

## Mássalhangzó-magánhangzó kapcsolatok automatikus osztályozása szubglottális rezonanciák alapján

Csapó Tamás Gábor<sup>1</sup>, Németh Géza<sup>1</sup>

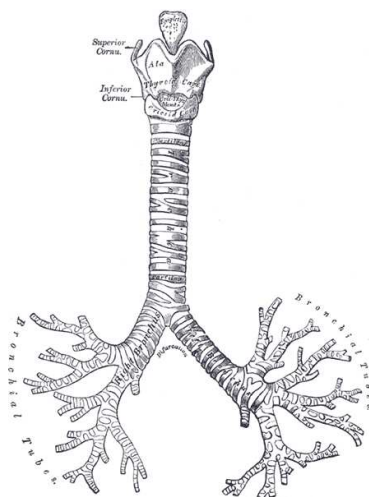
<sup>1</sup>Budapesti Műszaki és Gazdaságtudományi Egyetem,  
Távközlési és Médiainformatikai Tanszék,  
Budapest, 1117, Magyar tudósok krt. 2.,  
e-mail: {csapot,nemeth}@tmit.bme.hu

**Kivonat** A nemzetközi szakirodalom az elmúlt években kezdett intenzíven foglalkozni a szubglottális rezonanciák vizsgálatával, melyek az alsó légutak rezonanciái. Korábbi kutatásokban kimutatták, hogy ezek a magánhangzókat természetes osztályokra tagolják. A mássalhangzó-magánhangzó kapcsolatokban a magánhangzó formánsértékei nem állandóak a koartikuláció miatt. A zárhangok például képzési helyüktől függően módosítják a szomszédos magánhangzó formánsait. A mássalhangzó végén és a magánhangzó közepén mérhető második formáns értékét összevetve rajzolható meg a locus egyenlet tér, melyben az egyes beszédhang-osztályok az artikulációs helyük szerint elkülönülve jelennek meg. Hipotéziseink szerint a csoportok elkülönüléséhez a szubglottális rezonanciák is hozzájárulnak, hasonlóan a magánhangzóknak okozott kategorikus elválasztáshoz. Jelen kutatás során egy magyar anyanyelvű beszélő alapján tovább vizsgáljuk a mássalhangzó-magánhangzó kapcsolatok helyét a locus egyenlet térben, valamint a szubglottális rezonanciák csoportelválasztó szerepét is elemezzük. Bemutatjuk egy automatikus osztályozó működését, amely a szubglottális rezonanciák és a második formáns viszonya alapján csoportosítja a mássalhangzó-magánhangzó beszédhangkapcsolatokat.

**Kulcsszavak:** szubglottális rezonancia, SGR, CV-kapcsolat, locus egyenlet

### 1. Bevezetés

A nemzetközi szakirodalom az elmúlt években kezdett intenzíven foglalkozni a szubglottális rezonanciák (SGR) vizsgálatával, melyek az alsó légutak (pl. tüdő, légcső, hörgők, l. 1. ábra) rezonanciái [15]. Ezek a formánsokhoz hasonlóan alakítják a zöngés hangok spektrumát, de a formánsokkal ellentétben nem erősítik a rezonanciafrekvencia körüli harmonikusokat, hanem gyengítik őket. Mivel az alsó légúti szervek viszonylag keveset mozognak a beszéd során, a rezonanciafrekvenciák közel állandóak egy-egy ember beszédében.



1. ábra. Az alsó légúti rendszer [4]. Tipikus rezonanciafrekvencia értékei 600 Hz, 1400 Hz és 2100 Hz körüliek [15].

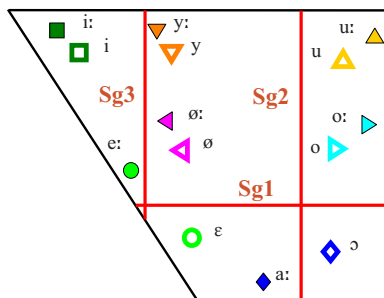
### 1.1. A szubglottális rendszer rezonanciáinak szerepe

Korábbi kutatásokban kimutatták, hogy a szubglottális rezonanciák a magánhangzókat kategorikusan természetes osztályokra tagolják [8]. Az angol nyelven végzett vizsgálatok alapján az derült ki, hogy a második szubglottális rezonancia ( $Sg2$ ) természetes határként (fonológiai megkülönböztető jegy, [14]) szolgál az elől és hátul képzett magánhangzók között: ha a második formáns frekvenciája ( $F2$ ) magasabb, mint a második alsó légúti rezonancia, akkor elől képzett magánhangzóként érzékeljük, ha alacsonyabb, akkor hátul képzettként. Az első ( $Sg1$ ) és harmadik ( $Sg3$ ) alsó légúti frekvencia elválasztó szerepére is utalnak bizonyos eredmények [8].

