The *Journal of Chinese Linguistics* vol.49, no.1 (January 2021): 106–141
 ISSN 0091-3723/ Statistical modeling of application completeness of two tone sandhi rules. By Si Chen and Bin Li.

STATISTICAL MODELING OF APPLICATION COMPLETENESS OF TWO TONE SANDHI RULES

Si Chen[1] **Bin Li**[2]

[1] *The Hong Kong Polytechnic University, Hong Kong*

[2] *City University of Hong Kong, Hong Kong*

ABSTRACT

This study examines the application of two Mandarin tone sandhi rules on real and wug words varying in degrees of phonological and semantic dependency. Using two statistical methods, we examined the surface f0 contours and underlying pitch targets. For the third tone sandhi, a lexical effect was discovered on the relationship between "word-likeness" of stimuli and completeness of rule application. The degree of application for the half-third sandhi tone, however, was less consistent. This study offers new insights in the debate between categorical and gradient views of sandhi rules. We propose three hypothesized situations and argue that the Mandarin tone sandhi rule application involves computation of sandhi forms, though it becomes more incomplete on wug words containing more illegitimate morphemes. Finally, between the two rules, the application of the third tone sandhi rule is less phonetically motivated and more biased in wug words, exhibiting differences between real words and wug words.

Acknowledgments This work is supported by grant [1-ZVHH] and [1-ZVHJ] from the Hong Kong Polytechnic University and partly supported by Early Career Scheme [No. 25602816] from the Research Grants Council of Hong Kong and the Language Fund [EDB(LE)/P&R/EL/175/12] from Standing Committee of Language Education and Research, Hong Kong.

Si Chen (陈思; corresponding author) [sarah.chen@polyu.edu.hk]; Department of Chinese and Bilingual Studies, The Hong Kong Polytechnic University, Hong Kong SAR, China;
https://orcid.org/0000-0003-3581-9561

KEYWORDS
Mandarin tone sandhi **G**rowth curve analysis **U**nderlying pitch target **C**omputation mechanism **W**ug test

1. INTRODUCTION

1.1 Mandarin Tones and Tone Sandhi

Mandarin Chinese has four tones, which can be described by Chao tone numbers (Chao 1948). The tone numbers reflect the starting and ending point of tones on a one-to-five scale, where one indicates the lowest pitch of a speaker, and five the highest, as shown in (1).

(1) Tone 1 (T1) 妈 /ma/ high-level (55) "mother"
Tone 2 (T2) 麻 /ma/ high-rising (35) "hemp"
Tone 3 (T3) 马 /ma/ low-dipping (213) "horse"
Tone 4 (T4) 骂 /ma/ high-falling (51) "to scold"

Tone sandhi is the tonal alternation of a tone triggered by certain phonological environment (e.g., Chen 2000). There are two well-known tone sandhi rules in Mandarin as described in many studies (e.g., Chao 1948, 1965; Cheng 1968). They are shown in (2).

(2) a. T3(213) → T2 (35)/___ T3(213) (third-tone sandhi)
b. T3(213) → 21/___{T1(55), T2(35), T4(51)} (half-third sandhi)

In the tone sandhi rules in (2), (a) describes that the first T3 in a disyllabic combination T3 + T3 changes into T2. Rule (b) states that the second half of the first T3 is truncated in disyllabic combinations T3 + T1, T3 + T2 and T3 + T4. These rules are considered to be phonological for two reasons (Zhang and Lai 2010): first, they concern language-specific tone changes, and secondly, they are not due to tone co-articulation, which influences the beginning and ending of a tone (e.g., Chen, Wiltshire and Li 2018; Xu 1997).

1.2 Neutralization of the Third Sandhi Tone and T2 on Real Words

Myers and Tsay (2003) note two existing views of Mandarin third-tone sandhi on real words—the categorical view and the gradient view—which differ at the phonetic level. They admit that the third-tone sandhi rule can be applied to novel phrases, and that the application seems sensitive to prosodic factors defining the domain of tone sandhi (Cheng 1968; Hsiao 1992; Shih 1986; Wan and Jaeger 1998). The categorical view holds that there is a substitution of T3 by T2, similar to the allomorphy "a/an" in English. However, the gradient view states that there is a modification of T3, which results in sounds more similar to the lexical T2.

These two views are largely dependent on the analysis of acoustic data. Three hypotheses on the degree of application were proposed in the literature: (1) the third tone sandhi is similar to the original T3; (2) the third tone sandhi is not neutralized with T2; (3) the third tone sandhi is neutralized with T2. Early studies argued for the first possibility. Kratochvil (1986) conducted a discriminant analysis considering f0 and amplitude at six points, and showed that sandhi tone was grouped with lexical T3 rather than with lexical T2.

Some studies found evidence supporting the second possibility that acoustic differences exist between the third sandhi tone and T2. Zee (1980) found that neutralization of T2 and the sandhi T3 did not occur where the third tone sandhi is 34 instead of 35. Shen (1990) reported that the average f0 sandhi tone was 6 Hz lower than T2, and with a shallower slope. Xu (1993) and Peng (2000) also found the sandhi tone to be lower in f0 with statistical significance. In a corpus of conversational speech, Yuan and Chen (2014) noted a difference in the magnitude of f0 rise and the percentage of f0 rise duration. However, the discovered differences found might be due to variation in intonation and intrinsic f0 effects (Zhang and Peng 2013).

The third hypothesis on the possibility of neutralization was tested in more recent studies. Zhang and Peng (2013) demonstrated that the f0 contour of the third sandhi tone was not significantly different from T2 after controlling the intrinsic f0 effect.0Xu and Prom-on (2014) used a Mandarin corpus and compared the synthesis accuracies from simulating three third tone sandhi hypotheses. Their results showed that the underlying tonal target of the sandhi tone has either changed to another tonal category or was the same as the rising tone, though the results slightly favored the former. In addition, Chen, Wiltshire and Li (2019) revealed that in real words, there were no statistically significant differences between the underlying pitch target of the third sandhi tone and that of T2 despite significant differences on the surface f0 contours.

1.3 Underlying Pitch Targets

It is beneficial to conduct analyses of tonal contours both on the surface and underlying pitch targets for a better understanding of phonetic and phonological processes (Chen, Zhang, McCollum and Wayland 2017; Chen, 2019) and especially in tone sandhi processes (Chen et al., 2019). A conceptual framework of underlying pitch targets was first proposed by Xu and Wang (2001), where they define an underlying pitch target as “the smallest articulatorially operable units associated with linguistically functional pitch units such as tone and pitch accent.” The pitch targets consist of two types, a static pitch target specifying the register such as [high] or [low], and a dynamic pitch target specifying a linear movement such as [rise] or [fall] (Xu and Wang 2001).

The quantitative Target Approximation (qTA) model was proposed by Prom-on, Xu and Thipakorn (2009), where communication functions such as lexical tones are modelled as the driving force of a linear system. A statistical testing procedure was later proposed by Chen et al. (2017) to test the differences in lexical tones. Chen et al. (2017) showed that despite phonetic variability on the surface f0 contours, the underlying pitch targets might barely show statistical differences. Xu and Prom-on (2014) also stated that although contextual variability can be found on surface f0 contours, the surface f0 contours converge to a corresponding underlying tone (high-level for the High tone, rising for the Rising tone, low-level for the Low tone and falling for the Falling tone). The current study uses statistical testing of both surface f0 contours and underlying pitch targets in examining tone sandhi rule application.

1.4 The Application of Tone Sandhi Rules on Wug Words

In addition to real words, the application of Mandarin tone sandhi rules was also examined on wug words (Berko 1958). Although tone sandhi rules in Taiwanese show inconsistent applications (Hsieh 1970, 1975, 1976; Zhang, Lai and Sailor 2011), the third tone sandhi in Mandarin can be applied to novel words (Xu 1997; Zhang and Lai 2010) and even in

Chinese-English code-switching sentences (Cheng 1968). Zhang, Xia and Peng (2015) highlighted the debate between a lexical and computation mechanism in production of tone sandhi rules, where the former believes that the encoding of tone sandhi only applies to real words since the mechanism involves lexical representations attached to phonological representations (Hsieh 1970, 1975, 1976), and the latter believes that the sandhi forms of tones are computed based on phonological context (Zhang and Lai 2010; Zhang and Peng 2013). Results from an ERP study suggested a computation mechanism governing the phonological encoding of third-tone sandhi (Zhang, Xia and Peng 2015).

Moreover, the application of tone sandhi rules may differ on real versus wug words. Zhang and Lai (2010) used a wug test to examine two patterns of tone sandhi rules (third-tone and half-third sandhi) in Mandarin Chinese. They found that the contour shapes of the third sandhi tones (the first tone in T3+T3) were significantly different between real words and wug words, and the wug words had a lower and later turning point and a longer duration. Zhang and Peng (2013) having tested third tone sandhi on wug words, showed that the average f0 on the sandhi tone was significantly lower than in T2. They also reported significantly different f0 contours in wug words but not in real words, which indirectly shows discrepancy in the application of the third tone sandhi rule on real and wug words. They believe that the subtle f0 differences are more phonetic than phonological. However, no agreement has been reached about acoustic differences of T2 and the third sandhi tone, which might be due to the degree of controlling other factors affecting f0 contours (Peng 2000; Xu 1997; Zee 1980; Zhang and Peng 2013). Moreover, none of these studies used more advance statistical techniques to capture variation on the surface f0 contours nor did they examine the differences of underlying pitch targets, which may prove similar between real and wug words. We aim to model the whole surface tonal contour as well as the underlying pitch targets of real and wug words, and statistically test whether they differ on real and wug words with stimuli highly controlled for consonants and vowels. We hypothesize three situations regarding the rule application on real words versus wug words. The first situation is plotted in Figure 1(a), where neither the underlying pitch targets nor the surface f0 contours has significant differences. The second situation in Figure 1(b) shows a situation where the surface contours are significantly different, but the underlying pitch targets have no significant differences. Figure 1(c) describes a situation where both the surface contours and underlying pitch targets are different. The first situation indicates that computation mechanism is involved in application of tone sandhi rules on wug words. The second and third situations indicate that the computation might not be complete, leading to differences either on the surface contours or underlying targets.

