Stellenbosch Papers in Linguistics Plus, Vol. 62(2), 2022, 113-136
doi: 10.5842/62-2-904

Mind the subtle f0 modifications: The interaction of tone and intonation in Sinitic varieties

Yiya Chen

Leiden University Centre for Linguistics, Leiden University, The Netherlands
E-mail: yiya.chen@hum.leidenuniv.nl

Abstract
Sinitic varieties are well known for their complex lexical tone systems. Lesser known is that these varieties also employ intonation for multiple communicative functions, ranging from indexing a speaker's socio-cognitive information (such as emotions and attitudes) to signaling various linguistic information (such as asking questions, marking focus, and encoding prosodic structure). This paper reviews the multiplexing of lexical tone and intonation into the same melodic f0 signal. The main empirical focus is on Standard Chinese; whenever possible, comparisons are made across Sinitic varieties (such as Shanghai Wu Chinese and Cantonese). I will show that lexical tone constrains the changes of f0 for intonation. How tone and intonation interact, however, varies across communicative contexts and language varieties, which is also reflected in how listeners utilize the f0 information to decode the melodic pitch signal during speech processing. From a cross-linguistic viewpoint, findings on intonation in Sinitic varieties suggest 1) the need for detailed acoustic and perceptual studies to understand the subtle f0 modifications for intonation in tone languages, and 2) the importance of a comparative approach to understanding the similarities and differences of intonation in tone languages.

Keywords: Chinese, Sinitic languages, tone and intonation, prosodic structure, emotion and attitude, focus and question

1. Introduction

Successful speech communication depends on the recognition of word meaning and comprehension of multiple layers of additional meanings that utterances may convey. In West-Germanic languages like English, much of the utterance-level information is expressed via a speaker's tone of voice. For example, saying "yes" (normally signaling affirmation) with a rising pitch contour[1] often implies a question. Pitch variation at the utterance level is known as intonation (e.g., Ladd 2008, Gussenhoven et al. 2013). Intonational meanings in speech have

[1] Pitch is the perception of fundamental frequency (f0; the rate at which vocal cords vibrate). The two are correlated and often used interchangeably. Voice quality has a small effect on listeners' judgment of pitch level but absolute f0 is the most important determinant (Bishop & Keating 2012).

been proposed to derive from anatomical and physiological effects on vocal production, known as the biological codes of vocal communication (Ohala 1984, Gussenhoven 2016). These codes are commonly believed to have evolved into language-specific intonation patterns that we observe, with mappings between voice pitch and communicative meanings that can be arbitrary and are often context-dependent.

Languages are fascinatingly diverse in how pitch variation is employed to convey different layers of meaning in speech. Let's again take English as an example. A wide range of messages is conveyed by intonation, such as asking questions, highlighting important information, signaling intention, and conveying attitude or emotion. However, close to half, if not more, of the world's languages also have lexical tones and rely on pitch variation to distinguish word/morpheme meanings (Yip 2002, Maddieson 2013). These languages are known as lexical tone languages. What makes tone languages intriguing is that in these languages, tone and intonation are conveyed in the same melodic pitch signal in speech.

Our knowledge of intonation in lexical tone languages has remained sporadic and compartmentalized, as noted by researchers who work on genealogically different tone languages (e.g., Brunelle et al. 2020 on Southeast Asian languages, DiCanio & Bennett 2020 on Mesoamerican languages, Kaufman & Himmelman 2020 on Austronesian languages, Hyman et al. 2020 on Sub-Saharan African languages, and Zhang et al. 2020 on Sinitic languages). Thus far, there are two landmark studies on intonation in tone languages. Gussenhoven (2004) is a monograph offering a phonological analysis of the interaction of tone and intonation in European languages (and Japanese); Downing & Rialland (2017) is an edited survey of intonation in African tone languages. Both are concerned with intonation for a subset of communicative meanings and are based on data collected by individual researchers with methods that are not directly comparable. Inspired by Laura Downing's significant contributions to our understanding of tone and intonation in African languages, I would like to take this opportunity and call for large-scale comparative research to investigate intonation with data from different lexical tonal systems in genealogically diverse languages.

This contribution aims to facilitate such cross-language comparisons. To this end, findings on intonation in Sinitic varieties will be introduced to readers who are not necessarily familiar with this language family. Sinitic varieties are also known as Chinese and Chinese dialects. It is worth noting that China is home to hundreds of related language varieties (that belong to different language families such as Sino-Tibetan and Tai-Kadai). Sinitic varieties are classified into ten major groups, each containing multiple subgroups (Wurm et al. 1987). These varieties are known to have tonal systems different from African tone languages (Yip 2002, Snider 2018), although they may have more similarities than commonly assumed (Evans 2008). My empirical focus in this review is on Standard Chinese.[2] Standard Chinese has four lexical tones: High (level) tone (H; T55), rising tone (R; T35), falling tone (F; T51), and low tone (L; T21), which is also known as a dipping tone given its (optional) rising tail (T214) when produced in

[2] Standard Chinese is also referred to as Mandarin Chinese or Standard Beijing Mandarin. Standard Chinese is based on northern Mandarin dialects, with Beijing Mandarin as its norm of pronunciation (Norman 1988: 135). Standard Chinese is considered a Mandarin variety but can be quite different from other Mandarin varieties not only in lexical tone (e.g., Li et al. 2019 on Tianjin Mandarin) but also in segment (e.g., Chen & Guo 2022 on Zhushan Mandarin).

isolation or at the final position of an utterance.[3] Following Chao 1968, Tone numbers here indicate pitch levels with 1 at the lowest end of a speaker's pitch range and 5 at the highest end. In addition, unstressed and reduced syllables do not carry any of the four lexical tones but are said to have a neutral tone (N).[4] The focus on Standard Chinese is because the intonation of this variety has been most extensively investigated. Data from other Sinitic varieties (such as Cantonese and Shanghai Wu Chinese) will be discussed when comparisons are possible.

As shown below, there is sufficient evidence to assume that in Sinitic varieties, intonation is present in every utterance, and it signals diverse linguistic functions and paralinguistic meanings. Section 2 will review findings on how intonation encodes the socio-cognitive information of a speaker. Sections 3-5 will zoom into and discuss the intonation for three linguistic functions: asking questions, marking focus, and encoding prosodic structure. My focus will be on pitch variation, but I will also briefly touch upon other intonation-related acoustic cues (e.g., intensity, duration, and voice quality).

2. Intonation for encoding socio-cognitive information

Pitch variation is employed in Sinitic varieties to encode a speaker's socio-cognitive information, such as emotion and attitude. Emotion reflects an individual's psychophysiological state (e.g., happiness and sadness). Vocal expression of emotion has primitive roots, as argued by Darwin (1872), and is commonly considered universal (Paulmann & Uskul 2014). Standard Chinese and Cantonese are the only two Sinitic varieties which have been subject to relatively detailed studies on the pitch encoding of emotion. One general conclusion is that lexical tone constrains the f0 manipulation for emotional speech (e.g., Li et al. 2011; Wang & Lee 2015). Li et al. (2011) investigated the f0 realization of lexical tones in Standard Chinese under seven emotional categories (i.e., anger, disgust, fear, sadness, happiness, pleasant surprise, and neutrality). Their results showed that tones are realized with a lowered pitch register for sadness and disgust, but a more expanded pitch range and raised pitch register for happiness and surprise. Note that in this study, only two speakers were recruited, but individual variations were nevertheless observed.

One particularly interesting observation about emotional intonation in Standard Chinese is that for certain emotions, speakers may annex additional pitch rises/falls to the lexical tone pitch contours. As shown in Figs. 1 and 2, such f0 annexation behaves like a boundary tone commonly observed in intonational languages.[5] (See similar observations in Chao (1968) with introspective observation, Mueller-Liu (2006) with dialogue data, and Li et al. (2011) with laboratory speech.)

Fig. 1 shows three f0 realizations of the same high (level) lexical tone in Standard Chinese, produced over the segmental syllable *ma* (/ma/) meaning 'mother' but uttered with different emotions, namely neutral, surprise, and anger (or annoyance; the two are not easy to differentiate). When uttered with surprise, there is a raised f0 with a slight rising contour; when

[3] In this paper, we will leave aside the issue of whether the primitives of lexical tones in Standard Chinese (and Sinitic varieties) consist of static high and low tones or dynamic rising and falling tones.
[4] Neutral tone is not marked with pitch specifications given its f0 variations after different lexical tones. For further details on the f0 patterns of neutral tone, see, e.g., Chen and Xu (2006).
[5] Unless otherwise noted, the figures are based on the speech production of one female speaker (born in the 70s), with a Sennheiser PC 131 headset and the standard recording settings in Praat (i.e., mono, 44100 Hz, and 16 bits).

spoken with anger/annoyance, there is an additive falling contour (in red circle) despite that the lexical high tone is typically realized with a level f0 contour (as shown in an emotionally neutral context).

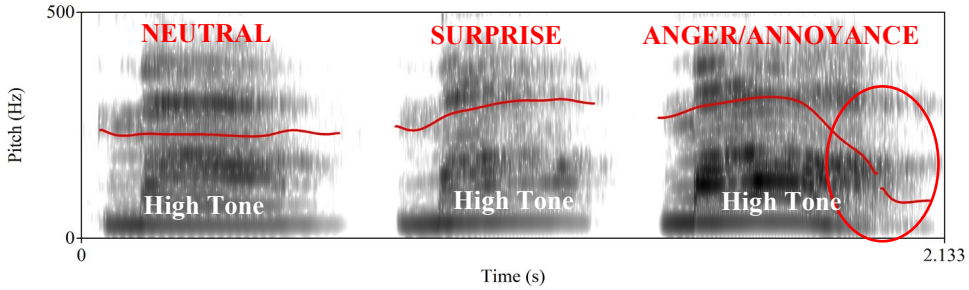

Fig 1. *ma* with a high (level) tone meaning 'mother' in Standard Chinese, uttered with different emotions (from left to right: neutral, surprise, and anger/annoyance)

In comparison, Fig. 2 shows three f0 realizations of the lexical falling tone (over *ma* meaning 'to scold'), produced with the same emotional categories as in Fig. 1. Note, in particular, the similarity of the f0 patterns when the high and falling tones are uttered with annoyance/anger. These figures demonstrate that distinct lexical tones under certain emotional contexts can be realized with rather similar f0 contours (in, e.g., f0 peak height and rise/fall slopes). Despite their similarities, the two tones are nevertheless distinguishable for native listeners and maintain their contrast.