Az eddigi eredmények szerint a szubglottális rezonanciák a formánsmenetekben a folytonosság megszakadását okozhatják [2], észrevehetőek a beszédpercepció számára [7], valamint hasznosak lehetnek a beszélőnormalizálásban [16,17]. Eddig azonban csak néhány nyelvre vizsgálták a magánhangzó-formánsok és SGR-ek kapcsolatát. Wang és kollégái angol-spanyol kétnyelvű gyermekek beszédével foglalkoztak [16]. Lulich egy felnőtt férfi és kilenc gyermek amerikai angol beszélő  $Sg2$  és  $F2$  kapcsolatát elemezte [8]. Madsack és társai az  $Sg1$ - $F1$  és  $Sg2$ - $F2$  közötti összefüggést kutatta két német dialektus néhány beszélőjén [11], Jung pedig hasonlólt végzett a koreai nyelvre [6].

A szubglottális rezonanciák magánhangzó-elkülönítő szerepével kapcsolatban magyar nyelvre eddig kezdeti kutatások történtek csak. Az első kísérletek alapján az  $Sg1$ ,  $Sg2$  és  $Sg3$  szerepet játszhat a beszédhangok produkciójában [3]. Az eredmények szerint az  $Sg1$  az alsó és nem alsó, az  $Sg2$  az elől és hátul képzett magánhangzók közötti határon található, míg az  $Sg3$  az elől képzett ajakréses nem alsókat különíti el a többi elől képzett magánhangzótól. A magyar magánhangzócsoportok között feltételezett elválasztó szerepet a 2. ábra mutatja.

A vízszintes és függőleges vonalak utalnak a szubglottális rezonanciák helyére a formánstérben. [3] eredményei szerint két férfi és két nő logatom-olvasása alapján nagyrészt teljesülnek ezek a hipotézisek.



2. ábra. A magyar magánhangzók elméleti formánstere. A vízszintes és függőleges vonalak a szubglottális rezonanciák által feltételezett elkülönülést mutatják.

A különböző nyelvekre történt kutatásokat viszonylag kevés adaton végezték el, de az eredmények konzisztensek abban, hogy mindegyik vizsgált nyelvben az alsó légúti rezonanciák határként szolgálnak különböző magánhangzó-csoportok között.

## 1.2. Formánsmenetek mássalhangzó-magánhangzó kapcsolatokban

A mássalhangzó-magánhangzó (CV) kapcsolatokban a magánhangzó formánsértékei nem állandóak a két hang közötti koartikuláció miatt [5]. A zöngés és zöngétlen zárhangok képzési helyüktől függően kisebb-nagyobb mértékben módosítják a szomszédos magánhangzó formánsait. A második formáns változása alapján ezen hangkapcsolatokat regressziós egyenesek (ún. locus egyenlet) segítségével jellemezhetjük [9]. A regressziós egyenesekből megrajzolható az ún. locus egyenlet tér, mely a zárhang végén és a mássalhangzó közepén mérhető második formáns értékét veti össze [9, 2. ábra]. Ezen ábrán az egyes beszédhangosztályok az artikulációs helyük szerint elkülönülő csoportokban jelennek meg az  $F2$  változása miatt. Néhány korábbi kísérletben kimutatták, hogy ezen csoportok elkülönüléséhez a szubglottális rezonanciák is hozzájárulnak, hasonlóan a magánhangzókban okozott kategorikus elválasztáshoz [9,10].

Jelen kutatás során tovább vizsgáljuk a mássalhangzó-magánhangzó kapcsolatok helyét a második formáns által meghatározott locus egyenlet térben, valamint az alsó légúti rezonanciák csoportelválasztó szerepét is bemutatjuk. A kísérleteink során egy magyar anyanyelvű beszélő hangfelvételeit és szubglottális felvételeit elemezzük. Bemutatjuk egy automatikus osztályozó eljárás működését, amely az alsó légúti rezonanciák és a második formáns viszonya alapján csoportosítja a mássalhangzó-magánhangzó beszédhangkapcsolatokat. Az eredmények segíthetik a fonológiai megkülönböztető jegyek szerepének megértését,

illetve alkalmazásra kerülhetnek a beszélőnormalizálásban és beszédfelismerésben.

## 2. Módszerek

A kísérleteink során egy magyar anyanyelvű beszélő ("B1", 29 éves, férfi) beszédfelvételeit és szubglottális felvételeit elemeztük. A rögzített hanganyagot elsősorban akusztikai szempontból vizsgáltuk.

