BALTISTICA LI(1) 2016 149–168
doi: 10.15388/Baltistica.51.1.2249

Juris GRIGORJEVS
Latvian Language Institute of the University of Latvia

ACOUSTIC CHARACTERISTICS OF THE LATVIAN DIPHTHONGS PRODUCED BY MALE AND FEMALE INFORMANTS

Abstract. The aim of the present study was finding acoustic characteristics of the Latvian diphthongs irrespective of the speaker's gender. The method designed by the author allowed to compare diphthongs of different duration, thus revealing common tendencies both in the formant structures and in the diphthong paths on the vowel plane. A comparison of the diphthong paths with the location of the long monophthongs on psychophysical F2′ vs. F1 plane showed that the paths begin in the vicinity of the corresponding monophthong but end with an undershoot, i.e., not completely reaching the target frequencies. Dynamics of the formant trajectories of the Latvian diphthongs produced in the zero context in most cases display a two-phase pattern with a relatively short steady state of the first component (the longest in [ie] and [uo]: about 1/3 of the total duration) and a long transitionary phase with no steady state of the second component. The duration of the Latvian diphthongs (W) is between that of the short (V) and the long (Vː) monophthongs. The durational ratio of the short monophthongs to the diphthongs and the long monophthongs is V : W : Vː = 1 : 1.5 : 2.
Keywords: Latvian; diphthongs; acoustic phonetics; formant structure; acoustic characteristics; dynamics of the formant; formant trajectories; relative duration.

1.0 Introduction

The vowel system of Standard Latvian consists of 12 monophthongs (/i, iː, e, eː, æ, æː, ɑ, ɑː, ɔ, ɔː, u, uː/) and 10 diphthongs (/ɑ̯i̯, ɑ̯u̯, e̯i̯, e̯u̯, i̯e̯, i̯u̯, ɔ̯i̯, ɔ̯u̯, u̯i̯, u̯o̯/). Working on the chapter "Phonetics and Phonology" of the new academic Grammar of the Latvian Language (Auziņa et al. 2013) the traditional description and classification of speech sounds was revised by the author of the present article including phonetic and phonological analysis of the Latvian diphthongs. The study of the diphthongs was continued in subsequent years (Grigorjevs 2014a; 2014b). A conclusion is that the Latvian diphthongs should be viewed as a type of long vowels instead of a combination of two short ones. In the present article the data acquired for male and female pronunciation are used to study the influence of the speaker's gender upon the acoustic structure of the diphthongs.

1.1 Background

Until recently all the Latvian diphthongs were treated as two monophthongs joined in one syllable (Laua 1997, 25; Skujiņa 2007, 97). According to this definition a diphthong is viewed as a tautosyllabic sequence of two short monophthongs as opposed to a vowel combination, where a sequence of two monophthongs appears in heterosyllabic position. Consequently all the Latvian diphthongs have traditionally been regarded as biphonemes, while some linguists have suggested considering the diphthongs /ie/ and /uo/ as monophonemes due to their structure and functions (Auziņa et al. 2013, 46; Markus, Bonda 2014, 68–72). According to the traditional approach both components of the Latvian diphthongs (except /ie/ and /uo/ regarding their final component) are realized as the corresponding short monophthongs in respect to their articulatory and acoustic qualities. In dynamic spectrograms monophthongs are usually represented by quasi-stationary formant structures (at least in their middle part) forming the so called *steady states*. The approach viewing the Latvian diphthongs as tautosyllabic sequences of two monophthongs is very close to the ***duality (sequence) view*** described by Pamela Arlund, where two short monophthongs comprising a diphthong are linked by a gliding segment or transition which "displays rapid formant change from the onset steady state to the offset steady state" (Arlund 2006, 2). In order to qualify as a diphthong (according to the duality view) there have to be two steady states with a transitionary section in between (Arlund 2006, 22).

In the review by Burton Rosner and Brian Pickering of numerous phonetic studies conducted by several researchers it has been concluded that "some disagreement exists between diphthong paths in F2/F1 space and the positions of their assumed monophthongal components" which in its turn has raised the question "whether the disagreement represents a problem of transcription or whether diphthongs are incorrectly conceptualized as containing two phonemic pieces identifiable as monophthongs" (Rosner, Pickering 1994, 293). The author of this paper has also questioned the vowel sequence (duality) view (Grigorjevs 2009; 2012b; 2014a; 2014b). Applying the theoretical assumptions of Luciano Canepari (2007, 138–

139) to Latvian provided ground for considering all the Latvian diphthongs as monophonemes (Grigorjevs 2012b, 89; 2014a, 98–99; 2014b, 66; Auziņa et al. 2013, 44–52). Phonologically this decision is based on the observation that the Latvian diphthongs, if viewed as heterogeneous vowels, are capable of distinguishing the meaning in minimal pairs in the same way as the short and the long monophthongs (Auziņa et al., 2013 46; Grigorjevs 2014b, 50). Phonetically the decision of the monophonemic character of the diphthongs is based on the differences observed between the quality of the 1st and the 2nd component of any diphthong and the quality of the corresponding short or long monophthong produced in the same phonetic context (Grigorjevs 2014b, 50). It was observed that when the Latvian diphthongs were pronounced in 4 different positions of real words embedded into carrier phrases the steady state phase of the diphthong's initial or the final component was not regularly executed, especially in unstressed positions (Grigorjevs 2009, 45–46). It led to the conclusion that the steady state phases were not essential for the perception of the Latvian diphthongs. Noting the differences between the quality of diphthong components and seemingly corresponding monophthongs led to the conclusion that the transition phase provides sufficient information to perceive the diphthong, and its endpoints do not necessarily have to coincide with the formant structures of the corresponding monophthongs. These findings support ***unity (single vowel) view***, "which states that diphthongs are simply a vowel that has continuously changing formant qualities" (Arlund 2006, 2), and the presence of two steady states corresponding to monophthongs is not necessary. In her thesis Arlund mentions that "monophthongs are also inherently dynamic and that perception of them depends on their dynamism and their endpoints more than on their static portions" (Arlund 2006, 31), and therefore the spectral change is not a unique feature of diphthongs. If so, both the most popular views (duality and unity) upon diphthongs have their weak points.

