Tone realization in Mandarin speech: a large corpus based study of disyllabic words

Yaru Wu, Lori Lamel, Martine Adda-Decker

Tone realization in Mandarin speech: a large corpus based study of disyllabic words

Yaru Wu[1,2]*, Lori Lamel*[1]*, Martine Adda-Decker*[1,2]

[1]Université Paris-Saclay, CNRS, LIMSI, 91400, Orsay, France
[2]Laboratoire de Phonétique et Phonologie (UMR7018, CNRS-Sorbonne Nouvelle), France
yaru.wu@sorbonne-nouvelle.fr, lamel@limsi.fr, madda@limsi.fr

Abstract

This study aims to increase our knowledge about tone realization in disyllabic words in continuous Mandarin speech. Automatic alignments of large speech corpora were carried out to enable the study of potential tone variants, with a special focus on variation factors such as prosodic position and right tonal context. The alignments without tone variants (V0, phonological representation) show that Tone 4 is more frequent in phrase-final position than in other prosodic positions, supporting the "declination line" pattern often observed in speech production. Tone 4 is also the most frequent lexical tone (>50%) in all prosodic positions. Alignments permitting tone variants (V1, phonetic realization) show an increase of Tone 1 in phrase-initial position, compared to V0. Tone realization is observed to be related not only to the prosodic position, but also to the within-word right tonal context. Unsurprisingly, the most notable change in tone realization happens for Tone 3 in the first syllable of disyllabic words when followed by another Tone 3 because of the well-known "tone sandhi rule" in which T3T3 disyllabic words become T2T3. Cross-word right tonal context is found to impact only Tone 3. However, the results in this study show that Tone 3 sandhi rule is more a tendency than an absolute rule.

Index Terms: Mandarin, tone realization, large corpora, continuous speech

1. Introduction

Mandarin Chinese (Standard Chinese) has four lexical tones: high (Tone 1), rising (Tone 2), low/dipping (Tone 3) and falling (Tone 4). Word meaning in a tonal language can depend on the tone, for example, “ma" (/ma/) means “mother" in Tone 1, “linen" in Tone 2, “horse" in Tone 3 and “scold" in Tone 4. Whereas most Mandarin[1] words carry a lexical tone, there is also a “neutral tone" (Tone 0) which occurs in some words in unstressed syllables. Despite the importance of tone, few studies report on how tones are realized in continuous speech as compared to their expected canonical form.

Different phenomena may influence tone realization in fluent Mandarin speech, namely tone sandhi [1, 2], hypoarticulation [3], and declination line [4, 5, 6] in common with many languages. In this study we aim to better understand how lexical tones in Mandarin are realized in continuous speech in light of these factors. We also aim to clarify how tone realization is affected by variation factors such as within- and cross-word right tonal context and prosodic position.

Tones vary according to their phonological contexts (i.e. Tone sandhi). The most prominent sandhi tone phenomenon in Mandarin concerns a consecutive sequence of Tone-3 syllables: when two syllables with Tone 3 follow one another, the tone of the first syllable is realized with Tone 2. For instance, "苦楚" (*kǔchǔ - Tone3Tone3, misery*) is pronounced kúchǔ (Tone2Tone3), according to the Tone-3 sandhi rule. Previous studies often treat the tone sandhi rule as a baseline theory and carry out linguistic analyses accordingly. For instance, [7] found Tone 3 sandhi different from Tone 2 in connected speech and [8] analyzed Tone 3 sandhi as a separate category and observed that overall f0 contour of Tone 3 sandhi changes dramatically as a function of different prosodic contexts. In this study, however, we explore to what extent the well-known Tone-3 sandhi rule [1] is respected[2], and question if in continuous speech the sandhi rule is limited to within-word sequences or if it also applies to cross-word ones. As prosodic position has been found to influence speech production acoustically [9] or articulatorily [10, 11], we also investigate the influence of prosodic position on tone realization. We suppose that, when low/dipping and falling lexical tones are not realized canonically, they will be realized more like high tones in phase-initial position and as low/dipping or falling tones when phrase-final.

