INTER-GESTURAL TIMING BETWEEN VOCALIC GESTURES AS A FUNCTION OF SYLLABLE POSITION: ACOUSTIC EVIDENCE FROM ROMANIAN

Ioana Chitoran

INTER-GESTURAL TIMING BETWEEN VOCALIC GESTURES AS A FUNCTION OF SYLLABLE POSITION: ACOUSTIC EVIDENCE FROM ROMANIAN

Ioana Chitoran

Linguistics and Cognitive Science
Dartmouth College

ABSTRACT: This study investigates the relative timing of vocalic gestures as a function of syllable position, manifested in the traditional distinction between hiatus and diphthong. The hiatus-diphthong distinction has been treated as one in syllable affiliation, therefore differences in their production can be interpreted as phonetic correlates of the syllable. I compare the acoustic realization of Romanian diphthongs [jV] and similar hiatus sequences [iV]. The latter have variable pronunciation, occasionally realized as diphthongs. A contrast is maintained between them, although individual speakers show tendencies toward a merger. Differences between diphthongs and hiatus sequences are found in: the vocalic portion duration, the F2 value at the onset of the vocalic portion, the duration of high energy frication at the release of a preceding [p]. I propose that all of these differences result from differences in articulatory timing between [i]/[j] and V. Inter-gestural timing in [jV] is tightly controlled, producing a short vocalic gesture for [j]. In a hiatus sequence [iV] timing is loose, allowing for variation. An occasional early onset of the V gesture can produce a glide percept. The results are interpreted in the framework of Browman and Goldstein's (2000) Bonding Strength Model.

INTRODUCTION

A growing body of literature has provided articulatory and acoustic evidence for syllable position effects in several languages (see Kochetov 2002 for a review). We now have a good understanding of positional variation as manifested in the inter-gestural timing of consonants involving multiple oral gestures, and in the coordination of consonantal and vocalic gestures in onset and coda position (Sproat and Fujimura 1993, Keating 1995, Fougeron and Keating 1997, Gick 1999, Kochetov 2002, among others). A question which has not yet been explored concerns the relative timing of vocalic gestures as a function of syllable position. In this context, an interesting object of study is the traditional distinction between hiatus and diphthong, which is understood as a distinction in syllable affiliation, and is expected to be reflected in production differences.

A language suitable for investigating this question is one which contains both diphthongs and hiatus sequences involving the same vocalic gestures. Several Romance languages meet this criterion, and the present study focuses on Romanian. The distribution of prevocalic [j] and [i] in Romanian can generally be predicted from the location of stress, except for the very specific environment #C_V, where V is a non-high vowel. In this environment [j] is always present in native words: *pjatrə* 'stone', *pjerde* 's/he loses', *mjere* 'honey', *bjatə* 'poor', *fjerbe* 'it boils', *vjerme* 'worm'. The rising diphthongs [ja] and [je] above developed historically from the diphthongization of Latin /ɛ/ under stress. A preceding coronal or velar further absorbed the palatal glide, resulting in a very restricted distribution of the diphthongs in Modern Standard Romanian – they only occur between a labial and a stressed vowel.

In the #C_V environment the vowel [i] is present in lexical items that are mostly loanwords, or that contained an iV sequence in Latin: *piastru* 'piaster', *pian* 'piano', *tiarə* 'tiara', *dietə* 'diet', *fiasko* 'fiasco', *siestə* 'siesta'. The hiatus pronunciation is not systematic, but is best described as variable, with [i] occasionally coming close to a glide. The presence of variation in hiatus sequences has been established in previous acoustic studies (Chitoran and Hualde 2002, Hualde and Chitoran 2003), and was found to be related to position in the word and to distance from stress/pitch accent. Regardless of this variation, native speakers of Romanian are aware of the distinction between diphthongs and sequences, as demonstrated by their consistent agreement on the syllabification of these vocalic combinations: diphthongs are systematically syllabified as tautosyllabic (jV), while sequences are systematically syllabified with hiatus (i.V).

