Relation of formants and subglottal resonances in Hungarian vowels

Tamás Gábor Csapó[1], *Zsuzsanna Bárkányi*[2], *Tekla Etelka Gráczi*[2], *Tamás Bőhm*[1,3], *Steven M. Lulich*[4]

[1]Department of Telecommunications and Media Informatics,
Budapest University of Technology and Economics, Budapest, Hungary
[2]Research Institute for Linguistics, Hungarian Academy of Sciences, Budapest, Hungary
[3]Institute for Psychology, Hungarian Academy of Sciences, Budapest, Hungary
[4]Speech Communication Group, MIT, Cambridge, MA 02139
`csapot@tmit.bme.hu, bzsu@budling.nytud.hu, graczi@nytud.hu,`
`bohm@tmit.bme.hu, lulich@speech.mit.edu`

Abstract

The relation between vowel formants and subglottal resonances (SGRs) has previously been explored in English, German, and Korean. Results from these studies indicate that vowel classes are categorically separated by SGRs. We extended this work to Hungarian vowels, which have not been related to SGRs before. The Hungarian vowel system contains paired long and short vowels as well as a series of front rounded vowels, similar to German but more complex than English and Korean. Results indicate that SGRs separate vowel classes in Hungarian as in English, German, and Korean, and uncover additional patterns of vowel formants relative to the third subglottal resonance (Sg3). These results have implications for understanding phonological distinctive features, and applications in automatic speech technologies.

Index Terms: subglottal resonances, Hungarian, quantal theory, vowels

1. Introduction

Since the advent of formal generative phonology [1], the role of phonetic and functional factors in shaping and explaining sound patterns and phonological processes has been approached from several theoretical perspectives (e.g. [2, 3, 4, 5, 6]). One of the most successful theories which tries to define the phonetic basis of phonological distinctive features is Quantal Theory (QT) [3, 4, 7]. QT relies on the claim that "the relation between an acoustic parameter that can be observed in the sound and an articulatory parameter that can be manipulated by a speaker takes a particular non-monotonic form" [4, p. 3]. This means that in some regions of the articulator space small movements lead to big acoustic changes, while in other regions large movements lead to small acoustic changes. The latter are the stable regions which underlie distinctive features. This also means that phonological systems will use the acoustically unstable regions as a dividing line between the $+$ and $-$ value of a distinctive feature. One set of unstable regions (for vowels) arises from acoustic coupling between the vocal tract and the subglottal airway. Since the subglottal system does not have moving articulators during speech production, subglottal resonances (SGRs) are fairly constant for a given speaker (typical values of SGRs can be found in [8, 9]). SGRs can distort the spectral peaks of formants if they fall nearby, such as when F2 and the 2nd subglottal resonance (Sg2) are close [7]. Therefore, it is hypothesized that speakers will avoid putting vowel formants in these regions. For instance, Sg2 forms a natural division between front and back vowels [7], Sg1 forms a division between low and non-low vowels [7], and Sg3 forms a division between the tense and lax front vowels [10]. For the same vowel produced at different times or in different contexts, formants are thought to be free to vary only within the frequency bands defined by the subglottal resonances [10, 11].

In recent studies it has been shown that SGRs 1) can cause discontinuities in formant trajectories [12], 2) are salient in speech perception [13], and 3) are useful in speaker normalization [14, 15]. However, the relationship between SGRs and vowel formants has been studied for only a few languages. Wang and colleagues [14, 15] tested speaker normalization applications on the speech of American English-Spanish bilingual children. Lulich [10] studied the relation between Sg2 and F2 in one adult native speaker and nine child speakers of American English. Madsack and colleagues [11] investigated the relation between Sg2 and F2, and Sg1 and F1, in two German dialects (High German and Swabian), and Jung [16] did the same in Korean. Although the data is still relatively sparse, in each case the results were consistent with the interpretation that SGRs form divisions between classes of vowels across languages.

