

# A perception–production study of Romanian diphthongs and glide–vowel sequences

Ioana Chitoran

Linguistics and Cognitive Science Program Dartmouth College  
*ioana.chitoran@dartmouth.edu*

This study compares two diphthongs ([ea], [oa]) and two glide–vowel sequences ([ja], [wa]) in Romanian. The diphthongs and the sequences are auditorily very similar, but they differ in their phonological patterning. An integrated production and perception experiment is conducted in search of perceptual and acoustic evidence for the different phonological representations proposed for the diphthongs and the sequences. Four acoustic parameters of the diphthongs and sequences are measured and compared in a production study. In addition, a perception experiment tests native speakers’ ability to correctly identify the two types of vocalic sequences. The results support the different representations proposed for [ea] vs. [ja], but not necessarily for [oa] vs. [wa]. This asymmetry is interpreted in the context of language-specific frequency differences, and of contrast maintenance. The study shows that detailed phonetic description is needed for a complete understanding of the phonological facts.

## 1 Introduction

Glide–vowel sequences and rising diphthongs coexist in many languages. The interest of the Romanian data examined in this study lies in the striking auditory similarity between two such pairs: [ja] vs. [ea] and [wa] vs. [oa]. In all four cases the first element is pronounced as a glide. Although the language contains only one perfect minimal pair, the first one shown in (1), [ja]–[ea] and [wa]–[oa] occur in identical consonantal and prosodic environments.

- |        |         |               |    |         |              |
|--------|---------|---------------|----|---------|--------------|
| (1) a. | bjátə   | ‘poor’ fem.   | b. | beátə   | ‘drunk’ fem. |
|        | kwártsu | ‘quartz’ def. |    | skoártə | ‘tree bark’  |
|        | polwáre | ‘pollution’   |    | paloáre | ‘pallor’     |

From a purely descriptive standpoint, all the vocalic combinations illustrated in (1) could be called diphthongs. However, since this paper argues for a representational difference between (1a) and (1b), a terminological distinction is also adopted, for the sake of clarity. Thus, the data in (1a) are said to contain glide–vowel sequences, and the data in (1b) diphthongs. The glide–vowel sequences differ in their phonological behavior from the diphthongs. The goal of this study is to determine whether the glide–vowel sequences [ja] and [wa] are also phonetically

different from the diphthongs [ea] and [oa], respectively. An integrated production and perception study is designed to answer this question.

The paper is organized as follows. Section 2 presents the relevant data, focusing on the phonological differences between glide-vowel sequences and diphthongs. Two different phonological representations are proposed for sequences and for diphthongs, and the phonetic study undertaken here is designed to test their accuracy. Section 3 outlines earlier phonetic studies of the Romanian glides and diphthongs. Section 4 describes the production experiment (4.1) and the perception experiment (4.2). Section 5 contains the discussion of the results, and section 6 the conclusion.

## 2 The distribution and phonological behavior of diphthongs and glide-vowel sequences

The phonemic vowel inventory of Romanian is given in (2).

(2)	i	ɨ	u	
	e	ə	o	
	ea	a	oa	glides: j w

Following Chitoran (2001), the diphthongs are treated as the low counterparts of front and back vowels, respectively. Some occurrences of the diphthongs are phonologically and morphologically conditioned, but others are phonemic, as shown by (near) minimal pairs such as the following:

(3)	témə	‘theme’	teámə	‘fear’
	sérə	‘greenhouse’	seárə	‘evening’
	flék	‘shoe heel’	fleák	‘trifle’
	grótə	‘cave’	groápə	‘ditch’
	póftə	‘appetite’	poártə	‘gate’

The diphthongs, whether phonemic or derived, always occur under stress, a consequence of their historical development through the diphthongization of stressed mid vowels. As for the morphologically conditioned diphthongs, their distribution correlates with the category of gender in nouns, and with person in verbs (see Chitoran (2001, to appear) for a complete phonological and morphological analysis). The analysis presented in this section is somewhat simplified for clarity of exposition.

The distribution of the two diphthongs and of the two glide-vowel sequences in the Romanian lexicon is relevant for the setup of the production and perception experiments. The first observation that can be made is that the front diphthong [ea] occurs in fewer environments and is subject to more phonological restrictions than [oa]. For example, the diphthong [ea] does not occur in absolute word-initial position, whereas [oa] does (*oálə* ‘pot’, *oáste* ‘army’). While both diphthongs can be preceded by a consonant, as shown in (3), only [oa] can also be preceded by a vowel (*vioárə* ‘violin’). Only [ea], however, can occur in absolute word-final position (*steá* ‘star’, *neá* ‘snow’, *kureá* ‘belt’). The distribution of the two diphthongs is summarized in table 1.

Both diphthongs are subject to metaphony, the process by which the vowel of an inflectional marker affects the height of the stressed vowel of the stem. There are two contexts in which metaphony occurs in Romanian. When the inflectional marker contains the high vowel [i], the stressed vowel of the stem cannot be a diphthong. It surfaces instead as the corresponding mid vowel [o] or [e], as shown in the singular/plural alternations in (4). The examples contain two

**Table 1** Distribution of [ea] and [oa].

	[ea]	[oa]
Absolute word-initial	–	yes
Word-internal postvocalic	–	yes
Word-internal postconsonantal	yes	yes
Absolute word-final	yes	–

plural markers: /-i/ surfacing as a secondary articulation [-j] for feminine nouns, and /-uri/ surfacing as [-ur<sup>j</sup>] for neuter and feminine nouns.

(4) Metaphony before a high front vowel

SINGULAR	PLURAL	
poártə	pórts <sup>j</sup>	‘gate’
seárá	sér <sup>j</sup>	‘evening’
treábə	trébur <sup>j</sup>	‘task’

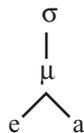
The second type of metaphony is triggered by the vowel /e/, but it only affects the front diphthong, [ea]. In the following examples, in (5), [-e] is a plural marker. Before [-e], the diphthong [ea] surfaces as [e] instead (cf. (5a)), but [oa] may surface as the stressed vowel of the stem (cf. (5b)).

(5) Metaphony before a mid front vowel

SINGULAR	PLURAL	
a. beátə	béte	‘drunk’ fem.
mireásə	mirése	‘bride’
b. koástə	koáste	‘rib’
koroánə	koroáne	‘crown’

Examples (4) and (5) illustrate a crucial property of the diphthongs, namely, their phonological and morphological alternation with the monophthongal vowels [e] and [o]. This particular behavior is captured by the phonological representation in (6), proposed for the diphthongs by Chitoran (2001).

(6) The phonological representation of diphthongs



In (6), both elements of the diphthong are represented as sharing a syllable nucleus. According to this representation, diphthongs are predicted to function as a single unit and a single segment.

Further support for this representation comes from the comparison with French diphthongs. A well-formed syllable onset in both French and Romanian consists of two elements or a maximum of three, where the first of the three is a sibilant. In addition, both languages have a glide formation process, which turns a high vowel into a glide when followed by another vowel. Glide formation applies if the onset contains only one segment (cf. (7a)),

but it is blocked if the onset already has two segments (cf. (7b)), such as an obstruent and a liquid (Morin 1976, Kaye & Lowenstamm 1984, among others).

