CHARACTERISING INTONATION IN PLASTIC MANDARIN USING POLYNOMIAL MODELLING

Chenzi Xu

University of Oxford
chenzi.xu@ling-phil.ox.ac.uk

ABSTRACT

This paper presents an experimental investigation of pitch variations of tone sequences under different contexts in Plastic Mandarin, a newly-formed contact-induced regional Mandarin variety. Instead of describing f_0 variations of a single tone-bearing unit, the f_0 contours of trisyllabic phrases varying in tonal composition, utterance position, and focus condition were examined by building orthogonal polynomial models on production data elicited from connected semi-spontaneous natural speech. The coefficients of the models that capture the shape of f_0 contours were subjected to further linear mixed effects analysis. This study reveals the synergistic effects of tone interactions, focus, and declination, exerted on the f_0 of the whole phrase in an integrated manner. The findings enhance our knowledge of a prevalent yet scarcely studied Mandarin variety and contribute to the on-going research into the dynamic realisation of tone and intonation in connected speech.

Keywords: Plastic Mandarin, polynomial modelling, intonation, contextual tonal variation.

1. INTRODUCTION

Prosodic information including pitch contours play a crucial role in discriminating language varieties [23]. Plastic Mandarin, for example, impressionistically differs from Standard Mandarin in pitch patterns. This study examines the contextual tonal variations in Plastic Mandarin, particularly downstep phenomena.

1.1. Plastic Mandarin

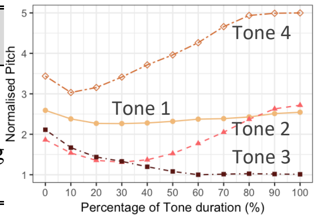
Language variation and change are concomitant with the official promotion of Standard Mandarin throughout China when it comes into contact with local dialects.

In the city of Changsha in Hunan province, China, a local Mandarin variety has emerged, termed *Plastic Mandarin*, in which *plastic* connotes the sense of being inauthentic. Plastic Mandarin here refers to a particular non-standard variety with distinct intonation in Hunan. It is predominantly used in schools in urban Changsha, serving as the *de facto* medium of communication, though Standard Mandarin is expected by policy. This urban youth speech has been crystallised and is used in an increasingly wide range of social circumstances.

The paucity of work on Plastic Mandarin prompts this study to perform instrumental phonetic analysis on naturalistic data. Plastic Mandarin has the same tonal categories as Standard Mandarin (1-4). The pitch patterns, however, are different. Our prior research established the canonical citation forms of Plastic Mandarin tones by analysing utterances of monosyllable /i/ in four tones by eight speakers. Table 1 summarises the tone system of Plastic Mandarin based on the empirical data, shown in the figure below where the average of normalised pitch values (in semitones scaled from 1 to 5) of eight speakers was plotted at 10% intervals. Chao's tone letters [3] were adopted in the description of pitch values.

Table 1: Descriptions of Plastic Mandarin citation tones based on empirical data

Tone	Pitch value		Description
1	˧	M	(lower) mid level
2	˩˧	LM	low to mid rising
3	˨˩	L	low-mid to low falling
4	˧˥	MH	mid to high rising



1.2. Studies on tone and intonation

The surface f_0 realisation of connected Mandarin speech is never as straightforward as concatenating contours of corresponding lexical tones [10] and the surface tone pattern of a morpheme varies in different prosodic contexts. Such f_0 variation has been a locus of integrated research involving phonological aspects of a language, biomechanical aspects of speech production, and information structure in communication. Previous research has demonstrated the effect of the preceding and following tones [18, 25], sentence types [26], utterance positions [15], focus and topic [4, 24], in various languages. Many studies, though, worked on individual factors without examining their interactions, and were based on laboratory read speech or sometimes nonsense carrier sentences (e.g. [25]).

The present study investigates the effect of tonal combination, focus, and utterance position on f_0 realisation of trisyllabic phrases in connected speech of Plastic Mandarin, using quantitative modelling.

