Romanian diphthongs /ea/ and /oa/: an articulatory comparison with /ja/ - /wa/ and with hiatus sequences

Stefania MARIN

Institute of Phonetics and Speech Processing
University of Munich
marin@phonetik.uni-muenchen.de

ABSTRACT
One notable feature in Romanian phonetics and phonology is the presence in its inventory of typologically rare diphthongs /ea/ and /oa/ («mid diphthongs»), which happen to contrast phonologically with both corresponding hiatus sequences /e.a/ - /o.a/, and with diphthongs /ja/and /wa/ («high diphthongs»). Structurally, on the basis of their phonotactic properties, mid diphthongs /ea/ and /oa/ have been assumed to form complex syllable nuclei, while high diphthongs /ja/ and /wa/ have been represented as onset-nucleus structures, and hiatus sequences as vowels in consecutive syllable nuclei. Little is however known on the articulatory properties characterizing the three-way contrast between mid and high diphthongs, and hiatus sequences.

Keywords: Romanian phonetics, diphtongs, hiatus.

[Recibido, septiembre 2013; aprobado, diciembre 2013]

Los diptongos rumanos /ea/ y /oa/: una comparación articulatoria con /ja/ - /wa/ y con secuencias de hiato

RESUMEN
Una característica destacable de la fonética y fonología rumanas es la presencia, en su inventario, de los diptongos tipológicamente raros /ea/ y /oa/ ("diptongos medios"), que contrastan fonológicamente tanto con las secuencias de hiatos /e.a/ - /o.a/ como con los diptongos /ja/ y /wa/ (diptongos altos). Estructuralmente, según sus propiedades fonotácticas, los diptongos medios /ea/ y /oa/ asumen la función de formar núcleos de sílaba complejos, mientras que los diptongos /ja/ y /wa/ se representan como estructuras iniciales de núcleo y las secuencias de hiato como vocales en núcleos de sílabas consecutivas. Se sabe poco, sin embargo, de las propiedades articulatorias que caracterizan esta triple comparación entre diptongos, medios, altos y secuencias de hiato.

Palabras clave: fonética rumana, diptongos, hiatos.

1. Introduction

One notable feature in Romanian phonetics and phonology is the presence in its inventory of typologically rare diphthongs /ea/ and /oa/ (henceforth «mid diphthongs»), which happen to contrast phonologically with both corresponding hiatus sequences /e.a/ - /o.a/, and with diphthongs /ja/and /wa/ (henceforth «high diphthongs»). Structurally, on the basis of their phonotactic properties (cf. Chitoran 2001; 2002a for a review) mid diphthongs /ea/ and /oa/ have been assumed to form complex syllable nuclei, while high diphthongs /ja/ and /wa/ have been represented as onset-nucleus structures, and hiatus sequences as vowels in consecutive syllable nuclei. Little is however known on the articulatory properties characterizing the three-way contrast between mid and high diphthongs, and hiatus sequences.

With respect to the contrast between mid and high diphthongs, previous experimental work on Romanian (Chitoran 2002b)has found that these diphthongs in the front condition (/ea/ - /ja/) were reliably distinguished perceptually and also differed in their acoustic properties. In terms of duration, /ja/ was longer than /ea/, providing empirical evidence for the structural analysis of /ja/ as an onset-nucleus sequence of two segments, and of /ea/ as a one-unit syllable nucleus. These diphthongs also differed in onset height, as reflected in F2 values (higher for /ja/)and F2 transition rates. The diphthongs in the back condition (/oa/-/wa/) on the other hand were undistinguishable from each other either acoustically or perceptually, suggesting a possible phonetic neutralization of the two categories, facilitated by the very low frequency of /wa/ (cf. Chitoran 2002b). This study suggests therefore that the difference between mid and high diphthongs, at least in the front condition, lies in the vocalic quality of its onset, a finding that matches these diphthongs' (orthographic) transcription, and justifies the terminology adopted here in designating them.

Regarding the contrast between mid diphthongs and hiatus sequences, previous work(Marin & Goldstein 2012; Marin 2007) has modeled this contrast in the front condition (/ea/ vs. /e.a/) in terms of differing timing relations between the articulatory movements for the two composing vowels. Extending general assumptions on the timing of vowels across syllable boundary (e.g. Öhman 1966; Smith 1995),hiatus /e.a/ was successfully modeled with single vowels /e/ and /a/ timed sequentially to each other. Diphthong /ea/ was modeled either with the single vowels overlapping for approximately 90% of their movement (Marin 2007) or with the single vowels timed synchronously and with the articulatory movement for vowel /a/ being given additional prominence relative to /e/ (Marin & Goldstein 2012). Modeling evidence suggests therefore that the difference between mid diphthongs and hiatus sequences lies primarily in how the two vocalic components are timed to each other.

The three categories (mid and high diphthongs, hiatus sequences) in both front and back conditions have not yet all been examined in a single study, nor have they all been compared to corresponding single vowels. The current study aims to fill this gap by providing an articulatory description of Romanian mid (/ea/, /oa/) and high (/ja/, /wa/) diphthongs, as well as of corresponding hiatus sequences (/e.a/, /o.a/) in relation to each other and to matching single vowels.

On the basis of previous findings, it is hypothesized that mid and high diphthongs differ mainly in actual target specifications and realizations, with a possible neutralization expected for the back diphthongs. At the articulatory level, these two diphthongs are expected to differ in tongue position at their onset. Mid diphthongs and hiatus sequences on the other hand are hypothesized to differ in relative timing of the two com-

posing vowels, with extended temporal overlap between the vowels (reflecting a more synchronous timing) expected for the diphthong compared to the hiatus. At the production level, this extended temporal overlap is expected to show up as an increased blend (intermediate tongue position values) between vowel targets for the diphthong compared to the hiatus. The difference in temporal overlap is also expected to show up as a duration difference between categories, with the less overlapped category (the hiatus) being longer than the more overlapped one (the diphthong).

