TAL2018, Sixth International Symposium on Tonal Aspects of Languages
18-20 June 2018, Berlin, Germany

Production and Perception of Rising Tone Sandhi in Mizo

Wendy Lalhminghlui[1] *and Priyankoo Sarmah*[1,2]

[1]Department of Humanities and Social Sciences
[2]Center for Linguistic Science and Technology
Indian Institute of Technology Guwahati, India
{wendy,priyankoo}@iitg.ac.in

Abstract

Mizo, a Tibeto-Burman language of the Kuki-Chin group is primarily spoken in the north-east Indian state of Mizoram in Northeast India. Mizo has four lexical tones, namely, high (H), rising (R) , falling (F) and low (L). Mizo tones are mostly dynamic, except the H tone which is a static tone. Previous researches have reported that the rising tone in Mizo changes into low tone when it is followed by either high tone or falling tone which is regarded as rising tone sandhi. The present study analyzes the production and perception of rising tone sandhi. The production data of rising tone sandhi is carried out by comparing the F0 contours of the derived low tone of rising tone sandhi in trisyllabic phrases with the citation form of low tone and low tone in phrases. Results have shown that the F0 contour of rising tone sandhi and the canonical low tone in Mizo are different in terms of F0 contour. The result of perception study in the form of identification test has shown that the speakers of Mizo could distinguish the low tone derived out of rising tone sandhi from the canonical low tone which indicates that tone sandhi in Mizo is perceptually categorical.

Index Terms: tones, tone sandhi, Mizo, Tibeto-Burman, Kuki-Chin

1. Introduction

Tone sandhi is a phonological process that changes the canonical tone of a syllable into a different one due to the occurence of a neighbouring tone [1]. Four characteristics of tone sandhi as reported in previous literature are summarized below [2, 3, 4]. Firstly, the occurrence of tone sandhi is consistent across speakers. Secondly, there is a categorical shift from the canonical tone to another tone or even to a new tone different from the other canonical tones. Thirdly, the phenomenon of tone sandhi should affect the entire tone, and lastly, the resulted in tone sandhi is perceptually not different from the canonical tone. The phenomenon is mostly attributed to Chinese languages as Chinese languages are rich in tonal inventory with complex contours which trigger tone sandhi in certain contexts. Also, it may be noted that most of the literature on tone sandhi are based on the Chinese languages. For example, the often cited example of tone sandhi is of Mandarin Chinese where the first tone in a sequence of two third tones (T3) in the language, changes into a second tone (T2).

It is also stated that Chinese tone sandhi patterns fall into two types: right-dominant ones and left-dominant ones. The right-dominant sandhi preserves the underlying tone on the final syllable in a sandhi domain and hence changes the tone of the non-final syllables. It also tends to involve local or paradigmatic tone change as in Standard Chinese. On the other hand, left-dominant sandhi preserves tone on the first syllable in a sandhi domain and tend to involve the extension of the initial tone rightward as in Changzhou (Northern Wu) [1]. Acoustic study of six tone sandhis attested in disyllabic words in Tianjin Chinese reported that only one tone sandhi (T1 + T1) is acoustically neutralized with T2 while the other sandhis (T3 + T3, T4 + T1, T3 + T2 and T3 + T4) are non neutralized, and the T4 + T4 sandhi is obsolete [5]. Another previous study also reported that two of the six tone sandhis of Tianjin Chinese where T3 becomes T1 when followed by either T2 or T4 (Two Half-Third Sandhi) are not entirely neutralizing, retaining the rising feature of T3 [6].

A perception study conducted on Tianjin Chinese shows that the participants could differentiate between T1 and sandhi T3 with an accuracy of above 85% [6]. Acoustic studies on the Third tone sandhi in Standard Chinese have reported that the Third tone did not become completely neutralized to T2 [7, 8, 9, 5]. At the same time, perception studies reported that the native speakers of Standard Chinese were unable to distinguish between T2 and the T3 sandhi [10, 11]. On the other hand, in case of Taiwanese Mandarin, acoustic studies have shown conflicting results. While T2 and T3 sandhi are found to be acoustically different in some studies [11], other studies show no significant difference between the two tones [12]. Hence, it is noticed in previous works that the resultant tone in a tone sandhi may or may not be acoustically similar to any of the phonological tones in the language. At the same time, studies have also revealed that acoustic differences in the sandhi tone may not be perceived categorically by native speakers of the tone language.

