

# Are Mandarin Sandhi Tone 3 and Tone 2 the Same or Different? The Results of Functional Data Analysis

Chierh Cheng<sup>1</sup>Jenn-Yeu Chen<sup>1</sup>Michele Gubian<sup>2</sup><sup>1</sup>Department of Chinese as a Second Language, National Taiwan Normal University, Taiwan<sup>2</sup>Center for Language and Speech Technology, Radboud University Nijmegen, Netherlands

chierh.cheng@gmail.com, psyjyc@ntnu.edu.tw, M.Gubian@let.ru.nl

## Abstract

Functional Data Analysis (FDA) is used to investigate Tone 3 sandhi in Taiwan Mandarin. Tone 3 sandhi is a tone change phenomenon that arises when two low tones occur in succession resulting in the first tone being realised as a rising tone. Tone dyads T2T3 and T3T3 were compared in terms of their  $F_0$  contours and velocity profiles. No difference was found between the  $F_0$  contours of the two tone dyads. In contrast, velocity profiles showed an increased difference in the later part of the seemingly similar rising movements of T2 and sandhi T3, with a steeper rising-falling movement in the former than the latter. This research demonstrates that FDA can elucidate more detail in the time dimension than that of conventional techniques commonly employed in the current phonetic literature.

## 1 Introduction

This paper aims to take a closer look at tone sandhi, and in particular Tone 3 sandhi, a traditional research question in Chinese linguistics. Tone sandhi occurs whenever a prosodic context is met for disyllabic units of certain tonal combinations; the tone of the first syllable is changed from its underlying tone in a systematic manner. Tone 3 sandhi and *yi* sandhi are the two major tone sandhi phenomena in Mandarin<sup>1</sup>. Tone 3 sandhi means that a low tone is realised as a rising one when followed by another low tone. For example, *zong<sub>3</sub> tong<sub>3</sub>* (‘總統’, means ‘president’) is realised as

*zong<sub>2</sub> tong<sub>3</sub>*. Tone 3 sandhi is also seen to be cyclically applied according to a hierarchical prosodic structure (Shih, 1997). The scope of *yi* sandhi, on the other hand, is somewhat limited. *yi* here means the number ‘one’ in Mandarin and its underlying high-level tone (Tone 1) undergoes respective changes according to its following morpheme being a classifier of a specific tone.<sup>2</sup> That is to say, *yi* sandhi requires matching phonological and morphological contexts.

Myers and Tsay (2003) supplemented their study on tone sandhi in general with *yi* sandhi and concluded that more empirical data is needed to elucidate the most ubiquitous but still unclear nature of Tone 3 sandhi in Chinese phonology. Xu (2004, p. 796) similarly speculated the following on the mechanism of Tone 3 sandhi,

“It is unclear whether the largely rising contour in the first L of L L is due to a complete change of the target to [rise] as in R or due to implementation of the same complex target as in isolated L with a time constraint. Further studies are needed to sort this out.”

This paper follows their advice and experiments with a method called Functional Data Analysis (FDA) to see whether further details can be exposed via this advanced statistic method.

### 1.1 Previous findings and implications

It is important to note that the aspect which remains unclear in the majority of acoustic research on Tone 3 sandhi, and which also

<sup>1</sup> Tone 1 (high-level, H), Tone 2 (rising, R), Tone 3 (low, L) and Tone 4 (falling, F) are the four main tone categories in Mandarin. There is also a neutral tone, Tone 5, which is more frequently used in mainland China than in Taiwan.

<sup>2</sup> For example, *yi*1 is realised as T2 when it is followed by a T4 classifier (e.g. *yi*2 *jian*4 ‘一件’, means ‘one article’) and is realised elsewhere as T4 when it is followed by T1, T2, or T3 (e.g. *yi*4 *zhang*1 ‘一張, one piece of’; *yi*4 *zhi*2 ‘一直, all along’; *yi*4 *zhong*3 ‘一種, a kind of’). *yi* sandhi is sometimes mentioned together with *qi*(七)、*ba*(八)、*bu*(不), all of them go through similar variations.

influences psycholinguistic research, is whether (1) speakers replace the first low tone with a rising one by applying a phonological rule prior to articulation (i.e. a categorical change) or (2) no rule is specifically defined prior to articulation and a seemingly ‘rising’ pattern in the first syllable followed by a low tone is attributed to an online phonetic process (i.e. a gradient phonetic implementation). The gradient phonetic differences of sandhi T3 to that of underlying T2 could be due to speakers’ inability to produce two low tones in quick succession, or the existence of an articulatory encoding stage to adjust the online articulatory operations prior to actual articulation (Chen and Chen, in press).

