INFLUENCE OF SPEAKING RATE IN SPANISH DIPHTHONGS

GUILLERMO A. TOLEDO

NORMA ANTOÑANZAS-BARROSO

Dept. of Linguistics, University of California at Los Angeles
Los Angeles, CA, USA 90024

ABSTRACT

The universal diphthong production model, i.e., invariance in F2 onset frequency and in the rate of change of F2 transition in spite of differences on speaking rate, was reconsidered. Three Argentinian Spanish speakers'data: a corpus of nonsense words, CVVC and CVVCV, consisting of diphthongs /ai ia au ua eu ue oi io/ embedded in a consonantal and syllable environment: initial /t/, final /s/ and /sa/ was analyzed through LPC analysis. A control of the acoustic measurements on three speaking rates (slow, moderate and fast) was taken into account. A specially-written computer program for the measurement of F2 slope (maximum: four steepest points in 30 ms, and normal: from the onset to the offset) was performed. Results indicated significant differences in F2 rate of change. However, F2 onset frequency showed invariance in some data, and significant differences in the rest. Spanish diphthong production would appear to have language- or diphthong-dependent patterns than universal ones.

INTRODUCTION

It has been claimed that the onset frequency and the rate of change of the second formant transition in the production of English diphthongs is constant in spite of differences on speaking rates /1/. In addition, it has been suggested that the rate of change, rather than absolute F2 frequency, is the primary acoustic cue for diphthong recognition /2/.

In an acoustic study, similar results were obtained by Manrique for the fourteen Spanish diphthongs /3/. She worked on corpus of natural words where diphthongs appear in a labial-dental or dental-labial environment. Tokens were emitted under two speaking rates: moderate and fast. F2 rate of change calculation (Hz/ms) through manual spectrographic measurements was made. In so doing, Manrique also observed that the rate of change of the second formant transition shows an invariant pattern under suprasegmental changes of speaking rate. In this previous research, however, three relevant observations should be made: 1) spectrographic analysis is inappropiate for measurements under fast speaking rate: formant crushing interferes with accurate results, 2) selection of a corpus of natural speech (words in a carrier sentence) would be unsuitable because of temporal and spectral distortions caused by different consonantal environments and 3) no statistical differences among results were reported. Besides, in recent literature, this universal model for diphthong production was reanalyzed. Dolan and Mimori /4/ reported two different calculations of F2 rate of change: one, in four points of the transition (30 ms), i.e., the steepest ones, and two, the rate of change of this F2 transition throughout its total duration. They observed that changes in speech rate influenced the F2 slope in English and in Japanese diphthongs even though in Japanese the influence of speech rate was less important than in English.

The purpose of this work, then, was the observation of Spanish diphthongs under changes in speaking rate, related to the universal model previously suggested.

PROCEDURE

Spanish material consisted of nonsense words, CVVC and CVVCV, made up by diphthongs /ai ia au ua eu ue oi io/ embedded in a consonantal and syllable environment: initial /t/, final /s/ and /sa/. These frames were chosen because alveolar or dental consonantal context has less acoustic influence over transitions than the rest of consonants /4/. The corpus, then, was composed of 16 words x 3 male adult Argentinian speakers x 3 repetitions x 3 speaking rates: slow, moderate and fast. Tokens were embedded in a carrier sentence to avoid context effects. Speakers were instructed to control changes in speed through previous listening of reference speech samples emitted by the first author who accurately controlled different speaking rates. The three speakers appeared to be fit for solving the task.

Sentences were recorded on magnetic tape in a sound-proof room. Then, Ss' recorded and selected tokens were sampled on a PDP11/23 Digital computer for acoustical analysis. A 10-KHz sampling rate with a 4.5KHz low-pass filter was used. A 25.6 ms Hamming window with 10 ms intervals was utilized for analysis through linear predictive coefficient (LPC) using UCLA WAVES speech analysis system programs. Tokens were measured by a special computer program. F2 onset and offset, transition duration, maximum rate of change of transition in the steepest four points with a duration of 30 ms (henceforth: MAXSLOPE), and rate of change of complete F2

TABLE I.- MEANS AND STANDARD DEVIATIONS (IN PARENTHESIS) OF F_2 MAXSLOPE AND AVESLOPE. MEANS (Hz) AND STANDARD DEVIATIONS (Hz, IN PARENTHESIS) OF F_2 ONSET, OFFSET AND RANGE.

