

# **BESZÉDKUTATÁS 2008**

**SOKSZÍNŰ BESZÉDTUDOMÁNY**

**BUDAPEST**







# **BESZÉDKUTATÁS 2008**

## **SOKSZÍNŰ BESZÉDTUDOMÁNY**

**Szerkesztette:  
GÓSY MÁRIA**

MTA Nyelvtudományi Intézet  
Kempelen Farkas Beszédkutató Laboratórium  
Budapest 2008

A szerkesztőbizottság elnöke: Gósy Mária

A szerkesztőbizottság tagjai: Hoffman Ildikó

Markó Alexandra

Olaszy Gábor

Siptár Péter

Vago, Robert

A korpusz szerkesztésében részt vettek: Bóna Judit

Gósy Mária

Grácsi Tekla Etelka

Gyarmathy Dorottya

Imre Angéla

Markó Alexandra

Az egyes tanulmányokat szakterületi kompetenciával rendelkező szakemberek lektorálták.

Technikai szerkesztő: Markó Alexandra

A kötet kiadását az MTA Könyv- és Folyóíratkiadó Bizottsága támogatta.

ISSN 1218-8727

© MTA Nyelvtudományi Intézet

Felelős kiadó: Kenesei István igazgató

Készült: Regiszter Kiadó és Nyomda Kft.

## TARTALOM

|                                                                                                                                   |     |
|-----------------------------------------------------------------------------------------------------------------------------------|-----|
| Gósy Mária: A zaj hatása a beszédre.....                                                                                          | 5   |
| Horváth Viktória: Nazális hatás a magánhangzóképzés függvényében .....                                                            | 22  |
| Grácz Tekla Etelka: Alveoláris spiránsok akusztikai fonetikai vizsgálata .....                                                    | 33  |
| Mády Katalin: Magyar magánhangzók vizsgálata elektromágneses<br>artikulográffal normál és gyors beszédben.....                    | 52  |
| Gyarmathy Dorottya: Fantomhangok a magyar spontán beszédben .....                                                                 | 67  |
| Markó Alexandra: A magyar eldöntendő kérdő mondat megvalósulásai<br>különböző beszédmódokban .....                                | 77  |
| Beke András: A felolvasás és a spontán beszéd alaphangszerkezeteinek<br>vizsgálata .....                                          | 93  |
| Böhm Tamás – Ujváry István: Az irreguláris fonáció mint egyéni<br>hangjellemző a magyar beszédben .....                           | 108 |
| Imre Angéla: A hangsúly megjelenésének percpéció és produkció<br>vizsgálata .....                                                 | 121 |
| Várad Viola: A virtuális mondatok műfaji meghatározottsága.....                                                                   | 134 |
| Bóna Judit: A nyújtás sajátosságai a dadogó és az ép beszédben .....                                                              | 148 |
| Horga, Damir: Repetitions in interrupted speech production.....                                                                   | 157 |
| Nikléczy Péter – Gósy Mária: A személyazonosítás lehetősége a beszéd<br>anyag időtartamának függvényében.....                     | 172 |
| Tóth Bálint – Németh Géza: Rejtett Markov-modell alkalmazása magyar<br>nyelvű gépi szövegfelolvasáshoz .....                      | 182 |
| Gósy Mária: Magyar spontánbeszéd-adatbázis – BEA.....                                                                             | 194 |
| Olasz Gábor – Bartalis Mátyás: Jelfeldolgozási és fonetikai<br>algoritmusok kombinációja a gépi hanghatárjelölés javítására ..... | 208 |
| „Nyelvbottlás”-korpusz.....                                                                                                       | 221 |
| 2. Morfológiai, szintaktikai hiba .....                                                                                           | 222 |
| 3. Kontamináció .....                                                                                                             | 223 |
| 4. Téves szótalálás.....                                                                                                          | 224 |
| 5. „Nyelvem hegyén van” .....                                                                                                     | 230 |
| 6. Változtatás.....                                                                                                               | 231 |
| 7. Újraindítás.....                                                                                                               | 231 |

|                                                                                                                                                                             |     |
|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----|
| 8. Perszeveráció .....                                                                                                                                                      | 232 |
| 9. Anticipáció .....                                                                                                                                                        | 234 |
| 10. Metatézis .....                                                                                                                                                         | 235 |
| 11. Egyszerű nyelvbontás.....                                                                                                                                               | 237 |
| 12. Többféle osztályozás.....                                                                                                                                               | 237 |
| A pedagógusjelöltek Kazinczy-versenye .....                                                                                                                                 | 240 |
| A pedagógusjelöltek Kazinczy Ferencről elnevezett szép magyar<br>beszéd versenyének 36. országos döntője • <i>Balázs Géza</i> .....                                         | 240 |
| A Kazinczy Ferencről elnevezett szép magyar beszéd verseny 2008-as<br>döntőjének tanulságai: az írott szöveg meghangosítása és a gyakori<br>hibák • <i>Bóna Judit</i> ..... | 242 |
| A magyar és a nemzetközi tudományos élet eseményei.....                                                                                                                     | 245 |
| Nyelv- és beszédtechnológiai platform • <i>Olaszy Gábor</i> .....                                                                                                           | 245 |
| Koutny Ilona: Natural Language Processing for Hungarian Speech<br>Synthesis (könyvismertetés) • <i>Olaszy Gábor</i> .....                                                   | 246 |
| Beszéd kutatás 2009 (konferenciafelhívás).....                                                                                                                              | 249 |
| A kötet szerzői .....                                                                                                                                                       | 251 |

## A ZAJ HATÁSA A BESZÉDRE

Gósy Mária

### Bevezetés

Az emberek közötti verbális kommunikáció szinte mindig valamilyen zajos környezetben történik. A zajok forrása, erőssége, típusa, akusztikai jellemzői nagyon különbözők; a természeti jelenségek, a közlekedés, maguk az épületek, a gépek, szórakoztató hangkeltő eszközök és még sok más hangfőléség kíséri mindennapi életünket (Lazarus 1990). Az emberek közötti beszélgetés egy része mások beszédének egyidejű jelenlétében zajlik. A környezeti hangok jellemzőitől és erősségétől függően reagálunk beszélgetés közben az adott akusztikai háttérre; felfigyelünk rá, kellemetlenül érint, kifejezetten zavar, avagy alig vesszük észre, és így tovább. A zajos környezet egyértelműen befolyásolja az ember verbális kommunikációját, annak mindkét oldalát, a beszédprodukciónak és a beszédmegértésnek is. Számos tanulmány foglalkozik a beszéd érthetőségével zajos környezetben. Megállapításaik szerint a felnőttek jobban képesek észlelni és megfelelően feldolgozni a zajban elhangzó beszédet, mint a gyermekek (Hygge 2003; Kloefer et al. 2006). A gyermekek beszédmegértését már az alacsony intenzitású háttérzaj is negatívan befolyásolja. A zaj mérhetőségére a múlt század ötvenes éveiben tettek kísérletet, 1959-ben Kryter határozta meg azt a skálát, amelyet később a nemzetközi használat számára adaptáltak (Kryter eredeti mérései a repülőgép zajhatására irányultak). Az ún. észlelt zajszint (Perceived Noise Level) a frekvencia és az intenzitás függvényében adható meg, a zajos környezet mérési egysége az 1 noy (ez 900 Hz és 1100 Hz között 40 dB-es intenzitásértéket jelent).

A zajhatás azonban nemcsak a beszéd érthetőségét érinti, hanem a beszélő személy beszédprodukciónak is. A zajnak a beszélésre gyakorolt hatását Lombard-hatásnak nevezik. Etienne Lombard (1868–1920) francia orvos, fül-orr-gégész, audiológus volt az első, aki egy 1911-ben megjelent tanulmányában leírta azt a jelenséget, hogy a beszéd intenzitása szignifikánsan megnövekszik a háttérzaj hatására. Az intenzitás növekedését eredményező artikulációs módok feltehetően automatikusak, azaz a beszélőnek nincs konkrét tudomása arról, hogy hangosabban beszél. A beszéd hangerejének növekedése tehát a környezeti zaj következtében fellépő ösztönös változtatás. A Lombard-hatást a hallásnak a beszédmegértésben betöltött funkciójával kapcsolatosan tárgyalja Lane és Tranel tanulmánya (1971). További vizsgálatok foglalkoztak annak a leírásával, hogy milyen módon hat a zaj a beszéd artikulációs gesztusainak megváltozására, valamint a beszéd akusztikai fonetikai jellemzőire



(Junqua 1996; Castellanos et al. 1996; Winkworth–Davis 1997; Wassink et al. 2007). Az utóbbi évtizedben azt is megállapították, hogy a környezeti zaj nem csupán az ember verbális kommunikációjára, de egyes állatok kommunikációjára is hatással van. A majmoknál, a fülémülnél, a macskáknál, a zebráknál és a delfineknél is igazolták a Lombard-hatást (vö. Rabin et al. 2003; Brumm 2004; Brumm–Slater 2006). A tamarinoknál (majomféle) nem csupán intenzitásváltozást, hanem a hangadási tempó ritmikus módosulását is tapasztalták (Egnor–Hauser 2006).

A Lombard-hatás egyértelműen magyarázható a belső hallási folyamatainkkal. E területen a Nobel-díjas Békésy György végzett kiemelkedő kutatásokat (1949). Tudjuk, hogy a saját beszédünket a légvezetési és a csontvezetési hallással egyaránt halljuk, és a kettőnek az intenzitása nagyjából azonos. A szájüregben a létrehozott beszéd intenzitása 40–45 dB-lel erősebb, mint amit a belsőfül érzékel. Ez magyarázza azt, hogy miközben a beszédünket a hallgató jól „hallja”, észleli, a beszélőnek mégsem okoz kellemetlen hangerejű hangzást. Békésy szerint ugyanis a középfül úgy működik, hogy a gégeben keletkezett hangot (ez a zönge), valamint a szájüregben módosult hangot (beszédhangok sorozata) nem olyan hatásosan továbbítja, mint a külvilágból érkező hangokat.

Amikor a beszélő érzékeli a környezeti zaj növekedését, ösztönösen parancsot küld az artikulációs szerveknek, hogy úgy működjenek, hogy nagyobb legyen a hangerő. A mindenkori cél a beszéd jobb érthetősége, egyfelől a beszélő számára, másfelől a beszélgető partner számára. Más megfogalmazásban a Lombard-hatás a grice-i maximák érvényesítésének egy speciális formája. Az állatoknál tapasztalt hasonló kommunikációs viselkedést a magasabb kognitív tevékenységükkel hozzák kapcsolatba.

A zajnak a beszédre gyakorolt hatása az automatikus beszédfelismerés szempontjából fontos gyakorlati probléma (Hansen 1996; Skowronski–Harris 2006). A Lombard-hatást mint speciális technikát alkalmazták abból a célból, hogy tudatosan megnöveljék a beszéd intenzitását az özofáguszbeszéd esetén, valamint azoknál, akik mesterséges szájpadlással beszélnek (Zeine–Brandt 1988).

A folyamatos környezeti zaj befolyásolja az analóg spontán beszéd folyamatait, és ez láncreakciót indít el, hiszen a zajhatás következményeként megváltozott beszéd módosítja a beszédfeldolgozási stratégiákat is. Korábbi vizsgálatunkban a különböző zajok hatását vizsgáltuk a beszédre (Balázs–Gósy 1988). Eredményeink igazolták a beszélők hangerejének növekedését, valamint a zöngképzés nemkívánatos változását, ami abból adódott, hogy a beszélők hol összepréselték a hangszalagjaikat, hol túlzottan levegősen, illetve zörejes hangképzéssel beszéltek. A kísérletben résztvevőknek hangosan kellett olvasniuk, és fülhallgatón keresztül hatféle zajt közvetítettünk háromféle intenzitásértéken (65, 80 és 95 dB-en). A hangos olvasást legnagyobb mértékben a metróállomás zaja, legkevésbé a zene és az emberi beszélgetés zavarta.

Kérdés, hogy vajon a spontán beszéd tervezési és kivitelezési folyamatait mennyire befolyásolja a környezeti zaj, ezek közül is a beszédzaj. A spontán beszéd esetében mások a tervezési folyamatok, mint olvasáskor (Gósy 2005). A leírtak meghangosítása szükségtelenné teszi a grammatikai, fonológiai (stb.) átalakítási folyamatokat, ekkor a betűsorok meghangosítása zajlik. Spontán beszéd esetén olyan megnyilatkozásokat formálunk, amelyeket korábban ugyanúgy talán még sohasem mondtunk. Ez utóbbi tehát jóval nagyobb összpontosítást, figyelmet igényel, hiszen az artikulációs műveleteket a rejtett tervezési műveletekhez kell igazítanunk. Minden zavaró tényező többszörös problémát jelenthet, illetve nehezíti a spontán korrekciós folyamatok működését (Postma–Kolk 1993, Gósy 2008). Vizsgálatot igényel az a kérdés, hogy milyen mértékben hat a zajos környezet ezekre a beszédtervezési folyamatokra.

Mindennapi tapasztalatunk az, hogy zajban nehezebb az elhangzó beszédet megérteni. Feltételezhető, hogy nemegyszer a percepciós stratégia módosítására is szükség van ahhoz, hogy zajos környezetben megfelelően észleljük és megértsük a rövidebb-hosszabb közléseket. A beszédfeldolgozás sikere számos tényezőtől, így a mindenkori zaj intenzitásától, a frekvenciaeloszlástól, a zajtípustól, az elhangzó beszéd akusztikai jellemzőitől, valamint a hallgató percepciós és pszichés sajátosságaitól, életkorától (stb.) is függ. A zajban elhangzó beszéd vizsgálata több tudomány érdeklődési területére esik. E témával foglalkozik a fonetika (pl. az invariáns jegyek vizsgálatakor), a pszicholingvisztika (beszédszlelési, megértési folyamatok megismerése), a pszichológia (pl. a zaj hatása a kognitív működésekre), valamint az eredmények közvetlen gyakorlati felhasználásában érdekelt diszciplínák, mint a pedagógia (a zaj hatása a tanulásra és a tanításra), az audiológia (nagyothallók beszédhallása zajos környezetben), a környezetvédelem (a zajszenyezésnek az emberre, illetve a kognitív teljesítményre gyakorolt hatása), a gerontológia (a zajhatás vizsgálata az idősek beszédfeldolgozására), valamint a beszédtechnológia (a beszédfelismerés problémái zajos környezetben, vö. Castellanos et al. 1996). Az eddigi eredmények igazolták, hogy *a*) a vizuális csatorna segít a zajos szövegek megértésében (pl. Rosenblum et al. 1996); *b*) az osztálytermi beszédértést, illetőleg a szövegek értelmezését nagymértékben nehezíti a környezeti zaj (pl. Hativa 2000); valamint *c*) hogy az időskori beszédértési zavarok oka feltehetően nem a kognitív folyamatok gyengülése, hanem a hallási folyamatok nem kielégítő működése (pl. Schneider et al. 2005). Újabb kutatások kimutatták, hogy a zajhatás és a beszédsszöveg szemantikai előrejelzése között kapcsolat van a neurális hálózatban, amely a zajban elhangzó szöveg feldolgozásakor aktiválódik (fMRI vizsgálatok: Ghazanfar et al. 2003). Kutatási kérdésünk e tekintetben az volt, hogy vajon a beszédzaj miként befolyásolja a szövegértést 16–18 éves korban. Valószínűsítettük, hogy ekkorra a beszédkommunikációs gyakorlottság olyan szintű, hogy a nem túlzottan intenzív zajjal elfedett szöveg megértése nem okoz nehézséget.



Kutatásunkban a fentiekben leírt problémákra igyekeztünk választ kapni a spontán beszéd és a beszédmegértés kísérletes vizsgálatával. A környezeti zaj valamennyi kísérletünkben társalgás, azaz beszédzaj volt. Fő kérdéseink a következők voltak. Az intenzitás az egyetlen tényező-e, amely megváltozik a zaj hatására, avagy más paraméterekben is tapasztalható változás? Kimondható-e, hogy a zaj az artikulációs gesztussorozat többszörös módosulását idézi elő? Megállapítható-e, hogy a zajhatás a beszédtervezés rejtett működéseire is hatást gyakorol? Igazolható-e, hogy középiskolás korban a relatíve kis intenzitású beszédzaj nem okoz feldolgozási gondot? Hipotéziseink szerint az intenzitás változása csak az egyik, bár kétségtelenül a legszembetűnőbb következménye a környezeti zajhatásnak. Feltételeztük, hogy más beszédjellemzők is módosul(hat)nak zajos körülmények között, valószínűsíthetően a beszéddallam, a tempóviszonyok, valamint a beszédtervezési folyamatok.

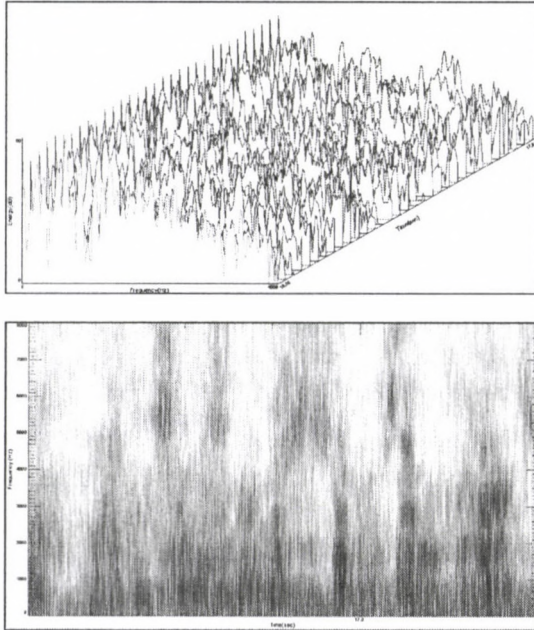
### **Kísérleti személyek, anyag, módszer**

#### **1. Beszédproduktions kísérlet**

Kétperces spontán beszédet rögzítettünk kétféle környezetben: „zajos” és „csendes” helyzetben. A kísérleti személyeket arra kértük, hogy a munkájukról, illetve a szabadidős tevékenységükről beszéljenek folyamatosan. Az instrukció szerint, bármilyen történést is, a téma kifejtését folytatniuk kellett (de nem közöltük velük, hogy zajos környezetet fogunk teremteni). Egy perc elteltével beszédzajt közvetítettünk szabad hangtérben, ezzel a mindennapokban gyakran előforduló kommunikációs helyzetet szimuláltunk. A beszédzaj természetesnek tekinthető, hiszen a mindennapokban gyakori; nem meglepő, nem ismeretlen zaj, előfordul, hogy váratlanul ilyen helyzetbe kerül a beszélő, tehát a szituáció nem idegenszerű. A beszédzaj spektrális eloszlása pedig alkalmas a saját beszéd kvázi-elfedésére (1. ábra). A háttérzaj kifejezetten e kísérlet céljára rögzített öt résztvevő társalgás volt. Minden kísérleti személy pontosan ugyanazt a háttérzajt hallotta. A társalgás eredetileg 10 percig tartott, ebből választottunk ki azt az egyperces időtartamú részt, amelyben a résztvevők egyidejűleg beszélnek. További egy percben a kísérleti személyeknek ebben a háttérzajban kellett beszélniük. Az átlagos jel/zaj arány 15 dB volt (a beszéd átlagos intenzitása 15 dB-lel volt magasabb, mint a háttérzajé).

A kísérleti személyek spontán beszédét mind a „zajos”, mind a „csendes” helyzetben úgy rögzítettük, hogy az anyag további akusztikai fonetikai elemzésekre alkalmas legyen (a mikrofon távolságának és árnyékolásának kialakításával). Ezzel biztosítottuk, hogy a hangfelvétel a háttérzajt nem tartalmazta. A beszédet közvetlenül számítógépre rögzítettük 44 kHz-es mintavételezéssel, majd a Praat szoftverrel (4.04-es verzió) elemeztük. Vizsgáltuk az intenzitást, az alaphangmagasság és a frekvenciartomány sajátosságait, változásait, a tempó alakulását és a megakadásjelenségek típusait és előfordulásukat. A szünetek elemzésekor mind a kitöltött, mind a kitöltetlen jelkimaradásokat (függetlenül attól, hogy ez utóbbiak légzési szünetek voltak-e vagy sem)

számításba vettük. Az előfordulásuk vizsgálatakor a szubjektíven észlelhető jelkimaradásokat adatoltuk, míg az időtartamuk mérésekor csak a 100 ms-os vagy annál nagyobb jelkimaradásokat vettük tekintetbe. A statisztikai elemzéseket az SPSS program 8.0-s verziójával végeztük, 95%-os szignifikanciaszinten (párosított  $t$ -próba, egytényezős ANOVA).



1. ábra

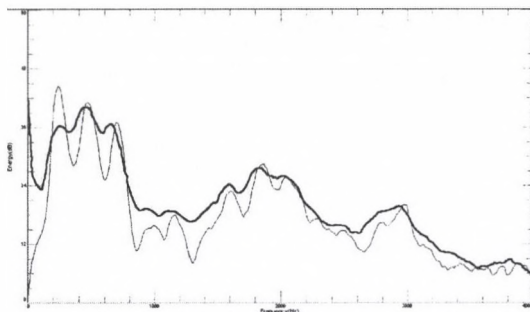
A háttérzajként használt társalgás egy részletének energiaspektruma (felső kép) és hangszíneke (alsó kép)

16 adatközlő, 8 nő és 8 férfi vett részt a kísérletben (20 és 30 év közöttiek). Ép hallásúak, egyiküknek sem volt beszédhibája, valamennyien érettségizettek, hatan felsőfokú végzettséggel rendelkeznek.

## 2. Beszédmegértési kísérlet

Tudományos ismeretterjesztő szöveget (az idő mérésének történetéről) női bemozdóval, átlagos beszédtempóban csendes helyiségben rögzítettünk közvetlenül számítógépre (44 kHz-es mintavételezéssel). A felvétel tartama 3,4 perc. Ezt a szöveget használtuk a kísérletek egyik részében, a másik részében ugyanezt a szöveget zajjal elfedve. Az előzőekben leírt társalgási zajt használtuk az elfedéshez; a jel/zaj viszony 15 dB volt (a háttérzaj tehát csekély mértékben fedte el a beszéd jellemző akusztikai kulcsait), vö. 2. ábra. A elfe-

dő beszédzaj hatására az átlagos energiagörbén a zaj nélküli állapothoz képest sokkal kiegyenlítettebbek a maximumpontok, aminek következtében az eredeti beszédhangok formáns helyei kevésbé dominálnak.



2. ábra

A zajos (vastag vonal) és a nem zajos (vékony vonal) tesztanyag 400 ms tartamú azonos beszédrészletének átlagolt energiaspektruma (LTA)

A kétféle hanganyagot (zaj nélküli és zajjal fedett felolvasást) 16 és 18 éves, ugyanazon budapesti gimnáziumba járó, összesen 88 diákkal hallgattattuk meg; majd kértük őket, hogy írásban 10 megértést ellenőrző kérdésre válaszoljanak. Négy osztály vett részt a kísérletben (osztályonként 22 fő), mindkét évfolyamban az egyik osztály a zaj nélküli, a másik pedig a zajos szöveget hallgatta meg. A helyes válaszok adatait statisztikailag elemeztük (mint a fenti kísérletben).

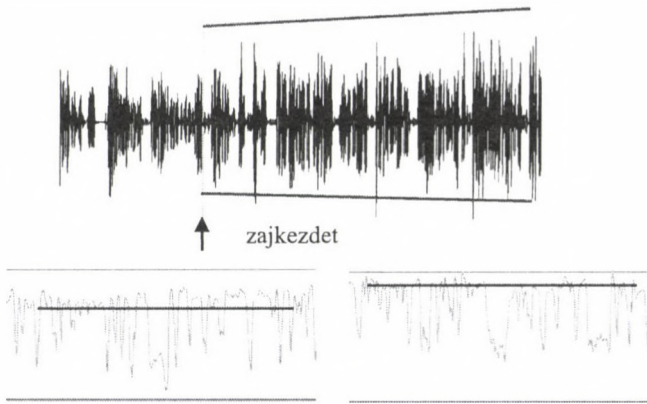
## Eredmények

### 1. Beszédproduktions kísérlet

A szakirodalomban újra és újra leírt, igazolt intenzitásváltozást, tehát a Lombard-hatást minden beszélőnél tapasztaltuk. Az egyes személyek között azonban jellegzetesek a különbségek, és eltérők az aktuális értékek (3. ábra). Az ábra regisztrátumai (rezgés kép és intenzitás görbék) szemléltetik az adott beszédrészlet intenzitásának alakulását csendben és zajban.

Az elemzett spontán beszéd átlagos intenzitása csendben 66,06 dB volt (az átlagos eltérés: 1,387 dB), míg zajban 78,46 dB (az átlagos eltérés 2,948 dB). A beszédzaj hatására tehát a beszélők mintegy 12 dB-lel növelték meg a beszéd átlagos hangerejét. A legkisebb intenzitásváltozást egy férfi beszélőnél tapasztalunk, mindössze 5 dB-et átlagosan, a legnagyobbat egy női beszélőnél, ez átlagosan 15 dB-nek adódott. A beszédintenzitás különbsége a csendes és zajos helyzetben szignifikánsan különbözött (a párosított  $t$ -próba adatai:  $t(15) = -9,946, p < 0,001$ ).





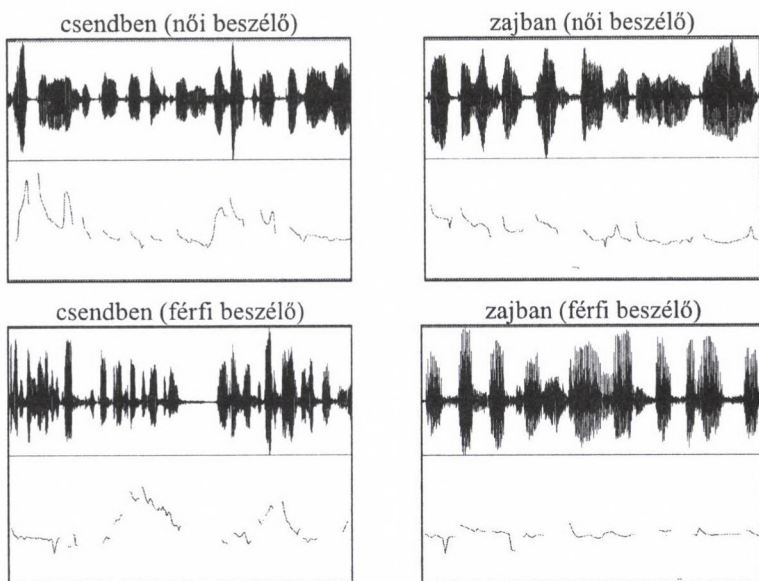
3. ábra

A beszéd intenzitásviszonyai csendes (bal oldali intenzitásgörbe, a vonal az átlagos 66 dB-t jelzi) és zajos (jobb oldali intenzitásgörbe, a vonal az átlagos 78 dB-t jelzi) környezetben

A beszéddallam elemzési eredményei további különbségeket igazoltak a csendes és a zajos környezetben elhangzó beszéd között. Az adatközlők monotonabban beszéltek a zaj hatására, alig-alig fordultak elő kiemelkedő dal-lameszcúcsok, a beszéddallam szinte nem mutatott változást az időben (vö. 4. ábra). Jellegzetesen módosult az a frekvenciatartomány is, amelyben a kísérleti személyek beszéltek. Csendben nagyobb mértékben használták ki az alaphangmagasság frekvenciatartományát, szélesebb tartományban valósult meg a beszédük. Zaj hatására szűkülés következett be.

Az alaphangmagasságban is módosulás volt tapasztalható, többnyire emelkedés. Ez bizonyos fokig az intenzitásnövekedés velejárója. A beszéddallam modulálása és a használt frekvenciatartomány mindazonáltal sokkal nagyobb egyéni különbségeket mutatott, mint az intenzitásváltozás. A férfi beszélőknél az alaphangmagasság átlagosan 9 Hz-cel növekedett zajos környezetben. Ez a változás azonban nem mindenkinél következett be, sőt akadtak olyan résztvevők is, akiknél csökkenést tapasztaltunk a csendes környezetben adatolt alaphangértékhez képest. Ha tekintetbe vesszük, hogy az intenzitás növekedése mindig a hangmagasság emelkedésével jár együtt, akkor mégis kimondható, hogy a zaj hatására valamiféle változás minden beszélőnél kimutatható volt. A férfiak átlagos alaphangmagassága csendben 98,14 Hz volt (az átlagos eltérés: 6,33 Hz), zajban 107,42 Hz (az átlagos eltérés: 8,28 Hz). Az értékek statisztikailag szignifikáns különbséget mutatnak (párosított  $t$ -próba:  $t(7) = -3,112$ ,  $p < 0,008$ ). A kísérletben részt vevő női beszélőknél azonban nem találtunk szignifikáns eltérést az alaphangmagasságban a csendes, illetve zajos

környezetben (párosított  $t$ -próba:  $t(7) = -2,211$ ,  $p < 0,063$ ). A nők alaphangmagasságának átlaga csendben 197,1 Hz volt (az átlagos eltérés: 9,96 Hz), zajban pedig 200,4 Hz (az átlagos eltérés: 8,32 Hz). Valamennyi beszélőnél szűkült a beszéddallam frekvenciasávja zaj hatására. Tizenöt beszélő esetében monoton(abb) lett a beszédprodukción zajban, mint csend esetén. Mindössze egyetlen (férfi) beszélőnél nem tapasztaltunk különbséget.



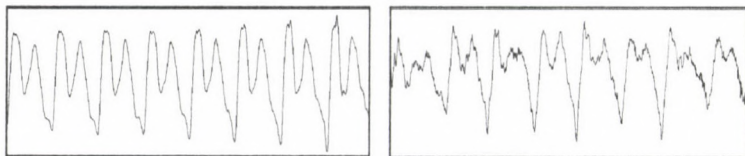
4. ábra

A beszéddallam változásai a csendes és a zajos környezet függvényében

A használt frekvenciasáv csendben 80 Hz és 210 Hz közötti volt a férfibeszélőknél és 100–290 Hz közötti a női beszélőknél. Zajos környezetben a férfiak beszéde 60–140 Hz-es sávra, a nők pedig 170–250 Hz-es frekvenciasávra módosult. A különbség a férfiaknál 50 Hz-es sáv, a nőknél pedig 110 Hz-es sáv, mindkét nemnél zajban tapasztaltuk a szűkebb frekvenciasávokat. A százalékban kifejezett különbség jelentős: a férfiaknál 11%, a nőknél viszont 49%. Ezek az értékek azt mutatják, hogy zaj hatására a nők beszéde nagyobb mértékben szűkül a frekvenciatartomány tekintetében, mint a férfiaké. Az abszolút értékek jelzik a jól ismert tényt, hogy a férfiak használt beszédfrekvencia-tartománya csendben szűkebb a nőkre jellemzőnél. Újabb vizsgálatok azt is kimutatták, hogy nem csupán a beszéd frekvenciatartomá-

nya, hanem a leggyakrabban használt frekvenciasáv is szűkebb a férfiaknál a nőkéhez viszonyítva (Beke 2008).

A zaj hatására a zöngéképzés is módosul, préseltebb, levegősebb lesz a megnövekedett intenzitással és a frekvenciatartomány szűkülésével együtt. Korábbi glottográfós vizsgálatok (Balázs–Gósy 1988) igazolták, hogy a kísérleti személyek nagyobbik részénél a glottogramok jelentős eltérést mutatnak a normálishoz képest. A zaj hatására préselt, feszített zöngét azonosítottak. Változott a nyakizomzati munka, a hangindítás és a levegővételi technika is. Az 5. ábra ugyanannak a beszélőnek azonos fonetikai helyzetű magánhangzójából kivágott öt periódusú szemléltet csendes és zajos beszédkörnyezetben. Csendes beszédkörülmenyek között a rezgések egyenletesebb periódusokat mutat, a zöreje jelenléte szinte nem tapasztalható, mindezt igazolja a spektrogram is. A zaj hatására módosult a zöngéképzés, s ennek a következménye látható a rezgésekben és a hangszínekben is. Zöreje superponálódott a zöngére, amely kevésbé egyenletes periódusokat tartalmazott, és jelentkezett a levegős zöngéképzés.



5. ábra

Az [e] magánhangzó nyolc-nyolc periódusának rezgése és hangszíne csendben (bal oldalon) és zajban (jobb oldalon)

Az artikulációs és a beszédtempót, illetőleg a szünethelyeket elemeztük mindkét beszédhelyzetben valamennyi kísérleti személynél. Szignifikáns különbséget találtunk a szünetek számában, azaz előfordulásában (párosított  $t$ -próba:  $t(15) = -10,193$ ,  $p < 0,001$ ) és időtartamában (független  $t$ -próba:  $F(1, 502) = 11,625$ ,  $p < 0,001$ ). A zaj hatására a beszélők gyakrabban és hosszabb szüneteket tartottak, mint csendes körülmények között (1. táblázat), bár az individuális eltérések elég nagyok. A szünetekben adatolható különbségek nem csupán a beszédtempó alakulására hatnak, hanem jelzik a beszédtervezési folyamatokban feltételezett eltéréseket. A több és hosszabb szünet arra utal, hogy a beszélők megszokott beszédproduktions stratégiáit megzavarja a szünet, mind az átalakítási, mind az előhívási folyamatok több időt vesznek igénybe, és nagyobb kontrollt igényelnek.

A háttérzaj hatására a beszélők többségénél a szünetek módosulásán és a megakadási jelenségek változásán túl számos egyéb megnyilatkozást is tapasztaltunk, kényszerű nevetéseket, köhécselest, torokköszörülést, lihegést.



1. táblázat: A szünetek előfordulásának és időtartamának adatai csendes és zajos körülmények között

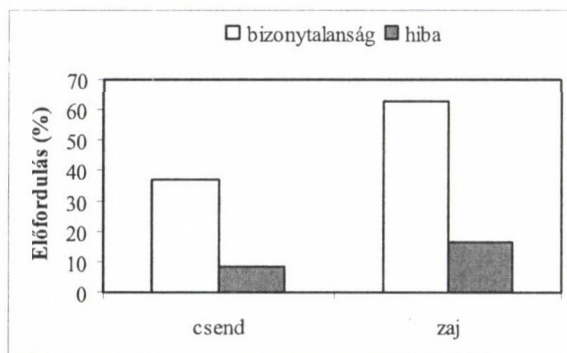
| Tényezők         | A spontán beszéd szüneteinek adatai |              |        |              |
|------------------|-------------------------------------|--------------|--------|--------------|
|                  | csendben                            |              | zajban |              |
|                  | Átlag                               | Átl. eltérés | Átlag  | Átl. eltérés |
| Előfordulás (db) | 22,93                               | 4,68         | 29,12  | 4,91         |
| Időtartam (ms)   | 204,18                              | 105,15       | 324,00 | 129,01       |

Az artikulációs tempó átlaga csendben 12,31 hang/s (az átlagos eltérés: 0,92 hang/s), míg zajban átlagosan 13,25 hang/s (az átlagos eltérés: 0,73 hang/s). A különbség ugyan kicsi, de szignifikáns (párosított  $t$ -próba:  $t(15) = -4,186$ ,  $p < 0,001$ ). Ez azt jelenti, hogy a beszédhangok sorozatainak képzésében a zajos környezetben bekövetkező gyorsulás nem véletlen, hanem a zajhatás következménye. Az egyéni különbségek várhatóan nagyok; vannak beszélők, akiknél alig jelentkezik eltérés a másodpercenként ejtett hangok számában, míg vannak, akiknél ez jelentős, akár 2-2,5 hang/s-mal is gyorsul az artikuláció zajos beszédhelyzetben. A beszédtempó (tehát a szüneteket és megakadásokat is tekintetbe vevő elemzés) értékei is matematikailag igazolható különbséget mutatnak (kétmintás  $t$ -próba:  $t(15) = 12,575$ ,  $p < 0,001$ ), ami előrejelezhető volt a szünetek hosszának és időtartamának ismeretében a kétféle beszédhelyzetben. A beszédtempó átlaga csendben 124,22 szó/perc (az átlagos eltérés: 9,63 szó/perc), míg zajban 109,26 szó/perc (az átlagos eltérés: 8,86 szó/perc). A különbség átlagosan 15 szó percenként. Az artikulációs tempó kisebb mértékben növekszik, mint amilyen mértékben a beszédtempó lassul, összességében tehát a zaj hatására lassulást tapasztalunk, ami a beszédtervezés érintettségére utal. Sajátos diszharmonia alakul ki. A beszélő tudat alatt igyekszik a kellemetlen helyzetből minél előbb szabadulni, ezért megpróbálja gyorsan megfogalmazni és kiejteni a gondolatait. A zavaró tényező (itt: a háttérzaj) miatt azonban a gondolatok nyelvi átalakításának folyamatai nehezítettek, amit a beszélő a gyorsabb artikulációval tovább nehezít. Mindennek eredményeként sűrűsödnek és hosszabbodnak a kitöltött és kitöltetlen szünetek.

Elemeztük a megakadásjelenségeket, mindkét típusúakat, a beszédtervezés bizonytalanságára utalókat és a hibákat. A bizonytalanságra utalók csendes és zajos helyzetben is előfordultak: ismétlések, hezitálások, nyújtások, újraindítások és töltelékszavak. A megjelenési gyakoriságuk azonban jelentősen megnövekedett a zajos környezetben (6. ábra).

A hibák (pl. nyelvtörténetek, téves kezdések, grammatikai hibák) előfordulása is jóval nagyobb mértékű a zajos környezetben rögzített spontán beszédanyagban. Nem található meg azonban minden lehetséges hibatípus. A gyakoriságon túl jellegzetes különbség a csendes és a zajos monológok között az,

hogy többféle hiba jelenik meg a zaj hatására. Sorrendiségi hiba például csendes körülmények között nem fordult elő.



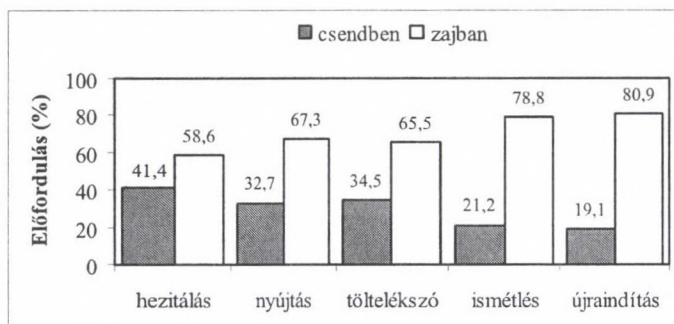
6. ábra

A megakadásjelenségek két típusának, a bizonytalanságnak és a hibának az előfordulása a kétféle beszédkörnyezetben

A megakadásjelenségek előfordulása egyértelműen összefügg a zaj megjelenésével, a két beszédhelyzetben rögzített spontánbeszéd-anyagok e tekintetben is szignifikáns különbséget mutatnak (egytényezős ANOVA:  $F(1, 31) = 12,517$ ,  $p < 0,001$ ). Az összes megakadás előfordulásának átlaga csendben 10,37 (az átlagos eltérés: 3,34), zajban pedig 18,50 (az átlagos eltérés 8,55). A jelen kísérletben részt vevő beszélők csendben átlagosan 1,41 megakadást produkáltak másodpercenként, míg zajban átlagosan 2,31 megakadást. A bizonytalanságra utaló megakadások közül csendben legnagyobb arányban a hezitálás fordult elő, míg zajban legtöbbször az újraindítás, majd a nyújtás, illetve az ismétlés (7. ábra). Néhány példa: *a m mindennapos programjaikat, ez ő ez ez családi, és aztán újra újakezdtém, jogászóók orvosóók meg mások, és nagyon sssok más*. Az ok feltehetően mindkét esetben az, hogy a nyelvi átalakítás a különböző beszédprodukciós szinteken nehezített, a nyújtások, az ismétlések és az újraindítások arra utalnak, hogy a kontroll mechanizmus kevésbé hatásosan működik.

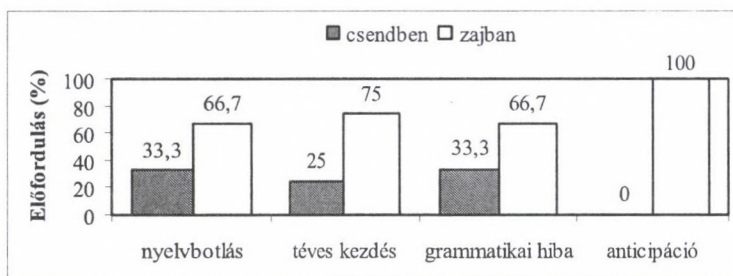
A legalább háromszor előforduló hibatípusokat szemlélteti a 8. ábra. Szembetűnő, hogy az anticipáció milyen gyakori zajos környezetben. Ez azt igazolja, hogy a gondolati tervezéssel a nyelvi tervezés nem tud megfelelően haladni, elővételez olyan nyelvi elemeket, amelyek az adott nyelvi struktúra egy későbbi időpillanatában szükségesek. Például: *gép idézőjelbe csak gépirni tudnak vagy szssz például ő számítógépes; hibás grammatikai forma: kérdezem én, akkor ne is tartják meg az órát?* (helyesen: *ne is tartják meg*). Meglehetősen nagy arányú a téves kezdések megjelenése, például *azon izgu-*

lok ho legjobban mert. Azt feltételeztük, hogy a nyelvbtlások jóval nagyobb mértékben fordulnak elő a zaj hatására, ez azonban nem igazolódott. A háttérzaj tehát nem feltétlenül az artikulációs kivitelezést, hanem legalább annyira a megelőző nyelvi tervezési folyamatokat bizonytalanítja el. Nem fordult elő ugyanakkor sok téves szótalálás (pl.: a *tézis*füzetet megszervezni, a szándékolt: *megszerkeszteni*).



7. ábra

A bizonytalanságra utaló megakadások arányai a kétféle beszédkörnyezetben



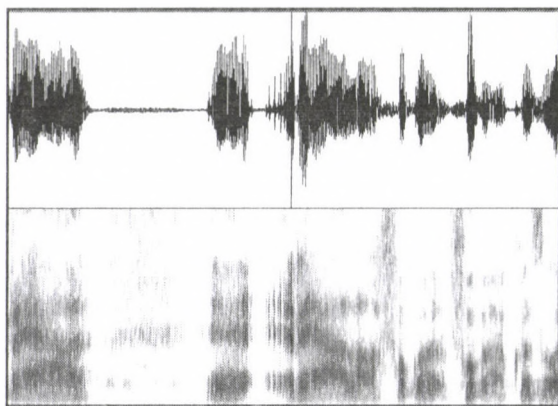
8. ábra

A hiba típusú megakadások arányai a kétféle beszédhelyzetben

A bizonytalanságra utaló és a hiba típusú megakadások láncszerűen is követhetnek egymást. A 9. ábra egy ilyen példát mutat be a rezgés kép és a hangszínkép segítségével, a beszédrészet: *reál reállal fogal foglalkozó*. A zajban beszélő önellenzési folyamata a kimondott *reál* szóról megállapítja, hogy hibás, ezt jelzi a hosszú, 673 ms-os néma, sóhajt tartalmazó szünet, majd az azt követő 250 ms-ös kitöltött szünet (hesztálás, az *ő* betűkkel jelölve). A 80 ms-os rövid néma szünet után egy 176 ms-os vokalizáció hallható, amit a korrigált szó, a *reállal* követ. A beszélő azonban alighogy túljut mindezen,



egy váratlan nyelvbtlás (kiesés) idéz elő ismét diszharmóniát, amelyet még a szó befejezése előtt észlel és azonnal javít (*fogal – foglalkozó*). Az önjavítás artikulációs gyorsulással jár együtt, amíg a *fogal* szókezdet ejtése 402 ms-ig tart, addig a *foglal* szórészlet tartama 326 ms. A 3,136 másodperces beszéd-részletben hatféle megakadás fordult elő.

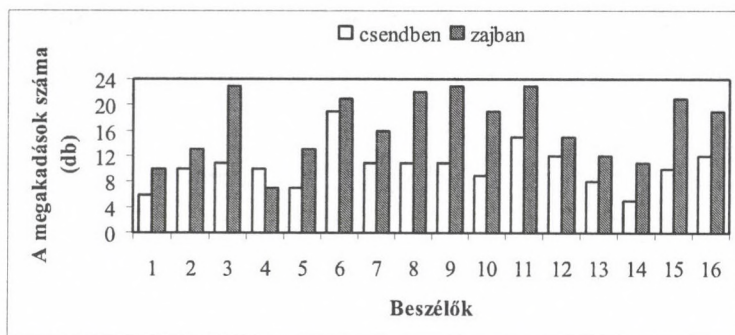


reál SÓHAJT őő NYÖG reállal fogal foglalkozó  
9. ábra

Megakadássorozat egy zajban elhangzó spontán beszédből

Az egyének közötti különbségek valamennyi megakadás esetében nagyok, pontosabban azok mennyiségében és típusainak megjelenésében az egyes beszélők jelentős eltéréseket mutatnak (10. ábra). Vannak olyanok, akiknél alig van különbség a megakadások számában a kétféle beszédhelyzetben, sőt egy férfi beszélőnél csendben valamivel több megakadást adatoltunk, mint zajban. A résztvevők nagyobb részénél azonban a beszédzaj hatására jelentősen megnövekedtek a diszharmóniás jelenségek, s ezek különféle megakadásokban voltak tetten érhetők.

A különböző zajoknak (fogfúrás, kutyaugatás, légkalapács, zene, fülhallgatón közvetítve, a zaj/jel viszony 80 dB volt) a spontán beszédre gyakorolt hatását vizsgálta kísérletében Gyarmathy (2008). Adatközlőinél kismértékben csökkent a megakadások száma a zaj hatására, kivéve a kutyaugatást. Az egyes megakadástípusok száma ugyanakkor megnövekedett a zajok hatására. A fogfúrás hatására például 18-féle megakadást produkáltak a kísérleti személyek. A beszélők által tartott szünetek hossza a különféle zajok hatására megnövekedett (ez alól csak a zene volt kivétel).



10. ábra

A megakadások száma személyenként a kétféle helyzetben

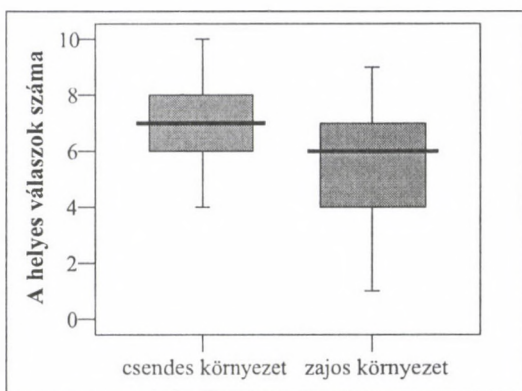
## 2. Beszédmegértési kísérlet

Említettük, hogy a háttérzaj nemcsak a beszédprodukcióra, hanem a beszédmegértésre is hatást gyakorolt. Az eddigi kutatások szerint a zajban csökkent beszédérthetőség miatt a hallgató megértése romlik, avagy csak fokozott figyelemkoncentrációval tudja ellensúlyozni a zavaró hatást. Saját vizsgálataink igazolták, hogy a zajjal fedett mondatok felismerése gyermekkorban nehéz, és életkor-specifikusan működik (Gósy–Horváth 2006). A jelen kísérletben 16 és 18 éves gimnáziumi tanulók szövegértését teszteltük. A szövegértési adatok mindkét korosztályban különbséget mutatnak a beszédzaj jelenlététől függően (3. táblázat).

3. táblázat: A szövegértés adatai zajos/nem zajos szöveg esetén

| Csoportok | Szövegértési teljesítmény (%) |              |        |              |
|-----------|-------------------------------|--------------|--------|--------------|
|           | csendben                      |              | zajban |              |
|           | Átlag                         | Átl. eltérés | Átlag  | Átl. eltérés |
| 16 évesek | 67,7                          | 17,7         | 51,3   | 21,4         |
| 18 évesek | 74,5                          | 13,7         | 59,5   | 18,8         |

A két korosztály között sem a zaj nélküli, sem a zajos szöveg megértésében nem volt szignifikáns különbség, azonban mind a 16, mind a 18 évesek matematikailag igazolható eltérést mutattak a zajos és a nem zajos szöveg megértésében (párosított  $t$ -próba 16 éveseknél:  $t(21) = -2,737$ ,  $p < 0,012$ ; 18 éveseknél:  $t(21) = -3,139$ ,  $p < 0,005$ ). A 11. ábra grafikonja összevontan szemlélteti a 88 diák beszédmegértési adatait a zajos/nem zajos szöveg függvényében; az eltérések jól láthatók.



11. ábra

A szövegértést ellenőrző kérdésekre adott helyes válaszok száma  
a zaj nélküli és a zajjal fedett szöveg esetén  
(az összes diák adata alapján, medián és szóródás)

### Következtetések

Zajos környezetben a beszélők hangosabban és kissé gyorsabban, szűkebb frekvenciasávban és monotonabban beszélnek. Ilyenkor a beszédfeldolgozás korlátozott, a megértés biztonsága csökken, az értelmezés pontossága gyengébb, mint csendben. Kimondható, hogy a zaj egyértelműen negatív, nemkívánatos hatást gyakorol mind a beszélőre, mind a hallgatóra. A zaj kellemetlen vegetatív és pszichés állapotot is teremt(het) a verbális kommunikációban résztvevők számára.

Az audiológia hallási adaptációnak nevezi azt a jelenséget, amikor az érzékenység csökken a megelőző hanginger következtében (Pauka 1982). Az adaptáció viszonylag rövid és kis intenzitású hangbehatás után is kimutatható. A Lombard-hatás az adaptációnak egy sajátos megnyilvánulása, amikor a beszédprodukciónak minőségét érinti a környezeti hanghatás (a zaj). Megállapítható, hogy a Lombard-hatás nem csupán az intenzitás megemelkedésében jelentkezik. Multifaktoriális jelenségről van szó, amely a beszédtervezés valamennyi szintjét érinti; a hallási visszacsatolás alapján a beszédtervezési és a beszédkivitelezési folyamatok működési diszharmonijához vezet. A zajhatás természetesen nem független a zajtípustól és az intenzitástól. Az impulzusos zaj sokkal kellemetlenebb, és bizonyos körülmények között súlyosabb halláskárosodást okoz, mint az állandósult zajállapot. Különösen fontos figyelemmel lenni például az osztálytermi zajszintre (ez nem haladhatja meg a 35 dB-t a környezeti zaj értékeléséről és kezeléséről szóló 280/2004-es kormányrendelet szerint). Adataink azt mutatták, hogy még a relative halk tárgyalás is beszédmegértési nehézségekhez vezet.



## Irodalom

- Amazi, Deborah K. – Garber, Sharon R. 1982. The Lombard sign as a function of age and task. *Journal of Speech and Hearing Research* 25. 581–585.
- Balázs Boglárka – Gósy Mária 1988. Környezetünk hangjelenségeinek hatása a beszédre. *Fül-Orr-Gégégyógyászat* 34. 145–150.
- Beke András 2008. A felolvasás és a spontán beszéd alaphangszerkezeteinek vizsgálata. *Beszéd kutatás* 2008. 94–108.
- Von Békésy, George 1949. The structure of the middle ear and the hearing of one's own voice by bone conduction. *Journal of the Acoustical Society of America* 21. 217–232.
- Brumm, Henrik 2004. The impact of environmental noise on song amplitude in a territorial bird. *Journal of Animal Ecology* 73. 434–441.
- Brumm, Henrik – Slater, Peter J. B. 2006. Animals can vary signal amplitude with receiver distance: evidence from zebra finch song. *Animal Behaviour* 72. 699–705.
- Brumm, Henrik – Voss, Katrin – Köllmer, Ireen – Todt, Dietmar 2004. Acoustic communication in noise: Regulation of call characteristics in a New World monkey. *Journal of Experimental Biology* 207. 443–448.
- Castellanos, Antonio – Benedí, José-Miguel – Casacuberta, Francisco 1996. An analysis of general acoustic-phonetic features for Spanish speech produced with the Lombard effect. *Speech Communication* 20. 23–35.
- Egnor, Roian, S. E. – Hauser, Marc D. 2006. Noise-induced vocal modulation in cotton-top tamarins (*Saguinus oedipus*). *American Journal of Primatology* 68. 1183–1190.
- Ghazanfar, Asif A. – Pinsk, Mark A. 2003. Speech perception: Linking comprehension across a cortical network. *Current Biology* 17. 29–33.
- Gósy Mária 2005. *Pszicholingvisztika*. Osiris Kiadó, Budapest.
- Gósy Mária – Horváth Viktória 2006. Beszédfeldolgozási folyamatok összefüggései gyermekkorban. *Magyar Nyelvőr* 130. 470–482.
- Gósy Mária 2008. Önellenőrzési folyamatok a beszédben. *Magyar Nyelv* (megjelenőben).
- Gyarmathy Dorottya 2008. Különböző zajok hatása a beszédprodukcóra. *Alkalmazott Nyelvtudomány* (megjelenőben).
- Hansen, John H. L. 1996. Analysis and compensation of speech under stress and noise for environmental robustness in speech recognition. *Speech Communication* 20. 151–173.
- Hativa, Niva 2000. *Teaching for effective learning in higher education*. Springer, Berlin-Heidelberg.
- Hygge, Staffan 2003. Classroom experiments on the effects of different noise sources and sound levels on long-term recall and recognition in children. *Applied Cognitive Psychology* 17. 895–914.
- Junqua, Jean-Claude 1996. The influence of acoustics on speech production: A noise-induced stress phenomenon known as the Lombard reflex. *Speech Communication* 20. 13–22.
- Kloepfer, Michael – Griefahn, Barbara – Kaniowski, Andrzej M. – Klepper, Gernot – Lingner, Stephan – Steinebach, Gerhard – Weyer, Heinrich W. – Wysk, Peter 2006. *Leben mit Lärm?* Springer, Berlin-Heidelberg.
- Kryter, Karl D. 1959. Scaling human reaction to sound from aircraft. *Journal of the Acoustical Society of America* 31. 1415–1429.

- Lane, Harlan – Tranel, Bernard 1971. The Lombard sign and the role of hearing in speech. *Journal of Speech and Hearing Research* 14. 677–709.
- Lazarus, H. 1990. New methods for describing and assessing direct speech communication under disturbing conditions. *Environment International* 16, 373–392.
- Lombard, Etienne 1911. Le signe de l'élévation de la voix. *Annales des Maladies de l'Oreille et du Larynx* 37. 101–119.
- Nonaka, Satoshi – Takahashi, Ryuji – Enomoto, Keiichi – Katada, Akihiro – Unno T. 1997. Lombard reflex during PAG-induced vocalizations in decerebrate cats. *Neuroscience Research* 29. 283–289.
- Pauka Károly 1982. A beszéd megértése. In Bolla Kálmán (szerk.): *Fejezetek a magyar leíró hangtanból*. Budapest, Akadémiai Kiadó, 175–233.
- Postma, Albert – Kolk, Herman J. 1993. The covert repair hypothesis: prearticulatory repair processes in normal and stuttered disfluencies. *Journal of Speech and Hearing Research* 36. 472–87.
- Rabin, Lawrence A. – McCowan, Brenda – Hooper, Stacie L. – Owings, Donald H. 2003. Anthropogenic Noise and its effect on animal communication: an interface between comparative psychology and conservation biology. *International Journal of Comparative Psychology* 16. 172–192.
- Rosenblum, Lawrence, D. – Johnson, Jennifer A. – Saldaa, Helena M. 1996. Point-light facial displays enhance comprehension of speech in noise. *Journal of Speech and Hearing Research* 39. 1159–1170.
- Schneider, Bruce A. – Daneman, Meredyth – Murphy, Dana R. 2005. Speech comprehension difficulties in older adults: Cognitive slowing or age-related changes in hearing? *Psychology and Aging* 20. 261–271.
- Skowronski, Mark D. – Harris, John G. 2006. Applied principles of clear and Lombard speech for automated intelligibility enhancement in noisy environments. *Speech Communication* 48. 549–558.
- Summers, W. Van – Pisoni, David B. – Bernacki, Robert H. – Pedlow, Robert I. – Stokes, Michael A. 1988. Effects of noise on speech production: Acoustic and perceptual analyses. *Journal of the Acoustical Society of America* 84. 917–928.
- Wassink, Alicia Beckford – Wright, Richard A. – Franklin, Amber D. 2007. Intraspeaker variability in vowel production: An investigation of motherese, hyper-speech, and Lombard speech in Jamaican. *Journal of Phonetics* 35. 363–379.
- Winkworth, Alison L. – Davis, Pamela J. 1997. Speech breathing and the Lombard effect. *Journal of Speech, Language and Hearing Research* 40. 159–169.
- Zeine, Lina – Brandt, John F. 1988. The Lombard effect on alaryngeal speech. *Journal of Communication Disorders* 21. 373–383.

## NAZÁLIS HATÁS A MAGÁNHANGZÓKÉPZÉS FÜGGVÉNYÉBEN

Horváth Viktória

### Bevezetés

A folyamatos beszédben a beszédhangok nem egymás után ejtett egységek, hanem különféleképpen hatnak egymásra, módosítják a szomszédos hangok artikulációs konfigurációját, ezáltal akusztikai szerkezetét. A hangképző szerveknek bizonyos időre van szükségük ahhoz, hogy az egyik hang artikulációs konfigurációjáról áttérjenek a másik hang képzési helyzetébe. Ez a jelenség a koartikuláció. A koartikuláció egyrészt fiziológiai okokra vezethető vissza, amelyek meghatározzák az egyes hangkapcsolatok megvalósulását. A beszédhangok egymásra hatása másfelől nyelvspecifikus, vagyis az egyes nyelvek között nagy különbségek lehetnek a koartikuláció megvalósulásában, és ezek nagymértékben függenek a hangkapcsolatot alkotó beszédhangok minőségétől. Továbbá az egyéni ejtés is eredményezhet sajátos koartikulációs megoldásokat (Gósy 2004). A koartikuláció akusztikai következménye a beszédben többféle módon érvényesül, ezek közül az egyik a magánhangzók formáns-szerkezetének változása különféle hangkörnyezetekben (Nikléczy 2006). A fonetikai vizsgálatok egyik, a nemzetközi szakirodalomban legtöbbször elemzett kérdésköre a nazális hatás, vagyis hogy a nazális mássalhangzók artikulációjának milyen következményei vannak a szomszédos beszédhangokra (Malécot 1960, Whalen–Beddor 1989, Chen et al. 2007, Gósy–Beke 2008). A nazális mássalhangzók első nyelvészeti szempontú leírása az i. e 5. századból származik; ezek képzésének, illetve akusztikai szerkezetének „műszeres” vizsgálata azonban csak a 20. században indulhatott meg. A nazális mássalhangzók percepció-s szempontú vizsgálata az 1950-es évekre tehető (vö. Delvaux et al. 2007). A nazális hangok létrehozásában az orrüreg mint rezonátor vesz részt, amely a garatüreggel és a szájüreggel ellentétben merev. A lágy szájpad „zárként” viselkedik: nazális zárhangok képzésekor leereszkedik, szabaddá téve a levegő útját, amely az orrüregen keresztül folyamatosan áramlik ki, miközben a szájüregben valahol, avagy az ajkak között zár keletkezik. A nazális mássalhangzók képzése sem független azonban a kontextustól. A magyarra vonatkozó legújabb kutatás eredményei szerint a VNC hangkapcsolatokban a nazális mássalhangzó időszerkezete attól függően változik, hogy milyen a követő mássalhangzó képzésmódja. Résképzésű mássalhangzó hatására a nazális időtartama rövidül, az előtte lévő magánhangzó időtartama – ellensúlyozva a nazális időtartamának csökkenését – nő (Gósy–Beke 2008).



Orális hangok képzésekor (ilyenek a magyarban a magánhangzók) az orr-üreg felé vezető út a lágy szájpad és az uvula működtetésével záródik; így a levegő a szájüregen keresztül távozik. A nazális mássalhangzók és a magánhangzók találkozásakor az utóbbiak artikulációs konfigurációja azonban különféle mértékben módosulhat. Ekkor a lágy szájpad leereszkedik, de nem képződik zár a szájüregben, a levegő tehát folyamatosan áramlik az orr- és szájüregen keresztül egyaránt (Rose 2002). Angol nyelvre vonatkozó műszeres elemzések azt mutatták, hogy a magánhangzó akkor is nazalizálódik, ha csupán az egyik szomszédos mássalhangzó nazális; ugyanakkor a teljes nazális környezet (NVN) nem eredményezi a magánhangzó nagyobb mértékű nazalizációját (Chen et al. 2007). A vokálisok és a nazálisok formánsértékeinek viszonyától függ az, hogy a koartikuláció mely frekvenciákon eredményez intenzitásváltozást. Az angol nyelvre elvégzett kutatások azt mutatták, hogy a felső nyelvállású magánhangzók akusztikai szerkezete nagyobb mértékben módosul a nazalizáció hatására, mint az alsó nyelvállású magánhangzóké (Bell-Berti 1993). A nazalizáció hatása a magánhangzók formánsértékeire nagymértékben függ továbbá a hangkapcsolatban részt vevő beszédhangok minőségétől, illetve a szótagszerkezettől (Pickett 1999). Olyan hangkapcsolatokban, ahol a magánhangzót nazális, majd nem nazális mássalhangzó követi (VNC), a nazalizáció nagyobb mértékű, ha a nem nazális mássalhangzó zöngétlen (Malécot 1960). Újabb kutatási eredmények szerint a nazalizálódó magánhangzó időtartama hosszabb abban az esetben, ha az azt követő nazális (VN) rövidebb időtartamban valósul meg (Beddor 2007).

A nazálisok hatásának átfogó leírását megnehezíti az a tény, hogy a jelenség az egyes magánhangzók akusztikai szerkezetére más-más módon hat (Ladefoged 2005). Az 1950-es években meginduló kutatások az angol és a francia nyelvben vizsgálták a nazális beszédhangokat és artikulációs hatásukat (vö. Delvaux et al. 2007). A fonetika eszköztárának fejlődése lehetővé tette a nazalizációra vonatkozó elemzések számának növekedését – a különféle nyelvek (Hardcastle–Hewlett 1999) mellett a magyarra is készült néhány kutatás e kérdéssel kapcsolatban. Az eredmények szerint a magánhangzó két nazális mássalhangzó között nazalizálódik (Földi 1989), de abban az esetben is kimutatható a koartikulációs hatás, ha a hangkapcsolatban csak egy nazális szerepel – a magyarban a progresszív nazalizáció erősebb hatású, mint a regresszív (Horváth 2005). Minden tudományos kérdésre még nem sikerült választ adni. A jövő feladatai közé tartoznak a több beszélőtől rögzített, statisztikai elemzésekkel alátámasztott, minden magyar magánhangzóra kiterjedő kutatások, illetve annak vizsgálata, hogy a nazalizáció hogyan valósul meg a spontán beszédben.

A jelen kutatás a magyar nyelv kardinális magánhangzói közül kettő, az [i] és az [a:] akusztikai szerkezetének kontextusfüggő módosulását vizsgálja. A kardinális magánhangzók viszonyítási pontokat képeznek az artikulációs rendszerben, amelyekhez a többi magánhangzó képzési konfigurációja viszo-

nyítható (Gósy 2004). Az egyik a legmagasabban és a legelőrébb képzett (a magyarban az [i]), a másik a leghátrább és legalacsonyabban képzett, az [a:]. Arra a kérdésre kerestük a választ, hogy a nazális környezet hatására milyen módosulás tapasztalható ezen magánhangzók akusztikai szerkezetében a nazálist nem tartalmazó kontextushoz képest. A többi magyar magánhangzó nazalizációja a későbbiekben összehasonlítható lesz a kardinálisoknál tapasztalt módosulással. A jelen kutatásban csak a három fonémaértékű nazális mássalhangzó koartikulációs hatását elemeztük, a /m/ és a /n/ allofónjai ([m] és [ŋ]) ugyanis ritkán fordulnak elő a spontán beszédben (megjelenésük 0,1%-nál alacsonyabb), megvalósulásuk fonetikai pozícióhoz kötött (Gósy 2004).

Hipotézisünk szerint a két nazális mássalhangzó közötti helyzet statisztikailag igazolható eltéréseket eredményez a magánhangzó formáns- és időtartamértékeiben. Azt feltételeztük a nemzetközi szakirodalom, valamint a magyarra vonatkozó korábbi elemzések alapján (Horváth 2005), hogy a három mássalhangzó eltérő mértékben befolyásolja a vokálisok akusztikai szerkezetét, továbbá az előre ható nazalizáció (amikor a nazális a magánhangzó előtt áll) erősebben érvényesül a hátra ható folyamathoz képest.

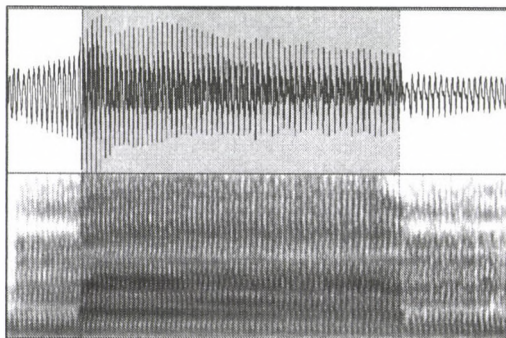
### Anyag és módszer

A vizsgálathoz CVC szerkezetű logatomokat hoztunk létre, amelyekben a két elemzett magánhangzó különböző hangkörnyezetekben szerepelt. A környező mássalhangzók a magyar fonémaértékű nazálisok voltak: a bilabiális [m], az alveoláris [n] és a palatális [ɲ]. A magánhangzók szerkezetét olyan hangsorok azonos magánhangzóival hasonlítottuk össze, amelyek egyáltalán nem tartalmaztak nazálist, ebben az esetben a mássalhangzó a [j] posztalveoláris zöngétlen réshang volt (*sis*, *sás*). (A későbbiekben a C mindig ezt a mássalhangzót jelöli, a nazális mássalhangzókat pedig N jelöli a hangkapcsolatokban). A magánhangzókat olyan logatomokban is rögzítettük, amelyekben csak az egyik mássalhangzó volt nazális képzésű (CVN és NVC). Az adatközlőket arra kértük, hogy a négy hangsort olvassák fel egymás után, például: *sis*, *mis*, *sim*, *mim*.

A két kardinális magánhangzó koartikulációját vizsgáltuk mind a három fonémaértékű nazálissal, tehát egy beszélőtől 24 hangsort rögzítettünk. A két mássalhangzó közötti magánhangzó akusztikai szerkezetét elemeztük a réshangkörnyezetben, a teljes nazális környezetben; valamint vizsgáltuk az első vokális szerkezetének változását attól függően, hogy a nazális megelőzi (NVC), vagy követi a magánhangzót (CVN). A nazális hatás természetesen nem írható le egyetlen akusztikai paraméterrel, a formánsszerkezet változásán kívül megjelenhet az első formánsamplitúdó csökkenésében, vagy egy új csúcs megjelenéseként a spektrumban 1000 Hz körül. A jelen kutatásban a formánsszerkezetre és az időtartamra vonatkozó méréseket végeztünk a Praat program 4.4-es verziójával. Az első három formáns értékét a magánhangzó tiszta fázisának közepén mértük a Praat formánselemzőjével automatikusan,



de minden esetben megtörtént a kontroll, irreális adat esetén leolvasással a spektrogramról. Meghatároztuk a magánhangzók teljes időtartamát a mássalhangzók közötti fonetikai pozícióban a második formáns mozgása, a rezgés-kép és a többszöri meghallgatás alapján (1. ábra).



1. ábra

A *mám* hangsor rezgés- és hangszínképe  
(A szürkített rész a magánhangzót mutatja.)

A kísérletben 20 női adatközlő vett részt. Mindannyian egyetemisták vagy egyetemet végzettek; 20–35 évesek. A 24 hangsort a húsz adatközlő bemondásában digitálisan rögzítettük (44 kHz, 16 bit). A mintegy 2000 adat statisztikai elemzése (egytényezős ANOVA, Tukey-féle post hoc teszt) a 13.0 verziószámú SPSS programmal történt.

### Eredmények

#### Módosulások az [i] akusztikai szerkezetében a nazális környezet hatására

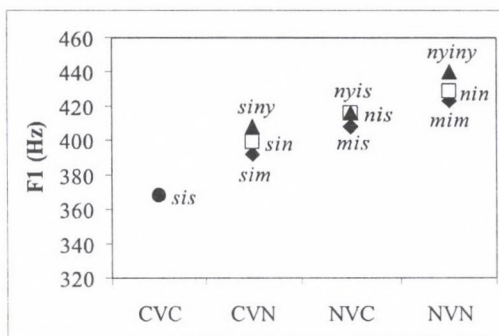
Az [i] magánhangzó formánsszerkezetére vonatkozó elemzés azt mutatta, hogy mindhárom fonémaértékű nazális mássalhangzó akkor változtatja meg a magánhangzó akusztikai szerkezetét, ha megelőzi azt (NVC helyzet), vagy ha az [i] teljes nazális környezetben fordul elő (NVN helyzet). Nem volt statisztikailag kimutatható eltérés a réshangkörnyezetben (CVC) mért értékek és a között, ha a magánhangzót követte a nazális (CVN-helyzet). A magánhangzó szerkezetére tehát kevésbé hat a követő hang minősége – a magánhangzó hasonló akusztikai szerkezettel valósul meg függetlenül attól, hogy nazális vagy réshang követi.

Az angol és a holland nyelvre vonatkozó kutatások (Hardcastle–Hewlett 1999) szerint a progresszív nazalizáció nagyobb mértékben módosítja a magánhangzó akusztikai szerkezetét, mint a regresszív. A jelen kutatás igazolta a hipotézist, ez a tendencia a magyar beszédre is igaz. A regresszív nazalizáció erősebb hatása független a nazális mássalhangzó minőségétől, vagyis mind-



három elemzett nazális esetében igazolható. A formánsszerkezetre gyakorolt nazális hatás a frekvenciaértékek emelkedésében realizálódik. A progresszív nazalizáció erősebb hatását az mutatja, hogy a réshangkörnyezethez képest az előre ható nazalizáció hatására a formánsok magasabb frekvencián valósultak meg, mint a hátra ható esetén (vö. 2–4. ábra).

Az [i] formánsai azonban eltérő mértékben változnak attól függően, hogy melyik nazális mássalhangzó hatása érvényesül. Az első formáns ( $F_1$ ) értékének változásában (2. ábra) csak a bilabiális nazális hatására mutatható ki statisztikailag szignifikáns különbség a réshangkörnyezethez képest [ $F(3, 79) = 2,669$ ;  $p = 0,041$ ], és csak abban az esetben, ha a magánhangzó két [m] között fordul elő (*mim*).



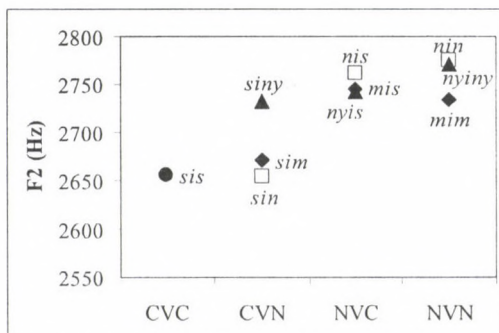
2. ábra

Az [i] első formánsának átlagértékei a különféle hangkörnyezetekben

Az [i] második formánsának ( $F_2$ ) értéke emelkedik két nazális között, de statisztikailag igazolható eltérés csak a [n] [ $F(3, 79) = 5,515$ ;  $p < 0,002$ ] és a [ɲ] [ $F(3, 79) = 3,764$ ;  $p < 0,014$ ] esetében mutatható ki, a bilabiális nazális nem változtatja meg az  $F_2$ -t szignifikáns mértékben (3. ábra). A [n] koartikulációs hatására az  $F_2$  akkor is nagymértékben (post hoc teszt:  $p < 0,006$ ) változik a réshangkörnyezethez képest, ha a nazális megelőzi a magánhangzót (*nis*) – a progresszív nazalizáció miatt az [i] második formánsának módosulása hasonló mértékű, mint amikor teljes nazális környezetben fordul elő (*nin*).

A magánhangzó harmadik formánsa ( $F_3$ ) a beszélő személy egyéni ejtéséről is tartalmaz információt, a magyarban általában csak másodlagosan jellemzi a magánhangzót (Gósy 2004). A nazalizáció nagymértékben érinti az [i] harmadik formánsát, amely mindhárom nazális hatására szignifikánsan változik a réshangkörnyezethez képest – [m]:  $F(3, 79) = 5,296$ ,  $p < 0,002$ ; [n]:  $F(3, 79) = 6,214$ ,  $p < 0,001$ ; [ɲ]:  $F(3, 79) = 17,232$ ,  $p < 0,001$ . Az  $F_3$  nemcsak teljes nazális környezetben változik, hanem az előre ható nazalizáció hatására is (post hoc teszt: [m]:  $p = 0,042$ ; [n]:  $p = 0,034$ ; [ɲ]:  $p < 0,001$ ). Szig-

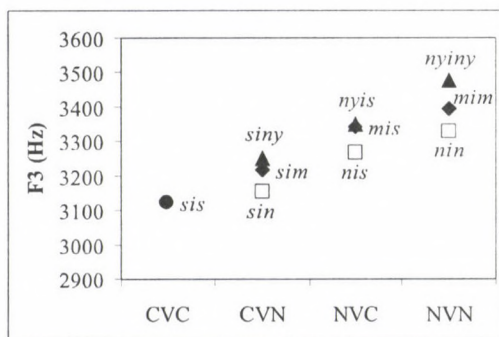
nifikáns különbséget találtunk továbbá a *sin* és *nin* (post hoc:  $p = 0,013$ ), valamint a *siny* és *niny* (post hoc:  $p < 0,001$ ) hangsorok magánhangzójának  $F_3$  értéke között. Ez azt mutatja, hogy a regresszív nazalizáció nincs hatással az [i] harmadik formánsára (a követő hang nem befolyásolja; hasonló frekvenciaértékekkel valósul meg attól függetlenül, hogy réshang vagy nazális követi).



3. ábra

Az [i] második formánsának átlagértékei a különféle hangkörnyezetekben

Megállapítható tehát, hogy az [i] esetében harmadik formáns érintett a legnagyobb mértékben a nazalizációban. (Korábbi kutatásainkban ezt a beszélők kevés száma miatt még nem sikerült igazolni.) A változás minden esetben az értékek emelkedésében nyilvánul meg a nazalist nem tartalmazó környezet-höz képest (4. ábra).



4. ábra

Az [i] harmadik formánsának átlagértékei a különféle hangkörnyezetekben

Az [i] időtartamát az egyes hangsorokban vizsgálva azt tapasztaltuk, hogy a hangkörnyezet nem befolyásolja nagymértékben a magánhangzó időtartamát; nem találtunk szignifikáns különbséget a réshangkörnyezetben és a nazális tartalmazó hangsorokban mért értékek között. A nazális hatás tehát a formánsszerkezetekben mutatható ki, de nem érinti a magánhangzó időtartamát (1. táblázat).

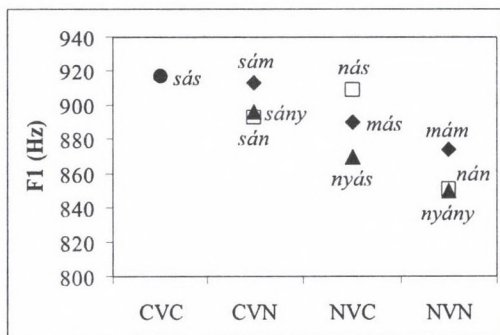
1. táblázat: Az [i] időtartama a különféle hangkörnyezetekben (ms)

|           | CVC  | CVN  |      |      | NVC  |      |      | NVN  |      |      |
|-----------|------|------|------|------|------|------|------|------|------|------|
|           |      | [m]  | [n]  | [ɲ]  | [m]  | [n]  | [ɲ]  | [m]  | [n]  | [ɲ]  |
| Átlag     | 125  | 134  | 134  | 138  | 123  | 126  | 122  | 120  | 127  | 121  |
| Átl. elt. | 26,7 | 25,3 | 27,6 | 37,3 | 26,8 | 32,8 | 38,1 | 25,4 | 25,0 | 27,7 |

Az időtartam-változás nem szignifikáns a hangkörnyezettől függően, de bizonyos temporális módosulások megfigyelhetők. Az [i] átlagos időtartama a CVN hangkapcsolatban a leghosszabb, vagyis abban az esetben, amikor a nazális mássalhangzó követi a magánhangzót. Ez a tendencia érvényesül a szintetizált beszéd alapján meghatározott specifikus időtartam esetén is (Olasz 2006).

#### Módosulások az [a:] akusztikai szerkezetében a nazális környezet hatására

Az [a:] nazalizációjára vonatkozó elemzések jóval árnyaltabb képet mutatnak a palatális magánhangzóhoz képest. Az [a:] első formánsának értéke szignifikánsan nem változik egyik nazális hatására sem, noha bizonyos módosulások kimutathatók (5. ábra). A módosulás fő tendenciája a frekvenciaértékek csökkenése (és nem az emelkedés, mint a palatális magánhangzónál). A legnagyobb csökkenés az  $F_1$  értékében a [n] hatására következik be a réshangkörnyezethez képest (a húszt beszélőnél mért értékek átlagát tekintve 73 Hz).

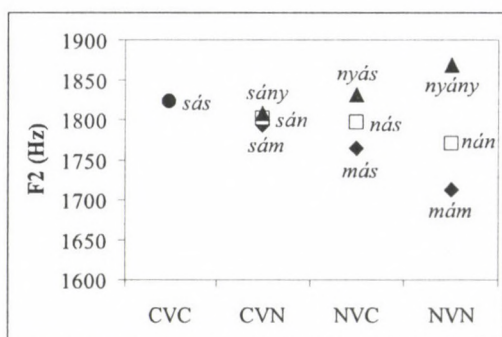


5. ábra

Az [a:] első formánsának átlagértékei a különféle hangkörnyezetekben



Az [a:] második formánsa csupán a bilabiális nazális hatására változik szignifikánsan, de csak abban a hangsorban, amelyben a vokális két nazális között fordult elő [ $F(3, 79) = 5,317, p = 0,002$ ], tehát sem a progresszív, sem a regresszív nazalizáció önmagában nem idéz elő jelentős változást az  $F_2$  értékeiben. Ennek ellenére a módosulás tendenciái kimutathatók: a formánsértékekben a [m] és a [n] hatására csökkenést, a [n] hatására pedig emelkedést tapasztaltunk (6. ábra). A progresszív nazalizáció az  $F_2$ -nél is nagyobb mértékű, mint a regresszív.

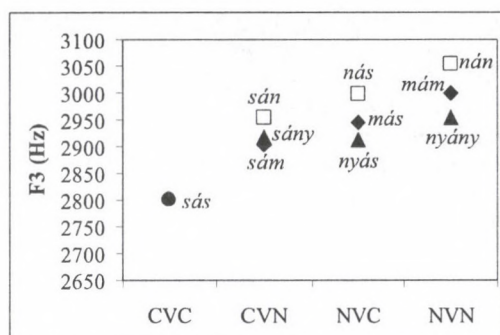


6. ábra

Az [a:] második formánsának átlagértékei a különféle hangkörnyezetekben

Az [a:] harmadik formánsa a legnagyobb mértékben érintett a nazalizációban – csakúgy, mint a palatális magánhangzóé. Mindhárom nazális környezetben szignifikánsan változik, a nazális hatás ismét az értékek emelkedésében nyilvánul meg, mint az [i] esetében (7. ábra). Az  $F_3$  változik a réshangkontextushoz képest a bilabiális nazális hatására [ $F(3, 79) = 4,908, p = 0,004$ ] abban az esetben, ha a vokalist két [m] fogja közre (mám, post hoc teszt:  $p = 0,002$ ), és a progresszív nazalizáció következményeként is (más, post hoc:  $p = 0,043$ ). A [n] is szignifikáns változást idéz elő a harmadik formáns értékében [ $F(3, 79) = 9,092, p < 0,001$ ]. A legerősebb ez a hatás két alveoláris nazális között (post hoc teszt:  $p = 0,001$ ) és a progresszív nazalizáció hatására (post hoc:  $p < 0,001$ ), de a regresszív nazalizáció esetén is kimutatható a sán hangsorban (post hoc:  $p = 0,014$ ). A palatális nazális koartikulációs hatása csak a két nazális közötti helyzetben idéz elő nagymértékű változást a magánhangzó harmadik formánsának értékében (sás – nyány post hoc:  $p = 0,026$ ).

Az [a:] magánhangzónál a progresszív nazalizáció kizárólagos dominanciájára vonatkozó hipotézisünket teljesen nem, de nagymértékben sikerült igazolni. Az elemzett esetek többségében (67%) az előre ható nazalizáció érvényesült erősebben (vagyis nagyobb mértékben módosultak a formánsértékek annál, mint amikor a nazális követi a magánhangzót).



7. ábra

Az [a:] harmadik formánsának átlagértékei a különféle hangkörnyezetekben

A különböző hangsorok [a:] magánhangzójának időtartamában jóval nagyobb különbségeket találtunk, mint a palatális magánhangzónál (2. táblázat). Az időtartam ugyan módosul a különböző kontextusokban, de szignifikáns különbséget nem találtunk egyik nazális mássalhangzó hatására sem – az [a:] nazalizációja sem jelenik meg az időtartamban.

2. táblázat:

Az [a:] időtartama a különféle hangkörnyezetekben (ms)

|           | CVC  | CVN  |      |      | NVC  |      |      | NVN  |      |      |
|-----------|------|------|------|------|------|------|------|------|------|------|
|           |      | [m]  | [n]  | [ɲ]  | [m]  | [n]  | [ɲ]  | [m]  | [n]  | [ɲ]  |
| Átlag     | 205  | 224  | 245  | 208  | 227  | 234  | 198  | 223  | 233  | 168  |
| Átl. elt. | 24,5 | 44,4 | 47,8 | 47,0 | 54,4 | 45,4 | 47,8 | 43,6 | 42,2 | 34,8 |

### Következtetések

A jelen kutatásban arra a kérdésre kerestük a választ, hogy a fonémaértékű nazális mássalhangzók koartikulációs hatása hogyan realizálódik a két elemzett magyar kardinális magánhangzó akusztikai szerkezetében. Az eredmények azt mutatták, hogy a nazalizáció az [i] esetében a formánsok frekvenciaértékének emelkedésében ragadható meg a réshangkörnyezethez képest. Az [a:] nazalizációjakor differenciáltabb a koartikulációs hatás, ennél a magánhangzónál az első két formáns értéke a bilabiális és az alveoláris nazális környezetében csökken. A két magánhangzó akusztikai szerkezete tehát nem azonos módon és mértékben változott az egyes nazális mássalhangzók hatására (3. táblázat).

3. táblázat: A nazalizáció hatása a magánhangzók akusztikai szerkezetében a teljes nazális környezetben (NVN) – az *igen* a szignifikáns változást, a *nem* a statisztikailag nem igazolható változást jelenti

|     | F <sub>1</sub> |      | F <sub>2</sub> |             | F <sub>3</sub> |             | Időtartam |      |
|-----|----------------|------|----------------|-------------|----------------|-------------|-----------|------|
|     | [i]            | [a:] | [i]            | [a:]        | [i]            | [a:]        | [i]       | [a:] |
| [m] | <b>igen</b>    | nem  | nem            | <b>igen</b> | <b>igen</b>    | <b>igen</b> | nem       | nem  |
| [n] | nem            | nem  | <b>igen</b>    | nem         | <b>igen</b>    | <b>igen</b> | nem       | nem  |
| [ŋ] | nem            | nem  | <b>igen</b>    | nem         | <b>igen</b>    | <b>igen</b> | nem       | nem  |

A nazális mássalhangzók környezetében a magánhangzók első formánsa a legkevésbé érintett a koartikulációs folyamatban, csak a *mim* hangsor magánhangzójának első formánsában mutatható ki szignifikáns változás a réshang-környezethez képest. A második formáns változásában van a legnagyobb különbség a két elemzett magánhangzó között, az F<sub>2</sub> eltérő módon érintett a nazalizációban a magánhangzó és a nazális mássalhangzó minőségétől függően. Mindkét magánhangzónál azt tapasztaltuk, hogy mindhárom fonémaértékű nazális szignifikáns mértékben megváltoztatja F<sub>3</sub> frekvenciaértékét a réshang-kontextushoz képest. A nazalizáció a formánsszerkezetben eredményez változást, a magánhangzók időtartamát nem befolyásolja.

Vajon mi eredményezi azt, hogy a két elemzett kardinális magánhangzó különböző mértékben és más módon változik a nazálisok hatására? A magánhangzók nazalizációjának eltérő megvalósulása a különböző artikulációs konfiguráció következménye. Az [i] képzéshelye közel esik a nazális mássalhangzók képzési helyéhez, ezért a koartikulációs folyamat közvetlenebbül megy végbe, a nazális hatás erősen érvényesül. Az [a:] képzésekor nagyobb az állkapocs nyitásszöge, a magánhangzó képzéshelye távolabb esik a mássalhangzóétól, megnő a koartikulációs „távolság”, a nazális hatás kevésbé érvényesül ezen a magánhangzón. Az [i] akusztikai szerkezete jobban érintett a nazalizációban, és a hatás szabályszerűbben leírható, mint a veláris magánhangzónál. Ez az eredmény hasonló tendenciát mutat ahhoz, mint amelyet az angol nyelvre igazoltak: a felső nyelvvállású magánhangzók akusztikai szerkezete nagyobb mértékben módosul a nazalizáció hatására, mint az alsó nyelvvállású magánhangzóké (Bell-Berti 1993).

A kutatás igazolta, hogy a magyarban a progresszív nazalizáció hatására nagyobb változás következik be mindkét magánhangzó akusztikai szerkezetében, mint a regresszív nazalizáció következtében. A fonológiai koartikulációs folyamatokhoz sorolt képzési mód szerinti hasonulás mellett a nazalizáció a másik progresszív irányban domináló koartikulációs folyamat.

A két kardinális magánhangzó nazalizációjának közös jellemzője, hogy az orrhangú mássalhangzók hatására a harmadik formáns módosul a legnagyobb mértékben. Az F<sub>3</sub> a beszéd egyéni ejtéséről is tartalmaz fontos információkat; a nazális mássalhangzók pedig szintén jellemzik a beszélőt (az orrüreg



alakja, mérete nem változik – csökken az egy beszélőtől való variációk száma). Az orrűreg jellemzői egyénenként eléggé különbözőek lehetnek ahhoz, hogy össze lehessen hasonlítani az egyes beszélőkre jellemző sajátosságokat a nazálisok tekintetében (Rose 2002). A nazálisok és a magánhangzókra gyakorolt hatásuk így jellemzők lehetnek az egyénre – a koartikulációs folyamatok megvalósulása a nazálisok tekintetében újabb támpontot adhatnak a kriminalisztikában a beszélő személy azonosításához.

### Irodalom

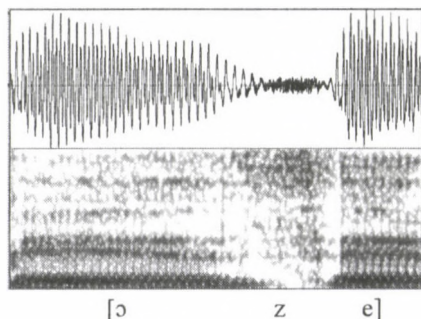
- Beddor, Patrice Speeter 2007. Nasals and nasalization: The relation between segmental and coarticulatory timing. In Trouvain, Jürgen – Barry, William J. (eds.): *Proceedings of the 16<sup>th</sup> International Congress of Phonetic Sciences. Saarbrücken 6–10 August 2007*. Pirrot GmbH., Dudweiler, 249–254.
- Bell-Berti, Fredericka 1993. Understanding velic motor control: studies of segmental context. In Huffmann, Marie K. – Krakow, Rena A. (eds.): *Nasals, nasalization and the velum*. Academic Press, New York, 63–85.
- Chen, Nancy F. – Slifka, Janet L. – Stevens, Kenneth L. 2007. Vowel nasalization in American English: Acoustic variability due to phonetic context. In Trouvain, Jürgen – Barry, William J. (eds.): *Proceedings of the 16th International Congress of Phonetic Sciences. Saarbrücken 6–10 August 2007*. 905–908.
- Delvaux, Veronique – Amelot, Angélique – Rossato, Solange 2007. Nasal Processes at the Interface Between Phonetics and Phonology. In Trouvain, Jürgen – Barry, William J. (eds.): *Proceedings of the 16<sup>th</sup> International Congress of Phonetic Sciences. Saarbrücken 6–10 August 2007*. Pirrot GmbH., Dudweiler, 245–248.
- Földi Éva 1989. Magánhangzó-nazalizáció, hosszú mássalhangzó vagy gemináta? *Egyetemi Fonetikai Füzetek* 2. 98–120.
- Gósy Mária 2004. *Fonetika, a beszéd tudománya*. Osiris Kiadó, Budapest.
- Gósy, Mária – Beke, András 2008. Anticipatory coarticulation in a specific context (megjelenőben).
- Hardcastle, William J. – Hewlett, Nigel 1999. *Coarticulation*. Cambridge University Press, Cambridge.
- Horváth Viktória 2005. A magánhangzók nazalizációjáról. *Beszéd kutatás* 2005. 51–62.
- Ladefoged, Peter 2005. *Vowels and consonants*. Blackwell Publishing, Oxford.
- Malécot, Andre 1960. Vowel nasality as a distinctive feature in American English. *Language* 36. 222–229.
- Nikléczy Péter 2006. A koartikuláció a magánhangzók tiszta fázisa és az átmenetek összefüggésében. *Beszéd kutatás* 2006. 23–31.
- Olaszy Gábor 2006. *Hangidőtartamok és időszervezeti elemek a magyar beszédben*. Nyelvtudományi Értekezések 155. Akadémiai Kiadó, Budapest.
- Pickett, James M. 1999. *The acoustics of speech communication*. Allyn & Bacon, London.
- Rose, Philip 2002. *Forensic speaker identification*. Taylor & Francis, London–New York.
- Whalen, Douglas H. – Beddor, Patrice Speeter 1989. Connections between nasality and vowel duration and height: elucidation of the Eastern Algonquian intrusive nasal. *Language* 65. 457–486.

