Analyzing Nanjing Tones and Sandhi: statistical modelling methods

Si Chen[1], Caroline Wiltshire[2], Bin Li[3], & Ratree Wayland[2]
[1]The Hong Kong Polytechnic University, [2]*University of Florida, &* [3]*City University of Hong Kong*

When documenting and analyzing a tone language, two major issues related to the tone system must be resolved: 1) what is the tone inventory? 2) what are the tone sandhi rules? The second question is particularly difficult, as tone sandhi and tonal coarticulation are hard to disentangle in an unknown system (Li & Chen, 2016). We argue that modelling the underlying pitch target for tones can answer both questions, and we apply statistical techniques to Nanjing Chinese data to resolve discrepancies in the early descriptions of tones and sandhi in this understudied variety.

The language: Nanjing, a city on the east coast of China, has a population of over 5.3 million and two main languages: Nanjing Chinese and Mandarin Chinese (Xu et al. 2007). Descriptions of Nanjing's tone system have differed. The general consensus is that there are five basic tones and five or six tone sandhi rules (Liu 1995, 1997; Song 2006; Sun 2003). However, Sun (2003) and Liu (1995, 1997) describe the basic tones with the following Chao values: T1 = 31 or 41, T2 = 24 or 13, T3 = 22 or 212 or 11, T4 = 44, T5 = 5 or 55; thus the values of the first three tones are still unclear. Moreover, they provide different sandhi rules. In a T3 + T3 combination, the first T3 becomes T2 in Liu (1995) but T1 in Sun (2003). In T5 + T5, Liu (1995) reports that the first T5 turns into a derived tone, but Sun (2003) argues that it becomes T4. Furthermore, Liu (1995) has one more sandhi rule (T4 + T5 = T1 + T5) than Sun (2003).

Participants and stimuli: Twelve native speakers of Nanjing Chinese (6F/6M), aged 35-65, participated in our study. All speakers were recorded in a quiet room using a digital recorder with a head-mounted microphone, and the recordings were transferred to a PC with a sampling rate of 44.1kHz. The speech materials consisted of 660 monosyllabic tones (11 monosyllables * 4 tones * 12 speakers) and 360 disyllabic tones (5 disyllabic words * 6 combinations * 12 speakers). Real words were used; most disyllables were chosen from Liu (1995) in consultation with native speakers.

Analytical methods: Using Praat, target words were segmented manually, and a script was run to extract F0. In each vowel, time-normalized F0 values were extracted with 20 time points, and the analysis window size was 25.6ms. We first performed a logarithmic Z-score normalization on F0 values (Rose 1987; Zhu 1999), and modelled the surface tonal contours and the underlying pitch targets (Prom-On 2009, Chen et al. 2017). We transform F0 values to Chao tone numbers, building on previous methods (Shi 1990; Zhu, Shi & Wei 2012) but based on the underlying pitch targets. To do so, we calculated four sample quantiles (20%, 40%, 60%, 80%) for all the fitted values of monosyllabic tones to obtain cut-off values for each sample quantile, which divide the acoustic tonal space evenly and allow a transformation from a fitted F0 value to an integer from 1-5. To determine sandhi (as opposed to tonal coarticulation), we tested underlying pitch targets by determining the optimal model for each pair of sandhi tone and the tone it turns into, appearing in the same or similar disyllabic contexts. Then, we fit the optimal model and obtained coefficients from each speaker to compare coefficients of underlying pitch targets of each sandhi tone and the citation tone, testing whether they were statistically different. Non-significant difference in the coefficients is taken as evidence of neutralized sandhi tones.

Findings: All tones can be best modelled by a linear underlying target except for T2, which requires a quadratic term. The initial and final points of the fitted values based on the optimal model of each tone are given in Table 1. Sample quantiles for all the fitted values of five monosyllabic tones were calculated, and the cut-off values corresponding to each quantile are as follows: 20%: -0.90; 40%: -0.02; 60%: 0.39; 80%: 0.97. Transformed integers appear in Table 1, along with the previous impressionistic values.

