

Effect of Speaking Rate on Diphthong Formant Movements

THOMAS GAY

Hunter College of the City University of New York, New York 10021

and

Haskins Laboratories, New York, New York 10017

Duration and formant frequency measurements of the five diphthongs, /ɔɪ, aɪ, aʊ, eɪ, oʊ/, were made for three conditions of speaking rate. Results indicate that onset target position and second-formant rate of change are fixed features of the diphthong formant movement, while offset target positions are variable across changes in duration. Other measurements show that increased diphthong duration due to the presence of a following voiced consonant is accomplished by longer steady-state targets for /ɔɪ, aɪ, aʊ/ and longer glide durations for /eɪ, oʊ/. The data provide a basis for an articulatory description of diphthong production.

INTRODUCTION

THE principle diphthongs of American English, /ɔɪ, aɪ, aʊ, eɪ, oʊ/, have been studied acoustically by several investigators¹⁻³ who compared the formants of the initial and terminal target areas to those of measured vowels. Generally, it was found that diphthong targets are not necessarily phonetically compatible with the vowels used to describe them; for example, the onset of /aɪ/ can vary from [a] to [æ] and the offset from [e] to [i]. Recently, however, experiments using synthetic speech⁴ have suggested another parameter of the diphthong movement, namely rate of formant-frequency change or in articulatory terms, speed of movement. These experiments indicated that the shift in perception from simple vowel to diphthong is a function of increased duration and that the distinction between /ɔɪ/ and /aɪ/ at least, is governed by the rate of change of the formant transition rather than the onset or offset target positions. Exploratory real-speech measurements supported these findings in showing that changes in the duration of the phoneme (produced by different speaker rates of production) are reflected by changes in onset or offset target positions rather than by modification of

formant rate of change. The purpose of the experiment reported here was to investigate fully the effects of different speaker rates of production on the formant movements of the five diphthongs, /ɔɪ, aɪ, aʊ, eɪ, oʊ/, and in addition, to determine the effects of consonant context on duration and target-frequency position.⁵

I. PROCEDURES

Both duration and formant-frequency measurements were made from spectrograms of words recorded in a sentence context. The word list contained 50 different CVC monosyllables with each diphthong represented ten times. Minimal pairs (for example, "pie-buy," "dice-dies") were used where possible and accounted for at least four of the 10 words per diphthong. Each word was placed in a stressed position in a different sentence. All sentences were random ordered into a master list.

This list was recorded by five young adult male undergraduates whose speech was typical of the New York City dialect area. Each recorded the list three times: first, at a normal conversational rate, then at a fast rate, and finally at a slow rate. The three rate conditions were dictated by the speakers' own estimation of natural production. Ample practice time preceded each condition.

¹ R. K. Potter and G. E. Peterson, "The Representation of Vowels and their Movements," *J. Acoust. Soc. Amer.* 20, 528-535 (1948).

² I. Lehiste and G. E. Peterson, "Transitions, Glides and Diphthongs," *J. Acoust. Soc. Amer.* 38, 268-277 (1961).

³ A. Holbrook and G. Fairbanks, "Diphthong Formants and their Movements," *J. Speech Hear. Res.* 5, 33-58 (1962).

⁴ T. Gay, "A Perceptual Study of American English Diphthongs," *Status Rep. Speech Res., Haskins Lab. Suppl.*, 1-80 (1967).

⁵ A distinction should be made between the classes /ɔɪ, aɪ, aʊ/ and /eɪ, oʊ/. The diphthongal nature of /ɔɪ, aɪ, aʊ/ is phonemically distinctive in most dialects of American English. /eɪ, oʊ/, on the other hand, alternate with the simple vowels [e, o], suggesting that their offglides carry no phonemic significance.

Recordings were made on an Ampex, AG-500 tape recorder through an Electrovoice, model 654 microphone in a quiet but not fully sound-treated room. Spectrograms were made on a Kay, model 661-B spectrograph using the 300-Hz filter setting. Measurements were made for onset steady-state, glide, and offset steady-state durations and first- and second-formant onset and offset frequency levels. Duration and frequency measurements were made to the nearest 5 msec and 20 Hz, respectively. A total of 50 words were analyzed for each diphthong under each of the three duration conditions.

II. RESULTS

A. Main Duration Effects

Table I shows the duration measurements for all three speaker conditions. In general, reduced over-all duration is reflected by shorter durations of each of the three segments. For the two slowest conditions, each phoneme contains both steady-state targets while at the fast condition, one or the other, either onset or offset, is negligible or not present.⁶ For all conditions, only /ɔɪ, aɪ/ contain relatively prominent steady states, at onset. Offset steady states are considerably shorter. Onset and offset steady states for /aʊ/ are comparable in duration to the offset steady states of /ɔɪ, aɪ/. Steady states are least prominent in /eɪ, oʊ/. Conversely, of course, glide durations are relatively longest for /eɪ, oʊ/ and shortest for /ɔɪ, aɪ/.

