Tone sandhi, prosodic phrasing, and focus marking in Wenzhou Chinese

Scholz, F.

Author: Scholz, Franziska
Title: Tone sandhi, prosodic phrasing, and focus marking in Wenzhou Chinese
Issue Date: 2012-10-18

Chapter 5

Tonal coarticulation as prosodic marker in Wenzhou Chinese

5.1 Introduction

5.1.1 Tonal coarticulation in Chinese

The implementation of lexical tones in tone languages is influenced by neighboring tones, just as segments coarticulate with other segments. A number of instrumental studies across different languages have shown that neighboring tones influence each other in a way that affects the realization of both tones, but have also pointed out cross-linguistic differences in the exact details of these coarticulatory influences.

For example, the realization of lexical tones in Thai (Gandour et al. 1994; Potisuk et al. 1997), Mandarin Chinese (Xu 1997), and Taiwan Min (Wang 2002) in naturalistic speaking conditions has been shown to be asymmetrically influenced by neighboring tones, with carryover coarticulation exerting a greater influence than anticipatory coarticulation. When Mandarin speakers were instructed to keep their speaking manner constant and not implement the natural stress difference between syllables, however, it could be shown that the two effects were similar in strength (Shen 1990b). For naturalistic speech in Malaysian Southern Min, it has also been argued that anticipatory and carry-over coarticulation are comparable in magnitude (Chang & Hsieh 2012).

In all cases, it was found that the coarticulation effect was mainly localized in the part of the syllable that was adjacent to the influencing context, and that the influence of the neighboring syllable's tone decreased with greater distance from that syllable. However, the exact nature of the coarticulation has also been shown to differ between languages. For Thai and Mandarin Chinese for example, it was argued that carryover coarticulation is assimilatory in nature, such that e.g. a high tonal offset before a rising tone on the following syllable raises the onset of that rise. In contrast, anticipatory coarticulation was found to be primarily dissimilatory, such that speakers magnified the differences between two successive tonal targets across syllable boundaries (Gandour et al. 1994; Xu 1997). In contrast, Peng (1997) for Taiwan Min, Han & Kim (1974) for

Vietnamese, and Wang, H. S. (2002) for Malaysian Southern Min showed that the manner of tonal coarticulation (assimilation/dissimilation) varied more between the individual tones than between the direction of coarticulation (anticipatory/carry-over).

Furthermore, languages also vary with respect to the aspect of the tonal production that is influenced by coarticulation. For Thai, it has been argued that coarticulation mainly affects the height aspect of tonal implementation (Gandour et al. 1994), whereas for Mandarin, slope seems to be the variable that is affected most (as in Xu 1994, but see Shen 1990b for different findings). For Vietnamese, both parameters seem to be affected by coarticulation (Han & Kim 1974). In sum, while neighboring tones affect the realization of tonal contours in all tone languages that have been studied in detail, the languages may vary in the exact type, extent, and direction of the influences from one tone onto another.

5.1.2 Contextual influence on coarticulation

For segmental coarticulation, a number of studies have shown that its extend depends on different contextual factors. For example, prosodic structure has been shown to influence the magnitude of coarticulation in several respects. In contexts where two segments are separated by prosodic boundaries of different levels (prosodic word, prosodic phrase, intonational phrase, utterance), it has been shown that they coarticulate less with each other if they are separated by a higher-level prosodic boundary, compared to a lower-level prosodic boundary (Byrd & Saltzman 1998; Cho 2004, 2006; Fougeron & Keating 1997; Jun 1998).

At the same time, it has also been shown that the strength of a syllable influences the magnitude with which it coarticulates with neighboring segments. Specifically, a greater coarticulation effect has been shown for unaccented compared to accented syllables (Cho 2004), and for unstressed compared to stressed syllables (Cabré & Prieto 2005; de Jong et al. 1993). While prosodic strength of a syllable and the magnitude of a neighboring prosodic boundary may be connected, it has also been shown that the two effects are to some extent independent of each other (Cho & Keating 2009). In that sense, segmental coarticulation serves as a marker for both prosodic boundary level and positional prosodic strength (Cho 2011).

Most research on the phonetic correlates of prosodic structure has been concerned with intonation languages like English or French. Only a few studies have looked at segmental effects in tone languages (Cao & Zheng 2006; Hayashi et al. 1999; Hsu & Jun 1998; Pan 2007a; Zheng et al. 2006) or included them as part of a cross-linguistic comparisons (Keating et al. 2003). All the above mentioned studies only included two or three speakers each, and some reported contradictory results. For example, of the two Taiwanese speakers investigated in Hayashi et al. 1999 and Keating et al. 2003, only one distinguished more than two levels of prosodic structure in the strength of the articulation of post-boundary consonants. On the other hand, Pan (2007a) reports that for her three Taiwanese speakers, the coarticulation of pre-boundary segments varied with the strength of the intervening boundary, with higher boundaries inducing less nasal coarticulation.

Segmental coarticulation was also found to be reduced across higher prosodic boundaries in Mandarin Chinese by Zheng and colleagues (2006), compared to lower prosodic boundaries. However, the question whether prosodic boundaries also affect tonal coarticulation has only been asked once by Pan & Tai (2006) in a small-scale study with three speakers. They report no statistics, but observe that the F_0 range of falling tones is greater when this tone precedes or follows an IP boundary than preceding or following a lower prosodic boundary.

The current chapter proposes to investigate prosodic structure by looking at the realization of the lexical tones in Wenzhou Chinese. The idea is as follows: if prosodic boundaries and prosodically strong positions (prosodic heads) can induce a strengthening and coarticulatory resistance for segments, the same should happen to lexical tones in tone languages. Consequently, if prosodic levels are the primary determiner of tonal coarticulation, lexical tones at the edges of higher prosodic constituents should coarticulate less with neighboring tones than lexical tones at the edges of lower prosodic constituents. If prosodic strength is the main factor to influence tonal coarticulation, tones in prosodically strong positions should show greater resistance to coarticulation, and be implemented in a more independent way compared to lexical tones in prosodically weak positions.

5.1.3 Focus effects on tonal coarticulation

Recent years have seen a number of studies investigating the effects of focus on the implementation of lexical tones. Across these studies, “focus” as a term is used both for contexts in which a speaker corrects or contrasts a part of an utterance (sometimes called ‘contrastive focus’), and for contexts in which a speaker answers a question containing a wh-expression (sometimes called ‘information focus’), and in both types, the constituent corresponding to the wh-expression or the corrected part of the context sentence is considered to be focused.

To mark these focused constituents, it has been found that, outside of lengthening, F_0 modification is applied to focused syllables and words in Mandarin or Standard Chinese (Wang & Xu 2006; Xu 1999), Cantonese (Gu & Lee 2007b), Shanghai Chinese (Chen 2009), and Taiwan Mandarin (Chen et al. 2009; Xu et al. 2012). In these dialects, the tones on focused constituents are implemented within an expanded F_0 range under focus, such that high targets were realized higher and low targets were realized lower than in control contexts.

However, taking a closer look at the mechanisms of F_0 expansion of tonal contours under focus in Chinese, Chen & Gussenhoven (2008) for Standard Chinese and Chen (2009) for Shanghai Chinese have argued that taken together, the F_0 adjustments lead to an enhanced implementation of the tonal contours that cannot just be reduced to F_0 expansion. For example, the rising/falling tonal trajectories in Chen & Gussenhoven 2008 were implemented with magnified and more distinct movement gestures under focus, and exhibited less coarticulatory influence from neighboring tones.

Similarly, Chen (2010) argues that tonal articulation in post-focal position, which has been assumed to be lowered or compressed compared to focused positions (Jin 1996; Xu 1999, 2005), is better accounted for as hypoarticulation. In a mirror image to the strengthening of tonal implementation under focus, tones in post-focal positions are better understood as being weakly implemented, similar to tones in prosodically weak elements (Chen & Xu 2006). A by-product of this weak implementation is a greater susceptibility to coarticulatory influence from neighboring tones (see also Chen & Gussenhoven 2008).

In sum, there is evidence that speaks for a strengthening of tonal implementation under focus, which also reduces the amount of coarticulatory influence of neighboring tonal contours onto the strengthened syllable. The current chapter will investigate whether this strengthening under focus is similar to the strengthening of segments in prosodically strong positions, and whether both types of strengthening exert a comparable influence onto the magnitude of tonal coarticulation.

5.1.4 Current experiments and hypotheses

The current study was designed to answer two connected research questions: (i) which contextual factors affect tonal coarticulation in Wenzhou Chinese, and (ii) how do these factors interact with the strengthening of tonal implementation induced by focus? In accordance with the findings in Xu 1994, 1999 for Standard Chinese, the amount of coarticulatory influence was measured as a change in tonal slope in the implementation of the tones on the target syllables. In the following, the first experiment is described, which tests the predictions on prosodic structure. The predictions on the influence of focus will be tested in a second experiment, presented in section 5.5.

Since it has been shown in the investigations on segmental coarticulation that prosodic boundaries (e.g. prosodic word vs. prosodic phrase boundary) and prosodic strength (such as stress) exert different types of influence on the coarticulation of adjacent segments (see section 5.1.2), both elements of prosody are taken into consideration here. Specifically, the test phrases were designed in a way that they would show different results if tonal coarticulation was mainly boundary-dependent, or if it was mainly dependent on prosodic strength.

In order to tease apart the prosodic boundary from prosodic headedness effects, two types of morphosyntactic structures involving a syntactic VP were investigated: adverb-verb structures and verb-object structures. In most syntax-to-prosody mapping algorithms, a distinction is made between arguments and adjuncts of a VP. While arguments can form a prosodic domain together with their heads, adjuncts are mapped onto a separate prosodic domain (Gussenhoven 1992; Samek-Lodovici 2005; Truckenbrodt 1995, 1999).[16] Under this assumption, the two structures outlined above would result in two different prosodic outputs. The adverb-verb structures would be separated by a prosodic phrase boundary, since the adverb is adjoined to the VP, while the verb-object structure would be mapped into one prosodic phrase together.[17]

---

[16] In impressionistic accounts of syntactically conditioned tone change in Chinese dialects, it has been argued that for the purpose of prosodic phrasing, VP-adverbs have to be treated differently from sentential adverbs to account for the differences in perceived tonal realization (Chen 1987, 2000; Lin 1994; Soh 2001). In these studies, observations about tonal realization were taken as evidence, and the syntactic analyses were based on these observations. In contrast, the current study presupposes an unambiguous analysis of the syntactic structures, and tests the influence of the derived prosodic structures in an instrumental and quantified way.

