



*Budapesti Műszaki és Gazdaságtudományi Egyetem
Távközlési és Telematikai Tanszék
Beszédtechnológiai Laboratórium*

Magyar nyelvű szöveg–beszéd átalakítás: nyelvi modellek, algoritmusok és megvalósításuk

Ph.D. értekezés tézisei

Írta:
Olaszi Péter

Tudományos vezetők:

Dr. Németh Géza
*Budapesti Műszaki és Gazdaságtudományi Egyetem
Távközlési és Telematikai Tanszék
Beszédtechnológiai Laboratórium*

Dr. Olaszy Gábor
*Magyar Tudományos Akadémia
Nyelvtudományi Intézet
Kempelen Farkas Beszédkutató Laboratórium*

2002.

1 Bevezetés

Magyarországon az elektronikus beszédelőállítás mintegy két évtizedes múltat tekint vissza. Az első elektronikus beszédkeltő berendezést a Magyar Tudományos Akadémia Nyelvtudományi Intézete Fonetikai Laboratóriumában (MTA NyI) hozták létre Olasz Gábor vezetésével 1980–82 között. A Hungarovox nevű rendszert DEC PDP–11/34 számítógépen OVE III formánsszintetizátor alkalmazásával valósították meg, elsősorban nyelvészeti és fonetikai alapokra támaszkodó kutatások céljára. A formánsszintézis a gerjesztés és a beszédjelet jellemző formánsszintézis vizsgálatán és mesterséges létrehozásán alapul.

A Nyelvtudományi Intézetben folyó munkával párhuzamosan a Budapesti Műszaki Egyetem Híradástechnikai Elektronikai Intézetében (BME HEI) Gordos Géza vezetésével távközléstechnikai motivációból is zajlottak beszédkeltéssel kapcsolatos kutatások. A két intézmény közötti együttműködés keretében készült el 1983-ban a Scriptovox beszédkeltő rendszer, amely a MEA 8000 formánsszintetizátoron alapult.

Az MTA NyI és a BME HEI közötti kapcsolat ezt követően folyamatosabbá vált. A személyi számítógépek teljesítményének növekedése lehetővé tette a valós idejű, kötetlen szótárak szöveg–beszéd átalakítás formánsszintézissel történő megvalósítását, így 1986-ban megkezdődött a Multivox többnyelvű beszéd-szintézis-rendszer fejlesztése Olasz Gábor és Németh Géza vezetésével. A program első változatai DOS operációs rendszeren működtek, és a beszédjelet a számítógéphez illesztett Philips PCF 8200 formánsszintetizátor chip segítségével hozták létre. A későbbi változatokban a hardver emuláció lehetővé tette a beszéd-szintetizátor tisztán szoftveres realizációját. Később elkészült a Windows operációs rendszeren működő változat is. A magyar nyelv formánsszintézisével kapcsolatos tapasztalatokat átfogó formában Olasz Gábor összefoglaló munkájában találjuk meg.

A 90-es évek közepétől kezdve mind külföldön, mind itthon a hullámforma-összefüzeses technológia került a kutatások homlokterébe. A személyi számítógépek folyamatos számítás- és tárolókapacitás-növekedésének köszönhetően ekkorra reális céllá vált az emberi beszédből kivágott hullámformaelemek összefüzésén alapuló valós idejű beszéd-szintézis. E technológia alapváltozatában két-két beszédhang kapcsolatát, úgynevezett diádelemeket tárolunk a hangadatbázisban. Egy diádelem egy beszédhang második felét és a következő beszédhang első felét tartalmazza. Az alkalmasan megválasztott vágási pontokkal biztosítható, hogy a diádokból összefüztetett beszédjel természetes hangzású

legyen. A BME Távközlési és Telematikai Tanszékén – amely a HEI utódjának tekinthető – 1998-ban készült el a Profivox szöveg–beszéd átalakító rendszer alapváltozata. Ennek jellemzője, hogy a beszédjelet a mintegy 1100 diádelemet tartalmazó hangadatbázisból építi fel.

Szintén a 90-es évek közepére tehető a beszédelőállítással kapcsolatos prozódiai és nyelvészeti kutatások jelentős nemzetközi és magyarországi térnyerése. A beszéd-szintézis-rendszerben prozódia alatt a hang magasságának, a hangsúlyozásnak, a beszéd ritmusának, időszerkezetének, a szüneteknek és az intenzitásnak a változásait értjük. A magyar nyelv prozódiájának leíró jellegű, elméleti vizsgálatát elsősorban Bródy Mihály, É. Kiss Katalin, Fónagy Iván, Hunyadi László, Kálmán László és Varga László munkáiban találjuk meg. A beszéd-megértés prozódiai szempontú vizsgálatával kapcsolatban Gósy Mária, Pléh Csaba és Vicsi Klára munkáit emelném ki. A mesterséges beszéd-keltés prozódiai vonatkozásait Olasz Gábor és Koutny Ilona munkái összegzik. Az írott szöveg számítógépes feldolgozásával kapcsolatban Prószéky Gábornak a számítógépes nyelvészet témakörében végzett munkásságát és gyakorlati eredményeit kell megemlítenünk.

A mesterséges beszéd-keltés magyarországi vonatkozásaival kapcsolatban említést érdemelnek az MTA Központi Fizikai Kutatóintézetében a 80-as évek közepétől kezdődően Arató András vezetésével folytatott, formánsszintézisen alapuló vakügyi beszédtechnológiai fejlesztések, a Király József által 1988-ban elkészített, DOS operációs rendszeren működő, tisztán szoftveres, de intonációt nem tartalmazó szöveg–beszéd átalakító program, illetve az utóbbi években a magyar Mindmaker cég kisebb részben kutatási, elsősorban profit-központú szemlélettel létrehozott szöveg–beszéd átalakító rendszere.

Én 1992-től kapcsolódtam be a BME Távközlési és Telematikai Tanszéke Beszédtechnológiai Laboratóriumának munkájába. Kezdetben a formánsszintézisen alapuló rendszer megismerését tűztem ki célul, ezt követően az alkalmazások megvalósításával foglalkoztam. Az angol nyelvű betű–beszédhang átalakítás problémájával összefüggésben ekkor kerültem kapcsolatba az MTA NyI részéről Siptár Péterrel, akitől a későbbiekben is számos iránymutatást kaptam fonológiai tárgyú kérdésekben. Érdeklődésem már ebben az időben a nyelvi elemzés és generálás témakörének irányába fordult. Koutny Ilonával (Eötvös Lóránd Tudományegyetem, Általános és Alkalmazott Nyelvészeti Tanszék) és Prószéky Gáborral (MorphoLogic Kft.) történő együttműködésünk nyomán diplomatervemben az egyszerű magyar mondatok szintagmatikus elemzésén alapuló automatikus prozódia-címkéző algoritmust ismerttettem. 1997-től doktorandusz hallgatóként önálló kutató-fejlesztő tevékenységbe kezdtem. A nemzetközi ku-

tatási irányvonalat követve, ekkor került kutatásaim homlokterébe a hullámforma-összefűzéses technológián alapuló beszédszintézis. A jelfeldolgozási algoritmusok tanulmányozásával párhuzamosan mind nagyobb figyelmet szenteltem a gépi beszéd-keltéssel kapcsolatos szövegelemzésnek, a nyelvi modellek megismerésének és új modellek alkotásának.

Jövőbeni tudományos céljaimat a többnyelvű beszédszintézis megvalósításában, a különböző szintézis-módszerek (formáns- LPC-, hullámforma-összefűzéses és egyéb technológiák) integrációjában, nagy mennyiségű természetes emberi beszédből történő automatikus prozódia-kivonatoló rendszer létrehozásában, a parametrikus szintézisen alapuló, igen nagy természetességű beszédszintézis kutatásában, és a beszédszintézissel összefüggésben a magyar mondatok teljesség igényével történő szintaktikai elemzésében jelölöm meg.