In the current study, we report the findings of production and perception experiments conducted on the rising tone sandhi in Mizo. Mizo is a Tibeto-Burman language of the Kuki-Chin subfamily, spoken in the North-East Indian state of Mizoram and its neighboring areas. It is reported that Mizo is also spoken in parts of Myanmar and Bangladesh. Mizo has four tones, namely, high (H), low (L), rising (R) and falling (F) [13, 14, 15, 16]. The averaged z-score normalized pitch contours of the four Mizo tones, extracted from 47082 iterations produced by 19 speakers, are presented in Figure 1. As noticed in Figure 1, while the high tone is static in nature, the other three tones (L, R and F) are dynamic [15, 17]. The high tone starts with a high F0 value, falls slightly at the beginning but stays around that level throughout the syllable. The rising tone starts at a low F0, then falls slightly before rising throughout the remainder of the syllable till it reaches an offset similar to the offset of the high tone. The low tone starts at an F0 value slightly above the onset of the rising tone and falls till it reaches its minimum before rising slightly towards the end. The falling tone starts with a high F0 value similar to that of the high tone and falls gradually to an offset just above the low tone.

It is observed that Mizo rising tones undergo tone sandhi when followed by a high or a falling tone [13, 14, 16]. The

10.21437/TAL.2018-23

Table 1: *Two sets of tonal minimal pairs in Mizo.*

Mizo	Meaning	Category	Tone
t^hay	famous	adjective	High
	gone	verb	Falling
	greasy	adjective	Low
	trap	noun	Rising
vai	chaff	noun	High
	wave	verb	Falling
	dazzle	verb	Low
	search	verb	Rising

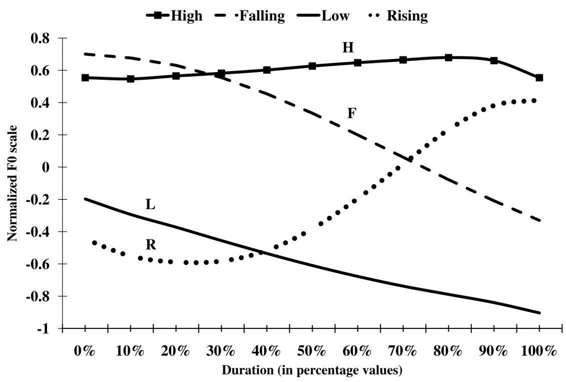

Figure 1: *Normalized F0 contours of Mizo tones*

rising tone in Mizo is reported to be neutralized to a low tone as a result of the tone sandhi process [13, 14]. Apart from the acoustic investigation, we also report parts of an attempt to automatically identify the neutralized rising tone in Mizo from a low tone using statistical modeling [18]. Finally, in this work, we report a perception experiment conducted on Mizo speakers where the speakers were tasked to identify a word undergoing tone sandhi, resulting in a low tone from a phonological low tone in the language.

The remainder of the paper is organized as follows: Section 2 provides a production test and the results of tone sandhi in Mizo. Section 3 describes the procedure and results of the perception study on Mizo tone sandhi. Section 4 reports the automatic identification of rising tone sandhi from the canonical low tone in Mizo. Lastly, Section 5 discusses and concludes the paper.

2. Production Experiment

In order to investigate the nature of rising tone sandhi in Mizo, a production test was conducted. The data collected in the production experiment were subjected to acoustic analysis and attempts were made to see the difference in F0 contours of a low tone emerging due to tone sandhi (RTS) and a phonological low tone (L).

2.1. Method

2.1.1. Materials

The data used for the production study consisted of Mizo meaningful trisyllabic phrases comprised of three monosyllabic words with all the possible combinations of the four Mizo lexical tones resulting in (4^3=) 64 distinct tone combinations. Five sentences each for the 64 tone combination, repeated three times by each speaker resulting in $18,240$ phrases with $54,720$ tokens tone tokens (19 speakers x 64 tonal combinations x 5 trisyllabic phrases x 3 monosyllables x 3 repetitions). However, only $49,957$ tokens were considered to be of suitable for F0 extraction.

