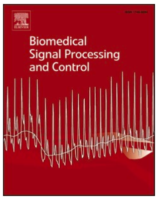
Contents lists available at ScienceDirect

Biomedical Signal Processing and Control

journal homepage: www.elsevier.com/locate/bspc



Applying Random Forest classification to diagnose autism using acoustical voice-quality parameters during lexical tone production

Chengyu Guo [a], Fei Chen [a,*], Yajie Chang [a], Jinting Yan [a,b]

[a] School of Foreign Languages, Hunan University, Changsha, China
[b] Cangzhou Research Centre for Child Language Rehabilitation, Cangzhou Normal University, Hebei, China

ARTICLE INFO

Keywords:
Autism
Random Forest classification
Voice quality
Lexical tone production

ABSTRACT

Atypical voice quality has been reported among children with autism spectrum disorder (ASD). Yet, it is unclear which acoustic parameters played a crucial role in discriminating the voice of children with ASD from that of their typically developing (TD) peers, especially those who speak a tone language. The current study carried out a preliminary investigation of voice quality in Mandarin-speaking children with ASD using multidimensional acoustic parameters, in an effort to seek the most robust cues using the Random Forest classification. Twenty Mandarin-speaking children with ASD and twenty age-matched TD children participated in the lexical tone production using a picture-naming task. Acoustic parameters included in this study were time-domain parameters: fundamental frequency (F0), the range of F0, the strength of excitation; spectral parameters: H1*-H2*, H2*-H4*, H1*-A1*, H1*-A2*, H1*-A3*; and signal aperiodicity parameters: cepstral peak prominence, harmonic-to-noise ratio, subharmonic-to-harmonic ratio, jitter, and shimmer. Results showed that except for HNR and F0 range, group differences (ASD vs. TD) were found in the other 11 parameters. Additionally, a 78.5% accuracy rate was obtained for classification analysis between voice-quality features of children with and without ASD, with shimmer and jitter as robust parameters. These results indicated that Mandarin-speaking children with ASD tended to overexert and overstrained their voices. Especially for Tone 3 production, they notably exhibited a higher F0 with a less creaky voice, losing the typical voice-quality feature of T3. Although no voice disorders were detected among Mandarin-speaking children with ASD, voice quality has the potential and supplementary value for diagnosing ASD.

1. Introduction

Autism Spectrum Disorder (ASD) could be deemed as a group of neurodevelopment disabilities characterized by significant difficulties in social communications [1], as well as restricted, repetitive patterns of behavior [2]. Individuals with ASD exhibit language and communicative impairments compared with their typically developing (TD) peers [3]. For instance, atypical voice quality among individuals with ASD was found in the speech of non-tonal language speakers [4–6]. However, it remains unknown about the voice quality of autistic individuals who speak a tone language in which the voice quality plays a crucial role in discriminating the tonal categories. This study provides an investigation of voice quality with multiple parameters in Mandarin-speaking children with ASD.

The prosody (e.g., stress, intonation, and rhythm) among individuals with ASD has been variously described as monotonous, robot-like [7], dull, wooden, sing-songy [4], over-exaggerated, stilted [8], hoarse, and harsh [9]. These subjective descriptions may reflect atypical vocal characteristics among autistic individuals, resulting in the negative perceptual assessment of listeners and eventually hampering social communication and interaction [10–12]. Especially for children with ASD, the communicative barrier probably exerts a long-termed and irreversible impact on the development of social ability [13]. Clinically, prosodic atypicality also has been listed as one of the diagnostic criteria in the Autism Diagnostic Observation Schedule, 2nd Edition (ADOS-2 [14]).

The most frequently used acoustic parameter was fundamental frequency (F0) when focusing on the prosody in ASD [6,11,12,15–17]. Higher F0 and greater F0 variations (standard deviation or range) were characterized as prosodic features of ASD in the literature. Recently, the strength of excitation (SoE) was also adopted in acoustic analysis for ASD, generally indicating the strength in voicing [18]. Similarly, higher

[*] Corresponding author at: Lushan South Road, Yuelu District, Changsha, Hunan Province Zip Code: 410082, China.
E-mail address: chenfeianthony@gmail.com (F. Chen).

https://doi.org/10.1016/j.bspc.2022.103811
Received 28 February 2022; Received in revised form 20 April 2022; Accepted 7 May 2022

SoE was found among individuals with ASD [19].

The voice quality is also quantified in terms of the spectral tilt [20–22]. To be more specific, the spectral tilt could be manifested by the amplitude difference between the first two harmonics in dB (H1-H2), and between the second and the fourth harmonics (H2-H4). Besides, the spectral tilt was also measured with an amplitude difference between the first harmonic and the harmonic nearest the first formant (H1-A1), the second formant (H1-A2), and the third formant (H1-A3), respectively [21]. A significantly lower spectrum tilt was found in the individuals with ASD [17], and it was reported that the value of "H1-A3" might be the robust distinctive parameter as spectrum tilt for toddlers with ASD [23].

Furthermore, signal periodicity parameters have been treated as important acoustic measurements in clinical assessment. The cepstral peak prominence (CPP), for instance, has been regarded as one of the measurements of dysphonia in recent years [24,25]. Lower CPP values indicated more noise perturbations in the vowel [26]. Additionally, the harmonic-to-noise ratio (HNR) indicates the amplitude relationship between the harmonic and noisy parts of a speech signal, lower HNR is associated with more noise [27,28]. Subharmonic-to-harmonic ratio (SHR) represents the amplitude ratio between subharmonics and harmonics [29], with higher SHR suggesting increased noise. Note that the noise perturbation from the vocal source might be caused by special phonation types such as creaky and breathy voices [30]. Furthermore, jitter and shimmer were recommended in non-invasive voice assessment by the European Laryngological Society [31]. Jitter and shimmer refer to the cycle-to-cycle perturbation in F0 and amplitude, respectively [32].

Among these parameters, higher absolute jitter was found in the spontaneous speech of autistic children, and higher jitter contributed to the "overall severity" of voice [33], measured by the Consensus Auditory-Perceptual Evaluation of Voice (CAPE-V [34]). Furthermore, it was reported that children with higher autism severity scores exhibited significantly increased variabilities of CPP, HNR, and jitter [5]. Although the distinctive voice quality has been reported, the consistent and quantitative evidence for ASD is still "extremely sparse" [16]. To the best of our knowledge, there is a paucity of research on the voice quality of tone-language-speaking children with ASD. Then, the question arises whether and how the voice quality of autistic children who speak a tone language is atypical compared with TD peers.

In tone languages like Mandarin Chinese, voice quality plays an indispensable role in discriminating lexical tones. The Mandarin tonal inventory includes four lexical tones: Tone 1 (T1, high-level tone), Tone 2 (T2, mid-rising tone), Tone 3 (T3, low-falling-rising tone), Tone 4 (T4, high-falling tone) [35], as shown in Fig. 1. Except for the high-level T1, other tones (T2, T3, and T4) are contour tones that have notable pitch varying across the whole syllable over time. Moreover, the creaky voice is linked to lexical tones with a relatively low pitch such as T3 and T4 (only at the end of vowels) [36]. It was believed that the creaky voice could act as the enhancement cue in discriminating T3 from T2, apart from the F0 [37,38].

Additionally, prosodic positions in Mandarin words may also exhibit different voice qualities. Specifically for each lexical tone, the final syllables showed a higher spectral tilt than non-final syllables [39]. Another tonal phenomenon concerning the prosodic position is T3 sandhi which only occurs in the non-final position [35]. In Mandarin disyllabic words with underlying T3-T1/T2/T4, the preceding T3 undergoes tone sandhi, realized as a low falling tone (namely T3 half-sandhi). However, when followed by another T3 (T3-T3), the preceding T3 changes to a rising tone perceptually close to T2 (namely T3 full-sandhi). Pitch contours of T3 half- and full-sandhi are also shown in Fig. 1. Still, the voice quality of T3 half-sandhi/full-sandhi remained unrevealed in literature.

To date, studies concerning voice quality have applied automatic classification analysis to predict the difference in voice features (e.g., [19,40]). Random Forest classification, for example, is an ensemble learning method for classification [41,42]. This classification algorithm makes it possible for us to understand the possibility of using acoustical voice-quality parameters to diagnose ASD, and which parameter matters most.

In a nutshell, the current study aimed to systematically explore the voice quality of lexical tone productions by Mandarin-speaking children with and without ASD via quantitative measures. We tried to answer the following research questions: (a) Are there any atypical voice-quality features in Mandarin-speaking children with ASD during lexical tone production? (b) If so, would certain Mandarin lexical tone (T1, T2, T3, and T4) or prosodic position [the first syllable (S1) and the second syllable (S2)] affect the voice-quality features in ASD? (c) Which are the relatively robust cues for diagnosing ASD?

