Mandarin Neutral Tone—does it Change Target

Xiaoluan Liu

Department of Speech, Hearing and Phonetic Sciences, University College London, London, UK

Email address:
LXL0803@gmail.com

Abstract: It is known that in Mandarin each of the five lexical tones can be assigned with an articulatorily functional target: [high] for tone 1, [rise] for tone 2, [low] for tone 3, [fall] for tone 4 and [mid] for tone 5 (the first four tones are known as full tones while tone 5 is called neutral tone). Given that the targets of full tones can change (e.g., from tone 3 to tone 2) in certain speech conditions (e.g. tone sandhi), it is natural to ask whether the same is true for Mandarin neutral tone. This is still an unresolved question, the solution of which can contribute to our understanding of articulatory "strength" as an index of speech communication which is less well explored than other areas of speech production. Motivated by the above concerns, this study uses speech production experiment to test whether the target of Mandarin neutral tone has similar target values (in terms of target slope, height, duration and strength) to those of other tones in Mandarin under three speech conditions: emotion (anger, happiness, disgust and neural emotion), sentence position of the neutral tone (sentence medial and final) and tones preceding the neutral tone (all full tones in Mandarin). The results reveal that the neutral tone is highly likely to change its target in certain combinations of the aforementioned three speech conditions. This study not only further supports previous studies on the impact of emotion, sentence position and tonal contexts on the target behavior of tones, but also highlights the possibility of Mandarin neutral tone changing from weak to strong in articulation for the purpose of effective communication, providing further evidence for "strength" as a communication index.

Keywords: Mandarin Neutral Tone, Tonal Targets, Articulatory Strength

1. Introduction

Mandarin is known as a tone language in which lexical entities are distinguished on the basis of not only segments but also more importantly suprasegments such as tones. There are five tone types in Mandarin: High (H), Rising (R), Low (L), Falling (F), and Neutral (N). Recent research [61-62, 64, 66] on Mandarin tones has suggested that the surface F0 of Mandarin tones can be associated with invariant underlying pitch targets which are defined as "the smallest articulatory operable units associated with linguistically functional pitch units such as tone and pitch accents" [66]. The surface F0 of Mandarin tones undergoes a process called Target Approximation (TA) to approach their underlying pitch targets demonstrated by the following TA model (Fig.1).

It is shown that the solid curve representing the surface F0 asymptotically approximates the dashed line representing the underlying pitch targets. All Mandarin full tones are therefore assigned with pitch targets: [high] for high tones and [low] for low tones; [rise] for rising tones and [fall] for falling tones (the neutral tone is to be discussed in the following section). The first two are static targets in that they have a static pitch register while the last two are dynamic targets because they move in a linear fashion.
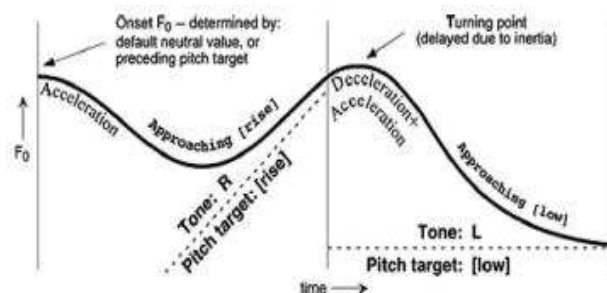

Figure 1. *Graphical display of the Target Approximation model (from 66]).*

Target slope, height, duration and strength are the parameters proposed to quantitatively capture the characteristics of pitch targets, which can be used for facilitating computational modelling of speech prosody (cf. [40, 43] for the mathematic calculation of those parameters).

2. Mandarin Neutral Tone and Motivations of this Study

2.1. Mandarin Neutral Tone

Apart from the four full tones in Mandarin mentioned above, a less well-known but equally important tonal category in Mandarin is the neutral tone. It is typically distributed in the forms of grammatical morphemes, lexical items, diminutive terms and reduplication [8]. The following examples (from [9]) correspond to the four functions just mentioned (the neutral tone is marked without tonal markings):

(a) *làde* 'something spicy'
(b) *bōli* 'glass'
(c) *mèimei* 'sister'
(d) *xiángxiang* 'to think (for a little while)'

Because of the fact that Mandarin neutral tone does not have a consistent F0 contour as do those of the four full tones (cf. [9] for references), is has been a subject of controversy in terms of whether it has a fixed target. A majority of the previous studies on Mandarin neutral tone suggest that it is targetless/toneless because of the variability of its surface F0 which has been seen as the product of the interpolation between two adjacent tones [30, 51, 56]. Nevertheless, recent findings [9] have demonstrated that Mandarin neutral tone does have a target [mid] which is somewhere in the middle of the range between the tonal targets [high] and [low].

The neutral tone is also found to have intrinsically weak articulatory strength which is reflected in the fact that it is not as fast and effective as the four full tones to overcome the influence of the preceding full tones. Hence the F0 variation of the neutral tone is not a result of the interpolation between two adjacent tones; rather, it is mainly affected by its preceding tone (especially in terms of velocity) rather than its following tone [9].

2.2. Motivations of this Study

What remains unknown, however, is whether this [mid] target of the neutral tone remains unchanged in any speech conditions, given the fact that surface F0 of tones tends to vary in different speech conditions. Such a phenomenon is known as tonal variation which is thought to have two types [61]: Target alteration and implementational variation. Target alteration "occurs in cases where the pitch target of a tone is presumably changed before being implemented in articulation" [61]. An example is that L tone in Mandarin occurring in non-sentence-final position tends to have a low-dip only, i.e., losing the final rise in its canonical form [8, 62]. Another example in Mandarin is the L to R rule in which the L tone when followed by another L tone changes into the R tone [61]. Implementational variation refers to cases where either different tonal contexts or articulatory effort make the acoustic realization of the tonal targets varied but not fundamentally different from the canonical form [61]. For instance, the amount of F0 drop in L tone in

non-sentence-final position varies according to speakers and contexts, although the underlying form is still L [8, 61].

Therefore, given the variant nature of tones in different contexts, it is reasonable to speculate that Mandarin neutral tone may not stick to an invariant form in all speech conditions, particularly when we consider factors such as emotions, the sentence position of the neutral tone and the category of the tones preceding the neutral tone (reasons for choosing these factors are to be given in the following paragraphs). Admittedly, one may argue that Mandarin full tones are also subject to the influence of those factors and hence there is no good reason to single out the neutral tone. However, the extent of the influence of the factors on different tonal categories may be quite various. For example, although the mean F0 and F0 range of Mandarin full tones do vary according to different emotions [31], they have never been reported to vary to such an extent that they undergo fundamental target change, i.e., changing from one tonal target (e.g. tone 1) to another (e.g. tone 2) in emotional speech. Therefore the F0 variations of Mandarin full tones do not categorically deviate from their canonical tonal contours even though in emotional speech. Mandarin neutral tone, on the other hand, is hard to believe to behave in the same way as the full tones since it is weaker in articulatory strength than full tones [9] and hence it is more prone to the impact of the aforementioned factors, making the extent of the surface F0 variability of the neutral tone likely to be greater than that of the full tones. Whether such greater variability may give rise to target change of the neutral tone is unknown, and hence it serves as the motivation of this study.

Emotion is chosen as a factor because of the well-known influence of emotion on segmental and suprasegmental aspects of speech, which has been systematically studied over a long period (cf. [47] for details). Despite the controversy as to the exact acoustic markers of each emotion, what seems consistent across the previous studies on emotion and speech is that the acoustic characteristics of speech do not stay the same in emotional contexts as in neutral contexts (cf. [48]). Therefore, emotion serves as a perfect platform to test the variations in the behavior of the neutral tone target. Since it is impossible to exhaust all the emotional categories in this study, only three are selected out of the "big six" emotions [14]: (Hot) anger, happiness and disgust. Given that anger has two types—hot and cold [46] with different acoustic characteristics [20], only hot anger is selected to avoid theoretical and practical complexities.