KEY : SLOW = S MODERATE = M FAST = F

		/ai/	/ia/	/au/	/ua/	/eu/	/ue/	/oi/	/io/
MAXSLOPE	S	10.51 (2.53)	-10.86 (1.99)	-5.20 (1.35)	6.07 (1.22)	-11.05 (8.24)	14.29 (2.87)	15.42 (3.05)	-18.90 (3.62)
	M	9.46 (0.89)	-10.46 (1.39)	-5.00 (1.11)	6.48 (1.78)	-14.44 (2.33)	14.47 (2.74)	17.13 (3.35)	-17.61 (3.27)
	F	10.62 (1.54)	-11.20 (1.39)	-5.28 (1.12)	6.10 (1.89)	-13.99 (2.59)	12.33 (8.64)	20.76 (6.02)	-17.50 (3.90)
AVESLOPE	S	6.10 (1.45)	-7.11 (1.40)	-3.52 (1.17)	4.11 (0.69)	-6.63 (4.78)	8.97 (1.29)	7.85 (1.58)	-8.47 (1.99)
	M	6.19 (0.59)	-6.65 (1.78)	-3.79 (0.77)	4.79 (0.97)	-9.20 (1.01)	9.44 (1.53)	9.07 (1.25)	-9.76 (1.52)
	F	7.08 (1.20)	-7.49 (1.97)	-4.73 (0.71)	5.02 (1.20)	-9.43 (1.88)	8.63 (5.86)	11.44 (1.75)	-10.53 (1.14)
ONSET	S	1390.30 (82.22)	2258.75 (76.37)	1302.04 (136.01)	801.13 (100.43)	1961.55 (135.31)	853.55 (143.35)	974.90 (81.68)	2194.82 (71.57)
	M	1434.03 (57.34)	2112.31 (127.15)	1254.49 (56.56)	818.42 (90.53)	1929.16 (126.98)	910.58 (142.87)	1009.25 60.07	2155.17 (87.29)
	F	1448.17 (56.50)	2085.03 (122.88)	1282.13 (46.89)	871.25 (136.94)	1847.90 (124.03)	959.29 (182.02)	1082.65 (66.55)	2118.69 (62.16)
OFFSET	S	2216.42 (95.03)	1496.33 (82.94)	908.72 (86.06)	1292.52 (63.01)	962.10 (144.05)	1935.23 (93.33)	2066.55 (126.16)	1010.96 (74.13)
	M	2122.71 (93.82)	1458.83 (79.10)	976.87 (94.61)	1215.83 (80.51)	1019.97 (136.51)	1864.23 (47.09)	2119.51 (129.88)	1109.25 (93.28)
	F	2103.65 (122.76)	1473.39 (87.32)	1030.23 (100.90)	1218.88 (59.78)	1100.23 (99.77)	1773.47 (154.39)	2032.18 (168.23)	1165.72 (75.58)
RANGE	S	826.12 (129.30)	762.42 (71.33)	393.32 (123.34)	491.39 (91.96)	999.45 (199.24)	1074.82 (229.56)	1091.65 (115.69)	1183.86 (89.56)
	M	688.68 (102.13)	653.48 (148.69)	277.62 (90.63)	397.41 (149.44)	909.19 (190.10)	953.65 (178.70)	1010.61 (154.50)	1045.92 (143.59)
	F	655.48 (89.20)	611.64 (158.24)	251.90 (69.76)	347.63 (181.61)	747.73 (150.27)	849.97 (202.90)	949.53 (205.12)	952.97 (85.38)

transition (henceforth: AVESLOPE) were calculated. Through this program a detection of transitional segments from steady-state portions was made. The starting point was indicated by differences over 15 Hz in between two consecutive windows, the preceding one with a 10 ms interval duration. Similar treatment for English (with 15 Hz acoustic difference, and for Japanese (with a 20 Hz value) was reported by Dolan and Mimori /4/.

