

# Tianjinese Lexical Tone, Tone Sandhi, and Prosody: Amplitude-f<sub>0</sub> Dependency under Prominence in Mandarin

Deborah S. Davison

Department of Research Compliance  
Stanford University, Stanford, CA 95070 USA  
ddavison@stanford.edu

## Abstract

In Tianjin Mandarin, f<sub>0</sub> and amplitude track together on low (L) tone T1, low rising (LH) tone T3, and high falling (HL) tone T4, as is claimed to be true generally of Standard Mandarin (see Yip 2002). In contrast, lexical high tone T2 f<sub>0</sub> tracks least systematically with amplitude, whereas tone-sandhi derived high tones on T3 and T2 co-vary with amplitude most consistently. The tone sandhi induced insertion of T2 high f<sub>0</sub>/db pitch accent between adjacent low tones co-occurs with phrase level prominence on the disyllabic word or phrase in which it appears. These facts are consistent with other evidence in Tianjinese of H pitch accent surfacing before L tones: f<sub>0</sub> on neutral tone 0 before L Tone 1 and LH Tone 3 syllables is high, on hypothesis in the same phrasal environments, and f<sub>0</sub>/amplitude excursions on T3 before L T1 often are as well. These data support the author's hypothesis (Davison in press) that a high f<sub>0</sub>/db pitch accent marks intonational phrase, contrary to the prevailing assumption that lexical tone languages such as Chinese do not permit phrasal pitch accents internal to intonational phrases.

## 1. Introduction

The data are from a passage in the modern Chinese novel *Ah!* by Feng Jicai read by WSW, a 40 year-old male TJ speaker. F<sub>0</sub>, amplitude, and duration were measured using Praat.

### 1.1. TJ tone and tone sandhi features

Tianjin Mandarin (TJ) dialect closely resembles Beijing Mandarin (BJ) in contrasting four lexical tone categories: T1, T2, T3, T4 or A, B, C, D respectively. Tone contour relative values are represented phonetically using a 1-5 low-high scale, phonologically as a series of Ls and Hs.

Figure 1: *BJ and TJ Basic Tones.*

BJ: A high level 55 'H'	TJ: A low level 21~11 'L'
B high rising 35 'LH'	B high 45~55 'H'
C low 11 'L'	C low rising 13 'LH'
D falling 51 'HL'	D falling 51 'HL'
0 'neutral tone' toneless	0 'neutral tone' toneless

Note that while the phonological representations of tone contours in the two dialects are identical (though differently distributed among the four tone categories), BJ has two high register tones A and B, whereas TJ has two low register tones A and C, in addition to having one falling and one high tone.

Figure 2: *TJ Examples from WSW data (Segmental transcription in Pinyin).*

A fei21 'not'; ta21 's/he, it'  
B lai45 'come'; bie45ren45 'other people'  
C jiu13 'wine, liquor'; hao13 'good'  
D hui51 'can, be able to'; you51 'again, both'

"Tone sandhi" (TS) refers to the replacement of one tone category by another on the first, 'target' syllable of disyllabic words and phrases, conditioned by the tone category of the second, 'trigger' syllable. While BJ has one TS rule, TJ is traditionally described as having four. Recent studies such as Milliken et al. 1997 have reanalyzed TJ TS as resulting from phonological auto-segmental association processes and the application of language universal constraints.

Figure 3: *TS in BJ and TJ Syllables are separated by '-'*.

