Egyetemi doktori (PhD) értekezés tézisei

A formánsmozgások statisztikai vizsgálata és modellezése a magyar magánhangzókban

Abari Kálmán

Témavezető: Dr. Várterész Magda

DEBRECENI EGYETEM
Matematika- és Számítástudományok Doktori Iskola
Debrecen, 2013

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.

Tartalomjegyzék – Contents

TÉZISEK MAGYARUL

1. Elméleti háttér, célkitűzések

A számítógépek teljesítményének és a háttérkapacitás további növekedésének köszönhetően mára a legtöbb tudományterület rendelkezik számítógépes megfelelőjével, amelyeket olyan nagy számítást igénylő feladatok jellemeznek, mint pl. a mintakeresés nagy adatbázisokban vagy nagyméretű szimulációk futtatása [7], [20]. A beszédtudomány területén is – a fonetikai kutatásokban és a beszédtechnológiai alkalmazásokban egyaránt – megfigyelhető a nagyméretű beszédadatbázisok használata, 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 algoritmusaik számára.

A nagy adatbázisok használata több szempontból is kihívás elé állítja a kutatót. Egyrészt biztosítania kell, hogy állításai ellenőrizhetők legyenek, így szükség van a mérések átláthatóvá és megismételhetővé tételére. Erre a reprodukálható kutatás adja meg a választ [6]. Másrészt új elemző technikákra van szükség az egyre bonyolultabb adatstruktúrák rendezésére, leírására és elemzésére. Az elmúlt két évtizedben megjelent funkcionális és objektum-orientált statisztikai módszerek ennek a kihívásnak felelnek meg [21], [26]. Az előbbi görbék elemzését és összehasonlítását ígéri, utóbbi esetében pedig tetszőlegesen összetett objektumok vizsgálatára nyílik lehetőség.

A fonetika egyik alapfeladata a hangok egymásra hatásának vizsgálata. Korán felismerték, hogy a magánhangzó teljes időtartamában a spektrális tulajdonságok változáson mennek keresztül [15]. Mégis az ezt követő évtizedekben a magánhangzók spektrális leírására főképp statikus modellt használtak. E modell szerint a magánhangzó-minőség teljes leírásához szükséges információ rendelkezésre áll a magánhangzó egy viszonylag állandó (steady-state) részében, a tiszta fázisban. A magánhangzó spektrális leírásának dinamikus modellje szerint azonban a magánhangzó azonosításához legalább annyi információt szerezhetünk a hangátmenetekből, mint a tiszta fázisból [14], [23]. Tanulmányok sora bizonyította, hogy sok esetben nem elegendő egyetlen időszelet spektrális információját feldolgozni, a hangátmenetek ugyanúgy hozzájárulnak a magánhangzó felismeréséhez, mint a tiszta fázis.

A magánhangzók dinamikus spektrális tulajdonságának legfontosabb oka a koartikuláció, amely a szomszédos hangok artikulációs gesztusainak egymásra hatását jelenti [11]. Az értekezésben a koartikuláció spektrális hatását a formánsfrekvencia értékek alapján vizsgálom.

A magánhangzók vagy magánhangzószerű mássalhangzók képzése során a tüdőből kiáramló levegő a hangszalagokat „megrezegteti", létrejön a zönge. Az alaphangból és felhangokból álló zönge a toldalékcsőben az adott képzési konfigurációtól függően módosul. A toldalékcső üregei a rezonanciafrekvenciáiknak megfelelő vagy ahhoz közel eső felharmonikusokat a zöngében felerősítik, így energiakoncentrációk keletkeznek, amelyeket a fonetikai szakirodalom formánsoknak nevez [11], [17], [24]. A formánsfrekvenciák jelölésére az F_1, F_2, …, F_N betűket használjuk.

Értekezésemben egy saját beszédadatbázis és a hozzá kapcsolódó formánsadatbázis létrehozását, valamint az azon végzett mérések és modellalkotások eredményeit mutatom be. A formánsadatbázis magánhangzónként öt mérési pont mindegyékéhez az F_1, F_2 és F_3 adatokat tartalmazza, így alkalmas a koartikulációs hatások

vizsgálatára a hangátmenetekben. Az adatbázis mintegy 5 millió magánhangzó formánsadatot tartalmaz. A magánhangzók dinamikus spektrális tulajdonságainak tanulmányozását a funkcionális adatelemzés módszerével végzem. Bemutatom azt a szoftverkörnyezetet is, amely biztosítja a vizsgálatok reprodukálható jellegét.

Új eredményeimmel hozzájárulok, hogy matematikailag is magyarázhatóvá váljanak a formánsokkal kapcsolatos eddigi fonetikai vizuális megfigyelések és leíró jellegű magyarázatok.

2. Az értekezés új tudományos eredményei

A disszertáció 4 téziscsoportot és összesen 8 tézist tartalmaz.

I. A megismételhető kutatás szempontjainak megfelelő saját beszédadatbázis létrehozásával kapcsolatos téziscsoport

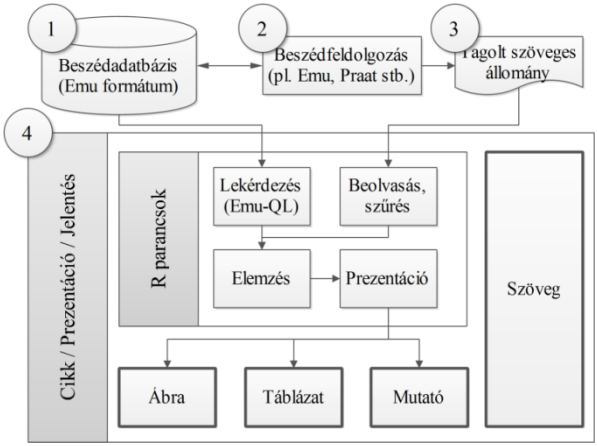
I.1. tézis. *Önmagában ismert eszközök és eljárások újszerű kombinációjával kidolgoztam az [1]-ben részletezett modellt, amely a beszédtudomány néhány területén a reprodukálható kutatás alapja lehet (1. ábra).*

Az általam javasolt modell komponensei: (1) Emu formátumú beszédadatbázis, amely integrált programcsomag annotált beszédadatbázisok létrehozására, lekérdezésére és elemzésére [5]; (2) a beszédjel elemzésére szolgáló szoftver eszközök és csomagok (pl. Praat[1], R `emu`[2] WaveSurfer/Snack[3]), amelyek hangelemzést és jelfeldolgozást végeznek; (3) a beszédjel elemzése során, az Emu

[1]A Praat széles körben használt fonetikai elemző program http://www.fon.hum.uva.nl/praat/
[2]http://cran.r-project.org/web/packages/emu/
[3]http://www.speech.kth.se/snack/

adatbázis struktúrájához nem illeszkedő adatok tabulátorral tagolt szöveges állománya; (4) a futtatható publikáció.

A modell központi része a futtatható cikk (1. ábra: 4), amely 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 használunk. A 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-rel kombinált használatát az [1] publikációban be is mutatom.



1. ábra. A reprodukálható beszédkutatás javasolt modellje

I.2. tézis. *A reprodukálható beszédkutatás fenti irányelveinek érvényesítésével nagyméretű beszédadatbázist hoztam létre, amelyben kialakítottam a formánsadatok tárolásának szerkezeti elveit. Eljárást dolgoztam ki, amellyel a korábbi kutatások beszédadatbázisát az új adatbázis-szerkezetbe konvertáltam.*

A beszédadatbázis alapja a BME TMIT Beszédakusztikai Laboratóriumban létrehozott FKM (Fonetikailag Kiegyensúlyozott

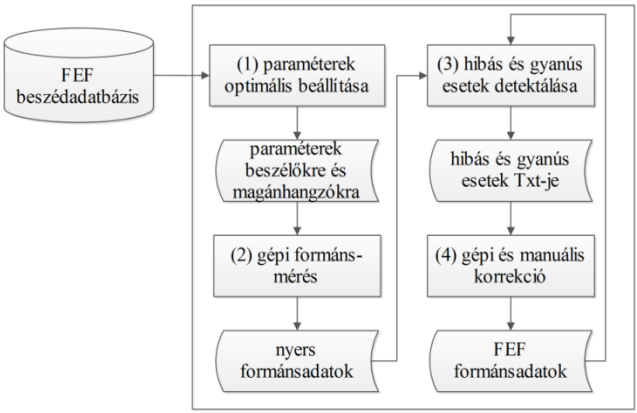
Mondatok) beszédadatbázis[4] [18]. Új eljárást dolgoztam ki, amellyel az FKM forrásadatbázist az 1. ábra szerinti Emu adatbázis-szerkezetbe konvertáltam. A saját FEF (Formánsokkal ellátott FKM) beszédadatbázisom 10 beszélő, 5 férfi és 5 nő bemondásait tartalmazza. A felolvasott szöveg mondatait a BABEL beszédadatbázis[5] szövegállománya adta [25]. A mondatgyűjtemény irodalmi művek gondosan kiválasztott részeit tartalmazza, a teljes lista összesen 1992 db különböző mondatot foglal magában. A FEF beszédadatbázis teljes ideje 1524 perc (25,4 óra), összesen 854 240 db beszédhangot és 332 827 db magánhangzót tartalmaz.

