THE EFFECT OF DURATION SHORTENING ON THE ACOUSTIC CHARACTERISTICS OF THE DIPHTHONGS IN HANGZHOU CHINESE

LI Jian

City University of Hong Kong
50008823@cityu.edu.hk

ABSTRACT

This study investigates how the shortening of duration caused by fast speech rate and different syllable structure affect the acoustic characteristics of the diphthongs in Hangzhou Chinese (HC). Diphthongs in the open and closed syllables produced in normal and fast speech are investigated. The results indicate that the two factors have different effect on the temporal structure and spectral targets of the diphthongs. While the diphthongs in closed syllables tend to undershoot their targets as compared with those in the open syllables, speech rate does not change the diphthongs targets in any significant way.

Keywords: diphthong, rate, acoustics, Chinese

1. INTRODUCTION

Similar to most Chinese dialects, the vowels and diphthongs in HC do not contrast in duration. But the vowels and diphthongs (D) in closed syllables ((C)DS) are phonetically shorter than those in open syllables ((C)D). Speech rate is another factor affecting vowel duration. An increase of speech rate reduces the duration of vowels. Since both (C)DS syllable typs and fast speech rate reduce the duration of the diphthongs in HC, we are interested in the question whether they both affect the temporal organization and spectral characteristics of the diphthongs, and if yes, whether they change the temporal and spectral characteristics of the diphthongs in the same way.

According to the target-undershoot model ([3]), shortening of vowel duration resulting from fast speech rate and lack of stress causes vowel reduction (defined as textual assimilation in [3], or centralization in [5]), owing to the "limitations inherent in the articulatory mechanism" [3]. While this model is widely quoted, the relationship between duration shortening and vowel reduction seems not to be so straightforward. [2] and [1], for instance, found no major effect of speaking rate or stress on vowel quality. In [6] and [7], the speakers achieved the same formant targets at both normal and fast rates.

Therefore, the present study also aims to re-examine the relationship between duration shortening and vowel reduction. If there is a direct correlation between duration shortening and vowel reduction, both speech rate and syllable type should have similar effect on the diphthongs, since both of them changes the duration of the diphthongs. On the contrary, if no such effect is found, the relationship between vowel reduction and duration shortening need to be explained differently.

2. METHOD

Six diphthongs ([ia iɛ iɔ ua uo ɥa]) occurring in both (C)D and (C)DS syllables are investigated. Meaningful monosyllabic words containing anyone of the diphthongs were selected as the test words (Table 1). Words associated with a mid-level tone for the (C)D syllables, or a high short tone for (C)DS syllables (The closed and open syllables are always associated with different tones in HC) without an initial consonant were preferred.

Table 1: Test words

D	(C)D syllable		(C)DS syllable	
	Test word	Meaning	Test word	Meaning
ia	ia ^{44}	crow	iɑʔ $^{\underline{54}}$	duck
iɛ	tɕiɛ ^{44}	street	iɛʔ $^{\underline{54}}$	one
iɔ	iɔ ^{44}	monster	iɔʔ $^{\underline{23}}$	jade
ua	ua ^{44}	frog	uaʔ $^{\underline{54}}$	dig
uo	uo ^{24}	say	uoʔ $^{\underline{23}}$	live
ɥa	tsɥa ^{44}	catch	sɥaʔ $^{\underline{54}}$	brush

2.1. Speakers and recording

The speech data were provided by 4 (two male and two female) native HC speakers, all born and grew up in the city of Hangzhou, with no reported speech or hearing disabilities. Digital audio recording was made in a quiet room, using a Sony PCM-R700 digital audio recorder and a Shure SM-58 microphone. Each test word was imbedded in a carrier sentence: [ŋo^{51} sɥəʔ^{5}__ pəʔ^{5} n̩i^{51} t^{h}in^{33}] (I

read __ for you (to) listen), so that the test diphthongs are always preceded and followed by a stop which will help segmentation. 4 repetitions of all the test materials were randomized in a list. The speakers were instructed to read the list in a clear and natural manner, first at a normal speech rate, then at a fast speech rate.