It appears there is still no single approach to the interpretation of monophthongs and/or diphthongs. Considering that only few vowels appear as single phonemes in everyday communication, monophthongs and diphthongs are usually studied in real words or utterances. Often the impact of the phonetic context, stress location, speech tempo and other factors influencing the vowel quality is neglected which leads to a diverse description of the vowel quality. The author of the present article strongly believes that while acquiring a language we do not store in memory the whole variety of different realizations of the same phoneme. Instead we form the percept of some prototypical realization of each phoneme and its variations depending on the speech tempo, phonetic context etc. To ascertain the quality of vowel prototypes the author has chosen to analyse vowels produced in isolation, i.e. in the zero contexts, because this is the closest approximation of the mental prototype in production. As a result of analyses of the acoustic and auditory characteristics of the prototypical Latvian monophthong targets it has been found that members in the pair of short vs. long monophthong differ only by quantity (Grigorjevs 2008, 100–101; 2012a, 157). The inherent spectral change of the monophthongs was not addressed in these studies, because monophthongs were treated as quasi-stationary constructs. In their extensive work on the vowel perception and production Rosner and Pickering have pointed to the possible solution of the problem regarding diphthongs as full-fledged members of the vowel system – "the acoustic paths of diphthongs produced by an individual speaker must be compared with F2/F1 positions of isolated vowels produced by the same speaker" (Rosner, Pickering 1994, 293). This suggestion was followed when in a recent study by Grigorjevs (2014b) the dynamic structure of the Latvian long isolated vowels (both long monophthongs and diphthongs) was investigated registering the mean pitch and formant frequency values for 10 durational sections of each vowel. The data acquired for the long monophthongs showed very little variation of formant frequencies in the middle part (about ⅔ of the total duration) of each monophthong, and therefore all the points based on the values obtained for 10 durational sections except for the initial and/or the final one fell inside 1 bark circles representing quality zone of each monophthong in F2′/F1 plane (Grigorjevs 2014b, 60; see also figures on 59–62) thus proving the quasi-steadiness of the formant structures. The data acquired for the diphthongs showed much greater variation where even the shortest diphthong path (the diphthong [ei] produced by informant 1) exceeded 2.5 bark interval "signalizing about the change of perceptual quality and duality of its terminal segments" (Grigorjevs 2014b, 62). The study of acoustic targets of the Latvian diphthongs produced by five male informants has allowed drawing conclusions that at least phonetically the diphthongs are long gliding vowels instead of being sequences of two successive short monophthongs (Grigorjevs 2014a, 100; 2014b, 65–66), because only the first component of most diphthongs forms a relatively short steady state phase, in most cases resembling the formant structure of the corresponding long monophthong, while the length of this phase appears to be speaker dependent. The duration of the diphthongs has been found shorter than that of the long monophthongs (Grigorjevs 2014a, 100; 2014b, 64–66) which also casts doubt on the interpretation of the Latvian diphthongs as a mere combination of two short monophthongs in one syllable (Laua 1997, 25).

1.2 Aims

The aim of this study is to examine whether the structure characteristic of the Latvian diphthongs produced in the zero contexts by female speakers is the same as that observed in male speakers' production, thus finding common acoustic characteristics of the Latvian diphthongs irrespective of the speaker's gender. To achieve this goal the diphthong paths in psycho-physical F2′/F1 plane have to be examined in relation to the corresponding monophthongal targets, and a comparison of the male and female data has to be made. Also the formant schemes based on the mean values representing the production of the diphthongs by groups of male and female informants need to be compared to find if speakers of both genders employ the same strategy when pronouncing the Latvian diphthongs.

2.0 Materials and methods

The informants were 10 speakers (5 male and 5 female) of the age between 16 and 39. The age group was evenly split into 5 subgroups (16-20, 21-25, 26-30, 31-35 and 36-39) and one male and one female informant from each subgroup was recorded to achieve as even coverage of the whole age group as possible. The acoustic differences caused by the age of informant were not expected, therefore any differences in production were regarded as individual.

The recordings were performed using a computer, the USB audio capture device *EDIROL UA–25* and the *headset* condenser microphone *AKG C520*. All the informants produced all the Latvian vowels in sequences (e.g., ***Saka baib arī*** – ***baib*** – ***ai*** or ***Saki dīd arī*** – ***dīd*** – ***ī***) – i.e., a monosyllabic word (***CVC***[1]) in the carrier phrase followed after a pause by a monosyllabic word (***CVC***), and after a pause – by a vowel in zero context (***#V#***). Each sequence

[1] ***CVC*** – symmetric syllable where the initial and final ***C*** is the same – any of the Latvian consonants /b, ʦ, ʧ, d, ʣ, ʤ, f, g, ɟ, x, j/ʝ, k, c, l, ʎ, m, n (ŋ), ɲ, p, r, s, ʃ, t, ʋ/v, z, ʒ/, ***V*** – any of the Latvian vowels /i, e, æ, ɑ, ɔ, u, iː, eː, æː, ɑː, ɔː, uː, ie, iu, ei, eu, ɑi, ɑu, ɔi, ɔu, ui, uo/

was repeated 3 to 4 times. The vowels were produced with so called "falling syllable intonation", i.e., unmarked realization.

To acquire information about the vowels that are the closest approximation to the vowel targets of the Latvian language, the material for the analysis included vowels produced in isolation, i.e., in the zero context (***#V#***). Four instances of each vowel's production in the zero context were acoustically analysed using software *WaveSurfer* (v. 1.8.5., Kåre Sjölander and Jonas Beskow). The values of each diphthong's total duration were registered, as well as the frequency values of the fundamental (f0) and the first four formants (F1, F2, F3 and F4) were obtained as mean values for 10 equal durational intervals. To allow comparison of formant trajectories and vowel paths of the long monophthongs with those of the diphthongs the same methodology was applied to obtain values for 10 durational intervals of the long monophthongs, too. The data acquired for the long monophthongs showed very little variation of formant frequencies in the middle part (about ⅔ of the total duration) (Grigorjevs 2014b, 59–62), therefore it was decided to view the long monophthongs in a more traditional manner, i.e., as quasi-stationary constructs. Using software *WaveSurfer* four productions of each long and short monophthong by each of the informants were measured marking the major part of the vowel (leaving out of the marked region the very beginning and the very end of it) and obtaining statistical mean values of the fundamental and the first four formants for the marked region.