## ALVEOLÁRIS SPIRÁNSOK AKUSZTIKAI FONETIKAI VIZSGÁLATA

Gráci Tekla Etelka

### Bevezetés

A jelen tanulmányban egy zöngésségi oppozícióban álló magyar spiráns fonémapár spontán beszédbeli megvalósulásainak akusztikai elemzését tűztük ki célul. A kutatás kérdései abból a tapasztalatból adódtak, hogy a zöngésségi oppozícióban álló fonémapárok realizációjában a hangszalagok működése nem minden esetben a fonéma adott fonetikai helyzetben elvárt realizációjának megfelelően történik (1. ábra).



1. ábra

Egy részben zöngétlen /z/-realizáció az *azért* szóban (női ejtés)

(Fent: oszillogram, lent: spektrogram 8000Hz-ig.)

Stevens (1998: 465) a zöngés obstruenseket a következőképp jellemzi: a felpattanó zárhangok és a spiránsok néhány csoportjánál a transzglottális nyomást, a hangszalagok állapotát és konfigurációját az jellemzi, hogy a szájüregben történő akadályképzésnek legalább egy részében hangszalagrezgés történik. A zöngés obstruensek esetén az akadály képzéséig és annak feloldásától van jelen a zöngé – ennek oka egy artikulációs kényszer, amely a hangszalagok működéséből kiindulva írható le. A zöngéképzéshez a beállító izmok bezárják és megfeszítik hangszalagokat, majd ezen zár alatt feltorlódik a tüdőből kiáramló levegő. Amikor a feltorlódott levegőrészecskék elérik a hangszalagok felnyitásához szükséges nyomást, azok felnyílnak, majd a nyomásesőkkedés hatására (a beállító izmok változatlan működése miatt)

visszaállnak a zárt állásba (Tonndorf 1925, idézi Gósy 2004). Annak azonban, hogy ez a rezgés valóban bekövetkezhesen, az a fizikai feltétele, hogy a szupraglottális nyomás alacsonyabb legyen a hangszalagok felnyitásához szükséges szubglottális nyomásnál. Az obstruensek képzése során a szájüregben létrehozott akadály mögött a levegő feltorlódik explozívák esetében a zár megléte, réshangok esetében pedig a réses képzés alatt fenálló szűk keresztmetszet miatt. A levegő feltorlódásának következtében a toldalékcsőbeli nyomás folyamatosan növekszik (Stevens 1998), így a hangszalagműködést egy adott nyomás elérése után megakadályozza. Emiatt Lisker (1977: 305) az explozívákról beszélve az akadályt a következőképp jellemzi: az artikulációs zár egyben zöngétlenítő gesztus is. Lisker ebben a munkájában felhívja a figyelmet arra, hogy ezt a fiziológiás kényszert nem lehet kizárólagos zöngétlenségi tényezőként kezelni, hiszen a fonológiai is zöngétlen fonémák esetében természetesen a hangszalagok beállítása szándékolt (1977: 305). Tehát az izomműködés eltérő a zöngésségi párok tagjai között, ami az artikulációs kényszer hasonló működése miatt a spiránsok esetében is figyelembeveendő. Ennek a hatásnak a feloldására, a zöngésség (tovább) történő fenntartására különböző artikulációs megoldásokat alkalmaznak a beszélők, amelyek következtében a szupraglottális üregrendszer részeinek térfogata megnövelhető, vagyis a hangszalagok feletti nyomás csökkenthető. Ez elérhető a lágy szájpad megemelésével, a pharynx tágításával, a gége megsüllyesztésével és ennek következtében a nyelv felületéhez közeli izmok elernyedésével, továbbá a nyelvgyök és az epiglottisz előrébb mozdításával (vö. pl. Stevens 1998). További lehetőség a spiráns fonéma realizációjára a képzésmód megváltozása lehet, vagyis approximáns jellegű képzés; illetve a rövidebb időtartamú megvalósulás (Maddieson 1999). Mindezen lehetőségek miatt a zöngés obstruensek zöngétlenedésének mértékében (hány százalékuk érintett és milyen arányban) nagy, beszélők közötti variabilitást találtak (pl. Smith 1997, idézi Jesus-Shadle 2003).

Mindezek miatt a spiráns fonéma realizációjának (és az obstruensekének általában) több univerzális vonása és velejárója van, melyeket ezzel az fiziológiás kényszerrel magyarázhatunk (vö pl. Stevens 1998; Maddieson 1999). Ilyen például a zöngétlen hangok zöngés párjukhoz viszonyítva hosszabb időtartamban történő realizációja (Baum–Blumstein 1987 – a szignifikáns eltérés mellett nagy arányú időtartam-átfedést is találtak –; Docherty 1992; Jesus–Shadle 2003; magyar adatokra pl.: Magdics 1966, Kassai 1982), melyre korábban több magyarázatot fogalmaztak meg. A több együttes ok között a zöngésségi oppozíció kiemelése és fenntartása is szerepel (részletesen vö. Maddieson 1999). Ezzel összefüggésben a konsonánst megelőző magánhangzó időtartama hosszabb, ha zöngés (tehát rövidebb) mássalhangzó követi, aminek oka egyfajta hangsorbeli időtartam-kiegyenlítés, vagyis a hasonló szerkezetű hangsor hasonló tartamban realizálódása (Maddieson 1999; magyar nyelvre kapott mérések összefoglalóan vö. Kassai 1982). Ezen univerzá-



lis jellemzők közé tartozik a mikrintonáció szintjén, hogy a zöngétlen obstruensek után az alaphang magasabb, mint a zöngések után. Ennek az az oka, hogy a hangszalagok fűvóállásba történő beállításával azok tapadási felületei távolabb kerülnek egymástól, így a hangszalagok hossza megnyúlik, rezgési frekvenciájuk pedig megemelkedik a zöngéképzés megindulásakor, mivel ezen izmok elernyedése hosszabb időt vesz igénybe (Maddieson 1999). A hangszalagok eltérő beállítása miatt a zöngességi párok tagjai között eltér a zöngétlen rész aránya is. A zöngétlen hang átlagosan nagyobb arányban marad zöngétlen a zöngétlenedett zöngés párjánál, vagyis a megelőző hang lecsengő és a követő hang meginduló zöngéje kisebb időtartamban van jelen esetlegesen az obstruens képzése során (Mair–Shadle 1996, idézi Jesus–Shadle 2003).

Az univerzális jellegzetességek mellett néhány nyelvspecifikus tulajdonság és kérdés is felmerül. Jesus és Shadle a portugál nyelv spiránsainak vizsgálatakor (2003) az egyes kísérleti korpuszok között eltérő tendenciákat adatolt a képzési helynek a zöngésségre gyakorolt hatásában (ti. minél hátrébb képzett, annál hosszabb tartamú a réshang) és a szóbeli pozíció és a zöngétlenedés összefüggésében (ti. a szó eleji réshang ritkábban, szó végi gyakrabban zöngétlenedik). Eltérő eredmények születtek a spiráns realizációjának időtartama és zöngéssége közötti összefüggésre. Crystal és House az amerikai angol spiránsok vizsgálatakor (1988) nem, Jesus és Shadle a portugálról szóló tanulmányában (2003) azonban talált korrelációt, vagyis minél hosszabb a zöngés résfőnéma megvalósulásának időtartama, annál nagyobb arányban zöngétlenedett. Különbözik továbbá az egyes nyelvek között, hogy milyen gyakorisággal és arányban zöngétlenednek a zöngés spiránsok; valamint ellentmondásos eredmények születtek a fonetikai környezet hatásáról (elsősorban az intervokális helyzet magánhangzó-minőségének függvényében – vö. Jesus–Shadle 2003).

A magyar nyelvben főként a /v/ zöngésségi jellegzetességét vizsgálták eddig. Fonetikai szempontból Böhm és Olszky (2007), fonológiai viselkedésének szempontjából fonetikai alapon Kiss és Bárkányi (pl. 2006) elemezték. A vizsgálatok alapján megmagyarázhatóvá vált például a labiodentális spiráns „különös” viselkedése a zöngésségi hasonulásban.

A jelen tanulmányban azt a kérdést tettük fel, hogy a magyar spontán beszédben hogyan realizálódnak a zöngésségi oppozícióban álló alveoláris fonémák. Milyen zöngésségi és időszerkezet jellemzi a megvalósulásukat, illetve ezeket milyen tényezők és hogyan befolyásolják. Gósy (2004) hanggyakorisági adatai alapján választottuk az alveoláris spiráns párokat, mert ezen képzéshelyű rövid spiráns zöngés és zöngétlen változata fordul elő a leggyakrabban a spontán beszédben. A nemzetközi és a magyar szakirodalom eddigi, többnyire olvasott beszédén (szó- és mondatlista) alapuló eredményei alapján állítottuk fel hipotéziseinket. Azt vártuk, hogy a zöngés és zöngétlen alveoláris spiráns párok időtartama szignifikánsan különbözik (a /s/ általában hosszabb időtartamban valósul meg), azonban jelentős átfedést mutat. Feltételeztük, hogy a

lazább artikuláció és az eltérő tervezési-kivitelezési folyamatok miatt nagyobb arányú zöngétlenedés jellemzi a spontán beszédbeli megvalósulásokat, mint a korábbi szakirodalom adatai. Hipotéziseink szerint a zöngétlenedett zöngésfonéma-realizációk zöngétlen része arányában kisebb a zöngétlen fonémák megvalósulásaiénál, míg a zöngésségi hasonulás következtében zöngétlenedett zöngés alveoláris spiránsok esetében a zöngétlen pár realizációjával lényegében azonos szerkezetet vártunk. Korábbi, olvasott szövegen alapuló vizsgálataink alapján (Gráci 2008) azt vártuk, hogy a fonetikai kontextus és helyzet befolyásolja a tartamot és a zöngésséget a beszédhangokban oly módon, hogy a követő hang képzéséhez szükséges függőleges nyelvmozgással függ össze (vagyis a magánhangzóknál a nyelválásfokkal – főként az elől képzettek esetében; a szonorok előtt a vokálisokkal szemben nagyobb arányú, az obstruenseknél pedig a legnagyobb arányú zöngétlenedést vártuk). Továbbá abszolút szakasz eleji helyzetben kisebb, abszolút szakasz végi helyzetben pedig nagyobb arányú zöngétlenedést vártunk a zöngés fonémák realizációiban.

### Anyag és módszer, kísérleti személyek

A kérdések megválaszolásához a „BEA” adatbázis (Gósy 2008) nyolc (négy női és négy férfi) beszélőjének (22–52 évesek) spontánbeszéd-anyagát dolgoztuk fel. A kísérletvezető az adatközlőket munkájukról, illetve tanulmányaikról kérdezte. A felvételek ezen részei 1'25"–5'10" hosszúak, összesen 28'54". A Praat 5.0 verziószámú szoftverrel elemeztük a hanganyagban /z/, /z:/ és /s/, /s:/ realizációinak teljes időtartamát és a zöngétlen rész ehhez viszonyított arányát és elhelyezkedését a képzés folyamatában. Az így kapott adatokon elemeztük, hogyan alakul az alveoláris spiránsok akusztikai szerkezete a fonológiai zöngésség tekintetében; illetve hogyan módosul a fonetikai környezet, illetve a zöngésségi hasonulás hatására. Illetve azon esetekben is, amikor a beszélő egy mássalhangzó-kapcsolatot nem annak alaprealizációjában valósít meg, vagyis a vizsgált spiránst megelőző vagy követő hang ejtése elmarad, illetve azt a spiránssal egy fonetikai egységként artikulálja. Elemeztük a tartam és a képzésmód összefüggését a zöngésséggel, valamint a fonológiai tartam és az alveoláris spiránsok akusztikai szerkezetének kapcsolatát. A statisztikai számításokat (független kétmintás *t*-próba, egytényezős varianciaanalízis, Pearson-féle korreláció – 95%-os konfidenciaszinten) az SPSS 15.0 szoftverrel végeztük azon esetekben, ahol a kapott adatok mennyisége ehhez elegendő volt.

Az eredmények elemzése során a realizálódó hang esetében elvárt hosszúság alapján két kategóriába osztottuk az alveoláris spiránsokat. Az egyik csoport azokat a fonémarealizációkat tartalmazta, amelyek esetében a rövid variánsban való megvalósulást vártuk. Ebbe a csoportba kizárólag olyan hangsorok kerültek, melyben maga a vizsgált fonéma tartama is rövid volt (pl. *beszéd, viszont, korosztály, nagyon szeretek, fizetik, intézethez, közgazdaság, közgazgatás*). A másik csoportba azokat a megvalósulásokat soroltuk, ame-



lyek *a)* alaprealizációja hosszú, és ebben realizálódtak az adott fonetikai környezetben (pl. *össze, hossza* – a zöngés alveoláris spiráns esetében ilyen nem találtunk); *b)* továbbá azokat a realizációkat, melyek két fonéma kapcsolatában koartikulációs hatás következtében egy hosszú réshangként jelentek meg (pl.: *gápszámla, Tacituszal, próbálkozzak, hozzá, azzal*). A két csoport elválasztására a „fonológiai rövid”–„fonológiai hosszú” megnevezést használjuk, az IPA-jelölésekben pedig a ferde – fonémákat jelölő – zárójelpárt, mivel mindkét csoport megvalósulásaiban a zöngésség és az időtartam is változó, így a hang időtartama és zöngéssége is megnehezíti a fonetikai kategóriánév és a jelölés megválasztását.

A zöngétlen réshangok esetében néhányszor előfordult, hogy rövid szünet-szerű részek jelentek meg. Ezek a vizsgált hang előtt, illetve után is előfordulhattak, időtartamuk akár a 30 ms-ot is meghaladta. Ezeknek a magyarázata a hang képzésében keresendő: a magánhangzó-spiráns hangsor képzésében a glottisz gyakran már a szájüregi akadály megformálása előtt szétnyílik. A spirantikus zörej megindulása azonban pontosan nem behatárolható (ti. az akadály megképződésének viszonylatában, vö. Shadle 1999: 44–45). Ebben a szakaszban a spiráns turbulens zöreje többnyire fokozatos intenzitás-változással jelent/szűnt meg, így pontos időtartammal ezt az átmeneti részt nem tudjuk jellemezni. A megelőző, illetve a követő hang zöngéje ezen szakaszban cseng le, illetve indul meg. Ezeket a részeket – mivel már nem a magánhangzóhoz tartoznak, hanem a spiráns képzésmozzanatai indulnak meg – a spiráns időtartamához mértük.

A nazálisokkal alkotott mássalhangzó-kapcsolatok esetében a koartikulációs néma fázist a vizsgált spiráns részeként elemeztük, mivel olyan nazális nem fordult elő a vizsgálati anyagban, mely részben zöngétlenedett volna. (Ebben a szakaszban lényegében a két konzonáns artikulációs konfigurációja közötti váltás zajlik – vö. Stevens 1998: 563–564, Olasz 2007.)

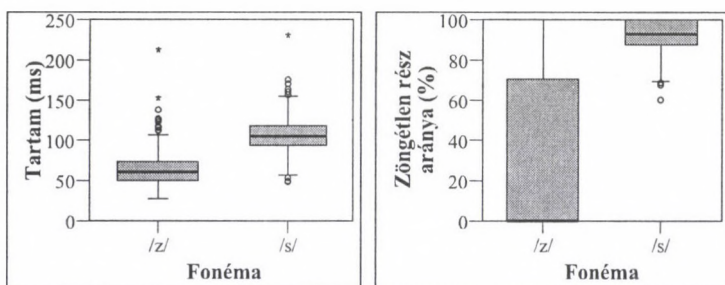
A vizsgált konzonánsokat megelőző vagy követő mássalhangzók képzése néhány mássalhangzó-kapcsolatban részben vagy egészben elmaradtak. Ezeket a realizációkat külön elemezzük.

### Eredmények

A /z/- és /s/-realizációk időtartamukban és a zöngésségi szerkezet megvalósulásában is jelentősen eltérnek egymástól (2. ábra). A zöngés alveoláris spiráns átlagidőtartama 64 ms volt, 27–213 ms közötti szóródással, az esetek közel háromnegyedében 40–75 ms között valósult meg. A zöngétlen réshang pedig 107 ms, 48–230 ms közötti szóródással, az esetek mintegy 75%-ában 75–125 ms között realizálódott. Ez a korábbi adatokhoz hasonlítva valamivel hosszabb hangidőtartamot jelent (Olasz 2006: 119: /z/ 64 ms, 27–149 ms; /s/ 91 ms, 35–165 ms), főként a szóródás tér el. A zöngésségi párok tagjai időtartamuk alapján szignifikánsan különböznek [független kétmintás *t*-próba:  $F(2, 449) = 2,860$ ;  $p < 0,001$ ]. Baum és Blumstein (1987) adataihoz hasonlóan azt



találtuk, hogy a zöngésségi párok az időtartamátlagok alapján elkülöníthetők, ez átlagosan mintegy 30%-nyi eltérést jelent (a korábbi szakirodalom olvasott anyag alapján 20–23%-ot mutatott ki – összefoglalóan vö. Kassai 1982), az átfedés azonban jelentős. Az időtartam az esetek jelentős többségében másodlagos megkülönböztető jegyként szerepelhet a percepció számára a spontán beszédben során is.



2. ábra

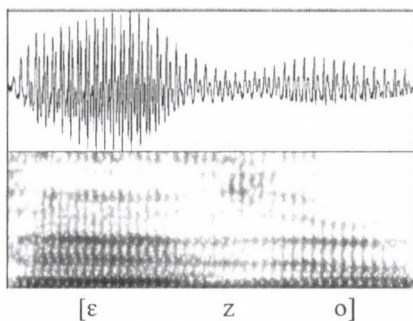
A fonológiai rövid alveoláris spiránsok időtartama (balra) és zöngéssége (jobbra)

A statisztika által extrém hosszú időtartamúnak talált értékek a zöngés spiráns esetében a következőképp alakultak. Az öt leghosszabb adat szakasz végi /z/ fonéma realizációja, egy esetben hezitáció is követi, négy esetben approximánsokkal alkotott mássalhangzó-kapcsolat első tagja, egy esetben pedig intervokális helyzetű. A /s/ esetében egy /s/-sel alkotott konzonáns-kapcsolat második tagja, egy abszolút szakasz végi és egy abszolút szakasz eleji, négy pedig intervokális helyzetű. A szakasz eleji és végi, illetve két intervokális helyzetű realizáció esetében egyértelmű a vizsgált spiráns megakadásjelenség miatti nyújtása. A /s/ esetében extrém rövid időtartamokat is találtunk. Ezek közül kettő /st/, egy pedig /stm/ fonémák mássalhangzó-kapcsolatában volt adatolható.

A két fonéma realizációi a zöngésség tekintetében is jelentős eltérést mutatnak, amint az várható volt. Az összes /z/-realizáció 52,61%-a volt teljesen (tehát időtartamának 100%-ában) zöngés. Időtartama nagy részében, 80–100%-ában pedig 55,11%-uk. A /s/ megvalósulásai 90,96%-ukban zöngétlenek tartamuknak legalább 80%-ában, annak 100%-ában, azaz teljesen zöngétlen pedig csak közel 35,54%-uk volt. Összességében az adatok a zöngés spiráns esetében átlagosan közel az időtartamnak 33%-nyi részében, a zöngétlenek pedig mintegy 92%-ában voltak zöngétlenek. Az eltérés a vártaknak megfelelően szignifikáns [független kétmintás *t*-próba:  $F(2, 449) = 916,817$ ;  $p < 0,001$ ]. A /s/ realizációi között a statisztikai elemzés három extrém hangot jelölt, amelyek a többi megvalósuláshoz képest zöngésebbek. Ezek két pontatlanabb

artikulációjú beszélőtől származnak, feltehetően ez magyarázza a nagyobb mértékű zöngésségüket.

A zöngés alveoláris spiráns esetében a zöngésség és az intenzív réses képzés fenntartása miatt azt vártuk, hogy a hosszabb időtartamú megvalósulások nagyobb arányban fognak zöngétlenedni, míg zöngétlen párjánál nem vártunk ilyen jellegű összefüggést, mivel a képzés eleve zöngétlen, a környező hangok zöngéjének megindulása és lecsengése pedig nem a vizsgált konzonáns időtartamának függvénye. Míg a /z/ esetében  $r = 0,413$ ,  $p < 0,001$ , tehát a korreláció közepesen erős, addig a /s/ realizációira  $r = 0,137$ ,  $p = 0,069$ , vagyis nincs összefüggés. A zöngés spiráns esetében alternatív approximáns jellegű (tehát magánhangzós szerkezetű zöngés) hangokat is adatoltunk (3. ábra), amelyek megjelenése leggyakrabban intervokális helyzetben volt jellemző, de szonoránssal és obstruenssel alkotott konzonánskapcsolatokban is előfordult [Magdics 1965-ös mássalhangzó-elemzésében már megemlíti, hogy a zöngés spiránsok néhány esetben magánhangzószerű szerkezettel realizálódnak (57)]. Ez a jelen anyagban a teljes tartományban zöngésen realizálódott rövid zöngés alveoláris spiránsok (148 db) 35,81%-ára volt jellemző (53 előfordulás). (A nyolc teljesen zöngésen megvalósult /z/ közül kettő esetében is adatolhattunk approximáns jellegű megvalósulást.) A teljesen zöngésen megvalósult rövid alveoláris zöngés spiránsok képzésmódjának függvényében nem találtunk szignifikáns eltérést az időtartam eloszlásában, az approximáns realizációk átlagos tartama 53,60 ms, a réses képzésűeké 56,52 ms volt [független kétmintás  $t$ -próba:  $F(2, 146) = 0,042$ ;  $p = 0,258$ ].



3. ábra

A /z/ approximáns képzésű megvalósulása a *szezonális* szóban (női ejtés)  
(Oscillogram és spektrogram 8000 Hz-ig.)

Elemeztük, hogy az alveoláris spiránsok szerkezetét hogyan befolyásolja a fonetikai környezet. Ez esetben a környező hangokat a következőképpen csoportosítottuk (1. táblázat): magánhangzó, szonoráns, obstruens, nincs megelőző/követő hang (tehát a vizsgált alveoláris spiráns szakasz eleji vagy végi

helyzetű). A csoportosítás során a realizálódott hangsort vettük figyelembe, azaz amennyiben egy környezeti fonéma nem alaprealizációjában jelent meg, akkor az adott realizáció kategóriájához soroltuk. Ez elsősorban az /r/-rel alkotott konzonánskapcsolatokat érintette, mivel az többször réses képzéssel realizálódott.

1. táblázat: Az alveoláris spiránsok előfordulása, tartama és zöngétlen részének aránya a fonetikai helyzet függvényében  
(C<sub>o</sub> = obstruens, C<sub>sz</sub> = szonoráns, \_ = alveoláris spiráns)

|                                 | /z/              |                  |                    | /s/              |                  |                    |
|---------------------------------|------------------|------------------|--------------------|------------------|------------------|--------------------|
|                                 | Előfordulás (db) | Átl. tartam (ms) | Átl. ztl. rész (%) | Előfordulás (db) | Átl. tartam (ms) | Átl. ztl. rész (%) |
| #_V                             | 3                | 58,69            | 77,73              | 9                | 118,49           | 96,52              |
| V_V                             | 189              | 60,34            | 32,06              | 84               | 110,01           | 92,01              |
| V_C <sub>sz</sub>               | 22               | 72,81            | 21,10              | 6                | 100,20           | 96,62              |
| V_C <sub>o</sub>                | 11               | 52,08            | 7,02               | 21               | 85,55            | 89,37              |
| V_#                             | 8                | 124,54           | 77,28              | 2                | 174,13           | 100,00             |
| V_ə                             | 3                | 106,60           | 14,64              | 0                | –                | –                  |
| C <sub>sz</sub> _V              | 12               | 63,90            | 40,52              | 15               | 110,67           | 88,41              |
| C <sub>sz</sub> _#              | 0                | –                | –                  | 1                | 90,93            | 69,29              |
| C <sub>o</sub> _V               | 19               | 68,38            | 44,02              | 14               | 99,99            | 95,78              |
| C <sub>o</sub> _C <sub>sz</sub> | 0                | –                | –                  | 1                | 114,34           | 95,50              |
| ə_V                             | 2                | 56,44            | 0,00               | 3                | 106,97           | 90,40              |

Az alveoláris spiránsok realizációinak átlagos időtartamát és a zönge nélkül képzett rész arányát a fonetikai környezet függvényében az 1. táblázat foglalja össze. Mindkét esetben a legrövidebb átlagos realizációt a magánhangzó és obstruens közötti megvalósulásokra kaptuk (pl. *Kosztolányi, kosztium, közgazgatás, közgazdaság*), a leghosszabb tartamban pedig a magánhangzót követő abszolút szakasz végi helyzetben (pl. *blúz#, kész#*). A /z/-megvalósulások esetében a V\_ə szakasz végi hangok (pl. *ez [ə]#*) esetében található még kiemelkedően (a legrövidebb időtartamnak mintegy kétszerese) hosszú időtartam. A konzonánskapcsolaton belüli sorrendiség mindkét esetben eltérő hatást mutat a két mássalhangzó-kategória függvényében, illetve a két vizsgált mássalhangzó esetében hatásuk is eltérő irányú. Meghatározó azonban a követő hang obstruens vagy szonoráns volta. Mindkét esetben az alveoláris spiráns hosszabb átlagos időtartamban realizálódott, amennyiben a követő konzonáns szonor volt (pl. *közlekedés, felvételizni, használ, Propertius Monobibloszáról*). A /s/ esetében mintegy 15%-kal, a /z/ esetében pedig közel 28%-kal hosszabb átlagidőtartamot adatoltunk, mint a követő obstruens esetében.

A továbbiakban azon típusú helyzeteket elemezzük, melyekből legalább 10 előfordulást tudtunk adatolni a hanganyagban. A legkevésbé a magánhangzó

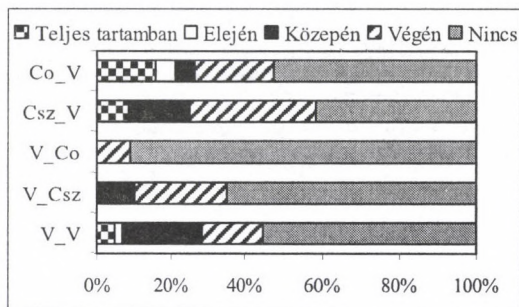


és obstruens közötti helyzetű /z/-realizációk (pl. *közgazdaság*, *kézbésít*) zöngétlenedtek (7,02%), a szonoráns által követett /z/ (pl. *az nagyon*) képzése már közel háromszor nagyobb arányban zöngétlen (21,10%), az intervokális helyzetű (pl. *az adóhivatalnál*, *ezen*) pedig tartamának ismét nagyobb arányában (másképpen hosszabban: 32,06%). A /z/-realizációk esetében nem mutatott jelentős hatást a megelőző konszonáns típusa. Mind az obstruensok (pl. *képzéseket*, *rögzít*, *végzem*), mind a szonorok esetében (pl. *elemzés*, *túlzottan*) ismét mintegy egy harmaddal nagyobb arányú (40,52%, ill. 44,02%) zöngétlenedést mértünk. A szakasz eleji és végi helyzetre csak tendenciákat állíthatunk fel. Az abszolút szakasz eleji és végi helyzetben mértük a legnagyobb arányban (> 70%) zöngétlenedett hangokat; amennyiben svá előzte meg, illetve követte azokat, a képzés nem látszik nagy mértékben zöngétlenedni (0%, ill. 14,64%). Ennek magyarázata feltehetően az, hogy a szakasz elején a zöngéképzés megindulásához hosszabb időre van szükség, amennyiben azonban már sváképzés történt, illetve egyéb zöngés hang megelőzte a vizsgált mássalhangzót, a zöngé már megindult, így a képzés során a zöngének leállni majd újraindulni kell a zöngétlenítéshez.

Elemeztük, hogy a vizsgált konszonánsokban a zöngétlenedés a képzésnek mely szakaszát érinti, és ez a fonetikai környezettel milyen összefüggést mutat. A 4. ábrán a /z/-realizációk zöngésségi szerkezetét ábrázoltuk. Az V\_V, a V\_C, illetve a C\_V helyzetű /z/-realizációk leggyakrabban teljes tartamukban zöngések. Amennyiben a vizsgált zöngés spiráns magánhangzó (pl. *kézbésít*, *az van*, *század*) vagy obstruens előzi meg (pl. *képzés*, *végzettség*), gyakoribb a teljesen zöngés realizáció, a követő hang azonban nagyobb hatást gyakorol rá. Követő obstruens esetén a /z/ (pl. *közgazdaság*, *közben*) az esetek 90,91%-ában maradt teljesen zöngés, szonor esetén (pl. *elhozni*, *az nagyon*) mintegy egyharmaddal kevesebb esetben (65,00%), V/z/V helyzetben pedig csupán 55,44%-ban (pl. *tanszékvezető*, *nehezebb*). Intervokális helyzetben a zöngétlenedés leggyakrabban (21,76%) a spiráns középső szakaszát érintette, míg mássalhangzó-kapcsolat esetében a képzés végét. A szakasz eleji realizációk esetében, amennyiben nem kitöltött szünet előzte meg a vizsgált hangot, a képzés elején, végén, illetve teljes tartamában találtunk zöngétlenedést. A teljes tartamban és a képzés elején a zöngétlenség oka nyilván a zöngéképzés megindításának késése, a képzés végének pedig a zöngé és a nagy intenzitású spiráns egyidejű artikulációjának összetettsége lehet. A kitöltött szünet utáni megvalósulás (pl. [ə] *zavarólag*) teljesen zöngés maradt, ami az artikulációs konfigurációk alapján feltehetően az intervokális realizációkhoz hasonlóan magyarázható.

A hezitációt megelőző szakasz végi /z/-k esetében (pl. *ez* [ə]#) a /z/ realizációja nem egyértelműen kapcsolható sem az intervokális, sem a V\_# előfordulásokhoz (pl. *blúz*#). Időtartamában (átl.: 16,60 ms) az abszolút szakasz végi (átl.: 124,54 ms) megvalósuláshoz áll közelebb (intervokális átl.: 60,34 ms), a zöngésség tekintetében (átl. zöngétlen rész: 14,64%) azonban inkább a V\_V

csoporthoz (átl. zöngétlen rész: 32,06%) sorolható (V\_# átl. zöngétlen része: 77,28%).



4. ábra

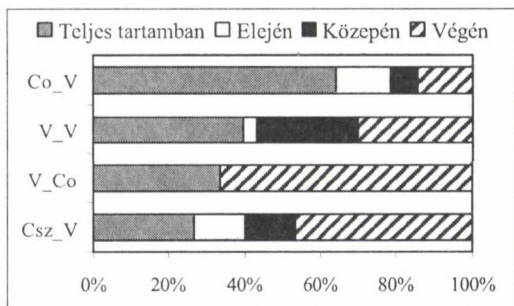
A /z/-realizációk zöngétlenedett szakaszának elhelyezkedése a képzés folyamatában a fonetikai kontextus függvényében (C<sub>o</sub> = obstruens, C<sub>sz</sub> = szonoráns, \_ = alveoláris spiráns.

Az ábra a legalább 10 adat alapján kapott eredményeket tartalmazza.)

A /s/-realizációk esetében a képzés a fonetikai környezettől lényegében függetlenül minden esetben átlagosan közel 90%-os zöngétlenséget mutatott. Az egyetlen 69,29%-os „átlag” egy adat eredménye. A /s/-ek esetében a kismértékű zöngésedés oka, hogy a megelőző hang zöngéjének lecsengése vagy a követő hang képzéséhez a zöngé megindulása rövid tartamban a /s/ képzési ideje alatt még/már történik. Ennek következtében azt vártuk, hogy a megelőző vagy követő zöngés (magánhangzó és szonoráns) hangok hatásaként zöngés hang után a /s/ képzésének kezdete, követő zöngés hang esetén pedig végső szakasza tartalmaz rövid idejű zöngét, követő obstruens esetében pedig minden esetben zöngétlen a képzés (mivel ez esetben csak a zöngétlen obstruensek jöhettek szóba a zöngésségi hasonulás eseteit kizárandó). Feltételezésünk azonban nem igazolódott (5. ábra). A lecsengő zöngé rövid idejű egyidejűsége a /s/ képzési idejével igazolódott, azonban a meginduló zöngé-képzés csak #/s/V- és C<sub>obstruens</sub>/s/V-szekcióra volt érvényes. Magánhangzó és szonoráns, illetve szonoráns és magánhangzó, valamint intervokális helyzetben a feltételezés alapján a részben zöngés realizációk között a képzés közepén várt zöngétlenség helyett minden esetben gyakrabban vagy közel azonos arányban a képzés vége volt zöngétlen.

A nemzetközi szakirodalom leggyakrabban az intervokális helyzetű spirán-sok szerkezetét elemzi, illetve hogy a környező magánhangzótípusok milyen hatást gyakorolnak rá. Ezért a magánhangzók közötti alveoláris réshangokat elemeztük a magánhangzó-minőség tekintetében. Mivel azonban a kísérleti anyag spontán beszéd, nem kiegyenlített arányú a fonetikai környezet. Csak

azokat az intervokális helyzeteket vettük figyelembe, amelyek legalább 6 előfordulásban szerepeltek (2. táblázat).



5. ábra

A /s/-realizációk zöngétlenedett szakaszának elhelyezkedése a képzés folyamatában a fonetikai kontextus függvényében (C<sub>o</sub> = obstruens, C<sub>sz</sub> = szonoráns, \_ = alveoláris spiráns.

Az ábra a legalább 10 adat alapján kapott eredményeket tartalmazza.)

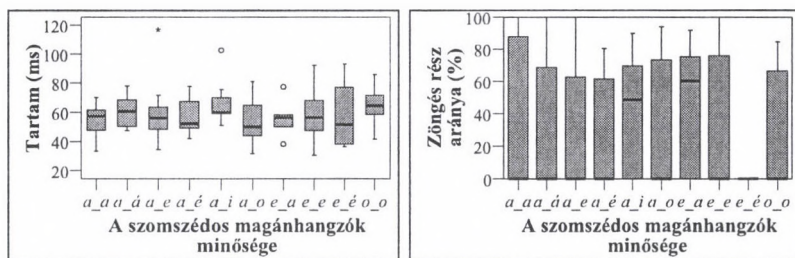
2. táblázat: A legalább hat előfordulású intervokális alveoláris spiránsok fonetikai környezetének eloszlása (db)

|        | /z/ | Példa               | /s/ | Példa                 |
|--------|-----|---------------------|-----|-----------------------|
| /ɔ ɔ/  | 10  | <i>az a</i>         | 7   | <i>esztétika szak</i> |
| /ɔ a:/ | 13  | <i>igazából</i>     | –   |                       |
| /ɔ e/  | 15  | <i>az első</i>      | –   |                       |
| /ɔ e:/ | 15  | <i>azért</i>        | –   |                       |
| /ɔ i/  | 7   | <i>az ilyen</i>     | –   |                       |
| /ɔ o/  | 12  | <i>azonosítás</i>   | –   |                       |
| /e ɔ/  | 8   | <i>ez abba</i>      | –   |                       |
| /e e/  | 27  | <i>kifejezetten</i> | –   |                       |
| /e e:/ | 7   | <i>jelentkezési</i> | 17  | <i>beszéd</i>         |
| /e: e/ | 0   | <i>intézethez</i>   | 10  | <i>természetben</i>   |
| /i o/  | 0   | <i>bizony</i>       | 6   | <i>iszonyat</i>       |
| /o o/  | 15  | <i>halmozottan</i>  | –   |                       |

A /s/ esetében a fonetikai kontextus megoszlása nem elegendő következtetések levonásához, ezért a magánhangzó-minőség hatását annak szerkezetére nem elemezzük. A /z/ esetében (6. ábra) a legnagyobb arányú eltérés mintegy 20%, az /azi/ (tartam: 67,46 ms; átl. eltérés: 17,19 ms) és /azo/ (tartam: 53,86 ms; átl. eltérés: 15,50 ms) között adódott. Ezekben a fonetikai kontex-



tusokban a megelőző hang azonos, a követő hang azonban nyelválásfokban, kerekítettségben és a nyelv vízszintes helyzete alapján is eltér, ezért vessük össze az /ɔzɔ/ és /ɔze/ környezettel. Ezek az előző két hangkörnyezettől páronként nyelválásfokban térnek el. Az /ɔzɔ/ esetében 55,09 ms-os (átl.elt.: 10,69 ms), az /ɔze/ esetében pedig 58,16 ms-os átlagtartamot (átl. elt.: 19,57 ms) kaptunk. A palatális követő hangok esetében a zártabb képzésűt megelőzve a vizsgált spiráns hosszabb tartamot mutat (mintegy 14%-kal), míg a veláris képzésűeknél a nyíltabb esetében ez az eltérés csupán 3%. Hasonló tendencia mutatkozik az /eze/ (58,36 ms, átl. elt.: 15,15 ms) és /eze/ (61,40 ms; átl. elt.: 22,51 ms) hangok között. Az eggyel magasabb nyelválásfokú környezetben (ennél azonban a fonológiai kvantitás is eltér!) mintegy 5%-kal hosszabb a zöngés spiráns realizációja. Tisztán kerektségi, illetve a nyelv vízszintes helyzete szerinti oppozíciót nem találunk ezen kontextusok között. Az /ɔzɔ/ (55,09 ms; átl.elt.: 10,69 ms), /ɔze/ (58,16 ms; átl. elt.: 19,57 ms), /eze/ (58,36 ms, átl. elt.: 15,15 ms) és /ezo/ (55,69 ms; átl. elt.: 10,99 ms) kontextusokra kapott adatokat elemezve azt mondhatjuk, hogy amennyiben a vizsgált konszonánst /ɛ/ követi, mintegy 5%-kal hosszabb, mint amennyiben /ɔ/ követi – függetlenül a megelőző hang /ɔ/, ill. /ɛ/ voltától. A követő magánhangzó minősége tehát nagyobb hatással van a /z/ realizációjának időtartamára.



6. ábra

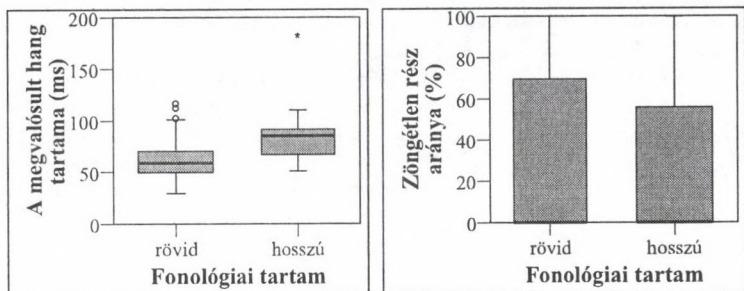
Az intervokális zöngés alveoláris spiránsok időtartama (balra)  
és zöngétlen részének aránya (jobbra)  
a megelőző és követő magánhangzók minőségének függvényében

A zöngés spiránsoknak több mint 56%-a teljesen zöngés marad, a tartamnak legfeljebb 20%-ában pedig összesen 57,67%-uk zöngétlenedik, míg közel teljesen (legalább a tartam 80%-ában) pedig az összes ezen helyzetben szereplő zöngés spiráns 13,23%-a. A zöngétlen rész teljes időtartamhoz viszonyított arányában a legnagyobb eltérést az /ezo/ (átl.: 45,42%; átl. elt.: 39,26%) és /eze/ (átl.: 9,72%; átl. elt.: 25,74%) magánhangzók közötti realizációk mutatták, ami közel 79%-ot jelent. A két fonetikai környezetben ismét a követő hang tér el, és ismét minden képzésjegyben (nyelválásfok, kerekí-

tettség, palatalitás és fonológiai kvantitás). Ismét az /ɔzɔ/ (átl.: 32,97%; átl.elt.: 45,46%), /ɔze/ (átl.: 29,00%; átl. elt.: 37,88%), /eze/ (átl.: 33,25%, átl. elt.: 40,45%) és /eɔz/ (átl.: 45,42%; átl. elt.: 39,26%) kontextusok összevetése alapján elmondható, hogy követő /e/ esetén kevésbé zöngés a képzés, ez közel 12 és 29%-nyi eltérést jelent, míg a hasonló magánhangzók között szereplő spiráns közel azonos arányban volt zöngétlen, vagyis a megelőző és a követő magánhangzó hatása feltételezhetően hasonló arányú. A fonológiai kvantitás és a nyelv vízszintes mozgásának befolyásoló szerepét a zöngésség tekintetében sem tudjuk elemezni, azonban a nyelvállásfok hatásáról az /ɔzɔ/ (átl.: 32,97%; átl. elt.: 45,46%) és /ɔzo/ (átl.: 26,81%; átl. elt.: 39,96%), illetve /ɔze/ (átl.: 30,54%; átl. elt.: 34,58%) és /ɔzi/ (átl.: 39,68%; átl. elt.: 39,22%) hangkörnyezetekre kapott adatok összevetésével elmondható, hogy míg a veláris kerekített pár esetében az alacsonyabb nyelvállásfokút megelőző /z/ kisebb arányban zöngétlenedik, a palatális réses magánhangzópár esetében az arány fordított. Felmerülhet, hogy a palatális hangok esetében azért a magasabb nyelvállásfokú vokális előtt nagyobb arányú a zöngétlenedés, mert ez esetben a vizsgált spiráns képzésében részt vevő nyelvterület magasabb állásban való maradása miatt a szájüregi nyomás csak lassabban csökkenhet. További befolyásoló tényező lehet a palatális magánhangzók esetében a fonológiai kvantitás is.

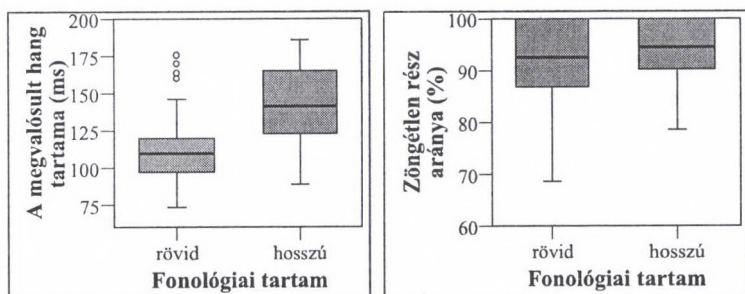
A realizálandó variáns hosszúságának függvényében a megvalósulások jelentősen eltértek az artikulációs-akusztikai tartamban is, az átfedés azonban jelentős a hosszúsági párok között (7. és 8. ábra). Mind a /z/, mind a /s/ esetében szignifikánsan hosszabb a hosszú realizációk időtartama [független kétmintás *t*-próba: /z/:  $F(2, 206) = 5,544, p = 0,007$ ; /s/:  $F(2, 113) = 7,460, p < 0,001$ ]. Míg a /z/-k időtartama átlagosan mintegy 60 ms (29–116 ms), addig a /z/-ké 86 ms (50–182 ms). A /s/ és /s:/ esetében ez az eltérés kisebb: mintegy 110 ms (73–175 ms) a fonológiailag rövid, 140 ms (88–185 ms) a fonológiailag hosszú zöngétlen alveoláris spiránsok átlagos időtartama. Az eltérés a korábbi adatoknál kisebb, Kassai (1982) mondatfelolvasáson alapuló tanulmányában 1 : 1,9-es arányt talált a hangsor belseji hosszúsági párok között. Az eltérés oka feltehetően a két beszédtypus tervezési és artikulációs pontossági eltéréseiből adódik. Feltételezésünkkel ellentétben a hosszú fonémák realizációinak időtartama kisebb eltérést mutat a zöngésségi párok között (intervokális rövid spiránsok: 45,46%; intervokális hosszú spiránsok: 38,57%). A zöngétlen rész arányának tekintetében feltételezéseinknek megfelelően nem található szignifikáns eltérés – /z/:  $F(2, 206) = 4,849, p = 0,268$  (32%, ill. 21%); /s/:  $F(2, 113) = 4,029, p = 0,204$  (92%, ill. 94%).

Elemeztük a zöngésségi hasonulás során zöngétlenedő /z/-realizációk akusztikai szerkezetét (9. ábra). (A vizsgálati anyagban nem jelent meg hasonlóan következőben zöngés variánsú /s/-előfordulás.) A hasonulást mint fonológiai koartikulációt a beszéd tervezési, és nem pedig artikulációs szakaszában bekövetkezőként tartják számon (Gósy 2004). A beszédképzés során azonban



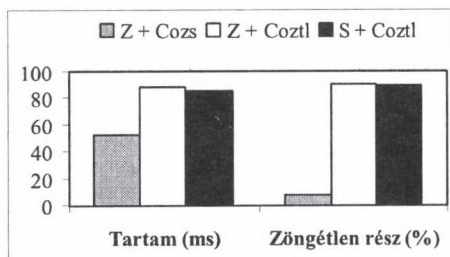
7. ábra

A zöngés alveoláris spiránsok időtartama (balra) és zöngétlen részének aránya (jobbra) a fonológiai tartam függvényében



8. ábra

A zöngétlen alveoláris spiránsok időtartama (balra) és zöngétlen részének aránya (jobbra) a fonológiai tartam függvényében



9. ábra

Az alveoláris spiránsok realizációja magánhangozó és obstruens közötti helyzetben a zöngésségi hasonulás függvényében  
(Co = obstruens, zs = zöngés, ztl = zöngétlen, Z = /z/, S = /s/.)



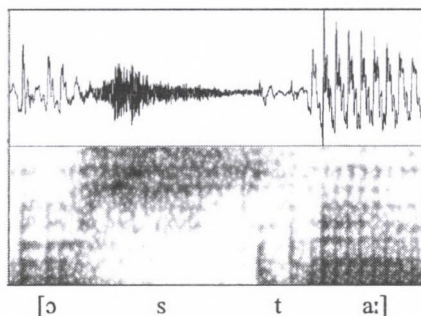
előfordulhat, hogy a zöngösségi hasonulás az elvárttól eltérően elmarad vagy csak részben történik meg, esetleg progresszív irányban hat (vö. pl. Adda-Decker–Hallé 2007, Bóna–Grácz–Markó 2008). Ez a magyar beszédben mintegy 10–15%-ban jellemző (Bóna–Grácz–Markó 2008). Felmerült ezért a kérdés, hogy a /z/ zöngösségi hasonulási helyzetben milyen realizációkat mutat. A vizsgálati anyagban a zöngétlen obstruens és követő /z/ kapcsolatában nem volt adatolható progresszív zöngösségi hasonulás. Míg a zöngés obstruenseket megelőző /z/-realizációk átlagosan 52,77 ms időtartamúak, addig a zöngétlen obstruensek előtti /s/ 88,55 ms-nyí. A zöngösségi hasonulás következtében zöngétlenedő /z/-k 88,25 ms átlagos időtartamban jelentek meg. A zöngétlen rész aránya az alapvariánszt feltételező /z/-k esetében 7,72% (tizenegy előfordulásból összesen egy volt részben – 77,24%-ban – zöngétlen), míg a koartikuláció hatására zöngétlenedők minden esetben a tartamuk több mint felében (50,40%), átlagosan 90,05%-ban. Az összes előfordulás (59 db) több mint 86%-a legalább az időtartamának 80%-ában zöngétlen volt. A zöngétlen alveoláris spiráns mindegyik realizációja (21 előfordulás) legalább 60%-ban (60,17%) zöngétlen volt, az esetek több mint 90%-ában 80% feletti időtartamban voltak zöngétlenek. Ez átlagosan 89,37%-ot jelent. Olvasott francia mondatok alapján hasonló eredményeket kapott Darcy és Kügler (2007) a felpattanó zárhangok vizsgálatában. A hasonulás miatt zöngétlenedett hangok átlagosan 13%-ukban voltak zöngések, tehát 87%-uk volt zöngétlen, a zöngétlen fonémák alaprealizációi pedig időtartamuk 15%-ában voltak zöngések, azaz 85%-ban zöngétlenek.

Elemeztük azon konzonánskapcsolatokat, melyek esetében az ejtés során a két fonéma (az alveoláris spiráns és a megelőző vagy követő konzonáns) egyetlen beszédhangként vagy nem egyértelműen szegmentálhatóan realizálódott (3. táblázat). A nehezen szegmentálható megvalósulás elsősorban az explozívakkal alkotott kapcsolatok esetén jelentkezett. Ekkor a zárképzés elmaradása, de a felpattanás megléte volt jellemző. A 10. ábra egy /zt/ fonémakapcsolat realizációját mutatja be példaként. Ebben a megvalósulásban a [s] hang spirantikus zöreje és a [t] hang (két) zár-felpattanása jól elkülöníthető, azonban a [t] zárszakasza nem elemezhető az ábrán. Mind a zöngés, mind a zöngétlen alveoláris spiráns esetében a követő homorgán felpattanó zárhang esetében következett be a legtöbb olyan megvalósulás, mely során az adott explozívának csak a zár-felpattanása valósult meg. A zöngés részfónémával alkotott kapcsolatokban a zöngösségi hasonulás minden esetben bekövetkezett, azok átlagosan a tartam mintegy 91,25%-ban zöngétlen hangként realizálódtak. A követő explozíva zárszakaszának elmaradását feltehetően a realizálandó hangok nagymértékű hasonlósága (homorganitás) okozza. Amikor a beszélő nem törekszik gondozott beszédre (3 beszélőnél adatolhattuk ezen előfordulásokat), a réshanghoz megképzett rés fokozatosan alakul át a követő zárhang zárjává, esetleg a rés szűkül, de a teljes zár megképzése elmarad; majd – a zár megléte esetén – a réshang miatti, már kellően magas szájüregbeli nyomás mi-

att a zár szinte azonnal felpattanhat; esetleg a feloldás módja is fokozatos, részes lehet. Mindezeket alátámasztja a realizálódott hang(ok) időtartama is, 97–174 ms-ban, vagyis egy hosszú hang és egy konzonánskapcsolat tartamának megfelelő hosszúságban valósultak meg. A nehéz szegmentálhatóság érintette a más spiránsokkal alkotott kapcsolatokat, azonban kisebb arányban. Egy /zfr/, illetve egy /s/ konzonánskapcsolat esetében a két réshang spirantikus zöreje nem volt egyértelműen elválasztható, mintegy átmenet jellegű szakasz volt felfedezhető a hangszínekben. Itt feltehetően a képzési helyek közötti fokozatos váltásról lehet szó.

3. táblázat: A nem két teljes hangként realizálódó konzonánskapcsolatok előfordulási gyakorisága az adott típus összes példányát tekintve 100%-nak („1 + ½”: a kapcsolódó C részben megjelenik; „1”: a kapcsolódó C nem realizálódik.)

|                  |          | /z/ + C    | C + /z/ | /s/ + C    | C + /s/    |
|------------------|----------|------------|---------|------------|------------|
| <b>Obstruens</b> | összesen | 23 (100%)  | 19      | 23 (100%)  | 22 (100%)  |
|                  | „1 + ½”  | 9 (39,13%) | –       | 1 (4,35%)  | 5 (22,73%) |
|                  | „1”      | 3 (13,04%) | –       | 2 (8,70%)  | 2 (9,09%)  |
| <b>Szonoráns</b> | összesen | 22         | 12      | 25 (100%)  | 24 (100%)  |
|                  | „1 + ½”  | –          | –       | –          | –          |
|                  | „1”      | –          | –       | 3 (12,00%) | 8 (33,33%) |



10. ábra

A /zt/ fonémakapcsolat megvalósulása az *aztán* szóban (férfi ejtés)  
(Fent: osszcillogram, lent: spektrum 8000 Hz-ig.)

Azon konzonánskapcsolatok, melyekben a kapcsolódó hang ejtése teljesen elmaradt, elsősorban a képzési hely alapján jellemezhetőek. A /z/ esetében csupán a követő /d/ és megelőző /r/ fonéma ejtése maradt el összesen 3 alkalommal, a /s/ esetében azonban változatosabb képet kapunk. A homorgán szonoránsok és obstruensok ejtése is elmaradhatott a vizsgált hanganyagban. Összesen négy esetben maradt el nem homorgán képzésű hang ejtése



[ʃ, j, ʒ, ɲ]. Ezek közül kettő kötőszó és /s/-szel kezdődő szó kapcsolatában fordult elő (*és szépigetni, hogy száz*). A kötőszók ejtésében gyakran elmarad az utolsó konzonáns ejtése (vö. pl. a *hogy* realizációjáról Tóth–Kocsor 2003). A kapcsolódó hang elmaradása esetében egy hangnyi időtartamban realizálódott alveoláris spiránsok jelentek meg (/z/: 63–84 ms; /s/: 95–147 ms).

Ezen eredmények alapján azt mondhatjuk, hogy a zöngétlen alveoláris részhang (tehát akár /z/, akár /s/ realizálódik [s]-ként) esetében marad el jellemzően a szomszédos konzonáns képzésének része vagy egésze, és ez elsősorban a homorgán képzéshelyű hangokat érinti.

### Következtetések

A nemzetközi és a magyar szakirodalom részletesen foglalkozik az obstruenek zöngétlenedésével, és a zöngés fonéma realizációjának vizsgálatával, illetve a zöngétlen pár összevetésével. A jelen tanulmányban a magyar spontán beszédbeli megvalósulásokat és az ezeket befolyásoló tényezőket elemeztük.

Az alveoláris spiránsok időtartamadatokai részben eltértek a korábban a magyar nyelvre kapott eredményektől, aminek egyik okaként a vizsgálati anyag különbözősége (olvasott vs. spontán beszéd) feltételezhető. A zöngésségi fonémapárok mind tartamuk, mind zöngésségük tekintetében a meglévő átfedések ellenére jól elkülöníthetőek. A /z/ megvalósulásai akár 0–100% közötti zöngétlenedést mutattak, azonban az előfordulások felében teljesen zöngések voltak (ez a portugál adatoknál jelentősen alacsonyabb arány, Jesus és Shadle [2003] mérései alapján a portugál nyelvben az előfordulások 74–78%-a érintett), és átlagosan a teljes időtartamnak mintegy harmada, míg a /s/ esetében több mint 90%-a zöngétlen. A legzöngésebb /s/-realizáció is több mint 60%-ban zöngétlen. Az összes /z/-előfordulásnak mintegy a negyede mutatott 60% feletti zöngétlenedést. A fonológiai kvantitás nem okozott szignifikáns zöngésségbeli eltérést a fonémapár egyik tagjánál sem. A zöngés spiránsnál azonban tendenciaszerűen kevésbé zöngés a pár fonológiaiilag hosszú tagja, aminek oka a hosszabb időtartam alatt kevésbé fenntartható zöngésség. Ezt igazolja a rövid konzonánsokra kapott tartam-zöngésség korreláció. Míg a /s/ esetében nincs összefüggés a két tényező között, addig a /z/ várhatóan annál nagyobb arányban zöngétlenedik, minél hosszabb.

A fonetikai helyzet és környezet hatással van a vizsgált konzonánsok realizációjára. A legnagyobb arányban zöngétlen realizációkat mindkét fonéma esetében az abszolút szakasz eleji, illetve végi helyzetekben kaptuk, ami a /z/ esetében a korábbi szakirodalmi adatoknak ellentmond (vö. pl. Jesus–Shadle 2003), mivel a szakaszeleji spiránsok esetében találták a legkevesebb zöngétlenedő megvalósulást. A zöngés részfonéma esetében – a korábbi, olvasott mondatokon alapuló vizsgálataink alapján (Grácz 2008) – azt feltételeztük, hogy a szakasz belseji hangok zöngésségének mértékét a megelőző és – főként – a követő hang befolyásolja, mégpedig oly módon, hogy minél magasabb nyelvhelyzetet igényel a kontextusban szereplő adott hang képzése, an-



nál nagyobb arányban lesz a /z/ megvalósulása zöngétlen, mivel a toldalékcsőbeli nyomás megnövekedése gyorsabb, illetve csökkenése lassabb. A követő hangok függvényében nem igazolódott a hipotézisünk. A megelőző hangtípus hatása a következőképpen alakult. A legkisebb zöngétlenedést az intervokális, a legnagyobbat pedig az obstruens előtti helyzetben kaptuk. Vagyis, a megelőző hang esetében igazolódott a feltételezésünk. A követő palatális magánhangzók esetében a magasabb nyelvéllásfokkal magasabb zöngétlenségi arány járt, ami a nyomás lassabb csökkenésével lehetne indokolható, azonban a mássalhangzós környezetekre kapott adatok ezt a magyarázati lehetőséget megkérdőjelezzik: azok alapján a megelőző magánhangzó minőségének kellene nagyobb befolyásoló szereppel bírnia. A megelőző vokálisok esetében azonban nem beszélhetünk (a jelen vizsgálat szempontjából lényeges) különbségekről, ezért lehetséges, hogy ez esetben a követő hang játszik jelentősebb szerepet. A veláris követő magánhangzók esetében ellentétes irányú hatást tapasztaltunk. Az eltérő tendencia egyik oka lehet, hogy az elől képzett magánhangzók esetében a nyelvnek az elülső része érintett nagyobb mértékben, amely az alveoláris képzésben is jelentős szereppel bír.

### Irodalom

- Adda-Decker, Martina – Hallé, Pierre 2007. Bayesian framework for voicing alternation and assimilation studies on large corpora in French. In Trouvain, Jürgen – Barry, William J. (eds.): *Proceedings of the 16th International Congress of Phonetic Sciences. Saarbrücken 6–10 August 2007*. Dudweiler, Saarbrücken, 613–617.
- Bárkányi Zsuzsanna – Kiss Zoltán 2006. A fonetikai célpontok összeférhetetlenségéről: a magyar v fonológiájának fonetikai alapú megközelítése. In Kálmán László (szerk.): *KB 120: A titkos kötet. Nyelvészeti tanulmányok Bánréti Zoltán és Komlósy András tiszteletére*. MTA Nyelvtudományi Intézet–Tinta Könyvkiadó, Budapest, 263–275.
- Baum, Shari R. – Blumstein, Sheila E. 1987. Preliminary observations on the use of duration as a cue to syllable – initial fricative consonant voicing in English. *Journal of the Acoustical Society of America* 82/3. 1073–1077.
- Bóna, Judit – Gráci, Tekla Etelka – Markó, Alexandra 2008. Coarticulation and speaking style dependency. (In press.)
- Böhm Tamás – Olasz Gábor 2007. A magyar [v] hang szerkezetének és zöreijességének fonetikai vizsgálata. *Beszédkutatás* 2007. 19–35.
- Crystal, Thomas H. – House, Arthur S. 1988. A note on the durations of fricatives in American English. *Journal of the Acoustical Society of America* 84/5. 1932–1935.
- Darcy, Isabelle – Kügler, Frank 2007. Phonological context effects for voicing and devoicing in French. In Trouvain, Jürgen – Barry, William J. (eds.): *Proceedings of the 16th International Congress of Phonetic Sciences. Saarbrücken 6–10 August 2007*. Dudweiler, Saarbrücken, 1257–1260.
- Docherty, Gerard J. 1992. *The timing of voicing in British English obstruents*. Foris Publications, Berlin, 1992.
- Gósy Mária 2004. *Fonetika, a beszéd tudománya*. Osiris Kiadó, Budapest.

- Gósy Mária 2008. Magyar spontánbeszéd-adatbázis – BEA. *Beszéd kutatás 2008*. 194–207.
- Grácz Tekla Etelka 2008. „Z”-öngétlenedés: beszélőfüggő paraméter? *II. Alkalmazott Nyelvészeti Konferencia 2008*. MTA Nyelvtudományi Intézet, Budapest. (Megjelent a kötetben.)
- Jesus, Louis M. T. – Shadle, Christine H. 2003. Temporal and devoicing analysis of European Portuguese fricatives. In Solé, Maria-Josep – Recasens, Daniel – Romero, Joaquin (eds.): *Proceedings of the 15th International Congress of Phonetic Sciences. Barcelona 3-9 August 2003*. Universitat Autònoma de Barcelona, Barcelona, 779–782.
- Kassai Ilona 1982. A magyar beszédhangok időtartamviszonyai. In Bolla Kálmán (szerk.): *Fejezetek a magyar leíró hangtanból*. Akadémiai Kiadó, Budapest, 115–154.
- Lisker, Leigh 1977. Factors in the maintenance and cessation of voicing. *Phonetica* 34. 304–306.
- Maddieson, Ian 1997. Phonetical universals. In Hardcastle, William J. – Laver, John (eds.): *The handbook of phonetic sciences*. Blackwell Publishers, Oxford, 619–639.
- Magdics Klára 1965. *A magyar beszédhangok akusztikai szerkezete*. Nyelvtudományi Értekezések 49. Akadémiai Kiadó, Budapest.
- Magdics Klára 1966. A magyar beszédhangok időtartama. *Nyelvtudományi Közlemények* 68. 125–139.
- Mair, Sheila J. – Shadle, Christine H. 1966. The voiced/voiceless distinction in fricatives: EPG, acoustic and aerodynamic data. *Proceedings of Institute of Acoustics* 18. 163–169.
- Olaszy Gábor 2006. *Hangidőtartamok és időszerkezeti elemek a magyar beszédben*. Nyelvtudományi Értekezések 155. Akadémiai Kiadó, Budapest.
- Olaszy Gábor 2007. A koartikulációs néma fázis jelensége. *Beszéd kutatás 2007*. 47–59.
- Shadle, Christine H. 1997. The aerodynamics of speech. In Hardcastle, William J. – Laver, John (eds.): *The handbook of phonetic sciences*. Blackwell Publishers, Oxford, 33–64.
- Smith, Caroline L. 1997. The devoicing of /z/ in American English: Effects of local and prosodic context. *Journal of Phonetics* 25/4. 471–500.
- Stevens, Kenneth N. 1998. *Acoustic phonetics*. The MIT Press, Cambridge, Massachusetts–London.
- Tonndorf, Wilhelm 1925. Die Mechanik bei der Stimmenlippenwungung und beim Schnarchen. *Zeitschrift für Hals-, Nasen- und Ohrenheilkunde* 12. 159–168.
- Tóth László – Kocsor András 2003. A Magyar Telefonbeszéd-adatbázis (MTBA) kézi feldolgozásának tapasztalatai. *Beszéd kutatás 2003*. 134–147.

## MAGYAR MAGÁNHANGZÓK VIZSGÁLATA ELEKTROMÁGNESES ARTIKULOGRÁFFAL NORMÁL ÉS GYORS BESZÉDBEN

**Mády Katalin**

### **Bevezetés**

A magyar magánhangzók artikulációját a korábbi évtizedben többen is vizsgálták különféle módszerekkel, így pl. legutóbb Bolla (1995) röntgenfelvételek, fotolabiogramok, valamint palato- és lingvogramok alapján. Az azóta eltelt időben elterjedtté vált néhány újabb módszer, amelyek lehetővé teszik az artikuláció pontosabb, többdimenziós, valamint dinamikus, azaz időbeli vizsgálatát. Erre teszünk kísérletet az alábbiakban az elektromágneses artikulográfia módszerével, különös tekintettel arra a kérdésre, hogy a magánhangzóhosszúság milyen szerepet játszik a magyar magánhangzók artikulációjában.

### **A magyar magánhangzók artikulációs jellemzése**

A magánhangzók artikulációját szegmentális szinten alapvetően három jegy alapján szokás leírni, ezek: a nyelv nyugalmi pozíciójához viszonyított vízszintes és függőleges helyzete, valamint az ajakaktivitás. Egyes szerzők a függőleges nyelvállással párhuzamosan, ennek szinonimájaként helyenként megadják a zártság fokát, valamint az ajakműködés tágasságát is (Kassai 1998).

A magyar magánhangzók pontos számát illetően nincs teljes egyetértés. A leírásokban közös, hogy megadnak három felső nyelvállású magánhangzót (elöl képzett ajakrétes /i/, elöl képzett kerekített /y/, hátul képzett kerekített /u/), valamint két középső nyelvállású hangzót (elöl képzett kerekített /ø/, valamint hátul képzett kerekített /o/). Ezek a hangok rövid és hosszú hangzóként egyaránt előfordulnak, a rövid hang többnyire kis mértékben nyitabb, tehát centralizált. Függetlenül attól, hogy ez a különbség a sematikus ábrázolásban kifejezésre jut-e vagy sem, a legtöbb szerző e hangokat egy hosszúsági oppozíció mentén helyezi el.

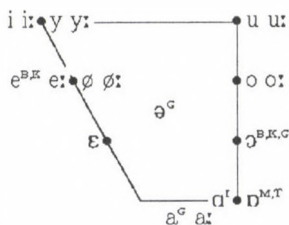
Kevésbé egységesek az álláspontok a fennmaradó magánhangzókra vonatkozóan a következő pontokban (Bolla 1995, Kassai 1998, Gósy 2004): 1. a rövid elöl képzett ajakrétes hangok száma, azaz az *e* graféma ejtészváltozatai; 2. az á vízszintes nyelvállása (centrális, ill. hátsó) és a hangszínben neki megfelelő

1. Ezek a paraméterek finomíthatók, valamint kiegészíthetők szupraszegmentális és tonális jegyekkel, mint pl. a hosszúság vagy a hangsúlyosság.



rövid hang státusza; 3. az *a* függőleges nyelvállása (alsó, ill. legalsó), valamint a vele járó ajakműködés; 4. a *svá* státusza a magánhangzórendszerben.

Az 1. ábra a fent említett különböző magánhangzókat, ill. különböző hangszíneket összesíti.<sup>2</sup>



1. ábra

A magyar magánhangzók összefoglaló ábrázolása

(Amennyiben egy hang nem szerepel minden leírásban, megadjuk a forrás kezdőbetűjét, ahol az adott hang előfordul: **B** = Bolla 1995, **G** = Gósy 2004,

**I** = IPA 1999, **K** = Kassai 1998, **M** = Maddieson 1984, **T** = Ternes 1987.)

### A kvantitás szerepe a magyar magánhangzórendszerben

Kovács (2002) és Gósy (2004) a magyar magánhangzórendszer felső és középső nyelvállású magánhangzóit egy hosszúsági oppozíció mentén felsorakozó öt magánhangzópárnak tekinti. Gósy (2004) az [a:] magánhangzó mellett feltételez egy azonos minőségű rövid [a] hangot is, amely gondozott beszédben, bizonyos idegen szavak ejtésekor fordul elő egyes beszélők nyelvhasználatában.

Más tanulmányok, amelyek eltekintenek a csupán bizonyos nyelvváltozatokban létező magánhangzóktól, és négy alsó, ill. legalsó nyelvállású magánhangzót tekintenek a rendszer részének, ezeket a magánhangzókat is párokba szervezik, szem előtt tartva a nyilvánvalóan fennálló hangszínbeli különbséget. Így tehát az [e:] rövid párja a nyíltabb [ɛ], a centrális vagy hátul képzett [a:] rövid megfelelője az alsó vagy legalsó, ajakkerekítéses vagy ajakréses [ɔ]/[ɒ]/[ɑ]. Ezzel a hosszúsági oppozíció a magánhangzórendszer egészére kiterjed, amely hét rövid-hosszú párként fogható fel. Emellett a következő érveket szokás felhozni: 1. a hosszúsági jegy **jelentésmegkülönböztető szerepe** (IPA 1999); 2. a **többségi elv**, amely a rendszer gazdaságosságát és szimmetriáját szolgálja (Ternes 1987); 3. valamint a fenti érvek kiegészítése a hét magánhangzó-párra érvényes **morf fonológiai alternációkkal** (Nádasdy-

2. Bolla (1995: 92) a fentiek mellett egy [ɛ̃] hangot is megad. Ez a diakritikum a Bolla által használt IPA jelölésrendszerben „megemelt” (raised) képzésmódra utal.

Siptár 1994), amelyek fényében a rövid és hosszú alsó/legalsó magánhangzók szembeállításuk is értelmessé válik.

Tekintettel arra, hogy az alább bemutatandó kísérletben felolvasott álszavakat vizsgálunk, a jelen tanulmányban kézenfekvő volt a magánhangzók írásképeiből kiindulni. A magyar helyesírás 14 magánhangzót feleltet meg külön grafémának, ehhez hét rövid magánhangzót, valamint ezek ékezettel jelölt hosszú megfelelőjét használja fel. Mivel tanulmányunk egyik tárgya a magánhangzók artikulációjának leírása, nem tűnik célszerűnek fonetikai jelekkel hivatkozni a beszédhangokra, hiszen ezzel állást foglalnánk olyan kérdésekben is, amelyeket csak az Eredmények fejezetben fogunk megválaszolni.<sup>3</sup>

### **Lindblom H&H-elmélete**

Amennyiben elfogadjuk, hogy mind az  $i - í$ , mind az  $e - é$  hangpár fonológiai szempontból egy hosszúsági oppozíció két eleme, felvetődik a kérdés, hogy az eltérő hosszúságból következő rövidebb vagy hosszabb tartam milyen szerepet játszik az artikulációs folyamatban. Az artikulációs kutatásban a kérdést gyakran Lindblom elmélete alapján szokták megvilágítani, amely a szerző 1990-es írását követően H&H-elmélet (Hyper and hypo speech – túlartikulált és alulartikulált beszéd) néven vált ismertté, és a beszédprodukción és -percepció összefüggéseit modellezi (Lindblom 1990).

Lindblom szerint a beszélő a kommunikációs helyzetben kétféle célt követ. Egyrészt érthetővé akarja tenni beszédét a hallgató számára, ezért igyekszik a szegmensek megkülönböztetéséhez elegendő jegyet produkálni. Amennyiben ez a cél dominál, hiperartikulálttá válik a beszéd (pl. az adott nyelvet rosszul ismerő külföldivel szemben, fontos információ közlésekor, vagy ha zajos hátter mellett kell megértetnünk magunkat). A beszélő egyszersmind törekszik a gazdaságos produkcióra, azaz az artikulációs befektetés minimalizálására.

A H&H-elmélet alapjául a **célkonfiguráció** fogalma szolgál, amely alatt az adott beszédhangra jellemző idealizált artikulációs és akusztikai mintát értjük. Míg a túlartikulált beszédben feltételezzük, hogy a beszédképző szervek elérik vagy megközelítik az adott hang célkonfigurációját, redundáns közlések esetén nagyobb fokú alulkonfiguráltságot, ún. célalulmúlást (target undershoot) várunk. Ez utóbbi egyrészt a centralizált képzésmódról, másrészt az erősebb koartikulációs befolyásról ismerhető fel, és értelemszerűen fokozatos, nem kategorikus jellemző.

A hosszúsági oppozíciót tartalmazó magánhangzórendszerek kapcsán felmerül a kérdés, hogy az egymásnak megfelelő hosszú és rövid magánhangzók célkonfigurációja azonos-e. Különböző modellek képzelhetők el: 1. a két magánhangzó célkonfigurációja azonos, rövid magánhangzók képzésekor a beszélőszervek mozgása gyorsabb; 2. a két magánhangzó célkonfigurációja azonos, de a rövid magánhangzók jellemzően alulkonfiguráltak; 3. a célalul-

3. Erre elsősorban az  $a$  különböző jelölésmódjai miatt van szükség.



múlás nem a rövidebb tartam következménye, hanem a fonológiai rendszer része, azaz a rövid magánhangzó hosszabb képzés esetén sem éri el a hosszú magánhangzó célkonfigurációját. Az 1. pont a tisztán kvantitativ nyelvekre jellemző (pl. cseh, finn, japán), a 3. pedig például azokra a nyelvekre, ahol a hosszúsági oppozíció feszességi oppozícióval jár együtt (pl. angol, német).

A korábban elmondottak alapján azt váránk, hogy a magyar magánhangzórendszerben a hosszúsági oppozíció különböző artikulációs mintákat hív elő. A felső és középső nyelvtávolságú magánhangzó párok feltehetően a 2. pontban vázolt elv szerint viselkednek, tehát a rövid magánhangzók tipikusan alulkonfiguráltak – erre utal szisztematikus redukáltságuk (lásd pl. Gósy 2004). Az *e* és *é* esetében feltételezzük, hogy nem célalulmulásról, hanem eltérő célkonfigurációról van szó. Valamivel összetettebb a kép az *a* hang esetén, mert a függőleges nyelvhelyzet a leírások tanúsága szerint valamivel közelebb van a magánhangzó-trapéz középpontjához, mint az *á* esetében (azaz legalsó helyett alsó nyelvtávolságról beszélünk), ami a célalulmulás jele is lehet, ám az ajakkerekítés mint aktív gesztus ellene szól a célalulmulás modelljének. Tovább bonyolítja a képet, ha az *á*-t nem hátul képzett, hanem centrális hangnak tekintjük, ez esetben ugyanis a hátul képzett *a* hang biztosan nem alulkonfiguráltság eredménye.

Egy hang akkor alulkonfigurált, ha képzésére a beszélő prozódiai, ritmikai vagy pragmatikai okokból nem tud vagy nem akar elegendő időt fordítani. A következő kísérletben a túl-, illetve alulartikulált beszédmodot a beszédtempó változtatásával vizsgáltuk. Lassú beszédben valószínűbb a célkonfiguráció elérése, valamint a kisebb koartikulációs befolyás, gyors beszédben pedig az alulkonfiguráltság mértékének növekedése.

### Anyag és módszer

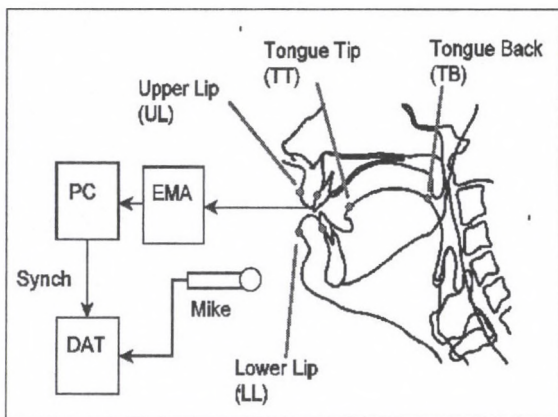
14 magyar magánhangzót vizsgáltunk *C\_Ca* szerkezetű álszavakban, amelyek mondatonként kétszer, szimmetrikus /p/, /t/, /k/ környezetben fordultak elő, azaz a kiejtendő mondatok a *Most a kéka meg a kéka volt* séma szerint épültek fel.

A kísérletben két köznyelvi normát beszélő egyetemi hallgató vett részt, egy nő és egy férfi, akik beszédüket, nyelvi készségüket, valamint hallásukat tekintve egészségesek voltak. A kísérleti személyek feladata az volt, hogy a képernyőn látott mondatot a megadott útmutatás szerint gyorsan, valamint kényelmes tempóban olvassák fel. A női beszélő, HM egy magánhangzót a három mássalhangzó-környezetben és a két tempóban összesen nyolcszor ejtett ki. Az ismétlések számát a jobb statisztikai megbízhatóság érdekében MI esetében tízre növeltük. A mondatokat a berlini Zentrum für Allgemeine Sprachwissenschaft beszédkutató laboratóriumában rögzítettük kétdimenziós elektromágneses artikulográffal és DAT-magnóval.



### Elektromágneses artikulográf

Az elektromágneses artikulográf (EMA)<sup>4</sup> alapelve, hogy a szájüregben és a fejen rögzített néhány milliméteres szenzorok mozgását rögzíti. A kísérleti személy a felvétel alatt egy keretben ül, amelynek a szélére hat adó van rögzítve. Az adók váltakozó pólusú mágneses teret hoznak létre, amely a szenzorokban váltóáramot indukál, ez lehetővé teszi a szenzorok helyzetének pontos követését (a berendezés szerkezetét a 2. ábra szemlélteti). A szenzorok szigetelésének és a váltóáram parányi feszültségének köszönhetően az eljárás az egészségre teljességgel ártalmatlan.



2. ábra

Az elektromágneses artikulográf sematikus ábrázolása

Az EMA nagy előnye, hogy magas időbeli felbontásának (200 kép/s) és a viszonylag alacsony mérési hibahatárnak (2 mm) köszönhetően igen megbízható képet ad a beszélszervek mozgásáról. Mivel az eljárás zajmentes, lehetőség nyílik egyidejű jó minőségű hangfelvételek készítésére is. Természetesen szem előtt kell tartani, hogy a rögzített beszéd nem teljesen természetes, hiszen a szenzorok és a hozzájuk csatlakozó kábelek kétségkívül befolyásolják a beszélszervek mozgását. A magánhangzók vizsgálatára mindezzel együtt az EMA a jelenleg legalkalmasabb módszer.

Vizsgálatunkban hat szenzor adatait elemezzük, ezek helyzete a következő: 1. alsó ajak kívül, 2. alsó fogmeder és alsó ajak között, 3. nyelvhegy, 4. nyelvperem, 5. a nyelv hát kemény szájpadlás felé eső része, 6. a nyelv hát lágy száj-

4. A kétdimenziós artikulográfot EMMA-ként is szokás emlegetni, az *Electromagnetic Midsagittal Articulography* elnevezésre utalva.

padlás felé eső része. Az 1. szenzor eszerint az ajakműködést (tehát a kerekítettséget) jelzi, a 2. az állkapocs nyíltsági fokát, a 3. és 4. a nyelv első részének vízszintes és függőleges mozgását, az 5. szenzor nagyjából a palatális képzéshelyhez tartozó nyelvhelyzetet, a 6. pedig a veláris képzéshelyhez tartozó nyelvhelyzetet.

Az EMA-felvételt egy előzetes vizsgálat előzte meg. Ennek során a kísérleti személynek megmutattuk a berendezést, és próbaképp rögzítettünk egy szenzort a nyelvéhez. Ezzel egyrészt azt teszteltük, természetes marad-e a beszéde a megváltozott körülmények között, másrészt hogy tolerálja-e a szájüreg átmeneti irritálását, főként a nyelvhát érintését. Az előkészítő vizsgálatok ellenére nem garantálható, hogy a felvétel értékelhető lesz – így vizsgálatunkból például utólag ki kellett zárni egy harmadik kísérleti személy adatait.

Az elemzés statikus, a magánhangzók tartamának felezőpontjában mért adatokra támaszkodik. A felvétel közben készített hanganyag alapján mértük a magánhangzók tartamát, valamint 1. és 2. formánsát.

## Eredmények

### Tartam

A gyors beszédtempó elsősorban a hosszú magánhangzók rövidülését vonja magával, ezek ugyanis HM esetében a 123 ms-os középtérről átlagosan 30 ms-mal csökkentek a megnövelt tempó hatására, a rövidek 63 ms-os középtérről viszont átlagosan csak 10 ms-mal (azaz 25%-kal, ill. 16%-kal). MI lényegesen hosszabb magánhangzói 181 ms-ról átlagosan szintén kb. 30 ms-mal rövidültek a gyorsabb tempó esetén (= 15%-os rövidülés), átlag 106 ms-os rövid magánhangzói viszont alig 3 ms-mal, azaz 3%-kal. HM rövid magánhangzói esetében a viszonylag kismértékű tempócsökkenés nem meglepő, hiszen a rövid magánhangzók tartamcsökkenése az érthetőség megőrzése mellett csak bizonyos mértékig lehetséges. MI rövid magánhangzói esetén ez a probléma nem áll fenn – úgy tűnik azonban, hogy e kísérleti alannak általában véve nehezebbé esett a gyorsítás.

A rövid és hosszú magánhangzók tartama HM esetében többnyire egyértelműen elkülönült. Ez alól kivételt képez az *i* lassú és gyors beszédben egyaránt, valamint a lassú tempójú *u*. Ezzel szemben MI magánhangzói többségükben nem különülnek el egyértelműen, ami alól csak az *a* és *á* kivétel. Ez egybecseng korábbi eredményeinkkel, miszerint a felső nyelvállású magánhangzók tartama jelentős átfedést mutat, szemben az egyértelműen elkülönülő *a* és *á* hangokkal (Mády–Reichel 2007). Természetesen az átfedés mértéke kisebb, ha a tartamokat a szomszédos mássalhangzók szerint elkülönítve vizsgáljuk.

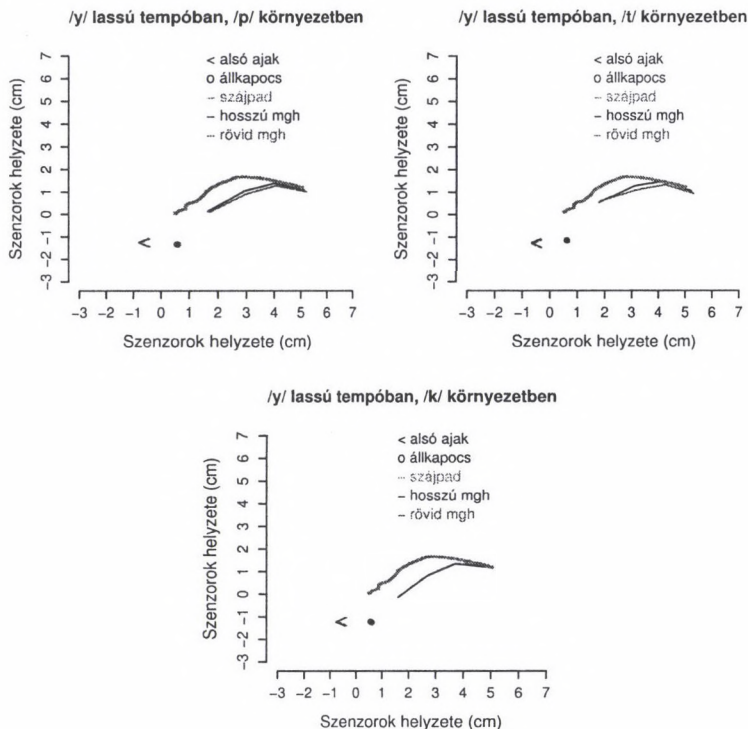
A kísérlet egyik célja az volt, hogy tartambeli átfedéseket generáljon a lassú, rövid és a gyors, hosszú magánhangzók között. Így ugyanis választ kaphattunk volna arra, hogy a felső és középső nyelvállású rövid magánhangzók hosszú párjukhoz képest azért centralizáltak-e, mert képzésükhöz rövidebb

idő áll rendelkezésre (ahogy ezt egy klasszikus kvantitatív nyelvben várnánk), vagy azért, mert célkonfigurációjuk eleve eltérő, tehát azonos tartam mellett sem fedik egymást (mint az angolban vagy a németben, lásd fent).

Kísérleti anyagunk erre a kérdésre nem ad választ: ha a környező mássalhangzót is figyelembe vesszük, egyik beszélő sem rövidítette le a hosszú magánhangzókat gyors beszédében olyan mértékben, hogy azok elérték volna a lassan ejtett rövid magánhangzók tartamát. Ezért a konfiguráció és tartam összefüggését egy következő kísérletben fogjuk vizsgálni.

### Artikulációs minták lassú beszédben

**Felső nyelvállású magánhangzók (*i, ü, u*):** A nyelv helyzete, az állkapocs nyíltsági foka, valamint az ajakműködés a hosszú magánhangzók ejtésekor igen kis mértékben tért el (a különbség többnyire nem volt szignifikáns). Az egyetlen lényeges különbség a /t/ koartikulációs befolyásában látszott, ami a



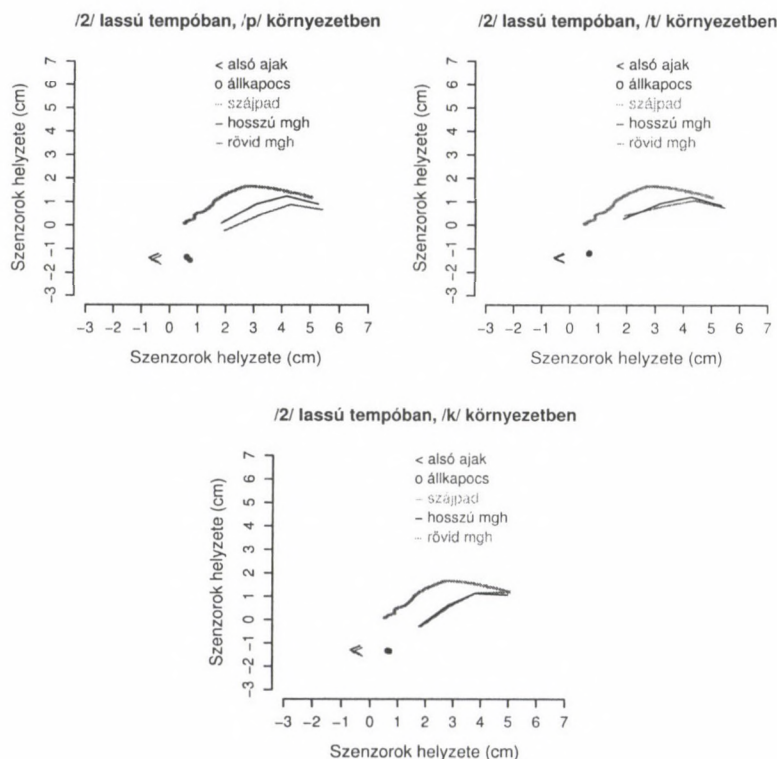
3. ábra

A beszélőszervek helyzete *ü* és *ű* képzésekor (medián értékek HM adataiból)



rövid magánhangzóknál az első két nyelvészor magasabb helyzetében mutatkozott meg. A toldalékcso kersztmetszete HM ü hangja esetében mutatta a legnagyobb különbséget a rövid és hosszú magánhangzó között (3. ábra). MI kerekített magánhangzói ejtésekor az 1., azaz az alsó ajak helyzetét jelző szenzor jóval nagyobb kitérést mutat az ajakréses magánhangzókhoz képest, mint HM ejtésében – ez a többi nyíltági fok esetében is fennáll.

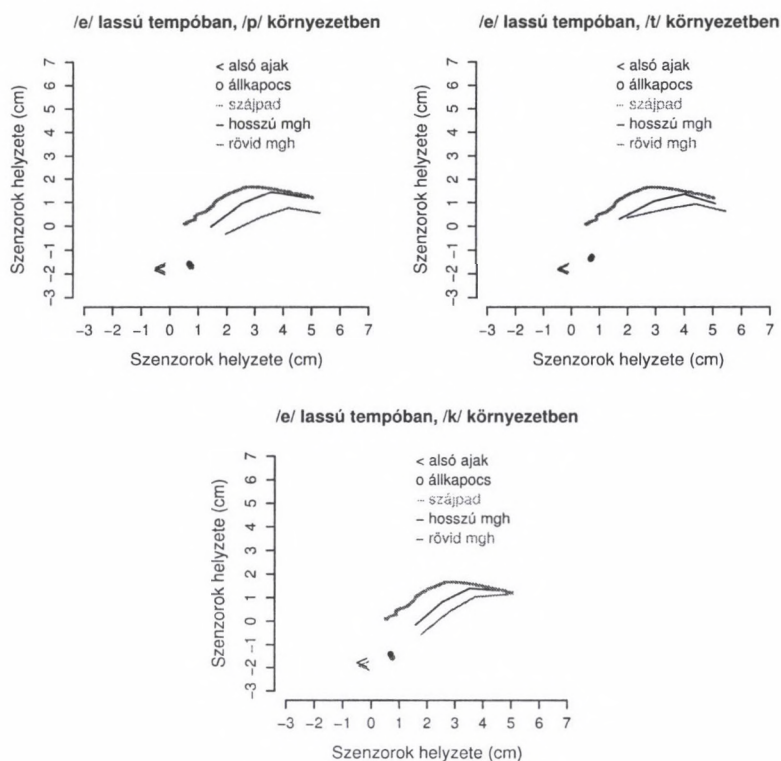
**Kerekített középső nyelvállású magánhangzók (ö, o):** A nyelv helyzete a rövid és a hosszú magánhangzó képzésekor egyértelműen eltér. Az ajakműködés részben, de nem minden esetben intenzívebb a hosszú magánhangzó képzésekor, egyes esetekben az állkapocs helyzetét jelző szenzor is előrébb található a rövid magánhangzóhoz képest (4. ábra). A veláris mássalhangzó koartikulációs hatása mindkét magánhangzóra jelentős, a rövid hangzó ejtésekor a nyelv valamivel hátrébb kerül a hosszúhoz képest.



4. ábra

A beszélőszervek helyzete ö és ő képzésekor (medián értékek HM adataiból)

**Az *e* és *é* hang:** A hosszú és rövid hang képzése egyértelműen eltérő nyelvhelyzetekhez kötődik. Feltűnő azonban, hogy a rövid hangra jellemző alacsonyabb nyelvállás nem jár együtt az állkapocs nagyobb nyíltsági fokával. HM esetében az *e* képzésekor az állkapocs függőleges kitérése a környező mássalhangzó függvényében átlagosan 0,9–1,5 mm-rel nagyobb, mint az *é* artikulációjában (5. ábra). MI esetében a különbségek valamivel nagyobbak, de a 3 mm-t nem lépik túl. (Emlékeztetőül: az EMA-ban 2 mm-es eltérések még mérési hibából is származhatnak.)



5. ábra

A beszélszervek helyzete *e* és *é* képzésekor (medián értékek HM adataiból)

Az itt megfigyelt jelenséget már Wood (1975) is leírta egy általánosabb összefüggésben: szerinte a magánhangzók nyíltsági fokának az artikulációban az állkapocs nyíltsága feleltethető meg. Ha két magánhangzóban az áll-

kapocs helyzete megegyezik, a nyelv függőleges helyzete viszont eltér, az a feszességi oppozíció korrelátuma.<sup>5</sup> Konkrét példával szemlélítve: az [i:] – [e:] szembenállása az állkapocs nyíltsági fokában, az [i] – [ɪ] szembenállása pedig a nyelvállásban manifesztálódik. Érdekes, hogy bár a magyarban a feszesség nem számít elsődleges fonológiai jegynek, az *e* és *é* mégis hűen tükrözi a feszességi oppozíció artikulációs mintáját.