II. Formánsfrekvenciák automatikus meghatározásáról és korrekciójáról szóló téziscsoport

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

Olyan új eljárást dolgoztam ki, amelynek alkalmazásával ellenőrzött formánsadatok rendezetten tárolt együttesét alakíthatjuk ki. (2. ábra). Hipotézisem szerint a kidolgozott többlépcsős formánsjavító eljárás nagy beszédadatbázisok esetén csaknem hibamentes formánsadatbázishoz vezet, azaz nagy részben helyettesítheti a manuális ellenőrzést.

[4] A forrásadatbázis használatához Titoktartási Nyilatkozat kitöltésére volt szükség.
[5] http://alpha.tmit.bme.hu/speech/hdbbabel.php



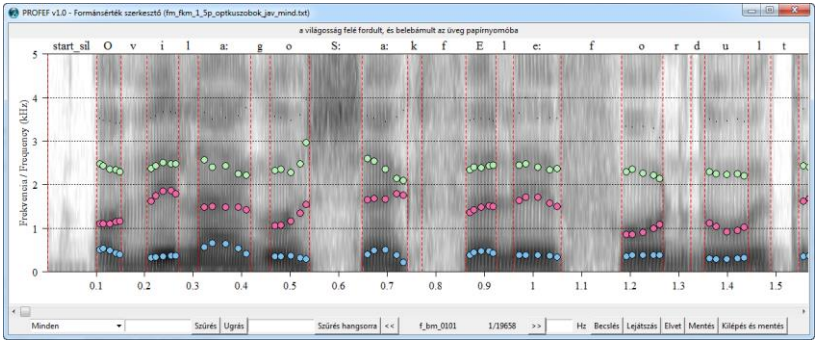
2. ábra. Az ellenőrzött formánsadatok előállításának lépései

A gépi formánsmérések automatikus végrehajtásához saját szkriptet készítettem, amely a FEF adatbázisban tárolt hangfelvételek és annotációk kezelése mellett az algoritmus optimális bemenő paramétereinek kezelését is elvégezte (2). A bemenő paraméterek optimális beállítására (1) a szakirodalomban javasolt egyik módszert implementáltam [8]. Az így kapott nyers formánsadatok lesznek a bemenetei egy saját fejlesztésű, többlépcsős, félautomatikus hibadetektáló és javító (3, 4) algoritmusnak. Korábbi kutatások ([19] és [3]) tapasztalataiból kiindulva tipizáltam a formánsmérések lehetséges tévesztéseit és 5 hibatípus automatikus felismerésére készítettem szabály alapú algoritmust. A hibatípusok: hiányzó adat, formánstávolság hiba, formánssorszám-tévesztés, kiugró adat és hibás formánsmenet. Az egyes hibák felismerése eltérő nehézségű, több közülük segédtáblázatok összeállítását igényli.

II.2. tézis. *Algoritmust és hozzá kapcsolódó vizuális, interaktív alkalmazást fejlesztettem, amellyel gépi és manuális hibajavításra, illetve az automatikusan javított formánsadatok kézi ellenőrzésére van lehetőség.*

A detektált hibák javítására eljárást dolgoztam ki, amelyhez vizuális grafikus felhasználói felülettel rendelkező alkalmazást fejlesztettem

ki (3. ábra). A PROFEF fantázianevű program a korábban létrehozott webes környezetben működő Interaktív Formánsérték Szerkesztő[6] [2] továbbfejlesztett változata. A PROFEF a hibák javítása során a kézi és az automatikus (gépi) módszert is támogatja.



3. ábra. A PROFEF alkalmazás munkaképernyője

Az ellenőrzött formánsadatok tárolására tabulátorral tagolt szöveges formátumú elrendezést dolgoztam ki, melynek felépítése a 1. táblázatban látható.

1. 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

[6] Elérhető: http://magyarbeszed.tmit.bme.hu/ifem

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 az F_1-F_3 formánsfrekvencia értékeket sorolja Hz-ben kifejezve.

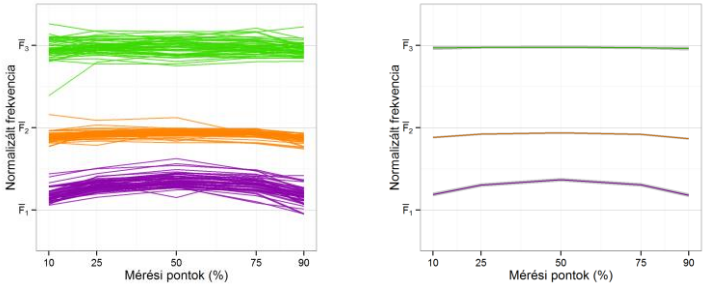
III. A magyar magánhangzók formánsterének modellezésével és vizsgálatával kapcsolatos téziscsoport

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

A beszédadatbázis különböző beszélőihez tartozó formánsértékek egységes kezeléséhez szükség van az adatok normalizálására. 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 [16] választottam, amely könnyen értelmezhető és jól használható a beszélők közötti különbségek eliminálására (pl. fiziológia, anatómiai különbségek), míg a hangsor beszélőfüggetlen része, a fonemikus információ, változatlan marad [9]. A Lobanov-féle módszer a matematikai statisztikában megszokott Z-transzformáción alapul.

A három különböző formánshoz tartozó normalizált formánsgörbék elviekben nem jeleníthetők meg egyetlen ábrán, mert az *y*-tengely normalizált formánsfrekvencia értékeket tartalmaz. 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 4. ábra a) része tartalmazza az egymásra halmozott normalizált formánsgörbéket. Az ábrán szereplő

három vízszintes vonal az egyes formánsátlagoknak felel meg. Az ábra 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 lentről 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 az egyes formánsfrekvencia értékek.



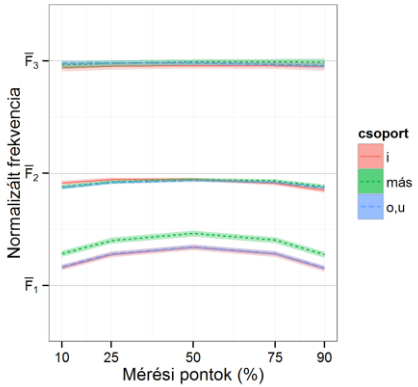
a) ábra *b) ábra*

4. ábra. A normalizált formánsgörbék az a) részben és a helyükbe lépő átlagos görbék a b) részben

Az SSANOVA módszere [12] segítségével a görbéinket egyetlen simított spline-nal helyettesíthetjük, és emellett a görbék szóródására a Bayes konfidencia-intervallum nagyságából következtethetünk. A 4. ábra b) részében megjelenített spline-ok alkalmasak tetszőleges magánhangzó tetszőleges hangkörnyezetbeli formánsmozgásának leírására. Hasonló eredményt kaphatunk a funkcionális varianciaelemzés végrehajtásával is [22].

A különböző hangkörnyezetekben megvalósuló formánsmozgások összehasonlítására szintén alkalmas az SSANOVA modell. Ha a bilabiális mássalhangzók közötti [a:] formánsmozgását vizsgáljuk különböző –2. hangkörnyezetben, akkor a könnyen általánosítható 5. ábrát kapjuk. Az összehasonlítandó hangkörnyezetek függvényében különböző csoportokat definiálunk és a csoportok formánsgörbéinek eltérését teszteljük. Az 5. 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. A görbék körül berajzolt Bayes konfidencia-intervallumok átlapolása mutatja meg, hogy eltérnek-e szignifikánsan az egyes csoportok, és ha eltérnek, akkor a görbe melyik részén. Az 5. 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.

Ezzel megmutattam, hogy a funkcionális adatelemzés módszere alkalmas a formánsmozgások modellezésére és összehasonlítására a normalizált formánstérben. Adaptáltam és továbbfejlesztettem a modell értékelésére eddig használt grafikus ábrázolást [10].