The statistical analysis (using *MS Excel* and *IBM SPSS Statistics 22*) of the acquired data was performed calculating the mean values and standard deviations, as well as the values of minimum and maximum for each analysed parameter. The obtained mean values were used to create graphical illustrations (Figure 1–10) in *MS Excel*.

3.0 Results and Discussion

The mean values of the fundamental frequency, formant frequencies and duration of each vowel were calculated for each informant individually, as well as for the groups of the male and the female informants.

3.1 Diphthong paths in psycho-physical F2'/F1 plane

To examine whether the structure characteristic of the Latvian diphthongs produced in the zero context by female speakers coincides with the one in the production by male speakers, first the plots in psycho-physical F2′/F1 plane were made separately for the vowels produced by the male (Figure 1) and the female speakers (Figure 2).

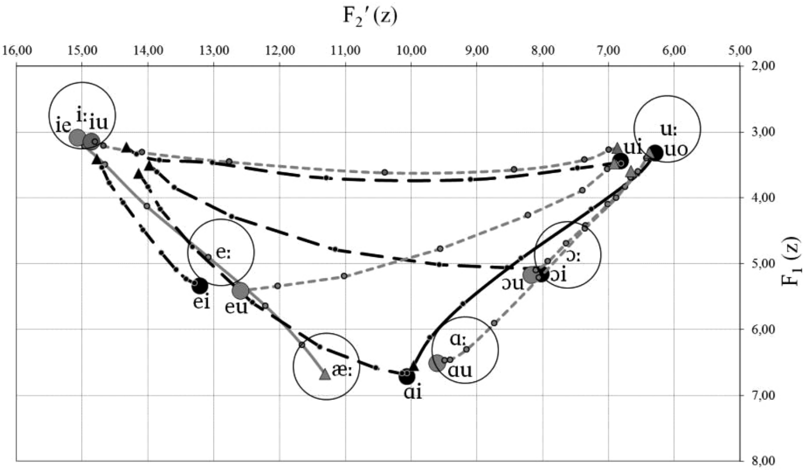


Figure 1. **The placement of the long Latvian vowels produced by the male informants in psycho-physical F_2'/F_1 vowel plane**

(all long monophthongs are shown by large white circles; the diphthong [ie] – by a grey solid line; the diphthong [uo] – by a black solid line; the diphthongs ending in [i]-like component – by black dashed lines; the diphthongs ending in [u]-like component – by grey dashed lines; the initial interval of each diphthong is marked by a large black or grey dot, the final interval – by a corresponding triangle, but the intermediate intervals – by small dots of the corresponding colour)

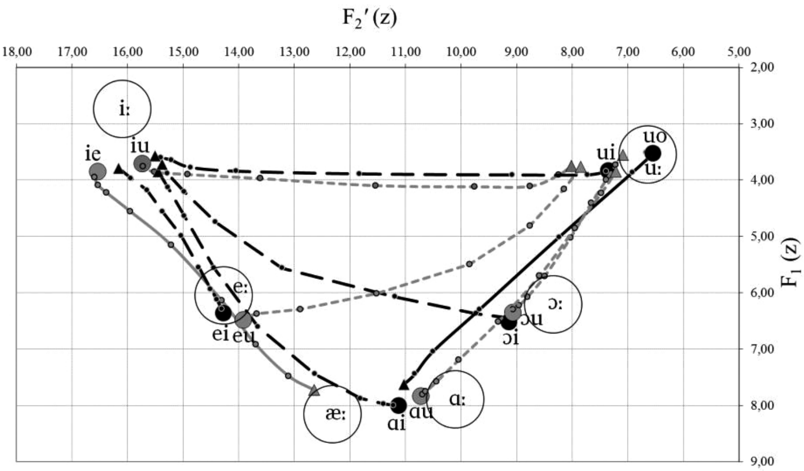


Figure 2. **The placement of the long Latvian vowels produced by the female informants in psycho-physical F_2'/F_1 vowel plane**

(all long monophthongs are shown by large white circles; the diphthong [ie] – by a grey solid line; the diphthong [uo] – by a black solid line; the diphthongs ending in [i]-like component – by black dashed lines; the diphthongs ending in [u]-like component – by grey dashed lines; the initial interval of each diphthong is marked by a large black or grey dot, the final interval – by a corresponding triangle, but the intermediate intervals – by small dots of the corresponding colour)

The long monophthongs (the mean data calculated for central ⅔) in these plots are represented by large circles while the diphthongs are represented by trajectories, i.e. vowel paths whose beginnings (large dots) correspond to the first, but the ends (triangles) – to the last of the measured 10 segments of equal duration.

To take into account the influence of F3 and F4 on the vowel quality the values of the effective second formant (F2′) calculated with Bladon and Fant's formulae (Bladon, Fant 1978, 3) were used instead of the measured values of F2. To balance the effect of F1 and higher formants upon the perceptual quality of vowels, before plotting all the values were transformed to psychophysical units *barks* (z) applying Traunmüller's formulae (1988, 97–100). The long monophthongs are represented by circles whose diameter is 1 z to demarcate areas of the same perceptual quality (Iivonen 1987).

The diphthong paths in psycho-physical F_2'/F_1 plane were examined in relation to the corresponding monophthongal targets. It can be observed that the diphthong paths reflecting the production of the male (Figure 1) and the female (Figure 2) informants generally follow the same pattern, with an exception of the diphthongs [ie] and [iu] pronounced by the females. Although the frequency of F1 of the first/initial section (represented by large dots in Figure 1 and 2) for the diphthongs [ie] and [iu] is increased in comparison to that of [iː] both in the male and the female production, the increase is substantially (approx. 0.5 z) larger for the female informants. Reviewing the mean data for each of the informants revealed that the increase of F1 value ranges from 0 z to 0.7 z for the male and from 0.8 to 1.1 z for the female informants. Thus it can be concluded that the female informants tend to produce the first component of the diphthongs [ie] and [iu] more open than the male informants while the strategy for producing the other diphthongs is the same. In general, the initial components of all diphthongs are more open than the corresponding monophthongs, and to some extent they are influenced by the quality of the final components. The final components demonstrate target undershoot (with an exception of [ɔu] ending in the zone of [uː]) which increases along with differences in the articulation of both components. These observations are relevant for diphthongs produced by all the informants (both male and female). Since measurements have been made at equal intervals, the diphthong paths in F_2'/F_1 plane provide also some information about the diphthong dynamics. Densely located points signal about the quasi-stationary phases where there is very little movement of the speech organs. In both figures (Figure 1 and 2) the close location of the points can be observed in the initial phases of the diphthongs, while the greatest distances between the points (signalling about the highest velocity of the articulatory movement) are observed for 5–6 points in the middle of each diphthong.