. Table 1. Beginning and ending points of fitted tone values

Tone	Initial F0, normalized (Chao's #)	Final F0, normalized (Chao's #)	Reported values in literature	**Proposed Tone values**
1	0.65 (4)	-0.82 (2)	31/41	**42**
2	-0.85 (2)	0.94 (4)	24/13	**24**
3	-0.94 (1)	-1.59 (1)	22/212/11	**11**
4	0.45 (4)	0.15 (3)	44	**43**
5	1.08 (5)	1.31 (5)	5/55	**55**

For sandhi, we found that the underlying pitch target of the first T3 in T3 + T3 did not neutralize with that of citation T1 (Sun 2003), but rather with that of T2 (Liu 1995). Moreover, the modelling of acoustic data lends support to Sun (2003)'s proposal of a rule T5 → T4/_T5, where the underlying pitch targets of T4 and T5 neutralized before T5, undermining Liu (1995)'s description of a derivational tone with a tone value transcribed as "3". Based on our modelling, we propose the rules in Table 2.

Table 2 Comparison of tone sandhi rules proposed in literature and in current study

Liu (1995)	Sun (2003)	**Proposed sandhi rules**
T1→T4/_T1 (41→44/_41)	T1→T4/_T1 (31→44/_31)	**T1→T4/_T1 (42→43/_42)**
T2→T3/_T5 (24→11/_5)	T2→T3/_T5(13→22/_5)	**T2→T3/_T5 (24→11/_5)**
T3→T2/_T1 (11→24/_41)	T3→T2/_T1(22→13/_31)	**T3→T2/_T1 (11→24/_42)**
T3→T2/_T3 (11→24/_11)	T3→T1/_T3(22→31/_22)	**T3→T2/_T3 (11→24/_11)**
T4→T1/_T5 (44→41/_5)	None	**T4→T1/_T5 (43→42/_5)**
T5→3/_T5 (5→3/_5)	T5→T4/_T5 (5→44/_5)	**T5→T4/_T5 (5→43/_5)**

Conclusion: We show that statistically testing underlying pitch targets reveals the categorical nature of tone sandhi. Moreover, acoustic studies involving statistical testing and modelling of underlying pitch targets can provide a quantitative basis for a more precise transcription of tones and sandhi, and these techniques can be applied to any undocumented or understudied language to provide a tonal analysis based on productions by the speech community.

Selected References

Chen, S., C. Zhang, A. McCollum, & R. Wayland. 2017. Statistical modelling of phonetic and phonologised perturbation effects in tonal and non-tonal Languages. *Speech Communication* 88, 17-38.

Li, Q., & Y. Chen. 2016. An acoustic study of contextual tonal variation in Tianjin Mandarin. *Journal of Phonetics* 54, 123 – 150.

Liu, D.Q. 1995. *Nanjing Fangyan Cidian (A dictionary of the Nanjing Dialect)*. Nanjing: Jiangsu Education Publisher.

Liu, D. Q. 1997. *Nanjing Dialect Phonetic Profile*. Shanghai: Shanghai Education Publisher.

Prom-On, S., Y. Xu, & B. Thipakorn. 2009. Modeling tone and intonation in Mandarin and English as a process of target approximation. *JASA 125*(1), 405 – 424.

Song, Y.D. 2006. *Nanjing Fangyan Shengdiao Shiyan Yanjiu*. (The acoustic study of tones in Nanjing dialect). Master's thesis, Nanjing Normal University, Nanjing.

Sun, H.X. 2003. Nanjing fangyan shengdiao de ruogan wenti. (A discussion on the tones of Nanjing dialect). *Journal of Nanjing Xiaozhuang College, 1, 34 – 40*.

Xu, C.,M. Liu. C. Zhang, S. An, W. Yu & J.M. Chen. 2007. The spatiotemporal dynamics of rapid urban growth in the Nanjing metropolitan region of China. *Landscape Ecology 22*, 925-937.

Zhu, X. N., D.F. Shi, & M. Y. Wei. 2012. Six level tones in Yuliang Miao and the Multiregister and Four-Level Tonal Model. *Minzu Yuwen, 4,* 1-10.