

Analyzing Nanjing Tones and Sandhi: Statistical Modelling Methods

Si Chen¹, Caroline Wiltshire², Bin Li³, Rtree Wayland²

¹ *The Hong Kong Polytechnic University*, ² *University of Florida*, ³ *City University of Hong Kong*

1 Introduction

Tone sandhi is a phonological and categorical process exhibiting contextually determined tone alternations at the juncture of words or morphemes, in contrast to the phonetic and gradient phenomenon of tonal coarticulation in which the acoustic realization of tones exhibits either a carryover (progressive) or anticipatory (regressive) effect of the preceding or the following tones, (e.g. Brunelle, 2009; Chen, Wiltshire & Li, 2018). Statistical modeling of the acoustic data is conducive to promoting our understanding of tone sandhi and providing phonological representations of underlying tone targets (Chen, accepted). The surface f0 contours are subject to more phonetic variation, and previous studies examining the surface contours report non-neutralization in the Mandarin third-tone sandhi rule (Peng, 2000) and tone sandhi rules in Tianjin Chinese (Zhang & Liu, 2011). Statistical modeling on the underlying pitch targets are shown to be able to statistically distinguish between phonetic and phonologised effects (Chen, Zhang, McCollum and Wayland, 2017), based on the model of underlying pitch targets proposed by Xu and Wang (2001). Since the characteristics of a tone sandhi process involve stability, categorical shift and influence on the entire tones, we test both surface f0 contours and the underlying pitch targets for Nanjing Chinese.

1.1 Nanjing Chinese The two main languages in Najing, China, are Nanjing Chinese and Mandarin Chinese. Nanjing Chinese used to be classified as one of the Wu dialects, but it is now considered to belong to Jianghuai Mandarin (Jin, 2010).

Tone values are often transcribed based on impressionistic data, resulting in some discrepancies among scholars, and phonological analyses can vary due to different transcription (Zhang & Liu, 2011). Although tone values for Nanjing Chinese are transcribed differently among researchers as in Table 1, the general consensus is that there are five basic tones and five or six tone sandhi rules (Liu, 1995, 1997; Song, 2006; Sun, 2003). Sun (2003) and Liu (1995, 1997) report the basic tones with the following values: T1 (31/41), T2 (24/13), T3 (22/212/11), T4 (44), T5 (5/55).

Tone	Syllable	Liu (1995)	Sun (2003)	Meaning
T1	/fu/	41	31	‘skin’
T2	/fu/	24	13	‘hold’
T3	/fu/	11	22/212	‘rotten’
T4	/fu/	44	44	‘negative’
T5	/fu/	5	55	‘fortune’

Table 1. Tones values in Nanjing Chinese according to Liu (1995) and Sun (2003)

* We are indebted to Nikolay Bliznyuk and Michael J. Daniels for help with statistical modelling. This work is supported by grant [1-ZVHH] from the Faculty of Humanities and grant [G-UAAG] from the Department of Chinese and Bilingual Studies at the Hong Kong Polytechnic University and partly supported by Early Career Scheme [No. T26023416] from the Research Grants Council of Hong Kong. We thank those who made comments at our poster presentation at the Annual Meeting on Phonology 2018 in San Diego. Any remaining errors are our own.

Liu (2011) notes that the values of certain tones are still debatable: (1) whether T3 is a contour or a level tone with different transcriptions, such as 212, 22 and 11; (2) whether T1 can be described as 31 or 41 and (3) whether T2 has the tone value 13 or 24.

Moreover, there are also some discrepancies concerning tone sandhi rules among scholars, as shown in Table 2. Liu (1995) has one more sandhi rule for the tonal combination T4 + T5 (22/212/11 + 5/55) than those proposed by Sun (2003). Also, Liu (1995) argues that the first T3 in T3 + T3 becomes T2, but Sun (2003) states that it becomes T1 instead. In the tonal combination T5 + T5, Liu (1995) argues that the first T5 turns into a derived tone, but Sun (2003) argues that it turns into T4.

