# Metrical and Tonal Prominence in Swatou

*Suki S.Y. Yiu*

University of Arizona, USA
University of Hong Kong, Hong Kong
syutji@hku.hk

## Abstract

This paper explores a metrical interpretation of prominence in Swatou (Teochew, Southern Min), a six-tone language with bi-directional sandhi spoken in the Chaoshan region of China, Hong Kong, and Southeast Asia. It attempts to provide phonetic evidence in terms of duration and intensity to see whether systematic patterns can be observed based on predictions of the Iambic-Trochaic Law. Acoustic data elicited from two native speakers of Swatou was processed with Praat, then submitted to Linear Mixed Effect Regression Models with R to generate estimates for disyllabic units at both positions (i.e. the first and second syllables) of two sandhi types (i.e. anterior and posterior sandhi) for determining whether the durational and intensity contrasts of each estimate point are significantly different. Model comparison was also made to examine whether sandhi type predicts duration and intensity contrasts. Results showed significant duration and intensity contrasts for anterior and posterior sandhi respectively, despite a tendency for higher intensity on the first syllable for both sandhi types. The results largely support a metrical interpretation of prominence in complex tone languages, in addition to the pitch-related cue for tonal prominence in terms of tone sandhi.

**Index Terms**: metrical prominence, tonal prominence, Iambic-Trochaic Law, tone and stress, Swatou, Teochew

## 1. Introduction

Prominence in tone languages is often defined by tonal stability and complexity [1]. Apart from the tone-related behaviors or pitch-related cues, we know little about the role of other phonetic correlates which are typically considered as manifestations of prominence in metrical theory [2, 3] in signaling rhythm in tone languages. This paper examines the metrical prominence of tone languages via duration and intensity. The metrical view hypothesizes that duration and intensity cues are present on top of tone sandhi to indicate prominence, and hence, to determine whether a language has iambic or trochaic-like rhythm. This paper provides first hand data of Swatou, commonly known as one of the Teochew dialects (Southern Min, Chaoshan), to illustrate the discussion.

Earlier works on Swatou, like other varieties of Teochew), generally focus on grammar description, comparison of dialectal sound variation, tone, and so on, instead of the metrical structure [4-7]. While providing valuable resources for studying the grammar and phonological system of Swatou, many like [8] noted that when acoustic data is available for closer examination, current transcriptions might subject to revision, and hence, the relevant analyses. This paper presents measurements on duration and intensity to examine the metrical prominence of Swatou.

Audio-recordings from two native speakers of Swatou were recorded in a sound booth, processed with Praat [9], and submitted to Linear Mixed Effect Regression (LMER) Models with R [10] to generate statistic estimates for determining whether the durational and intensity contrasts of estimate points within disyllabic units of both sandhi types are significant. Comparison of LMER models was also made to see whether sandhi type predicts duration and intensity contrasts. Based on the statistic comparisons of phonetic measurements, this paper discusses the metrical prominence of Swatou with reference to hypotheses and predictions of the Iambic-Trochaic Law [2, 3].

## 2. Tonal and metrical prominence

### 2.1. Tone inventory and bi-directional sandhi

As listed in Table 1, there are six lexical tones in Swatou for sonorant-final syllables: H(igh), M(id) and L(ow), HM, MH and LM, with two checked tones as shorter versions of the high and low tones not listed here.

Table 1: *Tones in Swatou.*

| Citation tone [4] | Tone letter |
|---|---|
| 33 | M |
| 53 | HM |
| 213 | LM |
| 55 | H |
| 35 | MH |
| 11 | L |

Each of the six tones undergoes sandhi to another tone depending on whether the tone is on the first or second syllable of a disyllabic unit, regardless of the tonal content of the trigger. Anterior sandhi refers to tonal changes observed on the first syllable whereas posterior sandhi on the second. For example, the disyllabic unit HM.MH becomes LM.MH after anterior sandhi but HM.L after posterior sandhi. As in Table 2, the anterior position holds four different tones (M, L, HM and MH), and the posterior position holds only two (L and LM).

