Article

FIRST LANGUAGE

First Language
2022, Vol. 42(5) 670–692

DOI: 10.1177/01427237221109342
journals.sagepub.com/home/fla
SAGE

The effect of tone hyperarticulation in Cantonese infant-directed speech on toddlers' word recognition in the second year of life

Luchang Wang
Xi'an Jiaotong-Liverpool University, China; The Chinese University of Hong Kong, China

René Kager
Utrecht University, The Netherlands

Patrick C. M. Wong
The Chinese University of Hong Kong, China

Abstract

The acoustic properties of infant-directed speech (IDS) have been widely studied, but whether and how young learners' language development benefits from individual properties remains to be confirmed. This study investigated whether toddlers' word processing was affected by tone hyperarticulation in the IDS of a tone language. Nineteen- and 23-month-old Cantonese-learning toddlers completed a familiar word recognition task and were tested (a) in the hyperarticulated-tone (HT) condition in which the tonal distances were exaggerated, and (b) in the non-hyperarticulated-tone (NT) condition with smaller tonal distances that resembled those in adult-directed speech. The 19-month-old toddlers performed significantly better in the HT condition than in the NT condition, while the 23-month-olds performed comparably well in both conditions. These findings suggest that tone language learners' word recognition can be facilitated by tone hyperarticulation in IDS, in the middle of the second year of life; as their language development proceeds, this facilitatory effect appears to largely diminish by the end of the second year of life.

Corresponding author:
Patrick C. M. Wong, Department of Linguistics and Modern Languages, The Chinese University of Hong Kong, G03, Leung Kau Kui Building, Hong Kong SAR, China.
Email: p.wong@cuhk.edu.hk

Keywords
Infant-directed speech, lexical tones, hyperarticulation, word recognition, Cantonese

Introduction

It has been found across numerous world languages (e.g. English, French, German, Japanese, Thai, and Chinese, to name but a few) that a special speech register, known as infant-directed speech (IDS; also parentese or baby talk), is often used when adults speak to young children (Fernald et al., 1989; Grieser & Kuhl, 1988; Kitamura et al., 2002). IDS is distinct from adult-directed speech (ADS) in many aspects. Prosodically, speakers tend to produce utterances in IDS with higher pitch, larger pitch range, and greater pitch variability (Fernald et al., 1989; Fernald & Simon, 1984). Moreover, acoustic modifications in IDS can also affect the realization of phonemically relevant information. For instance, an expansion of acoustic vowel space (i.e. the area of the vowel triangle determined by the first and second formants of the three most peripheral vowels /a/, /i/, and /u/) has been widely found in IDS compared with ADS (often called 'vowel hyperarticulation') in a variety of languages, such as English, French, Japanese, and Mandarin Chinese (Burnham et al., 2002; Dodane & Al-Tamimi, 2007; Kuhl et al., 1997; Liu et al., 2003; cf. Cristia & Seidl, 2014; Englund & Behne, 2005, for different findings). Lexical tones, another type of phonemic contrast, have also been found to show signs of hyperarticulation in IDS of various tone languages, including Mandarin Chinese, Cantonese, and Hakka, in comparison with ADS (Cheng & Chang, 2014; Han et al., 2018; Liu et al., 2007; Rattanasone et al., 2013).

More importantly, empirical studies have shown that children benefit from the use of IDS, especially in lexical processing (Floccia et al., 2016; Foursha-Stevenson et al., 2017; Graf Estes & Hurley, 2013; Ma et al., 2011; Singh et al., 2009; cf. Han, 2019; Mani & Pätzold, 2016). In particular, among the acoustic-phonetic characteristics of IDS, the prosodic exaggeration and vowel hyperarticulation have specifically been reported to facilitate word processing (e.g. prosody: Graf Estes & Hurley, 2013; Ma et al., 2011; vowels: Song et al., 2010). However, little research has been conducted so far with respect to lexical tones. This study aimed to investigate the effect of tone hyperarticulation on tone language learning toddlers' word recognition. Considering the uniqueness of lexical tones, this study is expected to enrich our understanding of the role of phonetic enhancements in IDS, in child language development.

Tone hyperarticulation in IDS

Lexical tones mark phonemic changes and alter lexical meanings in tone languages akin to vowels, whereas they are unique in terms of being primarily instantiated by changes of pitch, a prosodic cue (Yip, 2002). As a result, to acquire lexical tones, children may face unique challenges to differentiate the lexically relevant pitch variation in lexical tones from those lexically irrelevant pitch changes at the utterance level. Young language learners have been found to place different weights on tone variation versus vowel variation when processing words. In comparison with vowels, their processing of lexical-tone information

appears to develop in a distinct, prolonged, and more complex trajectory (Ma et al., 2017; Singh et al., 2015; Singh & Fu, 2016).

Rattanasone et al. (2013) provided an important piece of evidence for tone hyperarticulation in IDS of Cantonese. Analogous to the vowel space, the tonal space was calculated in this study as the area of the tone triangle; the tone triangle was formed when the onset and offset fundamental frequency values of the three peripheral lexical tones of Cantonese, that is, T1, T2, and T4, were plotted in two-dimensional space (the onset on the *x*-axis and the offset on the *y*-axis). An expansion of the tonal space was found in IDS compared with ADS, suggesting an acoustic exaggeration of the distances among tone categories.

Tone hyperarticulation has also been detected in other tone languages such as Mandarin Chinese (Han et al., 2018; Liu et al., 2007) and Hakka (Cheng & Chang, 2014). Liu et al. (2007), as the first empirical study on characteristics of lexical tones in IDS, reported an enhancement of pairwise acoustic contrasts between tone categories in IDS of Mandarin. Specifically, they found that the differences of pitch height in most of the tone pairs were greater in IDS than in ADS, the differences of pitch range between the level tone and the contour tones increased in IDS compared with ADS, and the differences of duration were larger in some tone pairs in IDS as well. The exaggeration of the pitch range between the level and contour tones in IDS of Mandarin Chinese has later been confirmed by Han et al. (2018). Cheng and Chang (2014), turning attention to another tone language, Hakka, measured the pitch differences at multiple time points along tone contours between every two tone categories. They also reported an enhancement of pairwise tone distances in IDS compared with ADS.

The effects of IDS on child word processing

The dual function of pitch in tone languages, in which it is used both at the utterance level and to differentiate word meanings, may pose particular challenges for young learners' use of tones in lexical processing. Tone language learning children must understand that certain changes of pitch patterns at the syllabic level (i.e. lexical tones), in contrast to general pitch movements at the utterance level, alter lexical meanings. Thus, an important question to address is whether these young learners benefit from tone hyperarticulation in IDS in word processing. Given the dual function of pitch and the challenges it may bring, tone hyperarticulation might be particularly beneficial to tone language learners. However, to our knowledge, research on this question is currently very limited. One relevant study (Xu, 2008) focused on the role of tone hyperarticulation in infants' lower level phonemic discrimination. It only provided some very preliminary and exploratory findings: among three mother–infant dyads, the infant whose mother most expanded the tonal space in IDS performed the best in discriminating consonants. Another study investigated the learning of Mandarin lexical tones by artificial neural networks instead of children (Gauthier & Shi, 2011). It reported no facilitatory effect of speech input with typical-IDS lexical tones. Neither of these studies paid attention to the role of tone hyperarticulation in children's word processing.

Despite the lack of work on lexical tones, a number of studies have been conducted to investigate the effects of other acoustic-phonetic enhancements in IDS on children's word processing. They have demonstrated that the exaggerated prosody and hyperarticulated

vowels in IDS can facilitate children's word learning (Foursha-Stevenson et al., 2017; Graf Estes & Hurley, 2013; Ma et al., 2011), word segmentation (Floccia et al., 2016), and word recognition (Singh et al., 2009; Song et al., 2010). For example, at a very early stage, 7- and 8-month-old English-learning infants succeeded in recognizing real words embedded in passages after a delay, only when the words were introduced in typical-IDS prosody with increased pitch height, pitch range, and duration (Singh et al., 2009). The exaggeration of these non-lexical prosodic cues in IDS has further been reported to support toddlers' learning of novel words in the second year of life (Graf Estes & Hurley, 2013; Ma et al., 2011). Toddlers in the second year of life were also found to benefit from an expansion of the vowel space in familiar word recognition. Specifically, 19-month-old toddlers performed significantly better in recognizing familiar words when the words were produced with expanded vowel space of typical IDS, as compared with when they were produced with ADS-like vowel space (Song et al., 2010). The facilitatory effects of vowel hyperarticulation on child language development have also been suggested by correlational studies: Significant positive associations were reported between the degree of vowel-space expansion in caregivers' IDS and the children's sound discrimination in the first year of life (Liu et al., 2003), and expressive and receptive language outcomes in the second year of life (Hartman et al., 2017; Kalashnikova & Burnham, 2018). Support has also come from brain studies: infants exhibited mature neural responses to /i/ extracted from IDS, but not ADS (Peter et al., 2016), and showed enhanced neural activities and synchronization when processing the formant-expanded vowel compared with the non-expanded one (Zhang et al., 2011).