[17] For the purpose of syntax-to-prosody mapping, it is irrelevant whether a preverbal adverb is analyzed as adjunct to the VP, or to a higher functional projection like IP or TP, since both structural configuations result in a similar prosodic mapping. Likewise, it does not matter for prosody whether adverbs are analyzed as adjuncts, or as specifiers in empty-headed functional projections. Both analyses would have the same consequence for syntax-prosody mapping, namely the pre-verbal adverb being positioned outside of the core VP. See for example Alexiadou 1997; Cinque 2004; É. Kiss 2009 for recent discussion of the syntactic analysis of adverbs.

At the same time, the two structures also differ in the prosodic positional properties of the two constituents involved, namely in the position of the prosodic heads. In intonation languages, prosodic headedness is usually manifested in stress or nuclear accents on prosodically strong positions (see e.g. Fougeron 1999 for an overview). In tone languages such as Chinese, prosodic heads are commonly associated with the preservation of tonal features (Chen 2000; Yip 1999). This means that lexical tones in prosodically strong positions tend to be preserved and articulated clearly, whereas tones on non-head syllables are prone to tone change induced by neighboring tones (Yip 1999).

For the two prosodic structures tested here, the difference in morphosyntactic configuration results in a different distribution of prosodic headedness. For the adverb-verb structures, which are assumed to map onto two prosodic phrases, each of the phrases has its own prosodic head (or prominent position), in line with the common assumption of a one-to-one relation between prosodic constituents and prosodic heads (Hayes 1995). In the case of the verb-object structures, which are phrased in one prosodic phrase together, the common assumption of "nonhead prominence" or "Non-Head Stress" (Duanmu 1995, 2005, 2007, 2012) determines that the constituent that is not the syntactic head is assigned prosodic prominence and thereby attains prosodic head status (Chen 2000; Cinque 1993).

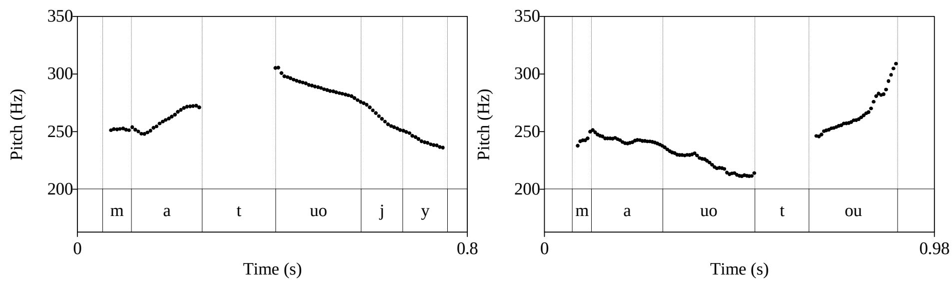
This means that the two structures tested in the current experiment not only differ along the dimension of prosodic boundary between the target word and the surrounding tonal context (prosodic word vs. prosodic phrase boundary), but also along the dimension of prosodic headedness. While the adverb-verb structure maps onto two prosodic phrases, each of which gets its own prosodic head, in the verb-object structure only the object is promoted to phrasal head status, as illustrated in Figure 5.1 (prosodic headedness marked by asterisk).

Morphosyntax:	ADV ${}_{VP}$(V)	${}_{VP}$(V NP)
	↓	↓
Prosodic headedness:	* *	*
Prosodic structure:	φ(ADV) φ(V)	φ(ω(V) ω(NP))

Figure 5.1: Schematic illustration of the assumed syntax-to-prosody mapping and the resulting difference in prosodic headedness.

In order to test the influence of prosodic boundaries and prosodic headedness on tonal coarticulation independently, the position of the target monosyllable was varied within the above structural configurations (target words shown in bold). For the adverb-verb structures, tonal coarticulation will be tested for monosyllabic target words that function as adverbs ((**ADV**)(V-V)) and as verbs ((ADV-ADV)(**V**)), and likewise for the verb-object structures, it will be tested both for monosyllabic target words that function as verbs (**V**(O-O)) and as objects ((V-V)**O**). Note that in all cases, the constituent order within the phrases remains unchanged.

In order to keep variations in sentence lengths and possible syntactic parsings as small as possible, the size of the stimuli was limited to three syllables. All trisyllabic stimuli consisted of a combination of a monosyllabic and a disyllabic lexical target word, whereby the monosyllabic target word could appear on either leftmost or rightmost within the stimulus phrase. In Wenzhou, disyllabic lexical words regularly undergo phonological tone change, whereby the tonal target of both syllables changes from the lexical tone to a specific tonal contour (disyllabic tone sandhi, see Chapter 2 of this thesis for a detailed description).



(1) a. <u>Compatible context</u>
(V (N- N))
ma *tuɔ* *jø*
24 (42. 31)
buy dumpling.soup
'to buy dumpling soup'

b. <u>Conflicting context</u>
(V (N- N)
ma *ʎuo* *tou*
24 (22. 35)
buy soy.beans
'to buy soy beans'

Figure 5.2: Example realization of the same experimental target syllable (initial) by the same speaker in compatible (left) and conflicting (right) context.

In order to test the influence of the context on tonal realization, the target words where elicited both in conflicting and in compatible contexts (see Xu 1994). Figure 5.2 illustrates an example, in which the same target word (*ma* with a rising tone) was followed by two different disyllabic context words. In the left example, the rising target word is followed by a high falling tone on the first syllable of the disyllabic context word, which represents a compatible context. In the right example, the rising target word is followed by a low level tone on the first syllable of the disyllabic context word, which represents a conflicting context.

Including both conflicting and compatible tonal contexts in the experiment serves to test the general prediction that the differences in tonal realization are connected to the adjacent tonal context. The expectation is that, in general, tones should be realized with steeper tonal contours in compatible context, because the adjacent tonal targets can be reached by the speakers without adjusting the tone realization on the target word. In comparison, tonal realization in conflicting contexts should induce adjustment in the implementation of the tones. Therefore, a comparison between the implementation in conflicting and compatible context can give a first indication of tonal coarticulation.

Furthermore, testing trisyllabic phrases with the two morphosyntactic structures outlined above, different predictions emerge with respect to prosodic boundary vs. prosodic head effects. For the first experiment, three different outcomes are conceivable.

(i) No tonal coarticulation difference between the two prosodic structures

Leftmost target words	Rightmost target words
(**ADV**)(VV) = ((**V**)(O-O))	(ADV-ADV)(**V**) = ((V-V)(**O**)

Possible interpretations:

- Tonal coarticulation in Wenzhou Chinese is not dependent on prosodic structure
- Or: Tonal coarticulation in Wenzhou Chinese is dependent on prosodic structure, but the two morphosyntactic structures that were tested map onto identical prosodic structures

(ii) More tonal coarticulation in verb-object structures than in adverb-verb structures on both sides

Leftmost target words	Rightmost target words
(**ADV**)(VV) < ((**V**)(O-O))	(ADV-ADV)(**V**) < ((V-V)(**O**)

Possible interpretations:

- Tonal coarticulation in Wenzhou Chinese is dependent on prosodic boundary strength

(iii) More coarticulatory adjustment in verb-object than in adverb-verb structures in the leftmost target words, but no coarticulatory adjustment difference between the two structures in the rightmost target words

Leftmost target words	Rightmost target words
* * * (**ADV**)(VV) < ((**V**)(O-O))	* * * (ADV-ADV)(**V**) = ((V-V)(**O**)

Possible interpretations:

- Coarticulatory adjustment is dependent on prosodic headedness

A comparison between the tonal implementation of the target words will show which of the predictions above accounts best for the experimental results.

5.2 Method

5.2.1 Stimuli

In order to test the hypotheses made in section 5.1.4, trisyllabic adverb-verb and verb-object structures were designed such that the monosyllabic target words appeared either in leftmost or in rightmost position. In order to control for the context-dependency of coarticulation, all structures were tested in both conflicting and in compatible context. For this purpose, the target words were coupled with disyllabic compounds which carried tone sandhi contours that started or ended with a low or high tonal target.

The monosyllabic target words carried either of the four lexical contour tones of Wenzhou Chinese: low rising, high rising, low falling, and high falling tone. The experiment was limited to contour tones under the assumption that the coarticulatory adjustment effects on slopes would be most clearly quantifiable for these tones, on basis of the steepness of their tonal trajectories. Examples are given in (2) and (3) (target word = bold, tones in Chao numbers).

(2) Conflicting context

	Structure	Hanzi	Wenzhou	Translation
a.	[ADV[V-V]	都喜欢	***ʔu42*** *(sz̩42.ɕø31)*	'like everything'
b.	[V[N-N]	喝汤圆	***ha42*** *(tʰuɔ42.jø31)*	'drink dumpling soup'
c.	[[ADV-ADV]V]	赶紧剁	*(kø44.tɕaŋ22)* ***tou42***	'chop hurriedly'
d.	[[V-V]N]	贩卖报	*(fa44.ma22)* ***pɜ42***	'sell the report'

(3) Compatible context

	Structure	Hanzi	Wenzhou	Translation
a.	[ADV[V-V]	必学习	***pi42*** *(ɦu24.zai31)*	'certainly learn'
b.	[V[N-N]	剁猪肉	***tou42*** *(tsɛi35.ɲou31)*	'chop pork'
c.	[[ADV-ADV]V]	干脆喝	*(kø22.tsʰai33)* ***hɜ42***	'simply drink'
d.	[[V-V]N]	接收布	*(hu22.lou33)* ***pɜ42***	'receive a report'

For each of the 32 combinations of factor levels (two prosodic structures*two contexts*two positions for the target word*four lexical tones), two lexically different examples were recorded, which brings the total stimulus count to 64 tokens per speaker. One example was later excluded, because the speakers realized it with a different tonal target than expected, which leaves 63 tokens per speaker for analysis. In addition to the 64 target tokens, the speakers read 80 trisyllabic filler tokens with varying morphosyntactic composition, so that every speaker saw 144 phrases per recording round.

All examples in the experiment were checked by a native speaker of Wenzhou Chinese and three more native Chinese speakers with different dialectal backgrounds for naturalness and grammaticality. The Wenzhou speaker was in the same age group as the participants of the experiment, but did not take part in the recordings himself. Furthermore, the selected phrases were screened for their segmental composition to minimize difficulties in the later analysis process. A full list of stimuli can be found in appendix 5.1 at the end of this thesis.

