

**BESZÉDKUTATÁS**  
**1995**



# **BESZÉDKUTATÁS**

**A Beszédkutatás '95 konferencia  
válogatott tanulmányai**



# BESZÉDKUTATÁS

A Beszédkutatás '95 konferencia  
válogatott tanulmányai

Szerkesztette:

Gósy Mária

A Magyar Tudományos Akadémia Nyelvtudományi Intézete  
Budapest 1995

Az egyes tanulmányokat lektorálták:

Gósy Mária  
Olaszy Gábor  
Siptár Péter

Technikai munkatárs:  
Szalai Enikő

ISBN 963 8461 97 7  
© MTA Nyelvtudományi Intézete

Felelős kiadó: Kiefer Ferenc igazgató  
MTA Nyelvtudományi Intézete

## TARTALOM

<b>MEGNYITÓ (Vértes O. András)</b> .....	1
<b>Tarnóczy Tamás:</b> Az egyfülű hallás síkbeli iránytól függő színképi torzításai.....	4
<b>Gósy Mária:</b> Szükséges és szükségtelen hangátmenetek .....	20
<b>Hunyadi László:</b> Mondathangsúly a magyarban.....	32
<b>Olaszy Gábor:</b> A kérés, a figyelmeztetés, a felszólítás és a kérdés prozodiája a kijelentő mondat tükrében .....	46
<b>Bakró-Nagy Marianne:</b> Hangtörténeti változások percepciós vonatkozásai.....	62
<b>Kovács Magdolna:</b> Időtartam-ítéletek függőségi viszonyai.....	71
<b>Szalai Enikő:</b> Az [u:] [a:]és [i] hangok koartikulációs mezőiről .....	83
<b>Banczerowski Janusz:</b> A nyelvi kompetenciáról .....	93
<b>Lajtha György:</b> A digitalizált beszéd minősége.....	105
<b>Ecsedi Csaba:</b> A gépi beszéddel kiegészített számítógépek alkalmazási lehetőségei vak és csökkentlátó emberek számára .....	116
<b>Ast László–Ács György–Kalotai Béla–Marosi Gyula– Tatai Péter:</b> Additív zaj hatása a beszéd lényegkiemelt paramétereire .....	124
<b>Hazai és nemzetközi támogatással végzett beszéd kutatások.....</b>	133
<b>Visszanéző (Siptár Péter)</b> .....	139



# MEGNYITÓ: VÉRTES O. ANDRÁS

## Hölgyeim és Uraim! Kedves Barátaim!

A hangzó beszédet nem hagyják változatlanul korszakunk jelentős tényezői, tendenciái. Ezek némelyikéről röviden a következőkben:

1. Korán eltávozott, kiváló tanítványom, Montágh Imre kéresemre leírta megfigyeléseit az "sz, s hang feminin, illetve maszkulin kiejtésé"-ről. Kézirata 1978-ból (vagy 1977-ből) való. Mint Montágh írta, "a korszellem a n e m i m a g a t a r t á s p o l a r i z á l t s á g á t (V. O. A. ritkította) összébb mosta: s fiatalabb férfiak s hangjai világosabbak, mint az ötven-hatvan éveseké. Az ember azt gondolná, hogy a női ejtésnek kellett volna a férfi formához közelíteni, de furcsa módon az ellenkezője tapasztalható, s bizvást állíthatom, hogy a magyar sz, s hangok világosodnak." Egyébként "az sz, z, c, dz és az s, zs, cs, dzs hangok kiejtése az átlagos női ejtésben lényegesen "világosabb" színű, mint az átlagos férfi hangoztatásban... nőies férfiak magasabb és férfias nők mélyebb sziszegőket ejtenek."

Ismeretes, például a szociológiából is, hogy a női és férfi viselkedésbeli, szokásbeli különbségek korunkban kisebbedtek, így az öltözködésben. Mikor annak idején idéztem Montágh Imre figyelemreméltó észrevételeit, nem szóltam azokról a mélyebbre ható változásokról, amelyeket a férfiak, illetőleg a nők pszichoszomatikus zavarai is visszatükröznek.

