C., Tager-Flusberg, H. B., Schlaug, G., & Guenther, F. H. (2021). A modeling-guided case study of disordered speech in minimally verbal children with autism spectrum disorder. *American Journal of Speech-Language Pathology*, *30*(3S), 1542-1557. https://doi.org/10.1044/2021_AJSLP-20-00121

Cheung, C. C. H., Rong, Y., Chen, F., Chen, S., Leung, M. T., Tang, T. P. Y., & Peng, G. (2020). Comprehension of literal statements and similes in Cantonese-speaking children with and without autism spectrum disorders. *Clinical Linguistics & Phonetics, 34*(4), 312-326. https://doi.org/10.1080/02699206.2019.1634765

Delattre, P. (1951). The physiological interpretation of sound spectrograms. *Publications of the Modern Language Association of America, 66*(5), 864-875. https://doi.org/10.2307/459542

Edwards, L. A. (2014). A meta-analysis of imitation abilities in individuals with autism spectrum disorders. *Autism Research, 7*(3), 363-380. https://doi.org/10.1002/aur.1379

Falter, C. M., Noreika, V., Wearden, J. H., & Bailey, A. J. (2012). More consistent, yet less sensitive: Interval timing in autism spectrum disorders. *Quarterly Journal of Experimental Psychology, 65*(11), 2093–2107. https://doi.org/10.1080/17470218.2012.690770

Flynn, N. (2011). Comparing vowel formant normalisation procedures. *York Papers in Linguistics, 2*(11), 1-28.

Hanson, H. M., & Stevens, K. N. (2002). A quasiarticulatory approach to controlling acoustic source parameters in a Klatt-type formant synthesizer using HLsyn. *The Journal of the Acoustical Society of America, 112*(3), 1158-1182. https://doi.org/10.1121/1.1498851

Hodges, R., Munro, N., Baker, E., McGregor, K., & Heard, R. (2017). The monosyllable imitation test for toddlers: influence of stimulus characteristics on imitation, compliance and diagnostic accuracy. *International Journal of Language & Communication Disorders, 52*(1), 30-45. https://doi.org/10.1111/1460-6984.12249

Huang, D., Yu, L., Wang, X., Fan, Y., Wang, S., & Zhang, Y. (2018). Distinct patterns of discrimination and orienting for temporal processing of speech and nonspeech in Chinese children with autism: an event-related potential study. *European Journal of Neuroscience*, *47*(6), 662-668. https://doi.org/10.1111/ejn.13657

James, D. G., Ferguson, W. A., & Butcher, A. (2016). Assessing children's speech using picture-naming: The influence of differing phonological variables on some speech outcomes. *International Journal of Speech-Language Pathology, 18*(4), 364-377. https://doi.org/10.3109/17549507.2015.1101159

Johnson, C. A., Weston, A. D., & Bain, B. A. (2004). An objective and time-efficient method for determining severity of childhood speech delay. *American Journal of Speech-Language Pathology, 13*(1), 55-65. https://doi.org/10.1044/1058-0360(2004/007)

Kent, R. D., & Vorperian, H. K. (2018). Static measurements of vowel formant frequencies and bandwidths: A review. *Journal of Communication Disorders, 74*, 74-97. https://doi.org/10.1016/j.jcomdis.2018.05.004

Kjelgaard, M. M., & Tager-Flusberg, H. (2001). An investigation of language impairment in autism: Implications for genetic subgroups. *Language and Cognitive Processes, 16*(2-3), 287–308. https://doi.org/10.1080/01690960042000058

Klinger, L. G., Dawson, G., Barnes, K., & Crisler, M. (2014). Autism spectrum disorder. In E. J. Mash & R. A. Barkley (Eds.), *Child psychopathology* (3rd ed., pp. 531–572). New York, NY: Guilford Press.

