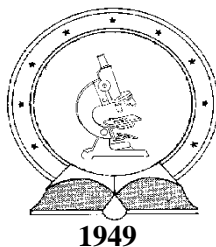


**DE TTK**



**A formánsmozgások statisztikai vizsgálata és modellezése a magyar magánhangzóknban**

Egyetemi doktori (PhD) értekezés

Abari Kálmán

Témavezető: Dr. Várterész Magda

DEBRECENI EGYETEM  
Természettudományi Doktori Tanács  
Matematika- és Számítástudományok Doktori Iskola  
Debrecen, 2013

*Ezen értekezést a Debreceni Egyetem Természettudományi Doktori Tanács Matematika- és Számítástudományok Doktori Iskola Számítástudomány és alkalmazásai programja keretében készítettem a Debreceni Egyetem természettudományi doktori (PhD) fokozatának elnyerése céljából.*

*Debrecen, 2013. július 1.*

*a jelölt aláírása*

*Tanúsítom, hogy Abari Kálmán doktorjelölt 2004.09.01- 2013.06.30. között a fent megnevezett Doktori Iskola Számítástudomány és alkalmazásai programjának keretében irányításommal végezte munkáját. Az értekezésben foglalt eredményekhez a jelölt önálló alkotó tevékenységével meghatározóan hozzájárult. Az értekezés elfogadását javasolom.*

*Debrecen, 2013. július 1.*

*témavezető aláírása*

Az értekezés elkészítését a TÁMOP-4.2.2/B-10/1-2010-0024 számú projekt támogatta.

A projekt az Európai Unió támogatásával, az Európai Szociális Alap társfinanszírozásával valósult meg.



**SZÉCHENYI TERV**

---

## **Köszönetnyilvánítás**

Köszönöm témavezetőmnek, dr. Várterész Magda egyetemi docens asszonynak, hogy doktori tanulmányaimat mindvégig figyelemmel kísérte, és mind elméleti, mind gyakorlati kérdésekben értékes tanácsokkal látott el, s dolgozatom összeállításában is segítségemre volt.

Köszönöm dr. Olasz Gábor egyetemi tanár úr önzetlen támogatását és a beszéd kutatással kapcsolatos nélkülözhetetlen tanácsait, amivel alapvetően hozzájárult szakmai fejlődésemhez és sikeres munkámhoz.

Köszönöm dr. Németh Géza egyetemi docens úrnak, hogy rendelkezésemre bocsátotta a BME Távközlési és Médiainformatikai Tanszék Beszédtechnológiai laboratóriumának 10 beszélős nagyméretű beszédadatbázisát, amely egyedülálló Magyarországon és kutatásom alapját képezte.

Köszönetet mondok dr. Máth János egyetemi docens kollegának a munka során nyújtott útmutató és hasznos tanácsaiért, valamint a jó hangulatban eltöltött munkanapokért.

Köszönöm a DE Pszichológiai Intézet és a Szociál- és Munkapszichológiai Tanszék vezetőinek és munkatársainak támogatását és segítségét.

Köszönettel tartozom családomnak és barátaimnak, akik biztattak és mellettem álltak a fokozatszerzéshez szükséges munka során.

# A formánsmozgások statisztikai vizsgálata és modellezése a magyar magánhangzókbán

Értekezés a doktori (Ph.D.) fokozat megszerzése érdekében  
a matematika- és számítástudományok tudományágban

Írta: Abari Kálmán okleveles programtervező matematikus

Készült a Debreceni Egyetem Matematika- és Számítástudományok Doktori  
Iskola (Számítástudomány és alkalmazásai programja) keretében

Témavezető: Dr. Várterész Magda

A doktori szigorlati bizottság:

elnök: Dr. ....  
tagok: Dr. ....  
Dr. ....

A doktori szigorlat időpontja: 20... ..

Az értekezés bírálói:

Dr. ....  
Dr. ....  
Dr. ....

A bírálóbizottság:

elnök: Dr. ....  
tagok: Dr. ....  
Dr. ....  
Dr. ....  
Dr. ....

Az értekezés védésének időpontja: 20... ..

## TARTALOMJEGYZÉK

1. Bevezetés.....	1
1.1. Motiváció.....	4
1.2. Az értekezés felépítése .....	4
1.3. Tézisek.....	5
2. Háttér, irodalmi áttekintés .....	7
2.1. A magánhangzók spektrális jellemzői: statikus és dinamikus szemlélet... 7	
2.2. A formánsok.....	10
2.3. Formánsmenet modellek .....	10
2.4. A magyar beszédhangok .....	11
2.5. A magyar formánskutatás.....	12
2.6. Reprodukálható kutatás .....	14
2.7. Beszédatbázisok kialakítása.....	18
2.8. Formánsadatbázisok kialakítása .....	20
2.9. Funkcionális adatelemzés.....	21
3. Anyag, módszer, kísérleti személyek .....	32
3.1. A forrásadatbázis (FKM).....	32
3.2. A formánsmérésekhez kialakított vizsgálati beszédatbázis (FEF).....	34
3.3. Formánsadatok .....	34
3.4. A PROFEF vizuális formánsadat-ellenőrző program.....	35
3.5. Magánhangzó normalizálás.....	36
3.6. Hangkategória definíciók .....	38
4. Reprodukálható beszédkutatás .....	40
4.1. A javasolt modell.....	40
4.2. Korlátok.....	43
4.3. Reprodukálhatóság a jelen kutatásban.....	44
5. A formánsadatbázis elkészítésére kidolgozott eljárás és alkalmazása .....	45
5.1. Gépi formánsmérés.....	47

5.2. A félautomatikus hibajavítás modellezése .....	56
5.3. A félautomatikus hibajavítás értékelése .....	68
5.4. Félautomatikus hibajavítás a FEF adatbázisra.....	70
5.5. A félautomatikus hibajavítás tapasztalatai, következtetések .....	71
6. Formánsmenetek vizsgálata .....	72
6.1. A hangsúly kezelése .....	72
6.2. Javasolt módszerek a formánsmozgások kontextusfüggő vizsgálatához	75
7. Összefoglalás.....	104
7.1. A kutatás eredményei és jelentősége .....	104
7.2. Jövőbeni tervek.....	108
8. Summary .....	109
8.1. Research results and their significance.....	109
8.2. Future plans .....	113
Függelék.....	114
Irodalomjegyzék.....	124

# 1. BEVEZETÉS

A beszédtudomány egy rendkívül széles, multidiszciplináris terület. Központjában az emberi beszédkommunikáció áll, amelynek célja, hogy gondolatokat juttassunk el más emberekhez hangzó formában. A beszédet legátfogóbban a fonetika vizsgálja, melynek kutatói annak a folyamatnak a megismerését és leírását tűzték ki célul, amely a beszéd agyi tervezésétől, a képzésen és a levegőben továbbított fizikai rezgéseken át, a beszéd észleléséig tart. A fonetikusok, szoros kapcsolatban állva más tudományterületek (például pszicholingvisztika, foniátria, fiziológia, fizikai akusztika, fül-orr-gégeszet, audiológia, matematika és informatika) művelőivel, azt tanulmányozzák, hogy (1) miként hozzuk létre gondolatainkból a beszédet az artikulációs szerveinkkel, (2) a levegő rezgései milyen formában tartalmazzák az eredeti gondolatot, illetőleg az annak megfelelő nyelvi jelentéseket, (3) az akusztikus jelekből hogyan állítjuk vissza az eredeti nyelvi tartalmat, illetőleg a beszélő eredeti gondolatát. A beszélő és a hallgató(k) között zajló, fent vázolt folyamat a beszéd körfolyamata vagy másképp a természetes beszédlánc [Denes & Pinson(1993)], [Gósy(2004):14–15].

A fonetika alapvetően kísérleti tudomány, így története jól nyomon követhető eszközeinek, használt mérési eljárásainak és kutatási módszereinek fejlődésével. A magyar fonetika történetének – nemzetközi vonatkozásokat is tartalmazó – áttekintését adja [Vértes O.(1980)], [Vértes O.(1982c)], [Gósy & Olaszky(1985)] és [Gósy(2004):312–327]. Az elmúlt században két olyan új eszközzel is gazdagodott a fonetika „fegyvertára”, amely forradalmi változást jelentett e tudomány életében: ez a hangspektrográf és a számítógép.

Az 1950-es években elterjedt hangspektrográf [Koenig et al.(1946)], [vö. Kent & Read(2002):57–61] jelentősen megváltoztatta a 19. század végétől induló kísérleti fonetikát.<sup>1</sup> A hangspektrográf az addig használt idő-amplitúdó viszonyt megjelenítő oscillogramokhoz képest jelentős előrelépést jelentett, mert a papíron megjelenő regisztrátumában – a spektrogramon – a beszéd három paraméterének változásai is leolvashatók voltak: az idő, a frekvencia és az amplitúdó. A hetvenes és nyolcvanas években megjelenő hazai és nemzetközi publikációk többnyire ezekre az ábrákra támaszkodtak [Gósy(2004):302].

---

<sup>1</sup> A kísérleti fonetika kezdete a francia Jean Pierre Rousselot abbé nevéhez fűződik (Principes de Phonétique Expérimentale, 1897–1901)



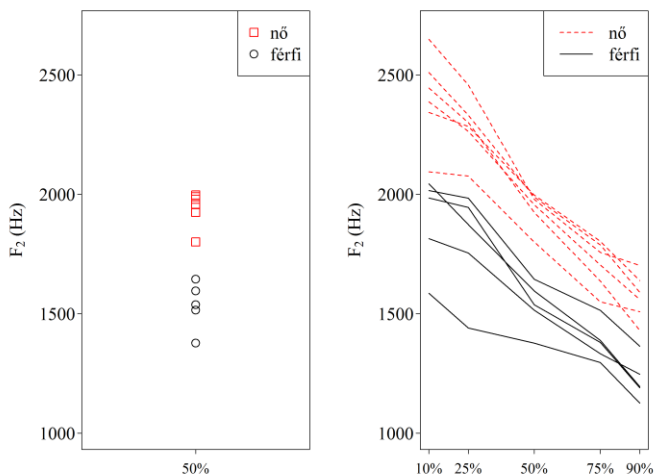
Az 1970-es évektől egyre jobban előtérbe kerülő számítógépek alapjaiban alakították át a használt mérési eljárásokat és kutatási módszereket. Lehetővé tették a beszédjel minden eddiginél kényelmesebb vizualizálását és sokrétű feldolgozását. Ezzel párhuzamosan kialakult két új diszciplína (1) a digitális gépi beszédfeldolgozás, amely a digitális jelfeldolgozás eredményeire is támaszkodva a természetes beszédlánc komponenseinek gépi vizsgálatát és modellezését tűzte ki célul, valamint (2) a beszédtechnológia, amely a természetes beszédlánc egyes elemeinek helyettesítésére, azok gépi megvalósítására törekedett [Németh & Olasz(2010):209].

Ma a fonetika egy harmadik, hatásában szintén forradalminak ígérkező változáson megy keresztül. A számítógépek teljesítményének és a háttérkapacitás további növekedésének köszönhetően, a fonetikai kutatásokban és a beszédtechnológiai alkalmazásokban a nagyméretű beszédatadátbázisok használata került előtérbe, szemben a kisméretű, speciális adatbázisok alkalmazásával. Így a beszédjelben lévő variáns és invariáns jegyek keresésére, illetve klasszikus fonetikai és fonológiai problémák vizsgálatára minden eddiginél szélesebb körben nyílik lehetőség a korpusz fonetika szemszögéből nézve. A beszédtechnológia oldaláról szemlélve pedig a beszéd- és beszélőfelismerő, valamint a gépi beszédet előállító programok jutnak rendezetten tárolt, nagy mennyiségű beszédmintához az algoritmusai számára. A nagy adatbázisok irányába való elmozdulás nemcsak a beszédtudományokra jellemző, hanem a tudományterületek többségére. Mára a legtöbb tudományterületnek ugyanis kialakult a számítógépes változata, amelyeket olyan nagy számítás igénylő feladatok jellemeznek, mint például a mintakeresés nagy adatbázisokban vagy nagyméretű szimulációk futtatása [Donoho et al.(2008)], [Peng(2011)].

A nagy adatbázisok használata több szempontból is kihívás elé állítja a kutatót. Egyrészt biztosítani kell, hogy állításai ellenőrizhetők legyenek, így szükség van a mérések átláthatóvá és megismételhetővé tételére. Erre a reprodukálható kutatás módszere adja meg a választ [Claerbout & Karrenbach(1992)]. Másrészt új elemző technikákra van szükség az egyre bonyolultabb adatstruktúrák leírására és elemzésére. Az elmúlt két évtizedben megjelent funkcionális és objektum-orientált statisztikai módszerek ennek a kihívásnak felelnek meg [Ramsay(2005)], [Wang & Marron(2007)]. Az előbbi görbék elemzését és összehasonlítását ígéri, utóbbi esetében pedig tetszőlegesen összetett objektumok vizsgálatára nyílik lehetőség.

A beszédjel időben változó akusztikus paraméterei a funkcionális adatelemzés számára megfelelő kiindulópontot jelentenek. Mára a mérőeszközök és a

számítógépes kapacitások fejlődésének következtében az alapfrekvencia, az intenzitás és a formánsfrekvencia, valamint különféle prozódiai jegyek egyszerűbben megragadhatók, időben akár több értékkel is. Ha ezeket a több időpillanathoz tartozó értékeket egyetlen megfigyelési egységnek tekintjük, akkor funkcionális adatról beszélünk (1.1. ábra). Egy jelenség jobb megértéséhez természetesen sok ilyen funkcionális adatra, nagy adatbázisra van szükségünk.



1.1. ábra. Az [n a: m] hangsor magánhangzójának második formánsa 11 bemondás alapján. A bal oldali ábrán bemondásonként egyetlen mérés áll rendelkezésre a magánhangzó közepéről, ezek skalár adatok. A jobb oldali ábrán 5 mérésből származó görbék jelentik a megfigyelési egységeket, ezek funkcionális adatok. Mindkét minta tehát 11 elemű

A fonetika egyik alapfeladata a hangok egymásra hatásának vizsgálata. A hangkapcsolatok fontosságát már a magyar polihisztor, Kempelen Farkas felfedezte beszélőgépezet készítése során, és leírta ezzel kapcsolatos – többnyire máig érvényes – gondolatait a Mechanismus der menschlichen Sprache c. művében 1791-ben [vö. Gósy(2010):37]. Azóta számos kísérlet történt a beszédhangok átmeneteinek leírására. Akusztikai szinten ez a kölcsönhatás leglátványosabban a formánsok mozgásában követhető nyomon. A kísérleti fonetika hagyományai szerint a magánhangzók formánsfrekvenciáit többnyire egy mérési pontban határozták meg (1.1. ábra bal oldali rész). Ennek előnyei és hátrányai is vannak. Azonban a magánhangzó akusztikai szerkezete a folyamatos beszédben dinamikusan, időben változik, így több mérési pont felvétele (1.1. ábra jobb oldali rész) pontosabbá teszi a hangátmenetek leírását.

Értekezésemben bemutatom egy nagyméretű beszédadatbázis és a hozzá kapcsolódó formánsadatbázis létrehozását, valamint az erre alapozott tudományos kutatás eredményeit. Maga a formánsadatbázis magánhangzónként öt mérési pontot tartalmaz, így alkalmas a hangátmenetek tanulmányozására, melyet a funkcionális adatelemzés módszerével végzek. Bemutatom azt a szoftverkörnyezetet is, amely biztosítja a vizsgálatok reprodukálható jellegét.

## 1.1. Motiváció

A formánsadatok vizsgálatával töltött néhány év tapasztalata alapján elmondhatom, hogy a számítástechnika gyors ütemű fejlődése nem csupán lehetővé teszi nagy mennyiségű mérések elvégzését és tárolását, hanem a megváltozott, új technológiai környezet jótékony módon ösztönzi, motiválja is erre a kutatót. Megfelelő mennyiségű hangfelvétel birtokában a beszédadatbázisok létrehozása és lekérdezése mára minden kutató számára elérhetővé vált, némi programozási ismeret birtokában pedig tömeges formánsmérések eredményeivel bővíthetjük ki adatbázisunkat. A szabadon elérhető eszközök eléggé általánosak és rugalmasak ahhoz, hogy bátorítást nyerjünk az első lépések megtételéhez azon az úton, amely rendkívül idő- és munkaigényes, valamint sok kintartást és erőt kíván.

Ha a kísérleti fonetikai vizsgálatok széles körét a magánhangzók formánsmozgásaira szűkítem, akkor Vértes O. András szavait kell felidézniem, aki az *agyagya* logatom második [ɔ] magánhangzóját elemezte férfi beszélő ejtésében, és a második formánsban a tiszta fázisra jellemző 1000 Hz-es érték előtt és után 1200 Hz-es és a fölötti részt talált. Ez alapján írja: „Az említett jelenség nem kivétel. Éppen ezért kérdéses, hogy mennyire jogos a magánhangzó közepére korlátozni a méréseket.” [Vértes O.(1982b):78]. Ugyanebben a tanulmányában Vértes O. András megállapítja, hogy „a különféle hangkapcsolatok formánstrukturája kevesebb átfedést mutat, mint az egyes hangoké” [Vértes O.(1982b):72]. A fenti gondolatokkal összhangban, a magánhangzók formánsszerkezetének több mérési pontos modellezésére teszek kísérletet.

## 1.2. Az értekezés felépítése

Az értekezés két részre osztható: a korábbi kutatások áttekintésére, majd a saját kutatásaim tárgyalására. A 2. fejezetben áttekintem a hazai és nemzetközi szakirodalmat. A kutatásaimhoz használt anyagokat és módszereket a 3. fejezetben tárgyalom. Itt részletezem a forrás- és vizsgálati beszédadatbázis

jellemzőit és a kísérleti személyek adatait. A 4. fejezetben bemutatom azt az általam kidolgozott modellt, amellyel a beszéd kutatás területén megfelelhünk a reprodukálható kutatás feltételeinek. Az 5. fejezetben a formánsadatbázis létrehozására kialakított eljárásom alkalmazását ismertetem. A 6. fejezetben formánsmenetek statisztikai vizsgálatából kapott eredményeimet és a formánsmenetek jellemzésére kialakított modell felépítését mutatom be. A 7. és 8. fejezet a magyar és angol nyelvű összefoglaló.

### **1.3. Tézisek**

A disszertáció 4 téziscsoportot és összesen 8 tézist tartalmaz.

**I. téziscsoport:** A megismételhető kutatás szempontjainak megfelelő saját beszédadatbázis létrehozása.

I.1. tézis: Önmagában ismert eszközök és eljárások újszerű kombinációjával kidolgoztam egy olyan modellt, amely a beszédtudomány néhány területén a reprodukálható kutatás alapja lehet.

I.2. tézis: A reprodukálható kutatás I.1. tézisben kidolgozott szempontjainak érvényesítésével nagyméretű beszédadatbázist hoztam létre, amely formánsadatok tárolására is alkalmas. Eljárást dolgoztam ki, amellyel a korábbi kutatások beszédadatbázisát az új adatbázis-szerkezetbe konvertáltam.

**II. téziscsoport.** Algoritmus a gépi formánsmérés hibáinak detektálására és korrekciójára.

II.1. tézis: Eljárást dolgoztam ki nagy tömegű, automatikusan meghatározott nyers formánsadat hibáinak gépi felismerésére.

II.2. tézis: Modellt dolgoztam ki hibás formánsadatok gépi és manuális hibajavítására, illetve az automatikusan javított formánsadatok kézi ellenőrzésére.

**III. téziscsoport.** A magyar magánhangzók formánsterének modellezése és statisztikai vizsgálata.

III.1. tézis: Megmutattam, hogy a funkcionális adatelemzés módszere alkalmas a formánsmozgások modellezésére és összehasonlítására a normalizált formánstérben. Adaptáltam és továbbfejlesztettem a modell értékelésére eddig használt grafikus ábrázolásokat.

III.2. A funkcionális adatelemzés segítségével igazoltam, hogy a magánhangzót követő 2. hang is befolyásolhatja a magánhangzóban létrejövő formánsmozgást.

**IV. téziscsoport.** A III. 1. tézisben szereplő modell mint új fonetikai mérőeszköz bemutatása.

IV.1. tézis: A formánsmenetek tanulmányozásához új vizsgálati eszközöket alakítottam ki, és ezek segítségével teljes körű áttekintést adtam a CVC kapcsolatok magánhangzóiban realizálódó formánsmozgásokra, a magyar nyelvre.

IV.2. tézis: Megoldást dolgoztam ki a formánsmozgások jóslására a magánhangzó és környezete ismeretében, amely lehetővé teszi a formánsmenetek becslését mérések végzése nélkül.

## 2. HÁTTÉR, IRODALMI ÁTTEKINTÉS

Disszertációmban a magyar beszédhangok hangzásbeli tulajdonságait tanulmányozom a folyamatos beszédben, azaz az elemzéseket az akusztikai fonetika keretein belül végzem. Vizsgálatom tárgya a magyar magánhangzó realizációkban előforduló formánsfrekvencia értékek változása az idő és a környező beszédhangok függvényében. A beszédhangok akusztikus jellemzőinek vizsgálata nagy hagyományokkal rendelkezik a magyar nyelvtudományban. A kísérleti fonetika, a beszédtechnológia és a laboratóriumi fonológia kutatói nagymértékben hozzájárultak ahhoz, hogy az artikuláció és az akusztikum közötti kapcsolat sok részlete mára ismertté vált.

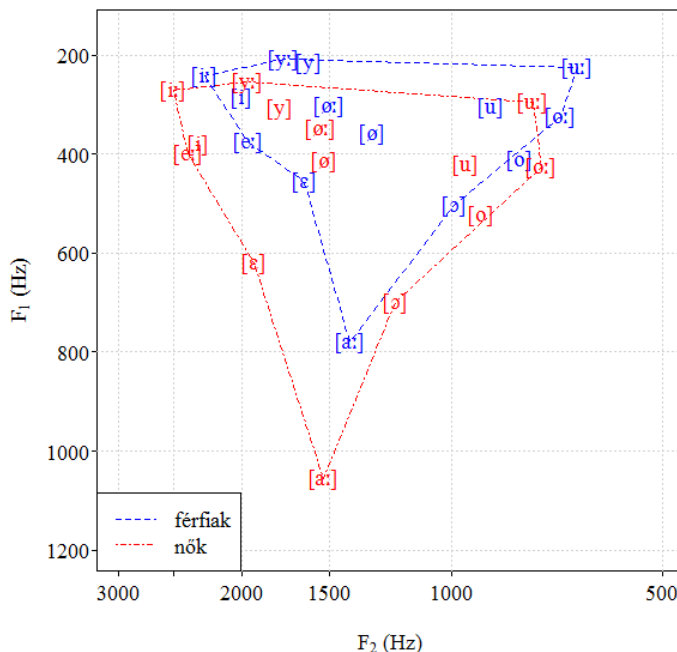
### 2.1. A magánhangzók spektrális jellemzői: statikus és dinamikus szemlélet

Korán felismerték, hogy a magánhangzó (magánhangzó jelölése V, mássalhangzó jelölése C) teljes időtartamában a spektrális tulajdonságok változáson mennek keresztül [Joos(1948)]. Mégis az ezt követő évtizedekben a magánhangzók spektrális leírására főképp statikus modellt használtak. E modell szerint – még ha expliciten nem is volt megfogalmazva – a magánhangzó-minőség teljes leírásához szükséges információ rendelkezésre áll a magánhangzó már egy viszonylag állandó (steady-state) részében, a tiszta fázisban. [Peterson & Barney(1952)] /hVd/ szótagok magánhangzóinak első három formánsát mérte meg. A formánsmérés „steady-state” helyét így határozta meg: „a magánhangzó azon része, amely a [h] hatását követi, a [d] hatását megelőzi és időtartama egy egyensúlyi állapotra terjed ki”<sup>2</sup>. A statikus spektrális modell használatát az alapozza meg, hogy a magánhangzó tiszta fázisa (egyensúlyi helye) megfelel a magánhangzóra jellemző idealizált artikulációnak, illetve annak akusztikus mintázatának, a célkonfigurációnak (vowel target). A tiszta fázis a magánhangzónak azon pontja vagy intervalluma, amely a spektrális változások legkisebb fokát mutatja, a legkevésbé befolyásolja a kontextuális hatás, tehát a leginkább a magánhangzóra jellemző rész. A tiszta fázis praktikusán a magánhangzó temporális középpontjában foglal helyet, de több próbálkozás ellenére sincs szabványos módszer a pontos meghatározására [Harrington(2010a):84]. A statikus spektrális modell grafikus ábrázolási módja a

---

<sup>2</sup> ‘Eredeti: „...a part of the vowel following the influence of the [h] and preceding the influence of the [d] during which a practically steady state is reached..” (177. oldal)

magánhangzó pontszerű megjelenítése az  $F_1 \times F_2$  síkon ( $F_1$ ,  $F_2$  definíciója a 2.2. fejezetben). Először [Essner(1947)] és [Joos(1948)] használta a magánhangzó-minőségek összehasonlítására, azóta szabványossá vált és tanulmányok egész sora vette át ezt a megjelenítési módot (2.1. ábra). Statikus modellt a jövőben is fognak használni: a mérések végrehajtása, javítása egyszerűbb, az ilyen adatok kezelése kényelmesebb.



2.1. ábra. [Magdics(1965):7] átlagos formánsadatai a hangsúlytalan magánhangzókra

A magánhangzó spektrális leírásának dinamikus modellje szerint a magánhangzó azonosításához legalább annyi információt szerezhetünk a hangátmenetekből, mint a tiszta fázisból [Harrington(2010a):87–89], [Strange & Jenkins(2013)]. Tanulmányok sora bizonyította, hogy sok esetben nem elegendő egyetlen időszelvény spektrális információját feldolgozni, a hangátmenetek ugyanúgy hozzájárulhatnak a magánhangzó felismeréséhez, mint a tiszta fázis.

Jenkins és munkatársai [Jenkins et al.(1983)] /bVb/ szótagok hangfelvételeit manipulálták: a magánhangzó közepét elnémították, kivágták, zajjal helyettesítették, illetve csak a magánhangzó közepét, induló vagy befejező fázisát hagyták meg. Az eredeti szótag és a néma maggal rendelkező szótag felismerése a percepciókísérletben egyaránt 93% volt. A csak a központi maggal rendelkező szótagnál ez lecsökkent 87%-ra, zajjal kitöltött szótagra 81%, illetve kivágott, kezdő és befejező szótagnál 70, 54 és 53% volt a felismerési

arány. A magánhangzó dinamikus természetét így fogalmazták meg: „E nézet szerint a mássalhangzó-magánhangzó koartikuláció nem «zaj» bevitelét jelenti az akusztikus jelbe. Éppen ellenkezőleg, a szótag fonémáinak koartikulációja egy olyan akusztikus elrendezést hoz létre, amelyben a magánhangzók és a mássalhangzók közösen vannak meghatározva az időben változó spektrális konfigurációban.”<sup>3</sup>.

Korábban Strange és munkatársai [Strange et al.(1976)] azt találták egy percepciókísérletben, hogy pontosabban azonosítható a magánhangzó CVC-ből, mint izolált V-ből. Hillenbrand és Nearey [Hillenbrand & Nearey(1999)] statikus spektrális tartalmú és formánsváltozásokat is tartalmazó szintetizált magánhangzók felismerését hasonlította össze percepciókísérletben. A felismerés hibája nagyobb volt statikus magánhangzók esetében.

Számos akusztikus kísérlet igazolja, hogy a magánhangzó osztályozásában az átmeneteknek is fontos szerep jut. [Suomi(1987)] és [Sussman(1990)] megmutatták, hogy ha a tiszta fázis helyett az átmeneti fázisból származó spektrális információt használjuk a magánhangzók osztályozására, akkor az eredmény nem változik. Ha pedig a tiszta fázis mellett az átmenetből is használunk információt, akkor a szétválasztás sikeresebb [Hillenbrand et al.(2001)], [Huang(1992)], [Zahorian & Jagharghi(1993)].

Mi okozza a magánhangzók dinamikus spektrális tulajdonságát? A legfontosabb ok a koartikuláció, amely a szomszédos beszédhangok egymásra hatása az artikulációban. A fonetikai koartikuláció azt jelenti, hogy a szomszédos beszédhangok valamilyen mértékben hatással vannak egymás artikulációjára függetlenül attól, hogy a koartikuláció következtében létrejövő módosult artikulációs konfiguráció definiál-e egy beszédhangot vagy nem [Gósy(2004):150]. Azonban nemcsak a kontextus, hanem maga a magánhangzó is lehet a spektrális változás forrása. A *magánhangzóhoz tartozó spektrális változás* (vowel inherent spectral change, VISC) fogalma először [Nearey & Assmann(1986)] munkájában jelent meg. A VISC értelmezése megfelel az eddig tárgyalt dinamikus spektrális tulajdonságnak, de annak okát magában a magánhangzóban keresi. Az elmélet szerint a VISC nemcsak a diftongusok és a triftongusok egyéni sajátja, hanem a monoftongusokat is jellemzi. A témakör fontosságát és népszerűségét jelzi, hogy külön tanulmánykötetet publikáltak e témában [Morrison & Assmann(2013)].

---

<sup>3</sup> Eredeti: „According to this view, coarticulation of consonants and vowels is not to be considered as the introduction of noise’ in the acoustic signal. On the contrary, the act of coarticulating phonemes in syllables gives rise to an acoustic array in which the consonants and vowels are cospecified in the time-varying spectral configuration” (449. oldal)



## 2.2. A formánsok

A spektrális információk vizsgálatát ebben az értekezésben a formánsfrekvencia értékekre korlátozom. Ugyan többen támogatják a teljes-spektrum megközelítést (vagy a formánsok amplitúdójának és sáv szélességének is a figyelembe vételét), tanulmányok sora igazolja, hogy kétségtelenül a formánsfrekvenciák biztosítják a legmarkánsabb információt a magánhangzó felismeréséhez, mind az akusztikus osztályozásban, mind a percepciók kísérletekben (részletesebben [Harrington(2010a):89]).

A magánhangzók vagy magánhangzószerű mássalhangzók képzése során a tüdőből kiáramló levegő a hangszalagokat „megrezegteti”, létrejön a zöngé. Ez a hangforrás szerepét tölti be a beszédképzés forrás-szűrő elméletében [Fant(1960)]. A modellben a szűrőt a gége feletti üregrendszer rezonancia tulajdonságai képviselik, ugyanis az alaphangból és felhangokból álló zöngé a toldalékcsőben az adott képzési konfigurációtól függően módosul. A toldalékcső üregei a rezonanciafrekvenciáiknak megfelelő vagy ahhoz közel eső felharmonikusokat a zöngében felerősítik, így energiakonzentrációk keletkeznek, amelyeket a fonetikai szakirodalom formánsoknak nevez [Gósy(2004):99], [Németh & Olasz(2010):51], [Subosits(2005):148]. Ez a definíció az akusztikumot helyezi előtérbe, így összhangban van [Fant(1960)] azon meghatározásával, mely szerint a formánsokat a spektrum lokális maximumai vagy későbbi megfogalmazásban a spektrum burkolójának lokális maximumai adják [vö. Gordos & Takács(1983):56]. A maximumhelyek frekvenciái a formánsfrekvenciák, amelyeket  $F_1, F_2, \dots, F_N$  betűkkel jelöli a szakirodalom. (A formánsstruktúrára még a sáv szélesség  $/B_1, B_2, \dots, B_N/$  és az amplitúdó  $/A_1, A_2, \dots, A_N/$  is jellemző.) A formánsok a rezonanciafrekvenciák becsléseinek foghatók fel [Stevens & House(1961)] és [Kent & Read(2002)]. A rezonanciafrekvencia az elmúlt évtizedben vonzó alternatívája lett a formánsnak, mert a rezonanciafrekvencia nem tűnik el, nem egyesül és nem válik szét, a beszéd során folyamatosan, még a szünetekben is jelen van [Deng et al.(2004)].

A toldalékcső alakja és a formánsok között ma ismert kapcsolatokat [Harrington(2010a)] foglalja össze. A magánhangzó időben változó spektrális leírására irányuló nemzetközi eredményeket [Strange & Jenkins(2013)] áttekintésében olvashatjuk.

## 2.3. Formánsmenet modellek

A formánsmenetek vizsgálatának előtérbe kerülése számos módszert hívott életre, de máig nincs konszenzus a kvantifikáció módjában és a statisztikai

elemzésükben. A formánsok mozgását leíró matematikai modellekre csak a nemzetközi szakirodalomban találtunk példákat. A magánhangzóban realizálódó mozgásokat [Lindblom(1963)] exponenciális függvényvel ragadja meg, [Stevens et al.(1966)] egy-egy másodfokú görbével írja le a két hangátmeneti részt, de van kísérlet a mozgási formák jellemzésére egyetlen parabolával [van Bergem(1993)], vagy diszkrét koszinusz-transzformációval [Watson & Harrington(1999)] is. A mássalhangzó-magánhangzó találkozások formánsváltozásainak könnyen implementálható megközelítése a locus egyenlet [Krull(1987)], [vö. Csapó & Németh(2009)], melyet főképp felpattanó zárhangok és magánhangzók átmeneteinek vizsgálatához alkalmaznak. A locus egyenlet a koartikulációs hatásokra két mérési pont alapján következtet. A módszer egyfajta általánosítását [Broad & Clermont(2010)] adja közre. Az általuk definiált lineáris modellben több mérési pont adata szerepelhet, például idézett tanulmányukban 11, egyenletesen elosztott pontot használnak a magánhangzó teljes hangidőtartamából a második formáns reprezentálásra.

## 2.4. A magyar beszédhangok

A disszertációban vizsgált beszédhangokat és azok jelöléseit a 2.1., 2.2. és 2.3. táblázat tartalmazza.

2.1. táblázat. A magyar magánhangzók betű- és hangszimbólumai

Betű	a	á	e	é	i	í	o	ó	ö	ő	u	ú	ü	ű
IPA-jel	[ɔ]	[aː]	[ɛ]	[eː]	[i]	[iː]	[o]	[oː]	[ø]	[øː]	[u]	[uː]	[y]	[yː]
SAMPA-jel	[O]	[aː]	[E]	[eː]	[i]	[iː]	[o]	[oː]	[2]	[2ː]	[u]	[uː]	[y]	[yː]

2.2. táblázat. A magyar mássalhangzók betű- és hangszimbólumai

Betű	b, bb	p, pp	d, dd	t, tt	gy, ggy	ty, tty	g, gg	k, kk	m, mm
IPA-jel	[b], [bː]	[p], [pː]	[d], [dː]	[t], [tː]	[j], [jː]	[c], [cː]	[g], [gː]	[k], [kː]	[m], [mː]
SAMPA-jel	[b], [bː]	[p], [pː]	[d], [dː]	[t], [tː]	[dʰ], [dʰː]	[tʰ], [tʰː]	[g], [gː]	[k], [kː]	[m], [mː]

Betű	n, nn	ny, nny	j, ly, jj, lly	h, hh	v, vv	f, ff	z, zz	sz, ssz	zs, zzs
IPA-jel	[n], [nː]	[ɲ], [ɲː]	[j], [jː]	[h], [hː]	[v], [vː]	[f], [fː]	[z], [zː]	[s], [sː]	[ʒ], [ʒː]
SAMPA-jel	[n], [nː]	[J], [Jː]	[j], [jː]	[h], [hː]	[v], [vː]	[f], [fː]	[z], [zː]	[s], [sː]	[Z], [Zː]

Betű	s, ss	dz, ddz	dzs, ddzs	c, cc	cs, ccs	l, ll	r, rr
IPA-jel	[ʃ], [ʃː]	[dʒ], [dʒː]	[dʒs], [dʒsː]	[tʃ], [tʃː]	[tʃs], [tʃsː]	[l], [lː]	[r], [rː]
SAMPA-jel	[S], [Sː]	[dz], [dzː]	[dZ], [dZː]	[ts], [tsː]	[tS], [tSː]	[l], [lː]	[r], [rː]

2.3. táblázat. A variánsokat jelölő betű- és hangszimbólumok

Betű	mv*, mf*	ng*, nk*	ch*, h
IPA-jel	[ŋ]	[ŋ]	[x]
SAMPA-jel	[F]	[N]	[x]
Példa	hamvas, kámfor	hangos, lankad	pech, doh

A 2.1. és 2.2. táblázatban a magyar köznyelv fonémaértékű (nyelvi funkcióval is rendelkező) hangjait soroltam fel. (Az újabb fonológiai rendszerezés nem ismeri el mindegyiket fonémának vö. [Siptár(2001)]). A beszédhangok jelölésére a magyar betűkkel jelölt formát, az IPA-szimbólumokat<sup>4</sup> és a SAMPA [Vicsi(1996)] jelölési rendszert<sup>5</sup> tartalmazzák a táblázatok. A beszédhangokat szögletes zárójelek határolják. A 2.1. és 2.2. táblázat alapkészletet kiterjesztettem a 2.3. táblázat jelöléseivel.

A kiterjesztett szimbólumok használatát az indokolja, hogy a hangok olyan variánsairól van szó, amelyek képzési helye eltolódik.

## 2.5. A magyar formánskutatás

A magyar beszédhangok formánsainak vizsgálata több mint 70 éves múltra tekint vissza, szakirodalma rendkívül gazdag. A magyar magánhangzók kísérletes vizsgálata még ennél is korábban kezdődött, Kohn Gyula már 1874-ben elvégezte vizsgálatát az akusztikai szerkezetük feltérképezésére, Gáti Béla pedig 1921-ben elsőként valamennyi magyar magánhangzó rezgésképét közreadja [Gósy & Olaszy(1985)]. A magyar magánhangzók spektrális vizsgálatát Tarnóczy Tamás nyitja meg, aki erősen zajcsökkentett környezetben, egyenletesen mozgó filmszalagra rögzítette a hanghullámot, majd ezen görbék Fourier-együtthatóit és formánshelyeit határozta meg [Tarnóczy(1941)]. Később, 4 női és 4 férfi beszélő hangmintái alapján formánsfrekvencia értékeket közöl, az adatokat  $F_1 \times F_2$  síkon is megjeleníti.

A magyar magánhangzók formánsszerkezetére vonatkozó első átfogó elemzést [Magdics(1965)] adta közre. Ennek egy részletét a 2.1. ábra mutatja be. Vizsgálatában 26 mássalhangzó és 8 magánhangzó összes CV és VC hangkapcsolódásának frekvenciaszerkezetét írja le. A CV hangpárok  $F_2$  és  $F_3$  formánsmozgásait ábrán követhetjük nyomon, de mindkét hangkapcsolat esetén

<sup>4</sup> Hangjelölő rendszer, amelyet a Nemzetközi Fonetikai Társaság (International Phonetic Association) hozott létre.

<sup>5</sup> A SAMPA (Speech Assessment Methods Phonetic Alphabet) hangjelölési rendszere az IPA nemzetközi fonetikus ábécé leírását könnyíti meg számítógépes környezetben.

(CV és VC) a formánsmozgás induló és cél frekvenciaértékeire számszerű adatokat is találunk. Megmutatta, hogy attól függően, hogy a szomszédos beszédhangok artikulációja mennyire tér el egymástól, a formánsfrekvencia értékek az időben kisebb-nagyobb mértékben változnak a magánhangzó képzési idején belül, azaz a formáns mozog a magánhangzón belül.

[Bolla(1978)] szintén teljes körű elemzést adott a magyar magánhangzók tiszta fázisát jellemző formánsokat illetően egy férfi és egy nő ejtésében, vizsgálataihoz beszédszintézist is használt.

Az artikuláció és az akusztikum közötti kapcsolatot foglalja össze [Vértes O.(1982a)], ugyanebben a kötetben áttekintést ad az addig folytatott magyar kutatások eredményeiről [Vértes O.(1982b)].

[Olaszy(1985)] a CV és VC kapcsolatokat vizsgálta analízis-formáns szintézis módszerrel, melyhez szóadatbázist készített (407 elem egyetlen férfi bemondóval). Törtvonalas közelítéssel megadta az  $F_1$ ,  $F_2$ ,  $F_3$  formáns mozgását a CV és a VC diád magánhangzójában 23 C-re és 9 V-re. Minden hangkapcsolódásra találunk adatot, tehát a teljes magyar hangkapcsolódási teret lefedi. A formánsmozgásokat számadatokkal írja le, amelyek korrektségét szintetizált szavakon igazolta.

A hangátmeneti részben bekövetkező formánsmozgás percepcióban betöltött szerepét [Gósy(1995)] vizsgálta CV hangkapcsolatokban. A hangátmenettől megfosztott hangkapcsolat mássalhangzójának felismerhetőségével kapcsolatos kísérlete alapján megállapította, hogy a hangátmenetek elsődleges vagy másodlagos akusztikus kulcsok az észlelésben, de sok esetben redundáns tényezőként funkcionálnak. Szerepük függ a mássalhangzók minőségétől és a kontextustól.

A magyar beszédhangok artikulációs és akusztikus sajátosságait tablószerűen mutatja be [Molnár(1970)], [Bolla(1995)].

[Szalai(1995)] három magánhangzó, az [u; a; i] formáns szerkezetét vizsgálja CVC hangkapcsolatokba ágyazva egyetlen férfi bemondó hangmintái alapján. A magánhangzók tiszta fázisában mért formánsértékeket tanulmányozza, az átmenetek kérdéskörével nem foglalkozik. A szerző közli az első három formánsra az átlagokat és a szélsőértékeket mindhárom magánhangzó esetén, de ezen túlmenően az  $F_1 \times F_2$  síkon is ábrázolja a különböző hangkörnyezetekben megvalósuló mérési adatokat magánhangzónként. Ez a koartikulációs mező, amelynek bemutatásával a mássalhangzó környezet hatását  $F_1$  és  $F_2$  viszonyrendszerében együtt tanulmányozhatjuk. Ehhez hasonlóan,

[Kovács(1998)] sűrűsödési ellipsziseket adott meg magyar magánhangzókra, melyek 50%-os valószínűséggel tartalmazzák a formánsadatokat.

Gósy Mária ([Gósy(2004):116]) a köznyelvi magyar magánhangzók első három formánsának jellemző középértékeit közli.

[Nikléczy(2006)] megvizsgálta az [ɔ ε] magánhangzók tiszta fázisában és az átmeneti részekben bekövetkező frekvenciaingadozásokat. Eredményét elsősorban nem a formánskontúrra, hanem részben az átlagos energiaspektrumra, részben a formánsadatokra (a 2 szélsőértékekre és a tiszta fázis adatára) alapozza.

[Grácsi & Horváth(2010)] a magyar magánhangzók első két formánsának realizációját elemzik spontán beszédben és kimutatják a vokálisok semlegesedését. Az átlagok mellett átlagos eltérést és határértéket is közölnek.

[Gósy(2010)] magánhangzók diftongizálódásának lehetőségét vizsgálta VC kapcsolatokban. A magánhangzó tiszta fázisából és a hangátmeneti részből származó két méréssel elemezte a koartikuláció következtében létrejött spektrális módosulásokat.

A fenti felsorolásba azokat a szakirodalmakat emeltem be, amelyek vagy átfogó módon tárgyalják a vokálisok spektrális tulajdonságait, vagy többpontos méréssel lépéseket tesznek a spektrális változás időbeli leírására.

## **2.6. Reprodukálható kutatás**

A reprodukálható kutatás (reproducible research) fogalma kétértelmű, megfelelő szövegkörnyezet nélkül félrevezető lehet. Tágabb értelemben a tudományos állítások megítélésének alapvető szabványa, amely a teljes kutatás független megismétlését (replikáció, [replication]) ígéri. Az elméleti tudományokban (például matematika) az eredmények teljes körű reprodukálása bárki számára elvégezhető a tétel mellett szereplő bizonyítás segítségével. A kísérleti tudományokban (például fizika, biológia vagy élettudományok) a hipotézis megfogalmazása és a kísérlet pontos leírása ad lehetőséget a publikált eredmények értékelésére. Ebben az esetben más kutatók a hipotézis érvényességét úgy ellenőrizhetik, hogy függetlenül gyűjtött adatok segítségével replikálják a kísérletet. Azonban a teljes kutatás független megismétlésére sokszor nincs lehetőségünk megfelelő erőforrás (például adatbázis nagysága, számítási teljesítmény) hiányában vagy más okok miatt (a kísérletek egy része sok időt és pénzt emészt fel). Szűkebb értelemben a reprodukálható kutatás azoknak a módszereknek az összessége, amely a kutatási eredmények mellett a felhasznált adatok és forráskódok publikálását is lehetővé teszi. Így más kutatók

a publikált adatokon a publikált kód futtatásával ellenőrizhetik az állítások érvényességét, így a reprodukálható kutatás egyfajta minimum szabvány szerepét tölti be a tudományos állítások megítélésében [Peng(2011)]. A reprodukálható kutatás fogalma – ebben az értelemben – Jon Claerbout-tól és munkatársától származik [Claerbout & Karrenbach(1992)]. Az értekezésben ebben az értelemben használom a reprodukálható kutatás fogalmát.

A reprodukálható kutatás különösen jelentős szerepet kap a nagyméretű adatbázisok előtérbe kerülésével. A tudományos módszertanok elméleti és empirikus ága mellett egy harmadik megközelítés – a számítástudomány – megjelenését idézte elő a digitális infrastruktúra gyors fejlődése [Donoho et al.(2008)]. Fel kellett ismerni azonban, hogy az e területről származó kutatási eredmények csak korlátozottan ellenőrizhetők a tudományos publikációk alapján, ugyanis a használt programkódok, azok bemenő paramétereinek és az adatbázisok tartalmának hiánya nem teszi lehetővé a kutatás megismétlését. A bonyolult elemzések során pedig megnő a nem szándékos hibák valószínűsége, amelyek hibás összefüggésekhez vezethetnek. Az új algoritmusok fejlesztését bemutató vagy a programok futási outputjára támaszkodó publikációk esetében az olvasók tehát igénylik a számítógépes kód megismerését, megosztását, mert az minden természetes nyelvű leírásnál részletesebben mutatja be (és egyben igazolja) az eredményeket. Azt mondhatjuk tehát, hogy a számítástudomány módszertanával kapott eredmények korlátozott szavahihetőségére szintén a reprodukálható kutatás nyújtja a megoldást [Donoho et al.(2008)], [Peng(2011)].

### **2.6.1. Út a reprodukálható kutatáshoz**

A reprodukálható kutatás elképzeléséhez nagyon közeli gondolatokat fogalmaz meg a literális programozás [Knuth(1984)]. Már itt megjelenik a forráskód és a (publikálandó) szöveg egymásba illesztése, amelynek célja Knuth szerint a programok lehető legjobb dokumentálása volt. A Knuth által kifejlesztett WEB lehetővé teszi, hogy a Pascal nyelv utasításait és a TeX formázó rendszer parancsait egyetlen állományban felhasználhassuk. Ebből az állományból a *weaving* eljárással a dokumentáció TeX szövegét, a *tangling* eljárással a Pascal forrásnyelvű szöveget tudjuk kinyerni. Később elkészült a WEB C nyelvre épülő verziója a CWEB [Knuth & Levy(1994)], ami kiegészült a C++ és Java programok dokumentálásának lehetőségével is. Célnyelvtől független dokumentációt tesz lehetővé a noweb rendszer [Johnson & Johnson(1997)], amely nagymértékben leegyszerűsítette a literális programozás alkalmazását. Azonban a literális programozás önmagában nem támogatja a reprodukálható kutatást, mivel a kódblokkok végrehajtása a dokumentumon belül nem

lehetséges. Forráskód végrehajtására csak a *tangling* folyamat végén, külön lépésben, fordító vagy interpreter futtatásával van lehetőségünk, de ennek a *weaving*-gel kapott szövegre nincs semmilyen hatása.

A reprodukálható kutatás (reproducible research) kifejezés először Claerbout és munkatársa munkájában jelent meg [Claerbout & Karrenbach(1992)]. Geofizikai kutatásukban vezettek be egy új eljárást, amellyel az elektronikus dokumentumok ábráit, a Makefile-hoz hasonló szisztema szerint, egyetlen gombnyomással reprodukálni tudták. Ezt a gondolatot vitte tovább Fomel és munkatársa [Fomel & Hennenfent(2007)] a nyílt forráskódú Madagascar szoftvercsomagban. A Makefile-hoz hasonló eszközzel (SCons) egyszerűbb és több platformról elérhető reprodukálhatóságot értek el. A wavelet-ek kutatáshoz új Matlab csomagot (WaveLab toolbox) fejlesztett Buckheit és Donoho [Buckheit & Donoho(1995)], amellyel a cikk mellett adatok, Matlab kódok (a cikkben megjelenő táblázatok és képek generálása) és dokumentációk publikálására is lehetőségünk van. A Madagascar és a WaveLab a kódot a cikk szövegétől elkülönítve tárolja. Az Sweave [Leisch(2002)] rendszerben a literális programozás elveinek megfelelően már a programkódok és a szövegek (R és LaTeX) egyetlen dokumentumban szerepelnek. Az Sweave ma a reprodukálható kutatás egyik legnépszerűbb eszköze, hatására számos hasonló eszköz jelent meg, amely a rendkívül erős R/LaTeX pár helyett más nyelveket támogat, például `odfWeave`<sup>6</sup> (R/ODF), `R2HTML`<sup>7</sup> (R/HTML), `R2wd`<sup>8</sup> (R/Word), `knitr`<sup>9</sup> (R/Markdown), `SASweave`<sup>10</sup> (SAS/LaTeX) és `StatWeave`<sup>11</sup> stb. A legnagyobb szabadságot ma az Emacs szövegszerkesztő Org-mode módja adja, amely teljes körűen támogatja a literális programozást és a reprodukálható kutatást is [Schulte et al.(2012)]. Egy egyszerű szöveges formanyelvbe az Emacs Babel kiterjesztésének köszönhetően 39 különböző nyelv utasításait kombinálhatjuk.

