Reverse-engineering Tones and Tone Sandhi in Wenzhou Dialect: Between-speaker Differences and Historical Development*

Phil Rose

Australian National University

Auditory and acoustic descriptions are presented for the tones and tone sandhi of two Wenzhou speakers in disyllabic words with tones from the historical ping + qu categories. Differences between the speakers both in isolation tones and tone sandhi are demonstrated. It is suggested that the opaque morphophonemic relationships between sandhi and isolation tones are better accounted for historically, rather than as synchronic operations on unique underliers. The effects of stress / metrical strength are an important factor.

Keywords: Wu dialects, Wenzhou, tone sandhi, between-speaker differences, reconstruction.

1. Introduction

This paper illustrates some of the fascinating and challenging aspects of disyllabic lexical tone sandhi in the Oūjiāng 蘇江 Wu dialect of Wēnzhōu 溫州, The Wu dialects of China are well known for their complex tone sandhi, and Wenzhou is no exception. One of the many sources of the complexity is the often opaque morphophonemic relationship between a dialect’s isolation tones and the tones when they occur on morphemes within a word. It is a goal to make tonological sense of these relationships: to understand how the tones on a word might arise, or might have arisen, from the combination of tones on its constituent morphemes. This paper illustrates an additional level of complexity in the tone sandhi: between-speaker differences. It documents some differences in tones and tone sandhi behaviour between a male Wenzhou speaker who was 34 years old when recorded in 1988, and a female speaker in her early twenties, recorded in 2006. The corpus and elicitation are described in Rose (2001), and were exactly the same for both speakers.

* Once again thanks to Professor William Ballard for generously making available his valuable field-work recordings. Thanks also to my colleagues Shen Zhongwei and Zhang Cuiling for helping with translation, and to Jiao Lei for suggesting references. The usual disclaimers apply.
This paper examines a very small, but nevertheless interesting, portion of Wenzhou disyllabic lexical tone sandhi: in words with the historical ping tonal category on the first syllable and the historical tonal category of qu on the second. For the bigger picture, at least of the male speaker’s tone sandhi patterns, an account of the morphotonemics in about a third of his disyllabic lexical tone sandhi can be found in Rose (2001). A further 16 combinations were analysed in Rose (2000a), another 12 in Rose (2002), and the remaining eight combinations in Rose (2004). It is four of these combinations—involving ping plus qu tones—that are the focus of this paper. It should be noted that differences between the two speakers are not confined to these combinations, but pervade all their tone sandhi behaviour. The two speakers’ isolation tones are described first, then their sandhi.

2. Isolation tones

An isolation tone is the tonal realisation of a morpheme when the latter occurs on its own. This may be when the morpheme is free and occurs as a monosyllabic word in normal speech, or when its Chinese character is read out. The term tone is used here in a sense analogous to phone, that is, as a constellation of audibly different properties, with pitch predominating but also including length and phonation type, that constitutes observation data for tonological analysis.

An auditory-phonetic description of the two speakers’ eight isolation tones is given in Table 1. The first three rows list the conventional names of the tones’ traditional categories, and underneath are descriptors of tonal pitch and length. Depression refers to the lowering of pitch at the onset of a tone. Thus one possible representation of a tone with a convex [341] pitch is as a depressed high fall [51]. Depression is a well-known phenomenon in the tonology of African tone languages, e.g. Zulu, but has recently also been shown to play an important part in Wenzhou tone sandhi (Rose 2001, 2002).
Reverse-engineering Tones and Tone Sandhi in Wenzhou Dialect

Table 1: Wenzhou isolation tones. Italics indicate tones involved in this paper.

<table>
<thead>
<tr>
<th></th>
<th>Ia</th>
<th>Ib</th>
<th>IIa</th>
<th>IIb</th>
<th>IIIa</th>
<th>IIIb</th>
<th>IVa</th>
<th>IVb</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yinping</td>
<td>Yangping</td>
<td>Yinshang</td>
<td>Yangshang</td>
<td>Yinqu</td>
<td>Yangqu</td>
<td>Yinru</td>
<td>Yangru</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>male speaker</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>mid level</td>
<td>depressed</td>
<td>short high</td>
<td>low rise</td>
<td>high</td>
<td>long</td>
<td>long</td>
<td>long</td>
<td>long</td>
</tr>
<tr>
<td>mid fall</td>
<td>rise</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>female speaker</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>high level</td>
<td>depressed</td>
<td>low level-rise</td>
<td>low</td>
<td>high</td>
<td>lower-mid</td>
<td>level</td>
<td>mid fall</td>
<td>depressed</td>
</tr>
<tr>
<td>high fall</td>
<td>rise</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