Hogy a XX. század első felében változtak a férfi és női személyiség típusok, azt már James L. Halliday megállapította, rámutatván a pszichoszomatikus zavarok nemek szerinti megoszlásának történeti változására: például a múlt században nők körében gyakoribb volt a gyomorfekély, a Basedow-kór és talán az essentialis hypertonia is; mindezek a XX. században egyre gyakrabban sújtották a férfi nemet; ezzel szemben a cukorbeteg,

amelynek prevalenciája régebben főleg férfiak között volt nagyobb, századunk első felében növekvő arányban fordul elő nőknél. Halliday ebből is jogosan következtet arra, hogy a női és a férfi személyiségi jegyek némileg semlegessé váltak: a férfiaké bizonyos tekintetben "femininek" lettek, a nőké pedig "maszkulinok". (Psychosocial Medicine. New York 1948, 66.)

Idevág az is, hogy Kosztolányi, miután egy férfi színésről azt írta, hogy "Föltétlenül nőies", de mindjárt hozzátette: "Vegyük azonban számba, hogy a tépelődés bizonytalanságától szenvedő századunk is az..." (Kosztolányi Dezső: Színház. Bp. 1943? 242.)

Nemcsak az artikulációban, hanem a gégében képzett hangnak (a "Stimmé"-nek) színezetében is lehetnek eltulodások. Több szakemberünk említette szóbeli közlésben falusi nők éles (vagy élesebb) hangját, így Ádám Jenő, Sztanó Pál, Magdics Klára (ő a Pápa vidéki nők élesebb hangjára emlékezett vissza). Elképzelhető, hogy az imént említett éles hangszínezet eltűnőben van vagy lesz, de ez vizsgálatra szorul.

2. Percepció s magatartásunk is változott. Szóvá teszi ezt Adorno is: "...a hallgatók maguk többé már egyáltalában nem is képesek koncentráltan hallgatni valamit. Képtelenek a fokozottan megfeszített figyelemre..." (Theodor W. Adorno: Fétis-karakter a zenében és a zenehallgatás regressziója. 1938. Ford. Horváth Henrik. In: Theodor W. Adorno: Zene, filozófia, társadalom. Bp. 1970, 256.) Bár Adorno itt zenéről szól, föltehető, hogy az európai kultúrában a beszédre való figyelem sem maradt érintetlen. Ebben a föltevésben megerősít Jürgen Rausch (Muttersprache 1960, 72), aki szerint a beszéd meghallgatásának kultúrája csökkenőben van, és támogat az a nagy pedagógiai-lélektani, illetve gyógyító pedagógiai irodalom, amely a gyermek- és ifjúkorban előforduló (általános) koncentrációs gyengeségről szól. Ez az irodalom nagyrészt a második világháború utáni évtizedek termése, amely föltehetően a koncentrációs zavarok növekedő számának köszönheti létét. Meg-

említhetjük, hogy Erich Fromm a koncentrált figyelem csökkenését a modern ember egyik jellemző vonásának tartja (*Die Kunst des Lebens*. Frankfurt/M.-Berlin-Wien 1972, 143).

3. A szupraszegmentális tényezők egyikének-másikának változását föltehetően módosította és módosítja a rádió és a televízió. Főleg a beszédtempó gyorsulására lehet hatással a két médium, de nemcsak erre, hanem a gyors beszéd megértésére is. A médiumokban hallható beszéd bizonyára kiegyenlítő hatással van az egyes nyelvjárásokban vagy kisebb településeken szokásos (egymástól különböző) beszédsebességekre.

4. Egy másik kérdés a sok közül, amelyre választ várunk. Korunkban változott a különféle konstitúció típusok gyakorisága. Feltűnő, hogy a leptoszóm alkat gyakoribb. Nem állítom, csak kérdezem, vajon hat-e ez a körülmény a beszédtempóra, a hangmagasságra?

