Phonetic enhancement of Mandarin vowels and tones: Infant-directed speech and Lombard speech

Ping Tang, , Nan Xu Rattanasone, , Ivan Yuen, and , and Katherine Demuth

Citation: The Journal of the Acoustical Society of America **142**, 493 (2017); doi: 10.1121/1.4995998
View online: http://dx.doi.org/10.1121/1.4995998
View Table of Contents: http://asa.scitation.org/toc/jas/142/2
Published by the Acoustical Society of America

Phonetic enhancement of Mandarin vowels and tones: Infant-directed speech and Lombard speech

Ping Tang,[a)] Nan Xu Rattanasone, Ivan Yuen, and Katherine Demuth
Department of Linguistics, ARC Centre of Excellence in Cognition and its Disorders, Macquarie University, Sydney, 16 University Avenue, Australian Hearing Hub, Balaclava Road, North Ryde, New South Wales 2109 Australia

(Received 28 October 2016; revised 1 July 2017; accepted 5 July 2017; published online 1 August 2017)

Speech units are reported to be hyperarticulated in both infant-directed speech (IDS) and Lombard speech. Since these two registers have typically been studied separately, it is unclear if the same speech units are hyperarticulated in the same manner between these registers. The aim of the present study is to compare the effect of register on vowel and tone modification in the tonal language Mandarin Chinese. Vowel and tone productions were produced by 15 Mandarin-speaking mothers during interactions with their 12-month-old infants during a play session (IDS), in conversation with a Mandarin-speaking adult in a 70 dBA eight-talker babble noise environment (Lombard speech), and in a quiet environment (adult-directed speech). Vowel space expansion was observed in IDS and Lombard speech, however, the patterns of vowel-shift were different between the two registers. IDS displayed tone space expansion only in the utterance-final position, whereas there was no tone space expansion in Lombard speech. The overall pitch increased for all tones in both registers. The tone-bearing vowel duration also increased in both registers, but only in utterance-final position. The difference in speech modifications between these two registers is discussed in light of speakers' different communicative needs.
[http://dx.doi.org/10.1121/1.4995998]

[SHF] Pages: 493–503

I. INTRODUCTION

Speakers use different speech styles (i.e., registers) in different conditions, and each speech style has its own acoustic characteristics. For example, when talking to an infant, speakers tend to use an exaggerated speech style, resulting in a unique speech register known as "infant-directed speech" or IDS (Kuhl *et al.*, 1997; Burnham *et al.*, 2002). In IDS, suprasegmental and segmental characteristics are modified. Relative to adult-directed speech (ADS), IDS exhibits higher overall pitch (Fernald *et al.*, 1989; Burnham *et al.*, 2002; Fernald and Simon, 1984), larger pitch variability (Fernald *et al.*, 1989), slower speaking rate (Fernald and Simon, 1984), and enhanced phonemic contrasts (Kuhl *et al.*, 1997; Burnham *et al.*, 2002; Liu *et al.*, 2003; Liu *et al.*, 2007). The general consensus is that speech modifications in IDS arise for the following reasons: to attract infants' attention, to communicate positive affect, and to facilitate the infants' language learning (cf. Song *et al.*, 2010, for discussion).

Another speech style occurs when people are talking in a noisy environment, which is known as Lombard speech (Lombard, 1911). This speech style also displays a set of acoustic enhancements. For example, relative to speech produced in quiet, Lombard speech exhibits higher pitch, greater intensity, longer vowel duration, higher first formant frequency (F1), and a tendency to enhance phonemic contrasts (see Junqua, 1996, for a review). It is suggested that talkers modify speech so as to maintain self-monitoring ability through the speech feedback loop and to improve the communicative effectiveness between speakers in a noisy environment (Zhao and Jurafsky, 2009).

It seems that these two speech registers undergo modifications with the common goal of accommodating to the communicative needs of the listener, and thus result in similar acoustic modification and speech hyperarticulation. However, to our knowledge, there is only one published study that has compared the acoustic modification between IDS and Lombard speech (Wassink *et al.*, 2007). In this study, Jamaican Creole and Jamaican English vowels were compared across three speech conditions: IDS, Lombard speech, and hyperspeech. Wassink *et al.* found similar pitch and intensity modifications in IDS and Lombard speech, while the two registers differed in how vowels and segmental duration were modified. It is unknown if, in a tonal language such as Mandarin Chinese which uses both vowels and lexical tones to convey linguistic meaning, these linguistic units also undergo different modifications across registers. This question was addressed in the current study by directly comparing vowel and lexical tone hyperarticulation in Mandarin IDS and Lombard speech.

In IDS, vowel contrasts are often enhanced, leading to vowel space expansion (e.g., Kuhl *et al.*, 1997; Burnham *et al.*, 2002). It has been argued that in this "stretched" vowel space, vowel categories are acoustically less overlapped and the vowel contrasts are maximized, which could provide infants with well-specified acoustic cues to help them develop vowel categories (Kuhl *et al.*, 1997). It has also been claimed that the expanded vowel space is a universal characteristic

[a)]Electronic mail: ping.tang1@students.mq.edu.au

employed to promote language learning (Kuhl *et al.*, 1997), and that IDS reflects caregivers' teaching effort, including the expression of positive affect or attracting infants' attention (Uther *et al.*, 2007). However, IDS studies are mixed in terms of the presence vs absence of vowel space expansion (see Benders, 2013, for a review), with some studies actually reporting a surprisingly compressed vowel spaces in IDS (Englund and Behne, 2006; Benders, 2013). Therefore, whether vowel hyperarticulation is a universal feature of IDS and whether the nature of vowel hyperarticulation is a result of caregivers' teaching effort are still open questions.

IDS has also been shown to alter suprasegmental aspects of speech. Lexical tone languages, which use contrastive tones to indicate the meanings of words (Yip, 2002), exhibit tone hyperarticulation in IDS as well. Evidence from Taiwan Mandarin, Hakka and Cantonese has shown that, relative to ADS, acoustic differences among tones—pitch height differences, pitch range differences, and pitch duration differences—are made more distinct in IDS (Liu *et al.*, 2007; Xu Rattanasone *et al.*, 2013; Cheng and Chang, 2014). It is therefore argued that, like vowel hyperarticulation, tone hyperarticulation in IDS could also increase infants' abilities to distinguish language-specific phonemic units and thus facilitate language learning (Liu *et al.*, 2007). It is also claimed that the larger pitch contours of IDS could improve infants' ability to discriminate vowels (Trainor and Desjardins, 2002). However, there is evidence showing that, in Thai IDS addressed to 3- to 12-month olds, the tonal contour was less identifiable than in ADS (Kitamura *et al.*, 2002). This divergence among studies challenges the universality of phonemic tone hyperarticulation in IDS, prompting further investigation of another tone language.

Although vowel hyperarticulation (Liu *et al.*, 2003) and tone hyperarticulation (Liu *et al.*, 2007) in IDS have been studied in Taiwan Mandarin, no study has explored IDS in Northern Mandarin, the dialect of Mandarin spoken in northern China (Li, 1973). Taiwan Mandarin and Northern Mandarin share the same vowel and tone inventory, but the phonetic realizations of vowels and tones are different across dialects. For instance, compared with Taiwan Mandarin, Northern Mandarin has a larger vowel space, a larger pitch range, and a higher mean pitch (Deng *et al.*, 2006). This implies that, relative to Taiwan Mandarin, vowels and tones in Northern Mandarin might be harder to hyperarticulate in speech registers such as IDS and Lombard speech. Examining possible vowel and tone hyperarticulation in Northern Mandarin therefore provides an ideal opportunity to further examine the universality of phonemic hyperarticulation in IDS.

Similar to IDS, Lombard speech also modifies vowels (Junqua, 1996) and tones (Zhao and Jurafsky, 2009; Kasisopa *et al.*, 2014). However, studies of Lombard speech have provided conflicting evidence for vowel hyperarticulation, and there are few reports on tone hyperarticulation. For example, there is evidence showing that, relative to speech produced in a quiet environment, vowel space is expanded in Lombard speech (Bond *et al.*, 1989), especially when an interlocutor was present (Cooke and Lu, 2010). However, Davis and Kim (2010) and Godoy *et al.* (2014) did not observe vowel space expansion in Lombard speech. In spite of the mixed vowel space expansion results, these studies uniformly reported that formant frequencies, especially F1, are increased in Lombard speech. These results imply that the manifestation of vowel hyperarticulation might be language-dependent. Perhaps speakers of some languages choose to maximize vowel contrasts, whereas others prefer to improve communicative efficiency by reinforcing the speech signal through higher F1.