5.2.2 Speakers

A total of 19 speakers (five males, 14 females) were recorded for this experiment. They all were born and raised in the inner-city Lucheng district of Wenzhou, and were of similar age (mean age = 23.7, SD = 3.0). None of them reported to have lived outside of Wenzhou for a significant amount of time within the last 5 years, and all of them considered themselves fluent speakers of the Wenzhou dialect. All of them were also fluent in Standard Chinese, but had no difficulty reading out aloud Chinese characters in their dialect. None reported any hearing or speech impediments.

Due to a technical error which sometimes cut off the recording of the stimulus before it was completely uttered, the recordings of six speakers had to be excluded because they showed too many gaps per condition. Data of the remaining 13 speakers was analyzed (four male, nine female; mean age = 23.0, SD = 2.8). Four of these speakers read the list of stimuli once, and the other nine speakers read the list twice.

5.2.3 Experimental procedure

Speakers were recorded in individual sessions in a sound-proofed recording studio in the TV and radio station in Wenzhou, and received a small payment for their participation. Each speaker was seated in front of a 13" monitor and given a Sennheiser pc130 headset. The experimenter ensured that the microphone of the headset was placed approximately 3 cm from the corner of the mouth of the subject. Via an external digitizer (UA-G1), the sound was recorded directly on the laptop (Acer Aspire 1810TZ) on which the stimuli were displayed to the speaker.

The speakers were first informed about the recording procedure. They were told that they were supposed to read out phrases and sentences presented on the screen using Wenzhou dialect, in a natural and clear fashion. If they were unsure how to pronounce a word or phrase, they could skip to the next item, and if they felt they had made a mistake, they could go back and repeat the recording of the previous item. They were told that they could interrupt or abort the recording at any point.

Before the actual recording, all speakers completed a practice series with 8 trisyllabic phrases that were not part of the actual experiment. This was done in order to familiarize the speakers with the self-managed recording procedure, during which they had to press a button to initiate the recording of sound.[18] After completing the practice items, the speakers were asked to indicate whether they understood the recording procedure and were ready to start the actual experiment. Upon confirmation, the experimenter started the actual experiment.

5.2.4 Data analysis

Before data analysis, all recordings were screened for tonal correctness. If a speaker produced a tonal contour on the disyllabic compound that was different from the expected realization in a way that the context was no longer conflicting/compatible, the token was excluded from further analysis. Likewise, all recordings that were incomplete or produced hesitantly with an audible pause within the phrase were excluded. If a speakers who read the list twice made a mistake on one of the recordings, this token was excluded from the analysis. If both recordings were correct, the average of the two recordings was computed and used for the analysis.

Since the target phrases were recorded in isolation, some of the recordings showed utterance-final breathiness and/or creakiness, which obscured parts of the F_0 contour on the final syllable. This concerned a total of 108 recordings, which were marked and subsequently excluded from the F_0 measurements and analysis. A total of 962 target tokens was retained for analysis.

Within the rhyme of the target syllable, F_0 values were sampled at 20 equidistant intervals with the help of the automated F_0 tracking algorithm in PRAAT (Boersma & Weenink 2001). Before data extraction, all F_0 contours were checked for tracking errors such as octave jumps, and these errors were manually corrected (26 instances). Additionally, the contours were smoothed before extraction, using the smoothing function at a 10 Hz bandwidth in PRAAT. A script performed the automated extraction of the duration information and F_0 measurements.[19]

5.3 Results

As described in section 5.1.4, three factors were expected to influence the amount of tonal coarticulation between the tones on the monosyllabic word and

[18] The script used for presenting and recording the stimuli was written by Jos Pacilly, and slightly modified by the author.
[19] The script used for segmenting and measuring the files was written by Jos Pacilly.

the disyllabic compound: (i) the position of the target word (leftmost/rightmost), (ii) the type of context (conflicting/compatible), and (iii) the prosodic structure of the stimulus phrase (ADV-V and V-O). The effects of these three factors will be presented in more detail in the following.

5.3.1 Position

Figure 5.3 illustrates the pitch trajectories in the different prosodic structures and contexts, split by tones and position of target word. In order to be able to average the F_0 values over speakers, all raw F_0 values were converted into semitones and scaled to the individual speaker's pitch range. The speaker's pitch range was set to the averaged values of the turning point in the low rising tone (baseline) and the fall onset in the high falling tone (topline) in rightmost position in compatible context. Figure 5.3 expresses the speaker pitch range on a scale between 0 and 100.

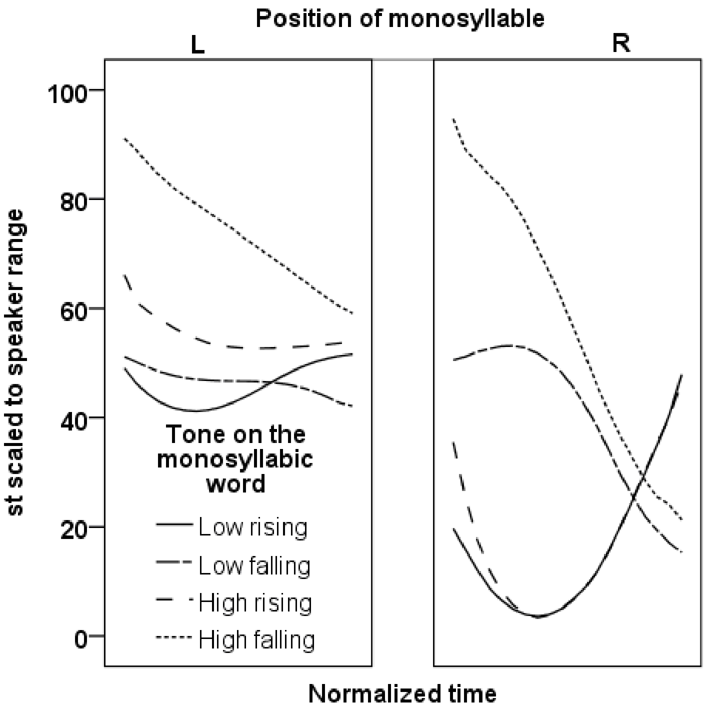


Figure 5.3: Averaged and time-normalized pitch contours of the four investigated tones, broken down by position (L = leftmost, R = rightmost).

As can be seen in Figure 5.3, the realizations of the tonal contours in leftmost and in rightmost position the trisyllabic phrases differ in the amount of the speakers' pitch ranges that they cover. More specifically, the rightmost tones cover a much wider portion of the speakers' pitch ranges than the leftmost tones, when averaged over the different contexts and prosodic structures.

5.3.2 Context

As discussed in section 5.1.2, tonal coarticulation has been found to be context-dependent. Tones coarticulate much more in conflicting contexts than in compatible contexts. In order to illustrate the effect of context, the following graphs displays the four lexical tones in conflicting (top) and compatible (bottom) context in both positions, averaged over the two prosodic structures.

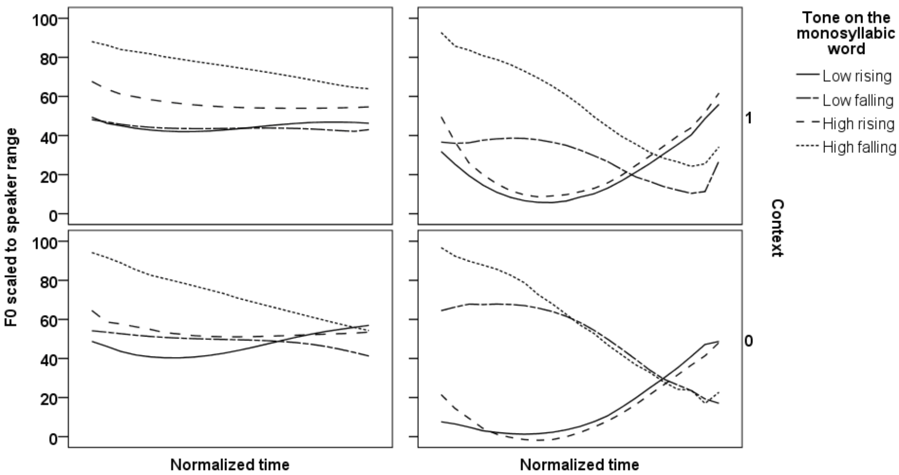


Figure 5.4: Frequency and time-normalized averaged pitch contours of the four investigated tones in leftmost/rightmost position, broken down across contexts.

As can be seen in Figure 5.4, the effect of coarticulation causes minor differences across the entire tonal trajectories, but is most clearly visible in the portion of the tone that is immediately adjacent to the context, i.e. the right edge for the leftmost tones and the left edge for the rightmost tones. In both positions, the context has consequences for the steepness of the tonal realizations. The two rising tones are largely deprived of their final rise in leftmost position in conflicting context compared to compatible context, making their tonal trajectories appear almost flat. In rightmost position, the rising tones start higher in conflicting compared to compatible context, which also flattens the overall slope. The falling tones both fall less steeply in leftmost position in conflicting compared to compatible context, and they start lower in rightmost position.

5.3.3 Prosodic structure

The previous graphs have indicated that the effect of coarticulation in trisyllabic phrases in Wenzhou is mainly local, and subsides with increasing distance from that tonal context. The investigation in this section will therefore concentrate on the respective halves of the tonal trajectories that are adjacent to the tonal context (i.e. the left half for rightmost targets, and the right half for leftmost targets).

In order to make the differences in the tonal trajectories more clearly visible, the following graphs will display linear approximations to the slopes of the respective halves of the tonal trajectories, rather than the trajectories themselves. In this way, it can be seen more easily whether the two structures induce a difference in tonal implementation. The slope values were computed by dividing the difference in F_0 between the first and last measurement of the trajectory part by half of the duration of the target syllable rhyme. For targets in which either the first or the last measurement were missing (for example because of creakiness or breathiness in the signal), the second or the last-but-one measurement were used to compute the slope value (with accordingly adjusted duration values). Targets which had more measurements missing were not included in the slope analysis.

5.3.3.1 Leftmost targets

In Figure 5.5, the slope values are displayed for the tones in leftmost position ((**ADV**)(V-V) with full lines and ((**V**)(O-O)) with dashed lines). Since coarticulation was mainly found in conflicting contexts, only the conflicting context values are shown.

It can be seen that all four tones display the same tendency: the tonal direction is more preserved in the adverb-verb structures (solid line) than in the verb-object structures (dashed line). For the two rising tones, the tonal trajectories in verb-object structures are almost horizontal (low rising) or even slightly falling (high rising), while the trajectories in the adverb-verb structures have the rising tonal direction preserved. For the falling tones, the falling trajectories are more steeply falling in the adverb-verb structures, and flatter or even rising in the verb-object structures.

