

**Peter B. Denes - Elliot N. Pinson: The Speech
Chain. The Physics and Biology of Spoken
Language**

2nd edition W.H., Freeman and Co. New York, N.Y. 1993.

Tarnóczy Tamás

A 30 évvel korábban megjelent első kiadás címe, sőt alcíme is azonos a maival. A két szerző annak idején a Bell Telephone Laboratories megbízásából - és részben műszaki propagandájaként - foglalta össze a szakmai ismereteket. Azóta azonban nagyot fejlődött a tudomány. Az újnnyomásnak tehát nem látván értelmét, a szerzők az alapos átdolgozást választották.

Peter B. Denes, azaz Dénes Péter, budapesti érettségi után Manchesterben szerzett diplomát, és Londonban doktorált. Fizikusként a University Collegeban Fry professzor mellett is dolgozott, de elsősorban a Fülészeti Klinikán szerzett kitűnő gyakorlati hallástani ismereteket. Fő kutatási területe később az önműködő beszédfélismerés és a mesterséges beszédösszetétel lett. Elliot N. Pinson fiatalabb kutató. Az ötvenes években Princentonban, majd az M.I.T.-ben végül Californiában szerezte tudományos fokozatait, majd ő is a Bell Laborhoz csatlakozott. Elsősorban az alkalmazott vezérlő rendszerek szakértője.

Mindketten 1961-ben léptek be a Laboratórium munkatársai közé, olyan elődök után, mint Harvey Fletcher, W. A. Munson, Homer Dudley, J.C.R. Licklider, vagy Ralph Potter. Ezt az örökséget vitték tovább munkájuk során.

A szerzők új könyve ezúttal meghökkentően modern grafikájú címlappal, önálló kiadásban, a régihez képest mintegy másfélszeres terjedelemben jelent meg, és foglalja össze a témára vonatkozó korszerű tudományos és gyakorlati ismereteket. A tárgyalás menete megmaradt, természetesen kiegészítve az újabb eredményekkel. Ennek folytán több olyan

téma is földolgozásra került, amely éppen az utolsó 30 év alatt fejlődött jelentőssé. Ezek az ismeretek nagyrészt az utolsó 3-4 fejezetben vannak összefoglalva, és a beszédmegértés, a digitális jelfeldolgozás, a beszéd-szintézis és a gépi beszéd-fölismerés témakörére vonatkoznak.

Általában értékelve az anyagot, ki kell emelnünk a fölépítését. A fejezetek váltakozva, fokozatosan mind mélyebb szinten tárgyalják a beszédképzés és a hallás, a vizsgálati eljárások és az agyi megismerési folyamatok ismeretkörét. A szerzők mindig igyekeznek az ellenőrzött és elfogadott adatokat közölni; nem merülnek mély fejtegetésekbe és nem alkalmaznak matematikai eszközöket. Rokonszenvesek a fejezetek bevezetőiben közölt rövid összefoglalások, valamint az a körülmény, hogy a szerzők a sokoldalú és széles jelenségcsoport összefüggéseit folyamatosan összekapcsolják egymással.

A könyv stílusa ahhoz a didaktikus alapelvhez illeszkedik, hogy a nem közvetlenül szakember vagy a kezdő kutató jó és megbízható áttekintést kapjon az időszerű tudományos állapotról, és szükség esetén később el tudjon mélyedni a témában. Az anyagnak tehát tanfolyam jellege van, amit szakszerűsége mellett világos és könnyen érthető stílusa különösen alkalmassá tesz ilyen célú fölhasználásra. A fejezetek végén közölt irodalmi anyag azonban kizárólag amerikai, ebben nálunk pótlására is szükség lehet. A tanfolyami célra utal az is, hogy a Kiadó által küldött kísérő szövegben többek közt ez olvasható: Reméljük, hogy a könyv anyagának fölhasználását figyelembe fogja venni saját kurzusában.

Ennek a gondolatnak a jegyében foglalkozunk a könyv hazai ismertetésével. Mivel pedig a 30 évvel ezelőtt megjelent első kiadás is rendelkezésünkre áll, érdekesnek tartjuk a fontosabb változtatások figyelemmel kísérését is.