2. Materials and methods

2.1. Participants

Twenty Mandarin-speaking children with ASD aged from 3;11 to 8;6 (year; month, M_{age} = 5;11, SD_{age} = 14 months, 17 females, 3 males)

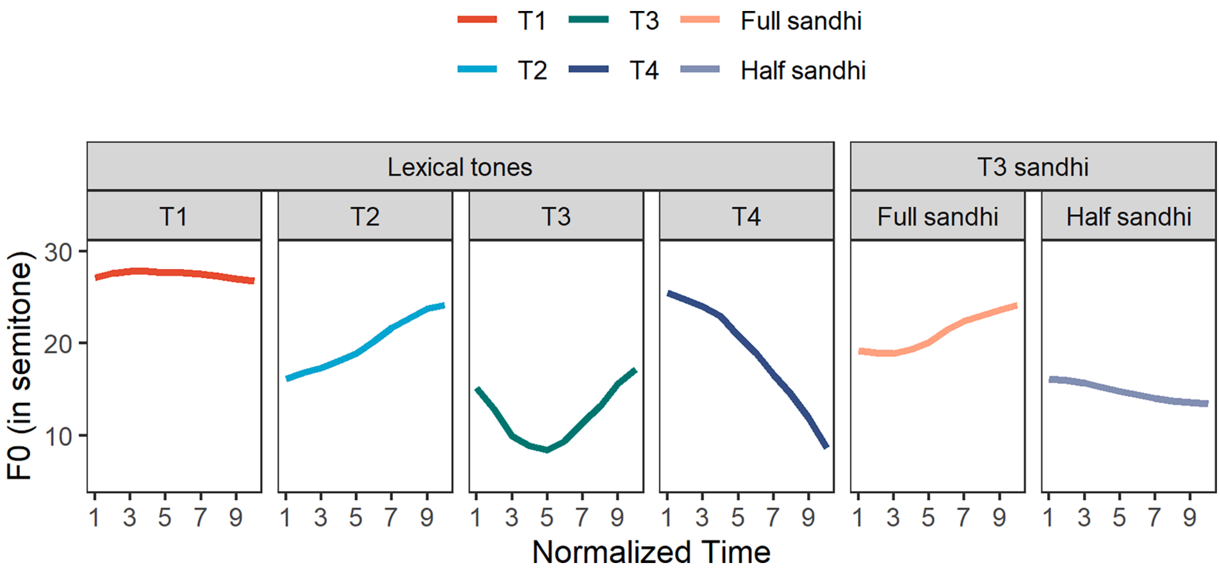


Fig. 1. Four lexical tones and two T3 sandhi forms (full-sandhi vs. half-sandhi) in the Mandarin tonal inventory.

were recruited from Cangzhou Research Centre for Child Language Rehabilitation. Correspondingly, twenty age-matched TD peers aged from 3;11 to 8;6 ($M_{\text{age}} = 5;12$, $SD_{\text{age}} = 14$ months, 17 females, 3 males) from Tuofu Kindergarten and Jiqing Primary School were recruited as the control group. All participants were born and raised in Cangzhou city, and they had no congenital hearing loss or related diseases. The parents of all child participants provided informed, written permission for their children to participate. Furthermore, this research was approved by the Research Ethics Committee of Cangzhou Normal University to ensure proper compliance with the informed consent procedure.

In the current study, the autistic participants had already been diagnosed with ASD clinically based on the DSM-5 [2] and ADOS-2 [14] by the pediatricians and child psychiatrists in local hospitals. Before the formal experiment, all participants took part in the test of language ability [43] and nonverbal intelligence [44]. Results showed that the children with ASD had lower language scores [$t\ (20) = -8.87, p < .001$] and nonverbal IQ scores [$t\ (37) = -8.41, p < .001$], compared with TD children.

2.2. Materials

Given that different lexical tones and prosodic positions in Mandarin are linked to acoustic realizations, we chose the stimuli containing the combination of four lexical tones (T1, T2, T3, and T4) at two syllabic positions (S1 and S2). Each combination included three disyllabic words such as toys, animals, fruits, and daily necessities, with 27 items in total (see Appendix A). Additionally, each combination generally contained both high and low vowels (i.e., [i] and [ɑ]). For lexical items with diphthongs, we mainly focused on the vowel segment of the nucleus. (i.e., low vowel [ɑ] in [ɑi] and [ɑu]).

Due to the tone sandhi of T3 in the S1, three words of T3 half-sandhi and full-sandhi were respectively selected. Note that the T3 half-sandhi [31] is a low-falling tone, and the T3 full-sandhi [45] is a rising tone. The T3 in S2, however, is uniformly realized as a dipping tone. Moreover, both syllables of the target words of [tɕʰi^{21} ɤ^{35}] (“penguin”), [lɑu^{35} xu^{214}] (“tiger”), and [tɕi^{55} mu^{51}] (“toy blocks”) were analyzed as different items. Therefore, the actual number of target words was 24.

2.3. Procedure

The tonal production was carried out in a sound-isolated room at the Cangzhou Research Centre for Child Language Rehabilitation. During the experiment, the speech sounds of participants were recorded by an external microphone (Shure MV51). The microphone was connected to a computer via the USB audio interface with a sampling rate modulated to 44,100 Hz.

A picture-naming task was carried out to elicit target lexical items. Twenty-four pictures corresponding to target words were presented randomly via the mobile application for both Android- and iOS-based platforms (c.f. [46]). All children were asked to produce the target words. During the experimental session, the Mandarin-speaking experimenter pointed to the picture, and asked the participant, “What is this?” without any hint of picture names. If the participant successfully produced the target words, the experimenter would guide them to repeat the word three times, and record their speech sounds. However, if the child participant cannot utter the lexical item, the experimenter would repeat the question. Occasionally, the participant still could not produce the target word after reconsideration, then the experimenter would continue this session by showing the next item picture. The ASD group uttered 21.60 out of 24 pictures ($SD = 2.85$) on average, and the elicited picture number was not different between ASD and TD groups [$t\ (25) = -1.31, p = 0.203$].

After data collection, based on the perceptual judgment, two trained phoneticians chose the best, accurate pronunciation (out of 3 times) as the representative token of each lexical item (Cohen’s kappa = 0.65). Since the speech sounds of children with ASD might not be stable due to their attention deficit, the best pronunciation for each item could represent the highest production ability, and was included in the subsequent data analysis. An example of a representative token from a Mandarin-speaking autistic child and a TD child is shown in Appendix B.

2.4. Measurement and parameters

To obtain reliable voice-quality data, we manually identified the vowel segment in each token and generated TextGrid files by Praat [47], based on the onset and offset of the second formant in the spectrogram [48].

The voice-quality features were acoustically measured by multidimensional parameters such as time-domain parameters, spectral parameters, and aperiodicity [20]. In the current study, specifically, we investigated F0, F0 range, strength of excitation (time-domain); H1*-H2*, H2*-H4*, H1*-A1*, H1*-A2*, H1*-A3* (spectral domain); CPP, HNR25, SHR, jitter, and shimmer (signal aperiodicity). Except for jitter and shimmer, the values of the other 10 parameters are time-varying, which could be extracted using the VoiceSauce [49]. Values of local jitter (%) and local shimmer (%) were extracted from each annotated vowel using Praat [47].

To be more specific, fundamental frequency (F0) was calculated using the STRAIGHT algorithm [50]. Given the relatively high F0 values produced by young children, we set the F0 range from 100 to 700 Hz during pitch tracking. Besides, all the raw F0 values (in Hz) were transformed into semitones (ST) with 100 Hz as the reference value for children [51]. It should be noted that we obtained the values of F0 and formants data first, then manually checked and corrected errors for further parameter extractions. For the spectral parameter, we selected the corrected values (e.g., H1*–H2*and H1*-A1*) based on the formant information [52]. Moreover, the HNR25 means the value measured from the frequency band of 0–2500 Hz, which is a common-used value within HNRs (henceforth HNR) (cf. [40]).

After choosing the parameters of voice quality, all data was automatically obtained from the VoiceSauce with a window size of 25 ms [49]. Then we normalized the data length by selected values at 11 equidistant time-points over each annotated vowel. To avoid irrelevant data fluctuation, the values at the first and last points were discarded, allowing 9 time-point data submitted to the statistical analysis.

2.5. Statistical analysis

The statistical analysis was conducted in R [53]. Firstly, to analyze group differences in each voice-quality parameter, a linear mixed-effect model (LMM) was conducted by using the lme4 package [54]. In each LMM, fixed factors were *Group* (ASD, TD), *Syllable* (S1, S2), *Tone* (T1, T2, T3, T4), and their interaction (including two-way and three-way interactions). Additionally, random factors were *Subject* and *Lexical item*, and random slope and intercept were incorporated to make it maximally generalizable across the data [55]. Therefore, for each parameter, the R code for the full model was below:

$$Parameter \sim Group*Syllable*Tone+(1+Group+Syllable+Tone|Subject)+(1+Group+Tone|Item).$$

In total, we built 13 LMMs for all thirteen voice-quality parameters. When fitting all models, the chi-square value and significance (*p*-value) of the main or interaction effect were generated with the likelihood ratio test, using the afex package [56]. If a significant *p*-value was detected in each LMM, post-hoc pairwise comparisons were performed using the lsmeans package with Tukey adjustment [57].

Secondly, after finding the voice-quality parameters discriminating between ASD and TD groups, we estimated the relative contribution among these parameters by conducting the Random Forest classification analysis [41], using the randomForest package [42]. Then, all data were randomly separated into training and testing samples, which were compiled with the machine learning methodology. To minimize the out-of-bag (OOB) prediction error, we sought the optimal number of decision trees (“ntree” in the randomForest package), and the size of the random features to split a node (“mtry” in the randomForest package) when growing the trees [58].

3. Results

The mean values and standard deviations of all parameters are listed in Table 1. Overall, except for the F0 range and HNR, the LMM results of other parameters showed a significant main effect of Group, or significant interaction effects such as *Group* × *Tone*, *Group* × *Syllable*, and *Group* × *Syllable* × *Tone* (see Table 2). More specific results of post-hoc pairwise comparisons between groups in each LMM are shown in Fig. 2 (time-domain and spectral parameters), and Fig. 3 (aperiodicity parameters).

3.1. Analysis of time-domain parameters

F0 The statistical result suggested a main effect of *Group* [$\chi^2(1) = 12.28$, $p < .001$], and an interaction effect of *Group* × *Syllable* [$\chi^2(1) = 10.37$, $p < .01$]. Results of post-hoc pairwise comparison (Fig. 2) showed that ASD group exhibited higher F0 pattern than TD group in both S1(β = 1.03, *SE* = 0.46, *t* = 2.24, $p < .05$) and S2 (β = 2.47, *SE* = 0.57, *t* = 4.35, $p < .001$).

F0 range The result of the F0 range, however, did reach any significant differences between the two testing groups (*ps* > 0.05; see Table 2).