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

Table 2. Tone Sandhi rules for Nanjing from Liu (1995) and Sun (2003)

1.2 Underlying Pitch Targets In this study, we examined surface contours as well as underlying pitch targets of tones in Nanjing Chinese. The models proposed to capture the underlying pitch targets are distinguished from those capturing surface f0 contours (Prom-On, Xu, & Thipakorn, 2009). The underlying pitch targets can include static and dynamic pitch targets, and the operation of the pitch targets adheres to pitch implementation rules. The statistical tests of the underlying pitch targets in this study are models of underlying pitch targets.

We statistically test whether the underlying pitch target is linear or quadratic, and whether the underlying pitch target of a citation tone has changed significantly in tonal contexts or not, based on previous quantitative models. Two models have been proposed for the underlying pitch targets. One is proposed to estimate T(t), representing the underlying target as follows:

$$(1) \begin{aligned} T(t) &= at + b \\ y(t) &= \beta e^{-\lambda t} + at + b \end{aligned}$$

where y(t) represents f0 values. The parameter λ represents the rate of approaching the target, the parameter a is the slope, and b is the intercept of the underlying target. Also, a third order critically damped system is proposed to estimate the underlying pitch target x(t) as follows:

$$(2) \begin{aligned} x(t) &= mt + b \\ f_0(t) &= (c_1 + c_2 t + c_3 t^2) e^{-\lambda t} + x(t) \end{aligned}$$

where $f_0(t)$ represents f0 values, and λ represents the rate of approaching the target. The initial f0 values, initial velocity and initial acceleration determine the three parameters c1, c2, and c3. The parameter m is the slope of the underlying pitch target, and the parameter b is the intercept of the pitch target. We used the R language (R Core Team, 2018) to conduct all the statistical analyses in this study.

2 Methodology

2.1 Participants and stimuli A total of 12 native speakers of Nanjing Chinese (six females and six males) participated in this study. Participants were between 35 and 65 years old and had lived in Nanjing for most of their lives. Literature on Nanjing Chinese (Liu 1995, Song 2006, Chen & Wiltshire 2013) documents differences across age groups, so we chose a specific age group for the current study. The age range contains only participants considered to speak a relatively new version of the dialect, compared with

speakers of more than 75 years of age who speak an older version as noted by Liu (1995). All our participants were recorded in a quiet room during the course of fieldwork using a Marantz PMD 660 digital recorder with a Shure SM2 head-mounted microphone, and the recordings were transferred to a PC with a sampling rate of 44.1kHz.

The speech materials selected in this study consisted of 660 monosyllabic tones (11 monosyllables * five tones * 12 participants) and 360 disyllabic tones (five disyllabic words * six combinations * 12 participants). Real words were used because the application of tone sandhi rules on nonsense words may be different from the application on real words (Zhang & Lai, 2010). Most disyllables were chosen from the *Dictionary of the Nanjing Dialect* (Liu, 1995) with consultations from native speakers of Nanjing Chinese. All the words were recorded at a normal speaking rate, and the speakers were instructed to adhere to the same intonation pattern they would use for declarative sentences.

2.2 F0 extractions and statistical analysis The target words were first segmented manually, using Praat (Boersma & Weenink, 2012), and a Praat script was run to extract f0 within each individual segment. 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, and modelled the surface tonal contours and the underlying pitch targets.

In addition, we applied a method of transformation from f0 values to Chao's tone numbers based on statistical modelling (Chen, accepted). The method we use in this study is based on the modelling of underlying pitch targets. We first calculated four sample quantiles (20%, 40%, 60% and 80%) for all the fitted values of monosyllabic tones to obtain cut-off values corresponding to each sample quantile. Using these sample quantiles, we may cut the acoustic tonal space evenly, and perform a transformation. For example, if a fitted value f0 falls in the range of the sample quantiles 0% - 20%, then this value is transformed to the integer 1 according to Chao's tone numbers.