2. Method

2.1. Corpus and alignments

Mandarin journalistic speech corpora (distributed by the LDC [12, 13]) of 850 hours were used (e.g., [14, 15]) in this study. The speech data was automatically segmented at the word and phone levels using a speech transcription system [16, 17, 18] in forced alignment mode. Phone-level representations including tones for each word were provided in the pronunciation lexicon. Guided by the orthographic transcriptions, the forced alignment selects the best matching pronunciation among the possible alternatives. Pauses, hesitations and breath were detected automatically by the system. Neutral tone was not distinguished in the pronunciation lexicon and is therefore not able to be selected during alignment. A minimal phone segment duration of 30 ms results from the acoustic modeling technical constraints of a 3-state model and a 10 ms (frame) step [19, 20].

The full corpora were aligned twice using two different pronunciation dictionaries in order to locate changes in tone realization according to our methodology. In the first set of alignments (V0, phonological representation of tones), the base pronunciation lexicon includes almost no variants. Words in the baseline dictionary with pronunciation variants (less than 4% of all word tokens) were excluded from this study. The sec-

[1]In this paper we use Mandarin to refer to Standard Mandarin.

[2]More complicated rules apply to "不" (bù, *not*) and "一" (yī, *one*) in Mandarin. In order to better focus on the third tone sandhi realization, we did not include "不" and "一" in this study.

ond round of alignments (V1, phonetic representation of tones) used an expanded pronunciation dictionary in which tone variants were introduced for the first and last syllable of each word, that is the first and last syllable can be aligned with any of the 4 tones. In this study, we decided to focus on disyllabic words since disyllabic words are the most typical word-type in journalistic Mandarin. By exclusively focusing on disyllabic words, we were able to analyze changes in tone for all syllables in V1 and better compare changes between V0 and V1. For instance, 皮肤(/pi2fu1/, *skin*) was not given additional variants in V0; however, the following variants were allowed for the word in V1: pi1fu1, **pi2fu1**, pi3fu1, pi4fu1, pi1fu2, pi2fu2, pi3fu2, pi4fu2, pi1fu3, pi2fu3, pi3fu3, pi4fu4, pi1fu4, pi2fu4, pi3fu4, pi4fu4). Words with pronunciation variants in V0 such as "啊"(a, *particle showing affirmation*) which can have all 4 tones were also excluded in the analyses with the expanded dictionary. These two alignments, without (V0) and with tone variants (V1), allow us to quantify changes in the realization of tone as a function of different variation factors.

2.2. Tone realization and variation factors

We first analyze tone realization in disyllabic words according to tone nature and prosodic position, and then analyze tone realization in the first syllable of disyllabic words as a function of tone nature, within-word right context (the 2nd (last) syllable of the disyllabic word) and prosodic position. Four prosodic positions (phrase-initial, word-initial, word-final and phrase-final) were analyzed for our analyses on tone realization in disyllabic words. For the first syllable of disyllabic words, the prosodic contexts are either phrase- or word-initial. Phrase-initial and final tone segments concern the first/last syllable of disyllabic words immediately following/preceding a pause (including silence, hesitation and breath) $\geq$ 100 ms. [3] Word-initial tone segments refer to tones in the first syllable of disyllabic words (phrase-initial tone segments excluded) and word-final tone segments refer to tones in the 2nd (last) syllable of disyllabic words (phrase-final tone segments excluded). Table 1 gives detailed information on different levels of the investigated variation factors.

Table 1: *Examined factors for tone realization. # stands for word boundary; ## for phrase boundary. Concerned syllables are circled in the rows attributed to each factor.*

Factors	Levels	Examples	
Tone nature	Tone1	(夫)(/fu1/)妻	(*husband and wife*)
	Tone2	(符)(/fu2/)号	(*symbol*)
	Tone3	(抚)(/fu3/)摸	(*touch*)
	Tone4	(富)(/fu4/)裕	(*prosperous*)
Prosodic position	Phrase-initial	##(富)裕	
	Word-initial	#(富)裕	
	Word-final	富(裕)#	
	Phrase-final	富(裕)##	
Right context	Within-word	幸(福)(/fu2/)	(*happiness*)
	Cross-word	提供#(服)(/fu2/)装	(*provide clothing*)

[3]The 100 ms threshold was empirically determined. Pause durations of 50, 100 and 200 ms gave comparable results on prosodic position.