*

*

*

Az egyes fejezetek rövid tartalma és terjedelme a következő:

1. A beszédlánc (1-10). Áttekintő összefoglalás a beszélő agyától a megfigyelő agyáig terjedő teljes pszichológiai, fiziológiai, anatómiai és akusztikai folyamatról, miközben a nyelv és a beszéd segítségével gondolatokat közlünk egymással. A szerzők nem feledkeznek meg a "nemakusztikus" kísérő jelenségek szerepéről sem. A bevezetés azért is érdekes, mert a régi kiadás szövegéhez semmit sem kellett hozzátenni, vagyis az eredeti szemléleten 30 év múlva sem kellett változtatni.

2. Nyelvi szervezettség (11-16). Rövid áttekintés a nyelvészeti alapfogalmakról a fonetikától a mondatszerkezetig, a nyelvtani szabályoktól a morfológiai és szemantikai szintig. Mindössze olyan fokon, ami a továbbiak megértéséhez okvetlenül szükséges. Itt is, mint a későbbi nyelvi vonatkozások során, a szerzők ragaszkodnak az amerikai-angol ismeretekhez és szabályokhoz, amit a 2.1 táblázatban közölt fonetikai rendszer is jellemez.

3. A hang fizikája(17-46). Ez a fejezet is alapismeretekkel foglalkozik, de részletesebb tárgyalással vezet be a fizikai tudnivalók elsajátításába. A rezgéstani alapismeretek terén az új szövegben új részleteket is találunk. Ilyen mindjárt a 3.2 ábra a rezgőmozgás lineáris átírására, ami egyúttal alkalmat ad a szerzőknek arra, hogy a régebbi c/s helyett bevezessék a szabványos Hz használatát. Megjegyzendő, hogy ugyanakkor még tekintettel az amerikai olvasóra, megmaradnak az inch használata mellett. Az új kiadásban bevezetik a tisztahang fogalmát; a rezonanciák és a Fourier-színkép fogalma változatlan. Részletesebben van kifejtve a decibel és a vele kapcsolatos összefüggések.

4. Beszédképzés (47-78). A fejezet anyaga már a szoros értelemben a témához tartozó ismertetés a beszédképző szervek szerepéről és működéséről, valamint a beszédhangok képzési sajátságairól. Elsősorban a hangszalag rezgéseivel és a felső üregek módosító hatásával létrehozott magánhangzókkal foglalkoznak, de a turbulens légáramlással létrehozott réshangok és a zárlatok megnyitásával keltett zárhangok elemzését is ismertetik. Az afrikai csettintett hangokat is megemlítik. A lényeges megjegyzések kiemelten mindig az angol fonetikára vonatkoznak.

Részletes - ábrákkal támogatott - anatómiai leírás és működési ismertetés található a larynx és a glottis szerepéről. A vizsgálati módszerek közül a gégetükrös, gyorsított filmfelvételi eljárást részletezik, a régebben szokásos periódusos megvilágítású sztroboszkópos módszertől eltekintenek. Egyik fontos gyakorlati megjegyzésük, hogy az erősen hangos hangadáskor a hangszalagok nyitása és zárása meredekebb, és a rés a periódusnak rövid részidejében van nyitva. Ezért ilyenkor a magasabb felhangok erőteljesebbek, a hang harsány és éles lesz. Ez a színképi forma akkor is a kiabálás jellegére utal, ha egyébként magát a hang intenzitását lecsökkentjük.

Részletesen foglalkoznak a felső üregek és az ajaknyílás hangformáló szerepével, valamint a nyelv és a lágyszájpad alakító hatásával. Érdekes, hogy a nazális képzésben szereplő nyelvcsapot nem mint külön anatómiai képletet, hanem mint a lágyszájpad egy részét határozzák meg. Ennyi előkészítés után bemutatják az úgynevezett kardinális magánhangzók nyelvállástól függő elhelyezkedését. Ebben a régi kiadáshoz képest némi jelölési korrekcióra kerül sor. Mindent egybevetve, az új kiadás fonetikai szempontjából előnyére változott. Érdekes észrevétel, hogy a korábbi szövegben az angolban hiányzó "ü" magyarázatához orosz példára hivatkoznak, az új kiadásban már a francia példa szerepel. Újdonság a központi semleges

magánhangzó bevezetése, az "ö" azonban nem került az angolban is kiejthető magánhangzók közé. A mássalhangzók artikulációs magyarázatában szintén több változtatást találunk az új kiadás előnyére, például az új 4.1 táblázat más rendzési elv szerinti fölosztás. Itt azonban kimaradt a "h" és nehezen érthető a "w" kettős elhelyezése. Az új "approximant" elnevezés ugyanis egy bilabiális mássalhangzó és egy veláris félmagánhangzó jellegű "w" kiejtést foglal össze. Az [r] hangból csak veláris változat található meg. A szerzők egyébként többször is hangsúlyozzák, hogy nem általános, hanem amerikai-angol fonetikát ismertetnek.