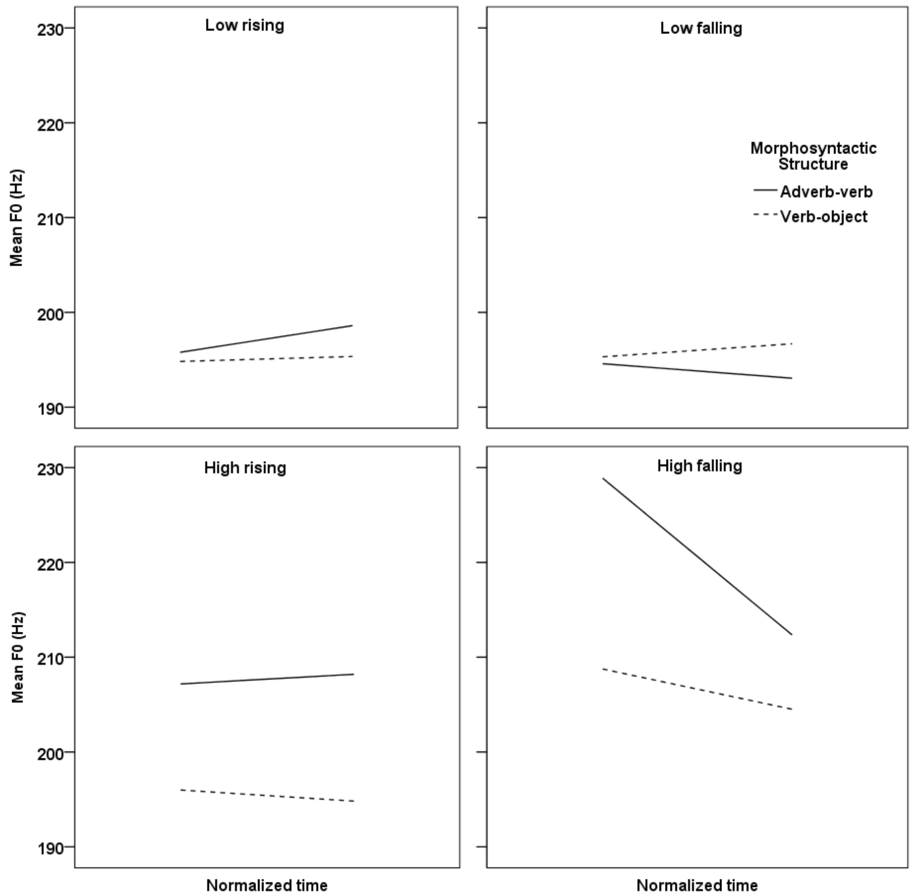


Figure 5.5: Linear representations of the tonal slopes (rightmost half) in leftmost position in conflicting context, broken down by tone on the target syllable and structure (dashed lines = verb-object, solid lines = adverb-verb).

In sum, it can be said that across all four tones, the same tendency is visible: the tonal trajectories are steeper and true to the original tonal direction in adverb-verb structures, but flattened or even slightly reversed in direction for the verb-object structures. In order to statistically test the validity of this observation, the slopes were first "adjusted" (in the sense of Xu 1994) and then pooled over all four tones. For the adjustment, the slope values for the falling tones were multiplied by (-1), so that for all four tones, a positive slope value would represent a preservation of tonal direction, and a negative slope value would represent a reversal of tonal direction.

Figure 5.6 graphically represents the pooled slope values, split by context (left = compatible context, right = conflicting context) and prosodic structure (light bars = ((**ADV**)(V-V), shaded bars = ((**V**)(O-O))) for the tones on target words in leftmost position.

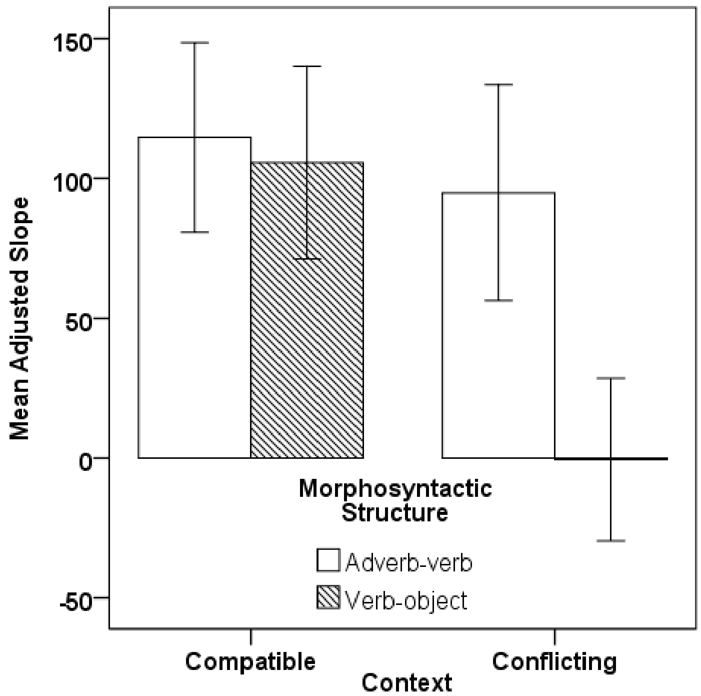


Figure 5.6: Means and error bars (±2 SE) for the adjusted slope values (right half) in leftmost position. Values broken down by context (left = compatible, right = conflicting) and structures (shaded = verb-object, white = adverb-verb).

In order to test whether the differences between the two contexts and the two prosodic structures is statistically significant, a by-subjects Repeated Measures (RM) ANOVA was conducted, with context and prosodic structures as the two main factors, and the adjusted slope values as the dependent variable.

Both the factor PROSODIC STRUCTURE [$F(1,12) = 7.36$, $p = 0.019$] and the factor CONTEXT [$F(1,12) = 11.59$, $p = 0.005$], as well as their interaction

[PROSODIC STRUCTURE*CONTEXT: $F(1,12) = 22.91$, $p < 0.001$] exert a significant influence on the tonal trajectories of the leftmost targets. Because a significant interaction was found, the two contexts were investigated separately for a difference between the prosodic structures. It turned out that the difference between the prosodic structures is highly significant in conflicting context [$F(1,12) = 34.92$, $p < 0.001$], but not significant in compatible context [$F(1,12) = 0.31$, $p = 0.59$, ns]. This confirms the impression that tones in leftmost position in verb-object structures are influenced by coarticulation to a different (greater) extent than those in adverb-verb structures.

5.3.3.2 Rightmost targets

For the tones on monosyllabic target words in rightmost position, the same slope computations were performed as for the tones in leftmost position. Figure 5.7 shows the first half of the tonal slopes in conflicting context, split between the two prosodic structures ((ADV-ADV)(**V**) = solid lines, ((V-V)(**O**)) = dashed lines) for the four lexical tones.

Figure 5.7 shows that in rightmost position, there is no systematic difference in the effect of coarticulation across all four tones. For the two rising tones (which in the first half of their trajectories have a falling slope in conflicting condition, compare Figure 5.4), the realization in adverb-verb structures seems to have reversed the tonal contour to a greater extent than the realization in verb-object structure for the low falling tone, but there is barely any difference for the high rising tones.

For the two falling tones, it seems that the tonal trajectories on the adverb-verb structures are more true to the original falling trajectories than the realization in verb-object structures. In sum, two of the tones show slightly more coarticulation for the verbs in adverb-verb structures, one of the tones shows slightly more coarticulation for the object in verb-object structure, and one of the tones shows very little difference between the two structures.

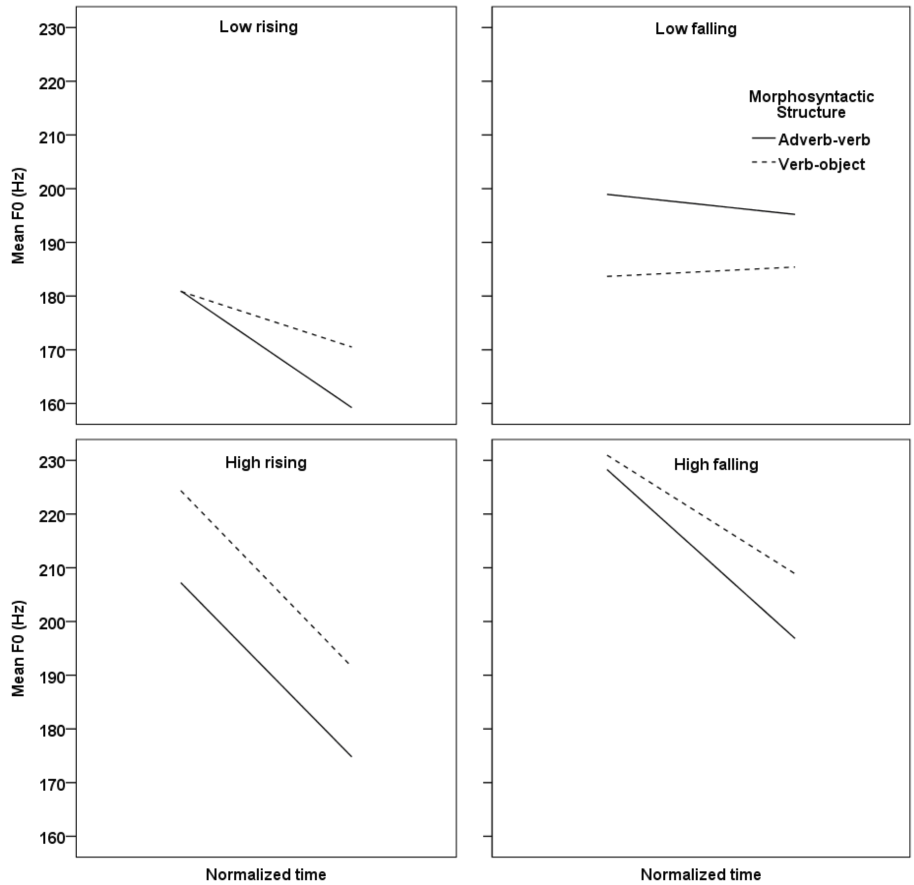


Figure 5.7: Linear representations of the tonal slopes (leftmost half) in rightmost position in conflicting context, broken down by tone on the target syllable and structure (dashed lines = verb-object, solid lines = adverb-verb).

