Proceedings of 20th International Congress on Acoustics, ICA 2010

23-27 August 2010, Sydney, Australia

Mandarin Tones Recognition by Segments of Fundamental Frequency Contours

Cheng-Yu Ho, Hui-Shan Chang, and Shuenn-Tsong Young

Institute of Biomedical Engineering, National Yang-Ming University, Taipei City, Taiwan

PACS: 43.72.Ar Speech analysis and analysis techniques; parametric representation of speech

ABSTRACT

Mandarin is one of the tonal languages. In Mandarin tones, there are four lexical tones (tone 1 to tone 4) with four different fundamental frequency (f0), such as flat and high, rising, falling and then rising, and falling, respectively. In order to process signal on lexical tone, at first we have to identify which tone is. We would like to find out an efficient approach to identify Mandarin tones by the segments of fundamental frequency contours. In this study, 3 male and 3 female participants engaged in recording speech materials. All participants are native Mandarin speakers, no history of any speech or hearing disorder, and all passed articulation and voice assessment. There are two target syllables (/ti/ and /tu/) of four lexical tones used for the materials. In our experiment, we analysed the signal features and the acoustics characteristics that included the range of f0, the average of f0 and so on. Then we could predict which tone is from what the segments told. The result of this study revealed that the segments of the contours could not target the corresponding tone correctly. The approach of this study may not provide a way for hearing-devices to predict Mandarin tone before signal processing. The further study for prediction by segments of f0 contours is required.

INTRODUCTION

In Mandarin tones, there are four tones (tone 1 to tone 4) with four different fundamental frequency (f0) contour patterns, such as flat and high, rising, falling and then rising, and falling [3]. They convey different lexical meaning with different fundamental frequency (f0) in Mandarin tones. Such as, /ba/ associated with Tone1 through Tone 4, with Tone 1 that means 'eight (八)', with tone 2 that means "uproot(拔)", with tone 3 that means 'grip (把)' and with Tone 4 that means 'father (爸)', respectively.

There are many methods for tone recognition; some use the entire portion of f0 contour for recognition, some use the segments of f0 contour. The former which is used for tone recognition takes large amount of speech corpus, such as artificial neural network (ANN) [1]. The later which is used segments of f0 contours provides a more effective way for tone recognition. Zhang and Hirose (2000) divided syllabic f0 contour into three segments: onset course, tone nucleus and offset course [2]. In this study, we proposed an efficient approach to identify Mandarin tones by the segments of fundamental frequency contours.

METHOD

Subjects

In this study, 3 male and 3 female participants (mean age was 26 years, 22 to 39 years of age) engaged in recording speech materials. All participants are native Mandarin speakers, no history of any speech or hearing disorder, and all of them passed articulation and voice assessment.

Recording Setup

Before recording, an 80 dB SPL speech noise that was generated from an audiometer through a loudspeaker and was recorded by the Computerized Speech Lab, KAY-Multi-Speech Model 3700 (CSL). The step functioned as a reference to convert relative intensity level into absolute level was used for volume calibration. 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 Analyser Type 2260, Brüel & Kjær) was placed next to the microphone of the CSL. The sound level meter was turned at C weighting.

All participants were seated in front of the microphone at 0 degree azimuth, 15 cm apart from the microphone. All recordings were conducted with a sampling rate of 11025 Hz and a 16-bit Hz resolution. The recordings were carried out in the sound booth in the Department of Speech and Hearing Disorders and Sciences in National Taipei College of Nursing.

Before recording procedure, each participant has different script order and was requested to repeat each target syllable for ten times. In recording, no pauses were introduced before or after the target syllables in the carrier phase. The target syllables were also not stress.

Materials

There are two target syllables (/ti/ and /tu/) of four lexical tones in a carrier phrases, 你說/ti1/這個字(You say the word "滴/ti1/"), used for the materials. The target syllable /ti/ associated with Tone1 through Tone 4; the meaning of Mandarin

/ti/ with Tone 1 is ‘to drop(滴)’, the Tone 2 version means ‘a bamboo flute(笛)’, and the Tone 3 and 4 meanings are ‘the bottom(底)’ and “the land(地)’ ,respectively. The other target syllable /tu/ associated with Tone1 through Tone 4, the meanings are ‘the capital(都)’ in Tone 1, ‘to read(讀)’ in Tone 2, ‘to gamble(賭)’ in Tone 3, and ‘the stomach(肚)’ in Tone 4, respectively.

Acoustic Features Extraction

The sound processing program, Praat (Version 5.1.31, Paul Boersma & David Weenink, 1992-2010), was used for extracting acoustics features. The acoustics features extraction included f0, and f0 duration. Then we divided f0 into 10 segments and got 11 points of values (0%, 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, and 100% of f0).

In this study, there are two acoustic features which we focus on. One is point-values of f0 contours, and the other is the slope between two points. The slope is expressed as

$$Slope = \frac{\Delta f_0}{\Delta T}, \qquad (1)$$

where Δf_0 is the difference between the n th 10% and $n+1$th 10% of f0 values; ΔT is the temporal interval between the n th 10% and $n+1$th 10% of f0.

Statistic Analysis

The SPSS (The Statistical Package for the Social Sciences Version 18.0) software was used for prediction and classification of Mandarin tones. We make use of the Discriminant Analysis to predict the tone from the points and the segments of the f0 contours. We took the acoustic features into groups: 0%-30%, 0%-50%, and 0%-100% points of f0 contour with and without sexual factor, and 0%-30%, 0%-50%, and 0%-100% slope of f0 contours with and without sexual factor, respectively, then analysis them.