In this paper we present evidence that SGRs define vowel distinctive features in Standard Hungarian. The following hypotheses were tested: 1) Sg2 is a boundary between front and back vowels; 2) Sg1 is a boundary between low and non-low vowels; 3) front rounded and front unrounded vowels can be differentiated in terms of Sg3.

1.1. The vowel system of Hungarian

The vowel system of Standard Hungarian contains 7 short vowels and 7 long vowels (all monophthongs). Phonologically, the short and long vowels are paired together, as illustrated in Table 1 [17, p. 47]. Note that the short vowel [ɛ] has the long pair [eː].

Phonetically, the low vowel pairs differ in quality as well as quantity. The differences in quality are based on differences in

Table 1: *Phonological classification of Hungarian vowels.*

	front				back	
	unrounded		rounded			
	short	long	short	long	short	long
high	i	iː	y	yː	u	uː
mid		eː	ø	øː	o	oː
low	ɛ				ɔ	aː

articulation, as illustrated in Table 2 [17, p. 44]. Acoustic differences are revealed in our analysis, as well (see below). There is no general agreement to what extent mid and high vowel pairs differ in quality (see [18, 19]).

Table 2: *Articulatory characteristics of Hungarian vowels (s. = short, l. = long).*

	front				central		back	
	unrounded		rounded		unrounded		rounded	
	s.	l.	s.	l.	s.	l.	s.	l.
high	i	iː	y	yː			u	uː
mid-high		eː	ø	øː			o	oː
mid-low	ɛ						ɔ	
low						aː		

2. Methods

2.1. Recordings

Acoustic data were collected from two male and two female adult native speakers of Standard Hungarian, aged between 22 and 38 years (denoted M1, M2, F1, F2). The speakers produced several utterances of “ɔCVCɔ” nonsense words, where the target vowel was inserted between two voiced plosives, and the first consonant was [b,d,g]. For speakers M1 and M2 the second consonant was fixed as [b], whereas for speakers F1 and F2 it was fixed as [d]. The target vowel was placed in the second (unstressed) syllable. Note that there is no vowel reduction in unstressed syllables in Hungarian. All Hungarian vowels ([ɔ,aː,o,oː,u,uː,ɛ,eː,i,iː,ø,øː,y,yː]) were included. The female participants repeated the nonsense words five times each, for a total of 15 utterances per vowel, while the males repeated each word three times, for a total of 9 utterances per vowel. The recordings were made in an anechoic chamber with a Monacor EMC 100 condenser microphone at a distance of approximately 15cm from the lips. The SGRs were recorded using a K&K HotSpot accelerometer attached to the skin of the neck below the thyroid cartilage. The two signals were digitized at 8kHz with a Terratec DMX 6 Fire USB external sound card and were recorded to separate channels using Wavesurfer (http://www.speech.kth.se/wavesurfer).

2.2. Measurements

The first three formants of each vowel were measured at the vowel midpoint from the microphone signal, and the first three SGRs were measured from the accelerometer signal. The microphone recordings were segmented semi-automatically with the Hungarian trained version of the MAUS forced-alignment program (http://www.phonetik.uni-muenchen.de/forschung/Verbmobil/VM14.7eng.html), and the formants were measured automatically in Praat (http://www.praat.org) and hand-corrected. SGRs were measured by hand at least 25 times for each speaker, using Wavesurfer. The median values are given in the last column of Table 3. Fig. 1 shows a spectrogram of the accelerometer signal from speaker M2, indicating that the measurement of SGRs is similar to reading off formants (c.f. [10, 12]).