Anger and happiness are selected and discussed together because they are the most widely studied and frequently encountered emotions in daily life [33]. Moreover, although the two emotions occupy the two extremes of the emotion continuum, the acoustic characteristics of the two are similar: The consensus of numerous studies seems to be that intensity, pitch range and speech rate are almost the same for anger and happiness [34, 47, 58]. This is further supported by the fact that anger and happiness are found to

be perceptually indistinguishable [29]. Physiologically, both of the two emotions belong to the "excitement" category which involves increased cardiorespiratory activities: Speeding up of the heart rate [44] together with fast and deep breathing [5]. That is why anger and happiness are ranked almost the same place on the scale of activation of emotion [12].

Although disgust is not as well and widely studied as anger and happiness, the interesting physiological response to disgust warrants special attention: Evolutionarily, disgust may have arisen as the physical act to reject toxic or rotten food [13]. Consequently, the sound of feeling disgusted is similar to the sound when one orally throws out nasty food (e.g. vomiting), accompanied by such cardiorespiratory activities as decrease in heart rate [15], slow or even suppression in breathing [4-5], pharynx tightening, F1 raising and possible devoicing [35].

The sentence position of the neutral tone is chosen as another factor that may affect the target behavior of the neutral tone. This is because as has been observed by many [21, 39, 61], the position of speech segments plays an important role in determining the surface phonetic realizations of the underling phonological forms (e.g., English boundary tone). With regard to lexical tone, sentence position also plays a role as supported by the evidence from Stockholm Swedish [6] in which the prenuclear position is the place where the canonical form of a lexical tone rests. Another motivation for taking into account the impact of sentence position is that it is related to articulatory strength. As has been suggested [26, 66],

articulatory strength tends to serve as an encoding scheme in communication, i.e., "all levels of strength carries information" [26]. A known example is sentence final lengthening and sentence initial strengthening which has been associated with speaker's intention to control articulation [26]. Given that the neutral tone is by nature weak as mentioned above, it is likely that different sentence positions may affect the surface F0 realizations of the underlying [mid] target of the neutral tone.

Apart from emotions and sentence position which can justifiably affect the surface F0 of the neutral tone as discussed above, the tones preceding the neutral tone are also likely to be a factor. This is because the source of the surface variation of the F0 of the neutral tone stems from the F0 and velocity of the full tone immediately preceding the neutral tone [9]. Hence there is a need to include all of the four full tones in Mandarin to investigate their contribution to the tonal variations of the following neutral tone. Motivated by the above considerations, this study attempts to address the following research question:

Is there a change of target of Mandarin neutral tone due to the influence of emotions, its position in sentences and its preceding full tones?

As discussed above, anger, disgust and happiness are selected. Neutral emotion is also included for the sake of comparison. Only sentence medial and final positions are selected to test the effect of sentence position on the neutral tone, since the neutral tone does not occur sentence initially [9]. Each of the four full tones in Mandarin is included as the tone immediately preceding the neutral tone.

Table 1. *The stimuli of the experiment in which the numbers of the syllables represent the five lexical tones in Mandarin: 1 for H (High tone), 2 for R (Rising tone), 3 for L (Low tone), 4 for F (Falling tone), and 5 for N (Neutral tone).*

	mao1 cat	*men1* cover			*mao1* cat	*men1* cover
xiao3 little	*mao2* fur	*men4* simmer	*yue4dui4 fa1xing2 le5* band release particle	*xiao3* little	*mao2* fur	*men4* simmer
	mao3 mortise	*men5* article			*mao3* mortise	*men5* particle
	mao4 hat				*mao4* hat	

Sentence meaning: The band *Xiaomaomen* has released the album *Xiaomaomen*.

3. Methodology

The stimuli (Table 1) in this study include three sets of Mandarin sentences with the target syllables (to be analyzed graphically and statistically) *mao* and *men* imbedded in sentence middle and final positions. The selection of *mao* and *men* is for the ease of phonetic segmentation while minimizing consonantal perturbation of F0, given the fact that non-sonorant consonants tend to raise or lower the F0 of the adjacent vowels (cf. [63]). In each set, there are four sentences with *mao* being assigned with four full tones respectively (high tone *mao1*, rising tone *mao2*, low tone *mao3* and falling tone *mao4*) as the syllable preceding *men*. *Men* apart from being assigned with a neutral tone (*men5*) is also assigned with a high tone (*men1*) and a falling tone (*men4*) to investigate whether the neutral tone will change its

target to that of full tones, based on the results of a pilot study showing that the neutral tone is likely to become similar in target to either a high tone or falling tone when the factors mentioned in section 2.2 are taken into account. Therefore, *men* is assigned with a high tone, falling tone and neutral tone respectively for set one, two and three of the sentences. Note that for the sake of semantic/pragmatic naturalness, *xiao* is added to *maomen* to form a compound word with three syllables denoting the name of a brand and their album, although *xiao*, *mao* and *men* when used separately have their own independent meanings.

Ten native Mandarin speakers with drama training background were recruited as subjects. They reported no speech or hearing problems. Emotion portrayal method was used to induce emotion, i.e., having subjects imagine themselves being in an emotional state as vividly as possible when saying each sentence (cf. [47-48] for the

justification of this method). The recording was conducted in a sound-controlled booth. All the sentences were repeated 3 times by 10 subjects in angry, disgust, happy and neutral emotion, resulting in 4 (*mao1/2/3/4*) * 3(*men1/4/5*) * 4 (emotions) * 10 (subjects) * 3 (repetitions) = 1440 sentences. A customized version of ProsodyPro [65], running under Praat [3], was used to extract and analyse F0 contours. The F0 contours were time-normalized in the sense that they were obtained by dividing the two syllables (*mao* and *men*) into 20 equal intervals (10 intervals for each syllable) and averaging across the three repetitions of the same sentence by 10 subjects. Then the F0 in Hertz were converted to semitones to reduce the bias towards higher pitch range over lower pitch range for the sake of statistical analyses before they were transformed back to Hertz for graphical demonstration. Similar methods have been used in [9, 32].Segmental boundaries were labelled by hands with visual inspection and listening validation.