## 2.6.2. Mi a reprodukálható kutatás?

A reprodukálható kutatás szerint – korábbi definíciókat tovább részletezve – egy kutatás végeredménye a publikáció, valamint a számítógépes környezetében lévő összes olyan adat és forráskód, amely a kutatás eredményeinek

---

<sup>6</sup><http://cran.r-project.org/web/packages/odfWeave/index.html> [Kuhn(2012)]

<sup>7</sup><http://cran.r-project.org/web/packages/R2HTML/index.html> [Lecoutre(2013)]

<sup>8</sup><http://cran.r-project.org/web/packages/R2wd/index.html> [Ritter(2012)]

<sup>9</sup><http://cran.r-project.org/web/packages/knitr/index.html> [Xie(2013)]

<sup>10</sup><http://homepage.cs.uiowa.edu/~rlenth/SASweave/> [Lenth & Hjosgaard(2007)]

<sup>11</sup> Több nyelvet támogat (R, S-plus, SAS, Stata, Maple) és különböző formátumokat (LaTeX, ODT). Elérhető: <http://homepage.cs.uiowa.edu/~rlenth/StatWeave/> [Lenth(2010)]

reprodukálásához szükséges<sup>12</sup>. Ezen elemek összecsomagolásával a reprodukálható kutatás kompendiumát [Gentleman & Temple Lang(2004)] kapjuk. Ennek részei:

- a kutatás publikációja (PDF és TeX/Word formátumban, bibliográfiával)
- adatok (a nyers és tisztított adatok, kód az adattisztításról, részletes adatszótár)
- számítási környezet (jól dokumentált forráskód, paraméterek, beállítások és a futtató platform leírása)
- eredmények (ábrák, táblázatok, numerikus adatok és ezek leírása)
- kiegészítő anyagok (például videó és hang fájlok)

A fenti komponensek mindegyike egy vagy több különálló állomány eltárolását jelenti. Azonban számos szoftvercsomag létezik, amely a literális programozás elve alapján a fenti komponensek közül többet egyetlen „futtatható” állományba tud sűríteni (lásd fent például *Sweave*, *Org-mode*). Ezen eszközök folyamatosan bővülő teljes listáját megtaláljuk a *Reproducible Research Planet!* honlapján (<http://www.rrplanet.com>). A szöveg és kód kombinálása egyetlen állományban, a reprodukálható kutatásban más célokat szolgál, mint a literális programozás során. Itt a publikálásra szánt dokumentum generálása közben (*weaving* eljárás) a programkódok végrehajtásra kerülnek, melyek eredménye (ábra, táblázat, számszerű adat stb.) azonnal megjelenik a szövegben.

A reprodukálható kutatás publikálása kétségtelenül nagyobb erőfeszítést igényel a szerzőtől, azonban haszonnal jár mind a szerző, mind az olvasó számára. Az egyik legnagyobb előnyt a szerző számára Jon Claerbout fogalmazza meg<sup>13</sup>: „A reprodukálható kutatás fő tanai közül az egyik az, hogy az idő egy másik személlyé változtatja mindegyikünket. Miközben erőfeszítéseket teszünk, hogy idegenekkel kommunikáljunk, segítünk magunknak, hogy jövőbeli önmagunkkal kommunikálhassunk.”<sup>14</sup> Alapvető követelmény tehát, hogy saját kutatásainkat reprodukálni tudjunk. Az olvasó és a tudomány számára talán a legfontosabb előny a kutatás transzparens volta.

---

<sup>12</sup>Understand Reproducible Research (<http://www.rrplanet.com/reproducible-research/reproducible-research.shtml>) [RRplanet.com(2013)]

<sup>13</sup><http://sepwww.stanford.edu/OLDWWW/research/redoc/IRIS.html> [Claerbout(2005)]

<sup>14</sup>Angolul: „One of the main tenets of reproducible research is that time turns each one of us into another person. By making an effort to communicate with strangers; we help ourselves to communicate with our future selves.”



### **2.6.3. Fonetikai kutatások és a reprodukálhatóság**

A kutatás független megismétlésének – azaz a teljes replikációnak – a beszéd kutatás területén nagy hagyománya van. Ugyanazt a fonetikai jelenséget több kutatócsoport vizsgálta és vizsgálja ma is többé-kevésbé eltérő eszközökkel és megközelítésekkel. A replikáció mellett a reprodukálható kutatásnak is helye van a beszédtudományokban. Ez egyrészt a bevezetőben említett nagy beszédatadabázisok használatának és az egyre komplexebb számítások, algoritmusok használatának köszönhető. A reprodukálhatóság ekkor a nem szándékos hibák elkerülését, ill. az algoritmusok számos paraméterének alapos dokumentálását segíti. Másrészt a reprodukálható kutatás fontosságát kiemeli az a tény, hogy az elemzésekben nagyon gyakoriak a matematikai statisztika elemei. A publikációk sokszor tartalmazzak leírást a mintavételről, az adatok transzformációjáról, jellemzők és származtatott adatok kinyeréséről, továbbá közlik a statisztikai mutatókat, a táblázatokat, a változatosan megszerkesztett ábrákat és a hipotézisvizsgálatok eredményeit. Ezekben az esetekben a reprodukálhatóság az olvasó számára a jobb megértésen túl biztosíthatja saját adatbázis felhasználását, kicserélését a szerző adatbázisával, valamint a statisztikai próbákhoz vezető transzformációk módosítását. Az adatok és kódok változtatása után az olvasók ellenőrizhetik a hipotézisvizsgálatokban, a megrajzolt táblázatokban és az ábrákban bekövetkező változásokat. A szerző számára egy ilyen soklépéses eljárás reprodukálhatóvá tétele a jövőbeni biztos végrehajtást biztosítja.

## **2.7. Beszédatadabázisok kialakítása**

### **2.7.1. A beszédatadabázis fogalma**

A beszédatadabázis Vicsi Klára meghatározása szerint „nagy méretű, nyelvileg és akusztikailag feldolgozott, tárolt hangadathalmaz, magyarázó jegyzetekkel, címkézéssel és átirásokkal ellátva, amelyet sokféle csoportosítás szerint hozhatunk létre” [Vicsi(2010):261]. A definíció implicit módon tartalmazza, hogy a beszédatadabázist számítógépen, digitális formában tároljuk és számítógéppel dolgozzuk fel. Ezt a tulajdonságot expliciten is tartalmazza [Gibbon et al.(1997):79] meghatározása, amely szerint „a beszédatadabázis a beszédminták bármely olyan gyűjteménye, amely a számítógép számára feldolgozható, valamint elégséges annotációt és dokumentációt tartalmaz ahhoz, hogy számunkra vagy más kutatók száma az adatbázis újrafelhasználható

legyen”<sup>15</sup>. Közös a két meghatározásban, hogy kizárja az annotálatlan és a megfelelő dokumentációval nem ellátott hangfelvételeket a beszédadatbázisok köréből. A második definíció hangsúlyt fektet a beszédadatbázisok újrafelhasználhatóságára is, amely mind a beszédkutató, mind a beszédtechnológia céljait szolgáló adatbázisok esetén nagy jelentőséggel bír.

### **2.7.2. A beszédadatbázis részei**

A beszédadatbázis a következő primer (P) és szekunder (Sz) forrásokat tartalmazhatja [Gibbon et al.(1997):80]:

- P1. beszédjel (például akusztikus beszédjel, de palatográfiai, labiográfiai, glottográfiai vagy szubglottális jelek/adatok is alkothatják)
- P2. annotáció (címkék és markerek, például ortografikus, fonemikus és fonetikus átírás stb.)
- P3. metaadatok (például a beszélő életkora, neme; a beszéd típusa: izolált szavak, mondatok, beszélgetés; a felvétel körülményei)
- Sz1. elemzések eredményei (FFT adatok, LPC adatok, alaphfrekvencia adatok, formáns adatok, statisztikai modellek és szabályok adatai)

### **2.7.3. Az adatbázis és a beszédadatbázis viszonya**

A beszédadatbázis szó használata némileg megtévesztő lehet, ha a szűken értelmezett, az informatikai szakmában elterjedt adatbázis fogalmat értjük rajta. Az adatbázis pontos technikai meghatározása megtalálható például Halassy 2002-ben írt könyvében [Halassy(2002):35]. A beszédadatbázisok irodalmában az adatbázis fogalmát kötetlenebb módon, tágabb értelmezéssel közelítik. Fő szempont az adatok szervezett tárolása, azok kényelmes és gyors kezelése a gépi feldolgozás számára, de az adatok egyidejű, közös használata, különböző jogosultságokhoz kötött elérése jelen pillanatban nem követelmény. Ennek ellenére történtek kísérletek a beszédadatok tárolására relációs, objektum-orientált és dedukciós adatbázisokban is [Draxler(1995)], de ezek a megoldások széleskörűen nem terjedtek el. Ez nem jelenti azt, hogy a beszédminták bizonyos szekunder adatai – például a beszédhangok hossza vagy a formánsfrekvenciák értéke – ne kerülhetnének ilyen formában tárolásra. Számos nyilvánosan is hozzáférhető, elsősorban szekunder adatokat tartalmazó relációs adatbázis jelent meg az elmúlt években. Ezek interneten keresztül, webes keresőfelület

---

<sup>15</sup> Eredetiben: „*spoken language corpus* is »any collection of speech recordings which is accessible in computer readable form and which comes with annotation and documentation sufficient to allow re-use of data in-house, or by scientist in other organisations«”.

segítségével érhető el. Ebbe a sorba illeszkedik a magyar hangkapcsolatok akusztikai szerkezetének bemutatása [Abari & Olasz(2007)] (<http://magyarbeszed.tmit.bme.hu/cvvc> [Olasz & Abari(2011)]), és az első magyar formánsadatbázis [Abari & Olasz(2011c)] (<http://magyarbeszed.tmit.bme.hu/formans> [Abari et al.(2011)]).

## 2.8. Formánsadatbázisok kialakítása

A formánsadatbázis egy rendezetten tárolt adathalmaz, amely minimálisan az első néhány formánsfrekvencia értékét  $\{F_1, F_2, F_3, \dots\}$  és esetleg a sáv szélességeket  $\{B_1, B_2, B_3, \dots\}$  valamint az amplitúdókat  $\{A_1, A_2, A_3, \dots\}$  tartalmazza. Az igazán értékes formánsadatbázisok nem egyszerűen beszédhanghoz tartozó formánsadatokat tartalmaznak, hanem a beszélőről és a mért hang környezetéről is szolgáltatnak információt. A formánsadatbázisok fontos jellemzője lehet a mérési pontra vonatkozó időkoordináta megadása a hullámforma időtengelyén, így a közölt formánsértéket hozzá tudjuk kapcsolni a hullámformához, az adott beszédhanghoz, azon belül pedig a beszédhang belső szerkezetéhez, illetve címkézett bemondások esetén azok annotációjához. Ebben az esetben a formánsadatbázis, mint származtatott információ, a beszédatbázis szerves részének tekinthető, és sok esetben a beszédkorpusz létrehozásával egyidőben, a korpusz által meghatározott szerkezetben épül be az adatbázisba.

A formánsadatbázisok tárolására több lehetőség adódik. Bináris vagy szöveges állományban, azonos időközönként vagy meghatározott időpontokban mért adatokat tartalmazó formánsadatbázisokra is találunk példát. A formánsadatokat eltárolhatjuk bemondásonként külön állományban, vagy egyetlen nagy adatállományban is.

A publikusan elérhető formánsadatbázisok közé tartoznak a Peterson és Barney klasszikus tanulmányában megjelenő formánsadatok<sup>16</sup> [Peterson & Barney(1952)]. Összesen 76 személytől (férfiak, nők és gyerekek) tíz szót rögzítettek kétszeri ejtésben angol nyelven. A tíz szó mindegyike /hVd/ szerkezetű volt, így csak a magánhangzóban tértek el egymástól. A formánsadatbázis szöveges formátumban áll rendelkezésre, amely  $76 \times 2 \times 10 = 1520$  sort tartalmaz. Minden sor tartalmazza a bemondó és a magánhangzó azonosítóját, valamint az alaphangfrekvenciát és az első három formánsfrekvenciát, a magánhangzó tiszta fázisából származó mérés alapján.

---

<sup>16</sup> <http://solon.cma.univie.ac.at/statdat.html> [Neumaier(2010)]

[Hillenbrand et al.(1995)] jelentősen nagyobb mintán<sup>17</sup> megismételte Peterson és Barney kutatását, 139-re növelte a beszélők, 12-re a vizsgált magánhangzók számát. A formánsadatokat továbbra is tagolt szöveges állományban állnak rendelkezésre, de a formánsfrekvencia értékek a magánhangzó teljes időtartamát 100%-nak tekintve, 10%-tól 80%-ig 10%-os lépésközzel összesen nyolc időpontból származnak. Minden mérési pont időkoordinátája és maguk a bemondások is elérhetők az adatbázisban.

Japán magánhangzók formánsadatait<sup>18</sup> közli [Mokhtari & Tanaka(2000b)] öt férfi beszélő adatai alapján. Az első négy formáns frekvenciaértékét és sáv szélességét tagolt szöveges állományban tárolja a formánsadatbázis. A beszélők 22-szer ismétlik a mért magánhangzókat tartalmazó szavakat.

[Deng et al.(2006a)] szintén az első négy formáns frekvenciaértékét és sáv szélességét közli<sup>19</sup>, de az adatokat HTK formátumú [Young(2006)] bináris állományban rögzítik, amely 10 ms-os periódusokban vételezett adatokat tartalmaz. A formánsadatokat – az eddig említett formánsadatbázisoktól eltérően – bemondásonként külön állományban vannak eltárolva. Az adatbázis 197 beszélő összesen 536 mondatát tartalmazza.

[Abari & Olasz(2011c)] webes lekérdező felület használatával is elérhető formánsadatbázist ad közre, amelynek alapja egy férfi és egy nő szólista felolvasásából összeállított, összesen közel 3000 szavas beszédadatbázis. Minden magánhangzón belül 3 mérési ponthoz – a teljes hang időtartamának 25, 50 és 75%-os pontjához – tartozó első négy formáns frekvenciaértékét relációs adatbázisban rögzítik (<http://magyarbeszed.tmit.bme.hu/formans> [Abari et al.(2011)]).

## 2.9. Funkcionális adatelemzés

A funkcionális adatelemzés azoknak a statisztikai eljárásoknak a gyűjtőneve, amelyek a görbéket mintaelemeknek tekintik, és azokkal végeznek elemzéseket. Ez a terminus először [Ramsay & Dalzell(1991)] tanulmányában jelent meg. A funkcionális adatelemzési technika mára nagyon népszerű lett, sikerét elsősorban a [Ramsay & Silverman(2005)] (az első kiadás 1997-ben jelent meg), [Ramsay & Silverman(2002)], [Ramsay et al.(2009)] és [Ferraty & Vieu(2010)] monográfiáknak köszönheti.

---

<sup>17</sup> <http://homepages.wmich.edu/~hillenbr/voweldata.html> [Hillenbrand(1995)]

<sup>18</sup> <http://www2.nict.go.jp/univ-com/mcc/people/parham/formant/formant.html> [Mokhtari & Tanaka(2000a)]

<sup>19</sup> <http://research.microsoft.com/en-us/groups/srg/downloads.aspx> [Deng et al.(2006b)]

A funkcionális adatelemzés számára számos módszert dolgoztak ki az elmúlt több mint 20 évben, szakirodalmi rendkívül gazdag, a használt módszerek és megközelítések szempontjából különböző iskolák alakultak ki [Ren(2005):3].

A funkcionális adatelemzés rokonságot mutat a sztochasztikus statisztikával (idősorok analízise, térstatisztika) és a longitudinális elemzésekkel is. Azonban ez utóbbiak inkább a paraméteres statisztikához tartoznak, a funkcionális adatelemzés nem-paraméteres természetű. Más jegyekben is eltér a modern statisztika ezen három ága. A funkcionális adatelemzés ugyanis alapvetően különbözik a hagyományos egy- és többváltozós statisztikától. A mintaelemek nem skalárok vagy vektorok, hanem végtelen dimenziós függvények. A gyakorlatban nem magukat a függvényeket, hanem azok diszkrét pontjait figyeljük meg, így az elemzés első lépéseként az adatpontokból fel kell építenünk magát a függvényt. Ramsay és Silverman (2002, 2005) által képviselt iskola számos simító technikát biztosít, amely a diszkrét megfigyelt adatokat valódi függvényre konvertálja.

### 2.9.1. A funkcionális adatok leírása: simítás

Egy  $N$  elemű funkcionális adatminta esetén minden egyes mintaelemet (görbét)  $n$  darab diszkrét időponthoz tartozó érték alapján írunk le:  $((t_1, y_{i1}), (t_2, y_{i2}), \dots, (t_n, y_{in}))$ , ahol  $i = 1, \dots, N$ . A megfigyelt  $\mathbf{y}_i = (y_{i1}, y_{i2}, \dots, y_{in})$  adatvektorok – mint a görbékbeli rendelkezésre álló pontok – a funkcionális adatelemzés számára egy látens  $x_i(t)$  függvény  $\mathbf{t} = (t_1, \dots, t_n)$  időpillanatokban  $\boldsymbol{\epsilon}_i = (\epsilon_{i1}, \dots, \epsilon_{in})$  hibával mintavételezett értékei, azaz

$$\mathbf{y}_i = x_i(\mathbf{t}) + \boldsymbol{\epsilon}_i, \quad i = 1, \dots, N$$

A funkcionális adatelemzés a görbék leírására használt  $x_i(t)$  függvényeket, bázisfüggvények lineáris kombinációjából állítja elő. Ha a leírásra  $K$  db bázisfüggvényt használunk  $\phi_k(t)$ ,  $k = 1, \dots, K$ , akkor a függvény kifejtése a következő:

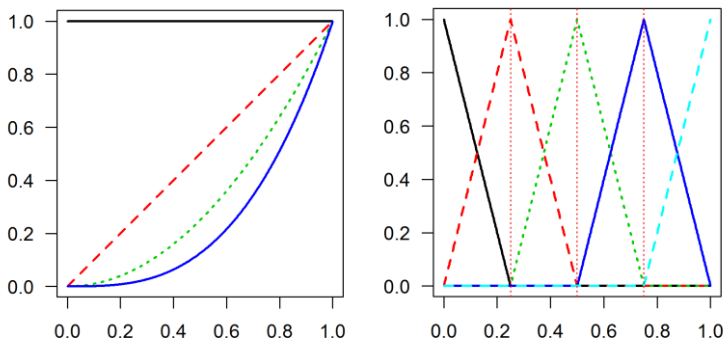
$$x_i(t) = \sum_{k=1}^K c_{ik} \phi_k(t), \quad i = 1, \dots, N \quad (2.1)$$

A  $\phi_k(t)$  bázisfüggvények különböző típusúak lehetnek: konstans, polinomiális, Fourier vagy B-spline bázisfüggvények. Mindegyik típus más-más előnnyel rendelkezik. Például a periodikus függvényeket leggyakrabban Fourier bázisfüggvényekkel, míg a nem-periodikusakat spline bázisfüggvényekkel

modellezik. A negyedrendű polinom alapú bázisfüggvények alakja:  $\phi_k(t) = t^{k-1}, k = 1, 2, 3, 4$  (2.2. ábra). A B-spline rekurzív definícióját a (2.2) összefüggés tartalmazza [de Boor(1986)].

$$\phi_{k,1}(t) = \begin{cases} 1, & \text{ha } t_k \leq t < t_{k+1} \\ 0, & \text{máshol} \end{cases} \quad (2.2)$$

$$\phi_{k,p}(t) = \frac{t - t_k}{t_{k+p-1} - t_k} \phi_{k,p-1}(t) + \frac{t_{k+p} - t}{t_{k+p} - t_{k+1}} \phi_{k+1,p-1}(t)$$



2.2. ábra. Negyedrendű polinomiális és másodrendű B-spline bázisfüggvények a [0,1] intervallumon

A (2.1) egyenlet ismeretlen  $c_{ik}$  együtthatóinak meghatározására a funkcionális adatelemzés szakirodalma számos módszert kínál. A regressziószámításból ismert közönséges legkisebb négyzetek módszere szerint az eredeti megfigyelések és a becült értékek eltérésének négyzetösszegét képezzük ( $SSE$ , hiba eltérés-négyzetösszeg), majd az együtthatókat úgy választjuk meg, hogy az  $SSE$  minimális legyen. Ha tekintjük az  $i$ . mintaelemet (görbét), akkor az eredeti megfigyeléseket az  $\mathbf{y}_i$  vektor tartalmazza (a görbe  $n$  darab megfigyelt értéke), az az ismeretlen együtthatókat pedig a  $\mathbf{c}_i = (c_{i1}, \dots, c_{iK})$  vektorral jelöljük. A (2.3) összefüggés az  $i$ . mintaelemre vonatkozó minimalizálandó kifejezést tartalmazza.

$$SSE(\mathbf{c}_i) = \sum_{j=1}^n [y_{ij} - x_i(t_j)]^2 = \sum_{j=1}^n [y_{ij} - \sum_{k=1}^K c_{ik} \phi_k(t_j)]^2 \quad (2.3)$$

Ha  $\Phi$  mátrix a  $\phi_k(t_j)$  értékeket tartalmazó  $(n \times K)$ -s mátrix, akkor a modellt a (2.4), az együtthatók becslését pedig a (2.5) összefüggés tartalmazza.

$$\mathbf{y}_i = \Phi \mathbf{c}_i + \boldsymbol{\epsilon}_i \quad (2.4)$$

$$\hat{\mathbf{c}}_i = (\Phi' \Phi)^{-1} \Phi' \mathbf{y}_i \quad (2.5)$$

A (2.4) összefüggésen alapuló modell feltételezi, hogy az  $\boldsymbol{\epsilon}_i$  hibatag független, normális eloszlású, 0 várható értékű valószínűségi változó. Funkcionális adatok esetében azonban a görbe egymáshoz közelebb lévő pontjai jobban korrelálnak egymással, mint két egymástól távolabb lévő pont. Ezért a legkisebb négyzetek elvét használó regressziós simítás modelljén túl számos további modellt dolgoztak ki a funkcionális adatok leírására.

Az „egyenletlenségi büntetés” (roughness penalty) megközelítés az illesztett görbe egyenletlenségének mértékét helyezi el a modellben, melynek meghatározása az  $x_i(t)$  illesztett görbefüggvényre, a második derivált négyzetének integráljával történik (2.6 összefüggés). Az integrálás a  $t$  legkisebb lehetséges értékétől a legnagyobbig terjed, amely így a görbe teljes görbületét adja.

$$\text{PEN}(x_i) = \int [D^2 x_i(t)]^2 dt \quad (2.6)$$

A (2.7) egyenletben szereplő modell tehát két tagot tartalmaz: az első az  $\mathbf{y}_i$  adatokra minél tökéletesebben igyekszik illeszkedni, míg a második bünteti azt az illeszkedést, amely nem megfelelő mértékben sima. A büntető rész miatt nem lesz tökéletes az illeszkedés az  $\mathbf{y}_i$  pontokra, de cserébe az eredményül kapott görbe megfelelően sima lesz.

$$F(\mathbf{c}_i) = \sum_{j=1}^n [y_{ij} - x_i(t_j)]^2 + \lambda \text{PEN}(x_i) \quad (2.7)$$

A (2.1) egyenlet ismeretlen  $c_{ik}$  együtthatóinak meghatározása most a (2.7) összefüggésben szereplő függvény minimalizálásával történik. Az együtthatók kiszámolása a (2.9) egyenlet alapján történik. A felhasznált egyenletlenségi büntetések  $\mathbf{R}$  mátrixának meghatározását a (2.8) egyenlet tartalmazza. A  $\lambda$  a becslési eljárás kritikus paramétere, mely automatikusan meghatározható a GCV (generalized cross validation) eljárással [Craven & Wahba(1979)]. Ha a  $\lambda$  paraméter értéke nagy, akkor a görbe simítottabb lesz, ha kicsi, akkor az egyes adatpontokra jobban illeszkedő, hullámosabb görbét kapunk.

$$\mathbf{R} = \int \boldsymbol{\phi}(t)\boldsymbol{\phi}'(t)dt, \text{ ahol } \boldsymbol{\phi}(t) = (\phi_1(t), \dots, \phi_K(t)) \quad (2.8)$$

$$\hat{\mathbf{c}}_i = (\boldsymbol{\Phi}'\boldsymbol{\Phi} + \lambda\mathbf{R})^{-1}\boldsymbol{\Phi}'\mathbf{y}_i \quad (2.9)$$

### 2.9.2. Feltáró módszerek

A funkcionális adatok leírása biztosította, hogy a diszkrét adatok helyett simított függvényekkel reprezentált mintaelemekkel dolgozhassunk tovább. Ramsay és Silverman (2002, 2005) részletesen beszámol azokról az adaptált többváltozós statisztikai eljárásokról, és számos új megközelítésről, amelyekkel feltáró és következtető jellegű elemzések is végezhetők funkcionális adatokon.

A funkcionális adatok átlagfüggvénye a pontonként vett görbeértékek átlaga.

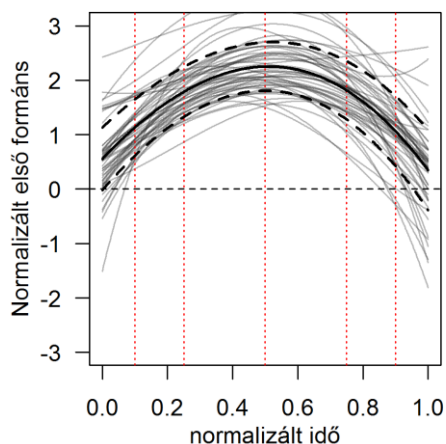
$$\bar{x}(t) = \frac{1}{N} \sum_{i=1}^N x_i(t) \quad (2.10)$$

Az átlagfüggvénnyel mind a varianciafüggvény (2.11 egyenlet), mind a szórásfüggvény értelmezhető (2.12 egyenlet).

$$var_x(t) = \frac{1}{N-1} \sum_{i=1}^N (x_i(t) - \bar{x}(t))^2 \quad (2.11)$$

$$sd_x(t) = \sqrt{var_x(t)} \quad (2.12)$$





2.3. ábra. A C.1[a:]C.1 hanghármas első formánsának mozgásai, átlagfüggvénye (folytonos vonal) és az átlagfüggvényhez hozzáadott és belőle kivont szórásfüggvény képe (szaggatott vonal). A pontozott függőleges vonalak a mérési pontok helyét jelölik.

A 2.3. ábrán a C.1[a:]C.1 hanghármas első formánsának mozgásait negyedrendű polinomiális bázisfüggvényekkel simítottam. A vastagított folytonos vonal az átlagfüggvény. Az átlagfüggvényhez hozzáadott és belőle kivont szórásfüggvényt szaggatott vonallal jelöltem.

### 2.9.3. Funkcionális lineáris modell

Ahhoz, hogy a funkcionális adatokban megmutatkozó variabilitást más változók segítségével magyarázhassuk, a klasszikus statisztika általános lineáris modelljének funkcionális környezetbe adaptált változatára van szükség. Funkcionális lineáris modellt kapunk, ha vagy a függő változó vagy a független változók valamelyike funkcionális természetű, illetve összetettebb esetekben akár mindkét oldal lehet funkcionális. Az értekezésemben a formánsmozgás, mint funkcionális változó a függő változó szerepét tölti be. A magyarázó változók a vizsgált magánhangzók környezetében lévő hangcsoportok kategóriái, így egy lényegesen egyszerűbb funkcionális lineáris modellel dolgozhatunk.

A (2.13) összefüggés a funkcionális varianciaelemzés modelljét tartalmazza, amely funkcionális függő változót és egy  $q$  különböző értékkel rendelkező kategorikus magyarázó változót tartalmaz. A formula a  $g$ . csoport  $i$ . függvényének modell szerinti megadását jelenti. A  $\mu(t)$  a teljes mintára vonatkozó átlagfüggvény, az  $\alpha_g(t)$  a  $g$ . csoport hatása, az  $\epsilon_i(t)$  a reziduum függvény.

$$y_i(t) = \mu(t) + \alpha_g(t) + \epsilon_i(t) \quad (2.13)$$

Az  $\mu(t)$  és  $\alpha_g(t)$  együttható függvények becslése egy 0 vagy 1 értékeket tartalmazó  $(N \times (q+1))$ -es  $\mathbf{Z}$  modellmátrix alapján történik. A  $\mathbf{Z}$  modellmátrix első oszlopa 1 értéket tartalmaz. Egy  $g$ . csoport  $i$ . függvényéhez tartozó  $\mathbf{Z}$ -beli sor, az első oszlopban szereplő 1 értéken kívül már csak egyetlen oszlopban, a  $(g + 1)$ . oszlopban tartalmaz 1 értéket, máshol nullát. A mátrix  $i$ . sorának  $j$ . elemét  $z_{ij}$ -vel jelölve és az együttható függvényeket átnevezve  $\beta_1 = \mu, \beta_2 = \alpha_1, \dots, \beta_{q+1} = \alpha_q$  alapján, a (2.13) modellel ekvivalens modellt kapunk a (2.14) formulában.

$$y_i(t) = \sum_{j=1}^{q+1} z_{ij} \beta_j(t) + \epsilon_i(t) \quad (2.14)$$

Az ismeretlen  $\beta_j(t)$  regressziós együttható függvények meghatározása a (2.15) összefüggésben szereplő függvény minimalizálásával történik. A folyamat végén a becsült  $\hat{\boldsymbol{\beta}} = (\hat{\beta}_1, \dots, \hat{\beta}_{q+1})$  paraméterfüggvények a (2.1) egyenlet kifejtése szerint állnak rendelkezésre. Ennek részleteit [Ramsay & Silverman(2005)] tartalmazza. A modell illesztése során a minden  $t$ -re teljesülő  $\sum_{j=2}^{q+1} \beta_j(t) = 0$  korlátozó feltételt használjuk.

$$SSE(\boldsymbol{\beta}) = \sum_{g=1}^q \sum_{i=1}^{N_g} \int \left[ y_i(t) - \sum_{j=1}^{q+1} z_{ij} \beta_j(t) \right]^2 dt \quad (2.15)$$

A együttható függvények becslése után a  $g$ . csoportra jellemző előrejelzett görbe a (2.16) két ekvivalens egyenlete alapján számolható, amely a csoportra prediktált görbét az összes görbe  $\hat{\beta}_1(t)$  (vagy  $\hat{\mu}(t)$ ) átlagára és egy  $\hat{\beta}_{g+1}(t)$  (vagy  $\hat{\alpha}_g(t)$ ) eltérésgörbe összegére bontja fel.

$$\begin{aligned} \hat{y}(t) &= \hat{\beta}_1(t) + \hat{\beta}_{g+1}(t) \\ \hat{y}(t) &= \hat{\mu}(t) + \hat{\alpha}_g(t) \end{aligned} \quad (2.16)$$

A modell illeszkedésének hagyományos fokmérője az  $R^2$  determinációs együttható, amely funkcionális környezetben függvényként jelenik meg. Az  $SSE$  hiba eltérés-négyzetösszeg függvény kiszámítását a (2.17) tartalmazza. A  $\mathbf{Z}_i$  a  $\mathbf{Z}$  mátrix  $i$ . sorát jelöli.

$$SSE(t) = \sum_{i=1}^N [y_i(t) - \mathbf{z}_i \hat{\boldsymbol{\beta}}(t)]^2 \quad (2.17)$$

A teljes eltérés-négyzetösszeg függvényt a (2.18) összefüggés tartalmazza.

$$SST(t) = \sum_{i=1}^N [y_i(t) - \hat{\mu}(t)]^2 \quad (2.18)$$

A két fenti összefüggés alapján az  $R^2$  kiszámítása:

$$R^2(t) = \frac{SST(t) - SSE(t)}{SST(t)} \quad (2.19)$$

Az  $R^2(t)$  a funkcionális modell magyarázó erejét mutatja a  $t$  teljes tartományában. A függvény értékei 0 és 1 között mozoghatnak.

Egyetlen numerikus értékkel is jellemezhetjük a modell magyarázó erejét, ha az  $R^2(t)$  átlagát tekintjük a  $t$  tartományán. Alkalmasan választott különböző  $N_R$  db alappontban vesszük a függvény értékét és azokat átlagoljuk (2.20 összefüggés).

$$R^2 = \frac{1}{N_R} \sum_{i=1}^{N_R} R^2(t_i) \quad (2.20)$$

#### 2.9.4. A simító spline ANOVA

A simító spline ANOVA (smoothing spline analysis of variance, SSANOVA) [Gu(2002)], [Wang(2010)], olyan statisztikai eljárás, amellyel eldönthetjük, hogy több görbecsoport alakja eltér-e egymástól vagy sem. A módszer matematikai háttérét a reprodukálható magú Hilbert-tér (Reproducing Kernel Hilbert Space, RKHS) adja, amely a 20. század közepén jelent meg és szivárgott be a valószínűségszámításba és a statisztikába [Aronszajn(1950)], [Parzen(1959)]. Wahba vezette be a spline függvényekkel történő simításban az RKHS használatát [Wahba(1990)], Gu pedig tenzor szorzatot fejlesztett a simító spline-okhoz, ezzel megadva az RKHS egy további felhasználását [Gu(2002)], amely az SSANOVA módszerhez vezetett.

A simító spline-ok a köbös spline-ok típusába tartoznak [Gu(2002)]. Felhasználásuk a görbék leírásában a (2.7) alapján történik. Az SSANOVA jelen értekezésben használt modellje két független változót tartalmaz, egy  $x_1$  diszkrét

változót, és egy  $x_2$  folytonos változót, ahol az  $x_2$  hatását simító spline modellezi [Wang(2010):102]. Az SSANOVA modellje a (2.21) összefüggést használja az  $x_1$  diszkrét változó  $g$ . csoportjára jellemző előrejelzett simító spline kiszámításhoz.

$$f = \mu + f_1(x_1) + f_2(x_2) + f_{12}(x_1, x_2) \quad (2.21)$$

A fenti modellben  $\mu$  képviseli a teljes átlagot. Az  $f_1(x_1)$  a csoport főhatás, amely mindegyik csoportra egy-egy simító spline-t ad a modellhez. A  $f_2(x_2)$  az összes adatpontra legjobban illeszkedő simító spline-t jelenti, ez az  $x_2$  főhatása. Az  $f_{12}(x_1, x_2)$  tag pedig a csoportokat leíró spline-ok (csoport főhatások) és az összes adatpontra illeszkedő  $f_2(x_2)$  spline-ja közötti különbséget reprezentálja, ezzel interakciós tagként szerepel a modellben.

A standard ANOVA-tól eltérően az SSANOVA nem tér vissza egy  $F$  értékkel, így az  $f_2(x_2)$  és az  $f_{12}(x_1, x_2)$  komponensek simító paramétereinek nagyságrendjét hasonlíthatjuk össze, hogy meghatározzuk a relatív hozzájárulásukat az egyenlethez. Ha e két paraméter nagyságrendje azonos, akkor a görbecsoportok különböznek egymástól, ellenkező esetben pedig nem.

Görbecsoportok eltérését tanulmányozhatjuk a csoportok főhatásait reprezentáló simító görbékre szerkesztett Bayes konfidencia-intervallumokkal is [Wang(2010):116]. Ahol ezek a konfidencia-intervallumok átfedik egymást, a különbség a két görbe között nem szignifikáns. Konfidencia-intervallumot a (2.21) modell bármely, simító spline-nal leírt komponense köré is rajzolhatunk.

Az SSANOVA előnye, hogy nemcsak teszteli, hogy van-e szignifikáns különbség a különböző görbecsoportok között, hanem megmutatja azokat a részeket is, ahol ezek eltérnek.

## 2.9.5. Funkcionális adatelemzés a fonetikai kutatásokban

A nyelvészeti és fonetikai alkalmazások szempontjából két fő irány figyelhető meg: a klasszikus funkcionális adatelemzési módszerek és az SSANOVA alkalmazása. A Ramsay és Silverman (2002, 2005) által kidolgozott klasszikus módszereket fonetikai adatstruktúrákra alkalmazta például [Gubian et al.(2009)], [Gubian(2010)], [Kondofersky(2010)] és [Gubian(2011)]. A beszédtudomány és a funkcionális adatelemzés kapcsolódási pontjainak egyik gyűjtőhelye [Gubian(2012)] számos tanulmányt, tutoriált és programkódot tartalmaz.

Jelen értekezésben csak azokat a szakirodalmakat említem részletesebben, amelyek az SSANOVA módszert alkalmazták a fonetikai adatok elemzésében. (Ezeket [Gubian(2012)] nem említi.) A szakirodalom két területre fókuszál: az

artikuláció vizsgálata a nyelvkontúr ultrahangos felvételei alapján, illetve a formánskontúrok vizsgálata különböző kontextusokban. Az artikuláció és az akusztikum kapcsolatát keresve kísérlet történt a két elemzés egyidejű alkalmazására is.

A nyelvészeti kutatásban először [Davidson(2006)] használta a funkcionális adatelemzés módszerét. A nyelv teljes midszaggittális<sup>20</sup> kontúráját vizsgálta ultrahangos felvétel segítségével. A felvételek alapján 64 ponttal írta le a nyelv teljes görbét a nyelvgyöktől a nyelv hegyéig. A [k g z] mássalhangzókat vizsgálta, melyek realizációihoz egyetlen nyelvgörbét rendelt. A vizsgált nyelvgörbék a „legszűkebb” nyelvéllésből származtak: a zár közepéről illetve a legszűkebb résállásból. Az adatokat simított spline és Bayes konfidencia-intervallum segítségével elemezte grafikusán, illetve az SSANOVA módszert alkalmazta az eltérő szópozícióban (szóbelseji és szóvégi) lévő nyelvgörbék összehasonlítására. A módszer alkalmasnak bizonyult például arra, hogy egy adott beszélő esetében a nyelv elülső részében (kb. az első harmadában) eltérő kontúrt mutasson ki a [g] szóvégi és szóbelseji ejtésében. Az adatfelvétel speciális jellege (például az adatközlő fejének szoros rögzítése), nem teszi lehetővé, hogy több beszélő adatát egyszerre elemezzük, sőt egy beszélő esetében sem lehetséges a több munkamenetben felvett adatok összehasonlítása.

[Lilienthal(2009)] a nyelvről készült ultrahangos felvételek segítségével koartikulációs hatást mutatott ki speciális pozícióban lévő magánhangzóknál, azokban az esetekben is, ahol az akusztikus eredményből nem volt megfigyelhető. A felhasznált eszközök és módszerek megegyeztek a [Davidson(2006)] tanulmányban leírtakkal, de itt az R statisztikai programcsomag [R Core Team(2012)] `assist` csomagja került felhasználásra (ott az S-Plus 2000 `assist` csomagja [MathSoft(2013)]).

[Grimaldi et al.(2010)] fonológiai jelenségek (harmónizáció) artikulációs alapjainak megerősítését mutatta ki ultrahangos felvételek és az SSANOVA vizsgálatával.

Az ultrahangos nyelvkontúrok elemzését a kardinális magánhangzók esetében a mandarin nyelvben [Yu & Hua(2011)] vizsgálta. Az izoláltan ejtett vokálisok nyelvgörbéinek összehasonlítására a `gss` csomagot használta. Grafikusán megmutatta, hogy a nyugalmi helyzethez képest hogyan változik meg a nyelv formája az [u i a] ejtésekben.

---

<sup>20</sup> A midszaggittális sík (vagy mediánszaggittális sík vagy mediánsík vagy középsík) a test szimmetriatengelyén van, azt bal és jobb oldali részekre osztja.

Formánskontúrok vizsgálatára először [Nycz & Decker(2006)] valamint [Baker(2006a)] használta az SSANOVA módszert. Nycz és De Decker az első három formánst vizsgálták 50 ponttal mért formánskontúr segítségével. Magánhangzók kezdő, középső és befejező részében kerestek eltéréseket a formánskontúrban az egyes hangkörnyezetekben. Felvetik annak a lehetőségét, hogy az egyponos formánsmérések által azonos csoportba sorolt variánsok megkülönböztetésére van lehetőségünk a formánskontúrok segítségével. Baker szintén az [æ] hangban rejlő formánskontúr eltéréseket kereste különböző hangkörnyezetek esetén (például a *bad-ban*, *ban-bam* szekvenciákban).

Az SSANOVA módszert alkalmazta [Chanethom(2011)] diftongusok esetén a dinamikus különbségek kimutatására. [Pritchard(2012)] bolgár CjV szekvenciák formánsmozgásait vizsgálta.

### **2.9.6. A funkcionális adatelemzés szoftverkönyezete**

A rendelkezésre álló szoftvereszközök mára biztosítják a SSANOVA egyszerű végrehajtását. Röviddel Davidson (2006) tanulmánya után Baker R kódot tett közzé [Baker(2006b)], amely két görbecsoport összehasonlítására alkalmas. Davidson és Baker az S-Plus 2000 illetve az R `assists` csomagját használta.

Az SSANOVA módszer grafikus megjelenítését bővíti [Fruehwald(2010)]. R-ben felhasználható csomagok az `assist` [Wang & Ke.(2011)] és a `gss` [Gu(2012)]. A klasszikus funkcionális adatelemzés csomagja az R-ben az `fda` [Ramsay et al.(2012)]. Disszertációmban a `gss` és az `fda` csomag eljárásait használtam fel.

### 3. ANYAG, MÓDSZER, KÍSÉRLETI SZEMÉLYEK

A formánsmenetek tanulmányozásához saját beszédadatbázist hoztam létre. A beszédadatbázis alapja a BME TMIT Beszédakusztikai Laboratóriumban létrehozott FKM (Fonetikailag Kiegyensúlyozott Mondatok) beszédadatbázis<sup>21</sup> [Olaszy(2013)]. Ezt forrásadatbázisnak nevezem. Először a forrásadatbázis lényeges tulajdonságait mutatom be, majd rátérek azokra a műveletekre, amelyek a FEF (Formánsokkal ellátott FKM) vizsgálati beszédadatbázis létrehozásához vezettek.

#### 3.1. A forrásadatbázis (FKM)

Az FKM forrásadatbázis 10 beszélő, 5 férfi és 5 nő felolvasott bemondásait tartalmazza, fontosabb jellemzőit a 3.1. táblázat foglalja össze.

3.1. táblázat. Az FKM forrásadatbázis jellemzői

Adatbázis neve	FKM (Fonetikailag kiegyensúlyozott mondatok)				
Hangfelvevő	Mikrofon				
Mintavétel	44100 Hz, 16 bit				
Rögzítési környezet	Hangstúdió				
Bemondás módja	Felolvasás				
A felolvasás alapegysége	Mondat				
Feldolgozás alapegysége	Beszédhang, szó				
Átírás	Fonetikus				
A beszédadatbázis teljes ideje (perc)	1524 (25,4 óra)				
A 10 beszélők azonosítója és életkora a felvétel időpontjában	Azonosító:	N1	N2	N3	N4 N5
	Életkor:	35	30	41	53 31
	Azonosító:	F1	F2	F3	F4 F5
	Életkor:	65	53	29	30 26
A mondatok (hangminták) száma	10×1992 mondat (részletesen a 3.2. táblázatban)				
A szavak száma	208 730 db				
A beszédhangok száma	854 240 db				
A magánhangzók száma	332 827 db				
Hangszimbólum-készlet	TMIT egyedi				
Eredeti annotálási formátumok	ortografikus (.txt), annotált (.TextGrid)				

A beszélők egynyelvűek, budapestiek, életkoruk a 3.1. táblázatból kiolvasható (átlagéletkor 39,3 év). A felvételek a BME Híradástechnikai Tanszék professzionális stúdiójában készültek.

<sup>21</sup> A forrásadatbázis használatához Titoktartási Nyilatkozat kitöltésére volt szükség.

A felolvasott szöveg mondatait a BABEL beszédatadtbázis<sup>22</sup> szövegállománya adta [Vicsi & Víg(1998)]. A mondatgyűjtemény irodalmi művek gondosan kiválasztott részeit tartalmazza, melyekből keverten mindig 1-1 mondat került a felolvasási listába, így a mondatok között nem volt tartalmi összefüggés. A teljes lista összesen 1992 db különböző mondatot tartalmaz. Ebből 115 db kérdő mondat (5,8%), a többi kijelentő mondat (1877 db, 94,2%). A 10 beszélő mindegyike ezt a mondatgyűjteményt olvasta fel, mindegyik mondatot egyszer. Az FKM adatbázis feldolgozása során néhány mondat szövegtévesztés miatt kiesett, így a tényleges bemondások száma az egyes beszélők esetén a 3.2. táblázatból olvasható ki.

3.2. táblázat. Az FKM forrásadatbázis beszélőire jellemző adatok

Azonosító:	N1	N2	N3	N4	N5
Bemondások száma (db):	1939	1992	1992	1992	1992
Beszédtempó (hang/s):	13,8	13,2	11,6	11,9	12,5
Artikulációs tempó (hang/s):	13,9	13,7	12,0	12,5	12,8

Azonosító:	F1	F2	F3	F4	F5
Bemondások száma (db):	1940	1942	1938	1938	1992
Beszédtempó (hang/s):	12,1	12,7	13,7	10,6	13,6
Artikulációs tempó (hang/s):	12,4	13,6	13,8	11,4	14,2

Öt beszélő esetében nem teljes az adatbázis (N1, F1, F2, F3 és F4), a hiányzó mondatok számának átlaga 52 mondat, melyek kb. 93%-ban megegyeznek egymással.

A BME TMIT munkatársai az FKM forrásadatbázist annotálták és címkékkel látták el. Első lépésként kényszerített gépi beszédfelismerést [Mihajlik et al.(2002)] alkalmaztak. Az eljárás bemenete az elhangzott mondat hullámformája (.wav) és a felismerendő beszéd ortografikus formája (.txt). A kényszerített gépi beszédfelismerés eredménye szöveges állományokban került rögzítésre (.TextGrid) az alábbiak szerint: 1) az időkoordinátákkal megadott hang- és szóhatárok, 2) a felismert hangszimbólumok sorozata (2.2. és 2.3. táblázat) és 3) különleges jelzések négy fajtája. A különleges jelzések közé tartozik a *start\_sil* és az *end\_sil*, amely a mondat első hangja előtti illetve az utolsó hangja utáni időintervallumot jelöli meg. A harmadik különleges jelzés a *sil*, amely a hangsor belsejében előforduló olyan események időintervallumát jelzi, amelyek nem tartoznak a fonemikus hangsorozathoz, például szünetek, idegen hangelemek. A *start\_sil*, *end\_sil* és *sil* jelöléseket összefoglalóan *sil*-jelöléseknek nevezem. Különleges jelzést kapott továbbá minden olyan hang,

<sup>22</sup> <http://alpha.tmit.bme.hu/speech/hdbbabel.php> [Vicsi & Víg(2011)]



amely valamilyen okból kivételes (például glottalizált, redukált). Ezt a hangszimbólum utáni kerek zárójel jelzi, így a negyedik fajta különleges jelzés alakja: hangszimbólum+zárójel.

A forrásadatbázis létrehozásának második lépésében a BME TMIT egy munkatársa a hanghatárok és a hangszimbólumok kézi ellenőrzését végezte el minden beszédhang esetén. A hanghatárok pontossága 10 ms-on belül van, amely a pontos formánsmérések nélkülözhetetlen előfeltétele.

### **3.2. A formánsmérésekhez kialakított vizsgálati beszédatbázis (FEF)**

A FEF vizsgálati adatbázist képezi kutatásaim alapját. Ezt saját céljaimnak megfelelően alakítottam ki, de a tervezése során lényeges irányelvem volt a szabványokhoz való igazodás, a formánsmérések egyszerű kivitelezésének lehetősége, az annotációs információk egyszerű lekérdezhetősége és nem utolsósorban a reprodukálható kutatás szempontjainak figyelembe vétele.

Első lépésben bemondásonként egyetlen .TextGrid állomány különböző rétegeibe integráltam az FKM adatbázisban korábban szeparáltan tárolt információkat: a mondat ortografikus átíratát, a hang- és szóhatárokat, hangszimbólumokat és a különleges jelöléseket.

A TMIT hangszimbólumok egyedi jelölését SAMPA kódra [Gibbon et al.(1997):713–714], [Vicsi(1996)] cseréltem. Az alapkészlet 14 magánhangzóját és 25 rövid-hosszú mássalhangzóparját a 2.1. és 2.2. táblázat tartalmazza.

A vizsgálati adatbázis létrehozása során az `nphonic` nevű .TextGrid réteg a finomabb, kiterjesztett hangszimbólum-készletet használja (2.3. táblázat). A speciális hangkörnyezetben megvalósuló szimbólumok átírását saját szkript segítségével végeztem.

### **3.3. Formánsadatok**

A FEF vizsgálati adatbázis létrehozásának utolsó lépéseként modellt dolgoztam ki a magánhangzók formánsadatainak tömeges meghatározására és strukturált eltárolására. A modellben a Praat program [Boersma & Weenink(2012)] formánsmeghatározó algoritmusát használtam. A formánsmérés eredményeit a modell szerint kialakított algoritmussal és kézi módszerekkel kijavítottam. Az így létrejött formáns adatokkal bővítettem ki a vizsgálati adatbázist. A

folyamatot részletesebben az 5. fejezetben írom le. A formánsadatokat szöveges állományban tároltam el, melynek a szerkezetét a 4. fejezetben ismertetem.

