

Mássalhangzó-magánhangzó kapcsolatok automatikus osztályozása szubglottális rezonanciák alapján

Csapó Tamás Gábor, Németh Géza
{csapot,nemeth}@tmit.bme.hu

BME Távközlési és Médiainformatikai Tanszék

VI. Magyar Számítógépes Nyelvészeti Konferencia,
2009. december 4.



1 Bevezetés

- A szubglottális rendszer
- Formánsmenetek hangkapcsolatokban

2 Módszerek

- Felvételek, mérések

3 Eredmények

- Locus egyenlet
- Osztályozás

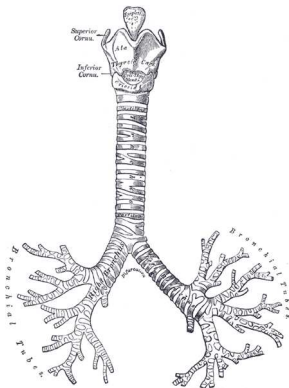
4 Összefoglalás

- Következtetések
- További kutatási irányok



A szubglottális rendszer, rezonanciái

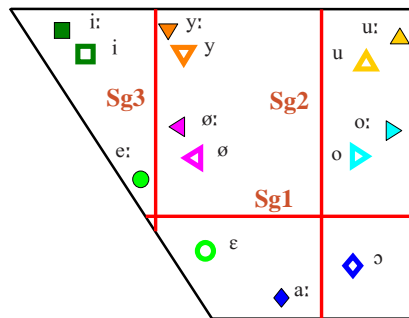
- alsó légúti rendszer (tüdő, légcső)
- **szubglottális rezonanciák (SGR)**
 - Sg1 ~600 Hz
 - Sg2 ~1400 Hz
 - Sg3 ~2100 Hz
- korábbi kutatások
 - SGR ↔ formáns
 - magánhangzók: kerülnek az SGR környéki frekvenciákat
 - angol, spanyol, német, koreai, magyar



Magyar magánhangzók

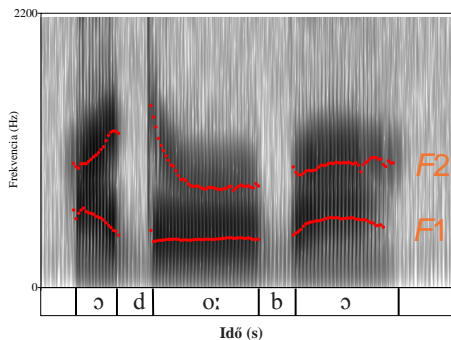
- korábbi eredmények, magyar nyelvre
- 2 férfi és 2 nő, logatomok
- SGR-ek: határ a magánhangzócsoportok között
 - Sg1: **alsó** ↔ **nem-alsó**
 - Sg2: **elöl** ↔ **hátsó** képzett
 - Sg3: **elöl képzett ajakkerek.** ↔ **ajakréses**

F1-F2 formánstér



Formánsmenetek CV kapcsolatokban

- koartikuláció
 - mássalhangzó (C) + magánhangzó (V)
 - változó F2 érték
- locus egyenlet
 - lineáris regresszió
 - $F2_{msh} = m \cdot F2_{mgh} + b$
 - képzési helytől függ
- locus egyenlet tér
 - CV csoportok
- hipotézis
 - SGR-ek \rightarrow CV csoportok elkülönülése



Felvételek, mérések (formánsok)

● felvételek

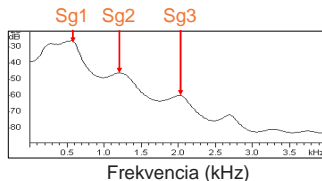
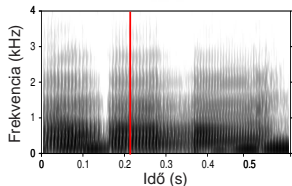
- 1 férfi beszélő ("B1", 29 éves)
- beszéd felvételek (logatom: "ɔCVbɔ")
- C: összes zárhang
 - alveoláris: [d,t], labiális: [b,p], palatális: [tʃ,c], veláris: [g,k]
- V: összes magánhangzó
 - első: [i,iː,e,ɛ,a,e,ø,øː,y,yː], hátsó: [ɔ,o,oː,u,uː]
- összesen 1120 logatom

