Context and speaker dependency in the relation of vowel formants and subglottal resonances

Evidence from Hungarian

INTERSPEECH 2011

Tekla Etelka Gráczi[1], Steven M. Lulich[2], Tamás Gábor Csapó[3], András Beke[1,4]

[1]Research Institute for Linguistics, Hungarian Academy of Sciences, Budapest, Hungary, graczi@nytud.hu

[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

Background: subglottal resonances

- Resonances of the human subglottal tract are fairly constant for a given speaker (no moving articulator) [1]: (Sg1~600Hz, Sg2~1400Hz and Sg3~2100Hz).
- Subglottal resonances (SGRs) can distort spectral peaks of formants [1], therefore ☞ speakers avoid putting vowel formants in these regions [2].
- SGRs have been claimed to be natural divisions between +/- values of several distinctive features, e.g.:
  - Sg1 is a boundary between low and non-low vowels,
  - Sg2 is a boundary between back and non-back vowels.
  - American English [2], Spanish [4], German [5], Korean [6], Hungarian [7,8]
- However, some vowel realizations contradict the subglottal hypothesis : Hungarian [7], [8]:
  - Some speakers produced some/most of their
    - low [ɛ] & [ɔ] with F1 < Sg1,
    - back [ɔ] with F2 > Sg2, and front [aː] & [ø] with F2 < Sg2.

☞ **Hypothesis:**
Besides speaker-dependency vowel target undershoot due to consonant coarticulation is responsible for masking the formant-SGR relations.

Methods: recordings & measurements

Recordings

- 2 male (m1, m2) and 2 female (f1, f2) adult native speakers of Standard Hungarian
- Utterances: "C_1VC_2V" nonsense words in the carrier sentence „*Most a C_1VC_2V szót olvasom.*" ('I am reading the word C_1VC_2V now.')
  - C_1 & C_2 : [b, d, ɟ, g],
    ⇨ 4 X 4 consonant contexts.
  - V: [ɔ, aː, ɛ, ø]
    - [ɔ]: low/mid-low, back, rounded
    - [aː]: low, unrounded, [+back], but phonetically front for most speakers [9,10]
    - [ɛ]: low/mid-low, front, unrounded
    - [ø]: mid, front, rounded
- 6 repetitions per nonsense words
- Simultaneous microphone and accelerometer recordings

Measurements:

- First and second formant (F1, F2) from microphone signals:
  - Measured semi-automatically using Praat + manual correction.
  - Measured 21 times during each vowel at regular intervals.
  - Measured at time of highest F1 (F1m, F2m –see fig. 2).
- First and second SGRs (Sg1, Sg2) from accelerometer signals :
  - Measured manually 25 times for each speaker.
  - Means were considered ground truth.

Means of SGRs (Hz)

	m1	m2	f1	f2
Sg1	607	555	624	500
Sg2	1290	1348	1536	1431

Results

Inter-speaker effects:

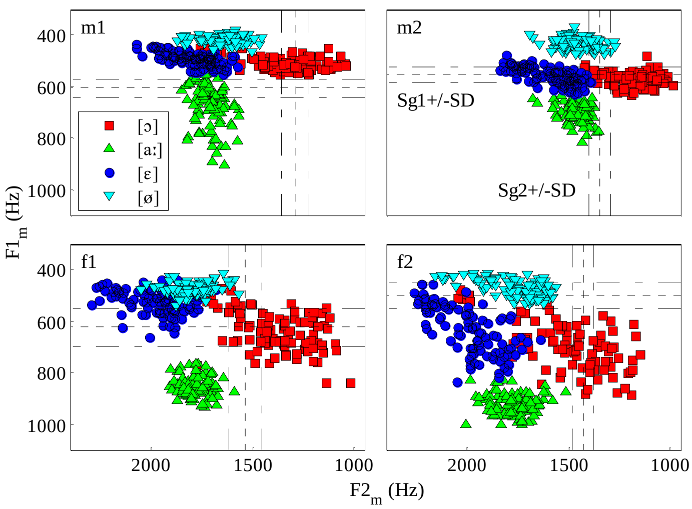


Figure 1. *Vowel spaces for F1m vs. F2m.* The vertical dashed lines indicate the speakers' mean Sg2 +/- one standard deviation. The horizontal dashed lines indicate the mean Sg1 +/- one standard deviation.