Acoustic data supporting either the categorical or gradient view of Tone 3 sandhi is generally acquired by comparing ‘underlying T2’ (e.g. T2 in a **T2T3** context) and ‘sandhi T3’ (e.g. **T3T3**→**T2T3**) in terms of several measurements as follows:  $F_0$  values (maximum, minimum, mean, etc.), duration, and  $F_0$  slope (for which the definition varies between studies). In general, if no statistically significant difference in the above measurements is found between underlying T2 and sandhi T3, then Tone 3 sandhi is suggested to be a ‘categorical change’. On the other hand, if there is a statistically significant difference in either measurement, then Tone 3 sandhi is considered ‘gradient’. If the results are of discernible differences, but these differences are not large enough to result in a statistical difference between the two rising  $F_0$  contours, Tone 3 sandhi process is sometimes suggested to be in the process of ‘merging’ from underlying T3 to T2, i.e. the differences between underlying T2 and sandhi T3 are not fully neutralized phonetically (Peng, 2000).

## 1.2 Purpose of this study

The purpose of this study is to investigate T3 sandhi phenomenon in the time dimension. Like in traditional studies, we will compare  $F_0$  contours extracted from ‘underlying T2’ (T2T3 dyads) and ‘sandhi T3’ (T3T3 dyads). Contrary to traditional studies, we will not rely on a few comprehensive measurements taken on  $F_0$  contours such as duration, slope, etc. Rather,  $F_0$  contours *as they are* will be input to statistical tools that perform a scan of the time axis and produce detailed results along the time dimension. This powerful extension of classic statistical analysis has two fundamental advantages. First, all the information carried by

the shape of  $F_0$  contours is preserved, since no selection of ‘special points’, such as peaks, is required. Second, results provide a detailed picture of the time dimension, rather than one comprehensive answer, e.g. in terms of overall significant difference. This approach is of particular interest for the investigation of T3 sandhi because, based on past studies, we expect that differences between tone dyad T2T3 and T3T3, if any at all, may be quantitatively small and localised in one or few specific phases of the whole articulatory gesture. The analysis will be carried out by applying Functional Data Analysis (FDA, Ramsay and Silverman, 2005; Gubian, 2013). In particular, a functional t-test will be employed, which extends the rationale of the well-known t-test to the time dimension in a principled way (Cheng et al., 2010).

To dig even deeper into the dynamics of the underlying articulatory gestures, our investigation will be carried out not only on  $F_0$  contours, but also on their respective velocity contours, i.e. their first derivative with respect to time. These curves show the slope of their respective  $F_0$  contours in time, i.e. take positive/negative values whenever the  $F_0$  contour is rising/falling. This representation offers further insight in the underlying tonal targets, since the intention of the speaker to reach a certain target might not be evident by looking at the  $F_0$  value at a certain time, but rather can only be revealed by looking at the rate with which this value is changing (Gauthier et al., 2007).

## 2 Methodology

### 2.1 Dataset

The data used is a subset from a systematically designed and collected corpus from the first author’s previous research on bi-tonal variation and time pressure. Six male Taiwan Mandarin speakers recorded disyllabic /ma+/ma/ sequences with a total of 16 (4x4) tonal combinations embedded in two carrier sentences. These two carrier sentences differ in the tone preceding the target sequence /ma+/ma/, being H or L. The tone following the target sequence is always H. Various conditions were imposed to elicit different degrees of tonal reduction, including positions in a carrier sentence (initial, mediate and final), repetition time (1<sup>st</sup>, 2<sup>nd</sup>, and 3<sup>rd</sup>) and speech rate (slow, normal and fast). Three reduction types were labelled according to the integrity of the intervocalic /m/,

*non-contracted* for a clear presence of an intervocalic nasal, *contracted* for a clear absence of nasal murmur and *semi-contracted* for intermediate cases. More details concerning the corpus can be found in Cheng et al. (2010).

In this study, we selected *non-contracted* T2T3 and T3T3 from a (T1#)/ma+/ma/(#T1) context, where # delimits the target sequence and its surrounding tones. We chose non-contracted realizations in order to avoid the complex issues involved in tonal contractions, which are not relevant here, and we selected only one tonal context for consistency. In total 119 T2T3 dyads and 106 T3T3 dyads were selected.

