Frequency Analysis of the Vowels in Cantonese from 50 Male and 50 Female Speakers

Eric Zee
Phonetics Lab, Department of Chinese, Translation and Linguistics
City University of Hong Kong
E-mail: eric.zee@cityu.edu.hk

ABSTRACT

The purpose of the study is to obtain measurements of the formant frequencies of the 7 long vowels [iː yː ɛː œː aː ɔː uː] in the (C)V: syllables and 7 half-long vowels [i y ɛ œ a ɔ u] in the (C)V:S syllables, and 4 short vowels [ɪ ɐ ɵ ʊ] in the (C)VS syllables in Cantonese. Results show that (1) in the F_1/F_2 plane the half-long vowels except [a] are centralized relative to the long vowels; (2) the center F-values for [ɪ ɐ ʊ] are similar to those for [ɛː aː ɔː], similar to what is reported in Ladefoged and Maddieson (1996); (3) both (1) and (2) are true for the male and female speakers; and (4) for both the long and half-long vowels, the distance between the mid vowels and the high vowels in the F_1/F_2 plane is larger than the distance between the mid vowels and the low vowel [a] for the female speakers, but for the male speakers the distances are similar.

1. INTRODUCTION

In Cantonese there are seven long vowels [iː yː ɛː œː aː ɔː uː] in the (C)V: syllables, seven half-long vowels [i y ɛ œ a ɔ u] in the (C)V:S syllables (S = stop [-p -t] or [-k]), and four short vowels [ɪ ɐ ɵ ʊ] in the (C)VS syllables [1]. The IPA vowel symbols used in this study are those that appear in the illustrations of IPA in the *Handbook of the IPA* [2].

A large-scale investigation of the formant frequency patterns of the vowels and diphthongs in Hong Kong Cantonese (HKC) has been carried out in the Phonetics Lab at the City University of Hong Kong, analyzing speech data from 100 native speakers of HKC, 50 male and 50 female. This paper reports some of the results of the spectral analysis of the vowels in HKC.

2. METHOD

2.1 Subjects

50 male and 50 female native speakers of Hong Kong Cantonese provided the speech data. They were all university students aged between 18 and 21 with no history of speech or hearing disorders.

2.2 Recording and test material

Audio recordings were made of subjects reading a list of 18 meaningful test monosyllables. The test monosyllables which contained the 7 long vowels [iː yː ɛː œː aː ɔː uː] were [ji˥] 'medicine', [jy˥] 'to grant', [ɛ˥] 'to complain', [hœ˥] 'boots', [ha˥] 'shrimp', [hɔ˥] 'harsh', and [wu˥] 'black'. All the test monosyllables are associated with a long '55' tone. Those contained the 7 half-long vowels [i y ɛ œ a ɔ u] included [hit˧] 'to rest', [hyt˧] 'blood', [hɛk˧] 'to eat', [sœk˧] 'to whittle', [hak˧] 'guest', [hɔt˧] 'thirsty', and [fut˧] 'wide'. All as the transcription indicates are associated with a mid '33' tone, rather than a high '55' tone. This is due to the fact that in Cantonese a large majority of the half-long vowels are only associated with a '33' tone. The test monosyllables [ɪk˥] 'hiccup', [sɵt˥] 'cricket', [hɐt˥] 'to beg', and [hʊk˥] 'to cry' which contain the 4 short vowels [ɪ ɐ ɵ ʊ] are all associated with a short high '5' tone. Each of the test words was embedded in a carrier sentence, [ŋɔ˩˧ jiu˧ tʊk˧ _____ pei˧˥ lei˩˧ tʰɛŋ˥] "I want (to) read _____ for you (to) listen.". The speakers were instructed to read the word list at a normal rate of speech. The order of the test words in the list was randomized. Five readings of the word list were recorded. The total number of the test tokens was 9,000 (100 speakers x 18 test monosyllables x 5 repetitions). The recordings were performed in a sound-proof booth (IAC) in the Phonetics Lab at the City University of Hong Kong.

2.3 Spectral processing

In this study, CSL4300B (Computerized Speech Lab), a speech analysis software by Kay Elemetric of USA, was used for spectral analysis of the vowels. The software provides 16-bit input A/D conversion and the hardware can capture speech signal over a variety of sampling rate. In this study, speech data were captured at a sampling rate of 10,000 samples per second, producing an upper frequency cutoff of 5,000 Hz. LPC was performed at the midpoint of the vowel using the pitch synchronous method. In the case of the short vowels, the midpoint, or the midpoint of the steady-state portion, of the formant trajectories was selected as the location for LPC analysis.