In Table 1, auditory descriptions are given separately for the two speakers, as there are some rather big differences between them in pitch height, contour, and merger:

- The most obvious difference between the two speakers is a structural one. Unlike most conservative Oujiang varieties, the female speaker does not distinguish between reflexes of IIA and IIb tones. Both her IIA and IIb reflexes are realised with the same [low level-rising] pitch. Compare this with the male’s separate reflexes of IIA and IIb, which present a typical Oujiang configuration of short high-rise vs. lower-rising pitched tones.

- The male’s Ia tone is located in the middle of his pitch range, while the female’s is higher.

- Their [depressed-falling] Ib tones are also located correspondingly higher.

- Whereas the male speaker has typical long complex falling-rising realizations for his IVa and IVb tones, the falling part being the more salient, the female lacks the rising component for both these tones and their onset is located higher in her pitch range. They consequently have higher and lower mid falling pitch. An interesting result of this is that the female’s IVb [depressed mid-falling] tone sounds very similar to the male’s Ib reflex.

- The female speaker’s reflex for IIIb is not as long as the male’s, for whom length appears in general to be an important auditory tonal dimension. (In addition to his long IIIb, his falling rising IVa/b tones sound overlong, and his high rising IIA tone is notable for its shortness.)

It can be seen from the above that the speakers share only two tones: the [high falling] reflex of IIIa, and the [depressed mid falling] reflexes of Ib/IVb. One could
probably distinguish therefore 14 different tonal pitch shapes from the two speakers' isolation tones.

This paper examines what happens to the falling and level pitched tones Ia, Ib, IIIa and IIIb, when they occur in disyllabic words. These tones are shown bold in Table 1. One of the important things the paper will show is that, although in Wenzhou these four tones group phonetically according to their pitch shape into two classes of level (Ia, IIIb) and falling (IIIa, Ib) pitched tones, the disyllabic lexical tone sandhi points to natural classes that require exactly the opposite pairing, namely mid/high-level and depressed mid/high-falling tones Ia and Ib; and high-falling and depressed lower-mid level tones IIIa and IIIb (Ballard 1980:86). This pairing, based obviously on the historical tonal ping (I) and qu (III) categories, of course wreaks immediate havoc with any analysis that tries to make sense of the sandhi in terms of rules operating on shared underlying tone features, since there is nothing that the two members of the natural classes share.

2.1 Morphophonemics

![Figure 1: Wave-form, wide-band spectrogram and superimposed F0 of two tokens of 白 white spoken by female speaker, showing free variation in initial consonant.](image-url)
The upper ("a"/yin) and lower ("b"/yang) Wenzhou isolation tones are distributed in typical Wu fashion with respect to several segmental and suprasegmental features of the syllable, in particular the manner of articulation of syllable-initial obstruents. Like most other Wu dialects, Wenzhou has three morphophonemically separate sets of syllable-initial stops (voiceless aspirated; voiceless unaspirated; and voiced), and two sets of syllable-initial fricatives (voiceless and voiced). The first two sets of stops occur on syllables with the upper ("a") tones and are realised by voiceless aspirated and voiceless unaspirated allophones.

The third set of stops, which co-occurs with the lower ("b") tones, has different realisation depending on position in the word, and also on the speaker. Word-internally both speakers have normal pulmonic voiced realisations. Word-initially, the male realises his third series of stops predominantly by voiceless, coincident VOT articulations, with a small amount of free variation with modal, VOT lead tokens. The female realises her third series of stops either pulmonically with coincident or lead VOT, or weak-implosively, with lead VOT. Realisations with VOT lead seem to be slightly more common. To illustrate this, Figure 1 shows the aligned wave-form, spectrogram and F0 of the female’s first and third repeats of 白 white which has [depressed mid falling] tone IVb and an initial bilabial stop. The first repeat has a clear coincident VOT [p]; the third shows the typical VOT profile of an implosive [6], with increasing peak-to-peak amplitude. Note also the typical depression of the F0 over about the first 10 centiseconds of the Rhyme in the first repeat: this example shows clearly that depression is not associated with initial consonant voicing, since there is none.