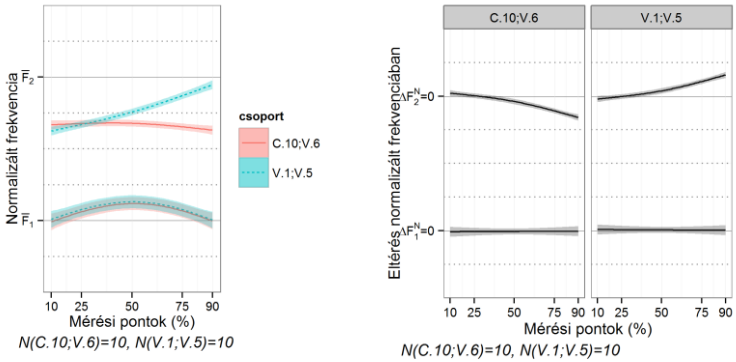


5. ábra. Az SSANOVA grafikus szemléltetése három görbecsoport esetén a három formánsra. Szignifikáns különbség mutatkozik a F_1-ben, ugyanis az [i] és [o, u] egységesen csökkenti az első formáns értékekét a többi –2. pozícióban lévő hanghoz viszonyítva

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

A III.1. tézisben szereplő grafikus ábrázolás segítségével megmutattam, hogy például az [ɔ] realizációk formánsának mozgása nem csak a közvetlenül szomszédos kapcsolódó hang függvénye, hanem függhet a +2. hangtól is. A mért mintákon a formánsmozgás szignifikáns eltérést mutat a két összehasonlított csoportban, amely a +2. hang hatásának köszönhető (6. ábra). A FEF adatbázisban a példa vizsgálatára a [vɔhi] és [vɔho] hangsorokra 10-10 mintát találtam, ez jelenti a két mért csoportot, amelyben csak az utolsó hang különbözik. A 6. ábrán az egyes görbék köré rajzolt 95%-os Bayes konfidencia-intervallumok segítenek eldönteni, hogy az eltérés a görbék mely pontján szignifikáns. A jobb oldali részen a csoport és interakciós hatás együttesen jelenik meg, amely explicitté teszi az eltérést. Az ilyen hatásokra más hangcsoportok esetében is vannak egyedi példák az adatbázisban, de nem elegendő számban

fordulnak elő a statisztikai elemzéshez. Tehát a jelenség további vizsgálatok tárgya lehet a jövőben.



6. ábra. A +2. hang szignifikáns hatásának bemutatása a magánhangzó második formánsára

IV. Új fonetikai mérőeszközök bemutatásáról szóló téziscsoport

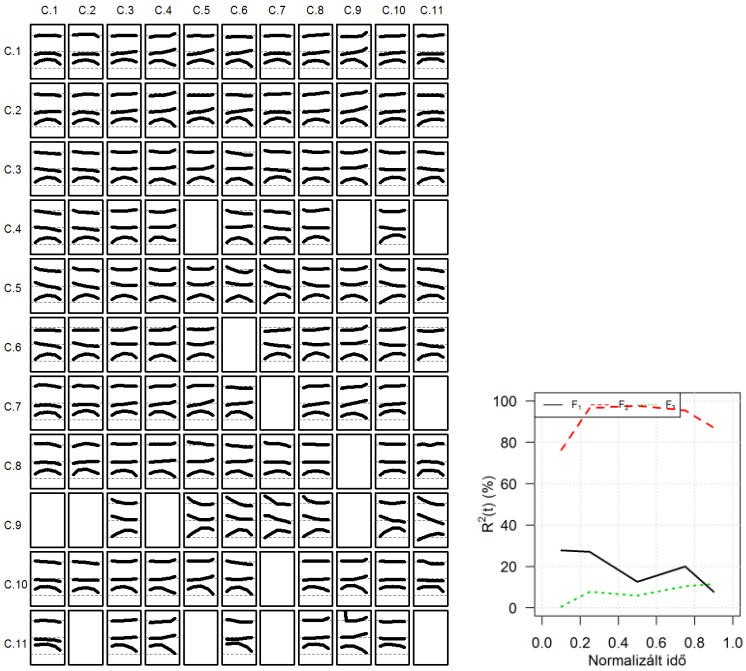
IV.1. tézis. *A formánsmenetek tanulmányozásához új vizsgálati eszközöket alakítottam ki, és ezek segítségével teljes körű áttekintést adtam magyar CVC kapcsolatok magánhangzóiban realizálódó formánsmozgásokra.*

A formánsmenetek tanulmányozásához a következő új vizsgálati eszközöket dolgoztam ki: formánstérkép, deformitási mutató és stabilitási vektor.

Formánstérkép alatt az egyes CVC hanghármasok jellemző F_1, F_2 és F_3 formánsmeneteinek együttes képét értjük egyazon ábrán megjelenítve [4]. Az egyes formánstérképeken az F_1-F_2-F_3-ra jellemző formánsmozgásokat a funkcionális varianciaelemzés módszerét alkalmazva készítettem el. A bázisfüggvények másodrendű B-spline-ok voltak. A formánstérképeket egy mátrixban, 11 sorba és 11 oszlopba rendezve jelenítem meg (7. ábra). A

sornevek a megelőző, az oszlopnevek a követő mássalhangzó kategóriáját adják meg.

A funkcionális varianciaelemzés modelljének illeszkedését az $R^2(t)$ determinációs együttható-függvény segítségével vizsgálom. Ezek áttekintése után azt mondhatjuk, hogy F_2 esetében a mássalhangzó környezetek jól meghatározzák a magánhangzóban megvalósuló második formáns alakját. A többnyire 80-90% fölötti determinációs együttható azt jelzi, hogy a formánsmozgásra vonatkozó bizonytalanságból szinte minden hatást megmagyaráz a megelőző és követő mássalhangzó környezet. Ugyanez a másik két vizsgált formáns esetén nem mondható el. A rendkívül alacsony (10-20%-os) $R^2(t)$ értékek azt jelzik, hogy a modell illeszkedése nagyon alacsony.



a) *b)*

7. ábra. Az a) részen [a:] formánstérképei láthatók. A b)-n az $R^2(t)$ determinációs együttható-függvény alakja három mássalhangzó minőségre. A folytonos vonal F_1-re, a szaggatott F_2-re és a pontozott F_3-ra vonatkozik

A *deformitási mutató*t a mássalhangzó környezetek formánsmozgató hatásának mérésére dolgoztam ki.

$$dm_{CV} = (t_2 - t_1)\big(\bar{x}(t_2) - \bar{x}(t_1)\big) + (t_3 - t_2)\big(\bar{x}(t_3) - \bar{x}(t_2)\big)$$

$$dm_{VC} = (t_4 - t_3)\big(\bar{x}(t_4) - \bar{x}(t_3)\big) + (t_5 - t_4)\big(\bar{x}(t_5) - \bar{x}(t_4)\big)$$

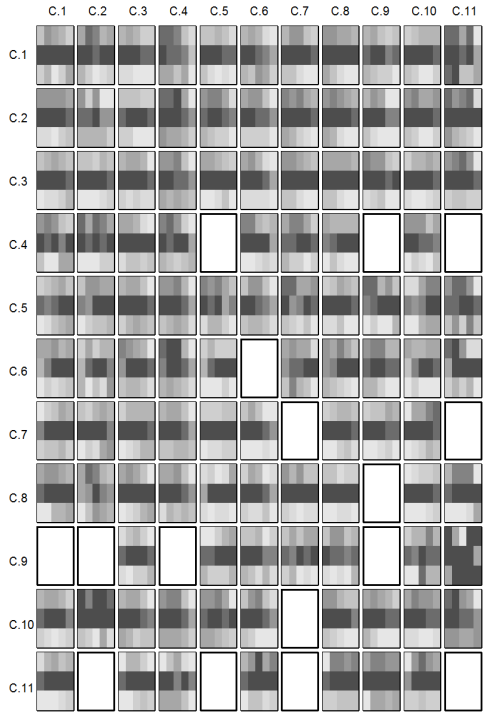
Legnagyobb deformitás F_1-ben a legalsó nyelválláshoz, majd csökkenő sorrendben az alsó, középső és felső nyelválláshoz társul. A második formáns mozgásai a hátulképzettekben nagyobb, az elölképzettekben kisebb változást szenvednek el.