### 2.1. Beszédfelvételek

A felvétel során "B1" beszélő "ɔCVbɔ" típusú logatomokat olvasott fel egy csendesszobában. A logatomok első mássalhangzója az összes zöngés és zöngétlen zárhangot tartalmazta (labiálisok: [b,p], alveolárisok: [d,t], velárisok: [g,k] és palatálisok: [j,c]). A középső hangsúlytalan szótagban mind a 14 magyar magánhangzó szerepelt ([ɔ,a,o,or,u,u:,e,e:,i,i:,ø,ø:,y,y:]). A logatomokat a beszélő véletlenszerű sorrendben olvasta fel, mindegyiket tízszer, így összesen 1120 logatomot kiejtve. A beszédhangot EMC 100 kondenzátor mikrofonnal rögzítettük, mely a felvétel során a beszélő ajkaitól kb. 15 cm-re helyezkedett el. Az elhangzott anyagot 48 kHz-es mintavételezéssel digitalizáltuk Terratex DMX 6 Fire USB külső hangkártyával, a Wavesurfer programot használva [13].

### 2.2. Formánsmérések

Az "ɔCVbɔ" logatomok hanghatárait a beszédfelvétel és a felolvasott szöveg alapján automatikus módszerrel határoztuk meg, egy beszédfelismerőt kényszerített üzemmódban használva [12]. A második formánsfrekvenciákat Praat segítségével [1] mértük automatikusan, a zárhang végén ( $F2_{msh}$ , a hangidőtartam 95%-ánál), valamint a második magánhangzó közepén ( $F2_{mgh}$ , a hangidőtartam 50%-ánál). Az automatikus formánsmérés eredményén ezután kézi ellenőrzést végeztünk: manuálisan megvizsgáltuk az átlagostól jelentősen eltérő eseteket, külön-külön az egyes CV csoportokra. A mássalhangzókból és magánhangzókból mért formánsértékek mediánjait az 1. táblázat tartalmazza.

### 2.3. Szubglottális felvételek

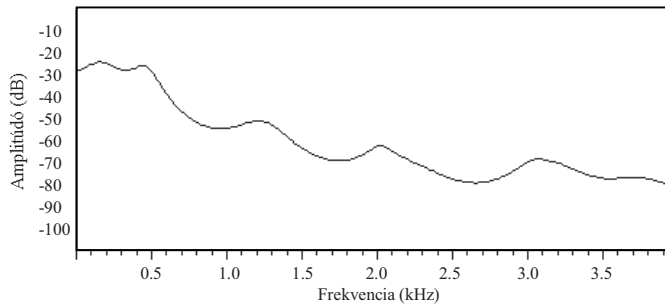
Egy másik felvétel során az alsó légúti rendszer jelét is felvettük csendesszobában, "B1" beszélőtől. Amíg a beszélő felolvasott néhány mondatot, a beszédhangját és alsó légúti jelét rögzítettük. A beszédfelvételeket jelen kísérlet során nem használtuk fel. A szubglottális jelet egy K&K HotSpot gyorsulásmérő eszköz segítségével vettük fel, amely a beszéd során a beszélő nyakára volt szorítva, a pajzsporc fölé. A jelet 8 kHz-es mintavételezéssel, Terratex DMX 6 Fire USB külső hangkártyával digitalizáltuk a Wavesurfer programmal.

1. táblázat. "B1" beszélő beszédfelvételein mért  $F2_{msh}$  és  $F2_{mgh}$  értékek mediánjai (az értékek Hz-ben értendők). Az  $F2_{msh}$  értékeket a zárhangok 95%-ánál, az  $F2_{mgh}$  értékeket a magánhangzók 50%-ánál mértük.

		$F2_{msh}$								$F2_{mgh}$							
		Labialis		Alveoláris		Veláris		Palatális		Labialis		Alveoláris		Veláris		Palatális	
		b	p	d	t	g	k	j	c	b	p	d	t	g	k	j	c
Hátso	o	1045	1435	1074	2022	1066	1001	1560	2058	1056	1251	1114	1295	1095	1037	1197	1322
	o	830	1304	843	1651	878	841	1514	1714	797	978	875	1003	845	786	958	1036
	o:	817	1374	853	1632	782	793	1499	1860	651	691	675	720	661	633	674	703
	u	805	1486	852	1703	807	789	1587	1899	691	878	749	919	712	686	849	976
	u:	825	1435	805	1690	784	798	1526	2035	619	712	640	691	644	552	678	728
	a:	1236	1655	1714	2001	1752	1266	1638	2106	1478	1506	1593	1560	1564	1504	1527	1541
Első	e	1518	1726	2021	2181	2101	1542	1753	2179	1678	1716	1798	1846	1795	1706	1678	1812
	ø	1348	1661	1374	2076	1524	1390	1695	2018	1433	1500	1475	1583	1477	1446	1500	1613
	ø:	1518	1726	1635	2055	1688	1525	1729	2007	1659	1680	1600	1702	1621	1602	1703	1663
	y	1594	1841	1730	2149	1809	1569	1860	2116	1803	1904	1740	1909	1782	1824	1975	1881
	y:	1708	1899	1796	2198	1961	1774	1934	2149	1953	2002	1824	1927	1849	1878	1911	1848
	e:	1769	1894	2112	2242	2299	1997	1880	2264	2278	2302	2288	2306	2300	2308	2287	2296
	i	1939	2022	2242	2225	2292	1956	1947	2244	2209	2281	2300	2255	2258	2240	2274	2190
	i:	2014	2025	2235	2217	2266	2308	1945	2309	2317	2380	2409	2334	2357	2312	2358	2358

