



HAL
open science

Vowels and diphthongs - The articulatory and acoustic structure of Romanian nuclei

Ioana Chitoran, Stefania Marin

► **To cite this version:**

Ioana Chitoran, Stefania Marin. Vowels and diphthongs - The articulatory and acoustic structure of Romanian nuclei. Mark Gibson and Juana Gil (Eds.). Contemporary Studies in Romance Phonetics and Phonology, Oxford University Press, 2018. hal-03127946

HAL Id: hal-03127946

<https://hal.science/hal-03127946>

Submitted on 14 Jul 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Vowels and diphthongs – The articulatory and acoustic structure of Romanian nuclei

Ioana Chitoran and Stefania Marin

X.1 Introduction

Within Romance typology, Romanian belongs to the group of languages that have undergone diphthongization, with diphthongs developing historically from several sources, as we detail below. As a result, Romanian distinguishes between typologically rare mid-diphthongs /ea/, /oa/ and high-diphthongs /ja/, /wa/. At the same time, Romanian is the only Romance language that has, to some extent, maintained a contrast between diphthongs and hiatus sequences of the same vocalic elements (Chitoran and Hualde 2007). Its lexicon is thus characterized by an abundance of glide-vowel and vowel-vowel sequences, which raises the question of the nature of a syllable nucleus in Romanian. In this study we address this issue by bringing together articulatory and acoustic evidence that identifies the differentiating properties of these three categories, limiting our observations to the front context: /ea/, /ja/, /e.a/, /i.a/.

The diphthongs that we study have the following historical sources:

(1) Historical diphthongs

/ja, je/ from Late Latin stressed /ɛ/ after labials:

[petra] > *[pjetra] > [pja_ɪtrɐ] ‘stone’

Palatalization of /l/ in velar-liquid clusters:

CLAVE > [k_ɪje] ‘key’, CLAMAT > [kja_ɪmɐ] ‘calls’

/ea / from Latin stressed /e /:

SERA > [searΛ] ‘evening’

In addition to diphthongs, we examine vowel-vowel sequences in hiatus, word-internally ([miop] ‘short-sighted’) and across word boundary ([krabi#arΛ], [krabj#arΛ]).

On the basis of their phonotactic properties (Cf. Chitoran 2001, 2002a), mid diphthong /ea/ has been assumed to form a complex nucleus, while high diphthong /ja/ has been represented as an onset-nucleus sequence, with hiatus sequences such as /e.a/ and /i.a/ occupying two consecutive syllable nuclei. We predict that this three-way phonological distinction should be reflected in differences at the articulatory and acoustic level. The diphthong /ea/ as a complex nucleus should exhibit the greatest temporal overlap between the two vowels and the largest coarticulation/blend between its targets. The hiatus sequences as vowels in two sequential syllables should show the least overlap and coarticulation, while the diphthong /ja/ should be intermediate, more similar to /ea/ than to hiatus sequences in terms of timing, because of it being a tautosyllabic vs. heterosyllabic sequence.

Previous empirical evidence has shown that indeed /ea/ and /ja/ are reliably distinguished perceptually and they also differ in their acoustic properties at onset (Chitoran 2002b). In terms of duration, /ja/ has also been shown to be longer than /ea/ providing empirical evidence for the proposed structural difference (an onset-nucleus sequence vs. a complex nucleus). A more recent articulatory study (Marin, accepted) examining the three-way contrast between /ea/, /ja/ and /e.a/ has reported a difference in articulatory position between the three categories. However, because of measurement limitations due to the stimuli recorded, this study could only indirectly infer possible timing differences between the categories. The current study, with a set of stimuli better

suites for an articulatory identification of vowel movement, aims to bring direct evidence on the timing differences between the categories, while at the same time providing both articulatory and acoustic evidence on the main characteristics differentiating /ja/ and /i.a/, two sequences not previously compared experimentally.

In the following sections of the paper we present the results of an articulatory (EMA) study and of an acoustic study that address these questions.

X.2 Articulatory study – Method

The articulatory study focuses on the following comparisons:

- diphthong /ea/ vs. diphthong /ja/. Contexts: [Cea#C], [C#jaC]. Both diphthongs are stressed.
- diphthong /ea/ vs. cross-word sequence [e#a] and monophthongs /e, a/
- diphthong /ja/ in [C#jaC] vs. cross-word sequences [Ci#a] and [Cj#a]. In the latter, the glide is a plural marker, the palatalization of the root-final consonant.