The effects of duration on first- and second-formant target frequencies are shown in Table II. For all diphthongs, there is little consistent change in onset frequency positions for the three duration conditions. Both first- and second-formant offsets, however, show consistent changes in frequency levels across changes in duration. For /ɔɪ, aɪ, eɪ/, first formant offsets are higher and second formant offsets are lower as phoneme duration decreases. For /aʊ, oʊ/, both first- and second-formant offsets are higher for conditions of shorter duration. Second-formant offset differences between the slow and fast conditions range from a high of 281 Hz for /aɪ/ to a low of 131 Hz for /aʊ/. Further, and perhaps more importantly, the second-formant rate of change for each diphthong remains relatively constant across changes in duration and distinct from the rates of change of the other diphthongs.⁷ These effects, which occurred consistently for all five speakers, are illustrated spectrographically in Fig. 1. As speaker rate of production increases, both onset steady-state and glide durations decrease. What is significant, however, is that the decrease in glide duration is accompanied by a decrease in second-formant offset frequency (and an increase in first-formant offset frequency) rather than

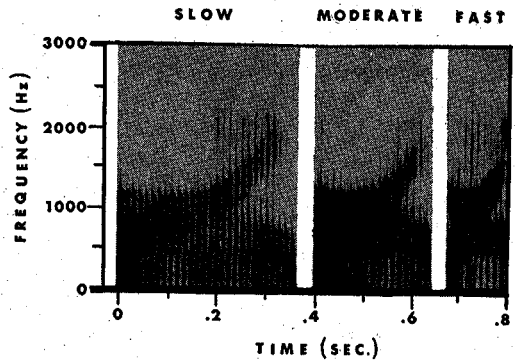


FIG. 1. Spectrograms of the word "boy" for three conditions of speaking rate. F2 offset is 1750 Hz (slow), 1600 Hz (moderate), 1340 Hz (fast). F2 rates of change are 7.0, 7.2, 7.2 Hz/msec, respectively.

a modification of second-formant rate of change. In other words, the glide course remains stable with the gesture simply terminating before reaching the offset target.

The effects of these changes on the phonemic identity of the targets are shown in Figs. 2 and 3, where the diphthong formants are plotted on an F1-F2 grid against positions of steady-state vowels produced by the speakers. Each diphthong, with the possible exception of /aʊ/, shows a clear change in offset identity. Onset positions, although generally preserved across changes in duration, are not clearly identifiable. The plots further illustrate that each diphthong follows a similar course across changes in duration with the shorter utterances terminating before reaching their targets. One exception is /ɔɪ/, where the tract for the slow production is characterized by lower first-formant onset and offset values. Second-formant values, however, are consistent with those of the moderate and fast productions, and the course of the slow production runs

TABLE I. Mean diphthong durations (in milliseconds) for three conditions of speaking rate.

		Onset steady state	Glide	Offset steady state	Over-all
/ɔɪ/	Slow	90	130	29	249
	Moderate	57	108	14	179
	Fast	21	91	...	112
/aɪ/	Slow	66	123	28	217
	Moderate	39	96	22	157
	Fast	17	88	...	105
/aʊ/	Slow	33	151	32	216
	Moderate	21	112	20	153
	Fast	6	98	2	106
/eɪ/	Slow	9	172	22	203
	Moderate	4	112	18	134
	Fast	...	84	7	91
/oʊ/	Slow	13	172	18	203
	Moderate	6	130	11	147
	Fast	2	93	...	95

⁶ A steady state was measured as such if it had a minimum duration of 15 msec.

⁷ Measurements for first formant rates of change were not made because of possible measurement error effects.

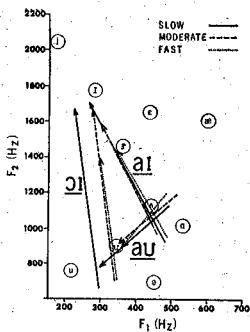


FIG. 2. F1-F2 grid for /ɔɪ, aɪ, aʊ/.

parallel to the others. The target frequency values reported here are not in full agreement with those of Lehiste and Peterson² and Holbrook and Fairbanks,³ nor would this be expected since different dialect areas were studied. One point in agreement, however, is that the targets of /ɔɪ, aɪ, aʊ, eɪ, oʊ/ are not necessarily compatible with the vowels used to describe them.

B. Other Duration Effects

Formant frequency and duration measurements were made for the diphthongs in various CVC consonant environments. The measurements did not indicate any consistent initial or terminal consonant effects on target frequency levels or initial consonant effects on duration. As would be expected, however, there was a terminal consonant voicing effect on duration. Table III shows these effects for 11 minimal pairs. Measurements are those for the moderate condition only, although the same effects are evident for the two other duration conditions. The increased duration of /ɔɪ, aɪ/ preceding a voiced consonant is accomplished primarily by a lengthening of the steady-state onset. Glide and steady-state offset durations are generally also longer but to

TABLE III. Mean diphthong durations (in milliseconds) at moderate speaking rate.