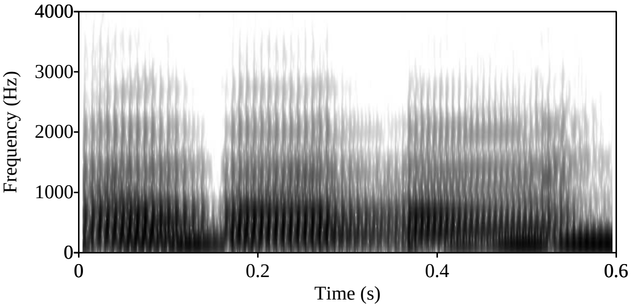

Figure 1: *Sample spectrogram of speaker M2’s accelerometer recording for the nonsene word “ɔdɔbɔ”.*

3. Results and Discussion

Vowel plots for each speaker are shown in Fig. 2, with the SGRs indicated by straight horizontal and vertical lines. In general, the vowel space is clearly divided by the subglottal resonances. Sg1 (the horizontal line) separates low and non-low vowels, while Sg2 (right vertical line) divides front and back vowels. Sg3 (left vertical line) divides the front unrounded vowels, [i,iː,eː], from their rounded counterparts, [ø,øː,y,yː].

There are some exceptions to these rules, however. For speaker M1, for instance, the vowels [ɛ] and [ɔ] have F1 values lower than Sg1, and F1 values for [ɛ] appear to be very close to Sg1 in speakers M2 and F2. There are three potential explanations for this. 1) The resonances in the accelerometer signal are not as clear as the formants in the microphone signal. This is due to the larger damping of the subglottal system [7], as well as the low-pass nature of the neck tissues. Often being close to F0, these factors make accurate measurement of Sg1 difficult. A further difficulty was noted for speaker M1, namely, that coupling between the vocal tract and the subglottal system was particularly strong, leading to a strong influence of the first formant on the spectrum in the region of Sg1. It is therefore possible that Sg1 was not measured accurately and the true Sg1 is at a lower frequency. For speaker M1 this is particularly likely because the measured Sg1 was disproportionately high compared with Sg2 and Sg3. 2) Even in laboratory speech, co-articulation is always present, and it is possible that the [ɛ] and [ɔ] tokens with low F1 were due to co-articulation with the consonants. 3) It is equally possible that the SGRs simply do not separate low and non-low vowels as cleanly as hypothesized.

In addition, note that the vowel [ɔ] has F2 lower than Sg2 in speakers M1, M2, and F2, but not in speaker F1. Madsack and colleagues [11] found that F2 in the German low vowel [a] was either categorically above Sg2 or categorically below it, depending on the speaker, and Jung [16] found that the F2 of [a] in Korean depended on the neighboring consonant place of articulation. It is possible that similar variations of low vowels will frequently be observed across languages.

Figures 3-5 show the frequency-normalized distributions of F1 and F2 relative to the SGRs for all four speakers. In Fig. 3, for instance, the raw F1 values for each speaker were normalized with respect to Sg1 and then pooled together. As can be seen, Sg1 separates low vowels from non-low vowels. In Fig. 4, raw F2 values were normalized with respect to Sg2. Front vowels and back vowels are separated by Sg2. In Fig. 5, raw F2 values were normalized with respect to Sg3. Unrounded non-low vowels, [i,iː,eː], are separated from their rounded counterparts, [y,yː,ø,øː], by Sg3.

The results suggest that the SGRs can serve as a threshold in classifying vowels into distinctive features by their formant frequencies. We therefore analyzed receiver operating characteristics (ROC) curves for each SGR separately for each speaker (graphs not shown). In each case, the ROCs showed that a range of frequencies optimally separated the different categories of vowels (e.g. front vs. back vowels). These ranges are given in Table 3.

