Dialect-geographical Acoustic-Tonetics: five disyllabic tone sandhi patterns in
cognate words from the Wu dialects of Zhèjiāng province

Phil Rose
Emeritus Faculty, Australian National University, Australia
philjohn.rose@gmail.com

Abstract

The typology of tone sandhi patterns on disyllabic tonally
cognate words in selected sub-groups of the Wu dialects of the
S. E. Chinese province of Zhejiang is investigated using data
from 48 sites collected over 45 years. A new method of
extrinsic z-score normalisation is demonstrated which permits
comparison of tones across dialects with different pitch
ranges. Five different but typical right-dominant word-tone
patterns are identified, acoustically quantified, and their
dialectological distribution specified. It is hypothesized that
changes in isolation tones and different types of dissimilation
of the first tone from the word-final tone are a possible origin
of the observed variation.

Index Terms: Wu dialects, dialect geography, lexical tone
sandhi, extrinsic F0 normalisation. tonal cognates

1. Introduction

The Chomskyan revolution, focusing on Language as
something in speakers’ brains, has tended to divert attention
from the more traditional, speaker-external view of Language
as a shared property of human societies, with an existence, and
variation, in space and time [1, pp. 9-12]. The first study of
how words vary as a function of geographical location –
Yang Xiong’s Fangyan – appeared early, in the Han dynasty
China of 53 BC [2, p.241]. The systematic discipline of
Dialect Geography arose much later, in late 19th century
Europe [3, pp.15–17], ultimately giving rise in the first part of
the 20th century to major surveys which plot the distribution of
dialectal features in the USA, Europe and the UK [3, pp.18–
23]. Chinese dialect surveys of this era usually go
mentioned, but much work has been published, with
accompanying maps. These include both large-scale surveys
of whole provinces e.g. [4,5,6,7,8] as well as smaller-scale
surveys of parts of provinces, e.g. [9,10,11,12]. This paper
returns to China, to investigate the geographical distribution of
a phonological feature: tone-sandhi types in the Wu dialects of
its east-central Zhejiang province.

Wu dialects show a great variety of citation tone shapes, the
geographical distribution of which has been described for
parts of the Wu area [13,14]. The Wu dialects are probably
best known, however, for their complex lexical tone sandhi.
This often results in tonal shapes for polysyllabic words which
can bear a complex, phonetically opaque relationship to the
isolation tones of their constituent morphemes [15, pp.81–84].
The complexity of the tone sandhi systems of individual sites
is of considerable tonological interest. Tone sandhi data from
Zhenhai, for example, has been adduced to demonstrate the
theoretically important point within autosegmental phonology
that tone contour can be copied [16], and Shanghai is often
touted as a typical Chinese example of left-to-right tone

Figure 1: Map of Zhejiang province showing Wu dialect
sub-groups, administrative districts, and location of sites
sampled. (Sub-)sub-groups and sites are colour-coded
thus: Taihu-Linshao 太湖-临绍, Taihu-Yongjiang
taihu 台州; Wazhou 嵊州; Oujiang
江; Chuzhou 池州, Chuqu-Chazhou
州, Symbols indicate sites with different word-
tone types discussed below.
1.1. Wu Dialects: Location, sites, subgroups

The Wu dialects of southern Zhejiang province, in East Central China, are conventionally divided into five main subgroups, although published sources [10,12,19] disagree somewhat in their details and names. Figure 1, based on the Language Atlas of China [19], shows a colour-coded map with the subgroups’ names and location, as well as indicating the sites sampled, and smaller administrative districts (xiàn 县). To save space, I have also included in this map some of the results, with symbols indicating the distribution of different word-tone shapes.

1.2. Wu lexical tone sandhi examples

The most common word-formation process in Chinese is compounding [20, p.73ff.], and part of this compounding process in Wu dialects involves changes in the tones of the word’s constituent morphemes (so-called lexical tone sandhi) [21, pp.41,42]. One result of this (perhaps its actual function) is that words are marked tonologically. Two examples from the data in this paper will help to clarify these morphotemic relationships.