A similar distribution applies mutatis mutandis to the two sets of fricatives: voiceless in upper, "a" tones; voiced word-internally, but voiceless in free variation with voiced word-initially in lower, "b" tones. Voiced or partially voiced fricatives are more common in the female. In this paper, the morphophonemically voiced series are transcribed with voiced symbols (so "b" for example stands for [b] intervocalically, and [p] in free variation with [b] or [6] word-initially).
2.2 Isolation tone acoustics

Figure 2: Mean F0 shapes of male (left) and female speakers’ isolation tones. x axis = duration (csec.), y axis = F0 (Hz). Thicker lines denote F0 of tones discussed in this paper. D = depressed.

Figure 2 shows the F0 values of the two speakers’ isolation tones plotted as a function of absolute duration. Each curve is a mean of ca. 5 tokens. To obtain these curves, tonal fundamental frequency was sampled from the Rhyme at a sufficiently high rate to derive a detailed picture of its time course. The F0 shapes of the tones discussed in this paper are plotted with thicker lines. The F0 shapes are largely as expected from the pitch descriptions, although the pitch of the female’s Ia tone sounds higher than implied by its F0 position, and the pitch of her Ib tone sounds to fall from lower than its F0 value implies. Note the relative F0 height difference between the two speakers’ Ia reflexes, and also between their two Ib reflexes. The similar F0 shapes for the female’s Ila and IIb reflexes are clear, with her Ila reflex onsetting slightly higher, and offsetting slightly lower, than the IIb reflex. The difference in the F0 of her two reflexes is far too small to be tonal, and is presumably conditioned by the difference in the consonant types associated with “a” and “b” tones, together with an inertial effect related to the rate of fall in the Ila which means that the Ila F0 does not rise quite as high as in the IIb reflex.

One conspicuous feature of the male’s F0 configuration, which also agrees well with the auditory impression, is the large between-tone differences in duration. For example his high rising tone Ila has less than half the duration of the falling-rising tones IVa and IVb. Duration differences are not so salient for the female: her IIIb reflex, for example is not overlong, as in the male.
Figure 3: Normalised F0 and duration of male and female speakers’ isolation tones. 

x axis = normalised duration (%), y axis = normalised F0 (%).

Figure 3 shows the speakers’ normalised tonal acoustics, in order to give a clearer indication of the way their F0 data relate. Because the speakers only appear to share two obvious tone shapes—the [high falling] IIIa tone and the [depressed mid falling] Ib/IVb reflexes—a simple percent-of-range normalisation was preferred for F0 to the normally superior z-score method (Rose 1987). The upper limit of the F0 range was defined as the peak value of the speakers’ [high falling] IIIa tone; the lower limit as the offset value in their [depressed mid/high falling tone] Ib (see Figure 2). F0 values were then normalised by subtracting the range minimum from the F0 value,
dividing by the range, and converting to percent. Duration, too, must be normalised, but not by equalization: it is important when comparing tonal acoustics not to equalize duration, as this can distort the relationship between F0 contours (Rose 1993). The reference duration for normalisation was taken as the mean duration of the speakers’ [high falling] and [mid/high level] tones IIIa and Ib. Tonal duration was then normalised as the percent ratio of a tone’s raw duration to the reference duration (Rose 2000b). For example, the male speaker’s mean IIIa and Ia tone duration was 27.1 csec., so the normalised duration for his [depressed low fall-rise] IVb tone, which was 47.7 csec. long, was ((47.7 / 27.1) * 100 =) 176%.

Figure 3 shows that the two speakers’ III tones (top left panel) have very similar normalised values, except for the greater duration in the male’s IIIb. Also note the similarity in the normalised F0 and duration of the male speaker’s Ib and the female’s IVb reflexes: a nice example of the “same tone, different category” phenomenon (Steed & Rose 2009) — although this time within a dialect. The remaining tones differ between the two speakers. In particular one can clearly see the relatively higher location of the female speakers’ I tones (top right panel). The higher location and truncated nature of the female’s IV tones (bottom right panel) are also clear, as is the female’s collapsing of IIa and IIb reflexes into a tone with a low level-rising pitch (bottom left panel). Note that this merged tone differs from the male’s low rising contour for IIb in its prolonged initial component.