X.2.1 Stimuli and data acquisition

Data were recorded from five Romanian native speakers (three female) with no reported speech, hearing or language problems. They all spoke standard Romanian and were naïve as to the purposes of the experiment. They were familiarized with the list of utterances prior to data collection and were instructed to speak at a comfortable rate. During data collection, speakers saw the target utterance on a computer screen and were visually cued when to speak. They repeated each utterance twice per trial in three randomized blocks, resulting in a targeted number of six repetitions per utterance. The stimuli, presented in

Table 1, consisted of target words containing diphthongs /ea/ or /ja/, corresponding cross-word vowel sequences, and control single vowels, all spoken in a constant sentence:

/spun ___/ “*I say ___*” for the /ea/ series, and /spuneam ___/ “*I was saying ___*” for the /ja/ series. The sentences for this experiment were interspersed with filler sentences constituting data sets for other experiments.

Because the aim of this experiment was to identify the vowels on the basis of their articulatory movement, it was necessary for the context preceding the target sequence to be labial. This requirement further limited our choice of stimuli, so that hiatus sequences had to be cross-word rather than within word, and one diphthong had to be word-final and the other word-initial.

Table 1. Target stimuli, with intervals of interest shown in bold face.

Category	Item	Stimulus
Diphthong	/ea/	ka.' fe a # fi.nʌ coffee fine <i>fine coffee</i>
	/ja/	'krab # ' ja .rʌ crab again
Cross-word vowel sequence	/e_ 'a/	'ga. fe # ' a .fi.ne blunders blueberries
	/i_ 'a/	'kra. bi # ' a .rʌ crabs-the plow <i>the crabs plow</i>
	/j_ 'a/	'kra. bj # ' a .rʌ crabs plow
Single vowel	/e/	'ga. fe # 'fi.ne blunders fine <i>fine blunders</i>
	/a/	'ga. fa # 'fi.nʌ blunder-the fine <i>the fine blunder</i>

The kinematic data were recorded at a sampling rate of 200 Hz using the electromagnetic articulography (EMA) system at the Institute of Phonetics in Munich (AG500, Carstens Medizinelektronik), at a sampling rate of 200 Hz. The system records articulatory movement over time by tracking, within an electromagnetic field, the position of sensors glued on various organs relevant for speech. In the current experiment, four sensors were placed on the tongue, spaced fairly equidistantly from tongue tip to tongue velar region: a

tongue tip (TT) sensor, an anterior tongue body (TB1) sensor, a posterior tongue body (TB2) sensor, and a velar/tongue dorsum (TD) sensor. Additional sensors were placed on upper and lower lips, and on the lower gums to determine jaw movement. Reference sensors were placed on the nose bridge, upper gums (maxilla) and behind the ears. All except those behind the ears were placed mid-sagittally. A palate trace was obtained for each speaker by sliding a sensor along the midline of the speaker's hard palate. The kinematic signals were filtered at 5 Hz cut-off frequency for the reference sensors, at 60 Hz for the TT sensor, and at 20 Hz for all other sensors. The data were corrected for head-movement on the basis of the reference sensors and rotated to each speaker's occlusal plane.

X.2.2 Measurement and analyses

For segmentation, we used the Matlab-based *Mview* software algorithm developed by Mark Tiede at Haskins Laboratories. Vowel articulation was determined on the basis of movement of the tongue body (TB1, TB2) sensors. The movement of the two sensors was similar so for convenience, the first vowel/glide in a sequence (/e/j/i/) was determined on the basis of TB1 movement, while the second (/a/) on the basis of TB2 movement.

Tongue body trajectory/movement was determined on the basis of the relative distance of the relevant sensor (TB1, TB2) to the palate trace, calculated as the minimal Euclidean distance of every time sample of the sensor to all sample points of the palate. Using information of the tongue position relative to the palate has the advantage of normalizing between speakers' different anatomies, compared to using the horizontal/vertical position coordinated. It also simplifies sensor position information to one dimension, by capturing

both horizontal and vertical sensor position in one single measurement reflecting constriction degree.

Kinematic events defining target achievement (Target), maximal constriction (Max) and release (Release) of a vowel were determined on the basis of changes in the tangential velocity profile of tongue body movement (cf. Figure 1). Maximal constriction was automatically detected as the minimum velocity between two velocity peaks (Peak 1, Peak 2) in the interval of interest determined by visual inspection, while target and release of the vowel were defined as the points in time at which velocity was 20% of the preceding and respectively following velocity peaks (cf. Figure 1). Maximal constriction was used for determining vowel-to-vowel timing, defined as the temporal lag between maximal constriction of the first vowel and maximal constriction of the second vowel: $\text{Timing} = \text{Max}_{V2} - \text{Max}_{V1}$. It was also the time-point at which the articulatory position of each vowel was determined in terms of distance of the relevant tongue body sensor to the palate trace. Duration of vowel constriction (Plateau duration) was determined as the interval between target and release of each vowel.