SoE The analysis only showed a main effect of *Group* [$\chi^2(1) = 14.76$, $p < .001$]. Post-hoc pairwise comparison indicated that ASD group exhibited higher SoE values than TD group (β = 0.02, *SE* = 0.01, *t* = 3.83, $p < .001$).

3.2. Analysis of spectral parameters

H1*–H2* A main effect of Group was found in analysis of H1*–H2* [$\chi^2(1) = 14.91$, $p < .001$]. Specifically, the children with ASD showed lower H1*–H2* values compared with their TD peers (*β* = -3.03, *SE* = 0.79, *t* = -3.82, $p < .001$).

H2*-H4* There was a main effect of Group in the LMM on H2*-H4* [$\chi^2(1) = 9.22$, $p < .01$]. Post-hoc pairwise analysis indicated that the ASD group had lower H2*-H4* values that TD group (*β* = -3.53, *SE* = 1.21, *t* = -2.92, $p < .01$).

H1*-A1* The analysis consistently yielded a main effect of Group on H1*-A1* [$\chi^2(1) = 14.55$, $p < .001$]. Similarly, the H1*-A1* values of ASD group were significantly lower than that of TD group (*β* = -4.60, *SE* = 1.25, *t* = -3.68, $p < .001$).

H1*-A2* The significant main effect of Group was found in the LMM on H1*-A2* [$\chi^2(1) = 34.13$, $p < .001$]. Relative to the TD group, the ASD group similarly exhibited lower values of H1*-A2* (*β* = -8.22, *SE* = 1.35, *t* = -6.10, $p < .001$).

H1*-A3* The result of the LMM exhibited a main effect of Group on H1*-A3* [$\chi^2(1) = 22.99$, $p < .001$]. Post-hoc pairwise analysis showed lower H1*-A3* values among children with ASD (*β* = -9.35, *SE* = 1.94, *t* = -4.83, $p < .001$).

Table 1
The mean values and standard deviations (in brackets) of 13 voice-quality parameters measured from disyllabic-word production by Mandarin-speaking children with autism spectrum disorder and typically developing children.

Group	Syllable	Tone	F0 (ST)	F0 range (ST)	SoE (%)	H1*–H2* (dB)	H2*-H4* (dB)	H1*-A1* (dB)	H1*-A2* (dB)	H1*-A3* (dB)	CPP (dB)	HNR (dB)	SHR (%)	Jitter (%)	Shimmer (%)
ASD	S1	T1	21.07	1.42	12.55	0.93	−6.90	7.08	7.76	−5.44	22.06	44.57	1.59	0.74	4.87
			(2.90)	(1.35)	(4.72)	(8.17)	(10.61)	(7.34)	(9.74)	(15.90)	(3.36)	(10.56)	(7.59)	(0.51)	(2.97)
		T2	18.26	1.76	11.00	3.55	−1.94	15.02	11.22	3.20	21.31	39.43	9.17	0.87	5.97
			(3.47)	(1.79)	(4.67)	(6.92)	(10.23)	(7.55)	(9.37)	(11.91)	(3.79)	(10.45)	(23.49)	(0.44)	(3.34)
		T3	18.44	1.86	10.75	2.10	−2.16	11.10	10.35	−0.31	21.82	42.12	10.53	0.94	5.39
			(3.94)	(1.83)	(4.84)	(8.10)	(9.78)	(8.34)	(10.09)	(13.97)	(3.91)	(11.24)	(26.63)	(0.59)	(2.63)
		T4	21.17	2.55	11.96	−0.44	−4.65	9.64	6.57	−3.05	21.79	36.06	3.71	1.01	5.10
			(3.56)	(2.05)	(4.78)	(7.62)	(11.88)	(8.77)	(10.53)	(14.81)	(3.77)	(11.74)	(13.43)	(0.66)	(2.37)
	S2	T1	21.40	1.59	13.44	0.26	−2.77	8.46	8.89	−0.99	22.89	40.88	3.34	0.68	4.71
			(2.96)	(1.70)	(4.55)	(8.62)	(10.60)	(8.62)	(10.15)	(18.29)	(3.92)	(11.32)	(15.60)	(0.42)	(2.34)
		T2	18.91	2.48	9.80	3.99	−2.28	12.47	11.67	−0.64	22.48	38.85	10.14	0.91	5.37
			(4.34)	(2.07)	(4.40)	(7.58)	(10.03)	(7.27)	(7.72)	(14.06)	(4.26)	(10.49)	(25.46)	(0.60)	(2.49)
		T3	16.54	1.24	9.85	4.26	−3.05	14.51	12.96	−0.40	21.93	44.25	16.95	0.94	5.10
			(5.07)	(1.27)	(4.50)	(6.75)	(11.05)	(8.11)	(9.60)	(15.83)	(4.11)	(10.12)	(31.30)	(0.95)	(1.79)
		T4	20.68	2.49	10.32	2.41	−1.64	14.04	11.08	2.03	22.79	34.81	5.93	0.98	5.79
			(4.04)	(1.87)	(5.10)	(7.46)	(13.47)	(8.86)	(9.81)	(14.17)	(3.85)	(9.39)	(18.29)	(0.47)	(2.94)
TD	S1	T1	20.50	2.10	11.82	2.67	−6.64	8.78	12.96	−1.84	21.94	40.51	2.85	0.53	5.53
			(3.58)	(4.52)	(5.21)	(9.35)	(12.18)	(10.58)	(10.23)	(15.46)	(3.20)	(13.29)	(13.55)	(0.36)	(4.40)
		T2	17.42	1.67	8.46	5.00	3.54	18.18	19.54	12.59	21.81	40.49	4.05	0.75	7.57
			(3.23)	(3.83)	(3.85)	(7.74)	(9.19)	(11.07)	(12.68)	(12.38)	(3.36)	(10.20)	(17.36)	(0.46)	(4.26)
		T3	17.26	1.45	8.65	5.80	0.65	16.16	19.73	10.35	21.00	39.18	11.81	1.12	7.63
			(3.58)	(2.38)	(4.49)	(7.55)	(9.76)	(12.04)	(11.40)	(15.71)	(3.73)	(11.06)	(28.06)	(1.37)	(4.11)
		T4	19.92	2.49	10.52	3.91	−1.10	13.43	16.25	4.07	22.37	36.30	5.52	0.93	7.62
			(3.63)	(4.65)	(5.18)	(7.65)	(12.23)	(13.23)	(12.59)	(18.85)	(2.69)	(13.69)	(18.95)	(0.45)	(3.91)
	S2	T1	19.24	1.81	11.04	4.49	−0.45	12.91	16.83	8.89	23.39	42.89	5.72	0.38	4.54
			(3.87)	(4.07)	(4.54)	(9.81)	(10.60)	(11.23)	(9.56)	(18.56)	(2.52)	(14.33)	(20.98)	(0.22)	(2.65)
		T2	17.11	2.02	6.93	6.59	2.38	17.88	19.88	10.11	22.60	41.06	12.73	0.68	5.90
			(3.77)	(2.92)	(3.79)	(5.60)	(10.40)	(8.07)	(9.16)	(12.24)	(3.10)	(10.43)	(29.83)	(0.29)	(2.84)
		T3	13.17	0.86	6.63	5.45	1.02	17.89	18.84	7.34	19.50	42.37	30.80	2.23	9.86
			(4.99)	(2.76)	(4.02)	(6.70)	(10.18)	(11.44)	(10.96)	(16.88)	(4.12)	(12.45)	(36.73)	(2.29)	(5.81)
		T4	18.55	1.77	8.45	5.39	2.75	20.29	20.11	14.89	22.95	31.94	5.44	1.04	7.44
			(3.53)	(2.20)	(3.74)	(7.56)	(12.00)	(11.94)	(12.65)	(13.28)	(2.99)	(9.81)	(19.15)	(0.68)	(4.42)

Note. ASD = autism spectrum disorder, TD = typically developing; ST = semitone; dB = decibel; F0 = fundamental frequency; SoE = strength of excitation; H_i*-H_j* = corrected amplitude difference between the (i)th harmonic and the (j)th harmonic; H_i*-A_j* = corrected amplitude difference between the (i)th harmonic and the harmonic closet to the (j)th formant; CPP = cepstral peak prominence; HNR = harmonic-to-noise ratio; SHR = subharmonic-to-harmonic ratio.

Table 2
Chi-Square results of model comparisons for the effect of Group, the two-way, and three-way interaction effects concerning the Group factor on the voice-quality parameters. Boldface indicates the significant findings.

Effect	F0	F0 range	SoE	H1*-H2*	H2*-H4*	H1*-A1*	H1*-A2*	H1*-A3*	CPP	HNR	SHR	Jitter	Shimmer
Group	**12.28*****	0.19$^{n.s.}$	**14.76*****	**14.91*****	**9.22****	**14.55*****	**34.13*****	**22.99*****	0.19$^{n.s.}$	0.87$^{n.s.}$	1.92$^{n.s.}$	0.67$^{n.s.}$	**11.16*****
Group × *Syllable*	**10.37****	0.74$^{n.s.}$	0.94$^{n.s.}$	0.00$^{n.s.}$	0.04$^{n.s.}$	0.42$^{n.s.}$	0.11$^{n.s.}$	0.53$^{n.s.}$	0.35$^{n.s.}$	0.05$^{n.s.}$	**5.21***	2.68$^{n.s.}$	0.03$^{n.s.}$
Group × *Tone*	0.77$^{n.s.}$	4.09$^{n.s.}$	2.03$^{n.s.}$	0.74$^{n.s.}$	1.38$^{n.s.}$	0.39$^{n.s.}$	1.10$^{n.s.}$	0.92$^{n.s.}$	**10.59***	2.76$^{n.s.}$	4.87$^{n.s.}$	**14.98****	**36.70*****
Group × *Syllable* × *Tone*	1.69$^{n.s}$	0.43$^{n.s.}$	1.17$^{n.s.}$	3.00$^{n.s.}$	0.40$^{n.s.}$	1.35$^{n.s.}$	2.11$^{n.s.}$	2.11$^{n.s.}$	1.25$^{n.s.}$	1.43$^{n.s.}$	7.80$^{n.s.}$	**20.11*****	**15.30****

Note. F0 = fundamental frequency; SoE = strength of excitation; H_i*-H_j* = corrected amplitude difference between the (i)th harmonic and the (j)th harmonic; H_i*-A_j* = corrected amplitude difference between the (i)th harmonic and the harmonic closet to the (j)th formant; CPP = cepstral peak prominence; HNR = harmonic-to-noise ratio; SHR = subharmonic-to-harmonic ratio.
n.s. = not significant.
* p <.05.
** p <.01.
*** p <.001.

