Budapesti Műszaki és Gazdaságtudományi Egyetem
Távközlési és Médiainformatikai Tanszék

Gépi beszédkeltés infokommunikációs rendszerekben

PhD disszertáció
Informatikai Tudományok Doktori Iskola

Zainkó Csaba
okl. mérnök-informatikus

Témavezetők:
Németh Géza, PhD
Olaszy Gábor, DSc

Budapest, 2010.

Verzió: 166-166 orig - none

Nyilatkozat önálló munkáról, hivatkozások átvételéről

Alulírott Zainkó Csaba kijelentem, hogy ezt a doktori értekezést magam készítettem és abban csak a megadott forrásokat használtam fel. Minden olyan részt, amelyet szó szerint, vagy azonos tartalomban, de átfogalmazva más forrásból átvettem, egyértelműen, a forrás megadásával megjelöltem.

Budapest, ...

Nyilatkozat nyilvánosságra hozatalról

Alulírott Zainkó Csaba hozzájárulok a doktori értekezésem interneten történő nyilvánosságra hozatalához az alábbi formában*:

- korlátozás nélkül
- elérhetőség csak magyarországi címről
- elérhetőség a fokozat odaítélését követően 2 év múlva, korlátozás nélkül
- elérhetőség a fokozat odaítélését követően 2 év múlva, csak magyarországi címről

Budapest, ...

*a megfelelő választást kérjük aláhúzni

Abstract

of the PhD Thesis of Csaba Zainkó,
„Automatic speech generation in infocommunication systems”

One of the research areas aiming to facilitate human-computer communication – and the one of my focus in research – is automatic speech generation and the related field of automatic text processing. My goal is to develop new algorithms and methods to improve the quality of communication between human and computer in infocommunication systems.

I developed a fast dictionary-based procedure for recovering the original form (with diacritics) of texts written without diacritics, which reached an accuracy of 95% in texts of emails. I improved the procedure up to 97.4% correct regeneration applying decision trees for reducing the errors made by the dictionary-based algorithm.

I determined the basic applicability conditions for the comparison of English-, German- and Hungarian-language word-based automatic speech technology methods. I have demonstrated the possibility of applications of the procedure.

By extending the notion of the letter, I developed a novel method for the qualification of texts for the purpose of speech synthesis. The novelty of the method is to take into consideration the connection between letters and phones. I have tested the method on the Hungarian National Corpus.

I developed procedures and algorithms for the text-to-speech conversion of Hungarian proper names, company names, and addresses. I proved the effectiveness of the name- and address-reading procedure by an intelligibility test.

I developed a procedure, based on virtual word intensity, to intensity-normalize the speech produced by a corpus-based synthesizer. The sound intensity distribution of Hungarian read speech was also determined based on recordings from multiple speakers as a basic research result foundation of the normalization procedure.

I proposed a procedure to produce speech with emotional content. It can transform neutral voice of the speech synthesizer to emotional voice. I applied this procedure directly on spoken human waveform and on corpus based unit selection, diphone/triphone waveform concatenative and HMM speech synthesizers.

Finally, a summary of the new results is given and the possible applications are presented.

Kivonat

Zainkó Csaba

„Gépi beszédkeltés infokommunikációs rendszerekben”
című PhD értekezéséhez

Az ember-gép kommunikációját megkönnyítő egyik kutatási témakör – amellyel kutatásaim során legtöbbet foglalkoztam – a gépi beszédelőállítás és a hozzá kapcsolódó gépi szövegfeldolgozás. Célom olyan új algoritmusok és módszerek kidolgozása, amelyek segítségével javítható az ember-gép közötti beszédkommunikáció minősége infokommunikációs rendszerekben.

Kidolgoztam egy olyan gyors szótár alapú eljárást, amely képes 95% pontossággal ékezet nélkül írt elektronikus szövegek ékezetes formájának visszaállítására. Ezt az eljárást továbbfejlesztettem döntési fák alkalmazásával, amely csökkenti a szótár alapú algoritmus hibáit. Az eljárást általános szövegen is teszteltem, a továbbfejlesztéssel 97,4% szópontosságot értem el.

Meghatároztam az angol, a német és a magyar nyelvű szó alapú gépi beszédtechnológiai módszerek összehasonlításához szükséges alapvető alkalmazhatósági sarokpontokat és bemutattam azok alkalmazási lehetőségeit.

A betű fogalmának kiterjesztésével új módszert dolgoztam ki szövegek beszédszintézis szempontjait figyelembe vevő minősítéséhez. A módszer újdonságeleme az, hogy a betű-hang kapcsolatot is figyelembe veszi a szöveg értékelésénél. A módszert a Magyar Nemzeti Szövegtáron teszteltem.

Eljárásokat és algoritmusokat dolgoztam ki magyar tulajdonnevek, cégnevek és magyarországi címek gépi felolvasásához. A név- és címfelolvasó eljárás eredményességét érthetőségi teszttel bizonyítottam.

Virtuális szóintenzitáson alapuló eljárást dolgoztam ki korpusz alapú szintetizátor beszédének intenzitás-kiegyenlítéséhez. A módszerhez vezető kutatások leágazó eredményeként meghatároztam több beszélőtől származó hangfelvételek alapján a felolvasott magyar beszédre jellemző hangintenzitás eloszlásokat, amelyek megalapozták a normalizáló eljárást. Igazoltam, hogy automatikus intenzitáskiegyenlítésnél kapott eredmények eredményesen felhasználhatók.

Beszédszintetizátorok hangjának érzelmi módosításához eljárást dolgoztam ki, amellyel a semleges hangú beszédszintezátor hangja megváltoztatható. Az eljárást három féle beszédszintetizálási technológián mutattam be, korpusz alapú elemkiválasztásos, diád és triád alapú hullámforma összefűzéses és HMM alapú beszédszintetizátorokon. Igazoltam továbbá, hogy természetes emberi beszéden is alkalmazható.

A disszertáció végén összefoglaltam az új eredményeimet és bemutattam az alkalmazási lehetőségeket.

Tartalomjegyzék

Függelékek

1. fejezet
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 Beszédtechnológia 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).

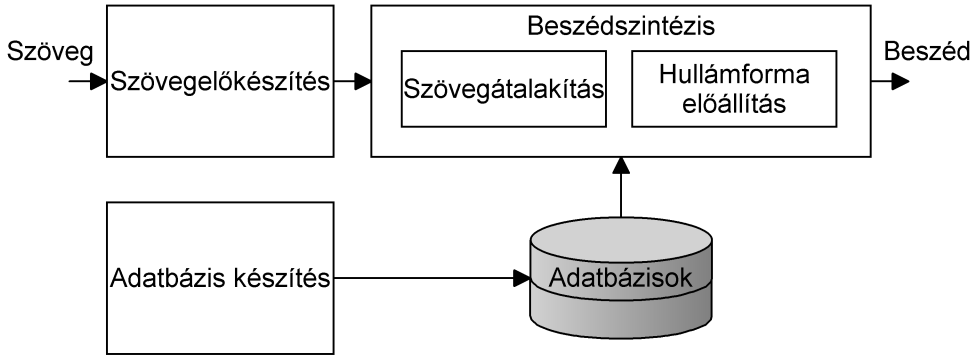
A magyar beszédszintézis területén kisebb cégek készítettek saját fejlesztésű zárt rendszereket, például a Speech Technology Kft. Ezekről a rendszerekről publikációk nem jelentek meg. Az utóbbi években külföldi fejlesztések is elindultak. Elkészült az MBROLA szabad felhasználású szintézis rendszer magyar 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 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.



1.1. ábra. Beszédszintézis folyamatának egyszerűsített blokkvázlata

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 egyszerűsített blokkvázlaton (1.1.ábra) látható főbb komponensekhez kapcsolódóan mutatom be, a disszertáció fejezeteit is ezen területek szerint osztottam fel: a 2. fejezetben a szövegelőkészítéshez kapcsolódó ékezetesítő eljárásokat ismertetem. A 3. fejezetben az adatbáziskészítéshez kapcsolódó eredményeimet foglalom össze. A 4., az 5. és a 6. fejezetekben a gépi beszédelőállításhoz kapcsolódó 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.

1.1. 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 tartom 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.

2. fejezet
Ékezet nélkül írt szövegek automatikus helyreállítása

A gépi beszédfeldolgozásban és főként a beszédszintézisben sokszor előfordul, hogy olyan szövegeket kell feldolgozni, amelyek részben vagy teljesen ékezetmentesek. Az ékezetmentes szövegek előállhatnak úgy, hogy a felhasználó nem használ ékezetes betűket, mert 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, és így bizonyos ékezetes betűk a konverziók során ékezet nélkülivé válnak.