	Onset steady state	Glide	Offset steady state	Over-all
voit	28	93	5	126
void	72	103	11	186
bois (terous)	41	81	12	134
boys	67	89	18	174
bite	9	91	28	128
bide	47	97	30	174
fife	19	99	21	139
five	32	96	15	143
dice	25	81	12	118
dies	43	89	17	149
bout	...	99	10	109
bowed	...	114	40	154
house	19	113	19	135
house (voiced)	28	124	48	229
gape	...	115	...	115
Gabe	...	120	...	120
bait	...	100	14	114
bade	...	108	19	127
dose	...	101	...	101
doze	...	144	...	144
boat	...	98	...	98
bode	...	106	...	106

much lesser degrees. The voicing effect for /aʊ/, on the other hand, is reflected by a longer steady-state offset with smaller steady-state onset and glide changes. These differences between /ɔɪ, aɪ/ and /aʊ/ voicing effects appear consistent with the presence of prominent steady states in /ɔɪ, aɪ/ and the absence of such in /aʊ/. Voicing effects for both /eɪ, oʊ/ are reflected primarily by changes in glide duration as steady-state onsets and offsets are not present (except for the "bait-bade" pair, which contains steady-state offsets). Further, the voicing effects for /eɪ, oʊ/ are generally smaller than those for /ɔɪ, aɪ, aʊ/.

III. DISCUSSION

The results of this study have indicated two features governing the diphthong formant movement: onset frequency position and second-formant rate of change. Although the synthesis data mentioned earlier⁴ showed only formant rate of change relevant to perception, a fixed-frequency onset position is not unexpected. This segment of the phoneme carries both stress and, especially for /ɔɪ, aɪ/, duration responsibilities, as well

TABLE II. Mean formant frequencies and formant rate of change for three conditions of speaking rate.

		F1 Onset (hertz)	F1 Offset (hertz)	F2 Onset (hertz)	F2 Offset (hertz)	F2 Rate (hertz/millisecond)
/ɔɪ/	Slow	296	226	658	1676	8.19
	Moderate	341	279	706	1664	8.95
	Fast	347	295	707	1400	8.13
/aɪ/	Slow	485	265	919	1709	5.83
	Moderate	459	302	970	1590	5.62
	Fast	487	346	951	1428	5.32
/aʊ/	Slow	493	299	1132	775	2.71
	Moderate	516	341	1200	875	2.95
	Fast	484	366	1198	906	2.91
/eɪ/	Slow	290	195	1612	1946	2.06
	Moderate	317	245	1569	1765	1.93
	Fast	339	285	1593	1740	1.90
/oʊ/	Slow	311	208	915	691	1.48
	Moderate	365	289	999	819	1.06
	Fast	331	318	955	853	1.38

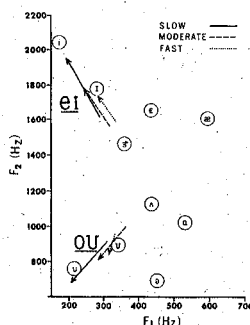


FIG. 3. F1-F2 grid for /eɪ, oʊ/.

as providing, perhaps by consequence of the glide, a convenient starting point. This starting point occurs in the vowel space at a place apparently unrelated to any one particular vowel position.

The classes /ɔɪ, aɪ, aʊ/ and /eɪ, oʊ/, although sharing some similar dynamics, should nonetheless be treated separately, at least for the dialect area studied here. Unlike /ɔɪ, aɪ, aʊ/, the diphthongal variants /eɪ, oʊ/ show little steady-state target influences and the significance of their glide properties is reduced since each has a simple vowel equivalent. Over-all descriptions then of the two classes are made separately.

/ɔɪ, aɪ, aʊ/ are each characterized primarily by onset steady state and continuously changing glide segments. Offset steady states, except for /aʊ/, are least prominent. The articulatory gesture begins at a fixed position in the vowel space. This position is held for a particular length of time, the duration of which is governed by the speaker's rate of production and/or the presence or absence of a following voiced consonant. The articulators then begin to move toward a final target position at a given speed. If speaker rate is slow, the target is reached and the gesture is completed; if speaker rate is fast, the movement, while on course, is cut off before

reaching the final target. /eɪ, oʊ/ behave similarly, except there is little or no onset steady-state influence, and changes in speaker rate are reflected primarily by changes in glide duration and offset frequency position.

These descriptions support a treatment of /ɔɪ, aɪ, aʊ, eɪ, oʊ/ as unit phonemes rather than as sequences of a vowel plus semivowel,⁸ or more traditionally, vowel plus vowel. The measurements reported here provide no indication of any phoneme sequencing in terms of either formant positions or over-all movement. The results of this study suggest, rather, a movement from one position in the vowel space toward, but not necessarily reaching, another position whose course might best be transcribed by using a superscript form, as in [ɔ^t, a^t, a^u, e^to^u].

ACKNOWLEDGMENT

The author wishes to express his thanks to Mrs. Celia F. Dorrow for her assistance in the collection and analysis of the data.

⁸ G. L. Trager and H. L. Smith, *An Outline of English Structure* (Battensburg Press, Norman, Oklahoma, 1951).