3.3. *Analysis of signal aperiodicity*

CPP The statistical analysis of CPP showed a significant interaction effect of *Group* × *Tone* [$\chi^2(3) = 10.59$, $p < .05$]. As shown in Fig. 3, results of pairwise comparison indicated that only when producing T3, children with ASD had higher CPP values than their TD peers (β = 1.62, *SE* = 0.63, *t* = 2.58, *p* < .05). Unlike the above spectral parameters, the group difference of CPP was only confined to T3.

HNR The result showed neither a main effect of *Group* nor the group-related interaction effect (*ps* > 0.5; see Table 2), indicating that HNR (0–2500 Hz) could not discriminate the voice quality between children with and without ASD.

SHR A significant interaction effect of *Group* × *Syllable* was found on SHR [$\chi^2(1) = 5.21$, $p < .05$]. Post-hoc pairwise comparison showed that ASD group only exhibited significantly lower SHR values when producing T3 in the S2 (β = -0.14, *SE* = 0.04, *t* = -3.46, *p* < .001), compared with TD peers.

Jitter The result showed a two-way interaction effect of *Group* × *Tone* [$\chi^2(3) = 14.98$, $p < .01$], and a three-way interaction effect of *Group* × *Syllable* × *Tone* [$\chi^2(3) = 20.11$, $p < .001$]. Results of pairwise comparison suggested that ASD group had significantly lower jitter values than TD peers only when producing T3 in the S2 (β = -0.013, *SE* = 0.003, *t* = -4.144, *p* < .001).

Shimmer The analysis yielded a main effect of *Group* [$\chi^2(1) = 11.16$, $p < .001$], two-way interaction effect of *Group* × *Tone* [$\chi^2(3) = 36.70$, $p < .001$], and three-way interaction effect of *Group* × *Syllable* × *Tone* [$\chi^2(3) = 15.30$, $p < .01$]. Specifically, in terms of the S1, the ADS group had lower shimmer values when producing T2, T3, and T4; for the S2, the ADS group exhibited lower shimmer values than the TD group when producing T3 and T4 (see Appendix C).

3.4. *The Random Forest classification analysis*

Based on the above results of the linear mixed model, the HNR and F0 range were excluded due to their inapplicability to differentiate ASD and TD groups in any syllables and tonal categories, allowing the other 11 parameters to enter the analysis of Random Forest classification [41]. Firstly, we wondered about the classification accuracy (ASD vs. TD) predicted by the voice-quality parameters. Then, within the 11 parameters, we aimed at seeking the relative contributions to the classification of children with and without ASD by using the Random Forest algorithm [42].

The input data was divided into train and test subsets with a 7:3 ratio, respectively. By using the “tuneRF” function of the randomForest package (in R), we found the optimal value of 4 as the number of variables randomly sampled at each split (“mtry”) to minimize the OOB prediction error. Then we set the number of trees as 500, given that the OOB errors often stabilized before this number when growing the trees [40]. This number was further checked by the “plot” function visually. After setting “mtry” (4) and “ntree” (5 0 0), the random forest model was trained and its error rate of OOB was 22.2%.

Then we used the “confusionMatrix” function of the caret package (in R), to evaluate the classification accuracy of the model with the separate test set. Results of the confusion matrix showed that ASD and TD were predicted with an accuracy rate of 78.2% and 78.7%, respectively. Overall, the test classification accuracy of the random forest model was 78.5%. The random forest model’s sensitivity and specificity were estimated to be 76.5% and 80.2%, respectively. Moreover, the three most important predictors differentiating ASD and TD groups were Shimmer, Jitter, and H1*-A2* according to the mean decrease accuracy (see Fig. 4).

4. Discussion

This study aimed at exploring the voice quality in lexical tone production by Mandarin-speaking children with ASD using various acoustic parameters. Based on the findings, we could answer the research questions and discuss the underlying clinical implications.

4.1. *Normal (nonpathological) voice in tone-language-speaking children with ASD*

Voice-quality parameters such as HNR, jitter, and shimmer have been recommended in non-invasive voice assessment by the European Laryngological Society [31]. For the voice quality of children with ASD, various subjective descriptions in literature [4,7] motivated us to wonder whether they had voice disorders. The current study measured the 13 voice-quality parameters of Mandarin-speaking children with ASD, offering a valuable window to evaluate their voice quality.

Table 1 shows the mean and SD values of parameters between ASD and TD groups among different tones and syllables. It is noteworthy that some of their SDs were even greater than the means, such as F0 range, SHR, and all spectral parameters. For spectral parameters (e.g., H1-H2), the values varied from negative to positive numbers with a bi-modal distribution. Likewise, the SHR data were also non-normally distributed under most circumstances (c.f. [29]). Besides, the relatively high SD of the F0 range might be attributed to the tonal development in 3- to 8-year-old child participants. It was reported that tonal productions by TD children varied according to age and were not adult-like before 5 years old [51], let alone children with ASD who exhibited substantially individual differences [59]. To make it less complicated, we only discussed the grand mean and SD values for each group as the baseline

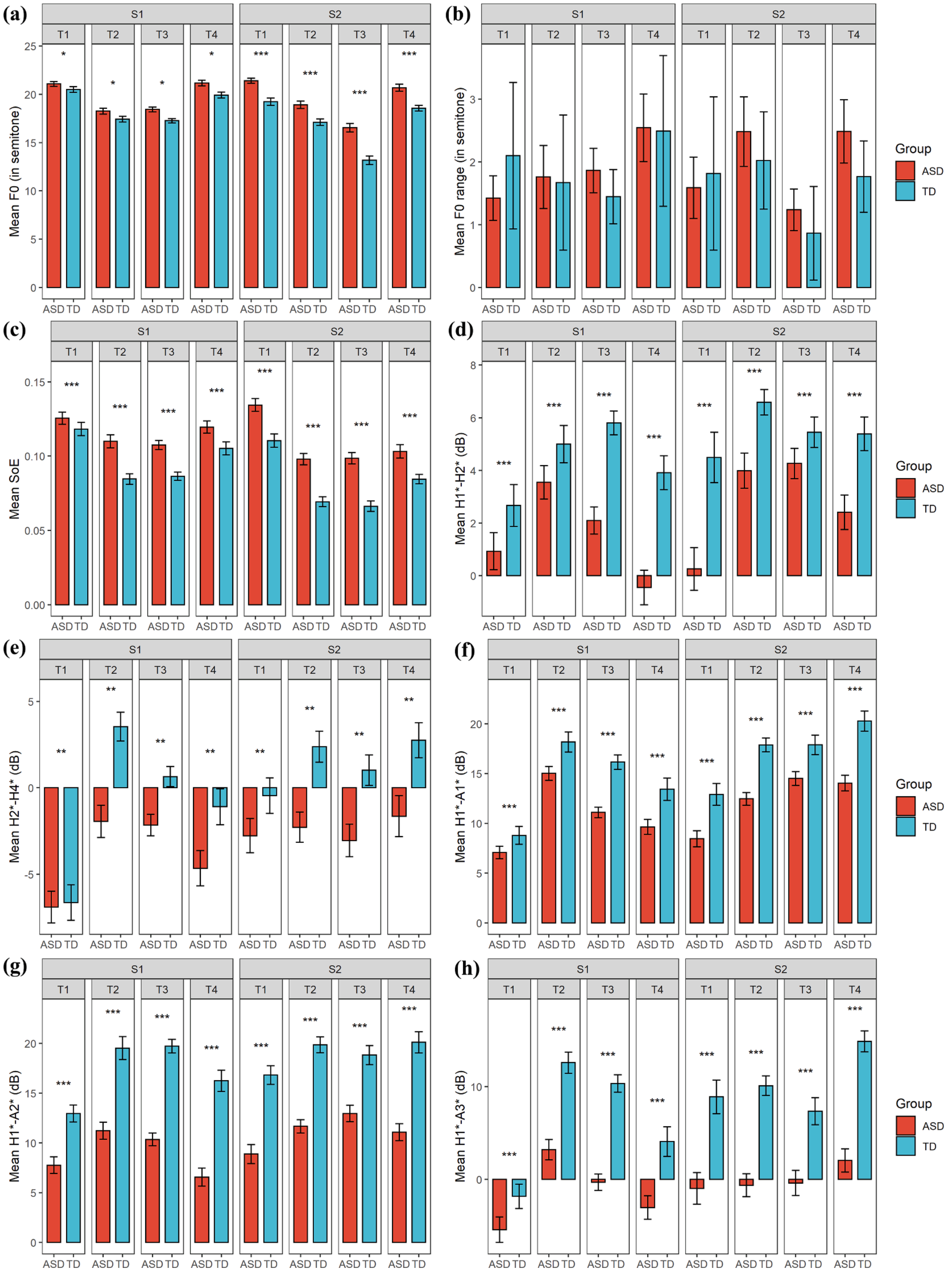


Fig. 2. The differences between ASD and TD groups in terms of F0 (a), F0 range (b), SoE (c), H1*-H2* (d), H2*-H4* (e), H1*-A1* (f), H1*-A2* (g), and H1*-A3* (h). The bars indicate the 95 % confidence interval. Asterisks indicate the significant findings in post-hoc test, *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$.

values as follows.