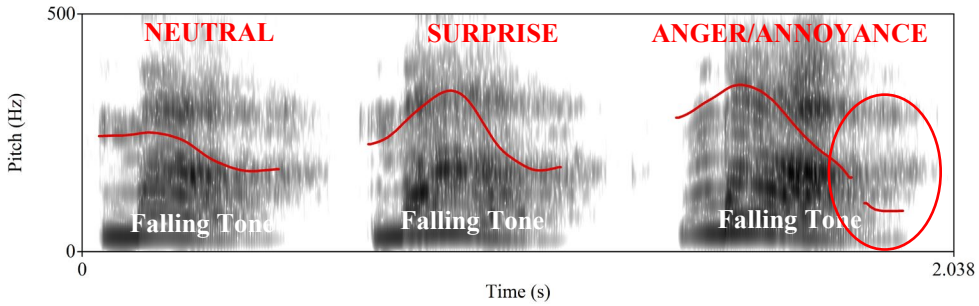

Fig 2. *ma* with a falling tone meaning 'to scold' in Standard Chinese, uttered with different emotions (from left to right: neutral, surprise, and anger/annoyance)

Another aspect worth noting is the salient lengthening of the final segments in the case of anger/annoyance, which accompanies the emotion-induced pitch annexation (or boundary tone). Such elongation is possibly due to both the effect of emotion on duration and the interaction of segmental structure and intonation, which, if confirmed with further investigation, lends support from lexical tone languages that although intonation and segmental structures are often separate (Pierrehumbert 1980), they are not wholly independent.[6]

Although pitch variation is utilized to distinguish emotions in Standard Chinese, its efficacy has been shown to reduce when compared to languages like English and French (Hirst et al.

[6] Interested readers are referred to Grice et al. (2018) for similar interdependency of intonation and segmental structures in intonation languages.

2013). It is important to note that despite the constraint of lexical tone on the f0 expression of emotion, Mandarin listeners' identification of emotion remains somewhat reliable. This is probably due to other cues that speakers are able to employ to convey emotion, such as intensity and voice quality (e.g., Chao 1968; Yuan et al. 2002; Wang & Lee 2015) as well as non-verbal cues (e.g., Huang et al. 2012).

Another category of socio-cognitive information besides emotion is attitude. Attitude refers to speakers' social effects and interpersonal stances such as sarcasm, humor, confidence, and friendliness (Wichmann 2000). Attitudes may be considered higher-level socio-cognitive information than emotions (which mainly reflect the internal and instinct states of a speaker). Compared to emotion, existing studies provide even less information about how the f0 contours of lexical tones are modified to convey different attitudes. Nevertheless, pitch has been shown to serve as an essential cue for attitudes in different Sinitic varieties. For example, sarcasm is produced with an elevated mean f0 but reduced f0 range (Cheang & Pell 2008, 2009 on Cantonese); friendliness induces a higher pitch register (Li & Wang 2004 on Standard Chinese). More generally speaking, f0 helps to contrast the positive vs. negative valence of a range of attitudes (e.g., friendly vs. hostile, polite vs. rude, and confident vs. uncertain) (Gu et al. 2011 on Standard Chinese). Other prosodic cues have also been reported, such as a slower speech rate for sarcasm and a lower harmonic-to-noise ratio (i.e., less vocal noise) for humor (Cheang & Pell 2009 on Cantonese). There is some evidence from Cantonese that speakers are able to successfully identify different attitudes via these prosodic cues (Cheang & Pell 2011).

More research is needed to replicate the existing findings and to investigate further the relationship between f0 and socio-cognitive information in tonal languages. Take friendliness, a category of attitude, as an example. It is commonly associated with high pitch, supporting the frequency code hypothesis (Ohala 1994; Gussenhoven 2004), which states that a high-pitch voice signals submissiveness through the vocal projection of body size (as smaller animals typically emit higher-pitched sounds). Chen et al. (2004) argued for a language-specific component in such universal tendencies. It is important to note that a simple contrast of pitch high vs. low in a tone language is not sufficient. As shown in Fig. 1 and Fig. 2, a high pitch in Standard Chinese does not just signal friendliness; other categories of socio-cognitive information, including surprise and anger, also correlates with a raised f0. So, pitch height is not informative and could subsume multiple distinct pitch patterns for different socio-cognitive (and other) meanings. More refined measurements of f0 trajectories are essential to reveal the vocal modulation of pitch for socio-cognitive information. Furthermore, systematic comparisons are crucial to understand their relationship. Their mapping is clearly not one-to-one but likely many-to-many; the open question is how they are mapped and what constrains the mapping(s).

3. Intonation for linguistic functions: Asking questions

Languages tend to develop grammatical means to convey communicative information. A case in point is the various sentence modes (e.g., declarative, yes-no question, *wh*-question) for different illocutionary forces (e.g., assertion, inquiry). The pitch marking of questions is quite conventionalized in languages, known as question intonation(s). The general tendency is to have rising pitch contours for questions and falling ones for statements (e.g., Bolinger 1978; Cruttenden 1981). This pattern has been attributed to the biological frequency code that associates size-related socio-cognitive meanings with the questioner being informationally

submissive (Ohala 1994; Gussenhoven 2016). There is increasing evidence against the universality of rising question intonation (e.g., Rialland 2009 for question intonation ending with a low tone and other phonetic features; Gunlogson 2003 for rising declarative intonation; and Warren 2016 for uptalk). Thus far, the limited experimental data on question intonation in Chinese confirm the general rising tendency and show an upward f0 trend, which becomes more pronounced towards the end of the utterance. However, the specific patterns of tone-intonation interaction are complex and vary across dialects, as illustrated below.

Before we zoom into the details of tone-intonation interaction, I will provide a brief overview of the grammatical means for conveying questions and question intonation, based primarily on observations in Standard Chinese. First, Chinese is a *wh-in-situ* language, in which *wh*-words remain at the base position as in their declarative counterparts (Cheng 1991). This is illustrated in (1), where the object of the verb *buy* is inquired. In English, the *wh*-word *what* is fronted, but in Standard Chinese, its equivalent *shénme* remains in the object position after the verb. Here, the diacritics denoting lexical tones in Pinyin[7] are included: High (¯) Rising (ˊ) Low(-dipping) (ˇ) and falling (ˋ). The tone sequence of the syllables in (1) is H, R, F, L, R, and N, respectively.

(1) 安妮要买什么?
ānní yào mǎi ***shénme***
/an ni jau mai ʃə mə/
H-R-F-L-R-N
Annie_want_buy_what
'**What** does Annie want to buy?'

Second, Sinitic varieties have also developed a rich set of question particles (e.g., 吧 *ba*, 吗 *ma*, 呢 *ne*) (Chao 1968; Li 2006). As these particles are often optional, an interrogative utterance may appear string identical to a declarative. The different string-identical clause types can be distinguished effectively via intonation (Lee 2005, Liu & Xu 2005, Yang et al. 2020).

Third, for the same interrogative construction, speakers can encode different illocutionary forces via f0, as shown in Zahner et al. (2022). For example, in Standard Chinese, the question in (2) can be interpreted as requesting information from the addressee (i.e., an information-seeking yes-no question) or committing the addressee to the answer that is presupposed (i.e., a rhetorical question). Compared to an information-seeking yes-no question (Fig.3; left), rhetorical questions (Fig.3; right) show an overall lowered f0 and longer duration. There could also be phonatory differences (more often with non-modal voice quality in a rhetorical question). These cues are modified in tandem as an *ensemble* rather than trading for each other. In this example, the tone sequence of the syllables is L-R-L-L-R-R, respectively.[8]

[7] Pinyin is the official romanization system for Standard Chinese.
[8] When a low tone is followed by another low tone, it is realized with a rising f0 contour (similar to that of a lexical rising tone), known as low tone sandhi. For more details on the acoustic realizations of low-tone sandhi in Standard Chinese, see Yuan & Chen (2014) and the references therein.

(2) 有人想买柠檬?
yǒurén xiǎng mǎi níngméng
/jou ɹən ɕaŋ mai niŋmən/[9]
L-R-L-L-R-R
anybody_want_buy_lemon
'Does anybody want to buy lemon?'

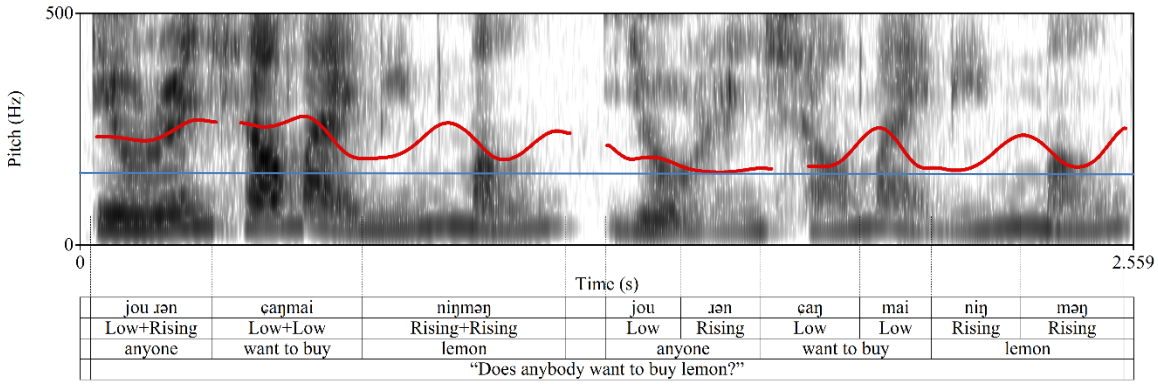

Fig 3. Information seeking yes-no question (left) and rhetorical question (right) of the same interrogative construction *yǒurén xiǎngmǎi níngméng?* 'Does anybody want to buy lemon?'

How do lexical tones and intonation interact? Experimental studies on the production and perception of Chinese intonation for question marking have been mainly conducted on Cantonese and Standard Chinese. So, in the following, I will focus on these two varieties and discuss the interactions of lexical tone and intonation in more detail.