A FEF beszédadatbázisban a fentiek alapján 3 különböző állománytípus tartozik minden bemondáshoz (bemondások száma = kb.  $10 \times 1992$ ): (1) .wav formátumú hanghullám, (2) .TextGrid formátum az annotációhoz és hangszegmentáláshoz (3) ellenőrzött formánsadatokat tartalmazó .txt állomány. A FEF formánsadatbázisa 4 992 405 db mért értéket tartalmaz.

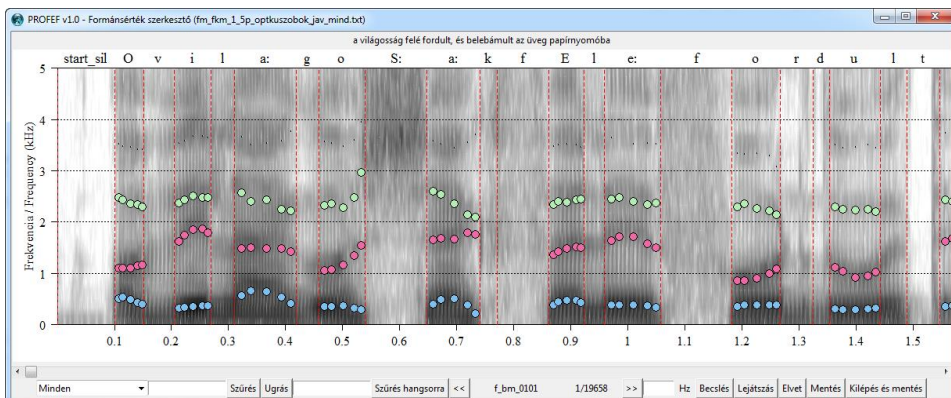
### **3.4. A PROFEF vizuális formánsadat-ellenőrző program**

A formánsadatbázis hibáinak felismerésére és javítására vizuális grafikus felhasználói felülettel rendelkező programot fejlesztettem ki. A PROFEF fantázianevű program (PROgram a FEF adatbázishoz) a korábban fejlesztett webes környezetben működő Interaktív Formánsérték Szerkesztő<sup>23</sup> [Abari & Olaszy(2011b)] továbbfejlesztett változata. A PROFEF asztali alkalmazás, futtatásához az R statisztikai programcsomag és az Active Tcl 8.4+ ingyenes szoftvercsomagok szükségesek. A 3.1. ábra az alkalmazás képernyőképét mutatja be. Az alkalmazás ablakának legnagyobb részét a beszédmintához tartozó spektrogram teszi ki, amelyen a mért formánsértékeket kis színezett pontok jelenítik meg. A hangszínek fölött a bemondás ortografikus és fonetikus formája jelenik meg. A képernyő alján a vezérlő gombok és a szöveges beviteli mezők helyezkednek el. A vezérlő elemekhez tartoznak a lapozást végző gombok, amelyek segítségével a beszédadatbázis összes eleme elérhető. Lehetőségünk van a beszédadatbázis lapozással elérhető elemeinek szűkítésére is, például egy adott beszélőre vagy egy tetszőleges hangsor, akár CVC szekvencia előfordulására is szűrhetünk, illetve egy tetszőleges beszédmintára is ugorhatunk.

A PROFEF legfontosabb tulajdonsága, hogy mind a hibák felismerése, mind a hibák javítása során a kézi és az automatikus (gépi) módszert is támogatja.

---

<sup>23</sup> Elérhető: <http://magyarbeszed.tmit.bme.hu/ifem> [Abari & Olaszy(2011a)]



3.1. ábra. A PROFEF alkalmazás munkaképernyője

A hibák manuális felismerését a spektrogramon megjelenő, formánsértékeket jelölő pontok függőleges pozíciójának vizuális ellenőrzésével végezhetjük el. Az automatikus hibafelismerést pedig egy újabb szűrő funkció biztosítja, amellyel a hibás hangmintákra szűkíthetjük a lapozható elemeket a beszédatadabázison belül. A különböző hibafajtákra (6.1. fejezet) külön-külön is szűrhetünk.

Detektált hibák javíthatók kézzel, ugyanis a formánsértékeket szimbolizáló pontok mozgathatók az egér segítségével. A mozgathatóság érzékenysége 15,3 Hz/pixel, azaz a legkisebb függőleges pontelmozdulás a hangszínképen ennyivel módosítja a formánsfrekvencia értékét. A felismert hibák automatikusan is javíthatók, az adatbázisra jellemző becült átlagérték segítségével. A javítás hatóköre a vizsgált bemondás, így az új formánsértékek, azaz a spektrogram pontjainak, azonnali vizuális ellenőrzésre is lehetőségünk van.

A formánsadatbázis létrehozása során több ponton használtam az itt bemutatott PROFEF alkalmazást (5.1. fejezet, 6.2. fejezet).

### 3.5. Magánhangzó normalizálás

A beszédatadabázis különböző beszélőihez tartozó formánsértékek egységes kezeléséhez szükség van az adatok normalizálásra. Minden formánsfrekvencia érték két koordinátával rendelkezik, egy idő és egy frekvencia komponenssel. A frekvenciatengelyen a sok szóba jöhető ún. magánhangzó normalizálási eljárás közül a Lobanov-féle módszert [Lobanov(1971)] választottam, amely könnyen értelmezhető és jól használható a beszélők közötti különbségek eliminálására (például fiziológia, anatómiai különbségek), míg a hangsor beszélőfüggetlen része, a fonemikus információ, változatlan marad [Flynn & Foulkes(2011)].

A Lobanov-féle módszer a matematikai statisztikában megszokott Z-transzformáción alapul:

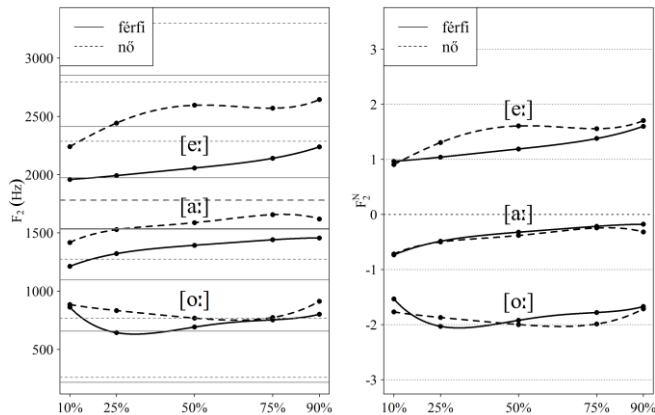
$$F_i^N = \frac{F_i - \bar{F}_i}{s_i}, i = 1,2,3 \quad (3.1)$$

A módszer alapján beszélőnként és formánsokként kiszámoljuk a formánsértékek átlagát ( $\bar{F}_i$ ) és szórását ( $s_i$ ), melyekkel korrigáljuk az összes magánhangzó megfelelő formánsértékeit. A 3.3. táblázatban összefoglaltam az adatbázisom magánhangzóira kiszámolt statisztikai mutatókat, 10 beszélőre és két formásra csoportosítva.

3.3. táblázat. A Lobanov-féle normalizálás során felhasznált átlag- és szórásadatok az összes magánhangzó alapján.

	Átlag ( $\bar{F}_i$ )		Szórás ( $s_i$ )	
	F <sub>1</sub> (Hz)	F <sub>2</sub> (Hz)	F <sub>1</sub> (Hz)	F <sub>2</sub> (Hz)
<b>N1</b>	540	1818	165	492
<b>N2</b>	517	1783	190	564
<b>N3</b>	502	1817	142	474
<b>N4</b>	517	1809	186	514
<b>N5</b>	550	1678	153	432
<b>F1</b>	473	1542	115	414
<b>F2</b>	441	1382	107	328
<b>F3</b>	431	1473	124	332
<b>F4</b>	486	1421	121	370
<b>F5</b>	488	1569	135	409

A normalizálás fontos szerepét a 3.2. ábra mutatja be egy példán keresztül. Három magánhangzó F<sub>2</sub> formánsmozgását láthatjuk egy női és egy férfi beszélő ejtésében. A bal oldali ábrán a normalizálás előtti állapotot, a jobb oldalin pedig a Lobanov-féle módszerrel korrigált mozgásokat láthatjuk.



3.2. ábra. A normalizálás hatása a formánsértékekre. Az átlagtól azonos szórásnyi távolságra lévő frekvenciaértékek (bal oldali ábra) a normalizált formáns térben azonos pozícióba kerülnek (jobb oldali ábra)

A normalizálás eredményének jobb megértését rácsvonalak segítik. A bal oldali ábrán a férfi beszélő F<sub>2</sub> formánsmozgás átlagát a folytonos vízszintes vastag vonal, továbbá az egy, két és három szórásra lévő frekvenciákat vékony folytonos vonal jelzi. A szaggatott vízszintes vonalak a női beszélőre vonatkoznak, hasonló jelentéssel. A jobb oldali ábra pontozott vonallal normalizált formáns tér -3, -2, -1, 0, 1, 2, és 3 pontjait emeli ki. Az egyes értékek az átlagtól való szórásnyi eltérések együtthatóit jelentik. A normalizálás úgy értelmezhető, hogy az átlagtól azonos szórásnyira lévő formánsértékek átfedik egymást a normalizált formáns térben. A 3.2. ábra bal oldali részén az [a:] hangra vonatkozó két formánsmozgás megmutatja, hogy a görbék a beszélőre jellemző saját átlaguknál csaknem azonos (saját szórásukban számolt) mértékben alacsonyabban helyezkednek el, így a jobb oldali ábrán ezek a görbék szinte átfedik egymást.

A normalizálás következtében az összes beszélő formánsadata egységesen kezelhető. A FEF adatbázis speciális jellege miatt, az időtengelyen eleve normalizáltan álltak rendelkezésre az adatok, így az  $x$  koordináta szerinti normalizálásra nem volt szükség. A normalizálási lépés után a mértékegység nélküli, normalizált formáns térben dolgozunk tovább, nem az eredeti Hz-ben mért formáns térben. Később bármikor visszatérhetünk az eredeti frekvencia értékekhez, ehhez csak a beszélőt és a formánsindexet kell ismernünk.

### 3.6. Hangkategória definíciók

A magánhangzókban létrejövő formánsmenetek meghatározása szempontjából a magánhangzók és az egyes hangkörnyezetek tipizálása egyszerűsíti a mérési

feladatot. A magánhangzók rövid és hosszú kategóriáit összevontam, akusztikai szerkezetüket egységesnek tételeztem fel (3.4. táblázat). A mássalhangzók akusztikai hatásának tipizálására a [Gósy(2004):79]-ben leírt mássalhangzó-kategóriákból kiindulva 11 csoportot alakítottam ki (3.5. táblázat). A 11 mássalhangzó kategória a szokásos képzési hely szerinti kategorizálásnál finomabb csoportokat ír elő, amely a csoportok elemszámának kiegyensúlyozásában is fontos szerepet játszik. A mássalhangzótipusok kialakításánál fontos elvnek tartottam a tisztán nazálisok megkülönböztetését. A C.2 a labiodentális, a C.4 posztalveoláris, a C.6 veláris és a C.11 laringális mássalhangzókat tartalmaz. A C.1 csoportban az orális bilabiálisok, a C.7 kategóriába az [m] nazális bilabiális szerepel. A C.3 típusban felsorolt alveolárisok között nem szerepel a C.8-ba áthelyezett [n] és a C.10-be átemelt [l]. A C.5 palatális kategóriából a C.9-be irányítottam az [ɲ] hangot.

3.4. táblázat. Magánhangzótipusok és típusnevek

V.1	V.2	V.3	V.4	V.5	V.6	V.7	V.8	V.9
[ɔ]	[a:]	[ɛ]	[e:]	[i, i:]	[o, o:]	[ø, ø:]	[u, u:]	[y, y:]

3.5. táblázat. Mássalhangzótipusok és típusnevek

C.1	C.2	C.3	C.4	C.5	C.6
[b, p]	[f, v, m]	[t, d, s, z, ts, dz, r]	[ʃ, ʒ, tʃ, dʒ]	[ç, j, ɲ]	[k, g, ŋ, x]

C.7	C.8	C.9	C.10	C.11
[m]	[n]	[ɲ]	[l]	[h]

A formánsmenetek vizsgálata során (6. fejezet) a magánhangzó- és hangkörnyezet-csoportokra a fenti típusnevek segítségével hivatkozom. Az összes magánhangzótipusra a  $V_m$ , az összes mássalhangzótipusra a  $C_t$  jelölést használom. A fonetikai szakirodalomban szokásos CVC vagy  $C_1VC_2$  helyett a  $C_{t,1}V_mC_{t,2}$  jelölést használom, amikor lényeges a típusok hangsúlyozása. A magánhangzótipus ismeretében a kategória neve helyett gyakran az öt alkotó hangok IPA jelét adom meg. Például  $C_{t,1}V.2C_{t,2}$  helyett  $C_{t,1}[a:]C_{t,2}$ , vagy  $C_{t,1}V.5C_{t,2}$  helyett a  $C_{t,1}[i, i:]C_{t,2}$  jelölést használom. Ezt a konvenciót használom 4 vagy 5 hangtípusból álló szekvenciákra is. Amennyiben a mássalhangzó környezet kihangsúlyozása a magánhangzó megjelölése nélkül bír jelentőséggel, akkor a  $C_{t,1};C_{t,2}$  jelölést használom (például egy szűkebb esetben C.1;C2).

## 4. REPRODUKÁLHATÓ BESZÉDKUTATÁS

Ebben a fejezetben [Abari(2012)] alapján a megismételhető beszéd kutatás feltételeit és a megvalósítás egy általam kidolgozott modelljét ismertetem.

### 4.1. A javasolt modell

Nagy merészség lenne a beszédtudomány összes területén megjelenő publikációra reprodukálható receptet adni. Azokra a kísérleti kutatásokra szűkíttem az általam kidolgozott modell hatókörét, amelyekben a következő adat és kód komponensek ragadhatók meg:

- a) nyilvános vagy legalább a kutatók (kutatócsoportok) között megosztható beszédatadatbázis (adat)
- b) beszédjel-feldolgozás az adatbázison (kód)
- c) beszédatadatbázis lekérdezése és lekérdezett eredmények (statisztikai) elemzése (kód)
- d) elemzési eredmények megjelenítése ábra, táblázat és mutató segítségével (kód).

A fenti 4 komponenssel megragadható kutatási környezet központjában a beszédatadatbázis áll, amely a hangminták annotációval és dokumentációval ellátott együttese. A beszédatadatbázis általában több beszélő hangmintáját tartalmazza egységes formátumban. A beszédatadatbázis formátuma sokféle lehet, mindig a kutatási cél határozza meg a legmegfelelőbb alakot (2.7. fejezet).

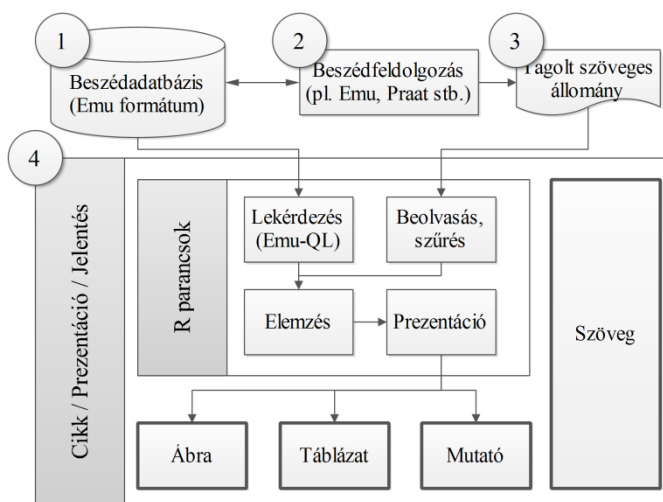
Az adatbázis beszédjel részét a digitális jelfeldolgozás segítségével szerteágazóan vizsgálhatjuk. Az algoritmusok eredményeinek egy részét a beszédjel mellett, az adatbázisban is eltárolhatjuk. A beszédfeldolgozásban felhasználható szoftver-eszközök és csomagok tárháza egyre bővül.

Az annotált beszédatadatbázisban lévő járulékos információk, a címkézési adatok és a jelfeldolgozással kapott eredmények sokféleképpen lekérdezhetők és felhasználhatók egy kutatásban. A megfelelően leszűrt és előkészített adatokon statisztikai tesztek segítségével ellenőrizhetjük hipotéziseinket.

A publikáció nélkülözhetetlen részei az ábrák, táblázatok és mutatók, melyek megjelenítéséről, szövegbe illesztéséről is gondoskodnunk kell.

A tipikusan fonetikai alapkutatásokban megjelenő, fenti adat és kód komponensek reprodukálható környezetbe implementálására a 4.1. ábrán látható modellt javaslom. Az úgynevezett végrehajtható cikk koncepciójára építék, mely

szerint egyetlen állományban (4.1. ábra: 4) szerepeltetjük a szöveget és a kódot, és a cikk „futtatásával” áll elő a kész publikáció. A cikk ábrái, táblázatai és numerikus mutatói az adatbázis (4.1. ábra: 1 és 3) tartalmának és a programkódoknak (4.1. ábra: 4, R code) megfelelően dinamikusan változhatnak az egyes „futtatások” során. A modell számítógépes környezete az R statisztikai programcsomag [R Core Team(2012)], amely Linux, Mac OS X és Windows környezetben is átfogó keretet ad a reprodukálható publikáció létrehozásához. Az R környezetet alapvetően statisztikai elemzések végrehajtására hozták létre, de tartalmaz egy általános célú, magas-szintű programozási nyelvet is, amelyet mára funkciók kimeríthetetlen tárházával láttak el. Az R ingyenes, szabad szoftver.



4.1. ábra. A reprodukálható beszéd kutatás javasolt modellje néhány területre

A reprodukálható környezet 4 elemet tartalmaz (az 4.1. ábrán használt sorszámok alapján):

1. Emu formátumú beszédatadtbázis
2. A beszédjel elemzésére szolgáló szoftver eszközök és csomagok
3. A beszédjel elemzésének eredménye tabulátorral tagolt szöveges állományban. Ide azok az eredmények kerülnek, amelyek nem illeszthetők be az Emu adatbázis struktúrájába.
4. A futtatható publikáció

Az Emu rendszer (Emu Speech Database System) egy platformfüggetlen (Linux, Mac OS X és Windows operációs rendszeren is futó), integrált programcsomag, amelyet annotált beszédatadtbázisok létrehozására, lekérdezésére és elemzésére használhatunk [Bombien et al.(2006)]. Szolgáltatásai közé tartozik a



hullámforma és a hangszínek alapján a beszédminták annotálása, valamint rendelkezik egy digitális jelfeldolgozó modullal is. Támogatja a hierarchikus annotációt, valamint az így tárolt információ lekérdezését a saját lekérdező nyelvével (Emu-QL). Az Emu használatát rendkívül megkönnyíti, hogy kezdetektől fogva transzparens interfésszel rendelkezik valamely statisztikai programcsomaghoz. Kezdetben az *S*, majd az *S-plus*, ma pedig az R statisztikai programcsomagnál érhetjük el az annotált beszédadatbázis egyes elemeit. A címkefájlok Praat és Emu között szabadon konvertálhatók, amely nagymértékben megkönnyíti az Emu formátumú beszédadatbázisok létrehozását.

A beszédben rejlő jellemzők kivonására számos eljárást fejlesztettek ki az elmúlt évtizedek folyamán. Ezek egy részét implementálták az Emu-ba (például F0 analysis, formant estimation, short-term autocorrelation function, short-term spectral analysis, Linear Prediction analysis, zero-crossing rates stb.). Az Emu elemzéseinek eredménye az Emu formátumú adatbázis szerves részét képezi, amelyek később az Emu-QL nyelv segítségével R-ből is lekérdeezhetőek. Ehhez az R `emu`<sup>24</sup> csomagját kell használnunk. Ha más jelfeldolgozó programcsomag szkriptjét (például Praat, WaveSurfer/Snack<sup>25</sup>) vagy saját fejlesztésű algoritmusokat szeretnénk futtatni, akkor azok eredményét tagolt szöveges állományban érdemes eltárolnunk, mert ez a formátum az R számára könnyen beolvasható. Megemlítjük, hogy maga az R is rendelkezik beszédfeldolgozó funkciókkal, számos csomag segíti a hangelemzést és a jelfeldolgozást. A *seewave*<sup>26</sup> rendkívül gazdag függvénykészletet biztosít a beszédelemzésre, manipulációra, megtekintésre és szerkesztésre. Kiválóan alkalmas hanghullámok és 2D és 3D spektrogramok megjelenítésére. A *tuneR*<sup>27</sup> csomag .wav fájlok, hangminták és zene kezelésére ad alapvető függvényeket. A *signal*<sup>28</sup> csomag Matlab/Octave kompatibilis jelfeldolgozó függvényeket biztosít. Továbbá az R *tcltk* csomagja lehetővé teszi a Tk grafikus felhasználói felületének használatát és *Tcl* parancsok kiadását az R-en belül. Azok a beszédfeldolgozó könyvtárak, amelyek rendelkeznek Tcl wrapper-rel (például WaveSurfer/Snack, ASSP) az R-ből közvetlenül elérhetőek.

A futtatható cikk (1. ábra: 4) R kódot és szöveget tartalmaz. Az R kódok biztosítják az adatok lekérdezését az Emu beszédadatbázisból és a tagolt szöveges állományokból. Továbbá a statisztikai elemzésekhez és a képek, táblázatok és numerikus mutatók megjelenítéséhez is R kódokat használnunk. A

---

<sup>24</sup><http://emu.sourceforge.net/> [Bombien et al.(2012)]

<sup>25</sup><http://www.speech.kth.se/snack/> [Sjölander(2006)]

<sup>26</sup><http://cran.r-project.org/web/packages/seewave/index.html> [Sueur et al.(2013)]

<sup>27</sup><http://cran.r-project.org/web/packages/tuneR/index.html> [Ligges(2013b)]

<sup>28</sup><http://cran.r-project.org/web/packages/signal/index.html> [Ligges(2013a)]

cikk másik részét a szöveg teszi ki, amelynek formátuma többfajta lehet. Az R kód számos formanyelvvel kombinálható. Három különböző formanyelv – a LaTeX, Markdown és Org – R-el kombinált használatát mutatja be [Abari(2012)].

## 4.2. Korlátok

[Vandewalle et al.(2009)] megemlíti, hogy az adatok publikálása nem minden esetben lehetséges. Ha a felhasznált adatbázisokat szerzői jog védi, vagy bizalmas információkat tartalmaznak, akkor (on-line) publikálásuk aggályos lehet. Ez fokozottan érvényes a beszédatadabázisokra, különösen a spontán beszédet tartalmazókra, valamint a beszédtechnológia speciális algoritmusaira, amelyek publikálása így nehézségekbe ütközik. Ekkor a reprodukálhatóság szükségszerűen háttérbe szorul. Peng (2011) a korlátokat egyrészt a tudományos kultúra hiányában látja, aminek pedig ki kellene követelnie minden tudományos állítás reprodukálhatóságát, másrészt az infrastruktúra elégtelenségében, amely jelenleg nem teszi lehetővé széles körben a reprodukálható kutatás egyszerű terjesztését.

Peng (2011) a tudományos állítások ellenőrzése szempontjából egy intervallumot határozott meg, amelynek egyik végén a nem reprodukálható, a másik végén a teljesen reprodukálható cikk szerepel. Köztük a reprodukálás különböző fokozatai jelennek meg, attól függően, hogy szerző az elemzéseikhez használt adatokból és kódokból mennyit tesz elérhetővé mások számára, illetve mennyire egyszerű rekonstruálni ezekből az eredményt. Egyes szerzők 3 fokozatot különböztetnek meg [Schwab et al.(1997)], mások 6 részre osztják a reprodukálhatósági skálát [Vandewalle et al.(2009)]. Ha tehát az adatok terjesztése korlátokba ütközik, például a titkossága miatt vagy annyira zajos, hogy a terjesztése nem praktikus, akkor a kódok önálló publikálása már egy lépés a reprodukálhatóság irányába.

A reprodukálható kutatás szükséges, de nem elégséges feltétele a jó kutatásnak. Önmagában a reprodukálhatóság nem garantálja a minőséget és az eredményeink érvényességét. Csak azt biztosítja, hogy ténylegesen az az eredmény, amit a publikációban állítunk.

A tudományos folyóiratok a kultúra megváltoztatásának motorjai lehetnek, mivel kidolgozott protokoll szerint várják a reprodukálható kutatási eredményeket. Ilyen folyóirat például a Nature<sup>29</sup>, The Insight Journal<sup>30</sup>, Annals

---

<sup>29</sup><http://www.nature.com/authors/policies/availability.html>

<sup>30</sup><http://www.insight-journal.org/>

of Internal Medicine<sup>31</sup>, Biostatistics<sup>32</sup>, IEEE Signal Processing Magazine<sup>33</sup>. Kimondottan a beszéddel és nyelvvel kapcsolatos reprodukálható kutatások eredményeinek on-line publikálására hozták létre a *The Journal of Experimental Linguistics* folyóiratot 2009-ben. (Jelenleg még nem érhetők el cikkek a weboldalon<sup>34</sup>.)

### 4.3. Reprodukálhatóság a jelen kutatásban

Értekezésemben a megismételhető kutatás irányelveit tartottam szem előtt. Ennek megfelelően a statisztikai elemzéseim mindegyike, *Rnw* vagy *Rmd* kiterjesztésű, „futtatható” LaTeX/HTML formátumban készült. A használt beszédatadatbázisokat szerzői jog védi, így azok megosztása nem lehetséges. A formásméréshez használt külső jelfeldolgozó programok mindegyike működtethető parancssoros üzemmódban, így ezek újrafelhasználása is problémamentes.

---

<sup>31</sup><http://annals.org>

<sup>32</sup>[http://www.oxfordjournals.org/our\\_journals/biosts/for\\_authors/msprep\\_submission.html](http://www.oxfordjournals.org/our_journals/biosts/for_authors/msprep_submission.html)

<sup>33</sup><http://www.signalprocessingsociety.org/publications/periodicals/tsp/>

<sup>34</sup><http://elanguage.net/journals/jel>

## 5. A FORMÁNSADATBÁZIS ELKÉSZÍTÉSÉRE KIDOLGOZOTT ELJÁRÁS ÉS ALKALMAZÁSA

Ebben a fejezetben azokat a mérési lépéseket, valamint hibakeresési és javítási lehetőségeket részletezem, amelyeket arra terveztem, hogy ellenőrzött formánsadatok rendezetten tárolt együttesét hozhassam létre. Hipotézisem szerint a kidolgozott többlépcsős formánsjavító eljárás nagy beszédatbázisok esetén csaknem hibamentes formánsadatbázishoz vezet, azaz nagy részben helyettesítheti a manuális ellenőrzést.

Kiindulásként vegyük sorra, hogy a magánhangzók formánsfrekvenciáinak mért adatait mely tényezők határozzák meg.

- a) A legfontosabb tényező maga a beszédprodukciónak: ki, mit, hogyan mond. Ezen belül megkülönböztethetjük a beszélőre jellemző egyéni sajátosságokat, a kiejtett hangsorra vonatkozó kontextusbeli befolyásoló tényezőket és a beszédmódra, a beszélés környezetére jellemző hatásokat. Az egyes vizsgálatok esetében érdemes a fenti tényezőkből többet konstans értéken tartani. Az értekezésemben felolvasott szöveget használok és a személyek közötti különbséget normalizálással egyenlítem ki. Érdemes megjegyezni, hogy ideális esetben csak az ebben a pontban említett beszédprodukciónak van hatása a formánsértékre és az azok vizsgálatából levont következtetésekre. A további b) és c) pontokban felsorolt két tényező hatását minimálisra kell csökkenteni, ugyanis ezek a mérés torzító faktoraiként jelennek meg.
- b) Mivel a beszéd sokoldalú vizsgálatára csak annak rögzítése után van lehetőségünk, a hangfelvétel körülményei is (környezeti zaj, mintavételi frekvencia, kvantálás) jelentősen befolyásolják a mért formánsfrekvenciák értékeit. A legtöbb esetben – így a FEF adatbázisban is – a csendesített körülmények között felvett, jó minőségű hangminták biztosítják ezen változó hatásának minimálisra csökkentését. Számos tanulmány számol be azonban a zajosabb és korlátozott sáv szélességgel rendelkező felvételek (például telefonbeszélgetések) formánsokra gyakorolt hatásáról (például [Byrne & Foulkes(2004)]). Általában azt mondhatjuk, hogy zajjal terhelt felvétel esetén még bizonytalanabbá válik a formánsbecslés.
- c) A formánsmérés módja, azaz a mérő algoritmus és annak paraméter beállításai is jelentősen befolyásolják a mért formánsfrekvenciák értékeit. A formánsmérésre matematikai eljárások állnak rendelkezésre. Mivel a

formánsok meghatározásához a beszédspektrum helyi maximumaira van szükség, a formánsbecslés gépi módszerei csúcsok keresésén alapulnak, a leggyakrabban STFT (Short-Time Fourier Transform) vagy LPC (Linear Predictive Coding) elemzésre támaszkodva. Az LPC egy egyszerűbb struktúrával közelíti a beszédspektrumot, így az elmúlt években kifejlesztett eljárásokban leggyakrabban ez fordul elő.

A fenti 3 tényezőcsoport ismertetése során említettem, hogy az a) pontban felsorolt beszédprodukciónál esetén a befolyásoló összetevők közül többet érdemes konstans értéken tartani a többi faktor hatásának korrekt vizsgálatához. A b) pontbeli hangfelvétel okozta torzítást az ellenőrzött körülmények biztosításával kiküszöbölhetjük ki. A harmadik tényezővel, a c) pontban leírt automatikus formánsméréssel kapcsolatban azonban nagyobb körütekintésre van szükség, ugyanis a formánsfrekvenciák értékének pontos meghatározására máig nincs kielégítő megoldás [Fulop(2011):1]. A legelterjedtebb LPC alapú formánsmérések során számos beállítási opció közül választhatunk, amelyek az algoritmus által meghatározott formánsértékekre is jelentős hatással vannak [Harrison(2004)]. A legfontosabb paraméter a használt együtthatók száma (fokszám), amely meghatározza, hogy hány darab együtthatóval modellezzük a toldalékeső pillanatnyi alakját. Minél alacsonyabb az érték, annál pontatlanabb a modell, míg a magasabb számú beállítás (egy pontig) pontosabb modellezést eredményez [Markel & Gray(1976)]. A második fontos paraméter a vizsgálatban az elemzési ablak szélessége. Ha az elemző ablak túl kicsi, akkor nem áll elegendő beszédinformáció rendelkezésre a formánsértékek pontos kiszámításához. Ha a hossza túl nagy, akkor a beszédjel nem lesz stacionáriusnak mondható az elemzés szélességében és a kiszámolt formánsok kevésbé lesznek pontosak. A rosszul beállított paraméterek hibás mérésekhez vezetnek. A hibák egyik fő oka a helytelenül megválasztott LPC fokszám [Vallabha & Tuller(2002)]. Gyakran hibáznak az eljárások abban az esetben is, ha nagyon alacsony a formánsfrekvencia értéke, vagy ha közel vannak egymáshoz a formánsértékek. Az automatikus formánsmérések hibáinak kiküszöbölésére szolgáló módszereket három csoportba sorolhatjuk.

a) A formánsmérő algoritmus pontosabbá tétele. Az értekezésben ezzel a ponttal nem foglalkozom, ugyanis szabadon hozzáférhető beszédelemző szoftverekben implementált algoritmusokat használok a formánsfrekvencia értékek meghatározására. A magyar és nemzetközi szakirodalomban számos kísérlet történt az ilyen jellegű javításra [vö. Böhm & Németh(2006)].

- b) Az algoritmus bemenő paramétereinek optimális beállítása mellett javítható a formánsmérés pontossága. A szakirodalom egyik módszerét implementáltam (5.1.1. fejezet).
- c) A kapott nyers (nem tisztított) formánsértékek hibáinak utólagos javítása manuális és/vagy gépi módszerrel (5.2. fejezet).

## 5.1. Gépi formánsmérés

Nagy tömegű formánsadat automatikus előállítására a beszédfeldolgozás gépi algoritmusai (FFT, LPC), és azokat implementáló – többnyire szabadon hozzáférhető – szoftvercsomagok állnak rendelkezésre (például *Praat*, *Snack/Wavesurfer*, *Advanced Speech Signal Processor (ASSP)*<sup>35</sup>, *Speech Filling System*<sup>36</sup>, *openSmile*<sup>37</sup> stb.). Ezek mindegyike szkriptelhető, azaz alkalmas nagy mennyiségű hangállomány automatikus feldolgozására, formánsainak meghatározására. Dolgozatomban a gépi formánsmérések automatikus végrehajtásához a Praat 5.3.19-es verziójú programot [Boersma & Weenink(2012)] használtam fel.

A legtöbb szabadon felhasználható fonetikai mérőprogram hullámformák és spektrumképek megjelenítését is támogatja, így a gépi mérés mellett kézi formánselemzést is végezhetünk. Szélsőséges esetben az egérkurzor segítségével is leolvashatjuk a formánsfrekvencia értékeket például a spektrogramról, de gyakoribb valamely beszédelemző algoritmus outputját felhasználni a pontos méréshez. Ekkor a manuális formánselemzés a következő lépéseket foglalja magában [Evanini et al.(2009)]:

1. egy kiválasztott formánsjósoló algoritmus felhasználásával a magánhangzók formánsainak nyomkövetése,
2. a jósolt formánsok helyességének eldöntése a spektrogram és a hangfájlok lehallgatásának segítségével,
3. ha a jósolt formánsok hibásak, az algoritmus paramétereinek módosítása, amíg azok nem lesznek helyesek.

A kézi formánsmérés lényeges jellemzője, hogy a kutató lehallgatással képes a magánhangzó azonosítására, így a formánsértékekről egy magánhangzóra jellemző (saját tudása alapján) várt modell szerint dönthet.

---

<sup>35</sup> <http://libassp.sourceforge.net/> [Bombien(2012)]

<sup>36</sup> <http://www.phon.ucl.ac.uk/resource/sfs/> [Huckvale(2013)]

<sup>37</sup> <http://opensmile.sourceforge.net/> [Eyben et al.(2013)]

A fenti manuális eljárás tanulsága egy pontos gépi formánsmeghatározó számára az a követelmény, hogy a beszédatbázis hangszintű, pontos címkézéssel legyen ellátva. A magánhangzó ismeretében ugyanis egy statisztikai modell alapján automatikus módszerekkel dönthetünk a formánsfrekvencia értékek helyességéről. Ez az alapja az általam kidolgozott félautomatikus hibajavításnak, amelyről az 5.2. fejezetben számolok be. A magánhangzó ismerete továbbá lehetővé teszi a magánhangzóra jellemző optimális bemenő paraméterek meghatározását is. Ez utóbbit a következő, 5.1.1. fejezetben részletezem.

Sok kutató előnyben részesíti a kézi méréseket az automatikusokkal szemben, a nagyfokú pontosság és a konzisztencia miatt [Evanini et al.(2009)]. Ez a legfőbb oka a kisméretű adatbázisok mai napig tartó kialakításának és használatának. Nagyméretű adatbázisok esetén azonban nem csak az időtényező indokolja a gépi mérések alkalmazását. [Duckworth et al.(2011)] megmutatta, hogy különböző kutatók/laboratóriumok kézi mérései ugyan általában azonos formánsmérésre vezetnek azonos hanganyagon, de előfordulnak kisebb eltérések, inkonzisztenciák is. Ahhoz, hogy a mérések manuálisan replikálhatóak legyenek (mindig azonos eredményre vezessenek), rögzíteni kell a használt beszédelemző algoritmust és annak néhány bemenő paraméterét, a mérés időpillanatát a magánhangzón belül, továbbá azt a narratívát, amelyek a kézi formánsdöntéseket meghatározzák. A gépi mérések ezzel szemben az algoritmus és a bemenő paraméterek rögzítése után mindig azonos eredményre vezetnek.

### **5.1.1. A nyers formánsadatok előállításának optimalizálása a maximális formáns beállításával gépi mérésnél**

A gépi formánsmérések automatikus végrehajtásához saját szkriptet készítettem, amely a FEF adatbázisban tárolt hangfelvételek és annotációk kezelése mellett a gépi algoritmus bemenő paramétereinek kezelését is elvégezte. Az eljárásban a Burg algoritmust használtam (a Praat szkriptben a `To Formant (burg)` parancs) és minden magánhangzó esetén az első három formánsra 5 mérést végeztem el a magánhangzó hosszának 10, 25, 50 75 és 90%-os pontjában. Az alapértelmezett paraméter-beállítások mellett mértem, azaz az ablak hossza 0,025 s; a formánsok maximális száma 5; az elő-kiemelés kezdőfrekvenciája 50 Hz; lépésköz az ablakok között 0,01 s. Az egyetlen kimaradt paraméter a maximális formáns. Escudero és munkatársai kis adatbázison megmutatták, hogy ezen paraméter optimális beállításával növelhető a formánsmérés pontossága [Escudero et al.(2009)]. Az általuk javasolt módszert implementáltam Praat és R szkriptek segítségével saját nagyméretű FEF

adatbázisomra, és a mérés során minden beszélőre és magánhangzóra külön megállapítottam a maximális formáns paraméterének értékét.

A Praat maximális formáns paramétere a formánskeresés felső határát határozza meg, alapértelmezett értéke a Praatban 5500 Hz női és 5000 Hz férfi beszélőre. A Praatban alapértelmezetten 5 formánst keresünk. A maximális formáns paraméter a beszédjelben lévő F<sub>5</sub> és F<sub>6</sub> elválasztására szolgál. Minél pontosabban állítjuk be ezt a paramétert az adott beszélő esetén, annál pontosabb formánsmérést tudunk végezni. A javasolt eljárásban a beszédatadtbázis minden beszélőjének minden magánhangzójára egy-egy optimális maximális formáns paramétert számítanak ki úgy, hogy a szisztematikusan változtatott paraméterbeállítások mellett a magánhangzók formánsfrekvenciáinak varianciáját mérik, néhány tíz (például 20 vagy 50) előre kiválasztott magánhangzó-példány alapján. Maga a mérés úgy zajlik, hogy női beszélők esetén 4500 Hz-től 6500 Hz-ig, férfi beszélők esetén 4000 Hz és 6000 Hz között, 10 Hz-enként változtatják a paraméter értékét, és mindegyiknél elvégzik a formánsmérést. A legkisebb variancia melletti maximális formáns lesz az adott magánhangzóra jellemző optimális érték.

A fenti eljárást végrehajtottam a FEF adatbázisra, beszélőnként és magánhangzónként 50 véletlenszerűen választott adatbázisbeli előfordulás alapján. A méréseket a magánhangzó közepéről vettem.

Az eljárás eredménye személyenként és magánhangzónként egyetlen optimális maximális frekvencia paraméter (5.1. táblázat), amelyet egy szöveges állományban rögzítettem. A gépi formánsmérés során ezt használtam fel.

5.1. táblázat. A maximális formáns paraméter optimalizált értékei a FEF adatbázisban az 5 női és 5 férfi bemondóra. A számértékek frekvenciákat jelölnek (Hz)

	[i:]	[i]	[e:]	[ɛ]	[a:]	[ɔ]	[o]	[o:]	[u]	[u:]	[ø]	[ø:]	[y]	[y:]
N1	6140	5420	6470	5470	5600	6210	5860	6280	5920	5790	6400	4920	6060	5920
N2	6490	5360	5350	5900	5460	4980	4780	5720	4710	5640	4860	4900	5120	5400
N3	5980	6010	6450	5820	6230	5490	5940	4880	5550	5460	6270	4960	5000	5200
N4	5850	6440	5400	6370	6440	5520	6330	4600	5670	4780	6410	4760	5060	5060
N5	5990	5890	6440	6500	6270	5880	5550	5500	5750	5500	5950	5540	6010	5840
F1	4320	5960	4520	4710	5180	4610	4520	4430	4990	4390	4350	4310	4290	4190
F2	5040	4830	4510	4460	5190	4570	4200	4480	4050	4010	4750	4100	4690	4720
F3	5300	4530	4990	5120	5370	5040	4790	4460	4850	4900	4900	4860	5060	5130
F4	5240	5270	5400	5390	4720	5210	4950	4260	4640	4810	4940	4930	4730	4570
F5	5360	5080	5430	4870	5470	4950	4600	4290	4060	5190	4830	4880	4520	4440



### 5.1.2. Kézi ellenőrzés

A FEF vizsgálati adatbázis formánsfrekvencia értékeinek 20%-ában kézi ellenőrzést és javítást végeztem a BME TMIT munkatársaival együtt. A beszédatadabázis ezen részére a FEF.20 jelölést használom. A kézi ellenőrzés célja kettős. Egyrészt a későbbi hibajavítás számára szolgáltató adatbázisspecifikus küszöbértékeket. Ettől a hibajavítás pontosságának növelését várjuk. Másrészt a hibák kiértékelését össze tudjuk vetni a gépi javítás összesített eredményével. Alapesetben azonos hibaeloszlást várunk.

A kézzel ellenőrzött rész hanganyagát az 1992 db mondat első 396 mondata alkotja, mind a 10 beszélő esetén. A FEF.20 így összesen 66 298 db magánhangzót tartalmaz. Magánhangzónként 15 db formánsfrekvencia értékkel kell számolni, mivel az első három formánst ( $F_1$ – $F_3$ ) öt mérési pontban mértük meg. Ez alapján összesen 994 470 db formánsfrekvencia értéket ellenőriztünk manuálisan.

A kézi ellenőrzés kényelmes végrehajtásához új algoritmust és programot terveztem. A PROFEF v1.0 részletes bemutatása a 3.4. fejezetben található. A program magánhangzónként az első három formáns 5 mérési pontban meghatározott frekvenciaértékének manuális javítását teszi lehetővé. A változtatásokat naplózza, így az előforduló hibák sokrétű elemzését is biztosítja.

A kézi módosítás során összesen 33 338 db formáns értéke változott meg, amely a FEF.20 adatbázis mért formánsainak 3,35%-a. A módosítások eloszlását az egyes beszélők esetén az 5.2. táblázat tartalmazza.

5.2. táblázat. A PROFEF vizuális formánsellenőrző alkalmazással módosított formánsok százalékos aránya a FEF.20 adatbázisban az egyes beszélők esetén (%)

N1	N2	N3	N4	N5	F1	F2	F3	F4	F5
4,47	6,65	2,42	5,72	1,35	2,37	0,78	2,50	1,77	5,36

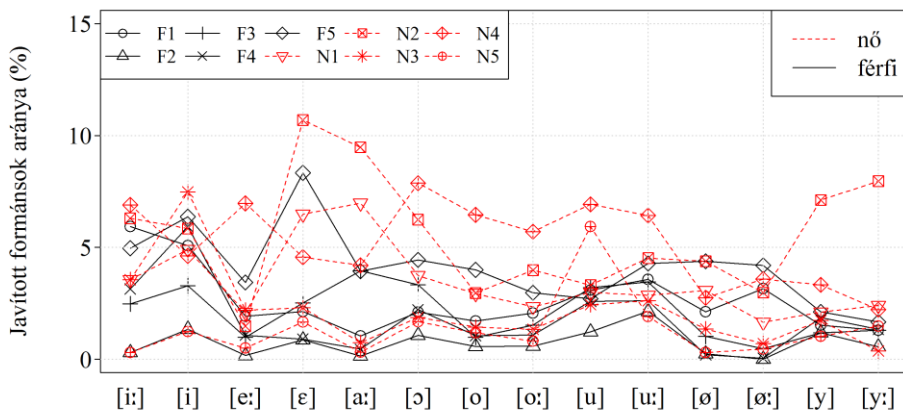
A legtöbb hiba N2, N4, F5 és N1 beszélők mért adataiban fordul elő. A legkevesebb hiba pedig az F2 beszélő esetében. Az előforduló hibák száma nem egyenletes az egyes beszélők esetén (Khi-négyzet próba:  $\chi^2 = 11\,134,56$ ,  $df = 9$ ,  $p < 0,000$ ).

Magánhangzónként vizsgálva a javítások eloszlását, hasonló eredményt kaptunk. Az előforduló hibák száma itt sem egyenletes (Khi-négyzet próba:  $\chi^2 = 2093,77$ ,  $df = 13$ ,  $p < 0,000$ ). A legtöbb hiba az [i, e, i:] vokálisokban fordul elő.

5.3. táblázat. A PROFEF vizuális formánsellenőrző alkalmazással módosított formánsok százalékos aránya a FEF.20 adatbázisban az egyes vokálisok esetén (%)

[i:]	[i]	[e:]	[e]	[a:]	[ɔ]	[o]	[o:]	[u]	[u:]	[ø]	[ø:]	[y]	[y:]
3,74	4,62	2,11	4,07	3,13	3,47	2,34	2,25	3,44	3,44	2,00	1,74	2,33	2,06

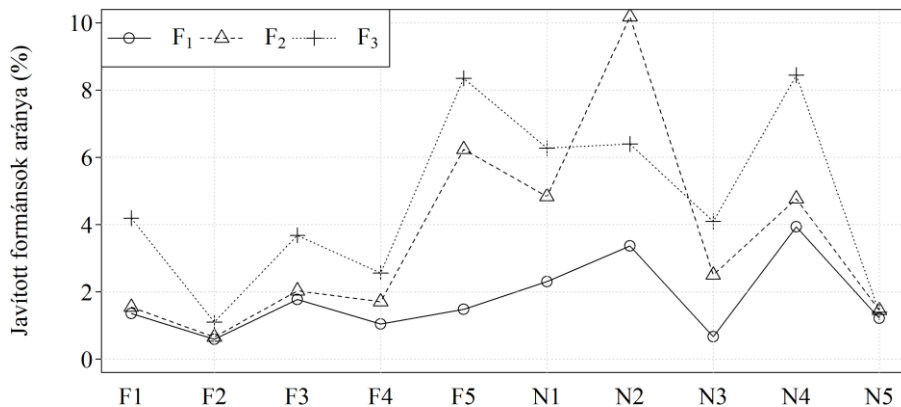
A manuális módosítások eloszlását személyenként és magánhangzónként együttesen az 5.1. ábra tartalmazza.



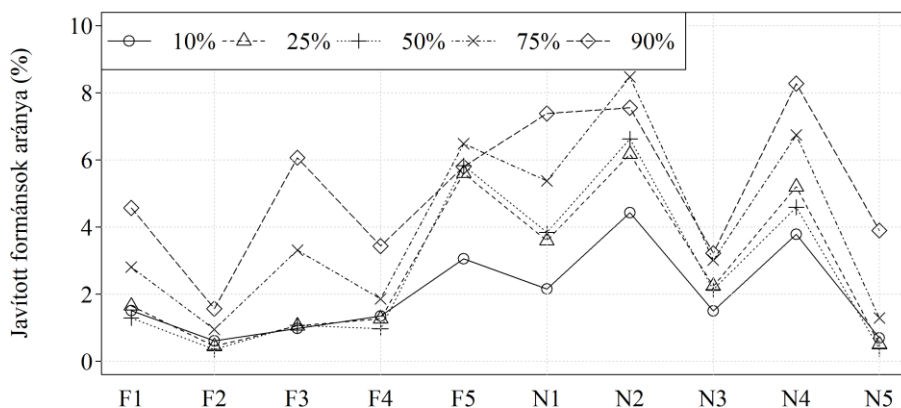
5.1. ábra. A FEF.20 adatbázisban a formánsmódosítások aránya személyenként és magánhangzónként

Az 5.1. ábra megerősíti azt a feltételezést, hogy a hibázások aránya inkább függ a személyektől, mint a magánhangzó minőségétől. A beszélőket megjelenítő összekötött vonalak eltérése nagyobbak mutatkozik, mint az egy vonalon belüli eltérés. Noha mindkét tényező esetén a Khi-négyzet próba megállapította az egyenletes eloszlástól való szignifikáns eltérést, a próbastatisztikák mintabeli értéke is erre enged következtetni, ugyanis a személyek esetében egy nagyságrenddel nagyobb volt az érték, mint a magánhangzóknál.

A formánsenkénti javítások megoszlását az 5.2. ábra szemlélteti. Általában elmondható, hogy a formánsorszámmal növekedésével a hibák száma is nő (átlagban F<sub>1</sub>: 1,78%; F<sub>2</sub>: 3,61% F<sub>3</sub>: 4,67%). Ez minden beszélő esetén teljesül, kivéve az N3 beszélőt. Az elvégzett hipotézisvizsgálat is igazolja a hibák arányának egyenlenségét (Khi-négyzet próba:  $\chi^2 = 4219,80$ ,  $df = 2$ ,  $p < 0,000$ ).



5.2. ábra. A FEF.20 adatbázisban a formásmódosítások aránya személyenként a három formás esetén



5.3. ábra. A FEF.20 adatbázisban a formásmódosítások aránya személyek és a mérési pozíciók függvényében

Az egyes mérési pontokban is megvizsgáltam a hibák arányát (5.4. táblázat). Megfigyelhető, hogy a vokális második felében megnő a szükséges javítások száma.