Values of jitter and shimmer are widely used in the automatic diagnostic tool or application of voice pathology detection [60]. Previous studies demonstrated that values of these two parameters were highly sensitive to confounding factors such as age, and gender [61,62]. It is easy to understand since the length and thickness of the vocal cord grow continually as a function of age and vary according to gender differences [63]. The empirical study indicated that values of jitter and shimmer tended to decrease with age, so that values of children likely exceeded the upper limited values of adults [62]. Therefore, it is more appropriate to compare our results with the reference values for children, rather than the standard values for adults.

In the present study, the mean jitter values of Mandarin-speaking children with and without ASD were 0.89% (*SD* = 0.6%) and 0.99% (*SD* = 1.18%), respectively. These results are generally within the typical ranges (0.27%-1.27%) in 5- to 9-years-old children without speech disorders [64]. As for shimmer, the mean values of Mandarin-speaking children with and without ASD were 5.30% (*SD* = 2.64%) and 7.14% (*SD* = 4.48), comparable with typical ranges (4.34%-14.40%) in children without speech disorders [64].

Moreover, other important diagnostic parameters for voice quality are CPP and HNR. CPP has been identified as a promising and potentially

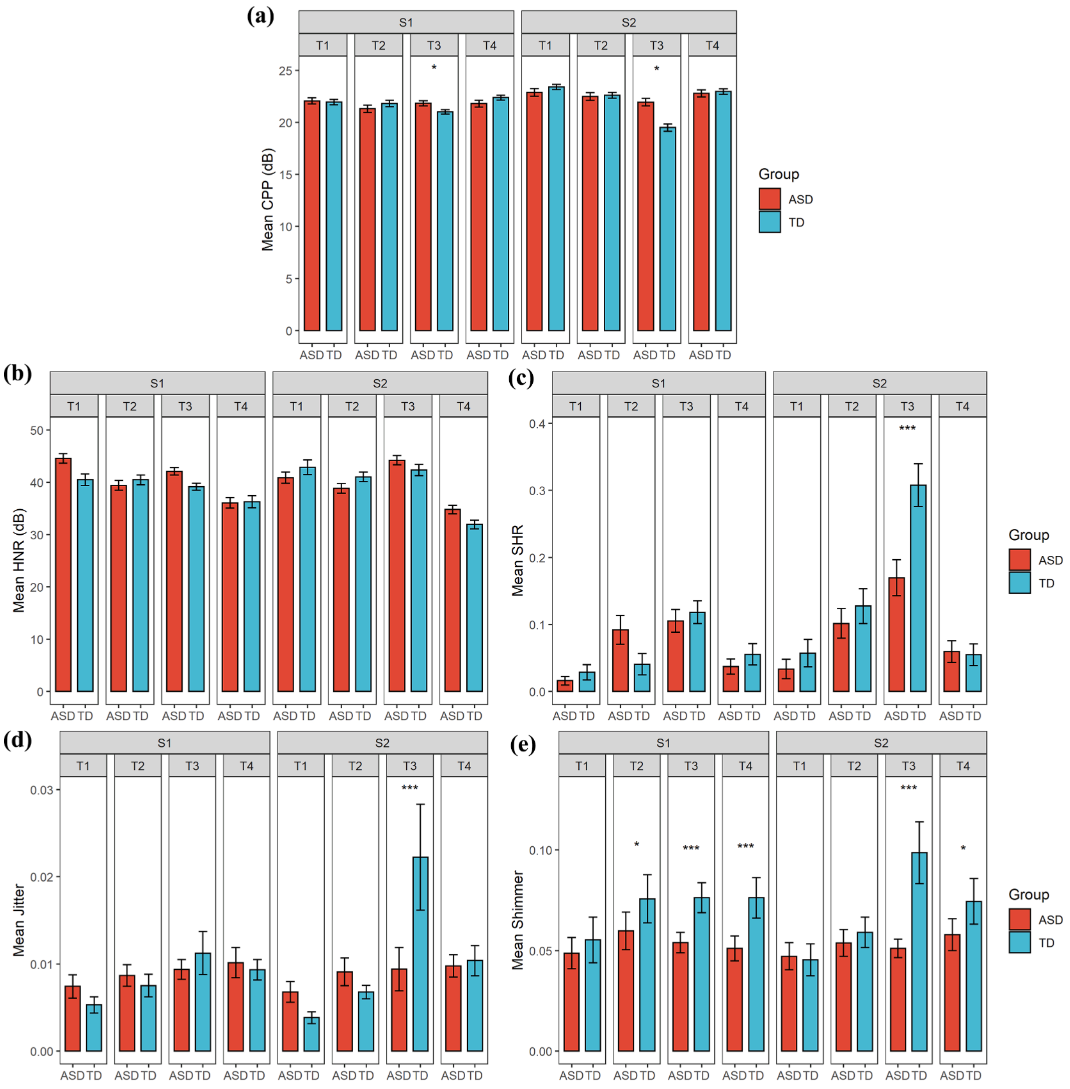


Fig. 3. The differences between ASD and TD groups in terms of CPP (a), HNR (b), SHR (c), jitter (d), and shimmer (e). The bars indicate the 95 % confidence interval. Asterisks indicate the significant findings in post-hoc test, $^{*}p < 0.05$, $^{***}p < 0.001$.

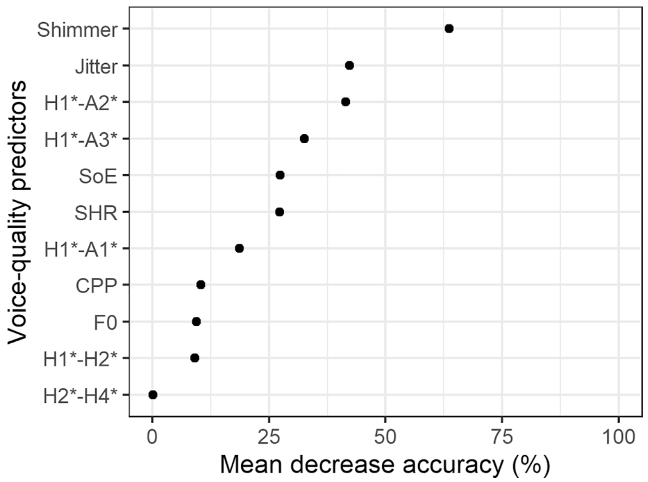


Fig. 4. The relative contribution of voice-quality predictors to the classification of children with and without ASD (ASD vs. TD), measured by the mean decrease in accuracy from the Random Forest algorithm.

robust measurement of dysphonia in recent years [24,25]. Based on the CPP values of 13- to 93-year-old patients in the voice disorders database, it has been proposed that a CPP value below 11.46 dB (for sustained vowels) might indicate the presence of voice disorder [65]. Our results showed that the mean CPP values of Mandarin-speaking children with and without ASD were 22.10 dB (SD = 3.91 dB) and 21.81 dB (SD = 3.50 dB), respectively, which were far beyond the threshold. On the other hand, it was reported that HNR values below 7 dB could be considered pathological [60]. Our results showed that the mean HNR values for Mandarin-speaking autistic children and their TD peers were 40.36 dB (SD = 11.21 dB) and 39.17 dB (SD = 12.29 dB), respectively. Thus, the CPP and HNR of both ASD and TD groups were generally beyond threshold values in the clinical assessment, indicating that both groups had no larynx diseases and disorders.

Judging from jitter, shimmer, CPP, and HNR, the voice quality of Mandarin-speaking children with ASD is not pathology-related, albeit with the mixed subjective descriptions towards them [4]. Their voices could be hardly considered dysphonic since the values of measurements were well beyond the cutoff threshold. Furthermore, more research should be done to set a database for a comprehensive vocal assessment for children covering different age ranges since most healthy voice-quality criteria were established based on infants and adults [61].

4.2. Vocal behavior- and language-related a typicality in Mandarin-speaking children with ASD

Despite the voice quality among Mandarin-speaking children with ASD was not pathological, their acoustical voice-quality signals were atypical during lexical tone production compared with TD children. Except for HNR and F0 range, group differences were widely found in multidimensional parameters within or across syllabic positions and lexical tones. The implications of these differences were discussed below:

4.2.1. Time-domain parameters

Results suggested that both F0 and SoE in children with ASD significantly exceeded those in TD children regardless of prosodic positions and lexical tones. Higher F0 mean in ASD are consistent with numerous studies [6,11,12,15–17], and match the subjective judgment such as "squeal" and "over-exaggerated" [8]. Additionally, significantly enhanced SoE among individuals with ASD was also reported in the literature [19]. In a nutshell, results indicated that Mandarin-speaking children with ASD abnormally exhibited greater strength in voicing, which could be treated as excitation source characteristics in ASD.

4.2.2. Spectral parameters

Analyses of all parameters indicated that spectral tilts (e.g., H1*-H2* and H1*-A1*) were lower in the ASD group than in the TD group. Lower values of spectral tilts, to some degree, suggested that children with ASD exhibited significantly rapid adduction of their vocal folds, perceptually judged as a strained (or pressed) voice [66]. Note that the strained voice was found across all syllabic positions and lexical tones, that is to say, this might be a general vocal fold vibratory feature among individuals with ASD.

4.2.3. Signal aperiodicity

The significant difference in signal aperiodicity between the two groups mainly occurred in the production of T3 (e.g., CPP and Shimmer), especially the T3 in the final syllable (e.g., SHR and jitter). Compared with TD children, Mandarin-speaking children with ASD showed higher CPP values but lower values of SHR, jitter, and shimmer when producing T3. Note that the T3 in Mandarin is a dipping tone accompanied by the feature of creaky voice as an important tone feature [36,37]. The creaky voice often results in irregular (aperiodic) waveform patterns, acoustically manifested by lower values of F0, CPP, and HNR, but higher jitter values and SHR values [45]. As Fig. 2 and Fig. 3 show, TD children produced T3 with creaky voices, while children with ASD could not fully realize this dipping-tone-discriminating feature as well as TD peers.

