9th International Conference on Speech Prosody 2018
13-16 June 2018, Poznań, Poland

Chanted Call Tune in Tianjin Mandarin: Disyllabic Calls

Cong Zhang

Language and Brain Laboratory, University of Oxford
cong.zhang@ling-phil.ox.ac.uk

Abstract

This paper examines the chanted call tune in Tianjin Mandarin in order to investigate the possibilities of intonational components, i.e. pitch accents, boundary tones, etc., in a tonal language. Six native Tianjin speakers' production of disyllabic names and kinship terms were recorded. The speech materials were composed of a set of left-prominent disyllabic names and a set of right-prominent disyllabic names. The results show that there is a L% boundary tone at the end of an intonational phrase, regardless of which lexical tone it has. Different from data from the intonational polar questions, the L% boundary tone was phonetically manifested and overrode the lexical tone contours. A H* pitch accent was found to be associated with the H of each lexical tone. Lengthening was also found in the chanted call tune. The chanted call tune in Tianjin Mandarin can be represented as follows:

$$[[\text{H*}]\text{sustained}]_{\text{higher register}} + \text{L\%}.$$

Index Terms: chanted call; tone; intonation; Tianjin Mandarin

1. Introduction

Tonal languages make use of pitch both lexically and post-lexically. Whether lexical tones can be overridden by intonation contours is a question that is often asked. [1] discovered that two of the major differences between syntactically unmarked polar questions and statements in Tianjin Mandarin are **register change** and the extra **floating H% boundary tone** of the intonational polar question tune (hereinafter "IntQ"). The floating H% boundary tone does not have phonetic representation but has a phonological function; the lexical tone contours thus are not overridden. However, does Tianjin Mandarin utilise such floating boundary tones for all tunes? Are actual boundary tones allowed in this tonal language? To answer the above questions, chanted call tune is studied in this paper.

Chanted call (hereinafter "CC") tune, a.k.a. vocative chant, calling contours, etc., is one of the most studied tunes [2]–[4]. The majority of the existing studies are of Indo-European languages ([3], [5]–[14]), and very few studies investigated the CC tune outside Indo-European languages ([15]–[17]); [18] is the only one that can be found that investigated the CC tune in a tonal language, Thai; however, only the observations were outlined without presenting acoustic data.

In many of the abovementioned languages, their CC tunes share the following features: (i) there is a H* pitch accent involved; (ii) the boundary tones are lower than the H* pitch accent. (iii) the boundary tones are mostly lengthened, or in another word, sustained. In general, the CC tunes in these languages have a scooped shape as in (1), i.e. LHL or LH!H. However, the specific associations and alignments are language-specific, and differ between the CC tunes of different pragmatic functions in the same language.

(1) typical CC tune contour

This paper investigates the CC tune in a tonal language, Tianjin Mandarin. Disyllabic names and kinship terms are examined in the current study since there is a preference for disyllabic combinations in modern Mandarin. Monosyllabic, trisyllabic, and quadrisyllabic calls are also possible in Tianjin Mandarin, but the monosyllabic and quadrisyllabic ones are relatively rarer.

Tianjin Mandarin has four lexical tones (Figure 1, also [19], [20]), which are symmetrically distributed – L Tone (T1, low and slightly falling), High Tone (T2, high and slightly rising), LH Tone (T3, low rising), and HL Tone (T4, high falling).

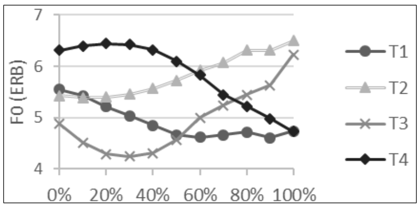

Figure 1. *Four Tianjin Citation Tones in ERB*

2. Methods

2.1. Speakers and Procedures

Six native speakers of Tianjin Mandarin (3 males and 3 females) were recorded. The age range of the informants was between 22 and 28. All were also capable of speaking standard Mandarin in formal situations, though with different degrees of Tianjin accent. None of them reported any speech disorders.