3 Results

3.1 Results of testing surface contours of Nanjing Chinese sandhi tones First we examined the surface values of tones which are reported to have undergone sandhi, and compare them with the surface value of the tones into which they are reported to change. In the following discussion, we are comparing the first tone in a two tone sequence, and we indicate this by listing the two tones followed by an 'a'. That is, T15a means that we are looking at the first tone ('a') in the sequence Tone 1-Tone 5, i.e. Tone 1 when it appears before Tone 5. Figure 1 plots each tone compared with the tone it is supposed to change into in a sandhi context, according to the descriptions of Liu (1995) and/or (Sun (2003). We chose the first T4 in the disyllabic context T4 + T4 (T44a) to compare with the first T5 in T5 + T5 (T55a) instead of T4 in T4 + T5 (T45a), because it is reported that T4 changes into T1 in T4 + T5 (Liu, 1995). The reason for choosing T4 + T4 is that the disyllabic context is relatively similar to T4 + T5, where T4 precedes a high and level tone T4, the most similar tone to T5.

In Figure 1, the shape of these pairs are fairly close, especially the slopes, except for the pair T33a and T13a in (d), which exhibit different signs of the slopes. However, most pairs do show small yet distinct acoustic differences on the surface contours visually. Statistical tests showed that all the pairs of tone sandhi reported in the literature were non-neutralized in the surface tonal contours. The differences mainly lie in the intercept and the slope terms.

One tone sandhi pair T4→T1/_T5 is reported by Liu (1995), but not by Sun (2003). The surface tonal contour for the sandhi T4 in T4 + T5 did deviate from T4 in isolation, but the existence of this tone sandhi is still questionable, since tonal coarticulation may also have affected the surface tonal contours. Therefore, we further tested whether neutralization may have occurred in the underlying pitch targets.

3.2 Underlying pitch targets of Nanjing Chinese monosyllabic tones We next calculated the underlying pitch target for each monosyllabic tone. Table 3 presents the selected model and the estimated parameters for each monosyllabic tone. Plugging in the estimated coefficients in Table 3 for each model of tones, we can plot the fitted tonal contours and the mean tonal contours from the original data for comparison as in Figure 2.