While most studies on Lombard speech have focused on vowel hyperarticulation, only two studies have examined tone hyperarticulation. On the one hand, Zhao and Jurafsky (2009) did not find enhanced tonal contrasts in Cantonese Lombard speech under 75 dB sound pressure level white noise, although overall pitch was higher for all tones in the noise condition. On the other, Kasisopa *et al.* (2014) reported enhanced tonal contrast in Thai Lombard speech under the same noise condition as in Zhao and Jurafsky (2009). Although the use of tone hyperarticulation in Thai Lombard speech differed from that in Cantonese, Thai Lombard speech also displayed an overall increase in pitch for all tones in the noise condition, with enhanced tonal contrasts towards the later part of the tone contours. These conflicting findings suggest the need to examine another tone language to test whether tone hyperarticulation is an intrinsic feature of Lombard speech.

The present study was therefore aimed at investigating the phonetic modification of Northern Mandarin vowels and tones in IDS and Lombard speech. In particular, we asked whether vowels and tones would be hyperarticulated in both IDS and Lombard speech in Northern Mandarin. If so, would they be expanded in the same way? To explore these issues, IDS, Lombard speech, and ADS were elicited from the same speakers (15 Northern-Mandarin speaking mothers of 12-month-old infants). Their vowel and tone productions were compared across registers, where the ADS served as the control for comparison. We had the following predictions.

Hypothesis 1 (H1): Vowel space would be expanded in Northern Mandarin IDS and Lombard speech.

Hypothesis 2 (H2): Point vowels would be hyperarticulated in both IDS and Lombard speech, but the direction of the hyperarticulation would differ between the two speech registers: vowels in IDS might be made maximally distinct in the closed-open (F1) and/or in the front-back (F2) dimension, whereas the vowels in Lombard speech might move towards a higher F1.

Hypothesis 3 (H3): Tone space would be expanded in Northern Mandarin IDS as well as in Lombard speech.

Hypothesis 4 (H4): The overall pitch and syllable/word duration for all tones would be increased in both IDS and Lombard speech (Zhao and Jurafsky, 2009; Kasisopa *et al.*, 2014).

II. METHOD

A. Participants

Fifteen Mandarin-speaking mothers and their 12-month-old infants [mean = 12 months, standard deviation (SD) = 0.99] were recruited in Sydney. All the mothers were born and raised in Mandarin-speaking families in Northern China (i.e., Beijing, Hebei Province, or northeastern China). Their age ranged from 18 to 34 years (mean = 24 years, SD = 4.93). At the time of study, their stay in Australia ranged from 1 to 8 years (mean = 5 years, SD = 3.27). All mothers were the main caregivers of their infants and spoke only Mandarin Chinese to their infants at home. According to parental report, no infants tested had speech, hearing, or language problems.

B. Stimuli

Six disyllabic Mandarin words were used to elicit the three target point vowels (/i/, /a/, and /u/) and the three target lexical tones (T1: level, T2: rising, and T4: falling).[1] All were nouns that could be illustrated with toys. The three-point vowels were selected to examine the area of vowel space across registers. T1, T2, and T4 were selected because they have simple tone contours and differ from the complex contour tone (T3). The three simple contour tones are distinguished in terms of pitch at the onset and offset of the syllable (T1: high pitch onset, high pitch offset; T2: mid pitch onset, high pitch offset; T4: high pitch onset, low pitch offset), which are used to define the tone space, in line with Barry and Blamey (2004) and Xu Rattanasone *et al.* (2013). If tone hyperarticulation occurs, the acoustic tone space will also expand.

The three point vowels /i/, /a/, and /u/ occurred in V_1 position of three $C_1V_1.C_2V_2$ disyllabic Mandarin words (see Table I). The target vowels were preceded by a bilabial aspirated stop consonant /p^h/ to minimize co-articulation between C_1 and V_1, while keeping the place of articulation as similar as possible in the following C_2 (i.e., a retroflex aspirated alveo-palatal affricate /tʂ^h/, an aspirated alveolo-palatal affricate/tɕ^h/, and an aspirated alveolar stop /t^h/). Although C_2 for all was not identical across items, we think that this would have minimal effect on the focus of the investigation, namely, the effect of register on the vowel. Having controlled for the preceding consonant and place of articulation of the following consonants, we also controlled for the tone of the first syllable (i.e., T2).

The three target tones T1, T2, and T4 were carried by the second syllable of the test words (see Table I). The tone of the first syllable was held constant to be T1 (i.e., level) so that the tonal coarticulation across the two syllables could be minimized, and the segment of the second syllable was held constant to be /tʂu/. See Table I for a list of all six test items.

TABLE I. Target vowels (/i/, /a/, and /u/) and lexical tones (T1, T2, and T4) in disyllabic words. The columns from left to right illustrate the three target vowels and tones, the target words written in Chinese characters, the corresponding meanings, and the phonetic transcriptions of the target words with the target syllables in bold.

Target vowels/tones	Target words	Meaning	Phonetic transcription
/i/	皮球	Ball	**/p^{h}i2** tɕ^{h}iou2/
/a/	爬虫	Worm	**/p^{h}a2** tʂ^{h}uŋ2/
/u/	菩提	Bodhi	**/p^{h}u2** t^{h}i2/
T1	珍珠	Pearl	/tʂən1 **tʂu1/**
T2	山竹	Mangosteen	/ʂan1 **tʂu2/**
T4	光柱	Light stick	/kuaŋ1 **tʂu4/**

C. Procedure

Prior to testing, mothers and their infants were invited into a sound-treated room to become familiar with the new environment. After about 10 min, the testing phase began, with each mother taking part in three speech production tasks in the same order: (1) IDS, (2) Lombard speech, (3) ADS. This helped ensure that data collection from the infants took place before they became fussy or tired.

In the IDS task, mothers were fitted with a head-mounted condenser microphone (AKG-C520) which was connected to a solid-state recorder (Marantz PMD661MKII) in a shoulder bag. Mothers wore the shoulder bag so that they were free to move around during the recording session. The same solid-state recorder and head-mounted microphone were used for recording in all three tasks. To elicit IDS, the mother and her infant engaged in a play session using a set of toys. Six toys corresponding to the six target words were provided, and each toy was labelled with the corresponding target word in written Chinese. These toys were randomly arranged into three cloth bags, to ensure counterbalancing on the order of the toys presented across mothers. The mothers were instructed to play with their infants using the toys as they normally would at home, and to use the written labels to refer to the toys when talking to the infant. Similar to the procedures used in previous IDS studies (i.e., Wassink *et al.*, 2007; Majorano *et al.*, 2013), a Mandarin-speaking experimenter (the first author) was present in the studio to ensure the experiment went successfully. At the same time, the experimenter kept count of the number of tokens the mother produced for each target word. When the mother had produced a minimum of ten tokens for each target word, the experimenter provided the mother with the second bag of toys. This continued until all three bags of toys were used during the play session.

In the Lombard speech task, only the mother and the experimenter were present in the room. The mother was encouraged to describe to the experimenter her experience in using the labels to refer to the toys in the play session with her infant. The mother and the experimenter talked to one another while listening to a 70 dBA Chinese eight-talker babble noise via open-ear headphones (AKG-K612 PRO). The Digitech-QM1591 decibel meter was used to calibrate the sound level played via the headphones. Before the task, the decibel meter was positioned on the headphones to make sure that the sound level of the auditory output from the headphones stayed at 70 dBA. The experimenter was keeping count of tokens the mother produced for each target word while talking to the mother, and the conversation continued until the mother had produced a minimum of eight repetitions for each toy. In cases when the mother did not produce at least eight tokens for a toy, the experimenter would ask several questions to help elicit more tokens from the mother, such as "What was the color of X?" and "Did your baby like X?," etc.

In the ADS task, the procedure and the minimal number of repetitions were identical to those in the Lombard speech task. The only difference was that the mother and the experimenter talked to one another in a quiet environment. All the recordings were made with a sampling rate of 44.1 kHz and a 16-bit quantization.

Mothers of 12-month olds were selected in order to compare our results with previous findings in Taiwan Mandarin IDS (Liu *et al.*, 2003; Liu *et al.*, 2007: 10- to 12-month olds). The 70 dBA ight-talker babble noise was used in the present study as the noise mask to elicit Lombard speech in accord with the observation that N-talker babble noise resembles everyday speech noise (Simpson and Cooke, 2005) and with the mean intensity level (70 dBA) of young females' conversional speech (Morris and Brown, 1994).