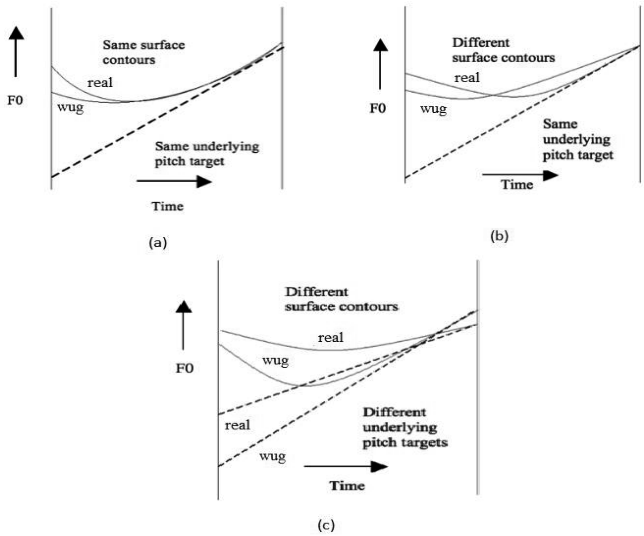

Figure 1 (a) Neither surface f0 contours nor underlying pitch targets differed significantly between real and wug words; (b) Surface f0 contours differ significantly between real and wug words, but the underlying pitch targets do not; (c) Both surface f0 contours and underlying pitch targets differ significantly; Solid lines stand for surface f0 contours and dotted lines represent underlying pitch targets.

The current study aims to answer the following research questions: (1) Are T2 and T3 neutralized on the surface f0 contours of different types of wug words? (2) Does the underlying pitch target of T2 differ from that of T3 on different types of wug words? (3) Does the third tone sandhi rule apply in the same manner on real versus wug words in terms of surface f0 contours and underlying pitch targets? (4) Does the half-third sandhi rule apply in the same manner on real versus wug words in terms of surface f0 contours and underlying pitch targets? (5) Does the third tone sandhi rule apply less accurately than the half-third sandhi rule on wug words?

2. METHODOLOGY

2.1 Stimuli

Based on the wug tests proposed by Zhang and Lai (2010) and Zhang and Peng (2013), we designed six blocks of minimal pairs by taking into consideration the phonological and morphological dependency of monosyllabic structures. Each block contains 8 pairs of bi-syllabic words contrasting in tones that the initial syllables bear (T2+T3 versus T3+T3). The six blocks are: (1) real words consisting of actual occurring morphemes, referred to as AO-AO, (2) non-occurring sequences consisting of real morphemes, referred to as *AO-AO, (3) sequences of a real morpheme and a syllable of an accidental gap, referred to as AO-AG, where AG stands for an accidental gap (The accidental gap refers to missing combination of a legal syllable with certain tones in the Mandarin lexicon, e.g., [s^{w}æn] is a legal syllable, which can bear T1 (55) "sour" 酸 and T4 (51) "garlic" 蒜, but it cannot bear the second or the third tone), (4) sequences of a syllable of an accidental gap and a real morpheme, referred to as AG-AO, (5) sequences of two syllables of accidental gaps, referred to as AG-AG, and (6) sequences of two pseudo syllables. Segmental components of such syllables are all legitimate in Mandarin, but their combination to form a syllable is not.

Our design of stimuli differs from that of previous research in that we included both accidental gaps as well as pseudo words to examine two Mandarin tone sandhi rules: the third tone sandhi and half-third sandhi rule. Also, the vowels across all six blocks were highly controlled and kept as similar as possible to reduce the intrinsic f0 effect of vowels (IF0). In order to minimize effects of tonal and lexical influence on participants, filler words consisting of real and wug words were designed to match the experimental stimuli (*AO-AO, AG-AG, etc.) in each block. There were a total of 1344 stimuli (6 block * 16 items * 14 speakers) and 1344 fillers (6 block * 16 fillers * 14 speakers). The test stimuli can be found in the Appendix.

2.2 Subjects

Participants were 14 native speakers of Mandarin (eight males and six females; age: 26 ± 4.86 (Mean ± SD)) who had lived in Beijing for most of their lives before coming to Hong Kong (20.14 ± 3.38 (Mean ± SD)). None of them reported any history of speaking, hearing or language difficulty. All participants wearing a GMH C 8.100 D headset were recorded at the speech lab of the Hong Kong Polytechnic University using Audacity on a laptop with a sampling rate of 44.1 kHz. All participants were paid for participating in the experiment and signed informed consent forms in compliance with a protocol approved by the Human Subjects Ethics Sub-committee at the Hong Kong Polytechnic University.

2.3 Experiment Procedure

Monosyllables of all the target stimuli and filler words were first read by a native speaker of Mandarin, who grew up in Beijing. Then we normalized the intensity of all the recorded monosyllables. Pairs of recorded monosyllables were presented to each speaker randomly using E-prime, with an Inter-Stimulus Interval (ISI) of 800ms. After hearing both monosyllables, participants were instructed to put them together to produce a disyllabic word in Mandarin. They were also instructed to speak at a normal speaking rate and were allowed to repeat themselves when necessary. All participants practiced this procedure in a training session before the experiment.

2.4 Methods of Data Analysis

We used both acoustic examination and statistical analysis to examine the degree of application of the two Mandarin tone sandhi rules.

We performed the acoustic examination in the following steps. First, vowels in the valid tokens were segmented and extracted for acoustic analysis. The segmentation criteria followed the procedure described in Jangjamras (2011), where the vowel onset is defined as "the first zero crossing at the beginning of voicing in the waveform," and the vowel offset is "at the downward zero crossing immediately following the final glottal pulse in the waveform". After proper segmentation, f0 values at 20 normalized time points in each segmented interval were obtained by the Praat script called Prosodypro (Xu 2013). We proceeded to perform a logarithmic Z-score normalization on the extracted f0 values to normalize variation in f0 values across gender (Rose 1987; Zhu 1999). The normalized f0 values were then subject to statistical analysis, including growth curve analysis of the surface f0 contours and a statistical procedure to compare underlying pitch targets.

Growth curve analysis, also known as "hierarchical linear modelling" and "mixed-effects model" (Mirman 2016, 22; Raudenbush and Bryk 2002), is usually used to capture detailed phonetic differences in contours on surface f0 contours (e.g., Chen et al. 2017).

In this study, we started to fit a simple linear mixed effects model with no quadratic terms as in equation 1 where i represents the i^{th} f0 contour and j is the j^{th} time point where the f0 value was extracted, γ_{00} stands for the population mean of the intercept, ζ_{0i} models variability of individual's intercept, γ_{10} is the population mean of the slope, ζ_{1i} models variability of individual's slope and ε_{ij} are the error terms. (Mirman, Dixon and Magnuson 2008).

$$Y_{ij} = (\gamma_{00} + \zeta_{0i}) + (\gamma_{10} + \zeta_{1i}) * Time_{ij} + \varepsilon_{ij} \quad (1)$$

In fitting each pair of surface f0 contours, we used orthogonal polynomials to avoid correlation between the linear and quadratic terms (Mirman 2016, 52). Then we included the quadratic terms as the fixed effect and the random effect modelling variation if they reached statistical significance by likelihood ratio tests. After optimizing the model for a pair of f0 contours, we conducted a likelihood ratio test to examine differences in each pair of surface f0 contours. Specifically, we fit a model treating the pair of surface f0 contours as the same and a model treating them as different, and we compared the two models using a likelihood ratio test. If significant differences were found in the model comparison, then it indicates a difference between the members of the pair under examination.

