Acoustic realization of Mandarin neutral tone and tone sandhi in infant-directed speech and Lombard speech

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

Citation: The Journal of the Acoustical Society of America **142**, 2823 (2017); doi: 10.1121/1.5008372
View online: https://doi.org/10.1121/1.5008372
View Table of Contents: http://asa.scitation.org/toc/jas/142/5
Published by the Acoustical Society of America

Acoustic realization of Mandarin neutral tone and tone sandhi in 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, 16 University Avenue, Australian Hearing Hub, Balaclava Road, North Ryde, New South Wales 2109, Sydney, Australia

(Received 23 February 2017; revised 29 September 2017; accepted 4 October 2017; published online 10 November 2017)

Mandarin lexical tones are modified in both infant-directed speech (IDS) and Lombard speech, resulting in tone hyperarticulation. However, it is unclear if these registers also alter contextual tones (neutral tone and tone sandhi) and if such phonetic modification might affect acquisition of these tones. This study therefore examined how neutral tone and tone sandhi are realized in IDS, and how their acoustic manifestations compare with those in Lombard speech, where the communicative needs of listeners differ. Neutral tone and tone sandhi productions were elicited from 15 Mandarin-speaking mothers during (1) interactions with their 12-month-old infants (IDS), (2) in conversation with a Mandarin-speaking adult in a noisy environment (Lombard speech), and (3) in conversation with a Mandarin-speaking adult in a quiet environment (adult-directed speech). The results showed that, although both contextual tones were modified in IDS and Lombard speech, their key tone features were maintained. In addition, IDS and Lombard speech modified these tones differently: IDS increased pitch height and modified pitch contour, while Lombard speech increased pitch height only. The realization of neutral tone and tone sandhi across registers is discussed with reference to listeners' different communicative needs.
 https://doi.org/10.1121/1.5008372

[TCB] Pages: 2823–2835

I. INTRODUCTION

Speakers modify their speech to cater to different communicative needs of the addressee, which result in various speech styles. For example, when talking to an infant, speech is usually hyperarticulated, resulting in a unique speech register known as infant-directed speech or IDS (Kuhl *et al.*, 1997; Burnham *et al.*, 2002). Speech hyperarticulation in IDS typically includes a series of suprasegmental and segmental modifications. Relative to adult-directed speech (ADS), IDS exhibits a higher pitch (Fernald *et al.*, 1989; Burnham *et al.*, 2002; Fernald and Simon, 1984), a larger pitch variability (Fernald *et al.*, 1989), a slower speaking rate (Fernald and Simon, 1984), and enhanced vowel contrasts (Kuhl *et al.*, 1997; Burnham *et al.*, 2002; Liu *et al.*, 2003). It is generally agreed that these modifications may help attract and maintain infants' attention, express positive affect, and possibly facilitate infants' language learning (Cooper *et al.*, 1997; Grieser and Kuhl, 1988; Singh *et al.*, 2002).

Another register exhibiting similar speech characteristics is known as Lombard speech, a speech style that people use when talking in a noisy environment (Lombard, 1911). Relative to speech produced in quiet conditions, Lombard speech typically exhibits a higher pitch, a greater intensity, a longer vowel duration, and increased formant frequencies (Summers *et al.*, 1988; Junqua, 1996). It is claimed that Lombard speech is used to maintain self-monitoring ability and transmission of speech signals more effectively in a noisy environment (Zhao and Jurafsky, 2009).

In tonal languages such as Mandarin Chinese, IDS and Lombard speech also modify lexical tones, resulting in tone hyperarticulation (Liu *et al.*, 2007; Tang *et al.*, 2017). Tone hyperarticulation can enhance the acoustic features of lexical tones and thus benefit listeners. However, in connected speech, there are many tone variants such as tone reduction or tone change. Two well-reported types of contextual variation are neutral tone and tone sandhi (Yip, 2002). Although a large body of work has explored the realization of neutral tone and tone sandhi under normal speech conditions, no study has yet directly examined how these contextual tones are realized in IDS and Lombard speech. It is therefore not known whether these contextual tones will also undergo modification in IDS and Lombard speech. It has been claimed that IDS can benefit young children in lexical tone acquisition by means of tone hyperarticulation (Liu *et al.*, 2007; Xu Rattanasone *et al.*, 2013). Indeed, some acquisition evidence has shown that Mandarin-learning children acquire lexical tones quite early, i.e., before age 3 (Li and Thompson, 1977; Zhu and Dodd, 2000). However, relative to lexical tones, contextual tones are reported to be acquired by children much later, i.e., only by the age of 6 (Wang, 2011), but the reason remains unclear. Later acquisition of tone sandhi processes compared to lexical tones is also reported in Bantu languages (e.g., Demuth, 1993), suggesting that learning tone sandhi processes may take longer to master in many tone languages. It is also possible that these contextual tones might not undergo phonetic exaggeration in IDS to the same extent that lexical tones do, making them harder to learn. Thus, the primary goal of the present study was to examine the acoustic realization of these contextual tones in IDS to

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

shed light on the possible reasons for the late acquisition of contextual tones.

Although lexical tones have been found to be hyperarticulated in both IDS and Lombard speech registers (e.g., Tang *et al*., 2017), the phonetic dimensions of tone hyperarticulation manifest differently across the two registers. For example, although pitch height is increased in both registers, the pitch contour is modified only in IDS (Liu *et al*., 2007; Zhao and Jurafsky, 2009; Tang *et al*., 2017). This subtle difference appears to be driven by the different communicative goals of these two registers. The increased pitch in Lombard speech, for example, might be a by-product of the increased vocal effort needed in order to speak more loudly, as proposed by Uchanski (2005). Similar proposals are suggested by Schulman (1989), who found that vocal effort increased dramatically with loud speech, and, as a consequence of an increase of vocal effort, the pitch was also increased. In contrast, acoustic modifications of pitch height and pitch contour in IDS have been associated with attentional, affective, and didactic functions (Fernald and Kuhl, 1987; Kitamura and Burnham, 2003; Trainor and Desjardins, 2002). For example, Fernald and Kuhl (1987) found that infants prefer to listen to IDS when it differs from ADS in terms of pitch (both pitch height and pitch range), but not when it differed in amplitude or duration. Similarly, when adults were asked to rate their impression of IDS, Kitamura and Burnham (2003) found that pitch height was associated with affective and attentional scales, but pitch range correlated only with attentional scales. However, Trainor and Desjardins (2002) provided some evidence suggesting that exaggerated pitch contours might help with language learning. In their study, 6–7-month-old infants were reported to better discriminate the vowels /i/ vs /ɪ/ with exaggerated pitch contours compared to monotone speech. However, since monotone prosody is quite unnatural, the observed effect of the exaggerated pitch contour might also be attributed to attracting the infants' attention.

Contextual tones might also undergo pitch modification in both IDS and Lombard speech in order to achieve certain communicative aims. If so, the modifications might be different between the two registers, i.e., pitch height might be increased in both registers, while the pitch range might be expanded in IDS only. Therefore, the second goal of the present study was to compare directly the role of register on the realization of contextual tones, to better understand the nature of tonal enhancement across these two registers.

A. Neutral tone

In Mandarin, neutral tone is called the "fifth tone" and often referred to as "toneless" or T0. Neutral tone has a short duration, and usually occurs in a final unstressed syllable of a disyllabic word. Depending on the preceding lexical tone, the pitch contour of the neutral tone varies (Cao, 1992). For example, neutral tone exhibits a falling pitch when preceded by syllables containing T1 (tone 1, the level tone), T2 (tone 2, the rising tone), and T4 (tone 4, the falling tone); however, its pitch contour rises when it is preceded by a syllable with T3 (tone 3, the dipping tone) (Yip, 2002). Neutral tones are thus often found in (1) certain monosyllabic particles (i.e., the diminutive particle /tsi0/, the nominalizer or possessive particle /tɣ0/, and the classifier particle /kɣ0/); (2) the second syllable of a disyllabic reduplicated word, such as /ma1 ma0/ (*mother*); or (3) the second syllable of some disyllabic words, such as /t^{h}ou2 fa0/ (*hair*) (Li and Thompson, 1977; Zhu and Dodd, 2000).

It has been reported in previous studies that children do not master neutral tone productions until around 4–6. The most common error that children make is to replace a neutral tone with a full lexical tone, i.e., using /tsi3/ for /tsi0/ (Li and Thompson, 1977; Zhu and Dodd, 2000). According to Zhu and Dodd (2000), these tone substitution errors are often found when using the diminutive particle /tsi0/ and in the final syllable of disyllabic words such as /t^{h}ou2 fa0/ (*hair*).