Például:

```
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 percepciós rendszerünk azonban már nem képes könnyen feldolgozni és megérteni, mivel az ékezetek elhagyása a legtöbb esetben hangváltozást is eredményez. A 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 a szövegek gépi felolvasása előtt az ékezetek meglétét, illetve hiányát géppel ellenőrizni, majd elvégezni az újraékezetesítést.

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), de leginkább a román területen kutatták ezt a problémát. A publikált megoldások általában gépi tanuláson alapultak, amelyek ugyan általános megoldást adnak, de nem veszik figyelembe az infokommunikációs rendszerek és a beszédszintézis követelményeit. Például nem vizsgálták azt, hogy az ismertetett eljárások milyen jellegű hibákat okoznak szószinten, amelyek a későbbiek során a szintetizált beszéd érthetőségét befolyásolhatják. Magyar nyelvre a MorphoLogic Kft. kínál valószínűleg nyelvi ellenőrzés alapú megoldást, de erről publikációk nem jelentek meg.

Ebben a fejezetben ismertetem, hogy infokommunikációs környezetben milyen okokból és környezetben fordul elő leggyakrabban ékezet nélküli szöveg. Ezt követően megadom az általam kidolgozott szótár alapú algoritmust, amely hatékonyságát és sebességét két teszthalmazon, elektronikus leveleken és általános nagyméretű szövegadatbázisokon vizsgáltam. Bemutatom továbbá a szótár alapú alapalgoritmus továbbfejlesztésére felhasznált

döntési fák alapalgoritmushoz való illesztését, és az ezek eredményességére vonatkozó tesztek adatait.

2.1. Ékezet nélküli szövegek forrásai infokommunikációs rendszerekben

Ékezet nélküli szövegek többfajta módon állnak elő. Régebben jellemző volt, hogy a számítógépes rendszerek tervezése és készítése Amerikában történt és a különböző nyelvi támogatások nem voltak elérhetők magyar nyelvre. Az ékezetes karakterek bevitelére sok esetben nem is volt lehetőség, vagy az ékezetes dokumentum tárolása, mozgatása vagy másolása közben nagy eséllyel sérültek az ékezetes karakterek. A kódolási problémák miatt nem csak az ékezetek eltűnésének volt realitása, hanem a dokumentum olvashatatlanná válásának is. A felhasználók – az előbb felsorolt veszélyek miatt – abban az esetben is inkább ékezet nélkül írtak, ha az ékezetes karakterek bevitelére egyébként lehetőségük volt. A széleskörű nyelvi támogatások elterjedésével a kódolási és beviteli korlátok fokozatosan csökkentek, de kisebb mértékben ma is léteznek. A felhasználók hozzászoktak az ékezet nélküli bevitelhez. Felismerték, hogy gyorsabban lehet gépelni 26 karakter használatával, így bizonyos felhasználási környezetekben – ahol megengedett ez a lazaság – még előfordul az ékezet nélküli szövegek írása. A 2.1. táblázatban láthatunk példákat az ékezethiány főbb okaira és az átvitt karakterekre. A táblázat harmadik oszlopában áthúzással jelöltem az át nem vitt karaktereket, és kiemeléssel az átvitteket. Megadtam a táblázatban, hogy az adott rendszerben milyen alternatív karakterek használata lehetséges.

2.1. táblázat. Ékezet nélküli szövegek keletkezésének okai

	Lehetséges ok	A leggyakrabban átvitt karakterek	Alternatív karakterek
Régi rendszer	nincs lehetőség	á,é,í,ó,ö,ő,ú,ü,ű,Á,É,Í,Ó,Ö,Ő,Ú,Ü,Ű	
Kényelmes felhasználó	gyorsabb bevitel	á,é,í,ó,ö,ő,ú,ü,ű,Á,É,Í,Ó,Ö,Ő,Ú,Ü,Ű	
SMS GSM 03.38	kódolási korlát	á,**é**,í,ó,**ö**,ő,ú,**ü**,ű,Á,**É**,Í,Ó,**Ö**,Ő,Ú,**Ü**,Ű	à,å,ì,ò,ù,Å
ISO-8859-1	nem jó kódlap	**á**,**é**,**í**,**ó**,**ö**,ő,**ú**,**ü**,ű,**Á**,**É**,**Í**,**Ó**,**Ö**,Ő,**Ú**,**Ü**,Ű	õ,ũ,Õ,Ũ

A mindennapi életben egyre szélesebb körben terjednek el a kis képernyős és általában csökkentett billentyűzettel rendelkező készülékek. A készülékek egy része – mint például a mobiltelefonok többsége – csak numerikus billentyűzettel rendelkezik, amelyen keresztül ugyan van lehetőség szöveg bevitelére, de használata lassú és körülményes. A karakterek beviteléhez a billentyűk többszöri megnyomására van szükség. Az ékezetes karakterek elhelyezése ebben a beviteli módban többnyire gyártófüggő, vagy az angol ábécé betűi közé, vagy azok utánra sorolták be. Ez a változatosság a felhasználót arra ösztönzi, hogy inkább kerülje az ékezetes karakterek használatát, mert minden készülékváltás esetén újra meg kell tanulnia a bevitel módját.

A mobiltelefonos rendszerekben előfordul, hogy SMS-ek esetében, a bevitt ékezetes karakterek a készüléken még megjelennek, de a továbbítás során az ékezet törlődik bizonyos betűkről. A felhasználó közvetlen visszacsatolást nem is kap arról, hogy az általa írt szöveg hogyan módosul.

A modernebb kis képernyős és általában érintőképernyős készülékek – okos telefonok, PDA-k – többnyire már lehetőséget biztosítanak az ékezetes karakterek teljes körű bevitelére is, de vannak olyan készülékek, ahol az ékezetes karakterek használata nagyságrenddel több időt

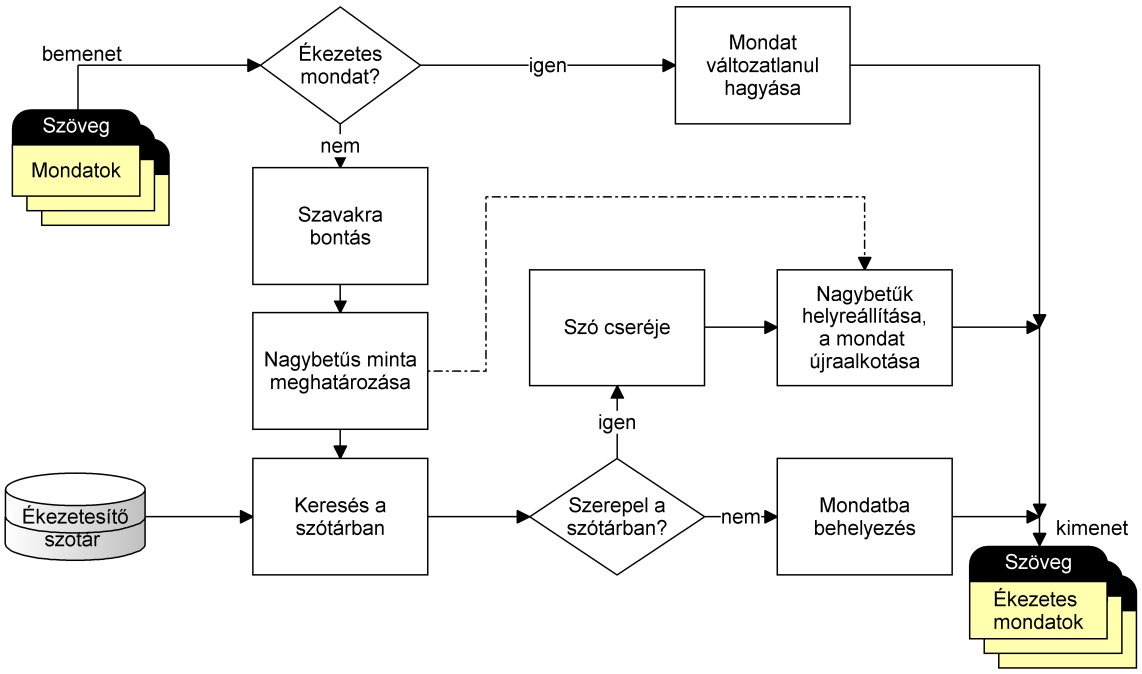
jelent, mint egy ékezet nélküli karakteré. Például alternatív beviteli billentyűzetre kell váltani, vagy az adott karakter hosszú lenyomása esetén választhatunk az ékezetes betűk közül.

Ékezet nélküli szövegek tehát több okból is keletkezhetnek, illetve az ékezet hiányának mértéke is változó. Például az SMS-ek esetén a felhasználó szokásai, az általa használt készülék és a szolgáltató rendszerének beállítása is okozhatja az ékezet nélküli szövegek létrejöttét.

2.2. Szótár alapú eljárás ékezet nélkül írt elektronikus szövegek ékezetes formájának visszaállítására

2.2.1. Az alapalgoritmus

Alapalgoritmusnak nevezem az általam kidolgozott szótár alapú ékezetesítő eljárást, amely szó alapú feldolgozást végez, azaz a szavakon belüli nyelvi egységeket (szótag, rag, jel) nem elemzi. A 2.1. ábrán mutatom be az alapalgoritmus működési folyamatát. Első lépésben az algoritmus



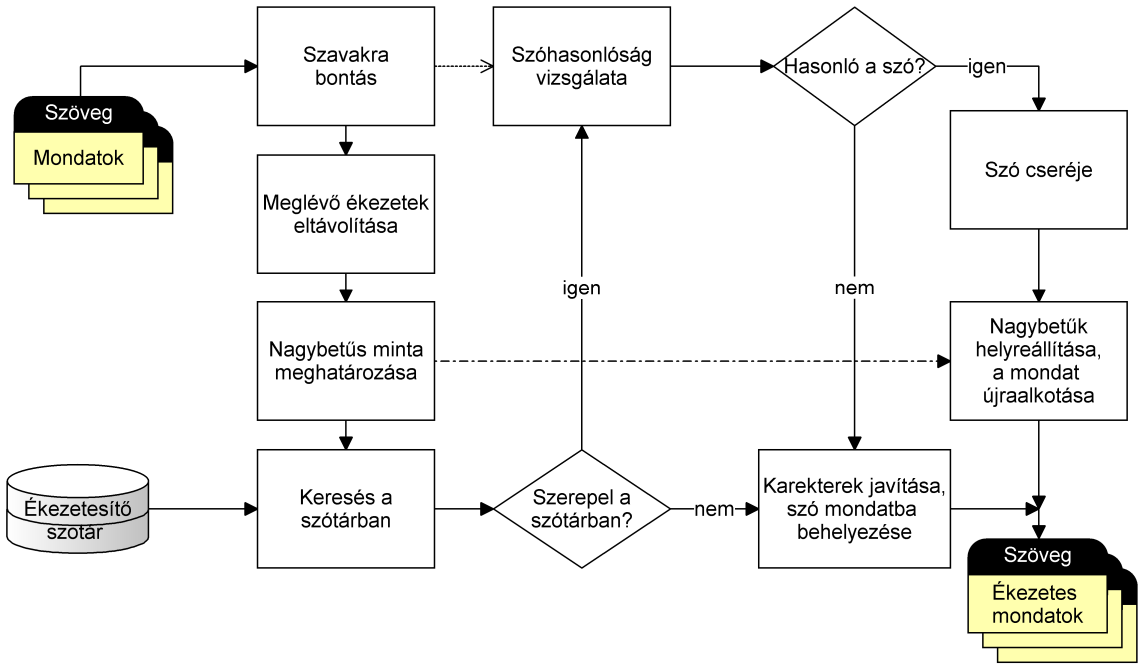
2.1. ábra. Szótár alapú ékezetesítés folyamata

még mondatszinten döntést hoz arról, hogy szükséges-e az ékezetesítés. Amennyiben a mondat már tartalmaz legalább egy ékezetes karaktert, akkor az ékezetesítést arra a mondatra már nem végezzük el, a mondat változatlan formában kerül a kimenetre. A mondat fogalmát a hagyományos értelmezésben használom, a pont, felkiáltójel és a kérdőjellel végződő feloldott – például a rövidítéseket nem tartalmazó – szósorozat.

Ha ékezetesítés szükséges, akkor a mondatokat a szóköz karakterek mentén szavakra bontja az eljárás és minden szót szekvenciálisan megvizsgál. A gépi szövegfelolvasás szempontjából jelentősége van a kis- és nagybetűk meglétének, prozódiai és egyéb feldolgozási szabályok működéséhez szükségesek ezek. Például az „*OK*” betűszó kiejtése nem egyezik meg az „*ok*” szó kiejtésével. Az ékezetesítés előtt tehát szükséges a kis- és nagybetűk szóra vonatkozó

mintázatának rögzítése. Ezután az ékezetesítő szótárban a szóalak keresése kezdődik. Az ékezetesítő szótár készítését a 2.2.2. fejezetben mutatom be részletesen. Ha nem található meg az ékezetesítő szótárban a keresett szó, akkor az változatlan formában kerül a kimeneti mondatba. Ha szerepel benne, akkor az ékezetes formára az előzőleg rögzített kis- és nagybetű mintázatát rávezeti az algoritmus és így illeszti be a kimeneti mondatba. A most ismertetett ékezetesítési folyamatot hívom a továbbiakban alapalgoritmusnak.

Az ékezetesítő eljárás nem csak teljes mondatokra alkalmazható, vannak olyan alkalmazások, ahol mondat szinten nem lehet meghatározni, hogy szükséges-e az ékezetesítés vagy nem. Ezért bevezettem a részben ékezetes szöveg fogalmát, valamint meghatároztam egy kétlépéses ékezet-visszaállítást. Részben ékezetes a szöveg akkor, ha a mondatokban szereplő szavak tartalmaznak ékezeteket, de az ékezetekből legalább egy hiányzik. Például SMS-ek esetén részben ékezetes szövegbemenettel lehet számolni. Ilyen esetek kezelésére az alapalgoritmus keresési és szócserélési részét módosítani kell. A keresés megindítása előtt az ékezetes szavak ékezet nélküli formáját kell előállítani. A keresés során ezt az ékezet nélküli formát használja az algoritmus, a részben meglévő vagy hasonló ékezetek (például a „*rõzse*” szó „*õ*” betűje) információtartalmát nem, így az alapalgoritmushoz épített szótár is felhasználható. Sikeres keresés esetén a szótárban szereplő ékezetes szót összehasonlítjuk a bemeneti ékezetes szóval. Megvizsgáljuk, hogy van-e ellentmondás a keresési eredmény és az eredeti szó ékezetei között. Ellentmondás akkor áll fenn, ha valamelyik meglévő ékezetes betű helyén a szótárban másfajta ékezetes betű áll, például a bemeneten szereplő részben ékezetes szó az „*ónoz*” és a találat pedig az „*önöz*” szó. Abban az esetben nem áll fenn ellentmondás, ha a meglévő ékezetes betű a szótárban szereplőnek egy alternatív írási módja, például a bemeneten a „*fõùt*” szerepel, a szótárban pedig a helyes „*főút*” szó van. Ha nincs ellentmondás, akkor a szótárban szereplő találatra cseréljük a bemeneti szót. Ellentmondás esetén a bemeneti szó marad, vagy a bemeneti javított formája (például a „*rõzse*” lecserélése „*rőzse*” szóra) kerül a kimeneti mondatba. A részben ékezetes szövegbemenet feldolgozásának főbb lépései a 2.2. ábrán láthatók.

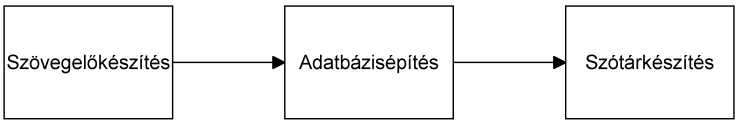


2.2. ábra. Részben ékezetes szöveg szótár alapú ékezetesítésének folyamata

Egyértelmű esetekben elégséges lenne – az ékezetesítő eljárást kihagyva – csak a szó javított formáját a kimeneti mondatba berakni, például „*vår*” alak helyett, a „*vár*” szót, de ezek csak kis számban fordulnak elő a magyar nyelvben, ezért külön kezelésük nem indokolt.

2.2.2. Szótár építése

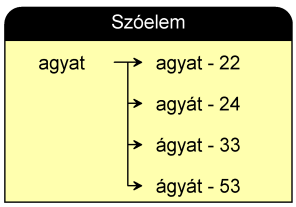
Az alapalgoritmus a szótárra támaszkodik. Ezt a szótárat nagy szövegadatbázis (tanító szövegkorpusz) alapján építjük fel , amely tartalmazza a különböző ékezet nélküli szóalakokat és a hozzájuk tartozó, nyelvileg korrekt ékezetes formákat is. A szótár építéséhez nagy mennyiségű szöveg szükséges. A pontos újraékezetesítéshez olyan szövegekből kell a szótárt felépíteni, amelyek az ékezetesítendő bemeneti szöveggel megegyező tematikájúak. Ez sok gyakorlati alkalmazás esetén nem lehetséges, mert például SMS-ek és elektronikus levelek – főleg adatvédelmi okok miatt – nagy tömegben nem elérhetőek. További problémát jelent az, hogy a szótárépítéshez használt tanító adatoknak nyelvtanilag helyesnek kell lenniük, mert a sok hibát tartalmazó szöveg statisztikai tulajdonságai nem megfelelőek. A gyakori hibás alakok hibás szótárt, ezáltal hibás kimenetet adnak.



2.3. ábra. A szótárépítés folyamata

A szótárkészítés első lépése a szövegelőkészítés (2.3. ábra). A tanító szövegek a szavakon kívül egyéb karaktereket, számokat, írásjeleket is tartalmaznak. Ezeket el kell távolítani. A szövegek szavakra bontásánál fontos, hogy az algoritmus ugyanazokat a szabályokat használja, mint amit később az ékezetesítő eljárás közben fog alkalmazni. Ha például a kötőjeles szavakat a szótárépítés külön szavakra bontja, akkor az ékezetesítő eljárás közben is így kell a mondatot feldolgozni, különben az eljárás a kötőjeles szavakat nem fogja tudni ékezetesíteni.

A második lépésben, az előkészített szavakból adatbázist kell építeni, amely tartalmazza az ékezet nélküli szóalakokhoz tartozó különböző alakokat (ékezetest és ha van, akkor az ékezet nélkülit is), és azok gyakorisági adatait. Például az „*agyat*” szó különböző alakjait és előfordulási gyakoriságukat láthatjuk a 2.4. ábrán. A gyakorisági adatok alapján a szavakat a



2.4. ábra. Példa egy több lehetséges értelmes alakkal rendelkező magyar szóalakra és előfordulási gyakoriságára

következő kategóriákba soroltam:

Egyalakú ékezet nélküli: Ezek a szavak a szótárépítés során csak egyetlen ékezet nélküli alakban fordulnak elő. Ilyen például a „*magyar*” szó, amelynek nincs ékezetes variációja.

Egyalakú ékezetes: Az olyan szavak, amelyek csak egyetlen ékezetes formában fordulnak elő, és az ékezet nélküli alakjuk a magyar nyelvben nem értelmes szó. Például: „*lépcső*”.

Több alakú (kétes eset): Azok a szavak, amelyeknek több értelmes ékezetes alakjuk van, de az ékezet nélküli formájuk értelmetlen. Például: „*hó-hő*”. Ide tartoznak azok az szavak is, amelyeknek az ékezet nélküli alakjuk is értelmes és még legalább egy ékezetes alakjuk is van. Például: „*mar-már*”. Ezeket a több alakú szavakat kétes eseteknek is hívom, mert csak önmagából a szóból nem döntehető el, hogy melyik alak – eset – a megfelelő.

A szótárépítéshez felhasznált tanító szöveg nagy valószínűséggel tartalmazni fog hibás szóalakokat, amelyek elgépelés, hibás helyesírás miatt vagy egyéb okból a szövegbe kerültek. Ha ezeknek a hibás alakoknak az előfordulási száma alacsonynak vélhető, akkor a statisztika feldolgozása során az adott szó alacsony gyakoriságú alakjait nem kell figyelembe venni. Amennyiben az ilyen hibás alakok száma nagy, akkor az adott szöveg nem használható szótárépítésre.

A szótárépítés utolsó lépése, hogy az előzőleg felépített adatbázisból elhelyezzük azokat a szavakat a szótárban, amelyek az adott ékezet nélküli szóalakhoz a leggyakrabban fordultak elő a tanító szövegben. A 2.4. ábrán látható „agyat” ékezet nélküli szóalakhoz a szótárba az „ágyát” szóalak fog bekerülni. Ha az ékezetesítő eljárás során csak teljesen ékezet nélküli szövegeket kell ékezetesíteni (2.1. ábra), akkor az előzőleg felsorolt kategóriák közül nem szükséges az összes típusú szó eltárolása. A szótárba nem kell tárolni az egyalakú ékezet nélküli szavakat és azokat a több alakú szavakat, amelyeknél az ékezet nélküli forma a leggyakoribb, mert az ékezetesítő algoritmus is ezeket az ékezet nélküli alakokat helyezi a kimeneti mondatba. A részben ékezetes mondatok kezelésére módosított ékezetesítő algoritmus esetében is hasonlóan csökkenteni lehet a szótár méretét, ami az ékezetesítő algoritmus keresési terét csökkenti, ezáltal gyorsítva a működést. Az olyan esetekben, amikor azokat a szavakat is kezelni kívánjuk, amelyek nem szerepelnek a tanító szövegekben, ez az egyszerűsítés nem hajtható végre. Meg kell különböztetni azokat a szavakat, amelyeknél az ékezet nélküli alak a leggyakoribb azoktól, amelyekhez egyáltalán nincs gyakorisági adat. Az ilyen algoritmus leírását a 2.3.2. fejezetben ismertetem. Az ékezetesítő szótárból részleteket az A. függelékben mutatok be.

A szótár építése közben egyszerűsítéseket végez az algoritmus. A szótárépítő eljárás egyformán kezeli a kis- és nagybetűket, minden szót kisbetűsre alakít át. Ezzel információt vesztünk, mert vannak olyan esetek, amelynél a nagybetűs írott forma segítene a megfelelő ékezetes forma kiválasztásánál. Például a „*SI*” szó a nemzetközi mértékegység angol rövidítése lehet, amelyet változatlanul kell hagyni, a „*si*” pedig a sportág lehet, amely esetében „*sí*” lesz a helyes alak. Ezzel szemben számos problémát is okoz a kis- és nagybetűk megkülönböztetése, amely miatt ezt az információt az algoritmus inkább eldobja. Például a tanító szövegadatbázisok elemzése során megállapítottam, hogy a szövegekben a felhasználók nem a magyar helyesírásnak megfelelően használják a kis- és nagybetűket. Kiemelésekre alkalmazzák, vagy internetes kommunikációban „kiabálásra” használják a csupa nagybetűs írásmódot. Személyes kommunikációban gyakran elhagyják a mondatkezdő, vagy más helyen szereplő nagybetűket és csupa kisbetűt használva írnak. A szótár építése során azonban az ilyen személyes kommunikációt tartalmazó írásformákat nem szabad kihagyni, mivel infokommunikációs alkalmazásokban az ilyen jellegű szövegek gyakran előfordulhatnak, például SMS-ek felolvasásakor. A másik ok, hogy elhagyjuk a kis- és

nagybetűk megkülönböztetését, hogy a szótárban túl ritkán fordulnának elő bizonyos szóalakok, ha minden szót az eredeti formájában vizsgálnánk.

Részben az infokommunikációs rendszerekben történő gyors működés érdekében illetve az előzőekben említett szövegformai hanyagságok miatt az algoritmus nem tartalmaz nyelvi elemzéseket, nem vizsgálja a toldalékokat.

2.2.3. Az alapalgoritmus tesztelése

Az alapalgoritmus tesztelését kétféle adathalmazon végeztem el, egyrészről elektronikus leveleken futtattam, másrészről általános nagyméretű adatbázison. Ezzel elértem, hogy mind szűk és mind általános tematikájú szövegek esetén rendelkezésre álljanak szóhelyességi eredmények.

2.2.3.1. Újraékezetesítés elektronikus levelekben

Az alapalgoritmust elektronikuslevél-felolvasó alkalmazásban teszteltem. A szótárépítéshez nem állt rendelkezésre nagy mennyiségű elektronikus levél, ezért az internetről gyűjtött elektronikus szöveganyagokat használtam fel, amelyek a következő forrásokból származnak: a Magyar Elektronikus Könyvtár és a Magyar Nemzeti Szövegtár 1999-es verziója ($MNSZ_{1999}$). A szövegek mérete megközelítőleg 24 millió szövegszó volt, ami közepes méretű adatbázisnak tekinthető. A két szöveganyagból elkészített ékezetesítő szótárban lévő rekordok száma 330 és 745 ezer lett. A két szöveganyag egyesített szótára 900 ezer rekord méretű, amely nem tartalmazta azokat az alakokat, amelyek esetében az ékezet nélküli alak a leggyakoribb. A szövegforrások főbb adatairól a 2.2. táblázatban adok összefoglalást.

2.2. táblázat. Elektronikuslevél-felolvasó szótárépítéséhez használt források

Forrás neve	Szavak száma	Különböző szavak száma	Forrásból épített szótár mérete
Magyar Elektronikus Könyvtár (MEK)	≈3 millió	≈340 000	328 498
Magyar Nemzeti Szövegtár ($MNSZ_{1999}$)	≈21 millió	≈750 000	734 725
Kombinált MEK+$MNSZ_{1999}$	≈24 millió	≈950 000	900 564

Az eljárás hatékonyságát egy 380 elektronikus levélből álló teszthalmazon teszteltem, amelyet ékezetes magyar nyelvű levelekből választottam ki véletlenszerűen. A leveleket egy távközlési szolgáltató bocsájtotta rendelkezésemre, anonimizálva, szintén véletlenszerűen kiválasztva aktuális levelezési forgalmából. A leveleket manuálisan ellenőriztem, hogy ne tartalmazzanak durva helyesírási hibát, vagy nagyobb mennyiségű idegen szöveget. A teszthalmaz leveleiről eltávolítottam az ékezeteket, majd lefuttattam az ékezetesítő eljárást. Szavakra vonatkoztatva vizsgáltam a hibákat, a szót helyesnek tekintettem, amennyiben az eredeti ékezetes szöveghez képest a szóban nem volt betűeltérés.

Az eljárást a 2.2. táblázatban részletezett három szövegkorpuszból készített szótárral is teszteltem és összehasonlítottam másik két módszerrel is. A teszthalmaz a szótár építő adatoktól teljesen független volt. Az eredmények a 2.3. táblázatban láthatók. A különböző méretű szótárak felhasználásának az volt a célja, hogy megvizsgáljam a szóhelyesség (amikor az algoritmus helyesen állítja helyre az ékezeteket) és a szótárméret közötti összefüggést. A

másik két – az alapalgoritmustól eltérő – módszerrel azért hasonlítottam össze az eljárásomat, hogy megvizsgáljam a nyelvi ellenőrzés alapú módszerek sebességét és eredményességét. A 2.3. táblázatban található különböző szótárak a 2.2. táblázatban felsorolt különböző méretű forrásokból épített ékezetesítő szótárak. A táblázat első, negyedik és ötödik sorában –

2.3. táblázat. Az újraékezetesítő algoritmusok sebességi és szóhelyességi adatai

Eljárás neve	Futásidő* [mp]	Helyes szavak aránya [%]
Szótár alapú (MEK)	**31**	**90,46**
Nyelvi ellenőrzés alapú	1900	89,75
Kombinált	863	94,90
Szótár alapú (MNSZ$_{1999}$)	**39**	**94,92**
Szótár alapú (MNSZ$_{1999}$+MEK)	**51**	**95,18**

* PC, Windows NT 4.0, 266MHz Pentium II CPU, 64Mbyte RAM.

megvastagítva – találhatók az ismertetett eljárásom eredményei. A nyelvi ellenőrzés alapú megoldást a MorphoLogic Kft. helyesírás ellenőrző modulja segítségével alakítottam ki. Az ékezetesítendő szó összes lehetséges formáját a helyesírás ellenőrző modul bemenetére adtam. Azt az ékezetes alakot helyettesítettem be az ékezet nélküli alakra, amelyre a helyesírás ellenőrző először adott helyes ítéletet. Ha nem volt helyes alak, akkor az ékezet nélküli alak maradt. A harmadik sorban megadott kombinált módszer esetében a helyesírás ellenőrző csak a táblázat első sorában szereplő, szótár alapú (MEK) eljárás azon szavaira futott le, amelyek nem szerepeltek abban a szótárban. Tehát a helyesírás ellenőrző modul csak azokat a szavakat ékezetesítette, amelyeket az alapalgoritmus változatlan formában hagyta, mivel nem szerepelt a szótárában.

A vizsgált eljárások közül az MNSZ$_{1999}$ és a MEK összesített forrásából épített szótár alapú eljárás adata a legtöbb helyes szóalakot. A három vizsgált szótár alapú eljárás közül tehát a legnagyobb méretű forrás esetében volt a helyes szavak száma a legmagasabb. A nyelvi ellenőrzés alapú megoldás futási ideje volt a leghosszabb. A kombinált megoldás gyorsította az ékezetesítést, de a tisztán szótár alapú megoldások egy nagyságrenddel rövidebb idő alatt futottak le. A nyelvi ellenőrzés alapú ékezetesítés továbbfejlesztésére nem volt lehetőségem, jobb eredményeket valószínűleg csak a helyesírás ellenőrző modul készítői (MorphoLogic Kft. fejlesztői) tudnának elérni.

A vizsgálatba bevont hardver eszköz egyenértékű azzal az erőforrással, mint ami egy mai mobil eszközben tipikusan rendelkezésre áll a párhuzamosan futó alkalmazások mellett. Mivel az ékezetesítés problémaköre jelenleg a mobil eszközök esetében a leggyakoribb, így a vizsgált hardver környezetben kapott értékek jól felhasználhatók a mobil alkalmazások futásidejének becslésére.

2.2.3.2. Újraékezetesítés nagyméretű tanító-adatbázissal

Az elektronikus levelek újraékézetesítése során (előző fejezet) megállapítottam, hogy a nagyobb méretű szótár esetén a helyesen ékezetesített szavak száma nagyobb volt. Ennek a tendenciának a teszteléséhez újabb vizsgálatot végeztem, az elektronikus leveleknél használt tanító szövegállomány helyett egy nagyságrenddel nagyobb szövegkorpuszt használtam fel. A Magyar Nemzeti Szövegtár későbbi, 2006-os verziója már 187 millió szövegszót tartalmazott. Ez az adatbázis több alkorpuszt is tartalmaz különböző témakörökben, az alkorpuszok

adatait a 2.4. táblázatban adom meg. A szótárépítésnél minden alkorpuszt felhasználtam. Az

2.4. táblázat. A Magyar Nemzeti Szövegtár ($MNSZ_{2006}$) felépítése. Az adatok millió szóban értendők. Forrás: `http://corpus.nytud.hu/mnsz`

	magyarországi	szlovákiai	kárpátaljai	erdélyi	vajdasági	összesen
sajtó	71,0	5,7	0,7	5,5	1,5	84,5
	A sajtószövegek a korpusz majdnem felét teszik ki. Széles skáláját mutatják be a nyelvi változatoknak, vertikálisan és horizontálisan is.					
szépirodalom (DIA)	35,5	1,4	0,4	0,8	0,2	38,2
	2005. őszén készült el a Digitális Irodalmi Akadémia anyagainak teljes feldolgozása. Ez adja a magyarországi szépirodalmi alkorpuszt.					
tudományos	20,5	2,3	0,7	1,6	0,3	25,5
	A magyarországi tudományos szövegek a Magyar Elektronikus Könyvtárból származnak.					
hivatalos	19,9	0,2	0,3	0,6	0,1	20,9
	Ezek a szövegek szabályokat, törvényeket, rendeleteket, parlamenti vitákat tartalmaznak.					
személyes	17,8	-	0,4	0,4	0,1	18,6
	Ez az alkorpusz internetes fórumok (az index.hu fórumainak és több kárpátaljai fórum) beszélgetéseit tartalmazza. Ez a nyelvi változat azért fontos, mert ez áll a legközelebb a spontán nyelvi kommunikációhoz, bizonyos esetekben nagyon hasonlít a beszélt, élő kommunikációhoz.					
összesen	164,7	9,5	2,5	8,9	2,0	187,6

algoritmus tesztelését a szótárépítésből kivont részhalmazzal végeztem, amely szintén minden alkorpuszból tartalmazott elemeket. Az ellenőrzést az elektronikus levelekhez hasonlóan végeztem, a teszthalmazról eltávolítottam az ékezeteket, majd az ékezetesítés lefuttatása után megszámoltam az eltérő szavak számát. Ebben az esetben azt a szót tekintettem hibásnak, amelyben legalább egy betűeltérés fennállt. Az ékezetesítést az $MNSZ_{2006}$-en keresztvalidációval vizsgálva a szóalapú helyesség 96%-os volt.

2.2.4. Az alapalgoritmus tulajdonságai

Az ékezetesítő alapalgoritmus a szótárépítésben szereplő lexikai egységekre nyelvtanilag helyes alakokat ad. Az algoritmus előnye, hogy csak az előkészítő fázisban (szótárépítés) használ nagyméretű szövegadatbázist, de a működéshez kialakított tudástár (maga az ékezetesítő szótár) már kis méretű és gyors keresést biztosít. Hátránya, hogy minden esetben csak a leggyakoribb alakra ékezetesít, ami hibákat eredményez. További hátrány, 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. Az eljárás magyar szabadalom, lajstromszáma: 226740 P 00 03443.

2.3. Az alapalgoritmus továbbfejlesztése

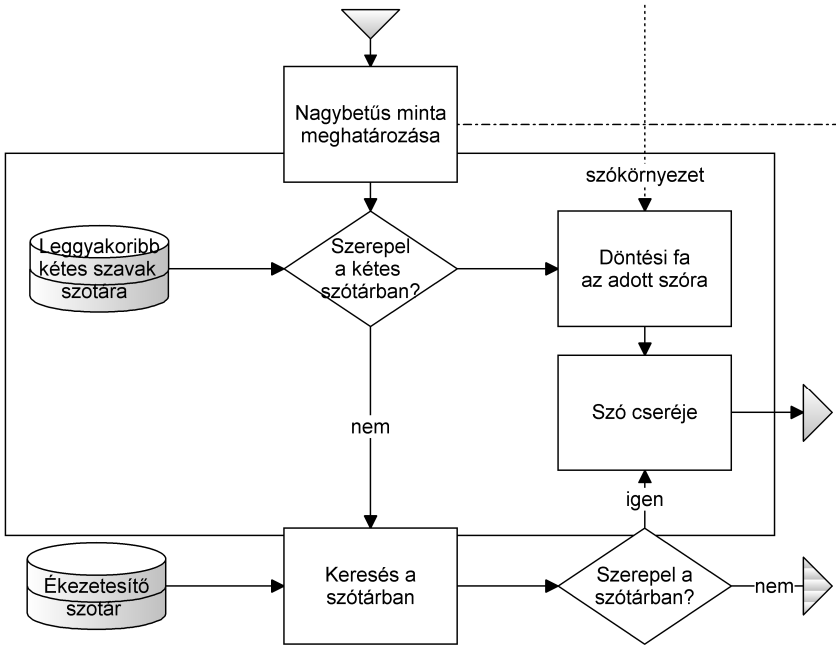
A továbbfejlesztéssel a célom az volt, hogy az alapalgoritmus két hátrányos tulajdonságát kiküszöböljem. Az egyik ilyen tulajdonság a kétes szavak egyszerűsített kezelése, a másik az általánosító képesség hiánya.

2.3.1. A szótár alapú algoritmus hibájának csökkentése

A szótár alapú algoritmus a kétes eseteket (több alakú szavakat) nem kezeli, minden esetben a leggyakoribb változatra dönt. A Magyar Nemzeti Szövegtár 2006-os verzióján vizsgáltam, hogy az ilyen több alakú szavak milyen gyakorisággal fordulnak elő. Mérésem szerint 21 757 ilyen különböző eset fordul elő a szövegtárban. Ezek az esetek 29 millió szövegszóból származnak, ami a teljes szövegtár 15,5%-a. Az alapalgoritmus az esetek nagy részében jól dönt, hibás döntés átlagosan az esetek kevesebb mint 10%-ában fordul elő. Azonban a konkrét szótól függően ezek a hibák nagyon zavaróak lehetnek, a hibás döntés miatt a mondat más értelmet kaphat. Ilyen például a következő mondat: „*Megvetette az ágyát.*” - „*Megvetette az agyát.*”. Az alapalgoritmus hibáinak elemzése során összeszámoltam, hogy hány szó esetén fordul elő hibás döntés. Megállapítottam, hogy a leggyakoribb kétes esetek jobb kezelésével a hibák száma jelentősen csökkenthető az alapalgoritmus módszeréhez képest. Az eredmény az, hogy a 21 757 kétes esetből a 100 leggyakoribb felelős a hibás döntések 60%-áért. A 100 leggyakoribb esetet az A. függelékben felsorolom.

Ezeknek a kétes eseteknek a kezelésére olyan megoldást dolgoztam ki, amely a leggyakoribb kétes eseteket egy döntési fa segítségével egyértelműsíti, a környezet alapján hoz döntést a lehetséges variációk közül. Az eljárásom a (Mihalcea–Nastase 2002) módszert fejleszti tovább, ami gépi tanulást használ fel karakter szinten. Mihalcea a gépi tanulást több nyelvre és 5 karakteres környezetre alkalmazta. A módszerét magyar nyelvre is kipróbálta, de csak kisméretű és homogén (csak irodalmi szövegek) segítségével tanította és ellenőrizte. A módszerét továbbgondolva adaptáltam azt a szó alapú feldolgozásra a következők szerint: A kétes esetek listája és azok gyakorisága rendelkezésre áll a szótárépítés eredményeként. A kapott listát sorba rendeztem aszerint, hogy átlagosan mekkora lenne a hibás döntések száma a szótár alapú ékezetesítés esetén. Például a „*meg*” szó 796 688-szer fordult elő az $MNSZ_{2006}$-ben, a „*még*” szó pedig 541 760 esetben. Az alapalgoritmus a „*meg*” szót minden esetben változatlanul hagyná, a „*még*” szót egyszer sem választaná, tehát több mint félmillió esetben hibás döntést hozna.