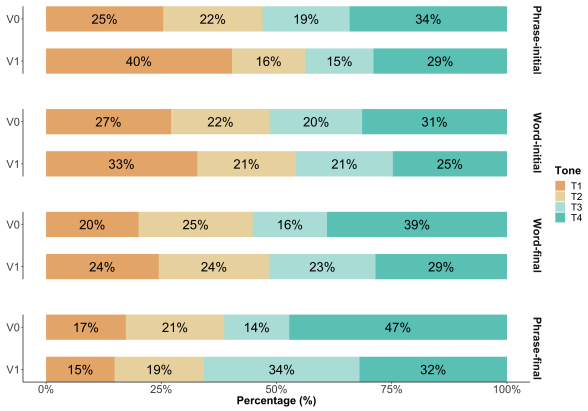

Figure 1: *Tones in disyllabic words grouped by prosodic position in alignments without* (V0) *and with* (V1) *tone variants.*

2.3. Statistical analyses

Generalized linear models (GLM) were used to carry out the statistical analyses in R [21]. Two GLMs were estimated. The first GLM was used to test the effects of tone nature and prosodic position on tone realization (changes vs. unchanged) in disyllabic words. Tone nature (Tone 1, Tone 2, Tone 3 and Tone 4, reference: Tone 1) and prosodic position (phrase-initial, word-initial, word-final and phrase-final; reference: phrase-initial) were the investigated factors. The second model was used to investigate the effect of within-word right context (i.e. 2nd (last) syllable of the disyllabic word, Tone 1, Tone 2, Tone 3 and Tone 4, reference: Tone 1) on tone realization of the first syllable of disyllabic words. Tone nature and prosodic position were included as control variables. Post-hoc tests based on each model were performed to obtain information for each level of the investigated factors.

3. Results

After a first analysis of tone occurrence distribution in Mandarin disyllabic words, observations of tone switch from the canonical lexical form as a function of word-internal and prosodic position are presented.

3.1. Tone distribution in disyllabic words

Figure 1 illustrates the distribution of Mandarin lexical tones in disyllabic words using alignments without (V0) and with (V1) tone variants, for all concerned prosodic positions. With respect to V0 (phonological representation), Tone 4 is the most frequently observed tone in all prosodic positions and Tone 3 is the least frequent tone in all prosodic positions. Interestingly, Tone 4 is more frequently observed in phrase-final position (almost 50%) than at the other three prosodic positions. Tone 1, on the other hand, is more frequent at initial positions (phrase-initial or word-initial) than at final positions (phrase-final or word-final). It is interesting to note, that the higher proportion of Tone 1 (high) tones phrase-initially and higher proportion of Tone 4 (falling) phrase-finally, is consistent with the f0 "declination line" pattern [5] in speech production. For V1 (forced alignment with tone variants, thus phonetic realization of tones), Tone 1 is the most frequent lexical tone in phrase-initial and word-initial positions (instead of Tone 4 in V0). This also supports the relatively higher pitch of "declination line" pattern by having 40% high tone (contour corresponding to

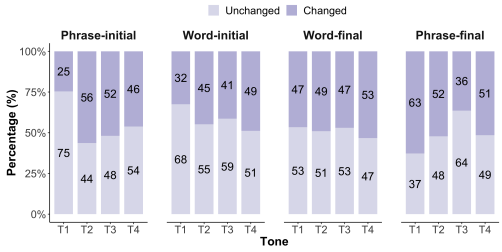

Figure 2: *Tone realization in disyllabic words (changed vs. unchanged from V0 to V1) for each prosodic position.*