Table 2: *Distribution of low tones (L) in phrases and rising tones undergoing tone sandhi (RTS)*

Tone	Initial	Medial	Final	Total
RTS	1751	1124	NA	2875
L	3703	3770	3878	11351

2.1.2. Subjects

For the production study, 19 native speakers of Mizo (10 male, 9 female), born and brought up in Mizoram, participated in the study. None reported any speech and hearing disabilities. All of them spoke Mizo fluently and could understand English. They were asked to read the sets of trisyllables from a list provided to them. Most of the times tones are not marked in Mizo orthography, hence, in order to eliminate ambiguity and elicit correct tones, English translations corresponding the trisyllables were provided. Each participant took about 60 minutes to read the wordlist. They were provided a remuneration of about \$6 for their participation.

2.1.3. F0 Analysis

The collected trisyllables were annotated using Praat [19] for tones. The annotations were done by native speakers of Mizo. Once the annotations were completed, F0 values were extracted for each tone with a Praat script at every 10% of the total duration of the pitch contour. After extracting raw F0 values, they were normalized for speaker effects using z-score normalization. The z-score normalization is considered to be the best method for speaker normalization [20]. This normalization is performed with the Equation 1.

$$x = \frac{x - \mu}{\sigma} \tag{1}$$

In Equation 1, μ is the mean F0, and σ is the standard deviation of the F0 values considered for mean F0. After z-score normalization, subjective F0 values in Hertz or Mel are replaced with speaker-independent height and shape of the original pitch contour. The pitch contours represented in z-score are plotted for visual examination.

2.2. Results

The z-score normalized F0 contour for the rising tone sandhi is presented on Figure 2. In the same figure, the rising and low tones, as shown in Figure 1 are added for comparison. As seen in the figure, the rising tone that had undergone tone sandhi is phonetically quite distinct from the canonical low tone. At the same time, the RTS contour does not have any similarity to the original rising tone either.

2.3. Discussion

Considering the F0 contours, we can assume that while it is claimed that a rising tone sandhi is neutralized to a low tone [14, 13], this low tone is quite distinct from the canonical low tone in Mizo. However, this phonetic difference does not necessarily warrant distinct perception of the low tone emerging of

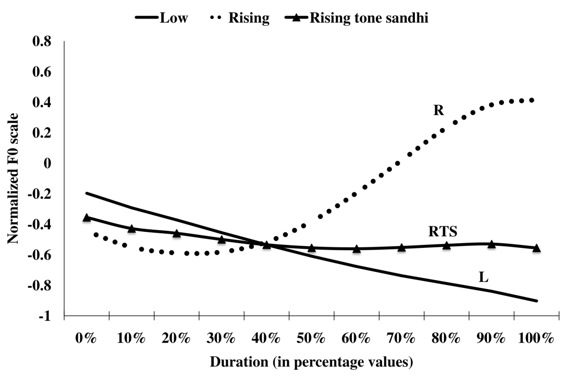

Figure 2: *F0 contour of rising tone sandhi in contrast with that of a low and a rising tone.*

tone sandhi, as in the Chinese languages. Hence, a follow up perception study conducted to resolve this issue is reported in the following section.

3. Perception Experiment

A perception study was carried out to see if Mizo speakers could identify the canonical low tone and the rising tone sandhi in Mizo. Words and phrases containing segmentally homophonous word pairs specified with RTS and L were provided to Mizo speakers for identification. The methodology of the two tests are described in the following subsections.