After adjusting the slope values for the rightmost targets and pooling over all four tones in the same manner as was done for the leftmost tones, it was found that the coarticulation effects cancel each other out, because they do not systematically point in the same direction as the effects on the leftmost tones do. In other words, when pooling over all four tones, the trajectories are overall flatter in conflicting than in compatible context, but there is no big difference between the prosodic structures on either side. Figure 5.8 illustrates this.

A by-subjects RM ANOVA confirms that there is a highly significant main effect of CONTEXT [$F(1,12) = 23.98$, $p < 0.001$], but no significant effect of PROSODIC STRUCTURE [$F(1,12) = 0.02$, $p = 0.907$, ns], or an interaction between the two factors [CONTEXT*PROSODIC STRUCTURE: $F(1,12) = 1.27$, $p = 0.281$, ns]. This shows that in rightmost position, the two prosodic structures are not systematically different in slope, which speaks for a coarticulation effect that is equally strong across the two prosodic structures.

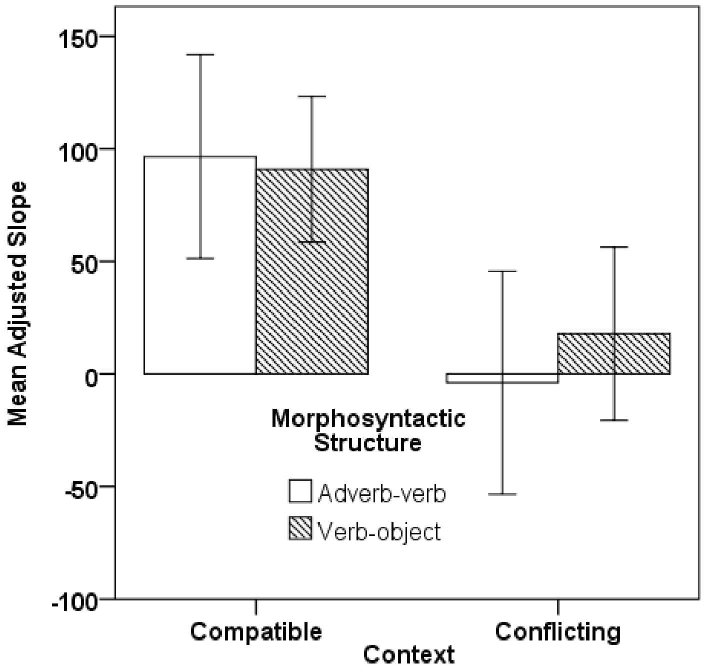


Figure 5.8: Means and error bars (±2 SE) for the adjusted slope values (left half) in rightmost position. Values broken down by context (left = compatible, right = conflicting) and structures (shaded = verb-object, white = adverb-verb).

5.3.4 Duration

In the above statistics, the factor duration is already included, in so far as the slope measurements have been calculated using the duration data for the individual conditions. However, it is also interesting to look at the duration data itself, since it can give insights into the type of tonal modification that occurs in the different contextual and prosodic conditions. Specifically, it can be investigated whether the relative steepness/slope of the tonal contours is directly covariant with lengthening, or whether the slopes are modified independent of the durational modification. Figure 5.9 represents the duration results graphically.

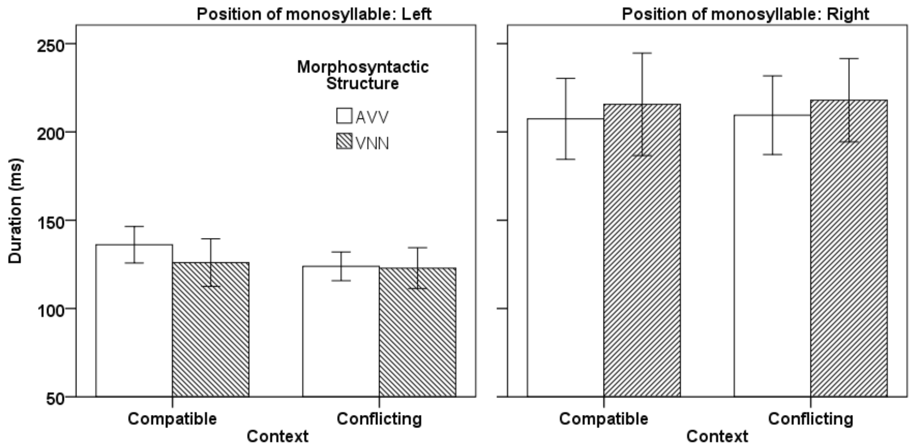


Figure 5.9: Means and error bars (±2 SE) for the duration values in leftmost (left graph) and rightmost (right graph) position. Values broken down by context (left = compatible, right = conflicting) and structures (shaded = verb-object, white = adverb-verb).

The duration of the target syllables was compared in a by-subjects RM ANOVA, with POSITION, CONTEXT, and PROSODIC STRUCTURE as main factors. The results show that of the three factors, only POSITION (monosyllabic target syllable leftmost/rightmost) had a significant effect on the duration of the target syllable [$F(1,12) = 35.17$, $p < 0.001$], whereas neither CONTEXT (conflicting/ compatible) [$F(1,12) = 1.01$, $p = 0.334$, ns] nor PROSODIC STRUCTURE (ADV-V vs. V-O) [$F(1,12) = 0.09$, $p = 0.765$, ns] exerted a significant effect.

The strong effect of POSITION meant that also the interaction POSITION*CONTEXT reached significance [$F(1,12) = 5.04$, $p = 0.044$], and the interaction POSITION*PROSODIC STRUCTURE approached significance [$F(1,12) = 3.94$, $p = 0.07$]. Neither the interaction CONTEXT*PROSODIC STRUCTURE [$F(1,12) = 3.04$, $p = 0.107$, ns] nor the three-way interaction POSITION*CONTEXT* PROSODIC STRUCTURE [$F(1,12) = 0.77$, $p = 0.399$, ns] reached significance.

Because there was a significant interaction POSITION*CONTEXT, it is worthwhile to examine the two positions separately for effects of the two other factors. For the targets in rightmost position within the phrase, neither the factor CONTEXT [$F(1,12) = 0.27$, $p = 0.613$, ns] nor the factor PROSODIC STRUCTURE [$F(1,12) = 1.23$, $p = 0.29$, ns] nor their interaction [CONTEXT*PROSODIC STRUCTURE $F(1,12) = 0.002$, $p = 0.961$, ns] reached significance. For the targets in leftmost position within the phrase, only the factor CONTEXT [$F(1,12) = 8.6$, $p = 0.013$] reached significance, while neither the factor PROSODIC STRUCTURE [$F(1, 12,) = 3.59$, $p = 0.082$, ns] nor the interaction between the two factors [CONTEXT*PROSODIC STRUCTURE $F(1,12) = 2.71$, $p = 0.126$, ns] reached significance.

5.4 Discussion: Tonal coarticulation and prosodic structure

The foregoing experiment tested whether prosodic structure influences tonal coarticulation in Wenzhou Chinese, and if so, whether prosodic boundaries or prosodic headedness play a larger role. In order to quantify the extent of the influence of neighboring tones onto the tonal trajectories of the target words, their realizations were compared between compatible context, in which no adjustment of tonal trajectories is expected, and conflicting context, in which the tonal trajectories should show influence from the adjacent tonal context.

As has been shown in the previous section, the tonal slopes were indeed significantly flatter in verb-object structures in leftmost position in conflicting context, compared with compatible contexts. In contrast, the adverb-verb structures showed similar tonal slopes in both conflicting and compatible context in leftmost position. This speaks for a greater autonomy of the targets in adverb-verb structures in their tonal realization, whereas the targets in verb-object structures were subject to significant influence from the neighboring tonal context.

Up to this point, the results are compatible with both hypothesis (ii) and hypothesis (iii) from section 5.1.4. In order to decide whether the observed effect is brought about by the prosodic boundary strength between the target word and its tonal context, or between the prosodic headedness distribution within the prosodic structure, the results in leftmost position have to be compared with the results in rightmost position. As discussed in the previous section, in rightmost position, the target monosyllables in both structures are implemented in a similar way in both compatible and in conflicting context. This leads to the interpretation that it is not prosodic boundary strength which determines the amount of coarticulation in the target words of this experiment, since this boundary strength is identical for target words in leftmost and rightmost position.

Rather, the results lead to conclude that the prosodic headedness of the target words is responsible for the differences in magnitude of coarticulation between the two prosodic structures. As laid out in section 5.1.4, in terms of prosodic headedness, both structures display a prosodically strong position on the right side, but only the adverb-verb structures also have a prosodically strong position on the left side. Therefore, finding a difference in tonal coarticulation in leftmost but not in rightmost position in the phrases strongly suggests that the effect is dependent on prosodic headedness, rather than on prosodic boundary strength.

Taking into account the durational data, it is interesting to note that the difference in tonal slopes in the leftmost tones did not correlate with a significant duration difference between the two structures. This indicates that the tonal implementation effect is not just due to e.g. greater pre-boundary lengthening in prosodic phrase compared to prosodic word context, where the longer duration of the syllable would allow a fuller implementation of the tone, compared to the truncated tonal realization in non-lengthened positions. Rather, the duration data shows that, while there was a slight, non-significant lengthening in the adverb-verb structures compared to the verb-object structures (see Figure 5.9), this difference alone cannot be responsible for the significant slope effect.

5.5 Tonal coarticulation under narrow focus

If tonal coarticulation in trisyllabic phrases is indeed influenced by prosodic structure and in particular by prosodic headedness, an interesting follow-up question to ask is: what happens to this effect under narrow focus? As laid out in section 5.1.3, research for other Chinese dialects has suggested that focus exerts a strengthening effect on tonal realization, which leads to magnified implementation of tonal contours and to greater resistance of focused tones to influence from neighboring tonal contours.

For Wenzhou, it has been suggested that the effect of focus is therefore best accounted for as prosodic prominence effect. Under focus, the syntactically derived prosodic structure is overridden, and the focused constituent becomes the prosodic head of the entire intonation phrase (Chen 2000: 511). Similar accounts have been proposed for other languages, in order to explain the effects of focus in an indirect way via its proposed influence on prosodic prominence (Büring 2010; Gussenhoven 1992; Truckenbrodt 1999). The idea is to relate focus to maximal prominence in its prosodic domain, and in cases where focus requirements collide with the syntactically derived prosodic structure, the prosodic structure is changed to fulfill the focus requirements.

For the results presented in the foregoing section, such an interpretation of focus would lead to the assumption that under focus, the effects of prosodic structure should be neutralized. More specifically, if the focused constituent is assumed to be the prosodic head, regardless of which constituent should be the prosodic head according to the morphosyntactic structure, it can be expected that any effect that is brought about by the morphosyntactic structure should disappear in the presence of focus.

