

TONE REPRESENTATION AND TONE PROCESSING IN SHANGHAINESE

Jianjing Kuang, Jia Tian

University of Pennsylvania
kuangj@sas.upenn.edu, jiatian@sas.upenn.edu

ABSTRACT

In Shanghainese, it has been assumed that contrastive sandhi tones at the word level are derived from base tones at the syllable level by a spreading rule. Does this mean that speakers have to access syllable tones before computing sandhi patterns? By examining the production latencies of a speeded production experiment, we show that Shanghainese speakers produce sandhi tones fairly fast, but are much slower in producing syllable tones. This result suggests that Shanghainese speakers in fact access sandhi tones directly without referring to base tones. By contrast, as the control group, Mandarin speakers produce words and syllables equally fast. In light of these findings, we will discuss how Shanghainese speakers represent and process the tonal contrasts at both syllable and word levels.

Keywords: tone sandhi, speech production, tone representation, speech processing, Shanghainese

1. INTRODUCTION

Shanghainese, like other Chinese Wu dialects, is known to have contrastive tonal patterns at both the syllable level and the word level. It is usually assumed that, similar to Mandarin, underlying tones of Shanghainese are contrastive at the syllable level [4, 16, 17, 19, 20], and that the sandhi patterns at the word level are the derived forms, because the overall contour of sandhi in Shanghainese is essentially the spreading of the tonal contour of the first syllable of the polysyllabic word, as illustrated below:

1. Shanghainese tone sandhi rules

- (a) T1-X: 53-X → 55-31
- (b) T2-X: 34-X → 33-44
- (c) T3-X: 13-X → 22-44
- (d) T4-X: 55-X → 33-44 (checked tone)
- (e) T5-X: 12-X → 11-13 (checked tone)

However, it is questionable whether these generative phonological rules reflect how tones are stored in memory and accessed during language production [10, 11]. To date, most experimental studies on tone and tone sandhi have mainly focused on Man-

darin [2, 3, 7–11, 21, 22], but very little is known about the representation and processing of tone in Shanghainese. Nonetheless, Zhang and Meng [23] found that the tone sandhi rules of Shanghainese can be productively applied to nonce words in wug tests by native speakers. Therefore, it seems possible that Shanghainese speakers productively derive the sandhi patterns from the base tones in tone production.

This paper aims to further explore the representation and processing of syllable tones and sandhi tones in Shanghainese. If we assume sandhi tones are derived from base tones in processing, it is predicted that in production, native speakers of Shanghainese should access the base tones before applying the tone sandhi rule. In other words, producing syllable tones should be faster or at least as fast as sandhi tones.

This study adopts the speeded production task, where speakers are asked to produce the prompt stimuli as fast and accurately as possible, using production latencies to probe the real-time processing of different levels of tonal contrasts. If syllable tones are indeed easier to access than sandhi tones, producing syllable tones should be faster or at least as fast as sandhi tones; however, if syllable tones are harder to access than sandhi tones, the production latencies for syllable tones should be longer. Moreover, as shown by Yan [18], frequency has a significant effect on tone sandhi production, and thus we considered several frequency conditions (summarized in Table 1) in the experiment. High-frequency syllable tones should be relatively easier to recognize and produce than low-frequency ones. Lastly, since sound change has been reported in Shanghainese – whereby the phonation contrast for the syllable tones has been merged for younger speakers [5, 14] – we also control for speakers' age and gender.

2. METHODS

2.1. Participants

60 participants who grew up and still lived in the urban area of Shanghai were recruited. Although

they all identified themselves as native speakers of Shanghaiese, we found that some younger participants were not able to speak Shanghaiese fluently, probably due to insufficient exposure and usage. We therefore excluded these participants from our analyses. Data from a total of 48 participants were analyzed. They were sorted into four age \times gender groups, namely old female (age range 44-78, mean age 60), old male (age range 38-73, mean age 60), young female (age range 17-30, mean age 22), and young male (age range 17-34, mean age 21).