**Az *a* és *á* hang:** Amint a bevezetésben elmondtuk, a rövid *a* minőségét illetően a leírások meglehetősen megosztottak. Adatainkból a következő kép rajzolódik ki: 1. a rövid és a hosszú hang képzésekor megfigyelt nyelvállás foka azonos; 2. az *á* és az *a* nyelvállása kismértékben alacsonyabb, mint az *e* hangé, vagy megegyezik azzal; 3. az ajakműködés az *a* esetében intenzívebb, tehát a hang kerekített – ez fokozottan érvényes MI artikulációjára; 4. az állkapocs az *á* képzésekor 2–4 mm-rel nyíltabb; 5. az *a* képzéshelye az *á* hangénál hátrébb, feltehetőleg a leghátsó szenzor helyzete mögött található, míg az *á* képzésekor a palatális nyelv szenzor függőleges kitérése nagyobb, mint a velárisé (6. ábra).

#### Artikulációs minták gyors beszédben

A gyors tempóban ejtett magánhangzók artikulációjában megfigyelhető tendenciák csupán az *a* és *á* hangok esetében különböznek a fent elmondottaktól: a rövid és hosszú magánhangzó függőleges nyelvállása még kis mértékben sem tér el egymástól, a különbség kizárólag az ajakműködésben, az állkapocs nyíltságában, valamint a vízszintes nyelvállásban manifesztálódik. Ugyanez érvényes a lassan ejtett *a* és a gyorsan ejtett *á* hangzóra is.

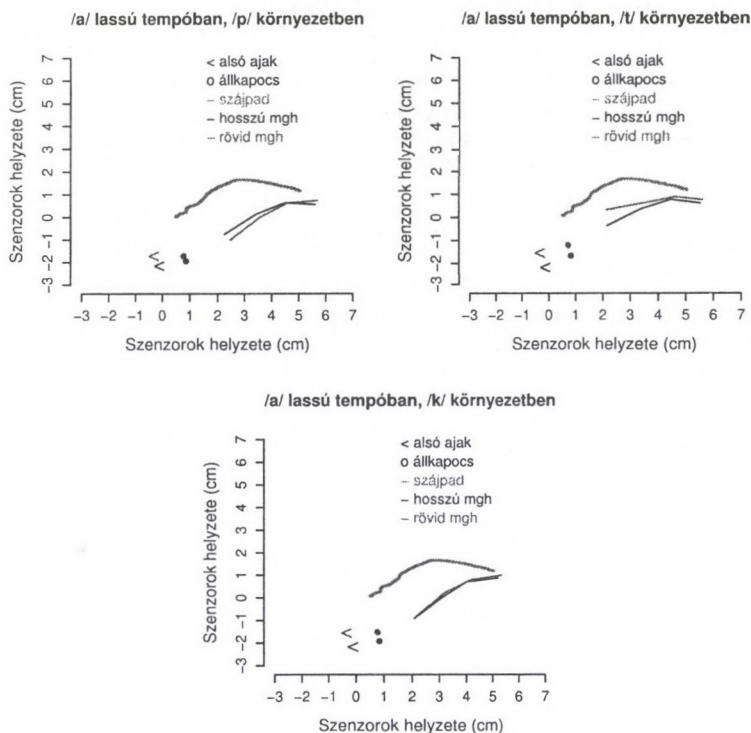
Adatainkban az *á* tartama azonos beszédtempó mellett egyértelműen különbözik az *a* tartamától, ami a két beszédtempó összehasonlításában is fennáll, sőt az *á* kiemelkedik hosszú tartamával a többi magánhangzó közül. Valószínű, hogy a lassú tempóban megfigyelt alacsonyabb nyelvállás ennek a ténynek köszönhető, vagyis annak, hogy a beszédképző szervek a rendelkezésre álló idő alatt jobban meg tudták közelíteni célkonfigurációjukat, mint az *a* vagy a gyorsan ejtett *á* esetén. Ez egyszersmind arra is utal, hogy az *a* és *á* célkonfigurációjában a függőleges nyelvállás nem játszik megkülönböztető szerepet, az ajakműködés, az állkapocs nyíltsági foka, valamint a vízszintes nyelvhelyzet viszont igen.

#### Alulkonfiguráltság

Amint a fenti ábrák mutatják, /t/ környezetében a nyelvhegy és a nyelvperem magasabban található, mint a nem apikális mássalhangzók szomszédságában. Ez a hatás jóval erősebb a rövid magánhangzók esetén, tehát elmond-

5. Ebben az összefüggésben meglehetősen érdekes, hogy az angol *high*, *mid-high*, *mid-low*, *low vowel* megnevezések magyar megfelelője *felső*, *középső*, *alsó* és *legalsó nyelvvállású magánhangzó*, hiszen Wood a nyelvvállást a laza-feszés oppozíció paramétereként fogja fel.





6. ábra

A beszélő szervek helyzete *a* és *á* képzésekor (medián értékek HM adataiból)

hatjuk, hogy a rövid magánhangzók a magyarban minden esetben alulkonfiguráltak. Ugyanakkor a /k/ hatása a rövid és hosszú hangzók ejtésében egyaránt kimutatható, és a nyelvállások különbségének semlegesítése irányában hat. Ez alól csupán az *e* és az *é* hang kivétel, mivel itt a függőleges nyelvállások különbsége jelentős a rövid és hosszú magánhangzó között.

A mássalhangzók koartikulációs befolyása régóta ismert tény, amint azt többek között Hillenbrand–Clark–Nearey (2001) is dokumentálják. Ez a hatás HM és MI formánsértékeiben is erősen tükröződik: normál beszédtempó mellett rövid hangok esetén az  $F_2$  értéke átlagosan 200 Hz-cel is eltérhet a környező mássalhangzók függvényében, a szélső értékek különbsége pedig akár 400 Hz is lehet. Az ingadozás kisebb a hosszú magánhangzók esetén, de MI magánhangzói itt is 100 Hz-ig terjedő eltérést mutatnak, HM hangzói valamivel kevesebbet.

### Az *a* artikulációs besorolása

Amint a Bevezetésben említettük, a magyar magánhangzórendszer leírásában vitás kérdés az *a* hangzó artikulációs jellemzése. Amint az Eredményekben láttuk, az *a* és az *á* közötti különbség az állkapocs nyíltsági fokában, valamint az ajakműködésben nyilvánul meg. Felmerül a kérdés, hogy az alsó ajak és az állkapocs mozgása egymástól független-e. Wood (1975) és Hoole-Kühnert (1996) vizsgálataiból kiderül, hogy azonos nyíltsági fokú magánhangzók esetén (a *nyelvállás* kifejezést itt szándékosan kerüljük el) a kerekített magánhangzó alacsonyabb nyelvállással és szűkebb állkapocsnyílással valósul meg. (Wood az [y] – [i], valamint az [o]/[ɔ] – [a]/[ɑ] magánhangzópárokat vizsgálta különféle nyelvekben.)

Az 1. és 2. táblázatban elől képzett kerekítetlen és kerekített magánhangzók állkapocsnyílását és függőleges nyelv helyzetét hasonlítjuk össze. HM adatai alapján az *ő* állkapocsnyílásának nem az *é*, hanem az egy fokkal magasabb nyelvállású *í*, valamint esetenként az *ü* nyílása felel meg. A nyelvhat helyzete igazodik a Wood (1975), valamint Hoole-Kühnert (1996) által megfogalmazott mintához: az *ü* és *ő* képzésekor a nyelvhat alacsonyabb, mint a megfelelő nyelvállásfokú kerekítetlen magánhangzó esetében. Az *ü* nyelvállása az *é*-hez képest a környező mássalhangzó koartikulációs hatásától függ, tehát nem írható le általános tendenciaként.

1. táblázat: Az állkapocs (Ák, 2. szenzor) és a nyelvhat (Nyh, 5. szenzor) függőleges kitérése, HM adatai alapján (centiméterben, átlagolva)

|          | <i>i</i> |      | <i>ü</i> |      | <i>é</i> |      | <i>ő</i> |      |
|----------|----------|------|----------|------|----------|------|----------|------|
|          | Ák       | Nyh  | Ák       | Nyh  | Ák       | Nyh  | Ák       | Nyh  |
| <i>p</i> | -1,43    | 1,49 | -1,34    | 1,39 | -1,56    | 1,47 | -1,34    | 1,24 |
| <i>t</i> | -1,26    | 1,52 | -1,19    | 1,46 | -1,37    | 1,36 | -1,20    | 1,25 |
| <i>k</i> | -1,36    | 1,47 | -1,25    | 1,34 | -1,45    | 1,40 | -1,33    | 1,15 |

2. táblázat: Az állkapocs (Ák, 2. szenzor) és a nyelvhat (Nyh, 5. szenzor) függőleges kitérése, MI adatai alapján (centiméterben, átlagolva)

|          | <i>i</i> |      | <i>ü</i> |      | <i>é</i> |      | <i>ő</i> |      |
|----------|----------|------|----------|------|----------|------|----------|------|
|          | Ák       | Nyh  | Ák       | Nyh  | Ák       | Nyh  | Ák       | Nyh  |
| <i>p</i> | -1,41    | 1,50 | -1,37    | 1,38 | -1,71    | 1,26 | -1,59    | 1,00 |
| <i>t</i> | -1,18    | 1,58 | -1,13    | 1,43 | -1,39    | 1,25 | -1,28    | 1,06 |
| <i>k</i> | -1,41    | 1,32 | -1,26    | 1,21 | -1,75    | 1,11 | -1,57    | 0,81 |

MI magánhangzóiban ezzel szemben nem találkozunk az állkapocs vagy a nyelv kitérésének kategóriaváltásával. Az *ő* magánhangzó az állkapocs kitérése szempontjából a következő sorrendet mutatják: *ü* – *i* – *ő* – *é*, a zártabbtól a nyíltabb hang felé. A nyelvhat helyzetét tekintve a felsőbbtől az alsóbb

nyelvállásokig a sorrend a következő:  $i - \tilde{u} - é - \tilde{o}$ . A megfigyelt tendencia nem változik a koartikuláció hatására.

A fentiekből következik, hogy az  $a$  és  $\acute{a}$  képzésekor mért állkapocsnylás különbsége önmagában nem jelenti azt, hogy a két hangot eltérő nyelvállásfokúnak kell tekintenünk. Ellenkezőleg: amennyiben a kerekített  $a$  alsó nyelvállású, akkor az állkapocs függőleges kitérése az eggyel felsőbb nyelvállású kerekítetlen  $\acute{e}$ -nek felel meg, ha pedig legalsó nyelvállású, akkor az alsó nyelvállású  $e$ -nek. (A nyelvhelyzet összehasonlításának az  $e$  és  $\acute{e}$ , ill. az  $a$  eltérő képzéshelye miatt nincs értelme.)

A 3. táblázatból kiderül, hogy HM artikulációjában az  $a$  képzésével járó állkapocsnylás nagyjából megegyezik az  $e$ -re jellemző adatokkal, az  $\acute{e}$ -hez képest viszont lényegesen nagyobb nyíltságot mutat. Ezzel szemben MI ejtésében (4. táblázat) az  $a$  állkapocsnylása az  $e$  és az  $\acute{e}$  állkapocsnylása közé esik, eltekintve a  $/p/$  környezettől, ahol az  $a$  kitérése nagyobb, mint az  $e$  vagy az  $\acute{e}$  ejtésekor tapasztalt érték. Nem világos, hogy az eltérést a nála tapasztalt lassabb beszédtempó okozza-e, vagy az, hogy MI ajakműködése kerekítéskor intenzívebb. Az adatok alapján azonban megkérdőjelezhető, hogy az  $a$  alsó nyelvállású magánhangzóként való besorolása helytálló-e, hiszen HM adatai alapján egyértelműen a legalsó kategóriához tartozik, és MI ejtésében  $/p/$  szomszédságában ugyanez a tendencia mutatkozik.

3. táblázat: Az állkapocs (Ák, 2. szenzor) függőleges kitérése, HM adatai alapján (centiméterben, átlagolva)

|     | $a$<br>Ák | $\acute{a}$<br>Ák | $e$<br>Ák | $\acute{e}$<br>Ák |
|-----|-----------|-------------------|-----------|-------------------|
| $p$ | -1,73     | -1,94             | -1,71     | -1,56             |
| $t$ | -1,21     | -1,64             | -1,26     | -1,37             |
| $k$ | -1,51     | -1,90             | -1,58     | -1,45             |

4. táblázat: Az állkapocs (Ák, 2. szenzor) függőleges kitérése, MI adatai alapján (centiméterben, átlagolva)

|     | $a$<br>Ák | $\acute{a}$<br>Ák | $e$<br>Ák | $\acute{e}$<br>Ák |
|-----|-----------|-------------------|-----------|-------------------|
| $p$ | -2,07     | -2,83             | -1,97     | -1,71             |
| $t$ | -1,44     | -2,19             | -1,49     | -1,39             |
| $k$ | -1,93     | -2,62             | -2,17     | -1,75             |

### Összegzés

Vizsgálatunk alapkérdése Lindblom H&H elméletéből kiindulva az volt, hogy a magyar rövid és hosszú magánhangzók célkonfigurációja különbözik-e, és ha igen, miben, valamint hogy a rövid magánhangzók leírhatóak-e hosszú párjaik alulkonfigurált gesztusaként.



A rendelkezésre álló adatok alapján a következőket állapíthatjuk meg: a koartikulációs hatás lényegesen erősebb a rövid magánhangzók esetén, nyelvvállástól függetlenül. A felső nyelvvállású és a kerekített középső nyelvvállású magánhangzókra igazolódott a várt tendencia, miszerint az előbbiek esetében az artikulációs különbség csekély, az utóbbiak esetében a nyelvvállások általában eltérnek, az ajakműködés pedig általában aktívabb a hosszú hangzó ejtésekor.

Az adatok – a korábbi leírásoknak megfelelően – jelentős különbséget mutatnak az eltérő hangszínű magánhangzók esetében: az *e* és az *é* között egyértelmű különbség mutatkozik, méghozzá elsősorban a függőleges nyelvvállás tekintetében. Az *á* és *a* állkapocsnyílásában megfigyelt különbség csak az *a* kerekítettségének figyelembevételével értelmezhető. Mind HM, mind MI adataiban megfigyelhető, hogy a kerekítettség a nyelv alacsonyabb helyzetével és az állkapocs csekélyebb függőleges kitérésével jár együtt. A két beszélő eltérő stratégiát alkalmaz a kerekítettség jelzésére: míg MI ajakműködése intenzívebb, HM nagyobb mértékben alkalmazza a kerekítettség társult jegeit. Annak kiderítésére, hogy az eltérő stratégiák személyfüggőek-e, avagy a két beszélő eltérő beszédtempójára vezethetőek-e vissza, csak egy kiterjedtebb vizsgálat keretében tudnánk válaszolni.

### Irodalom

- Bolla Kálmán 1995. *Magyar fonetikai atlasz. A szegmentális hangszerkezet elemei.* Nemzeti Tankönyvkiadó, Budapest.
- Gósy Mária 2004. *Fonetika, a beszéd tudománya.* Osiris Kiadó, Budapest.
- Hillenbrand, James M. – Clark, Michael J. – Nearey, Terrance M. 2001. Effects of consonant environment on vowel formant patterns. *Journal of the Acoustic Society of America* 109/2. 748–763.
- Hoole, Phil – Kühnert, Barbara 1996. Tongue-jaw coordination in German vowel production. In *Proceedings of the 1<sup>st</sup> ESCA Tutorial and Research Workshop on Speech Production Modelling. 4<sup>th</sup> Speech Production Seminar.* Autrans, France. 97–100.
- IPA (ed.) 1999. *Handbook of the International Phonetic Association.* Cambridge University Press, Cambridge.
- Kassai Iлона 1998. *Fonetika.* Nemzeti Tankönyvkiadó, Budapest.
- Kovács Magdolna 2002. *Tendenciák és szabályszerűségek a magánhangzó-időtartamok produkciójában és percepciójában.* Doktori disszertáció. Debreceni Egyetem.
- Lindblom, Björn 1990. Explaining phonetic variation: a sketch of the H and H theory. In Hardcastle, William J. – Marchal, Alain (eds.): *Speech production and speech modeling.* Kluwer, Dordrecht, 403–439.
- Maddieson, Ian 1984. *Patterns of sounds.* Cambridge University Press, Cambridge.
- Mády Katalin – Reichel, Uwe D. 2007. Quantity distinction in the Hungarian vowel system—just theory or also reality? In Trouvain, Jürgen – Barry, William J. (eds.): *Proceedings of the 16<sup>th</sup> International Congress of Phonetic Sciences. Saarbrücken 6–10 August 2007.* Pirrot GmbH., Dudweiler, 1053–1056.

- Nádasdy Ádám – Siptár Péter 1994. A magánhangzók. In Kiefer Ferenc (szerk.): *Strukturális magyar nyelvtan 2. Fonológia*. Akadémiai Kiadó, Budapest, 42–182.
- Ternes, Elmar 1987. *Einführung in die Phonologie*. Wissenschaftliche Buchgesellschaft, Darmstadt.
- Wood, Sidney 1975. The weakness of the tongue-arching model of vowel articulation. *Lund Working Papers* 11. 55–107.

Köszönöm Christian Gengnek és Phil Hoole-nak,  
hogy az adatok felvételekor és feldolgozásakor számíthattam a támogatásukra.  
Hálával tartozom továbbá Markó Alexandrának konstruktív javaslataiért,  
és mindennekelett az Alexander von Humboldt Alapítványnak,  
hogy kutatásaimat ösztöndíjjal támogatta.

# FANTOMHANGOK A MAGYAR SPONTÁN BESZÉDBEN

Gyarmathy Dorottya

## Bevezetés

A spontán beszéd szegmentális szerkezetének elemzése számos problémát felvet. Gyakori tapasztalat, hogy a hangsorban nem jelenik meg a vizsgálni kívánt beszédhang; ez a jelenség több okra is visszavezethető. A beszéd nem individuális beszédhangok egymáshoz illesztett sorozatának eredménye, egy hangsor létrehozása nem hasonlítható ahhoz a folyamathoz, mikor az írógép vagy a számítógép billentyűinek leütésével létrehozunk egy betűsorozatot. A beszédhangok kölcsönösen hatnak egymásra, befolyásolják egymás képzését és hangzását. Ugyanazon fonéma a hangkörnyezettől függően különféleképpen realizálódik azért, hogy minél hasonlóbbá váljon a szomszédos beszédhangokhoz. A [k] képzési helye például a *kincs* szóban kissé előrébb, míg a *kantár* szóban kissé hátrább tolódik (Gósy 2004a). Erre a jelenségre már Kempelen Farkas is felfigyelt a 18. században, beszélőgépiének építése során. Amikor egy szótagban több hangot akart összekötni, szembesült azzal, hogy bármilyen kis szünetet tart is a két hang között, azok túlságosan elkülönülnek egymástól; ha pedig nem tart szünetet, a hangok összefolynak, és a hangsor érthetatlenné válik (Kempelen 1791/1989). A természetes artikulációban a fonémarealizációk egyfajta átmenettel kapcsolódnak egymáshoz, beszédképző szerveinknek – rugalmatlanságuk következtében – ugyanis időre van szükségük ahhoz, hogy felvegyék az adott beszédhang képzési konfigurációját (Gósy 2004a). Általános elvként kimondható, hogy minél nagyobb a két hang képzésbeli különbsége, annál nagyobb mértékű az egyes hangok egymáshoz való alkalmazkodása (Kassai 1998). A toldalékcso fisiológiai módosulásának folyamata, a szomszédos beszédhangok egymásra hatása a koartikuláció, melynek mértéke a módosuló beszédhangot tekintve sokféle lehet. A hangok egymásra hatásában megkülönböztetjük a fonetikai és a fonológiai koartikuláció fogalmát. A fonetikai koartikuláció térben és időben fokozatosan realizálódik, a szomszédos beszédhangok valamilyen mértékben hatást gyakorolnak egymás artikulációjára. Ennek folyamán létrejöhetnek hangátmenetek, amelyek az egyik hangból a másikba tartó artikulációs mozgások következményei; módosult artikulációs konfiguráció, amikor a szomszédos beszédhangok úgy hatnak egymás artikulációjára, hogy az egyik beszédhang artikulációs konfigurációja módosul; illetve a koartikuláció következtében meg is változhat az eredeti beszédhang. A fonológiai koartikuláció alakítani köthettségű, időben és térben nem fokozatos, az eredményeként létrejött artikulációs



konfiguráció pedig egy másik fonéma alaprealizációjának feleltethető meg. Ide tartozik a hiástöltés jelensége, a hasonulás, a rövid-hosszú oppozíció változása és a kiesés (Gósy 2004a).

A koartikulációs következményen túl a spontán beszédben tapasztalható hangkiesések okaként meg kell említeni az aktuálisan létrehozott artikulációs mozzanat „megbotlását” is. Ilyenkor a feltehetően jó motoros program ellenére a kivitelezés mégis hibás lesz (Gósy 2004b).

Egyfajta hangkiesésnek tekinthetjük továbbá a beszéd folyamat azon részeit is, amelyek percepciós szempontból egyértelműen észlelhetők, a rezgés-, illetve hangszínképen mégsem jelennek meg tipikus formájukban. Részletes akusztikai fonetikai elemzésük szükséges annak igazolására, hogy ez a jelenség összefüggésben van-e valamilyen koartikulációval. Az effajta beszédhangkiesésekkel a magyar nyelvű fonetikai kutatások eddig még nem foglalkoztak. A jelenségre új fonetikai szakkifejezésként a jelen tanulmányban a *fantomhang* kifejezést kívánjuk bevezetni. A *fantom* szó egyik jelentése az ÉKsz. szerint „kísértet” (343), tehát olyan dolog, amelynek fizikai jelenléte nem igazolható, mégis érzékeljük, vagy érzékelni véljük. Így például a *fantomfájdalom* az orvosi szaknyelvben „műtétileg eltávolított testrész helyén érezni vélt fájdalom” jelentésben használatos (ÉKsz.: 343). Ennek analógiájaként a fantomhangok olyan beszédhangok, amelyek percepciósan észlelhetők, fizikailag azonban annyira sincsenek jelen, mint azt a koartikulációs helyzet szokásosan indokolná; és nem tekinthetők az adott beszédhang artikulációs tervezésekor bekövetkező hiba egy formájának, a kiesésnek (a kiesés típusú megakadásoknál a hang hiánya ugyanis hallható). A korszerű hangszínképelemző szoftverek nagyban hozzájárultak a jelenség felfedezéséhez. Ezek ugyanis lehetővé tették a fonetikusok számára, hogy egyidejűleg hallják az elemezni kívánt hangot, és lássák az ennek megfelelő oszcillogramot és spektrogramot. A fonetikai hanganalizáló programok továbbá korlátlan számú visszahallgatást is lehetővé tesznek, és ez az ellenőrzés szempontjából elengedhetetlen. A koartikuláció nem tipikus sajátosságainak pontosabb megismerése a fantomhangok definiálását is pontosabbá teheti.

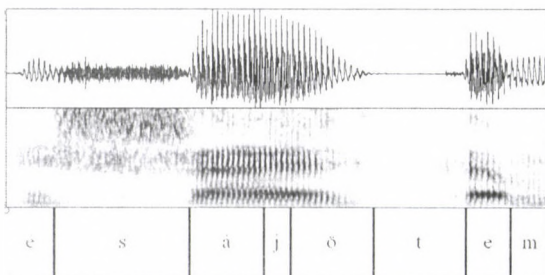
A magyar és a nemzetközi szakirodalomban találkozhatunk hasonló megnevezésekkel, úgymint *fantomszegmentum* (Szigetvári–Törkenczy 2006) vagy *fantom fonéma* (Huet 2003, 2006; Kashino 2006), azonban ezek egyike sem az általunk meghatározott értelemben használatos. Szigetvári és szerzőtársa (2006) a *fantomszegmentum* kifejezést olyan hangokra alkalmazza, amelyek perceptuálisan nem észlelhetők. Az athabasz nyelvét beszélő nyugat-kanadai indiánok bizonyos, észlelhetően magánhangzóra végződő szavakról úgy vélik, hogy mássalhangzóra végződnek. A *díni*-nek ejtett szó ’ez’ jelentésben használva véleményük szerint magánhangzóra, míg ’hangot ad’ jelentésben mássalhangzóra végződik. Gérard Huet (2003, 2006) a szanszkrit nyelv számítógépes szegmentálásában a szandhi (egyfajta hasonulási mód) leírására vezet be egy *fantom* fonémát. Az általunk használt fogalomhoz a

legközelebb a Makio Kashino (2006) által leírt fantom fonéma effektus áll. Ez az agynak a hanghelyreállító hatását jelenti, amellyel már számos korábbi kutatás foglalkozott (pl. Warren 1970; Samuel 1981; Dankovics – Pléh 2001). Ha egy lexikai egység egyik fonémáját törölték, és azt köhögéssel vagy egyéb széles frekvenciájú zajjal helyettesítették, a hallgatók a hiányzó beszédhangot többszöri lehallgatás után is tisztán észlelték, még akkor is, amikor már tudatában voltak, hogy az fizikailag nincs jelen.

A jelen tanulmányban a magyar spontán beszéd **fantomhangjait** elemezzük. Feltételeztük, hogy nem minden fonéma realizálódhat fantomhangként, és ezek megjelenése függ a beszélő artikulációs biztonságától (az adott kontextusban), a fonológiai hangsúlyos vagy hangsúlytalan helyzetben, illetve tartalmas szavakban vagy funkcióelemekben való előfordulásuktól. Választ kerestünk továbbá arra, hogy *a)* mely beszédhangok válnak leggyakrabban fantomhangokká; *b)* megjelenésüket befolyásolja-e, és ha igen, hogyan a beszédhangok spontán beszédben való gyakorisága, valamint a hangkörnyezet; illetve *c)* mely tényezők felelősek megjelenésükhöz.

### Anyag, módszer, kísérleti személyek

Kutatásunkban a BEA magyar spontánbeszéd-adatbázis (fonetikailag megtervezett adatbázis, melynek felvételi körülményei finom fonetikai elemzésekre adnak lehetőséget) 8 beszélőjének (4 nő, 4 férfi) hangfelvételeit elemeztük. Adatközlőink átlagéletkora 46 év volt, voltak közöttük gyakorlott (színész, pedagógus), illetve kevésbé gyakorlott (köztisztviselő, bérletszámoló, könyvtáros) beszélők. Összesen 1 óra 30 percnyi spontán beszédet elemeztünk a Praat akusztikai fonetikai elemző szoftver 5.0 verziójával. Ahhoz, hogy a fantomhangokat adatolni tudjuk, a felvételek beszédhangszintű elemzését kellett elvégeznünk, melynek során 162 olyan esetet tudunk rögzíteni, amikor a hang percepciósan egyértelműen azonosítható volt, a spektrogramon azonban nem jelent meg (1. ábra).



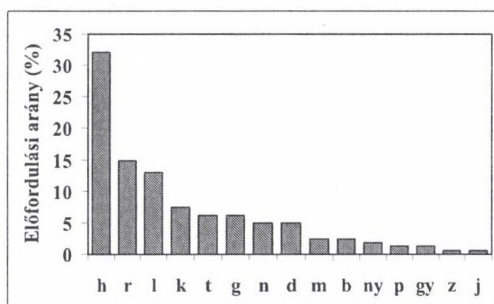
1. ábra

Az *és* és *rájöttem* hangsor rezgés- és hangszínképe: a [r] fantomhang (a [j] mássalhangzó nem tartalmaz a pergőhangra jellemző akusztikai elemet)

A statisztikai vizsgálatokban Pearson-féle korrelációanalízist és binominális nem parametrikus tesztet alkalmaztunk az SPSS 16.0-ás verziójával (99%-os szignifikanciaszinten).

### Eredmények

Hipotézisünknek megfelelően nem minden magyar beszédhang realizálódott fantomhangként. A jelenség az általunk elemzett 8 adatközlő mindegyikénél kizárólag a mássalhangzókat – ezeknek 53%-át – érintette (2. ábra).



2. ábra

A fantomhangok gyakorisága

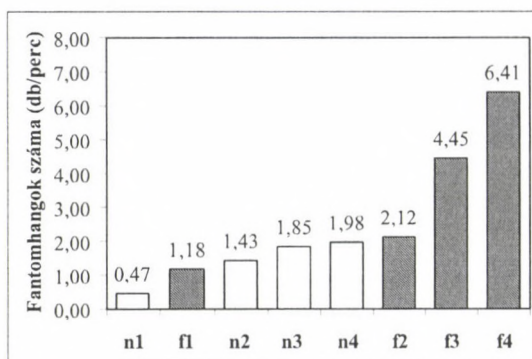
Elemeztük, hogy a fantomhangok gyakorisági rangsora milyen összefüggést mutat az adott mássalhangzók gyakoriságával a spontán beszédben (Gósy 2004). A Pearson-féle korrelációanalízis (értéke:  $r = 0,656$ ), közepesenél kissé erősebb korrelációt eredményezett, amely szignifikáns ( $p = 0,008$ ). Ez azt jelenti, hogy a vizsgált adatok között lineáris, egyenes arányú összefüggés van. Minél gyakrabban fordul elő az adott mássalhangzó a spontán beszédben, annál nagyobb valószínűséggel válik fantomhanggá. A tendencia azonban nem feltétlenül érvényesül minden esetben (a leggyakoribb fantomhang, a [h] például csak a nyolcadik leggyakoribb mássalhangzónk).

Az egyes adatközlőknél elemezve a fantomhangok előfordulási gyakoriságát megállapíthatjuk, hogy a beszédben való jártasság, az artikulációs biztonság, valamint feltehetően az ejtési tudatosság, illetve odafigyelés nagymértékben befolyásolja a jelenség előfordulását. A feszes artikulációjú beszélőknél percenként kisebb (az átlag 1,38 hang/perc), míg a lazábban artikulálókénál (f2, f3, f4) nagyobb számban jelentkeztek fantomhangok (4,32 hang/perc átlagosan) – vö. 3. ábra.

A szótövek és a toldalékok feltehetőleg külön elemekként tárolódnak a mentális lexikonban. Ez magyarázza, hogy a toldalékolt alakok hozzáférése nehezebb, mint a szótöveké (Stanners et al. 1979; Caramazza et al. 1985 – idézi: Gósy 2005a). Megvizsgáltuk, hogy a fantomhangok a szótövekben vagy a tol-



dalékokban fordulnak-e elő gyakrabban. Az eredmények azt mutatják, hogy a vizsgált hangok döntő hányada (83,33%-a) a szótövekben jelenik meg, az eltérés szignifikáns ( $p = 0,001$ ). Lehetséges magyarázatot erre a beszédproduktions folyamat működésében találhatunk. A produkció során a tervezési folyamat előbbre tart az aktuális kiejtésnél. Egy szó kimondásakor már tudatában vagyunk annak, hogy mi lesz a következő, illetve az arra következő szó (Gósy-Markó 2006). Jórészt tisztában vagyunk továbbá a szándékolt közlés grammatikai struktúrájával, illetve az egyes szavak toldalékolási szabályaival. Toldalékoltszavak esetében tehát, amikor a szótót sikerül megfelelően aktiválni, a beszélő figyelme valószínűsíthetően már a tö és a toldalékmorféma összeillesztésére fókuszál. Ez pedig kedvezhet a fantomhangok megjelenésének.



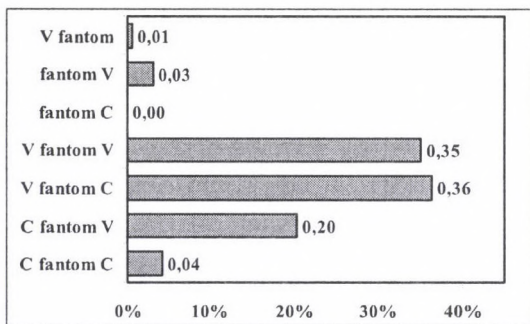
3. ábra

A fantomhangok percenkénti gyakorisága a beszélőknél  
(a fehér oszlopok a női, a szürkék a férfi adatközlőket jelzik)

Beszédünkben eltérő arányban fordulnak elő tartalmas szavak (ige, főnév, melléknév stb.), illetve funkcióelemek (névmás, névelő, igekötő stb.). Szende Tamás 1973-as adatai alapján a spontán beszéd 39,25%-a tartalmas szó, 60,75%-a pedig funkcióelem. Ehhez hasonló eredményeket kapott Keszler Borbála (1983) kötetlen beszélgetéseket vizsgálva. Az általa elemzett felvételekben a tartalmas szavak előfordulási gyakorisága 41,21%, míg a funkcióelemeké 58,79% volt. A BEA-n végzett, még nem publikált kutatások, amelyek jól tükrözik a jelen nyelvallapotot, a spontán beszédben a tartalmas szavak 37,92%-os, míg a funkcióelemek 62,08%-os előfordulását mutatják. Feltételeztük, hogy a funkciószavakban gyakrabban fordulnak elő fantomhangok is, részben a gyakorisági tényező miatt, részben azért, mivel készen tárolódnak mentális a lexikonban (Gósy 2005a). A feltételezett könnyebb előhívhatóság és a klisészerű tárolás miatt az artikulációjuk pontatlanabb lehet. A kapott eredmények azonban némiképp ellentmondanak hipotézisünknek, hiszen

a fantomhangok kisebb hányada (46,3%) fordult elő funkcióelemekben, míg nagyobb részük (53,7%) tartalmaz szavakban jelent meg. Az eredményeket statisztikailag is elemeztük, mely szerint a két csoport közötti különbség nem szignifikáns ( $p = 0,388$ ). A fantomhangok tartalmazó szavakban, illetve funkcióelemekben való előfordulása tehát teljesen véletlenszerű.

Elemeztük, hogy a mássalhangzós vagy a magánhangzós környezet kedvez-e jobban a hangok fantomizálódásának. Az anyagunkban adatolt fantommássalhangzók főként magánhangzós környezetben, nemegyszer intervokális helyzetben fordultak elő (4. ábra).



4. ábra

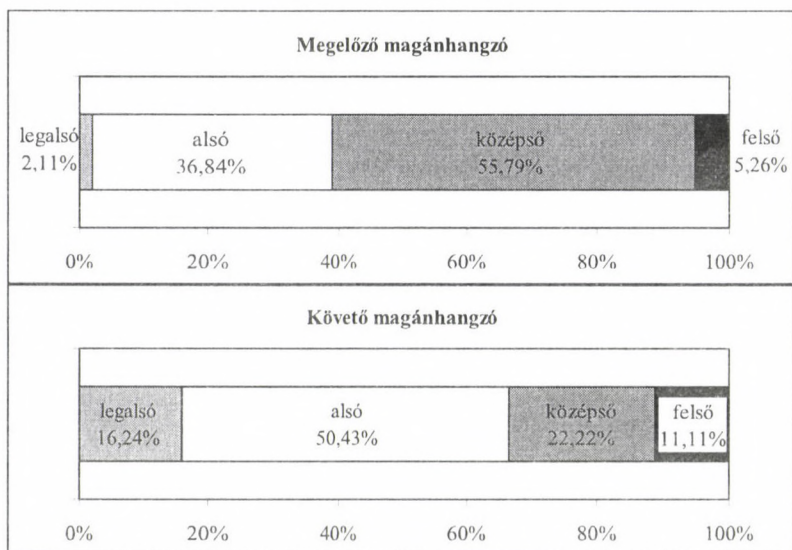
A fantomhangok hangkörnyezete

Elemeztük, hogy a megelőző, illetve a követő magánhangzók nyelvvállásfoka miként befolyásolja a fantommássalhangzók megjelenését. Azt találtuk, hogy a nyelv függőleges állása szerinti szélsőbb pozíciók kevésbé hatnak kedvezően a fantomhangok megjelenésére (5. ábra).

Az általunk adatolt három leggyakoribb fantomhang ([h], [r], [l]) esetében a magánhangzók képzési helye szerint is elemeztük a szomszédos beszédhangok és a fantomhang viszonyát. Valószínűsítettük, hogy a szomszédos beszédhangok képzési helye meghatározó lesz a hangok fantomhanggá válásának szempontjából. A leggyakoribb fantomhang az elemzett beszédanyagban a [h] volt (52 előfordulás), vö. 6. ábra. Az elemzett példák mindegyikében kizárható a vokalizálódott [h] lehetősége.

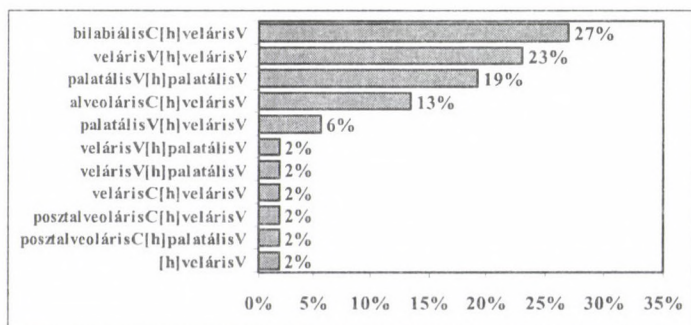
A [h] hang több szempontból is problematikus mássalhangzó. A képzési módját tekintve réshang, a többféle képzési hely és hangszalagműködés több mássalhangzót definiál (Gósy 2005b). Már Kempelen Farkas is megemlíti, hogy „e betűnek van egy különös sajátossága, amely minden más betűtől megkülönbözteti. Az, hogy nincsen saját fekvése, hanem mindig azét a magánhangzóét veszi föl, amely utána következik” (1791/1989: 215). Későbbi kutatások megállapították, hogy magánhangzóközi helyzetben a magyar /h/ foné-

mának több allofónja ismert. Ezek egy része zöngés, másik részük zöngétlen, de előfordul olyan eset is, amikor vokalizálódik, tehát a mássalhangzó lényegében elveszti zöreijhang tulajdonságát. A [h]-nak ez a „kaméleon” tulajdonsága, illetve a vokalizálódás vezethet a fantomizálódáshoz; további elemzést igényel, hogy ilyenkor mi idézi elő a hallgató hangérzetét. A [h] észleletét biztosíthatja például a megelőző és a követő hangok közötti intenzitáscsökkenés.



5. ábra

A megelőző és a követő magánhangzók eloszlása nyelvvállásfok szerint

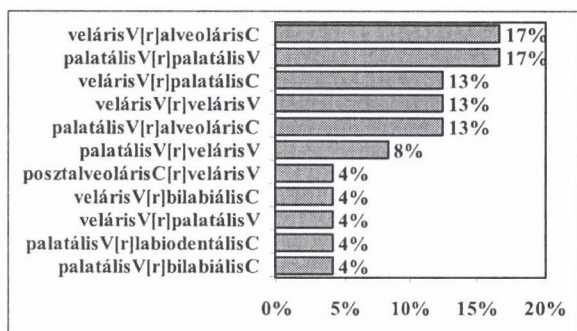


6. ábra

A [h] hangkörnyezete



A második leggyakoribb hang a [r] volt (24 db), amely fantomhangként a következő beszédhangok között fordult elő legtöbbször: két palatális magánhangzó között (*ismeretség*), palatális és veláris magánhangzó között (*ember aki*), két veláris magánhangzó között (*korom*), palatális magánhangzó és alveoláris mássalhangzó között (*írni*), veláris magánhangzó és alveoláris mássalhangzó között (*várnak*), illetve veláris magánhangzó és palatális mássalhangzó között (*tárgyak*) – vö. 7. ábra. Ezekben az esetekben a hang fantomizálódásának lehetséges oka az egyes hangok képzési helyének egymástól való relatív távolsága, illetve a [r] ejtési különbözősége. A fantomhanggá válás azonban nem következett be például palatális magánhangzó és mássalhangzó között (*környezetükben*), veláris magánhangzó és labiodentális mássalhangzó között (*elárvultak*), posztalveoláris mássalhangzó és palatális magánhangzó között (*másrészről*). A [r] mássalhangzóval foglalkozó tanulmány megemlíti a hangképzési hiányt vagy más néven a „percepciós pergőhangot” mint a [r] egyik lehetséges realizációját. Gósy Mária a mássalhangzónak az aktuális képzés hiánya ellenére bekövetkező észleletét a megelőző és a követő hang közötti hirtelen intenzitásváltással magyarázza, továbbá azzal, hogy ez a hangátmenet a percepcióban a perdület akusztikai élményét kelti (2008).



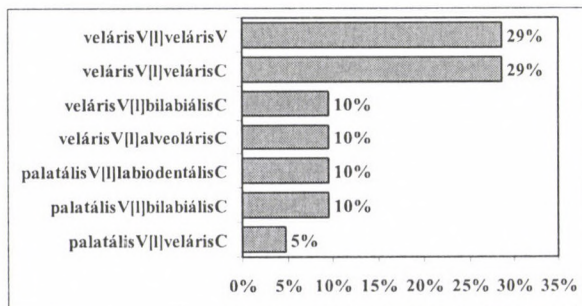
7. ábra

A [r] fantomhang hangkörnyezete

Az elemzett hanganyagban a harmadik leggyakoribb (21 db) fantommássalhangzó, a [l] veláris magánhangzó és mássalhangzó között (*dolgozom*), illetve két veláris magánhangzó között (*választottam*) fordult elő (8. ábra). A [l] mássalhangzóból nem lett fantomhang palatális magánhangzó és alveoláris mássalhangzó között (*beszélnek*), palatális és veláris magánhangzó között (*egyedülállóan*), illetve veláris magánhangzó és labiodentális mássalhangzó között (*múlva*).

Eredményeinkkel igyekeztünk szemléltetni azokat a tényezőket, amelyek szerepet játszanak, játszhatnak a hangok fantomizálódásában. Annak megvá-

laszolója azonban, hogy az adott hang hiányában mi kelt mégis hangérzetet, már nehezebb, további akusztikai fonetikai elemzéseket igényel. Egyik lehetséges okként megemlíthető a megelőző és a követő hang közötti hirtelen intenzitásváltás – ami az adatok többségénél megfigyelhető volt –, illetve a hangátmenet sajátosságai.



8. ábra

A [l] fantomhang hangkörnyezete

### Következtetések

A jelen tanulmány elsőként foglalkozott a magyar beszédhangok fantom realizációjával. Fő célunk a fantomhang jelenségének bemutatása, előfordulási sajátosságai feltárása volt. A magyar beszédhangok közül a mássalhangzók egy része realizálódik fantomhangként. Igazolható az a tendencia, hogy minél gyakrabban fordul elő az adott mássalhangzó a spontán beszédben, annál nagyobb a fantomrealizáció valószínűsége. A gyakran ejtett hangok pontos artikulációja elnagyolt lehet, s erre visszavezethető a fantomhang kialakulása. Az eredmények alapján elmondható, hogy a beszélők artikulációs biztonsága, illetve az ejtési tudatosság szerepet játszik a jelenség létrejöttében. A szótövekben szignifikánsan magasabb arányú volt a mássalhangzók fantomrealizációja, mint a toldalékokban. Ebből arra következtethetünk, hogy a beszédprodukciónál a szótó sikeres aktiválását követően a figyelem átkerül a tö- és a toldalékmorféma összeillesztésére. Az intervokális hangkörnyezet, ezen belül is a középső és alsó nyelvtálasú magánhangzók jobban elősegítik a mássalhangzók fantomrealizációit. A leggyakoribb három fantomhang részletesebb elemzéséből megállapíthattuk, hogy a képzési helyek közti viszonylag nagy távolság is szerepet játszhat a jelenség kialakulásában. A fantomhangok megjelenése függetlennek tűnik attól, hogy tartalmas szavakban vagy funkcióelemekben fordul-e elő. A kutatás folytatásával adható majd válasz arra, hogy az észlelet számára mi biztosítja az objektíven nem létező beszédhang akusztikai élményét.

## Irodalom

- Dankovics Natália – Pléh Csaba 2001. Hangrestaurációs jelenségek és alaktani feldolgozás a magyarban: azt halljuk-e, amit várunk? In Pléh Csaba – Lukács Ágnes (szerk.): *A magyar morfológia pszicholingvisztikája*. Osiris Kiadó, Budapest, 55–83.
- Gósy Mária 2004a. *Fonetika, a beszéd tudománya*. Osiris Kiadó, Budapest.
- Gósy Mária 2004b. A spontán magyar beszéd megakadásainak hallás alapú gyűjteménye. *Beszéd kutatás 2004*. 6–17.
- Gósy Mária 2005a. *Pszicholingvisztika*. Osiris Kiadó, Budapest.
- Gósy Mária 2005b. A /h/ zöngésedése két magánhangzó között. *Beszéd kutatás 2005*. 5–21.
- Gósy Mária – Markó Alexandra 2006. Szegmentumsorok hibás kivitelezése a beszéd-produkcióban. *Magyar Nyelvőr 130*. 198–214.
- Gósy Mária 2008. „R” hangok: kiejtés, hangzás, funkció. *Magyar Nyelvőr 132*. 1–18.
- Huet, Gérard 2003. Towards computational processing of Sanskrit. In *Proceedings of the International Conference on Natural Language Processing, ICON-2003*, Mysore, India. <http://yquem.inria.fr/~huet/PUBLIC/icon.pdf>
- Huet, Gérard 2006. Lexicon-directed segmentation and tagging of Sanskrit. In Tikkanen, Bertil – Hettrich, Heinrich (eds.): *Themes and tasks in old and middle Indo-Aryan linguistics*. Motilal Banarsidass, Delhi, 307–325.
- Kassai Ilona 1998. *Fonetika*. Nemzeti Tankönyvkiadó, Budapest.
- Kashino, Makio 2006. Phonemic restoration: The brain creates missing speech sounds. *Acoustical Science and Technology 27/6*. 318–321.
- Keszler Borbála 1983. Kötetlen beszélgetések mondat és szövegtani vizsgálata. In Rácz Endre – Szathmári István (szerk.): *Tanulmányok a mai magyar nyelv szöveg-tana köréből*. Tankönyvkiadó, Budapest, 164–202.
- Kempelen Farkas 1791/1989. *Az emberi beszéd mechanizmusa, valamint a szerző beszélőgépezet leírása*. Szépirodalmi könyvkiadó, Budapest.
- Samuel, Arthur G. 1981. Phonemic restoration: Insights from a new methodology. *Journal of Experimental Psychology: General 110*. 474–494.
- Szende Tamás 1973. *Spontán beszédanyag gyakorisági mutatói*. Nyelvtudományi Értekezések 81. Akadémiai Kiadó, Budapest.
- Szigetvári Péter – Törkenczy Miklós 2006. Pszeudo-Kratülosz. In Kálmán László (szerk.): *KB 120: A titkos kötet. Nyelvészeti tanulmányok Bánréti Zoltán és Komlósy András tiszteletére*. MTA Nyelvtudományi Intézet–Tinta Kiadó, Budapest, 263–275.
- Warren, Richard M. 1970. Perceptual restoration of missing speech sounds. *Science 167*. 392–393.



## A MAGYAR ELDÖNTENDŐ KÉRDŐ MONDAT MEGVALÓSULÁSAI KÜLÖNBÖZŐ BESZÉDMÓDOKBAN

Markó Alexandra

A szakirodalom szerint a magyar, lexikálisan jelöletlen, eldöntendő kérdések egyetlen markere az ún. kérdő intonáció. Azaz mivel semmilyen más (morfológiai, szintaktikai, szórendi stb.) nyelvi jegy nem különbözteti meg az ilyen kérdéseket az állításoktól, a sajátos dallam felelős a kommunikációs funkció azonosíthatóságáért (vö. pl. Deme 1962: 505).

A magyar kérdő intonáció leírásában a szerzők megkülönböztetnek *a)* három vagy több, *b)* két és *c)* egy szótagú formákat (vö. pl. Gósy 2004: 194).

A legalább három szótagos eldöntendő kérdések szökő-eső dallamzárral realizálódnak, ahol az emelkedés az utolsóelőtti, a lelépés az utolsó szótagon valósul meg (vö. pl. Deme 1962: 505–506, Olaszgy 2002: 91).

A két szótagú eldöntendő kérdésekben a szakirodalom szerint „a fellépés is, a lelépés is a második szótagban következik be” (Fónagy–Magdics 1967: 40, ill. vö. Olaszgy 2002: 90), ezzel szemben egy, a spontán beszéd kérdéseit vizsgáló korábbi tanulmányban azt találtuk, hogy a lelépés nem látszik kötelezőnek (Markó 2007: 65).

Fónagy és Magdics (1967) szerint az egy szótagú kérdő mondatban „a lelépés elmarad” (40) – Varga (2002: 137) kísérleti adatai is ezt támasztották alá. Ugyanakkor a spontán közlésekben a lelépés lehetséges (Markó 2007: 64). Olaszgy (2002) szerint „a magánhangzó elején az emelkedés enyhe, majd fokozatosan meredekebb lesz” (89). Az általunk vizsgált realizációk alapján a spontán beszédben 1. a közlés elején nem jelenik meg kötelezően az emelkedés, a nyitás lebegő, sőt ereszkedő is lehet; 2. a meredekség fokozódása sem tart a közlés végéig, lebegő, sőt ereszkedő zárlat is megvalósulhat (vö. Markó 2007: 64).

Akár fonológiai, akár fonetikai nézőpontból közelítik meg a szerzők a magyar eldöntendő kérdő mondatokat, akár egyes mondatok, akár konstruált szövegek (pl. minidialógusok) felolvasatása vagy eljátszatása révén jutnak adatokhoz, akár hallás utáni elemzésre, akár objektív akusztikai adatokra hagyatkoznak, akár eltérnek a megállapításaik bizonyos részleteket tekintve (lásd fentebb), akár nem, az minden korábbi forrásban közös, hogy az eldöntendő kérdés funkciójának kifejeződését a kérdő intonációnak, azaz egy sajátos dallamformának tulajdonítják. Mivel azonban a hangtan elmélete a dallamot hagyományosan (és értelemszerűen) artikulációsan a hangszalagok rez-

gésének következményeként, akusztikailag pedig „az alaphfrekvencia ( $F_0$ ) folyamatos és célzott változtatásának (modulálásának)” (Gósy 2004: 187) eredményeként határozza meg, felmerül a kérdés, hogy hordozhatnak-e kérdő funkciót olyan közlések, amelyeknek létrehozásakor a hangszalagok nem vagy a normáltól eltérően rezegnek. Bár ez a kérdés látszólag marginális, a kérdésre kapott válasz hozzájárulhat a dallam funkciójának, a dallamérzet okainak jobb megismeréséhez, valamint ezáltal a fonetika alapfogalmainak további tisztázásához.

A jelen tanulmányban ezért olyan (a szabály szerint kérdő intonációval ejtendő) eldöntendő kérdéseket elemzünk akusztikai és percepciós nézőpontból, amelyek nem a szokásos zöngéhasználat révén jönnek létre. A vizsgált közlések suttogott, nyelöcsőbeszéddel létrehozott, valamint a kérdő magban ún. érdes (creaky) zöngével realizálódott kérdések. A percepciós tesztlés eredményei megmutatják, hogy a különböző artikulációs eltérések mennyiben befolyásolják a kérdések azonosíthatóságát; az akusztikai elemzés pedig arra irányul, hogy milyen (másodlagos) akusztikai kulcsok teszik/tehetik lehetővé az említett eltérések ellenére a kérdő funkció kifejeződését.

### 1. A kérdő magban érdessé váló zöngé

Ez a csoport olyan kérdő mondatokat tartalmaz, amelyek részben (az ún. kérdő magban – vö. Fónagy–Magdics 1967: 42) vagy egészükben ún. érdes (creaky) zöngével realizálódtak. Ennek a sajátosságnak az oka a hangszalagok rezgésének szabálytalansága (szemben a normál zöngé esetén tapasztalható kváziperiodikus rezgéssel). Az angol nyelvű szakirodalom alapján Böhm (2006: 197) glottalizációnak nevezi a jelenséget, és „érzékelhetően irreguláris hangszalagrezgés”-ként definiálja. Ugyanakkor meg kell jegyeznünk, hogy a nem fonetikus fül számára a rezgésnek ez a fajta szabálytalansága nem feltétlenül érzékelhető. Ezért szerencsésebbnek tartjuk az akusztikai megközelítést, amelynek alapján az érdes zöngé esetén 1. szélsőségesen mélyen futó alaphanggörbét látunk a regisztrátumon (vö. Ní Chasaide–Gobl 1997, Esling–Harris 2003), általában egy indokolatlan ugrás vagy szakadás után (1. ábra), ahol az alacsony frekvenciaértéket a beszélő alaphangjának lokális mélysége nem magyarázza, mert – különösen a női beszélőknél – a mérési tapasztalatok szerint irreálisak ezek az adatok. 2. A periódusok jóval ritkábbak (vö. pl. Böhm et al. 2008). 3. Az ezeken a szakaszokon mért harmonicitás-zaj viszony (az angol nyelvű szakirodalomban *harmonics-to-noise ratio* – *HNR*) átlagos értéke egészen alacsony (egy példában 6,53 dB – szemben a szabályos zöngé esetén mért 14,08 dB-es átlaggal, vö. Markó 2005). A szakirodalom is megerősíti, hogy a zöngé minőségének érzete összefügg a harmonicitás-zaj viszony értékével: minél magasabb a HNR értéke, annál „világosabbnak”, „lágyabbnak”, minél alacsonyabb, annál „durvábbnak”, „érdesebbnek” észleljük a beszélő hangját (vö. pl. Wagner–Braun 2003).



1. ábra

A *Tizenöt?* kérdés rezgéseke és periódusai (fent), valamint dallamgörbéje (lent) – jól látható a zárlatban az érdessé váló zöngre következtében a periódusok megritkulása és az alapvonalra csúszó dallamgörbe

A szakirodalom alapján a glottalizáció egyrészt beszélőfüggő sajátosság: van, akinek a beszédére nagyon jellemző, míg másokéra alig (vö. pl. Redi-Shattuck-Hufnagel 2001, Slifka 2000 – idézi Böhm 2006; a magyarra vö. Markó 2005, Böhm-Ujváry 2008). Másrészt az artikuláció fiziológiájából következően jellemzőbb a szakaszok végén, mint az elején, bár van olyan tanulmány is, amely az amerikai angolra vonatkozóan intonációs frázisok elején találta gyakoribbnak (Redi-Shattuck-Hufnagel 2001 – idézi Böhm 2006). Ez felveti, hogy lehetnek nyelvspecifikus vonatkozásai is, azaz egyes nyelvek beszélői gyakrabban élnek vele, mint más nyelvűek.

Az érdes zöngre megjelenési arányára nézve sokat mondó adat, hogy egy kétórás társalgásban annotált zaj- és érzelemmentes (három vagy több szótagú) eldöntendő kérdések 20%-a (45-ből 9) ezen a módon realizálódott (Markó 2007). Ez a tény és különösen az előfordulási gyakoriság felveti tehát azt a kérdést, hogy ha az eldöntendő kérdés differentia specificája a kérdő intonáció, és az intonációt megtestesítő zöngre torzul, honnan tudja a hallgató, hogy kérdést hallott (ha a kontextus/szituáció ebben nem segíti).

## 2. Suttogott kérdések

A suttogás olyan – természetes – beszédmód, amely közben a hangszalagok nem rezegnek, így beszéddallam sem keletkezik. A hangszalagok suttogó állása azt jelenti, hogy szinte teljes hosszukban bezáródnak, csak a kannaporcok közötti szűk résen áramlik át a levegő. Ez okozza a suttogás zöreijességét és kis intenzitását (részletesebben lásd Gósy 2002 és 2004). A beszéd mégis nagymértékben érthető marad, bár a különböző beszédhangok és fonológiai



jegyek felismerhetősége eltérő (vö. Grácz 2005, 2006). „A hallgató benyomása az, hogy bizonyos magassági változásokat érzékel, tehát például felismeri a kérdést kizárólag a dallam alapján is. Ez úgy magyarázható, hogy az intenzitásváltozás, valamint a formánsok frekvenciájának és sávzélességének alakulása eredményezi a magasságbeli változás élményét” (Gósy 2002: 23). Ugyanakkor esetenként-egyénenként eltérő módon hordozzák ezek a jegyek a kérdő funkciót: formánsemelkedéssel, a formáns magasabb fekvésével, de olyan is tapasztalható, hogy nincs ilyen különbség a kijelentés és a kérdés között, stb. (Fónagy 1968). A korábbi kutatások eredményei tehát a suttogott kérdés kifejezése tekintetében nem egységesek.

Fónagy Ivánnak a „suttogott dallam”-ról szóló közleménye (1968) közismert a fonetikai irodalomban, ezért felmerülhet a kérdés, hogy miért szükséges a suttogott kérdések új vizsgálata. Fónagy is végzett percepciós tesztet, de a tanulmányból nem derül ki, hány adatközlővel és milyen módszerrel; illetve úgy tűnik, hogy az aprólékos akusztikai elemzés eredményeit nem minden esetben hozta kapcsolatba a felismerés sikerességével. Kutatásunkban nem az ő több szempontú elemzését kívánjuk felülbírálni, mindössze nagyobb súllyal vesszük figyelembe a percepciós szempontot, és eredményeinket más, zöngé nélküli beszédmódok elemzési adataival is összevetjük.

### 3. Nyelőcsőbeszéddel létrehozott kérdések

Magyarországon évente több mint négyszáz esetben végeznek teljes gégeeltávolítást. Ezt követően – hangszalagok hiányában – a két leggyakoribb „zöngé”-képzési mód a nyelőcsőbeszéd (latin eredetű kifejezéssel *özofáguszbeszéd*) és a külső gerjesztés, azaz az elektromos gége (a legerterjedtebb készüléktípus gyártója után *SERVOX*-nak is hívják). A nyelőcsőbeszéd megtanulása a betegek számára annak a technikának a begyakorlását jelenti, hogy a gyomorba nyelt levegőt a nyelőcsőn át kiengedve, és a nyelőcső felső szakaszán lévő nyálkahártya-redőzetet mintegy póthangszalagként használva hozzák létre a zöngét helyettesítő alaphangot. A betegek egy része olyan jó minőségű hangot tud így előállítani, hogy a hallgatók úgy érzékelik, mintha csak egy kissé rekedt lenne. A külső gerjesztés használata esetén a beszélő a nyaki lágyrészehez illeszti az elektromos gégét, ez a készülék hoz létre egy állandó (100 Hz körüli) frekvenciájú alaprezgést, miközben a beszélő némán artikulál. Ez a beszédmód mindig gépi hang érzetét kelti a hallgatóban, hiszen nincs mód a természetes beszédre jellemző modulációkra.

A vizsgálatok azt mutatják, hogy ezen beszédmódok esetében a beszédjel akusztikai szerkezete jelentősen megváltozik, így a percepciós sajátosságai is eltérnek a normál fonációjú beszédétől (vö. Balázs–Gósy–Szabó 1996, Gósy 2002, Markó–Grácz–Fent 2007). A nyelőcsőbeszéd gerjesztési sajátosságait, az így előállított közlések, beszédhangok akusztikai szerkezetét, a temporális viszonyokat elemző kutatás már született a magyarra vonatkozóan (Gósy 2002), a kérdő funkció kifejezését, illetőleg a beszéddallamnak megfelelően jegyeket vizsgáló tanulmány azonban még nem.

### **Anyag, kísérleti személyek, módszer**

A vizsgálat sorozat anyagát részben társalgásból származó eldöntendő kérdések, részben a teszteléshez célzottan felvett (felolvasott vagy utánmondott) kérdések adták. A célzott felvételek előnye, hogy kijelentés-kérdés páronként összevethető a közlések akusztikai szerkezete. Ezzel a rögzítési móddal éltünk tehát a suttogás és a nyelőcsőbeszéd esetében. Az érdes zöngé produkciója azonban nem akaratlagos. Bár a gyakorlott beszélők egyfajta „présseléssel” képesek glottalizálni, félő volt, hogy az eredmény túljátszott, természetellenes lesz. Ezért döntöttünk úgy, hogy valódi társalgásból vágunk ki érdes zöngével realizálódott kérdő mondatokat. Így a lejátszott mondatok természetessége biztosított volt, ugyanakkor hátrányt jelentett, hogy ezeken a kérdéseken nem tudtunk kijelentő párjukkal összehasonlítva akusztikai elemzést végezni.

A társalgásfelvétel közel kétórás (113'25"), négy 21 éves egyetemi hallgatóval rögzítettük (2 nő és 2 férfi). A kísérleti személyek mindannyian budapesti születésűek, a köznyelvi sztenderdet beszélik, nem éltek hosszabb ideig sem magyar nyelvjárási területen, sem külföldön. Egynyelvű családból származnak, de mindannyian jól beszélnek két vagy több idegen nyelven. Nincs sem beszédhibájuk, sem ismert hallásproblémájuk. A beszélgetést egy professzionális rádióstúdióknak használt csendesített szobában vettük fel, térmikrofonnal, és közvetlenül számítógépre rögzítettük digitális formában. A mintavétel egycsatornás felvételtől történt, 22 050 Hz-en, 16 biten. Az ezen a felvételen szereplő zajmentes kérdő mondatok akusztikai elemzését egy korábbi tanulmányban (Markó 2007) végeztük el. A jelen vizsgálat sorozathoz ebből a korpuszból választottuk ki a 103 közlést, amelyet egy percepciós teszt keretében minősítettünk hallgatókkal. A spontán kérdések közül csak három vagy több szótagúak kerültek bele a tesztbe (az egy és két szótagos kérdések kis száma miatt). Ebben a tesztben 50 egyetemi hallgató (43 nő és 7 férfi, átlagéletkoruk 21 év, végyesen magyar és más szakosok) vett részt.

A suttogott és a nyelőcsőbeszéddel létrehozott mondatokat egy másik percepciós tesztel vizsgáltuk. A 20-20 tesztmondat (10 állítás és 10 kérdés) nyelvi szerkezete páronként azonos volt (pl. *Rendben van. – Rendben van?; Találkoztak Angélával. – Találkoztak Angélával?; Igen. – Igen?*). A kérdéseket és állításokat egy, a nyelőcsőbeszédet nagyon jó szinten alkalmazó férfi beszélő mondta be, a suttogott változatokat pedig egy ép gégejű férfi beszélő. A véletlenszerű sorrendben lejátszott mondatokat 50 egyetemi hallgató hallgatta meg (46 nő és 4 férfi, átlagéletkoruk ugyancsak 21 év, szintén magyar és más szakosok – a két tesztben részt vett hallgatók egy része azonos volt).

Mindkét percepciós teszt csoportokban zajlott, számítógépről ún. aktív hangfalon keresztül, monóban játszottuk le a hanganyagot. A hallgatók tesztlapot kaptak, amelyen a mondatok helyesírási formában szerepeltek, és arra kértük őket, hogy a mondat után szereplő pont vagy kérdőjel bekarikázásával jelöljék, állításnak vagy kérdésnek érzékelték az elhangzott közlést.



A percepció tesztek kiértékelését, majd statisztikai próbákkal ellenőriztük az eredményeket.

Az érdes zöngével realizálódó kérdések esetében – mivel itt az egyes közlések tartalma és nyelvi szerkezete eltérő (nem tervezett, hanem spontán korpuszról lévén szó) – az akusztikai elemzés eredményeit nem tudtuk a megfelelő kijelentés paramétereivel összevetni.

A suttogott és a nyelőcsöbeszéddel létrehozott mondatok közül azokon végeztünk akusztikai elemzést, ahol az állítás és a kérdés esetében is 70% feletti arányban ismerték fel a hallgatók a közlés modalitását. Így valószínűleg valóban olyan akusztikai jegyeket találhatunk, amelyek szerepet játszanak a kérdés és az állítás elkülönítésében. Az azonos szerkezetű kijelentő és kérdő mondatok esetében vizsgáltuk és összevetettük *a)* a teljes mondat artikulációs tempóját; *b)* a kérdő mag szempontjából kitüntetett (a 3 és több szótagú kérdések esetén az utolsóelőtti) szótag magánhangzójának időtartamát és formánsstruktúráját. A formánsszerkezet paraméterei közül az első 4 formánsnak a magánhangzó tiszta fázisának közepén mért frekvencia-középtértékét és sáv szélességét, valamint a formánsmenetet (az adott formáns frekvenciájának alakulása 6,25 ms-os mintavételezési gyakorisággal) adatoltuk. Az adatok számossága a formánsmenetek esetében tette lehetővé a statisztikai analízist.

A nyelőcsöbeszéd esetében vizsgáltuk az alaphang modulációjának lehetőségét, azaz a „dallam” esetleges megvalósulásait is.

A mondatok akusztikai elemzéséhez a Praat 5.0-s verzióját alkalmaztuk. A statisztikai vizsgálatokat 95%-os konfidenciaszinten az SPSS 13.0 szoftverrel végeztük el.

## **Eredmények**

### **1. A kérdő magban érdessé váló zönge**

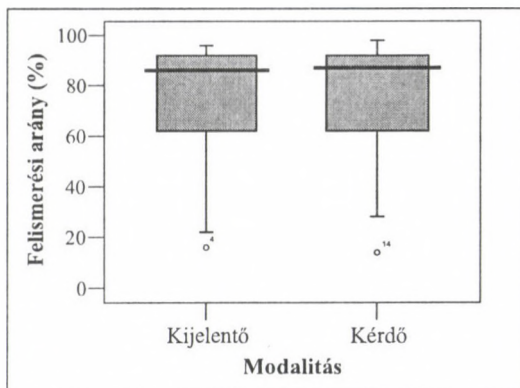
A glottalizált kérdések percepciója a tesztkben 100%-os volt mind a 9 mondat esetében, ami az jelenti, hogy a zönge irreguláris volta nem befolyásolja a kérdő funkció azonosítását. Ez azzal magyarázható, hogy az érdes zönge jellegében megtartja a dallamérzetet, és a hangmagasság változását annak ellenére észleljük, hogy akusztikai elemzéssel nem mutatható ki. A minőségváltozást inkább a hangszínezetben (a beszédhang torzsa válásában) tapasztalja a hallgató. Valamely másodlagos akusztikai kulcs jelenléte ugyancsak feltételezhető.

### **2. Suttogott kérdések**

A suttogott kijelentések és kérdések között sem a felismerési arány átlagában, sem a szóródásában nincs különbség: a kijelentésekre az átlag 71,0% (az átlagos eltérés [SD] = 29,9%), a kérdések felismerése átlagosan 73,0% (SD = 29,5%) – vö. 2. ábra. Ennek megfelelően a párosított *t*-próba sem mutat szignifikáns eltérést. Ez azt jelenti, hogy a kísérleti személyek ugyanolyan arányban észlelték kérdésnek az állítást, mint fordítva (állításnak a kérdést), ami arra utal, hogy az akusztikai szerkezet információhiányai miatt a hallgatók



egyes esetekben tippelnek. Ez a jelenség azonban általánosságban nem rontja nagymértékben az azonosítási arányt, hiszen a felismerési átlagértékek a véletlen találaténál jóval magasabbak.



2. ábra

A modalitás felismerésének sikeressége a szűrt mondatok esetében (szóródás és medián)

A percepció tesztek eredményei alapján választottuk ki azt a 3 mondatpárt, amelyen az akusztikai elemzés eredménye releváns információkhoz vezethet. Mivel szűrés esetén nincs zöng (vagy annak a funkcióját betöltő alaphang), így dallam sem mutatható ki, és az intenzitásgörbe alapján sem lehetett különbségeket találni a kijelentések és a kérdések között, ezért elsősorban a szegmentális szerkezetben kerestünk tendenciaszerű eltéréseket. Az 1. táblázat mutatja be a számszerű adatokat. Az akusztikai elemzést az utolsó előtti szótag magánhangzóján végeztük el, ezeket a táblázat első sorában félkövérrel emeltük ki. Természetesen ilyen kevés mondatpár alapján bármilyen megállapítás csak a felvetés szintjén fogalmazható meg, ennek ellenére a további kutatási irány kijelölésére már ezek az adatok is alkalmasak.

Mindhárom vizsgált kérdés nagyobb artikulációs tempóval realizálódott, mint az azonos szerkezetű állítások. Ez önmagában lehet pusztán a bemondó egyéni beszéd sajátossága vagy a felvétel tényéből adódó jellemző, a tempókülönbséggel együtt ugyanakkor az is látható, hogy a kérdő magbéli magánhangzó – ha csak kismértékben is – hosszabban realizálódott, mint a kijelentő mondatbéli megfelelője. Tehát a magánhangzó a tempógyorsulás ellenére nyúlik meg, ami jelölt pozícióra utal. Különösen az egy szótagú kérdésben tűnik markánsnak ez az eltérés.

A formánsszerkezet paraméterei közül az állításokban és a kérdésekben egymásnak megfeleltethető formánsszépítőértékek ( $F_{1-4}$ ) és a sáv szélességada-

tok ( $B_{1-4}$ ) eltérései nem rendeződnek egységes mintázatba. A formánsmenetek vizsgálata ugyanakkor több jellegzetes különbséget is kimutatott. Az egyes formánsok frekvencia-középértékeit független kétmintás  $t$ -próbával vettük össze. A táblázatban a  $p$  értékek az egyes állítások és kérdések magánhangzóformánsainak statisztikai elemzési eredményeit mutatják ( $p_{F1-F4}$ ). Félkövérrel emeltük ki a szignifikáns eltérést mutató eredményeket. Ezek közül szürke cellákban láthatók azok, amelyeknek az esetében a különbség háttérben az áll, hogy a kérdő mondat magánhangzójának a formánsa magasabban jelentkezik, mint a kijelentő mondaté.

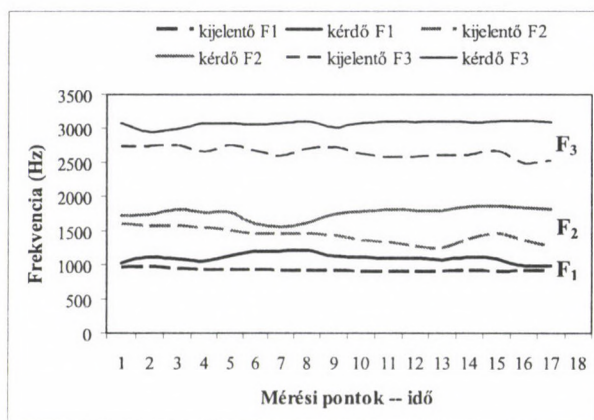
1. táblázat: A 70%-nál nagyobb arányban azonosított szuttogott mondatok akusztikai paraméterei

| Mondat                                                          |                | <i>Csak akkor kell<br/>menned, ha te is<br/>akarod</i> |      | <i>Találkoztak Angé-<br/>lával</i> |      | <i>Zöld</i> |      |
|-----------------------------------------------------------------|----------------|--------------------------------------------------------|------|------------------------------------|------|-------------|------|
| Modalitás                                                       |                | kijelentő                                              |      | kijelentő                          |      | kijelentő   |      |
| Felismerés (%)                                                  |                | 86                                                     | 84   | 96                                 | 92   | 90          | 76   |
| Artikulációs tempó<br>(hang/s)                                  |                | 11,4                                                   | 14,1 | 11,4                               | 13,6 | 5,8         | 6,4  |
| A kérdő mag szempontiából ki-<br>tűntetett szótag magánhangzója | Időtartam (ms) | 61                                                     | 67   | 100                                | 106  | 189         | 226  |
|                                                                 | $F_1$ (Hz)     | 804                                                    | 798  | 921                                | 1084 | 759         | 574  |
|                                                                 | $B_1$ (Hz)     | 177                                                    | 522  | 478                                | 421  | 534         | 511  |
|                                                                 | $p_{F1}$       | 0,316                                                  |      | 0,000                              |      | 0,000       |      |
|                                                                 | $F_2$ (Hz)     | 1350                                                   | 1306 | 1426                               | 1751 | 1638        | 1639 |
|                                                                 | $B_2$ (Hz)     | 229                                                    | 675  | 421                                | 530  | 478         | 591  |
|                                                                 | $p_{F2}$       | 0,583                                                  |      | 0,000                              |      | 0,321       |      |
|                                                                 | $F_3$ (Hz)     | 2528                                                   | 3089 | 2650                               | 3068 | 2486        | 2566 |
|                                                                 | $B_3$ (Hz)     | 223                                                    | 394  | 393                                | 534  | 506         | 506  |
|                                                                 | $p_{F3}$       | 0,000                                                  |      | 0,000                              |      | 0,000       |      |
|                                                                 | $F_4$ (Hz)     | 3802                                                   | 3910 | 4205                               | 4038 | 3998        | 3776 |
|                                                                 | $B_4$ (Hz)     | 428                                                    | 380  | 798                                | 740  | 384         | 384  |
|                                                                 | $p_{F4}$       | 0,267                                                  |      | 0,081                              |      | 0,000       |      |

A formánsmenetek alakulását tekintve minden mondatpárban találunk jellegzetes eltéréseket: az első kérdő mondatban az  $F_3$ , a másodikban az első három formáns, a harmadikban pedig az  $F_1$ , az  $F_3$  és az  $F_4$  mutat szignifikáns eltérést a kijelentőhöz képest. Ezek közül öt esetben a kérdő mondatbeli formáns magasabban fut, mint a kijelentőbeli, és az  $F_3$  az, amelyik mindhárom kérdésben így realizálódik. A *Találkoztak Angélával* mondatok utolsóelőtti szótagbeli  $a$  magánhangzójának első három formánsának alakulását mutatja a 3. ábra.

Ellenőrzésképpen megvizsgáltunk egy olyan mondatpárt is, amelynek a kérdő változatát jóval 50% alatt azonosították kérdő modalitásúnak a hallgatók. Az *Igen?* közlés felismerése 28%-os volt, míg az állító párjáé 96%-os.

Az akusztikai paramétereit tekintve azonban nagyon hasonló a fent bemutatottakhoz. Az artikulációs tempó itt is a kérdésben gyorsabb (a két érték 11,2 és 13,0 hang/s), az *e* hang időtartambeli különbsége (a fenti hosszabb mondatokéhoz hasonlóan) minimális (94 és 97 ms). A formánsmenetek különbsége statisztikailag szignifikáns az  $F_1$ ,  $F_3$  és  $F_4$  esetében (éppen úgy, mint a fenti egy szótagú mondatnál), és mindegyikre jellemző, hogy a kérdésbeli formáns magasabb frekvencián realizálódik, mint a kijelentésbeli. Mindez azt jelenti, hogy ezek az akusztikai paraméterek – bár nagy valószínűséggel felelősek a kérdő modalitás kifejezéséért a suttogott beszédben – önmagukban nem elegendőek ahhoz, hogy a modalitás egyértelműen azonosítható legyen.



3. ábra

Az utolsóelőtti szótagi magánhangzó első három formánsának alakulása a *Találkoztak Angélával* kijelentő és kérdő változatában

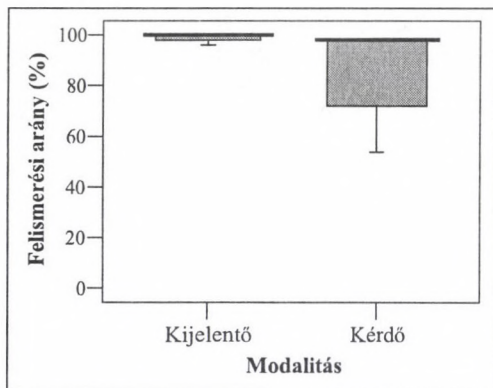
### 3. Nyelőcsöbeszéddel létrehozott kérdések

A nyelőcsöbeszéd esetében közel 15%-os eltérés tapasztalható az állítások és a kérdések felismerésének sikeressége között: az előbbi átlagosan 98,8%-ban ( $SD = 1,8\%$ ), az utóbbit 84,0%-osan ( $SD = 20,2\%$ ) azonosították a hallgatók. A 4. ábra jól mutatja a szóródásbeli eltérést is.

Nyelőcsöbeszéd révén ugyan zöngé nem állítható elő, a jó technika megfelelő automatizálásával mégis jó minőségű és érthetőségű beszéd hozható létre (vö. Balázs–Gósy–Szabó 1996, Markó–Grácz 2007), amelynek vannak többé-kevésbé a periodikushoz közelítő részei. Ennek köszönhetően az akusztikai elemző szoftver megfelelő beállításával a periodikushoz közelebb álló szakaszokon „dallamgörbe” jelenik meg a regisztrátumon (vö. Markó–Grácz megjelenőben). Ezért a három magas felismerési arányszámú mondatpár akusztikai elemzését előbb a „dallam” szempontjából végeztük el. Azt talál-

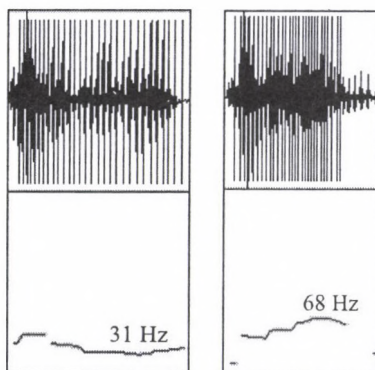


tuk, hogy mindegyik közlés esetében olyan mértékű a nyelvcsőgerjesztés periodicitása, amely alapján auditíven és a szoftver által megjelenítve vizuálisan is valóban dallam érzékelhető (vö. 5. ábra).



4. ábra

A modalitás felismerésének sikeressége a nyelvcsőbeszéddel létrehozott mondatok esetében (szóródás és medián)



5. ábra

A *Rendben van* közlés nyelvcsőbeszéddel létrehozott kijelentő (balra) és kérdő (jobbra) változata: rezgésképp és periódusok (fent), valamint dallamgörbe (lent)

Ezeknek a megvalósulásoknak az esetében tehát nincs szükség kiegészítő akusztikai információkra, a begyakorlott nyelvcsőtechnika önmagában képes a normál zöngéhez hasonló hanglejtsémintázatot előállítására. Ettől az ered-

ménytől függetlenül itt is megvizsgáltuk azokat az akusztikai paramétereket, amelyek (hipotéziseink és a korábbi szakirodalom alapján) másodlagos észlelési kulcsként funkcionálhatnak. Az eredményeket a 2. táblázat foglalja össze. A kérdő mag eltolódása figyelhető meg a *Könnnyű Katát táncba vinni?* mondatban, azaz a hangsúly realizációja – akárcsak a normál fonáció esetében – befolyásolja a dallamstruktúrát.

2. táblázat: A 70%-nál nagyobb arányban azonosított, nyelőcsőbeszéddel létrehozott mondatok akusztikai paraméterei

| Mondat                                                      |                     | <i>A piros toll Danié</i> |       | <i>Könnnyű Katát táncba vinni</i> |       | <i>Rendben van</i> |       |
|-------------------------------------------------------------|---------------------|---------------------------|-------|-----------------------------------|-------|--------------------|-------|
| Modalitás                                                   |                     | kijelentő                 | kérdő | kijelentő                         | kérdő | kijelentő          | kérdő |
| Felismerés (%)                                              |                     | 98                        | 72    | 100                               | 98    | 100                | 98    |
| Artikulációs tempó (hang/s)                                 |                     | 12,4                      | 10,0  | 11,5                              | 12,0  | 11,7               | 12,7  |
| A kérdő mag szempontjából ki-tüntetett szótag magánhangzója | Időtartam (ms)      | 106                       | 99    | 92                                | 79    | 111                | 99    |
|                                                             | F <sub>1</sub> (Hz) | 410                       | 313   | 597                               | 364   | 413                | 369   |
|                                                             | B <sub>1</sub> (Hz) | 119                       | 41    | 362                               | 321   | 323                | 154   |
|                                                             | p <sub>F1</sub>     | 0,000                     |       | 0,017                             |       | 0,592              |       |
|                                                             | F <sub>2</sub> (Hz) | 2138                      | 2168  | 1460                              | 1560  | 1566               | 1528  |
|                                                             | B <sub>2</sub> (Hz) | 163                       | 99    | 196                               | 267   | 703                | 47    |
|                                                             | p <sub>F2</sub>     | 0,186                     |       | 0,157                             |       | 0,619              |       |
|                                                             | F <sub>3</sub> (Hz) | 3254                      | 2869  | 2422                              | 2805  | 2763               | 2762  |
|                                                             | B <sub>3</sub> (Hz) | 1289                      | 639   | 1112                              | 248   | 646                | 259   |
|                                                             | p <sub>F3</sub>     | 0,000                     |       | 0,050                             |       | 0,982              |       |
|                                                             | F <sub>4</sub> (Hz) | 4080                      | 3645  | 3852                              | 4319  | 3996               | 4012  |
|                                                             | B <sub>4</sub> (Hz) | 596                       | 453   | 234                               | 2442  | 701                | 357   |
|                                                             | p <sub>F4</sub>     | 0,092                     |       | 0,012                             |       | 0,896              |       |

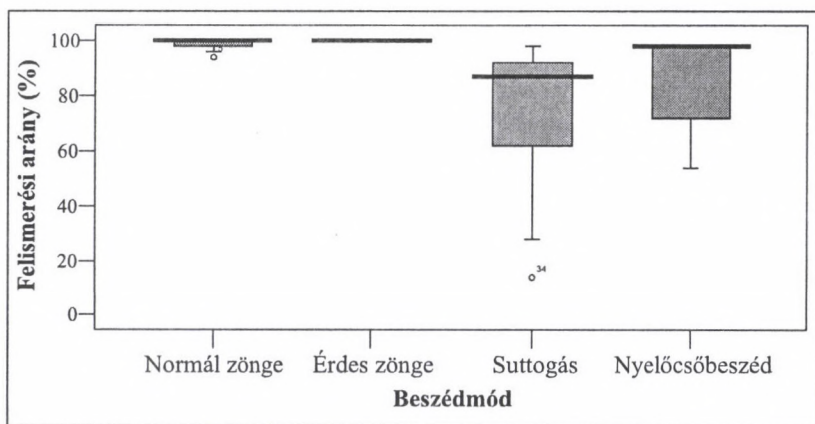
A kijelentő és kérdő mondatok artikulációs tempójának eltérései nem mutatnak szabályosságot. A továbbiakban ezúttal is a dallamcsúcsot hordozó szótag magánhangzójának időtartamát és formánsstruktúráját elemeztük. A kérdésekben a magánhangzó valamivel rövidebb (azaz éppen ellenkező tendencia látszik, mint a suttogásnál). A formánsközéptételek és a sávzsélességs adatok itt sem térnek el valamely modalitás irányában következetesen. A formánsmenetek statisztikai elemzése (független kétmintás *t*-próba) 12-ből 5 esetben adott szignifikáns eredményt (ezeket félkövérrel jelöltük), ugyanakkor csak a második mondat 3. és 4. formánsa tér el a kijelentésben és a kérdésben oly módon, hogy a kérdő mondatbeli formánsmenet magasabban fut, mint a kijelentésbeli.

#### A különböző gerjesztéssel létrehozott kérdések összehasonlítása

Mindegyik kérdéstípus percepciósi eredményeit a szabályos kérdő intonációval realizálódó, érdelemmentes spontán eldöntendő kérdések felismerési

adataival vetjük össze. Ez az érték 98,8% (94 és 100% között szóródik mondatonként, 17 kérdés esetében).

A vizsgált (nem – szokásos – zöngével realizálódó) kérdőmondat-típusok közül a suttogással létrehozottak bizonyultak a legnehezebben azonosíthatóknak. Míg a zöngés fonációval létrehozott semleges kérdések felismerési átlaga 98,8%, és a szórás igen alacsony ( $SD = 1,9\%$ ), addig a suttogott kérdések felismerése átlagosan 73,0%, és a szórásérték magas ( $SD = 29,5\%$ ). Volt olyan suttogott kérdés, amelynek a modalitását mindössze 14%-ban azonosították helyesen az adatközlők – vö. 6. ábra. A nyelőcsőtechnikával előállított kérdések felismerési átlaga 84,0% ( $SD = 20,2\%$ ) volt, itt a legrosszabb eredmény 54%.



6. ábra

A különböző gerjesztéssel létrehozott kérdések percepciósi eredményei (szóródás és medián)

Az egytényezős varianciaanalízis szignifikáns eltérést mutatott ki a csoportok között [ $F(3, 37) = 6,806, p = 0,001$ ]. A post hoc próba a suttogott kérdéseknek a normál szabályos kérdő mondatoktól ( $p = 0,001$ ) és az érdes zöngével realizálódókéktól ( $p = 0,004$ ) való eltérését jelezte szignifikánsnak.

### Következtetések

A percepciósi tesztek eredményei alapján képet kaphattunk arról, hogy a magyar eldöntendő kérdő mondat különböző – normál zöngés, glottalizált, suttogott és nyelőcsőbeszéddel létrehozott – megvalósulásait (kontextustól függetlenül, azaz egyéb háttér-információ nélkül) milyen arányban azonosítják ép halló fiatal felnőttek kísérleti szituációban. Ez természetesen nem jelenti azt, hogy például társalgási helyzetben, ahol a helyzet és a szövegkör-



nyezet segíti a feldolgozást, hasonló eredményeket kapnánk. A vizsgálatunk arra irányult, hogy a rendelkezésre álló, kizárólag akusztikai információ mennyire teszi lehetővé a modalitásazonosítást, illetve (ahol és amennyire lehetett) megkíséreltük meghatározni azokat az akusztikai paramétereket, amelyek a kérdő funkció kifejezéséért (mint másodlagos kulcsok) felelősek lehetnek.

A zöngé érdessé válása nem befolyásolta a kérdő mondat felismerését, ugyanakkor a vizsgált mondatok kis száma és az akusztikai elemzés hiánya miatt csak óvatos következtetések fogalmazhatók meg. Valószínűsíthetjük, hogy az érdes zöngé jellegében megtartja a dallamérzetet, és a hangmagasság változását annak ellenére észleljük, hogy akusztikai elemzéssel nem mutatható ki. Ennek ellenőrzésére további kísérlet(ek)re van szükség, még hozzá a kísérleti technika nagymértékű megújításával: beszédtechnológiai módszerek bevonása, azaz az érdes zöngének szintézis útján történő előállítása (vö. Böhm et al. 2008) lehetővé tenné az állítás és a kérdések összehasonlítását.

Auttogás és a nyelvcsőbeszéd összevetésében jelentős különbségnek tartjuk, hogy míg az előbbiben a kijelentések és a kérdések felismerése között nem volt szignifikáns eltérés, addig az utóbbiban igen. Vagyis míg auttogás esetén ugyanolyan arányban fordult elő, hogy a hallgatók kérdésnek észlelték az állítást, mint fordítva, a nyelvcsőbeszédben erre alig volt példa. Ez többféleképpen is magyarázható. Auttogásban az akusztikai szerkezet információvesztése olyan nagy mértékű, hogy a hallgató nagyobb energiával mozgósítja az intuíciját, és az észlelése olyan paramétereket is kulcsként hasznosít, amelyeket zöngés fonációval létrehozott hangsor feldolgozása esetén figyelmen kívül hagyta. Ezt a hatást a kísérleti helyzet fel is erősítheti. Az is felmerülhet, hogy a bemondó a hangfelvétel minősége érdekében kifejezőbben artikulál, így a szakaszvégek is „feszesebben” hangzanak, amely tulajdonságot a hallgató tudat alatt a kérdésjelleghez asszociálja.

Auttogásban a kérdő magnak megfelelő magánhangzó a vizsgált kérdésekben valamivel hosszabb időtartamban realizálódott, mint a kijelentésekben, miközben a kérdés artikulációs tempója nagyobb volt. A nyelvcsőbeszédben ilyen tendenciát nem láttunk, a magánhangzók a kérdésekben inkább rövidebben realizálódtak, a tempó pedig mondatpáronként eltérően alakult. Ez alapján lehetséges, hogy az időzítési viszonyok a bemondók beszédstílusára vezethetők vissza, de az is, hogy a fonációs mód különbsége testesül meg az eltérésekben. További – több beszélővel rögzített mondatokon végzett – vizsgálat adhat választ a felmerült kérdésekre.

Auttogott kérdések kitüntetett magánhangzójának harmadik formánsa mindhárom vizsgált (nagy arányban azonosított) közlésben magasabb frekvencián realizálódott, mint a kijelentő mondatbeli párjáé, ugyanakkor az ellenpróba (egy kis arányban azonosított kérdés vizsgálata) cáfolta, hogy az  $F_3$  alakulása elegendő lenne a modalitás biztos felismeréséhez.

Mindhárom, nagy arányban felismert, nyelöcsőbeszéddel létrehozott kérdés dallamérzetet idézett elő a hallgatók percepciójában (az elemző szoftver érzékenységeinek növelésével a hanglejtésgörbe a regisztrátumokon is megjeleníthető). Emellett a vizsgált szegmentális szerkezeti jellemzők kevesebb szabályos eltérést mutattak, mint a suttogott mondatokban. Felvetődik, hogy a szegmentális szerkezetben megjelenik egyfajta többlet jelöltség akkor, amikor dallam vagy a zöngét helyettesítő alaphang hiányában a kérdő funkció kifejezése nem lehetséges a szupraszegmentális szinten.

Az eredmények alapján egyértelműnek látszik, hogy (a fonetikai jellemzők közül) nem kizárólagosan a kérdő dallam (amely a hangszalagrezgés adott szabályok szerint történő frekvenciaváltozásában ragadható meg) határozza meg az eldöntendő kérdés kifejeződését. A nyelöcsőbeszéddel létrehozott kérdések nagyarányú azonosítása azt igazolja, hogy a hangszalagok tevékenysége helyettesíthető, és ezzel a technikával a beszéddallam jó minőségű „imitálására” van lehetőség. Nemcsak az észlelés számára megragadható, hanem objektív elemző eszközzel is kimutatható ez a moduláció, ami azt jelenti, hogy bármilyen, a periodikushoz közelítő rezgés – még ha a zöngékváziperiodicitásától távol áll is – alkalmas arra, hogy nyelvi funkciókat fejezzen ki.

Az érdes zöngé esetében ugyancsak dallamot észlel a feldolgozó rendszer – sőt a kérdésazonosítás sikeressége a normál zöngével előállított közlésekével megegyező volt. Itt is arról van szó, hogy a zöngé rezgése irreguláris, ezt az eltérést azonban az akusztikai elemző szoftver oly mértékűnek detektálta, hogy nem volt kimutatható periódus a glottalizált szakaszban. Az érdes zöngé percepciója még feltáratlan (vö. Böhm et al. 2008), de a nagyarányú modalitásfelismerés arra enged következtetni, hogy a dallamérzet megmarad annak ellenére, hogy a fonáció aperiodikus akusztikai eredményt hoz létre.