## 2.2 F<sub>0</sub> extraction

Extraction of F<sub>0</sub> contours was first carried out with the vocal cycle marking of the Praat program (Boersma and Weenink, 2013) and then with manual repair of octave jumps and other distinct irregularities using a Praat script (Xu, 2013). For each target curve, 40 measurement points were generated, 20 equidistant points per syllable. This time sampling scheme implicitly produced an alignment of all contours on their syllabic boundary, because the 21st F<sub>0</sub> sample marks the beginning of the second syllable in all contours. This convenient fact will be used in the following steps. F<sub>0</sub> values are converted to semitones and the average of the 40 samples subtracted from all contours, which helps reduce variability owing to speaker identity.

## 2.3 From F<sub>0</sub> samples to functions of time

Functional Data Analysis (FDA) refers to a set of tools that extend ordinary multivariate statistics to the domain of functions. Any FDA session requires that the input curves are represented as functions, in our case functions of time  $f(t)$  (hence the term “functional” in FDA). Standard smooth interpolation techniques were customarily applied at this stage. In particular, we applied B-splines-based smoothing with roughness penalty (Ramsay and Silverman, 2005). B-splines are a family of polynomial functions that allow one to smoothly interpolate arbitrary contours defined by a finite number of samples (de Boor, 2001). Roughness penalty refers to a regularization procedure where fitting error, i.e. the difference between the exact sample values and the corresponding one read on the interpolating function  $f(t)$ , is traded for smoothness of  $f(t)$ . The latter is enforced in order to reduce unnecessary and rapidly varying shape

detail, which in the case of F<sub>0</sub> contours is likely to be due to the limited precision of the F<sub>0</sub> tracker and to irrelevant microprosodic effects. An example of smoothing is shown in Figure 1.

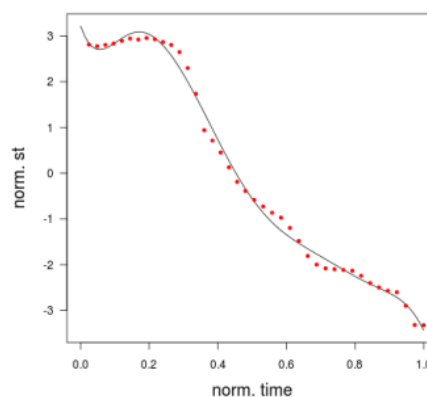


Figure 1: An example of smoothing applied to one of the F<sub>0</sub> contours in our data set. Red dots indicate original F<sub>0</sub> samples, the solid curve is the interpolating function  $f(t)$ .

A crucial aspect of the functional representation of data lies in the fact that the same time axis must be used in all curves. This means that all curves are normalised to have the same duration. Since the original tone dyads do not have the same duration, a proper normalization has to be carried out. In our case, we have divided the common time axis in two equal halves, the first half spans the first syllable, the second half the second syllable. This normalization was based on the sampling scheme described above. The total duration is conventionally indicated as 1 (i.e. *not* the average duration in seconds).

Velocity curves  $v(t)$ , i.e. first derivatives of the F<sub>0</sub> curves, were computed directly from the analytical form of  $f(t)$  before the application of time normalization. That is, the values of  $v(t)$  curves are from the original un-normalised F<sub>0</sub> contours (and are thus in units of semitones per second) before themselves being normalised in time.

## 2.4 Functional t-test

Functional t-test is a technique that extends the concept of t-test to continuous functions. Ordinary t-test compares the mean of two sets of numbers and determines whether those means are different, at a certain confidence level. Functional t-test compares the mean functions of two groups of functions sharing the same time axis and determines *where* those mean curves differ, at a certain confidence level. The mean

curve is the curve obtained by averaging the values of all curves at each corresponding point in time. Operationally, a battery of ordinary t-tests is carried out in the time dimension. The confidence value for the test statistics are obtained by resampling techniques (see Ramsay and Silverman, 2005 for further details).

In this work, two separate functional t-tests have been conducted, one comparing the 119 T2T3  $F_0$  contours against the 106 T3T3  $F_0$  contours, and another comparing their respective velocity curves.

### 3 Results

The results of the two functional t-tests described above are shown in Figures 2 and 3 for  $F_0$  contours and their velocity curves respectively. Figure 2a shows the mean  $F_0$  contours for the T2T3 and T3T3 dyads. As mentioned earlier, the time axis is normalized to a common unitary duration, and the middle point  $t = 0.5$  marks the boundary between the first and the second syllable for all contours, i.e. the onset of the second /m/ sound in all /mama/ sequence. Also the vertical axis is normalized, since the mean semitone value in time has been subtracted from each curve. Both curves appear very similar, the general trend is an initial fall, since the preceding tone is a high tone in all cases, followed by a rise, which realises the first target, and ending with a sharp fall, which realises the second target. The outcome of the functional t-test is shown in Figure 2b. The solid red line indicates the value of the test statistic along the time axis, which is the same axis as in Figure 2a. Blue curves in Figure 2b represent time dependent significant thresholds, the higher flat line representing a more conservative threshold. The test statistic locally reaches significance at 5% confidence whenever the red curve reaches above one of the blue ones. Figure 2b shows that there is no point in time where the small differences between the average T2T3 and T3T3 curves are significant, since the red curve always remains below both thresholds.