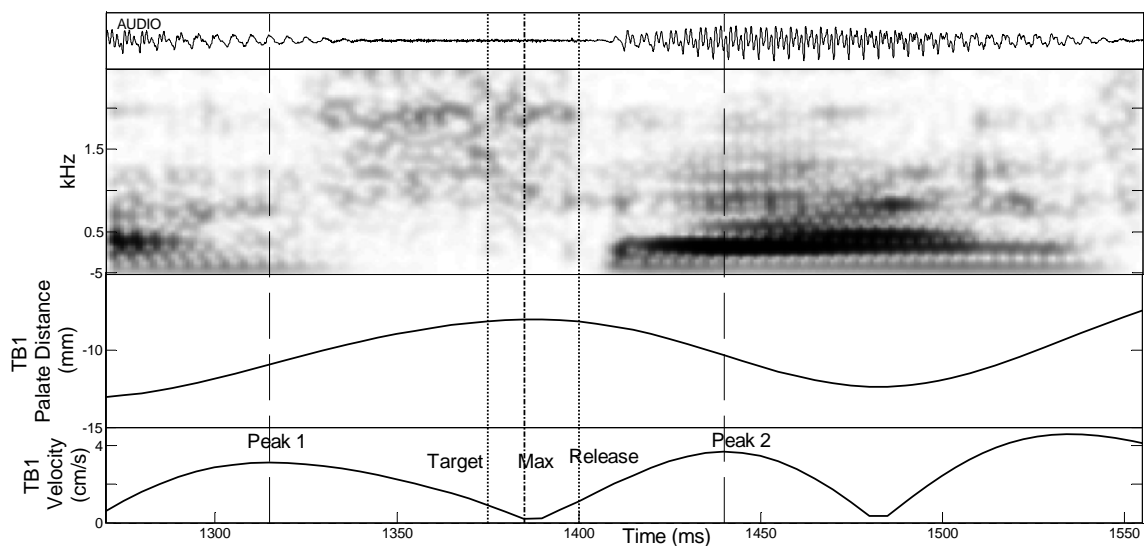


Figure 1. Example measurement of vowel articulatory movement for vowel /e/ in diphthong /ea/ in one repetition of word /kafea/. The visible part is the interval [fea].

For statistical analyses, mixed linear models were computed using the *lme4* package for R, with *p*-values being determined by comparing a model including the factor of interest with a model with no fixed factor (cf. Bates, 2010). This method circumvents the difficulty in estimating denominator degrees of freedom for mixed linear models. For post-hoc comparisons, *p*-values were determined using the Tukey adjusted contrast in the *multcomp* package for R, a package appropriate for mixed models (Hothorn et al. 2008).

Given the phonological analysis of diphthongs /ea/ and /ja/ and of the corresponding hiatus sequences, it is expected that the diphthongs differ both in timing (temporal overlap) and articulatory position from corresponding hiatus sequences. It is also expected that /ea/ as a complex nucleus will differ in its timing and degree of blending between its targets from /ja/.

X.2.3 Results

Figure 2 shows vowel-to-vowel timing across the two diphthongs and respective cross-word sequences. A mixed linear model on dependent variable: Timing Lag, with fixed factor: Item (levels: /ea/, /e_a/, /ja/, /i_a/, /j_a/) and random factor: Speaker, showed that vowel-to-vowel timing differed as a function of category ($F = 22.333, p < .001$). Post-hoc comparisons showed that diphthong /ea/ had significantly shorter vowel-to-vowel timing lags than all cross-word sequences ($p < .05$). The lags for diphthong /ea/ were also tendentially shorter than for diphthong /ja/ ($p = .07$). Diphthong /ja/ had comparable

timing lags to sequence /e_ 'a/ ($p = .9$), but significantly shorter than sequences /i_ 'a/, /j_ 'a/ ($p < .001$). Sequences /j_ 'a/ and /i_ 'a/ also showed longer vowel-to-vowel timing lags than sequence /e_ 'a/ ($p < .001$). The results thus show a clear difference in vowel-to-vowel timing between the cross-word sequences and the diphthongs, and also tendentially between the two diphthongs.