## 2.4. Szubglottálisrezonancia-mérés

A szubglottális jelből manuális módon, a Wavesurfer program segítségével mértük az első három szubglottális rezonancia értékét. A 3. ábra egy példa spektrumot mutat "B1" gyorsulásmérő felvételéből, melyen látható, hogy az SGR-mérés a formánsméréshez hasonlóan, a spektrumbeli csúcsok leolvasásával történik. Az SGR meghatározásának módszeréről részletesebb leírás olvasható [2,8]-ben. A hullámformában 20 helyen mértük meg az SGR-értékeket, az összesített adatok a 2. táblázatban találhatóak.



3. ábra. Példa LPC spektrum "B1" beszélő gyorsulásmérő felvételéből. A spektrális csúcsok (454 Hz, 1211 Hz, 2023 Hz és 3067 Hz) a szubglottális rezonanciák értékei. Az ábrán látható  $Sg1$  értéke meglehetősen alacsony a szakirodalmi adatokhoz képest [15].

2. táblázat. "B1" beszélő gyorsulásmérő felvételében mért SGR értékeinek adatai. A továbbiakban a medián értékeket használtuk fel.

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

### 3. Eredmények

A formánsmérések alapján megvizsgáltuk "B1" beszélő locus egyenlet terét, a szubglottális rezonanciák szerepét kiemelve. Ezután egy osztályozó segítségével vizsgáltuk a különböző CV-csoportok elválaszthatóságát.