Koo, T. K., & Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine, 15*(2), 155-163. https://doi.org/10.1016/j.jcm.2016.02.012

Kuhl, P. K. (2000). A new view of language acquisition. *Proceedings of the National Academy of Sciences, 97*(22), 11850–11857. https://doi.org/10.1073/pnas.97.22.11850

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: tests in linear mixed effects models. *Journal of Statistical Software, 82*(13), 1-26. https://doi.org/10.18637/jss.v082.i13

Lee, J., Shaiman, S., & Weismer, G. (2016). Relationship between tongue positions and formant frequencies in female speakers. *The Journal of the Acoustical Society of America, 139*(1), 426-440. https://doi.org/10.1121/1.4939894

Li, C. N., & Thompson, S. A. (1977). The acquisition of tone in Mandarin-speaking children. *Journal of Child Language*, *4*(2), 185–199.

https://doi.org/10.1017/S0305000900001598

Li, X., & To, C. K. (2017). A review of phonological development of Mandarin-speaking children. *American Journal of Speech-Language Pathology, 26*(4), 1262-1278. https://doi.org/10.1044/2017_AJSLP-16-0061

Lisker, L., & Abramson, A. S. (1964). A cross-language study of voicing in initial stops: Acoustical measurements. *Word, 20*(3), 384-422. https://doi.org/10.1080/00437956.1964.11659830

Liu, H. M., Kuhl, P. K., & Tsao, F. M. (2003). An association between mothers' speech clarity and infants' speech discrimination skills. *Developmental Science, 6*(3), F1-F10. https://doi.org/10.1111/1467-7687.00275

Liu, H. M., Tsao, F. M., & Kuhl, P. K. (2009). Age-related changes in acoustic modifications of Mandarin maternal speech to preverbal infants and five-year-old children: a longitudinal study. *Journal of Child Language, 36*(4), 909-922. https://doi.org/10.1017/S030500090800929X

Martin, J. S., Poirier, M., & Bowler, D. M. (2010). Brief report: Impaired temporal reproduction performance in adults with autism spectrum disorder. *Journal of Autism and Developmental Disorders, 40*(5), 640–646. https://doi.org/10.1007/s10803-009-0904-3

McCleery, J. P., Tully, L., Slevc, L. R., & Schreibman, L. (2006). Consonant production patterns of young severely language-delayed children with autism.

Journal of Communication Disorders, 39(3), 217–231. https://doi.org/10.1016/j.jcomdis.2005.12.002

McKeever, L., Cleland, J., & Delafield-Butt, J. (2019). Aetiology of speech sound errors in autism. In S. Fuchs, J. Cleland, & A. Rochet-Capellan (Eds.), *Speech production and perception: Learning and memory* (pp. 109-138). Peter Lang.

McLeod, S. & Masso, S. (2019). Screening children's speech: The impact of imitated elicitation and word position. *Language, Speech, and Hearing Services in Schools, 50*(1), 71-82. https://doi.org/10.1044/2018_LSHSS-17-0141

Mody, M., Shui, A. M., Nowinski, L. A., Golas, S. B., Ferrone, C., O'Rourke, J. A., & McDougle, C. J. (2017). Communication deficits and the motor system: exploring patterns of associations in autism spectrum disorder (ASD). *Journal of Autism and Developmental Disorders, 47*(1), 155-162. https://doi.org/10.1007/s10803-016-2934-y

Most, T., Amir, O., & Tobin, Y. (2000). The Hebrew vowel system: Raw and normalized acoustic data. *Language and Speech, 43*(3), 295-308. https://doi.org/10.1177/00238309000430030401

Naigles, L. (Ed.). (2016). *Innovative investigations of language in autism spectrum disorder*. Berlin, Germany: Walter de Gruyter GmbH.