3. Sandhi

The sandhi combinations dealt with in this paper are shown in Table 2. It can be seen that they consist of the two I tones on the first syllable, followed by the two III tones. It will be shown that both the I tones and the III tones constitute separate natural classes by virtue of their common behavior in the tone sandhi.

<table>
<thead>
<tr>
<th>First syllable tone</th>
<th>Second syllable tone</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ia 阴平</td>
<td>IIIa 陰去</td>
</tr>
<tr>
<td>high/mid level</td>
<td>high fall</td>
</tr>
<tr>
<td></td>
<td>+</td>
</tr>
<tr>
<td>Ib 阳平</td>
<td>IIIb 陽去</td>
</tr>
<tr>
<td>high/mid depressed fall</td>
<td>depressed lower-mid level</td>
</tr>
</tbody>
</table>
3.1 Auditory description

Table 3: Phonetic transcription of examples of Wenzhou disyllabic words with input III tones on the second syllable, and input I tones on the first.

<table>
<thead>
<tr>
<th></th>
<th>Ia + IIIa 陰平 + 陰去</th>
<th></th>
<th>Ib + IIIa 陽平 + 陰去</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>male</td>
<td>female</td>
<td>gloss</td>
</tr>
<tr>
<td></td>
<td>['tsʰŋ]</td>
<td>tsʰŋ</td>
<td>green</td>
</tr>
<tr>
<td>vegetables</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>22(3)</td>
<td>4</td>
<td>11</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Ia + IIIb 陰平 + 陽去</th>
<th></th>
<th>Ib + IIIb 陽平 + 陽去</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>male</td>
<td>female</td>
<td>gloss</td>
</tr>
<tr>
<td></td>
<td>['sa2</td>
<td>dǒŋ?</td>
<td>cave</td>
</tr>
<tr>
<td></td>
<td>sa</td>
<td>dǒŋ?</td>
<td></td>
</tr>
<tr>
<td></td>
<td>22(3)</td>
<td>4</td>
<td>11</td>
</tr>
</tbody>
</table>

The auditory characteristics of words with I tones on the first syllable morpheme and III tones on the second are given in Table 3. For these I + III combinations, both speakers have a similar low-high pitch pattern, but differ in important ways. The male speaker’s words with these input tones consist of a long first syllable with pitch which is low if the input tone is [depressed mid-falling] Ib, and lower-mid if the input tone is [mid level] Ia. The male thus appears to clearly distinguish tone Ia from Ib in this position. The pitch contour of his first syllable often has a slight rise at the end. This rise is most salient when there is low pitch on the first syllable, and looks to be part of a commonly encountered intrinsic-suspendible anticipatory assimilation of the first syllable to the second in rate of cord vibration.

In contrast to this, there is neutralization of the difference between [high falling] and [depressed lower-mid level] input tones IIIa and IIIb on the second syllable. This carries a short high or high rising pitch, the latter being again part of an intrinsic perseverative assimilation aided by the presence of the voiced intervocalic consonant. The second syllable invariably ends in a glottal stop, which is part of the tone and not segmental. One very salient feature of the male speaker’s production of these combinations is their trochaic nature, the clear prominence on the first syllable being realized in its much greater length, and much shorter length of the second syllable. No other combinations in his disyllabic lexical sandhi have this trochaic nature.

The female differs from the male firstly in that her first syllable Ia morphemes sound to have the same low level pitch as her Ib morphemes, and so she appears to have neutralized the tonal contrast between her [high level] Ia and [depressed high
falling] Ib in word-initial position. Secondly, the female has much more evenly distributed prominence, with both syllables having similar lengths, and she thus lacks the male’s trochaic pattern. Like the male, she does not contrast IIIa and IIIb tones on the second syllable, but, presumably as a consequence of the greater length of her second syllable Rhyme, it is easier to hear a perseverative rise in second syllable pitch. This rise is more salient when there is a voiced intervocalic consonant.