Recordings were made in a soundproof booth in Tianjin University, China. The informants were asked to read the materials, presented as Chinese characters, without context in a Microsoft PowerPoint presentation, as naturally as they could. The informants were asked to produce a statement when seeing a Chinese full stop "。" at the end of the utterances, and call the names out when seeing a Chinese exclamation mark "！". For the calls, they were told to imagine that the person with the name on the PowerPoint is at the other side of a valley in a mountain, and the informants needed their attention. All the materials were randomised. To avoid any potential list effect, each utterance was placed on a separate slide. Each item was non-consecutively repeated twice.

Originally, the participants were asked to call the names as if they are in close vicinity (downstairs), following many existing studies (e.g. , [5]: for dinner, [7]: in front of the speaker, or in the next room). However, such scenario did not create any

10.21437/SpeechProsody.2018-106

melodious chanted call; instead, the informants solely relied on the change of amplitude. Once the instruction was changed to the mountain scenario which depicted a farther distance, the prosody changed notably. Borras-Comes et al. [7] also found the physical distance to be a significant factor on the production of CC tune in Catalan. In Tianjin Mandarin, the distance needs to be much larger for the CC tune to be different from the statement tune. This indicates that, for a tonal language, the alternation of the intonational tune may be an additional strategy and would not be necessary unless more intensive methods are needed. Another factor behind the lack of change in the first setting may be the small space of the recording booth which limited the imagination of the informants, since in daily life or in a drama, chanted calls are not particularly rare.

2.2. Materials

Two sets of disyllabic materials were used in this study: left-prominent disyllabic calls and right-prominent disyllabic calls.

Set 1 (left-prominent disyllabic calls) was composed of left-prominence names and kinship terms as shown in Table 1, which contained the monosyllabic names, but doubled. This is a common way in Mandarin to make a nickname or an intimate call, such as /mama/ ‘mom’, /weiwei/ a nickname for ‘Wei’. The lexical tones of the second syllable for this set of data were all T0, i.e. the neutral tone that does not have any underlying tonal contour. The T0 syllables were reduced syllables that did not carry any lexical prominence.

Table 1. Materials for Set 1

	L Tone (Tone 1) T1T0	**H Tone (Tone 2) T2T0**	**LH Tone (Tone 3) T3T0**	**HL Tone (Tone 4) T4T0**
Kinship Terms	/mama/ ‘mom’	/jeje/ ‘grandpa’	/nainai/ ‘granny’	/meimei/ ‘sister’
Given Names	/weiwei/ ‘*Weiwei*’	/leilei/ ‘*Leilei*’	/ ʐweiʐwei/ ‘*Ruirui*’	/ʐweiʐwei/ ‘*Ruirui*’

Set 2 (right-prominent disyllabic calls) contained disyllabic names which were composed of a surname and a given name, e.g. Wang Wei, Li Lei. Both the first syllable and the second syllable had one of the four lexical tones. Two tokens were designed for each tonal combination. Therefore, a total of 32 names were used.

3. Results

In order to accurately capture the components of the CC tune, the contours of different tones were observed in comparison with the statement counterparts. Pitch scaling and pitch alignment are subsequently examined to provide quantitative support from phonetic details. Statistical analyses in this section were conducted using R (RStudio-Team, 2017), and a mixed-effect model was used. TYPE (Statement, Question), TONE (L-Tone, H-Tone, LH-Tone, HL-Tone), and SYLLABLE (First syllable, Second syllable) were taken as fixed-effect factors. SUBJECT (six different informants), ITEM (two different tokens), SUBJECT-specific, and ITEM-specific slopes were held as random-effect factors.

3.1. Visual Inspection

Examples from Set-1 data are presented in Figure 2. On the left, the tone of the first syllable is a low-falling T1, while the second syllable has T0 (neutral tone). The one on the right is T3T0 (LH + T0). Regardless of the lexical tones of the first syllable, both of the T0 syllables are falling. A L% can be observed from these figures.

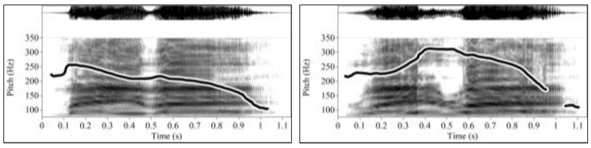

Figure 2. *CC tune of T1T0 (left) and T3T0 (right)*

Instances from Set 2 are presented in Figure 3 to compare between the statement tune (left) and the CC tune (right). A L% can also be found at the end of the intonational phrase of the CC tune. Furthermore, the overall register of the CC tune is also considerably higher than the statement tune. The contours are much more exaggerated, which lead to the possibility of a H*.

