THE PHONATION FACTOR IN THE CATEGORICAL PERCEPTION OF MANDARIN TONES

Ruo-Xiao Yang

The Chinese University of Hong Kong, Hong Kong
yangruoxiao@cuhk.edu.hk

ABSTRACT

This study presents a new way of observing the categorical nature of Mandarin tone perception by considering the phonation factors in the experiment design. In order to demonstrate the reliability of experiment materials, a pitch synchronous overlap-add (PSOLA) method is used to produce the speech stimuli, and an inverse-filtering method is applied to assess the invariability of sound source of the stimuli. Results provide evidence for the prediction that phonation factors are adopted in perceiving tones by Mandarin speakers. Especially for the tone which has a special voice characteristic, such as the third tone in Mandarin, the perception difference among continua with different voice qualities is much more obvious. Results also suggest that there is categorical perception for each pair of two tones in Mandarin. These results lead to a discussion about the necessity of a wider and more integrated theory model for describing tones in tone languages, in which the phonation factors should be considered.

Keywords: phonation perception, voice quality, categorical perception, Mandarin tones

1. INTRODUCTION

Categorical perception (CP) is an important phenomenon in human cognition because people sort out surrounding things incessantly every day. Roughly, the differences among things can be perceived with two modes: the “continuous perception” and the “categorical perception (CP)”. CP have been formally thought to be peculiar to speech and color of human perceivers, but it turns out to be much more general by later research on human infants and animals [4, 8, 9]. CP has been investigated in different levels of language. In human speech, the general conclusion about vowels as continuous perception and consonants as categorical perception has been widely accepted, whereas the categorical nature of tone in tone languages is still controversial.

Tone is a critical element for tone languages.. Generally, tone is produced by the vibration of vocal folds, which is often measured as fundamental frequency (F0) acoustically and perceived as pitch. When tone is described or perceived, the F0, sometimes as well as intensity and duration, is taken as its most critical physical correlates. However, this standpoint is only based on the physical or acoustic representations of tone but overlook its articulator, the vocal folds.

Vocal folds are the source of each sound, which connect the dynamics from lung beneath and the resonance from vocal tract above. The vibration of vocal folds for speech production can be described not only as its vibrating velocity per second, i.e. the fundamental frequency (F0), but also its vibrating mode which is often referred as phonation types (or loosely, voice qualities). Since a speech sound can be divided into segmental (consonants and vowels) and super-segmental parts under the backgrounds of modern phonology and phonetics, the phonation part should be sorted to be a super-segmental feature.

Related to surveys on the CP nature of tones, the usual pervious hypothesis has suggested that F0 is the main or even the most crucial correlate for perceiving tones. Perception experiments are thus designed based on this viewpoint. However, if tone is observed from the “phonation” perspective, F0 might not be the only correlate for distinguishing tones as different phonemic units. Furthermore, more and more evidence show that some tone languages (such as Hani language and Jingpo language) indeed use different phonation types as distinctive features rather than F0/pitch (Kong, [6]). Thus, it is important and necessary to discuss the CP nature of tones within a phonation perspective.

Several previous studies have explored the phonation modes of Mandarin tones from a linguistic viewpoint. Kong used EGG (Electroglottographic) signal [6] and high-speed digital imaging [7] to investigate the voice of the four basic tones of Mandarin/standard Chinese in single syllables in isolated monosyllabic words. These research suggested that F0 and phonation features all contribute to the perception of Mandarin tones. Keating and Esposito [5] also took some preliminary measurements on F0 and other parameters related to phonation types of the four basic tones of Mandarin. They also pointed out some special phonation features in the low falling tone and the end of the falling tone of Mandarin, i.e., the creaky voice can be heard on the low falling tone and visible at the end of the falling tone. Moreover, they suggested a more general conclusion that all “tones” (in all languages) may have some correlated variation in phonation, only most of them being subtle within the modal ranges and some owning phonological status, but it was probable that listeners in many tone languages use the phonation information in recognizing their tones especially out of context. Even though the relationships of phonation and tones have attracted academic attention, no studies have been done to discuss how the phonation information related to tones can be perceived. Moreover, no previous research on CP of tones and perceiving phonation information has attempted to consider the phonation information contributing to perception of tones. We argue that it is important to look at the phonation factor and CP of tones concurrently in one experiment design in order to achieve a better understanding of CP and how to define the phonological status of tones in tone languages. Therefore, in this experiment, the phonation factor has been considered concurrently with CP testing. This design permits us to examine whether CP exists in arbitrary two tones of the four basic tones in Mandarin by making continua as stimuli between each pairs of two tones and to determine whether phonation information is used in differentiating tone categories by synthesized stimuli.