3.2 Acoustic characteristics

Figure 4 shows the speakers’ mean acoustics (F0 and duration) corresponding to these auditory characteristics. F0 is plotted on the first and second syllable Rhymes, and on their intervocalic consonant if voiced. Each curve is the mean of at least three different words. In each of the panels, two F0 curves are shown, corresponding to a different second syllable input tone: one (lines with triangles) for the mean value with tone IIIa on the second syllable, and one (lines with dots) for the mean value with tone IIIb on the second syllable. Thus the F0 shape with dots in the top left panel of Figure 4 shows mean values for input tone Ia before input IIIb, and the F0 shape with triangles shows mean values for input tone Ia before IIIa. This graphing permits examination for possible assimilatory effects between the two syllables’ F0 shapes. The F0 shapes have been aligned with respect to Rhyme onset in first syllable, as the male’s first syllable is clearly the dominant one from the trochaic patterning. Thinner lines in the middle of the F0 time-course indicate F0 on the voiced intervocalic consonant, which was also sampled in mid hold. F0 is shown on the intervocalic consonant only in combinations with IIIB on the second syllable, because as explained in section 2.1, syllable-initial consonants on morphemes with “a” tones are voiceless.

For the male, it can be seen in Figure 3 that his first syllable Rhyme is on average about three times as long as the second. This no doubt contributes to the trochaic impression of these combinations. As far as his F0 is concerned, on the first syllable are seen two pairs of mildly concave shapes, separated by about 20Hz. The upper pair, in panel A, corresponds to input tone Ia, and the lower pair, in panel B, to input tone Ib. On the second syllable are four short F0 shapes with the same high peak F0 value (ca. 170 Hz) and falling offset perturbations. F0 shapes on second syllables with voiceless initial consonants (i.e. with input tone IIIa) are effectively level; F0 shapes on syllables with voiced initial consonants (tone IIIB) have rising F0. The difference between the onset F0 values on the second syllable Rhyme conditioned by the +/- voiced feature of the intervocalic consonant is about 20-25 Hz. These features are as expected from the auditory impression.
Figure 4: Mean F0 shapes for Wenzhou disyllabic words with input tones IIIa and IIIb on the second syllable preceded by input tone Ia (A) and input tone Ib (B) on the first. Top row shows the male values; bottom the female. Lines with dots show combinations with IIIb on second syllable; lines with triangles show combinations with IIIa on the second syllable. Axes = F0 (Hz) and duration (csec.).

On the female speaker’s first syllable Rhyme, although there are clear differences between the F0 time course associated with her two different input tones Ia and Ib, there is no acoustic evidence that she distinguishes these two tones in initial syllable position. For both input tones Ia and Ib, the F0 falls and then levels out, with a very slight anticipatory rise at the end. The initial fall is more abrupt with input tone Ia, and the quasi-level portion also lies some 10 Hz higher, a difference that t-tests show to be statistically significant at the 5% level at many points along the F0 contour. However, these differences, which are reminiscent of those between her IIa and IIb reflexes in isolation (Figure 2), are much more likely to have arisen as a function of a difference in syllable-initial consonants and/or depression associated with the two tones (section 1.2). A 10 Hz difference is about 1/12th of the female’s F0 range and would be
unusually small to underlie a tonal contrast. Contrast this with the difference between the Ia and Ib reflexes in the male which are separated by about 25 Hz. This is about 1/5th of his F0 range and clearly big enough to underlie a tonal contrast.

Unlike the male, the female speaker’s first syllable Rhyme is only marginally longer than the second. This presumably contributes to the relatively equal prominence of both her syllables and to the lack of trochaic percept. The female’s rising F0 shapes on her second syllable Rhyme are also clear. That the rising F0 can be heard as rising pitch is also presumably related to the length of the Rhyme over which the rise takes place. The difference between the onset F0 values on the second syllable Rhyme conditioned by the +/- voiced feature of the intervocalic consonant is about 30-40 Hz. It can be seen that the difference between the overall acoustic shapes of the two speakers’ I + III combinations lies not so much in the time course of the F0 which is quite similar in both, but rather in the differential timing relationship of the first and second syllable Rhymes to the F0.