● formánsmérések

- automatikus címkézés (BME TMIT) + kézi ellenőrzés
- automatikus formánsmérés (Praat) + kézi ellenőrzés
- $F2_{msh}$, mássalhangzó végén (C, 95%)
- $F2_{mgh}$, magánhangzó közepén (V, 50%)

Felvételek, mérések (SGR)

- felvételek
 - 1 férfi beszélő ("B1", 29 éves)
 - folyamatos beszéd
 - gyorsulásmérő felvételek
- szubglottális rezonancia
 - manuális mérés (Wavesurfer, LPC)
 - Sg1, Sg2 és Sg3
 - 20 mérési pont / SGR



	Sg1	Sg2	Sg3
Átlag	545 Hz	1241 Hz	2027 Hz
Medián	554 Hz	1244 Hz	2022 Hz
Szórás	60 Hz	42 Hz	145 Hz

Locus egyenlet tér

- 1 L, V + H
- 2 A, P + H
- 3 A, L, V + E1
- 4 A, L + E2
- 5 P + E1
- 6 P + E2

A = alveoláris = [d,t]

L = labiális = [b,p]

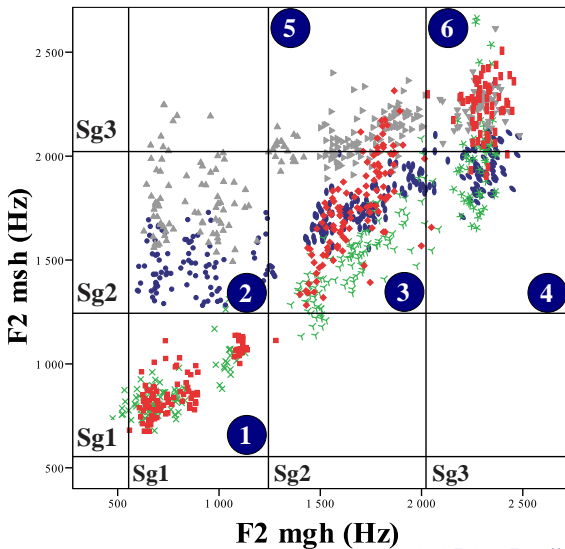
P = palatális = [j,c]

V = veláris = [g,k]

E1 = első/1 = [a,e,ø,ø:,y,y:]

E2 = első/2 = [i,i:,e:]

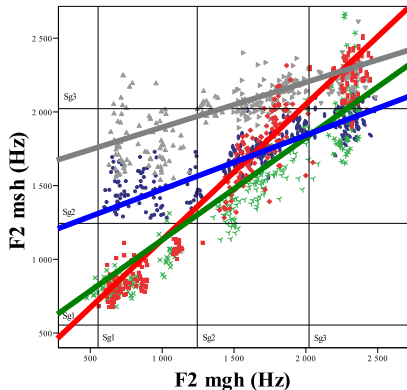
H = hátsó = [ɔ,o,ɔ:,u,u:]



Locus egyenletek (C-osztályok)

● $F2_{msh} = m \cdot F2_{mgh} + b$

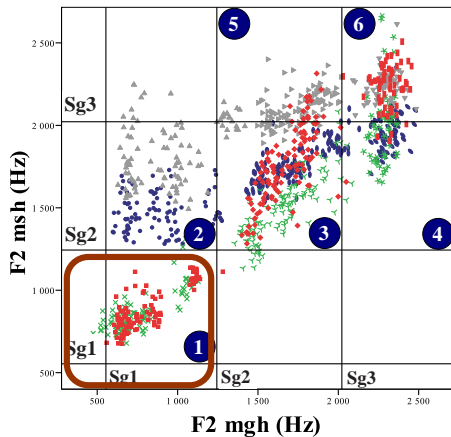
	m	b	R^2
Alveoláris	0,33	1184,35	0,76
Labiális	0,73	301,22	0,91
Palatális	0,30	1552,82	0,62
Veláris	0,91	179,19	0,93



