The interaction between tone and prosodic focus in Mandarin Chinese

Ting Wang,[1] Jun Liu,[2,3] Yong-hun Lee[4] and Yong-cheol Lee[3]
[1] Tongji University | [2] Liaocheng University | [3] Cheongju University |
[4] Chungnam National University

This study characterized focused tones in Mandarin Chinese through a production experiment using phone number strings. The results revealed that, although phonation cues had little effect on any focused tone, prosodic cues exhibited various patterns of distribution. Duration played an important role for each focused tone, but intensity had a relatively less salient role. Among pitch-related parameters, the raising of pitch register was an important cue when a level tone (tone 1) was focused. By contrast, due to the interaction between tone and intonation, absolute slope and pitch range had less effect on tone 1 focus. These cues, however, were prominent when contour tones (tones 2 and 4) were in focus. Unlike other focused tones that raised pitch, tone 3 focus exhibited the opposite pattern, lowering its pitch target. In the aggregation of all focused tones, it was found that only primarily pitch-related parameters were selected as the main variables discriminating one from another. The results of this study, therefore, suggest that the prosodic marking of focus is not uniform, even within a single language, but clearly differs by tone type. Accordingly, prosodic marking of focus should be considered multimodal in a tonal language.

Keywords: prosody, corrective focus, tone, Mandarin Chinese, multimodal

1. Introduction

During conversation, focus – a communicative function – highlights a particularly important piece of information within a sentence (Ladd 1984; Xu & Xu 2005). Due to its significance in discourse, focus attracts prosodic prominence. Conventional wisdom on the phonetic universal of focus holds that a word under focus exhibits greater duration, intensity, and pitch range than a word without focus (Cooper et al. 1985; Xu 1999; Lee & Xu 2010). However, focus has also

https://doi.org/10.1075/lali.00063.wan | Published online: 2 April 2020
Language and Linguistics 21:2 (2020), pp. 331–350. issn 1606-822x | e-issn 2309-5067

been documented to perform differently in different languages, according to a language's prosodic structure (Lee 2015). This is especially true in tonal languages, such as Mandarin Chinese. As a core part of the study, we first describe tone and intonation in Mandarin Chinese, followed by examining the effects of prosodic focus observed in this language.

1.1 Tone and intonation in Mandarin Chinese

There are four lexical tones in Mandarin Chinese: high (tone 1), rising (tone 2), low/dipping (tone 3), and falling (tone 4). Tones 1, 2, 3, and 4 are conventionally labeled [55], [35], [214], and [51], respectively (Chao 1968), where the numbers from [5] to [1] refer to pitch levels from high to low. Of the four, tone 3 in Mandarin undergoes tone sandhi, becoming tone 2 when it is followed by another tone 3. In addition, the dipping version (i.e. [214]) of tone 3 only occurs in citation forms or at boundary positions, but when in connected speech, its final rise is entirely absent (Xu 2004). The four tones can be further classified according to whether their pitch targets are static or dynamic (Xu & Wang 2001). Tones 1 and 3 have a static pitch target, and tones 2 and 4 have a dynamic one. Furthermore, the initial portion of each tone is affected by the final portion of the preceding tone, so the underlying pitch levels for each tone only appear sometime after the midpoint of the tone. These alterations are known as CARRYOVER EFFECTS (Xu 1997).

In Mandarin Chinese, lexical tones are integrated into sentence intonation. Chao (1968: 39) describes the interaction between tone and intonation as "small ripples riding on large waves." The small ripples are the lexical tones, and the large waves are sentence intonation. This description views intonation as independent, although it inextricably interacts with tone at the local level of pitch contour. A different description can be found in Gårding et al. (1983), who depicts Mandarin intonation patterns as grids with different directions and ranges. Lexical tones are represented by local pitch contours between the upper and lower grids, whereas sentence intonation is marked by rising and falling patterns at a global level, as well as changes within the range of the grid. Likewise, Shen (1985) claims that intonation in Mandarin Chinese is characterized by two lines, top and bottom, which have different prosodic functions. The top line signals semantically related prominence within a sentence, represented by pitch accent. When a word is in focus within a sentence, the focused word is expressed by a raised top line. The bottom line marks the boundary of a prosodic or rhythmic unit. A larger prosodic unit shows a lower bottom line at the boundary position than a smaller one does, meaning that the larger the prosodic unit, the lower the bottom line is towards the boundary position.

1.2 Effects of prosodic focus in Mandarin Chinese

Previous studies have shown important features about the prosodic correlates of focus in Mandarin Chinese. Instead of changes in the shape of pitch contours, as in English, prosodic focus expands pitch range in Mandarin to maintain the tonal structure (Ouyang & Kaiser 2015). Focus exhibits different characteristics, in particular conforming to Mandarin Chinese's tonal system (Lee et al. 2016). A high pitch target is raised when tones 1, 2, or 4 are in focus, whereas a low pitch target is lowered when tone 3 is in focus (Xu 1999). Besides pitch range variation, increased duration and intensity play important roles in marking prosodic focus (Shih 1988; Chen & Gussenhoven 2008), but they are not considered as significant as pitch range movement (Cao 2002; Xu 2005).

Several researchers have attempted to account for how pitch range is altered in marking prosodic focus in Mandarin Chinese. Chao (1968) uses the image of an ELASTIC TRANSPARENT SHEET to illustrate pitch range variation. When a word is in focus, the ELASTIC TRANSPARENT SHEET is stretched vertically to vary the pitch range of the target syllable. In a similar way, the ELASTIC TRANSPARENT SHEET is stretched horizontally to lengthen the focused syllable. Shen's (1985) model, which exhibits two lines – top and bottom – represents prosodic focus by modifying the top line of pitch contour, which complements Shih's (1988) viewpoint. The work of Wu (2001) and Cao (2002) views prosodic focus as realized by the manipulation of pitch register. The pitch register is raised or lowered for focused syllables, depending on the type of tone. Under focus, a high tone is higher, but a low tone is lower.

