

Zora JACHOVA
Lidija RISTOVSKA
Ljudmil SPASOV

UDK:159.937.24:159.946.3
Original research paper

FORMANT TRANSITIONS AS ACOUSTIC CUES FOR PLACE OF ARTICULATION IN SPEECH PERCEPTION

Abstract

Formant transitions reflect the overall change in the shape of the vocal tract during speech production. The objective of the study was to analyze F1 and F2 transition in Macedonian vowels in consonant context. This study included 12 native speakers of Macedonian (mean age of 42.7 ± 11.9 years).

Vowel /o/ in dento-alveolar plosive context /do/ had falling F2 transition in the duration of 38 ms and transition extent 772 Hz, and in velar plosive context /go/ had falling F2 transition in the duration of 25 ms and transition extent 85 Hz. Vowel /a/ in alveolar fricative context /fa/ had falling F2 transition in the duration of 25 ms and transition extent 324 Hz, and in labial fricative context /fa/ had rising F2 transition in the duration of 38 ms and transition extent 214 Hz. Vowel /u/ in palatal fricative context /ju/ had falling F2 transition in the duration of 48 ms and transition extent 1084 Hz. F2 locus for dento-alveolar consonant was 1900 Hz, for velar 2100 Hz, for alveolar 1700 Hz, for labial 1200 Hz, and for palatal consonant was 2600 Hz.

F2 locus is the highest for palatal consonants and the lowest for labial consonants. F2 transition is an acoustic cue for the place of articulation.

Keywords: COARTICULATION, FORMANT TRANSITION, ARTICULATION, SPEECH PERCEPTION

Introduction

Coarticulation is generally defined as the influence of one phonetic segment on another (Marchal, 2009). It is an overlap between the articulations of neighboring phonemes. The articulators are constantly moving as we talk, so the shape of the vocal tract for a particular phoneme is influenced by the shapes for the phonemes that both precede it and follow it (Goldstein, 2010). Two subdivisions of coarticulatory effects are generally made: left-to-right or carry-over effects, in which properties of a segment carry over to influence those of following segments, and right-to-left or anticipatory effects, in which properties of a segment influence those of earlier segments (Tatham and Morton, 2006).

The measurement of formants is one of the elements of the acoustic analysis of vowels. The first two formants are the most important for identifying a vowel (Raina, Chakraborty, and Velankar, 2014). A formant is a concentration

of acoustic energy around a particular frequency in the speech wave and corresponds to a resonance in the vocal tract. Formants can be seen very clearly in a wideband spectrogram, where they are displayed as dark bands. They are labeled as F1, F2, F3, F4, etc. starting with the lowest frequency (Gunasekar et al., 2017). F1 is related primarily to tongue height, with high vowels having a low F1 frequency and low vowels having a high F1 frequency. F2 relates mostly to tongue advancement, with back vowels having a low F2 frequency and front vowels having a high F2 frequency (Ludlow, Kent and Gray, 2019).

Vocal tract movements produce transitions of the formant frequencies. Formant transitions reflect the overall change in the shape of the vocal tract during speech production (Story and Bunton, 2010). The middle of the vowel, where the frequency is relatively stable, is called a *steady state* (Maltby, 2002). A formant transition is defined as a change of 20 Hz or greater over a 20 ms interval. Alternatively, a vowel steady-state condition is assumed to exist if the rate of change in the formant is less than 20 Hz over a 20 ms interval (Blomgren and Robb, 1998). A number of studies have demonstrated the importance of the formant transitions of the following vowel to the identification of a consonant's place of articulation (Baken and Orlikoff, 2000; Kishon-Rabin, Dayan and Michaeli, 2003; Kerdpol, 2012).

An important concept in describing formant transitions is the formant locus. The locus is an apparent point of origin of the formant for each place of consonant articulation (Ladefoged and Johnson, 2011). F2 locus is of particular interest because it helps to define the F2 transitions that play an important role in phonetic judgments and speech intelligibility. The locus is a presumed frequency for a particular place of articulation. Formant transition duration is the interval during which a substantial portion of a formant transition occurs. The transition can take different shapes, such as falling, nearly flat, or rising (Ludlow, Kent and Gray, 2019). In general, F2 transition with low-frequency locus cues the perception of a labial sound, one with higher locus cues the perception of an alveolar sound, and a variable, vowel-dependent locus cues a palatal or velar sound (Raphael, Borden and Harris, 2011).

The objective of the study was to analyze F1 and F2 transition in Macedonian vowels following the consonants with different places of articulation, as well as, to compare the formant frequencies in consonant context with the formant frequencies in vowels pronounced in isolation.

