International Symposium on Chinese Spoken Language Processing (ISCSLP 2000)
Fragrant Hill Hotel, Beijing
October 13-15, 2000

FREQUENCY ANALYSIS OF THE VOWELS IN CANTONESE

Eric ZEE
Phonetics Lab, Dept. of CTL, City University of Hong Kong, Hong Kong
ctlzee@cityu.edu.hk

ABSTRACT

The study investigates the spectral characteristics of the vowels in Cantonese. Results show that (1) the vowels in the (C)V:S syllables undershoot in the formant frequencies relative to the canonical target formant pattern associated with the same vowels in the (C)V: syllables; (2) the center formant frequency values for the vowels in the (C)VS syllables are not representative of the quality of the vowels due to short vowel duration; and (3) the center formant frequencies for the vowels in the (C)V: and (C)V:S syllables can be useful in terms of vowel transcription.

1. INTRODUCTION

It is generally agreed among scholars that phonetically in Cantonese there are seven long vowels in the (C)V: syllables, seven half-long vowels in the (C)V:S syllables (S = stop [-p -t] or [-k]), and four short vowels in the (C)VS syllables (Jones and Woo, 1912; Wong, 1941; Chao, 1947; Yuan, et al., 1960; Kao, 1971; Hashimoto, 1972). However, as shown in Table 1, they differ in transcribing some of the vowels, and the difference lies mainly in the transcription of the short vowels, for instance, the Cantonese short, low vowel has been transcribed as [ɑ] (Jones and Woo, 1912; Chao, 1947), [a] (Yuan, et al., 1960), or [ɐ] (Wong, 1941; Kao, 1971; Hashimoto, 1972). The Cantonese short, high, front vowel which occurs before [-k] or [-ŋ] has been transcribed as [i] (Jones and Woo, 1912; Wong, 1941), [e] (Chao, 1947), or [ɪ] (Yuan, et al., 1960; Kao, 1971; Hashimoto, 1972). And, the Cantonese short, high, back vowel which also occurs before a velar stop or nasal has been transcribed as [u] (Jones and Woo, 1912; Wong, 1941; Chao, 1947) or [ʊ] (Yuan, et al., 1960; Kao, 1971; Hashimoto, 1972). As for the short, mid, round vowel, it has been transcribed as [œ] (Jones and Woo, 1912; Wong, 1941; Yuan, et al., 1960), [ɵ] (Chao, 1947), or [ø] (Kao, 1971; Hashimoto, 1972). As no articulatory or acoustical information about these vowels are given to substantiate the choice of the IPA vowel symbol in any of these studies, the inconsistencies in the description of the vowels are assumed to have been caused by the authors' differing impression of the speech samples from perhaps speakers of different socio-biological backgrounds. The purpose of this study is to investigate the spectral characteristics of the vowels in Cantonese.

Past studies	Long vowels in the (C)V: syllables	Half-long vowels in the (C)V:S syllables	Short vowels in the (C)VS syllables
Jones & Woo (1912)	[i y ɛ ɑ u ɔ œ]	[i y ɛ ɑ u ɔ œ]	[i ɑ u œ]
Wong (1941)	[i y ɛ a u ɔ œ]	[i y ɛ a u ɔ œ]	[i ɐ u œ]
Chao (1947)	[i y ɛ ɑ u ɔ œ]	[i y ɛ ɑ u ɔ œ]	[e ɑ u ɵ]
Yuan, et al. (1960)	[i y ɛ a u ɔ œ]	[i y ɛ a u ɔ œ]	[ɪ a ʊ œ]
Kao (1971)	[i y ɛ ɑ u ɔ œ]	[i y ɛ ɑ u ɔ œ]	[ɪ ɐ ʊ ø]
Hashimoto (1972)	[i y ɛ A u ɔ ø]	[i y ɛ A u ɔ ø]	[ɪ ɐ ʊ ø]

Table 1. Cantonese vowels in the (C)V:, (C)V:S, and (C)VS syllables as transcribed in the past studies.

2. METHOD

2.1 **Subjects**. Speakers who provided the speech data consisted of 10 male university undergraduate students, all native speakers of Hong Kong Cantonese, with no history of speech or hearing disorders.

