AJSLP

Research Article

Pitch Variation in Children With Childhood Apraxia of Speech: Preliminary Findings

Eddy C. H. Wong,[a] Shelley L. Velleman,[b] Michael C. F. Tong,[c,d] and Kathy Y. S. Lee[c,d]

Introduction: Pitch variation, which refers to one's ability to vary fundamental frequency (F0) within or between syllables when speaking, has not been investigated in children with childhood apraxia of speech (CAS). However, pitch variation plays an important role in tone languages, as varying F0 patterns communicate different lexical meanings. This study investigated pitch variation abilities in individuals with CAS via the tone-sequencing tasks (TSTs), focusing on task performance and the effects of syllable structure, lexical status, and tones.

Method: Three Cantonese-speaking children with CAS (aged 3;7–5;8 [years;months]) and six children without CAS participated in the study. Children without CAS were divided into two control groups, comprising those with speech and/or language impairment or typical development. TSTs consisted of 56 sets of five repetitions of stimuli. The stimuli varied in syllable structure, lexical status, and tones. Percentage of tones correct (PTC), consistency scores, F0 values, and acoustic repetition duration were measured.

Results: The CAS group performed more poorly than the control groups on the TST with respect to tone accuracy, consistency, and repetition duration. No interaction effects between group and syllable structure or group and lexical status were found. No significant difference was found on F0 values across time between Tone 1 and Tone 2 syllables in the CAS group. However, interaction effects between group and time points of F0 values on Tone 2 syllables were found.

Discussion: The results suggest that children with CAS have difficulty with pitch variation, which was revealed on the TST with respect to tone accuracy, consistency, and repetition duration. Moreover, children with CAS have difficulty in varying F0 values to produce high-rising tones and tend to use high-level tones to substitute. Clinically, the TST may be useful to assist in the diagnosis of CAS. Isolated vowel stimuli may be useful to test young children or children with severe impairment. Future investigations and development of a normed tool for children with CAS are suggested.

Childhood apraxia of speech (CAS) is a motor speech disorder characterized by speech movement-sequencing difficulties that result in errors in speech sound production and prosody (American Speech-Language-Hearing Association [ASHA], 2007). Prosodic errors reflect deficits in using suprasegmental elements in speech, such as rate, stress, intonation, tone, and rhythm. A lexical stress deficit is the most common prosodic symptom of CAS in English (ASHA, 2007).

Stress comprises three perceptual parameters, namely, length, loudness, and pitch. Speakers of English with CAS produce stress errors at the lexical level (ASHA, 2007; Iuzzini-Seigel & Murray, 2017; Kopera & Grigos, 2020). Specifically, stress is either perceptually equal across syllables or wrongly shifted onto nearby syllables, with these being respectively described as "equal stress" and "lexical stress errors" (Shriberg et al., 2011). Lexical stress errors that are detected perceptually are considered to be a potential diagnostic marker for discriminating CAS from speech delay and dysarthria (Murray et al., 2015; Shriberg et al., 2003).

Physical correlates of the three main parameters of lexical stress have been used to examine stress acoustically. Duration, intensity, and fundamental frequency (F0) are

[a]Department of Chinese and Bilingual Studies, The Hong Kong Polytechnic University, China
[b]Department of Communication Sciences and Disorders, The University of Vermont, Burlington
[c]Department of Otorhinolaryngology, Head and Neck Surgery, The Chinese University of Hong Kong, China
[d]The Institute of Human Communicative Research, The Chinese University of Hong Kong, China

Correspondence to Eddy C. H. Wong: eddychwong@gmail.com

Editor-in-Chief: Jessica E. Huber
Editor: Yunjung Kim

Received May 27, 2020
Revision received August 23, 2020
Accepted February 22, 2021
https://doi.org/10.1044/2021_AJSLP-20-00150

Publisher Note: This article is part of the Special Issue: Selected Papers From the 2020 Conference on Motor Speech—Clinical Science and Implications.

Disclosure: *The authors have declared that no competing interests existed at the time of publication.*

parameters that describe the length, loudness, and pitch components of stress, respectively (Terband et al., 2019). Though lexical stress errors in children with CAS have been studied empirically, there have been limited independent investigations of pitch variation skills in children with CAS. Kopera and Grigos (2020) examined the acoustic parameters underlying lexical stress production in children with CAS. In the study, the stress patterns of three groups of children (children with CAS, speech delay, and typical development [TD]) were compared with respect to F0 and duration independently by calculating the pairwise variability index. No significant group differences were found for any acoustic variables. The authors suggested that differences in the acoustic manifestations of lexical stress between children with and without CAS may only be seen when all of the acoustic parameters are studied in a collective fashion, such as by using the lexical stress ratio (Shriberg et al., 2003).

However, in children with CAS, more attention should also be paid to pitch variation, which represents the skill of varying F0 within or between syllables. There are two reasons for this. First, appropriate production of prosody requires motor control of the laryngeal muscles (Liang & Du, 2018). Given that children with CAS have difficulty in sequencing the spatiotemporal parameters of muscles for the accurate production of segments and suprasegments (ASHA, 2007), children with CAS may be unable to coordinate movements of the oral articulators with simultaneous laryngeal movements to achieve perceptually accurate pitch production when speaking. Thus, children with CAS may demonstrate deficits in pitch variation.

Second, pitch plays an important role in lexical meaning in tone languages. For example, Cantonese has six distinctive tones (Yip & Matthews, 2011): Tones 1, 3, and 6 are high-, mid-, and low-level tones, respectively, whereas Tone 2 is high rising, Tone 5 is low rising, and Tone 4 is low falling. These tones are different with respect to both the levels and the contours of F0. In English orthography, tones are indicated with a number after the word, as in /jɐu4 sœy2/ ("swim"), which has a low-falling tone on the first syllable and a high-rising tone on the second. The variations in F0 for the six tones indicate different lexical meanings. For example, /jɐu4 sœy2/ means "swim" and /jɐu4 sœy3/ means "convince." The two verbs share identical segmental features and sequences, and the tones differentiate their meanings. Miscommunication would result if a speaker produced the high-rising tone (Tone 2) in the second syllable as a mid-level tone (Tone 3). Subtle changes in F0 levels and/or contour can thus alter lexical meanings, and there is therefore an urgent need for an investigation of pitch (F0) variation in CAS, especially in the context of tone languages.

The tone-sequencing task (TST) was designed to investigate pitch variation skills in Cantonese-speaking children with CAS (Wong et al., 2019). The TST requires children to rapidly imitate seven sets of words/nonwords following the verbal instruction "repeat after me as fast as possible." The sets are each formed of five repetitions of a real or nonsense word stimulus, with these stimuli being monosyllabic, disyllabic, or trisyllabic. In Wong et al.'s (2019) work, each syllable was formed by a consonant–vowel (CV) structure (i.e., [mɔ]) containing one of the three early-developing Cantonese tones (i.e., Tone 1 [high level], Tone 2 [high rising], or Tone 4 [low falling]). There were three monosyllabic stimuli (i.e., [mɔ1], [mɔ2], and [mɔ4]), three disyllabic stimuli (i.e., /[mɔ1mɔ2], [mɔ2mɔ4], and [mɔ4mɔ1]), and one trisyllabic stimulus (i.e., [mɔ1mɔ2mɔ4]). Wong et al. (2019) found that children with CAS had difficulty with pitch variation and that this was observable on the TST. Specifically, tone accuracy, consistency, and set duration of production on the TST in two children with CAS were compared to those of two children with non-CAS speech delay (i.e., phonological impairment). The results showed that, compared with the children in the control group, those in the CAS group had low percentages of correct tones, low production consistency, and longer durations of five repetitions. The study was confounded by co-occurring language impairment in the speakers with CAS. Consequently, it was not clear whether poor TST performance by the CAS group was caused by degraded speech motor planning and/or programming skills, co-occurring language impairment, or both.