A leggyakoribb hibás döntést adó kétes szavakra egy J48-as döntési fát építettem, amely a szó környezete alapján egyértelműsíti, hogy a több lehetséges ékezetes alak közül melyik a legvalószínűbb. A J48-as a C4.5 döntési fa szabadon felhasználható Java alapú implementációja (Quinlan 1993). A C4.5 döntési fa pedig az ID3 algoritmus kibővítése. A választásom azért esett erre a döntési fára, mert képes volt kezelni a rendelkezésre álló nagy mennyiségű tanító adatot. A használt döntési fa lehetővé tette a gyors működést, ami infokommunikációs rendszerekben alapkövetelmény. Továbbá az algoritmus kevésbé érzékeny a tanulási példák hibáira, így zajos tanulás mellett is használható. A döntési fának az alapalgoritmusba való beillesztését a 2.5. ábrán mutatom be. Az ékezetesítő szótárban történő keresés előtt megállapítja az algoritmus, hogy az adott szóra készült-e döntési fa. Ha igen, akkor a döntési fa bejárásával választja ki az algoritmus a megfelelő szóalakot, amelyhez az adott szó környezetét használja fel. A környezet mérete 20 karakter (10 előtte, 10 utána). A környezet méretének meghatározása során Mihalcea eredeti kísérleteit nem tudtam felhasználni, mert a munkája során – infokommunikációs rendszerek szempontjából – nem reprezentatív adatbázison dolgozott (irodalmi szövegek). A vizsgálandó kétes szavak környezetének meghatározásához egy másik kutatásom adatait használtam fel [J5]. A környezet méretét a magyar nyelv statisztikai tulajdonságait figyelembe véve úgy választottam meg, hogy átlagosan a magyar szavak 90%-a beleférjen ebbe a méretbe. Ezt a következő fejezet 3.3.-as ábrája segítségével



2.5. ábra. A szótár alapú alapalgoritmus döntési fával való kiegészítése. A kiegészítést a bekeretezés jelöli.

határoztam meg. A magyar szóalakok súlyozott görbéjét kiintegrálva 10 karaktert kapunk. Ezt csak az adott mondaton belül értelmezem, mondathatár esetén a túlnyúló környezet szóköz karakterekkel van kitöltve. A döntési fa építéséhez a Rapidminer programot használtam (`http://www.rapidminer.com`).

Az ismertett döntési fát felhasználó algoritmus jellegében illeszkedik az alapalgoritmushoz, mert a nagy számításigényű feldolgozási műveleteket külön választottam a gyors alkalmazási eljárástól. A szótárépítés memória igényes, és nagy – több 10 millió szövegszó – méretű szöveg feldolgozása szükséges. Az alkalmazási fázisban viszont már csak egy kisméretű döntési fa bejárása történik, ami gyorsan elvégezhető. Az algoritmus egyik korlátját az jelenti, hogy rendelkezésre kell állnia nagyméretű tanító adatnak minden egyes kétes esetre. Az algoritmus előnye, hogy a szótár alapú eljáráshoz képest már korlátozott általánosító képességgel is bír a kezelt kétes esetekre, mert olyan környezetben is döntést tud hozni, amelyekre tanító adatok nem állnak rendelkezésre. Ez az általánosító képesség azonban még nem kezeli az olyan adatokat, amelyek maguk nem szerepelnek a tanító adatokban, erre a következő részben leírt továbbfejlesztés nyújt megoldást.

Az ismertetett döntési fa beépítésével az alapalgoritmusba olyan eljárást hoztam létre, amelynek segítségével a Magyar Nemzeti Szövegtár ($MNSZ_{2006}$) anyagával tanított algoritmussal a 100 leggyakoribb kétes esetet kezelni lehetett, ami az összes kétes eset 60%-a. A kiegészítést a 2.2.3.2. fejezetben ismertetett módon teszteltem. Ezeket a leggyakoribb kétes eseteket 93%-ban (2.5.táblázat) helyesen ékezetesítette az algoritmus [C10]. Ebben a vizsgálatban is szóhelyességet mértem.

2.3.2. Általánosító képesség szótár alapú újraékezetesítő alapalgoritmushoz

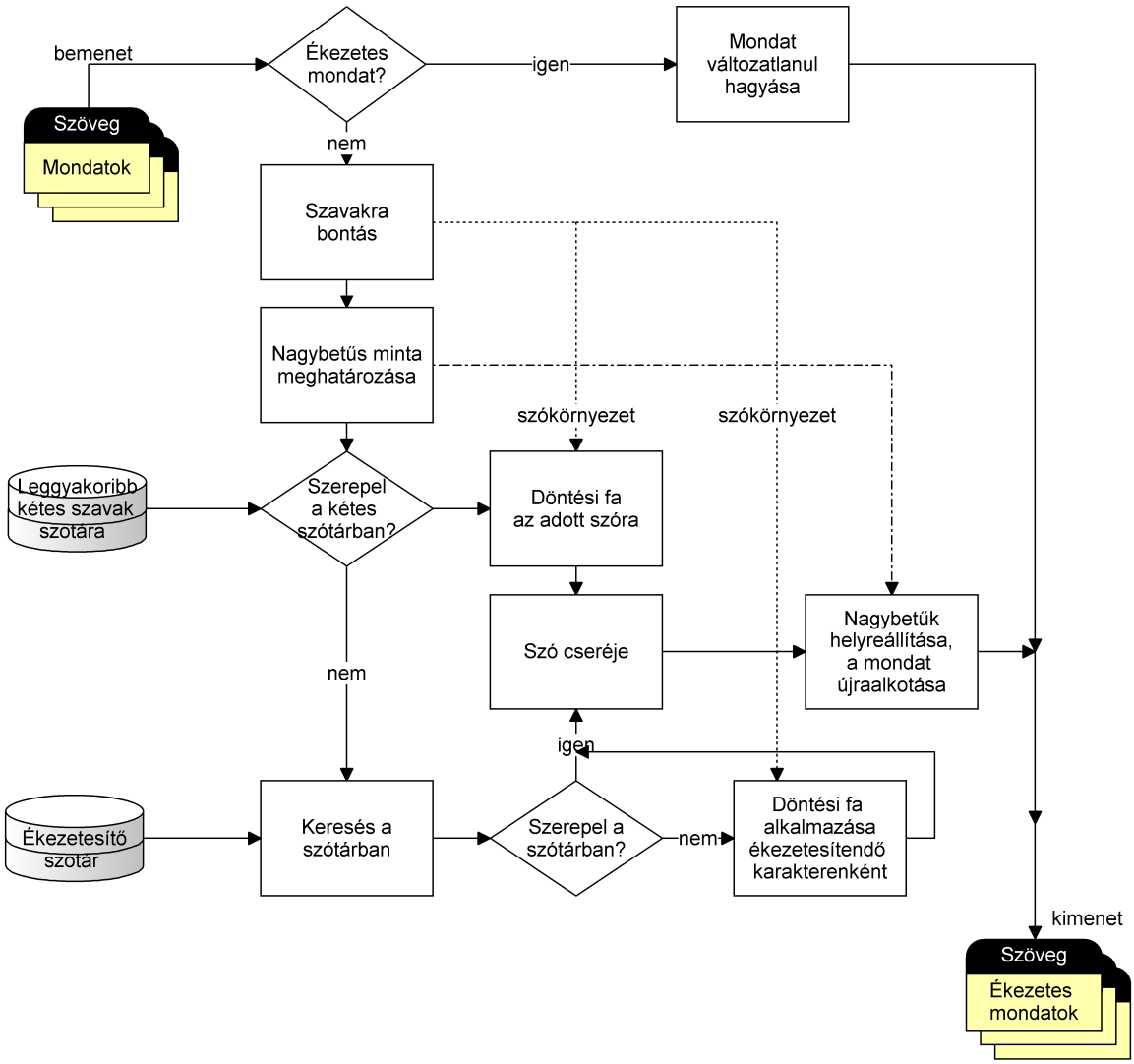
A döntési fával kiegészített ékezetesítő algoritmus együttesen sem kezeli az ékezetesítés során előforduló összes ékezet nélküli szót. Mihalcea–Nastase (2002) módszerét módosítottam a maradék esetekre – az ékezetesítő szótárban és a leggyakoribb kétes szavak között nem szereplő esetekre – karakter környezet alapján működő döntési fát hoztam létre. Mihalcea módszerét ebben az esetben kisebb mértékben módosítottam, mint a kétes szavak kezelésénél. A kétes szavak döntéséhez képest ebben az esetben tehát nem a szavak, hanem a karakterek környezetére építettem döntési fát, hasonlóan az eredeti módszerhez. A karakterkörnyezet méretének meghatározása azonban itt is másként történt, mint Mihalecea eredeti megoldásában. A 10 karakteres méret itt is a magyar szavak hosszából lett meghatározva a kétes szavaknál már megadott gondolatmenet szerint. Szintén J4.8-as döntési fát építettem a lehetséges ékezetesítendő karakterek 20 méretű környezete (10 előtte, 10 mögötte) alapján. Például a következő mondatban az „*u*” hang környezete: „*A vona[t balatonf]u[zfon all m]eg.*” A környezet közepén a következő magánhangzók állhatnak a magyar nyelv esetében: „*a, e, i, o, u*”. Az így kapott döntési fákat – magyar esetében 5 db-ot – azoknak a szavaknak a betűire alkalmazom, amelyeket az alapalgoritmusnak és a leggyakoribb kétes eseteket kezelő megoldásnak nem sikerült feldolgozni. Az ékezetesítő algoritmus kibővített folyamatábrája a 2.6. ábrán látható.

Az így kibővített algoritmust a Magyar Nemzeti Szövegtár különböző alkorpuszain teszteltem. A magyar nyelv kétes eseteire vonatkozó tesztelési eredmények a 2.5. táblázatban láthatók. Az első sorban a Digitális Irodalmi Akadémiára (DIA) vonatkozó, a másodikban az $MNSZ_{2006}$ személyes alkorpuszára, a harmadik sorban pedig a teljes szövegtárra vonatkozó eredményeket láthatjuk. A korpuszok részletes adatai korábban a 2.4. táblázatban adtam meg. Az első oszlopban a kétes esetek szótár alapú alapalgoritmusának eredményei vannak. Ezek az értékek azt mutatják, amikor mindig a leggyakoribb kétes esetet választjuk, tehát az alapalgoritmus a kétes eseteket 75%-os szóhelyességi aránnyal tudja ékezetesíteni. A célom az volt, hogy ezeknek az eseteknek a külön kezelésével ezt az értéket javítani tudjam. A második oszlopban a kigyűjtött 100 leggyakoribb kétes eset döntési fa alapú megoldásának értékei szerepelnek. Az utolsó oszlopban az összes kétes eset kezelésének értékét látjuk, amelyben szerepelnek a leggyakoribb kétes esetek és az olyan kétes esetek eredményei is, amelyben az ékezetesítendő betűket a környezetük alapján kezeltem. Az értékek szóhelyességet mutatnak. Az alapalgoritmus továbbfejlesztésére alkalmazott döntési fák tehát a kétes esetek esetében 83%-os pontosságot értek el, tehát javult az alapalgoritmushoz képest.

2.5. táblázat. Kétes esetekre vonatkozó teszteredmények

	Szótáralapú	100 leggyakoribb kétes	Kétes esetek
$MNSZ_{2006}$ DIA	75,9%	92,7%	82,4%
$MNSZ_{2006}$ személyes	71,4%	88,5%	80,8%
$MNSZ_{2006}$ (összes)	75,2%	92,9%	82,9%

További előnye a továbbfejlesztésnek, hogy az algoritmus általánosító képességgel bír. Az alapalgoritmus csak olyan szavak ékezetesítésre volt alkalmas, amely a szótárépítő szövegben szerepelt. Ha szótáron kívüli elem fordult elő, akkor az ékezetesítést nem hajtotta végre. Ezzel szemben a továbbfejlesztett algoritmus a szótáron kívüli szavak esetében is tud ékezetesítést végezni, bármilyen szóra alkalmazható, olyanra is, amely a tanító adatbázisban nem szerepelt.



2.6. ábra. Az ékezetesítő algoritmus döntési fákkal kibővített folyamatábrája

Ez lehetővé teszi, hogy a nyelvben ritkán használt szavak, vagy nem gyakran használt ragozott alakok esetében is az ékezetek helyreállítása megtörténjen. A magyar nyelvre bemutatott eredmények felhasználhatóak más nyelvekre is, azonban az általánosító képesség kiaknázása szempontjából a következőkre figyelemmel kell lenni: a magyar esetében az általánosító képességet biztosító döntési fákat a kétes szavakkal és azok környezetével tanítottam. Ezt azért tehettem meg, mert az alapalgoritmus szótára nagy, jól lefedi a nyelvet, így az ismeretlen (szótáron kívüli) elemek előfordulása várhatóan kicsi, nem szükséges külön döntési fa ezekhez. Amennyiben egy nyelvben nem áll rendelkezésre olyan szöveganyag, amely használata esetén a szótáron kívüli elemek előfordulása kicsi, akkor az általánosító képességet biztosító döntési fát nem a kétes szóalakból kell építeni, hanem a teljes szöveganyagból.

A döntési fákkal a továbbfejlesztés a szótár alapú ékezetesítő eljárás hibáinak 60%-át javította általános szövegkörnyezetben [C10], így a teljes algoritmus az MNSZ_{2006}-en (keresztvalidációval) vizsgálva 97,4%-os pontosságot ért el.

2.4. Összegzés

Kidolgoztam egy ékezetesítő alapalgoritmust, amely segítségével olyan szövegek is felhasználhatók a gépi beszédkeltés során, amelyekben az ékezetek hiányoznak. Az alapalgoritmushoz meghatároztam a szótárépítés folyamatát és a felépített különböző szótárakkal kísérleteket végeztem. Az alapalgoritmust döntési fákkal egészítettem ki, amely csökkentette a hibás ékezetesítések számát. Az elért helyes szavak aránya – a szótárépítés forrásától és a vizsgált szövegek tematikájától függően – 95,2–97,4% (2.6. táblázat). Az

2.6. táblázat. Eredmények összefoglalása

Elnevezés	tanító adatbázis	tesztelő adatbázis	tesztelés formája	eredmény [%]
Szótáralapú	$MNSZ_{1999}$ + MEK	elektronikus levelek	független teszthalmaz	95,2
Szótáralapú	$MNSZ_{2006}$	$MNSZ_{2006}$ (tanítóból kivett)	keresztvalidáció	96
Szótáralapú + kiegészítés	$MNSZ_{2006}$	$MNSZ_{2006}$ (tanítóból kivett)	keresztvalidáció	97,4

algoritmust és a kísérleteket is magyar nyelvre mutattam be, de az eljárás alkalmazható más ékezetes betűket használó nyelvekre is.

3. fejezet
Magyar és idegen nyelvű írott szövegek vizsgálata a gépi beszédkeltés szempontjaiból

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 felhasználhatók beszédszintetizátorok alapját képező beszédadatbázisok előállításához és a szintetizálási algoritmusok teszteléséhez. Több nyelv adatainak összehasonlítása pedig segítséget ad az idegen nyelvű algoritmusok magyar nyelvű adaptálásához és az eredmények nemzetközi összehasonlításához. A magyar nyelvre az internet világa előtt nem állt rendelkezésre olyan mennyiségű írott szöveg, amely könnyen feldolgozható volt és közel tükrözte az aktuális nyelvhasználatot. Nemzetközileg a nagy világnyelvek vizsgálata megkezdődött, de a magyar nyelv ilyen jellegű feltérképezése és összevetése más nyelvekkel a hazai kutatótársadalomra maradt. Ennek a feladatnak – a teljes nyelv szempontjából apró, de a beszédtechnológiai kutatások szempontjából fontos – a szavak statisztikájával foglalkozó részével kezdtem az ilyen irányú kutatásaimat. A beszédtechnológiában sok algoritmus kapcsolódik a szó méretű elemekhez. Ennek egyik oka például az, hogy a szavak szintjén szupraszegmentális – például hangsúly, hangerő, dallammenet – jellemzők megadása kellően részletes leírást adhat a beszéd szintetizáláshoz. A szavak ugyanakkor eléggé kisméretűek ahhoz, hogy a hozzájuk rendelt kisebb egységek – például szótagok vagy hangok – ne alkossanak túl összetett struktúrát, tehát könnyen lehessen kezelni őket. A magyar szavak statisztikai vizsgálatát indokolja az is, hogy a legtöbbet vizsgált angol nyelvben a nyelv tulajdonságai miatt sok a szó alapú megközelítés. Ezek magyar nyelvre való átültetési lehetőségeinek megvizsgálása a hazai kutatók feladata.

A statisztikai vizsgálatok felhasználhatóak például a kötött szótáras beszédfelismerés magyar nyelvre történő felhasználhatóságának vizsgálatára, vagy az elemkiválasztásos beszédszintetizátorok szövegkorpuszának meghatározásához. Egy 20 ezer szavas rendszer az angol nyelv esetében szinte általános tematikák lefedésére alkalmas mind beszédszintézis, mind beszédfelismerés terén. Vajon magyar nyelv esetében egy ilyen rendszer mire képes? Alkalmazható-e az angol nyelvre publikált módszer a magyarra? Hány szó szükséges a magyar nyelvben ugyanahhoz a tematikához? Hogyan lehet egy szövegről eldönteni, hogy mennyire alkalmas beszédszintézishez? Ezekre és az infokommunikációs rendszerek két fő technológiájához – a beszédszintézishez és a beszédfelismeréshez – kapcsolódó hasonló kérdésekre adnak választ a következőkben ismertetett kutatásaim.

3.1. A magyar nyelvre jellemző alapvető szó-gyakorisági eloszlások

A magyar nyelvben – annak ragozó tulajdonsága miatt – a nyelvtanilag helyes és értelmes szóalakok száma rendkívül nagy, különböző becslések milliárdos nagyságrendet határoznak meg (például egyetlen igének 1000 ragozott alakja lehet (Prószéky 1988) és 1 millió lexémára vonatkoztatjuk (Kenesei et al. 1984)). A valóságban használt szóalakok száma ennél kisebb. Mindezen felül sok idegen szó is előfordul nyelvünkben.

Szóalakok vizsgálatánál szónak tekintem azokat a folytonos karaktersorozatokat, amelyek a magyar ábécé betűiből és a magyar írott szövegekben előforduló idegen betűkből állnak. A szavak vizsgálatánál szótövezést és egyéb változtatást nem végeztem, egy szó különböző ragozott alakjai különböző szavaknak számítanak.

3.1.1. Felhasznált adatbázisok

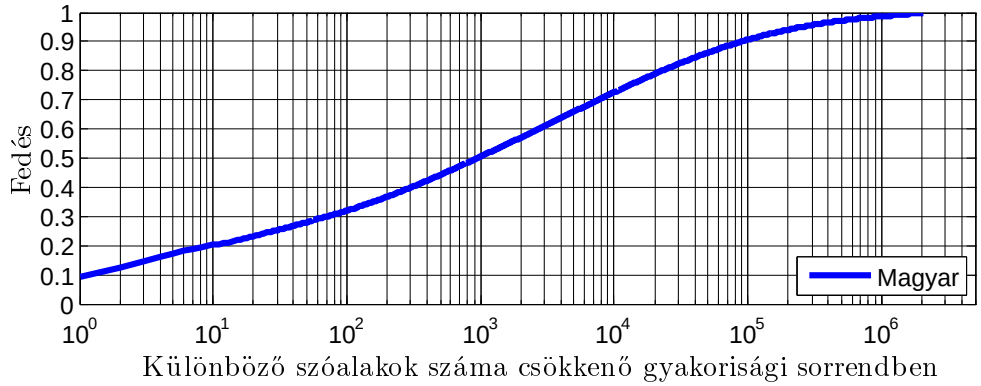
A szóalakok számának és azok eloszlásának meghatározásához a következő elektronikus források szövegeit használtam fel: a Magyar Elektronikus Könyvtár magyar nyelvű művei, a Digitális Irodalmi Akadémia művei, online folyóiratok cikkei és a Magyar Nemzeti Szövegtár. A források egyesítve kb. 80 millió szövegszavas gyűjteményt tesznek ki. A 3.1. táblázatban látható az adatbázisok mérete és a különböző szóalakok száma. A források közül a Magyar

3.1. táblázat. Adatbázisok főbb adatai

jelölés	Adatbázis neve	Szavak száma	Különböző szóalakok száma
DIA	Digitális Irodalmi Akadémai (2002)	13 294 786	835 809
MH	Magyar Hírlap	2 054 777	196 965
MH2001	Magyar Hírlap 2001	5 478 451	368 884
MH2002	Magyar Hírlap 2002	5 636 642	365 868
MN	Magyar Nemzet	4 373 412	345 657
MN2001	Magyar Nemzet 2001	4 039 096	320 596
MN2002	Magyar Nemzet 2002	12 050 613	561 407
HVG	HVG online	4 091 732	311 578
MEK	Magyar Elektronikus Könyvtár	6 799 701	522 431
$MNSZ_{2002}$	Magyar Nemzeti Szövegtár (2002)	21 139 882	691 159

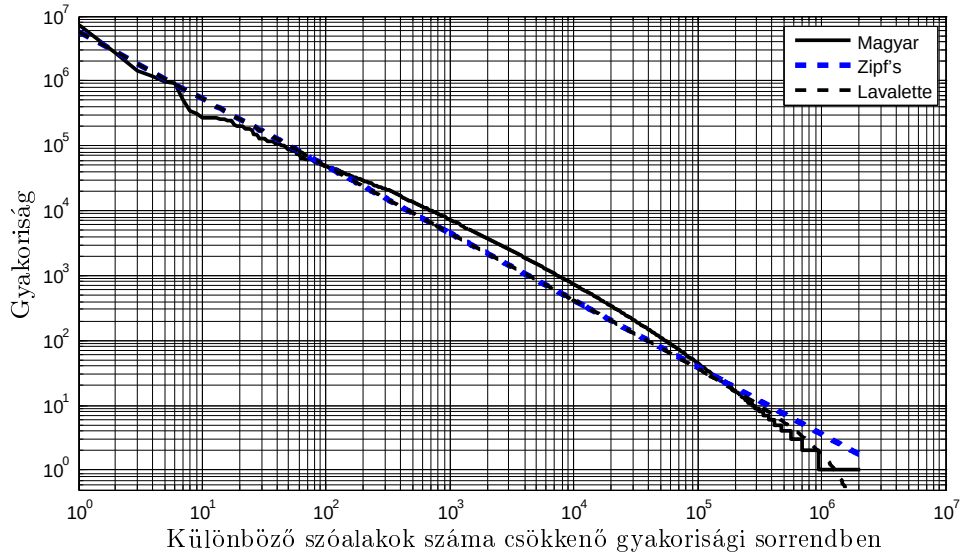
Nemzeti Szövegtár (MNSZ) a legnagyobb, amelynek 2002-es változata még nem tartalmazta a Digitális Irodalmi Akadémiát (DIA). A DIA hozzávetőlegesen az $MNSZ_{2002}$-nek a kétharmadát teszi ki, mégis 145 000-rel több különböző szóalakot tartalmaz, ami arra utal, hogy az irodalmi nyelvezet változatosabb, mint az $MNSZ_{2002}$ köznyelvi gyűjteménye. A feldolgozott források három online újság (MH, MN, HVG) cikkeit is tartalmazzák. A napilapok esetében az adatok éves bontásban találhatók a táblázatban. A szövegek feldolgozásakor törekedtem arra, hogy kiszűrjem az ismétlődő részeket, amelyek hírekben gyakran előfordultak. Például az online újság felépítésétől függően a rövid összefoglaló a cikk oldalán többször is előfordult. Figyelmet fordítottam arra is, hogy a nem szövegszerű részeket – például totó eredmények, tőzsde árfolyamok – eltávolítsam. A Magyar Elektronikus Könyvtár (MEK) anyagaiból a magyar nyelvű írásokat dolgoztam fel. A MEK-ben szereplő művek sok szerzőtől származnak és a témaköreik is változatosak. A forrásokat egyesítve közel 80 millió szövegszavas gyűjteményt hoztam létre, amely 2 034 634 különböző szóalakot tartalmazott.

3.1.2. Gyakorisági megfigyelések



3.1. ábra. A magyar nyelv fedési görbéje 80 millió szavas adatbázis alapján

A 3.1. ábrán adom meg a megvizsgált szövegadatbázis fedési görbéjét. Ez a görbe azt adja meg, hogy az első n leggyakoribb szó a teljes állomány hányad részét fedi le. A fedési görbe segítséget nyújthat abban, hogy különböző elvárt fedési értékekhez megbecsüljük a szükséges szavak számát, illetve hogy adott számú szóalakkal mekkora fedés érhető el. Például, ha egy algoritmus 1000 szót tud kezelni, akkor a görbe alapján a legjobb esetben a magyar nyelvben a szavak 50%-át tudjuk lefedni. Egy másik megjelenítési módja a gyakorisági adatoknak a 3.2. ábrán látható. A 80 millió szavas adatbázis adatait folyamatos vastag vonallal jelöltem az ábrán. A vízszintes tengely azonos a fedési görbe vízszintes tengelyével, a függőleges tengelyen pedig a gyakorisági adatok találhatók logaritmikus skálán.



3.2. ábra. Összefüggés a szavak gyakorisága és a sorrendjük között

A gyakorisági adatokat összevetettem Zipf – nyelvek vizsgálatakor gyakran alkalmazott – törvényével (Li 1992), amelyet az ábrán vastag szaggatott vonallal szintén jelöltem. A gyakorisági adatok követték Zipf törvényét ($f(r) = C \cdot r^{-b}$) ahol C egy normalizációs

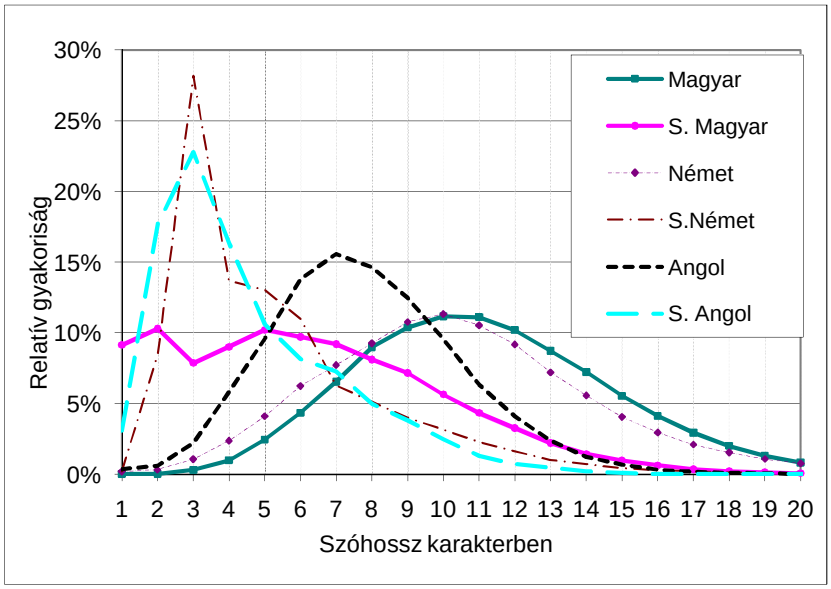
konstans, a b általában 1 körüli érték. Az ábrán mindkét tengely logaritmikus beosztású, így az exponenciális lefutású görbe egyenesként ábrázolható. A gyakorisági adatokra illesztett függvény konstansai a következők: $C = 10^{6.7573}$, $b = 1.033$. A Zipf által meghatározott görbe a kis gyakoriságú elemek esetén eltér a mért adatoktól, ennek kiküszöbölésére a Lavalette formulát alkalmaztam (Popescu 2003), amely a 3.1. egyenlettel írható le. Az alkalmazott konstansok azonosak a Zipf formula konstansaival.

$$f(r) = C \cdot \left(\frac{r_{max} r}{r_{max} - r + 1} \right)^{-b} \tag{3.1}$$

A meghatározott Lavalette görbe szintén a 3.2. ábrán látható vékony szaggatott vonallal.

3.1.3. A magyar szavak karakter szerinti eloszlása

A szó alapú algoritmusok vizsgálata esetén a szavak hossza és azok eloszlása is fontos adat. A 80 milló szavas szöveg elemzése alapján megállapítottam a különböző szavak hosszának



3.3. ábra. Szóhossz karakterben különöző nyelvekre (S=súlyozott)

eloszlását. Ez a 3.3. ábrán látható folytonos vonallal és rajta kicsi négyzettel van jelölve. Ha a szavak hosszát a gyakoriságukkal súlyozom, akkor az eloszlás a rövidebb szavak irányába csúszik el (folytonos vonallal és rajta ponttal jelölve). A különböző szavak átlagosan 11,2 karakter hosszúak, a súlyozott átlag pedig 6,2 karakter.

3.2. Az angol, a német és a magyar nyelv szó-gyakorisági adatainak összehasonlítása

3.2.1. Felhasznált adatbázisok

3.2. táblázat. Az adatbázisok főbb adatai

jelölés	nyelv	Adatbázis neve	Szavak száma	Különböző szavak száma
Magyar	magyar	Magyar Elektronikus Könyvtár 50k<	2 508 371	293 928
Német	német	Gutenberg projekt	3 100 225	148 989
Angol	angol	Magyar Elektronikus Könyvtár angol művei	3 458 861	62 141
Angol (BNC)	angol	Britsh National Corpus	98 661 019	845 147
Magyar2	magyar	Összesített magyar adatbázis	78 959 087	2 034 634

A három nyelv összehasonlításához különböző szövegkorpuszokat dolgoztam fel. A felhasznált szövegek főbb adatai a 3.2. táblázatban olvashatók. A három nyelvhez azonos, közelítőleg 20 millió karakter méretű szövegkorpuszt használtam. A magyar nyelvhez (Magyar) a Magyar Elektronikus Könyvtár 50 000 karakternél nagyobb műveit használtam fel. A német nyelvhez (Német) a Gutenberg projekt anyagát használtam. Az angol nyelvű (Angol) anyagokat a Magyar Elektronikus Könyvtár angol műveiből állítottam össze. A szövegek változatos tartalmúak, irodalmi anyagokat, híreket és egyéb műveket tartalmaznak.

Nagyobb méretű korpuszok elemzéséhez az angol nyelvű (Angol BNC) British National Corpus 89 millió szövegszót tartalmazó írott korpuszát használtam fel. A magyar nyelv (Magyar2) esetében a 3.1. fejezetben összeállított szövegkorpuszt használtam fel.

A motivációja annak, hogy különböző méretű adatbázisok segítségével is elvégeztem ugyanazokat az összehasonlításokat az, hogy így meg tudtam vizsgálni az összehasonlító módszer adatbázistól való méretfüggését is.

A korpuszok vizsgálatánál a szóalakok definíciója megegyezik a magyar nyelvre készített statisztikáknál használttal.

3.2.2. Gyakorisági megfigyelések és összehasonlításuk

A 3.4. ábrán látható a három kisméretű (Magyar, Német, Angol) és a két nagyméterű (Magyar2, Angol(BNC)) szövegkorpusz fedési görbéje. Az angol nyelv görbéje balra található a német görbétől, tehát kevesebb szóval lehet ugyanakkora fedést elérni. A magyar nyelv görbéje az angol és a német görbétől jobbra található, a magyar esetén ugyanakkora a fedéshez több szóra van szükség. A 3.3. táblázatban látható három példa, amely megadja, hogy az adott fedési hányadhoz a három nyelv esetében hány szó szükséges.

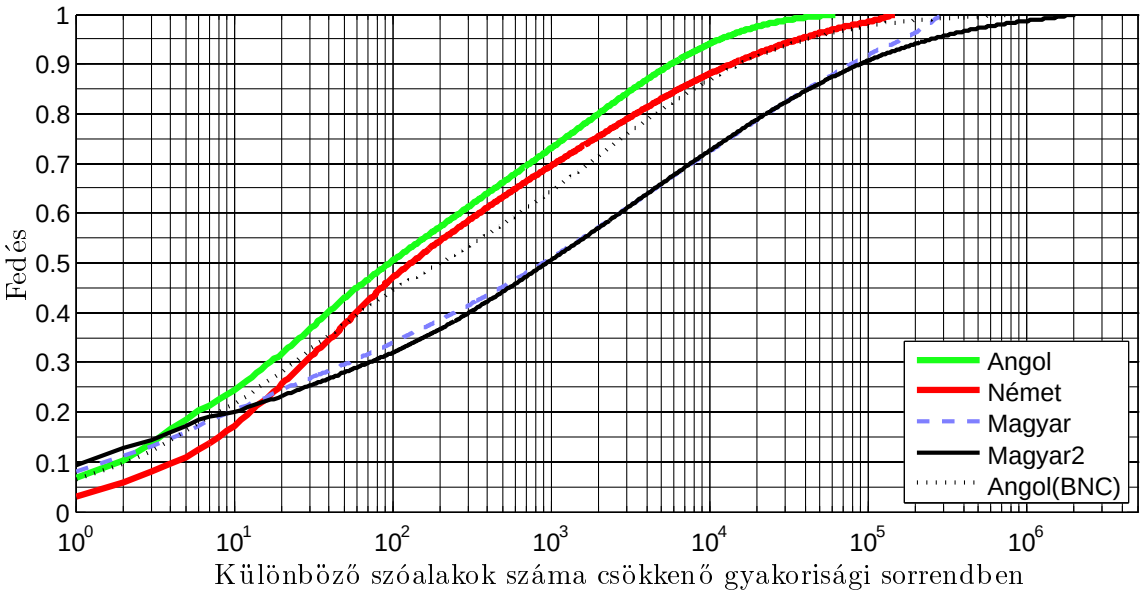
A táblázatból látható például, hogy 90%-os fedés esetén a magyar nyelvhez több mint 12-szer annyi szó szükséges, mint angol esetén, és a némethez képest is majdnem 5-ször annyi szó kell. A kapott adatok elemzésének másik lehetséges formája, amikor az adott szószám alapján hasonlítjuk össze a különböző nyelvek fedési adatait. Erre láthatunk példákat a 3.4. táblázatban. Azonos szószám mellett – nagyobb mint 20 szó – a magyar nyelv fedése a legkisebb, míg az angolé a legnagyobb. A két nagyobb (Magyar2, Angol(BNC)) szöveg esetén a fedési görbék a

3.3. táblázat. A szükséges 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

3.4. táblázat. Korpusz fedés adott leggyakoribb 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%



3.4. ábra. A vizsgált nyelvek fedési görbéi

vártnak megfelelően jobbra csúsznak, az angol esetében nagyobb mértékben, a magyar esetében kevésbé. A magyar görbe esetén inkább a görbe végén van jelentősebb különbség. A két nyelv között a számszerű különbség változott, de az egymáshoz viszonyított jelleg nem.

3.2.3. Azonos tematikájú szöveg

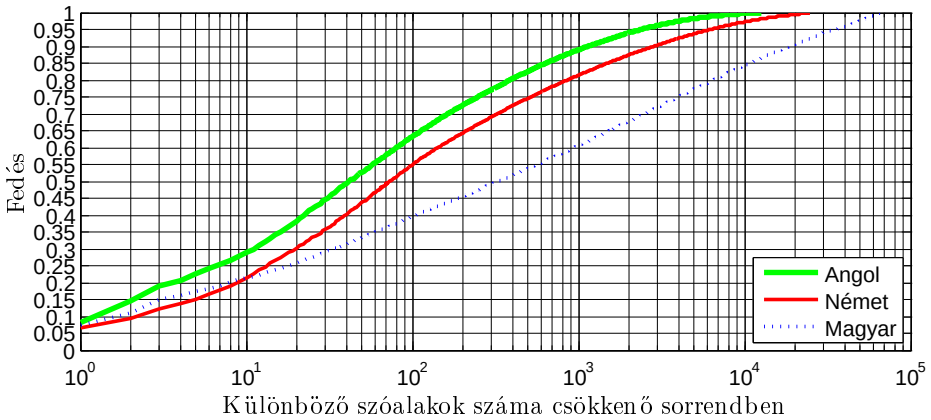
Három nyelv Bibliáinak adatait a 3.5. táblázatban mutatom be. Az angol esetében két Bibliát is megvizsgáltam, de a közel azonos fedési görbéjük miatt csak az egyik (KJV) adatait jelenítettem meg.

3.5. táblázat. A Bibliák főbb adatai

jelölés	nyelv	Adatbázis neve	Szavak száma	Különböző szóalakok száma
KJV	angol	King James Bible	788 657	12 543
ASV*	angol	American Standard Version	775 409	12 088
ELB	német	Elberfelder Bibel	740 735	24 719
KAT	magyar	Katolikus Biblia	607 763	65 941

* A grafikonon nincs megjelenítve.

A kapott fedési görbék a 3.5. 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 3.3. táblázatban látható 90%-os fedését összevetve a 3.5. ábra 90%-os vonalával, látható, hogy az angol és a német nyelv közötti háromszoros arány megegyezik, és az angol és a magyar közötti egy nagyságrendi különbség is.



3.5. ábra. A három különböző nyelvű Biblia fedési görbéi

3.2.4. *Korpuszközi eltérések*

A görbék megadásakor már utaltam a görbék közötti kapcsolatokra, elsősorban a nyelvek közötti szempontból, most a különböző korpuszok közötti eltéréseket elemzem. A korpuszokhoz tartozó fedési görbék számszerű adataiban természetesen jelentős eltérés tapasztalható, de a görbék jellege minden esetben szinte azonos. A lineáris-logaritmikus tengelyen ábrázolt fedési görbék, többnyire rendelkeznek egy laposabb kezdettel, amely az első 10–100 szóig tart, majd egy meredekebb szakasszal, amely 0,85–0,9-es fedési értékig tart. A korpuszok közötti eltérés leginkább a görbe utolsó szakaszán figyelhető meg. A görbék egy része az 1,0-ás fedési értékhez lassan, a másik része meredeken konvergál. A görbe ilyen jellegű viselkedését az határozza meg, hogy a nagyon ritkán előforduló szóalakok száma hogyan viszonyul a szövegkorpusz méretéhez. Meredek görbe vége a kisméretű magyar korpusz és a magyar Biblia esetén figyelhető meg, míg a többi korpusz lassan éri el az 1,0-ás fedési értéket. A meredek vég ritkán előforduló szavak hiányára utal.

3.2.5. *Beszédtechnológiai módszerek összehasonlítása*

A beszédfeldolgozásban alkalmazott módszereket a legtöbb esetben az alkalmazási technológia szűk keresztmetszeteihez – tároló kapacitás, processzor sebesség, csatorna kapacitás – kell optimalizálni, tehát sok esetben csak korlátozott méretű szótárakat lehet alkalmazni. Ezek leggyakrabban szavakat tartalmaznak vagy szavak számával jól jellemezhetők. A szótárak méretéhez – a kutatásomban meghatározott adatok alapján – meg lehet becsülni, hogy

nyelvenként milyen fedés érhető el. A minimális fedés konkrét értéke az alkalmazástól függ. Például hang- és videóanyagok keresését szolgáló címkézéshez 40–50%-os fedés mellett is használható eredményt ad egy beszédfelismerő, míg egy diktáló rendszer esetén 90–95%-os fedés is kevés lehet a hatékony működéshez. A fedési értékek függnek a tematikától is. A kutatásom során egy általános korpuszra adtam meg az összehasonlításhoz szükséges információkat, de ezt szűkebb korpuszokra is elvégezhető. Ezek közül a Bibliák témakör részletes adatait megadtam, illetve néhány olyan speciális tematika adatait, amelyhez rendelkezésemre állt megfelelő vizsgálati korpusz. Ezekhez a következő fejezetben megtalálhatók a fedési görbék: magyarországi vezetéknevek (4.1. ábra), magyar keresztnevek (4.2. ábra), magyarországi településnevek (4.3. ábra).

Más infokommunikációs technológia, például a koprusz alapú beszédszintetizátor esetében szintén felhasználható az általam ismertetett módszer a szintetizálási eljárás adott nyelvű és tematikájú szintézisének értékeléséhez. A korpuszban szereplő szavak száma alapján nyelvenként megállapítható a fedési érték, amely a legtöbb esetben korrelációban van a minőséggel. Minél kisebb a fedés, annál több esetben fogja a beszédszintetizátor kisebb méretű elemekből előállítani a beszédet. A több elem, több illesztést jelent, amely több illesztetlenségi hiba lehetőségét adja. Az illesztetlenségi hibák általában hallhatóak, amelyek közvetlenül a megítélt minőség romlásához vezetnek. Angol nyelvre készültek általános koprusz alapú beszédszintetizátorok, amelyek közel emberi minőségben szólnak. Ha megvizsgáljuk a 3.4. ábrát, látható, hogy angol esetében 97,5%-os fedést 20 100 szó biztosít. Ehhez képest magyar esetében ugyanilyen fedéshez, tehát megközelítőleg ugyanahhoz a beszédminőséghez 400 000 szó szükséges. Ez magyarázza azt is, hogy eddig nem készült magyar nyelvre jó minőségű korpusz alapú beszédszintetizátor, a feladatot más úton kell megoldani.

Diktáló rendszereket vizsgálva megállapítható, hogy angol nyelv esetében 20 ezer szavas nagyszótáras beszédfelismerők az adatok szerint átlagosan 97,5%-os fedést biztosítanak. Ugyanez a megoldás magyar nyelv esetében már azt eredményezi, hogy elvi korlátként átlagosan minden 5. szó hibás felismerési eredményt ad, amely a felhasználást már nem teszi lehetővé. A szó alapú beszédfelismerők magyar esetében csak korlátozott tematikára működnek, vagy más kisebb méretű nyelvi egységek, például morfémák felhasználása szükséges.

3.3. A betű fogalmának kiterjesztése szövegek minősítéséhez