Methods

This study included 12 native speakers of Macedonian, six males, and six females, aged 26 to 64 years (mean age of 42.7 ± 11.9 years). They pronounced Macedonian vowels, and the following monosyllable and disyllable words: *dozd* (rain), *fal* (scarf), *farma* (farm), *gotvi* (cook), and *jug* (south). The words were re-recorded using the microphone Sennheiser e840 (Sennheiser electronic, Germany). Acoustic analysis of WAV files was performed in the computer program

Praat, Version 6.0.43 (Boersma and Weenink, 2018). We set the analysis parameters *Maximum formant* (Hz) and *View range* (Hz) in spectrogram: 5000 Hz for males and 5500 Hz for females. We measured fundamental frequency (Fo) and F1 to F5, and we analyzed F1 and F2 transition in consonant context. Segmentation of vowels in words was performed in a manual approach.

Results

Formant transitions in consonant-vowel sequences from real words were analyzed. Consonants with different places of articulation were included in the analysis. We used International Phonetic Alphabet (IPA) symbols for Macedonian sounds. The waveform and the spectrogram of the word [doʒd] (English: rain) pronounced by the female speaker are displayed in Figure 1. Vowel /o/ in consonant context /do/ was analyzed. Macedonian consonant /d/ is dento-alveolar plosive.

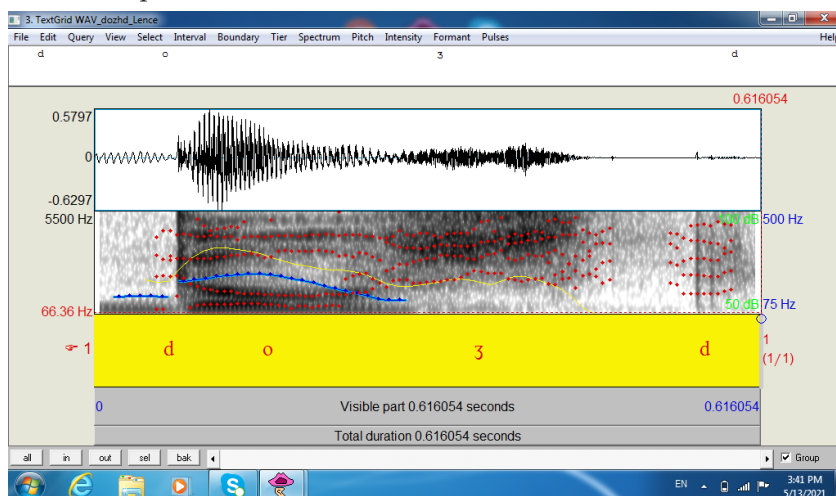


Figure 1. Waveform and spectrogram of the word [doʒd/English: rain]

There was a rising F1 transition and a falling F2 transition in the vowel /o/. The vowel was segmented in a duration of 175 ms. F1 transition duration was 25 ms, and the transition extent was 234 Hz, from 218 to 452 Hz. F2 transition duration was 38 ms. In this period, the formant frequency decreased by 772 Hz, from 1921 to 1149 Hz. F2 locus for dento-alveolar consonant is probably around 1900 Hz. We displayed formants of the vowel /o/ in the word [doʒd] with a noticeable greater F2 transition extent (Figure 2).

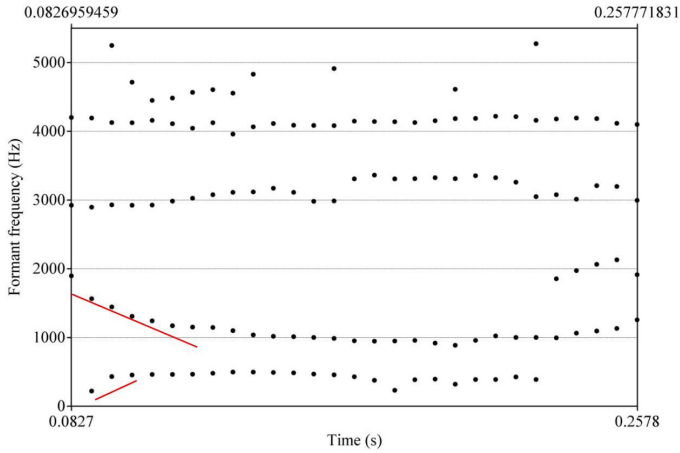


Figure 2. Formants of the vowel /o/ in the word [doʒd/English: rain]

In Table 1 we displayed F_0 and formant frequencies of the vowel /o/ in plosive context /do/ and isolated vowel /o/ in female speakers. F_2 of the vowel /o/ in plosive context /do/ was 1114 Hz, and F_2 of isolated vowel /o/ was 1097 Hz.