3. RESULTS AND DISCUSSION

3.1 Result I: male speakers

Figure 1 shows the positions of the vowel ellipses for the 7 long Cantonese vowels [iː yː ɛː œː aː ɔː uː] in the (C)V: syllables in the F_1/F_2 plane for 50 male speakers. Each of the vowel ellipses is drawn with radii of two standard deviations along the two principal components of each vowel cluster [3]. Each vowel cluster consists of 250 data points (5 repetitions x 50 speakers) in the F_1/F_2 plane for a particular vowel. Each data point represented by an IPA vowel symbol in the chart is determined by F_1 and F_2 of a vowel with F_1 plotted on the ordinate and F_2 on the abscissa on Bark scales. Thus, each vowel ellipse indicates the dispersion of 250 data points or utterances of a vowel in the F_1/F_2 plane for 50 male speakers. As can be seen in Figure 1, the vowel ellipses for the vowels [ɛː œː aː ɔː uː] in the (C)V: syllables occupy distinct locations in the acoustical vowel space. As for the vowel ellipses for the vowels [iː] and [yː], they overlap partially. The main

ISBN 1-876346-48-5 © 2003 UAB

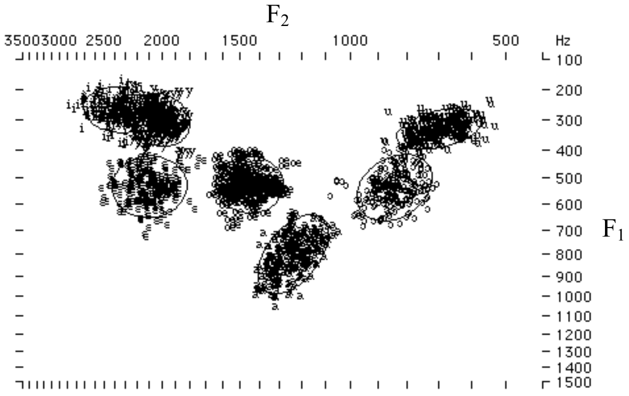

Figure 1: The vowel ellipses for the 7 long vowels [iː yː ɛː œː aː ɔː uː] in the (C)V: syllables (50 male speakers).

Vowel	F_1		F_2		F_3	
	Mean	s.d.	Mean	s.d.	Mean	s.d.
[iː]	264.82	38.06	2368.88	165.70	3327.04	184.45
[yː]	297.91	44.41	2026.07	120.59	2400.19	142.09
[ɛː]	533.58	57.65	2104.27	151.15	2810.02	170.54
[œː]	526.16	52.25	1452.96	89.03	2471.16	159.17
[aː]	805.49	86.75	1240.99	80.16	2817.78	220.86
[ɔː]	537.21	62.23	844.28	65.44	2872.82	240.36
[uː]	329.44	34.60	697.81	66.64	2710.73	218.26

Table 1: Mean F_1, F_2, and F_3 values (in Hz) (n = 250) and their standard deviations for the 7 long vowels [iː yː ɛː œː aː ɔː uː] in the (C)V: syllables in Cantonese (50 male speakers).

difference between [iː] and [yː] in the (C)V: syllables is in F_3 (3327.04 Hz for [iː]; and 2400.19 Hz for [yː]) as shown in Table 1.

An observation is made about the number of vowel heights in Cantonese. Based on the positions of the vowel ellipses in Figures 1, 2, and 3, there are only three discernible vowel heights, i.e., high [iː yː uː i y u], mid [ɛː œː ɔː ɛ œ ɔ ɪ ɵ ʊ], and low [aː a ɐ].