Table 2: *Anterior and posterior sandhi [4].*

| Tone letter | Anterior sandhi | Posterior sandhi |
|---|---|---|
| M | M | L |
| HM | MH | LM |
| LM | H | L |
| H | L | L |
| MH | L | L |
| L | L | L |

With bi-directional tone sandhi [8] within one language, not only can we make within sandhi type comparison between the first and second syllables, but also across sandhi type comparison between the citation syllables in both types, between the sandhi syllables in both types, and between the sandhi syllables and their citation counterparts in the other type. This offers a six-way comparison in terms of duration and intensity to unveil the relative metrical prominence of syllables in different positions within a disyllabic unit of both sandhi types.

### 2.2. Hypotheses and predictions

According to the Iambic-Trochaic Law [2, 3], iambic and trochaic rhythmic types tend to adopt different acoustic cues to signal prominence: Iambs use length for final prominence while trochees use intensity for initial prominence. This predicts that the second syllable is longer than the first syllable in iambic units, whereas the first syllable is louder than the second syllable in trochaic units. In the case of tone sandhi, based on the assumption of the relationship between tonal stability and prosodic prominence [1], syllables with citation tones are considered to be more prominent than those with sandhi tones. Since anterior sandhi changes the tone on the left, it is expected to pattern with the iambic rhythm. As the posterior sandhi changes the tone on the right, it is expected to behave like trochees.

The metrical view hypothesizes that duration and intensity cues are present on top of tone sandhi to indicate prominence. It predicts a durational asymmetry between the two sandhi types, whereas the null hypothesis predicts no pattern between the sandhi types in terms of duration and intensity. As presented in the below spaces, hypothesis 1 reflects direct consequences of the Iambic-Trochaic Law [2, 3], and hypotheses 2 and 3 can be indirectly inferred from it.

Hypothesis 1 tests the durational asymmetry where initial prominence is signalled by intensity and final prominence is signalled by duration. It predicts that anterior sandhi units adopt duration and posterior sandhi units adopt intensity to signal prominence.

Hypothesis 2 states that sandhi syllables are less prominent than citation syllables in the same syllable position. It tests the relative prominence of the syllables in sandhi positions to those in citation positions. It predicts that syllables carrying sandhi tones are less prominent than those carrying citation tones.

Hypothesis 3 states that the type of prominence that the prosodic heads of each rhythmic type adopts depends on the rhythmic type that they pattern with. It compares the relative prominence between the citation syllables of both sandhi types, and between the sandhi syllables of both sandhi types. It predicts that citation tone syllables of anterior sandhi are longer but quieter than that of posterior sandhi while sandhi syllables of anterior sandhi are shorter but louder than that of posterior sandhi.

## 3. Methodology

### 3.1. Subjects

Audio-recordings were obtained from two native speakers of Swatou in their 20s. They were both born and raised in the Swatou speaking area of the Chaoshan region, China.

### 3.2. Stimuli

The data was elicited with a wordlist of seventy-two disyllabic units, covering all possible tonal combinations for both sandhi types. For example, citation tone H was paired with each of the six citation tones, so as to obtain data of all six sandhi forms when in combination with another syllable, resulting in six combinations for each tone, and thirty-six combinations for all six tones. After repeating the same process for the two sandhi types, this provides seventy-two combinations. The wordlist was randomized, and produced five times by each speaker. This gives us a total of 1440 tokens of data (6 tones x 6 tones x 2 syllables x 2 sandhi types x 5 iterations x 2 speakers).

### 3.3. Data collection and analysis

The data was recorded in a sound booth using a Shure earset microphone, attaching to a Marantz portable digital recorder at a sampling frequency of 44100Hz, monitored by the experimenter with a Sennheiser headphone set. The subjects were presented with the randomised words on a laptop one at a time, and were asked to read the words in a natural way. Subjects were also asked to fill in a language background questionnaire after the recording session so as to provide information on the specific Teochew dialect that they speak and potential interference from other dialects or languages to the one under investigation.