Table 1.

F₀ and formants of the vowel /o/ (in Hz)

Fo / Formants	Vowel /o/ in context /do/ (female speaker)	Isolated vowel /o/ (female speaker)	Isolated vowel /o/ (mean±SD)
F ₀	222	210	209±18
F ₁	454	718	558±76.5
F ₂	1114	1097	1205±166.1
F ₃	3230	3158	3133±95.2
F ₄	4125	3699	3937±185.8
F ₅	4670	4448	4848±278.4

We analyzed the word [fal] (English: scarf) pronounced by a male speaker. The waveform and the spectrogram of the word [fal] are displayed in Figure 3.

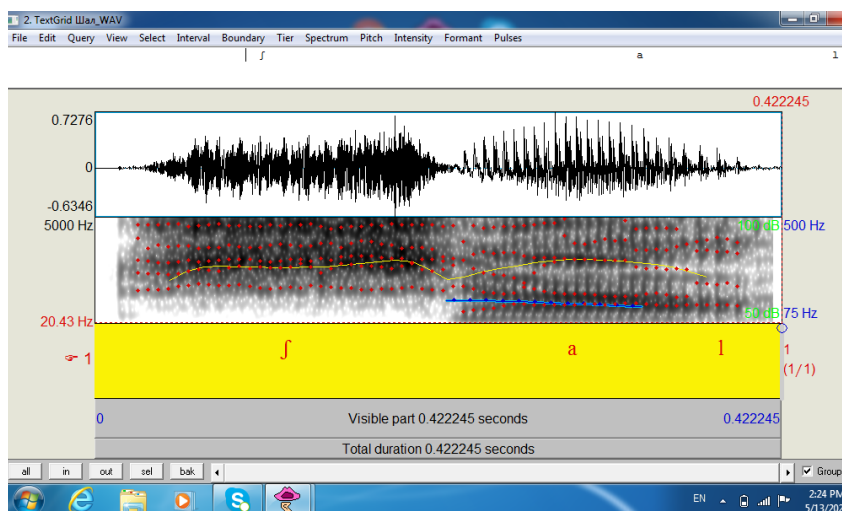


Figure 3. Waveform and spectrogram of the word [fal/English: scarf]

Vowel /a/ in alveolar fricative context /fa/ was analyzed. There was a rising F1 transition and falling F2 transition in the vowel /a/. The duration of the word [fal] was 422 ms, and the segmented vowel /a/ had a duration of 127 ms. F1 transition duration was 37 ms. In this period F1 increased by 299 Hz, from 529 to 858 Hz. F2 transition duration was 25 ms and the formant frequency decreased for 324 Hz, from 1678 to 1354 Hz. F2 locus for alveolar consonant is probably around 1700 Hz. In Figure 4 we displayed formants F1 to F5 of the vowel /a/ in the word [fal]. There is a different transition shape between the first two formants and a short “steady state” of F2.

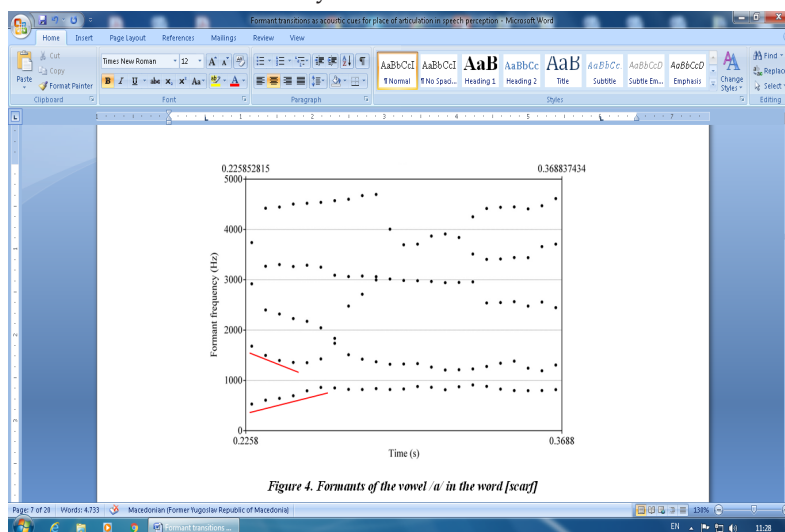


Figure 4. Formants of the vowel /a/ in the word [scarf]

Figure 4. Formants of the vowel /a/ in the word [fal/English: scarf]

We displayed F_0 and formant frequencies of the vowel /a/ (Table 2). F_2 of the vowel /a/ in fricative context /fa/ was 1425 Hz, and F_2 of isolated vowel /a/ was 1323 Hz.