3.3 Other descriptions

Published descriptions agree in describing a neutralisation on the second syllable, but differ on its realisation. Zhengzhang’s (1964:108, 1980:248, 2008:94) description is the closest to the data described in this paper. He describes the second syllable as a “light” 輕 [53] or [43], which becomes [5] in fast speech. Qian’s (1992:655) notation of [52] for pitch on the second syllable again implies a neutralisation of IIIa and IIIb, but in favour of an input IIIa high falling tone. As far as the first syllable is concerned, Qian’s (1992:655) descriptions are closest to the male speaker in this paper, recognising a distinction in pitch height between Ia, and Ib before IIIa/b. He transcribes Ia as [44] and Ib as [22], so his pitches are a little higher than observed here, and they also imply isolation realisation, at least for Ia. Zhengzhang’s transcriptions on the other hand imply a neutralisation of Ia and Ib on the first syllable realised as a low dipping or low rising pitch notated variously as [213], [11] [13], [113] and [214], the latter two said to be characteristic of slow speech. These are similar therefore to the female speaker.

3.4 Tonological interpretation

Accounting for these disyllabic shapes is not simple. The results above have shown that a solution has to answer several questions. In the male’s Ib + IIIa combinations, for example, how might one relate his isolation tonal values of [depressed mid falling] + [high falling] to sandhi values of [low level] + [short high];
or his isolation values of [mid level] + [lower-mid level] to [lower-mid level] + [short high] for his Ia plus IIIb combinations? Why the neutralisation of the very different [high falling] IIIa and [depressed lower-mid level] IIIb tones on the second syllable? Why the realisation of this neutralisation as a tone with short and high pitch? Whence the low level pitch corresponding to the [depressed mid-falling] I b pitch isolation tone on the first syllable? Why does the [mid level] isolation Ia tone have a lower-mid pitch on the first syllable?

It is possible to account for the relationship between the isolation tones and the sandhi shapes with a synchronic derivation using the isolation tones as unique underliers (Rose 2004). For example, the short high pitch of the second syllable tone which realises both IIIa and IIIb can be accounted for in the following way. Assume, non-controversially, tonological representation of IIIa as a high register tone with falling contour, i.e. [+U, HL] (Yip 1980, Bao 1990). Assume further, and also corresponding to the phonetic reality, that IIIb is [-U, H]. The short high pitch of the second syllable tone in I + IIIa/IIIb combinations, which is [+U, H], can then be derived by first changing the register of the IIIb tone to +U — such a register change is common in other Wu dialects and can be independently motivated in Wenzhou (Rose 2001). This results a sequence of HH autosegments for IIIb, and HL for IIIa tones, from which the second autosegment can then be deleted, conditioned by stress on the previous syllable, to yield the desired output of [+U, H].

Although they work formally to generate the correct output, I think it is highly unlikely that such complex derivational accounts can represent synchronic cognitive reality (for some more examples in another Wu dialect—Zhenhai—see Rose (1990). Perhaps the answer is a diachronic one. What would such a solution look like? It has been shown that in I + IIIa/IIIb combinations, there is an apparent synchronic neutralization of two phonetically very different [high falling] IIIa and [depressed lower-mid level] IIIb tones on the second syllable. The crucial information, of course, is that these phonetically different tones are nevertheless historically related, both being reflexes of the same historical III tone. Assume that the proto-form of this tone was similar to the modern [high falling] IIIa reflex, so that modern IIIa derives from a proto *HL with voiceless syllable-initial consonants, and modern IIIb comes from proto *HL with voiced. The occurrence of a short high pitch on the last syllable of words with historical III tones would then cease to be a problem: it would have arisen simply as the result of the historical *[high falling] III tone losing its falling portion under the influence of the preceding stressed syllable: *[σ + HL]w > *[σ + H]w.

What would need to be explained in this scenario is the isolation shape of the modern IIIb reflex, namely [depressed lower-mid level]. Such an explanation would require two parts. Firstly it would need to be demonstrated how a proto [high falling]
tone could change into a [depressed lower-mid level] tone. This is not particularly problematic and can be envisaged as a set of gradual changes of the form *[high fall] $^1$ $>$ *[341] $>$ *[231] $>$ *[233] $>$ [depressed lower-mid level], conditioned by the syllable-initial consonant, or depression, or both.