RESULTS

Table I shows means and standard deviations of F2 rate of change on MAXSLOPE and AVESLOPE calculations, and means and standard deviations of F2 onset, offset and range. Results of ANOVA indicated significant differences among speech rates in MAXSLOPE (F (12, 419)= 97.62, p>0.0001), and in AVESLOPE (F (12, 419)= 76.82, p>0.0001). Similarly, ANOVA calculations showed significant differences of range among the three speaking rates (F (12, 419)= 182.40, p>0.0001). ANOVA results also revealed that AVESLOPE and range were strongly influenced by rate: the F ratio for rate related to AVESLOPE was 34.77 (2 and 419 df and p>0.0001), and the F ratio for rate connected with range was 76.22 (2 and 419 df and p>0.0001). However, ANOVA calculations yielded a less significant influence of rate in MAXSLOPE (F (2, 419)= 2.83, p>0.0599).

Table II shows statistical differences (Student's t test) between F2 onset frequencies influenced by different speaking rates. In the same manner the table displays statistical differences between F2 offset frequencies. Inspection of the table reveals

TABLE II.- STATISTICAL DIFFERENCES (t-test) ON F_2 ONSET AND F_2 OFFSET INFLUENCED BY CHANGES IN SPEAKING RATE

KEY: ns = NOT SIGNIFICANT x = p < 0.0001 xx = p < 0.02 xxx = p < 0.05

S = SLOW M = MODERATE F = FAST

	F_2 ONSET								F_2 OFFSET							
	/ai/	/ia/	/au/	/ua/	/eu/	/ue/	/oi/	/io/	/ai/	/ia/	/au/	/ua/	/eu/	/ue/	/oi/	/io/
F vs S	ns	xxx	ns	ns	ns	ns	xxx	xxx	xx	ns	xx	ns	ns	xxx	ns	ns
F vs M	ns	ns	ns	ns	x	xxx	xx	xx	ns	ns	ns	ns	xx	ns	ns	ns
S vs M	ns	ns	ns	ns	ns	ns	ns	ns	xx	ns	xx	ns	ns	ns	ns	ns

that only F2 onset in diphthongs /ai au ua/ remained invariant and uninfluenced by speaking rate. The reverse was true in diphthongs /ia eu ue oi io/. Calculations on F2 offset frequencies resulted in an invariant trend in diphthongs /ia ua oi io/ and, on the contrary, resulted in significant differences in diphthongs /ai au eu ue/.

Correlation coefficients (Fig. 1 and Fig. 2), on the other hand, showed that MAXSLOPE was highly

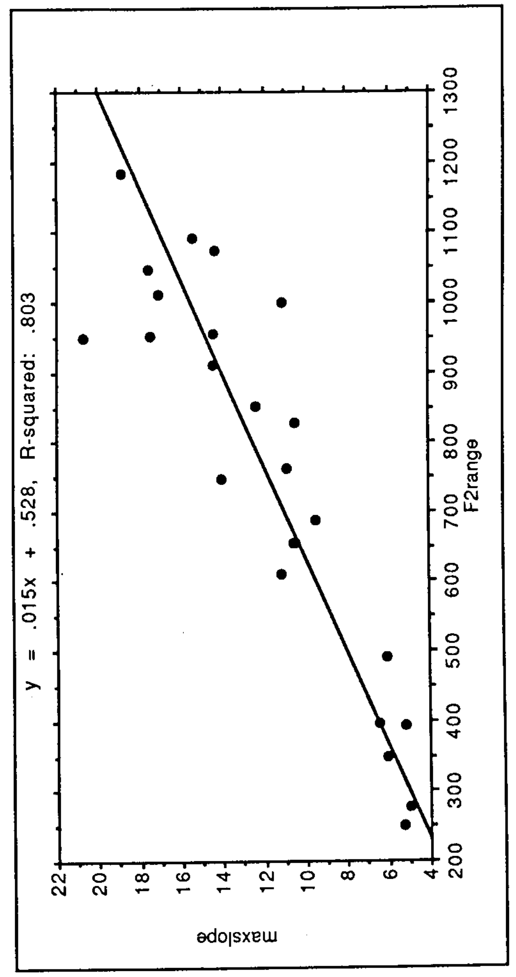

FIGURE 1. CORRELATION COEFFICIENTS BETWEEN MAXSLOPE AND F2 RANGE

correlated with range: R= 0.803, p> 0.0001, but poorly correlated with transition duration: R= 0.113, p> 0.0183. AVESLOPE showed a similar trend: a high degree of correlation with range: R= 0.654, p>0.0001, but a low one with transition duration: R= 0.119, p> 0.0001.