Figure 2 shows the vowel ellipses for the 7 long vowels [iː yː ɛː œː aː ɔː uː] (thick line and larger IPA symbols) in the (C)V: syllables (Figure 1) superimposed onto the vowel ellipses for the 7 half-long vowels [i y ɛ œ a ɔ u] (thin line and smaller IPA symbols) in the (C)V:S syllables, displaying the relative positions of the vowel ellipses for the two sets of vowels in Cantonese in the F_1/F_2 plane. As can be seen in Figure 2, the positions of ellipses for the 7 half-long vowels in the (C)V:S syllables relative to those for the 7 long vowels in the (C)V: syllables are all lowered due to an increase in F_1. And, with the exception of [a], the relative positions of the ellipses for the other 6 half-long vowels [i y ɛ œ ɔ u] to those for the long vowels are centralized in the F_1/F_2 plane, i.e., an increase in F_1 and decrease in F_2 for [i y ɛ œ] and an increase in both F_1 and F_2 for [u ɔ]. The positions of the vowel ellipses for the 6 half-long vowels relative to those for the long vowels may be considered as a case of vowel undershoot of the "bull's eye formant pattern" [4] which is associated with the longer vowels (approximately 350 ms in the case of the Cantonese long vowels) in the optimal context of the (C)V:

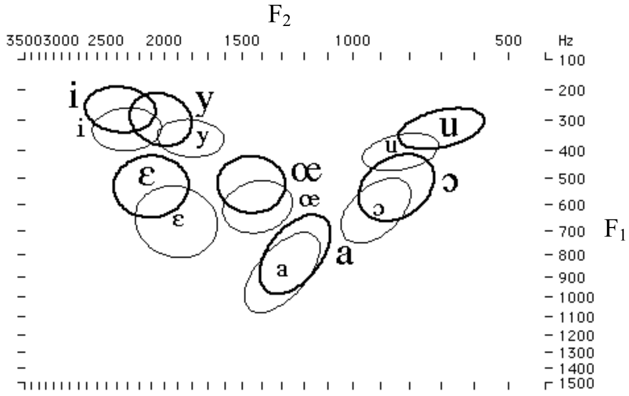

Figure 2: The vowel ellipses for the 7 long vowels [iː yː ɛː œː aː ɔː uː] in the (C)V: syllables (thick line and large IPA symbols) superimposed on the vowel ellipses for the half-long vowels [i y ɛ œ a ɔ u] in the (C)V:S syllables (thin line and small IPA symbols) (50 male speakers).

syllables. The undershoot is assumed to have been caused by the shorter vowel duration (approximately 200 ms) of the half-long Cantonese vowels in the (C)V:S syllables. However, this cannot explain why the half-long [a] does not undergo centralization as do the other half-long vowels. Quite on the contrary, the half-long [a] may be considered as a case of vowel overshoot relative to the long [a]. It is not clear why the half-long [a] should behave differently when the duration is shortened. Despite the differences in the locations of the vowel ellipses, the 7 half-long vowels have not been transcribed differently from the 7 long vowels [1]. It is assumed that the differences in perceptual quality between the corresponding members of the two sets of vowels are not large enough to warrant different IPA symbols.

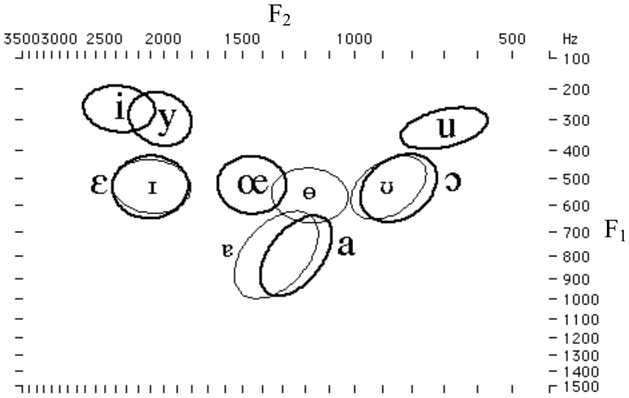

Figure 3: The vowel ellipses for the long vowels [iː yː ɛː œː aː ɔː uː] in the (C)V: syllables (thick line and large IPA symbols) superimposed on the vowel ellipses for the short vowels [ɪ ɐ ɵ ʊ] in the (C)VS syllables (thin line and small IPA symbols) (50 males speakers).