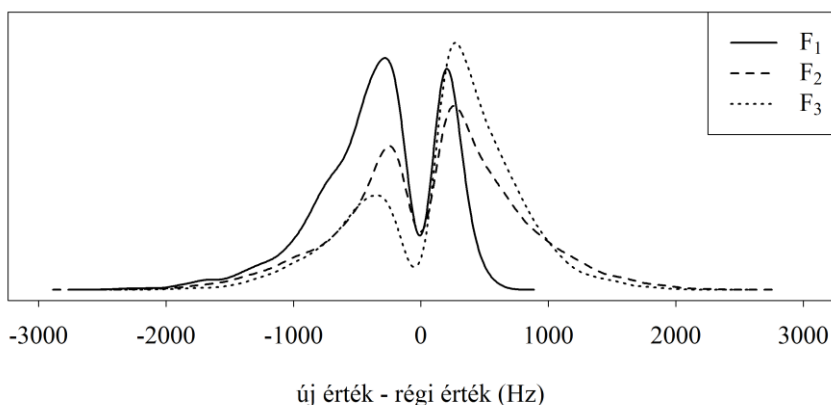
5.4. táblázat. A PROFEF vizuális formásmellenőrző alkalmazással módosított formások százalékos aránya az egyes mérési pozíciókban (%)

Mérési pont:	10%	25%	50%	75%	90%	Átlag
Javítások aránya (%):	2,01	2,79	2,72	4,05	5,19	3,35

Az öt kategóriában a javítások számának egyenetlenségét statisztikai próbával bizonyítottam (Khi-négyzet próba:  $\chi^2 = 3763,66$ ,  $df = 4$ ,  $p < 0,000$ ).

Az 5.3. ábra a személyek és a mérési pozíciók függvényében mutatja be a javítások arányát. Az ábráról leolvasható, hogy minden személy esetében a 75 és 90%-os mérési pontokban a hibázás aránya meghaladja a többi pontban mért formánsra vonatkozó hibázási arányt.

A módosítások mértékét és irányát az 5.4. ábra mutatja be. A módosított új formánsfrekvencia értékéből kivontam a régi értéket, és az így kapott különbség eloszlását rajzoltam az ábrára. Mindhárom simított eloszlásgörbéből kiolvasható, hogy egyaránt tartalmaz felfelé és lefelé módosított frekvencia értékeket. Ezek aránya  $F_1$  esetében 29% és 71%, vagyis a formánsfrekvenciák csökkentése a jellemző a javítások során. Az első formánsban az emelés átlaga 231 Hz, a csökkentése  $-547$  Hz. A második formáns esetében sokkal kiegyenlítettebb az emelések és csökkentések aránya (58%, 42%), az átlagok 561 Hz és  $-526$  Hz volt. Az  $F_3$  hasonlóan az első formánshoz eltérő módosítási arányt hozott, de most az emelések kerültek többségbe (70%, 30%). Az átlagok 503 Hz és  $-572$  Hz.



5.4. ábra. A FEF.20 adatbázisban a formánsmódosítások eloszlása a három formánsra

Megvizsgáltam, hogy a végrehajtott módosítások hány százaléka jelentős, vagyis különbözik nullától. A használt küszöbszámok  $F_1$  esetén 40 Hz,  $F_2$  és  $F_3$  esetén 100-100 Hz volt, az ennél nagyobb abszolút eltéréseket tekintettem nullától különbözőnek. A fenti küszöb feletti módosítások aránya  $F_1$  esetén 99% feletti,  $F_2$ -ben 97% és  $F_3$ -ban 98% volt.

A kézi ellenőrzéssel helyessé tett FEF.20 formánsadatbázis alkalmas az optimalizált bemenő paramétereket használó módszer értékelésére. A FEF.20 formánsadatbázist referenciának jelöltem ki. A FEF adatbázis FEF.20 adatbázissal közös hanganyagán többfajta mérési algoritmust futtattam (ezek halmaza *alg*), majd az így kapott formánsfrekvencia értékeket az (5.1) képletben leírt átlagos abszolút hiba (Mean Absolute Error) mérőszám segítségével összevettem a FEF.20 referencia formánsadatbázis frekvenciaértékeivel.

$$AE_{f_s,ab} = \frac{\sum_{i=1}^n |f_{f_s,ab,i} - f_{f_s,FEF.20,i}|}{n}, \quad (5.1)$$

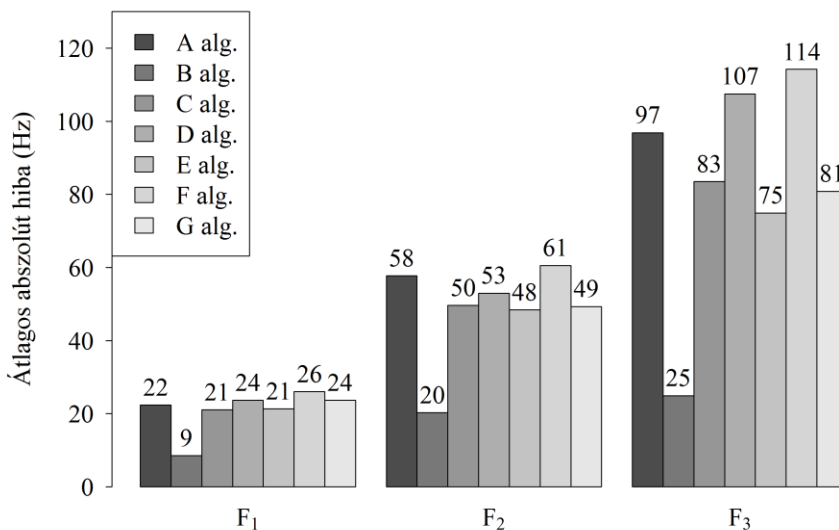
$f_s = \{F_1, F_2, F_3\}, alg = \{A, B, \dots G\}$

A futtatott algoritmusok a következők (*alg* = {*A, B, C, D, E, F, G*}):

- A* a Praat alapértelmezett maximális formáns beállításával futtatott formánsmérés, nem optimalizált mérés
- B* a Praat 5.1.1. fejezetben részletezett optimális maximális formáns beállításával futtatott formánsmérés, beszélőnként és magánhangzónként 50 véletlenszerűen választott magánhangzó-előfordulás alapján
- C* a *B* algoritmussal kapott formánsértékek kiinduló pontjai voltak a FEF.20 referencia adatbázisnak, így az objektív összehasonlítás miatt egy másik 50 véletlen magánhangzón alapuló optimális maximális formáns beállítással futtatjuk az 5.1.1. fejezetben részletezetteket
- D* ugyanaz, mint a *C* algoritmus, de csak  $F_1$  és  $F_2$  értékeit vesszük figyelembe az optimális paraméter meghatározásakor
- E* ugyanaz, mint a *C*, de normalizált formánsfrekvencia értékeket használtam az optimális maximális formáns meghatározásához (normalizálásról a 6.1. fejezetben lesz szó)
- F* ugyanaz, mint a *C*, de csak magánhangzótípusonként és csak 20 véletlen magánhangzó-előfordulással
- G* ugyanaz, mint a *C* algoritmus, de 100 véletlen magánhangzó-előfordulással

A fent felsorolt 7 algoritmus átlagos abszolút hibáját formánsenként az 5.5. ábra mutatja be. Kiindulópontnak használhatjuk az *A* algoritmus  $AE_{F_1,A}$ ,  $AE_{F_2,A}$  és  $AE_{F_3,A}$  értékeit, amelyeket az 5.5. ábra első oszlopai jelenítenek meg, hiszen ezek alapértelmezett, tehát nem optimalizált bemenő paramétereket használtak a mérés során. A *B–G* algoritmusok mindegyike optimalizált paramétert használ. Ha tehát hatékony az optimalizált paraméterekkel történő formánsmérés, akkor a

*B–G* algoritmusokat szimbolizáló oszlopok alacsonyabbak, mint az *A*-t jelentő első oszlop magassága. A *B* algoritmussal kapott hiba a legkisebb, de mivel a referencia adatbázist *B* algoritmus eredményéből kiindulva kaptuk, ezt nem vesszük figyelembe. Az [Escudero et al.(2009)] dolgozatban leírt optimális maximális formáns beállítással nyert algoritmus értékelésére a *C* oszlop enged következtetni. Az egyes formánsorszámnál megjelenő hibák alig kisebbek, mint az alapértelmezett paraméterbeállításokkal nyert *A* algoritmushoz tartozó hibák. A javulás kis mértéke hívta életre a *D–G* algoritmusokat, amelyek az Escudero és munkatársai által leírt módszer módosításainak tekinthetők. A *D* megnövekedett hibája jelzi, hogy az  $F_3$  nem hagyható ki az optimális paraméterbeállításokból. A legkisebb hiba az *E* algoritmusnál következett be, amely a normalizált formánsfrekvencia értékek használatának fontosságát mutatja. Az *E* algoritmus csökkentette az optimális paraméter meghatározásához szükséges magánhangzó-előfordulások számát. Ennek következménye a legnagyobb előforduló hiba lett. A *G* algoritmus pedig megduplázott magánhangzó-előfordulásokkal dolgozik, amely ennek ellenére sem járult hozzá a hiba lényeges csökkenéséhez.



5.5. ábra. A manuálisan javított FEF.20 adatbázis egyes formánsfrekvencia értékeitől mért átlagos abszolút hiba a különböző algoritmusokkal mért frekvenciaértékekhez képest

### **5.1.3. A kézi ellenőrzés tapasztalatainak összefoglalása**

A kézi javítás tapasztalataiból levonható következtetések az optimális paraméterbeállításokkal végrehajtott mérésre a következők:

- az optimális paraméterek nagy adatbázis esetén is kis mértékben növelik a mérés pontosságát, a legnagyobb javulás a módszer egy módosított változatával, a normalizált formánsfrekvencia-értékekkel valósítható meg
- a mérés során egyaránt kapunk a helyes érték felett és alatt is hibás értékeket, tehát nincs tendencia a hibázásban
- a hibák száma a formánsorszám növekedésével nő
- a hibák száma nagyobb a magánhangzó második felében
- a hibák száma beszélő- és hangfüggő, véleményem szerint inkább függ a beszélőtől, mint a magánhangzótól.

A fentiek ellenére az optimalizált maximális formáns frekvencia értékének helyes megállapítása nagy körültekintést igényel, nagy adatbázisok esetén óvatosan használjuk ezt a megoldást, esetleg egészítsük ki magánhangzó-normalizálással.

## **5.2. A félautomatikus hibajavítás modellezése**

A gépi mérés nyers formánsértékeiben az optimalizált paraméterek használata ellenére előfordulnak hibák. Az előző fejezet manuális hibajavítási lépése alapján ez a mért formánsok 3,35%-át érinti. A FEF adatbázis esetében ez kb. 167 ezer formánsértéket jelent. A nagy tömegű formánsadat vizsgálatához, a hibás formánsértékek felderítésére olyan gépi detektáló eljárást dolgoztam ki, amely megbízhatóan csak a hibásnak vélt helyeket mutatja meg és ajánlja fel vizuális vizsgálatra, manuális vagy gépi javításra.

### **5.2.1. Hibás és gyanús esetek gépi detektálása**

A korábbi, internetes magyar referencia formánsadatbázis [Olaszy et al.(2009)] és [Abari & Olaszy(2011c)] létrehozásának tapasztalataiból kiindulva tipizáltam a formánsmérések lehetséges tévesztéseit és a következő hibatípusok automatikus felismerését tűztem ki célul:

- h1 hiányzó adat: a formánsmérő algoritmus nem szolgáltat eredményt az adott mérési ponton (például az ejtésből adódóan),

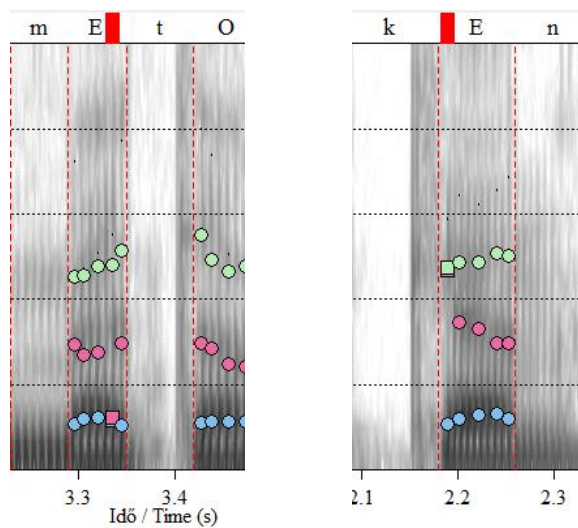
- h2 formánstávolság hiba: túl közel kerülhetnek egymáshoz az azonos mérési ponthoz tartozó formánsértékek számadatai (például az  $F_2-F_1$  különbség kicsi),
- h3 formánsorszám-tévesztés: formánsok között egy adott mérési pontban elcsúszás keletkezik valamelyik irányban: például hamis formáns detektálásakor (fantomformánsokat mér a rendszer), melyek hatására a valódi  $F_2$ -érték  $F_3$ -ra kerül; vagy ha a formánsmérő nem ismer fel egy formánst, akkor a valódi  $F_3$ -érték  $F_2$ -re kerül,
- h4 kiugró adat: a várt határértéken kívülre kerül a mért formánsérték (számítási pontatlanság, ejtési sajátosság, zavaró esemény),
- h5 hibás formánsmenet: a magánhangzóban megvalósuló formánsmozgás tendenciájából kilóg egy-egy mérési ponton a formánsérték, azaz nem felel meg a hangkörnyezet figyelembe vételével várt formánsmenetnek .

A fenti hibák gépi detektálása nem egyforma nehézségű. Míg az első két hibatípus (h1, h2) felismerése könnyen automatizálható, a többi (h3–h5) megvalósításához segéd táblázatokra és/vagy bonyolultabb algoritmusokra van szükség.

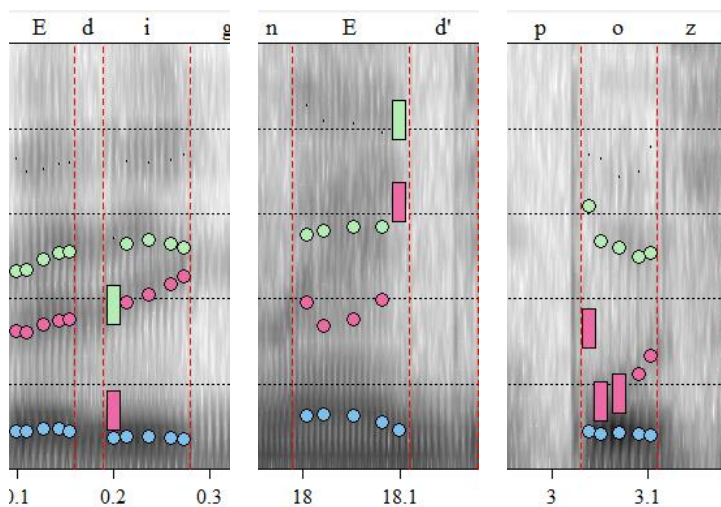
A hiányzó adatok (h1) a formánsmérő algoritmusok szöveges eredménylistájában speciális karaktorsorozattal szerepelnek, tehát könnyen felismerhetők. A Praat az „-- undefined --” címkét használja a hiányzó formánsérték jelölésére.

A túl közeli formánstávolságok (h2) felismerése küszöbszámok rögzítése után szintén eredményesen automatizálható. Az  $F_2-F_1$  távolságra az 50 Hz-es, az  $F_3-F_2$  különbségre pedig a 100 Hz-es küszöbszámot határoztam meg, amelyen belül a formánsértékeket túl közelinek minősítettem. Az 5.6. ábra formánstávolság hibára mutat két példát.





5.6. ábra. Gépi mérési hibák gépi detektálása. Formánstávolság hiba. A bal oldali ábrán az  $(F_2-F_1)$  távolság, míg a jobb oldali részen az  $(F_3-F_2)$  távolság túl kicsi. A függőleges tengely a frekvencia, amely 0–5000 Hz-ig jelenik meg



5.7. ábra. Gépi mérési hibák gépi detektálása. A hibákat téglalap jelöli. A bal oldali ábrán fantomformáns jelenik meg az [i]-ben. A középső ábra nem helyesen felismert  $F_2$  és  $F_3$  formánsmutat példát a magánhangzóban. A jobb oldali ábrán  $F_2$ -ben formánsmenethibára látunk példát. A függőleges tengely a frekvencia, amely 0–5000 Hz-ig jelenik meg

A sorszámtevesztés (h3) felismeréséhez a kulcsot két egymást követő, azonos irányba kiugró formánsérték adja ugyanazon a mérési ponton. Az 5.7. ábra bal

oldali részében példát láthatunk egy ilyen esetre, amikor egy fantomformánst észlel a mérő algoritmus az [i]-ben, a 10%-os mérési ponton a valódi  $F_1$  és  $F_2$  formánsérték között. Így frekvenciában lefelé tolódik a formánsok emelkedő sorrendje  $F_1$  felett. Az ábra középső részében a magánhangzó 5. mérési pontján az  $F_2$  elméleti helyén (2100 Hz) nem detektált formánst a program, így frekvenciában felfelé tolódik a formánsok emelkedő sorrendje. A sorszámtevesztés (h3) és a következő két hibatípus, a kiugró adatok (h4) és a hibás formánsmenetek (h5) felismerését az nehezíti meg, hogy szükség van a kiugró értékek pontos azonosítására. Azokat a formánsfrekvencia értékeket kell kiugróként megjelölnünk, amelyek a formánsorszám által meghatározott lehetséges frekvenciaintervallumba már nem férnek bele.

#### 5.2.1.1. Döntés a kiugró értékekről

A kiugró érték statisztikai értelemben feltűnően nagy vagy kicsi értéket jelent, amelynek szokásos definíciója az alsó és felső kvartilisre támaszkodik. Ha  $Q_1$  az alsó,  $Q_3$  a felső kvartilist,  $IQR$  pedig a kettő különbségét ( $Q_3 - Q_1$ ) jelöli, akkor a

$$[Q_1 - 1,5 \cdot IQR; Q_3 + 1,5 \cdot IQR] \quad (5.2)$$

intervallumon kívül eső értékeket kiugrónak minősítjük [Chambers et al.(1983): 21]. Az (5.2) egyenlettel meghatározott intervallumot *statisztikai intervallumnak* fogom nevezni.

A statisztikai intervallum határai meghatározhatók a pontos formánsadatokat tartalmazó magyar referencia formánsadatbázisból [Olaszy et al.(2009)] és [Abari & Olaszy(2011c)], a FEF.20 kézzel javított adatbázisból, vagy akár a nyers mért értékeket tartalmazó FEF adatbázisból is.

A kiugró értékek azonosítására másik lehetőség a szakirodalomban fellelhető határértékek figyelembe vétele lehet. A magyar szakirodalomban több helyen találunk „elméleti” minimum- és maximumértékeket nemenként, magánhangzónként és formánsokként csoportosítva. Az ott szereplő értékeknél szélsőségesebb eseteket minősíthetjük kiugrónak. A szélsőértékeket a pontos formánsértékeket tartalmazó magyar referencia formánsadatbázisból vagy a FEF.20 adatbázisból is kinyerhetjük. Az így meghatározott intervallumokat (*pontos*) *szélsőértéken alapuló intervallumnak* nevezem.

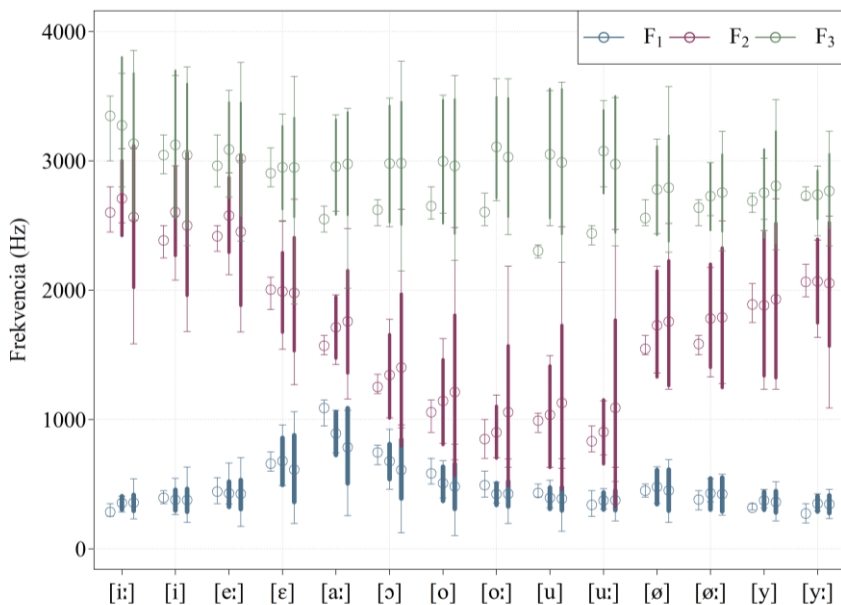
Az 5.8., 5.9. és 5.10. ábrák, a fent felsorolt lehetőségeket mutatják be női beszélők adatai alapján. Mindegyik ábra 3-3 különbözőképpen összeválogatott frekvenciaérték együttes származtatott mutatóit hasonlítja össze. Az ábrákon a mintaátlagot körrel, a szélsőértéken alapuló intervallumokat talpak közötti vékony vonallal és a statisztikai intervallumot vastagított vonallal jelöltem.

Az 5.8. ábra sorrendben (1) [Magdics(1965)] adatait, (2) a referencia formánsadatbázis 50%-os mérési pontjából illetve (3) az összes mérési pontjából származó adatokat tartalmazza. Az 5.9. ábra rendre (1) a referencia formánsadatbázis 50%-os mérési pontjából származó adatokra, (2) a FEF.20 kézzel mért adatbázis 50%-os pontjára, illetve (3) az összes mért pontjára támaszkodó származtatott értékeket tartalmazza. Az 5.10. ábrán ezek szerepelnek: (1) a FEF.20 kézzel mért adatbázis 50%-os adatai, (2) a FEF adatbázis 50%-os és (3) az összes mérési pontjából származó adatai.

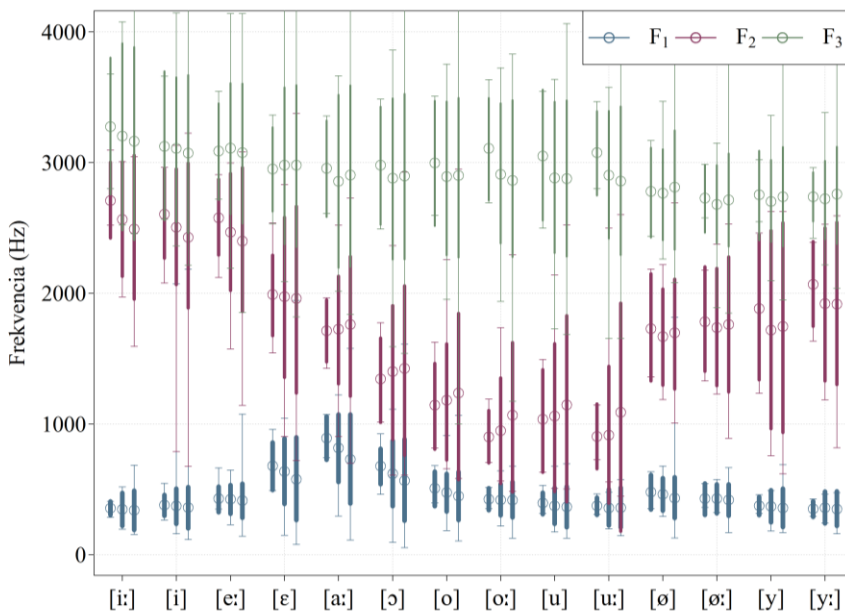
Az 5.8.(1) esetén csak mintaátlag és szélsőértékek jeleníthetők meg az ábrán, adatminta hiányában statisztikai intervallum nem rajzolható. Az 5.8., 5.9. és 5.10.(1) ábra kézzel mért adatokon alapul, csak az 5.10.(2-3) támaszkodik nyers formánsértékekre.

Az 5.8. ábráról leolvasható, hogy az évtizedekkel korábbi szakirodalmi adatok felhasználása nehezen képzelhető el a kiugró értékek feltárásában [vö. Gósy(2012)]. Az 5.8.(1) intervallumai lényegesen kisebbek a referencia formánsadatbázison alapuló (2-3) intervallumoknál. A több beszélőn alapuló, FEF.20 adatbázisból származó intervallumok pedig tovább nőnek, ahogyan az az 5.9.(1) és 5.9.(2-3) összehasonlításából leolvasható.

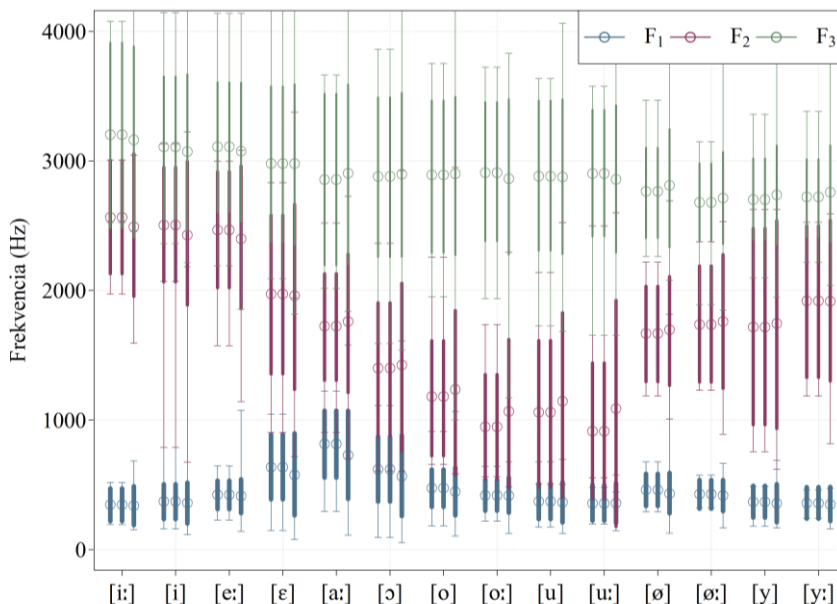
Ha az 50%-os mérési ponton túl, több mérési pontra támaszkodva számoljuk ki a lefedő intervallumokat, akkor az rendszerint szélesebb lesz. Ha összevetjük az ábrák (2.) és (3.) adatsorait, akkor látható, hogy ez minden esetben teljesül. Az 5.10. ábrán a kézzel mért adatokon (1) alapuló intervallumok és a nyers adatok összehasonlítására van lehetőség. Látható, hogy a statisztikai intervallumok alig változnak, míg a szélsőértéken alapuló intervallumok (2-3) esetében nagyon megnyúlnak.



5.8. ábra. Női beszélők formánsfrekvencia értékei. Magdics, RFA 50%, RFA összes.



5.9. ábra. Női beszélők formánsfrekvencia értékei. RFA 50%, FEF.20 50% és FEF.20 összes.



5.10. ábra. Női beszélők formánsfrekvencia értékei. FEF.20 50%, FEF 50% és FEF összes.

Bármelyik módszert is választjuk, a küszöbértékeket tartalmazó segédtablázat minimálisan  $2 \times 14 \times 3 \times 2 = 168$  elemet tartalmaz (beszélők neme: 2; magánhangzók száma: 14; formánsok száma: 3; szélsőértékek száma: 2). Azonban ennél jóval nagyobb elemszámú segédtablázat is készíthető – amennyiben bemondókra és/vagy hangkörnyezetekre csoportosítunk –, de csak abban az esetben, ha az adatbázis mérete megengedi, azaz ha a finomabb felbontás kategóriái elegendően nagy elemszámúak (például legalább 20 eleműek).

A további vizsgálatokban a kiugró értékről a FEF.20 adatbázis szélsőértékei alapján hoztam döntést: amelyik formánsfrekvencia érték nincs az így tekintett intervallumban, kiugrónak minősült. Ennek legfontosabb oka, hogy a kézzel mért adatbázis esetében az statisztikai intervallum rendszerint kisebb, mint a szélsőértéken alapuló.

### 5.2.1.2. Döntés a hibás formánsmenetekről

Az elvárt formánsmenettől való eltérés ( $h_5$ ) detektálásához a hangátmenetek várt formánsmeneteinek ismeretére van szükség. A magyar magánhangzók körében már történtek vizsgálatok [Abari & Olaszky(2012)], melyek eredményeiből a detektáló algoritmus számára egy másik segédtablázatot állítottam össze.

Az elvárt formánsmenettől való eltérés ( $h_5$ ) detektálására (5.7. ábra c) rész) kidolgozott megoldásunkban a mért V első felére (CV, VV), valamint külön a

másodikra (VC, VV) határoztuk meg a jellemző mozgásokat, és ezek segítségével végzünk becslést. Öt formánsmozgási kategóriát különböztettünk meg a magánhangzó feléig (a 10%, 25%, 50% mérési pontokból számolva), illetve a felétől a végéig (az 50%, 75%, 90% mérési pontokból számolva) terjedő hangszakaszra: szinttartó=0, növekvő=+, enyhén növekvő=(+), csökkenő= – és enyhén csökkenő= (–). A hangátmenetek várt formánsmeneteit a referencia formánsadatbázis [Abari & Olasz(2011c)] alapján határoztam meg. Például az 5.5. táblázat a CV hangátmenetek várt formánsmozgásait tartalmazza F<sub>1</sub>-re.

5.5. táblázat. A CV hangátmenetben elvárt formánsmozgások az F<sub>1</sub>-ben a C képzési helyének függvényében. BL: bilabiális, LD: labiodentális, AL: alveoláris, PAL: posztalveoláris, PA: palatális, VE: veláris. Két menet megadása esetén, például 0(+), bármelyik előfordulását elfogadjuk

F <sub>1</sub>	[a:]	[ɔ]	[o, o:]	[u, u:]	[e:]	[ɛ]	[ø, ø:]	[y, y:]	[i, i:]
<b>BL</b>	+	+	0(+)	0	(+)0	0	0(+)	0	0
<b>LD</b>	+	+	0(+)	0	(+)0	0	0(+)	0	0
<b>AL</b>	+	(+)0	0	0	(+)0	0	0(+)	0	0
<b>PAL</b>	+	+	(+)0	0	(+)0	0	(+)0	0	0
<b>PA</b>	+	+	0(+)	0+	(+)0	0(+)	+	0	0(–)
<b>VE</b>	+	+	0(+)	0+	+	(+)0	(+)0	0	0

### 5.2.1.3. Hibás és gyanús eseteket tartalmazó szöveges adatállomány

A hibás és gyanús esetek detektálásának eredménye egy szöveges állomány, amely azokról a magánhangzókról és hozzá tartozó mérési pontokról tartalmaz információt, amelyekben a fenti h1–h5 hibák valamelyikét detektáltuk. A szöveges állomány az ezt követő lépés, a vizuális ellenőrzés és hibajavítás kiinduló pontja lesz, azaz a PROFEF vizuális formánselemző bemeneti állománya.

A szöveges állomány szerkezeti felépítését az 5.6. táblázat mutatja. Az állomány *file* oszlopában a beszédminta azonosítója szerepel. A *label* a magánhangzó szimbóluma SAMPA jelöléssel. A *num* azonosítja a beszédmintán belül a magánhangzó pozícióját: hányadik szegmentált elem a sorban. A *time* egy pont az időtengelyen (másodpercben mérve, tízezred pontossággal), amely a mérés helyét jelöli meg. A *pos* egy azonosító, amely egy magánhangzón belül a mérési pontokat különbözteti meg (például a 10 a 10%-os mérési pontot jelenti). Az eddigi oszlopok a hiba pontos helyét jelölik ki. Az *f* oszlop a formáns sorszámát adja meg, a *type* pedig a hiba jellegét mutatja. Az utolsóként szereplő *predict*

oszlop egy becslést ad a hibás érték helyettesítésére. Ezt a PROFEF programban jóváhagyhatjuk vagy felül is bírálhatjuk. A becslő algoritmus működését az 5.2.1.4. fejezetben részletezem.

5.6. táblázat. A hibás vagy gyanús gépi formánsméréseket tartalmazó szöveges állomány fejléce és első 6 sora. A type oszlopban a hiba típusa, az előtte lévő oszlopokban a hiba helye, a predict oszlopban pedig egy becslült érték található, amely akár automatikusan is alkalmazható a javításra

file	label	num	time	pos	f	type	predict
f_og_0102	i	32	2,5070	10	f2	MENETHIBA	2216
f_og_0102	i	32	2,5175	25	f2	MENETHIBA	2211
f_og_0102	i	32	2,5350	50	f2	MENETHIBA	2141
f_og_0102	O	8	0,7862	50	f1	MENETHIBA	475
f_og_0102	O	8	0,8057	75	f1	MENETHIBA	455
f_og_0102	O	8	0,8174	90	f1	MENETHIBA	437

#### 5.2.1.4. Becslő algoritmus a javításra

A kézi hibajavítás fáradságos munkáját nagy adatbázisok esetén szükséges lehet gépi módszerrel is megtámogatni, mellyel nagy tömegű formánsadat automatikus javítására nyílik lehetőség. Ezért a hibák felismerésével egyidőben a helyes értékre vonatkozó becslést is meg kell adnia. Modelletem úgy alakítottam ki, hogy a hibásnak vélt adat feljegyzése mellett a becslült érték is bekerül a hibákat tartalmazó állományba (5.6. táblázat *predict* oszlopa).

5.7. táblázat. A becslült értékeket tartalmazó 3. segédtáblázat néhány sora

Beszélő	Magánhangzó (SAMPA kód)	Mérési pont (%)	Előző hang kategóriája	Követő hang kategóriája	Becslült F <sub>1</sub> (Hz)	Becslült F <sub>2</sub> (Hz)	Becslült F <sub>3</sub> (Hz)
N3	2	10	#	C.3	432	1676	2762
N3	y	50	C.2	C.10	367	1566	2739
N3	E	90	C.1	#	260	1705	2728
N3	E	50	C.3	C.9	545	2086	2978
N3	u:	75	C.6	#	413	688	2861
N3	e:	10	C.7	C.5	379	2217	3011
N3	i	10	C.10	C.10	376	2469	3053
N3	o	50	C.4	C.3	437	1370	2764
N3	O	50	C.2	#	651	1319	2847
N3	o:	90	C.6	C.3	414	1520	2994
N3	i	90	V.1	C.4	301	2543	3133

A becsült értékek mintaátlagok, amelyeket a FEF adatbázis kézzel javított formánsértékeiből (FEF.20) határoztam meg. A becsült értékeket tartalmazó segéd táblázat átlagolással nyert értékeinél figyelembe vesszük a beszélő személyét, a magánhangzót, a mérési pozíciót és a közvetlen hangkörnyezetet is. A 5.7. táblázat a segéd táblázat néhány sorát mutatja.

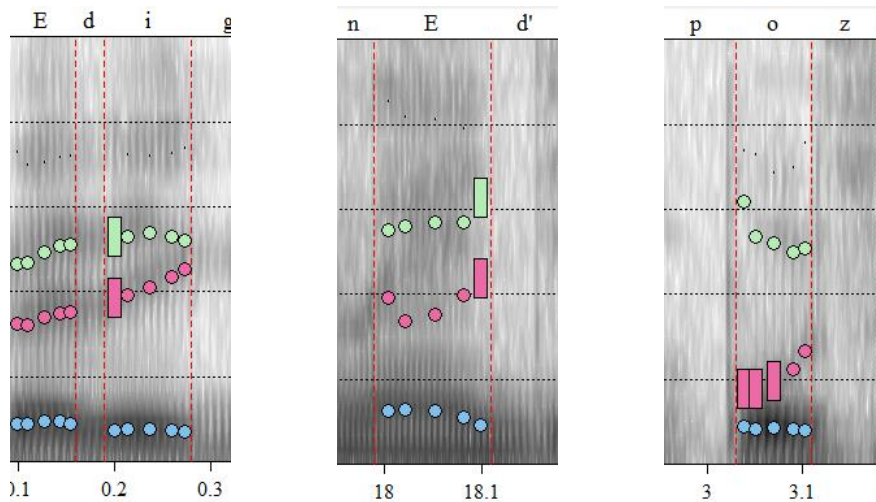
Amennyiben az N3 beszélő [y] hangjának 50%-os mérési pontjában mért becsült értékre van szükségünk, olyan hangkörnyezetben, amelyben a megelőző hang C.2 típusú a követő pedig C.10 kategóriájú, akkor a segéd táblázatból az F<sub>1</sub>-re a 367 Hz-et használjuk a becslésre (5.7. táblázat 2. sor). Mivel a FEF.20 nem fed le minden hanghármaszt (például nincs V.5[ɔ]C.6 vagy C.6[ε]C.3 hanghármas), így a hangkörnyezet hatását nem lehetséges minden esetben érvényesíteni a segéd táblázat felhasználása során. Ha tehát becslést szeretnénk adni az [ɔ] hangra V.5;C.6 környezetben, akkor első körben a megelőző környezetet hagyjuk el, és ez alapján keresünk bejegyzést a segéd táblázatban. Ha sikerrel járunk, akkor a találatok átlaga lesz a becsült érték. Ha nincs ilyen bejegyzés, akkor a követő hangkörnyezetet hagyjuk ki a feltételekből, de ha így sincs találat, akkor mindkét környező hangot kihagyjuk.

Sorszámtevesztés esetében, mivel többnyire a helyes értékek elcsúszásáról van szó, a becsült értékek közé a hibás sorszámmal felismert, de értékükben pontos, sorszámukban korrigált formánsértékek kerülnek.

### **5.2.2. A formánsértékek vizuális ellenőrzése, korrekciója**

A formánsérték vizuális ellenőrzésére és félautomatikus korrekciójára kifejlesztett alkalmazást a 3.4. fejezetben ismertettem. A PROFEF a FEF adatbázis hangmintáira, annotációira és az előre elkészített spektrogramokra épül. További bemenő paraméter a hibásként felismert formánsértékeket tartalmazó szöveges állomány, amely lehetővé teszi a hibás és nem hibás formánsfrekvenciák vizuális elkülönítését, és kézi vagy a becsült értékek alapján az automatikus javítását. Ez utóbbi esetben egyetlen lépésben egyszerre sok hibát korrigálhatunk. A becslés elfogadásával végrehajtott javítás eredményét láthatjuk az 5.11. ábrán. Az így vizsgált bemondásokhoz már előáll a precíz formánsadatokat tartalmazó adatállomány.





5.11. ábra. A becslés gépi elfogadásával kijavított formánsadatok (az 5.8. ábra alapján). A függőleges tengely a frekvencia, amely 0–5000 Hz-ig jelenik meg

### 5.2.3. A formánsadatbázis szerkezete

A formánsadatokat tabulátorral tagolt szöveges állományban tárolom, melynek felépítése az 5.8. táblázatban látható.

5.8. táblázat. Néhány sor a formánsadatbázisból

file	label	num	time	pos	f1	f2	f3
f_og_0101	o:	3	0,3893	10	401	1607	2246
f_og_0101	o:	3	0,4096	25	440	1266	2267
f_og_0101	o:	3	0,4435	50	456	1025	2445
f_og_0101	o:	3	0,4774	75	457	1046	2602
f_og_0101	o:	3	0,4978	90	282	1081	2425

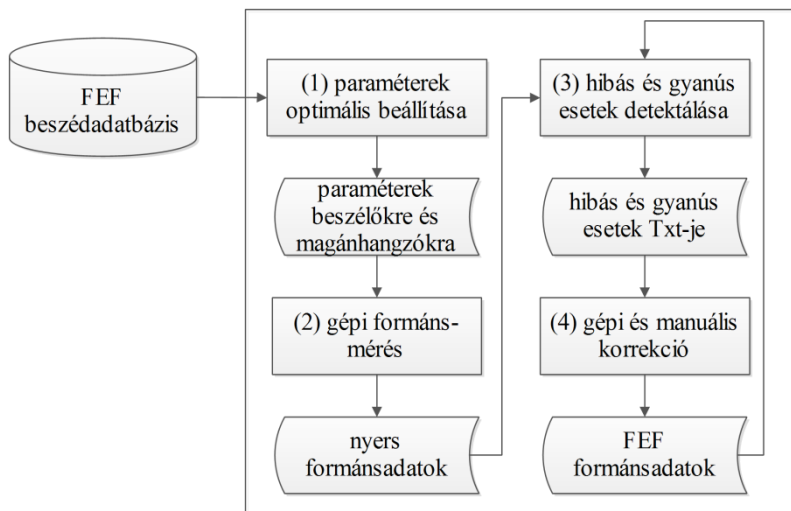
A formánsadatokat tartalmazó szöveges állomány *file* oszlopában a beszédminta azonosítója szerepel. A *num* azonosítja a beszédmintán belül a magánhangzó pozícióját: hányadik szegmentált elem a sorban. A *label* a magánhangzó szimbóluma SAMPA jelöléssel. A *time* egy pont az időtengelyen (másodpercben mérve), amely a mérés helyét jelöli meg. A *pos* egy azonosító, amely egy magánhangzón belül a mérési pontokat különbözteti meg (esetünkben például a 10 a 10%-os mérési pontot jelenti). Az utolsó három oszlop (f1–f3) az  $F_1$ – $F_3$  formánsfrekvencia értékeket sorolja Hz-ben kifejezve.

Jelen pillanatban a formánsadatbázis egyetlen tabulátorral tagolt szöveges állományban foglal helyet, melynek mérete kb. 110 MB. Amennyiben még

nagyobb adatbázissal szeretnénk dolgozni és a megnövekedett formánsadatbázis kezelése nehézséget okozna, akkor hangmintánként feldarabolhatjuk a formánsadatbázis sorait, több különálló állományba. A hangminta, az annotáció és a formánsadatok kapcsolatát – az Emu adatbázisfelépítésének megfelelően – a közös állománynév biztosítja. A formánsadatbázist alkotó állományok szerkezete ettől nem változik.

#### **5.2.4. Lépcsőzetes eljárás hibafelismerésre és hibajavításra**

Az ellenőrzött formánsadatok előállítására kidolgozott algoritmusom lépéseit az 5.12. ábra foglalja össze. A paraméterek optimális beállítása (1) és a gépi formánsmérés (2) nyers formánsadatokat eredményez. A hibás és gyanús esetek detektálása (3) és a gépi és manuális korrekció (4) lépések végrehajtása azt eredményezi, hogy a FEF beszédadatbázist megbízható formánsadatokkal bővíthetjük ki. A (3) és (4) lépéseket, vagyis a vélt hibák keresését, illetve javítását több fokozatban érdemes végrehajtani (lásd a két lépés közötti visszacsatolást az 5.12. ábrán). Ez azt jelenti, hogy első lépésben a h1 típusú hibákat keressük, majd javítjuk, ezt követi a h2 típusú hibák felismertetése és javítása és így tovább. A lépcsőzetes hibafelismerés és -javítás egyik előnye, hogy az egyes hibatípusok keresése és javítása leegyszerűsödik, csak az éppen vizsgált hibatípusra kell figyelniük. Például a h3 formánsorszám-tévesztés javítása során már nem várunk hiányzó értékből vagy túl közeli formánsokból adódó hibát, vagy a h5 menetvizsgálat esetén már semmilyen h1–h4 hibára nem számítunk. Végeredményben az ötféle hibatípusnak megfelelően, egymás után ötször hajtjuk végre a detektálás (3) és hibajavítás (4) lépéseket. Az eljárás másik előnye, hogy minden hibadetektálás során a becslést adó formánsfrekvenciát a formánsadatbázis pillanatnyi állapota alapján újraszámolhatja az algoritmusunk, így a hibák fokozatos csökkenésével egyre pontosabb értékekre támaszkodhatunk a félautomatikus hibajavítás során. Ez a lehetőség akkor jelent igazán előnyt, ha nem tudunk kézi mérések eredményeire támaszkodni. Esetünkben a kézi méréseket tartalmazó FEF.20 adatbázis rendelkezésre áll, így ez utóbbi lehetőséget nem használjuk ki.



5.12. ábra. Az ellenőrzött formánsadatok előállításának algoritmusja

### 5.3. A félautomatikus hibajavítás értékelése

A bemutatott gépi formánshiba-felismerő eljárást kétféle megközelítéssel értékelem. Egyrészt kiszámolom, hogy a felismert hibák hány százaléka tényleges hiba (helyesség), másrészt a tényleges hibák felismerésének arányát is megadom (teljesség). Az elemzéshez a FEF.20 formánsadatait használom fel, mivel abból a nyers példány (FAD.1, formánsadatok első változat) és a manuálisan javított változat (FAD.2) is rendelkezésre áll. Az objektivitás miatt a hibakereséshez szükséges szélsőértékeket a FAD.2 formánsadatainak 20%-a alapján határoztam meg. Ennek az az oka, hogy a teljes FEF-re vonatkozó hibajavítás a manuálisan javított FEF.20-ra fog épülni.

A hibakeresés folyamatában az egyes hibatípusok felismerésének helyességét az 5.9. táblázat tartalmazza. A táblázat adatai azt mutatják meg, hogy a nyers formánsokat tartalmazó FAD.1-ben a gépileg felismert hibák milyen aránya tényleges (FAD.2-ben kézzel javított) hiba. FAD.1-ben hiányzó érték (h1) nem fordult elő, de ennek a legegyszerűbb hibatípusnak a helyessége és a felismerési pontossága is egyaránt 100%-os. Az algoritmusom által felismert formánstávolság-hiba 99,7%-a takar tényleges hibát. A formáns-sorszám hiba felismerésének helyessége 83,6%, a kiugró adatoké 88,5%. Ez utóbbi két hibafajta detektálása nagymértékben támaszkodik a szélsőértékeket tartalmazó segéd táblázatra. A hibás formánsmenet felismerésének helyessége alacsonynak mondható. A hibaként felismert esetek kb. 65%-ában a kézi átvizsgálás során jóváhagytuk a formánsmozgást. Ez azt mutatja, hogy a hibás formánsmenetek

detektálásához finomabb mássalhangzó-kategóriák és szélesebb hangkörnyezet figyelembe vétele szükséges.

5.9. táblázat. A hibafelismerés helyessége

Hiba típusa	$100 \cdot \frac{\text{felismert tényleges hiba}}{\text{felismert hiba}} (\%)$
h1 - hiányzó adat	-
h2 - formánstávolság hiba	99,7
h3 -formánsorszám-tévesztés	83,6
h4 - kiugró adat	88,5
h5 - hibás formánsmenet	35,8

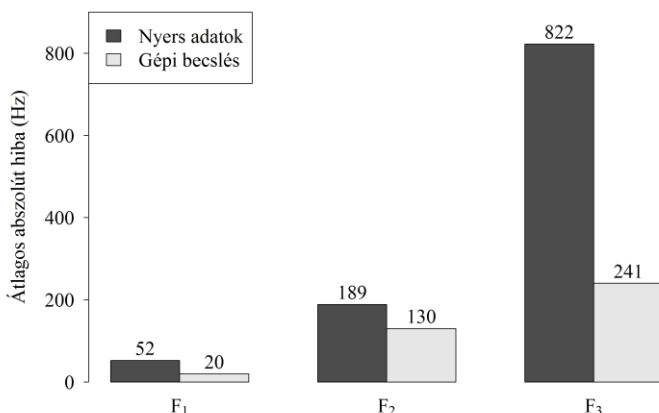
A gépi formánshiba-felismerő megítélésének másik aspektusa a teljesség, ami azt határozza meg, hogy a tényleges hibák hány százalékát sikerült felismerni. Mindhárom formásra nézve ez az arány 42%, az első két formásra pedig 66% a felismerés pontossága (az első három formásra ezek rendre: 74%, 62% és 15%). A harmadik formánsban mutatkozó hibák felismerésének sikertelensége nagyon lerontja az algoritmus hatékonyságát. A kézi hibajavítás során a formánshibák frekvencia értékeit az y koordináta (Hz) mentén változtattuk, emeltük vagy csökkentettük. A módosítások előjel nélküli átlagát tartalmazza az 5.10. táblázat, az első három formásra. Külön sorban szerepelnek a hiba-felismerő algoritmusom által sikeresen felismert és fel nem ismert hibák átlagai. A táblázatból leolvasható, hogy a felismerő képes a nagyobb frekvencia-eltéréseken alapuló hibák gépi detektálására, de a finomabb tévesztések egy részét nem találja meg.

5.10. táblázat. A manuálisan javított hibák átlagos javítása frekvenciában (Hz)

	F <sub>1</sub>	F <sub>2</sub>	F <sub>3</sub>
Felismert hiba	546	625	869
Nem felismert hiba	293	451	476

Végül megvizsgáltam, hogy az automatikus becslések milyen mértékben segítenek a korrekt formánssadatbázis előállításában. A formánsmérés során a talált hibák gépi javítására becslések állnak rendelkezésre (5.2.1.4. fejezet), amelyek automatikusan alkalmazhatók a kézi javítás felgyorsítására. Ennek hatékonyságát úgy mértem, hogy a FAD.1 és a FAD.2, valamint a becslött értékek és a FAD.2 között átlagos abszolút eltérést számoltam azokon a pontokon, ahol hibafelismerés történt (ugyanis csak itt áll rendelkezésre automatikusan becslött érték). Az 5.13. ábra oszlopai ezeket az átlagos eltéréseket tartalmazzák az első három formánsban. A gépi becslésekhez tartozó oszlopok alacsonyabbak, ami azt mutatja, hogy a becslések hozzájárulnak a

pontosabb formánsadatbázis előállításához. A legnagyobb arányú csökkenés  $F_3$ -ban figyelhető meg, de  $F_1$ -ben is lényeges közeledés van a kézel javított formánsértékek felé.



5.13. ábra. A manuálisan javított (FAD.2) formánsértékek átlagos abszolút eltérése a nyers formánsmérés eredményétől, illetve a gépi becsléssel kapott formánsadatoktól

A fentiekben leírt félautomatikus hibakereső és -javító eljárás jelentősen hozzájárulhat nagyméretű és ellenőrzött formánsadatbázis létrehozásához. A felismert hibák ténylegesen hibák (az első négy hibatípusra 83% feletti a helyesség), és a hibák nagy százalékát (az első két formánsban 66%-át) valóban felismeri. Az eljárás további javítása a finomabb hibák pontosabb detektálásának megoldását jelenti. Ez a formánsmozgások – jelen értekezésben is tárgyalt (6.2.2. fejezet) – precízebb leírásával és a küszöbértékek még pontosabb beállításával lehetséges.