B. Tone sandhi

Tone sandhi (also known as tone 3 sandhi) refers to a phonological process in which a lexically-specified tone 3 syllable (T3) is realized as a rising tone (T2: full sandhi) when followed by another T3 syllable or realized as a low-falling tone (half sandhi) when followed by different tones: T1, T2, or T4 (Yip, 2002). The (full) sandhi rule can also be applied recursively to multiple T3 words, and it depends on the prosodic structure of the phrase (Shih, 1997). Consider the underlying forms /ɕiau3 ma3 ji3/ (*small ant*) and /ma3 ji3 tɕiau3/ (*ant's feet*). The former has a right-branching prosodic structure [σ [$\sigma\sigma$]] and the latter has a left-branching prosodic structure [[$\sigma\sigma$] σ]. These structures trigger rightward parsing [see (1)] and leftward parsing [see (2)], respectively, leading to two different surface realizations of the underlying tone sequence "/ma3 ji3/" → T2T3 vs T2T2, respectively (Shih, 1997).

(1) A right-branching trisyllabic noun phrase [σ [$\sigma\sigma$]]

[ɕiau [ma ji]]	(*Small ant*)
T3 T3 T3	Underlying tone
T3 (T2 T3)	Tone sandhi rule applies within prosodic domain
(T3 T2 T3)	Incorporation, no additional tone sandhi rule applied; surface tone

(2) A left-branching trisyllabic noun phrase [[$\sigma\sigma$] σ]

[[ma ji] tɕiau]	(*Ant's feet*)
T3 T3 T3	Underlying tone
(T2 T3) T3	Tone sandhi rule applies within prosodic domain
(T2 T2 T3)	Incorporation, tone sandhi rule applied again; surface tone

The tone sandhi rule is challenging for children to acquire, with children at age 6 still not achieving productive knowledge of adult-like tone sandhi in production (Wang, 2011). It requires children to have knowledge of both the tonal and prosodic context in which the rule applies (Wang, 2011).

As overall prosodic features are modified in IDS and Lombard speech, these modifications, especially the slow speaking rate, might have additional repercussions for the acoustic representation of both neutral tone and tone sandhi as input to children. For example, neutral tone, realized on an unstressed syllable, is mainly characterized by short duration. Thus the need to preserve short duration for neutral

tone will go against the need to speak slowly (with increased duration) in both registers. The conflicting demands between speech register and the acoustic characteristics of neutral tone might thus result in production of ambiguous tone. If so, neutral tone might be realized more like a full lexical tone, i.e., /tsi0/ being realized as /tsi3/, similar to the type of error that children tend to make in producing neutral tone syllables (Zhu and Dodd, 2000). A recent study found greater utterance-final lengthening in IDS and Lombard speech relative to ADS (Tang *et al.*, 2017). Since neutral tone syllables mainly occur in the final syllable of a word, IDS and Lombard speech could also potentially lengthen neutral tone productions, especially in the utterance-final position, resulting in tone ambiguity.

Similar to neutral tone, tone sandhi rule application may also be affected by the slow speaking rate of IDS and Lombard speech, especially for some polysyllabic noun phrases. For example, as mentioned above, in a normal speaking rate, the surface tones of a T3 right-branching phrase [σ [$\sigma\sigma$]] vs a left-branching phrase [[$\sigma\sigma$] σ] would be T3T2T3 and T2T2T3, respectively. However, in a slow speaking rate, a prosodic boundary could be inserted between the disyllabic word [$\sigma\sigma$] and the monosyllabic word [σ] in both cases. Although an inserted boundary would not affect tone sandhi application of the right-branching phrase [see (3)], it would influence tone sandhi application in the left-branching case, resulting in reduced tone sandhi application [see (4)], and the potential for different meanings. In other words, slow speech in IDS and Lombard speech could undermine the didactic aim of IDS and the clarification goal of Lombard speech. Furthermore, it has recently been found that different speech registers can affect both the pitch height and pitch contour of lexical tones differently (Tang *et al.*, 2017). This raises the possibility that certain phonetic aspects of contextual tones (neutral tone and tone sandhi) might be differentially modified to suit the didactic vs clarification aims of IDS vs Lombard speech, respectively, resulting in different types of modifications for these two registers.

(3) [σ [$\sigma\sigma$]] right-branching phrase

T3T3T3	Underlying tone
T3#T3T3	Boundary inserted
T3#T2T3	Tone sandhi applies; surface tone

(4) [[$\sigma\sigma$] σ] left-branching phrase

T3T3T3	Underlying tone
T3T3#T3	Boundary inserted
T2T3#T3	Tone sandhi applies; surface tone

It is also possible that the resulting speech modifications on the neutral tone and the tone sandhi rule in IDS and Lombard speech may distort their acoustic/linguistic characteristics and thus be a potential hindrance for the listeners. If so, this might lead to insufficient input for children, and therefore help explain the reported later acquisition of these contextually determined tones. It may also decrease communicative effectiveness by making it more difficult for listeners to perceive these tones in noisy environments. These outcomes would then go against the communicative aims of these registers, i.e., the language-teaching aim of IDS and the clarification aim of Lombard speech.

The aim of the present study was therefore to investigate the phonetic modification of Mandarin contextual tones—neutral tone and tone sandhi—in IDS and Lombard speech. In particular, we asked whether IDS and Lombard speech would modify neutral tone and tone sandhi productions, and if so, whether the two speech registers would differ in the acoustic implementation of these changes. To explore these issues, neutral tone and tone sandhi productions were elicited across three conditions: IDS, Lombard speech (while listening to babble noise), and ADS (in quiet, as a control). Neutral tone vs full tone minimal pairs were employed to investigate the acoustic realization of neutral tone using the full lexical tone as a baseline control; tone sandhi productions were elicited in three different contexts (a disyllabic word, a right-branching disyllabic noun phrase, and a left-branching disyllabic noun phrase) to examine the tone sandhi realization across different prosodic contexts. Neutral tone and tone sandhi productions were also compared between registers to explore the register effect on these tones. We predicted that:

Hypothesis 1 (H1): Neutral tone syllables would be modified in terms of duration and pitch and produced as full tone in IDS and Lombard speech.
Hypothesis 2 (H2): Tone sandhi syllables would be modified in terms of duration and pitch and produced with the underlying tone (T3) in IDS and Lombard speech.
Hypothesis 3 (H3): Relative to ADS, IDS would increase the pitch height and modify the pitch contour of neutral tone and tone sandhi syllables, while Lombard speech would increase the pitch height only, as previously observed for lexical tones in IDS (Liu *et al.*, 2007) and Lombard speech (Zhao and Jurafsky, 2009).

II. METHOD

A. Participants

Fifteen mothers and their 12-month-old infants [Mean = 12 months, standard deviation (SD) = 0.99] were recruited from Sydney, Australia. Mothers of 12-month-olds were selected to compare our results to previous IDS studies of (Taiwan and Northern) Mandarin lexical tones, which used similar age groups (Liu *et al.*, 2007; Tang *et al.*, 2017).

All mothers were raised in Northern-Mandarin speaking families (e.g., Beijing, Hebei province, and Northeastern China) before 18 yrs of age, had been in Australia from 1 to 8 yrs (Mean = 5 yrs, SD = 3.27), and spoke both Mandarin and English. Their age ranged from 18 to 34 yrs (Mean = 24 yrs, SD = 4.93). All were the main caregivers of their infants, speaking only Mandarin to their infants at home.

B. Stimuli

Seven Mandarin nouns were chosen as target stimuli, all illustrated with toys. For the neutral tone conditions, two neutral tone vs full tone minimal pairs were selected

(see Table I). All were disyllabic words where the second syllable was the target syllable, which carried either the neutral tone or the full tone counterpart. The neutral tone-bearing unit was the Chinese word /t^hou/ (*head*) in pair 1 and a diminutive particle /tsi/ (*piece*) in pair 2. These two types of neutral tone were selected because they have been reported to be acquired last by Mandarin-learning children, compared to the other neutral tone semantic categories (e.g., the nominalizer or possessive particle /tɣ0/ and the classifier particle /kɣ0/). The neutral tone syllables in these two types of words are easily confusable with their full tone counterparts in children's productions, i.e., /ʂɣ2 t^hou0/ (*tongue*) confused with / ʂɣ2 t^hou2/ (*snake's head*) and /tɕ^{h}i2 tsi0/ (*flag*) confused with /tɕ^{h}i2 tsi3/ (*chess piece*) (Li and Thompson, 1977; Zhu and Dodd, 2000). Using minimal pairs thus allowed for better control in comparing neutral tone with full tone productions (as a baseline control).

For the tone sandhi comparisons, a disyllabic T3T3 word "/ma3 ji3/" (*ant*) and two trisyllabic T3T3T3 noun phrases were selected (see Table II). These two noun phrases shared the same syllables "/ma3 ji3/" (*ant*) but differed in their prosodic structures, i.e., right-branching /ɕiau3 **ma3 ji3**/ (*little ant*) and left-branching **/ma3 ji3** tɕiau3/ (*ant's feet*). The syllable /ma3/ and syllable /ji3/ were treated as the target syllables. If the tone sandhi rule is correctly applied, the expected surface tone of /ma3 ji3/ would be T2T3 in the disyllabic word and the right-branching noun phrase, and T2T2 in the left-branching noun phrase (Shih, 1997).