To further investigate the pitch variation abilities of children with CAS via the TST, in addition to comparing their performance with that of children without CAS, three possible linguistic factors need to be examined, namely, syllable structure, lexical status, and tones. First, syllable structure may play a role in tone production in children with CAS, as they have difficulty in producing relatively complex syllable structures (ASHA, 2007; Wong et al., 2020a). In particular, the empirical literature indicates that CAS is characterized by difficulty in transitioning articulatory gestures from phoneme to phoneme (ASHA, 2007). Although Cantonese tone is carried on the vowel portion of a syllable, it is unknown whether the presence of an initial consonant would affect the pitch production of the vowel. From a motor perspective, children with CAS may have more difficulty in producing accurate pitch in CV syllables than in isolated vowels.

Second, lexical status may play a role in pitch variation abilities. Lee et al. (2002) found that 2- and 3-year-old Cantonese learners perceived tones more accurately in words than in nonwords, although the stimuli were not controlled for syllable shape. For adult Cantonese speakers, high lexical frequency has a priming effect on tone production, such that their tone performance on familiar words is better than on nonwords (Zhao & Jurafsky, 2007). It is possible that the priming effect works on children with CAS, as well. In other words, children with CAS may perform better on word stimuli than on nonword stimuli.

Last, tones are produced with different F0 contours, and this pitch variation requires rapid laryngeal muscle movements to regulate these F0 changes. In Mandarin, another tone language spoken by Chinese, visible neck movements that occur during tone production were suggested as a facilitative strategy for tone identification (Chen & Massaro, 2008). This suggests that there are significant spatiotemporal movements for different tone productions. The movements required for Cantonese tones may challenge the abilities of children with CAS due to their core deficits in planning and/or programming spatiotemporal parameters of movement sequences for speech (ASHA, 2007). Specifically, they may have difficulty in coordinating the spatiotemporal parameters of muscles to raise or lower the F0 contours within a syllable while simultaneously producing consonant and vowel sounds. In particular, rising pitch appears to be more challenging than falling pitch. Snow (1998) analyzed acoustic data from ten 4-year-old English-speaking children with no history of hearing, speech, or language impairment. The results showed that the children replaced rising pitch patterns with falling ones when adult models were given for imitation. When the contour direction of sentence-final rising tones was correctly imitated by the children, accompanying lengthened word durations suggested that the speed of pitch change was slower than that of adults in rising tones. This finding suggests that production of rising pitch patterns may be more difficult than falling ones with respect to F0 values and duration in typically developing young children. Thus, it is logical to predict that the rising tones in Cantonese words may be more challenging than falling tones for Cantonese-speaking children, especially children with CAS.

Thus, the purpose of this study was to (a) compare the performance on tone accuracy, tone consistency, and repetition duration of children with and without CAS and (b) examine the effects of three factors (i.e., syllable structure, lexical status, and tone) on pitch variation skills in children with CAS. We compared two perceptual ratings (percentage of tones correct [PTC]) and consistency scores) and two acoustic variables (F0 values and durations of sets of five repetitions of stimuli) among children with CAS, children with non-CAS speech and/or language impairment (nCAS), and children with TD. The following hypotheses were proposed:

1. The CAS group will show difficulty with pitch variation, characterized by reduced PTC and consistency scores and lengthened repetition durations, compared with the control groups on the TST. In particular, the CAS group will perform more poorly than the controls on these measures.
2. Syllable structure will have no effect on the control groups but will affect the CAS group. Better performance will be observed on vowel (V) stimuli than on CV stimuli.
3. Lexical knowledge will facilitate pitch variation skills in both the CAS and control groups, as compared with performance in response to nonlexical stimuli.
4. Stimuli with rising and falling tones, especially rising tones, will challenge the children with CAS, whereas those in the other groups will not have significant difficulty with these.

Method

Participants

This study was approved by the Joint Chinese University of Hong Kong–New Territories East Cluster Clinical Research Ethics Committee (Reference Number 2018.265). Three Cantonese-speaking children with CAS (two boys and one girl) participated in this study. Table 1 lists their assessment results. Their ages ranged from 3;7 to 5;8 (years;months; M = 54.33 months, SD = 12.67 months). The inclusion criteria were (a) age between 3;0 and 6;11, (b) Cantonese as the main language used in daily life, (c) no diagnosis of autism spectrum disorder or other medical conditions that affect speech, (d) no structural abnormalities that affect speech, and (e) diagnosed with or suspected of having CAS by a qualified speech therapist. Children who did not imitate sounds were excluded. Six age- and gender-matched children (four boys and two girls) were recruited as controls. Their ages ranged from 3;4 to 6;2 (M = 53.50 months, SD = 13.77 months). Based on the results of the initial assessments, the total of nine children were assigned to three groups, namely, the CAS group (n = 3), the nCAS group (n = 3), and the TD group (n = 3).

An independent research assistant, who had experience conducting acoustic analysis of Cantonese speech, participated in the reliability measures of manual vowel segmentations. A qualified speech and language pathologist who had more than 10 years of clinical experience in Cantonese-speaking children's speech sound disorders (SSDs) participated in this study as an independent rater. Both the research assistant and the rater were blind to the aims of this study and the diagnoses of the children.

Procedure

Initial Assessment

All of the children were assessed by the first author. The assessment protocol comprised a hearing screening, a speech and language sample, a standardized receptive vocabulary test, a standardized articulation screening, and a standardized tone identification test. The standardized expressive language test was only administered to children who had a history or suspicion of expressive language difficulties. Thus, no formal expressive language assessment was done with children in the TD group. In addition, no standardized expressive language test was administered to children with severe SSDs. Because of the effects on their speech accuracy, it was felt that examiners would be at risk of wrongly recording or interpreting some of the children's productions on standardized expressive tasks, resulting in invalid test results. Instead, informal assessments of expressive language skills were administered. All of the children passed a 20-dB pure-tone screening at 1000, 2000, and 4000 Hz (American Academy of Audiology, 1997).

The assessment results are listed in Table 1. The confirmation process for CAS diagnoses is described in the next section. In terms of language skills, within the CAS group, Participant 3 had a moderate impairment in receptive vocabulary skills, and Participants 1 and 2 had receptive vocabulary skills within normal limits (WNL). For Participant 3, expressive language difficulties were reported by the parents. However, due to the child's severe SSD, no standardized expressive language scale was administered.

Table 1. Participant assessment results.

								HKCAT							
				HKCRVT		RDLS-ES		IC	V/Di	FC	T		CANTIT		
No.	Age	Sex	Group	*SD*	Severity	*SD*	Severity	Percentile				Severity	Percentile	*z* scores	Severity
1	3;7	M	CAS	−0.5 < *SD* < 0.0	WNL	NA	NA	0.1	0.1	0.1	0.1	Severe	< 8[a]	< −1.35	Mild
2	5;8	F	CAS	0.0 < *SD* < +0.5	WNL	NA	NA	0.1	0.1	9	0.1	Severe	52	0	WNL
3	4;4	M	CAS	−2.5 < *SD* < −2.0	Moderate	NA	NA	0.1	0.1	0.1	0.1	Severe	< 4[a]	< −2.24	Moderate
4	3;4	M	nCAS	0.0 < *SD* < +0.5	WNL	−1.4 < *SD* < −1.0	Mild	25	63	63	50	Mild	34	−0.62	WNL
5	6;2	F	nCAS	−1.5 < *SD* < −1.0	Mild	NA	NA	0.1	0.1	0.4	50	Severe	12	−1.58	Mild
6	3;10	M	nCAS	−1.0 < *SD* < −0.5	WNL	−1.4 < *SD* < −1.1	Mild	75	63	84	50	WNL	42	−0.47	WNL
7	3;6	M	TD	0.0 < *SD* < +0.5	WNL	NA	NA	75	75	63	50	WNL	42	−0.47	WNL
8	5;6	F	TD	+0.5 < *SD* < +1.0	WNL	NA	NA	63	63	91	50	WNL	96	1.70	WNL
9	4;5	M	TD	0.0 < *SD* < +0.5	WNL	NA	NA	75	67	75	50	WNL	98	1.34	WNL

Note. HKCRVT = Hong Kong Cantonese Receptive Vocabulary Test (Lee et al., 2009); RDLS-ES = Expressive scales of Reynell Developmental Language Scales–Cantonese (Reynell & Huntley, 1987); HKCAT = Hong Kong Cantonese Articulation Test (Cheung et al., 2006); CANTIT = Cantonese Tone Identification Test (Lee, 2012); IC = initial consonants; V/Di = vowels or diphthongs; FC = final consonants; T = tones; M = male; CAS = childhood apraxia of speech; WNL = within normal limits; NA = not applicable; F = female; nCAS = non-CAS speech and/or language impairment; TD = typical development.