We also recruited 30 Mandarin speakers as the control group (7 male, 23 female, age range 20-50, mean age 28 in 2018). All Mandarin participants were recruited in the U.S.

2.2. Stimuli

81 monosyllabic stimuli and 67 disyllabic stimuli were presented to Shanghaiese speakers in Chinese characters. Each monosyllabic stimulus has one of the three unchecked tones (i.e., T1, T2, and T3, or 53, 34, and 13 in Chao numbers). Each disyllabic stimulus is a disyllabic word which undergoes tone sandhi in Shanghaiese. The first syllable of the word stimulus has one of the three unchecked tones. Given that there is no existing frequency corpus of Shanghaiese, a native Shanghaiese linguist estimated the usage frequency of all Shanghaiese stimuli. A summary of the stimuli is given in Table 1.

The same stimuli were presented to Mandarin participants.

Table 1: Summary of the stimuli.

Shanghaiese: single syllable		
Group	#	Frequency
H	33	High in isolation and word-initially
M	20	Low in isolation; high word-initially
L	28	Low in isolation and word-initially
Shanghaiese: disyllabic sandhi word		
Group	#	Frequency
High	34	High
Low	33	Low

2.3. Procedure

Participants were tested individually in a quiet room. Participants were seated about 60 cm from the computer screen. A practice session was conducted prior to the actual experiment. For each trial of the experiment, a target stimulus was displayed on the computer screen in Chinese characters. When the stimulus appeared on the screen, a 100-ms warning beep

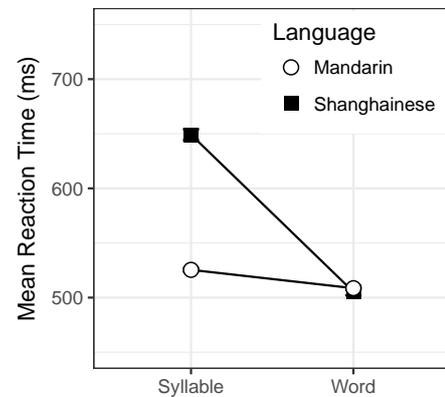
was played. The participants were instructed to read each stimulus aloud as quickly and accurately as possible. All word stimuli were presented after all syllable stimuli. Shanghaiese and Mandarin participants were instructed to read the stimuli in Shanghaiese and Mandarin, respectively. Response latencies or reaction times were measured as the time lag between the onset of the warning beep and the onset of the participants' response.

3. RESULTS

Statistical analyses based on linear mixed-effects modeling were conducted in R [12], using the `lmer` function in the `lme4` package [1]. The `lmerTest` package [6] was used to compute *p*-values. Statistical analyses were conducted on log-transformed reaction times because inspection of response latencies revealed a skewed distribution. While statistical analyses were conducted on log-transformed data, raw reaction times were used in figures for the ease of interpretation.

3.1. Effect of tone type

Figure 1: Reaction time for syllables and words, in Shanghaiese and Mandarin.



A linear mixed-effects model was built to examine the effect of syllable vs. word-level tones on reaction time in both Shanghaiese and Mandarin. Tone type (syllable vs. word), language, and their interaction were included as fixed effect predictors. Note that all these words undergo tone sandhi in Shanghaiese, so for Shanghaiese, the comparison is actually between syllable and sandhi tones. Random effects included random intercepts for participants and items, a by-participant random slope for condition, and a by-item random slope for language. Log-transformed reaction time was the response variable. All fixed effect predictors were dummy coded. Word

was the reference category for tone type. Mandarin was the reference category for language. Table 2 summarizes the model output. Mean reaction times are shown in Fig. 1.

Table 2: The output of the linear mixed-effects model of reaction time for syllable and word stimuli.