\*

Csábító az ellenkező irányú hatásról szólni, ti. arról, hogy a hangzó beszéd miként hat beszédén kívüli tényezőkre, így a lelkiismeretre. A lelkiismeret szavát ugyanis akusztikus jellegűnek fogjuk föl, Hans-Martin Gauger szerint "vokal-akustisch"-nak (*Neue Rundschau* 1974, 582). A szó azonban – és talán különösen a kimondott szó – veszített értékéből, érzelmi velejárájából (vö. MNy LXXXII, 292). Így érthető, hogy a svájci Hans Zbinden szerint korunk lelkiismeretét fenyegeti a szó elértéktelenedése is (*Der bedrohte Mensch*. Bern – München, 1959, 170). De tüzetesebb vizsgálódás szükséges, hogy tisztábban láthassuk a hangzó beszéd és a lelkiismeret kapcsolatát. Einstein szava jut eszembe, aki azt mondta, hogy nem az atomerő a legnagyobb probléma, nagyobb ennél az emberiség lelkiismeretéé.

Mindezekre gondolván megnyitom a konferenciánkat.

# AZ EGYFÜLŰ HALLÁS SÍKBELI IRÁNYTÓL FÜGGŐ SZÍNKÉPI TORZÍTÁSAI

Tarnóczy Tamás

MTA Természettudományi Kutatólaboratóriumok, Budapest

A monaurális hallás irányfüggő torzítását vizsgáltuk egy forgatható Brüel-torzón. A vizsgáló hang rózsaszínű zörej volt, az érzékelt hatást 1/12 oktáv szélességű sávokban mértük. Az alapfeltételezés szerint a különböző irányból érkező és a dobhártya helyén jelentkező színeképek egymásból kivonhatók, így a külső hallójárat hatása kiesik, és csak a fülkagyló és a fej torzító hatását mérjük.

## Az előzmények áttekintése

A beszéd és a hallás és a beszédmegértés szoros agyi kapcsolatban áll egymással. Bármelyikük teljes megismeréséhez a többiek működésének ismerete is szükséges. Ezért nem különös, ha a beszédkutatókkal foglalkozó beszámolók közé egy hallástani részletkérdés tárgyalása is beékelődik.

Mielőtt az akusztikai jel adatfeldolgozásra a hallószervbe beérkezik, a hangtér az ember jelenléte miatt torzulást szenved. Az agyi ítélet a torzított hangtéri adatok alapján születik meg.

Az emberi fülkagyló (1. ábra) bonyolult és egyéni formázású alakzatától nem várható, hogy általános érvényű, pontos átviteli adatokat szolgáltatson. Fő szerepe nyilván impedanciaillesztés a szabad hangtér és a külső hallójárat akusztikai ellenállása között. A nagyfrekvenciás jellegű mássalhangzók elsősorban a kétfülű intenzitáskülönbség alapján ismerhetők föl, hiszen a fej árnyékoló hatása leginkább ilyen hangokra érvényesül. A magánhangzók alsó formánsainak fölismerésében az idő- vagy fáziskülönbségre vagyunk érzékenyek, a magasabb formánsok érzékelésében kombináltan reagálunk az idő-, az intenzitás- és a torzításkülönbségek kétfülű megjelenítésére.

Egyelőre még nem értjük, hogy a külső füllel szabályzott előkészítés mennyit segít a végleges agyi földolgozásban, mégis a század eleje óta többen utalnak a kétfülű hallás ezen szerepére. Az első jelentősebb kísérletek a 40-es évektől Francis Wienertől, Békésy jó barátjától származnak (Wiener–Ross 1946; Wiener 1947). Ettől kezdve mind a kétfülű, mind az egyfülű hallással bőven foglalkoztak, például Kietz (1953), Jahn (Jahn–Vogelsang 1959), Schirmer (1963), Shaw (1966) és Blauert (1968) de csaknem mindig a teljes külső fül viselkedését vizsgálták. Ebbe a fejtől és a fülkagylótól a külső hallójáraton keresztül a dobhártyáig számos összetevő tartozik.