A következő részben a beszédhangok akusztikai szerkezete kerül sorra. Részletezik a hullámforma és a színekép kapcsolatát, definiálják a formánsokat, megmutatják a hangmagasság okozta színeképi módozatok formánsfüggetlenségét és egyes vokálisok ("i", "o", "u") üregformájának és formánsszerkezetének összefüggését. Magyarazatként a Fant-féle csőrezonancia elméletet fogadják el.

5. Hallás (79-110). A fül anatómiája és fiziológiája, valamint a hangfelfogás pszichológiai vonzatai kerülnek sorra. A fül szerkezetét a szokásos leírással világítják meg, jelentős terjedelemben foglalkoznak a középfül transzformáló szerepével és a belső fülhöz való illeszkedés fizikai magyarázatával. A Corti-féle szervet a szokott formában ismertetik. Békésy nevének említése nélkül tárgyalják a haladó hullámú elméletet, amely ma legelfogadottab magyarázat az alaphártya hidrodinamikai gerjesztésre. A régi kiadás 5.8 ábrája helyett az újban 5.7 helyen mutatják be az alaphártyán a 200 Hz-re vonatkozó haladó hullámú kitérések összesített burkolóját (ezúttal idézve Békésyt).

A receptorok szerepének taglalása átdolgozott anyag. Szerepel benne a sztereociliák valószínű működése, valamint az,

hogy az általuk létesített elektromos aktivitás indítja meg a hallóidegekben tovaterjedő elektrokémiai folyamatokat. Itt került be az új szövegbe az utolsó évtizedben fölismert (Kemp, 1978) és jelenleg is tisztázás alatt álló, otoakusztikus jelenségcsoport, vagyis a csigában, pontosabban a külső szőrsejteken keletkező periódusos mechanikai begerjedés, amely visszahat a dobhártyára, és így gyöngö, de mérhető hanghatást kelt. A ma szokásos kifejezéssel élve: az otoakusztikus emisszió valójában a régen ismert fülcsengésnek a hallójáratában mérhető objektív megfelelője.

A fejezet további részében a hangfelfogás hagyományos összefoglalását kapjuk. A hallásélesség, hangosság és hangmagasság, valamint ezek összefüggése a fizikailag mérhető adatokkal a téma. A két utolsó alfejezet az elfedés és a kétfülű hallás rövid ismertetése. Ezen a területen nem tartották elérkezettnek időt az 1950-es években leülepedett ismeretek bővítésére, bár a "coctail party" hatást és a sztereotechnikát röviden megemlítik.

6. Az idegek és az agy a beszédláncban (111-137).

Lényegében ebben a fejezetben indul meg a régi koncepció átforgalmazása. Ebből kapunk jelentős ízelítőt a szokásos bevezetőben, ahol a számítógépi szemlélet veszi át az uralmat. Itt említik meg először az integrált áramkörök legkorszerűbb formáját. (VLSI), valamint a legújabb vizsgálati módszereket, például a pozitron kibocsátási tomográfiát. Az új kiadás már 10 milliárd agyi neuronnal számol, bár olvashatók már 50 milliárdos becslések is. Az idegi anatómia magyarázata a régi változatnál részletesebb, az ábrák nagyobbak, világosabbak és nagyon tanulságosak. Az idegimpulzusok vezetési folyamatainak részletes ismertetésében új annak kihangsúlyozása, hogy a velőhüvelyes és Ranvier-féle befűződésekkel ellátott idegrostokban az impulzusok terjedési

sebessége elérheti a 10 cm/ms értéket. A gyakorlati terjedési sebességet természetesen csökkenti a szinapszisok (idegi összekapcsolódások) késleltető hatása.

