Budapesti Műszaki és Gazdaságtudományi Egyetem
Távközlési és Médiainformatikai Tanszék

# Gépi beszédkeltés infokommunikációs rendszerekben

PhD tézisfüzet
Informatikai Tudományok Doktori Iskola

Zainkó Csaba
okl. mérnök-informatikus

Témavezetők:
Németh Géza, PhD
Olaszy Gábor, DSc

Budapest, 2010.

# 1. Bevezetés

Az infokommunikációs rendszerekben egyre fontosabb szerep jut az ember és gép közötti természetes kommunikációnak. A felhasználók és a felhasználás jellege megváltozott, ma már nem csak a számítógépes szakemberek kezelik a rendszereket, hanem emberek tömegesen váltak mindennapi felhasználóvá. A sikeres ember-gép kapcsolat kialakításához szükség van olyan megoldásokra, amelyek lehetővé teszik, hogy ne csak szakemberek értsék meg a gép utasításait, közléseit, illetve tudjanak utasítást adni, hanem bárki.

Az ember-gép kommunikációját megkönnyítő egyik kutatási témakör – amellyel kutatásaim során legtöbbet foglalkoztam – a gépi beszédelőállítás és a hozzá kapcsolódó gépi szövegfeldolgozás. Ezeken a területeken a problémák és a megoldások jelentős része nyelvfüggő, ezért egy-egy kérdéskörrel akkor is foglalkozni kell, ha az más nyelvekre – főleg angolra – már megoldott és publikált. A magyar nyelv tulajdonságai miatt más típusú nyelvekre kidolgozott megoldások nem, vagy csak részben alkalmazhatók, illetve máshol elvetett eljárás magyarra működőképes lehet. Ugyanakkor a magyar nyelvű megoldások kutatása nemzetközileg is hasznosítható, a kidolgozott eljárások, megállapítások hasonló ragozó nyelvekre általában könnyen adaptálhatók, amennyiben a nyelvfüggő részek jól körülhatároltak.

Az első beszédszintézisre vonatkozó kísérletekről több mint 200 éve 1791-ben Kempelen Farkas számolt be (Kempelen 1969 (eredeti kiadás: 1791)). Az első elektronikus beszédszintézis azonban csak a XX. században valósult meg a Bell Laboratóriumban 1939-ben (Homer et al. 1939) kézi vezérléssel. A gépi beszédképzést Fant (1960) alapozta meg, az első formánsszintézis alapú szövegfelolvasót Rabiner (1968) ismertette. A magyar kutatások egy évtizeddel később kezdődtek a témakörben és az első magyar nyelvű számítógépes beszédszintetizátor 1980-ban valósult meg az MTA Nyelvtudományi Intézetének Fonetikai Laboratóriumában Hungarovox néven (Kiss–Olaszy 1984). Az első szintetizátorok hangja nagyon robotos és adott környezetben viszonylag érthető volt. Jelentős kutatás és fejlesztés folyt a BME-n is Gordos Géza vezetésével (Gordos–Takács 1983, Gordos–Sándor 1985). Ezekben az időkben angol nyelvterületen a MITalk (Allen et al. 1987) és a DecTalk (Hallahan 1995) rendszerek készültek el. Szintén meghatározó eredmény a PSOLA algoritmus (Hamon et al. 1989) kidolgozása a prozódia gépi módosítására, amelyet sok későbbi kutatás felhasznált. A ’90-es években a számítógépek teljesítményének növekedése lehetővé tette a tisztán szoftver alapú megoldásokat is, amelyek hangminősége már megközelítette az emberi beszédet. Ezen munkák összefoglalása megtalálható Olaszi Péter PhD értekezésében (Olaszi 2002). A szövegfeldolgozás fejlődésének nagy lehetőséget adott az interneten található nagy mennyiségű írott anyag. Angol nyelvterületen a British National Corpus munkálatai szövegadatbázis gyűjtésére 1991-ben indultak (Burnard 1995). A magyar nyelv vonatkozásában 1998-ban kezdődött el az MTA - Nyelvtudományi Intézetének Korpusznyelvészeti Osztályán a Magyar Nemzeti Szövegtár készítése (Váradi 1999). A BME Távközlési és Telematikai Tanszék Beszédtech-

nológia Laboratóriuma munkájába 1997-ben kapcsolódtam be, amikor a Profivox diád alapú hullámforma-összefűzéses szintetizátor készült (Olaszy et al. 2000).

Angol nyelvterületen meghatározó szintézisrendszer volt a CHATR (Black–Taylor 1994) illetve a szabad forráskódú Festival rendszer (Black et al. 2006). Az ezredforduló után a természetes hangzás elérése volt a legfőbb cél. Ennek egyik sikeres megoldása a Rhethorical rVoice nevű elemkiválasztásos rendszere volt, amelyhez több nyelvre is készültek hangadatbázisok (Rutten et al. 2002, Rutten–Fackrell 2003). Szintén a 2000-es évek elején indult el egy új irányzat, amely a beszédfelismerésben már bevált rejtett Markov modelleket (HMM) alkalmazó technikát ültette át a beszédszintézisbe (Tokuda et al. 2000, Zen et al. 2007). A természetes hangzású beszéd szintézise mellett az érzelmi töltetű beszéd előállításának a kutatása is megkezdődött (Scherer 2003).

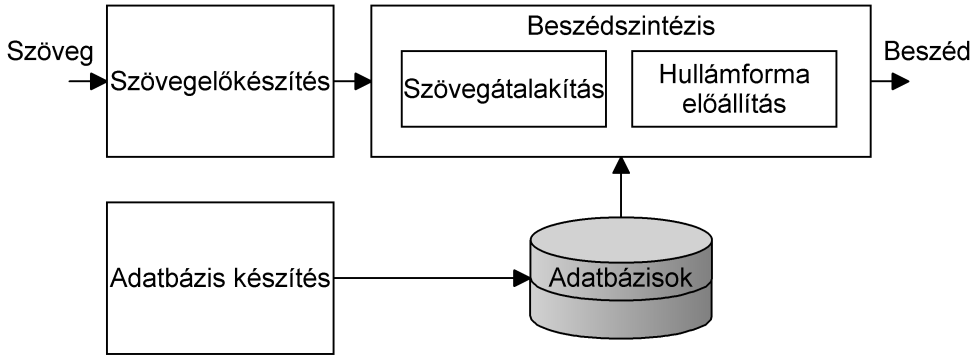
A magyar beszédszintézis területén kisebb cégek készítettek saját fejlesztésű zárt rendszereket, például a Speech Technology Kft. Ezekről a rendszerekről publikációk nem jelentek meg. Az utóbbi években külföldi fejlesztések is elindultak. Elkészült az MBROLA szabad felhasználású szintézis rendszer magyar nyelvű adaptációja (MBROLA 2006). 2010-ben jelentek meg nagyobb cégek termékeként magyar nyelvű beszédszintetizátorok, mint például az SVOX vagy a Nuance beszédszintetizátorai. Az utóbb felsorolt rendszerekről sem találtunk publikációt.

Az ezredforduló után folyamatos magyar nyelvű beszédszintézis kutatás csak a BME Távközlési és Telematikai/Médiainformatikai Tanszékén folyt. Elkészült a név- és cím felolvasására alkalmas hibrid rendszer [C4], majd a természetes beszéd minőségét megközelítő korpusz alapú szintetizátor [C12]. A nemzetközi kutatási irányokat követve a tanszék Beszédtechnológia Laboratóriuma bekapcsolódott a HMM alapú beszédszintézis kutatásába és magyar nyelvű megoldást készített (Tóth–Németh 2008). A semleges beszéd szintézise mellett elkezdődött az érzelem kifejezésére alkalmas megoldások kutatása is [C5,C9].

## A beszédszintézis

A beszédszintézis folyamata során a bemeneti szöveg előkészítése az első lépés. Az 1.1. ábrán a beszédszintézis egyszerűsített blokkvázlata látható. A szövegelőkészítés lépésben a bemenetre érkező karaktersorozatból állítunk elő egy olyan betűsorozatot, amely már csak az ábécé betűit, a szóköz karaktert és a központozást tartalmazza. Az előkészítés során feloldjuk a rövidítéseket, az idegen neveket és szavakat felolvasható formára írjuk át. A számokat szintén átírjuk szöveges formára, figyelembe véve, hogy éppen telefonszámról, pénzösszegről, dátumról vagy tő- és sorszámnévről van-e szó. Ebben a lépésben végezzük el például az ékezet nélkül írt szövegekben az ékezetes karakterek visszaállítását is.

Az előkészített szöveg kerül a beszédszintetizátor központi blokkjának bemenetére. Ez a központi rész két főbb egységre tagolható. Az első a szövegátalakítás, a második a hullámforma előállítás. A szövegátalakítás során a bemeneti betűsorozatból egy hangsorozatot és a hangsorozathoz rendelhető prozódiai információkat




*1.1. ábra.* Beszédszintézis folyamatának egyszerűsített blokkvázlata

állítunk elő. Az így elállított adatok alapján előállítjuk a hullámformát, amely a beszédszintézis technológiájától függően vagy a végleges beszéd, vagy egy nyers hangsorozat. A nyers hangsorozatot különböző további jelfeldolgozási lépésekkel alakítjuk át a végleges beszéddé.

A beszédszintetizátorok általában különböző típusú adatbázisokból dolgoznak. Ezek az adatbázisok többnyire valamilyen hullámforma vagy szövegadatbázisok, vagy például HMM szintetizátor esetében különböző paraméteradatokat tartalmaznak.

A kutatásaim ismertetését az 1.1. ábrán látható főbb komponensekhez kapcsolódóan mutatom be. A téziseket ezen területek szerint osztottam téziscsoportokra: Az I. téziscsoportban a szövegelőkészítéshez kapcsolódó ékezetesítő eljárásokat ismertetem. A II. téziscsoportban az adatbáziskészítéshez kapcsolódó eredményeimet foglalom össze. A III. és a IV. téziscsoportban a gépi beszédelőállításhoz kapcsolódó új megoldásaimat ismertetem. A kutatási témáim tehát a beszédszintézis folyamat különböző területeihez kötődnek, de mindegyik azt a célt szolgálja, hogy a bemeneti szövegből jobb minőségben, vagy kisebb erőforrással tudjuk előállítani a szintetizált beszédet.

# 2. Kutatási célkitűzések

Célom olyan új algoritmusok és módszerek kidolgozása, amelyek segítségével javítható az ember-gép közötti beszédkommunikáció minősége infokommunikációs rendszerekben.