Áttekintéséhez két szerző kiváló összefoglaló munkáira hivatkozhatunk, Edgar Shaw 1974. és 1980. évi dolgozatára és Jens Blauert ugyanezen időben megjelent, ma már klasszikussá vált monográfiáira (1974; 1985). A frekvenciafüggő átviteli függvényt a kísérletek során általában a hallójáratba bevezetett szondamikrofonnal határozták meg, de Shaw és Teranishi (1968) modellvizsgálatokat is végeztek. A hangforrás diszkrét vagy folyamatosan változó szinuszos hang, esetleg terc sávzélességű zöreje vagy különféle időtartamú és összeállítású hangimpulzus, a hangtér pedig síkhullám vagy közeli gerjesztésű pontforrás volt, végül a vizsgálat a vízszintes vagy a fejközép síkjában több téri irányra vonatkozott. A modellvizsgálatok kivételével nem fordult elő, hogy a teljes külső fül átviteli függvényét a hallójárat akusztikai hatásától el tudták volna különíteni. Blauert (1985, 55) úgy nyilatkozik, hogy a teljes külső fül egyetlen rezonátor rendszer, amelynek egyes sajátrezonanciái a hangforrás irányától és távolságától függenek. Shaw és Teranishi (1968) modellvizsgálataiból arra a következtetésre jut, hogy a fülkagyló bizonyos értelemben a hallójárat meghosszabítása és mintegy nyílaskorrekció a szerepe. Mi fordított értelemben arra gondoltunk, hogy a concha a külső hangtér impedanciáját transzformálja (mintegy illeszti) a hallójárat impedanciájához. Eb-

ben az elképzelésben a különböző irányokból érkező síkhullám számára minden esetben azonos feltételek mellett látja el az impedanciaillesztést a terelő fülkagyló, és a fülkagyló legmélyebb részét kitöltő, a hallójáratba nyíló *cavum conchae*. Egyelőre kevés részletkérdés van tisztázva ahhoz, hogy ez a munkahipotézis pontosabban leírható legyen.

Wiener és Ross (1946) azonban megállapította, hogy a külső hallójárat átvitele a beeső hanghullám irányának függvényében 8 kHz alatt invariáns. Később ezt az elméleti határt Shaw (1969) 10-15 kHz-ig is érvényesnek tekintette. Sajnos, a hallójárat bejárati nyílása előtti említett rezonanciaterbe, a conchába helyezett mikrofonsonda, amellyel a kísérleteket végezni lehetett, annyira megzavarta az ott kialakult hangteret, hogy Wiener és Ross korszakalkotó jelentőségű méréseit sem lehet teljesen hitelesnek elfogadni.

### A kísérleti módszer

Saját kísérleteim során (Tarnóczy 1992) feltételeztem, hogy a hallójárat átviteli invarianciája valóban fennáll, vagyis hogy nagy valószínűség szerint a külső belépési nyílástól a dobhártyáig vezető csőalakú rezonátor (*meatus externus*) minden beérkező hangjelben azonos változást okoz. A torzítások más része a fej és a fülkagyló hatására jön létre, és az természetesen irányfüggő. Ha tehát a hallójáratval szemről és valamilyen más irányból beérkező hangjeleknek a dobhártya helyén mérhető színeképeit egymásból kivonjuk, a hallójárat azonos értékű módosítható hatása kiesik és csak a fej és a fülkagyló okozta torzítást regisztráljuk. Erre pedig legegyszerűbb kísérleti eljárás a műfejes hangmérés, például a Brüel és Kjaer 4128 típusszámú torzójával. Ezt a bábút a 3922. számú forgóasztallal a vízszintes síkban bármilyen irányba be lehet állítani, azaz bármilyen irányú hangjelnek a dobhártya helyén kialakult színeképét lehet általa rögzíteni. Esetünkben a bábu

mindkét, plasztikus műanyagból készült fülkagylója a helyén volt, de csak a jobb fülbe érkező hang szinképét regisztráltuk.

Vissza kell utalnunk arra a tényre, hogy a fülkagyló alakja, nagysága és dőltégi foka személyenként nagy változatosságot árul el, ezért természetes esetben az irányfüggő kísérleteknek ez okból is nagy a szórása. Az általunk javasolt műfülkagylóval ez elesik és a kísérletek bármikor megismételhetők és összehasonlíthatók. A hangforrás alakjától és időállandóságától, továbbá az elemzési és értékelési eljárástól függően még így is elég hibalehetőség marad.

