Formulating the Identification of Mandarin Tone2 and Tone3 in multi-dimensional spaces

Wu, J.

Citation

Wu, J. (2011). Formulating the Identification of Mandarin Tone2 and Tone3 in multi-dimensional spaces. *International Congress Of Phonetic Sciences (Icphs 2011)*, 2165-2168. Retrieved from https://hdl.handle.net/1887/19071

FORMULATING THE IDENTIFICATION OF MANDARIN TONE2 AND TONE3 IN MULTI-DIMENSIONAL SPACES

Junru Wu

Dept. of Chinese Language and Literature, Peking University, Beijing, China
yuerr@163.com

ABSTRACT

This study examines and formulates the effect of acoustic parameters on the identification of Mandarin Tone2 and Tone3, which are timing of turning point (T_t), falling of F0 (F0_△), duration (T_d), and precursor register (Register). Three stimulus matrixes of tone tokens are set up with T_t and F0_△. Subjects identify tone tokens. In Experiment 1, matrixes are different in T_t or T_t/Td. In Experiment 2 a same target matrix is preceded by different precursor registers. Identification rate of Tone2 (P_I) is calculated. Repeated ANOVA is applied on mean P_I across each matrix. Logistic regressions between P_I and different predictors are adopted. The result shows: (1) both T_t and F0_△ are negatively relevant to P_I and interacts in a quasi-linear way; (2) both T_t and T_t/T_d, with negative effects, should be taken as independent predictors; (3) negative effect of Register is confirmed, but it works as a quadratic reciprocal predictor.

Keywords: tone, Mandarin, identification, logistic regression

1. INTRODUCTION

One of the challenges of psychological science and speech science is how tonal categories are identified. One well established approach for this question is the "Categorical perception" (CP) paradigm[8]. However, most of the generalizations over the tonal categoricality have been based on a single specific acoustic continuum, while several acoustic measurements have been proved to affect tonal perception, such as fundamental frequency cues (F0) [9], timing [4, 5], amplitude [4, 13], and (4) harmonic structure [4, 10, 13], (5) Phonation [15]. How to incorporate necessary factors into one formulation remains a problem.

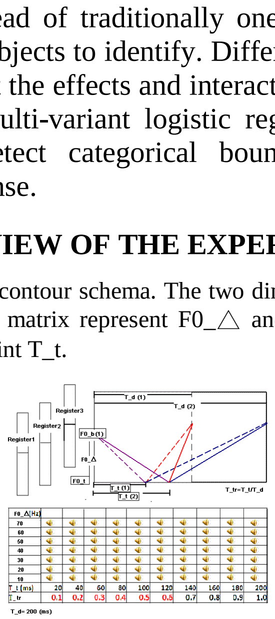
The goal of the present study is to formulate the identification of Mandarin Tone 2 and Tone 3 in a multi-dimensional space.

Both Mandarin Tone2 and Tone3 have a falling phase followed by a rising phase of F0. According to former studies, F0 is both necessary and sufficient [4, 13] for tonal recognition and its contour should be long enough for its detection [7]. More specifically, the following acoustic parameters have been reported responsible: (1) timing of the turning point, negatively related to the identification of Tone2 [10, 12, 17]; (2) the change of F0 of falling phase – the larger it is, the less likely the stimuli is identified as Tone2 [11, 12]; (3) duration – by elongating the syllable while keeping the proportion between the falling phase and the syllable duration, the identification of Tone2 decreased; this was interpreted as evidence for the essentiality of absolute duration of the falling phase [1]; (4) Precursor register– following a higher precursor, a same stimulus is more likely to be identified as Tone 3 [11]; (5) phonation types can also bias tonal perception [15] , but it is not discussed in the present study.

In the present study, 2-D acoustic stimulus matrixes (instead of traditionally one continuum) are built for subjects to identify. Different matrixes are built to test the effects and interactions of other parameters. Multi-variant logistic regressions are applied to detect categorical boundaries in a generalized sense.

2. OVERVIEW OF THE EXPERIMENTS

Figure 1: F0 contour schema. The two dimensions of each stimulus matrix represent F0_△ and timing of the turning point T_t.



In the experiments, stimulus matrixes vary in two acoustic continua: (1) the change of F0 in falling phase F0_△ and (2) absolute timing of turning point T_t. F0 of the turning points F0_t and ending points F0_e are fixed at 150 Hz and 220Hz in accordance with Moore, et.al. [11]. With the F0 contour schema shown in Fig.1, we look for the acoustic differences that cause perceptual differences.

2.1. Experiment 1: duration effect

Table 1: Conditions of matrixes in Experiment 1.