Segmentation of the recordings, and extraction of syllable duration and maximal intensity were performed in Praat [9]. The duration and intensity measurements were compiled as a main data file to submit to LMER models in R [10]. lmerTest [11] was adopted in order to generate estimates for disyllabic units at both positions (i.e. the first and second syllables) of two sandhi types (i.e. anterior and posterior sandhi) for determining whether the durational and intensity contrasts of each estimate point are significantly different from each other in the expected direction under the Iambic-Trochaic Law [2, 3]. *P*-values were calculated with the Satterthwaite approximation, one of the most conservative methods of *p*-value calculation for LMER models having a small number of subjects [12]. Model comparison was also made to see whether sandhi type predicts the contrasts.

## 4. Results

Section 4.1 displays the duration and intensity measurements of both speakers by token for a general description. Section 4.2 shows the LMER results based on those duration and intensity measurements. It reports the estimates for disyllabic units at both positions of two sandhi types for determining whether the durational and intensity contrasts of each estimate point are significantly different. It also reports the results of model comparison for whether sandhi type predicts the contrasts in duration and intensity.

### 4.1. Duration and intensity measurements

Figures 1 and 2 show the actual tokens of duration (ms) and intensity (dB) measurements respectively. In Figure 1, the left panel displays duration over a disyllabic domain (1st and 2nd syllables) of anterior sandhi, while the right panel displays that of posterior sandhi. Data of Speakers 1 and 2 is shown in empty dots and solid dots respectively.

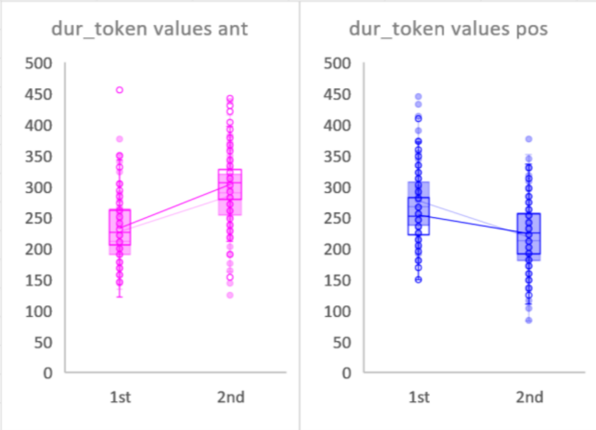


Figure 1: *Duration (ms) of disyllabic units of anterior (left panel) and posterior (right panel) sandhi for Speaker 1(empty dots) and Speaker 2 (solid dots).*

For duration, anterior sandhi and posterior sandhi display an opposite pattern with the position of the syllable. Anterior sandhi has a longer second syllable than first syllable, while posterior sandhi has a longer first syllable than second syllable. Speakers 1 and 2 share a very similar use of duration contrast for both sandhi types in terms of the absolute values, the means, and the spread of the tokens.

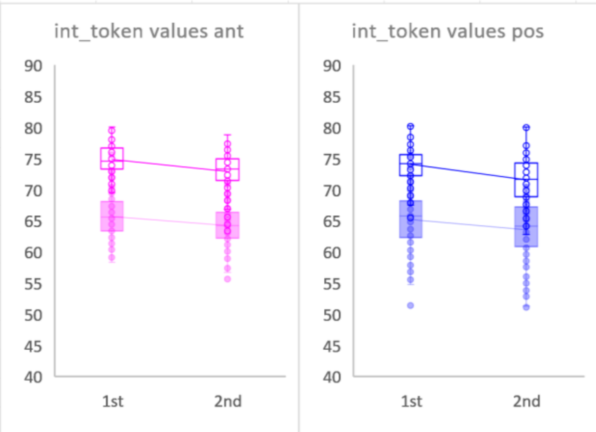


Figure 2: *Intensity (dB) of disyllabic units of anterior (left panel) and posterior (right panel) sandhi for Speaker 1(empty dots) and Speaker 2 (solid dots).*

For intensity, even though the first syllable is louder for both sandhi types, the negative slope for posterior sandhi is steeper than that of the anterior sandhi, indicating that the intensity contrast is bigger for posterior sandhi than anterior sandhi. Although the intensity difference may seem small in absolute values, studies in music production show that the Just Noticeable Difference is around 1 dB for humans [13-15], mediated by frequency and initial loudness of the sound. Again, both speakers share a similar pattern for intensity despite Speaker 1 being louder than Speaker 2.