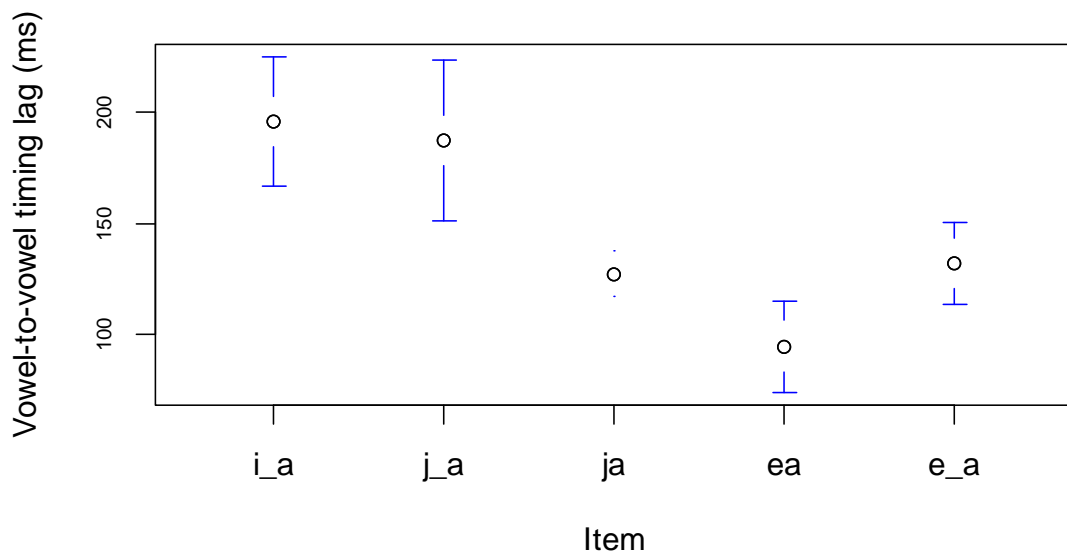


Figure 2: Vowel-to-vowel timing lag, representing the lag between maximal constriction of the two vowels.

Articulatory position during the first vowel (V1) in a sequence is shown in Figure 3, and during the second vowel (V2) in the sequence in Figure 4. For V1 distance to the palate, the statistical analysis (fixed factor: Item and random factor: Speaker) indicated a significant difference in V1 position as a function of item ($F = 81, p < .001$). Post-hoc comparisons showed that this effect was due to the high vowel/glide having a smaller

distance to the palate than the mid vowel ($p < .001$), but no other differences were observed. Thus, in terms of V1 position, a difference as a function of vowel height was observed, but not as a function of category (diphthong vs. cross-word sequence).

A statistical analysis on the position (distance to the palate) of V2, with fixed factor: Item and random factor: Speaker also indicated a significant difference between items ($F = 18.918, p < .001$). A post-hoc analysis further revealed that the position of /a/ was affected by category. Thus /a/ in both /'ea/ and /'ja/ was significantly closer to the palate (less open) ($p < .01$) than in the comparable cross-word vowel sequences (/e_'a/ and /i_'a/-/j_'a/ respectively). Furthermore, vowel /a/ in diphthong /'ea/ was significantly higher than in diphthong /'ja/ or singleton /a/ ($p < .001$). Position of vowel /a/ in diphthong /'ja/ and sequence /e_'a/ was comparable to position of singleton vowel /a/ ($p > .05$).

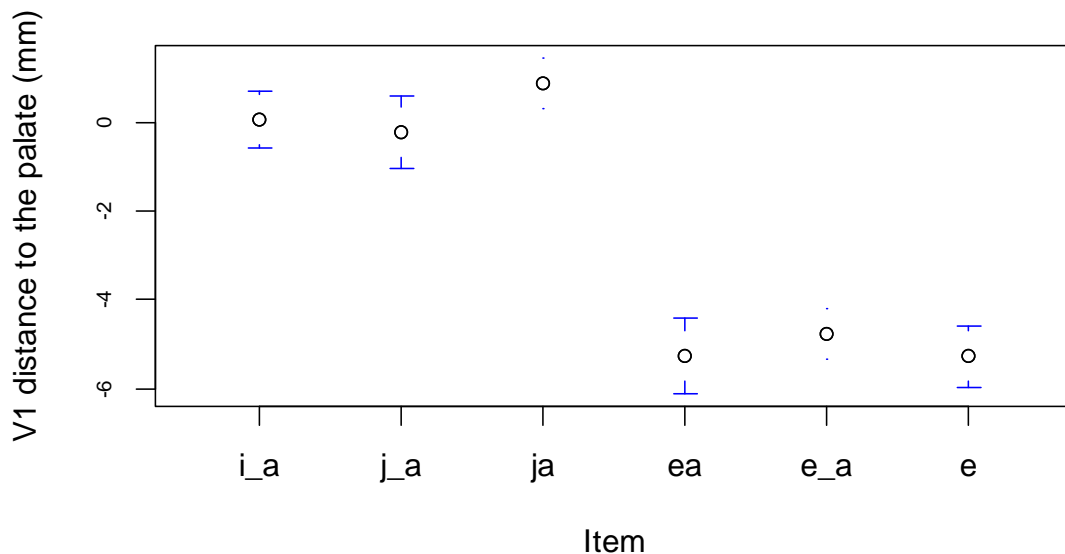


Figure 3. Position of V1.