Matrix	F0_△ (Hz) Step=10	T_t (ms)	T_t step (ms)	T_d (ms)	T_tr= T_t/T_d
1	10-70	20-200	20	200	0.1-1
2	10-70	20-200	20	400	0.05-0.5
3	10-70	40-400	40	400	0.1-1

In Experiment 1, three matrixes (Table1.) are built to decide the status of the absolute T_t and duration T_d in the formulation. All three matrixes share a same set of F0_△. Matrix 1 and Matrix 2 share a same set of T_t which is different from Matrix 3. Matrix 1 and Matrix 3 share a same set of T_tr [T_tr=T_t/T_d], which is different from Matrix 2. Here are the hypotheses and their predictions. (1) If T_t has effects but T_tr doesn't, significant difference shows between Matrixes 1-3 as well as between Matrixes 2-3, but not between Matrixes 1-2. (2) If T_tr has effects but T_t doesn't, significant difference shows between Matrixes 1 -2 as well as between Matrixes 3-2, but not between Matrixes 1-3. (3) If both T_t and T_tr have their independent effect, all three matrixes shows significant differences between each other.

2.1.1. Method

Subjects: 15 subjects participated in this experiment. All are native Mandarin speaking Beijingers. Subjects received payments.

Apparatus: the stimuli were presented by DMDX program [3] on a computer through both-ear speaker of a headphone in a sound treated booth. Subjects' responses were collected by the same computer and program.

Stimuli: the stimuli were three matrixes of /tu/ (du), each containing 70 different stimuli that vary along two acoustic continua of F0_△ and T_t as shown in Table1. Each matrix was repeated three times in a random order to a subject. We applied LF model [2] for the generation of voice periods, Klatt Synthesizer [6] for the formants, and overlapping concatenate methods for the gradual change of formants. The spike was adapted from the first period of the vowel.

Procedure: The three matrixes of stimuli formed three blocks of Experiment 1. The blocks and stimuli within each block were represented in random order. On each trial, subjects identified and rated the goodness of a single stimulus. For the identification task, subjects judged whether the stimulus sounded like /tu2/ (du2) "read" or /tu3/ (du3) "block up". After that, subjects were asked to rate how good an example the stimulus was of that category on a scale from 1 ("bad") to 5 ("good"). Only the identification data were used in this study. In each block each subject completed a practice block of 10 trials with stimuli randomly selected from the current matrix. If a stimulus was identified as Tone2, "1" was assigned to the identification score (P_I), otherwise "0" was assigned.

2.1.2. Results

A repeated ANOVA by subjects on mean P_I of each matrix (Table 2) yielded a significant main effect [F (2) =29.695, p<0.05]. Paired-Samples T-test revealed that mean P_I of Matrix 2 was significantly greater than that of Matrix 1 [t (14) =5.245, p<0.05] and 3 [t (14) =6.070, p<0.05], while that of Matrix 1 was significantly greater than that of Matrix 3 [t (14) =4.047, p<0.05]. All three matrixes showed significant difference between each other. Hence the result supported the third hypothesis that both T_t and T_tr (=T_t/ T_d) have their independent effect in the identification of Tone2 and Tone 3.

Based on the binomial distribution of P_I and the sigmoid shape of the response function, a logistic regression in Eq.(1) between P_I and two repeated measure predictors (T_t & F0_△) was adopted to obtain the mean identification function for each matrix. This method was adapted from Xu, et.al.[14]

$$\text{logit}(P_I) = \log_e\left(\frac{P_I}{1-P_I}\right) = b_0 + b_1 T_t + b_2 F0_{\triangle} \tag{1}$$

By generalizing the categorical boundary from a single value to a function, we derived the mean position of the categorical boundary in each matrix from the function involving T_t and F0_△ corresponding to the 50% identification score. See Eq. (2).

$$b_0 + b_1 T_t + b_2 F0_{\triangle} = \log_e\left(\frac{0.5}{1-0.5}\right) = 0 \tag{2}$$

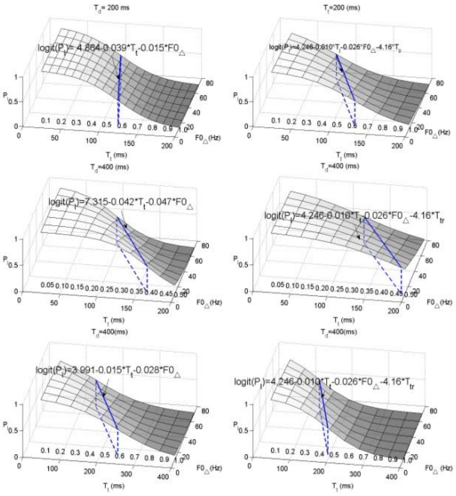
$$\Rightarrow \quad b_0 + b_1 T_t + b_2 F0_{\triangle} = 0$$

The estimated regression coefficients are presented in Table 2. These are applied for the left plots in Fig. 3.