In this study I compare temporal and spectral differences between diphthongs and hiatus sequences, to determine whether they can be attributed to differences in articulatory timing corresponding to syllable affiliation. The results of this study are relevant for testing speech production models attempting to capture syllable affiliation. I propose an interpretation of these results in the framework of Browman and Goldstein's (2000) Bonding Strength Model, which was developed as a phonological model of the temporal patterning of consonantal and vocalic gestures within and across syllables.

In sections 2 and 3 I report and discuss the results of comparisons in duration (experiment 1), and in F2 values (experiment 2). The theoretical interpretation is proposed in section 4, followed by conclusions in section 5.

EXPERIMENT 1: DURATION

Methods

Data from four native speakers of Standard Romanian were recorded in a sound proof booth, using a Sony DAT recorder, at a sampling rate of 44kHz. The acoustic analysis was done using the Praat speech analysis software. Seven repetitions of each word were recorded, embedded in the frame sentence "Spune ____ de trej or[j]" 'Say ____ three times'. The sentences were read in randomized order. The data consist of diphthongs and hiatus sequences (a total of 161 per speaker). The words were also controlled for stress, number of syllables, the quality of the non-high vowel (mid or low), and as much as possible the quality of the flanking consonants.

Table 1. Sample data

diphthongs	hiatus sequences
fjerbe	fiordu
fjartə	fiasko
pjarə	pianu
bjete	dieta
pjatrə	piastru
mjere	miopu

Two duration measurements were taken, based on waveform and wideband spectrogram. The total duration of the vocalic portion [jV] and [iV] was measured from the onset of F1 to the offset of F2. The duration of high energy frication at the release of initial [p] into the narrow constriction of [i] or [j] was measured from the end of the [p] release burst to the onset of F1 of [i] or [j].

Hypotheses

The comparison of the vocalic portion duration tests the hypothesis that the occasional glide percept in hiatus sequences is due to a shorter acoustic duration of [i] or of the transition between [i] and V. It is expected, however, that a contrast is maintained between diphthongs and hiatus sequences, and consequently the total duration of the monosyllabic diphthongs [jV] is hypothesized to be significantly shorter than that of disyllabic sequences [iV].

The comparison of the frication portion duration at the release of [p] tests the hypothesis that the occasional glide percept in hiatus sequences is due to a narrower constriction of [i]. The duration of frication is partly determined by the degree of constriction of the following vowel. Release into a narrow constriction generates more frication (in energy and duration) than release into a more open constriction. The [p] frication portion preceding the glide of a diphthong will therefore be significantly longer than that preceding the high vowel of a hiatus sequence. This measure is clearly not unambiguous, since it also includes some amount of aspiration at the release of a voiceless stop. However, a qualitative comparison of [p] releases before non-high vowels for the same speakers shows a much shorter, often absent, frication portion in those environments. Longer and higher energy frication may therefore be due primarily to release into a narrower constriction.

Results

Eight pairs of words containing hiatus sequences and diphthongs, like those illustrated in Table 1, were included in the total duration comparison (a total of 112 words per speaker). The total duration of the vocalic portion is evaluated as the dependent variable in an analysis of variance with Type as fixed factor (2 levels: sequence, diphthong). For all four speakers, hiatus sequences (iV) are significantly longer than diphthongs (jV), which suggests that contrast is maintained between them. This result is especially robust, given that the words containing sequences actually have an extra syllable compared to those containing diphthongs. The total duration results are presented in Table 2.

Table 2. Vocalic portion duration in sequences vs. diphthongs (ms). Means and standard deviations.