Table 2.
F₀ and formants of the vowel /a/ (in Hz)

Fo / Formants	Vowel /a/ in context /fa/ (male speaker)	Isolated vowel /a/ (male speaker)	Isolated vowel /a/ (mean±SD)
Fo	166	145	115±14.7
F1	831	853	733±78
F2	1425	1323	1161±81.1
F3	2817	2574	2667±118.1
F4	3508	3278	3413±173.1
F5	4657	4304	4410±328

Formant transitions in the vowel /a/ were analyzed in labial fricative context /fa/ in the word [farma] (English: farm) pronounced by a female speaker. The waveform and the spectrogram of the word [farma] are displayed in Figure 5. There was a rising transition of both F_1 and F_2 .

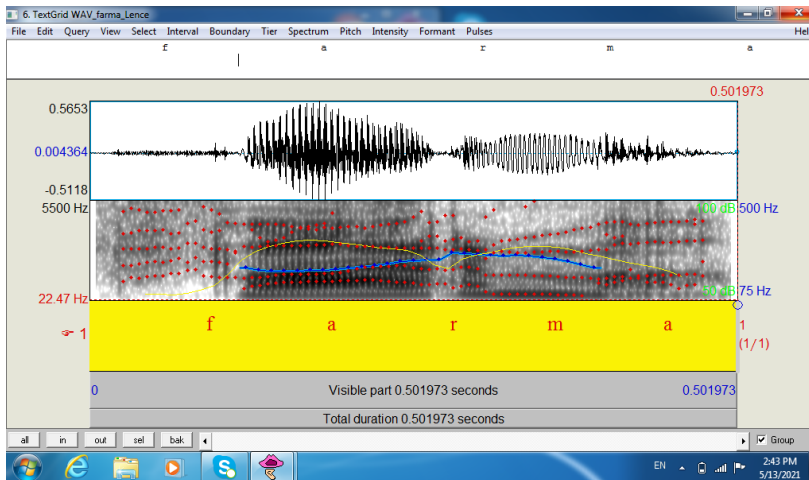


Figure 5. *Waveform and spectrogram of the word [farma/English: farm]*

The duration of the word [farma] was 502 ms, and the duration of the first vowel /a/ was 149 ms. The first formant transition duration was 29 ms. In this period F_1 increased for 296 Hz, from 644 to 940 Hz. F_2 transition duration was 38 ms and the formant frequency increased for 214 Hz, from 1169 to 1383 Hz. F_2 locus for labial consonant is probably around 1200 Hz. In Figure 6 we displayed formants of the vowel /a/ in the word [farma].

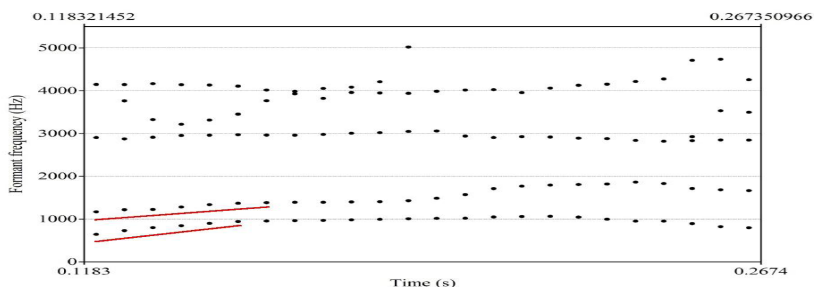


Figure 6. Formants of the first vowel /a/ in the word [farma/English: farm]

In Table 3 we displayed F_0 and formant frequencies of the vowel /a/ in fricative context /fa/ and isolated vowel /a/. F_2 of the vowel /a/ in fricative context /fa/ is 1557 Hz, and F_2 of isolated vowel /a/ is 1448 Hz.

Table 3.

F_0 and formants of the vowel /a/ (in Hz)

Fo / Formants	Vowel /a/ in context / fa/ (female speaker)	Isolated vowel /a/ (female speaker)	Isolated vowel /a/ (mean \pm SD)
F₀	199	198	204 \pm 31
F₁	994	1035	869 \pm 129.6
F₂	1557	1448	1467 \pm 162.9
F₃	2952	2905	2896 \pm 187.1
F₄	3915	3639	3866 \pm 247.4
F₅	4198	4307	4842 \pm 316.3

We analyzed formant transitions in the word [gotvi] (English: cook). The waveform and the spectrogram of the word [gotvi] pronounced by a male speaker are displayed in Figure 7.