	Estimate	SE	t	p
(Intercept)	6.222	0.032	195.759	<0.001
ToneSyl	0.033	0.027	1.208	0.23
LangSH	-0.051	0.042	-1.235	0.22
ToneSyl:LangSH	0.192	0.037	5.182	<0.001

The effect of tone type was not significant ($p > 0.05$), indicating that Mandarin participants produced syllable and word stimuli equally fast. The effect of language was also not significant ($p > 0.05$), indicating that Shanghainese and Mandarin participants produced word stimuli equally fast. The interaction between condition and language, however, was highly significant ($p < 0.001$), indicating that Shanghainese participants produced syllable stimuli significantly slower than word stimuli. As shown in Fig. 1, response time for syllable tones was about 150 ms slower than sandhi tones for Shanghainese speakers.

3.2. Frequency effect

To further examine the frequency effect on the response time, a linear mixed-effects model was built with frequency as the fixed effect. Random effects included random intercepts for participants and items, and a by-participant random slope for frequency. Log-transformed reaction time was the response variable. Frequency was coded using backward difference coding. In backward difference coding, the mean of the dependent variable for one level of the categorical variable is compared to the mean of the dependent variable for the prior adjacent level. In our model, the first comparison compared the mean of M-frequency syllables (low frequency as stand-alone monosyllabic words but high frequency as the word-initial syllable) with the mean of L-frequency syllables (cannot stand alone and also low frequency as the word-initial syllable). The second comparison compared the mean of H-frequency (high frequency in isolation) syllables with the mean of M-frequency (high frequency word-initially) syllables, etc. Table 3 summarizes the model output. Frequency effects on mean reaction times are visually represented in Fig. 2 (c.f. section 3.3 for age \times gender effects).

As expected, frequency significantly affected the response times. High frequency syllables were

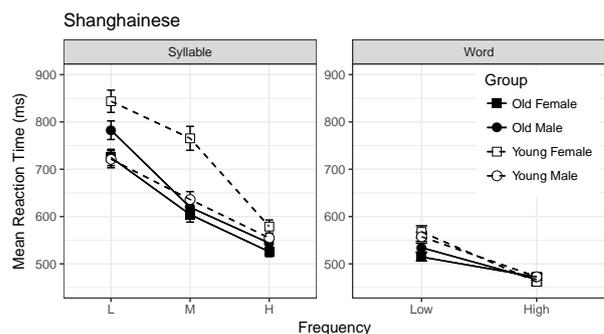
produced significantly faster than Low frequency words. For syllable tones, H-frequency syllables (high frequency as stand-alone monosyllabic words) are produced significantly faster than M-frequency syllables (low frequency as stand-alone words, but high frequency as the word-initial syllables), and L-frequency syllables are the slowest in production.

Moreover, there is no significant difference in response time between H-frequency syllables, the fastest situation for syllable tones, and Low-frequency words, the slower type of word tones. Or, in other words, syllable tones can be at most as fast as the low-frequency sandhi tones. This result further suggests that it is much more difficult for Shanghainese speakers to retrieve the forms of syllable tones.

Table 3: The output of the linear mixed-effects model of reaction time for stimuli of different frequencies. H_Syl: high-frequency syllables both in isolation and word-initially; M_Syl: syllables high-frequency word-initially, but low-frequency in isolation; L_Syl: syllables low-frequency in either position; H_Wd: high-frequency words; L_Wd: low-frequency words.

	Estimate	SE	t	p
(Intercept)	6.311	0.030	208.135	<0.001
M_Syl v. L_Syl	-0.150	0.035	-4.260	<0.001
H_Syl v. M_Syl	-0.159	0.034	-4.664	<0.001
L_Wd v. H_Syl	-0.008	0.030	-0.281	0.779
H_Wd v. L_Wd	-0.136	0.029	-4.635	<0.001

Figure 2: Reaction times at different prosodic levels, for speakers in different age and gender groups.



3.3. Age and gender effects

This section further examines if response time is modulated by age and gender. Separate linear mixed-effects models were built for syllable and sandhi stimuli. In these models, log-transformed reaction time was the dependent variable. Frequency,

age \times gender group, and their interaction were fixed effects. Random effects included random intercepts for participants and items, and a by-participant random slope for frequency. Frequency and age \times gender groups were dummy coded such that high-frequency syllables were set as the reference category for syllable stimuli, high-frequency words were set as the reference category for word stimuli, and young male was the reference category for age \times gender group. Table 4 and 5 summarizes the model output. Due to limited space, only significant effects are reported here. Mean reaction times are visually represented in Fig. 2.