2. EXPERIMENT MATERIALS AND METHODS

2.1. Materials

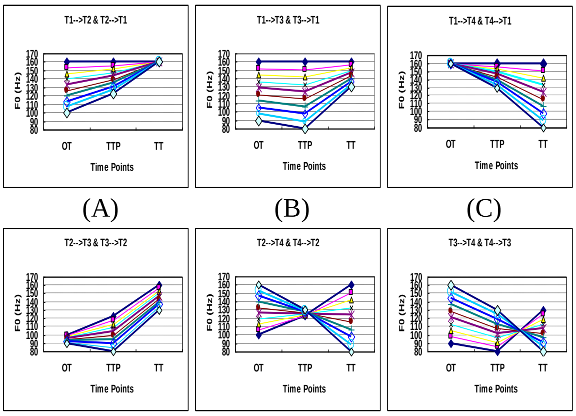
The stimuli were manipulated based on four Mandarin syllables with four basic tones out of context. To make the syllable more real and keep better phonation information (which may be lost in context), the four original syllables are produced separately by a male speaker whose native language is Beijing Mandarin. Moreover, in order to keep as much sufficient information of original syllables as possible and avoid losing information during manipulation, the syllable structure was controlled and thus only [ta] syllable was used. Previous studies suggest that [ta] is the best syllable structure for keeping the formant information when manipulating F0.

For assessing the phonation information during perceiving Mandarin tones, 12 10-step continua were synthesized as stimuli by manipulating F0 through the “pitch synchronous overlap-add” method (i.e., PSOLA) in Praat [1].

Table 1: Syllable with four Mandarin tones and F0 parameters of four tones for synthesis.

Four Tones	OT F0 (Hz)	TTP F0 (Hz)	TT F0 (Hz)
Tone1 (T1)	160	160	160
Tone2 (T2)	100	123	160
Tone3 (T3)	90	80	130
Tone4 (T4)	160	130	80

Figure 1 (A-F): Diagrams of manipulating F0 in 12 continua: (A) T1→T2/T2→T1, (B) T1→T3/T→T1, (C) T1→T4/T4→T1, (D) T2→T3/T3→T2, (E) T2→T4/T4→T2, (F) T3→T4/T4→T3.



(A) (B) (C)

(D) (E) (F)

Taylor [11] explained that TD-PSOLA is attempting to mimic the process of safely changing the pitch without changing filter by separating the effects of each pulse. Thus, under the most ideal condition, TD-PSOLA will keep the spectral envelope characteristics and therefore keep the original phonation/voice information from the sound source. Furthermore, Lemmetty [10] and Upperman [12] also proposed that the speech signal resulted from PSOLA has the same spectrum as the original signal but with a different F0. Esposito [3] also suggested that PSOLA changes the F0 of a signal without changing other properties of the voice and applied it to normalize F0 of stimuli for testing the perception of phonation types.

2.2. Subjects

19 participants from Peking University in China

(11 female and 8 male, aged 19-25 years) participated for a small amount of money in the experiment: an identification test followed by a discrimination test. All of them were native fluent speakers of Mandarin / Putonghua with different Chinese dialect backgrounds.