3.1. Method

3.1.1. Materials

The perception study contained stimuli consisting of phrases with low tones and rising tones in Mizo followed by high and falling tones. As shown in Table 3, five phrases were chosen containing low tone monosyllables followed by high tone monosyllables (set L-H) and five rising tone monosyllables followed by high tone monosyllables (set S-H). The S-H set is expected to trigger rising tone sandhi. As shown in Table 4 another set was made with five low tone monosyllables followed by falling tone monosyllables (set L-F) and five rising tone monosyllables followed by falling tone monosyllables (set S-F). Again the S-F set is expected to trigger rising tone sandhi. In total, 20 unique meaningful Mizo disyllabic phrases were constructed. As seen in Table 3 and Table 4, each row of the tables represent a segmentally homophonous phrases that differs in only their tonal specification. A female native speaker of Mizo recorded the stimuli in a sound attenuated recording booth. The target stimuli, i.e. the first monosyllables of each phrases, which are phonologically low and the derived low tone from the sandhi process are then extracted from the phrases. These initial syllables from the phrases are used as stimuli in the perception test.

3.1.2. Subjects and procedure

For the perception test, 11 Mizo native speakers (6 male, 5 female) participated. All reported normal speech and hearing abilities and the age range of the participants was from 24 to 35 years. All of them were born and brought up in Mizoram and were at least bilingual, with Mizo as their first language and English as the second language.

Table 3: *Stimuli set with a high tone in the second syllable*

Mizo phrases	L-H (Low + High)	S-H (Rising + High)
lei var	white bridge	white soil
ral hma	front of the otherside	before it gets over
tho hma	wake up face	face of fly
kir rei	takes time to return	curly for a long time
kawr thar	new drainage	new shirt

Table 4: *Stimuli set with a falling tone in the second syllable.*

Mizo phrases	L-F (Low + Falling)	S-F (Rising + Falling)
tlan chhuak	saved (redeem + out)	escape (run + out)
vai nghal	become dark soon	search soon
bul tawk	very short	very near
kang lut	arrive with fire	entered while lifting
eng en	what do you see	looking at the light

A multiple forced choice experiment was designed using Praat, where only the initial syllables of the 20 unique stimuli, as shown in Table 3 and Table 4, were randomly played. In total there were 10 repetitions for each syllable, resulting in 200 tokens. The experiment was designed with three response buttons displayed on the screen of the computer where the first two response buttons indicated the possible meaning of the syllable a participant heard. A third button was marked ”Both”, in case the participants categorized the stimuli into either of the categories.

The test window displayed a prompt to click anywhere on the screen once the participant is ready to take the perception test. When the participant clicked on the screen, the experiment console randomly played one of the 200 phrases (see Subsection 3.1.1) and prompted the participant to give a response from the three choices provided to her.

The participants sat in a sound attenuated booth and listened to the stimuli with a pair of headphones. The stimuli were presented on a laptop PC running Windows XP with the experiment running on Praat [19]. The participants clicked on the screen of the computer specifying the most suitable response as per the syllable he or she heard. The participants were allowed to replay sounds for a maximum of two times. No trial test was allowed. After a test with one subject is over, her responses were exported to a spreadsheet for further analysis. Average time taken to complete the task was about 60 minutes.

3.1.3. Analysis

The responses of the subjects were analyzed for their correctness. Percentage of correctness for all the 11 participants were averaged and plotted as a bar graph for further analysis.

3.2. Results

The results of the perception test demonstrated strong categorical perception of the stimuli by the native speakers of Mizo. The fact that there were only 2.4% of ”Both” responses in the perception test, it can be assumed that Mizo speakers are aware of the phonological differences between a tone sandhi and a low tone.

As far as individual results for rising tone sandhi and low tones are concerned, the speakers do marginally better in perceiving the sandhi tones. As far as their performance in the lexical identification task is concerned, they are able to identify tone sandhi correctly more than 65% of the times. However, for

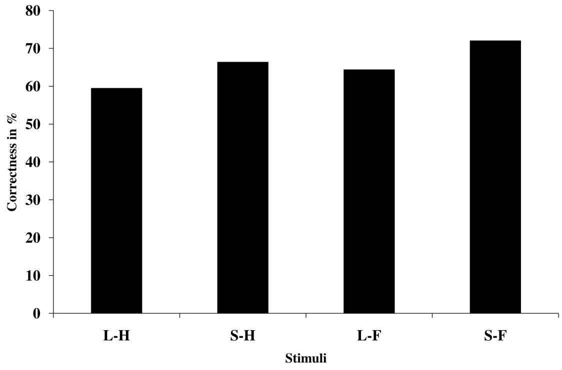

Figure 3: *Correct responses of four contrastive stimuli sets.*

low tones, they are able to identify them more than 60% of the times.