### 4.2. LMER results

Figures 3 and 4 display the plots of LMER estimates of duration and intensity respectively. Anterior sandhi (ant) is in pink and posterior sandhi (pos) is in blue. The estimates and *p*-values are listed in Tables 3 and 4 following their corresponding figures. For the convenience of description, I will refer to the first and second syllables of anterior sandhi as A1 and A2; and those of posterior sandhi as P1 and P2.

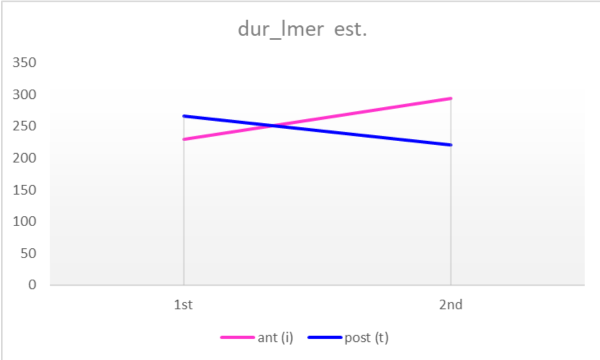


Figure 3: *Plot of LMER estimates of duration (ms) for anterior (pink) and posterior sandhi (blue).*

Table 3: *Significance values of LMER estimates for duration (ms).*

| | est. 1 | est. 2 | *p*-value | sig. |
|---|---|---|---|---|
| A1 vs. A2 | 229.733 | 294.414 | < 2E-16 | *** |
| P1 vs. P2 | 266.041 | 221.064 | < 2E-16 | *** |
| A1 vs. P1 | 229.733 | 266.041 | 7.37E-11 | *** |
| A2 vs. P2 | 294.414 | 221.064 | < 2E-16 | *** |
| A1 vs. P2 | 229.733 | 221.064 | 1.04E-01 | n.s. |
| P1 vs. A2 | 266.041 | 294.414 | 2.20E-07 | *** |

For duration, taking the data into LMER models which contained a fixed effect of sandhi type and syllable position, and random effects of speaker, word (as a subset of speaker) and iteration, results show that A2 is longer than A1; P1 is longer than P2; P1 is longer than A1; A2 is longer than P2; P2 is not significantly longer than A1; A2 is longer than P1. Overall, five out of the six comparisons in duration are highly significant in the way that a metrical view would directly or indirectly predict. We will discuss the results in Section 5.

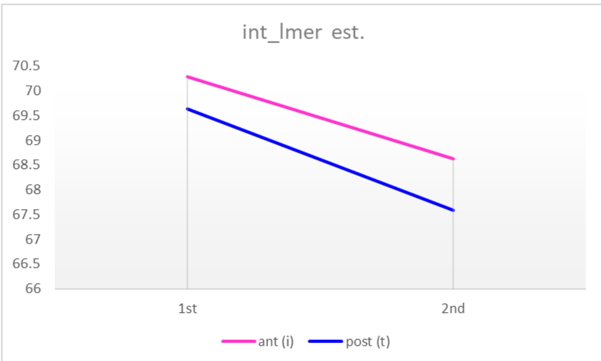


Figure 4: *Plot of LMER estimates of intensity (dB) for anterior (pink) and posterior sandhi (blue).*