DISCUSSION

ANOVA results obtained from Spanish data dueweight to reject Gay's model /1/: changes in speech rates resulted in significant statistical differences in F2 rate of change. Similar conclusions were reported by Dolan and Mimori for English and Japanese complex vowels /4/. In addition, F2 onset frequencies showed significant variations in diphthongs /ia eu ue oi io/ because of changes in speaking rate.

According to Gay's model, no correlation should appear between slope and either range and transition duration if the F2 rate of change remain invariant and uninfluenced by different speech rates. Correlation coefficients in Spanish data reject this notion, at least in one aspect: the high degree of correlation between slope and range.

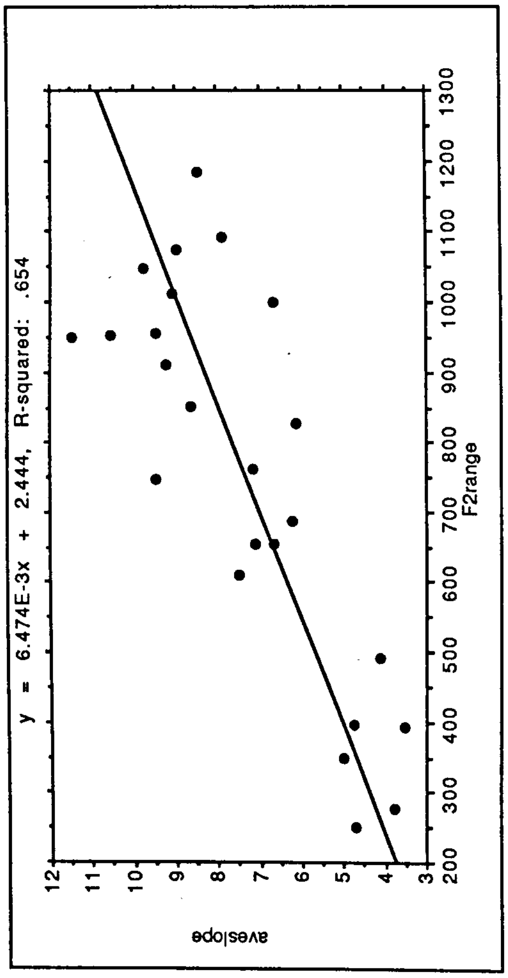

FIGURE 2. CORRELATION COEFFICIENTS BETWEEN AVESLOPE AND F2 RANGE

In brief, the production of diphthongs in Argentinian Spanish shows a language-specific trend rather than an universal-specific one previously suggested by Gay /1/ and by Manrique /4/. This position is supported by recent research on diphthongs in different languages: English and Japanese /4/, Hausa /5/, Chinese /6/, and -in a cross-phonetic study- Hausa, Arabic and Chinese /7/.

REFERENCES

/1/ Gay, T. (1968). Effect of speaking rate on diphthong formant movements, The Journal of the Acoustical Society of America 44, 6, pp. 1570-1573.

/2/ Gay, T. (1970). A perceptual study of American English diphthongs, Language and Speech 13, 2, pp. 65-88.

/3/ Manrique, A.M.B. (1979). Acoustic analysis of the Spanish diphthongs, Phonetica 36, 3, pp. 194-206.

/4/ Dolan, W. and Mimori, Y. (1986). Rate-dependent variability in English and Japanese complex vowel F2 transitions, The Journal of the Acoustical Society of America 80, Suppl.1, S96.

/5/ Lindau-Webb, M. (1985). Hausa vowels and diphthongs, Studies in African Linguistics 16, 2, pp. 161-182.

/6/ Ren, H. (1986). A truncation model for the F2 trajectory in syllables with complex vocalic components, The Journal of the Acoustical Society of America 80, Suppl.1, S98.

/7/ Lindau, M., Norlin, K. and Svantesson, J. (1985). Cross-linguistic differences in diphthongs, UCLA Working Papers in Phonetics 61, pp. 40-44.

ACKNOWLEDGEMENTS

This work has been supported by a grant awarded to the first author, Senior Scholar Fulbright Program: Advanced Research Award in the United States (N°85-08139), by the Council for the International Exchange of Scholars (CIES) affiliated with the American Council on Education. This research has been also supported by the Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Argentina through a salary appointment to the first author at the Laboratorio de Investigaciones Sensoriales, Buenos Aires, Argentina.

We are particularly indebted to Peter Ladefoged, Mona Lindau, Ian Maddieson and William Dolan for fruitful comments, guide and inspirational help.