Tone 1) in phrase initial position. T3 is seen to increase a lot in phrase-final position. This may be related to the observations reported that creaky voice appears more often at the end of utterances [8] and that low/dipping Tone 3 is frequently tied with creaky voice [22, 23, 24]. Tone 4 has similar occurrence frequencies in all four investigated prosodic positions. This result supports the theory of hypoarticulation on speech variation [3]. The falling contour is preferred in phrase-final position in canonical form of speech production (V0) and a high pitch contour is favored in phrase-initial position in real life speech productions (V1). GLM results show that the probability that a tone is realized as another one is significantly larger for rising Tone 2 (log odds ratio = 0.334915, |Z| = 144.61, $p<0.001$), low/dipping Tone 3 (log odds ratio = 0.215012, |Z| = 86.86, $p<0.001$) and falling Tone 4 (log odds ratio = 0.477769, |Z| = 226.52, $p<0.001$) than for high Tone 1. Tones are significantly less often realized as another tone in word-initial position (log odds ratio = -0.068812, |Z| = 14.49, $p<0.001$). Changes in tone realization are significantly more frequent in word-final (log odds ratio = 0.211480, |Z| = 44.52, $p<0.001$) and phrase-final (log odds ratio = 0.251374, |Z| = 40.08, $p<0.001$) positions, than in phrase-initial position. Post-hoc results based on the GLM model show that all pairwise-comparisons of the four tones are significant ($p<0.001$) as well as all comparisons of prosodic positions ($p<0.001$).

3.2. Tone realization in disyllabic words

Figure 2 shows the percentage of changed (dark purple) and unchanged (light purple) tones obtained in V1 (phonetic realization) segmentations by comparison to the V0 baseline, for each lexical tone and for each prosodic position. Interestingly, according to V1, almost 50% of the tones were not realized canonically. Tone 1 is the least often realized as another of the four lexical tones in phrase-initial ($\sim$25% changes in Tone 1 vs. $>$45% for Tones 2, 3 and 4) and word-initial positions ($\sim$32% vs. $>$40% for Tones 2, 3 and 4). It is realized as another tone mostly ($>$60%) at the end of phrases . This might also be related to the "declination line" pattern mentioned earlier: high tone might become less high and realized otherwise in phrase-final position. Tone 3 realized as other tones the least in phrase-final position than at other prosodic positions. The amount of changes in Tone 2 and Tone 4 are relatively stable among four prosodic positions.

Table 2 shows the realization (V1) of Mandarin lexical tones (V0) grouped by four prosodic positions. The highest rate is always the tone itself, regardless of the prosodic position. Moreover, interesting patterns are found on tone realization at phrase-initial and phrase-final positions. All lexical tones tend to be realized as similar to high tone (T1) when not realized

Table 2: *Tone realization rates for four prosodic positions. The rows correspond to the V0 tones and the columns to the V1 tones. For each prosodic position, the highest rate for each lexical tone is shown in bold and the second in light blue.*

V0 \ V1	Phrase-initial				Word-initial				Word-final				Phrase-final			
	T1	T2	T3	T4	T1	T2	T3	T4	T1	T2	T3	T4	T1	T2	T3	T4
T1	**76**	5	6	13	**67**	11	9	13	**54**	15	12	18	**38**	15	22	25
T2	28	**44**	12	16	18	**54**	17	11	17	**52**	18	13	10	**48**	31	11
T3	19	13	**51**	17	13	14	**59**	14	12	15	**53**	21	8	11	**64**	18
T4	32	8	6	**54**	25	12	12	**51**	20	14	19	**47**	11	10	31	**47**

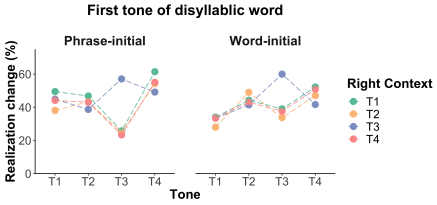

Figure 3: *Change of lexical tone realizations in the first syllable of disyllabic words, grouped by the within-word right context, for each prosodic position.*

canonically in phrase-initial position and they tend to be realized as similar to low/dipping Tone 3 or falling Tone 4 when not realized canonically in phrase-final position (see numbers in light blue for these two prosodic positions).

