AN ACOUSTICAL ANALYSIS OF THE DIPHTHONGS IN CANTONESE

Eric Zee
Phonetics Lab, Department of Chinese, Translation, and Linguistics, City University of Hong Kong

ABSTRACT

The study is an acoustical analysis of the diphthongs in Hong Kong Cantonese spoken by young adult undergraduates, male and female. The onset and offset formant measurements were made as well as the durations of the onset and offset steady states and the transition between them. Results show that (1) formant frequency data provide a basis for diphthong transcription, (2) the diphthongs may be separated into two categories according to temporal organisation, and (3) the relative formant values for some of the diphthongs and vowels differ in male and female speech.

1. INTRODUCTION

It is generally agreed that in Cantonese there are 10 diphthongs and 7 long vowels in open syllables [1, 2, 3]. However, the diphthongs and vowels are transcribed differently in the earlier studies of the language as shown in Table 1. An additional diphthong [E…U] which occurs only in a few monosyllabic words has been identified recently in the speech of Cantonese speakers in Hong Kong. As no arti-culatory or acoustical data of the diphthongs and vowels are

	Diphthongs	Long Vowels
Chao [1]	[A…i, A…u, Ai, Au, i…u, u…i, ei, ou, O…i, •y]	[i, y, E, A, u, O, {]
Kao [3]	[ai, au, åi, åu, iu, ui, ei, ou, Oi, {y]	[i, y, E, A, u, O, {]
Hashimoto [2]	[Ai, Au, ai, au, iu, ui, ei, ou, Oi, Py]	[i, y, E, A, u, O, {]

Table 1. Cantonese diphthongs and long vowels in open syllables [1, 2, 3]; the symbol "[A]" represents the vowel between Cardinal Vowels [a] and [A] in [2].

presented in any of these studies, it is believed that the discrepancies in the transcriptions are mainly due to the difference in the individual author's auditory impression. The present study analyses the acoustical properties of the diphthongs in the speech of the young adult speakers of Hong Kong Cantonese. The onset and offset formant frequency measurements are made as well as the durations of the onset and offset steady states and the transition between them.

2. PROCEDURE

2.1. Subjects

Speakers who provided the speech data consisted of 10 male and 10 female university undergraduate students, all native speakers of Hong Kong Cantonese between 18 and 21 years of age.

2.2. Test material and recording

Audio recordings were made of subjects reading a randomised list of 18 meaningful test monosyllables which consisted of 11 diphthongs and 7 long vowels in Cantonese. The F-values for the long vowels were also analysed. It is assumed that a meaningful interpretation of the F-values for the elements of the diphthongs is possible only if the values are considered taking into account of their relation to the F-values for the vowels in open syllables in the language. The suggested IPA symbols for the diphthongs and vowels in this study are [a…I, a…U\a…o, ai, au\aU, i…U, u…Y, Ei, Ou, E…U, O…Y, Äy] and [i, y, e\E, •\Ä, a, o\O, u]. Table 2 shows the test monosyllables contain-ing 11 diphthongs and 7 long vowels. Each test word was embedded in a carrier sentence, [No\NOï ji…Uî

Diphthongs	Long Vowels
[a…I] in [ha…Iâ] "to touch"	[i] in [jiâ] "medicine"
[a…U]\[a…o] in [ha…Uâ]\[ha…oâ] "to roast"	[y] in [jyâ] "to grant"
[ai] in [aiâ] "to plead"	[e]\[E] in [eâ]\[Eâ] "to complain"
[au]\[aU] in [hauâ]\[haUâ] "to watch"	[•]\[Ä] in [h•â]\[hÄâ] "boot"
[i…U] in [hi…Uâ] "arrogant"	[a] in [haâ] "shrimp"
[u…Y] in [wu…Yâ] "to bake"	[o]\[O] in [hoâ]\[hOâ] "harsh"
[Ei] in [hEiâ] "diluted"	[u] in [wuâ] "black"
[Ou] in [hOuâ] "a kind of vegetable"	
[E…U] in [tE…Uô] "to throw away"	
[O…Y] in [hO…Yâ] "to open"	
[Äy] in [hÄyâ] "marketplace"	

Table 2. The test monosyllables containing 11 diphthongs [a…I, a…U\a…o, ai, au\aU, i…U, u…Y, Ei, Ou, E…U, O…Y, Äy] and 7 long vowels [i, y, e\E, •\Ä, a, o\O, u]; the IPA symbols after the slashes representing sounds in female speech.

tok\tOkô _____ pEiü lEiï tÓeN\tÓENâ] "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 1,800 (20 speakers x 18 test monosyllables x 5 repetitions). The recordings were performed in a sound-proof booth (IAC), using a Nakamichi DR-1 Cassette recorder and a Shure dynamic microphone.