D. Coding and measurements

The data were coded in PRAAT (Boersma and Weenink, 2016) with the aid of spectrograms and waveforms, as outlined below. Ten percent of the items were re-coded by a second trained native speaker. Interrater reliability was 95.56% for target vowel carriers and 95.77% for target tone carriers, as measured by the correlation of vowel duration between raters.

1. Vowels

The three target point vowels (/i/, /a/, /u/) across the three register types (IDS, Lombard speech, ADS) were annotated. To identify the beginning and end of the target point vowels, clear F2 onset and offset, respectively, were used. On the basis of the annotated vowel intervals, the trajectories of F1 and F2 were extracted in PRAAT. Averaged values of the F1 and the F2 for each target point vowel were calculated from the middle portion (from 40% to 60%) of the vowel interval (Fig. 1), as the vowel target (i.e., the most stable formants with minimal influence from the formant transitions) is typically reached towards the middle part of the vowel. The Burg method (Burg, 1975) was used in PRAAT to track the formant values, calculated over a range from 0 to 5500 Hz. The length of the analysis window was 25 ms, and frequencies above 50 Hz were pre-emphasized. Formant values which were mistracked were hand-corrected in PRAAT, although the influence of this mistracking was small (a correlation analysis of formant values extracted before and after hand-correction yielded high correlation coefficients for F1 and F2: 0.87 and 0.9). All the formant values extracted were then transformed to Bark values to match the scale of human perception, using the following formula (Zwicker and Fastl, 1980):

$$\text{Bark} = 13 * \arctan(0.76 * \text{Hz}) + 3.5 * \arctan(\text{Hz}/7.5)^2.$$

The averaged F1 and F2 values of /i/, /a/, and /u/ were used to derive the vowel space area for each mother/participant in the IDS, Lombard speech, and ADS conditions using the following formula (Liu *et al.*, 2003).

Vowel space area $= |\{[F1i * (F2a - F2u) + F1a * (F2u - F2i) + F1u * (F2i - F2a)]/2|$, where F1i, F1a, and F1u are the F1 values of /i/, /a/, and /u/, respectively; F2i, F2a, and F2u are the F2 values of /i/, /a/, and /u/, respectively.

2. Tones

The three target lexical tones (T1, T2, T4) in the disyllabic words and the three register types (IDS, Lombard speech, ADS) were annotated. Since the realization of Mandarin tones is highly influenced by the tonal context, i.e., the preceding and the following tones (Xu, 1994), the positions of the target tones within utterances were also annotated. The target tones were always carried by the second syllable of the target disyllabic words, so the target tones appeared either in the "utterance-medial" position or the "utterance-final" position. To track the pitch contour of each tone, the onset and offset of the tone-bearing vowel were first identified on the basis of clear F2. Two pitch points (pitch onset and offset) were then measured from the vocalic portion of the target word, with the pitch onset extracted at 5% point and the pitch offset extracted at the 95% point (Fig. 2) so that tonal coarticulation and pitch perturbation from neighbouring consonants could be minimised. Pitch was tracked using a short-term autocorrelation algorithm in PRAAT. Pitch values were checked

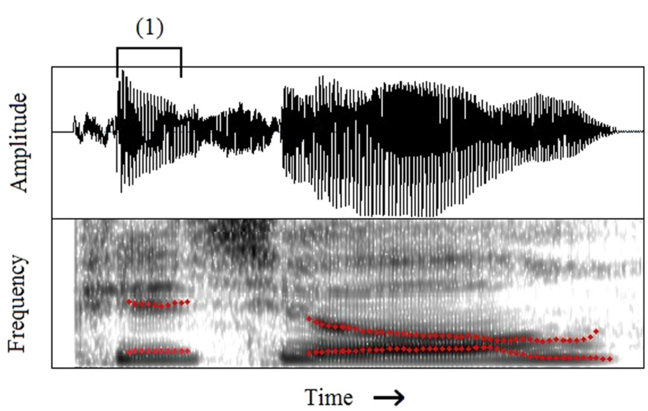

FIG. 1. (Color online) Waveform and spectrogram for the word /p^{h}i2 tɕ^{h}iou2/. This token was produced by a subject M1 in Lombard speech condition. (1) illustrates the vowel portion of target syllable /p^{h}i2/. Red dotted lines from the bottom up represent F1 and F2, respectively. The averaged F1 and F2 values of target vowel /i/ were extracted from the 40%–60% portion of (1).

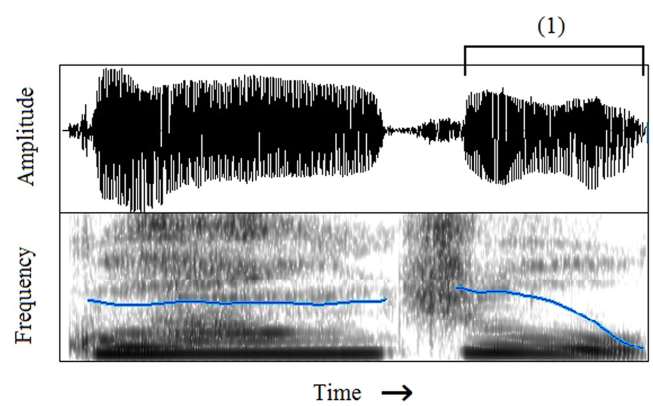

FIG. 2. (Color online) Waveform and spectrogram for the word /kuaŋ1 tʂu4/. This token was produced by subject M1 in the Lombard speech condition. (1) illustrates the vowel portion of target syllable /tʂu4/. Blue lines represent the pitch contours of two syllables. The pitch onset and offset of the target tone were extracted from the 5% and the 95% point of (1), respectively, and the tonal duration was measured across the duration of (1).

and manually revised to correct for the "doubling" or "halving" errors in pitch tracking. In the analysis, the pitch values were transformed to semitones from observed Hz values with 50 Hz as the reference to match the scale of human perception, using the following formula:

$$\text{Semitone} = 12 * \log 2(\text{target Hz}/50).$$

The averaged pitch onset and offset values of T1, T2, and T4 were then used to derive the tone space area in the IDS, Lombard speech, and ADS conditions for each mother/participant. The same formula for calculating the vowel space area was used to compute the tone space area, as below.

Tone space area $= |\{[\text{T1onset} * (\text{T4offset} - \text{T2offset}) + \text{T2onset} * (\text{T1offset} - \text{T4offset}) + \text{T4onset} * (\text{T2offset} - \text{T1offset})]/2|$. The tonal duration information was also extracted as the duration of the vowel interval (ms) in the target word.

E. Statistical analysis

A total of 2828 tokens were included in the analysis, with 1431 vowel tokens and 1397 tone tokens, as illustrated in Table II. An additional 141 tokens were excluded from the analysis for the following reasons: overlap with another sound, such as the infant's vocalization or noise made by toys or other environmental disturbance; the mother laughing or singing when producing the token; the token having been whispered; or mispronunciation.

The data were analysed using R (R Core Team, 2016). A linear mixed effect model was adopted for the comparison of vowel and tone parameters across registers, using LME4 package (Bates *et al.*, 2015) and LMERTEST package (Kuznetsova *et al.*, 2013). When a significant main effect of a multi-level factor or a significant interaction effect was observed, Tukey-HSD *post hoc* comparisons were then also performed on the multi-level factor, as well as interactions.

III. RESULTS

A. Vowels

1. *Vowel space area*

The F1-F2 two-dimensional vowel space areas of the three registers are presented in Fig. 3. To test H1, vowel space areas were compared across registers. A fixed factor "Register" (IDS, Lombard speech, and ADS) and a random factor "Subject" (15 subjects) were included in the model with a random intercept and a random slope for the fixed factor. The significance of the random slope in model fitting was first tested by a likelihood ratio test which showed that the exclusion of the random slope did not significantly affect the fit [$\chi^2(5) = 2.571$, $p = 0.766$]. Therefore, to keep the model parsimonious, the random slope was excluded from the model in further analyses.[2]

TABLE II. Number of vowel and tone tokens included in each register for further analysis. Vowel space and tone space were calculated based on the averaged data per participant.