2 Az értekezés célja

Az értekezés célja a magyar nyelvű szöveg–beszéd átalakító rendszer létrehozásával kapcsolatos kérdések elemzése, a meglévő nyelvi és jelfeldolgozási modellek és algoritmusok integrációja és új modellek alkotása, az eredmények gyakorlatba történő átültetése, a működő rendszer vizsgálata és az eredmények értelmezése. Ennek érdekében az alábbi célokat tűztem ki magam elé:

- Kidolgozom egy hullámforma-elem-összefűzéses technológián alapuló magyar nyelvű szöveg–beszéd átalakító rendszer tervét.
- A beszédszintézis-rendszerben alkalmazható algoritmust dolgozok ki a magyar nyelvű szövegekben előforduló számok betűkké alakítására.
- A rendelkezésre álló modellek alapján automatikus prozódiai címkéző algoritmust valósítok meg, amely mind önállóan, mind a beszédszintézis-rendszer részeként alkalmazható.
- A klasszikus generatív fonológia keretében meghatározott többszintű fonológia-modellt adaptálom a beszédszintézis céljára.
- Fonetikai-fonológiai alapokon univerzális beszédhang-ábrázolási rendszert dolgozok ki többnyelvű beszédszintézishez.
- Kidolgozom a hangkettősöket, hanghármasokat és ezeknél hosszabb elemeket is tartalmazó hullámforma-hangadatbázis szerkezetét, és algoritmust határozok meg a beszédhangsor fenti adatbázis-elemekkel történő optimális felépítésére.

- Bemutatok egy, a fogyatékos emberek kommunikációjában és rehabilitációjában alkalmazott piktogram–beszéd átalakító rendszert, és ezzel összefüggésben ismertetek egy kötött szótáron és korlátozott nyelvtanon alapuló magyar nyelvű morfológiai és szintaktikai generáló rendszert.

3 A kutatás és a vizsgálatok módszere

Munkám során az elméleti kutatási eredmények elérését gyakran a gyakorlati alkalmazás motiválta. Máskor éppen ellenkezőleg, a kutatási eredmények indukálták a gyakorlati alkalmazásokat. Kutatásaimat részben önállóan, részben tudományos vezetőimmel és munkatársaimmal együtt végeztem.

Kutatásaim egyik módszere a rendelkezésre álló szakirodalom alapján és saját eredményeink felhasználásával a meglévő modellek integrációja és új modellek alkotása. Ezt a módszert követtem például a beszéd-szintézis-rendszer szerkezetének kialakításakor, a grammatikai elemzésen alapuló prozódiai címkéző eljárás kidolgozásakor, vagy a kötött szótáras morfológiai és szintaktikai generáló rendszer elkészítésénél.

A korpuszelemzés módszerét alkalmaztam a számjegy–betű átalakító algoritmus megtervezésekor, nagy mennyiségű, számokat is tartalmazó elektronikus levelet vizsgálata során.

A prozódiai szabályrendszer kialakításakor konzulenseimmel a szintézis által történő analízis iteratív módszerét alkalmaztuk: létrehoztuk a szabályrendszer prototípusát, majd a meghallgatásos vizsgálatok alapján megjelöltük modell fejlesztési pontjait. A modell iteratív javításával fokozatosan alakítottuk ki a végleges változatot.

A beszédhangsor diádelemnél nagyobb elemekből történő felépítésének problémáját gráfelméleti algoritmusra vezettem vissza, és a hangsorépítés algoritmusát analitikus módon állítottam elő.

A szöveg–beszéd átalakító rendszert munkatársaimmal mind laboratóriumi, mind valós körülmények között vizsgáltuk. Az elkészült programot a stabil működés ellenőrzésére terheléses tesztnek vetettük alá. Ipari körülmények között a rendszert kísérleti elektronikuslevél-felolvasó szolgáltatás részeként vizsgáltuk, és a felhasználók észrevételeit folyamatosan regisztráltuk.

4 Új tudományos eredmények és műszaki megvalósításuk

Az új tudományos eredményeket az alábbi tézisek formájában összegzem:

1. Kidolgoztam a hullámforma-diádelemek összefűzésén alapuló komplex szöveg–beszéd átalakító rendszer tervét, és beszéd-szintézis-programmotor formájában megvalósítottam a rendszert.
2. A számok kiejtését vizsgáltam magyar nyelvű szövegekben. Eljárást dolgoztam ki a különböző szöveggörnyezetben előforduló számok betűsorozattá alakítására, beszédkeltés céljára.
3. Automatikus prozódiai címkéző eljárást hoztam létre. Elkészítettem az eljárásnak a felszíni jelenségek vizsgálatán alapuló prototípusát, és a komplex, morfológiai és szintagmatikus elemzésen alapuló változatát.
4. Tanulmányoztam a klasszikus generatív fonológiai leíráson alapuló többszintű fonológia-modellt. A modellt a beszéd-szintézis szempontjai szerint formalizáltam.
5. Nyelvfüggetlen, univerzális beszédhang-reprezentációt dolgoztam ki gépi beszédkeltéshez. Ezzel megteremttem a lehetőségét a rugalmasan bővíthető többnyelvű beszéd-szintézis-hangadatbázisok létrehozásának.
6. Kidolgoztam a diádnál nagyobb elemeket is tartalmazó komplex hullámforma-hangadatbázis szerkezetét. A beszédhangsor optimális felépítéséhez szükséges elemek kiválasztásának problémáját gráfelméleti algoritmusra vezettem vissza.
7. Kialakítottam a beszéd-sérült emberek rehabilitációjában alkalmazható Blissvox beszélő kommunikációs segédeszköz szerkezetét. Létrehoztam a kötött nyelvtanra működő Mondgen magyar morfológiai és szintaktikai generáló rendszert.

Az alábbiakban a tézisek szerinti felosztásban részletesen bemutatom tevékenységem folyamatát. Kitérek azokra a vizsgálatokra is, amelyek nem hoztak tézis szintű eredményeket, azonban véleményem szerint a kutatások szempontjából jelentőséggel bírnak.

1. Kidolgoztam a hullámforma-diádelemek összefűzésén alapuló komplex szöveg–beszéd átalakító rendszer tervét, és beszédszintézis-programmotor formájában megvalósítottam a rendszert.

Munkám során a formánszintézis-technológián alapuló Multivox 4-es szöveg–beszéd átalakító rendszerből indultam ki. Tanulmányoztam a beszédszintetizátor DOS alatt futó változatát, és Windows NT operációs rendszeren implementáltam a programot. Elkészítettem a módosított rendszer dokumentációját, és ez alapján dolgoztam ki a Multivox 5-ös, diádelemek összefűzésére épülő változatának rendszertervét [R2], [R3].

A szöveg–beszéd átalakító rendszer általam felépített modelljében szigorúan elkülönítettem az analízis és szintézis szakaszait. Az analízis-szakasz bemenete *grafémák*, azaz betűk, számok és egyéb írásjelek sorozata. A szintézis-szakasz kimenete a létrehozott beszédjel. A két szakasz között a szegmentális beszédhangsort és a szupraszegmentális vezérlő információt leíró paramétersorozat teremt kapcsolatot. Ezt a paramétersorozatot Olaszky és Németh terminológiájával *prozódiai mátrixnak* nevezem. A feldolgozási szakaszok elhatárolásával lehetővé válik egy rendszeren belül különböző szintézis-technológiák alkalmazása.

Felismertem, hogy az *írott szöveg* → *beszédhangsor-reprezentáció* → *prozódiai vezérlő paramétersorozat* → *beszédjel* átalakítás szigorúan szekvenciális, és hogy a folyamat modellezésére a csővezeték-elv a legelőnyösebb. Ennek alapján a feldolgozási lánc alábbi öt logikai részegységét határoztam meg: (1) graféma szintű előfeldolgozás, (2) graféma–beszédhang átalakítás, (3) beszédhang szintű műveletek, (4) beszédhang–hullámforma átalakítás és (5) hullámforma szintű műveletek.

A szakaszok között leírtam az adatfolyam állapotait, és definiáltam az interfészeket. Ezzel egzakt módon lehetővé tettem a feldolgozási folyamat lépésenkénti nyomon követését, és a folyamatban beavatkozási pontokat biztosítottam.

A logikai részegységek megvalósítása során az alább ismertetett részfeladatokat végeztem el.

A bemeneti szövegbe ágyazható továbbfejlesztett vezérlőparancs-készletet dolgoztam ki. E parancskészlet segítségével például módosítható a hangmagasság, a beszédtempó, a hangerő, vagy prozódiai vezérlő információ helyezhető el a szövegben. Munkám során a Multivox 4 vezérlő parancsaiból indultam ki, felhasználtam a DecTalk beszédszintetizátor és a Microsoft Speech Application Programming Interface (MS SAPI) vezérlő parancsait, illetve figyelembe vettem az Extendible Markup Language (XML) és a Speech Synthesis Markup Language (SSML) ajánlásait is. A parancskészletet az [R6] elektronikus feljegyzésben ismertettem.

Algoritmust dolgoztam ki magyar nyelvű szövegben a számjegyek betűsorozattá alakítására. Vizsgálataimról külön tézis formájában számolok be.