Table 4: *Significance values of LMER estimates for intensity (dB).*

| | est. 1 | est. 2 | *p*-value | sig. |
|---|---|---|---|---|
| A1 vs. A2 | 70.2868 | 68.62 | 7.55E-15 | *** |
| P1 vs. P2 | 69.6401 | 67.5854 | < 2E-16 | *** |
| A1 vs. P1 | 70.2868 | 69.6401 | 0.0667 | . |
| A2 vs. P2 | 68.62 | 67.5854 | 0.00355 | ** |
| A1 vs. P2 | 70.2868 | 67.5854 | 5.25E-13 | *** |
| P1 vs. A2 | 69.6401 | 68.62 | 0.00403 | ** |

For intensity, results of the same LMER model, which contained a fixed effect of sandhi type and syllable position, and random effects of speaker, word (as a subset of speaker) and iteration, show that A1 is louder than A2; P1 is louder than P2; A1 is not significantly louder than P1; A2 is louder than P2; A1 is louder than P2; P1 is louder than A2. Five out of the six comparisons in intensity are significant in the way that a metrical view would predict.

Apart from the above, additional models and ANOVA tests were run to to check whether sandhi type predicts the contrasts in duration and intensity. The models contained a fixed effect of syllable position in a disyllabic unit, and random effects of speaker, word (as a subset of speaker) and iteration. These models were used to compare with models where sandhi type interacts with syllable position under ANOVA tests. Model comparisons confirmed that sandhi type significantly predicts level of duration over a disyllabic domain ($\chi^2$ (2) = 456.89, p < 2.2E-16). Posterior sandhi results in an increase of duration by 36.308 ms ± 5.281 ms (standard error), relative to anterior sandhi. Sandhi type also significantly predicts level of intensiy ($\chi^2$ (2) = 8.5813, p = 0.0137). Posterior sandhi results in a decrease of intensity by 0.6467 dB ± 0.3497 dB (standard error), relative to anterior sandhi.

## 5. Discussion

First, based on the within-sandhi type comparisons, there is a duration asymmetry between the two sandhi types. Anterior sandhi prefers duration contrast to signal final prominence whereas posterior sandhi has a stronger preference for intensity contrast to signal initial prominence. Anterior and posterior sandhi map neatly with iambic and trochaic rhythmic types respectively.

Second, results of sandhi-citation comparisons show that A2 is more prominent than P2 in terms of duration. This agrees with the metrical view that prosodic heads are more prominent than non-heads. However, the same comparison between head and non-head is found to be not significant for intensity since P1 is not significantly different from A1.

Third, sandhi-sandhi comparisons show that A1 is louder than P2, indicating that the non-head of posterior sandhi is quieter than that of anterior sandhi, even though the non-head of anterior sandhi (A1) is not significantly longer that of posterior sandhi (P2). Citation-citation comparisons show that A2 is longer but quieter than P1, indicating that the head of anterior sandhi is longer than that of posterior sandhi and the head of posterior sandhi is louder than that of anterior sandhi. Since iambs prefer uneven feet, the head needs to be longer and the non-head needs to be shorter to meet the rhythmic ideal. Also, given trochees' tendency to adopt intensity to signal prominence, the head needs to be louder and the non-head needs to be quieter. This again is what the metrical view would predict.

Thus far, the metrical view of prominence based on duration and intensity contrasts largely match the tonal view based on tonal behaviors.

Given the above discussion on how anterior and posterior sandhi map with iambic and trochaic rhythmic types following the predictions based on the Iambic-Trochaic Law [2, 3], one way to interpret the model comparison results which confirm sandhi type predicting duration and intensity contrasts is that the metrical alternative of understanding prominence largely goes along with the tonal view of defining prominence, and that tonal prominence defined by sandhi types has already embedded metrical factors in terms of duration and intensity. While the tonal behaviors are much better documented and discussed than the metrical manifestations of prominence in the literature, this paper offers statistical analyses based on phonetic evidence for a metrical alternative of understanding rhythm in tone languages.

