Proceedings of 20th International Congress on Acoustics, ICA 2010

23-27 August 2010, Sydney, Australia

Effects of the Acoustic Characteristics on the Emotional Tones of Voice of Mandarin Tones

Hui-Shan Chang (1, 2), Shuenn-Tsong Young (1) and Kevin Yuen (3)

(1) Institute of Biomedical Engineering, National Yang-Ming University, Taipei, Taiwan

(2) Institute of Speech and Hearing Disorders and Sciences, National Taipei College of Nursing, Taipei, Taiwan

(3) Department of Otorhinolaryngology, Head & Neck Surgery, The Chinese University of Hong Kong, Hong Kong, China

PACS: 43.10Ce, 43.10.NQ, 43.70.FQ

ABSTRACT

Objectives: In Mandarin Chinese, a tonal language, the pitch pattern of syllables conveys lexical meaning. There are four tone patterns (Tone 1 through Tone 4) defined by the pitch pattern, or the fundamental frequency pattern. Each of them has very unique acoustic characteristics, including fundamental frequency, amplitude, duration, tempo, pausing, distribution of energy in the frequency spectrum, and formant location. Emotional tone of voice is defined as a vocal expression of emotion conveying affective information. Each of them also has very unique acoustic characteristics. The acoustic characteristics of Mandarin tones change the listener's perception of the meaning of an utterance. Emotional tone of voice is alike in perception, when acoustic characteristics of the emotional tone of voice changes, the perception of the listener may vary. Therefore, the purpose of this study is to investigate the variation of the acoustic characteristics of the four tones across different emotional tones of voice in Mandarin.

Study Design: 11 male subjects and 11 female subjects (mean age: 26) participated in this study. All subjects were native Mandarin speakers, had no history of any speech and hearing disorder, and passed articulation and voice assessment successfully. All speech samples designed by the author were used in this study. There were a total of forty speech samples: two phonetic contexts (/ti/ and /tu/), four lexical tones, and five different emotional tones of voice (angry, sad, fear, neutral, and happy tones of voice). Acoustic analysis of each target syllable included the mean F0, the mean amplitude, and the duration of the target syllables.

Results: The mean F0 and the mean amplitude for four tones of the target syllables are significant different across five different emotional tones of voice, but the duration for four tones of the target syllables of neutral tone of voice are similar with those of the other four emotional tones of voice. This is a preliminary study to investigate the variation of the acoustic characteristics of the four tones across different emotional tones of voice in Mandarin, and the furthering investigation is needed.

INTRODUCTION

Mandarin, a language spoken by more people than any other tonal languages[15], has four tone patterns. The F0 contours of Tones 1-4 are (1) flat and high; (2) rising; (3) falling and then rising; and (4) falling, respectively[15].

The four tones differ in duration and amplitude. Howie (1974)[6] observed that vowel or syllable durations in Mandarin were longer with Tone 2 than Tone 4. Tseng (1990) reported the duration measurement of vowels (/i/, /u/, and /a/) of Tone 1 to Tone 4, with the results shown in Table 3. The results revealed that Tone 3 has the longest vowel duration, while Tone 4 has the shortest. Xu and Pfingst (2002) investigated that the duration of the syllables /ba/ and /yi/ produced by a native male speaker of Tone 1 to Tone 4. They also observed that Tone 3 has the longest syllable duration, while Tone 4 has the shortest.

The peak amplitudes of the syllables /ba/ and /yi/ produced by a native male speaker of Tone 1 to Tone 4 were reported by Xu and Pfingst (2002). They found that the peak amplitude of Tone 3 is generally the lowest, while Tone 4 is the highest.

Emotional tone of voice is defined as a vocal expression of emotion conveying affective information[3]. Each of the emotions including anger, happiness, sadness, fear, surprise, scorn, and so on, have very unique acoustic characteristics[2, 3, 4].

Previous research on the acoustics characteristics of angry tone of voice showed that it had higher mean F0[4, 12]; the higher mean amplitude[4, 12].

Several studies showed that happy tone of voice leads to an increase in the mean F0[4].

Öster and Risberg (1986)[4] found that the utterance duration for happy tone of voice was longer, but Davitz (1964b)[4] reported the reverse. They found that the utterance duration for happy tone of voice was with shorter duration. In addition, there was higher mean amplitude in this emotional tone of voice[4].

Sad tone of voice was found to have similar mean F0 and a longer duration compared with neutral tone of voice[4], lower mean amplitude[2], and a slight falling F0 contour[2, 12] compared with neutral tone of voice.