Automatikus prozódiai címkéző eljárást hoztam létre. Elkészítettem az eljárásnak a felszíni jelenségek vizsgálatán alapuló prototípusát, és a komplex, morfológiai és szintagmatikus elemzésen alapuló változatát. Eredményeimről külön tézis formájában számolok be.

Eljárást dolgoztam ki a grafémák szintjén a rövidítések szótárának hatékony ábrázolására. A szótárhoz szófa-struktúrán alapuló gyors keresési eljárást terveztem. A graféma szintű szabályrendszer kezeléséről az [R8] elektronikus feljegyzésben számoltam be.

Felismertem, hogy általános célú többnyelvű szöveg–beszéd átalakító rendszer megvalósításához a beszédhangok nyelvfüggetlen ábrázolása szükséges. Ennek érdekében univerzális beszédhang-ábrázolási sémát dolgoztam ki. Vizsgálataimról külön tézis formájában számolok be.

Kidolgoztam a graféma–beszédhang átalakítási szabályok és kiejtési szótár ábrázolási formáját [R11]. A szótárban a keresés hatékonyságát a szófa-struktúrájú szervezés biztosítja, nagy szótárméret mellett is. Gyakorlati szempontból kiemelem, hogy az általam bevezetett szótár-szervezéssel megszűnt a beszéd-szintézis-rendszer egyik korábbi korlátja, ami a kiejtési szótárba felvető szavak számát erősen korlátozta.

Fonetikus jelkészletet dolgoztam ki a beszédhangok megjelenítésére és a graféma–beszédhang átalakítási szabályok egyszerű módosíthatósága céljából [R12]. A beszéd-szintézis-rendszer korábbi változatában a kiejtési szabályokban a beszédhangokat számjegyes formában, kódjaikkal kellett felsorolni. Ez a szótárt áttekinthetlenné, a szerkesztést nehézkessé tette. Az általam kidolgozott fonetikus jelkészlet egyszerű, áttekinthető módot biztosít a szótár szerkesztéséhez és a beszéd-szintézis-rendszerben a feldolgozási lépések követéséhez.

Kísérletet tettem angol nyelvű graféma–fonéma átalakító algoritmus kidolgozására. A fonológiai szabályrendszer figyelembe vette a szavakban a szótagok számát, vizsgálta az ismert elő- és utótagokat, és beépített hangsúlyozási minták alapján bejelölte a hangsúlyok helyét is. A szabályrendszer kidolgozásában és az algoritmus tesztelésében hathatós segítséget kaptam Siptár Pétertől. A vizsgálatokat a NETalk kiejtési szótár felhasználásával végeztem. Azt találtam, hogy az általam kidolgozott algoritmus a szavak 58%-ára adta meg a pontos kiejtést. Az eredmény elmaradt a várakozásaimtól, és alátámasztotta azt a későbbi felismerésemet, hogy az angol nyelvű betű–beszédhang átalakítást gyakorlati célokra hatékonyan a szavaknak kiejtési szótárban történő felsorolásával lehet megoldani. Ebbéli meggyőződésemet megerősítette a lexikális fonológia mo-

delljének tanulmányozása során az a konklúzióm is, miszerint a lexikális fonológiai szabályrendszer a beszédszintézis-rendszerben rejtve maradhat, elegendő csupán a szavak felszíni alakjainak kezelése. Az eredményeimet a [B1] könyvcikkben publikáltam, illetve a [P1] zártkörű előadás formájában prezentáltam.

Szabályrendszert dolgoztam ki a beszédhangok szintjén fellépő hasonulás (zöngésedés, zöngétlenedés, képzés helye szerinti hasonulás, összeolvadás, hangkiesés és rövidülés) kezelésére és megjelenítésére. A fenti szabályrendszert kísérleti jelleggel implementáltam.

Eljárást dolgoztam ki a diádelemeket tartalmazó hangadatbázis formai és tartalmi vizsgálatára, és az adatbázis-hibák automatikus detektálására és javítására. A hangadatbázis vizsgálatának kérdéseit [R14] alatt tárgyaltam.

Megvalósítottam az Olaszgy által kidolgozott, a beszédhangok szintjén működő specifikus hangidőtartam és -intenzitás módosító szabályrendszert [R13]. Megoldottam a tagmondat-dallamegységek és intonációs vezérlő jelek beszédhangokra történő ráültetését, amely a dallamsémákat a szegmentumok szintjére vetíti. Elkészítettem a prozódiai vezérlésen alapuló elemösszefűző algoritmust és implementáltam a hangidőtartam, hangmagasság és hangerő módosító eljárásokat. Bevezettem az *intonációs profil* fogalmát, amely a hangmagasság-változások alapzöngé-periódusonkénti kezelését teszi lehetővé. Olaszgyval közös eredményeinket a [B2] könyvcikkben publikáltuk.

Implementáltam a kimeneti hullámforma 16 bites PCM, illetve 8 bites A-törvényű és μ -törvényű kódolási változatai közötti konverziós eljárásokat. A hordozhatóság figyelembevételével függvénykönyvtárat hoztam létre a szintetizált beszédminták lejátszására.

A szöveg–beszéd átalakító rendszert beszédszintézis-programmotorként (önállóan nem futtatható; programozói interfész segítségével alkalmazásokba építhető formában) implementáltam Windows NT operációs rendszeren [R4], [R5]. A létrehozott programmotor elnevezése: Multivox 5. Az erre épülő, beszédszintézis-szerver alkalmazás neve: Profivox.

A rendszeren sebesség- és stabilitási méréseket végeztem. Vizsgálataimat Pentium II 266 MHz processzorral, 64 MB memóriával rendelkező személyi számítógépen végeztem, Windows NT operációs rendszer alatt. A vizsgálatok egyik csoportjában a magyar szavak szintézisének sebességét vizsgáltam. Azt találtam, hogy a rendszer egy közel 800 000 szót tartalmazó gyakorisági szótár tartalmát hozzávetőlegesen 240 szó/másodperc sebességgel dolgozta fel. A vizsgálatok másik csoportjában a rendszer stabilitását ellenőriztem. Egy konkrét teszt során szöveges és bináris állományokat egyaránt tartalmazó fájlokat, összesen mintegy 1,4 GB mennyiségű adatot adtam a rendszer bemenetére. A ki-

sérlet során a rendszer valamennyi bemenő adatot hibajelzés nélkül képes volt értelmezni.

A programmotor Microsoft Speech API-ba épített változatát a keretrendszerben definiált terheléses tesztnak vetettük alá. A vizsgálatok során a rendszer bemenetére értelmes szavakat, mondatokat, és véletlen karaktersorozatokot adtunk. A mérések során jelentkező futási hibákat fokozatosan kiküszöböltük. A beszédszintézis-rendszeren számos meghallgatásos tesztet is elvégeztünk. A vizsgálatok során született szubjektív véleményeket a nyelvi modellek iteratív fejlesztése során figyelembe vettük.

A beszédszintézis-programmotort beépítettük a tanszéken kifejlesztett elektronikuslevél-felolvasó rendszerbe is. A kísérleti rendszer 1999 decembere óta a Westel Mobil Távközlési Rt. *Mailmondó* szolgáltatásaként üzemel. A hozzánk eljuttatott felhasználói észrevételeket folyamatosan regisztráljuk. A teljes rendszer stabilitást jól jellemzi, hogy a szolgáltató visszajelzése szerint a telepítéstől eltelt két év alatt szoftverhiba miatt egyszer sem állt le a szolgáltatás.

A tézisben megfogalmazott eredmények közül kiemelném a beszédszintézis-rendszer moduláris felépítését és áttekinthető szerkezetét, ezáltal biztosítva a rendszer jövőbeni továbbfejleszthetőségét és bővíthetőségét. A hullámforma-összefűzéses technológia megvalósításával és a prozódia hatékony kezelésével lehetővé tettem az emberi beszédet megközelítő természetességű gépi beszéd előállítását.

Tekintettel a rendszer ipari alkalmazására, a technikai részeredményeket, algoritmusokat elsősorban a rendszer fejlesztői számára hozzáférhető belső elektronikus feljegyzések formájában publikáltam. Az elkészült beszédszintézis-rendszerről a [J3] cikkben számoltunk be. Az elektronikuslevél-felolvasó rendszer kísérleti változatról a [B3] könyvrészletben, a konkrét szolgáltatásban működő rendszerről munkatársaimmal a [J4] cikkben adtunk számot. A szövegbeszéd átalakító rendszeren alapuló kötött szótáras üzenetszerkesztő rendszert a [C3] és [C4] konferencia-cikkekben ismertettünk.

A beszédszintézis-rendszer fent bemutatott architektúrája a BME Villamosmérnöki és Informatikai Karán a Beszédinformációs rendszerek tantárgy tematikájának részét képezi.