2. Method

2.1. Stimuli and data acquisition

Articulatory and acoustic data were recorded from five Romanian speakers (three female) with no reported speech, hearing or language problems, and naïve as to the purposes of the experiment. The participants spoke standard Romanian without any pronounced dialectal features; they were familiarized with the list of utterances prior to data collection, and were instructed to speak at a comfortable rate. The stimuli were presented on a computer screen and the speakers were visually cued when to speak. Each utterance was repeated 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/, /oa/, /ja/ and /wa/, hiatus sequences /e.a/, /o.a/, as well as relevant single vowels /a/, /e/, /i/, /o/, /u/. The target words were embedded in carrier phrases: /'zik pu.'tsin/ *I say a little* for the front condition; /spu.'neam me.'rew/ *I was saying always*, for the back condition. The sentences for this experiment were interspersed with filler sentences constituting data sets for other experiments.

Category	Front condition	Back condition
Vowel /a/	/'sa.ra/ *proper name* /'da.mə/ *lady* /'bat/ *I beat*	/'ka.la/ *calla lily*
Mid vowel	/'se.ra/ *the greenhouse* /'te.mə/ *theme* /'pet/ *plastic bottle*	/'ko.la/ *Coca Cola*
High vowel	/'bit/ *byte*	/'kub/ *cube*
Mid diphthong	/'sea.ra/ *the evening* /'tea.mə/ *fear* /'bea.tə/ *drunk*, Fem. /'beat/ *drunk*, Masc.	/'koa.la/ *the sheet* /'koan.də/ *propername*
High diphthong	/'bja.tə/ *poor, Fem.*	/'kwan.tə/ *quantum*
Hiatus	/'se.at/*proper name* /'te.am/ I have you	/'ko.a.la/ *koala bear*

Table 1. Stimuli: intervals of interest are underlined.

The kinematic data were recorded at a sampling rate of 200 Hz using the AG500 (Carstens Medizinelektronik) electromagnetic articulography (EMA) system at the Munich Institute of Phonetics. The system records articulatory movement over time by tracking, within an electromagnetic field, the position of sensors glued at various points on the speaker's vocal tract. The acoustic data were recorded simultaneously at a sam-

pling rate of 32 kHz and synchronized with the kinematic signals during post-processing.

For the articulatory recordings, four sensors were placed on the tongue, spaced fairly equidistantly from tongue tip to tongue velar region: a tongue tip (TT) sensor (attached approximately 1 cm behind the actual tongue tip), an anterior tongue body (TB1) sensor, a posterior tongue body (TB2) sensor, and a tongue dorsum (TD) sensor. Additional sensors were placed on the upper and lower lips, and on the lower gums. Reference sensors were placed on the nose bridge, upper gums (maxilla), and behind the ears. All sensors, except for 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 a 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.

2.2. Analyses

Articulatory position and velocity data were extracted at five temporal landmarks determined on the basis of the acoustic signal: at the acoustic onset (0%) and offset (100%) of the vowel/diphthong/vowel sequence interval, and at 25%, 50% and 75% within this interval. The acoustic onset and offset timepoints also served in calculating the acoustic duration of the interval of interest. The measurement points were determined on the basis of the acoustic signal as it is impossible to identify on the kinematic signal vowel movement landmarks such as onset of movement, achievement of target etc., for consecutive vowels (cf. for example Harringtonet al. 2011; Hoole 1999, who also found it more reliable to use acoustically-defined landmarks to determine tongue properties during vowels even though in their case, the targets were singleton vowels in controlled symmetrical consonantal contexts).

Position of the tongue was defined as the relative distance (POSDIST) of the tongue sensors to the palate trace, calculated as the minimal Euclidean distance of every time sample of each sensor to all sample points on the palate. The tongue sensors used for the analysis were TB1, TB2, and TD, since these were the sensors placed in regions of the tongue relevant for forming vowel constrictions. For the back condition, where rounding is also a factor of potential interest, position and velocity information of the lip aperture (LA), defined as the Euclidean distance between the upper and lower lip sensors, was also extracted. Using information of the tongue position relative to the palate, and of the lips relative to each other has the advantage of normalizing between speakers' different anatomies, as well as to some extent between their variable positioning within the electromagnetic field.

For statistical analyses, the tongue position of each token at each temporal landmarkwas further quantified by calculating its Mahalanobis (M) distance to the centroids of appropriate single vowels (M_a, M_e, M_i, M_o, M_u) on the basis of the POSDIST of TB1, TB2, TD. The advantage of the Mahalanobis method over other distance calculation methods is that it takes into account the distribution's shape when determining the distance to it. These distances were calculated separately for each speaker and for the front/back sets, and so they served as a further means of normalization between speakers. Relative proximity (P) of each token to either /a/ or /e/ (P_{ae}), /a/ or /i/ (P_{ai}), /a/ or /o/ (P_{ao}) and /a/ or /u/ (P_{au}) was calculated as the difference between the two Mahalanobis distances on the logarithmic scale, using the following formulas: $P_{ae} = \log(M_a) –$

$\log(M_e)$, $P_{ai} = \log(M_a) - \log(M_i)$, $P_{ao} = \log(M_a) - \log(M_o)$, $P_{au} = \log(M_a) - \log(M_u)$. When P is 0, the token is equidistant between the two respective singleton vowels; when it is negative, the token is closer to /a/, and when it is positive, the token is closer to /e/, /i/, /o/, or /u/ respectively. P_{ae} was used to compare /ea/ and /e.a/ to the single vowels /e/ and /a/ and to each other; P_{ai} was used to compare /ja/ (and /ea/) to the single vowels /i/ and /a/ (and to each other); P_{ao} was used to compare /oa/ and /o.a/ to the single vowels /o/ and /a/ and to each other; and finally P_{au} was used to compare /wa/ (and /oa/) to the single vowels /u/ and /a/ (and to each other). This measure has the advantage not only of reducing the tongue position data to one value per token and landmark, but also of directly assessing relative similarity of the complex categories to the simplex categories.