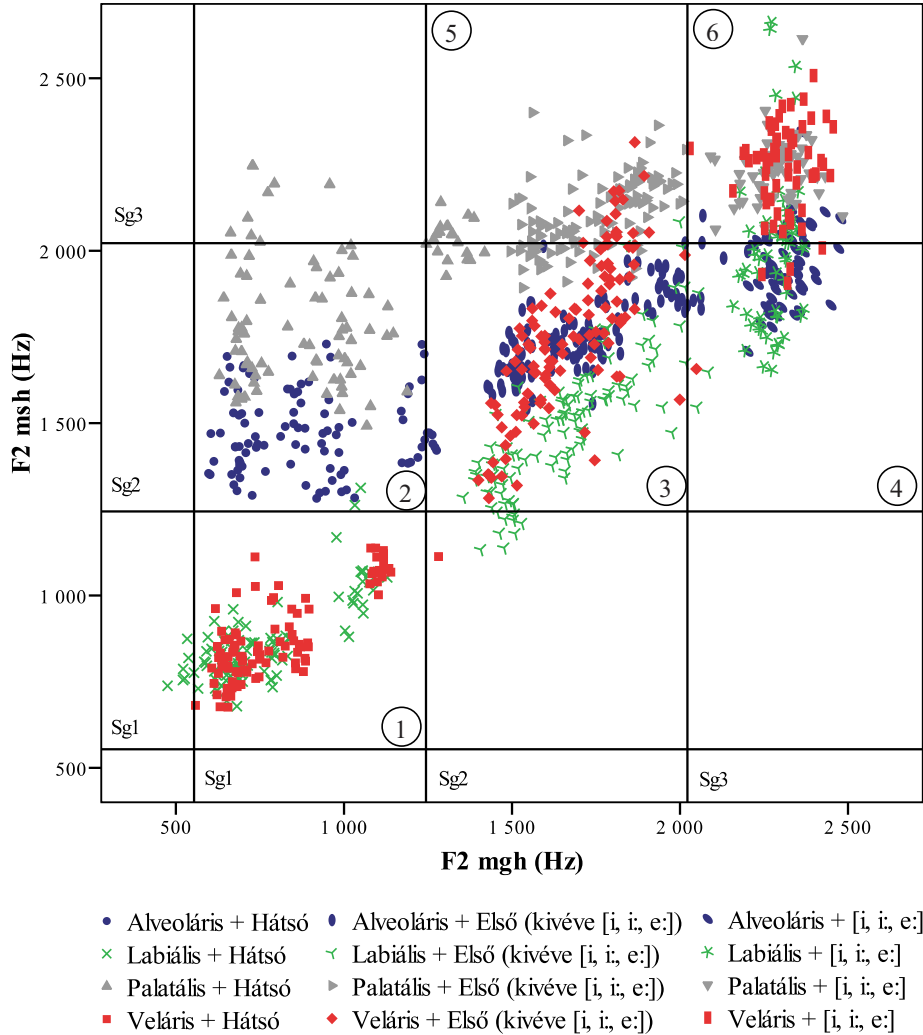
#### 3.1. Locus egyenlet tér

"B1" beszélő  $F2$ - és SGR-adatai alapján elkészítettük a locus egyenlet terét, amely a 4. ábrán látható módon veti össze az  $F2_{msh}$  és  $F2_{mgh}$  értékeket. Amint az ábra mutatja, a locus egyenletek terében a CV-kapcsolatok elkülönülnek, az  $F2_{msh}$ - $F2_{mgh}$  párok a mássalhangzó és a magánhangzó képzési helyének megfelelő csoportokban jelennek meg. Ezen csoportokat a szubglottális rezonanciák határolják: a függőleges  $Sg2$  az elől, illetve hátul képzett magánhangzók közé ékelődik, az  $Sg3$  az elől képzett ajakréses nem alsó magánhangzókat választja el a többi elől képzettől. A vízszintes  $Sg2$  azokat a labiális és veláris mássalhangzókat különíti el, amelyeket hátul képzett magánhangzó követ. A vízszintes  $Sg3$  szerepe kisebb mértékű.

Az ábrán hat tartományt jelöltünk számokkal, mindegyik téglalap egy-egy CV-osztálynak felel meg, melyeket az SGR-ek határolnak:

1. Labiális és veláris mássalhangzók, hátul képzett magánhangzókkal
2. Alveoláris és palatális mássalhangzók, hátul képzett magánhangzókkal
3. Alveoláris, labiális és veláris mássalhangzók, elől képzett magánhangzókkal, kivéve [i, i:, e:]
4. Alveoláris és labiális mássalhangzók, elől képzett ajakréses nem alsó magánhangzókkal ([i, i:, e:])
5. Palatális mássalhangzók, elől képzett magánhangzókkal, kivéve [i, i:, e:]
6. Palatális és veláris mássalhangzók, elől képzett ajakréses nem alsó magánhangzókkal ([i, i:, e:])

Ezek a tartományok részben különböznek az amerikai angol nyelven végzett kísérletben bemutatotthoz képest [10]. Az angol nyelvben a veláris mássalhangzó - elől képzett magánhangzó kapcsolatokban a  $F2_{msh}$  érték nagyobb, mint az  $Sg3$ . A magyar nyelvre végzett kísérletünkben csak az [i, i:, e:] magánhangzók esetén igaz ez, a többi veláris - első kapcsolatban az  $F2_{msh}$  érték kisebb  $Sg3$ -nál. A palatális mássalhangzókat is vizsgáltuk kísérletünkben, amelyek az angol nyelvben nem fordulnak elő.



4. ábra. "B1" beszélő locus egyenlet tere. 1120 adatpont látható, melyek a logatomokban vizsgált CV-kapcsolatok második formánsai alapján kerültek ábrázolásra. A különböző képzési helyű mássalhangzókat és magánhangzókat eltérő színnel és alakkal jelöltük. A CV-kapcsolatok  $F2_{msh}$ - $F2_{mgh}$  párjai a mássalhangzó és magánhangzó képzési helyének megfelelően elkülönülő csoportokban jelennek meg, melyeket 1–6 számokkal jelöltünk. A vízszintes és függőleges vonalak a mért szubglottális rezonanciák helyét jelzik.

A 4. ábra alapján az SGR-ek jól elkülöníthető csoportokra osztják a CV-kapcsolatokat második formánsuk alapján. Ez csak néhány kisebb CV-halmaz esetén nem teljesül. A palatális - hátsó kapcsolatok az  $F2_{msh}$  irányban nagy teret foglalnak el, néhány adatpont esetén az  $F2_{msh}$  érték magasabb az  $Sg3$ -nál. A palatálisok egy jól elkülönülő csoportja található a függőleges  $Sg2$  és a vízszintes  $Sg3$  között (melyekre az  $F2_{mgh}$  érték nagyobb  $Sg2$ -nél). A legtöbb palatális - elől képzett CV-kapcsolat esetén az  $F2_{msh}$  nagyobb  $Sg3$ -nál, míg a "palatális - első (kivéve [i, i:, e:])" csoportnak körülbelül harmada nyúlik  $Sg3$  alá. Azokban a CV-kapcsolatokban, amelyekben a magánhangzó [i, i:, e:] volt, a mássalhangzók  $F2$  értéke 1600–2600 Hz között szóródik, így a 4-es és 6-os tartomány adatpontjai nehezen elkülöníthetők (elsősorban a labiálisok találhatóak meg a tartomány szélső értékeinél is).

