Acoust. Sci. & Tech. **24**, 4 (2003)

ACOUSTICAL LETTER

Analysis of vowel formant frequency variations between focus and neutral speech in Mandarin Chinese

Zhenglai Gu, Hiroki Mori and Hideki Kasuya*

Graduate School of Engineering, Utsunomiya University

(*Received 3 December 2002, Accepted for publication 24 January 2003*)

Keywords: Focus, Formant frequencies, Acoustic correlates, Chinese
PACS number: 43.70.Fq, 43.72.Ja [DOI: 10.1250/ast.24.192]

1. Introduction

It is widely known that acoustic correlates such as fundamental frequency (F_0), duration and intensity are reliable prosodic cues to signaling focus in speech across many languages [1–7]. F_0 movements, moreover, are generally regarded as the most reliable cue to focus.

It is also considered, however, that the prosodic cues are neither necessary nor sufficient to both signal and perceive focus. Spectrum correlates that are related to the sound color are shown as additional reliable cues to detect focus in several languages. For example, many investigations of the spectral tilt of focus vs. neutral speech indicate that there is relatively more energy in the higher frequency domain of the focused speech in general [8–10]. In addition to the spectral tilt, the focus effects on the first two formant frequencies (F_1, F_2) were investigated recently. A study reported that systematic changes in the formant frequencies of several focused vowels were observed in Japanese [3]. Erickson *et al.* compared the formant frequencies and jaw/tongue positions for two focused English vowels /æ/ and /i/ and three focused Japanese vowels /a/, /e/ and /i/ [11]. Their results suggest that there are significant differences between the focused and non-focused vowels in terms of formant frequencies and jaw/tongue positions in English similar to the results in Japanese, while there are still some language-specific differences both in formants and jaw/tongue positions. However, these investigations on formants are limited in quantity, and the same issue in Mandarin Chinese is poorly understood. No study has been done in this area in Mandarin Chinese.

Thus, the present study attempts to clarify whether and how the focused vowels systematically vary in terms of the first two formants (F_1, F_2) in Mandarin Chinese.

2. Method

2.1. Speech materials

We prepared a carrier sentence, "Wo3Shuo1 TW Zhe4 Ge4 Zi4. (I say the word TW)", into which target words (TW) were inserted. The numbers in the carrier sentence identify the four tones.

In the above sentence, the target words were underlined for indicating the focus position. Five monosyllabic target words with the same tone 1 and the same consonant /d/ except the different vowels including /a/, /i/, /e/, /o/ and /u/ were selected for the analysis of formant frequencies.

*e-mail: kasuya@klab.jp

2.2. Speakers

Five native Chinese speakers born in Beijing, two males and three females, participated in the experiment. None of them had any previous acoustic or phonetic knowledge, and they were not told the purpose of the experiment.

2.3. Recording

The speakers read each sentence five times with focus on the target words as well as with no focus on the target words. Thus, 250 (= 5 words × 1 focused and 1 neutral × 5 repetitions × 5 speakers) utterances were recorded on DAT in a soundproof room before being digitized at a sampling rate of 11.025 kHz.

2.4. Formant measurements

Formant frequencies (F_1, F_2) of all the vowels in the five monosyllabic target words were estimated by a robust ARX analysis method [12], under the conditions of equation orders $p = 12$ and $q = 0$, frame length of 35 ms, and frame shift of 5 ms. Formant frequencies of a target vowel were defined as the mean estimated values of 10 frames at the middle part of the vowel.

The ARX analysis method is a novel speech analysis-synthesis method based on an auto-regressive with exogenous input (ARX) speech production model. The method has been proved to be superior to conventional LPC-based methods in estimating low formants for very high-F_0 voices in which the estimation of low formants tend to be biased by high F_0s. The ARX method was of great benefit because in the focused speech F_0s were often raised to very high values (up to 410 Hz or more).

Besides the high F_0 problem, we also encountered an additional problem of spurious formants in some cases. We manually pruned the spurious formants, if any, based on the segmental and bandwidth information.

2.5. Results

Figure 1 shows the first two formant frequencies in the (F_1, F_2) plane. A point means an average value of the five repetitions of the same vowel by the same speaker. The arrows indicate the directions of the formants movement from neutral to focus. As can be seen, the focused vowel formants plotted in the (F_1, F_2) plane tend to move apart from those of neutral in general.

The difference values of the first two formants of each speaker between focus and neutral are shown in Table 1. A positive value means some increase of formant frequency by the effect of focus. We conducted a one-paired *t*-test for the mean difference value of the same vowel across speakers. The

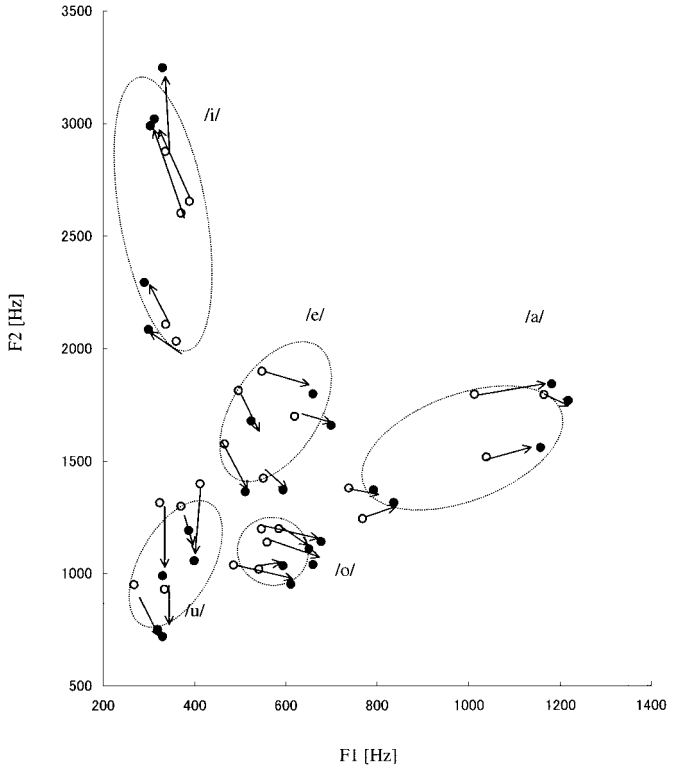

Fig. 1 Distribution of the formants (F_1, F_2) of the focused and neutral vowels uttered by 5 speakers.