2.2 **Recording and Test Materials**. Audio recordings were made of subjects reading a list of 18 test meaningful monosyllables which consisted of 7 Cantonese long vowels [i y ɛ œ a ɔ u] in the (C)V: syllables [ji˥ jy˥ ɛ˥ hœ˥ ha˥ hɔ˥ wu˥], 7 Cantonese half-long vowels [i y ɛ œ a ɔ u] in the (C)V:S syllables [hit˧ hyt˧ hɛk˧ sœk˧ hak˧ hɔt˧ fut˧], and 4 Cantonese short vowels [ɪ ɵ ɐ ʊ] in the (C)VS syllables [ɪk˥ sɵt˥ hɐt˥ hʊk˥]. Each test word 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. Five readings of the word list were recorded. Thus, the total number of the test tokens was 900 (10 speakers x 18 test monosyllables x 5 repetitions). The order of the test words in the list was randomized. 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, CSL4300 (Computerized Speech Lab), a speech analysis software by Kay Elemetric of USA was used for spectral analysis of the vowels. CSL 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 which was performed at the midpoint of the vowel provided the resonance frequencies (F_1 F_2 F_3), 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

Figure 1 shows the positions of the vowel ellipses for the 7 long Cantonese vowels [i y ɛ œ a ɔ u] in the (C)V: syllables in the acoustical vowel chart for 10 male speakers. Each of the vowel ellipses is drawn with radii of two standard deviations along the two principal

components of each vowel cluster (Disner, 1983). Each vowel cluster contains 50 data points (5 repetitions x 10 speakers) in the F_1/F_2 plane for a particular vowel. Each data point represented by an IPA vowel symbol in the

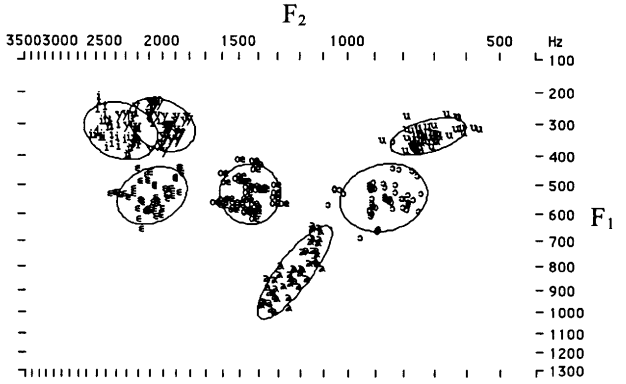

Figure 1. Vowel ellipses for the 7 Cantonese long vowels [i y ɛ œ a ɔ u] in the (C)V: syllables (10 male speakers).

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 50 data points or utterances of a vowel in the F_1/F_2 plane for 10 male speakers. As can be seen in Figure 1, the vowel ellipses for the vowels [ɛ œ a ɔ u] in the (C)V: syllables occupy distinct locations in the acoustical vowel space. As for the vowel ellipses for the vowels [i] and [y], they overlap partially. Statistical data show that the difference in the mean F_1 and F_2 values between the vowels are significant at 0.05 level for F_1 and 0.0001 level for F_2. The phonetic difference between [i] and [y] in the (C)V: syllables is enhanced by the large difference in the mean F_3 value (3,213 Hz for [i]; and 2,412 Hz for [y]).

Figure 2 shows the positions of the vowel ellipses for the 7 Cantonese half-long vowels [i y ɛ œ a ɔ u] in the (C)V:S syllables. The pattern of the positions of the vowel ellipses for these Cantonese vowels in the (C)V:S

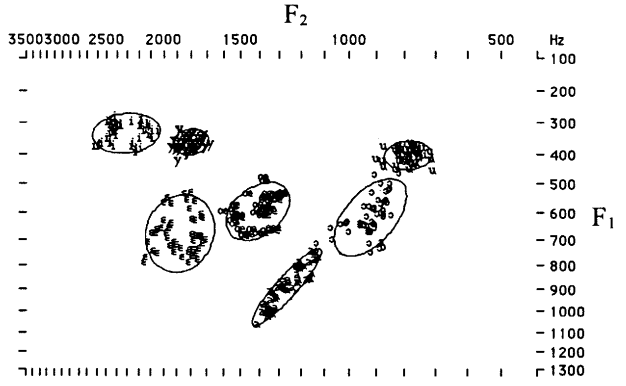

Figure 2. Vowel ellipses for the 7 Cantonese half-long vowels [i y ɛ œ a ɔ u] in the (C)V:S syllables (10 male speakers).