CV kapcsolatok osztályozása

- tartományok, határok

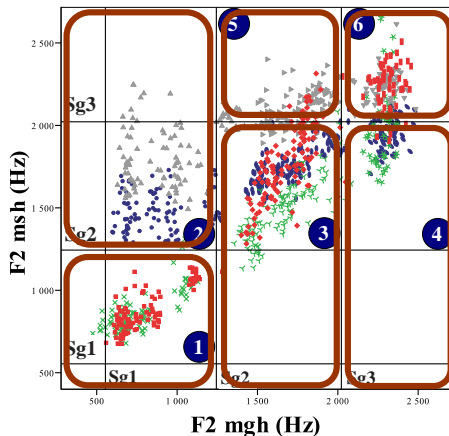
1.	$F2_{msh} < Sg2$ $F2_{mgh} < Sg2$
----	--------------------------------------



CV kapcsolatok osztályozása

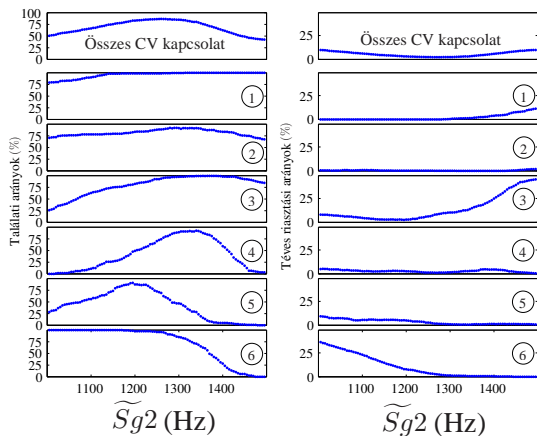
- tartományok, határok

1.	$F2_{msh} < Sg2$ $F2_{mgh} < Sg2$
2.	$Sg2 < F2_{msh}$ $F2_{mgh} < Sg2$
3.	$F2_{msh} < Sg3$ $Sg2 < F2_{mgh} < Sg3$
4.	$F2_{msh} < Sg3$ $Sg3 < F2_{mgh}$
5.	$Sg3 < F2_{msh}$ $Sg2 < F2_{mgh} < Sg3$
6.	$Sg3 < F2_{msh}$ $Sg3 < F2_{mgh}$



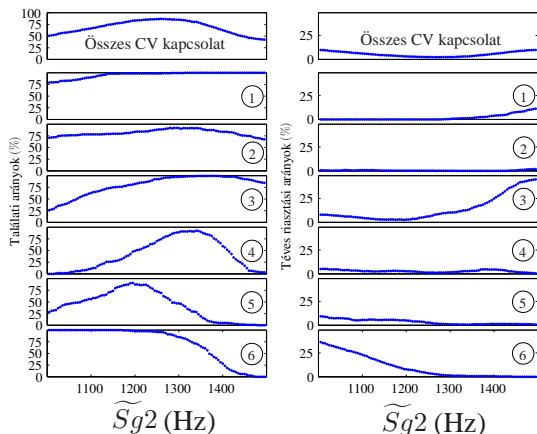
Osztályozás eredménye

- $\widetilde{Sg2} = 1000 \dots 1500 \text{ Hz}$
- $\widetilde{Sg3} = 1,6254 \cdot \widetilde{Sg2}$
("B1" adatai alapján)



Osztályozás eredménye

- $\widetilde{Sg2} = 1000 \dots 1500 \text{ Hz}$
- $\widetilde{Sg3} = 1,6254 \cdot \widetilde{Sg2}$
("B1" adatai alapján)
- tartományonként eltérő optimum
- optimális osztályozás:
 $\widetilde{Sg2} = 1260 \text{ Hz}$



Osztályozás eredménye - összesítve

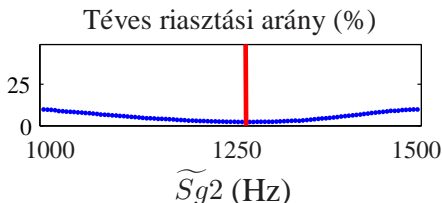
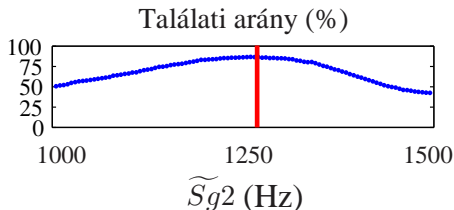
- $\tilde{S}g2 = 1260 \text{ Hz}$