3.3. Tone realization in the first syllable of disyllabic words

Different from Figure 2 which presented pooled results for the two syllables, Figure 3 focuses on the first syllable of disyllabic words and shows the changes in tone realization for the first syllable of disyllabic words grouped by the within-word right context of the syllable (here, the second/last syllable of the disyllabic word) for the two prosodic positions of interest. Right contexts Tones 1, 2 and 4 have similar tone realization changing patterns for each prosodic position. Tone realization change has a distinguishing pattern while the right context is Tone 3 for both prosodic positions, comparing with the other three right contexts. We notice that the Tone 3 has the most notable tone changing rate while its right context is Tone 3. This is very likely an outcome of the third tone sandhi rule.

Figure 4 illustrates the four lexical tones (V0) as a function of their realizations in V1, grouped by the within-word right context for both concerned prosodic positions. In phrase-initial position, Tone 1 is more frequently switched to Tone 2 than to Tones 3 and 4, when followed by Tones 1, 3 and 4. Tone 1 is similarly distributed for all four within-word right contexts in word-initial position (all $<$ 15%). Tone 2 of the first syllable of disyllabic word is the most frequently replaced by Tone 3 in phrase-initial position, regardless of the right context. However, Tone 2 is the most frequently replaced by Tone 1 in word-initial position. When Tone 3 is followed by Tones 1, 2 and 4, the realization changing rate is distributed similarly among Tones 1, 2 and 4 in both phrase-initial and word-initial positions. With regard to Tone 3 followed by Tone 3, Tone 2 is the most frequently observed switch in phrase-initial position (29%) and Tones 1 and 2 are almost equally frequently observed as Tone 3's replacement in word-initial position (22% for observed Tone 1 and 23% for observed Tone 2). However, it is worth pointing out that $\sim$40% of the first-syllable Tone 3 in "Tone 3+Tone 3" disyllabic words are aligned with "Tone 3+Tone 3" at both phrase-initial and word-initial posi-

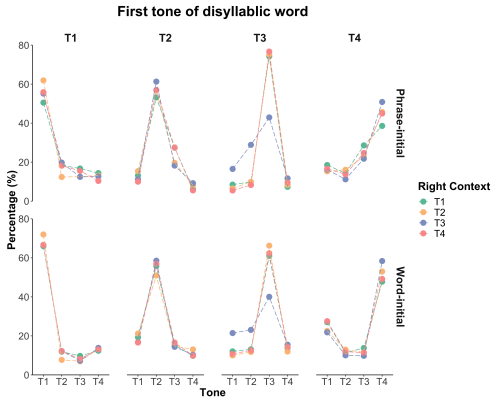

Figure 4: *Percentage of each lexical tone (V0) as a function of its realization (V1) in the first syllable of disyllabic words grouped by within-word right context. Top: Phrase-initial position; Bottom: Word-initial position.*

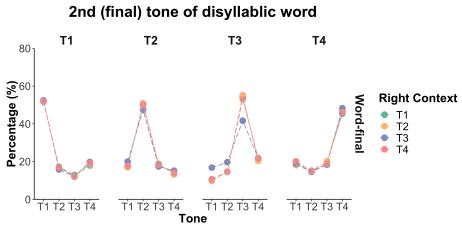

Figure 5: *Percentage of each lexical tone (V0) as a function of its realization (V1) in the second (i.e. last) syllable of disyllabic words by cross-word right context (word-final prosodic position). Each panel shows a lexical tone in V0 and its realization (selected variant) in V1.*

tions. The results on "Tone 3+Tone 3" in disyllabic words seems to suggest the application of Tone 3 sandhi rule for only part of the sequences (29% in phrase-initial position and 23% in word-initial position) and that Tone 3 sandhi rule is more a tendency than an absolute rule. With regard to Tone 4, it is the most frequently replaced by Tone 3 in phrase-initial position and is most frequently replaced by Tone 1 in word-initial position. The probability that tone realizes as another tone decreases significantly for within-word right contexts Tone 2 (log odds ratio = -0.155580, $|Z| = 44.741$, $p<0.001$) and Tone 4 (log odds ratio = -0.051623, $|Z| = 16.355$, $p<0.001$), and increases significantly for within-word right context Tone 3 (log odds ratio = 0.022349, $|Z| = 5.812$, $p<0.001$), with respect to that observed for within-word right context Tone 1. Post-hoc results based on the model show that all pairwise-comparisons of the within-word right contexts are significant ($p<0.001$).