syllables is similar to the pattern for the Cantonese long vowels (Figure 1), except for the fact that the vowel ellipses for [i] and [y] in the (C)V:S syllables do not overlap. Figure 3 shows the vowel ellipses for the long vowels [i y ɛ œ a ɔ u] (represented with larger IPA symbols) in the (C)V: syllables (Figure 1) superimposed onto the vowel ellipses for the half-long vowels [i y ɛ œ a ɔ u] (represented with smaller IPA symbols) in the (C)V:S syllables (Figure 2), displaying the relative positions of the vowel ellipses for the two sets of vowels in Cantonese in the acoustical vowel space. As can be seen in Figure 3, the ellipses for the 7 half-long vowels in the (C)V:S syllables relative to those for the long vowels in the (C)V: syllables, with the exception of [a], are centralized in the acoustical vowel space, i.e., an increase

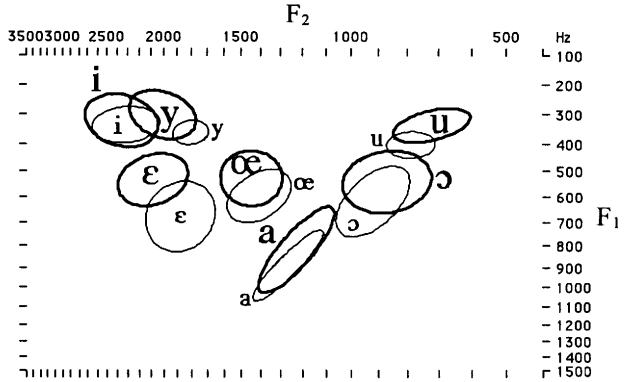

Figure 3. The vowel ellipses for the long vowels [i y ɛ œ a ɔ u] in the (C)V: syllables (represented with larger IPA symbols) superimposed on the vowel ellipses for the half-long vowels [i y ɛ œ a ɔ u] in the (C)V:S syllables (represented with smaller IPA symbols).

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]. As for the vowel ellipse for the half-long vowel [a], it is lowered in comparison with the vowel ellipse for the long vowel [a]. The positioning of the vowel ellipses for the half-long vowels relative to those for the long vowels is a case of vowel undershoot of the "bull's-eye formant pattern" (Lindblom, 1963, p. 1779) which is associated only with the long vowels (approx. 350 ms) in the optimal context of the (C)V: syllables. The undershoot is due to the shorter vowel duration (approx. 200 ms) of the half-long vowels in the (C)V:S syllables. Despite being undershot, the half-long vowels have not been transcribed differently from the long vowels in the past studies of Cantonese (Jones and Woo, 1912; Wong, 1941; Chao, 1947; Yuan, et al., 1960; Kao, 1971; Hashimoto, 1972). It is assumed that the differences in phonetic quality between the corresponding members of the two sets of vowels are not large enough to warrant different IPA symbols.

As can be seen in Figure 3, the vowel ellipses for the vowels [ɛ œ ɔ], whether they are long in the (C)V: syllables (Figure 1) or half-long in the (C)V:S syllables (Figure 2), are located in between the vowels ellipses for [i y u] and the vowel ellipse for [a]. Thus, in Cantonese only three levels of vowel height, i.e., high, mid, and low may be distinguished. The close-mid and open-mid levels which are differentiated in the IPA vowel chart are not identified in Cantonese. The IPA vowel chart basically has four levels of vowel height, i.e., high, close-mid, open-mid, and low, plus an extra mid level exclusively for the mid-central schwa [ə].