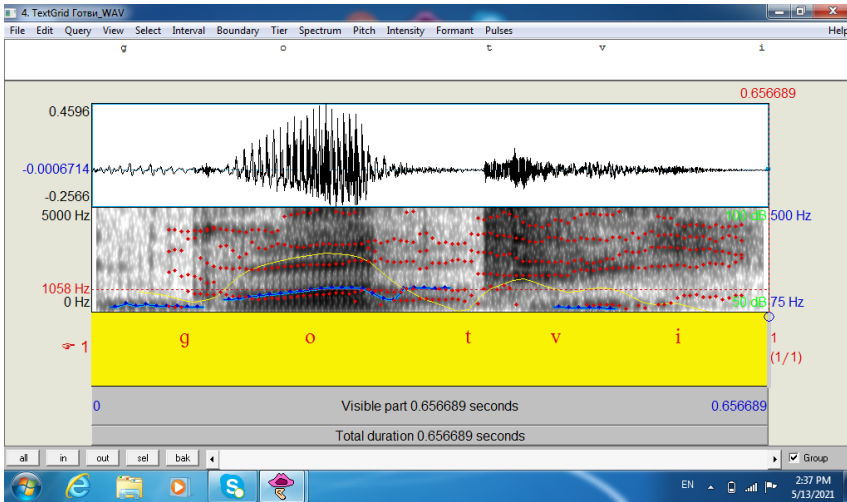


Figure 7. Waveform and spectrogram of the word [gotvi/English: cook]

Vowel /o/ in velar plosive context /go/ was analyzed. The duration of the word [gotvi] was 657 ms, and the duration of the vowel /o/ was 135 ms. F1 transition duration was 26 ms and the formant frequency decreased for 131 Hz, from 692 to 561 Hz. F2 transition duration was 25 ms and the formant frequency decreased for 85 Hz, from 2056 to 1971 Hz. F2 locus for velar consonant is probably around 2100 Hz. In Figure 8 we displayed formants of the vowel /o/.

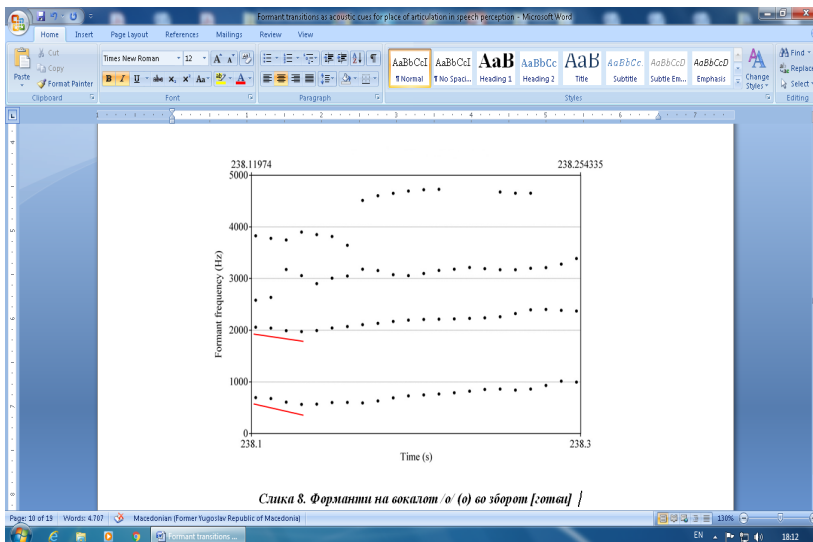


Figure 8. Formants of the vowel /o/ in the word [gotvi/English: cook]

In Table 4 we displayed F_0 and formant frequencies of the vowel /o/ in plosive context /go/, isolated vowel /o/ in a male speaker who pronounced the word [gotvi], and mean F_0 and formant frequencies in all male speakers. F_2 of the vowel /o/ in plosive context /go/ was 2040 Hz. F_2 of isolated vowel /o/ was 1506 Hz in the duration of the vowel of 284 ms.

Table 4.
F₀ and formants of the vowel /o/ (in Hz)

Fo / Formants	Vowel /o/ in context /go/ (male speaker)	Isolated vowel /o/ (male speaker)	Isolated vowel /o/ (mean±SD)
F ₀	127	146	123±13.8
F ₁	688	662	528±88.9
F ₂	2040	1506	1087±274.4
F ₃	3002	2822	2697±212.1
F ₄	4098	3779	3598±122.2
F ₅	4651	4049	4457±207.6

The vowel /u/ was analyzed in palatal fricative context /ju/. The word [jug] (English: south) was pronounced by a female speaker. The waveform and the spectrogram of the word [jug] are displayed in Figure 9. There was rising F_1 and falling F_2 transition in the vowel /u/. The duration of the word [jug] was 416 ms, and the duration of the vowel /u/ was 125 ms.