2.2. Data analysis

Speech data were digitalized at a sampling rate of 11,250 samples per second and analyzed using Praat 4.3.19. The formant trajectories for each diphthong were analyzed using linear-prediction-based formant track overlaid on the wide-band spectrogram in Praat. The durations of the 1st and 2nd elements of the diphthongs and the transitions were measured from the waveforms and the formant trajectories, with reference to the wide-band spectrograms. The total duration of the diphthongs were then calculated. The frequencies of the first two formants of the target elements were measured at the midpoint of the steady state portions of the diphthong elements. There are cases where no steady-state can be identified for the diphthong elements. For these cases, the durations of the elements are defined as zero, and the formants are measured from the beginning (for the 1st elements) or the end (for the 2nd elements) of the diphthongs.

3. RESULTS

3.1. Temporal measurements

The mean durations of the 1st elements (V1), the transitions, the 2nd elements (V2) and the whole diphthongs in (C)D and (C)DS syllables produced in normal (N) and fast (F) speech are shown in table 2 and plotted in figure 1. Comparisons are made between (C)D and (C)DS syllables and between normal and fast speech rate respectively.

3.3.1. (C)D vs. (C)DS

The overall duration of the diphthongs are significantly shorter in (C)DS as compared with (C)D syllable. The duration of the diphthongs in (C)DS syllables are 56% shorter in normal speech and 47% shorter in fast speech as compared with those in (C)D syllables (see table 3).

The diphthongs in the two syllable types are different in their internal temporal structure (see table 2 and figure 1). For the diphthongs in (C)D syllables, the 2nd elements have a longer steady state than the 1st elements (except for [iɛ] which has no steady state on the 2nd element); while for those in (C)DS syllables, the 2nd elements have no or a very short steady state, and the 1st element have a relatively long steady state (except for [ɳa] with no steady state on the 1st element). It is the 2nd elements that sacrifice most for the duration shortening.

Table 2: The mean durations (in ms) of the 1st elements (V1), transitions, 2nd elements (V2) and the whole diphthongs in (C)D and (C)DS syllables produced in normal (n) and fast (f) speech (standard deviations in parenthesis).

D	condition		V1	transition	V2	Total
ia	(C)D	N	55(25)	162(28)	186(32)	402(49)
		F	31(12)	135(32)	91(24)	256(29)
	(C)DS	N	69(22)	110(23)	10(10)	190(34)
		F	48(15)	87(14)	5(7)	141(12)
iɛ	(C)D	N	163(77)	220(37)	2(8)	385(53)
		F	91(45)	130(40)	0	221(20)
	(C)DS	N	75(22)	94(16)	7(11)	176(30)
		F	60(15)	77(12)	2(6)	138(14)
iɔ	(C)D	N	56(22)	153(27)	242(40)	451(67)
		F	33(15)	126(23)	114(25)	273(25)
	(C)DS	N	55(12)	117(22)	21(20)	172(45)
		F	32(9)	110(15)	15(18)	137(21)
ua	(C)D	N	42(24)	171(36)	196(58)	409(67)
		F	25(7)	125(33)	98(26)	248(38)
	(C)DS	N	50(16)	118(14)	15(9)	182(14)
		F	26(17)	104(18)	7(10)	137(26)
uo	(C)D	N	64(38)	170(43)	103(29)	337(49)
		F	48(15)	120(25)	77(21)	245(26)
	(C)DS	N	56(14)	139(23)	19(6)	214(21)
		F	34(12)	114(10)	6(8)	154(16)
ɳa	(C)D	N	0	199(52)	203(36)	402(61)
		F	0	139(32)	96(24)	235(19)
	(C)DS	N	3(6)	103(21)	11(10)	117(17)
		F	0	74(12)	5(7)	79(11)

Figure 1: The temporal structures (in ms) of the diphthongs in (C)D (left) and (C)DS (right) syllables (upper section) and the corresponding percentage-wise data (lower section). For each diphthong, the upper bar shows the normal speech data, the lower fast speech data.