The facilitatory effects of IDS seem to decrease as children get older. A few studies on children at around 2 years of age and older reported comparable performance in their word processing with IDS and ADS stimuli (Han, 2019; Ma et al., 2011; Robertson et al., 2017). Moreover, children's reliance on IDS input in word processing appears to be affected by their vocabulary size. Ma et al. (2011) reported that although 21-month-olds as a group failed to learn word–object pairs without IDS-like exaggerated prosody, a subgroup of them with vocabulary size larger than the median succeeded.

The findings of the facilitatory effects of IDS, especially with respect to vowel hyperarticulation, pose the question of whether such effects can be generalized to lexical tones. Although lexical tones are phonemes that distinguish one word from another like vowels, young language learners at certain ages have been found to treat tone variation in word processing with different sensitivity as compared with vowel variation.

Lexical tones in child word processing

Infants who are exposed to a tone language have been shown to attend to tone changes at a very early stage. In Singh and Foong (2012), 11-month-old English-Mandarin bilinguals recognized both pitch-matched and pitch-mismatched words in English, whereas they only recognized lexical-tone-matched words in Mandarin. Despite the early sensitivity to tone variation, tone representations undergo considerable subsequent refinement along an extended and complex trajectory as compared with vowels (Singh & Fu, 2016). While monolingual Mandarin-learning toddlers at 17 to 18 months were able to integrate some lexical tones into new words in a switch task (Singh et al., 2016), they failed in a

similar task when presented with a certain tone contrast (a rising-falling contrast; Burnham, Singh, Mattock, et al., 2018). Moreover, in familiar word recognition, 17-month-old toddlers learning Japanese with a pitch-accent language did not consistently perform better in the pitch-matched condition than the pitch-mismatched condition in a preferential looking paradigm, indicating unstable sensitivity to pitch variation at this age (Ota et al., 2018). Ma et al. (2017) compared the influence of tone and vowel variation on older monolingual Mandarin-learning children's recognition of both novel words (i.e. word learning) and familiar words using the preferential looking paradigm. Two-year-old children were tested with a novel word recognition task; their performance was found to be hindered by mispronunciations of both lexical tones and vowels. By contrast, 3-year-old children were tested with not only the novel word recognition task but also a familiar word recognition task; their performance was hindered by mispronunciations of vowels but not lexical tones in both the two tasks. The results suggested that, while the children at 2 years of age were sensitive to vowel and lexical-tone variation alike, those at 3 years of age seemed to be only sensitive to vowel variation but not lexical-tone variation.

There are also studies examining English-Mandarin bilingual children's sensitivity to lexical-tone and vowel variation in familiar word recognition. Wewalaarachchi et al. (2017) reported that 2-year-old bilinguals demonstrated more sensitivity to vowel (and consonant) variation than to lexical-tone variation. Singh et al. (2015) found that 2.5- to 3.5-year-olds instead had greater sensitivity to lexical-tone variation than to vowel (and consonant) variation, while 4- to 5-years-olds on the contrary were more sensitive to vowel (and consonant) variation than to lexical-tone variation. From 2 to 5 years of age, there seems to be a bell-shaped developmental trajectory of bilingual tone language learners' sensitivity to lexical-tone variation in familiar word recognition in comparison with vowel variation. A more recent study (Singh & Wewalaarachchi, 2020) turned attention to tone language learning children's sensitivity to native versus nonnative lexical tones and found greater sensitivity to native tone variation in both toddlerhood and preschool age.

Non-tone language learners have been found to remain open to the use of tones in lexical processing until a quite late developmental stage. Fourteen-month-old English-learning toddlers were able to learn novel word pairs that contained the same segmental components but carried different Mandarin lexical tones (Hay et al., 2015), as long as one of the novel words carried a tone with a rising pitch contour (Hay et al., 2019). Such flexibility was found to be lost for the monolingual English-learning toddlers at 17 months of age and older (Hay et al., 2015). Bilingual children learning two non-tone languages demonstrated an even longer period of flexibility. They did not fail to learn novel word pairs, only contrasting in lexical tones until 22 months of age (Graf Estes & Hay, 2015). When comparing non-tone language learners' sensitivity with tone and vowel variation, Singh et al. (2014) reported that both monolinguals and bilinguals at 18 months of age still treated tone substitutions as mispronunciations in novel word learning as they did for vowel substitutions. They lost the sensitivity to tone variation late at 24 months of age.

All these studies on young learners' processing of lexical tones have revealed a complex developmental trajectory. On one hand, infants display early sensitivity to lexical tones in word processing. On the other hand, compared with vowels, it takes a longer time for tone representations to be refined and stabilized. Moreover, there seems to be a

late, unexpected, decrease in children's sensitivity to tone variation as compared with vowel variation (Ma et al., 2017; Singh et al., 2015). It appears that, at some ages, tone language learning children do not treat lexical tones and vowels with the same sensitivity in word processing. The more complex and extended developmental trajectory of tone acquisition may result from the use of the prosodic cue pitch in lexical tones, which is likely to pose particular challenges to tone language learners as aforementioned. The uniqueness of lexical tones has prompted us to pay attention to the role of tone hyperarticulation in IDS in early language development.

The current study

This study investigated whether tone hyperarticulation in IDS would facilitate Cantonese-learning toddlers' familiar word recognition in the second year of life. Previously, Song et al. (2010) reported a facilitation of 19-month-old toddlers' familiar word recognition by vowel hyperarticulation in IDS. Accordingly, this study started by targeting toddlers of the same age in a similar word recognition task to determine whether a facilitatory effect of lexical-tone hyperarticulation would be found (Experiment 1). Toddlers' word recognition was compared between a hyperarticulated-tone (HT) condition (in which the tonal distances in word pairs were exaggerated as in IDS) and a non-hyperarticulated-tone (NT) condition (with smaller tonal distances that resembled those in ADS). The segmental components (i.e. vowels and consonants) in each word pair were specifically matched as much as possible in the present study to examine whether tone exaggeration would facilitate toddlers' recognition between similar sounding words.

To preview the results, a facilitatory effect of tone hyperarticulation was found for toddlers at 19 months of age. Therefore, to explore whether such an effect would diminish as children's language development proceeds, Experiment 2 was conducted to examine the same question in toddlers at around 2 years of age. This age was targeted because previous studies had demonstrated a decrease of IDS effects at around this time point (e.g. Han, 2019; Robertson et al., 2017) as aforementioned.

Moreover, as mentioned above, vocabulary size has been reported to affect children's reliance on the phonetic enhancements in IDS (Ma et al., 2011). Therefore, the vocabulary size of the participants was measured in this study through parental reports to examine whether this factor also modulated the effect of tone hyperarticulation on toddlers' word recognition.

Experiment 1

Method

Participants. Data were collected from 21 Cantonese-learning toddlers aged 19 months (M = 563.71 days; SD = 14.09 days; 10 males and 11 females). The language background of the toddlers was investigated through an interview with the parents: their parents were all Cantonese native speakers and spoke Cantonese to them, and there was little exposure to other languages in their home environment. Two additional toddlers were tested but excluded due to fussiness during the experiment. As an effect of vowel

hyperarticulation on toddlers' familiar word recognition had been detected from 16 participants (Song et al., 2010), we pursued a sample size of no less than 16 participants. The participants were all Chinese, recruited in Hong Kong by advertising in WhatsApp groups. The parents reported no mental, sensory, or language deficits of the participants. Written informed consent approved by The Joint Chinese University of Hong Kong - New Territories East Cluster Clinical Research Ethics Committee was obtained from the parents. The study was conducted in accordance with the principles of the Helsinki Declaration.