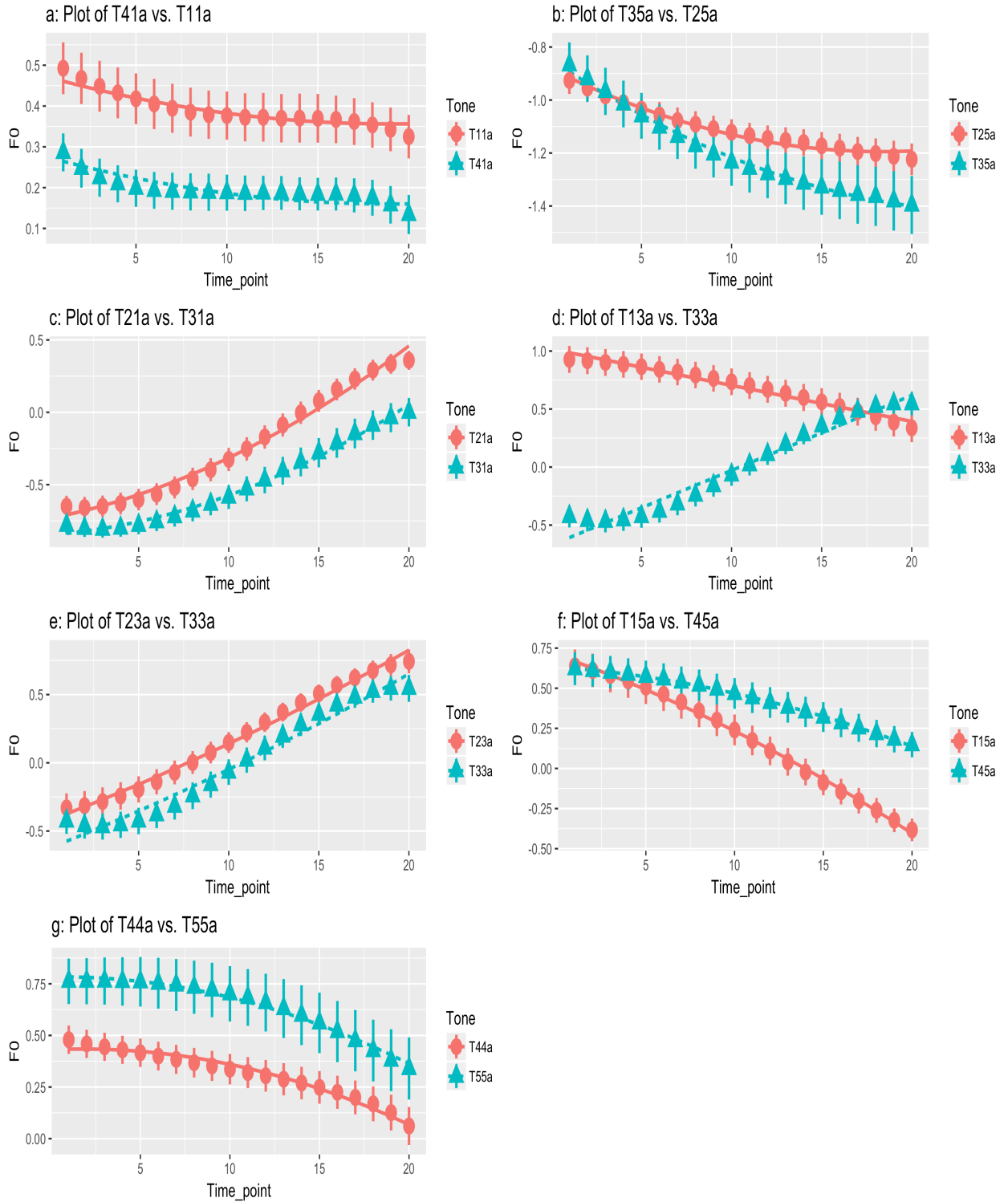


Figure 1. Growth curve analysis in Nanjing Chinese: the mean and error bar of the f0 data in each normalized time point are plotted, and the solid and dotted lines represent the fitted values of the growth curve models representing two tonal contours (a: T41a vs. T11a; b: T35a vs. T25a; c: T21a vs. T31a; d: T13a vs. T33a; e: T23a vs. T33a; f: T15a vs. T45a; g: T44a vs. T55a)