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, valamint ahhoz, 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. Ha rendelkezésre áll egy ilyen statisztika készítő módszer és egy nyelvre jellemző statisztika, akkor ismeretlen szövegek gyors jellemzésére és összehasonlítására nyílik lehetőség. Az ilyen összefüggések megállapítására végzett kutatások eredményeit ismertetem a következőkben.

3.3.1. Betűstatisztika a hangalak figyelembevételével

A mérésekhez gépi gyűjtő és szortírozó algoritmusokat készítettem, kifejezetten ehhez a kutatáshoz. [C8] A betű fogalmát kiterjesztettem és a célkitűzéshez alakítottam. A betűstatisztikához a vizsgált leghosszabb betűsorozat a szó volt, a nem betű típusú karaktereket figyelmen kívül hagytam (számok, relációs jelek stb.).

3.3.1.1. A betű fogalmának kiterjesztése

A betű fogalmának kiterjesztése azt jelentette, hogy 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 rendelkezésre áll az új típusú betűstatisztikában a [j] hangként kimondott *j* és *ly* betű, vagy az [i] hangként kimondott *i* és történelmi nevek végén gyakran szereplő *y* betű.

A betűstatisztika készítésekor betűként a következő tulajdonságú karaktert vagy karaktersorozatokat értem:

- a 44 betűs magyar ábécé tagjai: *a, á, b, c, cs, . . . , q,. . . , s, sz,. . . , w, x, y, z, zs*
- régi magyar családnevekben, idegen szavakban gyakran előforduló betűkombinációk: *cz, ch*, amelyeknek többféle ejtése is lehet (*technika, charter*).

Néhány ilyen betű jelölését ki kellett bővíteni, hogy érzékeltessem a belőle keletkezett hangot. Ezt a jelölést betű-hangjelnek nevezzük.

A *ch* betűkapcsolatból háromféle hangot vizsgáltam ([x], [t͡ʃ] és [k]). Ennek megfelelően ezekre különböző jelölést alkalmazok: ch_x, ch_cs, ch_k (ezeknél a jelöléseknél az aláhúzás utáni betű jelöli a kiejtési formát).

A *h* betű hangalakja is többféle lehet. Néma lehet szó végén (*cseh* [t͡ʃ ɛ]). Jelölése: $h_{néma}$.

A zöngésen ejtett forma a [ɦ] hang, ami esetenként intervokális helyzetben fordul elő. Jelölése: $h_{zöngés}$. Mássalhangzó kapcsolatban lehet velarizált zöngétlen réshang [x] (*sahnak*).

A *j* betűt egyes esetekben zöngétlen [ç] réshangként ejtjük. Jelölése: $j_{zöngétlen}$.

Az *sch* német eredetű betű is gyakran előfordul, a 3 karakteres hossza miatt fontos a külön kezelése. Jelölése: sch.

Az *y* betű többnyire régi nevekben és idegen eredetű szavakban fordul elő általában [i] vagy [j] hangként valósul meg ejtéskor. Jelölésük: y_i, illetve y_j. Nem vizsgáltam azokat az eseteket, amikor az *y* betű [ji] formában valósul meg, mint például a *Fáy* szóban.

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 használtam (*Ri-chárd*, *Mün-chen*, *Ben- czúr)*, miszerint ezek a betűk nem elválaszthatóak. A döntéseket a magyar elválasztási minta-gyűjtemény szószedete (Nagy 2008) alapján hoztam meg. Az algoritmusom figyelembe veszi a két karakterből álló betűk mellett (*gy, ty, ny, sz, zs, cs*) azok hosszú változatát is. A hosszú változatokat két betűnek tekintettem (*zzs= zs + zs*) a statisztikai feldolgozás során.

3.3.1.2. A statisztika készítés módszere

A kiterjesztett betű értelmezésével el kell készíteni a bővített statisztikát. Ehhez a szövegekből előállítottam a hangalakot (hangszimbólumok írott sorozatát) beszédtechnológiai gépi módszerek alkalmazásával. Három hangátírási eszközt használtam, egy szabály alapú algoritmust (a Profivox szövegfelolvasó rendszer fonetikai átíróját és szabálygyűjteményét (Olaszy et al. 2000)), a magyar elektronikus kiejtési szótárat (Abari–Olaszy 2006) és a névmondó tulajdonnév kiejtési gyűjteményt [C4]. A kiejtési forma meghatározásához kialakított szabályok a magyar nyelvi normát képviselik, a fonetikai átíró több éves iteratív kutatási és fejlesztési munka eredménye. Az írott formához hozzárendelt hangsorozat segítségével meghatározhatók a kiterjesztett betűk statisztikái.

3.3.1.3. A módszer tulajdonságai

Az újfajta betűstatisztika elkészítéséhez kidolgozott módszer megismételhető vizsgálatokra alkalmas, mivel számítógépes támogatással készült, az algoritmusok többször is futtathatók, nem szükséges kézi feldolgozás. Újszerű tudományos vizsgálatok is végezhetők. Összekapcsolható a beszédhang és a karakter reprezentáció.

Az eljárás pontosságát számos tényező befolyásolhatja. Számolni kell azzal, hogy a gépileg gyűjtött és ellenőrzött szöveg tartalmaz(hat) hibákat. A feldolgozott szövegkorpusz nagy mérete miatt manuális ellenőrzés nem jöhet szóba. A felhasznált kiejtési kivétel szótárak szintén részben gépi módszereken alapulnak, ezért tartalmazhatnak hibákat vagy hiányosak is lehetnek. A vizsgált betűk meghatározása önkényes, a gépi beszédfeldolgozás egyes szempontjait tartotta szem előtt, más felhasználás esetén a vizsgált betűk kiválasztása korlátozást jelenthet. Például a régi írásmódú betűk vizsgálata nem teljes körű, amely a beszédfeldolgozási szempontjából többnyire megengedhető, de névelemzés esetén már esetleg nem.

3.3.2. Módosított betűstatisztika magyar nyelvre

Az első átfogó fonémastatisztikát Szende (1976) kézi méréseiből ismerjük. A szerző 80 000 fonémát felölelő beszédanyagot dolgozott fel. Magyar betű statisztikáról nem találtam szakirodalmi adatot.

Kutatásomban a vizsgált nyelvi anyag a Magyar Nemzeti Szövegtár 2006-os verziójának teljes szöveganyaga volt [C8]. A $MNSZ_{2006}$-ról a részletes ismertetést korábban a 2.4. táblázatban megadtam.

Az új statisztika összehasonlításához elkészítettem egy hagyományos karakter- és egy hangstatisztikát. A hangstatisztikát mondatokra vonatkoztatott egységekből készítettem, figyelembe véve a szóhatárokon a folyamatos ejtésből eredő hangváltozásokat is.

A vizsgált szöveg és hangalak statisztikai elemzése három formában készült. Az eredmények a 3.6. táblázat oszlopaiban láthatók. Az első két oszlop a karakterstatisztika, a második kettő a betűstatisztika, az utolsó két oszlop a hangstatisztika eredményeit mutatja. A táblázatban szereplő számértékek megadják, hogy átlagosan 1000 elemből hány adott elem fordul elő. Az üres mezők azt jelentik, hogy az adott típusú statisztikában az elem nem szerepelt.

3.6. táblázat. Magyar karakter-, betű- és hangstatisztika

Karakter	1000-ből	Betű	1000-ből	Hang	1000-ből	Karakter	1000-ből	Betű	1000-ből	Hang	1000-ből
a	89,37	a	92,85	[ɔ]	90,21	o	40,93	o	40,21	[o]	42,26
á	35,95	á	37,58	[aː]	37,99	ó	10,03	ó	10,49	[oː]	9,95
b	19,66	b	20,56	[b]	18,28	ö	10,90	ö	11,39	[ø]	11,75
c	7,64	c	3,97	[t͡s]	6,10	ő	8,94	ő	9,35	[øː]	9,68
		cs	3,91	[t͡ʃ]	3,85	p	11,14	p	11,65	[p]	12,42
d	19,74	d	20,42	[d]	19,49	q	0,04	q	0,04		
		dz	0,03			r	42,47	r	44,41	[r]	44,02
		dzs	0,02			s	60,35	s	39,08	[ʃ]	35,89
e	98,70	e	101,31	[ɛ]	106,59			sz	19,27	[s]	24,55
é	33,46	é	35,02	[eː]	35,69	t	79,42	t	82,72	[t]	81,58
f	9,18	f	9,59	[f]	9,04			ty	0,27	[c]	4,09
g	33,80	g	22,69	[g]	19,82	u	10,18	u	10,73	[u]	11,19
		gy	12,70	[ɟ]	11,45	ú	3,01	ú	3,06	[uː]	2,69
h	15,32	h	13,07	[h]	17,56	ü	5,51	ü	5,85	[y]	5,54
i	44,06	i	46,39	[i]	47,28	ű	1,86	ű	1,86	[yː]	1,74
í	5,82	í	5,60	[iː]	5,51	v	19,89	v	20,80	[v]	21,52
j	11,19	j	11,98	[j]	14,27	w	0,28	w	0,29		
k	49,22	k	51,46	[k]	53,63	x	0,36	x	0,38		
l	62,27	l	60,78	[l]	58,46	y	22,71	y	0,21		
		ly	3,77			z	43,48	z	26,48	[z]	24,51
m	35,00	m	36,56	[m]	36,61			zs	0,73	[ʒ]	2,21
n	58,12	n	53,78	[n]	54,37						
		ny	7,02	[ɲ]	8,21						

A speciális betűk statisztikáját a 3.7 táblázatban adom meg, ebben a számértékek 1 millió elemre vonatkoznak. Az összes vizsgált betűre vonatkozó gyakorisági sorrendet a 3.8. táblázat mutatja.

A különböző statisztikák elkészítése nagyságrendileg eltérő erőforrást igényelt. A karakterstatisztika másodpercek alatt elkészült, a betűstatisztika több tíz perc, míg a hangstatisztika elkészítése 3-4 órát vett igénybe (PC, 3 GHz-es processzor, 4 Gbyte memória). A karakterstatisztika használata tehát akkor előnyös, ha a gyors működés elengedhetetlen.

A karakterstatisztika csak 36 karakterre tartalmaz információkat. Ahol ugyanazon karakter több betűreprezentációban is előfordulhat, ezt figyelembe kell venni a számadatok értelmezésénél.

Például a 3.6. táblázat *s* karakteréhez és betűjéhez tartozó gyakoriságokat összevetve látható, hogy az *s* karakter jóval gyakrabban fordul elő, mint az *s* betű. Ennek oka a kettős betűk szétdarabolása. Betűstatisztika helyett ezért a karakterstatisztika fenntartásokkal használható. Ennek ellenére az egyszerű programozhatóság miatt sok helyen így használják.

A megadott hangstatisztika is tartalmaz egyszerűsítéseket, csak 38 beszédhang szerepel benne (nem kezeli külön a hosszú-rövid mássalhangzókat, mivel közöttük csak az időtartam különbség van, spektrális nincs). Ennek ellenére a karakterstatisztikához képest, jobban tükrözi a nyelv tulajdonságait, mert az egyszerűsítések fonetikailag megengedhető helyeken történtek. A betűstatisztikával összehasonlítva az elemek hasonló gyakorisággal szerepelnek.

Szende adataival összehasonlítva a magánhangzók esetében az [yː] és a [y] estében 50 %-os eltérést tapasztaltam. A legkisebb különbség az [i] esetében volt megfigyelhető. Ha a teljes statisztikára vetítve vizsgáljuk az eltérést, akkor 1 százalékpont volt a legnagyobb eltérés, amely elhanyagolható. A leggyakoribb mássalhangzókat vizsgálva az adatok eltérőek, Szende 68,8

[n] hangot számolt meg 1000-ből, míg itt csak 54,4 hang volt. A [t] hang esetében fordított a helyzet, Szende 65 hangjával szemben itt 81,6 hang szerepel. Gósy (2004) spontán beszédre készített hangstatisztikát, amelyben a magánhangzó-mássalhangzó arány 43% és 57% volt. Itt ez az arány 42% és 58%. A leggyakoribb hangot összehasonlítva szintén hasonló számokat kaptunk, az [ɛ] hang Gósy statisztikájában 11,4%-os gyakoriságú, itt 10,7%. A 3.6. táblázat betűstatisztika részében mind a 44 betű gyakoriságát megtalálhatjuk. Ez a 44 betű a vizsgált szövegkorpusz 99%-át alkotja, míg a speciális betűk csak 1%-ot tesztnek ki. A ábécé betűi közül a *dz, dzs, q* szerepel nagyon ritkán, a 1 millió szóban átlagosan 20-40 db található meg.

Az *y* betű [j] hangként való realizációja gyakoribb, mint az [i] hangként való megjelenése. Ez abból adódik, hogy idegen nevek többször szerepelnek (például *Toyota*), mint a történelmi nevek (például *Dessewffy*).

A *ch* betű leggyakrabban [x] hangként jelenik meg, a [t͡ʃ] hang a második leggyakoribb formája, [k] hangként ritkán ejtjük.

3.7. táblázat. Betűstatisztika speciális betűkre

Betű-hangjel	1000000-ból
ch_cs	28,12
ch_h	129,04
ch_k	2,33
ck_k	4,96
cz_c	25,24
$h_{néma}$	60,27
$h_{zöngés}$	2470,19
sch	85,09
ts_cs	34,25
tz_c	11,84
y_i	65,27
y_j	111,27
$j_{zöngétlen}$	3,59

3.3.3. Alkalmazási lehetőségek

Az új típusú betűstatisztika használható kutatási feladatokra, a magyar szöveges állományok statisztikai tulajdonságainak vizsgálatára (Milyen gyakran jelöl a *ch* betűkapcsolat [x] hangot?). Az algoritmus felhasználható beszédadatbázisok készítésekor a felolvasandó szövegállományok elemzésére, válogatására. Például megbecsülhető, hogy egy adott szöveg felolvasása esetén a felolvasott szöveg egy kiválasztott hangból elegendő számút tartalmaz-e.

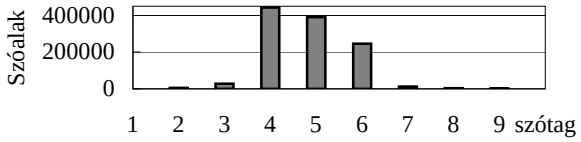
3.4. A magyar szóalakok szótagszám szerinti eloszlása

A magyar ragozó nyelv, ezért egy-egy szótőhöz számos rag, jel, képző kapcsolható. Ennek következtében az egy szótagú szavaktól egészen az igen hosszú szóalakig terjed a hosszúsági skála. Az alábbi szóstatisztika azt mutatja meg, hogy milyen a magyar szóalakok gyakorisága a szótagszám függvényében. A mérés nyelvi anyaga egy 80 millió szóból álló szövegkorpusz volt,

3.8. táblázat. Betűstatisztika gyakorisági sorrendben

Betű	db/1000	Betű	db/1000	Betű	db/1000	Betű	db/1000
e	101,31	d	20,42	ú	3,06	y_i	0,065
a	92,85	sz	19,27	$h_{zöngés}$	2,47	$h_{néma}$	0,060
t	82,72	h	13,07	ű	1,86	q	0,040
l	60,78	gy	12,70	zs	0,73	ts_cs	0,034
n	53,78	j	11,98	x	0,38	ch_cs	0,028
k	51,46	p	11,65	w	0,29	dz	0,026
i	46,39	ö	11,39	ty	0,27	cz_c	0,025
r	44,41	u	10,73	y	0,21	dzs	0,017
o	40,21	ó	10,49	ch_h	0,13	tz_c	0,012
s	39,08	f	9,59	y_j	0,11	ck_k	0,005
á	37,58	ő	9,35	sch	0,09	$j_{zöngétlen}$	0,004
m	36,56	ny	7,02			ch_k	0,002
é	35,02	ü	5,85				
z	26,48	í	5,60				
g	22,69	c	3,97				
v	20,80	cs	3,91				
b	20,56	ly	3,77				

amiből kiválogattam a szóalakokat (betűkép szerinti válogatással). Különbözőnek tekintettem két szóalakot, ha két szóköz közötti betűsor egyetlen karakterben is eltért. A szóalakok kézi vizsgálatát Olaszy (2006) végezte el és ennek eredményeként kaptam egy 1,5 millió szóalakból álló korpuszt, amelyben minden szóalak különbözött, nyelvtanilag helyes, idegen szavaktól mentes volt, és mindegyik szóalak egyszer fordult elő. Minden szóalaknak meghatároztam a szótagszámát. Mivel itt a vizsgálatban elsődlegesen magára a szóalakokra koncentráltam és nem azoknak a folytonos szövegben való megjelenésükre, ezért csak ennél a vizsgálatnál a határozott névelőket nem jelenítettem meg a statisztikában. Ebből készült a magyar szóalakok



3.6. ábra. Magyar szóalakok szótagszám szerinti eloszlása az összes szóalak függvényében

statisztikája a szótagszám függvényében (3.6. ábra). A szótagok száma szerinti eloszlás képe azt mutatja, hogy nyelvünkben a 4 és 5 szótagú szavakból van a legtöbb, majd a 3 és 6 szótagúak következnek. Legkevesebb az 1 és 2 szótagú, valamint a 8 és 9 szótagú szavakból van.

A magyar szavak szótagszám szerinti gyakoriságát a Magyar Nemzeti Szövegtár anyagán mértem meg (187 millió szó). A gyakoriságot a 3.7. ábra mutatja. A mérésbe nem számoltam bele az *a*, *az* névelőket. A gyakorisági adatok szerint a leggyakrabban két szótagú szavakat használunk a szövegekben, mármint akkor, ha a határozott névelőktől eltekintünk.

3.7. ábra. Magyar szavak előfordulásának gyakorisága szövegekben a szótagszám függvényében, ha a határozott névelőket nem vesszük bele a mérésbe

3.5. Összegzés

A magyar nyelvre megállapítottam a szóalakok gyakorisági sorrendjét és azt, hogy az adott számú leggyakoribb szó a vizsgált korpusz mekkora részét fedi le. Ezeket az adatokat grafikusan és formalizálva is leírtam. Összehasonlítottam három nyelvet több különböző méretű szövegkorpusz segítségével. Az elemzéseket magyarra, angolra és németre végeztem el, de a fedési és gyakorisági görbék használatával más nyelvek is hasonlóan összehasonlíthatók. Eljárást – egy módosított betűstatisztikát – dolgoztam ki magyar nyelvre, amely figyelembe veszi a gépi beszédkeltés szempontjait.

4. fejezet
Magyar tulajdonnevek, cégnevek és magyarországi címek gépi felolvasása

A beszédszintetizátorok által előállított beszéd minősége függ a létrehozásnál alkalmazott technológiától és modellezéstől, illetve attól is, 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.

A név- és címfelolvasás témaköre az általános szövegek felolvasásához képest szűkebb terület, de a nevek természetéből adódóan nem korlátos. A folyamatosan megjelenő idegen eredetű személynevek, illetve a nyelvi határokat figyelembe nem vevő cégnevek megkívánják a tetszőleges betűsorozat felolvasásának képességét. A jobb hangminőség érdekében kihasználható, hogy a felolvasandó részek tartalmaznak különböző gyakran előforduló elemeket is. Magyar nyelvre ilyen témában nem született más publikált megoldás és idegen nyelvre is csak részproblémákkal foglalkoztak később érintőlegesen (Aylett 2004).

A név- és címfelolvasásban három gépi beszédelőállítási technológiát kombináltam, a triád alapú általános szövegfelolvasást [J7], a számfelolvasást (Olaszy–Németh 1999) és a kötöttszótár alapú szövegfelolvasó módszert. Az adott közlés előállítása során tehát váltakoznak a gépi beszédelőállítási módszerek. A kombinált eljárás lényege és újszerűsége, hogy a felolvasott névben és címben mindig azt a technológiát alkalmazom, amelyik a legjobb minőséget tudja biztosítani az adott résznek. A név- és címfelolvasás során a különböző típusú kiejtendő elemek csak meghatározott pozícióban szerepelhetnek a hangsorozatban, így a prozódiájuk előre meghatározható. Ez alapján a gyakori elemeket külön előre felolvasva és beillesztve a legjobb minőség biztosítható.

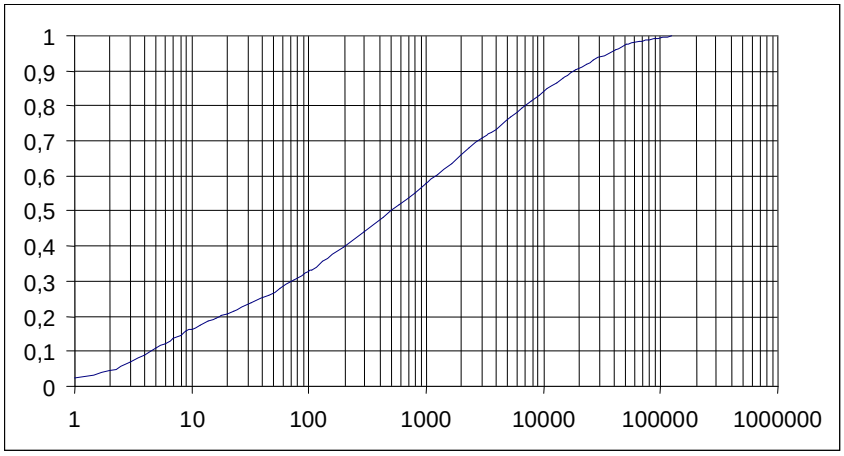
A megoldás kialakítása során elvégeztem a szövegfelolvasó bemeneti szövegeinek elemzését. Ehhez rendelkezésemre állt egy 3 millió telefon előfizetői rekordot tartalmazó név- és címlista (amit továbbiakban röviden adatbázisnak nevezek). Az elemzéshez a neveket és a címeket részekre bontottam és különböző statisztikai tulajdonságaikat így vizsgáltam. Címként a továbbiakban mindenhol a következő típusú karaktersorozatokat értelmezem: „*1111 Budapest, Műegyetem rkp. 3-9.*”.

4.1. Nevek

A nevek vizsgálatakor a rekordokat kettéosztottam cégnevekre és természetes személyek neveire. Cégnévnek azokat a rekordokat tekintettem, amelyek magukban foglaltak valamilyen cégformát (például *kft,bt*), intézménynevet (*. . . hivatal*) vagy egyéb elnevezést, amely utal arra, hogy nem természetes személy nevéről van szó (*ABC, bolt, kereskedés*).

4.1.1. Vezetéknevek

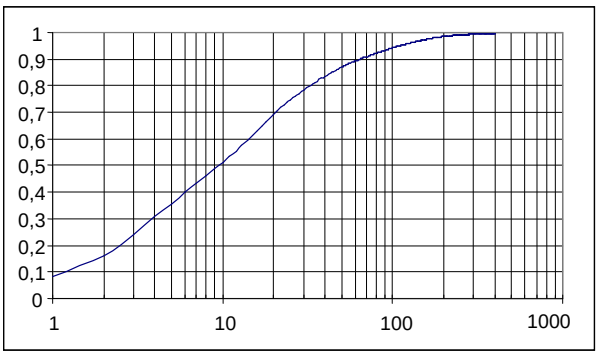
Az adatbázisban 2,6 millió személynév található, amelyek között 103 850 különböző vezetéknevet találtam. A leggyakoribb vezetéknevek és gyakorisági adataik megtalálhatók a függelék B.1. táblázatában. A vezetéknevek statisztikai tulajdonságait a 3.1. fejezetben alkalmazott fedési görbével vizsgáltam. A 4.1. ábrán látható, hogy a fedési görbe csak lassan emelkedik, tehát sok, ritkán előforduló név található a vezetéknevek között. A leggyakoribb 1000 csak az összes vezetéknév 57%-át fedi le. A 95%-os fedés eléréséhez pedig 40 000 név szükséges.



4.1. ábra. A leggyakoribb vezetéknevek fedési görbéje

4.1.2. Keresztnevek

A keresztnevek vizsgálatához különböző források (internetes keresztnévadó oldalak) alapján összeállítottam egy listát, amely 2834 különböző magyar keresztnevet tartalmazott. Az elemzett adatbázisban ezekből 1797-et találtam meg, és további 183 egyéb keresztnevet azonosítottam, amelyek idegen keresztnevek voltak. A leggyakoribb keresztnevek és gyakorisági adataik megtalálhatók a függelék B.2. táblázatában. A leggyakoribb keresztneveket vizsgálva azt tapasztaltam, hogy a női keresztnevek előfordulása kisebb, mint az a demográfiai adatokból várható lenne. A leggyakoribb női név a *Mária*, amelyik a 23. leggyakoribb keresztnév. A leggyakoribb 30 keresztnév között is csak 3 női név szerepel még: *Éva, Katalin, Erzsébet*. Az okok valószínűleg szociolingvisztikaiak, a telefonkészülékek többnyire a férfi nevén vannak bejegyezve, illetve a nők felveszik a férjük nevét. A férjezett nevek esetén a *-né* toldalékot eltávolítva vizsgáltam a keresztneveket. A keresztnevek fedési görbéjét a 4.2. ábrán mutatom be. Látható, hogy a keresztnevek sokkal kevésbé változatosak, mint a vezetéknevek. A 10 leggyakoribb az összes keresztnév 50%-át lefedi. A 95%-os fedéshez már elegendő hozzávetőlegesen 100 név.



4.2. ábra. A leggyakoribb keresztnevek fedési görbéje

4.1.3. Titulusok és egyéb előtagok

A személynevek vizsgálatakor 82 különböző titulust és előtagot találtam. A leggyakoribb előtagok a *dr, dr.-né, ifj, id, özv* voltak. A nevek között még előfordultak egy betűből álló előtagok is. A titulusokra jellemző volt, hogy több pozícióban is előfordultak, amely a felolvasás prozódiája szempontjából előnytelen. Például a „*Dr. Tóthné Dr. Kiss Márta*” esetében, a két „*Dr.*” eltérő prozódiával rendelkezik, így ugyanaz az elem nem használható a két pozícióban.

4.1.4. Cégnevek

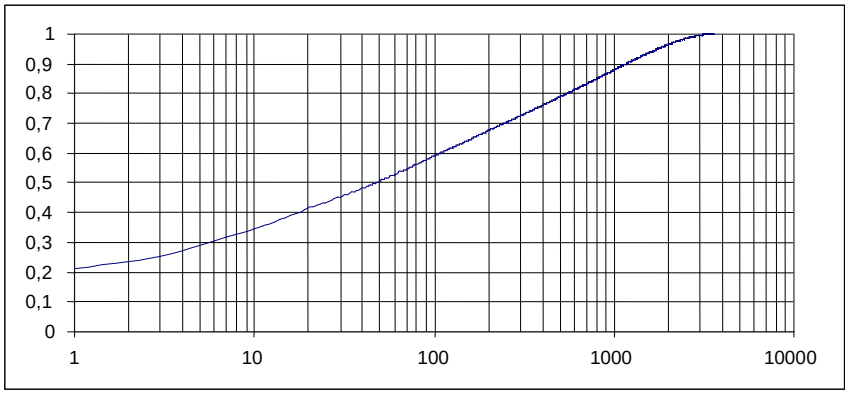
A vizsgált adatbázisban a cégnevek felolvasási szempontból az esetek háromnegyedében egyszerű szerkezetűek: a cég elnevezése után a cégforma következik. A cégnevek 10%-ban tartalmaztak legalább egy keresztnevet. Szintén 10%-uk valamilyen betűszót, 9%-uk számelemeket és 2,5%-uk valamilyen speciális jelet tartalmazott.

4.2. Címek

Az adatbázis vizsgálat során a címeket a következő részekre tudtam bontani: irányítószám, település neve, településrész neve, közterület neve, közterület fajtája, házszám, lépcsőház, emelet, ajtó. A címek irányítószámai mindig 4 jegyűek, kezelésük egyszerű. Ugyanazon településnevek és a településrészek neveinek írása többféle volt, az adatbázis ebből a szempontból nem volt egységes. Egyes rekordoknál ezeket a neveket különírták, máshol egybe, és voltak olyan rekordok, ahol a településrész neve a közterület elnevezéséhez került. A közterület elnevezése az „*Üllői út*” esetében az „*Üllői*”, a közterület fajtája az „*út*”. A közterület elnevezés és a cím hátralévő részében bizonyos rekordok hiányosak voltak, vagy nem voltak egyértelműen feldolgozhatók. Például hiányos cím a „*1117, Budapest, Kiss Lajos utca 16, 4 102/a*” esetében a „*4*”, amelyről nem eldönthető, hogy épületet, emeletet, vagy lépcsőházat jelent.

4.2.1. Településnevek

A Magyar Posta irányítószám-jegyzékében 3562 különböző településnév szerepel, amelyből a vizsgált adatbázisban 3523 fordult elő. További 28 elnevezést találtam. Ezek valamelyik településrész pontosabb megnevezésére szolgáltak. A településnevek fedési görbéje a 4.3. ábrán látható. A településneveknél önmagában *Budapest* 21%-ot lefedett. Az 1000 leggyakoribb településnév majdnem 90%-os fedést ad, a 95%-os fedéshez közel 1800 név szükséges.



4.3. ábra. A leggyakoribb településnevek fedési görbéje

4.2.2. Közterületnevek

A települések közterületi nevei között 15 932 különböző elnevezést találtam az adatbázisban. A közterületek fajtáját tekintve az *utca* a leggyakoribb, ez az esetek háromnegyedében fordult elő. A második leggyakoribb közterület fajta az *út* (15%) és a *tér* (3%).

4.3. A felolvasandó elemek meghatározása a kötött szótáras beszédszintézishez

A kötött szótáras beszédszintetizátor hullámforma-elembázisának létrehozásához meghatároztam az ember által felolvasandó elemeket. Ezek általában 5–15 szótag hosszúságú elemek. Az adatok elemzése során a 4.1. táblázatban szereplő kategóriákat vizsgáltam meg abból a szempontból, hogy mennyire alkalmasak arra, hogy bemondóval felolvastatva a szintetizált közlés hangminőségét javítsák.

A felsorolt különböző kategóriákban szereplő szavakra tehát meghatároztam a gyakorisági sorrendet és a fedési görbéket. A kategóriákban előforduló összes szó (mintegy 120 ezer) felolvasása nem lehetséges jó minőségben, 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 órányi felvétel). A 4.1. táblázatban szereplő második oszlop az elemzés során talált szavak számát, az utolsó oszlop a felolvasott szavak számát jelenti.

4.1. táblázat. A vizsgált kategóriák és tulajdonságaik

Kategória	Különböző elnevezések száma	Felolvasott elemek száma
Vezetéknevek	103 850	0
Keresztnevek	1 797	313
Cégformák	8	8
Településnevek	3 523	1000
Közterület elnevezések	15 932	14
Közterület fajták	14	14
Betűzés	36	36

A szükséges felolvasandó szavakat a fedési görbe segítségével határoztam meg, a cél a 95%-nál nagyobb fedés elérése volt, de azzal a kikötéssel, hogy ne legyen 1000-nél több felolvasandó tétel. Az 1000 elemes korlát azért szükséges, hogy a felvett elemek hangzása egyforma legyen. Túl sok elem esetén a teljes felolvasási procedúra elhúzódik, amely során nehezen biztosítható az azonos minőség, de ezzel a darabszám korlátozással megoldható a 4 órás felvételi korlát betartása. A vezetékneveknek nagy a változatossága, a 95%-os fedéshez 40 000 név felolvasására lenne szükség, ami jóval nagyobb, mint a korábban meghatározott felső korlát. Továbbá a leggyakoribb nevek felolvastatása a szintézis során jó minőségben megoldható a triádos szintetizátorral, mert ezekhez a nevekhez tartozó hangkapcsolódási sorozatok a szintetizátor alapelemének számító CVC beszédelemekkel jól lefedhetők, így külön emberi felolvasásuk ezen okból sem indokolt.

A keresztnevek statisztikai vizsgálata alapján elegendő lenne a 95%-os fedési kritérium eléréséhez 100 név emberi felolvasása. Azonban a keresztnevek a hangsorépítés utolsó elemeként szerepelnek a névben, ezért minőségük jobban hat a felhasználó szubjektív minőségi érzetére. A stúdióban felolvasott 313 elem 99,8%-os fedést ad a keresztnevekre.

A cégformák kis száma miatt minden elem felolvasása indokolt, illetve a keresztnevekhez hasonlóan ezek az elemek is hangsorvégi pozícióban szerepelnek.

A településnevek esetében az 1000-es korlát szabja meg a felolvasott elemek számát, ami kb. 90%-os fedést biztosít.

A közterületek típusai kis változatosságot mutatnak, itt minden – adatbázisban szereplő – elem felolvasása megoldható.

A betűzés pozíciója változatos a különböző címek és nevek esetén. Külön felolvasásukat nem a statisztikai előfordulásuk száma indokolja, hanem az, hogy rövidségük miatt az ilyen elemek nehezen érthetők, így a legjobb minőségű felolvasás indokolt.

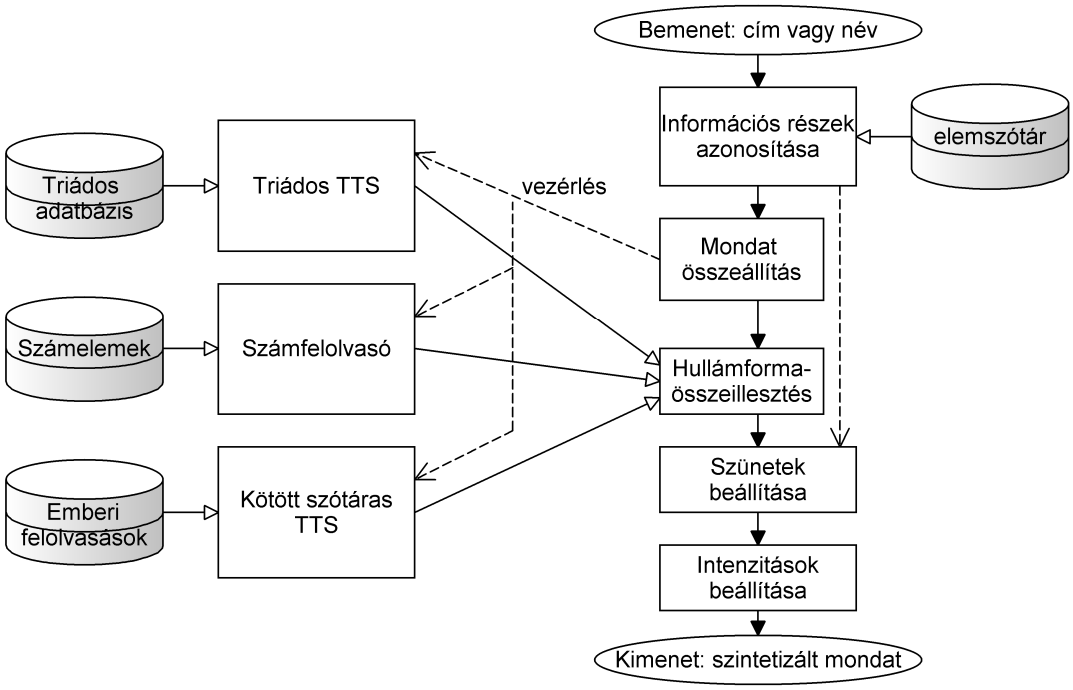
4.4. Nevek és címek szintetizálása

4.4.1. Az információs elemek meghatározása

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

Személynevek: titulus, vezetéknév, keresztnév
Cégnevek: cégnév, cégforma

Címek: irányítószám, település neve, településrész neve, közterület neve, közterület fajtája, házszám, lépcsőház, emelet, ajtó



4.4. ábra. A nevek és címek szintetizálásának főbb lépései

A szintetizálás főbb lépéseit a 4.4. ábrán szemléltetem. A bemeneti szintetizálandó cím vagy név feldolgozásának első lépése az, hogy az információs elemeket azonosítom. Ez a következő három paraméter szerint történik: az elem pozíciója a rekordban, az elemszótárban való szereplés, már azonosított információs elemekhez viszonyított pozíciója. A feldolgozáshoz a különböző elemszótárakat a megvizsgált adatbázis alapján állítottam össze, amelyek a következők:

- Ez az elemszótár a cégformákat (például: „*bt.*”), és azokat az elnevezéseket (például: „*kiskereskedés*”) tartalmazza, amelyek alapján eldönthető, hogy személynévről vagy cégnévről van-e szó. A cégforma elemszótára 303 szót tartalmaz, amelyet a B. függelékben soroltam fel.
- A szintetizálandó személynevek időszerkezetének meghatározásához szükséges a kereszt- és vezetéknév elemek azonosítása, amely a keresztnév elemszótár segítségével történik. 2834 db keresztnév szerepel az elemszótárban.
- A közterületek típusait tartalmazó elemszótár segítségével a címben azonosítható a közterület neve és típusa, és a mögötte lévő egyéb elemek. Az elemszótár 70 szót tartalmaz. Az elemszótár szavait a B. függelékben megadom.

Címek: A címek esetében az irányítószám és a város az első két pozícióban szerepel. Amennyiben a bemeneten ez nem biztosítható, akkor a városnév elemszótár építése és felhasználása segítségével azonosítható. A cím közterületnév része a közterülettípus segítségével meghatározható. A házszám és a hozzá kapcsolódó elemek a közterülettípus után található. Az ezen a részen használt rövidítések („*em.*”, „*lph.*”, „*ép.*”) alapján lehet az