Stimuli. The speech stimuli were Cantonese utterances that contained target words familiar to the toddlers of this age. Table 1 lists the pairs of target words. The visual stimuli were pictures corresponding to each of the target words. The two pictures corresponding to the words in each pair would appear simultaneously on the screen, and the toddlers would be instructed by the speech stimuli to look at one of the pictures. Table 2 represents an example of the speech and visual stimuli used in one trial.

Pairs of target words carrying either high-level (T1) versus mid-level (T3) tones or high-level (T1) versus low-level (T6) tones were used. Cantonese as a tone language is unique in having three level tones that are differentiated from each other only by the pitch height. A close examination of pairwise tone contrasts in IDS of Cantonese found that the tone hyperarticulation appeared to be mostly driven by larger pitch differentiation between the high-level tone and the lower level tones (Xu, 2008).

The degree of familiarity of each target word was determined using CCDI (Tardif et al., 2008); it was calculated as the percentage of children (aged 14–16 months) who were reported by the parents to understand the word. CCDI is the Cantonese version of the MacArthur-Bates Communicative Development Inventories (Fenson et al., 1993). All the target words achieved no less than 60% degree of familiarity, that is, 60% and more children at 14 to 16 months of age were reported to understand the words. The target words in each pair were matched as far as possible in their degree of familiarity.

Efforts were also made to match the segmental components of the target words in each pair. Ideally, it would be best to use minimal pairs of words that only contrast with each other in terms of the lexical tone. However, in reality, it is impractical to find enough

Table 1. The list of the pairs of target words used in the experiment.

Tone pairs	Word pairs	Gloss	IPA	Familiarity	Condition
T1–T6 (VM)	粥(zuk)-肉(juk)	Porridge-meat	[tsʊk̚]-[jʊk̚]	67%–60%	HT
	筆(bat)-襪(mat)	Pen-sock	[pɐt̚]-[mɐt̚]	75%–86%	NT
T1–T3 (CM)	窗(coeng)-菜(coi)	Window-vegetable	[tsʰœːŋ]-[tsʰɔːy]	60%–76%	HT
	羹(gang)-鏡(geng)	Spoon-mirror	[kɐŋ]-[kɛːŋ]	74%–67%	NT
T1–T3 (MP)	燈(dang1)-凳(dang3)	Lamp-stool	[tɐŋ]-[tɐŋ]	71%–79%	Both

Note. The last column represents one possibility of the condition that each word pair was tested for an individual participant. CM = initial consonant matched; MP = the minimal pair; VM = vowel matched; HT = hyperarticulated-tone condition; NT = non-hyperarticulated-tone condition; IPA: International Phonetic Alphabet.

Table 2. An illustration of the procedure of the experiment with the visual and speech stimuli demonstrated (taking one trial as an example).

	Duration (s)	Left side	Center	Right side	Speech stimuli
Attention getter			Baby face		Music
Familiarization phase	3				呢个系乜来噶？睇下呢个啊! (What is this? Look at this!)
	3				呢个系乜来噶？睇下呢个啊! (What is this? Look at this!)
Salience phase	3				呢的系乜来噶？ (What are these?)
Test phase – the first part	2				燈系边度啊？燈！ (Where is the lamp? The lamp!)
Test phase – the second part	2.5				

minimally contrastive pairs of this kind that are all familiar to Cantonese-learning toddlers at this age. We only managed to find one such minimal pair (MP), that is, 燈 ([tɐŋ˥])-凳([tɐŋ˧]). The other word pairs were matched as much as possible either with respect to the vowel (vowel-matched [VM]) or the initial consonant (consonant-matched [CM]), whichever was available.

The speech stimuli were originally recorded by a female Cantonese native speaker who was required to speak as if interacting with a young child. The recording was made in a soundproofed booth, using a condenser head-mounted microphone (Audio-Technica BP894), an audio interface (Roland Quad-Capture), and a laptop (MacBook Air) with 44100 Hz sampling rate and 16 bits sampling precision. The recordings were selected and acoustically manipulated in Praat (Boersma & Weenink, 2017) into two different conditions, that is, the HT condition and the NT condition. Specifically, the pitch of each target word was manipulated by altering the height of the overall pitch contour while maintaining the original shape of the contour.

To decide how much the T1 to T3 and T1 to T6 pairs should contrast in the HT and NT conditions, we analyzed pitch of words carrying T1, T3, and T6 in IDS and ADS collected from 10 Cantonese-speaking mothers when speaking to their children (of a similar age to the participants in this experiment) and an adult. The tone contrasts were calculated as the difference in the averaged pitch height between T1 and T3/T6. On average, the pitch contrast between T1 and T3 was 4.16 semitones ($SD = 0.84$) in their IDS and 2.26 semitones ($SD = 1.42$) in their ADS, and the pitch contrast between T1 and T6 was 5.51 semitones ($SD = 0.82$) in IDS and 3.2 semitones ($SD = 2.27$) in ADS. Based on these results, we decided to round the averaged values of the tone contrasts down (except for the T1–T6 contrast in IDS, which was far from both 5 and 6 semitones) and manipulate the T1 to T3 and T1 to T6 contrasts to differ by 4 semitones and 5.5 semitones, respectively, in the HT condition, and by 2 semitones and 3 semitones, respectively, in the NT condition. With the original pitch height of the recorded target words (see Table 3)

Table 3. The acoustic details of the speech stimuli in the hyperarticulated-tone (HT) condition and the non-hyperarticulated-tone (NT) condition.

Tone contrasts	Target words	Before manipulation	After manipulation		Degree of pitch manipulation (semitones)	Duration of the target words (milliseconds)
		Pitch of the target words (Hz)	Pitch of the target words (Hz)	Pitch contrasts between the target words (semitones)		
HT condition						
T1–T6	zuk	347.2	340	5.5	0.36	123
	juk	256.4	247		0.65	110
	bat	339.1	340	5.5	0.05	100
	mat	258.5	247		0.79	108
T1–T3	coeng	358.9	340	4	0.94	300
	coi	294.3	270		1.49	280
	gang	336.4	340	4	0.18	223
	geng	270.3	270		0.02	250
	dang1	322.1	340	4	0.94	201
	dang3	277.7	270		0.49	206
NT condition						
T1–T6	zuk	347.2	303	3	2.36	123
	juk	256.4	255		0.09	110
	bat	339.1	303	3	1.95	100
	mat	258.5	255		0.24	108
T1–T3	coeng	358.9	303	2	2.93	300
	coi	294.3	270		1.49	280
	gang	336.4	303	2	1.81	223
	geng	270.3	270		0.02	250
	dang1	322.1	303	2	1.06	201
	dang3	277.7	270		0.49	206

Note. The pitch values before and after manipulation were both reported. The degree of pitch manipulation was calculated as the difference in the pitch of the target words before and after manipulation in semitones. HT = hyperarticulated-tone condition; NT = non-hyperarticulated-tone condition.

taken into consideration, all the tokens of T1, T3, and T6 were modified to be 340 Hz, 270 Hz, and 247 Hz, respectively, in the HT condition. In contrast, for the NT condition, the pitch height of all the T1, T3, and T6 tokens was modified to be 303 Hz, 270 Hz, and 255 Hz, respectively. The degree of pitch manipulation was calculated as the difference in the pitch of a target word before and after manipulation (see Table 3). On average, the degree of pitch manipulation for the HT condition was 0.59 semitones, while it was 1.24 semitones for the NT condition. The smaller degree of pitch manipulation in the HT condition was expected as the stimuli were originally recorded in an IDS style. To ensure that the different degrees of pitch manipulation did not lead to a difference in the naturalness of the HT versus NT stimuli, a small experiment was conducted with 10 adult

Cantonese native speakers. They were required to rate the degree to which each stimulus sounded like natural speech on a 10-point scale ('1' for *very unnatural speech*; '10' for *very natural speech*). A paired *t* test showed no significant difference in the naturalness of the stimuli between the two conditions ($t(9) = .87$, $p = .4$, $d = .051$; HT condition: $M = 8.52$, $SD = 1.13$; NT condition: $M = 8.46$, $SD = 1.18$). The duration of the target words was not normalized, but was controlled for each word pair as much as possible by selecting the tokens with similar duration from the original recordings. The intensity of all the speech stimuli was normalized at 65 dB. The acoustic details of the stimuli are represented in Table 3.