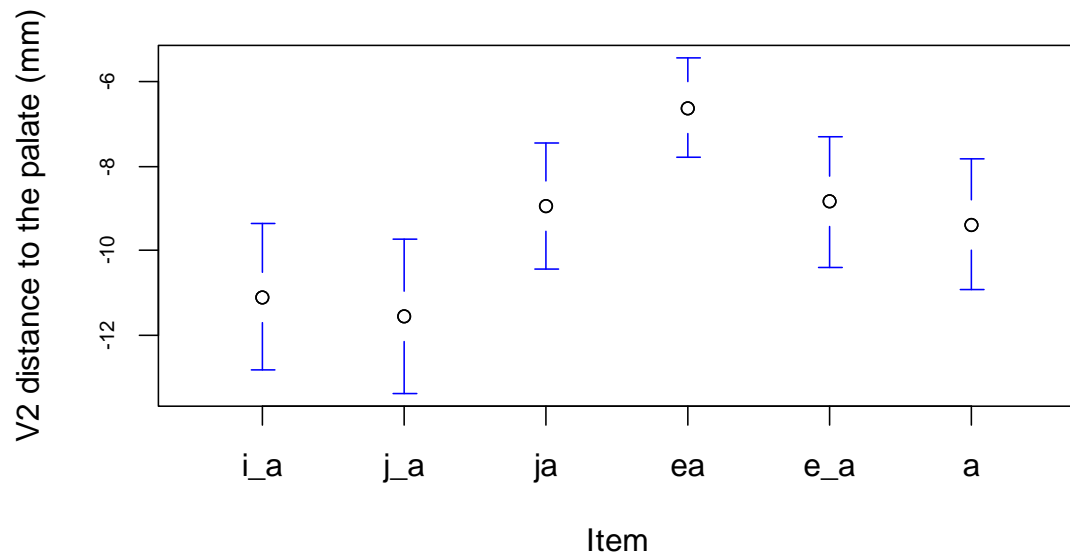


Figure 4. Position of V2.

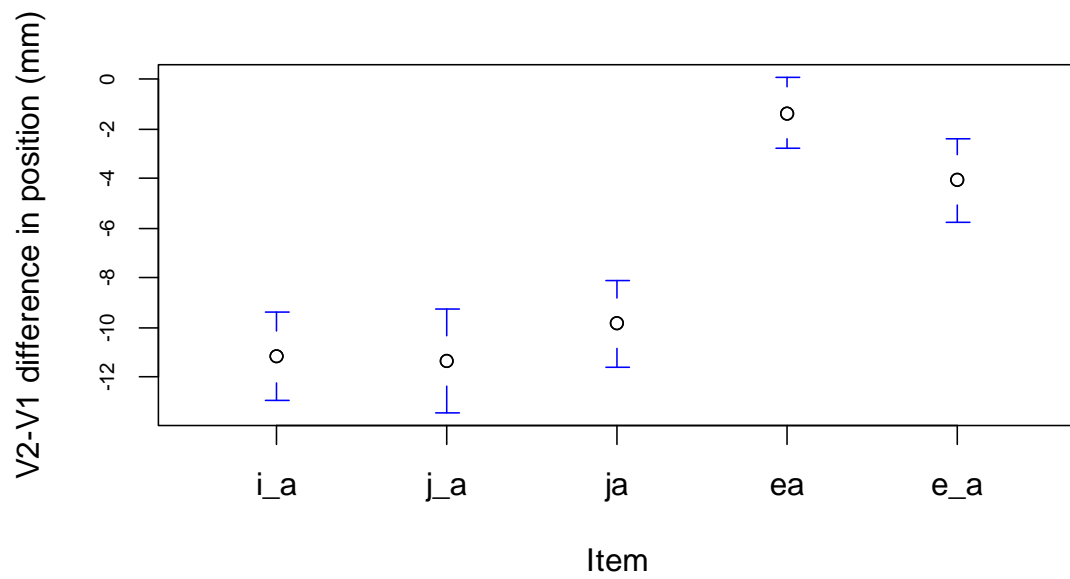


Figure 5. Difference in position between the two vowel targets at the point of maximum constriction for each.

An analysis of the difference in position between the two vowels (plotted in Figure 5) indicated that the categories changed by different amounts ($F=82.66$, $p < .001$), and that diphthong /'ea/ exhibited the least change in constriction degree from V1 to V2. The statistical analysis confirmed that position of the vowels in /'ea/ changed significantly less than in all other categories ($p < .01$).