- F1m on the expected side of Sg1:
  - **[ɔ]: 0–96.9%,**
  - [aː]: 79.1 for m1, 100% for the other 3 speakers,
  - **[ɛ]: 0–93.8%,**
  - [ø]: 76.1 for f2, 100% for the other three speakers.
- F2m on the expected side of Sg2:
  - **[ɔ]: 40.6–90.7%,**
  - [aː]: 100% for all speakers,
  - [ɛ]: 100% for all speakers,
  - [ø]: 90.8% for m2, 100% for the other 3 speakers.

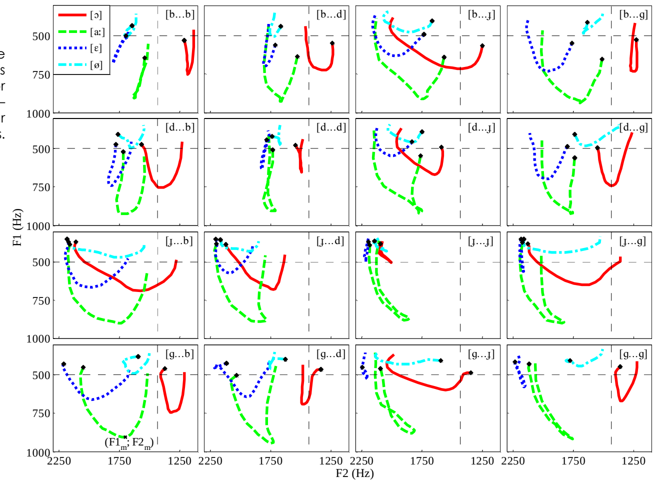
☞ F2-values of [aː] & [ɛ] show no recalcitrance in the present study.
☞ F1-values of [aː] & both F1 & F2 of [ø] contradicts the subglottal hypothesis only in one speaker's pronunciation. (F1-values appear on the wrong side in above 20%, F2 of [ø] in less than 10%).
☞ F1 of [ɛ] and both F1 & F2 of [ɔ]: speaker dependency:
- ☑ : F is usually on the expected side of the Sg,
- ☒: F is usually on the unexpected side of the Sg,
- ☑☒: F is sometimes on the expected, sometimes on the unexpected side of Sg.

	Speaker			
	m1	m2	f1	f2
F1 of [ɔ]	☒	☑	☑☒	☑
F1 of [ɛ]	☒	☑☒	☒	☑
F2 of [ɔ]	☑☒	☑	☑	☑☒

The Chi square analysis for F1 of [ɛ] & [ɔ], and F2 of [ɔ] ($p < 0.001$ in all cases; χ^2: between: 101.641and 244.242; Cramer's *V*: between 0.265 and 0.779) shows significant effect of speaker-dependency.

Context effects:

Fig 2.: Example average formant trajectories based on the results for speaker f2 – representative for our present subjects.



[ɟ] & [g] pull the F1 of [ɛ] closer to / on the wrong side of the SGR.

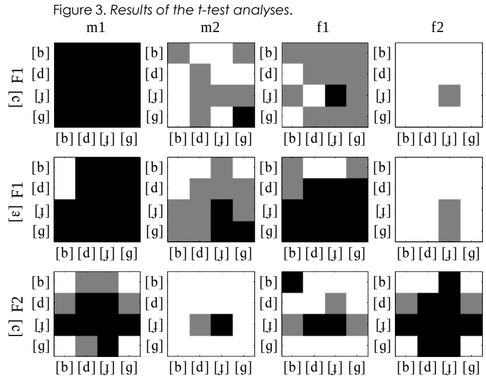
[ɟ] & [d] pull the F2 of [ɔ] closer to / on the wrong side of the SGR.

Both effects are enhanced if both consonants are [d] and/or [ɟ].