Különösen föl kell figyelniünk egy kisebb alfejezetre, amelynek címe: a gondolattól a beszédig. A második fele különben is teljesen új, de ettől függetlenül az egész alapgondolat olyan rövid, tömör összefoglalását kapjuk, amelyhez hasonlót a nagy klasszikus könyvekben is alig találunk. A szerzők hitet tesznek amellett, hogy a beszéd sokkal több, mint komplex motoros tevékenység. Tudni kell hozzá a nyelvi kódokat, a szavaknak a tárgyakkal és fogalmakkal való kapcsolatait, és biztosítani a mondattani és jelentéstani rendszerek fölötti teljes uralmat. Mindeközben a legmagasabb agyi szinten állandó kapcsolatot kell tartani az elraktározott adatokkal és az akaratlagos aktivitással. A végeredmény egy bonyolult izommozgatási folyamat, amelynek akusztikai vetülete összevág az eredeti gondolat rögzített lenyomatával. Ennek a folyamatnak a teljes végigjárása a beszéd kutatás feladata, és ennek teljes megértésében eddig kevés felhasználható eredmény született.

Az okát is sejtjük, hogy miért. A központi idegrendszer szerkezeti (anatómiai) kutatásban sok értékes eredmény született, ugyanakkor a rendszer szervezési (funkcionális) megismerése egyelőre legyőzhetetlen akadályokba ütközik. Az eredménynek még messze kevesen vannak a végleges magyarázathoz. Hiába ismerjük a bal oldali agyfélteke Broca- és Wernicke-féle területeinek helyét, a pozitronon kisugárzási tomográfia kitűnő vizsgálati eredményeit, nagyon kevés ismeretünk van a sokmilliárd neuron kapcsolatáról, működéséről és hatásáról. A fejezet végén a halláselméletekről, helyesebben azok klasszikus formáiról esik szó. Sokat foglalkoznak a Helmholtz-féle régen megcáfolt, de történelmi jelentőségű rezonanciaelmélettel, sőt még az újra elővett telefon

elméletet is megemlítik. Viszonylag kisebb súllyal foglalkoznak a haladó hullámú elmélettel, és még nem tekintik elérkezettnek az időt a Corti-féle szerv működésének legújabb (például R. Pujol, *Ier Congr. Franç. d' Acoustique* C2-99, 1990) magyarázataival foglalkozni.

7. A beszéd akusztikai szerkezete (139-151). Ebben a rövid fejezetben a beszédszinkép, a beszédhangok szinképi szerkezete, végezetül a folyamatos beszéd szinképi ábrázolása a tárgyalás anyaga. A szinképelemzők számítógépes változatainak elvi működése és képeken is bemutatott szinképelemzési eredmények alkotják a fejezet néhány lapjának tárgyát.

8. Beszédfölfogás (154-183). Ebben a fejezetben jelentős az eltérés a régi tárgyalástól. Részben új, részben más benne a logikai sorrend, tehát lényeges átdolgozásról van szó. Mindjárt az elején összefoglalják a beszédmegértés paramétereinek háromféle megközelítési vizsgálati módját. A normális beszéd érthetőségi vizsgálata az első lehetőség. Mélyebben lehet a fölfogás részleteit megismerni, ha mesterségesen torzított formákat közlünk a hallgatóval. Ilyen eset például a zajhatás, a szűrőkörökkel torzított, vagy korlátozott amplitudójú beszéd. Végül különleges eredményeket érhetünk el, ha összeállított (szintetizált) hangokat vagy beszédet vizsgálunk. E módszerekkel az érthetőség mellett a szöveg minőségi adatait, mint a természetesség, a hanghordozás, a beszédinamika stb. is kutatható. A mesterségesen összeállított anyagban talán a legjobban ellenőrizhetők a beszédmegértési folyamatban legfontosabb elemek.

E rövid áttekintés után főként a magánhangzók és a zárhangok fölismerhetőségének paramétereit vizsgálják. A tárgyalás során Cooper és munkatársai (JASA, 1952) híres dolgozata alapján közlik az angol zárhangok "biztos"

fõl ismerhetõségi adatait. Szerencsére a régi kiadás sematikus 8.7 ábrája ezúttal elmaradt, de maguk a szerzõk is megjegyzik, hogy a teljes fõl ismeréshez más tényezõket is figyelembe kell venni. Rá kell mutatnunk arra is, hogy a 171. lapon korszerû eszme futtatást kapunk a kontinuos beszéd folyamat és a szekvencia szerû fonémásor azonosításának rendkívüli nehézségeirõl. A szerzõk valahogy úgy fogalmazzák, hogy nem mondható, hogy egy zárhangból és magánhangzóból álló szótag elválasztható szegmentuma pontosan megfelelne akár a kezdõ mássalhangzó, akár a követõ magánhangzó fõl ismerési adatának. Mégis, a nyelvirõl az akusztikai szintre való átalakítás során a diszkrét beszédhangok (fonémák) sora áttanszformálható folyamatos beszédhangokká, a folyamatos akusztikai hullám azonban egyelőre nem szegmentálható egymás utáni különálló fonémákká.