In the dialect of Lishui (Chuzhou site purple 4 in figure 1), the word for melon 西瓜 is pronounced [s] kuə 44.323, with a high level pitch on the first syllable and a low-dipping pitch on the second. Panel A of figure 2 shows the tonal acoustics for this word, as well as two other words with the same pitch, in a young male Lishui speaker. It can be seen that the mean F0 trajectory, in red, corresponds well to the high-level + low-dipping pitch. The word melon is composed of two morphemes. The first morpheme is {east} /s/ and the second is {gourd} /kuə/. In isolation, both morphemes have the same, low-dipping tone, the mean acoustics of which can be seen in panel C of figure 2. It is marked la and plotted with a thick red line. The mean F0 of four of his other tones is also shown to indicate the location of the low-dipping tonal F0 in his F0 range. In the formation of this Lishui word melon, then, it appears that the tone on the word-final syllable has not changed, but the tone on the first has changed to high level. This kind of sandhi, where the tone on a word-final syllable remains the same as the tone on a single syllable word/morpheme, but the tone changes on syllables toward the beginning of the word, is called right dominant. It is supposed to be common in the Southern Zhejiang dialects under examination in this paper [18].

Genglou (Taizhou site blue 2 in figure 1) lies, as the crow flies, some 125 kilometers east-north-east of Lishui. (It is considerably further as the crow hops, as the two locations are separated by mountains.) In Genglou dialect the word for melon 西瓜 is pronounced [ɕi kua 35.731] with a first-syllable pitch normally rising above the speaker’s modal pitch range to become falsetto, and a high-falling pitch on the second. Panel B of figure 2 shows the tonal acoustics for this word, as well as two others with the same pitch, in a young male speaker. Comparison with his citation tonal F0 in panel D shows that, again, it is the word-final tone that is preserved and the tone on the first syllable changed.

It can be seen that the same, i.e. cognate, words are used in figure 2 for both dialects. What links the Lishui and the Genglou data and makes them comparable is that, despite the considerable variation, all the relevant tones in the examples from these two sites derive historically from the same single Middle Chinese tone *Ia. In this paper, only words composed of reflexes of this tone (“Ia+Ia words”) are used. This tonal combination was chosen because previous research [e.g.18,22,23,24,25] has shown that (along with la+1b, lb+la, lpb+l) it tends to preserve its different sandhi shape and is less subject to mergers with other tonal combinations.

2. Procedure

2.1. Corpora

Two main sources of data were used: tape-recordings made by Prof. Zhu Xiaonong and his assistant in the mid to late 90’s as part his post-doctoral survey of Wu tones and tone sandhi funded by a large Fellowship Grant from the Australian Research Council; and tape-recordings made about a decade earlier by Prof. William Ballard in early 1988 as part of his survey of tones and tone sandhi in southern Wu varieties, funded by the US National Endowment for Humanities [26]. I record my gratitude to Profs. Ballard and Zhu for making their valuable recordings available for digitization and analysis. This material was supplemented by recordings made by my students and myself over about the last 45 years, but consisting primarily of tape-recordings I made in the mid-70’s, of older speakers from the N.E. of the Zhejiang area around Ningbo. All three sets of recordings were primarily intended to elicit data for analyzing disyllabic Wu tone sandhi, as well as, of course, tones on monosyllables. Because of this, the same words are often found in all three sets. All recordings used in this paper may be listened to, and the acoustic results viewed, on my web-site: https://philjohnrose.net. The data are also available upon request to anyone interested in further acoustic-phonetic research of the speech of this area.
2.2. Extraction and measurement

Recordings were digitised at 22.05k and examples of Ia+Ia words identified and edited out for further processing. The number of Ia+Ia words available for each speaker ranged from 3 to 20+. Each speaker’s Ia+Ia words were then transcribed with respect to segments and pitch (perceived pitch, not F0 c.f.[27]), and any non-modal phonation type. Transcription is an indispensable part of the procedure. Even though one is often dealing with complete cognates, they can still differ considerably in segmental quality across the area. For example, the same proto-Wu segment *t, has developed variously into [n], [t], [d], [d] and [f], so the Ia+Ia word east wind 东风, can be heard as [tɔŋ fʊŋ] in the north-east of the area, [tʊ̃əm fʊm] in the centre, and [dɪn ɨŋ] in the south-west. Such differences in sonorance, voicing and initiation can be expected to exert considerable intrinsic effect on F0; there are in addition differences in vowel quality and phonation type which will also have intrinsic effects. Because of these intrinsic effects it is impossible to interpret variation in F0 shapes without knowledge of the accompanying segmentals. Note, too, there is no indication from the F0 alone that the high rising F0 on the first syllable of the Genglou pattern in panel B of figure 2 represents an excursion above the speakers’ modal F0 range: only an auditory transcription will supply that information. More importantly still, not all Ia+Ia words for a given site were always said with the same tonal pitch: consistent with gradual diffusion through the lexicon, some were obviously changing to another sandhi shape. A prior pitch transcription is therefore necessary to avoid mechanically extracting mean F0 values that correlate different tonal targets.