In addition to the pitch range expansion of a syllable under focus, it is also widely known that prosodic focus is followed by post-focus compression, which was discovered by Xu and his colleagues and published in several previous works (e.g., Xu et al. 2012). The pitch ranges of post-focus syllables are significantly compressed, relative to the same syllables uttered with no prosodic focus. This phenomenon has also been observed in many other languages, such as English (Xu 2005), Japanese (Lee & Xu 2012), and Korean (Lee & Xu 2010).

1.3 The current study

Although previous studies have produced important findings about prosodic focus in Mandarin Chinese, the understanding of its precise nature remains far from complete. One of many reasons for this relates to several uncertainties that remain to be clarified. Lee et al. (2016) demonstrate different levels of pitch range expansion for each focused tone: a tone 4 focus produced the greatest level of pitch range expansion, followed by tone 2, tone 1, and tone 3, in descending order.

However, it is not yet clear which parameter compensates for the smallest level of pitch range expansion for tone 3 focus. It is also unclear which acoustic parameter, among many, plays the most important role for each focused tone. To put it differently: do acoustic parameters have a similar effect on the focused tone, or are other parameters more important?

Another factor that may hinder the creation of a clear picture of prosodic focus in Mandarin Chinese is that the acoustic parameters tested in previous studies seem insufficient to characterize the focused tone, even though prosody provides a fluent repertoire of linguistic functions (Xu 2005; Wang & Lee 2015). As a function of focus, three prosodic cues (duration, intensity, and pitch-related features) are generally measured (Ouyang & Kaiser 2015). Many studies have included such prosodic cues (inter alios, Chen & Gussenhoven 2008; Ouyang & Kaiser 2015; Lee et al. 2016), and some studies have included creakiness, a phonation cue, in examination of tone 3 focus (Cao & Zhang 2008; Chen & Gussenhoven 2008). One potential problem with previous research is that it appears to investigate primarily whether each dependent variable tested is significant in marking prosodic focus. There still remain some noticeable gaps: (i) determining the relative importance of each variable in encoding prosodic focus for each focused tone, and (ii) doing the same task by aggregating all focused tones.

The expression of prosodic focus varies with tones in Mandarin Chinese. In other words, when tone 1 is in focus, pitch does not fluctuate much because tone 1 has a static pitch target. When tone 2 is in focus, the pitch scales from mid to high. When tone 3 is in focus, the pitch is scaled toward a low pitch target, meaning that it begins high at the onset of tone 3 focus and becomes lower toward the offset. When tone 4 is in focus, the pitch scale is from high to low. Unlike other individual cues, such as intensity and duration, that change independently under focus, slope does not change independently, and it conveys information about dynamic changes in pitch range expansion (vertically) and duration (horizontally) at the same time. Slope must, therefore, be included as a function of focus to achieve more finely grained pitch-scaling information for each focused tone.

This study is intended to improve the understanding of prosodic focus in Mandarin Chinese by quantifying the prosodic characteristics of focus and determining whether the four lexical tones behave similarly or differently when focused. We hypothesize that, as stated above, the prosodic realizations of focus differ, conforming to the structures of individual tones in the language. Certain acoustic parameters would thus be of greater importance than others for each focused tone. This hypothesis motivates us to rank the acoustic parameters contributing to the prosodic characteristics of each focused tone. We also attempted to determine the relative importance of each acoustic parameter in marking prosodic focus, aggregating all the focused tones. To test the hypothesis, we examined eight acoustic cues (duration, intensity, maximum pitch, minimum pitch, pitch range, mean absolute slope, cepstral peak prominence, and H1–H2), ranked by conducting a random forest analysis. Phonation cues, such as cepstral peak prominence and H1–H2, were included because creakiness is thought to be a crucial property that distinguishes tone 3 focus from other focused tones (Zheng 2006; Chen & Gussenhoven 2008; Yin & Kong 2010). This study was limited to focus positions, though focus entails different prosodic adjustments in the pre- and/or post-focus positions. Furthermore, focus can be used differently in different contexts; this study elicited corrective focus by correcting an incorrectly given digit in a digit string. This method has the advantage that prosodic focus can be placed in any position in the string (Lee et al. 2015).

2. Method

2.1 Stimuli

We used a Python script to create 100 randomized 10-digit phone number strings, grouped as (NNN)-(NNN)-(NNNN), in the style of American phone numbers. The design of the 100 strings was based on two criteria: each digit (i.e. 0 through 9) appeared equally often at each string position and each pair of digits (i.e. 0–1, 1–2, …, 8–9) appeared equally often across each pair of string positions. Please note that the numbers (0 through 9) can be classified into four groups by tone: 1, 3, 7, 8 (tone 1); 0 (tone 2); 5, 9 (tone 3); and 2, 4, 6 (tone 4).

Speakers were instructed to read out the 100 randomized target stimuli in isolation for broad focus as background reading, as in (1), below. To elicit corrective focus on a single digit, the same target sequences were embedded in a Q&A structure, as in (2), below. In the Q&A structure, pre-recorded prompts asked whether a given phone number is correct, and the speakers answered by giving the phone number with one corrected digit, as indicated below by underlined boldface type. This experimental design enabled us to directly compare a target digit between broad-focus and corrective-focus conditions. Broad focus here indicates a condition in which no single element receives narrow focus within a sentence, while corrective focus refers to a condition in which a certain element receives narrow focus in the correction of wrong information in the previous sentence (Gussenhoven 2008). It should be noted that no focused tone 3 digit ever appeared before another tone 3 digit in our stimuli, meaning that this study completely excluded tone sandhi.

(1) Broad-focus condition
李梅的電話號碼是267-800-8717。
'Li Mei's number is 267-800-8717.'

(2) Corrective-focus condition
Q: 李梅的電話號碼是367-800-8717。是嗎?
'Li Mei's number is 367-800-8717. Right?'
A: 不是。李梅的電話號碼是<u>2</u>67-800-8717。
'No, Li Mei's number is <u>2</u>67-800-8717.'

2.2 Subjects

In all, 10 speakers participated in this study. Of these, three females and two males (of the respective ages 26, 26, 24, 24, 26 years; mean age: 25.2 years, SD: 1.1) were recruited at the University of Pennsylvania in Philadelphia, USA. Two females and three males (of the respective ages 33, 39, 39, 28, 22 years; mean age: 32.2 years, SD: 7.3) were recruited at Cheongju University in Cheongju, South Korea. All participants were native speakers of Mandarin Chinese and reported neither speech nor hearing disorders. After the experiment, each participant recruited in the US received 10 dollars, and each one recruited in South Korea received 10,000 won (approximately 10 dollars) for taking part in the experiment.