## 5.4. Félautomatikus hibajavítás a FEF adatbázisra

A FEF adatbázison az 5.12. ábra lépéseit végrehajtottam, amelynek eredménye egy ellenőrzött adatokat tartalmazó nagy formánsadatbázis. A hibadetektálási eljárás előző fejezetben ismertetett sikeressége miatt úgy döntöttem, hogy a vizuális ellenőrzést nem végzem el a FEF megmaradt részére (FEF.20-on túli 80%-ára). Az automatikusan detektált hibákat gépileg javítottam, azaz a hibás frekvenciaértékek javítására a becslési értékeket automatikusan elfogadtattam. A jövőben a FEF félautomatikus ellenőrzését tervezem.

## 5.5. A félautomatikus hibajavítás tapasztalatai, következtetések

Olyan formánsérték-meghatározó eljárást dolgoztam ki, amelyik megbízható pontosságú formánsadatokkal láthat el egy beszédadatbázist. Az eljárás értékelését a FEF.20 adatbázis alapján végeztem, amely 10 beszélő, összesen 66 298 db magánhangzójának  $F_1$ ,  $F_2$  és  $F_3$  formánsán alapul. A kidolgozott formánsmeghatározás gépi és manuális elemeket is tartalmaz. Ez a kombinált vizsgálat teszi lehetővé, hogy a kapott formánsadatok hitelessége garantált legyen. Az eljárás alkalmazása munkaigényes. Eljárásomban a szakirodalomból ismert paraméter-optimalizációs eljárást implementáltam, majd a nyers formánsértékek detektálására és félautomatikus hibajavítására adtam szabály alapú megoldást.

Az 5.12. ábrán összefoglalt formánsérték-meghatározó eljárás sikeresen alkalmazható nagyméretű beszédadatbázisokra. Az ismertetett formánsmérő módszer nagyon keveset követel attól a beszédadatbázistól, amelyet formánsértékekkel szeretnénk ellátni: csupán a .wav állományokra és a hozzá tartozó hangszintű szöveges átiratra, valamint pontos hanghatár címkékre van szükség. A kifejlesztett gépi hibadetektáló algoritmus, valamint a javítást támogató becslő eljárás pontossága biztosítja, hogy az eljárás végére kialakuló formánsadatok és formánsmenetek pontosnak tekinthetők, tehát referenciának elfogadhatók. Az eljárás más nyelvekre is könnyen adaptálható.

Az eljárás számos pontján küszöbértékek segítségével ítélem hibásnak vagy gyanúsnak az egyes formánsméréseket. A küszöbparaméterek és a hibakeresés mélyebb kapcsolatának megértése további elemzést igényel.

További eredmény, hogy eljárásom olyan formánsadatokat tartalmazó adatállományt állít elő, amely referenciaként használható, nyilvános adattárként is közreadható (a hozzá tartozó hullámforma és .TextGrid állományokkal együtt). Bárki betekinthez a beszédadatbázis kiválasztott részletének hangsorába, tanulmányozhatja a hanghatárok elhelyezését, megnézheti a formánsadatokat a spektrális képpel együtt, és akár kontroll vagy egyéb méréseket is tervezhet, végezhet. A formánsadatok és a spektrogram együttes megjelenítése a fonetika és a beszédtechnológia oktatásában is jól használható [Abari & Papp(2008)]. A címkerendszer tartalmazza ezen túl a szóhatárok időkoordinátáit is, tehát szószintű egységekre is lehet méréseket végezni például a szó hossza, helyzete stb. alapján.

## 6. FORMÁNSMENETEK VIZSGÁLATA

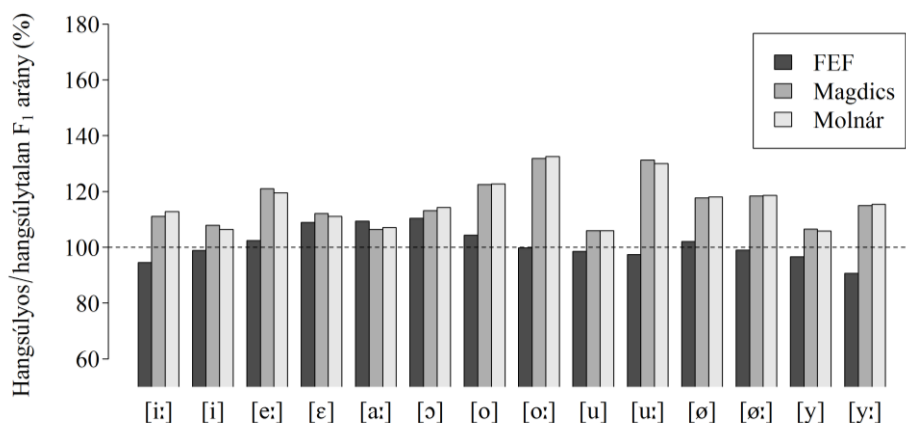
### 6.1. A hangsúly kezelése

A szakirodalom a magánhangzók formánsainak számszerű bemutatásánál hagyományosan különbséget tesz hangsúlyos (H) és hangsúlytalan (NH) helyzet között. Ezért a FEF adatbázis minden magánhangzójához a BME TMIT Beszédakusztikai Laboratórium munkatársaival együttműködve a fenti két címke (H, NH) valamelyikét rendeltem hozzá. A címkézés alapja a Profívox beszédszintetizátor és szövegfelolvasó rendszer szupraszegmentális modulja volt [Olaszy et al.(2000)]. A Profívox a kijelentő mondatok szavait ötféle jelöléssel látja el: fókusz, emfatikus hangsúly, hangsúly, neutrális, negatív hangsúly. Az első három kategóriához a H, a maradék két kategóriához az NH címkéket rendeltük a FEF 1877 db kijelentő mondatában. Az automatikusan megállapított címkéket manuálisan is ellenőriztük és javítottuk, majd a 10 beszélő bemondásait tartalmazó hangfelvételek annotációjában elhelyeztük. A kijelentő mondatok 314 986 db magánhangzójából 51 593 db (16%) kapott H hangsúlycímkét és 263 393 db (84%) NH hangsúlytalan címkét.

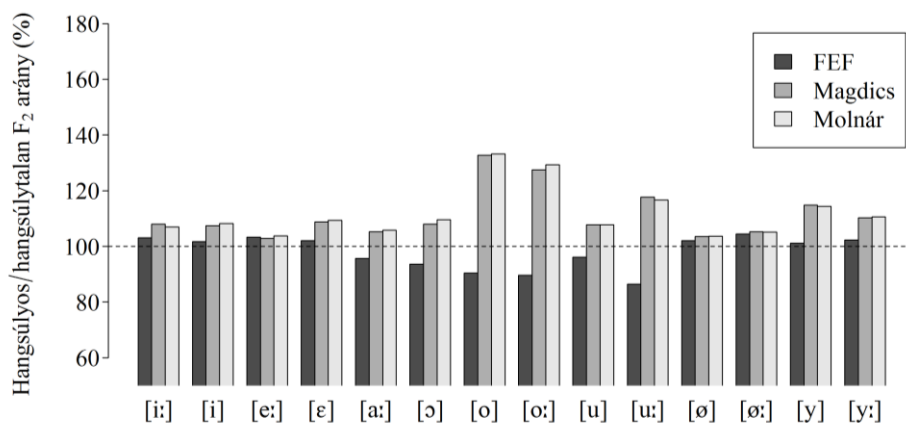
Megvizsgáltam mindkét nemben külön-külön, hogy a FEF adatbázis hangsúlyos magánhangzóikban az 50%-os mérési ponthoz tartozó formánsfrekvencia értékek nagysága hány százaléka a hangsúlytalan helyzetben mértékhez viszonyítva. Női bemondók esetén például, az [i:] -ben  $F_1$  átlaga hangsúlyos helyzetben 331 Hz, míg hangsúlytalan helyzetben 349 Hz. Így a hangsúlyos [i:]  $F_1$  értékének százalékos aránya a hangsúlytalanhoz viszonyítva 95%. Ez az érték olvasható le a 6.1. ábra legelső oszlopáról. Közvetlenül mellette a [Magdics(1965)] és [Molnár(1970)] által mért adatokból kiszámítható hasonló százalékos arány látható szintén a [i:] -re. Ezek 100% fölöttiek, azaz mindkét szerző mérése szerint a hangsúlyos [i:]  $F_1$  értéke magasabb a hangsúlytalantól. Az ábra további részében az [i:] -hez hasonló szerkezetben kapjuk a többi magánhangzóra vonatkozó adatokat. Az összehasonlítás összes vokális esetén elvégezhető, tehát egyrészt a 100%-os pontba húzott vízszintes szaggatott vonallal, másrészt a szakirodalmi adatokkal. A 6.2. és 6.3. ábrán az  $F_2$  és  $F_3$  formánsfrekvencia értékekre látunk oszlopdigramokat, hasonló jelentéssel. A 6.4., 6.5. és 6.6. ábra a férfi beszélők adatait mutatja be. Az ábrákról leolvasható, hogy a két szakirodalmi mérés minden esetben magasabb formánsfrekvencia értéket mutat hangsúlyos esetben a hangsúlytalanhoz képest. Ezzel szemben a FEF adatbázis H-val jelölt magánhangzóiról ilyen nem állíthatunk. A 100%-os

aránytól nagyon kicsi az eltérés, és az eltérés iránya is változó. Ennek az lehet az oka, hogy a felolvasás során nem volt a mondatok előtt felvezetés, kontextusba helyezés. Magdics rövid helyzetképpel vezette be a felolvasott mondatokat [Magdics(1965):4]. Molnár tanulmányában nem szerepel erre vonatkozó adat.

Ezek alapján a FEF adatbázisban a hangsúlyos/hangsúlytalan helyzetek megkülönböztetése a formánsfrekvenciák szempontjából nem indokolt.

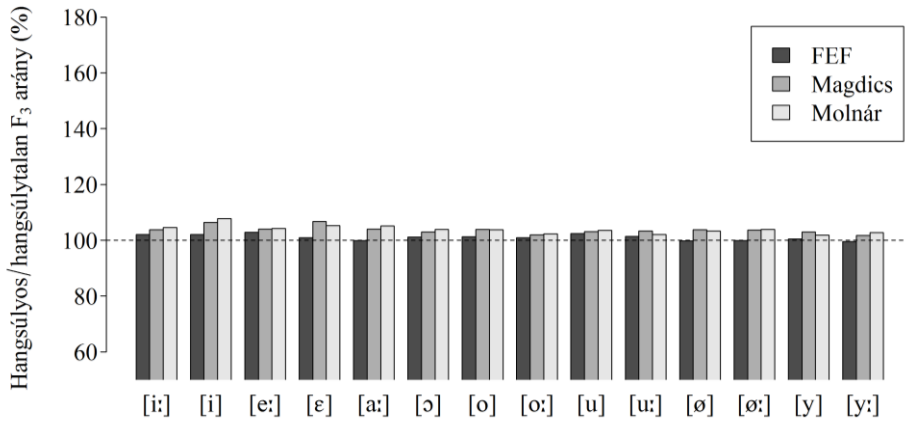


6.1. ábra. Női beszélők esetén a hangsúlyos/hangsúlytalan F<sub>1</sub> arány. A 100%-nál húzott szaggatott vonal feletti oszlopok esetén hangsúlyos pozícióban magasabb a frekvenciaérték a hangsúlytalanhoz viszonyítva

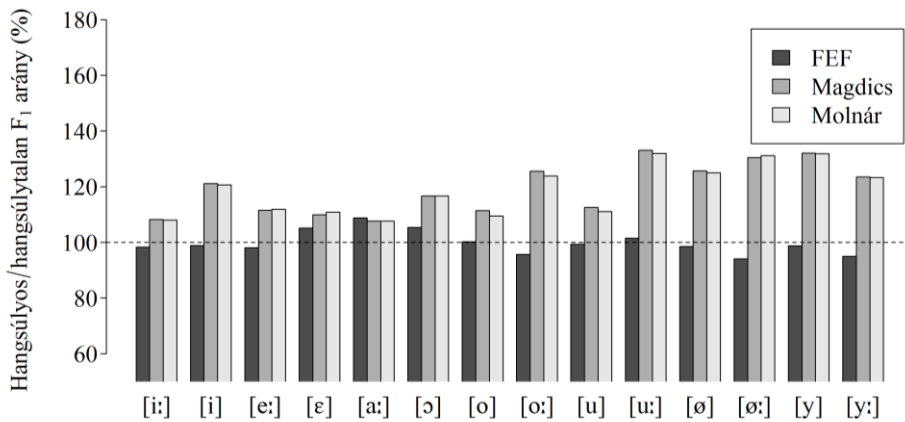


6.2. ábra. Női beszélők esetén a hangsúlyos/hangsúlytalan F<sub>2</sub> arány.

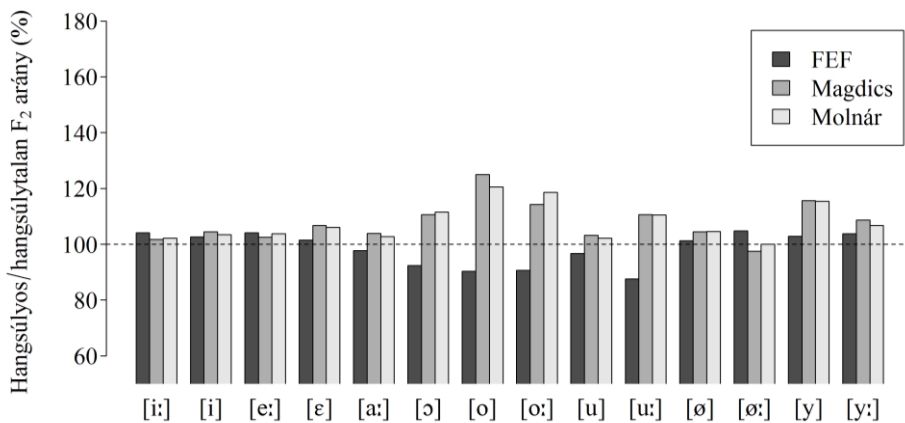




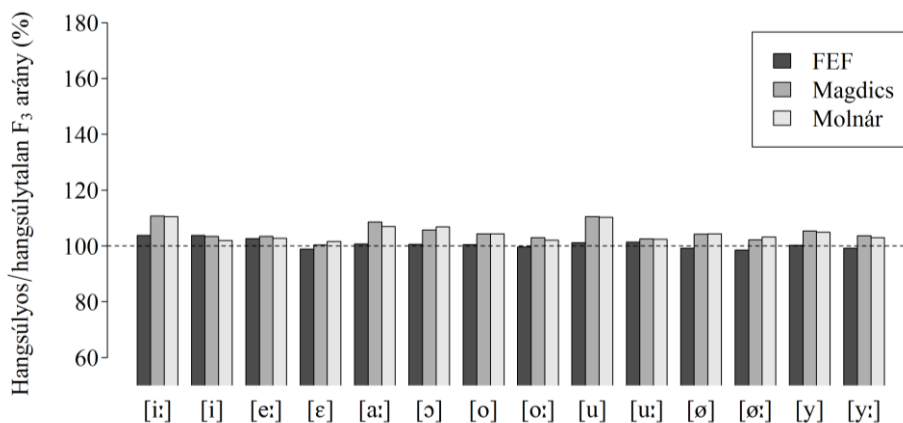
6.3. ábra. Női beszélők esetén a hangsúlyos/hangsúlytalan  $F_3$  arány.



6.4. ábra. Férfi beszélők esetén a hangsúlyos/hangsúlytalan  $F_1$  arány.



6.5. ábra. Férfi beszélők esetén a hangsúlyos/hangsúlytalan  $F_2$  arány.



6.6. ábra. Férfi beszélők esetén a hangsúlyos/hangsúlytalan  $F_3$  arány.

## 6.2. Javasolt módszerek a formánsmozgások kontextusfüggő vizsgálatához

Szisztematikus leírási sémát dolgoztam ki a 3, 4 vagy 5 elemű hangsor központi magánhangzójában az első három formáns mozgásának jellemzésére a FEF adatbázisból származó adatok segítségével. A szisztematikus leírás az összes szimmetrikus  $C_{t,1}V_mC_{t,2}$  hangkapcsolatra kiterjedhet, amelyre elegendő minta van az adatbázisban. A leírás mellett a közvetlen hangkörnyezet (-1. és +1. hang) és a távolabbi hangkörnyezet hatását is vizsgálom (-2. és +2. hang). Továbbá a különböző hangkörnyezetekben lévő formánsmozgások összehasonlítását is bemutatom. A formánsmenetek modellezésére a funkcionális adatelemzés és az SSANOVA módszert használom.

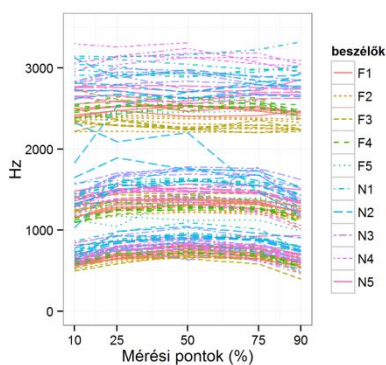
A kitarottnan ejtett magánhangzó formánsai egy adott beszélő adott ejtésében – a toldalécső artikulációs mozgásainak hiányában – közel konstans értéket vesznek fel. A természetes beszéd során a magánhangzó időben változó formánsfrekvencia értékekkel jellemezhető. Ha szigorúan a magánhangzóra jellemző, különböző hangkörnyezetben bekövetkező formánsmozgások rendszerét szeretnénk vizsgálni, akkor a modellben paraméterként csak a magánhangzó és a hangkörnyezet szerepelhet, minden más változót ki kell iktatni. Minden „zavaró” paraméter kiiktatása azonban nem lehetséges, hiszen ismert, hogy egy adott beszélő adott magánhangzójának azonos kontextusban ejtett adatai is többé-kevésbé eltérnek egymástól, a formáns indexszámának emelkedésével egyre nagyobb mértékben [Gósy(2004):118].

A FEF adatbázis egységes abban az értelemben, hogy felolvasott szöveg hangbemondásait tartalmazza, így a beszédmód változóval nem kell

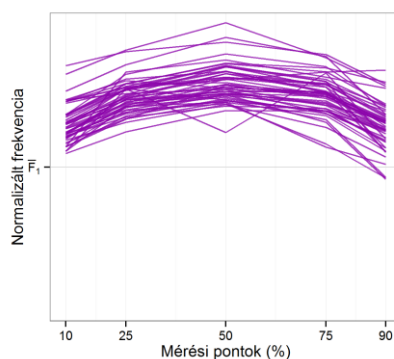
foglalkozunk. A beszélők közötti különbségek (artikulációs tempó, hangszín, felhangtartalom, életkor, nem, pillanatnyi érzelmi állapot stb.) egy része kiküszöbölhető a magánhangzó normalizálással, a megmaradt változók hatásától ebben az értekezésben eltekintünk.

### **6.2.1. Új grafikus ábrázolások**

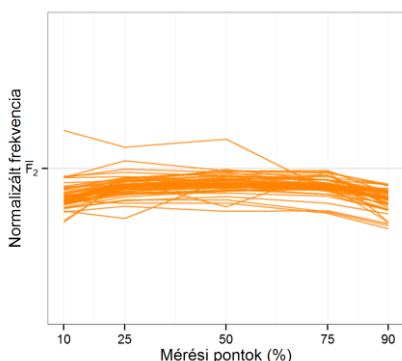
A 6.7. ábra a 3.5. fejezetben részletezett magánhangzó normalizálás folyamatát mutatja be, az adatbázisból származó [a:] hang bilabiális mássalhangzó környezetében. Az a) ábrán a kiinduló állapotot látjuk: a 10 beszélő összes adatát a Hz-ben mért formánsfrekvencia értékekkel. A b), c) és d) ábrák a formánsindexek szerint elvégzett, a normalizálás eredményeképpen kapott formánsmeneteket tartalmazzák. Az y tengely a normalizált formánsfrekvencia értékeket tartalmazza. A középben berajzolt vízszintes vonal az átlagos értéknek felel meg az adott formáns esetében, ami a normalizált formánstérben a 0-t jelenti. A könnyen áttekinthető a) ábra helyett így három különálló ábrát kaptunk a normalizált formánstérben, mivel közös y koordináta nem feleltethető meg a három formánsindexnek. Ennek ellenére egy pusztán technikai lépéssel egymás fölé rajzolhatjuk a normalizált görbéket a formánsindexek növekvő sorrendjében. Ez azt jelenti, hogy az első formáns értékeihez képest 4 egységgel feltolva jelennek meg a második formáns normalizált értékei, a harmadik formáns görbéi pedig 8 egységgel lesznek feltolva az elsőhöz képest (azaz 4 egységgel a másodikhoz képest). A 6.8. ábra a) része tartalmazza az egymásra halmozott normalizált formánsgörbéket. Az ábrán szereplő három vízszintes vonal az egyes formánsátlagoknak felel meg. A normalizálás kiinduló pontját tehát a 6.7. ábra a) része, végpontját pedig a 6.8. ábra a) része tartalmazza.



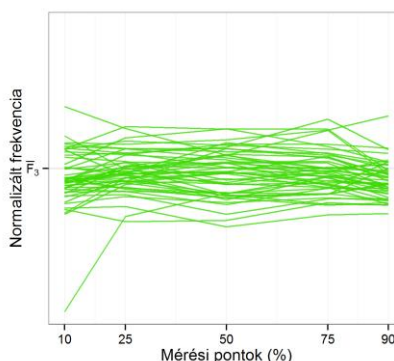
a) ábra



b) ábra



c) ábra



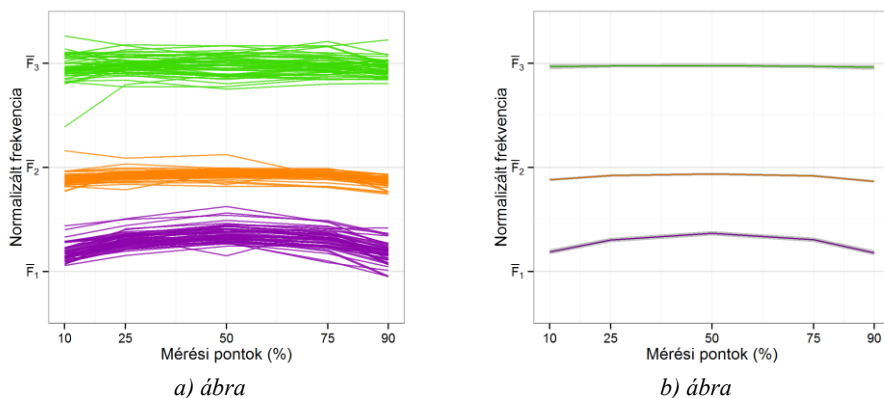
d) ábra

6.7. ábra. A normalizálás folyamata. Az a) tartalmazza az összes beszélő mindhárom formánsának adatait. A b) ábra az  $F_1$  normalizált frekvencia adatokat tartalmazza. A középső vízszintes vonal a normalizált térben 0-nak felel meg, ami az összes  $F_1$  érték átlagát jelenti. A c) és d) ábra hasonlóan értelmezhető  $F_2$ -re és  $F_3$ -ra.

A 6.8. ábra a) része több szempontból is informatív. Egyrészt minden beszélő adatát egyesítve tartalmazza, ami jelentős egyszerűsítést jelent, hiszen nem kell például nem vagy életkor szerinti bontással szaporítani az ábrák számát. A formánsindexek természetes sorrendjükben letről felfelé növekvő sorrendben helyezkednek el. A berajzolt átlagvonalak jelzik, hogy az adott formánsmenetek az átlaghoz képest hol helyezkednek el. Esetünkben az  $F_1$  értékek például jóval az átlag felett foglalnak helyet. A különböző formánsindexekhez tartozó görbék egymáshoz viszonyított helyzete is releváns, arra a kérdésre ad választ, hogy mennyire helyezkednek el közel vagy távol egymáshoz képest az egyes formánsfrekvencia értékek.

Az SSANOVA módszere segítségével a görbéinket egyetlen simított spline-nal helyettesíthetjük, és emellett a görbék szóródására a Bayes konfidencia-intervallum nagyságából következtethetünk. A 6.8. ábra b) része az átlagos

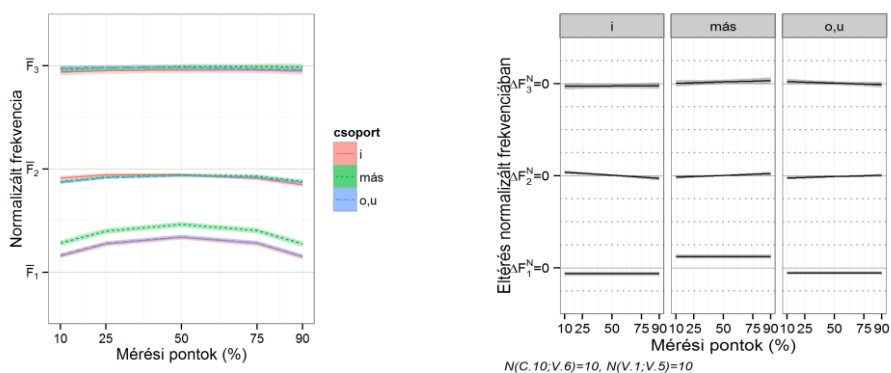
simított spline-t tartalmazza a három formánsra, valamint a köré rajzolt Bayes konfidencia-intervallumot.



6.8. ábra. A normalizált formánsgörbék az a) részben és helyükbe lépő átlagos simított spline-ok és a Bayes konfidencia-intervallumok a b) részben.

A 6.8. ábra b) részében megjelenített spline-ok alkalmasak tetszőleges magánhangzó tetszőleges hangkörnyezetbeli formánsmozgásának leírására. Kiegészítésként az ábra numerikus információkat is megjeleníthet, például a mintaelemek (görbék) számát, vagy a formánsmozgás stabilitására vonatkozó adatokat is.

A különböző hangkörnyezetekben megvalósuló formánsmozgások összehasonlítására szintén alkalmas az SSANOVA modell. Ha példánkat folytatva, a bilabiális mássalhangzók közötti [a:] formánsmozgását vizsgáljuk különböző –2. hangkörnyezetben, akkor a könnyen általánosítható 6.9. ábrát kapjuk. Az összehasonlítandó hangkörnyezetek függvényében különböző csoportokat definiálunk és a csoportok formánsgörbéinek eltérését teszteljük. A 6.9. ábrán bemutatott példa három csoportot definiál a –2. hang előfordulása alapján, ami lehet [i], vagy [o, u] és a harmadik csoport bármilyen más egyéb hang előfordulását jelenti. Az ábra bal oldali részében a görbék köré rajzolt 95%-os Bayes konfidencia-intervallumok átlapolása mutatja meg, hogy eltérnek-e szignifikánsan az egyes csoportok, és ha eltérnek, akkor a görbe melyik részén. A 6.9. ábráról leolvasható, hogy a megelőző [i] valamint [o, u] szignifikánsan csökkenti az  $F_1$  értékét, itt ugyanis a konfidencia-intervallumok nem fedik egymást. Az  $F_2$ -ben a magánhangzó kezdő fázisában mutatkozik [i] hatására némi frekvencia-emelkedés.  $F_3$ -ban nincs eltérés a három csoport tekintetében.



6.9. ábra. Az SSANOVA grafikus szemléltetése három görbecsoport esetén a három formásra. Szignifikáns különbség mutatkozik a  $F_1$ -ben, ugyanis az [i] és [o, u] egységesen csökkenti az első formáns értékeit a többi -2. pozícióban lévő hanghoz viszonyítva.

A 6.9. ábra jobb oldali része még szemléletesebbé teszi a csoportok közötti különbséget. A három csoporthoz tartozó egy-egy kis ábra a (2.21) modellből származó csoport és interakciós hatás együttes görbéit szemlélteti, azaz formánsenként és csoportonként az  $f_1(x_1) + f_{12}(x_1, x_2)$  összeget. Az egyes görbék köré rajzolt 95%-os Bayes konfidencia-intervallumok segítenek eldönteni, hogy az eltérés a görbe mely pontján szignifikáns. Az ábrán az y tengely egyes formánsokhoz tartozó nulla pontjait vízszintes, folytonos rácsvonal jelöli. Ez képviseli az összes mintaelemet legjobban leíró  $f_2(x_2)$  simító spline-t. A görbe azon pontjai térnek el szignifikánsan egymástól, amelyek 95%-os Bayes konfidencia-intervalluma nem tartalmazza a nulla pontot.

Az új grafikus ábrázolások kidolgozásakor [Fruehwald(2010)] ábráiból indultam ki, amelyeket a normalizált formánstérben a három formánsmenet egyidejű megjelenítésével egészítettem ki. A fenti módszerrel tetszőleges számú csoport formánsgörbéi összehasonlíthatók, eltérésük szignifikanciája mindhárom formáns esetében a görbe teljes terjedelmében az ábráról leolvasható.

## 6.2.2. Új vizsgálati módszerek a magánhangzó formánsmozgásának vizsgálatára

Az előzőekben bemutatott célzott vizsgálatok mellett fontos, hogy a formánsmozgások alakjáról átfogó képet kapjunk a lehetséges hangkörnyezetek figyelembe vételével. Ebben a fejezetben a CVC hanghármasokra szűkíttem a vizsgálatot, és a magánhangzóban bekövetkező formánsmozgást a funkcionális adatelemzés módszerével és új mutatók bevezetésével írom le.

Az átfogó vizsgálat egyes lépéseit az [a:] magánhangzó kapcsán ismertetem. Megadom az egyes mutatók és ábrák pontos kiszámítási és értelmezési módját. A összes magánhangzó-minőség – beleértve az itt részletezett [a:]-t is – formánsmeneteinek jellemzését a függelékben, tablószerűen mutatom be. Az összes magánhangzó-minőségre vonatkozó vizsgálat eredménye a 6.2.3. fejezetben található.

### 6.2.2.1. A formánstérképek a kutatás támogatására

A vizsgálat elsődleges célja, hogy a 11 mássalhangzó-környezet (3.5. táblázat) által meghatározott formánsmozgásokat  $F_1$ - $F_2$ - $F_3$ -ra egy-egy jellemző görbe együttes megadásával írjam le. Ennek megfelelően, például az [a:] magánhangzóra elméletileg 121 különböző  $C_{t,1}[a:]C_{t,2}$  hanghármas(típus) definiálható. ( $C_{t,1}$  a megelőző,  $C_{t,2}$  a követő mássalhangzótípus). Az összes magánhangzó-minőséget ( $V_m$ ) figyelembe véve pedig  $9 \times 121$  db  $C_{t,1}V_mC_{t,2}$  hanghármas(típus) írható fel. *Formánstérkép* alatt az egyes  $C_{t,1}V_mC_{t,2}$  hanghármasok jellemző formánsmeneteinek együttesét értjük [Abari & Olaszky(2012)]. Az [a:] -ra így elvileg 121 formánstérkép határozható meg, amelyeket egy mátrixban, 11 sorba és 11 oszlopba rendezve jeleníthetünk meg (6.10. ábra). A sornevek a megelőző, az oszlopnevek a követő mássalhangzó kategóriáját adják meg. Egy mátrixos elrendezés adott magánhangzóhoz kötődik (esetünkben [a:] -hoz), és az összes lehetséges 121 db  $C_{t,1};C_{t,2}$  kombinációhoz biztosít egy sor-oszlop találkozást – cellát –, amelyet a 6.10. ábrán a megfelelő formánstérképekkel töltöttem ki. Az üresen maradt cellák azt jelzik, hogy a FEF adatbázisban nincs minta az illető  $C_{t,1}[a:]C_{t,2}$  hanghármasra.

Az egyes formánstérképeken az  $F_1$ - $F_2$ - $F_3$ -ra jellemző formánsmozgásokat a funkcionális varianciaelemzés módszerével a (2.16) összefüggés alapján rajzoltam meg. A bázisfüggvények másodrendű B-spline-ok voltak. A formánstérképek értelmezése megegyezik a 6.8. ábra b) részénél elmondottakkal, de konfidencia-intervallumot nem jeleníték meg.

A funkcionális varianciaelemzés modelljét a (2.19) összefüggés alapján meghatározott  $R^2(t)$  determinációs együttható-függvény segítségével értékelem. Az  $R^2(t)$  értéke azt fejezi ki, hogy a szomszédos hangkörnyezettől való függés a formánsmozgások teljes variabilitásából hányad részt magyaráz meg. A 6.10. ábra b) részéből kiolvasható, hogy a második formáns alakját a hangkörnyezet rendkívül nagymértékben megmagyarázza (minél magasabb a függvényérték, annál erősebb a magyarázat). A magánhangzó két szélén – a 10 és 90%-os mérési pontokban – kicsit csökken a magyarázóerő mértéke, de még mindig elég

magasnak mondható.  $F_1$ -ben és  $F_3$ -ban a hangkörnyezetek nem magyarázzák ilyen karakteresen az előforduló formánsmozgásokat. Ennek az az oka, hogy a (2.19) összefüggésben az  $SSE(t)$  hiba eltérés-négyzetösszeg függvény és az  $SST(t)$  teljes eltérés-négyzetösszeg függvény értékei hasonló nagyságrendűek, ahogyan ez a 6.1. táblázatból is kiolvasható. Leegyszerűsítve ez azt jelenti, hogy az [a:] összes formánsmozgása egyforma távol van mind a teljes átlaggörbéhez, mind a saját  $C_{1,1}; C_{1,2}$  csoportjukra előrejelzett görbéhez viszonyítva. Így ha ez a távolság relatíve alacsony, mint  $F_1$  esetében, akkor az elenyésző magyarázóerő a formánsmozgások [a:]-beli hasonlóságát fejezi ki. Ha pedig olyan magas, mint amit  $F_3$  mutat, akkor az valóban a modell elégtelenségét jelenti.

6.1. táblázat. Az [a:]-ra végrehajtott funkcionális varianciaelemzést kísérő  $SSE(t)$  és  $SST(t)$  függvények 5 mérési pontban átlagolt értékei

[a:]	$SSE(t)$ átlaga	$SST(t)$ átlaga
$F_1$	25 008	28 911
$F_2$	5 571	78 908
$F_3$	3 488 606	3 505 539

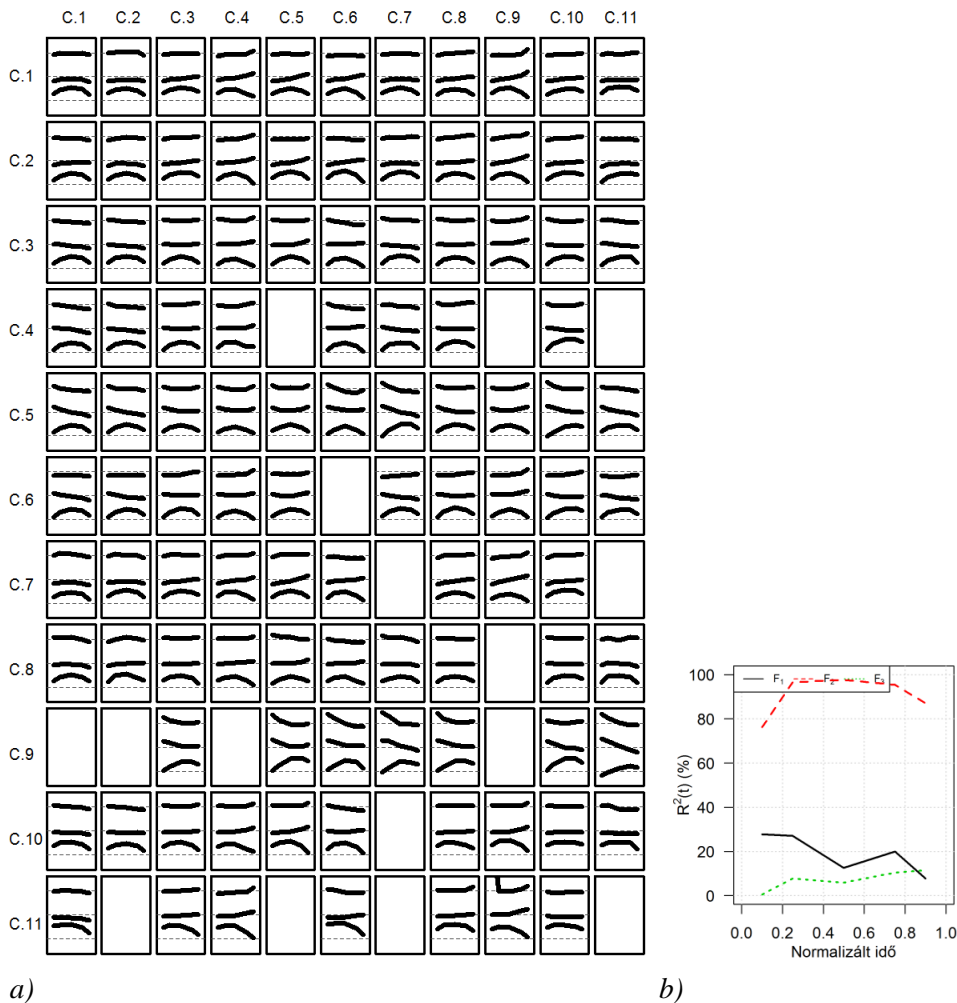
A formánstérképek vizsgálata során elsődleges, hogy az egyes formánsgörbék az  $y$  koordináta mentén hol helyezkednek el, ezt nevezzük  $y$ -eltolásnak. Ezt az átlaggörbe középső pontjával jellemezhetjük. Fontos vizsgálni a formánsmenet alakját is, vagyis a szomszédos hangok hatását. Ezen tényező számszerűsítésének érdekében *deformitási mutatókat* vezetek be. Egy CVC hangkapcsolatot két deformitási mutató jellemez, a CVC hangkapcsolat első felét leíró  $dm_{CV}$  és a második felét jellemző  $dm_{VC}$ . Kiszámításukat a (6.1) összefüggések tartalmazzák. Mindkét mutató 2 tag összege, melyek a szomszédos mássalhangzó hatását számszerűsítik.

$$\begin{aligned} dm_{CV} &= (t_2 - t_1)(\bar{x}(t_2) - \bar{x}(t_1)) + (t_3 - t_2)(\bar{x}(t_3) - \bar{x}(t_2)) \\ dm_{VC} &= (t_4 - t_3)(\bar{x}(t_4) - \bar{x}(t_3)) + (t_5 - t_4)(\bar{x}(t_5) - \bar{x}(t_4)) \end{aligned} \quad (6.1)$$

6.2. táblázat. Az [a:] formánstérképein szereplő formánsmenetek  $y$ -eltolásának leíró statisztikai adatai

	Átlag	Szórás	Min.	Max.
$F_1$	1,93	0,25	0,53	2,49
$F_2$	-0,15	0,2	-0,59	0,31
$F_3$	-0,16	0,25	-0,9	0,39





6.10. ábra. a) Az [a:] formánsstérképei. b) Az  $R^2(t)$  determinációs együttható-függvény alakja három formánsra. A folytonos vonal  $F_1$ -re, a szaggatott  $F_2$ -re és a pontozott  $F_3$ -ra vonatkozik.

A 6.10. ábra formánsmeneteinek  $y$ -eltolása egységes képet mutat:  $F_1$ -ben erős emelkedés,  $F_2$ -ben és  $F_3$ -ban a szaggatott vonalakkal jelzett átlaghoz képest kisebb csökkenés olvasható ki. Ezt számszerűsíti a 6.2. táblázat. A kis szóródási mutatók jelzik az  $y$ -eltolás állandóságát a  $C_{t,1}[a:]C_{t,2}$  hanghármasokban.

A deformitási mutató szerint sorrendbe állíthatjuk a megelőző és a követő mássalhangzó-kategóriákat. A 6.3. táblázat közli a mássalhangzó környezetek sorrendjét, a deformitási mutató szerint csökkenő mértékben. A kategória mellett a hozzá tartozó  $dm_{CV}$ , illetve a  $dm_{VC}$  érték látható. Pozitív deformitási mutató növekvő meredekségű változást okoz, míg a negatívhoz csökkenő meredekség társul. Az első formáns CV felében markáns növekedés, VC felében pedig

átlagosan ugyanolyan mértékű csökkenés olvasható ki. A környezetek deformáló hatása sokkal kisebb az  $F_2$  és  $F_3$  formánsoknál. A nulla körüli átlagos deformitási mutatók jelzik, hogy [a:]-ban csak néhány mássalhangzó kategória képes a második és harmadik formáns jelentős mozgására.

6.3. táblázat. Az [a:] deformitási mutatói szerint csökkenő sorrendbe rendezett mássalhangzó kategóriák és deformitási mutatók

	$F_1$		$F_2$		$F_3$	
	CV	VC	CV	VC	CV	VC
1.	C.9 0,34	C.6 -0,29	C.9 -0,22	C.9 0,16	C.9 -0,28	C.9 0,15
2.	C.6 0,25	C.9 -0,28	C.5 -0,15	C.5 0,14	C.11 -0,23	C.4 0,11
3.	C.5 0,24	C.4 -0,27	C.6 -0,08	C.4 0,1	C.5 -0,12	C.1 -0,07
4.	C.4 0,23	C.1 -0,21	C.7 0,08	C.1 -0,09	C.3 -0,07	C.3 0,05
5.	C.3 0,22	C.2 -0,19	C.2 0,07	C.7 -0,08	C.4 -0,07	C.6 -0,05
6.	C.2 0,22	C.5 -0,19	C.1 0,07	C.11 -0,07	C.10 -0,06	C.2 -0,05
7.	C.8 0,18	C.3 -0,17	C.4 -0,06	C.6 0,07	C.2 0,04	C.10 0,05
8.	C.7 0,18	C.8 -0,15	C.3 -0,03	C.2 -0,05	C.7 0,03	C.8 0,05
9.	C.1 0,16	C.7 -0,15	C.11 0,02	C.3 0,04	C.1 0,02	C.7 -0,03
10.	C.10 0,11	C.11 -0,11	C.8 0,01	C.8 0,02	C.8 -0,02	C.5 0,03
11.	C.11 0,05	C.10 -0,07	C.10 0	C.10 0,01	C.6 0,01	C.11 -0,01
Átlag	0,20	-0,19	-0,02	0,02	-0,06	0,02

### 6.2.2.2. A formánsmozgások stabilitása

Minden formánstérkép esetén megvizsgálhatjuk, hogy a benne megvalósuló formánsmozgások milyen mértékben hasonlók egymáshoz. Erre a (2.11) és (2.12) összefüggésekkel leírt varianciafüggvény és szórásfüggvény adhatja meg az alapot. A varianciafüggvényhez nagyon hasonló ún. stabilitási mutatót ( $SM$ ) vezettünk be [Abari & Olasz (2012)]-ben. Egy adott  $C_{t,1}V_mC_{t,2}$  hanghármasra és formánsorszámra (például  $F_1$ -re) jellemző stabilitási mutató kiszámításához az átlaggörbétől mért átlagos euklideszi távolságot használtuk fel. Ez a mutató egyetlen számértékbe sűríti az adott hanghármas  $F_1$ -re vonatkozó formánsmozgásainak hasonlóságát. A kisebb érték stabilabb formánsmozgás halmazt jelent a cellán belül.

Bevezetjük a *stabilitási vektort* ( $SV$ ), amely figyelembe veszi, hogy a funkcionális adatelemzés mutatói többnyire függvények, így az  $SV$  a mérési pontokban értelmezett – esetünkben 5 elemű – vektor. Kiszámítása (6.2 egyenlet) a normalizált formánstérhez kötődik: a vizsgált 5 mérési pontban az átlaggörbétől fél szóráson belüli (összesen egy szóráson belüli) pontok aránya.

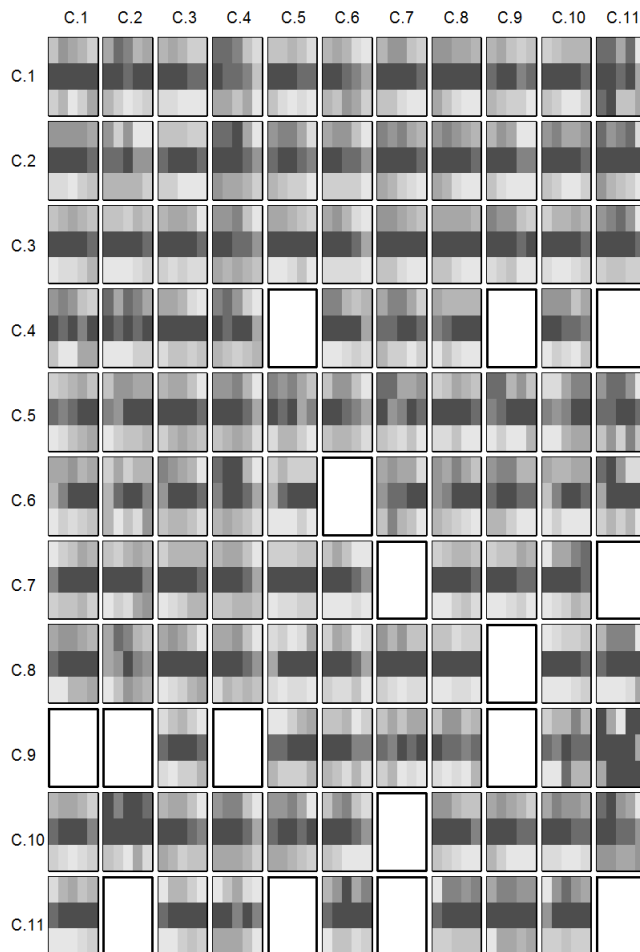
$$SV(t_j) = \frac{\left| \left\{ y_{ij} \mid \bar{x}(t_j) - \frac{1}{2} \leq y_{ij} \leq \bar{x}(t_j) + \frac{1}{2} \right\} \right|}{N}, \quad (6.2)$$

$$i = 1, 2, \dots, N, j = 1, 2, 3, 4, 5$$

Ezzel cellánként és formánsenként 5 adattal jellemezhetjük a formánsmozgások hasonlóságát. Minél nagyobb a hasonlóság, annál közelebb van az  $SV(t_j)$  mutató értéke a 100%-hoz. Ennek előnye, hogy könnyen összehasonlíthatóak a különböző hanghármasokhoz tartozó fenti arányok. Továbbá könnyen implementálható hozzá grafikus megjelenítés, amely alapján dönthetünk, hogy melyek a nagyon vagy kevésbé hasonló formánsmozgásokkal rendelkező cellák.

A 6.11. ábrán az [a:] formánsmozgásainak stabilitását grafikusán ábrázoltam. A cellák elrendezése megegyezik a 6.10. ábra formánstérképeinek elhelyezkedésével, de itt cellánként  $3 \times 5$  db szürke téglalap jelzi, hogy az adott formáns (3 db), adott mérési pontjában (5 db) mennyire hasonlóak a formánsértékek. Egy cellán belül az alsó öt téglalap az  $F_1$ -hez tartozó  $SV(t_1)$ ,  $SV(t_2)$ , ...,  $SV(t_5)$  értékek grafikus megfelelőjét tartalmazza. A második, felette lévő téglalapsor, az  $F_2$ -höz, a legfelső sor pedig az  $F_3$ -hoz tartozó  $SV(t_j)$  értékeknek megfelelő színeket. A nagyobb hasonlóság sötétebb színű téglalappal van jelölve.

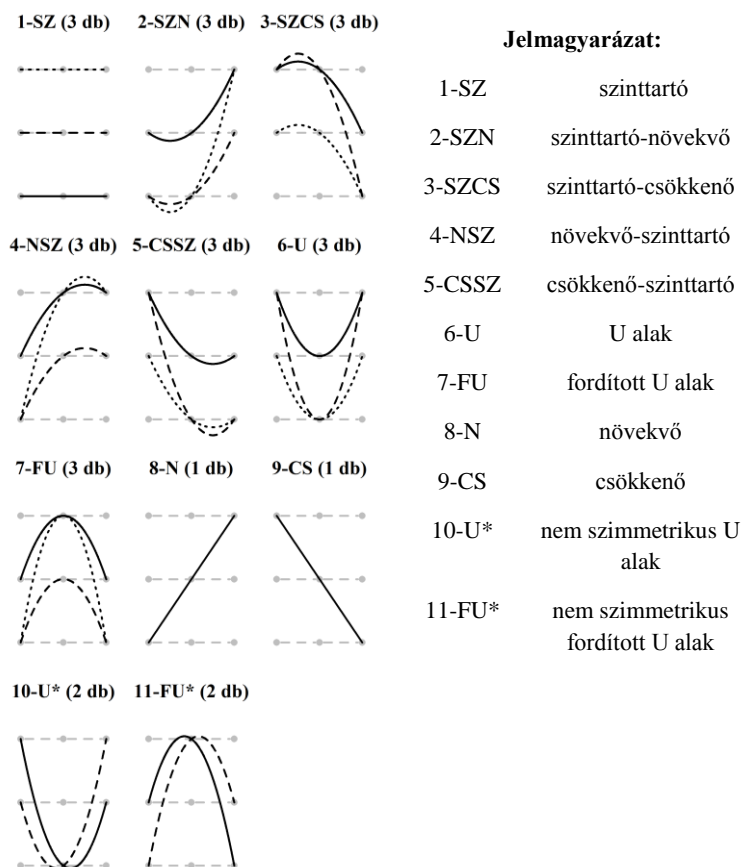
A 6.11. ábra megerősíti azt a hipotézist, hogy a legstabilabb formánsmozgások  $F_2$ -ben jelennek meg, míg a másik két ábrázolt formánsban sokkal gyengébb a stabilitás.