Secondly, and more importantly, we would have to assume that the morphotonemic relationship between the sandhi shape of the tone and its isolation shape became more and more obscure until it was no longer recoverable for native speakers, and that they consequently had different mental representations for both. In other words, speakers would start to learn the shape of tones on a word, and construct tonal representations accordingly. The sandhi shape of a tone could continue to exist relatively unchanged while the isolation shape underwent changes. There is already evidence for this kind of development from the fact that there appears to be less variation in the sandhi shapes within dialect groups than in the isolation tones (Ballard 1980, 1988)—a fact which suggests that it may be better to base proto forms on the sandhi tones rather than their isolation variants (cf. also Ting 1982). Moreover, I have yet to hear of a case where Wu speakers have started to unpack their tone sandhi into its constituent isolation tones: rather, changes in tone sandhi seem to usually involve simplification, along the lines demonstrated by the female in this paper (unlike the male, she also merges her second syllable Ia and Ib tones in I + I combinations).

A possible scenario for the development of I + III combinations is fleshed out in the tableau in Table 4. Morphemes taken from Table 3 are used as examples. On the right of the tableau are shown disyllabic words with the second morpheme carrying a historical III tone: 山洞 cave and 青菜 vegetables. Monomorphemic words are shown on the left.

The first stage represents a period when the proto form for tone III was *HL irrespective of whether the morpheme began with a voiced or voiceless stop, as in *dun cave 洞, or *tse vegetables 菜, and the morphotonemic relationship between isolation tones and tones in words was therefore clear: the morpheme had the same tonal representation irrespective of whether it occurred in isolation or on the final syllable of a polysyllabic word. In the second stage, the shape of the isolation tone begins to change. The reason for this may be no more than, as Lass (1997:325 ff.) entertains, that change is simply sublumary: it’s just what happens. I suspect, however, that the indexical function of Language may be implicated. Tonal detail may be like subtle differences in vowel quality in Germanic languages, manipulated by speakers to convey information on their group membership. Changes to the isolation shape continue in stage three. While the relationship between the tone’s isolation shape and

---

$^1$ For examples of proto-Oujiang tones reconstructed on the basis of their acoustics, see Rose (2009).
the sandhi shape remains clear, it can be assumed that the speaker constructs the same mental representation for both, and the realisation of the surface forms can be accounted for according to your preferred paradigm.

**Table 4:** Hypothetical stages in historical development of Wenzhou tone sandhi in I + III combinations. w = phonological word.

<table>
<thead>
<tr>
<th>monosyllabic words</th>
<th>disyllabic words</th>
</tr>
</thead>
<tbody>
<tr>
<td>山</td>
<td>洞</td>
</tr>
<tr>
<td>*sa</td>
<td>*duŋ</td>
</tr>
<tr>
<td>2</td>
<td>*[HL]w → *[431]</td>
</tr>
<tr>
<td>3</td>
<td>*[HL]w? → *[231]</td>
</tr>
<tr>
<td>4</td>
<td>*[LM]w → *[23]</td>
</tr>
<tr>
<td>5</td>
<td></td>
</tr>
</tbody>
</table>

There comes a point, however, when the isolation shape has changed so much that the morphophonemic relationship between its shape and the sandhi shape is too opaque (this is at stage 4). At this point it is reasonable to assume the speaker ceases to construct the same mental representation for the isolation and sandhi shapes. In effect, they can now be said to learn two different shapes for the same historical tone: one for the monosyllabic and one for the polysyllabic word. They may still have a rule which relates the two shapes, as was demonstrated by Hsieh (1970) for tone sandhi in Min, but this rule too may be lost. In the same way as a reason is needed for the changes in isolation shapes, one also needs to be advanced for the existence, integrity and perseverance of sandhi shapes. Perhaps there are two: tone sandhi in Wu has a clear demarcative function, signaling phonological words and syntactic boundaries, analogous to stress in English pairs like *blackbird* and *black bird*. However, like isolation tones, it is also possible that differences in sandhi shapes between dialects may also serve an indexical function.

At stages 5 and 6, changes are shown occurring to the disyllabic word tone shape as the result of first syllable stress. The first syllable becomes longer and its pitch lower; and the second syllable becomes shorter and loses its falling component. These changes, which are the natural consequences of the so-called positive and negative effects of stress (Kratochvil 1968), have been discussed in detail in Rose (2004).
caused the stress shift, or when it occurred, is of course not clear. I have introduced it at stage 5 because some other Ou dialects still show a high fall for tones IIIa/b in this position. Stage 6 represents the male speaker’s modern forms: his second syllable tone has lost its fall as the result of the negative affects of stress, and a thoroughly opaque relationship between isolation IIIa and IIIb tones and their word-final realisation in disyllabic words obtains.