2.3 Recording procedure

Stimuli were recorded in sound-attenuated booths at the University of Pennsylvania and Cheongju University. Recordings were made using PRAAT software at a sampling frequency of 44,100 and were saved directly onto the hard disk drive of the recording laptop. Each speaker sat comfortably in front of the screen of the laptop and wore a head-mounted microphone. The same laptop and microphone were used for recordings in both locations. The speech materials were presented in the center of the laptop screen on PowerPoint slides. Before the recordings began, the speakers performed three sample trials for each focus condition to become familiar with the stimuli and the recording procedure. Speakers were instructed to read target stimuli for broad focus as naturally as possible. For the corrective-focus condition, they were instructed first to listen to a pre-recorded question and provide a correct answer with one corrected digit. During the actual recordings, the broad-focus conditions were always recorded before the corrective-focus ones for the speakers recruited at the University of Pennsylvania. To counterbalance the effect of order, however, corrective-focus recordings always preceded broad-focus ones for the speakers recruited at Cheongju University. A

5-minute break was given to each speaker during the transition between the two focus conditions. The entire duration of the recording time was about 50 minutes for each speaker.

The production experiments conducted in the two locations produced a total of 2,000 10-digit phone number strings (100 strings × 10 speakers × 2 focus conditions). Grouped by lexical tone, our recordings included: 800 digit strings for tone 1; 200 digit strings for tone 2; 400 digit strings for tone 3; and 600 strings for tone 4. This breakdown comes from the fact that, as noted above, there are four tone 1 digits (1, 3, 7, 8), one tone 2 digit (0), two tone 3 digits (5, 9), and three tone 4 digits (2, 4, 6) in Mandarin Chinese.

2.4 Overview of pitch contours

We looked at time-normalized pitch contours to determine the general picture for prosodic marking of corrective focus in Mandarin Chinese. In this study, we manually labeled every digit in every digit string, and we obtained time-normalized pitch contours by measuring and averaging 10 equidistant pitch points for every digit in every digit string, using ProsodyPro (Xu 2013). Figure 1 illustrates the pitch contours for each focused tone, averaged by all relevant string positions spoken by 10 speakers in the two focus conditions (broad vs. corrective). We observed that, as established by previous work, corrective focus was manifested by lowering the low pitch point for tone 3 focus and raising the high pitch point for the other focused tones. Although all tone types displayed a similar expansion of pitch range as a function of corrective focus, the different focused tones featured several other characteristics. First, pitch range expansion began at different time points: at point 1 for tones 1 and 4 foci, but not until point 7 for tone 2 focus and not until point 5 for tone 3 focus. Next, a closer examination enables us to observe that the size of pitch range expansion differs by tone type: tone 4 focus exhibited the greatest pitch range expansion; tone 3 focus shows the least; and tones 1 and 2 foci fall in between. We found the steepest slopes for tones 2 and 4 foci and a gradual slope for tone 3 focus. It seems, however, that pitch contour was virtually flat for tone 1 focus. These different characteristics for different focused tones suggested that prosodic marking of corrective focus is not universal, even within a single language. Instead, it should be considered multimodal, particularly in a tone language.

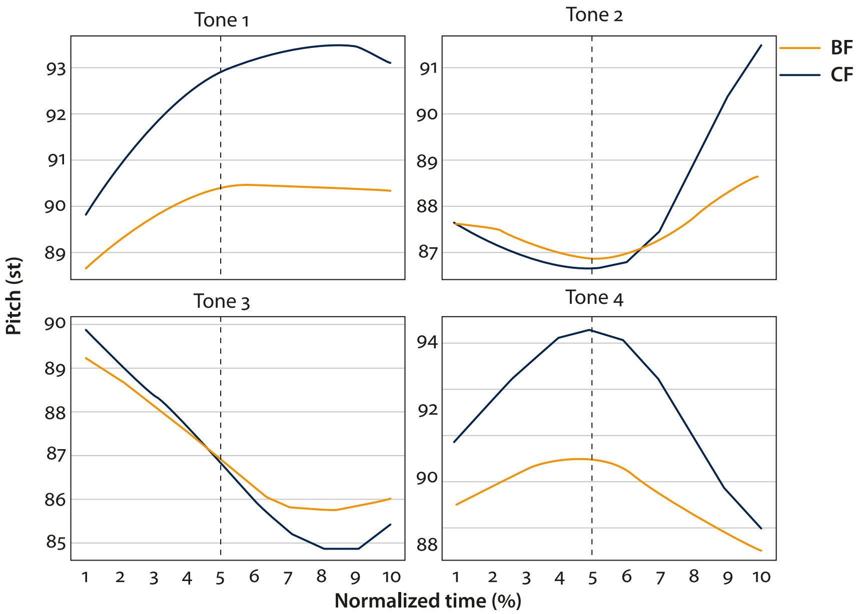

Figure 1. Time-normalized pitch contours of tones 1–4 in broad focus (BF) and corrective focus (CF)

2.5 Acoustic measurements

A visual inspection of Figure 1 shows that the following pitch-related parameters were measured: maximum pitch (MaxP), minimum pitch (MinP), pitch range (RNG), calculated by the difference between MaxP and MinP, and mean absolute slope (MAS), measured by two adjacent pitch values in distances among the normalized time points. For these parameters, we first obtained pitch values in Hertz (Hz) and then converted them into semitones (st), using the following formula: $st = 12\ log_2\ Hz$ (Lee 2017; Lee et al. 2018; Xu & Wang 2009). We also included duration (Dur) in milliseconds (ms) and mean intensity (Int) in decibels (dB) to measure the function of corrective focus. Finally, since tone 3 focus is often distinguished by creakiness, we measured two phonation cues: cepstral peak prominence (CPP), representing harmonics-to-noise ratio, and the amplitude difference between the first and second harmonics (H1H2). Following Keating et al. (2015), we expected CPP and H1H2 to be lower in creaky voice, resulting in pitch lowering.