3.4. Tone realization in the last syllable of disyllabic words

Figure 5 illustrates the four lexical tones (V0) as a function of their realization (V1) and their cross-word right context. There is roughly a 20% switch of tone for all four lexical tones. In contrast to Figure 4, Figure 5 shows few differences related to cross-word right context except for when the last syllable of a disyllabic word has Tone 3 and is followed by a word starting with Tone 3. In this case, Tone 3 is realized as Tone 1, Tone 2 and Tone 4 each about 20% of the time. Note that the remaining 40% are realized as Tone 3 even though the cross-word right context is also Tone 3.

Differences in the effect of within-word and cross-word right context suggest the importance of word boundaries on tone realization in spoken Mandarin, as far as disyllabic words are concerned. Our results suggest that tone sandhi on "Tone3Tone3" sequences is activated both within and across words.

4. Summary and discussion

In this study, tone distribution and tone realization in disyllabic words were analyzed according to different factors, namely tone nature and prosodic position, in large corpora of fluently journalistic Mandarin. Tone realization of the first syllable of disyllabic words and the second (last) syllable were also investigated separately as a function of tone nature, the right tonal context and the prosodic position.

Analyses of selected tone variants (V1 compared to V0) using automatic speech-text alignments provides supplementary information on how tone segments are realized in fluent speech. In disyllabic words of Mandarin continuous speech, Tone 4 is frequent in the phonological representation in phrase-final position. Nevertheless, the high pitch contour (Tone 1) is favored in phrase-initial position in real-life production. The results support the "declination line" pattern in Mandarin, which is observed in speech production of many languages.

High Tone 1 less often stays in phrase-final position than at the three other prosodic positions. Tone 3 is the tone the least susceptible to change in phrase-final position. Analyses of tone realization in the first syllable of disyllabic word suggest that tone realization is influenced by prosodic position and within-word right context. Tone 2 has very different patterns in phrase-initial and word-initial positions (c.f. Fig. 4). It changes more often to Tone 3 than to Tone 1 or 4 in phrase-initial position, but changes more frequently to Tone 1 than to Tones 2 or 3 in word-initial position. Tone 3 is frequently realized as other tones when it is followed by a second Tone 3, regardless of the prosodic position. This suggests that tone sandhi rule (i.e. Tone 3 is realized Tone 2 when followed by Tone 3) is more a tendency than an absolute rule. Tone 4 changes more to T3 than to other tones in phrase-initial position and it changes mostly to T1 in word-initial position, regardless of the right context. Few differences are found related to the influence of cross-word right context on the realization of the second (last) syllable of disyllabic words, except for when the last syllable of a disyllabic word has Tone 3 and is followed by a word starting with Tone 3. In this case, Tone 3 is realized as Tone 3 the most and Tones 1, 2 and 4 each about 20% of the time.

These results allow us to better understand tone realization in real-life continuous speech (corpus of 850 hours). As a tonal language, continuous Mandarin still leaves traces of declination line commonly observed in speech production and variation factors, such as prosodic position and neighboring contexts, still play an important role in tone realization. This study also investigated a method using forced alignments with tone variants to assess tone realizations. Indeed, forced alignment with variants has already been used to analyze various linguistic phenomena, such as schwa or liaison in French [25]. Nonetheless, to our knowledge, this is the first time this method was applied to study the realization of tone variants.

5. Acknowledgements

This work was supported by the by the French Investissements d'Avenir - Labex EFL program (ANR-10-LABX-0083).

6. References

[1] C. Shih, "Mandarin third tone sandhi and prosodic structure," *Linguistic Models*, vol. 20, pp. 81–124, 1997.

[2] J. Zhang and Y. Lai, "Testing the role of phonetic knowledge in mandarin tone sandhi," *Phonology*, pp. 153–201, 2010.

[3] B. Lindblom, "Explaining phonetic variation: A sketch of the h&h theory," in *Speech production and speech modelling*. Springer, 1990, pp. 403–439.