2. METHOD

2.1. Recordings

The recordings were obtained in my fieldwork in Changsha, Hunan, China in September 2017. This study focuses on twelve trisyllabic phrases concatenated into complete sentences, as displayed in Table 2. In each phrase, the first and the third syllables bear tone 4, which has the highest pitch target among the four tones, while the medial syllable was varied. Syllables with high nuclear vowels including [i], [ɪ], [ʊ], [u], and [y] were mainly selected for the medial syllables to minimise variability induced by vowel intrinsic f_0 [12]. Efforts were made to include as few aspirated or voiceless prevocalic obstruent consonants as possible, in order to avoid their f_0 perturbation effects [27].

Table 2: Trisyllabic Stimuli.

	Tone Sequence	Test phrase	IPA	Gloss
Character	4 1 4	yi gong jiang	[i kʊŋ ʨiaŋ]	Craftsman Yi
	4 2 4	yi ni jiang	[i ni ʨiaŋ]	Mason Yi
	4 3 4	yi tie jiang	[i tiɛ ʨiaŋ]	Blacksmith Yi
	4 4 4	yi mu jiang	[i mu ʨiaŋ]	Carpenter Yi
Location	4 1 4	zai yi yuan	[ʦaɪ i yɛn]	in the hospital
	4 2 4	zai ting yuan	[ʦaɪ tʰɪŋ yɛn]	in the courtyard
	4 3 4	zai ying yuan	[ʦaɪ ɪŋ yɛn]	in the cinema
	4 4 4	zai xi yuan	[ʦaɪ ɕi yɛn]	in the theatre
Action	4 1 4	kan su ju	[kʰan su ʨy]	watch Suzhou opera
	4 2 4	kan lu ju	[kʰan lu ʨy]	watch Luzhou opera
	4 3 4	kan wu ju	[kʰan u ʨy]	watch dance opera
	4 4 4	kan yu ju	[kʰan y ʨy]	watch Henan opera

Recordings of 7 speakers (4 females and 3 males) were used in the present study. They were high school students aged 16 in Changsha Nanya Middle School. The elicitation was conducted through a card game, which provided a relaxing context for speakers to comfortably speak Plastic Mandarin, and kept them maximally engaged. Formal elicitation was avoided because students were likely to switch to Standard Mandarin in a formal interview setting, as they have been trained for years. Participants in pairs were given three sets of cards: Character, Location, and Action, and a fully meaningful test phrase as in Table 2 was printed on each card. In this game, one participant picked one card from each set to form a sentence and displayed only two to their friend, and the other participant guessed at the hidden card. The procedure is exemplified in Table 3: in each response (B), one test word was in focus (underlined), and the narrow focus was placed on the medial syllable in a phrase (bolded), which is the new information in this discourse situation.

Factors of tonal combination, positions in an utterance (initial, middle, and final), and focus conditions (focus and non-focus) were therefore manipulated in our analysis of contextual tonal variations. In total, there are 4 (tonal combinations) × 3 (position) × 2 (focus) = 24 utterance patterns. Due to the uncertain nature of the guessing game, we obtained different number of repetitions of each pattern, and 232 tokens in total. Sound files were labelled and segmented in Praat [2].

Table 3: Demonstration of the card game.

	Conversation	Translation
A	[ʂeɪ ʦaɪ ɕi yɛn kʰan u ʨy]	Who is watching dance opera in the theatre?
B	[i **mu** ʨiaŋ ʦaɪ ɕi yɛn kʰan u ʨy]	Carpenter Yi is watching dance opera in the theatre.
A	[pu tweɪ]	Not correct.
B	[i **ni** ʨiaŋ ʦaɪ ɕi yɛn kʰan u ʨy]	Mason Yi is watching dance opera in the theatre.
A	[tweɪ la]	Correct!