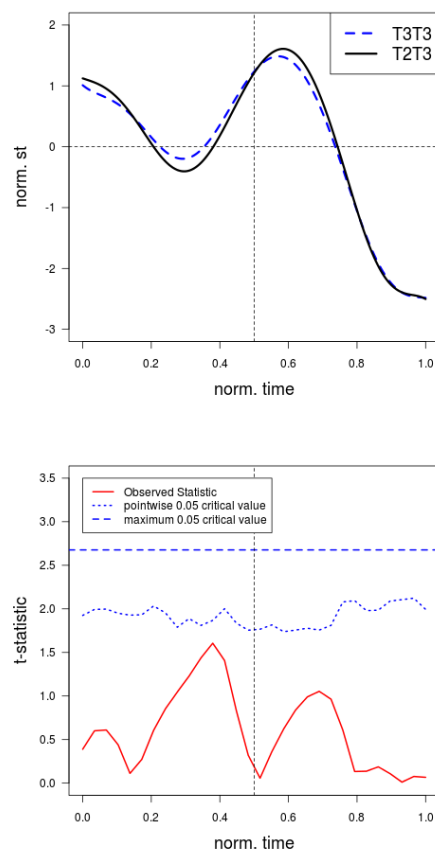


Figure 2: top (a) Normalised average  $F_0$  contours of T3T3 (blue) and T2T3 (black); bottom (b) Functional t-test on the normalised averaged  $F_0$  contours of T3T3 and T2T3.

Figures 3a and 3b show results for the velocity curves. Figure 3a shows that the velocity curves are qualitatively very similar on average. Crucially, their signs are almost always the same, meaning that at every instant in time, on average the  $F_0$  contours are either both rising (positive velocity) or both falling (negative velocity). However, small quantitative differences are visible, especially around 0.5 normalised time units. Figure 3b shows that the mean velocity curves of Figure 3a differ significantly in an interval localized around 0.5 normalised time units, since the red curve surpasses one of the empirical confidence levels (i.e. pointwise 0.05 critical value). Returning to Figure 3a, this means that the difference in slope between T2T3 and T3T3 localized around the syllable boundary is systematic. T2T3  $F_0$  curves tend to increase more rapidly than sandhi T3T3 curves (i.e. T2T3 velocity curves reach a higher value) in the process of completing the realization of the first target and preparing for the second one.

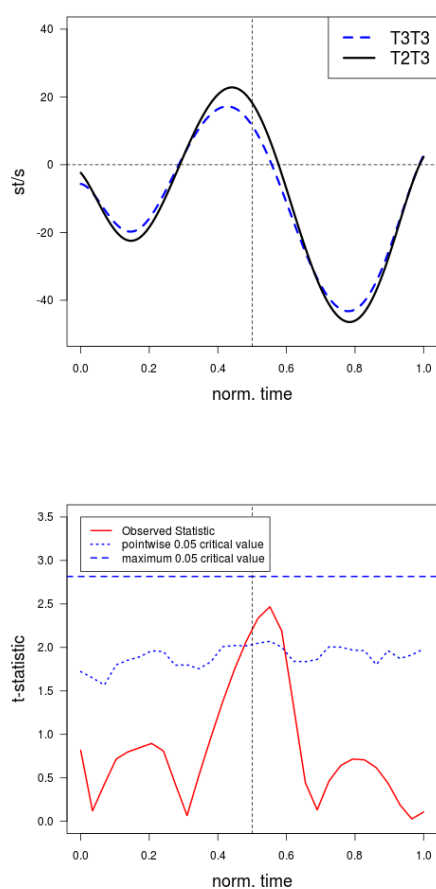


Figure 3: top (a) Normalised average  $F_0$  velocity contours of T3T3 (blue) and T2T3 (black); bottom: (b) Functional t-test on the normalised averaged  $F_0$  velocity contours of T3T3 and T2T3.

Finally, we check whether this result obtained using velocity contours could be the consequence of an overall difference in duration between the T2T3 and T3T3 syllable groups. The mean duration of the first syllable is 0.194 second for T2T3 and 0.181 second for T3T3, while the duration of the second syllable is almost identical for both dyads (0.175 second). The small duration difference in the first syllable is not significant ( $p$ -value  $> 0.05$ ). That is, the higher velocity of the  $F_0$  contour towards the end of the first syllable in the T2T3 group is unlikely to be attributed to a longer duration. From this, we conclude that the localised effect seen on velocity is real.