X.2.4 Discussion

Vowels in diphthongs are timed closer together than in sequences, with the shortest timing lag observed for diphthong /'ea/. In terms of position, vowel /a/ in the two diphthongs had a significantly higher tongue position than in the comparable vowel sequences, suggesting a higher degree of coarticulation with the preceding vowel for the the diphthongs (expected given the closer timing). Furthermore, tongue position of /a/ in diphthong /'ea/ was higher than in diphthong /'ja/. If both diphthongs had been equally influenced by coarticulation, the opposite would have been expected: since glide /j/ had, as expected, a significant higher tongue position than vowel /e/, it should have determined a higher tongue position for /a/ in /'ja/ than in /'ea/. The observed pattern on the other hand suggests a higher degree of coarticulation for /'ea/ than for /'ja/: this is not surprising given the timing pattern (with a shorter vowel-to-vowel timing for /'ea/ than for /'ja/, and also given the phonological organization predicted for /'ea/ (a complex nucleus), compared to /'ja/ (phonologically assumed to be an onset-nucleus sequence).

The fact that tongue position changes the least during /'ea/ points to the same conclusion: diphthong /'ea/ is a complex vowel with two articulatory targets, but

overlapping and coarticulating to such an extent that the two targets blend into one, supporting its analysis as a monophthong.

The absence of difference in the first part of the vowel sequence for either the high (/ʲa/ - /j_ 'a/ - /i_ 'a/) or the mid (/ea/ - /e_ 'a/) stimuli is possibly due to experimental design limitations. Thus, /j/ in diphthong /ʲa/ was in word-initial position and in a stressed syllable, but was compared to /i/-/j/ in word-final unstressed position. Likewise, /e/ in /ea/ was in a stressed syllable, while /e/ in /e_ 'a/ and as a singleton was in unstressed position. Stress asymmetries may have thus obscured potential differences in position as a function of category (diphthong vs. hiatus). However, it may also be the case that in terms of degree of aperture (reflected in the distance to the palate measure) the categories do not differ articulatorily in the first part of the vowel, but only in the second. High vocalic elements in V1 (/i, j/) pattern together in having a smaller degree of aperture than mid vocalic elements, and this major difference between them affects the degree of aperture of V2.

X.3 Acoustic study – Method

The goal of the acoustic study is to compare the historical diphthongs /ja, je/ and the hiatus sequences /ia, ie, io/ in a stressed word-initial context: [#CjV] vs. [#CiV]. Chitoran and Hualde (2007) have shown that in several Romance languages, including Romanian, hiatus sequences of this kind vary in duration as a function of prosodic factors such as position in the word and proximity to stress. Sequences are longer in a word-initial than in a word-medial syllable. Sequences are also longer the closer they are to primary stress. They are thus longer in the immediately pretonic syllable (dia'mant 'diamond') than in a

pre-pretonic one (**di**apa'zon 'tuning fork'). The short /ia, ie/ sequences fall in the duration range of the historical /ja, je/ diphthongs, and it is proposed that variability along this acoustic parameter is exploited in production-perception in a potential phonologization of a diphthong-hiatus contrast.

In the present study we narrow down the comparison to the context where the difference between hiatus sequence and diphthong is enhanced, namely in word-initial position. We predict that diphthongs, in addition to being shorter than hiatus sequences, will show stronger coarticulation between the onset consonant and the nucleus vowel. The acoustic parameter of coarticulation we examined is the F2 value at the onset of the vocalic sequence and diphthong. In the hiatus sequence the nucleus vowel is assumed to be /i/. This should result in a higher F2 onset value compared to the diphthong. In the diphthong, the nucleus vowel is assumed to be /a/, the V2 of the diphthong, so F2 should have a lower value at onset.

X.3.1 Stimuli and participants

Data from four native speakers of Standard Romanian were recorded in a sound proof booth, using a Sony DAT recorder, sampled at 44 kHz. Eight words paired for diphthong-sequence conditions were used for spectral analysis. Seven repetitions of each word were recorded, embedded in the frame sentence "Spune ___ de trei ori", 'Say ___ three times', for a total of 70 words per speaker. The sentences were read in randomized order. The words were controlled for stress, number of syllables, the quality of the non-high vowel, the preceding consonant and, whenever possible, the following consonant.