In conclusion, the atypical voice quality of Mandarin-speaking children with ASD is reflected by overexerting and overstraining their voice in lexical tone production. Therefore, this general result has responded to our first research question. In regard to the second research question, especially when producing the T3 in the final syllable, children with ASD exhibited higher F0 with a less creaky voice, losing the typical tone features in terms of pitch height and phonation type of T3 [37,38].

4.3. Potential and supplementary value for diagnosing tone-language speakers with ASD

The result of the Random Forest classification provided us with three crucial pieces of information, that is, shimmer, jitter, and H1*-A2* contributed higher to discriminating the voice quality between ASD and TD groups. Moreover, voice quality might be considered as a potential acoustic biomarker for Mandarin-speaking children with ASD, since the classification accuracy rate reached 78.5%.

Traditionally, many studies demonstrated the prosodic atypicality in ASD using various acoustic parameters such as F0 (F0 variations) and HNR [10–12], but without a conclusion on detecting the most useful parameters to identify ASD. To date, automatic classification systems have been adopted to clarify ASD by acoustic parameters (e.g., [19,67,68]). The Random Forest classification, for instance, allows us to understand the hierarchy of importance of acoustic parameters [41]. This study found that the shimmer and jitter were the most crucial two parameters identifying the voice quality in Mandarin-speaking children with ASD. When producing T3, Mandarin-speaking children with ASD showed fewer perturbations of both pitch and intensity, resulting in the absence of inherently low (and irregular) F0 and creaky voice of T3. These might be the most robust acoustic cues for Mandarin speakers with ASD, corroborating the lowest rating towards their T3 production (vs. T1, T2, and T4) by trained phoneticians [69].

Regrettably, a classification accuracy rate of 78.5% was obtained in our study, slightly lower than 80%. To account for this, we should take the well-known individual differences of ASD into consideration [59]. The voice description of autistic individuals is not consistent, variously described as "monotonic, song-singy, harsh, hoarse, etc" [4,7]. Besides, the number of participants with ASD and lexical items in this study was also limited, and a larger data size is needed to obtain a more reliable result.

In summary, atypical voice quality could be seen as the potential biobehavioral marker for ASD. Clinically, shimmer and jitter in T3 production might be the most robust cues for Mandarin-speaking children with ASD. That is also the answer to our third research question raised before. For Mandarin-speaking young children exhibiting early signs and symptoms of ASD, the non-invasive voice-quality assessment would have the supplementary value for diagnosing ASD.

5. Conclusions

This study investigated the voice quality of Mandarin-speaking children with ASD and their TD peers using 13 acoustic parameters, i. e., F0, F0 range, SoE (time-domain parameters), H1*-H2*, H2*-H4*, H1*-A1*, H1*-A2*, H1*-A3* (spectral tilt), CPP, HNR, SHR, jitter, and shimmer (signal aperiodicity). Then, the parameters discriminating voice quality between two groups were utilized for automatic classification by using the Random Forest algorithm to find out robust acoustic parameters to diagnose ASD. Results of statistical analysis showed that except for HNR and F0 range, notable group differences were found in other 11 parameters either across all prosodic contexts or within a certain lexical tone. Besides, an accuracy rate of 78.5% was obtained in Random Forest classification, indicating shimmer and jitter were the two crucial parameters that contributed mostly when diagnosing ASD. Based on these parameters, although Mandarin-speaking children with ASD had no obvious voice disorders, they tended to overexert and overstrain their voices, resulting in atypical voice quality, especially when producing the dipping tone of T3. Clinically, the voice-quality assessment has potential and supplementary value for diagnosing Mandarin speakers with ASD.

Author contributions

Chengyu Guo and Fei Chen conceived and designed the study, participated in the statistical analysis, interpreted the data, and wrote the first draft of the manuscript; Jinting Yan collected the data; Chengyu Guo and Yajie Chang annotated and analyzed the recording data.

Funding

This work was partly supported by the Social Science Foundation of Hunan Province (No. 20ZDB003), the Fundamental Research Funds for the Central Universities, Hunan University (No. 531118010660), and the MOE (Ministry of Education of China) Youth Foundation Project of Humanities and Social Sciences (No. 21YJC740015).

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendices

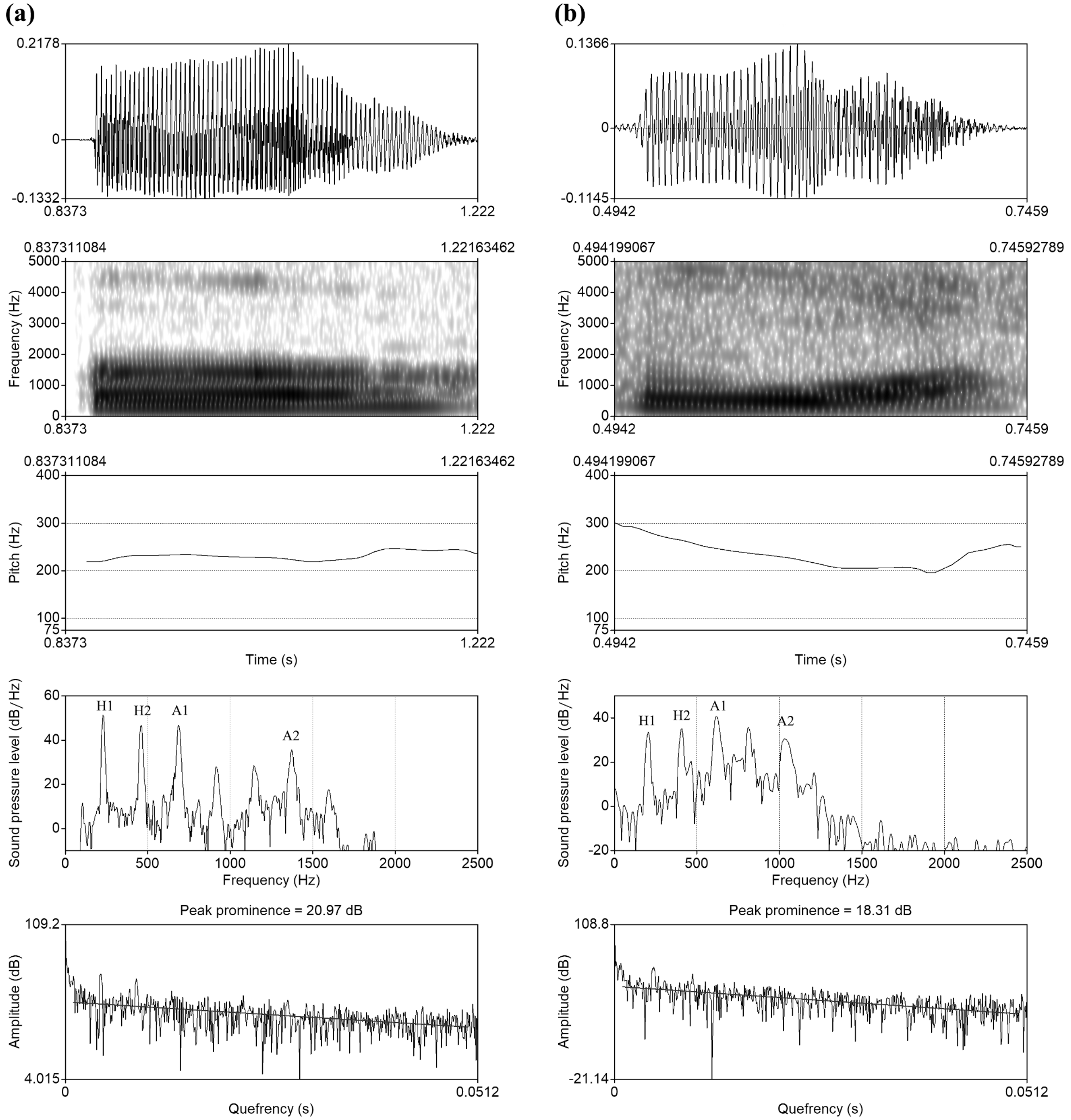
Appendix A

The target words were classified into four lexical tones (T1, T2, T3, and T4) and two prosodic positions (S1 and S2). In the initial syllable, the tonal representations of T3 include half-sandhi and full-sandhi. Boldface indicates the target syllable.

Syllable	T1	T2	T3 (half-sandhi/ full-sandhi)		T4
S1	**积 [tɕi^{55}]** 木	**葡 [pʰu^{35}]** 萄	**企 [tɕʰi^{21}]** 鹅	**雨 [y^{35}]** 伞	**气 [tɕʰi^{51}]** 球
	Toy blocks	Grape	Penguin	Umbrella	Balloon
	乌 [u^{55}] 龟	**魔 [mo^{35}]** 方	**土 [tʰu^{21}]** 豆	**马 [mʌ^{35}]** 桶	**兔 [tʰu^{51}]** 子
	Tortoise	Magic cube	Potato	Toilet	Rabbit
	菠 [po^{55}] 萝	**麻 [mʌ^{35}]** 花	**薯 [ʂu^{21}]** 片	**老 [lɑu^{35}]** 虎	**大 [tʌ^{51}]** 象
	Pineapple	Bread twist	Potato chips	Tiger	Elephant
S2	滑梯 **[tʰi^{55}]**	拼图 **[tʰu^{35}]**	玉米 **[mi^{214}]**		积木 **[mu^{51}]**
	Slide	Puzzle	Corn		Toy blocks
	橙汁 **[tʂʅ^{55}]**	草莓 **[mei^{35}]**	老虎 **[xu^{214}]**		白菜 **[tsʰai^{51}]**
	Orange juice	Strawberry	Tiger		Cabbage
	风车 **[tʂʰɤ^{55}]**	企鹅 **[ɤ^{35}]**	斑马 **[mʌ^{214}]**		手套 **[tʰɑu^{51}]**
	Windmill	Penguin	Zebra		Gloves

Appendix B

The waveform, spectrogram, F0 contour, spectral slice (in the middle point of a vowel), and cepstrum (in the middle point of a vowel) of the final syllable in “tiger” [xu214] from a Mandarin-speaking autistic child (a) and a Mandarin-speaking TD child (b).