A suttogás akusztikai szerkezetét és percepcióját már több tanulmányban vizsgálták (a magyarra lásd Fónagy 1968; Gósy 2002; Grácz 2005, 2006 stb.), és néhányuk a nyelvi funkciók kifejezésére is irányult. Az itt vizsgált beszéd/fonációs módok közül a suttogott kérdés észlelése volt a legnehezebb, ami azzal függhet össze, hogy a dallamérzet kiváltása itt lényegében a szegmentális szerkezet feladata, miközben a fonáció jellege miatt már nagyfokú információhiány jelentkezik az akusztikai szerkezetben. Mégis 70% fölötti a modalitás felismerése, tehát a szegmentális szerkezet az esetek nagyobb részében képes átvállalni ezt a funkciót.

Az adatok alapján felvetődik, hogy a magyar eldöntendő kérdő mondat formai jellemzőit a percepció fonetika nézőpontja felől volna érdemes megközelíteni. Ekképpen a vizsgált beszédmódok közös jegyét úgy határozhatjuk meg, hogy az adott közlésben a kérdésre jellemző moduláció észlelhető. Ennek a perceptuális élménynek a hátterében a fonációs módtól függően eltérő artikulációs gesztusok és akusztikai (szupraszegmentális és szegmentális!) paraméterek állhatnak. Ugyan az akusztikai kulcsok pontos és biztos megha-



tározása további vizsgálatokat igényel, az elvégzett kísérletek alapján több olyan kérdés és hipotézis is megfogalmazódott, amelyek kijelölik a kutatás folytatásának irányát.

### Irodalom

- Balázs Boglárka – Gósy Mária – Szabó Iván 1996. A gége nélküli beszéd fonetikai sajátosságai. *Beszédkutatás* 1996. 58–71.
- Böhm Tamás 2006. A glottalizáció szerepe a beszélő személy felismerésében. *Beszédkutatás* 2006. 197–207.
- Böhm, Tamás – Audibert, Nicolas – Shattuck-Hufnagel, Stefanie – Németh, Géza – Aubergé, Véronique 2008. Transforming modal voice into irregular voice by amplitude scaling of individual glottal cycles. *Proceedings of Acoustics '08, June 29-July 4, 2008, Paris*. Paris, 6141–6146.
- Böhm Tamás – Ujváry István 2008. Az irreguláris fonáció mint egyéni hangjellemző a magyar beszédben. *Beszédkutatás* 2008 108–120.
- Deme László 1962. Hangsúly, szórend, hanglejtés, szünet. In Tompa József (szerk.): *A mai magyar nyelv rendszere. Leíró nyelvtan II*. Akadémiai Kiadó, Budapest, 457–522.
- Esling, John H. – Harris, Jimmy G. 2003. An expanded taxonomy of the states of the glottis. In Solé, Maria-Josep – Recasens, Daniel – Romero, Joachim (eds.): *Proceedings of the 15th International Congress of Phonetic Sciences. Barcelona 3-9 August 2003*. Universitat Autònoma de Barcelona, Barcelona, 1049–1052.
- Fónagy Iván 1968. Suttogott dallam? *Magyar Nyelvőr* 92. 253–262.
- Fónagy Iván – Magdics Klára 1967. *A magyar beszéd dallama*. Akadémiai Kiadó, Budapest.
- Gósy Mária 2002. Beszédképzés zöngé nélkül. *Beszédkutatás* 2002. 18–37.
- Gósy Mária 2004. *Fonetika, a beszéd tudománya*. Osiris Kiadó, Budapest.
- Grácz Tekla Etelka 2005. Suttogott mássalhangzók zöngésségének észlelése. *Beszédkutatás* 2005. 78–90.
- Grácz Tekla Etelka 2006. Suttogott ejtésű magánhangzók észlelése. *Beszédkutatás* 2006. 86–100.
- Markó Alexandra 2005. *A spontán beszéd néhány szupraszegmentális jellegzetessége. Monologikus és dialogikus szövegek összevetése, valamint a hűmmögés vizsgálata*. Doktori értekezés. ELTE, Budapest.
- Markó Alexandra 2007. Kérdő funkciójú hanglejtésformák a spontán beszédben. *Beszédkutatás* 2007. 59–74.
- Markó Alexandra – Grácz Tekla Etelka 2007. Gégeeltávolításon átesett betegek beszédének hallgatói megítélése. *Alkalmazott Nyelvtudomány VII/1–2*. 39–55.
- Markó Alexandra – Grácz Tekla Etelka megjelenőben. Az eldöntendő kérdő modalitás megjelenése a nyelöcsőbeszédben. *Alkalmazott Nyelvtudomány VIII*.
- Markó Alexandra – Grácz Tekla Etelka – Fent Zoltán 2007. Különleges zöngékepzési módok hatása az észlelésre. *Magyar Tudomány* 167. évf. 2007/3. 262–264.
- Ní Chasaide, Ailbhe – Gobl, Christer 1997. Voice source variation. In Hardcastle, William J. – Laver, John (eds.): *The handbook of phonetic sciences*. Blackwell Publishers, Oxford, 427–461.



- Olaszy Gábor 2002. A magyar kérdés dallamformáinak és intenzitás szerkezetének fonetikai vizsgálata. *Beszéd kutatás 2002*. 83–99.
- Redi, Laura – Shattuck-Hufnagel, Stefanie 2001. Variation in the realization of glottalization in normal speakers. *Journal of Phonetics* 29. 407–429.
- Slifka, Janet 2000. *Respiratory constraints on speech production at prosodic boundaries*. PhD-disszertáció. MIT, Massachusetts.
- Varga László 2002. Az egy szótagú magyar eldöntendő kérdések intonációja. In Hunyadi László (szerk.): *Kísérleti fonetika – laboratóriumi fonológia 2002*. Debreceni Egyetem Kossuth Egyetemi Kiadója, Debrecen, 127–140.
- Wagner, Anita – Braun, Angelika 2003. Is voice quality language-dependent? Acoustic analyses based on speakers of three different languages. In Solé, Maria-Josep – Recasens, Daniel – Romero, Joachim (eds.): *Proceedings of the 15th International Congress of Phonetic Sciences. Barcelona 3-9 August 2003*. Universitat Autònoma de Barcelona, Barcelona, 651–654.

## A FELOLVASÁS ÉS A SPONTÁN BESZÉD ALAPHANGSZERKEZETEINEK VIZSGÁLATA

**Beke András**

### **Bevezetés**

A beszéd pszicholingvisztikai szempontból két nagy típusra különíthető el. Az egyik a nem spontán beszéd (pl. felolvasás), míg a másik a spontán beszéd. Ez utóbbi produkciója az a folyamat, amely a megszólalás szándékától a kiejtésig tart (Gósy 2005). A spontán beszédben a gondolatok kialakulása, nyelvi formává alakítása, valamint a kivitelezés az adott helyzetben történik meg. Ezzel szemben a felolvasásban a már meglévő, kész szöveg artikulációs kivitelezése történik meg. A beszéd típusa meghatározza a prozódiai megvalósítást. A spontán beszéd létrehozását számos tényező befolyásolja: genetikai adottságok, az artikulációs biztonság, a szókincs nagysága, illetve a hozzáférési folyamatok, az anyanyelvi ismeretek biztonsága, a gyakorlottság, a beszédtema, s végül nem elhanyagolható a szorongás mértéke (Gósy 2004). A spontán beszédben a virtuális mondatok, illetőleg a „bekezdések”, azaz a több virtuális mondatból álló egységek szintaktikailag és szemantikailag meghatározó összefüggésrendszert mutatnak, és a szupraszegmentumok jellegzetes érvényesülésén keresztül különülnek el (Gósy 2003). A szupraszegmentális szerkezet pontos meghatározásával kapcsolatban sok bizonytalanság található mind a hazai, mind a nemzetközi szakirodalomban. A szegmentális szerkezetre épülő, ún. szupraszegmentális (vagy prozódiai) jelenségekre az elmúlt másfél évtizedben nagyobb figyelmet fordítottak, mint korábban (vö. Gósy–Horváth 2007, Markó 2007). A prozódiai jegyek meghatározásával megismerhetők azok a paraméterek, amelyek a beszéd kifejező erejét adják, s amelyekkel érzékeltetjük a mondat fajtáit, a lelkiállapotunkat, és azt is, hogy mit tartunk fontosnak a mondanivalóból vagy mit nem, valamint hogy milyen szituációt akarunk hangban a hallgató elé tárni (Olaszy 2005). A technikai fejlődésnek és a korábbi ismereteknek köszönhetően a spontán beszéd szegmentális és szupraszegmentális szerkezeteinek leírása folyamatosan bővül (Gósy 2003, 2004; Horváth 2006; Markó–Bóna 2006; Markó 2007). A hazai szakirodalomban többen is foglalkoztak különféle megközelítésben a szupraszegmentális szerkezetekkel (Varga 1994; Elekfi–Wacha é. n. [2003]; Markó 2005, Olaszy 1995, 1996, 2005; Imre 2005). A különböző beszéd típusok szupraszegmentális szerkezeteinek – főként temporális – összevetésével többen is foglalkoztak (Bolla 1998, Bóna 2007; Markó 2005); az eltérő beszéd típusok alaphangmagasság struktúrájának átfogó fonetikai összehasonlítása a

magyar beszédben azonban még nem történt meg. Vizsgálatunk ezt a hiányt igyekszik pótolni. Jelen kutatásunk célja az, hogy a felolvasás és a spontán beszéd összevetésével olyan paramétereket határozzunk meg és számszerűsítsünk, amelyek jól jellemzik a két beszéd-típus közti különbségeket az alaphangszerkezet tekintetében. A feltételezésünk az volt, hogy a spontán beszédben a tervezéstől a kivitelezésig tartó folyamat nemcsak a temporális szerkezetet vagy a megakadásjelenségeket érinti, hanem az alapfrekvencia szerkezetére is kihat.

### **Anyag, módszer, kísérleti személyek**

A kísérletben 5 adatközlő spontán beszédét és felolvasását használtuk fel a BEA spontánbeszéd-adatbázisból [2 férfi (F) és 3 nő (N)]. Valamennyien 25–30 év közöttiek, ép hallók és nem beszédhibások. Az adatközlőknek egy 13 kijelentő, érzelmileg semleges mondatból álló összefüggő szöveget kellett felolvasniuk. A szöveg a zöldségekben és a gyümölcsökben felhalmozódott mérgek okozta megbetegedésekről szólt. A felolvasás időtartama a két férfi adatközlőnél 3'51" és 2'51". A három nőnél 1'50", 1'45", 2'15".

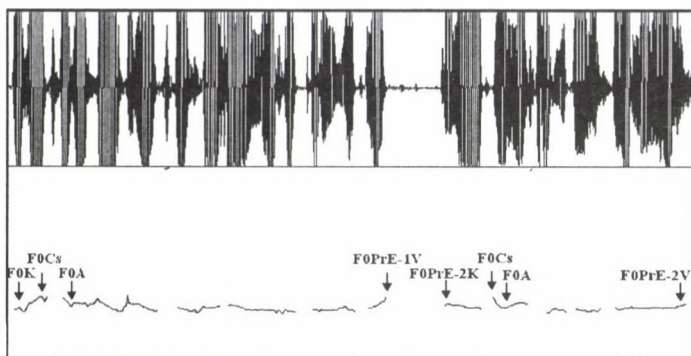
A spontán beszédben az adatközlők a saját életükről, munkájukról beszéltek. A rögzített spontán beszédben a két férfi esetében 9, illetve 16 virtuális mondatot különítettünk el (lásd alább), amelyek időtartama 4, illetve 5 perc volt. A nőknél 16, 12, 13 virtuális mondatot jelöltünk ki, amelyek időtartama 4, 5, illetve 5 perc. Az elemzésekhez a Praat 4.6 beszédelemző szoftvert használtunk, a statisztikai elemzéshez pedig az SPSS 13.0 szoftvert. A statisztikai módszerek közül egytényezős ANOVA-t alkalmaztunk 95%-os szinten, illetve Pearson-féle korrelációanalízist végeztünk.

A jelen kutatásban olyan módszert alkalmaztunk, amelynek segítségével számokkal leírhatóvá és összevethetővé válik a két beszéd-típus közötti különbség. A vizsgálatban a legnagyobb egység, amelyben az alapfrekvencia-szerkezetet vizsgáltuk, a mondat volt, spontán beszédben a virtuális mondat. A mondat alapfrekvencia-szerkezetének leírásában Olasz (2005) módszerét követtük, így a mondatokat „prozódiai egység”-ekre (PrE) osztottuk (1. ábra). Ezeknek a megítélése a beszélő által tartott szünetektől és a szintaktikai tagolástól függött. A felolvasásban a prozódiai egységek legtöbbször akkor fordultak elő, amikor a szövegben is tagoló jelzés (például vessző) volt. A szövegben a PrE-eket pontosvesszővel jelöltük. A PrE határainak megítélése a felolvasásban egyértelmű volt, míg a spontán beszédben Gósy (2003) virtuális mondatokra megadott szempontjait vettük figyelembe; a határok jelölését meghallgatással korrigáltuk.

Az elemzésben az alaphangmagasság mozgását a következő paraméterekkel írtuk le:

- a beszélőre jellemző alapfrekvencia átlagértéke Hz-ben;
- a mondatkezdés alapfrekvenciája ( $F_0K$ ), amely az első hangsúlyos szón mérhető frekvenciacsúcs Hz-értékét jelenti;





1. ábra

*A zöldségekre és gyümölcsökre olyan mérgek kerülhetnek (1. PrE) amelyek még mosás után is hormonális... (2. PrE) mondat prosódiai egységei*

- a mondat befejezésekor mérhető érték ( $F_0V$ ) az utolsó szótagban;
- a mondat belseji prosódiai egységek indítása ( $F_0PrE-xK$ ), ami az alapprofundencia-csúcs értékét jelenti, azaz a prosódiai egységben található az első hangsúlyos szó első szótagján;
- a mondat belseji prosódiai egység befejezése ( $F_0PrE-xV$ ) a prosódiai egység utolsó szótagján;
- a hangsúlyozás az alapprofundencia-görbe kicsúcsosodásaként is realizálódik (Olaszy 2005). A hangsúly függ attól is, hogy mennyi idő alatt következik be. A legtöbb esetben két szótagnyi területen jön létre (Olaszy 2002), így a hangsúly mérését erre a területre terjesztettük ki;
- a hangsúlyokban mérhető alapprofundencia-csúcs értéke ( $F_0Cs$ );
- a hangsúly csúcsa után az alapprofundencia visszatérési értéke ( $F_0A$ );
- a hangsúly ( $H$ ) paraméter az  $F_0Cs$  és az  $F_0A$  értékekből számíthatjuk ki:  $H = F_0Cs / F_0 - F_0A / F_0$ . Minél nagyobb a  $H$  értéke, annál erőteljesebben észlelhető a hangsúly. A hangsúlyozási szint két szélső értékét egy bemondón belül a  $H_{max}$ ,  $H_{min}$  értékekkel adjuk meg;
- a hangterjedelem a beszélő  $F_0$ -értékeinek maximuma és minimuma közötti tartomány;
- a gyakran használt tartomány a hangterjedelemnek az a része, amely a beszélő beszédének a legjellemzőbb frekvenciasávjait tartalmazza. Ennek jellemzésére nagy mennyiségű mintát, azaz  $F_0$ -értéket vettünk a felolvasási anyagból (minden 3 ms-ban mérhető Hz-értéket adatoltunk, így egy beszélőnél átlagosan 3000 mérést végeztünk). A könnyebb összehasonlíthatóság érdekében egy skálát hoztunk létre, amely 70 Hz-től 350 Hz-ig terjedt, s amely 10 Hz-es sávokra van felosztva. A skála minimum és maximum értéke az

adatközlők  $F_0$ -értékeinek szélsőértéke. Azt a frekvenciasávot határoztuk meg gyakran használtkak, amely magában foglalja az  $F_0$ -értékek legalább 70%-át;

– a beszéd alapfrekvenciájának gyakoriságeloszlása azt jelenti, hogy a beszélő hangtartományának hány százalékát használja a leggyakrabban.

– a hangtartomány és a gyakran használt sáv aránya azt jelenti, hogy a beszélő teljes hangtartományán belül mely frekvenciasávokat használja legalább 70%-os gyakoriság mellett.

## Eredmények

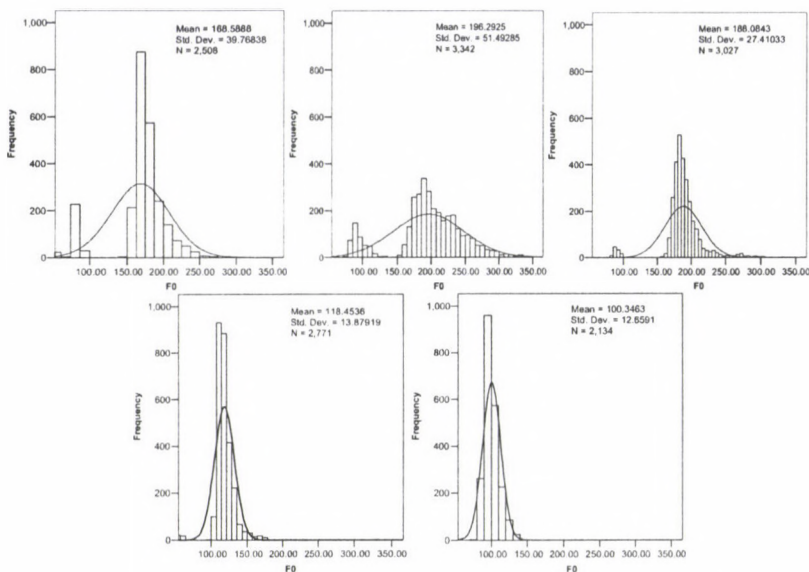
### 1. Felolvasás

Minden beszélőre jellemző egy olyan **hangtartomány**, amelyet a beszédében használ, és amelyet az  $F_0$ -minimum és az  $F_0$ -maximum értékével adhatunk meg. Az elemzések során megfigyeltük, hogy a beszélő hangtartományán belül vannak olyan frekvenciasávok, amelyek gyakrabban használatban vannak, és amelyek függenek a beszéd típusától is (1. táblázat). Az N3 adatközlőnél például ez azt jelenti, hogy a beszédéből 3033  $F_0$ -mintát vettünk, amelyeknek értéke 160–260 Hz közötti; ez tehát a beszélő hangtartománya. Ezen belül a 170–200 Hz-es sávot használja leggyakrabban, vagyis a mért 3033  $F_0$ -minta 94%-a ideesik, ami 2862 mintát jelent.

1. táblázat: A férfi és női hangtartomány, a hangtartományban leggyakrabban használt sáv arányai a felolvasásában

|                                                              | Adatközlők |        |         |         |         |
|--------------------------------------------------------------|------------|--------|---------|---------|---------|
|                                                              | F1         | F2     | N1      | N2      | N3      |
| <b>Hangtartomány (Hz)</b>                                    | 90–170     | 80–140 | 160–260 | 150–290 | 160–260 |
| <b>A leggyakrabban használt tartomány (Hz)</b>               | 100–130    | 80–110 | 160–200 | 170–250 | 170–200 |
| <b>használatának gyakorisága %-ban</b>                       | 85         | 94     | 93      | 96      | 94      |
| <b>A hangtartomány és a gyakran használt sáv %-os aránya</b> | 37,5       | 50     | 40      | 57      | 30      |

A vizsgálatban részt vevő férfiakra és nőkre jellemző beszédfrekvencia-értékek közel megegyeznek a korábban publikáltakkal (Gósy 2004): a nők alaphangmagasságnak értéke 160–260 Hz, míg a férfiaké 80–130 Hz. A méréseink szerint a férfiaké 80–180 Hz, míg a nőké 150–290 Hz. A férfi hangtartomány átlagosan 100 Hz széles sáv, míg a nőké 140 Hz. Sokan megállapították már, hogy nők hangtartománya szélesebb, mint a férfiaké. Azt, hogy ennek a hangtartománynak átlagosan hány %-át használják ki a férfiak és nők, még nem vizsgálták. Méréseink szerint a férfiak a 30%-át használják, míg a nők az 50%-át. Ennek az alapfrekvencia-gyakoriságnak az eloszlása a 2. ábráról olvasható le.



2. ábra

A hangtartományon belüli gyakoriságeloszlás a felolvasásban  
(fent: N1, N2, N3; lent: F1, F2)

A 2. ábrán az is látható, hogy a hangterjedelmen belüli hangeloszlás sokkal egyenletesebb a nők beszédében, azaz jóval több Hz-sávot használnak közel ugyan olyan gyakorisággal. A nők tehát jobban kihasználják a rendelkezésükre álló hangterjedelmet. Az ábrákon az is megfigyelhető, hogy az alacsony frekvenciatarományokban megjelenik egy-egy kisebb csúcs az eloszlásgörbén. Ezek a csúcsok a beszéd érdes zöngéje miatt jönnek létre (ugyanis sok helyen a beszélő alaphangjának lokális mélysége nem indokolja a túlzottan alacsony frekvenciákon való realizációt, mivel ezek az értékek a hangszalagok irreguláris működéséből származnak – vö. Böhm 2006; Markó 2007). A görbe ilyen jellegű kicsúcsosodása nem része a hangterjedelemnek, de jelzi, hogy az adatközlő beszédében milyen mértékben jelenik meg az érdes zöngé.

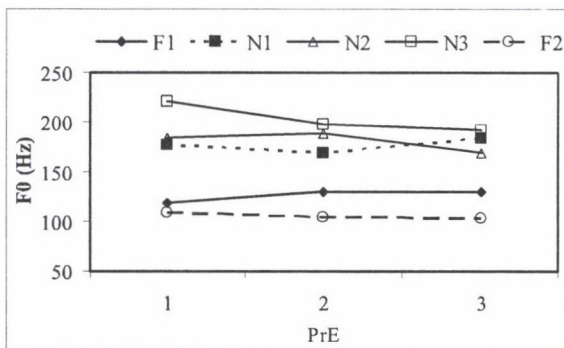
A **mondatkezdő alaphangfrekvencia**, azaz a mondat kezdő szavának első szótagjában mérhető frekvenciaérték attól függ, hogy a mondat első szava információs értékét tekintve mennyire fontos. Az esetek 81%-ában a leghangsúlyosabb szó a mondat elején (az első három szótag valamelyikén) realizálódik. Az alaphangfrekvencia-csúcs 28%-ban a mondat első szótagján valósul meg, ekkor a mondat elején nem találunk névelőt. 42%-ban a mondat második szótagján realizálódik a legmagasabb F<sub>0</sub>Cs, ekkor névelős főnév áll a mondat élén. 10%-ban a harmadik szótagon jelenik meg. A mondat belseje



felé haladva egyre kisebb a valószínűsége annak, hogy ott valósuljon meg az  $F_0$ Cs. Az alapprofrekvencia-csúcs értéke és a mondat szótagjainak helyi értéke között erős a korreláció (Pearson-korreláció: negatív irányú lineáris kapcsolat  $r = 0,607$ ,  $p = 0,008$ ).

A mondatindítások Hz-értékei széles skálán mozognak. A férfiaknál a legmagasabb érték (ekkor hangsúlyos helyzetben van) közel 170 Hz, a legalacsonyabb 118 Hz. A nőknél ez az érték 167–290 Hz.

Olaszy (2005) szerint: ha több prozódiai egységből áll egy mondat, akkor a prozódiai egységeket indító Hz-értékek mindig magasabban realizálódnak, mint a prozódiai egységeket záró Hz-érték. A jelen vizsgálatban számos esetben a mondat belseji prozódiai egységeket kezdő Hz-értékek alacsonyabbak, mint a zárók. A szakirodalom szerint (Olaszy 2005) a mondatot indító  $F_0$ K-értékek sok esetben magasabbról indulnak, mint az utánuk következő PrE-ek  $F_0$ K értékei. A jelen vizsgálatban mért adatok szerint nem mindig ez a forma valósul meg. Ez adódhat a közlők és a szöveg tartalmának különbözőségéből is (3. ábra).

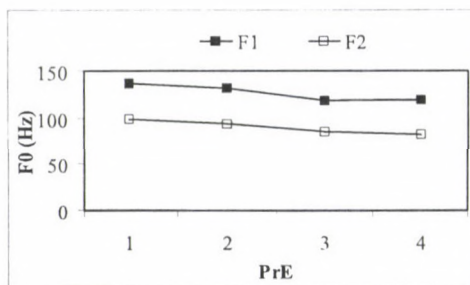


3. ábra

Egy három prozódiai egységből álló összetett mondat kezdő frekvenciaértékeinek viszonyai

Az, hogy milyen Hz-értéken realizálódik az első hangsúlyos szótag kezdete, nem véletlen. Az alapprofrekvencia értéke függ attól, hogy mondatkezdő vagy mondatvégi pozícióban van-e [F1:  $F(1, 49) = 9,296$ ,  $p = 0,004$ ; N1:  $F(1, 49) = 11,707$ ,  $p = 0,001$ ; N2:  $F(1, 49) = 12,618$ ,  $p = 0,001$ ; N3:  $F(1, 49) = 4,526$ ,  $p = 0,039$ ; F2:  $F(1, 49) = 7,633$ ,  $p = 0,008$ ]. A mondatkezdő Hz-érték átlagosan magasabb értéken realizálódik, mint a mondat végső  $F_0$  értéke [férfi:  $F_0K = 120$  Hz;  $F_0V = 206$  Hz; női:  $F_0K = 106$  Hz;  $F_0V = 150$  Hz;  $F(1, 124) = 249,76$ ,  $p < 0,001$ ].

A **mondatvég alaphfrekvenciája**, azaz a mondat utolsó szava utolsó szótagjának magánhangzójában mérhető Hz-érték attól függ, hogy az adott PrE-et követi-e még további ilyen prosódiai egység. Az is befolyásolja, hogy a két PrE (feltételezve, hogy két tagmondat határán van) milyen szintaktikai viszonyban állnak egymással (vö. Deme 1963). A prosódiai egységeket záró Hz-értékek mindig esnek (4. ábra), a PrE-ek számának növekedésével egyre kisebb frekvenciával valósulnak meg, majd az utolsó PrE-et záró frekvenciával közel azonos értéket vesznek fel. Az első prosódiai egységben mért végső  $F_0$ -érték átlaga a férfiaknál 112 Hz, míg a nőknél 185 Hz. A több prosódiai egységből álló mondatban, ahogy a PrE-ek számával haladunk a mondat vége felé, az  $F_0$ -V-értékek átlaga kisebb: férfiaknál 100 Hz, nőknél 143 Hz. A férfiak esetében ez a különbség nem szignifikáns [ $F(1, 49) = 3,638, p = 0,062$ ], míg a nőknél igen [ $F(1, 74) = 9,452, p = 0,003$ ].

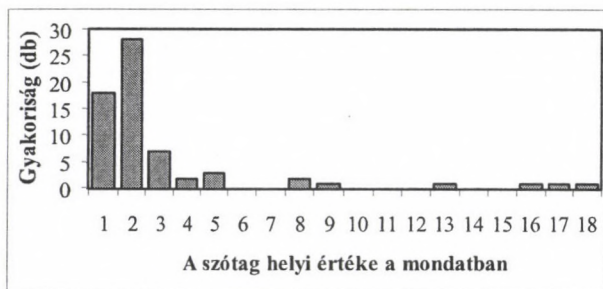


4. ábra

A prosódiai egységek záró Hz-értékeinek csökkenése a mondaton belül

Az **alaphfrekvencia kicsúcsosodása** ( $F_0$ Cs) függ a beszélő tagolási szándékától. Az alaphfrekvencia kicsúcsosodásának értéke egy automatikus folyamat és egy tervezett folyamat kettősségéből adódik (Gósy 2004). Az  $F_0$ Cs-értékek a több PrE-ből álló mondatban mindig ereszkedő mozgást mutatnak. Az első PrE-ekben megvalósuló  $F_0$ Cs-értékek átlaga a férfiaknál 147 Hz, míg a nőknél 277 Hz. Az utánuk következő PrE-ek  $F_0$ Cs értékeinek átlaga férfiaknál 127 Hz, nőknél 251 Hz. Az első prosódiai egység és az utána következő prosódiai egység  $F_0$ Cs értékei között szignifikáns a különbség [férfiak:  $F(1, 49) = 13,534, p = 0,001$ ; nők:  $F(1, 74) = 9,667, p = 0,003$ ].

A legmagasabb  $F_0$ Cs-érték előfordulási gyakorisága egy bizonyos szótagon, valamint a mondat szótagjainak helyi értéke jól korrelál egymással (Pearson-korreláció: negatív irányú lineáris kapcsolat  $r = 0,607, p = 0,008$ ). Az, hogy az  $F_0$ Cs a mondatban milyen gyakorisággal fordul elő egy bizonyos szótagon, függ a szótag helyi értékétől a mondatban. A felolvasásban az esetek 81%-ában a legmagasabb  $F_0$ Cs-érték az első három szótagban valósult meg (5. ábra).

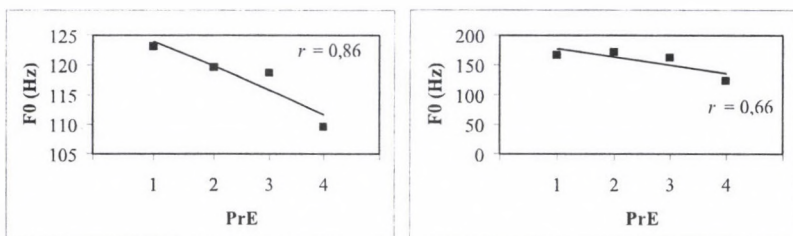


5. ábra

Az alaphfrekvencia kicsúcsosodásának előfordulási gyakorisága a mondatban

A H szerkezet a PrE-ekben nem esik egybe az  $F_0$ Cs-értékek szerkezetével, mivel a H értéke nemcsak az  $F_0$ Cs értékétől függ, hanem attól is, hogy az alaphfrekvencia kicsúcsosodásáról milyen mértékben esik vissza az  $F_0$  értéke ( $F_0$ A). Általában az olyan mondatokban, ahol több PrE van, ott az első PrE-ben van a legnagyobb H-érték, és a továbbiakban egyre csökken az értéke. Előfordul azonban olyan helyzet is, amikor három PrE-ből álló mondat közepső PrE-ének a legnagyobb H-értéke.

A felolvasásban a prozódiai egységek **átlagos alaphfrekvencia-értékei** lineárisan csökkennek a mondat vége felé haladva (6. ábra).



6. ábra

Az átlagos alaphfrekvencia-érték lineáris csökkenése egy négy prozódiai egységből álló mondatban (bal oldalon: N1-es, jobb oldalon: F1-es adatközlő)

A mondatban az első prozódiai egység alaphfrekvenciája a legmagasabb (férfiakra átlagosan 114 Hz, nőknél 192 Hz), az utánuk következő PrE-ek átlagos alaphfrekvencia-értékei csökkennek, ami szignifikáns különbséget mutat (férfiakra 107 Hz, nőknél 178,9 Hz). [ $F_1$ :  $F(1, 24) = 17,974$ ,  $p < 0,001$ ;  $N_1$ :  $F(1, 24) = 6,375$ ,  $p = 0,019$ ;  $N_2$ :  $F(1, 24) = 12,020$ ,  $p = 0,002$ ;  $N_3$ :  $F(1, 24) = 5,506$ ,  $p = 0,028$ ;  $F_2$ :  $F(1, 24) = 8,195$ ,  $p = 0,009$ ]. A PrE átlagos alaphfrek-



vencia-értéke tehát attól függ, hogy a mondatban milyen pozícióban helyezkedik el.

## 2. Spontán beszéd

A spontán beszéd **hangterjedelme** a felolvasáshoz képest átlagosan 60 Hz-cel szűkebb, és ezzel együtt a leggyakrabban használt sáv is szűkebb. A 2. táblázatból leolvasható, hogy a spontán beszédben a férfiak hangtartományuknak 20%-át használják 90%-os gyakorisággal, míg a nők 43%-át. Ez jelentős eltérés a felolvasáshoz képest, amelyet a 7. ábrán is láthatunk. A nők ebben a beszédtypusban is jobban kihasználják a rendelkezésükre álló hangterjedelmet. A spontán beszédben a hangterjedelem kihasználtságának mértéke azért kisebb a felolvasáshoz képest, mert a beszélőnek az a tudatos vagy ösztönös stratégiája, hogy a felolvasás során a mondatok prozódiajának változatosnak kell lennie (Olaszy 2005).

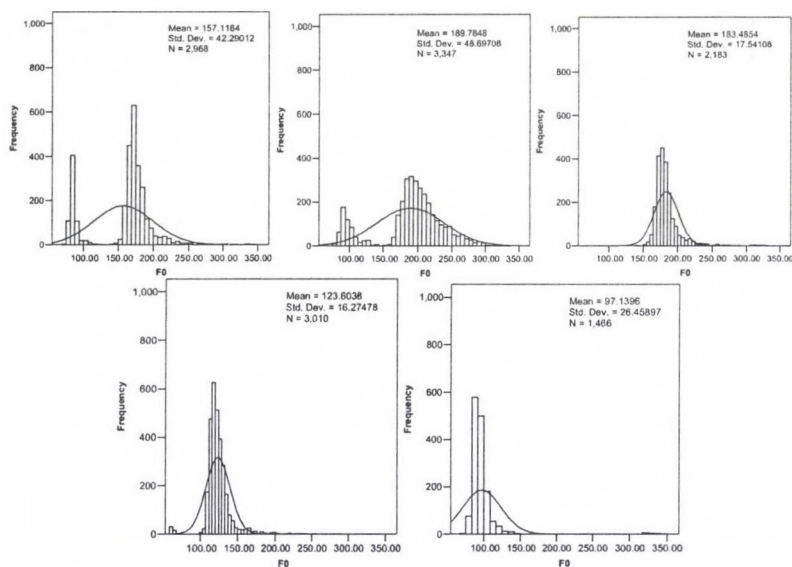
2. táblázat: A férfi és a női hangtartomány és a leggyakrabban használt sávok jellemző határértékei és %-os arányai a spontán beszédben

|                                                       | Adatközlők |        |         |         |         |
|-------------------------------------------------------|------------|--------|---------|---------|---------|
|                                                       | F1         | F2     | N1      | N2      | N3      |
| Hangtartomány (Hz)                                    | 90–170     | 80–120 | 160–260 | 150–290 | 160–260 |
| A leggyakrabban használt tartomány (Hz)               | 100–130    | 80–110 | 160–190 | 150–220 | 160–190 |
| használatának gyakorisága %-ban                       | 93         | 81     | 76      | 82      | 98      |
| A hangtartomány és a gyakran használt sáv %-os aránya | 20         | 33     | 30      | 50      | 30      |

Az eloszlásdiagramokból (7. ábra) számolt eredmények és a szakirodalomban írtak (Rose 2002) alapján elmondható, hogy a beszédtypusból származó különbség mértéke kisebb, mint a beszélők közötti különbség mértéke. Ezért a hangtartományon belüli gyakoriságeloszlás egy lehetséges eszköz lehet a beszélőfelismerésben.

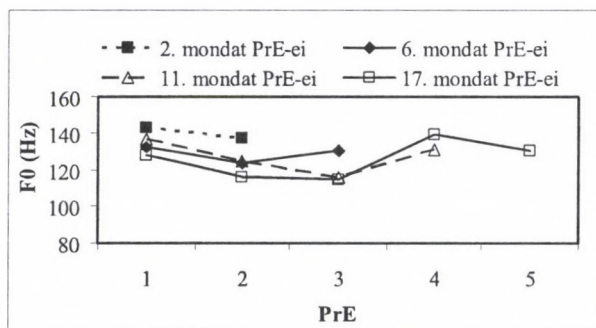
A spontán beszédben a virtuális **mondatot kezdő alaphangfrekvencia-érték** ugyancsak attól függ, hogy hangsúlyos szón található-e. A felolvasásban láthattuk, hogy nagy arányban realizálódik a hangsúly az első szótagon, ezért az  $F_0K$  értéke magas lesz. A spontán beszédben az esetek 22%-ban az első szótagon valósul meg a leghangsúlyosabb rész a virtuális mondatban.

Ebben a beszédtypusban négy dallammenetet lehet tipizálni a virtuális mondatot indító és a virtuális mondatban lévő PrE-eket indító  $F_0$ -értékek viszonyában. Az első dallammenetben az  $F_0K$  értéke a legmagasabb, és a prozódiai egységek kezdő értékei a PrE-ek folyamán egyre csökkennek. Ilyenkor az alaphangfrekvencia-csúcs az első prozódiai egységben az első szótagon van (vö. 8. ábra az F1 beszélő 2. virtuális mondatának PrE-eiben). A második esetben a mondatot indító Hz-érték a legmagasabb, majd a következő PrE-et indító Hz-érték



7. ábra

A hangtartományon belüli gyakoriságeloszlás a spontán beszédben  
(fent: N1, N2, N3; lent: F1, F2)



8. ábra

Az F1 adatközlőnek a mondatot és a prozódiai egységet kezdő Hz-értékei egy 2, 3, 4 és 5 PrE-ből álló virtuális mondatban

csökken, a harmadik PrE kezdő értéke pedig a két érték ( $F_{0K}$  és  $F_{0PrE-2K}$ ) közé emelkedik (8. ábra: a 11. virtuális mondat F1-nél). Ekkor az  $F_{0Cs}$  nem eshet az első szó első szótágjára. A harmadik eset, amikor az  $F_{0K}$ -értéktől

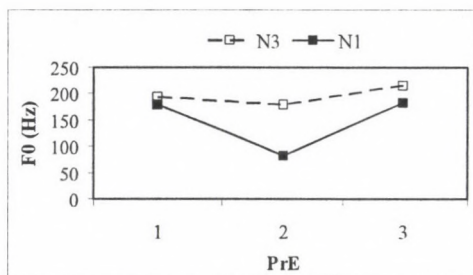
kezdve folyamatosan csökken az  $F_0\text{PrE-xK}$  értéke az utolsó  $F_0\text{PrE-xK}$  értékéig, ahol a virtuális mondat legmagasabb Hz-értékét veszi fel (lásd a 8. ábrán, a 6. virtuális mondat PrE-eiben F1 beszélő esetében). Ilyenkor sem állhat az  $F_0\text{Cs}$  az első szó első szótágján. A negyedik esetben az  $F_0\text{K}$ -tól kezdve ereszkedő tendenciát vesznek fel az  $F_0\text{PrE-xK}$ -értékek, majd elérik az utolsó előtti PrE-ben a legmagasabb virtuális mondatot kezdő Hz-értéket, és az utolsó PrE-ben az  $F_0\text{K}$  és az utolsó előtti  $F_0\text{PrE-xK}$  érték közé ereszkedik az alaphangfrekvencia (8. ábra: az F1 beszélő 17. virtuális mondatának PrE-eiben). Ilyenkor egyik virtuális tagmondatban sem állhat az  $F_0\text{Cs}$  a PrE első szavának első szótágján.

Az  $F_0\text{K}$ -érték változatosságot mutat ebben a beszéd típusban is, és széles skálán mozog. A férfiaknál a legmagasabb érték 157 Hz, ami csak akkor fordulhat elő, ha hangsúlyos szón realizálódik, a legalacsonyabb 94 Hz. A nőknél ez az érték 277 Hz és 97 Hz. A spontán beszédben is sokszor előfordul, ahogy a felolvasásban is, hogy az  $F_0\text{PrE-xV}$ -érték magasabb, mint az  $F_0\text{PrE-xK}$ . A férfiaknál az  $F_0\text{K}$  értéke átlagosan 124 Hz, a  $K_0\text{V}$  értéke 111 Hz [ $F(1, 60) = 5,226$ ,  $p = 0,026$ ]. A nőknél az  $F_0\text{K}$  értéke átlagosan 192 Hz, a  $K_0\text{V}$  értéke 176 Hz [ $F(1; 115) = 3,548$ ,  $p = 0,062$ ]. A nőknél ez a különbség nem szignifikáns.

A spontán beszédben mind a mondatot, illetve a prozódiai egységeket kezdő átlagos Hz-értékek kisebb értéken realizálódnak, mint a felolvasásban. A nőknél a felolvasásában átlagosan 206 Hz, a férfiaknál 120 Hz. A spontán beszédben a nőknél 184 Hz, a férfiaknál 113 Hz (öt beszélőből kettőnél szignifikáns ez a különbség [F1:  $F(1, 62) = 0,770$ ,  $p = 0,783$ ; N1:  $F(1, 68) = 7,569$ ,  $p = 0,080$ ; N2:  $F(1, 64) = 5,899$ ,  $p = 0,018$ ; N3:  $F(1, 56) = 2,560$ ,  $p = 0,115$ ; F2:  $F(1, 48) = 6,242$ ,  $p = 0,016$ ].

**A mondatzáró  $F_0$  értéke** a férfiaknál 70–157 Hz, a nőknél 100–293 Hz között van. Ekkor egy PrE-ből áll a virtuális mondat. Ha több PrE-ből áll, akkor férfiaknál 70–182 Hz, míg a nőknél 100–277 Hz. Az átlagos  $F_0\text{V}$ -értékek férfiaknál 79–129 Hz, nőknél 150–204 Hz. A felolvasáshoz képest a spontán beszédben a virtuális mondatot záró átlagos  $F_0$ -érték magasabb (az öt beszélőből kettőnél szignifikáns a különbség [F1:  $F(1, 62) = 6,831$ ,  $p = 0,001$ ; N1:  $F(1, 68) = 0,595$ ,  $p = 0,443$ ; N2:  $F(1, 64) = 17,066$ ,  $p < 0,001$ ; N3:  $F(1, 56) = 0,660$ ,  $p = 0,807$ ; F2:  $F(1, 48) = 1,928$ ,  $p = 0,172$ ]. Ez azért lehet, mert a spontán beszédben igen fontos kommunikációs szerepe van a virtuális mondatvég kiemelésének; jelzi, hogy a beszélő még folytatni kívánja a közlést (Gósy 2003). A két PrE-ből álló virtuális mondatok legjellemzőbb  $F_0\text{PrE-xV}$  szerkezete ereszkedő, a virtuális mondatzáró  $F_0\text{V}$ -érték felé haladva a  $F_0\text{PrE-xV}$  értéke csökken. A három PrE-ből álló virtuális mondatok  $F_0\text{PrE-xV}$  szerkezetére az jellemző, hogy az első PrE-ben lévő  $F_0\text{PrE-1V}$ -érték után a második PrE-ben a  $F_0\text{PrE-2V}$  értéke csökken, majd a harmadikban az első  $F_0\text{PrE-1V}$ -érték fölé emelkedik. Így egy V alakú mozgást mutat (9. ábra). Ez a típusú szerkezet nagyban különbözik a felolvasásban a három PrE-et tartalmazó mondat  $F_0\text{PrE-xV}$  szerkezetétől.

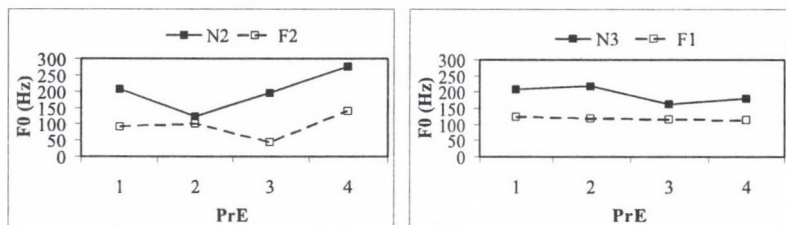




9. ábra

A három prozódiai egységből álló virtuális mondatok prozódiaiegység-záró értékeinek alaphangszerkezete

A négy PrE-ből álló szerkezetekre két típus a legjellemzőbb. Az elsőben a  $F_0$ PrE-xV szerkezet hasonlít a három PrE-et tartalmazó  $F_0$ PrE-xV szerkezet-hez. Hiszen a négy PrE-ű virtuális mondat is V szerkezetű, csak a szélesebb [10.a) ábra]. Ez a típus különbözik a felolvasás négy PrE-et tartalmazó virtuális mondat  $F_0$ PrE-xV szerkezetétől. A másik típusban a  $F_0$ PrE-xV szerkezete lépcsőzetes, és inkább hasonlít a felolvasás négy PrE-ből álló virtuális mondat  $F_0$ PrE-xV szerkezetére [10.b) ábra].

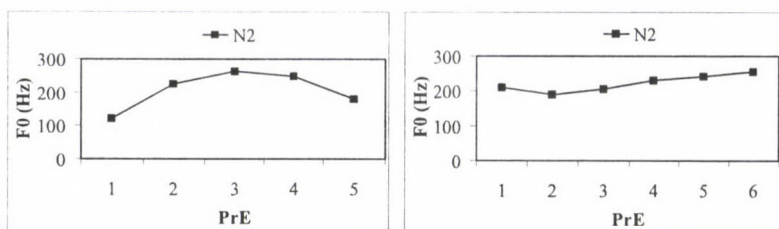


10. ábra

A négy prozódiai egységből álló virtuális mondat prozódiaiegység-záró értékeinek viszonya: a) V szerkezet, b) lineáris

Az öt PrE-et tartalmazó virtuális mondat  $F_0$ PrE-xV szerkezete fordított U alakot ad [11.a) ábra]. A hat PrE-ből álló virtuális mondatok  $F_0$ PrE-xV szerkezete lépcsőzetes, ahogy a négy PrE-ből összetevődőké [11.b) ábra].

Az **alaphfrekvencia-kicsúcsosodás** értéke a spontán beszédben alacsonyabb, mint a felolvasásban. A férfiak átlagos  $F_0$ Cs értéke a felolvasásban 137 Hz, a spontán beszédben 119 Hz, míg a nőké 264 Hz a felolvasásban, 203 Hz a spontán beszédben [ $F_1$ :  $F(1, 62) = 26,065, p < 0,001$ ;  $N_1$ :  $F(1, 68) = 63,645, p < 0,001$ ;  $N_2$ :  $F(1, 64) = 108,579, p < 0,001$ ;  $N_3$ :  $F(1, 56) = 94,149, p < 0,001$ ;  $F_2$ :  $F(1, 48) = 32,517, p < 0,001$ ].



11. ábra

Az a) öt és a b) hat prozódiai egységből álló virtuális mondat prozódiaiegység-záró értékeinek szerkezete

A legmagasabb  $F_0$ Cs-érték előfordulási gyakorisága egy bizonyos szótagon, illetve a virtuális mondat szótagjainak helyi értéke korrelál egymással (Pearson-korreláció: negatív irányú lineáris kapcsolat:  $r = 0,785$ ,  $p < 0,001$ ). Ez azt jelenti, hogy az, hogy milyen gyakorisággal fordul elő a legmagasabb  $F_0$ Cs értéke egy bizonyos szótagon, függ a szótag helyi értékétől a virtuális mondatban. A spontán beszédben az esetek 58%-ában a legmagasabb  $F_0$ Cs-érték az első négy szótagon valósult meg, ami kevesebb, mint a felolvasásban. A spontán beszédben a legmagasabb  $F_0$ Cs-érték beljebb kerül, és az első hat szótag valamelyikén jelenik meg ugyanolyan gyakorisággal, azaz 81%-ban, ahogy a felolvasásban. Ez azért lehet, mert azokban a virtuális mondatokban, ahol beljebb realizálódik a legmagasabb  $F_0$ Cs-érték, ott ezt a szakaszt bevezetés előzi meg, amely nem tartozik szervesen a virtuális mondat tartalmi részéhez, inkább hezitálásnak tűnik. Például az F1 beszélő 10. virtuális mondatában a 6. szótagon realizálódik (félkövér kiemelés) ez a legmagasabb  $F_0$ Cs-érték. Ezt megelőzi egy tartalmatlan, hezitálásnak tűnő rész: *Nem tudom, ugye, olvasni szeretek, filmeket nézni. A Nem tudom, ugye elindítja a közlést, de nem tartozik szorosan a virtuális mondat szemantikai keretéhez. A legmagasabb  $F_0$ Cs-érték attól is függ, hogy melyik PrE-ben valósul meg. Az első PrE-ben realizálódó  $F_0$ Cs-érték mindig az első három szótagban található. A virtuális mondatban egyre beljebb lévő PrE-eknek a legmagasabb  $F_0$ Cs-értéke egyre beljebb kerül a PrE-ben. Ez azzal magyarázható, hogy ilyenkor több olyan szó előzheti meg, amely visszautaló szóként funkcionál a PrE-ben. Például az F2 beszélő 5. virtuális mondatában: **Valószínűleg be fogok jönni vasárnap is dolgozni; mert jövő héten egyik nap nem tudok bejönni; és akkor azt így általában így szoktam megoldani.** A legmagasabb  $F_0$ Cs-érték helyei: 1. PrE-ben az elsőn, 2. PrE-ben a hatodikon, a 3. PrE-ben pedig a tizenharmadikon. Az első PrE-ben az első szótagon 29%-ban jelenik meg a legmagasabb  $F_0$ Cs-érték, ugyanúgy, mint a negyedikén. A második PrE-ben azonban a legmagasabb  $F_0$ Cs-érték előfordulási gyakorisága csökken, csak 20%, míg az ötödiken 30%. Ahogy haladunk a mondat vége felé, úgy csökken az esélye annak, hogy a legmagasabb  $F_0$ Cs-érték az első szótagon valósuljon meg a PrE-ben.*

A PrE-ek átlagos alapfrekvencia-értéke a virtuális mondaton belül lépcsőzetesen csökkenő tendenciát mutat, ugyanúgy, ahogy a felolvasás mondataiban. Emellett olyan szerkezetek is találhatók (leggyakrabban a több PrE-ből álló virtuális mondatokban), amelyben az előzőtől eltérő módon realizálódik a virtuális mondat átlagos alapfrekvenciájának szerkezete. A spontán beszédben a PrE-ek átlagos alapfrekvenciájának értéke alacsonyabb, mint a felolvasásban (a férfiaknál felolvasásban 110 Hz, spontán beszédben 104 Hz; a nőknél 185 Hz a felolvasásban, míg spontán beszédben 174 Hz). [F1:  $F(1, 62) = 0,535, p = 0,468$ ; N1:  $F(1, 68) = 11,760, p = 0,001$ ; N2:  $F(1, 64) = 4,002, p = 0,050$ ; N3:  $F(1, 56) = 10,711, p = 0,002$ ; F2:  $F(1, 48) = 39,177, p < 0,001$ ].

### Összefoglalás

A jelen kutatásunkban számszerűsített adatokkal igazoltuk azt, hogy a spontán beszéd és a felolvasás alapfrekvencia-szerkezetei különbözők. Ezzel bizonyítottuk azt is, hogy az alapfrekvencia szerkezetének megvalósulásában fontos szerepet játszik a beszéd típusa. Alapvető különbségként megállapítottuk, hogy a spontán beszéd hangtartománya szűkebb, és annak kihasználtsága korlátozottabb, mint a felolvasásé. Ez magyarázható azzal, hogy a beszélő automatikusan változatosabban valósítja meg a felolvasás prozódiaját. Kutatásunkban megerősítést nyert az ismert tétel, hogy a nők alapfrekvenciája a teljes és a leggyakrabban használt tartományt tekintve is szélesebb skálán jelentkezik. Elemeztük az alapfrekvencia-szerkezetek tipikus meneteit, átlagértékeit, és azt is, hogy milyen struktúrákkal lehet jellemezni az egyes beszéd-típusokat. A hangtartományon belül megállapítottunk olyan mérőgörbékét, amelyek használható paraméterek lehetnek a beszélőfelismerésben. Ennek igazolására további kutatásokat tervezünk.

### Irodalom

- Bolla Kálmán 1998. Spontán szöveg és kifejező elmondása. *Egyetemi Fonetikai Füzetek* 22. 7–23.
- Bóna Judit 2007. *A felgyorsult beszéd produkciós és percepciós sajátosságai*. PhD-értekezés. ELTE, Budapest.
- Böhm Tamás 2006. A glottalizáció szerepe a beszélő személy felismerésében. *Beszéd-kutatás 2006*. 197–208.
- Deme László 1963. Hangsúly, szórend, hanglejtés, szünet. In: Tompa József (szerk.): *A mai magyar nyelv rendszere. Mondattan*. Akadémiai Kiadó, Budapest, 115–123.
- Elekfi László – Wacha Imre é. n. [2003]. *Az értelmes beszéd hangzása. Mondatfonetika – kitekintéssel a szövegfonetikára*. Szemimpex Kiadó, Budapest.
- Gósy Mária 2003. Virtuális mondatok a spontán beszédben. *Beszéd-kutatás 2003*. 19–43.
- Gósy Mária 2004. *Fonetika, a beszéd tudománya*. Osiris Kiadó, Budapest.
- Gósy Mária 2005. *Pszicholingvisztika*. Osiris Kiadó, Budapest.
- Gósy Mária – Horváth Viktória 2007. Fonetikai elemzések a spontán beszédben. *Beszéd-kutatás 2007*. 7–19.



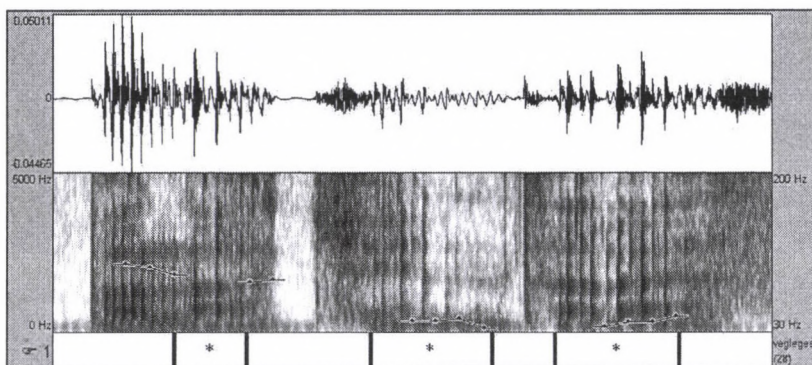
- Horváth Viktória 2006. A spontán beszéd és a beszédfeldolgozás összefüggései gyerekeknek. *Beszéd kutatás 2006.* 134–147.
- Imre Angéla 2005. Különböző műfajú szövegek szupraszegmentális jellemzői. *Magyar Nyelvőr* 129. 510–520.
- Markó Alexandra 2005. A temporális szerkezet jellegzetességei eltérő kommunikációs helyzetekben. *Beszéd kutatás 2005.* 63–77.
- Markó Alexandra – Bóna Judit 2006. A spontán beszéd lejegyzésének néhány kérdése. *Beszéd kutatás 2006.* 124–134.
- Markó Alexandra 2007. Kérdő funkciójú hanglejtésformák a spontán beszédben. *Beszéd kutatás 2007.* 59–75.
- Olaszy Gábor 1995. A kérdés, a figyelmeztetés és a kérdés prozódiaja a kijelentő mondatok tükrében. *Beszéd kutatás 1995.* 46–61.
- Olaszy Gábor 1996. Szabályrendszer prozódiai elemek gépi megvalósításához. *Beszéd kutatás 1996.* 72–78.
- Olaszy Gábor 2002. Az alaphangfrekvencia és a hangsúlyozás kapcsolata a magyarban. In Hunyadi László (szerk.): *Kísérleti fonetika, laboratóriumi fonológia.* Debreceni Egyetem Kossuth Kiadója, Debrecen, 73–84.
- Olaszy Gábor 2005. Prozódiai szerkezetek jellemzése a hírfelolvasásban, a mesemondásban, a novella és a reklámok felolvasásában. *Beszéd kutatás 2005.* 21–50.
- Rose, Philip 2002. *Forensic speaker identification.* Taylor and Francis, London–New York.
- Varga László 1993. *A magyar beszéd dallamok fonológiai, szemantikai és szintaktikai vonatkozásai.* Nyelvtudományi Értekezések 135. sz. Akadémiai Kiadó, Budapest.

## AZ IRREGULÁRIS FONÁCIÓ MINT EGYÉNI HANGJELLEMZŐ A MAGYAR BESZÉDBEN

Bóhm Tamás – Ujváry István

### Bevezetés

A hangszalagok általában majdnem tökéletesen reguláris (kváziperiodikus) rezgése néha irregulárisává válik. Ilyenkor a hangszalagok nem állandó, hanem periódusról periódusra nagymértékben változó időközönként csapódnak össze, egyes periódusok „kimaradoznak”. Akkor beszélhetünk irreguláris hangszalagrezgésről, ha a periodicitástól való eltérés jóval meghaladja a normális jitter vagy shimmer<sup>1</sup> értékeket (1–2. ábra) és ez az eltérés az ép hallású személyek számára határozottan érzékelhető<sup>2</sup>. Irreguláris fonációnak tekintjük azt a jelenséget is, amikor az alapfrekvencia hirtelen jelentősen, a beszélő jellemző hangterjedelme alá csökken és ezáltal érzékelhetően megváltozik a hangszínezet. Így az irreguláris alapperiódusok hossza lehet akár 20–40 ms is, de akár a normál tartományba is eshet.

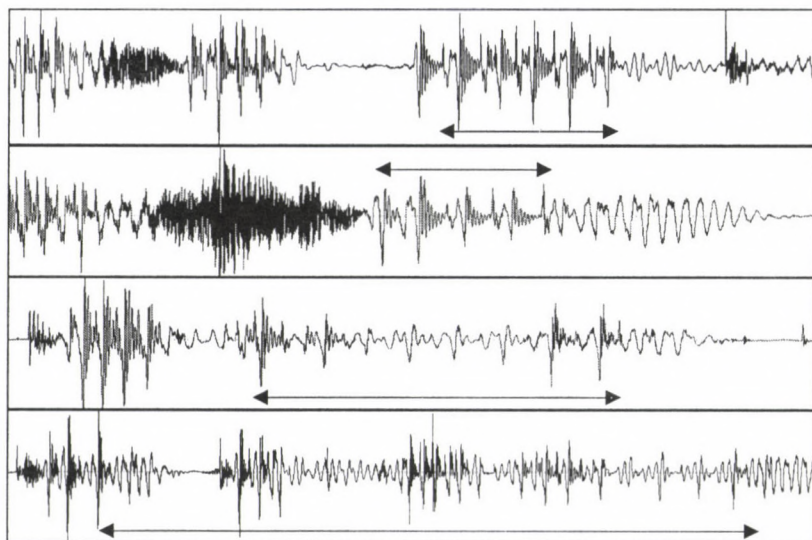


1. ábra

Irreguláris fonáció a *palacsinta lesz* szavak kiejtésében (\*-gal jelölt részek)

1. A jitter az alapfrekvencia, a shimmer az amplitúdó ingadozását méri. Mivel még a reguláris zöngé se tökéletesen periodikus, az alapfrekvencia és az amplitúdó ebben az esetben is kis mértékben ingadozik.

2. A zöngemínőségekben járatlan személyek esetén nem feltétlenül tudatosan.



2. ábra

Irreguláris fonáció férfi bemondások rezgésképeiben (\*-gal jelölt szakaszok)  
(Fentről lefelé: *az ebéd, eső, (Ka)talinnak, (ku)tyákat bevenni.*)

Sajnos definíciónk valamelyest szubjektív, mert nem tudunk olyan határértéket adni az alaphékvencia és az amplitúdó ingadozására, amely objektíven elválasztaná a normális, reguláris hangszalagrezgést az irreguláristól (ha tudnánk, akkor automatikusan detektálható lenne az irreguláris fonáció, a kézi címkézés szükségtelen lenne). Egyrészt még nem tudjuk, hogy a zöngeminőségek természetes kategóriákba sorolhatók, vagy folytonos skálát alkotnak (Gordon–Ladefoged 2001). Az utóbbi esetben nincs éles határ közöttük, így a határértékek keresése értelmetlen lenne. Másrészt az érzékelhető hangszínváltozáshoz különböző beszélők esetén, sőt egy beszélőn belül is, változó jitter és shimmer értékek szükségesek.

Az irreguláris fonációt egyes források *recsegő* (Gósy 2004: 30), *érdes, rekedtes* (Fónagy–Magdics 1967), *nyikorgó zöngé*-nek (Markó 2005a: 61), mások *laringalizált, csikorgó beszéd*-nek (Tóth–Kocsor 2003) vagy éppen *glottalizáció*-nak nevezik (Böhm 2006). Az angol szaknyelv is számos különböző kifejezést használ a jelenségre (pl. *creaky voice, vocal fry, pulsed phonation, laryngealization, glottalization*).

Az irreguláris fonáció produkcióját gyakran a hangszalagok szoros összehúzásával magyarázzák, ami a rezgést instabillá teszi. A tüdőből kiperéselt levegő ezt az erős zárat csak ritkábban és rövidebb ideig tudja felfeszíteni, és akkor se teljes hosszában. Így a reguláris fonációhoz képest jelentősen keve-



sebb levegő áramlik át a hangrésen egy időegység alatt (Laver 1980: 122–126). Slifka (2006) azonban párhuzamos akusztikai és fiziológiai mérésekkel kimutatta, hogy a hangszalagok széthúzásával is lehet irreguláris alaphangot kelteni. Ebben az esetben az átlagos légáram nem csökken, hanem erősödik a reguláris hangszalagrezgéshez képest.

Az irreguláris fonációjú beszédet érdes, rekedtes hangként érzékeljük. Bár a pontos pszichofizikai összefüggéseit még nem ismerjük, a szakirodalomban számos hasonlaltat szemléltették az általa kiváltott érzetet. Többek között forró olaj sercegéséhez (Ishi et al. 2008), kukorica pattogásához (Henton–Bladon 1987) és egy fémkerítésen végighúzott rúd hangjához (Catford 1964 – idézi Laver 1980) hasonlították. Hollien és Wendahl (1968) szerint az egyes glottális impulzusok (hangszalag-összecsapódások) külön-külön érzékelhetők.

Patológiás hangszalag-elváltozások (például aszimmetrikus működés, csomó, bénulás) egyik lehetséges tünete az állandóan irreguláris hangszalagrezgés (Hirano 1981; Gósy 2004: 258). Ebben a tanulmányban normális, egészséges hangszalagokkal képzett beszéddel foglalkozunk. Ezek esetén is előfordul irregularitás, de általában csak időszakosan jelentkezik, és előfordulása összefüggésbe hozható bizonyos nyelvi (szegmentális és szuprasegmentális) és nem nyelvi üzenetekkel (pl. érzelmi töltet). Egyes nyelveken az irreguláris fonáció fonológiailag kontrasztív: például a mexikói jalapa mazatecben (Gordon–Ladefoged 2001) és több szudáni nilusi nyelvben (Laver 1994: 195). A dánban a jelentésmegkülönböztető irreguláris fonációt *stød*-nak hívják (Laver 1994: 330–331). Az amerikai angolban pedig zöngétlen felpattanó zárhangok allofónja lehet (Pierrehumbert–Talkin 1991). Szintén az amerikai angolban gyakran előfordul intonációs frázishatárokon és hangsúlyos szókezdő magánhangzókban (Dilley et al. 1996). Fónagy és Magdics (1967) többféle érzelmet kifejező beszéd esetén tapasztalt érdes, rekedtes zöngét. Gobl és Ní Chasaide (2003) formánsszintézissel kimutatta, hogy az irreguláris fonáció megváltoztathatja az érzékelt érzelmi töltetet.

Nem ismerünk olyan kutatást, amely magyar beszédben az irreguláris fonáció előfordulásának szabályszerűségeit kísérlete volna meg felderíteni. Ez a tanulmány ilyen irányú munkánk módszereit és kezdeti eredményeit ismerteti. Mivel könnyű elsiklani az irreguláris fonáció felett (vagy valamilyen „furcsaságnak”, a felvétel hibájának tekinteni), egyes szakemberek a magyarban elhanyagolhatónak tekintik ezt a jelenséget. Valószínűleg azonban az irreguláris fonáció előfordulása a magyar beszédben nem elhanyagolható, mert egyrészt saját munkánk során gyakran találkoztunk vele, másrészt több tanulmányban találtunk rá utalásokat (Fónagy–Magdics 1967; Tóth–Kocsor 2003; Markó 2005a; Markó–Bóna 2006; Olasz 2007; Markó 2007; Olasz–Bartalis 2008). Markó (2005b) szerint nonverbális közlésekben is előfordul. Néha az – egyébként nem a zöngeminőséggel foglalkozó – tanulmányokban közölt rezgésekepen is látható az irreguláris fonációból adódó akusztikai szerkezet (pl. Gósy 2005, 5d. és 6b. ábra), bár ezeken az érzékelhetőség krité-

riumát nem tudtuk ellenőrizni. Az irreguláris fonációt említő tanulmányokban azonban nem vizsgálták szisztematikusan annak előfordulását, csak a számítógépes beszédelemzést megnehezítő jelenségeként írnak róla. Például dallammenetek kiszámításakor megakadályozza a hanglejtésformák kinyerését (Markó 2005a; Markó–Bóna 2006; Markó 2007), és kényszerített beszédfelismerésnél a hanghatárok hibás bejelölését okozza (Tóth–Kocsor 2003; Olaszy 2007; Olaszy–Bartalis 2008).

A fő kutatási kérdésünk az irreguláris hangszalagrezgés személyfüggőségére vonatkozik. Több angol nyelvre és egy magyar nyelvre végzett vizsgálat is nagy különbségeket mutatott az egyes beszélők fonációs szokásaiban (az angol nyelvű kísérletekben olvasott, míg a magyar nyelvűben spontán beszédet elemeztek). Slifka (2006) kísérletében a négy adatközlő magánhangzóra végződő bemondásai végeinek 0%, 51%, 85% és 85%-ában talált irreguláris zöngéképzést. Redi és Shattuck-Hufnagel (2001) tanulmányában a 14 amerikai személy közül volt, aki a vizsgált pozíciók 88%-ában képzett irreguláris zöngét és volt, aki csak 13%-ában. Dilley és szerzőtársai (1996) öt rádióbemondó szökezdő magánhangzóit vizsgálták, 13% és 44% közötti arányról számoltak be. Henton és Bladon (1987) a 79 brit beszélő közül 10 esetében gyakran tapasztalt irreguláris fonációt, míg másoknál csak ritkán (kvantitatív adatokat nem közöltek). Markó (2005a: 61) egy magyar adatközlő spontán beszédében gyakran, a többi három felvételeiben csak néhányszor észlelt „nyikorgó zöngét”.

Ezekből a tanulmányokból arra következtethetünk, hogy az irreguláris fonáció gyakorisága jellemző a beszélő személyre. A hivatkozott kísérletek célja azonban nem a személyfüggőség vizsgálata volt, így többnyire kevés adatközlővel dolgoztak, és csak a beszélők közötti különbségeket mérték, az egy személyen belüli változékonyságot (azaz, hogy a beszélők több felvételen hasonló arányban használnak-e irreguláris fonációt) nem. Kiemelten foglalkozunk a mondatok végén előforduló irreguláris fonáció személyfüggőségével.

Hipotéziseink tehát az alábbiak: az irreguláris fonáció *a)* a magyar beszéd részét képezi, előfordulása nem elhanyagolható; *b)* előfordulási aránya jellemző egyes személyek beszédére.

### **Anyag, módszer, kísérleti személyek**

Vizsgálatainkhoz 12 személy [2 nő, 10 férfi, életkoruk 23–33 év, kivéve egy férfi beszélőt (F3), aki 64 éves] beszédét rögzítettük. Minden adatközlő felolvasott egy 11 mondatból (7 kijelentő, 3 kérdő, 1 felkiáltó) álló szöveget. A felolvasást rövid szünet után még kétszer megismételték. A szöveget egy korábbi tanulmányból vettük át (Olaszy 2006), a mondatok 5–22 szótagból (átlag: 10,7 szótag) álltak. A félreolvasott vagy megakadást tartalmazó mondatokat a beszélők újraolvasták. A felvételeket hangszigetelt szobában készítettük (44,1 kHz, 16 bit).



A felvételeken bejelöltük az irreguláris zöngével képzett beszédszakaszokat. A címkézést egy akusztikai és egy auditív kritérium alapján végeztük. Az akusztikai kritérium szerint egy beszédszakasz akkor irreguláris, ha az alapperiódusok időtartama vagy amplitúdója hirtelen, periódusról periódusra jelentős változásokat mutat. Szintén teljesíti az akusztikai kritériumot az a szakasz, amelyben az alapfrekvencia hirtelen jelentősen lecsökken. Az auditív kritérium szerint akkor irreguláris egy beszédrésztlet, ha a hangszínezet egy zöngeminőségekben valamelyest jártas szakember számára érzékelhetően megváltozott, a beszéd érdeessé, rekedtessé vált. (Tapasztalataink szerint a témában kevés tapasztalattal rendelkezők könnyen elsiklanak az irreguláris fonáció felett: valószínűleg észlelik a jelenséget, de az esetleg nem tudatosul.)

Azokat a beszédszakaszokat, ahol mindkét kritérium teljesül, \*-gal jelöltük (egyértelműen irreguláris zöngéképzés). Ahol viszont csak az egyik kritérium teljesül, vagy a kritériumok teljesülése nem dönthető el egyértelműen, ott kérdőjelet (?) használtunk. Így például a nagyon rövid vagy nagyon halk szabálytalan hangszalagrezgéseket, amelyek nem eredményeznek érdes hangzást, kérdőjellel jelöltük. A címkék elejét és végét a felvétel tetszőleges időpontjára elhelyezhettük (nem illesztettük a hanghatárokhoz).

Csak azokkal az esetekkel foglalkoztunk, ahol a hangszalagok rezgéséből adódó irregularitás figyelhető meg. A beszédjel ugyanis akkor is eltér a periodicitástól, ha pergőhangot ejtünk vagy ha zaj adódik a zöngéhez (kevert gerjesztésű mássalhangzók vagy levegős fonáció esetén), de a hangszalagok regularisan rezegnek. Ezt nem tekintjük irreguláris fonációnak. Viszont ha a zaj irreguláris zöngéhez adódik, akkor a fonáció irreguláris. Erre utal, ha a zöngperiódusok időzítésében és amplitúdójában olyan mértékű hirtelen változások láthatóak, amelyek nem magyarázhatóak pusztán hozzáadott zajjal (pl. Gósy 2005, 5d. ábra). Ebben az esetben a beszédrésztlet teljesíti az akusztikai kritériumot.

A címkézést Praatban végeztük, ahol a címkésáv mellett a rezgésképet, a spektrogramot és a dallammenetet jelenítettük meg (1. ábra, a Praat által számított dallammenetet a spektrogram sávjában ábrázoltuk). Az akusztikai kritérium ellenőrzéséhez elsődlegesen a hullámformát és a spektrogramot használtuk. Az alapfrekvencia-görbe csak kiegészítő információként szolgált, mert ez irreguláris fonáció esetén az átlagosnál jóval több mérési hibát tartalmaz. A máshol is gyakran előforduló oktávhibák mellett a Praat algoritmus sokszor zöngétlennek osztályozza az irreguláris részeket vagy az első formánst tekinti alapfrekvenciának.

A következetes címkézés érdekében konszenzusos eljárást alkalmaztunk: a korpusz összes felvételén mindkét szerző egy rögzített kritériumrendszer alapján megjelölte az irreguláris beszédszakaszokat, majd a párhuzamos címkéállományokat összehasonlítottuk. Az összehasonlítás során az eltéréseket közösen kielemeztük, és kialakítottuk a végleges címkéket. A konszenzusos címkézési eljárással számos figyelmetlenségi hibát ki tudtunk szűrni.



Az esetek túlnyomó többségében mindketten ugyanazokat a beszédrészleteket jelöltük irregulárisnak. Viszont az irregularitás pontos kezdetében és végében általában kisebb-nagyobb eltérések voltak. Ezek az eltérések nem meglepőek, ha figyelembe vesszük, hogy egyrészt a címkék határait tetszőleges időpontokra helyezhettük, másrészt az irreguláris fonációnak (a beszédhangokhoz hasonlóan) nincs éles határa. Ugyanezen két okból egy másik típusú eltérést is tapasztaltunk a címkék összehasonlítása során: ha egy hosszabb beszédrészletben rövid megszakításokkal több irreguláris szakasz is található, azt néha egy összevont címkével, néha több egymás utáni címkével jelöltük.

Az irreguláris fonáció mennyiségét mérhetnénk a címkék összsidótartamában vagy a címkék számában. Ezek a mértékek azonban a címkék határainak és összevonásának bizonytalansága (lásd előző bekezdés) miatt meglehetősen megbízhatatlanok. A két említett bizonytalansági tényező egy-egy címke esetén legfeljebb néhányszor tíz ezredmásodpercnyi eltérést okozott, így az irreguláris fonációt tartalmazó szótagok<sup>3</sup> megszámlálása megfelelő mértéknek tűnik. Az adatelemzés során minden címkéhez megállapítottuk, hogy hány szótagot érint. Egy címke akkor „érint” egy szótagot, ha legalább egy zöngperiódusnyi része a szótagra esik. Így előfordul, hogy egy rövidebb címke csak egy szótagot érint, míg egy hosszabb többet is (amelyikben kezdődik, amelyeken áthalad, és amelyikben véget ér).

A további feldolgozás megkönnyítése érdekében a végleges címkéket egy táblázatban összesítettük. A táblázat minden címkéhez tartalmazza az adatközlő nevét, a bemondás sorszámát (első, második vagy harmadik), a címke típusát (\* vagy ?), hosszát (szótagban), mondatbeli pozícióját (kezdő, belső vagy záró), valamint hogy milyen modalitású mondatban fordult elő. Hosszabb címkék esetén egyszerre többféle pozíció is lehetséges: például, ha egy rövid mondat teljes hosszában irreguláris, akkor a megfelelő címke egyszerre mondatkezdő, mondat belseji és mondatzáró. Egy címkét akkor tekintettünk mondatkezdőnek, ha az érinti legalább egy mondat első szótagját. Mondatzárók azok a címkék, amelyek érintik egy mondat utolsó vagy utolsóelőtti szótagját.

Az irreguláris fonáció mellett alapfrekvenciára vonatkozó méréseket is végeztünk. Az alapfrekvencia átlaga és tartománya (a hangterjedelem) nagymértékben jellemző a beszélő személyére (Rose 2002). Így összehasonlítási alapként szolgálhat, valamint vizsgálható az egyes személyek jellemző alapfrekvenciája és fonációs szokásai közötti összefüggés.

Minden adatközlő mindhárom felolvasására külön-külön kiszámítottuk az alapfrekvencia átlagát és szórását (a teljes felolvasásra vonatkoztatva). A mértést Praattal végeztük, nők esetén 100–300 Hz-es, férfiaknál 50–250 Hz-es keresési tartományt beállítva. Az irregulárisként címkézett beszédrészleteket

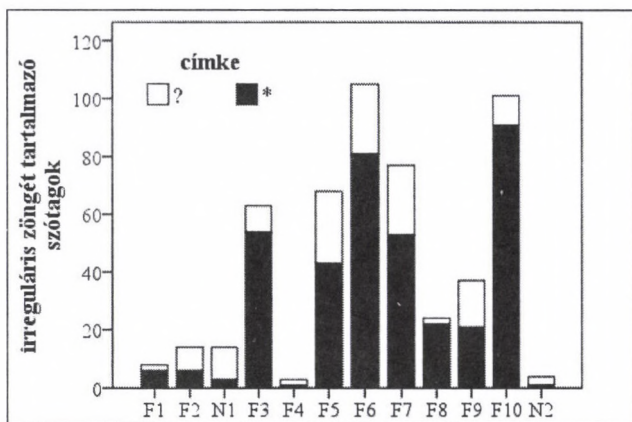
3. A szótagolást a magyar akadémiai helyesírás szabályai szerint végeztük.

előzetesen kitöröltük a felvételekből. A hangterjedelmet nem a dallamgörbe legmagasabb és legmélyebb pontjának különbségeként számoljuk, mert ez nagyon érzékeny az alapfrekvencia-számítás hibáira (pl. oktávhiba). Ehelyett az összes mért  $F_0$ -érték szórását használjuk. Az adatokat SPSS 16.0-val elemeztük.

### Eredmények

518 szótagban, az összes szótag 12%-ában fordult elő irreguláris fonáció. Ezek többsége, az összes szótag 9%-a, egyértelmű eset volt (\* címke). Az előfordulási arányok meglepően magasak, ha figyelembe vesszük, hogy egy eddig többnyire figyelmen kívül hagyott jelenségről van szó. A hangszalagok fáradásával megváltozhat a zöngeminőség (Linville 1995), de esetünkben nem ez okozza a magas előfordulási arányt: a három felolvasás során nagyjából azonos számú szótagban fordult elő (160, 188 és 170 szótag). Tehát eredményeink azt mutatják, hogy az irreguláris zöngékepzés nem ritka jelenség magyar beszédben: az általunk vizsgált beszédkorpuszban átlagosan minden 11. szótagban előfordult ez a fonációs típus (csak az egyértelmű esetekkel számolva).

A 3. ábrán látható, hogy az irreguláris fonáció előfordulási aránya nagymértékben eltérő az egyes adatközlők esetén. Öt beszélő esetén gyakran, másik öt esetén pedig ritkán fordult elő ez a zöngékepzési mód (átlagosan 63,6, illetve 8,6 szótagban; a maradék két adatközlő 27,0 szótagos átlagához viszonyítva harmadannyi, illetve több mint kétszer annyi szótagban).

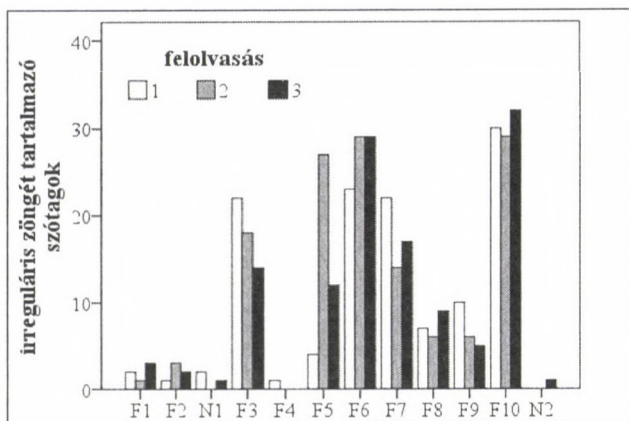


3. ábra

Az irreguláris zöngét tartalmazó szótagok száma személyenként  
(Nők: N1–N2; férfiak: F1–F10.)

A különbségek még markánsabbak, ha csak az egyértelmű (\*-gal jelölt) címkéket tekintjük (F8 és F9 adatközlők 18,5 szótagos átlagához viszonyítva öt beszélő esetén kevesebb mint ötödannyi, átlagosan 3,4 szótagban fordult elő irreguláris fonáció, a másik öt beszélő esetén több mint két és félszer annyi, átlagosan 47,8 szótagban). Ennek az az oka, hogy az alacsony előfordulási arányokkal rendelkező adatközlők esetén a kevés címkének is csak egy töredéke egyértelmű. Így a személyfüggőség további vizsgálatához csak a \*-gal jelölt címkéket elemezzük. F10 beszélő a kiejtett szótagjai 26,0%-át, míg N1 és F4 csak 0,3%-át képezte irreguláris zöngével.

Azonban a beszélőfüggőség megállapításához az is szükséges, hogy egy-egy személy következetesen használjon vagy ne használjon irreguláris fonációt. A 4. ábra alapján a beszélők a három felolvasás során hasonló arányban képeztek irreguláris zöngét. 9 személy esetén a három bemondás nagyjából ugyanannyi irreguláris szótagot tartalmaz, míg 2 személy (F3 és F7) esetén közepes, egy személy (F5) esetén pedig nagyobb eltérések láthatóak.



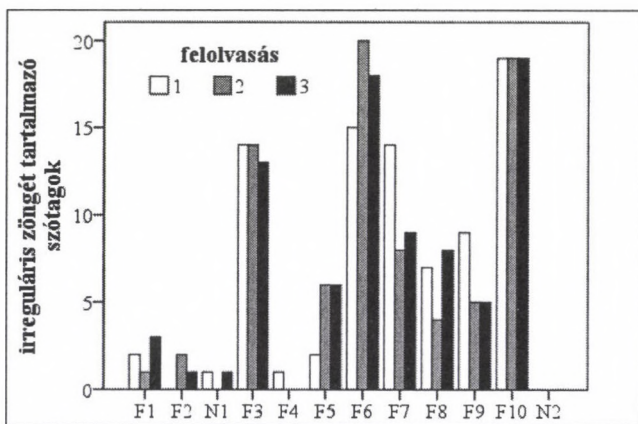
4. ábra

Az irreguláris zöngét tartalmazó szótagok száma személyenként  
(a három felolvasásra bontva, csak egyértelmű címkék)

A Bevezetésben már említettük Slifka (2006) eredményét, mely szerint irreguláris fonációt kétféle módon is lehet képezni: az egyik esetén csökken, a másik esetén megnő az átlagos légáram a hangrésnél. Slifka (2007) szerint a gyenge légárammal képzett irregularitás feltételezhetően akaratlagos, és így például nyelvi információ kifejezésére alkalmas. Ezzel szemben az erős légárammal járó eset a bemondások végén jellemző, és valószínűleg nem akaratlagos. Így elképzelhető, hogy a beszélő személyek sajátosságai az ilyen típusú irreguláris fonáció előfordulását vizsgálva jobban megfigyelhetők. Az 5. ábra



ezért csak a mondatok végén levő irreguláris fonáció előfordulási arányait ábrázolja (a 192 mondatvégi címke közül 20 darab mondat belsejében lévő szótagot is érintett). Látható, hogy F3 és F5 esetén a 4. ábrához képest nagyjából eltűntek a három felolvasás közti különbségek. Ezek alapján a mondatvégi irreguláris fonáció előfordulási aránya valószínűleg jellemző a beszélő személyére.



5. ábra

Az irreguláris zöngét tartalmazó mondatvégi (utolsó vagy utolsóelőtti) szótagok száma személyenként (a három felolvasásra bontva, csak egyértelmű címkék)

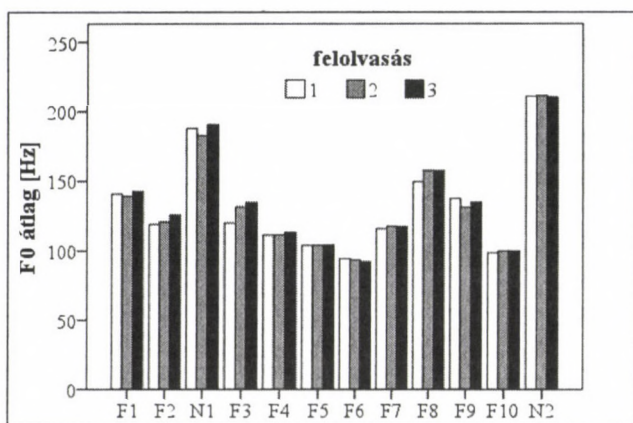
Az eredmény egy lehetséges magyarázata a mondatzáró pozíció fiziológiailag kitüntetett szerepe. A mondat végén ugyanis lezárul egy intonációs frázis, a beszédképzés hosszabb-rövidebb időre szünetel, gyakran egyben a kilégzés végét is jelenti, és belégzés követi. A mondat végéhez közeledve így a hangszalagok zöngéállásból fokozatosan nyugalmi állapotba (légzőállásba) mennek át. A fonáció fenntartásához szükséges három feltétel (glottális nyomáskülönbség, a hangszalagok feszessége, valamint azok összeszorítása) fokozatosan megszűnik (Stevens 1998: 80–82). Ekkor olyan köztes állapotba kerülhetnek a hangszalagok, amely azok irreguláris rezgését eredményezheti, és ez az állapot viszonylag hosszú ideig fennállhat. Elképzelhető, hogy az egyes zöngéképzési feltételek megszűnésének sorrendje személyenként változó, így

4. A zöngéállás a mondat ejtése során nem folyamatos, zöngétlen hangok képzésekor is szétnyílnak a hangszalagok, de csak átmenetileg, rövid időre.

egyes beszélők esetén általában előfordul irreguláris fonációt lehetővé tevő állapot, míg másoknál nem.

Bár a mondatok elején ugyanez a folyamat zajlik le időben visszafelé, itt általában nem figyelhető meg irreguláris fonáció: a közlési szándék miatt a nyugalmi állapotból a hangszalagok nagyon gyorsan váltanak zöngéállásba.

Az átlagos alapfrekvenciát – többek között – markáns személyes ismertetőjegyként tartják számon (Rose 2002). Saját adataink is ezt támasztják alá: a 6. ábrán látszik, hogy mindegyik kísérleti személy esetén a három felolvasásra szinte azonos átlagos  $F_0$ -értéket kaptunk. A hangterjedelem tekintetében – amelyet itt az  $F_0$  szórásával mérünk – nagyobb, de még mindig többnyire kicsi vagy közepes különbségek jelentkeztek a három felvétel között (7. ábra).



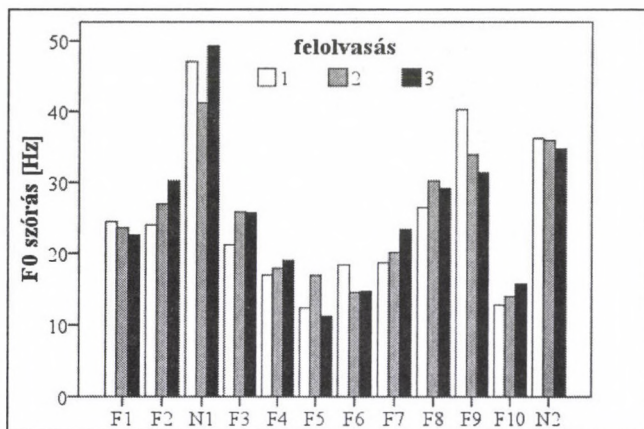
6. ábra

Az alapfrekvencia átlaga személyenként, a három felolvasásra

Az ábrák alapján az átlagos  $F_0$  a beszélő személy egy jóval erősebb ismertetőjegye, mint a mondatvégi irreguláris fonáció. A hangterjedelem és az irreguláris fonáció azonban hasonló személymegkülönböztető képességű, mert hasonló mértékűek a beszélők közötti és a beszélőn belüli különbségek is.

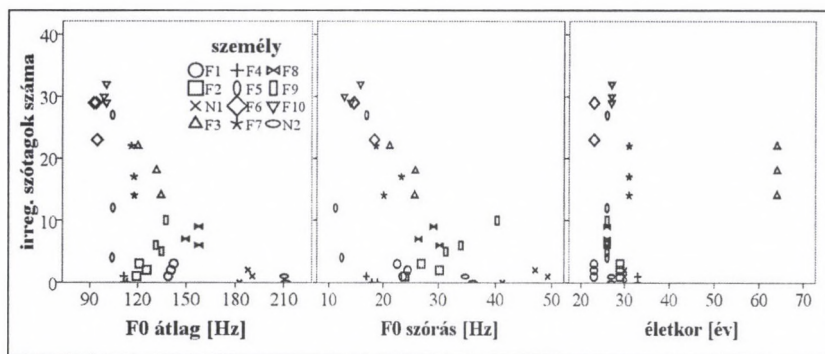
A 8. ábra a két  $F_0$ -statisztika, valamint a beszélő életkora és az irreguláris fonáció összefüggését mutatja. Az irreguláris fonáció közepes mértékű negatív, szignifikáns korrelációt mutat az átlagos alapfrekvenciával ( $r = -0,621$ ;  $p < 0,0005$ ) és a hangtartománnyal is ( $r = -0,573$ ;  $p < 0,0005$ ). Ezek szerint egy mélyebb hangú személy beszédében várhatóan többször találunk irreguláris fonációt, mint egy magasabb hangúéban. Ez összhangban van a foniátriában patológiás beszédre kapott eredményekkel. Az irreguláris zöngé előfordulása nem korrelál a beszélő személy életkorával ( $r = 0,122$ ; nem szigni-

fikáns). Így az egyéni különbségeket feltételezhetően nem a hangszalagok öregedése, fáradékonyága okozza.



7. ábra

Az alapfrekvencia szórása személyenként, a három felolvasásra



8. ábra

Az irreguláris szótagok száma, az  $F_0$ -átlag, az  $F_0$ -szórás és az életkor szórásképei (minden pont egy felolvasásnak felel meg)

### Következtetések

Eredményeink alátámasztották hipotéziseinket, de azok hitelt érdemlő igazolásához további vizsgálatok szükségesek. A vizsgált magyar nyelvű beszédkorpuszban gyakran előfordult irreguláris fonáció, így ez nem tekinthető marginális jelenségnek. Az eredmények azt mutatták, hogy összhangban a



Bevezetésben említett angol kutatásokkal, a magyarban is valószínűleg jellemző az irreguláris fonáció gyakorisága az egyes beszélőkre, különösen a mondatok végén. Ugyanannak a szövegnek a három felolvasását összehasonlítva, az egyes adatközlők mindhárom alkalommal a kiejtett szótagok körülbelül ugyanakkora hányadát képezték irreguláris zöngével. Az adatok alapján az irreguláris fonáció az átlagos alapfrekvenciánál kisebb, de a hangterjedelemből hasonló mértékben jellemző a beszélőre (ennek megerősítéséhez további, kvantitatív kísérleteket kell végezni). Az irreguláris fonáció a beszéd-korpuszunkban a mély hangú személyekre jellemzőbb, mint magasabb hangokra (de a korreláció csak közepes mértékű).

Ez a tanulmány egy hosszabb vizsgálat kezdeti eredményeit ismerteti. Lehetséges, hogy az adatközlők pillanatnyi egészségi állapota vagy pihentsége befolyásolta eredményeinket. A jövőben így ugyanezekkel a személyekkel további felvételeket szükséges készíteni, hogy kizárhassuk ezeket a tényezőket, és összehasonlíthassuk az adatközlők beszédét néhány hónap eltéréssel. További női beszélők vizsgálata pedig nemek közötti összehasonlításokat tenne lehetővé (pl. az alapfrekvenciával való összefüggés a két nemre külön vizsgálható lenne). Felmerül az a kérdés is, hogy mennyire általánosíthatóak az eredmények spontán beszédre.

Eredményeink azt mutatják, hogy az irreguláris fonáció mint egyéni hangjellemző valószínűleg felhasználható a gépi beszélőfelismerésben, a beszélmódosításban és az igazságügyi hangazonosításban.