[a]In the test manual of the CANTIT (Lee, 2012), the percentiles of the lowest rank in each age group are presented as a range (e.g., < 8). Child 1 and Child 3 performed at the lowest rank in their age groups on tone identification skills.

His expressive language skills were observed during the assessment session. The child mainly used two-word utterances (e.g., subject–verb) for communication. Some three-word utterances (e.g., subject–verb–object) with limited verbs (e.g., "play" and "eat") and nouns (e.g., "this" for a toy) were recorded during the assessment session. Expressive language impairment was suspected. In the nCAS group, Participant 4 had WNL receptive vocabulary skills and mild expressive language impairment, Participant 5 had mild impairment in receptive vocabulary skills, and Participant 6 had mild expressive language impairment only. For Participant 5, no results of standardized expressive language scales were reported due to the coexisting severe SSD that could have confounded the test results. An informal story-retelling task was administered to collect a language sample from this participant for further analysis. Twenty-one utterances were recorded, with a mean length of utterance of 2.04. Expressive language impairment was suspected. In terms of speech production skills, all children in the CAS group had severe SSD. Participants 4 and 5 in the nCAS group had mild and severe phonological disorders, respectively. Phonological processes such as stopping, fronting, and gliding were found.

The groups performed differently on the receptive tone identification test. In the CAS group, Participant 2 had WNL tone identification skills, and Participants 1 and 3 showed significant tone identification problems. In the nCAS group, Participant 5 had a mild deficit on tone identification skills, and Participants 4 and 6 had WNL tone identification skills. All the children in the TD group had WNL tone identification skills.

Confirmation of CAS Diagnoses

Children who were diagnosed with or suspected as having CAS received an additional motor speech assessment from the first author. The 2-hr assessment protocol comprised collection of a case history and a speech sample, an assessment of the ability to imitate polysyllabic words or phrases, and an oral motor assessment that included a diadochokinetic (DDK) task. The first author, who has more than 10 years of experience in differential diagnosis of pediatric SSDs, confirmed the CAS diagnoses based on the following criteria, with modifications of the assessment tasks and criteria for Cantonese speakers.

The criteria modified for Cantonese speakers related to the fact that this language does not have lexical stress (Wong et al., 2020a), but it does have tones. Lexical tone errors have been identified as one of the clinical features of CAS in Cantonese (Wong et al., 2019, 2020a, 2020b), as a parallel for lexical stress in English. Accordingly, "inappropriate prosody, especially in the realization of lexical or phrasal stress," (ASHA, 2007) and poor lexical stress matches (Murray et al., 2015) were modified to relate to lexical tone errors. In addition, "poor polysyllabic production accuracy" in Murray et al. (2015) included both poor segmental and poor prosodic accuracy in polysyllabic words. In this study, only segmental accuracy in polysyllabic words was considered because of reduced prosodic information in Cantonese words, in which every syllable has almost the same duration (Aberchrombie, 1967) and pitch is lexical instead of prosodic. The identification of "accuracy on repetition of [pətəkə] on DDK task" (Murray et al., 2015, p. 53) was also different. Murray et al. adopted the procedure proposed by Robbins and Klee (1987), in which the number of correct three-sound sequence productions was counted in a 3-s interval. In the current study, this procedure was simplified by counting the number of correct phones in each of the sets of five repetitions.

The resulting criteria were the following:

1. Three of the core CAS features from the ASHA (2007) technical report were observed in three of the four assessment tasks, namely, a speech sample, a standardized articulation test, an imitation of polysyllabic words, and DDK tasks. Each feature was required to be observed at least once. They were "inconsistent errors on consonants and vowels in repeated productions of syllables or words," "lengthened and disrupted coarticulatory transitions between sounds and syllables," and "inappropriate lexical tones" (in lieu of "inappropriate lexical stress").
2. Four clinical features modified from Murray et al. (2015) were monitored throughout the assessment: (a) syllable segregation, (b) poor lexical tone matches (in lieu of poor lexical stress matches), (c) reduced accuracy in multisyllabic words, and (d) reduced phone accuracy on DDK tasks. The first two features were observed at least once each in three of the four assessment tasks. The third feature was observed in two of the three tasks, excluding the DDK task, and the last feature was observed in the DDK task at least once.

A working definition was devised in advance for each of the clinical features. Table 2 summarizes the definitions and the tasks used.

TST

All of the children performed the TST after their initial assessment. The TST requires rapid imitation: sets of five repetitions for each of the words and nonwords ranging from one to three syllables in length following the verbal instruction "repeat after me as fast as possible" and a demonstration. A total of 56 sets were given for imitation, with each set containing five repetitions of a stimulus. There were seven stimuli that can be divided into three types: monosyllabic (Tone 1, Tone 2, and Tone 4), disyllabic (Tones 1 and 2, Tones 2 and 4, and Tones 4 and 1), and trisyllabic (Tones 1, 2, and 4). Each stimulus syllable carried one of the three early Cantonese tones. Four V and four CV structures were selected as the syllable structures for the stimuli. Among the 56 sets, there were 24 monosyllabic sets, 24 disyllabic sets, and eight trisyllabic sets. Within the monosyllabic sets, there were 11 word stimuli and 13 nonword stimuli. Within the disyllabic sets, there were two word stimuli (i.e., [ma4ma1] ["mother"]

Table 2. Definitions of clinical features and tasks to confirm childhood apraxia of speech diagnoses.

Clinical features	Definitions	Assessment tasks: Speech sample	HKCAT	DDK	Imitation of polysyllabic words
1. Inconsistent errors on consonants and vowels in repeated productions of syllables or words	Token-to-token inconsistency across productions of the same words in sequence		✓	✓	✓
2. Lengthened and disrupted coarticulatory transitions between sounds and syllables	Syllable segregation—noticeable gaps between syllables Difficulty achieving initial articulatory configurations and transitions into vowels—within-speech groping, false starts, restarts, and hesitation	✓	✓	✓	✓
3. Lexical tone errors[a]	Substitution or distortion of target lexical tones	✓	✓	✓	✓
4. Reduced accuracy in multisyllabic words[a]	PPC in polysyllabic words or sentences < 60%	✓	✓		✓
5. Reduced phone accuracy on DDK task[a]	PPC in stimuli with three different syllables on DDK tasks < 70%			✓	

Note. HKCAT = Hong Kong Cantonese Articulation Test (Cheung et al., 2006); DDK = diadochokinesis; PPC = percentage of phonemes/phones correct.

[a]Diagnostic Criteria 3, 4, and 5 were modified from Murray et al. (2015).

and [pa4pa1] ["father"]) and 22 nonword stimuli. All trisyllabic sets were nonword stimuli. Table 3 shows the V and CV stimuli and the semantic meanings used for generating the 56 sets in the TST.

The first author administered the TST to all children. The 56 sets were presented in eight blocks. Each block included seven sets that were formed of either V or CV structure with three tones. All the children received the blocks in the same order. The seven sets in each block were presented in the same order, which started with three sets of monosyllabic stimuli, followed by three sets of disyllabic stimuli and one trisyllabic stimulus, respectively. Before each set of five imitations, the children were asked to imitate the stimulus once as a baseline. Then, the children were asked to imitate five repetitions of the stimulus as fast as possible. All productions of all children were video-recorded. The average time for completing the TST was about 15 min per child. Fatigue behaviors were not observed in any of the children.

Data Analyses

Perceptual and acoustic analyses were performed to address the research questions. Perceptual analysis was performed to rate production accuracy and consistency. The first author viewed the recordings and transcribed the

Table 3. Stimuli used in the tone-sequencing task.

