

Németh Géza – Olaszy Gábor

**A magyar beszéd. Beszédkutatás, beszédtechnológia, beszédinformációs rendszerek.**

(Akadémiai Kiadó: Budapest, 2010. 708 pp.)

*Bodnár Ildikó*

Az a hétszáz oldalnál is terjedelmesebb könyv, amelyről az alábbiakban szólni szeretnék, 2010-ben jelent meg az Akadémiai Kiadónál, de maga a munka a Budapesti Műszaki és Gazdaságtudományi Egyetem Távközlési és Médiainformatikai Tanszékén született meg. Írói gárdáját a két szerkesztő, Németh Géza és Olaszy Gábor mellett további tizenhét személy alkotja; Gordos Géza áttekintő szerkesztője lett a kiadványnak. Az alkotók valamilyen módon a beszédkutatás kiváló szakemberei, zömükben a Budapesti Műszaki és Gazdaságtudományi Egyetem munkatársai, s a legfiatalabbakat kivéve magas tudományos fokozattal rendelkező személyek.

Ám a BME mellett jó néhány más intézmény munkatársait is megtaláljuk a szerzők között. Így Abari Kálmán a Debreceni Egyetem Pszichológiai Intézetét, Czap László a Miskolci Egyetem Automatizálási Tanszékét, Olaszy Gábor az MTA Nyelvtudományi Intézetét, Takács György a Pázmány Péter Katolikus Egyetem Információs Technológiai Karát képviseli. Campbell Nick beszédkutató több amerikai kutatóhely (IBM, Bell Laboratories) után Japánban végez többek között a nemverbális beszédkommunikációra, a beszédadatbázisok készítésére, ill. a prozódiai modellezésre vonatkozó kutatásokat. Az egyetlen női munkatárs a szerzők között Vicsi Klára akusztikus, aki az MTA doktora, s aki megteremtője volt a beszédfelismerési munkák alapját képező magyar nyelvű beszédadatbázisoknak. A szerzők munkásságának tömören megfogalmazott – s fényképpel is illusztrált – bemutatása rögtön az *Előszó* után olvasható; ugyanitt megtaláljuk azt az áttekintő táblázatot is, amelyből megtudhatjuk, hogy a tizennyolc közreműködő közül ki, melyik fejezeteknek volt a szerzője, illetve társszerzője.

A szerzők legnagyobb része a műszaki indíttatású beszédkutatást képviseli, tehát például híradástechnikai szakmérnök, akusztikus, műszaki informatikus, de van a könyv írói között programtervező matematikus és több villamosmérnök is. Villamosmérnökként végzett a kötetet szerkesztő fonetikus nyelvész is, aki egyik kifejlesztője lett a ProfiVox magyar szövegfelolvasó szoftvernek.

Egyértelműen állítható, hogy igen magas felkészültségű gárda közös munkájaként készült el a huszonegyedik század magyar tudományosságát a beszédkutatás terén méltán reprezentáló, igényes tartalmú és színvonalas kiállítású könyv. Szinte mindegyik fejezetét – a témától függően – jól értelmezhető, a jelenségek lényegét kiemelő rajzok, diagramok, számítógépes ábrák, jól követhető spektrogramok, dallamképletek, táblázatok vagy éppen képletek tucatjai egészítik ki. S akkor még nem szóltunk azokról az anyagokról, amelyeket *A magyar beszéd* könyv honlapján találunk meg: <http://magyarbeszed.tmit.bme.hu>.