Cantonese has six lexical tones: a falling tone (T21), two rising tones (T23 and T25), and three level tones (T55, T33, and T22) (Bauer & Benedict 1997). So, compared to T25, T55 and T33 may be grouped into higher-register tones and the rest to lower-register tones (T21, T23, T22). When uttered as a question, T21, T23, and T22 are realized with a rising tail and appear similar to T25, as illustrated in Fig. 4. The other two higher-register tones (T55 and T33) may also be realized with a rising tail. As shown in Fig. 5, these two tones remain relatively distinct from the other tones even in question, in part due to their higher-register tonal onset.[10]

When the lexical tones are embedded in a sentence, Gu et al. (2005) and Ma et al. (2006) have shown that question intonation in Cantonese is marked by a relatively subtle raising of f0 for the non-final syllables and a salient f0 rise localized at the end of an utterance for all tones, sometimes at the cost of their distinctive f0 contours, as illustrated in Fig. 4. Perceptually, the local final f0 rise in Cantonese has been shown to serve as a reliable cue for question intonation (Ma et al. 2011; Xu & Mok 2011) and consequently, to incur difficulties in spoken word recognition (Kung et al. 2014). In short, Cantonese seems to favor cueing intonation at the risk

[9] Note that we have transcribed *xiang* as /ɕaŋ/. It is to be further investigated whether the glide /j/ should be posited. Chinese dialects also seem to differ in whether /j/ is necessitated in the underlying representation of syllables with a palatal/alveolo-palatal onset.
[10] The Cantonese data plotted in Fig. 4 and Fig. 5 were produced by a male speaker of Cantonese (born in the 90s), with Sony MDR-ZX310AP headphones and the standard recording settings in Praat.

of lexical spoken word misidentification. The production and perception data on tone and intonation together suggest that the final f0 rise for Cantonese questions is salient. It is possible that such f0 modification for question intonation may have been grammaticalized into a local high boundary tone, which in turn serves as an optimal cue for question intonation.

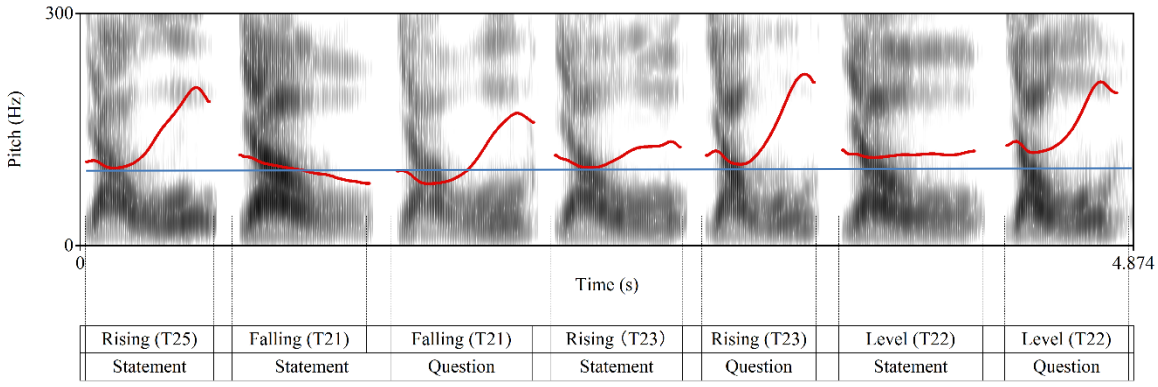

Fig. 4. f0 contours of low falling (T21), low rising (T23), and low level (T22) tones in Cantonese, produced as a statement or a yes-no question, in comparison to the rising (T25) tone in statement intonation. (The tone-carrying syllable is /jɐu/, with T25 meaning 'oil',[11] with T23 meaning 'to have', with T22 meaning 'again', and with T21 meaning 'by'.)

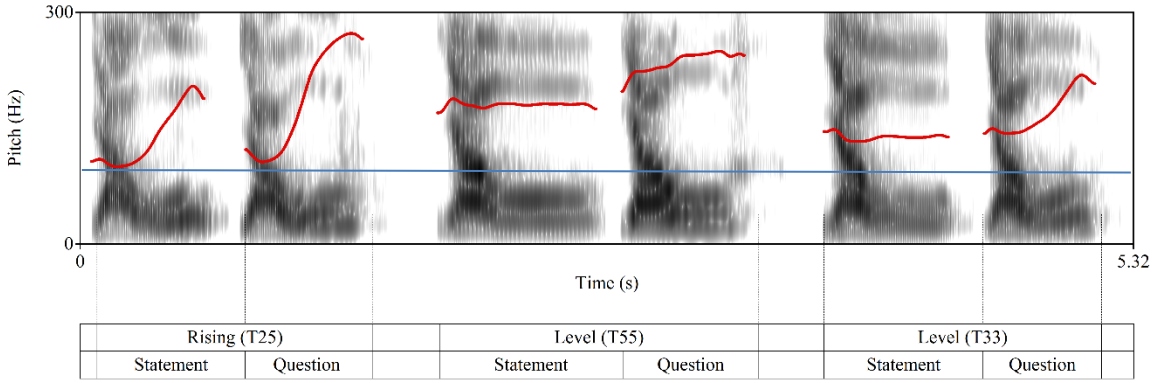

Fig. 5. f0 contours of high level (T55) and mid (T33) tones in Cantonese, produced as a statement or a yes-no question, in comparison to the rising (T25) tone. (The tone-carrying syllable is /jɐu/, with T25 meaning 'oil', with T55 meaning 'to rest', and with T33 meaning 'young'.)

[11] This tone specification relates to morphological alternation. It is an instance of morphologically derived tone (known as *pinjam*), where /jɐu/ with T21 'to grease' is belied to change to /jɐu/ with T25 'oil' via verb nominalization. See Yu (2007) for further discussion on *pinjam* in Cantonese and whether the phonetic realizations of the lexical T25 and the morphological T25 are identical.

Now, let's turn to Standard Chinese. Generally speaking, question intonation in Standard Chinese is also encoded via a globally raised f0 with an accelerating local utterance-final f0 rise (Yuan 2004), as in Cantonese. What differs from Cantonese is that the magnitude of global f0 raising in Standard Chinese seems larger, while the utterance-final local f0 rise shows more respect towards the lexical tone f0 contours. As a result, all four lexical tones in question retain their distinctive f0 contours.

Illustrated in Fig. 6 are the f0 contours of two tonal sequences. One is composed of a high tone followed by a rising tone, and the other consists of a high tone followed by a falling tone. Both were produced as either a statement (left) or a yes-no question (right). At the end of a question, a falling tone, often slightly raised, maintains its falling f0 contour (but would not fall as low as in a declarative sentence). In contrast, a rising tone is realized with its characteristic and more exaggerated rising f0 contour. Both show a raised mean f0 compared to that of a statement.

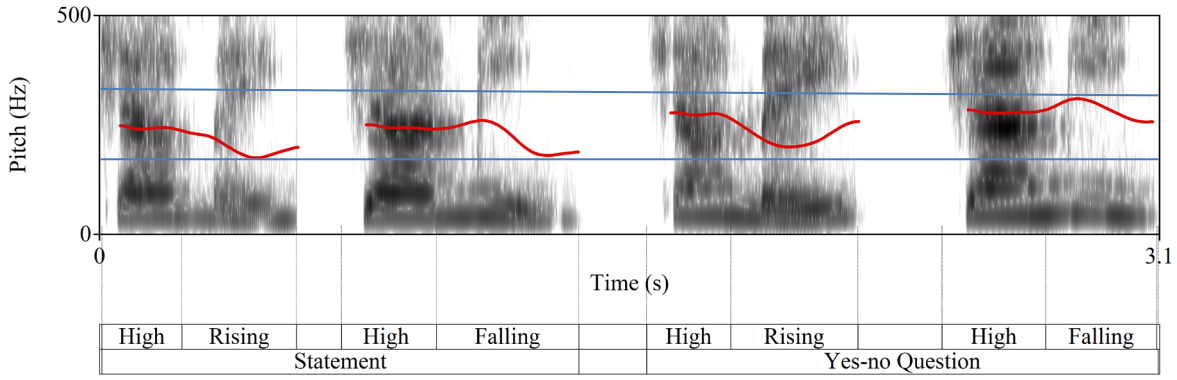

Fig. 6. f0 contours of two disyllabic tonal sequences (high-rising and high-falling) in Standard Chinese, produced as a statement (left) or a yes-no question (right). (The stimuli are: *jiā méi* /tɕa.mei/ 'to add coal'; *jiā mì* /tɕa.mi/ 'to add honey'.)

Standard Chinese thus seems to favor preserving tonal identity at the lexical level even at the risk of intonation misidentification. This is also reflected in the high accuracy rate in native listeners' lexical tone identification regardless of intonation type and their difficulty in recognizing statement vs. question, especially when the utterance-final tone is T35 (Yuan 2011, Liu et al. 2021). A similar asymmetry in intonation identification (between rising and falling tones) has also been reported in neural processing studies (Ren et al. 2009, Ren et al. 2013, Liu et al. 2016).

What is striking is that even when an utterance ends with a neutral tone syllable, we see a similar question-induced f0 raising to that observed in lexical-tone sequences. The neutral tone in Standard Chinese has been argued to have a phonetic pitch target at the mid-low level. Due to the weak phonetic implementation in an unstressed syllable, its f0 realization is subject to significant influence from the preceding lexical tone and it takes several neutral tone syllables to observe the convergence towards their mid-low phonetic target (Chen & Xu 2006).

Figs. 7-9 illustrate the f0 contours of the four lexical tones followed by a varying number of neutral tone syllables: One neutral-tone syllable in Fig. 7, two in Fig. 8, and three in Fig. 9.