Neumeyer, V., Harrington, J., & Draxler, C. (2010). An acoustic analysis of the vowel space in young and old cochlear-implant speakers. *Clinical Linguistics & Phonetics,* *24*(9), 734-741. https://doi.org/10.3109/02699206.2010.491173

Ozonoff, S., Boodlin-Jones, B., & Solomon, M. (2005). Evidence-based assessment of autism spectrum disorder in children and adolescents, *Journal of Clinical Child and Adolescent Psychology, 34*(3), 523–540. https://doi.org/10.1207/s15374424jccp3403_8

Pang, E. W., Valica, T., MacDonald, M. J., Taylor, M. J., Brian, J., Lerch, J. P., & Anagnostou, E. (2016). Abnormal brain dynamics underlie speech production in children with autism spectrum disorder. *Autism Research*, *9*(2), 249-261. https://doi.org/10.1002/aur.1526

Peng, G., & Chen, F. (2020). Speech Development in Mandarin-Speaking Children. In H.-M. Liu, F.-M. Tsao, & P. Li (Eds.), *Speech perception, production and acquisition* (pp. 219-242). Singapore: Springer. https://doi.org/10.1007/978-981-15-7606-5_12

R Core Team. (2019). *R: A language and environment for statistical computing*. Vienna, AT: R Foundation for Statistical Computing.

Rapin, I., Dunn, M. A., Allen, D. A., Stevens, M. C., & Fein, D. (2009). Subtypes of language disorders in school-age children with autism. *Developmental Neuropsychology, 34*(1), 66–84. https://doi.org/10.1080/87565640802564648

Schoen, E., Paul, R., & Chawarska, K. (2011). Phonology and vocal behavior in toddlers with autism spectrum disorders. *Autism Research, 4*(3), 177–188.

https://doi.org/10.1002/aur.183

Schopler, E., Lansing, M. D., Reichler, R. J., & Marcus, L. M. (2005). *Examiner's manual of psychoeducational profile* (3rd ed.). Austin, TX: Pro-ed Incorporation.

Shi, F., & Wen, B. Y. (2007). Vowel development of Mandarin-speaking children. *Studies of the Chinese Language, 5,* 444–454.

Shriberg, L. D., Paul, R., Black, L. M., & van Santen, J. P. (2011). The hypothesis of apraxia of speech in children with autism spectrum disorder. *Journal of Autism and Developmental Disorders, 41*(4), 405–426. https://doi.org/10.1007/s10803-010-1117-5

Shriberg, L. D., Paul, R., McSweeny, J. L., Klin, A., Cohen, D. J., & Volkmar, F. R. (2001). Speech and prosody characteristics of adolescents and adults with high-functioning autism and Asperger syndrome. *Journal of Speech, Language, and Hearing Research, 44*(5), 1097–1115. https://doi.org/10.1044/1092-4388(2001/087)

Si, Y. Y. (2006). Mandarin phonological acquisition: A case study. *Contemporary Linguistics*, 2006(1), 1-16.

Song, J. Y., Demuth, K., & Morgan, J. (2010). Effects of the acoustic properties of infant-directed speech on infant word recognition. *The Journal of the Acoustical Society of America, 128*(1), 389–400. https://doi.org/10.1121/1.3419786

Strand, E. A., Duffy, J. R., Clark, H. M., & Josephs, K. (2014). The Apraxia of Speech

Rating Scale: a tool for diagnosis and description of apraxia of speech. *Journal of Communication Disorders, 51*, 43–50. https://doi.org/10.1016/j.jcomdis.2014.06.008

Tager-Flusberg, H., Paul, R., & Lord, C. (2005). Language and communication in autism. In D. Cohen & F. Volkmar (Eds.), *Handbook of autism and pervasive developmental disorders* (3rd ed., pp. 335–364). New York, NY: Wiley.