	Register		
Target vowels/tones	IDS	Lombard speech	ADS
/a/	169	151	146
/i/	241	143	144
/u/	162	140	135
Total vowels	572	434	425
T1	199	152	139
T2	176	139	136
T4	182	135	139
Total tones	557	426	414

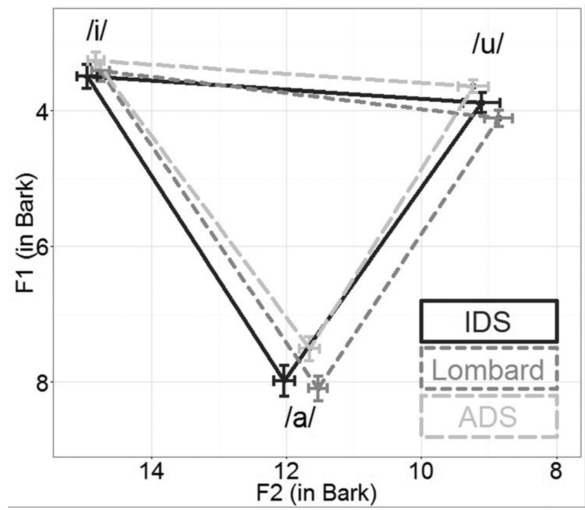

FIG. 3. The vowel space, defined by F1 and F2 in Bark, with the three point vowels /i/, /a/, and /u/ in infant-directed speech (IDS), Lombard speech (Lombard), and adult-directed speech (ADS). The corners of each space represent the group means of averaged F1 and F2 values for /i/, /a/, and /u/, respectively, with error bars showing the standard error of the group means of F1 and F2 values.

The results of the comparison are presented in Table III. Consistent with H1, the vowel space was expanded in both IDS (mean = 12.52 Bark^2, SD: 1.98; IDS – ADS: $\beta = 1.23$, $SE = 0.48$, $t = 2.56$, $p < 0.05$) and Lombard speech (mean = 12.65 Bark^2, SD = 1.43; Lombard speech – ADS: $\beta = 1.63$, $SE = 0.48$, $t = 2.82$, $p < 0.01$) relative to ADS (mean = 11.28 Bark^2, SD = 1.68).

2. *Formant frequencies*

Table IV gives the F1 and F2 values across vowels and registers. To examine H2, F1 and F2 values of /i/, /a/, and /u/ were compared across the three registers. Two fixed factors Register and "Vowel" (/i/, /a/, and /u/) and the random factor Subject were included in models for F1 and F2 comparison,

TABLE III. Results of the test of effects of the fixed factors Register (IDS, Lombard speech, and ADS, where ADS was the reference level for comparison) on vowel space using a linear-mixed effects model. Estimated differences across registers, standard errors, degrees of freedom, t values, and *p* values are provided.

	Estimate	SE	df	t	*p*
(Intercept)	11.28	0.44	31.66	25.52	<0.001[a]
Register (IDS)	1.23	0.48	28	2.56	0.016[a]
Register (Lombard)	1.36	0.48	28	2.82	0.009[a]

[a]Statistical significance with an alpha value of $p = 0.05$.

TABLE IV. Mean F1 and F2 (Bark) values (SD) across vowels (/a/, /i/, /u/) and registers (IDS, Lombard speech, ADS).

	F1			F2		
Register	/a/	/i/	/u/	/a/	/i/	/u/
IDS	7.92 (0.76)	3.41 (0.56)	3.92 (0.51)	11.98 (0.63)	14.88 (0.53)	9.10 (1.33)
Lombard speech	8.06 (0.63)	3.44 (0.45)	4.11 (0.40)	11.49 (0.44)	14.75 (0.44)	8.89 (0.90)
ADS	7.54 (0.75)	3.29 (0.42)	3.67 (0.34)	11.64 (0.62)	14.80 (0.39)	9.25 (1.05)

with random intercept and random slopes for the fixed factors. Since vowel duration has an effect on F1 and F2 (Gay, 1978), "Vowel Duration" (the duration of vowels, computed as the interval between vowel onset and offset) was also entered into the models as a covariate to minimise this effect. The random slopes were checked for significance by the likelihood ratio test for each model. In both models, the random slope of Subject for Register was excluded since it did not reach significance in model fitting [in the model for F1 comparison: $\chi^2(9) = 16.03$, $p = 0.066$; in the model for F2 comparison: $\chi^2(9) = 10.087$, $p = 0.344$].[3]

The results of the comparisons are presented in Table V.[4] The main effect of Register was significant for both F1 and F2, which means that vowel formant frequencies differ across registers. Since a significant two-way interaction between Register and Vowel was observed, a Tukey-HSD *post hoc* test was then performed on the model to compare the register difference in this interaction. The results of the *post hoc* test are presented in Table VI.

According to Table VI, IDS showed higher F1 and F2 values relative to ADS for /a/ and /u/, thus leading to a downward and forward shift of the acoustic vowel space. This was inconsistent with H2, which predicted vowels in IDS could be made maximally distinct in the closed-open (F1) and/or the front-back (F2) dimensions. In contrast, Lombard speech showed a higher F1 value for all three vowels, which led to a downward shift of the acoustic vowel space, consistent with H2.

B. Tones

1. Tone space area

The pitch onset-offset two-dimensional tone space areas in different positions (utterance-medial and utterance-final) across the three registers are presented in Fig. 4. To test H3, the tone space areas were compared across registers. Two fixed factors Register and "Position" (medial and final) and the random factor Subject were included in the model, where the random factor had its own intercept as well as random slopes for fixed factors. The significance of the random slopes was first checked and showed that all slopes were significant in model fitting. Therefore, all random slopes were included in the model.[5]

The results of the comparisons are presented in Table VII. The results showed that, relative to ADS, tone space was expanded in the utterance-final position of IDS. To further examine the tone space expansion across registers and positions, a Tukey-HSD *post hoc* test was also performed on the model, and the results of the *post hoc* test are presented in Table VIII.

Partially consistent with H3, tone space expansion was only observed in IDS (utterance-medial: mean = 2.17 St^2, SD = 0.58; utterance-final: mean = 8.18 St^2, SD = 2.55) in the utterance-final position compared with ADS (utterance-medial: mean = 1.10 St^2, SD = 0.17; utterance-final: mean = 2.29 St^2, SD = 0.52). Unexpectedly, Lombard speech (utterance-medial: mean = 1.33 St^2, SD = 0.26; utterance-final: mean = 4.03 St^2, SD = 0.64) did not show tone space expansion relative to ADS in either position.

2. Pitch height and tonal duration

Table IX gives the mean pitch and duration of the tone-bearing vowel across tones and registers. To test H4, the mean pitch and duration were compared across registers. Two models were built for the comparison of mean pitch and tone-bearing vowel duration, respectively. In both

TABLE V. Results of the test of effects of the fixed factors Register (IDS, Lombard speech, and ADS, where ADS was the reference level for comparison) and Vowel (/i/, /a/, and /u/, where /a/ was the reference level for comparison) on F1 and F2 using a linear-mixed effects model. Estimated differences, standard errors, degrees of freedom, t values, and *p* values are provided.

	F1					F2				
	Estimate	SE	df	t	*p*	Estimate	SE	df	t	*p*
(intercept)	6.38	0.16	96.44	40.04	<0.001[a]	11.82	0.17	175.83	70.63	<0.001[a]
Register (IDS)	1.12	0.14	1396.08	7.79	<0.001[a]	0.40	0.16	1391.59	2.43	<0.05[a]
Register (Lombard)	0.86	0.16	1388.72	5.33	<0.001[a]	−0.25	0.19	1397.00	−1.34	0.182
Vowel (/i/)	−3.12	0.21	166.44	−15.02	<0.001[a]	2.70	0.21	729.97	13.04	<0.001[a]
Vowel (/u/)	−2.79	0.19	111.53	−14.48	<0.001[a]	−0.93	0.21	171.62	−4.43	<0.001[a]
Register (IDS): Vowel (/i/)	−1.01	0.20	1402.32	−5.14	<0.001[a]	−0.16	0.22	1339.74	−0.72	0.470
Register (IDS): Vowel (/u/)	−0.74	0.18	1399.91	−4.00	<0.001[a]	−0.60	0.21	1403.29	−2.85	<0.01[a]
Register (Lombard): Vowel (/i/)	−1.04	0.23	1391.58	−4.51	<0.001[a]	0.20	0.27	1399.06	0.77	0.442
Register (Lombard): Vowel (/u/)	−0.49	0.22	1396.13	−2.27	<0.05[a]	0.17	0.25	1396.29	0.68	0.500

[a]Statistical significance with an alpha value of $p = 0.05$.