BJ: CC->BC L-L -> LH-L    TJ: AA->CA L-L ->LH-L  
CC->BC LH-LH ->H-LH  
DA->BA HL-L ->H-L  
DD->AD HL-HL ->L-HL

Figure 4: *Examples of TJ TS from the data.*

AA->CA deng21guang21 'lamplight' -> deng13guang21  
CC->BC ling13dao13 'leader' -> ling45dao13  
DA->BA lei51hua21 'teardrop' -> lei45hua21  
DD->AD xian51zai51 'now' -> xian21zai51

## 2. Tones, Tone Sandhi, and Prominence

As shown in Davison (in press), in natural read speech TJ TS does not apply in all phonologically sanctioned disyllabic environments. TJ TS effects low register dissimilation within disyllables by insertion of a single H tone at the internal boundary between the two syllables. Only TJ tones B and D have underlying high register Hs. TJ BB H-H disyllables a single H target results from autosegmental association to both syllables. That low register but not high register disyllabic sequences must dissimilate supports a discourse functional interpretation. I propose that when disyllabic lexical or syntactic phrases occur as Prehead (prenuclear) or Head (Nuclear) constituents of the intonational phrase (IP) in which they occur, they exhibit the tone sandhi effects described above. Though the evidence will not be provided here (see Davison to appear), the neutral tones as well as sandhi context constituents are equally subject to H pitch accent insertion, putatively at a prosodically defined slot adjacent to IP boundaries. Evidence that underlying lexical H in TJ B and D tone syllables also may 'attract' pitch accent, identifiable by

increased amplitude to f0, is also examined, in a forthcoming paper. Taken together these facts support an interpretation that H f0 is perceptually more prominent than L f0, making H the more natural candidate to function as a prosodic nucleus marker, hence its phonological ‘insertion’ in the environments mentioned above. The present study suggests that high amplitude co-varies with H f0 just in case prosodic prominence is being conveyed. Based on impressionistic observation, it appears likely that high amplitude also co-occurs with prosodically prominent L f0 lexical tones. The relative distribution of high amplitude H and L f0 tones in TJ will be examined in future work.

In WW’s speech several possible permutations of DD HL-HL exist: the final syllable may reduce and become toneless, its H replaced by continuation of the contour declining from the first syllable’s high, as in yun51dong51 ‘political campaign’ -> yun53dong21; it may delete initial H via DD->AD TS, as in xian51zai51-> xian21zai51 ‘now’ above; and, less frequently, the leftmost internal L may delete creating a BD H-HL type contour, as in jiu51 yao51 ‘then must’ -> jiu45 yao41. As such DD sandhi for this speaker is more complex than generally reported and is not further examined in this paper. C’s H is often realized as mid under influence of the preceding L, though it too can be realized as super-high f0/db in appropriate prosodic environments.

### 3. Prominence Domains Defined

Davison (in press) argues that TJ natural speech is parsed into right-headed footless prosodic phrasal domains. Further analysis of the phrasal organization of the same read text by speaker WSW in this paper suggests that the phrasal domains may contain a minimum of one and a maximum of two (three? See below) prominence-attracting heads, referred to as Pre-Heads and Heads, at left and right edges respectively. Only the Head may be followed by a pause. I interpret pauses to mark IP boundaries; as such they tend to be isomorphic with syntactic boundaries but are not always so, as is typical of prosodic domain boundaries. Prosodic phrases may be one to several syllables in length. In long phrases, a phrase-internal “pivot” syllable marked by relatively long duration and/or high amplitude, usually coincident with a sentence-intermediate syntactic boundary, sometimes occurs, introducing the complication of potential limited hierarchical structure as well as indeterminacy regarding Head identification.

#### 3.1. Domain Elements

Phonetically, Pre-heads and Heads are characterized by domain-internal local amplitude peaks as well as optional extended duration. Weakly stressed extrametrical syllables may precede or follow them and may be longer in duration than their heads especially when adjacent to phrase boundaries. When only one head occurs in an IP it is analyzed as the Head. It can occur at either left or right phrase edge: #liang51jing32jing21de1# ‘sparkling’ vs. ding13le0 ni13 ‘[I criticized you’ -> #ding12le5ni12#.