3.2 Formant schemes

To make more precise judgements about the duration and quality of the quasi-stationary phases and formant changes during the transitionary phases it was decided to use a schematic representation of the first four diphthong formants based on the calculated mean values (Figure 3–10).

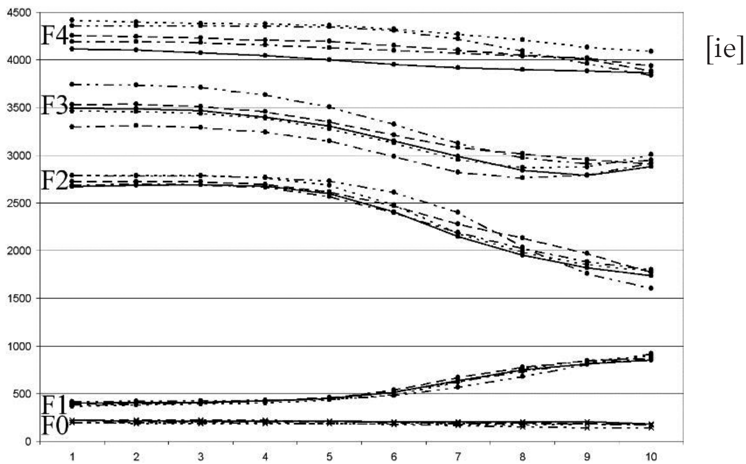
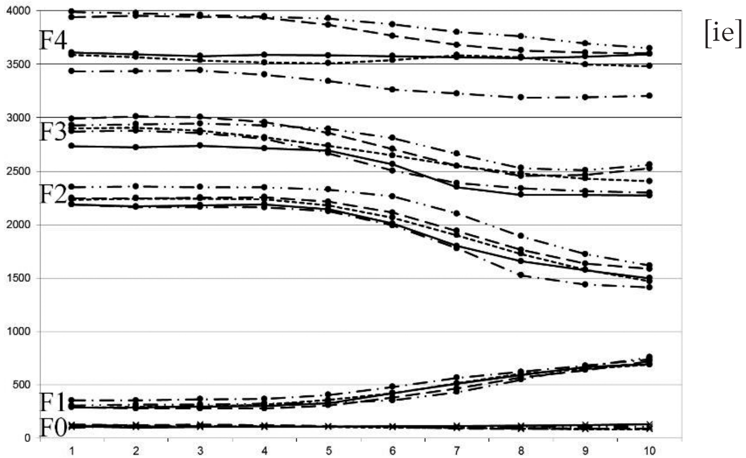


Figure 3. **The formant schemes of the Latvian diphthong** [ie] **produced by five female informants**

(Fem1 – a solid line, Fem2 – a dashed line, Fem3 – a dotted line, Fem4 – a dashed-dotted line, Fem5 – a dashed-double-dotted line; each line represents the mean values calculated from 4 productions of the diphthong; the tick marks at the bottom demarcate 10 durational segments, F0 – the fundamental)

Examining the formant schemes of the diphthong [ie] produced by five female informants (Figure 3) it can be observed that the most essential interspeaker differences concern F3 and F4 although the general tendencies are similar for all the informants.

The interspeaker variation in the production of the diphthong [ie] is even more pronounced for the male speakers (Figure 4), however, the overall



F i g u r e 4. **The formant schemes of the Latvian diphthong** [ie] **produced by five male informants**

(M1 – a solid line, M2 – a dashed line, M3 – a dotted line, M4 – a dashed-dotted line, M5 – a dashed-double-dotted line; each line represents the mean values calculated from 4 productions of the diphthong; the tick marks at the bottom demarcate 10 durational segments, F0 – the fundamental)

pattern is very close to the formant schemes of the diphthong pronounced by the female informants (cf. Figure 3 and 4). Analysing the formant schemes of the diphthong [ie] (Figure 3–4) it can be concluded that although differing in the frequency values the formant trajectories of both the male and female informants follow nearly the same pattern. It can also be concluded that despite slight interspeaker differences within each gender group the mean values calculated for this group produce formant trajectories that reflect general tendencies in the production of each individual informant. On the basis of this conclusion it was decided to use the mean values (calculated from 20 productions of each diphthong) to compare the male and female production.

To make a more rational comparison of the male and female data, it was decided to perform the uniform normalization of the female data using the coefficient k=21% calculated from the mean F2 and F3 values of the monophthongs [i][2] and [i:][3] according to F a n t's formula $\boldsymbol{k=\frac{1}{2}(k_{2i}+k_{3i})}$ (1975,

[2] For the male speakers F_2=2205 Hz, F_3=2784 Hz; for the female speakers F_2=2666 Hz, F_3=3345 Hz.

[3] For the male speakers F_2=2230 Hz, F_3=2800 Hz, for the female speakers F_2=2703 Hz, F_3=3430 Hz.