3.3. Discussion

It can be concluded from the perception test that the rising tone sandhi is perceived categorically by Mizo speakers. Also, the F0 contour differences observed in the production of tone sandhi is probably phonologized by the Mizo speakers.

4. Evidence from automatic tone recognition

In a related work by the same group of authors as in this paper [18], an attempt is being made to model the Mizo tones using Deep Neural Network (DNN) [21] for automatic tone recognition. The DNN with 3 hidden layers was trained for automatic Mizo tone recognition using various features identified for the purpose. A learning rate of 0.01 was used to train the network for 100 epochs with a batch size of 10. Out of 37,982 iterations for Mizo tones 26,616 were used for training and 11,366 were used for testing. While the overall tone identification accuracy will not be discussed in this work, relevant parts related to the recognition of tone sandhi will be discussed in the results section.

The DNN based tone recognition performs fairly well in automatically identifying Mizo tones [18]. While the low tones are identified with an accuracy of 81.4%, the rising tone sandhi is identified with an accuracy of 79.4%. In other words, automatic tone modeling is being able to build robust model of the tone sandhi, distinctly from the low tone. More importantly, error analysis of the wrongly identified low tones and tone sandhi revealed that the system does not confuse any low tone as tone sandhi. In case of tone sandhi, it recognizes only 4.6% of the tone sandhi stimuli as the low tone, which again, is not a significantly high amount of confusion and reiterates our claim that the canonical low tone and the 'low tone' emerging due to rising tone sandhi are phonetically different.

5. General discussion

The current study provides some insight into the tone sandhi phenomenon of a hitherto less studied language of the Tibeto-Burman language family, spoken in North East India. The study contradicts the findings reported in previous studies that Mizo rising tone sandhi is merely a toneme change phenomenon, changing a rising tone to a low tone when followed by a high or a falling tone. In this work we demonstrate that the rising tone that had undergone tone sandhi, has not simply become a low tone in the language. It has in fact resulted in a new tone that has distinct phonetic characteristics. These phonetic cues associated with the sandhi is so robust that Mizo speakers are able to categorically identify the tone sandhi from a low tone. Our observations here fit into the increasing body of literature on prominent tone sandhi phenomena, where acoustic differences between the tone sandhi and the canonical toneme are observed. For example, it is demonstrated for Mandarin Chinese that second tone in the language and the second tone emerging as a result of tone sandhi have subtle acoustic differences [22]. The same study concludes that the third tone sandhi is not a mere change of one toneme to another toneme in Standard Chinese, as suggested by phonological accounts.

As in the case of many Chinese languages, rising tone sandhi in Mizo is a right-dominant sandhi where the tone of a syllable before the final syllable undergoes tone change resulted in contour simplification of rising tone. Hence, the rising tone sandhi in Mizo is dissimililatory. This kind of dissimilation pattern in tone sandhi is widely attested in right-dominant sandhi systems, like in Tianjin Chinese [5]. That is, even though a canonical rising tone in Mizo reaches a high target, when the same tone is followed by a high or a falling tone, both with high initiation F0, there is a preference for lowering the preceding F0 contour, rather than assimilating. This also indicates that there are strong physiological underpinnings associated with the rising tone sandhi phenomenon, possibly due to the need of gestural adjustments before producing a tone initiated with high F0.

Although the initiation of the F0 contour of rising tone sandhi is the same as of a rising tone (see Figure 2, the pitch contour of the low tone derived from sandhi is not neutralized with the canonical low tone. The pitch contour of sandhi begins with the same F0 value of the canonical rising tone and undergoes a gradual fall till 50% of the total duration and meets the canonical low tone and rising tone at about 40% of the total duration. Then it deviates from this point with a level pitch contour resulted in a little higher F0 value than the low tone with a slight fall at the end. The 'effect' of tone sandhi seems to be starting from about 40% of the total duration. This is also consistent with the previous findings reported for Mizo where contextual effects of the following tone begins from about 50% of the total duration of the tone [16].