#### 3.2. Maximum One or Two Heads Per Phrase?

A question arises whether the data are better described by breaking up two-headed prosodic phrases into two

independent phrases, as implied by my earlier analysis (Davison in press), that intonational phrases are right-headed. This approach has the appeal of resolving the hierarchical structure question in the negative. Other arguments in favor include: a) both Prehead and Head are characterized phonetically by local amplitude and durational peaks; b) prepausal lengthening may be an orthogonal pause conditioned feature: only pause obligatorily (neutral tone 0 optionally, sometimes followed by pause) forces phrase formation; d) it can be difficult to determine whether a series of local amplitude peaks preceding a large phrase boundary are related to each other as head to head or pre-head to head, and which is which; e) prominence on pre-heads can be analyzed as phrase-initial strength assigned the leftmost lexical trochee, cf. Shih 1997 and Duanmu 2002.

These are strong arguments that are beyond the scope of this paper to refute. However, an important outstanding problem is how best to analyze trisyllabic words and phrases. While not as frequent as disyllabic constructions, they are not unusual in this text and in the language generally, and are often pronounced with strong-weak-strong (SWS) rhythm. In context of her discussion of Chen 2000’s prosodic constraint on BJ tone sandhi: “NOSTRADDLING: Immediate constituents must be in the same MRU [minimum rhythmic unit]”, Yip notes that BJ exhibits a “clear...preference for MRUs to be two to three syllables long” (Yip 2002:124). Trochaic analysis would seem to require a two-trochee, hence two phrase, interpretation of trisyllabic rhythmic units, rather than, in the terms of this paper, a prehead-head one. Consider the example phrases bounded by pauses on each edge: [[wu45] [zhong51yi51]] ‘surname +personal name; Zhongyi Wu’ -> ##wu45zong21yi41, and ##[[hao51] [zu13ren51]] ta21## ‘Chairman Hao he...’. The 2<sup>nd</sup> and 3<sup>rd</sup> syllables in each form a word over which a trochee should be built, but is not: both are audibly rhythmically SWS(W). Thus I assume with Yip that for purposes of this paper both disyllabic and trisyllabic rhythmic units are basic in TJ Mandarin, alternating SW/WS rhythms are observed language universally (Hayes 1995), and, e.g., in the second phrase above, trisyllabic phrases may exhibit both pre-head and head prominence.

Table 1: *Amplitude peak, f0 trajectory, and syllable duration in two sample trisyllabic phrases.*

Wu 45	Zong51 >zong21	Yi51	Hao 51	Zhu13 >zu13	Ren 51	Ta21
79db	68db	70db	85db	76db	82db	80db
128-170-145 Hz	88-115-119	119-138-115	140-134-135	134-148	151-159-157	129-134-127
.241Sec	.207	.154	.262	.093	.193	.349

Zhong51 (pronounced ‘zong’ in TJ, similarly Zhu13 ‘zu’) has undergone DD->AD TS, hence its low f0 relative to following Yi51. Notice the low amplitude peaks on Zong51>21 and Zu13, the middle syllables of the putative SWS trisyllabic phrases in each example. These two syllables are audibly weaker in intensity than the other syllables in each phrase. For Zong51 low amplitude is accompanied by low f0. However, since Hao51 and Zu13 in the second example have comparable if inverse pitch contours and f0 values, low amplitude/duration is the likely auditory cue to SWS rhythm for this example.

#### 3.3. Relationship of Amplitude to Pitch

Yip 2002:292 notes, “In some languages, like Mandarin Chinese, there is a close correlation between pitch contour and amplitude contour...if not only pitch but also duration cues are removed, while amplitude contours are retained, a relatively high degree of pitch recognition of about 60 per cent is still possible....The contribution of duration [is] found to be non-significant except in the case of the third tone, which is known to be much longer than the other tones in isolated syllables. ...We conclude then that F0 is indeed the primary cue for the discrimination of tones in natural languages, but that to the extent that the F0 contour is mirrored in the amplitude, amplitude too is a useful cue.”