Apparatus and procedure. The experiment was conducted in a sound-attenuated booth. A Tobii TX300 eye-tracking system was employed for both stimuli presentation and eye-movement data collection. The toddlers sat on their caregiver's lap, facing the Tobii monitor at a distance of approximately 65 centimeters. The visual stimuli were displayed on the screen of the monitor, while the speech stimuli were presented through two loudspeakers behind and to the left and right of the monitor, respectively. A projector and a high-resolution camera were built into the Tobii monitor. The projector emitted invisible near-infrared light, and the camera recorded the participants' visual fixation on the screen according to the light reflected by their cornea relative to their pupils using certain mathematical algorithms. The experimenter monitored the presentation of the stimuli and the toddlers' real-time visual fixation through another computer out of the toddlers' sight, and controlled the experiment. At the start of each experiment, a 5-point infant calibration process was conducted, and repeated if necessary. Before the experiment, the caregivers were asked to wear a pair of sunglasses with the lenses covered by black tape or to close their eyes throughout the experiment to avoid any possible influence from them on the toddlers' performance.

We employed the intermodal preferential looking paradigm (IPLP) and, in particular, followed the experimental design in Song et al. (2010). Each trial began with the face of a smiling baby at the center of the screen along with music to attract the toddlers' attention. The trial only began after the toddlers turned their attention to the image on the screen, controlled by the experimenter. A trial consisted of three phases (see Table 2). First, in the familiarization phase, toddlers were sequentially shown two pictures of objects (one located at the left edge of the screen and the other at the right edge) corresponding to the two words in a word pair. Each picture was presented for 3 seconds along with the speech stimuli 'What is this? Look at this!'. The familiarization phase was followed by a salience phase lasting for 3 seconds in which both of the two objects were presented on the screen simultaneously, with one at the left edge, the other at the right edge, and approximately 27 centimeters between them; the speech stimuli 'What are these? was played along with the presentation of the visual stimuli. The salience phase would reveal any object and position preferences that the participants had, and thus served as the baseline for their test performance. Finally, the test phase followed, which was separated into two parts. In the first 2-second part, a picture of a child holding a question marker appeared to ensure that the toddlers fixated at the center of the screen. The speech stimuli 'Where is ___? The ___!' were presented in which the blanks were filled with one of the two words to instruct the toddlers to look at its corresponding object

(this word and the corresponding object would be the target in this trial). The second part of the test phase started as soon as the speech stimuli stopped. In the second part, the two pictures of the objects appeared on the screen simultaneously in the same positions as in the salience phase, and remained visible for 2.5 seconds. The side of the screen that each object was presented on was not counterbalanced across trials. However, this should not influence the pattern of results due to the use of the salience phase as the baseline. The two-part design of the test phase (i.e. presenting the speech stimuli first with an attention grabber and subsequently presenting the objects) was done to ensure that participants were fixating a neutral position (rather than either of the objects) during the playing of the speech stimuli. In total, each trial lasted for 13.5 seconds.

The toddlers completed both the HT and NT conditions. For an individual participant, one of the two CM and one of the two VM word pairs were tested in the HT condition; the other CM and VM pairs were tested in the NT condition. The condition in which a CM/VM word pair was tested was counterbalanced across participants (see the last column in Table 1 for an example). The one minimally contrastive word pair (MP) was instead tested in both conditions, considering that it is an ideal test material for our purpose. Within a word pair, each word served as the target once; thus, every CM/VM word pair was tested twice, while the one minimal pair word was tested 4 times, twice for each condition. In total, a toddler completed 12 trials throughout the experiment, with six trials in each condition in which there were three word pairs (one CM word pair, one VM word pair, and plus the one MP), each tested twice. The order of presentation of the two conditions was counterbalanced across participants, and the order of presentation of the trials within each condition was semi-randomized with no word pair tested twice in a row.

Vocabulary assessment. To measure the toddlers' vocabulary size, their caregivers were asked after the behavioral experiment to fill out the 'Words and Sentences' checklist from CCDI (Tardif et al., 2008) to indicate whether their child could produce each word in the list. The toddlers on average obtained a score of 125.81 ($SD = 135.26$) for vocabulary production. To prepare for later analyses, following Ma et al. (2011), we divided the toddlers into two groups, that is, the high vocabulary (HV) group and the low vocabulary (LV) group, based on a median split of their CCDI scores ($Mdn = 93$).

Data analysis. The data were originally exported from the Tobii studio, which recorded whether the toddlers fixated on the left, right, or neither side of the screen at 3-millisecond intervals. They were then processed by a self-written MATLAB script. The data in the salience phase and the test phase (the second part) were extracted for analysis. There was, on average, a 16% ($SD = 13\%$) gaze loss per trial across participants. We excluded the trials with gaze loss more than 30% (Singh et al., 2014); in total, 91% of the trials were included in the analyses. For each trial, we first calculated the percentage of the toddlers' fixation time to the target in the salience phase and the test phase separately. An increase of fixation time to the target in the test phase, compared with the salience phase, can indicate successful recognition of the word. Therefore, the difference in the percentage of fixation time to the target between the two phases was computed, to index the toddlers' performance in this trial. Specifically, we subtracted the percentage of fixation on the target in the salience phase from the percentage of fixation on the target in the test

phase. The fixation time in the salience phase was used as the baseline to control for any possible picture or position biases. This measurement has also been employed in Song et al. (2010). The difference scores of all the trials in each condition were then averaged for individual participants to represent their performance in this condition.

Results

To test (a) whether the exaggeration of tone contrasts would lead to a significant difference in the toddlers' performance, and (b) whether the effect of tone hyperarticulation would be modulated by the toddlers' vocabulary size, a two-way analysis of variance (ANOVA) was conducted with condition (HT vs. NT) as the repeated-measures variable, vocabulary size (HV vs. LV) as the between-subjects variable, and the toddlers' difference scores as the dependent variable. The results showed significant main effects of both condition ($F(1, 19) = 5.31$, $p = .033$, $\eta^2 = .12$) and vocabulary size ($F(1, 19) = 4.49$, $p = .048$, $\eta^2 = .08$), but no significant interaction ($F(1, 19) = .94$, $p = .34$, $\eta^2 = .022$). The difference scores for the HT and NT conditions are represented in Figure 1 separately for the HV and LV groups. The significant effect of condition suggested that the 19-month-old toddlers performed significantly better in the HT condition containing enhanced tone contrasts than in the NT condition with smaller tone contrasts.

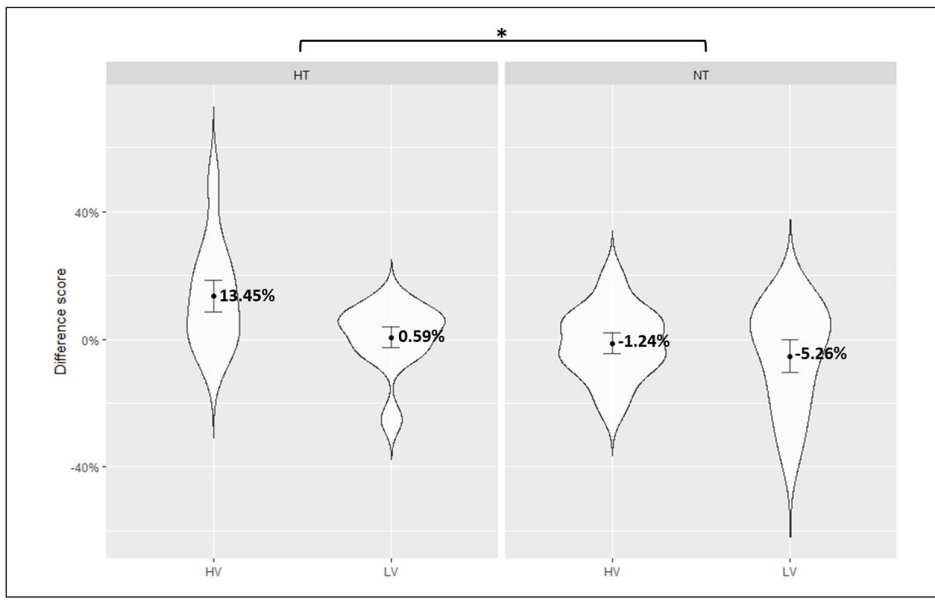

Figure 1. Violin-plots of the difference scores (i.e. the difference in the percentage of fixation time to the target between the test and salience phases) of the 19-month-old toddlers separated into the high vocabulary (HV) group and the low vocabulary (LV) group in the hyperarticulated-tone (HT) condition versus non-hyperarticulated-tone (NT) condition.
The dot and error bar represent the mean (labeled with the mean value) and standard error of the data.
*$p < .05$.