Fairbanks and Pronovost (1939)[4] and Fónagy (1978, p.62)[4] compared "fear tone of voice" with the four other emotions, and they noted that fear tone of voice has the highest mean F0.

Fairbanks and Hoaglin (1941)[4] noted a shorter duration in fear tone of voice, but Sulc (1977) [4] noted a longer duration.

METHOD

Subjects

Twenty-two subjects, eleven male and eleven female, participated in this study. Their mean age was 26 years old, with the range from 22 to 39 years old. All subjects were native Mandarin speakers, had no history of any speech and hearing disorder, and passed articulation and voice assessment successfully.

Speech Samples

All speech samples designed by the author were used in this study (Appendix 1). Each speech sample consisted of a target syllable (/ti/ or /tu/) with tokens of a four different lexical tones (Tone 1 through Tone 4). The target syllables were embedded in a carrier phrase: 你說讀/tu2/這個字 (You say the word "讀/tu2/").

There were a total of 40 conditions: two phonetic context, four lexical tones, and five different emotional tones of voice (angry, sad, neutral, fear, and happy tones of voice).

Procedure

An 80 dB SPL speech noise generated by an Audiometer (GSI 61) through a loudspeaker and was recorded by the Computerized Speech Lab, KAY-Multi-Speech Model 3700 (CSL) and functioned as a reference to convert relative target syllable intensity level into absolute level.

The microphone (SHURE BETA87, SX12-48 VDC, Illinois, USA) of the CSL was placed 1 meter from the loudspeaker during recording. At the same time, the microphone of the sound level meter (Modular Precision Sound Analyzer Type 2260, Brüel & Kjær) was placed right next the microphone of the CSL. The sound level meter was a C weighting.

All speech samples were recorded with the subjects seated in a sound treated booth at the Department of Speech and Hearing Disorders and Sciences at National Taipei College of Nursing, at 0 degree azimuth, 15 cm from microphone of CSL. The sampling rate of recording was 44,100 Hz.

Before the recording session starts, the subjects were given time to become familiar with the all speech samples. All subjects were asked to produce ten repetitions of each speech sample. During recording, no pauses were introduced before and after the target syllables in the carrier phrase. The target syllables were not stressed.

The speech samples were selected for further analysis only if they passed a rating test. The rating test was aimed at identifying those speech tokens that conveys the intended emotion satisfactorily. The speech samples were presented to eleven normal-hearing native Mandarin-speaking adults. They were asked to provide a score on whether the speech sample matched with the intended emotion. The scores were integers between 1 and 5 (Likert's 5-point scale), with 1 being "completely not match", 2 being "not match", 3 being "fairly", 4 being "match", and 5 being a "completely match" score. Only the speech samples with average subjective score from 3 to 5 were further analyzed in this study.

Acoustic Analysis

Acoustic analysis was performed with the Praat sound processing program (Version 5.1.31, Paul Boersma & David Weenink, 1992-2010). The values of the amplitude levels of the signal were arbitrary and were not in the sound pressure level (SPL) scale. A correction factor was applied on the test items. In this study, the sound pressure level of the calibration signal (the speech noise) obtained before recording was 80 dB SPL, but the mean amplitude measured in the praat program was 66.2 dB. Therefore, 13.8 dB should be added to the amplitude measurement in the praat program to obtain the SPL reading. For example, if the amplitude of the syllable /ti/ in the praat program was 52 dB, the SPL level should be 65.8 dB (52 dB + 13.8 dB = 65.8 dB).

Acoustic analysis of each tone of target syllable included (1) the mean F0, (2) the duration, and (3) the mean amplitude of the target syllable. The duration of the target syllable is defined as the time between the last vocal pulsing of the waveform before the target syllable and the last one of the target syllable.

Statistic Analysis

A three-way analysis of variance (ANOVA) was conducted to investigate the main effects for five emotional tones of voice, phonetic context, and lexical tones on the different acoustic characteristics of each tone of the target syllables including: (1) the mean F0, (2) the duration, and (3) the mean amplitude. Post-hoc Tukey Honest Significant Difference (HSD) Test was performed for each significant main effect. The Statistical Package for the Social Sciences (SPSS) 18.0 software was used to perform the analyses.