4. Results

4.1. Graphical Presentations of Time-Normalized F0 Contours

Figs.2-5 display the time-normalized F0 contours of *mao1/2/3/4* + *men1/4/5* at sentence medial and final positions in neutral, angry, happy and disgust speech. It can be observed that firstly, the neutral tone *men5* tends to become more similar to full tones (either *men1* or *men4*) in emotional than neutral speech, particularly in angry (Fig.3) and disgust speech (Fig.5). Secondly, the tonal categories of the preceding full tones appear to have affected the behavior of the neutral tone as well: Following the rising tone (*mao2*), the F0 contour of *men5* in many cases is very similar to that of *men4*; following the low tone (*mao3*), *men5* in emotional speech tends to display an F0 contour similar to that of *men1*. Thirdly, the extent of similarities between the neutral tone and full tones seems greater in sentence medial position than sentence final position.
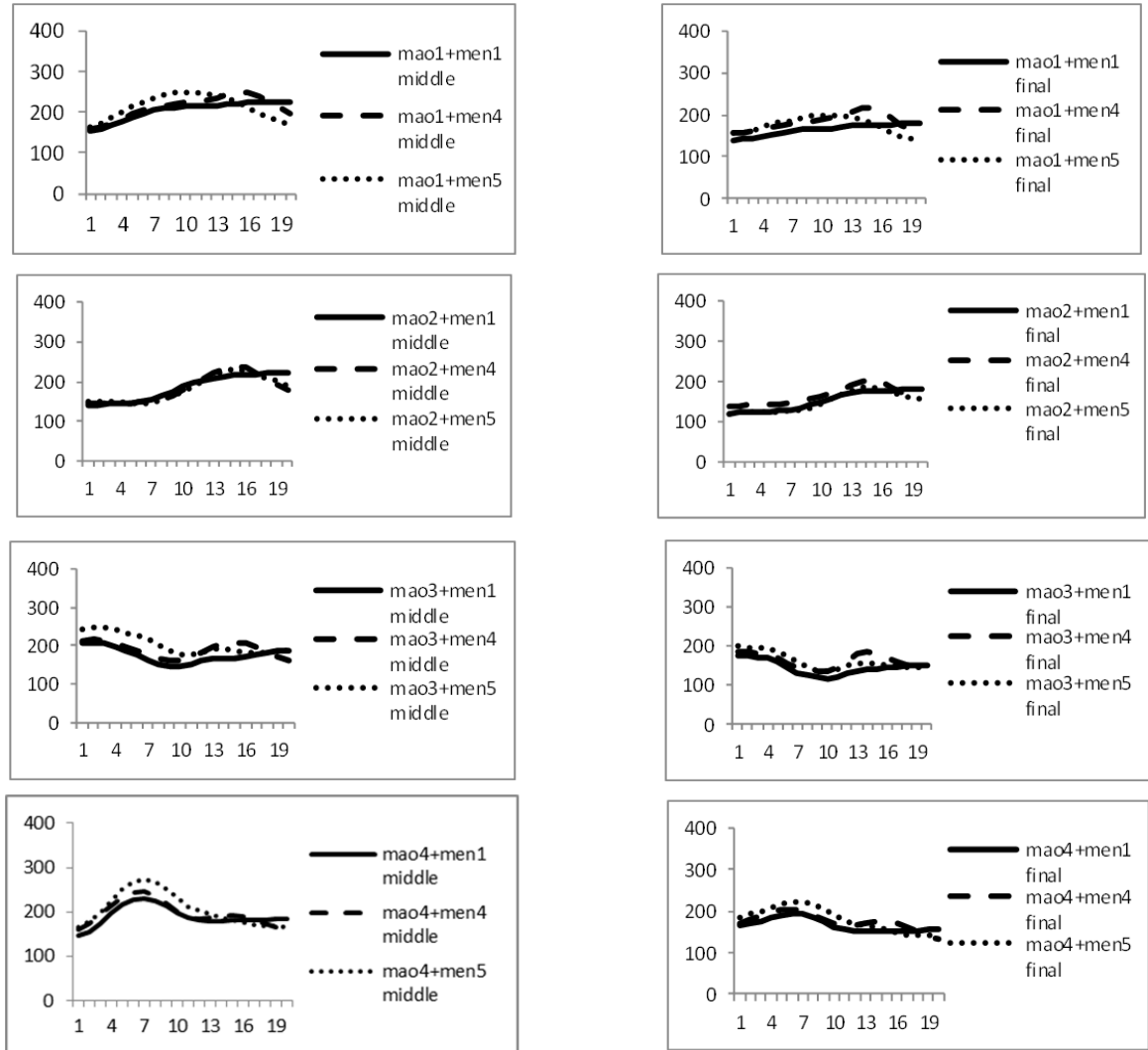