Next, based on this idea of separating surface f0 contours and underlying pitch targets, we proceeded to investigate differences in underlying pitch targets. In this study, we followed a recently proposed statistical procedure in testing underlying pitch targets (Chen et al. 2017). The procedure is based on quantitative models (Sun 2001; Prom-On et al. 2009) proposed to quantify Xu and Wang's (2001) conceptual model. Both Sun (2001) and Prom-On et al. (2009) propose a linear underlying pitch target in the form of equation 2 where the parameter t stands for time point, a is the slope and b is the intercept of the underlying target.

$$T(t) = at + b \qquad (2)$$

Sun (2001) used a second order critically damped system, as shown in equation 3 where $f_0(t)$ stands for the f0 contour on the surface, the parameter t stands for time point, the coefficient β is the distance between f0 contour and the underlying pitch target when t = 0, the parameter λ represents the rate of approaching the target.

$$f_0(t) \;=\; \beta exp(-\lambda t) \;+\; at + b \qquad (3)$$

Prom-On et al. (2009) used a third order critically damped system as in equation 4 where $f_0(t)$ is the response of frequency, λ represents the rate of approaching the target. The three parameters are determined by the initial f0 values, initial velocity and initial acceleration.

$$f_0(t) = (c_1 + c_2 t + c_3 t^2)e^{-\lambda t} + at + b \qquad (4)$$

Since the underlying pitch target might be curvilinear in some languages (Xu 2005), Chen et al. (2017) proposed a model selection procedure comparing models that use quadratic and linear pitch targets. In addition, Prom-On et al. (2009) showed the fourth and higher order linear systems did not improve the model fitting compared to lower order linear system, and that a critically damped linear system was better than an overdamped system. Chen et al. (2017) also took both second and third critically damped system into consideration. Therefore, we chose to fit the following four models as proposed by Chen et al. (2017) and selected the optimal one, which has the least Akaike's Information Criterion (AIC; a smaller AIC indicates a better fit (Kim and Timm 2006)):

1) The second order linear system with linear underlying targets (sim_1) as in equation 5 where $f_0(t)$ represents f_0 values, λ is the rate of approaching the underlying pitch target, a is the slope of the underlying target, and b is the intercept of it.

$$f_0(t) = \beta e^{-\lambda t} + at + b \quad (5)$$

2) The third order linear system with linear underlying targets (com_1) as in equation 6 where $f_0(t)$ represents f_0 values, λ represents the rate of approaching the underlying pitch target, a is the slope of underlying pitch target, b is the intercept of it, c_1, c_2 and c_3 are transient coefficients.

$$f_0(t) = (c_1 + c_2 t + c_3 t^2)e^{-\lambda t} + at + b \quad (6)$$

3) The second order linear system with polynomial of the second degree in the underlying targets (sim_2) as in equation 7 where $f_0(t)$ represents f_0 values, λ represents the rate of approaching the underlying pitch target, d is the quadratic coefficient of the underlying pitch target, a is the slope of it, and b is the intercept of it.

$$f_0(t) = \beta e^{-\lambda t} + dt^2 + at + b \quad (7)$$

4) The third order linear system with polynomial of the second degree in the underlying targets (com_2) as in equation 8 where $f_0(t)$ stands for f_0 values, λ is the rate of approaching the underlying pitch target, d is the quadratic coefficient of the underlying pitch target, a is the slope of it, b is the intercept of it, c_1, c_2 and c_3 are transient coefficients.

$$f_0(t) = (c_1 + c_2 t + c_3 t^2)e^{-\lambda t} + dt^2 + at + b \quad (8)$$

We used nonlinear regression to fit all the above four models. A nonlinear regression model requires initial estimates of parameters, and the model was iteratively solved (Huet et al. 2006). The initial estimates of parameters were found by plotting the function and adjusting the values so that the shape of the contour was closer to the mean f0 curve. The model fit was then improved by adjusting the values of initial estimates until the improvement is small and negligible. After optimizing the underlying pitch target model, we statistically tested pairs of underlying pitch targets following the procedure proposed in Chen et al. (2017). We first fit the chosen optimal model for each speaker and obtained all parameters of the underlying pitch target for each speaker. We excluded any outliers of the coefficients identified by the Hampel identifier. Then we proceeded to conduct a non-parametric Wilcoxon signed rank test to test whether the obtained coefficients of the pair of underlying targets were statistically significant. We used the R language (R Core Team 2017) to conduct all the statistical analyses.

Note that in solving traditional phonological problems like tone sandhi, the experimental design usually requires that the phonological environment to be more controlled. Compared to fitting the underlying pitch targets using data containing more contextual variation, fitting the underlying pitch targets using data with less contextual variation may be more likely to have overfitting problems. However, the current paper used the analysis of underlying pitch targets to better examine the tone sandhi rule application on different types of wug words and compare the underlying pitch targets to those of real words. The aim of the current paper is not to derive invariant underlying pitch targets from various contexts, but to conduct a comparison of underlying pitch targets on real and wug words, where the tonal contexts are more controlled.

3. RESULTS

3.1 Surface Contours of the Third-Tone Sandhi

First, we analysed surface f0 contours of the first T3 in T3 + T3 (henceforth T33a) in real and wug words. Specifically, we used growth curve analysis to test whether surface f0 contours of T33a in real words were statistically different from those of T33a in each type of wug word. The fitted contours are plotted in Figure 2.

For the comparison of f0 contours of real words versus *AOAO wug words, no term was found statistically significant. For the comparison between real words versus AOAG (slope: $\chi^2(1) = 12.94$, $p < 0.001$) and real words versus AGAO (slope: $\chi^2(1) = 9.10$, $p = 0.003$), only the slope term was significantly different. For real words versus AGAG (quadratic: $\chi^2(1) = 19.45$, $p < 0.001$) and real words versus Pseudo words (quadratic: $\chi^2(1) = 13.39$, $p < 0.001$), only the quadratic term reached significance.

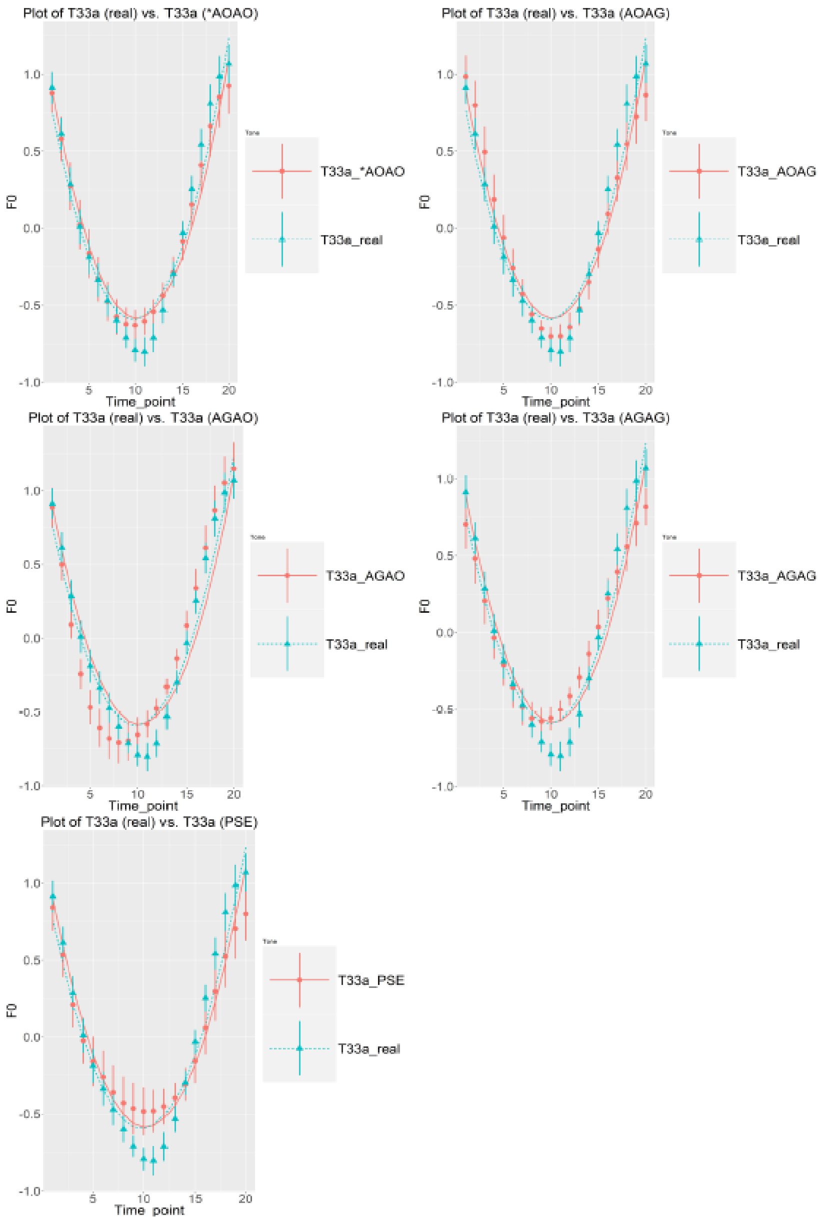

Figure 2 Growth curve analysis of T33a (real) versus T33a in each type of wug word; the mean and error bar of observed f0 at each time point are plotted; the solid and dotted lines represent the fitted values of the growth curve models.