2. A számok kiejtését vizsgáltam magyar nyelvű szövegekben. Eljárást dolgoztam ki a különböző szövegkörnyezetben előforduló számok betűsorozattá alakítására, beszédeltés céljára.

A számjegyek lehetséges kiejtési változatait vizsgáltam magyar nyelvű szövegekben. Tanulmányoztam a magyar tőszámneveket, törtszámneveket és sor-

számneveket alkotó számelemeket, azok számjegyekkel, illetve betűkkel történő írásmódját és az ejtési változatokat.

Vizsgáltam a számok környezetében előforduló írásjelek szerepét sorszámokban, tizedes törtekben, dátumokban. Külön foglalkoztam az ezres határoló és a mondatvégi ponttal, illetve a felsorolást, tagmondat határát jelölő vesszővel.

Tanulmányoztam a telefonszámok írásmódját és prozódiaját. A telefonszámformátumokat 20 000 elektronikus levélből automatikus kivonatolással nyert telefonszám-lista alapján állapítottam meg. Azt találtam, hogy a vizsgált mintában a telefonszámoknak mintegy 70 különböző írásmódja azonosítható. Eljárást dolgoztam ki a leggyakoribb telefonszám-formátumok kezelésére.

Tanulmányoztam a dátumok és időpontok leggyakoribb írásmódjait. Eljárást dolgoztam ki a leggyakoribb formátumok betűkkel történő átírására.

Külön foglalkoztam a számjegyek kiejtését módosító toldalékokkal. Ezek között vizsgáltam a sorszámnev és dátum képző toldalékokat, a közönséges törtek írásmódját, és a számjegyekhez kapcsolódó névszói esetragok módosító hatását. Ilyenek például a nyílt-magánhangzó-nyúlás, a töváltozás, illetve a szövegkörnyezettől függő rövid és hosszú szóalakok. Algoritmust készítettem exponenciális alakban megadott számok betűkké történő átalakítására.

Olaszy és Németh szempontjai alapján vizsgáltam a számelemek kiejtésekor felmerülő prozódiai kérdéseket. Megvalósítottam az általuk kidolgozott, a számokhoz automatikusan prozódiai információt rendelő algoritmusokat.

Itt jegyzem meg, hogy Olaszy magyar, német és angol tőszámnevek felolvasásához tervezett számelem-készletet és hangadatbázist, amely a koartikulációs hatásokat is figyelembe vette. Esetemben a koartikulációs hatásokkal nem kellett számolnom, mert célom számjegy–betű átalakító eljárás kidolgozása volt. A koartikulációt és egyéb posztlexikális fonológiai jelenségeket a feldolgozás későbbi szintjén kezeli a beszédszintézis-rendszer. Másrésről a célom nem csupán tőszámnevek, hanem tetszőleges szövegkörnyezetben előforduló számok felolvasása volt.

Háromszintű modellt dolgoztam ki a számokat betűsorozattá alakító algoritmus részeként. A modell legalsó szinten helyezkedik el az alapszintű, a számokat rögzített szabály szerint átalakító eljárás. A középső szinten történik a formázott alakok, például telefonszámok, tizedes törtek átalakítása. A legfelső szinten történik a számokat és betűket vegyesen tartalmazó szöveg értelmezése. A szám–betű átalakító eljárást a fenti modell szerint valósítottam meg.

A számjegy–betűsorozat átalakító modult a beszédszintézis-rendszerbe építve vizsgáltam. A vizsgálatokhoz elektronikus leveleket választottam, amelyek összesen 110 000 szót tartalmaztak. E levelekből 560 olyan kifejezést válogat-

tam ki, amelyekben különböző szövegösszefüggésben álló számok fordultak elő: dátum, idő, tő- és sorszámnevek, pénzegységek, tizedes törtek, szoftver verziószámok, toldalékolt alakok, kötőjellel és egyéb jelekkel határolt alakok, számok és betűk vegyesen szerepeltek a feldolgozott anyagban. Az ellenőrző vizsgálatok eredménye szerint a rendszer az esetek 96 százalékában helyesen alakította át a szövegbe ágyazott számokat. A hibákat az iteratív modell-fejlesztési eljárás módszerével azonosítottuk és javítottuk ki.

A számjegy–betűsorozat átalakító algoritmról először az [R18] belső feljegyzésben számoltam be. A tudományos eredményeimet [J1] alatt publikáltam.

3. Automatikus prozódiai címkéző eljárást hoztam létre. Elkészítettem az eljárásnak a felszíni jelenségek vizsgálatán alapuló prototípusát, és a komplex, morfológiai és szintagmatikus elemzésen alapuló változatát.

A beszéd-szintézis-rendszerben prozódia alatt a hangidőtartamok, a szünetek, a hangmagasság és az intenzitás változásait értjük. Ezeket a beszédhangsorra, mint szegmentumokra, ültetett szupraszegmentális információnak fogjuk fel.

A beszéd-szintézis-rendszerben a prozódia-kezelést az Olasz, Németh és Koutny által kidolgozott háromszintű modell szerint értelmeztem, amely következő lépésekből áll: (1) a grafémák szintjén elvégzett nyelvtani analízis és a magas szintű prozódiai vezérlés előállítására, (2) a beszédhangok szintjén elvégzett elemzés és alacsony szintű prozódiai vezérlés, és (3) a beszédjelen az előírt paraméterek alapján a prozódia beállítása.

Olasz modellje alapján implementáltam a prozódiai paraméterek szegmentális szempontból vett háromszor hármas tagozódását: 1. Dallammenet: 1a. Mondatok, tagmondatok, kifejezések szintjén jelentkező globális dallammenet. 1b. Mondaton belüli szóhangsúlyok. 1c. Beszédhang szintű mikrointonáció. 2. Időtartam-szerkezet: 2a. A mondat globális időtartam-szerkezete. 2b. Szavak időtartam-térképe. 2c. Beszédhangok környezetfüggő specifikus időtartamai. 3. Intenzitás-szerkezet: 3a. Mondatszintű intenzitás paraméterek. 3b. Szószintű intenzitás-térkép. 3c. Beszédhangok szintjén jelentkező intenzitás-változások.

A prozódiai címkéző eljárás Olasz által kidolgozott prototípusa alapján rendszereztem a leírt nyelvi anyagot, és kidolgoztam az algoritmust, amely a mondat mélyre ható elemzése nélkül határozza meg a magas szintű prozódiai vezérlő paramétereket. Ez az eljárás a mondatbeli funkciószavak, jellemző szókapcsolatok azonosításán alapul.

Koutny nyelvi modellje alapján megvalósítottam a prozódiai címkéző eljárás komplex változatát, amely a szavak morfológiai elemzésén és a szintagmák azonosításán alapuló prozódiai címkéző szabályrendszerre épül.

A morfológiai elemzéshez a *Humor* gyors morfológiai elemző programot használtam. Morfológiai szótárt hoztam létre, amely kiegészíti a morfológiai elemző lexikonját. Eljárást készítettem a hibás morfológiai elemzések korrigálására. Algoritmust dolgoztam ki az alternatív elemzési változatok párhuzamos kezelésére és a morfológiai jegyek hatékony ábrázolására.

A szintagmatikus elemzés keretét Koutny Ilona függőségi nyelvtanon alapuló szintaktikai elemzőjének modellje adta, amely alapján meghatároztam a kifejezések unifikáción alapuló ábrázolási struktúráját. A nyelvtani szerkezetek azonosításához a morfológiai elemzés kimenetét felhasználó, gyors kifejezés-minta-illesztő eljárást terveztem.

A prozódiai címkéző eljárás elkészítéséhez gyakorlati szempontok szerint formalizáltam a Koutny által kötött jelentéstartalmú, egyszerű mondatokra kidolgozott prozódiai címkéző szabályrendszert, amely a morfológiai és szintagma-szintű információ ismeretében meghatározza az egyes szavak hangsúlyszintjeit, kijelöli a kifejezések határait, dallammintákat rendel a kifejezésekhez, és meghatározza a szünetek helyét és időtartamait.

A prozódiai címkéző eljárásnak mind a prototípusát, mind a komplex algoritmust megvalósítottam a beszéd-szintézis-rendszer részeként. Az algoritmusokat nagyszámú meghallgatásos vizsgálat során tökéletesítettük.

A tézis jelentőségét abban látom, hogy a prozódiai címkéző eljárással bővített rendszer által létrehozott beszédjel természetessége jelentősen javult, ezáltal lehetővé vált a mondattípusok gazdagabb skáláját tartalmazó – akár irodalmi – szövegek felolvasása is.