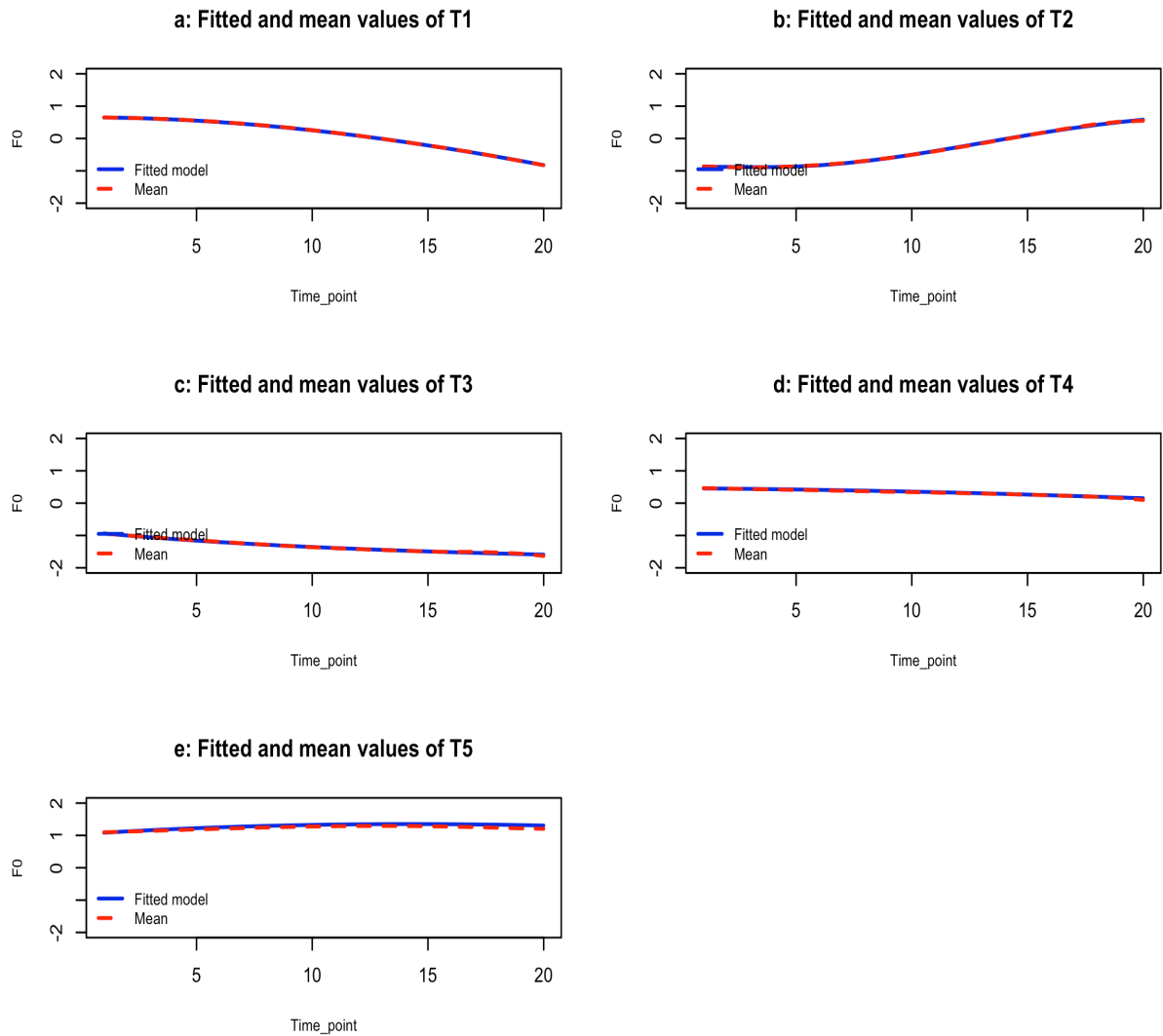


Figure 2. Fitted and mean values of Nanjing Chinese T1-T5: the dotted lines plot the averaged values from the original observed data, and the solid line is the fitted values from the optimal model for each tone.

Tone	Model Selected	$c1$	$c2$	$c3$	β	λ	a	b
1	Second order	NA	NA	NA	-7.80	0.035	-0.27	8.45
2	Third order	-11.85	-1.30	-0.09	NA	0.14	-0.27	10.96
3	Second order	NA	NA	NA	0.71	0.09	-0.0065	-1.58
4	Second order	NA	NA	NA	-5.21	0.015	-0.083	5.67
5	Second order	NA	NA	NA	-9.30	0.02	-0.14	10.34

Table 3. Models chosen for each tone and estimated coefficients

The beginning and ending points of the fitted values based on the optimal model of each tone were calculated first. The four sample quantiles (20%, 40%, 60% and 80%) for all the fitted values of five monosyllabic tones were then calculated. The cut-off values corresponding to each sample quantile is as follows: 20%: -0.90; 40%: -0.02; 60%: 0.39; 80%: 0.97. If the fitted f_0 is 0.65, which falls in the range 0.39 (60%)-0.97 (80%), then this value is transformed to the integer four according to Chao's scale. We therefore proposed the following transcribed values for Nanjing Chinese tones: Tone 1(42), Tone 2(24), Tone 3(11), Tone 4(43), Tone 5(55). The resulting system is presented in Table 4, which also provides the previously reported values for each tone.