Olaszy Gábor alább idézett megfogalmazása egyértelműen jelzi a sokféle kutató együttes munkájának szükségességét: „A hang fizikai terjedésével és az azt befolyásoló környezeti tényezőkkel az akusztika foglalkozik. Ennek egyik ága a pszichoakusztika, amely a fizikai hanginger és a szubjektív hangészlelés közötti kapcsolatokat tárja fel. A beszédjel fizikai feldolgozásával, átvitelével kapcsolatos problémakör műveléséhez mate-

matikai, villamosmérnöki, informatikai és digitális jelfeldolgozási ismeretekre van szükség. A természetes beszédlánc (emberi dialógus) egyes elemeinek gépi megvalósítása a beszédtechnológia tudományához kapcsolódik (gépi beszédszintézis, -felismerés, -megértés, -azonosítás, -módosítás, beszéd-tömörítés). Az átviteli közeg térbeli és időbeli kiterjesztésével (telefonálás, hangfelvétel készítése és lejátszása) a híradástechnika foglalkozik” (10).

A kezdetektől jól lehetett követni a könyv azon irányultságát, hogy lett légyen szó bármilyen egyszerű vagy bonyolult elméleti kérdésről, ezek tárgyalása során a szerzők a kitűzött gyakorlati célt: a beszédtechnológiai, beszédinformatikai rendszerek megvalósítását tartották szem előtt, azaz elméleti írásaik a létrejött gyakorlati megoldásokra irányulnak, így azokra gyakran utalnak is előre, valamint a későbbi, gyakorlat-orientált fejezetek köréből ugyancsak többször visszautalást találunk az elméleti részekhez. Példaként említem a beszéd és az írás kapcsolatát bemutató fejezetet, ahol az ortografikus és a fonemikus alak mellett nem véletlenül jelenik meg az ún. fonetikai átírat. Ennek E2-hangjelekkel írott változata érthető a számítógép számára (78–83). A téma négyszáz oldallal később, a kötetnek az automatikus szövegfelolvasással foglalkozó részeként tér vissza (472–488).

A munka négy nagy témakörre tagolódik: (1) *Ember, beszéd, nyelv* (3–92); (2) *A beszéd szerkezeti elemzése* (95–205); (3) *Beszédtechnológia* (209–522) és végül (4) *Beszédtechnológiai alkalmazások* (525–651) címmel, melyeket egy mintegy harmincoldalas *Irodalomjegyzék* és egy rövid *Függelék* követ, valamint nem hiányzik a kötet legvégéről a könyvben való eligazodást segítő *Tárgymutató* sem. A felsorolt négy nagy rész összesen 14 fejezetre oszlik, ezek folyamatosan számozódnak, vagyis a második témakör az ötödik fejezettel, a negyedik pedig a tizenegyedik fejezettel indul.

Egy-egy külön fejezetet kapott az első részben *A beszéd és az információs társadalom* és *A beszéd komplex szerkezete*. A harmadik, *Fiziológiai, fizikai alapok* című fejezeten belül egy-egy alfejezet képviseli mind a beszédképzési, mind a hallási folyamatnak, mind pedig a beszéd fizikai jellemzésének a bemutatását. Negyedik fejezetként *A beszéd és az írás* címen olvashatunk több, a tárgykörbe tartozó alfejezetet, pl. *Írásrendszerek, Hangjelölés*, ill. *Az írott szöveg és a hangalak kapcsolata* címen, ill. megismerhetünk a magyar nyelvre vonatkozó hang-, betű- és szóstatisztikai adatokat is.