#### 4 Discussion and Conclusion

Looking only at the statistical results obtained from performing a functional t-test on the  $F_0$

contours, shown in Figure 2b, no significant difference is seen between underlying T2 and sandhi T3. However, applying a functional t-test to the first derivative of the  $F_0$  contours (velocity) reveals interesting information in the time dimension—specifically, a significant difference is seen between tone dyads T3T3 and T2T3 velocity curves towards the end of the first syllable (Figure 3b). A significant difference in this region may suggest speakers' late modification of articulatory implementation. The increased difference in velocity also indicates a steeper rising-falling movement in T2T3 than that of T3T3. This may be an indication that the articulatory target for sandhi T3 is not as clearly defined as that of an underlying T2 followed by a low tone, thus resulting in a less 'prepared' rising-falling trajectory in the T3T3 context.

Responding to the question raised in the title of this paper, 'are Mandarin sandhi T3 and underlying T2 the same or different?', through analysing the velocity profiles along the time dimension, the current results demonstrate 'a difference' at a later part of the implementation of sandhi T3 to that of an underlying T2.

It is interesting to note that if the FDA comparison was made using only  $F_0$  contours, sandhi T3 and underlying T2 could have been interpreted as 'the same'. It is worth mentioning that  $F_0$  contours contain surface variations, for which articulatory movements are less exposed (compared to  $F_0$  velocity) and should be taken with caution in pitch related research.

To conclude, in this paper, techniques from Functional Data Analysis have been applied to look at Tone 3 sandhi in Taiwan Mandarin. Applying a functional t-test to  $F_0$  velocity contours demonstrated a difference between 'underlying T2' (T2T3) dyads from 'sandhi T3' (T3T3) dyads; the former  $F_0$  contours exhibit a steeper rise than the latter near the end of the first syllable. This small yet systematic difference has not been previously reported in the literature. It is hoped that such quantitative findings on Mandarin Tone 3 sandhi can shed further light on the mechanism of tone sandhi in general.

#### Acknowledgements

Writing of this paper was supported by "Aim for the Top University Plan" of the National Taiwan Normal University and the Ministry of Education, Taiwan, R. O. C. The authors would like to thank Dr. Yi Xu and Dr. Rhodri Nelson for their

valuable comments and suggestions to improve the quality of this paper.

## References

- Boersma, Paul and Weenink, David. 2013. Praat: doing phonetics by computer [Computer program]. Version 5.3.52, retrieved 12 June 2013 from <http://www.praat.org/>
- Chen, Jenn-Yeu and Chen, Train-Min. In press. Mechanism and Locus of Tone 3 Sandhi during Mandarin Chinese Spoken Word Production. *Journal of Chinese Linguistics*.
- Cheng, Chierh, Xu, Yi, and Gubian, Michele. 2010. Exploring the mechanism of tonal contraction in Taiwan Mandarin. *Proc. Interspeech, 2010: 2010-2013*.
- de Boor, Carl. 2001. *A Practical Guide to Splines*, Revised Edition. Springer, New York.
- Gauthier, Bruno, Shi, Rushen, and Xu, Yi. 2007. Learning phonetic categories by tracking movements, *Cognition*, 103: 80-106.
- Gubian, Michele. 2013. Functional data analysis for speech research. [Online]. Available: <http://lands.let.ru.nl/FDA>
- Myers, James and Tsay, Jane. 2003. Investigating the phonetics of Mandarin tone sandhi. *Taiwan Journal of Linguistics*, 1(1): 29-68.
- Peng, Shu-Hui. 2000. Lexical versus 'phonological' representations of Mandarin Sandhi tones. *Papers in laboratory phonology V: Acquisition and the lexicon*: 152-167.
- Ramsay, Jim and Silverman, Bernard. 2005. *Functional Data Analysis—2nd Ed.* Springer.
- Shih, Chilin. 1997. Mandarin third tone sandhi and prosodic structure. *Linguistic Models*, 20: 81-124.
- Xu, Yi. 2004. Understanding tone from the perspective of production and perception. *Language and Linguistics*, 5.4: 757-797.
- Xu, Yi. 2013. ProsodyPro—A Tool for Large-scale Systematic Prosody Analysis. *Proc. Tools and Resources for the Analysis of Speech Prosody (TRASP 2013)*, Aix-en-Provence, France.