Ezért találkozunk olyan törekvésekkel, hogy találjunk invariáns akusztikai paramétereket a beszédérthetõségre. Ezek az invariáns jegyek függetlenek volnának a hangkörnyezettõl, és pontosan egyetlen fonémára utalnának. A kérdés még nem dõlt el, és különösen azért vita tárgya, mert a valóságos beszéd és a mûvi (összeállított) beszéd értékelése agyunkban valószínûleg más program szerint bonyolódik le.

A fejezet vége felé még két érdekes téma kelti fõl a figyelmet, és pedig a "szupraszegmentális" elemek és nem akusztikai jellegû kísérõjelenségek erõtlen hangsúlyozása. Az elõbbiek a beszéd emocionális jellegû velejárái, az utóbbiak a hallgató oldalán jelentkező nyelvi (statisztikai) és agyi (asszociációs vagy tárolási jellegû) segítõ elemek. Ezek egy nyelven belül annyira jelentõsek lehetnek, hogy meg nem értett akusztikai jegyek fõl ismerését is pótolhatják.

9. Beszédhangok digitális földolgozása (186-202).
Lényegében ez a fejezet is új. Elõljáróban a szerzõk a digitális

eljárások számítástechnikai előnyeit ismertetik. 1990-ben egy nagyteljesítményű, de már asztali méretű, VLSI rendszerű számítógép 16 millió elemet (karaktert) tárolhat és 40 millió számtani műveletet végez el 1 másodperc alatt. Amellett 30 év óta a teljesítmény növekedése és az ár egyidejű csökkenése együtt 250 milliószoros kedvező irányú helyzetet teremtett. Ezért vezették be mind intenzívebben a beszédkutatásban (gépi fölismerés, fordítógép stb.).

E rövid bevezetés után kerül sor a "bit" és "byte" fogalom ismertetésére, és az analóg-digitális átalakítás legfontosabb alapelemeinek, mint például a mintavételi eljárás, a mintavételi szám, a kvantálási hiba stb. bemutatására. A digitális szűrők és a szinképelemzés tárgyával zárul a fejezet. Bemutatják a szerzők azt is, hogy az így létrehozott, időben változó szinkép mennyiben különbözik egy régi spektrogramtól. Például egy 10 ms hosszúságú ablakon a gyors Fourier-tranformáció 10000/sec mintavételi sebesség folytán éppen 100 minta jelenik meg. Amellett az érzékenység is nagyobb és a változtathatóság (kivágás, nagyítás, színezés) összehasonlíthatatlanul jobb vizsgálati lehetőségel jár.

10. Beszédösszetétel (203-218). A jelenlegi nagy lehetőségek fölillantása után a szerzők rövid áttekintést nyújtanak a mechanikus-akusztikus rendszerekről, elsősorban Kempelen híres beszélőgéperől. Röviden említik, az első elektromos próbálkozásokat, sőt a Dudley-féle Voder - bár ez szerintünk éppen a Kempelen-féle alap gondolat megfelelője - ki is maradt a fősorolásból, csak a formáns-szintetizátorok összefoglaló alcím alatt sejthető. Viszont a Vocoder rendszerek telefontechnikai jelentőségük miatt részletesen föl vannak sorolva.

A témához szorosabban hozzátartozó, számítógépes program alapján működő text-to-speech rendszerek alapelvei és

megoldásai alkotják a fejezet gerincét. Érdekes, hogy a kiindulási alap elsősorban nyelvi, és csak másodsorban akusztikai. Mindez a 10.1 táblázatban van összefoglalva, amelynek egyes pontjait az alábbiakban kivonatossan felsoroljuk.

1. Kell egy kiejtési szótár, amelyben minden címszó fonetikai átírási részleteket tartalmaz a szótaghangokra, hangsúlyra és nyelvtani alakra. Még arra is, hogy a szó tartalmi vagy szerkezeti jellegű-e.

2. A szintaktikai szerkezet elemzése, a frázis és mellékmondat határainak kijelölése, a központosítás, és további szótári rögzítés a szavak beszédbeli jellegéről.

3. Ugyancsak rögzítendő a szavak hanglejtése és hangsúlya.

4. A difonéma jellegű akusztikus szótagokat vissza lehessen hívni, hogy átmeneti kapcsolatokat lehessen szerkeszteni az érintkező fonémák között.

