

Context and speaker dependency in the relation of vowel formants and subglottal resonances – Evidence from Hungarian

Tekla Etelka Grácz¹, Steven M. Lulich², Tamás Gábor Csapó³, András Beke^{1,4}

¹ Research Institute for Linguistics, Hungarian Academy of Sciences, Budapest, Hungary

² Department of Psychology, Washington University, Saint Louis, Missouri

³ Department of Telecommunications and Media Informatics, Budapest University of Technology and Economics, Budapest, Hungary

⁴ Eötvös Loránd University, Budapest, Hungary

graczi@nytud.hu, slulich@wustl.edu, csapot@tmit.bme.hu, beke.andras@gmail.com

Abstract

Subglottal resonances are claimed to divide front/back vowels and low/high vowels in several languages, including Hungarian. However, some ‘recalcitrant’ vowels appear to resist this mould. We therefore performed a careful analysis of the role of coarticulation and speaker-dependent effects that might play in the recalcitrance of these vowels in Hungarian. The present analysis focused on various stop contexts in order to see the place of articulation triggered effects. It is shown that the subglottal resonances indeed divide the vowel space as claimed, and that the recalcitrance of certain vowels is due to coarticulation with specific consonants. The magnitude of the coarticulation effect is speaker dependent.

Index Terms: subglottal resonances, coarticulation, vowels, Hungarian

1. Introduction

Several studies in recent years [1, 2, 3, 4, 5] have explored the possible role of subglottal resonances (SGRs) in defining vowel and consonant categories, as originally hypothesized by Stevens [6] (we will refer to this as “the subglottal hypothesis”). In general, it has been found that the relation between the first subglottal resonance (Sg1) and the first formant (F1) defines the contrast between [+low] and [-low] vowels, and that the relation between the second subglottal resonance (Sg2) and the second formant (F2) defines the contrast between [+back] and [-back] vowels. Specifically, for [+low] vowels F1 is at a higher frequency than Sg1, and for [-low] vowels F1 is lower. Similarly, for [+back] vowels F2 is at a lower frequency than Sg2, and for [-back] vowels F2 is higher.

These relations for [low] and [back] vowels were shown to hold in general for Hungarian, although the vowels /*ɛ*/, /*ɔ*/, /*a*/, and /*ø*/ exhibited some exceptional behavior [4, 5]. (For a description of the phonological and phonetical categorization of these vowels, see [4]). Some of this behavior was explained on the basis of speaker-specific patterns. For instance, some speakers produced the majority of their [+low] /*ɛ*/ vowels with F1 higher than Sg1, while others produced the majority of their /*ɛ*/ vowels with F1 lower than Sg1. This was also true of the [+low] vowel /*ɔ*. At the same time, the [+back] vowel /*ɔ*/ was frequently produced with F2 less than Sg2, but for some speakers there was also a significant number of productions with F2 higher than Sg2. The vowels /*ø*/ and /*a*/, which are phonetically [-back] (although /*a*/ patterns with [+back] vowels phonologically [7, 8, 9]) were similarly frequently

produced with F2 higher than Sg2, but for some speakers there was a significant number of productions with F2 lower than Sg2.

It is possible that “the subglottal hypothesis” simply does not hold categorically for all vowels and all speakers, at least in Hungarian. On the other hand, it is possible that coarticulation between the vowels and their adjacent consonants masked the categorical nature of the case. Since CV- and VC-transitions with varying consonants may be more or less coarticulated and result in variable amounts of vowel target undershoot (depending on the consonants themselves, the speaker, and the speaking style, among other things, [10]), we hypothesized that a detailed examination of consonant context effects would yield evidence that coarticulation does indeed play a significant role in masking the categorical nature of “the subglottal hypothesis” in Hungarian. This paper investigates that possibility by examining formant transitions and their extrema for Hungarian vowels in a variety of voiced stop consonant contexts.