Figure 2. Time-normalized mean F0 contours of mao1/2/3/4 + men1/4/5 (at sentence medial and final positions) in neutral speech.
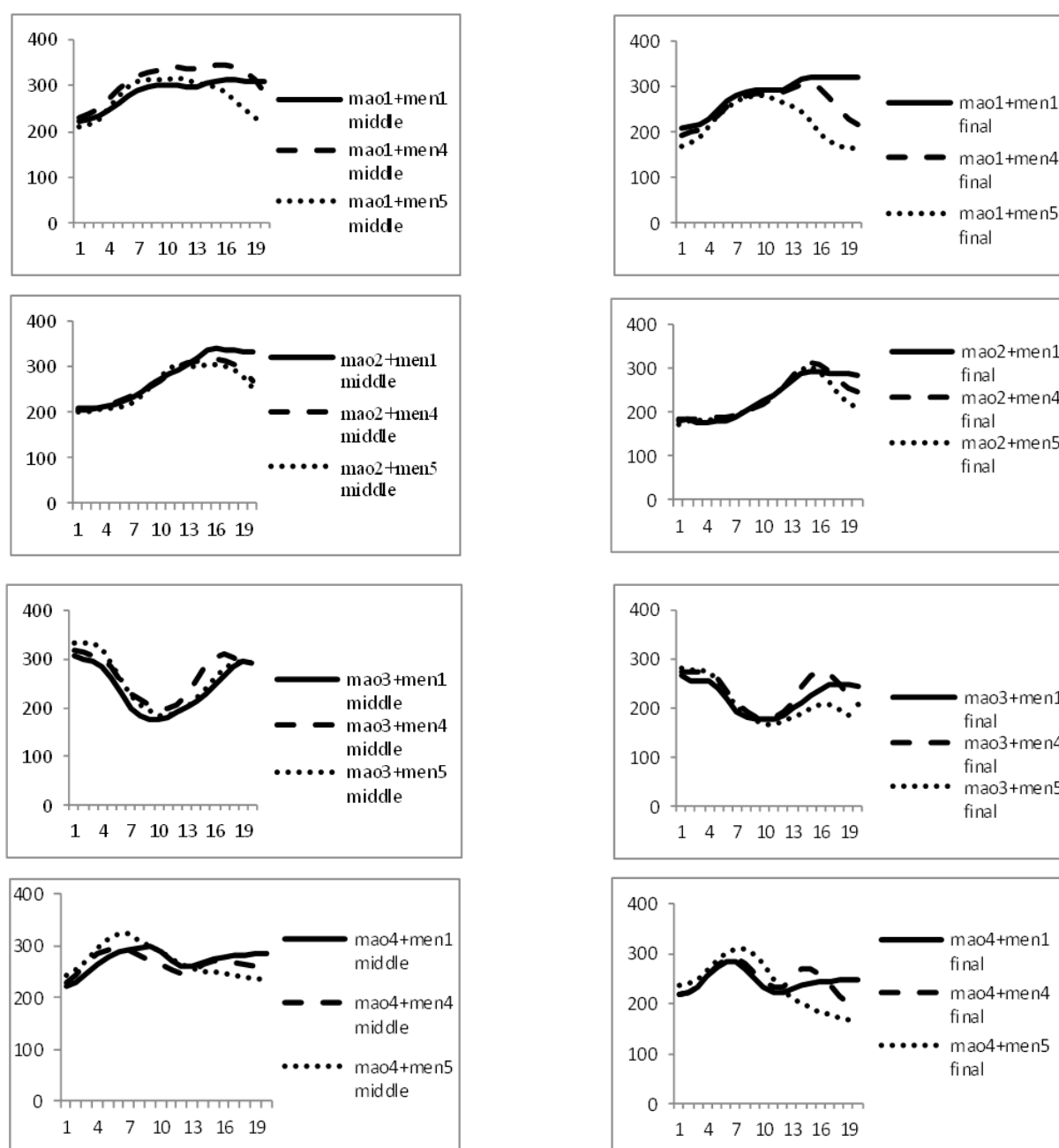
Figure 3. *Time-normalized mean F0 contours of mao1/2/3/4 + men1/4/5 (at sentence medial and final positions) in angry speech.*
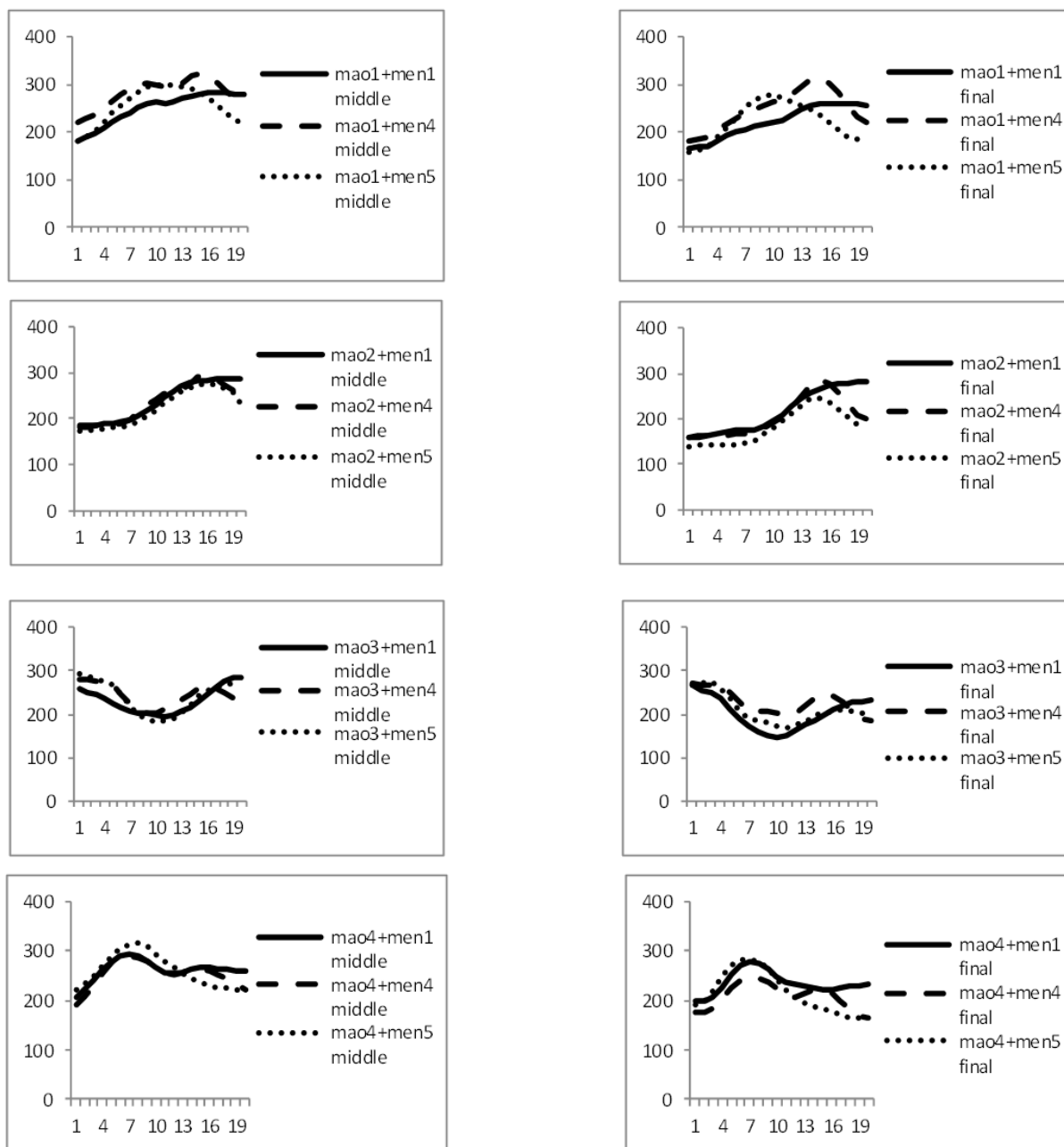
Figure 4. *Time-normalized mean F0 contours of mao1/2/3/4 + men1/4/5 (at sentence medial and final positions) in happy speech.*
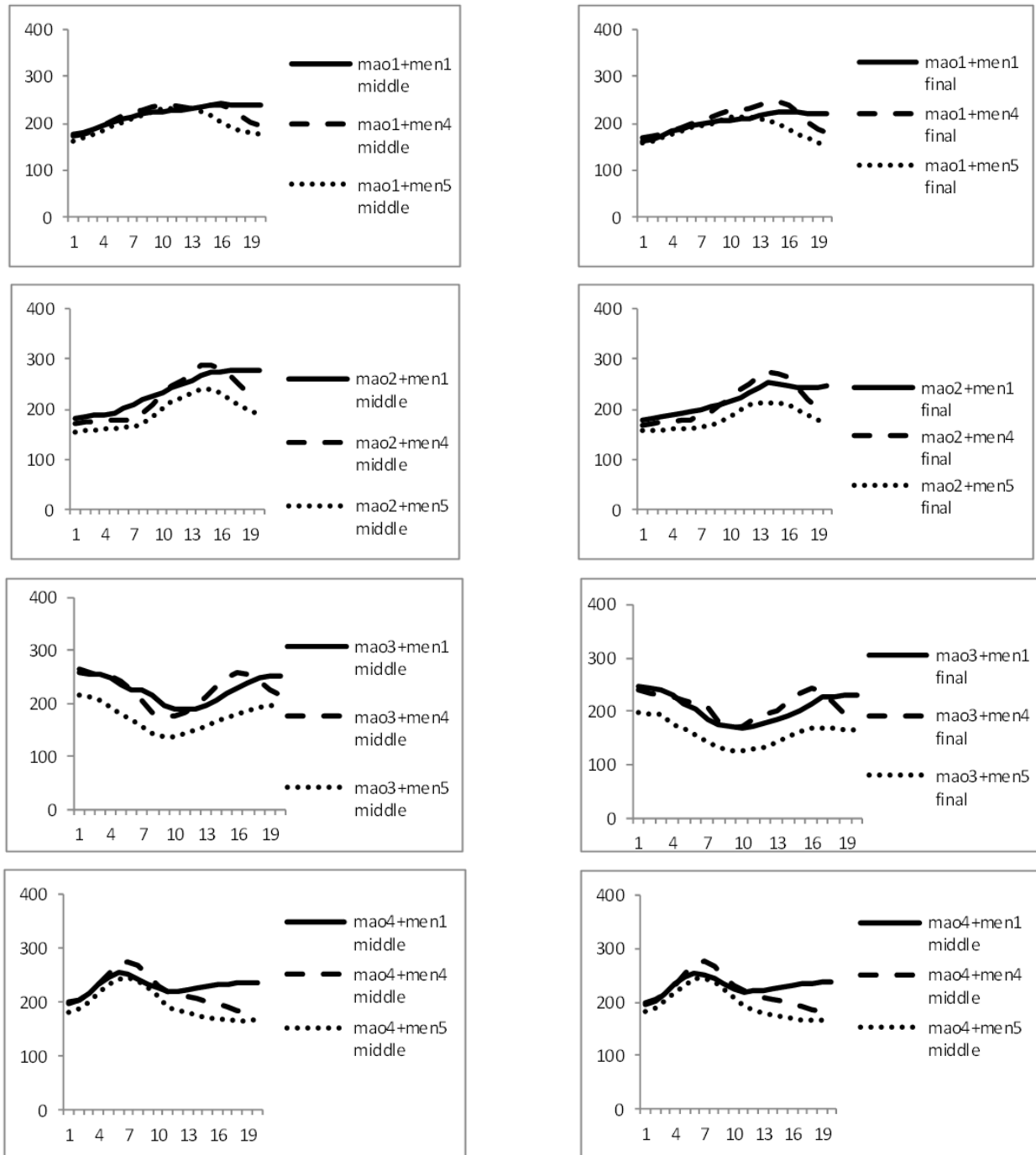
Figure 5. *Time-normalized mean F0 contours of mao1/2/3/4 + men1/4/5 (at sentence medial and final positions) in disgust speech.*

4.2. Quantitative Analyses

To formally assess whether the above observations of the graphical displays of the tonal targets are significant, statistical tests were performed on the parameters (i.e., slope, height, duration and strength) of the tonal targets, using PENTATrainer1 [43] and PENTATrainer2 [42] to obtain the values of the parameters.