### 3.2. Mássalhangzóosztályok locus egyenletei

A különböző artikulációs helyű CV-hangkapcsolatokra jellemző formánsmeneteket lineáris regresszió segítségével vizsgáltuk. Ezen locus egyenletek együtthatói, valamint a korrelációs mérőszámok a 3. táblázatban találhatóak. A lineáris regressziós vizsgálatok eredményeként kiderült, hogy a formánsmenetet leíró egyenlet meredeksége ( $m$ ) és y-metszete ( $b$ ) eltérő a különböző mássalhangzócsoporthoz. Az alveolárisok és palatálisok meredeksége 0,3 körüli, míg a labiálisok és velárisok esetében ez az érték 1-hez közelít. A labiálisok és velárisok  $F2_{msh}$  és  $F2_{mgh}$  értékei között erősebb a korreláció, melyet a 4. ábrán látható lineárishoz közelítő elhelyezkedésük is mutat.

3. táblázat. A különböző artikulációs helyű mássalhangzóosztályok locus egyenleteinek lineáris regressziós együtthatói és Pearson-féle korrelációs mérőszámai.

$$F2_{msh} = m \cdot F2_{mgh} + b$$

	$m$	$b$	$R^2$
Alveoláris	0,333	1184,350	0,768
Labiális	0,732	301,220	0,915
Palatális	0,307	1552,820	0,628
Veláris	0,912	179,195	0,936

### 3.3. CV-kapcsolatok osztályozása

A kísérletek során a [10]-ben bemutatott osztályozó algoritmust használtuk fel, melynek segítségével lehetséges a CV-kapcsolatok automatikus osztályozása, szubglottális rezonanciák alapján. Mivel a magyar nyelv mássalhangzómagánhangzó kapcsolatai részben különböznek az angolétól, az algoritmuson kisebb változtatásokat végeztünk, így például a palatális mássalhangzókat is vizsgáltuk. Az osztályozás során a 4. ábrán látható 1–6 tartományokat vettük



figyelembe. Ezen régiók határait a formánsok ( $F2_{msh}$  és  $F2_{mgh}$ ) és szubglottális rezonanciák ( $Sg2$  és  $Sg3$ ) közötti egyenlőtlenségek segítségével írhatjuk le, melyek a 4. táblázatban találhatóak.

4. táblázat. A CV-osztályok határait megadó egyenlőtlenségek.

Tartomány	CV-osztály	1. egyenlőtlenség	2. egyenlőtlenség
1	Labiális, Veláris + Hátsó	$F2_{msh} < Sg2$	$F2_{mgh} < Sg2$
2	Alveoláris, Palatális + Hátsó	$Sg2 < F2_{msh} < Sg3$	$F2_{mgh} < Sg2$
3	Alveoláris, Labiális, Veláris + Első	$F2_{msh} < Sg3$	$Sg2 < F2_{mgh} < Sg3$
4	Alveoláris, Labiális + [i, i:, e:]	$F2_{msh} < Sg3$	$Sg3 < F2_{mgh}$
5	Palatális + Első	$Sg3 < F2_{msh}$	$Sg2 < F2_{mgh} < Sg3$
6	Palatális, Veláris + [i, i:, e:]	$Sg3 < F2_{msh}$	$Sg3 < F2_{mgh}$

Célunk az 1–6 tartományok optimális klasszifikációjának megtalálása volt. Ennek érdekében az  $\widetilde{Sg2}$  értékét 1000 és 1500 Hz között folyamatosan növelve vizsgáltuk az osztályozás találati és téves riasztási arányait. Az  $\widetilde{Sg3}$  értékét "B1" beszélő  $Sg2/Sg3$  aránya alapján számítottuk ( $\widetilde{Sg3} = 1,6254 \cdot \widetilde{Sg2}$ ).