RESULTS

Mean F0

ANOVA results for Tone 1 ($p < 0.001$), Tone 2 ($p < 0.001$), Tone 3 ($p < 0.001$), and Tone 4 ($p < 0.001$), reveal that there are significant main effects of mean F0 from emotional tones of voice. Post-hoc Tukey HSD test (Figure 1) reveals that the mean F0 of Tone 1 of neutral tone of voice is higher than sad tone of voice and lower than angry, fear, and happy tones of voice. Tone 2, Tone 3, and Tone 4 of sad and neutral tones of voice are lower than the other three emotional tones of voice.

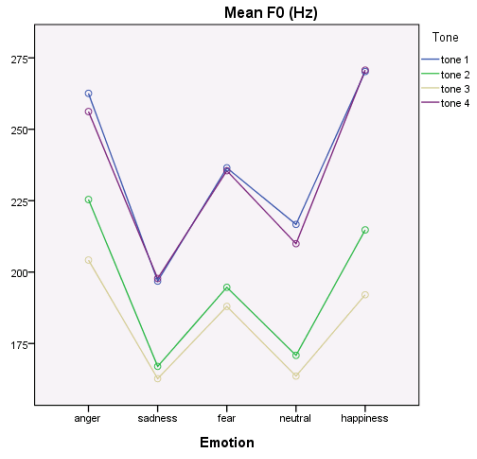

Figure 1 The mean F0 of target syllables of five emotional tones of voice and Mandarin tones.

Duration

ANOVA results show that that there are significant main effects of duration for Tone 1 ($p < 0.001$), Tone 2 ($p < 0.001$), and Tone 4 ($p < 0.001$) except Tone 3 ($p = 0.522$), reveal that there are significant main effects of duration from emotional tones of voice. Post-hoc Tukey HSD test (Figure 2) reveals that the duration of (1) Tone 1 and Tone 2 of neutral and sad tones of voice are longer than the other three tones of voice. (2) Tone 4 of neutral, fear, and happy tones of voice is shorter than anger tones of voice, and shorter than sad tone of voice.

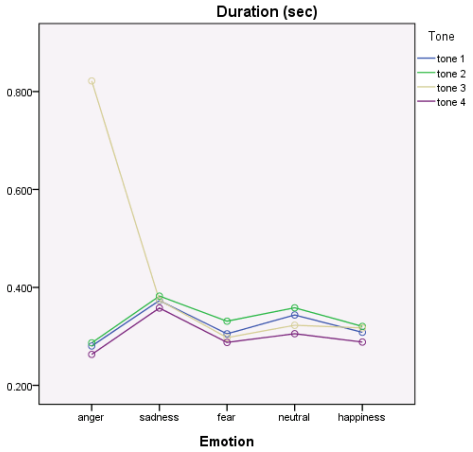

Figure 2 The duration of target syllables of five emotional tones of voice and Mandarin tones.

Mean Amplitude

ANOVA results show that there are significant main effects of the mean amplitude for Tone 1 ($p < 0.001$), Tone 2 ($p < 0.001$), Tone 3 ($p < 0.001$), and Tone 4 ($p < 0.001$), reveal that there are significant main effects of mean amplitude from emotional tones of voice. Post-hoc Tukey HSD test (Figure 3) reveals that the mean amplitude of Tone 1, Tone 2, and Tone 3 of neutral and fear tones of voice are higher than sad tone of voice and are lower than angry and happy tones of voice. Tone 4 of neutral, sad, and fear tones of voice are lower than angry and happy tones of voice.

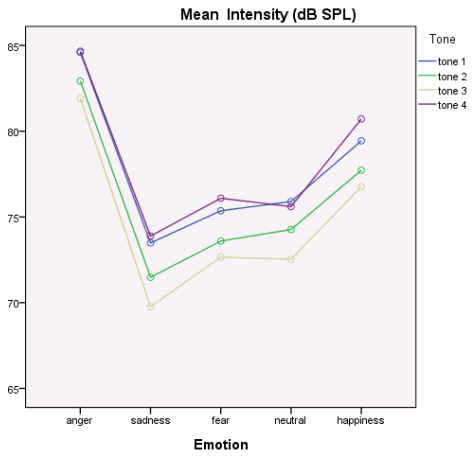

Figure 3 The mean intensity of target syllables of five emotional tones of voice and Mandarin tones.

DISCUSSIONS

Mean F0

Most of early researches and this study showed that the mean F0 of the neutral tone of voice is lower than angry, fear, and happy tones of voice[2, 4].

Duration

In this study, the duration of the target syllables of neutral tone of voice is not lower than angry tone of voice is consistent with those from previous studies[4].