Ennek megfelelően az egyik célkitűzésem a magyar nyelv statisztikai feltérképezése a gépi beszédszintézis szempontjainak figyelembe vételével. Ennek alapján vizsgálható többek között a más nyelvekre kidolgozott módszerek és eljárások adaptálhatósága is. Ezen a területen korábban a számítógépes technikák korlátai miatt nagyméretű adatbázisokon nem tudtak vizsgálatokat végezni. Fontosnak tar-

tom azt, hogy az új feldolgozási lehetőségekkel a korábbi kisebb adatbázisokon meghatározott adatokat ellenőrizzem és új összefüggéseket tárjak fel.

Másik célkitűzésem, hogy a korábbi beszédszintézis módszereket új eredményekkel egészítsem ki, amelyek tovább szélesítik a felhasználási lehetőségeket. Célom, hogy a szintetizált beszéd hangzása egyre jobban hasonlítson az emberi beszédhez, a robotos, gépi jelleg minél kevésbé legyen érezhető. A gépi beszédszintézist új attribútumokkal kívánom bővíteni, mint például az érzelmi töltetű beszédelőállítás módozataival.

A kutatásaim során az eljárások nagy részét úgy dolgoztam ki, hogy magyar nyelven igazoltam működőképességüket, de mindig szem előtt tartottam azt, hogy az eredményeket nemzetközileg is hasznosítani lehessen. Külön figyelmet fordítottam arra, hogy a megoldások nyelvfüggő részei elkülöníthetőek, vagy könnyen adaptálhatóak legyenek más nyelvekre.

A kutatási céljaim elérése során mindig az infokommunikációs rendszerek adta lehetőségeket és korlátokat is figyelembe veszem. Az infokommunikációs rendszerek, mint például a széles körben alkalmazott interaktív hangválasz (angol terminológiában Interactive Voice Response, IVR) rendszerek megkívánják a valós idejű, skálázható megoldásokat. Az erőforrások sok esetben korlátozottak, figyelmet kell fordítani arra, hogy akár 240 párhuzamos csatornát is ki tudjon egyidejűleg szolgálni egyetlen PC.

## 3. Módszertan

A értekezésben bemutatott kutatások során alapvetően a beszédtechnológiában és a kapcsolódó tudományágakban elterjedt módszereket használtam.

A kutatásom első fázisában felhasznált szövegadatbázisok az internetről gyűjtöttem. A későbbi során a Magyar Nemzeti Szövegtárat is használtam, amely lehetőségeihez mérten reprezentatívan tartalmazza a mai magyar nyelv jellegzetes megnyilvánulásait. A feldolgozáshoz és elemzéshez többnyire általános rendezési és osztályozási algoritmusokat használtam.

A beszédadatbázisokon végzett méresekhez nem állt rendelkezésemre nagyméretű magyar nyelvű olvasott beszédkorpusz. Ezért ezeket a Beszédlaboratórium különböző kutatási és fejlesztési munkái során készített hangfelvételekből állítottam össze. Beszédadatbázisok és hullámformák vizsgálatára a Praat szoftver (Boersma–Weenink 2010) függvényeit használtam.

Az algoritmusok és eljárások megvalósításához a MATLAB és a RapidMiner eszköztárát alkalmaztam, illetve szükség esetén C nyelven implementáltam, amennyiben a korlátozott erőforrás ezt indokolttá tette.

# 4. Új eredmények

## 4.1. I. téziscsoport: Eljárások ékezet nélkül írt szövegek automatikus helyreállítására

A gépi beszédfeldolgozásban és főként a beszédszintézisben sokszor előfordul, hogy olyan bemeneti szövegeket kell feldolgozni, amelyek részben vagy teljesen ékezetmentesek. Az ékezetmentes szövegek előállhatnak úgy, hogy a felhasználó nem használ ékezetes betűket, mert vagy kényelmetlen részére, vagy az adott eszköz, konfiguráció nem biztosít számára megfelelő beviteli lehetőséget. A másik eset az, amikor a tárolás vagy adatátvitel során lép fel veszteség, bizonyos ékezetes betűk a konverziók során ékezet nélkülivé válnak.

Példa:

```
Agyunk a beszedet nem onmagaban dolgozza fel,
hanem az osszes erzekszervunkbol kapott informaciot
kombinalja es ertelmezi.
```

Az ember számára az ékezet nélküli szövegek elolvasása nem jelent különösebb nehézséget, olvasáskor a szöveg eredeti tartalmát szinte teljesen vissza tudja állítani az agy. Ilyenkor a vizuális észlelésünk a nyelvi tudásunkra alapozva kikövetkezteti a hiányzó ékezeteket. Az ékezet nélküli, géppel felolvasott magyar szövegeket hallás után azonban már nem képes könnyen feldolgozni és megérteni a percepciós rendszerünk, mivel az ékezetek elhagyása a legtöbb esetben hangváltozást is eredményez. torzított hangtestet nem értjük meg, vagy félreértjük. Például a „*mögött*” szó helyett a „*mogott*” elhangzása értelmetlen. Az „*ágyát*” helyett az „*agyat*” szó pedig más jelentése miatt okozhat zavart a megértésben. Ezért szükséges szövegek felolvasása előtt ellenőrizni az ékezetek meglétét, illetve hiányát, majd elvégezni az újra ékezetesítést.

Az ékezetek helyreállításának problémaköre nem csak a magyar nyelvet érinti, más ékezeteket használó nyelv esetében is fennáll a jelenség kezelésének szükségszerűsége (Mihalcea–Nastase 2002). Nemzetközi szinten más ékezeteket használó nyelvekre születtek publikációk (Mihalcea–Nastase 2002, De Pauw et al. 2007, Ungurean et al. 2008). Leginkább román területen kutatták ezt a problémát. A publikált megoldások általában gépi tanuláson alapultak, amik általános megoldást adnak, azonban nem veszik figyelembe az infokommunkációs rendszerek és a beszédszintézis követelményeit. Például nem vizsgálták azt, hogy az ismertetett eljárások milyen jellegű hibákat okoznak szószinten, – például nyelvtanilag helytelen szót hoz létre, – amely a későbbiek során a szintetizált beszéd érthetőségét befolyásolhatja.

### *I.1. tézis: Szótár alapú eljárást dolgoztam ki ékezet nélkül írt elektronikus szövegek ékezetes formájának visszaállítására. [P1,B2,J1,J2,J3,C1]*

A szótár alapú megoldás lényege, hogy nagy szövegadatbázis alapján szótárat építünk, amely tartalmazza a különböző ékezet nélküli szóalakokat és a hozzájuk tartozó, nyelvileg korrekt ékezetes formákat. A szótár építésénél használjuk a gyakorisági adatokat is. Ahol egy ékezet nélküli szónak több lehetséges szóalakja is előfordulhat, ott a statisztikailag legvalószínűbbet választja az algoritmus. A módszert magyar nyelvre ellenőriztem, amellyel 95%-os szópontosság érhető el elektronikus levelek szövegeiben [C1], 96%-os pontosság általános témakörben [C10]. Az eljárás magyar szabadalom [P1], lajstromszáma: 226740 P 00 03443.

A megoldás előnye, hogy a szótárépítésben szereplő lexikai egységekre nyelvtanilag helyes alakokat ad. Az algoritmus az előkészítő fázisban nagyméretű szövegadatbázist használ, de a működéshez kialakított tudástár (szótár) már kis méretű és gyors keresést biztosít. Hátránya, hogy nincs általánosító képessége, a szótáron kívüli elemeket nem képes ékezetesíteni. Ha olyan szót kell ékezetesíteni, amely a statisztikák készítésénél használt szövegadatbázisban nem szerepelt, akkor adat hiányában az ékezet nélküli verziót fogja meghagyni, amely értelmetlen szó is lehet.

### *I.2. tézis: Eljárást dolgoztam ki a szótár alapú algoritmus hibájának csökkentésére[C10]*

Az I.1. tézisben ismertetett eljárás a kétes eseteket (szavakat) nem kezeli, mindenkor a leggyakoribb esetre dönt. Kétes esetnek nevezzük azokat az alakokat, amelyeknél több helyes forma is ugyanahhoz az ékezet nélküli szóhoz tartozik, mint például a „*veres - verés - véres*”. Ezt a hiányosságot olyan eljárással küszöböltem ki, amely a leggyakoribb kétes eseteket egy döntési fa segítségével egyértelműsíti, a környezet alapján hoz döntést a lehetséges variációk között. Az eljárásom a Mihalcea–Nastase (2002) módszerét fejleszti tovább, ami gépi tanulást használ fel karakter szinten. Ezt adaptáltam a szó alapú feldolgozásra a következők szerint:

A kétes esetek listája és azok gyakorisága rendelkezésre áll az I.1. tézis eljárás eredményeként. Ezekre a szavakra egy J48-as döntési fát építettem, amely a szó környezete alapján egyértelműsíti azt, hogy a több lehetséges ékezetes alak közül melyik a legvalószínűbb. A J48-as döntési fa a C4.5 szabadon felhasználható java alapú implementációja (Quinlan 1993).

Az algoritmus illeszkedik az I.1. tézis eljárásához, mert a nagy számításigényű feldolgozási műveleteket külön választottam a gyors alkalmazási eljárástól. A szótárépítés memória igényes, és nagyméretű szöveg – több 10 millió szövegszó – feldolgozása szükséges. Az alkalmazási fázisban viszont már csak egy kisméretű döntési fa bejárása szükséges, ami gyorsan elvégezhető. Az algoritmus egyik korlátját az jelenti, hogy a tanításhoz rendelkezésre kell állnia nagyméretű tanító

adatnak minden egyes kétes esetre. Az algoritmus előnye, hogy az I.1.-es tézis eljárásához képest már általánosító képességgel is bír a kezelt kétes esetekre, mert olyan esetekben is döntést tud hozni, amilyen környezetre a tanító adatok nem állnak rendelkezésre.

Ezzel a hiba csökkentő eljárással a Magyar Nemzeti Szövegtár anyagával tanított algoritmus a kétes esetek 60%-át kezeli és ezeket 93%-ban helyesen ékezetesítette [C10]. A javulás az összes hibához képest kevés, de a magyar nyelv esetén a beszédértést nehezítő eseteket kezeli, például: „*Megvetette az ágyát.*" - „*Megvetette az agyát.*"

Az I.1. tézis algoritmusa továbbá nem kezeli azokat az eseteket, amelyek nem szerepelnek az I.1. tézis algoritmusának szótárában. Az ilyen esetek kezelésére döntés fát építettem, amely a karakterkörnyezet alapján működik.