A *stabilitási vektort* a vizsgált 5 mérési pontban az átlaggörbétől fél szóráson belüli (összesen egy szóráson belüli) pontok arányát méri:

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

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

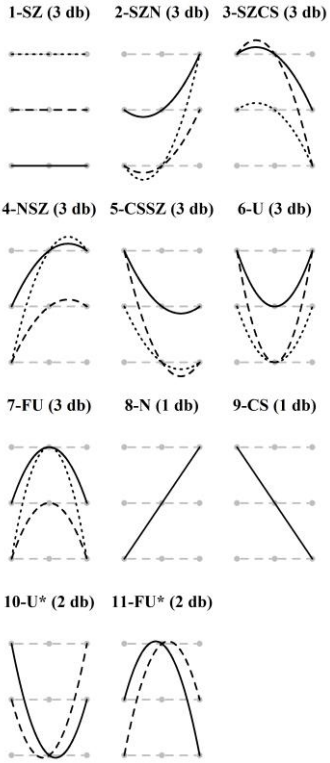
Ezzel cellánként és formánsonként 5 adattal jellemezhetjük a formánsmozgások hasonlóságát. Minél nagyobb a hasonlóság, annál közelebb van 100%-hoz az $SV(t_j)$ mutató értéke. Ennek előnye, hogy könnyen összehasonlíthatóak a különböző hanghármasokhoz tartozó fenti arányok. Továbbá könnyen implementálható hozzá grafikus megjelenítés (8. ábra), amely alapján dönthetünk, hogy melyek a nagyon vagy kevésbé hasonló formánsmozgásokkal rendelkező cellák. Méréseim eredménye szerint, 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. 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.



8. ábra. A formánsmozgások stabilitása [aː]-ban. A sötétebb helyek stabilabb formánsmozgást jeleznek

A felsorolt mérőeszközök segítségével teljes körű áttekintést adtam a CVC kapcsolatok vokálisaiban realizálódó formánsmozgásokra a környező mássalhangzó típusok függvényében, melyek rendszerét tablószerűen mutattam be.

IV.2. tézis. *Bemutattam, hogy egy egyszerűsített modell felhasználásával hogyan jósolhatók a formánsmozgások a hangkörnyezet ismeretében anélkül, hogy méréseket végeznénk.*



Jelmagyarázat:

1-SZ	szinttartó
2-SZN	szinttartó-növekvő
3-SZCS	szinttartó-csökkenő
4-NSZ	növekvő-szinttartó
5-CSSZ	csökkenő-szinttartó
6-U	U alak
7-FU	fordított U alak
8-N	növekvő
9-CS	csökkenő
10-U*	nem szimmetrikus U alak
11-FU*	nem szimmetrikus fordított U alak

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

A 9. ábrán a matematikailag lehetséges ideális formánsmenetek rendszerét mutatom be [4] alapján. 27 db lehetséges ideális formánsmozgást definiáltam, amelyeket 11 formánsmenet-típusba soroltam. Távolságot definiálhatunk két formánsmenet-típus között azzal, hogy minimálisan hány pontnyi mozgással juthatunk el az egyik típusból a másikba. Az 1 távolságra lévő típusok nagyon hasonló mozgást írnak le. A formánstérképeken szereplő jellemző

görbéket a 11 formánsmenet-típus valamelyikével kívánom helyettesíteni. Az [4]-ben részletezett, paraméterek minimalizálásán alapuló eljárással, minden egyes formánsmozgáshoz hozzárendelhetünk egy formánsmenet-típust, a lehetséges 11+(1 hiba)=12 kategóriából. A cellához tartozó jellemző formánsmenet típus (1-12) meghatározásához a cellában lévő N darab görbe típusainak megfigyelt gyakoriságait $(O_k, k = 1, \dots, 12)$ használtam fel: a cellához azt a k típust rendelem hozzá, amelyből a legtöbb volt a cellában, azaz, amelyre az O_k/N hányados maximális értéket ad. Ezt a hányadost *szigorú stabilitási mutató*nak nevezem.

$$SM_2 = \max_k O_k/N\,, k = 1, \dots, 12$$

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

2. táblázat. Az összes magánhangzó formánsmozgásainak formánsonként vett szigorú stabilitás átlaga és az 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

A hozzárendelés 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_2mutatók nem tekinthetők túl magasnak. Összesítve a 2. 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 bizonyítja, hogy a 2. táblázat formánsmeneteivel jelentősen egyszerűsíthetjük a formánstérkép jellemző formánsmeneteit.

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.

Ezek alapján azt mondhatjuk, hogy a 9. ábrán leírt formánsmenet modellel a $C_{t.1}V_mC_{t.2}$ hanghármasok F_1 és F_2 formánsmozgásai jól leírhatók, de a formánsmozgások folytonosan töltenek ki egy határozottan körülírható részt a normalizált formánstérben.

THESES IN ENGLISH

1 Theoretical background, objectives

As the efficiency of computers and the mass storage capacity of computers have further grown, most of the scientific disciplines have computational versions, which are characterized by projects requiring grand computation, like data mining for subtle patterns in vast databases or massive simulation of a physical system [7], [20]. The use of large-scale speech databases is also apparent in the field of speech science – in phonetic research and in speech technology as well – opposed to the use of small-scale, special, one research oriented databases. Thus, from the point of view of corpus phonetics, the search for variant and invariant characteristics in the speech signal and the analysis of classical problems of phonetics and phonology become possible in an ever wider scope. As regards speech technology, speech and speaker recognition programs as well as ones producing synthetic speech can obtain large-scale speech sample stored orderly for their algorithms.

The use of large-scale databases challenges researchers from several points of view. On the one hand, they have to ensure that their claims can be checked, thus measurements need to be made transparent and repeatable. Reproducible research gives an answer to the above demand [6]. On the other hand, new techniques of analysis are needed to describe and analyse the more and more complicated data structures. The functional and object oriented statistical methods of the last two decades respond to this challenge [21], [26]. The above promises the analysis and comparison of curves, while the latter offers the possibility of examining arbitrarily complex objects.

One of the basic tasks of phonetics is to examine the effect of sounds on each other. It was soon discovered that the spectral features go through change during the full length of the vowel [15]. However, in the decades following this, mainly a static model was used for the spectral description of vowels. According to this model, the information necessary for the full description of vowel quality is available in a relatively steady-state part of the vowel. However, according to the dynamic model of the spectral description of the vowel, sound transitions provide as much information for identifying the vowel as the steady-state phase [14], [23]. Several studies have proved that it is not sufficient to process the spectral information of a single time slice in many cases, as sound transitions contribute to vowel identification the same way as vowel target do.

The most important reason for the dynamic spectral characteristic feature of vowels is coarticulation, which comes about due to the neighboring phonemes affect on the articulatory configuration of a phoneme [13]. The spectral effect of coarticulation is examined on the basis of formant frequency values in the thesis.

During the articulation of vowels or vowel-like consonants, the air expelled from the lung „oscillates" the vocal cords, and voicing comes about. The voice comprising fundamental frequency and overtones is modified according to the given articulatory configuration in the vocal tract. The cavities of the vocal tract strengthen the overtones in voicing in line or near their resonance frequencies, thus energy concentrations come about, which are named formants in the specialized literature of phonetics [11], [17], [24]. The letters F_1, F_2, …, F_N are used to denote formant frequencies.

I present the creation of an own speech database and the related formant database in my thesis. The formant database contains five measurement points per vowel, thus it is suitable for analysing sound transitions, and the dynamic spectral characteristics of vowels as well, which is conducted according to the method of functional data

analysis. I also present the software environment which ensures the reproducible nature of the examinations.

2 New scientific results of the thesis

The dissertation is comprised of 4 thesis groups and 8 theses altogether.

Thesis group I Generating an own speech database fulfilling the criteria of reproducible research

Thesis I.1 *Model for reproducible speech research*

With the novel combination of tools and processes known in themselves, I elaborated the model detailed in [1], which can provide a basis for reproducible research in some fields of speech science (Figure 1). The components of the model I propose are: (1) an Emu-format speech database, which is an integrated collection of software tools for generating, querying and analysing annotated speech databases [5]; (2) software tools and packages for analysing the speech signal (e.g. Praat[7], R `emu`[8] WaveSurfer/Snack[9]), which conduct sound analysis and signal processing; (3) in the process of speech signal analysis, the tab delimited text file which do not fit the structure of the Emu database; (4) the executable publication.

The central part of the model is the executable article (Figure 1, 4), which comprises R codes and text. R codes ensure the query of data from the Emu speech database and from the delimited text files. Furthermore, R codes are used for statistical analyses and the visualization of images, tables and numerical indicators as well. The other part of the article is the text, whose format can vary. The R

[7]Praat is the most widely used free scientific software program for the analysis of speech: http://www.fon.hum.uva.nl/praat/
[8]http://cran.r-project.org/web/packages/emu/
[9]http://www.speech.kth.se/snack/

code can be combined with several markup languages. I present the use of three different form languages – LaTeX, Markdown and Org – combined with R in publication [1].