Az általános ábrázolási szokásoktól eltérően  $0^\circ$ -nak a fülbe egyenesen érkező hang tengelyirányát tekintettük, és minden más irányt ehhez viszonyítottunk. A 2. ábrán bemutatjuk, hogy Blauert és Shaw iránybeosztása mellett sajátunk hogyan értelmezendő: a szemből érkező hang iránya nálunk  $+90^\circ$ , a hátulról érkező  $-90^\circ$  (elől-hátul érzékelés egyfülű hallással). További formabontás, hogy a forgatható műfüles bábút öltözet nélkül és kötött kabátban, parókával ellátva is bevontuk a kísérletekbe (3. ábra). Erre azért volt szükség, hogy a ruházat hangnyelő hatását is kimutathassuk.

Minden korábbi eljárástól eltérő két kísérleti újításunkat tartjuk még fontosnak kiemelni. Az előző abban áll, hogy mindig az adott folyamatos, nem szinuszos és nem impulzus jellegű zörejang teljes átviteli szintjének alakzatát vizsgáljuk meg. A süketszobában 3 m távolságból érkező rózsaszínű zörejang teljes szintje 75 dB volt, 200 Hz és 16 kHz között  $\pm 6$  dB ingadozással. Nagyobb eltérés csak a hangszóró két sugárzó sávjának kereszteződési frekvenciája helyén, 3,3 kHz környékén jelentkezett. A frekvenciafüggő ingadozások az esetek nagy részében nem zavartak, mert a különbségképzés, vagyis a szinképek kivonása miatt, a feldolgozás során úgyis kiestek.

A másik kísérleti újdonság a Brüel és Kjaer 2133 típusú kétsatornás azonosidejű elemző használata, amelynek 1/12 oktáv sávszűrős algoritmusát vettük igénybe, vagyis a gép az elemzést zenei félhargonként végezte el. Ezt a számítástechnikai feladatot

az ingadozások kiegyenlítésére esetünkben általában 16 másodperc alatt bonyolította le. Világos, hogy ezzel az eljárással nem lehet folyamatos beszédet elemezni, de az állandósult állapotú zöreijhang irányától függő torzulásainak fölismerésére kiválóan alkalmas. A 15°-onkint fölvelt színekép rögzítése, elemzése, összehasonlítása, vagy egymásból való kivonása szolgáltatja azt az 1600 adatból álló eredménylistát, amelynek eddig csak egy részét közöltem (Tarnóczy 1992) és amelyről röviden be kívánok számolni.

A kivonási eljárás nem mindig volt azonos irányú. Általában a közvetlen irányban ( $\varnothing \neq 0^\circ$ ) fölvelt színeképet tekintettük alapnak, és vontuk ki belőle az egyéb irányban kapott színeképi adatsort. A kísérlet során ez volt a természetes megoldás, mert így az egyszer fölvelt alapszínekép minden további műveletnél a helyén maradhatott. Ha ilyen esetben valamelyik frekvenciatartományban a kivonandó meghaladja az alap színeképi adatát, negatív érték az eredmény, ami valójában azt jelenti, hogy ott a ferdén érkezett hangjel a dobhártya helyén nagyobb szinten jelenik meg, vagyis a negatív érték éppen hangerősödést jelent. Ilyesmi, persze, ritkán fordul elő; a ferdén érkező hang intenzitásának értékelése általában csökken, amit a kivonási hangszínekép növekedése jellemez. De nem minden esetben alkalmaztuk ezt a kivonási irányt, ezért a szintváltozás értelmezését esetenként pontosítjuk.

Alapvető módszerünket a 4. ábrán mutatjuk be. A felső színekép a hangszóró hangja szabad hangterében, a később odahelyezendő műfej helyén. A 3 m távolságban elhelyezett hangszóró hangja itt 2% pontossággal síkhullám állapotú, átlagos szintje 75 dB, ingadozásáról már szó volt. A középső színekép a  $\varnothing = 0^\circ$  irányból érkező hang dobhártya helyén mért értéke, ha a műfüles bábu már a helyén van. A mélyhangú összetevők láthatólag 15 dB-lel csökkentek, de érdekes módon 5 kHz környékén széles sávban intenzitásnyereség tapasztalható. Ez a nyereség a korábbi irodalmi adatok szerint a meatus okozta rezonanciás hangerősítés jele. Látható az is, hogy a színeképi érzékelés elég pontosan követi