[4] A. Cohen, R. Collier *et al.*, "Declination: construct or intrinsic feature of speech pitch?" *Phonetica*, vol. 39, no. 4-5, pp. 254–273, 1982.

[5] D. R. Ladd, "Declination.: a review and some hypotheses," *Phonology*, vol. 1, pp. 53–74, 1984.

[6] J. Yuan and M. Liberman, "F0 declination in english and mandarin broadcast news speech," *Speech Communication*, vol. 65, pp. 67–74, 2014.

[7] Y. Chen and J. Yuan, "A corpus study of the 3rd tone sandhi in standard chinese," in *Eighth Annual Conference of the International Speech Communication Association*, 2007.

[8] J. Kuang, "Creaky voice as a function of tonal categories and prosodic boundaries." in *INTERSPEECH*, 2017, pp. 3216–3220.

[9] J. Pierrehumbert and D. Talkin, "Lenition of/h/and glottal stop," *Papers in laboratory phonology II: Gesture, segment, prosody*, pp. 90–117, 1992.

[10] C. Fougeron and P. A. Keating, "Demarcating prosodic groups with articulation," *The Journal of the Acoustical Society of America*, vol. 97, no. 5, pp. 3384–3384, 1995.

[11] C. Fougeron and P. A. Keating, "Articulatory strengthening in prosodic domain-initial position," *UCLA Working Papers in Phonetics*, vol. 97, pp. 61–87, 1996.

[12] S. M. Strassel, C. Cieri, A. Cole, D. DiPersio, M. Liberman, X. Ma, M. Maamouri, and K. Maeda, "Integrated linguistic resources for language exploitation technologies." in *LREC*, 2006, pp. 185–190.

[13] A. Morris, B. Antonishek, X. Li, and S. Strassel, *HAVIC MED Progress Test–Videos, Metadata and Annotation*. Linguistic Data Consortium, University of Pennsylvania, 2019.

[14] K. Walker, C. Caruso, K. Maeda, D. DiPersio, and S. Strassel, "Gale phase 3 chinese broadcast news speech ldc2015s13," *Web Download. Philadelphia: Linguistic Data Consortium*, 2015.

[15] M. Glenn, H. Lee, S. Strassel, and K. Maeda, "Gale phase 3 chinese broadcast news transcripts ldc2015t25," *Web Download. Philadelphia: Linguistic Data Consortium*, 2015.

[16] J.-L. Gauvain, L. Lamel, and G. Adda, "The limsi broadcast news transcription system," *Speech communication*, vol. 37, no. 1-2, pp. 89–108, 2002.

[17] M. Adda-Decker and L. Lamel, "The use of lexica in automatic speech recognition," in *Lexicon Development for Speech and Language Processing*. Springer, 2000, pp. 235–266.

[18] S. Huang, J. Liu, X. Wu, L. Wu, Y. Yan, and Z. Qin, "Mandarin broadcast news speech (hub4-ne)," *Linguistic Data Consortium*, 1998.

[19] L. Lamel, J.-L. Gauvain, V. B. Le, I. Oparin, and S. Meng, "Improved models for mandarin speech-to-text transcription," in *2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2011, pp. 4660–4663.

[20] L. Chen, L. Lamel, G. Adda, and J.-L. Gauvain, "Broadcast news transcription in mandarin," in *Sixth International Conference on Spoken Language Processing*, 2000.

[21] R Development Core Team, *R: A La nguage and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, 2019. [Online]. Available: https://www.R-project.org/

[22] A. Belotel-Grenié and M. Grenié, "Phonation types analysis in standard chinese," in *Third International Conference on Spoken Language Processing*, 1994.

[23] A. Belotel-Grenié and M. Grenié, "The creaky voice phonation and the organisation of chinese discourse," in *International symposium on tonal aspects of languages: With emphasis on tone languages*, 2004.

[24] R.-X. Yang, "The phonation factor in the categorical perception of mandarin tones." in *ICPhS*, 2011, pp. 2204–2207.

[25] M. Adda-Decker, P. Boula de Mareüil, and L. Lamel, "Pronunciation variants in french: schwa & liaison," in *Proceedings of the XIVth International Congress of Phonetic Sciences*, 1999, pp. 2239–2242.