A társzerzőimtől független önálló eredményemnek tekintem a szabályrendszer konzisztencia-vizsgálatát és egyszerűsítését, a szabályok szigorú formalizálását, az elméleti modellhez kapcsolódó algoritmusok kialakítását, azok gyakorlatba történő átültetését.

Az alapváltozat Olaszgy által kidolgozott egyszerű szabályrendszerét az [R7] belső elektronikus feljegyzésben formalizáltam. A komplex rendszer kialakításával kapcsolatos munkákról a [T1] diplomaterv-dolgozatomban, és az [A1], [A2] és [C2] konferencia-előadásokon, illetve a [P2] zártkörű előadáson számoltam be. A technikai kérdésekről az [R19] kutatási zárójelentésben adtam számot. A beszéd-szintézis-rendszerbe illesztett háromszintű prozódia modellt a [C5] konferencia-előadáson ismertettük.

A Portói Egyetemen végzett kutatási program keretében tanulmányt készítettem, amelyben a mondatprozódia szintagmatikus elemzésen alapuló automatikus meghatározásának általános, nyelvfüggetlen kérdéskörét tárgyaltam. Munkámat az [R1] kutatási jelentés formájában foglaltam össze.

4. Tanulmányoztam a klasszikus generatív fonológiai leíráson alapuló többszintű fonológia-modellt. A modellt a beszéd-szintézis szempontjai szerint formalizáltam.

A beszéd-szintézis-rendszer korábbi változatait tanulmányozva szembesültem azzal a problémával, hogy a rendszer a beszédhangsort közvetlenül az izolált szavak írott formájából próbálta meghatározni, és így nem tudott számot adni a szavak határán átnyúló, tisztán a beszédhangsor (és nem a betűsorozat) szintjén működő kiejtési szabályokról. Kiefer, illetve Durand és Siptár munkáin keresztül tanulmányoztam a lexikális fonológia-modellt, és felismertem, hogy egy teljesség igényével megtervezett beszéd-szintézis-rendszerben elkerülhetetlenül szükség van a lexikális fonológia keretein belül kidolgozott többszintű szabály-rendszer kialakítására.

A lexikális fonológia modelljében a szabályok három szintre oszthatók: az első csoportban a lexikon szintjén működő fonológiai és morfológiai szabályok előállítják a szó szótári ábrázolását. A második szinten helyezkednek el a lexikális beillesztés által vezérelt szabályok, amelyek a szóalakok mondatba illesztett változatát hozzák létre, végül a harmadik szinten szerepelnek a felszíni alakokat előállító posztlexikális fonológiai szabályok. Ez utóbbiak a lexikális elemekből összeállított hangsonon működnek, és az eredeti szó- illetve morféma-határokon is átnyúlnak.

Először a modell első két szintjét vizsgáltam. Felismertem, hogy a lexikális fonológiai szabályok jelentős részét a beszéd-szintézis-rendszerben nem indokolt megvalósítani, mert a) a szavak kiejtett (fonetikus) alakjai a lexikonban megadhatók, és b) a magyar nyelv fonemikus jellege miatt az írott forma nagy mértékben tükrözi a kiejtett alakot. A lexikális és a szintaktikai beillesztés szabályait a graféma–fonéma átalakítás szintjén ábrázoltam. A szabályrendszer magában foglalja az egyszerű és kettős betűkre vonatkozó betű–beszédhang átalakítási szabályokat, a betűzést, a magánhangzók időtartam-változásait, a lexikális *h*-törlést, a lexikális palatalizációt és a *j*-hasonulási szabályokat. Szintén itt tároltam a rendhagyó írásmódú családnevek, idegen szavak és rövidítések kiejtését.

Ezt követően a fonológia-modell posztlexikális harmadik szintjén álló szabályokat formalizáltam. E szabályokat szintén a klasszikus generatív fonológia megkülönböztető jegyes ábrázolásában tanulmányoztam, de törekedtem a gyakorlati szempontoknak jobban megfelelő leírasmódra. A szintézis-rendszer beszédhang szintű feldolgozási lépésében kerül tehát sor a mássalhangzó-ikeredés (gemináció) és degemináció, a nazálisok képzés szerinti hasonulásának, az abszolút szóvégi *j* palatalizációjának, a posztlexikális palatalizációnak, a zöngés-

ségi hasonulásnak, hangkiesésnek, összeolvadásnak, a szinkópa-jelenségnek és minden más, posztlexikális szinten jelentkező szabályoknak a kezelésére.

A tézis jelentőségét abban látom, hogy az általam formalizált szabályrendszerrel a beszédszintézis-rendszerben a teljesség igényével számot adhatunk valamennyi elméletileg leírt fonológiai jelenségről. További előny, hogy az egzakt leírással csökkenthetők a fonológiai szabályok közötti egymásra hatások, illetve minden szabályról a maga helyén (a lexikálisokról a lexikonban, a posztlexikálisokról a beszédhangok szintjén) adhatunk számot. A szabályrendszert fordítási időben előkészítve és a kiejtési szótár formátumában tárolva ki tudjuk aknázni az utóbbi évek számítástechnikai tárhelykapacitás-növekedésében rejlő lehetőségeket a futási idő csökkentésére.

A lexikális és szintaktikai beillesztés szintjén működő szabályokat az [R11], a posztlexikális fonológiai szabályokat az [R13] feljegyzésekben formalizáltam.

5. Nyelvfüggetlen, univerzális beszédhang-reprezentációt dolgoztam ki gépi beszédkeltéshez. Ezzel megteremttem a lehetőségét a rugalmasan bővíthető többnyelvű beszédszintézis-hangadatbázisok létrehozásának.

A Multivox beszédszintetizátor 3-as, többnyelvű változata nyolc nyelvre készült el. Ebben a rendszerben a beszédhangokat úgynevezett hangkódok azonosították. A beszédhangok száma nyelvenként 30–50 között mozgott. Egy adott nyelvhez tartozó hangkódok egy másik nyelv esetében sokszor eltérő beszédhangokat jelöltek. Gondot okozott a meglévő rendszerek új beszédhangokkal történő bővítése: a hangkódok kiosztási szisztémája miatt megtörténhetett, hogy „elfogytak” a szabad hangkódok – a rendszerbe nem lehetett újabb beszédhangot felvenni. Ezek a problémák gátját szabták több nyelv egyetlen beszédszintézis-rendszerbe történő integrálásának.

Ki kell emelnem, hogy a beszédszintézis-rendszer korábbi változatai az adatbázis méretének minimalizálása érdekében gyakran megelégedtek a nyelv *fonémáinak* kezelésével, és csak később merült fel a *beszédhang*-szintű leírás igénye. A fonéma a nyelven belül megkülönböztető erővel bíró egység, de csupán a beszélő szándékát kifejező absztrakt kategória. A fonéma egy konkrét realizációját beszédhangnak nevezzük. Így például az /n/ fonéma a *len* szóban az [n] beszédhang formájában realizálódik, míg a *leng* szóban [ŋ] alakban szerepel. Úgy is mondhatjuk, hogy az [n] és az [ŋ] beszédhang az /n/ fonéma egy-egy *allofónja*, azaz ejtési változata. Az ejtési változatok ábrázolására a beszédszintézis-rendszerben az utóbbi időben merült fel igény.

A fenti kérdés megoldására univerzális beszédhang ábrázolási sémát dolgoztam ki. Tanulmányoztam a Nemzetközi Fonetikai Társaság (IPA) ajánlásaként közzétett fonetikai ábécét, és a témában elérhető elektronikus dokumentumokat.

A beszédhang-reprezentációs séma megtervezésekor az alábbi szempontokat vettem figyelembe: egy kód egyértelműen ábrázoljon egy beszédhangot; az ábrázolás legyen a lehető legtömörebb; a kódrendszer képes legyen az egymáshoz fonetikai-fonológiai szempontból hasonló beszédhangok „hasonlóságát” ábrázolni; a leírás legyen átfogó (például adjon lehetőséget az allofónok kezelésére is); végül a leírás az elméleti rendszerezésekben szereplő beszédhang-jellemzőket ortogonális rendszerben tartalmazza (más szóval, a beszédhang egy bizonyos tulajdonság-jegyét egy jellemző írja le, és egy jellemző csak egy tulajdonság-jegyét ábrázolja).