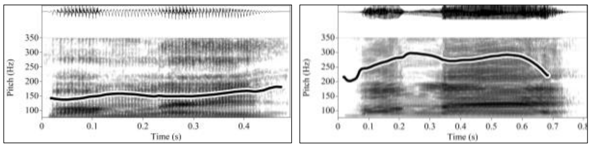

Figure 3. *Statement (left) and CC tune (right) of T2T2*

3.2. Duration

In both sets of data, the CC tune was significantly longer than the statement tune (Set1: TYPE ($\chi^2(1) = 33.15$, $p < 0.001$ ***) ; Set 2: TYPE ($\chi^2(1) = 42.79$, $p < 0.001$ ***)). The second syllables were longer than the first syllables in the CC tune in both sets of data, despite the fact that in T1 the second syllables were shorter in the statement tune. In Set 1, the first syllables of the CC tune were longer than the statement counterparts by 33.7ms on average, while the second syllables of the CC tune were longer than their statement counterparts by 164.8ms. In Set 2, the differences between the CC tune and the statement tune on the first syllables were 17.8ms, while those were 116.7ms on the second syllables.

3.3. Pitch Scaling

Table 2. Pitch Scaling

	Mean F0 (ERB)		**MaxF0 (ERB)**		**MinF0 (ERB)**	
	SET1	SET2	SET1	SET2	SET1	SET2
CC vs S	CC	CC	CC	CC	CC	CC
Syl.1 vs Syl.2	1st	2nd	2nd	2nd	1st	1st
Syl.1: CC-S	1.94	1.82	2.18	1.83	1.59	0.75
Syl.2: CC-S	1.9	1.88	1.86	1.79	0.99	0.63
CC: Syl.2-Syl.1	-0.25	0.16	-0.20	0.56	-1.20	-1.40
S: Syl.2-Syl.1	-0.21	0.10	0.12	0.59	-0.60	-1.28

In this study, we use mean pitch, values of F0 maxima and minima to analyse the scaling of the pitch. The scaling data and important comparisons are presented in Table 2. The second row compares which syllable of the disyllabic utterances produces higher values of the examined factors. The following rows are comparisons between the CC tune and statement tune in the same syllables, and between the two syllables within the same tunes.

In both sets of data, the CC tune is higher than the statement tune, in terms of all three factors. These results indicate a register lift – both the lowest part and the highest part of the tune were raised together.

3.4. Pitch Alignment

Through visual inspection, a H* was observed to be associated with the H part of each lexical tone. This section displays distribution plots of the F0 maxima and minima to show the alignment of the pitch peak and valley, with the orange dots and shades representing the CC tune, and the blue crosses the statement tune. On the basis of Figure 4, we further examine whether this association applies to most instances instead of merely to the example figures by means of displaying the maximum F0 alignment points for all instances in Set 2. Set 1 data are consistent with the results from Set 2. In the first syllables, most of T2 and T3's F0 maxima were aligned at the end of the first syllables, since they were both rising tones. T1 and T4 had more F0 maxima points towards the middle of the syllables, which were also where the H in the rhymes starts; however, due to tonal dissimilation, T1 (falling L) and T4 (HL) became T2 (rising H) in T1T1 and T4T1, so the distribution was more scattered. Further examination of T1T1 supports this analysis and thus strongly suggest that the H* is associated with the H of the lexical tone.

In the second syllables, the alignment of maximum F0 in TxT1 and TxT4 with the CC tune were similar to their counterparts in statements –T1 in TxT1 aligned at the beginning of the syllable, and T4 in TxT4, at the middle of the syllable, which were consistent with the F0 maxima alignment of the lexical tones in their forms. In the instances ending in rising tones, i.e. T2 and T3, however, the F0 maxima alignment moved away from the end of the syllable by a small but noticeable margin. The change was due to the insertion of the L% boundary tone, which will be further elaborated in the F0 minima alignment data in Figure 5.

a.
b.