A második rész első, ám a teljes munka ötödik fejezeteként olvasható része *A beszéd szegmentális szerkezete* címet viseli, és a magyar beszédhangok igen részletes, sokoldalú jellemzését adja. Bár a könyv céljának az akusztikai megközelítés felel meg a legjobban, nem hanyagolják el a szerzők az artikuláció leírását sem. Mindez már a fejezet legelején, az artikuláció akusztikai vetülete részben magyarázatot nyer: „Az akusztikai vetület egyfajta kapcsolatot teremt az artikulációs csatorna pillanatnyi térbeli formációja és a keletkezett hang között. Ha változnak az artikulációs csatorna fizikai méretei, változik az akusztikai vetület is” (95). A beszéd mesterséges előállításának igénye olyan elemek vizsgálatát követeli meg, mint a hangok és hangkapcsolatok időtartama, ill. a magánhangzó-magánhangzó kapcsolódások (beleértve a hiátustöltés jelenségét is), a különféle két-, három-, sőt négyelemű mássalhangzó-kapcsolatok, vagy a mássalhangzó-magánhangzó-mássalhangzó kapcsolódások területe. Az ezekre vonatkozó mérési adatokat részletes táblázatok, kör- és oszlopos diagramok mutatják be, s rendkívül szemléletesek a felvonultatott, a beszéd változó hangszínképét élénk állító spektrogramok, ill. az azokról készült spektrografikus rajzok is. A gyakorlati megvalósítás igénye miatt kerülhetett sor a zöngéesség – zöngétlenség vonatkozásában a *gerjesztési forma* szakszó használata, ill. az akusztikai szempontból *egyszerű és összetett szerkezetű hangok* elkülönítése. A fentiek egy rövidebb, a *suttogásról*

szóló alfejezet követi, melynek egyik tanulsága: „...A suttogott beszéd gépi felismerése a normál beszédre kidolgozott módszerekkel nem végezhető el” (167).

A hatodik, *A beszéd szupraszegmentális szerkezete* című fejezet a szokásos módon tagolódik a beszéddallamokat, a hangsúlyozást, a beszédtempót, a szüneteket, a ritmust, sőt a hangszínezetet bemutató részekre. Itt is jellemző az időszerkezeti tényezők, valamint a kapcsolódások igen erős hangsúlyozása, továbbá a minél hűbb ábrázolás kérdésének a vizsgálata. A hangsúly-meghatározás és -kijelölés, a beszéddallam és az időtartam kérdései ugyancsak részletesen előkerülnek a már említett későbbi, az *Automatikus szövegfelolvasás* címet viselő fejezetben, a 429–458. oldalakon. De a prozódia kérdéseivel találkozunk az ún. *akusztikai arculatot* vizsgáló alfejezetekben is (532–539).

A könyv fő kérdésköreit tárgyaló harmadik, *Beszédtechnológia* című rész a legterjedelmesebb, ennek oldalszáma a háromszáz oldalt is meghaladja. Hasonlóan az első témakörhöz, ez is négy nagy fejezetre tagolódik, amelyek rendre *A beszédtechnológia tudománya*, az *Adatbázisok a beszédtechnológia szolgálatában*, *A beszéd gépi észlelése és felmérése*, ill. *A beszéd gépi előállítása* címet viselik.

A *beszédtechnológia* szó benne van a munka alcímében, szó esik róla már az *Előszó* oldalain is, ám a hetedik fejezet elején szabatos meghatározását is megismerhetjük: „A beszédtechnológia az a tudomány, amelyik az emberi beszédtevékenység körfolyamatából valamely komponens(ek) modellezésével és gépi megvalósításával foglalkozik.”

Továbbá megtudjuk, hogy: „A beszédtechnológia lényeges szerepet kap az infokommunikációs rendszerekben, egyrészt automatizált gépi szolgáltatásokat tesz lehetővé (automatikus tudakozó), másrészt kényelmi szolgáltatásokat is nyújthat (beszédalapú tárcsázás, sms-felolvasás, sms-diktálás)” (209). Később részletesen szó esik a beszédtechnológiának orvosi célú, ill. a fogyatékkal élők számára jelentős segítséget nyújtó oldalairól is.