To further assess (absence of) tongue and lip movement over time (i.e. gliding from one vowel target to another), velocity profiles were also analyzed. For tongue movement, a global velocity measure (MeanVel) was used, which was the average of the individual TB1, TB2, TD velocities. For lip movement, velocity of the lip aperture was used.

The difference between mid and high diphthongs has been hypothesized to be instantiated mainly in terms of targets, and hence in terms of tongue (and lips) position differences. Relative proximity measures are expected to capture such differences: for example, at onset /ja/ but not /ea/ is expected to be similar (closer) to /i/. On the other hand, the contrast between mid diphthongs and hiatus sequences is expected to be instantiated in terms of different timing relations between the composing vowels. Because articulatory temporal landmarks (such as onset/offset of movement, achievement of target) cannot be reliably measured for consecutive vowels on the kinematic signal, relative timing between the vowels in the complex categories (diphthongs and hiatus) cannot be directly determined. Instead, timing differences must be inferred indirectly from analyzing tongue positions: if two vowels that control the same articulator (tongue body/dorsum or the lips) overlap extensively (as is hypothesized for mid diphthongs), it is expected that the resulting tongue position would be somewhat intermediate between the tongue positions when no such overlap exists. Relative proximity measures should therefore also capture timing differences in as much as they can reveal less extreme/more intermediate articulatory positions for the diphthongs than for the hiatus over time. In terms of movement patterns, if the two vowel targets overlap extensively, any movement/gliding from one target to the next should occur earlier than if the targets overlap less, and hence any velocity increases should be observed earlier in the extensively overlapped category (the mid diphthong) than in the less overlapped category (the hiatus).

For the statistical analyses, mixed linear models were computed using the *lme4* package for R(Bates 2010), with speaker and word as random factors. The advantage of a mixed linear model is not only that it allows for crossed random effects, but it is also robust for unequal sample sizes(Baayen et al. 2008). One known difficulty with mixed models is the calculation of the denominator degrees of freedom, and hence of *p*-values in the customary way. To determine *p*-values for the main effect, a model including the fixed factor of interest was compared with the same model with no fixed factor, on the assumption that a significant difference between the two models indicates that the fixed factor contributes significantly to the model (cf. Bates 2010). The *p*-value thus obtained is reported along with the F-value of the mixed linear model; because denominator degrees of freedom are difficult to estimate, and furthermore, no longer play a role in calculating the *p*-values, they are not reported. For post hoc comparisons, the *p*-values

were determined using the Tukey contrast in the *multcomp* package (Hothornet al. 2008).

3. Results

3.1. Duration

The hypothesized difference in temporal overlap between mid diphthongs and hiatus sequences is expected to be reflected in a durational difference between the two categories. Additionally, based on previous results (Chitoran 2002b), it is expected that mid and high diphthongs also differ in duration, at least in the front condition. Finally, if mid diphthongs are composed of two vowels timed (almost) synchronously, they should be comparable in duration to a matching single vowel, i.e. the two extensively overlapping vowel targets should not take more time than a comparable single vowel target. Duration means and standard deviations are reported in Table 2.

Mixed linear models with fixed factor: Duration and random factors: Speaker and Word, showed that the categories differed significantly for both the front ($F = 33.3$, $p < .001$), and back conditions ($F = 63.7$, $p < .001$).Post-hoc analyses showed that in both conditions, the vowels in hiatus were significantly longer than either mid or high diphthongs ($p < .001$), while the two diphthong types had comparable durations ($p > 0.7$). In the front condition, the diphthongs were comparable in duration to vowel /a/*(*$p > 0.3$), butlonger than /e/ or /i/ ($p < .001$). In the back condition, the diphthongs were longer than all single vowels ($p < .001$). It must be noted that although overall vowel /a/ in the front condition was longer than /a/ in the back condition, the difference was not statistically significant ($p = 0.13$).

Front	Mean	Std. Deviation	Back	Mean	Std. Deviation
a	121.10	28.13	a	96.38	18.52
e	96.16	25.14	o	86.93	20.04
i	73.83	11.96	u	78.84	29.49
e.a	193.78	67.06	o.a	181.31	47.37
ea	128.78	33.89	oa	120.86	22.42
ja	143.25	23.97	wa	132.61	37.68

Table 2. Mean durations and standard deviations.

3.2. Position

Tongue (POSDIST) and lips (LA) positions for the vowel categories over time are shown in Figure 1. Overall, it can be observed that in the front condition the categories are well differentiated on the basis of the position of the three tongue sensors (the three sensors exhibit similar patterns, with the observation that the differences are less pronounced on the TD measure). Impressionistically, hiatus /e.a/ and diphthong /ea/ start with tongue positions similar to those for vowel /e/, by 25% the tongue is already lowered towards /a/ for diphthong /ea/, and at 50% the tongue is intermediate between vowels /e/ and /a/ for both diphthong and hiatus. At 75% they approach the position for vowel /a/, but even at 100%, the tongue body position for the diphthong (TB1, TB2) is

still distinct from that of /a/. For /ja/, tongue position remains similar to /i/ till half-way in the diphthong, when it is roughly equidistant between /i/ and /a/, and then approaches an /a/-like position. Diphthongs /ea/ and /ja/ differ both in position at starting point (more /i/-like for /ja/ and /e/-like for /ea/), and in how fast they start moving from /e/ or /i/ towards /a/: while at 25% /ea/ is already distinct from /e/, /ja/ is at that landmark still very similar to /i/.