The median SGRs mostly fall within one standard deviation

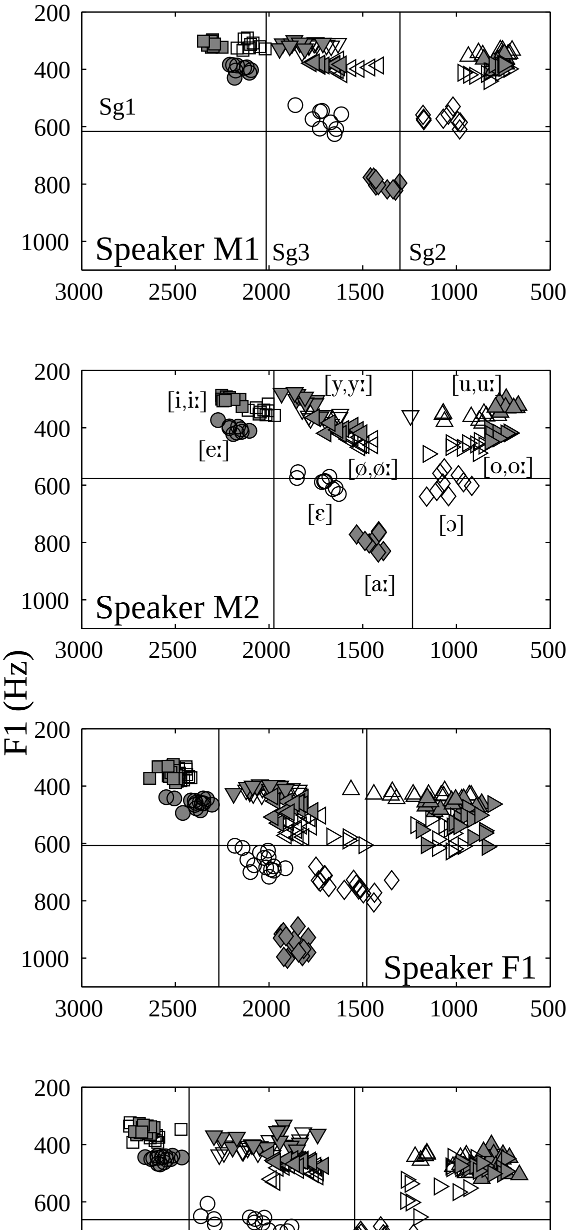

Figure 2: *Vowel plots of the four speakers. Horizontal and vertical lines indicate subglottal resonance frequencies. Empty symbols represent short vowels, and filled symbols are long vowels. The vowel identities are labeled in the vowel plot for speaker M2.*

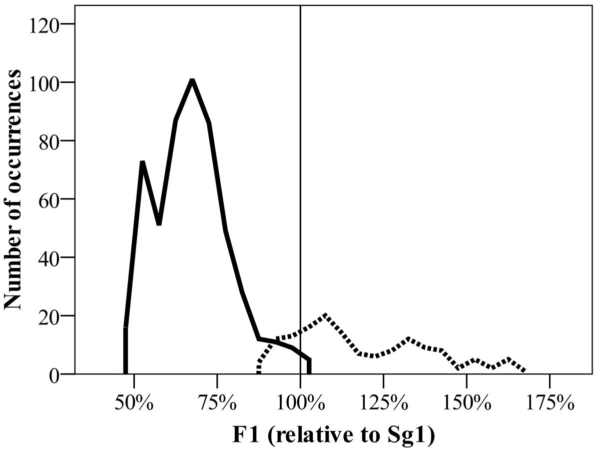

Figure 3: *Normalized distributions of F1, for all speakers. The solid line corresponds to all low vowels, while the dashed to all of the non-low ones.*

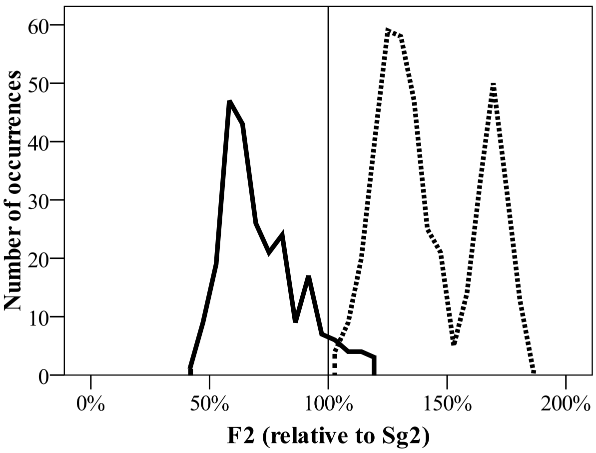

Figure 4: *Normalized distributions of F2, for all speakers. The solid line corresponds to all front vowels, while the dashed to all of the back ones.*