2.2. Polynomial modelling

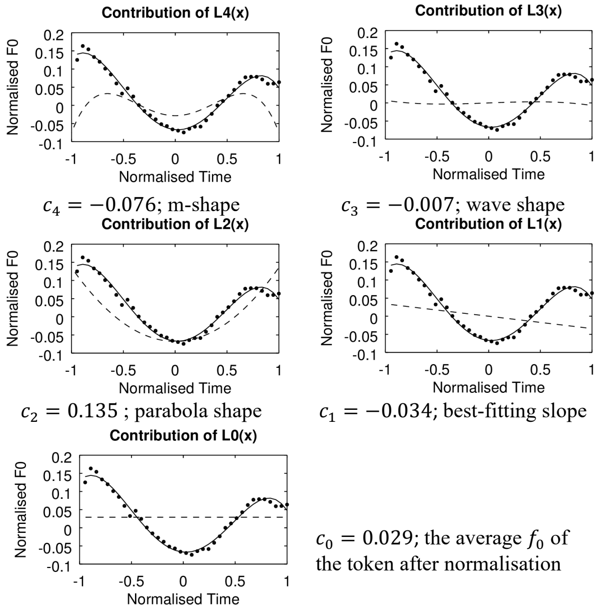
Polynomials have recently been employed to model f_0 contours in speech [1, 7]. This study adopted the approach in [7] using orthogonal (Legendre) polynomials, raising the number of coefficients to five to model the shape of these trisyllabic phrases. Alternative models including PENTA model [19] and CR model [8] were not chosen given their predetermined assumptions about f_0 production.

The model was specified by a best-fit sum of coefficients, c_i, that multiply Legendre polynomials normalised to unit variance, $L_i(t)$:

$$\hat{f}_0(t) = \sum_i c_i \cdot L_i(t) \tag{1}$$

f_0 estimates in 10 millisecond intervals of voiced regions were obtained using the *get_f0* program from the ESPS package [6]. The corresponding time values were linearly scaled to the range of -1 to 1.Prior to the modelling, all f_0 values of a speaker were normalised to his/her mean f_0 and centred around zero.

Figure 1: Contribution of the five Legendre polynomials in characterising the pitch contour of the token [kʰan˥ su˧ ʥy˥]



$c_4 = -0.076$; m-shape

$c_3 = -0.007$; wave shape

$c_2 = 0.135$; parabola shape

$c_1 = -0.034$; best-fitting slope

$c_0 = 0.029$; the average f_0 of the token after normalisation

The interpretation of the coefficients is illustrated through the contribution of corresponding Legendre polynomial terms (dashed lines) to the f_0 shape in Figure 1. Each plot shows the original f_0 measures in black dots and the fitted quartic polynomial in a black line as a reference. Higher-ranking polynomials successively pick out varying properties on a smaller scale [7].

2.3. Statistical analysis

The five coefficients of a polynomial model transform qualitative descriptions of a shape of pitch contour into interpretable quantitative estimates. Statistical models of f_0 contour shape predicted by contextual factors were built to investigate the way speakers manipulate f_0 to convey lexical meanings and discourse salience.

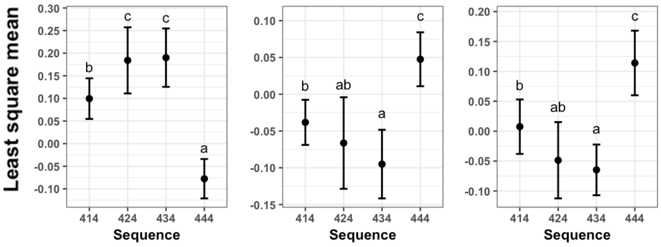
Linear Mixed Effects (LME) models were adopted because they are suitable for a repeated-measures design from multiple subjects and can deal with uneven sample sizes or unbalanced design. The data contained an uneven number of tokens of each type of utterance due to the uncertainties in the card game. In each LME model, the independent fixed-effect predictors were (1) TONE SEQUENCE (4 levels: 414, 424, 434, and 444), (2) POSITION (3 levels: Initial, Medial, and Final), (3) FOCUS (2 levels: Focus and Non-focus), with SUBJECT being a crossed random factor. The dependent variables were the Legendre coefficients (i.e. c_4, c_3, c_2, c_1, or c_0). Interactions between all possible pairs of predictors were also included in the model. The final models were, derived from top-down modelling, the most complex models properly supported by the amount of data. For significant variance components, predictions in Least-squares Means from LME models and post hoc Tukey-adjusted pairwise comparisons were obtained to summarise the effects of factors and their contrasts, using the R [20] package 'emmeans' [13].