6.11. ábra. A formánsmozgások stabilitása [a:]-ban. A sötétebb helyek stabilabb formánsmozgást jeleznek

### 6.2.2.3. Az egymástól különböző formánsmenetek száma

Ha arra a kérdésre szeretnénk válaszolni, hogy hány különböző formánsmenettel jellemezhetők az [a:]-k formánsmozgásai, akkor először a formánsmozgások tipizálását kell elvégezni. A formánsmenet-típusok feltérképezésére kisebb adatbázison, 9 magánhangzó-minőség és 6 képzési hely szerinti mássalhangzó kategória alapján már történt kísérlet [Abari & Olaszky(2012)]. Jelen vizsgálatban lényegesen nagyobb adatbázison, finomabb mássalhangzó-környezetek figyelembe vételével van lehetőségem elemzést végezni.



6.12. ábra. A formánsmenet-típusok 11 elméleti kategóriája. Zárójelben a típus által lefedett formánsmenetek száma látható

Az 6.12. ábrán a matematikailag lehetséges ideális formánsmenetek rendszerét mutatom be [Abari & Olaszy(2012)] alapján. 27 db lehetséges ideális formánsmozgást definiáltunk, amelyeket 11 formánsmenet-típusba soroltunk. Minden formánsmenet-típus 9 pontot tartalmaz, 3 sorban és 3 oszlopban elrendezve. A 3 oszlop jelentése: a  $C_{t,1}V_mC_{t,2}$  hármass egyes elemeinek megfelelő artikulációs konfiguráció akusztikus vetülete. A három sor pedig a frekvenciaértékek 3 lehetséges szintje, melynek alapja, hogy a mássalhangzó-környezetbe helyezett magánhangzók frekvenciaértékeinek 2, ill. 3 fokozata figyelhető meg [Magdics(1965):24]. Az 6.12. ábra első formánsmenet-típusának felirata: 1-SZ (3db), a neve szinttartó, és összesen 3 db lehetséges formánsmenetet tartalmaz (a pontozott, szaggatott és folytonos vonaltípusok különböztetik meg őket). Az egy típusba tartozó formánsmenetek a grafikon 9

pontjából úgy kötnek össze 3 pontot, hogy a 3 pont által meghatározott két átmeneti szakaszban, formánsmenetenként azonos legyen a mozgás iránya. A mozgás iránya lehet szinttartó, növekvő vagy csökkenő. Az 1. formánsmenet-típus esetén mindkét átmenet szinttartó mindhárom formánsmenetben. A 2–5. grafikonokon további 3-3 ideális formánsmenetet látunk, az átmenetek itt: szinttartó-növekvő, szinttartó-csökkenő, növekvő-szinttartó és csökkenő-szinttartó. A 6. és 7. grafikon 3-3 szimmetrikus menete a csökkenő-növekvő és növekvő-csökkenő eseteknek felel meg. A 8. és 9. grafikon egy-egy ideális formánsmenete növekvő-növekvő illetve csökkenő-csökkenő mozgást mutat. A 10. és 11. csillaggal jelölt formánsmenetek a nem szimmetrikus U alakokat írják le. Az 6.12. ábra egy formánsmenete tehát kiindul valamely hangátmeneti frekvenciából, átlagostól (középen), átlag felettiből vagy átlag alattiból (3 eset), elér egy tiszta fázisbeli értékhez (itt is 3 eset van: átlagos (középen), átlag feletti és átlag alatti) és zárul egy hangátmeneti frekvencia értékkel (szintén 3 eset). Az ideális formánsmenetek száma így  $3 \times 3 \times 3 = 27$  darab.

6.4. táblázat. A 11 formánsmenet-típus egymáshoz mért távolsága. Az utolsó sorban a teljesen kitöltött mátrix oszlopainak összege látható. A sor és oszlopnevek magyarázata a 6.12. ábrán szerepel

	1	2	3	4	5	6	7	8	9	10	11
<b>1-SZ</b>	0	1	1	1	1	1	1	2	2	2	2
<b>2-SZN</b>		0	2	1	2	1	2	1	3	1	3
<b>3-SZCS</b>			0	2	1	2	1	3	1	3	1
<b>4-NSZ</b>				0	2	2	1	1	3	3	1
<b>5-CSSZ</b>					0	1	2	3	1	1	3
<b>6-U</b>						0	3	2	2	1	3
<b>7-FU</b>							0	2	2	3	1
<b>8-N</b>								0	4	2	2
<b>9-CS</b>									0	2	2
<b>10-U*</b>										0	4
<b>11-FU*</b>											0
<b>Σ</b>	<b>14</b>	<b>17</b>	<b>17</b>	<b>17</b>	<b>17</b>	<b>18</b>	<b>18</b>	<b>22</b>	<b>22</b>	<b>22</b>	<b>22</b>

A 6.12. ábra 11 darab formánsmenet-típusa a görbét alkotó 3 pont mozgásával átvihető egymásba. Távolságot definiálhatunk két formánsmenet-típus között: minimálisan hány pontnyi mozgással juthatunk el az egyik típusból a másikba (6.4. táblázat). A távolság szimmetrikus tulajdonsága miatt a 6.4. táblázatban elegendő a felső háromszöget kitölteni. Az utolsó,  $\Sigma$  sorból azt olvashatjuk ki, hogy az adott típushoz képest az összes többi típus együttesen milyen távol van. Minél kisebb ez a szám, annál hasonlóbb a többihez az adott típus (a szinttartó van legközelebb a többihez). A  $\Sigma$  érték kiszámításához a felső

háromszögből tükrözzük az adatokat az alsó háromszögbe, majd összegezzük az oszlopokat.

Amennyiben rendszerezni szeretnénk az adatbázisbeli formánsmozgásokat a fenti modell szerint, akkor minden konkrét formánsmozgást a 6.12. ábrán megjelenő 11 formánsmenet-típus valamelyikébe kell besorolnunk. Nem gondoljuk, hogy a fenti séma minden formánsmozgást tökéletesen leír, hiszen alapvetően egyenes (1, 8, 9), illetve parabola jellegű mozgással (2–7, 10, 11) igyekszünk leírni a mozgásformákat. Ismert, hogy a formánsmozgások mindegyike pontosan nem írható le parabolikus görbével [Harrington(2010b):143], és már a mi 5 ponton alapuló mozgásaink között is elképzelhető a 6.12. ábrán látható formánsmenetektől lényegesen eltérő forma. Szükség van tehát egy 12. egyéb kategóriára, amelybe azok a formánsmozgások kerülnek, amelyek egy adott hibaküszöböt meghaladó mértékben illeszkednek csak az 1–11 típusokra. A módszerünk annál hatékonyabb, minél kisebb elemszámú a 12. kategória, vagyis minél kevesebb elemet kell figyelmen kívül hagynunk a további elemzésből. A módszer másik sarkalatos pontja a kategóriák közötti döntés biztossága. Látjuk, hogy sok 1 távolsággal rendelkező típuspár létezik, az ezek közötti döntés nagy körültekintést igényel.

Az [Abari & Olasz(2012)]-ben részletezett, paraméterek minimalizálásán alapuló eljárásunkkal, minden egyes formánsmozgás esetén hozzárendelhetünk egy formánsmenet-típust, a lehetséges  $11+1=12$  kategóriából. A cellához tartozó jellemző formánsmenet-típus (1–12) meghatározásához a cellában lévő  $N$  darab görbe típusainak megfigyelt gyakoriságait ( $O_k, k = 1, \dots, 12$ ) használtuk fel: a cellához azt a  $k$  típust rendeljük hozzá, amelyből a legtöbb volt a cellában, azaz, amelyre az  $O_k/N$  hányados maximális értéket ad. Ezt a hányadost *szigorú stabilitási mutató*nak nevezzük.

$$SM_2 = \max_k O_k/N, k = 1, \dots, 12 \quad (6.3)$$

Értéke 0 és 1 között változhat, minél közelebb van az egyhez, annál stabilabban írja le a formánsmenet-típust az adott cella adott formánsát.

A 6.5. táblázatban az [a:] magánhangzó egyes celláihoz rendelt formánsmenet-típusok eloszlását mutatom be a három vizsgált formánsra. A táblázatban szereplő számok az adott formánsmenettel leírható formánsmozgások százalékát jelentik a teljes adatbázisban. Az oszlopnevek magyarázata a 6.12. ábrán szerepel.

6.5. táblázat. Az 1–11 formánsmenet-típusok eloszlása és a 12. kategória százalékos gyakorisági táblázata  $F_1, F_2$  és  $F_3$  esetén. Az adatok egész %-ra kerekítettek. A cellákban szereplő számok az adott formánsmenettel leírható formánsmozgások százalékát jelentik a teljes adatbázisban

	1	2	3	4	5	6	7	8	9	10	11	12	
$F_1$	0	0	2	0	0	0	77	2	4	0	15	0	100%
$F_2$	30	5	1	0	1	6	5	29	24	0	0	0	100%
$F_3$	15	8	0	1	3	23	3	22	26	0	0	0	100%

A legalább 5%-ot elérő formánsmenet-típusok  $F_1$  esetén a 7-FU (77%) és a 11-FU (15%). Az [a:]  $F_1$ -beli mozgásai tehát jól leírhatók a fordított U alak két változatával, ami az egyes  $C_{t,1}[a:]C_{t,2}$  hanghármasokhoz illesztett mozgások 92%-át jelenti. Az  $F_2$  és  $F_3$  jóval több formánsmenet-típussal írható le.  $F_2$  esetében már 6 formánsmenet-típus éri el az 5%-ot: 1-SZ, 2-SZN, 6-U, 7-FU, 8-N, 9-CS. A harmadik formáns 5 ilyen kategóriával rendelkezik: 1-SZ, 2-SZN, 6-U, 8-N, 9-CS. A mássalhangzó-környezet tehát elsősorban a magánhangzó második és harmadik formánsának mozgására hat.

A cellákat jellemző formánsmenet-típusok meghatározásakor felhasználtuk a szigorú stabilitási mutatót. Az  $SM_2$  értéke jelzi, hogy a kapott formánsmenet milyen mértékben jellemzi a cellában megvalósuló tényleges mozgásokat. Formánsenként átlagolva a 6.6. táblázat első oszlopa sorolja fel a szigorú stabilitási mutató átlagát. Ezek az ideálisnak tekinthető 1-hez képest messze vannak. Például  $F_1$  estében az 1 átlag az jelentené, hogy a  $C_{t,1}[a:]C_{t,2}$  hanghármasok minden egyes adatbázisbeli formánsmozgásának a kategóriája egy cellán belül megegyezik egymással: például a cellák egyik részében csak 7-FU kategóriájú  $F_1$  mozgások, a másik részében csak 11-FU típusú mozgások találhatóak. A valóságos 0,58 azt jelenti, hogy a cellákban megvalósuló mozgások kategóriája és a cella formánsmenet-típusának egyezése átlagosan csak 58 %-os.

6.6. táblázat. Az [a:] formánsmozgásainak formánsenként vett szigorú stabilitás átlaga az első oszlopban, az összevont szigorú stabilitási mutató értéke a második oszlopban található.

[a:]	$SM_2$ átlag	Összevont $SM_2$ átlag
$F_1$	0,58	0,79
$F_2$	0,55	0,74
$F_3$	0,31	0,49

A formánsorszám növekedésével ez az arány láthatóan tovább csökken. Ha azonban a cellák második leggyakoribb formánsmenetét is megvizsgáljuk, és amennyiben azt találjuk, hogy az 1 távolságra van a cella meghatározó



formánsmenet-típusához (távolságokhoz lásd a 6.4. táblázatot), akkor nem követünk el túl nagy hibát, ha összevonjuk ezt a két típust. Az 1 távolság ugyanis biztosítja, hogy a mozgás alakján lényegesen nem változtatunk. Összevonást tudunk végrehajtani  $F_1$ -ben az esetek 89%-ában,  $F_2$ -ben 65% és  $F_3$ -ban 56%-ban. Az egyesítés megnövelte az egyezési arányt, ahogyan azt a 6.6. táblázat második oszlopa is tartalmazza. Ezek alapján azt mondhatjuk, hogy a 6.12. ábrán leírt formánsmenet modellel a  $C_{t,1}[a:]C_{t,2}$  hanghármasok  $F_1$  és  $F_2$  formánsmozgásai jól leírhatók, de a formánsmozgások folytonosan töltenek ki egy határozottan körülírható részt a normalizált formánstérben.

#### **6.2.2.4. A $C_{t,1}V_mC_{t,2}$ kapcsolat C eleminek jóslása a formánstérkép ismeretében**

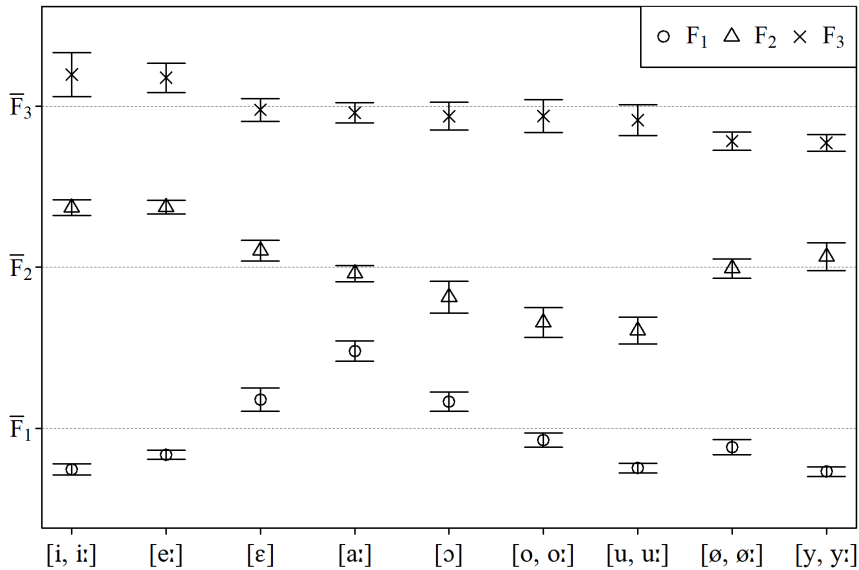
Végül feltehetjük azt a kérdést, hogy az [a:] (vagy bármely más magánhangzó) valamely formánstérképének ismeretében, milyen mértékben tudjuk megjósolni a  $C_{t,1}[a:]C_{t,2}$  hanghármasot. A formánstérképeket most az  $F_1$ - $F_2$ - $F_3$  jellemző formánsmenet-típusával írjuk le. Összesen 44 db kombinációban valósulnak meg a formánsmenet-típus hármasok. A 6 leggyakoribb kombináció: 7-FU.8-N.8-N (11%), 7-FU.9-CS.9-CS (10%), 7-FU.1-SZ.6-U (9%), 7-FU.1-SZ.1-SZ (7%), 7-FU.9-CS.6-U (7%), 7-FU.1-SZ.9-CS (7%). Ez a 6 kombináció a lehetséges  $C_{t,1}[a:]C_{t,2}$  cellák közel 50%-át jellemzi (összesen 52 db cellát). Ezekben az esetekben a formánsmenetek ismerete kevésbé járul hozzá a hangkörnyezetek jóslásához. Azonban 27 db kombináció (28%) mindössze egyszer fordul elő, például a 9-CS.1-SZ.1-SZ, 7-FU.9-CS.1-SZ, 3-SZCS.8-N.2-SZN és a 7-FU.9-CS.4-NSZ. Az összes cella 40%-a leírható maximum 2 előfordulással rendelkező menet-kombinációval, így ezekben az esetekben pontosabb jóslás adható.

#### **6.2.3. A kontextusfüggő vizsgálatok kiterjesztése**

Az előző 4 alfejezetben a  $C_{t,1}[a:]C_{t,2}$  hanghármas vizsgálata kapcsán mutattam be az új vizsgálati eszközöket: a formánstérképet, a deformitási mutatót, a stabilitási vektort, az (összevont) szigorú stabilitási mutatót és a mássalhangzó-környezet jóslására vonatkozó egyszerű számításokat. Ebben a fejezetben kiterjesztem a vizsgálatot az összes magánhangzó-minőségre.

### 6.2.3.1. A formánstérképek értékelése

A formánstérképek vizsgálata módot ad az  $y$ -eltolás és a deformitási mutatók sokoldalú vizsgálatára. A magánhangzóra jellemző  $y$ -eltolás magánhangzónkénti átlaga és szórása jelenik meg a 6.13. ábrán.



6.13. ábra. A 9 magánhangzó-minőség átlaga és szórása az első három formánsban

Az ábrán három vízszintes vonallal az összes beszélőre vonatkoztatott normalizált formánsfrekvencia nulla pontját (beszélőnkénti átlagát) jelöltem meg a három formánsra. Ez alapján képet kaphatunk arról, hogy a magánhangzót leginkább jellemző  $y$ -eltolás mely esetekben átlag alatti vagy átlag feletti. Az ábra visszaigazolja a szakirodalomban található mérési eredményeket (vö. [Magdics(1965):10] és azt az ismeretet, mely szerint magasabb nyelválláshoz kisebb  $F_1$  érték és palatális területen képzett vokálshoz nagyobb  $F_2$  érték tartozik. A szórás mindhárom vizsgált formáns esetében kicsi, de a formánssorszám emelkedésével nő. Az ábra igazolja, hogy a normalizált formánsfrekvenciák alapján levont következtetéseink visszavezethetők a Hz-ben mért értékekre. A nyelv vízszintes és függőleges mozgása szerint megkülönböztett magánhangzó kategóriákat [Gósy(2004):67] tartalmazza.

A formánstérképek mellett a függelékben megadtam az  $R^2(t)$  determinációs együttható-függvény grafikus képét is. Ezek áttekintése után azt mondhatjuk, hogy  $F_2$  esetében a mássalhangzó környezetek jól meghatározzák a magánhangzóban megvalósuló második formáns alakját. A (2.20) összefüggés

az 5 mérési pontban kiszámolt determinációs együttható-függvényt egyetlen számba sűríti. Ezek átlaga az összes magánhangzóban  $F_2$ -re 78,5%, ami azt jelzi, hogy a formánsmozgásra vonatkozó bizonytalanságból jelentős részt megmagyaráz a megelőző és követő mássalhangzó környezet. Ugyanez a másik két vizsgált formáns esetén nem mondható el. A rendkívül alacsony értékű, hasonló módon kiszámolt illeszkedési értékek ( $F_1$ : 14,4%,  $F_2$ :13,4%) azt jelzik, hogy a modell magyarázó ereje nagyon alacsony, azaz a teljes eltérés-négyzetösszeg és a hiba eltérés-négyzetösszeg túl közeli értékeket vesz fel. Ennek oka alapvetően megegyezik a 6.2.2.1. fejezetben az [a:] -ra adott magyarázattal, de most minden magánhangzó-minőségre kiterjesztem a vizsgálatot.

6.7. táblázat. Az magánhangzó-minőségek átlagolt  $SSE(t)$  és  $SST(t)$  függvényei (az átlagolás a mérési pontokra vonatkozik)

[i, i:]			[e:]			[ɛ]		
	SSE	SST		SSE	SST		SSE	SST
$F_1$	0,63	0,66	$F_1$	0,24	0,27	$F_1$	0,64	0,75
$F_2$	0,23	5,05	$F_2$	0,16	3,47	$F_2$	0,17	0,85
$F_3$	0,89	1,15	$F_3$	0,64	0,83	$F_3$	160,58	161,11

[a:]			[ɔ]			[o, o:]		
	SSE	SST		SSE	SST		SSE	SST
$F_1$	0,90	1,04	$F_1$	0,58	0,65	$F_1$	0,46	0,51
$F_2$	0,20	2,85	$F_2$	0,15	1,44	$F_2$	0,18	0,98
$F_3$	126,18	126,80	$F_3$	92,69	92,82	$F_3$	1,09	1,22

[u, u:]			[ø, ø:]			[y, y:]		
	SSE	SST		SSE	SST		SSE	SST
$F_1$	0,41	0,45	$F_1$	0,31	0,35	$F_1$	0,28	0,30
$F_2$	0,30	0,52	$F_2$	0,12	0,63	$F_2$	0,19	1,59
$F_3$	1,15	1,30	$F_3$	0,58	0,69	$F_3$	0,70	0,83

A 6.7. táblázat az egyes magánhangzóra vonatkozó  $SST(t)$  és  $SSE(t)$  függvények mérési pontokra átlagolt értékeinek és a magánhangzó-minőség elemszámának a hányadosát tartalmazza. A táblázat egyes értékeit hasonlóan kapjuk, mint az [a:] -ra vonatkozó 6.1. táblázat adatait, de itt a különböző magánhangzó-minőségek közötti összehasonlíthatóság kedvéért az előforduló elemszámmal korrigálunk. A táblázatból kiolvasható az  $F_1$  és az  $F_3$  alacsony  $R^2(t)$  függvényértékei közötti különbség minden magánhangzó-minőségre. A relatíve alacsony átlagos  $SST(t)$  és  $SSE(t)$   $F_1$ -ben azt jelzi, hogy az első formánsban magánhangzónként nagyon hasonló formánsmozgások realizálódnak. A relatíve magas átlagos  $SST(t)$  és  $SSE(t)$   $F_3$ -ban azt jelenti, hogy az esetek egy részében nem tudunk jó illeszkedéssel görbét prediktálni a  $C_{t,1}V_mC_{t,2}$  hanghármasokra, mert azokban nagyon különböző formánsmozgások figyelhetők meg. A harmadik formáns viszonylag elfogadható hibával prediktál

görbét az [ø, ø:], [e:] [y, y:], [i, i:], [o, o:] és [u, u:] minőségekre, míg a többi esetben kiugróan nagy hibával. A 6.7. táblázat minden magánhangzó-minőségben az  $F_2$ -re vonatkozó sor a  $F_1$  sorral összehasonlítva megmutatja, hogy magánhangzónként változatosabb formánsmozgások jelennek meg a második formánsban az elsőhöz viszonyítva (*SST* oszlop), de a  $C_{t,1}V_mC_{t,2}$  hanghármasok prediktált görbéi  $F_2$ -ben már jóval kisebb hibával (*SSE* oszlop) jellemzik az illető  $C_{t,1}V_mC_{t,2}$  cella formánsmozgásait,  $F_1$ -hez képest.

Az  $R^2(t)$  determinációs együttható-függvény elemzése alapján megállapíthatjuk, hogy a második formáns prediktált görbéi jól jellemzik az egyes hanghármasokat. Az első formánsban bekövetkező mozgásokat kevesebb görbével lehet leírni  $F_2$ -höz képest az egyes magánhangzóknál, de a leírás pontossága némileg csökken. Az  $F_3$ -ban a prediktált görbék csak bizonyos magánhangzók esetén járulnak hozzá a formánsmozgás jobb jellemzéséhez.

### 6.2.3.2. A deformitási mutató értékelése

A deformitási mutató a mássalhangzó környezet formánsmozgató hatását méri előjelhelyes módon. Nagyobb abszolút értékű mutató nagyobb formánsmozgató hatást jelent. A mutató alapján a magánhangzók sorrendbe állíthatók. A 6.8. táblázatban a deformitási mutató nagyságát vettük csak figyelembe, az irányát nem. A magánhangzókat az összes előjel nélküli  $dm_{CV}$  és  $dm_{VC}$  átlagával jellemezve, a táblázatbeli csökkenő sorrendet kapjuk.

6.8. táblázat. A magánhangzók átlagos abszolút deformitási mutató szerinti sorrendje az első három formánsban

$F_1$	[a:] 0,19	[ε] 0,12	[ɔ] 0,12	[o, o:] 0,05	[ø, ø:] 0,05	[e:] 0,03	[u, u:] 0,03	[i, i:] 0,03	[y, y:] 0,03
$F_2$	[u, u:] 0,11	[o, o:] 0,09	[ɔ] 0,08	[a:] 0,08	[e:] 0,07	[y, y:] 0,07	[i, i:] 0,07	[ε] 0,06	[ø, ø:] 0,06
$F_3$	[ɔ] 0,11	[i, i:] 0,11	[e:] 0,09	[ε] 0,08	[y, y:] 0,08	[a:] 0,07	[u, u:] 0,07	[ø, ø:] 0,06	[o, o:] 0,06

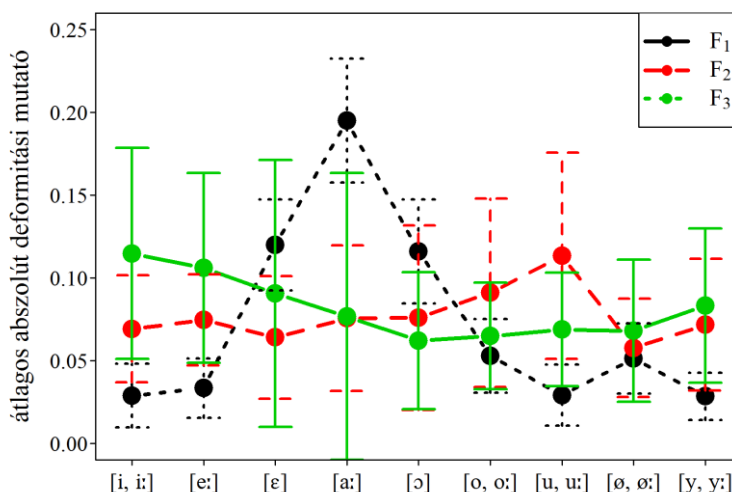
A 6.8. táblázat adatai azt mutatják, hogy az  $F_1$ -re és  $F_2$ -re vonatkozó sorokban a magánhangzó-minőségek deformitási mutató szerinti sorrendje éppen a nyelv vízszintes, illetve függőleges mozgása szerint megkülönböztetett kategóriákat követik. Legnagyobb deformitás  $F_1$ -ben a legalsó nyelválláshoz, majd csökkenő sorrendben az alsó, középső és felső nyelválláshoz társul. A második formáns mozgásai a hátulképzettekben nagyobb, az előlképzettekben kisebb változást szenvednek el. Az átlagos abszolút deformitási mutatókat a 6.9. táblázat tartalmazza az egyes magánhangzó-kategóriákra. Egyszempontos

varianciaelemzéssel [Welch(1951)] az egyes csoportok között szignifikáns különbséget mértem (Az  $F_1$ -re vonatkozó deformitás a nyelv függőleges mozgása szerint:  $F(3, 344,8)=1036,9$ ;  $p < 0,00$ ; az  $F_2$ -re vonatkozó deformitás a nyelv vízintes mozgása szerint:  $F(1, 648,5)=38,7$ ;  $p < 0,00$ ).

6.9. táblázat. A magánhangzók átlagos abszolút deformitási mutató szerinti sorrendje az első három formánsban

Függőleges nyelvállás	Átlagos abszolút deformitási mutató $F_1$ -re	Vízszintes nyelvállás	Átlagos abszolút deformitási mutató $F_2$ -re
felső	0,028	elöl képzett	0,067
középső	0,045	hátsó képzett	0,087
alsó	0,117		
legalsó	0,194		

Ha azt a kérdést vizsgáljuk, hogy melyik formánsban nagyobb a környezetek hatása a formánsmozgásokra, akkor egy kétszemponos varianciaelemzés modelljébe [Montgomery(2006)] a formáns és a magánhangzó-minőség tényezőket emelhetjük be. Mindkét tényező hatása szignifikánsnak bizonyult, és az interakciónak is van hatása. Ez azt jelenti, hogy különböző magánhangzók esetén más-más a három formánsban a deformitási mutató nagysága. (Magánhangzó-minőség hatása:  $F(8, 2628)=46,17$ ;  $p < 0,00$ ; formáns hatása:  $F(2, 2628)=3,50$ ;  $p = 0,030$ ).



6.14. ábra. A 9 magánhangzó-minőség deformitási átlaga és szórása az első három formánsban

A 6.14. ábra magánhangzónként és formánsenként mutatja be a deformitási mutató előjel nélküli átlagait (hasonlóan a 6.8. táblázathoz). A függőleges vonalak a szórást mutatják meg, amely a formánsorszám emelkedésével nő.

A mássalhangzó kategóriákra a deformitási mutató alapján hasonló sorrend állapítható meg. Itt arra keresem a választ, hogy melyek a legnagyobb változást eredményező mássalhangzó kategóriák. Külön vettem figyelembe az előjel nélküli  $dm_{CV}$  és  $dm_{VC}$  átlagokat, így a 6.10. táblázatban a sorok száma megnövekedett. A C.9 kategória első helyen szerepel szinte minden esetben.

6.10. táblázat. A mássalhangzók átlagos abszolút deformitási mutató szerinti sorrendje az első három formánsban

F <sub>1</sub>	CV	C.9 0,13	C.6 0,10	C.5 0,09	C.4 0,08	C.2 0,08	C.3 0,07	C.8 0,07	C.7 0,07	C.1 0,06	C.11 0,05	C.10 0,04
	VC	C.9 0,14	C.6 0,13	C.1 0,10	C.4 0,09	C.5 0,08	C.7 0,07	C.3 0,07	C.2 0,07	C.8 0,06	C.11 0,05	C.10 0,04
F <sub>2</sub>	CV	C.9 0,17	C.5 0,12	C.7 0,08	C.4 0,06	C.2 0,06	C.1 0,05	C.8 0,05	C.6 0,05	C.3 0,05	C.10 0,05	C.11 0,04
	VC	C.9 0,14	C.5 0,13	C.4 0,10	C.1 0,09	C.7 0,09	C.3 0,07	C.2 0,07	C.10 0,07	C.8 0,06	C.6 0,06	C.11 0,05
F <sub>3</sub>	CV	C.9 0,18	C.7 0,11	C.5 0,10	C.11 0,08	C.1 0,07	C.2 0,07	C.8 0,07	C.4 0,06	C.10 0,06	C.3 0,06	C.6 0,04
	VC	C.4 0,13	C.9 0,13	C.7 0,10	C.1 0,10	C.5 0,09	C.2 0,08	C.3 0,07	C.6 0,07	C.11 0,06	C.10 0,06	C.8 0,05

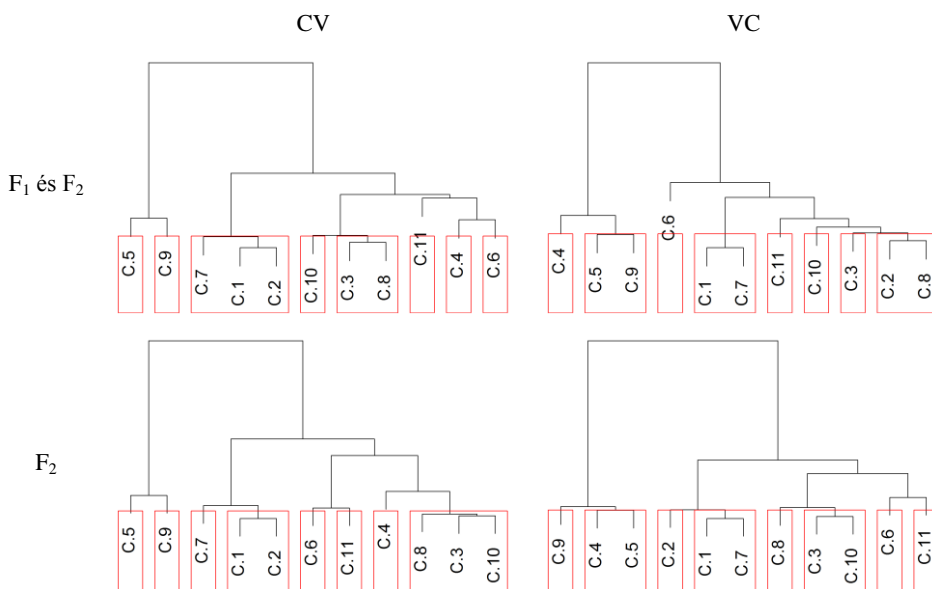
A legnagyobb összdeformitással rendelkező hanghármaszt is meghatároztam. A 6.11. táblázatból leolvasható, hogy mindhárom formáns esetében szimmetrikus környezetekről van szó.

6.11. táblázat. Az első hat legnagyobb abszolút deformitási mutatóval rendelkező hanghármas az első három formánsban

F <sub>1</sub>	C.6[a:]C.6 0,25	C.9[a:]C.9 0,22	C.7[a:]C.7 0,21	C.4[a:]C.4 0,20	C.5[a:]C.5 0,20	C.1[a:]C.1 0,20
F <sub>2</sub>	C.9[u, u:]C.9 0,20	C.5[u, u:]C.5 0,19	C.9[o, o:]C.9 0,16	C.5[o, o:]C.5 0,15	C.4[u, u:]C.4 0,15	C.5[ɔ]C.5 0,13
F <sub>3</sub>	C.9[a:]C.9 0,19	C.4[ɛ]C.4 0,19	C.1[i, i:]C.1 0,17	C.9[e:]C.9 0,16	C.7[e:]C.7 0,16	C.2[e:]C.2 0,15

Hierarchikus klaszterelemzés segítségével megvizsgáltam, hogy vannak-e hasonló formánsmozgást okozó mássalhangzó környezetek, összevonhatók-e bizonyos mássalhangzó-csoportok a deformitási mutató alapján. Külön vettem figyelembe a  $dm_{CV}$  és a  $dm_{VC}$  mutatókat, vagyis a megelőző és a követő mássalhangzó okozta hatásokat. Ha például a magánhangzó előtti C.1-hez keresünk hasonló formánsmozgást okozó másik, vele összevonható mássalhangzó-típust, akkor a C.1-et jellemző változókat kell számba venni. Ezek a lehetséges C.1V<sub>m</sub>C<sub>t,2</sub> kombináció  $dm_{C,1V}$  deformitási mutatói, amely összesen 9×11=99 db változót jelentenek formánsenként. Az összes további

mássalhangzó-kategóriát is ezekkel a változókkal írrom le. Vizsgálatomban euklideszi-távolságot számoltam a 11 mássalhangzó-kategória között, és a csoportok összevonására a Ward-féle eljárást használtam [Gordon(1999)]. A 6.15. ábra négy dendrogramja 8 mássalhangzó-klaszter létrehozását mutatja be. A megelőző és a követő mássalhangzóhoz tartozó dendrogramok az első (CV), illetve a második (VC) oszlopban láthatók. Ha az  $F_1$  és  $F_2$  formánsmozgásait is figyelembe vesszük, akkor  $2 \times 99 = 198$  változó írja le a mássalhangzó-kategóriát. Ezek az ábrák az első sorban találhatók. A második sor ábrái az  $F_2$ -höz tartozó deformitási mutatókon alapulnak.



6.15. ábra. A hierarchikus klaszterelemzés dendrogramjai

A 6.15. ábra bal felső dendrogramja a megelőző mássalhangzó-környezeteket veszi figyelembe és az  $F_1$ -ben és  $F_2$ -ben számolt deformitási mutatók alapján végzi a csoportosítást. Az ábráról leolvasható összevonási sorrend a következő: (1) C.1-C.2 (az orális bilabiálisok és a labiodentálisok kategóriája); (2) C.3-C.8 (az alveoláris és az [n] összevonása); (3) az elsőt összevont csoporthoz (bilabiális és labiodentális) sorolja az [m] hangot is.

Ha csak az  $F_2$  deformitási mutatóit használjuk (bal alsó ábra), akkor nagyon hasonló összevonást látunk: a C.1-C.2 újra egy csoportba kerül, és visszakapjuk a teljes alveoláris csoportot az [n] és [l] hangokkal. A követő mássalhangzó deformitási tényezőit hasonló módon vizsgálhatjuk (jobb oldali ábrák). Itt a C.1-C.7 (bilabiális és [m] hang összevonása a közös). Mindkét formáns mozgását

figyelembe véve (jobb felső ábra), C.5-C.9 a velárisok közé helyezi a [ɲ] hangot, illetve a labiodentálisok közé az [ɲ] hangot. Ha csak a második formáns mozgásait figyeljük (jobb alsó ábra), akkor az alveoláris csoportba kerül az [l] hang, illetve a posztalveoláris és a palatális egy klaszterbe kerül.

Megállapítható, hogy azonos pozíció esetén (CV vagy VC rögzítése után) a mássalhangzó-kategóriák összevonása a deformitási mutató alapján nagyon hasonló eredményre vezet, ha csak önmagában az  $F_2$ , illetve az  $F_1$ - $F_2$  formánsmozgásait vesszük figyelembe. Azonban a különböző pozíciókban eltérő összevonási sorrendet figyeltem meg. Tehát a mássalhangzó-kategóriákat jellemző deformitási mutatók különbözhetnek attól függően, hogy magánhangzó előtt vagy után állnak. Az CV esetben az átlagos deformitási mutató 0,075, míg VC pozícióban 0,083. (A különbség szignifikánsnak bizonyult Wilcoxon-próbával [Hollander & Wolfe(1973)],  $p < 0.00$ ). Ha két nazális mássalhangzó között vizsgálom az első három formáns deformitási mutatóit, azaz szűkíték a C.7, C.8. és C.9 kategóriákra, akkor már a progresszív hatás tendenciája érvényesül:  $C_{naz}V$  pozíció 0,097,  $VC_{naz}$  0,086 [vö. Horváth(2005)].

A fentieket összefoglalva, a deformitási mutatók számszerűsített módon bemutatták, hogy a magánhangzók képzése során a felső nyelvállástól az alsóig egyre növekedő mértékű deformitás jellemzi a formánsmozgásokat  $F_1$ -ben, valamint a hátulképzett magánhangzók formánsmenetei  $F_2$ -ben nagyobb alakváltozást szenvednek el, mint az elől képzetteké. A deformitási tényező nagysága függ a formánsindextől és a magánhangzó-kategóriától is, és ezek hatása nem additív. A mássalhangzó pozíciója (CV vagy VC) meghatározza, hogy az milyen mértékben változtat a követő vagy megelőző magánhangzó formánsmozgásán. A mássalhangzó hatása inkább hátraható, a VC pozícióban szignifikánsan nagyobb a deformitási mutató.

### **6.2.3.3. A stabilitási vektor értékelése**

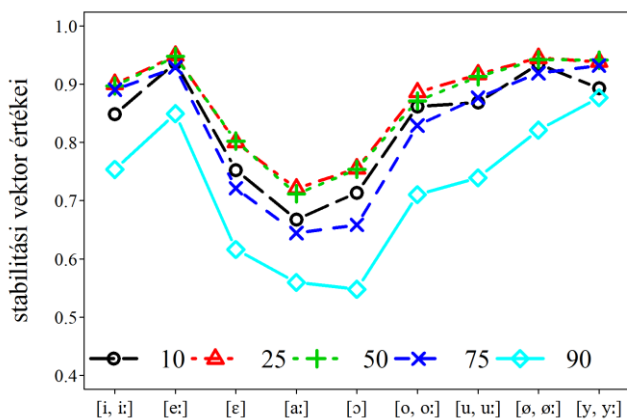
A stabilitási vektor az eddigi eredményeket árnyalja abból a szempontból, hogy a felhasznált átlaggörbe, mennyire jellemzi jól az illető hanghármast. A függelék tablóinak b) ábrájáról mindez leolvasható. A sötétebb téglalapok jobb, a világosak gyengébb stabilitást jeleznek. Számszerűen összesítve az adatokat, a legnagyobb stabilitással  $F_2$ -ben helyezkednek el a formánsmozgások (0,87 az átlagos, 5 pontra összesített mutató értéke), míg  $F_1$  esetében ez 0,81. A harmadik formáns egy-szóráson kívüli pontjainak átlagos aránya 0,55. Tehát az előző fejezetben tett, deformitási mutatóval kapcsolatos megállapításaink leginkább az első két formánsra helytállóak.



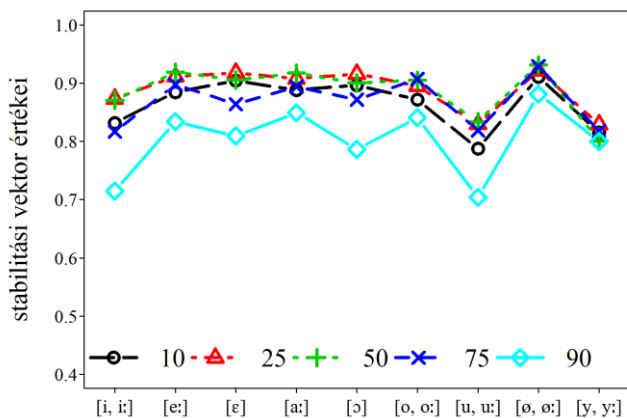
A különböző  $C_{t,1}V_mC_{t,2}$  hanghármasokon belül tehát az  $F_2$  formánsmozgásai egységesebbek az  $F_1$  formánsmozgásaihoz viszonyítva, és e kettőnél jóval nagyobb változatosság jellemzi a harmadik formáns mozgásait. A hangkörnyezet egymáshoz hasonlóbb mozgásformákat alakít ki az  $F_2$ -ben és  $F_1$ -ben  $F_3$ -hoz képest.

A stabilitási vektor egyes mérési pontjaiban (10%, 25%, 50%, 75%, 90%) számolt átlagok rendre a következők: 0,75, 0,77, 0,77, 0,75, 0,67. Egyedül a 90%-os mérési pont esetén tapasztalható némi esés.

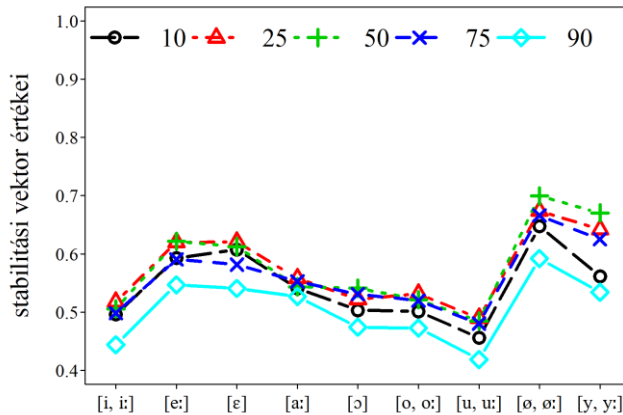
Részletesebb vizsgálathoz a magánhangzók bevonására is szükség van.



6.16. ábra. A 9 magánhangzó-minőség stabilitás vektorainak átlagos értéke az első formánsban



6.17. ábra. A 9 magánhangzó-minőség stabilitás vektorainak átlagos értéke a második formánsban



6.18. ábra. A 9 magánhangzó-minőség stabilitás vektorainak átlagos értéke a harmadik formánsban

Mindhárom formáns esetében megvizsgáltam kétszemponos varianciaelemzés módszerével, hogy a magánhangzó-minőség és a mérési pontok milyen hatással vannak a stabilitási vektorra. A két faktor által meghatározott csoportok átlagait a 6.16., 6.17. és 6.18. ábrák szemléltetik. Mindkét hatás mindhárom formáns esetén szignifikánsnak bizonyult. Az első két formáns esetében az interakció is szignifikáns volt, míg a harmadik formáns esetében ez nem teljesült. Az ábrákról leolvasható, hogy a 90%-os mérési ponthoz tartozó stabilitási vektor értéke egységesen, mindhárom formáns összes magánhangzója esetén a legalacsonyabb értékű. Ha ezt kivesszük a vizsgálatból, akkor a maradék négy mérési pontban a stabilitási vektor értékét a magánhangzók azonos irányba mozgatják, mindhárom formáns esetén.

A stabilitási vektor értékeinek elemzése alapján kimutattam, hogy az átlaggörbéhez nagyon hasonló formánsmozgások jelennek meg  $F_2$ -ben és  $F_1$ -ben (a mutató értéke 0,87 és 0,81), míg a harmadik formánsban a formánsmozgások nagy szóródást mutatnak. A mérési pontok közül kiemelkedik a 90%-os pont, amely az összes magánhangzó mindhárom formánsában a legalacsonyabb stabilitási értékkel rendelkezik. A magánhangzónként kiszámolt átlagos stabilitásokról kiolvasható, hogy főképp az első formáns esetében a nagy  $y$ -eltolással rendelkező magánhangzók, alacsony stabilitás vektor értékkel rendelkeznek (6.16. ábra). Ez felveti azt a lehetőséget, hogy a deformitási mutató és a stabilitási vektor értékei között kapcsolatot keressünk. Ezt a 6.2.3.5. fejezetben vizsgálom meg.

#### 6.2.3.4. A szigorú stabilitási mutató értékelése

A szigorú stabilitási mutatót az egyes formánstérképek és a 6.12. ábrán definiált formánsmenet-típusok összerendelésének mérésére hoztam létre.

A 6.12. táblázatban összefoglaltam, hogy a  $C_{t,1}V_mC_{t,2}$  hanghármasokat milyen arányban és milyen átlagos stabilitással jellemzik az egyes formánsmenet-típusok. Az első sor első oszlopában lévő 46 és 0,57 érték azt fejezi ki, hogy az  $F_1$ -ben az 1-SZ formánsmenet-típus a lehetséges 885 db  $C_{t,1}V_mC_{t,2}$  hanghármas 46%-át fedi le és a cellák átlagos  $SM_2$  értéke 0,57. A 6.12. táblázat alapján azt mondhatjuk, hogy a  $C_{t,1}V_mC_{t,2}$  hanghármasokhoz rendelt formánsmenet-típusok  $F_2$ -ben és főképp  $F_3$ -ban változatosabbak, mint  $F_1$ -ben, és a hozzárendelés stabilitása is nagyobb  $F_2$ -ben és  $F_1$ -ben  $F_3$ -hoz képest. Ha az 5 százalékot elérő kategóriákat tekintjük, akkor az első formáns 4 formánsmenet-típussal, az  $F_2$  és  $F_3$  5-5 kategóriával jellemezhető.

6.12. táblázat. A formánsmenet-típusok eloszlása és az átlagos  $SM_2$  értékek a  $C_{t,1}V_mC_{t,2}$  hanghármasok között (a cellák első adata egész %-ra kerekített)

Formánsmenet-típusok													Össz.
1- SZ	2- SZN	3- SZCS	4- NSZ	5- CSSZ	6- U	7- FU	8- N	9- CS	10- U*	11- FU*	12- EGYÉB		
$F_1$	46	0	10	0	0	0	33	6	3	0	2	0	100 %
	,57	-	,41	,33	-	,27	,51	,45	,43	-	,42	-	
$F_2$	36	2	1	0	0	5	7	21	26	0	0	0	100 %
	,54	,46	,42	-	-	,48	,48	,65	,64	,40	-	-	
$F_3$	20	4	2	1	1	13	11	20	26	0	0	9	100 %
	,32	,30	,32	,29	,29	,32	,33	,37	,40	-	-	-	

A kategorizálás sikeresnek mondható, hiszen a 12. egyéb (hibakategória) csak  $F_3$ -ban fordult elő 9%-os arányban. A cellákhoz rendelt formánsmenet-típusok szigorú stabilitását jelző  $SM_2$  mutatók nem tekinthetők túl magasnak. Összesítve a 6.13. táblázatban láthatók. Azonban ha a cella formánsmenetéhez a második leggyakoribb típust is hozzávesszük (feltéve, hogy köztük 1 a távolság), akkor az összevont mutató értéke jelentősen megnő. Ez igazolja, hogy a 6.12. ábra formánsmenet-típusaival jelentősen egyszerűsíthetjük a formánstérkép jellemző formánsmeneteit.

6.13. táblázat. Az összes magánhangzó formánsmozgásainak formánsenként vett szigorú stabilitás átlaga és összevont szigorú stabilitási mutató értéke

$V_m$	$SM_2$ átlag	Összevont $SM_2$ átlag
$F_1$	0,52	0,73
$F_2$	0,58	0,78
$F_3$	0,35	0,56

### 6.2.3.5. A kontextusfüggő vizsgálatok összefoglalása

A függelékben minden magánhangzó-minőség formánstérképét, a  $C_{t,1};C_{t,2}$  környezet szerinti mátrixos formában megadom. Ez megteremti a lehetőséget, hogy az adott magánhangzó összes tipikus, kontextustól függő formánsmozgását egyetlen ábrán vizsgálhassuk. Az egyes formánstérképek értelmezése a bevezetett új ábrázolási formának felel meg, azaz jelentősége van a formánsmenet-görbe és az átlagvonal közötti, valamint a két formáns közötti távolságnak is. Értékezésemben az  $y$ -eltolás viszonyait nem elemzem, ebben a magyar szakirodalom igen gazdag, például [Magdics(1965)], [Molnár(1970)], [Bolla(1978)]. A görbe  $y$  koordinátától mentes alakját a deformitási mutató írja le. A görbéből levonható következtetések megbízhatóságát a stabilitási vektor mutatja.

Az adatok alapján úgy tűnik, hogy nagyobb deformitási mutatóhoz kisebb stabilitás tartozik. A tendencia mindegyik formáns esetén teljesül (szignifikáns a Spearman-féle rangkorreláció [Hollander & Wolfe(1973)]), de a rangkorrelációs együttható értéke  $F_1$  esetében a legnagyobb  $r_s = -0,67$ , míg  $F_2$  esetén csak  $r_s = -0,39$ ,  $F_3$ -ban mindössze  $r_s = -0,11$ .

Az  $F_1$  és  $F_2$  formánsmeneteire vonatkozó megállapítások sokkal megbízhatóbbak, mint az  $F_3$ -ra vonatkozók.  $F_2$ -ben nagyobb kiugrás nélkül stabil, adott  $C_{t,1}V_mC_{t,2}$ -n belül nagyon hasonló formánsmozgásokkal találkozunk.  $F_1$ -ben némileg csökken a stabilitás, amit a szélsőségesen nagy formánsfrekvenciákkal rendelkező alsó és legalsó nyelvállású magánhangzók rendkívül alacsony stabilitása okoz. Ahogyan a rangkorreláció is bizonyította a legnagyobb dinamizmust is ebben a magánhangzó csoportban tapasztaljuk.

Mindhárom formánsra teljesül, hogy az utolsó, 90%-os mérési pontban a legnagyobb a bizonytalanság. A többi 4 mérési pozíció stabilitási vektor értéke az egyes magánhangzókban nagyon hasonló.