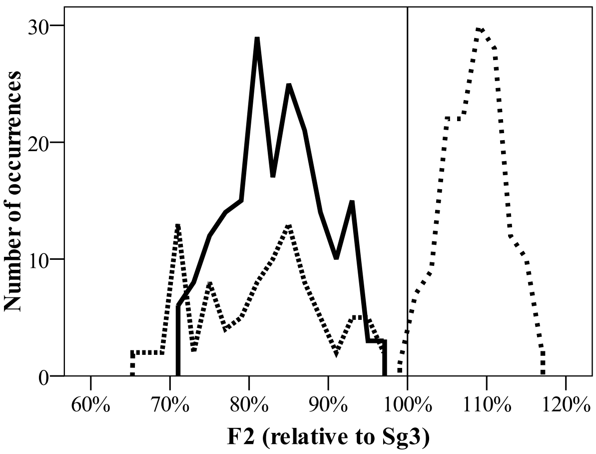

Figure 5: *Normalized distributions of F2 for front vowels, for all speakers. The solid line corresponds to all rounded vowels, the left dashed to all of the unrounded low vowels, while the right dashed line to all of the unrounded non-low.*

Table 3: *Results of ROC analysis (all numbers in Hz).*

SGR	Speaker	Optimal range	SGR ± 1 S.D.	Median SGR
Sg1	F1	609 – 616	576 – 638 *	607
Sg1	F2	596 – 606	601 – 723 *	662
Sg1	M1	430 – 525	527 – 707	617
Sg1	M2	488 – 541	496 – 658 *	577
Sg2	F1	1735 – 1766	1393 – 1563	1478
Sg2	F2	1523 – 1688	1445 – 1641 *	1543 *
Sg2	M1	1175 – 1304	1251 – 1351 *	1301 *
Sg2	M2	1151 – 1245	1152 – 1318 *	1235 *
Sg3	F1	2182 – 2304	2070 – 2466 *	2268 *
Sg3	F2	2328 – 2465	2389 – 2463 *	2426 *
Sg3	M1	1927 – 2020	1933 – 2097 *	2015 *
Sg3	M2	1863 – 1971	1898 – 2050 *	1974

of the optimal threshold range. In six cases out of 12, the median SGR is within the optimal range (denoted by asterisks in the Median SGR column). In four of the remaining six cases, the median SGR is within one standard deviation of the optimal range (denoted by asterisks in the SGR ± 1 S.D. column). The two cases in which the median SGR is not within one standard deviation of the optimal range are due to the same exceptions discussed above with regard to the vowels [ɛ] and [ɔ]. It appears that, with some exceptions, SGRs are near optimal division lines between the formants of low vs. non-low, front vs. back, and front unrounded non-low vs. front rounded non-low vowels.

4. Conclusions

Our results confirm that Sg2 is a reliable boundary between front and back vowels in Hungarian. It is interesting to note that the (mid-)low back vowel [ɔ] for speaker F1 appears in the "front" region. This speaker's vowel space occupies a much smaller region in F1-F2 space than we observe for the other speakers (and is usually assumed). The vowel space in this case as well is neatly divided by Sg2. This finding is in accordance with earlier findings that the relative frequency of F2 and Sg2 for the low back vowel [a] in German is speaker dependent [11]. A study on a much larger set of data would shed light on whether speaker F1's small vowel space is exceptional or follows a general pattern. Note that phonologically both [ɔ] and [aː] are back vowels while for this speaker they are acoustically (central)/front.

The dividing line between low and non-low vowels is somewhat less clear in our data. A reason for this might be that it is much more difficult to measure Sg1 than Sg2. "The effect on the vowel spectrum of the lowest subglottal resonance is generally not as pronounced as that of the second subglottal resonance since the partially open glottis introduces acoustic losses that reduce the degree of prominence of this lowest resonance" [7, p. 300]. It is also very close to the fundamental frequency which makes accurate measurement difficult.