The tonal acoustics of each disyllabic word (F0 and duration) were measured by first generating its wideband spectrogram in Praat, together with its aligned wave-form and superimposed F0 contour. The good time-domain resolution of the spectrogram was then used to hand-segment the disyllabic words into first syllable Rhyme, intervocalic consonant, and second syllable Rhyme. Criteria for segmentation are described in [18]. The F0 over each Rhyme was then extracted with Praat, and modeled in R with an eighth-order polynomial, from which F0 was then sampled with a sufficiently high frequency to capture the details of its time-course: at 10% points of the Rhyme as well as 5% and 95%. F0 was also measured in mid-duration of any voiced intervocalic consonant. The mean F0 trajectory was then calculated over all tokens. Figure 2 contained typical results of this quantification. Conventional dialect-geographical procedure was followed in using established descriptive parameters – here, the transcribed pitch contour – to assign a speaker’s tokens to a particular type (e.g. “mid-level + high-falling”). The geographical distribution of these types was then plotted in a so-called display map [3, p. 29]. Figure 1 is an example.

2.3. Mean-adjusted Extrinsic Z-score Normalisation

In order to obtain quantified representations of the acoustic properties of the different types of Ia+Ia tone patterns, their mean acoustics were z-score normalised [28]. Although this method of tonal normalisation has been shown to be clearly superior for removing speaker-dependent characteristics [29], the use of intrinsic normalisation parameters (NPs) – i.e. NPs derived from the data to be normalised – will be inappropriate for comparison between the tone sandhi patterns observed. This is because, as has been observed in the Genglou data, the different types of Ia+Ia tone patterns clearly occupy different parts of their speakers’ pitch range, and therefore their NPs of mean and standard deviation F0 will not necessarily accurately reflect the location of the Ia+Ia word-tone in the speaker’s overall F0 range, and thus not permit comparison between the different pattern types. The data were therefore first normalised using extrinsic NPs derived from quasi long-term distributions reflecting the whole of a speaker’s F0 range [30]. These were obtained from all the examples of a speaker’s disyllabic tone sandhi combinations elicited in the session: usually between ca. 1.25 and 1.75 minutes net speech. Normalisation was done in the log10F0 domain to counteract the long-term F0 distributions’ typically positive skew. Speakers’ extrinsic normalised mean values were then adjusted by linear shifting relative to the normalised mean to give a better idea of the clustering. The results of this process are shown for one sandhi type, with high-level + low-dipping pitch, in panels A & B of figure 3. It can be seen that the mean extrinsic normalised value for the sandhi type representing the position in the F0 range (thick red line) remains the same, but the clustering of speakers around the mean is reduced by the linear shifting of the individual speakers’ normalised value. This clustering is comparable to that achieved with an intrinsic normalisation of the same data, shown in panel C of figure 3. It can also be seen that the intrinsic normalisation makes the word-fnal tone appear slightly lower in the speakers’ range than it actually is.

3. Results & Discussion

The first finding – unsurprising, given the extent of the area sampled – is that there are indeed different patterns for word tone: tonal acoustics for five patterns are shown in figure 3. Secondly, these patterns are not randomly distributed, but largely group geographically as shown on the map in figure 1. Finally, as already hinted at in figure 2, some of these patterns appear rather different from each other. The six patterns are described briefly below, naming them either after the main district where they are found, or a main city in the area.

Lishui pattern (fig.1: ⌂) This pattern, found in the central south of the area, has high-level pitch on the first syllable and low-dipping pitch on the word-final syllable, e.g. Lishui [tʊŋ ɨŋ 44.32] east wind. Panel B of figure 3 gives normalised acoustics for 16 speakers. This pattern occurred mostly in varieties in Lishui and Yunhe districts (Chuzhou sub-sub-group), but was also found further south in Taishun (site purple 9) and the Wenzhou sub-sub-group (site cyan 3).