2.3. Analysis

Kay CSL4300B was used for spectral analysis of the diph-thong and vowel formants. In this study, speech data were captured at a sampling rate of 10,000 samples per second, producing an upper frequency cut-off of 5,000 Hz. LPC was performed, using the pitch synchronous method, at the midpoints of the vowels or the steady states of the first and second elements of the diphthongs. Formant trajectories of the diphthongs were obtained for determining the steady states of the onset and offset. Measurements of the durations of the first element, transition, and second element of the diphthongs were made with the aid of (1) the formant trajectories of the diphthongs, (2) auditory judgement of the investigator, and (3) visual inspection of the waveforms.

3. RESULT 1: DIPHTHONG FORMANTS

Figures 1a-1e show the diphthong formants of the 11 Cantonese diphthongs [a…I, a…U, ai, au, i…U, u…Y, Ei, Ou, E…U, O…Y, Äy] in the F_1-F_2 plane for the 10 male speakers. The diph-

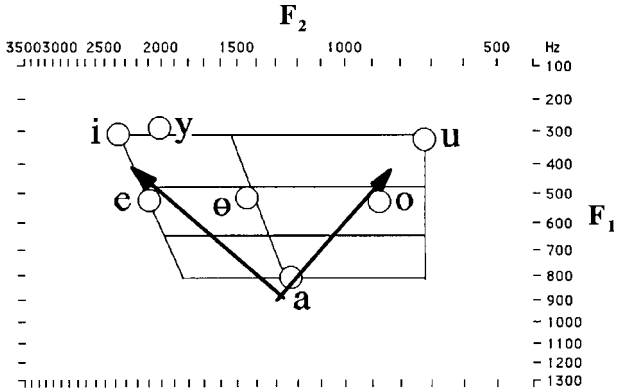

Figure 1a. Formants (in Hz) of the Cantonese diphthongs [a…I] and [a…U] for 10 male speakers.

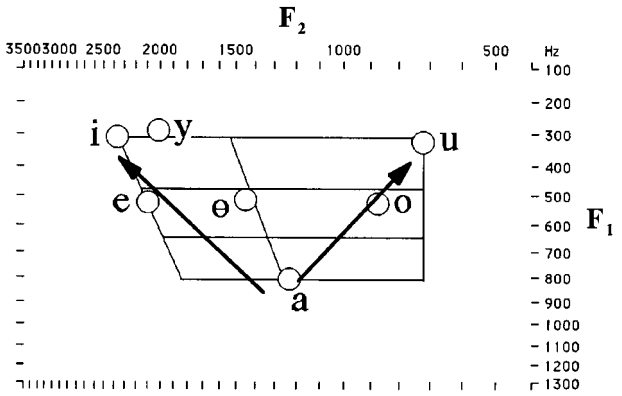

Figure 1b. Formants (in Hz) of the Cantonese diphthongs [ai] and [au] for 10 male speakers.

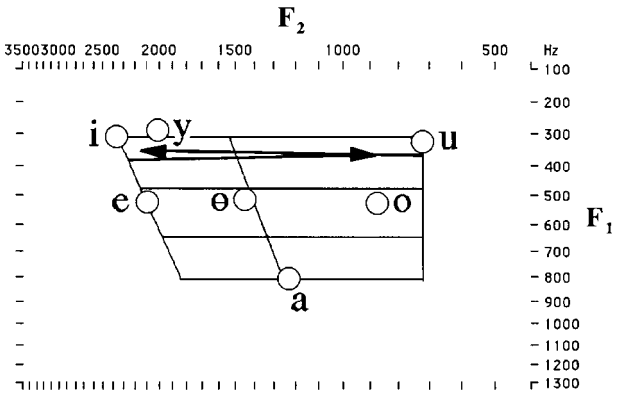

Figure 1c. Formants (in Hz) of the Cantonese diphthongs [i…U] and [u…Y] for 10 male speakers.