Az IPA fonetikus jelkészletének előnye, hogy igyekszik valamennyi beszédhang ábrázolására jelölést adni. Hátránya, hogy a jelrendszer nem mutat kohéziót, más szóval az ejtésben „közeli” beszédhangok jelképei nem tükrözik a hasonlóságot. További probléma, hogy a jelkészlet az ember által történő olvashatóság érdekében kényszerű kompromisszumot köt: míg egyik-másik beszédhangot a kiejtést sejtető szimbólum ír le, addig más hangokat olyan (esetleg találmokra választott) jel ábrázol, amely nem sugallja a kiejtést. Olykor különböző mellékjelekkel egészülnek ki a szimbólumok.

A SAMPA fonetikus jelkészletet az IPA szimbólumok számítógépen történő ábrázolására és továbbítására hozták létre. Itt minden IPA-jelnek egy vagy több 7 bites ASCII karakter felel meg. Ebben a rendszerben egy beszédhang ábrázolásához sokszor 3–4 karakter is szükséges. A feladat szempontjából tehát a SAMPA jelkészletre is érvényesek az IPA jelkészletnél ismertetett hátrányok.

Az általam kidolgozott, fonetikai-fonológiai jegyeken alapuló beszédhang-ábrázolási kódrendszer a magánhangzókban hat tulajdonság-csoportot különböztet meg: kerekesség, a nyelv vízszintes és függőleges pozíciója, hanghosszúság, nyelvállás és nazalizáltság szerint. A mássalhangzók esetében a következő tulajdonság-csoportokat határoztam meg: zöngesség, képzés helye és módja, szótagalkotóság, és a képzést módosító egyéb jegyek az aspirált, palatalizált, velarizált, stb. kiejtési változatok megkülönböztetésére. Egyetlen kóddal jelöltem mássalhangzók esetében az affrikátákat, de magánhangzóknál a diftongusok leírásakor két kód ábrázolja a kettőshangzót. Az utóbbiak ábrázolásához bevezettem a „kapcsolva ejtendő” jelölést. Ezzel a hármashangzók (trifongusok) kezelése is könnyen lehetővé válik.

A javasolt kódrendszer jellegéből adódóan egyáltalán nem törekszik ember által olvasható megjelenésre, azonban könnyűszerrel leképezhető akár az IPA jelkészletre, akár másik alkalmasan megválasztott fonetikus ábécére.

A tézis jelentőségét abban látom, hogy a kidolgozott beszédhang-leírási rendszer egységes ábrázolást biztosít mind a többnyelvű beszédshintézis esetében, mind pedig egy nyelv szintézise során a különböző ejtésváltozatok és allofónok kezelésében. A hullámforma-hangadatbázis komplex változatában az itt bemutatott beszédhang-leírási rendszert alkalmaztam. A kidolgozott leírást az [R9] és [R10] belső feljegyzésekben publikáltam. A széleskörű publikációt megelőzően igyekszem az elméleti leírás használhatóságát minél több gyakorlati alkalmazásban bizonyítani.

6. Kidolgoztam a diádnál nagyobb elemeket is tartalmazó komplex hullámforma-hangadatbázis szerkezetét. A beszédhangsor optimális felépítéséhez szükséges elemek kiválasztásának problémáját gráfelméleti algoritmusra vezettem vissza.

A diádelem-összefűzéses technológia továbbfejlesztéseként Németh és Olaszy kezdeményezésére olyan rendszer kidolgozását kezdtük meg, amely a diádok mellett képes hanghármasok, -négyesek, illetve ennél több beszédhangból álló elemek kezelésére is. Az ilyen elemeket tartalmazó adatbázist komplex hangadatbázisnak nevezzük.

Munkám során a diádos hangadatbázis szerkezetéből indultam ki. Eljárást dolgoztam ki az adatbázis szerkezeti és tartalmi ellenőrzésre, és automatikus javító algoritmust konstruáltam az adatbázis-inkonzisztencia kiküszöbölésére. Kidolgoztam az új, komplex hangadatbázis szerkezetét.

A diádelemeket és a CVC (mássalhangzó-magánhangzó-mássalhangzó) triádokat tartalmazó komplex hangadatbázisban az elemek száma megközelíti a nyolcezret. A beszédhangsor felépítése során futási időben kell az adatbázisból kiválasztani a hangsort optimálisan lefedő elemet. Optimálisnak tekintem azt a lefedő algoritmust, amely a megadott beszédhangsort a lehető legkevesebb adatbázis-elem felhasználásával építi fel. Eljárást dolgoztam ki a komplex adatbázisban történő gyors keresésre, és több lehetséges elem-jelölt esetében az optimális elem kiválasztására. A problémát visszavezettem Dijkstra legrövidebb út keresési gráfelméleti algoritmusára. Az eredeti eljárásból kiindulva lineáris idejű lefedési algoritmust konstruáltam, ami az adatbázis-elemek nagy száma miatt bír kiemelt jelentőséggel. Megmutattam, hogy a kidolgozott algoritmus minden esetben az optimális hosszúságú elemet választja ki az adatbázisból. A kereső

algoritmust a komplex adatbázis kísérleti változatával együtt implementáltam. A szegmentális beszédjelen alapszintű meghallgatásos kísérleteket végeztem.

Bevezettem a beszédhang-helyettesítés fogalmát. A beszédjel-előállítás során előfordulhat, hogy az hangadatbázis nem tartalmazza a felépítendő hangsor valamelyik beszédhangját vagy az előírt hangkapcsolatok valamelyikét. Ebben az esetben a hangsor felépítéséhez olyan adatbázis-elemeket kell kiválasztania a rendszernek, amelyek a lehető legkisebb mértékben térnek el az előírt beszédhangoktól. Az előírt beszédhangokat a beszédhang-helyettesítő eljárás képezi le a rendelkezésre álló hangkapcsolatokra. Ha a kimondandó szövegben előírt beszédhang nem található meg az éppen használt hangadatbázisban, akkor a hangot az azt fonetikai-fonológiai szempontból legjobban közelítő hanggal helyettesítjük. A helyettesítés előre rögzített táblázat alapján történik.

A komplex hangadatbázis szerkezetét az [R15], a belső hangadatbázis-ábrázolást az [R16], az adatbázis-változatok kezelésének összehasonlítását az [R17] elektronikus feljegyzésekben összegeztem.

A tézisben ismertetett módszer jelentősége, hogy (1) a hangadatbázis-elemeket szófa-struktúrába szerveztem, amely igen nagy elemszámú hangadatbázis esetén is gyors keresést biztosít, és (2) a hangsorépítéshez kidolgozott algoritmus tetszőleges adatbáziselem-hossz esetén is beszédhangsor optimális lefedését biztosítja.

7. Kialakítottam a beszédsérült emberek rehabilitációjában alkalmazható Blissvox beszélő kommunikációs segédeszköz szerkezetét. Létrehoztam a kötött nyelvtanra működő Mondgen magyar morfológiai és szintaktikai generáló rendszert.

A beszédszintézis-rendszer és az írott nyelv-technológia ötvözéseként, a BME TTT Beszédtechnológiai Laboratórium és a Segítő Kommunikáció Módszertani Központ (SKMK) közötti együttműködés keretében hoztuk létre a Blissvox kommunikációs segédeszközt, amelyet mozgássérült, értelmi fogyatékos gyermekek rehabilitációjában alkalmaznak. A számítógépes segédeszköz a fogyatékosok kommunikációjában elterjedt Bliss jelképrendszeren alapul. A felhasználók egyszerű jelképek, piktogramok egymás mellé illesztésével üzeneteket állíthatnak össze, amelyeket a rendszer a hozzá kapcsolt formánsszintetizátor segítségével magyar nyelven felolvas. A rendszer első változatát Kálmán Zsófia doktornő, az SKMK vezetője és Olaszgy Gábor elképzelései alapján valósítottam meg 1994-ben.

A Bliss jelképrendszert eredetileg angol nyelvterületen, angol nyelvű felhasználók számára dolgozták ki. Az üzenetekben egymás mellé illesztett jelké-

pek eredeti angol jelentését összeolvasva nyelvtanilag közel helyes angol mondatot kapunk. A Blissvox rendszerben a jelképek magyar jelentését alkalmaztuk. A felhasználók által készített üzenetekben az egymás után felsorolt szavakból összeállított mondat a toldalékok hiányában magyartalan volt, és sok esetben a szórend sem felelt meg nyelvünk szabályainak. E probléma kiküszöbölésére hoztuk létre a Mondgen rendszert, amely különálló szótövekből és toldalékokból kötött szótár és korlátozott nyelvtan alapján megkísérli előállítani a nyelvtanilag helyes mondatot.