Chi-square test:
significant effect of context across all speakers:

	Both Cs	C_1		C_2	
	p	χ^2	Cramer's *V*	χ^2	Cramer's *V*
F1 of [ɔ]	< 0.05	10.885	0.143	12.518	0.157
F1 of [ɛ]	< 0.03	9.000	0.152	13.241	0.184
F2 of [ɔ]	< 0.001	114.185	0.545	41.165	0.327

Figure 3. *Results of the t-test analyses.*



T-test:

- T-tests between F1m and Sg1, and between F2m and Sg2 (cf. Figure 3)
  - Significant difference in wrong direction: black squares.
  - Significant difference in right direction: white squares.
  - No significant difference: gray squares.
- F1 of [ɛ] & [ɔ] is more likely to be on the wrong side in a velar or palatal context.
- F2 of [ɔ] is more likely to be on the wrong side in an alveolar or palatal context.

☞ The effect of the context are the same order or smaller than the inter-speaker effects, but they are also more consistent across speakers.

F1: articulations utilizing the tongue body ([ɟ, g]) require increased jaw height (which correlates inversely with F1) and palatal articulation requires a relatively long constriction which further constrains the jaw height (for Hungarian palato- and linguographic results: [11])
F2: alveolar and palatal stops have a high F2 locus able to exert coarticulatory pressure on F2m [2,12,13]

Conclusions & further questions

- The recalcitrant vowels [ɛ] & [ɔ] showed **both speaker and context dependency**:
  ☞ The data suggest that coarticulatory context effects can mask the formant-SGR relations. However, more data from more speakers are needed to be able to draw general conclusions.
- The vowels were on the wrong side of the SGRs less often than in previous Hungarian studies [7,8].
  ☞ This could be due to interspeaker differences or to differences in phonetic context (unstressed syllables in the former studies, stressed syllables in the present study).

?: What impact do these context- and speaker-dependent effects have on vowel perception?

Key references

[1] Stevens, K.N., *Acoustic Phonetics*, Cambridge, MA: MIT Press,1998.
[2] Lulich, S.M., "Subglottal resonances and distinctive features,"J. Phon., 2010.
[3] Gray, H., *Anatomy of the human body*. Philadelphia:Lea & Febiger, 1918.
[4] Wang, S., Alwan, A., Lulich, S.M., "Speaker normalization based on subglottal resonances," *ICASSP*, 4277–4280, 2008.
[5] Madsack, A., Lulich, S. M., Wokurek, W., Dogil, G., "Subglottal resonances and vowel formant variability: A case study of high German monophthongs and Swabian diphthongs," in *Proc. LabPhon*, 11:91–92, 2008.
[6] Jung, Y., "Subglottal effects on the vowels across language: Preliminary study on Korean," JASA 125:2638, 2009.
[7] Csapó, T.G., Bárkányi, Zs., Gráczi ,T.E., Bőhm, T., Lulich, S.M., "Relation of formants and subglottal resonances in Hungarian vowels", in Proc. Interspeech 484–487, 2009.
[8] Csapó, T.G., Gráczi T.E., Bárkányi, Zs., Beke, A., Lulich, S.M., "Patterns of Hungarian vowel production and perception with regard to subglottal resonances", in The Phonetician (in print)
[9] Siptár, P., Törkenczy, M., "The Phonology of Hungarian", Oxford University Press, 2000.
[10] Kovács, M., "Az artikulációs konfiguráció módosulásának kérdése", talk on the VII. Nemzetközi Magyar Nyelvtudományi Kongresszus, Budapest, 2004.
[11] Bolla, K., "Magyar fonetikai atlasz. A szegmentális hangszerkezet", Nemzeti Tankkönyvkiadó, 1995.
[12] Lulich, S.M., "On the relation between locus equations and subglottal resonances", POMA 5:060003, 2009.
[13] Csapó, T.G., Németh, G., "Mássalhangzó-magánhangzókapcsolatok automatikus osztályozása szubglottális rezonanciák alapján", MSZNY, 226-237, 2009.

Acknowledgements

T. E. G. was supported **by ISCA-Google grant**.
S. M. L. was supported in part by **NSF Grant No. 0905250**,
T. G. Cs. Was supported 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.