## 6. Conclusions

This paper briefly illustrates a metrical alternative to diagnose prominence motivated by the Iambic-Trochaic Law [2, 3] via phonetic measurements most relevant to defining the rhythmic type, that is duration and intensity. Results show that prominence indicated by duration and intensity contrasts largely behaves the way as expected from the metrical perspective, supporting a metrical interpretation of prominence in Swatou, in addition to the pitch-related cue in terms of tone sandhi.

Regarding the stimuli, while the disyllabic units are controlled for tonal combinations, the segmental make-up is largely imperfect. This paper has presented the basic analysis focusing on reporting the results of the main effects. Further analyses using models with different statistical means for reducing such 'noise' should also be employed. Future analysis will look into the morphosyntactic and semantic structure of the two sandhi types, and longer prosodic domains.

## 7. Acknowledgements

I sincerely thank Diana Archangeli and Stephen Matthews for many valuable discussions. My gratitude also goes to Natasha Warner, Jonathan Yip, John Culnun, and Seongjin Park for specific discussions on statistics and visualization of data. Many thanks to Elaine Lau and Weiling Lu for identifying subjects, and to the Swatou subjects for providing the recordings and sharing their native knowledge with me. Comments by two anonymous reviewers of TAL2018, and feedback from mfm25 and Chula-ISSSEAL2017 on two related studies helped improve the content. All errors are my own.

This project has received generous funding by a Fulbright Fellowship and a Universitas 21 Collaborative Research Award.

## 8. References

[1] M. Chen, *Tone Sandhi: Patterns across Chinese Dialects*. Cambridge: Cambridge University Press, 2000.

[2] B. Hayes, "Iambic and trochaic rhythm in stress rules," *Proceedings of the Annual Meeting, Berkeley Linguistics Society*, vol. 11, pp. 429–446, 1985.

[3] B. Hayes, *Metrical Stress Theory: Principles and Case Studies*. Chicago: University of Chicago Press, 1995.

[4] L. Lin and X. Chen, *Guangdong Min fangyan yuyin yanjiu [On the phonology of Min dialects in Guangdong]*. Shantou, Shantou University Press, 1996.
[5] Z. Bao, *On the Nature of Tone*. PhD dissertation, MIT, 1990.
[6] J. M. Cai, *Chaozhou Fangyan Cihui [A Dictionary of Chaozhou Dialect]*. Institute of Chinese Culture, Chinese University of Hong Kong, 1991.
[7] Z. Bao, “Teochew Tone Sandhi and the Representation of Tone,” *Pan-Asiatic Linguistics: Proceedings of the Fourth International Symposium on Languages and Linguistics*, vol. 1, pp. 25–46, 1996.
[8] J. Zhang, “A directional asymmetry in Chinese tone sandhi systems,” *Journal of East Asian Linguist*, vol. 16, pp. 259–302, 2007.
[9] P. Boersma and D. Weenink, Praat: doing phonetics by computer [computer program]. Version 6.0.36, 2017. http://www.praat.org. (accessed 21 February 2018)
[10] R Core Team, R: A language and environment for statistical computing. Version 3.4.3, 2017. (last accessed 21 February 2018).
[11] A. Kuznetsova, P. B. Brockhff, and R. H. B. Christensen, “lmerTest package: Tests in linear mixed effects models,” *Journal of Statistical Software*, vol 82, no. 13, pp. 1–26, 2017. doi: 10.18637/jss.v082.i13
[12] S. G. Luke, “Evaluating significance in linear mixed-effects models in R,” *Behavior Research Methods*, vol. 49, no. 4, pp. 1494–1502, 2017. doi: 10.3758/s13428-016-0809-y
[13] A. W. Mills, “Lateralization of high-frequency tones,” *Journal of the Acoustical Society Of America*, vol. 32, pp. 132–134, 1960.
[14] J. C. Middlebrooks and D. M. Green, “Sound Localization by Human Listeners,” *Annual Review of Psycholog*y, vol. 42, pp. 135–159, 1991.
[15] K. Forinash and W. Christian, Sound: An Interactive EBook, 2016. Retrieved February 23, 2018, from https://www.compadre.org/books/SoundBook