Table 4. The percentage of the 19-month-old toddlers' fixation time to the target in the salience phase and the test phase (the second part) in the hyperarticulated-tone (HT) condition and the non-hyperarticulated-tone (NT) condition, averaged across participants.

	Percentage of fixation time to the target		Paired *t* test
	The salience phase	The test phase	
HT condition	$M = 47.54\%$; $SD = 9.89\%$	$M = 54.86\%$; $SD = 13.49\%$	$t(20) = 2.2$, $p = .04$, $d = .62$
NT condition	$M = 53.42\%$; $SD = 8.39\%$	$M = 50.26\%$; $SD = 10.59\%$	$t(20) = -1.08$, $p = .29$, $d = -.33$

Note. The mean and standard deviation were reported. Comparisons were made between the salience and test phases using paired *t* tests. The results were reported along with the effect size (Cohen's *d*). HT = hyperarticulated-tone condition; NT = non-hyperarticulated-tone condition.

The HV group in general outperformed the LV group, as indicated by the significant effect of vocabulary size. In addition, it seemed that the HV group took more advantage of the enhanced tone contrasts in the HT condition than the LV group, according to direct observation of the difference-score plots in Figure 1. The effect of tone enhancement on the LV group's performance was limited, shifting it from a slight preference away from the target in the NT condition to a chance level in the HT condition. However, we failed to detect a significant interaction effect; statistically, no evidence could be provided that the effect of tone enhancement on the 19-month-olds' word recognition was modulated by their vocabulary size.

As the ANOVA had exhibited a significant difference in the toddlers' performance between the HT and NT conditions, we further checked within each condition whether the toddlers successfully recognized the words; we compared their percentage of fixation time with the target between the salience phase and the test phase with paired *t* tests. The results are reported in Table 4 in detail. In the HT condition, the toddlers significantly increased the percentage of their fixation on the target in the test phase ($p = .04$), indicating successful recognition of the words. By contrast, in the NT condition, the percentage of fixation on the target did not increase and even slightly decreased in the test phase compared with the salience phase; the result of the *t* test showed that the difference was not significant ($p = .29$). There was no sign of successful word recognition in the NT condition.

Experiment 2

As the performance of the 19-month-old toddlers was found to be facilitated by the enhanced tone contrasts, we further conducted Experiment 2 with an older group of toddlers, at around the end of the second year of life, to investigate whether they reduced their reliance on tone hyperarticulation in recognizing the similar sounding words.

Method

In this experiment, data were collected from 33 Cantonese-learning toddlers aged 23 months ($M = 690.33$ days; $SD = 22.21$ days; 17 males and 16 females). As the

facilitatory effect of tone hyperarticulation was predicted to decrease for the older group, more subjects were recruited compared with Experiment 1 to reduce the risk of a type II error. According to the parental interview, the parents of the participants were all Cantonese native speakers and spoke Cantonese to their children, and there was little exposure to other languages in the home environment. Six additional toddlers were tested but excluded due to fussiness ($n = 4$) or inattentiveness ($n = 2$) during the experiment. The participants were all Chinese recruited in Hong Kong by advertising in WhatsApp groups. The parents reported no mental, sensory, or language deficits of the toddlers. Written informed consent approved by The Joint Chinese University of Hong Kong - New Territories East Cluster Clinical Research Ethics Committee was obtained from the parents.

The stimuli, apparatus, procedure, and data analysis of this experiment were all the same as Experiment 1. For this group of participants, there was, on average, a 22% ($SD = 16\%$) gaze loss per trial across participants; 85% of the trials were included in the analyses after excluding the trials with gaze loss more than 30% (Singh et al., 2014). The 23-month-old toddlers on average obtained a score of 318.27 ($SD = 198.04$) for vocabulary production. Their vocabulary scores were significantly larger than those of the 19-month-old toddlers ($t(51.68) = 4.24$, $p < .001$, $d = 1.09$). We also divided these toddlers into the HV and LV groups based on a median split of their CCDI scores for vocabulary production ($Mdn = 312$).

Results

With this group of older toddlers, the two-way ANOVA (condition * vocabulary size) only found a main effect of vocabulary size ($F(1, 31) = 7.75$, $p = .0091$, $\eta^2 = .1$) on the toddlers' difference scores; there was neither a main effect of condition ($F(1, 31) = .11$, $p = .75$, $\eta^2 = .0017$) nor an interaction between the condition and vocabulary size ($F(1, 31) = .48$, $p = .49$, $\eta^2 = .0076$). The difference scores for the HT and NT conditions are represented in Figure 2 separately for the HV and LV groups of older toddlers. The 23-month-old toddlers performed comparably in the two conditions. The tone hyperarticulation did not lead to a significant difference in their recognition of the familiar words.

Similar to the younger group, the older toddlers with higher vocabulary size also outperformed those with lower vocabulary size. In light of the insignificant interaction, the effect of tone hyperarticulation on the toddlers' performance was not found to be modulated by vocabulary size.

We also compared the toddlers' percentage of fixation time with the target between the salience and test phases within each condition with paired t tests. The results are reported in Table 5 in detail. In the HT condition, similar to what was found for 19-month-olds, we observed a significant increase of the percentage of fixation time to the target in the test phase as compared with the salience phase ($p = .006$), indicating successful recognition of the words. As for the NT condition, different from the performance of the 19-month-olds, the percentage of fixation time to the target also increased in the test phase compared with the salience phase; the result of the t test reported a p value at the margin of statistical significance ($p = .062$).

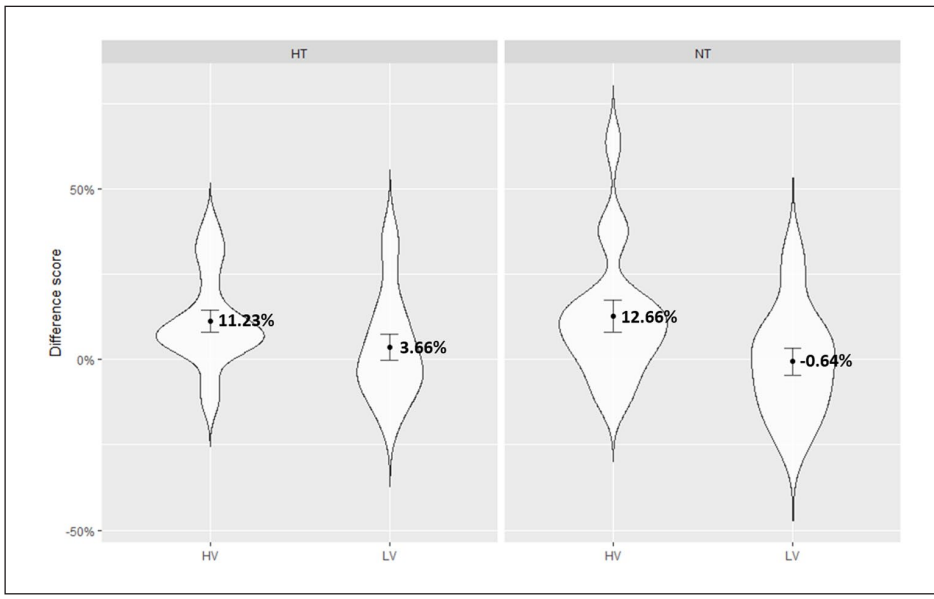

Figure 2. Violin-plots of the difference scores (i.e. the difference in the percentage of fixation time to the target between the test and salience phases) of the 23-month-old toddlers separated into the high vocabulary (HV) group and the low vocabulary (LV) group in the hyperarticulated-tone (HT) condition versus non-hyperarticulated-tone (NT) condition.
The dot and error bar represent the mean (labeled with the mean value) and standard error of the data.

Table 5. The percentage of the 23-month-old toddlers' fixation time to the target in the salience phase and the test phase (the second part) in the hyperarticulated-tone (HT) condition and the non-hyperarticulated-tone (NT) condition, averaged across participants.