Note. H1 = the first harmonic; H2 = the second harmonic; A1 = the harmonic nearest the first formant; A2 = the harmonic nearest the second formant.

Appendix C

Tukey adjusted post-hoc comparison of shimmer values between the ASD group and TD group across prosodic positions (S1, S2) and lexical tones (T1, T2, T3, T4). Boldface indicates the significant findings.

Parameter	Syllable	Tone	contrast	estimate	*SE*	df	*t*	*p*
Shimmer	S1	T1	ASD-TD	−0.007	0.007	140	−0.926	0.3561
		T2	**ASD-TD**	**−0.017**	**0.008**	**168**	**−2.238**	**<0.05***
		T3	**ASD-TD**	**−0.023**	**0.006**	**71**	**−3.694**	**<0.001*****
		T4	**ASD-TD**	**−0.025**	**0.007**	**140**	**−3.470**	**<0.001*****

(*continued on next page*)

(*continued*)

Parameter	Syllable	Tone	contrast	estimate	*SE*	df	*t*	*p*
	S2	T1	ASD-TD	0.002	0.008	186	0.241	0.810
		T2	ASD-TD	−0.006	0.007	148	−0.837	0.404
		T3	**ASD-TD**	**−0.048**	**0.007**	**143**	**−6.503**	**<0.001*****
		T4	**ASD-TD**	**−0.017**	**0.007**	**145**	**−2.268**	**<0.05***

References

[1] P. Evans, S. Golla, M. Ann Morris (Eds.), Spectrum Disorders: Clinical Considerations. Rosenberg's Molecular and Genetic Basis of Neurological and Psychiatric Disease, Fifth Edition, Academic Press, 2015, pp. 197–207, https://doi.org/10.1016/B978-0-12-410529-4.00018-8.

[2] American Psychiatric Association, Diagnostic and Statistical Manual of Mental Disorders, American Psychiatric Association (2013), https://doi.org/10.1176/appi.books.9780890425596.

[3] H. Tager-Flusberg, Understanding the language and communicative impairments in autism. In International review of research in mental retardation, 23, Elsevier, 2000, pp. 185–205.

[4] C.A.M. Baltaxe, J.Q. Simmons, Prosodic Development in Normal and Autistic Children, in: E. Schopler, G.B. Mesibov (Eds.), Communication Problems in Autism, Springer, US, 1985, pp. 95–125, https://doi.org/10.1007/978-1-4757-4806-2_7.

[5] D. Bone, C.-C. Lee, M.P. Black, M.E. Williams, S. Lee, P. Levitt, S. Narayanan, The Psychologist as an Interlocutor in Autism Spectrum Disorder Assessment: Insights From a Study of Spontaneous Prosody, J. Speech, Language, Hearing Res. 57 (4) (2014) 1162–1177, https://doi.org/10.1044/2014_JSLHR-S-13-0062.

[6] S.J. Sheinkopf, P. Mundy, D.K. Oller, M. Steffens, Vocal Atypicalities of Preverbal Autistic Children, J. Autism Dev. Disord. 30 (4) (2000) 345–354, https://doi.org/10.1023/A:1005531501155.

[7] L. Kanner, Autistic disturbances of affective contact, Nervous Child 2 (3) (1943) 217–250.

[8] A. Järvinen-Pasley, S. Peppé, G. King-Smith, P. Heaton, The Relationship between Form and Function Level Receptive Prosodic Abilities in Autism, J. Autism Dev. Disord. 38 (7) (2008) 1328–1340, https://doi.org/10.1007/s10803-007-0520-z.

[9] W. Pronovost, M.P. Wakstein, D.J. Wakstein, A Longitudinal Study of the Speech Behavior and Language Comprehension of Fourteen Children Diagnosed Atypical or Autistic, Exceptional Children 33 (1) (1966) 19–26, https://doi.org/10.1177/001440296603300104.

[10] A.-M.-R. Depape, A. Chen, G.B.C. Hall, L.J. Trainor, S.A.E. Kotz, L. Kuchinke, Use of prosody and information structure in high functioning adults with Autism in relation to language ability, Front. Psychol. 3 (2012) 72, https://doi.org/10.3389/fpsyg.2012.00072.

[11] R. Paul, L.D. Shriberg, J. McSweeny, D. Cicchetti, A. Klin, F. Volkmar, Brief Report: Relations between Prosodic Performance and Communication and Socialization Ratings in High Functioning Speakers with Autism Spectrum Disorders, J. Autism Dev. Disord. 35 (6) (2005) 861–869, https://doi.org/10.1007/s10803-005-0031-8.

[12] L.D. Shriberg, R. Paul, J.L. McSweeny, A. Klin, D.J. Cohen, F.R. Volkmar, Speech and Prosody Characteristics of Adolescents and Adults with High-Functioning Autism and Asperger Syndrome, J. Speech, Language, Hearing Res. 44 (5) (2001) 1097–1115, https://doi.org/10.1044/1092-4388(2001/087).

[13] A.S. Warlaumont, J.A. Richards, J. Gilkerson, D.K. Oller, A Social Feedback Loop for Speech Development and Its Reduction in Autism, Psychol. Sci. 25 (7) (2014) 1314–1324, https://doi.org/10.1177/0956797614531023.

[14] C. Lord, M. Rutter, P.C. DiLavore, S. Risi, K. Gotham, S. Bishop, R.J. Luyster, W. Guthrie. *Autism Diagnostic Observation Schedule*, Second Edition (ADOS-2), Western Psychological Services, 2012.

[15] F. Chen, C.-H. Cheung, G. Peng, Linguistic Tone and Non-Linguistic Pitch Imitation in Children with Autism Spectrum Disorders: A Cross-Linguistic Investigation, J. Autism Dev. Disord. 52 (5) (2021) 2325–2343.

[16] R. Fusaroli, A. Lambrechts, D. Bang, D.M. Bowler, S.B. Gaigg, Is voice a marker for Autism spectrum disorder? A systematic review and meta-analysis, Autism Res. 10 (3) (2017) 384–407, https://doi.org/10.1002/aur.1678.

[17] D.K. Oller, P. Niyogi, S. Gray, J.A. Richards, J. Gilkerson, D. Xu, U. Yapanel, S. F. Warren, Automated vocal analysis of naturalistic recordings from children with autism, language delay, and typical development, Proc. Natl. Acad. Sci. 107 (30) (2010) 13354–13359, https://doi.org/10.1073/PNAS.1003882107.

[18] A.J. Chong, M. Risdal, A. Aly, J. Zymet, P. Keating, Effects of consonantal constrictions on voice quality, J. Acoust. Soc. Am. 148 (1) (2020) EL65–EL71, https://doi.org/10.1121/10.0001585.

[19] A. Mohanta, V.K. Mittal, Analysis and classification of speech sounds of children with autism spectrum disorder using acoustic features, Comput. Speech Lang. 72 (2022), 101287, https://doi.org/10.1016/J.CSL.2021.101287.

[20] C. d'Alessandro, Voice source parameters and prosodic analysis, in: S. Sudhoff, D. Lenertova, R. Meyer, S. Pappert, P. Augurzky, I. Mleinek, N. Richter, J. Schließer (Eds.), Methods in empirical prosody research, De Gruyter, 2006, pp. 63–87, https://doi.org/10.1515/9783110914641.

[21] M. Gordon, P. Ladefoged, Phonation types: a cross-linguistic overview, J. Phonet. 29 (4) (2001) 383–406, https://doi.org/10.1006/JPHO.2001.0147.

[22] J. Hillenbrand, R.A. Houde, Acoustic Correlates of Breathy Vocal Quality: Dysphonic Voices and Continuous Speech, J. Speech, Language, Hearing Res. 39 (2) (1996) 311–321, https://doi.org/10.1044/jshr.3902.311.

[23] J.F. Santos, N. Brosh, T.H. Falk, L. Zwaigenbaum, S.E. Bryson, W. Roberts, I. M. Smith, P. Szatmari, J.A. Brian, Very early detection of Autism Spectrum Disorders based on acoustic analysis of pre-verbal vocalizations of 18-month old toddlers, in: 2013 IEEE International Conference on Acoustics, Speech and Signal Processing, 2013, pp. 7567–7571, https://doi.org/10.1109/ICASSP.2013.6639134.

[24] R. Fraile, J.I. Godino-Llorente, Cepstral peak prominence: A comprehensive analysis, Biomed. Signal Process. Control 14 (2014) 42–54, https://doi.org/10.1016/j.bspc.2014.07.001.

[25] R.R. Patel, S.N. Awan, J. Barkmeier-Kraemer, M. Courey, D. Deliyski, T. Eadie, D. Paul, J.G. Švec, R. Hillman, Recommended Protocols for Instrumental Assessment of Voice: American Speech-Language-Hearing Association Expert Panel to Develop a Protocol for Instrumental Assessment of Vocal Function, Am. J. Speech-Language Pathol. 27 (3) (2018) 887–905, https://doi.org/10.1044/2018_AJSLP-17-0009.

[26] B. Blankenship, The time course of breathiness and laryngealization in vowels, University of California, Los Angeles, 1997.

[27] P. Boersma, Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound, Proc. Inst. Phonetic Sci. 17 (1993) 97–110.

[28] R. Montaño, F. Alías, The role of prosody and voice quality in indirect storytelling speech: Annotation methodology and expressive categories, Speech Commun. 85 (2016) 8–18, https://doi.org/10.1016/J.SPECOM.2016.10.006.