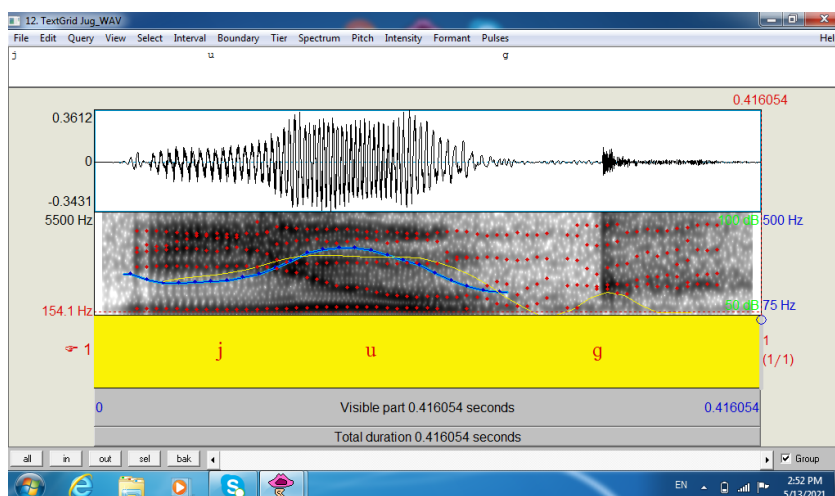


Figure 9. *Waveform and spectrogram of the word [jug/English: south]*

F_1 transition duration was 27 ms and the formant frequency increased for 63 Hz, from 283 to 346 Hz. F_2 transition duration was 48 ms and the formant

frequency decreased for 1084 Hz, from 2537 to 1453 Hz. F2 locus for palatal consonant is probably around 2600 Hz.

In Figure 10 we displayed formants of the vowel /u/. It is noticeable great F2 transition extent in comparison to F1 transition extent.

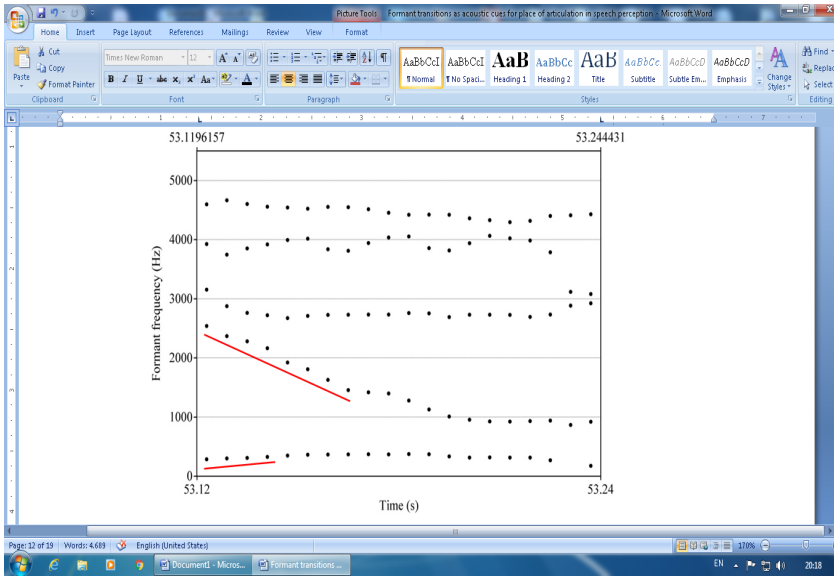


Figure 10. Formants of the vowel /u/ in the word [juɡ/English: south]

In Table 5 we displayed Fo and formant frequencies of the vowel /u/ in fricative context /ju/ and isolated vowel /u/. The second formant of the vowel /u/ in fricative context /ju/ was 1542 Hz in the duration of the vowel of 125 ms.

Table 5.
Fo and formants of the vowel /u/ (in Hz)

Fo / Formants	Vowel /u/ in context / ju/ (female speaker)	Isolated vowel /u/ (female speaker)	Isolated vowel /u/ (mean±SD)
Fo	258	238	255±22.8
F1	352	386	416±52
F2	1542	1320	1395±181.3
F3	2781	3093	3129±235.8
F4	3903	3992	4182±202.8
F5	4469	4286	4821±253.1

The second formant frequency of isolated vowel /u/ was 1320 Hz in the duration of the vowel of 249 ms. The mean second formant frequency in all female speakers was 1395 ± 181.3 Hz.