Before moving on, we should note that, as discussed earlier, the initial portion of each tone is influenced by the final portion of the preceding tone, the CARRY-OVER EFFECTS (Xu 1997). In this study, therefore, we included only the portions of each target syllable after point 5 – indicated by an orange line in Figure 1 – in order to best approximate the pitch values for each tone. We thus computed the following measures related to pitch after point 5: MaxP, MinP, RNG, MAS, CPP, and H1H2.

3. Analyses and results

Our simple analysis created a direct comparison between the broad and corrective-focus conditions using the aggregates of the eight prosodic and phonation measures. For statistical analysis, we conducted a linear mixed model with the *lmerTest* package (Kuznetsova et al. 2013) in R (R Core Team 2017). In this model, we treated focus (broad, corrective), tone (tones 1–4), and their interaction as fixed effects; speaker (10 speakers), digit (each tone's different digits, except tone 2), and string position (positions 1–10) as random effects; and the eight acoustic parameters as dependent variables. To determine the significance of each fixed effect and their interaction, we used the ANOVA function of the *lmerTest* package. Furthermore, given that each focused tone reveals unique characteristics for prosodic marking of corrective focus, we posited that different acoustic parameters would feature each focused tone differently. To identify the exact nature of each focused tone, we performed a random forest analysis implementing the *randomForest* package (Liaw & Wiener 2002) in R. This method classifies the most important factors contributing to the differences between broad and corrective-focus conditions for each focused tone.

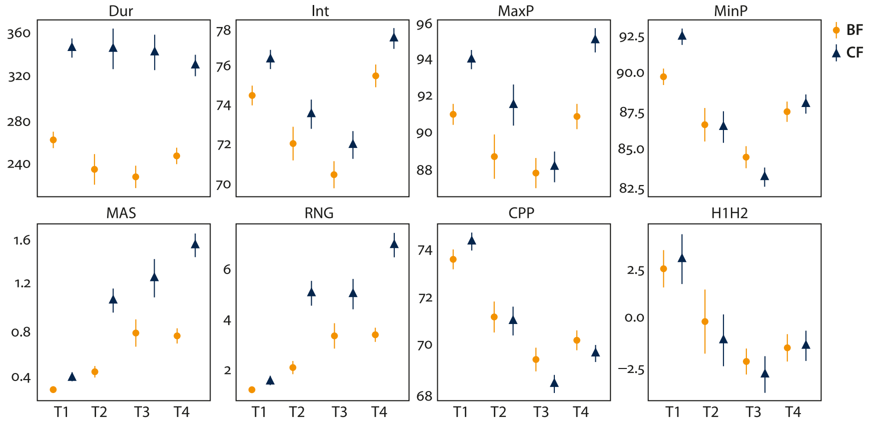

Figure 2. Means and 95% confidence intervals for each acoustic parameter (T1: tone 1, T2: tone 2, T3: tone 3, T4: tone 4; BF: broad focus, CF: corrective focus). Points and error bars indicate means and confidence intervals, respectively

Figure 2 gives the aggregate means and 95% confidence intervals of the eight acoustic parameters in the two focus conditions, separated by tone type. At first glance, this figure appears to show that broad-focus and corrective-focus conditions yielded different values for all acoustic parameters. For example, corrective-focus conditions showed clearly increased values for Dur and Int unlike their broad-focus counterparts. Upon closer examination, however, with the exception of Dur and Int, different focused tones exhibited different patterns of distribution for other parameters. For MaxP, focused tones, with the exception of tone 3 focus, produced increased pitch values in marking corrective focus. For MinP, tone 1 focus produced a greater pitch value, but tone 3 focus showed the opposite pattern. Tones 2 and 4 foci exhibited minimal differences. Turning to MAS and RNG, tones 2–4 foci showed greater values in corrective focus than in broad focus, whereas tone 1 focus featured a negligible difference for these parameters. For the next two parameters, CPP and H1H2, the directions of the differences were quite mixed, so their interpretations did not seem straightforward.

To confirm our visual observation from Figure 2, we turned to the results of the ANOVA with focus as a fixed effect in Table 1, in which boldface type and shaded cells refer to insignificance. Please note that this statistical test was conducted using tone type. We therefore here present the results one by one, from tone 1 to tone 4. For tone 1, focus had a significant effect on all of the parameters, with the exception of H1H2. For tone 2, the effect of focus was significant for Dur, Int, MaxP, MAS, and RNG, but its effect on MinP, CPP, and H1H2 did not reach significance. For tone 3, the effect of focus on the majority of parameters was significant, with the exceptions of MaxP and H1H2. Finally, for tone 4, focus had a significant effect on all prosodic cues, but not on any phonation cues.

Table 1. The effect of focus on each acoustic parameter ($df = 1$)

	Tone 1		Tone 2		Tone 3		Tone 4	
	X^2	*p*-value	X^2	*p*-value	X^2	*p*-value	X^2	*p*-value
Dur	377.57	< .001	124.67	< .001	247.9	< .001	262.87	< .001
Int	90.14	< .001	23.05	< .001	28.15	< .001	80.19	< .001
MaxP	609.36	< .001	81.91	< .001	1.37	**= .24**	445.59	< .001
MinP	410.91	< .001	0.62	**= 0.43**	16.83	< .001	10.88	< .001
MAS	45.18	< .001	98.97	< .001	32.41	< .001	180.86	< .001
RNG	20.45	< .001	97.29	< .001	27.74	< .001	167.59	< .001
CPP	13.14	< .001	0.30	**= .59**	9.21	< .01	3.53	**= .06**
H1H2	2.06	**= .15**	1.25	**= .26**	0.70	**= 0.40**	0.01	**= .94**

The ANOVA results suggested that different focused tones did not exhibit similar trends in marking corrective focus prosodically. Put another way, each focused tone employed eight acoustic parameters quite distinctly for prosodic marking of corrective focus. This motivates us to conduct a two-way analysis of the interaction of tone and focus to identify whether a significant interaction effect could be found for each acoustic parameter. Table 2 shows that the interaction between tone and focus had a significant effect on all parameters. This means that prosodic marking of corrective focus varied according to differences in focused tones, as the acoustic parameters tested here were manifested differently by tone type.