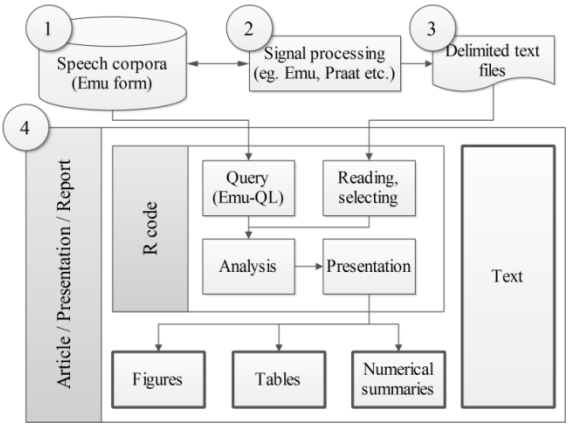


Figure 1 The proposed model of reproducible research in speech science

Thesis I.2 *Generating an own speech database (FEF)*

Along the above principles of reproducible research, I generated a large-scale speech database, which is capable of storing formant data as well. The basis of the speech database is FKM (Phonetically Balanced Sentences) speech database[10] generated in BME TMIT Laboratory of Speech Acoustics [18]. I elaborated a new process, which enabled the FKM source database to be converted into an Emu database structure as shown in Figure 1. My own FEF speech database comprises of the utterances read by 10 speakers, 5 men and 5 women. The sentences of the text read out were provided by the text file of the BABEL speech database[11] [25]. The collection of sentences is made up of carefully chosen parts of pieces of literature, the complete list comprises a total of 1992 different sentences. The

[10] A confidentiality declaration had to be filled and signed for the use of the source database.
[11] http://alpha.tmit.bme.hu/speech/hdbbabel.php

full length of the speech database is 1524 minutes (25.4 hours), it contains a total of 854 240 speech sounds and 332 827 vowels.

Thesis group II Automatic measurement and correction of formant frequencies

Thesis II.1 *Automatic determination of correct formant frequencies, search of mistakes and correction in large speech databases*

I elaborated a new process which leads to the orderly stored collection of checked formant data (Figure 2). According to my hypothesis, the elaborated multi-step formant improving process leads to an almost error-free formant database in the case of large-scale speech databases, that is, it may largely substitute manual checking.

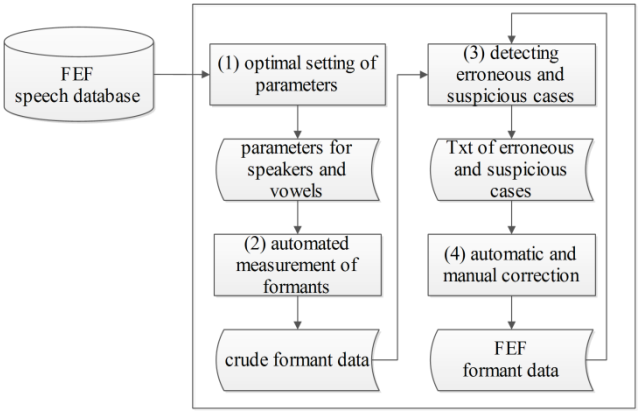


Figure 2 Steps for generating checked formant data

I prepared an own script for the automatic execution of automatic formant measures, which processed the optimal input parameters of the algorithm too, beyond the sound records and annotations stored in the FEF database. I implemented one of the methods recommended by the specialized literature for the optimal setting of

input parameters [8]. The resulting crude formant data shall be the input to a multi-step, half automatic error detecting and correcting algorithm of own development. Starting out from experience from previous research ([19] és [3]) I typified the possible mistakes of formant measurements and I created a rule-based algorithm for the automatic recognition of 5 error types.

Thesis II.2 *Method for the correction of found incorrect formant frequencies*

I developed a program with visual graphic user interface for correcting errors detected (Figure 3). The program called PROFEF is the improved version of the Interactive Formant Editor[12] [2], which operates in a web-based environment and was developed earlier. PROFEF supports both manual and automatic methods in correcting errors.

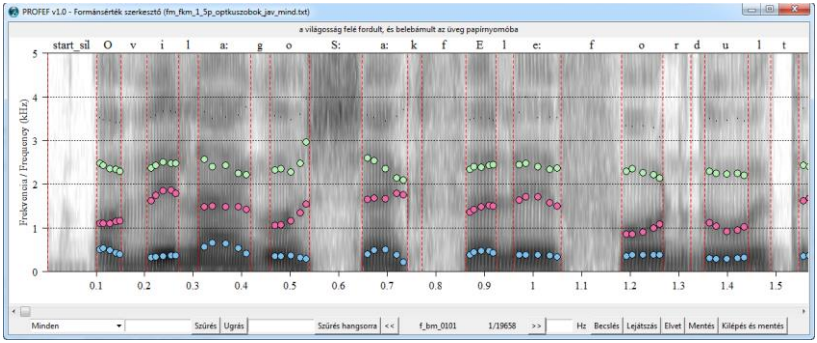


Figure 3 A screenshot of PROFEF application

I use a tab delimited text file to store checked formant data. Table 1 shows the set-up of the text file.

Column *file* of the formant data text file shows the identifier of the speech sample. *Num* identifies the position of the vowel within the speech sample: which segmented element it is in the line. *Label* is

[12] Available at: http://magyarbeszed.tmit.bme.hu/ifem

the symbol of the vowel in SAMPA notation. *Time* is a point on the time axis (measured in seconds), which marks the place of the measurement. *Pos* is an identifier which differentiates points of measurement within a vowel (in our case 10, for example, means the 10% measurement point). The last three columns list F_1-F_3 formant frequency values in Hz.

Table 1 A few lines from the formant database

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

Thesis group III Modelling and analysing of the formant space of Hungarian vowels

Thesis III.1 *Functional data analysis in the normalized formant space, new figures*

The unified treatment of formant values related to the different speakers of the speech database requires data to be normalized. Every formant frequency value has two coordinates, a time and a frequency component. Of the many possible so-called vowel normalization processes for the frequency axis, I chose Lobanov's method [16], which is easily interpretable and well applicable for eliminating differences between the speakers (for example, physiological or anatomical differences), while the speaker-independent part of the sound string, the phonemic information, remains unchanged [8]. Lobanov's method is based on the Z-transform used in mathematical statistics.

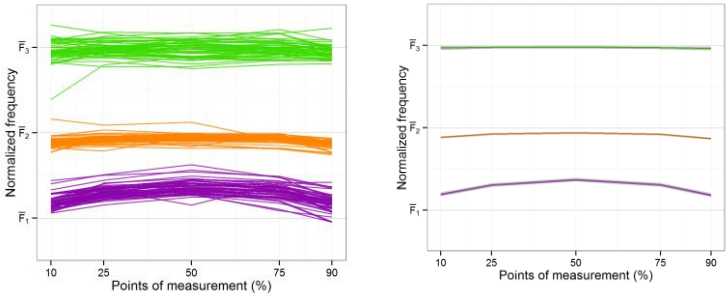


figure a) *figure b)*

Figure 4 Normalized formant curves in part a) and the average curves taking their places in part b)

In theory, the normalized formant curves related to the three different formants cannot be visualized in a single figure because axis y comprises normalized formant frequency values. Nonetheless, in a merely technical step, the normalized curves can be drawn above each other in the ascending order of formant indices. It means that the normalized values of the second formant appear 4 units pushed up as compared to the values of the first formant, and the curves of the third formant are 8 units pushed up as compared to the first (that is, 4 units above the second). Part a) of figure 4 contains the normalized formant curves accumulated on each other. The three horizontal lines in the figure show the individual formant means. The figure is informative in many respects. On the one hand, it contains the data of every speaker in a united form, which means a significant simplification as the number of figures does not need to be further multiplied by ones on age or gender. Formant indices are set out in their natural order from bottom to top, in an ascending order. The average lines drawn indicate where the given formant trajectories are compared to the average. In our case, F_1 values, for example, are set well above the average. The position of curves related to different formant indices as compared to each other is also relevant, as it gives

an answer to the question how close or far the given formant frequency values are from each other.

The method SSANOVA [12] allows us to replace our curves with a single smoothing spline, and besides this, we may conclude on the deviation of the curves from the size of the Bayes confidence intervals. The splines in part b) of figure 4 are capable of describing the formant movements of any vowel in any sound environment. A similar result can be achieved through the implementation of variance analysis as well [22].

The SSANOVA model is appropriate for comparing formant movements coming about in different sound environments. If the [a:] formant movement between bilabial consonants is examined in different –2. sound environments, the result is the easily generalizable figure 5. We define different groups depending on the sound environments to be compared, and we test the deviation of the formant curves of the groups. The example presented in figure 5 defines three groups on the basis of the occurrence of sound –2., which can be [i], or [o, u] and the third group means the occurrence of any other sound. The overlap of the Bayes confidence intervals drawn around the curves shows whether the individual groups significantly deviate, and if so, at which part of the curve. It is visible from figure 5 that the previous [i] and [o, u] significantly decrease the value of F_1, as the confidence intervals do not cover each other here. In F_2, there shows some frequency increase upon the effect of [i] in the beginning phase of the vowel. There is no deviation in F_3 as regards the three groups.