Discussion

We analyzed the first and the second formant transition in the consonant context in real words. According to the place of articulation, Macedonian consonants are classified into: labial, dento-alveolar, alveolar, palatal, and velar consonants (Bojkovska et al., 2008). Acoustic cues to the perception of speech segments might be divided into those important to the perception of the manner of articulation, place of articulation, and voicing distinctions. The acoustic cues to the place of articulation are: formant spacing, frequency of the noise component of a consonant, and F2 transition (Raphael, Borden and Harris, 2011).

The acoustic analysis of speech samples was done in the computer program Praat. Praat is software for analyzing, synthesizing, and manipulating speech. Praat is not a text-to-speech system, but one can generate many types of sounds with Praat (Boersma and Weenink, 2001). The waveforms and spectrograms of monosyllable and disyllable words were analyzed. The spectrogram is a spectro-temporal representation of the sound. The horizontal direction of the spectrogram represents time, and the vertical direction represents frequency (Proverbio et al., 2016). The level of darkness in the spectrogram represents amplitude, such that a dark portion states that there is significant energy at that frequency at that time (Taylor, 2009).

We performed manual segmentation of vowels in order to analyze the formant transition. Manual speech segmentation is based on listening and visual judgment in order to detect the required boundaries (Al-Manie, Alkanhal and Al-Ghamdi, 2010). The vowel boundaries refer to vowel onset and offset points. The vowel offset point is marked as the instant at which the energy of the second formant drops abruptly (Yadav and Rao, 2013).

In the analysis of vowel /o/ in dento-alveolar plosive context /do/, we found a rising F1 transition and a falling F2 transition. There was a high-frequency F2 locus, probably around 1900 Hz. The second formant of the vowel /o/ in consonant context was higher than F2 in isolated vowel /o/. The formant transitions are determined by the movements of the tongue body and mandible as these structures are displaced toward the vowel from configurations that they must assume to form the stop consonant closure (Stevens, 2000).

We analyzed formant transitions in vowel /a/ in alveolar fricative context /ʃa/. Fricative articulations are the result of two articulators being in close approximation with each other. This is a degree of stricture whereby the articulators are held close enough together for air to pass between them, but because the gap between them is small, the airflow becomes turbulent and creates friction noise (Ogden, 2009). There was a rising F1 transition and a falling F2 transition. F2 locus was around 1700 Hz. The vowel /a/ second formant in fric-

ative context /ja/ was higher than F2 of the vowel /a/ pronounced as an isolated sound. In English, F2 transition with high-frequency locus cues the perception of an alveolar sound (Raphael, Borden and Harris, 2011). According to Ladefoged and Johnson (2011) F2 locus for alveolar sounds is around 1700-1800 Hz.

In the analysis of vowel /a/ in labial plosive context /fa/ we found both F1 and F2 rising transitions. The second formant of the vowel /a/ in plosive context /fa/ is higher than F2 of the vowel /a/ pronounced as an isolated sound. F2 locus is probably around 1200 Hz. The locus for labial consonant is lower in comparison to F2 locus for dento-alveolar and alveolar consonant. In general, F2 transition with low-frequency locus cues the perception of a labial sound (Raphael, Borden and Harris, 2011). According to Ladefoged and Johnson (2011) locus of both second and third formants in bilabial context are comparatively low.

Analysis of vowel /o/ in velar plosive context /go/ showed both F1 and F2 falling transition. There was a high-frequency F2 locus for velar consonant around 2100 Hz. F2 of the vowel /o/ in fricative context /go/ is higher than F2 of the vowel /o/ pronounced as isolated sound. In English, there is a high F2 locus in velar plosive context (Ladefoged and Johnson, 2011).

We analyzed formant transitions in vowel /u/ in palatal fricative context /ju/. There was a rising F1 transition and a falling F2 transition with a great transition extent. We determined a high-frequency F2 locus for palatal consonant around 2600 Hz. In English, F2 transition with variable, vowel-dependent locus cues the perception of a palatal or velar sound (Raphael, Borden and Harris, 2011). The second formant of the vowel /u/ in fricative context /ju/ is higher than F2 of the vowel /u/ pronounced as isolated sound. The great F2 transition extent can be explained with the acoustic structure of the sound /j/. In terms of its articulatory and acoustic characteristics, this sound is similar to vowel /i/ (Ristovska et al., 2018).

The second formant transition duration in a plosive context in our study was 25 and 38 ms, and in a fricative context was 25, 38, and 48 ms. For English stop consonants, transition durations are in the range of 50 to 80 ms (Ludlow, Kent and Gray, 2019).