A J48-as döntési fa építése a lehetséges ékezetesítendő karakterek 20 méretű környezete alapján történik. Magyar nyelvben csak a magánhangzókat ékezetesítjük, betűjelük: „*a, e, i, o, u*". Az így kapott döntési fákat – magyar esetében 5 db-ot – azoknak a szavaknak a betűire alkalmazom, amelyeket korábban ismertetett algoritmusok nem kezeltek.

A döntési fa alapú algoritmus általánosító képessége nagy, bármilyen szóra alkalmazható, olyanra is, amely a tanító adatbázisban nem szerepelt. Az I.2. tézis eljárásai az I.1.-es tézis eljárásának hibáit 60%-ban javította általános szövegkörnyezet esetében [C10]. Az eddig nem kezelt kétes eseteket 83%-os pontossággal ékezetesítette [C10].

## 4.2. II. téziscsoport: Magyar és idegen nyelvű írott szövegek vizsgálata és összehasonlító eljárásai elsősorban gépi beszédkeltés támogatására

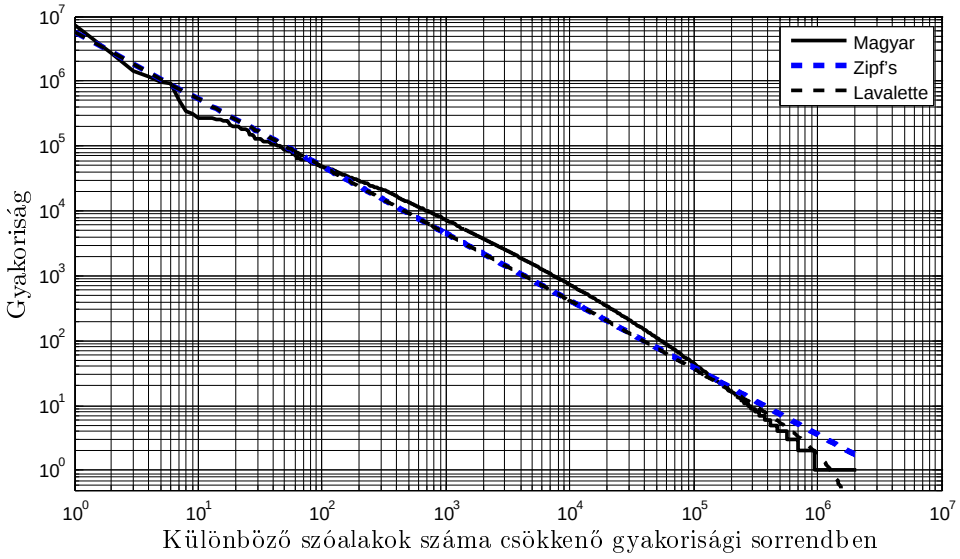
A gépi szövegfelolvasás bemenete a legtöbb esetben az írott szöveg. A jó minőségű beszédszintetizáláshoz a bemeneti szöveg tulajdonságainak ismerete szükséges. Az írott szövegek statisztikai adatai felhasználhatóak beszédszintetizátorok alapját képező beszédadatbázisok előállításához, a szintetizálási algoritmusok teszteléséhez is. Több nyelv adatainak összehasonlítása pedig segítséget ad az idegen nyelvű algoritmusok magyarra történő adaptálásához és az eredmények nemzetközi összehasonlításához.

### *II.1. tézis: Megállapítottam az angol, a német és a magyar nyelvű szó alapú gépi beszédtechnológiai módszerek összehasonlításhoz szükséges alapvető alkalmazhatósági sarokpontokat[B3,J4,J5,C2,C3]*

A magyar nyelv ragozó tulajdonsága miatt a nyelvtanilag helyes és értelmes szóalakok száma rendkívül nagy, különböző becslések milliárdos nagyságrendet határoznak meg (1 millió lexéma (Kenesei et al. 1984) és például egy igének 1000

ragozott alakja is lehet (Prószéky 1988). A valóságban használt szavak száma ennél kisebb, illetve különböző idegen eredetű szavakat is használunk a nyelvben. A használt szóalakok számának és azok eloszlásának meghatározásához a következő források szövegeit használtam fel: A Magyar Elektronikus Könyvtár magyar nyelvű művei, a Digitális Irodalmi Akadémia művei, online folyóiratok cikkei és a Magyar Nemzeti Szövegtár. A források egyesítve 80 milliós szövegszavas gyűjteményt tesznek ki.

A magyar nyelvre megállapítottam a szavak gyakorisági sorrendjét és azt, hogy az adott számú leggyakoribb szó a korpusz mekkora részét fedi le. A szó gyakoriságok elemzésének eredménye a 4.1. ábrán látható folyamatos vastag vonallal jelölve.




*4.1. ábra.* Összefüggés a szavak gyakorisága és a sorrendjük között

A gyakorisági adatokat összevetettem – nyelvek vizsgálatakor gyakran alkalmazott – Zipf törvényével (Li 1992), amelyet az ábrán szintén jelöltem vastag szaggatott vonallal. A gyakorisági adatok követték Zipf törvényét ($f(r) = C \cdot r^{-b}$) ahol $C$ egy normalizációs konstans, a $b$ általában 1 körüli érték. Az ábrán mindkét tengely logaritmikus beosztású, így az exponenciális lefutású görbe egyenesként ábrázolható. A gyakorisági adatokra illesztett függvény konstansai a következők: $C = 10^{6.7573}$, $b = 1.033$. A Zipf által meghatározott görbe a kis gyakoriságú elemek esetén eltér a mért adatoktól, ennek kiküszöbölésére a Lavalette formulát alkalmaztam (Popescu 2003), amely a 4.1. egyenlettel írható le. Az alkalmazott konstansok azonosak a Zipf formula konstansaival.

$$f(r) = C \cdot \left( \frac{r_{max} r}{r_{max} - r + 1} \right)^{-b} \tag{4.1}$$

A meghatározott Lavalette görbe szintén a 4.1. ábrán látható vékony szagatott vonallal.

A három nyelv összehasonlításához különböző szövegkorpuszokat használtam fel. Az angol esetében a British National Corpus (BNC) 89 millió szövegszót tartalmazó verzióját és egy gyűjtést használtam fel, amely a Magyar Elektronikus Könyvtár angol nyelvű szövegeiből állt. A német nyelv esetén a Gutenberg projekt anyagát használtam. A magyar nyelvű adatok a már elemzett 80 milliós szövegkorpuszból származnak. Továbbá mindhárom nyelven rendelkezésre állt a Biblia fordítása. Ez utóbbi vizsgálatát külön végeztem, mivel tartalmi és méreti szempontból egységesek voltak.

Megállapítottam, hogy a szavak gyakorisági sorrendben vizsgálva a korpusz mekkora részét fedik le. A 4.1. táblázatban látható három kiragadott példa, amely megadja, hogy az adott fedési hányadhoz a három nyelv esetében hány szó szükséges.

*4.1. táblázat.* A szükséges leggyakoribb szavak száma a korpusz fedésének függvényében

| Nyelv | Fedés | | |
|---|---|---|---|
| | 75 % | 90% | 97,5% |
| Angol | 1250 | 5800 | 20 100 |
| Német | 2000 | 14 550 | 80 000 |
| Magyar | 10 650 | 70 000 | 400 000 |

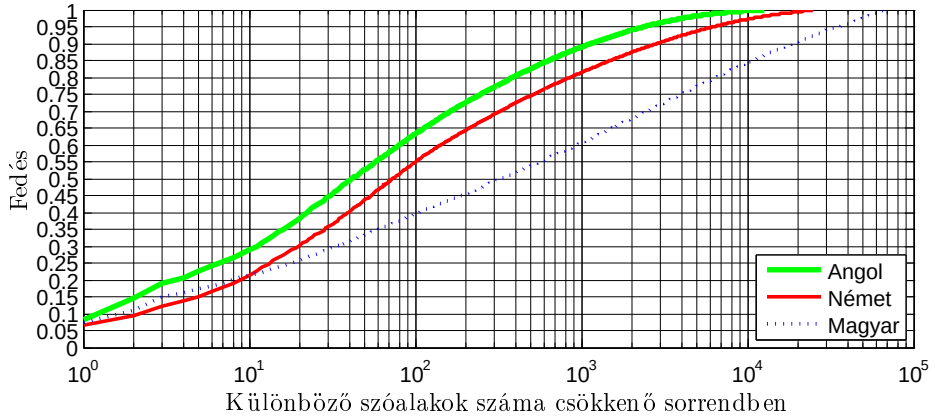
A kapott adatok elemzésének másik lehetséges formája, amikor az adott szószám alapján hasonlítjuk össze a különböző nyelvek fedési adatait. Erre láthatunk példát a 4.2. táblázatban.

*4.2. táblázat.* Korpusz fedés adott szószám mellett

| Nyelv | Leggyakoribb szavak száma | | |
|---|---|---|---|
| | 1000 | 20 000 | 100 000 |
| Angol | 72,8% | 97,5% | (100%) |
| Német | 69,1% | 91,8% | 98,1% |
| Magyar | 51,8% | 80,7% | 92,0% |

A részletes ábrák és adatok a [J5]-ben és a disszertációban találhatók.

Az összehasonlítást elvégeztem azonos tematikájú szövegeken is, a Biblia három változatában. A kapott fedési görbék a 4.2. ábrán láthatók. A konkrét gyakorisági számok eltérőek a korábbi nagy vizsgálati korpuszok számaitól, de a nyelvek közötti arány megmaradt. A magyar adatokat pontsor, az angolt a vastag, a németet pedig a vékony vonal jelzi. Például a nagy korpuszok 4.1. táblázatban látható 90%-os fedését összevetve a 4.2. ábra 90%-os vonalával, látható, hogy az angol és a német nyelv közötti háromszoros arány megegyezik, és az angol és a magyar közötti 1 nagyságrendi különbség is.




*4.2. ábra.* A 3 különböző nyelvű Biblia fedési görbéi

A módszeremmel és a publikált adatokkal tehát szó alapú algoritmusok nyelvek közötti átültethetőségére lehet vizsgálatokat végezni. A vizsgálati módszerem nem csak a bemutatott három nyelvre alkalmazható, más nyelvek közötti összefüggések kimutatására is használható.