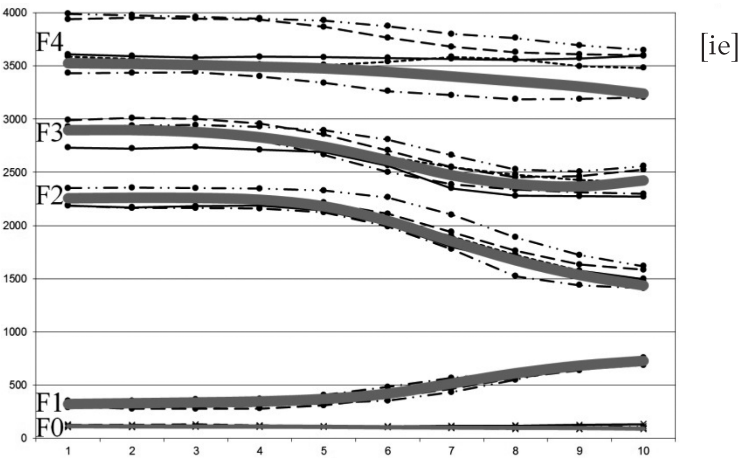


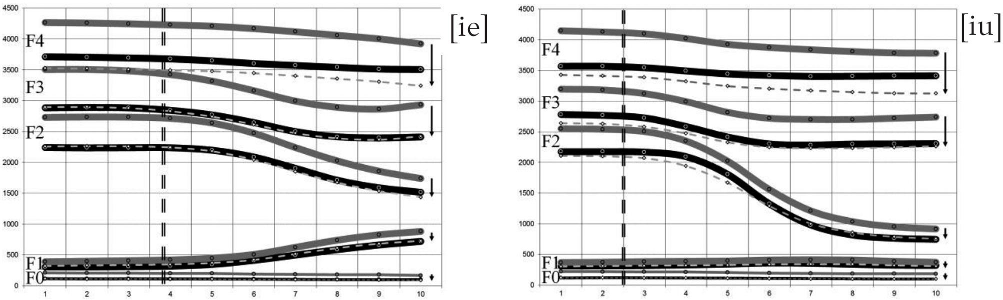
Figure 5. **The formant schemes of the Latvian diphthong** [ie] **produced by five male informants with the superimposed formant scheme of** [ie] **representing the normalized mean values of female informants**
(M1 – a solid line, M2 – a dashed line, M3 – a dotted line, M4 – a dashed-dotted line, M5 – a dashed-double-dotted line; each line represents the mean values calculated from 4 productions of the diphthong; the thick grey lines represent the mean values calculated from 20 productions of the diphthong by five females and normalized with k=21%)

8). The coefficients $\boldsymbol{k_{2i}}$ and $\boldsymbol{k_{3i}}$ were calculated from the mean F2 and F3 values of both monophthongs [i] and [iː] using the formulae $\boldsymbol{k_{2i}=F_{2i(female)}/F_{2i(male)}}$ and $\boldsymbol{k_{3i}=F_{3i(female)}/F_{3i(male)}}$. For the monophthongs [i] and [iː] the coefficient normalizing F2 values was equal to $k_{2i(:)}$=21% while the coefficients normalizing F3 values differed (k_{3i}=20% for [i] and $k_{3i:}$=22% for [iː]) giving the same average value $k_{3i(:)}$=21%. This value is close to the value characterizing the length differences between male and female vocal tracts. Uniform normalization of the female formant data with k=21% results in formant patterns close to those of the male with a slightly lower (over-normalized) F4 still being in range of the interspeaker differences observed for the speakers of the same gender (Figure 5).

All the further analysis of the formant patterns of the Latvian diphthongs in the present study is carried out using the mean values for each of the gender groups. To find out if the speakers of both genders employ the same strategy in the pronunciation of the diphthongs, the formant trajectories of the male (the thick black lines) and the female (the thick grey lines) production of the same diphthong are plotted in the same graph along with the trajectories (the thin dashed grey lines) based on the normalized (k=21%) female data (Figure 6–10).

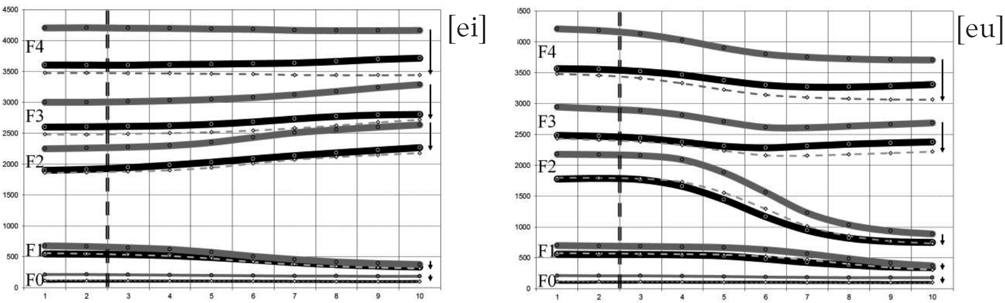
The trajectories representing the fundamental frequency are plotted at the bottom of these graphs (the male – the thin black lines, the female – the thin grey lines) and marked by F0.

The downward pointing arrows on the right of each figure show the displacement of each formant track after normalization of the female data



F i g u r e 6. **The formant schemes of the Latvian diphthongs** [ie] **(left) and** [iu] **(right)**

(the thick black lines represent the male production, the thick grey lines – the female production, the thin dashed grey lines – the female production normalized with k=21%; the dashed vertical double line demarcates the end of the steady state of the first component)



F i g u r e 7. **The formant schemes of the Latvian diphthongs** [ei] **(left) and** [eu] **(right)**

(the thick black lines represent the male production, the thick grey lines – the female production, the thin dashed grey lines – the female production normalized with k=21%; the dashed vertical double line demarcates the end of the steady state of the first component)

with k=21%. The double vertical dashed lines in these plots demarcate the end of the quasi-stationary phase of the first component basing the judgement about the steadiness (a rather rough one) upon the trajectories of F1 and F2 reflecting the male and normalized female data. It is observed that if the judgement about the quasi-stationary phase had to be based upon the data of all four measured formants (Figure 6–10), the length of this phase would seldom exceed $^{1}/_{10}$ of a diphthong.

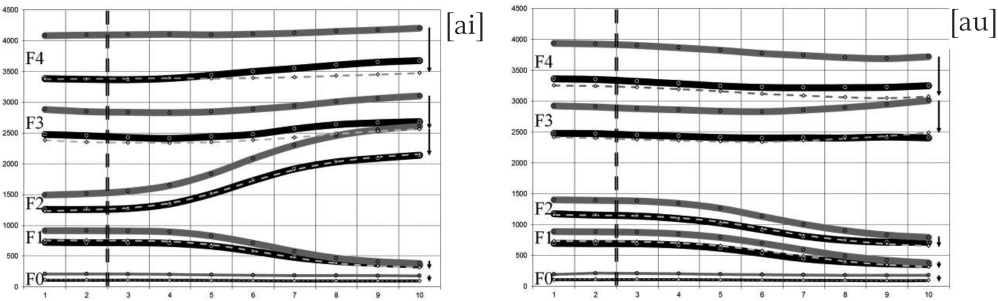


Figure 8. **The formant schemes of the Latvian diphthongs [ɑi] (left) and [ɑu] (right)**