TABLE VI. Tukey-HSD pairwise comparisons of the different levels of Register (IDS, Lombard speech, and ADS) and the two-way interaction effect between Register and Vowel (/i/, /a/, and /u/) on F1 and F2. The estimated difference between the two levels of Register, standard error, t values, and HSD-adjusted p values are provided.

Register difference	F1				F2			
	Estimate	SE	t	p	Estimate	SE	t	p
				/i/				
IDS-ADS	0.14	0.06	2.17	0.077	0.01	0.07	0.17	0.984
Lombard-ADS	0.19	0.07	2.86	0.012[a]	−0.12	0.08	−1.48	0.299
IDS-Lombard	−0.06	0.05	−1.17	0.471	0.13	0.06	2.25	0.063
				/a/				
IDS-ADS	0.47	0.05	8.66	<0.001[a]	0.39	0.06	6.23	<0.001[a]
Lombard-ADS	0.45	0.06	7.53	<0.001[a]	0.13	0.07	1.83	0.160
IDS-Lombard	−0.01	0.06	−0.19	0.981	0.51	0.07	7.57	<0.001[a]
				/u/				
IDS-ADS	0.24	0.06	4.21	<0.001[a]	0.23	0.07	3.42	0.002[a]
Lombard-ADS	0.47	0.06	7.89	<0.001[a]	−0.02	0.07	−0.31	0.949
IDS-Lombard	−0.22	0.05	−4.32	<0.001[a]	0.21	0.06	3.43	0.002[a]

[a]Statistical significance with an alpha value of $p = 0.05$.

models, three fixed factors Register, Position, and "Tone" (T1, T2, and T4) and the random factor Subject were included, where the random factor had its own intercept as well as random slopes for fixed factors. The random slopes were checked for significance by the likelihood ratio test for each model. In the model for mean pitch comparison, the random slope of Subject for Position was excluded since it did not reach significance in model fitting[6] [$\chi^2(6) = 4.857$, $p = 0.562$]; in the model for tone-bearing vowel duration, the random slope of Subject was excluded for Tone since it did not reach significance in model fitting [$\chi^2(11) = 17.833$, $p = 0.086$].[7]

The results of the comparisons are presented in Table X. Since a significant two-way interaction between Register and Position was observed in both models, a Tukey-HSD *post hoc* test was performed on the two models to further compare the register difference across positions. The results

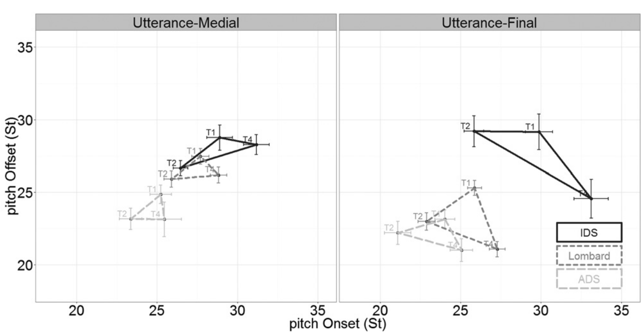

FIG. 4. The tone space, defined by pitch onset and offset in semitones, with the three lexical tones T1, T2, and T4 in utterance-medial and -final positions, in IDS, Lombard (speech), and ADS. The corners of each space represent the group means of averaged pitch onset and offset values for T1, T2, and T4, respectively, with error bars showing the standard errors of the group means of pitch onset and offset values.

TABLE VII. Results of the test of the effects of Register (IDS, Lombard speech, and ADS, where ADS was the reference level for comparison) and Position (utterance-medial and utterance-final, where the utterance-medial was the reference level) on the tone space area (St^2) using a linear-mixed effects model. Estimated differences, standard errors, degrees of freedom, t values, and p values are provided.

	Estimate	SE	df	t	p
(intercept)	1.87	2.09	40.75	0.89	0.377
Register (IDS)	2.42	3.88	26.37	0.62	0.538
Register (Lombard)	0.53	2.68	74.10	0.20	0.845
Position (final)	2.79	3.15	48.78	0.89	0.380
Register (IDS): Position (final)	13.33	3.78	75.00	3.53	<0.001[a]
Register (Lombard): Position (final)	2.76	3.78	75.00	0.73	0.467

[a]Statistical significance with an alpha value of $p = 0.05$.

are presented in Tables XI and XII. Other interactions were not further examined since they were not the focus of the present study.

Partially consistent with H4, the mean pitch of tones was highest in IDS, followed by Lombard speech, and then the lowest for ADS. The only exception was observed in utterance-medial position, when there was only a trend for the mean pitch to be higher in IDS than Lombard speech ($p = 0.068$). The tone-bearing vowel duration was also longer in both IDS and Lombard speech than in ADS, but only in utterance-final position.

IV. DISCUSSION

The results of this study showed that, in Northern Mandarin, vowels were hyperarticulated in both IDS and Lombard speech in terms of vowel space expansion, while the patterns of vowel-shift were different between the two registers: IDS showed an increase in both F1 and F2 of /a/ and /u/, and Lombard speech showed an increase in F1 for all three point vowels. IDS displayed tone space expansion in utterance-final position but not utterance-medial position, whereas Lombard speech showed no tone space expansion. In spite of the difference in tone space expansion between the two registers, the overall pitch increased for all tones in both registers. The tone-bearing vowel duration also increased in both registers but only in utterance-final position. These results indicate that IDS and Lombard speech modify vowels and tones in somewhat different ways.

TABLE VIII. Tukey-HSD pairwise comparisons on different levels of Register (IDS, Lombard speech, and ADS) in the two-way interaction effect between Register and Position (utterance-medial and utterance-final) on tone space area. The estimated difference between the two levels of Register, standard error, t values, and HSD-adjusted p values are provided.

	Utterance-medial				Utterance-final			
Register difference	Estimate	SE	t	p	Estimate	SE	t	p
IDS-ADS	2.42	4.02	0.62	0.808	15.75	3.88	4.06	0.001[a]
Lombard-ADS	0.53	2.68	0.20	0.979	3.30	2.68	1.23	0.440
IDS-Lombard	1.90	3.76	0.51	0.870	12.46	3.76	3.32	0.007[a]

[a]Statistical significance with an alpha value of $p = 0.05$.

TABLE IX. Mean pitch height (St) and tonal duration (s) values (SD) for tones (T1, T2, T4), positions (utterance-medial and utterance-final), and registers (IDS, Lombard speech, ADS).

		Pitch height			Tonal duration		
Register	Position	T1	T2	T4	T1	T2	T4
IDS	Medial	28.57 (4.45)	26.33 (3.68)	29.43 (4.56)	0.17 (0.08)	0.16 (0.05)	0.18 (0.07)
	Final	29.53 (4.97)	26.92 (5.06)	29.46 (6.29)	0.31 (0.15)	0.30 (0.15)	0.30 (0.16)
Lombard speech	Medial	27.74 (2.43)	25.74 (2.17)	27.49 (2.29)	0.16 (0.06)	0.17 (0.08)	0.18 (0.07)
	Final	25.27 (3.15)	22.54 (3.25)	24.10 (3.34)	0.29 (0.06)	0.31 (0.07)	0.26 (0.06)
ADS	Medial	24.97 (3.34)	23.32 (3.47)	24.37 (5.25)	0.16 (0.05)	0.16 (0.06)	0.17 (0.06)
	Final	23.71 (4.36)	20.93 (4.56)	23.19 (4.47)	0.26 (0.06)	0.29 (0.07)	0.22 (0.07)

The expanded vowel space in IDS and Lombard speech found in the current study is consistent with previous studies on IDS (e.g., Kuhl *et al.*, 1997; Liu *et al.*, 2003) and Lombard speech (e.g., Bond *et al.*, 1989). Our results show that, although Northern Mandarin has a larger vowel space than Taiwan Mandarin, the vowel space still expands in Northern Mandarin IDS, suggesting that the extent of vowel contrast enhancement in IDS is independent of vowel inventory size. This is similar to a previous study of clear speech (Smiljanić and Bradlow, 2005) which observed equivalent vowel space expansion in English and Croatian clear speech, despite the difference in vowel inventory size between the two languages. In the case of Lombard speech, the expanded vowel space also suggests that vowels are adjusted to be maximally contrastive so as to provide listeners with more distinctive vowel information and improve the speech intelligibility when the communicative environment is noisy.