Table 2. Target stimuli

Category	Item	Stimulus
Diphthong	/ˈja/	'fjart Λ boiled (fem.) 'pjatr Λ stone 'pjar Λ perish (subj.) 'pjaz Λ omen 'pjar Δ lose (subj.)
Hiatus sequence	/i'a/	fi'asko fiasco pi'astru piaster pi'anu piano

The acoustic analysis was done in Praat (Boersma & Weenink). The analysis consists of running formant tracks over the vocalic portion of each sequence and diphthong, and extracting the value of F2 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 Praat algorithm computes the LPC coefficients. The extracted F2 values were then compared to the corresponding spectrograms. Deviating values and the corresponding tokens were excluded from the statistical analysis.

X.3.2 Results

Our results, plotted in Figure 6, showed a difference between categories both in terms of duration and F2. Paired-samples t-tests showed that the diphthong was shorter than the hiatus ($t(3) = 3.367$, $p = 0.036$) and the diphthong had a lower F2 value at its onset than the hiatus sequence ($t(3) = 3.97$, $p = 0.029$). The scatterplot in Figure 7 shows that subjects differed in the extent to which they realized a contrast between the two categories. Speaker 2 shows the least difference in either duration or F2 values, while speaker 3 shows the greatest contrast in terms of duration, and speaker 1 in terms of F2.

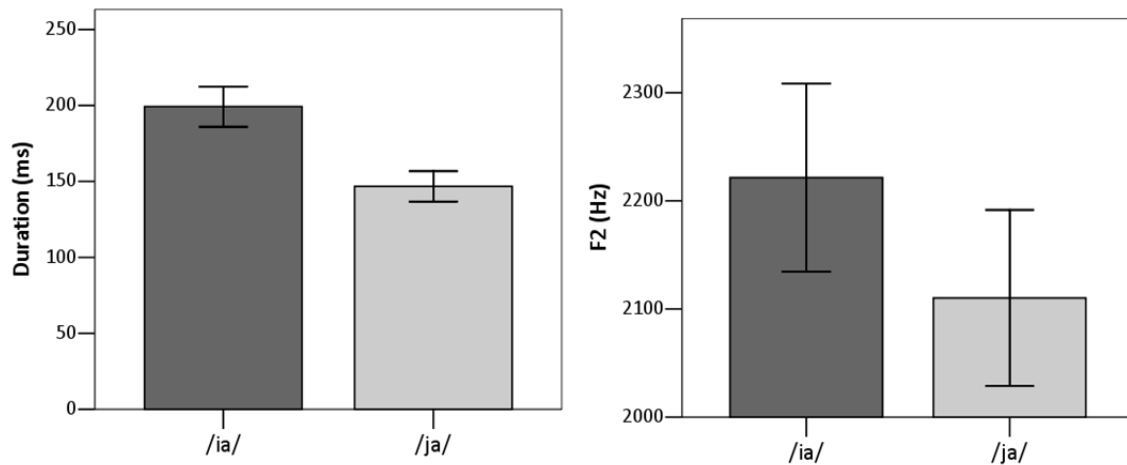


Figure 6. Mean and 1 standard deviation for acoustic duration (left) and F2 values at vowel onset time point (right) as a function of category.

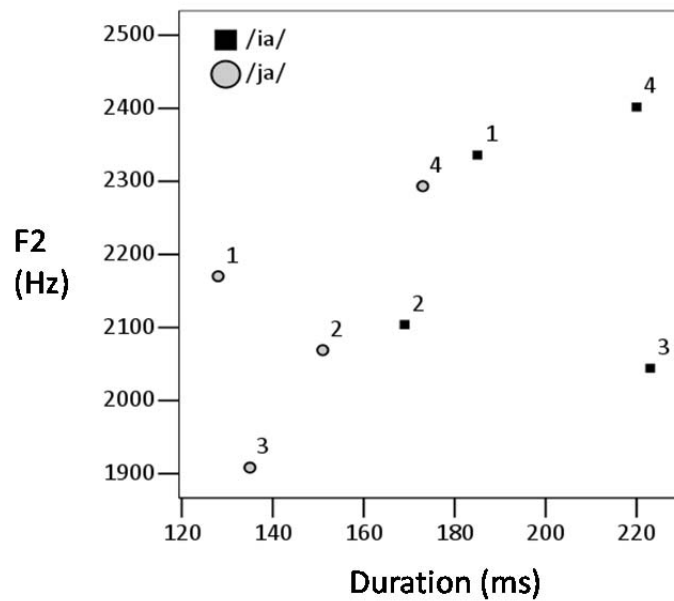


Figure 7. Scatterplot of mean durations by mean F2 as a function of speaker.

Figure 7 shows that all four speakers maintain a contrast between diphthong [ja] and hiatus sequence [i.a]. Speaker 2 shows a tendency towards merging the two categories, representative of inter-speaker variation.