Table 2. The interaction effect of tone and focus on each acoustic parameter ($df = 7$)

	Tone × Focus	
	X^2	*p*-value
Dur	1003.40	< .001
Int	202.20	< .001
MaxP	967.12	< .001
MinP	96.85	< .001
MAS	428.16	< .001
RNG	430.77	< .001
CPP	36.50	< .001
H1H2	23.73	< .001

The results of the random forest analysis, which are shown in Figure 3, identified the most important factors contributing to the differences in each tone under study with corrective focus and the same tone with broad focus in normal utterance. As is clear from Figure 3, MaxP was selected as the most important cue for tone 1, distinguishing corrective from broad focus, followed by (in descending order) MinP, Dur, Int, MAS, H1H2, RNG, and CPP. Next, Dur functioned as the most important parameter in tone 2 focus, followed by MAS, RNG, MaxP, Int, CPP, MinP, and H1H2. For tone 3, Dur was found to be what critically distinguished broad from corrective focus, followed by RNG, Int, MAS, MinP, CPP, MaxP, and H1H2. Last, for tone 4, MaxP was chosen as the crucial determinant of the differences between the two focus conditions. The remaining parameters fell into the order of Dur, MAS, RNG, MinP, Int, H1H2, and CPP, from second most important to least. The random forest analysis clearly indicated that speakers of Mandarin Chinese used different parameters in the production of corrective focus, depending on tone type.

The results of the ANOVA and the random forest analysis must be compared. Table 1 and Figure 3 show that the results of the two methods generally agreed,

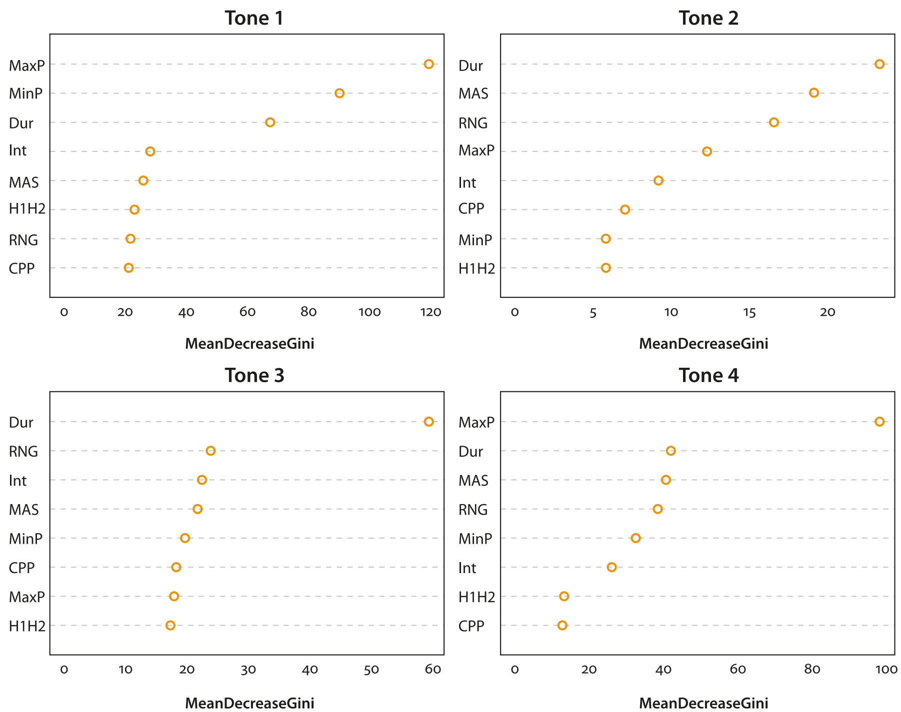

Figure 3. Results of random forest analysis with highest mean decrease in Gini, stratified by tone

increasing the reliability and validity of the statistical analyses. For example, H1H2 did not function well in distinguishing broad from corrective focus in all tone groups. We observed that H1H2 was ranked fairly low, showing a low Gini value for each tone. Likewise, MinP and CP did not have significant differences between the two focus conditions in tone 2, and thus they were ranked low in the random forest analysis for tone 2. MaxP for tone 3 and CPP for tone 4 exhibited similar patterns.

Finally, it appeared that the different rankings shown in Figure 2 were related to the different phonological properties of the four lexical tones. We thus aggregated all tone types and identified the relative importance of each variable in marking corrective focus. As Figure 4 illustrates, RNG, MaxP, MinP, and MAS were selected as the most important variables in distinguishing all focused tones in Mandarin Chinese. Note that these parameters are all pitch-related. On the other hand, other parameters not related to pitch, such as H1H2, Dur, CPP, and Int, were chosen as (relatively) less important cues that can be used to discriminate all focused tones from one another. The fact that only primarily pitch-related parameters contribute to differences in marking corrective focus for all tone types is discussed below.

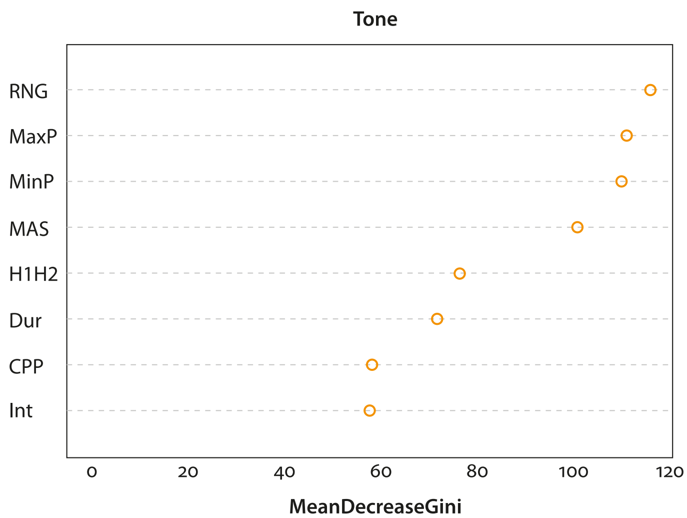

Figure 4. Random forest analysis results with highest mean decrease in Gini, aggregated by all tones

4. Discussion and conclusion

This study explored interactions between tone and corrective focus in Mandarin Chinese, using the experimental paradigm of 10-digit phone number strings. We hypothesized that prosodic marking of corrective focus would be different for different tones. Our experimental design enabled us to conduct a systematic investigation of interactions between tone and corrective focus. The results indicated that prosodic marking of corrective focus was not completely represented by one acoustic dimension alone (e.g., MaxP) but rather by dynamic changes in several dimensions (e.g., RNG, MaxP, Dur). Since even the same dimension contributed differently to the prosodic marking of corrective focus for different tones in Mandarin Chinese, the prosodic realizations of corrective focus appear to be complex. That is, the dynamic system is signaled by different acoustical cues in multidimensional space. We call this mechanism multimodal. The results of our experiment supported our hypothesis, namely, that prosodic marking of corrective focus in Mandarin Chinese is multimodal, due to the presence of lexical tone.