	Stimulus (semantic meaning, if relevant)			
Monosyllabic sequences				
V structure	[a1] (particle for exclamation) [a2] (dumb) [a4] (tooth/teeth)	[ɔ1] (excretion) [ɔ2] (goose) [ɔ4] (goose)	[u1] [u2] [u4]	[ɛ1] [ɛ2] [ɛ4]
CV structure	[ma1] (mother) [ma2] [ma4] (grandmother)	[pa1] (scar or father) [pa2] (target) [pa4] (father)	[hu1] [hu2] [hu4]	[pu1] [pu2] [pu4]
Disyllabic sequences				
V structure	[a1a2] [a2a4] [a4a1]	[ɔ1ɔ2] [ɔ2ɔ4] [ɔ4ɔ1]	[u1u2] [u2u4] [u4u1]	[ɛ1ɛ2] [ɛ2ɛ4] [ɛ4ɛ1]
CV structure	[ma1ma2] [ma2ma4] [ma4ma1] (mother)	[pa1pa2] [pa2pa4] [pa4pa1] (father)	[hu1hu2] [hu2hu4] [hu4hu1]	[pu1pu2] [pu2pu4] [pu4pu1]
Trisyllabic sequences				
V structure	[a1a2a4]	[ɔ1ɔ2ɔ4]	[u1u2u4]	[ɛ1ɛ2ɛ4]
CV structure	[ma1ma2ma4]	[pa1pa2pa4]	[hu1hu2hu4]	[pu1pu2pu4]

Note. Within the 28 V stimuli, there are six word stimuli (with semantic meanings) and 22 nonword stimuli. Within the 28 CV stimuli, there are seven word stimuli (with semantic meanings) and 21 nonword stimuli. V = vowel; CV = consonant–vowel.

productions perceptually. Accuracy was recorded as the PTC, which was calculated by the following equation:

$$\frac{\text{number of correct tones}}{\text{total number of tones produced}} \times 100. \quad (1)$$

Production consistency was rated using the scoring system proposed by Williams and Stackhouse (2000), in which 4 points and 1 point were given for highly consistent and highly inconsistent productions, respectively, with 3 points and 2 points as intermediate scores. The five repetitions of the stimulus in each set were compared with the baseline one by one to determine the consistency level (see Williams & Stackhouse, 2000, p. 277, for more details). An overall consistency score was given for each set of productions.

To determine the inter- and intrarater reliability of the perceptual ratings of tone, the independent rater and the first author rerated 27.7% of all TST tone productions by viewing the video recording 3 months after the initial rating to minimize possible bias from the initial rating. All of the items for the reliability measures were selected randomly. A two-way random intraclass correlation coefficient [ICC (2, 1)] was calculated for interrater reliability between the first author and the independent rater, and a two-way mixed ICC (3, 1) was calculated for the first author's intrarater reliability for perceptual ratings of tone accuracy. High degrees of inter- and intrarater reliability were found based on 140 repeated scores (27.7% of all scores). The average measure ICC (2, 1) was .987, with a 95% confidence interval of .982 to .991, $F(139, 139) = 75.223$, $p < .001$, for interrater reliability between the first author and the independent rater. The average measure ICC (3, 1) was .968, with a 95% confidence interval of .955 to .977, $F(139, 139) = 31.165$, $p < .001$, for the intrarater reliability of the first author.

Acoustic analysis was performed to measure the F0 values and duration of each set (repetition duration) using Praat software (Version 6.1.16; Boersma & Weenink, 2020). The audio segments of the videos were imported to the Praat software. By following the procedure suggested by Ma et al. (2006), the vocalic segment of each syllable was identified manually from a wideband spectrogram and an amplitude waveform visually displayed in the software. F0 was estimated at five evenly spaced time points from the beginning to the end of the vocalic segment of each syllable using an autocorrelation algorithm (0%, 25%, 50%, 75%, and 100% of the total duration of each syllable). No manual corrections were conducted to extract F0 contour. The repetition duration was measured by calculating the difference in time between the 0% time point of the first syllable (beginning of the set) and the 100% time point of the last syllable (end of the set).

Since an autocorrelation algorithm was used with no manual corrections, inter- and intrareliability of manual vowel segmentations was measured by the strength of association of the generated F0 values. Two out of the nine (22%) participants' productions on the TST task were randomly selected for the first author to redo the manual vowel segmentation 1 month after initial segmentation. A different, two-out-of-the-nine participants' performance on the TST task was randomly selected for an independent research assistant, who had experience conducting acoustic analysis of Cantonese speech and was blind to the purpose of this study to perform manual vowel segmentations. A Pearson product–moment correlation was run to determine the relationship between the F0 values generated by the first author and between the first author and the independent research assistant. There was a strong, positive correlation between the values of the two manual analyses of the first author, which was statistically significant ($r = .946$, $n = 4,800$, $p < .01$). A strong and positive correlation between the first author and the independent research assistant was also statistically significant ($r = .917$, $n = 4,780$, $p < .01$).

Twelve stimuli ([a1], [a2], [a4], [u1], [u2], [u4], [pa1], [pa2], [pa4], [pu1], [pu2], and [pu4]) were selected to examine the effects of syllable structure and lexical status in order to balance the number of stimuli for data analyses. Three-way analyses of variance (ANOVAs; Group × Structure × Lexical Status) were used to examine the main and interaction effects on two perceptual dependent variables (PTC, consistency score) and one acoustic dependent variable (duration). Group (CAS, nCAS, and TD) was the between-subjects factor; syllable structure (V vs. CV) and lexical status (words vs. nonwords) were the within-subject factors. Kruskal–Wallis tests were performed to determine the between-groups differences when significant group differences were found on the PTC, consistency scores, and/or repetition durations, which showed main effects.

Apart from the 12 monosyllabic stimuli, 12 additional disyllabic stimuli ([a1a2], [a2a4], [a4a1], [u1u2], [u2u4], [u4u1], [pa1pa2], [p2pa4], [pa4pa1], [pu1pu2], [pu2pu4], and [pu4pu1]) were added to examine the effects of tone on F0 values, for a total of 24 stimuli. Two-way mixed ANOVAs (Tone × Time Point) were used to examine the main and interaction effects of three Cantonese tones (i.e., Tone 1, Tone 2, and Tone 4) on F0 values across five time points for each group of participants. Two-way mixed ANOVAs (Group × Time Point) were used to examine the main and interaction effects of group (i.e., CAS, nCAS, and TD) on F0 values across five time points for each Cantonese tone syllable. The F0 values of each syllable from all the monosyllabic and disyllabic stimuli were divided according to the targeted tone (i.e., Tone 1, Tone 2, and Tone 4). Since the positions of the tones in trisyllabic stimuli were fixed and unbalanced (i.e., Tone 1 was the first syllable, Tone 2 was the second syllable, and Tone 4 was the last syllable), we excluded the trisyllabic stimuli from this analysis. Kruskal–Wallis tests and post hoc tests were performed to identify possible main effects on between-tones differences across five time points within groups and on tone differences across five time points between groups.

Results

Table 4 lists the means and standard deviations of three dependent variables (i.e., PTC, consistency scores,

Table 4. Descriptive statistics for three dependent variables in response to different types of stimuli in different groups.

Dependent variables	Structure	Lexical status	*M* (*SD*)		
			CAS	nCAS	TD
PTC (%)	V	W	44.44 (45.58)	91.11 (26.67)	100.00 (0.00)
		NW	51.11 (41.37)	86.67 (28.28)	100.00 (0.00)
	CV	W	42.22 (29.06)	95.56 (13.33)	100.00 (0.00)
		NW	37.78 (39.30)	91.11 (26.67)	100.00 (0.00)
Consistency scores (%)	V	W	77.78 (23.20)	97.22 (8.33)	100.00 (0.00)
		NW	69.44 (24.30)	91.67 (17.68)	100.00 (0.00)
	CV	W	69.44 (11.02)	97.22 (8.33)	100.00 (0.00
		NW	58.33 (25.00)	94.44 (16.67)	100.00 (0.00)
Duration (s)	V	W	2.79 (1.21)	1.81 (0.44)	1.37 (0.28)
		NW	2.54 (1.23)	2.00 (0.60)	1.58 (1.03)
	CV	W	2.55 (0.59)	1.92 (0.50)	1.02 (0.13)
		NW	2.01 (1.05)	1.58 (0.22)	1.11 (0.09)

Note. CAS = childhood apraxia of speech; nCAS = non-CAS speech and/or language impairment; TD = typical development; PTC = percentage of correct tones; V = vowel; W = word; NW = nonword; ; CV = consonant–vowel.

and repetition duration) in response to different aspects of the stimuli (i.e., syllable structure and lexical status) in the three participant groups (i.e., CAS, nCAS, and TD).