Walley, A. C., & Carrel, T. D. (1983). Onset spectra and formant transitions in the adult's and child's perception of place of articulation in stop consonants. *The Journal of the Acoustical Society of America, 73*, 1011-1022. https://doi.org/10.1121/1.389149

Wang, X. Y., Wang, S. P., Fan, Y. B., Huang, D., & Zhang, Y. (2017). Speech-specific categorical perception deficit in autism: An event-related potential study of lexical tone processing in Mandarin-speaking children. *Scientific Reports*, *7*, Article 43254. https://doi.org/10.1038/srep43254

Wang, Y. J., & Shangguan, X. N. (2004). How Japanese learners of Chinese process the aspirated and unaspirated consonants in standard Chinese. *Chinese Teaching in the World, 3,* 54–66.

Whitehill, T. L. (2010). Studies of Chinese speakers with dysarthria: Informing theoretical models. *Folia Phoniatrica et Logopaedica, 62*(3), 92–96.

Wolk, L., & Brennan, C. (2013). Phonological investigation of speech sound errors in children with autism spectrum disorder. *Speech, Language and Hearing, 16*(4), 239–246. https://doi.org/10.1179/2050572813Y.0000000020

Wolk, L., Edwards, M. L., & Brennan, C. (2016). Phonological difficulties in children with autism: An overview. *Speech, Language and Hearing, 19*(2), 121–129. https://doi.org/10.1080/2050571X.2015.1133488

Wolk, L., & Giesen, J. (2000). A phonological investigation of four siblings with childhood autism. *Journal of Communication Disorders, 33*(5), 371–389. https://doi.org/10.1016/S0021-9924(00)00021-6

Wong, E., Lee, K., & Tong, M. (2020). The applicability of the clinical features of English childhood apraxia of speech to Cantonese: A modified Delphi survey. *American Journal of Speech-Language Pathology, 29*(2), 652–663. https://doi.org/10.1044/2019_AJSLP-19-00118

Wu, H., Lu, F., Yu, B., & Liu, Q. (2020). Phonological acquisition and development in Putonghua-speaking children with autism spectrum disorders. *Clinical Linguistics & Phonetics*, *34*(9), 844-860. https://doi.org/10.1080/02699206.2019.1702720

Xie, H. (2009). Consonant development of Mandarin-speaking children [Master's thesis, Hunan University]. China Master's Theses Full-text Database. https://eng.oversea.cnki.net/kns55/brief/result.aspx?dbPrefix=CMFD

Yang, J. (2018). Development of stop consonants in three- to six-year-old Mandarin-speaking children. *Journal of Child Language, 45*(5), 1091-1115.

https://doi.org/10.1017/S0305000918000090

You, R. S., Serniclaes, W., Rider, D., & Chabane, N. (2017). On the nature of the speech perception deficits in children with autism spectrum disorders. *Research in Developmental Disabilities*, *61*, 158-171. https://doi.org/10.1016/j.ridd.2016.12.009

Yu, L., Fan, Y., Deng, Z., Huang, D., Wang, S., & Zhang, Y. (2015). Pitch processing in tonal-language-speaking children with autism: An event-related potential study. *Journal of Autism and Developmental Disorders, 45*(11), 3656-3667. https://doi.org/10.1007/s10803-015-2510-x

Yu, L., Zhu, X., Shek, D. T. L., Zou, X. B., Deng, H. Z., & Au Yeung, P. W. H. (2019). Validation of the simplified Chinese psychoeducational profile third edition in Mainland China. *Journal of Autism and Developmental Disorders, 49*, 1599-1612. https://doi.org/10.1007/s10803-018-3827-z

Zhou, H., Xu, X., Yan, W., Zou, X., Wu, L., Luo, X., Li, T., Huang, Y., Guan, H., Chen, X., Mao, M., Xia, K., Zhang, L., Li. E., Ge, X., Zhang, L., Li, C., Zhang, X., Zhou, Y., ... & LATENT-NHC Study Team (2020). Prevalence of autism spectrum disorder in China: A nationwide multi-center population-based study among children aged 6 to 12 years. *Neuroscience Bulletin*, *36*, 961–971. https://doi.org/10.1007/s12264-020-00530-6

Zhu, H. (2002). *Phonological development in specific contexts——Studies of Chinese-speaking children*. Bristol, UK: Multilingual Matters.