C. Procedure

Every mother-infant dyad was tested in a sound-attenuated room. During the familiarization phase, the mother was instructed to wear a head-mounted microphone [AKG-C520, Acoustic and Cinema Equipment (AKG), Vienna, Austria] which was connected to a solid-state recorder (Marantz PMD661MKII, Kanagawa, Japan). The recorder was placed in a shoulder bag to allow for the mother's free movement in the test phrase. The recording was made at a sampling rate of 44.1 kHz and a 16-bit quantization. In the test phase, three speech production tasks were conducted to elicit the three registers: (1) IDS, (2) Lombard speech, and (3) ADS. The order of these tasks was consistent across participants: IDS first, followed by Lombard speech, and then ADS. This order ensured that IDS data were collected before the infant became fussy.

In the IDS task, the mother and her infant were engaged in a play session. The seven target stimuli were labelled on the corresponding toys using Chinese characters, and the toys were randomly allocated to three cloth bags, minimizing noise. These bags were adopted to counterbalance the order of presentation of the toys/test items across participants. The mother was provided with one bag at a time and asked to play with the toys while interacting with her infant as they normally would at home. From the control room, a Mandarin-speaking experimenter (P.T.) kept count of the tokens that the mother produced (at least ten repetitions) for each toy in each bag.

TABLE I. Stimuli for eliciting neutral tone productions, including two neutral vs full tone minimal pairs.

Pair	Type	Stimuli	Target syllable	Target tone
Pair 1	Neutral tone	/ʂɣ2 t^hou0/ (*tongue*)	/t^hou0/	T0
	Full tone	/ʂɣ2 t^hou2/ (*snake's head*)	/t^hou2/	T2
Pair 2	Neutral tone	/tɕ^{h}i2 tsi0/ (*flag*)	/tsi0/	T0
	Full tone	/tɕ^{h}i2 tsi3/ (*chess piece*)	/tsi3/	T3

In the Lombard speech task, the mother was asked to converse with the experimenter about her experience in the play session with her infant. The mother was instructed to use the same written labels to refer to the toys. The mother and the experimenter both wore open-ear headphones (AKG-K612 PRO), through which a 70-dBA Chinese 8-talker babble noise was played during the Lombard speech sessions. Open-ear headphones were adopted since they allowed the participants to hear both the babble noise played via headphone and the speech produced by the interlocutor. This was done to ensure that babble noise did not interfere with the sound recordings of the participant's speech. A Digitech QM1591 Decibel Meter was used to calibrate the sound level played via headphones. Before testing, the decibel meter was placed on the headphone to make sure the sound level of the auditory output was at 70-dBA. The conversation continued until the mother produced at least eight repetitions for each toy (compared to ten repetitions for each toy in IDS, where we expected a higher exclusion rate due to overlap with infant vocalizations). In case the mothers failed to produce the minimal eight repetitions for a toy, the researcher would prompt them to produce more repetitions by asking questions such as "what is the color of X?" or "does your infant like X?," etc.

In the ADS task, the procedure and the minimal number of repetitions were identical to that for the Lombard speech task. The only difference was that the conversation in the ADS task took place in a quiet environment.

D. Coding and measurements

Mothers' production data were coded by a trained native speaker of Mandarin Chinese (P.T.) in Praat (Boersma and Weenink, 2016), with the aid of spectrograms and waveforms. All the target words were first identified and segmented from the raw speech file for further analysis. Vowel onset and offset of the target words were annotated based on clear *F*2 regardless of utterance position, though position information was also labelled (see Sec. II D 1). Ten percent of the tokens were recoded by another trained native Mandarin Chinese speaker for a reliability check to ensure that the annotated vowel interval is consistent across different coders. This is important because the duration and pitch information were extracted automatically within this annotated interval. Correlations were conducted between coders for the annotated vowel duration, resulting in a Pearson's correlation of 0.90 for neutral tone words and 0.89 for tone sandhi words.

1. Neutral tone

Target syllables (the second syllable of each target word) across the two neutral tone vs full tone minimal pairs

TABLE II. Stimuli for eliciting tone sandhi productions, including a lexical disyllabic tone sandhi word, a right-branching trisyllabic noun phrase, and a left-branching trisyllabic noun phrase. The underlying tone and the surface tone of the target words are also provided.

Stimuli	Type	Target syllables	Underlying tone	Surface tone
/ma3 ji3/ (*ant*)	Disyllabic	/ma3 ji3/	T3T3	T2T3
/ɕiau3 ma3 ji3/ (*little ant*)	Right-branching	/ma3 ji3/	T3T3	T2T3
/ma3 ji3 tɕiau3/ (*ant's feet*)	Left-branching	/ma3 ji3/	T3T3	T2T2

(pair 1: /ʂγ2 **t^{h}ou0**/ vs /ʂγ2 **t^{h}ou2**/; /tɕ^{h}i2 **tsi0**/ vs /tɕ^{h}i2 **tsi3**/) and three registers (IDS, Lombard speech, ADS) were annotated. The utterance position of the target syllable was also annotated. Since the neutral tone and its full tone counterpart were always carried by the second syllable of a target disyllabic word, the target tones appeared either in utterance-medial position (when medial in an utterance) or in utterance-final position (when final in the utterance). The absolute duration (in milliseconds) of both syllables (first and second syllable) of each target word were measured and the normalized duration was computed for the target syllable (second syllable), using the following formula:

$$\text{Normalized duration} = \text{2nd vowel duration}/(\text{1st vowel duration} + \text{2nd vowel duration}). \quad (1)$$

In addition, four fundamental frequency parameters were measured from the vocalic part of the target second syllable: *f*0 onset, *f*0 offset, *f*0 minimum, and mean *f*0. *F*0 onset and *f*0 offset were measured as the *f*0 values (in Hertz) of the 5% and 95% points from the vocalic portion so that tonal coarticulation and micro prosodic perturbation from neighboring consonants could be minimized; the *f*0 minimum and mean *f*0 were then measured from 5% to 95% of the vocalic portion, respectively (see Fig. 1).

*F*0 points measured by Praat were checked and mistracked points were manually revised, to correct for the "doubling" or "halving" errors in pitch tracking. In the analysis, *f*0 values were transformed to semitones from observed Hertz values with 50 Hz as the reference, using the following formula:

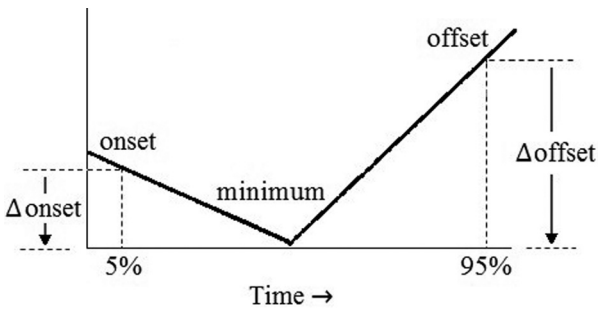

FIG. 1. Pitch parameters schematized for a contour tone (T3), including Δ onset and Δ offset. Δ onset and Δ offset were computed as the differences between *f*0 onset/offset and the minimal *f*0. *F*0 onset and offset were measured as the *f*0 values (in Hertz) of the 5% and 95% points from the vocalic portion; minimal *f*0 was measured as the minimal *f*0 value from 5% to 95% of the vocalic portion.

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

Extracted *f*0 onset, *f*0 offset, and *f*0 minimum values were then used to compute Δ onset (changes in tone onset) and Δ offset (changes in tone onset), using the following formulas (see Fig. 1):

$$\Delta\,\text{Onset} = f0\,\text{onset} - f0\,\text{minimum};$$
$$\Delta\,\text{offset} = f0\,\text{offset} - f0\,\text{minimum}. \quad (3)$$

Δ onset and Δ offset are two parameters that can effectively quantify the pitch contour of Mandarin tones (Shen *et al.*, 1993; Moore and Jongman, 1997). Δ onset indicates the size of the falling component of the pitch contour, and Δ offset indicates the size of the rising component. For example, according to Yip (2002), neutral tone exhibits a falling pitch when following a T2 syllable (as in /ʂγ2 **t^{h}ou0**/ and /tɕ^{h}i2 **tsi0**/ of our stimuli), so the *f*0 minimum was the same as *f*0 offset of the pitch contour, resulting in a larger Δ onset and a smaller Δ offset. A full set of the description of Δ onset and Δ offset patterns across tones is presented in Table III.

Among all these parameters, four duration and fundamental frequency parameters were adopted in the later analysis: normalized duration, mean *f*0, Δ onset, and Δ offset. These four parameters can quantify both the duration and pitch contour of tone production, which are two of the most important dimensions in distinguishing between neutral tone and full tone productions (Li *et al.*, 2014).

2. Tone sandhi

Target tone sandhi syllables (/ma3 ji3/) across three contexts (disyllabic word, right-branching noun phrase, and left-branching noun phrase) and three registers (IDS, Lombard speech, ADS) were annotated. Three fundamental frequency parameters were derived from tone sandhi target syllables for later analysis: *f*0 mean, Δ onset, and Δ offset, using the same method as described above. These parameters can quantify the pitch contour of tone production, which is the most important dimension in distinguishing between the underlying tone (T3, with a dipping pitch) and the surface tone (T2, with a rising pitch) of tone sandhi syllables (Moore and Jongman, 1997).