These tonal sequences are produced as a statement (left) or a yes-no question (right). Let's focus on the neutral tones uttered in a statement (in solid circles). In Fig. 7, we observe that when there is only one neutral tone syllable, the f0 realization of the neutral tone varies significantly as a function of the preceding lexical tone. When the number of neutral tone syllable increases to two (Fig. 8), we observe slowly converging f0 realizations (except for the neutral tone after a Low tone). Fig. 9 shows that by the end of the third neutral tone syllable, there is better convergence towards a low phonetic pitch target.[12]

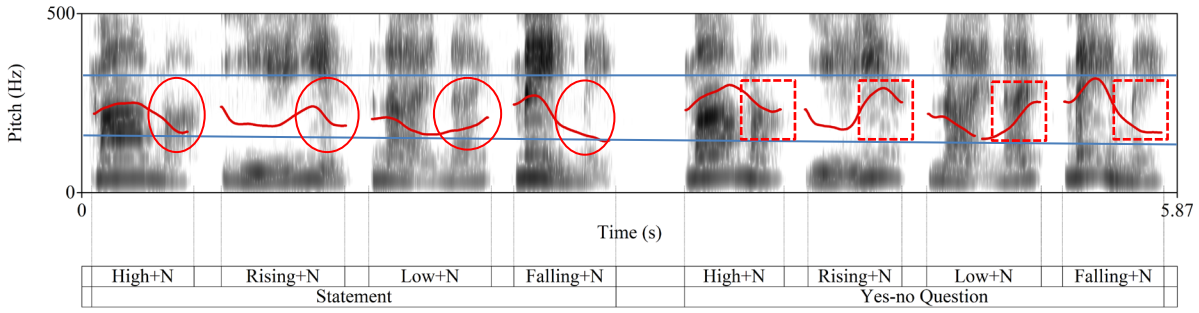

Fig. 7. f0 contours of disyllabic tonal sequences in Standard Chinese, with each of the four lexical tones followed by one neutral tone syllable, produced as a statement (left) or a yes-no question (right). (The stimuli are: *māma* /ma.ma/ 'mother', *yéye* /je.je/ 'grandpa', *năinai* /nai.nai/ 'grandma', and *mèimei* /mei.mei/ 'sister'.)

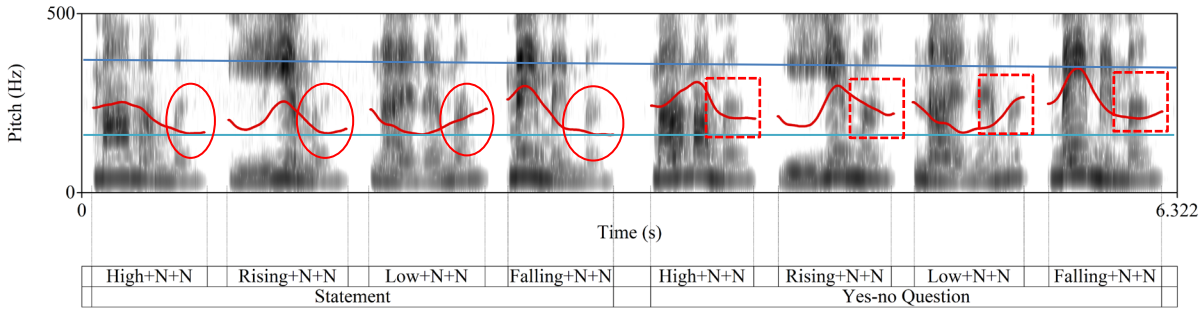

Fig. 8. f0 contours of tri-syllabic tonal sequences in Standard Chinese, with each of the four lexical tones followed by two neutral tone syllables, produced as a statement (left) or a yes-no question (right). (The stimuli are: *māmamen* /ma.ma.mən/ 'mothers', *yéyemen* /je.je.mən/ 'grandpas', *năinaimen* /nai.nai.mən/ 'grandmas', and *mèimeimen* /mei.mei.mən/ 'sisters'.)

[12] In the final position, the neutral tone seems to be realized with a similar f0 trajectory as in an utterance-medial position but within a lower pitch range. Whether such observed low f0 realization in Standard Chinese is due to a general utterance-level declination effect or a low pitch target specified for neutral tone in a final position (i.e., different from an utterance-medial position) needs further investigation.

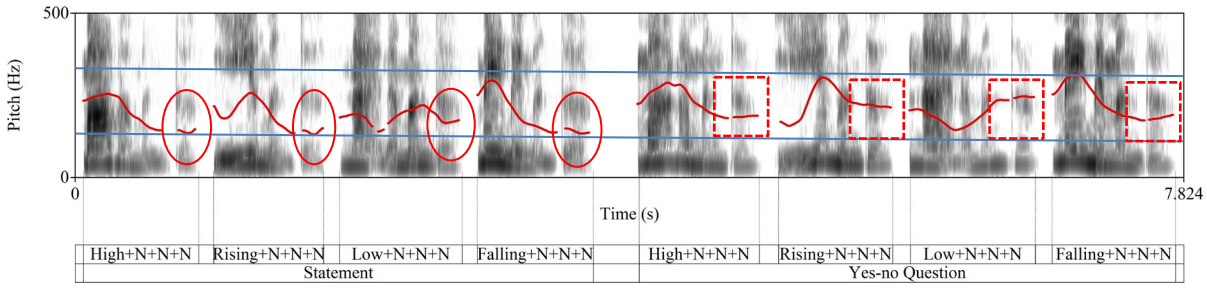

Fig. 9. f0 contours of quadri-syllabic tonal sequences in Standard Chinese, with each of the four lexical tones followed by three neutral tone syllables, produced as a statement (left) or a yes-no question (right). (The stimuli are: *māmamende* /ma.ma.mən.də/ 'of mothers', *yéyemende* /je.je.mən.də/ 'of grandpas', *năinaimende* /nai.nai.mən.də/ 'of grandmas', and *mèimeimende* /mei.mei.mən.də/ 'of sisters'.)

This pattern of variance and constancy in neutral tone f0 realization remains the same regardless whether the utterance is a statement or a question. In questions (Figs. 7-9; right), we also observe f0 realizations of the neutral tone (in dotted circles) as a fucntion of the preceding lexical tones and the number of neutral tone syllables. Furthermore, the general raising of f0 (in terms of higher f0 peak and raised f0 register) in yes-no questions is similar to what we have observed in the lexical tone sequences in Fig. 6. This casts doubts on the existence of a specific boundary tone for neutral-tone syllables.

In short, one consistent observation in the literature is that tone and intonation interact in both Standard Chinese and Cantonese, but with clear differences. In the face of competition for f0 cues for both lexical tone and intonation, Standard Chinese seems to opt to signal tonal identity even at the expense of intonation recognition, which renders the utterance-final syllable a less reliable f0 cue-bearer for interrogativity than it is in Cantonese. Consequently, question-induced final f0 rise in Standard Chinese is unlikely to have been grammaticalized into a local high boundary tone.[13] The two dialects also seem to differ in the time course of question perception. No effect of early intonation detection has been observed for Cantonese. In Standard Chinese, however, Jiang & Chen (2011) reported above-chance identification (74%) with only the subject (gated from a subject-verb-object utterance) in a two-way (question vs. statement) forced-choice task (but see Gryllia et al. 2020 for an early identification effect based on prosodic cues other than f0). Further research is needed to replicate the existing findings and to investigate the tone-intonation interaction in other Sinitic tonal varieties (e.g., Tianjin Mandarin in Zhang 2018). Moreover, it is important to examine the production-perception link, especially regarding how the f0 cues found in the production data are utilized, weighted, and integrated by listeners for question perception.

[13] One may posit a boundary tone for question intonation in Sinitic varieties (e.g., Lin 2006) and attribute the f0 variations to the phonetic realizations of this boundary tone. The issue that arises is how to model the complex mapping between the varied f0 modifications over different lexical tones and the phonological boundary tone, which could become intractable.

4. Intonation for linguistic functions: Marking focus

Another well-studied case of grammatical means for communicative meanings is focus constructions. Focus here refers to new and contrastive information in the discourse.[14] Take English as an example. In answer to the question of what Mary teaches (3A), MATH (3B) is focused and pronounced with prominence (indicated with capital letters). Focus is signaled mainly via salient pitch variations (known as pitch accent) (Ladd 1980, Gussenhoven 1992), in addition to durational and intensity changes (Breen et al. 2010). Languages may also grammaticalize the melodic marking of focus (Frota 2000).

(3) A: What does Mary teach?
B: Mary teaches MATH.

In various Chinese dialects, despite the rich inventory of sentence-final particles and grammatical constructions for different information structure notions (Chen et al. 2016 and references therein), focal prominence has nevertheless been reported as being cued via pitch variation (e.g., Xu 1999 on Standard Chinese), durational lengthening (e.g., Chen 2006 on Standard Chinese), intensity increase (e.g., Lee et al. 2016 on Standard Chinese, cf. Jin 1996), and hyper-articulated segmental contrasts (e.g., Chen 2008a on Shanghai Wu Chinese).

Fig. 10 illustrates the f0 realizations of the four lexical tones in Standard Chinese (over the segmental syllable *miao*) excised from a template sentence where the target tone is preceded by a high tone and followed by a rising tone. (The data were reported in Chen & Gussenhoven 2008.) The target tones were either produced in a pre-focus condition or an on-focus condition, which show focus-induced longer duration, more exaggerated f0 contours that are characteristic of the lexical tones and, less visibly, higher intensity and more hyper articulated segments. For low pitch targets, focus also induces more creakiness, especially in the low tone.

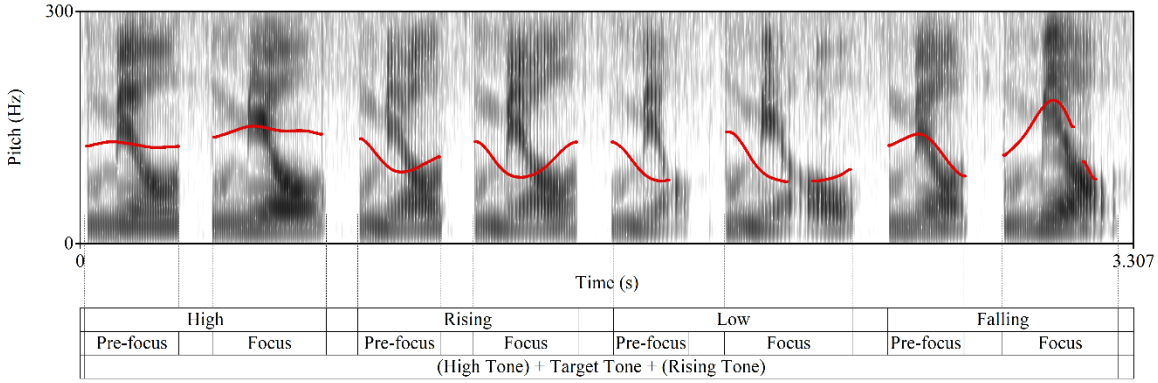

Fig 10. f0 contours of the four lexical tones (over the segmental syllable *miao*), excised from a template sentence where the target tone is preceded by a high (level) tone and followed by a rising tone. The target tones were produced in a pre-focus condition (left) or an on-focus condition (right).