5.5.1 Stimuli, speakers, experimental procedure, data analysis

In order to test this hypothesis, a second experiment was conducted, using the same materials as the first experiment. In the second experiment, however, the trisyllabic phrases were presented in the context of an alternative question, which induced narrow contrastive focus on the monosyllabic target word, as exemplified in (4).

(4)	Q:	到	中国	或者	走	中国？
		tɜ42	*tɕoŋ35.kai42*	*va22.tsɜ33*	*tsau35*	*tɕoŋ35.kai42*
		reach	China	or	walk-to	China
		‘‘To reach China’ or ‘to walk towards China’?’				
	A:	到	中国			
		tɜ42	*tɕoŋ35.kai42*			
		reach	China			
		‘To reach China.’				

Similarly to the first experiment, the stimulus phrase in the answer constitutes its own utterance, and therefore its realization should, for all intents and purposes, be comparable to the realization of the stimuli in the first experiment, apart from the additional influence of contrastive focus on the target monosyllable.[20] The

[20] In the following, the results of the second experiment reported in section 5.5, which deals with narrow contrastive focus, will be compared to the results of the first experiment discussed in sections 5.2-5.4, which presented the items without a specific information-structural context. It is a matter of debate whether a presentation out of the

same 19 speakers as in the first experiment took part in the second experiment, but in order to further increase the comparability of the two experiments, only the data from the same 13 speakers as in the first experiment was analyzed and will be reported. Each of the speakers recorded the trisyllabic phrases under focus after completing the recording of the trisyllabic phrases in isolation, so the speakers were already familiar with the materials.

The recording procedure for the trisyllabic phrases under focus was identical to the procedure used to obtain the recordings in isolation (see section 5.2.3). After excluding incomplete and erroneous renditions, 1265 target tokens remained for analysis. These were segmented and labeled in the same way as the target tokens from the first experiment, and subjected to the same F_0 and duration data extraction procedure (see section 5.2.4). Ninety-three tokens with utterance-final breathiness/creakiness were excluded from the data measurements, and the automatic F_0 measurements were hand-corrected for tracking errors such as octave jumps in eight cases.

5.5.2 Results

In order to facilitate the comparison to the results of the first experiment, the graphs representing the realizations of the monosyllabic target words under focus will be set up in the exact same way as the graphs in section 5.3. All data was transformed in the same way (for example, for the following to graphs, the measurements have been scaled to the individual speaker's pitch ranges before being averaged across speakers).

5.5.2.1 Position

Figure 5.10 represents the tonal trajectories of the monosyllabic words under contrastive focus, averaged across repetitions, speakers, contexts, and prosodic structures, and split by tones and position within the trisyllabic phrase. As can be seen, the general trend from the unfocused data can also be confirmed for the focused data, namely that the tonal trajectories are more pronounced and cover a

---

blue evokes so-called "broad" or neutral focus that is comparable to an "unfocused" condition, or whether it just puts the entire utterance in focus, as argued for example in Lambrecht 1994. Since the current experiment is concerned with narrow contrastive focus on one of the three syllables in the trisyllabic phrases, the isolated context will be referred to as "non-focused" or "unfocused" for ease of comparison. However, this should not be taken to imply the complete absence of any (broad) focus in the isolated examples in the first experiment.

wider area of the speakers' pitch ranges in rightmost compared to leftmost position.

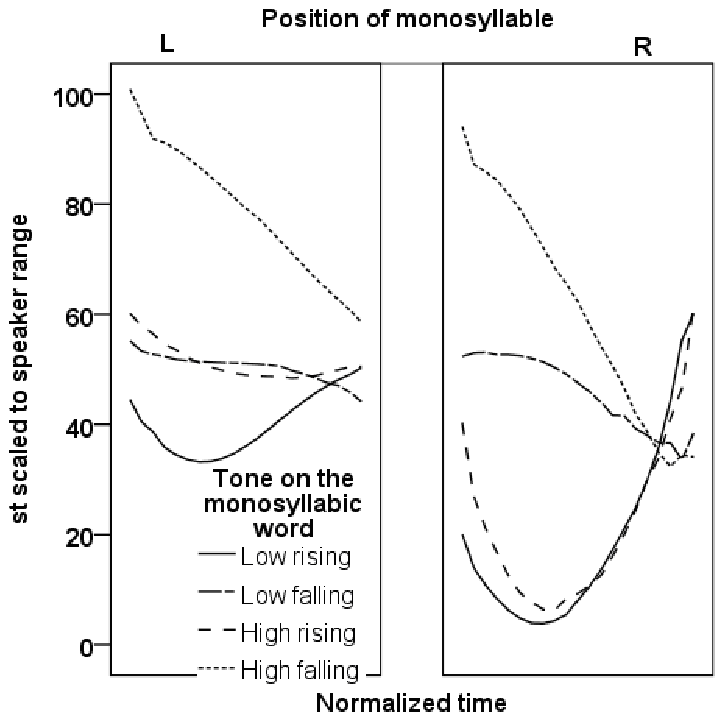


Figure 5.10: Averaged and time-normalized F_0 contours of the four lexical tones under narrow focus, broken down by position (L = leftmost, R = rightmost).

5.5.2.2 Context

Because of the above-mentioned split between tonal realizations in leftmost vs. rightmost position, the following analysis will again present the results for the both positions separately. The graph on the left side of Figure 5.11 shows the tonal trajectories of the tones in leftmost position, split by conflicting (top) and compatible (bottom) context, and the graph on the right shows the same for the tonal trajectories in rightmost position.

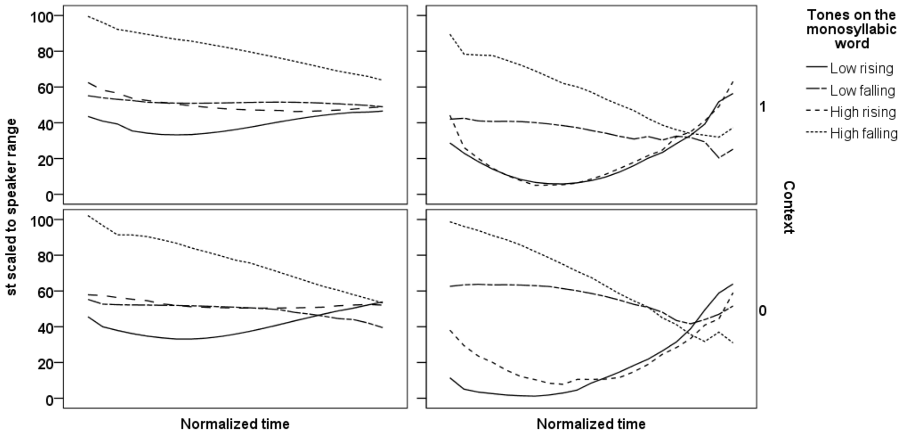


Figure 5.11: Averaged and time-normalized F_0 contours of the four investigated tones in initial (left) and final (right) position, broken down across contexts.

Again, the differences between the tonal realizations in the two contexts in focused condition mirror the effects found in the unfocused condition. The difference between the trajectories in the two contexts again pertains mainly the half of the contours that is immediately adjacent to the context tones, namely the right half for the leftmost tones and the left half for the rightmost tones. The effects for the specific tones are also comparable: rising tones are largely deprived of their final rising portion in leftmost position, and start higher in rightmost position. Falling tones are also flattened in leftmost position in conflicting context, and start from a less high starting point in rightmost position in conflicting context compared to compatible context.

5.5.2.3 Prosodic structure

The same calculations were performed on the data as in the first experiment to visualize the influence of the two prosodic structures onto tonal coarticulation. Specifically, the slope values were computed for the right half of the leftmost tones, and for the left half of the rightmost tones, to zoom into the area that is most affected by tonal coarticulation.

5.5.2.3.1 Leftmost targets

As Figure 5.12 shows, the difference for the leftmost trajectories between the two prosodic structures that could be seen in the unfocused data cannot be found back in the data under focus. Apart from a slight difference in the low rising structure, the two slopes run more or less parallel. This indicates that there is no significant difference in slope between the two prosodic structures for leftmost targets.

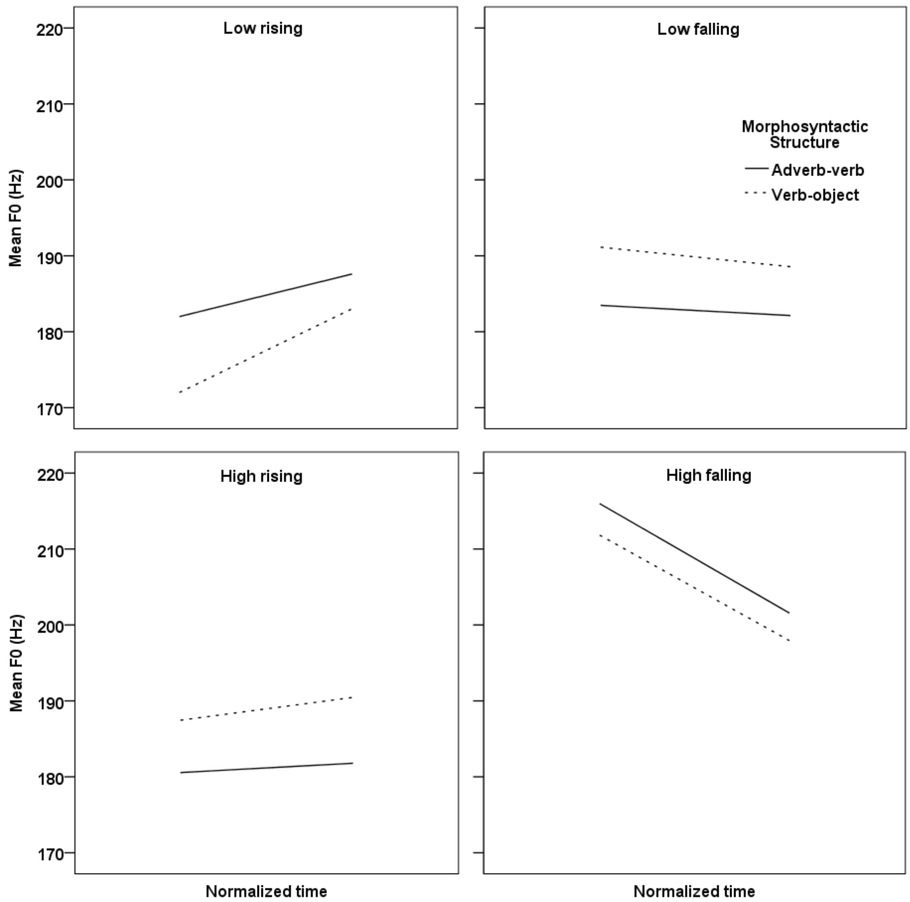


Figure 5.12: Linear representations of the tonal slopes (rightmost half) in leftmost position in conflicting context under focus, broken down by tone on the target syllable and structure (dashed lines = verb-object, solid lines = adverb-verb).