In the back condition, the distinction in terms of tongue position between categories is much smaller overall (within a 4 mm range for /u/ vs. /a/), with observable differences restricted mostly to position of TB2. Diphthong /oa/ and hiatus /o.a/ start with TB2 positions intermediate between /o/ and /a/, and by the 25% landmark they already have positions very close to /a/. Diphthong /wa/ starts with a TB2 position close to /u/, is fairly equidistant between /u/ and /a/ at 25%, and very close to /a/ at mid-point.Diphthongs /oa/ and /wa/ are further apart at onset and by 50% they reach very similar tongue positions. In terms of lip aperture, /oa/, /o.a/ and /wa/ start with lips closer together, typical for rounding (in the vicinity of /o/, /u/), with /oa/ exhibiting the most extreme lip closure, and by the 75% temporal landmark the lips are more open, although only for /o.a/ do they impressionistically reach a position similar to /a/, while /oa/ and /wa/ have more intermediate positions between rounded /o/, /u/ and unrounded /a/.

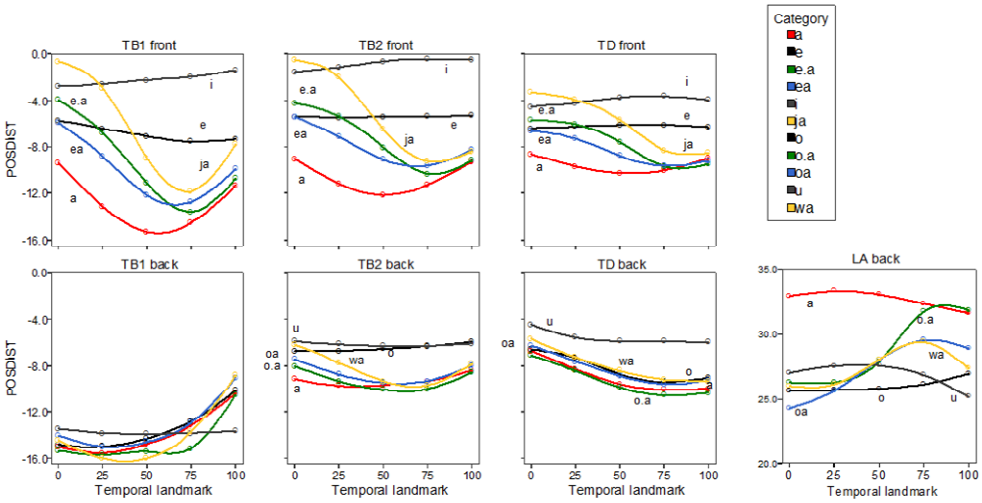

Figure 1. Articulatory positions for vowel categories at five temporal landmarks, in the front (top) and back (bottom) conditions. The values represent averages across speakers and words.

For statistical modeling, to limit the number of tests to the extent possible, proximity indices calculated on the basis of the position of the tongue sensors were used rather than the individual tongue positions. Recall that by this measure, the global tongue position of a given token at a given time point was evaluated in relation to the average tongue position of a single vowel category (see method section for details). Thus, for example the proximity index P_{ae} of a given /ea/ token at a given time point evaluated whether the respective token was more similar in terms of tongue position to vowel /e/ (indicated by a positive value), to vowel /a/ (indicated by a negative value) or equidis-

tant between the two (indicated by a value of zero). Box plots of these proximity indices are shown in Figure 2 for the front condition and Figure 3 for the back condition.For the statistical analyses, mixed linear models on the various proximity measures were computed with fixed factor: Category, and random factors: Speaker and Word. Main effects of these analyses are reported in Table 3.

Measure	Temporal landmark				
	0	25	50	75	100
P_{ae}	F = 12.1**	F = 33.3**	F = 103**	F = 174**	F = 36.9**
P_{ai}	F = 39.7**	F = 66.9**	F = 105**	F = 266**	F = 61.5**
P_{ao}	F = 47.5**	F = 43.5**	F = 39.1**	F = 20.6**	F = 43.8**
P_{au}	F = 34.7**	F = 34.0**	F = 45.3**	F = 73.4**	F = 137**
LA_{back}	F = 5.50*	F = 4.94*	F = 2.93	F = 2.99	F = 3.97
$TB2_{back}$	F = 13.7**	F = 19.3**	F = 27.5**	F = 31.1**	F = 15.1**

Table 3. Statistical results for the mixed linear models with dependent variable: Proximity measures/Lip aperture (back condition)/TB2 (back condition), with fixed factor: Category, and random factors: Speaker and Word. ** indicates$p \leq .001$, * indicates $p \leq .01$; otherwise, $p > .05$.

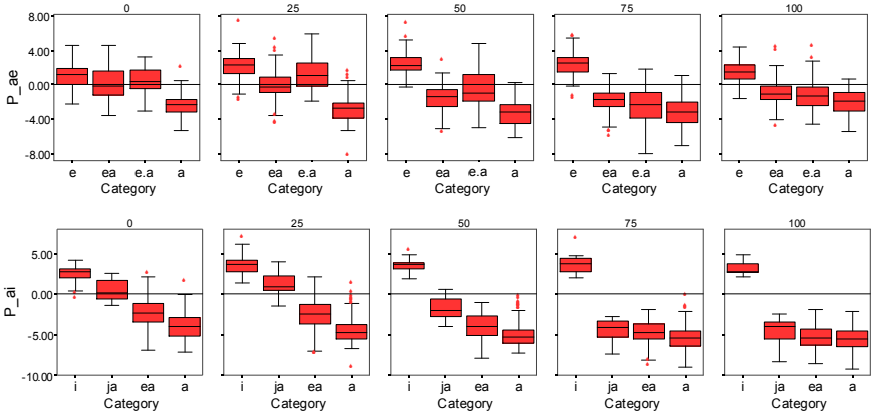

Figure 2. Box plots of relative proximity indices for the front condition. P_{ae} measures relative proximity to vowels /a/ and /e/, and P_{ai} measures relative proximity to vowels /a/ and /i/. Negative values indicate that a token was closer (more similar) to /a/ than to /e/ or /i/, while values around zero indicate that a token was equidistant between categories.