Table 4: The output of the linear mixed-effects model of reaction time for word stimuli produced by different age and gender groups.

	Estimate	SE	t	p
(Intercept)	6.122	0.064	95.939	< 0.001
FreqLow	0.150	0.037	4.072	< 0.001

The word model (Table4) did not show significant effects of age \times gender group, nor were age \times gender group by frequency interactions found, indicating that reaction times of all participants patterned similarly.

Table 5: The output of the linear mixed-effects model of reaction time for syllable stimuli produced by different age and gender groups.

	Estimate	SE	t	p
(Intercept)	6.277	0.073	86.424	< 0.001
FreqL	0.231	0.059	3.936	< 0.001
GrpYF:FreqM	0.138	0.045	3.109	0.003
GrpYF:FreqL	0.146	0.070	2.074	0.044

The syllable model (Table5) did not show significant effects of age \times gender group, indicating that all participants produced H-frequency syllables equally fast. However, there were significant interactions of GroupYoungFemale \times FrequencyM and GroupYoungFemale \times FrequencyL, indicating that young female participants produced M- and L-frequency syllables (syllables that rarely appear in isolation) much more slowly than other age \times gender groups.

4. DISCUSSION AND CONCLUSION

A speeded production experiment provides new evidence regarding the nature of tone representation and tone processing in Shanghainese. Significantly longer response latencies were found for syllable stimuli among speakers of all age and gender groups (by a huge effect size of 150 ms on average). By contrast, Mandarin participants who produced the same

stimuli exhibit no difference in response latency between syllables and words. Therefore, our results strongly suggest that it is unlikely that Shanghainese speakers access the syllable tones before they compute the sandhi tones. Speakers should be able to access sandhi tones directly.

However, at the same time, we found that all speakers were able to produce the correct syllable tones even though it took them much longer to retrieve the forms. This suggests that speakers still have the knowledge of the syllable tones. Moreover, productivity studies [23] suggested that Shanghainese speakers clearly know the link between syllable tones and sandhi tones, as they can apply this link to nonce words. How do Shanghainese speakers know the mapping relationship between the sandhi tones and syllable tones, if the sandhi forms are not computed from the syllable forms?

We propose that it is possible that Shanghainese speakers store multiple variants (i.e., monosyllabic form, disyllabic form, trisyllabic form, and so on) for each tonal category of their language. For example, Tone 1 of Shanghainese have both 53 (monosyllabic) and 55+31 (disyllabic) in the mental representation. Each variant is mapped to the underlying/abstract tonal category (e.g., Tone 1), while there is no direct link between the monosyllabic form (e.g., 53) and the sandhi form (e.g., 55+31). In a wug test, the mapping between a syllable tone (e.g, 53) and a sandhi form (e.g., 55+31) is realized through their mapping with the abstract tonal category. In a speeded production task, the reaction time is determined by the frequency of the variant.

The fact that Shanghainese speakers generally have difficulties in producing the syllable tones is not very surprising. Tone sandhi is an important part of the word-formation process in Shanghainese. Since over 70% of words in modern Chinese (across all language varieties) are composed of two or more syllables [13, 15], Shanghainese speakers hear the sandhi tones substantially more often than the syllable tones. Monosyllabic verbs are essentially the only contexts where monosyllabic tones can be produced. This can explain why the accessibility of tonal patterns is strongly affected by frequency.

Lastly, effects of age and gender were also found in our study. While there was no reaction time difference for word stimuli and H frequency syllable stimuli across speakers in different age and gender groups, younger female speakers produced syllables which less frequently occur in isolation significantly slower. The processing of lower frequency syllable tones is becoming increasingly difficult for Shanghainese speakers.