The results of Leven's test of equality of error variances showed that the assumption of homogeneity of variance had been violated ($p < .005$), which means that the null hypothesis is incorrect and the variances are significantly different. Normal quantile transformation was performed using Blom's formula, followed by three-way ANOVAs on the normal scores (Conover & Iman, 1981).

Between-Groups Differences on Tone Accuracy, Consistency, and Repetition Duration

Significant main effects of group were observed on PTC, $F(2, 96) = 63.791$, $p < .005$; consistency scores, $F(2, 96) = 62.225$, $p < .005$; and repetition duration, $F(2, 96) = 41.585$, $p < .005$. Table 5 lists the results of the Kruskal–Wallis test. Significant between-groups differences were found with respect to PTC, consistency scores, and repetition duration. Specifically, the CAS group performed significantly lower with respect to PTC and consistency scores than the nCAS and TD groups. They also demonstrated significantly longer repetition durations than the nCAS and TD groups. Moreover, the nCAS group demonstrated significantly longer repetition durations than the TD group, with no significant differences on PTC or consistency scores.

Effects of Syllable Structure and Lexical Status

A significant main effect of structure was found on repetition duration, $F(1, 96) = 8.236$, $p < .05$. This suggests that children used a shorter time to produce five repetitions of CV structures than V structures when the effects of lexical status and the presence of CAS were controlled, despite the fact that the consonants were included in the duration measurements. No significant differences between V and CV structures were found with respect to PTC, $F(1, 96) = 0.024$, $p > .05$, or consistency scores, $F(1, 96) = 1.058$, $p > .05$. No interaction effects between group and structure were observed with respect to PTC, $F(2, 96) = 0.404$, $p > .05$; consistency scores, $F(2, 96) = 1.890$, $p > .05$; or repetition duration, $F(2, 96) = 1.558$, $p > .05$.

No significant differences between words and nonwords were found with respect to PTC, $F(1, 96) = 0.007$, $p > .05$; consistency scores, $F(1, 96) = 1.732$, $p > .05$; or repetition duration, $F(1, 96) = 1.140$, $p > .05$. No interaction effects between group and lexical status were observed with respect to PTC, $F(2, 96) = 0.326$, $p > .05$; consistency

Table 5. Results of Kruskal–Wallis test of percentage of tones correct (PTC), consistency score, and repetition duration among the participant groups.

Dependent variables	*H*	Post hoc test		
		CAS vs. nCAS	CAS vs. TD	nCAS vs. TD
PTC	62.881^	–38.000^	–45.250^	–7.250
Consistency scores	62.706^	–37.583^	–45.042^	–7.458
Repetition duration	47.365^	34.500^	49.542^	15.042*

Note. CAS = childhood apraxia of speech; nCAS = non-CAS speech and/or language impairment; TD = typical development.

^$p < .01$. *$p < .05$.

scores, $F(2, 96) = 0.546$, $p > .05$; or repetition duration, $F(2, 96) = 2.507$, $p > .05$.

No interaction effects between structure and lexical status were observed with respect to PTC, $F(1, 96) = 0.001$, $p > .05$; consistency scores, $F(1, 96) = 0.089$, $p > .05$; or repetition duration, $F(1, 96) = 0.413$, $p > .05$. No interaction effects among group, structure, and lexical status were observed with respect to PTC, $F(2, 96) = 0.039$, $p > .05$; consistency scores, $F(2, 96) = 0.038$, $p > .05$; or repetition duration, $F(2, 96) = 1.626$, $p > .05$.

Effects of Tone on F0 Values

Figure 1 shows the mean F0 values of the three Cantonese tones across five time points in the participant groups, with Tone 1 on the left, Tone 2 in the middle, and Tone 4 in the right column. Two-way mixed ANOVAs (Tone × Time Point) were run to examine the main and interaction effects of three Cantonese tones (i.e., Tone 1, Tone 2, and Tone 4) on F0 values across five time points for each group of participants. The significance values of Mauchly's test of sphericity ($p > .05$) suggest that the assumption of sphericity was violated. Greenhouse–Geisser corrections were applied. Significant main effects of tones on F0 values across time were found for the CAS group, $F(2, 537) = 21.836$, $p < .001$; the nCAS group, $F(2, 537) = 31.806$, $p < .001$; and the TD group, $F(2, 537) = 38.815$, $p < .001$. This suggests that all of the groups produced different F0 values to indicate the three different Cantonese tones. Significant main effects of time on F0 values were found for the nCAS group, $F(2.728, 1464.969) = 3.512$, $p < .05$, and the TD group, $F(2.282, 1225.319) = 4.436$, $p < .05$, but not for the CAS group, $F(2.039, 1094.918) = 1.708$, $p > .05$. This implies that the nCAS and TD groups produced significantly different F0 values across the five time points, but the CAS group did not. A significant interaction effect between tone and time was found for the nCAS group, $F(5.456, 1464.969) = 15.118$, $p < .001$, and the TD group, $F(4.564, 1225.319) = 17.257$, $p < .001$, but not for the CAS group, $F(4.078, 1094.918) = 1.050$, $p > .05$. This suggests, again, that the nCAS and TD groups produced significantly different F0 values across time to indicate the three Cantonese tones, but the CAS group did not. A Bonferroni test was performed to examine if there were significant differences among the tone syllables produced by the CAS group. The results showed that Tone 1 and Tone 2 syllables were not significantly different ($p > .05$) with respect to F0 in the CAS group. However, the F0 values of Tone 1 and Tone 4 syllables and Tone 2 and Tone 4 syllables were significantly different ($p < .01$) in the CAS group. The results of the Kruskal–Wallis test and pairwise comparisons are summarized in Table 6. They confirm that the F0 values produced by the CAS group for Tone 1 and Tone 2 syllables were not significantly different.

Two-way mixed ANOVAs (Group × Time Point) were used to examine the main and interaction effects of three groups (i.e., CAS, nCAS, and TD) on F0 values across

Figure 1. Mean fundamental frequency (F0) values of the three Cantonese tones across five time points in the participant groups. CAS = childhood apraxia of speech; nCAS = non-CAS speech and/or language impairment; TD = typical development.