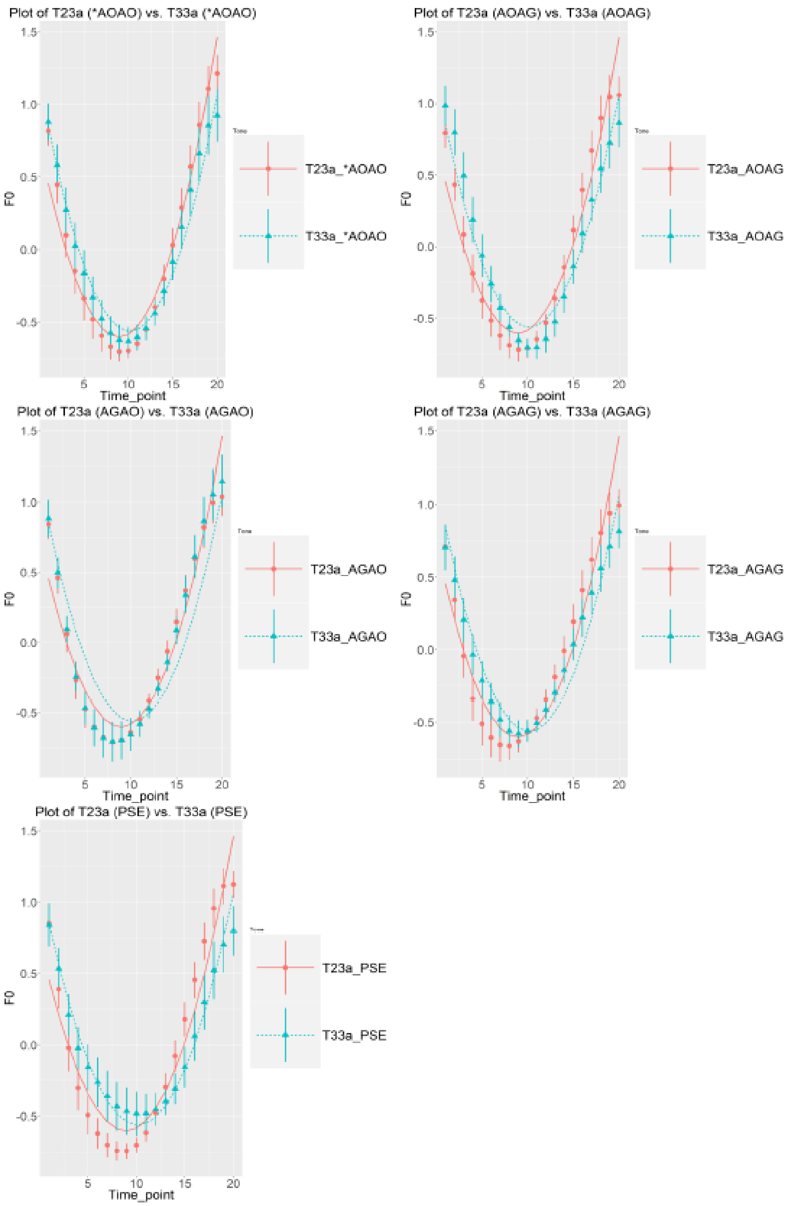

Figure 3 Growth curve analysis of T23a versus T33a in each type of wug word the mean and error bar of observed f0 at each time point are plotted; the solid and dotted lines represent the fitted values of the growth curve models.

Secondly, we found that surface f0 contours of T33a were statistically different from those of T23a in wug words in general since the slope term reached significance ($\chi^2(1) = 72.51$, $p < 0.001$). We further tested each type of wug word to assess whether the third tone sandhi T33a and T23a were neutralized in the surface f0 contour. Statistical significance was found in the slope term of the type *AOAO ($\chi^2(1) = 11$, $p < 0.001$), AOAG ($\chi^2(1) = 38.55$, $p < 0.001$), and AGAG ($\chi^2(1) = 16.13$, $p < 0.001$), and in both the slope ($\chi^2(1) = 30.06$, $p < 0.001$) and quadratic terms ($\chi^2(1) = 6.42$, $p = 0.01$) of the type pseudo words. Figure 3 shows a comparison of T23a and T33a in each type of wug word.

3.2 Underlying Pitch Targets of the Third-Tone Sandhi

Following the procedure proposed by Chen et al. (2017), we fitted four models to f0 values extracted from each pair (e.g., T33a (real) and T33a (*AOAO)) and chose the optimal model by referencing the AIC.

The model "com_2" as depicted in equation 8 was chosen for each pair, and the fitted parameters are presented in Table 1.

Table 1 Fitted coefficients of underlying pitch targets for T33a

	λ	d	a	b
T33a (real)	0.24	-0.098	4.87	-61.83
T33a (*AOAO)	0.18	-0.13	7.33	-112.31
T33a (AOAG)	0.17	-0.13	7.44	-116.53
T33a (AGAO)	0.15	-0.11	6.94	-114.74
T33a (AGAG)	0.15	-0.09	5.24	-84.48
T33a (pseudo)	0.16	-0.13	7.92	-134.31

We proceeded to compare the underlying pitch targets of T33a in real words with those in each type of wug word. The results showed that underlying pitch targets were different only among three types of wug words, as reported in Table 2. All of the three types of wug words showed significant differences in the quadratic term and some types in the slope and intercept terms.

Finally, in order to test whether T23a and T33a are neutralized in the underlying pitch targets, we fitted four models to f0 values extracted from each pair (e.g., T23a (AOAG) and T33a (AOAG)) and chose the optimal model. The model "com2" for each pair was chosen and the fitted parameters are shown in Table 3. We proceeded to compare the underlying pitch targets of T23a and T33a in each type of wug word. The results showed that underlying pitch targets were different only among pseudo words. The quadratic (V= 5, p = 0.001), slope (V= 97, p =0.003) and intercept (V= 13, p = 0.01) terms all reached significance. All the other types of wug words showed neutralization of T23a and T33a in their underlying pitch targets.

Table 2 A comparison of underlying pitch targets of T33a on real versus wug words

Pair	Quadratic term	Slope	Intercept
Real versus AGAO	(D) V = 14, p = 0.01	(D) V = 91, p = 0.01	(S) V = 30, p = 0.17
Real versus AGAG	(D) V = 11, p = 0.0067	(D) V = 98, p = 0.002	(D) V = 20, p = 0.04
Real versus Pseudo	(D) V = 12, p = 0.009	(D) V = 93, p = 0.009	(D) (marginal significance) V = 25, p = 0.09

Same: S; Different: D

Table 3 Fitted coefficients of underlying pitch targets for T23a

Tone	λ	d	a	b
T23a (*AOAO)	0.17	-0.14	8.16	-126.17
T23a (AOAG)	0.20	-0.15	8.06	-112.82
T23a (AGAO)	0.16	-0.11	6.20	-98.63
T23a (AGAG)	0.15	-0.12	6.88	-112.74
T23a (pseudo)	0.18	-0.17	9.08	-132.61

3.3 Surface Contours of Half-Third Sandhi

We tested whether f0 contours of T3 in the disyllabic combinations of T3 + T1, T3 + T2 and T3 + T4 (T3T124a henceforth) were statistically different in real words and in each type of wug word. Figure 4 shows a plot of T3T124a for all types of words.

The results showed significant differences on the quadratic term for real words versus *AOAO ($\chi^2(1)$ = 4.39, p = 0.036) and real words versus pseudo words ($\chi^2(1)$ = 4.52, p = 0.03). For a comparison between real words versus *AOAO ($\chi^2(1)$ = 12.78, p < 0.001), real words versus AOAG ($\chi^2(1)$

= 4.61, p = 0.032) and real words versus AGAO ($\chi^2(1) = 4.74$, p = 0.029), the slope term showed statistical significance. For a comparison between real words versus AOAG ($\chi^2(1) = 4.12$, p = 0.04), real words versus AGAO ($\chi^2(1) = 8.22$, p = 0.004), real words versus AGAG ($\chi^2(1) = 28.69$, p < 0.001) and real words versus pseudo words ($\chi^2(1) = 32.50$, p < 0.001), the intercept term reached significance. The degree of incompleteness is shown in wug words and is similar as that found in the third tone sandhi rule: the rule application to pseudo words resulted in significantly different surface f0 contours from those arising from the application to real words, on higher-order quadratic terms. However, greater dissimilarities in stimuli between real words and wug words did not all lead to more different surface f0 contours in half-third sandhi rule application. For example, the wug type AGAG showed fewer differences on the surface contour than in AOAG and AGAO.

3.4 Underlying Pitch Targets of Half-Third Sandhi

We fitted four models to f0 values extracted from each pair (e.g., T3T124a (real) and T3T124a (*AOAO)) and chose the optimal model. The model "com2" for each pair was chosen and the fitted parameters are presented in Table 4.

Table 4 Fitted coefficients of underlying pitch targets for T3T124a

Tone	λ	d	a	b
T3T124a (Real)	0.037	0.28	-18.27	-6.21
T3T124a (*AOAO)	0.04	0.29	-17.85	-6.70
T3T124a (AOAG)	0.035	0.26	-17.85	-5.47
T3T124a (AGAO)	0.037	0.28	-18.26	-6.45
T3T124a(AGAG)	0.036	0.27	-18.25	-5.80
T3T124a (pseudo)	0.04	0.32	-19.73	-8.57

We tested the underlying pitch target of T3T124a in real words and of T3T124a in all types of wug words. The underlying pitch targets on *AOAO, AOAG and AGAO were not statistically different from those of the real words. The intercept term of the underlying pitch target in AGAG (V= 13, p = 0.04) and the slope (V= 17, p = 0.09) term of the underlying target in pseudo words were significantly or marginally different from that on real words.