We used the P_{ae} measure to compare /ea/ and /e.a/ to each other and to the respective single vowels at each temporal landmark. The main effect for this proximity index was significant at all time points (cf. Table 3). Post hoc tests showed, as expected, that the P_{ae} index for single vowels differed at all temporal landmarks ($p < .001$), with /e/ tokens being closer to the /e/ category and /a/-tokens to the /a/ category. Diphthong /ea/ started with a similar tongue position to /e/ at onset, indicated by statistically comparable proximity indices ($p = .6$), but was already different from it at 25% ($p < .001$),

where a median around 0 suggests that the diphthong's tongue position was equidistant between the two end vowels. Although starting to be more similar (closer in distance) to /a/ quite early (at the 25% temporal landmark), /ea/ differed from /a/ at all temporal landmarks($p < .001$). This quantitatively confirms the observation from Figure 1 that the tongue for /ea/ has an intermediate position between /e/ and /a/ quite early on and that it maintains this intermediate position between categories throughout: it moves from being closer to /e/ at onset to being closer to /a/ at offset, without however being statistically similar to it.

Hiatus /e.a/ was similar to /e/ at onset and 25% ($p > .5$), differed from both /e/ and /a/ at 50% and 75% ($p < .001$), and was similar to /a/ at 100% ($p = .1$). Unlike the diphthong, the hiatus became dissimilar from /e/ only at the 50% time point, and at offset reached an /a/-like position. The diphthong and hiatus also differed from each other at the 25% and 50% time points ($p < .03$).

To assess diphthong /ja/ in relation to vowels /i/ and /a/, the proximity P_{ai} index was used. Again, as expected, the single vowels differed on this measure at all temporal landmarks ($p < .001$). At onset and at the 25% landmarks, diphthong /ja/ was closer to /i/ but different from it ($p < .05$). By 50%, it was closer to /a/ (negative proximity values), but still different from it ($p < .001$), a difference maintained at the 75% landmark as well ($p < .001$). Finally, at offset /ja/ and /a/ were statistically undistinguishable from each other. On this measure, /ea/ and /ja/ differed from onset to the 50% landmark ($p < .001$), and at the 75% and 100% temporal landmarks they no longer differed statistically ($p > .7$), corroborating the observations made on the basis of Figure 1.Likewise, the observation that /ja/ maintained for longer an /i/-like position than /ea/ maintained an /e/-like position could be inferred from the fact that at 25% /ja/ was still closer to /i/ than to /a/ (positive proximity values), while /ea/ was equidistant between /e/ and /a/ (values around zero). In summary, the complex categories in the front condition /ea/, /e.a/ and /ja/ could be distinguished in terms of their global tongue position in relation to corresponding single vowels. Diphthong /ea/ and hiatus /e.a/ differed in time point at which gliding from a vowel position to another occurred, as well as in end positions achieved. Diphthongs /ea/ and /ja/ differed in timing of gliding from one vowel position to another and in tongue position at onset through midpoint.

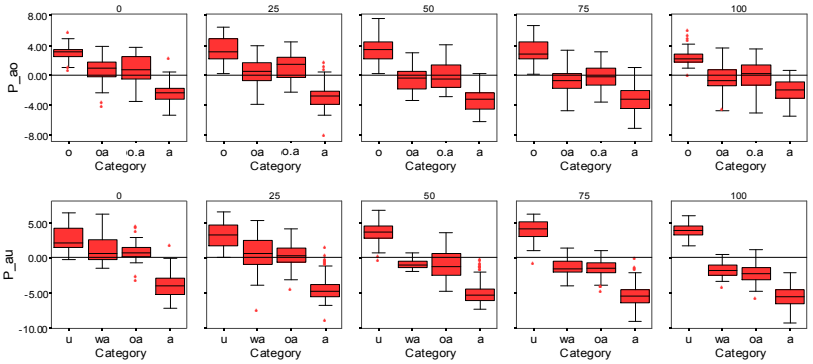

Figure 3. Box plots of relative proximity indices for the front condition. P_{ao} measures relative proximity to vowels /a/ and /o/, and P_{au} measures relative proximity to vowels /a/ and /u/. Negative values indicate that a token was closer (more similar) to /a/ than to /o/ or /u/, while values around zero indicate that a token was equidistant between categories.

For the back context, the P_{ao} measure was used to compare diphthong /oa/ to hiatus /o.a/ and to vowels /o/ and /a/. The single vowels differed from each other at all temporal landmarks ($p < .001$). Diphthong /oa/ and hiatus /o.a/ differed from both single vowels at all temporal landmarks ($p \leq .001$), and did not differ from each other ($p > .8$). They both started closer to, but significantly different from /o/ (positive proximity values), and by 50% they were closer to, but different from /a/ (negative proximity values). On the P_{au} measure, comparing vowels /a/ and /u/ and diphthongs /oa/ and /wa/, the single vowels differed from each other at all landmarks ($p < .001$). Diphthong /wa/ differed from both vowels at all landmarks ($p \leq .002$), and diphthongs /wa/ and /oa/ were statistically comparable to each other ($p > .7$). On the proximity measures, the diphthongs and hiatus in the back condition were all different from the corresponding single vowels, and undistinguishable from each other.