Through the above I have presented that the method of functional data analysis is capable of modelling and comparing formant movements in normalized formant space. I have adapted and further developed the graphic depiction used so far for evaluating the model [22].

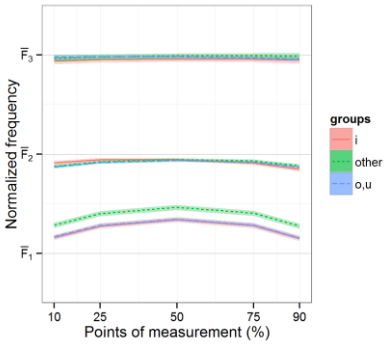


Figure 5 SSANOVA's graphic depiction in the case of three curve groups for the three formants. There is a significant difference in F_1, as [i] and [o, u] uniformly decrease the value of the first formant compared to the other sounds in position –2.

Thesis III.2 *Analysing the effect of sound +2 following the vowel on formant movement*

With the help of the graphic outline in thesis III.1., I presented that the second formant movement of [ɔ] realizations, depends on not only from the following consonant, but from the +2nd sound too. In the measured sound sequences the F_2 movement, show significant deviation in the two groups, which is due to the effect of sound +2 (Figure 6). 10-10 samples [vɔhi] and [vɔho] have been analysed in the FEF database. The difference in the samples is represented only in the fourt sound. As a supplement, the figure visualizes numerical information as well: the number of sample elements (curves). In the right side of figure 6, the group and interaction effect appear together, which makes the deviation explicit.

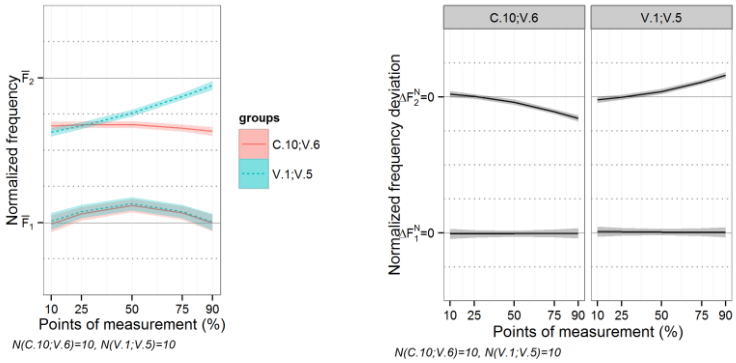


Figure 6 Presentation of the significant effect of sound +2.on the second formant of the vowel

Thesis group IV New phonetic tools for measurements

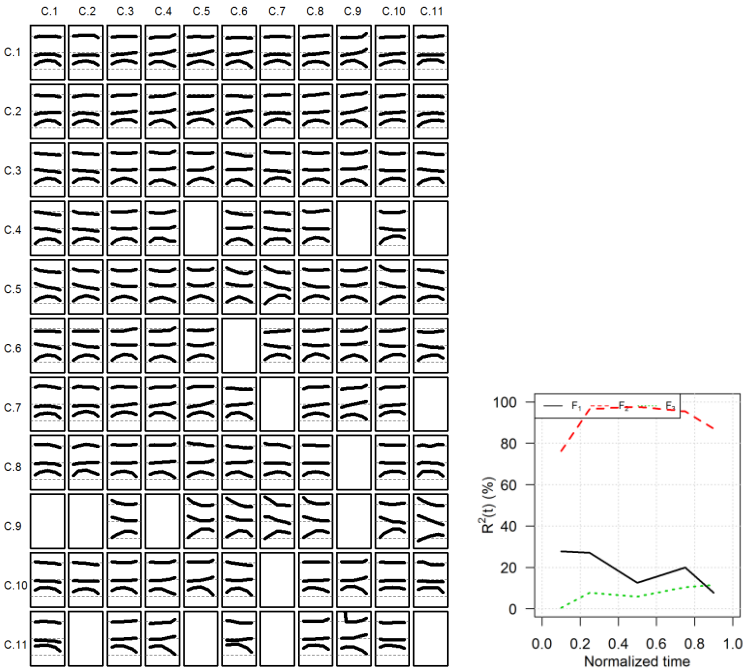
Thesis IV.1 *New tools for analysis and presenting their application*

I created new tools for analysis to examine formant trajectories. These are the formant map, deformity index and the stability vector.

Formant map is the collection of formant trajectories typical of individual sound triplets [4]. I drew the formant movements characteristic of F_1-F_2-F_3 in the individual formant maps with the help of functional variance analysis. Base functions were second-rate B-splines. I present formant maps in a matrix, arranged in 11 rows and 11 columns (Figure 6). Row labels show the category of the preceding consonant, whereas column labels give that of the consequential consonant.

I analyse the fitting of the model of functional variance analysis with the help of $R^2(t)$ determination coefficient function. Having reviewed them, we are able to state that in the case of F_2, the form of the second formant shaping up in the vowel is well determined by the consonant environments. The coefficient of determination mostly above 80-90% means that almost all effects from the instability

concerning formant movements are explained by the preceding and consequential consonant environment. The same cannot be said for the other two formants examined. The extremely low (10-20%) $R^2(t)$ values signify that the matching of the model is very low.



a) *b)*

Figure 7 Part a) shows the formant maps of [aː]. Part b) presents the form of $R^2(t)$ determination coefficient function on three consonant qualities. The solid line refers to F_1, the dashed line refers to F_2, and the dotted line to F_3

I elaborated a *deformity index* to measure the formant moving effect of consonant environments.

$$dm_{CV} = (t_2 - t_1)\big(\bar{x}(t_2) - \bar{x}(t_1)\big) + (t_3 - t_2)\big(\bar{x}(t_3) - \bar{x}(t_2)\big)$$
$$dm_{VC} = (t_4 - t_3)\big(\bar{x}(t_4) - \bar{x}(t_3)\big) + (t_5 - t_4)\big(\bar{x}(t_5) - \bar{x}(t_4)\big)$$

Deformity indices are of greatest value for vowels that have extreme y-offset. In the case of consonant types, type C.9, C.6 and C.7 are outstanding.

The *stability vector* measures the ratio of points within a half deviation from the average curve (within one deviation altogether) in the five measurement points:

$$SV(t_j) = \frac{\left|\left\{y_{ij}\middle|\bar{x}(t_j) - \frac{1}{2} \le y_{ij} \le \bar{x}(t_j) + \frac{1}{2}\right\}\right|}{N},$$

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

The above allows us to characterize the similarity of formant movements with 5 data per cell and formant. The greater the similarity is, the closer the value of $SV(t_j)$ index is to 100%. Its advantage is that the above ratios related to the different sound triplets become easily comparable. Furthermore, a graphic visualization is easily implementable to it, on the basis of which we can decide which are the cells with very similar or less similar formant movements. According to the result of my measurements, formant movements have greatest stability in F_2 (0.87 is the value of the average index cumulated on 5 points), while in the case of F_1 this value is 0.81. The average ratio of the points of the third formant which are beyond one deviation is 0.55. Within the different $C_{t.1}V_mC_{t.2}$ sound triplets, the formant movements of F_2 are more unified compared to the formant movements of F_1, and the movements of the third formant are characterized by significantly greater diversity than that of the above two.

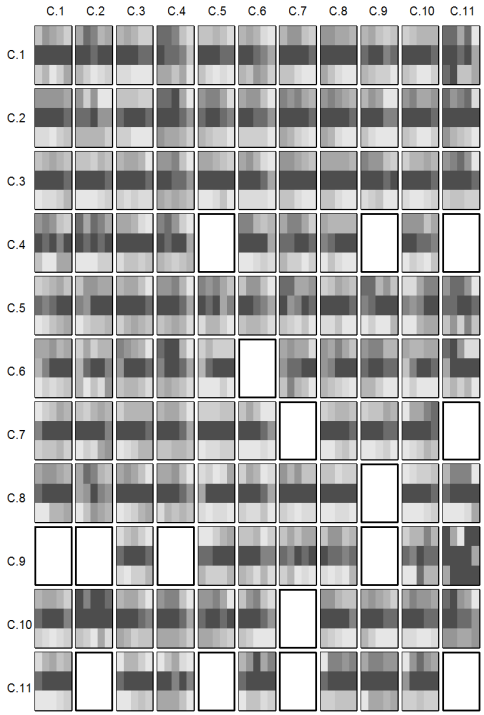
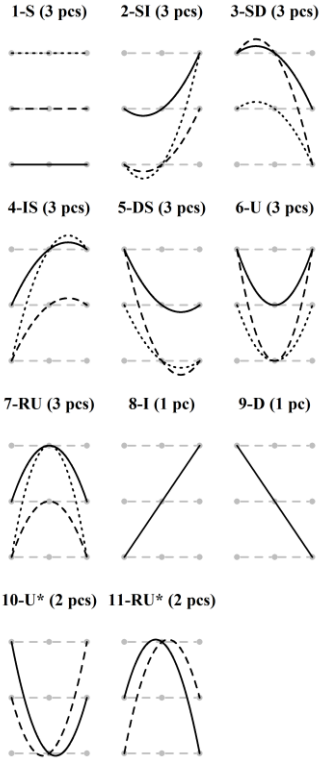


Figure 8 The stability of formant movements in [a:]. The darker spots indicate more stable formant movements

With the help of the tools for measurement listed above, I provided a comprehensive overview on the formant movements realized in the vocals of CVC relations, whose system I presented in a tableau-like format.