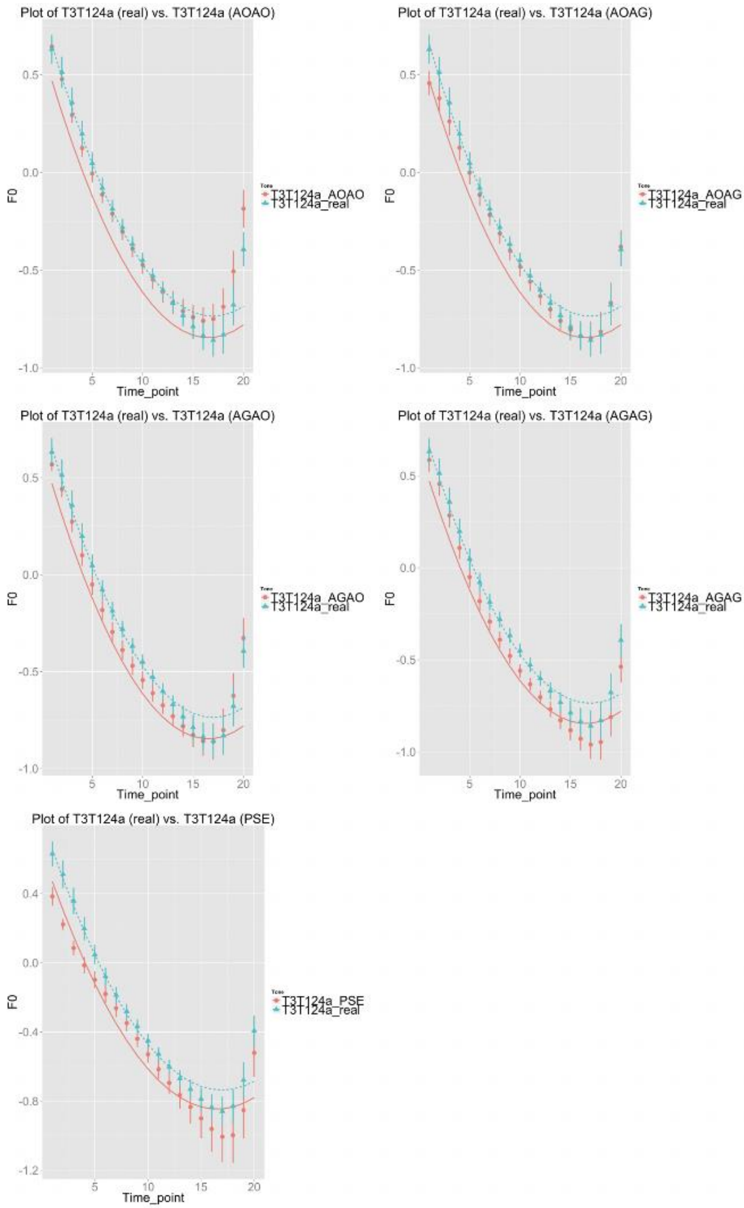

Figure 4 Growth curve analysis of T3T124a (real) versus T3T124a on each type of wug word. The mean and error bar of observed f0 at each time point are plotted. The solid and dotted lines represent the fitted values of the growth curve models.

4. DISCUSSION

4.1 The Third Tone Sandhi Rule Application

Our results suggest that the application of the third tone-sandhi rule might not be complete in some types of wug words. Based on the analysis of surface f0 contours, the degree of tone sandhi rule application varies for different types of wug words. Except for *AOAO, all the other types of wug words exhibited significant differences.

The more distant phonetically and morphologically the stimuli were from real words, i.e., containing more non-legitimate components in Mandarin, the more incomplete the third tone sandhi rule application became. Recall that the type *AOAO is an illegal sequence of two legit morphemes, whereas AGAO and AOAG both contain a non-legitimate morpheme AG (accidental gap involving missing tonal combination with legal segments).

Both AGAO and AOAG demonstrated significant differences in the lower-order slope terms. When both morphemes are non-existing in Mandarin lexicon, as in AGAG or in pseudo words, the higher-order quadratic terms showed significant differences. In sum, the degree of completeness of the third tone sandhi rule on different types of wug words can be represented as *AOAO > AGAO, AOAG > AGAG, pseudo (">" means more complete application and more similar to the application on real words).

Xu (2005) proposes that speech melody could be understood better if two aspects are considered: (1) communicative information that speech conveys; (2) biophysical properties of human articulators. The surface acoustic forms are achieved by converting the communicative function such as tones and focus through target approximation. Underlying pitch targets are assumed to be associated with communicative functions such as lexical tones, the driving force of the linear system (Prom-On et al. 2009). Then, significant differences in the underlying pitch targets indicate differences in the communicative function, i.e., tones in our case. If the underlying pitch targets are different, then the implementation of them, which are surface f0 contours are also likely to be different. However, surface f0 contours are more subject to phonetic perturbation (Chen et al. 2017) or contextual variation (Xu 1993, 1994; Xu and Wang, 2001). Therefore, even if the surface f0 contours are different, the underlying pitch targets may still remain similar with no statistical difference. The analyses of underlying pitch targets further demonstrated that the application of tone sandhi rules on three types of wug words (AGAO, AGAG and pseudo words) was not complete since not only the surface f0 contours differed from those of real words, but also the underlying pitch targets were different. Therefore, we can infer from the results that although there might be computation involved in applying the third tone sandhi rules on wug words, the degree of application differs on different types of wug words.

Moreover, this study is among the first to test whether the sandhi T3 and T2 were neutralized on all types of wug words. For the surface contours, the two tones were only neutralized for AGAO types of wug words. For the underlying pitch targets, the sandhi T3 and T2 were different only on pseudo words, where all the quadratic, slope and intercept terms were found to be significant. The results confirmed that the application of the third tone sandhi rule was the most different on the real versus pseudo words because (1) the sandhi T3 (T33a) differed on real versus pseudo words on the surface f0 contours and underlying pitch targets; (2) the sandhi T3 and T2 were neutralized in the underlying pitch targets on all other types of wug words but not on pseudo words.

4.2 The Half-Third Sandhi Rule Application

Similar to the third tone sandhi rule application, we also examined surface f0 contours and underlying pitch targets of the half-third sandhi rule application on real and all types of wug words. We found that the application of this rule was not complete for some types of wug words. For the surface f0 contours, all types of wug words showed significant differences from real words. However, the degree of incompleteness in the rule application did not increase as consistently as the third tone sandhi rule when more non-real components were involved. Specifically, both *AOAO and pseudo words showed significant differences on the higher order quadratic term. The types AOAG and AGAO had significantly different slope and intercept terms, but the type AGAG involving two non-real morphemes only showed a significantly different intercept term.

The analysis on the underlying pitch targets revealed significant or marginally significant differences on two types of wug words AGAG and pseudo words. The conclusion from the results of both surface f0 contour and the underlying pitch targets was that computation mechanism might be involved in the application on wug words, but incompleteness was shown on wug words that are made up of more non-real components. A summary of degree of application in the two sandhi rules is presented in Table 5.

Table 5 Degree of application completeness across stimuli types and levels of analysis

(a) The third tone sandhi rule

Stimuli type	Neutralization with the sandhi tone in real syllables (surface f0 contour)	Neutralization with the sandhi tone in real syllables (underlying pitch target)
Wug:*AOAO	Y	Y
Wug: AOAG	N	Y
Wug: AGAO	N	N
Wug: AGAG	N	N
Wug: pseudo	N	N

(b) The half-third tone sandhi rule

Stimuli type	Neutralization with the sandhi tone in real syllables (surface f0 contour)	Neutralization with the sandhi tone in real syllables (underlying pitch target)
Wug:*AOAO	N	Y
Wug: AOAG	N	Y
Wug: AGAO	N	Y
Wug: AGAG	N	N
Wug: pseudo	N	N

4.3 Theoretical Implications

Neutralization of Mandarin T2 and the sandhi T3 in tone sandhi cases has been investigated extensively. Previous studies reported that the sandhi T3 and T2 were not neutralized on the surface f0 contours. The sandhi T3 was either reported to have a lower mean f0 than T2 (e.g., Peng 2000; Shen 1992) or it differed from T2 in the magnitude of f0 rise and the percentage of f0 rise duration (Yuan and Chen 2014). Zhang and Peng (2013) found that the sandhi T3 showed significantly lower f0 than T2 on pseudo words but not on real words. More recently, Xu and Prom-on (2014) compared the synthesis accuracies obtained from simulation for three hypotheses about the result of the tone sandhi rule: (1) T3 stays unchanged; (2) T3 becomes another derivational tone; or (3) T3 becomes T2. They found that it is unlikely that T3 remains unchanged, and T3 may become another derivational tone or T2. Chen et al. (2019) also found neutralization in the underlying pitch targets of T2 and T3 despite significant differences on surface f0 contours.

The current study tested both the surface f0 contours and the underlying pitch targets of the sandhi T3 and T2 on various types of wug words for the first time. Our results confirmed Zhang and Peng's (2013) findings of non-neutralization of surface f0 contours on pseudo words. More specifically, we found that the sandhi T3 and T2 were only neutralized for AGAO on the surface f0 contours. However, the underlying pitch target of the sandhi T3 and that of T2 were significantly different on pseudo words, but not other types of wug words. The results also showed that the types of wug words have a strong influence on the results of neutralization. If the wug words only contain real morphemes (AO) or some accidental gaps (AG) that are made up of real syllables with a tone it cannot bear accidentally, then the third tone sandhi rule is likely to apply more categorically with neutralized underlying pitch targets despite variation on the surface. If a wug word is genuinely pseudo, i.e., composed of illegal combination of Mandarin consonants and vowels, then the tone sandhi rule may not apply categorically, nor result in tonal neutralization, in either surface f0 contours or underlying pitch targets. In sum, the third tone sandhi rule applied more categorically on wug words consisting of legitimate CV syllables (AO or AG type of wug words) than on genuine pseudo words.