Table 2: Mean identification scores and estimates of regression coefficients of each matrix in Eq. (1).

Matrix	mean P_I	b0	b1	b2	b1/b2
1	0.5038	4.864	-0.039	-0.015	2.6
2	0.5919	7.315	-0.042	-0.047	0.89
3	0.4294	3.991	-0.015	-0.028	0.54

Figure 3: Upper - Matrix1. Middle - Matrix2. Lower - Matrix3. Left - Eq.(1). Right- Eq. (3). Solid lines - Categorical boundaries. Dash line - boundaries projected on the stimulus matrix plane.



Greater T_d leads to smaller b1/b2 in Eq. (1). This indicates that T_d interacts with b1/b2. Hence a new logistic regression between the P_I and three repeated measure predictors (T_t, F0_△ & T_tr instead of T_d) was adopted to obtain one mean identification function as Eq. (3). The categorical boundary is a function involving T_t, F0_△ and T_tr corresponding to the 50% identification score. The estimated regression function is presented in Table 3.

$$\text{(3)}\quad \mathrm{logit}(P_I) = b_0 + b_1 T_t + b_2 F0_{\triangle} + b_3 T_{tr} = b_0 + b_1 T_t + b_2 F0_{\triangle} + b_3 T_t / T_d$$

Table 3: Estimates of regression coefficients applicable for all three matrixes in Eq. (3).

b0	b1	b2	b3
4.246	-0.010	-0.026	-4.160

With T_tr= T_t/T_d included, a longer duration (a greater T_d) results in a smaller absolute value of [b1+b3/T_d] in Eq.(3), and hence leads to a smaller [b1/b2] in Eq. (1). In this way, Eq. (3) incorporates T_d's effect on mean P_I of matrix as well as its interaction with b1/b2.

2.2. Experiment 2: register effect

Experiment 2 confirms the negative effects of precursor register. "Register" is practically defined as the average of the highest and lowest F0 of precursors. Target matrix is the same Matrix 2 in Experiment 1, preceded by three different precursor conditions and hence three different registers, each is about a half octave (six semitones) higher than another (Table 4).

Table 4: Registers in Experiment 2.

	Register (Hz)	Range (Hz)	precursor 1 (Hz)	precursor 2 (Hz)
1	139	92	185-93	185-185
2	196	92	242-150	242-242
3	277	92	323-231	323-323

2.2.1. Method

Similar method like Experiment 1 was applied, except that subjects identified one target syllable preceded by two precursors, a fall followed by a level (fall – level - target).

2.2.2. Results

A repeated ANOVA by subjects on mean P_I of each group yielded a significant main effect [F (2) =26.975, p<0.05]. Paired-Samples T-test revealed that mean P_I of the target matrix was significantly different between each two registers [t (14) =6.08, p<0.05, Register1-2], [t (14) =6.520, p<0.05, Register 1-3], [t (14) =2.322, p<0.05 for Register 2-3]. As shown in Table 5, the higher the register is, the lower the mean P_I is. This result confirms the negative effect of Register.

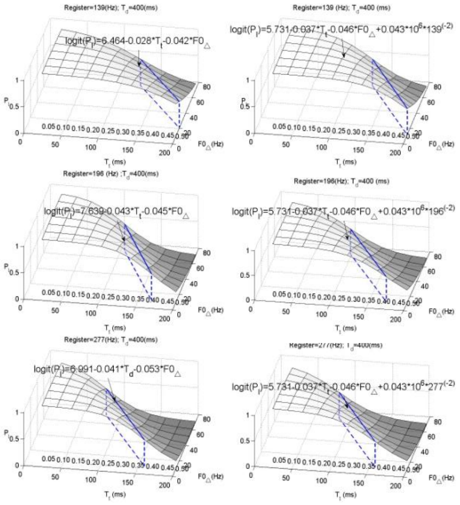
A logistic regression like Eq. (1) was adopted to obtain the mean identification function for the matrix after different registers. The estimated regression coefficients for the functions are presented in Table 5. These are applied for the left plots in Fig. 4.

Table 5: Mean identification scores and estimates of regression coefficients of each matrix in Eq. (1).

Register (Hz)	mean P_I	b0	b1	b2	b1/b2
(1) 139	0.7454	6.464	-0.028	-0.042	0.667
(2) 196	0.6154	7.639	-0.043	-0.045	0.956
(3) 277	0.5463	6.991	-0.041	-0.053	0.774

To include register effect, we adopted an additional predictor R in Eq. (4). Register is negatively relevant to mean P_I, but the relation is nonlinear, similar to a quadratic reciprocal function, we defined R as [10^6*Register^{-2}], estimated regression coefficients presented in Table 6, displayed in right plots Fig.4.