4.2.1. Target Slope and Height

A three-way repeated measures ANOVA was performed to test the effect of emotion, preceding full tones and sentence position on the target slope and height of the neutral tone *men5*. Emotion has four levels (anger, disgust, happiness and neutral); preceding full tones have four levels (*mao1*, *mao2*, *mao3*, *mao4*); sentence position has two levels (sentence medial and final). The results as demonstrated in Table 2 show that all the three factors as well as their interactions have a significant impact on the target (slope and height) of the neutral tone. Post-hoc Tukey tests further suggest that the neutral tone can be significantly similar to full tones in terms of target slope and height under conditions shown in Table 3.

4.2.2. Target Duration

A three-way (emotions, preceding full tones and sentence position) repeated measures ANOVA was performed (Table 4) showing that all three factors and the interaction between them play a significant role in determining the duration of *men* (*men1/4/5*). Post-hoc Tukey tests were further conducted to examine in which condition the neutral tone does not differ significantly from the full tones in terms of duration. Table 5 shows that firstly, *men5* becomes more similar to either *men1* or *men4* in emotional speech (anger and disgust in particular) than neutral speech and this occurs more often in sentence medial position than sentence final position. Secondly, it can be seen that there are more durational similarities between the neutral tone and full tones when they follow *mao2* and *mao3* than when they follow *mao1* and *mao4*.

4.2.3. Target Strength

A three-way (emotions, preceding full tones and sentence position) repeated measures ANOVA was performed. The results (Table 6) show that all three factors and the interaction between them play a significant role in determining the target strength of *men* (*men1/4/5*).

Post-hoc Tukey tests were further conducted to examine in which condition the neutral tone does not significantly differ from the full tones in terms of target strength. The results shown in Table 7 are similar to those in Table 5: Firstly, *men5* becomes more similar to either *men1* or *men4* in emotional speech than neutral speech. Moreover, this

pattern occurs more often in sentence medial position than final position. Secondly, there are more strength similarities between the neutral tone and full tones when preceded by *mao2* and *mao3* than preceded by *mao1* and *mao4*.

5. Discussion

5.1. Is there a Target Change of the Neutral Tone

A caveat is in order here before proceeding to the final discussion: The variations of the surface F0 contour of the Mandarin neutral tone reported above are different from the well-known carry-over effect of tones ([18] for Thai; [59-60] for Mandarin). This is because tones under the influence of carry-over effects can still approach their own tonal targets by the end of their carrier syllables [62], although the initial part of the F0 contour is heavily assimilated to the contour of the previous tone. Target change, on the other hand, means that other things being equal, instead of approaching its own tonal target, a tone takes on a target shape similar to that of another tone even by the end of its carrier syllable [61].

The results of this study suggest that the variations of the neutral tone do not fall into the carry-over category because the target of the neutral tones significantly similar to some of the full tones as summarized in Table 8, taking into account the comparisons of target slope, height, duration and strength between the neutral tone and full tones discussed in the above sections.

Table 2. *Results of the three-way repeated measures ANOVA on the effects of emotions, preceding full tones and sentence position on the target slope and height of the neutral tone men5.*

Significant effects	Target slope			Target height		
	F	df	*p*	F	df	*p*
emotions	3.22	3, 27	<0.05	3.39	3, 27	<0.05
preceding full tones	3.85	3, 27	<0.05	4.01	3, 27	<0.05
sentence position	8.97	1, 9	<0.05	7.76	1,9	<0.05
emotions * preceding full tones * sentence position	2.13	9, 81	<0.05	2.39	9, 81	<0.05

Table 3. *Pairs of men (sentence medial and final) that are not significantly different in terms of target slope and height by post-hoc Tukey tests.*

		Slope				Height			
		medial *men*		final *men*		medial *men*		final *men*	
		men5 vs. *men1*	*men5* vs. *men4*	*men5* vs. *men1*	*men5* vs. *men4*	*men5* vs. *men1*	*men5* vs. *men4*	*men5* vs. *men1*	*men5* vs. *men4*
mao1	anger		✓		✓				
	disgust		✓		✓				
	happiness		✓		✓				
	neutral		✓		✓				
mao2	anger	✓	✓	✓	✓		✓		✓
	disgust	✓	✓	✓	✓		✓		✓
	happiness	✓	✓	✓	✓		✓		✓
	neutral	✓	✓	✓	✓		✓		✓

		Slope				Height			
		medial *men*		final *men*		medial *men*		final *men*	
		men5 vs. *men1*	*men5* vs. *men4*	*men5* vs. *men1*	*men5* vs. *men4*	*men5* vs. *men1*	*men5* vs. *men4*	*men5* vs. *men1*	*men5* vs. *men4*
mao3	anger	✓			✓	✓	✓		
	disgust	✓		✓		✓	✓	✓	✓
	happiness	✓		✓		✓			✓
	neutral					✓	✓	✓	✓
mao4	anger						✓		✓
	disgust		✓		✓		✓		✓
	happiness						✓		✓
	neutral						✓		✓

Table 4. *The significant effects of emotions, preceding full tones and sentence position on the duration of men1/4/5.*

Significant effects	Duration		
	F	df	p
emotions	3.46	3, 27	<0.05
preceding full tones	4.05	3, 27	<0.05
sentence position	7.62	1, 9	<0.05
emotions* preceding full tones * sentence position	2.15	9, 81	<0.05

Table 5. *Pairs of men (sentence medial and final) that are not significantly different in terms of duration by post-hoc Tukey tests.*

		Sentence middle		Sentence final	
		*men1*vs. *men5*	*men4*vs. *men5*	*men1*vs. *men5*	*men4*vs. *men5*
mao1	anger				
	disgust		✓		✓
	happiness				
	neutral				
mao2	anger		✓		✓
	disgust		✓		✓
	happiness				
	neutral				
mao3	anger	✓			
	disgust	✓			✓
	happiness	✓		✓	
	neutral	✓			
mao4	anger		✓		✓
	disgust		✓		
	happiness				
	neutral				

Table 6. *The significant effects of emotions, preceding full tones and sentence position on the target strength of men1/4/5.*

Significant effects	Target strength		
	F	df	p
emotions	3.31	3, 27	<0.05
preceding full tones	3.99	3, 27	<0.05
sentence position	8.36	1, 9	<0.05
emotions* preceding full tones * sentence position	2.39	9, 81	<0.05

Table 7. *Pairs of men (sentence medial and final) that are not significantly different in terms of target strength by post-hoc Tukey tests.*

		Sentence middle		Sentence final	
		*men1*vs. *men5*	*men4*vs. *men5*	*men1*vs. *men5*	*men4*vs. *men5*
mao1	anger				
	disgust				
	happiness				
	neutral				
mao2	anger		✓		✓
	disgust		✓		✓
	happiness				
	neutral				
mao3	anger	✓		✓	
	disgust	✓		✓	
	happiness	✓			
	neutral				
mao4	anger	✓			
	disgust		✓		
	happiness				
	neutral				

Table 8. *Paris of tones (N=neutral tone, H=high tone, F=falling tone) that have significantly similar targets when preceded by the four full tones (T1=high tone, T2=rising tone, T3=low tone, T4=falling tone) in angry, disgusted, happy and neutral speech.*

	Sentence medial				Sentence final			
	Anger	Disgust	Happiness	Neutral	Anger	Disgust	Happiness	Neutral
following T1								
following T2	N=F	N=F			N=F	N=F		
following T3	N=H	N=H	N=H					
following T4		N=F						

Does it mean the neutral tone in those cases (Table 8) has undergone a process of *complete* target change to the full tones? The answer is a tentative yes with caution. This is mainly because evidence of a complete target change may require a *complete* merger in shape of the target between the neutral tone and full tones, which is somehow not the case for all of the pairs listed in Table 8, as can be observed from the graphical presentations in section 4.