Because Figure 1 suggested that most of the difference between the complex categories (/oa/, /o.a/, /wa/) in the back condition was in terms of TB2 position, they were also compared on the basis of this variable alone (rather than global tongue position). Statistically, /oa/ and /o.a/ did not differ on this measure at any time point ($p > .3$), while /oa/ and /wa/ differed at onset ($p = .02$), with TB2 for /wa/ being higher (closer to the palate) than for /oa/. The significant main effect for TB2 at all temporal landmarks (Table 3) was due to vowel /a/ differing from both /o/ and /u/ ($p < .001$).

In terms of lip position, as assessed by the lip aperture measure, a main effect could only be observed for the onset and25% temporal landmarks. Post-hoc tests at the time points where the main effect was significant showed that /o/ and /u/ expectedly differed from /a/ ($p < .03$), and did not differ from each other ($p > .9$). Diphthongs /oa/ and /wa/ and hiatus /o.a/ did not differ from each other ($p > .9$), or from the round vowels ($p > .9$), but they differed from vowel /a/ ($p < .01$). The back complex categories therefore could not be robustly distinguished between each other (or from the round single vowels) in terms of lip position. Interestingly, beyond the 25% landmark no categories differed significantly in terms of lip aperture: given that when looking strictly at means across speakers and words, differences between categories were apparent at all temporal landmarks (cf. Figure 1), it may be inferred that large speaker/item variability on this measure rendered these differences statistically weak. In summary, diphthong /oa/ and hiatus /o.a/ could not be distinguished in terms of tongue or lip positions, nor in terms of timing of gliding from one vowel position to another. Diphthongs /oa/ and /wa/ could only be distinguished at onset on the basis of TB2 position, but not on the basis of global tongue position or lip aperture.

3.3. Velocity

Proximity indices could offer indirect evidence about the timing of the two vowels composing diphthongs and hiatus sequences, by indicating at what time point the respective diphthong/hiatus moved from being closer to one end vowel to being closer to the other end vowel (and also at what time points it was intermediate between the two). Velocity measures can further shed light on this issue, on the premise that increasing velocity indicates movement from one position to another, while a decrease or plateau in velocity profile indicates a stationary state of the articulators. Thus, if such velocity increase is observed comparatively earlier, it may be inferred that gliding from one vowel target to the other occurs earlier in time, and hence that the two targets overlap more (i.e. they are timed more synchronously). This measure is of interest for the com-

plex categories (diphthongs and hiatus) and comparisons are made within one category across time points.

Velocity profiles for each complex category are shown in Figure 4 for tongue movement, and Figure 5 for lip movement. For the statistical analyses, mixed linear models for each complex category on MeanVel and LA_Vel were computed with fixed factor: Temporal Landmark, and random factors: Speaker and Word. Main effects of these analyses are reported in Table 4.

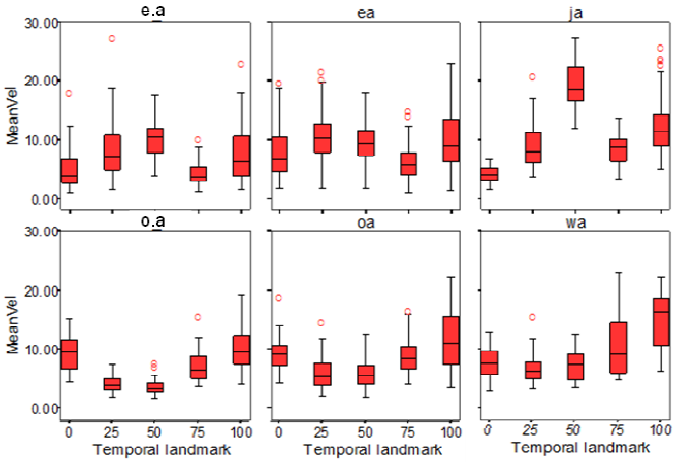

Figure 4. Box plots of tongue velocity profiles. MeanVel is the average velocity across TB1, TB2, and TD.

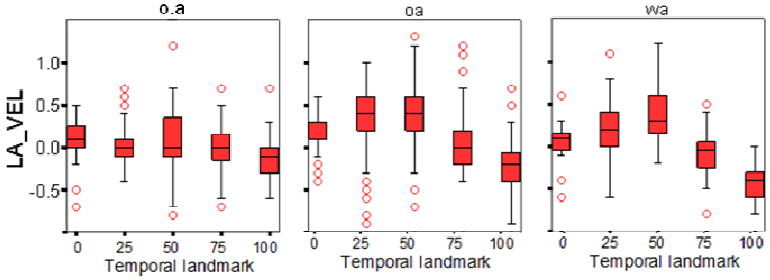

Figure 5. Box plots of lip aperture velocity profiles (LA_Vel).

Measure	Temporal landmark					
	/ea/	/e.a/	/ja/	/oa/	/o.a/	/wa/
MeanVel	F = 26.7	F = 22.2	F = 81.7	F = 15.3	F = 4.7	F = 31.9
LA_Vel	n/a			F = 35.8	F = 8.9	F = 21.4

Table 4. Statistical results for the mixed linear models with dependent variable: MeanVel/LA_Vel (back condition), with fixed factor: Landmark, and random factors: Speaker and Word. For all tests, $p \leq .001$.

Post-hoc analyses showed that for diphthong /ea/ tongue velocity significantly increased between onset and 25% ($p < .001$), remained constant between 25% and 50% ($p = .4$), and then decreased after 50% ($p < .001$), suggesting that the gliding from one tongue position to another occurs quite early, by 25% in the diphthong interval. For /e.a/ velocity significantly increasedfrom onset through 50% ($p < .05$), and then significantly decreased ($p < .001$), suggesting a later gliding. Likewise, for diphthong /ja/, there was a significant velocity increase till 50% ($p < .001$), followed by a significant velocity decrease ($p < .001$). Diphthong /ja/ and hiatus /e.a/ had similar velocity patterns over time, with the observation that the change in velocity for /ja/ was more pronounced than for /e.a/ (or /ea/), reflecting the greater change in tongue position required in gliding from an /i/-like target to an /a/-like target, compared to an /e/-/a/ gliding. It must also be noted that a significant velocity increase ($p < .001$) was again observed for all categories between the 75% and 100% landmarks, associated with the tongue movement for the following consonant.