E. Statistical analysis

A total of 3285 tokens were included in the analysis, including 1821 neutral tone vs full tone productions and 1464

TABLE III. Pitch contours of lexical tones and neutral tone (following T2) and their corresponding Δ onset and Δ offset patterns: "+" indicates a relative large value and "−" indicates that the value is relatively small or close to zero.

Tone	Pitch contour	Δ onset	Δ offset
T1	level	−	−
T2	rising	−	+
T3	dipping	+	+
T4	falling	+	−
T0 (following T2)	falling	+	−

tone sandhi productions. The 1821 neutral tone vs full tone productions included 926 neutral tone syllables and 895 full tone counterparts. The 1464 tone sandhi productions included 565 trisyllabic tone sandhi words, 442 right-branching noun phrases, and 457 left-branching noun phrases. An additional 254 neutral tone productions and 111 tone sandhi productions 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 produced in whisper; mispronunciation.

The data were analyzed using R (R Core Team, 2016). A linear mixed-effects model was performed to compare the acoustic characteristics across tone categories and across registers, using the "lme4" package (Bates *et al.*, 2015) and the "lmerTest" package (Kuznetsova *et al.*, 2013). The *post hoc* test was performed on these models to conduct pairwise comparisons, using the "lsmeans" package (Lenth, 2016).

III. RESULTS

A. Neutral tone

Figures 2 and 3 summarize the pitch contours and the acoustic characteristics (normalized duration, mean *f*0, Δ onset and Δ offset) of the neutral tone vs full tone syllables across the two utterance positions (medial and final), the three registers (IDS, Lombard speech, ADS) and the two minimal pairs (pair 1: T0 vs T2; pair 2: T0 vs T3). Means and SDs for these parameters are provided in the online supplementary material.[1]

To test H1 regarding the acoustic realization of neutral tone syllables, these acoustic parameters were compared between neutral tone and full tone productions across registers in the two minimal pairs. Two separate linear mixed-effect models for pair 1 and pair 2 were performed on these parameters with two fixed factors: Tone (neutral tone and full tone), Position (medial and final), and Register (IDS, Lombard speech, and ADS). A random factor was also included: Subject (15 subjects). To keep the model optimal for generalizing the data, random slopes of Subject for main effects of all fixed factors were included (Barr *et al.*, 2013), i.e., random slopes of Subject on Tone, Position, and Register.[3] The results are summarized in Appendix A.

Our results showed that, for both pairs, the main effect of Tone was significant for all four acoustic parameters, the two-way interaction of Tone × Register was not significant for most parameters except for mean *f*0, Δ onset and Δ offset of syllables in pair 2, and the three-way interaction of Tone × Position × Register was not significant for most parameters except for mean *f*0 of syllables in pair 2 and Δ offset of syllables in both pairs. Pairwise comparisons were adopted to further compare the neutral tone and full tone syllables across positions and registers (see Figs. 2 and 3). The results showed that, for IDS, neutral tone and full tone syllables exhibited significant duration (normalized duration) and pitch height (mean pitch) differences in both medial and final positions, and they exhibited significant pitch contour (Δ onset and Δ offset) differences in the final position; for Lombard speech, neutral tone and full tone syllables showed significant duration (normalized duration) differences in

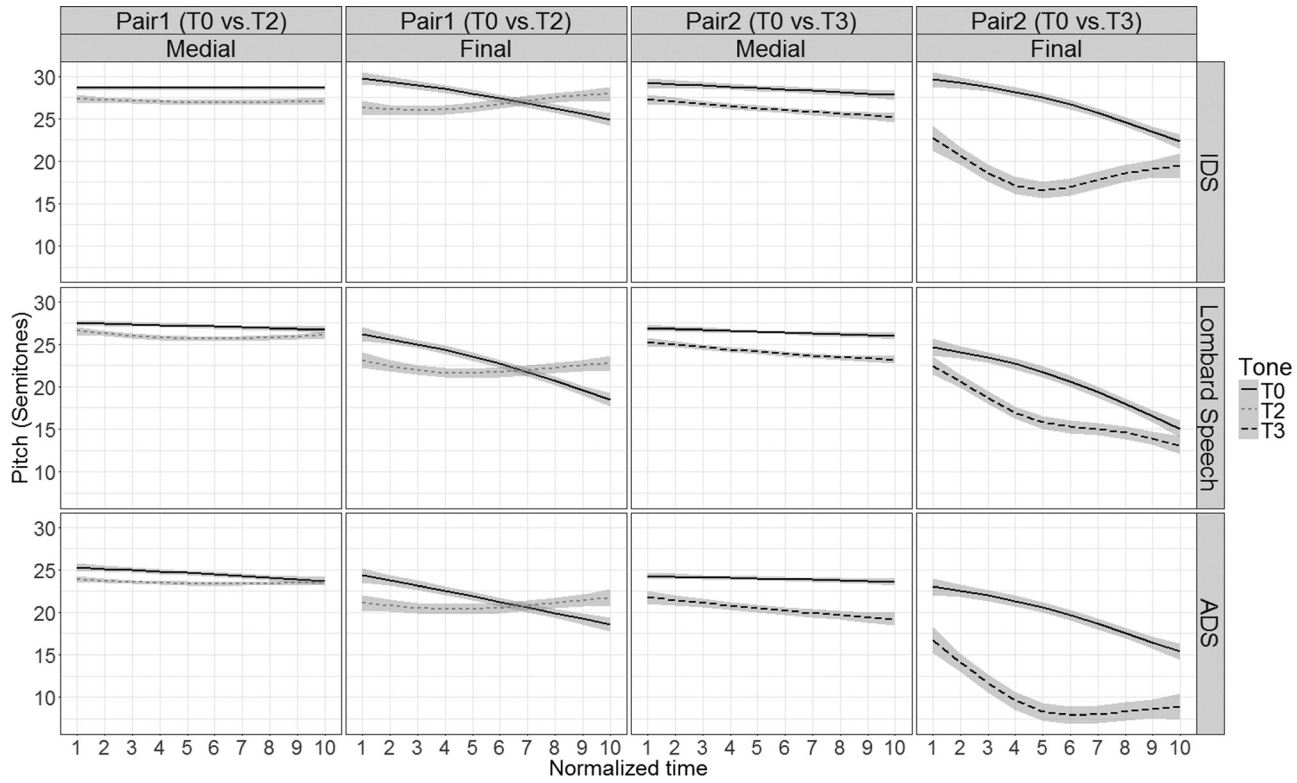

FIG. 2. Pitch contours of neutral tone and full tone syllables across three registers (IDS, Lombard speech, and ADS) and two utterance positions (medial and final) in two neutral tone vs full tone minimal pairs (pair 1: T0 vs T2; pair 2: T0 vs T3).