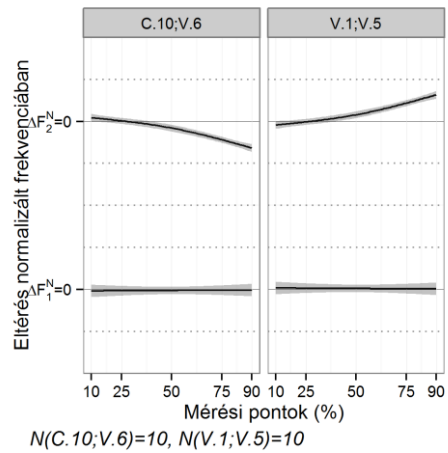
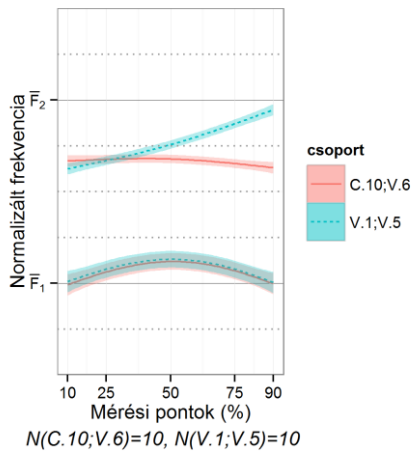
A 11 egyszerűsített formánsmenetet tartalmazó modell a FEF adatbázisomban is visszaigazolta korábbi kutatási eredményemet [Abari & Olaszy(2012)]. Az  $F_1$  kevesebb,  $F_2$  és  $F_3$  több ideális típusal jellemezhető. A jellemzés ereje csak  $F_1$  és  $F_2$  esetében megfelelő,  $F_3$  ideális típusokkal nem írható le kielégítően.

Az eredmények azt mutatják, hogy a magánhangzók  $F_1$  és  $F_2$ -ben megvalósuló formánsmozgásai jól szétválaszthatók a  $C_{t,1};C_{t,2}$  szerinti osztályozással: mind a funkcionális varianciaelemzéssel kapott átlaggörbe, mind az egyszerűsített modellel hozzárendelt formánsmenet-típus jól jellemzi a  $C_{t,1}V_mC_{t,2}$  hanghármaszt. A jellemző formánsmenetek leolvashatók a formánstérképről, vagy a deformitási mutató segítségével számszerűen is jellemezhetők. Konkrét döntést a mozgások

megbízhatóságáról az adott  $C_{t,1}V_mC_{t,2}$  hanghármas esetén a stabilitási vektor számszerű értékével, vagy a függelékben közölt szürkeárnyalatos megjelenítés segítségével hozhatunk.

#### 6.2.4. A tágabb hangkörnyezet hatása

Az előző fejezetekben háromelemű hangkapcsolatokban vizsgáltam a hangok egymásra hatását. Jelen fejezetben kiterjesztettem a „szomszédos hang” fogalmát a mért hangtól távolabbi hangokra is: a  $-2.$  és a  $+2.$  hangra. Így a háromelemű szekvenciák (triád) helyett hangnégyes vagy hangötös sorozatokat vizsgálok, amelyekben a köztes magánhangzó formánsmeneteinek alakulása a kérdés, a  $-2.$  és  $+2.$  hang függvényében. Hipotézisem szerint, vannak esetek, amikor a  $-2.$ , illetve a  $+2.$  hang is befolyásolja a vizsgált magánhangzóban a formáns mozgási irányát, mértékét, és ez az SSANOVA kiterjesztett grafikus ábrájával is kimutatható. A hipotézishez kapcsolható az a tény is, hogy az utóbbi évtizedben kidolgozott statisztikai parametrikus elvekre támaszkodó beszéd szintetizátorokban is 5 elemű hangkapcsolatokra készítik el a modellt, ami a beszédelőállításához szükséges paraméter sorozatot szolgáltatja [Zen et al.(2009)]. Az előző pontban bevezetett SSANOVA módszer és az általam kidolgozott, hozzá kapcsolódó képi megjelenítés már alkalmas konkrét 5 elemű hangsor vizsgálatára is. A FEF adatbázisban [vøhi] és [vøho] hangsorokra 10-10 mintát találtam (beszélőnként 1-1 realizáció a *szavahihetőségéről* és a *tolva hozzá* bemondásrészekből). Az [ø] realizációk második formánsának menete szignifikáns eltérést mutat a két csoportban, amely a  $+2.$  hang hatásának köszönhető (6.19. ábra). A példában tehát a [h] hangon átívelő hatás látható, és ezt a hatást a  $+2.$  hang formánsszerkezete határozza meg. Az [i] hang felfelé húzza az  $F_2$  mozgását, az [o] pedig lefelé. Hasonló hatás mutatható ki más ötelemű kapcsolatokban is, azonban a FEF adatbázis nem elég nagy ahhoz, hogy ezeket statisztikai vizsgálat alá vonjuk. A fenti példához hasonló átívelő hatás figyelhető meg a  $C_1$  vagy  $C_2$  pozíciójú [v], [b], [d], hangokban is, amennyiben a  $-2.$ ,  $+2.$  hang az [i] vagy az [o, u]. Ennek szisztematikus vizsgálatára érdemes külön beszédadatbázist tervezni, mert akkor hasonló módon lehet tipizálásokat végezni, mint ahogy azt a  $C_{t,1}V_mC_{t,2}$  hanghármasokkal kapcsolatokban tettem.



6.19. ábra. A +2. hang szignifikáns hatásának bemutatása a magánhangzó második formánsára. A bal oldali ábrán a spline-okkal modellezett jellemző görbék láthatók a két csoportban. A jelmagyarázat a -2. és +2. hang kategóriáját tartalmazza (C.10;V.6=[vɔho]; V.1;V.5=[vɔhi]). A jobb oldali ábrán a csoport és interakciós hatás együttesen jelenik meg

## 7. ÖSSZEFOGLALÁS

Értekezésemben a magyar magánhangzók formánszerkezetének több mérési ponton alapuló statisztikai modellezésére tettem kísérletet. A magánhangzók időben változó spektrális szerkezetének kontextusfüggő leírását a magánhangzó teljes időtartamából vett 5 mérési pontra (10, 25, 50, 75 és 90%-os) és az első három formánsra alapoztam. Kutatásom alapját, az eredmények általánosíthatósága miatt, egy saját készítésű, nagyméretű beszédadatbázis képezte, amely közel 5 millió formánsadatot tartalmaz.

### 7.1. A kutatás eredményei és jelentősége

Kidolgoztam egy reprodukálható kutatási modellt a beszéttudomány néhány területére. A modell a futtatható cikk koncepciójára épül, amely szerint a publikálandó szöveget és azokat a programkódokat, amelyek a publikáció ábráit, táblázatait és számszerű végeredményeit alkotják, egyetlen állományban összefogva kell tárolni. Az általam kidolgozott modellben a programsorok R parancsokat jelentenek, a publikáció szövege pedig LaTeX vagy Markdown formátumú. A modell fontos összetevője az Emu rendszer, amelyben a beszédadatbázisunkat fejleszthetjük. Az Emu közvetlen kapcsolattal rendelkezik az R felé, sőt a beszédjel sokoldalú feldolgozását is lehetővé teszi.

Ez a modell jól illeszkedik a fonetikai alap kutatások természetéhez, felhasználása főképp ezen a területen várható. A fonetikai kutatások alapja a beszédadatbázis, amelyet a modell szerint Emu formátumban kell létrehozunk. Ez nem jelent nagy megszorítást, mert az Emu nyitott más rendszerek felé, így akár a kutatás utolsó fázisában is konvertálhatunk Emu-ba, ha más rendszerben kezdtük az adatbázis építését. Az alap kutatás része a beszédjel akusztikus paramétereinek kinyerése, amit maga az Emu is támogat, de modellem szerint tetszőleges „külső” eszközt is igénybe vehetünk. A kinyert akusztikus adatok statisztikai feldolgozását, valamint táblázatokba, ábrákba és mutatókba sűrítését R parancsok segítségével végezhetjük. Utolsó lépésként, a publikáció generálását is az R biztosítja. A modell alkalmazásának legfontosabb akadálya a beszédadatbázis korlátolt elérése, terjesztése lehet. Ebben az esetben a reprodukálható modell saját kutatásunk jövőbeni megismétlését biztosítja, aminek jelentőségét nem szabad alábecsülni.

A formánsmenetek tanulmányozására létrehozott nagyméretű adatbázisom a BME TMIT Beszédakusztikai Laboratóriumban létrehozott beszédadatbázison

alapul. Általános eljárást dolgoztam ki, amely ebből a formátumból Emu szabványnak megfelelő beszédadatbázist állít elő. Az algoritmus egyes lépései R szkriptek formájában rendelkezésre állnak a laboratórium munkatársai számára, így a jövőben más adatbázisok konvertálása is elvégezhető.

Nagy adatbázis esetén a formánsértékek meghatározása csak gépi eszközökkel valósítható meg. Eljárást dolgoztam ki, amely egyrészt a gépi mérés eredményeiben képes a hibákat detektálni, másrészt a megtalált hibákat kijavítani. A gépi mérések eredményesebb végrehajtására a Praat formánsmérő algoritmusát optimális bemenő paraméterekkel láttam el. Korábbi kutatások tapasztalataiból kiindulva tipizáltam a formánsmérések lehetséges tévesztéseit és 5 hibatípus automatikus felismerésére készítettem szabály alapú algoritmust. A hibatípusok: hiányzó adat, formánstávolság hiba, formánsorszám-tévesztés, kiugró adat és hibás formánsmenet. Az egyes hibák felismerése eltérő nehézségű, több közülük segéd táblázatok összeállítását igényli. Bizonyítottam a hibák felismerésének hatékonyságát. A detektált hibák javítására algoritmust és hozzá kapcsolódó vizuális grafikus felhatalmazó felülettel rendelkező programot fejlesztettem ki. Az alkalmazás a hibák javítása során a kézi és az automatikus (gépi) módszert is támogatja.

A formánsmérés hibáinak javítása után rendelkezésre áll az első nagyméretű magyar formánsadatbázis, amely csaknem 5 millió formánsfrekvencia-értéket tartalmaz, magánhangzónként  $3 \times 5 = 15$  db-ot. Ez az adatállomány egyetlen tabulátorral tagolt szöveges állományban foglal helyet, amelynek általam kialakított szerkezete biztosítja az optimális tárolást. Ugyanis a választott szöveges állomány formátum nemcsak a gépi feldolgozást könnyíti meg, hanem a kutató számára is egyszerű betekintést enged az adatokba. Magában a formánsadatbázisban csak a legfontosabb információkat tárolom: a magánhangzó SAMPA kódja, a mérés pontos ideje ms-ban, a mérési pozíció és a mért három formáns frekvenciaértéke. Ezen túl a hangminta azonosítója is szerepel az állományban, így a hanghullám és az annotáció is elérhetővé válik. A kialakított szerkezet rugalmasan illeszkedhet az épp aktuálisan folyó kutatáshoz. A mérési pozíciók száma (amely most 5) tetszőlegesen csökkenthető vagy növelhető. Ha túl nagyra nőne az állomány mérete, akkor hangmintánként szétbontást alkalmazhatunk anélkül, hogy a szerkezeten változtatni kellene.

Az ellenőrzött formánsadatokon a Lobanov-féle módszerrel magánhangzó normalizálást végeztem, amely könnyen értelmezhető és jól használható a beszélők közötti különbségek eltávolítására. A beszélő személy mint változó (és így a személy neme változó is) kikerült a vizsgálatból, minden további eredmény a normalizált formánstérre vonatkozik. A normalizálás megbízhatóságát ábrával



szemléltettem, a magánhangzó 50%-os mérési pontjainak összehasonlítása alapján. Egy új ábrázolási módot vezettem be, mellyel egyetlen ábrán jeleníthetjük meg a három normalizált formánstérbe konvertált formánst. Egymás fölé rajzoltam a normalizált görbéket a formánsindexek növekvő sorrendjében úgy, hogy az első formáns értékeihez képest 4 egységgel feltolva jelennek meg a második formáns normalizált értékei, a harmadik formáns görbéi pedig 8 egységgel lesznek feltolva az elsőhöz képest (azaz 4 egységgel a másodikhoz képest). Ez az ábrázolás több szempontból is informatív. Egyrészt minden beszélő adatát egyesítve tartalmazza, ami jelentős egyszerűsítést jelent, hiszen nem kell például nem vagy életkor szerint bontással szaporítani az ábrák számát. A formánsindexek természetes sorrendjükben lentől felfelé növekvő sorrendben helyezkednek el. A berajzolt átlagvonalak jelzik, hogy az adott formánsmenetek az átlaghoz képest hol helyezkednek el. Továbbá a görbék egymáshoz mért távolságának is van jelentősége. Ezt az ábrázolást alkalmaztam például az SSANOVA grafikus eredményeiben és a formánstérképek megrajzolásánál is.

Az új ábrázolási mód könnyen adaptálható, és minden olyan esetben felhasználható, amikor eredendően különböző skálájú formánsok együttes ábrázolása a követelmény.

Az SSANOVA ismert grafikus eredményábráira alkalmaztam az új grafikus ábrázolási módot, és így statisztikai eszközökkel is kimutattam, hogy a magánhangzót követő +2. hang befolyásolhatja a magánhangzóban létrejövő formánsmozgást.

A CVC hanghármasok magánhangzóiban realizálódó formánsmenetek tanulmányozásához új vizsgálati eszközöket alakítottam ki. Ezek a formánstérkép, a deformitási mutató és a stabilitási vektor. A formánstérkép a CVC-re jellemző három formáns tipikus alakját mutatja meg. Az egyes formánstérképeken az  $F_1$ - $F_2$ - $F_3$ -ra jellemző formánsmozgásokat a funkcionális varianciaelemzés módszerével rajzoltam meg. A funkcionális varianciaelemzés modelljének illeszkedését az  $R^2(t)$  determinációs együttható-függvény segítségével vizsgáltam. A mássalhangzó környezetek formánsmozgató hatásának mérésére deformitási mutatót dolgoztam ki. Minden CVC-t két deformitási mutató ír le: a CV-re és a VC-re vonatkozó. A stabilitási vektort a vizsgált 5 mérési pontban az átlaggörbétől fél szóráson belüli (összesen egy szóráson belüli) pontok arányát méri, azaz a CVC-n belüli formánsmozgások hasonlóságát méri.

Az egy magánhangzóra vonatkozó formánstérképeket mátrixos formába rendeztem, ezzel megteremttem a lehetőséget, hogy az adott magánhangzó összes tipikus, kontextustól függő formánsmozgását egyetlen ábrán vizsgálhassuk. A további két újonnan bevezetett eszköz a formánstérképről vizuálisan leolvasható következtetések számszerű bizonyítását teszi lehetővé. A deformitási mutató a formánsmozgás alakját, a stabilitási vektor az eredmény megbízhatóságát alakítja mutatóvá. A két mutató értékeinek nagysága alapján sokféle rangsor felállítható a magánhangzókra, a környező mássalhangzókra vagy magukra a CVC hanghármasokra. Ezeket az értekezésemben publikálom. Ezek összesítéséből kiolvasható, hogy az  $F_1$  és  $F_2$  formánsmeneteire vonatkozó megállapítások sokkal megbízhatóbbak, mint az  $F_3$ -ra vonatkozók.  $F_2$ -ben nagyobb kiugrás nélkül stabil, adott CVC-n belül hasonló formánsmozgásokkal találkozunk.  $F_1$ -ben némileg csökken a stabilitás, amit a szélsőségesen nagy formánsfrekvenciákkal rendelkező alsó és legalsó nyelvállású magánhangzók rendkívül alacsony stabilitása okoz. A legnagyobb dinamizmust is ebben a magánhangzó csoportban tapasztaltam.

A bevezetett új vizsgálati eszközök más nyelvekre is adaptálhatók. A környezetek számát emelhetjük, például magánhangzók bevonásával, de ekkor a formánstérképek mátrixos elrendezését érdemes átgondolni.

Bemutattam, hogy egy egyszerűsített modell felhasználásával hogyan jósolhatók a formánsmozgások a hangkörnyezet ismeretében anélkül, hogy méréseket végeznénk. Ideális formánsmenetek rendszerét definiáltam, amelyben a lehetséges formánsmenet-típusok 11 kategóriáját, egyenesek és parabolák felhasználásával írtam le. Távolságot számoltam két formánsmenet-típus között úgy, hogy minimálisan hány pontnyi mozgással juthatunk el az egyik típusból a másikba. Az 1 távolságra lévő típusok nagyon hasonló mozgást írnak le. A célom az volt, hogy a formánstérképeken szereplő jellemző görbéket a 11 formánsmenet-típus valamelyikével helyettesítsem. A helyettesítés hatékonyságát a szigorú stabilitási mutatóval és annak gyengített változatával az összevont szigorú stabilitási mutatóval mértem (ez utóbbi esetben az 1 távolságra lévő formánsmeneteket is számba vettem). A jóslás mértéke  $F_1$  és  $F_2$  esetében 73 és 78% pontosságú, míg  $F_3$ -nál ez 56%. Az  $F_3$  alacsony stabilitásának megfelelően a jóslása nehézségekbe ütközik adott CVC hanghármasban. Az  $F_1$  és  $F_2$  esetében a jóslás mértéke kielégítő.  $F_1$  mozgásai kevesebb formánsmenet-típussal írhatók le  $F_2$ -höz és  $F_3$ -hoz képest.

Új eredményeimmel hozzájárultam, hogy matematikailag is magyarázhatóvá váljanak a formánsokkal kapcsolatos eddigi fonetikai vizuális megfigyelések és leíró jellegű magyarázatok. A formánsmozgások rendszerezése további

formánsvizsgálatok kiindulópontjául szolgálhat, például a koartikulációs hatások pontosítása terén. Az eredményeket az alkalmazott fonetikában is felhasználhatjuk (beszédszintézis, beszédfelismerés, beszélő személy felismerése, beszédhangképzés zavarainak vizsgálata). A formánstérképek jó alapot adnak hatékonyabb hiba-felismerő és formánsérték-javító algoritmusok kidolgozásához, így egyre kevesebb manuális munka árán juthatunk egyre nagyobb méretű sokbeszélős formánsadatbázishoz.

## **7.2. Jövőbeni tervek**

A fonetikai jelenségek vizsgálata során a kutatók törekednek az artikuláció-akusztikum-percepció hármasságát megteremteni. Ez az értekezés a formánsmozgások elemzését az akusztikum szempontjainak figyelembe vételével kísérelte meg. A vizsgálat természetes folytatása lehet a jövőben annak megállapítása, hogy milyen formánsmozgás eltérések relevánsak a percepció számára.

A formánsmozgások teljes körű, minden részletre kiterjedő vizsgálata nem lehetett ennek az értekezésnek a tárgya. Sok olyan kérdés merülhet fel, amely a rendelkezésre álló eszközök segítségével, láthatóan, már vizsgálható lenne. Ezek közé tartozik például a magánhangzó kvantitás hatása a formánsmozgásra, vagy a rövid és hosszú magánhangzók szétbontott vizsgálata. Szinte kimeríthetetlen számú vizsgálatot végezhetünk a kis stabilitású (nagy szórású) cellák esetében a második hangkörnyezet figyelembe vételével. A szerző bízik ezen kérdések jövőbeni megválaszolásában.

## **8. SUMMARY**

In my thesis, I have tried to provide a statistical model of the formant structure of Hungarian vowels. The context-dependent description of the spectral structure of vowels changing in time was modelled by 5 points of measurement (10, 25, 50, 75 and 90%) taken from the full duration of the vowel and on the first three formants. In order to conclude general purpose results, I based my research on a large-scale speech database of own development (about 5 million formant data).

### **8.1. Research results and their significance**

A reproducible research model was planned and realised for some fields of speech science. The model is based on the concept of executable articles, according to which the text to be published as well as the program codes that make up the figures, tables and final numerical results of the publication are stored together in a single file. In the model I developed, the program lines mean R commands, and the text of the publication is of LaTeX or Markdown format. An important component of the model is the Emu system, in which the speech database can be developed. Emu has direct contact toward R, and it makes possible the manifold processing of the speech signal.

This model fits well the nature of basic phonetic research, and its application is expected mainly in this field. The basis of phonetic research is the speech database, which have to be generated in Emu format according to the model. This does not mean great restriction, as Emu is open towards other systems, which means that it is possible to convert to Emu even in the last phase of the research, if we started to build the database in another system. A part of the basic research is the obtainment of the acoustic parameters of the speech signal, which is supported by Emu itself, but according to my model it is also possible to use optional „external” tools. The statistic processing of the acoustic data obtained as well as their concentration in tables, figures and indices can be implemented with the help of R commands. As a final step, the generation of the publication is also provided by R. The most significant obstacle in applying the model is the limited access to and dissemination of the speech database. In this case, the reproducible model ensures the future repetition of our own research. We should not underestimate its significance.

The large-scale database I created for analysing formant trajectories is based on the speech database generated in BME TMIT Laboratory of Speech Research. I

elaborated a general procedure which creates a speech database fulfilling Emu standards from the aforementioned format. Certain steps of the algorithm are available in the form of R scripts for colleagues of the laboratory, thus other databases can also be converted in the future.

In the case of large speech databases, the determination of formant values can only be executed by computerized tools. I elaborated a process which, on the one hand, is capable of detecting errors in the results of automatic measurements, on the other hand it is able to correct errors found. I provided optimal input parameters to the formant measuring algorithm of Praat to implement automatic measurements more successfully. Starting out from experience provided by previous research, I typified the possible mistakes of formant measurements, and I elaborated a rule-based algorithm for the automatic recognition of 5 formant value error types. The error types are: missing data, formant distance error, bad formant sequential number, outlier formant and incorrect formant trajectory. The recognition of individual errors is of different complication, many of them need auxiliary tables to be compiled. I proved the efficiency of error recognition. I developed an algorithm and a related program with visual graphic user surface to correct errors detected. The application supports both the manual and automatic method in correcting errors.

Following the correction of the errors of formant measurement, the first large-scale Hungarian formant database is available, which contains almost 5 million formant frequency values,  $3 \times 5 = 15$  pieces per vowel. This data file is in a tab delimited text file, and its structure I set up ensures the most optimal storing. The chosen text file format facilitates not only machine processing, but it also provides easy insight in the data for the researcher. I store only the most important information in the formant database itself: the SAMPA code of the vowel, the exact time of the measurement in ms, the measurement position, and the frequency values of the three formants measured. Beyond the above, the file contains the identifier of the sound sample as well, thus the sound wave and the annotation become available, too. The structure set up can flexibly fit the research conducted at the given time. The number of measurement positions (currently 5) can be decreased or increased optionally. If the size of the file grew too large, we can apply division per sound samples without having to change the structure.

Using Lobanov's method, I conducted vowel normalization on the formant data checked, which is good interpretable and well applicable for eliminating differences among speakers. The speaking individual as variable (and thus the individual's gender variable as well) was taken out of the examination, and all

further results regard the normalized formant space. I illustrated the reliability of normalization with a graph, on the basis of comparing the 50-percent measurement points of vowels. I introduced a new method of representation by which it is possible to present the three formants converted in normalized formant space in one single graph. I drew normalized curves above each other, in the ascending order of formant indices, in a way that the normalized values of the second formant appear 4 units pushed up compared to the values of the first formant, and the curves of the third formant are 8 units pushed up compared to the ones of the first (that is, 4 units above the second). This representation is informative from several points of view. On the one hand, it contains the data of all speakers in a unified form, which means significant simplification, as the number of figures does not need to be multiplied by division according to gender or age. Formant indices are in their natural order, from bottom to top in an ascending order. The average lines drawn in show the place of the given formant trajectories compared to the average. Furthermore, the distance of curves from each other also has significance. I used this depiction in the graphic results of SSANOVA, for example, and also in drawing the formant maps.

The new method of depiction is easily adaptable and it can be used in any cases where formants of originally different scales are required to be represented together.

I applied the new graphic representation for the known graphic result figures of SSANOVA, and thus it was also with statistical tools that I presented that sound +2 following the vowel can influence the formant movement coming about in the vowel.

I set up new tools for analysis to examine formant trajectories coming about in the vowels of the CVC triphones. These are the formant map, the deformity index and the stability vector. The formant map shows the typical form of the three formants characteristic of the vowel in a CVC structure for all possible C types. I drew the formant movements characteristic of  $F_1$ - $F_2$ - $F_3$  on the individual formant maps with the method of functional variance analysis. The fitting of the model of functional variance analysis was examined with the help of  $R^2(t)$  determination coefficient function. I elaborated a deformity index to measure the formant moving effect of consonant environments. Every CVC is described by two deformity indices: one referring to the CV and the other to the VC. The stability vector measures the ratio of points within a half deviation from the average curve (altogether within one deviation) in the 5 measure points examined, that is, it measures the similarity of formant movements within the CVC.

I arranged formant maps referring to the given vowel in a matrix format, thereby creating the possibility of examining all the typical, context-dependent formant movements of the given vowel in a single graph. One cell of the matrix represents an actual CVC combination. In one formant map (121 cell drawings show the formant movements inside the vowel similarly to the spectrograms but only with solid curves. Altogether 9 formant maps – for the 9 vowel types – describe the full CVC data for Hungarian concerning the  $F_1$ ,  $F_2$  and  $F_3$  formant movements inside the vowels.

The further two newly introduced tools allow to prove numerically the conclusions visually apparent from the formant map. The deformity factor transforms the form of formant movement into an index, whereas the stability vector does likewise with the reliability of the result. On the basis of the values of the two indices, various hierarchies can be set up for the vowels, the surrounding consonants and to the CVC sound triplets as well. I publish them in my thesis. The total of them show that the statements concerning the formant trajectories of  $F_1$  and  $F_2$  are much more reliable than the ones concerning  $F_3$ . We find stable formant movements without significant deviation in  $F_2$ , which are similar within a given CVC. Stability somewhat decreases in  $F_1$ , which is due to the extremely low stability of near-open and open vowels with extremely huge formant frequencies. I experienced the greatest dynamism in this group of vowels.

The new tools for analysis introduced can be adapted to other languages as well. The number of environments can be increased by the introduction of vowels, for example, but the matrix arrangement of formant maps is worth reconsidering in this case.

With the application of a simplified model, I presented how formant movements can be predicted in awareness of the sound environment but without taking measurements. I defined the system of ideal formant trajectories, in which 11 categories of the possible formant trajectory types were described with the help of lines and parabolas. I calculated the distance between two formant trajectory types by determining the number of points of movement to be taken to get from one type to the other. The types at a distance of 1 show very similar movements. My aim was to substitute the characteristic curves in the formant maps by one of the 11 formant trajectory types. I measured the efficiency of the substitution by the strict stability index and its weakened version, the joint strict stability index (in the latter case, formant trajectories at a distance of 1 were also taken into consideration). The accuracy of prediction in  $F_1$  and  $F_2$  was 73 and 78%, while in  $F_3$  it was 56%. In line with the low stability of  $F_3$ , its prediction also proves

difficult in a given CVC sound triplet. In the case of  $F_1$  and  $F_2$ , the degree of prediction is satisfactory. The movements of  $F_1$  can be described by fewer formant trajectory types compared to  $F_2$  and  $F_3$ .

With my new results, I contributed to making mathematically explainable the described visual observations by phoneticians and descriptive explanations related to formants.

## **8.2. Future plans**

Researchers try to create the unity of the articulation-acoustics-perception triad when analysing phonetic phenomena. The present thesis aimed to analyse formant movements while taking into account the aspects of acoustics. The natural continuation of the analysis can be the future statement of formant movement deviations relevant for perception.

The comprehensive, fully-detailed analysis of formant movements could not be the subject of the present thesis. There may occur many questions which could be examined visibly with the tools already available. The effect of vowel quantity on formant movement, for example, or the separate analysis of short and long vowels are all such fields. An almost inexhaustible number of analyses could be conducted in the case of cells of little stability (of great deviation), taking the second sound environment into account. The author hopes to answer these questions in the future.

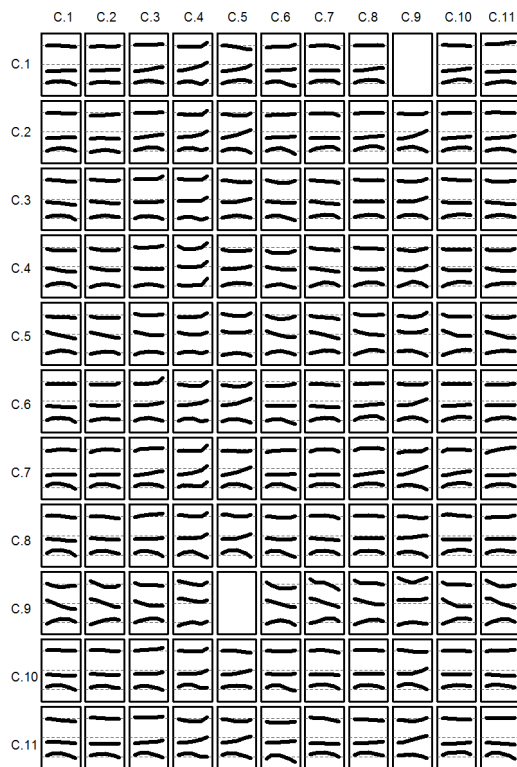


## FÜGGELÉK

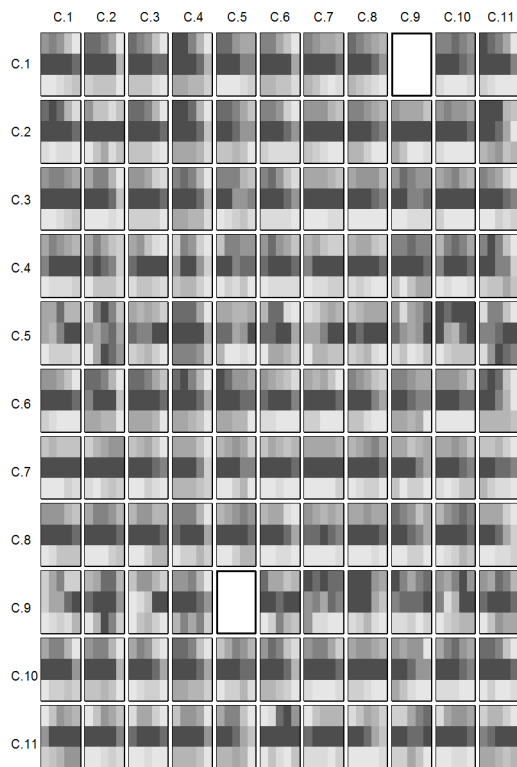
A függelék a 6. fejezetben ismertetett vizsgálatok legfontosabb eredményeit tablószerű formátumban mutatja be. A következő 9 oldalon a 9 db magánhangzó-minőség tablója szerepel. Az egyes tablók a következő elemeket tartalmazzák:

- a) A magánhangzó-minőség  $C_{t,1}; C_{t,2}$  hangkörnyezetekhez tartozó formánstérképei mátrixos elrendezésben. A sornevek a megelőző, az oszlopnevek a követő mássalhangzó kategóriáját adják meg. Ha a FEF adatbázisban nincs minta az adott  $C_{t,1}V_m C_{t,2}$  hanghármasra, akkor üres (fehér színű) ábra jelenik meg a formánstérkép helyén. A formánstérképek három vízszintes vonala,  $F_1$ ,  $F_2$  és  $F_3$ -ra a beszélőnkénti átlag közös 0 pontját jelentik a normalizált formánstérben. Az ábra részletesebb értelmezése a 6.2.2.1. fejezetben található.
- b) A magánhangzó-minőség stabilitási vektorainak grafikus képe az a) pont formánstérképeivel megegyező elrendezésben. Cellánként  $3 \times 5$  db szürke téglalap jelzi, hogy az adott formáns ( $F_1$ ,  $F_2$  és  $F_3$ , 3 db), adott mérési pontjában (5 db) mennyire hasonlók a formánsértékek. A nagyobb hasonlóságot sötétebb színű téglalap jelöli. Az ábra részletesebb értelmezése a 6.2.2.2. fejezetben található.
- c) Az  $R^2(t)$  determinációs együttható-függvény grafikonja a három formánsra. Az a)-ban közölt formánstérképek prediktált görbéinek létrehozásához a funkcionális varianciaelemzés modelljét használtam fel. Az  $R^2(t)$  a modell illeszkedését méri. Az  $R^2(t)$  ismertetése a 2.9.3. fejezetben, az ábra részletesebb értelmezése a 6.2.2.1. fejezetben található.
- d) A 11 formánsmenet-típust tartalmazó egyszerűsített modell és a magánhangzó illeszkedésének értékelése. Az első táblázat a legalább 5%-os gyakorisággal rendelkező formánsmenet-típusokat sorolja fel. A második táblázat a szigorú stabilitási mutatók átlagát és az összevont szigorú stabilitási mutatók átlagát tartalmazza, amely az illesztés pontosságát becsüli meg. Az egyszerűsített modelltől részletesebben a 6.2.2.3. fejezetben olvashatunk.

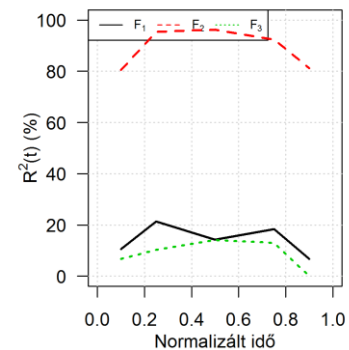
1. tábló. [c]



a) A formánstérképek



b) A formánsmozgások stabilitása



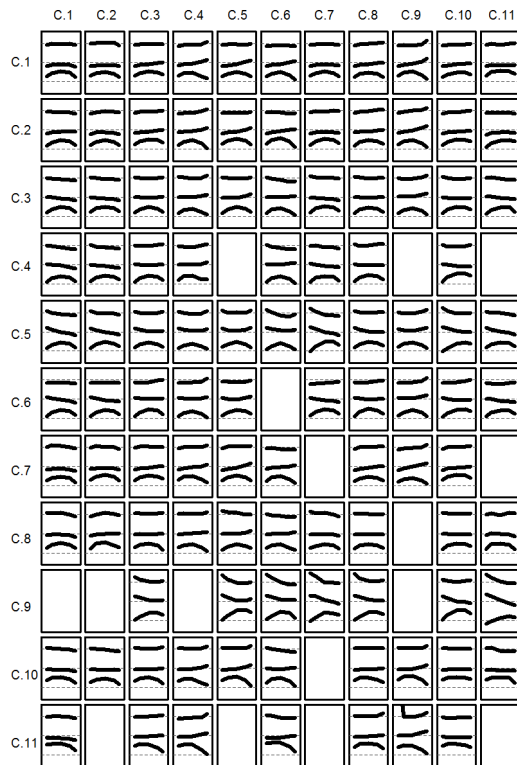
c) A hangkörnyezet magyarázó hatása

F <sub>1</sub>	7-FU (65%), 3-SZCS (27%)
F <sub>2</sub>	1-SZ (45%), 8-N (26%), 9-CS (23)
F <sub>3</sub>	9-CS (27%), 6-U (23%), 1-SZ (21%), 8-N (13%), 2-SZN (7%), 7-FU (6%)

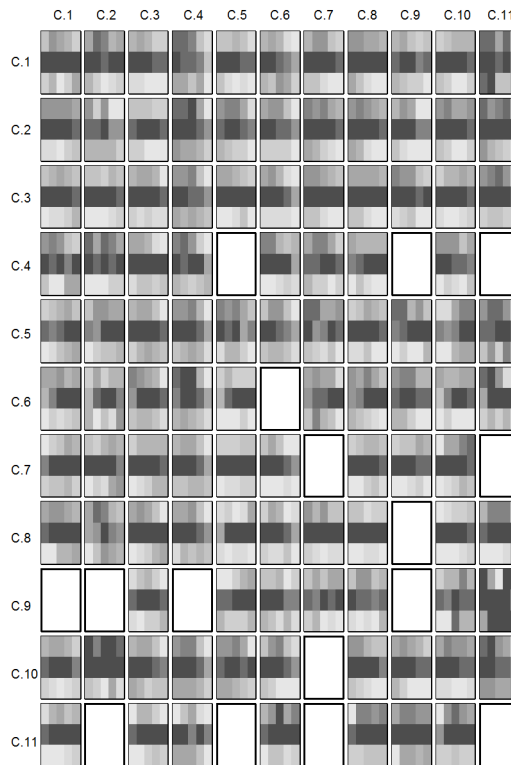
	SM <sub>2</sub> átlag	Összevont SM <sub>2</sub> átlag
F <sub>1</sub>	0,44	0,67
F <sub>2</sub>	0,66	0,84
F <sub>3</sub>	0,28	0,48

d) A jellemző formánsmenet-típusok és a cellák szigorú stabilitási mutatóinak átlaga

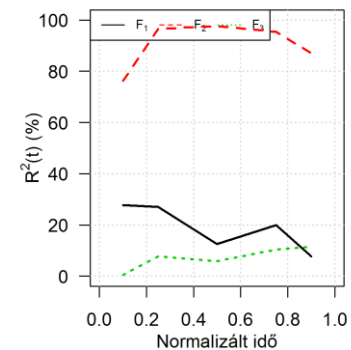
2. tabló. [a:]



a) A formánstérképek



b) A formánsmozgások stabilitása



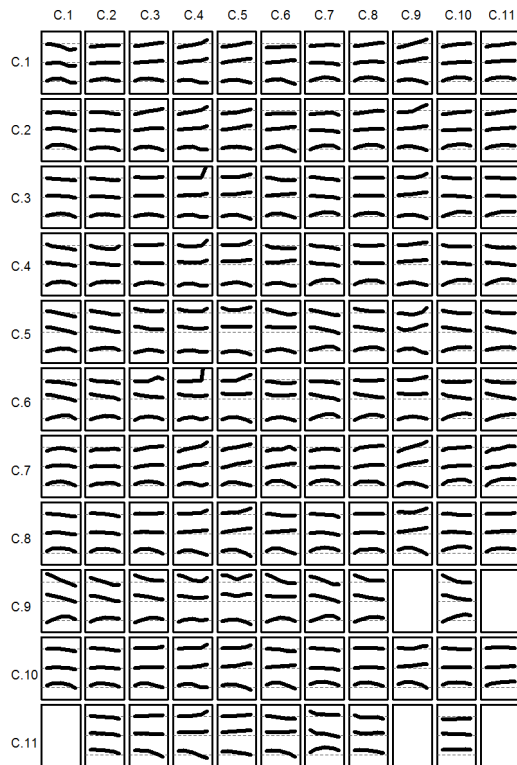
c) A hangkörnyezet magyarázó hatása

F <sub>1</sub>	7-FU (77%), 11-FU (15%)
F <sub>2</sub>	1-SZ (30%), 8-N (29%), 9-CS(24%), 6-U (6%), 2-SZN (5%), 7-FU (5%)
F <sub>3</sub>	9-CS (26%), 6-U (23%), 8-N (22%), 1-SZ (15%), 2-SZN (8%)

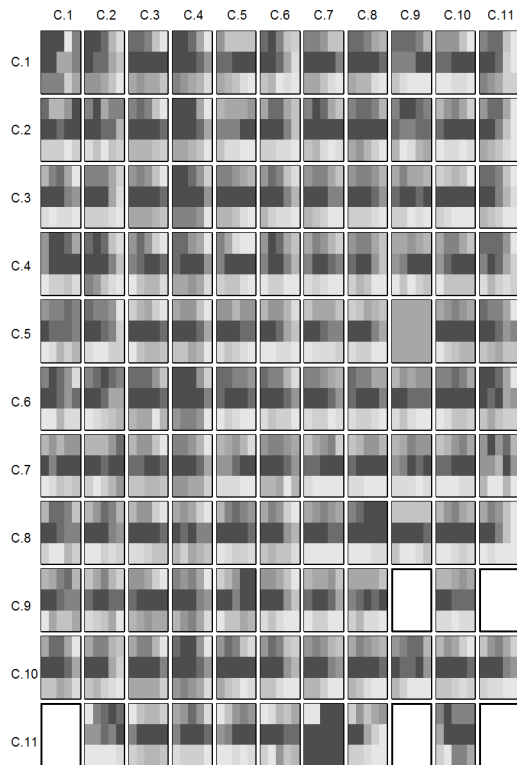
	SM <sub>2</sub> átlag	Összevont SM <sub>2</sub> átlag
F <sub>1</sub>	0,58	0,79
F <sub>2</sub>	0,55	0,74
F <sub>3</sub>	0,31	0,49

d) A jellemző formánsmenet-típusok és a cellák szigorú stabilitási mutatóinak átlaga

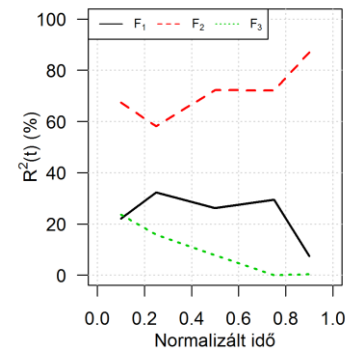
3. tabló. [ε]



a) A formánstérképek



b) A formánsmozgások stabilitása



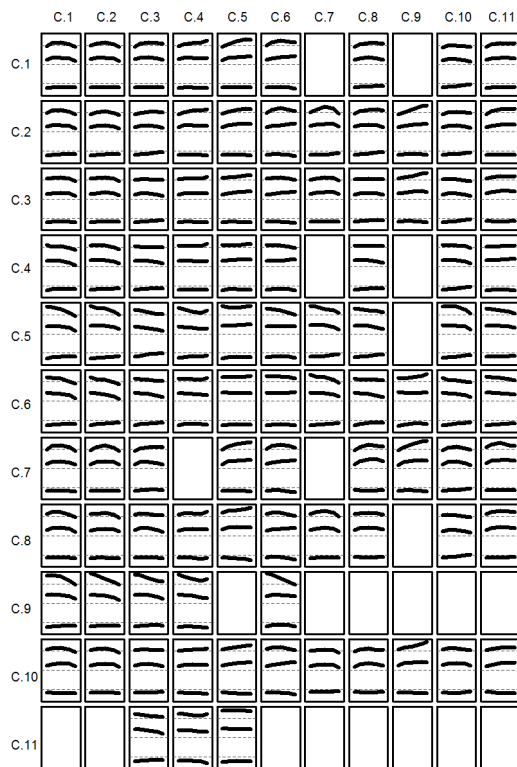
c) A hangkörnyezet magyarázó hatása

F <sub>1</sub>	7-FU (56%), 3-SZCS (29%), 8-N (9%)
F <sub>2</sub>	1-SZ (37%), 9-CS (28%), 8-N (25%), 3-SZCS (5%)
F <sub>3</sub>	9-CS (36%), 8-N (32%), 1-SZ (16%), 6-U (7%), 7-FU (5%)

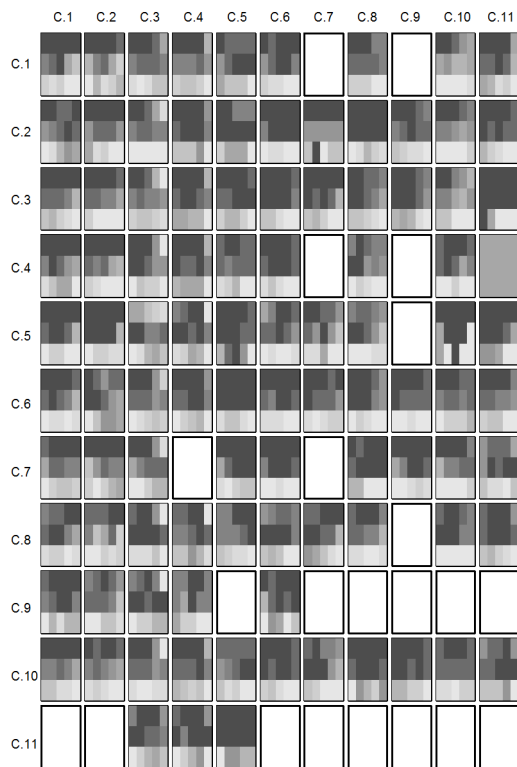
	SM <sub>2</sub> átlag	Összevont SM <sub>2</sub> átlag
F <sub>1</sub>	0,50	0,72
F <sub>2</sub>	0,57	0,75
F <sub>3</sub>	0,40	0,57

d) A jellemző formánsmenet-típusok és a cellák szigorú stabilitási mutatóinak átlaga

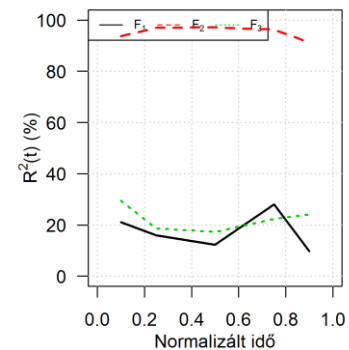
4. tabló. [e:]



a) A formánstérképek



b) A formánsmozgások stabilitása



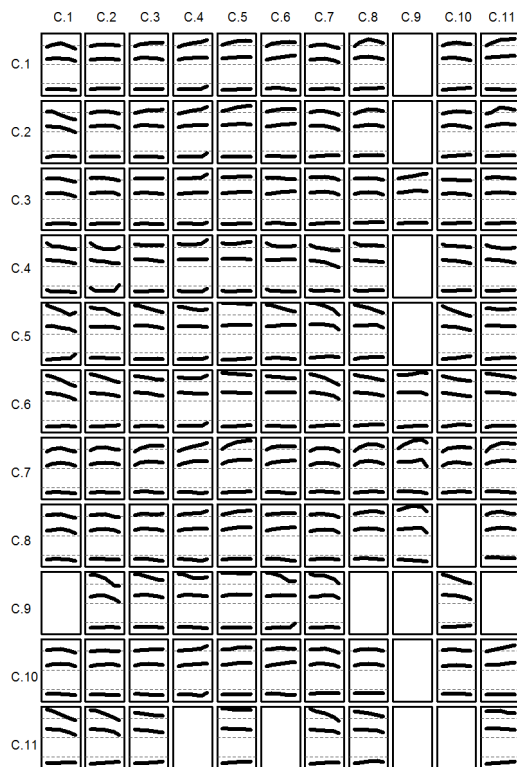
c) A hangkörnyezet magyarázó hatása

F <sub>1</sub>	1-SZ (76%), 8-N (15%), 7-FU (6%)
F <sub>2</sub>	7-FU (31%), 1-SZ (27%), 9-CS (25%), 8-N (16%)
F <sub>3</sub>	7-FU (29%), 9-CS (26%), 1-SZ (16%), 8-N (15%), 3-SZCS (10%)

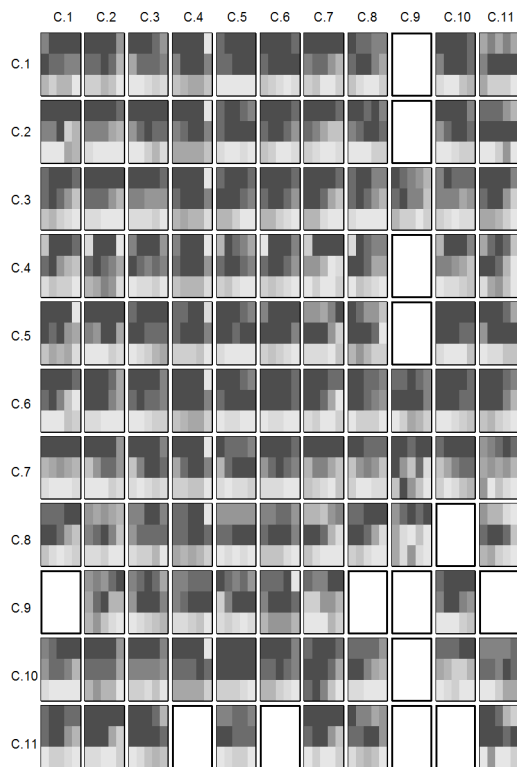
	SM <sub>2</sub> átlag	Összevont SM <sub>2</sub> átlag
F <sub>1</sub>	0,57	0,82
F <sub>2</sub>	0,58	0,80
F <sub>3</sub>	0,43	0,64

d) A jellemző formánsmenet-típusok és a cellák szigorú stabilitási mutatóinak átlaga