itt található elemeket pontosabban behatárolni. A címekben előforduló rövidítések listája és feloldása bővebben a B. függelékben található.

Nevek: A nevek feldolgozásánál első lépésben a cégforma elemszótár alapján történik keresés. Amennyiben szerepel az elemszótárban a bemenet valamelyik szava, akkor a szintetizálandó szöveg cégnévként értelmezendő, ha nem, akkor személynévként. A személynevekben a keresztnév meghatározása az elemszótár alapján történik. Az ábrán szaggatott vonallal jelöltem, hogy az itt megállapított információkat a prozódia kialakítása részben használjuk fel.

4.4.2. Az információs elemek szintetizálása

A szintetizálás következő lépése a különböző információs elemek szintetizálása, majd az elemhez tartozó hullámformák egymáshoz illesztése (triádos szintetizált hullámforma, számfelolvasóval generált számok és külön felolvasott elemek). A 4.2. táblázatban összefoglaltam, hogy a különböző információs elemek szintézisből vagy külön felolvasott elemből származnak-e. A táblázatban jelölt elsődleges módszer adja az adott elemre a jobb

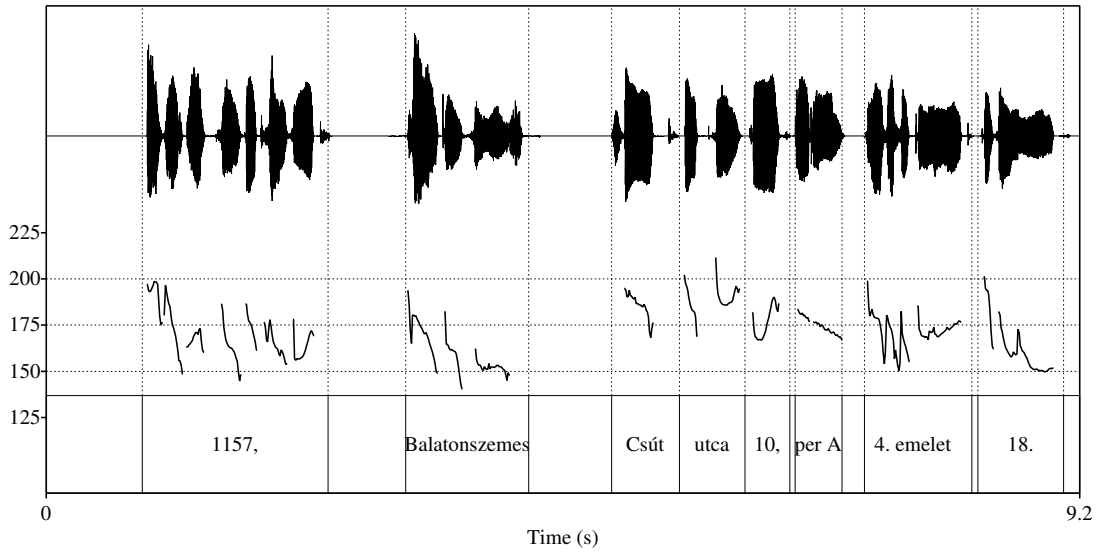
4.2. táblázat. Információs részek gépi felolvastatásának tulajdonságai

Kategória	Elsődleges módszer	Helyettesítő módszer	Dallammenet
Titulus	természetes ejtésből	triádos TTS	mondatkezdő
Vezetéknevek	triádos TTS	-	mondatközbeni/monoton
Keresztnevek	természetes ejtésből	triádos TTS	mondatvégi
Cégnevek	triádos TTS	-	mondatközbeni/monoton
Cégformák	természetes ejtésből	triádos TTS	mondatvégi
Irányítószámok	számfelolvasó	-	mondatközbeni
Településnevek	természetes ejtésből	triádos TTS	mondatközbeni
Közterület elnevezések	triádos TTS	-	mondatközbeni/monoton
Közterület fajták	természetes ejtésből	triádos TTS	mondatközbeni
Házszámok, ajtószám, épületszám	számfelolvasó	-	mondatközbeni vagy mondatvégi
„per A”	természetes ejtésből	triádos TTS	mondatvégi
„. . . emelet”	természetes ejtésből	triádos TTS	mondatközi
„. . . épület”	természetes ejtésből	triádos TTS	mondatközi

beszédminőséget, a helyettesítő módszer biztosítja, hogy az elsődleges módszer sikertelensége esetén is az információt beszéddé alakítsuk. Ilyen eset például az, amikor nem szerepel a természetes ejtésből felvett elemek között az adott szó. A felolvasási stratégiát tehát úgy alakítottam ki, hogy hibás információs elem azonosításából eredő hiányok ne okozzanak felolvasási hibát, minden esetben legyen olyan helyettesítő módszer, amely elő tudja állítani az adott elemet. Például, ha egy nevet hibásan keresztnévnek azonosít a rendszer, de az adott keresztnév nem lett rögzítve a természetes ejtéssel, a szintetizált névbe a triádos TTS fogja előállítani az adott elemet. Ilyen esetben a minőség és az érthetőség csökkenhet, de szerepelni fog minden elem a szintetizált beszédben.

4.4.3. Prozódia kialakítása

A prozódia klasszikus komponensei közül a ritmus, a dallam és a hangsúlyozás paramétereire dolgoztam ki szabályrendszert. Az elemek szintetizálása után következik a szünetek beállítása. A 4.5. ábrán példát adok egy cím felolvasásának időszerkezetére és alapfrekvencia-menetére. A felolvasás időszerkezetére jellemző, hogy egy természetes emberi ejtéshez képest a szintetizált mondat több szünetet tartalmaz. Ennek a célja az, hogy az így kialakított tagolás növelje az információ érthetőségét. Az ábrán látható cím esetében például természetes ejtésnél a házszám és a mögötte álló elem között általában nem tartunk szünetet, de egy rövid szünet segítségével a házszám jobban érthetővé válik (saját percepciós tesztek szerint). A szünetek elhelyezése szabályalapon, az azonosított információs elemek szerint történik. A szünetezés szabályai a B. függelékben megtalálhatók.



4.5. ábra. Példa egy szintetizált cím idő- és dallamstruktúrájára

A prozódia alapfrekvencia komponensének változtatására csak a triádos hullámforma összefűzéses szintetizátor esetében van lehetőség. Ezek a szintetizált elemek a pozíciójuknak megfelelő dallammenettel készülnek, a példánkban a „*Csút*” közterületnév készült triádos szintetizátorral. A természetes bemondásokhoz való jobb illeszkedés érdekében a mondatbelseji helyzetben a dallammenet megegyezik a triádos adatbázis eredeti monoton ejtésével, mert így nincs hangszínezetet is módosító jelfeldolgozás. A mondatvégi pozícióban viszont az eső dallammenetet meg kell valósítani, így ott az alapfrekvencia változtatást elvégezzük. A számfelolvasó esetében a számelemek 2 fajta, egy számsor végi és egy számsor közbeni prozódiával állíthatók elő, így mondatbelseji és mondatvégi helyzetben is a megfelelő dallammenet biztosítható a számoknál. A példában a hangsor belsejében szereplő számelemek vesszővel, a hangsor végén szereplők ponttal vannak jelölve.

A teljes közlemény prozódiájára kidolgozott szabályok a kijelentő mondatokra jellemző dallammenetet hozzák létre. A prozódia intenzitás komponenseként egy egyszerűsített modellt használtam, amely csak az utolsó elem esetében ír elő csökkentést, a többi elemet azonos szintre egyenlíti ki. Az emberi felolvasásból rögzített elemeket a tervezett mondatbeli helyzetük

és az annak megfelelő prozódia szerint rögzítettem. Ezt célzott vivőmondatok alkalmazásával értem el. Itt bizonyos esetekben kompromisszumot kellett kötni, például a „*per A*” elem kerülhet a cím belső pozíciójába is és a végére is. Mivel ez az elem az adatbázis vizsgálata szerint többnyire cím végén szerepel, ezért a hangfelvételkor az elemet a cím végi pozíció szerinti kiejtéssel rögzítettem, de csak enyhén eső dallammal. Ezzel biztosítottam azt, hogy belső pozícióban is lehessen használni. A 4.5. ábrán is láthatunk egy ilyen megoldást, a „*per A*” alapfrekvenciája az elem végére lecsökken, de nem annyira, hogy a következő elem alapfrekvenciája jelentősen eltérjen.

4.5. Érthetőségi teszt

4.5.1. A vizsgálat módja

A név- és címfelolvasó eredményességének vizsgálatára egy érthetőségi tesztet készítettem. A tesztben véletlenszerűen kiválasztott neveket és címeket használtam fel, összesen 20–20 darabot. A nevek megtalálhatók a 4.3. táblázatban az eredmények mellett, a címek a függelék B.3. táblázatában. A neveket és címeket leszintetizáltam az ismertetett rendszerrel illetve a Profivox általános beszédszintetizátorral [J7]. Azért választottam a Profivox rendszert, mivel ez a rendszer a név- és címfelolvasóban szereplő triádos TTS-nek az általános teljes prozódiai modullal rendelkező változata, amely alternatíva lehet a magyar nyelvű név- és címfelolvasásban. A bemeneti szövegek annyiban tértek el a két rendszerben, hogy a Profivox által nem ismert kifejezéseket feloldottam és központozással láttam el a megfelelő szünetezés érdekében. A vizsgált mintákat újra-mintavételeztem 8000Hz-cel, hogy a telefonos környezethez hasonló hangzást érjek el.

A meghallgatási teszt webes felületen történt. A tesztelők 20 db nevet és 20 db címet hallgattak meg. Egy kategórián belül a mondatok sorrendje és a szintetizátor fajtája is véletlenszerű volt. A tesztelő minden szintetizált mondatot csak egyszer hallgathatott meg, visszalépésre, újrahallgatásra nem volt lehetőség. A tesztelő feladata az volt, hogy a meghallgatott nevet vagy címet egy szövegdobozba gépelje be. A gépelés ideje nem volt korlátozva. A következő mintára való továbblépés egy gomb megnyomására történt. A tesztelő minden mondatot csak egyszer hallott, amelyet vagy az egyik, vagy a másik szintetizátor generált.

4.5.2. Eredmények

A tesztet 22 alany végezte el, 21-62 éves korig. A tesztelők között 17 férfi és 5 nő volt. A tesztelők közül 6 fő hangszórón keresztül, 16 fő pedig fej- vagy fülhallgatón keresztül hallgatta meg a mintákat. A meghallgatásos teszt átlagosan 11,5 percet vett igénybe.

Az eredmények kiértékelése kézzel történt. A különböző információs elemek pontos leírása nem elvárható, az elgépelések és a különböző írásmódokat azonosan helyesnek fogadtam el. Például: „*Bátorfi, Bátorfy*”, „*Fsz/3,földszint 3, fszt 3*”. A címek írását akkor tekintettem helyesnek, ha minden információs elemet tartalmazott és nulla vagy egy hibás volt közte. Több hiba esetén a cím hibás volt. Nevek esetén azt számoltam meg, hogy a név hány része

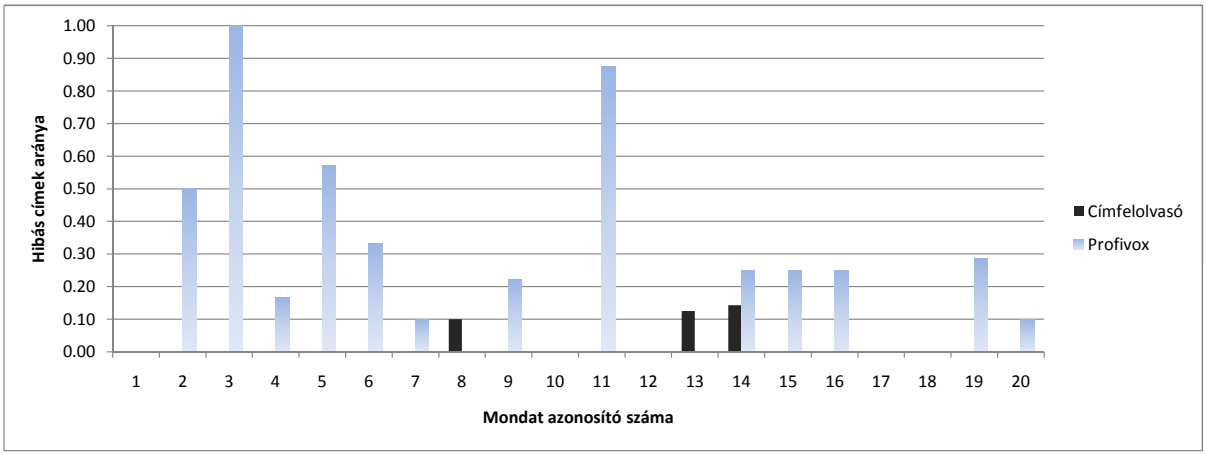
volt hibátlan az összes részhez képest. Például, ha a személynév egy vezetéknévből és egy keresztnévből állt, akkor hibás vezetéknév és helyes keresztnév esetén a név 0,5-es értéket kapott.

4.3. táblázat. Nevek érthetősége

Névfelolvasó	Profivox	Különbség	Név
0,90	0,10	0,80	Halmschläger Trade Rt.
1,00	0,47	0,53	Soft-Way Team Kft.
0,92	0,56	0,36	Szilléry Éva
0,86	0,50	0,36	Schürlein Andor
1,00	0,68	0,33	Oxy-Med Kft./Vaszari Béla
1,00	0,72	0,28	Bagossy József
1,00	0,79	0,21	Poly-Stop Bt.
0,75	0,57	0,18	Hellto Kft.
0,71	0,57	0,14	Chi & Brother Kft
0,88	0,79	0,09	Czverkon József
1,00	0,94	0,06	Ihracska Gábor
1,00	1,00	0,00	Bátorfi Szabolcs
1,00	1,00	0,00	Székely Árpádné
1,00	1,00	0,00	XII.ker.Önkormányzat Gamesz
1,00	1,00	0,00	Kálmán Imréné
1,00	1,00	0,00	Váczi Béla
1,00	1,00	0,00	Szalontay Szilvia
1,00	1,00	0,00	Gyarmatiné Fekete Csilla
0,80	0,90	-0,10	CYBERSYS Kft
0,64	0,75	-0,11	Green Line 5042 Kft

A nevek helyességének mért adatait a 4.3. táblázatban adom meg. Az első oszlopban az látható, hogy a névfelolvasó esetében a tesztelők az adott nevet mennyire sikeresen tudták leírni. 1,0-s érték esetén minden tesztelő helyesen jegyezte le a nevet. A második oszlopban az általános szövegfelolvasó Profivox esetében mért helyes észlelések aránya látható. A harmadik oszlop a két technológia értékeinek különbségét tartalmazza. Pozitív érték mellett a névfelolvasó, negatív érték mellett a Profivox által szintetizált mondatokat tudták a tesztelők nagyobb arányban helyesen lejegyezni. A 20 név közül 11 esetben a névfelolvasó, 2 esetben a Profivox volt érthetőbb, 7 esetben azonosan helyesen ismerték fel a két technológiával szintetizált neveket. Az eredmények alapján tehát a névfelolvasó érthetőbb (átlag: 0,92), mint a Profivox (átlag: 0,77), főleg az összetettebb nevek esetében.

A címek eltérő kiértékelése miatt az összesített táblázat helyett az eredményeket a 4.6. ábrán összesítettem. A címek és a teszt részletes eredményeit a függelék B.3. táblázatában adtam meg. Az ábrán a hibás lejegyzések arányát jelenítettem meg oszlopdiagramm formájában. A sötét oszlopok mutatják a címfelolvasó, a világos oszlop a Profivox hibáinak arányát az adott mondaton. A címfelolvasó mondatai közül 3 esetben voltak hibás lejegyzések, míg a Profivox rendszerében 13 cím esetében. A Profivox esetében 4 cím esetén legalább a tesztelők fele hibásan jegyezte le a címet. A címek érthetőségi tesztje alapján a címfelolvasó érthetőbb, mint a Profivox rendszerrel készített mondatok.



4.6. ábra. Mondatok hibás megértésének aránya a két szintetizátor esetén

4.6. Összegzés

A név- és címfelolvasás gépi megoldását újfajta módon közelítettem meg, több szintézis módszert kombináltam. A magyar nyelvű név- és címfelolvasáshoz a triádos szövegfelolvasót, a számfelolvasót és 1385 db előre felvett emberi ejtésű elemet (szót, szókapcsolatot) kombináltam. A felvett elemeket a rendelkezésre álló adatbázisból a fedési görbék segítségével úgy határoztam meg, hogy a hangfelvételre kerülő elemek száma a meghatározott korlátok alatt maradjon. A név- és címfelolvasáshoz meghatároztam az elemek sorrendjét, és az elemekhez hozzárendeltem a szintézis módszerét. Kidolgoztam a prozódia megvalósítására szolgáló szabályrendszert és megvalósítottam a működő megoldásban. Az eljárás eredményességét érthetőségi teszttel bizonyítottam.

5. fejezet
Korpusz alapú beszédszintetizátor hangminőségének javítása

A beszédszintetizátorok működéséhez használt hangadatbázisok készítése általában is összetett folyamat. A korpusz alapú szintetizátorok beszédadatbázisának mérete a néhány órától több száz óráig is terjedhet, amelyek felvételi körülményei változatosak lehetnek. A jobb szintetizált beszédminőség elérése érdekében a hangadatbázis intenzitáskiegyenlítése is szükséges. Ehhez első lépében megvizsgáltam a magyar olvasott beszéd hangintenzitás-viszonyait, majd algoritmust dolgoztam ki a korpuszos beszédszintetizátor intenzitáskiegyenlítéséhez.

5.1. A magyar olvasott beszéd beszédhangjainak szóra vetített intenzitástérképe

A magyar beszédhangok intenzitásviszonyaira vonatkozó korábbi mérések csak korlátozottan állnak rendelkezésre (Olaszy 1989), nagy mennyiségű adat feldolgozásáról szakirodalmi közleményt nem találtam. A beszédhangok intenzitás viszonyait eddig csak szavakon és példamondatokon vizsgálták. Ezért célom volt, hogy több adatközlőtől származó összefüggő, nagymennyiségű beszéden is végezzek intenzitás méréseket. A vizsgálat egységének a szót választottam, mivel ez a megközelítés illeszkedik a korpusz alapú beszédszintetizátor működési elvéhez. Másik oka enne a döntésnek, hogy az eredményeket össze tudtam vetni a – kis számú – korábbi kutatások adataival.

5.1.1. A vizsgálatba bevont hangadatbázisok

A felhasznált hangadatbázisok három helyszínen készültek. A felvételek legnagyobb részét egy professzionális stúdióban (Stúdió I.), a másik részét egy rádióállomás stúdiójában (Stúdió II.), a harmadik részét pedig az egyik mobil szolgáltató félprofesszionális stúdiójában (Stúdió III.) készítettük. Az első két helyen a felvételeket hangmérnök felügyelte, a harmadik helyen a bemondó maga végezte a hangrögzítést. Mindhárom felvételi környezet zajmentes volt és a felvételek nem tartalmaznak egyéb beszédet, zenét vagy más zavaró jelet. A felvételek jel/zaj viszonya 40-60 dB nagyságú.

Ahhoz, hogy ezeket a hangfelvételeket egységes beszédadatbázisba tudjuk szervezni, utófeldolgozást végeztem. A hangfelvételek intenzitáskiegyenlítése a felvétel fajtájától függően különböző beszédrészleteken történt, a legkisebb egység a mondat, a legnagyobb egy egész bekezdés volt.

A felvételek intenzitását -20 dB-re egyenlítettem ki (0 dB a maximálisan ábrázolható kivezérlés). Az intenzitást effektív érték (RMS) számításával határoztam meg, a bekapcsolási idő 200 ms, a kikapcsolási idő szintén 200 ms volt. A ki- és bekapcsolási küszöb -45 dB volt. A normalizációs algoritmus a csúcsok levágását dinamika kompresszorral küszöböli ki, amely csak közvetlenül a csúcsot és szűk környezetét befolyásolja. Beszédjel esetén -20 dB-re történő RMS normalizálás esetén ez a dinamika kompresszor általában nem lép működésbe, mert túlvezérlés gyakorlatilag nem jelentkezik. Az egyenszint kiegyenlítést szintén elvégeztem.

A hangadatbázisokat két csoportra osztottam, az első csoportban különböző tartalmú szövegek, a másodikban ugyanaz a szöveg felolvasása található minden beszélőnél.

5.1.1.1. Hangadatbázisok, 1. csoport

Az adatbázisok olvasott beszédet tartalmaznak. A mondatok átlagos hossza 17 szó (87 hang) volt. Az adatbázisok nagy része különböző interaktív hangválaszú rendszerek (IVR) bemondásait, időjárás előrejelzéseket és híreket tartalmaz. Két női és három férfi bemondó felvételeit dolgoztam fel. A felvételi körülmények nem voltak teljesen egyformák. A hangadatbázisok részletes adatait az 5.1. táblázatban adom meg.

5.1. táblázat. Az 1. csoport hangadatbázisainak fontosabb adatai

	Női_Zs	Női_E	Férfi_KMI	Férfi_Sz	Férfi_K
Mondatok száma	7 864	9 122	4 193	1 096	741
Szavak száma	122 344	144 136	83 434	21 793	15 221
Hangok száma	631 070	659 029	413 103	105 104	73 219
Méret [óra]	14,7	15	9,5	2,7	1,8
Tartalom	promptok	promptok és időjárás előrej.	hírek	hírek	hírek
Felvételi hely	stúdió I.	stúdió III.	stúdió II.	stúdió II.	stúdió II.

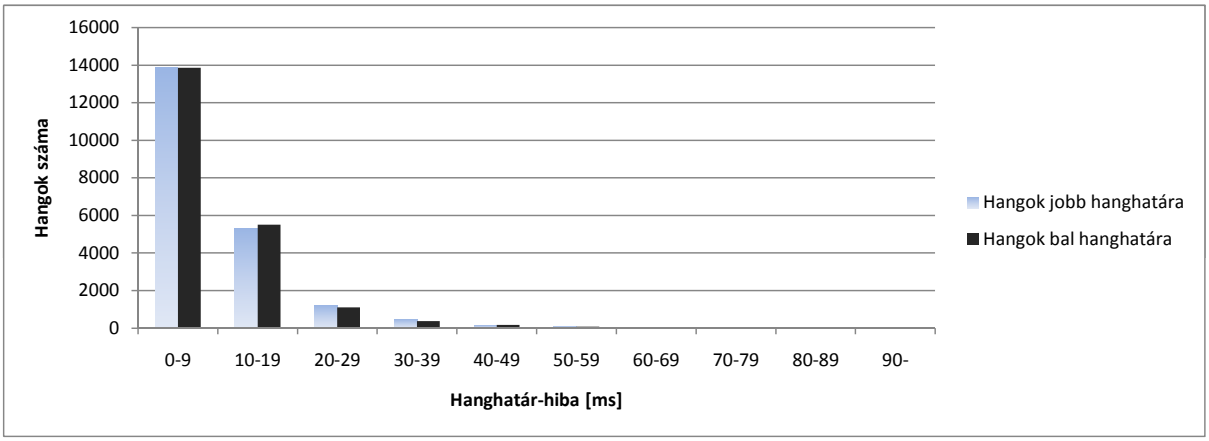
5.1.1.2. Hangadatbázisok, 2. csoport

Az 1. csoport adatbázisaiban a különböző beszélők más-más tartalmú szöveget olvastak fel. Azért, hogy a beszélők közötti különbséget könnyebb legyen vizsgálni, speciális adatbázisokat készítettünk. A 2. csoport adatbázisaiban a beszélők fonetikailag gazdag mondatokat olvastak fel. Ezek lefedik a magyar nyelv leggyakoribb hangkapcsolatait. Az adatbázisok egy női és négy férfi beszélővel készültek. Az adatbázisok tartalmi szempontból egyforma méretűek, 1940 mondatot tartalmaznak. Ez adatbázisonként mintegy 20 000 szót és 82 000 hangot jelent. Egy-egy ilyen beszédadatbázis átlagosan 2,7 órányi beszédet tartalmaz. A mondatok átlagos hossza 10 szó (42 hang). A felvételek a Stúdió I.-ben készültek.

5.1.1.3. Hanghatárok

Az adatbázisokban a hanghatárokat automatikus módszerrel határoztuk meg, ami gépi beszédfelismerésen alapul (Mihajlik et al. 2002). Mivel a felvételek olvasott beszédet tartalmaznak, a szövegátirat a felvételek nagy részéhez rendelkezésre állt. A maradék részhez (hírek) kézzel írtuk le a szöveget. A beszédfelismerő kényszerített felismerési módban

határozta meg a hanghatárokat. Az ablakméret 30 ms volt, 10 ms-os lépésközzel. A felismerés után a hanghatár-hibák nagy része speciális algoritmussal [C6, J10] és kézzel került javításra. A hanghatárok hibájának mértéke nem határozható meg pontosan, mert beszélő és hanganyagfüggő. A teljes adatbázis kézi címkézése lenne szükséges, amely a rendelkezésre álló erőforrások mellett nem megvalósítható. A különböző hanghatár-hibák másként hatnak az intenzitás kiszámolásra. Azokon a helyeken, ahol az intenzításviszonyok lassan változnak, a felismerés pontatlanabb, de a hiba hatás kevésbé érzékelhető. A gyorsan változó helyeken az intenzitás kiszámolása érzékenyebb a hanghatár-hibákra, de ezeken a helyeken a hanghatár meghatározás pontosabb.



5.1. ábra. Automatikus hanghatár-meghatározás pontossága referencia adatbázison

Az automatikus hanghatár meghatározás pontosságát egy referencia adatbázison ellenőriztem. A referencia adatbázis a vizsgálatban szereplő egyik női beszélőtől származó 200 db időjárás témájú mondatot tartalmaz. Ezen az adatbázison kézzel bejelöltem a hanghatárokat. A bejelölt hanghatárokat egy másik személy leellenőrizte. Ugyanezeken a mondatokon automatikusan is elvégeztem a hanghatárok bejelölését és összehasonlítottam a kézi jelöléssel. Az eredményeket az 5.1. ábrán jelenítettem meg grafikus formában. A vízszintes tengelyen a hiba nagysága található ms-ban, függőlegesen a hibatartományba eső hangok száma. A sötét oszlopok jelzik a bal, a világos oszlopok a jobb hanghatár pontosságát. A legtöbb hanghatár 10 ms-nál pontosabb és a hanghatárok 97%-a 30 ms-nál pontosabb. A hanghatárok nagyobb eltolódását több esetben elvi problémák okozzák, például hangsorkezdő pozícióban álló zöngétlen zárhang esetében a hang kezdete nem meghatározható, gyakorlatban a zárfelpattanás előtt kb. 80 ms-mal helyezzük el a hanghatárt.

Az automatikus hanghatár meghatározás az aktuális feladathoz elegendő pontosságú, figyelembe véve azt is, hogy az utólagos félautomatikus javítás tovább pontosítja a gépi felismerés eredményeit.

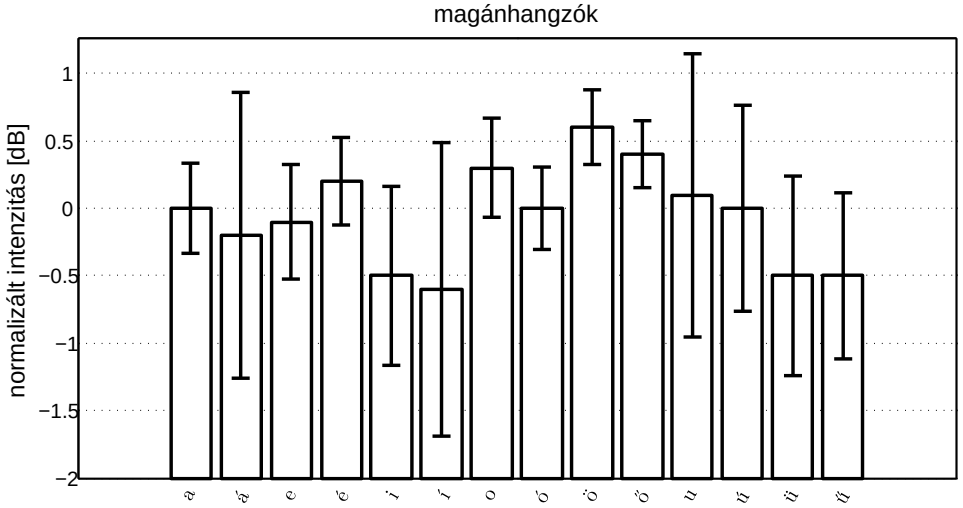
5.1.1.4. A hangok intenzitása

A hangok intenzitásának mérésére effektív értéket (RMS-t) használtam, az 5.1. képlet szerint, ahol $x[n]$ a beszédjel, N_i az i-edik hang mintáinak száma, a $b[i]$ az i-edik hang kezdete.

$$RMS = \sqrt{\frac{1}{N_i} \sum_{n=b[i]}^{b[i]+N_i} (x[n])^2} \tag{5.1}$$

A hangok intenzitását tehát a két hanghatár közötti jelre számítom ki, függetlenül attól, hogy milyen a hang belső felépítése. Az intenzitások kiszámításánál egységesen kezeltem a hangokat, függetlenül attól, hogy egy vagy több szóhoz tartoznak. Szavak határánál az első szó utolsó és a második szó első hangja realizáció szempontjából egyetlen hang lehet. Az ilyen esetekben az adott hang, mindkét szóhoz azonos intenzitással lesz beszámítva. Mivel az intenzitásviszonyok számítása szavanként készül, ezért nevezem szóra vetített intenzitástérképnek az elkészített statisztikát.

5.1.1.5. Hangintenzitások



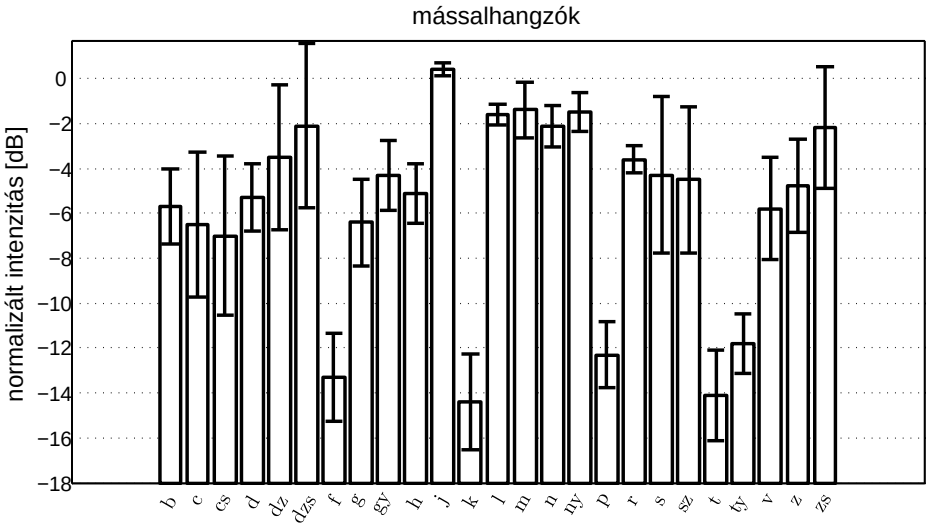
5.2. ábra. A magánhangzók intenzitásátlagai több beszélő folyamatos, felolvasott beszédében

A címkézett adatbázison az intenzitásviszonyokat gépi úton mértem le, az összesített eredmények az 5.2. és az 5.3. ábrán láthatók. A mérésekben az 1-es és a 2-es csoport hangadatbázisai is szerepeltek. 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.

A beszélők közötti szórás mértékét megvizsgáltam külön a 2. csoport hangadatbázisai között is, ahol a felolvasott szöveg azonos volt. Az 5.2. táblázatban láthatók azok a hangok, amelyek intenzitása a beszélők között legkevésbé szórt. Az 5.3. táblázatban azokat a hangokat tüntettem fel, amelyek változatossága a legnagyobb volt.

5.2. táblázat. A legstabilabb hangok az intenzitás szempontjából (csak a 2. csoport hangadatbázisaiból)

Hang	ő	o	ó	j	é	r	ö	a	u	l
Szórás [dB]	0,19	0,21	0,22	0,22	0,26	0,3	0,31	0,32	0,37	0,38



5.3. ábra. A mássalhangzók intenzitásátlagai több beszélő folyamatos, felolvasott beszédében

5.3. táblázat. Legnagyobb változatosságot mutató hangok az intenzitás szempontjából (csak a 2. csoport hangadatbázisaiból)

Hang	sz	c	s	cs	h	z	zs	t	gy	k
Szórás [dB]	3,16	2,69	2,28	1,92	1,57	1,49	1,19	1,18	1,04	1,03

A mért átlag-intenzitásértékek eltérnek Olaszy (1989) munkájától, az adatok kiegyenlítettebbek. Ezt az okozza, hogy olvasott beszéden végeztem a vizsgálataimat, amíg Olaszy csak szavakon és példamondatokon. A vizsgált beszédadatbázisok hangjainak intenzitástérképei az átlagos intenzitásokon alapulnak, amely lehetőséget ad a mondatprozódia során megvalósuló intenzitás eltérések további kutatására. Mivel ez a kutatás elsősorban a korpusz alapú beszédszintetizátorok intenzitás-kiegyenlítéséhez készült, ezért a hangsúlyok és egyéb intenzitást befolyásoló tényezők vizsgálatával jelen kutatásomban nem foglalkoztam, azok hatásai az átlagolás miatt külön nem jelentkeznek.

5.2. Korpusz alapú beszédszintetizátor beszédjelének intenzitás-kiegyenlítése virtuális szóintenzitással

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. Ezért a felvételek intenzitáskiegyenlítése szükséges, amely a felvételi körülmények kontrollálásával, illetve utólagos korrekciókkal részlegesen megoldható. Amennyiben a beszédadatbázis inhomogén – kisméretű önállóan felolvasott elemeket, például csak szavakat vagy csak számokat tartalmaz – akkor ezeknek az elemeknek a kiegyenlítése összetett feladat. A kiegyenlítés már nem valósítható meg a jelfeldolgozó rendszerek általános normalizációs algoritmusaival – például úgy, mint hangadatbázisok mérésénél alkalmazott RMS normalizálással – mert ha az elvileg kiegyenlített elemeket hangsorba illesztjük szintéziskor, akkor 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 – szintetizált eredményt okoz. Az ilyen elemekből származó beszédhangok máshova történő beillesztése intenzitásugráshoz vezet.

5.2.1. Virtuális intenzitás

A korpuszos szintetizátorok nem csak szavakból, szókapcsolatokból építkezhetnek, hanem hang méretű elemekből is (Taylor 2009). A hangok intenzitásai jobban ingadoznak, mint a szavaké. Az ilyen kisméretű építőelemek kiegyenlítéséhez bevezetem a virtuális szóintenzitást, amely megadja, hogy az a szó, amelyben a hang szerepel, milyen intenzitású lenne, ha minden hangja az 5.1. fejezetben 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, $I_{ph}^{average}(ph)$ a hangadatbázisokból kiszámított átlag intenzitás és a $ph(i)$ a szó i.-dik hangja:

$$I_{word}^{virtual} = \frac{\sum_{i=1}^{N} t_{ph}(i) I_{ph}^{average}(ph(i))}{\sum_{i=1}^{N} t_{ph}(i)} \tag{5.2}$$

Ezek alapján meghatározható az adott beszédhang intenzitásmódosításának a mértéke:

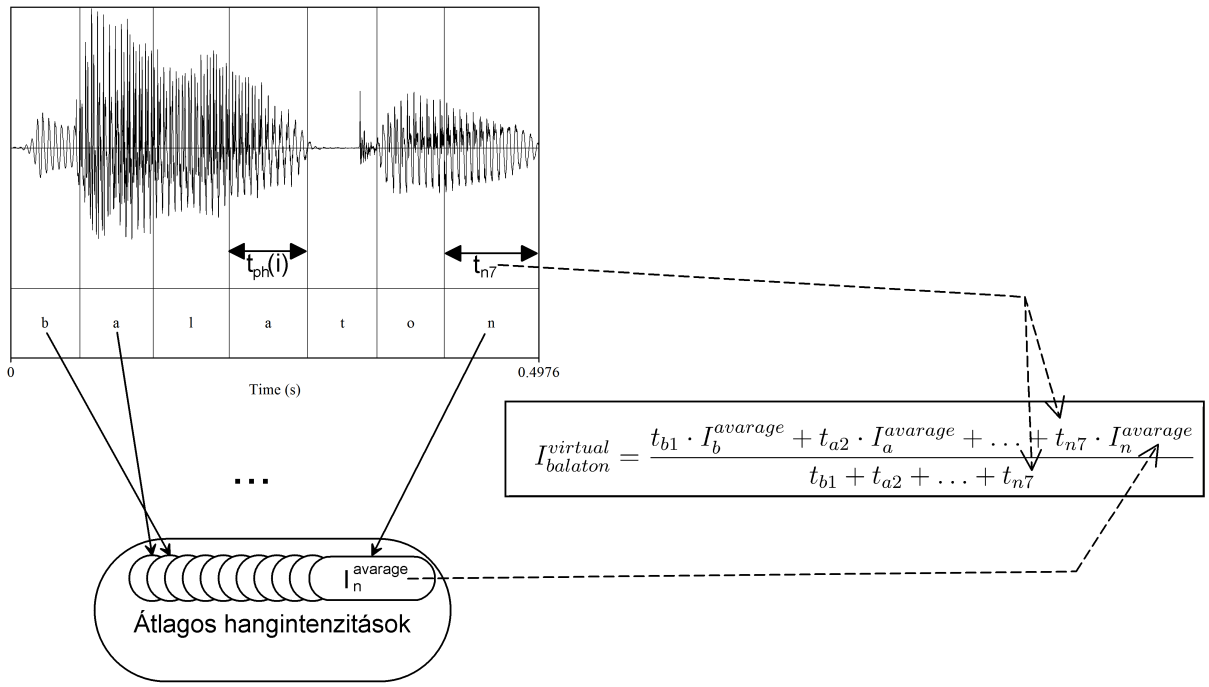
$$gain_{ph} = \frac{gain_{prosody} \cdot gain_{base} \cdot I_{word}^{virtual}}{I_{word}^{real}} \tag{5.3}$$

A $gain_{base}$ állandó adja a mondat átlagos intenzitását, amely megakadályozza az összeállított szintetizált mondat túlvezérlését. A $gain_{prosody}$ pedig a prozódia által meghatározott eltérést adja, amely a hang prozódiai egységen belüli pozíciójától és a hangot tartalmazó szó hangsúlyától függ. Az ismertetett eljárás szemponjából a $gain_{prosody}$ paraméter mint külső bemenetként jelenik meg. Az $I_{word}^{real}(ph)$ adja meg annak a szónak az intenzitását, amelyben a kiegyenlítendő hang szerepel. Az 5.4. ábra illusztrálja a „*Balaton*” szó virtuális intenzitásának kiszámítását.

A virtuális intenzitásnál a hangadatbázisok átlagai alapesetben az 5.1. fejezetben meghatározott értékek. Amennyiben az adott beszélőtől rendelkezésre áll nagyobb hanganyag, akkor az átlagok származtatása történhet csak az adott beszélő adatbázisaiból. Például a korábban bemutatott adatbázisoknál, az 1. csoportban szereplő női bemondók hanganyagaiból önállóan is számítható a hangok átlagos intenzitása (5.1. táblázat).

5.2.2. Percepciós teszt

Az intenzitás kiegyenlítés módszerének jóságát meghallgatásos teszttel vizsgáltam. A tesztben három féle szintetizált mondatot hallgattak meg a tesztelők. Az első változatban kiegyenlítés nélküli mondatok voltak, amelyeket referenciaként hallgattak meg a tesztelők. A második változatban csúcsra kiegyenlített mondatok szerepeltek, ahol a hangokat a bennük szereplő abszolút legnagyobb kivezérlés szerint egyenlítettem ki. A harmadik változat mondataira a virtuális intenzitás alapú eljárást alkalmaztam. A meghallgatásos teszt web alapú volt, 16



5.4. ábra. A „*Balaton*” szó virtuális intenzitás számításának bemutatása

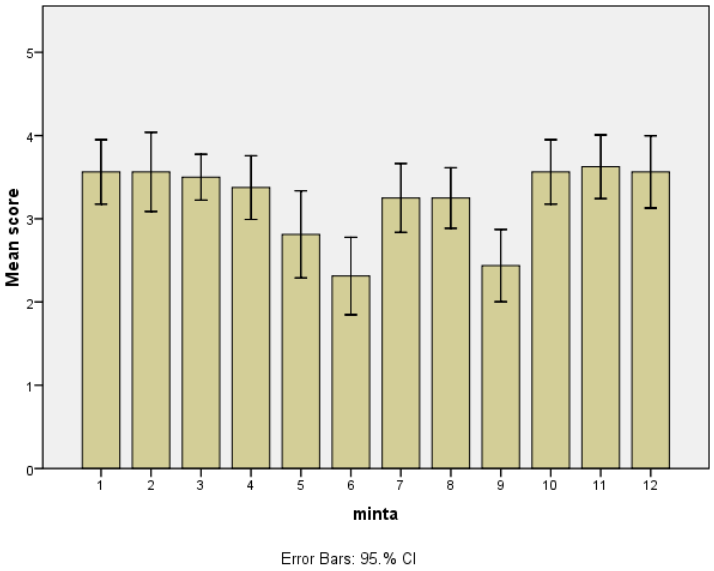
fő végezte el. A teszt két részből állt, az egyik részben 1-5-ös (MOS) skálán értékelték a minőséget (12 mondat), a másik részben pedig páros összehasonlítást végeztek ugyanazokon a különböző módszerrel kiegyenlített mondatokon (18 mondatpár). A páros összehasonlításban *az első sokkal jobb, az első jobb, egyforma, a második jobb, a második sokkal jobb* lehetőségek közül lehetett választani.

5.2.3. Eredmények

A percepciós teszt eredményei az 5.5. és az 5.6. ábrákon láthatók. Az ábrákon található számozás magyarázata a mellette elhelyezett táblázatban olvasható. A mondatok jelölésénél a „*nincs*” verzió jelöli azt a mondatot, amelynél intenzitáskiegyenlítés nem történt. A „*csúcs*”-csal jelölt mondatok esetén a kiegyenlítés a csúcsra történt. A „*virtuális*” mondatok a korábban ismertetett algoritmussal kiegyenlített mondatok, ahol a hangok átlagos intenzitásai saját adatbázisokból számítottak. A „*más*” jelölővel ellátott mondatoknál annyi változtatás történt a jelölő nélküli mondatokhoz képest, hogy nem a saját adatbázisuk átlagos intenzitásértékeivel történt a virtuális intenzitás számítása.

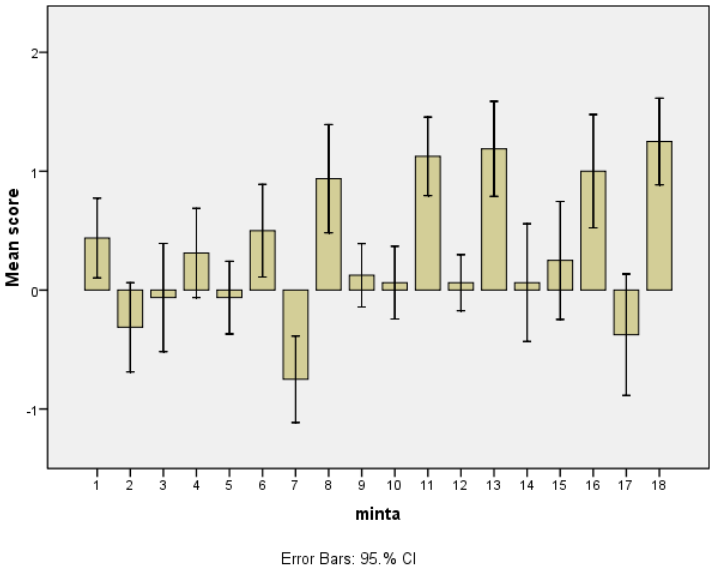
A mondatok első (MOS tesztben: 1–4. mondat, Páros összehasonlításnál: 1–6. mondat) csoportjába tartozó férfi hangadatbázisának rögzítése felügyelt körülmények között történt. A felolvasott mondatok egyforma intenzitásúak, ezért a vártnak megfelelően a különböző verziók között nem mutatkozott szignifikáns ($p<0.05$) eltérés.