2.3. Experiment procedures

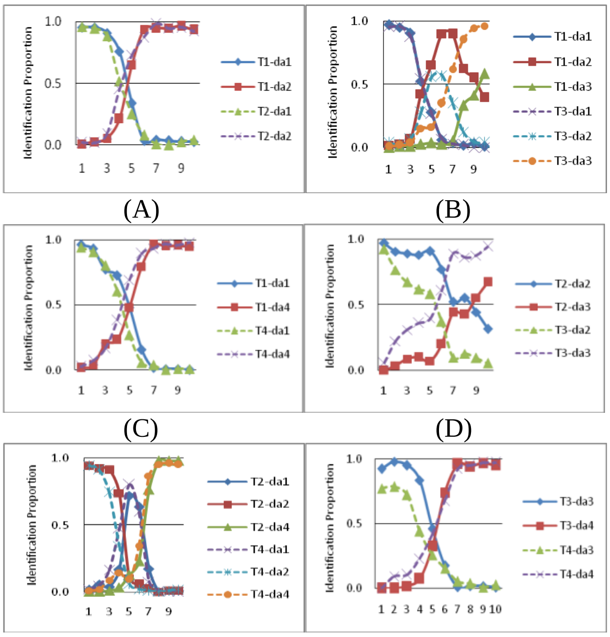
The whole experiment was consisted of an identification task and a AX discrimination task. In the identification task, subjects listened to stimuli presented in isolation. They were instructed to press one of the four buttons on the computer screen by using the left button of the mouse, each of which was labeled with the four original syllables with Mandarin tones from real speech in Pinyin, i.e., “da1 ([ta55], ‘put’)”, “da2 ([ta35], ‘answer’)”, “da3 ([ta214], ‘hit’)” and “da4 ([ta51], ‘big’)”. Each of the 120 stimuli occurred 6 times (720 trials) in a random order.

In the discrimination task, stimuli were presented in pairs with a 0.5 second interstimulus interval (ISI). Each pair of stimuli consisted of two different stimuli separated by two steps on each of the 12 continua (i.e., 1-3, 2-4, 3-5, 4-6, 5-7,6-8, 7-9, 8-10). Therefore, totally 96 pairs of different stimuli were made. Each of them was presented 6 times in random order (576 trails). After hearing each pair, subjects were instructed to judge whether the two stimuli were the same or the different, and to respond by pressing two buttons labeled with “Same” and “Different” on the computer screen through the left mouse button.

3. RESULTS

Figure 2 (A-F) showed the identification curves of subjects’ responses to pairs of continua with the same F0 model but different voice qualities. As a whole, curves of the six pairs of continua revealed a similar feature that the continuum based on a specific voice quality inclined to be perceived as the tone with the specific quality. For example, for the pairs of continua of T1 and T2 based on the voice of T1 and T2 in Figure 2 (A), the identification boundary for judging the stimuli as T1 based on the voice of T1 (solid lines) was earlier than the one based on the voice of T2 (dashed lines). The similar phenomenon also occurred in other pairs of continua if one of the two tones within each pair was used as the datum mark. That is to say, the continuum with specific voice quality was much easier to be perceived as the tone with this voice quality; therefore, it could be roughly concluded here that voice quality or phonation information indeed influenced the categorical perception of Mandarin tones.

Figure 2 (A-F): Identification curves for responses of 19 subjects in 12 continua to show the differences between pairs of continua with different voice qualities: (A) T1→T2 based on the voice of T1 (solid line) and T2 (dashed line); (B) T1→T3 based on the voice of T1 (solid line) and T3 (dashed line); (C) T1→T4 based on the voice of T1 (solid line) and T4 (dashed line); (D) T2→T3 based on the voice of T2 (solid line) and T3 (dashed line); (E) T2→T4 based on the voice of T2 (solid line) and T4 (dashed line); (F) T3→T4 based on the voice of T3 (solid line) and T4 (dashed line).



Furthermore, continua with T3 presented much more characteristics, which may be related to its special voice quality. As mentioned above, the third tone in Mandarin is often produced with vocal fry [6, 7] or creaky voice [5]. It makes the third tone much more unique when heard, differentiating with other three tones. It can be observed that the continua synthesized based on T3, i.e., owning the voice quality of T3, can be perceived as T3 much easier than the continua based on other voice qualities. Especially, in continua of T1→T3 based on the voice of T1 and T3, the continuum based on T3 can be perceived more as T3 than the continuum based on T1, and the “new category of T2” was much smaller in the continuum with voice of T3 than the one without it, which may be attributed to the unique voice quality of T3. The same situation also occurred in the continua of T3→T4 based respectively on the voice of T4 and T3. For the continuum based on the voice of T4, the perception of T3 seemed to be much more difficult and even might be judged as T1 and T2 in a larger proportion, while for the continuum based on the voice of T3, the proportion was much smaller and easier to be perceived as T3.