	$Sp1_{(DS)}$	$Sp2_{(LD)}$	$Sp3_{(BS)}$	$Sp4_{(AG)}$
Sequences (iV)	185 (24.16)	169 (25.91)	223 (33.69)	220 (35.53)
Diphthongs (jV)	128 (21.77)	151 (28.38)	135 (21.26)	173 (30.67)
Effect of Type	$F(1,109)=169.418$ $p<.001$	$F(1,110)=12.219$ $p<.01$	$F(1,103)=243.325$ $p<.001$	$F(1,110)=55.449$ $p<.001$

An investigation of token-by-token duration values reveals interesting differences among individual speakers. Overall, all four speakers maintain a contrast between hiatus sequences and diphthongs, but they differ in the degree to which they do so. Speakers 1, 3, and 4 show a clear contrast. Speaker 2, however, shows much less separation. As many as 41 of her sequences are short enough to fall within the duration range of diphthongs. Impressionistically, her sequences often sound like diphthongs. These results suggest that duration is manipulated in the hiatus/diphthong contrast, as predicted by the first hypothesis. The occasional glide percept is due to a shorter acoustic duration of the vocalic portion. A qualitative comparison of the spectrograms shows a much shorter, often non-existent steady-state portion for [j], and a shorter F2 transition from [j] to V. The F2 trajectory in diphthongs looks almost flat, closer to that of a monophthongal vowel.

For the comparison of the frication portion duration, 4 pairs of words were analyzed, namely those containing hiatus sequences and diphthongs preceded by [p] (a total of 56 words per speaker). Only two of the four speakers show a statistically significant difference between the duration of frication in [piV] vs. [pjV]. The results are presented in Table 3.

Table 3. Duration of high energy frication at the release of [p] in sequences vs. diphthongs (ms). Means and standard deviations.

	$Sp1_{(DS)}$	$Sp2_{(LD)}$	$Sp3_{(BS)}$	$Sp4_{(AG)}$
Sequences (iV)	21 (9.34)	44 (19.8)	23 (13.8)	17 (7.31)
Diphthongs (jV)	26 (16.45)	48 (7.53)	35 (10)	54 (21)
Effect of Type	$p>.05$	$p>.05$	$F(1,52)=12.73$ $p<.01$	$F(1,51)=71.639$ $p<.001$

As predicted by the second hypothesis, the frication portion is longer in diphthongs, before a narrower glide constriction, than in sequences, before a vowel constriction. However, since the difference is statistically significant for only two of the speakers, I conclude that constriction size is less important in maintaining a hiatus/diphthong contrast.

EXPERIMENT 2: F2 ONSET

Methods

Five pairs of words were used for spectral analysis: *fjartə/fiasko*, *pjatrə/piastru*, *pjarə/pianu*, *pjazə/pianu*, *pjardə/piastru*, a total of 70 words per speaker. The analysis consists of running formant tracks over the vocalic portion of each sequence and diphthong, and extracting the value of the second formant at the onset of the sequence/diphthong. A short-term spectral analysis is performed in a 25 ms window, at a time step of 2.5 ms. For each analysis window the algorithm in Praat computes the LPC coefficients. The F2 values extracted by the algorithm were compared to the corresponding spectrograms. Deviating values and the corresponding words were excluded from the statistical analysis. For each sequence/diphthong pair, the formant tracks generated for each individual token (14 per pair) were then overlaid on top of one another and saved to a picture file, for overall qualitative comparison.

Hypothesis

It is hypothesized that hiatus sequences and diphthongs will show different coarticulatory patterns with the preceding consonant. The occasional diphthong percept in sequences is predicted to be due to specific coarticulatory patterns between the consonant and the following high vowel, as manifested in the F2 value at the onset of the vocalic portion. Specifically, it is predicted that the F2 value will be higher before [i] than before [j]. An onset consonant is expected to show more coarticulation with the nucleus vowel [i], seen in a higher F2 onset value, corresponding to a high front vowel. In a diphthong, the nucleus is a non-high vowel, and coarticulation with this vowel is expected to lower the F2 onset value even across an intervening [j].

Results

The F2 value at the onset of the vocalic portion is evaluated as the dependent variable in an analysis of variance with Type as fixed factor (2 levels: sequence, diphthong). As predicted, the F2 onset values are higher in hiatus sequences than in diphthongs, although for one of the four speakers the difference is not statistically significant. The results are reported in Table 4.