Nevertheless, the significant statistical results clearly show a strong tendency of the neutral tone to become highly similar in target to the full tones under the conditions listed in Table 8. Therefore it can be tentatively concluded that the target of Mandarin neutral tone is highly likely to be similar to that of the full tones under conditions shown in Table 8, although there may not be a 100% merger in pitch contour between them. Also note that the results do not contradict the fact that Mandarin neutral tone has a static [mid] target in neutral speech as proposed in [9], since it is clearly shown in Table 8 that the target change occurs only in emotional speech.

5.2. The Impact of Emotions

The results of this study suggest that the target change of Mandarin neutral tone occurs in emotional speech only, lending further support to previous studies on emotion being an important mechanism for varying the phonetic behavior of speech segments.

The reason can be associated with the physiological mechanisms of emotions. As has been mentioned in the introduction section, anger and happiness tend to trigger excited cardiorespiratory activities such as speeded up heart rate and blood pressure [44], fast and deep breathing [5], etc. Therefore it is reasonable to speculate in angry and happy speech the increased cardiorespiratory effort can give rise to greater aerodynamic supply to the lungs and larynx, which as a result channels more "power" to the articulated speech segments. Under such physiological mechanisms, the neutral tone which should be weak by nature can gain strength and hence become stronger in articulation. As a result, its acoustic parameters such as surface F0 can change from one typical of a weak speech segment to one characteristic of a strong speech segment such as full tones in Mandarin. What is noteworthy of the

results is that there is only one case of neutral tone target change in happy speech. The reason may be that the participants in this study produced happiness in a more "subdued" way, so to speak—rather than laughing out loud or grinning widely, they just smiled slightly to convey happiness (probably due to shyness typical in Chinese culture). Therefore the happiness in this study borders more on pleasantness/calm happiness [38] which is more in line with feeling good than exhilarated. As a result, the physiological response mentioned above may not be as strong as that of the heartfelt happiness, hence the few cases of target change.

In terms of disgust, although it is often associated with such cardiorespiratory activities as decrease in heart rate [15], slow breathing [4-5], pharynx tightening [35], etc., it does not mean there should be a decrease in the aerodynamic energy supply for speech sounds produced in the emotion of disgust. This is because disgust is more often expressed in the form of affect bursts [2] which usually sound similar to vomiting, e.g., "*ah*", "*buah*", etc. It is observed that in this study the neutral tone *men5* was produced by the speakers in a burst-like way, i.e., it does not sound as smooth and uninterrupted as *men5* produced in other emotional speech conditions (i.e., angry, happy and neutral speech). Hence, it is reasonable to speculate that such "brute vocalizations" [49] tend to make the corresponding speech sounds become the "outlet", so to speak, for the oral response to vomiting. Thus there should be an increase rather than a decrease in the aerodynamic energy supply for the production of speech sounds in disgust so as to meet the physiological requirement for them to be the "outlet" for vomiting. This mechanism enables the neutral tone to gain extra power to become strong and take on a target shape similar to that of full tones as shown in this study.

5.3. The Impact of Preceding Full Tones

Another important factor in shaping the target of the neutral tone has been found to be the tonal categories of the full tones preceding the neutral tone. The most evident example is when the preceding tone is a rising tone: It is shown that in anger and disgust, the neutral tone immediately following the rising tone in sentence medial and final position uniformly takes on a target similar to that of a falling tone. The possible explanation lies in the tonal nature of the rising tone. It is known that the rising tone, in order to approximate its target [rising] according to the Target Approximation model [64, 66], needs to take on the form of a sharp rise which involves overcoming a momentum requiring more vocal effort [64]. Physiologically, such process calls for an increase in the active contraction of the cricothyroid (CT) muscle so as to raise F0 [22, 45]. At the end of the rising tone, there tends to be a CT peak (i.e., highly activated CT muscle contraction) when compared with other Mandarin tones according to the results of an EMG study [45]. After such high activation of the muscle, the elastic nature of the vocal

folds tends to call for a spontaneous relaxation of the CT muscle, which will agree with the direction of the following tone if the following tone has a falling F0 contour [45]. It is reasonable to speculate that the tenser the muscle, the greater the extent of relaxation following the muscle contraction according to physical laws. Hence there will be more "agreement", so to speak, between the state of the CT muscle after producing a rising tone and the production of the following neutral tone given the fact that the neutral tone generally takes on the direction of a descending F0 contour. Naturally, such greater extent of CT muscle relaxation tends to make the descending F0 contour fall even more sharply, thus making the following neutral tone take on a steeper F0 contour than usual, which as a result makes the neutral tone take on a target similar to that of a falling full tone as demonstrated in this study.

Another impact of the preceding full tones comes from the low tone which enables the following neutral tone to have a target similar to that of a high tone. This is associated with the nature of the low tone: In Mandarin and Cantonese the low tone has been found linked to the "post-low bouncing" mechanism in which tones, especially the neutral tone immediately following the low tone tends to have a rising F0 contour throughout the syllable [9, 19, 50]. This is particularly evident when the low tone is produced with focus [9], i.e., with extra articulatory effort. The reason has been proposed to be associated with laryngeal activity [9] due to the fact that such laryngeal muscles as sternohyoids, sternothyroids, and thyrohyoids tend to be activated if a lower than mid-level F0 is produced [1, 16, 36]. The result of such activation generated by larynx lowering is that the vocal folds tend to become shortened to some extent [1, 24] which is likely to trigger readjustment of the balance between laryngeal muscles [41]. Such mechanism tends to build up vocal fold tension which provides extra F0 raising force for speech elements [41]. Therefore the raising of the surface F0 of the neutral tone observed in this study can be regarded as being activated by such post-low bouncing mechanism. The reason why in this study such bouncing phenomenon only occurs in emotional speech rather than neutral speech may be that firstly, emotions as mentioned above tend to generate greater cardiorespiratory responses and thus greater energy supply for F0 raising in the neutral tone. Secondly, the post-low bouncing as suggested in [41] does not have to be present in all cases where L is the preceding tone. Individual differences in terms of physiology, psychology and speech preferences can all play a role in determining the final surface F0 contour of the neutral tone. So in this study the lack of post-low bouncing of the neutral tone after L tone in neutral speech may be the results of the interaction of various factors.