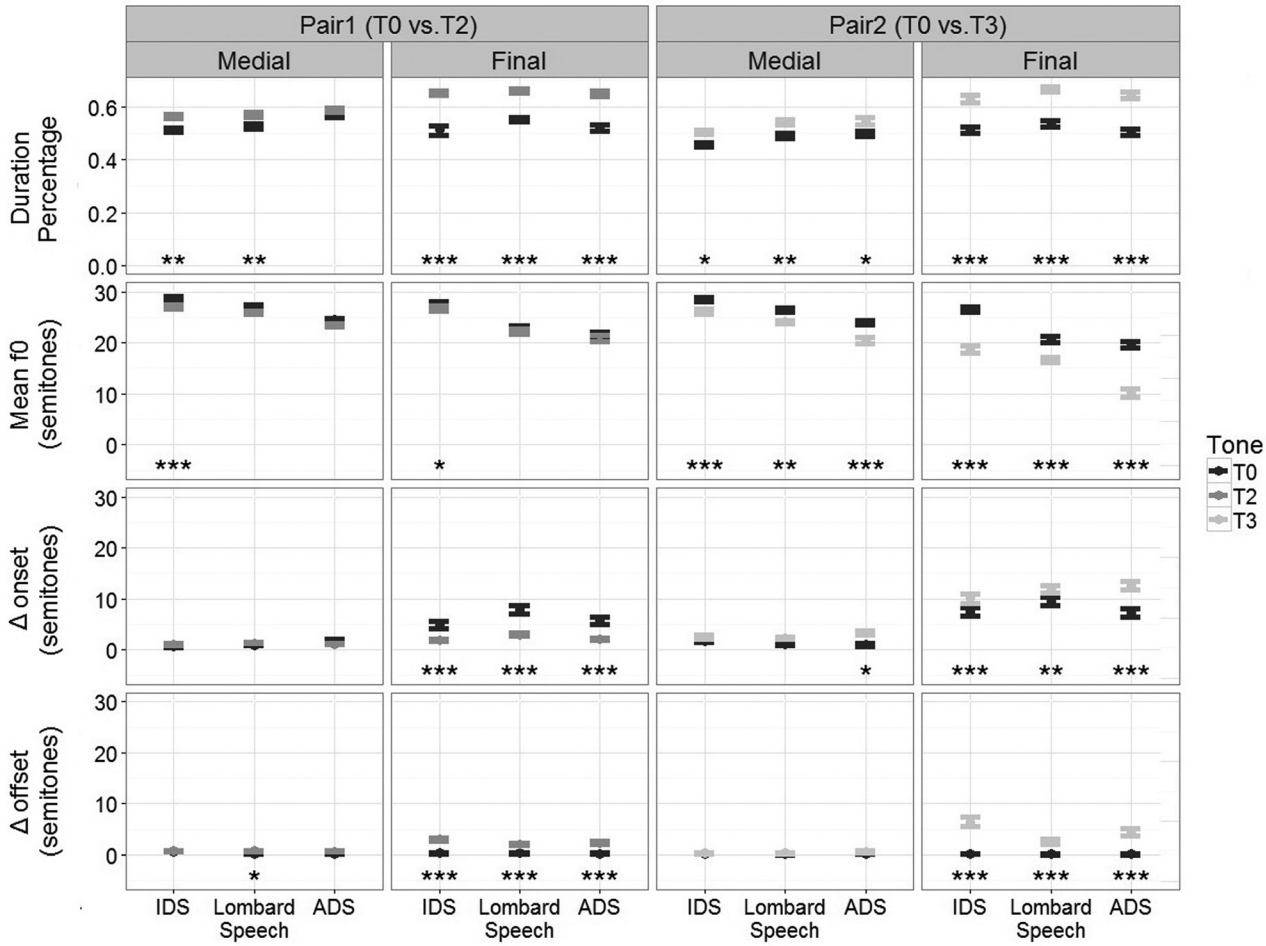

FIG. 3. Acoustic characteristics of neutral tone and full tone syllables across three registers (IDS, Lombard speech, and ADS) and two utterance positions (medial and final) in two neutral tone vs full tone minimal pairs (pair 1: T0 vs T2; pair 2: T0 vs T3). Four acoustic parameters are included: (1) normalized duration, (2) mean *f*0, (3) Δ onset, and (4) Δ offset. Results of significant pairwise comparisons between neutral tone and full tone syllables are illustrated by asterisks, and the number of which indicates the level of statistical significance between neutral tone and full tone syllables: $p < 0.05$*; $p < 0.01$**; $p < 0.001$ ***.

both medial and final positions, and they exhibited significant pitch contour differences in the final position.

We also ran a separate analysis to compare the absolute raw duration (in milliseconds) of both neutral tone and full tone syllables across registers. The results showed that, relative to ADS, IDS exhibited a longer absolute duration (in milliseconds) for both the neutral tone syllable [$\beta = 29.99$, standard error (SE) = 9.64, $t = 3.11$, $p < 0.05$] and its preceding full tone syllable ($\beta = 33.82$, SE = 9.64, $t = 3.51$, $p < 0.01$). Lombard speech, in contrast, did not show this pattern. Relative to ADS, neither the neutral tone syllable nor its preceding full tone syllable (neutral tone: $\beta = 17.93$, SE = 7.86, $t = 2.28$, $p = 0.07$; full tone: $\beta = 11.48$, SE = 7.85, $t = 1.46$, $p = 0.32$) was lengthened. This result indicated that, in IDS, speakers lengthened both neutral tone and full tone syllables while maintaining their durational difference; in Lombard speech, in contrast, speakers did not change duration of either neutral tone syllables or full tone syllables, and the durational difference between neutral tone and full tone syllables was maintained as well.

These results did not support H1, which predicted that, relative to ADS, neutral tone syllables would be acoustically realized as full tone syllables (i.e., their full tone counterparts) in IDS and Lombard speech. Rather, our results demonstrated that, in both registers, neutral tone and the full tone counterparts were distinct in terms of normalized duration, pitch height, and pitch contour, and this distinction was exaggerated in the final position.

To test H3 regarding the registers' difference in modifying neutral tone syllables, we performed another linear mixed-effects model on the acoustic parameters of the two neutral tone syllables only (/t^{h}ou0/ and /tsi0/). Three fixed factors Syllable (/t^{h}ou0/ and /tsi0/), Position (medial and final), and Register (IDS, Lombard speech and ADS) and the random factor Subject (15 subjects) were included. Random slopes of Subject for main effects of all fixed factors were included and were kept in the model as well.[4] The results are summarized in Appendix B.

The results showed that the main effect of Register was significant for mean *f*0 and Δ offset, and the interaction of Position × Register was significant for these parameters as well. This interaction indicates neutral tone syllables were modified differently across registers in terms of mean *f*0 and Δ offset. Pairwise comparisons on register differences in the interaction of Position × Register indicated that (1) IDS exhibited a higher mean *f*0 for neutral tone syllables than Lombard speech and ADS in both positions, while Lombard

speech exhibited a higher mean *f*0 than ADS only in the medial position ($\beta = 2.5$, SE = 0.58, $t = 4.34$, $p < 0.001$) rather than the final position ($\beta = 1.07$, SE = 0.58, $t = 1.86$, $p = 0.16$); (2) IDS exhibited a larger Δ offset of neutral tone syllables in the medial position than Lombard speech ($\beta = 0.27$, SE = 0.08, $t = 3.37$, $p < 0.01$) and ADS ($\beta = 0.28$, SE = 0.08, $t = 3.35$, $p < 0.01$). These results supported H3, which predicted that IDS will modify both pitch height and pitch contour of neutral tone syllables, while Lombard speech will modify the pitch height only.

B. Tone sandhi

Figures 4 and 5 summarize the pitch contours and the acoustic characteristics (mean *f*0, Δ onset and Δ offset) of tone sandhi syllables /ma/ and /ji/ across contexts (disyllabic word, right-branching noun phrase, and left-branching noun phrase) and registers (IDS, Lombard speech, ADS). Means and SDs for these parameters are provided in the online supplementary material.[2]

To test H2 regarding the acoustic realization of the tone sandhi syllables, the acoustic parameters of /ma/ and /ji/ were compared across contexts and registers. Two separate linear mixed-effect models for /ma/ and /ji/ were performed on these parameters with two fixed factors: Context (disyllabic word, right-branching noun phrase, and left-branching noun phrase) and Register (IDS, Lombard speech, and ADS). The random factor Subject was also included (15 subjects). All random slopes were included in the model to keep the model optimal generalizing the data.[5] The results are summarized in Appendix C.

The results showed that, for both /ma/ and /ji/, the main effect of Context was significant for all parameters, and the interaction of Context × Register was not significant. These results indicate that, although the pitch height and pitch contour of tone sandhi syllables varied across contexts, this variation did not change as a function of register.

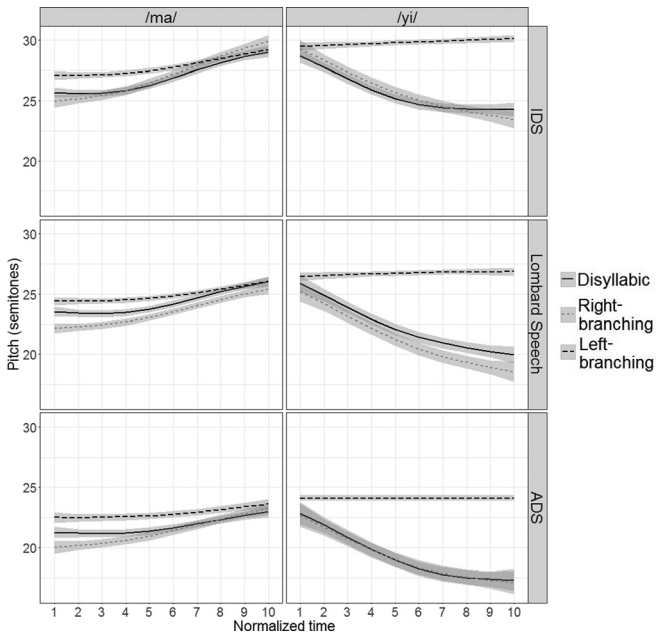

FIG. 4. Pitch contours of tone sandhi syllables /ma/ and /ji/ across three contexts (disyllabic word, right-branching noun phrase, and left—branching noun phrase) and three registers (IDS, Lombard speech, and ADS).