	Percentage of fixation time to the target		Paired *t* test
	The salience phase	The test phase	
HT condition	$M = 48.79\%$; $SD = 10.92\%$	$M = 56.34\%$; $SD = 10.19\%$	$t(32) = 2.95$, $p = .006$, $d = .72$
NT condition	$M = 49.37\%$; $SD = 11.7\%$	$M = 55.58\%$; $SD = 11.58\%$	$t(32) = 1.93$, $p = .062$, $d = .53$

Note. The mean and standard deviation were reported. Comparisons were made between the salience and test phases using paired *t* tests. The results were reported along with the effect size (Cohen's *d*). HT = hyperarticulated-tone condition; NT = non-hyperarticulated-tone condition.

General discussion

IDS as a special speech register is widely used when talking to young children and has been found to play an important role in early language acquisition (Cristia, 2013). In particular, a number of studies have been conducted on the prosodic exaggeration and vowel hyperarticulation in IDS and reported some facilitatory effects on young learners'

word processing (e.g. Ma et al., 2011; Song et al., 2010). However, previous work has not considered whether children's lexical processing is similarly facilitated by lexical-tone hyperarticulation of IDS.

This study demonstrated that the 19-month-old Cantonese-learning toddlers performed differently in recognizing similar sounding familiar words with and without tone enhancement. The toddlers as a group succeeded when the acoustic contrasts of lexical tones in word pairs were enlarged as in IDS, but failed when tone contrasts were smaller, resembling those in ADS. In contrast, the 23-month-old toddlers performed comparably well in the same word recognition task with and without the tone-contrast enhancement.

The findings from the 19-month-olds suggest that, in the middle of the second year of life, tone language learning toddlers' ability to recognize familiar words can be facilitated by tone hyperarticulation in IDS. This is consistent with the previous findings reported by Song et al. (2010) with regard to vowel hyperarticulation in IDS. Although children appear to treat vowel and tone variation with different sensitivity at certain ages and follow dissimilar developmental trajectories for tone and vowel processing (Ma et al., 2017; Singh et al., 2015; Singh & Fu, 2016), their word recognition in the middle of the second year of life can benefit from an exaggeration of both vowel and tone contrasts. These findings on lexical tones from tone language learners provide cross-language support for the facilitatory effects of the phonetic enhancements in IDS. In addition, our findings may have some practical implications for intervention studies on language disorders. There are disorders of language that are closely related to deficits in processing tones (see also Wong et al., 2012 and Wong et al., 2020, for the potential neurogenetic basis of tones). Mandarin speakers with congenital amusia, for example, displayed reduced categorical perception of lexical tones and had difficulties in extracting context-dependent tonal categories (Liu et al., 2021). The amusics' abilities in tone processing could be improved through auditory training (Liu et al., 2017). In light of the facilitatory effect of tone exaggeration detected in this study, such a feature could be incorporated into future intervention for amusia and other language disorders in connection with tone processing.

The findings from the 23-month-olds further demonstrate a developmental decrease of the facilitatory effect of tone hyperarticulation between 19 and 23 months of age. At around the end of the second year of life, the enhancement of the tone contrasts may have little influence on the toddlers' performance in familiar word recognition. Previous studies on IDS effects in word processing have reported that children aged 2 years and older stopped relying on IDS, especially with regard to the prosodic exaggeration (Han, 2019; Ma et al., 2011; Robertson et al., 2017). However, there is a lack of such research on vowel hyperarticulation with children older than 19 months of age. Little is known about the later developmental changes in the effects of enhancements of phonemic contrasts in IDS. Our findings demonstrate that the facilitatory effects of phonemic tone enhancements on child word processing may decline in a manner similar to the effects of prosodic exaggeration.

The results from the analyses within each experimental condition suggest that it might be the toddlers' increased proficiency in doing the task that led to the developmental decrease of the facilitatory effect of tone hyperarticulation. The 19-month-olds were able to recognize the words in the HT condition but not the NT condition, indicating that the

toddlers at this age may not be very proficient in recognizing similar sounding words. Therefore, they may require more exaggerated tone contrasts to help them perform the task. By contrast, the 23-month-olds appeared to be able to recognize the similar sounding words in both conditions (although to a marginally significant degree in the NT condition). The older toddlers may become more proficient in performing this challenging task and thus can overcome the difficulty that 19-month-olds may experience when presented with non-exaggerated tone contrasts. It seems that the enhancement of tone contrasts facilitates toddlers' word recognition mainly when they are not proficient enough in doing the task.

The toddlers' improvement in performing the task may be attributed to increases in their processing efficiency during word recognition in the second year of life. Fernald et al. (1998) reported that in a familiar word recognition task, 24-month-old children could make rapid decisions based on incomplete phonetic information, that is, they showed a sign of successful recognition even before the whole word was spoken; in contrast, 15-month-olds could only do so after the end of the spoken word. According to Fernald et al.'s findings, it is reasonable to suppose that toddlers' processing efficiency of phonetic information got developed from the age of 19 to 23 months, which supported more proficient recognition of the similar sounding words; as a result, the toddlers reduced their dependence on the exaggerated tone contrasts. It is hard to confirm this supposition with our data due to the limitation of the experimental design. To accurately measure the efficiency of online phonetic processing in word recognition in a preferential looking paradigm, it is important to present the spoken words and pictures simultaneously in the test period, as Fernald et al. (1998) did. However, in this experiment, following Song et al. (2010), the pictures appeared after the speech stimuli ended in the test phase; moreover, because of the syntactic structure of Cantonese, the target word was even placed at the beginning of the sentence naming it and thus appeared long before the presentation of the pictures.

Another factor that might account for the increase of the toddlers' proficiency in performing this task is with respect to their sensitivity to tone variation in word recognition. In this experiment, the target word pairs contained segmental components that were either similar (the CM/VM pairs) or even identical (the minimal pair). Therefore, the participants' performance might be particularly influenced by their sensitivity to the suprasegmental component, that is, lexical tones. Neural research has shown that Cantonese-learning children's tone encoding without lexical context keeps improving during the ages of our interest (Novitskiy et al., 2022); unfortunately, it cannot be determined whether their sensitivity in encoding tone variation in lexical processing develops between the two ages that we tested. For the older age group, as elaborated above, Ma et al. (2017) reported that 2-year-old Mandarin-learning monolinguals showed good sensitivity to tone variation in novel word recognition as they did to vowel variation. However, such study is lacking for the younger toddlers. Although Singh and Foong (2012) found early sensitivity to tone variation in word recognition from 11-month-old English-Mandarin bilinguals, the task that they used did not require these prelexical infants to associate the familiar words to their corresponding objects. Ota et al. (2018) employed a word recognition task similar to Ma et al. (2017) and focused on toddlers in the middle of the second year of life, but it investigated learners of Japanese, a

pitch-accent language rather than a tone language. In pitch-accent languages, the word accent indicated by pitch can change lexical meanings, but it is usually assigned to one syllable in a word; this is different from the lexical tone that can be carried by each syllable in tone languages. Ota and colleagues found that 17-month-old Japanese toddlers' familiar word recognition was not consistently better in the pitch-matched condition than in the pitch-mismatched condition. They thus argued that Japanese toddlers' ability to encode lexically relevant pitch variation might not be stabilized at this age. Considering the differences between the pitch accent and tone language systems, the results from Ota et al. (2018) cannot be directly generalized to tone language learners. More research is needed to draw clear developmental trajectories of children's sensitivity to tone variation in word recognition, so as to test whether this is a factor underlying the developmental decrease of the effect of tone hyperarticulation. Moreover, according to Ma et al. (2017), monolingual Mandarin learners' sensitivity to tone variation in word recognition appeared to decrease between 2 and 3 years of age; 3-year-olds were only sensitive to vowel variation but not tone variation in both novel and familiar word recognition. In light of their findings, a question that is worth further investigation arises: Will the facilitatory effect of tone hyperarticulation rebound for children aged 3 years as a consequence of the decline of their sensitivity to tone variation?

This study also investigated whether the effect of tone hyperarticulation on toddlers' word recognition would be mediated by the toddlers' vocabulary size. Despite the main effect of vocabulary size, we did not find statistically significant interaction between the experimental condition and the toddlers' vocabulary within each age group. Although the toddlers with larger vocabulary outperformed those with smaller vocabulary, little statistical evidence could be provided that the vocabulary size modulated their dependence on enhanced tone features in word recognition. This is different from the abovementioned Ma et al.'s (2011) findings that 21-month-olds with larger vocabulary could learn new words without exaggerated prosody, while those with smaller vocabulary could not. As the two studies also differ in the word processing task employed (familiar word recognition in this study vs. novel word learning in Ma et al., 2011), it cannot be determined whether the dissimilar findings reflect a difference between the facilitatory effects of tone hyperarticulation and prosodic exaggeration in IDS.