The ANOVA results indicated that focused tones exhibited different patterns of prosodic marking of corrective focus. Furthermore, the random forest analysis provided rankings of the acoustic parameters contributing to the differences between broad and corrective focus. Here we select the four most important acoustic cues in each focused tone to see the multimodality of prosodic marking of corrective focus in Mandarin Chinese, where the > symbol illustrates descending order.

Tone 1: MaxP > MinP > Dur > Int
Tone 2: Dur > MAS > RNG > MaxP
Tone 3: Dur > RNG > Int > MAS
Tone 4: MaxP > Dur > MAS > RNG

The varied rankings for each focused tone indicate that prosodic marking of corrective focus was found to differ by tone type.

These results may prompt one to ask why prosodic marking of corrective focus varies by tone type in Mandarin Chinese. We posit that this is because Mandarin Chinese is a tone language, indicating that prosodic modulation by corrective focus is expressed by conformity to prosodic structure. Therefore, as shown in Figure 4, primarily pitch-related parameters played crucial roles in distinguishing among the focused tones in this language. To be more specific, tone 1 has a level pitch target, meaning that MAS and RNG are suppressed due to tone and intonation interaction. By contrast, the entire pitch register was raised for the prosodic marking of corrective focus; that is, MinP and MaxP were increased throughout the entire region of tone 1 focus. By contrast, for tones 2 and 4, which have a dynamic pitch target, instead of increasing the register of the whole pitch, broad focus and corrective focus most likely differed toward the high pitch point of these focused tones. As a result, MaxP, MAS, and RNG played crucial roles in distinguishing broad focus from corrective focus. Finally, tone 3 focus was manifested by pitch lowering toward its low pitch target. Thus, MaxP was one of the least important cues, whereas MAS, RNG, and MinP were prominent cues for this focused tone.

As we noted in the introduction, conventional wisdom on prosodic focus views duration, intensity, and pitch range as concomitantly increased in prosodic realization of focus. What is striking here, however, is that in our results, Int did not play a key role in marking corrective focus prosodically, except for in tone 3. For example, for tones 2 and 4, Int was not ranked even as one of the four most important cues affecting the differences between broad and corrective focus. For tone 1, although Int was included in the top four acoustic cues, a notable difference was seen between the third (Dur) and fourth (Int) rankings. By contrast, Dur was consistently used for all tone types and was one of the most reliable cues for distinguishing corrective from broad focus. According to Ma (2017), prosodic focus can be realized not only through the vertical dimension of pitch range expansion but also by the horizontal dimension of duration lengthening. Here, the vertical dimension is restricted by tone type due to tone and intonation interaction – for example, the pitch contour of tone 1 is suppressed because it is a level tone. The horizontal dimension, however, shows no such restriction. A functional hypothesis claims that the prosodic space available to languages is finite (Zhu 2013). If a language uses duration to distinguish between vowels, then the duration parameter will play a less important role in the marking of stress, which in other languages depends on duration (Berinstein 1979; Potisuk et al. 1997). Another piece of evidence comes from Mandarin Chinese, in which pitch is used primarily for lexical tones, and thus, less room remains in pitch for the conveyance of paralinguistic features, such as emotions (Zhu 2013; Wang 2015). Likewise, in Mandarin Chinese, pitch may have a smaller effect on the signaling of corrective focus in certain cases. Therefore, duration may play a more important role in marking corrective focus, which supports our finding that each tone clearly increased in duration in the prosodic marking of corrective focus.

This study demonstrated that phonation cues (CPP and H1H2) were not affected by prosodic marking of corrective focus in Mandarin Chinese – even for tone 3 focus, which, according to some studies (Zheng 2006; Chen & Gussenhoven 2008; Yin & Kong 2010), is distinguished by creakiness. This means that prosodic marking of corrective focus does not usually involve changing voice quality, but instead it most typically occurs with modal voice quality. It is important to note that prosodic marking of corrective focus is realized primarily by prosodic cues, and such cues are modulated differently by tone type. However, we must be cautious in coming to a firm conclusion until natural stimuli are comparatively examined; it is premature to assert that our results will apply equally well to natural stimuli.

Although this study yielded interesting findings on prosodic marking of corrective focus in Mandarin Chinese, it also had several limitations; these results point to directions for future research. As our results were drawn from phone number strings, they may not be generalizable to natural stimuli, thus requiring a comparative study between the two types of stimuli. In addition, more speakers should be recruited for future studies, for two reasons: to increase the reliability of the study and to increase the statistical power of any production experiment. With respect to further directions, future studies should take our results as a guide to testing other tone languages, such as Cantonese and Thai, to determine whether these languages show different facets of prosodic marking for corrective focus, depending on tone type. Cantonese and Thai also have a high level tone. It would be interesting to see, when this tone interacts with prosodic focus in those languages, whether the prosodic realizations of focus are compatible with the prosodic patterns of Mandarin Chinese. In addition, our results have relevance for the field of machine learning. We suggest that the multimodal characteristics of corrective focus in Mandarin Chinese be applied to the development of machine-learning techniques, with special attention to corrective focus. Put differently, machine-learning techniques must consider an ample amount of broad-focus vs. corrective-focus data and predict optimal patterns for each focus type for each tone.