X.3.3 Discussion

Confirming the pattern observed in the articulatory study, the acoustic study showed that the difference between diphthong and hiatus sequences was reflected in the duration of the sequence: diphthong /ja/ was acoustically significantly shorter than the hiatus /ia/ sequence. The acoustic study confirms that the difference in overlap (reflected in acoustic duration) is not restricted to diphthong /ja/ vs. a cross-word sequence (as we could observe in the articulatory study), but it also applies to within-word hiatus sequences (cf. also Marin, accepted, for a similar finding).

Although in the articulatory data there was no difference in position between /j/ in the diphthong and /j/, /i/ in a cross-word sequence, the better controlled acoustic data clearly indicated a qualitative difference between the two (indicative of a difference in articulatory position of the two). The lack of a difference in our articulatory study may thus have been due to confounding factors resulting from the experimental design (hiatus sequences were cross-word, with the first part of the sequence being word final and in an unstressed syllable, unlike the diphthong, which was word-initial and in a stressed syllable).

X.4 Conclusions

In our study we tested the phonological hypothesis that vocalic elements in Romanian can combine to form three different categories of syllabic nuclei: (i) single complex nucleus (diphthong /ea/); (ii) tautosyllabic onset-nucleus sequence (diphthong /ja/); (iii) heterosyllabic hiatus sequence (/i.a/, /e.a/). The three types of nuclei are predicted to show different degrees of coarticulation between the vocalic elements. The vocalic

elements sharing the same nucleus are predicted to show the highest degree of blending, and the hiatus sequence, the lowest. The results of our articulatory and acoustic studies support the hypothesis.

The predictions are borne out first by temporal measures. Articulatory vowel-to-vowel timing is found to be the shortest for /ea/, tendentially longer for /ja/ and /e_a/, and significantly longer for /i_a and j_a/. In terms of acoustic duration, hiatus /i.a/ is significantly longer than the diphthong /ja/.

Articulatory and acoustic measures of coarticulation both support the hypothesis. In the analysis of kinematic data, coarticulatory measures are based on: V1 distance to the palate, V2 distance to the palate at the point of maximum constriction, and the difference between the two points. The results for V1 show that in terms of degree of constriction/aperture, /i/ and /j/ are similar. They both have a higher degree of constriction than mid /e/. There is no difference in terms of the type of nucleus. However, measures of V2 distance to the palate show that /a/ in /ea/ has a higher degree of constriction than in /ja/, and higher still than in /j_a, i_a, e_a/. These results indicate that V2 /a/ is most strongly coarticulated with V1 in /ea/, somewhat less so in /ja/, and even less in cross-word and hiatus sequences. The diphthong /ea/ also has the smallest change in distance from the palate between V1 and V2.

In the acoustic data, coarticulation is measured at the C-V transition. The results indicate that the location of the lingual constriction at that moment is more anterior in /i.a/ than in /ja/. This acoustic result suggests that V1 also coarticulates with V2, but the locus of this coarticulation is earlier than the point of maximum constriction of V1. We conclude that the different phonological behavior of diphthongs and vowel sequences is

reflected in their acoustic and articulatory structure. The articulatory and acoustic data provided complementary information which, taken together, supports the presence of three different types of syllable nucleus organization in Romanian.

Acknowledgements

The second author's work has been supported by DFG Grant PO 1269/1-1 and the European Research Council under the European Union's Seventh Framework Programme (FP/2007-2013)/ERC Grant Agreement n. 283349-SCSPL. We thank Susanne Walzl for help with the articulatory data collection.

References

- Bates, D. (2010). *lme4: Mixed-effects modelling with R*. Springer.
- Boersma, P., & Weenink, D. (n.d.). *Praat: doing phonetics by computer*. Available for download from <http://www.praat.org/>.
- Chitoran, I. (2001). *The Phonology of Romanian: A Constraint-Based Approach*. Berlin, New York: Mouton de Gruyter.
- Chitoran, I. (2002a). The phonology and morphology of Romanian diphthongization. *Probus*, 14: 205–246.
- Chitoran, I. (2002b). A perception-production study of Romanian diphthongs and glide-vowel sequences. *Journal of the International Phonetic Association*, 32: 203–222.
- Chitoran, I., & Hualde, J. (2007). From hiatus to diphthong: the evolution of vowel sequences in Romance. *Phonology*, 24: 37-75.
- Hothorn, T., Bretz, F., Westfall, P. 2008. Simultaneous Inference in General Parametric Models. *Biomedical Journal*, 50: 346–363.
- Marin, S. (accepted). Romanian diphthongs /ea/ and /oa/: An articulatory comparison with /ja/ - /wa/ and with hiatus sequences. To appear in *Revista de Filología Románica*.