Hence, the present study finds that the low tone derived from a tone sandhi has different pitch contour realization than the phonologically low tone in Mizo. This observation is not only confirmed by acoustic analysis, but also by automatic tone recognition based on statistical modeling and perception test conducted on native speakers of Mizo.

6. Acknowledgements

This work is part of the project on Acoustic and Tonal Features Based Analysis of Mizo funded by the Department of Electronics & Information Technology (DeitY), Ministry of Communication & Information Technology (MC&IT), Government of India. The authors would like to thank Siami Ngente of the Center for Linguistic Science and Technology, IIT Guwahati for her contribution in setting up the perception test. The authors also would like to thank the Mizo speakers who participated in this study.

7. References

[1] J. Zhang, "Tones, tonal phonology, and tone sandhi," *The handbook of Chinese linguistics*, pp. 443–464, 2014.

[2] X. S. Shen, "On tone sandhi and tonal coarticulation," *Acta Linguistica Hafniensia*, vol. 25, no. 1, pp. 83–94, 1992.

[3] M. Y. Chen, *Tone sandhi: Patterns across Chinese dialects*. Cambridge University Press, 2000, vol. 92.

[4] S. Chen, C. Wiltshire, and L. Bin, "An updated typology of tonal coarticulation properties," *Taiwan Journal of Linguistics*, 2017.

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

[6] M. Qiu-wu and J. Yuan, "Two new third tone sandhi rules in Tianjin dialect: A critical re-analysis," *Journal of Tianjin Normal University (Social Science)*, vol. 1, p. 011, 2006.

[7] E. Zee, "A spectrographic investigation of Mandarin tone sandhi," *UCLA working papers in phonetics*, vol. 49, no. 9, 1980.

[8] X. S. Shen, "Tonal coarticulation in Mandarin," *J. Phonet.*, vol. 18, pp. 281–295, 1990.

[9] Y. Xu, "Contextual tonal variation in Mandarin Chinese," 1993.

[10] W. S. Wang and K. P. Li, "Tone 3 in Pekinese," *Journal of Speech, Language, and Hearing Research*, vol. 10, no. 3, pp. 629–636, 1967.

[11] S.-H. Peng, "Lexical versus 'phonological' representations of Mandarin sandhi tones," *Papers in laboratory phonology V: Acquisition and the lexicon*, pp. 152–167, 2000.

[12] J. Myers and J. Tsay, "Investigating the phonetics of Mandarin tone sandhi," *Taiwan Journal of Linguistics*, vol. 1, no. 1, pp. 29–68, 2003.

[13] A. Weidert, *Componential analysis of Lushai phonology*. John Benjamins Publishing, 1975, vol. 2.

[14] L. Chhangte, "Mizo syntax," Ph.D. dissertation, University of Oregon, 1993.

[15] L. Fanai, "Some aspects of the lexical phonology of Mizo and English: An autosegmental approach," Ph.D. dissertation, CIEFL, Hyderabad, India, 1992.

[16] P. Sarmah, L. Dihingia, and W. Lalhminghlui, "Contextual variation of tones in Mizo," in *Interspeech*, 2015, pp. 983–986.

[17] P. Sarmah and C. R. Wiltshire, "A preliminary acoustic study of Mizo vowels and tones," *Journal of the Acoustical Society of India*, vol. 37, no. 3, pp. 121–129, 2010.

[18] P. Gogoi, A. Dey, W. Lalhminghlui, P. Sarmah, S. R. M. Prasanna, R. Sinha, and S. R. Nirmala, "Deep Neural Network based Mizo tone recognition using acoustic features," in prep.

[19] P. Boersma, "Praat, a system for doing phonetics by computer." *Glot International*, vol. 5, no. 9/10, pp. 341–345, 2001.

[20] P. J. Rose, "Considerations on the normalization of the fundamental frequency of linguistic tone," *Speech Communication*, vol. 10, no. 3, pp. 229–247, 1991.

[21] M. Chen, Z. Yang, and W. Liu, "Deep neural networks for Mandarin tone recognition," in *2014 International Joint Conference on Neural Networks (IJCNN)*, July 2014, pp. 1154–1158.

[22] 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.