For statistical comparison, the slope values were adjusted and averaged across the four tones, as shown in Figure 5.13. A by-subjects RM ANOVA on the adjusted slope values shows that the only significant difference is brought about by the factor CONTEXT [$F(1,12) = 14.87$, p = 0.002], while both PROSODIC STRUCTURE [$F(1,12) = 3.6$, p = 0.82, ns] and the intercept between the two factors [CONTEXT*PROSODIC STRUCTURE: $F(1,12) = 0.92$, p = 0.356, ns] turn out to not be significantly different from each other. This confirms what the inspection of the graphical slopes already led to assume: the effect of prosodic structure upon the magnitude of coarticulatory adjustment in the tonal realization on initial target words disappears under contrastive focus.

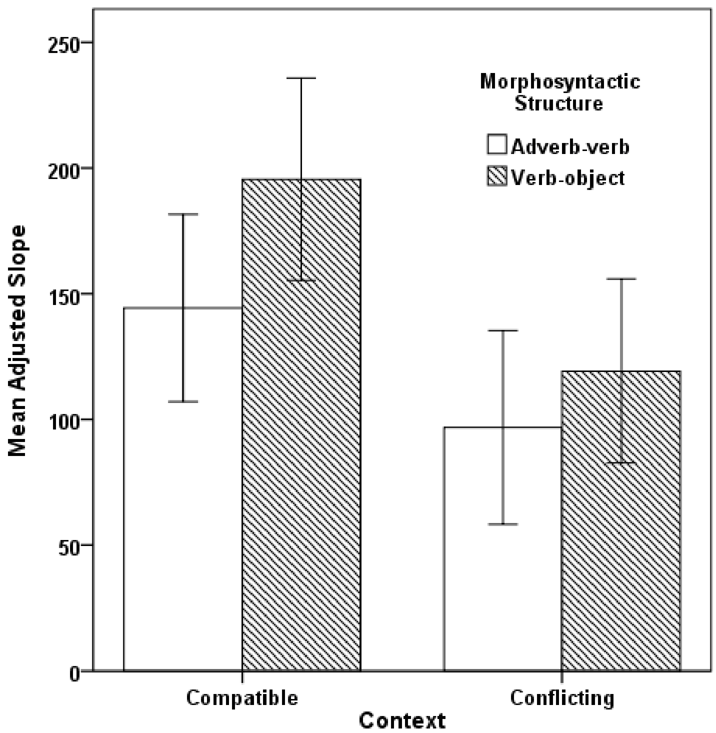


Figure 5.13: Means and error bars (±2 SE) for the adjusted slope values in leftmost position under contrastive focus. Values per tone broken down by context (left = compatible, right = conflicting) and structures (shaded = verb-object, white = adverb-verb).

At the same time, the Figure 5.13 shows an important difference between the leftmost targets in the first and in the second experiment. In the first experiment, the slopes in conflicting contexts were almost flat for the verb-object structures. In contrast, in the second experiment, the tonal trajectories are relatively steep in both conflicting and compatible context. There is still a difference between the two contexts, and as in experiment 1, it is a significant one, in that the tones in compatible context are steeper across the board than the tones in conflicting context. However, under focus, the targets are strengthened in their tonal realization in both contexts, in that the slopes even in compatible context are relatively steeper under focus than in the first experiment (compare Figures 5.6 and 5.13).

5.5.2.3.2 Rightmost targets

Figure 5.14 illustrates the tonal slopes in rightmost position under focus. As in the non-focused condition, the tones on the right side show no consistent influence of prosodic structure onto tonal realization. For the two rising tones, the tones in verb-object structure seem to be less falling and consequently less influenced by the preceding context than the tones in adverb-verb structure. For the falling tones, the picture is reversed. Here, it is the adverb-verb structures that show steeper falling contours and thereby less influence from coarticulation. In sum, the picture that was shown in non-focused condition is repeated, namely no systematic difference in slope between the two prosodic structures.

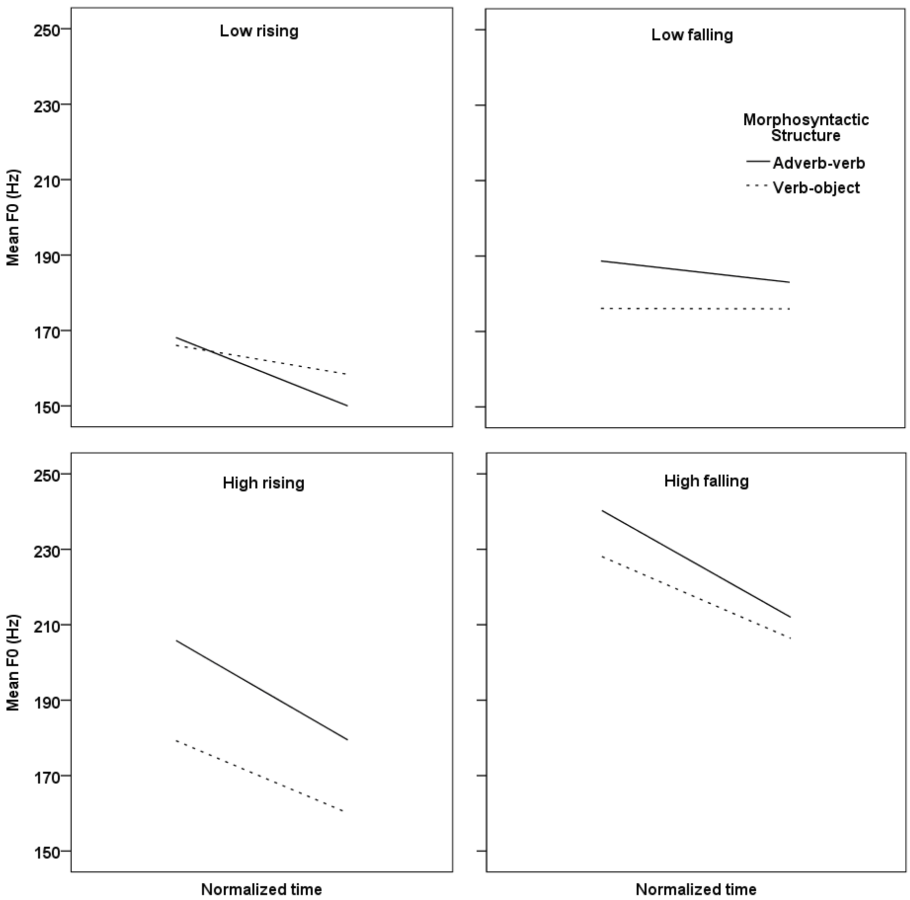


Figure 5.14: Linear representations of the tonal slopes (leftmost half) in rightmost position in conflicting context under focus, broken down by tone on the target syllable and structure (dashed lines = verb-object, solid lines = adverb-verb).

Also for the tones in rightmost position, the slope values were adjusted and then averaged across tones, as shown in Figure 5.15. A by-subjects RM ANOVA shows that neither CONTEXT [$F(1,12) = 3.19$, p = 0.1, ns] nor PROSODIC STRUCTURE [$F(1,12) = 1.17$, p = 0.301, ns] or the interaction between the two factors [CONTEXT*PROSODIC STRUCTURE: $F(1,12) = 0.03$, p = 0.866, ns] induces a significant difference in the slopes of the tonal realizations.

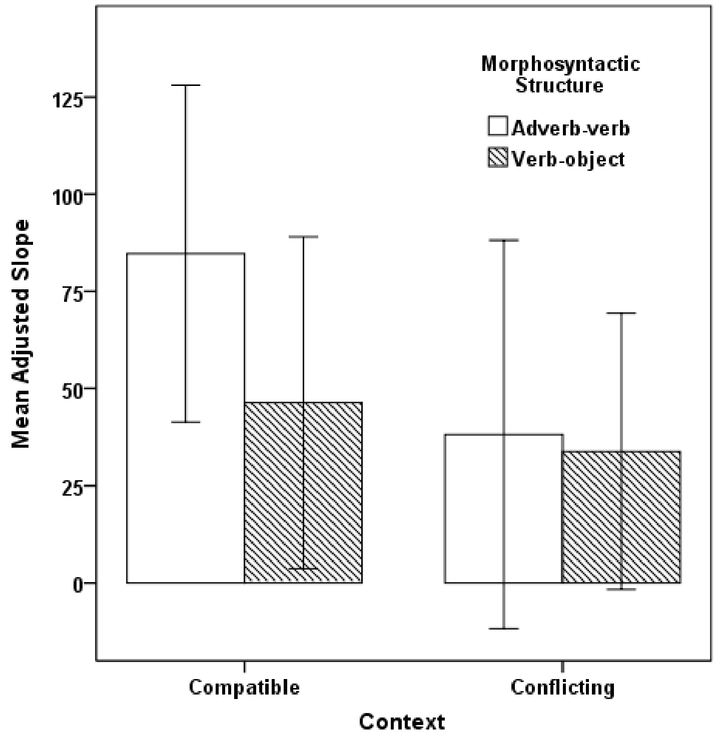


Figure 5.15: Means and error bars (±2 SE) for the adjusted slope values in rightmost position under contrastive focus. Values per tone broken down by context (left = compatible, right = conflicting) and structures (shaded = verb-object, white = adverb-verb).

5.5.2.4 Duration

As in the previous experiment, the duration measurements of the target monosyllables under focus were averaged over speakers, and then tested in a by-subjects RM ANOVA. Figure 5.16 illustrates the duration results under focus graphically.

Similar to the previous experiment, only the factor POSITION (leftmost/rightmost within the phrase) exerts a significant influence on the duration data [$F(1,12) = 135.95$, $p < 0.001$], whereas neither the factor CONTEXT (conflicting/compatible) [$F(1,12) = 0.11$, $p = 0.744$, ns] nor the factor PROSODIC STRUCTURE (ADV-V vs. V-O) [$F(1,12) = 1.65$, $p = 0.223$, ns] exerted a significant difference on their own. Furthermore, a significant two-way interaction CONTEXT*PROSODIC STRUCTURE [$F(1,12) = 17.78$, $p = 0.001$] was found, as well as a significant three-way interaction POSITION*CONTEXT* PROSODIC STRUCTURE [$F(1,12) = 8.93$, $p = 0.011$]. None of the other interactions reach significance [POSITION*PROSODIC STRUCTURE ($F(1,12) = 0.38$, $p = 0.55$, ns), POSITION*CONTEXT ($F(1,12) = 0.34$, $p = 0.569$, ns].