(the thick black lines represent the male production, the thick grey lines – the female production, the thin dashed grey lines – the female production normalized with k=21%; the dashed vertical double line demarcates the end of the steady state of the first component)

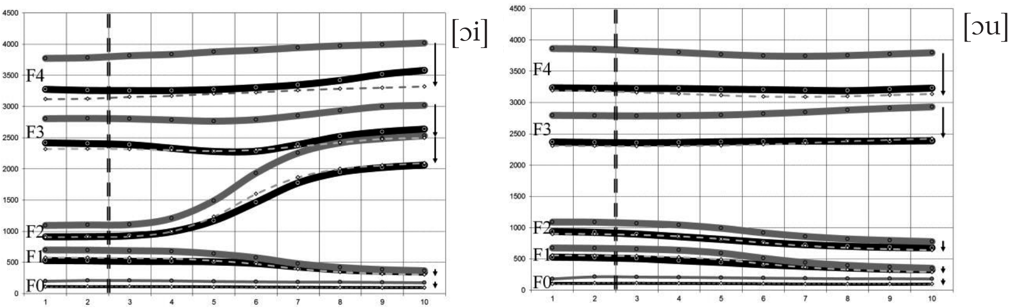
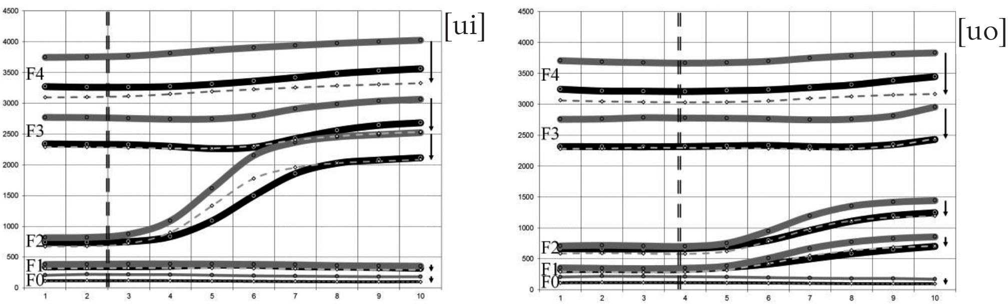


Figure 9. **The formant schemes of the Latvian diphthongs [ɔi] (left) and [ɔu] (right)**

(the thick black lines represent the male production, the thick grey lines – the female production, the thin dashed grey lines – the female production normalized with k=21%; the dashed vertical double line demarcates the end of the steady state of the first component)

The uniform normalization of the female data results in a close approximation of the trajectories of F1, F2, and in many cases also of F3 to the male data, while F4 in most cases is over-normalized and has lower frequency than in the male pronunciation.

As mentioned before, this difference in the frequency of F4 (Figure 6–10) is not greater than the variation observed for different speakers of the same gender (Figure 3–5). In fact, in many cases it is not greater than the differences between the formant trajectories obtained for 4 productions of the same diphthong by the same informant (not illustrated in this article).



F i g u r e 10. **The formant schemes of the Latvian diphthongs** [ui] **(left) and** [uo] **(right)**

(the thick black lines represent the male production, the thick grey lines – the female production, the thin dashed grey lines – the female production normalized with k=21%; the dashed vertical double line demarcates the end of the steady state of the first component)

In general, the formant trajectories of the diphthongs produced by the male and female informants follow the same pattern (Figure 6–10). It does not matter if the formant trajectories reflecting the measured (the thick grey lines) or normalized (the thin dashed grey lines) female data are compared with the formant trajectories representing the male data (the thick black lines). The differences between F1 and F2 trajectories of the diphthongs pronounced by the male and the female informants can mostly be attributed to the differences in the length of the resonator since the uniform normalization eliminates them nearly completely. Although after normalization the F3 trajectories of the female diphthongs do not completely overlap with the corresponding male trajectories, they display the same pattern (with an exception of the diphthong [ɑu], where the female values of F3 at the end of the diphthong are higher than in the beginning, while in the male pronunciation they are lower at the end).

If the timing of the articulatory gestures is evaluated on the basis of the formant schemes, it can be concluded that the transitionary phase starts a little earlier in the female diphthongs [ie] (Figure 6*left*), [iu] (Figure 6*right*), [ɑi] (Figure 8*left*) and [ui] (Figure 10*left*) than in the same diphthongs produced by the male informants, while it starts a little later in the female diphthong [eu] (Figure 7*right*) and [ɔu] (Figure 9*right*) production.

The formant schemes reviewed in this study (Figure 6–10) suggest that the Latvian diphthongs do not display the conventional three phase structure (the steady state phase of the first component – the transitionary phase – the steady state phase of the second component) that could confirm the duality (sequence) view (Arlund 2006, 22). They display a two phase structure (the steady state phase of the first component – the transitionary phase) instead, where the transitionary phase takes at least $^2/_3$ of the diphthong's total duration and therefore cannot be considered an offglide from the base vowel. The duration of the steady state phase of the first component in most diphthongs is about $^1/_5$ of the diphthong's total duration, while it is longer (about $^1/_3$ of the total duration) in the diphthongs [ie] and [uo] (Figure 6*left* and Figure 10*right*). The steadiness of the first component is questionable because in normal speech conditions diphthongs very rarely appear without a phonetic context. Taking into account that the transition from the preceding consonant constitutes about 0.05 initial seconds of a vowel, we have to conclude that the major part of the steady state phase of the diphthong's initial component (Figure 6–10) includes this transition[4] leaving a very short or no quasi-stationary phase at all.