Ningbo pattern (fig.1: ◊) This is found in the N.E. corner of the area and has mid-level pitch on the first syllable and high falling pitch on the word-final syllable, e.g. Zhenhai [tʊŋ ɨŋ 33.51] east wind. Panel D of figure 3 gives normalised acoustics for 21 speakers. Many of the varieties with this word-tone are found in the Yongjiang sub-sub-group (red sites 1-6), but the pattern is not coterminal with it, being also found in the sub-group of varieties of Shaoshing and Cixi (brown sites 1, 2). There is also an occurrence in Huzhou (yellow site 8), just within the Wenzhou sub-group.

Linhai pattern (fig.1: ⬇) The high-rising-falsetto – high-falling word tone exemplified in figure 2 is found in a group of sites in the Taizhou sub-group in the east, e.g. Linhai [tʊŋ 35°41] east wind. Panel E of figure 3 gives normalised acoustics for 5 speakers. Note that the high point of the first-syllable F0, as part of the falsetto realisation, extends over three standard deviations above the mean.
Chuzhou, progressing from the dialects' tonal cognate words in Zhejiang Wu dialects. Because only tonally cognate words were used in this experiment, all the documented variation in word-tone pattern has – to quote the foundational document of historical linguistics [1, p.106,107] – sprung from some common source, namely, the single Middle Chinese *Ia tone. It is of obvious interest to explain how such ex uno pluribus variation arose.

This paper has demonstrated six different word-final tones which divide categorically into two groups. One group appears to embody gradual variation in contour, progressing from high-rise (Qingtian ⚫), to low-dipping (Lishui ⚫), to laryngealised low-dipping (Tonary Ω) to laryngealised low-falling (Tonary Θ). The other group has two high-falling variants which differ mainly by rate of F0 fall: derivative increasing (Ningbo ⚤) and decreasing (Linhai ⚤). If it is the case that all the modern varieties examined have right-dominant tone sandhi, with their word-final tone similar to their isolation tone, a sensible hypothesis is that the synchronic variation observed in word-final tone is referable to prior variation in isolation tone, and the explanandum then will be how the isolation tones diverged. One of the next stages of the investigation is thus to document the dialects’ isolation tones, but currently, given the variety in observed shapes, no proto-form can be sensibly reconstructed.

There are also considerable differences in the words’ first-syllable tone, of which the results show four types: high-level, mid-level, high-rising-to-falsetto, and mid-falling. As one might expect [32] some of these first syllable shapes are clearly the result of dissimilation. The data also suggest, however, that different types of dissimilation can be triggered by different tonal properties of the following syllable. Thus a following high rise appears to have triggered a low fall in the Qingtian ⚫ pattern, and a following high fall appears to have triggered a high rise in Linhai ⚤. Thus further investigation might yield a typology of dissimilation. Certainly, the apparent phonetic motivation of the first-syllable tone changes means reconstruction is possible: a proto-mid-level tone would be the rational choice.

4. Summary

This paper has identified and acoustically quantified five different putatively right-dominant word-tone patterns on disyllabic cognate words in Zhejiang Wu dialect sub-groups. It has shown how they distribute geographically, and speculated on their historical development.

The results represent, of course, only a miniscule portion of the overall picture of the distribution of Wu tone sandhi in Zhejiang. There are additional Ia+Ia types not described in this paper; and, given that many dialects have eight citation tones, Ia+Ia combinations represent theoretically only a small proportion of the possible shapes in a single variety.

Urban dialectology and forensic voice comparison have shown that variation in speech is ubiquitous and multidimensional, showing correlations with, inter alia, age, sex, social status and style as well as location. Space precludes detailed discussion but it is clear that there are some age differences in the corpus, although they appear to correlate with location. For example, there are very little differences between speakers of the Ningbo pattern separated by ca. 80 years real time, whereas in the Oujiang sub-group (not described in this paper) there do appear to be differences in word tone correlating with age [33]. Other correlations remain to be investigated.

A final caveat is that, given some recently documented tonal changes in northern Wu [34], the non-contemporaneity of the material does not guarantee the paper’s findings represent the current state of affairs, which is likely to be less complicated.
5. References