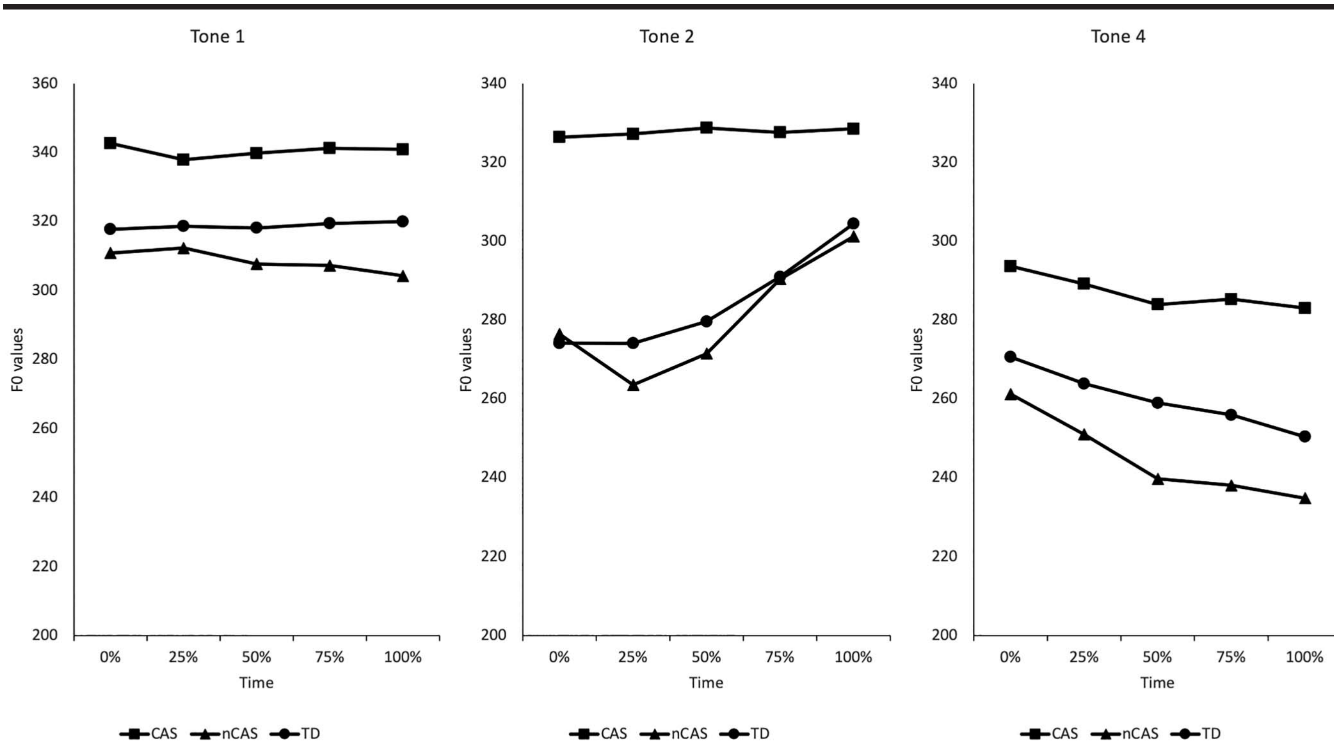

Table 6. Results of Kruskal–Wallis test of the fundamental frequency values across five time points in the childhood apraxia of speech (CAS) group.

			Post hoc test (pairwise comparison)		
Group	Time point	*H*	Tone 1 vs. Tone 2	Tone 1 vs. Tone 4	Tone 2 vs. Tone 4
CAS	0%	30.006^	1.679	5.355^	3.676^
	25%	45.155^	2.430	6.641^	4.211^
	50%	58.748^	2.494	7.524^	5.030^
	75%	46.102^	2.109	6.644^	4.534^
	100%	43.235^	1.855	6.391^	4.535^

^$p < .01$.

five time points for each tone syllable. The significance values of Mauchly's test of sphericity ($p > .05$) suggested that the assumption of sphericity was violated. Greenhouse–Geisser corrections were applied. Significant main effects of group on F0 values were found with respect to Tone 1 syllables, $F(2, 537) = 7.289$, $p < .01$; Tone 2 syllables $F(2, 537) = 19.970$, $p < .01$; and Tone 4 syllables, $F(2, 537) = 8.678$, $p < .01$. This indicates that the F0 values of all three tone syllables were significantly different among groups. Specifically, the CAS group produced significantly higher F0 values than the control groups on all three tone types. Significant main effects of time point on F0 values were found with respect to Tone 2, $F(2.438, 1309.338) = 22.737$, $p < .01$, and Tone 4 syllables, $F(2.453, 1317.310) = 28.511$, $p < .01$, but not for Tone 1 syllables, $F(2.266, 1216.949) = 0.168$, $p > .01$. This implies that the F0 values of Tone 2 and Tone 4 syllables across five time points were significantly different among the participants, irrespective of their diagnoses. No interaction effects between group and time point were found with respect to Tone 1 syllables, $F(4.532, 1216.949) = 0.168$, $p > .05$, or Tone 4 syllables, $F(4.906, 1317.310) = 1.446$, $p > .05$. The Group × Time Point interaction was significant with respect to Tone 2 syllables, $F(4.876, 1309.338) = 7.130$, $p < .01$, indicating that F0 values differed among the groups across the five time points when Tone 2 productions were required. The results of the Kruskal–Wallis tests and post hoc tests, summarized in Table 7, showed that the F0 values produced by the CAS group significantly differed from those of the nCAS and TD groups on each of the five time points with respect to Tone 2 syllables. Specifically, the CAS group produced less varied F0 values on Tone 2 syllables across five time points when compared with the control groups. The Tone 2 (high-rising) syllables were produced similarly to Tone 1 (high-level) syllables. The variations of the F0 values produced by the CAS group on Tone 1 and Tone 4 (low-falling) syllables across five time points were not significantly different from those of the control groups.

Discussion

Pitch Variation Skills in Children With CAS

The aim of this study was to use the TST to examine the pitch variation skills of Cantonese-learning children with CAS. The results showed that children with CAS performed more poorly than the control groups on the TST with respect to tone accuracy and consistency, with statistically significant differences in both perceptual (PTC and consistency scores) and acoustic data (repetition duration). Specifically, children with CAS produced five repetitions within each stimulus set with statistically significantly lower PTC and consistency scores and longer repetition durations than the control groups. Thus, the first hypothesis was supported. The results suggest that the children with CAS had more difficulty varying their vocal pitch for tone than children without CAS. This difficulty was evidenced not only in their lower accuracy rates (PTC) but also in lower consistency and longer repetition durations. The fact that their productions were slower suggests that it may have taken them longer to coordinate segments with tones, although children with CAS tend to speak more slowly in intonation languages as well, so it is also possible that the longer durations were not specifically related to the tones. The results were consistent with those of the previous pilot study (Wong et al., 2019) and provide support for the documented expert opinion that holds that Cantonese-speaking children with CAS have difficulty in producing Cantonese tones (Wong et al., 2020a). It is notable that the current result is not consistent with the findings of Kopera and Grigos (2020), in which

Table 7. Results of Kruskal–Wallis test of the fundamental frequency values across five time points on Tone 2 syllable.

			Post hoc test (pairwise comparison)		
Syllable	Time point	*H*	CAS vs. nCAS	CAS vs. TD	nCAS vs. TD
Tone 2	0%	79.138^	7.235^	8.100^	0.865
	25%	89.998^	7.737^	8.623^	0.886
	50%	73.500^	7.062^	7.740^	0.678
	75%	34.808^	4.792^	5.376^	0.584
	100%	15.901^	3.443^	3.464^	0.021

Note. CAS = childhood apraxia of speech; nCAS = non-CAS speech and/or language impairment; TD = typical development.
^$p < .01$.

no significant between-groups differences were found on the acoustic variables (F0 and duration) when they were studied independently using the pairwise variability index. This may be due to differences between lexical stress and tones. Stress patterns vary when pitch, intensity, and duration vary simultaneously, while tones vary only by pitch. The current results imply that tones may provide a medium to demonstrate the challenge of producing pitch variation on its own for children with CAS. Lastly, the differences between the CAS and control groups suggest that further studies of the TST should be carried out to further explore its usefulness for identifying tone differences between children with CAS and without CAS. The TST may also be used to investigate pitch variation skills independently in children, apart from the pairwise variability index used previously (Kopera & Grigos, 2020). Further investigations applying similar tasks to English-speaking children with CAS may enrich our understanding of more specific aspects of prosody control (i.e., pitch uncoupled from other aspects of lexical stress) in children learning intonation languages.

Effects of Linguistic Elements on Pitch Variation

This study aimed to examine three factors that might affect pitch variation abilities in children with CAS by testing three specific hypotheses (Hypotheses 2–4). The second hypothesis concerned the effects of syllable structure. We predicted that there would be no phonotactic effect on the control groups but that there would be an impact on the CAS group, with these children performing better on V stimuli than on CV stimuli. No within-group effects were documented for phonotactic impacts on perceptual ratings with respect to either PTC or consistency scores, but there were phonotactic effects on acoustic durations of stimulus repetitions. The participants produced shorter durations on CV stimuli than V stimuli. Irrespective of their diagnosis, the children produced five repetitions of CV stimuli with shorter durations than the same number of repetitions of V stimuli. This is consistent with the finding that English-speaking adults produce CV syllables faster than V syllables (Mackay, 1974). It was observed that the children produced the sequences of V syllables with pauses between syllables, while no such pauses were observed on the sequences of CV syllables. No interaction effect between group and syllable structure was found. This suggests that the impact of syllable structure is similar for the CAS and control groups with respect to tone accuracy, consistency, and repetition duration between V and CV stimuli. Thus, the second hypothesis was not supported. It is possible that the difference between V and CV stimuli is subtle for a task of this sort and does not reflect the phonotactic difficulties demonstrated by children with CAS on other tasks.