[14] Readers are referred to, e.g., Chafe (1974), Rooth (1996), Gussenhoven (2007), and Krifka (2008) for more detailed discussions of focus and other information structure notions.

Experimental studies on the f0 marking of focus in Standard Chinese have repeatedly reported that under various types of focus conditions (e.g., emphasis, correction, new or contrastive information), on-focus lexical tones are realized within an expanded f0 range grid, and out-of-focus tones within a compressed range grid (e.g., Gärding et al. 1983, Shih 1988, Jin 1996, and Xu 1999). What has remained an open question is whether focus should be viewed as directly encoded via a tri-zone pitch range control: expansion under focus, compression after focus, and little or no change before focus (Xu & Xu 2005; Xu 2005). One alternative is the prominence-marking view of focus expression, explored in Chen (2003, 2009 & 2010) and Chen & Gussenhoven (2008). (See review in Chen 2012 and references therein from a cross-linguistic perspective.) Under this view, the phonological reflex of focus is utterance-level prosodic prominence, the acoustic expression of which is contingent upon the lexical and prosodic properties of the focused constituent. Under focus, greater articulatory force applies to both lexical tone and segments. As shown below, there are within- and cross-dialect variations in both on-focus f0 range expansion and post-focus f0 compression, which are not expected under the view of a strict tri-zone focus range manipulation, lending support to the prominence-marking view.

Concerning on-focus f0 effects, it has been shown that in some dialects, f0 range expansion is only reliable and salient when the distinctness of contrastive lexical tones is not compromised. Take Shanghai Chinese (a Wu dialect) as an example. It has five lexical tones: high-register rising tone (T34), low-register rising tone (T23), falling tone (T53), short high-register level tone (T5), and short low-register rising tone (T12) (Xu & Tang 1988).[15] Chen (2009) showed that among the five lexical tones, the high-register rising tone does not show significant f0 range expansion under focus, presumably to ensure its distinctness from the low-register rising tone. This is because significant f0 range expansion of the high-register rising tone would likely result in overlapping the f0 spaces for the two rising tones, which, consequently, would become less distinguishable. Taiwanese (Min) seems to parallel Shanghai Chinese and lacks consistent focus-induced f0 range expansion across lexical tones. Pan (2007) showed salient f0 raising/range expansion in the HH and HL tones but not in the MM and ML tones. These studies converge on the influence of lexical tonal properties on the magnitude (or absence) of focus-induced f0 range manipulation for intonation.

In the post-focus position, both Cantonese (Man 2002, Wu & Xu 2010) and Min dialects (Pan 2007, Xu et al. 2012, Chen et al. 2009) have been documented to lack f0 compression. Even for Mandarin, which is known for a general post-focus f0 compression effect, lack of f0 compression in specific tonal contexts has been observed (Xu 1999 on Standard Chinese; Shen & Xu 2016 on Lanyin Mandarin; Duan, Jia & Ran 2013 on Jiaoliao Mandarin). Chen (2010) showed that in Standard Chinese, a post-focus falling tone after a focused high-ending lexical tone is realized with an expanded pitch range but without the characteristic f0 contours of a falling tone observable in a prominent condition. This is presumably due to the influence of tonal co-articulation from the preceding tone and the weak implementation of the target tone in a post-focus non-prominent condition. In Honggu Mandarin, a Lanyin Mandarin dialect, tonal neutralization has also been observed in a post-focus condition (Ge & Li 2020), in line with the view that post-focus lexical tones are weakly implemented. These observations cannot be accounted for via tri-zone f0 range manipulation.

[15] See Chen & Gussenhoven (2015) on the disagreements among researchers regarding the numerical pitch values of the five tones in Shanghai Wu Chinese.

The f0 reflex of focus seems also sensitive to prosodic domains. In Wenzhou Wu Chinese, a disyllabic word serves as the domain for contextual tonal alternation (known as tone sandhi) (Chen 2000). F0 range expansion is distributed over the entire sandhi domain even when corrective focus is elicited over just one syllable within the domain. This suggests that the disyllabic tone sandhi domain serves as the minimal domain for focus expression (Scholz 2012; Scholz & Chen 2014). Shanghai Wu Chinese shows a similar sensitivity of focus expression to the tone sandhi domain: corrective focus affects the f0/duration patterns of the whole tone sandhi domain rather than the individual pragmatically focused syllables (Chen 2009; Ling & Liang 2017). Thus, these Wu dialects contrast with Standard Chinese, in which corrective focus can be easily elicited over one single syllable within a multi-syllable plae name (Chen 2006).[16]

In short, what has transpired in the existing literature is that under focus, tones are typically realized with magnified f0 contours, characteristic of the lexical tone identity.[17] They are therefore maximally distinguishable when produced under focus and perceived as prominent. What remains to be further researched is the prosodic marking of different types of focus. In Standard Chinese, there is some evidence that new informational focus is different from corrective focus (Chen & Braun 2006, Wang & Xu 2011, Ouyang & Kaiser 2013). Focus types are also reflected in their different cognitive processing patterns as revealed by readers' eye movements (Chen et al. 2012) and brain responses (Chen et al. 2014). More cross-dialect comparisons are needed to better understand the prosodic encoding of different types of focus and particularly, the interaction of tone and intonation in focus marking. Furthermore, the investigation of focus also needs to be situated in our understanding of the prosodic realization of other information structure notions such as topic. Topic identifies the entity in a discourse about which a speaker provides further information (or comment). Chinese is known as a topic-prominent language (Li & Thompson 1981). A range of syntactic structures has been identified as topic constructions (Xu 2006). When different topics contrast, more robust f0 raising and gradual f0 lowering over the rest of the utterance have been reported (Chen 2009 on Shanghai Wu Chinese; Wang & Xu 2011 on Standard Chinese). Future research needs to establish possible prosodic differences between focus and contrastive topic.

5. Intonation for linguistic functions: Encoding prosodic structure

In West-Germanic languages, a prominent role of intonation is to encode the prosodic structure of utterances. Prosodic phrasing reflects the syntactic structure of an utterance, though not necessarily in a one-to-one mapping (see, e.g., Shattuck-Hufnagel & Turk 1996, Selkirk 1995, and Féry 2016 for reviews on the prosodic structure and syntax-prosody mapping). Prosody facilitates the parsing of an incoming acoustic signal (Beckman 1996).

In Standard Chinese, a combination of acoustic cues such as pitch modification, lengthening, and pause duration also helps encode the prosodic structure of an utterance (Li 2002, Li & Yang 2009). Fig. 11 illustrates two possible prosodic groupings of the same string of words, *yā-bù-chī-le* (literally meaning 'duck-not-eat-aspect marker'), which can encode two different syntactic structures. (The tone sequence of the syllables is H-F-H-N.) On the left, *yā 'duck'* is

[16] Note that probably due to influence from Mandarin, younger-generation Shanghainese speakers have also been observed to break up a typical tone sandhi domain and express focus on a single syllable.

[17] Tonal realization under focus respects contextual tonal coarticulation when the neighboring tones are associated with syllables which are not across major prosodic boundaries. See Li and Chen (2019) on the interaction of focus, tonal coarticulation, and prosodic boundary in Tianjin Mandarin.

grouped with the rest as an intonational phrase to encode a subject-predicate structure, giving rise to the interpretation that the duck under discussion has stopped eating. On the right, *yā* '*duck'* is lengthened while the rest of the words are phrased together, signaled via a raised f0 onset, strengthened stop onset release, and longer duration of the phrase-initial syllable. Such a phrasing facilitates a topic-comment interpretation.

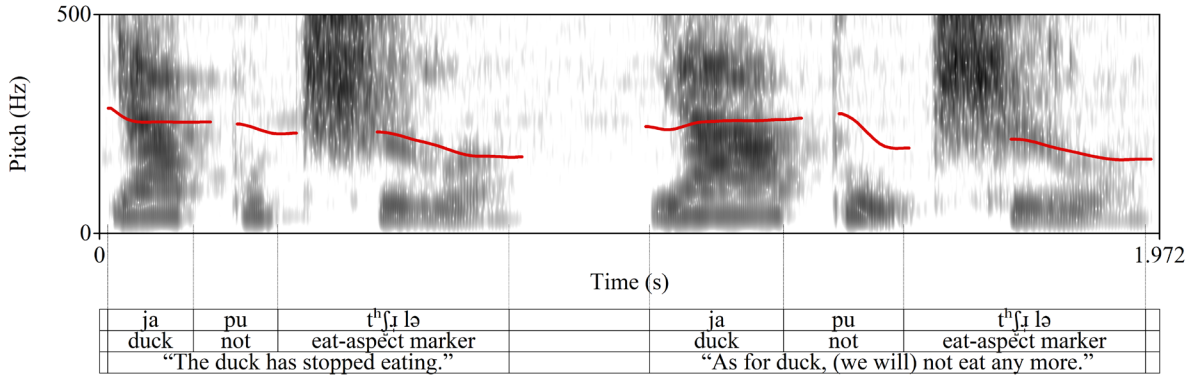

Fig 11. Spectrograms and f0 contours of *yā-bù-chī-le* /ja pu tʰʃɻ lə/. On the left is a subject-predicate construction where '*duck'* serves as the agent. On the right is a topic comment construction where '*duck'* serves as the patient.

Prosodic grouping can also be directly encoded via lexical tone sandhi in certain tonal sequences. In Standard Chinese, tone sandhi helps to disambiguate underlying syntactic structures. Take (4) as an example. 好 *hǎo* 'well/good' can be phrased in two ways. It can be phrased with 管 *guǎn* 'to manage', meaning 'to manage well', or with 酒家 *jiǔjiā* 'restaurant', meaning 'good restaurant'. As illustrated in Fig. 12, the f0 realizations of the lexical tones (i.e., L-L-L-H) of this same string of syllables differ significantly, surfacing as SR-SR-L-H and L-SR-L-H, respectively. (SR represents the sandhi rising variant of the lexical low tone.[18]) More specifically, the f0 encoding of the prosodic structures in (4) is expressed via both tone sandhi change over *guǎn* (in circle) and f0 scaling over *jiǔ*, in addition to other prosodic cues).