Figure 3 show the vowel ellipses for the long vowels [iː yː ɛː œː aː ɔː uː] (thick line and larger IPA symbols) in the (C)V: syllables (Figure 1) superimposed onto the vowel ellipses for the short vowels [ɪ ɐ ɵ ʊ] (thin line and smaller IPA symbols) in the (C)VS syllables, displaying the relative positions of the vowel ellipses for the two sets of vowels in Cantonese in the F_1/F_2 plane. As can be seen, the vowel ellipses for the short vowel [ɪ] and long vowel [ɛː] overlap extensively. This is also true for the ellipses for the short [ʊ] and long [ɔː]. As for the vowel ellipse for the short vowel [ɐ], it also overlaps with the vowel ellipse for the long vowel [aː], although the degree of overlap is less than that for the other two cases. And, as shown in Table 2, the differences in the mean F_3 value between [ɪ] and [ɛː], [ʊ] and [ɔː], and [ɐ] and [aː] are also small. Despite the similarities in F-values, [ɪ] and [ɛː], [ʊ] and [ɔː], or [ɐ] and [aː] have been transcribed differently in most of the past studies [1, 5, 6].

Vowel	F_1		F_2		F_3	
	Mean	s.d.	Mean	s.d.	Mean	s.d.
[ɪ]	529.27	48.82	2099.81	156.35	2,700.54	213.04
[ɛː]	533.58	57.65	2104.27	151.15	2810.02	170.54
[ʊ]	533.34	58.04	881.07	64.63	2817.99	246.10
[ɔː]	537.21	62.23	844.28	65.44	2872.82	240.36
[ɐ]	799.32	94.48	1336.08	100.48	2782.43	193.31
[aː]	805.49	86.75	1240.99	80.16	2817.78	220.86

Table 2: Mean F_1, F_2, and F_3 (in Hz) values (n = 250) and their standard deviations for the short vowels [ɪ], [ʊ], and [ɐ] in the (C)VS syllables and the long vowels [ɛː], [ɔː], and [aː] in the (C)Vː syllables in Cantonese (50 male speakers).

Ladefoged and Maddieson [7] pointed out that there are "discrepancies between acoustic plot and traditionally linguistic classification. In particular, the vowels ɪ and ʊ, which are traditionally classed as high, are acoustically closer to the mid-vowels ɛ and o rather than to i and u" (p. 285). Thus, the discrepancies between the center F-values and vowel transcription of the Cantonese vowels is not an isolated case. As the transcription of the Cantonese short vowels such as [ɪ ʊ ɐ ɵ] in the past studies [1, 5, 6] is assumed to be based on the auditory impression of the vowel quality rather than the center F-values for any vowels, an explanation may be that the center F-values for any one of the short vowels in question are not representative of the spectral characteristic of the entire vowel. Due to a short duration of approximately 120 ms and to the fact that the short vowels in Cantonese are usually preceded by a consonant and always followed by a consonant, the steady-state portions of the formant trajectories are short (20-40 ms). Tape-cutting experiments of Schatz as well as Harris show that for the vowels in the (C)VC syllables the "vowel quality cannot be discretely localized in any single portion of the syllable, but is distributed throughout the period during the voicing is present" (from Strange and Verbrugge [8], p. 213). It follows that the perceptual impression of the short vowels in Cantonese are not expected to correspond to their center F-values, which explains why those long and short vowels which have similar F-values were transcribed differently by the authors of the past studies [1, 5, 6].

3.2 Result II: male and female speakers

Due to the differences in the size and length of the vocal tract, the F-values for the vowels are expected to be different between the female and male speakers. In American English [9], Dutch [10, 11] and Korean [12], the F-values for the vowels are larger for the female than male speakers. In this section, the F-values for the vowels in Cantonese for the female speakers are compared with those for the male speakers. Figure 4 shows the positions of the vowel ellipses for the 7 long Cantonese vowels [iː yː ɛː œː aː ɔː uː] in the (C)Vː syllables in the F_1/F_2 plane for 50 male (thick line and large IPA symbols) and 50 female (thin line and small IPA symbols) speakers. As shown in the figure, the positions of the vowel ellipses for the female speakers are shifted toward the lower left corner in the acoustical vowel space relative to those for the male speakers, with the exception of [u], for which there is no observable difference in the F_1 dimension. Table 3 shows the mean F_1, F_2, and F_3 values for the 7 long vowels for both male and female speakers. As shown in the table, in all the cases the mean F-value is larger for the female speakers than the male speakers, except for [u] the mean F_1 value for the female speakers (319.34 Hz, s.d. = 43.74) is in fact 10.1 Hz smaller than that for the male speakers (329.44 Hz, s.d. = 34.60). An examination of the F_1 values for the 100 individual speakers shows that it is not always the case in which [u] has a smaller F_1 value for the female speakers than the male speakers. However, result of the two-tailed grouped data t-test shows that the difference in F_1 value between the male and female speakers is significant ($p < 0.05$). In any case, it is not clear as to why the F_1 value for [u] does not follow the general pattern.