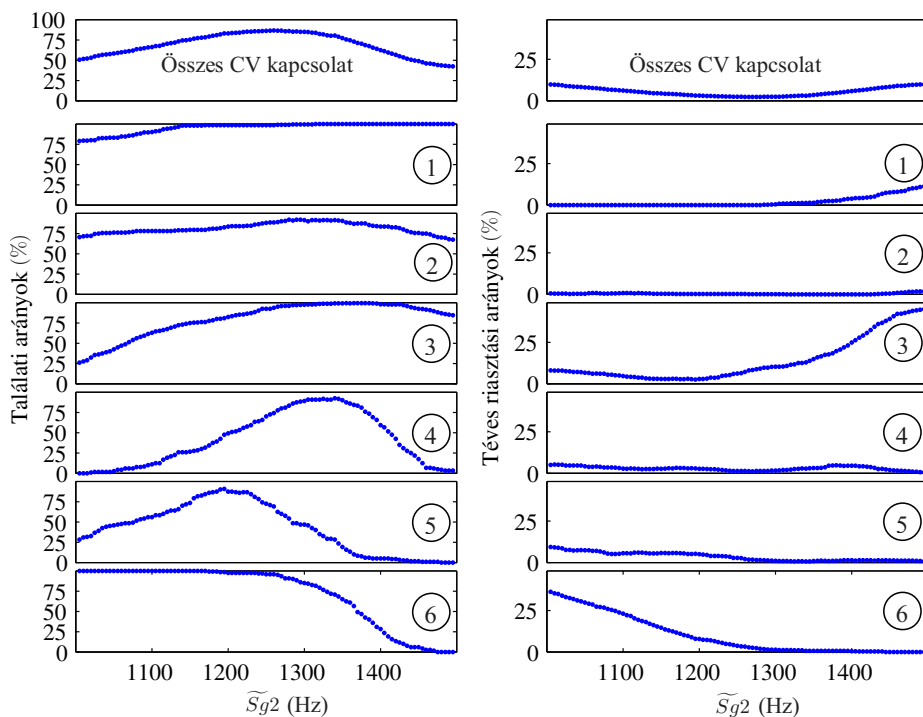
Az 5. ábra mutatja az osztályozás találati és téves riasztási arányait az  $\widetilde{Sg2}$  függvényében. Az összes CV-kapcsolatra vonatkozó optimális klasszifikáció 1260 Hz esetén történik. Emellett eltérő a legmagasabb találati arány a 4-es tartomány esetén. A 4. ábrát megvizsgálva azt láthatjuk, hogy a 4-es és 6-os tartományban lévő CV-kapcsolatok között sok az átfedés, emiatt fordulhat elő, hogy a 4-es tartomány esetén az optimális szeparáció magasabb  $\widetilde{Sg2}$  esetén (kb. 1350 Hz) megy végbe.

Az optimális osztályozás ( $\widetilde{Sg2}=1260$  Hz) esetén az összesített eredményeket az 5. táblázat mutatja. Az összes CV-kapcsolatra a találati arány 86,6%, míg a téves riasztási arány 2,3%. A 4-es tartomány esetén a legalacsonyabb a találati arány, a korábban leírtak miatt.

Ezután megvizsgáltuk az osztályozást "B1" beszélő szubglottális jelében manuálisan mért SGR-értékek mediánjai alapján ( $Sg2=1244$  Hz,  $Sg3=2022$  Hz). A találati és téves riasztási arányokat a 6. táblázat mutatja külön-külön az egyes kategóriákra, illetve összesítve is. Az összes vizsgált CV-kapcsolatra a találati arány 85,5%, míg a téves riasztási arány 2,4%. Ezek az értékek nagyon közel vannak az optimális elválasztáshoz, mivel a mért  $Sg2$  értéke (1244 Hz) szinte megegyezik az optimális osztályozás során kapott  $\widetilde{Sg2}$ -vel (1260 Hz).

#### 4. Következtetések

Jelen kutatás során egy kísérletsorozatot mutattunk be, amely egy beszélő logatomfelvételeiből származó CV-kapcsolatok által definiált locus egyenlet teret elemzett, illetve vizsgálta a szubglottális rezonanciák által okozott elválasztást. Először megvizsgáltuk "B1" beszélő locus egyenlet terét, majd a CV-csoportok



5. ábra. Az osztályozás eredménye az  $\widetilde{Sg2}$  függvényében. A legfelső részábrák mutatják az összesített találási és téves riasztási arányokat, az alsóbb ábrák pedig az 1-6 tartományokhoz tartozó eredményeket.

5. táblázat. CV-kapcsolatok osztályozásának találási és téves riasztási arányai, az optimális  $\widetilde{Sg2}$  értékkel számolva. ( $\widetilde{Sg2} = 1260$  Hz, CV jelöli az összes eredményt, 1-6 az egyes tartományokat.)

	CV	1	2	3	4	5	6
Találási arány	86,6%	98,5%	88,5%	93,6%	74,2%	69,2%	95,8%
Téves riasztási arány	2,3%	0%	0,2%	7,1%	1,3%	2,1%	3,2%

6. táblázat. CV-kapcsolatok osztályozásának találási és téves riasztási arányai, a mért SGR-értékekkel számolva. ( $Sg2=1244$  Hz)

	CV	1	2	3	4	5	6
Találási arány	85,5%	98,5%	87%	91,7%	65%	74,2%	96,7%
Téves riasztási arány	2,4%	0%	0,2%	5,7%	1,3%	2,9%	4,3%

artikulációs helye alapján hat tartományt definiáltunk, melyek hipotéziseink szerint az SGR-ek segítségével elkülöníthetőek. A különböző mássalhangzó osztályok lineáris regressziós egyenleteit (ún. locus egyenlet) is vizsgáltuk. A [10]-ben bemutatott osztályozó algoritmust a magyar nyelv hangjainak megfelelően módosítottuk, és alkalmaztuk a CV-adathalmazra. A mért SGR-értékek alapján történő osztályozást összehasonlítottuk az optimális találati arányt és téves riasztási arányt okozó klasszifikációval. A szubglottális rezonancia alapú automatikus mássalhangzó-magánhangzó hangkapcsolat osztályozás "B1" beszélő esetén az optimálishoz képest mindössze 1%-kal alacsonyabb a találati arányt eredményezett. Ez a [10]-ben bemutatott amerikai angol kísérlethez hasonló eredményt jelent.