Tone	Initial F0, normalized (Chao's #)	Final F0, normalized (Chao's #)	Values in literature	Our 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

Table 4. Tone values for Nanjing Chinese tones

In addition, we tested underlying pitch targets in Nanjing Chinese. We first determined the optimal model for each pair of sandhi tone and the reported tone it turns into, appearing in the same or similar disyllabic contexts. Then, we fit the chosen optimal model and obtained coefficients from each speaker to compare the coefficients of underlying pitch targets. The results showed that most of the underlying pitch targets of the sandhi tones did become neutralized with the reported citation tone. The underlying pitch target of the first T3 in T3 + T3 did not neutralize with that of the reported citation T1 (Sun, 2003), but it neutralized with that of T2 (Liu, 1995). Moreover, the modelling of acoustic data lends support to Sun (2003)'s proposal of the rule T5 → T4/_T5, where the underlying pitch target of T55a neutralized with T45a, not supporting Liu (1995)'s description of a derivational tone with the tone value transcribed as "3".

Figure 3 plots the mean and error bars of the observed data as well as the fitted values of the model, which can be regarded as the surface tonal contours regenerated based on the fitted underlying pitch target. In sum, based on the modelling of acoustic data consisting of monosyllabic tones and tone sandhi, we propose the following tone sandhi rules with the transcription: T1→T4/_T1 (42→43/_42); T2→T3/_T5 (24→11/_5); T3→T2/_T1 (11→24/_42); T3→T2/_T3 (11→24/_11); T4→T1/_T5 (43→42/_5); T5→T4/_T5 (5→43/_5).

The use of underlying pitch target models leads us to confirm the first three rules, on which Liu (1995) and Sun (2003) agreed, while supporting two of Liu's further three rules, and the final rule of Sun's, as shown in Table 5.

Liu (1995)	Sun (2003)	Our 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)