## Irodalom

- Böhm Tamás 2006. A glottalizáció szerepe a beszélő személy felismerésében. *Beszédkutatás* 2006. 197–207.
- Catford, John C. 1964. Phonation types: the classification of some laryngeal components of speech production. In Abercrombie, David – Fry, Dennis B. – MacCarthy, Peter A. D. – Scott, N. C. – Trim, John L. M. (eds.): *In honour of Daniel Jones*. Longmans Green, London, 26–37.
- Dilley, Laura – Shattuck-Hufnagel, Stefanie – Ostendorf, Mari 1996. Glottalization of word-initial vowels as a function of prosodic structure. *Journal of Phonetics* 24. 423–444.
- Fónagy Iván – Magdics Klára 1967. *A magyar beszéd dallama*. Akadémiai Kiadó, Budapest.
- Gobl, Christer – Ní Chasaide, Ailbhe 2003. The role of voice quality in communicating emotion, mood and attitude. *Speech Communication* 40. 189–212.
- Gordon, Matthew – Ladefoged, Peter 2001. Phonation types: a cross-linguistic overview. *Journal of Phonetics* 29. 383–406.
- Gósy Mária 2004. *Fonetika, a beszéd tudománya*. Osiris Kiadó, Budapest.
- Gósy Mária 2005. A /h/ zöngésedése két magánhangzó között. *Beszédkutatás* 2005. 5–20.
- Henton, Caroline – Bladon, Anthony 1987. Creak as a sociophonetic marker. In Hyman, Larry M. – Li, Charles N. (eds.): *Language, speech and mind*. Routledge, London, 3–29.

- Hirano, Minoru 1981. *Clinical examination of voice*. Springer, Vienna.
- Hollien, Harry – Wendahl, Ronald W. 1968. Perceptual study of vocal fry. *Journal of the Acoustical Society of America* 43. 506–509.
- Ishi, Carlos T. – Sakakibara, Ken-ichi – Ishiguro, Hiroshi – Hagita, Norihiro 2008. A method for automatic detection of vocal fry. *IEEE Transactions on Audio, Speech, and Language Processing* 16. 47–56.
- Laver, John 1980. *The phonetic description of voice quality*. Cambridge University Press, Cambridge.
- Laver, John 1994. *Principles of phonetics*. Cambridge University Press, Cambridge.
- Linville, Sue E. 1995. Changes in glottal configuration in women after loud talking. *Journal of Voice* 9. 57–65.
- Markó Alexandra 2005a. *A spontán beszéd néhány szuprasegmentális jellegzetessége. Monologikus és dialogikus szövegek összevetése, valamint a hőmmögés vizsgálata*. PhD-disszertáció, ELTE, Budapest.
- Markó Alexandra 2005b. „Szavak nélkül”. Nonverbális vokális közlések fonetikai elemzése. *Magyar Nyelvőr* 129. 88–104.
- Markó Alexandra 2007. Kérdő funkciójú hanglejtésformák a spontán beszédben. *Beszédkutatás* 2007. 59–74.
- Markó Alexandra – Bóna Judit 2006. A spontán beszéd lejegyzésének néhány módszertani kérdése. *Beszédkutatás* 2006. 124–133.
- Olaszy Gábor 2006. *Hangidőtartamok és időszerkezeti elemek a magyar beszédben*. Nyelvtudományi Értekezések 155. Akadémiai Kiadó, Budapest.
- Olaszy Gábor 2007. Fonetikai algoritmus a hanghatárok gépi meghatározásának javítására nagyméretű beszédatadabázisokban. In Alexin Zoltán – Csendes Dóra (szerk.): *V. Magyar Számítógépes Nyelvészeti Konferencia*. Szegedi Tudományegyetem, Szeged, 12–23.
- Olaszy Gábor – Bartalis Máttyás 2008. Jelfeldolgozási és fonetikai algoritmusok kombinációja a gépi hanghatárjelölés javítására. *Beszédkutatás* 2008. 208–220.
- Pierrehumbert, Janet – Talkin, David 1991. Lenition of /h/ and glottal stop. In Docherty, Gerald J. – Ladd, D. Robert (eds.): *Papers in Laboratory Phonology II*. Cambridge University Press, Cambridge, 90–117.
- Redi, Laura – Shattuck-Hufnagel, Stefanie 2001. Variation in the realization of glottalization in normal speakers. *Journal of Phonetics* 29. 407–429.
- Rose, Philip 2002. *Forensic speaker identification*. Taylor & Francis, London.
- Slifka, Janet 2006. Some physiological correlates to regular and irregular phonation at the end of an utterance. *Journal of Voice* 20. 171–186.
- Slifka, Janet 2007. Irregular phonation and its preferred role as a cue to silence in phonological systems. In Trouvain, Jürgen – Barry, William J. (eds.): *Proceedings of the 16<sup>th</sup> International Congress of Phonetic Sciences. Saarbrücken 6–10 August 2007*. Pirrot GmbH., Dudweiler, 229–232.
- Stevens, Kenneth N. 1998. *Acoustic phonetics*. MIT Press, Cambridge.
- Tóth László – Kocsor András 2003. A Magyar Telefonbeszéd-adatbázis (MTBA) kézi feldolgozásának tapasztalatai. *Beszédkutatás* 2003. 134–146.

A szerzők ezúton fejezik ki köszönetüket Olaszy Gábornak és a két bírálónak a kézirat korábbi változatával kapcsolatos részletes és hasznos észrevételeikért.



## A HANGSÚLY MEGJELENÉSÉNEK PERCEPCIÓS ÉS PRODUKCIÓS VIZSGÁLATA

Imre Angéla

### Bevezetés

A fonetikai szerkezetek definiálásában (az intonáció mellett, illetőleg gyakran azzal együttesen) meghatározó a hangsúly, valamint a szünet. A beszéd folyamatot értelmessé tevő és tagoló eszközök közül a legtöbbször a hangsúlyt említjük, mert rendszerint ezt valósítjuk meg beszéd közben a leg tudatosabban. A hangsúly a fonetikában egy szótagnak a környezetéhez képest egyfajta kiemelését jelenti. Fontos nyelvi szerepe van, de mivel több elemből tevődik össze, mérésénél is több paramétert kell figyelembe vennünk. A köznapi beszédben a **hangerőtöbblettel** (nyomaték) rendszerint együtt jár a hanglejtésnek bizonyos mértékű változása, általában emelkedése, valamint a hangsúlyozott szónak (a beszéd többi részéhez viszonyítva) bizonyos mértékű lassítása is. Gyakran előfordul, hogy a kiemelendő szó előtt és után még **szünetet** is tartunk (Fónagy 1958; Hunyadi 2003; Elekfi–Wacha é. n. [2003]; Gósy 2004).

A beszédprodukciós lehetőségek közül az egyes nyelvek eltérően „válogatnak” a hangsúly létrehozásának módozataiban, aminek következtében a hallgató hangszíntélménye is eltérő. A magyarban elsősorban a megnövekedett intenzitás kelt hangszíntélményt; de a percepciós döntést egyéb tényezők is befolyásolják, mint az alaphangmagasság (és változása), szemantikai tényezők, szünethelyek, sőt a beszédtempó is. A magyar beszéd folyamatában a hangsúly feltételeesen helyhez kötött, azaz ha valamely szó mondattani vagy egyéb okból hangsúlyossá válik a beszédben, akkor ez a hangsúly általában az első szótagra esik. Ezt nevezzük szóhangsúlynak. A mondatot nyelvi és tartalmi szempontból szakaszokra tagoló hangsúlynak szakaszhangsúly a neve, a legkiemelkedőbb szakaszhangsúly pedig a mondathangsúly (Elekfi–Wacha é. n. [2003]). A hangsúly a monotonia megtörése mellett többek között szintaktikai határjegyként működik; észlelésében szerepet játszanak az adott közlemény szemantikai összefüggései is (Szende 1995). A beszédben a hangsúlynak három fokozatát lehet megkülönböztetni: a főhangsúlyt, a mellékhangsúlyt és a zérus hangsúlyt. Az első kettő abban különbözik egymástól, hogy a főhangsúly mindig intonációs szereppel is bír. A beszédben a hangsúlyosságot a szóval jelölt fogalomnak a beszélés folyamatában betöltött fontossága, illetőleg az adott szintaktikai szerkezet dönti el. Ez azt jelenti, hogy a



hangsúlyviselés nem elsősorban a szónak, hanem annak a szerkezetnek a sajátja, melyben a szó megjelenik (Kálmán–Nádasdy 1994).

A pszicholingvisztikai szakirodalom említi a hangsúlyeltolódás jelenségét, amely összefüggésbe hozható a megakadásjelenségekkel (Fromkin 1973, Cutler 1980), ill. feltételezhetően a fokozott kiemelésre való törekvéssel (Gósy 2002). Felvetődik a kérdés, hogy vajon a hangsúly más szótagra való eltolódása mennyiben befolyásolja a beszédfeldolgozás folyamatait. A témával kapcsolatban kevés kísérlet ismeretes. A spanyol nyelvben a hangsúly olyan szupraszegmentális elem, amely az előhívás folyamata során a versengő aktivált szavak közül kizárhat elemeket. Soto-Faraco és munkatársai (2001) azt találták, hogy amikor a kísérleti személyeknek egy mondatban hiányosan megadott szót kellett kiegészíteniük, a rossz hangsúlyozású szó megnövelte a reakcióidőt. Ezzel szemben egy hasonló angol tesztelés azt az eredményt mutatta, hogy a rossz hangsúly nem okoz komolyabb gátat a lexémák aktiválásában (Cooper, Cutler és Wales 2002). Ez azt jelentheti, hogy az angol nyelvben kevésbé meghatározó a hangsúly a szavak helyes észlelésében és értésében, mint a spanyolban. Ismeretes, hogy a holland nyelv a kötetlen hangsúlyozású nyelvek közé tartozik. A spanyol kísérletet alapul véve van Donselaar, Koster és Cutler (2005) kísérletsorozatukban azt kutatták, hogy amikor a holland beszélők egy szónak, szókapcsolatnak csak egy részét hallják, a szupraszegmentális (ezen belül is elsősorban a hangsúlylyal kapcsolatos) információk befolyásolják-e a lexikális aktiválást és az aszociációt. A kísérlethez olyan szópárokat választottak ki, ahol az első vagy az első két szótag szegmentálisan megegyezett, azonban a szóhangsúly eltért, így a kiválasztott megfigyelendő szótag a szópáron belül az egyik szóban hangsúlyos volt, a másikban pedig nem. A kísérleti személyeknek az első szótag alapján kellett eldönteniük, hogy melyik lehetett a keresett szó. A spanyol nyelvű kísérlethez hasonlóan a kutatók azt találták, hogy a holland nyelvben is van gátló, szűkítő szerepe a szupraszegmentális információnak a lexikális aktiválás folyamatában. Ugyanakkor meg kell jegyeznünk, hogy Jongenburger és van Heuven (1995) korábbi kísérlete ellentmond ennek a megállapításnak, mivel ők azt tapasztalták, hogy az első szótag alapján a hollandok csak akkor tudják a keresett szó meghatározásához a hangsúllyal kapcsolatos információkat sikeresen felhasználni, ha a második szótag magánhangzójából valamennyit hallhatnak.

A magyar nyelvre vonatkozóan Bóna és Imre kutatásukban azt vizsgálták, hogyan hat a helytelen hangsúlyozás a beszédészlelésre, ill. a megértésre (2007). Egy 3 perces spontán narratíva meghallgatásakor azt a feladatot adták a kísérleti személyeknek, hogy jelöljék be a vélt hibás hangsúlyokat, egy másik kísérletben pedig egy 2 perces *a)* helyes hangsúlyozású, *b)* hibás hangsúlyozású, ill. *c)* hangsúlyozás nélküli szöveg meghallgatása után megértést ellenőrző kérdésekre kellett válaszolniuk. A kutatók feltételezték, hogy a hangsúlyeltolódás feltűnő, és zavaróan hat a hallgató beszédészlelésére. Ezzel

szemben percepció kísérletük eredményei szerint a hangsúlyeltolódást a hallgató sokszor észre sem veszi. A szövegértést azonban szignifikánsan rontja a hibás hangsúlyozás, mivel ellentmond a hallgató előfeltevéseinek, és ezáltal lassabbá és bizonytalanabbá válik a beszédfeldolgozás.

A magyar nyelv a kötött hangsúlyozású nyelvek közé tartozik, de mivel a hangsúly nem csak szószinten jelentkezhet, előfordul, hogy két szegmentálisan megegyező virtuális mondat szupraszegmentálisan eltér egymástól, s ezzel különbözik a mondatok jelentése. Jelen kutatásunkban arra keressük a választ, mennyiben határozza meg vagy befolyásolja egy adott mondat jelentés-tartalmát az egyedi szóhangsúlyjelölés, a szófajhangsúly, a mondatrészhangsúly, valamint a mondat szerkezet. Mi határozza meg a hangsúlyt az átlagos beszélő beszédében? Mi dönti el, hogy egy szó főhangsúlyt visel-e vagy sem? A kísérleti csoportok életkor szerinti megfelelő kialakításával az is vizsgálhatónak tűnik, mikor sajátítjuk el anyanyelvünk hangsúlyozási szabályait. Feltételezésünk szerint az életkor függvényében még problémát okozhat az azonos szintaktikai szerkezettel rendelkező, de különböző hangsúllyal megvalósított (tag)mondatok értelmezése, a hangsúlyszabályok alkalmazásakor pedig olyan produkció megvalósítással is fogunk találkozni, amelyek újabb példái a mai magyar nyelvben mostanában megjelent sajátos hangsúlyeltolódásnak.

A magyar nyelvben érvényesülő hangsúlyszabályokat két aspektusból vizsgáltuk, a produkció és a percepció szempontjából. Kísérletet végeztünk arra vonatkozóan, hogy a hallgató milyen mértékben képes azonos szegmentális szerkezettel rendelkező egységek jelentését megkülönböztetni az eltérő hangsúlyozás következtében. Majd egy produkció teszt segítségével azt vizsgáltuk, hogy egy szöveg megfelelő értelmezése mennyiben követhető nyomon az adott szöveg meghangosításában.

### **Anyag, módszer, kísérleti személyek**

A kísérletsorozatunkban 60 magyar személy vett részt, életkoruk szerint három csoportba soroltuk őket: 20-20 fő tíz és tizenhárom éves általános iskolás gyermek, valamint 20 fő húszéves bölcsészhallgató.

### **Percepció kísérlet**

Kísérleti anyagként összeállítottunk egy olyan 10 mondatpárból, azaz 20 mondatból álló listát, ahol a pároknál a mondatok eleje szintaktikailag megegyezett, a mondatok befejezése azonban eltért egymástól. A helyes hangsúlyozást a mondatok egészének értelmezése határozta meg. Például:

*Tegnap Júliával veszekedett Péter;*

*a) ma Annával.*

*b) ma már együtt mentek moziba.*

Ha a fenti példamondatban a *Júliával* szót hangsúlyoztuk, akkor az *a)* megoldás, ha pedig a *veszekedett* szót, abban az esetben a *b)* megoldás volt a helyes. Mind a húsz mondatot Sony MZ-R900 minidisc-felvevővel és a gyár-



tó által hozzá biztosított irányított mikrofonnal, csendesített szobában rögzítettük egy női beszélő bemondásában, majd minden párból kiválasztottuk az egyik mondatot, és abból kivágtuk a közös részt Audacity 1.3.5 hangszerkesztő programmal. Így kaptunk 10 mondatkezdést valamilyen szupraszegmentális megvalósításban; ez képezte a percepciók kísérlet hanganyagát. A kísérleti személyek írásban megkapták a 10 mondat lehetséges befejezéseit (tehát az *a*) és *b*) lehetőséget, a mondat elejét azonban nem). Az így elkészített hanganyagot számítógépről szabad hangtérben lejátszottuk, s a mondatok elejének egyszeri lejátszása után a kísérletben résztvevőknek el kellett dönteniük, hogy melyik lehet a két lehetőség közül a mondat befejezése. A mondatpárok kiválasztásakor törekedtünk arra, hogy azok értelmesek, grammatikailag megfelelőek legyenek.

### Produkciós kísérlet

A percepciók tesztet kiegészítettük egy produkciós kísérlettel is, ahol azt vizsgáltuk, hogy átlagos beszélők felismerik-e egy adott szövegben a megfelelő „hangsúlyhelyeket”, és mely beszédprodukciós lehetőség segítségével hozzák létre a hangsúlyt. Természetesen a legjobb megoldás lett volna az elsajátítás ellenőrzésére a természetes mintavétel, de az valószínűleg nem eredményezett volna kellő mennyiségű releváns adatot, ezért választottuk a felolvasást. Az összefüggő szöveg olyan szavakat, ill. szerkezeteket tartalmazott, amelyek előfordultak az észlelési tesztben is. A kísérleti személyek először némán végigolvashatták a szöveget, értelmezheték az olvasottakat, majd a hangos felolvasásukat csendesített szobában minidisc-felvevővel rögzítettük. A felolvasás előtt mindig megkérdeztük, hogy értik-e a szöveget, ha ismeretlen szóval vagy kifejezéssel találkoztak, annak megadtuk a magyarázatát. A rögzített felolvasások elemzésénél lejegyeztük a felolvasó által létrehozott hangsúlyhelyeket. Az értékelésnél a szövegben az itt félkövéren szedett részek megvalósítását vettük figyelembe, de természetesen a kísérleti személyek a szöveget mindenféle kiemelés nélkül kapták meg. A produkciós teszt szövege:

*Tegnap a barátom Prágában dolgozott, és én is elkísértem őt. Sajnos, drága autóval külföldre utazni, mi mégis ezt választottuk, mert nagyon sietnünk kellett, és már nem kaptunk vonatjegyet. Persze majd elájultam, amikor a végén ki kellett fizetni a benzinszámlát! Indulás előtt még felhívott telefonon, és elmondta, hogy öltözzek fel, mert én az ottani szokásokat nem annyira ismerem. Nyelvi gondjaink nem voltak, hiszen ő sokáig élt Csehországban. Szeretem hallgatni, amikor az emberekkel tárgyal, természetesen beszél csehül, nem affektál. Meg nem mondaná róla senki, hogy nem ott született! A szállással szerencsére nem volt gond: az egyetem területén építettek egy lakókomplexumot a vendégprofesszoroknak; a négy emeletes és egy földszintes épület közül mi választhattuk ki, hol szeretnénk aludni. A barátom beszédtanár. Kurzusa elején a legtöbb hadaró hallgatót alig lehetett követni, de most már lassan beszélnek, így lassan én is megérttem, mit akarnak mondani.*



## Eredmények

### A percepció teszt eredményei

Ha jobban megfigyeljük, a látszólag hasonló felszíni szerkezetek mögött eltérő hangsúlyozási szabályok fedezhetők fel. Tesztmondataink összeállításakor elsősorban Kálmán-Nádasdy (1994) csoportosítását vettük figyelembe, a példák részben az ő tollukból származnak.

A percepció tesztben összességében a 10 évesek 60,5%-ban, a 13 évesek 65,5%-ban, a felnőttek pedig 82,0%-ban döntöttek helyesen a hangsúlyok alapján a mondat jelentéséről (vö. 3. ábra), ami azt mutatja, hogy míg a pubertáskor körül nagyon meghatározó a felismerésben a mondat szerkezete, addig a felnőttek – a számneves szerkezet kivételével – már nagy biztonsággal felismerik a szupraszegmentumokhoz kapcsolható jelentéseket.

Az egyes tesztmondatok felismerése eltérő nehézségűnek bizonyult.

*Rudi meg nem várta soha*

a) az öccsét edzés után.

b) , hogy az apja csinálja meg helyette a leckéjét.

A fentebbi példában a *meg* különböző viselkedését figyelhetjük meg. Az első mondatban igeikötőként szerepel, de mivel a tagadószó elválasztja a *várta* szótól, ezért önálló hangsúlyt kap, míg a második mondatban *pedig* értelemben használjuk, a mondat első részében a tagadószó lesz hangsúlyos, így a *meg* hangsúlytalanná válik. Az értékeléskor végül ezt a példát figyelmen kívül hagytuk, mert még a tesztmondat bemondója is alig tudott árnyalatnyi különbséget tenni a két értelmezés között, így azt láttuk, hogy a kísérleti személyek gyakorlatilag nem tudták megkülönböztetni a *meg* szó egymástól különböző jelentéseit (a válaszok inkább találgatásnak tündek, ezért nem volt érdemes azokat értékelni).

Feltételezhetően a teszt eredménye attól is függ, milyen a kísérleti személyek elvárása, amit egy adott szerkezet gyakorisága is befolyásol. Ezért tudatosan a tesztmondatok közé azokat a szerkezeteket választottuk be nagyobb számban, amelyek ritkábban fordulnak elő a mindennapi kommunikációban. Ezzel a módszerrel várhatóan jobban ellenőrizhető, hogyan használják fel a kísérleti személyek a hangsúlyról szerzett ismereteiket, hiszen így kevésbé tudják megoldani a feladatokat rutinból. A *hogy* különböző szituációkban történő előfordulása erre kiváló példa.

*Megparancsolta, hogy öltözzek fel,*

a) de nekem az ő ízlése túl modern.

b) mert megfázom.

A *hogy* szót a leggyakrabban kötőszóként használjuk, ilyenkor nem is kap hangsúlyt. A másik funkciójában kérdőszóként szerepel, de ha alárendelt mondatban alkalmazzuk, akkor is ritkán áll önmagában, azaz a *hogy* kötőszó nélkül: *Szeretném tudni, hogy aludtál.* Főleg a beszélt nyelvben sokkal jel-

lemzőbb, hogy a kérdőszót megelőzi a *hogy* kötőszó, így egymás mellé kerül alakilag két azonos, de funkcióját tekintve különböző szó: *Szeretném tudni, hogy hogy aludtál.* (Ezt a megoldást egyébként a legtöbb nyelv nem ismeri, hiszen kérdőszó mellett más kötőszó általában nem állhat.) A hangsúlyozás szempontjából a *hogy* kérdőszó akár önállóan, akár a *hogy* kötőszóval áll, mindenképpen hangsúlyt visel. A tesztelésben az *a)* mondatbefejezéshez tartozó közös részt játszottuk le. Az eredmények azt mutatják, hogy az életkor előrehaladtával egyre pontosabban felismerjük ezt a viszonylag ritkább használatot is: míg a 10 éveseknek csak 30%-a, addig a 13 évesek 65%-a, a felnőtteknek pedig a 85%-a helyesen azonosította a megfelelő jelentést.

A mondatok között több olyan példát is találhatunk, ahol a mondatpárban megegyező, a tesztelés szempontjából kiválasztott szó azonos szófajú (pl. mindkettő melléknév), de az adott szó mindkét mondatban mondattanilag más szerepet tölt be.

*Kicsit varrt a nagymama,*

*a) aztán kiment a kertbe ásni.*

*b) ezért az unokája nem tudja hordani az új szoknyát.*

A *kicsit* szó mind a két mondatban tárgyragos melléknév, de az elsőben háttérzőként ('rövid ideig') szerepel, s ebben a jelentésben hangsúlytalan, a másodikban pedig tárgyként ('kis szoknyát'), és így a szó hangsúlyos. A tesztelésben a második példát játszottuk le. Ennek a felismerése nehézséget okozott a kísérletben résztvevőknek, a felnőttek sem azonosították megfelelően a hallottakat. A 10 évesek ennél a mondatnál 70%-ot, a felnőttek 60%-ot, a 13 évesek azonban csak 35%-ot értek el. Meglepő, hogy a legfiatalabbak sokkal jobban teljesítettek a 13 éveseknél, de erre külön magyarázatot nem találtunk.

Egy másik példában a hangsúly mellett a szünet is szerepet kapott:

*Nagyon drága kocsival*

*a) ne parkolj a ház előtt!*

*b) Mongóliába menni.*

A *drága* melléknév egyszer jelzőként szerepel, s a példamondatban megelőzi a *nagyon* szó, ilyenkor hangsúlytalan lesz, ha pedig a mondatban a névszói állítmány szerepét tölti be, akkor hangsúlyos. Az első mondatban a melléknév és az utána következő főnév egy egységet alkot, így nem iktatunk szünetet a két szó közé, míg a másodikban ez nem áll fenn, így a szünettel is jelezzük, hogy a *drága* és a *kocsival* szavakat külön egységként kell kezelünk. A tesztelésben a *b)* mondatbefejezéshez tartozó közös részt játszottuk le. A különböző mondati szerep felismerése a 10 évesek még több mint felének gondot okozott, csak 40%-uk azonosította jól a mondat jelentését. Az életkor előrehaladtával ez már döntően nem jelentett nehézséget: a 13 évesek 95%-ot, a felnőttek 90%-ot értek el.



A melléknevek egy külön csoportját képezik a számnevek. Az irodalmi adatok alapján a megfelelő hangsúlyozásuk nem egyszerű feladat, különösen, ha mértékegység előtt vagy idővel kapcsolatos kifejezésben (a *perc*, *óra* stb. kifejezésekkel együtt) használjuk, ahogy ezt később meg is állapíthattuk.

*Három[]emeletes házat építettek,*

a) *de nincs benne lift.*

b) *de egyikben sem lakik senki.*

Az első mondatban egy házról van szó, amelyben három emelet van (*háromemeletes ház*), a másodikban viszont három különböző házról, amelyekről csak annyit tudunk, hogy nem földszintesek (*három emeletes ház*). Mindhárom kísérleti csoportunk meglepően jó eredményt ért el a beazonosításban: az általános iskolások 75%-os, a felnőttek 90%-os teljesítményt nyújtottak.

Az alábbiakban ismertetjük a legjobban és a legkevésbé felismert hangsúlyhelyzeteket.

*Lassan menjetek fel a színpadra,*

a) *nehogy elessetek!*

b) *nemsokára kezdődik az előadás!*

Az első mondatban a *lassan* szó módhatározóként szerepel, és a *menjete*k igével szorosan képez egy egységet, így hangsúlyos helyzetben van, egyszer pedig mondathatározó vagy módosító határozó, és az egész mondatot módosítja. Szófaji hovatartozását tekintve mindkét esetben határozószó, de máshogy viselkednek a mondatban (a kontextus szerint módosul a jelentésük), így más lesz a hangsúlyozásuk. A tesztelésben a b) mondatbefejezéshez tartozó közös részt játszottuk le. A három kísérleti csoportból mindössze csak 3 fő nem ismerte fel a hangsúly alapján a *lassan* szó jelentését (egy 10 és két 13 éves), a felnőttek mindegyike hibátlanul értelmezte a hallottakat (1. ábra).

A másik jól felismert szerkezetben pedig a jelentéskülönbséget kizárólag a hangsúly helyének különbsége valósítja meg, azaz a két alak percepció szempontból ún. hangsúlyozási minimálpár.

*A[]rab lány*

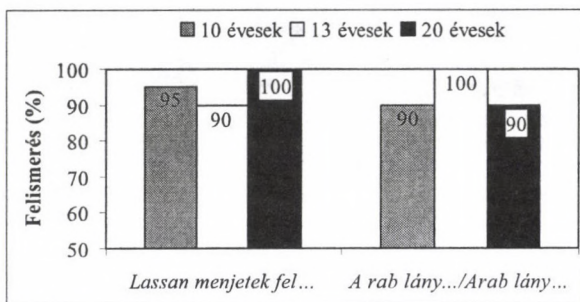
a) *alig várja, hogy kiszabaduljon a börtönből.*

b) *nem sétálhat az utcán kendő nélkül, európai igen.*

Ilyen alakpárokat viszonylag könnyen találhatunk, ahol a hangsor vagy valamelyik mutató névmással, vagy pedig határozott névelővel kezdődik. Ezekben az esetekben a különbség érzékeltethető például kisebb szünettel, de a szakirodalom szerint ez nem kötelező és nem is szokásos (Kálmán–Nádasdy 1994: 400). Ennek ellenére a szerkezetet egy 10 éves és egy felnőtt kivételével mindenki jól felismerte (1. ábra). A tesztelésben az a) mondatbefejezéshez tartozó közös részt játszottuk le. A teljesítmény azért is kiemelkedő, mert



nagyon rövid szegmentum alapján kellett dönteni; a kísérleti személyek csak az [ɔrɒla:n] hangsort észlelték, más kontextus nem segített az értelmezésben.



1. ábra

A legjobban felismert szerkezetek

Az azonosításban a legnagyobb nehézséget a számneves szerkezet jelentette. A strukturális leírások szerint, ha az időpontot adjuk meg, akkor az időegységet jelölő főnévnek hangsúlytalanul kell állnia, tehát *a) négy óra van* (félkövérrel jelöljük a hangsúlyos szótagot). Ha viszont az *óra* szó időtartamot fejez ki, vagy 'időtartamot megadó szerkezet' értelemben használjuk, akkor ez is hangsúlyt visel: *b) négy óra van* (a lakásban). A tesztelésben a *b)* mondatbefejezéshez tartozó közös részt játszottuk le. Az általános iskolások közül mindössze 1 diák hozott jó döntést, de a felnőtteknek is csak a 25%-a ismerte fel helyesen, hogy a hallott hangsorban az *óra* szó milyen funkcióban és jelentésben szerepel (2. ábra). Az eredmények azt mutatják, hogy ilyen helyzetben a megértésben inkább a kontextusra hagyatkozunk, a kísérleti személyek csak a szupraszegmentum alapján nem tudtak sikeresen dönteni. Kérdés azonban, hogy a gyenge felismerés, ill. megkülönböztetés a nem megfelelő elsajátításnak köszönhető-e, vagy inkább annak, hogy az említett fonológiai szabályok alkalmazása a gyakorlatban nagyon kis nyomatékkülönbséget eredményez (a számnév mindenképpen hangsúlyos, így talán kevésbé figyelünk arra, hogy az utána következő főnév hangsúlyos helyzetben van-e), ezért a jelentés azonosításánál inkább a kontextusra támaszkodunk. Feltételezésünk szerint inkább az utóbbi megállapítás a valószínűbb, mivel az eredményekben felnőttkorra sem láttunk jelentős javulást.

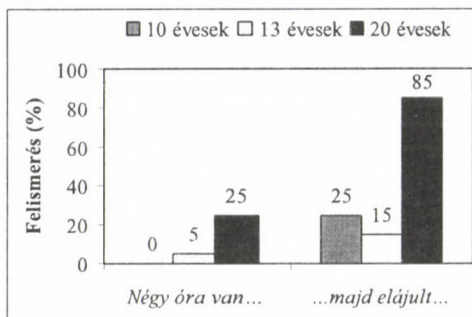
Nehézséget okozott a következő mondatpár szerkezetének felismerése is:

*A színész kilépett a kórház ajtaján, majd elájult,*

*a) mert hirtelen leesett a vérnyomása.*

*b) mennyi rajongója várt rá!*

Meglepő módon a fiatalabbak nem ismerték fel a *majd* szó különböző jelentéseit. A *majd* szó, ha hangsúlyos, akkor azt jelenti, 'kis híján'. Tehát a megvalósítás: *majd elájult*. Ha pedig nem visel hangsúlyt, akkor a jelentése 'azután', így a hangsúlyozása: *majd elájult*. Láthatjuk, hogy a hangsúlyt a szerkezet határozza meg, megváltoztatásával más jelentést hozunk létre. A tesztelésben a b) mondatbefejezéshez tartozó közös részt játszottuk le. Ennek felismerése a felnőtteknek már nem jelentett problémát, ők 85%-ban jól értelmezték a hallott szerkezetet. Ugyanakkor az általános iskolások kevesebb, mint negyede (a 10 évesek 25%-a, a 13 évesek 15%-a) oldotta meg a tesztfeladatot hibátlanul (2. ábra). Ennek az is lehet az oka, hogy feltételezhetően a gyermekek egy része nem ismeri, s ezért nem is tudja megfelelően azonosítani a *majd* szót mint a *majdnem* rövidített alakját.



2. ábra

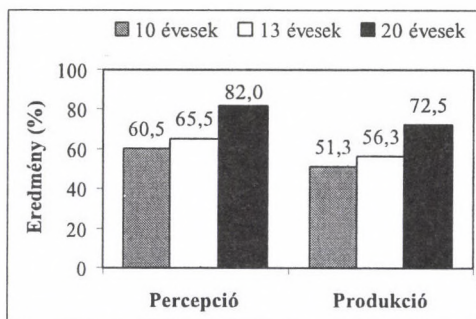
A legkevésbé felismert szerkezetek

### A produkciós teszt eredményei

Ha egy szöveget írásban kapunk meg, láthatjuk a mondatokat, s ha nem értünk meg valamit, ha nem világos számunkra a mondatok összefüggése, többször is újraolvashatjuk azokat. A beszédet hallgató ember számára – az újraolvasás, de akár az újrhallgatás lehetőségének hiányában is – mindezeket az összefüggéseket hanggal kell érzékelhetővé, felismerhetővé tenni. A prozódiai eszközöknek az a feladatuk a beszéd során, hogy lehetővé tegyék, elősegítsék, sőt megkönnyítsék a hallgató számára az elhangzott gondolatok felfogását és megértését. A jelen percepció tesztben szereplő, különböző hangsúllyal megvalósítható szavak és szerkezetek segítségével összefüggő szöveget állítottunk össze. A kísérleti személyek ezt a szöveget felolvasták, azt értékeltük, hogy képesek-e a hangsúlyok megfelelő alkalmazásával értelmezni, érthetővé tenni a rövid történetet. Jól ismert, hogy a nem gyakorlott olvasók másként hangosítanak meg egy szöveget, mint ahogy az

spontán beszédben elhangzana. A felolvasás előtti néma olvasás lehetőségével feltehetően csökkentettük a szöveg félreértésének valószínűségét.

A produkciós teszt átlageredményei a 3. ábrán láthatók. Az összesítés jól mutatja, hogy a felnőttek a hangsúlyhelyeket már nagyobb biztonsággal valószínűsítják meg, mint az általános iskolások, és a három csoport teljesítménye között akkora a különbség, mint a percepciós résznél.



3. ábra

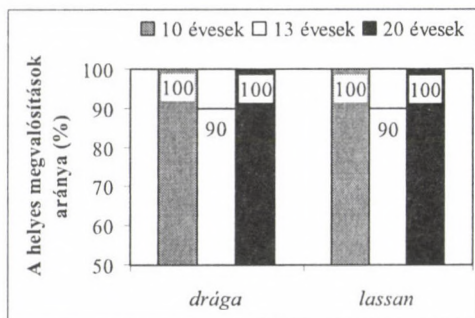
A percepciós és a produkciós teszt átlageredményeinek összevetése

Ha az egyes hangsúlyhelyeket megvizsgáljuk, azt látjuk, hogy szinte hibátlanul oldotta meg a feladatot mind a három csoport, amikor a *drága* szóról el kellett dönteni, hogy határozóként vagy jelzőként szerepel: *Sajnos, drága autóval külföldre utazni*, ill. amikor a *lassan* szó az egész mondatra vonatkozott: *így lassan én is megértem, mit akarnak mondani*. Két-két 13 éves kivételével mindkét mondat esetében minden kísérleti személy 100%-os teljesítményt nyújtott (4. ábra). Ez összefüggést mutat a percepciós teszt eredményével is, hiszen a 10 évesek kivételével minden kísérleti csoportnak azoknak a szerkezeteknek a felismerése sikerült a legjobban, amelyek tartalmazták a *lassan* és a *drága* kifejezést (*a rab lány/arab lány* szerkezet nem fordult elő a produkciós tesztben, ezért nem hasonlítható össze ebben az esetben a percepciós és produkciós teljesítmény).

Kévs kivétellel mindenkinek nehézséget okozott, hogy felismerje a szövegben a *hogy* szó aktuális jelentését. A gyakoribb előfordulással ellentétben a *hogy* szó a szövegben nem kötőszóként szerepelt. Ezt a kísérleti személyek többsége nem ismerte fel, pedig ebben a szöveggösszefüggés is segít: *és elmondta, hogy öltöztek fel, mert én az ottani szokásokat nem annyira ismerem*. Ha felolvasáskor az *öltöztek* szót hangsúlyozzuk, a mondat úgy értelmezhető, hogy a telefonhívás elmaradása esetén meztelenül utaztam volna el. A példa és az említett hangsúlyozás azt mutatja, hogy mivel a *hogy* szó önállóan (tehát más kötőszó nélkül) összehasonlíthatatlanul gyakrabban fordul elő kötő-

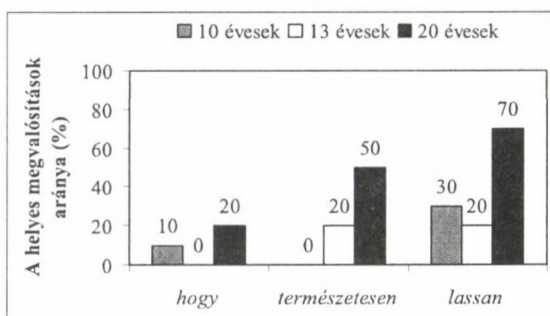


szó funkcióban, mint kérdőszóként, így a kísérleti személyek az erre jellemző hangsúlyozást helyezték előnybe. Minden bizonnyal a jelentésbeli gyakoriság hatással van az észlelési biztonságra. Helyes megoldást mindössze két 10 évestől és négy felnőttől kaptunk, a 13 évesek közül ez senkinek sem sikerült (5. ábra).



4. ábra

A legsikeresebben megvalósított szerkezetek



5. ábra

A legkevésbé sikeresen megvalósított szerkezetek

A kísérleti személyeknek problémát okozott a *természetesen* szó értelmezése is: *természetesen beszél csehül, nem affektál*. Ha a *természetesen beszél csehül* szerkezetben a *beszél* szót hangsúlyozzuk, akkor a mondat jelentése: 'természetes dolog, hogy beszél ezen a nyelven'. A *nem affektál* kiegészítés viszont utalt rá, hogy itt a helyes megoldás a *természetesen* (azaz 'természetes módon') *beszél csehül* lenne. Ezt a jelentést a felnőtteknek is csak a fele, a 13 éveseknél a csoport 20%-a ismerte fel, a legfiatalabbak közül senki (5. ábra).

Végül az utolsó hibás megvalósítás: a *lassan* szó hangsúlyozása, amikor az szoros szerkezetet alkot a *beszél* szóval: *a legtöbb hadaró hallgatót alig lehetett követni, de most már lassan beszélnek*. Itt a *hadaró* kifejezés utalt rá, hogy a *lassan* szó szembeállítását fejez ki, ezért hangsúlyozni kell. Az általános iskolások többsége a hangsúlyhelyet ilyen módon valószínűsítette meg: *de most már lassan beszélnek*, pedig ez azt jelenti, hogy eddig a *hadaró* hallgatók egyáltalán nem beszéltek (5. ábra).

### Következtetések

A percepció és produkció tesztek segítségével az volt a célunk, hogy megvizsgáljuk, mennyiben határozza meg vagy befolyásolja egy adott mondat jelentéstartalmát az egyedi szóhangsúly-, szófajhangsúly-, mondatrészhangsúly-megvalósítás, és mennyiben döntő a szerkezet. További kérdés volt, hogy a kísérleti személyek milyen mértékben képesek azonos szegmentális szerkezettel rendelkező egységek jelentését megkülönböztetni az eltérő hangsúlyozás következtében, ill. egy szöveg megfelelő értelmezése mennyiben követhető nyomon az adott szöveg meghangosításában. A kísérletben részt vevő 3 csoport életkori összetétele lehetőséget nyújtott arra is, hogy megvizsgáljuk: az anyanyelv-elsajátítás során mikorra sajátítottuk el anyanyelvünk szupraszegmentális szerkezetének összetevőit (jelen esetben konkrétan a hangsúlyozási szabályokat).

A hangsúly a monotónia megtörése mellett többek között jelöli a mondat egy-egy szavának, szószerkezetének a beszélés folyamatában elfoglalt fontos (vagy annak érzett) helyzetét, elkülöníti a releváns információt az irrelevánstól, valamint érzékelteti a mondatok belső szerkezeti tagolását. A beszédben az egyes szó (vagy szószerkezet) nem attól függően hangsúlyos vagy hangsúlytalan, hogy mely szófajhoz tartozik vagy milyen mondatrészként áll; a hangsúlyosságot a szóval jelölt fogalomnak a beszélés folyamatában betöltött szerepe, az aktuális vagy kontextuális jelentés, illetőleg az adott szintaktikai szerkezet dönti el.

A percepció tesztben elért teljesítmények azt mutatták, hogy általános iskolás korban még sokkal nagyobb szerepe van a hangsúlyról hozott helyes döntésben a mondat szerkezetének, s gyermekkorban még nagyobbak az egyéni teljesítmények közötti eltérések, ami jól jelzi, hogy az elsajátítás egyéni különbségeket mutat. A felnőttek már nagy biztonsággal felismerik a szupraszegmentumok változtatásával létrehozott különböző jelentéseket. A produkció teszt átlageredményei megközelítőleg 10 százalékponttal lettek gyengébbek a percepció teszt eredményeinél. Megállapítható, hogy minél gyakrabban fordul elő egy szerkezet a mindennapi beszédben, annál nagyobb sikerrel ismerik fel a kísérleti személyek a tesztelés során, s annál pontosabban valószínűsítik meg a helyes hangsúlyhelyeket. A vizsgálatok végeredménye azt sugallja, hogy a hangsúly megvalósítását – tudatosan vagy nem tudatosan – bonyolult elemzés előzi meg. Az eredmények azt a véleményt támasztják

alá, hogy a szupraszegmentumok elsajátítása a (pre)pubertás korban még nem fejeződött be. Beszélőként figyelünk kell arra, hogy a nem megfelelő szupraszegmentális eszközök alkalmazása lassíthatja a feldolgozást. Ez különösen fontos kérdés a nyelvművelés szempontjából is, mert a helytelen hangzás nehezíti a beszéd dekódolását.

### Irodalom

- Bóna Judit – Imre Angéla 2007. A hangsúlyeltolódás hatása a beszédfeldolgozásra. *Beszédkutató* 2007. 75–82.
- Cooper, Nicole – Cutler, Anne – Wales, Roger 2002. Constraints of lexical stress on lexical access in English: Evidence from native and non-native listeners. *Language and Speech* 45. 207–228.
- Cutler, Anne 1980. Errors of stress and intonation. In Fromkin, Victoria A. (ed.): *Errors in linguistic performance. Slips of the tongue, ear, pen, and hand*. Academic Press, New York–London, 67–80.
- Van Donselaar, Wilma – Koster, Mariëtte – Cutler, Anne 2005. Exploring the role of lexical stress in lexical recognition. *Quarterly Journal of Experimental Psychology* 58/2. 251–274.
- Elekfi László – Wacha Imre é. n. [2003]. *Az értelmes beszéd hangzása. Mondatfonetika – kitekintéssel a szövegfonetikára*. Szemimpex Kiadó, Budapest.
- Fónagy Iván 1958. *A hangsúlyról*. Akadémia Kiadó, Budapest.
- Fromkin, Viktoria A. 1973. The non-anomalous nature of anomalous utterances. In Fromkin, Viktoria A. (ed.): *Speech errors as linguistic evidence*. Mouton, The Hague–Paris, 215–242.
- Gósy Mária 2002. A hangsúlyeltolódás jelensége. In Balázs Géza – A. Jászó Anna – Koltói Ádám (szerk.): *Eltető anyanyelvünk*. Tinta Könyvkiadó, Budapest, 193–198.
- Gósy Mária 2004. *Fonetika, a beszéd tudománya*. Osiris Kiadó, Budapest.
- Hunyadi László 2003. A hangsúlyról a fonológiában és a fonetikában. In Hunyadi László (szerk.): *Kísérleti fonetika, laboratóriumi fonológia a gyakorlatban*. Kosuth Egyetemi Kiadó, Debrecen, 65–92.
- Jongenburger, Willy – van Heuven, Vincent J. 1995. The role of lexical stress in the recognition of spoken words: Prelexical or postlexical? In *Proceedings of the 13<sup>th</sup> International Congress of Phonetic Sciences*. Stockholm, 368–371.
- Kálmán László – Nádasdy Ádám 1994. A hangsúly. In Kiefer Ferenc (szerk.): *Strukturális magyar nyelvtan 2. Fonológia*. Akadémiai Kiadó, Budapest, 393–467.
- Soto-Faraco, Salvador – Sebastián-Gallés, Núria – Cutler, Anne 2001. Segmental and suprasegmental mismatch in lexical access. *Journal of Memory and Language* 45. 412–432.
- Szende Tamás 1995. *A beszéd hangszerelése. Idő, hangmagasság, hangerő és határjelzés a közlésben*. Linguistica, Series A, Studia et Dissertationes 13. MTA Nyelv- tudományi Intézet, Budapest.



## A VIRTUÁLIS MONDATOK MŰFAJI MEGHATÁROZOTTSÁGA

Váradi Viola

### Bevezetés

A spontán beszédet alapvetően két csoportra oszthatjuk, a félszponán és a szponán beszédre. A kettő között lényeges a különbség a beszédtervezési folyamatot illetően, mivel az első esetben a beszélő előzetesen felkészül a mondanójára, átgondolja, rendszerezi, hogy mit is akar mondani, addig a második esetben semmilyen beszédtervezés sem előzi meg az aktuális közlést (Gósy 2005). Természetes beszédhelyzetben, az igazi élőszóban a gondolkodás, a szövegalkotás és a szöveg hangoztatása egyidejű, szimultán folyamatban zajlik le. Ilyenkor nincs mód a gondolatok írásos megfogalmazására (Wacha 1999).

Az elhangzó hosszabb közlések megértésének és értelmezésének egyik alapfeltétele a megfelelő tagolás (Gósy 2003). A befogadónak nincs szüksége minden esetben a szegmentálásra, például az írott beszéd esetén karakterhiány, központosítás jelzi az olvasó számára az egyes egységek – szavak, mondatok stb. – határát. A folyamatos beszédben azonban nem ilyen egyértelmű a határjelzés, nem jelzik szünethelyek a szemantikai, szintaktikai egységek kezdetét vagy befejezését (Gósy 2005). A beszélő nincs mindig tudatában annak, hogy mely szegmentális és szupraszegmentális tényezőket alkalmazza tagoló funkcióban, vagyis az artikulációs megvalósítás általában ösztönös, és nagyok az egyéni különbségek is (Kohler 1983). Gyakorlatlan beszélőknél a tagolást gyakran fiziológiai tényezők – például a légzés – szervezik. A hallgató ettől függetlenül arra kényszerül, hogy kisebb egységekre bontsa, tagolja az elhangzottakat, s ezek mentén ismerje fel az összefüggéseket.

A spontán beszéd tagolása nagy eltéréseket mutat az interpretatív szövegekéhez képest, hiszen felolvasáskor nincs szükség a beszédproduktions mechanizmus kezdeti folyamatainak a tervezésére. A szövegalkotás és a szövegmondás nem egyszerre zajlik, mivel a felolvasó olyan szöveget szólaltat meg, melyet már korábban valaki nyelvi formába öntött (Imre 2005). A beszélő sokkal tudatosabban valósíthatja meg a meghangosítás jellemzőit, így a tagolásokat is (Gósy 2005). Megfelelően központosított szövegek felolvasása esetén a mondathatárok jelölése a hallgatók részéről csaknem 90%-os egyöntetűséget mutat (vö. Batliner et al. 1998).

A tagolási egység lehet akár mondatnyi hosszúságú, vagy annál hosszabb közlésrészlet, összefüggő szövegrész is. Ez utóbbit bekezdésnek is nevezzük,

bár e terminus használata nem feltétlenül egyértelmű, mert elsősorban az írott szövegek bizonyos egységeinek tipográfiai elkülönítésére szolgál (Lehiste 1979). Ha eltekintünk a bekezdés fogalmától, akkor szembekerülünk azzal a problémával, hogy mit tekinthetünk a spontán beszéd legkisebb egységének. A mondatnak többféle – logikai, formális, lélektani, funkcionális – megközelítése, meghatározása létezik a nyelvészeti szakirodalomban. Gósy bevezette a virtuális mondat fogalmát, a *virtuális* jelzőt a 'benne rejlő' értelmében használva. Ezek az egységek többé-kevésbé azonos paraméterekkel jellemezhetők, melyeket a hallgatók képesek tagoló funkcióban használni spontán közlések hallgatásakor (Gósy 2003).

A szegmentálás problémája nemcsak a hangok, szavak, hanem a mondatok szintjén is felvetődik a beszédészlelésben- és megértésben. Felmerülhet a kérdés, hogy a beszédnek melyek azok az összetevői, amelyek határjelző funkcióval bírhatnak a hallgató számára, valamint a szupraszegmentális tényezők közül – beszéddallam, hangerő, hangsúly, artikulációs és beszédtempó, ritmus, szünet és egyéni hangszínezet – melyek játszanak elsődleges szerepet a beszédfolyam kisebb egységekre, virtuális mondatokra való szegmentálásában. A szupraszegmentumok közül a szünetek az elsődleges határjelzők, míg az alaphangmagasság-változások csak a másodlagosak közé sorolhatók. A szintaktikai, szemantikai és prozódiai tényezők együttesen segítik a sikeres szegmentálást, részben együttműködnek, részben hierarchikusan szerveződnek a beszédmegértésben (Gósy–Kovács 2007). A humán mechanizmus nem konkrét értékek, hanem viszonyok alapján dönt, amelyek egy bizonyos egységen belül vagy egységek sorozatában realizálódnak. A spontán beszéd bármely sajátossága, amely a virtuális egység befejezését jelezheti, önmagában nem, csak bizonyos összefüggésekben bír határjelző funkcióval, s csak akkor, ha a szemantikai sajátosságok ezt a határt megerősítik (Gósy 2003).

Korpuszokon végzett vizsgálatok eredményei azt mutatták, hogy a prozódiai és egyéb felismerési kulcsok különböznek a szöveg típusától függően. Olaszky (2005) jellemezte a prozódiai szerkezeteket a hírfelolvasásban, mesemondásban, a novella és a reklámok felolvasásában, és azt tapasztalta, hogy az egyes felolvasási stílusok között konkrét számadatokkal is kimutatható különbségek vannak. Úgy tűnik, hogy a beszélőkben kódolva vannak a műfaji sajátosságok, melyek nemcsak felolvasáskor, hanem spontán beszéd során is befolyásolhatják a produkciót. A percepció oldaláról közelítve meg a műfaji jellemzőket, gyakran tapasztaljuk, hogy nem okoz nehézséget, még kevésbé gyakorlott beszélők számára sem, hogy felismerjék a hallott szöveg műfaját. Természetesen segít ebben a szöveg tartalma, felépítése és stílusa is. A tartalomnak és a mondatszerkezetnek megfelelően változhat a beszéd tempója, ritmusa, kiszélesedhet a hangterjedelem, változhat a közbeékeltszünetek hossza és típusa. Wacha Imre felvetette, hogy a különböző műfajú szövegeket nemcsak irodalmi, hanem nyelvészeti szempontból (morfológia, lexikológia, mondattan, fonetika) is érdemes lenne elemezni. A szövegek műfa-



jonkénti csoportosítását nagymértékben befolyásolja a nyelvi forma és a szövegmondó személye. A szövegeknek az írott stílusuk mellett megvan a maguk akusztikus stílusuk is (Imre 2005).

Hipotézisünk szerint a műfaji különbségek nemcsak a felolvasásban, hanem a spontán beszéd folyamatban is megjelenhetnek. A beszéd folyamat egy-egységekre bontását, szegmentálását nagymértékben segítik a szupraszegmentumok, ezért azt feltételeztük, hogy a műfajok különbségeiből fakadó eltérések a prozódiai szerkezetben hatással lehetnek a mondatathár-észlelésre.

A különböző műfajú szövegek között nemcsak a szupraszegmentális szerkezetben, hanem a szupraszegmentumok határjelző funkciójában is feltételeztünk eltéréseket. Hipotézisünk szerint a virtuális mondatok észlelését befolyásolja az elhangzott szöveg műfaja. Legtagoltabbnak a mesét feltételeztük, mivel a gyerekekkel való kommunikáció során a beszélő általában törekszik a hallgató percepciói működéseinek megkönnyítésére, illetve a beszélő számára egész mondat szerkezetek is adottak lehetnek, melyeket nem kell spontán módon megalkotnia. Feltételeztük, hogy a szövegek között előforduló lehetséges különbségek – a beszédszakaszok<sup>1</sup> hossza, a beszédszakaszok és szintaktikai szerkezetek egybeesése; a szünetek időtartama, gyakorisága, típusa; hangterjedelem, dallammenetek – befolyásolhatják a mondatathár-észlelést. Úgy gondoltuk, hogy az adatközlők abban a műfajban érzik a legkönnyebbnek a mondatvégek jelölését, melyben a legegyszerűbbek voltak a jelöléseik.

### Kísérleti személyek, anyag és módszer

Adatközlőnk 27 éves, egyetemet végzett nő, nincs sem beszédhibája, sem hallásproblémája, a felvétel készítésének tudatában volt. Három, különböző műfajú rövidebb spontán beszédet rögzítettünk vele monologikus helyzetben. Először megkértük, hogy ossza meg velünk nyári élményeit (narratíva), majd el kellett mesélnie a *Piroska és a farkas* című közismert mesét (mese), végül be kellett mutatnia lakhelyét (leírás). A felvétel csendesített szobában történt minidisc-hangfelvevővel. A narratíva időtartama 185 s, a meséé 217 s, a leírásé 108 s.

A hanganyagok elsődleges feldolgozásához a Praat 4.3 szoftvert használtuk, ezzel végeztük el az annotálást, az adatolást és a méréseket. Szegmentáltuk a hanganyagokat, a szöveges szakaszokat fonetikusán lejegyeztük (vö. Markó 2005). A szövegeket helyesírásban is lejegyeztük középpontozás és szünet helyek, szóköz nélkül, és ezt a tesztlapot kapták meg a percepciói tesztben résztvevők.

Az annotáció alapján kimértük a beszédszakaszok, néma és kitöltött szünetek időtartamát, átlagot számítottunk belőlük, és megvizsgáltuk az egyes idő-

1. A beszédszakasz fogalmán két szünet közötti szöveges szakaszt értünk (vö. Markó 2005).



intervallumok gyakorisági eloszlását folyamatos akusztikai visszacsatolás és ellenőrzés mellett. Vizsgálatunkban nem határoltuk el a különböző funkcióban (levegővétel, gondolkodási szünet, hatásszünet) megjelenő szüneteket. A jelen vizsgálati anyagban a szünethossz alsó határának olyan értéket tekintettük, melyet az átlagos hallgatók mintegy 60%-a képes észlelni. Ez az érték a szakirodalmi adatok alapján 80 ms (vö. Gósy 2000).

Mindhárom szövegben kiszámoltuk a beszédszakaszok artikulációs tempóját, illetőleg a műfajonkénti átlagos artikulációs és beszédtempót. A beszédallam leírásában figyelembe vettük az alaphang terjedelmét, illetőleg a beszédszakaszok határán jellemző hangmenetek típusát az egyes szövegekben.

A percepció kísérletben 36 fő vett részt, 18–54 év közöttiek, átlagéletkoruk 30 év, 19 nő és 17 férfi. Adatközlőink azt a feladatot kapták, hogy szövegenként kétszeri meghallgatás során függőleges vonallal jelöljék meg a teszt-lapon azokat a helyeket, ahol mondatvéget észleltek. A kísérleti személyeknek nem okozott nagyobb nehézséget a feladat megoldása, többségük már az első meghallgatás során megjelölte a vélt mondathatárokat. Végül meg kellett ítélniük, hogy mely szövegben volt számukra a legkönnyebb, illetve a legnehezebb a mondathatár jelölése, egytől (legkönnyebb) – háromig (legnehezebb) kellett osztályozniuk.

A kísérleti személyek által jelölt mondathatárokat összegeztük, majd összevetettük, hogy műfajonként hány helyen észlelt legalább egy adatközlő virtuális mondatot, illetve kiszámoltuk, hogy átlagosan hányszor jelöltek mondathatárt műfajonként. Elemeztük továbbá a beszédszakaszok végén és a szerkezethatárokon előforduló jelöléseket és az akusztikai sajátosságok összefüggéseit. A szünetek időtartamának, illetve az alaphangmagasság változásának és a mondatjelöléseknek az összefüggéseit is vizsgáltuk. A kapott adatokon az SPSS 15.0 szoftverrel végeztük el a statisztikai próbákat.

## Eredmények

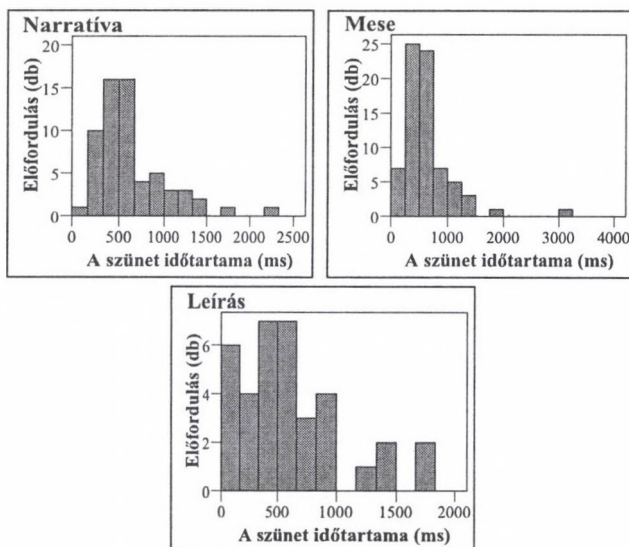
### A spontán szövegek fonetikai elemzése

A nyári élmények felelevenítése, illetve a lakhely leírása természetes szituáció, mindennapi beszédtevékenységünk során is gyakran előforduló témák. Feltételezhető, hogy e műfajok esetén közelít a beszédszakaszok hossza a beszélő természetes beszédjellemzőihez. Az átlagadatok mindhárom műfaj esetén nagyon hasonlóak egymáshoz, a legrövidebbek a szakaszok a narratívában (2228 ms), majd a mesében (2304 ms), a leghosszabbak a leírásban (2348 ms). A statisztikai elemzések (egytényezős varianciaanalízis) szerint a különbség nem szignifikáns [ $F(2, 171) = 0,065$ ;  $p = 0,937$ ], tehát a beszédszakasz hosszát nem befolyásolja a szöveg műfaja.

A beszédszakaszok nem minden esetben estek szerkezethatárra, ezért megvizsgáltuk a szakaszvégek és a szerkezethatárok összefüggéseit is. A narratívában a szerkezethatárok kisebb részben (a beszédszakaszok 48%-a) esnek

szakaszvégre; a leírásban (beszédszakaszok 53%-a) és a mesében (beszédszakaszok 65%-a) a beszédszakaszok több mint fele esik szerkezethatárra.

A néma és a kitöltött szüneteket, illetve ezek kombinációját együtt vizsgáljuk a szünetadatok tárgyalásakor. A három műfaj sem a szünetek gyakorisága (1. ábra), sem az átlagos időtartamuk tekintetében nem tér el egymástól jelentősen, a statisztikai elemzések sem mutattak szignifikáns különbséget [egytényezős varianciaanalízis:  $F(2, 167) = 0,388$ ;  $p = 0,679$ ]. A leghosszabb szünetidőt, mely két néma és egy kitöltött szünet kombinációja, a mesében mértük (3190 ms), a legrövidebbet (116 ms) a narratívában.

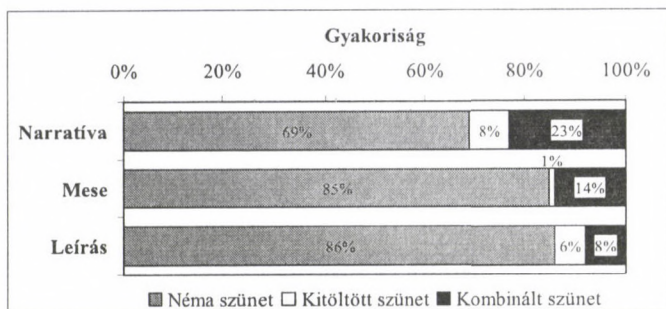


1. ábra

A szövegek szünetidőtartamainak gyakorisági eloszlása a narratívában, a mesében és a leírásban

A kitöltött szünetek előfordulása műfajonként változik. Leggyakoribb realizációi [ə], [m], [əm] vagy [ym]. Önállóan és néma szünettel kombinálva – néma szünet + hezitáció, hezitáció + néma szünet, néma szünet + hezitáció + néma szünet – is megjelenhetnek. A szünetek jelentős része mindhárom műfaj esetén néma, legkisebb része pedig kitöltött. A mesében csak egyetlen kitöltött szünet jelent meg önállóan (460 ms). Sokkal inkább jellemző mindhárom műfajban a kombinált szünetek megjelenése (2. ábra). Feltételezésünk szerint ez a beszélőre jellemző egyéni sajátosság lehet, időnyerés céljából inkább kombinálja a néma és a kitöltött szüneteket ahelyett, hogy hosszabb ideig hezitáljon.





2. ábra

A néma, a kitöltött és a kombinált szünetek aránya a spontán szövegekben

A különböző szövegekben mért beszédtempóértékek között nincs jelentős különbség (narratíva: 11,01 hang/s; mese: 10,03 hang/s; leírás: 11,35 hang/s), ám az artikulációs tempók értékei közül a meséé a legalacsonyabb (narratíva: 14,13 hang/s; mese: 12,80 hang/s; leírás: 14,23 hang/s). A beszélő artikulációs tempóján sokkal kevésbé tud akaratlagosan változtatni (vö. Bóna 2005), vagyis a műfaji eltérésekből fakadó különbségek inkább – a szünetezés eltérései miatt – a beszédtempóban jelentkezhetnek. Felmerül a kérdés, hogy miért az artikulációs tempóknál mértünk jelentősebb különbséget a műfajok tekintetében. Mesét mondani jellemzően gyerekeknek szoktak, és kifejezetten kértük rá a beszélőt, hogy képzelje magát olyan szituációba, hogy gyermekeknek meséli el a Piroska és a farkast. A gyermekekkel való kommunikáció során általában megpróbálunk lassabban, érthetőbben, jobban artikulálva beszélni. Előfordulhat, hogy beszélőnk artikulációs tempójának csökkenése a mesélés során ennek köszönhető. A szövegek artikulációs tempói között mért különbségek szignifikánsak [egytényezős varianciaanalízis:  $F(2, 171) = 5,095$ ;  $p = 0,007$ ]. A szövegek a szünetek arányát illetően nem különböznek egymástól jelentősen, a mesében és a narratívában a teljes beszédidő 22%-a, míg a leírásban 20%-a szünet. A beszélőre jellemző szünetezési szokásokat nem befolyásolja jelentős mértékben, hogy nyári élményeit vagy a *Piroska és a farkas* című mesét meséli-e el, illetve a lakását írja-e le.

A narratíva és a leírás között nem feltételeztünk jelentős eltéréseket az alaphangfrekvencia-adatokban. A mese műfaji sajátosságaihoz azonban sokkal inkább hozzátartozhat, hogy a beszélő „dallamosabban” beszél, ezért feltételeztük, hogy a beszélő hangterjedelme szélesebb lesz. Vizsgálatunk köréből ki kellett rekeszteni azokat a hangmeneteket, amelyekben nyikorgóvá, érdesé válik a zöngé, vagy amelyeket nevetés közben mond adatközlőnk. Ez szélsőségesen alacsonyan, illetve magasan futó alaphangmenetben jelentkezik. Percepció szempontból sem a hang mélyülése, illetve magasodása, hanem nyikorgóvá válása, illetve nevetés érzékelhető (vö. Markó 2005).

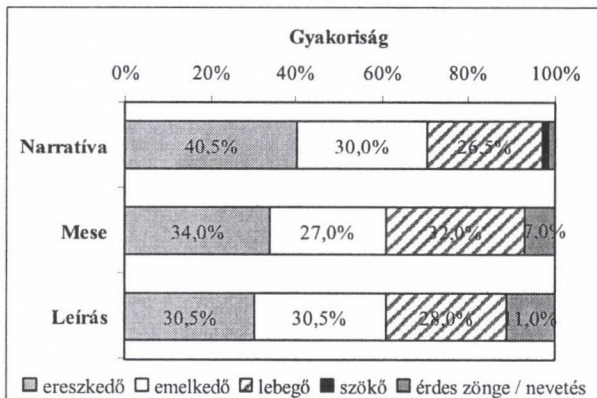


Az 1. táblázatban láthatjuk, hogy az alaphangmagasság minimum értéke hasonló mindhárom műfajban, de a maximum értékekben nagyobb különbségeket figyelhetünk meg. A beszélő hangterjedelme a mesében és a narratívában a legszélesebb, a leírásban a legszűkebb.

1. táblázat: A szövegek alapfrekvencia-adatai

|                  | F <sub>0</sub> -minimum<br>(Hz) | F <sub>0</sub> -maximum<br>(Hz) | F <sub>0</sub> -átlag<br>(Hz) | Hangterjedelem<br>(F <sub>0</sub> -max/F <sub>0</sub> -min) |
|------------------|---------------------------------|---------------------------------|-------------------------------|-------------------------------------------------------------|
| <b>Narratíva</b> | 148,37                          | 497,44                          | 203,32                        | 3,35                                                        |
| <b>Mese</b>      | 142,32                          | 483,22                          | 222,91                        | 3,40                                                        |
| <b>Leírás</b>    | 145,67                          | 443,54                          | 200,84                        | 3,04                                                        |

Megvizsgáltuk, hogy a szakaszvégi dallammenetek milyen típusúak (ereszkedő, emelkedő, lebegő stb.) a következő kérdésekre keresve a választ: 1. milyen arányban fordulnak elő a különböző típusú hangmenetek; 2. ezek az arányok összefüggenek-e a szöveg műfajával. A hangmenetek típusának megállapításakor azt a gyakorlatot követtük, hogy 10%-os frekvenciaváltozásig lebegőnek, 50%-os frekvenciaváltozásig ereszkedőnek (ha negatív irányú), illetve emelkedőnek (ha pozitív irányú) minősítettük a dallammenetet (vö. Markó 2005).



3. ábra

A szakaszvégi hangmenetek típusa a szövegekben

A 3. ábrán láthatjuk, hogy a narratívában (40,5%) és a mesében (34,0%) a szakaszvégi hangmenetek legjelentősebb hányada ereszkedő, a leírásban az ereszkedő és az emelkedő hangmenetek egyenlő arányban fordultak elő szakaszvégeken (30,5%). A narratívában magasabb az emelkedő hangmenetek

(30,0%) aránya, mint a lebegőké (26,5%) a beszédszakaszok végén. A mesében a sorrend ennek a fordítottja: lebegő (32,0%), emelkedő (27,0%).

Érdes zöngével, illetve nevetéssel szakaszvégeken a legnagyobb számban a leírásban (11,0%) találkoztunk, majd a mesében (7,0%) és a narratívában (1,5%). Szökő dallammenet csak a narratívában fordult elő, ott is csak egyetlen esetben. Az eredmények azt mutatják, hogy a szakaszzáró hangmenetek típusát a műfaji sajátosságok befolyásolják.

### A virtuális mondatok észlelése

Az összes mondatvégjelölés, ahol legalább egy adatközlő észlelt mondat-határt, a narratívában 91, a mesében 96 és a leírásban 46. A szövegek eltérő időtartama miatt a továbbiakban nem a darabszámokkal dolgoztunk, hanem az adott műfajban jelölt legtöbb mondat-számot vettük száz százaléknak, és ehhez viszonyítottuk a többi jelölést. Ha a százalékban számolt átlagos mondat-határ-jelöléseket hasonlítjuk össze, akkor láthatjuk, hogy a leírásban jelölték a legtöbb és a narratívában a legkevesebb virtuális mondatot (2. táblázat). A narratíva és a mese jelölt mondat-számai közelebb állnak egymáshoz, a leírásé jelentősebben eltér a másik két műfajtól. Ez fakadhat műfaji különbségekből, hiszen a narratíva és a mese olyan műfajok, melyekben van történet-mesélés, a leírásban azonban nincs.

2. táblázat: A virtuális mondatvégek jelölése a szövegekben

|           | Átlagos mondatvég-jelölés (db) | Szóródás (db) | Átlagos mondatvég-jelölés (%) | Szóródás (%) |
|-----------|--------------------------------|---------------|-------------------------------|--------------|
| Narratíva | 19,72                          | 6–44          | 44,78                         | 14–100       |
| Mese      | 25,89                          | 12–53         | 48,89                         | 23–100       |
| Leírás    | 14,63                          | 6–25          | 58,56                         | 24–100       |

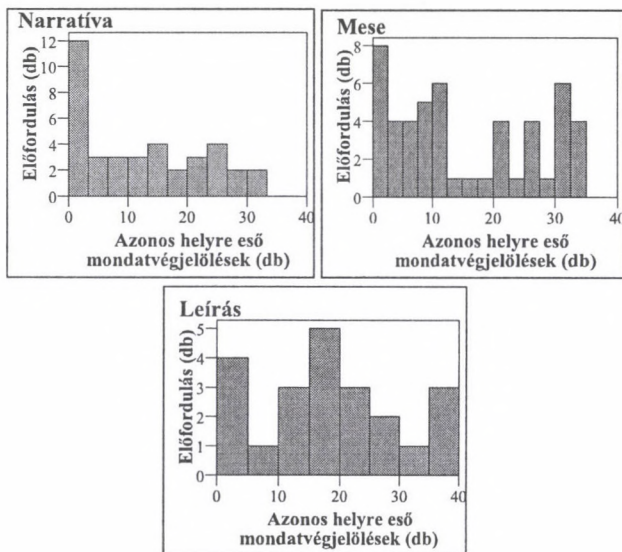
### A virtuálismondat-észlelés és a beszédszakaszok, szerkezethatárok összefüggései

Elemeztük azokat a mondatvégjelöléseket, amelyek beszédszakasz végén, vagyis szünetet megelőzően fordultak elő. A narratívában a beszédszakaszok 55%-ánál, a mesében 59%-ánál, a leírásban 61%-ánál jelölt legalább egy adatközlő mondatvéget. Megvizsgáltuk, hogyan oszlanak el a jelölések a különböző műfajokban. Láthatjuk, hogy a narratíva hisztogramja tér el leginkább a másik kettőtől, ebben a műfajban a beszédszakaszok végén is gyakoriak az egy adatközlő részéről történő jelölések (4. ábra). A másik két műfaj esetén egyenletesebb a jelölések eloszlása.

A mesében a legmagasabb azon jelölések száma, melyek szerkezethatáron fordultak elő. Összehasonlítottuk a beszédszakasz végén és a szerkezethatáron történő jelöléseket, és azt tapasztaltuk, hogy a szerkezethatár sokkal inkább kelti a befejezettség érzetét, mint a szakaszvég.

Elemeztük a szövegekben azokat a jelöléseket, amelyek beszédszakasz- és szerkezethatáron fordultak elő egyszerre, 100%-nak tekintettük az adott szó-

vegben az összes olyan helyet, ahol a beszédszakasz vége és a szerkezethatár egybeesik (narratíva: 31 db, mese: 48 db, leírás: 19 db). A mesében volt a legtöbb olyan jelölés (50%), amely beszédszakasz- és szerkezethatárra esett, majd a leírásban (41,30%), a legkevesebb pedig a narratívában volt (34,07%). A narratíva esetén fordultak elő a legnagyobb számban azok a jelölések, amelyek nem szakasz- vagy szerkezethatáron vannak. Valószínűleg azért, mert ebben a műfajban volt legkevésbé jellemző, hogy a szakasz- és szerkezethatárok egybeesnek.



4. ábra

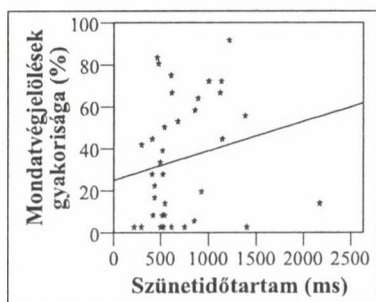
A mondatvégjelölések eloszlása a beszédszakaszok végén

#### A virtuálismondat-észlelés és a szünetek összefüggései

A beszédszakaszok és a mondatvégek összefüggéseinek az elemzése során – mivel a beszédszakasz két szünet közötti egység – a szünetek és a mondatvégjelölések összefüggéseit is vizsgáltuk. A narratívában nem volt egyetlen olyan szünet sem, amely előtt az összes adatközlő mondatthatárt jelölt. A leghosszabb szünetnél – több mint két másodperc időtartamú kombinált szünet – csak öt adatközlő (13,89%) jelölt mondatthatárt. Egy 1218 ms időtartamú kombinált szünet előtt 33 adatközlő (91,67%), egy 464 ms időtartamú néma szünet előtt az adatközlők 83,33%-a (30 fő) jelölt mondatvéget. Mindkét szünet szerkezethatáron fordult elő.



Elemeztük, hogy a szünet időtartama befolyásolta-e a hallgatók virtuális-mondatvég-észlelését. Feltételeztük, hogy minél hosszabb a szünet, annál nagyobb mértékben kelti a befejezettség, lezárttság érzetét, vagyis annál valószínűbb, hogy mondathatárt jelölnek a kísérleti személyek. A Pearson-féle korreláció ( $r = 0,177$ ;  $p = 0,300$ ) azonban nem mutatott ki összefüggést a narratíva műfajában a szünetek hossza és a jelölések száma között (5. ábra).



5. ábra

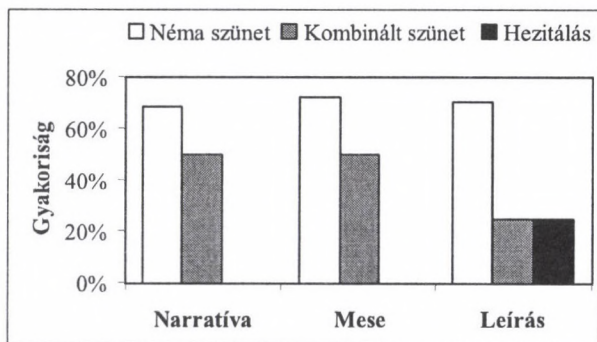
A szünetidőtartamok és a mondatvégjelölések összefüggései a narratívában

A mesében is elemeztük a szünetek hosszának és a mondatvég-észleléseknek az összefüggéseit. Ebben a műfajban nagyobb számban fordult elő, hogy a szünet előtt az adatközlők jelentős hányada, közel 90%-a mondatvéget észlel, de egy esetben sem fordult elő teljesen egyöntetű (az adatközlők 100%-a) jelölés. Két szünet előtt – 305 ms-os néma szünet, 720 ms-os néma szünet előtt – a kísérleti személyek 97,22%-a (35 fő) mondatvéget jelölt. További két esetben az adatközlők több mint 90%-a (33, 34 fő) – 423 ms-os kombinált szünet (néma szünet + hezitáció + néma szünet), 1234 ms-os néma szünet előtt – jelölt mondathatárt. Három szünet előtt – 3191 ms-os kombinált szünet, 642 ms-os és 521 ms-os néma szünet – az adatközlők közel 90%-a (32 fő) jelölt mondatvéget. Ennél a műfajnál is feltételeztük, hogy a szünet hossza befolyásolhatja a jelölések számát. A Pearson-féle korreláció ( $r = 0,149$ ;  $p = 0,305$ ) azonban itt sem mutatott ki értékelhető összefüggést. Úgy tűnik, hogy nemcsak a narratívában, hanem a mesében sem befolyásoló tényező önmagában a szünet hossza a virtuális mondatok észlelésében.

A leírásban egy 1700 ms-os kombinált és egy 622 ms-os néma szünet előtt az összes adatközlő egyöntetűen mondathatárt jelölt, továbbá egy 1296 ms-os néma szünetnél egy kísérleti személyt kivéve, az adatközlők 97,22%-a észlelt mondatvéget. A statisztikai elemzések nem támasztották alá ezt a feltételezést sem, a Pearson-féle korreláció ( $r = 0,406$ ;  $p = 0,068$ ) szerint a leírásban gyenge, de nem szignifikáns a korreláció a szünetek hossza és a mondathatár-észlelések között.

Megvizsgáltuk, hogy a szünetek előtt milyen arányban voltak egyöntetűek<sup>2</sup> a mondathatár-jelölések, 100%-nak tekintettük az adott műfajban az összes olyan helyet, ahol legalább egy adatközlő mondatvéget észlelt. Ez az arány a leírásban és a mesében a legmagasabb (14,29%), és a legalacsonyabb a narratívában (2,78%). Nem igazolódott az a hipotézisünk, hogy minél hosszabb a szünet időtartama, annál valószínűbb, hogy mondathatárt észlelnek az adatközlők. Sem a narratívában, sem a mesében, sem a leírásban nem találtunk összefüggést a szünet hossza és a mondathatár-jelölések között.

Célszerűnek tartottuk elemezni, hogy a szünet időtartamán túl a típusa – néma szünet, hezitálás vagy kombinált szünet – befolyásolja-e az adatközlők mondatvégjelöléseit a különböző műfajú szövegekben. Mindhárom műfajban 100%-nak tekintettük az adott típusú szünet összes előfordulását, és ehhez viszonyítottuk, hogy összesen hány jelölés történt legalább egy adatközlő részéről a különböző típusú szünetek előtt (6. ábra).



6. ábra

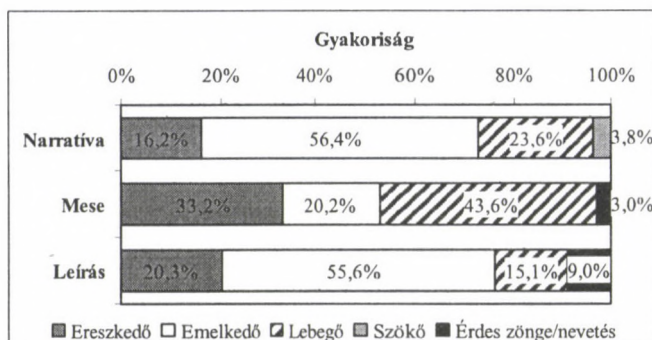
A szünettípus és a mondathatár-észlelés összefüggései

A néma szünet bírt a legerősebb határjelző funkcióval, ezt követte a kombinált, legvégül a kitöltött szünet. A néma szünetek túlnyomó része után legalább egy adatközlő jelölt mondatvéget, a narratívában és a mesében a hezitálás előtt egyszer sem észleltek mondathatárt. A leírásban mind a kombinált, mind a kitöltött szünetek előtt az adatközlők ugyanolyan arányban jelöltek mondatvéget. Az eredmények azt mutatják, hogy műfajtól függetlenül a néma szünet keltette leginkább a befejezettség érzését. A narratívában és a mesében egyszer sem jelöltek mondathatárt hezitálás előtt a kísérleti személyek.

2. Egyöntetű jelöléseken azokat értjük, melyeknél az adatközlők közel 90%-a (32 fő) jelölt mondatvéget.

### A virtuálismondat-észlelés és az alaphangmagasság változásainak összefüggései

Elemeztük, hogy az alaphangmagasság változásai befolyásolják-e a hallgatók mondatthár-észlelését. Holland anyanyelvűekkel végzett kísérletek alapján azt feltételeztük, hogy a beszélő és a hallgató egyaránt használja a befejezettség érzékeltetésére az alaphangváltozást (vö. Swerts–Collier 1991). Az egyes műfajokban megvizsgáltuk a dallammenetek típusát beszédszakaszok végén azokon a helyeken, ahol a kísérleti személyek közül legalább egy mondatvéget jelölt (7. ábra). 100%-nak tekintettük a különböző típusú dallammenetek összes előfordulását.



7. ábra

A dallammenetek típusa és a mondatthár-jelölések összefüggései

Beszédszakaszok végén a narratívában (56,4%) és a leírásban (55,6%) emelkedő dallammenet után jelölték a legtöbb mondatvéget a kísérleti személyek. A mesében (43,6%) lebegő dallammenet keltette a befejezettség érzetét. Érdes zöngé vagy nevetés után a leírásban (9,0%) jelölték a legtöbb mondatthart a három műfaj közül. Szökő dallammenet utáni jelöléssel csak a narratívában találkoztunk (3,8%), mivel csak ebben a műfajban fordult elő ilyen típusú hangmenet.

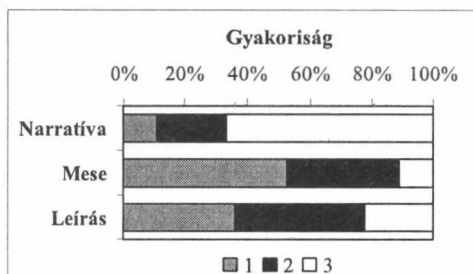
#### A szövegek nehézségének megítélése

Az adatközlőkkel osztályoztattuk a szövegeket a mondatthár-jelölés nehézségének függvényében, 1: amelyben a legkönnyebbnek, 3: amelyben a legnehezebbnek érezte a feladatot (8. ábra). Az adatközlők a narratívát találták a legnehezebbnek (átlag: 2,56), majd a leírást (átlag: 1,86), legkönnyebbnek a mesét (átlag: 1,58).

A nehézségi fok megítélését összevetettük a jelölések között tapasztalt szignifikáns különbségekkel. A mondatthárjelölések a mesében voltak a leggyöntetűbbek, majd a narratívában és a legkevésbé egyöntetűek a leírásban. Feltételezésünk, hogy az adatközlők azt a műfajt fogják a legkönnyebbnek



ítélni, melyben jelöléseik a legegyszerűbbek voltak, esetünkben a mesét, igazolódott.



8. ábra

A szövegek nehézségének megítélése a tagolás szempontjából

### Következtetések

A jelen vizsgálatban egy adatközlő három különböző műfajú spontánbeszéd-anyagát hasonlítottuk össze fonetikailag, és az eredmények azt mutatják, hogy a műfaj objektíve adathalható szupraszegmentális sajátosságai nem feltétlenül tükröződtek a hallgatók tagolásában.

A szövegek átlagos beszédszakaszhozsa között nem találtunk szignifikáns különbséget, de az eredmények azt mutatják, hogy a szakaszvégek nem egyforma arányban esnek szerkezet végére a különböző műfajú szövegekben. A narratívában a beszédszakaszok csaknem fele esik csak szerkezethatárra, a leírásban és a mesében ez az arány magasabb. Beszélőknek nagyobb nehézséget okozott megfelelően tagolni, rendszerezni nyári élményeit. A leírás és a mese esetén kevesebb tervezésre volt szüksége, valószínűleg azért, mert ezekben a műfajokban már többször használt, a mentális lexikonból könnyen előhívható klisékkel dolgozott.

A beszédtempó-értékek között nem találtunk jelentős különbségeket a műfajok tekintetében, az artikulációs tempó értékeiben azonban igen. A legmagasabb a leírásban, majd a narratívában, a legalacsonyabb a mesében volt. A beszélő artikulációs tempóján kevésbé tud akaratlagosan változtatni (Bóna 2005), vagyis a műfaji eltérésekből fakadó különbségek inkább a beszédtempó-értékekben jelentkezhetnek. A szünetek arányát elemezve nem mértünk jelentős különbséget a szövegek között, nem fakadhattak ebből a beszédtempó értékeinek különbségei. Ez a különbség abból fakadhat, hogy a beszélő megpróbált érthetőbben, kifejezőbben artikulálni a mesemondás során, hogy ezzel könnyítse meg a hallgató – a mesék hallgatója általában gyerek – percepciók működését.

Feltételezésünk, hogy a szövegek tagolhatóságát befolyásolja a műfajuk, igazolódott, és ezek a műfaji különbségek a percepciók tesztben részt vevők

mondathatár-észleléseit is befolyásolták. Nagyobb nehézséget okozott számukra, hogy a narratívában jelöljék a mondathatárokat, mint a mesében vagy a leírásban. Az eredmények azt mutatják, hogy a szerkezethatár, vagyis egy szemantikai-grammatikai egység vége sokkal inkább kelti a befejezettség érzetét, mint önmagában a szünet, vagyis a beszédszakasz vége. Az a feltételezésünk, hogy a hosszabb szünet nagyobb határjelző funkcióval bír, nem igazolódott. Mindhárom műfajban a néma szünetek keltették leginkább és a kitöltött szünetek legkevésbé a befejezettség érzetét. Az eredmények azt mutatják, hogy a hezitálás, ellentétben a néma szünettel, a gondolatok nyelvi kifejezését összekötő, nem tagoló jelenség. Egyik szövegben sem az ereszkedő, hanem a narratívában és a leírásban az emelkedő, a mesében a lebegő dallammenet bírt erősebb határjelző funkcióval. Ha nem felolvasunk egy szöveget, akkor kevésbé tudatosan alkalmazzuk a szupraszegmentumokat tagoló funkcióban, a beszéd folyamatosságának érdekében emelkedő vagy lebegő dallammenetek fordulnak elő nagyobb arányban a beszédszakaszok végén.

### Irodalom

- Bóna Judit 2005. A hadaró és a gyors beszéd temporális sajátosságai. *Magyar Nyelvőr* 129. 235–242.
- Batliner, Anton – Kompe, Ralf – Kiessling, Andreas – Mast, Marion – Niemann, Heinrich – Nöth, Elmar 1998. M = Syntax + Prosody: a syntactic-prosodic labelling scheme for large spontaneous speech databases. *Speech Communication* 25. 193–222.
- Gósy Mária 2000. A beszédszünetek kettős funkciója. *Beszédkutatás* 2000. 1–14.
- Gósy Mária 2003. Virtuális mondatok a spontán beszédben. *Beszédkutatás* 2003. 19–43.
- Gósy Mária 2004. *Fonetika, a beszéd tudománya*. Osiris Kiadó, Budapest.
- Gósy Mária 2005. *Pszicholingvisztika*. Osiris Kiadó, Budapest.
- Gósy Mária – Kovács Magdolna 2007. Virtual sentences of spontaneous speech: boundary effects of syntactic-semantic-prosodic properties. In Gardner-Bonneau, Daryle – Blanchard, Harry E. (eds.): *Human factors and voice interactive systems*. Second edition. Springer, New York, 361–379.
- Imre Angéla 2005. Különböző műfajú szövegek szupraszegmentális jellemzői. *Magyar Nyelvőr* 129. 510–520.
- Kohler, Klaus J. 1983. Prosodic boundary signals in German. *Phonetica* 40. 89–134.
- Lehiste, Ilse 1979. Perception of sentence and paragraph boundaries. In Lindblom, Björn – Öhman, Sven (eds.): *Frontiers of speech communication research*. Academic Press, London–New York–San Francisco, 191–201.
- Markó Alexandra 2005. *A spontán beszéd néhány szupraszegmentális jellegzetessége. Monologikus és dialogikus szövegek összevetése, valamint a hűmmögés vizsgálata*. PhD-disszertáció, ELTE, Budapest.
- Olasz Gábor 2005. Prozáiai szerkezetek jellemzése a hírfelolvasásban, a mesemondásban, a novella és a reklámok felolvasásában. *Beszédkutatás* 2005. 21–50.
- Swerts, Marc – Collier, René 1991. On the controlled elicitation of spontaneous speech. *Speech Communication* 12. 35–48.
- Wacha Imre 1999. *Szöveg és hangzása*. Kodolányi János Főiskola, Székesfehérvár.



## A NYÚJTÁS SAJÁTOSSÁGAI A DADOGÓ ÉS AZ ÉP BESZÉDBEN

Bóna Judit

### Bevezetés

A dadogás az egyik legsúlyosabb beszédhiba, amelynek két típusa van: a klónusos és a tónusos dadogás. Az előbbi a szókezdő hangok vagy szótagok ismétlése jellemzi, míg az utóbbinak a legjellemzőbb tünetei között a szó eleji beszédhangok görcsös megnyújtása szerepel. A két típus kombináltan is előfordulhat, ez a klonotónusos dadogás (Sheehan 1974; Subosits 2001). A nyújtás azonban az ép beszédű személyek beszédére is jellemző jelenség, amely egyéntől függően igen gyakori lehet a spontán beszédben.

A magyar nyelvben a nyújtások döntő többsége a spontán narratívákban a határozott névelőkön jelentkezik, és sokkal gyakrabban fordul elő magánhangzókon, mint mássalhangzókon (Gósy 2003); a társalgásban azonban ez az arány lecsökken, és a kötőszavakban, illetve a főnevekben előforduló nyújtások aránya megnő (névelőn: 20,6%; kötőszón: 22,1%; főnevekben: 17,6% – vö. Horváth 2007).

Egy percepciókísérletünk szerint a hallgatók a megakadások közül a legkevésbé a nyújtás jelenségére figyelnek fel (Bóna 2006). Egy másik, kimondottan a nyújtás percepcióját vizsgáló kutatásunk szerint (Bóna 2007) *a)* a nyújtás észlelésében nem pusztán a hangidőtartamok növekedése játszik szerepet, bár ez az elsődleges akusztikai kulcsok egyike a jelenség azonosításához; *b)* a nyújtás észlelése függ a magánhangzó minőségétől: az elől képzett magánhangzók nyújtását könnyebben felismerjük; valószínűleg azért, mert a spontán beszédben kisebb az időtartamaik szóródása; *c)* a névelőn előforduló nyújtásokat kevésbé észleljük, mint a más szófajú szavakban előfordulókat; mivel ezek a leggyakoribbak a spontán beszédben; *d)* a jelenség észlelésében szerepet játszik a beszédszakaszban elfoglalt pozíció, a beszéddallam és az intenzitás is.

Az ép beszédben előforduló nyújtás lehet egyéni beszéd-sajátosság, stílus-jegy; vagy a bizonytalanságból származó megakadások közé sorolható: gyakran a mentális lexikon aktiválásának nehézségére utal, avagy azt jelzi, hogy a beszélő nehezen találja a megfelelő grammatikai formát (Giannini 2003; Peters 2003; Gósy 2005). A dadogó beszédben ugyanakkor a beszédképzés görcösségét jelzi; a kezdődő dadogás első tünete az egyes hangok hosszabb realizációja. Kérdés azonban, hogy milyen időtartam tekinthető a hangok kóros megnyújtásának, hiszen az ép beszéd esetén is mértünk olyan nyújtást,



amelynek az időtartama több mint ötszöröse volt az adott beszédhangra jellemző átlagos időtartamnak (Bóna 2007).