(4) 管好酒家
guǎn_hǎo_jiǔjiā
L-L-L-H (citation tones of the morphemes)
/kwan hau tɕou tɕa/
manage_well/good_restaurant
'to manage a restaurant well' or 'to manage a good restaurant'

[18] The f0 realization of the low tone here may be modulated by speaking rate. For example, at a slow speaking rate with pause, the second low tone may not surface as SR. Low tone is often realized with creakiness/glottalization as shown in *jiǔ* in Fig. 12.

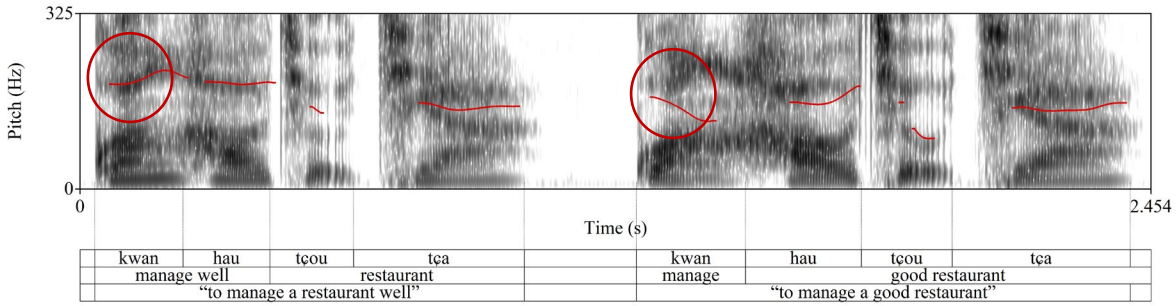

Fig 12. Spectrograms and f0 contours of *guǎn_hǎo_jiǔjiā* /kwan hau tɕou tɕa/. On the left, *hǎo* is phrased as a complement of the verb, meaning 'to manage a restaurant well'. On the right, *hǎo* is phrased as a modifier of the following noun meaning 'to manage a good restaurant'.

Worth noting is that there is evidence from both behavioral and neurophysiological studies that listeners of Standard Chinese rely on the subtle and nuanced f0 modifications/alternations at the sentence level, together with other cues, to disambiguate otherwise identical strings of words or to predict upcoming speech during online sentence comprehension. For further details, see, e.g., Speer et al. (1989) on the effect of low tone sandhi in speech processing, Li & Yang (2009) on the perception of prosodic hierarchical boundaries, and Li et al. (2011) on the role of phrasing and prominence in disambiguating relative clause from verb-object construction. One may argue for an equivalence view that intonation in Standard Chinese functions similarly to that in non-tonal languages in effectively guiding natural speech processing, with similar time courses and neural responses. Further research is needed to test this possibility with more data from other Sinitic tonal varieties. To this end, what is urgently needed is data on the f0 modifications and tonal alternations that jointly encode the prosodic and information structures.

6. Conclusion and future research

I hope that this brief introduction to intonation in Sinitic varieties serves well to call for more systematic and comparative research on intonation in typologically different tone languages.

My review has focused mainly on Standard Chinese, venturing into other Sinitic varieties whenever possible. It is important to emphasize that Chinese dialects differ significantly in their lexical tonal systems and in particular, contextual tonal variations. For example, in Wu dialects, lexical tones are often neutralized within a tone sandhi domain (e.g., Chen 2008b on Shanghai Chinese). Thus, Wu dialects may provide better docking sites for tonal events that are designated for post-lexical intonation, in comparison to Mandarin and Cantonese (where lexical tone realization is syllable-based and more dense). More research is needed to understand the range of strategies that different Sinitic varieties adopt to incorporate intonational pitch features into the same melodic pitch signal where lexical tones also reside. Furthermore, granted with such post-lexical f0 modifications for intonation, the question that arises is in what ways they are different from non-tonal languages, where salient f0 variations necessitate the presence of intonational tonal structures such as prominence-lending f0 movements (known as pitch accents) and edge marking f0 movements (known as phrase accents and boundary tones).

In addition to the open questions and knowledge gaps discussed above, it would also be important to place the varying patterns of pitch marking (or lack thereof) for intonation in a

broader perspective. The main focus of this review has been on the multiplexing of lexical tone and intonation into the same melodic f0 signal. One issue to be considered in the future is the possible functional trade-off between pitch variation and other linguistic means of marking sentence-level meanings. For example, it is important to consider f0 variation together with other speech cues (such as voice quality and intensity), grammatical constructions and particles, as well as non-verbal information (such as gestures and facial expressions). Another crucial development is to understand how the different layers of meaning (such as emotions, attitudes, and linguistic functions) are intertwined and how those connections may be manifested differently and jointly in the same melodic pitch signal. For both, comparisons across different lexical tonal systems as well as between tonal and intonational languages are essential.

Given the complexity of intonation f0 marking in tone languages, it is also a methodological challenge to ensure that subtle and nuanced pitch variations can be adequately captured with replicable patterns. Only with solid empirical data revealing the intricate interactions of tone and intonation can we achieve a proper understanding of the cross-language diversity and universality of melody in speech.

Acknowledgments

I would like to thank Yifei Bi, Promise Dodzi Kpoglu, Priscilla Lam, Menghui Shi, Charlie Xu, Xinyi Wen, and Tingting Zheng for their questions and comments on different versions of the manuscript.. I'd also like to thank the anonymous reviewers for their feedback, and Kate Huddlestone and Cédric Patin for their editorial assistance and patience. This work was supported by The Netherlands Organization for Scientific Research (NWO-VI.C.181.040).

References

Bauer, R.S. and P.K. Benedict. 1997. *Modern Cantonese Phonology*. Berlin: Mouton de Gruyter. https://doi.org/10.1515/9783110823707

Beckman, M.E. 1996. The parsing of prosody. *Language and Cognitive Processes* 11: 17-67. https://doi.org/10.1080/016909696387213

Bishop, J. and P. Keating. 2012. Perception of pitch location within a speaker's range: fundamental frequency, voice quality and speaker sex. *J. Acoust. Soc. Am. 132*: 1100-1112.

Bolinger D. 1978. Intonation across languages. In J.H. Greenberg, C.A. Ferguson and E.A. Moravcsik (eds.) *Universals of Human Language, vol 2: Phonology*. Palo Alto, CA: Stanford University Press. pp. 471–524.

Breen, M., E. Fedorenko, M. Wagner and E. Gibson. 2010. Acoustic correlates of information structure. *Language and Cognitive Processes* 25(7): 1044-1098. https://doi.org/10.1080/01690965.2010.504378

Brunelle, M., J. Kirby, A. Michaud and J. Watkins. 2020. Prosodic systems of Mainland Southeast Asia. In C. Gussenhoven and A. Chen (eds.). *The Oxford Handbook of Language*

Prosody. Oxford: Oxford University Press. pp. 344-354. https://doi.org/10.1093/oxfordhb/9780198832232.013.22

Chafe, W.L. 1974. Language and consciousness. *Language* 50(1): 111-133. https://doi.org/10.2307/412014

Chao, Y. 1968. *A Grammar of Spoken Chinese*. Berkeley: University of California Press.

Cheang, H.S. and M.D. Pell. 2008. "The sound of sarcasm". Speech Communication 50: 366–381.

Cheang, H.S. and M.D. Pell. 2009. Acoustic markers of sarcasm in Cantonese and English. *The Journal of the Acoustical Society of America* 126(3): 1394-1405. https://doi.org/10.1121/1.3177275

Cheang, H.S. and M.D. Pell, 2011. Recognizing sarcasm without language: A cross-linguistic study of English and Cantonese. *Pragmatics & Cognition* 19(2): 203-223. https://doi.org/10.1075/pc.19.2.02che

Chen, A., C. Gussenhoven and T. Rietveld. 2004. Language-specificity in the perception of paralinguistic intonational meaning. *Language and Speech* 47: 311–349. https://doi.org/10.1177/00238309040470040101

Chen, L., X. Li and Y. Yang. 2012. Focus, newness, and their combination: Processing of information structure in discourse. *PLOS ONE* 7(8): e42533. https://doi.org/10.1371/journal.pone.0042533

Chen, L., L. Wang and Y. Yang. 2014. Distinguish between focus and newness: An ERP study. *Journal of Neurolinguistics* 31: 28-41. https://doi.org/10.1016/j.jneuroling.2014.06.002

Chen, M. 2000. *Tone sandhi: Patterns across Chinese dialects*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511486364

Chen, S.-W., B. Wang and Y. Xu. 2009. Closely related languages, different ways of realizing focus. *Proceedings of Interspeech 2009*: 1007-1010. https://doi.org/10.21437/Interspeech.2009-298

Chen, Y. 2003. *The Phonetics and Phonology of Contrastive Focus in Standard Chinese*. PhD dissertation. Stony Brook: Stony Brook University.

Chen, Y. 2006. Durational adjustment under corrective focus in Standard Chinese. *Journal of Phonetics* 34: 176-201. https://doi.org/10.1016/j.wocn.2005.05.002

Chen, Y. 2008a. The acoustic realization of vowels of Shanghai Chinese. *Journal of Phonetics* 36: 629-648. https://doi.org/10.1016/j.wocn.2008.03.001

Chen, Y. 2008b. Revisiting the phonetics and phonology of Shanghai tone sandhi. In P.A. Barbosa, S. Madureira, and C. Reis (eds.), *Speech Prosody 2008,* Fourth International Conference, Campinas, Brazil, May 6-9, 2008. ISCA Archive. pp. 253-256.

Chen, Y. 2009. Prosody and information structure mapping: Evidence from Shanghai Chinese. *Chinese Journal of Phonetics* 2(1): 123-133.

Chen, Y. 2010. Post-focus *f0* compression - Now you see it, now you don't. *Journal of Phonetics* 38: 517-525. https://doi.org/10.1016/j.wocn.2010.06.004

Chen, Y. 2012. Message-related variation: Tonal variation. In A. Cohn, C. Fourgeron and M. Huffman (eds.), *Oxford Handbook of Laboratory Phonology*. Oxford: Oxford University Press. pp. 103-115.