Ez a *beszédtechnológia tudománya* című hetedik fejezet olyan szakkifejezések (kvantálás, Fourier-sor, Fourier-transzformáció, kepsztrum, ablakoló függvények stb.), továbbá képletek sorával van tele, amelyek a szűkebb tudói kör számára hozzáférhetők; magam a legnagyobb részt Vicsi Klára, ill. Olasz Gábor által írott nyolcadik, az *Adatbázisok a beszédtechnológia szolgálatában* fejezet részeit elérve tudtam ismét bekapcsolódni a könyv elmélyedt olvasásába. Itt rendkívül érdekes alfejezetek sora mutatta be a beszédadatbázisok készítéséhez kapcsolódó feladatoknak, illetve maguknak a beszédadatbázisoknak a sokféleségét, a betanítás feladatát, a szám- és szövegfelolvasás megoldandó, s napjaink kutatásainak köszönhetően már nagyrészt meg is oldott tenger problémáját. (Az más kérdés, hogy a felhasználók hanyagságból – az akusztikai arculat iránti közömbösségüktől, mint később ez is kiderült –, továbbá takarékosági okokból csak ritkán veszik igénybe a jó minőségű, az élő beszédet nagymértékben megközelítő programokat. (Az érdeklődők ilyen hangsorok, beszélő programok meghallgatásáig is eljuthatnak a már közölt honlapon kutatva, az *Interaktív anyagok* menüpontra kattintva, a *Profivox magyar beszéd szintetizátor bemutatása hangdemonstrációkkal* címnél.)

Számomra nehezebben, ritkán könnyebben érthető részek váltakoztak *A beszéd gépi észlelése és felmérése* című következő fejezetben, melynek utolsó részében találkoztam az *Érzelemfelismerés* című, ugyancsak Vicsi Klára által írott alfejezettel, valamint a Czaplász László készítette *Beszéd felismerés támogatása multimodális paraméterekkel* alfejezettel. Ez utóbbinak fontos gondolatát jelentik a 402. oldalon olvasható sorok: „A vizuális modalitás előnyei az emberi beszéd felismerésben elsősorban három területen mutatkoznak meg: segíti

a hangforrás, a beszélő helyének meghatározását, megkönnyíti az akusztikai jel szegmentálását, kiegészítő információval szolgál az artikuláció helyének meghatározásához”. Fontos szerepet szánunk a szerzők a képi ábrázolásnak, az ún. beszélő fejnek: „... A modalitásokat egymás kiegészítésére használjuk. Ha a hang gyenge minőségű, vagy hallássérült a megfigyelő, jobban hagyatkozik a szájról olvasásra.” Maga a beszélő fej, annak modellezése a következő nagy fejezetben jelenik meg, s megtudjuk azt is, mi minden szükséges az élethű beszédmodellezéshez: a) száj- és nyelvmozgás beszéd közben; b) fejmozgás; c) szem- és szemöldökmozgás; d) érzelmek kifejezése bizonyos gesztikulálással” (459). Az egyes hangokhoz ezen a szinten ún. vizéma-készlet tartozik, amelyben az egyes hangok az ajakszélesség, az ajaknyílás és a szájnyílás világossága alapján különülnek el egymástól. A szájról olvasás egyébként azért igen nehéz, mert a hangok többségének vizéma-készlete egybeesik egy vagy több másik hangéval. Vagyis míg beszédhangjaink (fonémáink) száma 39, a vizémáké mindössze 15. Azonos többek között az *o* - *ö*, *u* - *ü* magánhangzópár, továbbá a *b* - *p* - *m*, *z* - *sz* - *dz* - *c* stb. mássalhangzók szájról leolvasható képe. Az ún. transzparens megjelenítés a nyelvállásra ugyan utal, de a zöngéesség adta különbség így sem érzékeltethető. A kísérletben született modellekre ugyancsak rá lehet keresni a megadott honlap alapján, közvetlenül pedig az alábbi módon érhető el az adott oldal: <http://mazzola.iit.unimiskolc.hu/~czap/mintak>.