RESULTS

In Discriminant Analysis results of the analysis of points of f0 contour shows that f0 0%-100% points of f0 contour with sexual factor is the most effective group. The result indicated that it’s about 90% to identify Tone 1 correctly, 98.3% to Tone 2, 82.5% to Tone 3, and 84.2% to Tone 4 with the sexual, respectively. Secondly, the analysis results that f0 0%-50% points of f0 contour with sexual factor are the second effective group. It identified correctly 87.5% to Tone 1, 96.7% to Tone 2, 80% and 81.7% to Tone 3 and Tone 4, separately. (Table 1.)

Table 1. Discriminant Analysis of points of f0 contour with and without sexual factor

	F0 0%~30%		F0 0%~50%		F0 0%~100%	
	with sex	without sex	with sex	without sex	with sex	without sex
Tone1	55.0	40.0	87.5	56.7	90.0	66.7
Tone2	56.7	42.5	96.7	85.8	98.3	95.0
Tone3	65.0	49.2	80.0	60.0	82.5	62.5
Tone4	71.7	41.7	81.7	66.7	84.2	70.8

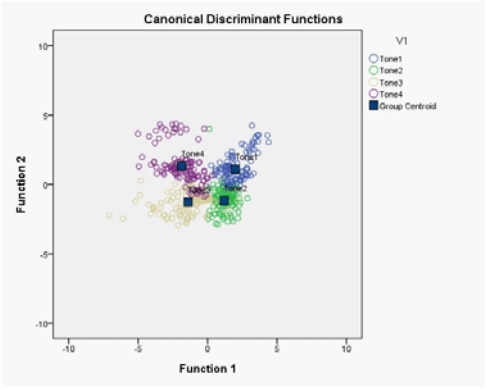

Figure 1. Canonical Discriminant Functions of the analysis of points of f0 contour shows that f0 0%-100% points of f0 contour with sexual factor.

In Discriminant Analysis results of the analysis of slope of f0 contours with and without sexual factor shows that f0 0%-100% points of f0 contour with sexual factor is the most effective group. The result indicated that it’s about 90% to identify Tone 1 correctly, 80.8% to Tone 2, 79.2% to Tone 3, and 60.8% to Tone 4 with the sexual, respectively. Secondly, the analysis results that f0 0%-100% points of f0 contour without sexual factor are the second effective group. It identified correctly 79.2% to Tone 1, 85% to Tone 2, 53.3% and 70% to Tone 3 and Tone 4, separately. (Table 2.)

Table 2. Discriminant Analysis of slope of f0 contour with and without sexual factor

	F0 Slope 0%~30%		F0 Slope 0%~50%		F0 Slope 0%~100%	
	with sex	without sex	with sex	without sex	with sex	without sex
Tone1	55.0	54.2	75.0	75.8	80.8	79.2
Tone2	48.3	42.5	76.7	72.5	79.2	85.0
Tone3	65.0	65.8	63.3	58.3	60.8	53.3
Tone4	16.7	22.5	69.2	68.3	73.3	70.0

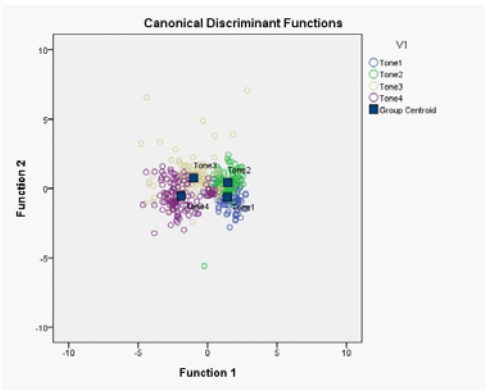

Figure 2. Canonical Discriminant Functions of the analysis of points of f0 contour shows that f0 0%-100% slope of f0 contour with sexual factor.

DISCUSSION

According to the result of Discriminant Analysis, both of points of f0 and slope of f0 contour with sexual factor indicated the most efficient approach for Mandarin tone identification. We could make a pre-process of sexual identification

before tone identification. It maybe makes a more efficient approach for tone recognition.

CONCLUSIONS

According to the results, the approach that we proposed to identify Mandarin tones by the segments of fundamental frequency contours implied that it failed to use only 3 to 4 point or segment of slope to predict Mandarin tones. The further study for prediction by segments of f0 contours is required.

REFERENCES

1 Chen S., Wang Y. (1995) Tone recognition of continuous Mandarin speech based on neural networks. IEEE Transactions on speech and audio processing 3:146-150.
2 Zhang J., Hirose K. (2000) Anchoring hypothesis and its application to tone recognition of Chinese continuous speech, IEEE; 1999.
3 Xu L., Tsai Y., Pfingst B. (2002) Features of stimulation affecting tonal-speech perception: Implications for cochlear prostheses. The Journal of the Acoustical Society of America 112:247.
4 Morton K., Torrione Jr P., Throckmorton C., Collins L. (2008) Mandarin Chinese tone identification in cochlear implants: Predictions from acoustic models. Hearing research 244:66-76.
5 Zhou N., Zhang W., Lee C., Xu L. (2008) Lexical tone recognition with an artificial neural network. Ear and hearing 29:326.