5. REFERENCES

- [1] Bates, D., Mächler, M., Bolker, B., Walker, S. 2015. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* 67(1), 1–48.
- [2] Chen, Y., Shen, R., Schiller, N. O. 2011. Representation of allophonic tone sandhi variants. *Proceedings of Psycholinguistics Representation of Tone. Satellite Workshop to ICPHS, Hongkong* 38–41.
- [3] Chien, Y.-F., Sereno, J. A., Zhang, J. 2016. Priming the representation of mandarin tone 3 sandhi words. *Language, Cognition and Neuroscience* 31(2), 179–189.
- [4] Duanmu, S. 1990. *A formal study of syllable, tone, stress and domain in Chinese languages*. PhD thesis MIT.
- [5] Gao, J. 2016. Sociolinguistic motivations in sound change: on-going loss of low tone breathy voice in Shanghai Chinese. *Papers in Historical Phonology* 1, 166–186.
- [6] Kuznetsova, A., Brockhoff, P. B., Christensen, R. H. B. 2017. lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software* 82(13), 1–26.
- [7] Li, X., Chen, Y. 2015. Representation and processing of lexical tone and tonal variants: Evidence from the mismatch negativity. *PLOS ONE* 10(12), 1–24.
- [8] Nixon, J. S., Chen, Y., Schiller, N. O. 2015. Multi-level processing of phonetic variants in speech production and visual word processing: evidence from mandarin lexical tones. *Language, Cognition and Neuroscience* 30(5), 491–505.
- [9] Politzer-Ahles, S., Schluter, K., Wu, K., Almeida, D. 2016. Asymmetries in the perception of Mandarin tones: Evidence from mismatch negativity. *Journal of Experimental Psychology: Human Perception and Performance* 42(10), 1547–1570.
- [10] Politzer-Ahles, S., Zhang, J. 2012. The Role of Tone Sandhi in Speech Production: Evidence for Phonological Parsing. *TAL2012-Third International Symposium on Tonal Aspects of Languages*.
- [11] Politzer-Ahles, S., Zhang, J. in press. Evidence for the role of tone sandhi in Mandarin speech production. *Journal of Chinese Linguistics*.
- [12] R Core Team, 2018. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing Vienna, Austria.
- [13] Sun, H., Sun, D., Huang, J., Li, D., Xing, H. 1996. The description on the corpus system of modern Chinese studies. In: Luo, Z., Yuan, S., (eds), *Studies of Chinese and Chinese character in the computer era*. Beijing: Tsinghua University Publisher.
- [14] Tian, J., Kuang, J. 2016. Revisiting the Register Contrast in Shanghai Chinese. *Tonal Aspects of Languages 2016* 147–151.
- [15] Wang, H. 1986. *Modern Chinese frequency dictionary*. Beijing: Beijing Language Institute Press.
- [16] Xu, B., Tang, Z. 1988. *Shanghai shiqu fangyan zhi (A grammar of inner city Shanghai)*. Shanghai: Shanghai Education Press.
- [17] Xu, Y. 1997. Contextual tonal variations in Mandarin. *Journal of Phonetics* 25(1), 61–83.
- [18] Yan, H. 2018. *The Nature of Variation in Tone Sandhi Patterns of Shanghai and Wuxi Wu*. Peking University Press and Springer Nature Singapore Pte Ltd.
- [19] Yip, M. 2002. *Tone*. Cambridge University Press.
- [20] Zee, E., Maddieson, I. 1980. Tone and tone sandhi in Shanghai: phonetic evidence and phonological analysis. *Glossa* 14, 45–88.
- [21] Zhang, C., Xia, Q., Peng, G. 2015. Mandarin third tone sandhi requires more effortful phonological encoding in speech production: Evidence from an erp study. *Journal of Neurolinguistics* 33, 149 – 162. Language, Brain, and Gene: The Chinese Context.
- [22] Zhang, J., Lai, Y. 2010. Testing the role of phonetic knowledge in Mandarin tone sandhi. *Phonology* 27(1), 153–201.
- [23] Zhang, J., Meng, Y. 2016. Structure-dependent tone sandhi in real and nonce disyllables in Shanghai Wu. *Journal of Phonetics* 54, 169–201.