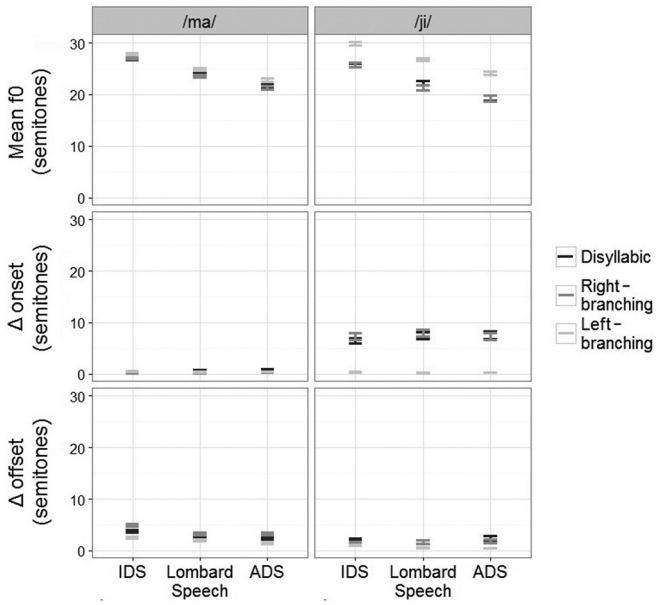

FIG. 5. Acoustic characteristics of tone sandhi syllables /ma/ and /ji/ across three contexts (disyllabic word, right-branching noun phrase, and left-branching noun phrase) and three registers (IDS, Lombard speech. and ADS). Three acoustic parameters are included: (1) mean *f*0, (2) Δ onset, and (3) Δ offset.

A Tukey HSD *post hoc* test was then performed on the main effect of Context to determine if tone sandhi rule was applied for tone sandhi syllables in correct contexts (/ma/ for all three contexts and /ji/ for the left-branching noun phrase). The results are presented in Appendix D, which shows that: (1) although Δ onset and Δ offset values of the syllable /ma/ differed across contexts, the syllable /ma/ exhibited a rising pitch contour in all contexts, as illustrated in Fig. 4. It indicates that speakers applied the tone sandhi rule on this syllable and produced it with the surface tone (T2); (2) the syllable /ji/ exhibited a higher mean *f*0 and a smaller Δ onset and a smaller Δ offset for the left-branching noun phrase than other contexts, which indicates that it exhibited a high-level tone in the left-branching noun phrase and a low dipping tone in other contexts (also see Fig. 4). It indicates that speakers applied the tone sandhi rule on the syllable /ji/ according to the prosodic structure, i.e., tone sandhi rule applied for /ji/ for the left-branching noun phrase, whereas the high-level pitch contour for /ji/ in this case was a result of the coarticulation effect that changed the surface tone (T2) of /ji/ to a high level tone (T1).[6]

These results did not support H2, which predicted that, relative to ADS, tone sandhi syllables would be modified and realized with the underlying T3 in IDS and Lombard speech. Thus, our results show that, although the acoustic realization of tone sandhi syllables varied across contexts, this variation did not change as a function of register, and speakers correctly applied tone sandhi across registers.

We also observed a significant main effect of Register on (1) mean *f*0 for syllable /ma/ and /ji/ and (2) Δ offset for syllable /ma/. To test H3 regarding differences in the modification of tone sandhi syllables /ma/ and /ji/ across registers,

we performed a *post hoc* test on these effects. The results showed that (1) for both /ma/ and /ji/, IDS exhibited a higher mean *f*0 than Lombard speech (/ma/: $\beta = 3.05$, SE = 0.52, $t = 5.82$, $p < 0.001$; /ji/: $\beta = 3.62$, SE = 0.3, $t = 12.27$, $p < 0.001$) and ADS (/ma/: $\beta = 5.4$, SE = 0.57, $t = 9.46$, $p < 0.001$; /ji/: $\beta = 6.26$, SE = 0.3, $t = 20.63$, $p < 0.001$), and Lombard speech exhibited a higher mean *f*0 than ADS (/ma/: $\beta = 2.35$, SE = 0.3, $t = 7.72$, $p < 0.001$; /ji/: $\beta = 2.64$, SE = 0.33, $t = 7.92$, $p < 0.001$); (2) for syllable /ma/, IDS exhibited a larger Δ offset than ADS ($\beta = 1.71$, SE = 0.36, $t = 4.75$, $p < 0.001$) and Lombard speech ($\beta = 1.76$, SE = 0.35, $t = 5$, $p < 0.001$). These results indicate that both IDS and Lombard speech increased the pitch height of tone sandhi syllables, and IDS exaggerated the rising component (Δ offset) of tone sandhi syllable /ma/, which had a rising pitch. These results supported H3, which predicted that IDS would modify both pitch height and pitch contour of the tone sandhi syllables, while Lombard speech would modify pitch height only.

IV. DISCUSSION

This study compared the acoustic characteristics of Mandarin neutral tone and tone sandhi productions in IDS, Lombard speech, and ADS. The results showed that, relative to ADS (in quiet, as a control), neutral tone and tone sandhi syllables were modified in IDS and Lombard speech to some extent, but the key tonal features of these contextual tones were maintained in both registers, including the short duration of the neutral tone and the rising pitch contour of the tone sandhi syllable. Additional register differences were found in the extent to which these contextual tones are modified. Specifically, IDS increased the pitch height and exaggerated the pitch contour of neutral tone (in utterance-medial position) and tone sandhi syllables, while Lombard speech increased the pitch height only.

The production of neutral tone and tone sandhi syllables in IDS and Lombard speech did not support our hypotheses (H1 and H2), which predicted that the slow speaking rate in the two registers might distort the two contextual tones to the extent that they would be realized as full tones. Thus, our results demonstrated that, in all registers, neutral tone is acoustically different from the full tone counterpart (controls) in tonal duration, pitch height (mean *f*0), and pitch contour (Δ onset and Δ offset). The tone sandhi rule was also correctly applied across registers. These results extend previous studies on the realization of Mandarin lexical tones in IDS (Liu *et al.*, 2007) and Lombard speech (Tang *et al.*, 2017), which observed that the register modifies the prosodic features of tones, but not at the expense of tonal contrast.

The durational difference between neutral tone and full tone syllables was maintained across registers, as reflected by the similar normalized duration (the duration proportion of neutral syllable in a disyllabic word). We hypothesized that, in both IDS and Lombard speech, the slow speaking rate might lengthen the neutral tone syllables, especially in the utterance-final position, and therefore distort the durational distinction between neutral tone and full tone syllables. However, our results showed that speakers maintain this durational distinction across positions and registers, though the strategies adopted were different: in Lombard speech, the raw/absolute duration of neutral tone syllables and full tone syllables were unchanged, and therefore the durational distinction was maintained; in IDS, in contrast, the raw/absolute duration of both neutral tone and full tone syllables was lengthened, but the durational difference was preserved. This (raw/absolute) durational difference between IDS and Lombard speech might be associated with different modifications on pitch contour across registers. In other words, it is possible that the exaggerated duration of IDS might be related to the need to implement its exaggerated pitch contour, i.e., the exaggerated contour in IDS might need more time to be implemented. This assumption is supported by a significant correlation found between the raw/absolute duration and pitch range (*f*0 maximum–*f*0 minimum) of the neutral tone and full tone syllables across registers (IDS: $r = 0.49$, $n = 1520$, $p < 0.001$; Lombard speech: $r = 0.44$, $n = 1146$, $p < 0.001$; ADS: $r = 0.38$, $n = 1030$, $p < 0.001$).

Similarly, the slow speaking rate in IDS and Lombard speech did not alter the realization of tone sandhi syllables. This result is not consistent with Speer *et al.* (1989), who claimed that the size (number of syllables) of the tone sandhi domain (the domain where the tone sandhi rule is applied) is one of the most important factors in determining how tone sandhi is realized. According to this view, the slow speaking rate is likely to induce boundaries between words and therefore break up the prosodic domain in which the tone sandhi rule applies. However, our results indicate that the speaking rate does not seem to influence the tone sandhi domain, or at least that the slow speaking rate of IDS and Lombard speech does not necessarily affect the size of the tone sandhi domain. This interpretation is in line with the findings in Kuo *et al.* (2007). They compared tone sandhi productions in slow, normal, and fast speech rate, and found that the size of the prosodic domain is larger in the slow speech rate than other speech rates. However, since the length of tone sandhi words in the present study were relatively short, i.e., either disyllabic or trisyllabic, it is possible that the effect of the speaking rate is more likely to influence the size of the tone sandhi domain at the sentence level. This is an issue that could be addressed in future studies.

Regarding the reason for children's late acquisition of neutral tone and tone sandhi, it seems unlikely that this is related to the exposure to acoustically distorted exemplars for these contextual tones in the language input learners hear. An alternative reason for their late acquisition may relate to the frequency of these contextual tones in the input. It has been shown that phoneme frequency in children's ambient language plays an important role in phonological development. For example, de Boysson-Bardies and Vihman (1991) conducted a cross-language study on French, English, Swedish, and Japanese infants' babbling and found a clear correlation between infants' babbling patterns and the distribution of consonantal place and manner categories of infants' ambient languages. Ingram (1988) also found that, relative to English-learning infants, word-initial /v/ is acquired much earlier by Swedish-, Estonian-, and Bulgarian-learning children, as it plays a more prominent role in the lexicon of these

languages. This evidence suggests a close relationship between language acquisition and phoneme frequency in the infant's ambient language. Therefore, it is possible that neutral tone and tone sandhi words do not occur very frequently in children's language input, resulting in the later acquisition of these tones. Some evidence suggests that the frequency of tone sandhi syllables in IDS is below 5%, and most of these syllables are disyllabic lexicalized items where productive application of sandhi rule might not be applied (Wang, 2011). Therefore, children may not receive enough input of the right type to allow them to learn the tone sandhi rule early. However, little is known about the frequency of neutral tone in children's language input and the relationship between tone input and children's tone acquisition: this needs to be further explored.