thong formants are plotted against the formants of the 7 long vowels produced by the speakers. Each of the two end points of the arrow is determined by the mean of 50 data points (10 speakers x 5 repetitions). Each of the data points is in turn determined by the F_1 and F_2 values for the first or second element of diphthongs, with F_1 plotted on the ordinate and F_2 on the abscissa on Bark scales. The circles in Figures 1a-1e represent the positions for the 7 Cantonese long vowels [i, y, e, a, •, o, u] in the F_1-F_2 plane for the male speakers. Similar to the elements of the diphthongs, the position of each

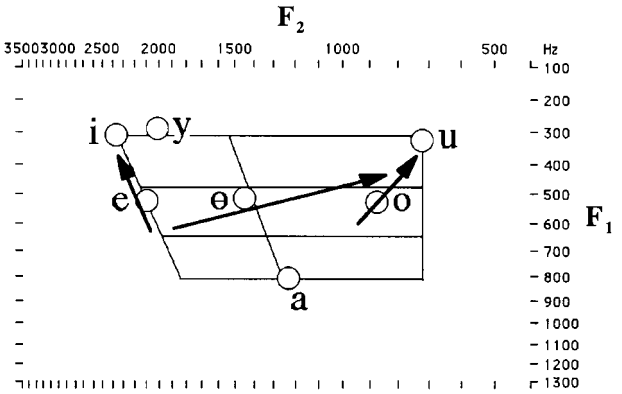

Figure 1d. Formants (in Hz) of the Cantonese diphthongs [Ei], [Ou], and [E…U] for 10 male speakers.

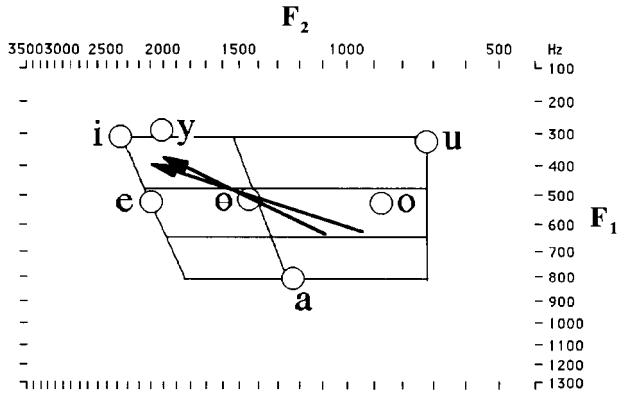

Figure 1e. Formants (in Hz) of the Cantonese diphthongs [O…Y] and [Äy] for 10 male speakers.

of the circles is determined by the two means of the 50 F_1 and 50 F_2 values for a particular vowel. The acoustical quadrilaterals in Figures 1a-1e are drawn for reference purpose. These trapezia with parallel top and bottom and right angles at top and bottom back are for the vowels and diphthongs produced by the male speakers. The positions of the four sides and the angle for the frontal side are determined by the F_1 or F_2 values for the vowels [i, e, a, u]. The slanting vertical centre line is drawn with two mid points of the parallel top and bottom. The upper mid point is determined by the sum of the F_2 value for [u] and one half of the difference between the F_2 values for the vowels [i, u]. The lower mid point is one half of the difference between the F_2 value for [u] and the F_2 value for the point where the frontal side and the bottom cross, plus the F_2 value for [u]. The frontal side is drawn with the F_2 values for the vowels [i] and [e]. The frequency differences between any two closest parallel horizontal lines in the acoustical quadrilaterals are equal.