2. Methods

Four native speakers of Standard Hungarian (2 males = m1, m2 and 2 females = f1, f2; aged 27–29 years) participated in the experiment. None of them reported any speech disorders or hearing problems. (Speaker m2 was also called speaker m4 in an earlier study [5].) They were recorded reading nonsense words in a carrier sentence. The nonsense words were CVCV disyllables, where the two vowels in each word were always identical (/*ɛ*, *ɔ*, *a*, *ø*/), and the two consonants varied independently among voiced stops (/b, d, j, g/). There were therefore a total of 64 nonsense words. The carrier sentence was “Most a CVCV szó olvasom.” [‘I am reading the word CVCV now.’]. The sentences were randomized, and repeated six times by each speaker. The first vowel in the CVCV was the target vowel for subsequent analysis. Sustained productions of all four vowels were also recorded from the four subjects.

Microphone recordings were made with an Audio-Technica AT 4040 microphone. Subglottal data were recorded simultaneously using a K&K HotSpot accelerometer pressed against the skin of the neck (by hand) below the thyroid cartilage. Both the microphone and accelerometer signals were recorded using an M-Audio Fast Track Pro external sound card in a sound-treated room at 22.1 kHz and 16 bits.

Table 1. Mean values and standard deviations (in parentheses) of the speakers' SGRs (Hz).

Speaker	Sg1	Sg2
m1	607 (36)	1290 (67)
m2	555 (30)	1348 (56)
f1	624 (73)	1536 (80)
f2	500 (50)	1431 (50)

The microphone recordings were automatically segmented [11], and the target vowel boundaries were manually corrected by the authors. Formant measurements (F1 and F2) were made automatically using Praat 5.1 [12], and manually corrected by the authors. Formants were measured 21 times in each vowel (at 0, 5, 10, ... 95, 100% of the vowel duration), thus reflecting the complete formant trajectories.

The first and second subglottal resonances (Sg1, Sg2) were manually measured from the accelerometer recordings using Wavesurfer [13]. Twenty-five vowel mid-points were selected for measurement from each speaker. The means and standard deviations were calculated (Table 1), since the SGRs are roughly invariant for a given speaker [2].

The relation between the formants and SGRs was analyzed statistically by measuring 1) the difference between Sg1 and the maximum (extremum) value of F1 in a given utterance, and 2) the difference between Sg2 and F2, measured at the same time point as the maximum F1 (see Fig. 2, lower left panel). We also measured the formants at the vowel midpoint. Crosstable analyses with chi-square and Cramer's V-statistics were run in order to analyze the categorical relations between the formants and the SGRs. Independent samples t-tests (at 95% confidence level) were carried out in order to analyze the relation of the distributions of the formant extrema (separated by speaker and consonant context) and the speakers' SGRs. All statistical analyses were carried out with SPSS 15.0.

3. Results and Discussion

3.1. The Vowel Space

Measurements of F1 and F2 at the vowel midpoint across speakers have distributions relative to the SGRs which are similar to those seen in previous studies [4,5] (data not shown). F1 for [a:] and [ø] appears on the expected side of Sg1 in over 93% of the tokens, while F1 for [ε] and [ɔ] does so only in 32.1% and 54.2% of the tokens, respectively. Likewise, F2 for [ε], [a:] and [ø] appears on the expected side of Sg2 in over 97.6% of the tokens, while F2 for [ɔ] does so only in 59.6% of the tokens.

Measurements of F1 and F2 at the time when F1 attained its maximum value (henceforth referred to as $F1_m$ and $F2_m$) showed similar overall distributions (see Figure 1), and the point in time (as a percentage of vowel duration) when these measurements were made was close to the vowel midpoint ([a:]: $53\% \pm 1.2\%$; [ø]: $46\% \pm 1.3\%$; [ε] and [ɔ]: not significantly different from 50%). These findings pooled across speakers are similar to what was previously reported [4,5].

Figure 1 shows the $F1_m$ vs. $F2_m$ vowel spaces for each speaker individually. In all four speakers, the vowel [ɔ] has an $F2_m$ value less than Sg2 in many cases, and this is nearly categorical for speakers m2 and f1.