Table 4. F2 values at the onset of the vocalic portion (Hz). Means and standard deviations.

	$Sp1_{(DS)}$	$Sp2_{(LD)}$	$Sp3_{(BS)}$	$Sp4_{(AG)}$
Sequences (iV)	2335.97 (134.3)	2103.95 (106.9)	2044.17 (91.3)	2401.69 (89.38)
Diphthongs (jV)	2170.13 (176.3)	2068.97 (141.1)	1908.2 (77)	2293.33 (155.17)
Effect of Type	$F(1,36)=14.264$ $p<.01$	$p>.05$	$F(1,32)=24.772$ $p<.001$	$F(1,64)=12.084$ $p<.01$

Interestingly, speaker 2, whose difference in F2 values is not statistically significant, is the same speaker whose sequences and diphthongs showed less of a separation in total duration (Table 2). The different patterns of two representative speakers are illustrated in Figures 1 and 2 below. These are plots of the vocalic portion duration and F2 onset value. They show overlapping values for Speaker 2 (Figure 1), whereas for speaker 3 they tend more toward a bimodal distribution (Figure 2).

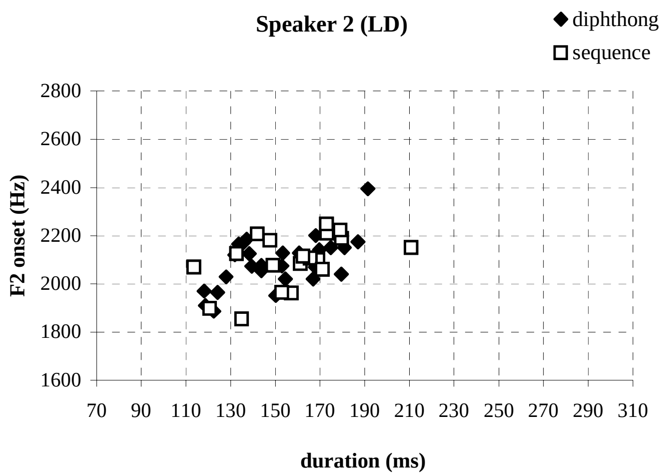

Figure 1. Plot of vocalic portion duration (x axis) and F2 onset value (y axis) for speaker 2.This speaker shows more overlap between diphthongs and hiatus sequences.

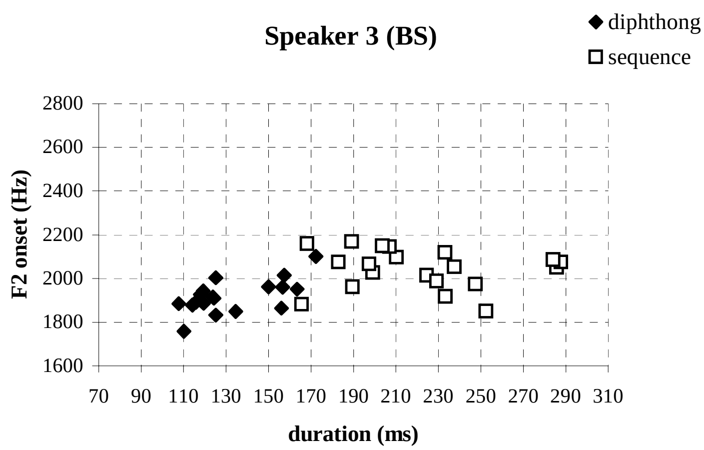

Figure 2. Plot of vocalic portion duration (x axis) and F2 onset value (y axis) for speaker 3.

The results support the hypothesis, suggesting that CV coarticulatory patterns as reflected in the onset value of F2 are important in encoding a sequence/diphthong contrast. A qualitative examination of the spectrograms reveals several other differences. In addition to the flatter F2 trajectory in [jV], the [p] release burst in [pja] is followed by a short formant rise into [j]. In the sequence [pia] the same formant rise is longer and higher. The lower F2 values at the onset of a diphthong suggest the presence of target undershoot in the glide as opposed to [i].