Another potential reason for the late acquisition of these contextual tones could be attributed to the challenge in learning the prosodic contexts that trigger neutral tone and tone sandhi. For example, the pitch contours of neutral tone syllables differ depending on the preceding lexical tones (Cao, 1992). Tone sandhi involves syllables being realized as a rising tone (full sandhi), a low falling tone (half sandhi), or a dipping tone (T3) in different tonal and prosodic contexts (Yip, 2002). These variations might be challenging for young learners, and require sufficient exposure to these different types of contexts for learning to take place.

With respect to the hypothesis that there would be register differences in the realization of neutral tone and tone sandhi (H3), neutral tone and tone sandhi syllables in IDS showed an overall raised pitch and an exaggerated pitch contour as compared with ADS. In contrast, neutral tone and tone sandhi syllables in Lombard speech exhibited an overall higher pitch only relative to ADS. These results are consistent with previous findings on lexical tones in IDS (Kitamura *et al.*, 2002; Liu *et al.*, 2007; Xu Rattanasone *et al.*, 2013) and Lombard speech (Zhao and Jurafsky, 2009; Boontham *et al.*, 2016).

The increased pitch height and exaggerated pitch contour of IDS observed in the present study are consistent with previous IDS studies which found similar modifications for lexical tones (i.e., Liu *et al.*, 2007; Tang *et al.*, 2017). It has been claimed that IDS mainly serves three functions: attracting and maintaining infants' attention, communicating positive emotion or affect between a caregiver and an infant, and (possibly) facilitating language acquisition (Song *et al.*, 2010). It is generally agreed that the increased pitch height in IDS is mainly associated with attentional and affective functions (Fernald and Kuhl, 1987; Kitamura and Burnham, 2003), and the exaggerated pitch contour in IDS is likely to be related to attentional and didactic functions (Trainor and Desjardins, 2002; Kitamura and Burnham, 2003; Uther *et al.*, 2007). This is in line with our observed effect of expanded pitch contour in IDS, but not in Lombard speech. It follows that this phonetic dimension of pitch was manipulated to cater to the needs of the addressee. That is, the pitch contour was exaggerated to facilitate tone learning by drawing the infants' attention through the overall pitch increase to the pitch movement of the contextual tones in IDS. The increased pitch height in Lombard speech, on the contrary, is consistent with the suggestion of Uchanski (2005) that it might be merely a by-product of an increased vocal effort. Evidence from Liénard and Di Benedetto (1999) also indicates that pitch (*f*0) increases with vocal effort as defined in terms of the physical distance between two persons. A similar argument has also been put forward in Uther *et al.* (2007) who compared the acoustic parameters across IDS, foreigner-directed speech (FDS, to adults), and ADS. The authors found that, on the one hand, vowel space was expanded in both IDS and FDS but not ADS, reflecting the didactic purposes of both registers; on the other, pitch was higher in IDS than FDS and ADS, and IDS was rated the highest with a positive effect. Since pitch increase can be associated with both attentional/affective (in IDS) and vocal effort (in Lombard speech) in the present study, it would be interesting to compare a Lombard version of IDS to Lombard speech in future studies to examine the effects of different communicative aims on pitch increase in the two registers. It is also possible that the pitch increase in the current study might be related to the involvement of the experimenter in eliciting Lombard speech. Perhaps the speaker might unconsciously increase her pitch in response to the experimenter. Therefore, future studies could do better by using a co-opted confederate to eliminate the potential influence from the experimenter.

V. CONCLUSION

Mandarin lexical tones are reported to be hyperarticulated in IDS and Lombard speech to benefit listeners. However, it has been unclear how contextual tones such as neutral tone and tone sandhi are realized in these registers, which might influence children's acquisition of these tones. The results of this study show that, although these contextual tones undergo certain modifications in both IDS and Lombard speech, these modifications do not distort their key tonal features, and they are still well-realized in these registers. These findings shed light on how Mandarin contextual tones are realized in the early language input to infants, and provide further insight into why these contextual tones might be later acquired. Registers differ in their modifications of these contextual tones between IDS and Lombard speech, but these appear to primarily reflect differences in addressees and communicative situations.

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 and strengthened this manuscript. This research was supported, in part, by a Macquarie University iMQRES scholarship to P.T., and the following grants: Grant Nos. ARC CE110001021 and ARC FL130100014 (K.D.). The equipment was funded by MQSIS 9201501719.

APPENDIX A

TABLE IV. Results of linear mixed-effects models on normalized duration, mean *f*0, Δ onset and Δ offset in two neutral tone vs full tone minimal pairs (pair 1: T0 vs T2; pair 2: T0 vs T3). Three fixed factors were included in the model: Tone (neutral tone and full tone), Position (medial and final), and Register (IDS, Lombard speech, and ADS). Asterisks indicate the level of statistical significance: $p < 0.05$*; $p < 0.01$**; $p < 0.001$***. The units are percentages for normalized duration and semitone for mean *f*0, Δ onset, and Δ offset.

			Acoustic parameters							
			Normalized duration		Mean *f*0		Δ onset		Δ offset	
Pair	Factors	DF	*F*	*p*	*F*	*p*	*F*	*p*	*F*	*p*
Pair 1	Tone	1	71.35	***	13.31	**	40.51	***	54.27	***
	Position	1	11.15	**	39.32	***	49.98	***	18.25	***
	Register	2	5.15	*	41.25	***	8.97	***	3.81	*
	Tone × Position	1	57.9	***	1.51	0.22	73.83	***	84.76	***
	Tone × Register	2	0.71	0.49	117	0.31	151	0.22	109	0.34
	Position × Register	2	5.41	**	19.93	***	5.99	**	1.88	0.15
	Tone × Position × Register	2	1.19	0.31	0.22	0.8	171	0.18	647	**
Pair 2	Tone	1	50.98	***	112.09	***	33.1	***	43.03	***
	Position	1	36.9	***	324.63	***	112.77	***	30.55	***
	Register	2	5.25	*	73.34	***	1.59	0.2	9.28	***
	Tone × Position	1	47.85	***	46.78	***	9.93	**	101.24	***
	Tone × Register	2	0.28	0.75	10.9	***	3.57	*	7.62	***
	Position × Register	2	3.84	**	10.38	***	386	*	8.72	***
	Tone × Position × Register	2	0.3	0.74	4.71	**	0.74	0.48	8.79	***

APPENDIX B

TABLE V. Results of linear mixed-effects models on normalized duration, mean *f*0, Δ onset and Δ offset for two neutral tone syllables (/t^{h}ou0/ and /tsi0/). Three fixed factors were included in the model: Syllable (/t^{h}ou0/ and /tsi0/), Position (medial and final), and Register (IDS, Lombard speech, and ADS). Asterisks indicate the statistical significance: $p < 0.05$*; $p < 0.01$**; $p < 0.001$***. The units are percentage for normalized duration and semitone for mean *f*0, Δ onset and Δ offset.

		Acoustic parameters							
		Normalized duration		Mean *f*0		Δ onset		Δ offset	
Factors	DF	*F*	*p*	*F*	*p*	*F*	*p*	*F*	*p*
Syllable	1	8.95	**	7.47	*	7.53	**	6.29	*
Position	1	0.66	0.43	76.81	***	83.33	***	0.15	0.71
Register	2	2.06	0.06	61.09	***	2.01	0.13	4.63	*
Syllable × Position	1	10.82	**	4.35	*	5.26	*	0.01	0.95
Syllable × Register	2	0.83	0.44	1.31	0.27	1.3	0.27	2.16	0.12
Position × Register	2	1.69	0.29	16.88	***	2.2	0.12	3.43	*
Syllable × Position × Register	2	0.53	0.59	0.37	0.69	0.39	0.72	1.6	0.2

APPENDIX C

TABLE VI. Results of the linear mixed-effects model on mean *f*0, Δ onset and Δ offset of two tone sandhi syllables /ma/ and /ji/ (presented in the upper and lower tables). Two fixed factors were included in these models: Context (disyllabic word, right-branching noun phrase, and left-branching noun phrase) and Register (IDS, Lombard speech, and ADS). Asterisks indicate the level of statistical significance: $p < 0.05$*; $p < 0.01$**; $p < 0.001$***. The unit is semitone for mean *f*0, Δ onset and Δ offset.

			Acoustic parameters					
			Mean *f*0		Δ onset		Δ offset	
Syllable	Factors	DF	*F*	*p*	*F*	*p*	*F*	*p*
/ma/	Context	2	11.58	***	14.31	***	25.94	***
	Register	2	53.68	***	1.62	0.23	14.59	***
	Context × Register	4	1.96	0.1	1.25	0.29	2.36	0.05
/ji/	Context	2	78.64	***	58.29	***	4.95	*
	Register	2	225.6	***	0.35	0.71	0.86	0.44
	Context × Register	4	1.53	0.19	1.15	0.33	1.21	0.3

APPENDIX D

TABLE VII. Tukey HSD pairwise comparison for the acoustic parameters of syllables /ma/ and /ji/ across contexts (1: disyllabic word; 2: right-branching noun phrase; 3: left-branching noun phrase). Asterisks indicate the level of statistical significance: $p < 0.05$*; $p < 0.01$**; $p < 0.001$***. The unit is semitone for mean *f*0, Δ onset and Δ offset.