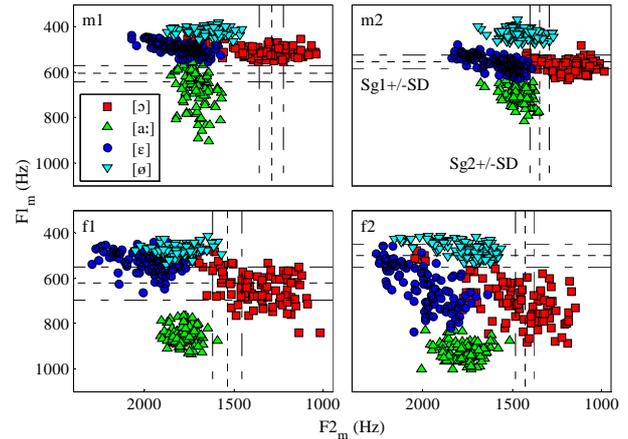


Figure 1. Vowel spaces for $F1_m$ vs. $F2_m$. The vertical dashed lines indicate the speakers' mean Sg2 +/- one standard deviation. The horizontal dashed lines indicate the mean Sg1 +/- one standard deviation.

Similarly, in all four speakers the vowel [a:] has an $F1_m$ value categorically greater than Sg1, the vowel [ø] has an $F1_m$ value less than Sg1, and the vowels [ε], [a:], and [ø] have an $F2_m$ value categorically greater than Sg2. For the vowels [ε] and [ɔ], the $F1_m$ value relative to Sg1 is speaker dependent. For speaker m1, $F1_m$ is less than Sg1 for both vowels. For speaker m2, $F1_m$ is usually greater than Sg1 for [ɔ], and for the vowel [ε] $F1_m$ is sometimes greater than Sg1 and sometimes less than Sg1. For speaker f1, $F1_m$ is usually less than Sg1 for [ε] and for [ɔ] it is sometimes less than Sg1 and sometimes greater than Sg1. For speaker f2, both vowels have $F1_m$ greater than Sg1.

3.2. Inter-Speaker Effects

In order to better understand the significance and magnitude of interspeaker effects, we performed Chi-square and Cramer's V analyses on the vowels [ε] and [ɔ]. The data were first parceled into two categories ($F > SGR$ and $F < SGR$). For $F1_m$ vs. Sg1 there were 5 tokens for which the measured $F1_m$ was equal to the mean Sg1, and these tokens were therefore discarded from this analysis. Similarly, for $F2_m$ vs. Sg2, 1 token was discarded. The results of the statistical analyses are not significantly affected by discarding these tokens, given their small number relative to the total.

For $F1_m$ vs. Sg1 and $F2_m$ vs. Sg2 there was a clear speaker dependency ($p < 0.001$ for both vowels using the Chi-square test). The Cramer's V statistic (an indication of the magnitude of the intercorrelation effect) ranged between 0.265 and 0.779 for $F1_m$ and $F2_m$ for the two vowels.

3.3. Context Effects

In order to better understand the possible effects of coarticulation with the preceding and following consonants, we performed an analysis on the formant trajectories. Figure 2 shows the averaged formant trajectories within the vowel space for each of the 16 consonantal contexts for speaker f2. The vowel [a:] always has $F1_m$ and $F2_m$ in the lower left quadrant, i.e. $F1_m$ greater than Sg1 and $F2_m$ greater than Sg2. Similarly, the vowel [ø] always has $F1_m$ and $F2_m$ in the upper left quadrant, i.e. $F1_m$ less than Sg1 and $F2_m$ greater than Sg2. For the vowels [a:] and [ø], speaker f2 is representative of all the speakers.