A második (MOS: 5–8., Páros: 7–12.) csoport női hangjai között a csúcsra húzás módszere adta a legrosszabb eredményt. Ez a hangadatbázis kevésbé homogén, mint az előző férfi adatbázis, így a kiegyenlítés nélküli módszert rosszabbnak ítélték a tesztelők. A kiegyenlítettek közül a virtuális szóintenzitáson alapuló módszert szignifikánsan jobbnak ítélték.



1	Nincs	Férfi
2	Csúcs	
3	Virtuális más	
4	Virtuális	
5	Nincs	Női
6	Csúcs	
7	Virtuális	
8	Virtuális más	
9	Nincs db1	Női
10	Nincs db2	
11	Csúcs	
12	Virtuális	

5.5. ábra. MOS teszt eredményei



1	Nincs	Csúcs	Férfi
2	Csúcs	Virtuális mas	
3	Virtuális mas	Virtuális	
4	Nincs	Virtuális mas	
5	Csúcs	Virtuális	
6	Nincs	Virtuális	
7	Nincs	Csúcs	Női
8	Csúcs	Virtuális	
9	Virtuális	Virtuális mas	
10	Nincs	Virtuális	
11	Csúcs	Virtuális mas	
12	Nincs	Virtuális mas	
13	Nincs db1	Nincs db2	Női
14	Nincs db2	Csúcs	
15	Csúcs	Virtuális	
16	Nincs db1	Csúcs	
17	Nincs db2	Virtuális	
18	Nincs db1	Virtuális	

5.6. ábra. Páros összehasonlítások eredményei

A harmadik (MOS: 9–12., Páros: 13–18.) csoportban két kiegyenlítés nélküli mondat is szerepelt. A „*db1*”-es adatbázis egy változatos adatbázis, a „*db2*”-es egy kisebb, a mondat témájához illeszkedő kiegyenlített homogén adatbázis. A kiegyenlítéses mondatok a változatos („*db1*”) adatbázisból készültek. A vártnak megfelelően a változatos adatbázis mondata rosszabbul szerepelt, a kiegyenlítés szignifikánsabban jobb eredményt ad.

A mondatok értékelésénél tehát a kiegyenlítés nélkülihez képest szignifikánsan (95%-os konfidencia intervallum mellett) jobb volt a másik két módszer, de a kiegyenlítettek között szignifikáns különbség nem volt kimutatható. A páros összehasonlításoknál T-próbát végeztem, a MOS teszt esetében ANOVA analízist. 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.

5.3. Összegzés

Olvasott beszédadatbázisokat felhasználva meghatároztam a magyar nyelv hangjainak intenzitástérképét. Módszert mutattam be a korpusz alapú elemkiválasztásos szintetizátor virtuális intenzitáson alapuló hang szintű intenzitáskiegyenlítésére.

6. fejezet
Beszédszintetizátorok hangjának érzelmi módosítása

A beszédszintetizátorok fejlesztése során az érthető, jó minőségű beszéd előállítása az alapvető cél. A szintetizátorok többnyire csak köznapi beszédet állítanak elő (közlések, hírek, időjárás-jelentés), amely érzelmi töltettel nem rendelkeznek. Ennek több oka is lehet, például a modellezés nem terjed ki ilyen területre, vagy a kiinduló beszédadatbázis semleges érzelmű köznapi beszédet tartalmaz. 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 módszer az, hogy a beszédet olyan transzformációnak vetjük alá, amely az érzelmi töltetűhöz hasonló beszédet állít elő. Ez az eljárás nem biztosít az első módszerrel azonos minőséget, de lehetőséget ad arra, hogy a már meglévő beszédadatbázisokból készítsünk érzelmi töltetűeket.

6.1. Érzelmi töltetű beszéd

Az érzelem kifejezése és felismerése összetett folyamat, sok modalitás együttese adja az azonosítható érzelmeket. A beszéd mellett a mimika, a gesztusok is érzelmet fejezhetnek ki. A továbbiakban csak az akusztikus beszédjel érzelmi töltetével foglalkozom.

A beszéd érzelmi töltete több komponensből áll össze. Az egyik a beszéd tartalmi része, a másik az akusztikai megvalósítása. A beszéd tartalmi része jelentősen befolyásolja az érzelem percepciós azonosítását. A gépi beszédkeltés során a tartalom sokszor külső paraméter, amelynek befolyásolására nincs lehetőség. Ezért munkám során az akusztikai területre koncentráltam. A tesztek, vizsgálatok alkalmával a tartalom befolyásoló hatását úgy csökkentettem, hogy a felhasznált mondatok általános témakörből származtak és nem tartalmaztak érzelemre utaló szavakat és kifejezéseket.

A legfontosabb akusztikai paraméterek, amelyek a különböző érzelmek kifejezésében részt vesznek a következők: dallammenet (az alapfrekvencia (F_0) változása), intenzitásmenet, artikulációs- és beszédsebesség, formánsmenetek, spektrális zajok, spektrális komponensek intenzitásának arányai és a glottalizáció mértéke (Scherer 2003).

A természetes érzelmes beszéd forrását két csoportba oszthatjuk. Az egyik átélt, a másik eljátszott érzelmet tartalmaz. Speciális esetben – például konkrét személy utánzásakor – az átélt érzelem megvalósítása a cél a gépi beszédszintézisben, de általános esetben az eljátszott érzelem megvalósítása a célravezető, mert az embereknek az átélt érzelmeket nehezebb azonosítani, mint az eljátszott érzelmet (Banse–Scherer 1996). Az érzelmes beszéd ember általi felismerhetősége sem 100%-os. A tanulmányok szerint (Scherer 2003, Tóth et al. 2007b, Laukka 2004) az érzelmek azonosítása a tartalmi információk nélkül 60-70%-os. Petrushin

(2000) kísérletében a beszélők saját érzelmeiket is csak 80%-os pontossággal ismerték fel. Ezeket az adatokat az érzelmi töltetű gépi beszédszintézis értékelésekor figyelembe kell vennünk, az elvárható azonosítási pontosság tehát legfeljebb 60-80% körül van.

6.2. Érzelmi töltetű gépi beszédszintézis

A gépi beszédszintézis területén több módszert alkalmaznak érzelmi töltetű beszéd előállításához. Az egyik megközelítés az, hogy a semleges hangadatbázist bővítik ki érzelmet kifejező elemekkel (Hamza et al. 2004). A semleges adatbázisból tetszőleges semleges mondat előállítható, míg a hozzáadott, érzelmet kifejező elemek önmagukban ilyenformán nem használhatók, csak kiegészítik a semleges adatbázist. Ezek az elemek lehetnek szavak, kifejezések, mondatok vagy egyéb kifejező hanghatások, mint például nevetés, sóhajtás vagy hümmögés (Baggia et al. 2006).

A másik megoldás az adatvezérelt érzelmi szintézis, ahol nagymennyiségű érzelmes beszédet rögzítenek (Campbell 2001). A felvételeket valós körülmények között készítik, az alany folyamatosan egy fejmikrofont és hangrögzítő eszközöket visel. Ezzel a módszerrel átélt érzelmek jó minőségben rögzíthetők, de az adatgyűjtés erőforrás-igényes, mind szervezési, mind technikai és utófeldolgozási szempontból.

A harmadik megoldás az, hogy érzelmi modellt dolgozunk ki és ez alapján módosítjuk a semleges szintetizált beszédet. A modell alkalmazása történhet a szintetizátor adatbázisain, a szintetizálás során, utólag a kész beszédjel módosításával vagy ezen szintek kombinációjával (Kawahara et al. 2005, Přibilová–Přibil 2006).

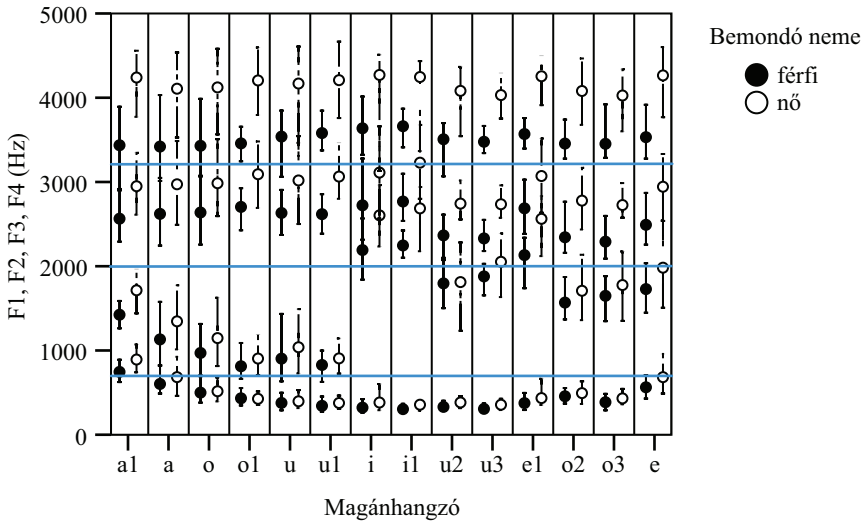
6.3. Nemlineáris frekvencia tartománybeli transzformáció

A beszéd érzelmi töltetének módosításához Přibilová–Přibil (2009) transzformációt hajtottak végre a frekvenciatartományban. Publikációjukban az érzelmes mondatok semlegessé konvertálását írták le. A módosítás azon a megfigyelésen alapul, hogy érzelmek hatására a beszédképzésért felelős vokális traktus megváltozik (Scherer 2003). Az első és a többi formáns értéke ($F_{2,3,4}$) érzelem hatására ellentétes irányba változik. Pozitív érzelmek hatására az F_1 csökken, a magasabb formánsok növekednek, míg negatív érzelem esetében F_1 nő és a magasabb formánsok csökkennek. A formánsok értéke nyelv- és hangfüggő. A transzformációt úgy alakították ki, hogy robusztus legyen, ne függjön a formánsanalízis pontosságától. Emiatt nem törekedtek a konkrét formánsfrekvenciák meghatározására, hanem megvizsgálták, hogy általában milyen tartományban fordulnak elő a formánsfrekvenciák (6.1. táblázat).

6.1. táblázat. Formánsfrekvenciák tartománya férfi beszélőre Přibilová–Přibil (2009) összegzésében

Formáns	F_1	F_2	F_3	F_4
Alsó határ	$\approx$ 250	$\approx$ 700	$\approx$ 2000	$\approx$ 3200
Felső határ	$\approx$ 700	$\approx$ 2000	$\approx$ 3200	$\approx$ 4000

A becsült formánstartományokat összehasonlítottam Olaszy (2010) által magyar nyelvre mért formánsfrekvenciákkal. A 6.1. ábrán látható a magyar nyelv 14 magánhangzójának 4 formánsfrekvenciája, amelyre vízszintes vonallal behúztam a Přibilová és Přibil által megállapított határokat (700 Hz, 2000 Hz, 3200 Hz). A meghatározott határfrekvenciák jól illeszkednek a magyar magánhangzók formánsfrekvenciájihoz, tehát magyar nyelvre is alkalmazhatók. Přibilová és Přibil a frekvencia-transzformációt egy nem lineáris függvénnyel



6.1. ábra. Az F_1, F_2, F_3, F_4 formánsok frekvenciasávjai egy férfi és egy női bemondó ejtéséből mérve a referencia-formánsadatbázis szóanyagán. A diagramok a magánhangzó 50%-os pontjában mért adatokat mutatják. Az ábra forrása: Olaszy (2010, 110.o.)

adták meg, amelyet a jobb kezelhetőség érdekében két részre bontottak. A transzformáció egyik részét a 6.1. egyenlettel írják le, amely logaritmikusan széthúzza az alacsony frekvenciákat a jobb kezelhetőség érdekében.

$$f(f_t) = a \cdot b^{f_t} + c \tag{6.1}$$

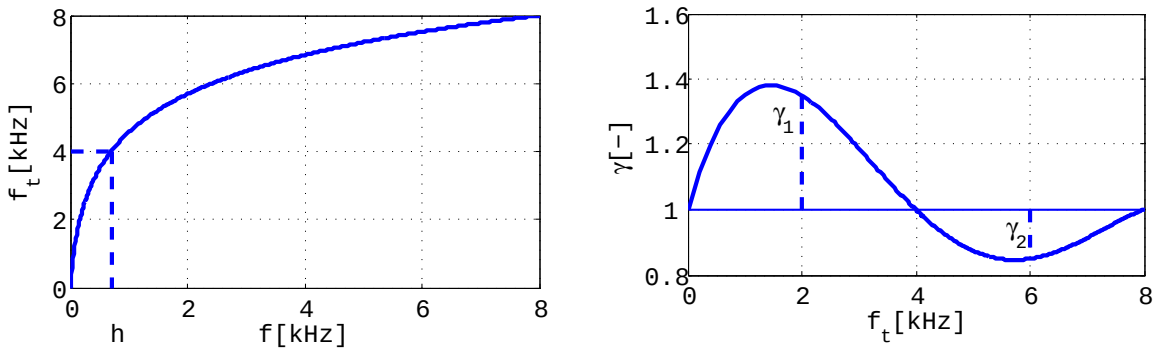
A három ismeretlen paramétert (a,b,c) három pont segítségével határozták meg: $[f_t, f]$ = [0 kHz, 0 kHz], [4 kHz, h kHz], [8 kHz, 8 kHz], ahol h az F_1 és az F_2 határfrekvenciája, amelyet férfi beszélő esetén 0.7 kHz-re tesz (6.1. táblázat). Az f_t és f jelöli a transzformáció cél és az eredeti frekvenciaértékeit. A megoldást a rögzített három pont mellett a 6.2. egyenlet adja meg.

$$a = \frac{h^2}{8-2h}, \qquad b = \sqrt[4]{\frac{8-h}{h}}, \qquad c = -a \tag{6.2}$$

A transzformált frekvencia kifejezhető a 6.3. egyenlet formájában. A függvény grafikus megjelenítése a 6.2a. ábrán látható, ahol a h értéke 0.7 kHz.

$$f_t(f) = \log_b \frac{f+a}{a} = \frac{\ln(\frac{f}{a}+1)}{\ln(b)} \tag{6.3}$$

A formánsok eltolásának mértékét egy negyedfokú polinom (6.4. egyenlet) adja meg, amelyet 5 pont rögzítésével határoz meg: $[f_t, \gamma]$ = [0 kHz,1],[0 kHz,γ_1],[4 kHz,1],[6 kHz,γ_2],[8 kHz,1], ahol γ az f_t frekvencia módosító szorzója.



(a) $f_t(f)$ *(b)* $\gamma(f_t)$

6.2. ábra. A spektrális transzformáció elemi függvényei

$$\gamma(f_t) = p \cdot f_t^4 + q \cdot f_t^3 + r \cdot f_t^2 + s \cdot f_t + t \tag{6.4}$$

A polinom konstansaira a következő értékeket kapjuk:

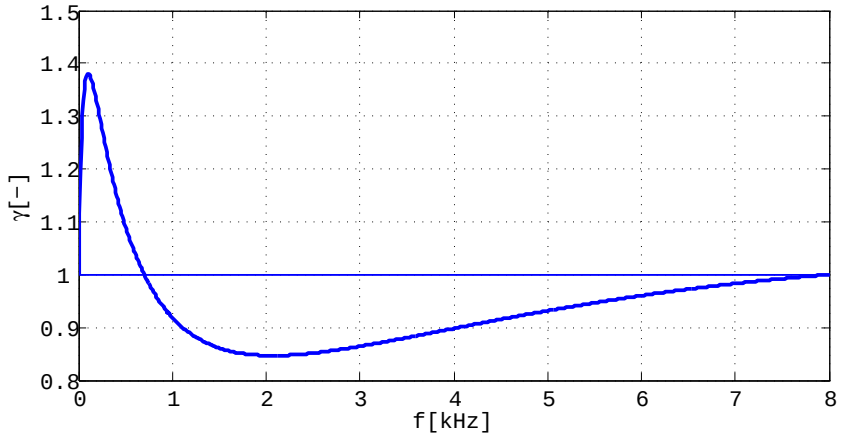
$$\begin{aligned} p &= \frac{1}{48} - \frac{1}{96}\gamma_1 - \frac{1}{96}\gamma_2, \qquad q = -\frac{1}{3} + \frac{3}{16}\gamma_1 + \frac{7}{48}\gamma_2, \\ r &= \frac{5}{3} - \frac{13}{12}\gamma_1 - \frac{7}{12}\gamma_2, \qquad s = -\frac{8}{3} + 2\gamma_1 + \frac{2}{3}\gamma_2, \qquad t = 1. \end{aligned} \tag{6.5}$$

A 6.4. függvény grafikus megjelenítése a 6.2b. ábrán látható, $\gamma_1 = 1.35$ és $\gamma_2 = 0.85$ értékeket behelyettesítve.

A módosított spektrum a 6.3. és a 6.4. egyenletek segítségével számítható a következő módon:

$$E'(f) = E\left(\frac{f}{\gamma(f_t(f))}\right) \tag{6.6}$$

A frekvenciakomponensek eltolásának grafikus megjelenítése a 6.3. ábrán látható, a konstansok megegyeznek a 6.2. ábra konstansaival.



6.3. ábra. A spektrális transzformáció egyesített függvénye

A módosítás 0-8 kHz-es tartományra adja meg a transzformációs függvényt, amely a beszédfeldolgozásban gyakran használt 16 kHz-es mintavételi frekvenciához illeszkedik. Magasabb mintavételi frekvencia esetén a függvény két módon is kiterjeszthető a 8 kHz feletti frekvenciákra. Az egyik megoldás, hogy a transzformációt csak 8 kHz-ig végezzük el, a magasabb frekvenciatartományokon spektrális módosítást nem végzünk. Ez a megoldás azon alapul, hogy a beszéd energiájának legnagyobb része a 8 kHz alatti részre összpontosul, ezért ez az egyszerűsítés végrehajtható. A másik megoldás, amit Přibilová–Přibil (2009b) későbbi munkájában szintén alkalmazott más mintavételi frekvenciák (F_s) esetében, hogy a transzformációs függvények felső határát kitolta $F_s/2$-ig.

6.4. Transzformáció a beszédjelen

Přibilová és Přibil a spektrális módosítást egy kepsztrum alapú szintetizátorral valósították meg, majd későbbi munkájukban LPC alapú algoritmusba integrálták a módszert (Přibilová–Přibil 2006, 2009b). Az általuk ismertetett eljárást fejlesztettem tovább, amely a PSOLA algoritmus analízis-szintézis modelljén alapul.

Az algoritmus első lépésében zöngeperiódus jelölőket – angol terminológiában pitchmarkokat – helyezünk el, amelyek a zöngés szakaszokon a hangszalag mozgásával szinkron amplitúdó csúcsokat jelölnek a beszédjelen. A jelölők meghatározásához autókorrelációs elven működő zöngeperiódus meghatározást alkalmazok (Boersma 1993). A zöngés szakaszokon tehát rendelkezésre állnak a zöngeszinkron jelek. Jelölésük: $t_a[i]$, ami megadja az i-edik zöngeperiódus szakasz kezdetét, így az eredeti jel zöngés szakasza felírható a zöngeperiódusok összegeként a 6.7. egyenlet formájában, ahol $x_i[n]$ az eredeti jelszakasz ablakozása.

$$x[n] = \sum_{i=-\infty}^{\infty} x_i[n - t_a[i]], \qquad x_i[n] = w_i[n]x[n] \tag{6.7}$$

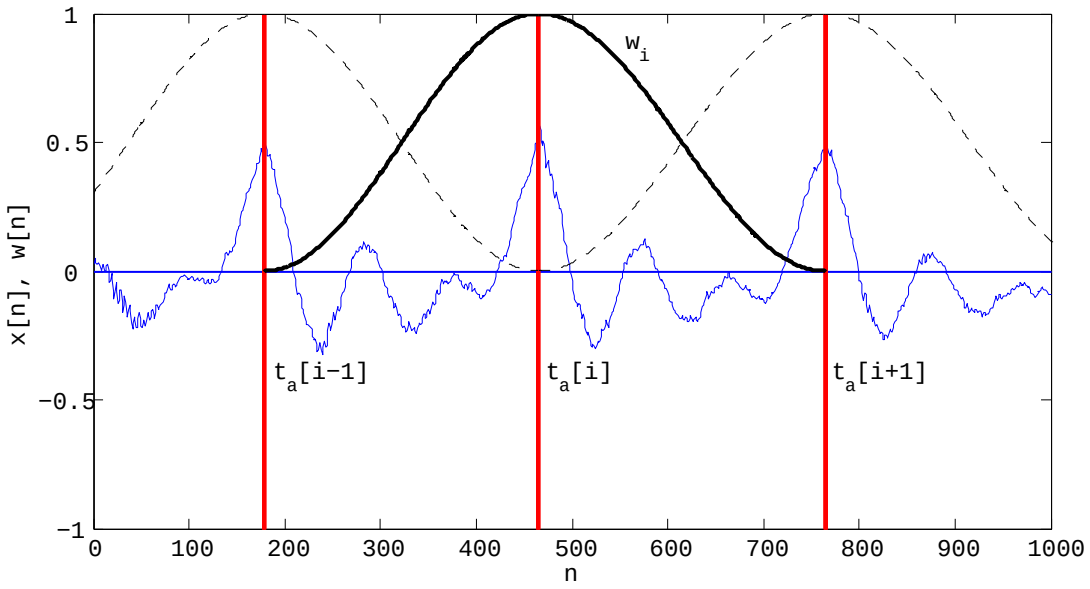
Az ablakozó függvényre teljesülnie kell a 6.8. egyenletnek.

$$\sum_{i=-\infty}^{\infty} w_i[n - t_a[i]] = 1 \tag{6.8}$$

Ablakozó függvényként Hann-ablakot használok, amely teljesíti az előző feltételt:

$$w[n] = 0,5\left(1 - \cos\left(\frac{2\pi n}{N-1}\right)\right) \tag{6.9}$$

A beszédjel ablakozását a 6.4. ábra illusztrálja. A kiablakozott $x_i[n - t_a[i]]$ jel diszkrét Fourier transzformálásával (DFT) áttérek frekvenciatartományba $X_i = \mathscr{F}(x_i[n - t_a[i]])$. A jelfeldolgozásban széles körben elterjedt FFT (Fast Fourier Transformation) itt nem használható, mert a zöngeperiódus jelölők távolsága tetszőleges lehet, így az ablakméret – a legtöbb esetben – nem 2 hatványa. A DFT algoritmus frekvenciafelbontása függ az alkalmazott mintavételi frekvenciától, az ablakméret pedig a beszélő alapfrekvenciájától. Mivel az ablakméret az alapfrekvenciához tartozó periódus idő kétszerese, a DFT felbontása a mintavételi frekvenciától függetlenül az alapfrekvencia fele lesz. Ez a zöngés beszédhangok szempontjából előnyös, mert így a zönge felharmonikusaira rendelkezünk információval. Az X jelet szétválasztom amplitúdó- (A^i) és fázisspektrumra (φ^i) majd végrehajtom a következő módosításokat:



6.4. ábra. A zöngés beszédjel ablakozásának illusztrációja

Transzformáció: A 6.3. fejezetben ismertetett transzformáció (6.6. egyenlet) inverzét végrehajtom a spektrumon, a fázisokat változatlanul hagyom.

Intenzitás: Az érzelemre jellemző intenzitásnak megfelelően konstans szorzóval a teljes spektrumot módosítom.

Spektrális energia eloszlás: Az alacsony és a magas frekvenciájú spektrális komponensek aránya változik a különböző érzelmek esetén. A módosításhoz rögzítek 2 frekvenciaértéket (f_{also}, f_{felso}) és az ezekhez tartozó szorzókat (g_{also}, g_{felso}), amelyek megadják az alsó és felső tartomány erősítési értékeit. A teljes frekvenciatartományra a 6.10. képlet szerint alkalmazom az erősítést, a két rögzített frekvencia közötti értékekre átmenetet alkalmazok a két erősítési érték között.

$$g(f) = \begin{cases} g_{also} & \text{ha } f \leq f_{also} \\ (g_{felso} - g_{also}) \cdot \left(\frac{f - f_{also}}{f_{felso} - f_{also}}\right) + g_{also} & \text{ha } f_{also} < f < f_{felso} \\ g_{felso} & \text{ha } f \geq f_{felso} \end{cases} \tag{6.10}$$

A módosítások elvégzése után inverz Fourier transzformációt végzek:

$$x'_i[n - t_a[i]] = \mathscr{F}^{-1}(X'_i) \tag{6.11}$$

ahol $X'_i = [A'^i, \varphi^i]$ a módosított amplitúdó- és az eredeti fázis-spektrumból áll össze. A módosított értékek miatt az inverz Fourier transzformáció után a jel két végpontján nem garantáltak a 0 értékek, amit a transzformáció előtt még w_i biztosított. Azért, hogy a későbbi összeadásnál ne keletkezzen ugrás a jelben, újabb ablakozást végzek el x'_i-n egy módosított Hann ablakfüggvénnyel. A végpontok környékén – az x'_i első és utolsó 10%-án – a Hann ablaknak megfelelő koszinusz görbe szerinti ablak található. A görbe a 10% alatt eléri az 1 értéket, ami a 90%-ig tart, majd koszinuszos lefutással a másik végponthoz szimmetrikusan lemegy 0-ra.

Az új szintetizált jelet a 6.12. egyenlet adja meg, ahol $y[n]$ az új jel, x'_j a módosított ablakozott jelszakasz, $t_s[j]$ pedig a j-dik jelszakasz új helye.

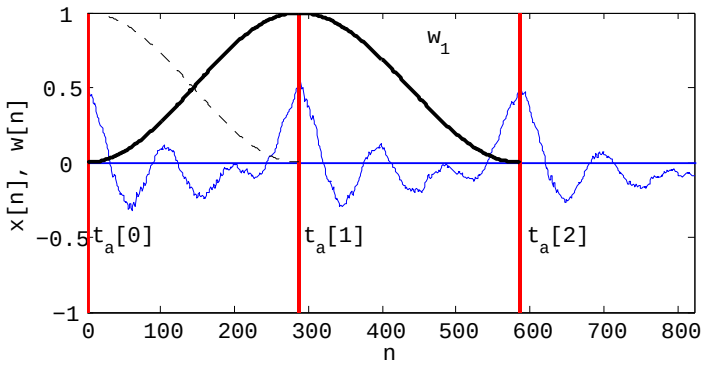
$$y[n] = \sum_{j=-\infty}^{\infty} x'_j[n - t_s[j]] \tag{6.12}$$

A $t_s[j]$ sorozat megválasztásával tudjuk a PSOLA algoritmus esetében az alapfrekvenciát és az időtartamot módosítani. Az eredeti jel periódusidejét $t_a[i] - t_a[i-1]$ adja meg. Amennyiben ezeknek a zöngeperiódus jelölőknek a távolságát módosítjuk az új szintetizált jel készítésekor, az alapfrekvencia változni fog, az új periódusidő $t_s[j] - t_s[j-1]$ lesz. Az időtartam módosítására, illetve az alapfrekvencia változtatás időtartam módosító hatásának kompenzálására x'_i kihagyható vagy megismételhető. Kihagyás esetén az időtartam csökken, ismétlés esetén az időtartam növekszik.

A zöngétlen szakaszok esetében két megoldás alkalmazható. Az egyik, hogy a transzformációt nem végezzük el, így az érzelmi módosítás nem teljes. Azonban mivel a zöngétlen hangok energiája többnyire a zöngéseknél lényegesen kisebb és jellemzően a módosított alacsony frekvenciájú komponensek gyengék, a spektrális transzformáció hiánya általában nem észlelhető. A zöngétlen szakaszok változatlan formában történő felhasználása esetén ezeken a szakaszokon a prozódiai változtatások egy része sem történik meg. A prozódia alapfrekvencia komponense ezen szakaszokon nem értelmezhető. Az energiakorrekció időtartományban is elvégezhető. Az időtartam korrekció viszont nem végezhető el ebben az esetben a PSOLA módszerrel. A hiányzó spektrális változtatás részét képező alacsony-magas frekvenciakomponensek energiaarányának módosítása a legtöbb zöngétlen hangnál nem is lenne elvégezhető, mert ezek a hangok alig tartalmaznak 1 kHz alatti komponenst.

A másik lehetőség az, hogy a zöngétlen részekre egyenlő távolságra virtuális zöngeperiódus-jelöléseket helyezünk el, és a zöngés részekkel azonos módon kezeljük. Ez a megoldás lehetővé teszi, hogy minden paramétert a zöngés hangokhoz hasonlóan módosítsunk a beszédjelen. A zöngétlen hangok ilyen kezelése kismértékben robotossá, fémessé teszi a beszédet, mivel olyan beszédszakaszok válnak a módosítás során periodikussá, amelyek egyébként nem azok. A fémesség lehetősége miatt az érzelemmódosító algoritmusomban ezt a megoldást nem alkalmaztam, kivéve a triádos szintetizátor esetében, ahol egy ilyen virtuális zöngeperiódus-jelölőn alapuló megoldás van a szintetizátorba építve.

A beszédjel a feldolgozás szempontjából belépő jel, illetve véges időre korlátos. A feldolgozás a legtöbb esetben olyan beszédszakaszokra történik, amelynek mindkét végén valamilyen szünet típusú szakasz található, amely nem igényel feldolgozást. A gépi beszédszintetizátorok feldolgozási egysége általában a mondat, amelyet mondatközi szünetek határolnak, így az algoritmus alkalmazható. Az algoritmus a jel kezdeti és végső részét nem módosítja, a legelső és a legutolsó ablakozás csak félig megvalósítható, a fél ablaknyi jelen a módosítás nem hajtható végre. Mivel a mondatok esetén a beszédjel eleje és vége szünet, így ez nem jelent problémát. A 6.5. ábrán látható egy olyan beszédjel kezdete és annak ablakozása, ahol a jel eleje nem szünettel indul. Ilyen beszédjelekkel a triádos szintetizátor hangadatbázisánál találkozhatunk, ahol az adatbáziselemek a beszédből kivágott rövid 100-200ms hosszú elemekből állnak. Ebben az esetben a hasznos jel kezdő és végső része csak részben lesz módosítva, az elem elején és végén tipikusan 5ms-os rész az ablakozó függvénynek megfelelően csak részben lesz transzformálva.



6.5. ábra. Egy zöngés beszédelem kezdetének ablakozása

6.5. A módosítás paraméterei

Az érzelmi töltetű beszéd előállításához szükséges paraméterek a 6.2. táblázatban láthatók. A módszer teszteléséhez a paramétereket [C9] és (Přibilová–Přibil 2009) munkájából származtattam. A táblázatban szereplő γ_1 és γ_2 értékek határozzák meg, hogy a h határfrekvencia alatt és feletti komponenseket a spektrális transzformáció merre tolja el. Ezeket a paramétereket meghagytam – a könnyebb összehasonlítás érdekében – Přibilová és Přibil eredeti értelmezésében, ami azt jelenti, hogy a transzformációs konstansok az érzelmes mondat semleges mondattá való konvertálásához rögzítettek. Az érzelmes mondatok előállítása során tehát a transzformáció inverzét kell elvégezni. Korábban a 6.2. és a 6.3. ábrákon a haragos érzelem konstansaihoz tartozó függvénymeneteket láthattuk. Haragos mondat transzformálása esetén tehát a 700 Hz-es határfrekvencia alatti komponensek és az F_1-es formáns csökken, az ennél magasabbak növekednek. A táblázatban a γ_1 és a γ_2 paraméteren kívül a többi paraméter

6.2. táblázat. Az egyes érzelmekhez tartozó konstansok

Érzelem	h határfrekvencia	γ_1	γ_2	F_0 szorzó	F_0 tartománya	f_{also}	f_{felso}	g_{also}	g_{felso}	energia	időtartam
Haragos	0.7kHz*	1,35**	0,85**	1,15	1,3	1kHz	2kHz	1,0	1,7	1,7	0,87
Örömteli	0.7kHz*	0,70**	1,05**	1,18	1,3	1kHz	2kHz	1,0	1,5	1,3	0,85
Szomorú	0.7kHz*	1,10**	0,90**	0,84	0,7	1kHz	2kHz	1,7	1,0	0,5	1,11

* Férfi beszélő esetén. ** Inverz transzformáció paraméterei

a semleges mondatból az érzelmes mondat előállításhoz szükséges állandó.

A nemlineáris transzformáció fenti paraméterekkel való alkalmazása módosítja például a magánhangzók formánsfrekvenciáit – amely az elsődleges célja az algoritmusnak, – de ezt csak akkora mértékben teszi meg, hogy a beszéd érthetősége ne csökkenjen jelentős mértékben. A különböző magánhangzók formánsfrekvencia értékei megmaradnak abban a tartományban, amely alapján még ugyanazt a beszédhangot tudja az ember azonosítani.

6.6. Kísérletek

6.6.1. Korpuszos beszédszintetizátor mondatai és a természetes beszéd

Az érzelemmel töltött mondatok teszteléséhez korpusz alapú, elemkiválasztásos beszédszintetizátort választottam. A célom az volt, hogy igazoljam, hogy a spektrális módosítással az érzelmek felé tolható el a beszéd hangzása. Kiválasztottunk egy természetes semleges emberi bemondást, és két olyan mondatot, amelyeket a korpusz alapú beszédszintetizátorral állítottunk elő. Ez a három mondat képezte a vizsgálat alapanyagát.

A korpusz alapú, elemkiválasztásos beszédszintetizátor - a továbbiakban korpuszos szintetizátor - egy olyan szövegfelolvasó, amely nagy mennyiségű előre rögzített beszédből (beszédkorpuszból) válogatja ki a megfelelő elemeket és azok összekapcsolásával állítja elő a szintetizált beszédet. A beszédkorpusz mérete nagy, több tíz-száz órányi beszédet is tartalmazhat [J8]. A szintetizált mondat hangminősége függ a beszédkorpusz méretétől és attól, hogy a mondat mennyire felel meg a korpusz tematikájának (szóhasználatának). A beszédkorpusz nagy mérete miatt az érzelmi módosítást nem érdemes elvégezni magán a korpuszon, helyette az utólagos módosítás módszerét választottam. A korpuszos szintetizátor nem végez jelfeldolgozást a beszéden, de az adatbázisában a zöngeperiódus-jelölők rendelkezésre állhatnak. A feldolgozás alapjául szolgáló szintetizált beszéden tehát már nem kell a zöngeperiódusokat meghatározni, ez már rendelkezésre áll.

A szintetizált mondatok egy férfi és egy női hangú szintetizátorral készültek. A módosító algoritmus minden esetben az előző fejezetben leírt volt. A módosítások paramétereit a 6.2. táblázatban már láthattuk. A mondatok szövege a következő (a természetes bemondás szövege megegyezik az első szintetizált mondatéval):

- „*A menüben minden szükséges információ elhangzik.*”
- „*Változóan felhős az ég, csapadékról nem érkezik jelentés.*”

A meghallgatásos vizsgálatot egyénileg végezték el az tesztelők, webes felületen keresztül, 25 magyar anyanyelvű, ismert halláskárosodással nem rendelkező személy. Az összesítésből 3 hallgatót zártunk ki, akik vagy nem fejezték be a tesztet, vagy véletlenszerűen választottak. A maradék 22 hallgató 15 férfiből és 7 nőből állt. A tesztelők átlagos életkora 38 év volt. 8 alany fej- és fülhallgatót használt, míg 14 tesztelő hangszórón hallgatta meg az anyagot. A meghallgatásos teszt 9,6 percet vett igénybe átlagosan. A hallgatóknak lehetőségük volt az adott tesztmondat újbóli meghallgatására, de a már elküldött értékelést nem módosíthatták, visszafelé nem léphettek. A lejátszási sorrend véletlenszerű volt minden hallgató számára.

Az eredmények a 6.3. táblázatban láthatók.

6.3. 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. A haragos érzelem felismerése kevésbé volt sikeres, erre vagy a módszer vagy a választott konstansok nem megfelelőek.

Elvégeztem a későbbiek során egy másik meghallgatásos tesztet is (továbbiakban második teszt), amelyben több szintézistechnológia vizsgálata mellett, a korábban már vizsgált korpuszos szintetizátor és a természetes mondatok újabb változatait is teszteltem. A vizsgálatot két felhasználói csoporton végeztem el. Az egyik csoportba (a továbbiakban I. csoport) azok a felhasználók tartoztak, akik még nem vagy csak alig találkoztak a beszédszintetizátor hangjával, a másik csoportba (a továbbiakban II. csoport) azok, akik a rendszeresen hallgatják ezt a beszédszintetizátort. Azokat az egyéneket, akik ismerték a érzelmi töltetű szintetizált mondatokat, kizártam a tesztelésből. A teszthez a következő mondatokat állítottam elő a korpuszos szintetizátorral:

- „*A réten át a fekete hajú lány közeledett feléje.*”
- „*A gimnázium régi épületét visszaigénylik.*”
- „*Legjobb belátásuk szerint cselekedhetnek.*”

A természetes ejtésű mondatok a következők voltak:

- „*A férfi mintha messzi távolról hallotta volna.*”
- „*Igen kétségtelen, ez a mostani tolvajnyelv monopolizált.*”
- „*Lerakodott a tálcájáról és rögtön enni kezdett.*”

A teszt értékeléses szakaszai előtt a tesztelőnek 2-3 semleges mondat meghallgatásával lehetősége volt megismerkednie a szintetizátor alapértelmezett semleges beszédével. A teszt 4 részből állt, 4 különböző technológiával készített 12-12 szintetizált mondatot kellett meghallgatni részenként. A tesztben a HMM, triádos és korpuszos szintetizátorral készült valamint természetes mondatok szerepeltek.

Az I. csoportban 27 tesztelő szerepelt, mindegyikük egyetemista hallgató, átlagos életkoruk 23 év volt. A tesztet közösen végezték el, a mondatok részenként véletlen sorrendben hangszórón keresztül hangzottak el. A tesztelők papírra rögzítették a válaszaikat, melyben 4 érzelmi megjelölés közül választottak: semleges (N), haragos(A), örömteli (H), szomorú(S).

A II. csoportba 8 fő tartozik, mindegyikük beszédszakértő és ismerik a szintetizátor semleges érzelmet kifejező beszédét. Itt a résztvevők webes tesztet hajtottak végre, ahol a teszt felépítése megegyezett az I. csoportban használttal.