Mean Amplitude

In this study, the mean amplitude of neutral tone of voice is lower than angry and happy tones of voice, which had been examined previously in the literature[4, 12].

CONCLUSIONS

The primary aim of this study was to investigate the variation of the acoustic characteristics across different emotional tones of voice in Mandarin Chinese tones. The results show that:

(1) Tone 1 and Tone 2 of neutral tone of voice on Mandarin Chinese have the lower mean F0 and lower amplitude than angry and happy tones of voice. The duration of tone 1 and tone 2of neutral tone of voice is similar with sad tone of voice, and longer than the other three emotional tones of voice.

(2) Tone 3 of neutral tone of voice on Mandarin Chinese has the lower mean F0 and lower amplitude than angry and happy tones of voice.

(3) Tone 4 of neutral tone of voice on Mandarin Chinese has the lower mean F0 and lower amplitude than angry and happy tones of voice. The duration of tone 4 of neutral tone of voice is longer than angry and happy tones of voice, and similar with the sad and fear tones of voice.

APPENDIX I

All target syllables are embedded in the carrier phrase including two syllables (/ti/ or /tu/) and four lexical tones (Tone 1 through Tone 4). All speech samples consist of a target syllable and one lexical tone. A total of 60 conditions used in this study are shown below:

(1) The twelve speech samples were: 你說 one of the target syllables (You say "one of the target syllables") with one of five different emotional tones of voice.

(2) Target syllables were 低/ti1/, 迪/ti2/, 底/ti3/, 弟/ti4/, 嘟/tu1/, 讀/tu2/, 賭/tu3/, and 杜/tu4/.

(3) Five different emotional tones of voice were angry, sad, fear, neutral, and happy tones of voice.

REFERENCES

1. Fu, Q.J., Zeng, F.G., Shannon, R.V. & Soli, S.D. "Importance of tonal envelop cues in Chinese speech recognition" *Journal of the Acoustical Society of America*, 104(1), 505-510 (1998).

2. Leinonen L., Hiltunen T., Linnankoski I., & Laakso M.L. "Expression of emotional-motivational connotations with a one-word utterance" *Journal of the Acoustical Society of America*, 102(3), 1853-1863 (1997).

3. Mullennix J.M., Bihon T., Bricklemyer J., Gaston J. & Keener J.M. "Effects of variation in emotional tone of voice on speech perception" *Language and Speech*, 45(3), 255-283 (2002).

4. Murray I.R. & Arnott J.L. “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(2), 1097-1108 (1993).

5. Murray I.R., Arnott J.L. & Rohwer E.A. “Emotional stress in synthetic speech: Progress and further directions” *Speech Communication*, 20, 85-91 (1996).

6. Tseng, Massaro, & Cohen. *Lexical Tone Perception in Mandarin Chinese: Evaluation and Integration of Acoustic Features* (In Linguistics, Psychology, and the Chinese Language, Kao, H.S.R. and Hoosain, R. Eds., University of Hong Kong Press, 1986) pp.91-104.

7. Tseng, C.Y. *An Acoustic Phonetic Study on Tones in Mandarin Chinese* (No. 94) (Nankang, Taipei: Institute of History & Philology, Academia Sinica, special publications, 1990).

8. Scherer, K.R. *Universality of emotion expression* (In D. Levinson, J. Ponzetti, & D. Jorgenson, 1999) pp. 669-674.

9. Scherer, K.R., Banse, R., & Wallbott, H.G. ”Emotion inferences from vocal expression correkate across language and cultures” *Journal of Cross-Cultural Psychology*, 32, 76-92 (2001).

10. Scherer, K.R. “Vocal Communication of Emotion: A Review of Research Paradigms” *Speech Communication*, 40, 227-256 (2003).

11. Schröder, M. “Emotional Speech Synthesis - A Review.” *Proc. Eurospeech 2001, Aalborg*, 1, 561-564 (2001).

12. Williams, C.E., Stevens, K.N. “Emotions and Speech: Some Acoustical Correlates” *Journal of the Acoustical Society of America*, 52, 1238-1250 (1972).

13. Whalen, D.H. & Xu, Y. “Information for Mandarin Tones in the amplitude contour and in brief segments” *Phonetica*, 49, 25-47 (1992).

14. Xu L., Tsai Y.H., & Pfingst B.E. “Features of stimulation affecting tonal-speech perception: Implications for cochlear prostheses” *Journal of the Acoustical Society of America*, 112(1), 247-258 (2002).