More comprehensive research is needed to better understand the role that tone hyperarticulation plays in child language development. As aforementioned, an adjustment to the experimental design (i.e. to present the speech and visual stimuli simultaneously) will enable an accurate measurement of participants' efficiency in online phonetic processing; then, it can be investigated whether this factor affects how much children rely on enhanced tone contrasts. In addition, it remains to be examined whether the facilitatory effect of tone hyperarticulation reappears for older children at 3 years of age when their sensitivity to tone variation declines as reported by Ma et al. (2017). Studies testing children of other age groups are also necessary to draw a fuller picture of the developmental trajectories of the tone hyperarticulation effect. To test how generalizable our findings are, future research can be conducted with learners of other tone languages with different tonal systems. A number of tone languages such as Mandarin Chinese and Thai have relatively more discriminable tone contrasts than Cantonese (Burnham, Singh, Kasisopa, et al., 2018). Young learners of these languages are supposed to face less

challenges in tone perception and acquisition. Therefore, they might rely less on tone exaggeration, or their reliance might diminish earlier. Last but not least, this study focused on toddlers' recognition of familiar words; how tone hyperarticulation may influence children's performance in other language tasks such as phonetic discrimination and novel word learning remains to be explored.

In conclusion, this study found a facilitation of Cantonese-learning toddlers' recognition of similar sounding familiar words by tone hyperarticulation in IDS in the middle of the second year of life, providing cross-language evidence for the facilitatory effects of phonetic enhancements in IDS. Meanwhile, a developmental decrease of this facilitatory effect from 19 to 23 months of age was observed. By the end of the second year of life, toddlers seemed to largely reduce reliance on the enhanced tone contrasts to recognize words, performing comparably well when the tone contrasts were not exaggerated.

Acknowledgements

We thank Dr. Marina Kalashnikova for her helpful advice on the study. The current research data were collected during the first author's Ph.D. study at The Chinese University of Hong Kong.

Author contribution(s)

Luchang Wang: Conceptualization; Data curation; Formal analysis; Investigation; Methodology; Writing – original draft; Writing – review and editing.
René Kager: Conceptualization; Supervision; Writing – review and editing.
Patrick C. M. Wong: Conceptualization; Funding acquisition; Project administration; Supervision; Writing – review and editing.

Declaration of conflicting interests

The author(s) declared the following potential conflicts of interest with respect to the research, authorship, and/or publication of this article: Patrick C. M. Wong is the founder of a technological startup company in Hong Kong supported by a Hong Kong government technological startup scheme for universities; the research reported here has no association with the company. All other authors declare no conflict of interest.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by the University Grants Committee (HKSAR) (Grant Number RGC14605119), The Chinese University of Hong Kong (CUHK) – Nanyang Technological University Singapore (NTU) – Western Sydney University (WSU) Joint Laboratory for Infant Research, and CUHK-UU Joint Center for Language, Mind, and Brain.

References

Boersma, P., & Weenink, D. J. M. (2017). *Praat: Doing phonetics by computer* [Computer program].

Burnham, D., Kitamura, C., & Vollmer-Conna, U. (2002). What's new, pussycat? On talking to babies and animals. *Science*, *296*(5572), 1435. https://doi.org/10.1126/science.1069587

Burnham, D., Singh, L., Kasisopa, B., Wong, P. C. M., Fu, C. S. L., Wewalaarachchi, D., Liu, L., Onsuwan, C., Chen, A., & Kalashnikova, M. (2018, May 17–19). The Tone Atlas, step 2 : Perceptual salience of Thai, Cantonese, Beijing and Singaporean Mandarin tones for tone and non-tone language listeners [Paper presentation]. *Abstracts of the 28th Annual Meeting of the Southeast Asian Linguistics Society, Wenzao Ursuline University of Languages*, Kaohsiung, Taiwan.

Burnham, D., Singh, L., Mattock, K., Woo, P. J., & Kalashnikova, M. (2018). Constraints on tone sensitivity in novel word learning by monolingual and bilingual infants: Tone properties are more influential than tone familiarity. *Frontiers in Psychology*, *8*, Article 2190. https://doi.org/10.3389/fpsyg.2017.02190

Cheng, M. C., & Chang, K. C. (2014). Tones in Hakka infant-directed speech: An acoustic perspective. *Language and Linguistics*, *15*(3), 341–390. https://doi.org/10.1177/1606822X14520662

Cristia, A. (2013). Input to language: The phonetics and perception of infant-directed speech. *Linguistics and Language Compass*, *7*(3), 157–170. https://doi.org/10.1111/lnc3.12015

Cristia, A., & Seidl, A. (2014). The hyperarticulation hypothesis of infant-directed speech. *Journal of Child Language*, *41*(4), 913–934. https://doi.org/10.1017/S0305000912000669

Dodane, C., & Al-Tamimi, J. (2007, August 6–10). An acoustic comparison of vowel systems in adult-directed-speech and child-directed speech: Evidence from French, English & Japanese [Conference session]. *16th International Congress of Phonetics Sciences*, Saarbrücken, Germany.

Englund, K. T., & Behne, D. M. (2005). Infant directed speech in natural interaction—Norwegian vowel quantity and quality. *Journal of Psycholinguistic Research*, *34*(3), 259–280. https://doi.org/10.1007/s10936-005-3640-7

Fenson, L., Dale, P., Reznick, J., Thal, D., Bates, E., Hartung, J., Pethick, S., & Reilly, J. (1993). *The MacArthur Communicative Development Inventories: User's guide and technical manual*. Paul H. Brookes.

Fernald, A., Pinto, J. P., Swingley, D., Weinberg, A., & McRoberts, G. W. (1998). Rapid gains in speed of verbal processing by infants in the 2nd year. *Psychological Science*, *9*(3), 228–231. https://doi.org/10.1111/1467-9280.00044

Fernald, A., & Simon, T. (1984). Expanded intonation contours in mothers' speech to newborns. *Developmental Psychology*, *20*(1), 104–113. https://doi.org/10.1037/0012-1649.20.1.104

Fernald, A., Taeschner, T., Dunn, J., Papousek, M., De Boysson-Bardies, B., & Fukui, I. (1989). A cross-language study of prosodic modifications in mothers' and fathers' speech to preverbal infants. *Journal of Child Language*, *16*(3), 477–501. https://doi.org/10.1017/S0305000900010679

Floccia, C., Keren-Portnoy, T., DePaolis, R., Duffy, H., Delle Luche, C., Durrant, S., White, L., Goslin, J., & Vihman, M. (2016). British English infants segment words only with exaggerated infant-directed speech stimuli. *Cognition*, *148*, 1–9. https://doi.org/10.1016/j.cognition.2015.12.004

Foursha-Stevenson, C., Schembri, T., Nicoladis, E., & Eriksen, C. (2017). The influence of child-directed speech on word learning and comprehension. *Journal of Psycholinguistic Research*, *46*(2), 329–343. https://doi.org/10.1007/s10936-016-9441-3

Gauthier, B., & Shi, R. (2011). A connectionist study on the role of pitch in infant-directed speech. *The Journal of the Acoustical Society of America*, *130*(6), EL380–EL386. https://doi.org/10.1121/1.3653546

Graf Estes, K., & Hay, J. F. (2015). Flexibility in bilingual infants' word learning. *Child Development*, *86*(5), 1371–1385. https://doi.org/10.1111/cdev.12392

Graf Estes, K., & Hurley, K. (2013). Infant-directed prosody helps infants map sounds to meanings. *Infancy*, *18*(5), 797–824. https://doi.org/10.1111/infa.12006

Grieser, D. A. L., & Kuhl, P. K. (1988). Maternal speech to infants in a tonal language: Support for universal prosodic features in motherese. *Developmental Psychology*, *24*(1), 14–20. https://doi.org/10.1037/0012-1649.24.1.14

Han, M. (2019). *The role of prosodic input in word learning: A cross-linguistic investigation of Dutch and Mandarin Chinese infant-directed speech* [Doctoral dissertation, LOT Publications].