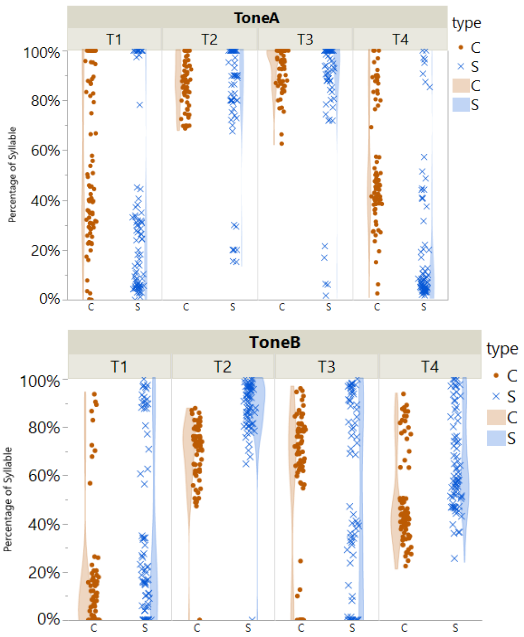

Figure 4. Alignment of Maximum F0 in Set 2: First Syllables (a) and Second Syllables (b)

Figure 5 presents the F0 minima alignment for Set-1 data, which only contains T0 syllables. In the statement tune, the majority of the F0 minima points were at approximate 60% of the T0 syllable, while the CC tune had much later F0 minima alignment, most of which were at around 90% of the syllable. Since T0 does not have an underlying tonal contour, the F0 minima points directly depicts the L% boundary tone. The fact that the L% is aligned more to the end of the T0 syllables in the CC tune than in the statement tune has provided strong evidence for the L% boundary tone analysis.

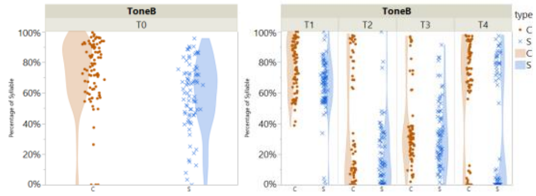

Figure 5. Alignment of Minimum F0 of the 2nd Syllables: Set 1 (left) and Set 2 (right)

4. Discussion

4.1. Lengthening

The results of the current study show that the duration of both syllables in the disyllabic words are significantly lengthened in the CC tune, with the second syllable being lengthened by a much larger margin. The sources of the lengthening are two: one is the side product of the L% boundary tone; the other one is lengthening solely for the purpose of chanting and calling.

The evidence for the first source, i.e. the L% boundary tone, can not only be found in the duration section, but also be gained from the analyses of F0 maxima alignment. As shown in Figure 4, the F0 maxima in T2 and T3 are aligned towards an earlier stage of the syllables with the CC tune than with the statement tune. The only explanation is that another tone was interpolated at the IP boundary.

The second source is from the tune itself. There are two types of lengthening – boundary-induced, and tune-induced "sustainment". This is one of the key points for the decision of whether to include the duration in the description of the tune or not, in order to delineate the principles that we should follow to determine whether lengthening is a component of a specific tune or not. Boundary-induced lengthening is a side product of the boundary tone, regardless of whether it is a floating boundary tone or an actual boundary tone. To identify such lengthening, both production data and perception data are useful. From the perspective of production, the durational differences do not take up a huge percentage, and removing the temporal difference would not create a distinctively different tune. From the perspective of perception, native listeners cannot consciously identify the temporal difference, irrespective of whether they are capable of distinguishing them subconsciously or not. The IntQ tune in Tianjin Mandarin in [1] is a typical case of such boundary-induced lengthening. If the durational difference is of this type, then it should not be included in its tune. However, if the temporal aspect serves the main purpose of forming a tune, then it should be included in the tune. For example, in the CC tune discussed in the current study, if it had the same length as a statement tune, the H* pitch accent would not sustain and thus perceptually could be regarded as a separate tune that served a different function.

It is regrettable that AM theory does not involve any temporal expression, which is nevertheless much needed. Only in the CC tune alone, due to the lack of temporal feature notation system, at least four different methods (using the boundary tone sign "%" alone, using "0%", labelling with the text "(sust)", using boundary tone in conjunction with a phrase tone) have been used to depict the "sustainment" of the boundary tone. Some languages change their intonation by means of changing the temporal aspect of the utterances. [21] lists a number of African languages that use lengthening as a question marker. The integration of the temporal descriptors into AM framework also prevents missing important information in depicting different pitch accents. For instance, [22] found significant temporal differences between L+H* and L*+H. Although the association was successfully captured, the temporal difference was not included in the tune.