A dadogás korai diagnosztikájában is fontos szerepet játszik a nyújtások vizsgálata. Ugyanakkor a dadogás kezdeti stádiumában nehéz különválasztani a még ép beszédre jellemző, illetve a már beszédhibára utaló nyújtásokat. Angol, német és holland gyermekekkel végzett kísérletek azt mutatták, hogy a dadogó gyermekek sokkal több (Zebrowski 1995; Natke et al. 2006), hosszabb időtartamú és görcsösebb, a dadogásra jellemző megakadást (nyújtást, ismétlést) produkálnak, mint a nem dadogó gyermekek (Boey et al. 2007). Gordon és Luper szerint a korai diagnózishoz a megakadások<sup>1</sup> gyakoriságának, típusának és időtartamának együttes vizsgálata szükséges (Gordon–Luper 1992a és 1992b – idézi Zebrowski 1995).

A dadogásra jellemző megakadások megjelenését ugyanakkor nyelvspecifikus sajátosságok is jellemzik. Amíg az angol dadogó beszédben növeli a megakadás valószínűségét a tartalmas szó kiejtése, ha a szó mássalhangzóval kezdődik, a mondat elején áll, és hosszú, addig a németben a mássalhangzóval kezdődő szavak nem okozzák a megakadások számának nagy mértékű megnövekedését. A mondat szerkesztési sajátosságok miatt (az ige bizonyos kötőszavak miatt gyakran a mondat végén áll) a német anyanyelvű dadogóknak az angolokkal ellentétben nem a mondat eleje, hanem a mondat vége okoz nehézséget (Dworzynski et al. 2003).

Subosits a dadogók hangidőtartamait vizsgálva megállapította, hogy a magánhangzók időtartama némileg megrövidül, a mássalhangzók tartamértéke pedig megnyúlik az ép beszédhez képest (Subosits 1975a). Ugyanő felolvasott szavakban a szókezdő két hang időtartamának együttes vizsgálata alapján arra jutott, hogy a dadogók ejtésében ezen hangok együttes időtartamának szórása mintegy hétszerese a normál beszélőknél tapasztalt szórásnak; illetve a kezdő hangkapcsolat megnyúlása leginkább a mássalhangzó-torlódásnál, legkevésbé pedig a réshangokkal, nazálisokkal és közelítőhangokkal kezdődő szavak esetében tapasztalható (Subosits 1975c).

A jelen kutatásunk célja annak meghatározása, hogy a magyarban mikor tekinthető már kórosnak a nyújtás mint megakadásjelenség. Vizsgálatunkban három felnőtt dadogó spontán beszédét elemeztük és vetettük össze az ép beszédével. A vizsgált személyek alacsony számát az indokolja, hogy a dadogók aránya a felnőttek között igen kicsi. Az iskolások mintegy 1–2%-a dadog; felnőttkorra pedig minden 300-ból egy dadogó marad meg (Gósy 2005).

Hipotéziseink szerint különbséget találunk az ép beszédű és a dadogó személyek között 1. a megnyújtott hangok időtartamában, 2. előfordulási gyako-

1. A *megakadás* terminus a dadogó beszéd kapcsán más értelemben használatos, mint az ép beszédben. A szakirodalmi hagyományoknak megfelelően azonban nem változtattunk a szóhasználaton.

riságában, 3. a szóban elfoglalt helyében, illetve 4. a szó szófájában, amelyben a nyújtás megjelenik. A dadogásra jellemző megakadások ugyanis a korábbi vizsgálatok szerint főként a szavak elején jelennek meg (Sheehan 1974; Subosits 1975c). A kóros nyújtások akusztikai sajátosságainak elemzése segítséget adhat a dadogás korai diagnózisában.

### **Anyag, módszer, kísérleti személyek**

Vizsgálatunkban nem szerint (2 nő és 4 férfi) és életkor szerint (21–32 évesek) megfeleltetett három, logopédus által diagnosztizált, ép halló dadogó és három ép beszédű kontroll személy vett részt. A dadogók között egy férfi súlyos, tónusos, egy nő és egy férfi pedig enyhe, klonotónusos dadogó volt.

A súlyos dadogó elmondása szerint gyermekkorától dadog, folyamatosan logopédiai kezelésekre jár. Megtanult a beszédhibával együtt élni, olyannyira, hogy sokat és szívesen beszél, gyakran szerepel nagyobb hallgatóság előtt is, aktív közösségi ember. A felvétel készítésekor is rendkívül kommunikatív volt, egy-egy kérdésre hosszan válaszolt. A családjában a beszédgyengeség öröklődik, az egyik nagymamája is dadogott.

Az enyhén dadogó férfi beszédhibája is gyermekkorra vezethető vissza. Ő csendesebb, visszahúzódnóbb típus, bár munkájához szükség van néha a nyilvános megszólalásra. Jelenleg logopédushoz jár. A hangfelvétel készítésekor készséges volt, de a kérdésekre igen röviden válaszolt.

Az enyhén dadogó nő beszédhibájára először 12–13 éves korában figyeltek fel. Bár járt logopédushoz, dadogása kissé súlyosbodott. Beszédprodukciója (a dadogásra általában jellemző módon) a kommunikációs helyzetnek és a fáradtságának megfelelően változik, családi körben egyáltalán nem dadog. Mind a magánéletében, mind a munkájában sikeres, az egyetem elvégzése után olyan munkát választott, amely igényli a rendszeres nyilvános megszólalást.

Mindegyik beszélőtől rögzítettünk egy rövid spontán beszédet (4–7 perc időtartamban), amelyet a Praat 5.0 szoftver segítségével annotáltunk két szinten. Az egyik szinten a beszédszakaszokat, a másikon a vizsgált beszédhangokat címkéztük fel. A beszédprodukciókban kiszámoltuk az artikulációs (AT) és a beszédtempót (BT); megmértük a nyújtások időtartamát; illetve kiválasztottunk minden megnyújtott hangnak megfelelő, 10–10 normál ejtésű beszédhangot, amelyeknek szintén megmértük az időtartamát. A hangidőtartamok mérésekor az időtartamot a hang kezdetétől a hang végéig mértük; ha a két szomszédos beszédhang között hangátmenet volt, a beszédhang határát az átmenet közepénél jelöltük meg.

A vizsgált beszédprodukciókban előforduló összes nyújtást elemeztük, magán- és mássalhangzókat egyaránt. Mivel a dadogásban jóval gyakoribbak a nyújtások, mint a normál beszédben, mintegy négyszer annyi dadogásból származó adatot kaptunk, mint a kontroll beszédből származót. Összesen, az az minden beszélőnél együttvéve 173 nyújtás és 380 normál ejtésű hang időtartamát elemeztük (ha a megnyújtott hangok között többször szerepelt

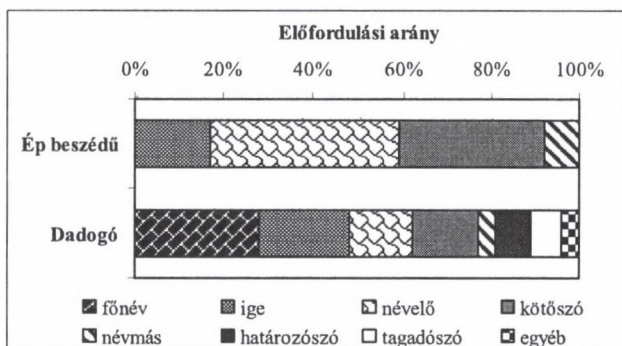


Az előfordulás szempontjából elemeztük, hogy mely hangokon, illetve milyen szófajú szavakban jelenik meg a nyújtás. A kontroll személyek háromszor olyan gyakran produkáltak magánhangzónyújtásokat, mint mássalhangzónyújtásokat (a nyújtások 77%-a fordult elő magánhangzókön, 23% mássalhangzókön), és leggyakrabban az *a* és *e* hangokat nyújtották meg. A dodo-



góknál ezzel szemben gyakrabban fordult elő mássalhangzónyújtás, mint a magánhangzók megnyújtása, bár náluk a magánhangzó- és mássalhangzónyújtások előfordulása között kisebb volt a különbség, mint a kontroll személyeknél (41% magánhangzókon, 59% mássalhangzókon). A legtöbb nyújtást (a szakirodalomnak megfelelően) az explozívák esetében adatoltuk, ezek a hangok bizonyultak a dadogók számára a legnehezebbnek. Ennek oka az explozívák képzésében keresendő, a dadogók a beszédsszervek görcsössége miatt nem tudják „időben” felpattintani a szájüregben vagy az ajkak között képzett zárat.

A kontroll személyek beszédében a leggyakrabban előforduló névelőnyújtások mellett igen gyakori volt a kötőszavak valamely hangjának megnyújtása (például *dee*, *haa*), illetőleg előfordultak névmásokon (például *amii*) és igéken (például *voltamm*) is nyújtások. A dadogó beszédben a legtöbb nyújtás a főneveken (28%) és az igéken (20%) fordult elő (1. ábra – az egyéb kategóriába a módosítószó és a névutó került).



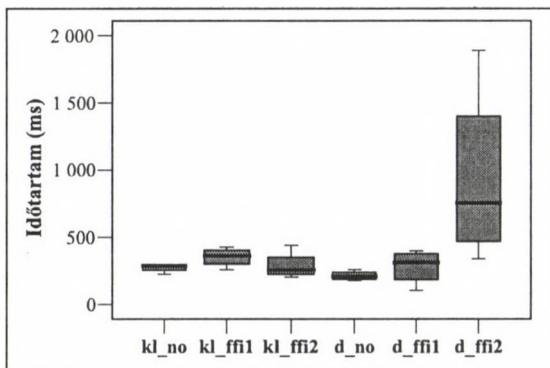
1. ábra

A nyújtások előfordulása a szófaj szerint

A kontroll, az enyhén dadogó és a súlyos dadogó beszédben előforduló nyújtások átlagos időtartama között csak a súlyos dadogó esetében ( $d\_ffi2$ ) találtunk szignifikáns különbséget a többi csoporthoz képest (az egytényezős varianciaanalízis szerint a csoportok között:  $F(2, 170) = 15,007$ ;  $p < 0,001$ ; a Tukey post hoc teszt alapján a  $d\_ffi2$  és a többi csoport között  $p < 0,001$ ). A legtöbb igen rövid (az összes többinél rövidebb) értéket a dadogó nőnél mértük ( $d\_no$ ); míg az egyik kontroll személy (kl\_ffi1) nyújtásainak átlagos tartama nagyobb volt, mint az enyhén dadogó személyeké (2. ábra).

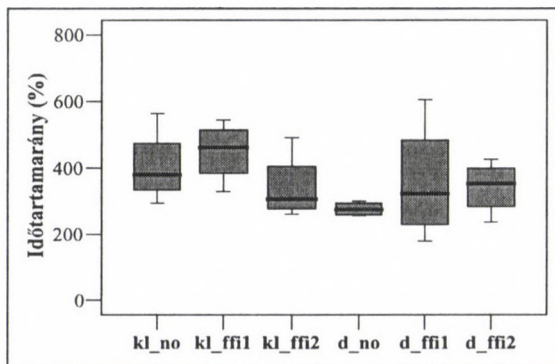
Mivel a nyújtás percepciója függ a beszélő artikulációs tempójától, a hangkörnyezettől, illetve a hangra jellemző átlagos időtartamoktól, ezért minden esetben meghatároztuk az átlagos időtartamokhoz viszonyítva a nyújtások re-

latív tartamát, vagyis azt, hogy százalékban megadva mennyivel hosszabb a nyújtás, mint az átlagos beszédhang. A 3. ábrán megfigyelhetjük, hogy bár a súlyos dadogó férfi (d\_ffl2) sokkal hosszabb nyújtásokat produkált, mint a többi beszélő (a leghosszabb nyújtása mintegy hatszor akkora, mint a többi beszélő által produkált nyújtások leghosszabbika), a lassabb artikulációs tempója miatt a nyújtásainak aránya az átlagos beszédhangjaihoz képest jóval kisebb különbséget mutat, mint az enyhén dadogó férfi (d\_ffl1), illetve két kontroll személy (kl\_no és kl\_ffl1) esetében. Ez azt jelenti, hogy a nyújtások teljes és relatív időtartama nem különbözik a vizsgált dadogó és kontroll beszédprodukciókban.



2. ábra

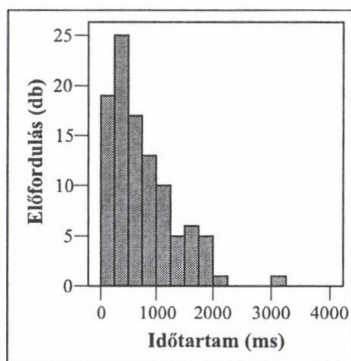
A nyújtások időtartama beszélőnként (szóródás és medián)



3. ábra

A nyújtások és az átlagos hangidőtartamok aránya személyenként (szóródás és medián)

A súlyos dadogó (d\_ff12) beszédprodukciónak részletesebben is elemeztük. Bár a beszédében gyakoriak voltak az egy másodpercnél hosszabb nyújtások (kb. 25%), döntő többségében egy másodpercnél rövidebb hangnyújtásokat produkált (4. ábra).



4. ábra

A nyújtások időtartamának és gyakoriságának összefüggései a súlyos dadogó beszédében

A súlyos dadogó nyújtásainak 60%-át turbulens zörej kísérte. Ez alátámasztja a szakirodalomban leírtakat, a dadogás súlyosságától függően a zörej gyakran megjelenhet; minél súlyosabb a dadogás, annál gyakoribb a megjelenése. A turbulens zörej ugyanis a beszédszervek görcsével áll összefüggésben (Subosits 1975b).

Elemzésünk szerint tehát csak a súlyos dadogó nyújtásainak akusztikai szerkezete különbözik az ép beszédétől, az enyhén dadogók beszédében előforduló nyújtások nem különböznek időtartamukat tekintve az ép beszédbeliektől. Ugyanakkor sokkal gyakrabban jelennek meg a beszédprodukciónban, ezzel a hallgató nyújtásélményét megsokszorozva.

### Következtetések

Hipotézisünk, amely szerint különbséget találunk a nyújtás időtartamában a kontroll beszélők és a dadogó személyek között, részben igazolódott; míg a nyújtások előfordulásának gyakoriságában és helyében egyértelmű különbséget figyelhettünk meg a két vizsgált csoport között.

1. A dadogó és az ép beszédben előforduló nyújtások időtartamukban csak a súlyosabb dadogás esetén különböznek, enyhe dadogás esetén az időtartamban nem találtunk szignifikáns különbséget. A súlyos dadogó nyújtásai gyakran hosszabbak voltak, mint egy másodperc, illetve nagyrészt turbulens zörej kísérte őket.



2. Az ép beszédben a nyújtások főként magánhangzókön fordulnak elő, míg a dadogók gyakrabban produkáltak mássalhangzónyújtásokat.

3. Az enyhe dadogásban és az ép beszédben előforduló nyújtások közötti fő különbségeket a gyakoriságban, a megnyújtott hang minőségében, illetőleg a szófaji kategóriákban találjuk, amelyekben a nyújtás megjelenik.

A nyújtások megjelenési formája tehát meghatározó lehet a dadogás korai felismerésében. Ugyanakkor a dadogás diagnózisában és a terápiájában is szem előtt kell tartani, hogy a nyújtások „jelentése” eltér a dadogó és nem dadogó beszédben. A nem dadogóknál a beszédtervezési bizonytalanságra és a mentális lexikonban való keresésre utalnak, illetve az egyéni beszédstílus jellemzői lehetnek. A dadogóknál ezek mellett a funkciók mellett a nyújtás a motoros kivitelezés problémáját és/vagy a megfelelő agyi parancs kiadásának nehézségét is jelöli. Végül mivel a dadogás kialakulása többnyire neurológiai, illetőleg pszichés okokra vezethető vissza (Schmidtné Balás 2005), a nyújtások nagy száma a pszichés gátlás, a beszéd görcsösségének mértékét is jelezheti.

### Irodalom

- Boey, Ronny A. – Wuyts, Floris L. – Van de Heyning, Paul H. – De Bodt, Marc S. – Heylen, Louis 2007. Characteristics of stuttering-like disfluencies in Dutch-speaking children. *Journal of Fluency Disorders* 32. 310–329.
- Bóna Judit 2006. A megakadásjelenségek akusztikai és percepciósi sajátosságai. *Beszédkutatás* 2006. 101–113.
- Bóna Judit 2007. Magánhangzó-nyújtások akusztikai-fonetikai paraméterei a spontán beszédben. *Beszédkutatás* 2007. 99–107.
- Dworzynski, Katharina – Howell, Peter – Natke, Ulrich 2003. Predicting stuttering from linguistic factors for German speakers in two age groups. *Journal of Fluency Disorders* 28. 95–113.
- Giannini, Antonella 2003. Hesitation phenomena in spontaneous Italian. In Solé, Maria-Josep – Recasens, Daniel – Romero, Joachim (eds.): *Proceedings of the 15th International Congress of Phonetic Sciences. Barcelona 3-9 August 2003*. Universitat Autònoma de Barcelona, Barcelona, 2653–2656.
- Gósy Mária 2003. A spontán beszédben előforduló megakadásjelenségek gyakorisága és összefüggései. *Magyar Nyelvőr* 127. 257–277.
- Gósy Mária 2005. *Pszicholingvisztika*. Osiris Kiadó, Budapest.
- Horváth Viktória 2007. Szófajfüggők-e a spontán beszéd megakadástípusai? Előadás a Beszédkutatás 2007. című konferencián.
- Natke, Ulrich – Sandrieser, Patricia – Pietrowsky, Reinhard – Kalveram, Karl Theodor 2006. Disfluency data of German preschool children who stutter and comparison children. *Journal of Fluency Disorders* 31. 165–176.
- Peters, Benno 2003. Multiple Cues for Phonetic Phrase Boundaries in German Spontaneous Speech. In Solé, Maria-Josep – Recasens, Daniel – Romero, Joachim (eds.): *Proceedings of the 15th International Congress of Phonetic Sciences. Barcelona 3-9 August 2003*. Universitat Autònoma de Barcelona, Barcelona, 1795–1798.
- Sheehan, Joseph G. 1974. Stuttering behavior: A phonetic analysis. *Journal of Communication Disorders* 7/3. 193–212.

- Schmidtné Balás Eszter 2005. *Képes könyv a dadogásról*. Közhasznú Alapítvány a Dadogókért, Budapest.
- Subosits István 1975a. A beszédhangok időtartama dadogók beszédében. *Gyógypedagógia* 1. 4–7.
- Subosits István 1975b. A görcsös hangadás hangszínképéről. In Göllesz Viktor (szerk.): *A Bárczi Gusztáv Gyógypedagógiai Tanárképző Főiskola évkönyve VIII*. BGGyTF, Budapest, 319–324.
- Subosits István 1975c. A szókezdő hangok időtartama dadogók beszédében. In Göllesz Viktor (szerk.): *A Bárczi Gusztáv Gyógypedagógiai Tanárképző Főiskola évkönyve VIII*. BGGyTF, Budapest, 309–318.
- Subosits István 2001. A beszéd rendellenességei. *Egyetemi Fonetikai Füzetek* 30. Budapest.
- Zebrowski, Patricia M. 1995. The topography of beginning stuttering. *Journal of Communication Disorders* 28/2. 75–91.

## REPETITIONS IN INTERRUPTED SPEECH PRODUCTION

Horga, Damir

### Introduction

#### Speech production as a naturally imperfect process

Generally human activities are not perfectly executed even though people intend to plan and perform actions that way. This fact is confirmed daily in newspaper reports about traffic accidents where the so-called "human factor" is often mentioned as the main cause of an accident. Apart from such extreme examples of imperfect human behaviour, there is an enormous number of not-so-severe cases of human actions executed imperfectly especially in our motor actions, such as typing, sports, the playing of musical instruments and various habitual movements. This is the case for speech and language production in their written or oral forms as well. Of course, human perceptual mechanisms are not perfect either, but that topic is outside the scope of this paper.

If two friends arranged to meet on Sunday and one of them, on leaving, said: *All right then, see you on Saturday*, the next moment he would realize that he had made a mistake and he would probably correct himself immediately saying: *I mean, on Sunday, as we arranged*. Although we make mistakes, if they are not fatal, we can correct them ourselves without being warned by somebody else. Speakers do that quite often. A self-repair is defined as a correction without an external warning. Therefore this kind of correction must be based on some kind of a mechanism of simultaneous execution, one that monitors speech production, notices the error and consequently repairs it. Speakers' reactions to their own errors are usually very fast, which means that the time window in which speakers control speech production is relatively narrow.

Signs of imperfections in the speech production mechanism can be manifested in three ways: firstly, as various forms of hesitations or disfluencies including hesitations and repetitions of parts of utterances, although in fact they were realised correctly, secondly, as real speech errors, and finally, as corrections of errors.

Nakatani and Hirschberg (1994) have found that in 10% of utterances in spontaneous speech there is some kind of repair, which implies that speech is very often changed during the process of speech production. But the process of speech error correction is also not perfect because speakers correct only



50% of produced errors (Nooteboom 1980). Errors can remain uncorrected when the mechanism of speech production control fails to notice them. Another reason for errors remaining uncorrected can be that although speakers notice an error, they choose not to correct it. This is so, because situational and language contexts are rich and redundant enough so the listener can correctly interpret the utterance and the intentions of the speaker even when the utterance is in an imperfect form.

The ways speech errors are caused and repaired have been analysed within the frameworks of various speech production models. These analyses help determine which components of the speech production process failed when a particular error occurred.

### **Speech production models**

Mechanisms of speech production have been explained by various theoretical models in which mechanisms are seen as hierarchically organised processes of information undergoing transformation on its way from one level to another by changing adequate codes and the material form characteristic for a particular level. Jaeger (2005: 6–11) in his “Representation and Processing Components Model” regards speech production as a system built on two kinds of components: on the one hand, the speaker must have built-in programmes of speech production and on the other, the possibilities available for effective realization of these programmes. Each of these two levels can be divided into several hierarchically organized sublevels. The first level is called representational, because in it the complete knowledge about the world is saved in long term memory, on the basis of which the speaker will organize the informational/content/conceptual component of his speech. This level also comprises knowledge about language and speech mechanisms which will enable the speaker to give linguistic form to the information and to produce it in speech. The second level is the processing level, which stores the mechanisms by which speakers activate, select and retrieve components from the long term memory and build them into the current utterance, modelling it into language and speech forms. On both levels the speaker can function imperfectly; either he has imperfect programs of speech production or having well-established programs, his realization of the programs in some components or in some moments can fail.

Different kinds of errors can be interpreted from the information or communication point of view. The speaker can produce errors because he does not know the code of a certain language i.e. semantic noise is present in the communication process. That is the case of foreign language speakers who are not skilful enough and do not know the linguistic and pragmatic rules of a language. Kurhila (2001), Rieger (2003) and Corder (1973) refer to these kinds of imperfect functioning of speech production mechanisms as “errors”. Mistakes which are caused by imperfect realization of the communication code are errors produced by native speakers of a language who know the

code of the language very well, but nevertheless make mistakes. Of course, this kind of error also often occurs in foreign language speakers, because mechanisms for code realization are not automated enough. Corder calls this kind of error "lapsus".

Another interpretation of errors is given by Ohlsson (1996). This interpretation agrees with the above mentioned dual view of speech error classification. He states that something can be considered as an error according to an external criterion like language correctness. These errors can be called objective errors. On the other hand speakers themselves can establish their own internal criterion and depending on the fact whether a produced utterance agrees with their expectations, they can consider it erroneous or not. These errors can be called subjective. Often subjective and objective errors coincide.

The complex process of speech production, from speakers' motivation to present linguistically formed information to their ability to articulate it in speech, is divided into several components in different models. Various authors synthesised those approaches in integral models of speech production (Fairbanks 1954, Borden and Harris 1980, Garret 1980, Dell 1986, Levelt 1995, Horga 1997, Postma 2000, Hartsuiker et al. 2001, Jaeger 2005, Gósy 2007). The model presented here (Figure 1) is based mainly on the model of Postma (2000) with some differences in components and their positions.

A major characteristic of that model is the hypothesis that speech production from the speaker's intention to communicate certain information articulated in the form of the speech sound signal is organized in three levels: informational, language and articulatory ones, which are hierarchically organized and controlled by a central monitor and a number of feedback connections. Furthermore, at each level the information is encoded by means of its rules, but the basic condition of the communication process must be considered, namely, that the information should remain identical from the moment of its initiation in the speaker's mind to the moment of its interpretation by the listener. But because speech production is, from the point of temporal organization, a continuous process, we can suppose that the speaker distributes the time he has at his disposal by interchanging the periods of time in which he plans the information, language and articulatory structure of the message with the moments when he performs what was planned. That means that the production of speech is a continuous and dynamic process of the creation of an utterance. In this complex process certain components of speech production mechanisms can fail to encode information correctly, so errors can occur representing the semantic noise produced by the speaker himself. The demands for the equivalence of information on its way from the speaker to the listener can be further complicated, because not only semantic noise but also real physical noise can occur in the communication channel, degrading the sound signal. These are the reasons why the listener must be an active participant in speech communication, interpreting the received message and

even reconstructing the destroyed parts of speech signal and correcting the errors made by the speaker, if he himself has not done that, which is very often the case in speech communication.

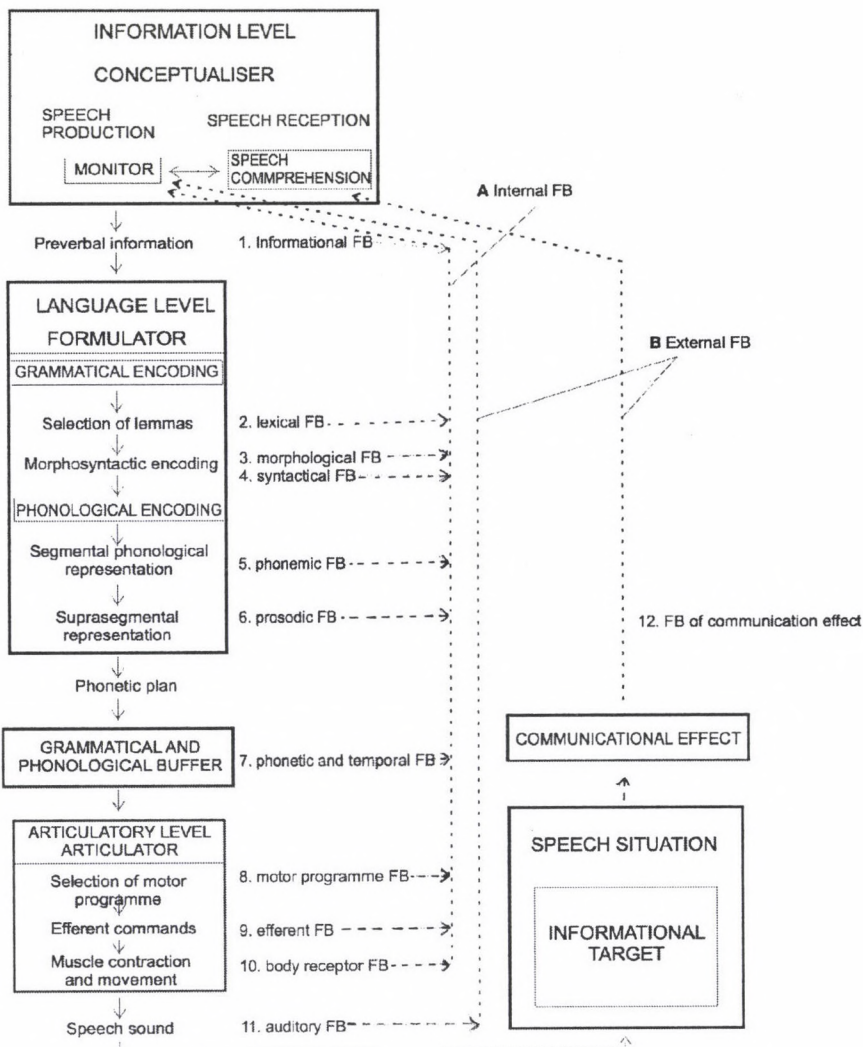


Figure 1.  
Speech production model (adapted from Postma 2000)



### **Feedback in the model of speech production**

It was already mentioned that the speaker controls and corrects his speech production by means of a number of feedback loops. They can be classified on the basis of various criteria. By one of them feedbacks can be classified as directive, corrective, and adaptive. Directive feedback (closed loops) is in fact feed forward. It is parallel and simultaneous with the movement of speech organs and the speaker can anticipate possible deviations from the desired target on the basis of feedback information about the current state of the articulators and immediately correct the movement before the error is made, thus avoiding the necessity of complete reprogramming and repeating of the movement (Abbs et al. 1984). That kind of feedback is very fast, just a few tenths of a second. Already produced errors requiring the reprogramming and correction are registered by means of corrective feedback. Those kinds of errors are characteristic of the informational and language levels of speech production. Finally, language learning and motor skills and harmonization of speech production with the general conditions in the communication channel or with sudden changes of these conditions are controlled by means of adaptive feedback. Borden and Harris (1980) classify speech production feedback loops as intrinsic or internal, which are active before the movement is performed by means of which the execution of the movement is controlled and which are based mostly on the proprioceptive feedback channels. According to their classification a motor movement just executed is controlled by means of external feedback. Škarić (1991) mentions the so-called internal feedback at the level of the central nervous system, extraauditory or proprioceptive, auditory and external or social feedback.

By means of feedback mechanisms speakers can correct their errors even during the process of speech production planning and before motor execution. Those are covered repairs. Speakers discover them by means of introspection or they are realised as disfluencies, which can be a sign of the speaker's engagement in correction of such kind of errors. If the error is already produced, then it is an open one and it can be registered.

We are going to describe 12 types of feedback (FB) connections mentioned in the present speech production model (Figure 1).

**1. Informational FB.** Speakers check whether the formulation of the information in the preverbal step corresponds to their intentions, whether the information corresponds to their ideas, attitudes, emotions, and whether it is adequate for the listener and for the speech situation, including the conditions in the communication channel.

**2. Lexical FB.** Speakers check whether the retrieval of lemmas is adequate regarding conceptual structure of the information.

**3. Morphological FB.** Speakers check whether morphological frames are adequately opened and filled and whether syntactic rules are obeyed.

**4. Syntactical FB.** Speakers check whether syntactic frames are adequately opened and filled.

**5. Phonemic segmental FB.** Speakers control the selection and sequencing of the phonemic structure of the word, including distinctive features.

**6. Phonemic prosodic FB.** Speakers control the adequacy of the realisation of phonemic prosodic components of tone, intensity and duration.

**7. Phonetic and temporal FB.** The utterance planned from the phonetic point of view is prepared and placed in the buffer and its elements are ready to be built in the utterance, which must be linear from the temporal point of view. At this level the sequencing of certain elements and their timing are controlled.

**8. Motor programme FB.** Speakers control the selection of an adequate motor programme to realize the phonetic plan. It can be supposed that at this point the rules of assimilation are activated and discrete units of the language level are marked as parametric speech units.

**9. Efferent FB.** Speakers control the adequacy of motor commands.

**10. Body receptor FB.** Various sensory modalities (tactile, kinaesthetic, vibrotactile, thermal, somatosensory) control the precision of the speech organ movement and its adequacy by means of direct feedbacks.

**11. Auditory FB.** Speakers listen to their own speech and control its adequacy.

**12. FB of communicational effect.** Speakers monitor the communicative effect of their speech and on the basis of verbal and nonverbal reactions of their listeners decide about the necessity to correct errors.

The versatility of feedback forms makes possible and safe the process of speech communication and, if necessary, the correction and specification of information.

In the present paper hesitations in speech production are analysed in the case when constituents of the motor programme are repeated depending on the sort of articulatory juncture. An articulatory juncture can be defined as a complex of two adjacent constituents which are either of the same kind (two sounds, two syllables, two morphemes, etc.), or of a different kind (a proclitic and the corresponding host-word and an enclitic and a corresponding host-word) plus the boundary between these constituents. Types of articulatory junctures investigated in this paper are: junctures between sounds, between syllables, between a stressed word and an unstressed word (enclitics and proclitics), and those between stressed words. An articulatory juncture is also referred to as an articulatory joint (Horga 2005).

### **Aim and method**

The aim of the investigation was to find out the characteristics of repetitions in speech production. Speakers often repeat parts of an utterance, so the question can be raised as to whether it is possible to interpret the repetition as



some kind of speech error which they want to repair, or it is just one kind of disfluency, in which speakers hesitate because they need additional time for the following planning and execution of speech. If hesitation is undesired, the repetition can be considered an error on the level of small speech units, especially if the repeated constituents are relatively short and if they are not accompanied by some other kind of hesitation to gain time, i.e. if they are repeated immediately without a pause. If repeated constituents are longer units and if they are accompanied by other forms of hesitations, the kind of repetition can be considered a disfluency.

The basic question in studies dealing with the mechanisms of the speech production is understanding the dichotomy between the representational, language level of the formulator and the physical, executive level of the articulator. The first level is described as a system of abstract, invariant and discrete units (phonemes) and the second one as a system of changeable, continuous and partly simultaneous patterns of speech movements by means of which the also changeable and indiscrete acoustical signal is produced. It is generally accepted that the variability and the difficulty in the parsing of a speech signal into discrete units are the result of the universal phenomenon of coarticulation. Two neighbouring phonemes make an articulatory juncture (Horga 2005). Boundaries between various phonetic and language units can be placed between neighbouring sounds. These boundaries are seen as integral parts of articulatory junctures. Boundaries between language units in the process of speech production are represented on the language level of formulator. Those elements of the representational level can be somehow present in the speech signal and marked by various physical markers thus making connections between physical, articulatory and acoustical levels on the one hand and representational, language and conceptual levels on the other. Articulatory junctures can be classified into nine groups, depending on the constituents of the juncture: 1. sounds, 2. syllables, 3. morphemes, 4. proclitics, 5. enclitics, 6. phonetic words, 7. clauses, 8. sentences and 9. paragraphs. Junctures of higher levels can be simultaneously junctures of lower levels, e.g. the proclitic juncture in the word *od#rasti* 'grow up' is, at the same time, a morphological, a syllabic and a sound juncture. Speech elements make junctures and it can be asked how strong certain junctures are, or in other words, how easy it is to realize boundaries between the two elements of the articulatory junctures in their clearest form, i.e. as a pause. It is natural that the juncture boundary is more easily detected in the speech signal if the constituents making up the juncture are of a higher level. So we can say that the sound juncture is stronger and the boundary between its members weaker than in the juncture of sentences. It can be expected that in the process of speech production, hesitations and disfluencies will occur in articulatory junctures of a higher level more often than in the junctures of a lower level.



Analysing Meringer's written corpus of speech errors from 1895, Nootboom (1980) formulated the principle rule applied by the speaker when the utterance is interrupted: interrupt the utterance immediately after noticing that the error is made. That implies that a **delay** in interrupting corresponds to a delay in discovering the error. But hesitation can occur even before the error is made. That is called covered repair, and in that case, it remains unknown what the reason of hesitation or the duration of latency was between the causes and the repair. In 18% of cases hesitations occurred within the word, in 51% immediately after the word in which the error occurred and in 31% one or several syllables occurred after the erroneous word. The hypothesis that the speaker delays the correction of the error up to some lexical boundary was not proved because when hesitation was zero, it occurred at the lexical boundary in 74% of the cases, while if hesitation was delayed, it occurred at the lexical boundary in 66% of the cases only. If the speaker delays his hesitation and the repair up to the lexical boundary, then the percentage of placement of hesitation on the lexical boundary should be higher than in the case of immediate hesitation. Besides that, in 22% of cases hesitation occurred within the word and in 39% it occurred in an incomplete syllable. All these data prove the already mentioned hypothesis about realization of hesitation immediately after noticing the error (Levelt 1995).

The present paper investigates hesitations of utterances when the speaker has not made an error and he does not repair anything but he continues the utterance repeating one part. Recorded radio programmes were live broadcasts, which enabled us to analyse spontaneous speech. The speech material consisted of five broadcasting programmes with a total duration of 80 minutes, with the participation of ten speakers. Duration measurements of analysed constituents were performed by CoolEdit software.

For the analysis of the recorded material, repetitions were noted, the number of repeated sounds and syllables counted, the duration of interrupted constituents, hesitations and repeated constituents were measured and the placement of hesitations was classified by type of articulatory juncture. The percentage of hesitation occurrence for different articulatory junctures was calculated. Hesitations were classified as: (1) having zero duration, if the repetition of the interrupted constituent was very fast and immediate; (2) a silent pause; (3) a filled pause (non-phonemic component); (4) having phonemic components or (5) a combination of these elements (Figure 2). The model is called the interrupted constituent, hesitation and repeated constituent model (IHR). (In Croatian the model is abbreviated to *PPP – prekinuti segment, prekid, ponovljeni segment*). Figure 2 shows the procedure of duration measurement of one IHR constituent.

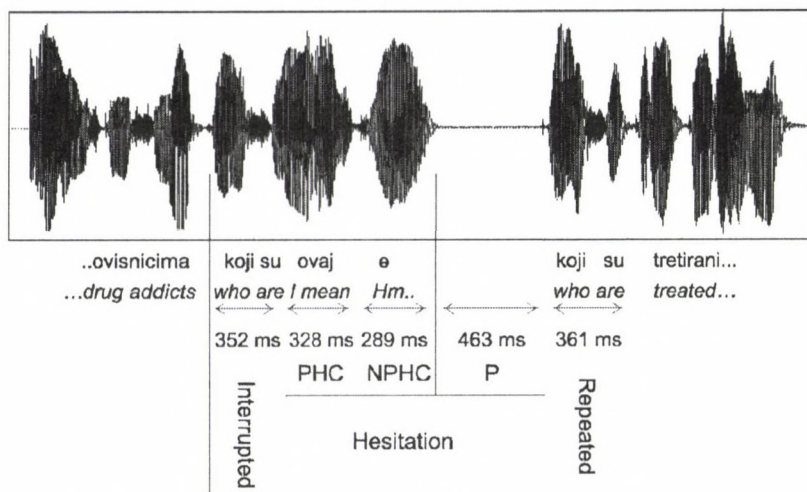


Figure 2.

Duration measurement of the components of an IHR constituent (PHC – phonemic component, NPHC – non-phonemic component, P – pause)

## Results and discussion

### Duration of the components of IHR-constituent

In the analysed speech sample IHR-constituents occurred 103 times, i.e. one IHR-constituent occurred every 48 seconds. In Figures 3 and 4 average durations of the components of the IHR constituent are shown. On average speakers interrupt their utterance after 3.43 sounds or 1.64 syllables or 346 milliseconds. Minimal duration of the interrupted constituent was one sound (*u#uvijek* – ‘always’) or one syllable, which can be one vowel, and the maximal duration was 14 sounds or 6 syllables (*koji će ga implementirati#implementirati u novoj organizaciji* – ‘which one will implement in new organization’). The duration of hesitation was 137 ms on average. But in 52% of the cases its duration was zero, because the speaker repeated the interrupted constituent immediately after hesitation (*go#golema* – ‘enormous’). If the average duration of hesitation is calculated with cases of zero hesitation (immediate repetition) left out of consideration, the average duration of hesitation was 1351 ms (*vratio bih se# još prije nego čujemo slušatelje #vratio bih se na pitanje* – ‘I would come back# before we hear the listeners# I would come back to the question’). It was mentioned that with regard to its structure, a hesitation could be classified as having zero duration, or it could be a silent pause, a filled pause with a non-phonemic component, a phonemic

component, or a combination of those three kinds of disfluencies. Silent pauses had an average duration of 165 ms, filled pauses had an average duration was 279 ms, hesitations with a phonemic component had an average duration 438 ms, and combined hesitations had an average duration of 761 ms. The repeated constituent naturally consisted of the same number of sounds and syllables as the interrupted one but its average duration was 277 ms, 69 ms shorter than that of the interrupted constituent, a statistically significant difference ( $t$ -test = 2.80,  $p < 0.005$ ).

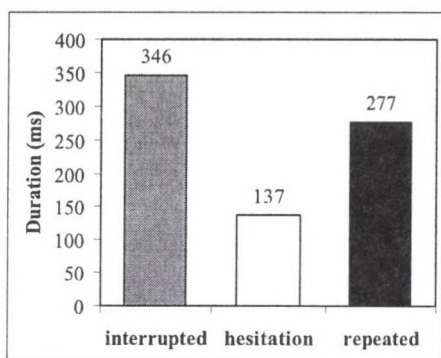


Figure 3.

The average duration of the components of the IHR constituent

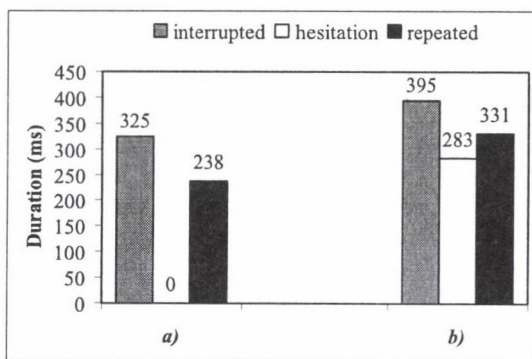


Figure 4.

The average duration of the components of the IHR constituent depending on the duration of hesitation: a) hesitation 0 ms, b) hesitation M = 283 ms



The duration of hesitation was related to the duration of the other two components of the IHR constituent, i.e. to the duration of the interrupted constituent and to the repetition (Figure 4). Two categories of IHR constituents were analysed. The first category consisted of the constituents in which the duration of hesitation was zero and the average duration of the interrupted component was 325 ms and that of the repeated component 238 ms. If the duration of hesitation had some value (the average was 283 ms) the interrupted component average value was 395 ms and repeated component 331 ms. The difference between the two categories was significant ( $p < 0.01$ ). In general, we can say that in the first category with zero duration of hesitation the other two components of the IHR constituent were shorter than in the second category. The explanation of the difference can again be that the former category involves the articulatory level but the second one the higher processing levels.

#### **The position of hesitation**

Apart from the duration of the components of the IHR constituent, the position of hesitation classified by articulatory juncture was also analysed. Gósy (2007) says that the pauses within a word in Hungarian occur mainly at morpheme boundaries and the duration of these pauses seem to be dependent on the word retrieval process. Their average durations were: between the prefix and stem 343.2 ms, within the compound 2190.8 ms, between stem and suffix 170.6 ms and within a stem 107.3 ms. In the present research the following percentages of hesitation realizations were obtained: between sounds 6.8%, between syllables 6.8%, between morphemes 5.8%, proclitic 32.0%, enclitic 10.7% and phonolexical 36.9%. The Chi-square test showed that the differences between realised percentages were statistically significant ( $p < 0.01$ ). When the values of percentages of hesitation constituents between sounds, between syllables and between morphemes were combined, a total percentage of 19.4% was obtained for hesitations within the lexical words, which was close to the percentage quoted by Levelt (1995) as the percentage of hesitations within the erroneous word. The remaining 78% of hesitations belong to hesitations between lexical units: 42.7% between clitics and the full word and 36.9% between full words. That fact can be explained by the influence of the language structure of the utterance on the occurrence of hesitation and the more frequent appearance of hesitations and repetitions between lexical units than within words. The average duration of hesitation depending upon the type of the articulatory juncture and the following values were obtained: between-sounds 24 ms, between syllables 19 ms, between morphemes 34 ms, proclitic 34 ms, enclitic 45 ms and phonolexical 57 ms. While these values were consistent with the trend of longer hesitation durations in the articulatory juncture of higher levels than in the junctures of lower levels, the differences between durations were not statistically significant.

What can be the explanation of the mechanisms underlying the IHR model? One possible explanation can be that the model simply represents a

complex motor programme, which can not be completely performed; at some point the programme is interrupted although up until that moment it was correctly executed and in the next attempt the speaker succeeds in completing it. The repeated constituent was 69 ms (20%) shorter on average than the interrupted constituent (277 ms vs. 346 ms). This could be explained on the one hand by the fact that the repeated constituent was already produced and so trained during the first execution before hesitation, and on the other hand by the psychological fact that the speaker was probably aware that hesitation caused the speaking rate to change and he tried to regain the lost time by faster rate of speaking in the repeated constituent. Finally, the repeated constituent was usually connected to the following larger constituents, which could have also contributed to the faster rate of speaking. The theory of activation radiation (Dell 1986) could possibly explain the occurrence of speech errors and the IHR model. The theory states that during speech production various nodes are activated at the various levels of the mental speech representation (distinctive features, phonemes, syllables, morphemes, words) and that incorrect activation of certain nodes causes speech error occurrence. Hesitation of the articulatory programme can be explained by the activation model, according to which the realization of a certain unit depends on its activation, but also on the activation of other units representing the associatively organized network. Units which are serially built into the utterance are also activated. From that point of view the unit which has been in the process of realization must be activated at a precise moment and with appropriate intensity. But it must also be deactivated with adequate timing and decreasing intensity to empty the place for the next unit. So the IHR model can be explained by the prolonged activation of the already realised unit and by a weak activation of the next unit to replace the previous one. That explanation refers particularly to cases in which the interrupted constituent is very short and where the duration of hesitation is zero, i.e. in cases in which there is no disfluency between the interrupted and the repeated constituent. These are the cases of inadequate activation of the articulatory programme. The characteristic of this kind of reactivation of the motor programme and avoiding hesitation is immediate and no special time is necessary to perform it. Cases in which the interrupted constituent is repeated several times could be included in that type of IHR model (*to bi≠bi, bi, bi≠bi koštalo* – ‘that would, **would**, would, **would**, would cost’).

The second type of IHR mechanism is represented by cases in which some kind of disfluency appears during hesitation, meaning that the speaker plans the following constituents of the utterance during the time of execution. For example in the utterance *zašto od≠hm od stigmatizacije* – ‘why from hm, from stigmatization’ the speaker repeats the proclitic *od* and in hesitation he inserts the non-phonemic component **hm** to choose and plan the pronunciation of the infrequent and, from the articulatory point of view, complicated



word *stigmatizacija*. Or in the already mentioned example with the longest hesitation (*vratio bih se# još prije nego čujemo slušatelje #vratio bih se na pitanje* – 'I would come back# before we hear the listeners I would come back to the question') the speaker adds new information on the conceptual level and that is why he interrupts the utterance.

### **The relationship between the interrupted constituent and hesitation**

The relationship between the duration of interrupted constituent and the hesitation type can also shed light on the mechanisms underlying the IHR model. The duration of the interrupted constituent, if it is followed by a hesitation of zero duration, is 313 ms on average, and if it is followed by a hesitation which has some kind of disfluency, 385 ms on average, 72 ms longer. Although that difference is not significant at the level  $p < 0.05$ , it is very close to that level ( $t = 1.73$ ,  $p < 0.09$ ). This could mean that in the first case the utterance is interrupted because of problems in the performance of the motor programme at the beginning of the utterance and immediately corrected, while in the second case the cause of hesitation can be found at the higher levels of speech production, so that the duration of the motor programme is longer—it is interrupted in order to gain more time. During the prolonged duration of hesitation, the speaker tries to find the way how to continue the realization. That means that errors can occur at the level of selection and at the level of realization. Correction of errors at the level of selection require abandoning the selected and incorrect programme and reprogramming, which is a serial process and involves the activity of the central brain mechanisms. Errors at the level of performance are of a milder character and their correction is possible serially during the execution and they can be corrected automatically at the lower levels of the speech production.

From the point of view of feedback mechanisms which are active in the IHR model, interrupted constituents and hesitations are relatively short and it can be supposed that the control of speech production for that constituent is performed by means of fast internal feedback, especially when a hesitation is of zero duration. But it is possible that when the interrupted constituent and hesitation are both relatively long, the auditory feedback is included. In the analysed speech sample when the IHR constituent occurred, listeners never warned the speaker about the problem in his speech production, indicating that the feedback of the communicational effect, which is relatively slow, was not activated and that the speaker himself controls mechanisms of the IHR model realization.

### **Conclusion**

In conclusion it is possible to say that the IHR constituent presents evidence from speech production indicating that speech production is not a perfect mechanism and shows the obstacles speakers must overcome on the level of performance, especially in the motor programme or in planning. The IHR



model is controlled mostly by means of the internal feedback, which is very fast, significantly faster than the feedback of the communicational effect. On the basis of the durational characteristics of the interrupted constituent, the hesitation and the repeated constituent, it is possible to classify IHR constituents into those in which the motor programme is interrupted and those in which constituents of disfluency occur, which can indicate problems in speech production at language and information levels. The relations between representational speech and language levels and their physical realisation were analysed depending on the position of hesitation in the articulatory juncture and it was concluded that the representational level influences the distribution of IHR constituents so that their frequency is higher in the articulatory juncture between larger language units than between smaller ones.

In this research only one kind of discourse was analysed, namely spontaneous conversation in radio broadcasting. It would be interesting to analyse the characteristics of the IHR constituents in other types of discourse. Also it would be interesting to compare the IHR constituents of various speakers and to see whether the IHR constituent represents one type of speech production strategy which is characteristic for some speakers, [in the same way as the frequency and kinds of disfluencies are. It would be valuable to perform this kind of research on a greater sample of speech material to obtain more reliable data about the duration and frequency of the IHR constituent.

### References

- Abbs, James H. – Gracco, Vincent L. – Cole, Kelly J. 1984. Control of multimovement coordination: sensorimotor mechanisms in speech motor programming. *Journal of Motor Behavior* 16/2. 195–231.
- Borden, Gloria J. – Harris, Katherine S. 1980. *Speech science premier*. Williams and Wilkins, Baltimore–London.
- Corder, Stephen Pit 1973. *Introducing applied linguistics*. Penguin Education, Harmondsworth, England.
- Dell, Gary S. 1986. A spreading-activation theory in sentence production. *Psychological Review* 93. 283–321.
- Fairbanks, Grant A. 1954. A theory of the speech mechanism as a servo system. *Journal of Speech and Hearing Disorders* 19. 133–139.
- Garret, Merrill F. 1980. Levels of processing in sentence production. In Butterworth, Brian (ed.): *Language production: Vol. 1. Speech and talk*. Academic Press, London.
- Gósy, Mária 2007. Disfluencies and self-monitoring. *Govor* 2. 91–110.
- Hartsuiker, Robert J. – Kolk, Herman H. J. – Lickley, Robin J. 2001. Stuttering on function words and content words: A computational test of the Covert Repair Hypothesis. In Maassen, Ben – Hulstijn, Wouter – Kent, Raymond – Peters, Hermann F. M. – van Lieshout, Pascal (eds.): *Speech motor control in normal and disordered speech*, 6569. Nijmegen: Uetgeverij Vantilt.
- Horga, Damir 1997. Samoispravljanje u govornoj proizvodnji [Self-correction in speech production]. *Suvremena lingvistika* 43–44. 91–104.

- Horga, Damir 2005. Izgovorni zglobovi u spontanom govoru [Articulatory joints in spontaneous speech]. In Pranjković, Ivo (ed.) *Od fonetike do etike. Zbornik o sedamdesetoj godišnjici prof. dr. Josipa Silića*. Disput, Zagreb, 63–74.
- Horga, Damir – Požgaj Hadži, Vesna 2005. Slušateljeva redakcija govornika [Listeners edit speakers]. In Granić, Jagoda (ed.): *Semantika prirodnog jezika i metajezik semantike*. HDPL, Zagreb–Split, 301–310.
- Jaeger, Jeri J. 2005. *Kids' slips. What young children's slips of the tongue reveal about language development*. Lawrence Erlbaum Associates Publishers, Mahwah, New Jersey and London.
- Kurhila, Salla 2001. Correction in talk between native and non-native speakers. *Journal of Pragmatics* 33. 1083–1110.
- Levelt, Willem J. M. 1983. Monitoring and self-repair in speech. *Cognition* 14. 41–104.
- Levelt, Willem J. M. 1995. *Speaking. From intention to articulation*. A Bradford Book and MIT Press, Cambridge, Massachusetts and London, England.
- Nakatani, Christine H. – Hirschberg, Julia 1994. A corpus-based study of repair cues in spontaneous speech. *Journal of the Acoustical Society of America* 95/3. 1603–1616.
- Nooteboom, Sieb G. 1980. Speaking and unspeaking: detection and correction of phonological and lexical errors in spontaneous speech. In Fromkin, Victoria A. (ed.): *Errors in linguistic performance: slips of the tongue, ear, pen and hand*. Academic Press, New York, 87–96.
- Ohlsson, Stellan 1996. Learning from performance errors. *Psychological Review* 103/2. 241–262.
- Postma, Albert 2000. Detection of errors during speech production: review of speech monitoring models. *Cognition* 77. 97–131.
- Postma, Albert – Kolk, Hermann – Povel, Dirk-Jan 1990. On the relations among speech errors, disfluencies and self-repairs. *Language and Speech* 33/1. 19–29.
- Rieger, Caroline L. 2003. Relations of self-repair strategies in English and German conversations. *Journal of Pragmatics* 35. 47–69.
- Schmidt, Richard A. 1988. *Motor control and learning: a behavioral emphasis*. Erlbaum, Hillsdale.
- Škarić, Ivo 1991. Fonetika hrvatskoga književnog jezika [Phonetics of the Standard Croatian]. In Babić, Stjepan – Brozović, Dalibor – Moguš, Milan – Pavešić, Slavko – Škarić, Ivo – Težak, Stjepko: *Povijesni pregled, glasovi i oblici hrvatskoga književnog jezika*. HAZU and Globus Nakladni zavod, Zagreb.

## A SZEMÉLYAZONOSÍTÁS LEHETŐSÉGE A BESZÉDANYAG IDŐTARTAMÁNAK FÜGGVÉNYÉBEN

Nikléczy Péter – Gósy Mária

### Bevezetés

A beszélő személy kilétének megállapítása – a beszéde alapján – közel két évtizede a fonetikusok és a beszédtechnológusok érdeklődésének középpontjában áll. A cél tehát a személy azonosságának a lehetőség szerinti legpontosabb meghatározása rögzített hangminta alapján. Az eddigi vizsgálatok több oldalról próbálták meghatározni azokat az univerzális és nyelvspecifikus beszédösszetevőket, amelyek a legjellemzőbbek az adott beszélőre (Nolan 1991; Hollien 2002; Rose 2002). Kutatások folytak az alaphang értékének meghatározó szerepére, a formánsstruktúra személyfüggőségére, a nazalitás azonosító szerepére, a hangszalagok működésének jellegzetességére, a szóhasználati sajátosságokra, a szövegben esetlegesen előforduló kiejtési hibákra, avagy éppen a beszélő személy életkorának, alkatának a beszédmintája alapján történő leírására (Nolan 1977; Hollien 1995; Nikléczy 1996; McDougall–Nolan 2007). Ezen kutatások mindegyike természetesen hozzájárult az egyre pontosabb személyazonosításhoz, azonban még mindig sok a kérdés, amelyek további kutatást igényelnek a megoldáshoz.

A korábbi eredményeket, illetve a jelenleg folyó kutatásokat is felhasználva a gyakorlati kriminalisztikában már jelentős mértékben teret hódított a fonetikai alapú személyazonosítás. A kriminalisztikai beszélőazonosítási vizsgálat eredményességét alapvetően meghatározza a rendelkezésre álló hangminta időtartama és akusztikai tisztasága. A gyakorlatban jellemzően előfordul, hogy egy korábban rögzített (inkriminált) hanganyagot kell összehasonlítani a későbbiekben a gyanúsítottnak tekintett személy beszédével. Ezekben az esetekben a gyanúsított személytől akár hetek, sőt hónapok elteltével vesznek hangmintát, amely alapjául szolgál a bűnügyi fonetikai vizsgálatnak. A vizsgálat sorozatot nehezíti, hogy sok esetben akusztikailag nem megfelelő körülmények között rögzített hangminta áll rendelkezésre. Eppen ezért nagy a jelentősége annak, ha megállapítható, hogy a kriminalisztikai beszélőazonosítási vizsgálatok elvégzéséhez milyen időtartamú hanganyagra van minimálisan szükség.

A jelen kutatásunknak az volt a célja, hogy megállapítsuk, a műszeresen elvégezhető fonetikai személyazonosítási vizsgálat sorozat milyen minimális időtartamú hangminta esetén ad nagy valószínűséggel megbízható ered-



ményt. Feltevésünk szerint, ha ugyanazon személy stúdióköörülmények között rögzített felolvasását egyre csökkenő időtartamokban összehasonlítjuk a teljes időtartamú anyaggal, akkor meghatározható az az időtartam, amely még éppen elfogadható azonosságot mutat a teljes időtartamú beszédanyaggal. Ez azt jelenti, hogy egy meghatározott időtartamnál rövidebb hangminta már nem lesz jól alkalmazható a beszélő személy azonosítására.

### **Kísérleti személyek, anyag, módszer**

A kutatáshoz stúdióköörülmények között rögzített, ugyanazon szöveg felolvasásait használtuk 12 férfi produkciójában (életkoruk 25 és 76 év közötti volt). A módszertani azonosságot a BEA adatbázis tette lehetővé (azonos felvételi körülmények, azonos szöveg, azonos feladat, vö. Gósy 2008). A felhasznált szöveget az 1. ábra tartalmazza.

#### **Méreg vagy vitamin?**

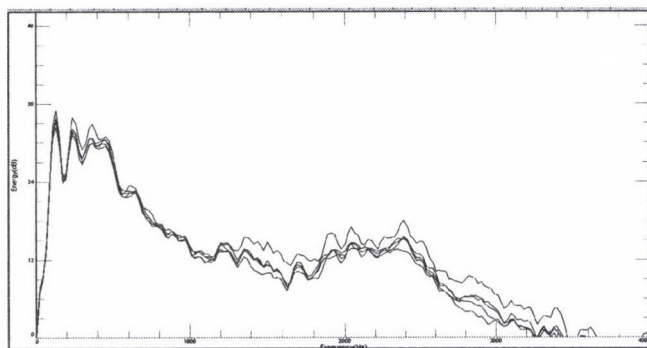
A zöldségeken és gyümölcsökön lévő növényvédő szerek egy része lemoshatatlan, ezért fogyasztás során a szervezetünkbe kerül, ahol immunrendszeri zavarokat, daganatos megbetegedést okozhat. Télen és kora tavasszal nehéz ellenállni a boltokban és a piacon kínált friss salátának, reteknek vagy éppen az epernek. Aki mégis beadja a derekát kétszeresen is csapódhat. Egyrészt a várt ízvilág messze elmarad a szezonális termékektől, viszont ennél talán súlyosabb problémát jelentenek a primőr zöldségekbe, gyümölcsökbe felszívódott lemoshatatlan növényvédő szerek, amelyeket gyanútlanul elfogyasztunk. Hazánkban évente közel 20 ezer tonna növényvédő szert használnak fel. Ez azt jelenti, hogy Magyarországon évente egy lakosra két kiló méreg jut. A legtöbb háztartásban télen és tél végén kedvelt vitaminforrás a primőr fejes saláta, amely a szakhatóságok elmúlt években végzett vizsgálatai alapján a legtöbb lemoshatatlan mérget tartalmazza. A kismamák sem sejtik, hogy a bébiételek húsz százalékában találtak a növényvédő szerekből viszsamaradt vegyi anyagokat. A hazai hatóságok rendszeresen vizsgálják a boltokban és piacokon kapható zöldségek, gyümölcsök, illetve a feldolgozott és a bébiételek növényvédőszer-tartalmát is. A vizsgálatok szerint a hazai eredetű termékekben valamivel kisebb, az importált árukban pedig magasabb arányban mutathatók ki a különféle növényvédő szerek. A zöldségekbe, a gyümölcsökbe olyan mérgek kerülhetnek, amelyek még mosás után is hormonális, immunrendszeri zavarokat, daganatos megbetegedéseket okozhatnak. A permetszerek hatóanyagainak jelentős része nem a kezelni kívánt növényekre és kártevőkre kerül, hanem a környezetbe jutva károsíthatja a környező területek élővilágát, szennyezheti a talajt és a vizeket – olvasható a Levegő Munkacsoport tanulmányában.

#### **1. ábra**

A felolvasott szöveg a BEA adatbázisból

A beszédszakaszok elemzésére a hosszú idejű analízis elnevezésű vizsgálatot (LTA) alkalmaztuk. Az LTA a különböző időtartamú beszédrészek átlagolt

energispektrumának függvényét adja meg. A kutatáshoz a KAY PENTAK Multi-Speech nevű analízáló programját használtuk. Az LTA-elemzés beállításai a következők voltak: elemzési méret 2048 pont, Blackman ablak használata, frekvenciatartomány: 0–3,2 kHz. Az így kapott energiaspektrumok numerikus értékeit (beszélőnként 152 adat) statisztikailag hasonlítottuk össze. Ez az eljárás lehetőséget ad az LTA-függvények összehasonlíthatóságára is, hiszen vizuálisan nem lehet egyértelműen eldönteni, hogy két vagy több görbe mikor tekinthető azonosnak (2. ábra). A kutatás során a teljes felolvasásokat minden beszélő esetén folyamatosan feleztük (vö. 3. ábra). A teljes hanganyag és az egyre csökkentett időtartamú minták LTA-értékeinek statisztikai összehasonlítása megadhatja azt a minimális időtartamú beszédmintát, amely a beszélő azonosításának szempontjából az eredeti (teljes) hanganyaggal még azonosnak tekinthető. Feltételeztük, hogy amennyiben az elemzési eredmények legkisebb időtartamértékei mind a 12 személynél azonos vagy nagyon hasonló értéket mutatnak, akkor meghatározható a minimálisnak tekintett, reálisan elemezhető beszédminta időtartama. Ezért megvizsgáltuk azt, hogy a 12 személy teljes hanganyagának LTA-adataiban jelezhető-e különbség (minden beszélőt minden beszélővel összevetve). A statisztikai elemzés (Spearman- és Kendall-féle korreláció) azt mutatta, hogy a beszélők különbözősége  $r = 0,9$ -es korrelációs együttható esetén igazolható. A további elemzések során ezért minden esetben ezt az értéket vettük figyelembe. Az időtartam-változtatás paramétereit és a magánhangzók előfordulását vizsgáltuk.



2. ábra

Az LTA-görbék alakulása különböző időtartamú beszédmintákban ugyanazon személy ejtése alapján

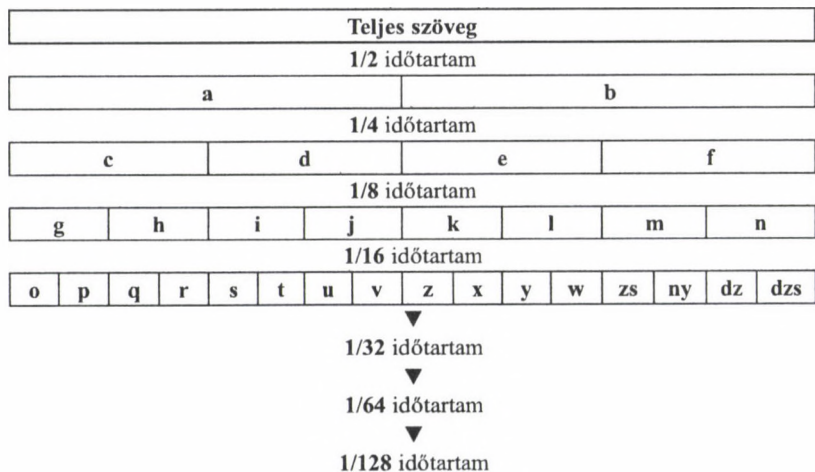
Meghatároztuk a szövegben előforduló magánhangzó-fonémák számát, illetve gyakoriságát is. Ebben az esetben a cél az volt, hogy kijelöljük annak a szövegrésznek a hosszát, amely a magánhangzó-előfordulás arányát tekintve statisztikailag azonosnak tekinthető a teljes anyaggal. A zöngés mássalhang-

zókát azért nem vettük figyelembe, mivel az artikuláció során gyakran zöngétlenedhetnek.

Kiindulásként megmértük a 12 személy által felolvasott szöveg időtartamát (a szünetekkel együtt), valamint meghatároztuk az adatközlő beszédtempóját (1. táblázat). (A különböző személyeket  $F_n$  kódokkal jelöltük.) A szöveg szavainak száma: 202 (a határozott és határozatlan névelők nélkül).

1. táblázat: A felolvasások időtartama és a beszélők beszédtempója

| Beszélők | A felolvasás időtartama (s) | Beszédtempó (szó/s) |
|----------|-----------------------------|---------------------|
| F1       | 124,68                      | 1,62                |
| F2       | 104,30                      | 1,93                |
| F3       | 150,00                      | 1,34                |
| F4       | 111,06                      | 1,81                |
| F5       | 115,00                      | 1,75                |
| F6       | 118,50                      | 1,70                |
| F7       | 127,22                      | 1,58                |
| F8       | 138,00                      | 1,46                |
| F9       | 115,10                      | 1,75                |
| F10      | 123,05                      | 1,64                |
| F11      | 137,00                      | 1,47                |
| F12      | 132,41                      | 1,53                |
| Átlag    | 124,69                      | 1,62                |



3. ábra

Az időtartam folyamatos felezésével kapott beszédminták jelölése



A teljes időtartamok azt az időt mutatják, amelyről a teljes LTA-vizsgálat készült. Ezt az időtartamot feleztük személyenként 1/128-ad időtartamig. A csökkentett időtartamú beszédszakaszokat a könnyebb azonosítás érdekében betűjelekkel láttuk el az 1/16-is időtartamokig, azt követően pedig törtszámokat használtunk (3. ábra). A különböző beszélők csökkentett beszédmintáit átlagoltuk.

A statisztikai vizsgálatok elvégzésére az SPSS 13. programot használtuk 99%-os szignifikanciaszint beállításával (a Spearman- és a Kendall-féle korrelációelemzést alkalmaztuk).

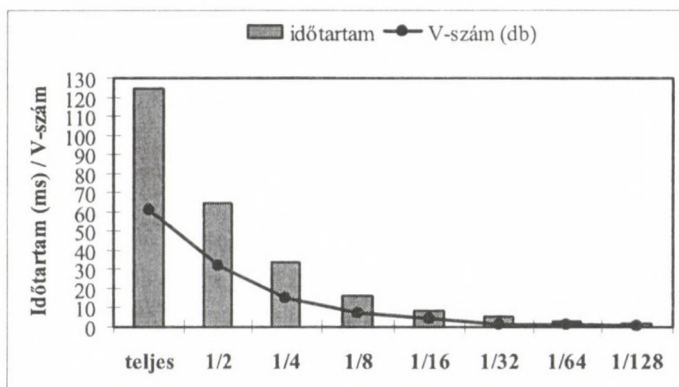
### Eredmények

Az elemzések során egyfelől a magánhangzó-előfordulásokat adatoltuk, majd pedig az LTA-értékeket összesítettük. A **magánhangzó-előfordulás** vizsgálatokor először meghatároztuk, hogy a teljes időtartam(ok) folyamatos felezésének eseteiben hol tart a beszélő személy a szöveg felolvasásában. A felezésekkel kapott szövegrészek időtartamát átlagoltuk az összes beszélő felolvasásában, majd valamennyi szövegrészben megszámoltuk a magánhangzókat, és ezeknek az adatoknak is meghatároztuk az átlagát (2. táblázat). A táblázat a felolvasott szövegek felezési időtartamait (a 12 beszélő esetében), valamint az előforduló magánhangzók számát összesíti az 1/8-ra csökkentett időtartamú szövegrészig.

2. táblázat: A beszédanyag felezési időtartam átlagai (1/8-os értékig) és a magánhangzók száma

| A minta jele és tartama (s) | A magánhangzók előfordulása (db) |    |     |    |    |   |    |   |    |    |    |   |    |   |
|-----------------------------|----------------------------------|----|-----|----|----|---|----|---|----|----|----|---|----|---|
|                             | a                                | á  | e   | é  | i  | í | o  | ó | ö  | ő  | u  | ú | ü  | ű |
| teljes (124,69)             | 155                              | 44 | 166 | 63 | 49 | 5 | 52 | 8 | 26 | 16 | 11 | 3 | 10 | 1 |
| a (64,49)                   | 82                               | 22 | 88  | 38 | 21 | 3 | 30 | 3 | 13 | 7  | 4  | 3 | 4  | 0 |
| b (59,85)                   | 73                               | 22 | 78  | 25 | 28 | 2 | 24 | 5 | 13 | 9  | 7  | 0 | 6  | 1 |
| c (30,55)                   | 39                               | 10 | 44  | 19 | 12 | 2 | 14 | 0 | 5  | 3  | 1  | 0 | 3  | 0 |
| d (33,94)                   | 43                               | 8  | 46  | 16 | 4  | 1 | 14 | 3 | 8  | 4  | 3  | 3 | 1  | 0 |
| e (29,88)                   | 45                               | 12 | 31  | 16 | 19 | 0 | 12 | 3 | 7  | 4  | 2  | 0 | 2  | 1 |
| f (29,97)                   | 24                               | 10 | 47  | 9  | 9  | 2 | 12 | 2 | 6  | 5  | 5  | 0 | 4  | 0 |
| g (16,11)                   | 22                               | 3  | 24  | 12 | 5  | 0 | 11 | 0 | 7  | 2  | 1  | 0 | 3  | 0 |
| h (16,03)                   | 17                               | 7  | 20  | 7  | 7  | 2 | 3  | 0 | 0  | 1  | 0  | 0 | 0  | 0 |
| i (14,80)                   | 15                               | 4  | 23  | 8  | 2  | 1 | 7  | 1 | 5  | 3  | 2  | 2 | 1  | 0 |
| j (17,52)                   | 28                               | 4  | 23  | 8  | 2  | 0 | 7  | 2 | 2  | 1  | 1  | 1 | 0  | 0 |
| k (15,01)                   | 24                               | 7  | 15  | 7  | 9  | 0 | 10 | 2 | 4  | 2  | 0  | 0 | 1  | 0 |
| l (14,87)                   | 21                               | 5  | 16  | 9  | 10 | 0 | 2  | 1 | 3  | 2  | 2  | 0 | 1  | 1 |
| m (14,89)                   | 18                               | 3  | 25  | 5  | 5  | 0 | 8  | 1 | 3  | 1  | 2  | 0 | 2  | 0 |
| n (15,10)                   | 17                               | 7  | 22  | 4  | 4  | 2 | 4  | 1 | 3  | 4  | 3  | 0 | 2  | 0 |

A felolvasások felezési időtartamainak további átlagos értékei a következők: 1/16-nál 7,79 s, 1/32-nél 3,89 s, 1/64-nél 1,99 s, 1/128-nál 0,97 s. A magánhangzók száma természetesen csökken a beszédminta időtartamának rövidítésével a nyelvspecifikus gyakorisági eloszlásnak megfelelően. A teljes időtartamú szövegben 609, a legrövidebb szövegben pedig 6 magánhangzó fordult elő. A statisztikai vizsgálatban a különböző szövegrészek magánhangzóinak előfordulását elemeztük (Spearman-korreláció alkalmazásával). Azt találtuk, hogy az adatok az 1/8-os felezésig nem mutatnak különbséget a teljes anyaghoz képest ( $r = 0,9$ ). Az ennél rövidebb időtartamú felolvasásrészekben a korrelációs együttható értéke 0,9 alatti. Ennek alapján kijelenthetjük, hogy a magánhangzók előfordulásának statisztikai vizsgálata szerint a személyazonosításhoz minimálisan 15,58 s időtartamú beszédmintára van szükség. Az ebben az időtartamban előforduló szavak száma átlagosan 30, illetőleg mintegy 74 magánhangzó (4. ábra). Az ennél rövidebb időtartamú felolvasásrészekben a korrelációs együttható értéke 0,9 alatti. A magánhangzók eloszlása a beszédminta időtartamának függvényében arra utal, hogy a 15,58 s időtartamnál rövidebb minta – a vizsgált nyelvi jelek szempontjából már nem elegendő a beszélő személy azonosításához.

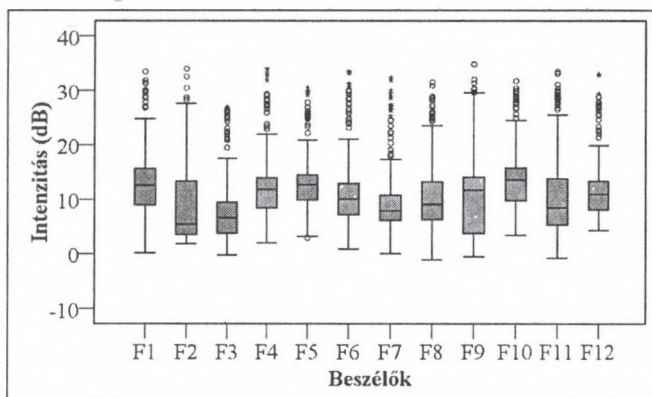


4. ábra

A beszédminta időtartamának és a magánhangzók számának összefüggése

A teljes időtartamú felolvasásokról készült **LTA-elemzések** értékei jellemzően eltérnek a **különböző beszélők** mintáiban. Az 5. ábra az LTA-értékek alapján kapott átlagot (medián) és a szóródást mutatja, szemléltetve a beszélők közötti eltéréseket. Mind az átlag, mind pedig a szóródási tartomány különbségeket igazolt a beszélők között. Ez azt jelenti, hogy az itt alkalmazott módszer jól használható a beszélő személy azonosításában mint egy újabb lehetőség a szokásos protokollban.

Az LTA-értékek statisztikai vizsgálatában (Kendall-féle korreláció) itt is az  $r = 0,9$ -es értéket állítottuk be a korábbiak alapján (ugyancsak 99%-os szignifikanciaszinten), amelyhez az egyes beszélők közötti anyagokat viszonyítottuk. A beszélők közötti összevetéshez a 12 személy teljes hanganyagáról készült LTA-elemzés függvényét viszonyítottuk az F1 jelű kísérleti személy beszédanyagához a korrelációs statisztikai elemzést alkalmazva. A kapott együtthatók értékeit a 4. táblázat mutatja.



5. ábra

Az LTA-értékek átlaga (medián) és szóródása az egyes beszélőknél

4. táblázat: A korrelációs együtthatók értékei az F1-es beszélő adataihoz viszonyítva

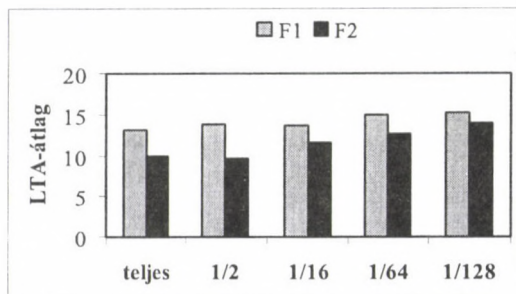
| Adatközlő  | F1   | F2   | F3   | F4   | F5   | F6   | F7   | F8   | F9   | F10  | F11  | F12  |
|------------|------|------|------|------|------|------|------|------|------|------|------|------|
| $r$ -érték | 1,00 | 0,55 | 0,57 | 0,68 | 0,67 | 0,50 | 0,50 | 0,64 | 0,65 | 0,53 | 0,66 | 0,50 |

A statisztikai eredmények igazolták, hogy az LTA-analízis az adott időtartamban diszkriminatív jellemzője a beszélő személynek. A beszélők közötti korrelációs együtthatók mindig jóval kisebbek voltak, mint az  $r = 0,9$ -es kijelölt érték.

Elemeztük a beszélők LTA-adatainak változását **ugyanazon személy** beszédmintái alapján a minta időtartamának függvényében. A csökkentett időtartamú beszédminták jellegzetes változásokat mutatnak egy beszélő ejtésén belül is; azonban ezek a változások kisebbek, mint amelyet a különböző beszélők között tapasztaltunk. Az LTA-értékek tehát – más adatokhoz hasonlóan – megerősítik, hogy a beszélők között nagyobbak az individuális különbségek a kiejtésben, mint ugyanazon beszélő beszédmintái között. Két adatközlő néhány szövegrészletének LTA-átlagáról készült grafikont mutat a 6. ábra. Mindkét



beszélőnél a felolvasott szöveg ugyanazon szakaszának mérési eredményeit ábrázoltuk. Ezekről az értékekről azonban csak statisztikai vizsgálattal állapítható meg, hogy mikortól tekinthetők különbözőeknek.



6. ábra

Két beszélő (F1 és F2) LTA-értékeinek átlaga  
különböző időtartamú beszédmintákban

Minden esetben a beszélő teljes hanganyagához hasonlítottuk a felezési időtartamokkal kapott saját rövidebb beszédmintáinak az értékeit. Az eredményeket az F1 és az F2 jelű beszélők esetén az 5. táblázatban összegeztük. Az áttekinthetőség érdekében nem az összes adatot közöljük, hanem minden időtartammintából egy-egy értéket.

5. táblázat: A különböző időtartamú beszédminták korrelációs együtthatója (a teljes időtartamú anyaghoz viszonyítva) két beszélőnél

| A beszédminta<br>időtartama | Korrelációs együttható ( $r$ -érték) |              |
|-----------------------------|--------------------------------------|--------------|
|                             | F1                                   | F2           |
| felezés                     | 0,933                                | 0,944        |
| negyedelés                  | 0,918                                | 0,907        |
| <b>1/8</b>                  | <b>0,913</b>                         | <b>0,916</b> |
| 1/16                        | 0,865                                | 0,843        |
| 1/32                        | 0,832                                | 0,817        |
| 1/64                        | 0,790                                | 0,800        |
| 1/128                       | 0,702                                | 0,789        |

A statisztikai vizsgálat (Kendall-korreláció) azt mutatja, hogy a különböző időtartamú szövegrészek korrelációs együtthatója 99%-os szignifikanciaszinten nem mutat különbséget az 1/8-os felezésig ( $r = 0,9$ ). (Említettük, hogy a 0,9-es együttható értéket az összes beszélő beszédmintájának elemzési tapasztalata alapján határoztuk meg.) A korrelációs együttható értéke csökken a beszédminta időtartamának rövidülésével. További felezés esetén az együtt-

ható értéke már kisebb, mint a 0,9-es érték (vö. 6. táblázat, a határnak tekintett érték félkövérel).

6. táblázat: A Kendall-féle korreláció együtthatója a beszédminta időtartamának függvényében (a teljes időtartamhoz viszonyítva)

| A beszédminta időtartama       | 1/2   | 1/4   | 1/8          | 1/16  | 1/32  | 1/64  | 1/128 |
|--------------------------------|-------|-------|--------------|-------|-------|-------|-------|
| Korrelációs együttható ( $r$ ) | 0,982 | 0,932 | <b>0,907</b> | 0,868 | 0,812 | 0,718 | 0,368 |

Valamennyi beszélőnél azt tapasztaltuk, hogy a teljes felolvasásuk 1/16-ára rövidített része az  $r = 0,9$ -es érték alá megy. Ez megerősíti tehát azt a feltételezést, hogy az 1/8-nál rövidebb beszédminta nem lesz alkalmas a beszélő személy azonosításának jó hatásfokú elvégzésére. Ez az 1/8-os csökkentés, az összes beszélő átlagát tekintve a 15,58 s időtartamú beszédminta, és tartamát tekintve megegyezik azzal a minimálisan használható időtartamú mintával, amelyet a beszédhangok statisztikai vizsgálatával kaptunk. Az elemzett szövegünk alapján ez nagyjából a következő nyelvi anyagot jelenti: *vitaminforrás a primőr fejes saláta, amely a szakhatóságok elmúlt éveiben végzett vizsgálata alapján a legtöbb lemoshatatlan mérget tartalmazza. A kismamák sem sejtik, hogy a bébiételek húsz százalékában találtak a növényvédő szerekből visszamaradt vegyi anyagokat.*

### Következtetések

Az elvégzett elemzések alapján igazoltuk, hogy minimálisan milyen időtartamú beszédminta szükséges ahhoz, hogy a beszélő személy azonosítása releváns eredményeket hozzon. A gyakorlatban azonban az esetek többségében az összehasonlítás alapjául szolgáló hangfelvételek nem érik el az általunk minimálisnak tartott, mintegy 16 s-nyi beszédidőtartamot, illetőleg nem tartalmazzák a szükségesnek ítélt kb. 30 elemezhető szót. A rövidebb (és esetlegesen rossz jel/zaj viszonytal rendelkezző) hanganyagok esetében a személyazonosítási vizsgálat elvégzése jóval bonyolultabb feladat, ami némely esetben kevésbé biztos eredményhez vezet.

### Irodalom

- Gósy Mária 2008. Magyar spontánbeszéd-adatbázis – BEA. *Beszédkutató 2008*. 194–207.
- Hollien, Harry 1977. Speaker identification by long-term spectra under normal and distorted speech conditions. *Journal of the Acoustical Society of America* 62. 975–980.
- Hollien, Harry 2002. *Forensic voice identification*. Academic Press, San Diego.
- McDougall, Kirsty – Nolan, Francis 2007. Discrimination of speakers using the formant dynamics of /u:/ in British English. In Trouvain, Jürgen – Barry, William J. (eds.): *Proceedings of the 16<sup>th</sup> International Congress of Phonetic Sciences*. Saarbrücken 6–10 August 2007. Pirrot GmbH., Dudweiler, 1825–1828.

- Nikléczy Péter 1996. Beszélő személy azonosítása szűk frekvenciás szavak alapján. *Beszédkutató* '96. 20–31.
- Nolan, Francis 1991. Forensic phonetics. *Journal of Linguistics* 27. 483–493.
- Nolan, Francis – Hollien, Harry – Braun, Angelika – Broeders, Peter 1995. Forensic phonetics. In Elenius, Kjell – Branderud, Peter (eds.): *Proceedings of the 13th International Congress of Phonetic Sciences*. 3. University of Stockholm, Stockholm, 130–142.
- Rose, Philip 2002. *Forensic speaker identification*. Taylor and Francis, London.



## REJTETT MARKOV-MODELL ALKALMAZÁSA MAGYAR NYELVŰ GÉPI SZÖVEGFELOLVASÁSHOZ

Tóth Bálint – Németh Géza

### Bevezetés

Napjainkban már számos automatikus szövegfelolvasási módszer létezik: a beszédkeltés mechanizmusát modellező formánsszintézistől kezdve a diádós és triádós hullámforma-összefüzéses szintézisen át az elemkiválasztó (korpusz) szintézisig. A beszédsszintetizátor által kiadott hangot érthetőség és természetesség szempontjából szokták minősíteni, magát a megoldást pedig műszaki paraméterekkel jellemzik. A formánsalapú szintetizátorok érthetősége jó, viszont hangzása nem természetes. A hullámforma-összefüzéses technológia kissé jobb eredményt nyújt, a legjobban érthető és a legtermészetesebbnek ítélt beszédet a korpuszalapú elemkiválasztásos módszerrel érték el. A műszaki jellemzők szoros kapcsolatban vannak a generált hang minőségével. Például, a legjobb minőséget nyújtó korpuszalapú szövegfelolvasó rendszerek adatbázisának a mérete igen nagy (gigabyte-os nagyságrendbe esik), az elemkiválasztás is sok számítási kapacitást igényel. A beszélő hangját az adatbázis meghatározza, azon változtatni nem lehet, viszont az adott hangon a mesterségesen előállított hangszínezet egészen természetes. A formánsszintetizátorok kisméretűek (2–10 kilobyte) és gyorsak, tehát olyan helyen is használhatók, ahol kis memóriakapacitás áll rendelkezésre. A technológia megengedi, hogy többféle hangon is megszólalhatnak. A hullámforma-összefüzéses eljárás ma a leggyakrabban alkalmazott módszer. Memóriaigénye viszonylag kicsi (2–20 megabyte), ezért sokféle gyakorlati alkalmazásban használják. Az ilyen rendszerek hangszínezete már közelít az emberihez, bizonyos korlátok között változtatható is. A jelen tanulmány olyan technológiát ismertet, amellyel növelni lehet a generált beszéd természetességét, és érthetősége is magas. Ilyen szempontból a tanulmányban bemutatott magyar nyelvű rejtett Markov-modellekkel (Hidden Markov Model, HMM) megvalósított beszédsszintetizátor a hullámforma-összefüzéses eljárás és a korpuszalapú szintézis közötti helyet foglalhatja el. A beszédépítési eljárása lényegesen különbözik az előbbi technológiáktól, mivel nem közvetlenül a hullámformával dolgozik, hanem a hullámformából spektrális és prozódiai jellemzők sokaságának kinyerése után (tanító fázis) ezekből válogatva alakítja ki a szintézishez szükséges adatsorozatot. A válogatást rejtett Markov-modellek végzik. A HMM-alapú beszédsszintézis számos előnye miatt lehet a jövő sikeres mesterséges beszédkeltési technológiája, mivel kisméretű (1,5–2 mega-

byte) beszédadatbázisból képes jó minőségű, érthető beszédet előállítani, amely hordozza a beszélő hangszínezeti tulajdonságait is (Yoshimura 2002). Magyar nyelvre eddig még nem alkalmazták az eljárást, az első, kísérleti eredményekről számolunk be.

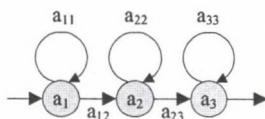
### A rejtett Markov-modell

A Markov-láncokat gyakran használják fizikai folyamatok modellezésére, ahol különböző megfigyelések alapján kell a folyamatot szimulálni. Ha egy adott megfigyelés egyértelműen azonosítja, hogy a folyamat milyen állapotban van, akkor a használt modellt Markov-láncnak nevezzük. Azonban számos olyan folyamat létezik, mint a beszéd is, mely esetben az állapotok jól definiálhatók, de rájuk a megfigyelések alapján mégsem tudunk egyértelműen következtetni. Ezeket modellezzük úgynevezett rejtett Markov-modellekkel. A beszédtechnológiában használt rejtett Markov-modell alkalmazásakor a beszédre jellemző, abból kinyert paramétereket kell eltárolni, mely jelentősen hatékonyabb, mint a hangmintaalapú rendszerek esetén a minták tárolása, hiszen a paraméterek jóval kevesebb helyet foglalnak el, mint az eredeti hullámformák. A paraméterek (például spektrális jellemzők) kinyeréséhez úgynevezett akusztikus modelleket alkalmaznak. Régebben fonémánkénti (ún. monofón) akusztikus modellt alkalmaztak, legújabban már a hangkörnyezetet is figyelembe vevő akusztikus modellek (pl. hanghármások, ún. trifónok) a leggyakoribbak (Mihajlik et al. 2006).

Napjainkban a beszédtechnológia területén a rejtett Markov-modelleket nagyrészt a beszédfelismerésben használják. A modell működését egy egyszerű példán keresztül mutatjuk be. A szavakat úgy tekintjük, hogy azok beszédhangok sorozataként állnak elő. Minden beszédhangra három állapotot feltételezünk: a hang eleje, közepe, vége. Az egyes állapotok között és az egyes állapotokból saját magukra mutató, úgynevezett élek határozzák meg, hogy az adott állapotból mely következő állapotokba lehet lépni (1. ábra). Az ábrán az  $a_1, a_2, a_3$  jelöli a beszédhang három belső állapotát, az  $a_{12}, a_{23}$  élek a három belső állapot közötti átmeneti valószínűségeket, végül az  $a_{11}, a_{22}, a_{33}$  pedig azt jelzi, hogy milyen valószínűséggel maradunk az adott belső állapotban. A modell betanítása során az élekhez valószínűségek rendelhetők, melyek a helyben maradás ( $a_{11}, a_{22}, a_{33}$ ), illetve továbblépés ( $a_{12}, a_{23}$ ) valószínűségét határozzák meg.

Az egyes állapotok tartalmazzák az akusztikus modellek készítése során becsült sokdimenziós Gauss-eloszlások paramétereit. Általában egy adott környezetben lévő beszédhang többször előfordul a tanító adatbázisban, a tanítás során pedig az ehhez tartozó spektrális paraméterhalmazt próbáljuk becsülni Gauss-eloszlással. A mintaillesztő eljárás ezen akusztikus modellekhez illeszti a bejövő paramétereket, hogy eldöntse, megegyezik-e az a felismerendő szóval. A témakörrel részletesen olvashatunk például Lawrence Rabiner (1989) klasszikus cikkében.





1. ábra

Három állapotú rejtett Markov-modell sematikus ábrája  
egy beszédhang három belső állapotának leírására

A rejtett Markov-modell alkalmazása a beszédszintézis területén az elmúlt évtizedben merült fel. Az erre kidolgozott eljárás két lényegi ponton tér el a beszédfelismerésre kidolgozott megoldástól. A legjelentősebb különbség az, hogy a két eljárás esetében a bemeneti és a kimeneti paraméterek felcserélődnek, tehát a végső lépésnél a mintaillesztés helyett (amikor felismerésnél a beszédhangokat jellemző paramétereket hasonlítjuk össze az akusztikus modell paramétereivel, tehát a hangok felismerésére adunk javaslatot) mintaválogatást hajtunk végre (azt javasoljuk, hogy melyik a legjobban jellemző mintasorozat a felolvasandó beszédhangra), majd a kiválasztott jellemző paraméterhalmazból a modell egy beszédkódoló eljárással beszédhangot állít elő, és így jön létre a szintetizált beszédhullám. A második fontos különbség, hogy a prozódia jellemző komponenseit (például hangmagasság, hangidőtartam) is modellezni kell a beszédszintézis esetében, és ezeket a feladatokat szintén végezhetik rejtett Markov-modellek.

A jelen kutatásban a spektrális és az alaphfrekvencia-paramétereket, valamint az állapot-időtartamokat is HMM-ekkel modellezzük.

### A HMM-alapú beszédszintézis

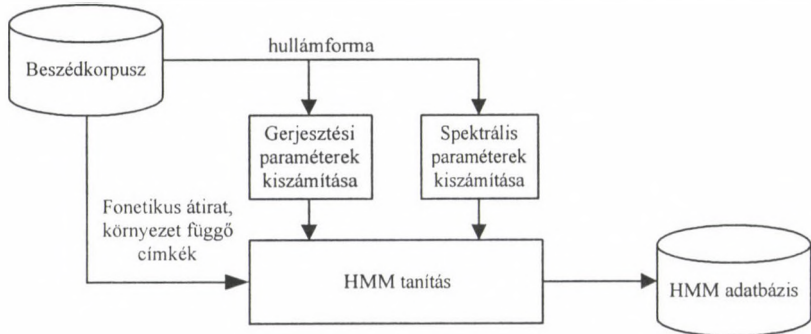
A HMM-alapú beszédszintetizátor elkészítése két részre bontható: a tanulási és a szintetizálási fázisra. A tanulási fázisban készítjük fel a szintetizátort a későbbi működésre. A tanulás során (2. ábra) tanítjuk be a rejtett Markov-modelleket egy nagy, gondosan megtervezett és felcímkézett beszédatadabázis segítségével. A tanítási folyamat végére egy kisméretű HMM modelladatbázis áll elő, amelyben a betanított beszédkorpuszra jellemző HMM-paraméterek találhatók. Ezekből válogatja majd ki a szintetizátor a szintetikus beszéd generálásához szükséges paramétereket.