The assumption about the absence of quasi-stationary phases when diphthongs are produced in a consonantal context is supported by results of an earlier study of the Latvian vowels (short and long monophthongs, and diphthongs) in different positions in real words embedded into carrier phrases (Grigorjevs 2009). These results suggest qualitative and quantitative changes of all the Latvian vowels caused by the speech tempo, phonetic context, location in the word in relation to the primary stress, etc. Some kind of steadiness of the initial components of diphthongs is preserved in the word initial position

[4] Considering the mean duration of isolated diphthongs (Table 1) a $^1/_{10}$ in the presented formant schemes represents the duration of about 0.03 s.

and of the final components – in the word final position, while it is lost in the word medial position. All the vowels of this study display the effect of the acoustic centralization in unstressed syllables. In stressed syllables the close and mid monophthongs and the corresponding initial components of the diphthongs demonstrate an increased opening while preserving approximately the same placement in front-back dimension as isolated vowels. The final components of these diphthongs demonstrate a larger undershoot of their acoustic targets than in isolated diphthongs (cf. Figure 1 in this article with 4. attēls (*Figure 4*) in G r i g o r j e v s 2009, 44). The acoustic differences from the monophthongs produced in isolation increase if the long monophthongs are produced in unstressed syllables; likewise increase the differences between the long monophthongs and the initial and the final components of the diphthongs. The fact that the qualities of the initial and the final phases of a diphthong are not equal with the qualities of the corresponding monophthongs supports the unity (single vowel) view (cf. A r l u n d 2006, 2).

T a b l e 1. **The duration of the Latvian diphthongs produced in the zero context**

Male		**Female**	
Vowel	Duration[a] (s)	Vowel	Duration[a] (s)
ie	0.284	ie	0.295
iu	0.303	iu	0.333
ei	0.301	ei	0.303
eu	0.318	eu	0.314
ɑi	0.315	ɑi	0.334
ɑu	0.311	ɑu	0.315
ɔi	0.311	ɔi	0.321
ɔu	0.310	ɔu	0.324
ui	0.298	ui	0.308
uo	0.295	uo	0.293
Mean[b]	**0.305**	***Mean***[b]	**0.314**
SD	*0.010*	***SD***	*0.014*
Min	0.284	Min	0.293
Max	0.318	Max	0.334

[a] *The mean values of each diphthong calculated from 20 measured values obtained from 4 productions by each of five informants of each gender;*

[b] *the values of mean, standard deviation (SD), minimum and maximum calculated from the above listed mean values of all diphthongs of each gender.*

Another factor that speaks in favour of this view is the diphthong duration. If a diphthong were a mere combination of two short monophthongs, one could expect its duration to be equal to the sum of both monophthong durations.

Table 2. **The duration of the Latvian monophthongs produced in the zero context**

Male				Female			
Vowel	Duration[a](s)	Vowel	Duration[b](s)	Vowel	Duration[a](s)	Vowel	Duration[b](s)
i	0.194	iː	0.416	i	0.184	iː	0.384
e	0.209	eː	0.425	e	0.196	eː	0.398
æ	0.220	æː	0.424	æ	0.206	æː	0.406
ɑ	0.202	ɑː	0.421	ɑ	0.198	ɑː	0.390
ɔ	0.221	ɔː	0.430	ɔ	0.197	ɔː	0.405
u	0.198	uː	0.421	u	0.188	uː	0.393
Mean[c]	**0.207**	***Mean***[d]	**0.423**	***Mean***[c]	**0.195**	***Mean***[d]	**0.396**
SD	*0.011*	***SD***	*0.005*	***SD***	*0.008*	***SD***	*0.008*
Min	0.194	Min	0.416	Min	0.184	Min	0.384
Max	0.221	Max	0.430	Max	0.206	Max	0.406

[a] *The mean values of each short monophthong calculated from 20 measured values obtained from 4 productions by each of five informants of each gender;*

[b] *the mean values of each long monophthong calculated from 20 measured values obtained from 4 productions by each of five informants of each gender;*

[c] *the values of mean, standard deviation (SD), minimum and maximum calculated from the above listed mean values of all short monophthongs of each gender;*

[d] *the values of mean, standard deviation (SD), minimum and maximum calculated from the above listed mean values of all long monophthongs of each gender.*

Since long monophthongs are sometimes regarded as combinations of two short monophthongs of the same quality, the duration ratio of short vs. long monophthongs could be expressed as 1 : 2. Indeed, the data of the present study (Table 2) and of an earlier study of isolated monophthongs (Grigorjevs 2008, 34–35) indicate that on the average the duration ratio of short vs. long monophthongs is about 1 : 2 (1 : 2.04 in the present study, 1 : 2.11 in Grigorjevs 2008). If two short monophthongs of equal quality contribute to the duration ratio 1 : 2, one could expect that the ratio of short monophthongs vs. diphthongs (a diphthong being treated as a combination of two monophthongs of different quality) will be even higher. The results of the present study indicate that the average ratio of short monophthongs

(Table 2) vs. diphthongs (Table 1) is about 1 : 1.5 (1 : 1.47 for the male and 1 : 1.61 for the female informants), while the average ratio of diphthongs (Table 1) vs. long monophthongs (Table 2) is about 1 : 1.3 (1 : 1.39 for the male and 1 : 1.26 for the female informants).

It can be seen that despite minor differences the duration of diphthongs produced by the male and female informants follows the same pattern – diphthongs are 1.5 times longer than short monophthongs and 1.3 times shorter than long monophthongs. In the author's opinion this result along with the other results mentioned above contradicts the assumption that diphthongs are compounds of two short vowels in one syllable.

5.0 Conclusions

The results of the present study indicate that both female and male informants use the same strategies producing the Latvian diphthongs with a minor exception that the female informants tend to produce the first component of the diphthongs [ie] and [iu] more open than the male informants. In general, the initial components of all diphthongs are more open than the corresponding monophthongs, and to some extent they are influenced by the quality of the final components. The final components demonstrate target undershoot (with an exception of [ɔu] ending in the zone of [uː]) which increases along with the differences in the articulation of both components.

The main differences in the acoustic quality of the Latvian diphthongs reflected in the formant schemes can be explained by the differences in the length of the male and female vocal tracts. The acoustic differences are nearly completely reduced by the uniform normalization with k=21%. The relative timing of the articulatory gestures is similar for the representatives of both genders. The Latvian diphthongs produced by all informants reveal the two phase structure consisting of a relatively short initial steady state (about $^1/_5$ of the total duration) and a long transitionary phase. The duration of the Latvian diphthongs is between that of the short and the long monophthongs. The duration values, as well as the predominance of the transitionary phase with no steady state of the final component in the formant schemes suggest viewing the Latvian diphthongs as long gliding vowels in accordance with the unity (single vowel) view.