The third hypothesis concerned the effects of lexical status. We predicted that lexical knowledge would facilitate pitch variation skills in both the CAS and control groups as compared with performance in response to nonlexical stimuli. No main effect was found for perceptual ratings of PTC, consistency scores, or acoustic duration. In a manner similar to the analyses of the responses to stimuli with different syllable structures, the effects of lexical status on the responses to stimuli were examined. The results suggest that children produce lexical and nonlexical tones with no significant differences in accuracy, consistency, or repetition duration. The third hypothesis was not supported. This result is not consistent with the fact that lexical knowledge has previously been shown to have a priming effect on tone production in adult Cantonese speakers (Zhao & Jurafsky, 2007). This may reflect a difference between child and adult tone productions. In addition to addressing lexical knowledge, our study addressed speech movement sequencing skills. Given that there is a lack of understanding of the relationship between linguistic elements and speech motor control, further investigation is needed to interpret the results from this study.

This study also aimed to investigate the effects of tone on pitch variation skills. The main effects of time on F0 values suggest that the pitch patterns of Tone 2 and Tone 4 syllables across time are significantly different among children, irrespective of their diagnoses. This provides evidence for using these Cantonese tones to examine pitch variation skills in children. In the last hypothesis, we predicted that stimuli with rising and falling tones, especially rising tones, would challenge the children with CAS, whereas those in the control groups would not have significant difficulty with these. The insignificant main effect of tone on F0 values across time and insignificant interaction effect between tone and time points on F0 values across time for the CAS group only suggested that this group produces Tone 1 and Tone 2 similarly, while the control groups produce Tone 1 and Tone 2 differently. Specifically, children with CAS tend to use the high-level Tone 1 to substitute for the high-rising Tone 2. The main effects of group on F0 values showed that children in different groups produced Tone 1, Tone 2, and Tone 4 syllables differently. Specifically, children with CAS produce higher F0 values than those without CAS. Interaction effects between group and time were found on Tone 2 syllables. The CAS group produced Tone 2 syllables significantly differently than the control groups. Less variation of F0 values was observed on CAS Tone 2 syllables across time points (see Figure 1). The results suggest that children with CAS have difficulty in producing high-rising tones (Tone 2), but not high-level (Tone 1) and low-falling tones (Tone 4), when compared with children without CAS. Specifically, children with CAS are not able to raise F0 values to produce a rising contour but are able to lower or maintain F0 values to produce falling and level tones, respectively. From a motor perspective, children with CAS may be unable to coordinate movements of the oral articulators with simultaneous laryngeal movements to achieve specific types of appropriate pitch variation within a syllable. The results of this study are consistent with the finding that production of rising pitch patterns is more difficult than producing falling tones for typically developing English-speaking young children (Snow, 1998).

Other Observations

Apart from the three possible factors listed here, it was observed that the levels of tone identification skills differed among the children within the different groups. Given that both tone perception and production abilities contribute to tone acquisition in Cantonese-speaking children (Mok et al., 2019), it is important to investigate a possible relationship between tone perception and production in children with CAS. In addition, although the accuracy of segments and syllable structures were not analyzed in detail in this study, it was observed that the children with CAS produced accurate tones with inaccurate segments, inaccurate tones with accurate segments, and both inaccurate tones and segments on all of the tasks, independent of the types of stimuli presented. This suggests that these speakers with CAS were having difficulty in simultaneously managing both segments and tones. Further investigations should include more detailed error analyses of the syllable structures, phonemes, and tones produced to understand potential patterns that may relate to existing speech production models.

Clinical Implications

The results of this study imply that the TST could be a potentially useful tool to assist in the diagnosis of CAS in terms of tone accuracy, consistency, and duration. The statistically significant differences on PTC, consistency scores, and repetition duration between the CAS and control groups suggest that the TST may be able to help differentiate children with CAS from those without CAS. The significant differences on tone accuracy, consistency, and repetition duration among the CAS and control groups support the claim that perceptual accuracy and consistency should be taken into consideration, along with acoustic duration for repetition tasks, as suggested by Williams and Stackhouse (2000), who demonstrated that accuracy and consistency measures are more sensitive than rate of production of syllable repetitions for preschool children. Moreover, the insignificant differences between V and CV stimuli with respect to PTC and consistency scores suggest that isolated vowels could be used as test stimuli for perceptual assessment of children with extremely limited verbal output.

Generalization of the current results to clinical practice with varied populations should be carried out with caution. Aside from the small sample size and other limitations described above and below, the TST in this study was administered to Cantonese-speaking children who had had exposure to a tone language and whose tone production skills had been nurtured from birth. It is therefore unknown whether the same results would be obtained from using this tool with speakers of nontonal languages, such as English-speaking children with CAS. English-speaking adults with no experience of tonal languages imitate Cantonese-level tones with appropriate pitch levels (Wu et al., 2016). Putting this claim and the results from this study together, we can hypothesize that the difficulty of producing appropriate pitch levels for Cantonese tones may also be present in English speakers with CAS. However, other factors, such as tone perception abilities and the differences between Cantonese tones and English stress, should be considered before testing English-speaking children with the TST. Any difficulty that speakers from nontonal languages have with producing tones may be due to the challenges of perceiving tones rather than a deficit in the production of tones. Also, productions of Cantonese tones by English speakers would require divorcing pitch from other aspects of lexical stress, such as syllable duration and loudness. This may impact the results. Therefore, at this time, it is not possible to judge whether this tool could be used to differentiate English-speaking children with CAS from those without CAS. Future investigation of pitch variation in other languages is recommended prior to implementation of the TST in clinical settings.

Limitations and Future Directions

There were only three children with CAS in this study. A larger sample size is recommended in future studies to obtain representative results for generalization. In addition, expressive language assessments were limited. Formal expressive language assessments were not done with the children in the TD group. Inconsistent informal expressive language measures were used with those children in the CAS and nCAS groups who had severe SSD. Also, the language samples collected were shorter than deemed appropriate for mean length of utterance calculation. Moreover, the nCAS group consisted of children with speech and/or language impairment. Further investigations should divide these children into two groups, one with speech impairment only and one with language impairment only, to have a clearer understanding of the impacts of speech and language skills on pitch variation skills. Another limitation is that we examined the effects of lexical status by comparing the children's performance on word versus nonword stimuli. However, all the words in this study had [a] or [ɔ] as the vowel, while all the nonwords had [u] or [ɛ]. Children's production performance might have been affected by this vowel difference. Future investigations should have a more balanced set of stimuli. Though both V and CV stimuli can be used for perceptual and acoustic TST-based assessment, further investigation should involve more complex syllabic structures (e.g., CVC) to explore the effects of syllable structure on pitch variation in more depth. Apart from the effects of syllable structure, future studies of the effects of lexical status on speech movement sequencing skills are also needed. Moreover, as mentioned, interactions among correctness of productions of syllable structures, phonemes, and tone identification skills were not analyzed in detail in this study. Analyses of possible relationships between tone perception and production should be carried out. Lastly, future investigations across languages are warranted to determine whether difficulty with pitch variation within a context unrelated to lexical stress is experienced by CAS speakers when using different languages. A future validation study of the TST for differentiating CAS from other types of speech and language disorders in nontone languages would be useful.

Conclusions

This study provides preliminary findings on the pitch variation skills of Cantonese-speaking children with CAS. The results suggest that children with CAS have difficulty with pitch variation. In particular, children with CAS produced five repetitions of different stimuli with significantly lower PTC, consistency scores, and longer repetition durations than the control groups. Main effect of syllable structure on repetition duration was found, while no interaction effects between group and these factors were found. An insignificant difference of F0 values across time between Tone 1 and Tone 2 syllables in the CAS group and significant interaction effects between group and time points of F0 values found on Tone 2 syllables suggest that children with CAS produce less varied F0 values on high-rising Tone 2 syllables. More specifically, children with CAS tend to substitute Tone 2 syllables with high-level Tone 1 syllables.