Thesis IV.2. *Prediction of formant movements in the function of the sound environment without measurements performed*



	Notation:
1-S	straight-line
2-SI	straight-increasing
3-SD	straight-decreasing
4-IS	increasing-straight
5-DS	decreasing-straight
6-U	U shape
7-RU	reverse U shape
8-I	increasing
9-D	decreasing
10-U*	non-symmetrical U shape
11-RU*	non-symmetrical reverse U shape

Figure 9 The 11 theoretical categories of formant trajectory types. The number of formant trajectories covered by type can be seen in brackets

Figure 9 presents the mathematically possible system of ideal formant trajectories, on the basis of [4]. I defined 27 possible ideal formant movements, which were categorized in 11 formant trajectory types. A distance can be defined between two formant trajectory types by determining the minimum number of points of

movement to be taken to get from one type to the other. Types at a distance of 1 have very similar movements. I intend to substitute the characteristic curves in the formant maps with some of the 11 formant trajectory types. According to the process detailed in [4] and based on the minimalization of parameters, a formant trajectory type of the possible 11+(1 error)=12 category can be assigned to every single formant movement. In order to determine the characteristic formant trajectory type (1-12) related to the cell, I used the observed frequency of N pieces of curve types in the cell ($O_k, k = 1, \dots, 12$): I assign type k to the cell which was most numerous in the cell, that is, to which O_k/N quotient gives the maximum value. I name this quotient *strict stability index*.

$$SM_2 = \max_k O_k/N\,, k = 1, \dots, 12$$

Its value can be between 0 and 1, the closer it is to one, the more stably the formant trajectory type describes the given formant of the given cell.

Table 2 The strict stability average per formant of the formant movements of all vowels and the value of *the joint strict stability index*

V_m	SM_2 average	Joint SM_2 average
F_1	0.52	0.73
F_2	0.58	0.78
F_3	0.35	0.56

The assignment is successful as the 12. other (error category) occurred only in F_3 in a 9%-ratio. The SM_2 indices indicating the strict stability of formant trajectory types assigned to cells cannot be considered too high. They can be seen in total in Table 2. However, if we add the second most frequent type to the formant trajectory of the cell (provided that the distance between them is 1), the value of the joint index significantly increases. This proves that the characteristic formant trajectories of the formant map can be significantly simplified by the formant trajectories of Table 2.

The formant trajectory types assigned to the $C_{t.1}V_mC_{t.2}$ sound triplets are more diverse in F_2 but especially in F_3 than in F_1, and the stability of the assignment is also greater in F_2 and F_1 compared to F_3. If the categories attaining 5 percent are considered, the first formant can be characterized by 4 formant trajectory types, and F_2 and F_3 by 5 categories each.

On the basis of the above, it can be stated that F_1 and F_2 formant movements of $C_{t.1}V_mC_{t.2}$ sound triplets are well characterizable by the formant trajectory model put forward in Figure 9, but formant movements continuously fill in a well-definable part in the normalized formant space.

HIVATKOZÁSOK – REFERENCES

[1] Kálmán Abari. Reproducible research in speech sciences. *International Journal of Computer Science Issues*, 9(6):43–52, November 2012.

[2] Kálmán Abari and Gábor Olaszy. Interaktív formánsérték-módosító fejlesztése. In Attila Tanács and Veronika Vincze, editors, *Magyar Számítógépes Nyelvészeti Konferencia*, pages 309–315, Szeged, 2011. Szegedi Tudományegyetem Informatikai Tanszékcsoport.

[3] Kálmán Abari and Gábor Olaszy. Magyar formánsadatbázis az interneten. In Mária Gósy, editor, *Beszédkutatás*, pages 73–82. MTA Nyelvtudományi Intézet, Budapest, 2011.

[4] Kálmán Abari and Gábor Olaszy. A formánsmenetek rendszere CVC kapcsolatok magánhangzóiban a C képzési helyének függvényében. In Mária Gósy, editor, *Beszédkutatás*, pages 70–93. MTA Nyelvtudományi Intézet, Budapest, 2012.

[5] Lasse Bombien, Steve Cassidy, Jonathan Harrington, Tina John, and Sallyanne Palethorpe. Recent Developments in the Emu Speech Database System. In Paul Warren and Catherine I. Watson, editors, *Proceedings of the 11th Australasian International Conference on Speech Science and Technology*, pages 313–316, Auckland, New Zealand, December 6 2006. Australian Speech Science & Technology Association.

[6] Jon Claerbout and Martin Karrenbach. Electronic documents give reproducible research a new meaning. In *Proc. 62nd Ann. Int. Meeting of the Soc. of Exploration Geophysics*, pages 601–604, 1992.

[7] David L. Donoho, Arian Maleki, Inam Ur Rahman, Morteza Shahram, and Victoria Stodden. Reproducible research in computational harmonic analysis. *Computing in Science & Engineering*, 11:8–18, 2009.

[8] Paola Escudero, Paul Boersma, Andréia Schurt Rauber, and Ricardo A. H. Bion. A cross-dialect acoustic description of vowels: Brazilian and european portuguese. *Journal of the Acoustical Society of America*, 126(3):1379–1393, 2009.

[9] Nicholas Flynn and Paul Foulkes. Comparing vowel formant normalization methods. *Proceedings of the 17th International Congress of Phonetic Sciences ICPhS 2011*, (August):683–686, 2011.

[10] Josef Fruehwald. SS ANOVA. http://www.ling.upenn.edu/ joseff/papers/fruehwald_ssanova.pdf, 2010. Letöltve: 2013. június. 20.

[11] Mária Gósy. *Fonetika, a beszéd tudománya*. Osiris Kiadó, Budapest, 2004.

[12] Chong Gu. *Smoothing Spline ANOVA Models*. Springer Series in Statistics. Springer, 2002.

[13] William J. Hardcastle and Nigel Hewlett. *Coarticulation: Theory, Data and Techniques*. Cambridge Studies in Speech Science and Communication. Cambridge University Press, 1999.

[14] Jonathan Harrington. *Acoustic Phonetics*, chapter 3, pages 81–129. Blackwell Publishing Ltd., 2010.

[15] Martin Joos. *Acoustic Phonetics*. Number 24 in Language Monograph. Linguistic Soc. of America, 1948.

[16] Boris M. Lobanov. Classification of russian vowels spoken by different speakers. *Journal of the Acoustical Society of America*, (49):606–608, 1971.

[17] Géza Németh and Gábor Olaszy. *A magyar beszéd: Beszédkutatás, beszédtechnológia, beszédinformációs rendszerek*. Akadémiai Kiadó, Budapest, 2010.

[18] Gábor Olaszy. Precíziós, párhuzamos, magyar beszédadatbázis fejlesztése és szolgáltatásai. In Mária Gósy, editor, *Beszédkutatás*. MTA Nyelvtudományi Intézet, Budapest, 2013.

[19] Gábor Olaszy, Zsuzsanna Zsófia Rácz, and Mátyás Bartalis. Formánsmérések automatizálása, formánsadatbázisok létrehozása. In Mária Gósy, editor, *Beszédkutatás*, pages 134–147. MTA Nyelvtudományi Intézet, Budapest, 2009.

[20] Roger D. Peng. Reproducible Research in Computational Science. *Science*, 334(6060):1226–1227, December 2011.

[21] James O. Ramsay. *Encyclopedia of Statistics in Behavioral Science*, volume 2, chapter Functional Data Analysis, page 675–678. John Wiley & Sons, Chichester, 2005.

[22] James O. Ramsay, Giles Hooker, and Spencer Graves. *Functional Data Analysis with R and MATLAB*. Use R! Springer, 2009.

[23] Winifred Strange and James J. Jenkins. Dynamic specification of coarticulated vowels. In Geoffrey Stewart Morrison and Peter F. Assmann, editors, *Vowel Inherent Spectral Change*, pages 87–115. Springer Berlin Heidelberg, 2013.

[24] István Subosits. *Hangtan*. TAS-11 Kft., Budapest, 2005.

[25] Klára Vicsi and Attila Víg. Az első magyar nyelvű beszédadatbázis. In Mária Gósy, editor, *Beszédkutatás*, pages 163–177. MTA Nyelvtudományi Intézet, Budapest, 1998.