To summarize, this study explored the interaction between tone and corrective focus in Mandarin Chinese, conducting a production experiment based on phone number strings. To identify whether prosodic marking of corrective focus differs by tone type, we analyzed eight prosodic and phonation cues and subjected them to random forest analysis. We observed that the effects of phonation cues (CPP and H1H2) were negligible in marking corrective focus prosodically in the language tested. The prosodic cues, on the other hand, exhibited clearly varied patterns of distribution, particularly for pitch-related parameters. For tone 1, for example, a raising of the pitch register was important for distinguishing between broad and corrective focus, unlike MAS and RNG, which were structurally suppressed due to tone and intonation interaction. But, MAS and RNG were important for contour tones, such as tones 2 and 4. For tone 3, corrective focus was expressed by a lowering of the pitch target, unlike other focused tones, which raised their pitch levels. Moreover, by aggregating all focused tones, we found that pitch-related parameters were the most important factors for distinguishing one from another. The results of this study allow us to conclude that prosodic marking of corrective focus is expressed differently according to the tonal structure of the language. Thus, we affirm that prosodic marking of corrective focus is multimodal in Mandarin Chinese.

Acknowledgements

We thank the anonymous reviewers for their constructive feedback, suggestions, and comments on earlier versions of this manuscript. This work was supported in part by the Chenguang Program (grant no. 16CG21) of the Shanghai Education Development Foundation and the Shanghai Municipal Education Commission, and Fundamental Research Funds for the Central Universities (grant no. WX0110020170698) of Tongji University, awarded to the first author. Partial production data used in this study were obtained from Lee (2015), but the data were here used to assess a different research goal.

Abbreviations

BF	broad focus
CF	corrective focus
CPP	cepstral peak prominence
dB	decibels
Dur	duration
H1H2	amplitude difference between first and second harmonics
Hz	Hertz
Int	mean intensity
MAS	mean absolute slope
MaxP	maximum pitch
MinP	minimum pitch
ms	milliseconds
RNG	pitch range
st	semitones

References

Berinstein, Ava E. 1979. A cross-linguistic study on the perception and production of stress. *UCLA Working Papers in Phonetics* 47. 1–59.

Cao, Wen & Zhang, Jinsong. 2008. Tone-3 accent realization in short Chinese sentences. *Tsinghua Science and Technology* 13(4). 533–539. https://doi.org/10.1016/S1007-0214(08)70085-3

Cao, Jianfen. 2002. Hanyu shengdiao yu yudiao de guanxi [The relationship between tone and intonation in Mandarin Chinese]. *Zhongguo Yuwen* [Studies of the Chinese Language] 2002(3). 195–202.

Chao, Yuen Ren. 1968. *A grammar of spoken Chinese*. Berkeley: University of California Press.

Chen, Yiya & Gussenhoven, Carlos. 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

Cooper, William E. & Eady, Stephen J. & Mueller, Pamela R. 1985. Acoustical aspects of contrastive stress in question-answer contexts. *The Journal of the Acoustical Society of America* 77(6). 2142–2156. https://doi.org/10.1121/1.392372

Gårding, Eva & Zhang, Jialu & Svantesson, Jan-Olof. 1983. A generative model for tone and intonation in standard Chinese. In Gösta, Bruce & Gårding, Eva (eds.), *Working papers in linguistics*, vol. 25, 53–65. Lund: Department of Linguistics and Phonetics, Lund University.

Gussenhoven, Carlos. 2008. Types of focus in English. In Lee, Chungmin & Gordon, Matthew & Büring, Daniel (eds.), *Topic and focus: Cross-linguistic perspectives on meaning and intonation*, vol. 82, 83–100. Dordrecht: Springer.

Keating, Patricia & Garellek, Marc & Kreiman, Jody. 2015. Acoustic properties of different kinds of creaky voice. In The Scottish Consortium for ICPhS 2015 (ed.), *Proceedings of the 18th International Congress of Phonetic Sciences*. Glasgow: University of Glasgow. (https://www.internationalphoneticassociation.org/icphs-proceedings/ICPhS2015/Papers/ICPHS0821.pdf) (Accessed 2019-10-21.)

Kuznetsova, Alexandra & Brockhoff, Per Bruun & Christensen, Rune Haubo Bojesen. 2013. lmerTest: Tests for random and fixed effects for linear mixed effect models (lmer objects of lme4 package). *R Package* (Version 1.2–0, 20 March 2013.) (http://www2.uaem.mx/r-mirror/web/packages/lmerTest/index.html) (Accessed 2019-10-31.)

Ladd, D. Robert. 1984. English compound stress. In Gibbon, Dafydd & Richter, Helmut (eds.), *Intonation, accent, and rhythm: Studies in discourse phonology*, 253–266. Berlin: Walter de Gruyter. https://doi.org/10.1515/9783110863239.253

Lee, Albert & Xu, Yi. 2012. Revisiting focus prosody in Japanese. In Ma, Qiuwu & Ding, Hongwei & Hirst, Daniel (eds.), *Proceedings of Speech Prosody 2012*, 274–277. ISCA Archive. (The 6th International Conference on Speech Prosody, Shanghai, 22–25 May 2012.) (https://www.isca-speech.org/archive/sp2012/sp12_274.html) (Accessed 2019-09-16.)

Lee, Yong-cheol. 2015. *Prosodic focus within and across languages*. Philadelphia: University of Pennsylvania. (Doctoral dissertation.)

Lee, Yong-cheol. 2017. Prosodic focus in Seoul Korean and South Kyungsang Korean. *Linguistic Research* 34(1). 133–161. https://doi.org/10.17250/khisli.34.1.201703.005

Lee, Yong-cheol & Wang, Bei & Chen, Sisi & Adda-Decker, Martine & Amelot, Angélique & Nambu, Satoshi & Liberman, Mark. 2015. A crosslinguistic study of prosodic focus. *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2015)*, 4754–4758. New York: IEEE (https://ieeexplore.ieee.org/document/7178873) (Accessed 2019-10-21.) https://doi.org/10.1109/ICASSP.2015.7178873

Lee, Yong-cheol & Nambu, Satoshi & Cho, Sunghye. 2018. Focus prosody of telephone numbers in Tokyo Japanese. *The Journal of the Acoustical Society of America* 143(5). EL340–EL346. https://doi.org/10.1121/1.5037360

Lee, Yong-cheol & Wang, Ting & Liberman, Mark. 2016. Production and perception of tone 3 focus in Mandarin Chinese. *Frontiers in Psychology* 7. 1-13 (Article no. 1058). (https://www.frontiersin.org/articles/10.3389/fpsyg.2016.01058/full) (Accessed 2020-02-07.)