a szabad hangtér frekvenciamenetét. Hogy mennyire, azt az alsó diagram mutatja. Ez ugyanis a két fölötté álló színeképe gépi kivonásával készült. A jól ismert 3...5 kHz közötti legnagyobb hallásérzékelés helye már a negatív tartományba kerül, ami azt jelenti, hogy ott a hangtér intenzitásához képest jobb a hallásérzékelés. A későbbiekben ezt a színeképet tekintjük szabványos nullszintnek, amikor más irányból érkező hang relatív érzékelési szintjét meg akarjuk állapítani.

### Néhány eredmény

Ha a 4. ábra alsó diagramjához hasonlítjuk a különféle irányból beérkező hangszíneképet, relatív színeképekről beszélünk. Előfordulhat azonban, hogy a középső diagramhoz hasonló közvetlen színeképeket is összehasonlíthatjuk egymással, és már ebből is le tudunk vonni bizonyos következtetéseket. Néhány adatot az 5. ábrán mutatunk be, ezeket a közvetlen (direct) megnevezéssel látjuk el. Az 5. ábra ilyen diagramok földolgozásának egyik gyakorlati módját szemlélteti, ahol a  $+45^\circ$  és  $-45^\circ$  irányból érkező és fölvetett közvetlen színeképek egymásra rajzoltatásával vagy gépi kivonásával jeleníthetjük meg a kettő között hallható különbséget. Ebben az esetben azonban a különbségi színekép az iránytól függetlenül általában pozitív, vagyis ha a  $+45^\circ$  irányú fölvetett szintből vonjuk ki a másikat, az előlről érkező hang nagyobb hangosságértéke a kivonási diagramon is nagyobb szintben jelenik meg.

A 6. és 7. ábrán a relatív módszer alkalmazása tanulmányozható a középállapottól előre, illetőleg hátrafelé növekvő irányokban  $15^\circ$ -onként. Mivel a kivonás mindig a  $0^\circ$ -ra vonatkozó állandó színeképből történik, a pozitív értékek ezekben a diagramokban hangerő csökkenést, a negatívok növekedést ábrázolnak. Valószínű hangerősítés csak pozitív irányban várható, hiszen a fülkagyló görbületének enyhe hanggyűjtő hatása ebben az irányban érvényesülhet. Ezt a várakozásunkat a 6. ábra színeképei alátámasztják, például a  $+45^\circ$  irányban 2,5 és 3 kHz között 4...5

dB értékű hangosságnövekedés észlelhető, ami  $60^\circ$ -nál éri el a maximumát és még  $75^\circ$ -nál is megmarad. Vagyis összesítve: pozitív irányban valóban tapasztalható némi hangerősítés, ami még a szemből,  $90^\circ$  szögirányból érkező hang esetén is észlelhető. Ha becslést akarunk végezni, a fülkagyló alakzatának részletméretei alapján (a helix, antihelix és a concha között 2-3 cm a távolság) arra gondotunk, hogy  $45^\circ$  irányban 10 és 15 kHz között is megjelenthethető némi interferenciás erősítő hatás. Valóban 18 kHz táján ennek is nyomát találtuk, viszont már  $+60^\circ$ -on 10-12 kHz táján erős intenzitásvesztés is mutatkozik. Érdekes, hogy ez a 10...15 dB-t elérő veszteség az irány további növelésével eleinte mélyebbre húzódik, később pedig nagyon kiszélesedik. Például  $+150^\circ$ -on 4-15 kHz között 10-től 28 dB-ig terjedő szinttel csökkenti az észlelt hangerősségeit. Itt már a fej árnyékoló hatása is erősen érvényesül.