Table 1 Neutral formant values (F_1, F_2) subtracted from focused values of 5 speakers as well as their mean difference values. The mean values in bold indicate significance by a one-paired t-test with $a = 0.05$.

Vowel	Formant	Speakers					Mean [Hz]
		Ma [Hz]	Mb [Hz]	Fa [Hz]	Fb [Hz]	Fc [Hz]	
/a/	F_1	68	54	44	118	209	**98**
	F_2	71	−8	−25	41	45	24
/e/	F_1	50	43	112	80	28	**62**
	F_2	−207	−44	−100	−40	−135	**−105**
/i/	F_1	−61	−47	−67	−77	−6	**−51**
	F_2	52	186	387	365	372	**272**
/u/	F_1	−4	51	−13	17	6	11
	F_2	−210	−200	−343	−107	−326	**−237**
/o/	F_1	126	42	160	71	148	**109**
	F_2	−130	16	−47	−58	−268	−97

results show that there are some significant differences between focus and neutral utterances in terms of vowel formant frequencies (F_1, F_2). The F_1 values of /a/ are significantly increased by the effect of focus but there is insignificant change in F_2 of /a/. Vowel /o/ tends to have a similar pattern to vowel /a/, except that the majority of speakers strongly decreased F_2 of /o/. The F_1 values are decreased and F_2 values are increased significantly for vowel /i/, which is in contrast to the change pattern for vowel /e/. For the focused vowel /u/, only the F_2 values are significantly decreased.

3. Discussion

The results of the present study indicate that formant frequencies (F_1, F_2) are additional reliable cues to signaling focus in Mandarin Chinese. The different change patterns for the vowels indicate some different ways of articulation for focus. For example, the significant increase of F_1 of focused vowel /a/ indicates that the speakers open their mouths more for focus utterance than for neutral utterance. The increased F_2 and decreased F_1 of focused vowel /i/ indicates more front and more closed articulation movement. Similar formant change tendencies for the two vowels were also reported in Japanese. However, the significant decrease of F_2 of focused vowel /u/ indicating more lip-rounding articulation movement was not observed for the same vowel /u/ in Japanese. These formant changes have proved to contribute to the perceived prominence [13].

References

[1] D. R. Ladd, *Intonational Phonology* (Cambridge University Press, Cambridge, 1996), p. 153, p. 180.
[2] G. Fant, A. Kruckenberg and J. Liljencrants, "Acoustic-phonetic analysis of prominence in Swedish," in *Intonation: Analysis, Modelling and Technology* (Kluwer Academic Publisher, Dordrecht, 2000), pp. 55–86.
[3] K. Maekawa, "Effects of focus on duration and vowel formant frequency in Japanese," in *Computing Prosody: Computational Models for Processing Spontaneous Speech* (Springer Verlag, New York, 1996), pp. 129–153.
[4] M. Heldner and E. Strangert, "Temporal effects of focus in Swedish," *J. Phonet.*, **29**, 329–361 (2001).
[5] E. Gårding, "Speech act and tonal pattern in standard Chinese constancy and variation," *Phonetica*, **44**, 13–29 (1987).
[6] Y. Xu, "Effects of tone and focus on the formation and alignment of f_0 contours," *J. Phonet.*, **27**, 55–105 (1999).
[7] Z. Gu, H. Mori and H. Kasuya, "Prosodic variations in disyllabic meaningful words focused with different patterns in Mandarin Chinese," *Acoust. Sci. & Tech.*, **24**, 111–119 (2003).
[8] M. Heldner, "Spectral emphasis as a perceptual cue to prominence," *Speech Music Hear.*, **42**, 51–57 (2001).
[9] N. Campbell and M. E. Bechman, "Stress, prominence, and spectral tilt," *ESCA Tutorial/Research Workshop on Intonation: Theory, Models, Applications*, Athens, Greece, pp. 67–70 (1997).
[10] A. Sluijter and V. van Heuven, "Spectral balance as an acoustic correlate of linguistic stress," *J. Acoust. Soc. Am.*, **100**, 2471–2485 (1996).
[11] D. Erickson, M. Hashi and K. Maekawa, "Ariticulatory and acoustic correlates of prosodic contrasts: A comparative study of vowels in Japanese and English," *Proc. Spring Meet. Acoust. Soc. Jpn.*, pp. 265–266 (2000).
[12] T. Ohstsuka and H. Kasuya, "Robust ARX-based speech analysis method taking voicing source train into account," *J. Acoust. Soc. Jpn.* (*J*), **58**, 386–397 (2002).
[13] Z. Gu, H. Mori and H. Kasuya, "Prosodic and segmental variations of focused vords in Mandarin Chinese: Analysis, synthesis and perception," *Tech. Rep. IEICE*, SP2002-94, pp. 1–6 (2002).