Figure 4: Upper-Register 1. Middle - Register 2. Lower - Register 3. Left - Eq.(1). Right - Eq. (4). Solid lines -- Categorical boundaries. Dash line – boundaries projected on the stimulus matrix plane.



(4)
$$\mathrm{logit}(P_l) = b_0 + b_1 T_t + b_2 F0_{\triangle} + b_4 R$$
$$= b_0 + b_1 T_t + b_2 F0_{\triangle} + 10^6\, b_4 \mathrm{Register}^{-2}$$

Table 6: Estimates of regression coefficients applicable for all three matrixes in Eq. (4).

b0	b1	b2	b4
5.713	-0.037	-0.046	0.043

3. CONCLUSION

When a syllable begins with a falling phase absolutely or proportionally shorter, when the initial F0 fall is smaller, and when it is preceded by precursors with lower registers, it is more likely to be identified as Mandarin Tone2 rather than Tone3. Confirming previous findings [11, 12, 17], a generalized CP model allows the formulation of relations between tonal identification and multiple acoustic parameters. Researchers have long been aware that if a syllable is short, low/ high instead of fall/ rise is more important [7, 16], but how to formulate this phenomenon remained a problem. The present study have formulated the interaction between duration and the slope of falling phase and hence provided a new insight into this problem.

4. ACKNOLEDGEMENTS

This research was funded by National Social Sciences Funds (Grant 10&ZD125) and Ministry of Science & Technology of China (Grant 2009BAH41B00). I thank two anonymous reviewers for the comments on the earlier version of the paper.

5. REFERENCES

[1] Blicher, D.L., Diehl, R.L., Cohen, L.B. 1990. Effects of syllable duration on the perception of the Mandarin Tone 2/Tone 3 distinction: Evidence of auditory enhancement. *Journal of Phonetics* 18, 37-49.

[2] Fant, G., Liljencrants, J., Lin, Q. 1985. A four-parameter model of glottal flow. *STL-QPSR* 4, 1-13.

[3] Forster, K.I., Forster, J.C. 2003. DMDX: A Windows display program with millisecond accuracy. *Behavior Research Methods Instruments and Computers* 35, 116-124.

[4] Fu, Q.J., Zeng, F.G. 2000. Identification of temporal envelope cues in Chinese tone recognition. *Asia Pacific Journal of Speech Language and Hearing* 5, 45-58.

[5] Howie, J.M. 1974. On the domain of tone in Mandarin. *Phonetica* 30, 129-148.

[6] Klatt, D.H., Klatt, L.C. 1990. Analysis, synthesis, and perception of voice quality variations among female and male talkers. *Journal of the Acoustical Society of America* 87, 820-857.

[7] Kratochvil, P. 1971. An experiment in the perception of Peking dialect tones. *A Symposium on Chinese Grammar: Scandinavian Institute of Asian Studies Monograph Series 6* Lund, Sweden, 7-31.

[8] Liberman, A.M., Harris, K.S., Hoffman, H.S., Griffith, B.C. 1957. The discrimination of speech sounds within and across phoneme boundaries. *J. Exp Psychol.* 54, 358-368.

[9] Lin, M.C. 1988. The acoustic characteristics and perceptual cues of tones in Standard Chinese. *Zhongguo Yuwen* 204, 182-193.

[10] Moore, B.C.J. 1997. Aspects of auditory processing related to speech perception. *The Handbook of Phonetic Sciences*, 539-565.

[11] Moore, C.B., Jongman, A. 1997. Speaker normalization in the perception of Mandarin Chinese tones. *Journal of the Acoustical Society of America* 102, 1864-1877.

[12] Shen, X.S., Lin, M. 1991. A perceptual study of Mandarin tones 2 and 3. *Language and Speech* 34, 145-156.

[13] Whalen, D.H., Xu, Y. 1992. Information for Mandarin tones in the amplitude contour and in brief segments. *Phonetica* 49, 25-47.

[14] Xu, Y., Gandour, J.T., Francis, A.L. 2006. Effects of language experience and stimulus complexity on the categorical perception of pitch direction. *Journal of the Acoustical Society of America* 120(2), 1063-1074.

[15] Yang, R. 2009. *A Study on Categorical Perception of Tones in Mandarin through a Phonation Perspective*. Master dissertation, Peking University, Beijing.

[16] Yip, M. 2008. Contour tones. *Phonology* 6, 149-174.

[17] Zue, V.W. 1976. Some perceptual experiments on the Mandarin tones. *Journal of the Acoustical Society of America* 60, S45.