In terms of the falling tone, this study shows that it tends to make the following neutral tone take on a target similar to that of a falling tone in disgust speech. The reason can be associated with the influence of subglottal pressure (Ps) on F0 control. It has been shown [1, 52, 55] that when vocal folds are thick and slack (e.g., in the case of producing

Mandarin falling tone), F0 variations tend to be dominantly controlled by aerodynamic forces generated by Ps rather than the tension of vocal folds. Since as mentioned in section 5.2 the neutral tone produced in disgust speech in this study sounds more like affect bursts which are supported by more aerodynamic energy, it is highly likely that after the production of the falling tone such extra subglottal pressure generated by disgust tends to make the F0 contour of the neutral tone fall deeper to the extent of becoming similar to that of a falling tone.

5.4. The Impact of Position

The position of the neutral tone in a sentence has also shown a tendency to affect the target behavior of the neutral tone. This study shows that target change occurs more at sentence medial than final position.

This is in agreement with the previous literature on the relations between articulatory strength of speech sounds and their positions in utterances. It is well known that domain final positions are often associated with lengthening [7, 37]. Moreover, there tends to be a declination of the extent of articulation as utterances progress [23, 28, 53, 57]. The reason has been attributed to the respiratory, laryngeal and supralaryngeal weakening of the articulation system [57] which triggers "a general 'winding down' in speech" as an utterance comes to a stop [28]. Sentence initial position, however, tends to be linked to strengthening as has been demonstrated by many studies [10-11, 17, 25, 27]. Hence, the earlier a speech sound occurs in an utterance, the more likely that it has strengthened articulation [17]. Therefore the reason why the neutral tone in this study is found to have undergone target change more at sentence medial than final position is that in the medial position there is more articulatory support of strength for the weak speech element to become strong than in the final position.

On the other hand, the presence of target change of the neutral tone at sentence final position in angry and disgust speech suggests that it is the other two factors (emotion and the preceding tones) that have played a role. Hence the position of the neutral tone in those cases does not weigh significantly in shaping its tonal target.

6. Conclusions

Articulatorily weak elements such as Mandarin neutral tone are shown in this study to have a strong tendency of undergoing a process of target change due to emotional, tonal and positional variations. The findings have not only lent further support to the influence of emotions, tones and sentence positions on shaping the phonetic behavior of articulatorily weak elements reported in previous literature, but more importantly have revealed the significance of articulatory strength as an index for speech communication. Since the relations between articulatory strength and speech communication are relatively less well explored than other arrears of speech [26, 62], it is hoped that this study can

serve as another piece of evidence showing the importance of this line of research. Due to the fact that this study has focused only on anger, happiness and disgust, other emotions such as fear, sadness and surprise may present a different picture in terms of the target behavior of the neutral tone. Other paralinguistic factors such as attitudes may also play a role. In addition, strength-related speech phenomena in other languages (e.g., schwa in English) would render the picture of articulatory strength more complete. Future work along those lines is thus needed to not only further improve our understanding of speech communication but also to better facilitate computational simulation of speech.

Acknowledgements

References

[1]   Atkinson, J. E. (1978). Correlation analysis of the physiological factors controlling fundamental voice frequency. *Journal of the Acoustical Society of America*, 63, 211–222.

[2]   Banse, R., and Scherer, K. R. (1996). Acoustic Profiles in Vocal Emotion Expression. *Journal of Personality and Social Psychology*, 70(3), 614-636.

[3]   Boersma, P., and Weenink, D. (2013). Praat: doing phonetics by computer [Computer program]. Version 5.3.59, retrieved 3rd Janurary 2013 from http://www.praat.org/

[4]   Boiten, F.A. (1998). The effects of emotional behaviour on components of the respiratory cycle. *Biological Psychology*, 49 (1–2), 29–51.

[5]   Boiten, F. A., Frijda, N. H., and Wientjes, C. J. E. (1994). Emotions and respiratory patterns: Review and critical analysis. *International Journal of Psychophysiology*, 17, 103-28.

[6]   Bruce, G. (1977). Swedish word accents in sentence perspective. In Malmberg, B. and Hadding, K. (eds.) *Travaux de l'institute de linguistique de lund XII*. Lund, Gleerup.

[7]   Byrd, D., Kaun, A., Narayanan, S., and Saltzman, E. (2000). Phrasal signatures in articulation. In Broe, M. B. and Pierrehumbert, J. B. (eds.), *Papers in Laboratory Phonology V,* pp.70-87. Cambridge: Cambridge University Press.

[8]   Chao, Y. R. (1968). *A Grammar of Spoken Chinese*. University of California Press, Berkeley, California.

[9]   Chen, Y., and Xu, Y. (2006). Production of weak elements in speech: Evidence from F(0) patterns of neutral tone in Standard Chinese. *Phonetica*, 63 (1), 47–75.

[10] Cho, T., McQueen, J. M., and Cox, E. (2007). Prosodically driven phonetic detail in speech processing: The case of domain-initial strengthening in English. *Journal of Phonetics*, 35, 210-243.

[11] Cho, T., and Keating, P. (2001). Articulatory and acoustic studies of domain-initial strengthening in Korean. *Journal of Phonetics*, 29, 155–190.

[12] Cowie, R., and Cornelius, R. (2003). Describing the emotional states that are expressed in Speech. *Speech Communication*, 40, 5-32.

[13] Darwin, C. (1872). *The Expression of the Emotions in Man and Animals*. London, England: John Murray.

[14] Ekman, P. (1999). Basic emotions. In Dalgleish, T. and Power, T. (eds.), *The handbook of cognition and emotion*, pp. 45-60. New York: John Wiley & Sons.

[15] Ekman, P., Levenson, R. W., and Friesen, W. V. (1983). Autonomic nervous system activity distinguishes among emotions. *Science*, 221, 1208-1210.

[16] Erickson, D. (2011). Thai tones revisited. *Journal of the Phonetic Society of Japan*, 15(2), 1-9.

[17] Fougeron, C., and Keating, P. A. (1997). Articulatory strengthening at edges of prosodic domains. *Journal of the Acoustical Society of America*, 101(6), 3728–3740.

[18] Gandour, J. T., Potisuk, S., and Dechongkit, S. (1994). Tonal coarticulation in Thai. *Journal of Phonetics*, 22, 477–492.

[19] Gu, W., and Lee, T. (2009). Effects of tone and emphatic focus on F0 contours of Cantonese speech—A comparison with standard Chinese. *Chinese Journal of Phonetics*, 2, 133–147.

[20] Hammerschmidt, K., and Jürgens, U. (2007). Acoustical correlates of affective prosody. *Journal of Voice*, 21, 531–540.

[21] Hawkins, S., and Smith, R. (2001). Polysp: A polysystemic, phonetically-rich approach to speech understanding. *Journal of Italian Linguistics – Rivista di Linguistica*, 13, 99–188.

[22] Herman, R., Beckman, M. E., and Honda, K. (1999). Linguistic models of *F*0 use, physiological models of *F*0 control, and the issue of "mean response time". *Language and Speech*, 42, 373-399.

[23] Hinton, V. A. (1996). Interlabial pressure during production of bilabial phones. *Journal of Phonetics*, 24, 337–349.

[24] Honda, K., Hirai, H., Masaki, S., and Shimada, Y. (1999). Role of vertical larynx movement and vertical lordosis in F0 control. *Language and Speech*, 42,401–411.

[25] Jun, S.-A. (1993). *The phonetics and phonology of Korean prosody*. Ph.D. dissertation, Ohio State University.

[26] Keating, P. A. (2006). Phonetic encoding of prosodic structure. In Harrington, J. and Tabain, M. (eds.), *Speech production: Models, phonetic processes, and techniques*, pp. 167–186. New York and Hove: Psychology Press.