[29] X. Sun, Pitch determination and voice quality analysis using Subharmonic-to-Harmonic Ratio, IEEE International Conference on Acoustics Speech and Signal Processing 333–336 (2002), https://doi.org/10.1109/ICASSP.2002.5743722.

[30] J.H. Esling, S.R. Moisik, A. Benner, L. Crevier-Buchman, Voice Quality: The Laryngeal Articulator Model, Cambridge University Press (2019), https://doi.org/10.1017/9781108696555.

[31] P.H. Dejonckere, P. Bradley, P. Clemente, G. Cornut, L. Crevier-Buchman, G. Friedrich, P. van de Heyning, M. Remacle, V. Woisard, A basic protocol for functional assessment of voice pathology, especially for investigating the efficacy of (phonosurgical) treatments and evaluating new assessment techniques, Eur. Arch. Otorhinolaryngol. 258 (2) (2001) 77–82, https://doi.org/10.1007/s004050000299.

[32] M. Asgari, A. Bayestehtashk, I. Shafran, Robust and Accurate Features for Detecting and Diagnosing Autism Spectrum Disorders, Interspeech 2013 (2013) 191–194.

[33] M.J. Boucher, M.V. Andrianopoulos, S.L. Velleman, L.A. Keller, L. Pecora, Assessing vocal characteristics of spontaneous speech in children with autism, American Speech-Language-Hearing Association Convention, San Diego, CA, 2011.

[34] G.B. Kempster, B.R. Gerratt, K. Verdolini Abbott, J. Barkmeier-Kraemer, R. E. Hillman, Consensus Auditory-Perceptual Evaluation of Voice: Development of a Standardized Clinical Protocol, Am. J. Speech-Language Pathol. 18 (2) (2009) 124–132, https://doi.org/10.1044/1058-0360(2008/08-0017).

[35] M. Yip, Tone. Cambridge University Press, 2002.

[36] A. Belotel-Grenié, M. Grenié, Phonation types analysis in standard Chinese, in: Proc. 3rd International Conference on Spoken Language Processing (ICSLP), 1994, pp. 343–346.

[37] J. Kuang, Phonation in Tonal Contrasts [University of California, Los Angeles], ProQuest Dissertations and Theses (2013). https://escholarship.org/uc/item/6n72p16n.

[38] R.-X. Yang, The Phonation Factor in the Categorical Perception of Mandarin Tones. Proc. 17th International Congress of Phonetic Sciences (ICPhS XVII), (2011), 2204–2207.

[39] P.R. Callier, Phonation and tone in conversational Beijing Mandarin, J. Acoust. Soc. Am. 135 (4) (2014), https://doi.org/10.1121/1.4877543, 2295–2295.

[40] S. Li, W. Gu, L. Liu, P. Tang, The Role of Voice Quality in Mandarin Sarcastic Speech: An Acoustic and Electroglottographic Study, J. Speech, Language, Hearing Res. 63 (8) (2020) 2578–2588, https://doi.org/10.1044/2020_JSLHR-19-00166.

[41] L. Breiman, Random Forests, Mach. Learn. 45 (1) (2001) 5–32, https://doi.org/10.1023/A:1010933404324.

[42] A. Liaw, M. Wiener, Classification and Regression by randomForest, R News 2 (3) (2002) 18–22.

[43] C. Ning, Test of language ability in Mandarin-speaking preschoolers, Tianjin University Press, 2013.

[44] D.J. Ehrler, R.L. McGhee, PTONI: Primary test of nonverbal intelligence, Pro-Ed Austin, TX, 2008.

[45] P. Keating, M. Garellek, J. Kreiman, Acoustic properties of different kinds of creaky voice, in: 18th International Congress of Phonetic Sciences, 2015, pp. 2–7.

[46] J. Yan, F. Chen, X. Gao, G. Peng, Auditory-Motor Mapping Training Facilitates Speech and Word Learning in Tone Language-Speaking Children With Autism: An

Early Efficacy Study, J. Speech, Language, Hear. Res. 64 (12) (2021) 4664–4681, https://doi.org/10.1044/2021_JSLHR-21-00029.

[47] P. Boersma, D. Weenink, Praat: doing phonetics by computer [Computer program (Version 6.1.56)] (2021) https://www.praat.org.

[48] A. Turk, S. Nakai, M. Sugahara, Acoustic segment durations in prosodic research: A practical guide, in: S. Sudhoff, D. Lenertova, R. Meyer, S. Pappert, P. Augurzky, I. Mleinek, N. Richter, J. Schließer (Eds.), Methods in empirical prosody research, De Gruyter, 2006, pp. 1–28, https://doi.org/10.1515/9783110914641.

[49] Y.-L. Shue, P. Keating, C. Vicenik, K. Yu, VoiceSauce – A program for voice analysis [Computer Program] (Version 1.37) (2011). http://www.phonetics.ucla.edu/voicesauce/.

[50] H. Kawahara, A. de Cheveigne, R.D. Patterson, An instantaneous-frequency-based pitch extraction method for high-quality speech transformation: revised TEMPO in the STRAIGHT-suite. Fifth International Conference on Spoken Language Processing (ICSLP), 1998.

[51] N. Xu Rattanasone, P. Tang, I. Yuen, L. Gao, K. Demuth, Five-Year-olds' Acoustic Realization of Mandarin Tone Sandhi and Lexical Tones in Context Are Not Yet Fully Adult-Like, Front. Psychol. 9 (2018), https://doi.org/10.3389/fpsyg.2018.00817.

[52] M. Iseli, Y.-L. Shue, A. Alwan, Age, sex, and vowel dependencies of acoustic measures related to the voice source, J. Acoust. Soc. Am. 121 (4) (2007) 2283–2295, https://doi.org/10.1121/1.2697522.

[53] R Core Team, R: A language and environment for statistical computing (Version 3.6.3), R Foundation for Statistical Computing (2020) https://www.R-project.org/.

[54] D. Bates, M. Mächler, B. Bolker, S. Walker, Fitting Linear Mixed-Effects Models Using **lme4**, J. Stat. Softw. 67 (1) (2015), https://doi.org/10.18637/jss.v067.i01.

[55] D.J. Barr, R. Levy, C. Scheepers, H.J. Tily, Random effects structure for confirmatory hypothesis testing: Keep it maximal, J. Mem. Lang. 68 (3) (2013) 255–278, https://doi.org/10.1016/J.JML.2012.11.001.

[56] H. Singmann, B. Bolker, J. Westfall, F. Aust, M.S. Ben-Shachar, afex: Analysis of factorial experiments, R Package Version (2015) 13–145.

[57] R.V. Lenth, Least-Squares Means: The *R* Package **lsmeans**, J. Stat. Softw. 69 (1) (2016), https://doi.org/10.18637/jss.v069.i01.

[58] M. Belgiu, L. Drăguţ, Random forest in remote sensing: A review of applications and future directions, ISPRS J. Photogramm. Remote Sens. 114 (2016) 24–31, https://doi.org/10.1016/j.isprsjprs.2016.01.011.

[59] D. Trembath, G. Vivanti, Problematic but predictive: Individual differences in children with autism spectrum disorders, Int. J. Speech-Language Pathol. 16 (1) (2014) 57–60, https://doi.org/10.3109/17549507.2013.859300.

[60] J.P. Teixeira, C. Oliveira, C. Lopes, Vocal Acoustic Analysis – Jitter, Shimmer and HNR Parameters, Procedia Technol. 9 (2013) 1112–1122, https://doi.org/10.1016/j.protcy.2013.12.124.

[61] A. Banik, S. Arya, A. Kant, Vocal Parameters in Children between 4 To 12 Years of Age: An Attempt to Establish a Prototype Database, Int. J. Sci. Res. Publ. 5 (11) (2015) 446–453, www.ijsrp.org.

[62] L.E. Glaze, D.M. Bless, P. Milenkovic, R.D. Susser, Acoustic characteristics of children's voice, J. Voice 2 (4) (1988) 312–319, https://doi.org/10.1016/S0892-1997(88)80023-7.

[63] Z. Zhang, Contribution of laryngeal size to differences between male and female voice production, J. Acoust. Soc. Am. 150 (6) (2021) 4511–4521, https://doi.org/10.1121/10.0009033.

[64] M. Brockmann-Bauser, D. Beyer, J.E. Bohlender, Clinical relevance of speaking voice intensity effects on acoustic jitter and shimmer in children between 5;0 and 9;11 years, Int. J. Pediatr. Otorhinolaryngol. 78 (12) (2014) 2121–2126, https://doi.org/10.1016/j.ijporl.2014.09.020.

[65] O. Murton, R. Hillman, D. Mehta, Cepstral Peak Prominence Values for Clinical Voice Evaluation, Am. J. Speech-Language Pathol. 29 (3) (2020) 1596–1607, https://doi.org/10.1044/2020_AJSLP-20-00001.

[66] C.M. Sapienza, B. Hoffman Ruddy, *Voice Disorders* (Third edition), Plural Publishing, 2018.

[67] S. Cho, M. Liberman, N. Ryant, M. Cola, R.T. Schultz, J. Parish-Morris, Automatic Detection of Autism Spectrum Disorder in Children Using Acoustic and Text Features from Brief Natural Conversations, Proc. Interspeech 2019 (2019) 2513–2517, https://doi.org/10.21437/Interspeech.2019-1452.

[68] E. Marchi, B. Schuller, S. Baron-Cohen, O. Golan, S. Bölte, P. Arora, R. Häb-Umbach, Typicality and emotion in the voice of children with autism spectrum condition: evidence across three languages, Proc. Interspeech 2015 (2015) 115–119, https://doi.org/10.21437/Interspeech.2015-38.

[69] H. Wu, F. Lu, B. Yu, Q. Liu, Phonological acquisition and development in Putonghua-speaking children with Autism Spectrum Disorders, Clinical Linguist. Phonet. 34 (9) (2020) 844–860, https://doi.org/10.1080/02699206.2019.1702720.