5. Be kell szabályozni a hangszeletek időtartamát, formánsait és alaphérvencióit, alkalmazkodva a hanglejtési és mondat szerkezeti kívánalmakhoz.

6. Végül a beszabályozott adatokkal vezérelni kell a beszédet utánzó hanggenerátort.

Az ismertetést készítő szerény véleménye mindössze annyi, hogy a fenti kívánalmak még mindig megoldhatatlan nehézségek elé állítják az igazi megoldások szerkesztőit. A fejezet további részében mégis részletes magyarázatot kapunk a kiejtési (angol) szótár megszerkesztésére, a szavak szerkezeti szerepére és a számítógépi program elkészítésére. A statisztika szerint egymillió szavas szövegben legfőljebb 50 000 szó van, sőt 2000 szó már a szöveg 75%-át tartalmazza. Mégis, biztonság okából a szöveg-beszéd rendszerű szótárak mintegy százezer szóval dolgoznak. Az ilyen szerkezetek még kezdetleges formában is hasznosak akár a vakok kultúrájának kiterjesztésében, akár a különféle betegek beszédgyakorlásának megkönnyítésében.

11. Önműködő beszéd felismerés (219-239). Nem véletlen, hogy az utolsó helyre a könyv a legfontosabb és legaktuálisabb fejezete került. Megértéséhez az eddigiek mellett teljesen új ismeretek kellene. A szerzők rokonszenves önkorlátozására utal, hogy ezt a nehéz feladatot is meg tudták oldani mintegy 20 lapos terjedelemben. Egy fél évszázadnak kellett eltelnie ahhoz, hogy világosan megfogalmazódhasson, hogy mit kellene és mit lehet elérni ebben a törekvésben. A fejezetben földolgozott és a további olvasmányként javasolt irodalmi anyag nagy része 4 évnél nem régebbi, tehát várható, hogy a szerzők pontosan a mai állapotot mutatják be. Mindjárt a szokásos bevezető összefoglaló elismerésre késztet a szerzők reális és mérsékelt bizakodó álláspontjával szemben, amelyben sem szakmai optimizmus, sem tudományos fantázia nem torzítja el a valóság reális értékelését. Érdeemes két mondatot szabadon idézni a 219. lapról. "Vannak olyan gépek, amelyek bizonyos szövegösszefüggéseket gyakorlati célokank megfelelő szinten és megfelelő költséggel képesek fölismerni. De nagyon messze vagyunk olyan gépi megoldásoktól, amelyek folyamatos beszédet ismernének föl azzal a pontossággal és hajlékonysággal, amely az emberi megértés során különösebb tudatos megerőltetés nélkül simán végbemegy".

Az áttekintésben először a személytől függő és a tőle független fölismerés feladatának, a kis és nagy szótári anyagot tartalmazó rendszerek összehasonlításának, a különálló szavak vagy a folyamatos beszéd fölismerési lehetőségeinek mint problematikának a rövid ismertetése kerül sorra. Ezután térnek rá az egyes részletkérdések taglalására. Például, hogy hogyan kell áttekinteni a gyors, és a színeképen szemmel aigha érzékelhető, de a számítógép sebességével követhető színeképi szerkezet elemeit információs adatokra. Például 10 ms-onkint rögzíthetők 12 frekvenciatartományban a színeképi amplitúdók, ugyanannyiban a két egymás utáni időköz átmeneti jellemzői,

továbbá a teljesítményszintek és további jellemzők. Mindezt több kiejtésre is meg kell határozni, miközben figyelembe kell venni a beszédsebesség és a kiejtés változatait, a hangfekvés eltéréseit, sőt az alapzajt és a felvételtechnikát is. A szerzők kétféle fölismerő szerkezetípust részleteznek: az úgynevezett "template matching" (magyarban talán *kirakós eljárásnak* lehetne honosítani) és a statisztikai beszédfölismerőket. Nem tartják még megérettnek az időt a néhány eredményt már fölmutatott neurális szerkezetek tárgyalására. Azt azonban megemlítik, hogy a nyelvi és nyelvtani modellek segédeszközeit figyelembe vevő szerkesztési elemek és módszerek minden esetben javítanak az elérhető eredményeken. Az egyik alaposan ismertetett rendszer elsősorban különálló szavak fölismerésére alkalmas. A gép az elhangzott szó szokásos 10 ms-os szeletelésével és az egyes szeleteknek a mintáival való összehasonlításával végzi munkáját. Ebben az esetben is fontos szerepet kap a "dynamic time warping" (magyarban a *kitűnő vetemítés* szót használjuk rá). A másik fontos gépi feladat a megfelelő mintával való összehasonlítás, ami - hála az igen gyors számítási műveleteknek - jó eredménnyel elvégezhető. Ha kapcsolt szavakról van szó, újabb nehézségként jelentkezik a szétválasztás, de sok esetben ez is megoldható. (Itt egy közvető megjegyzést kell tennünk. A példaként közölt 11.1 ábrában nemcsak az elválaszthatóság és az összefolyás jelent különbséget, hanem az alsó példában a magasabb felhangok hiánya is. Lásd a "the" szó kétféle alakját).