Han, M., de Jong, N. H., & Kager, R. (2018). Lexical tones in Mandarin Chinese infant-directed speech: Age-related changes in the second year of life. *Frontiers in Psychology*, *9*, Article 434. https://doi.org/10.3389/fpsyg.2018.00434

Hartman, K. M., Ratner, N. B., & Newman, R. S. (2017). Infant-directed speech (IDS) vowel clarity and child language outcomes. *Journal of Child Language*, *44*(5), 1140–1162. https://doi.org/10.1017/S0305000916000520

Hay, J. F., Cannistraci, R. A., & Zhao, Q. (2019). Mapping non-native pitch contours to meaning: Perceptual and experiential factors. *Journal of Memory and Language*, *105*, 131–140. https://doi.org/10.1016/j.jml.2018.12.004

Hay, J. F., Graf Estes, K., Wang, T., & Saffran, J. R. (2015). From flexibility to constraint: The contrastive use of lexical tone in early word learning. *Child Development*, *86*(1), 10–22. https://doi.org/10.1111/cdev.12269

Kalashnikova, M., & Burnham, D. (2018). Infant-directed speech from seven to nineteen months has similar acoustic properties but different functions. *Journal of Child Language*, *45*(5), 1035–1053. https://doi.org/10.1017/S0305000917000629

Kitamura, C., Thanavishuth, C., Burnham, D., & Luksaneeyanawin, S. (2002). Universality and specificity in infant-directed speech: Pitch modifications as a function of infant age and sex in a tonal and non-tonal language. *Infant Behavior and Development*, *24*(4), 372–392. https://doi.org/10.1016/S0163-6383(02)00086-3

Kuhl, P. K., Andruski, J. E., Chistovich, I. A., Chistovich, L. A., Kozhevnikova, E. V., Ryskina, V. L., Stolyarova, E. I., Sundberg, U., & Lacerda, F. (1997). Cross-language analysis of phonetic units in language addressed to infants. *Science*, *277*(5326), 684–686. https://doi.org/10.1126/science.277.5326.684

Liu, F., Jiang, C., Francart, T., Chan, A. H., & Wong, P. C. (2017). Perceptual learning of pitch direction in congenital amusia: Evidence from Chinese speakers. *Music Perception: An Interdisciplinary Journal*, *34*(3), 335–351. https://doi.org/10.1525/mp.2017.34.3.335

Liu, F., Yin, Y., Chan, A. H., Yip, V., & Wong, P. C. (2021). Individuals with congenital amusia do not show context-dependent perception of tonal categories. *Brain and Language*, *215*, Article 104908. https://doi.org/10.1016/j.bandl.2021.104908

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

Liu, H. M., Tsao, F. M., & Kuhl, P. K. (2007). Acoustic analysis of lexical tone in Mandarin infant-directed speech. *Developmental Psychology*, *43*(4), 912–917. https://doi.org/10.1037/0012-1649.43.4.912

Ma, W., Golinkoff, R. M., Houston, D. M., & Hirsh-Pasek, K. (2011). Word learning in infant-and adult-directed speech. *Language Learning and Development*, *7*(3), 185–201. https://doi.org/10.1080/15475441.2011.579839

Ma, W., Zhou, P., Singh, L., & Gao, L. (2017). Spoken word recognition in young tone language learners: Age-dependent effects of segmental and suprasegmental variation. *Cognition*, *159*, 139–155. https://doi.org/10.1016/j.cognition.2016.11.011

Mani, N., & Pätzold, W. (2016). Sixteen-month-old infants' segment words from infant- and adult-directed speech. *Language Learning and Development*, *12*(4), 499–508. https://doi.org/10.1080/15475441.2016.1171717

Novitskiy, N., Maggu, A. R., Lai, C. M., Chan, P. H., Wong, K. H., Lam, H. S., . . .Wong, P. C. (2022). Early development of neural speech encoding depends on age but not native language status: Evidence from lexical tone. *Neurobiology of Language*, *3*(1), 67–86. https://doi.org/10.1162/nol_a_00049

Ota, M., Yamane, N., & Mazuka, R. (2018). The effects of lexical pitch accent on infant word recognition in Japanese. *Frontiers in Psychology*, *8*, Article 2354. https://doi.org/10.3389/fpsyg.2017.02354

Peter, V., Kalashnikova, M., Santos, A., & Burnham, D. (2016). Mature neural responses to infant-directed speech but not adult-directed speech in pre-verbal infants. *Scientific Reports*, *6*, Article 34273. https://doi.org/10.1038/srep34273

Rattanasone, N. X., Burnham, D., & Reilly, R. G. (2013). Tone and vowel enhancement in Cantonese infant-directed speech at 3, 6, 9, and 12 months of age. *Journal of Phonetics*, *41*(5), 332–343. https://doi.org/10.1016/j.wocn.2013.06.001

Robertson, V. S., von Hapsburg, D., & Hay, J. S. (2017). The effect of hearing loss on novel word learning in infant- and adult-directed speech. *Ear and Hearing*, *38*(6), 701–713. https://doi.org/10.1097/AUD.0000000000000455

Singh, L., & Foong, J. (2012). Influences of lexical tone and pitch on word recognition in bilingual infants. *Cognition*, *124*(2), 128–142. https://doi.org/10.1016/j.cognition.2012.05.008

Singh, L., & Fu, C. S. L. (2016). A new view of language development: The acquisition of lexical tone. *Child Development*, *87*(3), 834–854. https://doi.org/10.1111/cdev.12512

Singh, L., Goh, H. H., & Wewalaarachchi, T. D. (2015). Spoken word recognition in early childhood: Comparative effects of vowel, consonant and lexical tone variation. *Cognition*, *142*, 1–11. https://doi.org/10.1016/j.cognition.2015.05.010

Singh, L., Hui, T. J., Chan, C., & Golinkoff, R. M. (2014). Influences of vowel and tone variation on emergent word knowledge: A cross-linguistic investigation. *Developmental Science*, *17*(1), 94–109. https://doi.org/10.1111/desc.12097

Singh, L., Nestor, S., Parikh, C., & Yull, A. (2009). Influences of infant-directed speech on early word recognition. *Infancy*, *14*(6), 654–666. https://doi.org/10.1080/15250000903263973

Singh, L., Poh, F. L. S., & Fu, C. S. L. (2016). Limits on monolingualism? A Comparison of monolingual and bilingual infants' abilities to integrate lexical tone in novel word learning. *Frontiers in Psychology*, *7*, Article 667. https://doi.org/10.3389/fpsyg.2016.00667

Singh, L., & Wewalaarachchi, T. D. (2020). Effects of age and bilingualism on sensitivity to native and nonnative tone variation: Evidence from spoken word recognition in Mandarin Chinese Learners. *Developmental Psychology*, *56*(9), 1642–1656. https://doi.org/10.1037/dev0001041

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

Tardif, T., Fletcher, P., Zhang, Z. X., & Liang, W. L. (2008). *The user manual of Chinese Communicative Development Inventories: Mandarin and Cantonese*. Peking University Medical Press.

Wewalaarachchi, T. D., Wong, L. H., & Singh, L. (2017). Vowels, consonants, and lexical tones: Sensitivity to phonological variation in monolingual Mandarin and bilingual English–Mandarin toddlers. *Journal of Experimental Child Psychology*, *159*, 16–33. https://doi.org/10.1016/j.jecp.2017.01.009

Wong, P. C. M., Chandrasekaran, B., & Zheng, J. (2012). The Derived Allele of ASPM Is Associated with Lexical Tone Perception. *PLoS ONE*, *7*(4), e34243. https://doi.org/10.1371/journal.pone.0034243

Wong, P. C., Kang, X., Wong, K. H., So, H. C., Choy, K. W., & Geng, X. (2020). ASPM-lexical tone association in speakers of a tone language: Direct evidence for the genetic-biasing

hypothesis of language evolution. *Science Advances*, *6*(22), Article eaba5090. https://doi.org/10.1126/sciadv.aba5090

Xu, N. (2008). *Tones and vowels in Cantonese infant directed speech: Hyperarticulation during the first 12 months of infancy* [Doctoral dissertation, University of Western Sydney].

Yip, M. (2002). *Tone*. Cambridge University Press. https://doi.org/10.1017/CBO9781139164559

Zhang, Y., Koerner, T., Miller, S., Grice-Patil, Z., Svec, A., Akbari, D., Tusler, L., & Carney, E. (2011). Neural coding of formant-exaggerated speech in the infant brain. *Developmental Science*, *14*(3), 566–581. https://doi.org/10.1111/j.1467-7687.2010.01004.x