### *II.2. tézis: A betű fogalmának kiterjesztésével új módszert dolgoztam ki szövegek beszédszintézis szempontjait figyelembe vevő minősítéséhez [C8]*

A nyelv írásos formája (betűkép) és a hangalak (kiejtés) szoros összefüggésben van egymással. A számítógépes nyelv- és beszédfeldolgozás felszínre hozta azt az igényt, hogy a statisztikai elemzéseknél vegyük figyelembe a két szint egymásra hatását is, hiszen egymásból következnek. Ez újfajta megközelítést igényel, olyat, amely alapjaiban kapcsolódik a szóstatisztikai adatokhoz, továbbá a szavakat felépítő betűk statisztikai feldolgozásához. Az is figyelembe veendő, hogy a betűkép milyen hang-szintű információkat tartalmaz. Teljes képet a nyelv statisztikai jellemzőiről csak akkor kaphatunk, ha mind a szövegszintű elemek, mind az elhangzó hangok szintjén, ugyanarra a nagyméretű nyelvi anyagra végzünk méréseket. Az osztályozáskor figyelembe vettem, a betűsorozathoz rendelhető hangalaki reprezentációt is. A mérésekhez gépi gyűjtő és szortírozó algoritmusokat készítettem, kifejezetten ehhez a kutatáshoz.

A betű fogalmát kiterjesztettem és a célkitűzéshez alakítottam. A karakter- és betűstatisztikához a vizsgált leghosszabb betűsorozat a szó volt, a nem betű típusú karaktereket figyelmen kívül hagytam (számok, relációs jelek stb.). A betű fogalmának kiterjesztése azt jelentette, hogy a beszédhang oldaláról is visszavetítettem elemeket az írás szintjére. Például a *pech* szó klasszikus értelemben vett *ch* betűkapcsolatát az *sz* betűhöz hasonlóan kezeltem, két karakterből álló betűnek tekintettem, ugyanakkor másként kezeltem például a *lánchíd ch* betűkapcsolatától, ahol külön *c* és *h* betű szerepel. Az új betű szintű osztályozás miatt megmaradnak olyan információk is, amelyek a fonetikus átírás közben elvesznek. Például ren-

delkezésre áll az új típusú betűstatisztikában a [j̯] hangként kimondott *j* és *ly* betű, vagy az [i] hangként kimondott *i* és történelmi nevek végén gyakran szereplő *y* betű.

Fontos megjegyezni, hogy ezek az osztályozások elsősorban beszédtechnológiai szempontok figyelembevételével történnek, nyelvészeti vonatkozásban bizonyos döntések hiányosnak tűnhetnek. A hangjelölések megállapítására és osztályozására az elválasztási szabályokra épített algoritmust használtuk (*Ri-chárd*, *Mün-chen*, *Ben- czúr)*, miszerint ezek a betűk nem elválaszthatóak. A döntéseket a magyar elválasztási minta-gyűjtemény szószedete (Nagy 2008) alapján hoztuk meg. Az algoritmus figyelembe veszi a két karakterből álló betűket is ( *gy, ty, ny, sz, zs, cs*), azok hosszú változatával egyetemben. A hosszú változatokat két betűnek tekintettük (*zzs*= *zs* + *zs*) a statisztikai feldolgozás során.

A szövegekből előállítottuk a hangalakot (hangszimbólumok írott sorozatát) beszédtechnológiai gépi módszerek alkalmazásával. Három eszközt használtunk, egy szabály alapú algoritmust (a Profivox szövegfelolvasó rendszer fonetikai átíróját és szabálygyűjteményét (Olaszy et al. 2000)), a magyar elektronikus kiejtési szótárt (Abari–Olaszy 2006) és a névmondó tulajdonnév kiejtési gyűjteményt [C4]. A kiejtési forma meghatározásához kialakított szabályok a magyar nyelvi normát képviselik.

A módszer segítségével megvizsgáltam a Magyar Nemzeti Szövegtár 2006-ös verziójának anyagát, amelynek részletes szöveg és hangalak statisztikai elemzése a [C8]-ban és a disszertációban található.

## 4.3. III. téziscsoport: Eljárások beszédszintetizátorok hangminőségének javításához

A beszédszintetizátorok által előállított beszéd minősége függ az egyes komponensek minőségétől, illetve attól, hogy a felolvasni kívánt szöveg mennyire korlátozott tematikájú. Általánosságban jellemző a gépi beszédszintézisre, hogy a tematika és a minőség szorzata állandó, tehát szűkebb témakörben jobb, tágabb témakörben kevésbe jó minőségű beszéd állítható elő. Ebben a téziscsoportban a gépi-beszédelőállítással kapcsolatos olyan eredményeimet foglalom össze, amelyek a szintetizált beszéd minőségének javítását eredményezte olyan környezetben, ahol a tematika behatárolt, de mégis nagy változatossággal bír.

### *III.1. tézis: Eljárásokat és algoritmusokat dolgoztam ki magyar tulajdonnevek, cégnevek és magyarországi címek gépi felolvasásához. [B1,C4,B5a]*

A név- és címfelolvasás témaköre az általános szövegek felolvasásához képest szűkebb terület, de a nevek természetéből adódóan nem korlátos. A folyamato-

san megjelenő idegen eredetű személynevek, illetve a nyelvi határokat figyelembe nem vevő cégnevek megkívánják a tetszőleges betűsorozat felolvasásának képességét. Azonban kihasználható, hogy a felolvasandó részek tartalmaznak különböző gyakran előforduló részeket. Magyar nyelvre ilyen témában nem született más publikált megoldás és idegen nyelvre is csak részproblémákkal foglalkoztak később érintőlegesen (Aylett 2004).

A név- és címfelolvasásban kombináltam a triád alapú beszédszintézist [J7], a számfelolvasást (Olaszy–Németh 1999) és a szótár alapú szintetizálást. Az eljárás lényege, hogy a szintetizálás során a különböző típusú elemek csak meghatározott pozícióban szerepelhetnek a hangsorozatban, így a prozódiájuk előre meghatározható és rögzíthető. Az előállított mondatban tehát vegyesen szerepelnek a triádos rendszer elemei, a számelemek és az külön felolvasott elemek. A felolvasandó adatok elemzése során a 4.3. táblázatban szereplő kategóriákat azonosítottam és vizsgáltam meg részletesebben.

Az elemzést egy 3 millió rekordot tartalmazó név- és címlistán végeztem, a II.1. tézisben használt módszerrel. A felsorolt különböző kategóriákban szereplő szavakra meghatároztam a gyakorisági sorrendet. A kategóriákban előforduló összes szó megfelelő minőségű felolvasása nem lehetséges, mert a bemondó képtelen egyenletes stílusban és hangon ilyen mennyiségű szó felolvasására. A szavak számát úgy szűkítettem, hogy a felolvasandó lista ne legyen nagyobb, mint amit egy képzett bemondó egy alkalommal fel tud olvasni (4 óra felvétel). A 4.3. táblázatban szereplő második oszlop az elemzés során talált szavak számát, az utolsó oszlop a felolvasott szavak számát jelenti. A szükséges felolvasandó szavak számát a fedési görbe segítségével határoztam meg, a cél a 95%-nál nagyobb fedés elérése volt, de azzal a kikötéssel, hogy ne legyen 1000 tételnél több (felvételi korlátok miatt).

*4.3. táblázat.* Vizsgált kategóriák és tulajdonságaik

| Kategória | Különböző szavak száma | Felolvasott elemek száma |
|---|---|---|
| Vezetéknevek | 103850 | 0 |
| Keresztnevek | 1797 | 313 |
| Cégformák | 8 | 8 |
| Településnevek | 3523 | 1000 |
| Közterület fajták | 14 | 14 |
| Betűzés | 36 | 36 |

A gépi felolvasás alkalmazásához meghatároztam azokat a szövegfeldolgozó szabályokat, amelyek a leggyakrabban előforduló szövegbemenetekben beazonosítják az egyes információs részeket, és a megfelelő sorrendben adják tovább a hullámforma előállító alrendszernek. Az információs részek a következők:

**Tulajdonnevek:** titulus, vezetéknév, keresztnév
**Cégnevek:** cégnév, cégforma
**Címek:** irányítószám, település neve, településrész neve, közterület neve, közterület fajtája, házszám, lépcsőház, emelet, ajtó

Az információs elemek azonosítása a felolvasandó rekordban a következő három paraméter szerint történt: az elem pozíciója a rekordban, elemszótárban való szereplés, már azonosított információs elemekhez viszonyított pozíciója.

Az információs elemeket az előbbi felsorolás szerint rendeztem, majd az elemhez tartozó hullámformák egymáshoz lettek illesztve (triádos szintetizált hullámforma, számfelolvasóval generált számok és külön felolvasott elemek). A prozódia alapfrekvencia komponensének változtatására a triádos hullámforma összefűzéses szintetizátor esetében van lehetőség. A számelemek 2 fajta prozódiával állíthatók elő, a felolvasott elemek pedig a tervezett prozódia szerint lettek rögzítve. A prozódia megvalósítása a kijelentő mondatokra jellemző dallammenetet követi. A prozódia intenzitás komponenseként egy egyszerűsített menetet használtam, amely csak az utolsó elem esetében ír elő csökkentést, a többi elemet azonos szintre egyenlíti ki. A prozódia idő komponenséből csak a szünetezés változtatására van lehetőség, amelyet az érthetőség figyelembevételével határoztam meg. A szünetek a felolvasás tagolási pontjaira kerültek, a szünetek hosszát percepciós kísérletekkel pontosítottam. Az név- és címfelolvasó eljárás eredményességét érthetőségi teszttel bizonyítottam.