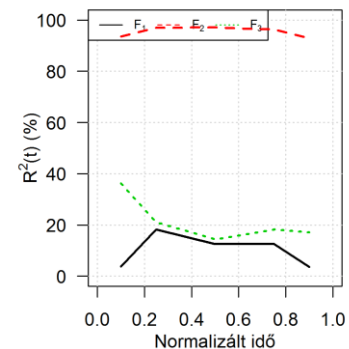
5. tabl. [i, i:]



a) A formánstérképek



b) A formánsmozgások stabilitása



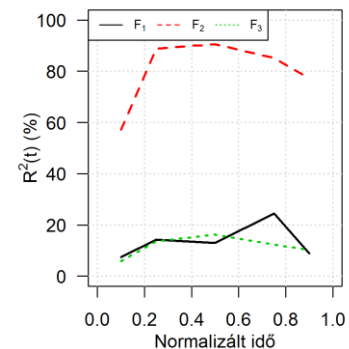
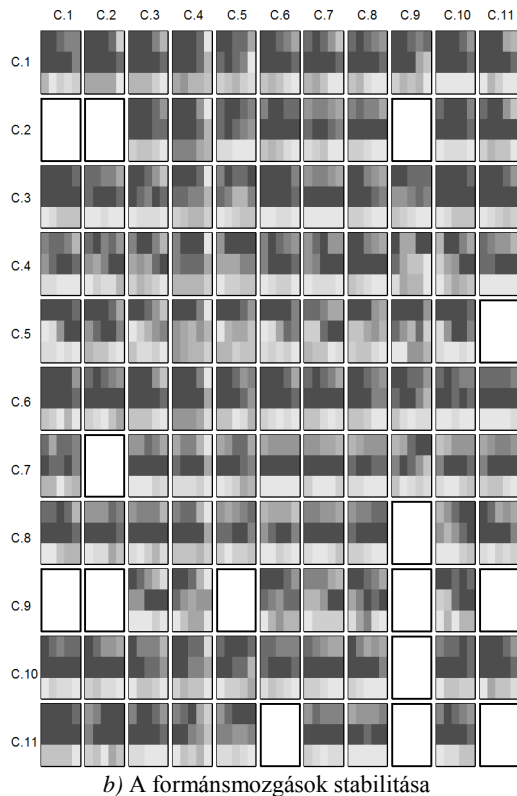
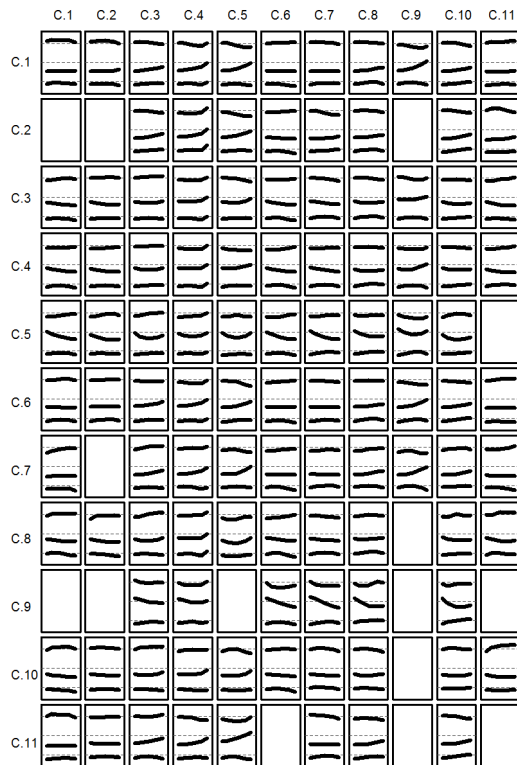
c) A hangkörnyezet magyarázó hatása

F <sub>1</sub>	1-SZ (92%)
F <sub>2</sub>	1-SZ (46%), 9-CS (25%), 7-FU (18%), 8-N (11%)
F <sub>3</sub>	9-CS (41%), 7-FU (22%), 8-N (21%), 1-SZ (9%)

	SM <sub>2</sub> átlag	Összevont SM <sub>2</sub> átlag
F <sub>1</sub>	0,58	0,78
F <sub>2</sub>	0,51	0,74
F <sub>3</sub>	0,39	0,60

d) A jellemző formánsmenet-típusok és a cellák szigorú stabilitási mutatóinak átlaga

6. tábló. [o, o:]

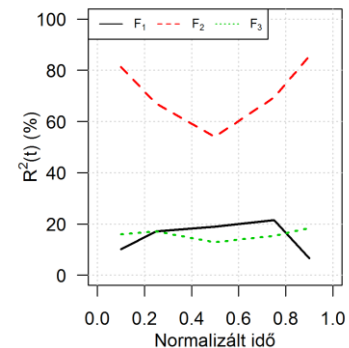
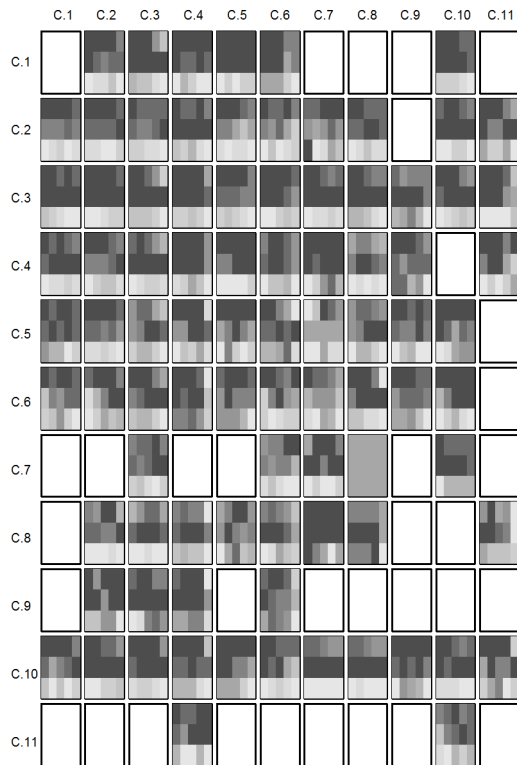
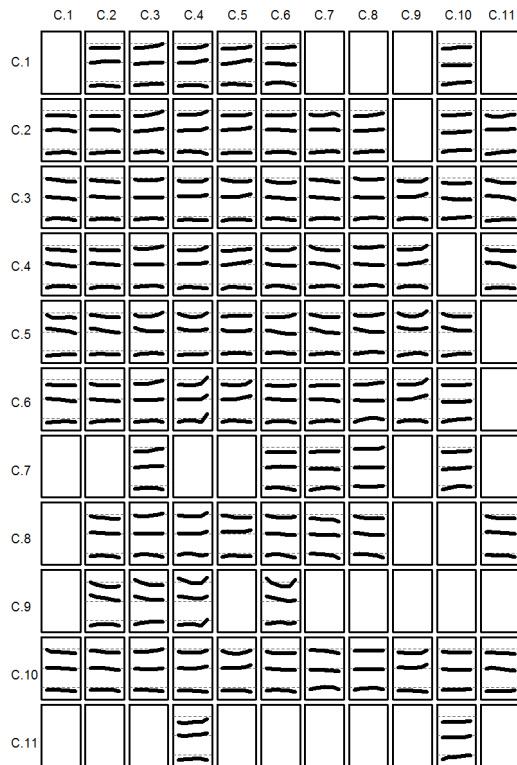


F <sub>1</sub>	1-SZ (42%), 7-FU (31%), 3-SZCS (15%), 8-N (9%)
F <sub>2</sub>	1-SZ (33%), 8-N (30%), 9-CS (25%), 6-U (9%)
F <sub>3</sub>	9-CS (24%), 1-SZ (23%), 8-N (21%), 7-FU (12%), 6-U (9%), 2-SZN (6%)

	SM <sub>2</sub> átlag	Összevont SM <sub>2</sub> átlag
F <sub>1</sub>	0,39	0,62
F <sub>2</sub>	0,68	0,84
F <sub>3</sub>	0,29	0,49

d) A jellemző formánsmenet-típusok és a cellák szigorú stabilitási mutatóinak átlaga

7. tábló. [ø, ø:]



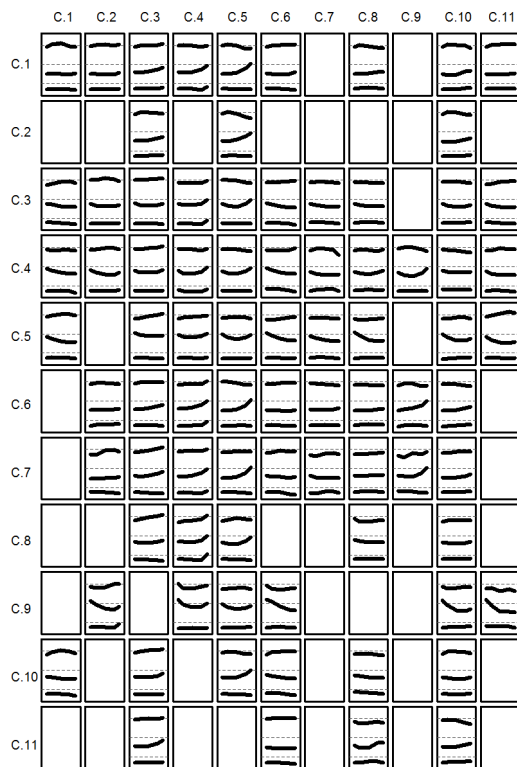
F <sub>1</sub>	1-SZ (56%), 7-FU (29%), 3-SZCS (7%), 8-N (7%)
F <sub>2</sub>	1-SZ (47%), 9-CS (29%), 8-N (11%), 2-SZN (8%), 6-U (5%)
F <sub>3</sub>	1-SZ (43%), 6-U (26%), 9-CS (15%), 8-N (14%)

	SM <sub>2</sub> átlag	Összevont SM <sub>2</sub> átlag
F <sub>1</sub>	0,49	0,76
F <sub>2</sub>	0,54	0,77
F <sub>3</sub>	0,40	0,62

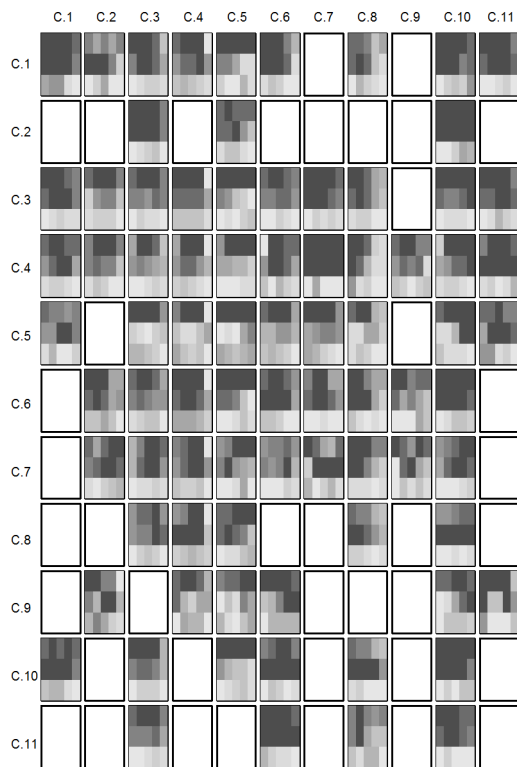
d) A jellemző formánsmenet-típusok és a cellák szigorú stabilitási mutatóinak átlaga



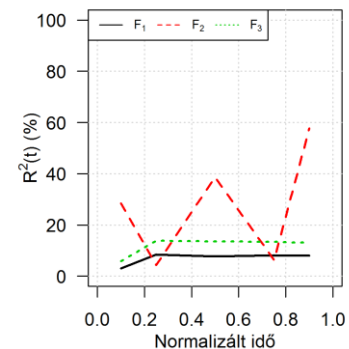
8. tabló. [u, u:]



a) A formánstérképek



b) A formánsmozgások stabilitása



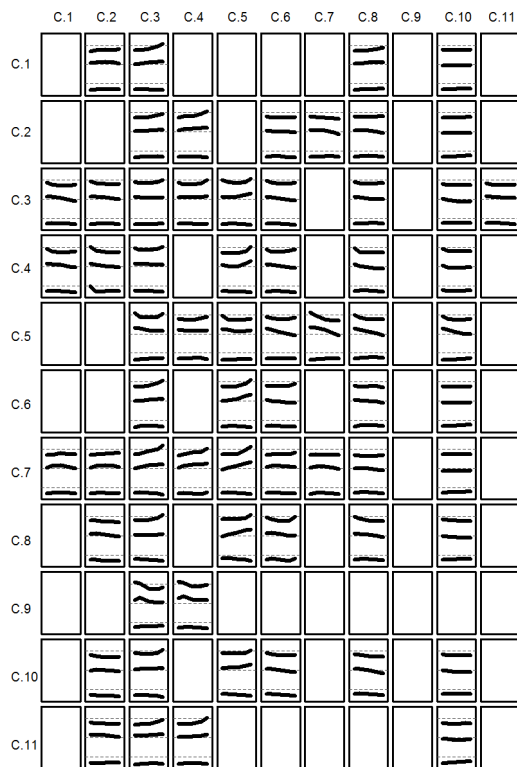
c) A hangkörnyezet magyarázó hatása

F <sub>1</sub>	1-SZ (91%), 9-CS (7%)
F <sub>2</sub>	6-U (28%), 8-N (26%), 9-CS (22%), 1-SZ (19%)
F <sub>3</sub>	8-N (31%), 7-FU (23%), 9-CS (15%), 1-SZ (11%), 6-U (10%), 3-SZCS (6%)

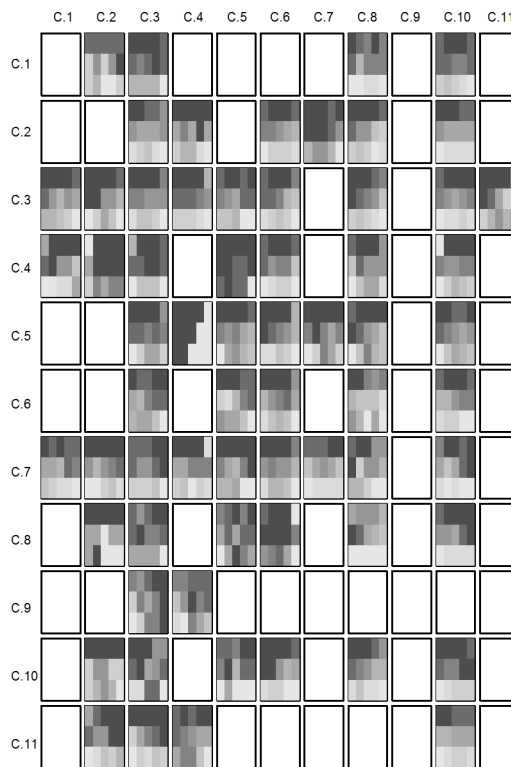
	SM <sub>2</sub> átlag	Összevont SM <sub>2</sub> átlag
F <sub>1</sub>	0,54	0,73
F <sub>2</sub>	0,55	0,75
F <sub>3</sub>	0,26	0,46

d) A jellemző formánsmenet-típusok és a cellák szigorú stabilitási mutatóinak átlaga

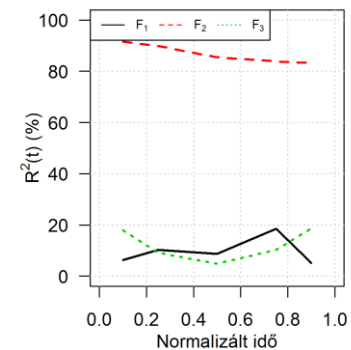
9. tabló. [y, y:]



a) A formánstérképek



b) A formánsmozgások stabilitása



c) A hangkörnyezet magyarázó hatása

F <sub>1</sub>	1-SZ (95%)
F <sub>2</sub>	1-SZ (37%), 9-CS (35%), 8-N (12%), 7-FU (8%), 3-SZCS (5%)
F <sub>3</sub>	1-SZ (37%), 6-U (23%), 8-N (15%), 9- CS (15%), 2-SZN (9%)

	SM <sub>2</sub> átlag	Összevont SM <sub>2</sub> átlag
F <sub>1</sub>	0,63	0,85
F <sub>2</sub>	0,50	0,77
F <sub>3</sub>	0,41	0,65

d) A jellemző formánsmenet-típusok és a cellák szigorú stabilitási mutatóinak átlaga

## IRODALOMJEGYZÉK

- [Abari(2012)] Abari, K. (2012). Reproducible research in speech sciences. *International Journal of Computer Science Issues*, 9(6), 43–52.
- [Abari & Olaszgy(2007)] Abari, K., & Olaszgy, G. (2007). A magyar beszéd hangkapcsolódásainak bemutatása az interneten. *Beszédkutató*, (pp. 178–186).
- [Abari & Olaszgy(2011a)] Abari, K., & Olaszgy, G. (2011). ifem: Interaktív Formánsérték Szerkesztő. <http://magyarbeszed.tmit.bme.hu/ifem>. Letöltve: 2013. május 19.
- [Abari & Olaszgy(2011b)] Abari, K., & Olaszgy, G. (2011). Interaktív formánsérték-módosító fejlesztése. In A. Tanács, & V. Vincze (Eds.) *Magyar Számítógépes Nyelvészeti Konferencia*, (pp. 309–315). Szeged: Szegedi Tudományegyetem Informatikai Tanszékcsoport.
- [Abari & Olaszgy(2011c)] Abari, K., & Olaszgy, G. (2011). Magyar formánsadatbázis az interneten. In M. Gósy (Ed.) *Beszédkutató*, (pp. 73–82). Budapest: MTA Nyelvtudományi Intézet.
- [Abari & Olaszgy(2012)] Abari, K., & Olaszgy, G. (2012). A formánsmenetek rendszere CVC kapcsolatok magánhangzóiban a C képzési helyének függvényében. In M. Gósy (Ed.) *Beszédkutató*, (pp. 70–93). Budapest: MTA Nyelvtudományi Intézet.
- [Abari et al.(2011)] Abari, K., Olaszgy, G., & Rácz, Zs. Zs. (2011). Első magyar formáns-adatbázis. <http://magyarbeszed.tmit.bme.hu/formans>. Letöltve: 2013. május 19.
- [Abari & Papp(2008)] Abari, K., & Papp, Z. (2008). Beszédtechnológiai alapismeretek az oktatásban. In A. Pethő, & M. Herdon (Eds.) *Informatika a felsőoktatásban 2008*. Debrecen.
- [Aronszajn(1950)] Aronszajn, N. (1950). Theory of reproducing kernels. *Transactions of the American Mathematical Society*, 68(3), 337–404.
- [Baker(2006a)] Baker, A. (2006). Quantifying diphthongs: A statistical technique for distinguishing formant contours. Paper presented at NWAV 35, Columbus, OH. <http://www.ling.ohio-state.edu/NWAV/Abstracts/Papr131.pdf>
- [Baker(2006b)] Baker, A. (2006). ssanova. Letöltve: 2013. június 20. <http://www.u.arizona.edu/~tabaker/ssanova/index.html>

- [van Bergem(1993)] van Bergem, D. R. (1993). Acoustic vowel reduction as a function of sentence accent, word stress, and word class. *Speech Communication*, (12), 1–23.
- [Boersma & Weenink(2012)] Boersma, P., & Weenink, D. (2012). Praat: doing phonetics by computer (version 5.3.19. <http://www.praat.org/>, computer program).
- [Bolla(1978)] Bolla, K. (1978). *A magyar magánhangzók akusztikai analízise és szintézise.*, vol. 1 of *Magyar Fonetikai Füzetek*. Budapest: MTA Nyelvtud. Int.
- [Bolla(1995)] Bolla, K. (1995). *Magyar fonetikai atlasz*. Budapest: Nemzeti Tankönyvkiadó.
- [Bombien(2012)] Bombien, L. (2012). libassp: Advanced Speech Signal Processor. <http://libassp.sourceforge.net/>. Letöltve: 2013. május 19.
- [Bombien et al.(2006)] Bombien, L., Cassidy, S., Harrington, J., John, T., & Palethorpe, S. (2006). Recent Developments in the Emu Speech Database System. In P. Warren, & C. I. Watson (Eds.) *Proceedings of the 11th Australasian International Conference on Speech Science and Technology*, (pp. 313–316). Auckland, New Zealand: Australian Speech Science & Technology Association.
- [Bombien et al.(2012)] Bombien, L., Cassidy, S., Harrington, J., John, T., & Palethorpe, S. (2012). The Emu Speech Database System. <http://-emu.sourceforge.net/>. Letöltve: 2013. május 19.
- [de Boor(1986)] de Boor, C. (1986). *B(asic)-Spline Basics*. Defense Technical Information Center.
- [Böhm & Németh(2006)] Böhm, T., & Németh, G. (2006). Algoritmus formánsok követésére, módosítására és szintézisére. *Híradástechnika*, 8, 11–16.
- [Broad & Clermont(2010)] Broad, D. J., & Clermont, F. (2010). Target-locus scaling methods for modeling families of formant transitions. *Journal of Phonetics*, 38(3), 337–359.
- [Buckheit & Donoho(1995)] Buckheit, J. B., & Donoho, D. L. (1995). WaveLab and reproducible research. [http://www-stat.stanford.edu/wavelab/-Wavelab\\_850/wavelab.pdf](http://www-stat.stanford.edu/wavelab/-Wavelab_850/wavelab.pdf). Letöltve: 2013. június 20.
- [Byrne & Foulkes(2004)] Byrne, C., & Foulkes, P. (2004). The 'mobile phone effect' on vowel formants. *Forensic Linguistics*, 11, 83–102.

- [Chambers et al.(1983)] Chambers, J. M., Cleveland, W. S., Kleiner, B., & Tukey, P. A. (1983). *Graphical Methods for Data Analysis*. New York: Chapman and Hall.
- [Chanethom(2011)] Chanethom, V. (2011). Dynamic differences in child bilinguals' production of diphthongs. In A. Botinis (Ed.) *Proceedings of the fourth ISCA Tutorial and Research Workshop on Experimental Linguistics*, (pp. 55–58). 25-27 May 2011, Paris, France.
- [Claerbout(2005)] Claerbout, J. (2005). Making Research Reproducible. <http://sepwww.stanford.edu/OLDWWW/research/redoc/IRIS.html>. Letöltve: 2013. május 19.
- [Claerbout & Karrenbach(1992)] Claerbout, J., & Karrenbach, M. (1992). Electronic documents give reproducible research a new meaning. In *Proc. 62nd Ann. Int. Meeting of the Soc. of Exploration Geophysics*, (pp. 601–604).
- [Craven & Wahba(1979)] Craven, P., & Wahba, G. (1979). Smoothing noisy data with spline functions. estimating the correct degree of smoothing by the method of generalized cross-validation. *Numerische Mathematik*, 31, 377–403.
- [Csapó & Németh(2009)] Csapó, T. G., & Németh, G. (2009). Mássalhangzó–magánhangzó kapcsolatok automatikus osztályozása szubglottális rezonanciák alapján. In *Magyar Számítógépes Nyelvészeti Konferencia*, (pp. 226–237.). Szeged.
- [Davidson(2006)] Davidson, L. (2006). Comparing tongue shapes from ultrasound imaging using smoothing spline analysis of variance. *Journal of the Acoustical Society of America*, 120(1), 407–415.
- [Denes & Pinson(1993)] Denes, P. B., & Pinson, E. N. (1993). *The speech chain: the physics and biology of spoken language*. New York: Worth Publishers.
- [Deng et al.(2006a)] Deng, L., Cui, X., Pruvencok, R., Huang, J., Momen, S., Chen, Y., & Alwan, A. (2006). A database of vocal tract resonance trajectories for research in speech processing. In *Proc. ICASSP*, (pp. 60–63).
- [Deng et al.(2006b)] Deng, L., Cui, X., Pruvencok, R., Huang, J., Momen, S., Chen, Y., & Alwan, A. (2006). Speech Downloads, VTR-Formant Data. <http://research.microsoft.com/en-us/groups/srg/downloads.aspx>. Letöltve: 2013. május 19.

- [Deng et al.(2004)] Deng, L., Lee, L. J., Attias, H., & Acero, A. (2004). A structured speech model with continuous hidden dynamics and prediction-residual training for tracking vocal tract resonances. In *International Conference on Acoustics, Speech, and Signal Processing*, vol. 1.
- [Donoho et al.(2008)] Donoho, D., Maleki, A., Rahman, I., Shahram, M., & Stodden, V. (2008). 15 years of reproducible research in computational harmonic analysis. <http://statistics.stanford.edu/ckirby/techreports/GEN/-2008/2008-02.pdf>. Letöltve: 2013. június 20.
- [Draxler(1995)] Draxler, C. (1995). Introduction to the verbmobil-phondat database of spoken german. In *Proc.of the Third International Conference on the Practical Application of Prolog*, (pp. 201–212). Paris.
- [Duckworth et al.(2011)] Duckworth, M., McDougall, K., DeJong, G., & Shockey, L. (2011). Improving the consistency of formant measurement. *International Journal of Speech Language and the Law*, 18(1), 35–51.
- [Escudero et al.(2009)] Escudero, P., Boersma, P., Rauber, A. S., & Bion, R. A. H. (2009). A cross-dialect acoustic description of vowels: Brazilian and european portuguese. *Journal of the Acoustical Society of America*, 126(3), 1379–1393.
- [Essner(1947)] Essner, C. (1947). Recherche sur la structure des voyelles orales. *Archives Néerlandaises de Phonétique Expérimentale*, 20, 40–77.
- [Evanini et al.(2009)] Evanini, K., Isard, S., & Liberman, M. (2009). Automatic formant extraction for sociolinguistic analysis of large corpora. In *INTERSPEECH*, (pp. 1655–1658). ISCA.
- [Eyben et al.(2013)] Eyben, F., Weninger, F., Woellmer, M., & Schuller, B. (2013). openSMILE: The Munich Versatile and Fast Open-Source Audio Feature Extractor. <http://opensmile.sourceforge.net/>. Letöltve: 2013. május 19.
- [Fant(1960)] Fant, G. (1960). *Acoustic theory of speech production*. Mouton De Gruyter.
- [Ferraty & Vieu(2010)] Ferraty, F., & Vieu, P. (2010). *Nonparametric Functional Data Analysis: Theory and Practice*. Springer Science+Business Media.
- [Flynn & Foulkes(2011)] Flynn, N., & Foulkes, P. (2011). Comparing vowel formant normalization methods. *Proceedings of the 17th International Congress of Phonetic Sciences ICPhS 2011*, (August), 683–686.

- [Fomel & Hennenfent(2007)] Fomel, S., & Hennenfent, G. (2007). Reproducible computational experiments using SCons. In *Acoustics, Speech and Signal Processing, 2007.*, vol. 4, (pp. IV–1257–IV–1260). ICASSP.
- [Fruehwald(2010)] Fruehwald, J. (2010). SS ANOVA. Letöltve: 2013. június. 20. [http://www.ling.upenn.edu/~joseff/papers/fruehwald\\_ssanova.pdf](http://www.ling.upenn.edu/~joseff/papers/fruehwald_ssanova.pdf)
- [Fulop(2011)] Fulop, S. A. (2011). *Speech Spectrum Analysis*. Springer Berlin Heidelberg.
- [Gentleman & Temple Lang(2004)] Gentleman, R., & Temple Lang, D. (2004). Statistical analyses and reproducible research. Tech. rep., Bioconductor Project.
- [Gibbon et al.(1997)] Gibbon, D., Moore, R., & Winski, R. (1997). *Handbook of Standards and Resources for Spoken Language Systems*. Mouton De Gruyter.
- [Gordon(1999)] Gordon, A. D. (1999). *Classification. Second Edition.*. London: Chapman and Hall / CRC.
- [Gordos & Takács(1983)] Gordos, G., & Takács, G. (1983). *Digitális beszédfeldolgozás*. Budapest: Műszaki Könyvkiadó.
- [Gósy(1995)] Gósy, M. (1995). Szükséges és szükségtelen hangátmenetek. In M. Gósy (Ed.) *Beszédkutatás*, (p. 20–32). Budapest: MTA Nyelvtudományi Intézet.
- [Gósy(2004)] Gósy, M. (2004). *Fonetika, a beszéd tudománya*. Budapest: Osiris Kiadó.
- [Gósy(2010)] Gósy, M. (2010). Magánhangzók diftongizálódása. In M. Gósy (Ed.) *Beszédkutatás*, (pp. 17–41). Budapest: MTA Nyelvtudományi Intézet.
- [Gósy(2012)] Gósy, M. (2012). *Az alsóbb nyelvválású magyar magánhangzók formánsszerkezete*, (pp. 43–66). Budapest: Akadémiai Kiadó.
- [Gósy & Olasz(1985)] Gósy, M., & Olasz, G. (1985). A magyar kísérleti fonetika első évtizedei. In *Nyelvtudományi Közlemények*, vol. 87, (pp. 109–121).
- [Grácz & Horváth(2010)] Grácz, T. E., & Horváth, V. (2010). A magánhangzók realizációja spontán beszédben. In M. Gósy (Ed.) *Beszédkutatás*, (pp. 5–16). Budapest: MTA Nyelvtudományi Intézet.

- [Grimaldi et al.(2010)] Grimaldi, M., Calabrese, A., Sigona, F., Garrapa, L., & Sisinni, B. (2010). Articulatory grounding of southern salentino harmony processes. In *INTERSPEECH*, (pp. 1561–1564).
- [Gu(2002)] Gu, C. (2002). *Smoothing Spline ANOVA Models*. Springer.
- [Gu(2012)] Gu, C. (2012). *gss: General Smoothing Splines*. R package version 2.0-11.
- [Gubian(2010)] Gubian, M. (2010). Functional data analysis as a tool for phonetic detail analysis. Tech. rep., Centre for Language and Speech Technology, Radboud University Nijmegen, the Netherlands.
- [Gubian(2011)] Gubian, M. (2011). Functional data analysis for phonetics research. In *Workshop on New Tools and Methods for Very-Large-Scale Phonetics Research*.
- [Gubian(2012)] Gubian, M. (2012). Functional Data Analysis for Speech Research. <http://lands.let.ru.nl/FDA/index.htm>. Letöltve: 2013. június 20.
- [Gubian et al.(2009)] Gubian, M., Torreira, F., Strik, H., & Boves, L. (2009). Functional data analysis as a tool for analyzing speech dynamics - a case study on the French word *c'était*. In *INTERSPEECH*, (pp. 2199–2202). ISCA.
- [Halassy(2002)] Halassy, B. (2002). *Adatmodellezés*. Budapest: Nemzeti Tankönyvkiadó.
- [Harrington(2010a)] Harrington, J. (2010). *Acoustic Phonetics*, chap. 3, (pp. 81–129). Blackwell Publishing Ltd.
- [Harrington(2010b)] Harrington, J. (2010). *The Phonetic Analysis of Speech Corpora*. Wiley-Blackwell. Letöltve: 2013. június 20. <http://www.phonetik.uni-muenchen.de/jmh/research/pasc010808/pasc.pdf>
- [Harrison(2004)] Harrison, P. (2004). *Variability of formant measurements*. Master's thesis, University of York, England.
- [Hillenbrand et al.(1995)] Hillenbrand, J., Getty, L. A., Clark, M. J., & Wheeler, K. (1995). Acoustic characteristics of American English vowels. *Journal of the Acoustical Society of America*, 97(5), 3099–3111.
- [Hillenbrand(1995)] Hillenbrand, J. M. (1995). Vowel database. <http://homepages.wmich.edu/hillenbr/voweldata.html>. Letöltve: 2013. május 19.



- [Hillenbrand et al.(2001)] Hillenbrand, J. M., Clark, M. J., & Nearey, T. M. (2001). Effects of consonant environment on vowel formant patterns. *Journal of The Acoustical Society of America*, 109.
- [Hillenbrand & Nearey(1999)] Hillenbrand, J. M., & Nearey, T. M. (1999). Identification of resynthesized /hVd/ utterances: effects of formant contour. *Journal of Acoustical Society of America*, 105(6), 3509–23.
- [Hollander & Wolfe(1973)] Hollander, M., & Wolfe, D. A. (1973). *Nonparametric Statistical Methods*. New York: John Wiley & Sons.
- [Horváth(2005)] Horváth, V. (2005). A magánhangzók nazalizációjáról. In M. Gósy (Ed.) *Beszédkutatás*, (pp. 73–82). Budapest: MTA Nyelvtudományi Intézet.
- [Huang(1992)] Huang, C. B. (1992). *Speech Perception, Production and Linguistic Structure*, chap. Modelling Human Vowel Identification Using Aspects of Formant Trajectory and Context, (pp. 43–61). Amsterdam: IOS press.
- [Huckvale(2013)] Huckvale, M. (2013). Speech Filing System. <http://www.phon.ucl.ac.uk/resource/sfs/>. Letöltve: 2013. május 19.
- [Jenkins et al.(1983)] Jenkins, J. J., Strange, W., & Edman, T. R. (1983). Identification of vowels in “vowelless” syllables. *Perception & Psychophysics*, 34, 441–450.
- [Johnson & Johnson(1997)] Johnson, A., & Johnson, B. (1997). Literate programming using noweb. *Linux Journal*, 1997(42), 64–69.
- [Joos(1948)] Joos, M. (1948). *Acoustic Phonetics*. No. 24 in Language Monograph. Linguistic Soc. of America.
- [Kent & Read(2002)] Kent, R. D., & Read, C. (2002). *The acoustic analysis of speech*. Singular/Thomson Learning.
- [Knuth(1984)] Knuth, D. E. (1984). Literate programming. *Computer Journal*, 27(2), 97–111.
- [Knuth & Levy(1994)] Knuth, D. E., & Levy, S. (1994). *The CWEB System of Structured Documentation: Version 3.0*. Boston, MA, USA: Addison-Wesley Longman Publishing Co., Inc., 1st ed.
- [Koenig et al.(1946)] Koenig, W., Dunn, H. K., & Lacy, L. Y. (1946). The sound spectrograph. *Journal of the Acoustical Society of America*, 18, 19–49.

- [Kondofersky(2010)] Kondofersky, I. (2010). *Analysis of Functional Phonetic Data*. Master's thesis, Ludwig-Maximilians-Universität München, Institut für Statistik.
- [Kovács(1998)] Kovács, M. (1998). A spektrális minőség és az időtartam szerepe a magánhangzók percepciójában. In M. Gósy (Ed.) *Beszéd kutatás*, (pp. 35–46). Budapest: MTA Nyelvtudományi Intézet.
- [Krull(1987)] Krull, D. (1987). Second formant locus patterns as a measure of consonant-vowel coarticulation. *Phonetic Experimental Research at the Institute of Linguistics University of Stockholm Sweden*, V, 43–61.
- [Kuhn(2012)] Kuhn, M. (2012). odfWeave: Sweave processing of Open Document Format (ODF) files. <http://cran.r-project.org/web/packages/odfWeave/index.html>. Letöltve: 2013.május 19.
- [Lecoutre(2013)] Lecoutre, E. (2013). R2HTML: HTML exportation for R objects. <http://cran.r-project.org/web/packages/R2HTML/index.html>. Letöltve: 2013. május 19.
- [Leisch(2002)] Leisch, F. (2002). Sweave: Dynamic generation of statistical reports using literate data analysis. In W. Hardle, & B. Rönz (Eds.) *Compstat 2002 – Proceedings in Computational Statistics*, (pp. 575–580). Physica Verlag, Heidelberg. ISBN 3-7908-1517-9.
- [Lenth(2010)] Lenth, R. V. (2010). StatWeave – Beta version. <http://homepage.cs.uiowa.edu/rlenth/StatWeave/>. Letöltve: 2013. május 19.
- [Lenth & Hjosgaard(2007)] Lenth, R. V., & Hjosgaard, S. (2007). SASweave: Literate programming using SAS. <http://homepage.cs.uiowa.edu/rlenth/SASweave/>. Letöltve: 2013. május 19.
- [Ligges(2013a)] Ligges, U. (2013). signal: Signal processing. <http://cran.r-project.org/web/packages/signal/index.html>. Letöltve: 2013. május 19.
- [Ligges(2013b)] Ligges, U. (2013). tuneR: Analysis of music and speech. <http://cran.r-project.org/web/packages/tuneR/index.html>. Letöltve: 2013. május 19.
- [Lilienthal(2009)] Lilienthal, J. (2009). The articulatory and acoustic impact of scottish english /r/ on the preceding vowel-onset. In *INTERSPEECH*, (pp. 2819–2822).
- [Lindblom(1963)] Lindblom, B. (1963). Spectrographic study of vowel reduction. *Journal of the Acoustical Society of America*, 35, 1773–1781.

- [Lobanov(1971)] Lobanov, B. M. (1971). Classification of russian vowels spoken by different speakers. *Journal of the Acoustical Society of America*, (49), 606–608.
- [Magdics(1965)] Magdics, K. (1965). *A magyar beszédhangok akusztikai szerkezete*. Budapest: Akadémiai Kiadó.
- [Markel & Gray(1976)] Markel, J. D., & Gray, A. H. (1976). *Linear prediction of speech*. Berlin: New York: Springer-Verlag.
- [MathSoft(2013)] MathSoft (2013). S-PLUS 2000. <http://s-plus-2000.software.informer.com/>. Letöltve: 2013. május 19.
- [Mihajlik et al.(2002)] Mihajlik, P., Révész, T., & Tatai, P. (2002). Phonetic transcription in automatic speech recognition. *Acta Linguistica Hungarica*, 49(3–4), 407–425.
- [Mokhtari & Tanaka(2000a)] Mokhtari, P., & Tanaka, K. (2000). A corpus of Japanese vowel formant patterns. <http://www2.nict.go.jp/univ-com/mcc/-people/parham/formant/formant.html>. Letöltve: 2013. május 19.
- [Mokhtari & Tanaka(2000b)] Mokhtari, P., & Tanaka, K. (2000). A corpus of japanese vowel formant patterns. *Bulletin of Electrotechnical Laboratory*, 64, 57–66.
- [Molnár(1970)] Molnár, J. (1970). *A magyar beszédhangok atlasza*. Budapest: Tankönyvkiadó.
- [Montgomery(2006)] Montgomery, D. C. (2006). *Design and Analysis of Experiments*. John Wiley & Sons.
- [Morrison & Assmann(2013)] Morrison, G. S., & Assmann, P. F. (2013). *Vowel Inherent Spectral Change*. Springer.
- [Nearey & Assmann(1986)] Nearey, T. M., & Assmann, P. F. (1986). Modeling the role of inherent spectral change in vowel identification. *The Journal of the Acoustical Society of America*, 80(5), 1297–1308.
- [Németh & Olasz(2010)] Németh, G., & Olasz, G. (2010). *A magyar beszéd: Beszédkutatás, beszédtechnológia, beszédinformációs rendszerek*. Budapest: Akadémiai Kiadó.
- [Neumaier(2010)] Neumaier, A. (2010). Statistical data sets. <http://solon.cma.univie.ac.at/statdat.html>. Letöltve: 2013. május 19.
- [Nikléczy(2006)] Nikléczy, P. (2006). A koartikuláció a magánhangzók tiszta fázisa és az átmenetek összefüggésében. In M. Gósy (Ed.) *Beszédkutatás*, (pp. 23–31). Budapest: MTA Nyelvtudományi Intézet.

- [Nycz & Decker(2006)] Nycz, J., & Decker, P. D. (2006). A New Way of Analyzing Vowels: Comparing Formant Contours Using Smoothing Spline ANOVA. Poster. NAWAV 35, Columbus, OH.
- [Olaszy(1985)] Olaszy, G. (1985). *A magyar beszéd leggyakoribb hangsorépítő elemeinek szerkezete és szintézise*. Nyelvtudományi Értekezések. Budapest: Akadémiai Kiadó.
- [Olaszy(2013)] Olaszy, G. (2013). Precíziós, párhuzamos, magyar beszédatbázis fejlesztése és szolgáltatásai. In M. Gósy (Ed.) *Beszédkutatás*, (pp. 261–270). Budapest: MTA Nyelvtudományi Intézet.
- [Olaszy & Abari(2011)] Olaszy, G., & Abari, K. (2011). Beszédminták adatbázisa a magyar beszédhang-kapcsolódások akusztikai szerkezetének bemutatására. <http://magyarbeszed.tmit.bme.hu/cvvc>. Letöltve: 2013. május 19.
- [Olaszy et al.(2000)] Olaszy, G., Németh, G., Olaszi, P., & Kiss, G. (2000). PROFIVOX - a legkorszerűbb hazai beszédszintetizátor és szüvegfelolvasó. In M. Gósy (Ed.) *Beszédkutatás*, (pp. 167–179). Budapest: MTA Nyelvtudományi Intézet.
- [Olaszy et al.(2009)] Olaszy, G., Rácz, Zs. Zs., & Bartalis, M. (2009). Formánsmérések automatizálása, formánsadatbázisok létrehozása. In M. Gósy (Ed.) *Beszédkutatás*, (pp. 134–147). Budapest: MTA Nyelvtudományi Intézet.
- [Parzen(1959)] Parzen, E. (1959). *Statistical Inference on Time Series by Hilbert Space Methods I*. Tech. rep., Department of Statistics, Stanford University, CA.
- [Peng(2011)] Peng, R. D. (2011). Reproducible Research in Computational Science. *Science*, 334(6060), 1226–1227.
- [Peterson & Barney(1952)] Peterson, G. E., & Barney, H. L. (1952). Control methods used in a study of the vowels. *Journal of the Acoustical Society of America*, 24, 175–184.
- [Pritchard(2012)] Pritchard, S. (2012). *A Cross-Language Study of the Production and Perception of Palatalized Consonants*. Ph.D. thesis, Department of Linguistics, Faculty of Arts, University of Ottawa, Canada.
- [R Core Team(2012)] R Core Team (2012). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0. <http://www.R-project.org/>

- [Ramsay(2005)] Ramsay, J. O. (2005). *Encyclopedia of Statistics in Behavioral Science*, vol. 2, chap. Functional Data Analysis, (p. 675–678). Chichester: John Wiley & Sons.
- [Ramsay & Dalzell(1991)] Ramsay, J. O., & Dalzell, C. J. (1991). Some Tools for Functional Data Analysis (with discussion). *Journal of the Royal Statistical Society.*, 53(3), 539–572.
- [Ramsay et al.(2009)] Ramsay, J. O., Hooker, G., & Graves, S. (2009). *Functional Data Analysis with R and MATLAB*. Use R! Springer.
- [Ramsay & Silverman(2002)] Ramsay, J. O., & Silverman, B. W. (2002). *Applied Functional Data Analysis: Methods and Case Studies*. Springer Series in Statistics. Springer London, Limited.
- [Ramsay & Silverman(2005)] Ramsay, J. O., & Silverman, B. W. (2005). *Functional Data Analysis*. Springer Series in Statistics. Springer.
- [Ramsay et al.(2012)] Ramsay, J. O., Wickham, H., Graves, S., & Hooker, G. (2012). *fda: Functional Data Analysis*. R package version 2.2.8.
- [Ren(2005)] Ren, H. (2005). *Functional Inverse Regression and Reproducing Kernel Hilbert Space*. Ph.D. thesis, Texas A&M University.
- [Ritter(2012)] Ritter, C. (2012). R2wd: Write MS-Word documents from R. <http://cran.r-project.org/web/packages/R2wd/index.html>. Letöltve: 2013. május 19.
- [RRplanet.com(2013)] RRplanet.com (2013). Understand Reproducible Research. <http://www.rrplanet.com/reproducible-research/reproducible-research.shtml>. Letöltve: 2013. május 19.
- [Schulte et al.(2012)] Schulte, E., Davison, D., Dye, T., & Dominik, C. (2012). A multi-language computing environment for literate programming and reproducible research. *Journal of Statistical Software*, 46(3), 1–24.
- [Schwab et al.(1997)] Schwab, M., Karrenbach, M., & Claerbout, J. (1997). Making scientific computations reproducible. In *Computing in Science & Engineering*, vol. 2, (pp. 61–67).
- [Siptár(2001)] Siptár, P. (2001). *Strukturális Magyar Nyelvtan. 2. kötet. Fonológia.*, chap. A mássalhangzók, (pp. 183–272). Budapest: Akadémia.
- [Sjölander(2006)] Sjölander, K. (2006). The Snack Sound Toolkit . <http://www.speech.kth.se/snack/>. Letöltve: 2013. május 19.
- [Stevens et al.(1966)] Stevens, K., House, A., & Paul, A. (1966). Acoustical description of syllabic nuclei: an interpretation in terms of a dynamic

- model of articulation. *Journal of the Acoustical Society of America*, 40, 123–132.
- [Stevens & House(1961)] Stevens, K. N., & House, A. S. (1961). An acoustical theory of vowel production and some of its implications. *Journal of Speech and Hearing Research*, 4(4), 303–320.
- [Strange & Jenkins(2013)] Strange, W., & Jenkins, J. J. (2013). Dynamic specification of coarticulated vowels. In G. S. Morrison, & P. F. Assmann (Eds.) *Vowel Inherent Spectral Change*, (pp. 87–115). Springer Berlin Heidelberg.
- [Strange et al.(1976)] Strange, W., Verbrugge, R. R., Shankweiler, D. P., & Edman, T. R. (1976). Consonant environment specifies vowel identity. *Journal of The Acoustical Society of America*, 60, 213–224.
- [Subosits(2005)] Subosits, I. (2005). *Hangtan*. Budapest: TAS-11 Kft.
- [Sueur et al.(2013)] Sueur, J., Aubin, T., & Simonis, C. (2013). seewave: Time wave analysis and graphical representation. <http://cran.r-project.org/web/packages/seewave/index.html>. Letöltve: 2013. május 19.
- [Suomi(1987)] Suomi, K. (1987). On spectral coarticulation in stop-vowel-stop syllables: Implications for automatic speech recognition. *Journal of Phonetics*, 15(1), 85–100.
- [Sussman(1990)] Sussman, H. M. (1990). Acoustic correlates of the front/back vowel distinction: A comparison of transition onset versus “steady state”. *The Journal of the Acoustical Society of America*, 88(1), 87–96.
- [Szalai(1995)] Szalai, E. (1995). Az [u:] [a:] és [i] hangok koartikulációs mezőiről. In M. Gósy (Ed.) *Beszédkutatás*, (pp. 83–93). Budapest: MTA Nyelvtudományi Intézet.
- [Tarnóczy(1941)] Tarnóczy, T. (1941). *A magyar magánhangzók akusztikai szerkezete*. Budapest: Kir. Magy. Pázmány Péter Tudományegyetem Általános Nyelvészeti és Fonetikai Intézete.
- [Vallabha & Tuller(2002)] Vallabha, G. K., & Tuller, B. (2002). Systematic errors in the formant analysis of steady-state vowels. *Speech Communication*, 38(1), 141–160.
- [Vandewalle et al.(2009)] Vandewalle, P., Kovacevic, J., & Vetterli, M. (2009). Reproducible research in signal processing - what, why, and how. *IEEE Signal Processing Magazine*, 26(3), 37–47.

- [Vértés O.(1980)] Vértés O., A. (1980). *A magyar leíró hangtan története az újgrammatikusokig*. Budapest: Akadémiai Kiadó.
- [Vértés O.(1982a)] Vértés O., A. (1982). Az artikuláció akusztikus vetülete. In K. Bolla (Ed.) *Fejezetek a magyar leíró hangtanból*, (pp. 155–164). Budapest: Akadémiai Kiadó.
- [Vértés O.(1982b)] Vértés O., A. (1982). A magyar beszédhangok akusztikai elemzésének kérdései. In K. Bolla (Ed.) *Fejezetek a magyar leíró hangtanból*, (pp. 71–115). Budapest: Akadémiai Kiadó.
- [Vértés O.(1982c)] Vértés O., A. (1982). A magyar leíró hangtan története az újgrammatikusoktól 1945-ig. In K. Bolla (Ed.) *Fejezetek a magyar leíró hangtanból*, (pp. 283–298). Budapest: Akadémiai Kiadó.
- [Vicsi(1996)] Vicsi, K. (1996). SAMPA - computer readable phonetic alphabet - Hungarian. <http://www.phon.ucl.ac.uk/home/sampa/hungaria.htm>. Letöltve: 2013. június 20.
- [Vicsi(2010)] Vicsi, K. (2010). *A magyar beszéd: Beszédkutatás, beszédtechnológia, beszédinformációs rendszerek*, chap. Adatbázisok a beszédtechnológia szolgálatában, (pp. 261–283). Budapest: Akadémiai Kiadó.
- [Vicsi & Víg(1998)] Vicsi, K., & Víg, A. (1998). Az első magyar nyelvű beszédadatbázis. In M. Gósy (Ed.) *Beszédkutatás*, (pp. 163–177). Budapest: MTA Nyelvtudományi Intézet.
- [Vicsi & Víg(2011)] Vicsi, K., & Víg, A. (2011). Hungarian speech databases. <http://alpha.tmit.bme.hu/speech/hdbbabel.php>. Letöltve: 2013. május 19.
- [Wahba(1990)] Wahba, G. (1990). *Spline Models for Observational Data*. Society for Industrial and Applied Mathematics.
- [Wang & Marron(2007)] Wang, H., & Marron, J. S. (2007). Object oriented data analysis: Sets of trees. *Annals of Statistics*, 35(5), 1849–1873.
- [Wang(2010)] Wang, Y. (2010). *Spline Smoothing: Methods and Applications*. Taylor & Francis.
- [Wang & Ke.(2011)] Wang, Y., & Ke., C. (2011). *assist: A Suite of S-Plus Functions Implementing Smoothing Splines*. R package version 3.1.2. <http://CRAN.R-project.org/package=assist>
- [Watson & Harrington(1999)] Watson, C., & Harrington, J. (1999). Acoustic evidence for dynamic formant trajectories in australian english vowels. *Journal of the Acoustical Society of America*, 106(1), 458–68.

- [Welch(1951)] Welch, B. L. (1951). On the comparison of several mean values: an alternative approach. *Biometrika*, 38, 330–336.
- [Xie(2013)] Xie, Y. (2013). knitr: A general-purpose package for dynamic report generation in R. <http://cran.r-project.org/web/packages/knitr/index.html>. Letöltve: 2013. május 19.
- [Young(2006)] Young, S. J. (December, 2006). *The HTK Book (for HTK Version 3.4)*. Cambridge, England: Cambridge University Engineering Department.
- [Yu & Hua(2011)] Yu, C., & Hua, L. (2011). Analysing Tongue Shape and Movement in Vowel Production Using SS ANOVA in Ultrasound Imaging. In L. Wai-Sum, & E. Zee (Eds.) *Proceedings of the 17th International Congress of Phonetic Sciences*, (pp. 124–127). Hong Kong, China.
- [Zahorian & Jagharghi(1993)] Zahorian, S. A., & Jagharghi, A. J. (1993). Spectral-shape features versus formants as acoustic correlates for vowels. *Journal of the Acoustic Society of America*, 94, 1966–1982.
- [Zen et al.(2009)] Zen, H., Tokuda, K., & Black, A. W. (2009). Statistical parametric speech synthesis. *Speech Communication*, 51(11), 1039–1064.