Parameter	Syllable	Group difference	β	SE	t	p
Mean *f*0	/ma/	1–2	0.45	0.34	1.32	0.4
		1–3	−0.79	0.28	−2.86	*
		2–3	−1.25	0.28	−4.42	***
	/ji/	1–2	0.26	0.4	0.65	0.79
		1–3	−4.5	0.38	−11.82	***
		2–3	−4.76	0.47	−10.06	***
Δ onset	/ma/	1–2	0.42	0.08	5.1	***
		1–3	0.22	0.1	2.29	0.08
		2–3	−0.19	0.07	−2.6	*
	/ji/	1–2	−0.33	0.54	−0.61	0.82
		1–3	6.96	0.66	10.54	***
		2–3	7.29	0.79	9.23	***
Δ offset	/ma/	1–2	−0.9	0.23	−3.87	**
		1–3	0.94	0.18	5.29	***
		2–3	1.84	0.27	6.85	***
	/ji/	1–2	0.48	0.25	1.89	0.17
		1–3	1.42	0.46	3.05	*
		2–3	0.94	0.43	2.16	0.1

[1]See Appendix A in supplementary material at http://dx.doi.org/10.1121/1.5008372 for the means and SDs for the acoustic parameters of neutral tone words.

[2]See Appendix B in supplementary material at http://dx.doi.org/10.1121/1.5008372 for the means and SDs for the acoustic parameters of tone sandhi words.

[3]The R code of this model is Normalized duration/ Mean *f*0/ Δ Onset/ Δ Offset ∼ Tone * Position * Register + (1 + Tone + Position + Register |Subject).

[4]The R code of this model is Normalized duration/ Mean *f*0/Δ Onset/Δ Offset ∼ Syllable * Position * Register + (1 + Syllable + Position + Register|Subject).

[5]The R code of this model is Mean *f*0/Δ Onset/Δ Offset ∼ Context * Register + (1 + Context + Register|Subject).

[6]According to Xu (1994), a T2 syllable will be realized as a high-level tone (T1) when it is preceded by a syllable with a high pitch offset (i.e., T2 or T4) and followed by another syllable. This is the case for the left-branching noun phrase (the surface tone is T2T2T3) in the present study, and this coarticulation effect leads to the surface tone T2T2T3 of left-branching noun phrase being realized as T2T1T3.

Barr, D. J., Levy, R., Scheepers, C., and Tily, H. J. (**2013**). "Random effects structure for confirmatory hypothesis testing: Keep it maximal," J. Mem. Lang. **68**, 255–278.

Bates, D., Maechler, M., Bolker, B., Walker, S., Christensen, R. H. B., and Singmann, H. (**2015**). "lme4: Linear mixed-effects models using Eigen and S4. R package version 1.1-9," https://CRAN.Rproject.org/package=lme4 (Last viewed September 1, 2017).

Boersma, P., and Weenink, D. (**2016**). "Praat: Doing phonetics by computer" [Computer program]. Version 6.0.19, http://www.praat.org/ (Last viewed June 13, 2016).

Boontham, C., Onsuwan, C., Saimai, T., and Tantibundhit, C. (**2016**). "An analysis of Lombard effect on Thai lexical tones: The role of communicative aspect," in *Proceedings of the 16th Speech Science and Technology Conference*, Sydney, Australia, pp. 149–152.

Burnham, D., Kitamura, C., and Vollmer-Conna, U. (**2002**). "What's new pussycat? On talking to babies and animals," Science **296**, 1435.

Cao, J. (**1992**). "On neutral-tone syllables in Mandarin Chinese," Canadian Acoust. **20**, 49–50.

Cooper, R. P., Abraham, J., Berman, S., and Staska, M. (**1997**). "The development of infants' preference for motherese," Infant Behav. Dev. **20**, 477–488.

de Boysson-Bardies, B., and Vihman, M. M. (**1991**). "Adaptation to language: Evidence from babbling and first words in four languages," Lang. **67**, 297–319.

Demuth, K. (**1993**). "Issues in the acquisition of the Sesotho tonal system," J. Child Lang. **20**, 275–301.

Fernald, A., and Kuhl, P. (**1987**). "Acoustic determinants of infant preference for motherese speech," Infant Behavior Dev. **10**, 279–293.

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.

Grieser, D. L., and Kuhl, P. K. (**1988**). "Maternal speech to infants in a tonal language: Support for universal prosodic features in motherese," Dev. Psychol. **24**, 14–20.

Ingram, D. (**1988**) "The acquisition of word-initial [v]," Lang. Speech **31**, 77–85.

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.

Kitamura, C., and Burnham, D. (**2003**). "Pitch and communicative intent in mother's speech: Adjustments for age and sex in the first year," Infancy **4**, 85–110.

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 Behavior 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.

Kuo, Y., Xu, Y., and Yip, M. (**2007**). "The phonetics and phonology of apparent cases of iterative tone change in Standard Chinese," in *Experimental Studies in Word and Sentence Prosody*, edited by C. Gussenhoven and T. Riad (Mouton de Gruyter, Berlin, Germany), pp. 211–237.

Kuznetsova, A., Brockhoff, P. B., and Christensen, R. H. B. (**2013**). "Lmertest: Tests for random and fixed effects for linear mixed effect models (lmer objects of lme4 package)," R package version 2.0-29, pp. 1–17. Available at https://cran.r-project.org/web/packages/lmerTest/lmerTest.pdf (Last viewed May 4, 2016).

Lenth, R. V. (**2016**). "Least-squares means: The R package lsmeans" J. Stat. Softw. **69**, 1–33.
Li, A., Gao, J., Jia, Y., and Wang, Y. (**2014**). "Pitch and duration as cues in perception of neutral tone under different contexts in Standard Chinese," in *Proceedings of the Asia-Pacific Signal and Information Processing Association Annual Summit and Conference* (*APSIPA*), Angkor Wat, Cambodia, pp. 1–6.
Li, C. N., and Thompson, S. (**1977**). "The acquisition of tone in Mandarin-speaking children," J. Child Lang. **4**, 185–199.
Liénard, J. S., and Di Benedetto, M. G. (**1999**). "Effect of vocal effort on spectral properties of vowels," J. Acoust. Soc. Am. **106**, 411–422.
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. Maladies Ear, Larynx, Nose, Pharynx **37**, 101–119.
Moore, C. B., and Jongman, A. (**1997**). "Speaker normalization in the perception of Mandarin Chinese tones," J. Acoust. Soc. Am. **102**, 1864–1877.
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 June 21, 2016).
Schulman, R. (**1989**). "Articulatory dynamics of loud and normal speech," J. Acoust. Soc. Am. **85**, 295–312.
Shen, X., Lin, M., and Yan, J. (**1993**). "F0 turning point as an F0 cue to tonal contrast: A case study of Mandarin tones 2 and 3," J. Acoust. Soc. Am. **93**, 2241–2243.
Shih, C. (**1997**). "Mandarin third tone sandhi and prosodic structure," Linguist. Models **20**, 81–124.
Singh, L., Morgan, J. L., and Best, C. T. (**2002**) "Infants' listening preferences: Baby talk or happy talk?," Infancy **3**, 365–394.
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.
Speer, S. R., Shih, C. L., and Slowiaczek, M. L. (**1989**). "Prosodic structure in language understanding: Evidence from tone sandhi in Mandarin," Lang. Speech **32**, 337–354.
Summers, W. V., Pisoni, D. B., Bernacki, R. H., Pedlow, R. I., and Stokes, M. A. (**1988**). "Effects of noise on speech production: Acoustic and perceptual analyses," J. Acoust. Soc. Am. **84**, 917–928.
Tang, P., Xu Rattanasone, N., Yuen, I., and Demuth, K. (**2017**). "Phonetic enhancement of Mandarin vowels and tones: Infant-directed speech and Lombard speech," J. Acoust. Soc. Am. **142**, 493–503.
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? A comparison of foreigner-and infant-directed speech," Speech Commun. **49**, 2–7.
Wang, C. Y. (**2011**). "Children's acquisition of tone 3 sandhi in Mandarin," Ph.D. dissertation, Michigan State University, East Lansing, MI, pp. 1–336.
Xu, Y. (**1994**). "Production and perception of coarticulated tones," J. Acoust. Soc. Am. **95**, 2240–2253.
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), 181 pp.
Zhao, Y., and Jurafsky, D. (**2009**). "The effect of lexical frequency and Lombard reflex on tone hyperarticulation," J. Phon. **37**, 231–247.
Zhu, H., and Dodd, B. (**2000**). "The phonological acquisition of Putonghua (Modern Standard Chinese)," J. Child Lang. **27**, 3–42.