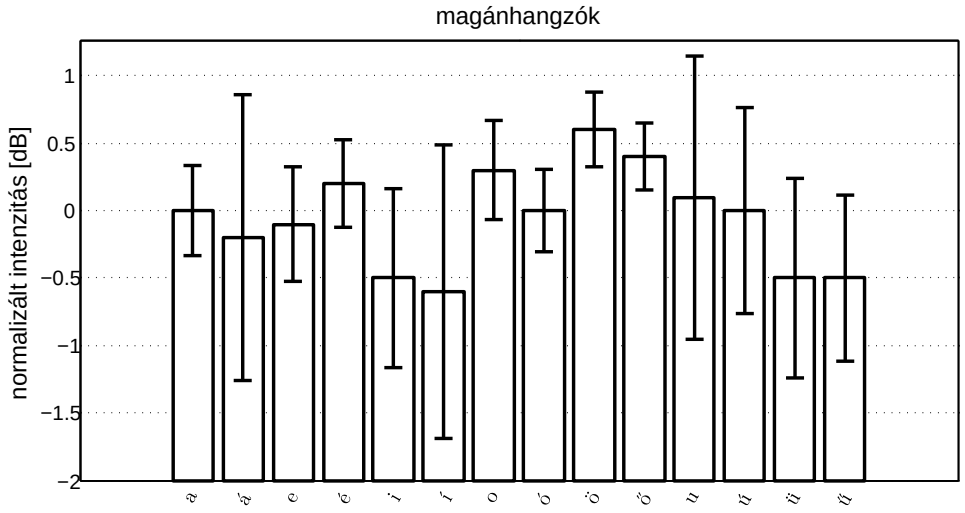
### *III.2. tézis: Virtuális szóintenzitáson alapuló eljárást dolgoztam ki korpusz alapú szintetizátor beszédének intenzitás-kiegyenlítéséhez. [B5b,J6]*

A magyar beszédhangok intenzitásviszonyaira vonatkozó korábbi mérések csak korlátozottan állnak rendelkezésre (Olaszy 1989), nagy mennyiségű adat feldolgozásáról irodalom nem található. A beszédhangok intenzitás viszonyát eddig csak példamondatokon vizsgálták.

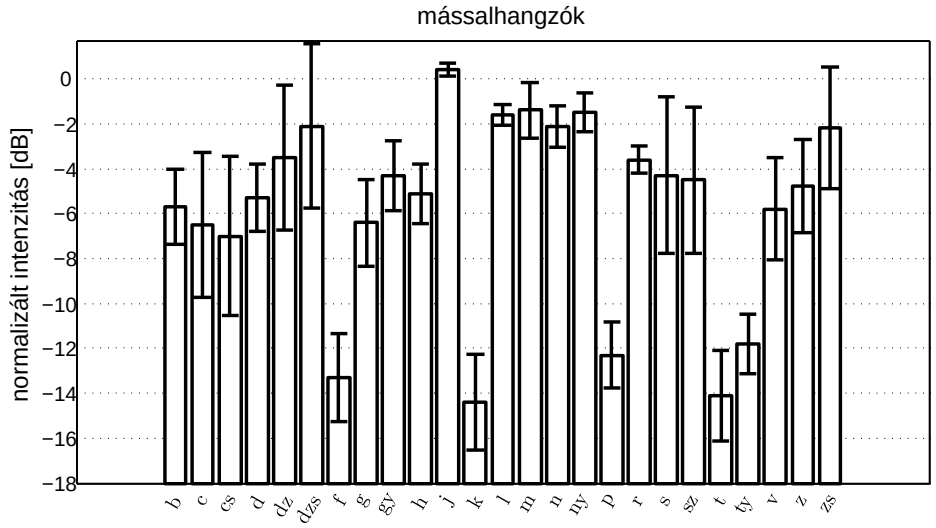
Vizsgálataimban saját gyűjtésű beszédadatbázisokat használtam fel, amelyek stúdiókban készültek [C12], így a felvételek jel/zaj viszonya 40-60 dB nagyságú, amely nem befolyásolja a mérést. A beszédadatbázis készítése során a felvételekből a hibás, nem az adatbázishoz tartozó vagy rontott részek eltávolításra kerültek. A beszédadatbázisban a hanghatárokat félautomatikus [C6,C7] eljárással határoztam meg, és különböző algoritmusok segítségével a jelentős hibák korrigálásra kerültek [J10]. A felvételekben szereplő mondatok átlagosan 15 szóból álltak, így az effektív érték (RMS) alapú mondat szintű intenzitáskiegyenlítés elvégezhető volt.

A mérést 9 beszélő (2 nő és 7 férfi) adatbázisán végeztem el, amelyek összesen 57 óra beszédet tartalmaztak. A címkézett adatbázison az intenzitásviszonyokat gépileg lemértem, az összesített eredmények a 4.3. és a 4.4. ábrán láthatók. Az intenzitást hangonként határoztam meg a hanghatárok között RMS segítségével. Az ábrákon az értékeket az *a* hang intenzitásához normalizáltam. A könnyebb olvashatóság érdekében a hangokat a hozzájuk tartozó betűképpel jelöltem. Az oszlopok tetején az egyes beszélők közötti szórás mértékét tüntettem fel.

Ezek az eredmények eltérnek Olaszy (1989) munkájától, a kapott értékek kiegyenlítettebbek. A hangok intenzitás térképei az átlagos intenzitásokon alapul-




*4.3. ábra.* A magánhangzók intenzitásátlagai több beszélő folyamatos, felolvasott beszédében




*4.4. ábra.* A mássalhangzók intenzitásátlagai több beszélő folyamatos, felolvasott beszédében

nak, amely lehetőséget ad a mondatprozódia során megvalósuló intenzitás eltérések további kutatására.

A beszédszintetizátorok hangadatbázisának készítése összetett folyamat. A korpusz alapú szintetizátorok beszédadatbázisának mérete a néhány órától több száz óráig is terjedhet. Ilyen méretű adatbázisok csak több lépésben, több hét vagy hónap alatt rögzíthetők. A felvételek intenzitás kiegyenlítése szükséges, amely részben a felvételi körülmények kontrollálásával, illetve utólagos korrekciókkal részlegesen megoldható. Amennyiben a beszédadatbázis inhomogén – tartalmaz kisméretű önállóan felolvasott elemeket, például csak szavakat vagy csak számokat – akkor ezek kiegyenlítése összetett feladat. A kiegyenlítés már nem valósítható meg a jelfeldolgozó rendszerek általános normalizációs algoritmusaival –

például úgy, mint a méréseknél alkalmazott RMS normalizálással – mert az elvileg kiegyenlített elemek hangsorba illesztésekor az intenzitás viszonyok eltérőek lesznek. Ha a rövid elem olyan hangokból épül fel, amelyek az átlagosnál gyengébb intenzitásúak, akkor a teljes szóra végzett kiegyenlítés hibás – intenzívebb – eredményt okoz. Az ilyen elemekből származó hangok máshova történő beillesztése intenzitásugráshoz vezet.

A korpuszos szintetizátorok hang méretű elemekből is építkezhetnek (Taylor 2009), amelyek intenzitásai ingadoznak. Az ilyen építőelemek kiegyenlítéséhez bevezetem a virtuális szóintenzitást, amely megadja, hogy az a szó amelyben a hang szerepel, milyen intenzitású lenne, ha minden hangja az ismertetett mérésekkel megállapított átlagos intenzitású lenne. A kiszámításnál a hangok időtartama szerinti súlyozott átlagot számolom ki a következő képlet szerint, ahol $N$ a szóban szereplő hangok száma, $t_{ph}$ a hang időtartama, $I_{ph}^{average}(ph)$ az ismertetett mérésben kiszámított átlag és a $ph(i)$ a szó $i$.-dik hangja:

$$I_{word}^{virtual} = \frac{\sum_{i=1}^{N} t_{ph}(i) I_{ph}^{average}(ph(i))}{\sum_{i=1}^{N} t_{ph}(i)} \tag{4.2}$$

Ezek alapján meghatározható az adott beszédhang módosításának a mértéke:

$$gain_{ph} = \frac{gain_{prosody} \cdot gain_{base} \cdot I_{word}^{virtual}}{I_{word}^{real}} \tag{4.3}$$

A $gain_{base}$ adja a mondat átlagos intenzitását, a $gain_{prosody}$ pedig a prozódia által meghatározott eltérést. Az $I_{word}^{real}(ph)$ adja meg annak a szónak intenzitását, amelyben a kiegyenlítendő hang szerepel. Az intenzitás kiegyenlítés módszerét meghallgatásos teszttel vizsgáltam.

A minőség értékelésénél a kiegyenlítés nélkülihez képest szignifikánsan ($p<0.05$) jobb volt a másik két módszer, de a kiegyenlítettek között szignifikáns különbség nem volt kimutatható. A páros összehasonlítás esetén is a másik két eljárás szignifikánsan jobb volt a kiegyenlítés nélküli módszernél. A két módszer között két teszthalmaznál nem volt szignifikáns különbség, egy teszthalmaznál a tézisben ismertetett módszer szignifikánsan jobb volt.

## 4.4. IV. téziscsoport: Eljárások beszédszintetizátorok hangjának érzelmi módosításhoz

A beszédszintetizátorok készítése során az érthető, jó minőségű beszéd előállítása az alapvető cél. A szintetizátorok többnyire csak semleges érzelmet tükröző beszédet állítanak elő, mert a kiinduló beszédadatbázisok semleges érzelmű olvasott beszédet tartalmaznak. Az érzelmi töltetű beszéd előállításának egyik lehetséges

útja az, amikor érzelmi töltetű beszédadatbázist készítünk (Douglas-Cowie et al. 2003). Ez idő és erőforrás-igényes feladat, ezért ilyen adatbázis sok esetben nem áll rendelkezésre. Egy másik út az, hogy a beszédet olyan transzformációnak vetjük alá, amely az érzelmi töltetűhöz hasonló beszédet állít elő. Az eljárás nem biztosít azonos minőséget, de lehetőséget ad a már meglévő beszédadatbázisok felhasználására.

A IV. téziscsoportban az érzelmi töltetű beszédszintézishez kapcsolódó új eredményeimet ismertetem.

### *IV.1. tézis: Eljárást dolgoztam ki érzelmi töltetű beszéd közvetlenül a hullámforma transzformációjával történő megvalósításához. [B4,C5,C9]*

A Přibilová–Přibil (2009) által ismertetett eljárást fejlesztettem tovább és kidolgoztam azt az algoritmust, amely tetszőleges beszédre alkalmazható. Az eljárás alapját egy spektrális transzformáció képezi, amely azon a megfigyelésen alapul, hogy a különböző érzelmű beszédben az alacsony és a magas spektrális komponensek aránya, továbbá az F1 és az F2-es formánsok távolsága megváltozik (Scherer 2003). Az eredeti módszert Přibilová és Přibil egy LPC alapú szintetizátorra dolgozta ki és alkalmazta cseh nyelvre.

A továbbfejlesztett eljárás a PSOLA algoritmust kombinálja Přibilová és Přibil módszerével. Az időtartománybeli jelet aszimmetrikus Hann ablak segítségével zönge-szinkron ablakozom, majd az ablakozott jelet DFT transzformációval átalakítom frekvenciatartományba. A frekvenciatartományban elvégzem a – disszertációban részletesen ismertetett – spektrális transzformációt, illetve az intenzitások beállítását. A transzformáció az ablakozott jelben olyan torzítást okoz, hogy az ablakok szélén a jel nem tart a nullához, így ismételt ablakozással korrigálom a torzítást. Az így módosított spektrumot inverz DFT-vel visszaalakítom időtartománybeli jellé, majd a PSOLA algoritmushoz hasonlóan átlapolva összeadom, amely közben a szükséges időtartam korrekciókat is elvégzem.