Az I. csoport válaszainak elemezése azt mutatja, hogy a semleges mondatot nem alkalmazták a hallgatók referenciaként az érzelmi kategória kiválasztásánál, hanem inkább az előző mondat alapján döntöttek. Ez a jelenség a beszédszintetizátorral készült mondatoknál jobban megfigyelhető volt, a természetes bemondásoknál kevésbé. Az I. csoport eredményei a 6.4. táblázatban láthatók. A II. csoport, amelyik már ismerte a beszédszintetizátor semleges hangját, sokkal egységesebben döntött. Az eredményeket a 6.5. táblázatban adtam meg.

A beszédszakértőkből álló II. csoport a semleges beszédet 79%-85%-ban jól ismerte fel. Az I. csoport ezzel szemben csak 49%-ban ismerte fel a semleges szintetizátor beszédét. A természetes beszédet a II. csoport már jobb eredménnyel azonosította, amit segíthetett az is, hogy a természetes beszéd a szintetizált beszéd után következett a tesztben és a szintetizátor ugyanazon a beszélő hangján beszélt. Ez a megnövelt adaptációs időszak okozhatja a jobb eredményt. Az első teszthez hasonlóan az öröm és a szomorú érzelmet azonosították jól a tesztelők, a haragot kevésbé.

6.4. táblázat. Az érzelmek tévesztési mátrixa az I. csoportra

		Felismert									
		N	A	H	S		N	A	H	S	
Tervezett	N	49%	4%	37%	10%	N	72%	6%	6%	16%	N=semleges
	A	37%	38%	22%	3%	A	13%	70%	15%	2%	A=haragos
	H	29%	11%	59%	1%	H	28%	34%	36%	2%	H=örömteli
	S	19%	3%	3%	75%	S	9%	0%	4%	87%	S=szomorú
		Korpuszos TTS-női					Természetes női				

6.5. táblázat. Az érzelmek tévesztési mátrixa a II. csoportra

		Felismert									
		N	A	H	S		N	A	H	S	
Tervezett	N	79%	0%	4%	17%	N	85%	5%	5%	5%	N=semleges
	A	18%	14%	68%	0%	A	19%	33%	43%	5%	A=haragos
	H	17%	17%	66%	0%	H	14%	24%	62%	0%	H=örömteli
	S	5%	5%	0%	90%	S	5%	9%	0%	86%	S=szomorú
		Korpuszos TTS-női					Természetes női				

6.6.2. *Érzelmi töltetű beszéd előállítása diád, triád alapú hullámforma összefűzéses rendszerrel*

Az érzelemmódosító eljárás 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 – az alapértelmezett semleges beszéd előállításakor is – és az ismertetett eljárás is végez alapfrekvencia, intenzitás és időtartam korrekciót, ezért az eljárások egyesítése indokolt, mind a minőség, mind a szintetizálás sebessége miatt.

A 6.4. fejezetben ismertetett eljárást két részre bontottam, a spektrális és a prozódiai módosításra. 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 hullámforma-elembázis elemen. A magyar nyelvű beszédszintézis 38 hang alkalmazásával megvalósítható. A diádok száma $38^2 = 1444$. A triádos adatbázisban a triádok száma elméletileg $38^3 = 54872$, de a gyakorlatban ennél jóval kevesebb elemet alkalmaznak (≈ 5000), többnyire csak olyan elemeket, amelyeknél középen egy magánhangzó áll. A diádos elemek 100 ms, a triádos elemek pedig 200 ms körüli hosszúságúak. 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.

A beszédadatbázis tervezési szempontjaiból és a szintetizátor működéséből következően az adatbáziselemek legnagyobb részénél az elemek elején és végén is zöngés rész áll. A 6.5. ábrán is bemutatott okok miatt az első zöngeperiódusnak a módosítása csak félig történik meg, illetve ugyanez a helyzet az elem utolsó zöngeperiódusa esetén is. Az érzelemmódosító algoritmus ezen tulajdonsága azonban itt nem jelent hátrányt, mert nem okoz illesztettlenségi hibát. A diádok és triádok ezeknél a zöngeperiódusoknál illeszkednek egymáshoz a szintetizálás során, így a beszédadatbázisban kialakított pontos kapcsolatokat nem módosítja a transzformáció.

Az érzelemmódosító eljárás másik részét azok a szabályok teszik ki, amelyek a prozódiát módosítják. Ezen szabályokat a prozódia előíró modul paraméterezésével valósítottam meg. A szintetizátor interfészén keresztül lehetőség van az alapfrekvencia és a beszédsebesség módosítására. Ezeket a 6.2. táblázatnak megfelelő értékekre állítottam be.

A módosított szintetizátort a 6.6.1. fejezetben ismertetett meghallgatásos tesztben vizsgáltam (második teszt). A triádos szintetizátorhoz 3 módosított adatbázist készítettem, majd általános témájú mondatokat szintetizáltam. A mondatok megegyeztek a korpuszos szintetizátor mondataival. Az I. csoport esetében a korpuszos és természetes beszédnél jobban jelentkezett

6.6. táblázat. Az érzelmek tévesztési mátrixa triádos szintetizátor esetében I.csoport

		Felismert				
		N	A	H	S	
Tervezett	N	52%	3%	2%	43%	N=semleges
	A	40%	21%	24%	15%	A=haragos
	H	30%	16%	49%	5%	H=örömteli
	S	2%	2%	0%	96%	S=szomorú
Triádos TTS-női						

az a jelenség, hogy a számukra ismeretlen hang semleges állapotát nem tudták a részek előtt elhangzó 2-3 mondat alapján rögzíteni, így értékeléskor nem a referencia mondatokhoz képest döntöttek, hanem az előző elhangzott mondathoz viszonyítottak. A semleges érzelmet csak 57%-ban találták el az összes részben. Az I. csoport tévesztési mátrixa a 6.6. táblázatban látható.

A II. csoport a szintetizátor hangjával rendszeresen találkozott, ezért jól ismerte azt. Az eredmények összefoglalása a 6.7. táblázatban látható.

6.7. táblázat. Az érzelmek tévesztési mátrixa triádos szintetizátor esetében II.csoport

		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						

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, amely arra utal, hogy az adott szintetizátorhang alapparaméterezése eltér a semlegestől.

6.6.3. Érzelmi töltetű beszéd előállítása HMM elvű beszédszintetizátorral

HMM beszédszintetizátor működése

A HMM elvű beszédszintetizátor működése gépi tanuláson alapszik. A beszéd információtartalmát paraméterek formájában kinyerjük, majd az így felhalmozott adatokból HMM modelleket

tanítunk és ebből állítjuk elő a szintetizált beszédet. A módszer két külön fázisra bontható, az első fázis a tanulás, a második maga a szintézis. (Tóth–Németh 2010)

A tanulás során nem közvetlenül a beszéd hullámformáját használjuk fel, hanem az ezekből kinyert spektrális és prozódiai jellemzők sokaságát. A jellemzők kinyerése több órás beszédadatbázisokból történik. A beszédadatbázisokban szükség van a beszéd fonetikus átírására és a beszédhangok határának jelölésére. Ezek mellett a beszéd szerkezeti és nyelvi részleteit jelzésekkel kell ellátni, környezetfüggő címkéket kell elhelyezni. A tanítás során az eredeti beszédadatbázishoz képest nagyságrendekkel kisebb HMM szintézis adatbázist kapunk. A tanítás során kétfajta HMM adatbázist állíthatunk elő. Ha egy beszélő hangjával tanítunk, akkor egy beszélőre jellemző adatbázist kapunk. Ezt felhasználva a beszélő hangkarakteréhez közeli szintetizált beszédet tudunk előállítani. Több beszélőtől származó felvételek alkalmazásával egy általános HMM szintézis adatbázist készíthetünk, amely nem annyira a beszélő személy jellegzetességeit, hanem az adott nyelvre vonatkozó információkat tartalmazza. Ez az általános HMM szintézis-adatbázis használatával később több hangkarakter is előállítható, ún. adaptáció segítségével. Ilyen esetben nem szükséges akkora beszédfelvétel a beszélő függő HMM szintézis-adatbázishoz, mintha csak az adott beszélő felvételeiből készítenénk adatbázist.

A HMM szintézis során a tanítás eredményét, a HMM szintézis adatbázist használjuk fel. A beszéd előállításához szükség van a szöveg fonetikus átírására és az ezekhez kapcsolható környezetfüggő címkékre. A várható hangidőtartamokat kinyerjük az állapot-időtartam valószínűségi sűrűségfüggvényekből, illetve a legvalószínűbb spektrális és gerjesztési paramétereket a HMM adatbázisból. Ezekből valamilyen beszédkódoló eljárással előállítjuk a szintetizált beszédet. A beszédhez kapcsolódó különböző paraméterek közvetlen manipulálására nincs lehetőség mert ezeket az eloszlásfüggvények tartalmazzák, a jellemzők a tanítás előtti adatok változtatásával módosíthatók.

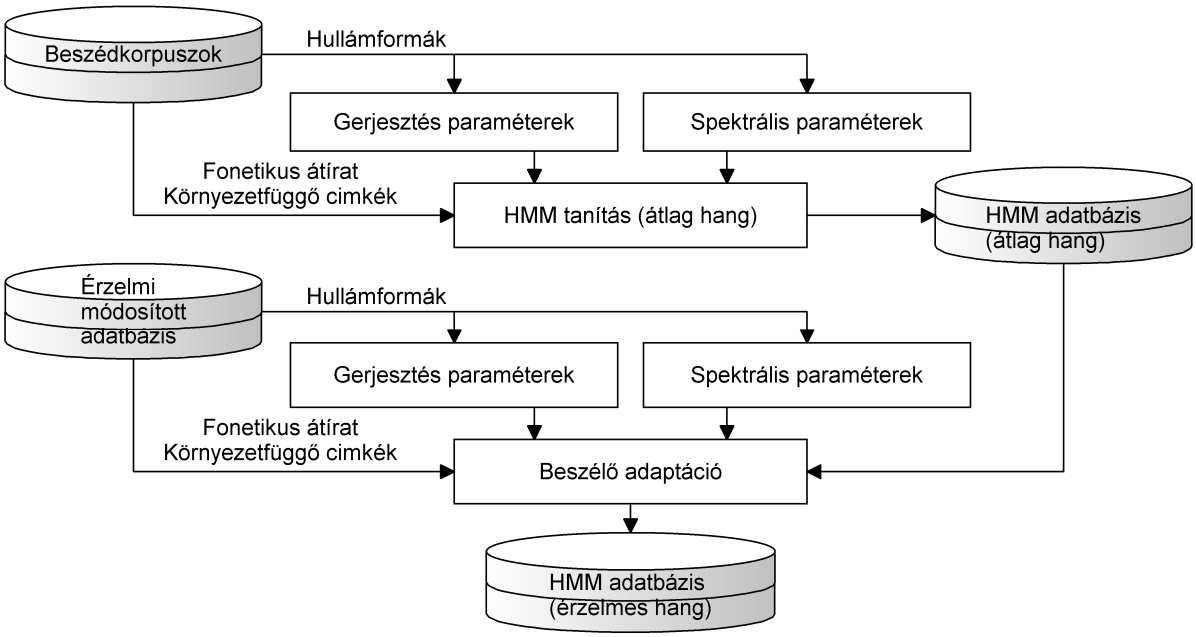
Érzelmi töltetű beszéd előállítása

A HMM szintetizátor kimenetét módosító eljárás működőképes, de az ismételt jelfeldolgozás és feldolgozási idő növekedése miatt – hasonlóan a triádos szintetizátorhoz – nem megfelelő módszer.

A HMM beszédszintetizátor tanító hangadatbázisai több tíz óra méretűek és a jobb minőségű HMM-ek esetében a tanítási fázis is időigényes (több hét nagyságrendű). A nagy erőforrás-igényű teljes tanítás helyett az átlaghang adaptációja segítségével (Tóth–Németh 2009) állítottam elő az érzelmes szintetizálási eredményt.

Az adaptációhoz első lépésben szükség van egy átlag hang előállítására. A tanítás menetét a 6.6. ábrán mutatom be. A nagyméretű beszédkorpuszok több embertől tartalmaznak hangfelvételeket. A hangfelvételek, a felvételekhez tartozó fonetikus átirat és a környezetfüggő címkék segítségével történik a HMM tanítás. A tanítás eredménye egy olyan HMM adatbázis, amely az átlaghang tulajdonságait tartalmazza és alkalmas adaptációra. Ez látható az ábra felső részén. Ennek az átlaghangnak az előállítása időigényes, az ezt követő adaptáció viszont már gyorsan végrehajtható. A beszélő adaptációhoz egy kisméretű – kb. 10 perc időtartamú – prozódiailag változatos természetes hanganyagot módosítottam úgy, hogy az adott érzelmet fejezze ki. A korábban ismertetett módosítás prozódiai és spektrális részét is végrehajtottam. Az adaptációs adatbázisban tehát módosítottam az alap prozódiai paramétereket: az alapfrekvenciát, a hangidőtartamokat és a hangenergiákat. A

spektrális módosítások teljes skáláját szintén az adaptációs adatbázison végrehajtottam: a nemlineáris frekvencia transzformációt és az alacsony és magas frekvenciatartományok energia-eloszlásának módosítását. A semleges érzelem mellett szomorú, örömteli és haragos adaptációs adatbázisokat állítottam elő. A beszélő adaptáció a HMM átlaghangját módosítja a kisméretű érzelmes adatbázis segítségével és állítja elő azt a HMM adatbázist, amellyel az érzelmes beszédszintézis végrehajtható. Ezek a lépések az ábra alsó részén követhetők.



6.6. ábra. HMM adaptáció érzelmes beszédadatbázissal

Az eljárást a triádos szintetizátorral közös meghallgatásos teszttel vizsgáltam. A teszt részletes menetének leírását a 6.6.1. fejezetben ismertettem (második teszt). A HMM szintetizátorhoz egy férfi hangot módosítottam a három érzelem szerint, majd általános témájú mondatokat szintetizáltam:

- „*A réten át a fekete hajú lány közeledett feléje.*”
- „*A gimnázium régi épületét visszaigénylik.*”
- „*Legjobb belátásuk szerint cselekedhetnek.*”

A vizsgálat során az I. csoport nem ért el értékelhető eredményt, csak a II. csoport, amelynek eredményeit a 6.8. táblázatban foglaltam össze.

6.8. táblázat. Az érzelmek tévesztési mátrixa HMM szintetizátor esetében (II. csoport)

		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. Az eljárás a triádos szintetizátorhoz hasonlóan a haragos érzelem kifejezésére nem működik megfelelően.

6.7. Összegzés

Az érzelmi töltetű gépi beszédkeltéshez Přibilová és Přibil módszerét adaptáltam. A spektrális és prozódiai módosításokat korpusz alapú elemkiválasztásos, diád és triád alapú hullámforma-összefűzéses és HMM alapú beszédfelolvasókhoz illesztettem. Az eljárást kísérletekkel teszteltem, amelyek eredménye szerint az örömteli és a szomorú érzelem jól felismerhető, a haragos kevésbé.

7. fejezet
Összefoglalás, tézisek rövid ismertetése

Ékezet nélkül írt szövegek automatikus helyreállítása: Kidolgoztam egy ékezetesítő alapalgoritmust, amely segítségével olyan szövegek is felhasználhatók a gépi beszédkeltés során, amelyek ékezet nélküliek. Az alapalgoritmushoz meghatároztam a szótárépítés folyamatát és a felépített különböző szótárakkal kísérleteket végeztem. Az alapalgoritmust döntési fákkal egészítettem ki, amely csökkentette a hibás ékezetesítések számát. Az elért helyes szavak aránya – a szótárépítés forrásától és a vizsgált szövegek tematikájától függően – 95–97,4%. Az algoritmust és a kísérleteket is magyar nyelvre mutattam be, de az eljárás alkalmazható más ékezetes betűket használó nyelvekre is.

I. téziscsoport: Eljárások ékezet nélkül írt szövegek automatikus helyreállítására

I.1. tézis: *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 (Quinlan 1993), amely a szó környezete alapján egyértelműsíti azt, hogy a több lehetséges ékezetes alak közül melyik a legvalószínűbb.

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].

Magyar és idegen nyelvű írott szövegek vizsgálata a gépi beszédkeltés szempontjaiból: A magyar nyelvre megállapítottam a szóalakok gyakorisági sorrendjét és azt, hogy az adott számú leggyakoribb szó a korpusz mekkora részét fedi le. Ezeket az adatokat grafikusan és formalizálva is leírtam. Összehasonlítottam három nyelvet több különböző méretű szövegkorpusz segítségével. Az elemzéseket magyarra, angolra és németre végeztem el, de a fedési és gyakorisági görbék használatával más nyelvek is hasonlóan összehasonlíthatók. Kidolgoztam magyar nyelvre egy módosított betűstatisztikát, amely figyelembeveszi a gépi beszédkeltés szempontjait.

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

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 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 80 milliós szövegszavas gyűjteményből származnak. Továbbá mindhárom nyelven rendelkezésre állt a Biblia fordítása. Ez utóbbiak vizsgálatát külön végeztem, mivel tartalmi és méreti szempontból egységesek voltak. Megállapítottam, hogy a szavak gyakorisági sorrendben vizsgálva a korpusz mekkora részét fedik le.

Az összehasonlítást elvégeztem azonos tematikájú szövegeken is, a Biblia három változatában. A konkrét gyakorisági számok eltérőek a korábbi nagy vizsgálati korpuszok számaitól, de a nyelvek közötti arány megmaradt. Például a nagy korpuszok és a Biblia korpuszainak 90%-os fedését összevetve, látható, hogy az angol és a német nyelv közötti háromszoros arány megegyezik, és az angol és a magyar közötti 1 nagyságrendi különbség is.

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 hangszintű 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 rendelkezésre áll az új típusú betűstatisztikában a [j] 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.

Beszédszintetizátorok hangminőségének javítása: A név- és címfelolvasás gépi megoldását újfajta módon közelítettem meg, több szintézis módszert kombináltam. A magyar nyelvű név- és címfelolvasáshoz a triádos szövegfelolvasót, a számfelolvasót és 1385 db előre felvett elemet kombináltam. A felvett elemeket a rendelkezésre álló adatbázisból a fedési görbék segítségével úgy határoztam meg, hogy a felvételre kerülő elemek száma a meghatározott korlátok alatt maradjon. A név- és címfelolvasáshoz meghatároztam az elemek sorrendjét, és az elemekhez hozzárendeltem a szintézis módszerét.

Olvasott beszédadatbázisokat felhasználva meghatároztam a magyar nyelv hangjainak intenzitástérképét. Módszert mutattam be a korpusz alapú elemkiválasztásos szintetizátor virtuális intenzitáson alapuló hang szintű intenzitáskiegyenlítésére.

III. téziscsoport: Eljárások beszédszintetizátorok hangminőségének javításához

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 folyamatosan 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.

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 a külön felolvasott elemek.

Az elemzést egy 3 millió rekordot tartalmazó név- és címlistán végeztem, a II.1. tézisben használt módszerrel. Az azonosított 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 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).

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 elemeket rendeztem, majd az elemhez tartozó hullámformák egymáshoz lettek illesztve (triádos szintetizált hullámforma, számfelolvasóval generált számok és külön felolvasott elemek). 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. A 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 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:

$$I_{word}^{virtual} = \frac{\sum_{i=1}^{N} t_{ph}(i) I_{ph}^{average}(ph(i))}{\sum_{i=1}^{N} t_{ph}(i)} \tag{7.1}$$

Ezek alapján meghatározható az adott beszédhang módosításának a mértéke:

$$gain_{ph} = \frac{gain_{prosody} \cdot gain_{base} \cdot I_{word}^{virtual}}{I_{word}^{real}} \tag{7.2}$$

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.

Beszédszintetizátorok hangjának érzelmi módosítása: Az érzelmi töltetű gépi beszédkeltéshez Přibilová és Přibil módszerét adaptáltam. A spektrális és prozódiai módosításokat korpusz alapú elemkiválasztásos, diád és triád alapú hullámforma-összefűzéses és HMM alapú beszédfelolvasókhoz illesztettem. Az eljárást kísérletekkel teszteltem, amelyek eredménye szerint az örömteli és a szomorú érzelem jól felismerhető, a haragos kevésbé.

IV. téziscsoport: Eljárások beszédszintetizátorok hangjának érzelmi módosításhoz

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. 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.

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.

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.

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.

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

7.1. 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.

7.2. Az eredmények értékelése

Az eredményeim az infokommunikációs rendszerek gépi beszédkeltéséhez kapcsolódnak. A téziscsoportok és a megfogalmazott tézisek segítségével – a célkitűzéseimnek megfelelően – a beszédszintézis különböző komponenseit javítottam. Az ékezetesítéssel a beszédszintetizátorok érthetőbben tudják felolvasni a szövegeket. A nyelvek összehasonlítása és a módosított betűstatisztika készítő eljárás segítségével jobb minőségű beszédadatbázisok, ezáltal jobb minőségű szintetizált beszéd állítható elő. A név- és címfelolvasó az általános beszédszintetizátornál jobban érthetőbb felolvasásra képes. A virtuális szó alapú intenzitáskiegyenlítés segítségével a korpuszos szintetizátor több forrásból származó beszédadatbázis esetén is egyenletes minőségű beszédet tud előállítani. Az érzelem módosító eljárással a beszédszintetizátorok felhasználási lehetőségeit bővítettem ki.

Az eredmények egy része már hasznosult, illetve a még nem teljesen hasznosított eredményekhez megadtam a lehetséges felhasználási területeket. Az eredményeket többnyire magyar nyelvre igazoltam, de a módszerek és eljárások kidolgozásánál más nyelvre történő adaptálhatóságot is szem előtt tartottam.

8. fejezet
Köszönetnyilvánítás

Köszönöm témavezetőimnek, Németh Gézának és Olaszy Gábornak, hogy figyelemmel kísérték és segítették munkámat, és a beszédkutatás érdekes területeire kalauzoltak. Biztosították a kutatáshoz szükséges feltételeket és ösztönözték az elért eredmények publikálását. A tudományos munka mellett a Beszédtechnológiai Laboratórium csapatát is összekovácsolták, így egy közösség tagjaként egymást ösztönző légkörben végezhettem a napi teendőket.

Köszönöm a labor volt és jelenlegi tagjainak: Bartalis Mátyásnak, Béres Andrásnak, Bőhm Tamásnak, Csapó Tamásnak, Fék Márknak, Kiss Gézának, Kosztyu Lászlónak, Laczkó Klárának, Olaszi Péternek, Tóth Bálintnak, Viktóriusz Ákosnak az együttműködést, a közös cikkeket és a közös élményeket.

Köszönöm Gordos Gézának és Sallai Gyulának a témámhoz és doktori munkámhoz nyújtott bölcs tanácsokat és útmutatásokat.

Köszönöm Fegyó Tibornak, Mihajlik Péternek és Tarján Balázsnak az együttműködést, melyben támogatás adtak, hogy a beszédfelismerést is alkalmazni tudjam kutatásaim során.

Köszönöm Prószéky Gábornak, hogy rendelkezésemre bocsájtotta a MorphoLogic Kft. helyesírás ellenőrző modulját.

Köszönöm Váradi Tamásnak, hogy a kutatásaimhoz a Magyar Nemzeti Szövegtár anyagát felhasználhattam.

Köszönöm Vicsi Klárának a kutatásaimhoz rendelkezésre bocsájtott anyagokat.

Köszönöm Farkas Richárdnak és Takács Györgynek a disszertációmhoz fűzött hasznos és értékes visszajelzéseit.

És végül köszönöm feleségem Vera és gyerekeim Noémi és Gergő odaadó segítségét és türelmét, amellyel támogatták munkámat.

A kutatásaimat többek között a következő programok is támogatták: Természetes beszédinformációs rendszerek: NKFP-2/034/2004; Beszélő mobiltelefon: GVOP-3.1.1-2004-05-0485/3.0; Ambiens intelligenciára épülő ipari alkalmazások kutatás-fejlesztése – BelAmi: ALAP2-00004/2005; Gyógyszervonal: GVOP-3.1.1 - 2004 - 05 - 0426 /3.0 ; Teleauto: OM-00102/2007; Ember és informatikai rendszerek kapcsolatának új, etológiai modell alapú generációja: TÁMOP-4.2.2-08/1/KMR-2008-0007

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.

Baggia, P. – Badino, L. – Bonardo, D. – Massimino, P. (2006): Achieving perfect TTS intelligibility. In *Originally presented at the AVIOS Technology Symposium, SpeechTEK West*.

Banse, R. – Scherer, K. (1996): Acoustic profiles in vocal emotion expression. 70. évf. *Journal of personality and social psychology*, 614–636.

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. (1993): Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound. In *Proceedings of the Institute of Phonetic Sciences*. vol. 17. 97–110.

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

Campbell, N. (2001): Building a Corpus of Natural Speech-and Tools for the Processing of Expressive Speech-the JST CREST ESP Project. In *Proceedings of the 7th European Conference on Speech Communication and Technology*. 1525–1528.

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.

Gósy M. (2004): *Fonetika, a beszéd tudománya*. Osiris, 85–90.

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.

Hamza, W. – Eide, E. – Bakis, R. – Picheny, M. – Pitrelli, J. (2004): The IBM expressive speech synthesis system. In *INTERSPEECH-2004*. Citeseer, 2561–2564.

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

Kawahara, H. – Cheveigné, A. – Banno, H. – Takahashi, T. – Irino, T. (2005): Nearly defect-free F0 trajectory extraction for expressive speech modifications based on STRAIGHT. In *Ninth European Conference on Speech Communication and Technology*. ISCA, 537–541.

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.

Laukka, P. (2004): *Vocal Expression of Emotion : Discrete-emotions and Dimensional Accounts*. Phd értekezés (Uppsala University, Department of Psychology).

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`.

Mihajlik P. – Révész T. – Tatai P. (2002): Phonetic Transcription in Automatic Speech Recognition. *Acta Linguistica Hungarica*, vol.49, no.3–4, 407–425.

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. (2010): A magyar magánhangzók. In Németh, G. – Olaszy, G. (szerk.) *A MAGYAR BESZÉD; Beszédkutatás, beszédtechnológia, beszédinformációs rendszerek*. Akadémiai, 106–113.

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. (2006): Hangidőtartamok és időszerkezeti elemek a magyar beszédben. In *Nyelvtudományi Értekezések*. Akadémiai Kiadó.

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

Petrushin, V. (2000): Emotion Recognition in Speech Signal: Experimental Study, Development, and Application. In *Sixth International Conference on Spoken Language Processing*, 2. köt. ISCA, 222–225.

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. (2006): Non-linear frequency scale mapping for voice conversion in text-to-speech system with cepstral description. 48. évf. 12. sz., *Speech Communication*, 1691–1703.

Přibilová, A. – Přibil, J. (2009b): Harmonic Model for Female Voice Emotional Synthesis. In Fierrez, J. – Ortega-Garcia, J. – Esposito, A. – Drygajlo, A. – Faundez-Zanuy, M. (szerk.) *Biometric ID Management and Multimodal Communication*. Lecture Notes in Computer Science sorozat, vol. 5707. Springer Berlin / Heidelberg, 41–48.

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.

Szende T. (1976): *A beszédfolyamat alaptényezői*. Budapest, Akadémiai Kiadó.

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.

Tóth B. – Németh G. (2010): A rejtett Markov-modellen alapuló gépu szövegfelolvasás. In Németh G. – Olaszy G. (szerk.) *A MAGYAR BESZÉD; Beszédkutatás, beszédtechnológia, beszédinformációs rendszerek*. Akadémiai, 512–518.

Tóth S. L. – Sztahó D. – Vicsi K. (2007b): Speech Emotion Perception by Human and Machine. In *Proc. COST Action 2102 International Conference*. Patras, Greece, 29-31 October., Springer, 213–224.

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 Informatikai 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.

A. függelék
Újraékezetesítés

Példa az ékezetesítő szótárra (részletek)

abrazolasuak ábrázolásúak
abrazolasukhoz ábrázolásukhoz
abrazolat ábrázolat
abrazolata ábrázolata
abrazolataban ábrázolatában
abrazolatai ábrázolatai
abrazolatja ábrázolatja
abrazolhat ábrázolhat
abrazolhatatlan ábrázolhatatlan
abrazolhatatlanna ábrázolhatatlanná
abrazolhatjak ábrázolhatják
abrazolhatna ábrázolhatna
abrazolhatnak ábrázolhatnak
abrazolhato ábrázolható

adoalanykent adóalanyként
adoalanynak adóalanynak
adoalanyok adóalanyok
adoalanyokat adóalanyokat
adoalanyokhoz adóalanyokhoz
adoalanyokka adóalanyokká
adoalanyokkal adóalanyokkal
adoalanyokkent adóalanyokként
adoalanyoknak adóalanyoknak
adoalanyokra adóalanyokra
adoalanyokrol adóalanyokról
adoalanyoktol adóalanyoktól
adoalanyonkent adóalanyonként
adoalanyra adóalanyra
adoalanyrol adóalanyról
adoalanyt adóalanyt
adoalanytol adóalanytól

folajanlasarol fölajánlásáról
folajanlasaval fölajánlásával
folajanlasok fölajánlások
folajanld fölajánld
folajanlja fölajánlja
folajanljuk fölajánljuk
folajanlom fölajánlom
folajanlott fölajánlott
folajanlotta fölajánlotta
folajanlottak fölajánlották
folajanlottam fölajánlottam
folajanltam fölajánltam
folajanlunk fölajánlunk
folajzasaban fölajzásában
folajzott fölajzott
folajzva fölajzva

hazassagokat házasságokat
hazassagokban házasságokban
hazassagokert házasságokért
hazassagokhoz házasságokhoz
hazassagom házasságom
hazassagomat házasságomat
hazassagomba házasságomba
hazassagomig házasságomig
hazassagomnak házasságomnak
hazassagomot házasságomot
hazassagon házasságon
hazassagos házasságos
hazassagot házasságot
hazassagpartiak házasságpártiak
hazassagra házasságra
hazassagrol házasságról
hazassagszedelgo házasságszédelgő

hazaszeretetenek hazaszeretetének
hazaszeretetrol hazaszeretetről
hazaszeretetuk hazaszeretetük
hazaszereto hazaszerető
hazaszeretok hazaszeretők
hazaszerzodne hazaszerződne
hazaszokott hazaszökött
hazaszoktem hazaszöktem
hazaszokunk hazaszökünk
hazaszol hazaszól
hazaszolitja hazaszólítja
hazaszolitottak hazaszólítottak
hazaszoljak hazaszóljak
hazat házat
hazatajan házatáján
hazatajarol házatájáról
hazatakarodasa hazatakarodása
hazatalal hazatalál
hazatalalas hazatalálás

lekotozes lekötözés
lekotozne lékötözne
lekotozott lekötözött
lekotozte lekötözte
lekotoztek lekötözték
lekotozve lekötözve
lekotrodjek lekotródjék
lekottazasanak lekottázásának
lekottazhatatlan lekottázhatatlan
lekottazott lekottázott
lekotve lekötve
lekotven lekötvén