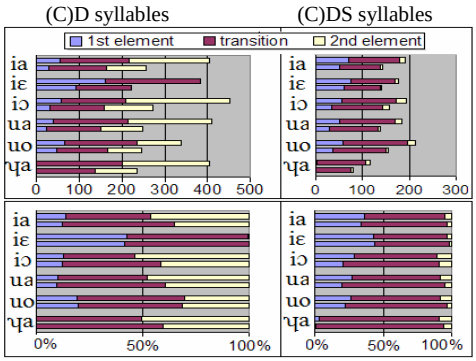


Table 3: Percentage of shortening of the diphthongs in (C)DS syllables as compared with (C)D syllables.

	ia	iɛ	iɔ	ua	uo	ɳa	Mean
normal	53%	54%	62%	56%	36%	71%	56%
fast	45%	38%	50%	45%	37%	66%	47%

Table 4: Percentage of shortening of the diphthongs in fast speech as compared with normal speech.

	ia	iɛ	iɔ	ua	uo	aɥ	M
CD	36%	38%	39%	39%	27%	41%	37%
CDS	26%	21%	21%	25%	28%	32%	25%

3.3.2. Normal vs. fast speech

The overall duration of the diphthongs are also significantly shortened in fast speech, which means that the speakers did increase their speech rate as instructed. Table 4 shows that the diphthongs in the (C)D and (C)DS syllables are shortened by a mean percentage of 37% and 25%, respectively.

As indicated in table 2 and figure 1, the overall reduction of the diphthong durations in fast speech is reflected by the shortening of each of the three segments, though the transition is shown to be more resistant to reduction, which is different from the effect of syllable type. The general temporal structures of the diphthongs are not changed by increasing speech rate. The temporal structural differences between the diphthongs in the two syllable types are not neutralized when the speech rate increases.

3.2. Formant measurements

Averaged formant values for each diphthong targets in different conditions are summarized in table 5. Two-way ANOVA was conducted using SPSS on the first two formants of the diphthong elements produced by the male and female speakers, to investigate whether the two factors (syllable type and speech rate) have significant effect on the formant values of the diphthong elements. Since the male and female speakers show similar pattern, only the male data are reported here to save space.

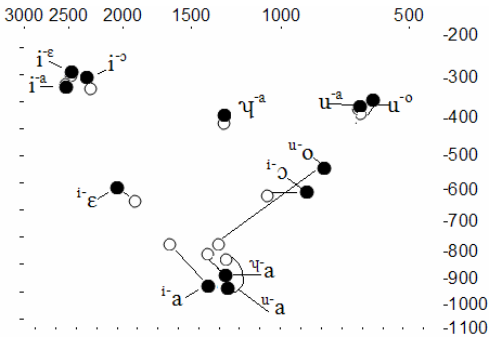
3.2.1 (C)D vs. (C)DS

Results show that syllable type has little effect on the 1st elements of the diphthongs except for F1 of [ɥa] (p<0.01) and [uo] (p<0.05), and F1 and F2 of [iɔ] (p<0.05), but significant effect on most of the 2nd elements (p<0.01 for F1 of [ia, ua, uo, ɥa], F2 of [ia, iɔ uo]; p<0.05 for F2 of [ɥa]). There is no significant interaction between the two factors except for F1 of [uo] (p<0.05). The first two formant frequencies of the diphthong targets in (C)D and (C)DS syllables are plotted in the F1/F2 plane to show their relative positions (see figure 2).