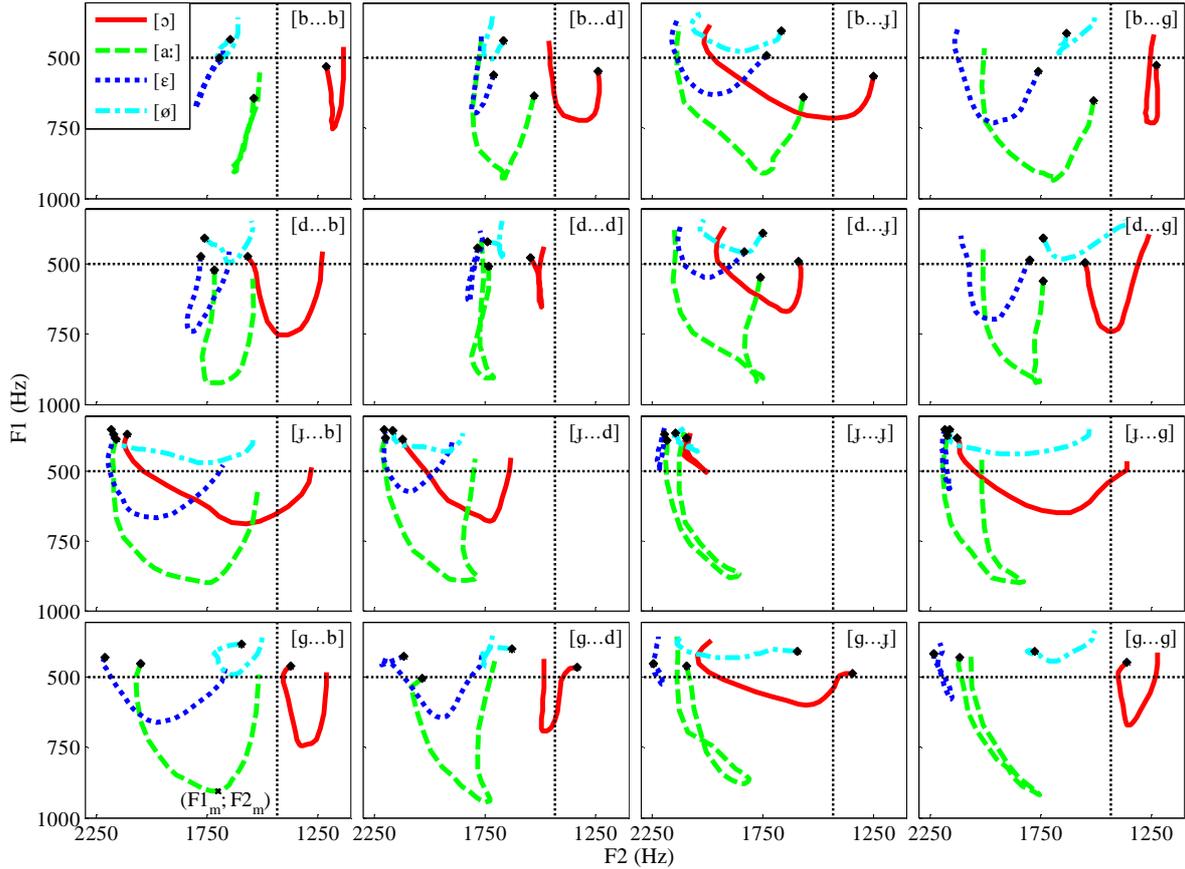


Figure 2: Average formant trajectories of the target vowel for each consonant context in the $F1$ vs. $F2$ vowel space for speaker $f2$. Horizontal lines indicate mean $Sg1$, vertical lines indicate mean $Sg2$. Black dots show the start of the formant trajectories. The lower left panel also shows an example of the $F1_m$ and $F2_m$ measurement point.

For the vowels $[\text{ɔ}]$ and $[\text{ɛ}]$, $F1_m$ is almost always greater than $Sg1$, and it is only in a velar or palatal context where $F1_m$ is less than (or near) $Sg1$. This context effect is strongest when both the preceding and following consonants are palatal. However, this context effect varies from one speaker to the next. For the vowel $[\text{ɛ}]$, across all speakers, the context effect is significant according to the chi-square test ($p < 0.030$), and the Cramer's V values for the effect of the preceding and following consonants are 0.152 and 0.184, respectively. The context effect for $[\text{ɔ}]$ across speakers is also significant ($p < 0.005$), and Cramer's V values are 0.143 and 0.157, respectively.

For the vowel $[\text{ɔ}]$ in Figure 2, $F2_m$ is context dependent. Velars and labials consistently lead to values of $F2_m$ less than $Sg2$, whereas alveolars and palatals lead to values of $F2_m$ greater than $Sg2$. Mixed contexts lead to values of $F2_m$ near $Sg2$. The context effect across speakers is significant ($p < 0.001$), and the Cramer's V statistics for the preceding and following consonants are 0.514 and 0.327 respectively. Context effects for $[\text{ɔ}]$ and $[\text{ɛ}]$ are thus of the same order or smaller than the inter-speaker effects, but (as shown below) they are also more consistent across speakers.