Moreover, our results may shed light on the debate of lexical versus computation mechanism. Recall that lexical mechanism refers to that surface sandhi tone contours are stored as exemplars. The computation mechanism, on the other hand, holds that the sandhi and non-sandhi forms (phonological alternation) are stored as part of the abstract tonal category. Our results suggested that for surface f0 contours, the third tone sandhi rule applied similarly on *AOAO, but not other types of wug words. The underlying pitch targets were not neutralized on three types of wug words, when we compared the underlying targets of real words and those of AGAO, AGAG and pseudo words. Taking into account the results of both surface f0 contours and underlying pitch targets, we may conclude that computation is more complete in wug words that have only a few non-legitimate components such as in *AOAO where real morphemes are put in non-existing combination to form a word. For the half-third sandhi tone, all types of wug words showed significant differences from real words on the surface, but only the underlying pitch targets of AGAG and pseudo words were different from real words. The results support three hypothesized situations we proposed in the introduction section (Figure 1). We can infer that computation mechanism is involved in applying these two rules on wug words. However, the computation of these two tone sandhi rules was not complete either for surface f0 contours or for underlying pitch targets if the wug words involved more non-real components such as AGAG (two accidental gaps) or non-existing syllables (pseudo words). For these two types of wug words, both the surface f0 contours and underlying pitch targets showed significant differences. The results we found may be attributed to that lexical representations being more available in encoding and accessing sandhi tones for legitimate real morphemes (AO) than AG types of wug or pseudo morphemes. Therefore, our acoustic analysis yielded results that indicate that more than one mechanism may be playing a role for native speakers in encoding and accessing sandhi tones.

Finally, we improved the experimental methods to effectively compare the results of these two tone sandhi rules. Regarding the third tone sandhi rule application, Zhang and Lai (2010) showed consistent differences in f0 contours between real words and wug words. However, such differences were not found consistently in the application of the half-third sandhi rule on the same real and wug words. They argued that the third tone sandhi rule is phonetically less motivated than the half-third sandhi tone, and then proposed that the wug-testing paradigm could help test how relevant phonetics is with respect to synchronic phonology. Their results of less accuracy in the third tone sandhi rule application than on the half-third sandhi tone rule indicated a bias against phonetically less motivated rules (i.e., the third tone sandhi rule in this case). We have improved the design of stimuli in the current study by adding more types of wug words and controlling phonetic environments across those types in most cases to reduce intrinsic F0 effects. The repeated-measures ANOVA used by Zhang and Lai (2010) has also been criticized for not being able to model individual deviation in the slope over time (Gibbons, Hedeker and DuToit 2010). We improved the statistical analyses by using a more recent statistical technique, growth curve analysis, to test the surface f0 contours in addition to a statistical procedure to test underlying pitch targets. The limitations of repeated measures of ANOVA exist for both longitudinal variations discussed by Gibbon et al. (2010) and the temporal variations in tone production. Essentially, both types of data have correlated measurements across time. Repeated measures ANOVA has a more restricted assumption about the structure of the correlation, which is called the compound-symmetric form, where the growth curve analysis has the advantage of not assuming the compound-symmetric form. Moreover, compared to the repeated measures ANOVA used in Zhang and Lai (2010), where only one interaction term was used to determine the tonal shape, we included both linear and quadratic terms to better determine the shape of tonal contours, in addition to one term determining the average of f0 values using growth curve analysis. An anonymous reviewer pointed out although Zhang and Lai (2010) did not add quadratic terms, these terms could also be subjected to ANOVA.

To compare the results of the two methods, we tested on T3 using traditional methods employed by Zhang and Lai (2010). They conducted a two-way repeated-measures ANOVA with two independent variables Word Group and Point. The Word Group variable indicates that the pitch values are from two different group (e.g., real versus wug). Therefore, a significant main effect suggests that the two word groups have different average pitches. The Point variable stands for the time points where f0 values are extracted. Thus, a significant interaction between the variables Word Group and Point suggests different shapes of the two curves. We have listed the results in Table S1 and Table S2 in the appendix. Using growth curve analysis, we may have two terms (quadratic and linear terms) to determine the shape of f0 contours, where there is only one term (interaction) claimed to determine the shape of f0 contours. Using repeated measures ANOVA, there was no difference between real words versus *AOAG, where the growth curve analysis yielded significant slope differences between the two. Using repeated measures ANOVA, we could obtain a similar trend turning from no significance to significance in f0 shapes as the wug type became less word-like, but the results were not so clear as the growth curve analysis, where more significant differences are indicated by significances in quadratic terms. Note that we do not intend to argue that obtaining more significant terms identified by growth curve analysis is necessarily an advantage of it.

From our results, we noted that both the third tone sandhi rule and the half-third sandhi rule showed significant differences between real words and most types of wug words. Further examination on the underlying pitch target revealed that more types of wug words were significantly different for the third tone sandhi rule than for the half-third sandhi. These new results are also important evidence in support of Zhang and Lai's (2010) claim that the third sandhi is less phonetically motivated and applies less accurately.

5. CONCLUSIONS

Our study set out to assess the degree of application of two Mandarin tone sandhi rules. This is among the first to apply quantitative methods to examine both the surface f0 contours and underlying pitch targets in order to examine applications of Mandarin tone sandhi rules on wug words. The study used five types of wug words and strictly controlled consonants and vowels in the design of stimuli. We found that the more dissimilar the stimuli were from real words, the more incomplete the third tone sandhi rule application became. However, the relationship between the degree of incompleteness and types of stimuli was not as consistent in the application of the half-third sandhi rule.

Moreover, this study offers new insights in many theoretical aspects. With respect to the debate of categorical and gradient view of the third tone sandhi rule, our results showed that for tone sandhi rule application on wug words, the types of wug words strongly affected whether the sandhi T3 and T2 neutralized in the underlying pitch target. On wug words containing real morphemes or accidental gaps, tone sandhi rule applied more categorically than on pseudo words.

In addition, our results confirmed the three hypothesized situations about the potential differences between tonal application on real and wug words. Our results also shed light on the debate of a lexical and computation mechanism in production of tone sandhi rules. The Mandarin tone sandhi rule application involves computation with neutralized underlying pitch targets between many types of wug words and real words. However, if the wug words contained more non-real morphemes, the computation was shown to be incomplete with non-neutralized underlying pitch targets. In addition, based on a comparison of the underlying pitch targets in applying the third tone sandhi and half-third sandhi rule on reals versus wug words, we confirmed that there is a bias in application on wug words for the less phonetically motivated third tone sandhi rule.

APPENDIX

Table S1 A comparison of T33a on real versus different types of wug words using repeated measures ANOVA

Pair	Wd Gr (F0 average)	Point	Wd Gr Point (F0 shape)
Real versus *AOAO	(S) $F(1, 13) = 0$, $p = 1$	(D) $F(19, 247) = 31.13$, $p < 0.001$	(S) $F(19,247) = 0.60$, $p = 0.90$
Real versus AOAG	(S) $F(1, 13) = 0$, $p = 1$	(D) $F(19, 247) = 38.94$, $p < 0.001$	(S) $F(19, 247) = 1.16$, $p = 0.30$
Real versus AGAO	(S) $F(1, 13) = 0$, $p = 1$	(D) $F(19, 247) = 41.16$, $p < 0.001$	(D) $F(19, 247) = 1.95$, $p = 0.011$
Real versus AGAG	(S) $F(1, 13) = 0$, $p = 1$	(D) $F(19, 247) = 34.94$, $p < 0.001$	(D) $F(19, 247) = 1.90$, $p = 0.015$
Real versus Pseudo	(S) $F(1, 13) = 0$, $p = 1$	(D) $F(19, 247) =21.26$, $p < 0.001$	(D) $F(19, 247) = 2.29$, $p = 0.002$

Table S2 T2/T3 + T3 (Real Words)

T2+T3	English gloss	T3+T3	English gloss	Transcription
凡响	Common/ normal music	反响	Feedback	fæn ɕaŋ
携手	Hand in hand	写手	A writer	ɕjə ʂəu
埋土	To bury with dirt	买土	To buy dirt	mai thwu
仁者	A wise humanitarian	忍者	A ninja	ʐən ʈʂɤ
神美	Spiritual beauty/ fairness	审美	Aesthetics	ʂən məi
梅酒	Plum wine	美酒	Quality wine	məi tɕjəu
油水	Profit	有水	Containing water	jəu ʂwəi
唯美	Perfectionism	伟美	Magnificent	wəi məi

Table S3 T2/T3 + T3 (Wug words *AO-AO)

T2+T3	T3+T3	Transcription
凡两	反两	fæn l^{j}aŋ
携否	写否	ɕee fəu
埋苦	买苦	mai k^{hw}uu
仁舍	忍舍	ʐən ʂɤɤ
神垒	审垒	ʂən ləi
梅朽	美朽	məi ɕəu
油毁	有毁	jəu x^{w}əi
唯匪	伟匪	wəi fəi

Table S4 T2/T3 + T3 (Wug words AO-AG)