Although the vowel space expands in both registers, the way this is achieved differs. In IDS, F1 and F2 increased for /a/ and /u/, but remained unchanged for /i/. This resulted in vowel space rotation around /i/, with a downward and forward shift (higher F1 and F2, respectively). Since /a/ moved further than /u/ during the vowel space rotation, this led to an expanded vowel space. In Lombard speech, F1 increased in all three point vowels, whereas F2 remained unchanged. The vowel space was therefore shifted downward in Lombard speech, relative to ADS. Since the extent of F1-increase differed across the three vowels, i.e., /a/ = /u/ > /i/, with a larger F1 shift in /a/ and /u/ than /i/, this also resulted in vowel space expansion (see Fig. 3).

In spite of the enhanced vowel contrasts in both IDS and Lombard speech, the different vowel-shift patterns observed for the two styles suggest that the mechanisms underlying these vowel shifts are different. The vowel modifications in IDS may be driven by mothers' articulatory adjustment to express positive affect to infants during the mother-infant interaction. The vowel modifications in Lombard speech, in contrast, may be the result of the speakers' articulatory adjustment to increase intensity in a noisy environment.

TABLE X. Results of the test of the effects of Register (IDS, Lombard speech, and ADS, where ADS was the reference level for comparison), Position (utterance-medial and utterance-final, where utterance-medial was the reference level for comparison), and Tone (T1, T2, and T4, where T1 was the reference level for comparison) on pitch height and tonal duration using a linear-mixed effects model. Estimated differences, standard errors, degrees of freedom, t values, and p values are provided.

	Pitch height					Tonal duration				
	Estimate	SE	df	t	*p*	Estimate	SE	df	t	*p*
(intercept)	25.08	0.80	22.9	31.21	<0.001[a]	0.16	0.01	140.8	15.57	<0.001[a]
Register (IDS)	3.60	0.82	37	4.40	<0.001[a]	0.01	0.01	114	0.90	0.373
Register (Lombard)	2.51	0.64	169	3.92	<0.001[a]	0.01	0.01	400	0.48	0.630
Position (final)	−1.68	0.64	378	−2.61	0.009[a]	0.10	0.02	109	6.24	<0.001[a]
Tone (T2)	−2.06	0.64	342	−3.24	0.001[a]	0.00	0.01	376	−0.27	0.784
Tone (T4)	−0.84	0.67	142	−1.26	0.211	0.01	0.02	137	0.80	0.423
Register (IDS): Position (final)	2.37	0.81	1339	2.93	0.003[a]	0.03	0.02	1343	2.08	0.016[a]
Register (Lombard): Position (final)	−0.50	0.86	1333	−0.59	0.556	0.02	0.02	1345	0.85	0.395
Register (IDS): Tone (T2)	−0.40	0.79	1331	−0.50	0.616	−0.01	0.02	1333	−0.61	0.540
Register (Lombard): Tone (T2)	0.06	0.84	1334	0.07	0.944	0.01	0.02	1339	0.47	0.639
Register (IDS): Tone (T4)	1.79	0.79	1333	1.16	0.261	−0.01	0.02	1340	−0.40	0.690
Register (Lombard): Tone (T4)	0.68	0.86	1336	0.79	0.428	0.00	0.02	1343	0.12	0.908
Position (final): Tone (T2)	−0.45	0.87	1336	−0.52	0.604	0.03	0.02	1337	1.65	0.099
Position (final): Tone (T4)	0.26	0.87	1335	0.30	0.768	−0.05	0.02	1337	−2.80	0.005[a]
Register (IDS): Position (final): Tone (T2)	0.48	1.16	1341	0.42	0.677	−0.01	0.03	1342	−0.38	0.708
Register (Lombard): Position (final): Tone (T2)	−0.49	1.23	1335	−0.40	0.691	−0.01	0.03	1346	−0.38	0.704
Register (IDS): Position (final): Tone (T4)	−1.13	1.15	1339	−0.98	0.325	0.04	0.03	1343	1.46	0.143
Register (Lombard): Position (final): Tone (T4)	−1.30	1.23	1337	−1.06	0.290	0.01	0.03	1346	0.48	0.628

[a]Statistical significance with an alpha value of $p = 0.05$.

TABLE XI. Tukey-HSD pairwise comparisons on different levels of Register (IDS, Lombard speech, and ADS) in the two-way interaction effect between Register and Position (utterance-medial and utterance-final) on pitch height. The estimated difference between two levels of Register, standard error, t values, and HSD-adjusted *p* values are provided.

Register difference	Utterance-medial				Utterance-final			
	Estimate	SE	t	*p*	Estimate	SE	t	*p*
IDS-ADS	4.06	0.70	5.77	<0.001[a]	6.22	0.71	8.73	<0.001[a]
Lombard-ADS	2.78	0.44	6.28	<0.001[a]	1.68	0.45	3.73	0.001[a]
IDS-Lombard	1.28	0.55	2.33	0.068	4.54	0.56	8.10	<0.001[a]

[a]Statistical significance with an alpha value of $p = 0.05$.

Infants are likely to be attracted to positive affect as expressed in IDS, such as joy (Singh *et al.*, 2002). There are studies suggesting that, during mother-infant interaction, positive affect is not only expressed through the mothers' voices but also their facial expressions (Stern, 1974; Chong *et al.*, 2003). It is claimed that two of the predominant facial expressions in IDS to convey positive affect are an open "surprised" mouth and a joyful smile, often with a slightly opened mouth (Benders, 2013). A more open month in IDS could result in jaw-lowering, thus leading to the increase in F1, especially for the low and back vowels (in our results, the F1 of /i/ in IDS is also affected by the jaw-lowering, although the effect is a trend: $p = 0.077$). The other facial expression is "spread lips" in IDS, which could lead to F2 increase, especially for /a/ and /u/, since the lips are already spread when producing /i/, at least in the Northern Mandarin in this study. The phonetic outcomes of these articulatory modifications, such as the increased F1 and F2 values for the low and back vowels in the current study, are also consistent with the acoustic features of happy speech (Kienast and Sendlmeier, 2000).

In contrast, when talking in a noisy environment, speakers usually speak more loudly to maintain self-monitoring ability and to transmit the speech signals more effectively (Zhao and Jurafsky, 2009). To increase loudness, speakers adjust their articulatory gestures with acoustic consequences. It is suggested that one of the most commonly used articulatory adjustments made to speak more loudly in noisy environments is to increase the degree of jaw-lowering (Schulman, 1989), which can lead to an overall increase in F1. Godoy *et al.* (2013) also suggested that, to increase loudness, another possible articulatory adjustment is to contract the vocal muscles and to move the vocal tract resonances closer together, which results in decreased separation between formant frequencies, i.e., a decrease in F1-F2 and F2-F3. The phonetic outcome of such articulatory adjustments is largely consistent with our results (i.e., an increased F1 and a decreased F1-F2, which was a result of the increased F1 and the unchanged F2), which implies that vowel modifications in Lombard speech are presumably the result of articulatory adjustments made to increase intensity in a noisy environment.

TABLE XII. Tukey-HSD pairwise comparisons on different levels of Register (IDS, Lombard speech, and ADS) in the two-way interaction between Register and Position (utterance-medial and utterance-final) on tonal duration. The estimated difference between two levels of Register, standard error, t values, and HSD-adjusted *p* values are provided.

Register difference	Utterance-medial				Utterance-final			
	Estimate	SE	t	*p*	Estimate	SE	t	*p*
IDS-ADS	0.01	0.01	0.66	0.786	0.05	0.01	4.65	<0.001[a]
Lombard-ADS	0.01	0.01	1.14	0.496	0.03	0.01	3.01	0.010[a]
IDS-Lombard	−0.01	0.01	−0.27	0.961	0.02	0.01	1.73	0.209

[a]Statistical significance with an alpha value of $p = 0.05$.

With respect to our results in tone hyperarticulation, it is surprising that the tone space of IDS only expands in utterance-final position. Note that previous Taiwan Mandarin and Cantonese IDS studies did not take the positional effect into consideration when reporting enhancement of tonal contrasts. There are two possible reasons for the position effect in tone hyperarticulation. First, as reported in previous studies, pitch information is one of the most important carriers of affect (Pell *et al.*, 2009). In a tonal language, such as Mandarin, when speakers express "happy" affect, overall pitch is increased, and the slope of the pitch contour of the final syllable of each prosodic word is exaggerated, especially for syllables in utterance-final position (Wang *et al.*, 2005). These acoustic features of Mandarin happy speech are highly consistent with our results. The tone hyperarticulation and vowel hyperarticulation findings converge, suggesting that speech modifications in IDS are mainly driven by the mothers' communicative goal to convey positive affect.