Conclusion

F2 locus is the highest for palatal consonants and the lowest for labial consonants. F2 transition is an acoustic cue for the place of consonant articulation. The second formant in vowels in fricative context is higher than F2 in vowels pronounced as isolated sounds. Formant transitions are context-dependent acoustic cues, and transition direction depends on neighboring consonants.

References

- Al-Manie, M.A., Alkanhal, M.I. and Al-Ghamdi, M.M. (2010). Arabic speech segmentation: Automatic versus manual method and zero crossing measurements. *Indian Journal of Science and Technology*, 3(12), pp. 1134-1138.
- Baken, R.J. and Orlikoff, R.F. (2000). *Clinical measurement of speech and voice*. 2nd ed. San Diego: Singular Thomson Learning, pp. 273-274.
- Blomgren, M. and Robb, M. (1998). How steady are vowel steady-states? *Clinical Linguistics & Phonetics*, 12(5), pp. 405-415.
- Boersma, P. and Weenink, D. Praat [Computer program]. Version 6.0.43. Retrieved 24.10.2018 from: <http://www.fon.hum.uva.nl/praat/>
- Boersma, P. and Weenink, D. (2001). Praat, a system for doing phonetics by computer. *Glott International*, 5(9/10), pp. 341-347.
- Bojkovska, S., Minova-Gjurkova, L., Pandev, D. and Cvetkovski, Zh. (2008). Opshta gramatika na makedonskiot jazik. Skopje: Prosvetno delo, pp. 102-103. [In Macedonian]
- Goldstein, E.B. (2010). *Sensation and perception*. 8th ed. Belmont: Wadsworth, Cengage Learning, p. 315.
- Gunasekar, C., Sabrigirinathan, C., Vinayagavel, K. and Ramkumar, K. (2017). The acoustic parameters for analyzing speech with complete dentures. *International Journal of Dental Research*, 5(2), pp. 115-120.
- Kerdpol, K. (2012). Formant transitions as effective cues to differentiate the places of articulation of Ban Pa La-u Sgaw Karen Nasals. *MANUSYA: Journal of Humanities Regular*, 15(2), pp. 21-38.
- Kishon-Rabin, I., Dayan, M. and Michaeli, O. (2003). Effects of second-formant transition on the perception of Hebrew voiced stop consonants. *Journal of Basic and Clinical Physiology and Pharmacology*, 14(2), pp. 151-164.
- Ladefoged, P. and Johnson, K. (2011). *A course in phonetics*. 6th ed. Boston: Cengage Learning, pp. 199, 204.
- Ludlow, C.L., Kent, R.D. and Gray, L.C. (2019). *Measuring voice, speech, and swallowing in the clinic and laboratory*. San Diego: Plural Publishing, pp. 181, 188, 194.
- Maltby, M.T. (2002). *Principles of hearing aid audiology*. 2nd ed. London: Whurr Publishers, p. 75.
- Marchal, A. (2009). *From speech physiology to linguistic phonetics*. London: ISTE, p. 173.
- Ogden, R. (2009). *An introduction to English phonetics*. Edinburgh: Edinburgh University Press, p. 17.
- Proverbio, A.M., Massetti, G., Rizzi, E. and Zani, A. (2016). Skilled musicians are not subject to the McGurk effect. *Scientific Reports*, 6:30423.

- Raina, D., Chakraborty, S. and Velankar, M.R. (2014). Automatic classification of instrumental music & human voice using formant analysis. *International Journal of Advanced Research in Computer Science & Technology*, 2(2), pp. 242-245.
- Raphael, L.J., Borden, G.J. and Harris, K.S. (2011). *Speech science primer: physiology, acoustics, and perception of speech*. 6th ed. Baltimore: Lippincott Williams & Wilkins, pp. 214-215.
- Ristovska, L., Jachova, Z., Spasov, Lj. and Balova, T. (2018). Acoustic characteristics of Macedonian vowels. *Journal of Special Education and Rehabilitation*, 19(3-4), pp. 40-50.
- Stevens, K.N. (2000). *Acoustic phonetics*. Cambridge: MIT Press, p. 377.
- Story, B.H. and Bunton, K. (2010). Relation of vocal tract shape, formant transitions, and stop consonant identification. *Journal of Speech, Language, and Hearing Research*, 53, pp. 1514-1528.
- Tatham, M. and Morton, K. (2006). *Speech production and perception*. Basingstoke: Palgrave Macmillan, p. 23.
- Taylor, P. (2009). *Text-to-speech synthesis*. New York: Cambridge University Press, pp. 159.
- Yadav, J. and Rao, K.S. (2013). Detection of vowel offset point from speech signal. *IEEE Signal Processing Letters*, 20(4), pp. 299-302.