Notice that locally amplitude does not uniformly track with f0, either across phrases—Wu45 has higher f0 than Ren51 but lower amplitude—or within them: Hao51 has lower f0 than Ren51, but higher amplitude, though both have underlying tone D 51 HL. The global tendency for the 350 syllable tokens examined in this paper is for amplitude and f0 to track together, at least in the sense that low tones rarely occur with high amplitude, see below. However, examples such as these from the data, in which f0 and amplitude do not track together within a local phrasal domain, are not unusual.

When amplitude and pitch do not track together, amplitude is not a ‘useful cue’ to tone discrimination. Instead, on hypothesis, amplitude in these two examples is a ‘useful cue’ to underlying rhythmic structure. Tracking of amplitude-f0 is examined across data tokens in the following section, in support of the argument that mismatch may be systematically employed to implement local stress and prominence contrasts while preserving phonemic tonal contrasts by f0.

#### 3.4. Neutralization of tone contrasts

The most important exception to the general tendency to preserve tonal contrasts in Mandarin is the existence and creation of phonemically toneless syllables, the former as lexicalized neutral tone syllables and the latter via loss of tone under de-emphasis, post-head and elsewhere. That neutral tone syllables are more significant for Mandarin Chinese than, for example, Cantonese Chinese is evidenced by the fact that in Cantonese reduced, toneless syllables are generally restricted to phrase final particles, whereas in BJ Mandarin neutral tone 0 is observed to occur in as many as 25-30% of syllable tokens in natural speech (Duanmu 2002) and is not restricted to phrase final position. Neutral tone is also prevalent in these read speech data as read—not written, since written Chinese, even Baihua Wen ‘Spoken language Writing’ style as in this novel, avoids use of toneless particles that are not essential to comprehension.

The other important exception to tone preservation in TJ Mandarin as well as most other Mandarin dialects is neutralization due to TS. Unlike neutralization under post-head emphasis, TS potentially applies throughout the utterance. According to my analysis of TJ, when the goals of expressing prosodic prominence relations and preserving underlying tonal contrasts conflict, in TS contexts tone neutralization occurs, in favor of isolating a H tone within the disyllabic domain to signal that the word containing it is prosodically prominent. In leftmost TS position, prominence marking is favored over contrast preservation, i.e. prominence wins.

Unlike neutralization under post-head emphasis, TS potentially applies throughout the utterance. Also, three of the

Given that modification of one syllable’s underlying tone enhances the word’s prominence at the expense of loss of lexical contrast, TS is thus predicted to be constrained or favored to occur in disyllables in phrasally prominent positions, viz., as heads or preheads.

DD is interesting because what is lost in both DD->AD and DD->BD TS permutations (see above) is preceding D’s fall. The Praat pitch extractor rarely extracted a falling pitch for tone D in my data either in isolation or in multi-syllabic (including TS) contexts, possibly due to limitations of its pitch extraction algorithm. Alternatively, the gesture producing falling pitch may be unique in some other way; historically it is argued to be a reflex of lost syllable-final \*-s.

A noticeable lacuna is the absence of TS for BB H-H. As suggested in Davison 1989 (also Duanmu 2002), pan-dialectally dissimilation of low tones is much more frequent than high tones, though exceptions exist. Dialects that dissimilate adjacent high falling tones are more numerous, those that dissimilate adjacent level high tones most rare. BJ/TJ T1/2 H respectively exemplify the latter general pattern.

Absence of dissimilating H-H is explained as follows: pitch accents in TJ are assigned to words, not individual syllables. Mandarin words are typically 1-3 syllables in length. Identification of an appropriate pitch accent target must be made for 2- and 3-syllable words/phrases. Assuming that autosegmental association of phonological H-H(-H) results in merger of adjacent Hs into a single rightmost H pitch target, the pitch accent target identification problem is resolved for the B-B(-B) environment. TJ TS similarly creates a relatively unambiguous target: a single, isolated H.