Second, tonal contrasts may not have been enhanced in utterance-medial position because of the tone context. It is well-known that Mandarin tone realization is heavily influenced by adjacent tones, as reported in the tone carry-over effect and the tone anticipatory effect (Xu, 1997). The tone contours of T2 and T4 are prone to modification when they occur in a "conflicting tone context" (Xu, 1994), leading to a possible "neutralization" of the tone contour. For instance, the conflicting context for T2 (rising tone) refers to a situation where the T2 syllable is preceded by a syllable with a high pitch offset and followed by a syllable with a low pitch onset; for T4 (falling tone), this refers to a context where a T4 syllable is preceded by a syllable with a low pitch offset and followed by a syllable with a high pitch onset. In the present study, the tone preceding the target tones was controlled (i.e., T1). Since we could not control for the tone following the targets in utterance-medial position, this led to conflicting tone contexts in some cases, such as when a target T2 is followed by a T2. This will minimize the tonal contrast between the target T2 and T1. As a result, the tone space was less distinct between the registers when the target tones occurred in the utterance-medial position.

It is worth noting that Trainor and Desjardins (2002) observed that, in English IDS, the exaggerated pitch contour could aid infants' acquisition of vowel categories, while the increased pitch height could serve affective or attentional functions. Therefore, it is also possible that the expanded tone space in the present study is related to mothers' didactic intention to help infants in learning vowels.

It is also surprising to find no tone space expansion in Lombard speech. It should be noted that the task of Lombard speech in our experiment is different from previous Lombard speech studies of Cantonese (Zhao and Jurafsky, 2009) and Thai (Kasisopa *et al.*, 2014), which only adopted a reading task with one speaker. The present study, in contrast, used a communicative task where two speakers are engaged in a conversation. The task difference may explain the inconsistent results between the present study and previous studies such as Kasisopa *et al.* (2014), which observed tone space expansion in Lombard speech. Perhaps, in the noisy environment, speakers increase their effort, and this manifests in terms of increasing the overall pitch of tones, rather than exaggerating the tonal contrast, to convey tone units more effectively. Alternatively, it is also possible that the exaggerated pitch contour found in IDS in the present study was simply a result of mothers expressing positive affect, which is expected in IDS but not Lombard speech.

The increased overall pitch in IDS and Lombard speech found in our study is consistent with other findings in the literature (e.g., Fernald *et al.*, 1989; Junqua, 1996). It is generally agreed that the increased pitch height in IDS is associated with the communication of positive affect to infants (see Xu *et al.*, 2013, for a review), while it is claimed the increased overall pitch in Lombard speech is merely an irrelevant by-product of an increased vocal effort that serves to increase the relative intensities of the higher-frequency components in the speech spectrum (Uchanski, 2005).

Although there was no difference in the overall tone-bearing word duration between the two registers, the tone-bearing word duration was longer in IDS and Lombard speech than in ADS in utterance-final position. This duration pattern resembles the well-documented final-lengthening phenomenon found in IDS (Morgan *et al.*, 1987). Perhaps mothers exaggerate word duration in utterance-final position in IDS to facilitate word segmentation for their infants, in line with the prosodic bootstrapping hypothesis (Morgan *et al.*, 1987). Similarly, the increase in utterance-final word duration in Lombard speech could also aid listeners in identifying word units in a noisy environment and thus improve communicative effectiveness.

However, it should be acknowledged that there are several potential limitations of the present study. First, the use of one token for each target vowel and target tone could potentially limit the extent of generalizability. We did our best to identify words in Mandarin Chinese that could meet the criterion of our stimuli: picturable disyllabic triads that only differed in vowels or tones. Second, the order of the three tasks was fixed: IDS, Lombard speech, and ADS. Keeping the ADS last may potentially increase the contrast between ADS and other registers due to previous mention or fatigue effects. However, most of the present results are consistent with previous findings regarding vowel and tone hyperarticulation. In addition, the focus of the present study was to examine the dimensions of vowel and tone hyperarticulation across registers rather than the degree of hyperarticulation.

V. CONCLUSION

Speech units such as vowels and tones are reported to be hyperarticulated in both IDS and Lombard speech, however, it has been unclear if the same acoustic parameters might be hyperarticulated in a similar manner between these two registers. The results of this study show that vowels and tones were hyperarticulated in different ways in Northern Mandarin between IDS and Lombard speech. Vowels and tones in IDS are modified to express positive affect, whereas vowel and tone modification in Lombard speech is mainly aimed at transmitting speech signals more effectively in noisy environments. According to Hazan and Baker (2011), speech production is listener-focused, and talkers modulate their speech according to their interlocutors' needs. Our results support this view and also reveal that speakers modify their speech to cater for different communicative needs of the addressee, thus providing a more in-depth understanding of the nature of these speech registers and speech enhancement more generally.

ACKNOWLEDGMENTS

We thank Titia Benders, Carmen Kung, the Child Language Lab, the Phonetics Lab, and the ARC Centre of Excellence in Cognition and its Disorders at Macquarie University for their comments, feedback and supports. We thank Xin Cheng for helping with the reliability check. We also acknowledge the comments and suggestions from the editor and anonymous reviewers which significantly improved the manuscript. This research was supported, in part, by a Macquarie University iMQRES scholarship to the first author, and the following grants: ARC CE110001021, ARC FL130100014. The equipment was funded by MQSIS 9201501719.

[1]Six target words were selected and elicited in the present study, and additional eight T3, neutral tone and tone sandhi words were used for another study. All these stimuli were elicited in the same experiment.
[2]The R code of this model is: Vowel space area ~ Register + (1 | Subject).
[3]The R code for these models is: F1/F2 ~ Register * Vowel * Vowel Duration + (1 + Vowel | Subject).
[4]The main effect of the covariate Vowel Duration and its interactions with other factors were not presented since they were not the main interest of the current study.
[5]The R code for this model is: Tone space area ~ Register * Position + (1 + Register + Position | Subject).
[6]The R code for this model is: Pitch Height ~ Register * Position * Tone + (1 + Register + Tone | Subject).
[7]The R code for this model is: Tonal duration ~ Register * Position * Tone + (1 + Register + Position | Subject).

Barry, J. G., and Blamey, P. J. (**2004**). "The acoustic analysis of tone differentiation as a means for assessing tone production in speakers of Cantonese," J. Acoust. Soc. Am. **116**, 1739–1748.
Bates, D., Mächler, M., Bolker, B., and Walker, S. (**2015**). "Fitting linear mixed-effects models using lme4," J. Stat. Softw. **67**, 1–48.
Benders, T. (**2013**). "Mommy is only happy! Dutch mothers realisation of speech sounds in infant-directed speech expresses emotion, not didactic intent," Infant Behav. Dev. **36**, 847–862.
Boersma, P., and Weenink, D. (**2016**). "Praat: Doing phonetics by computer" [computer program], version 6.0.19, http://www.praat.org/ (Last viewed 6/13/2016).
Bond, Z. S., Moore, T. J., and Gable, B. (**1989**). "Acoustic-phonetic characteristics of speech produced in noise and while wearing an oxygen mask," J. Acoust. Soc. Am. **85**, 907–912.
Burg, J. P. (**1975**). "Maximum entropy spectrum analysis," Ph.D. thesis, Stanford University, Stanford, CA.
Burnham, D., Kitamura, C., and Vollmer-Conna, U. (**2002**). "What's new pussycat? On talking to babies and animals," Science **296**, 1435.