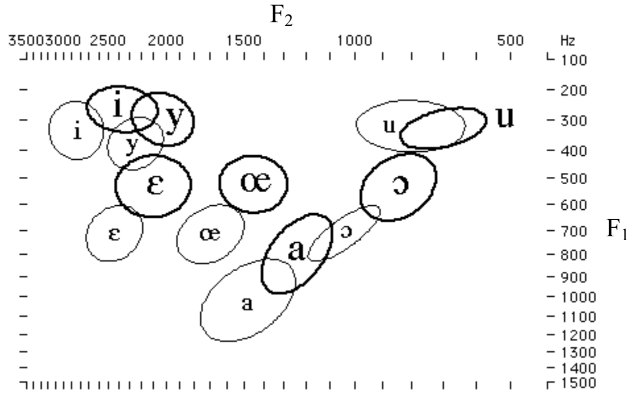

Figure 4: The vowel ellipses for the long vowels [iː yː ɛː œː aː ɔː uː] in the (C)Vː syllables for the male (thick line and large IPA symbols) and female (thin line and small IPA symbols) speakers (50 male and 50 female speakers).

Vowel	F_1		F_2		F_3	
	Male	Female	Male	Female	Male	Female
[iː]	264.82	333.84	2368.88	2850.34	3327.04	3630.78
[yː]	297.91	378.82	2026.07	2253.70	2400.19	2874.45
[ɛː]	533.58	714.90	2104.27	2445.91	2810.02	3269.49
[œː]	526.16	718.10	1452.96	1702.18	2471.16	2977.74
[aː]	805.49	1019.05	1240.99	1487.60	2817.78	3239.25
[ɔː]	537.21	715.47	844.28	1045.09	2872.82	3120.08
[uː]	329.44	319.34	697.81	806.24	2710.73	3172.24

Table 3: Mean F_1, F_2, and F_3 values (in Hz) values (n = 250; standard deviations not shown) for the 7 long vowels [iː yː ɛː œː aː ɔː uː] in the (C)VːS syllables in Cantonese (50 male and 50 female speakers).

Based on the F-values in Table 3, it is observed that the difference in F_1 value between the male and female speech

ISBN 1-876346-48-5 © 2003 UAB

is negatively correlated with vowel height. The observation is based on the fact that the differences between the male and female speech in F_1 value for the high vowels [iː yː uː] (69.02 Hz, 80.91 Hz, 10.1 Hz, respectively) are smaller than those for the mid vowels [ɛː œː ɔː] (181.32 Hz, 191.94 Hz, 178.25 Hz, respectively) and the difference is the largest for the low vowel [a] (213.56 Hz). Furthermore, the differences in F_2 value between the male and female speech are smaller for the back vowels [uː ɔː] (108.43 Hz, 201.81 Hz, respectively) than the non-back vowels [iː yː ɛː œː aː] (481.46 Hz, 227.63 Hz, 341.46 Hz, 249.22 Hz, 246.61 Hz, respectively). Thus, the difference in F_2 value between the male and female speech is negatively correlated with vowel backness.

Figure 5 and Figure 6 show the vowel ellipses for the 7 half-long vowels [i y ɛ œ a ɔ u] in the (C)V:S syllables and the 4 short vowels [ɪ ɐ ɵ ʊ] in the (C)VS syllables, respectively, for the male (thick line and large IPA symbols) and female (thin line and small IPA symbols) speakers.

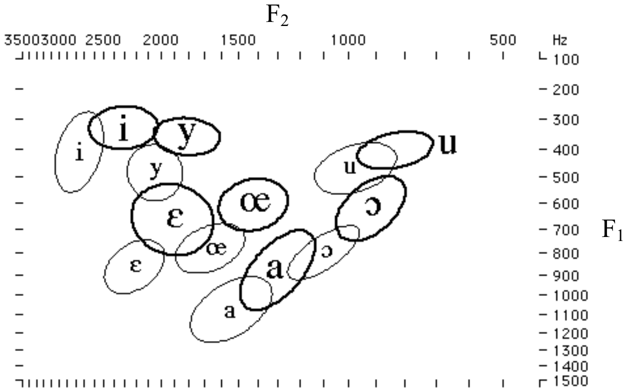

Figure 5: The vowel ellipses for the half-long vowels [i y ɛ œ a ɔ u] in the (C)V:S syllables for the male (thick line and large IPA symbols) and female (thin line and small IPA symbols) speakers (50 male and 50 female speakers).