(7) Glide formation in French and Romanian

French			Romanian			
a.	lje ~ lije	‘to bind’	pjan		‘piano’	
	fjakø ~ fjakø	‘carriage’	italjanə		‘Italian’ fem.	
	nwe ~ nue	‘to tie’	lwa ~ lua		‘to take’	
b.	plije	* <i>plje</i>	‘to fold’	klijent	* <i>kljent</i>	‘client’
	klue	* <i>klwe</i>	‘to nail down’	triumf	* <i>trjumf</i>	‘triumph’

Based on the blocked glide formation in obstruent-liquid onsets, the syllables in (8), which do surface in French, have been explained by treating the glide-vowel portion as a diphthong fully contained in the nucleus. The following sequences thus contain diphthongal nuclei:

(8) French diphthongs (complex onsets)

[wa]	tɔwa	‘three’
	pɔwa	‘prey’
[ɥi]	pɥi	‘rain’

Similarly, in Romanian [ea] and [oa] can be preceded by obstruent-liquid clusters, unlike [ja] and [wa].

(9) Romanian diphthongs (complex onsets)

[oa]	broaskə	‘frog’
	gloatə	‘crowd’
[ea]	treaz	‘awake’
	zdreantsə	‘rag’

These data therefore support the representation in (6), according to which [ea] and [oa] constitute diphthongal syllable nuclei.

Another important aspect of the proposed representation is the presence of a single mora. Although it is true that the Romanian diphthongs occur only in stressed syllables, these syllables can be either open or closed, and there is no evidence that diphthongal nuclei have extra weight compared to monophthongal ones.

Turning now to the glides [j] and [w], they are rendered in the orthography as ⟨i⟩ and ⟨u⟩, respectively. The orthography therefore encodes a distinction between ⟨ia⟩, the glide-vowel sequence, and ⟨ea⟩, which corresponds to the diphthong. Similarly, orthography distinguishes the sequence ⟨ua⟩ and the diphthong ⟨oa⟩. The distinction made in the orthography is significant to the extent to which it reflects a phonological contrast. The question asked in this study is whether native speakers actually produce and perceive these contrasts ([ja] vs. [ea] and [wa] vs. [oa]). Impressionistically, [ja] and [ea] have very similar pronunciations, and [wa] sounds even more similar to [oa], if not identical.

The distribution of the glides [j] and [w] in the language is asymmetrical. The front glide has a wider distribution than the back one. This asymmetry is not surprising, given that the palatal glide is also typologically more common than the labio-velar one. In Romanian, [j] can be found in a syllable onset in absolute word-initial position (*járnə* ‘winter’, *jépure* ‘rabbit’, *júte* ‘quick’), after a consonant (*pjátrə* ‘stone’, *ffér* ‘iron’) and word-internally after another vowel (*báját* ‘boy’). It can also surface in a syllable coda both word-internally (*hájnə* ‘coat’) and word-finally (*púj* ‘chicken’). The glide [w], however, never occurs in a word onset (there

**Table 2** Distribution of [j] and [w].

	[j] spelled ⟨i⟩	[w] spelled ⟨u⟩
Word onset	yes	–
Syllable onset postvocalic	yes	–*
Syllable onset postconsonantal	yes	yes
word-internal coda	yes	–
word-final coda	yes	yes

\* Except at the end of a word, before the feminine desinence vowel [-ə] or the definite article [-a]: *roúw-ə* 'dew' (indef.), *roúw-a* 'dew' (def.).

are no words beginning orthographically with the sequence #⟨ua⟩). As a syllable onset, [w] is always followed by the vowel [a], whereas [j] can also be followed by other vowels, as seen in many of the preceding examples. As a syllable coda, [w] only surfaces in word-final position (*bów* 'ox'), never word-internally, except for exactly two words (*áwgust* 'August' and *awgúr* 'augur'). The distribution of [j] and [w] is summarized in table 2.

A crucial difference in the phonological behavior of the glide-vowel sequences is the fact that [ja] and [wa] do not alternate with monophthongal vowels, the way the diphthongs do. The singular–plural alternations in (10), for example, still involve metaphony, but this time only the nucleus vowel is affected.

- |      |          |         |             |
|------|----------|---------|-------------|
| (10) | SINGULAR | PLURAL  |             |
|      | pjátrə   | pjétre  | 'stone'     |
|      | bjátə    | bjéte   | 'poor' fem. |
|      | polwáre  | polwér̃ | 'pollution' |

The representation in (11) has therefore been proposed as one that best captures this behavior of glide-vowel sequences (Chitoran 2001).

- (11) The phonological representation of glide-vowel sequences



This representation differs from that proposed for the diphthongs in the fact that only the vowel belongs to the nucleus, while the glide belongs to the onset. This syllable structure predicts the resistance to glide formation following an obstruent-liquid cluster, illustrated earlier in (7) for French and Romanian.

The distribution and the phonological behavior of the Romanian diphthongs compared to glide-vowel sequences reveal several important observations. First, we see that the back diphthong [oa] has fewer phonological restrictions than [ea], and thus occurs in more phonological environments. Second, the back glide [w] occurs less frequently in a syllable onset than [j]. Third, throughout the Romanian lexicon, the sequence [wa] is also much less frequent than the diphthong [oa]. At the same time, the distribution of [ja] and [ea] in the vocabulary is much more balanced. Frequency differences must be taken into consideration in a phonetic study, since they may affect native speakers' production and perception of a contrast between glide-vowel sequences and diphthongs.

In addition to the low frequency of [wa], it should be noted that this sequence occurs predominantly in loanwords, mostly from French. This explains the ⟨ua⟩ orthography, which is based on the French spelling. A list of native and non-native words containing the orthographic sequence ⟨ua⟩ is given in (12a, b). The native Romanian examples in (12b) constitute an exhaustive list. A dash (–) indicates a morpheme boundary.

(12) a. NON-NATIVE		b. NATIVE	
kwarts	‘quartz’	alwat	‘dough’
kwantə/kwantum	‘quantum’	lw–a	‘to take’
kwartet	‘quartet’	lw–at	‘taken’
kwaternar	‘quaternary’		
akwarelə	‘watercolor’		
gwafe	‘gouache’		
nwantə	‘nuance’		
swav	‘suave’		
an–w–al	‘yearly’		
an–w–ar	‘yearbook’		
akt–w–al	‘current’		
tekst–w–al	‘textual’		
kontjept–w–al	‘conceptual’		
spirit–w–al	‘spiritual’		
polw–are	‘pollution’		
polw–at	‘polluted’		
dilw–at	‘diluted’		
eventwal	‘possibly’		

The shorter words, such as *swav* ‘suave’ and the forms of the verb *lua* ‘to take’, show some speaker variation, possibly related to differences in speech rate. The pronunciations *suav* and *lua* are also attested. In fact, Academia Română (1995), the Handbook of Orthography of the Romanian Academy, which is also a pronunciation guide, prescribes the disyllabic pronunciation [u.a] for all the words in (12), with the exception of the first six words in (12a), which should be pronounced with monosyllabic [wa]. It is also interesting to notice that in all of these six words the glide-vowel sequence is preceded by a velar stop.

The difference in orthography and the prescribed pronunciation have important consequences for a phonetic study. The data in (12) show that it is virtually impossible to construct a well-balanced wordlist, with lexical items contrasting [wa] and [oa]. A number of predictions can also be made based on these two factors. If native speakers’ pronunciation is affected by the ⟨ua⟩ spelling and by the prescribed pronunciation of this vocalic sequence ([u.a]), then we expect to find a significant difference in the production of the glide-vowel sequence relative to the diphthong. Specifically, we expect the glide-vowel sequence to be significantly longer than the diphthong, simply because the sequence may count as two syllables (two vowels with hiatus), while the diphthong always counts as one.

If, on the contrary, native speakers are not sensitive to the spelling difference and do not follow the prescribed pronunciation, then it is less likely that a difference will be found in the production and perception of [wa] and [oa], since the wider phonological distribution of the diphthong [oa] may in some sense eclipse the sequence [wa].

The question of a possible contrast between diphthongs and similar glide-vowel sequences has been addressed previously by Romanian phoneticians. The next section summarizes several relevant studies.