Zhu, H., & Dodd, B. (2000). The phonological acquisition of Putonghua (Modern Standard Chinese). *Journal of Child Language, 27*(1), 3–42. https://doi.org/10.1017/s030500099900402x

Zwicker, E., & Terhardt, E. (1980). Analytical expressions for critical-band rate and critical bandwidth as a function of frequency. *The Journal of the Acoustical Society of America, 68*(5), 1523-1525. https://doi.org/10.1121/1.385079

Appendix A. Chronological age and standard scores of PEP-3 subtests of each child with ASD.

Subject	Age	Communication			Motor			Maladaptive behaviour			
		CVP	EL	RL	FM	GM	VMI	AE	SR	CMB	CVB
1	4.08	12	15	13	10	7	8	12	11	12	6
2	4.67	13	15	14	13	13	13	12	12	13	13
3	7.00	7	6	7	9	11	8	10	8	10	4
4	5.17	11	13	12	12	13	12	11	11	14	9
5	4.17	9	8	9	10	11	12	10	13	11	4
6	4.92	3	12	12	1	1	3	5	5	1	3
7	6.00	7	7	7	5	2	6	7	7	8	7
8	4.50	3	13	13	1	1	3	5	5	1	3
9	3.67	13	15	15	12	12	11	12	13	14	10
10	5.00	9	10	11	9	8	10	9	10	9	7
11	3.33	15	14	14	12	12	13	13	13	14	7
12	3.33	9	11	11	9	8	11	10	10	9	7
13	4.25	15	15	15	13	12	12	12	14	15	12
14	4.75	9	11	11	8	9	11	9	9	10	6
15	4.00	9	10	9	6	6	9	10	8	11	5
16	4.50	11	13	13	12	14	11	12	11	14	12
17	5.00	12	11	13	12	12	11	12	10	11	10
18	4.85	10	10	12	8	10	11	10	11	12	5
19	4.13	5	7	8	5	6	6	10	7	9	5
20	3.56	9	11	11	8	9	11	10	9	10	6
21	6.38	3	5	4	1	1	2	4	4	1	3
22	5.77	6	6	8	10	9	5	7	9	6	3
23	5.33	8	10	10	11	12	9	12	12	9	8

24	6.58	8	8	9	11	9	11	10	10	9	4
25	3.53	8	9	11	7	5	12	7	6	5	5
26	5.96	11	11	11	11	13	12	13	13	13	10
27	4.10	6	7	8	5	2	9	5	5	2	3
Mean (SD)	4.76 (1.01)	8.93 (3.33)	10.78 (3.13)	10.96 (2.84)	8.57 (3.79)	8.65 (4.12)	9.04 (3.34)	9.74 (2.54)	9.65 (2.76)	9.74 (4.13)	6.74 (3.05)
Range	3.33-7	3-15	5-15	4-15	1-13	1-14	2-13	4-13	4-14	1-15	3-13

Note. ASD = Autism Spectrum Disorder, PEP-3 = Psychoeducational Profile-Third Edition, CVP = cognitive verbal/preverbal, EL = expressive language, RL = receptive language, FM = fine motor, GM = gross motor, VMI = visual-motor imitation, AE = affective expression, SR = social reciprocity, CMB = characteristic motor behaviour, CVB = characteristic verbal behaviour.

Appendix B. Chronological age of each typically developing child.

Subject	Age	Subject	Age	Subject	Age
1	3.42	11	4.25	21	5.67
2	3.42	12	3.92	22	4.17
3	3.92	13	4.00	23	3.83
4	4.00	14	5.92	24	4.50
5	4.25	15	5.92	25	4.00
6	4.25	16	4.00	26	3.92
7	5.33	17	5.50	27	4.00
8	5.67	18	5.50	28	4.17
9	5.33	19	3.33	29	4.25
10	7.00	20	4.17	30	3.92