3. RESULTS

According to the LME models, the tone of the varied middle syllable in each phrase is one of the significant contributors to shape differences in the aspects of Parabola (F(3, 3.091)=20.02, p<.05), Overall Slope (F(3, 3.14)=13.34, p<.05), and Average f_0 (F(3, 3.05)=14.72, p<.05) of the f_0 contour, apart from its significant interactions with position and focus.

Figure 2 shows the predicted c_2, c_1, and c_0 values (c_3, c_4 are not significant) for the four types of sequences and their pairwise comparison results. Phrases with three consecutive Tone 4s are very distinguished from the other types of phrases, with a large rise in the centre of the phrase ($c_2 < 0$), rising overall slope ($c_1 > 0$), and significantly higher average f_0 ($c_0 > 0.1$). In the following sections, we examine firstly how tones interact in general and then the effects of utterance position and focus.

Figure 2: Tonal sequence of phrases distinguished significantly by the coefficients of c_2, c_1, c_0.



c_2 Parabola c_1 best-fit slope c_0 Average f_0

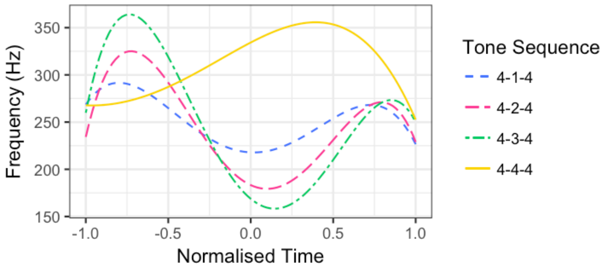
Error bars indicate 95% confidence interval of the Least Square Mean. Means sharing a same letter are not significantly different.

3.1. Downstep and upstep

Downstep, the lowering of a high tone following a sequence of a high tone and a non-high tone, has been observed in many languages such as Kikuyu [5], English [16], Japanese [17], Mandarin [22], and Yoruba [11], among others.

That the c_1 predictions in the model are all negative for non-444 sequences suggests similar downstep phenomena in Plastic Mandarin. All tones other than the higher tone 4, no matter rising or falling, triggered the downstep effect, especially tone 3 (˨˩) with the lowest pitch level which has the smallest c_1, suggesting the greatest downstep effect. This is exemplified in Figure 3, presenting the polynomial models for different sequences produced at sentence-initial position in focus by one speaker. The second peaks of non-444 sequences are lower than the first peaks, most saliently for the 434 sequence, in which the second peak is about 90 Hz lower than the first.

Figure 3: Utterance-initial focused phrases by a speaker

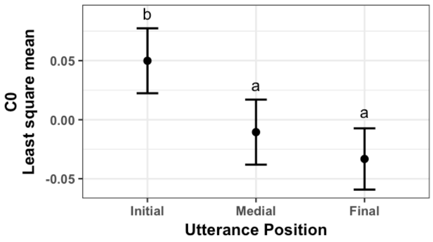


In this figure, the f_0 contour of the 444 sequence indicates an upstep effect, where the middle high tone has a higher tone target than two other high tones.

3.1. Interaction of downstep and position

The factor POSITION significantly predicts the variability of c_0 (p<.001). Figure 4 illustrates that the average f_0 of phrases at utterance-initial position is significantly higher than at other positions, with sequence and focus effects being averaged. It shows a significant f_0 drop after the initial phrase, and overall f_0 decreases over the utterance.

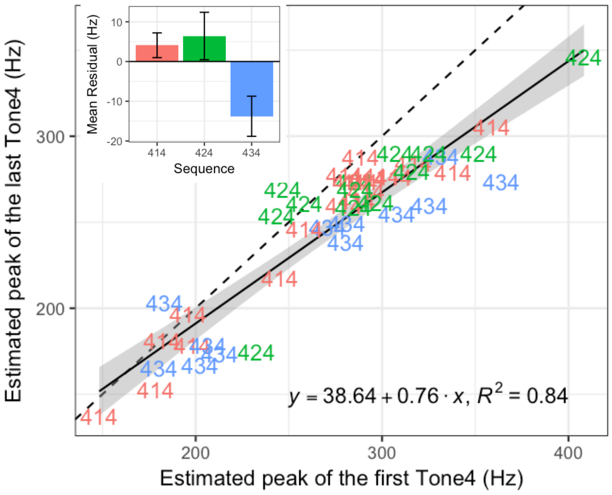
Figure 4: Predictions and pairwise comparisons of c_0 in three utterance positions



Error bars indicate 95% confidence interval of the Least Square Mean. Means sharing a same letter are not significantly different.