T2+T3	T3+T3	Transcription
AO-AG		
凡 niang	反 niang	fæn n^{j}aŋ
携 cou	写 cou	ɕee ts^{h}əu
埋 su	买 su	mai s^{w}uu
仁 me	忍 me	ʐən mɤɤ
神 hei	审 hei	ʂən xəi
梅 diu	美 diu	məi t^{j}əu
油 dui	有 dui	jəu t^{w}əi
唯 hei	伟 hei	wəi xəi

Table S5 T2/T3 + T3 (Wug words AG-AO)

T2+T3	T3+T3	Transcription
suan 响	suan 响	s^{w}æn ɕaŋ
mie 手	mie 手	m^{j}ee ʂəu
sai 土	sai 土	sai t^{hw}uu
sen 者	sen 者	sən ʈʂɤɤ
te 美	te 美	t^{h}ɤɤ məi
hei 酒	hei 酒	xəi tɕəu
diu 水	diu 水	t^{j}əu $ʂ^{w}$əi
dui 美	dui 美	t^{w}əi məi

Table S6 T2/T3 + T3 (Wug words AG-AG)

T2+T3	T3+T3	Transcription
suan niang	suan niang	sʷæn nʲaŋ
mie cou	mie cou	mʲee tsʰəu
sai su	sai su	sai sʷuu
sen me	sen me	sən mɤɤ
te hei	te hei	tʰɤɤ xəi
hei diu	hei diu	xəi tʲəu
diu dui	diu dui	tʲəu tʷəi
dui hei	dui hei	tʷəi xəi

Table S7 T2/T3 + T3 (Pseudo words)

T2+T3	T3+T3	Transcription
muan diang	muan diang	mʷæn tʲaŋ
fie bou	fie bou	fʲee pəu
fai fong	fai fong	fai fʷuŋ
ten fe	ten fe	tʰən fɤɤ
fe rei	fe rei	fɤɤ ʐəi
rei miu	rei miu	ʐəi mʲəu
fiu lui	fiu lui	fʲəu lʷəi
lui rei	lui rei	lʷəi ʐəi

Table S8 T3 + T1/T2/T4 (Real words)

Tones	Chinese diagram	Transcription		Tones	Chinese diagram	Transcription	
3+1	每天 every day	məi	tʰʲæn	3+4	美丽 beauty	məi	lʲii
	转身 turn around	tʂʷæn	ʂən		转动 roll	tʂʷæn	tʷuŋ
	产生 produce	tʂʰæn	ʂəŋ		反正 anyhow	fæn	tʂəŋ
	首先 first	ʂəu	ɕæn		手术 surgery	ʂəu	ʂʷuu
	整天 all day	tʂəŋ	tʰʲæn		整个 entire	tʂəŋ	kɤɤ
	可惜 unfortunate	kʰɤɤ	ɕii		可怕 terrible	kʰɤɤ	pʰaa
	假装 pretend	tɕaa	tʂʷaŋ		假币 Counterfeit money	tɕaa	pʲii
	水晶 crystal	ʂʷəi	tɕəŋ		水面 water surface	ʂʷəi	mʲæn

Table S8 (*continued*)

Tones	Chinese diagram	Transcription	
3+2	美国 U.S.A	məi	kʰʷoo
	转移 transfer	ʈʂʷæn	ii
	反而 contrary	fæn	ər
	手足 brothers	ʂəu	tsʷuu
	整洁 tidy	ʈʂəŋ	tɕee
	可能 possible	kʰɤɤ	nəŋ
	假如 if	tɕaa	ʐʷuu
	水平 level	ʂʷəi	pʰʲəŋ

Table S9 T3 + T1/T2/T4 (Wug words *AO+AO)

Tones	Chinese diagram	Transcription		Tones	Chinese diagram	Transcription	
3+1	每聪	məi	tsʰʷuŋ	3+4	美特	məi	tʰɤɤ
	转跟	ʈʂʷæn	kən		转不	ʈʂʷæn	pʷuu
	产参	ʈʂʰæn	tsʰæn		反脆	fæn	tsʰʷəi
	首插	ʂəu	ʈʂʰaa		手否	ʂəu	fəu
	整催	ʈʂəŋ	tsʰʷəi		整哈	ʈʂəŋ	xaa
	可慌	kʰɤɤ	xʷaŋ		可害	kʰɤɤ	xai
	假森	tɕaa	sən		假略	tɕaa	lᶣee
	水黑	ʂʷəi	xəi		水奏	ʂʷəi	tsəu
3+2	美别	məi	pʲee				
	转仇	ʈʂʷæn	ʈʂʰəu				
	反存	fæn	tsʰʷən				
	手能	ʂəu	nəŋ				
	整狼	ʈʂəŋ	laŋ				
	可勾	kʰɤɤ	kəu				
	假林	tɕaa	lʲəŋ				
	水朝	ʂʷəi	ʈʂʰau				

Table S10 T3 + T1/T2/T4 (Wug words AO+AG)

Tones	Chinese diagram		Transcription		Tones	Chinese diagram		Transcription	
3+1	每	den	məi	tən	3+4	美	dei	məi	təi
	转	liang	ʈʂʷæn	lʲaŋ		转	zhua	ʈʂʷæn	ʈʂʷaa
	产	mei	ʈʂʰæn	məi		反	keng	fæn	kʰəŋ
	首	mu	ʂəu	mʷuu		手	nin	ʂəu	nʲin
	整	mai	ʈʂəŋ	mai		整	qiu	ʈʂəŋ	tɕʰəu
	可	kuo	kʰɤɤ	kʰʷoo		可	ca	kʰɤɤ	tsʰaa
	假	re	tɕaa	ʐɤɤ		假	sen	tɕaa	sən
	水	shun	ʂʷəi	ʂʷən		水	mang	ʂʷəi	maŋ
3+2	美	mie	məi	mʲee					
	转	ta	ʈʂʷæn	tʰaa					
	反	ka	fæn	kʰaa					
	手	suan	ʂəu	sʷæn					
	整	dui	ʈʂəŋ	tʷəi					
	可	diu	kʰɤɤ	tʲəu					
	假	te	tɕaa	tʰɤɤ					
	水	cui	ʂʷəi	tsʰʷəi					

Table S11 T3 + T1/T2/T4 (Wug words AG+AO)

Tones	Chinese diagram		Transcription		Tones	Chinese diagram		Transcription	
3+1	hei	天	xəi	tʰʲæn	3+4	hei	丽	xəi	lʲii
	shuan	身	ʂʷæn	ʂən		shuan	动	ʂʷæn	tʷuŋ
	pan	生	pʰæn	ʂəŋ		pan	正	pʰæn	ʈʂəŋ
	cou	先	tsʰəu	ɕæn		cou	术	tsʰəu	ʂʷuu
	sen	天	sən	tʰʲæn		sen	个	sən	kɤɤ
	te	惜	tʰɤɤ	ɕii		te	怕	tʰɤɤ	pʰaa
	xia	装	ɕaa	ʈʂʷaŋ		xia	币	ɕaa	pʲii
	dui	晶	tʷəi	tɕəŋ		dui	面	tʷəi	mʲæn
3+2	hei	国	xəi	kʰʷoo					
	shuan	移	ʂʷæn	ii					
	pan	而	pʰæn	ər					
	cou	足	tsʰəu	tsʷuu					
	sen	洁	sən	tɕee					
	te	能	tʰɤɤ	nəŋ					
	xia	如	ɕaa	ʐʷuu					
	dui	平	tʷəi	pʰʲəŋ					

Table S12 T3 + T1/T2/T4 (Wug words AG+AG)

Tones	Chinese diagram		Transcription		Tones	Chinese diagram		Transcription	
3+1	hei	den	xəi	tən	3+4	hei	dei	xəi	təi
	shuan	liang	ʂʷæn	lʲaŋ		shuan	zhua	ʂʷæn	ʈʂʷaa
	pan	mei	pʰæn	məi		pan	keng	pʰæn	kʰəŋ
	cou	mu	tsʰəu	mʷuu		cou	nin	tsʰəu	nʲin
	sen	mai	sən	mai		sen	qiu	sən	tɕʰəu
	te	kuo	tʰɤɤ	kʰʷoo		te	ca	tʰɤɤ	tsʰaa
	xia	re	ɕaa	ʐɤɤ		xia	sen	ɕaa	sən
	dui	shun	tʷəi	ʂʷən		dui	mang	tʷəi	maŋ
3+2	hei	mie	xəi	mʲee					
	shuan	ta	ʂʷæn	tʰaa					
	pan	ka	pʰæn	kʰaa					
	cou	suan	tsʰəu	sʷæn					
	sen	dui	sən	tʷəi					
	te	diu	tʰɤɤ	tʲəu					
	xia	te	ɕaa	tʰɤɤ					
	dui	cui	tʷəi	tsʰʷəi					

Table S13 T3 + T1/T2/T4 (Wug words Pseudo words)

Tones	Chinese diagram		Transcription		Tones	Chinese diagram		Transcription	
3+1	lui	fai	lʷəi	fai	3+4	lui	fai	lʷəi	fai
	fie	diang	fʲee	tʲaŋ		fie	mui	fʲee	mʷəi
	ten	fie	tʰən	fʲee		ten	fong	tʰən	fʷuŋ
	bou	fe	pəu	fɤɤ		bou	ten	pəu	tʰən
	len	mui	lən	mʷəi		len	diang	lən	tʲaŋ
	be	fiu	pə	fʲəu		be	fie	pə	fʲee
	fia	ten	fʲaa	tʰən		fia	bou	fʲaa	pəu
	nui	len	nʷəi	lən		nui	fe	nʷəi	fɤɤ
3+2	lui	fong	lʷəi	fʷuŋ					
	fie	ten	fʲee	tʰən					
	ten	fe	tʰən	fɤɤ					
	bou	fai	pəu	fai					
	len	mui	lən	mʷəi					
	be	fie	pə	fʲee					
	fia	bou	fʲaa	pəu					
	nui	diang	nʷəi	tʲaŋ					

REFERENCES

BERKO, Jean. 1958. The child's learning of English morphology. *Word* 14(2): 150–177.