A Blissvox részeként kialakított mondatgeneráló rendszert a 2500 jelképet tartalmazó Bliss-szótár szavaira és korlátozott nyelvtanra dolgoztuk ki. A mondatgenerátort úgy terveztük meg, hogy az ne csak a Blissvox részeként, hanem önálló programként is működjön, és beépíthető legyen más alkalmazásokba is. A Mondgen rendszert Koutny Ilona elképzelései alapján és nyelvészeti vezetésével készítettem el. A névszók szintézise Papp Ferenc, az igék szintézise Elekfi László munkáin alapult, amelyeket Prószéky Gábor rendszerezésében tanulmányoztunk.

A nyelvtani generálási folyamat négy fő lépésre bontható: előfeldolgozás, a névszói és igei csoportok kialakítása, a mondatszerkezet kialakítása, és a morfológiai generálás fázisára.

A szavak és toldalékok azonosítására szótárt terveztem. Algoritmust készítettem a szótárban nem szereplő szavak szófaji, ragozási és hangrendi csoportba sorolására. Formalizáltam a szintagmák felépítését és a mondatszerkezetet kialakító szabályrendszert. Külön algoritmusokat készítettem a névszók és az igék szintézisére, amik az előírt tőtípus, ragozási típus, hangrend, nyelvtani szám és személy alapján létrehozták a mondatba illeszkedő szóalakokat. A Mondgen rendszert a Blissvox részeként megvalósítottam.

A komplex rendszer az SKMK-ban végzett kísérletek tanúsága szerint beváltotta a hozzá fűzött reményeket. A Blissvox rendszer alapváltozatáról a [C1] konferencia-előadásban és az [S1] Tudományos Diákköri Konferencián (TDK) számoltam be. A [B5] könyvfejezetben a konferencia-előadás magyar fordítását adtam közre. A Mondgen rendszerről először az [S2] TDK előadásban adtam hírt. A Blissvox rendszer mondatgenerátorral bővített változatát a [J5] folyóirat-cikkben mutattuk be.

Köszönetnyilvánítás

Köszönöm doktori témám tudományos vezetőinek, Olasz Gábornak és Németh Gézának, hogy megismertették velem a beszédtechnológia számos területét, és a Beszédtechnológiai Laboratóriumban a kezdetektől fogva figyelemmel kísérték és tanácsaikkal segítettek munkámat, biztosították a kutatásaimhoz szükséges feltételeket, és eredményeim publikálására ösztönöztek.

Köszönöm Koutny Ilonának, hogy a mondatgeneráló rendszer és a prozódiai címkéző eljárás kidolgozása során a vonatkozó nyelvi elméletekkel megismertette, és az elmúlt évek során elindított a magyar nyelvten mélyebb tanulmányozásának irányába.

Külön köszönet illeti Prószéky Gábort, aki kutatásainkhoz rendelkezésünkre bocsátotta a *Humor* morfológiai elemző programot, és emellett számos gyakorlati tanáccsal is segítette a szintagmatikus mondatelemző megvalósítását.

Az angol betű–beszédhang átalakítási szabályok kidolgozása során Siptár Pétertől kaptam hathatós segítséget, és ő volt az, aki időt és fáradságot nem kímélve nagy részt vállalt az algoritmus tesztelésében, és számos hibára hívta fel a figyelmemet. A nyelvfüggetlen beszédhang-ábrázolási rendszer kidolgozásakor szintén támaszkodhattam véleményére és tanácsaira. Személyes beszélgetéseink során és könyveiből is sokat tanultam.

Személyes köszönettel tartozom Gordos Géza professzor úrnak, aki egyetemi éveimtől fogva a BME Távközlési és Telematikai Tanszéke vezetőjeként figyelemmel kísérte és támogatta a tanszéki munkámat.

5 A tézisek témaköréből készített publikációk

A publikációk jelölésére az alábbi konvenciót alkalmazom:

B	Cikk szerkesztett könyvben
J	Külföldön megjelent idegen nyelvű folyóiratcikk
C	Nemzetközi konferencia-kiadványban megjelent idegen nyelvű előadás
A	Csak kivonatban megjelent konferencia-előadás
P	Csak szóban elhangzott előadás
R	Könyvtárakban el nem helyezett kutatási jelentés, elektronikus feljegyzés
S	Tudományos Diákköri dolgozat
T	Diplomaterv-dolgozat

5.1. Cikk szerkesztett könyvben

- [B1] **Olaszi P.** (1994). Számítógépes algoritmus angol szöveg fonetikus átalakítására. In: Gósy Mária (Ed.): *Beszéd kutatás '94*, MTA Nyelvtudományi Intézete, Budapest. pp. 183–197.
- [B2] Olaszy G., **Olaszi P.** (1998). Hangidőtartamok mesterséges változtatása periódusok kivágásával, megismétlésével. In: Gósy Mária (Ed.): *Beszéd kutatás '98*, MTA Nyelvtudományi Intézete, Budapest. pp. 151–162.
- [B3] Németh G., Zainkó Cs., Bogár B., Szendrényi Zs., **Olaszi P.**, Ferenczi T. (1998). Elektronikuslevél-felolvasó. In: Gósy Mária (Ed.): *Beszéd kutatás '98*, MTA Nyelvtudományi Intézete, Budapest. pp. 189–203.
- [B4] Olaszy G., Kiss G., Németh G., **Olaszi P.** (2000). Profivox: a legkorszerűbb hazai beszéd szintetizátor. In: Gósy Mária (Ed.): *Beszéd kutatás 2000*, MTA Nyelvtudományi Intézete, Budapest. pp. 167–179.

5.2. Külföldön megjelent idegen nyelvű folyóiratcikk

- [J1] **Olaszi P.** (2000). Analysis of Written and Spoken Form of Hungarian Numbers for TTS Applications. In: Olaszy G., Gardner-Bonneau D. (Eds.): *International Journal of Speech Technology*. Kluwer, Boston. pp. 177–186.

- [J2] Koutny I., Olaszy G., **Olaszi P.** (2000). Prosody Prediction from Text in Hungarian and its Realization in TTS conversion. In: Olaszy G., Gardner-Bonneau D. (Eds.): *International Journal of Speech Technology*. Kluwer, Boston. pp. 187–200.
- [J3] Olaszy G., Németh G., **Olaszi P.**, Kiss G., Zainkó Cs., Gordos G. (2000). Profivox—a Hungarian Text-to-Speech System for Telecommunications Applications. In: Olaszy G., Gardner-Bonneau D. (Eds.): *International Journal of Speech Technology*. Kluwer, Boston. pp. 201–215.
- [J4] Németh G., Zainkó Cs., Fekete L., Olaszy G., Endrédi G., **Olaszi P.**, Kiss G., Kis P. (2000). The Design, Implementation and Operation of a Hungarian E-mail Reader. In: Olaszy G., Gardner-Bonneau D. (Eds.): *International Journal of Speech Technology*. Kluwer, Boston. pp. 217–236.
- [J5] **Olaszi P.**, Koutny I., Kálmán S. L. (2002). From Bliss Symbols to Grammatically Correct Voice Output: A Communication Tool for People with Disabilities. In: Gardner-Bonneau, D. (Ed.): *International Journal of Speech Technology*. Vol. 5:1. Kluwer, Boston, pp. 49–56.

5.3. Nemzetközi konferencia-kiadványban megjelent idegen nyelvű előadás

- [C1] Olaszy G., Kálmán Zs., **Olaszi P.** (1994). BLISSVOX – Voice Output Communication System for Teaching, Rehabilitation and Communication. In: Zagler, W.L., Busby, G., Wagner, R.R. (Eds.): *Computers for Handicapped Persons, 4th International Conference*. Springer-Verlag, Wien. pp. 421–428.
- [C2] **Olaszi P.**, Koutny I., Olaszy G., Németh G. (1998). Syntactic Analysis of Hungarian Sentences to Predict Prosodic Information for Speech Synthesis. In.: *Proceedings of the 1998 Polish-Czech-Hungarian Workshop on Circuit Theory, Signal Processing and Applications*. Krakow, Poland. pp. 49–54.
- [C3] Olaszy G., Németh G., **Olaszi P.**, Gordos G. (1999). Interactive, TTS Supported Speech Message Composer for Large, Limited Vocabulary, but Open Information Systems. In.: *Eurospeech '99*. Vol. 2., Budapest, Hungary. pp. 943–946.

- [C4] Olaszy G., Németh G., **Olaszi P.** (1999). Preparation of Limited Vocabularies with an Interactive Open TTS Based Development System. In: *Proceedings of the 1999 Polish-Hungarian-Czech Workshop on Circuit Theory, Signal Processing and Applications*. Herbertov, Czech Republic. pp. 73–76.
- [C5] Olaszy G., Németh G., **Olaszi P.** (2001). Automatic Prosody Generation – A Model for Hungarian. In.: *Eurospeech 2001*. Vol. 1., Aalborg, Denmark. pp. 525–528.