Table 5. Sandhi rules of Nanjing Chinese

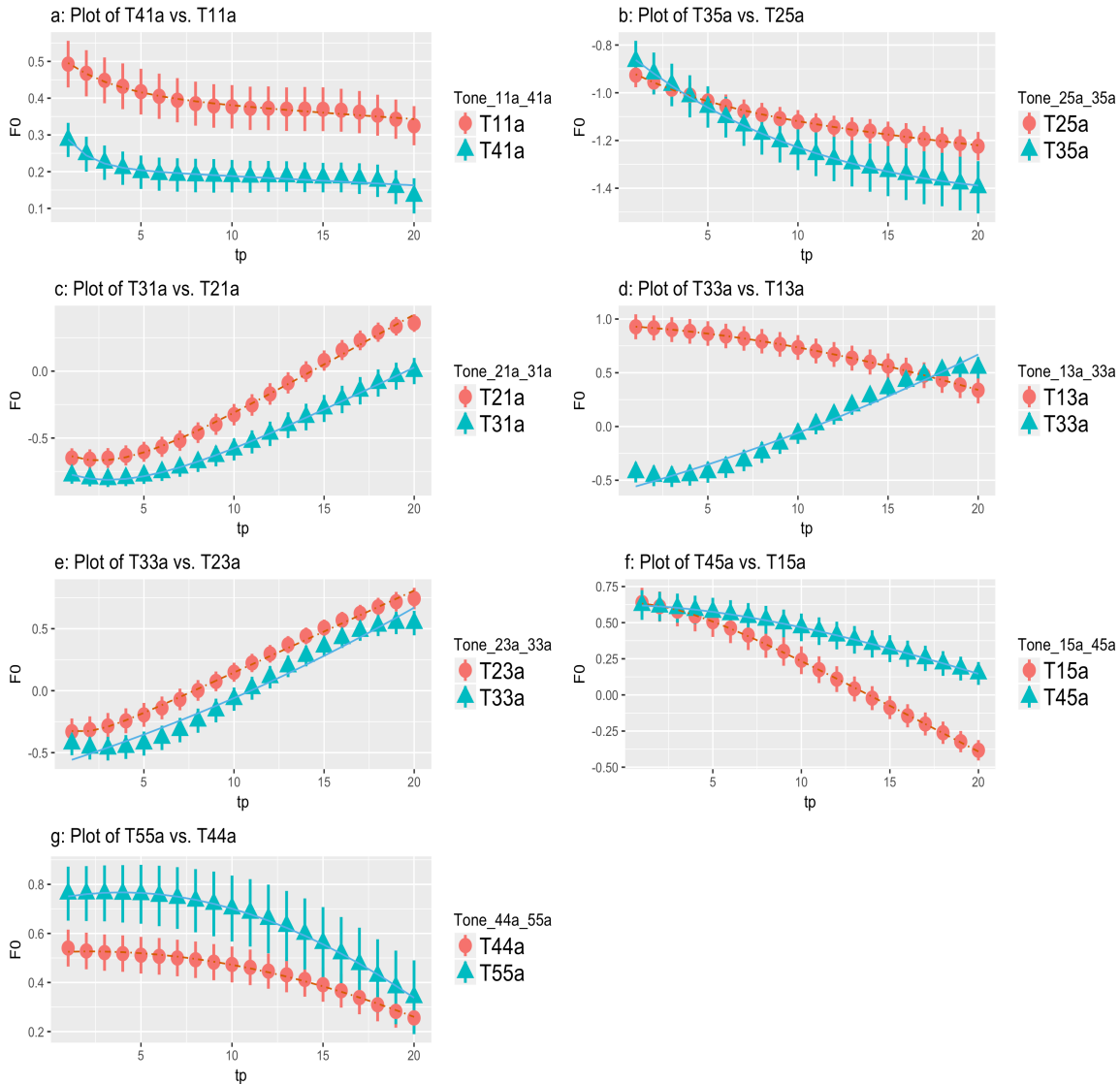


Figure 3. Fitted underlying pitch target models in Nanjing Chinese (a: T41a vs. T11a; b: T35a vs. T25a; c: T21a vs. T31a; d: T13a vs. T33a; e: T23a vs. T33a; f: T15a vs. T45a; g: T44a vs. T55a)

4 Discussion

Transcriptions in Chao's scale were previously proposed for monosyllabic tones in Nanjing Chinese based on impressionistic data in the literature. This study proposes transcriptions of Nanjing Chinese monosyllabic tones to represent the onset and offset of each tone, which corresponded well with the plots of tones and statistical models of acoustic data.

The tonal shapes, including tones with a rising or a falling slope, correspond fairly well to traditional descriptions, though the specific integers chosen for Chao's transcription based on acoustic data in this study exhibit some variation from traditional descriptions. After examining the slope and intercept values for the underlying pitch target as well as the plots for regenerated f0 contours and mean f0 contours, a transcription of each monosyllabic tone that we propose based on statistical analysis of acoustic data is as follows: T1 (42), T2 (24), T3 (11), T4 (43), T5 (5). The transformation to Chao's tone numbers based on statistical modelling has some advantages (Chen, accepted): (1) Statistical modelling may help determine whether a tone should be represented as a straight or circumflex tone; (2) Statistical modelling helps predict

tonal contours of the entire speech community based on samples of speech production; (3) The transformation is based on underlying pitch targets, conforming to articulatory mechanisms (Xu & Prom-on 2014), whereas previous transformation methods rely extensively on the raw data or some transformation of it without statistical modelling.

Moreover, we showed that examining both surface and underlying pitch targets may improve our understanding of tone sandhi rules. Using growth curve analysis to analyze Nanjing Chinese tone sandhi rules, we found statistical differences in the intercept or the linear terms used to capture the surface contours. These results indicate that growth curve analysis is able to capture variability in the surface f_0 contours, and is sensitive to phonetic perturbation.

The procedure of testing underlying pitch targets was useful in identifying phonological tone sandhi processes because statistical significance is obtained only when there is a consistent and significant difference between the two underlying pitch targets in question, consistent with the characteristics of tone sandhi, namely stability and a categorical shift. Consequently, this procedure can be extended to help examine tone sandhi phenomena in other under-studied languages. If the underlying pitch target has varied extensively from that of the original citation tone in the same or similar disyllabic context, then a categorical shift has likely occurred, and thus tone sandhi rather than tonal coarticulation is responsible for the observed variation. On the other hand, if the underlying pitch target is not significantly different from that of the original citation tone, then the observed surface perturbation is likely due to contextual coarticulation or perturbation on the surface.