A comparison of the vowel symbols used to represent the long vowels in the (C)V: syllables or half-long vowels in the (C)V:S syllables among the past studies of Cantonese

listed in Table 1 shows that the low vowel is transcribed as [a] in Wong (1941) and Yuan, et al. (1960), [ɑ] in Jones and Woo (1912), Chao (1947), and Kao (1971), or [A] in Hashimoto (1972). [A] is not an IPA vowel symbol. It has been used widely by the Chinese linguists to refer to a low, central vowel positioned in between the Cardinal Vowels 4 [a] and 5 [ɑ]. Based on the positions of the vowel ellipses for the long, low vowel in the (C)V: syllables (Figure 1) and the half-long, low vowel in the (C)V:S syllables (Figure 2) in Cantonese, the vowel in question is more appropriately treated as a low, central vowel. It is assumed here that F_1 and F_2 are correlated with the tongue positions for the production of vowels, as "a tongue elevation ... provides an F_1-decrease but a fairly constant F_2, and a shift of the reference point in a direction parallel to this surface causes an F_2-increase at fairly constant F_1" (Fant, 1970, p. 112). In Lee (1985), a spectrographic analysis of the formant frequencies of the Cantonese vowels from six male speakers, the low vowel in question is considered a back [ɑ]. An examination of Lee's spectrographic data shows that treating the vowel as a low, back vowel is unwarranted. The mean F_2 values for the low vowel estimated from his spectrographic data for six speakers are approximately 1,125 Hz, 1,230 Hz, 1,350 Hz, 1,375 Hz, 1,200 Hz, and 1,320 Hz. These figures are similar to the mean F_2 values for the long (1,229 Hz) and half-long (1,270 Hz) low vowels obtained in this study. As shown in Table 2, in languages such as American English (Peterson and Barney, 1952), Swedish

	[a] in Cantonese (this study)		[ɑ] in Cantonese (Lee, 1985)	[ɑ] in American English (P&B, 1952)	[ɑ] in Swedish (Fant, 1970)	[ɑ] in Russian (F, 1970; R, 1928)
	CV:	CV:S				
F_1	827	896	705	730	616	680
F_2	1,229	1,270	1,230	1,090	1,070	1,070

Table 2. Mean values of F_1 and F_2 (in Hz) for the low vowel in Cantonese (this study; Lee, 1985) and [ɑ] in American English (Peterson and Barney (P&B), 1952), Swedish (Fant, 1970), and Russian (Fant (F), 1970; Russell (R), 1928).

(Fant, 1970), and Russian (Fant, 1970; Russell, 1928), the F_2 are all under 1,100 Hz. Figure 4 shows the positions of the Cantonese long vowels [i ɛ a ɔ u] in the

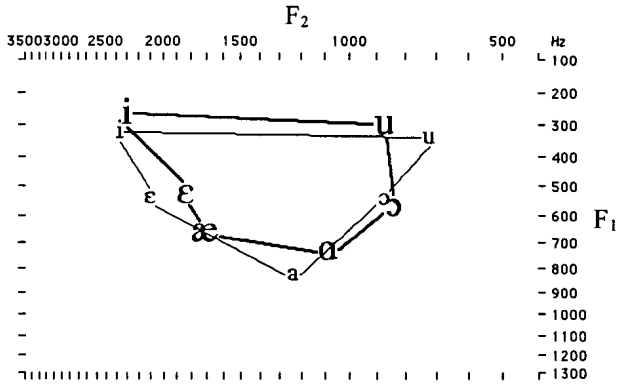

Figure 4. The pattern of the relative positions of the long vowels [i ɛ a ɔ u] in Cantonese in this study (smaller IPA symbols connected by the thin lines) superimposed onto the pattern of the relative positions of the vowels [i ɛ æ ɑ ɔ u] in American English in Peterson and Barney (1952) (larger IPA symbols connected by the thick lines).

(C)V: syllables in the F_1/F_2 plane, based on the mean F_1 and F_2 values for the five vowels in this study, superimposed onto the positions of the vowels [i ɛ æ ɑ ɔ u] in American English (Peterson and Barney, 1952), also based on the mean F_1 and F_2 values for the six vowels. As can be seen, the distance in the horizontal axis between the Cantonese low vowel [a] and back vowels [ɔ] and [u] in the acoustical vowel space is greater than that between the vowel [ɑ] and back vowels [ɔ] and [u] in American English. This and the larger absolute F_2 value for the Cantonese vowel in question support the contention that the long, low vowel in the (C)V: syllables in Cantonese should be treated as a low, central vowel [a], rather than a low, back [ɑ] as in American English. What has been said about the Cantonese long, low vowel [a] in the (C)V: syllables is also true for the half-long, low vowel [a] in the (C)V:S syllables in Cantonese, as the half-long vowel has an even larger mean F_2 value than its long counterpart in the (C)V: syllables.