#### 4. Discussion: When Amplitude Tracks Pitch

To examine whether f0 and amplitude track independently in TJ, acoustic measurements using Praat were made of 350 tokens of full-tone syllables and db/f0 plotted for first and second data points, left edge of extracted pitch and first excursion if any, respectively. Sandhi change tones were recorded as tokens of the tone changed to, not basic tone value. The wide T to f0 category range reflects the data’s origins in natural read speech. T4 HL’s H shows the strongest db/f0 correlation, T2’s H the weakest (Figures 5 and 6). T4 f0 ranges across db categories are most discrete, T2’s least (Table 2). 151 tokens were then re-measured and TS changed and ‘basic’ lexical tones broken out. Figures 7 and 8 show that for ‘basic’ tones T2 and T3 (also T1, data not shown), f0/db are weakly correlated if at all. In Figure 9 the TS changed tone ‘target’ and ‘trigger’ rightmost conditioning sandhi tone f0/db tokens are combined. As expected, only 4 of the lowest 16 and 10/13 highest f0 tokens are targets, since the TS change is to a high pitch. The last 16 tokens show the strongest positive f0/db correlation of the second data set. TS target-trigger pairs across pre-head/head environments were also compared. No conclusive correlation with head type was found in support of the dual head hypothesis, though more data should be compared in future studies.

#### 5. Conclusion

An intonation model of TJ read speech is outlined based on right-headed prosodic phrases with optional leftmost Preheads. Heads are characterized by local relative high amplitude and optional pause adjacency induced extended

duration. TJ TS promotes identification of Head. The implications of Fig. 9 are, contrary to hypothesis though consistent with a prominence-marking function, that TS target co-occurs with high f0/db. Analysis of additional data, speakers, pitch accent, and H insertion environments is needed to determine whether f0/db dependency, instead of being pervasive at the local phonological level, is systematically selective of discourse prominent environments.

Table 2: Number & f0 range of tokens/T/data point/db

#range	Pt.	70-74db	75-79db	80-85 db
T1	1 <sup>st</sup>	26(75-140)	20(105-153)	10(108-150)
T1	2 <sup>nd</sup>	27(70-140)	18(108-148)	10(100-140)
T2	1 <sup>st</sup>	19(75-160)	21(108-180)	12(140-180)
T2	2 <sup>nd</sup>	20(88-170)	22(108-180)	11(138-180)
T3	1 <sup>st</sup>	29(90-160)	26(100-158)	18(108-170)
T3	2 <sup>nd</sup>	20(90-170)	22(140-180)	11(138-180)
T4	1 <sup>st</sup>	17(110-145)	25(135-175)	24(152-180)
T4	2 <sup>nd</sup>	19(110-150)	27(130-180)	25(135-180)