### 3 Previous phonetic studies of the [ja]–[ea], [wa]–[oa] distinction

The question of the acoustic distinction between the two types of vocalic sequences, [ja]–[ea] and [wa]–[oa] in Romanian was addressed early on by Rosetti (1955, 1959), in an acoustic, articulatory and perception study. The acoustic study consists of duration measurements made on kymograph recordings of (near) minimal pairs. Rosetti finds that [ja] and [wa] are distinct from [ea] and [oa], although they have a similar acoustic structure. In all four sequences, two elements can be clearly distinguished, the first of which is always a glide, whether high or mid.

A first set of comparisons involves the front diphthong and glide. The duration of [ea] is compared to that of [ja], then [ea] and [ja] are each compared to the vowel [a]. Rosetti found that the total durations of [ea] and [ja] are comparable, although the first element in [ea] is shorter than the [j] in [ja]. The duration of [a] also varies, making up for the difference. It should be noted that the target words were spoken in isolation, not embedded in carrier phrases, and that the results were not subjected to a statistical analysis.

In the second set of measurements, the back diphthong [oa] is compared to the monophthongs [a], [u] and [o]. Here the first element of the diphthong can be clearly delimited acoustically and identified as corresponding in height to the back mid vowel [o].

More convincing evidence for the difference between [ea] and [ja] comes from articulatory data, from comparisons of palatograms of [Cea] and [Cja] sequences. More contact is visible for the sequence [Cja], whereas [Cea] shows contact only on the edges of the palate.

Finally, in a perception experiment, subjects were asked to listen to words containing [ja] and [ea], played backwards. Rosetti reports that in most cases the listeners were able to identify the two elements in reverse order ([ae] or [aj]), and to distinguish between [j] and [e] as the final portion of the sequence. The perception experiment for [oa] and [wa] was different, and involved splicing. When the [o] portion was replaced by the vowel [u], subjects reported perceiving a different diphthong rather than [oa] (Rosetti 1959: 41).

A later acoustic study by Ulivi (1975) considers only the first element of the diphthongs, [e] and [o], respectively, and compares it to [j] and [w]. The comparisons are based on wideband spectrograms of the recordings of three speakers. The data consisted of words read in isolation, with the target segments in different positions in the word. The author reports a difference in duration, with the [e] and [o] portions being shorter than [j] and [w]. This result is consistent with Rosetti's, but Ulivi's study also does not include a statistical analysis. The acoustic descriptions are nevertheless valuable, since they do suggest a difference between [j], [w] and [e], [o]. For example, the author interprets [j] and [w] as being consonantal, based on the larger amount of friction seen on the spectrogram, compared to [e] and [o]. She concludes that [j], [w], [e] and [o] are all glides, but the first two are consonantal ('semi-consonants'), while the diphthongs contain vocalic glides ('semi-vowels').

The two studies do not provide a definitive answer, but they do support the phonetic distinction between diphthongs and similar glide-vowel sequences in Romanian. The present experiment was designed to further clarify this question: do native speakers distinguish between [ja]–[ea] and [wa]–[oa], or not? If they do, how many and what particular acoustic cues do they rely on? Does the phonetic realization support the phonological analysis which argues for a distinction between [ja]–[ea] and [wa]–[oa]? The following section describes the production and perception experiments undertaken to answer these questions.

### 4 Experiments

Section 2 presented evidence that the diphthongs and glide-vowel sequences behave differently in the phonology. Diphthongs alternate with monophthongal vowels in singular-plural pairs (*beátă*–*béte* 'drunk' fem.), while the glide portion of a glide-vowel sequence remains

unaffected in the same context, and only the nucleus vowel changes (*bjátə–bjéte* ‘poor’ fem.). Consequently, the diphthongs are analyzed as single segments, the low counterparts of front and back vowels, respectively, whereas glide-vowel sequences are analyzed as a sequence of two segments, an onset glide and a nucleus vowel. Romanian orthography also maintains a distinction between all four: ⟨ea, oa, ia, ua⟩.

Based on the differences outlined above, and on the results of previous studies by Rosetti and Ulivi, the following hypothesis can be formulated regarding the phonetics of diphthongs and glide-vowel sequences: [ja] and [wa] are produced differently from [ea] and [oa]. The first portion is in all cases a glide, but they differ in height, and this difference should be reflected in the acoustic structure. At the same time, if frequency and prescribed pronunciation play a role, then it is possible that [ja] and [wa] are phonetically identical to [ea] and [oa], in spite of the phonological difference. In this case we do not expect native speakers to be able to identify four different sequences [ja], [wa], [ea] and [oa].

If the first hypothesis is correct, native listeners should be able to identify four different vocalic sequences, based on differences in several acoustic parameters. The first parameter is acoustic duration. The assumption here is that the phonology-to-phonetics mapping assigns individual duration to individual segments. Therefore, if diphthongs constitute one segment, and glide-vowel sequences two segments, then, phonological weight being equal, the total duration of diphthongs is expected to be shorter than that of glide-vowel sequences (see Ham 1998 for a proposed integrated phonological and phonetic timing model which takes into account both syntagmatic and hierarchical effects).

A second acoustic parameter expected to differ is the transition portion characteristic of a glide, the change in frequency between the onset of the glide and the onset of the vowel portion following it. The duration of this transition can also be affected by the one- vs. two-segment difference. Glide-vowel sequences are expected to have a longer transition duration than the diphthongs. If [j] and [w] differ from [e] and [o] in height, this difference is also expected to affect the second formant transition, from the first vocalic element to the following vowel [a]. More specifically, the difference in height can affect the onset of F2 at the beginning of the glide, and the transition rate. The onset of F2 is expected to be higher for [j] than for [e], and lower for [w] than for [o], corresponding to the high-mid difference. At the same time, a faster transition rate is expected for glide-vowel sequences than for diphthongs. F2 has a steeper slope in glide-vowel sequences, going from a high vowel to a low one, than in diphthongs, where it goes from a mid to a low vowel. Therefore the rate at which the second formant drops or rises is faster in glide-vowel sequences.

The production and perception experiments are described in the following two subsections, and the results are presented.

## 4.1 Experiment 1: Production

### 4.1.1 Methodology

For the acoustic study, data from four native speakers of Romanian were collected, three male and one female. All four speakers are originally from the Bucharest area, speak what is considered to be the standard dialect, and have had no speech or hearing disorders. Two speakers were recorded in a soundproof booth in the USA. They had been in the USA for less than a year and for two years, respectively, at the time the recording was made. The other two speakers were recorded in Bucharest, in a quiet room. The recordings were made on a high quality Marantz analog tape recorder, using an AKG microphone, model D310.

The speakers read a wordlist consisting of near minimal pairs of words which contained the sequences [ja], [ea], [wa], [oa] in contexts as similar as possible. The sequences that were compared were preceded by the same consonants, and the syllable count and stress pattern were held constant. The wordlist is given in (13). Each word was embedded in the carrier phrase: [spuneá \_\_\_\_fip leká] ‘he said \_\_\_\_ and left’. (#) indicates a word boundary.