4.2. Register

The register change is the clearest difference between the tunes in question and their statement counterparts, and is directly detectable through visual inspection. In both sets of data, the prominent syllable was higher than the non-prominent syllable. The maximum F0 was higher in the second syllables and the minimum F0 was lower in the second syllable, which means the range of the first syllables was much smaller than the second syllables. The register is another feature that needs to be included in the transcription system of AM framework. For a contour tone language, the register is an important feature that distinguishes tunes both in production and in perception. Not only the results of the current study highlight this need, other studies of non-tonal languages also suggest such absence. For instance, [8] reported the only difference between a default chanted call and a chanted call in anger is that the H* of the anger CC tune is higher in the register.

4.3. Pitch Accent

In the study of monosyllabic CC tune, we discovered a H* pitch accent, the association of h was yet to be ascertained. The current study provides new insight on the association. Based on visual inspection and the F0 maxima alignment results, there is strong evidence for the association of H* pitch accent with the H of every lexical tone, including the high parts of the level lexical tones. In the case of T0, which does not have an underlying tone, the H* is only associated with the first syllable, but spreads to the second syllable, as illustrated in (2):

(2) Examples of H* association rules:

H*H*	H*
LH HL	LH
Tx Tx	Tx T0

Comparing the results cross-linguistically, the results are directly in contrast with the CC tune in Thai, another contour tone language. The H* pitch accent of the CC tune in Tianjin Mandarin magnifies the contours of the lexical tones, while in Thai [18], the CC tune levels the lexical tones: for example, a high lexical tone, which has a continuous rise and is followed by a fall, becomes a mid-high level tone in the CC tune.

4.4. Boundary Tone

As reported in the results section, the current study of the disyllabic chanted call tune confirms the observation of the L% boundary tone. Clear evidence from visual inspection, together with the results of duration, F0 maxima alignment, and F0 minima alignment in particular, suggests that there is a L% boundary tone.

The discovery of the L% boundary tone in Tianjin Mandarin is noteworthy, in that it is an actual boundary tone, instead of a floating boundary tone as in the IntQ tune reported in [1]. This shows that it is entirely possible to have an extra tone interpolated at the IP boundary in Tianjin Mandarin – even for rising lexical tones, a L% boundary tone can be attached to the end. Such boundary tones have already been reported in Cantonese [23], this paper is the first to report such a phenomenon in Tianjin Mandarin.

Although both L% and sustainment occur in the CC tune in Tianjin Mandarin as they do in the non-tonal languages described in [13]-[15], the CC tune in Tianjin Mandarin differs from those languages in that they have sustained boundary tones, rather than lengthened pitch accents. As discussed in the lengthening section (4.1), the durational differences are not solely brought by the boundary tone in Tianjin Mandarin. Evidence from F0 extrema alignment can also serve as evidence that the boundary tone only took a small percentage of the whole tune. This suggests that even though a boundary tone is interpolated in this tonal language, the contours of the lexical tones have to be kept in their original form as much as possible. Thus, inserting a new boundary tone which takes a substantially large proportion of the tune is not likely.

5. Discussion and Conclusions

The three major findings of the chanted call tune in this study are raised register, sustained H* which associates with the H of each lexical tone, and L%.

It goes through the following processes and can be analysed as (3):

a. Raise the register;
b. Lengthen every syllable, with the biggest amount on the last syllable;
c. Associate H* to every H in each lexical tone;
d. Spread H* to syllable without lexical tone, if any;
e. Attach L% at IP boundary.

(3) Association of H* and L% in monosyllabic and disyllabic CC tunes:

Set1: TXT0	Set2: TATB
H* L%	H*H* L%
[LH]$_{IP}$	[LH HL]$_{IP}$
Tx T0	Tx Tx

Summarising all results from this paper, the CC tune in Tianjin Mandarin is:

[[H*]sustained]$_{\text{higher register}}$ + L%.

6. References

[1] C. Zhang, "Tones and Tunes in Tianjin Mandarin," in *Tone and Intonation in Europe*, 2016.

[2] K. L. Pike, *The Intonation of American English*. Ann Arbor, Michigan: University of Michigan Press, 1945.

[3] D. R. Ladd, "Stylized intonation," *Language*, vol. 54, no. 3. pp. 517–540, 1978.