Minden mai fölismerő szerkezet fő nehézsége, hogy a jó hatásfokú megoldás érdekében hosszasan "be kell tanítani". Ez személytől függő esetben aránylag egyszerűbb, de akkor is napokat vesz igénybe. A gépek bonyolultságától függően ezekkel a megoldásokkal 100-1000 szavas, szakmai szókészletet fölismerő rendszerek léteznek. A fölismerés

százaléka egyetlen betanult gép és begyakorolt személy esetében elég magas, de csak különálló szavakra.

Jobb eredmények érhetőek el az 1980 óta egyre gyakrabban alkalmazott, statisztikus módszerekkel működő fölismerő szerkezetekkel. A módszer titka a matematikából jól ismert Markov-féle eljárás. Régen lánc volt a neve, ma modell formában emlegetik. A híres HMM (hidden Markov model) a szeletelt színeképi elemeket valószínűségi elosztásként kezeli, amely az időben változik. Az egyes állapotok változásai, valamint visszacsatolással való megerősítése elvezet a legvalószínűbb hasonlóság kiválasztásához. Az eljárás "gátolt" kifejezése arra mutat, hogy nem a beérkező adatok fölismeréséről, hanem az egyes pillanatképek szerkezeti változásáról kapunk fölvilágosítást.

Az úgynevezett folyamatos beszéd fölismerése természetesen további nehézségekkel jár. Hiszen ilyenkor sem a szóhatárokra, sem a koartikulációs hatásokra nem támaszkodhatunk. Az említett módszerrel tehát 100 szavas szótáraknál nagyobb korpuszt nem foghatunk át. Ezért a legegyszerűbb eljárások szavaknál kisebb szegmentumok (szótagok vagy éppen fonémák) meghatározásával foglalkoznak. A szerzők az utolsó 4 lapon az ilyen rendszerek taglásába is bebocsátkoznak. Ezúttal sem a fejlesztő vállalatok adatait tartják elfogadható eredménynek, hanem valós adatokat közölnek. Egy közel 1000 szavas szakszótár (hajózás, földrajzi helyzet) 3500 mondatából 100 beszélő szövegmondása alapján 1991 végén 80%-os mondatfölismerés volt az eredmény.

A recenzens megjegyzése, hogy 1993-ig a helyzet sokat javult, de még egyszer sem sikerült egy begyakoroltatott szűk témakörön kívül eső szavakat eredményesen fölismertetni. Ez pedig azt jelenti, hogy a jelenlegi, nyelvtől függő megoldási rendszerek messze nem képesek még a saját nyelvükön belül sem eljutatni a teljes beszédfölismerés ideáljához.

A könyvet 6 oldalas tárgymutató egészíti ki, amelyben nemcsak a szövegre, hanem a bő ábraanyaga is hivatkoznak.

* * *

A továbbiakban néhány értékelő megjegyzést sorolunk föl, közülük ismételten megemlítünk néhány előremutató tanulságot. Kiemelésre kívánczok, hogy számos részletet találunk a kutatás előterébe kerülő újabb feladatokról, megoldásokról és vizsgálati eljárásokról. Közülük néhány: az otoakusztikus jelenség (93), a pozitron emissziós tomográfia (127), az integrált áramkörök adta korszerű lehetőségek (188) és a Markov-modell (229) alkalmazása. Értékes része az anyagnak a beszédmegértés nem-akusztikai jellegű összetevőinek taglalása (179-182) és jóleső tudomásul szolgál a 8. fejezet címének Speech recognition-ról Speech perception-ra változtatása. Az előbbi kifejezés ugyanis inkább a gépi azonosítás terminusa, amint a 11. fejezet címe is mutatja. További értékes bővülés az összeállított (szintetikus) beszédjellegű információk vizsgálati eljárásaival kapcsolatos. Ez a módszer a labilis fizikai egyensúlyi helyzetek éles befolyásolási lehetőséghez hasonlít.