Másként alakul a fülkagylóhoz képest negatív irányból érkező hang sorsa. A 7. ábra tanúsága szerint  $-15^\circ$ -on némi hangosságnövekedés észlelhető a 12-16 kHz-es sávban, ami az irányváltozás növekedésével lassan eltűnik. A hangosságnövekedés talán úgy magyarázható, hogy az árnyékoló jellegű kagyló hangelhajlást okoz, és a hang a 2-3 cm méretű akadályrendszeren átbukva interferenciás erősítést szenved. A fülkagyló belső szerkezete ez esetben nem oka a folyamatnak. Különben is, ez a fajta hatás már  $-45^\circ$  mellett eltűnik, és egyre erősödik 4 kHz göccel egy jelentős hanggyöngítő folyamat, amely a nagyobb szögirányokban kiszélesedik és 15...20 dB-re fokozódik. A legnagyobb csökkenés természetesen  $180^\circ$  irányból várható, de  $+150^\circ$  és  $-150^\circ$  között aránylag kevésbé változik, súlypontja 1-3 kHz és 4-16 kHz területére esik. Az utóbbi helyen a 25...30 dB-t is eléri. Tájékozódásul a 8. ábrán néhány jellegzetes frekvenciafüggő szintváltozást tüntettünk föl a hangérkezési irány paramétere szerint. Ezúttal a valóságnak megfelelő irányú hangosságváltozási adatokat rajzoltuk be félhangonként.

Néhány további kérdésre is megkíséreltünk feleletet kapni. Ilyen például a testről visszaverődő hang hatása. A színeképi fölvételek nem erősítették meg az úgynevezett vállvisszaverődés hangosító hatást. A kísérletekből kiderült, hogy  $0^\circ$ -on a bábu ruhás és ruhátlan állapota között alig van változás. Az 1...2 dB növekedés néhány diszkrét frekvenciahelyen nem sokat mond. Különösen döntőnek hatott az a diagram, amelyről a bábu vállára egy 5 cm vastag szivacsréteget helyeztünk, hogy csökkentsük a visszaverődésnek hatását. A változás itt is jelentéktelen. A ruhátlan/ruhás különbség jobban vizsgálható  $90^\circ$ -on és még inkább  $180^\circ$ -on (9. ábra). Ennek természetes oka, hogy az érzékelő füllel szemben, a fej túlsó oldaláig elhajló hang nagyobb elnyelést szenved a ruhás bábun, mint a ruhátlanon. Az elnyelési különbség már 800 Hz-en mutatkozik és legnagyobb értéke 2-5 kHz között a 8 dB-t is elérheti. Az összehasonlító ábra arra is fényt vet, hogy a ruhás bábu elnyelési többlete csak 4-5 kHz-ig érvényesül, viszont kevéssé magyarázható módon már 6 kHz-en lecsökken. Talán ezen a nagy frekvencián már a bábu anyagának is elegendően nagy az elnyelése. Ez a jelenség egy oktávval magasabban megismétlődik.

Egy további érdekes és eddig is észlelt összefüggésre is gyűjtöttem néhány új adatot, mégpedig az elől-hátul érzékelés különleges frekvenciatartományainak objektív kimutatásával. Ezzel a kérdéssel bizonyos értelemben egy korábbi dolgozatomban (Tarnóczy 1958) már foglalkoztam. A hangjel akkor magashangú impulzushang (taps) volt, és a plasztilin műfülkagylókat előre és hátra irányítva emberi fül elé illesztettük. Ebben a tájékozódó jellegű binaurális vizsgálatban az derült ki, hogy az elől-hátul érzékelés a fülkagylók irányának megfordításával erre a hangjelre szintén megfordul. A jelen munkában monaurálisan azt kívántuk tisztázni, hogy ekkor is vannak-e olyan frekvenciatartományok, amelyek előnyösen befolyásolják az elől-hátul hangérzékelést. Ezt binaurálisan Blauert (1974) igazolta, a két fül szimmetriája miatt