(13) Wordlist			
[Coá]		[Cwá]	
koan.də	proper name	kwan.tə	'quantum'
skoar.tsə	'tree bark'	kwar.tsu	'quartz' def.
loa.zə	'louse'	lwa.tə	'taken' fem.
pa.loa.re	'pallor'	po.lwa.re	'pollution'
o.noa.re	'honor'	a.nwa.re	'yearbooks'
[Ceá]		[Cjá]	
bea.tə	'drunk' fem.	bja.tə	'poor' fem.
kli.pea.lə	'blink'	ko.pja.tə	'copied' fem.
ka.fea.wa	'coffee' def.	ka#fja.ra	'like beast' def.
fu.mea.zə	'he smokes'	a.mja.zə	'afternoon'
ves.teaw	'they announced'	ves.tjar	'locker room'
po.tjea.lə	'ugliness'	spe.tjja.lə	'special' fem.
pe#d3ea.nə	'on eyelash'	bel.d3ja.nə	'Belgian' fem.
ju.tsea.lə	'speed'	sko.tsja.nə	'Scottish' fem.
pə.kə.lea.lə	'joke'	i.ta.lja.nə	'Italian' fem.
plo.ko.nea.lə	'brown nosing'	ko.lo.nja.lə	'colonial' fem.
plik.ti.sea.lə	'boredom'	par.na.sja.nə	'Parnassian' fem.
grə.mə.dea.lə	'crowd'	ka.na.dja.nə	'Canadian' fem.

The list contains twelve [ja]–[ea] pairs and only five [wa]–[oa] pairs, due to the lower frequency of [wa] sequences. Three repetitions of each word were recorded for each speaker, resulting in a total of 60 [wa] sequences, 60 [oa] diphthongs, 144 [ja] sequences and 144 [ea] diphthongs.

According to Academia Română (1995)'s Handbook of Orthography, only two of the five words containing orthographic ⟨ua⟩ have the prescribed pronunciation [wa]: *kwantə*, *kwartsu*. Of the twelve words containing ⟨ia⟩, only three have the prescribed [ja] pronunciation: *amjazə*, *bjatə*, *fjarə*. The other words are supposed to be pronounced with high vowels instead of glides ([i.a] and [u.a]). Although the Handbook of Orthography does not explain it, it is possible that the prescribed [i.a] pronunciation is based purely on morphological arguments. In these words the [ja] sequence spans a morphological boundary. Impressionistically, speakers do not seem to observe the prescribed pronunciation in hiatus, especially in fluent speech. It is possible, however, that the duration of a glide-vowel sequence is more variable than that of a diphthong, and this detail can be verified in an acoustic analysis.

#### 4.1.2 Analysis

The recorded sentences were digitized on a SPARC station LX at a sampling rate of 11 kHz, and processed by the software package ESPS/waves+. Measurements were made on waveforms and wideband spectrograms. For the measurement of the total duration of glide-vowel sequences and diphthongs, the following landmarks were used. The onset of the sequence/diphthong was determined on the waveform as the onset of periodicity after a stop burst or frication portion, and as the first larger period after a nasal or liquid. On the spectrograms, the onset of F1 was taken to mark a vocalic onset. The offset of the sequence/diphthong was determined on the waveform as the last period before a following stop closure or before the smaller periods of a nasal or liquid, and on the spectrogram as the offset of F2. A labeled waveform and spectrogram for the token [koandə] (proper name) are shown in figure 1.

The transition duration was measured according to the following criteria, based on Ren (1986) for Chinese diphthongs. For [ja] and [ea] the transition onset is chosen to be the highest

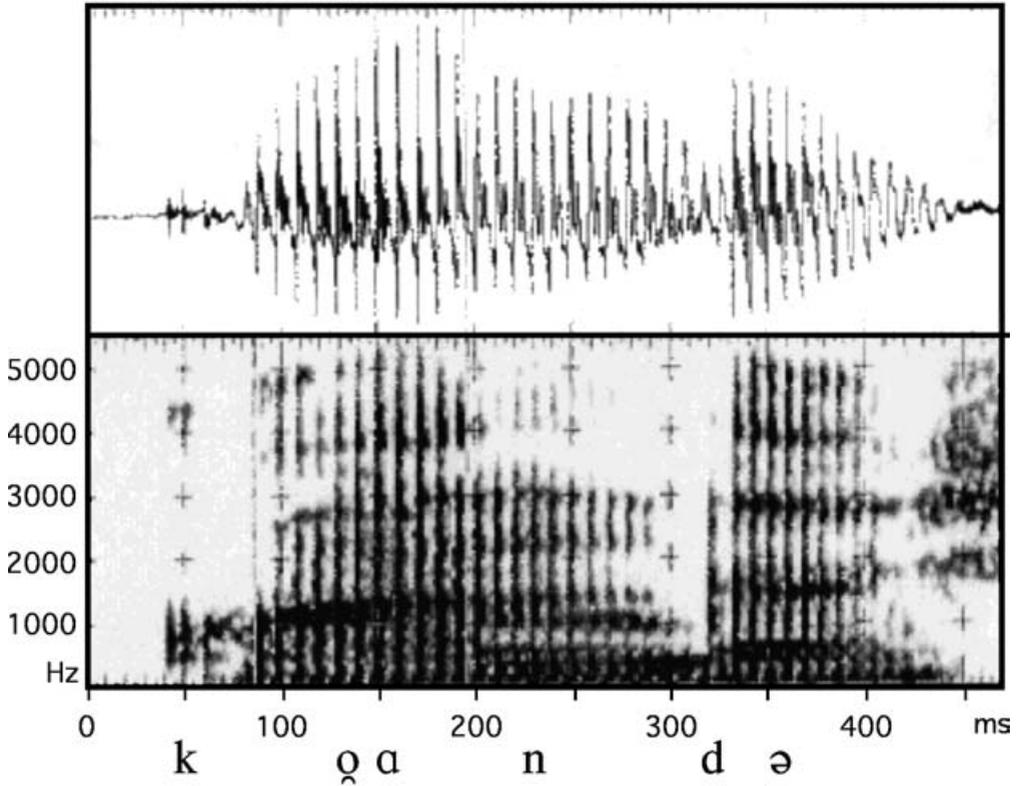


Figure 1 Waveform and spectrogram of the token [koandə].

F2 value at the beginning of the sequence/diphthong, before it falls by at least 20 Hz. For [wa] and [oa] it is the lowest F2 value, marked at the point where it rises by at least 20 Hz. The F2 values were determined by running a formant track over the entire sequence/diphthong, set to compute the F2 value every 5 ms. The offset of the transition is marked by the turning point from a falling F2 for [ja] and [ea], and a rising F2 for [wa] and [oa], to an F2 steady-state.

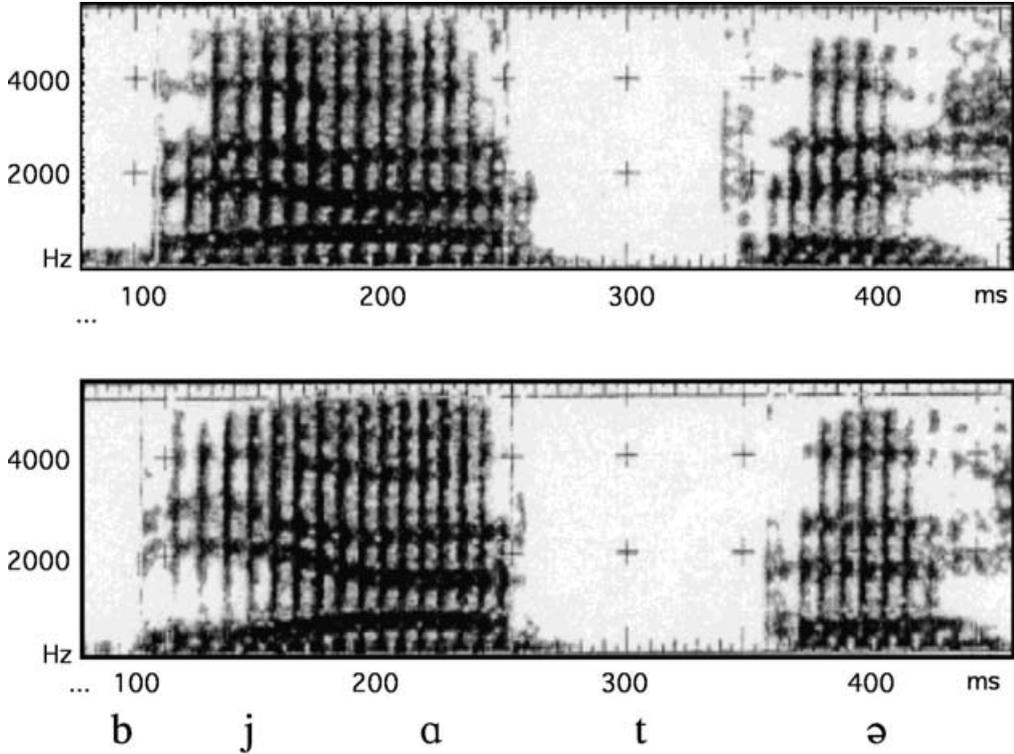
The values for the F2 transition onset were also compared, as well as the F2 rates of transition. The transition rate was calculated by subtracting the lowest F2 value from the highest F2 value, and dividing the result by the transition duration. The results were evaluated by two-tailed t-tests.