Az említett kiváló ismereti anyag mellett bizonyos megkötöttséget jelent az alkalmazásokban a tárgyalás kizárólagos angol nyelvi alapja. Ugyanis jól tudjuk, hogy éppen az amerikai-angol fonetika, valamint az ehhez kapcsolódó többértelmű beszéd-szöveg kapcsolat teszi alig megoldhatóvá a gépi beszédfölismerés feladatát, ami egy szilárdabb fonetikájú nyelvben, mint a japáni, olasz, finn stb. sokkal kisebb nehézségekkel járna.

Ennek ellenére a könyv elolvasása után a szakmai ismeretekkel rendelkező kutató is sok, szemléletileg vagy összefüggéseit tekintve újszerű megállapítással gazdagodik. Különösen az agyi értékelésről, a korszerű digitális

jelföldolgozásról és a gépi beszédfölismerésről írottak kelthetik fel érdeklődését.

Néhány eszmefuttatásról, megállapításról vitatkozni lehet, amint ezekről az irodalomban is vita van. Ilyen például az érthetőség többértelmű fogalma. A fonetikai, szótag-, szó- és mondatérthetőség közti összefüggések elég határozatlanok. A nyelvi ismeretek, figyelem, intelligencia mellett a gyakorlat döntően befolyásolja az eredményeket. Tulajdonképpen az érthetőség, mint jól definiálható fogalom, nem is létezik; objektíven csak a fonetikai- és a szövegerthetőség vizsgálható.

Hasonlóképpen nehezen fogadható el nyelvektől függetlenül egyértelmű kritériumnak a könyvben ismertetett mechanizmus a zárhangok megkülönböztetésére. Várni kell még a nyelvi sajátosságoktól független, pontosabb meghatározásra, hiszen az ismertetett túl egyszerűsített fölismerési mechanizmus az angol nyelvere sem mindig igaz.

Végül szabadjon néhány módosításra érdemes adatra felhívnom a szerzők és az olvasó figyelmét. Kívánatos volna a mértékegységek SI-rendszerének következetes használata. Régiesen hat néhány hosszúság inch-ben és néhány sebesség feet/sec-ban megadott értéke (52, 54, 58, 118, 120), valamint a dyn/cm² használata a pascal (Pa) helyett. A mikron (118) helyett ma inkább a mikrométer (μm) szokásos.

Az ábrák egy része valószínűleg szövegszerkesztési okokból a részletekhez képest túl nagyra sikerült (pl. 3.9-3.11, 4.4, 4.8, 5.2, 5.5, 6.9, 8.7, 8.8), egyesek meg vannak ismételve (4.17a = 4.18a = 4.19, 5.6 = 6.9), végül 7.2-ből hiányzanak a jelölések, 8.1 pedig számszerűleg nem helyes.

Kifogásolható a ma szabványos hangosság szint görbék helyett a régi Fletcher-Munson-féle görbék közlése (5.10); hasonló megjegyzés vonatkozik a son-skála (5.11) és részben a mel-skála (5.12) grafikonjára is.

A lány szájpád hátsó lebenye, a nyelvcsap (velum) az ábrákon (4.8, 4.10 és 4.12) nincs külön megnevezve, de a 172. lapon előfordul a velum megnevezés, amely így nem azonosítható az ábrákkal. A külön megnevezés az ábrákon is indokolt.

Helmholtz rezonancia-elmélete kétségkívül úttörő jelentőségű (132), de ma már túlhaladott. Viszont a haladó-hullámú elmélet első kísérleti kimutatója (Békésy, Phys. Zs. 1928, 793-810) név szerint nincs megemlítve, pedig munkásságáért 1961-ben Nobel-díjat kapott. Egyébként a könyv 5.6 és a vele azonos 6.9 ábrája éppen Békésy egyik dolgozatából (Akust. Zs. 1943. 66-76) származik.

Kétségtelen, hogy ez a néhány kritikai megjegyzés semmit sem von le a könyv tudományos és didaktikai értékéből, ezért ismételten csak elismerésünket fejezhetjük ki a Szerzőknek és a Kiadónak a tanulságos és szép kiállítású könyvért. Hazai kollégáinknak pedig melegen ajánljuk elolvasását és munkájukban való fölhasználását.