[4] D. Gibbon, *Perspectives of intonation analysis*, vol. 9. Bern: Lang., 1976.

[5] A. Arvaniti, M. Żygis, and M. Jaskuła, "The Phonetics and Phonology of the Polish Calling Melodies," in *Phonetica*, 2016, no. 73, pp. 338–361.

[6] C. Gussenhoven, "The Dutch foot and the chanted call," *J. Linguist.*, vol. 29, no. 1, pp. 37–63, 1993.

[7] J. Borràs-Comes, R. Sichel-Bazin, and P. Prieto, "Vocative Intonation Preferences are Sensitive to Politeness Factors," *Lang. Speech*, vol. 58, no. 1, pp. 68–83, 2015.

[8] N. Sadat-Tehrani, "The intonational grammar of Persian," University of Manitoba, 2007.

[9] Z. Fagyal, "Chanting intonation in French," *U. Penn Work. Pap. Linguist.*, vol. 4, no. 2, pp. 77–90, 1997.

[10] B. Hayes and A. Lahiri, "Durationally specified intonation in English and Bengali," in *Proceedings of the 1990 Wenner-Gren Center Conference on Music, Language, Speech and the Brain*, J. Sundberg, L. Nord, and R. Carlson, Eds. Wenner-Gren Center International Symposium Series. Palgrave, London, 1992, pp. 78–91.

[11] A. Arvaniti and M. Baltazani, "Intonational analysis and prosodic annotation of Greek spoken corpora," in *Prosodic Typology : The Phonology of Intonation and Phrasing*, S.-A. Jun, Ed. Oxford: Oxford University Press, 2005, pp. 1–52.

[12] M. Grice, S. Baumann, and R. Benzmüller, "German Intonation in Autosegmental Metrical Phonology," in *Prosodic typology: The phonology of intonation and phrasing*, S.-A. Jun, Ed. Oxford: Oxford University Press, 2005, pp. 55–83.

[13] S. Frota, M. Cruz, F. Svartman, G. Collischonn, A. Fonseca, C. Serra, O. Pedro, and M. Vigário, "Intonational variation in Portuguese: European and Brazilian varieties," in *Intonation in Romance*, S. Frota and P. Prieto, Eds. Oxford: Oxford University Press, 2015, pp. 235–283.

[14] S. Frota, "The intonational phonology of European Portuguese," in *Prosodic Typology II: The Phonology of Intonation and Phrasing*, S.-A. Jun, Ed. Oxford: Oxford University Press, 2014, pp. 6–42.

[15] L. Varga, "The calling contour in Hungarian and English," *Phonology*, vol. 25, no. 3, pp. 469–497, 2008.

[16] I. Abe, "Intonation in Japanese," in *Intonation Systems: A Survey of Twenty Languages*, D. Hirst and A. Di Cristo, Eds. Cambridge, UK: Cambridge University Press, 1998, pp. 363–378.

[17] A. Rialland and S. Robert, "The Intonational System of Wolof," *Linguistics*, vol. 39, no. 375, pp. 893–940, 2001.

[18] S. Luksaneeyanawin, "Intonation in Thai," in *Intonation Systems: A Survey of Twenty Languages*, D. Hirst and A. Di Cristo, Eds. Cambridge, UK: Cambridge University Press, 1998, pp. 376–394.

[19] Q. Li, Y. Chen, and Z. Xiong, "Tianjin Mandarin," *J. Int. Phon. Assoc.*, pp. 1–20, 2017.

[20] J. Zhang and J. Liu, "Tone sandhi and tonal coarticulation in Tianjin Chinese," *Phonetica*, vol. 68, no. 3, pp. 161–191, 2011.

[21] A. Rialland, "Question prosody: an African perspective," in *Tones and tunes*, vol. 1, T. Riad and C. Gussenhoven, Eds. Walter de Gruyter, 2007.

[22] A. Arvaniti and G. Garding, "Dialectal variation in the rising accents of American English," in *Papers in Laboratory Phonology 9*, J. Cole and J. H. Hualde, Eds. Berlin, New York: Mouton de Gruyter, 2007, pp. 547–576.

[23] J. K.-Y. Ma, V. Ciocca, and T. L. Whitehill, "Effect of intonation on Cantonese lexical tones," *J. Acoust. Soc. Am.*, vol. 120, no. 6, pp. 3978–3987, 2006.