Figures 1a and 1b show the diphthong arrows for [a…I, a…U] and [ai, au], respectively. Despite the similarity in formant movement between the two sets of the diphthongs, the second elements of [ai] and [au] are closer than those of [a…I] and [a…U] to the positions of the circles for the vowels [i] and [u], which justifies transcribing the second elements differently. In addition, the first element of [a…U] has larger mean F_1 and F_2 values than the first element of [au], and the first element of [a…I] has a larger mean F_1 but a smaller mean F_2 value than the first element of [ai], although the differences are only small. Furthermore, as will be demonstrated, the durations of the first elements of [a…I] and [a…U] are longer than [ai] and [au], whereas the durations of the second elements of [a…I] and [a…U] are shorter than [ai] and [au]. The differences in temporal structure are reflected in the transcriptions. Figure 1c shows the diphthong arrows for [i…U] and [u…Y]. While the second element of [i…U] is not close to the monophthong [u] enough to be transcribed as [u], the position of the second element of [u…Y] seems to match with [y]. However, due to its high mean F_3 value (2,724 Hz) relative to the mean F_3 value (2,412 Hz) for the monophthong [y], the second element is transcribed as [Y]. The temporal structures of [i…U] and [u…Y] are similar to those of [a…I] and [a…U]. The diphthong formants in Figure 1d are in general compatible with the symbols [E…U], [Ei], and [Ou]. [E…U] shares the temporal characteristics with [a…I, a…U, i…U, u…Y]. As for [Ei] and [Ou], their internal temporal structures are similar to those of [ai] and [au]. Figure 1e shows the diphthong arrows for [O…Y] and [Äy]. As can be seen, the arrows start and end in the same area, in addition to the fact that they cross each other. Despite close proximity, the first elements of the two diphthongs are transcribed differently as [O] and [Ä], as one of the elements is closer to the back vowel, whereas the other closer to the central vowel. However, as shown in the figure the two first elements are not quite compatible with the vowels used to describe them. That the second elements of these diphthongs are also transcribed differently is due to a higher mean F_3 value (2,663 Hz) for the second element of [O…Y] than the mean F_3 value (2,466 Hz) for the second element of [Äy]. It should be added that the mean F_3 value for the monophthong [i] in Cantonese for the male speakers is 3,213 Hz.

Figures 2a-2e show the diphthong formants of the 11 Cantonese diphthongs in the F_1-F_2 plane for the 10 female speakers. The acoustical quadrilaterals for the female speakers are drawn in the same way as those for the male speakers. A comparison of Figures 1a-1e with Figures 2a-2e shows that the acoustical quadrilaterals for the female speakers are larger than those for the male speakers. This indicates that the ranges of F_1 and F_2 of the vowels and diphthongs are wider for the female speakers than the male speakers. There is a difference in vowel production between the male and female speakers. As shown in the figures, the positions of the circles representing the mid vowels for the female speakers are lower relative to the positions of other vowels compared to the male speakers (Figures 1a-1e). For the female speakers, the distances between the circles representing the high vowels and mid vowels are greater than the distances between the circles representing the mid vowels and low vowels. It is not clear as to why the difference in the production of the mid vowels should occur between the male and female speakers. Two sets

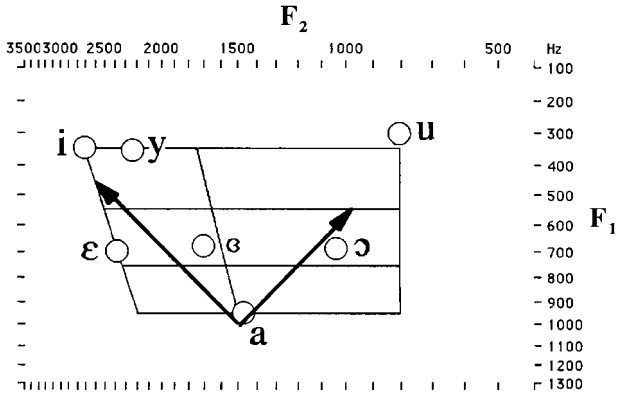

Figure 2a. Formants (in Hz) of the Cantonese diphthongs

[a…I] and [a…o] for 10 female speakers.

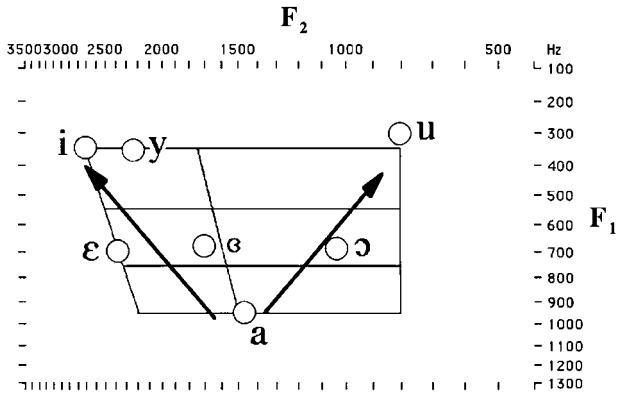

Figure 2b. Formants (in Hz) of the Cantonese diphthongs [ai] and [aU] for 10 female speakers.