Találat (%)	86,6%
Téves (%)	2,3%

- $Sg2_{meres} = 1244 \text{ Hz}$

Találat (%)	85,5%
Téves (%)	2,4%

Összes CV kapcsolat



Összefoglalás, következtetések

- "B1" beszélő
 - formánsok
 - szubglottális rezonanciák
- locus egyenlet tér: $F2_{msh} - F2_{mgh}$
 - CV csoportok képzési hely szerint
- osztályozás
 - 6 elkülönülő tartomány
 - SGR közel az optimumhoz

- SGR-ek szerepe a beszédben
 - határ a magánhangzócsoporthoz
 - határ a mássalhangzócsoporthoz

További kutatási irányok

- lehetséges alkalmazási területek
 - beszélő normalizálás
 - beszélő felismerés
- további kutatás
 - alternatív SGR mérési módszerek
 - percepciós kísérletek
 - több beszélő vizsgálata
 - más nyelvek vizsgálata

További kutatási irányok

- lehetséges alkalmazási területek
 - beszélő normalizálás
 - beszélő felismerés
- további kutatás
 - alternatív SGR mérési módszerek
 - percepciós kísérletek
 - több beszélő vizsgálata
 - más nyelvek vizsgálata

- Köszönöm a figyelmet!
- csapot@tmit.bme.hu

Irodalomjegyzék

- Boersma, P. and Weenink, D.: Praat (Version 5.1.19), <http://www.praat.org>, (2009).
- Chi, X. and Sonderegger, M.: Subglottal coupling and its influence on vowel formants, *JASA* 122, 1735–1745, (2007).
- Csapó, T. G., Bárkányi, Zs., Grácsi, T. E., Bóhm, T. and Lulich, S. M.: Relation of formants and subglottal resonances in Hungarian vowels, *Proc. Interspeech*, 484–487, (2009).
- Gray, H.: *Anatomy of the human body*, Philadelphia: Lea & Febiger, (1918).
- Gósy, M.: *Fonetika, a beszéd tudománya*, Osiris Kiadó, Budapest, (2004).
- Jung, Y.: Subglottal effects on the vowels across language: Preliminary study on Korean, *JASA* 125, 2638, (2009).
- Lulich, S. M., Bachrach, A. and Malyska, N.: A role for the second subglottal resonance in lexical access, *JASA* 122, 2320–2327, (2007).
- Lulich, S. M.: Subglottal resonances and distinctive features, *J. Phon.*, doi:10.1016/j.wocn.2008.10.006, (2009).
- Lulich, S. M.: On the relation between locus equations and subglottal resonances, *POMA* 5, 060003, (2009).
- Lulich, S. M. and Chen, N. F.: Automatic classification of consonant-vowel transitions based on subglottal resonances and the second formant, *POMA* 6, 060005, (2009).
- 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, *Proc. LabPhon11*, 91–92, (2008).
- Mihajlik, P., Révész, T. and Tatai, P.: Phonetic Transcription in Automatic Speech Recognition, *Acta Linguistica Hungarica*, Vol. 49. (3-4), 407–425, (2002).
- Sjölander, K. and Beskow, J.: Wavesurfer (Version 1.8.5), <http://www.speech.kth.se/wavesurfer>, (2009).
- Stevens, K. N.: On the quantal nature of speech, *J. Phon.* 17, 3–45, (1989).
- Stevens, K. N.: *Acoustic Phonetics*, MIT Press: Cambridge, MA, (1998).
- Wang, S., Lulich, S. M. and Alwan, A.: A reliable technique for detecting the second subglottal resonance and its use in cross-language speaker adaptation, *Proc. Interspeech*, 1717–1720, (2008).
- Wang, S., Lee, Y.-H. and Alwan, A.: Bark-shift based nonlinear speaker normalization using the second subglottal resonance, *Proc. Interspeech*, 1619–1622, (2009).
- Wokurek, W. and Madsack, A.: Comparison of Manual and Automated Estimates of Subglottal Resonances, *Proc. Interspeech*, 1671–1674, (2009).