Figure 5 examines the downstep effect in non-444 sentence-initial sequences by measuring the peaks or local maximum f_0 values of tone 4, which are approximated by the crest or endpoint of rising or falling in fitted curves. A regression line was fitted to all data points.

Figure 5: Peak f_0 values in sentence-initial phrases by 5 speakers and mean residual values from the regression in phrases by sequence.



Plotted against the dashed reference line $y = x$. The black line is a regression line fitted to all data points with shaded area being the 95% confidence interval. The vertical lines in the bar graph indicate the standard error of each mean. (No complete data for the other 2 speakers).

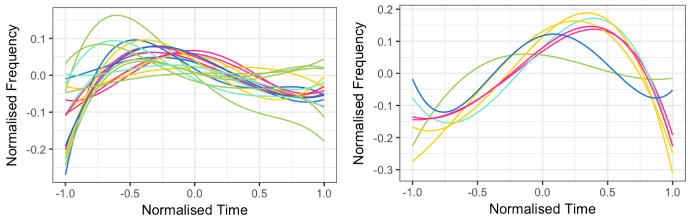
Utterances at sentence-medial and sentence-final positions were also examined and compared. The regression function is $y = 35.97 + 0.78x$, $R^2 = 0.78$ and $y = -0.38 + 0.92x$, $R^2 = 0.89$ respectively.

All regression line slopes are less than 1, and the majority of the data points are below the reference line $y = x$, indicating a generally lower second peak. The regression line slope for sentence-final phrases (0.92) is larger than its sentence-initial counterpart (0.76) in this dataset, suggesting a reduced peak difference in utterance final positions. Such decrease of the downstep effect over an utterance was also observed in English, and modelled as a decaying exponential function by [14].

3.2. Interaction of upstep and focus

The effect of focus is most salient for 444 sequences, manifested by significantly different c_2, c_1, c_0 values predicted by the LME models. Focus on the middle syllable strengthens the upstep effect, resulting in significantly higher overall f_0, and a larger and delayed rise in the centre of the phrase, demonstrated in Figure 5. The f_0 contour of the focused middle syllable *mu* resembles the prototypical citation form˥.

Figure 5: Focus effect for all utterance-initial tokens of '*yi*˥ *mu*˥ *jiang*˥' by 5 speakers



Non-Focus Focus

4. DISCUSSION

This paper concentrated on the downstep and upstep phenomena in Plastic Mandarin by creating various tri-tonal contexts within a phrase.

Tone contours are sensitive to local tonal contexts, one of the major sources of the influence on tonal variation [9,10]. There are differences in the direction and magnitude of coarticulatory effects in the interaction of tones (e.g. [18], [21]), parallel to cross-linguistic differences in segmental coarticulation. Our analysis of tone interactions suggests that the varied middle tone can have both rightward and leftward (anticipatory) effects on the f_0 of the adjacent tones in Plastic Mandarin. The non-high middle tone triggers downstep so that the peak of the following high tone is lower than the preceding one, and its pitch value influences the degree of downstep. Lower middle tones often relate to a higher maximum f_0 of the preceding high tone. Our data also indicate downdrift over an utterance. Further research can concentrate on its interaction with downstep.

Utterance position and focus are neither simply translated to f_0 declining or rising, nor uniformly applied to any sequences. They interact with different tonal sequences. However diversely realised in form, the function of focus tends to be similar—boosting contrasts or distinctiveness of tones. This is consistent with the finding of [4] on Standard Mandarin that f_0 contours are more informative under focus. Further research may involve a comparison between Standard Mandarin and Plastic Mandarin intonation.

5. REFERENCES

[1] Andruski, J. E., Costello, J. 2004. Using polynomial equations to model pitch contour shape in lexical tones: An example from Green Mong. *Journal of the International Phonetic Association,* 34(2), 125-140.

[2] Boersma, P., Weenink, D.1992-2017. Praat: doing phonetics by computer [Computer program]. Version 6.0.25, retrieved from http://www.praat.org/.