THEORETICAL INTERPRETATION

The theoretical interpretation must explain why, given two contrasting linguistic elements – diphthongs and hiatus sequences – one of them allows for occasional variation which leads to its perceptual confusion with the other. I propose that a suitable model for the interpretation of these data is Browman and Goldstein's (2000) Bonding Strength Model. The authors propose that temporal patterning is the result of weighted competition between a number of gestural coordination relations. The relevant one here is the C-V relation, which defines syllable onsets. According to it, onset consonantal gestures tend to be produced simultaneously with the nucleus vocalic gesture.

According to this model, in hiatus sequences only one coordination relation is present, between the onset C and the high vowel [i]. In diphthongs, which are tautosyllabic, the C and the glide form a complex onset, which as a whole bears a relation to the following V. Two coordination relations are therefore present in diphthongs: a C-V and a glide-V relation. Consequently, in hiatus sequences the relative timing between gestures is more variable and less tightly controlled than in diphthongs, where two coordination relations are active, and the onset as a whole bears a relation to V.

The model predicts that a possible source for the occasional glide percept in sequences lies in the absence of a [i]-V relation across the syllable boundary. In sequences, the V gesture is thus allowed to have an early or a late onset relative to the [i] gesture. If it starts early, it results in a shorter or absent [i] steady-state, and a shorter vocalic portion overall. In diphthongs the C-V and [j]-V relations reinforce each other, and the C and [j] gestures each tend to be timed simultaneously with the V gesture. This coordination pattern results in tight articulatory timing, and an early onset of the V gesture. The same configuration can explain the lower F2 onset values in diphthongs. By allowing a very short time window in which a glide can be acoustically realized, it may be responsible for the F2 target undershoot in [j].

CONCLUSIONS

In this study I compared duration and spectral properties of Romanian hiatus sequences and diphthongs. Based on the results, it was established that the difference between them is manifested primarily in two acoustic parameters: the vocalic portion duration and the F2 value at the onset of the vocalic portion. These acoustic differences can be explained by differences in articulatory timing, with looser timing in hiatus sequences, allowing for variation. A consequence of this variation is a possible early or late onset of the V gesture relative to the [i] gesture. An early onset can produce a glide percept. The results of this production study still have to be supported by perception tests.

ACKNOWLEDGEMENTS

I thank Louis Goldstein, José Ignacio-Hualde, John Kingston, and Donca Steriade for valuable comments on this study. Jesse Beach has been instrumental in writing Praat scripts. All remaining errors are my own.

REFERENCES

Browman, C. & Goldstein, L. (2000) “Competing constraints on intergestural coordination and self organization of phonological structures” *Bulletin de la Communication Parlée* **5** 24-34.

Chitoran, I. & Hualde, J.I. (2002) “Variability in hiatus resolution: a phonetic study of [CiV] sequences in two Romance languages” Poster presented at Laboratory Phonology 8, Haskins Laboratories and Yale University, June 2002.

Fougeron, C. & Keating, P.A. (1997) “Articulatory strengthening at edges of prosodic domains” *Journal of the Acoustical Society of America* **101** 3728-3740.

Gick, B. (1999) “The articulatory basis of syllable structure: a study of English glides and liquids” PhD dissertation, Yale University.

Hualde, J.I. & Chitoran, I. (2003) “Explaining the distribution of exceptional hiatus in Spanish and Romanian” *Proceedings of the 15th International Congress of Phonetic Sciences* 1683-1686.

Keating, P. (1995) “Effects of prosodic position on /t,d/ tongue/palate contact” *Proceedings of the 13th International Congress of Phonetic Sciences* 432-435.

Kochetov, A. (to appear) “Syllable position effects and gestural organization: articulatory evidence from Russian” *Laboratory Phonology* **8**.

Sproat, R. & Fujimura, O. (1993) “Allophonic variation in English /l/ and its implication for phonetic implementation” *Journal of Phonetics* **21** 291-311.