How do the female speaker’s forms fit into this scenario? Recall that she did not have a trochaic pattern for I + III combinations; merged IIIa and IIIb on the second syllable as rising pitch; and also merged Ia and Ib on the first as a low level pitch. At least some of these differences can be accommodated fairly well if we assume (which is plausible, given the speakers’ age difference) that her forms represent a further stage in the development (shown at stage 7). Once again stress is the key. It was pointed out above that emergence of a rising pitch on the second syllable could be seen as a natural result of destressing the first syllable and restoring equal stress. The emergence of a rising pitch on the female’s second syllable is very important, for the following reason. If it were the case that the male’s short high [+U, H] pitch on second syllable in I + III combinations were synchronically derived from a HL tone under conditions of destressing, then it would be reasonable to expect that, if the equal stress were restored, the fall would re-emerge. But it does not: the female’s equally stressed I + III combinations carry a rising pitch on the second syllable, not a fall. I think this is possibly another indication that she does not have a relationship between her isolation IIIa and IIIb shapes and the sandhi shape in I + III combinations, but has simply learnt the shape of I + III combinations as a whole.

The development of I tones in these combinations is somewhat less clear. If one assumes a proto [mid level] pitch for Ia, the relationship between sandhi and isolation shapes for Ia remains fairly transparent until the stage represented in the female speaker with a [high level] tone in isolation and a [low level] tone in sandhi. The male may well have a dissimilatory rule relating his sandhied lower-mid level Ia to its mid level isolation form, and thus have the same representation for Ia in both isolation and sandhi forms. It is not clear whether the female would have a rule relating her two high and low level forms, which are far more different than the male’s. This could be tested along the lines suggested by Hsieh’s (1970) “experimental phonological” method.

---

2 In Xiàngyáng 象陽 to the east, for example, the second syllable tone in I + III lexical sandhi combinations has a high falling pitch, like isolation IIIa. The pitch of the first syllable tone rises (from mid if the input is Ia, from lower if Ib). There is no trochaic impression.
For Ib, if its proto form were *[mid level] as with Ia, it could also have dissimilated in sandhi (before the high falling III) to [low level], and thus there could have been a stage when both sandhi forms and isolation forms had the same mental representation. A later change in isolation from *[mid level] to the [depressed mid fall] seen in the male and [depressed high fall] in the female would have obscured the morphotonemic relationship.

4. Summary

In his discussion of the phonological and phonetic processes underlying the wide range of tone sandhi phenomena in Chinese dialects, Chen (2000:81) states that “... there remains a vast assortment of tonal alternations that defy classification and description, let alone explanation”. This paper has presented acoustic and auditory descriptions of lexical tone sandhi in ping plus qu combinations from two Wenzhou speakers who differ in both their isolation tones and tone sandhi. It has shown that the relationships between the isolation shapes and sandhi shapes of the tones are indeed complicated, but has speculated that this is probably because the sandhi shapes and the isolation shapes have long since parted company, such that they may no longer share the same mental representations, speakers learning them separately. If this is the case—and Ballard (1980, 1988) made this point long ago—it really makes no sense to try to account synchronically for the different tonal shapes in derivational (or constraint-based) approaches in terms of realisations of a unique underlying tonological unit, although derivational analyses may still be of use in pointing out possible historical developments. It will make sense, however, to attempt to reconstruct the sandhied shapes of tones on polysyllabic words, rather than just their shapes on monosyllabic words. One final, and important, point is that this kind of insight cannot be won from synchronic data alone—one needs to be able to identify historical tonal categories to be able to interpret the synchronic data correctly.
References


Qian Nairong. 钱乃荣. 1992.《当代吴语研究》。上海：上海教育出版社。


Reverse-engineering Tones and Tone Sandhi in Wenzhou Dialect


Zhengzhang Shangfang. 郑张尚芳. 1964. 〈温州方言的连续变调〉，《中国语文》 1964.2:106-152。

Zhengzhang Shangfang. 郑张尚芳. 1980. 〈温州方言尾词的语音变化一〉，《方言》1980.4:245-262。

Zhengzhang Shangfang. 郑张尚芳. 2008. 《温州方言志》。北京：中華書局。