Eljutva a tizedik – *A beszéd gépi előállítása* című – fejezethez, azt tapasztalhatjuk, hogy ebben gyakorlatilag a közelmúlt és napjaink valósága jelenik meg. Mégis történeti bevezetővel indít a fejezet első részét szerző Olasz Gábor, mert – mint írja – „A gépi beszéd-előállításra vonatkozó kísérletek már több mint 200 évre nyúlnak vissza” (412). Kempelen Farkas volt az a 18. századi magyar fonetikus, aki korát messze meghaladóan nemcsak leírta a beszédhangok sokaságát, hanem beszélőgéppel hangokat, szavakat, sőt egy-két rövid mondatot is előállított. Leírást is olvashatunk Kempelen Farkas gépéről, s azt is megtudjuk, hogy a múltban – sőt a jelenben is – két megoldási stratégia áll egymás mellett a mesterséges beszéd előállítását illetően. Az első szerint az alap „az emberi hangképzés és artikuláció utánzása” (ilyen volt például a Kempelen-féle beszélőgép, s ilyen napjainkban a legtöbb robot), a második szerint pedig „az akusztikai produktum utánzása” történik (412–413).

A huszadik század első felében már tudtak géppel folyamatos beszédet előállítani, így 1939-ben a Bell Laboratórium gépével, bár ennek hangja még nehezen volt csak érthető. A nagy áttörést a számítógép alkalmazása jelentette ezen a területen (is). A 422. oldal táblázata tartalmazza a Magyarországon 1980 és 2008 között előállított szövegfelolvasó rendszereket, beszéd szintetizátorokat, utalva többek között a megvalósítás évére, a beszédkeltő készülékre, a hangadás alapjára, a felolvasó szoftverre, a hangminőségre, nem utolsósorban pedig a kifejlesztőkre. Magyarul beszélő, ingyenes szövegfelolvasó szoftver tölthető le a <http://speechlab.tmit.bme.hu> címen.

A fejezetnek a *Beszélő fej modellezése* című részében ismét Czapl László szól eredményeiről: „A 3D rendszerek életszerű, fotorealistikus finomságú modellek kidolgozását teszik lehetővé... Hallássérültek beszélni tanításában segíthet a helyesen artikuláló virtuális bemondó, amely esetlegesen átlátszóvá tett arcával a természetes beszélőnél jobban megmutathatja a hangképzés nem látható részleteit is” (458).

A különféle alkalmazott nyelvtudományi ágazatok – így az alkalmazott fonetika, beszédtechnológia – társadalmi hasznossága nagymértékben megmutatkozik ezekben a fejezetrészekben. A hallássérült gyerekek a beszélő fej, továbbá a *Beszédtanítás és beszéd-*

*technológia* alfejezetben megismerhető különféle játékos modellek révén juthatnak mind közelebb a helyes beszédhez. Az adott rész szerzője Vicsi Klára; modelljét számos iskolában használják, ahol hallássérült gyerekekkel foglalkoznak. Az automatikusan felolvasott szövegek pedig a látásukat veszített embereknek jelentenek igen nagy segítséget.

A tizedik fejezet számos alfejezete bizonyítja, milyen körültekintő munkára volt szükség a mai eredményekhez való eljutásig, ld. pl. a: *Hangminőségi skála*, a *Hangsúly-meghatározás a szöveg alapján*, a *Fonetikai átíró magyar nyelvre*; az *Ékezetek gépi helyre-állítása*; az *Ékezetesítő eljárások* stb. című fejezeteket, illetve azokat a részeket, amelyekben a beszéddallam megvalósításáról, a gépi beszéd időszerkezetének az előállításáról, vagy éppen az érzelmes szövegfelolvasásról van szó.

Azok az eredmények, amelyekről a *Beszédtechnológiai alkalmazások* összefoglaló címet viselő negyedik rész számol be, mindennapjaink részét képezik. Hiszen itt tárgyalja a könyv pl. a telefonon bonyolított ember-gép dialógusok kérdéskörét, pl. a menetrend-lekérdezés esetét egy-egy konkrét utazásra vonatkozóan, avagy a telefonról elérhető e-levél felolvasását, a mobiltelefonba épített SMS-felolvasókat, az automatikus szám szerinti tudakozókat, a gyógyszervonalakat, a különféle közlekedési alkalmazásokat, mint pl. a vasútállomási utastájékoztatókat.