### 4.1.3 Results

The acoustic measurements revealed an asymmetry between the two sets of pairs, [ja]–[ea] and [wa]–[oa]. A statistically significant difference was found between [ja] and [ea] in all four parameters measured, whereas the parameters measured for [wa] and [oa] were comparable.

Below are representative spectrograms of two pairs of tokens that were compared, containing the sequence [ja] vs. the diphthong [ea] (figure 2), and the sequence [wa] vs. the diphthong [oa] (figure 3).

For the sequence [ja] and the diphthong [ea] a significant difference was found in all four parameters. The sequence [ja] was found to be significantly longer than [ea] ( $p = .000$  for all four speakers). The total duration of [wa] and [oa] is comparable for all speakers ( $p > .05$ ). The statistically significant results are indicated with an asterisk in table 3.



**Figure 2** Spectrograms of the tokens [beatə] (above) and [bjatə] (below).

**Table 3** Total duration (ms) – average values.

Speaker	[ja]	[ea]	
1	154	115	$t(35) = -7.1^*$
2	147	117	$t(35) = -8.7^*$
3	166	125	$t(35) = -9.7^*$
4	157	126	$t(35) = -5.8^*$
	[wa]	[oa]	
1	125	116	$t(14) = -1.7$
2	140	135	$t(14) = -.37$
3	140	136	$t(14) = -.25$
4	146	131	$t(14) = -1.6$

The transition duration of [ja] was found to be significantly longer than that of [ea] for all four speakers ( $p = .000$ ), while for [wa] and [oa] no statistically significant difference was found ( $p > .05$ ).

In terms of the transition rate, the sequence [ja] showed a significantly higher F2 transition rate than the diphthong [ea], as predicted. F2 has a steeper slope in [ja] than in [ea], and thus

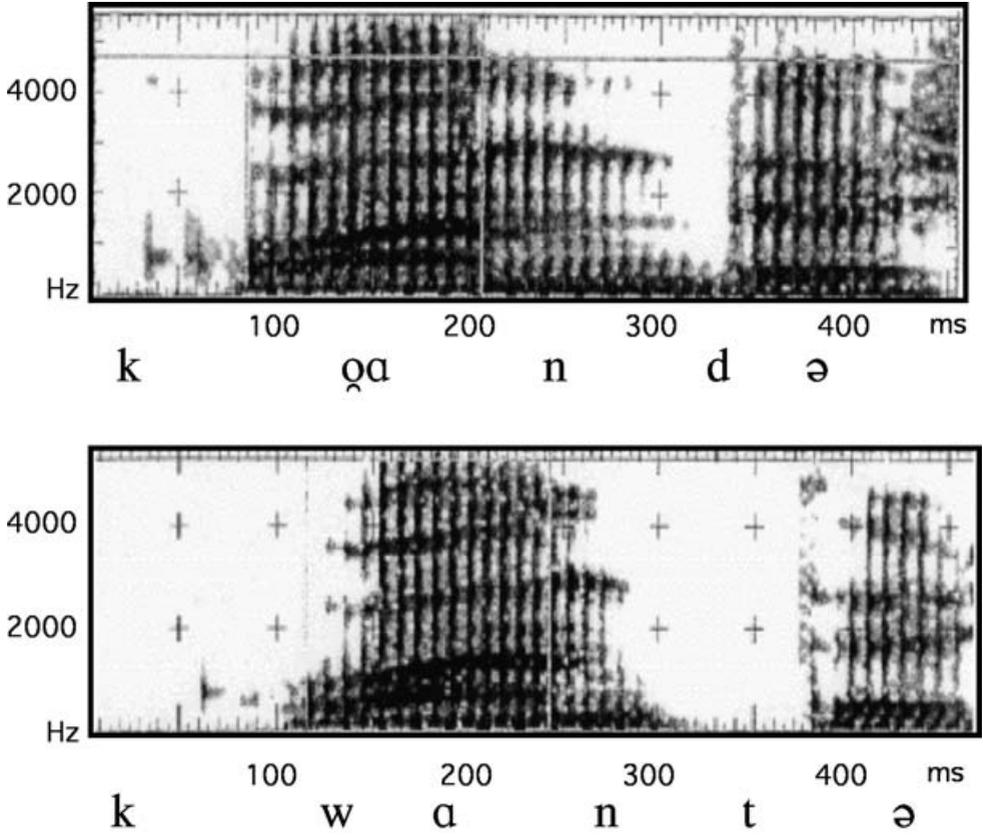


Figure 3 Spectrograms of the tokens [koandə] (above) and [kwantə] (below).

the rate at which the second formant drops is faster in the glide-vowel sequence. A higher value indicates a faster transition rate. The difference was statistically significant for all four speakers ( $p = .000$ ). For [wa] and [oa], however, the transition rate turned out to be comparable ( $p > .05$ ).

The F2 onset values also turned out to be significantly different for [ja] and [ea] for all four speakers, with F2 starting higher for [j] than for the [e] of the diphthong ( $p = .000$ ). The F2 onset values were comparable for [wa] and [oa] for three of the four speakers. Interestingly, one speaker showed a statistically significant difference ( $p = .005$ ), but in the unexpected direction. For the second speaker the F2 onset of [w] was found to be significantly higher than that of [o] in the diphthong. The F2 of [w] is actually predicted to be lower, since the glide [w] may involve more pronounced lip rounding than [o]. This difference is treated simply as an idiosyncratic aspect of his speech. The results are summarized below.

To summarize, for [ja] and [ea] all the parameters measured showed a statistically significant difference, supporting a different phonological analysis for glide-vowel sequences and diphthongs. The different phonological behavior of [ja] and [ea] is reflected in their different phonetic realizations. The shorter total duration and transition duration of [ea] are consistent with the representation of the diphthong as a single segment contained in a syllable nucleus. The longer duration of [ja] supports its representation as a sequence of two segments, filling an onset and a nucleus. Moreover, the higher F2 onset of [ja] and its faster transition rate show that the first elements of [ja] and [ea] have different vowel qualities.