A további kutatás célja más SGR mérési lehetőségek keresése. [10] szerint a beszédfelvételtől is meghatározható az  $Sg_2$  értéke, de ez viszonylag pontatlan, távol van a gyorsulásmérővel mért SGR-értékektől. [18]-ban egy egyedi eszközt készítettek a szubglottális jel felvételére, és az alsó légúti rezonanciák automatikus mérésével kísérleteztek, azonban ez távol volt a manuálisan mért értékektől. Amennyiben a szubglottális rezonanciák mérése egyszerűbben megoldható lesz, az itt bemutatott SGR alapú CV-osztályozás tetszőleges beszélő esetén alkalmazhatóvá válik.

Az itt bemutatott kísérlet során csak egy magyar nyelvű beszélő felvételeit vizsgáltuk. A továbbiakban érdemes lenne több beszélő hangfelvételeit is elemezni, megvizsgálva mások locus egyenlet terének és szubglottális rezonanciáinak kapcsolatát.

A szubglottális rezonanciákat már sikerrel alkalmazták a beszélőnormalizálásban [16,17], az eredményeink ezen kívül hozzájárulhatnak a beszédfelismeréshez is.

## 5. Köszönetnyilvánítás

A szerzők köszönetüket fejezik ki a támogatóknak (NKFP 2/034/2004, Jedlik OM-00102/2007, TÁMOP-4.2.2-08/1/KMR-2008-0007), a kísérletben részt vevő adatközlőnek, valamint Bóhm Tamásnak a hangfelvételek rendelkezésre bocsátásáért. Külön köszönet illeti Steven M. Lulichot a szubglottális rezonanciák témájának részletes ismertetéséért, valamint a cikk javítására irányuló javaslataiért és megjegyzéseieiért.

## Hivatkozások

1. Boersma, P., Weenink, D.: Praat (Version 5.1.19). <http://www.praat.org> (2009)
2. Chi, X., Sonderegger, M.: Subglottal coupling and its influence on vowel formants. *JASA* 122 (2007) 1735–1745
3. Csapó, T. G., Bárkányi, Zs., Grácsi, T. E., Bóhm, T., Lulich, S. M.: Relation of formants and subglottal resonances in Hungarian vowels. In: *Proc. Interspeech* (2009) 484–487
4. Gray, H.: *Anatomy of the human body*. Philadelphia: Lea & Febiger. (1918)

5. Gósy, M.: *Fonetika, a beszéd tudománya*. Osiris Kiadó, Budapest. (2004)
6. Jung, Y.: Subglottal effects on the vowels across language: Preliminary study on Korean. *JASA* 125 (2009) 2638
7. Lulich, S. M., Bachrach, A., Malyska, N.: A role for the second subglottal resonance in lexical access. *JASA* 122 (2007) 2320–2327
8. Lulich, S. M.: Subglottal resonances and distinctive features. *J. Phon.* doi:10.1016/j.wocn.2008.10.006 (2009)
9. Lulich, S. M.: On the relation between locus equations and subglottal resonances. *POMA* 5, 060003 (2009)
10. Lulich, S. M., Chen, N. F.: Automatic classification of consonant-vowel transitions based on subglottal resonances and the second formant, *POMA* 6, 060005, (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. LabPhon11* (2008) 91–92
12. Mihajlik, P., Révész, T., Tatai, P.: Phonetic Transcription in Automatic Speech Recognition. *Acta Linguistica Hungarica*, Vol. 49. (3-4), (2002) 407–425
13. Sjölander, K., Beskow, J.: *Wavesurfer* (Version 1.8.5). <http://www.speech.kth.se/wavesurfer> (2009)
14. Stevens, K. N.: On the quantal nature of speech, *J. Phon.* 17 (1989) 3–45
15. Stevens, K. N.: *Acoustic Phonetics*. MIT Press: Cambridge, MA. (1998)
16. 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* (2008) 1717–1720
17. Wang, S., Lee, Y.-H., Alwan, A.: Bark-shift based nonlinear speaker normalization using the second subglottal resonance. In: *Proc. Interspeech* (2009) 1619–1622
18. Wokurek, W., Madsack, A.: Comparison of Manual and Automated Estimates of Subglottal Resonances. In: *Proc. Interspeech* (2009) 1671–1674