For the back categories, tongue velocity did not significantly increase in the first part of the interval: for /o.a/ and /wa/, no significant differences could be observed between successive temporal landmarks from onset to 50% ($p > .4$), while for /ea/ a significant decrease in velocity was observed from onset to 25% ($p < .001$). This may be due to the smaller tongue position changes observed for the back compared to the front complex categories, changes that may presumably be achieved without large tongue movements, and hence without increases in velocity. In terms of tongue velocity changes, no meaningful difference was apparent between /oa/, /o.a/ and /wa/. The velocity pattern for TB2 only was qualitatively similar to the global tongue velocity.

On the other hand, lip velocity for the back complex categories mirrors the pattern observed for the tongue movement of the front categories. For diphthong /oa/, lip velocity significantly increased between onset and 25% ($p < .001$) suggesting an increased lip movement at this time point, indicative of a gliding movement from a rounded to an unrounded vowel. This was followed by a constant velocity between 25% and 50% ($p = .8$), and then by a decrease in velocity ($p < .001$). For hiatus /e.a/, a constant velocity between onset and 25% ($p = 1$) was followed by increased lip movement between the 25% and 50% landmarks ($p < .001$), and a decrease between 75% and 100% ($p = .002$). Diphthong /wa/ patterned similarly to /o.a./: a constant velocity to 25% ($p = .3$), followed by a significant increase between 25% and 50% ($p < .001$), and a significant decrease afterwards ($p < .001$). On this measure, /o.a/ and /wa/ patterned together showing a later lip movement associated with unrounding compared to diphthong /oa/. Thus although in terms of lip positions, the differences between the complex categories were too small/variable to be robust statistically, the lip velocity pattern suggests that /oa/ is characterized by an early gliding from a rounded to an unrounded vowel, distinguishing it from both hiatus /o.a/ and diphthong /wa/, which are characterized by a later unrounding. Furthermore, this matches the tongue velocity pattern observed for the front category.

4. Discussion

The duration analysis showed that diphthongs were significantly shorter than hiatus sequences, suggesting that the two categories exhibit different degrees of temporal

overlap between the composing vowels.[1]Furthermore, diphthongs in the front condition were comparable to vowel /a/, suggesting that the diphthong components at least in the case of /ea/, /ja/ were timed more or less synchronously to each other since no additional length beyond a matching single vowel was added. In the current data, the two diphthong types in the front condition had similar durations, differing from the pattern reported by Chitoran (2002b), but matching earlier results obtained by Rosetti et al. (1955).The discrepancy of results may be due to the fact that Chitoran included in the /ja/ category not only words where the orthographic sequence «ia» was invariably pronounced as diphthong /ja/, but also words where «ia» could be pronounced as either diphthong /ja/ or hiatus /i.a/.

For the articulatory analyses, the patterns of results are different for the front vs. back condition. In the front condition, in terms of tongue position, /ea/ and /e.a/ start to diverge at the 25% temporal landmark within the acoustic interval, with tongue position for /ea/ being already equidistant between the position typical for /e/ and that for /a/, while /e.a/ still maintains a tongue position more similar to /e/. Also, while /e.a/ reaches a position typical for /a/ by the acoustic offset, /ea/ is still different from /a/ at that landmark. Diphthong /ea/ tongue position – somewhat intermediate between that of a typical /e/ and of a typical /a/ at all timepoints, except onset of the interval, suggests an extensive co-production (blending) between the two vowel targets, explainable if the two articulatory movements extensively overlap during the diphthong production. In the case of hiatus, if the two targets overlap less, their targets blend (are intermediate) at fewer of the landmarks, and additionally beginning and end points are similar to typical corresponding single vowels. A similar picture emerges from the velocity analysis, where an earlier tongue movement increase for /ea/ suggests that the gliding between targets occurs earlier for /ea/ than for /e.a/. It can therefore be concluded that the main difference between diphthong /ea/ and hiatus /e.a/ is in the degree of temporal overlap between the two targets. This can be understood as a reflection at production level of different timing specifications between categories, an analysis compatible with the previous articulatory model proposed fordistinguishing /ea/ and /e.a/ in terms of synchronous vs. sequential timing between the composing vowel targets (Marin & Goldstein 2012).

Diphthongs /ea/ and /ja/differed in tongue position throughout the 75% landmark, and they also differed in time point at which they glided from one vowel endpoint to another. Thus, while at 25% /ea/ was already equidistant between /e/ and /a/ vowel categories, /ja/ was still more similar to /i/ than to /a/. This difference in timing of gliding is also reflected in the velocity patterns with an earlier velocity increase (reflecting increased tongue movement) for /ea/ than for /ja/. The articulatory results suggest that /ea/ and/ja/ differ not only in articulatory targets for most of the acoustic interval, but also in timing of the gliding between targets within that interval. Phonologically, /ea/ has been shown to function as a complex nucleus, while /ja/ as an onset-nucleus sequence (Chitoran 2001; 2002a); the difference in timing observed in the current data may be a reflection of this structural difference between the two diphthong types.

---

[1] The same duration result would obtain if the vowels in the diphthong would be truncated compared to singleton vowels/vowels in hiatus. The tongue position analyses, which for example show intermediate positions between /e/ and /a/ for /ea/ earlier than for /e.a/, suggest that varying degree of overlap rather than truncation is responsible for the duration difference.