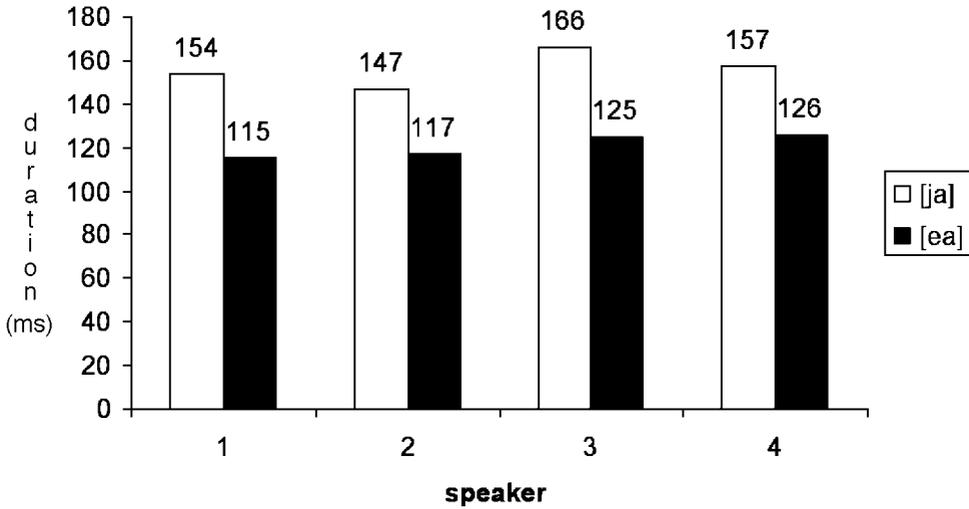


Figure 4 Total duration of [ja] vs. [ea] for all four speakers ( $p < .01$ ).

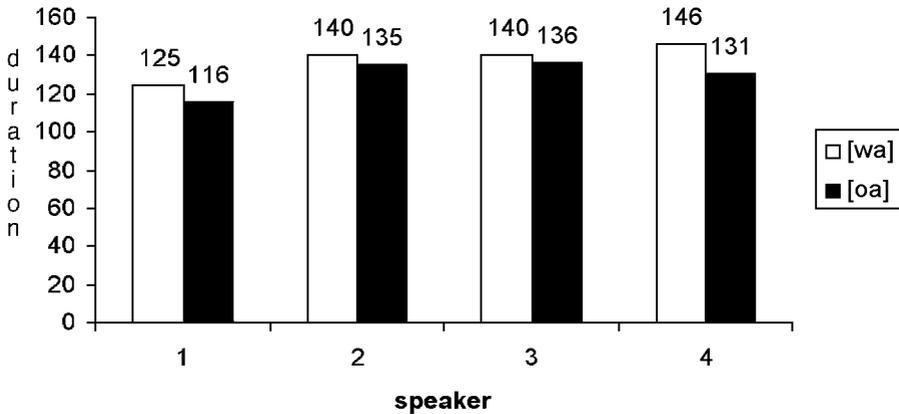


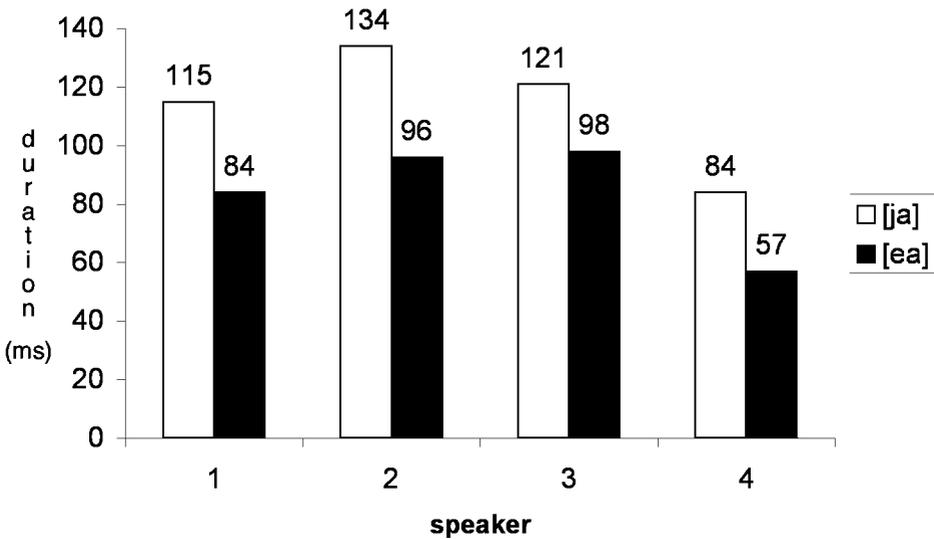
Figure 5 Total duration of [wa] vs. [oa] for all four speakers ( $p > .05$ ).

At the same time, however, the different phonological behavior of [wa] and [oa] is not reflected in the acoustics. The results of the acoustic measurements do not support the distinction proposed in the phonological analysis of the back glide-vowel sequence and diphthong.

Before moving on, it is important to perform a further test, to determine whether this asymmetry between the two pairs is not perhaps due to the asymmetries in distribution between sequences and diphthongs outlined at the beginning of the paper. It was pointed out in section 1 that, due to the low number of words containing the orthographic ⟨ua⟩ sequence, a balanced wordlist pairing [wa] and [oa] is impossible to construct, unless it is limited to very few pairs of words. Hence the relatively short [wa]–[oa] list used in this study, and presented in (13). Out of the five [wa] words in the list, two have the prescribed, invariant [wa] pronunciation, but they represent close to half of the list. In the longer [ja]–[ea] list, however, three out of twelve [ja] words, a lower ratio, have the invariant [ja] pronunciation. The remaining nine

**Table 4** Transition duration (ms) – average values.

Speaker	[ja]	[ea]	
1	115	84	$t(35) = -7.7^*$
2	134	96	$t(35) = -9.9^*$
3	121	98	$t(35) = -4.5^*$
4	84	57	$t(35) = -5.8^*$
	[wa]	[oa]	
1	113	97	$t(14) = 1.7$
2	104	91	$t(14) = -1.6$
3	128	119	$t(14) = -.64$
4	93	87	$t(14) = .47$



**Figure 6** Transition duration of [ja] vs. [ea] for all four speakers ( $p < .01$ ).

words have the prescribed [i.a] pronunciation. Although the speakers were closely monitored during the recording sessions, and impressionistically none were heard to produce hiatus, it is true that this pronunciation remains an option for them. The [ja]–[ea] list, therefore, contains more words in which the glide can potentially be realized as a vowel, thus with a possibly longer duration. If the speakers produced something closer to the vowel [i] than to a glide, that alone could have caused the [ja] duration to be longer.<sup>1</sup>

This possibility was tested by re-doing the statistical analyses on just a subset of the [ja]–[ea] list, that exactly matches the composition of the [wa]–[oa] list. This time, two words contain the invariant [ja] pronunciation (the first two on the right), and the other three contain the prescribed hiatus:

<sup>1</sup>Thanks to Donca Steriade for pointing out this possibility.

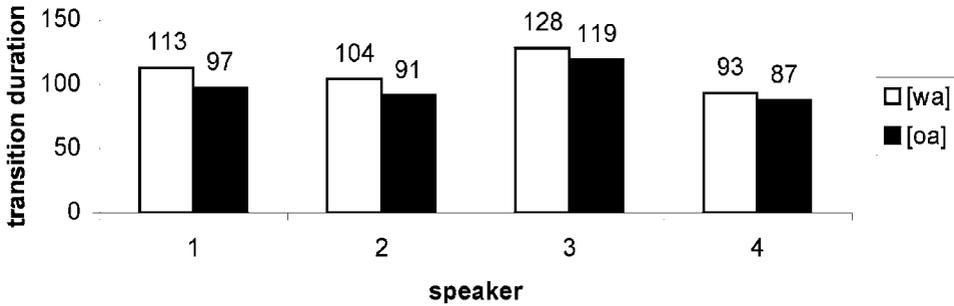


Figure 7 Transition duration of [wa] vs. [oa] for all four speakers ( $p > .05$ ).

Table 5 F2 transition rate (Hz/ms) – average values.