Az érzelmi módosítás esetén szükséges az alapfrekvencia módosítása is, amely az átlapolások eltolásával – ahogy a PSOLA esetén – megvalósítható. A zöngétlen szakaszok esetében két megoldás alkalmazható. Az egyik, hogy zöngétlen hangok esetén a transzformációt nem végezzük el, így az érzelmi módosítás nem teljes. Azonban mivel a zöngétlen hangok energiája kicsi és jellemzően a módosított alacsony frekvenciájú komponensek gyengék, a transzformáció hiánya percepciósan nem észrevehető. A másik lehetőség az, hogy a zöngétlen részekre egyenlő távolságra virtuális zöngeperiodus-jelölést helyezünk el, és a zöngés részekkel azonos módon kezeljük. A tesztelés során az első megoldást alkalmaztam.

Az eljárás igazolására meghallgatásos tesztet készítettem, ahol 1 természetes személy és 2 korpuszos szintetizátor által generált mondatra alkalmazott érzelmi módosítást teszteltem. A 3 különböző bemondást 3 érzelem szerint – szomorú, örömteli és haragos – módosítottam. A tesztelőknek a meghallgatott mondat érzelmi töltetét kellett meghatározniuk. A tesztben az érzelmes bemondások még a

módosítás mértéke szerint is változtak. Az eredmények a 4.4. táblázatban láthatóak. A meghallgatásos teszt részletes adatai a disszertációban olvashatóak.

*4.4. táblázat.* Az érzelmek tévesztési mátrixa

| | Felismert | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | N | A | H | S | | N | A | H | S | | N | A | H | S | |
| Tervezett | N | 82% | 0% | 14% | 6% | N | 50% | 23% | 0% | 27% | N | 77% | 0% | 18% | 5% | N=semleges |
| | A | 27% | 27% | 41% | 5% | A | 36% | 41% | 0% | 23% | A | 45% | 32% | 14% | 9% | A=haragos |
| | H | 27% | 5% | 68% | 0% | H | 40% | 23% | 14% | 23% | H | 9% | 4% | 82% | 5% | H=örömteli |
| | S | 9% | 5% | 0% | 86% | S | 14% | 5% | 0% | 81% | S | 22% | 5% | 5% | 68% | S=szomorú |
| | Korpuszos TTS-női | | | | | Korpuszos TTS-férfi | | | | | Természetes női | | | | | |

Az eredmények alapján női természetes és szintetizált mondatok esetén a szomorú és az örömteli érzelem felismerése volt a legbiztosabb, ezek az ismertetett módszerekkel tehát előállíthatóak.

### *IV.2. tézis: Adaptáltam a IV.1. tézis szerinti eljárást diád, triád alapú hullámforma összefűzéses rendszerekhez [B5b,C11]*

A IV.1. tézis eljárása alkalmazható a diád és triád alapú hullámforma összefűzéses szintetizátor (a továbbiakban röviden triádos szintetizátor) kimenetén is, de a szintetizált beszéd minősége az ismételt jelfeldolgozás miatt rosszabb lesz. Mivel a triádos szintetizátor és a IV.1. tézis eljárása is végez alapfrekvencia, intenzitás és időtartam korrekciót, ezért az eljárások egyesítése indokolt, mind a minőség mind a szintetizálás sebessége miatt.

A IV.1. tézis eljárását két részre bontottam. A spektrális összetevők módosítását a szintetizátor hullámforma-adatbázisán végeztem el. A triádos szintetizátor beszédadatbázisa diádokból és triádokból (két fél illetve egy fél egy egész és egy fél beszédhangból) áll, a spektrális módosítást egyesével kell elvégezni minden elemen. A különböző érzelmekhez külön beszédadatbázist kell készíteni. Az adatbázisok mérete mintavételi frekvenciától és elemszámtól függően 2-100 Mbyte tartományban van, tehát a megnövekedett tárigény nem jelentős a mai technológiák mellett. Az érzelem módosító eljárás másik részét azok a szabályok teszik ki, amelyek a prozódiát módosítják. Ezen szabályokat a prozódia előíró modul paraméterezésével valósítottam meg.

Az eljárást a IV.1. tézishez hasonló meghallgatásos teszttel vizsgáltam. A triádos szintetizátorhoz 3 módosított adatbázist készítettem, majd általános témájú mondatokat szintetizáltam. Az eredmények összefoglalása a 4.5. táblázatban látható. A teszt részletes adatai a disszertációban olvashatóak.

A triádos szintetizátor esetén a szomorú és az örömteli érzelem kifejezése a módszerrel megvalósítható, a haragos nem. A szintetizátor alapértelmezett semleges beállítása esetén a szintetizált mondatokat szomorú érzelműként azonosították a tesztelők.

*4.5. táblázat.* Az érzelmek tévesztési mátrixa triádos szintetizátor esetében

| | | Felismert | | | | |
|---|---|---|---|---|---|---|
| | | N | A | H | S | |
| Tervezett | N | 25% | 0% | 0% | 75% | N=semleges |
| | A | 29% | 17% | 29% | 25% | A=haragos |
| | H | 25% | 0% | 67% | 8% | H=örömteli |
| | S | 0% | 0% | 0% | 100% | S=szomorú |
| | | Triádos TTS-női | | | | |

### *IV.3. tézis: Adaptáltam a IV.1. tézis szerinti eljárást HMM elvű beszédszintetizátor rendszerhez [B5b,C11]*

A HMM elvű beszédszintetizátor működése gépi tanuláson alapszik, a különböző paraméterek közvetlen manipulálására nincs lehetőség. A HMM szintetizátor kimenetét módosító eljárás (IV.1. tézis) működőképes, de az ismételt jelfeldolgozás és feldolgozási idő növekedése miatt – hasonlóan a IV.2. tézisben szereplő triádos szintetizátorhoz – nem megfelelő módszer.

A HMM beszédszintetizátor tanító hangadatbázisai több tíz óra méretűek és a tanítási fázis is időigényes (több hét nagyságrendű). A nagy erőforrás-igényű teljes tanítás helyett a tanítás adaptációja segítségével (Tóth–Németh 2009) állítottam elő az érzelmes szintetizálási eredményt. Egy kis méretű – kb. 10 perc időtartamú – prozódiailag változatos hanganyagot módosítottam az IV.1. tézis eljárásának segítségével. A módosítás prozódiai és spektrális részét is végrehajtottam. A semleges érzelem mellett szomorú, örömteli és haragos mondatokat állítottam elő. A módosított felvételekből beszédadatbázist készítettem [J10] amelyet a meglévő tanításokhoz adaptáltam.

Az eljárást az IV.1. tézishez hasonló meghallgatásos teszttel vizsgáltam. A HMM szintetizátorhoz egy férfi hangot módosítottam a 3 érzelem szerint, majd általános témájú mondatokat szintetizáltam. Az eredmények összefoglalása a 4.6. táblázatban látható.

*4.6. táblázat.* Az érzelmek tévesztési mátrixa HMM szintetizátor esetében

| | | Felismert | | | | |
|---|---|---|---|---|---|---|
| | | N | A | H | S | |
| Tervezett | N | 71% | 4% | 8% | 17% | N=semleges |
| | A | 42% | 33% | 21% | 4% | A=haragos |
| | H | 21% | 0% | 75% | 4% | H=örömteli |
| | S | 17% | 0% | 0% | 83% | S=szomorú |
| | | HMM TTS-férfi | | | | |

A HMM szintetizátor esetén a szomorú és az örömteli érzelem kifejezése a módszerrel megvalósítható, a haragos nem.

# 5. Az eredmények alkalmazhatósága

Az eredmények nagy részét közvetlenül fel lehet használni jobb minőségű és változatosabb gépi beszédelőállításra. A statisztikai eredmények szélesebb területen is alkalmazhatók, például a beszédfelismerésben is. Az eredmények alkalmazhatóságára már a tézisek ismertetése során is példákat adtam, de a következő felsorolásban az egyes téziscsoportokra további felhasználási lehetőségeket mutatok be:

Az I. téziscsoport ékezetesítő eljárását (I.1. tézis) már alkalmaztam elektronikus levelek és SMS-ek felolvasásának segítésére. Alkalmazható továbbá a jelenleg széles körben elérhető mobil készülékekben, ahol a korlátozott beviteli lehetőségek miatt újra gyakran használnak a felhasználók ékezet nélküli szövegeket.

A II. téziscsoport eredményei széles körben alkalmazhatóak, a megállapított fedési adatok nyelvi technológiák lokalizálása során jól felhasználhatóak. A módosított betűstatisztika a gépi beszédszintézis további kutatása során, vagy tesztelési eljárások kidolgozásához ad segítséget.

A III. téziscsoport már gyakorlati környezetben is alkalmazásra került, de a nevekhez és címekhez hasonló komplexitású információk jó minőségű felolvasásához szintén felhasználható. Az intenzitás kiegyenlítés problémája a korpuszos szintetizátorok esetében szintén alkalmazásra került már, de a kapott eredmények további prozódiai kutatások kiindulópontjaként is szolgálhatnak.

A IV. téziscsoport eredményei alkalmazhatóak ember-gép interfészek kialakítása során, ahol szükséges a tartalmi információn túl érzelmi adatok átvitele is. Az érzelemi jellegű módosítás eredményei a bemutatott szintetizálási technikákon túl, további, általam még nem vizsgált szintetizálási technikákhoz is alkalmazhatók.

# 6. Hivatkozások

Abari K. – Olaszy G. (2006): Internetes beszédadatbázis a magyar mássalhangzó kapcsolódások akusztikai szerkezetének bemutatására. In Alexin Z. – Csendes D. (szerk.) *IV. Magyar Számítógépes Nyelvészeti Konferencia*. Szeged, 213–222.

Allen, J. – Hunnicutt, M. – Klatt, D. – Armstrong, R. – Pisoni, D. (1987): *From text to speech: The MITalk system*. London, Cambridge University Press.

Aylett, M. (2004): Merging data driven and rule based prosodic models for unit selection TTS. In *Fifth ISCA Workshop on Speech Synthesis*. Citeseer, 55–59.

Black, A. – Taylor, P. (1994): CHATR: a generic speech synthesis system. In *Proceedings of the 15th conference on Computational linguistics-Volume 2*. Association for Computational Linguistics, 983–986.