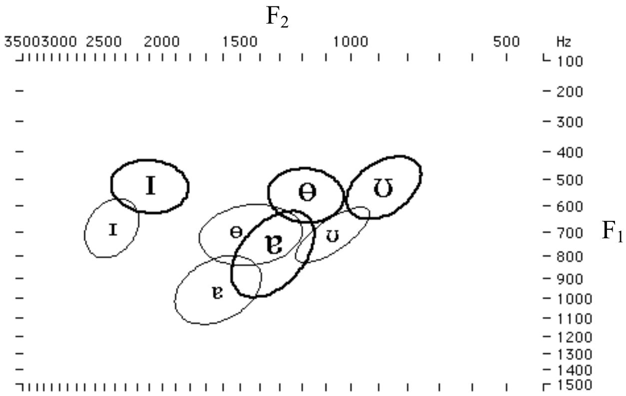

Figure 6: The vowel ellipses for the short vowels [ɪ ɐ ɵ ʊ] in the (C)VS syllables for the male (thick line and large IPA symbols) and female (thin line and small IPA symbols) speakers (50 male and 50 female speakers).

Again, the positions of the vowel ellipses for the 7 half-long and 4 short vowels for the female speakers are shifted toward the lower left corner in the acoustical vowel space, relative to those for the male speakers.

Aside from the differences in the absolute F-values between the male and female speakers, the F-value patterns for the two groups of speaker also differ. As can be seen in Figure 4 and Figure 5, for the female speakers the distance between the vowel ellipses for the mid and high vowels is larger than the distance between the vowels for the mid and low vowels. This is true for both the long and half-long vowels. As for the male speakers, the distances are similar. The phenomenon is also found in the vowel system in another major Chinese dialect, Beijing Mandarin [13].

4. SUMMARY

The study has presented the spectral characteristics of vowels in Hong Kong Cantonese based on the results of formant frequency analysis of the speech data from 50 male and 50 female speakers.

.

REFERENCES

[1] Y.-R. Chao, *Cantonese Primer*, New York, Greenwood Press, 1947.

[2] E. Zee, "Chinese (Hong Kong Cantonese)", *Handbook of the International Phonetic Associaion*, Cambridge, Cambridge University Press, 1999.

[3] S.F. Disner, *Vowel Quality: the Relation Between Universal and Language-Specific Factors*, *UCLA WPP*, 58, UCLA, Los Angeles, 1983.

[4] B.E. Lindblom, "Spectrographic study of vowel reduction", *JASA*, 35.11.1773-1781, 1963.

[5] D. Kao, *Structure of the Syllable in Cantonese*, The Hague, Mouton, 1971.

[6] A. Hashimoto, *Cantonese Phonology.* Cambridge, Cambridge University Press, 1972.

[7] P. Ladefoged and I. Maddieson, *The Sounds of the World's Languages*. London, Blackwell, 1996.

[8] W. Strange and R.R. Verbrugge, "Consonant environment specifies vowel identity", *JASA*, 60.1.213-224, 1976.

[9] J. Hillenbrand, L.A. Getty, M.J. Clark and K. Wheeler, "Acoustic characteristics of American English vowels", *JASA*, 97.1.3099-3111, 1995.

[10] L. Pols, H. Tromp and R. Plomp, "Frequency analysis of Dutch vowels from 50 male speakers', *JASA*, 50.1093-1103, 1973.

[11] D.J.P.J. van Nierop, L.C.W. Pols and R. Plomp, "Frequency analysis of Dutch vowels from 25 female speakers", *Acustica*, 29.110-118, 1973.

[12] B. Yang, "An acoustical study of Korean monophthongs produced by male and female speakers", *JASA*, 91(4).2280-2283, 1992.

[13] E. Zee and W.S. Lee, "An Acoustical analysis of the vowels in Beijing Mandarin", *Proceedings of Eurospeech 2001, Scandinavia*, September 3-7, 2001, Aalborg, Denmark, 643-646.

[*The research was supported by a Strategic Research Grant from the City University of Hong Kong*.]