Table 5: Averaged formant frequencies (in Hz) for diphthong elements in (C)D and (C)DS syllables produced in normal and fast speech (data from the male speakers)

		v1F1	v1F2	v2F1	v2F2
ia					
CD	N	315 (31)	2518 (234)	920 (35)	1366 (88)
	F	306 (19)	2524 (124)	909 (42)	1390 (46)
CDS	N	296 (43)	2477 (219)	771 (45)	1656 (125)
	F	276 (19)	2503 (191)	730 (38)	1672 (143)
iɛ					
CD	N	297 (38)	2422 (194)	593 (96)	2039 (316)
	F	311 (44)	2425 (240)	553 (99)	2011 (272)
CDS	N	308 (32)	2499 (225)	629 (30)	1880 (42)
	F	282 (30)	2537 (184)	580 (35)	1930 (157)
iɔ					
CD	N	296 (41)	2323 (105)	610 (58)	882 (127)
	F	300 (20)	2320 (89)	641 (32)	942 (123)
CDS	N	329 (47)	2186 (163)	638 (44)	1086 (164)
	F	334 (25)	2063 (365)	632 (27)	1159 (220)
ua					
CD	N	384 (61)	677 (46)	900 (30)	1254 (62)
	F	396 (23)	614 (89)	894 (21)	1270 (52)
CDS	N	376 (57)	664 (61)	818 (23)	1282 (20)
	F	378 (48)	623 (85)	793 (30)	1272 (35)
uo					
CD	N	361 (52)	609 (41)	529 (21)	811 (61)
	F	380 (23)	609 (54)	535 (24)	853 (39)
CDS	N	417 (30)	693 (64)	832 (94)	1348 (120)
	F	386 (20)	684 (90)	784 (90)	1316 (61)
ɥa					
CD	N	373 (70)	1292 (88)	900 (53)	1296 (84)
	F	400 (47)	1259 (124)	903 (30)	1299 (57)
CDS	N	438 (37)	1292 (83)	852 (46)	1355 (65)
	F	469 (47)	1246 (95)	831 (60)	1347 (26)

Figure 2: Diphthong targets in (C)D (solid circles) and (C)DS (empty circles) syllables. Based on the male data.



As can be seen, the 1st elements [i, ɥ, u] do not show much difference between the two syllables types, despite the cases of significant differences reported in the statistic analysis. As to the 2nd elements in (C)DS as compared to (C)D syllables, [a] in [ia, ɥa, ua] are higher, [ɔ] in [iɔ] more central, [o] in [uo] much lower in the acoustic vowel space. The 2nd elements in [ia, ɥa, ua, iɔ] seem to be assimilated to the 1st elements and undershoot their targets, i.e. their positions get

closer to the 1st elements in (C)DS syllables, which is compatible with the vowel undershoot model. However, it is hard to explain why the 2nd element of [uo] gets lower in (C)DS syllable. As noticed from the formant data here and the duration data in previous section, the diphthong [uo] and [ua] in (C)DS syllable are very similar in terms of formant and temporal values. It might be possible that these two diphthongs are in a process of merging.

3.2.2 Normal vs. fast speech

The positions of the diphthongs elements produced in normal and fast speech overlap extensively and no effect of speech rate can be observed visually from the F1/F2 plane (not shown in the paper). Two-way ANOVA shows no significant effect of speech rate on the first two formant values of the 1st and 2nd elements of the diphthongs, except for F2 of the 1st element of [ua] (p<0.05).

3.3. Formant rate-of-change (RoC)

RoC is calculated to characterize the dynamic spectral changes of the diphthongs:

$$RoC = (F_{n\text{-}1st\ element} - F_{n\text{-}2nd\ element})/D\ (Hz/ms),$$

where F_n denotes F1 or F2 (in Hz), and D (in ms) the duration of the transition. Means and standard deviation of RoC are summarized in table 6.

Table 6: Means and standard deviation of Formant RoC (in Hz/ms).