In conclusion, the methods chosen to test significant changes in underlying pitch targets may better help us identify tone sandhi processes. Based on statistical modelling of underlying pitch targets, we confirm some of the results reported based on impressionistic data, and propose the tone sandhi rules with transcriptions as in Figure 5 above. A numerical approach can function as a useful tool, and thus helps validate and modify research results based on impressionistic data.

References

- Brunelle, Marc. (2009). Northern and southern Vietnamese tone coarticulation: A Comparative Case Study. *Journal of Southeast Asian Linguistics*, 1, 49 – 62.
- Chen, Si. (accepted). Phonological representations based on statistical modelling in tonal languages. In H-M. Zhang & Y-Y. Qian (Eds.) *Prosodic Studies: Challenges and Prospects*. Routledge.
- Chen, Si., & Wiltshire, Caroline. (2013). Tone realization in younger vs. older speakers of Nanjing dialect. In J. S. Zhuo (Eds.), *Increased empiricism: new advances in Chinese linguistics* (pp. 147 – 170). Amsterdam and Philadelphia: Benjamins.
- Chen, Si, Wiltshire, Caroline. & Li, Bin. (2018). An updated typology of tonal coarticulation properties. *Taiwan Journal of Linguistics*, Vol. 16.2.
- Chen, Si, Zhang, Caicai, McCollum, Adam, & Wayland, Rtree. (2017). Statistical modelling of phonetic and phonologised perturbation effects in tonal and non-tonal Languages. *Speech Communication*, 88, 17-38. doi: 10.1016/j.specom.2017.01.006
- Liu, Dan Qing (1995). *Nanjing Fangyan Cidian* [A dictionary of the Nanjing Dialect]. Nanjing: Jiangsu Education Publisher.
- Liu, Dan Qing (1997). *Nanjing Dialect Phonetic Profile*. Shanghai: Shanghai Education Publisher.
- Liu, Shun. (2011). Sanshinian lai de Nanjing fangyan yanjiu [An overview of studies within recent thirty years in Nanjing dialect]. *Nanjing Shenji Xueyuan Xuebao* [Journal of Nanjing Audit University], 8, 92 – 97.
- Peng, Shu-hui. (2000). Lexical versus 'phonological' representations of Mandarin sandhi tones. In M. Broe & J. Pierrehumbert (Eds), *Papers in Laboratory Phonology 5: Acquisition and the Lexicon* (pp. 152 – 167). Cambridge: Cambridge University Press.
- Prom-On, Santitham., Xu, Y., & Bundit Thipakorn, B. (2009). Modeling tone and intonation in Mandarin and English as a process of target approximation. *Journal of the Acoustical Society of America*, 125(1), 405 – 424. doi: 10.1121/1.3037222
- R Core Team (2018). R: A language and environment for statistical computing. R foundation for statistical computing, Vienna, Austria. (Available online at <http://www.R-project.org/>)
- Song, Yi Dan. (2006). *Nanjing Fangyan Shengdiao Shiyuan Yanjiu*. [The acoustic study of tones in Nanjing dialect].
- Sun, Hua Xian. (2003). Nanjing fangyan shengdiao de ruogan wenti. [A discussion on the tones of Nanjing dialect]. *Journal of Nanjing Xiaozhuang College*, 1, 34 – 40.

- Xu, Yi., & Wang, Emily. (2001). Pitch targets and their realization: evidence from Mandarin Chinese. *Speech Communication*, 33, 319 – 337. doi: 10.1016/S0167-6393(00)00063-7
- Zhang, Jie., & Lai, Yu Wen. (2010). Testing the role of phonetic knowledge in Mandarin tone sandhi. *Phonology*, 27(1), 153 – 201. doi: 10.1017/S0952675710000060
- Zhang, Jie., & Liu, Jiang. (2011). Tone sandhi and tonal coarticulation in Tianjin Chinese. *Phonetica*, 68(3), 161 – 191. doi: 10.1159/000333387