Figure 5: F0/db, First data point T4: db rises with f0

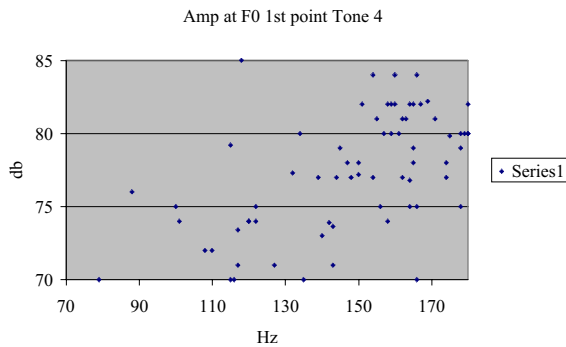


Figure 6: f0/db, lexical T2 only; no correlation to db

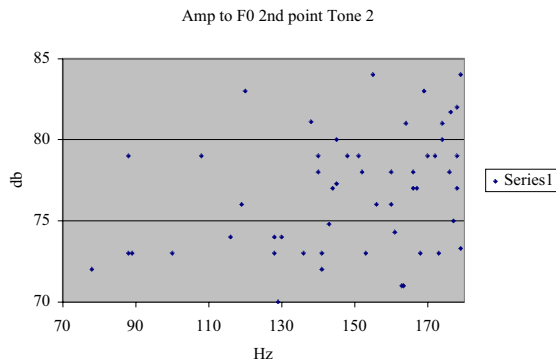


Figure 7: Lexical T2 f0/db does not correlate

## 6. References

[1] Chen, M., 2000. *Tone Sandhi*. Cambridge U Press.  
 [2] Davison, D., 1989. *Lexical Prosodies of Mandarin: Comparative Evidence from Northern Chinese Dialects*. Doctoral Dissertation, U. California, Berkeley.  
 [3] Davison, D., in press. *Tianjin Tone Sandhi in Action*. Tianjin: *Proceedings of the Chinese National Phonetic Assn*.  
 [4] Duanmu, S., 2000. *The Phonology of Standard Chinese*. Oxford: Oxford University Press.

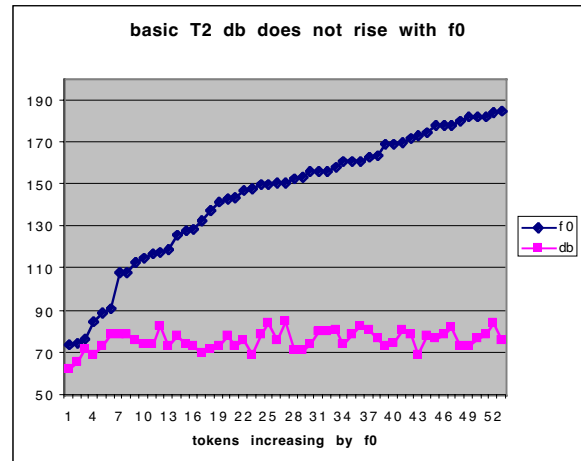


Figure 8: Lexical T3 shows slight f0/db correlation

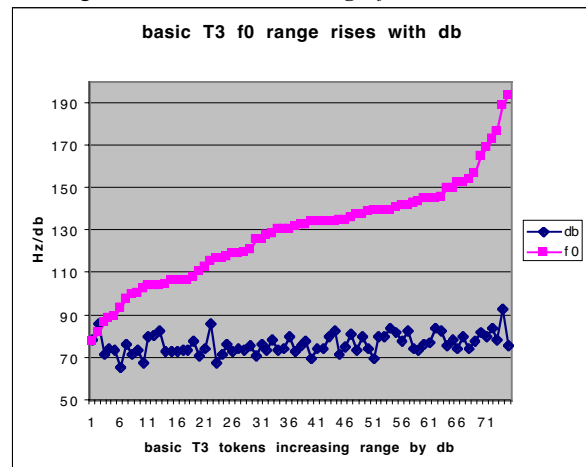
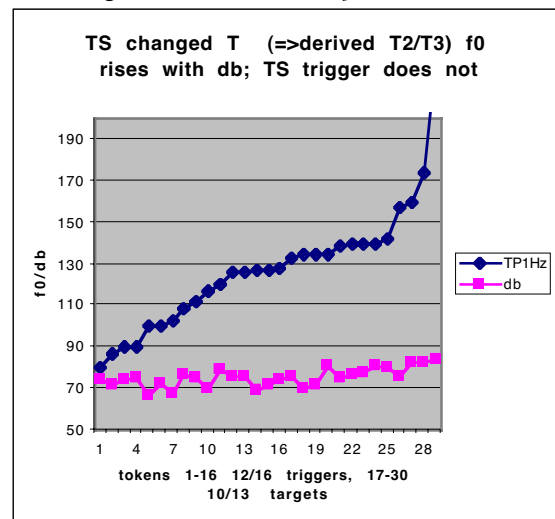


Figure 9: TS derived T2/T3 f0 rises with db



[5] Hayes, B., 1995. *Metrical Stress Theory*. U. Chicago Press.  
 [6] Wang, J., 2002. *Youxuanlun he Tianjinhua liandu biandiao ji qingsheng* "On tone sandhi and neutral tone in the Tianjin dialect within the framework of Optimality Theory." Beijing: *Zhongguo Yuwen No. 4*.  
 [7] Yip, M., 2002, *Tone*. Cambridge University Press.