In contrast to the front condition, the articulatory differences between categories in the back condition were less pronounced. Thus, while both tongue position and velocity patterns distinguished diphthong /ea/ from both hiatus /e.a/ and from diphthong /ja/, neither distinguished /oa/ from /o.a/ or from /wa/. Diphthong /oa/ and hiatus /o.a/ were distinguished exclusively in terms of duration (on which measure they patterned like front /ea/-/e.a/), and in terms of velocity patterns for the lips' movement (with an earlier velocity increase, suggesting earlier unrounding, for /oa/ than for /o.a/). The small articulatory differences in the back condition are compatible with the view that the second (unrounded) vowel target is timed earlier in the diphthong condition, suggesting hence an increased temporal overlap between the two targets in the diphthong compared to the hiatus (which can also explain the observed durational difference between the two). The fact that no difference in tongue body position was observed may be due to the fact that the lingual vowel targets for /o/ and /a/ are overall much closer to each other than the targets for /e/ and /a/, and hence even if more temporal blending may have been present for one category than for the other, it did not result in robustly different tongue positions.

Likewise, the back diphthongs /oa/ and /wa/ could only be distinguished by tongue body (TB2) position differences at acoustic onset (a higher tongue body for /wa/ than /oa/), and by the velocity pattern in the movement of the lips, which suggested an earlier unrounding for /oa/ compared to /wa/. These results mirror to some extent the distinctions observed in the front condition: a distinction in tongue position in the first part of the diphthong (albeit reduced to one sensor and one time point for the back diphthongs), and an earlier gliding for the mid than for the high diphthong (albeit only observed for the lip movement).These articulatory differences in the back condition are much subtler than those observed in the front condition, which explains why previous acoustic and perceptual results suggested a phonetic neutralization between the back diphthongs. While speakers may still produce the two diphthongs slightly differently, the small articulatory differences may not have large enough acoustic consequences to be robustly captured in terms of formant differences, and may not provide reliable cues for perception. Importantly, the current results revealed however that speakers do produce /wa/ and /oa/ differently (in spite of the low frequency of /wa/, cf. Chitoran 2002b).

Overall, the three categories (mid and high diphthong, hiatus) are more difficult to differentiate in the back condition (at least with the measures employed here, and restricting the analysis to only five distinct time points), but those differences that could be observed are similar qualitatively to the more robust differences in the front condition: in both front and back condition, there is evidence that gliding occurs earlier in the mid diphthong than in the high diphthong or the hiatus sequence, that the two diphthongs have different tongue positions at the acoustic onset, and that hiatus sequences are longer than diphthongs. In terms of phonological distinctions, the current data therefore do not speak against different representations for the categories in the front vs. back condition: the mid diphthong vs. hiatus distinction can be understood mainly in terms of different degrees of temporal overlap between the two vowel targets (hence an earlier gliding and shorter duration for the diphthong), while the mid vs. high diphthong distinction can be understood mainly in terms of target specifications (hence the difference in tongue position at onset), with speakers maintaining at least some differences between categories in production.

5. Acknowledgements

Work 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. I thank Susanne Waltl for help with data collection.

6. Bibliography

BAAYEN, R. (2008). *Analyzing Linguistic Data: A Practical Introduction to Statistics Using R*. Cambridge UK/New York, Cambridge University Press.

BAAYEN, R.& DAVIDSON, R. H. & BATES, D. (2008). «Mixed-effects modeling with crossed random effects for subjects and items». *Journal of Memory and Language*59: 390-412.

BATES, D. (2010). *lme4: Mixed-effects modelling with R*. http://lme4.r-forge.r-project.org/lMMwR/lrgprt.pdf.

CHITORAN, I. (2001). *The Phonology of Romanian: A Constraint-Based Approach*. Berlin/New York,De Gruyter.

_____ (2002a). «The phonology and morphology of Romanian diphthongization». *Probus*14: 205-246.

_____ (2002b). «A perception-production study of Romanian diphthongs and glide-vowel sequences». *Journal of the International Phonetic Association*32(2): 203-222.

HARRINGTON, J.& HOOLE, P.& KLEBER, F.& REUBOLD, U. (2011). «The physiological, acoustic, and perceptual basis of high back vowel fronting: Evidence from German tense and lax vowels». *Journal of Phonetics*39: 121-131.

HOOLE, P. (1999). «On the lingual organization of the German vowel system». *Journal of the Acoustical Society of America*106(2): 1020-1032.

HOTHORN, T.& BRETZ, F.& WESTFALL, P. (2008). «Simultaneous Inference in General Parametric Models». *Biometrical Journal*50(3): 346-363.

MARIN, S. (2007). «An articulatory modeling of Romanian diphthong alternations». In: J. TROUVAIN & W. J. BARRY,eds.,*Proceedings of the16th International Congress of Phonetic Sciences, Saarbrücken, 6-10 August 2007*. Saarbrücken, University of Saarbrücken:453-456.

MARIN, S.& GOLDSTEIN, L. (2012). «A gestural model of the temporal organization of vowel clusters in Romanian». In:P. HOOLE& L. BOMBIEN& M. POUPLIER& C. MOOSHAMMER& B. KÜHNERT,eds.,*Consonant Clusters and Structural Complexity*. Berlin/Boston, De Gruyter: 177-203.

ÖHMAN, S. E. G. (1966). «Coarticulation in VCV utterances: spectrographic measurements». *Journal of the Acoustical Society of America*39: 151-168.

ROSETTI, A.& AVRAM, A.& COCIAN, C.&GHITU, G.& SUTEU, V.& ZAMFIRESCU, I.& COCENAI, S. (1955). «Cercetări experimentale asupra diftongilor româneşti». *Studii şi Cercetări Lingvistice* 6: 7-27.

SMITH, C. L. (1995). «Prosodic patterns in the coordination of vowel and consonant gestures». In: B. Connell & A. Arvaniti,eds.,*Phonology and Phonetic Evidence: Papers in Laboratory Phonology IV*. Cambridge, UK, Cambridge University Press: 205-222