Lee, Yong-cheol & Xu, Yi. 2010. Phonetic realization of contrastive focus in Korean. *Proceedings of Speech Prosody 2010* (Paper 033). ISCA Archive. (https://www.isca-speech.org/archive/sp2010/sp10_033.html) (Accessed 2019-10-21.)

Liaw, Andy & Wiener, Matthew. 2002. Classification and regression by randomForest. *R News*, vol. 2/3, 18–22. (https://cran.r-project.org/doc/Rnews/Rnews_2002-3.pdf) (Accessed 2019-10-31.) https://doi.org/10.1057/9780230509993

Ma, Qiuwu. 2017. The focus accents in Chinese intonation: Prosodic implementation and types. In Feng, Shengli (ed.), *Studies in prosodic grammar 2017*, vol. 2, 32–72. Beijing: Beijing Language and Culture University Press.

Ouyang, Iris Chuoying & Kaiser, Elsi. 2015. Prosody and information structure in a tone language: An investigation of Mandarin Chinese. *Language, Cognition and Neuroscience* 30(1–2). 57–72. https://doi.org/10.1080/01690965.2013.805795

Potisuk, Siripong & Gandour, Jackson & Harper, Mary P. 1997. Contextual variations in trisyllabic sequences of Thai tones. *Phonetica* 54(1). 22–42. https://doi.org/10.1159/000262208

R Core Team. 2017. *R: A language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing. (http://www.r-project.org) (Accessed 2019-10-21.)

Shen, Jiong. 1985. Beijinghua shengdiao de yinyu he yudiao [Pitch range of tone and intonation in Beijing dialect]. In Lin, T. & Wang, L. (eds.), *Beijing yuyin shiyanlu* [Working papers in experimental phonetics], 73–130. Beijing: Peking University Press.

Shih, Chilin. 1988. Tone and intonation in Mandarin. *Working Papers of Cornell Phonetics Laboratory (WPCPL)* 3. 83–109.

Wang, Ting. 2015. *Shengdiao yuyan de qinggan yuyin shiyan yanjiu* [An experimental study of emotional speech in tonal languages]. Shanghai: Tongji University. (Doctoral dissertation.)

Wang, Ting & Lee, Yong-cheol. 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–EL123. https://doi.org/10.1121/1.4904916

Wu, Zongji. 2001. Putonghua yuyin hecheng zhong youguan zirandu de yunlv bianliang wenti [Naturalness of prosodic parameters in the speech synthesis of Mandarin]. In Tsai, Lian-hong & Zhou, Tong-chun & Tao, Jian-hua (eds.), *Xinshiji de xiandai yuyinxue: Di wu jie quanguo xiandai yuyinxue xueshuhuiyi lunwenji* [Modern phonetics of the new century: Proceedings of the 5th National Conference on Modern Phonetics], 299–302. Beijing: Tsinghua University Press.

Xu, Yi. 1997. Contextual tonal variations in Mandarin. *Journal of Phonetics* 25(1). 61–83. https://doi.org/10.1006/jpho.1996.0034

Xu, Yi. 1999. Effects of tone and focus on the formation and alignment of f_0 contours. *Journal of Phonetics* 27(1). 55–105. https://doi.org/10.1006/jpho.1999.0086

Xu, Yi. 2004. Understanding tone from the perspective of production and perception. *Language and Linguistics* 5(4). 757–797.

Xu, Yi. 2005. Speech melody as articulatorily implemented communicative functions. *Speech Communication* 46(3–4). 220–251. https://doi.org/10.1016/j.specom.2005.02.014

Xu, Yi. 2013. ProsodyPro – A Tool for large-scale systematic prosody analysis. In Bigi, Brigitte & Hirst, Daniel (eds.), *Proceedings of Tools and Resources for the Analysis of Speech Prosody* (TRASP 2013), 7–10. Aix-en-Provence: Laboratoire Parole et Langage.

Xu, Yi & Xu, Ching X. 2005. Phonetic realization of focus in English declarative intonation. *Journal of Phonetics* 33(2). 159–197. https://doi.org/10.1016/j.wocn.2004.11.001

Xu, Yi & Wang, Maolin. 2009. Organizing syllables into groups – Evidence from F_0 and duration patterns in Mandarin. *Journal of Phonetics* 37(4). 502–520. https://doi.org/10.1016/j.wocn.2009.08.003

Xu, Yi & Wang, Q. Emily. 2001. Pitch targets and their realization: Evidence from Mandarin Chinese. *Speech Communication* 33(4). 319–337. https://doi.org/10.1016/S0167-6393(00)00063-7

Xu, Yi & Chen, Szu-wei & Wang, Bei. 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

Yin, Jide & Kong, Jiangping. 2010. Putonghua yuju zhongyin de sangyin fasheng biaoxian yanjiu [The phonation characteristics of sentence focus in Mandarin]. *Proceedings of the 9th Phonetic Conference of China*, 35–44. (http://cpfd.cnki.com.cn/Article/CPFDTOTAL-EGYO201005001097.htm) (Accessed 2019-10-22.)

Zheng, Xiaoju. 2006. Voice quality variation with tone and focus in Mandarin. *Proceedings of the Second International Symposium on Tonal Aspects of Languages (TAL 2006)*, 132–136. ISCA Archive. (https://www.isca-speech.org/archive/tal_2006/tal6_132.html) (Accessed 2019-10-22.)

Zhu, Yinyin. 2013. *Expression and recognition of emotion in native and foreign speech: The case of Mandarin and Dutch* (LOT Dissertation Series 341). Utrecht: Netherlands Graduate School of Linguistics.

Authors' addresses

Yong-cheol Lee (corresponding author)
Department of English Literature and Language
Cheongju University
Cheongju 28503
South Korea
yongcheol@cju.ac.kr

Publication history

Date received: 31 May 2017
Date accepted: 25 June 2018
Published online: 2 April 2020