Clinically, the TST may be a useful tool to assist in the diagnosis of CAS, as suggested by the statistically significant differences observed between the CAS and control groups on TST-based assessments of tone accuracy, consistency, and duration in this limited study. Moreover, the insignificant difference between CV and V stimuli suggests that V stimuli on the TST may be used to test young or severely impaired children. Future investigations should include larger samples, complex syllable structures, tone perception analyses, and/or children who speak different languages. Further validation of the TST tool for the diagnosis of children with CAS is also indicated.

Acknowledgments

This study was funded by the Office of Research and Knowledge Transfer Services at The Chinese University of Hong Kong and the Social Innovation and Entrepreneurship Development Fund, which was granted to the first, third, and last authors. The reference number is KPF18HLF26. The authors are sincerely grateful to all of the participants and their families as well as to Christine So and Miss Yiting Chen for completing the interrater reliability measures.

References

Aberchrombie, D. (1967). *Elements of general phonetics*. Edinburg University Press.

American Academy of Audiology. (1997). *Identification of hearing loss & middle-ear dysfunction in preschool & school-age children* [Position statements]. https://www.audiology.org/publications-resources/document-library/identification-hearing-loss-middle-ear-dysfunction-preschool

American Speech-Language-Hearing Association. (2007). *Childhood apraxia of speech* [Technical report]. http://www.asha.org/policy

Boersma, P., & Weenink, D. (2020). *Praat 6.1.16: A system for doing phonetics by computer* [Computer software]. University of Amsterdam. https://www.fon.hum.uva.nl/praat/

Chen, T. H., & Massaro, D. W. (2008). Seeing pitch: Visual information for lexical tones of Mandarin-Chinese. *The Journal of the Acoustical Society of America, 123*(4), 2356–2366. https://doi.org/10.1121/1.2839004

Cheung, P., Ng, K. H., & To, C. (2006). *Hong Kong Cantonese Articulation Test*. Language Information Sciences Research Centre and City University of Hong Kong.

Conover, W., & Iman, R. L. (1981). Rank transformation as a bridge between parametric and nonparametric statistics. *The American Statistician, 35*(3), 124–129. https://doi.org/10.2307/2683975

Iuzzini-Seigel, J., & Murray, E. (2017). Speech assessment in children with childhood apraxia of speech. *Perspectives of the ASHA Special Interest Groups, 2*(2), 47–60. https://doi.org/10.1044/persp2.SIG2.47

Kopera, H. C., & Grigos, M. (2020). Lexical stress in childhood apraxia of speech: Acoustic and kinematic findings. *International Journal of Speech-Language Pathology, 22*(1), 12–23. https://doi.org/10.1080/17549507.2019.1568571

Lee, K. Y. S. (2012). *Hong Kong Cantonese Tone Identification Test*. Department of Otorhinolaryngology, Head and Neck Surgery, The Chinese University of Hong Kong.

Lee, K. Y. S., Chiu, S. N., & van Hasselt, C. A. (2002). Tone perception ability of Cantonese-speaking children. *Language and Speech, 45*(4), 387–406. https://doi.org/10.1177/00238309020450040401

Lee, K. Y. S., Lee, L. W., & Cheung, P. S. P. (2009). *Hong Kong Cantonese Receptive Vocabulary Test*. Institute of Human Communicative Research, The Chinese University of Hong Kong.

Liang, B., & Du, Y. (2018). The functional neuroanatomy of lexical tone perception: An activation likelihood estimation meta-analysis. *Frontiers in Neuroscience, 12,* 495. https://doi.org/10.3389/fnins.2018.00495

Ma, J. K., Ciocca, V., & Whitehill, T. L. (2006). Effect of intonation on Cantonese lexical tones. *The Journal of the Acoustical Society of America, 120*(6), 3978–3987. https://doi.org/10.1121/1.2363927

Mackay, D. G. (1974). Aspects of the syntax of behavior: Syllable structure and speech rate. *Quarterly Journal of Experimental Psychology, 26,* 642–657. https://doi.org/10.1080/14640747408400457

Mok, P. P. K., Fung, H. S. H., & Li, V. G. (2019). Assessing the link between perception and production in Cantonese tone acquisition. *Journal of Speech, Language, and Hearing Research, 62*(5), 1243–1257. https://doi.org/10.1044/2018_JSLHR-S-17-0430

Murray, E., McCabe, P., Heard, R., & Ballard, K. J. (2015). Differential diagnosis of children with suspected childhood apraxia of speech. *Journal of Speech, Language, and Hearing Research, 58*(1), 43–60. https://doi.org/10.1044/2014_JSLHR-S-12-0358

Reynell, J., & Huntley, M. (1987). *Reynell Developmental Language Scales–Cantonese (Hong Kong Version)*. NFEF-Nelson.

Robbins, J., & Klee, T. (1987). Clinical assessment of oropharyngeal motor development in young children. *Journal of Speech and Hearing Disorders, 52*(3), 271–277. https://doi.org/10.1044/jshd.5203.271

Shriberg, L. D., Campbell, T. F., Karlsson, H. B., Brown, R. L., McSweeny, J. L., & Nadler, C. J. (2003). A diagnostic marker for childhood apraxia of speech: The lexical stress ratio. *Clinical Linguistics & Phonetics, 17*(7), 549–574. https://doi.org/10.1080/0269920031000138123

Shriberg, L. D., Potter, N. L., & Strand, E. A. (2011). Prevalence and phenotype of childhood apraxia of speech in youth with galactosemia. *Journal of Speech, Language, and Hearing*

Research, 54(2), 487–519. https://doi.org/10.1044/1092-4388(2010/10-0068)
Snow, D. (1998). Children's imitations of intonation contours: Are rising tones more difficult than falling tones? *Journal of Speech, Language, and Hearing Research, 41*(3), 576–587. https://doi.org/10.1044/jslhr.4103.576
Terband, H., Namasivayam, A., Maas, E., van Brenk, F., Mailend, M. L., Diepeveen, S., van Lieshout, P., & Maassen, B. (2019). Assessment of childhood apraxia of speech: A review/tutorial of objective measurement techniques. *Journal of Speech, Language, and Hearing Research, 62*(8S), 2999–3032. https://doi.org/10.1044/2019_JSLHR-S-CSMC7-19-0214
Williams, P., & Stackhouse, J. (2000). Rate, accuracy and consistency: Diadochokinetic performance of young, normally developing children. *Clinical Linguistics & Phonetics, 14*(4), 267–293. https://doi.org/10.1080/02699200050023985
Wong, E. C. H., Lee, K. Y. S., & Tong, M. C. F. (2019). *Tone sequencing in childhood apraxia of speech: A pilot study* [Paper presentation]. 31st World Congress of the IALP, Taipei, Taiwan.
Wong, E. C. H., Lee, K. Y. S., & Tong, M. C. F. (2020a). 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
Wong, E. C. H., Lee, K. Y. S., & Tong, M. C. F. (2020b). *Sequencing lexical tones in childhood apraxia of speech: Preliminary results* [Paper presentation]. 2020 Motor Speech Conference, Santa Barbara, CA, United States.
Wu, M., Baker, B., Fletcher, J., & Bundgaard-Nielsen, R. (2016). How pitch moves: Production of Cantonese tones by speakers with different tonal experiences. *Proceedings of the 5th International Symposium on Tonal Aspects of Languages, 2016,* 133–137. https://doi.org/10.21437/TAL.2016-29
Yip, V., & Matthews, S. (2011). *Cantonese: A comprehensive grammar* (2nd ed.). Routledge.
Zhao, Y., & Jurafsky, D. (2007). The effect of lexical frequency on tone production. In J. Trouvain & W. J. Barry (Eds.), *Proceedings of the 16th International Congress of Phonetic Sciences* (pp. 477–480). International Phonetic Association.