[26] Haonan Wang and J. S. Marron. Object oriented data analysis: Sets of trees. *Annals of Statistics*, 35(5):1849–1873, 2007.

PUBLIKÁCIÓS LISTA – LIST OF PAPERS

Referált publikációk – Referred publications

1. Abari, Kálmán (2012): Reproducible research in speech sciences. In *International Journal of Computer Science Issues.* 9, 43–52. (http://ijcsi.org/papers/IJCSI-9-6-2-43-52.pdf).

   Impact Factor: 0,242

2. Abari, Kálmán–Rácz, Zsuzsanna Zsófia–Olaszy, Gábor (2011): Formant maps in Hungarian vowels - online data inventory for research, and education. In *Proc. of the International Conference on Speech and Technology (Interspeech)*. 1609–1612.

   DBLP: conf/interspeech/2011

3. Tamm, Anne–Abari, Kálmán–Olaszy, Gábor (2007): Accent Assignment Algorithm in Hungarian Based on Syntactic Analysis. In *Proc. of the International Conference on speech and technology (Interspeech)*, Antwerp, 466–469.

   DBLP: conf/interspeech/2007

4. Abari, Kálmán–Páli, Gábor (2007): Problems of acoustic echo cancellation. In *ICAI 2007 - 7th International Conference on Applied Informatics.* Eger, 335–342.

   Ref. szám: Zentralblatt MATH Database Zbl 1179.94029

5. Hunyadi, László–Abari, Kálmán–Tóth, Enikő (2003): Forensic Linguistics: its Contribution to Humanities Computing. *Literary and Linguistic Computing*, Vol. 18, No. 1, 49-62.

   DBLP: journals/lalc/HunyadiAT03

Angol nyelvű lektorált publikációk – Peer-reviewed publications in English

1. Abari, Kálmán (2008): From text-to-sound - A Hungarian word level pronunciation database using IPA symbols. *The Phonetician* 97-98, 93–104.

2. Olaszy, Gábor–Rácz, Zsuzsanna Zsófia–Abari, Kálmán (2008): A formant trajectory database of Hungarian vowels. *The Phonetician* 97-98, 7–14.

Magyar nyelvű lektorált publikációk – – Peer-reviewed publications in Hungarian

3. Abari Kálmán–Olaszy Gábor (2012). Új módszer nagyméretű beszédadatbázisok formánsadatokkal történő ellátására. In Gósy Mária (szerk.) *Beszéd, adatbázisok, kutatások*. Budapest, Akadémiai Kiadó, 251-267.

4. Abari Kálmán–Olaszy Gábor (2012): A formánsmenetek rendszere CVC kapcsolatok magánhangzóiban a C képzési helyének függvényében. In Gósy Mária (szerk.) *Beszédkutatás*. Budapest, MTA Nyelvtudományi Intézet, 70–93.

5. Abari Kálmán–Olaszy Gábor (2011): Magyar formánsadatbázis az interneten. In Gósy Mária (szerk.) *Beszédkutatás*. Budapest, MTA Nyelvtudományi Intézet, 73–82.

6. Abari Kálmán–Olaszy Gábor (2011): Interaktív formánsérték-módosító fejlesztése. In Tanács Attila–Vincze Veronika (szerk.) *Magyar Számítógépes Nyelvészeti Konferencia*. Szeged, Szegedi Tudományegyetem Informatikai Tanszékcsoport, 309–315.

7. Abari Kálmán (2010). Intelligens beszédhang-időtartam mérő. In Németh Géza, Olaszy Gábor (szerk.) *A magyar beszéd - beszédkutatás, beszédtechnológia, beszédinformációs rendszerek*. Budapest, Akadémiai Kiadó, 643–646. ISBN: 9789630589666.

8. Olaszy Gábor–Abari Kálmán (2010): A magyar hangkapcsolódások akusztikai bemutatása szavakban. In Németh Géza, Olaszy Gábor (szerk.) *A magyar beszéd - beszédkutatás, beszédtechnológia, beszédinformációs rendszerek*. Budapest, Akadémiai Kiadó, 315–319. ISBN: 9789630589666.

9. Olaszy Gábor–Abari Kálmán (2010): Elektronikus kiejtési szótár IPA-jelekkel és hangidőtartamokkal. In Németh Géza, Olaszy Gábor (szerk.) *A magyar beszéd - beszédkutatás,*

beszédtechnológia, beszédinformációs rendszerek. Budapest, Akadémiai Kiadó, 320–324. ISBN: 9789630589666.

10. Abari Kálmán–Papp Zoltán (2008): „Beszédtechnológiai alapismeretek" az oktatásban. In: Pethő Attila, Herdon Miklós (szerk.) *Informatika a felsőoktatásban 2008.* Debreceni Egyetem, Informatikai Kar, Debrecen. ISBN 978-963-473-129-0

11. Abari Kálmán–Olaszy Gábor (2007): A magyar beszéd hangkapcsolódásainak bemutatása az interneten. In Gósy Mária (szerk.) *Beszédkutatás 2007*. MTA Nyelvtudományi Intézet, 178-186.

12. Abari Kálmán–Tamm Anne–Gábor Kata–Olaszy Gábor (2007): Számítógépes összehasonlító szövegelemzés ügyfélszolgálati tájékoztatók legfontosabb prozódiai elemeinek a meghatározására. In Alexin Zoltán–Csendes Dóra (szerk.) *V. Magyar Számítógépes Nyelvészeti Konferencia.* Szegedi Tudományegyetem Informatikai Tanszékcsoport, Szeged, 24-33.

13. Abari Kálmán (2006): Hangos jelölőnyelvek. Jelölőnyelvek a beszéd alapú alkalmazások fejlesztésében. *Híradástechnika,* 2006/3. Hírközlési és Informatikai Tudományos Egyesület, Budapest, 2-7.

14. Abari Kálmán–Olaszy Gábor (2006): Internetes beszédadatbázis a magyar mássalhangzó kapcsolódások akusztikai szerkezetének bemutatására. In Alexin Zoltán–Csendes Dóra (szerk.). *IV. Magyar Számítógépes Nyelvészeti Konferencia.* Szegedi Tudományegyetem Informatikai Tanszékcsoport, Szeged, 213-222.

15. Abari Kálmán–Olaszy Gábor–Kiss Géza–Zainkó Csaba (2006): Magyar kiejtési szótár az Interneten. In Alexin Zoltán–Csendes Dóra (szerk.) *IV. Magyar Számítógépes Nyelvészeti Konferencia.* Szegedi Tudományegyetem Informatikai Tanszékcsoport, Szeged, 223-230.

16. Olaszy Gábor–Abari Kálmán (2005): Adatbázisok és számítógépprogramok a magyar beszéd időszerkezeti vizsgálatához. *Alkalmazott Nyelvtudomány* V/1-2, 41-62.

További publikációk – Other publications

17. Máth János–Abari Kálmán (2011): Knowledge spaces and historical knowledge in practice. *Applied Psychology in Hungary* 2011(1), 124–150.

18. Abari Kálmán–Máth János (2010): A történelmi tudás mérése a tudástér-elmélet segítségével. In Münnich Ákos, Hunyady György (szerk.) *A nemzeti emlékezet vizsgálatának pszichológiai szempontjai*. Budapest, ELTE Eötvös Kiadó, 191–216. ISBN: 9789633120354.

19. Boncsér Ágnes–Polonyi Tünde–Abari Kálmán (2010): Szövegértő olvasás a két tannyelvű oktatásban résztvevő diákok körében. *Alkalmazott Pszichológia* XII(3-4), 5–30.

20. Polonyi Tünde–Abari Kálmán–Nótin Ágnes (2009). Mesterséges nyelvtanulás első benyomás alapján, *Alkalmazott Pszichológia*, XI, 1-2, 5-26.

21. Abari Kálmán (2008): A tehetségdiagnosztika adatkezelésbeli alapjai R-környezetben. In Mező Ferenc (szerk.) *Tehetségdiagnosztika.* Kocka Kör – Faculty of Central European Studies, Constantine the Philosopher University in Nitra, Debrecen. 105-130.

22. Mező Ferenc–Máth János–Abari Kálmán (2008): Iskolai hatásvizsgálatok. In Mező Ferenc (szerk.) *Tehetségdiagnosztika.* Kocka Kör – Faculty of Central European Studies, Constantine the Philosopher University in Nitra, Debrecen. 131-203.

23. Münnich Ákos-Nagy Ágnes-Abari Kálmán (2006): *Többváltozós statisztika pszichológus hallgatók számára*. Debrecen: Bölcsész Konzorcium (http://psycho.unideb.hu/statisztika). ISBN: 963 9704 04 0