De idesorolhatók a vakok és a gyengénlátók számára megvalósuló szolgáltatások, így pl. a képernyőolvasás, az ún. hangoskönyvek rendszere, a beszélő bankautomaták, továbbá az egyik legújabb fejlesztés, az ún. NaviSpeech – a beszélő navigátor – a látássérült gyalogosoknak. A siketek kommunikációjának segítésére szánt mozgó szájat, ill. a beszédtanító programot is ebben a fejezetben írják le, mutatják be részletesen a szerzők. Az előbbi esetben fontos, hogy felolvasóknak képzett jeltolmácsokat választottak a program kidolgozói, mivel ők a legjobban tudnak alkalmazkodni – mind a pontos artikuláció, mind a megfelelő beszédtempó vonatkozásában – siket és nagyothalló felhasználók igényeihez.

Bemutatja a fejezet még a szintetikus beszéddel történő hallásfelmérést, így a már 1984 óta igen sokszor alkalmazott, s többek között Gósy Mária által kidolgozott magyar szabadalmat, az ún. GOH mérési rendszer elveit és gyakorlatát is.

A két utolsó fejezet címe önmagáért beszél: a 13. fejezeté ugyanis ez: *Interfészek, szabványok, honlapok, programok* (631–651) és a 14. fejezeté pedig: *A beszédtechnológia jövője* (653–655) – ez utóbbi valóban csak néhány oldalba sűrítve – a beszéd kutatás jövőjéről szól.

A recenzens feladatai közé tartozik, hogy megmondja, kinek ajánlható a könyv. De ezt már maguk a szerzők megtették, amikor a bevezetőben kijelölték célközönségüket, amelyet egyébként igen széles körűen jelöltek ki – a komoly kutatóktól a kutatásokba majd csak a jövőben bekapcsolódó középiskolás diákokig. Az alábbiakban – zárásként – szívesen idézek e bevezetőből: „Célközönségünk az egyetemek, főiskolák, valamint minden olyan oktatási hely, ahol informatikusokat képeznek. A könyv jó támogatást adhat távközlési fejlesztőknek és döntéshozóknak, a beszédtechnológiai fejlesztések szakembereinek, új tartalomszolgáltatási, egészségipari és rehabilitációs szolgáltatások tervezőinek... Segítheti a humán területek oktatását is (fonetika, beszédelemzés, nyelvészet és a beszéd kapcsolata, beszédpszichológia, egészségügyi betegségmegelőzés és rehabilitáció, tájékoztatás). Ajánlhatjuk továbbá a középiskolások felső tagozatának is, valamint mindenkinek, akit érdekel a témakör (például fizikusok, nyelvészek, rádiósok, televíziósok, filmesek, tudományos média szakemberei). Átfogó tartalma miatt hasznos információkat találnak benne a fenti szakmák művelői, a mérnököktől a bölcsészekig.” (Előszó, xvi).

Mint bölcsész végzettségű olvasó elmondhatom, hogy a munka legnagyobb része – a sokféle számadat, a gyakori műszaki leírások ellenére – bölcsész értelmem számára is jól megközelíthetőnek bizonyult, számos fejezete kifejezetten élvezetes olvasmányt jelentett. A beszédtechnológia résztől feltűnő számos matematikai képlettel viszont már nem igazán tudtam mit kezdeni, de ez semmiképpen nem a szerzők hibája.

"A bemutatott kutató munka a TÁMOP-4.2.1.B-10/2/KONV-2010-0001 jelű projekt részeként az Európai Unió támogatásával, az Európai Szociális Alap társfinanszírozásával valósul meg"

"This research was carried out as part of the TAMOP-4.2.1.B-10/2/KONV-2010-0001 project with support by the European Union, co-financed by the European Social Fund."