Speaker	[ja]	[ea]	
1	5.85	4.01	$t(35) = -7^*$
2	4.03	2.48	$t(35) = -6.6^*$
3	6.52	3.05	$t(35) = -7.8^*$
4	8.48	4.89	$t(35) = -7.1^*$
	[wa]	[oa]	
1	3.73	3.6	$t(14) = -.69$
2	3.47	3.35	$t(14) = -.61$
3	4.65	5.26	$t(14) = 1$
4	5.67	5.47	$t(14) = -.31$

(14) Subset of [ja]–[ea] list

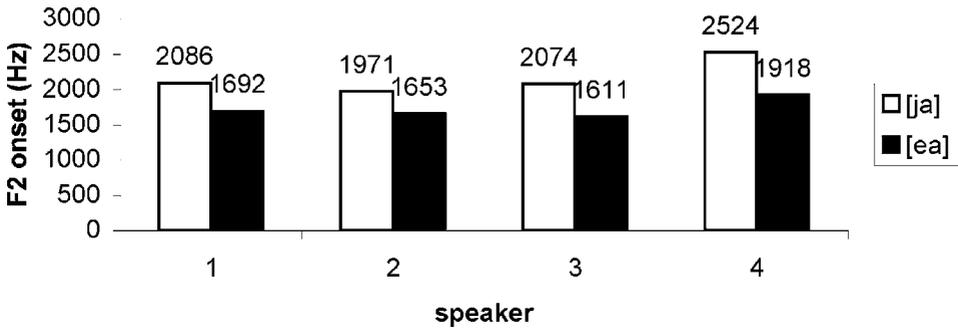
[Ceá]		[Cjá]	
beatə	‘drunk’ fem.	bjatə	‘poor’ fem.
kafəawa	‘coffee’ def.	ka fjata	‘like beast’ def.
klipealə	‘blink’	koɣjatə	‘copied’ fem.
vestəaw	‘they announced’	vestjar	‘locker room’
grəmədəalə	‘crowd’	kanadjanə	‘Canadian’ fem.

The new comparison yielded very similar results to those based on the longer list. All four speakers showed a statistically significant difference between the total duration of [ea] and [ja] ( $p < .01$ ), the transition rate ( $p < .05$ ), and the F2 onset ( $p < .01$ ). Two of the speakers also showed a statistically significant difference in the F2 transition duration ( $p < .01$ ), whereas speakers 2 and 3 did not ( $p > .05$ ). Upon re-examining the data of speakers 2 and 3, it could be seen that a large number of their [ea] tokens had nearly flat F2 trajectories. For this reason, in these tokens the duration of the F2 transition was taken to be the same as the total duration of the diphthong, because there was no point during the diphthong where F2 fell by more than 20 Hz. The flatness of the F2 trajectory in the data of speakers 2 and 3 thus accounts for larger individual values for the F2 transition duration in many of the [ea] tokens.

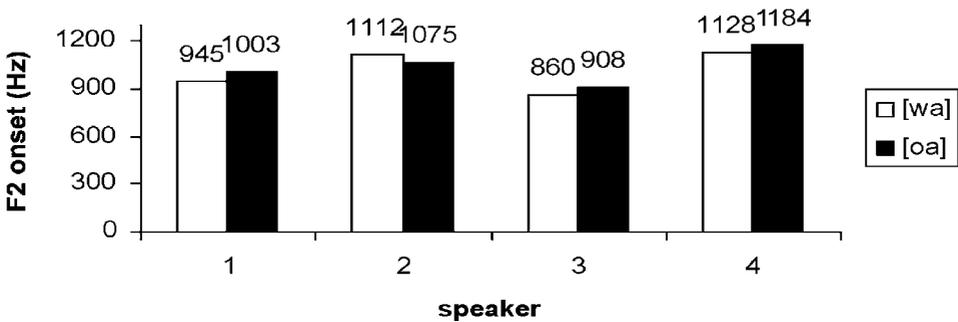
What is important for the present study, however, is the fact that a significant difference in the total duration of sequences and diphthongs is still present in the subset of the tokens for

**Table 6** F2 onset (Hz) – average values.

Speaker	[ja]	[ea]	
1	2086	1692	$t(35) = -16.6^*$
2	1971	1653	$t(35) = -10.1^*$
3	2074	1611	$t(35) = -12.1^*$
4	2524	1918	$t(35) = 13.5^*$
	[wa]	[oa]	
1	945	1003	$t(14) = 1.9$
2	1112	1075	$t(14) = 3.2^*$
3	860	908	$t(14) = 1.9$
4	1128	1184	$t(14) = .9$



**Figure 8** F2 onset of [ja] vs. [ea] for all four speakers ( $p < .01$ ).



**Figure 9** F2 onset of [wa] vs. [oa].  $p < .01$  for speaker 2;  $p > .05$  for remaining speakers.

all four speakers. This finding further strengthens the evidence that speakers did not resort to the hiatus pronunciation prescribed for some of the ⟨ia⟩ sequences, and validates the earlier results, based on the complete wordlist.

The results of the production study will be further tested in a perception experiment, which is described in the next subsection.

## 4.2 Experiment 2: Perception

### 4.2.1 Methodology

The goal of the perception experiment is to determine whether native speakers can perceive the difference between glide-vowel sequences and diphthongs. Based on the results of the acoustic study, separate predictions can be made about the front and back vocalic sequences, respectively. Listeners should be able to correctly identify [ja] and [ea], whose phonetic realization was found to be different. Given that no difference was found in the production of [wa] and [oa], the prediction is that listeners will not be able to correctly identify the glide-vowel sequence and the diphthong.

The tokens of one male speaker, speaker 1, were used in the perception experiment. The [ea], [ja], [wa], [oa] portions were excised from each word, using the segmentation criteria described in section 4.1, for the total duration of the sequence/diphthong. The files were transferred to a different computer, where they were converted into audio files and randomized for the perception experiment using the Bliss perception software (Mertus 1985). Two separate tests were set up, one containing the [ja] and [ea] tokens, the other containing the [wa] and [oa] tokens. Two repetitions of each token, chosen randomly, were included in a practice test (24 [ja] sequences, 24 [ea] diphthongs, 10 [wa] sequences and 10 [oa] diphthongs). All the tokens recorded, with additional three repetitions of each, were included in the actual perception test (108 [ja] sequences, 108 [ea] diphthongs, 45 [wa] sequences, 45 [oa] diphthongs). The two tests were transferred onto analog tapes on the Marantz tape recorder.

The perception tests were administered in Bucharest. The subjects were fourteen native speakers of Romanian, five male and nine female. Two of them were teenagers (aged 13 and 15 years), and the rest were aged between 25 and 60 years. Some speakers took the tests individually, others took it in small groups, on different days. The two tests ([ja]–[ea] and [wa]–[oa]) were given in different order to different groups of listeners. The test tapes were played on the Marantz tape recorder, in a quiet room, and no headphones were used.

The subjects were asked to listen to the practice test first. For the actual tests they were asked to perform an identification task. This particular task was chosen rather than a discrimination task precisely in order to make the test more difficult. The intention was to test not only the listeners' ability to distinguish between glide-vowel sequences and diphthongs, but also their ability to reliably identify the sequences and the diphthongs, based on the acoustic parameters studied in experiment 1. Each listener was given a sheet on which two different orthographies were marked for each token: ⟨ia⟩ and ⟨ea⟩ for one test, ⟨ua⟩ and ⟨oa⟩ for the other. The listeners were asked to circle the spelling they considered appropriate for each token they heard.

### 4.2.2 Results

The answers to the two perception tests were analyzed by a binomial sampling distribution test. The two predictions based on the production study were both borne out. The sequence [ja] and the diphthong [ea] were reliably correctly identified at a significance level of .05 ( $z = 33$ ). The sequence [ja] was correctly identified 89% of the time ( $z = 39$ ), and the diphthong [ea] 78% of the time ( $z = 28$ ).