Black, A. – Taylor, P. – Caley, R. (2006): The Festival Speech Synthesis System, 1994–2006, Manual and source code avaliable at `http://www.cstr.ed.ac.uk/projects/festival`.

Boersma P. – Weenink D. (2010): Praat: doing phonetics by computer [Computer program]. Version 5.1.43, retrieved 4 August 2010 from http://www.praat.org/.

Burnard, L. (1995): The BNC Users Reference Guide. *British National Corpus Consortium, Oxford, May*.

De Pauw, G. – Wagacha, P. – de Schryver, G.-M. (2007): Automatic Diacritic Restoration for Resource-Scarce Languages. In *Text, Speech and Dialogue*. Lecture Notes in Computer Science sorozat, vol. 4629. Springer Berlin / Heidelberg, 170–179.

Douglas-Cowie, E. – Campbell, N. – Cowie, R. – Roach, P. (2003): Emotional Speech: Towards a New Generation of Databases. 40. évf. *Speech Communication*, 33–60.

Fant, G. (1960): *Acoustic theory of speech production*. Mouton De Gruyter.

Gordos G. – Sándor L. T. (1985): A limited vocabulary speech synthesiser terminal. In *Proc. of the Finnish-Hungarian symposium on information technology*. Helsinki, 3–10.

Gordos G. – Takács Gy. (1983): *Digitális beszédfeldolgozás*. Budapest, Műszaki Könyvkiadó.

Hallahan, W. (1995): DECtalk software: Text-to-speech technology and implementation. 7. évf. 4. sz., *Digital Technical Journal*, 5–19.

Hamon, C. – Mouline, E. – Charpentier, F. (1989): A diphone synthesis system based on time-domain prosodic modifications of speech. In *ICASSP89*. 238–241.

Homer, D. – Ries, R. – Vatkins, S. (1939): A synthetic speaker. 227. évf. *J. Franklin Institute*, 739–764.

Kempelen F. (1969): *Az emberi beszéd mechanizmusa (Mechanismus der Menschlichen Sprache)*. Budapest (Wien), 1969 (eredeti kiadás: 1791), Szépirodalmi Kiadó.

Kenesei I. – Kelemen J. – Pap M. – Pléh C. – Radics K. – Réger Z. – Rohonci K. – Szabolcsi A. (1984): *A nyelv és a nyelvek*. Akadémiai Kiadó.

Kiss G. – Olaszy G. (1984): A Hungarovox magyar nyelvű, szótár nélküli, valós idejű párbeszédes beszédszintetizáló rendszer. 2. sz., *Információ Elektronika*, 98–111.

Li, W. (1992): Random texts exhibit Zipf's-law-like word frequency distribution. 38. évf. 6. sz., *IEEE Transactions on Information Theory*, 1842–1845.

MBROLA, s. (2006): The Festival Speech Synthesis System, 1996–2006, Manual and source code avaliable at `http://tcts.fpms.ac.be/synthesis`.

Mihalcea R. – Nastase V. (2002): Letter Level Learning for Language Independent Diacrtitics Restoration. In *Proc. Computational Linguistics*. 1–7.

Nagy B. (2008): Huhyphn: magyar elválasztásiminta-gyűjtemény, `http://www.tipogral.hu/`.

Olaszi P. (2002): *Magyar nyelvű szöveg-beszéd átalakítás: nyelvi modellek, algoritmusok és megvalósításuk*. Phd disszertáció (Budapesti Műszaki és Gazdaságtudományi Egyetem).

Olaszy G. – Németh G. (1999): IVR for banking and residential telephone subscribers using stored messages combined with a new number-to-speech synthesis method. In Gardner-Bonneau D. (szerk.) *Human Factors and Voice Interactive System*. Kluwer, 237–256.

Olaszy G. (1989): *Elektronikus beszédelőállítás, A magyar beszéd akusztikája és formánsszintézise*. Műszaki Kiadó.

Olaszy G. – Kiss G. – Németh G. – Olaszi P. (2000): Profivox: a legkorszerűbb hazai beszédszintetizátor. In *Beszédkutatás'2000*. Budapest, MTA Nyelvtudományi Intézet, 167–179.

Popescu, I. (2003): On a Zipf's law extension to impact factors. 6. évf. *Glottometrics*, 83–93.

Prószéky G. (1988): Hungarian-a Special Challenge to Machine Translation? In *New directions in machine translation: conference proceedings, Budapest, 18-19 August, 1988*. 219.

Přibilová, A. – Přibil, J. (2009): Spectrum modification for emotional speech synthesis. *Multimodal Signals: Cognitive and Algorithmic Issues*, 232–241.

Quinlan, J. (1993): *C4. 5: programs for machine learning*. Morgan Kaufmann.

Rabiner, L. (1968): Digital-Formant Synthesizer for Speech-Synthesis Studies. 43. évf. *J. Acoust. Soc. Am.*, 822–828.

Rutten, P. – Aylett, M. – Fackrell, J. – Taylor, P. (2002): A statistically motivated database pruning technique for unit selection synthesis. In *Seventh International Conference on Spoken Language Processing*. ISCA, 125–128.

Rutten, P. – Fackrell, J. (2003): The application of interactive speech unit selection in TTS systems. In *Eighth European Conference on Speech Communication and Technology*. 285–288.

Scherer, K. (2003): Vocal communication of emotion: A review of research paradigms. 40. évf. 1-2. sz., *Speech communication*, 227–256.

Taylor P. (2009): *Text-to-Speech Synthesis*. Cambridge University Press.

Tokuda, K. – Yoshimura, T. – Masuko, T. – Kobayashi, T. – Kitamura, T. (2000): Speech parameter generation algorithms for HMM-based speech synthesis. In *Acoustics, Speech, and Signal Processing, ICASSP'00. Proceedings. 2000 IEEE International Conference on* , vol. 3. IEEE, 1315–1318. ISBN 0780362934.

Tóth, B. – Németh, G. (2008): Rejtett Markov-Modell Alapú Mesterséges Beszédkeltés Magyar Nyelven. LXIII. köt. 2–6.

Tóth B. – Németh G. (2009): Rejtett Markov-modell alapú szövegfelolvasó adaptációja félig spontán magyar beszéddel. In Alexin Z. – Csendes D. (szerk.) *VI. Magyar Számítógépes Nyelvészeti Konferencia*. Szeged, 213–222.

Ungurean, C. – Burileanu, D. – Popescu, V. – Negrescu, C. – Dervis, A. (2008): Automatic diacritic restoration for a TTS-based e-mail reader application. 70. évf. 4. sz., *University" Politehnica" of Bucharest Scientific Bulletin, Series C: Electrical Engineering and Computer Science*, 3–12. ISSN 1454-234X.

Váradi, T. (1999): On developing the Hungarian national corpus. In *Proceedings of the Workshop Language Technologies-Multilingual Aspects, 32nd Annual Meeting of the Societas Linguistica Europea, Ljubjana, Slovenia*. 57–63.

Zen, H. – Nose, T. – Yamagishi, J. – Sako, S. – Masuko, T. – Black, A. – Tokuda, K. (2007): The HMM-based speech synthesis system (HTS) version 2.0. In *Proc. of Sixth ISCA Workshop on Speech Synthesis*. Citeseer, 294–299.

# A szerző tudományos közleményei

## A tézispontokhoz közvetlenül kapcsolódó

### *Szabadalom*

**[P1]** Zainkó Cs, Németh G, Olaszy G, Gordos G.: Eljárás magyar nyelven ékezetes betűk használata nélkül készített szövegek ékezetes betűinek visszaállítására. Lajstromszám: P 0003443
Közzététel éve: 2000 Benyújtás helye: Magyarország

### *Fejezetek szerkesztett könyvben vagy könyvrészlet*

**[B1]** Németh G, Zainkó Cs, Kiss G, Olaszy G, Fekete L, Tóth D: Replacing a Human Agent by an Automatic Reverse Directory Service.In: Magyar G, Knapp G, Wojtkowski W, W Wojtkowski G, Zupančič J (szerk.)
Advances in Information Systems Development.Springer, 2007. pp. 321–328.

**[B2]** Németh Géza, Zainkó Csaba, Bogár Balázs, Szendrényi Zsolt, Olaszi Péter, Ferenczi Tibor: Elektronikus-levél felolvasó.In: Gósy M (szerk.)
Beszédkutatás 98: Beszéd, spontán beszéd, beszédkommunikáció.Budapest: MTA Nyelvtudományi Intézet, 1998. pp. 189–203.

**[B3]** Németh G, Zainkó Cs: Statisztikai szövegelemzés automatikus felolvasáshoz.In: Gósy M (szerk.) Beszédkutatás 2000: Beszéd és társadalom.Budapest: MTA Kiadó, 2000. pp. 156–166.
Független idézők száma: 1

**[B4]** Zainkó Cs, Fék M: Beszédadatbázis prozódiájának szerepe a gépi beszéd hangzásában és érzelmi tartalmak kifejezésében. Vidám avagy szomorú a beszédszintetizátor? In: Gósy M (szerk.)
Beszédkutatás 2006.Budapest: MTA Kiadó, 2006. pp. 208–217.

**[B5a]** Németh G, Zainkó Cs: Automatikus szám szerinti tudakozó; In: Németh G, Olaszy G (szerk.) A MAGYAR BESZÉD; Beszédkutatás, beszédtechnológia, beszédinformációs rendszerek. Budapest: Akadémiai Kiadó, 2010. pp.

561–562.