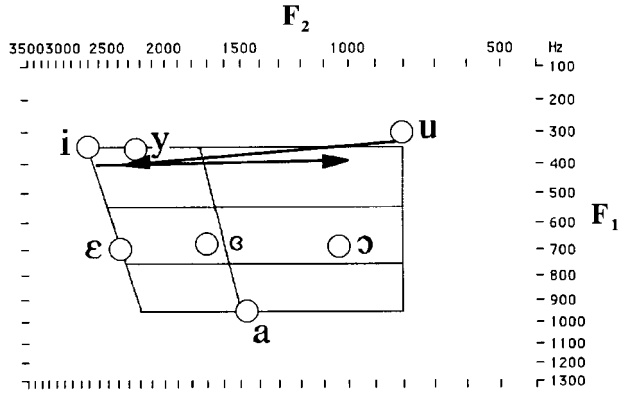

Figure 2c. Formants (in Hz) of the Cantonese diphthongs [i…U] and [u…Y] for 10 female speakers.

of symbols are thus used to transcribe the mid vowels in this study, i.e., close mid [e, •, o] for the male speakers and open mid [E, Ä, O] for the female speakers. As for the diphthong formants and temporal structures, the patterns for the female speakers are similar to the patterns for the male speakers. There are however two cases where the relative formant values for the second elements differ between male and female speech. The second element of the diphthong formant on the right in Figure 2a does not reach the area for [U] in the quadrilateral for

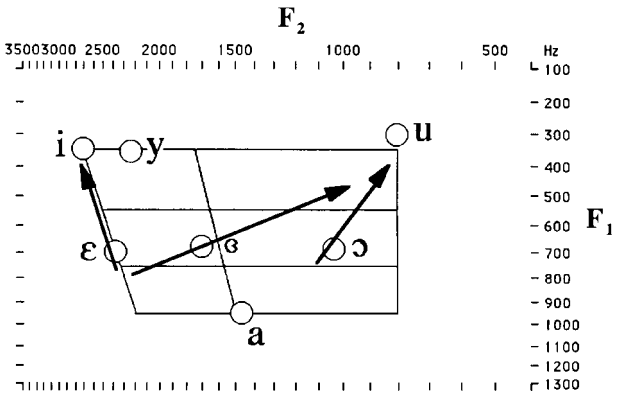

Figure 2d. Formants (in Hz) of the Cantonese diphthongs [Ei], [Ou], and [E…U] for 10 female speakers.

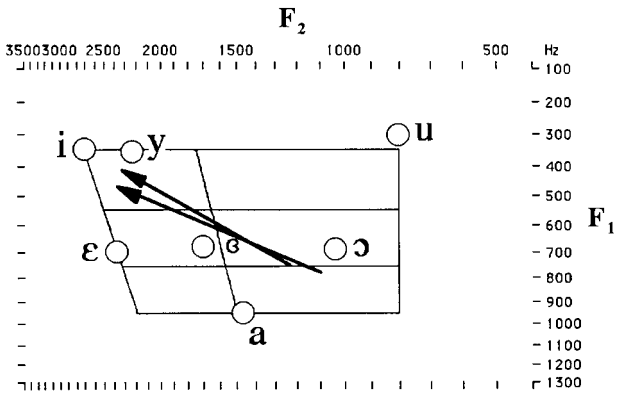

Figure 2e. Formants (in Hz) of the Cantonese diphthongs [O…Y] and [Äy] for 10 female speakers.

the female speakers, as it does for the male speakers (Figure 1a). Thus, the diphthong is transcribed as [a…o] for the female speakers, as opposed to [a…U] for the male speakers. By the same token, [au] for the male speakers (Figure 1b) is trans-cribed as [aU] for the female speakers (Figure 2b).