```
lekovetem lekövetem
lekozlendo leközlendő
lekozli leközli
lekozlik leközlik

pasztorain pásztorain
pasztorainak pásztorainak
pasztorainknal pásztorainknál
pasztorainkrol pásztorainkról
pasztorakent pásztoraként
pasztoranak pásztorának
pasztorara pásztorára
pasztorat pásztorát
pasztorbot pásztorbot
pasztorbundanal pásztorbundánál
pasztordal pásztordal
pasztoreleme pásztoreleme
pasztorember pásztorember

vervizsgalatot vérvizsgálatot
vervolgyi vérvölgyi
vervonal vérvonal
vervonala vérvonala
vervonalakat vérvonalakat
vervonalara vérvonalára
vervonalat vérvonalat
vervonalbol vérvonalból
vervonalhoz vérvonalhoz
vervonalu vérvonalú
vervoros vérvörös
```

Leggyakoribb kétes szavak

gyak	hibas	szó1	gyak1	szó2	gyak2	szó3	gyak3	szó4	gyak4	szó5	gyak5
1338448	541760	meg	796688	még	541760						
257438	58741	fel	198697	fél	58741						
82149	40358	őt	38001	öt	41791	ót	36				
204413	39239	ő	165174	ó	9794						
125281	34436	hat	34436	hát	90845						
100274	33556	akar	33553	akár	66718						
336723	19505	úgy	317218	ügy	17132						
49057	18449	no	18195	nő	30608						
37206	16333	területen	16311	területén	20873	térületén	3				
52095	15324	szinten	15316	szintén	36771						
63592	13359	fele	11239	felé	50233	féle	2099	félé	21		
59890	13231	lévő	46659	levő	12379						
59742	13149	ok	12753	ők	46593	ók	17				
29368	10143	lehetőségét	9610	lehetőséget	19225						
363442	9897	mert	353545	mért	9897						
70602	9877	hozza	9877	hozzá	60725						
29797	9828	figyelmet	19969	figyelmét	9828						
20967	9776	vettek	9775	vették	11191						
26301	9556	nevet	9550	nevét	16745	névét	3				
20440	9493	szeretne	9493	szeretné	10947						
21014	9368	tudtak	9368	tudták	11646						
18537	9254	tettek	9254	tették	9283						
22987	9131	eletet	57	életét	13856	életet	9069				
17192	8874	kor	8318	kör	7943	kór	914				
26976	8726	helyet	18250	helyét	8726						
37816	8646	helyen	29170	helyén	8646						
53499	8606	mely	44893	mély	8606						
22938	8594	ülésen	8557	ülésén	14344						
21498	8163	téren	13335	terén	8140	térén	2				
25913	7905	szerepet	18008	szerepét	7905						
16534	7779	értek	7720	érték	8755						
15675	7766	tartottak	7909	tartották	7766						
39028	7513	köze	7436	közé	31515						
14956	7381	tevékenységet	7376	tevékenységét	7575						

gyak	hibas	szó1	gyak1	szó2	gyak2	szó3	gyak3	szó4	gyak4	szó5	gyak5
19576	7260	helyzetet	12316	helyzetét	7260						
17389	7154	adtak	10235	adták	7153						
13744	6458	hova	6454	hová	7286	hóvá	3				
16425	5837	találtak	10588	találták	5690	tálalták	77	tálaltak	51		
11636	5737	hoztak	5736	hozták	5899	hóztak	1				
22202	5651	ugye	16551	ügye	5481	ügyé	5				
11778	5527	melyen	3077	mélyén	2449	mélyen	6251	melyén	1		
15321	5396	szoba	5393	szóba	9925						
25043	5281	haza	19762	háza	5252						
17878	5244	kaptak	12634	kapták	5244						
11263	5237	kerek	4176	kerék	1059	kérek	6026				
10838	5132	akartak	5130	akarták	5706						
15861	5092	arany	10769	arány	5092						
28880	5035	ön	23845	ón	99						
17061	4795	kar	4795	kár	12266						
165188	4598	e	160590	é	4598						
11271	4561	tó	6710	tő	507						
10252	4450	kezdtek	5802	kezdték	4450						
10927	4385	teli	4385	téli	6542						
13442	4299	erő	9143	érő	3971						
8116	4192	kertek	896	kérték	3924	kértek	3296				
8884	4095	tudna	4095	tudná	4789						
8352	4087	teve	671	téve	3359	tévé	4265				
12601	4031	felek	8570	félek	3969	félék	62				
11930	4019	ügyét	4008	ügyet	7911						
10108	3996	írók	6112	írok	3561						
7973	3893	értéket	3885	értékét	4080	értekét	1				
14843	3850	kezet	3847	kezét	10993	kézét	3				
7398	3849	köré	3549	kőre	322	köre	3433	koré	16	kóré	66
7552	3794	írtak	3758	írták	3679						
7628	3761	szoktak	3867	szokták	3759						
8446	3696	arat	481	árát	4750	árat	3190	arát	25		
9176	3632	vad	5544	vád	3632						
7627	3562	elek	4065	élek	3558						
14715	3490	összeget	11225	összegét	3489	ősszegét	1				
18561	3435	nemi	3435	némi	15126						
9602	3292	élén	6310	élen	3265						
9305	3282	címet	6023	címét	3166						
15408	3253	véget	12155	végét	3235						
7793	3193	veres	2803	véres	4600	verés	387				
16545	3162	kert	3162	kért	13383						
10645	3122	környékén	7523	környéken	3110						
6943	3113	sort	3830	sört	3111						
7294	3097	történetét	4197	történetet	3089						
8486	3079	látták	5407	láttak	3072						
7455	3000	címen	4455	címén	2955						
19675	2977	velem	16698	vélem	2976						
39139	2957	sor	36182	sör	2945						
9411	2940	ahova	2940	ahová	6471						
118581	2888	való	115693	váló	2581						
7355	2873	tervet	4482	tervét	2873						
11538	2869	roma	8669	róma	2847						
6362	2820	elve	2807	élve	3542	elvé	13				
7471	2801	szöveget	4670	szövegét	2779	szóvéget	7				

gyak	hibas	szó1	gyak1	szó2	gyak2	szó3	gyak3	szó4	gyak4	szó5	gyak5
7432	2781	vittek	2781	vitték	4651						
488722	2773	vagy	485949	vágy	2773						
5519	2736	adok	2730	adók	2783						
29971	2728	bank	27243	bánk	2728						
5580	2685	többségét	2679	többséget	2895						
6300	2675	elénk	2670	élénk	3625						
6472	2669	állítottak	2635	állították	3803						
16561	2657	mondjak	2657	mondják	13904						
603963	2652	mar	2652	már	601311						
59925	2630	sőt	57295	sót	1046						
10756	2585	könyvet	8171	könyvét	2544						
11567	2549	nevén	2507	néven	9018	névén	1				

B. függelék
Név- és címfelolvasás

Névgyakoriságok

B.1. táblázat. Leggyakoribb magyar vezetéknevek

Vezetéknév	Abszolút gy.	Relatív gy.	Vezetéknév	Abszolút gy.	Relatív gy.
Nagy	66311	2,40%	Simon	10934	0,40%
Tóth	60938	2,21%	Szűcs	10426	0,38%
Szabó	60787	2,20%	Fekete	9882	0,36%
Kovács	60646	2,20%	Szilágyi	8370	0,30%
Horváth	49247	1,79%	Török	7793	0,28%
Kiss	39370	1,43%	Rácz	7466	0,27%
Varga	37282	1,35%	Oláh	7452	0,27%
Molnár	30388	1,10%	Fehér	7096	0,26%
Németh	27475	1,00%	Gál	7082	0,26%
Farkas	19578	0,71%	Pintér	6987	0,25%
Balogh	17474	0,63%	Kocsis	6648	0,24%
Papp	16046	0,58%	Balázs	6630	0,24%
Takács	15103	0,55%	Fodor	6428	0,23%
Juhász	14267	0,52%	Hegedűs	6272	0,23%
Mészáros	11506	0,42%	Magyar	5926	0,21%

B.2. táblázat. Leggyakoribb magyar keresztnevek

Keresztnév	Abszolút gy.	Relatív gy.	Keresztnév	Abszolút gy.	Relatív gy.
József	219305	8,19	Béla	50937	1,9
István	218368	8,16	Gábor	49309	1,84
László	200853	7,5	Péter	38529	1,44
János	183358	6,85	Pál	37164	1,39
Ferenc	130244	4,87	Miklós	36111	1,35
Sándor	126285	4,72	Attila	35587	1,33
Lajos	79312	2,96	Antal	29165	1,09
Zoltán	76528	2,86	Mária	26325	0,98
Imre	74166	2,77	Tamás	25850	0,97
György	62373	2,33	Csaba	24174	0,9
Gyula	58836	2,2	Éva	22059	0,82
Tibor	55403	2,07	Katalin	21472	0,8
Mihály	52426	1,96	Géza	20897	0,78
András	52187	1,95	Zsolt	20847	0,78
Károly	52092	1,95	Erzsébet	18065	0,67

Szünetszabályok

- Amennyiben a cégformajelölés valós cégformát jelöl (az elem szótárban megtalálta a rendszer) akkor a cégforma elé egy szünet jelölést helyez el a rendszer.
- Csupa nagybetűs szavak után szünet jelölést rak.
- Írásjelek egyéb karakterek esetében, ha azok szerepelnek a karakterek szótárban akkor szünet jelölést rak utánuk.
- A következő karakterek nem szerepelnek a szótárban, de szünet jelölés lesz a helyükön: „/” perjel, „(” zárójel, „)” zárójel , „„” vessző, „-” kötőjel.
- Minden kimondás legelejére előír a rendszer 400 ms szünetet.
- Minden elem (szó) közé 100 ms szünetet ír elő.
- Amennyiben már előzőleg szünet jelölés lett elhelyezve, akkor azt helyettesíti egy 200 ms-os szünettel.
- Ha szám szerepel a bemondásban, akkor 200 ms hosszú szünetet rak utána.
- Ha egy szó hossza kisebb, vagy egyenlő 2 karakterrel, akkor 200 ms hosszú szünetet rak utána.
- Ha egy szó szerepel a „nincs szünet” szótárban, akkor nem rak szünetet az adott szó elé.
- Ha egy szó szerepel a keresztnév szótárban, akkor az elé 200 ms hosszú szünetet rak.
- A bemondás végére 400 ms szünetet ír elő.
- Az irányítószám után 700 ms szünet van.
- A városnév után 600 ms szünet van.
- Ha a városnévben nem betű és nem szám típusú karakter szerepel, akkor 400 ms szünettel helyettesíti azt.
- A közterület után 100 ms szünet van előírva.
- Az emelet után 100 ms szünetet ír elő.
- A lépcsőház után 200 ms szünetet ír elő.
- A per . . . után 200 ms szünetet ír elő.

Közterülettípus elemszótára

utca	park	árok
út	kapu	határút
tér	körtér	üzletközpont
körút	liget	lépcső
útja	puszta	sgt
ltp	kert	stny
köz	sziget	u
sor	udvar	ut
tanya	major	külterület
sétány	rakpart	rét
sugárút	lejtő	útfél
dűlő	hegy	krt
fasor	utcája	ln
tere	part	lkn

lakónegyed
lnd
szer
alszer
alsószer
baksaszer
belsőszer
besőszer
csárdaszer
csörgőszer
felsőszer
gyöpszer
keletiszer
keserűszer
kovács-szer
kovácsszer
nemesszer
papszer
pomperszer
siskaszer
templomszer
tószer
városszer
zsohárszer
keserűszer
égésszer
híd
hrsz

Cégforma elemszótára

Club
rt
kft
Kft
KFT
RT
gmk
kkt
kht
Kft
Kft
Bt
BT
Kkt
Kht
KHT
Kht
Rt
RT
Alapítvány
háza
Háza
Lelkisegélyszolgálat
Szerkesztősége
ABC
Taxi
Ételbár
Panzió
Pizzéria
Körömstúdió
Tömörülése
Kisebbségekért
Reklámügynökség
Magánklinikája
Könyvkötészet
Főiskola
állomás
Boltja
Képviselet
Rendelő
Panzio
Iroda
Üzem
Kertszövetkezet
Bútorszalon
Lakásotthonok
Szövetség
Intercom
Kávéház
Kisáruház
Társaság
Gmk
Biztosító
Egyesület
Gyeremekdivat
Galéria
Pékség
Szövetkezet
Szolárium
Reklámstúdió
Csempediszkont
Borozó
Stúdió
Bútorház
Elektronika
Szalon
Bolt
Húsbolt
Élelmiszerbolt
Vegyesbolt
Autósbolt
Kereskedő
Computer
Étterem
Lambéria
szolgálat
Baromfiboltja
Vendéglő
Cukrászda
Kertészet
Árufuvarozó
Fagyizó
Fodrászszalon
Virágüzlet
Zöldségház
Márkabolt
Kör
Rövidárú
üzlet
Alkatrészbolt
Divat
Kozmetika
Market

Rövidárubolt
Büfé
Porcelánbolt
Shop
Cég
Magángyógyszertár
Camping
Fitness
Sportcentrum
Szolgálat
Hivatal
Önkormányzat
Tankönyvcentrum
Vállalkozás
Praxis
Nevelőotthon
Adatbank
Szolgáltató
Üzlete
Sport
sportbolt
Csárda
Adófelügyelőség
férfidivat
Ruhák
Adótanácsadás
Trade
Butik
Ékszerbolt
Pressó
Szépségközpont
Iskolája
Studio
Kiadó
Egyházterület
Egyházelnök
Gondnok
Elnökség
gyülekezet
Egyház
Szeretetotthon
Alapkezelő
Bisztró
Áruház
Iskola
Könyvesbolt
Ház

Sajtóalapítvány
Könyvtár
Igazgató
Aerocaritas
Vezérképviselet
Gyógyszertár
bolt
étterem
Áfész
ÁFÉSZ
Szépségszalon
Center
Múzeum
Drogéria
Gyülekezet
Kocsma
Szakbizottsága
Plusz
Fogadó
Barlang
kiállítás
Park
Fodrászat
Depó
BÜFÉ
Lakberendezés
Tours
Eszpresszó
Intézet
Központ
Vállalat
Obszervatórium
Kivitelező
Agroporta
Kirendeltség
Kereskedés
kft
Üzlet
GmbH
Planet
Nagykereskedelem
Tár
ÁFÉMSZ
Agrárcenter
Agrárkamara
Agrárszövetkezet
Agrárszövetség

Skála
Műszaki
Agrobutik
Agrodiszkont
Egyesülés
Kereskedelmi
Planet
Szervezetek
Szerkesztőség
Hotel
Presszó
Ambulancia
söröző
Titkárság
Nyomda
Hangstúdió
Akkumulátor
Akkuvill
készítés
Mérnökiroda
Szőnyegáruház
Espresso
Filmstúdió
Főmérnökség
Tangazdaság
Tábor
Vizsgaközpont
Vagyonvédelem
Hungary
Diszkont
Klub
Táncegyesület
Albérletcentrum
Strandfürdő
Autósiskola
Kontakt
váltók
Solárium
Fordítóiroda
Kölcsönző
Nyelvstúdió
Rádió
Szálloda
Vállalkozó
Ltd
Autószalon
Spa

Ruhakölcsönző
Stop
csoportja
Gondozóház
Alkotóház
titkárság
osztály
Állami
Bölcsöde
Állategészségügy
Állateledel
Állatgyógyszertár
Begyűjtő
Állatkereskedés
Állatkert
Állatklinika
Állatkórház
Állatmenhely
Állatorvos
Állatpatika
Kutatóintézet
Állomás
Hungaria
Könyvkereskedés
Vegyesélelmiszer
Almássytex
Kastélyhotel
Söröző
virágszalon
Vadásztársaság
Ingatlanközvetítő
Sörbár
Alpintechnika
Vásáriroda
Ellátó
Hivata
Intézmények
Takarékszövetkezet
Részönkormányzat
Kultúrház
Ügynökség
Zeneiskola
iskola
tagozat
iroda
iskolák
School
Formaöntészet
Corporation
Patika
Szervíz
Irodája
Kamara
Accord
Agrober
Bár
International
Gmbh
Szerzetesrend
Műhely
Kongregációja
Antikvárium
Csecsemővédelem
ege
et
Ege
EGE
ET
Et

Rövidítések a címekben

u utca
ut út
sgt sugárút
stny sétány
pf postafiók
hrsz helyrajziszám
mfszt magasföldszint
mf magasföldszint
mfsz magasföldszint
Fszt földszint
fszt földszint
Fsz földszint
fsz földszint
ép épület
em emelet
sz szám
lph lépcsőház
lép lépcsőház
lépcsh lépcsőház
lh lépcsőház
krt körút
ltp lakótelep
ln lakónegyed
lkn lakónegyed
lnd lakónegyed

Az érthetőségi tesztben használt címek

B.3. táblázat. Névfelolvasó érthetőségi tesztjének részletes eredményei

Cím	Névfelolvasó			Profivox			Helyesek
	Helyes	Kicsit	Hibás	Helyes	Kicsit	Hibás	különbsége
7741, Pilis Hont F. u. 14/B. 2/9.	1.00	0.38	0.00	0.00	0.00	1.00	1.00
5661, Cegléd Vaszary Kolos utca 26/b	1.00	0.00	0.00	0.13	0.00	0.88	0.88
6050, Zalaszántó Battai út 2. 1. em 1a	1.00	0.38	0.00	0.43	0.29	0.57	0.57
7741, Nemesszalók Soltész Nagy Kálmán u.13.Fszt./10.	1.00	0.29	0.00	0.50	0.38	0.50	0.50
1054, Tamási Bástya Ltp. 14. 4/12.	1.00	0.00	0.00	0.67	0.33	0.17	0.33
4244, Békéscsaba József u. 148/a	1.00	0.00	0.00	0.67	0.22	0.33	0.33
7300, Debrecen Halastó szél 3.	1.00	0.13	0.00	0.71	0.14	0.29	0.29
2085, Dány Babóchay u. 14.	1.00	0.14	0.00	0.75	0.00	0.25	0.25
1191, Szob Schönhercz Z. u. 5.	1.00	0.09	0.00	0.75	0.00	0.25	0.25
8561, Tarján Posztó u.2. IV/4.	1.00	0.17	0.00	0.78	0.22	0.22	0.22
7675, Hajdúszoboszló Tóvárosi lnd.49.fsz/3.	0.86	0.29	0.14	0.75	0.13	0.25	0.11
6455, Szentes Napvirág u. 19. fszt. 2.	1.00	0.20	0.00	0.90	0.10	0.10	0.10
9144, Sátoraljaújhely Wesselényi Miklós utca 101.	1.00	0.40	0.00	0.90	0.00	0.10	0.10
3246, Békésszentandrás Pozsonyi utca 56.	1.00	0.00	0.00	1.00	0.10	0.00	0.00
4064, Csanytelek Hegymászó u. 3.	1.00	0.30	0.10	1.00	0.00	0.00	0.00
5661, Cegléd Böszörményi út 68. II. 210.	1.00	0.29	0.00	1.00	0.00	0.00	0.00
7960, Komló Madách liget 11. 4/3.	1.00	0.13	0.00	1.00	0.00	0.00	0.00
1192, Tab Vörösmarty u. 13.	1.00	0.00	0.00	1.00	0.17	0.00	0.00
8100, Tiszaújváros SZÉCHÉNYI I.TÉR 1. 2/3.	1.00	0.00	0.00	1.00	0.17	0.00	0.00
4075, Csemo Almássy telep 6.	0.88	0.00	0.13	1.00	0.00	0.00	-0.13