A működési fázisban már csak a tanítás eredményét használjuk, tehát ez a rész a tényleges beszédszintetizátor. A bemeneti szöveg alapján meghatározzuk, hogy milyen hangsorozatot kell generálni, és a HMM modelladatbázisban tárolt paraméterekből kiválogatjuk azt a paramétersorozatot, amelyik legjobban reprezentálja az előállítani kívánt hangsorozatot. Ezekből állítjuk vissza a spektrális jellemzőket, az időtartamokat, a szüneteket és az alaphfrekvenciát, majd ezek alapján a beszédkódoló eljárással a szintetizált hullámformát.



### A HMM-ek tanítása

A tanításhoz több órányi beszédet tartalmazó, nagyméretű beszédkorpuszt kell használni. A tanítás típusától függően ez lehet egy vagy több beszélőtől.



2. ábra

A HMM-alapú szövegfelolvasó tanítása

A korpusz tartalma a következő: a hullámforma digitalizált változata, a felolvasott szöveg fonetikai átírata és a hang- és szóhatárok pontos pozíciója. A feldolgozás egysége a mondat. A hullámformából 25 ms-os ablaklással mintákat veszünk, és azok paramétereit tároljuk, összesítjük, optimalizáljuk. Minden ablakolt részből kinyert valamennyi paraméterhez tartozik legalább egy HMM. A HMM-alapú beszédszintetizátort egyszer kell tanítani, mely eredményeként egy kisméretű paraméteradatbázis áll elő (HMM modelladatbázis), amelyet a beszéd szintetizálásakor fog használni a rendszer.

A rejtett Markov-modellek tanításához a beszédkorpuszból származtatott paraméterek sokaságára van szükség. Ezek a következők: a hullámforma spektrális tartalmára utaló, úgynevezett MFCC (Mel Frequency Cepstrum Coefficients) adatok, ezek első és második deriváltjai, továbbá az alapfrekvencia ( $F_0$ ), valamint annak első és második deriváltjai.

Ezen túl még szükség van a beszédkorpusz fonetikai átíratából képzett hangkörnyezetfüggő címkékre. A környezetfüggő címkék írják le egy adott beszédhang környezetét (pl. előtte és utána következő beszédhangok, hangsúlyos szótagok jelölése, szótagszámok stb.). A környezetfüggő címkékről bővebben alább szólunk. A fenti paramétereket jelfeldolgozási és matematikai szoftverekkel automatikusan lehet előállítani. A tanító beszédadatbázis címkézési pontosságát célszerű kézi módszerekkel is ellenőrizni.

Amennyiben minden szükséges adat a rendelkezésünkre áll, elkezdődhet a HMM-ek tanítása. A tanítás célja, hogy az egyes hangokhoz és környezetükhöz rendelt paraméterek segítségével előállítsunk egy azokat minél pontosabban becsülő függvényt (esetünkben Gauss-eloszlást használtunk). Ezen adato-

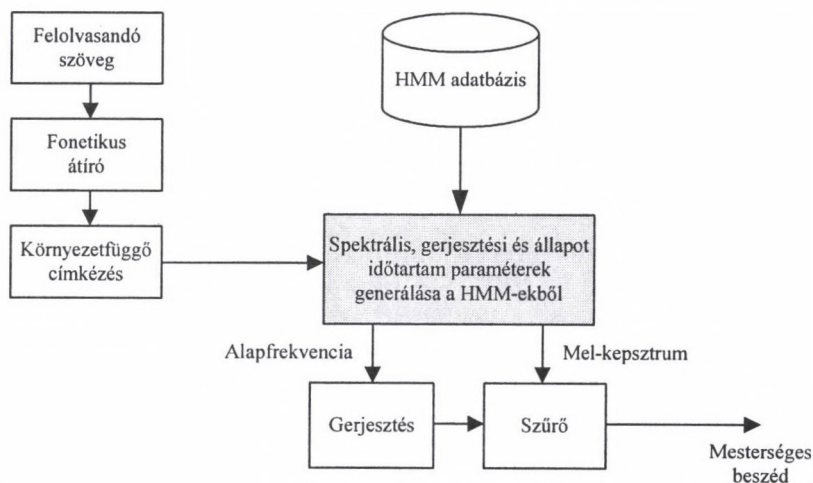
kat tároljuk el a HMM modelladatbázisban, melynek mérete (kb. 1-2 megabyte) nagyságrendekkel kisebb az eredeti beszédkorpusz méreténél (1-2 gigabyte).

A betanításhoz kétfajta módszert lehet használni. Betaníthatjuk a HMM-eket egy adott beszélőtől származó 2–4 órás beszédadatbázissal (ilyenkor erre a személyre jellemző hangszínezettel és stílussal beszél a szövegfelolvasó), illetve betaníthatjuk több beszélőtől gyűjtött adatbázisokkal (beszélőnként 1-1,5 óra hanganyag, minimum 3-4 különböző beszélő hangja). A tanítás után ekkor még szükséges a beszédhangot egy konkrét bemondónak a hangjára adaptálni. A személyre szabást egy 5–8 perces, tetszőleges személytől felvett beszédadatbázissal végezhetjük (Masuko 1997, Tamura 2001). Ezután ennek a személynek a hangjához hasonló hangszínnel és stílussal fog megszólalni a rendszer. Ezen túl még számos módszer létezik a beszédhang jellemzőinek a megváltoztatására (Yoshimura 1997, Tachibana 2005).

### Beszéd előállítása a betanított HMM-ek segítségével

A beszéd előállítása során első lépésként elkészítjük a szöveg fonetikai átiratát környezetfüggő címkékkel (lásd alább), majd kinyerjük a várható hang-időtartamokat az állapot-időtartam valószínűség sűrűségfüggvényekből, ezután pedig a legvalószínűbb spektrális és gerjesztési paramétereket nyerjük ki a HMM modelladatbázisból. Ezen paraméterek alapján állítjuk elő a mesterséges beszédet beszédkódoló eljárás segítségével (Imai 1983, Maia 2007).

A HMM alapú beszéd szintetizátor felépítését a 3. ábra mutatja.



3. ábra

A HMM-alapú szövegfelolvasó beszéd-előállítási mechanizmusa

### Magyar nyelvű adaptáció

A kísérleteket a HTS keretrendszer segítségével végeztük el (Zen et al. 2007), amely egy szabadon felhasználható, programozható HMM-rendszer. A magyar nyelvű változat elkészítéséhez a következő fő elemekre és adatokra volt szükség: nagyméretű beszédadatbázis címkékkel ellátva, a beszédadatbázis mondatainak fonetikai átíráta, egy hangkörnyezetfüggő címkéző eszköz és a magyar nyelvre jellemző döntési fák elkészítéséhez szükséges kérdések összeállítása. A következő pontokban áttekintjük a magyar változat létrehozásának ezen lépéseit.

A magyar nyelvű adaptáció során a következő beszédhangokat definiáltuk (a hangokat itt helyesírási betűjelükkel adjuk meg): magánhangzók: *a, á, o, ó, u, ú, ü, ű, i, é, ő, ö*; rövid és hosszú mássalhangzók: *b, d, gy, g, p, t, ty, k, m, n, ny, j, h, v, f, sz, z, c, dz, zs, s, cs, dzs, l, r*. A hanghosszúságot a mássalhangzóknál az időparaméterrel különböztetjük meg.

### A beszédkorpusz előkészítése

A tanításhoz 600 időjárás-jelentést tartalmazó mondatot használtunk, melyeket professzionális női bemondó olvasott fel (2 óra hanganyag). A mondatokat 16 000 Hz-es mintavételezéssel, 16 bites felbontással tároltuk. A mondatok fonetikus átíratát elkészítettük, és a hang- és szóhatárokat bejelöltük először automatikus módszerekkel (Mihajlik et al. 2003), majd ezeken fél-automatikus módszerrel finomítottunk (Olaszy–Bartalis 2008).

### Környezetfüggő címkézés

Annak érdekében, hogy a HMM-ek a legmegfelelőbb elemeket válasszák majd ki a beszéd előállítás során, számos fonetikai jellemzőt adunk meg paraméterként. A jellemzőket minden egyes beszédhangra kiszámoljuk. A legfontosabb jellemzőket az 1. táblázat foglalja össze.

A szótagokat a szótagmagok alapján keressük, számoljuk és jelöljük, tehát nem a nyelvi szótagolási szabályokat vesszük figyelembe. A szótaghatárokat saját programmal állapítjuk meg. A programban a *A magyar helyesírás szabályai* (Akadémiai Kiadó 1985) című kiadvány *A szótagolás szerinti elválasztás* szabályait algoritmizáltuk.

A táblázatban megadott jellemzőket nevezzük címkéknek. Tehát a beszédadatbázisban található összes mondat összes hangjához minden paramétert ki kell számolni, amely az 1. táblázatban található. Egy hanghoz összesen 38 környezetfüggő címkét rendelünk, így például egy 100 hangból álló mondat-hoz 3800 címkére van szükségünk. A címkézést automatikusan végezzük, mely néhány esetben (pl. hangsúlyos szótagok meghatározása) hibás lehet. A Profivox címkézője 70%-os pontosságú, nyelvmodell-alapú kísérletek sem adnak jobb eredményt (Tamm–Olaszy 2005) megfelelő nyelvi modell hiányában. Ez azonban nem okoz jelentős problémát, hiszen a beszéd előállításakor is ugyanazt az algoritmust használjuk, így a HMM hibás címkézés esetén is következetesen fogja az adott hangoknak megfelelő paramétereket kiválasztani.



1. táblázat: A környezetfüggő címkézéshez használt jellemzők

|                                                  |                                                                                                                                                                                                                                                                                                                                                                          |
|--------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>Hangok</b>                                    | – Az aktuális hang, valamint a megelőző és követő két-két hang (kvinfón). A szüneteket is hangként jelöljük.                                                                                                                                                                                                                                                             |
| <b>Szótag</b>                                    | – A szótaghangsúlyok jelölése (hangsúlyos/hangsúlytalan) az aktuális, az előző, és a követő szótagban. Erre a Profivox-rendszer hangsúly-meghatározó algoritmusát használtuk (Olaszy et al. 2000).<br>– A beszédhangok száma az aktuális/előző/következő szótagban.<br>– A szótagok száma az előző/következő hangsúlyos szótagtól/szótagig.<br>– A szótag magánhangzója. |
| <b>Szó</b>                                       | – A szótagok száma az aktuális/előző/következő szóban.<br>– Az aktuális szó pozíciója a mondatrészben (előlről és hátulról is számítva). Mondatrésznek tekintünk két vessző, pontosvessző, gondolatjel közötti szövegrészt a mondatban.                                                                                                                                  |
| <b>Mondatrész (két írás-jel közötti szakasz)</b> | – A szótagok és szavak száma az aktuális/előző/következő mondatrészben.<br>– Az aktuális mondatrész pozíciója a mondatban (előlről és hátulról is számítva).                                                                                                                                                                                                             |
| <b>Mondat</b>                                    | – A szótagok száma az adott mondatban.<br>– A szavak száma az adott mondatban.<br>– A mondatrészek száma az adott mondatban.                                                                                                                                                                                                                                             |

### Döntési fák

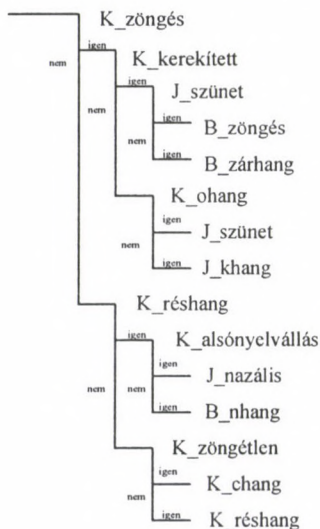
Az imént láthattuk, hogy számos környezetfüggő tulajdonság létezik, melyek összes lehetséges kombinációja óriási szám. Ha csupán a kvinfónok lehetséges változatait számoljuk meg, az is több mint 160 millió, de ezt a számot a többi környezetfüggő tulajdonság még exponenciálisan növeli. Ezért lehetetlen egy olyan, adott nyelvre jellemző, a tanításhoz szükséges beszédadatbázis-szöveget előállítani, melyben minden lehetséges kombináció szerepel. E probléma leküzdése érdekében be kellett vezetni a döntésifa-alapú klaszterezést (Odell 1995, Shinoda–Watanabe 2000), hogy a beszédszintézis során a HMM modelladatbázisban az adott környezetfüggő címkékhez legjobban hasonlító elemeket válassza ki a rendszer. Mivel a különböző tulajdonságok hatnak mind a spektrális, mind az alaphfrekvencia-paraméterekre és az állapot-időtartamokra is, ezért ezeket külön-külön csoportokra bontottuk, így összesen ez a háromféle klaszterezés van jelen a paraméterek kiválasztásánál. A 2. táblázat mutatja, hogy milyen, a magyar nyelvre jellemző tulajdonságokat (Gósy 2004) használtunk fel a döntési fák építésein.

A 4. ábrán egy példát láthatunk a spektrális paraméterekre vonatkozó döntési fára. A döntési fákot a program automatikusan készíti el. A döntési fa mindig a legelőnyösebb csoportra bontást próbálja megvalósítani. A K\_ előtag azt je-

lenti, hogy a szabály az épp középső hangra igaz, a J\_ előtag a középső hangot követőre (jobbra lévőre), a B\_ előtag pedig a középső hangot megelőzőre (tőle balra lévőre) vonatkozik. A bemutatott példában azt láthatjuk, hogy a középső hang zöngés-zöngétlen csoportokra bontása volt a legelőnyösebb (a K\_zöngés tulajdonság került a döntési fa legfelsőbb szintjére).

2. táblázat: A döntési fák építéséhez használt jellemzők

|                     |                                                                                                                                                                                                                                                                                                                                                                  |
|---------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>Beszédhangok</b> | – Magánhangzó/mássalhangzó.<br>– Zöngés/zöngétlen.<br>– Rövid/hosszú.<br>– A képzés helye a magánhangzóknál (hátul, középen, elől).<br>– Nyelvállás a magánhangzókra (felső, középső, alsó).<br>– Ajakállás a magánhangzókra (kerekített, nem kerekített).<br>– A képzés módja mássalhangzóknál (zárhang/részhang/zár-részhang/pergőhang/nazális, közelítőhang). |
| <b>Szótag</b>       | – Hangsúlyos/hangsúlytalan.                                                                                                                                                                                                                                                                                                                                      |
| <b>Szó</b>          | – Az adott szótagra vonatkozó számszerű adatok (lásd 1. táblázat).                                                                                                                                                                                                                                                                                               |
| <b>Mondatrész</b>   | – Az adott mondatrészre vonatkozó számszerű adatok (lásd 1. táblázat).                                                                                                                                                                                                                                                                                           |
| <b>Mondat</b>       | – Az adott mondatra vonatkozó számszerű adatok (lásd 1. táblázat).                                                                                                                                                                                                                                                                                               |



4. ábra

A spektrális paraméterekhez tartozó egy lehetséges döntési fa

### Meghallgatásos teszt

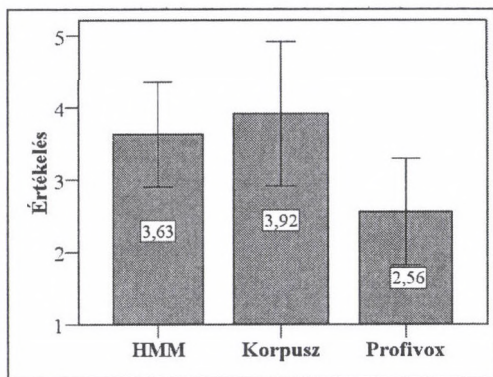
Annak érdekében, hogy objektíven tudjuk értékelni a magyar nyelvű HMM-alapú beszédszintézis minőségét egy MOS (Mean Opinion Score) meghallgatásos tesztet készítettünk el. A tesztben három beszédszintetizáló rendszer vett részt, egy triádokkal működő, egy korpuszos és a HMM-alapú szövegfelolvasó.

A teszt elején minden rendszertől 3-3 mondatot játszottunk le véletlenszerűen, amelyeket a tesztalanyoknak még nem kellett értékelni. Ez azt a célt szolgálta, hogy az alanyok hozzászokjanak a mesterséges hangokhoz, és hallják előre, hogy nagyjából milyen minőségre számíthatnak.

Ezután minden rendszer mintáiból 29 mondatot játszottunk le, minden tesztalany esetén más-más sorrendben, így zárva ki az esetleges „memóriahatásokat” (Santen 1993). A tesztmondatok tartalma időjárás-jelentés volt. A triádalapú rendszer kötetlen témakör szintézisére készült. A HMM-rendszer időjárás-jelentés tartalmú mondatokkal volt tanítva, illetve a korpuszos rendszer adatbázisa is időjárás-jelentéseket tartalmazott. Minden rendszerrel ugyanazt a 29 mondatot generáltuk, de egyik rendszer esetén sem szerepeltek ezek a mondatok az adatbázisban. A tesztalanyok a mondatokat egytől ötig értékelhették (egy volt a legrosszabb, öt a legjobb).

A meghallgatásos tesztet 12 tesztalany végezte el.

Az eredményt a 5. ábra mutatja.



5. ábra

A MOS meghallgatásos teszt eredménye az átlag és szórás értékekkel

A tudomány mai állása szerint a korpusz elvű beszédszintézis-módszer képviseli a legjobb hangminőséget kötött témakörre, ez tükröződik az eredményben is. A HMM-alapú rendszer megközelítette a korpuszost, tehát a beszédminősége szintén nagyon jónak mondható. A különbség a két rendszer között a műszaki paraméterekben viszont számottevő. A korpuszos rendszer



adatbázisa közel 11 órányi hanganyagot tartalmaz, míg a HMM-ek tanításához elegendő volt 1,5 órányi hanganyag, és tanítás után a HMM-szövegfelolvasó esetén az adatbázis mérete 2 megabyte alatt marad (szemben a korpuszos rendszer több mint egy gigabyte-os adatbázisával). A triádalapú rendszer általános témakörök lefedésére készült, semmilyen témakör-specifikus információ nem került bele. Ez is magyarázhatja az alacsonyabb értékelést. Az eredmények abszolút értéke kevésbé mérvadó, inkább az egymáshoz viszonyított arányok hordoznak érdemi információt.

### Jövőbeli tervek

A jelen tanulmány a magyar nyelvű, HMM-alapú mesterséges beszédkeltés első kísérleti változatát ismertette. A jövőben számos továbbfejlesztési irányt tűztünk ki célul, melyek közül első lépésként az adaptív tanításhoz szeretnénk további beszédkorpuszokat rögzíteni, így érve el természetesebb hangzást, továbbá ezáltal lehetőségünk nyílik kis (5-8 perces) adatbázisok segítségével új beszédhangokat és érzelmeket betanítani (Krstulovic et al. 2007) a rendszerrel.

A kis adatbázisméret előnyei és a jó minőségű beszédhang miatt szeretnénk a rendszert mobil eszközökön is megvalósítani. Ennek érdekében, amennyiben a jelenlegi rendszer nem képes valós idejű működésre a limitált platformon, optimalizálni fogjuk az eljárást mobil környezetre, és mérésekkel meghatározzunk azt a minimális hardverigényt, amely szükséges a megfelelő működéshez.

### Összefoglalás

A tanulmányban bemutattuk a rejtett Markov-modell alapú szintézis működésének az elvét, a magyar változat létrehozásának a lépéseit és az első magyar HMM-alapú beszédkeltéssel kapcsolatos kísérlet végén végzett meghallgatásos teszt eredményeit. A HMM-alapú szövegfelolvasó rendszerek igazi előnye, hogy kis adatbázis méretek mellett képesek jó minőségű beszédhangot előállítani. Az eljárás alkalmas arra is, hogy az eredeti hang karakterét megváltoztassuk más beszélő hangjára, illetve érzelmes beszédet is előállítsunk. Célunk, hogy ipari alkalmazásokban is használható magyar nyelven beszélő szövegfelolvasó rendszerré fejlesszük tovább a jelenlegi kísérleti rendszert.

Néhány HMM alapú szintézissel készített mondat meghallgatható a következő Internet címen: <http://speechlab.tmit.bme.hu/hmm/>

### Irodalom

- A magyar helyesírás szabályai* 1985. 11. kiadás. Akadémiai Kiadó, Budapest.  
<http://mek.oszk.hu/01500/01547/index.phtml> (A letöltés ideje: 2008. július 3.)  
Gósy Mária 2004. *Fonetika, a beszéd tudománya*. Osiris Kiadó, Budapest.

- Imai, Satoshi 1983. Cepstral analysis synthesis on the mel frequency scale. *Proceedings of ICASSP 1983*. 93–96.
- Krstulovic, Sacha – Hunecke, Anna – Schröder, Marc 2007. An HMM-based speech synthesis system applied to German and its adaptation to a limited set of expressive football announcements. In: *Proceedings of Interspeech 2007*. 1897–1900.
- Maia, Rannieri – Toda, Tomoki – Zen, Heiga – Nankaku, Yoshihiko – Tokuda, Keiichi 2007. A trainable excitation model for HMM-based speech synthesis. In: *Proceedings of Interspeech 2007*. 1909–1912.
- Masuko, Takashi – Tokuda, Keiichi – Kobayashi, Takao – Imai, Satoshi 1997. Voice characteristics conversion for HMM-based speech synthesis system. In: *Proceedings of ICASSP 1997*. 1611–1614.
- Mihajlik, Péter – Révész, Tibor – Tatai, Péter 2003. Phonetic transcription in automatic speech recognition. *Acta Linguistica Hungarica* 49/3–4. 407–425.
- Mihajlik Péter – Fegyő Tibor – Tatai Péter 2006. Új eljárás a gépi beszédfelismerés környezetfüggő beszédhangmodelljeinek kialakítására. *Beszéd kutatás 2006*. 218–230.
- Odell, Julian James 1995. *The use of context in large vocabulary speech recognition*. PhD thesis. Cambridge University, Cambridge.
- Olaszy Gábor – Bartalis Mátyás 2008. Jelfeldolgozási algoritmusok kombinációja a gépi hanghatárjelölés javítására. *Beszéd kutatás 2008*. 208–220.
- Olaszy Gábor – Németh Géza – Olaszi Péter – Kiss Géza – Zainkó Csaba – Gordos Géza 2000. Profivox – a Hungarian TTS system for telecommunications applications. *International Journal of Speech Technology* 3/3–4. 201–215.
- Rabiner, Lawrence R. 1989. A tutorial on hidden Markov models and selected applications in speech recognition. In: *Proceedings of the IEEE*. 257–286.
- van Santen, Jan P. H. 1993. Perceptual experiments for diagnostic testing of text-to-speech systems. *Computer Speech and Language* 7. 49–100.
- Shinoda, Koichi – Watanabe, Takao 2000. MDL-based context-dependent subword modeling for speech recognition. *Journal of Acoustical Society of Japan (E)* 21/2. 79–86.
- Tachibana, Makato – Yamagishi, Junichi – Takashi, Masuko – Kobayashi, Takao 2005. Speech synthesis with various emotional expressions and speaking styles by style interpolation and morphing. *IEICE Transactions Information and Systems* 88/11. 2484–2491.
- Tamm Anne – Olaszy Gábor 2005. Kísérlet automatizált szövegelemzési módszerek kialakítására a szóhangsúlyok meghatározásához. In Alexin Zoltán – Csendes Dóra (szerk.): *III. Magyar Számítógépes Nyelvészeti Konferencia*. Szegedi Tudományegyetem, Szeged, 383–393.
- Tamura, Masatsune – Masuko, Takashi – Tokuda, Keiichi – Kobayashi, Takao 2001. Adaptation of pitch and spectrum for HMM-based speech synthesis using MLLR. In: *Proceedings of ICASSP 2001*. 805–808.
- Yoshimura, Takayoshi – Tokuda, Keiichi – Masuko, Takashi – Kobayashi, Takao – Kitamura, Tadashi 1997. Speaker interpolation in HMM-based speech synthesis system. In: *Proceedings of Eurospeech 1997*. 2523–2526.
- Yoshimura, Takayoshi 2002. *Simultaneous modeling of phonetic and prosodic parameters, and characteristic conversion for HMM-based text-to-speech systems*. PhD thesis. Nagoya Institute of Technology, Nagoya.

Zen, Heiga – Nose, Takashi – Yamagishi, Junichi – Sako, Shinji – Masuko, Takashi – Black, Alan W. – Tokuda, Keiichi 2007. The HMM-based speech synthesis system version 2.0. In: *Proceedings of ISCA SSW6*. Bonn. (CD)

Ezúton szeretnénk kiemelten köszönetet mondani Olasz Gábornak szakmai tanácsaiért. Köszönjük a szubjektív kiértékelésben részt vevő tesztelőknek aktivitásukat. Külön köszönet illeti Bartalis Mátyást a web-es tesztfelület elkészítéséért és Mihajlik Pétert a magyar nyelvű beszédfelismerő eszközök használatához nyújtott segítségéért. A kutatást az NKTH a NAP projekt keretében (OMFB-00736/2005) részben támogatta.



## MAGYAR SPONTÁNBESZÉD-ADATBÁZIS – BEA

Gósy Mária

### Beszédatadatbázisok

Az adatbázis vagy más szóval korpusz a nyelvészetben az írott, illetőleg lejegyzett szövegek, valamint a rögzített beszédanyagok gyűjteménye. A korpusznyelvészet egyértelműen a technológia fejlődésének, a számítógépnek köszönhető (Leech 1992, McEnery–Wilson 1996, Váradi 2000). Az a cél ugyanis, hogy olyan mennyiségű adatot gyűjthessünk, amely kellőképpen reprezentálja egy közösség nyelvi jellemzőit, illetve nyelvhasználatát, és ezen az anyagon végezzék a különféle elemzéseket, csak a számítógépes technológia segítségével érhető el.

A beszédkorpusz a beszélt nyelv nagy mennyiségű rögzített felvételét jelenti, amely meghatározott céllal és metodológiával készült. A legtöbb beszédkorpusz ún. szövegfájlokat is tartalmaz, ezek a rögzített beszédanyag különböző szintű átírásának anyagai. A beszédkorpuszok két nagy típusra oszthatók. Az egyikbe az olvasott szövegek felvételei, a másikba a spontán beszédanyagok tartoznak. A felolvasásokat tartalmazó korpuszok lehetnek: *a)* könyvek részleteinek meghangosításai, *b)* rádióhírek, *c)* szölisták, *d)* számlisták, *e)* beszédrészletek (pl. szótagok). Az ismert spontánszöveg-korpuszok a következők: *a)* párbeszéd és társalgások; *b)* narratívák; *c)* térképmódszerrel (vö. Horváth 2004) rögzített szövegek; *d)* randevűfeladatok (ekkor két embernek különböző napi programleírást adnak, és ezek ismeretében kell egy olyan időpontot találniuk, amikor találkozniuk; *e)* élethelyzet modellálása (pl. jegyzendelés utazáshoz).

Mik az előnyei a beszéd kutatásban a korpuszok használatának? Mindenekelőtt megtakarítja a kutató(k)nak az időigényes felvételek készítését, hiszen ezt a nagy munkát a korpuszépítés során elvégezték. Hatalmas adathalmazt biztosít sokféle célú kutatáshoz. A nyelv valós használata alapján végezhetők a vizsgálatok, elemzések, az adatok különböző összevetésekre nyújtanak lehetőséget, továbbá egyéb tényezők is tekintetbe vehetők, például a nemek közötti esetleges eltérések, életkori sajátosságok, bizonyos szociolingvisztikai jellemzők. A beszédkorpuszoknak különös jelentőségük van a mesterséges beszéd-felismerésben és a beszéd szintézisben. Az előnyök mellett azonban adódhatnak hátrányok is. A spontánbeszéd-korpuszok felvételi minősége nem mindig tökéletes, némelykor túl soknak tűnik az információ, a beszélők nem kontrolláltak (a laboratóriumi beszéd bemondói rendszerint gyakorlott beszélők), továbbá nem minden nyelven léteznek beszédkorpuszok.

A korpusznyelvészet sok nyelven indult fejlődésnek az utóbbi másfél évtizedben, a legismertebbek közül említjük meg az alábbiakat: az amerikai TIMIT (1990), a Switchboard (Godfrey et al. 1992) és a Verbmobil (Geumann et al. 1997); részletesen lásd Bael et al. 2007. Beszélt nyelvi adatbázisok az angol, német, francia nyelveken kívül elérhetők már például arab, bolgár, kínai, japán, koreai, orosz, sőt vietnami nyelven is. Az egyes beszédkorpuszok nagysága, a fejlesztésük célja, feldolgozottságuk mértéke, illetve rögzítési módjuk változó.

Tájékoztató adatok néhány korpuszról: A *British National Corpus* (BNC) 100 millió szóból álló gyűjtemény, amely írott és beszélt szövegeket egyaránt tartalmaz. A *London–Lund* korpusz 50 dialógusból áll, és 170 000 szót tartalmaz. A *CallHome* nevű korpuszban 120 párbeszéd található, átlagosan 30 percesek, családi beszélgetések telefonon keresztül. Főként skót angol beszédet rögzít 62 személytől az a 18 órányi adatbázis (*HCRC Map Task*), amely a térképmódszert használja. Az ausztrál pizzarendeléses korpusz 3 óra 54 perc hosszúságú, egy év alatt 162 megrendelést rögzítettek (Hutchinson–Pereira 2001). A Stanford Egyetem (USA) *Switchboard* elnevezésű beszédkorpusza 2400 párbeszédet tartalmaz, amelyben 543 beszélő vett részt (számos amerikai dialektus reprezentált az anyagban, egy-egy beszélő több dialógusban is részt vett). Különbő kutatócsoportok hetenkénti értekezleteit rögzítették (*ICSI meetings*, Stanford), ami egy 72 órás beszédanyagot eredményezett 53 beszélővel. Az ATIS adatbázisban 36 beszélő repülőjegy-rendeléssel kapcsolatos beszélgetése található. A *Kiel Corpus of Spontaneous Speech*<sup>1</sup> olvasott mondatokat és szövegeket (53 személy), spontán beszédet (42 beszélő) és dialógusokat tartalmaz. Az egyik dialógustípusban a beszélgetőknek egy találkozást kell egyeztetniük (16 résztvevő, 240 perc), a másikban a majdani dialógus résztvevői külön-külön szobában egy-egy videofilmet láttak, amelyek hasonlóak, de nem teljesen azonosak voltak (15 percesek). A megtekintés után meg kellett beszélniük a látottakat (időtartama 80 perc, a beszélők száma 6). A BAS (*Bayerisches Archiv für Sprachsignale*<sup>2</sup>) német adatbázis több (al)korpuszból áll, különféle célok elérését kívánja lehetővé tenni az anyagaival. Például az olvasott korpuszban utcanevek felolvasása, a spontán beszéd gyűjteményben pedig a taxi diszpécser-szolgálatának beszélgetései találhatók. Tartalmaz narratívákat, gyermek- és kamasznyelvi, illetve nyelvjárási felvételeket is. Felolvasásokat (44 óra, a beszélő száma 561) és spontán telefonbeszélgetéseket (18 órányi anyag, a beszélők száma 344) tartalmaz a holland beszédkorpusz (*Corpus Gesproken Nederlands*, Oostdijk 2002). Igen nagy adatbázis a CSJ (*Corpus of Spontaneous Japanese*), amelyben 661 órányi beszédanyag található 1395 beszélőtől (olvasott és különböző spontán beszéd-

1. <http://www.ipds.uni-kiel.de/ipds/forsch.htm#KielCorpus>

2. <http://www.phonetik.uni-muenchen.de/forschung>



anyagok), a szavak száma 7,2 millió. Tanult, köznyelvi beszélőkkel rögzítettek a felvételeket, amelyeken már számos, különféle kutatást folytattak (Maekawa 2003).

A *RELATOR* elnevezésű beszédkorpusz európai projekt eredménye, beszélt és írott anyagokat tartalmaz, jelenleg dán, holland, angol, francia, német, görög, olasz, portugál és spanyol nyelven. A *22 language* elnevezésű beszédkorpusz (Oregoni Egyetem, USA) 50 191 fájlt tartalmaz, 2068 a beszélők száma, és közel 100 órányi beszédet rögzítettek benne. Ez a beszédanyag telefonbeszélgetésekből áll, egyenként 5-7 perc időtartamúak, a hívott félnek 29 kérdésre kell válaszolnia. Az anyag mintegy fele van helyesírásban átírva. A leírás szerint magyar anyag is található benne.<sup>3</sup> Végül megemlítjük az *UCLA Phonetic Data* (korábban: *Sounds of the World's Languages*) elnevezésű adatbázisát, amely több mint 100 nyelv beszédhangjainak kiejtését tartalmazza<sup>4</sup>. (A korpuszuk egy része kereskedelmi forgalomban kapható.)

A *Budapesti Szociolingvisztikai Interjú* (BUSZI) egyike azoknak a magyar nyelvi korpuszoknak, amelyek spontán beszédet (is) tartalmaznak (Kontra 1988; Váradi 2003). Megbízható adatokat szolgáltat a magyar nyelvnek a fővárosban beszélt változatairól. 1987 és 1989 között, 250 beszélővel magnetonon rögzített, egyenként 2-3 órás interjút tartalmaz. Számítógépes lejegyzése és kódolása is megtörtént<sup>5</sup>. A BABEL az első magyar beszédadatbázis, amely nemzetközi szabvány alapján készült (Vicsi–Vig 1998). Célja a magyar köznyelv reprezentálása hanganyaggal; hangkapcsolatokat, szavakat, számokat, öt mondatból álló mondatfüzereket tartalmaz, továbbá megtalálható benne 120 bekezdés fonetikai szinten címkézett és szegmentált anyaga. A hanganyag felvétele olvasott szöveg alapján készült (60 bemondóval). A mesterséges beszéd előállításához különféle beszédadatbázisok készültek (pl. diádós adatbázis, kötött szótáras rendszerekhez készített beszédelemtár, kevert felépítésű beszédadatbázisok); a hanganyag minden esetben egy-egy bemondóval rögzített felolvasás alapján készült (vö. Olaszy 1999). Az MTBA magyar telefonbeszéd-adatbázis vezetékes és mobiltelefonról rögzített ún. telefonos beszédkorpusz, amely a magyar beszédtechnológiai kutatások és fejlesztések támogatására készült. Az adatbázis 500 adatközlő hangfelvételeit tartalmazza; a beszélők olvasták a szövegeket (Vicsi et al. 2002). Vannak egyéb rögzített spontán magyar beszédgyűjtemények is (nyelvjárási archívumok, pl. a Hegedűs-archívum, vö. Horváth 2007), osztálytermi felvételek (Keszler 1983), az ELTE Fonetikai Tanszékének, illetve az MTA Nyelvtudományi Intézet Fonetikai Osztályának spontánbeszéd-felvételei.

3. [http://cslu.cse.ogi.edu/corpora/22lang/protocol\\_hungarian.html](http://cslu.cse.ogi.edu/corpora/22lang/protocol_hungarian.html)

4. <http://www.linguistics.ucla.edu/faciliti/sales/software.htm>

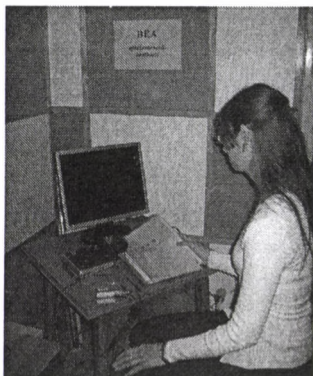
5. [www.nytud.hu/buszi/bsi.htm](http://www.nytud.hu/buszi/bsi.htm)



### A BEA fejlesztése

Mindenekelőtt a fonetikai elemzések igénye, de a szélesebb értelemben vett beszédkutatás új feladatai is szükségessé tették egy a modern technológián alapuló, a minőségi hangrögzítés minden kritériumát teljesítő, elsősorban fonetikai megalapozottságú és célú, többfunkciós korpusz létrehozását.

Az MTA Nyelvtudományi Intézet Fonetikai Osztályán 2007 őszén indult fejlesztés első szakaszában megtörtént az adatbázis tervezése, a korszerű felvételi (technikai) körülmények létrehozása (1. ábra), a tartalmi kritériumok meghatározása, a hangfelvételek készítőinek betanítása, a hangzó anyag lejegyzési stratégiáinak kialakítása, a lejegyzők betanítása, tizenöt próbafelvétel rögzítése és a próbalejegyzések elindítása. A 15 próbafelvétel sok szempontú elemzése után történt meg a felvételi protokoll véglegesítése.



1. ábra

A BEA felvételi körülményei

(MTA Nyelvtudományi Intézet, Fonetikai Osztály, csendesített helyiség)

2008-ban megkezdődtek a BEA-t kialakító beszédfelvételek. Jelenleg hat kutató, fejlesztő vesz részt a munkálatokban (adatközlők szervezése, felvételek készítése, részvétel a társalgásban, archiválás, biztonsági másolatok, dokumentálás, etikai vonatkozások biztosítása, lejegyzés, lejegyzések ellenőrzése, rendszerezés, a kutatói igények segítése).

### A BEA tartalma

Noha spontán beszédet tartalmazó adatbázisról van szó, a tervezési koncepció és a fonetikai célok kielégítése szükségessé tette, hogy – az összehasonlíthatóság érdekében – ne csak spontán beszédanyagok felvételére kerüljön sor. A BEA ezért felolvasásokat és ismétléseket is tartalmaz. A spontánbeszéd-rész is többféle, mivel tekintetbe veszi a teljes beszédproduktions folyamatot. A hangfelvételi protokoll (tartalmilag) a következő.

1. Többféle típusú spontán beszéd: *a)* narratíva a beszélő munkájáról, életéről, családjáról, hobbijáról, *b)* véleménykifejtés és *c)* tartalomösszegzés. A véleménykifejtés egy a kísérletvezető által megadott, éppen aktuális témának (amely tudást nem igényel) a véleményezése. A tartalomösszegzést megelőzően az adatközlő egy rövid tudományszerűsítő cikket (174 szavas, időtartama: 1,2 perc), majd egy történetet hall (270 szavas), mindkettőt monológként kell elmondania.

2. Mondatok utánmondása (ismétlése). 25 egyszerű és összetett mondatot – például *A farsangi bálban mindenkinek szép jelmeze volt.* – kell az adatközlőnek egyszeri meghallgatás után azonnal megismételnie.

3. Felolvasás. Az adatközlőnek az egyik felolvasáskor a 25 korábban ismelt mondatot kell meghangosítania, majd pedig a korábban hallott tudományszerűsítő cikket kell felolvasnia.

4. A felvétel mindig egy háromfős társalgással zárul. Az adatközlő és a kísérletvezetőn kívül a Fonetikai Osztály egy további (a BEA fejlesztésében folyamatosan közreműködő) tagja vesz részt a beszélgetésben. A társalgás témája változó, az élet mindennapjaihoz kapcsolódik.

### A BEA adatközlői

A korpuszban részt vevő beszélők egynyelvű, budapesti felnőttek, életkoruk 20 és 70 év közötti. Nem beszélnek nyelvjárást. Ép hallók. Az adatközlők – a koncepció szerint – arányosan oszlanak meg életkor szerint, valamint az iskolázottság tekintetében (nyolc általános, középiskolai, felsőfokú végzettség). (Az MTA Nyelvtudományi Intézetének a „*Humán vizsgálatokon alapuló nyelvészeti kutatások etikai szabályozása*”-ban leírtakat a fejlesztés minden tekintetben szigorúan betartja.) A felvételek mindig azonos körülmények között készülnek, a Fonetikai Laboratórium csendesített helyiségében. A felvevő mikrofon SONY gyártmányú, típusa ECM–MS 907. A rögzítés digitális, közvetlenül a számítógépre történik a GoldWave hangedítáló szoftverrel 44,1 kHz-es mintavételezéssel. Tárolás: 16 bit, 86 kbyte/s, monó. Az adatközlők kódolása (betű- és számsorok) rendszert alkot. A felvételek mellett rögzítjük a beszélők életkorát, esetleges beszédhibáját, iskolai végzettségét, foglalkozását, termetét, súlyát és azt, hogy dohányzik-e. Törekszünk arra, hogy az adatközlők különféle foglalkozást képviseljenek.

2008 szeptemberéig mintegy 35 órányi beszédanyag rögzítése történt meg (a becsült szószám: 285 000), az adatközlők száma 45, közülük 20 nő, 25 férfi (a munkát folyamatos). A rögzített felvételek jelenleg 9,28 GB-t tesznek ki; biztonságból DVD-ken és egy külső HDD-n is archiválva vannak. A távlati cél 500 személy beszédének rögzítése.

### A BEA felvételeinek lejegyzése

A különböző nyelvek korpuszai – mint láttuk – részben vagy egészen írásban is rögzítettek; az írásbeli tükröztetés azonban nagy különbségeket mutat



(csak helyesírási átírás, fonetikai transkripció, az intonáció és egyéb szuprasegmentumok jelölése).

A BEA felvételeinek lejegyzése a jelen fázisban a további kutatások számára legalkalmasabbnak ítélt elsődleges írásos tükröztetés. Ez azt jelenti, hogy helyesírásban, központozás nélkül írják le a lejegyzők a hanganyagot, jelezve a későbbi feldolgozás szempontjából fontosnak ítélt adatokat (megakadásjelenségek, fiziológiai hangadások stb.). A próbalejegyzések tapasztalatai alapján, többszöri megbeszéléssel és egyeztetéssel alakult ki a végleges lejegyzési útmutató.

A BEA lejegyzései tehát helyesírásban rögzítik a hangfelvételeket, ezen túl azonban számos további jelölést is tartalmaznak, azokat, amelyek az anyagot felhasználó kutatót segítik a saját munkájában. Jelenleg nem tekintjük célnak a finomabb (fonetikai) átírást, avagy a hangsúlyok jelölését. Ennek az az oka, hogy úgy véljük, a BEA-t használó kutatóknak a legnagyobb segítséget az ún. durva átírás vagy elsődleges leírás jelenti. Ez lehetővé teszi a további, finomabb fonetikai meghatározásokat, a saját szempontú és célú átírást. Az elsődleges átírások biztosan nem tartalmaznak nagy hibákat, a leírt szöveg nem „vezeti félre” a (későbbi) kutatót, ami finomabb átírás esetén az adott lejegyző és az ellenőrző személy individuális percepciójából adódóan bekövetkezhet(ne).

A rögzített szövegek lejegyzése (a Praat vagy más jó minőségű fonetikai hangelemző program használatával) nehéz feladat; megfelelő adottságot, jó beszédszlelést, kiváló fonetikai ismereteket, átírási gyakorlatot és rengeteg türelmet kíván. Egy 10 perces monológ leírása egy gyakorlott átírónak mintegy 1 órát vesz igénybe (további legalább 30 perc ugyanennek az ellenőrzése), a párbeszéd írásos tükröztetése hosszabb ideig tart, a közbevágások, egyszerre beszélések miatt. Egy átlagos hosszúságú BEA-felvétel leírása és ellenőrzése akár 12 órás munkát is jelenthet. A következő alfejezet a mindenkori lejegyzőnek szóló útmutató fő pontjait mutatja be.

### **Lejegyzési útmutató a BEA-hoz**

1. Egy beszélő teljes hanganyagát az egyes feladatrészeknek megfelelően külön hangfájlonként kell elemezni, és a lejegyzésben fel kell jegyezni az egyes részek **időtartamát**. (Mindig a kezdő beszédhangtól az utolsó beszédhang végéig.)

2. A .doc fájl neve megegyezik a hangfájl nevével (felvétel sorszáma).

3. A lejegyzés 2,5 cm-s margókkal, 1,5-ös sorközzel, 12-es Times New Roman betűtípusban történik.

4. A felvételen hallható személyek jelölése a következő: T1 (társalgó partner 1), A (adatközlő), T2 (társalgó partner 2).

5. A lejegyző a fejlécben csak a „Felvétel sorszáma”, „Lejegyző”, „A felvétel teljes hossza”, illetve „A lejegyzés dátuma” sorokat tölti ki.



6. A „mondathatárok”-at sosem jelöljük (nem alkalmazunk semmilyen központosítást), tehát lejegyzés közben nem alkalmazzuk helyesírási funkciójukban a következő írásjeleket: ; ; ! ?

7. Nagybetűt **csak** a tulajdonnevek írásakor használunk.

8. A különböző, nem a beszédhez tartozó, de a beszélő által produkált hangokat (köhögés, torokköszörülés, nyelvcsettintés, nyammogás, belégzés, kilégzés, hehezetes ejtés, nevetés stb.) a megfelelő helyen egységesen felkiáltójellel (!) jelöljük (például: *ki!mentünk!* = ez azt jelenti, hogy a *ki* után volt egy nevetés, az egész szó után pedig torokköszörülés).

9. Ha több személy egyszerre beszél, akkor az azonos időben hangzó szövegeket zárójelben ( ) jegyezzük le (amennyiben érthetők), például:

*A elnevettem*

*(A magam)*

*(T1 és arra)*

*T1 emlékszel amikor*

10. A lejegyzés alapvetően helyesírás szerint történik, ez azt jelenti tehát, hogy a kiejtés és a helyesírás eltéréseit nem jelöljük. A *játszik* szót helyesírásban jegyezzük le, azaz nem érvényesítjük az összeolvadás (vagy sorozatos hasonulás) szabályát (tehát **nem** *jáccik*). További lejegyzési szempontok:

a) Ha a beszélő olyan értelmetlen szót ejt, amelyet a lejegyző nem tud feloldani, nem értelmezhető, akkor azt két \* közt írjuk le a kiejtés szerint. Például: *ketten voltak és a \*dromek\**.

b) A megakadásjelenségeket félkövér betűtípussal jelöljük, a helyes szóalakot, amennyiben a beszélő maga nem javította, utána []-ben adjuk meg. Például: *akkor **berép** [belép] a diri*.

c) Ha a beszélő megnyújt egy hangot, azt úgy jelöljük, hogy az adott hangot kettőzve, félkövéren szedve írjuk le (például: *mikoor*).

d) Újraindítás esetén le kell jegyeznünk a szótöredéket félkövéren szedve, majd az elhangzott szót (például: ***mege**-megettem*).

e) Téves szókezdésnél vastagon szedve jegyezzük le az elhangzott töredéket (amennyiben egyértelmű, hogy a beszélő mely szót akarta kimondani, azt utána []-ben megadjuk). Például: ***mege** [megettem] mehettem*.

f) Téves szótalálásnál ugyancsak az előbbi szabály alkalmazandó (például: ***zárd be csukd be az ajtót***), amennyiben a beszélő nem javít, a helyes alakot []-ben adjuk meg.

g) Ha hiányzik a szó eleje vagy a vége, szintén félkövéren szedjük (például: *és **okt** végén vagy és **tóber** végén*). Ha egyértelmű, hogy milyen szó szakadt félbe, a helyes szóalakot []-ben adjuk meg.

h) A betűejtés szintén megakadásjelenségnek tekintendő, ilyenkor az adott hangokat vastagon szedjük (például: *és úgy **látszott** hogy*).

i) A szünet a szóban jelenséget az adott szóban megfelelő helyen elhelyezett tapadó □-tel jelöljük (például: *ki□mentem*).

j) Ha egy szó a köznyelvben használatos, de nem szótári alakjában fordul elő, akkor azt úgy jegyezzük le, ahogyan elhangzott, majd utána []-ben megadjuk a helyes alakot (például: *asszem [azt hiszem], nemtom [nem tudom], mé [miért]*).

k) Ha a beszélő nem magyar (idegen) szót ejt, azt először kiejtés szerint jegyezzük le, majd szögletes zárójelben [] megadjuk a szót az adott nyelvnek megfelelő írásmódban (például: *a jüzerném [username] rubrikába azt írod, hogy; az úticélunk Bajrajt [Bayreuth] volt*). Ugyanez érvényes az idegen nevekre stb. (például: *Dzsakomó Puccsini [Giacomo Puccini]*). Ha a lejegyző számára nem ismert az adott idegen szó írásmódja, akkor csak a kiejtés szerinti alakot adjuk meg, utána szögletes zárójelbe kérdőjelet téve (?) jelezzük az ismeretlenséget (például: *zső tem [?]*).

l) A rövidítéseket, betűszókat, mozaikszavakat is az idegen szavakhoz hasonlóan jelöljük, azaz a kiejtés szerint jegyezzük le (például: *ápévé érté [ÁPV Rt.], emtéa [MTA]*).

m) Ha a beszélő hezitál, azt a hezitálásnak megfelelően *ööö*-vel, *eee*-vel, *mmm*-mel stb. jelöljük (mindig az adott hang triplázásával, hogy ne keverhessük össze a nyújtással), és félkövéren szedjük (megakadás). Ha szó belsejében fordul elő, akkor folyamatosan írjuk a szóval (például: *kiööömentünk*).

n) A hűmmögéseket, igenléseket stb. az annak megfelelő hangokkal írjuk le (például *ühüm, hmm*).

o) Minden jól hallható néma szünetet nem tapadó □-tel jelölünk. Például: és akkor □ lejöttünk a hegyről hogy □ elkerüljük a vihart.

A lejegyzett szövegeket a mindenkori lejegyzőnek legalább kétszer célszerű ellenőrzés céljából ismételtlen meghallgatni. A 2. ábra lejegyzésrészleteket mutat egy monológból, a 3. ábra pedig egy társalgásrészletet (az eredeti 1,5-ös sorközt itt szűkítettük).

A *ööö* □ ! [tát] tehát szerintem szerintem nagyon *ööö* □ hasznos dolog hogy van *ööö* van olyan hogy elektronikus könyvtár □ *dee* □ *de* azért *ööö* mondjuk elolvasni egy regényt képernyőről az azért elég kemény dolog lenne tát [tehát] én nem hiszem hogy □ mondjuk még a maradék élvezeti *ööö* érték is elveszne szerintem *ööö* abban az esetben hogyha így kéne elolvasnom egy □ egy hosszabb lélegzetű dolgot tát [tehát] én akkor már inkább maradnék *aa* □ a papírnál □ arra az esetre hogyha nekem mondjuk *ööö* □ én mondjuk fel akarok használni valamilyen konkrét munkához □ *ööö* dolgozathoz ilyesmihez □ egy könyvet vagy szeretnék belőle idézni vagy *ööö* *mmm* hasonlók **akkor akkor** ez nagyon praktikus tud lenni meg □ meg hogyha □ tényleg nincs más lehetőségem beszerezni az adott szöveget **akkor akkor** az nagyon jó de □ de azért olvasni mégiscsak könyvből jobb!

## 2. ábra

Lejegyzésrészlet a BEA anyagából: monológ

A *egér* meg *eee mmm* *ilyesmin* [ilyesmit] nem esznek meg de *aa a av- av- v- a* *va*-  
dászati ösztön azér [azért] az benne van hogy addig ott leskelődött amíg el nem  
kapta a

A (rigót)

T1 (!)

A úgyhogy □ ! azér a macskák nem olyanok mint a kutyák

T1 hát nem *me* [mert] ők öntörvényűek nagyon *vele-* [velünk?] *va* [vagy] nekünk is  
a kollégiumba van macska □ azokkal galambokat ritkítják így napi rendszeressé-  
gel de *deő* [de] ők meg is eszik őket tehát így ! le!sben állnak a galambra ami  
majnem [majdnem] akkora mint ők és *egyszerűe* [egyszerűen] lecsapnak rájuk és  
megeszik !

T2 \*nám\* tesómék előző macskája az folyton hozta a □ megfogott ételt és felkínál-  
ta nekünk

T1 (!)

T2 (úgyhogy így)

T2 mindig ott lerakta az ajtó elé ! sógornóm mindig sikítva

T2 (elszaladt)

T1 (!)

T2 hogy na ezt ne □ de hát igen hát a jóltartásért cserébe ő is hozott egy kis rág-  
csálnivalót

T1 *mmm* de

T1 (\*kedvő\*)

T2 (hozzá-)

### 3. ábra

Lejegyzérszlet a BEA anyagából: társalgás

#### A lejegyzők

A BEA lejegyzői fonetikai kurzusokat végzett egyetemi hallgatók és dok-  
toranduszok, jelenleg összesen 6 fő, akik folyamatosan dolgoznak. A lejegyző-  
ket az előkészítő szakaszban betanítottuk, majd a próbalejegyzések közös érté-  
kelése után kezdődtek meg a már „hasznos” leírások. A lejegyzők által elké-  
szült anyagokat a Fonetikai Osztály munkatársai ellenőrzik és véglegesítik.

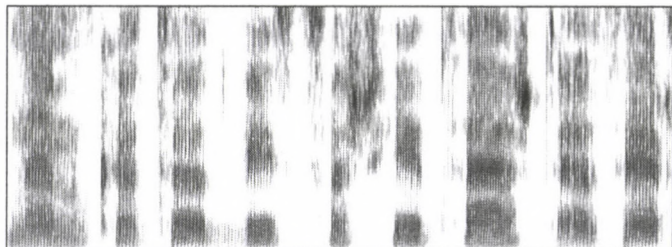
#### A BEA anyagainak felhasználásával folyó kutatások áttekintése

A spontán beszéd akusztikai szerkezete jelentős különbségeket mutat a la-  
boratóriumi vagy az olvasott beszédhez képest. Míg az utóbbiak esetében az  
adatközlőnek csak a fonetikai tervezéstől működik kvázi-tudatosan a beszéd-  
produkciós folyamata, addig a spontán beszédkor már a szándék pillanatától.  
Ennek egyértelmű lenyomata a módosult artikuláció akusztikai következmé-  
nye (4. ábra).

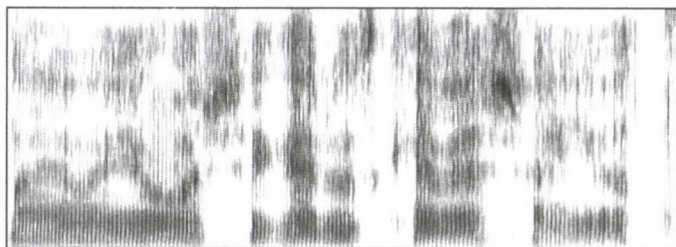
A *nem kötött biztosítást ezért* mondatrészet ismétlés, férfi adatközlője  
egyértelműen ejti a hangsorokat (átlagos beszélő, 29 éves). A néhány perccel  
később rögzített, monologikus spontán beszédrészletében (*nagyon jól sikerült*  
*kevesen jött*) a beszédhangok artikulációja kevésbé pontos, jellemzők a nem



nyelvspecifikus koartikulációk, a temporális viszonyok kiegyenlítetlenebbek, az intenzitásviszonyok erősen változók, még a zöngéképzés is kevésbé határozott.



*nem kötött biztosítást ezért*



*nagyon jól sikerült kevesen jötte*

#### 4. ábra

Mondatisméltés (felső regisztrátum) és spontán beszéd (alsó regisztrátum) alapján készült hangszíneképek szemléltetik a különböző beszédtervezés artikulációs sajátosságainak akusztikai következményeit

Noha a spontánbeszéd-korpusz végleges formája még sok évnyi munkát igényel, a már meglévő beszédanyagok alapján megindulhattak a kutatások. A legtöbb ismert beszédkorpusz is így működik (pl. Maekawa 2003). Példaként közöljük a BEA három spontán beszédanyagának elsődleges szófaji feldolgozásából kapott adatokat (e kutatás folyamatban van), vö. 1. táblázat.

A táblázat három beszélőjének adataiból máris néhány jellegzetes megállapítás tehető. Az egyes szófaji kategóriákba osztályozott szavak arányai nem függetlenek a feldolgozott korpuszrészlet nagyságától, illetőleg az összes szó számától. Szembetűnő, hogy a viszonyos aránya a legnagyobb mindhárom beszélőnél, a főnevek már jóval ritkábbak, őket a névmások követik. A tárgyas ragozású igék mintegy másfélszer gyakoribbak, mint az alanyi ragozásúak. A melléknevek használatának aránya jellemző lehet a beszélőre (5,93%, 4,45% és 8,85% az elemzett három adatközlőnél).

1. táblázat: Spontán beszédanyagok szófaji sajátosságai

| Szófaji kategóriák    | Arányok (%) |           |           |
|-----------------------|-------------|-----------|-----------|
|                       | Beszélő 1   | Beszélő 2 | Beszélő 3 |
| Ige (alanyi ragozás)  | 1,87        | 4,43      | 4,06      |
| Ige (tárgyas ragozás) | 3,64        | 6,85      | 7,29      |
| Főnév                 | 18,67       | 20,50     | 24,68     |
| Melléknév             | 5,93        | 4,45      | 8,85      |
| Számnév               | 3,28        | 2,15      | 2,63      |
| Névmás                | 17,85       | 15,99     | 9,33      |
| Viszonyzó             | 43,55       | 39,20     | 36,00     |
| Módosítószó           | 1,67        | 1,47      | 2,03      |
| Főnévi igenév         | 2,23        | 3,49      | 2,39      |
| Melléknévi igenév     | 1,14        | 0,80      | 1,79      |
| Határozói igenév      | 0,17        | 0,67      | 0,95      |
| Összes szó (db)       | 3405        | 744       | 836       |

A továbbiakban a jelenlegi kutatások három fő irányát és a célokat ismeretjük röviden. Megjegyezzük, hogy számos további vizsgálat is tervezhető a BEA anyagain, a jelen áttekintéssel azokat a kutatásokat kívánjuk bemutatni, amelyek a BEA nélkül nem vagy csak igen nehezen lennének elvégezhetőek. Az itt bemutatandó kutatások egy része már befejeződött, a kapott eredmények tanulmányok formájában e kötetben (is) olvashatók.

### 1. A beszéd szegmentális és szupraszegmentális elemzése

A BEA beszédfelvételei alapján elsőként válik lehetővé magyar nyelven is az összes magánhangzó koartikulációs mezőinek meghatározása spontán beszédben. (Hozzáteesszük, hogy ez felolvasás alapján sem áll rendelkezésre.) Az eredmények azt fogják megmutatni, hogy egy-egy magánhangzó képzése mennyire stabil a szövegkörnyezettől függően, illetőleg hogy melyek a koartikulációs mezők fonetikai tulajdonságai. Az eredmények alapján megállapítások tehetők többek között a koartikuláció erősségéről, az egyéni hangképzési sajátosságokról, a koartikuláció beszédhang- és kontextusspecifikus működéséről.

A zöngés mássalhangzók zöngésségének elemzése az egyéni ejtémód, valamint a szövegkörnyezeti hatás összefüggésében pontos információt nyújt a valóságos nyelvi realizációkról, illetőleg a hangszalagműködés sajátosságairól.

A beszédben tapasztalható hangzónyújtások természetével, akusztikai fonetikai sajátosságaival, tényezőivel foglalkozik az a kutatási irány, amely ezen nyújtásos jelenségek funkcionális hátterét igyekszik felderíteni (hangképzés, nyomaték, megakadásjelenség). Mindez összefüggésben van a spontán beszéd szegmentumainak, hangkapcsolatainak és hangsorainak temporális viszonyaival.



A spontánbeszéd-adatbázis szembesíti a kutatót az ún. „fantomhangokkal”. Ezek a beszédhangok egyértelműen észlelhetők, percepciósan relevánsak, fizikailag azonban mégsem léteznek, a rezgés- és a hangszínképeken nem láthatók, a szegmentálás során nem hallhatók. A kutatás eredménye a fantomhangok sajátosságainak leírása, ennek alapján tipizálásuk kialakítása, majd pedig az akusztikai fonetikai kontextus definiálása lesz.

Az artikulációs tempó különböző lehet a spontán beszédben és a felolvasásban. Ennek jellemzésével és a beszédhangok ejtésére gyakorolt következményeivel foglalkozik egy most induló vizsgálat sorozat. Az ismétlés, a felolvasás és a spontán beszéd összetevő fonetikai elemzésének részét képezi ez a vizsgálat. A kutatássorozatot a beszédtervezés és a beszédképzés kapcsolatáról nyújt új információkat.

A beszélők alaphangmagasságának elemzésével foglalkozó kutatás arra keres választ, hogy melyek a legjobban és melyek a legkevésbé használt frekvenciasávok a spontán beszédben, van-e jellegzetes különbség a női és a férfibeszélők között, illetve a felolvasás és a spontán beszéd alaphangfrekvencia-szerkezete mutat-e eltérést.

A BEA hanganyagai – technikai minőségük következtében – alkalmasak a hosszú idejű átlagolt energiaspektrumú (LTA) elemzések végzésére. E kutatás célja annak a megállapítása, hogy egy beszélő személy beszédének energiaspektruma statisztikai alátámasztással miként különíthető el egy másik beszélő beszédének energiaspektrumától. Döntő jelentőségű a kriminalisztikai fonetika számára annak megállapítása, hogy a beszélők közötti energiaspektrum-különbség milyen időtartamú hangminta esetén igazolható.

A mai beszélők kiejtését összevetve a mintegy 50 évvel ezelőtti Hegedűs-archívum felvételeivel (Horváth 2008), választ kaphatunk arra, miként változtak fél évszázad alatt az elemzett beszédjellemzők.

## **2. A koartikulációs folyamatok elemzése**

Számos koartikulációs folyamat vizsgálatához nélkülözhetetlen a spontánbeszéd-adatbázis. Az elméleti megállapítások és a laboratóriumi beszéd elemzésével kapott tények módosulhatnak a spontán közlések esetében. A nazális mássalhangzók progresszív és regresszív, akár több hangon át is érvényesülő koartikulációs hatásának megismerése mind a beszédképzés pontos leírásához, mind pedig az egyéni sajátosságok meghatározásához hozzájárul.

A magánhangzót követő nazális bizonyos mértékben (néhánykor teljes egészében) nazalizálja a magánhangzót. Az orrhangú mássalhangzó után ejtett obstruensek és approximánsok pedig hatással vannak mind a megelőző nazálisra, mind a magánhangzóra. Az időviszonyokra és a formánsfrekvenciákra gyakorolt hatásuk elemzése további szabályszerűségeket igazol a vizsgálható többszörös koartikulációk működésének megértéséhez.

Nagy mennyiségű spontán beszédben vizsgálható megbízhatóan a zöngéségi hasonulás érvényesülése a beszédstílus függvényében. A kutatás célja an-



nak felderítése, hogy mekkora a beszélők közötti ejtéskülönbség, valamint hogy a zöngésedés és a zöngétlenedés folyamatában mutathatók-e ki eltérések.

A beszédhangok egymásra hatása a nyelvspecifikus fonetikai és fonológiai folyamatoktól eltérően is megvalósulhat. Ezeket a változatokat különféle tényezők idézik elő. Fonetikai vizsgálatuk hozzájárul az artikuláció, az akusztikai következmény és a beszédészlelés összefüggéseinek pontosabb megértéséhez.

### **3. A spontán beszéd sajátosságainak elemzése**

A beszédtervezés és a beszédkivitelezés összefüggései elemezhetők a nagy számú beszélővel rögzített hasonló témájú monológokon. E kutatásban az ugyanannak a szövegnek a meghallgatását követő spontán beszédanyagok összevető fonetikai vizsgálata történik, az eredmények számos beszédtervezési kérdésre adnak választ.

A spontán beszéd folyamatosságát megszakító megakadásjelenségek tanulmányozása megbízhatóan végezhető rögzített beszédadatbázison. A BEA többféle ilyen jellegű kutatást tesz lehetővé. Folynak a mérésalapú elemzések annak megállapítására, hogy az ismétlések és az újraindítások milyen ejtési összefüggést mutatnak, valamint hogy miként működnek az önellenőrzési folyamatok.

A spontán beszédben előforduló egyes töltelékszavak vizsgálati adatai a szinkron változás folyamatát mutatják. Funkcióváltás igazolható bizonyos kötőszók használata, funkciója és kiejtésének összefüggései tekintetében akusztikai fonetikai adatok alapján (időviszonyok, formánsszerkezetek).

A téves kezdések és a téves szótalálások korrekciói a rejtett és a felszíni önellenőrzés következményei. A kétféle önellenőrzési folyamat elkülönítéséhez nagy mennyiségű spontán beszédanyagon végzett fonetikai elemzés szükséges.

A beszédhang, hangkapcsolat, szótag vagy szó szintű anticipációk vizsgálata a beszéd előre tervezésének, illetve előre tervezhetőségének sajátosságait szemlélteti. Rálátást nyújt a rejtett beszédtervezési folyamatok működésére, illetve az előre tervezésnek az artikulációra gyakorolt hatására.

### **Kitekintés**

A BEA spontánbeszéd-korpusz a közvetlen beszédkutatásokon túl alkalmas kriminalisztikai fonetikai (például a beszéd alapfrekvencia-értékeinek gyakorisági eloszlásával foglalkozó kutatás), bizonyos foniátriai (például a zöngképzés vizsgálata dohányzóknál) és szociofonetikai vizsgálatokra, sőt a mesterséges beszédfelismerés egyes kérdéseinek tanulmányozására.

## Irodalom

- Bael van, Christophe – Boves, Lou – van der Heuvel, Henk – Strik, Helmer 2007. Automatic phonetic transcription of large speech corpora. *Computer Speech and Language* 21. 652–668.
- Ferber, Rosa 1993. *Wie valide sind Versprechersammlungen?* Peter Lang, Bern.
- Geumann, Anja – Oppermann, Daniela – Schaeffler, Felix 1997. *The conventions for phonetic transcription and segmentation of German used for the Munich Verbmobil Corpus.* Verbmobil Memo 127. University of Munich, Germany.
- Godfrey, John – Holliman, Edward – McDaniel, Jane 1992. SWITCHBOARD: Telephone speech corpus of research and development. In *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. San Francisco, 737–740.
- Horváth Viktória 2008. A Hegedüs-archívum (1942–1962) feldolgozásának alapelvei. In Zelliger Erzsébet (szerk.): *Nyelv, területiség, társadalom*. A 14. Élőnyelvi konferencia előadásai. Magyar Nyelvtudományi Társaság, Budapest, 270–276.
- Keszler Borbála 1983. Kötetlen beszélgetések mondat- és szövegtani vizsgálata. In Rácz Endre – Szathmári István (szerk.): *Tanulmányok a mai magyar nyelv szövegtana köréből*. Tankönyvkiadó, Budapest, 164–187.
- Kontra Miklós (szerk.) 1988. *Beszélt nyelvi tanulmányok*. Linguistica, Series A, 1. MTA Nyelvtudományi Intézet, Budapest.
- Lee, Chin-Hui – Soong, Frank K. – Paiwal, Kuldip K. 1999. *Automatic Speech and Speaker Recognition*. Kluwer, Boston–Dordrecht–London.
- Leech, Geoffrey 1992. Corpora and theories of linguistic performance. In Svartvik, Jan (ed.): *Directions in corpus linguistics*. Mouton de Gruyter, Berlin, 105–122.
- Maekawa, Kikuo (2003): Corpus of spontaneous Japanese: its design and evaluation. In *Proceedings of ISCA and IEEE Workshop on Spontaneous Speech Processing and Recognition (SSPR 2003)*. Tokyo, 7–12.
- McEnery, Tony – Wilson, Andrew 1996. *Corpus linguistics*. Edinburgh University Press, Edinburgh.
- Olaszy Gábor 1999. Beszédatadatbázisok készítése gépi beszédelőállításához. *Beszédkutatás* '99. 68–90.
- Oostdijk, Nelleke 2002. The design of the Spoken Dutch Corpus. In Peters, Pam – Collins, Peter – Smith, Adam (eds.): *New frontiers of corpus research*. Rodopi, Amsterdam, 105–112.
- TIMIT: *Acoustic-phonetic continuous speech corpus* 1990. National Institute of Standards and Technology Speech Disc 1–11. NTIS (USA).
- Váradi Tamás 2000. Modern nyelvi technológiák a magyar nyelvért. In Kiefer Ferenc – Gósy Mária (szerk.): *Helyzetkép a magyar nyelvtudományról*. MTA Nyelvtudományi Intézet, Budapest, 146–156.
- Váradi Tamás 2003. A Budapesti Szociolingvisztikai Interjú. In Kiefer Ferenc – Siptár Péter (szerk.): *A magyar nyelv kézikönyve*. Akadémiai Kiadó, Budapest, 339–359.
- Vicsi Klára – Tóth László – Kocsor András – Gordos Géza – Csirik János 2002. MTBA – magyar nyelvű telefonbeszéd-adatbázis. *Híradástechnika* 8. 35–39.
- Vicsi Klára – Vig Attila 1998. Az első magyar nyelvű beszédatadatbázis. *Beszédkutatás* '98. 163–178.



## JELFELDOLGOZÁSI ÉS FONETIKAI ALGORITMUSOK KOMBINÁCIÓJA A GÉPI HANGHATÁRJELÖLÉS JAVÍTÁSÁRA

Olaszy Gábor – Bartalis Mátyás

### Bevezetés

A beszéd kutatás eszköztára egyre gazdagodik a számítógépes támogatás tekintetében. Számos, ma már ingyen is hozzáférhető szoftver segíti a kutatókat, hogy minél több adatot kapjanak a beszédjel vizsgálatából. Mindezek mellett vannak olyan kérdéskörök, amelyeknél a számítástechnikai segítséget ki kell egészíteni más természetű vizsgálati eszközökkel is, hogy megfelelő eredményt kapjunk. Ilyen terület a koartikuláció során létrejövő hangátmenet jellemzése, azaz két hang kapcsolódási pontjának meghatározása az időtengelyen. A koartikuláció folyamatossága miatt a hanghatárokat nem lehet pontosan megállapítani, azok bizonyos időintervallumhoz kötődnek (Gósy 2004). Az utóbbi évtizedekben kialakult a hanghatárok kijelölésének manuális módszere, amikor a jelölő személy meghallgatással, valamint a vizualizált akusztikai paraméterek alapján dönti el a hanghatár helyét. Az így jelölt hanghatárokat tekintjük optimális jelölésnek. Ilyen beszédadatbázis hozzáférhető a <http://fonetika.nytud.hu/cccc> honlapon (Abari–Olaszy 2006; Olaszy 2007). A digitális beszédfeldolgozás fejlődése, valamint az a követelmény, hogy egyre nagyobb beszédadatbázisokat kell készíteni bizonyos kutatásokhoz, azt eredményezte, hogy gépi módszereket is elkezdtek kidolgozni a hanghatárok szoftveres megjelölésére. Az ilyen munkát a gépi beszédfelismerésnél már jól bevált matematikai modellek támogatják. Lényegük, hogy statisztikai alapon működnek, a beszédjel spektrális összetevőinek változását figyelik, és ennek alapján jelölik be a hanghatárt a hullámformában. Az ilyen programok pontatlanabbak, mint a kézi feldolgozás, viszont nagyméretű beszédadatbázisokban gyorsan elvégzik a jelöléseket. A jelen kutatással az a célunk, hogy olyan jelölési rendszert dolgozzunk ki, amelyik megtartja a kézi jelölés pontosságát, ugyanakkor nagyryésben automatizálható is. Ezt olyan hibrid eljárással valósítjuk meg, amelyben a gépi jelölés különböző formáit kombináljuk, az emberi közreműködést minimalizáljuk, ezért a hanghatárok megállapítása költséghatékonyan végezhető el, akár nagyméretű beszédadatbázisokban is. Ilyen rendszer megvalósítására ez az első kísérlet Magyarországon. A kidolgozott módszerrel egyrésztől ténylegesen meg lehet mérni, hogy a gépi beszédfelismerési algoritmussal készített, automatikus címkézés hatásfoka milyen, másrésztől gyakorlatilag hibátlaná és hitelessé lehet tenni az adatbázis címkézését.



### A vizsgálati anyag és módszer

A kísérletekhez egy korábban rögzített, női (N) bemondó által felolvasott beszédatadabázist használtunk (Fék et al. 2006), valamint a témához előkészített korábbi munkát (Olaszy 2007a). Az N adatbázis 5821 mondatból (488 093 hang) áll, tematikus tartalmú, időjárás-jelentések szövegeit tartalmazza. Az adatbázisból 197 mondatot emeltünk ki (21 388 hang), és kézi módszerrel bejelöltük a hangokat és a hanghatárokat, valamint a szüneteket. Ezt tekintettük referencia-adatbázisnak (NR), a jelöléseket pedig etalonoknak.

Ennek a referencia-adatbázisnak a hullámformáján végeztük a kutatást, futtattuk a kísérleti algoritmusokat, próbáltuk ki a technológiai lépéseket. A futtatási eredményeket összehasonlítottuk a kézi címkézés eredményeivel, és megállapítottuk a két címkézés közötti eltéréseket. A fejlesztési munka során többször végeztük el a kísérleti algoritmusok futtatását, és többször javítottuk a fonetikai szabályokat. Akkor tekintettük elfogadhatónak a kidolgozott címkézési eljárást, amikor az NR adatbázisban az összehasonlítások során nem voltak 20 ms-nál nagyobb hanghatárjelölési eltérések a gépi és kézi jelölések között, valamint az összes hangdefiníció helyes hangot takart. Ekkor tértünk át az N adatbázis teljes anyagának a feldolgozására. E munka során alakult ki az ismertetett eljárás végleges formája. Az algoritmusok által bejelölt hanghatárok vizuális ellenőrzésére a Praat 4.0 fonetikai elemzőt használtuk. Ugyanazzal végeztük el a vizsgált beszédhullámok zöngés-zöngétlen címkézését is.

A fejlesztés során a beszédfelismerőt és a fonetikai algoritmusokat külön egységként kezeltük. A beszédfelismerő felismerési tulajdonságait nem változtattuk, azt mint kész programot fogadtuk el. A fonetikai algoritmusokat alakítottuk úgy ki, hogy a felismerő esetleges hibázásait ki tudjuk velük küszöbölni.

### Jelfeldolgozás a hanghatárok jelöléséhez

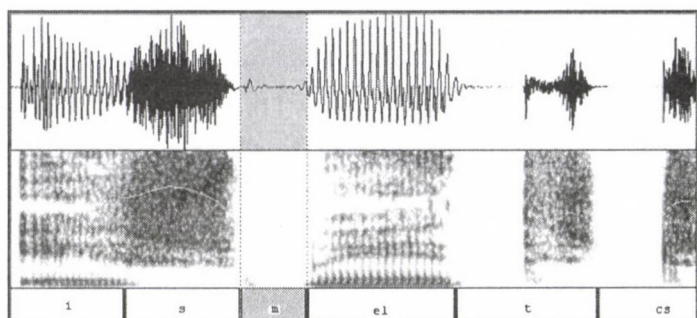
A jelen kutatásnál a Budapesti Műszaki és Gazdaságtudományi Egyetem (BME) Távközlési és Médiainformatikai Tanszékén (TMIT) kifejlesztett statisztikai alapú beszédfelismerési algoritmust használtuk. Az eljárás bemenő paramétere a szöveg, valamint az annak megfelelő hanghullám. A feldolgozási egység a mondat. A hanghullámon végzett jelfeldolgozás független a beszélőtől. Az algoritmus kiindulásként a szöveget hangszimbólumok sorozatává alakítja. Az eljárás kezeli a hasonulásokat, hangkieséseket, hangbetoldásokat. A kapott hangszimbólum-sorozatra támaszkodva végzi a program a hangfelismerést, figyelembe véve a vizsgált hang hangkörnyezetét is. A hangszimbólumok (lásd az ábrákon) megegyeznek az adott hangot reprezentáló betűvel, kivéve az ékezetes betűket, amelyeknél a hangszimbólum alapját a fő karakter és az ékezetnek megfelelő szám kombinációja adja. Példa:  $o = o$ ,  $o1 = \acute{o}$ ,  $o2 = \ddot{o}$ ,  $o3 = \tilde{o}$ ,  $a = a$ ,  $a1 = \acute{a}$ . Hangszintű építőelemnek tekintjük a *sil* jelölést is. Ez eredetileg a szünetet hivatott megjelölni, de – mint az eredmények ellenőrzése során kiderült – ez tulajdonképpen egy általánosabb jelölés, ame-

lyet minden olyan esetben is alkalmaz a beszédfelismerő, amikor bizonytalanság jelentkezik a döntésében (például glottalizáció esetén a glottalizált szakaszra *sil* jelölést tehet). A sikeres hangazonosításhoz a hangszimbólumoknak és a hanghullámban szereplő hangoknak pontos szinkronban kell lenni. A gépi felismerés folyamán a hangszimbólum-sorozat alapján úgynevezett kényszerített illesztést alkalmazunk a hanghullámra. Ebből következik, hogy annyi hanghatárt helyez el a felismerő a hanghullámban, ahány hangszimbólumot határoztunk meg a vizsgált mondatra (a felismerő által elhelyezett *sil* jelölés nem számít beszédhangnak, tehát a *sil*-ek száma nem befolyásolja a hanghatárok számát). A felismerési hálózat a szavak lineáris szekvenciájából adódik, a felismerő egyszeri futtatásával előállnak mind a hanghatárok, mind a hangidentitások (párhuzamosítva a hanghullámmal). A jelölések pontosságáról nem ad semmilyen információt a beszédfelismerő. A feldolgozás teljesen automatikus. Lényeges tulajdonsága továbbá ennek a hangfelismerési matematikai eljárásnak, hogy nem vizsgálja sem a zöngés/zöngétlen állapotot, sem a hangidőtartamot, csak spektrális jellemzők alapján dönt. A gépi címkézési algoritmust részletesen tárgyalja Mihajlik–Tatai (2001) és Mihajlik–Tatai–Gordos (2006).

Mikor téveszthet a beszédfelismerő program? A jó működés attól függ, hogy a felismerő betanítására használt sokbeszélős adatbázis címkézése mennyire pontos. Az ottani pontatlanságok felismerési hibákhoz vezetnek. A statisztikai elvből is következik, hogy néhány hanghatár pontatlan lehet. Meg kell emléteni, hogy ez a beszédfelismerő telefonsávi beszéddel lett betanítva, a magas frekvenciakomponensekben gazdag hangok, mint a [s], [ts] határainak pontos felismerése elvileg is lehetetlen. A gépi felismerő tévesztései nagyban függenek attól is, hogy a beszélő személy hogyan beszél. Gyors és laza artikuláció esetén több címkézési hiba lesz, ellenkező esetben kevesebb. Hibaforrás lehet továbbá a koartikulációból adódó olyan hangelem, amelyik fonemikus szinten nem köthető beszédhanghoz, mint például a svá (Gósy 2004), a glottalizált beszédszakasz (Böhm 2006), illetve a koartikulációs néma fázis (Olaszy 2007b). Az 1. ábrán példát láthatunk az utóbbira.

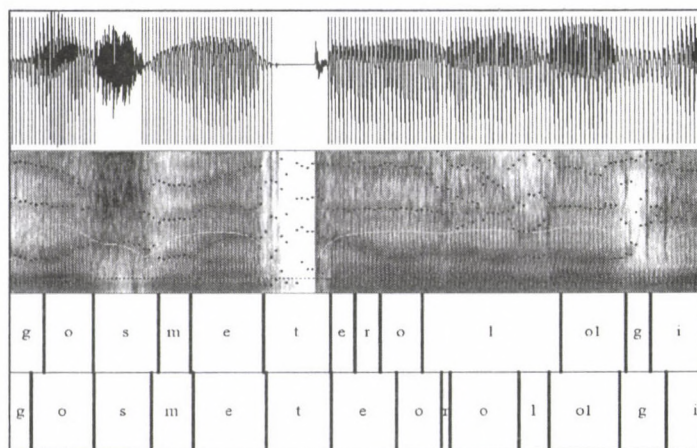
Hibát okozhat az is, amikor a hangképzési állapotok eltérnek az elméleti osztályozásoktól (pl. zöngétlen laterális mássalhangzó is kialakulhat, ilyen hangot nem definiál a magyar hangrendszer, noha a spektrális összetevőiben hasonló szerkezettel bír, mint az ugyanilyen zöngés hang). Végül címkézési hibát okozhat a szinkronitástól való eltérés is. Ilyenkor durva elcsúszások lehetnek a hangjelölésben. A 2. ábrán bemutatott hiba oka, hogy a szövegben kimaradt a *meteorológia* szó első *o* betűje, így a hangszimbólumok sorozatában csak 11 hang képviselte a szót, a hanghullámban pedig 12, hiszen a beszélő átsiklott a betűhibán és helyesen ejtette a szót.





1. ábra

Hanghatárhiba a koartikulációs néma fázis hatására (szürke jelölés). A látható hangsor: *ismét cs...*

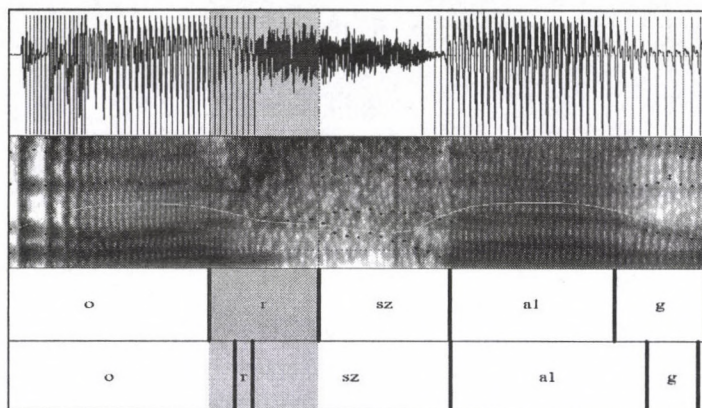


2. ábra

Példa a szinkronitási hibából eredő gépi (fent) és a helyes kézi hanghatárjelölésre (lent).

Szinkronhiba esetén a szinkront helyre kell állítani, és a felismertetést újból el kell végezni. A gépi beszédfelismerő hibáinak feltárására mérést végeztünk. Az NR adatbázis jelöléseit és a gépileg felismertett ugyanazon 197 mondatban elhelyezett hanghatárjelöléseket hasonlítottuk össze. A hibákat két csoportba soroltuk: durva hibák (rossz hangjelölés), illetve hanghatár-elcsúszási, kisebb hibák egy-egy hangon belül. Összesen 148 durva hibát találtunk. Példaként bemutatunk egy hangjelölési hibát, amely a rossz akusztikai felismerés eredménye a pergőhangra az *ország* szóban (3. ábra).



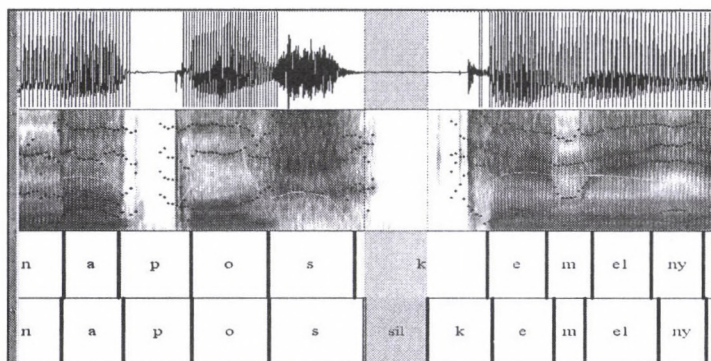


3. ábra

Példa a gépi hangjelölés (fent) és a kézi (lent) összehasonlítására.

A pergőhangot a gép rosszul jelölte a hanghullámban

A detektált hanghibákból sok a hanghullámban elhelyezett *sil*-ekhez kötődik. A felismerő minden olyan ponton *sil*-t helyezett el, ahol bizonytalanná vált a döntési folyamata. Ilyenek lehetnek a koartikulációs néma fázisok értelmezései, a glottalizációs hangrészek, de volt olyan *sil* jelölés is, amikor egy zárhang néma fázisát jelölte szünetnek a program. Példát találtunk arra is, hogy a beszédfelismerő nem detektálta a beszélő szünettartását, mivel a szünet után egy zöngétlen zárhang következett szóindításként (4. ábra). Ebben az esetben a *sil* jelölés hiánya jelentette a hibát.



4. ábra

A gépi felismerő (fent) és a kézzel jelölt hanghatárok (lent). A *napos* szó után tartott bemozdí szünetet a gép nem jelölte. A *sil* hiánya jelenti a hibát

Az összehasonlítás második fázisában azt mértük meg, hogy a gépi hanghatárjelzések milyen mértékben térnek el a kézi jelöléstől. Az eddigi ábrákon is láthattuk, hogy a gépi és kézi jelölések ritkán esnek teljesen egybe. A hibahatárt 20 ms-ban húztuk meg, ennyi eltérést még nem számítottunk hibának. A mérés eredményét az 1. táblázatban láthatjuk.

1. táblázat: A gépi és kézi hanghatárjelölések közötti különbség mennyisége a 20-ms-os küszöb felett az NR adatbázisban

| Eltérés (ms) | 20–29 | 30–39 | 40–49 | 50–59 | 60–69 | 70–79 | 80–89 | 90– |
|--------------|-------|-------|-------|-------|-------|-------|-------|-----|
| Darab        | 1112  | 374   | 170   | 81    | 43    | 16    | 9     | 22  |

Összesen 1827 hanghatárnál volt több az eltérés, mint a küszöbérték, ami azt jelenti, hogy a referencia-adatbázison futtatott gépi felismerési algoritmus mintegy 93%-os pontossággal működik kényszerített felismerésben. Az 1. táblázat részleteiből láthatjuk, hogy a kézi és gépi jelölésű hanghatárok közötti időkülönbség növekedésével egyre csökken a hibák száma. Ezt a tendenciát viszont a 90 ms-nál nagyobb eltérések száma nem követi. A 22 eset mindegyike hangjelölési hibát takart, tehát itt nem hanghatár-elcsúszásról volt szó, hanem durvább hibáról.

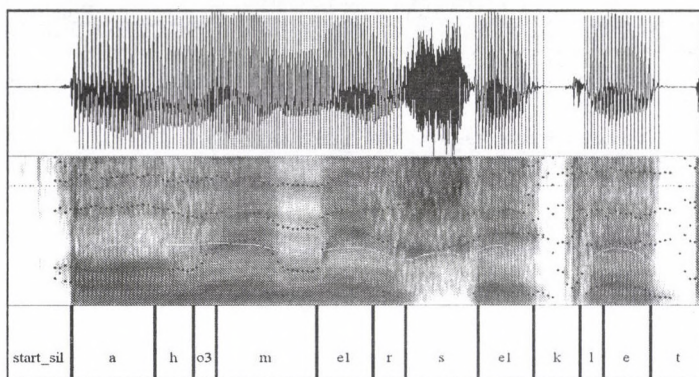
### Fonetikai algoritmusok a jelölések pontosítására

A fonetikai algoritmusok kialakítását a referencia-adatbázisban végzett mérések alapján kezdtük el. A fennmaradó 7%-nyi hibát kívánjuk javítani. A hibafeltárást és -javítást két részre osztottuk. Az első lépésben ( $F_1$ ) a gépi címkézés esetleges durva, hangjelölési hibáit kell megtalálni, majd kijavítani. Ez az alapja a második javító fázisnak ( $F_2$ ), amelyben – a már hangilag korrekt hangsoron – a kisebb mértékű hanghatár-elcsúszásokat keressük meg, majd javítjuk ki. Számítottunk arra, hogy mindkét technológiai lépés végén maradnak olyan hibák, amelyeket csak kézzel lehet kijavítani.

Kézi javításra akkor kerül sor, amikor a feltárt hiba egyedi (nem érdemes algoritmust készíteni a javítására), illetve akkor, ha géppel nem lehet eldönteni, hogy mi a helyes megoldás. A szinkronitási pontatlanságok esetén is manuálisan kell a hibát kijavítani.

Az elemzéshez a fonetikai algoritmusok a következő paramétereket használják: gerjesztési mód, hangintenzitás, a beszédhangokra vonatkozó szerkezeti formák, specifikus hangidőtartamok a hangkörnyezet függvényében (Olaszky 2006). Ezeket a paramétereket összevetik a vizsgált hangra vonatkozó értékekkel. A hangidőtartamok segítségével történő ellenőrzések főleg a tisztán zöngés szakaszok belsejében lévő hangoknál adnak bizonyos támpontot. Az 5. ábrán egy ilyen hibát mutatunk be. A program hibásan jelölte a *hőmérséklet* szóban a második hang végét, a hosszú magánhangzó feltűnően rövid, a nazális hang pedig feltűnően hosszú.





5. ábra

A beszédfelismerő hangjelölési hibája a *hőmérséklet* szó második hangjában

### Hangellenőrzési fázis

Az első lépés tehát a beszédfelismerő által jelölt hangok azonosítása, ugyanis nem biztos, hogy a beszédfelismerő által a hanghullámba bejelölt hangok és a hozzájuk rendelt hangszimbólumok megfelelnek egymásnak (lásd az 1–5. ábrákon). A fonetikai hangellenőrző ( $F_1$ ) algoritmus feladata, hogy a beszédfelismerő által adott eredményeket ellenőrizze, megadja a hangsornak azon pontjait, ahol hangjelölési durva hibát vél felfedezni. Az ellenőrző program a hanghullám összes beszédhangját akusztikailag összehasonlítja (két hanghatár közötti részt) a hozzá rendelt hangszimbólummal. Az összehasonlítást a hangokra összeállított fonetikai kategorizálás alapján (2. táblázat) teszi meg. A kategorizálásnál az oszlopok azokat a hangokat reprezentálják, amelyek a megjelölt hangtulajdonságok szerint egyformák, illetve közel esnek egymáshoz. Egyszerű szerkezetű hangoknak tekintjük azokat, amelyek folyamatosan is ejthetők, a kezdetük és végük között nincs lényegi hangzási változás.

A táblázatban megjelölt hangtulajdonságok, valamint az időtartamok gépi eszközökkel jól mérhetők a beszédhullámon. Az algoritmus működése a következő: megkapja a vizsgálandó hang szövegből származtatott szimbólumát, és ezzel egyidőben a táblázati tulajdonságok és az időtartam adatok alapján mérést végez a hullámforma hangot reprezentáló részén. Ha a táblázati besorolás és a talált, mért tulajdonságok legalább 75%-ban megegyeznek, akkor jónak ítéli meg a hangjelölést. Ellenkező esetben hibát jelez.