Figure 5 shows the positions of the vowel ellipses for the Cantonese short vowels [ɪ ɐ ɵ ʊ] in the (C)VS syllables. To show the relative positions of the vowel ellipses for these short vowels in the (C)VS syllables and

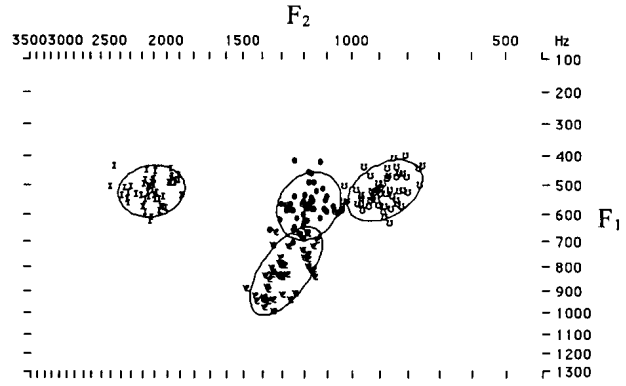

Figure 5. Vowel ellipses for the 4 Cantonese short vowels [ɪ ɐ ɵ ʊ] in the (C)VS syllables (10 male speakers).

the long vowels [i y ɛ a œ ɔ u] in the (C)V: syllables (Figure 1), the vowel ellipses for the two sets of vowels are superimposed in Figure 6. As can be seen, the vowel ellipses for the short vowel [ɪ] and long vowel [ɛ] overlap

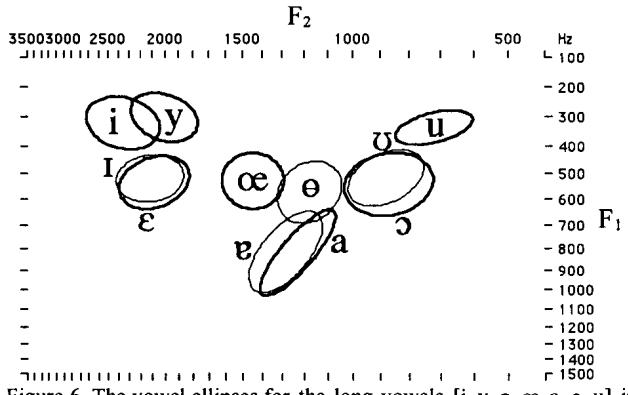

Figure 6. The vowel ellipses for the long vowels [i y ɛ œ a ɔ u] in the (C)V: syllables (represented with thick lines) superimposed on the vowel ellipses for the short vowels [ɪ ɵ ɐ ʊ] in the (C)VS syllables (represented with thin lines).

extensively. This is also true for the short [ʊ] and long [ɔ]. As for the vowel ellipse for the short vowel [ɐ], it also overlaps with the vowel ellipse for the long vowel [a], although the degree of overlap is less than the other two cases. And, as shown in Table 3, the differences in the mean F_3 value between [ɪ] and [ɛ], [ʊ] and [ɔ], and [ɐ] and [a] are also small. Despite the similarities in F-values, [ɪ] and [ɛ], [ʊ] and [ɔ], or [ɐ] and [a] have been transcribed differently in most of the past studies (Jones and Woo, 1912; Wong, 1941; Chao, 1947; Yuan, et al., 1960; Kao, 1971; Hashimoto, 1972). Ladefoged and Maddieson (1996) point 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" (Ladefoged and Maddieson, 1996, p. 285). Thus, the discrepancies between the center F-values and vowel transcription of the Cantonese vowels in question is not an isolated case. But, how should the discrepancies arise? As the transcription of the Cantonese short vowels, such as [ɪ ʊ ɐ ɵ] in the past studies is assumed to have been 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 property of the entire vowel. Due to short duration (approx. 120 ms) and to the fact that in most test words used in this study the short vowels are preceded by a consonant and always followed by a stop consonant, the steady-state portions of the formant trajectories are extremely short (20-40 ms). The short vowels are realized more like short diphthongs or triphthongs than monophthongs. Tape-cutting experiments of Schatz (1954) and Harris (1953) 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, 1976, p. 213). It follows that the perceptual impression of the short vowels in Cantonese should not be expected to correspond to their center F-values. The short [ɪ] and the long [ɛ] in Cantonese are assumed to be perceptually distinct, despite similarity in center F-values. Similar statements can be made about the short [ʊ] in relation to the long [ɔ] as well as the short [ɐ] in relation to the long [a] in Cantonese.