Chen, Y. and B. Braun. 2006. Prosodic realization of information structure categories in Standard Chinese. In R. Hoffmann and H. Mixdorff (eds.), *Speech Prosody 2006,* Third International Conference, Dresden, Germany, May 2-5, 2006. ISCA Archive. paper 051.

Chen, Y. and C. Gussenhoven. 2008. Emphasis and tonal implementation in Standard Chinese. *Journal of Phonetics 36*(4): 724-746. https://doi.org/10.1016/j.wocn.2008.06.003

Chen, Y. and C. Gussenhoven. 2015. Shanghai Chinese. *Journal of the International Phonetic Association* 45: 321-337. https://doi.org/10.1017/S0025100315000043

Chen, Y., P. Lee, and H. Pan. 2016. Focus and topic marking in Chinese. In C. Féry and S. Ishihara (eds.), *Oxford Handbook of Information Structure*. Oxford: Oxford University Press. pp. 733–752. https://doi.org/10.1093/oxfordhb/9780199642670.013.34

Chen, Y. and Y. Xu. 2006. Production of weak elements in speech – Evidence from F0 patterns of neutral tone in standard Chinese. *Phonetica* 63: 47–75. https://doi.org/10.1159/000091406

Chen, Yiya, and Li Guo. 2022. Zhushan Mandarin. *Journal of the International Phonetic Association* 52 (2): 309–27.

Cheng, L.L. 1991. *On the Typology of WH-Questions*. PhD dissertation, MIT.

Cruttenden, A. 1981. Falls and rises: Meanings and universals. *Journal of Linguistics* 17(1): 77-91. https://doi.org/10.1017/S0022226700006782

Darwin, C. 1872. *The Expression of the Emotions in Man and Animals*. London: John Murray. https://doi.org/10.1037/10001-000

DiCanio, C. and R. Bennett. 2020. Prosody in Mesoamerican languages. In C. Gussenhoven and A. Chen (eds.). *The Oxford Handbook of Language Prosody*. Oxford: Oxford University Press. pp. 408-427. https://doi.org/10.1093/oxfordhb/9780198832232.013.25

Downing, L. and A. Rialland (eds.). 2017. *Intonation in African Tone Languages*. Berlin/Boston: De Gruyter Mouton. https://doi.org/10.1515/9783110503524

Duan, W., Y. Jia and Q. Ran. 2013. The common and different features of the prosodic realization of focus in Shandong dialects: Evidence from Jinan, Liaocheng, and Zibo dialects. *Journal of Tsinghua University (Natural Science Edition)* 53(6): 835-838.

Evans, J. 2008. 'African' tone in the Sinosphere. *Language and Linguistics* 9(3): 463-490.

Féry, C. 2016. *Intonation and Prosodic Structure*. Cambridge: Cambridge University Press. https://doi.org/10.1017/9781139022064

Frota, S. 2000. *Prosody and focus in European Portuguese. Phonological phrasing and intonation*. New York: Garland Publishing.

Gärding, E., J. Zhang and J. Svantesson. 1983. A generative model for tone and intonation in Standard Chinese based on data from one speaker. *Lund Working Papers* 25: 53-65.

Ge, C. and A. Li. 2020. Tonal variations in Honggu Chinese. *Proceedings of Speech Prosody 2020*: 245-249. https://doi.org/10.21437/SpeechProsody.2020-50

Grice, M., M. Savino and T. B. Roettger. 2018. Word final schwa is driven by intonation – the case of Bari Italian. *The Journal of the Acoustical Society of America (JASA)* 143: 2474-2486. https://doi.org/10.1121/1.5030923

Gryllia S., J. Doetjes, Y. Yang and L. L. Cheng. 2020. Prosody, clause typing, and wh-in-situ: Evidence from Mandarin. *Laboratory Phonology* 11(1): 19. https://doi.org/10.5334/labphon.169

Gu, W., K. Hirose and H. Fujisaki. 2005. Analysis of the effects of word emphasis and echo question on F0 contours of Cantonese utterances. *Proceedings of Interspeech 2005*, 556. https://doi.org/10.21437/Interspeech.2005-556

Gu, W., T. Zhang and H. Fujisaki. 2011. Prosodic analysis and perception of Mandarin utterances conveying attitudes. *Proceedings of Interspeech 2011*: 402. https://doi.org/10.21437/Interspeech.2011-402

Gunlogson, C. 2003. *True to form: Rising and falling declaratives and questions in English*. London: Routledge. https://doi.org/10.4324/9780203502013

Gussenhoven, C. 1992. Sentence accents and argument structure. In I. Roca (ed.), *Thematic Structure: Its Role in Grammar*. Berlin: Foris. pp. 79-106. https://doi.org/10.1515/9783110872613.79

Gussenhoven, C. 2004. *The Phonology of Tone and Intonation*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511616983

Gussenhoven, C. 2007. Types of focus in English. In C. Lee, M. Gordon and D. Büring (eds.), *Topic and Focus: Cross-linguistic Perspectives on Meaning and Intonation.* Heidelberg, New York, London: Springer. pp. 83-100. https://doi.org/10.1007/978-1-4020-4796-1_5

Gussenhoven, C. 2016. Foundations of intonational meaning: Anatomical and physiological factors. *Topics in Cognitive Science* 8(2): 425-434. https://doi.org/10.1111/tops.12197

Gussenhoven, C, Y. Chen, S. Frota and P. Prieto. 2013. Intonation. In M. Aronoff (ed.), *Oxford Bibliographies in Linguistics*. New York: Oxford University Press. https://doi.org/10.1093/obo/9780199772810-0072

Hirst, D., J.C. Wakefield, and H.T. Li. 2013. Does lexical tone restrict the paralinguistic use of pitch? Comparing melody metrics for English, French, Mandarin and Cantonese. *First International Conference on the Phonetics of Languages (ICPLC-2013)*, City University of Hong Kong.

Huang, C., S. Hsiao, H. Hwu, and S. Howng. 2012. The Chinese facial emotion recognition database (CFERD): A computer-generated 3-D paradigm to measure the recognition of facial emotional expressions at different intensities. *Psychiatry Research* 200 (2): 928–32. https://doi.org/10.1016/j.psychres.2012.03.038.

Hyman, L., H. Sande, F. Lionnet, N. Rolle and E. Clem. 2020. Prosodic systems of Sub-Saharan African. In C. Gussenhoven and A. Chen (eds.), *The Oxford Handbook of Language Prosody*. Oxford: Oxford University Press. pp. 183-194. https://doi.org/10.1093/oxfordhb/9780198832232.013.11

Jiang, P. and A. Chen. 2011. Representation of Mandarin intonation: Boundary tone revisited. In Z. Jing-Schmidt (ed.), *Proceedings of the 23rd North American Conference on Chinese Linguistics, Volume 1*. Ohio: Ohio State University. pp. 97-109.

Jin, S. 1996. *An Acoustic Study of Sentence Stress in Mandarin Chinese*. PhD dissertation. Columbus: OSU.

Kaufmann, D. and N.P. Himmelmann. 2020. Suprasegmental phonology. In A. Adelaar and A. Schapper (eds.), *The Oxford Guide to the Western Austronesian Languages*. Oxford: Oxford University Press.

Krifka, M. 2008. Basic notions of intonation structure. *Acta Linguistica Hungarica* 55: 243-276. https://doi.org/10.1556/ALing.55.2008.3-4.2

Kung, C., D. Chwilla and H. Schriefers. 2014. The interaction of lexical tone, intonation and semantic context in on-line spoken word recognition: An ERP study on Cantonese Chinese. *Neuropsychologia* 53: 293-309. https://doi.org/10.1016/j.neuropsychologia.2013.11.020

Ladd, R.D. 1980. *The Structure of Intonational Meaning: Evidence from English*. Bloomington: Indiana University Press.

Ladd, R.D. 2008. *Intonational Phonology* (Second Edition). Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511808814

Lee, O. J. 2005. *The Prosody of Questions in Beijing Mandarin*. PhD dissertation, Ohio State University.

Lee, Y.C., T. Wang, M. Liberman. 2016. Production and perception of T3 focus in Mandarin Chinese. *Frontiers Psychology*. https://doi.org/10.3389/fpsyg.2016.01058.

Li, A. 2002. Chinese prosody and prosodic labeling of spontaneous speech. In B. Bel and I. Marlien (eds.), *Proceedings of Speech Prosody 2002*. Aix-en-Provence: Laboratoire Parole et Langage. pp. 39-46.

Li, A., Q. Fang and J. Dang. 2011. Emotional intonation in a tone language: Experimental evidence from Chinese. *Proceedings of ICPhS XVII*: 1198-1201.

Li, A. and H. Wang. 2004. Friendly speech analysis and perception in Standard Chinese. *Proceedings of Interspeech 2004*: 897-900. https://doi.org/10.21437/Interspeech.2004-324

Li, B. 2006. *Chinese final particles and the syntax of the periphery*. PhD dissertation, Leiden University.

Li, C.N. and S.A. Thompson. 1981. *Mandarin Chinese: A Functional Reference Grammar.* Berkeley: California University Press.

Li, W. and Y. Yang. 2009. Perception of prosodic hierarchical boundaries in Mandarin Chinese sentences. *Neuroscience* 158: 1416. https://doi.org/10.1016/j.neuroscience.2008.10.065

Li, Q. and Y. Chen. 2019. Prosodically Conditioned Neutral-Tone Realization in Tianjin Mandarin. *Journal of East Asian Linguistics* 28(3): 211–42. https://doi.org/10.1007/s10831-019-09194-4.

Li, Q., Y. Chen, and Z. Xiong. 2019. Tianjin Mandarin. *Journal of the International Phonetic Association* 49 (1): 109–28. https://doi.org/10.1017/S0025100317000287.

Li, X., Y. Chen and Y. Yang. 2011. Immediate integration of different types of prosodic information during online spoken language comprehension: An ERP Study. *Brain Research* 1386: 139-152. https://doi.org/10.1016/j.brainres.2011.02.051

Lin, M. 2006. Interrogative vs. declarative and boundary tone. *Studies of the Chinese Language* 4:364–376.