Cheng, M.-C., and Chang, K.-C. (**2014**). "Tones in Hakka infant-directed speech: An acoustic perspective," Lang Ling. **15**, 341–390.
Chong, S. C. F., Werker, J. F., Russell, J. A., and Carroll, J. M. (**2003**). "Three facial expressions mothers direct to their infants," Infant Child Dev. **12**, 211–232.
Cooke, M., and Lu, Y. (**2010**). "Spectral and temporal changes to speech produced in the presence of energetic and informational maskers," J. Acoust. Soc. Am. **128**, 2059–2069.
Davis, C., and Kim, J. (**2010**). "Is speech produced in noise more distinct and/or consistent?," in *Proceedings of the 14th Australasian International Conference on Speech Science and Technology*, Sydney, Australia, pp. 46–49.
Deng, D., and Shi, F. (**2006**). "The comparison on phonetic system between Putonghua and Taiwan Mandarin," in *Mountain Lofty, River Long: Festschrift in Honor of Professor Pang-hsin Ting on His Seventieth Birthday*, edited by D. A. He (Academia Sinica, Taipei, Taiwan), pp. 371–393 (in Chinese).
Englund, K., and Behne, D. M. (**2006**). "Changes in infant directed speech in the first six months," Infant Child Dev. **15**, 139–160.
Fernald, A., and Simon, T. (**1984**). "Expanded intonation contours in mothers' speech to newborns," Dev. Psychol. **20**, 104–113.
Fernald, A., Taeschner, T., Dunn, J., Papousek, M., de Boysson-Bardies, B., and Fukui, I. (**1989**). "A cross-language study of prosodic modifications in mothers' and fathers' speech to preverbal infants," J. Child Lang. **16**, 477–501.
Gay, T. (**1978**). "Effect of speaking rate on vowel formant movements," J. Acoust. Soc. Am. **63**, 223–230.
Godoy, E., Mayo, C., and Stylianou, Y. (**2013**). "Linking loudness increases in normal and lombard speech to decreasing vowel formant separation," in *Proceedings of INTERSPEECH*, Lyon, France, pp. 133–137.
Godoy, E., Koutsogiannaki, M., and Stylianou, Y. (**2014**). "Approaching speech intelligibility enhancement with inspiration from Lombard and Clear speaking styles," Comput. Speech Lang. **28**, 629–647.
Hazan, V., and Baker, R. (**2011**). "Acoustic-phonetic characteristics of speech produced with communicative intent to counter adverse listening conditions," J. Acoust. Soc. Am. **130**, 2139–2152.
Junqua, J.-C. (**1996**). "The influence of acoustics on speech production: A noise-induced stress phenomenon known as the Lombard reflex," Speech Commun. **20**, 13–22.
Kasisopa, B., Attina, V., and Burnham, D. (**2014**). "The Lombard effect with Thai lexical tones: An acoustic analysis of articulatory modifications in noise," in *Proceedings of INTERSPEECH*, Singapore, pp. 1717–1721.
Kienast, M., and Sendlmeier, W. F. (**2000**). "Acoustical analysis of spectral and temporal changes in emotional speech," in *Proceedings of the ISCA Workshop on Emotion and Speech*, Belfast, Northern Ireland, pp. 92–94.
Kitamura, C., Thanavishuth, C., Burnham, D., and 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 Behav. Dev. **24**, 372–392.
Kuhl, P. K., Andruski, J. E., Chistovich, I. A., Chistovich, L. A., Kozhevnikova, E. V., Ryskina, V. L., Stolyarova, E. I., Sundberg, U., and Lacerda, F. (**1997**). "Cross-language analysis of phonetic units in language addressed to infants," Science **277**, 684–686.
Kuznetsova, A. (**2013**). "Lmertest: Tests for random and fixed effects for linear mixed effect models (lmer objects of lme4 package)," R package version 2.0-29, https://cran.r-project.org/web/packages/lmerTest/lmerTest.pdf (Last viewed 5/4/2016), pp. 1–17.
Li, F. K. (**1973**). "Languages and dialects of China," J. Chinese Ling. **1**, 1–13, available at http://www.jstor.org/stable/23749774.
Liu, H.-M., Kuhl, P. K., and Tsao, F.-M. (**2003**). "An association between mothers' speech clarity and infants' speech discrimination skills," Dev. Sci. **6**, F1–F10.
Liu, H.-M., Tsao, F.-M., and Kuhl, P. K. (**2007**). "Acoustic analysis of lexical tone in Mandarin infant-directed speech," Dev. Psychol. **43**, 912–917.
Lombard, E. (**1911**). "Le signe de l'elevation de la voix" ("The sign of the rise in the voice"), Ann. Malad. Oreille Larynx Nez Pharynx **37**, 101–119.
Majorano, M., Rainieri, C., and Corsano, P. (**2013**). "Parents' child-directed communication and child language development: A longitudinal study with Italian toddlers", J. Child Lang. **40**, 836–859.
Morgan, J. L., Meier, R. P., and Newport, E. L. (**1987**). "Structural packaging in the input to language learning: Contributions of prosodic and morphological marking of phrases to the acquisition of language," Cogn. Psychol. **19**, 498–550.
Morris, R. J., and Brown, W. S., Jr. (**1994**). "Age-related differences in speech intensity among adult females," Folia Phon. Logop. **46**, 64–69.
Pell, M. D., Paulmann, S., Dara, C., Alasseri, A., and Kotz, S. A. (**2009**). "Factors in the recognition of vocally expressed emotions: A comparison of four languages," J. Phon. **37**, 417–435.
R Core Team (**2016**). "R: A language and environment for statistical computing" [computer program], version 3.3.1, https://www.R-project.org/ (Last viewed 6/21/2016).
Schulman, R. (**1989**). "Articulatory dynamics of loud and normal speech," J. Acoust. Soc. Am. **85**, 295–312.
Simpson, S., and Cooke, M. P. (**2005**). "Consonant identification in N-talker babble is a nonmonotonic function of N," J. Acoust. Soc. Am. **118**, 2775–2778.
Singh, L., Morgan, J. L., and Best, C. T. (**2002**). "Infants' listening preferences: Baby talk or happy talk?," Infancy **3**, 365–394.
Smiljanić, R., and Bradlow, A. (**2005**). "Production and perception of clear speech in Croatian and English," J. Acoust. Soc. Am. **116**, 2627–2627.
Song, J. Y., Demuth, K., and Morgan, J. (**2010**). "Effects of the acoustic properties of infant-directed speech on infant word recognition," J. Acoust. Soc. Am. **128**, 389–400.
Stern, D. (**1974**). "Mother and infant at play: The dyadic interaction involving facial, vocal and gaze behaviors," in *The Effect of the Infant on its Caregiver*, edited by M. Lewis, and L. Rosenblum (Wiley, New York), pp. 187–232.
Trainor, L. J., and Desjardins, R. N. (**2002**). "Pitch characteristics of infant-directed speech affect infants' ability to discriminate vowels," Psychon. Bull. Rev. **9**, 335–340.
Uchanski, R. M. (**2005**). "Clear speech," in *Handbook of Speech Perception*, edited by D. B. Pisoni and R. E. Remez (Blackwell Publishers, Malden, MA), pp. 207–235.
Uther, M., Knoll, M. A., and Burnham, D. (**2007**). "Do you speak E-NG-LI-SH? Similarities and differences in speech to foreigners and infants," Speech Commun. **49**, 1–7.
Wang, H., Li, A., and Fang, Q. (**2005**). "F0 contour of prosodic word in happy speech of Mandarin," in *Proceedings of International Conference on Affective Computing and Intelligent Interaction*, edited by J. Tao, T. Tan, and R. W. Picard (Springer, Berlin), pp. 433–440.
Wassink, A. B., Wright, R. A., and Franklin, A. D. (**2007**). "Intraspeaker variability in vowel production: An investigation of motherese, hyperspeech, and Lombard speech in Jamaican speakers," J. Phon. **35**, 363–379.
Xu, N., Burnham, D., Kitamura, C., and Vollmer-Conna, U. (**2013**). "Vowel hyperarticulation in parrot-, dog-and infant-directed speech," Anthrozoös **26**, 373–380.
Xu, Y. (**1994**). "Production and perception of coarticulated tones," J. Acoust. Soc. Am. **95**, 2240–2253.
Xu, Y. (**1997**). "Contextual tonal variations in Mandarin," J. Phon. **25**, 61–83.
Xu Rattanasone, N., Burnham, D., and Reilly, R. G. (**2013**). "Tone and vowel enhancement in Cantonese infant-directed speech at 3, 6, 9, and 12 months of age," J. Phon. **41**, 332–343.
Yip, M. (**2002**). *Tone* (Cambridge University Press, Cambridge), p. 1.
Zhao, Y., and Jurafsky, D. (**2009**). "The effect of lexical frequency and Lombard reflex on tone hyperarticulation," J. Phon. **37**, 231–247.
Zwicker, E., and Terhardt, E. (**1980**). "Analytical expressions for critical band rate and critical bandwidths as a function of frequency," J. Acoust. Soc. Am. **68**, 1523–1525.