6 Nem publikációértékű munkák

6.1. Cikk szerkesztett könyvben

- [B5] Olaszy G., Kálmán Zs., Olaszi P. (1994). A Blissvox – beszélő kommunikációs rendszer. In: Gósy Mária (Ed.): *Beszéd kutatás '94*, MTA Nyelvtudományi Intézete, Budapest. pp. 228–236. (A [C1] konferencia-cikk magyar fordítása.)

6.2. Csak kivonatban megjelent konferencia-előadás

- [A1] **Olaszi P.** (1998). Syntactic Analysis of Hungarian Sentences to Predict Prosodic Information for Speech Synthesis. In: *Conference of Ph.D. Students in Computer Science*. Konzulensek: Koutny I., Olaszy G., Németh G., "Best Talk of the Session" award. JATE, Szeged, Hungary. <http://www.inf.u-szeged.hu/~cscs/cscs.ps.gz> pp. 81.
- [A2] **Olaszi P.** (1998). Syntactic Analysis of Simple Sentences to Predict Prosodic Information for Speech Synthesis. In: *Doximp 1998 – 3rd Doctoral Symposium*. Konzulensek: Koutny I., Olaszy G., Németh G. Eötvös Lóránd Tudományegyetem, Budapest.

6.3. Csak szóban elhangzott előadás

- [P1] **Olaszi P.** (1995.07.14). The Multivox English TTS Presentation. *Zártkörű tanszéki előadás a Nemzetközi Tolmácsszövetség képviselői előtt*. BME TTT Beszédtechnológiai Laboratórium. `tts_pres.arj`
- [P2] **Olaszi P.** (1997.06.12). Syntactic parsing of Hungarian sentences for presenting prosody information for speech synthesis. *Zártkörű előadás az Ericsson svéd képviselői előtt*. BME TTT Beszédtechnológiai Laboratórium. `ericsson.doc`
- [P3] **Olaszi P.** (2001.02.19). Speech database size reduction by vector quantization. *Zártkörű előadás*. Nippon Telegraph and Telephone Corporation, Cyber Space Laboratories. Yokosuka, Japan. `ntt_final_presentation.doc`

6.4. Könyvtárakban el nem helyezett kutatási jelentés, elektronikus feljegyzés

- [R1] **Olaszi P.** (1997). Sentence prosody generation for speech synthesis by means of syntactic analysis – a Case Study for Hungarian. *Kutatási jelentés*. Faculdade de Engenharia da Universidade do Porto, Portugal.
- [R2] **Olaszi P.** (1998.06.14). Javaslat a Multivox 5-ös formánsszintetizátor alapú változatának architektúrájára. *Belső elektronikus feljegyzés*. BME TTT Beszédtechnológiai Laboratórium. mv5arch.txt
- [R3] **Olaszi P.** (1999.11.04). A Multivox beszéd szintetizátor 5-ös változatának architektúrája. *Belső elektronikus feljegyzés*. BME TTT Beszédtechnológiai Laboratórium. mv5arch2.txt
- [R4] **Olaszi P.** (1998.06.08). Javaslat a Multivox 5-ös változat API függvényekre és a bemeneti szövegben elhelyezhető vezérlő szekvenciákra. *Belső elektronikus feljegyzés*. BME TTT Beszédtechnológiai Laboratórium. mv5api.txt
- [R5] **Olaszi P.** (1998.12.16). Javaslat a Multivox 5-ös változat API függvényeire. *Belső elektronikus feljegyzés*. BME TTT Beszédtechnológiai Laboratórium. mv5api2.txt
- [R6] **Olaszi P.** (2000.07.25). Javaslat a Multivox beszéd szintetizátor bemeneti szövegében megadható vezérlő szekvenciákra. *Belső elektronikus feljegyzés*. BME TTT Beszédtechnológiai Laboratórium. control.txt
- [R7] **Olaszi P.** (1999.05.04). Egyszerű prozódiai szabályok a Multivox rendszerhez Olasz Gábor útmutatása alapján. *Belső elektronikus feljegyzés*. BME TTT Beszédtechnológiai Laboratórium. pr_szab.txt
- [R8] **Olaszi P.** (2000.05.16). Graféma–graféma átalakítási szótár kezelése. *Belső elektronikus feljegyzés*. BME TTT Beszédtechnológiai Laboratórium. gtg_dic.txt
- [R9] **Olaszi P.** (1998.06.17–2000.07.06). Language independent sound properties (Nyelvfüggetlen beszédhang-tulajdonságok). *Multivox 5 forrásfájl*. BME TTT Beszédtechnológiai Laboratórium. sound.h
- [R10] **Olaszi P.** (1998.06.14). Javaslat a Multivox 5-ös változatában a beszédhangok ábrázolására. *Belső elektronikus feljegyzés*. BME TTT Beszédtechnológiai Laboratórium. mv5hang.txt

- [R11] **Olaszi P.** (2000.04.27). Graféma–hang átalakítási szabályok magyar nyelvre. *Belső elektronikus feljegyzés*. BME TTT Beszédtechnológiai Laboratórium. gts_hu.txt
- [R12] **Olaszi P.** (2000.03.03). Fonetikus ábécé a magyar beszédhangok ábrázolására. *Belső elektronikus feljegyzés*. BME TTT Beszédtechnológiai Laboratórium. hangkod.txt
- [R13] **Olaszi P.** (1999.06.02). Hang–hang átalakítási szabályok magyar nyelvre. *Belső elektronikus feljegyzés*. BME TTT Beszédtechnológiai Laboratórium. sts_hu.txt
- [R14] **Olaszi P.** (1999.03.10, 1999.09.27). A Multivox beszéd szintetizátor hullámforma alapú diádós hangadatbázisának formai és tartalmi vizsgálata. *Belső elektronikus feljegyzés*. BME TTT Beszédtechnológiai Laboratórium. dbase1.txt
- [R15] **Olaszi P.** (2000.03.29). A Multivox beszéd szintetizátor 2-es változatú (hullámforma alapú, triádokat vagy több hangot tartalmazó) hangadatbázisának formai és tartalmi vizsgálata. *Belső elektronikus feljegyzés*. BME TTT Beszédtechnológiai Laboratórium. dbase2.txt
- [R16] **Olaszi P.** (2000.03.29). A Multivox beszéd szintetizátor 6-os változatától kezdve alkalmazott belső hangadatbázis ábrázolás. *Belső elektronikus feljegyzés*. BME TTT Beszédtechnológiai Laboratórium. dbase.txt
- [R17] **Olaszi P.** (1999.12.12). A hangadatbázist kezelő függvények változásáról. *Belső elektronikus feljegyzés*. BME TTT Beszédtechnológiai Laboratórium. db_diff.txt
- [R18] **Olaszi P.** (1999.04.28–2000.05.09). A számok magyar szöveggé alakításáról. *Belső elektronikus feljegyzés*. BME TTT Beszédtechnológiai Laboratórium. szamok.txt
- [R19] **Olaszi P.** (1998.10.22). Prozódiai előrejelző a Profivox beszédvel való szoló rendszerhez. *Kutatási zárójelentés*. BME TTT Beszédtechnológiai Laboratórium. zj981022.doc
- [R20] **Olaszi P.** (2001.02.21). Speech database size reduction by vector quantization. *Belső projekt zárójelentés*. Nippon Telegraph and Telephone Corporation, Cyber Space Laboratories. Yokosuka, Japan. ntt_final_report.doc

6.5. Tudományos Diákköri dolgozat

- [S1] **Olaszi P.** (1994). BLISSVOX – Beszélő kommunikációs rendszer rehabilitációhoz és oktatáshoz. *BME Villamosmérnöki és Informatikai Kar, Tudományos Diákköri Konferencia*, III. díj. Konzulensek: Olaszy G., Kálmán Zs.
- [S2] **Olaszi P.** (1995). Kötött szótáras mondatgenerátor magyar nyelvre; illesztése a Blissvox kommunikációs és rehabilitációs programhoz. *BME Villamosmérnöki és Informatikai Kar, Tudományos Diákköri Konferencia*, II. díj. Konzulens: Koutny I.

6.6. Diplomaterv-dolgozat

- [T1] **Olaszi P.** (1997). Algoritmusok az egyszerű magyar mondat szerkezetének meghatározásához prozódiai szerkezetek előrejelzése céljából. *Diplomaterv dolgozat*. BME Villamosmérnöki és Informatikai Kar.