We also found that Sg3 reliably differentiates front rounded vowels, [ø,øː,y,yː], from front unrounded vowels, [i,iː,eː]. The vowel [ɛ] is exceptional in all speakers in this regard, which is presumably related to the fact that it differs from [eː] in quality (height) as well as quantity.

These results have implications for understanding phonological distinctive features, and applications in automatic speech technologies. The latter may include speaker normalization and other subproblems of automatic speech recognition.

5. Acknowledgements

We are grateful to Katalin Nagy and Péter Herman for their help in the data analysis and to György Szaszák and Dávid Sztahó for their assistance in the recordings. T. G. Cs. and T. B. were supported by the Hungarian government under contracts NKFP 2/034/2004, Jedlik OM-00102/2007 and TÁMOP-4.2.2-08/1/KMR-2008-0007. Zs. B. was supported by HNRF PD050018 grant. T. G. Cs. was partially supported by the Bizáki Puky Péter Memorial Foundation.

6. References

[1] Chomsky, N. and Halle, M., *The Sound Pattern of English*, Cambridge, MA: MIT Press, 1968.

[2] Liljencrants, J. and Lindblom, B., "Numerical simulation of vowel quality systems: The role of perceptual contrast," Language 48:839–862, 1972.

[3] Stevens, K. N., "The quantal nature of speech: Evidence from articulatory-acoustic data," in P. B. Denes and E. E. David, Jr. [Eds], *Human Communication, A Unified View*, New York: McGraw-Hill, 54–66, 1972.

[4] Stevens, K. N., "On the quantal nature of speech," J. Phon 17:3–45, 1989.

[5] Ohala, J. J., "The origin of sound patterns in vocal tract constraints," in P. F. MacNeilage [Ed] *The Production of Speech*, New York: Springer-Verlag 189–216, 1983.

[6] Hayes, B., Kirchner, R., Steriade, D., *Phonetically-Based Phonology*, Cambridge: Cambridge University Press, 2004.

[7] Stevens, K. N., *Acoustic Phonetics*, Cambridge, MA: MIT Press, 1998.

[8] Ishizaka, K., Matsudaira, M., Kaneko, T., "Input acoustic-impedance measurement of the subglottal system," JASA 60:190–197, 1976.

[9] Cranen, B. and Boves, L., "On subglottal formant analysis," JASA 81:734–746, 1987.

[10] Lulich, S. M., "Subglottal resonances and distinctive features," Journal of Phonetics, doi:10.1016/j.wocn.2008.10.006, 2009.

[11] 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.

[12] Chi, X. and Sonderegger, M., "Subglottal coupling and its influence on vowel formants", JASA 122:1735–1745, 2007.

[13] Lulich, S. M., Bachrach, A., Malyska, N., "A role for the second subglottal resonance in lexical access," JASA 122:2320–2327, 2007.

[14] Wang, S., Alwan, A., Lulich, S. M., "Speaker normalization based on subglottal resonances," in *Proc. ICASSP*, 4277–4280, 2008.

[15] Wang, S., Lulich, S. M., Alwan, A., "A reliable technique for detecting the second subglottal resonance and its use in cross-language speaker adaptation," in *Proc. Interspeech* 1717–1720, 2008.

[16] Jung, Y., "Subglottal effects on the vowels across language: Preliminary study on Korean," JASA 125:2638, 2009.

[17] Nádasdy, Á. and Siptár, P., "A magánhangzók," [Vowels], in F. Kiefer [Ed] *Strukturális magyar nyelvtan 2. Fonológia [A Structural Grammar of Hungaran 2. Phonology]*, Budapest: Akadémiai Kiadó (Publishing House of the Hungarian Academy of Sciences), 42–181, 1994.

[18] Kovács, M., "Tendenciák és szabályszerűségek a magánhangzó-időtartamok produkciójában és percepciójában," [Trends and regularities in the production and perception of vowel length], PhD dissertation, University of Debrecen, 2002.

[19] Mády, K. and Reichel, U. D., "Quantity distinction in the Hungarian vowel systems - Just theory or also reality?," in *Proc. ICPhS*, Saarbrücken, 1053–1056, 2007.