To investigate further the context effects for the vowels $[\text{ɔ}]$ and $[\text{ɛ}]$, we performed independent samples t-tests for each of the 16 consonant contexts, comparing $F1_m$ with $Sg1$, and (for $[\text{ɔ}]$ only) $F2_m$ with $Sg2$. The results are shown in Figure 3. Black squares indicate that the formant values were significantly different from the corresponding subglottal

resonance values, but in the direction contrary to expectation (we expect $F1_m > Sg1$ for both vowels, and $F2_m < Sg2$ for $[\text{ɔ}]$). White squares indicate that the formant values were significantly different from the corresponding subglottal resonance values, and in the expected direction. Gray squares indicate that the formant values were not significantly different from the corresponding subglottal resonance values. The preceding consonant context is given along the vertical axis, and the following consonant context is given along the horizontal axis.

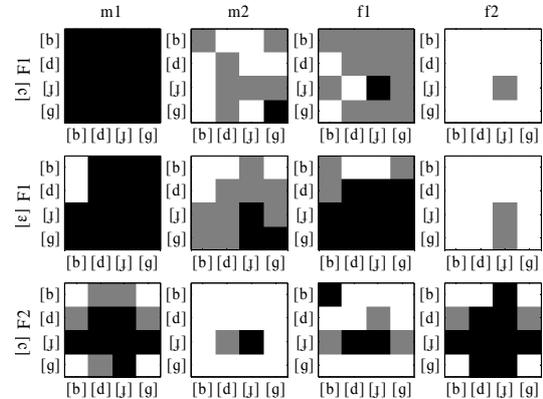


Figure 3. Results of the t-test analyses. Each column presents data for a given speaker. The first two rows show the results for $F1_m$ vs. $Sg1$. The bottom row shows the results for $F2_m$ vs. $Sg2$.

For $F1_m$, there is a trend across all speakers (although there is a strong speaker dependence) that the formant values are more likely to be less than $Sg1$ (i.e. more gray/) as the consonant contexts change from labial (upper left) to palatal and velar (lower right). This can be explained on the basis that 1) articulations utilizing the tongue body require increased jaw height (which correlates inversely with $F1$) and 2) palatal articulation requires a relatively long constriction which further constrains the jaw height (for Hungarian palato- and linguographic results see [14]). The jaw height in a vowel with velar and palatal contexts is therefore less likely to reach a value commensurate with a high $F1$.

For $F2_m$, the patterns are different, but it is still consistent across all speakers. Speakers $m1$ and $f2$ are the most representative cases: contexts involving labials and velars are more likely to result in $F2_m$ values less than $Sg2$, in accordance with expectation. Conversely, contexts involving alveolars and palatals are more likely to result in $F2_m$ values greater than $Sg2$, contrary to expectation. Mixed contexts (in which one consonant is labial or velar and the other is alveolar or palatal) result in the intermediate category in which $F2_m$ values are likely to be near $Sg2$. These results yield the black/gray cross-shaped pattern with white corners, and are consistent with previous studies in English and Hungarian showing that alveolar and palatal stops have a high $F2$ locus able to exert coarticulatory pressure on $F2_m$ [2,15,16].

Although $F1_m$ for speaker $m1$ is usually contrary to expectation in both vowels in almost all contexts, analysis of his sustained vowels (without any consonantal context) showed that $F1$ was greater than $Sg1$ for both vowels, especially [ε], which had a mean $F1$ value roughly 186 Hz higher than $Sg1$ ($F1$ was roughly 36 Hz higher than $Sg1$ for [ɔ]).

On the basis of these data, especially the context dependence of $F1_m$ and $F2_m$, we conclude that the “subglottal hypothesis” holds true for Hungarian in the abstract, but that coarticulation, the strength of which is speaker dependent, can frequently alter the relations between formants and SGRs for some vowels, particularly the vowels [ɔ] and [ε].

4. Conclusion

In this study we investigated the hypothesis that coarticulation with consonants can prevent some Hungarian vowels from consistently realizing “the subglottal hypothesis”, depending on the speaker. Although vowel target undershoot may be due to a variety of factors (see [10] and [17] for a discussion of such factors), it appears that articulatory constraints placed on the vowel by adjacent consonants can account for the majority of tokens with formants on the “wrong side” of the corresponding SGRs. These results therefore strengthen the conclusions of previous studies of Hungarian vowels [4, 5], and extend that work by demonstrating the effects of consonant context on vowel production with respect to subglottal resonances, and by showing that the context effects are themselves speaker dependent. These results are based on only 4 speakers. In the future, more speakers are needed to confirm these findings.