Ling, B. and J. Liang. 2017. Focus encoding and prosodic structure in Shanghai Chinese. *The Journal of the Acoustical Society of America* 141(6): EL610-616. https://doi.org/10.1121/1.4989739

Liu, F. and Y. Xu. 2005. Parallel encoding of focus and interrogative meaning in Mandarin intonation. *Phonetica* 62: 70-87. https://doi.org/10.1159/000090090

Liu, M., Y. Chen and N. Schiller. 2016. Online processing of tone and intonation in Mandarin: Evidence from ERPs. *Neuropyschologia* 91: 307-317. https://doi.org/10.1016/j.neuropsychologia.2016.08.025

Liu, M., Y. Chen and N. Schiller. 2021. Context matters for tone and intonation processing in Mandarin. *Language and Speech* 65(1): 52-72. https://doi.org/10.1177/0023830920986174

Ma, J., V. Ciocca and T. Whitehill. 2006. Effect of intonation on Cantonese lexical tones. *The Journal of the Acoustical Society of America* 120: 3978-3987. https://doi.org/10.1177/0023830920986174

Ma, J., V. Ciocca and T. Whitehill. 2011. The perception of intonation questions and statements in Cantonese. *The Journal of the Acoustical Society of America* 129: 1012-1023. https://doi.org/10.1121/1.3531840

Maddieson, I. 2013. Tone. In M. Dryer and M. Haspelmath (eds.) *The World Atlas of Language Structures*. Leipzig: Max Planck Institute for Evolutionary Anthropology. Online: http://wals.info/chapter/13 (Accessed 2018-08-26)

Man, V. 2002. Focus effects on Cantonese tones: An acoustic study. *Proceedings of the first international conference on speech prosody*: 467-470.

Mueller-Liu, P. 2006. Signaling affect in Mandarin Chinese - the role of utterance-final non-lexical edge tones. *Proceedings of Speech Prosody 2006*, 048.

Norman, J. 1988. *Chinese*. New York: Cambridge University Press.

Ohala, J.J. 1984. An ethological perspective on common cross-language utilization of F0 of voice. *Phonetica* 41(1): 1-16. https://doi.org/10.1159/000261706

Ohala, J.J. 1994. The frequency code underlies the sound-symbolic use of voice pitch. In L. Hinton, J. Nichols and J.J. Ohala (eds.), *Sound symbolism*. Cambridge, UK: Cambridge University Press. pp. 325-347. https://doi.org/10.1017/CBO9780511751806.022

Ouyang, I. and E. Kaiser. 2013. Prosody and information structure in a tone language: an investigation of Mandarin Chinese. *Language and Cognitive Processes* 28: 1-16. https://doi.org/10.1080/01690965.2013.805795

Pan, H. 2007. Focus and Taiwanese unchecked tones. In C. Lee, M. Gordon and D. Büring (eds.), *Topic and Focus: Cross-linguistic Perspectives on Meaning and Intonation*. Dordrecht: Springer. pp. 195-213. https://doi.org/10.1007/978-1-4020-4796-1_11

Paulmann, S. and A. K. Uskul. 2014. Cross-cultural emotional prosody recognition: Evidence from Chinese and British listeners. *Cognition & Emotion* 28(2): 230-244. https://doi.org/10.1080/02699931.2013.812033

Pierrehumbert, J. 1980. *The Phonology and Phonetics of English Intonation*. PhD dissertation, MIT, Cambridge, MA.

Ren, G., Y. Yang and X. Li. 2009. Early cortical processing of linguistic pitch patterns as revealed by the mismatch negativity. *Neuroscience* 162: 87-95. https://doi.org/10.1016/j.neuroscience.2009.04.021

Ren, G., Y. Tang, X. Li and X. Sui. 2013. Pre-attentive processing of Mandarin tone and intonation: Evidence from event-related potentials. In F. Signorelli and D. Chirchiglia (eds.), *Functional Brain Mapping and the Endeavor to Understand the Working Brain.* London: IntechOpen. https://doi.org/10.5772/56503

Rialland, A. 2009. The African lax question prosody: Its realisation and geographical distribution. *Lingua* 119(6): 928-949. https://doi.org/10.1016/j.lingua.2007.09.014

Rooth, M. 1996. Focus. In S. Lappin (ed.), *The Handbook of Contemporary Semantic Theory*. Oxford: Blackwell. pp. 271-297. https://doi.org/10.1111/b.9780631207498.1997.00013.x

Selkirk, E. 1995. Sentence prosody: intonation, stress, and phrasing. In J. Goldsmith (ed.), *The Handbook of Phonological Theory*. Oxford: Blackwell Publishing. pp. 550–569.

Shattuck-Hufnagel, S. and A. E. Turk. 1996. A prosody tutorial for investigators of auditory sentence processing. *Journal of Psycholinguistic Research* 25: 193-247. https://doi.org/10.1111/b.9780631207498.1997.00013.x

Scholz, F. 2012. *Tone sandhi, Prosodic Phrasing, and Focus Marking in Wenzhou Chinese*. PhD dissertation, Leiden University.

Scholz, F. and Y. Chen. 2014. The independent effects of prosodic structure and information status on tonal coarticulation: Evidence from Wenzhou Chinese. In J. Caspers, Y. Chen, W. Heeren, J. Pacilly, N.O. Schiller and E. van Zanten (eds.), *Above and Beyond the Segments: Experimental Linguistics and Phonetics*. Amsterdam: John Benjamins. pp. 275-287. https://doi.org/10.1075/z.189.22sch

Shen, C. and Y. Xu. 2016. Prosodic focus with post-focus compression in Lan-yin Mandarin. *Proceedings of Speech Prosody 2016*: 340-344. https://doi.org/10.21437/SpeechProsody.2016-70

Shih, C. 1988. Tone and intonation in Mandarin. *Working Papers of the Cornell Phonetics Laboratory 3: Stress, Tone and Intonation*. Ithaca: Cornell University. pp. 83-109.

Snider, K. 2018. *Tone Analysis for Field Linguists*. Texas: SIL International.

Speer, S., C. Shih, and M.L. Slowiaczek. 1989. Prosodic structure in language understanding: evidence from tone sandhi in Mandarin. *Language and Speech* 32: 337-354. https://doi.org/10.1177/002383098903200403

Wang, B. and Y. Xu. 2011. Differential prosodic encoding of topic and focus in sentence-initial position in Mandarin Chinese. *Journal of Phonetics* 37: 502-520. https://doi.org/10.1016/j.wocn.2011.03.006

Wang, T. and Y.C. Lee. 2015. Does restriction of pitch variation affect the perception of vocal emotions in Mandarin Chinese? *The Journal of the Acoustical Society of America*, 137(1): EL117-123. https://doi.org/10.1121/1.4904916

Warren, P. 2016. *Uptalk: The Phenomenon of Rising Intonation*. Cambridge University Press.

Wichmann, A. 2000. The attitudinal effects of prosody and how they relate to emotion. *Proceedings of the ISCA Workshop on Speech and Emotion*: 143–148.

Wu, W. L. and Y. Xu. 2010. Prosodic focus in Hong Kong Cantonese without post-focus compression. *Proceedings of Speech Prosody 2010*, 040.

Wurm, S. A., R. Li, T. Baumann and M. W. Lee. 1987. *Language Atlas of China*. Hong Kong: Longman.

Xu, B. and P. Mok. 2011. Final rising and global raising in Cantonese intonation. *Proceedings of the 17th International Congress of Phonetic Sciences*: 2173-2176.

Xu, B. and Z. Tang. 1988. *Shanghai Shiqu Fangyanzhi* [A Description of Shanghainese Spoken in the Urban Districts of the Shanghai City]. Shanghai: Shanghai Education Press.

Xu, L. 2006. Topicalization in Asian languages. In M. Everaert and H. van Riemsdijk (eds.), *The Blackwell Companion to Syntax*. Oxford: Blackwell. pp. 137-174. https://doi.org/10.1002/9780470996591.ch73

Xu, Y. 1999. Effects of tone and focus on the formation and alignment of *f0* contours. *Journal of Phonetics* 27: 55-105. https://doi.org/10.1006/jpho.1999.0086

Xu, Y. 2005. Speech melody as articulatorily implemented communicative functions. *Speech Communication* 46: 220-251. https://doi.org/10.1016/j.specom.2005.02.014

Xu, Y., S. Chen, and B. Wang. 2012. Prosodic focus with and without post-focus compression: A typological divide within the same language family? *The Linguistic Review* 29(1): 131-147. https://doi.org/10.1515/tlr-2012-0006

Xu, Y. and C.X. Xu. 2005. Phonetic realization of focus in English declarative intonation. *Journal of Phonetics* 33: 159-197. https://doi.org/10.1016/j.wocn.2004.11.001

Yang Y., S. Gryllia and L.L. Cheng. 2020. Wh-question or wh-declarative? Prosody makes the difference. *Speech Communication* 118: 21-32. https://doi.org/10.1016/j.specom.2020.02.002

Yip, M. 2002. *Tone*. Cambridge: Cambridge University Press.

Yu, A. 2007. Understanding near mergers: The case of morphological tone in Cantonese. *Phonology* 24(1): 187-214.

Yuan, J. 2004. *Intonation in Mandarin Chinese: Acoustics, Perception, and Computational modeling*. PhD dissertation, Cornell University.

Yuan, J. 2011. Perception of intonation in Mandarin Chinese. *The Journal of the Acoustical Society of America* 130: 4063-4069. https://doi.org/10.1121/1.3651818

Yuan, J. and Y. Chen. 2014. 3rd tone sandhi in Standard Chinese: A corpus approach. *Journal of Chinese Linguistics* 42: 218-237.

Yuan, J., L. Shen and F. Chen. 2002. The acoustic realization of anger, fear, joy and sadness in Chinese. *Proceedings of ICSLP 2002*, 556. https://doi.org/10.21437/ICSLP.2002-556

Zahner-Ritter, K., Y. Chen, N. Dehé, and B. Braun. 2022. The prosodic marking of rhetorical questions in Standard Chinese. *Journal of Phonetics* 95: 101190. https://doi.org/10.1016/j.wocn.2022.101190.

Zhang, C. 2018. *Tones and Tunes in Tianjin Mandarin*. PhD thesis. University of Oxford.

Zhang, J., S. Duanmu, and Y. Chen. 2020. Prosodic systems of Cihina and Siberia. In C. Gussenhoven, C. and A. Chen (eds.). *The Oxford Handbook of Language Prosody*. Oxford: Oxford University Press. pp. 332-344. https://doi.org/10.1093/oxfordhb/9780198832232.013.21