**[B5b]** Zainkó Cs: Magyar hang-, betű- és szóstatisztika; Érzelmi töltetű beszéd modellezése; Elemkiválasztás-alapú szövegfelolvasó; Érzelmes szövegfelolvasás In: Németh G, Olaszy G (szerk.) A MAGYAR BESZÉD; Beszédkutatás, beszédtechnológia, beszédinformációs rendszerek. Budapest: Akadémiai Kiadó, 2010. pp. 86–92., 466–467., 505–512., 518–520

### *Folyóratcikkek*

**[J1]** Zainkó Cs, Németh G, Bogár B, Szendrényi Zs: E-levél felolvasó. HÍRADÁSTECHNIKA 49:(11-12) pp. 61–76. (1998)

**[J2]** Németh G, Zainkó Cs, Fekete L, Olaszy G, Endrédi G, Olaszi P, Kiss G, Kiss P: The design, implementation and operation of a Hungarian e-mail reader. INTERNATIONAL JOURNAL OF SPEECH TECHNOLOGY 3-4: pp. 216–228. (2000)
Független idézők száma: 1

**[J3]** Németh G, Zainkó Cs, Fekete L: Statistical analysis used for e-mail reader development and enhancement. HÍRADÁSTECHNIKA LVI:(4) pp. 29–36. (2001)

**[J4]** Németh G, Zainkó Cs, Fekete L: Statisztikai elemzések felhasználása e-levél felolvasó kialakításában és továbbfejlesztésében. HÍRADÁSTECHNIKA LVI:(1) pp. 23–30. (Pollák–Virág díjas) (2001)

**[J5]** Németh G, Zainkó Cs: Multilingual Statistical Text Analysis, Zipf's Law and Hungarian Speech Generation. ACTA LINGUISTICA HUNGARICA 49:(3-4) pp. 385–405. (2002)
Független idézők száma: 2

**[J6]** Zainkó Csaba: Magyar nyelvű, kötött témájú korpusz-alapú beszédszintézis.: és a kötetlenség felé vezető út vizsgálata. HÍRADÁSTECHNIKA LXIII:(5) pp. 12–17. (2008)

### *Konferenciacikkek*

**[C1]** Németh G, Zainkó Cs, Olaszy G, Prószéky G: Problems of creating a flexible e-mail reader for Hungarian. In: European Conference on Speech Communication and Technology (Eurospeech 1999). Budapest, Magyarország,

1999.09.05-1999.09.09.(2) Budapest: pp. 939–942.
Független idézők száma: 4

**[C2]** Németh G, Zainkó Cs: Word Unit Based Multilingual Comparative Analysis of Text Corpora. In: Paul Dalsgaard, Borge Lindberg, Henrik Benner, Zheng-hua Tan (szerk.)
European Conference on Speech Communication and Technology (Eurospeech 2001). Aalborg, Dánia, 2001.09.03-2001.09.07.Aalborg: pp. 2035–2038.
Független idézők száma: 3

**[C3]** Zainkó Cs, Németh G: Statistical Text Processing for Automatic Synthesis of Speech. In: EURASIP Conference on Digital Signal Processing for Multimedia Communications and Services (ECMCS2001). Budapest, Magyarország, 2001.09.11-2001.09.13.Budapest: pp. 301–304.

**[C4]** Németh G, Zainkó Cs, Kiss G, Fék M, Olaszy G, Gordos G: Language Processing for Name and Address Reading in Hungarian. In: IEEE International Conference on Natural Language Processing and Knowledge Engineering (IEEE NLP-KE 2003). Beijing, Kína, 2003.10.26-2003.10.29.Beijing: pp. 238–243.(ISBN: 0-7803-7902-0)

**[C5]** Fék Márk, Zainkó Csaba, Németh Géza: Érzelmes beszéd gépi előállítása érzelem specifikus beszédadatbázisok felhasználásával. In: Alexin Zoltán, Csendes Dóra (szerk.)
Magyar Számítógépes Nyelvészeti Konferencia. Szeged, Magyarország, 2007.12.06-2007.12.07.Szeged: Szegedi Tudományegyetem Informatikai Tanszékcsoport, pp. 34–43.

**[C6]** Németh G, Zainkó Cs, Fék M, Olaszy G, Bartalis M: Promptgenerátor - Ügyfélszolgálati hangos üzenetek automatikus gépi előállítása egy adott bemondó hangjára. In: Alexin Zoltán, Csendes Dóra (szerk.)
Magyar Számítógépes Nyelvészeti Konferencia. Szeged, Magyarország, 2007.12.06-2007.12.07.Szeged: Szegedi Tudományegyetem Informatikai Tanszékcsoport, pp. 3–11.

**[C7]** Géza Németh, Csaba Zainkó, Mátyás Bartalis, Gábor Olaszy, Géza Kiss: Human Voice or Prompt Generation? Can They Co-Exist in an Application? In: Interspeech 2009: Speech and Intelligence. Brighton, Nagy-Britannia, 2009.09.06-2009.09.10.ISCA, pp. 620–623.

**[C8]** Zainkó Csaba: A magyar nyelv betűstatisztikája beszédfeldolgozási szempontok figyelembevételével. In: VI. MAGYAR SZÁMÍTÓGÉPES NYELVÉSZETI KONFERENCIA:. Szeged, Magyarország, 2009.12.03-

2009.12.04.Szeged: pp. 238–245.

**[C9]** Csaba Zainkó , Márk Fék, Géza Németh: Expressive Speech Synthesis Using Emotion-Specific Speech Inventories. LECTURE NOTES IN COMPUTER SCIENCE 5042, Proc. of COST 2102: pp. 225–234. Paper 17. (2008)
Független idézők száma: 1

**[C10]** Csaba Zainkó , Tamás Gábor Csapó, Géza Németh: Special Speech Synthesis for Social Network Websites. LECTURE NOTES IN COMPUTER SCIENCE 6231, Proc. of TSD 2010: pp. 455–463. (2010)

### *Konferencia előadás*

**[C11]** Csaba Zainkó , Géza Németh: Emotional modification for verbal communication. In: The PINK COST 2102 International Conference on Analysis of Verbal and Nonverbal Communication and Enactment: The Processing Issues Budapest, Sep. 7–10, (2010)

## A szerző további tudományos közleményei

### *Fejezet szerkesztett könyvben, könyvrészlet*

**[B5c]** Németh G, Zainkó Cs: Telefonról elérhető e-levél felolvasó; In: Németh G, Olaszy G (szerk.) A MAGYAR BESZÉD; Beszédkutatás, beszédtechnológia, beszédinformációs rendszerek. Budapest: Akadémiai Kiadó, 2010. pp. 555–557.

**[B5d]** Zainkó Cs, Bartalis M, Németh G: Automatikus áru- és árlista-felolvasó In: Németh G, Olaszy G (szerk.) A MAGYAR BESZÉD; Beszédkutatás, beszédtechnológia, beszédinformációs rendszerek. Budapest: Akadémiai Kiadó, 2010. pp. 569–573.

**[B5e]** Zainkó Cs, Németh G: Ékezetek gépi helyreállítása; SMS-felolvasó vezetékes telefonra; Időjárás-előrejelzés írott szöveges és hangos modalitással; Vasútállomási utastájékoztató In: Németh G, Olaszy G (szerk.) A MAGYAR BESZÉD; Beszédkutatás, beszédtechnológia, beszédinformációs rendszerek. Budapest: Akadémiai Kiadó, 2010. pp. 485–488., 557–560., 575–576., 579

**[B6]** Németh G, Olaszy G, Bartalis M, Kiss G, Zainkó Cs, Mihajlik P, Haraszti Cs: Beszédkommunikáció az ember és a gép között. In: Talyigás Judit (szerk.) Mozaikok a hazai telematika eredményeiből.Budapest: Hírközlési és Informa-

tikai Tudományos Egyesület, 2007. pp. 37–52.

**[B7]** Németh G, Olaszy G, Bartalis M, Kiss G, Zainkó Cs, Mihajlik P, Haraszti Cs: Automated Drug Information System for Aged and Visually Impaired Persons. In: Miesenberger K, Klaus J, Zagler W, Karshmer A (szerk.)
Computers Helping People with Special Needs.Springer-Verlag, 2008. pp. 238–241.

**[B8]** Zainkó Cs, Németh G: Az automatikus SMS-felolvasás problémai. In: Gósy Mária (szerk.)
Beszédkutatás 2002: Kísérleti beszédkutatás.Budapest: MTA Nyelvtudományi Intézet, 2002. pp. 197–211.

**[B9]** Németh G, Kiss G, Zainko Cs, Olaszy G, Tóth B: Speech Generation in Mobile Phones. In: Gardner-Bonneau D, Blanchard H. (szerk.)
Human Factors and Interactive Voice Response Systems: Speech Generation in Mobile Phones. Springer, 2008. pp. 163–191.
Független idézők száma: 1

### *Folyóiratcikkek*

**[J7]** Olaszy G, Németh G, Olaszi P, Kiss G, Zainkó Cs, Gordos G: Profivox - a Hungarian TTS System for Telecommunications Applications. INTERNATIONAL JOURNAL OF SPEECH TECHNOLOGY 3–4: pp. 201–215. (2000)
Független idézők száma: 7

**[J8]** Fék M, Pesti P, Németh G, Zainkó Cs: Generációváltás a beszédszintézisben. HÍRADÁSTECHNIKA LXI:(3) pp. 21–30. (2006)

**[J9]** Olaszy G, Németh G, Bartalis M, Kiss G, Zainkó Cs, Fegyó T, Árvay G, Szepezdi Zs, Terplánné Balogh M: Kísérleti gyógyszerinformációs rendszer beszédmodulokkal. HÍRADÁSTECHNIKA LXI:(3) pp. 8–13. (2006)

**[J10]** Németh Géza, Olaszy Gábor, Bartalis Mátyás, Zainkó Csaba, Fék Márk, Mihajlik Péter: Beszédadatbázisok előkészítése kutatási és fejlesztési célok hatékonyabb támogatására. HÍRADÁSTECHNIKA LXIII:(5) pp. 18–24. (2008)

**[J11]** Tamás Gábor Csapó, Csaba Zainkó, Géza Németh: A Study of Prosodic Variability Methods in a Corpus-Based Unit Selection Text-To-Speech System. INFOCOMMUNICATIONS JOURNAL LXV:(1) pp. 32–37. (2010)

### *Konferenciacikkek*

**[C12]** Fék M, Pesti P, Németh G, Zainkó Cs, Olaszy G: Corpus-Based Unit Selection TTS for Hungarian. LECTURE NOTES IN COMPUTER SCIENCE 4188, Proc. of TSD 2006: pp. 367–373. (2006)
Független idézők száma: 2

**[C13]** Abari K, Olaszy G, Kiss G, Zainkó Cs: Magyar kiejtési szótár az Interneten. In: Alexin Zoltán, Csendes Dóra (szerk.)
Magyar Számítógépes Nyelvészeti Konferencia. Szeged, Magyarország, 2006.12.07-2006.12.08.Szegedi Tudományegyetem Informatikai Tanszékcsoport, pp. 223–230.

**[C14]** G Németh, G Olaszy, M Bartalis, G Kiss, Cs Zainkó, P Mihajlik: Speech based Drug Information System for Aged and Visually Impaired Persons. In: Interspeech 2007 - Eurospeech: 9th European Conference on Speech Communication and Technology. Antwerpen, Belgium, 2007.08.27-2007.08.31.ISCA, pp. 2533–2536.