Preliminarily, a two-way mixed design repeated-measures ANOVA was conducted on the identification data, with voice quality group as the between-subject factor and tone continuum as within-subject factor. Corrections for violations of sphericity were made, where appropriate, using the Greenhouse–Geisser method. Significant main effects of voice quality group for T3 series were found, i.e., T3 and T1 ($F (1, 34) = 49.11, p < 0.01$); T3 and T2 ($F (1, 34) = 26.34, p < 0.01$); T3 and T4 ($F (1, 34) = 11.96, p < 0.01$), suggesting that the voice quality of T3 plays an important role in the differentiation of T3 with the other three tones. Moreover, two main effects of voice quality group for T2 series were reported, with T2 and T3 ($F (1, 34) = 25.82, p < 0.01$) and T2 and T4 ($F (1, 34) = 15.71, p < 0.01$). No significant main effects of voice quality group for T1 series and T2 series were found.

4. DISCUSSION

Generally, F0 (acoustically) or pitch (auditorily) is considered to be the primary physical correlate of tone. However, in some tone languages, especially the tone languages in Asia, three or four contrastive tone levels are quite common, which leads phonologists to think about the sufficiency of theory for describing tones as distinctive features with the backgrounds of generative phonology, since if tones are to be represented by distinctive features, one needs to convert multiple levels into binary features. Duanmu [2] suggested that Register is independently related to the voicing of the onset consonant and thus it is possible that voice quality is the main perceptual cue for distinguishing [+upper] and [-upper] Register, such as creaky voice being related to [-upper] Register in Mandarin tones. These theoretical arguments about Register in tone system in phonology have indeed touched the necessity of introducing phonation factors into description of tone systems. Moreover, the vocal folds as a voice source being the sound excitation of both tones and other segments in speech, the phonation process actually connects different segments of speech into an integrated process, which may provide more possibilities to interpret the process of sound change, as well as its related underlying physiological bases.

5. CONCLUSION

The perceptual experiments proposed in this article are designed under a categorical perception experiment scheme and provide empirical evidence for the fact that phonation factors are adopted in perceiving tones by Mandarin speakers. Especially for the tone which has special voice characteristics, such as the third tone in Mandarin, the different perception patterns with different voice qualities are much more obvious. These results lead to a discussion about a wider and more integrated theory model for describing tones, in which the phonation factors should be incorporated.

6. REFERENCES

[1] Boersma, P., Weenink, D. 2008. Praat. *http://www.fon.hum.uva.nl/praat/*

[2] Duanmu, S. 2000. Tone: An overview. In Cheng, L.L.S., Sybesma, R. (eds.), *The First Glot International State-of-the-article Book: The Latest in Linguistics*. Berlin: Mouton de Gruyter, 251-286.

[3] Esposito, C.M. 2006. *The Effects of Linguistic Experience on the Perception of Phonation.* Unpublished PhD Dissertation. UCLA.

[4] Harnad, S. 2003. Advanced topics: Categorical perception (CP). *http://users.ecs.soton.ac.uk/harnad/at.html*

[5] Keating, P.A., Esposito, C. 2006. Linguistic voice quality. Paper presented at the *11th Australasian International Conference on Speech Science and Technology*. *http://www.linguistics.ucla.edu/people/keating/keating.htm*

[6] Kong, J. 2001. *On Language Phonation* (1st ed.). (In Chinese). Beijing: The Central University of Nationalities Press.

[7] Kong, J. 2007. *Laryngeal Dynamics and Physiological Models: High Speed Imaging and Acoustical Techniques* (1st ed.). Beijing: Peking University Press.

[8] Kuhl, P.K. 1987. The special-mechanisms debate in speech research: Categorization tests on animals and infants. In Harnad, S. (ed.), *Categorical Perception: the Groundwork of Cognition*. Cambridge: Cambridge University Press, 355-386.

[9] Kuhl, P.K. 2004. Early language acquisition: Cracking the speech code. *Nat. Rev. Neurosci* 5(11), 831-843.

[10] Lemmetty, S. 1999. *Review of Speech Synthesis Technology*. Unpublished Master Thesis, Helsinki University of Technology

[11] Taylor, P. 2009. *Text-to-Speech Synthesis* (1st ed.). Cambridge: Cambridge University Press.

[12] Upperman, G. 2004. Changing Pitch with PSOLA for Voice Conversion. *http://cnx.org/content/m12474/latest/*