It should be noted that not all of the recalcitrant vowel tokens could be traced to context effects: the realizations of the vowel [ø] with $F1 > Sg1$ are still exceptional. However, such tokens occur infrequently (2.4% of tokens).

The context effects raise further questions, such as whether manner of articulation has a similar effect, or what

the perceptual consequences of context effects might be. We intend to explore such questions in our future work.

5. Acknowledgements

This work was supported in part by NSF Grant No. 0905250, by the BelAmi: ALAP2-00004/2005, the TÁMOP-4.2.2-08/1/KMR-2008-0007 and by the TÁMOP-4.2.1/B-09/1/KMR-2010-0002 projects.

6. References

- [1] Jung, Y., “Acoustic articulatory evidence for quantal vowel categories: The features [low] and [back]” PhD dissertation, MIT, 2009.
- [2] Lulich, S.M., “Subglottal resonances and distinctive features”, *J. of Phonetics* 38(1):20–32, 2010.
- [3] Madsack, A., Lulich, S. M., Wokurek, W., and Dogil, G. “Subglottal resonances and vowel formant variability: A case study of high German monophthongs and Swabian diphthongs”, in *Proc. of LabPhon 11:91–92*, 2008.
- [4] Csapó, T. G., Bárkányi, Zs., Grácz, T. E., Böhm, T. and Lulich, S. M. “Relation of formants and subglottal resonances in Hungarian vowels”, in *Proc. of Interspeech*, 484–487, 2009.
- [5] Csapó T.G., Grácz, T.E., Bárkányi, Zs., Beke, A. and Lulich, S.M., “Patterns of Hungarian vowel production and perception with regard to subglottal resonances”, *The Phonetician*, (in print), 2010/2009.
- [6] Stevens, K.N., “Acoustic Phonetics”, Cambridge: Cambridge University Press, 1998.
- [7] Kovács, M., “Az artikulációs konfiguráció módosulásának kérdése” [On the change of articulatory configuration], talk on the VII. Nemzetközi Magyar Nyelvtudományi Kongresszus [7th International Congress on Hungarian Linguistics], Budapest, 2004.
- [8] Beke, A. and Grácz, T.E. “A magánhangzók semlegesedése a spontán beszédben” [Vowel neutralization in spontaneous speech], in Navracsics, J. [ed.], *Nyelv, beszéd, írás. Pszicholingvisztikai tanulmányok 1* [Language, speech, writing. Psycholinguistic studies 1.] Tinta Könyvkiadó, 57–65, 2010.
- [9] Siptár, P. and Törkenczy, M. “The Phonology of Hungarian”, Oxford University Press, 2000.
- [10] Farnetani, E., “Coarticulation and connected speech processes”, in Hardcastle, W.J. and Laver, J. (ed.) *The Handbook of Phonetic Sciences*. Blackwell, 371–404, 1999.
- [11] Mihajlik, P., Révész, T. and Tatai, P., “Phonetic transcription in automatic speech recognition”, *Acta Ling. Hun.* 49(3-4):407–425, 2002.
- [12] Boersma, P. and Weenink, D., “Praat [Computer program] (Version 5.0.20)”, <http://www.praat.org>, accessed on 11 Dec 2008.
- [13] Sjölander, K. and Beskow, J., “Wavesurfer [Computer program] (Version 1.8.5)”, <http://www.speech.kth.se/wavesurfer/>, accessed on 3 Apr 2009.
- [14] Bolla, K., “Magyar fonetikai atlasz. A szegmentális hangszerkezet” [Phonetic atlas of Hungarian], Nemzeti Tankönyvkiadó, 1995.
- [15] Lulich, S. M., “On the relation between locus equations and subglottal resonances”, *POMA* 5:060003, 2009.
- [16] Csapó, T. G., and Németh, G., “Mássalhangzó-magánhangzó kapcsolatok automatikus osztályozása szubglottális rezonanciák alapján”, [Consonant-vowel classification based on subglottal resonances], *Magyar Számítógépes Nyelvészeti Konferencia*, Szeged, 3–4th December, 226-237, 2009.
- [17] Lindblom, B., “Spectrographic study of vowel reduction”, *J. of Acoust. Soc. Am.*, 35:1773–81, 1963.