Vegyük sorra az 1., 3., 4., 5. ábrákon bemutatott hanghibák megtalálásának folyamatát. Az 1. ábrán a [m] hangot a beszédfelismerő olyan szakaszra jelölte be, amelynek intenzitása kicsi, és gerjesztése sem zöngés. Ez hibás jelölés. A specifikus időtartamon alapuló hibatalálatra mutat példát a 3. ábra [r] hangja. Ebben a mássalhangzó-kapcsolatban a pergőhangra jellemző maximális időtartam 40 ms, a réshangra pedig 122 ms körüli. A kettő közötti arány tehát



mintegy háromszoros. A beszédfelismerő által megadott jelölés ettől lényegesen eltér, ezért a vizsgáló algoritmus ezen a ponton hibát jelez. A 4. ábrán szintén a specifikus időtartamokkal való összehasonlítás vezet el oda, hogy ez a [k] hang helytelen jelölésű. Ilyen hanghelyzetben ez a zöngétlen zárhang 80-90 ms körüli értékű lehet (a hangsorból tudjuk, hogy nem [k:] -ről van szó). A gépi jelölési pontok közötti rész viszont 210 ms-nyi. Ez azt jelenti, hogy a beszélő valószínűleg egy rövid szünetet tartott a zárhang előtt (ezt a kézi jelölés igazolta is). Az 5. ábra hibás jelölését is a specifikus hangidőtartamokkal való összevetés alapján találtuk meg. A folyamatosan zöngés szakasz belsejében elcsúszott egy hanghatárjelölés, és ezzel felborultak a hangidőtartam-arányok.

2. táblázat: A magyar beszéd elemeinek osztályozása. A hangokat a betűjelükkel jelöltük.

(V = magánhangzó,  $h^*$  = zöngés  $h$  hang, sil = szünet, illetve értelmezhetetlen hangelem)

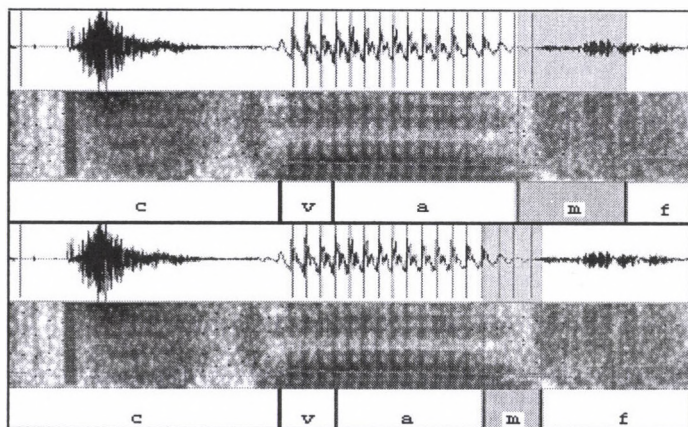
| Kategória               | 1              | 2                         | 3    | 4                         | 5                                 | 6          | 7           | 8           | 9           | 10  | 11  |
|-------------------------|----------------|---------------------------|------|---------------------------|-----------------------------------|------------|-------------|-------------|-------------|-----|-----|
| A hang és tulajdonságai | V<br>és<br>$j$ | $b$<br>$d$<br>$g$<br>$gy$ | $ny$ | $p$<br>$t$<br>$k$<br>$ty$ | $m$<br>$n$<br>$v$<br>$l$<br>$h^*$ | $h$<br>$f$ | $z$<br>$zs$ | $sz$<br>$s$ | $c$<br>$cs$ | $r$ | sil |
| Zöngés                  | X              | X                         | X    |                           | X                                 |            |             |             |             | X   |     |
| Zöngétlen               |                |                           |      | X                         |                                   | X          |             | X           | X           |     | X   |
| Vegyes gerj.            |                |                           |      |                           |                                   |            | X           |             |             |     |     |
| Intenzív                | X              |                           |      |                           |                                   |            |             |             |             |     |     |
| közepes                 |                |                           | X    |                           | X                                 |            | X           | X           | X           | X   |     |
| Gyenge                  |                | X                         |      | X                         |                                   | X          |             |             |             |     | X   |
| Egyszerű szerk.         | X              |                           |      |                           | X                                 | X          | X           | X           |             |     |     |
| Összetett szerk.        |                | X                         | X    | X                         |                                   |            |             |             | X           | X   |     |

A hangellenőrző szoftver hibalistát generál, amelyben megadja a hiba hangsorbeli helyét, hangkörnyezetét és fajtáját. A listában sokféle hiba szerepelhet, számuk általában nagy. Az  $F_1$  technológiai folyamathoz tartozik, hogy a listát manuálisan kell ellenőrizni. A hibákat kategóriákba soroltuk, majd a javító algoritmusokat ezek alapján alakítottuk ki. Tipikus hibának tekintettük, ha ugyanaz a hiba legalább 12-szer megjelent a listában. A továbbiakban manuális, vizuális és auditív vizsgálatot végeztünk a többször előforduló egyforma hibákra (5-7 ugyanolyan hibát vizsgáltunk) annak megállapítására, hogy algoritmizálható-e a hiba kijavítása. Ha egységes képet mutattak a vizsgált esetek, akkor egy-egy egyszerű algoritmust dolgoztunk ki a hiba javításához. Ezt lefuttatva a beszédadatbázison az adott hibák megszűntek, és az összesített hibaszám csökkent.

A kísérletek során néhány esetben a jó hangjelölést rossznak ítélte az 1. táblázat alapján elkészített fonetikai program. Például az intervokális helyzetű zöngés zárhangokat (*fejében*) néhány helyen hibás jelölésűnek vélte, azokban a hangkapcsolatokban, ahol a zöngés zárhangban nem lehetett periodikus elemeket detektálni (az elméleti fonetikai besorolás és az ejtett változat egy paraméterben különbözött, mert így ejtette a beszélő). Ez valójában nem hiba. Az ilyen hibás döntések elkerülésére részletesebb vizsgálatot írtunk elő az algoritmus eme szakaszára. A szabálykiegészítés a következő: VCV hangkörnyezeti esetekben a beszédfelismerő által bejelölt zöngés zárhangot minden esetben zöngésnek kell tekinteni, függetlenül a hanghullámban mérhető állapotoktól.

A kísérletek azt mutatták, hogy nem lehet olyan, általános hibajavító programot készíteni, amellyel minden hangjelölési hibát korrigálni lehet. Egyedileg vizsgáltunk meg minden hibacsoportot, és annak javítására célzott algoritmust készítettünk. Ezzel a módszerrel azt is biztosítani tudtuk, hogy az ilyen javítások során a javító program nem rontotta el az adatbázis más részét, az eredmény ellenőrzését jobban kézben lehetett tartani.

Összesen 18 fajta fonetikai javító szubrutint dolgoztunk ki a hangjelölési hibák csökkentésére. Egy példát mutatunk be ezek közül. A hiba leírása: a beszédfelismerő az [ɔmf] hangkapcsolat nazális hangját teljes egészében az őt követő zöngétlen réshangba jelölte, ami hibának minősül (6. ábra felső része).



6. ábra

Példa a nazális hang rossz címkézésére a [tsvɔmf] hangsorrészben (fent).

Az automatikus javítás utáni állapotot a lenti részen láthatjuk

A javító fonetikai algoritmus a következő: a [m] jobb oldali határát a [f] kezdetére (zöngés-zöngétlen váltási pont) kell balra elcsúsztatni, a bal oldali



határt pedig az [ɔm] hangkapcsolat belsejébe kell balra elcsúsztatni a következők szerint:

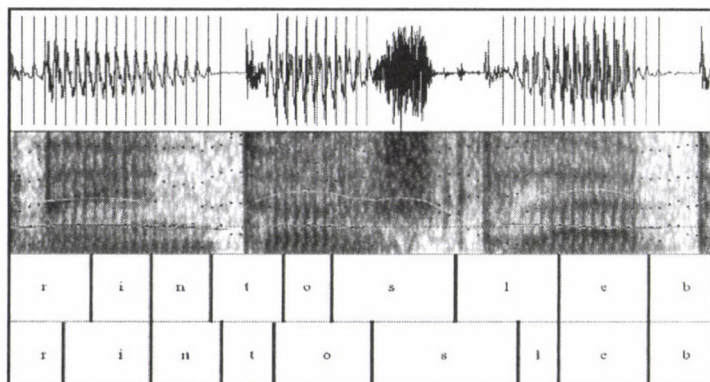
– Amennyiben az [ɔ]-t [v] előzi meg, akkor a [vɔm] teljes zöngés szakaszt az időtengely mentén 20-60-20%-ban kell felosztani (6. ábra lent) és ezek szerint kell a bal oldali hanghatárokat megjelölni.

– Ha az [ɔ]-t zöngétlen hang előzi meg, akkor az [ɔm] teljes zöngés szakaszt az időtengely mentén 50-50%-ban kell felosztani.

A 18 fonetikai algoritmus lefuttatása után megmaradt hangjelölési hibákat kézzel kell kijavítani. Itt fejeződik be az  $F_1$  javítási folyamat. Az eddigi hibafeltárással és -javításokkal elértük azt az állapotot, hogy a feldolgozott beszédadatbázisban minden jelölt hang akusztikai tartalma megfelel a vele párhuzamosan jelölt hangszimbólumnak. Innen léphetünk tovább a kisebb mértékű hanghatár-elcsúszások feltáráására és javítására.

#### A hanghatár-elcsúszások korrekciója

A fonetikai algoritmusok második csoportjának ( $F_2$ ) feladata a kismértékű hanghatár-elcsúszások megkeresése és javítása (ezen a ponton már csak beszédhangnál kisebb elcsúszások lehetnek). Itt is ugyanazon módszert alkalmaztuk, mint amelyet a hangjelölési hibáknál (lista, tipizálás, szubrutin készítés). A kialakított hibakereső algoritmusok és a javítók itt is külön programcsoportot alkotnak.



7. ábra

Példa a hanghatár-elcsúszásra (felső jelölt hangsor) és annak fonetikai algoritmussal való javítására (alsó jelölt hangsor)

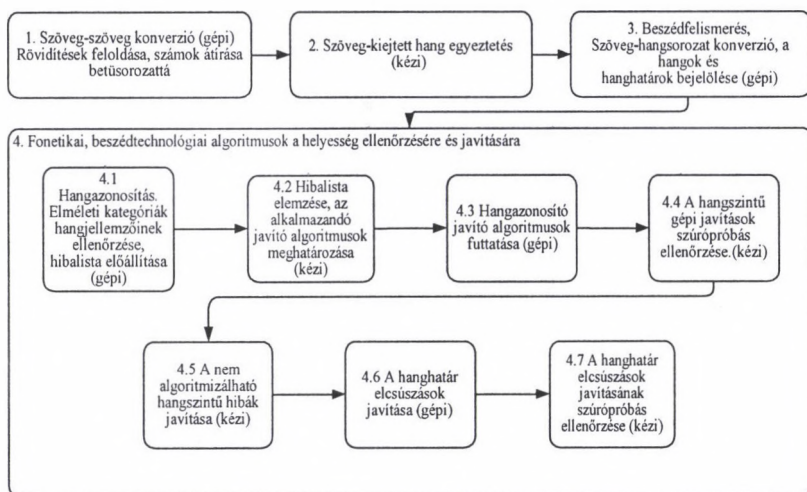
Milyen hanghatár-elcsúszásokat találhatnak meg ezek az algoritmusok? Egyrészt olyanokat, amelyek a zöngés-zöngétlen átmenetek határán jelentkeznek olyan formában, hogy a tényleges átmeneti ponttól (ami 10-15 ms-os sávban helyezhető el) távolabb van megjelölve a hanghatár. Másrészt olyan



hanghatár-elcsúszások is felfedezhetők (a specifikus hangidőtartamoktól való durva eltérés esetén), amelyek egy hosszabb zöngés szakasz belsejében vannak. Lássunk egy példát az első esetre. A 7. ábrán a *forintos lebeszélhetőség* szövegrész szóhatárán az [j] hang kezdetét az [o] hang belsejébe jelölte a beszédelfismerő (zöngés része is van a zöngétlen réshangnak). A javító algoritmus az ilyen esetekre a következő: A zöngétlen hang bal oldali hanghatárát eltoljuk az utolsó zöngéperiódusra jobbra. A zöngétlen hang jobb oldali hanghatárát pedig az öt követő zöngés hang első zöngés periódusának kezdetére állítjuk be.

### Eredmények

A fenti kutatás első és legfontosabb eredménye, hogy az ismertetett algoritmusok felhasználásával kialakítottuk azt a technológiai munkafázis-sorozatot, amellyel megállapíthatók és korrektil bejelölhetők a hanghatárok egy beszédadatbázisban. A rendszer blokksmáját a 8. ábrán mutatjuk be. Az egyes blokkoknál jelöltük a feldolgozás módját (gépi, kézi), valamint az elvégzendő feladatot. A feldolgozási sorrend szigorú. A végeredmény olyan beszédadatbázis, amelyben a hanghatárok 99,9%-a pontosan van bejelölve.



8. ábra

Beszédadatbázis címkézésére szolgáló technológiai műveletsor

Az első két lépés fontos előkészítő fázisa a feldolgozásnak. El kell érni, hogy a szöveg azt a betűsorozatot tartalmazza, amelyet a beszélő ténylegesen kiejtett (számokat, betűszavakat stb. fel kell oldani). Látható, hogy a 2. lépés-

ben az egyeztetést kézi módszerrel kell elvégezni, ez a legmunkaigényesebb része a feldolgozásnak. Az ellenőrző személynek a szöveget kell hozzáidomítania a beszédjelhez (csak a szövegben végezhet javítást). Ha a szövegben szinkronitási hibák maradnak, akkor hibás hang- és hanghatárjelölések kerülnek az adatbázisba, és a 4.1 elemzője ki fogja mutatni ezeket a hibákat. Ilyenkor a 3. lépést a hiba kijavítása után újból el kell végezni. El kell érni, hogy szinkronitási hiba ne legyen a beszédadatbázisban. A 3. lépés véglegesítése után előáll az a beszédadatbázis, amely a kiindulási változat lesz a további hibakeresésekhez, javításokhoz.

A 4.1 hangelemzője hibalistát generál, amelyben megadja a hiba hangsorbeli helyét, hangkörnyezetét és fajtáját. Az ilyen listában sokféle hanghiba szerepel, számuk általában nagy. Minden adatbázis javításánál embernek kell végignézni a listát (4.2) és dönteni a kategorizálásról és a javítás formájáról (gépi, kézi). A javító algoritmusok futtatása (4.3) után ellenőrizni kell, hogy a javítás korrekt-e, és csak ott történt meg, ahol hiba volt (4.4). A 4.5 munkafázis után az adatbázis hangjelöléseit korrektnek fogadjuk el. Ezután végezzük a hanghatár-elcsúszások javítását (4.6 és 4.7).

Az N jelű adatbázison végzett mérések, javítások eredményei a következők. A félmilliónyi hangból közel 700 hangjelölési hibát találtunk és javítottunk. Ebből 173-at kellett kézi módszerrel korrigálni. A hanghatár-elcsúszások száma 18 674 volt, ezek javítása algoritmusokkal történt meg, nem volt szükség kézi javításra. Valószínűsítjük, hogy a módszert kismértékben hozzá kell igazítani minden beszélőhöz. Elegendő beszélővel végzett adaptálás után viszont kialakulhatnak a gépi jelölésre általánosan jellemző hibacsoportok, így egy javítóalgoritmus-tár hozható létre a magyar nyelvre. Ez a munka folytatásának iránya.

### Összefoglalás

Az ismertetett új eljárás lehetőséget ad arra, hogy hanghatár-címkézési szempontból szinte hibamentes beszédadatbázisokat készítsünk. Az eljárással számszerűen is minősíteni lehet a beszédfelismerővel végzett gépi címkézés hatásosságát, amire eddig nem volt lehetőség. Az eljárás fejlesztése során világossá vált, hogy a beszédjel korrekt címkézéséhez gépi és kézi erőforrásokat is fel kell használni. Szükséges továbbá minden olyan információ, amely jellemzi a beszédjelet, tehát a pontosabb címkézéshez kombinálni kell a jelfeldolgozási és a fonetikai algoritmusokat. Az emberi tényező miatt a címkézés nem végezhető el teljesen automatikusan. Az új eljárással mintegy kétnapi munkával el lehet végezni egy új, több órányi hanganyagot tartalmazó beszédadatbázis címkézését. Az ilyen feldolgozás nagypontosságú címkézést eredményez, ami a későbbi használat során fontos tényező lehet. A pontosan címkézett beszédadatbázisok sokféle célra felhasználhatók a kutatásban, a konkrét fejlesztésekben és az oktatásban is.

## Irodalom

- Abari Kálmán – Olaszy Gábor 2006. Beszédatadatbázis a magyar mássalhangzókapcsolódások akusztikai szerkezetének bemutatására. In Alexin Zoltán – Csendes Dóra (szerk.): *IV. Magyar Számítógépes Nyelvészeti Konferencia*. Szegedi Tudományegyetem Informatikai Tanszékcsoport, Szeged, 213–222.
- Böhm Tamás 2006. A glottalizáció szerepe a beszélő személy felismerésében. *Beszédkutatás* 2006. 197–207.
- Fék, Márk – Pesti, Péter – Németh, Géza – Zainkó, Csaba – Olaszy, Gábor 2006. Corpus-based unit selection TTS for Hungarian. In Sojka, Petr – Kopeček, Ivan – Pala, Karel (eds.): *Text, Speech and Dialogue*. Springer, Brno, 367–374.
- Gósy Mária 2004. *Fonetika, a beszéd tudománya*. Osiris Kiadó, Budapest.
- Mihajlik Péter – Tatai Péter 2001. Automatikus fonetikus átírás magyar nyelvű beszédfelismeréshez. *Beszédkutatás* 2001. 172–185.
- Mihajlik, Péter – Tatai, Péter – Gordos, Géza 2006. Automatic phonetic transcription and its application in speech recogniser training: A case study for Hungarian. In Divenyi, Pierre – Greenberg, Steven – Meyer, Georg (eds.): *Dynamics of Speech Production and Perception*. (NATO Science Series I: Life and Behavioural Sciences.) IOS Press, Amsterdam, 245–262.
- Olaszy Gábor 2006. *Hangidőtartamok és időszerkezeti elemek a magyar beszédben*. Nyelvtudományi Értekezések 155. Akadémiai Kiadó, Budapest.
- Olaszy Gábor 2007a. Fonetikai algoritmus a hanghatárok gépi meghatározásának javítására nagyméretű beszédatadatbázisokban. In Alexin Zoltán – Csendes Dóra (szerk.): *V. Magyar Számítógépes Nyelvészeti Konferencia*. Szegedi Tudományegyetem Informatikai Tanszékcsoport, Szeged, 12–23.
- Olaszy Gábor 2007b. *Mássalhangzó-kapcsolódások a magyar beszédben*. Tinta Könyvkiadó, Budapest.



## „NYELVBOTLÁS”-KORPUSZ

### 5. rész

Az első magyar valós idejű, jegyzeteléses technikával gyűjtött megakadás-jelenség-korpusz első részét 5139 tétellel 2004-ben publikáltuk (*Beszédkutatás 2004.* 19–186); második részét 761 tétellel 2005-ben (*Beszédkutatás 2005.* 145–173); harmadik részét 388 tétellel 2006-ban (*Beszédkutatás 2006.* 231–247); negyedik részét 240 tétellel pedig 2007-ben (*Beszédkutatás 2007.* 187–198). A gyűjtés és az osztályozás azonban nem ért véget, a korpusz további részeit folyamatosan közzé kívánjuk tenni – e helyt újabb 444 megakadás olvasható. A kategorizálás elméleti hátterét Gósy Mária bevezető tanulmánya mutatja be részletesen: A spontán magyar beszéd megakadásainak halálás alapú gyűjteménye (*Beszédkutatás 2004.* 6–18).

Az adatokat kategóriánként, azon belül pedig ábécésorrendben közöljük. A kategóriák eredeti sorszámozását (vö. *Beszédkutatás 2004.*) megtartottuk. A „Megakadás” oszlopban találhatók a megakadást tartalmazó nyelvi adatok, a „Szándékolt közlés” oszlopban pedig a beszélő eredeti tervének megfelelő nyelvi formák – amennyiben a beszélő a téves megnyilatkozásban nem javított, és így abban az eredeti szándék nem jelenik meg. A □ jellel a szünetet jelöltük, a három pont (...) pedig az adott közlés további részleteinek ismeretlen voltára utal. A × jelet a kontaminációknál a vegyülésben szerepet játszó eredeti formák szemléltetésére alkalmaztuk.

## 2. Morfológiai, szintaktikai hiba

| Megakadás                                                                                                                 | Szándékolt közlés    |
|---------------------------------------------------------------------------------------------------------------------------|----------------------|
| A fejük mozog, azon is alapvetően az, amivel beszélnek, vagyis a szája.                                                   | szájuk               |
| A városrésznevek, amely a török háborúkkal kapcsolatosak...                                                               | amelyek              |
| Ahhoz, hogy jól tudjanak egymással dolgozni, az kell, hogy jól érezzék egymást, egymással jól érezzék magukat.            |                      |
| Behozza Perez, utána meg majd jön a cserék.                                                                               | jönnek               |
| Bezártál az ajtót?                                                                                                        | bezártad             |
| De jó, mackós bögréből iszi a kávé!                                                                                       | issza                |
| Ebbe a bugyorba tartozik a szükséges tárgyi ismeretek is.                                                                 | tartoznak            |
| ...én ne kérjek tőle, hogy gyűjtsön 150 megakadást...                                                                     | kérjem               |
| Ez nem az, mert ennek ronda a gombjai.                                                                                    | rondák               |
| Ezek a két ábra a jól teljesítők eredményeit mutatja.                                                                     | ez                   |
| Fütenek meg ilyenek, de így is megfagyik.                                                                                 | megfagy              |
| Itt most egy hosszú szövegrészek fognak következni.                                                                       | szövegrész fog       |
| Kellenek tehát □ nagyon sok felismerés hozzá.                                                                             | kell                 |
| Két ferrarisok nyugodtan várják.                                                                                          | ferraris (...) várja |
| Kovalainen mind a kétszer ott volt az élben.                                                                              | élen                 |
| Már tudom, hogy kell lótvészvirágból parfümöt csinálni, már csak azt kellene tudnom, hogy a lótvészvirágot honnan szedek! | szedem               |
| Melyek azok a szabályok, amelyek a képzés összeállításához szükséges?                                                     | szükségesek          |
| Ne szólj már hozzá, hát nem látod, hogy olvasik!                                                                          | olvas                |
| Olyan meleg van, hogy meggyulladom!                                                                                       | meggyulladok         |
| Pályában először „pole”-sorból indulva                                                                                    | pályáján             |

### 2. a) Újraindítás nélküli morfológiai változtatás

| Megakadás                                                                              | Szándékolt közlés |
|----------------------------------------------------------------------------------------|-------------------|
| Az unokatestvére rével találkozott.                                                    | unokatestvérével  |
| Egy időben az volt a hobbim, hogy felkerestem első világháborús ban harcolt katonákat. | világháborúban    |
| Ezt sokan elhisz □ ték, köztük én is.                                                  | elhitték          |
| Lapra ról olvasta az egész kiselőadását.                                               | lapról            |

**Megakadás**

...ötször indulhatott „pole”-pozícióban □ból.  
 Pici hibákka tól eltekintve jó volt a kaja.  
 Valami fogalmunk mam lehet ebből.

**Szándékolt közlés**

pozícióból  
 hibáktól  
 fogalmam

**2. b) Ismétlés toldalékjavítással****Megakadás**

Egyelőre azt az információt kapjuk kaptuk...  
 Ez a □ ezt az alapkérdést ő a Balzactól veszi át.  
 Ezt még láttuk, hogy itt megszállik □ megszáll.  
 Helyből kikap fél másodpercet másodperccel a csapattársától.  
 Itt ez a bosszantó, hogy most megint Räikkönen □ Räikkönent éri a pech...  
 Jellemző egyfajta formai homogenizmus őö homogenitás.  
 Jól megcsinálta az életét. Közben meg jó kis családja, családjja is van.  
 Nagy eltérések van a vannak a téves szótalálások között.  
 Szemnek őö szemet gyönyörködtető képekkel van teli ez a kódex.

**3. Kontamináció****Megakadás**

A pszicholingvistikák szerint...  
 Ami még nehéz □ teszi...  
 Annak nagyon boldog voltál.  
 De hát ez így egyáltalán nem éri meg az energiafeszítést!  
 Elég sok szakszókincse van.  
 Elvégre nem tehettem mást.  
 Ezt mi e nélkül is rájöttünk.  
 Fáj a feje, mégis rászív egy cigire őö vagy rágyújt egy cigire!  
 Hamilton kerékcseréjét ismételték vissza.  
 Igazságot lett szolgáltatva.  
 Így Vettelnek aztán esélytelen arra, hogy megverje Buttont.  
 Imádom, ahogy gyűlik a mosogatóban a colog □ dolog.

**Szándékolt közlés**

pszicholingvisták ×  
 pszicholingvisztikák  
 nehéz benne × nehezzé teszi  
 annak nagyon örültél × nagyon boldog voltál  
 energiaráfordítást × erőfeszítést  
 sok szakszót ismer × nagy a szakszókincse  
 elvégre × végül  
 erre e nélkül is rájöttünk ×  
 ezt e nélkül is tudjuk  
 rágyújt egy cigire × elszív egy cigit  
 ismételték meg × játszották vissza  
 igazságot szolgáltatnak × igazság lett szolgáltatva  
 Vettel esélytelen × Vettelnek nincs esélye  
 cucc × dolog



**Megakadás**

Itt van Kovalainen és a róla jellemző adatok.

Legalább el lehet különböztetni, hogy honnan jön.

Létre lehet állítani egyéni diasort.

Magam előtt képzelem, ahogy a botkezevel gyúrja a tésztát.

Magyarországon az a lényeg, hogy mi a minimum, amit szabad.

Mi az, hogy jóban lettek? Összeszerettek, akarom mondani egymásba szerettek!

Most már lüszek.

Na, akkor nekifekszek, aztán kiszülök valamit.

Nekem is nullán lennék...

Nem érted, hogy ezen döntik el?

Nem is húzom a szót.

Nincs ehhez korai?

Ő is Mátyásnak hívták.

Pica, veszel nekem csokit?

Räikkönen is belekapcsolt izé rákapcsolt.

Sok a vas van benne.

Sötétben olvasol? Miért nem kapcsoltad föl a redőnyt?

Spongyolát rá.

Szóval, a fél heteskor jössz.

Tovább teltett több időbe teltett.

Van időm összejönni.

Volt egyszer egy páros vetkőzős römi, de ha azt bulinak lehet tekinthető.

**Szándékolt közlés**

rá jellemző adatok × róla az adatok

el lehet különíteni × meg lehet különböztetni

létre lehet hozni × elő lehet állítani

magam elé képzelem × magam előtt látom

a minimum, amit kötelező × a maximum, amit szabad

összejöttek × egymásba szerettek

leülök × eszek

megszülök × kitalálok

nekem is nulla lenne × én is nullán lennék

ez alapján döntik el × ezen dől el

húzom az időt × szaporítom a szót

nincs ehhez korán × nem korai ez

ő is Mátyás × őt is Mátyásnak hívták

pici × cica

belehúzott × rákapcsolt

sok vas van × sok a vas

húztad föl a redőnyt × kapcsoltad föl a lámpát

spongyát × pongyolát

fél hétkor × a fél hetessel

tovább tartott × több időbe teltett

hazajönni × összepakolni

lehet tekinteni × tekinthető

**4. Téves szótalálás****Megakadás**

A békák kirúgtak a húsos fazékból.

A cicának nemcsak magát kell ellátnia, hanem még hat embert!

**Szándékolt közlés**

kiugrottak

hat kiscicát

## Megakadás

## Szándékolt közlés

A falon lógó festmények előtt drogéria öö dra-  
péria volt a 18. században.

A folyadék fölveszi a víz alakját.

az edény alakját

A hosszúság az halva született □ vagy bizonyta-  
lanságra született jegy.

A kevésbé iskolázatlan embereket lehet nehe-  
zebben beszervezni. Nahh, a kevésbé iskolá-  
zottakat.

A megmondó vagy bemondó emberek mégsem  
hibáztak olyan sokat.

A nápolyit se vedd föl! Ne edd meg.

A norma és a sztenderd fogalmak együttes  
használata elfogadott, ami nem akar öö vagy  
okoz problémát.

A rajongóim összegyűjtötték azokat a szövegré-  
szeket, amiket midig elfelejtettem. Ezt leírták  
és odaadták nekem összegyűrve.

összegyűjtve

A számítógép most helyettesítette az adatokat.

helyesbítette

A tejföl fel fog úszni kis csíkokba pöttyökbe a  
leves tetejére.

A tesóm is áttette a világítást! Világítást? A vizs-  
gáit 19-ére.

A tudományos forradalmak elmélete öö bocsá-  
nat □ szerkezete...

Ahogy halad előre az időszak vagy idő, ez egyre  
inkább elterjed.

Akármennyire is brillírozott a középső szekta □  
a középső szakaszban...

Ami a „Nagy könyv” mozgalomnál is kimutat-  
kozik vagy megmutatkozik, hogy az emberek  
nem feltétlenül magas irodalmat olvasnak.

Any erre paprikás krumplit csinált sonkával.

szalonnával

Átadom neked a staférungot stafétát.

Áttért az ellenség, átsétált az ellenség oldalára.

Az APEH nem szokta tőlem azt mondani, hogy  
honnan van a pénz.

kérdezni

Az ausztrál rádiónak □ illetve a tévének...

Az italautomatákat üzembe kell hagyni, javítani,  
szervizelni kell.

üzembe kell helyezni

Az tartja a földszintet is, illetve az emeletet.

Az volt a baj, hogy a barátomnak volt a barátja  
□ vagyis a barátnője, bocsánat.

| Megakadás                                                                 | Szándékolt közlés          |
|---------------------------------------------------------------------------|----------------------------|
| Aztán az anyja mondja neki, hogy menjél ki aludni őö játszani.            |                            |
| Aztán elindult a hajó teliholdas napon öö éjszákán.                       |                            |
| – Azt kérdezték, hogy hány fok volt. – Hát mit kérdezzenek egy hőmérőtől? | egy időjóstól              |
| Becsukom itt is az ablakot.                                               | ajtót                      |
| Behajtanád az ajtót?                                                      | ablakot                    |
| Csak menne rendbe az ország ügye.                                         | jönne                      |
| Csiszol csihol apád tüzet?                                                |                            |
| Csörög az eső!                                                            | csöpög                     |
| Csukja már be ott a fűnyíró!                                              | az ablakot                 |
| De aztán a katonaság is megtelt eltelt.                                   |                            |
| De ha már kimondjuk, hogy beteg, az egy mínuszos jelző.                   | negatív jelző              |
| De úgy is eltörténhetett az egész szituáció...                            | (meg)történhetett          |
| Dugulj odébb a székkal!                                                   | gurulj odébb               |
| ...egy egy egy vidék □ családi házban...                                  |                            |
| Elbújt a nyilvánosság előtt elől.                                         |                            |
| Elkészítjük a pudingot. A pudingot? A desszertet.                         |                            |
| Elvesztettem a szememet. Elvesztettem a szemüvegemet.                     |                            |
| Ennek van hármas típusa...                                                | három                      |
| Ennyi áron mondhatjuk, hogy soha ne együnk, mert úgyis megéhezünk újra.   | ennyi erővel               |
| És a Szovjetunióban sem □ mi! Oroszországban sem.                         |                            |
| Ez a démon nekem inkább arról tanúskodik □ árulkodik...                   |                            |
| Ezt szerettem volna megmondani megmutatni.                                |                            |
| Fél egytől az F-ben lesz épületem.                                        | az F-[épület]ben lesz órá  |
| Finom ez a saláta? Saláta? Vagy sajt.                                     |                            |
| Főleg, ha nem a kezembe kavarod, Robi. Vagy mi, öntöd.                    |                            |
| Fölvihetitek a mosógépet a padlásra kiteríteni!                           | vizes ruhát                |
| Funkciójának vagyis céljának megfelelően...                               |                            |
| Ha nem azonos nevű, akkor nem különböztetik meg.                          | nemű                       |
| Hát az hihetetlen volt. Nem akartunk a fülünknek hallani!                 | a fülünknek hinni          |
| Hol van a rizs? Fent a szobában.                                          | fent a (konyha)szekrényben |



## Megakadás

## Szándékolt közlés

20 percben is lehet lényegeset mondani, meg 60 percben is jókat. Hülyeségeket akartam.

Iszom, mint a görény.

Itt van a husi!

Itt van vége a pontnak a könyvnek.

Jobb oldalának választotta meg.

Június kettedikén írja...

Katonának el kellett menni, mert köteles volt.

Kész, elváltak. Kivégeztek egymással.

Kitéve vagy betéve tudtam egy verseskötetet.

Kubica és mögötte ő előtte Kovalainen.

Látni, hogy a májával van beteg □ baja.

Lehet ilyen félhivatalos kérdés formájában is megemlíteni a dolgot. (...) Ja, azt akartam mondani, hogy félhivatalos felszólítás.

Lenyeljük a tejfölös kanalat.

Ma, azt mondták, melegebb lesz, mint holnap.

Magdiéknál én vagyok a legfiatalabb.

Már csak ki kell bele hozni székét.

Margarin a sütőből ő a hűtőből?

Megszüntette □ megszöktette a lányt.

Minden vasárnap, nem minden vasárnap, hanem minden karácsonykor elmondta ezt a történetet.

Mit is kell a keresőbe írni? Harangozó? Nem! Hanghordozó!

Mondtam Csillának, hogy majd te hívod Robit □ nahh □ vagyis Balázst.

– Most át lehet menni, most nem jár a vonat. □ Aaaa metró. – A metró? – Jaaaj, a villamosra gondoltam.

Most már fegyverek közé odaértik, beleértik a kamerát is.

Na most gyulladnak ki a piros zászlók, vagy mi? A lámpák.

Nagy baj nem lesz, ha letöltöm ő vagy letöltöm.

Nagyobb volt a stressz, többre vártunk ő rosszabbra vártunk, de mégis megviselt a tengernek a mozgása, vagyis a hajónak a mozgása.

Ne egyetek még krumplis tésztát vagy pogácsát, mert még meleg!

gödény  
levestésza

jobb kezének  
másodikán  
kötelező  
végeztek egymással

lenyaljuk  
mint tegnap  
legidősebb  
párnát

krumplis pogácsát

## Megakadás

## Szándékolt közlés

Ne éles kést, pizzavágót vegyél nekem, tudod,  
azt a rolósat. Nem rolósat, gurulósat.

Ne gyere be kutya □ vagy cica!

– Ne haragudj, hogy még nem adtam vissza a  
CD-t! – Semmi baj, úgy is megvan otthon  
fénymásolatban, mármint másolatban!

Ne kötöggessél már!

Nekem is vannak kedvenc dolgaim, mint ez a  
sütő □ vagy ez a serpenyő.

Nem fogtam komolyan a dolgot, ezért rontottam  
le az átlagomat.

Nem hozott senki e-mailt?

Nem jutok sóhoz.

– Nem látok villamost és sokan várnak. – Igen,  
ilyenkor szoktam átmenni az Árkád oldalába  
□ vagyis aaaaaz Ikea oldalába.

Nem telt ellen őö el néhány perc, és minden el-  
fogyott.

Nem viszed ki a Magdát a macskát?

Nem zavar, ha rágózik, de csak, ha előtte ül egy  
széles vásznú gyerek □ széles vállú.

Nincs rajtam a gyűrűsujjam, ezért nem tudom  
megmondani, merre forduljatok!

Olyan dugó volt a templomban □ vagy hogy  
hívják □ tömeg...

Olyan utat kell kivállalni, ami se nem a felső,  
hanem a őö a középutat.

Ott mászott a combomon a hagyma. A hangya.

Ott ugrabugráltak a gyerekek az utcán az udva-  
ron.

Pfúj! Ez a szag már az én fülemet is bántja!

Rajtam nincs kifli. Késztyű.

Remélem ki fogunk kerekedni keveredni belőle.

Röviddel a halála után írta ezt a verset.

Sokszor azért kell elmennem őö azért kell haját  
mosnom...

Szakemberek és Lajosok őö laikusok...

Szeretem az uborkát, szeretem a levest, a thai-  
jal mint olyannal van dolgom bajom.

Szerintem oda van rajzolva □ ízé ragasztva van.

Szóval ez ilyen gyorstalpaló hely, igaz?

Tegnap □ Tegnap?! Tavalay Alonso is...

ne kössél már belém

nem vettem komolyan

küldött

szóhoz

gyűrűm

kiválasztani

orromat

előtt

talponálló

**Megakadás****Szándékolt közlés**

További kutatásra határozottan érdekes őő ér-  
demes...

Tudtam tanulni, mert ott volt a négynapos ün-  
nep négynapos szünet.

Valamilyen semantikus őő szemantikus nyelvi  
szerkezetben jelölte az anafora viszonyát.

Van banános csoki, bonbon, Túró Rudi. Ott van  
még a keksz, a hűtőben. Őőő, a fiókban.

Van ott tolvaj, tehát kalózfelvétel is.

Van-e a magyarban hosszú □ ja, az van, rövid á.

Végül is ebbe a pohárba föl lehet vagy be lehet  
tölteni mindent.

...világosított rá a Külügyminisztérium.

világított rá

**4. a) Téves szókezdés****Megakadás**

A ka a helikoptert nem mutatták.

A kabaréba a legérde-, legegyszerűbb szerepet kapni, mert ha kiállsz és van  
közönséged, akkor nyeregben vagy.

A kísérl □ a kutatás második része...

A különböző tudományos diszc- □ elméletek alapján...

Az előző mondatból kiala kiindulva...

Az volt a legjobb, hogy egy srá egy csávó megszólalt.

Azokat az episztemológiai őő pő elveket.

De nagyon gáz volt, mert a csá a srác volt az egyetlen, akit érdekelt a téma.

Ebben nincs Svédo □ Finnország.

Egy dara □ egy gyerek bonyolódott bele a szövegébe.

Ez egy futó □ mégiscsak egy felfüggesztéstörés lehetett.

Hamilton győzelmét nem fe □ veszélyezteteti?

Hoz a Jé a nyuszika valamit.

Képzeld el, a kislány zúgott egyet, de a könnyecsképeket felál felitattuk már.

Ki lehetett volna segí egészíteni a mondatot.

Mi lenne, ha össze visszahallgatnánk a videofelvételt?

Milyen □ zene milyen □ hangszereket használnak?

Most már tudják, miért változ választottam ezt a verset!

Most ráfordul a rajt célegyenesre.

Né né hétéves a szomszéd kislány, jövőre megy iskolába.

Nehezen érezhe érthető.

Nem csináltam volna ilyen világo □ vagy ilyen színesre.

Nem sokat kellett volna már a legeslegvégén kiharcol □ kibírnia.

Nem volt olyan, hogy ne köpő kötött volna bele valakibe.



**Megakadás**

Nem, mintha Ologun olyan nagy haj ☐ tempót diktálna.  
 Önöknek, tévénezőknek pedig csak ez kell, egy kis pi ☐ egy kis bátorság.  
 Te jössz, Boglár öő Borbála.  
 Tojá ☐ izével, kukoricás rizzsel kérem.  
 Tök jó a brin, bri, biciklid!  
 Valóba ☐ valójában nem erről van szó.  
 Webber há ☐ egy harmincöt...

**4. b) Malapropizmus****Megakadás****Szándékolt közlés**

...hogy adoptálódunk az adott kommunikációs  
 szituációhoz.

adaptálódunk

Nem, mert konzisztensen mindig ennyi köretet  
 küld.

konzekvensen

**5. „Nyelvem hegyén van”****Megakadás****Szándékolt közlés**

A freudi sűrítés, eltolás öő na! ☐ Hogy mondják  
 ezt? – Freudi elfojtás.

...a genetikától elkezdve aamm micsodákon át.

hormonokon

A kezdő kórus kórzus kurzusban az angolkerin-  
 gőt is fogjuk tanulni.

Ellepl hogy mondják? Elfedi a bajt.

Elővennéd a radiátort öő nem radiátort, hanem  
 hogy hívják? Ventilátort.

Erre is kell levél öő boríték ☐ bélyeg.

Föls ☐ föls ☐ hogy hívják ezt ☐ fölsős gyere-  
 kek...

Játszott azzal ☐ játszott ☐ mérközött azzal a  
 Keiji Ozakival...

Már csak kövön sült tengeri malac kell. Köme-  
 dence, na mi az?

kemence

Na, hogy hívják? Laza, nem, hogyis? Közvetlen  
 köszönési formák.

Nehogy fölszálljon vagy fölszaladjon vagy föl-  
 ugorjon a macska az asztalra!

Nem jut eszembe, hogy is hívják azt, ami az an-  
 gyal feje fölött van? Tudod, ilyen kerek és  
 sárga.

glória

### Megakadás

### Szándékolt közlés

Nézd, megint kirakják a virágos standokat! Csak azt tudnám, minek szedték össze a karácsonyiakat. Jaj, a húsvétiakat. Franc, a Valentin-napra gondoltam.

Tedd a mosó öö mosó öö a szemetesbe!

Teljes izé ☐ zavar.

De szeretem ezt a pasit! Hogy is hívják? Neil ☐

Neil ☐ Neil ☐ kicsoda?

Liam Neeson

Vettünk egy ilyen kis tálkát, aminek olyan formája van, mint a tea ☐ tea ☐ nem teafilter ☐ na öö mm teáskanna.

## 6. Változtatás

### Megakadás

Az, hogy ☐ ahhoz, hogy ez eredményes legyen...

Kerekítsél ebből egy ☐ történet ☐ mesét!

Le kellett másolnia egy csomó ☐ száz példányban.

## 7. Újraindítás

### Megakadás

A következő nagy alkat ☐ alkategóriának a szervezet lesz.

A legeredetibb mű mű műve...

A tudományfilozófiának ezek a műv ☐ művelői...

A versnek viszonylag az elején szere szerepel egy nő, akinek az idegeivel bajok vannak.

Akkor nevezhetné bibliográfiának, ha annak a témának az össze ☐ az összes létező, eddig megjelent szakirodalmát fetüntetné.

Amikor az irodalmilag iroda ☐ irodalmilag kevésbé képzett X megkérdezte...

Arr arra érdemes lesz odafigyelni.

Az a lényeg, hogy Kovalainen jó jól legyen.

Az utolsó szek szektorban vagy utolsó részben...

Azt a kapsola ☐ kapcsolatát fogja megírni, nem?

De most ezen a versenyen Ste Stefan Bradl...

Ebből arra lehet így kö ☐ következtetni

Egészen előzékenye ☐ előzékenyek egymással az emberek.

Ezt közölte vele a mérnöke, hogy két kétszer jön ki a boxba.

Ha valami nem sikerül, akko ☐ akkor én azt kudarcént élem meg.

Hát az anyukájával má mászkáltunk...

Hét ész ☐ észlelési és öt megértési folyamatot...

Hogy a tudomány fe ☐ fe ☐ felől hogy kell ezt mondani...

**Megakadás**

Hogy kiderüljön, ki le ☐ ki lesz majd a nigériai ellenfele.  
 Hogy sehol nincsen e e ☐ említve.  
 Így van ez a külsődleges, külsőd öö külsődleges elemekkel.  
 Mara ☐ marad-e ☐ marad-e elég üzemanyaga Sebastian Vettelnek...  
 Me merre menjek?  
 Mert ez társaa társadalomtörténeti kérdés.  
 Mondjuk azért ho holnap azért hátra van a futam.  
 Nem hiszem el, hogy csak öt cent ☐ hogy csak öt centi ment le.  
 Nem lehet, hogy al allergiás vagy?  
 Szinte ugyana, ugyanazzal az idővel.  
 Tehát a ha hatvanas évektől...  
 Tehát nee ☐ nem gondolom...  
 Többek között emberkereskedelemmel sze ☐ szerzi a vagyonát.  
 Úgy tudom, előtte matróz volt. Ho hogyan lett matróz?  
 Ugyanúgy sik sikoltva nevet a közönség.  
 Ugye ezer ezerkilencszáznolcvanhétben jelent meg.  
 Vagy vagyis föl és jobbra.

**7. a) Újraindítás változtatással****Megakadás**

A két férfi összeadott idő idő idő idejét is túlbeszélte.  
 Az emberek ezekkel a problémákkal fordultak a pszichoa ☐ mentek pszichoa-  
 analízisre.  
 Egy közösség ☐ egy elméletet alkotó tudós közösség...  
 És ez olyan belülről ☐ úgy érzem, belülről érkezett kritika...  
 Igen, ő azt szerintem ő azt mondja szerintem azt hiszi, én zaklatom őt.  
 Pedig elvileg más színű más színekkel van jelölve.  
 Tehát nem ☐ tehát pontosan onnan jön a hang.

**8. Perszeveráció****Megakadás****Szándékolt közlés**

A filmben reggel lett. És ez miért nem jutott  
 eszükbe reggel? Vaaagy este, na.  
 A mi fánkon több a fá. A fá? A virág.  
 Akkor szent a béka?  
 Az arab varázslásra esz kötelező volt  
 Az erőszakszeksze ☐ az erőszakszervezetek  
 monopóliuma...  
 Azon fáradozik, hogy az oktatás színvonala  
 jobb legyen legyen.

béke  
 ez



| Megakadás                                                                                                 | Szándékolt közlés        |
|-----------------------------------------------------------------------------------------------------------|--------------------------|
| Azt hiszem, kinyitom nyálunk nálunk is.                                                                   |                          |
| Beválasztották a bizottsági ülésbe. El kellett neki választania őö vállalnia a tagságot.                  |                          |
| Egész hetedikig elváltja, aztán harmadikba fordulja.                                                      | harmadikba fordul        |
| Ezen a beszédhangok kapcsolódásai sorát értjük.                                                           | kapcsolódási             |
| Ezért van itt, hogy a zs zsöngés.                                                                         | zsöngés                  |
| Ha van gyereke, akkor gyereksze őö anyaszerep az, amit be kell töltenie.                                  |                          |
| Heidfeld, akit a képet képen látunk.                                                                      |                          |
| Julieről nekem nincs véleményem, elvagyunk egymásról, és ennyi.                                           | elvagyunk egymással      |
| Kapsz egy csont üres lakást, amit még bebútorozni is nektek kell                                          | nektek                   |
| Kérlek, pillants rá egy picsit!                                                                           | picit                    |
| Kicsi a bors, de bo őö erős.                                                                              |                          |
| Lehet-e ennek annak köze, hogy mikor ezeket a tesztek elvégezték...                                       | lehet-e ennek ahhoz köze |
| Megpróbálom felolvasni saját gyerekkori emlékeimre haja hagyatkozva.                                      |                          |
| Mennyi tojást kérjek, tizet vagy inkább ötöt?                                                             | ötöt                     |
| Mert amikor ott van nálam, mindennek elmondja, mingyen retkesnek.                                         | minden                   |
| Mintha az ő fordításuk hübbek lebbének □ hübbek lennének.                                                 |                          |
| ...mondatkezdő töpik...                                                                                   | topik                    |
| Most ebben a japánosban volt a pás pasi.                                                                  |                          |
| Ne azt a fakakanalat vedd elő a fiókból!                                                                  | fakanalat                |
| Nem nagyon szeretem a vörösbört.                                                                          | vörösbort                |
| Nem tudom elképzelni, hogy olyan gusztustalanságot csinálnának, amibe én ne tudnék belehere, beleharapni. |                          |
| Olyankor maradj nyugodj □ nyugodt!                                                                        |                          |
| Räikkönennek volt egy-két köre Massának őö Massához képest.                                               |                          |
| Tényleg jobban megszívják magukat ezeket a piskóták.                                                      | ezek a piskóták          |
| Úgy, oda dördölözhetsz.                                                                                   | dörgölözhetsz            |
| ...utolsó előtti sör □ sor...                                                                             |                          |
| Vége van már a vacsoracsacsának?                                                                          | vacsoracsatának          |

| Megakadás                                                     | Szándékolt közlés |
|---------------------------------------------------------------|-------------------|
| ...viszonya az alannyal lehet minősítő vagy azonosító...      | azonosító         |
| Visszahívjunk, Saci? Jó, akkor visszahívjunk!                 | visszahívunk      |
| Voltak olyanok, amilyen amikre nagyon jó válasszok születtek. |                   |

## 9. Anticipáció

| Megakadás                                                                                                     | Szándékolt közlés         |
|---------------------------------------------------------------------------------------------------------------|---------------------------|
| A két lány <input type="checkbox"/> Goriót két lányának <input type="checkbox"/> azért kell a pénz...         |                           |
| A McLaren Mercedes pontnak csapatnak 14 pontja van.                                                           |                           |
| A reneszánsz perspektíva összehasonlítása...                                                                  | reneszánsz                |
| A túlméretezett tananyag e <input type="checkbox"/> viszonylag egyoldalúan fejleszti a tanulót.               |                           |
| Abszolút kitoltak Mick Heidfelddel a McLarenesek, Nick Heidfelddel a McLarenesek.                             |                           |
| Ahogy a fejlesztések hala előre haladnak...                                                                   |                           |
| Akkor bejutott ő befutott egy magyar film. És ha egy főiskolás bejut egy ilyen filmbe, akkor az sikeres lesz. |                           |
| Az biztos, hogy tizenkettő-tizenegy lesz most.                                                                | biztos                    |
| Az esélyünk így is több a egy a többmillióhoz.                                                                |                           |
| ...belekalkulálja...                                                                                          | belekalkulálja            |
| De nézd meg, a legtöbb jó angol (...) a legtöbb jó főzős műsor az angol.                                      |                           |
| Ebben van a krézkém.                                                                                          | kézkrem                   |
| Egy olyan címszó <input type="checkbox"/> tudományfilozófiai címszó alatt...                                  |                           |
| Eléggé ramakul ramatyul nézett ki.                                                                            |                           |
| Éppen az em az imént említett dolgok miatt.                                                                   |                           |
| Ez a szó <input type="checkbox"/> ez a rész a lelkiismeretéről szól.                                          |                           |
| Ez az elsőszegély ő segélyszer minden tünetre használható!                                                    |                           |
| Ez jólnap jó lesz reggel a műzlire.                                                                           | holnap jó lesz reggel     |
| Ezt úgy kell elképzelni, hogy ilyen proszt <input type="checkbox"/> posztraumás sérülésekben szenvednek.      |                           |
| Fantasztikus, látod, hogy hogy vág hozzá a vágáshoz?                                                          | hogy fog hozzá a vágáshoz |
| Lila pótó (nevet) lila póló, hátul tépett: dobbermanom nem akart elengedni.                                   |                           |

| Megakadás                                                                                                                                                                                                                                                                     | Szándékolt közlés                    |
|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------|
| Már a következő szakasban is elengedhetetlen.<br>Mégvárom, míg kiszorsolják a számaimat.<br>Melyik megállóban kell leszállni? Blajza Lujza?<br>Mert nem lehet őö nincs egy szilárd pont, amire<br>támaszkodni lehet. Nem lehet eldönteni, ki-<br>nek van igaza.               | szakaszbán<br>kisorsolják            |
| Mi lesz Hamiltonnal a vég a legvégén?<br>Moszt lesz nyilvánvaló, amikor az előszövegek-<br>kel való kapcsolatot fel tudjuk tárni.<br>Múlt órán volt szó a szonoren □ ánsokról, utána<br>az obstruensekről beszélt<br>Nagyon gyorsan tudok gépezet □ ékezet nélkül<br>gépelni. | most                                 |
| 4 őö 2 évig voltunk együtt, 4 éve mentünk szét.<br>...40 szokos kolisobában...                                                                                                                                                                                                | fokos                                |
| Nem csak egyféle delvin delfin van.<br>...nem dagdogó gyerekek megakadásai...                                                                                                                                                                                                 | dadogó<br>szörszál                   |
| Nem leszek sörszálhasogató!<br>Nem sár nem piros, hanem sárga az a rózsza.<br>Nem szfúj a szél!                                                                                                                                                                               | fúj                                  |
| Nem tudott mit kezdeni Webbellel a középső<br>etapban.                                                                                                                                                                                                                        | Webberrel                            |
| Nézd meg a két hasán □ a két oldalát! Mindjárt<br>szülni fog ez a macska.                                                                                                                                                                                                     | Nézd meg a hasának a két<br>oldalát! |
| Nézd, sznerelmes Snoopy-m van!<br>Nick Feidfeld berongyolt közékük.                                                                                                                                                                                                           | szerelmes<br>Nick Heidfeld           |
| Őök jobban ér jobban átérzik, mi az emo.<br>Őöt oldalon keresztyül osztják...                                                                                                                                                                                                 | keresztül                            |
| Próbáltuk jól érezni magunkkal egymással.<br>Remélhetőleg is képesek is leszünk erre.                                                                                                                                                                                         | magunkat egymással.                  |
| Sikerült a kanyálnak olyan részét lenyalni, ami<br>még nagyon sós volt.                                                                                                                                                                                                       | kanálnak                             |
| Só sincs arról, hogy ez jegyre megy<br>Tehát azért egy kicsit szna □ snassz volt.                                                                                                                                                                                             | szó sincs arról                      |

## 10. Metatézis

| Megakadás                                                                 | Szándékolt közlés |
|---------------------------------------------------------------------------|-------------------|
| A dolgozatban minden feladathoz három plédát<br>kell írni.                | példát            |
| A piros görbédbe, börgédbe, na a piros bögréd-<br>be töltöttem ki a teád! |                   |



| Megakadás                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          | Szándékolt közlés                                                                                                                                                                                                                                                                                                                                       |
|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| A síeség hülyelés ááááh a síelés hülyeség.<br>András szerintem családi kertes kertes családi házban lakik.<br>Arra azért jól következtetni lehet.<br>Azt mondta, ellenőrizhetteti. Ja nem, ellenőriztetheti.<br>Csillog benne ez a grilland.<br>De anyu, az jobb, mert az öntöttvas rácsból van.<br>Elég torelánsak amúgy is.<br>Eltörött a bal lábszáram és a csípsonatom.<br>Hány főzig perced a polentát?<br>Hozza ki tévéjéből a legjobbat, és egy kávé csésze? Én ezt nem értem.<br>Itt nagyon két különböző tárgyról van szó.<br>Leejtetted a parip zsebkendődet!<br>Már rehotom van a poli □ herótom van a politikától.<br>...megnyírválba a folyamatokat...<br>Mentség korom őö korom mentségére szolgál, hogy már túl vagyok a negyedik ikszen.<br>Mindig eltakarta a szemeivel a kezét.<br>Nem látom a fát az erdőtől vagy őö az erdőt a fától.<br>Ő szánta Nosferatut, szerette volna a mellét a fejére venni.<br>Ők addig nyilvánvalóan rendeni fogják a bontot.<br>Úgy éreztem, ehhez teljes nyugalom kell: lekapcsoltam a kajámat, és megettem a párael-szívót.<br>Úgy riogattak tak minkat, minket az érettségivel. | jól lehet következtetni<br>girland<br>öntöttvas rácsból<br>toleránsak<br>sípcsontom<br>hány percig főzöd<br>egy csésze kávé<br>két nagyon különböző<br>papír zsebkendődet<br><br>megnyírbálva<br><br>kezeivel a szemét<br><br>fejét a mellére<br>bontani fogják a rendet<br>lekapcsoltam a párael-szívót, és megettem a kajámat<br><br>riogattak minket |

**10. a) Az egyik felcserélt elem nem jelenik meg (zéró elem, vagy a beszélő a kimondása előtt megáll)**

| Megakadás                                                                                                    | Szándékolt közlés |
|--------------------------------------------------------------------------------------------------------------|-------------------|
| Képzeld, lesz holnap bulinál Timi!<br>Leül a szobában, és a gépében ott lenne őö az ölében ott lenne a gépe. | Timinél buli      |

**11. Egyszerű nyelvbottlás****11. b) Csere**

| <b>Megakadás</b>                                                                                         | <b>Szándékolt közlés</b> |
|----------------------------------------------------------------------------------------------------------|--------------------------|
| Apának a fizikai munkára nem nagyon jön be a salonna.                                                    | szalonna                 |
| ...és a sz <input type="checkbox"/> sima nyelvbottlás...                                                 |                          |
| Hé, te tökre olyan vagy, mint a vadalma!                                                                 | olyan                    |
| Kaphatnék még egy kis csukrot?                                                                           | cukrot                   |
| Nekik ugye nem volt lehetőségük, hogy próbálkozzanak itt az időmérő edzés közsepső közsepső szakaszában. |                          |
| Szegény Vettyel!                                                                                         | Vettel                   |
| Te jó ég! Már fél hizenkettő van?                                                                        | tizenkettő               |

**11. c) Kiesés**

| <b>Megakadás</b>                                                                               |
|------------------------------------------------------------------------------------------------|
| A női beszéd és a férfi között, férfi beszéd között sok különbség van.                         |
| Ez a háromféle korfere őő koreferenciaviszony megvalósulása.                                   |
| Ilyenek például a funkcó <input type="checkbox"/> funkcionális szavak. Ez a nevük a magyarban? |
| Kicsit tre <input type="checkbox"/> stre <input type="checkbox"/> stresszes vagyok.            |
| – Mi volt a főétel? – Ríííí <input type="checkbox"/> grízestészta.                             |

**12. Többféle osztályozás****A magánhangzó-harmónia megsértése vagy túlszabályozása; egyszerű nyelvbottlás**

| <b>Megakadás</b>                                  |
|---------------------------------------------------|
| A múlt héttől hasmenése van.                      |
| Iszonyú ordítással rárontott Hatvanyre Hatvanyra. |

**Anticipáció; metatézis; perszeveráció**

| <b>Megakadás</b>                                                      |
|-----------------------------------------------------------------------|
| Azzal a szebés sebéskész, sebés, sebészkéssel vágták meg.             |
| Háromezer külkeres nehezen fél <input type="checkbox"/> fér el rajta. |

**Anticipáció; perszeveráció**

| <b>Megakadás</b>                                                        | <b>Szándékolt közlés</b> |
|-------------------------------------------------------------------------|--------------------------|
| A mentális éren <input type="checkbox"/> érintettség kihat a tanulásra. |                          |
| A szórólaposokat megutatja megugatja                                    |                          |

| Megakadás                                                                                            | Szándékolt közlés |
|------------------------------------------------------------------------------------------------------|-------------------|
| A szöveg első rése minősítéssel kezdődik.                                                            | része             |
| A szuahélinek szok száz éves múltja van.                                                             | sok száz          |
| Az egyéb kategóriába sorolható a Farkatrét őő Farkasrét.                                             |                   |
| Az első néhány karakban kanyarban...                                                                 |                   |
| Az igemód és a tagada tagadó alakok...                                                               |                   |
| – Azt mondtad, hogy megköstölöd az almát, amit hoztam. – Meg fogom, de nem közvetlenül vacsora után. | közvetlenül       |
| Csak a hosszú mássalhangzók előtt rövidelnek meg.                                                    | rövidülnek        |
| És ebből kifolyólag nem tanultam meg bánni a fegyverrel.                                             | kifolyólag        |
| ...fono □ fonetikai folyamatot...                                                                    |                   |
| Ha megnézzük a konstruktórit, ott 28 pontja van.                                                     | konstruktórit     |
| Ha nem kezd sipáinkelni, megússza.                                                                   | sipáinkelni       |
| Ha valaki esetleg így ejti a nevét...                                                                | esetleg           |
| Itt is kidőlték az izék, húú, de hogy kitőlt!                                                        | kidőlt            |
| Jelentős rése része is a sztenderdet beszéli.                                                        |                   |
| Még volt egy Élak Élak Észak-Olaszországról szóló könyv is.                                          |                   |
| Merőben ragaszkodunk az előre ele eltervezett-hez.                                                   |                   |
| Négy három a hálynetek száma.                                                                        | helyzetek         |
| Nem azt mondtad, hogy sütünk szalonnát?                                                              | sütünk            |
| ...nyelvének a rendszeléről...                                                                       | rendszeréről      |
| Olyan iszonyú adatokat kell fordít fordítani.                                                        |                   |
| Személyes szemszögödből szemszögödből mit érzel?                                                     |                   |
| ...tehát akcentusos beséd során...                                                                   | beszéd            |
| Volt ilyen szitutok? Már a pasikot kérdezem most.                                                    | pasikot           |

### Egyszerű nyelvbontás; téves szótagolás

|                                                   |                          |
|---------------------------------------------------|--------------------------|
| <b>Megakadás</b>                                  | <b>Szándékolt közlés</b> |
| Akkor holnap találunk találkozunk!                |                          |
| Én az auditívet hozt húztam.                      |                          |
| Ha erről van szó, akkor kitér belőle öö előle.    |                          |
| Kuporgott egy kis pénzt, és összejött a 100 ezer. | kuporgatott              |
| Régen nem is volt konnektor!                      | konvektor                |



**Morfológiai, szintaktikai hiba; téves szótalálás**

| Megakadás                                                                                                                                        | Szándékolt közlés                       |
|--------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------|
| Az egyes és a het □ nem, az egyes és a kettesre szavaztak.                                                                                       |                                         |
| Az egyik csoporttársam holnap a hétvégi víkendházukhoz mennek le, őö most azt mondtam, hogy hétvégi hétvégi házukba, vagy víkend- víkendházukba. | csoporttársamék (...) a hétvégi házukba |
| Az elvárások nagyon kemények és elég őö meg van szabva, mit kell csinálni.                                                                       | eléggé                                  |
| Ezzel megelégszenek-e az Toro Rossónál?                                                                                                          | a <sup>1</sup>                          |
| Harcolnak, hogy ki a erősebb.                                                                                                                    | az                                      |
| Lassan elgondolkodhatok azon, hogyan fogom kivízkövezni, azaz víz □ vízköteleníteni.                                                             |                                         |

**Perszeveráció; téves szótalálás****Megakadás**

Kicsi vagyok, helyhi helytakarékos.  
 Szent Patrik az írek szintje, szintje? Szent őö szentje.

1. Ez a típusú megakadás (a határozott névelő *a/az* alakjának tévesztése) többször jelentkezett, más-más nyelvi kontextusban, de a korábban meghatározott elveknek megfelelően (vö. Gósy Mária: A spontán magyar beszéd megakadásainak hallás alapú gyűjteménye. *Beszéd kutatás* 2004. 6–18) a korpuszban csak egyszer-egyszer szerepeltek.

## A PEDAGÓGUSJELÖLTEK KAZINCZY-VERSENYE

### A pedagógusjelöltek Kazinczy Ferencről elnevezett szép magyar beszéd versenyének 36. országos döntője

Szombathely, 2008. április 18–20.

2008-ban a szép magyar beszéd verseny országos döntőjét a Nyugat-magyarországi Egyetem Savaria Egyetemi Központja rendezte meg Szombathelyen. Korábban már háromszor látta vendégül a szombathelyi Berzsényi Dániel Főiskola a szép magyar beszéd versenyt, ám a 2008. január 1-jével létrejött egyetem első alkalommal.

A szervezés munkálatait korábban a Magyar Nyelvészeti, most az Alkalmazott Nyelvészeti Tanszék vállalta. A verseny szervezőbizottsága: Balaskó Mária főiskolai docens (NYME), Balázs Géza egyetemi tanár (ELTE–NYME), Kovács László főiskolai adjunktus (NYME), Németh Balázs egyetemi hallgató, mb. előadó (NYME), Szibler Boglárka intézeti gazdasági ügyintéző (NYME) volt.

2007 szeptemberében (még a Berzsényi Dániel Főiskolán) megszületett a Felhívás, amely a szokásos információs csatornákon keresztül eljutott az ország felsőoktatási intézményeihez, valamint a határon túli pedagógusképző intézményekhez. Ebben a Felhívásban már szerepelt a Kazinczy-verseny számára Kovács László által készített embléma (logó) – amelyet a továbbiakban is lehet használni.

A szokásos helyi döntőket 2008 elején tartották a küldő intézmények. 2008 februárjában és márciusában értesítették a szervezőket a résztvevők a versenyzők személyéről. A döntőbe 39 hallgató jutott az ország különböző felsőfokú tanárképző intézményéből, valamint határon túli hasonló intézményekből. A döntőre 37-en érkeztek, többségük felkészítő tanára kíséretében.

A szervezőbizottság 2008 elején fölkerlte a kilencfős bírálóbizottságot, amelynek elnöke Gósy Mária volt (ELTE Fonetikai Tanszék), a tagok között pedig a tavalyi rendezők, a jövő évi rendezők, ismert fonetikus szakemberek: Bóna Judit (ELTE Fonetikai Tanszék), Imre Angéla (ELTE Fonetikai Tanszék), Jancsó Miklós (Babeş–Bolyai Tudományegyetem, Kolozsvár), N. Markó Julianna (NYME SEK Magyar Nyelvészeti Tanszék), Molnár Ildikó (Mozgássérültek Pető András Nevelőképző és Nevelőintézete), Nábrádiné Jurasek Ibolya (Vitéz János Római Katolikus Tanítóképző Főiskola), Steklács János (Kecskeméti Tanítóképző Főiskola). Valamint részt vett a zsűri munkájában az Alkalmazott Nyelvészeti Tanszék hallgatói képviselője is.

A versenyen jelen volt Bocz Zsuzsa, az Oktatási és Kulturális Minisztérium küldötte is.

A rendezvényt Gadányi Károly professzor, a Savaria Egyetemi Központ elnöke nyitotta meg. Ezután rövid tudományos előadások következtek:

Gósy Mária (ELTE Fonetikai Tanszék): A nagyothallás pedagógiai vonatkozásai;

Bóna Judit–Imre Angéla (ELTE Fonetikai Tanszék): A hangsúlyeltolódás jelensége a spontán beszédben;

Markó Alexandra – Gráczki Tekla Etelka (ELTE Fonetikai Tanszék – MTA Nyelvtudományi Intézet): A hangszalag nélküli beszéd megítélése;

N. Markó Julianna (Nyugat-Magyarországi Egyetem Savaria Egyetemi Központ): Hallgatóink spontán beszédének sajátosságai.

A verseny a szokásos mederben zajlott. Szombaton délután a szabadon választott és a felkészülés során begyakorolt szövegeket mondták el, vasárnap a reggel kiadott kötelező szöveget.

A sikeres verseny végén a 37 versenyző közül 10-en kaptak Kazinczy-érmet és pénzjutalmat: Farkas Tünde (Selye János Egyetem), Gelencsér Erika (SZTE BTK), Hevér Lóránt (SZTE JGYPK), Holics Réka (ELTE), Kriston Annamária (ELTE), Mátrai Mirella Hajna (Corvinus Egyetem), Nagy Zsófia (Nyíregyházi Főiskola), Pölcz Ádám (ELTE), Sándor Szilvia (PPKE VJK), Tóth Melinda (NYME SEK BTK). Minden versenyző kapott a könyvkiadók által följajánlott, szervezők által összegyűjtött könyvekből ajándékot.

Újdonság volt a versenyek történetében, hogy a verseny történéseit fotókkal dokumentálták. A mintegy hetven fényképet a következő internetes címen lehet megtekinteni: [picasaweb.google.com/kazinczy2008](http://picasaweb.google.com/kazinczy2008).

A rendezvényt – szokás szerint – az Oktatási és Kulturális Minisztérium anyagilag támogatta. További támogatók voltak: NYME Alkalmazott Nyelvészeti Tanszék, Apáczai Kiadó, ELTE Fonetikai Tanszék, Heiling Média Kft., Inter Kultúra-, Nyelv- és Médiakutató Központ Kht., Magyar Nyelvstratégiai Kutatócsoport, Magyar Nyugat Kiadó, Magyar Szemiotikai Társaság, MTA Nyelvtudományi Intézet Kempelen Farkas Beszédkutató Laboratórium, NYME Savaria Egyetemi Központ, Pazu-Westermann Kiadó, Rubicon-Ház Bt., Savaria University Press, Tinta Könyvkiadó, NYME Uralisztikai Tanszék, Vasi Szemle.

*Balázs Géza*



## **A Kazinczy Ferencről elnevezett szép magyar beszéd verseny 2008-as döntőjének tanulságai: az írott szöveg meghangosítása és a gyakori hibák**

A Kazinczy-versenyek döntője nemcsak versengés, hanem ünnep is, a szép magyar beszéd ünnepe. Így volt ez 2008-ban a 36. országos döntőn is, ahol töretlen lelkesedéssel, nagy számban vettek részt a magyarországi és a határon túli pedagógusképző intézmények legszebben beszélő diákjai. A versenyzők ismét nehéz feladat elé állították a zsűrit, hiszen a nagy többségük a verseny szellemiségéhez méltón, tiszta artikulációval, helyesen értelmezve olvasta fel a szövegeket.

Az írott szöveg meghangosítása során a tartalom pontosan ismert, így a felolvasónak kevesebb beszédtervezési műveletet kell végrehajtania, mint a spontán beszédben. A spontán beszéddel ellentétben ilyenkor nincs szükség a gondolatok összegyűjtésére, a lexikai válogatásra, a grammatikai formák keresésére; ezáltal a felolvasás során több idő jut a prozódianak és a szöveg érthető-értelmező interpretálásának megtervezésére.

A Kazinczyról elnevezett szép magyar beszéd versenyeken elvárás a tiszta artikuláció, a hibátlan olvasás, a helyes hangsúlyozás és intonáció, illetve a szövegegész gondolatainak érzékletes megjelenítése. Egy 1993-as tanulmány (Molnár 1993) azonban arról számol be, hogy az értelmező szövegmondás háttérbe szorította mind az olvasás pontosságát, mind a helyes kiejtés iránti igényt. Hasonló eredményt hoztak Adamikné Jászó Anna (2000) kutatásai, aki mintegy 10–15 év különbséggel végzett vizsgálatokat pedagógusnak készülő főiskolai hallgatók hangos olvasásáról. Ezek a felmérések arra hívják fel a figyelmet, hogy a magyar kiejtés egyre gyorsul, pongyolább lesz, és az olvasás színvonala romlik még a tanárjelöltek körében is.

A szöveg helyes interpretációjához a felolvasónak mindenekelőtt meg kell ismernie a leírtak tartalmát, értelmeznie kell azt, be kell fogadnia a szöveget. (A Kazinczy-versenyek kötelező szövegének esetében erre 20 perc áll a versenyzők rendelkezésére.) A tartalom mellett meg kell állapítani a szöveg műfaját, stílusát, hangulatát stb. is. A szöveg értelmezésében a tipográfiai jelölések (pl. bekezdés) és a központosítás (vesszők, gondolatjelek, kettőspont stb.) segítenek; ezek mintegy útjelzőül szolgálnak a szöveg meghangosítása során is.

Az első lépés a néma olvasás, a teljes szöveggel való megismerkedés. Ezután következik a fő részek és a kulcsszavak megkeresése; majd a részletek – azaz a mondatok közötti kapcsolatok és az egyes mondatok önálló jelentésének – értelmezése. Ezeket a felolvasónak érdemes már félhangosan, hangosan olvasnia.

A szöveg helyes meghangosításakor alapkövetelmény, hogy az olvasás az élőbeszéd hangzására hasonlítson, ne legyen mesterkélt, a felolvasó ne változtassa meg természetes hangfekvését. Továbbá elengedhetetlen a hibátlan

olvasás; a szövegértésnek megfelelő hangvétel; a megfelelő hangerő; a szöveg megfelelő tagolása beszédszakaszokra, megnyilatkozásokra, szövegegységekre (tekintettel a közléstartalomra és a grammatikai megszerkesztettségére); a jó légzéstechnika; a tiszta artikuláció, a beszédhangok helyes képzése; a kiejtési szabályok megtartása, a hangkapcsolatok helyes ejtése; a közismert idegen szavak, nevek helyes ejtése; illetve a szupraszegmentális eszközök (hangsúlyozás, beszéddallam, szünet, tempó, ritmus, hangszínezet, hangerő) adekvát alkalmazása. Ezek az értékelés szempontjával is szolgálnak.

A Kazinczy-verseny országos döntőjébe olyan egyetemi, főiskolai hallgatók kerülnek, akik saját oktatási intézményükben már bebizonyították, hogy szépen beszélnek, legalábbis szebben, mint diáktársaik. Ennek ellenére megjelennek bizonyos típushibák, amelyek évről évre több versenyzőnél is előfordulnak. Természetesen a versenyhelyzet miatt kialakuló pszichés feszültség is ronthatja a felolvasók teljesítményét. Így volt ez a 2008-as döntőn is.

A három fő hibatípus, amelyek a leggyakrabban előfordultak mind a kötelező, mind a szabadon választott szövegekben: az olvasási hiba (hangok, szótagok, rövidebb szavak betoldása, kihagyása, nagyon ritkán cseréje); a tagolási hiba (nem tagol a felolvasó, illetve rossz helyen tagolja a szöveget – ez főként a szünetek nem megfelelő alkalmazását jelenti); illetve a hangsúlyhiba (rossz hangsúlyozást hallhattunk a tulajdonneveknél – a felolvasó a vezeték-név helyett a keresztnévet hangsúlyozta, például *Szabó Katalin* – és a jelzős szerkezeteknél – a tulajdonnevekhez hasonlóan a jelző helyett a jelzett szóra került a hangsúly –, néhol pedig hiányoztak a hangsúlyok, ettől a felolvasás kissé monotonná vált).

Sajnálatos módon több olyan versenyző volt, akinek az artikulációja pontatlan volt. A legtöbbször a sziszegők és a susogók hibájával találkozunk, de előfordult az is, hogy a versenyző zárt szájjal beszélt, vagy éppen túlartikulálta a hangokat.

A versenyen a hagyományoktól kissé eltérő módon nem szépirodalmi vagy publicisztikai szöveget választottak kötelező szövegnek a szervezők; az idén egy egyetemi tankönyvből származó szövegrészletet kellett a résztvevőknek meghangosítani. Ez több szempontból is megnehezítette a felolvasást. Néhány esetben a versenyző nem találta meg a szöveghez illő hangvételt, és az internet használatáról szóló szöveget szépirodalmi szöveggént, pátosszal olvasta fel. Egy-két esetben előfordult az is, hogy a felolvasó kissé halkan vagy túl hangosan olvasott.

Bár nem a szöveg megfelelő hangvételéhez tartozik, de fontos hibaként megemlítenünk, hogy a szabadon választott szövegek bemutatásakor két versenyző fejből mondta el az általa választott részletet. A szöveg megtanulása és előadása igen komoly munkát, nagy felkészülést kívánt meg ezen versenyzőktől, ám a zsűri nem tudta megfelelően értékelni teljesítményüket. Ez a verseny ugyanis a felolvasásról, és nem a színészi teljesítményről szólt.

Végül két szupraszegmentális elem hibás alkalmazását kell megemlítenünk. Több versenyzőnél előfordultak dallamhibák (különösen a hosszabb mondatok megformálása okozott nehézséget), illetve tempóhibák. Volt, aki rossz tempóváltásokat alkalmazott. Volt, aki túl lassan olvasta a szöveget, ami dallamhibával járt együtt. Legtöbbször azonban az fordult elő, hogy a versenyző túl gyorsan olvasott. Ez egyrészt tagolási, hangsúlyozási, illetve artikulációs hibákhoz vezethet (és vezetett is), másrészt megnehezítette a hallgatóság szövegértését.

Az említett hibákat a versenyzők csak ritkán produkálták a felolvasásokban, így öröm volt zsűritagként részt venni a versenyen. A résztvevők bebizonyították, hogy a mai pedagógusjelöltek között is sokan vannak, akik szépen, helyesen tudnak magyarul beszélni.

### Irodalom

Adamikpé Jászó Anna 2000. Változott-e húsz év alatt a főiskolások kiejtése és olvasása? *Beszédkutató 2000.* 124–131.

Molnár Ildikó 1993. Tévesztés és megakadás kötött szöveg felolvasásában. *Beszédkutató 1993.* 175–184.

Bóna Judit



## A MAGYAR ÉS A NEMZETKÖZI TUDOMÁNYOS ÉLET ESEMÉNYEI

### Nyelv- és beszédtechnológiai platform

2008 nyarán megalakult a Nyelv- és Beszédtechnológiai Platform a Nemzeti Kutatási és Technológiai Hivatal (NKTH) támogatásával. A Platform működési ideje 2 év.

A nyelv- és beszédtechnológia egy feltörekvő, új iparág világszerte. A természetes nyelvi kommunikációt az ember számára a beszéd, illetve az írott szöveg biztosítja. A jövő társadalma igényli, hogy a nyelv- és beszédtechnológia egyre természetesebb megoldásokat nyújtson a jövő emberének az információ megszerzésében, kinyerésében, megadásában, a gépi rendszerekkel való kommunikációban, akár élőszóval is.

A nyelv- és beszédtechnológia olyan interdiszciplináris terület, amely statisztikai és szabályalapú módszereket felhasználó számítógépes alkalmazásokkal modellezi a nyelv írott és elhangzó formáját. Eredetileg a nyelv- és beszédtechnológiát informatikai kutatók művelték, akik számítógépes alkalmazások használatára specializálódtak a természetes nyelv feldolgozása témakörében. A jelenlegi kutatások fényében már tudjuk, hogy az emberi nyelv sokkal komplexebb, mint azt korábban gondolták, ezért a nyelv- és beszédtechnológusok gyakran interdiszciplináris kutatócsoportok tagjaiként dolgoznak, amelyek nyelvészekből (speciális képzettségekkel rendelkező nyelvészek) és informatikai szakemberekből állnak. A nyelv- és beszédtechnológiában érintett tudományterületek közül talán a legfontosabbak a következők: nyelvészet és fonetika, informatika, akusztika, mesterséges intelligencia, kognitív pszichológia, matematika és logika.

A Nyelv- és Beszédtechnológiai Platform célja, hogy felmérje a jelenlegi helyzetet, ismertesse azt széles körben a gazdaság és a politika döntéshozóival, továbbá hogy kidolgozzon egy olyan stratégiai kutatási és megvalósítási tervet, egyfajta jövőképet, amelynek köszönhetően egyrészt összehangoltabb és hatékonyabb lesz a területen működő szervezetek kutatási-fejlesztési tevékenysége, másrészt beindulhatnak azok a folyamatok, amelyek ennek az új iparágak további, dinamikus fejlődését biztosítják.

A platform jelenleg nyolc ipari és kutatási partnert tömörítő társulás.

Az MTA Nyelvtudományi Intézete végzi a platform munkájának koordinálását. További tagok: AITIA Zrt., Alkalmazott Logikai Laboratórium (ALL), Budapesti Műszaki és Gazdaságtudományi Egyetem Távközlési és Médiainformatikai Tanszéke (TMIT), valamint Média Oktatási és Kutatási Központja

(MOKK), Kilgray Fordítástechnológiai Kft., Morphologic Kft., Szegedi Tudományegyetem (SZTE) Informatikai Tanszékcsoport.

A jelenlegi platformtagok régóta együttműködő és dolgozó szakemberek, akik egyben várják további partnerek jelentkezését gazdasági, kutatási, fejlesztési területekről.

A platform programja a következő.

1. Felméri a jelenlegi irányzatokat, és ezek felhasználásával összefoglaló tanulmányt készít a jövőkép kialakításáról a nyelv- és beszédtechnológia területére vonatkoztatva. Ezzel segíti a felkészülést a jövő feltörekvő, új iparágának hatékony kiépítésére és működtetésére, valamint feltérképezi az iparág eredményeit felhasználó leendő szereplőket is.

2. A platform össze kívánja fogni a jelenlegi kutatásokat, javaslatot tesz bizonyos integrálásra annak érdekében, hogy elősegítse új kutatási és fejlesztési eredmények létrehozását, a nyelv- és beszédtechnológia súlyának erősítését a hazai szakmai érdekközösségekben.

3. Megfelelő PR munkával bemutatja a nyelv- és beszédtechnológiai kutatások, fejlesztések eredményeit, rendszerezett formában közzéteszi a működő szolgáltatások palettáját, a jövő lehetőségeit. E tevékenység keretében a következőket végezzük el:

– Internetes honlap létrehozása és működtetése. A honlap címe: <http://hlt-platform.hu> (Human Language Technology).

– Konferenciák szervezése a témakör ismertetésére. A cél, hogy aktív és interaktív formában ismertessük meg a gazdasági élet vezetőit és a politikai döntéshozókat eme új technológia lehetőségeivel, a jövő irányával, rámutatva arra, hogy ez a feltörekvő iparág a jövőben döntő szerepet fog játszani az elektronikus társadalom működési mechanizmusában, a kommunikáció, a hatásos munkavégzés, a gép-ember dilógus kialakításában.

– Sajtó, médiakommunikáció, riportok, ismeretterjesztő anyagok, műsorok létrehozása.

*Olaszy Gábor*

### **Koutny Ilona: *Natural Language Processing for Hungarian Speech Synthesis***

**(Természetes nyelvi feldolgozás a magyar beszéd-szintézishez)**

**Adam Mickiewicz University Press. Seria Językoznawstwo Stosowane 22.  
Poznan, Poland, 198 pp.**

Az angol nyelvű könyv a természetes nyelvi feldolgozás hagyományos módszereinek összekapcsolását taglalja a gépi beszédfeldolgozás egyik széles területével, a szövegek automatikus felolvastatásával (Text-to-Speech, TTS)



összekapcsolva. Fő célja, hogy összefüggéseket állapítson meg a szintaxis és prozódia között, és kidolgozzon egy szintaxis-prozódia interfészt a magyar nyelvre, azaz olyan algoritmusokat határozzon meg, melyek majd kialakítják a szintetizált beszéd prozódiaját. Ilyen összekapcsolásra eddig szinte nem végeztek kutatást. Az elemzés jól formált magyar mondatokon (amilyenek a semleges, informatív szövegekben találhatóak) és a beszélt köznyelvi magyar nyelvi normán (melyet a felolvasott információk és utasítás jellegű közlések is követnek) alapul. Az eredmény az a modell, amely lehetővé teszi, hogy természetes nyelvi elemzésre támaszkodva meghatározzuk a szintetizálandó mondat hangtestét, valamint a hangsúly- és dallamstruktúráját. A kidolgozott szabályok képesek szintagmahatárok megállapítására, ahol az intonáció változhat, és szünet fordulhat elő, továbbá a mondatbeli hangsúlyok elosztására.

Az így előállított prozódia a géppel összeállított beszéd hangzását természetesebbé teszi, így megkönnyíti a megértést. A munka beszédtechnológiai szempontból tehát hiányt pótló. A bevezető részben a szerző vázolja, hol van a nyelvfeldolgozás helye a TTS-típusú beszéd-szintézisben, milyen feladatokat kell megoldani, és milyen nehézségek fordulnak elő különböző nyelvekben.

Ezt követően röviden bemutatja a magyar nyelvet, majd a függőségi nyelvtant. A magyar mondat szerkezetek leírása ennek keretében történik, ugyanis a szerző meg van győződve arról, hogy ez a viszonylag szabad szórendű magyar nyelv jobb és egyszerűbb leírásához vezet. Mindazonáltal nem hagyta figyelmen kívül az uralkodó generatívista-strukturalista megközelítés jelentős eredményeit. A topik-fókusz szerkezet meghatározására tett kísérlet a munka egyik gyakorlati eredménye.

A magyar prozódia jellemzőinek a tanulmányozása és kategorizálása után a szerző szabályokba foglalta a mondat szerkezet és a prozódia között megállapított korrelációkat, majd tesztelte azokat oly módon, hogy az elemzés eredményeit beszéd-szintézissel szimulálta a mondatkészletre és további tesztmondatokra is, végül pedig meghallgatásos ellenőrzést végzett. Az alkalmazott módszer lényeges elemei a következők: kísérleti, természetes ejtésű mondatár (magyar férfi bemondóval felvéve), valamint a Profivox magyar szövegfelolvasó beszéd-szintetizátor kutatást is támogató interaktív kísérleti szoftverrendszere. A kísérletek eredményei igazolták a mondat szerkezet és a prozódia között megállapított kapcsolatokat. A szerző egy háromszintű intonációs modellt fejlesztett ki a magyar nyelvre: a megnyilatkozás intonációja úgy tekinthető és számolható ki, mint 1. a megnyilatkozás alap- vagy globális dallamának, 2. a benne található dallamszakaszoknak (tone groups – melyeket a mondat szerkezet határoz meg), valamint 3. a szóhangsúlyoknak – beleértve a fókuszhangsúlyt is – az interferenciája.

A könyv különböző mondatfajták prozódiaját írja le a háromszintű modell segítségével (a gyakorlatban is megvalósításra került szintetizált példamondatok meghallgathatók a [http://www.hung.amu.edu.pl/index\\_pliki/Page845.htm](http://www.hung.amu.edu.pl/index_pliki/Page845.htm) lengyel honlapon).



A megnyilatkozás utolsó dallamszakaszának a szerepe jól ismert, de a belső dallamszerkezet pontosabb leírása csak most történt meg. Vizsgálatra került a topik kiejtésének a kommenthez történő igazodása is. A szerző összehasonlította a természetes és mesterséges ejtések is. Ez a szabály alapú modell a mondatok nagy részét képes megfelelően leírni prozódiai szempontból.

A szövegelemzés irányában is új eredményeket tartalmaz a munka. Ezek kiegészítik a szerző által korábban végzett mondatelemzést, bizonyos helyzetekben módosítani tudják a mondat szintű prozódiaát (a fókusz meghatározásához további kapaszkodók kerültek elő), ez is javítja a szövegből generált szintetizált mondatok hangzását.

Végül, a könyv a magyar beszédszintézis alkalmazását mutatja be a számítógépes nyelvoktatásban és a fogyatékosok kommunikációját segítő Bliss rendszerben. A szerző egy korlátozott nyelvi modellt dolgozott ki a Bliss magyarul beszélő változatához, a Blissvox-hoz.

A könyvben foglalt kutatás fő irányvonalára még nem kapunk teljes megoldást, a kitűzött cél azonban teljesül. A teljessé tételhez további kutatások szükségesek a mondat szerkezet, a szövegszerkezet és diskurzus területén. A valenciaszótár bővítésére is szükség van, valamint a szemantikai elemzés bevonására.

A könyvet a magyar nyelv szerkezete és a beszédszintézis iránt érdeklődő diákok és kutatók figyelmébe ajánljuk.

*Olaszy Gábor*

## **BESZÉDKUTATÁS 2009**

### **A spontán beszéd**

**A konferencia ideje: 2009. október 16–17.**

**Helyszín: MTA Nyelvtudományi Intézet, földszinti előadóterem**

Jelentkezni és érdeklődni az alábbi elérhetőségeken lehet:

Horváth Viktória – e-mail: horviki@nytud.hu, telefon: 321-4830/172,  
postai cím: Budapest VI., Benczúr u. 33., 1068

Kérjük, hogy a konferenciára jelentkezés tartalmazza az alábbi adatokat:

- az előadó neve, címe, e-mail és telefonadatai;
- az előadás címe, (preferált) típusa (szóbeli vagy poszter),
- valamint csatolandó az előadás tartalmi kivonata (10 sor).

**Jelentkezési határidő: 2009. május 31.**

Az absztrakt alapján az előadás elfogadásáról 2009. június 30-ig küldünk értesítést. A lektorok által elfogadott kéziratok a következő évi *Beszédkutatós-ban* jelennek meg. A kéziratok leadási határideje minden évben május 31.

Örömmel jelentjük olvasóinknak, hogy a European Science Foundation (ESF – Európai Tudományos Alap) összeállított egy referencialistát a nemzetközi tudományosság szempontjából értékelhető folyóiratokról (ERIH = European Reference Index for the Humanities). A *Beszédkutatás* kiadványunkat 2007-ben felvették ebbe a listába, aminek az értékét növeli az, hogy annak ellenére szerepel, hogy magyar nyelvű.



## A kötet szerzői

| Név                | Intézmény                                                                            | E-levél                       |
|--------------------|--------------------------------------------------------------------------------------|-------------------------------|
| Balázs Géza        | NYME Alkalmazott Nyelvészeti Tanszék                                                 | balazsge@duc.hu               |
| Bartalis Mátyás    | BME Távközlési és Média-informatikai Tanszék                                         | bartalis@tmit.bme.hu          |
| Beke András        | ELTE Fonetikai Tanszék                                                               | beke.andras@gmail.com         |
| Bóna Judit         | ELTE Fonetikai Tanszék                                                               | bona_judit@freemail.hu        |
| Böhm Tamás         | BME Távközlési és Média-informatikai Tanszék                                         | bohm@tmit.bme.hu              |
| Gósy Mária         | MTA Nyelvtudományi Intézet                                                           | gosy@nytud.hu                 |
| Grácz Tekla Etelka | MTA Nyelvtudományi Intézet                                                           | graczi@nytud.hu               |
| Gyarmathy Dorottya | MTA Nyelvtudományi Intézet                                                           | gyarmathyd@nytud.hu           |
| Horga, Damir       | University of Zagreb, Department of Phonetics                                        | dhorga@ffzg.hr                |
| Horváth Viktória   | MTA Nyelvtudományi Intézet                                                           | horviki@nytud.hu              |
| Imre Angéla        | ELTE Fonetikai Tanszék                                                               | angelai@freemail.hu           |
| Mády Katalin       | Ludwig-Maximilians-Universität München, Institut für Phonetik und Sprachverarbeitung | mady@phonetik.uni-muenchen.de |
| Markó Alexandra    | ELTE Fonetikai Tanszék                                                               | markoxa@ludens.elte.hu        |
| Németh Géza        | BME Távközlési és Média-informatikai Tanszék                                         | nemeth@tmit.bme.hu            |
| Nikléczy Péter     | MTA Nyelvtudományi Intézet                                                           | nikleczy@nytud.hu             |
| Olaszy Gábor       | BME Távközlési és Média-informatikai Tanszék                                         | olaszy@tmit.bme.hu            |
| Tóth Bálint        | BME Távközlési és Média-informatikai Tanszék                                         | toth.b@tmit.bme.hu            |
| Ujváry István      | BME Távközlési és Média-informatikai Tanszék                                         | ujvary@alpha.tmit.bme.hu      |
| Varadi Viola       | ELTE Fonetikai Tanszék                                                               | varadiviola@gmail.com         |

2008 OKT 04