4. RESULT 2: TEMPORAL ORGANISATION

The 11 Cantonese diphthongs may be separated into two categories according to temporal structure. The first category consists of [i…U, u…Y, a…I, a…U\a…o, O…Y, E…U], in which the duration of the first element is longer than the second. The second category consists of [Ei, Ou, ai, au\aU, Äy], in which the duration of the first element is shorter than the second. For both categories, the durations of the transitions are similar. As for the total duration, the first category [i…U, u…Y, a…I, a…U\a…o, O…Y, E…U] is longer than the second category [Ei, Ou, ai, au\aU, Äy]. Two typical cases showing the temporal organisation of the Cantonese diphthongs produced by a male speaker and a female speaker are shown in Figures 3a and 3b. As can be seen, the patterns of the temporal organisation are similar for the male and female speakers.

5. CONCLUDING REMARKS

The three major works on Cantonese [1, 2, 3] differ in the transcription of a number of Cantonese diphthongs (Table 1). In fact, some of the transcriptions are considered questionable. The inconsistency is assumed to have been caused mainly by

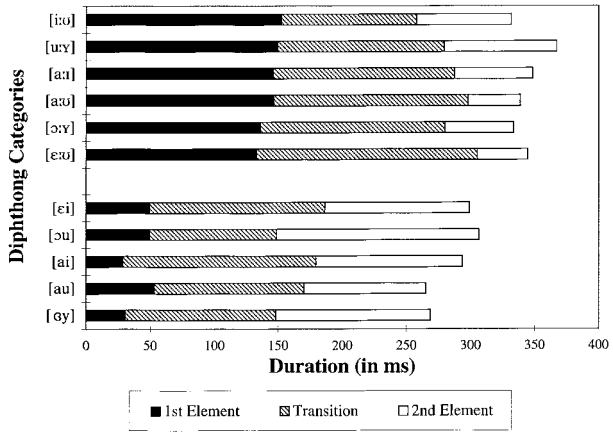

Figure 3a. The temporal organisation of the 11 Cantonese diphthongs [i…U, u…Y, a…I, a…U, O…Y, E…U, Ei, Ou, ai, au, Äy] (male speaker).

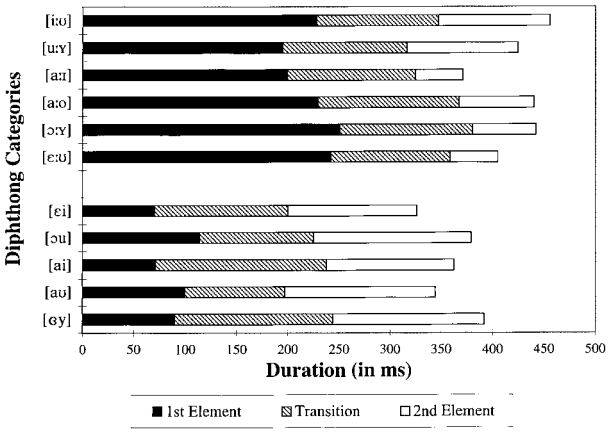

Figure 3b. The temporal organisation of the 11 Cantonese diphthongs [i…U, u…Y, a…I, a…o, O…Y, E…U, Ei, Ou, ai, aU, Äy] (female speaker).

the authors' differing impression of the speech data from perhaps speakers of different backgrounds. The acoustical data presented in this study provide useful and reliable information for reasonably accurate description and transcription of the Cantonese diphthongs, although a fuller understanding of the phonetic properties of the diphthongs under discussion can be gained with additional perceptual data. The results of the analysis also show that in Cantonese (1) the relative formant frequency values for some of the diphthongs differ in male and female speech, and (2) two categories of diphthongs may be distinguished in terms of temporal organisation.

ACKNOWLEDGEMENT

This research is part of a larger project supported by a Strategic Research Grant from the City University of Hong Kong.

REFERENCES

[1] Chao, Y. R. 1947. *Cantonese Primer*. New York: Greenwood Press.
[2] Hashimoto, O.-K. Y. 1972. *Phonology of Cantonese*. Cambridge: Cambridge University Press.
[3] Kao, D. L. 1971. *Structure of the Syllable in Cantonese*. The Hague: Mouton.