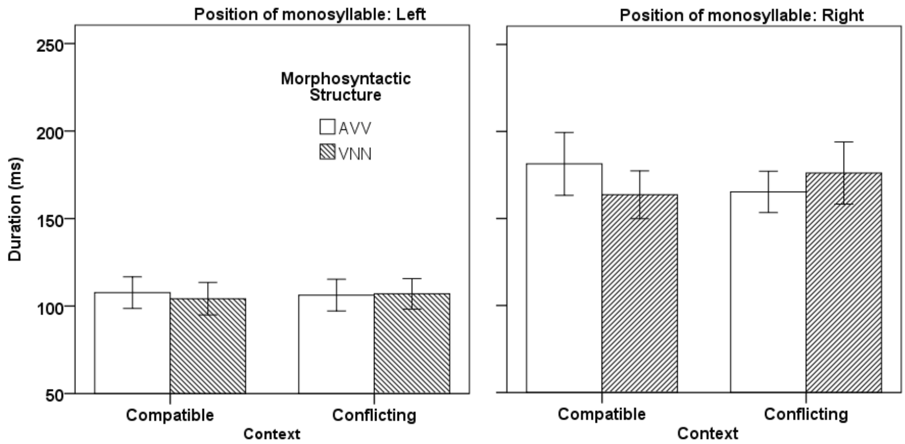


Figure 5.16: Means and error bars (±2 SE) for the duration values in leftmost (left graph) and rightmost (right graph) position under focus. Values broken down by context (left = compatible, right = conflicting) and structures (shaded = verb-object, white = adverb-verb).

A second RM ANOVA compared the duration results from the first experiment (recording in isolation) directly with those of the second experiment (recording under focus), with the added factor FOCUS as a within-speaker variable. First of all, the factor FOCUS was found to induce a significant main effect between the duration measurements of the first experiment and those of the second experiment [$F(1,12) = 55.33$, $p < 0.001$]. Closer inspection of the duration values of the target syllables under focus and in isolation showed however that the values were actually lower in focus condition than in isolation.

This can be explained by the presence of the focus-inducing contrastive sentence in the second experiment, which preceded the target phrases and led to a higher on-average speech rate across speakers. Comparison of the average duration of the entire trisyllabic phrases confirms this assumption: in isolation, the average duration of the trisyllabic phrases is 0.74 seconds (SD: 0.11 seconds), whereas the average duration of the same trisyllabic phrases in the focus condition is 0.6 seconds (SD: 0.09 seconds). At the same time, the ratio between the duration of the target monosyllable in relation to the duration of the entire trisyllabic phrase is similar in both experiments, namely 23% (isolation: 0.17 seconds, SD: 0.05 seconds, focus: 0.14 seconds, SD: 0.04 seconds). Therefore, the significant effect of FOCUS on the duration data can actually be explained by the speech rate, which masks any focus-induced lengthening that might be present on the target syllables in the second experiment.

Furthermore, the factor POSITION (leftmost/rightmost within the phrase) also exerted a significant main effect on the duration data when comparing the first with the second experiment [$F(1,12) = 66.17$, $p < 0.001$]. However, the two-way interaction FOCUS*POSITION turned out to be not significant [$F(1,12) = 2.81$, $p = 0.119$, ns], which shows that the effect of position is the same in the two experiments (namely longer duration of the target syllables in rightmost than in leftmost position). The factor PROSODIC STRUCTURE did not exert a significant difference on the duration measurements when comparing the first to the second experiment [$F(1,12) = 0.006$, $p = 0.94$, ns], nor did the factor CONTEXT exert a significant difference on its own [$F(1,12) = 1.56$, $p = 0.186$, ns].

In terms of interaction, only two factor combinations reached significance. A significant effect of the two-way interaction PROSODIC STRUCTURE*CONTEXT [$F(1,12) = 15.08$, $p = 0.002$] indicates that there is a difference in the duration measurements between the first and the second experiment that depends on prosodic structure in covariance with context. The three-way interaction FOCUS*POSITION*PROSODIC STRUCTURE [$F(1,12) = 5.31$, $p = 0.04$] indicates that the duration measurements differ between the two experiments when taking position and prosodic structure into account in combination. No other interaction reached significance.

5.5.3 Discussion: Tonal coarticulation under focus

The results of the first experiment showed that tonal implementation and its susceptibility to influence from neighboring tones in Wenzhou were influenced by prosodic structure, specifically by prosodic prominence in prosodic head positions. It was shown that tones in leftmost position in the target phrases, in which there is a difference in prosodic headedness between the two prosodic structures, showed a different amount of influence from neighboring tonal targets, with the tones in prosodically weak positions coarticulating more with the tonal context than tones in prosodically strong positions. In contrast, the tones in rightmost position in the target phrases showed no such difference between the two prosodic structures, which was explained with the fact that the rightmost syllables are prosodically strong in both contexts.

The second experiment tested whether this effect of prosodic strength could be influenced by focus, and specifically, whether the effect of prosodic structure would simply be nullified under focus, or whether focus would induce its own strengthening, even on those targets that were already in prosodically strong positions. The results presented in the foregoing sections point in the direction of the latter assumption.

The second experiment found that under focus, the difference in tonal slopes between the two prosodic structures that was found in the first experiment disappeared. Both the targets in rightmost and leftmost position showed similar slopes for the adverb-verb structures and for the verb-object structures. A superficial examination of these findings could lead to the assumption that under focus, the prosodic structure is changed to reflect the focus structure, and that focus takes over the task of assigning prosodic prominence and thereby overrides the prosodic structure that is built on basis of the morphosyntactic structure.

However, closer inspection of the results of the second experiment, and a comparison with those of the first experiment, shows that such an account cannot explain all the findings of the current data. For the targets in leftmost position, it can be seen that the slope values in compatible context are higher for both prosodic structures under focus compared to isolation. If focus were simply a re-allocation of prosodic headedness, it would not be expected that it should make a difference in the implementation of targets that are not in conflict with their tonal environment.

Even more so, a comparison between the tonal slopes in rightmost position in the two experiments also shows a clear effect of focus. In both experiments, there is no difference in prosodic headedness between the two prosodic structures in rightmost position, since both structures have a prosodic prominence on the right side. Nonetheless, there is a significant difference between tones in conflicting and compatible context even in rightmost position, with the tones in both structures being influenced by the preceding tonal context in conflicting position.

Under focus, however, context ceases to affect the implementation of tones in these prosodically strong positions. As Figure 5.15 shows, focus further strengthens the implementation of tones, even if they already are in prosodically prominent positions, so that they are realized with relatively greater autonomy from the conflicting tonal context. Under an account which simply interprets focus as prosodic prominence allocation, this finding would be completely unexpected. If focus simply shifts prosodic headedness, tones that already are in a prosodic head position morphosyntactically would be predicted to be unaffected by focus. In other words, focus would be predicted to strengthen prosodically weak positions, but it would not be expected to further strengthen prosodically strong positions.

This, however, is exactly what the comparison of the findings of the two experiments indicate. Under focus, tonal implementation is strengthened across the board, even for tones in already prosodically strong positions. This means that an explanation of focus as prosodic headedness is inadequate for the findings of the two experiments presented here. Rather, the influence of prosody onto tonal coarticulation and the influence of focus onto tonal coarticulation have to be kept apart, even if the two factors induce similar effects onto the implementation of tonal contours.

5.6 Conclusion

In the present chapter, two research questions were investigated, as laid out in section 5.1.4: (i) which contextual factors affect tonal coarticulation in Wenzhou Chinese, and (ii) how do these factors interact with the strengthening of tonal implementation induced by focus? In order to test the first question, the implementation of rising and falling tones in rightmost and leftmost position in trisyllabic phrases in Wenzhou Chinese was investigated. By comparing the implementation of tonal contours in conflicting and compatible contexts, it was measured to what extent the tonal trajectories were affected by the adjacent tonal context.

In order to specify the exact nature of the contextual influence, two different prosodic structures were tested, which differed along two prosodic criteria: prosodic boundary strength and prosodic head position. Since these two criteria were non-overlapping, specific predictions could be made that would allow to test which of the two criteria was responsible for the prosody effect on tonal coarticulation.

The results of the first experiment indicated that it was prosodic headedness which could best explain the differences in tonal coarticulation between the two prosodic structures. In prosodically strong positions, it was found that tones were coarticulated relatively less with adjacent tones than in prosodically weak positions. In these weak positions, it was found that the tonal slopes, particularly the slopes that were immediately adjacent to the tonal context, were implemented with flattened or directionally flipped tonal trajectories.

At the same time, the durational results showed that this flattening of tonal trajectories in prosodically weak position was not a direct by-product of durational truncation. Compared to prosodically strong positions, speakers took a comparable amount of time for the realization of the tonal trajectories, but implemented them in a more distinct way in strong positions.

Under focus, what could be observed is best described as a general strengthening effect that boosted the tonal implementation both in prosodically weak and in prosodically strong positions. In the weak position, this meant that the difference between the two prosodic contexts in terms of tonal coarticulation disappeared, and both prosodic structures showed a similar amount of tonal coarticulation. However, tonal implementation was also strengthened in prosodically strong positions. Across the board, under focus, the amount of difference in tonal implementation between conflicting and compatible contexts was reduced. This means that tones were realized more autonomously and independently from the adjacent tonal context, even if this context conflicted with the tonal targets, under focus.

These findings have important consequences for theoretical accounts of the interaction between prosodic structure and focus. Particularly, the current results show that focus should not be conceptualized as being implemented as prosodic prominence. The effects of prosodic prominence on tonal coarticulation, as presented in the current chapter, are independent from the effects of focus on tonal coarticulation, even if the two factors influence tonal implementation in a similar fashion.

In that respect, the current findings present an argument against an extreme version of the view that focus influences tonal realization only indirectly, via modifying the prosodic prominence status of the focused constituent (see similar proposals in Chen 2009; Féry 2010; Féry & Ishihara 2010). Rather, the current findings suggest that both focus and prosodic structure affect the strength and autonomy of tonal implementation, but do so in a separate way.