CHAO, Yuen Ren. 1948. *Mandarin Primer: An Intensive Course in Spoken Chinese*. Cambridge, MA: Harvard University Press.

———. 1965. *A Grammar of Spoken Chinese.* Oakland: University of California Press.

CHEN, Matthew Y. 2000. *Tone Sandhi: Patterns across Chinese Dialects* (*Cambridge Studies in Linguistics* 92). Newcastle upon Tyne: Cambridge University Press.

CHEN, Si. 2019. Phonological representations based on statistical modelling in tonal languages. In *Prosodic Studies: Challenges and Prospects,* edited by Hongming Zhang and Youyong Qian, 159–197. Oxfordshire: Routledge.

CHEN, Si, Caroline R. Wiltshire, Bin Li, and Ratree Wayland, R. 2019. A quantitative analysis of tone sandhi in Standard Mandarin and Nanjing Mandarin based on surface pitch contours and underlying pitch targets. *International Journal of Chinese Linguistics* 6(2): 183–220.

CHEN, Si, Caroline R. Wiltshire, and Bin Li. 2018. An updated typology of tonal coarticulation properties. *Taiwan Journal of Linguistics* 16(2): 79–114.

CHEN, Si, Caicai Zhang, Adam G. McCollum, and Ratree Wayland. 2017. Statistical modelling of phonetic and phonologised perturbation effects in tonal and non-tonal languages. *Speech Communication* 88 (April): 17–38.

CHENG, Chin-Chuan. 1968. English stresses and Chinese tones in Chinese sentences. *Phonetica* 18(2): 77–88.

GIBBONS, Robert D., Donald Hedeker, and Stephen DuToit. 2010. Advances in analysis of longitudinal data. *Annual Review of Clinical Psychology* 6: 79–107.

HSIEH, Hsin-I. 1970. The psychological reality of tone sandhi rules in Taiwanese. In *6th Meeting of the Chicago Linguistic Society,* 489–503.

———. 1975. How generative is phonology. *The Transformational-Generative Paradigm and Modern Linguistic Theory*: 109–144.

———. 1976. On the unreality of some phonological rules. *Lingua* 38(1): 1–19.

HSIAO, Yuchau E. 1992. *Syntax, Rhythm and Tone: A Triangular Relationship*. Taipei: Crane.

HUET, Sylvie, Anne Bouvier, Marie-Anne Poursat, and Emmanuel Jolivet. 2006. *Statistical Tools for Nonlinear Regression: A Practical Guide with S-PLUS and R Examples*. Berlin: Springer Science & Business Media.

JANGJAMRAS, Jirapat. 2011. Perception and production of English lexical stress by Thai speakers. PhD diss., University of Florida.

KIM, Kevin, and Neil Timm. 2006. *Univariate and Multivariate General Linear Models: Theory and Applications with SAS*. Boca Raton: CRC Press.

KRATOCHVIL, Paul. 1986. The case of the third tone. In *Wang Li Memorial Volumes, English Volume,* edited by The Chinese Language Society of Hong Kong, 253–277. Hong Kong: Joint Publishing.

MIRMAN, Daniel. 2016. *Growth Curve Analysis and Visualization Using R*. Boca Raton: CRC Press.

MIRMAN, Daniel, James A. Dixon, and James S. Magnuson. 2008. Statistical and computational models of the visual world paradigm: Growth curves and individual differences. *Journal of Memory and Language* 59(4): 475–494.

MYERS, James, and Jane Tsay. 2003. A formal functional model of tone. *Language and Linguistics* 4(1): 105–138.

PENG, Shu-Hui. 2000. Lexical versus 'phonological' representations of Mandarin sandhi tones. In *Papers in Laboratory Phonology V: Acquisition and the Lexicon*, edited by M. Broe and J. Pierrehumbert, 152–167. Cambridge: Cambridge University Press.

PROM-ON, Santitham, Yi Xu, and Bundit Thipakorn. 2009. Modeling tone and intonation in Mandarin and English as a process of target approximation. *The Journal of the Acoustical Society of America* 125(1): 405–424.

R Core Team. 2017. R: A Language and Environment for Statistical Computing. *R Foundation for Statistical Computing*. https://www.R-project.org.

RAUDENBUSH, Stephen W., and Anthony S. Bryk. 2002. *Hierarchical Linear Models: Applications and Data Analysis Methods*, 2nd ed. Newbury Park: Sage.

ROSE, Phil. 1987. Considerations in the normalisation of the fundamental frequency of linguistic tone. *Speech Communication* 6(4): 343–352. doi: 10.1016/0167-6393(87)90009-4.

SHEN, Xiaonan Susan. 1990. Tonal coarticulation in Mandarin. *Journal of Phonetics,* no.18: 281–295.

———. 1992. On tone sandhi and tonal coarticulation. *Acta Linguistica Hafniensia* 25(1): 83–94.

SHIH, Chih-Lin. 1986. The prosodic domain of tone sandhi in Chinese. PhD diss., University of California, San Diego.

SUN, Xuejing. 2001. Predicting underlying pitch targets for intonation modeling. In *Proceeding of the 4th ISCA Tutorial and Research Workshop (ITRW) on Speech Synthesis*, 143–148. ISCA Archive. https://www.isca-speech.org/archive_open/archive_papers/ssw4/ssw4_126.pdf, August 29 – September 1, 2001.

XU, Yi. 1993. Contextual tonal variations in Mandarin. PhD diss., University of Connecticut.

———. 1994. Asymmetry in contextual tonal variation in Mandarin. *Advances in the Study of Chinese Language Processing* 1: 383–396.

———. 1997. Contextual tonal variations in Mandarin. *Journal of Phonetics* 25(1): 61–83.

———. 2005. Speech melody as articulatorily implemented communicative functions. *Speech Communication* 46(3): 220–251.

———. 2013. *ProsodyPro—A Tool for Large-Scale Systematic Prosody Analysis*. Laboratoire Parole et Langage, France.

XU, Yi, and Q. Emily Wang. 2001. Pitch targets and their realization: Evidence from Mandarin Chinese. *Speech Communication* 33(4): 319–337.

XU, Yi, and Santitham Prom-On. 2014. Toward invariant functional representations of variable surface fundamental frequency contours: Synthesizing speech melody via model-based stochastic learning. *Speech Communication* 57: 181–208.

WAN, I-Ping, and Jeri Jaeger. 1998. Speech errors and the representation of tone in Mandarin Chinese. *Phonology* 15(3): 417–461.

YUAN, Jiahong, and Yiya Chen. 2014. 3rd tone sandhi in Standard Chinese: A corpus approach. *Journal of Chinese Linguistics* 42(1): 218–237.

ZEE, Eric. 1980. A spectrographic investigation of Mandarin tone sandhi. *UCLA Working Papers in Phonetics* 49(9): 98–116.

ZHANG, Caicai, and Gang Peng. 2013. Productivity of Mandarin third tone sandhi: A wug test. In *Eastward Flows the Great River: Festschrift in Honor of Prof. William SY. Wang on His 80th Birthday*, edited by Gang Peng and Feng Shi, 256–282. Hong Kong: City University of Hong Kong Press.

ZHANG, Caicai, Quansheng Xia, and Gang Peng. 2015. Mandarin third tone sandhi requires more effortful phonological encoding in speech production: Evidence from an ERP study. *Journal of Neurolinguistics* 33: 149–162.

ZHANG, Jie, and Yuwen Lai. 2010. Testing the role of phonetic knowledge in Mandarin tone sandhi. *Phonology* 27(1): 153–201.

ZHANG, Jie, Yuwen Lai, and Craig Sailor. 2011. Modeling Taiwanese speakers' knowledge of tone sandhi in reduplication. *Lingua* 121(2): 181–206.

ZHU, Xiaonong. 1999. Shanghai tonetics. PhD diss., Australian National University.

两组变调规则应用程度的统计建模

陈思 **李彬**
香港理工大学 香港城市大学

摘要
本研究探讨了两组普通话变调规则在真词和假词上的应用。我们设计了语音和语义上不同种类的假词，并使用了两种统计方法，计算了表层基频曲线和底层声调目标。研究发现三声变调中假词越类似真词，规则应用程度越完整。然而，与三声变调相比，半三声变调规则的应用程度与假词是否类似真词并不一定相关。变调规则向来有范畴和渐进两种对立的观点，此项研究提供了新的思路。我们提出了三种假设情境，认为普通话变调规则涉及变调调型的计算，但应用于含多个非汉语语素的假词时，计算则变得不完整。最后，在两组变调规则中，我们进一步证实了与半三声变调相比，三声变调比较缺乏语音上的变调动机，因此真词与假词上的规则应用差异较大。

关键词
普通话三声变调 **增**长曲线分析 **底**层音调目标 **计**算机制 **假**词测试