[27] Keating, P. A., Cho, T., Fougeron, C., and Hsu, C. (2003). Domain-initial strengthening in four languages. In Local, J., Ogden, R., and Temple, R. (eds.), *Papers in laboratory phonology 6: Phonetic interpretations*, pp.145–163. Cambridge, UK: Cambridge University Press.

[28] Krakow, R. A., Bell-Berti, F., and Wang, Q. E. (1994). Supralaryngeal declination: evidence from the velum. In Bell-Berti, F. and Raphael, L. (eds.), *Producing Speech: a Festschrift for Katherine Safford Harris*, pp.333-353.

Woodbury NY: AIP press.

[29] Kwon, O. W., Chan, K., Hao, J., and Lee, T. W. (2003). Emotion Recognition by Speech Signals. In *Proceedings of Eurospeech*, 125-128. Geneva, Switzerland.

[30] Li, Z. (2003). *The phonetics and phonology of tone mapping in a constraint-based approach*. PhD dissertation. MIT, Cambridge, Massachusetts.

[31] Li, A., Fang, Q., Hu, F., Zheng, L., Wang, H., and Dang, J. (2010). Acoustic and Articulatory Analysis on Mandarin Chinese Vowels in Emotional Speech. In *7th International Symposium on Chinese Spoken Language Processing*, 38-43. Tainan, Taiwan.

[32] Liu, F., Xu, Y., Prom-on, S., and Yu, A. C-L. (2013). Morpheme-like prosodic functions: Evidence from acoustic analysis and computational modelling. *Journal of Speech Sciences*, 3 (1), 85-140.

[33] Morrison, D., Wang, R., and De Silva, L.C. (2007). Ensemble methods for spoken emotion recognition in call-centres. *Speech Communication*, 49, 98–112.

[34] Murray, I. R., and Arnott, J. L. (1993). Toward the simulation of emotion in synthetic speech: A review of the literature on human vocal emotion. *Journal of the Acoustical Society of America,* 93, 1097-1108.

[35] Noble, L., and Xu, Y. (2011). Friendly speech and happy speech: Are they the same? In *Proceedings of the 17th International Congress of Phonetic Sciences*, 1502–1505. Hong Kong.

[36] Ohala, J. J. (1972). How is pitch lowered? *Journal of the Acoustical Society of America*, 52, 124.

[37] Oller, D. K. (1973). The effect of position in utterance on speech segment duration in English. *Journal of the Acoustical Society of America*, 54**,** 1235–1247.

[38] Philippot, P., Chapelle, C., and Blairy, S. (2002). Respiratory feedback in the generation of emotion. *Cognition and Emotion*, 16, 605-627.

[39] Pierrehumbert, J., and Talkin, D. (1992). Lenition of /h/ and glottal stop. In Docherty, D. R. and Ladd, D. (eds.), *Papers in Laboratory Phonology II*, pp. 90-117. London: Cambridge University Press.

[40] Prom-on, S., Liu, F., and Xu, Y. (2011). Functional modeling of tone, focus and sentence type in Mandarin Chinese. In *Proceedings of the 17th International Congress of Phonetic Sciences*, 1638–1641. Hong Kong.

[41] Prom-on, S., Liu, F., and Xu, Y. (2012). Post-low bouncing in Mandarin Chinese: Acoustic analysis and computational modelling. *Journal of Acoustical Society of America*, 132, 421-432.

[42] Prom-on, S., and Xu, Y. (2012). PENTATrainer2: A hypothesis-driven prosody modeling tool. In *Proceedings of the 5th ISEL Conference ExLing*, 27-29, Athens, Greece.

[43] Prom-on, S., Xu, Y., and Thipakorn, B. (2009). Modeling tone and intonation in Mandarin and English as a process of target approximation. *Journal of the Acoustical Society of America*, 125, 405-424.

[44] Rainville, P., Bechara, A., Naqvi, N., and Damasio, A.R. (2006). Basic emotions are associated with distinct patterns

of cardiorespiratory activity. *International Journal of Psychophysiology*, 61, 5–18.

[45] Sagart, L., Halle, P., de Boysson-Bardies, B., and Arabia-Guidet, C. (1986). Tone production in Modern Standard Chinese: An electromygraphic investigation. *Cahiers Linguistique Asie-Orientale*, 15, 205-221.

[46] Scherer, K. R. (1986). Vocal affect expression: A review and a model for future research. *Psychological Bulletin*, 99, 143-165.

[47] Scherer, K. R. (2003). Vocal communication of emotion: A review of research paradigms. *Speech Communication*, 40, 227-256.

[48] Scherer, K. R. (2013). Vocal markers of emotion: Comparing induction and acting elicitation. *Computer Speech and Language*, 27(1), 40-58.

[49] Scherer, K. R., and Zentner, M. (2001). Emotion effects of music: Production rules. In Juslin, P. and Sloboda, J. (eds.), *Music and emotion: Theory and research*, pp. 361–392. Oxford, England: Oxford University Press.

[50] Shen, J. (1994). Hanyu yudiao gouzao he yudiao leixing (Intonation structures and patterns in Mandarin). *Fangyan (Dialect)*, 3, 221–228.

[51] Shih, C. (1987). *The phonetics of the Chinese tonal system*. Bell Laboratories, Tech. Merevt.

[52] Stevens, K. N. (1963). Theory of Vocal-Cord Vibration and its Relation to Laryngeal Features. *Journal of the Acoustical Society of America*, 55, 383(A).

[53] Vaissière, J. (1986). Comment on Abbs's Paper. In Perkell, J.S. & Klatt, D.H. (eds.), *Invariance and Variability in Speech Processes*, pp. 220-222. Hillsdale, NJ: LEA.

[54] van Bezooijen, R. (1984). *The characteristics and recognizability of vocal expressions of emotion*. Dordrecht, The Netherlands: Foris.

[55] van den Berg, J. W., and Tan, T. S. (1959). Results of Experiments with Human Larynxes. *Practica oto-rhino-laryngologica*, 21, 425-450.

[56] van Santen, J., and Möbius, B. (1999). A quantitative model of F0 generation and alignment. In Botinis, A.(ed.), *Intonation: Analysis, Modeling and Technology*, pp.269-288. Dordrecht, Netherlands: Kluwer Academic Publishers.

[57] Vayra, M., and Fowler, C. (1992). Declination of supralaryngeal gestures in spoken Italian. *Phonetica*, 49, 48–60.

[58] Ververidis, D., and Kotropoulos, C. (2006). Emotional speech recognition: Resources, features, and methods. *Speech Communication*, 48, 1162-1181.

[59] Xu, Y. (1997). Contextual tonal variations in Mandarin. *Journal of Phonetics*, 25, 61-83.

[60] Xu, Y. (1999). Effects of tone and focus on the formation and alignment of F0 contours. *Journal of Phonetics*, 27, 55–105.

[61] Xu, Y. (2004). Understanding tone from the perspective of production and perception. *Language and Linguistics*, 5, 757-797.

[62] Xu, Y. (2005). Speech melody as articulatorily implemented communicative functions. *Speech Communication*, 46, 220-251.

[63] Xu, Y. (2006). Tone in connected discourse. In K. Brown (ed.), *Encyclopedia of Language and Linguistics*, 2nd Ed. Oxford: Elsevier. 12: 742-750.

[64] Xu, Y. (2009) Timing and coordination in tone and intonation—An articulatory-functional perspective. *Lingua*, 119, 906-927.

[65] Xu, Y. (2005–2013). ProsodyPro.praat, http://www.phon.ucl.ac.uk/home/yi/ProsodyPro

[66] Xu, Y., and Wang, Q. E. (2001). Pitch targets and their realization: Evidence from Mandarin Chinese. *Speech Communication*, 33, 319–337.