[3] Chao, Y. 1930. A system of tone letters. *Le maître phonétique,* 45, 24-27.

[4] Chen, Y., Gussenhoven, C. 2008. Emphasis and tonal implementation in Standard Chinese. *Journal of Phonetics,* 36, 724–746.

[5] Clements, G. N., Ford, K. C. 1979. Kikuyu tone shift and its synchronic consequences. *Linguistic Inquiry,* 10(2), 179-210.

[6] Entropic Signal Processing System, release of the Phonetics Lab, University of Oxford. Downloaded from: http://www.phon.ox.ac.uk/releases.

[7] Grabe, E., Kochanski, G., Coleman, J. 2007. Connecting Intonation Labels to Mathematical Descriptions of Fundamental Frequency. *Language and Speech,* 50(3), 281-310.

[8] Fujisaki, H., Wang, C., Ohno, S., Gu, W. 2005. Analysis and synthesis of fundamental frequency contours of Standard Chinese using the command-response model. *Speech Communication,* 47, 59-70.

[9] Kochanski, G. P., Shih, C. 2000. Stem-ML: Language -Independent Prosody Description. *6th International Conference on Spoken Language Processing.* Beijing: ISCA Archive.

[10] Kochanski, G., Shih, C., Jing, H. 2003. Hierarchical structure and word strength prediction of Mandarin prosody. *International Journal of Speech Technology,* 6, 33-43.

[11] Laniran, Y. O., Clements, G. 2003. Downstep and high raising: interacting factors in Yoruba tone production. *Journal of Phonetics,* 31, 203-250.

[12] Lehiste, I., Peterson, G. E. 1961. Some basic considerations in the analysis of intonation. *The Journal of the Acoustical Society of America,* 33(4), 419-425.

[13] Lenth, R. 2016. Least-Squares Means: The R Package lsmeans. *Journal of Statistical Software,* 69(1), 1-33.

[14] Liberman, M., Pierrehumbert, J. 1984. Intonational invariance under changes in pitch range and length. In M. Aronoff, & R. T. Oehrle, *Language sound structure: studies in phonology | presented to Morris Halle by his teacher and students* (pp. 157-233). Cambridge: MIT Press.

[15] Pierrehumbert, J. 1979. The perception of fundamental frequency declination. *Journal of the Acoustical Society of America,* 66(2), 363-369.

[16] Pierrehumbert, J. B. 1980. *The phonology and phonetics of English Intonation. Doctoral dissertation.* Cambridge, Massachusetts: MIT.

[17] Poser, W. J. 1985. The phonetics and phonology of tone and intonation in Japanese. Retrieved from DSpace@MIT: http://hdl.handle.net/1721.1/15169.

[18] Potisuk, S., Gandour, J., Harper, M. P. 1997. Contextual variations in trisyllabic sequences of Thai tones. *Phonetica,* 54, 22-42.

[19] Prom-on, S., Xu, Y., Thipakorn, B. 2009. Modelling tone and intonation in Mandarin and English as a process of target approximation. *The Journal of the Acoustical Society of America,* 125(1), 405-424.

[20] R Core Team. 2017. R: A language and environment for statistical computing. Retrieved from https://www.R-project.org.

[21] Shen, X. S. 1992. On tone sandhi and tonal coarticulation. *International Journal of Linguistics,* 25(1), 83-94.

[22] Shih, C. 1988. Tone and intonation in Mandarin. *Working Papers of the Cornell Phonetics Laboratory,* 3, 83-109.

[23] Vicenik, C., Sundara, M. 2013. The role of intonation in language and dialect discrimination by adults. *Journal of Phonetics* 41, 297-306.

[24] Wang, B., Xu, Y. 2011. Differential prosodic encoding of topic and focus in sentence-initial position in Mandarin Chinese. *Journal of Phonetics,* 39, 595-611.

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

[26] Xu, Y., Xu, C. X. 2005. Phonetic realization of focus in English declarative intonation. *Journal of Phonetics,* 33, 159–197.

[27] Xu, Y., Wang, M. 2009. Organizing syllables into groups — Evidence from F0 and duration patterns in Mandarin. *Journal of Phonetics,* 37, 502-520.