Acknowledgements

This study has been financially supported by the grant awarded by the Latvian Council of Science to the research project N 148/2012 *"The acoustic* *characteristics of the sound system of Standard Latvian by age groups (5–15, 16–39, 40–59, 60–80)*". The author expresses his gratitude to Professor Maija Brēde for reading and commenting on an earlier draft of this paper.

VYRŲ IR MOTERŲ IŠTARTŲ LATVIŲ KALBOS DIFTONGŲ AKUSTINĖS CHARAKTERISTIKOS

Santrauka

Šio straipsnio tikslas – nustatyti ir aprašyti būdingąsias spektrines latvių kalbos diftongų charakteristikas, kurių nelemia informantų lytis. Autoriaus sukurta metodika leidžia tiriamuosius skirtingos trukmės izoliuotai ištartus garsus kuo objektyviau palyginti įvairiais aspektais: aptariami ne tik bendrieji diftongų formantinės sandaros ypatumai, formančių dinamika (pirmosios keturios formantės, išreikštos hercais), bet ir bandoma nustatyti garsinę tiriamųjų segmentų sudėtį pateikiant diftongų formančių kitimo trajektorijas psichofizinėje latvių kalbos balsių erdvėje (ordinačių ir abscisių ašyse atidėjus atitinkamai pirmąją ir vadinamąją efektyviąją antrąją formantes, išreikštas barkais).

Lyginamoji gautų rezultatų analizė rodo, kad tiriamųjų diftongų formančių trajektorijos psichofizinėje erdvėje prasideda panašioje vietoje kaip ir ilgieji atitinkamos kokybės monoftongai, tačiau diftongų formančių trajektorijų pabaiga paprastai nesutampa ar nesiekia atitinkamų monoftongų lokalizacijos.

Formančių dinamikos analizė rodo, kad daugeliu atvejų izoliuotai ištarti diftongai sudaryti iš šių dviejų komponentų: 1) gana trumpos stacionariosios dalies (ilgiausia ši dalis [*ie*] ir [*uo*] atveju – sudaro maždaug 1/3 viso diftongo) ir 2) ilgo kintamojo elemento (t. y. tokio, kuris neturi stacionariosios dalies, formančių dinamika nuolat kinta). Nustatyta, kad latvių kalbos diftongai ([W]) yra ilgesni už trumpuosius ([V]), tačiau trumpesni už ilguosius ([V:]) latvių kalbos monoftongus. Apibendrintais duomenimis, kiekybės santykiai tokie: [V] : [W] : [V:] = 1 : 1,5 : 2.

REFERENCE LIST

Arlund, Pamela 2006, *An acoustic, historical, and developmental analysis of Sarikol Tajik diphthongs*, Theses for the Degree of Doctor of Philosophy, The University of Texas at Arlington. https://dspace.uta.edu/bitstream/handle/10106/438/umi-uta-1588.pdf?sequence=1 (27.03.2012).

Auziņa, Ilze, Juris Grigorjevs, Inese Indričāne, Dace Markus 2013, Fonētika un fonoloģija [Phonetics and Phonology], in Daina Nītiņa, Juris Grigorjevs (eds.), *Latviešu valodas gramatika* [Grammar of the Latvian Language], Rīga: LU Latviešu valodas institūts, 22–137.

Bladon, Anthony R. W., Gunnar Fant 1978, A two-formant model and the cardinal vowels, *Speech Transmission Laboratory – Quarterly Progress and Status Report,* 1978(1), 1–8.

Canepari, Luciano 2007, *Natural phonetics and tonetics (Articulatory, auditory & functional)*, München: LINCOM GmbH.

Fant, Gunnar 1975, Non-uniform vowel normalization, *Speech Transmission Laboratory – Quarterly Progress and Status Report* 1975(2–3), 1–19.

Grigorjevs, Juris 2008, *Latviešu valodas patskaņu sistēmas akustisks un auditīvs raksturojums* [Acoustic and auditory characteristics of the Latvian monophthong system], Rīga: LU Latviešu valodas institūts.

Grigorjevs, Juris 2009, Latviešu valodas divskaņu akustisks raksturojums [Acoustic characteristics of the Latvian diphthongs], *Latvijas Universitātes Raksti* 746 *("Valodniecība: Latvistika un somugristika")*, 40–47.

Grigorjevs, Juris 2012a, Acoustic and auditory characteristics of the Latvian monophthong system, *Linguistische Berichte* 2012 (230), 155–182.

Grigorjevs, Juris 2012b, Latviešu valodas divskaņu fonētiskā un fonoloģiskā interpretācija [Phonetic and phonological interpretation of the Latvian diphthongs], *Vārds un tā pētīšanas aspekti* 16, 78–89.

Grigorjevs, Juris 2014a, Latviešu valodas divskaņu akustiskais raksturojums informantu vecuma grupā no 16 līdz 39 gadiem [Acoustic characteristics of the Latvian diphthongs for informants in the age group from 16 to 39 years], *Vārds un tā pētīšanas aspekti* 18(1), 88–100.

Grigorjevs, Juris 2014b, Dynamics of the Latvian long vowels, *Linguistica Lettica* 22, 48–67.

Iivonen, Antti 1987, The critical band in the explanation of the number of possible vowels and psychoacoustical vowel distances. (English summary), *Helsingin yliopiston fonetiikan laitoksen monisteita* 12, Helsinki: University of Helsinki.

Laua, Alīse 1997, *Latviešu literārās valodas fonētika* [Phonetics of the Standard Latvian] (4th ed.), Rīga: Zvaigzne ABC.

Markus, Dace, Dzintra Bonda 2014, *Ievads fonoloģijā* [Introduction to Phonology], Rīga: Zinātne.

Rosner, Burton, John Brian Pickering 1994, *Vowel perception and production*, New York: Oxford University Press.

Skujiņa, Valentīna (ed.) 2007, *Valodniecības pamatterminu skaidrojošā vārdnīca* [Explanatory dictionary of the key terms of Linguistics], Rīga: Valsts valodas aģentūra.

Traunmüller, Hartmut 1988, Analytical expressions for the tonotopic sensory scale, *PERILUS* 8, 93–102.

Juris GRIGORJEVS
LU Latviešu valodas institūts
Akadēmijas laukums 1
LV-1050 Rīga
Latvia
[*jugrig@latnet.lv*]