The [wa]–[oa] test, as expected, presented more difficulty. The number of correct answers is not significant at the  $p = .05$  level ( $z = 3$ ). The sequence [wa] was correctly identified 46% of the time ( $z = 4$ ), and the diphthong [oa] 48% of the time ( $z = 2$ ). The results are summarized in table 7, where the statistically significant differences are marked with an asterisk.

The results of the perception experiment are consistent with those of the production study in reflecting the phonological difference between [ja] and [ea]. The glide-vowel sequence and the diphthong are produced differently and perceived as different. The phonological difference between [wa] and [oa], however, is not reflected in their acoustic realization, and thus the sequence and the diphthong cannot be reliably identified.

**Table 7** Identification of glide-vowel sequences vs. diphthongs – averages.

	% correct answers	z-value	Standard deviation (for H <sub>0</sub> )
[ja]	89	39*	19.4 (1.2%)
[ea]	78	28*	19.4 (1.2%)
overall [ja]/[ea]	83	33*	27.5 (.9%)
[wa]	46	4	11.6 (2%)
[oa]	48	2	11.6 (2%)
overall [wa]/[oa]	47	3	16.4 (1.5%)

## 5 Discussion

The first piece of information revealed by the results presented above concerns the use of prescribed pronunciation. The findings show that native speakers do not follow the prescribed pronunciation, and the reliable evidence comes from the study of the [wa]–[oa] pair. Three of the five ⟨ua⟩ words in the wordlist contain [u.a] prescribed pronunciations, and if the pronunciation had been observed, there should have been a statistically significant difference in all four acoustic parameters, primarily in the total duration and the transition duration. The fact that no such difference was found confirms that all the orthographic ⟨ua⟩ sequences were pronounced as monosyllables, [wa]. We know from earlier work (e.g. Liberman et al. 1956) that manipulating the formant transition leads to changes in perception. Specifically, increases in the transition duration while keeping the onset and offset values constant determine judgment changes from a glide-vowel sequence to a vowel-vowel sequence. The results of the current study do not, therefore, support the presence of [u.a] sequences in the data.

The repeated test, using a subset of the [ja]–[ea] list, also confirmed that the prescribed [i.a] pronunciation was not observed either. This strengthens the validity of the results for [ja] and [ea] and confirms that the statistically significant duration differences are not due to prescribed pronunciation, but to structural differences between glide-vowel sequences and diphthongs.

It is, therefore, safe to conclude that different phonological representations for [ja] and [ea] are supported by the results of the production-perception study. The phonological behavior of [ja] as a sequence of two segments and of [ea] as one segment is supported by the duration differences. The total duration and transition duration of glide-vowel sequences is significantly longer than that of diphthongs. The first elements of [ja] and [ea] are not identical, but differ in height, as shown by significant differences found in the F2 onset and in transition rate. The sequence [ja] has a higher F2 onset, corresponding to a high glide, and a faster transition rate, whereas the diphthong [ea] has a lower F2 onset, corresponding to a mid glide, and a lower transition rate. Moreover, native speakers of Romanian can reliably identify [ja] and [ea]. These results are also consistent with the acoustic and articulatory descriptions of [ja] and [ea] by Rosetti (1955, 1959) and Ulivi (1975).

The phonological difference between [wa] and [oa], however, is not directly encoded in the phonetics. The glide-vowel sequence and the diphthong are comparable in all four acoustic parameters measured, and native speakers could not identify them in the perception experiment. These results go against the earlier interpretation of the acoustic data by Rosetti and Ulivi. Does this mean that Romanian does not actually distinguish between a back glide-vowel sequence and a back diphthong although the phonology and the orthography suggest that they both exist?

The existence of only one phonetic realization for both [wa] and [oa] can be explained by two factors, one language-specific and one universal. It has already been shown that in

Romanian the sequence ⟨ua⟩ occurs in considerably fewer lexical items than the diphthong ⟨oa⟩, and primarily in loanwords. The relatively limited distribution of ⟨ua⟩ may be responsible for the phonetic neutralization between the glide-vowel sequence and the diphthong. At the same time, the acoustic difference between two back rounded glides is harder to maintain than one between two front glides. Back vowels and glides are characterized by a low second formant, and the effect of pronounced lip rounding characteristic of back glides further reduces the distance between the first two formants. This means that the possibility of maintaining a qualitative difference between back rounded glides is limited relative to a distinction between front glides. We may conclude that the phonological analysis of [wa] and [oa] can be maintained, but in the case of back glide-vowel sequences and diphthongs a process of phonetic neutralization takes place, motivated by the relatively low frequency of [wa] sequences, and by the difficulty of maintaining a contrast between two back rounded glides [w] and [o].

## 6 Conclusion

The integrated production and perception study presented in this paper illustrates the importance of considering both phonetic and phonological information in testing a proposed phonological analysis and providing an accurate phonological description of a language. The distinct phonological representations proposed for front glide-vowel sequences and diphthongs in Romanian are supported by acoustic data and by the results of a perception experiment. At the same time, the study revealed a phonetic neutralization that takes place between the back glide-vowel sequence [wa] and the back diphthong [oa]. Acoustically there is no difference between the sequence and the diphthong, and listeners cannot reliably identify either one. This neutralization can be explained by a language specific difference in frequency, and by the difficulty of maintaining a contrast between two back rounded glides. It is therefore proposed that underlying /ua/ sequences undergoing glide formation have the same phonetic realization as the diphthong [oa].

## Acknowledgements

I would like to thank Donca Steriade and an anonymous reviewer for helpful comments. Any errors are my own.

## References

- ACADEMIA ROMÂNĂ (1995). *Îndreptar ortografic, ortoepic și de punctuație* [Handbook of orthography, orthoepy, and punctuation]. Bucharest: Institutul de lingvistică ‘Iorgu Iordan’, Univers enciclopedic.
- CHITORAN, I. (2001). *The Phonology of Romanian: A Constraint-based Approach*. Berlin & New York: Mouton de Gruyter.
- CHITORAN, I. (to appear). The phonology and morphology of Romanian diphthongization. *Probus*.
- HAM, W. (1998). *Phonetic and Phonological Aspects of Geminate Timing*. Ph.D dissertation, Cornell University. (Published by Routledge Outstanding Dissertations in Linguistics, 2001.)
- KAYE, J. D. & LOWENSTAMM, J. (1984). De la syllabicité. In Dell, F., Hirst, D. & Vergnaud, J-R. (eds.), *La forme sonore du langage*, 123–159. Paris: Herman.
- LIBERMAN, A., DELATTRE, P., GERSTMAN, L. & COOPER, F. (1956). Tempo of frequency change as a cue for distinguishing classes of speech sounds. *Journal of Experimental Psychology* 52, 127–137.
- MERTUS, J. (1985). Brown Lab Interactive Speech Software (BLISS).
- MORIN, Y-C. (1976). Phonological tensions in French. In Hensey, F. & Luján, M. (eds.), *Current Studies in Romance Linguistics*. 37–49. Washington, DC: Georgetown University Press.

- REN, H. (1986). *On the Acoustic Structure of Diphthongal Syllables*. Ph.D. dissertation, UCLA. (Published as *UCLA Working Papers in Phonetics* 65.)
- ROSETTI, A. (1955). Cercetări experimentale asupra diftongilor românești [Experimental studies of the Romanian diphthongs]. *Studii și cercetări lingvistice* 5, 7–27.
- ROSETTI, A. (1959). *Recherches sur les diphtongues roumaines*. Bucarest & Kopenhagen: Editura Academiei RPR & Munksgard.
- ULIVI, A. (1975). Observații asupra structurii acustice a semiconsoanelor românești [y] și [w] [Remarks on the acoustic structure of the Romanian semi-consonants [y] and [w]]. *Fonetica și dialectologie* 9, 107–112.