	F_1		F_2		F_3	
Vowel	Mean	s.d.	Mean	s.d.	Mean	s.d.
[ɪ]	520	44.3	2,127	140.5	2,710	148.9
[ɛ]	537	49.9	2,088	142.7	2,804	165.2
[ʊ]	518	52.6	882	66.9	2,922	115.5
[ɔ]	544	59.4	871	79.1	2,894	117.5
[ɐ]	820	90.7	1,287	86.0	2,846	120.1
[a]	827	97.0	1,229	86.6	2,852	138.0

Table 3. Mean values of F_1, F_2, and F_3 (in Hz) and their standard deviations for the short vowels [ɪ], [ʊ], and [ɐ] in the (C)VS syllables and the long vowels [ɛ], [ɔ], and [a] in the (C)V: syllables in Cantonese for 10 male speakers.

4. CONCLUDING REMARKS

The paper has presented the acoustic characteristics of the vowels in the (C)V: ([i y ɛ œ a ɔ u]), (C)V:S ([i y ɛ œ a ɔ u]), and (C)VS ([ɪ ɵ ɐ ʊ]) syllables in Cantonese. The center formant frequency values for the vowels provide useful information for characterizing the vowels with a long steady-state in the (C)V: and (C)V:S syllables (approx. 150 ms or more). However, their usefulness diminishes when the vowels have a very short (approx. 20-40 ms) or no steady-state in the (C)VS syllables. The short vowels in the (C)VS syllables should be treated as diphthongs or triphthongs and analyzed for their formant movements in order to have a complete picture of the their spectral characteristics.

5. REFERENCES

Chao, Y.R. *Cantonese Primer*. New York: Greenwood Press, 1947.

Disner, S.F. "Vowel quality: the relation between universal and language-specific factors". *UCLA Working Papers*, 58, 1983.

Fant, G. *Acoustic Theory of Speech Production*. The Hague: Mouton, 1970.

Harris, C.M. "A study of the building blocks in speech". *Journal of the Acoustical Society of America*, 25.962-969, 1953.

Hashimoto, A. *Cantonese Phonology*. Cambridge: Cambridge University Press, 1972.

Jones, D. and K.T. Woo. *A Cantonese Phonetic Reader*. London, 1912.

Kao, D. *Structure of the Syllable in Cantonese*. The Hague: Mouton, 1971.

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

李行德 (Lee, H.-T.T.). 廣州話元音的音值及長短對立. 方言 (*Fangyan*), 1.28-38, 1985.

Lindblom, B.E. "Spectrographic study of vowel reduction". *Journal of the Acoustical Society of America*, 35.11.1773-1781, 1963.

Peterson, G.E. and H.L. Barney. "Control methods used in a study of the vowels". *Journal of the Acoustical Society of America*, 24.2.175-184, 1952.

Russell, G.O. *The Vowel*. Columbus: OSU Press, 1928.

Schatz, C. 1954. "The role of context in the perception of stop". *Language*, 30.47-56.

Strange, W. and R.R. Verbrugge. "Consonant environment specifies vowel identity". *Journal of the Acoustical Society of America*, 60.1.213-224, 1976.

袁家驊等 (Yuan, et al.). 漢語方言概要. Beijing: Wenzi Gaige Chubanshe, 1960.

Wong, S.L. *A Chinese Syllabary Pronounced According to the Dialect of Cantonese*. Canton: Lingnam University, 1941.

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