		F1 RoC		F2 RoC	
		CD	CDS	CD	CDS
ia	N	4.34 (0.25)	5.2 (0.55)	8.2 (2.14)	9.78 (1.65)
	F	5.29 (0.83)	5.77 (1.13)	9.08 (1.27)	10.5 (1.9)
iɛ	N	1.47 (0.42)	2.13 (0.37)	1.83 (0.46)	3.9 (1.52)
	F	3.77 (0.49)	4.28 (0.46)	7.09 (1.28)	8.75 (1.14)
io	N	2.36 (0.6)	3.18 (0.53)	10.5 (0.94)	12.8 (1.71)
	F	2.49 (0.38)	2.96 (0.35)	8.81 (1.66)	9.81 (4.34)
ua	N	3.75 (0.9)	4.44 (0.94)	4.16 (0.75)	5.71 (0.83)
	F	3.95 (0.96)	4.43 (1.27)	5.48 (0.83)	6.71 (1.09)
uɔ	N	0.88 (0.21)	1.44 (0.32)	1.13 (0.58)	2.23 (0.55)
	F	3.32 (0.75)	3.65 (1.13)	5.25 (1.3)	5.74 (0.89)
ɥa	N	3.11 (1.26)	4.49 (0.95)	0.25 (0.18)	0.97 (0.74)
	F	4.89 (0.75)	5.63 (1.19)	1.49 (1.1)	1.5 (1.26)

Two-way ANOVA shows that syllable type has significant effects on and F1 and F2 RoC of [iɛ] and [uo], and F2 of [ua] (p<0.01), with higher F1 and F2 RoC in (C)D than in (C)DS syllables. For the other diphthongs, the F1 and F2 RoC in the two syllable types do not show significant differences. On the other hand, speech rate has significant effect on the F1 RoC of [iɛ, iɔ, ɥa] (p<0.01) and [uo] (p<0.05), and F2 RoC of [iɛ, ua] (p<0.01) and [ia, iɔ] (p<0.05). A faster rate causes a general increase in F1 and F2 rate-of-change. There is no significant interaction between the two factors.

4. Conclusion and discussion

The duration of the diphthongs in HC are significantly shortened in both (C)DS syllables and fast speech. But according to the present results, the two factors have different effect on the temporal structure and spectral characteristics of the diphthongs.

In (C)DS syllables, mainly the second elements sacrifice for duration shortening, resulting in a change in temporal structure of the diphthongs. The 2nd elements of most of the diphthongs undershoot their targets as compared with those in (C)D syllables.

In fast speech, all the three components of the diphthongs contribute to duration shortening, keeping the temporal structure unchanged. Therefore, the (C)D-diphthongs in fast-speech have different temporal structure from the (C)DS-diphthongs in normal-speech. It is somewhat surprising that no significant undershoot as an effect of speech rate increase occur as predicted by the vowel undershoot model. The RoC data show a general increase of the rate of formant change during the transition when produced in fast speech, which may explain the lack of target undershoot. That is, the speakers increased the velocity of movement of the articulators to achieve the diphthong targets in a short duration.

5. REFERENCES

[1]. Fourakis, M. 1991. Tempo, stress, and vowel reduction in American English. *J. Acoust. Soc. Am.* 90, 1816-1827.

[2]. Gay, T. 1978. Effect of speaking rate on vowel formant movements. *J. Acoust. Soc. Am.* 63, 223-230.

[3]. Lindblom, B. 1963. Spectrographic study of vowel reduction. *J. Acoust. Soc. Am.* 35, 1773-1781.

[4]. Lindblom, B. 1990. Explaining phonetic variation: a sketch of the H and H theory. In Hardcastle, W. and Marchal, A. (eds), *Speech Production and Speech Modeling*. Dordrecht: Kluwer Academic Publishers, 403-439.

[5]. Miller, J. L. Effects of speaking rate on segmental distinctions. In P. D. Eimas & J. L. Miller (Eds.), *Perspectives on the study of speech*, LEA, Hillsdale: NJ, 39-74.

[6]. Van Son and Pols, 1990, Formant frequencies of Dutch vowels in a text, read at normal and fast rate. *J. Acoust. Soc. Am.* 88, 1683-1693.