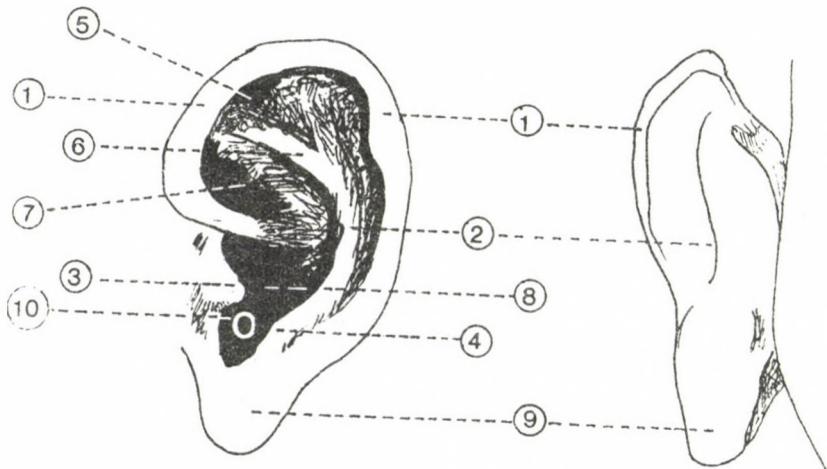
azonban az a helyzet előnyösebb. Saját kísérletünk magyarázatához tekintsük a 10. ábrát. A tarkó felől ( $\varnothing = -90^\circ$ ) érkező relatív színekben ilyen helyeket nem találunk, a homlok irányából ( $+90^\circ$ ) érkezőben azonban kettőt is: egy határozottabbat 3 kHz-en és egy kevésbé meggyőzőt 12 kHz-nél. Világosabbá válik a helyzet színekpek kivonása után (alsó diagram). A különbozeti színekpek elemzése szerint a szemből érzékelésnek három hangsáv kedvez: 400-1500 Hz, 2-6 kHz és 12-14 kHz. Viszont két hangsáv hátulról érkező hang hatását érzékelteti. Ezek a helyek nem olyan szélesek és 1,5 kHz és 8 kHz környéken található. Úgy képzelhető, mintha bizonyos jól definiált frekvenciasávok jelenléte vagy uralkodása befolyásolná az érzékelés irányát. Blauert egyenesen iránymeghatározó sávokról beszél, mivel kétfülű kísérleteiben hallgatóinak nagy része bármely irányból érkező határozott frekvenciájú hangsávokra nagy százalékban "elő", "hátul" vagy "főnt" érzettel válaszoltak.

### Kitekintés

Az utolsó helyen említett összehasonlítás fölveti az egyfülű és kétfülű vizsgálatok eltérő jellegéből folyó különböző lehetőségeket. A monaurális kísérletek eredményesebben tisztázhatják a fülkagyló hallásmeghatározó szerepét. Az eredményeknek nem csak az egyik fülre nagyothalló veheti hasznát, hanem a két fülét más-más célra igénybe vevő tolmács, szövegfordító vagy akár telefonkezelő is. Mivel azonban a normális hangérzékelés kétfülű, általános esetekben két, egymástól eltérő információt kell a két fülnek egyeztetni vagy összehasonlítani. Ehhez szolgáltat adatokat a monaurális vizsgálatok sora. A kétfülű elől-hátul-érezékelés szélsőséges eset, mert ekkor sem a kétfülű idő-, sem az intenzitáskülönbség nem működik, ezért jut nagyobb szerephez a jellegzetes frekvenciasávok iránymeghatározó hatása. Más esetekben azonban sokkal bonyolultabbá válik a kétfülű érzékelés. Ha például  $\varnothing = +60^\circ$  irányból érkezik az egyik fülbe, akkor a másik

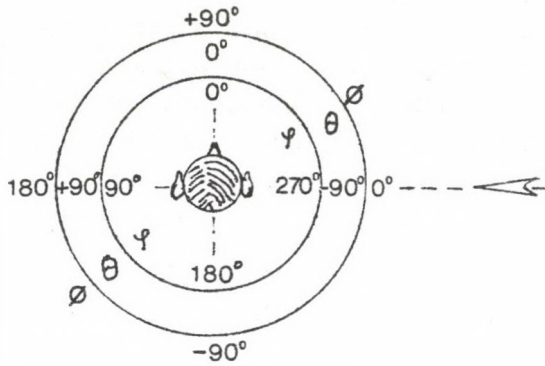
fülben úgy hat, mintha  $+120^\circ$  irányból érkezett volna. A kétféle frekvenciatorzító hatás mellett idő- és intenzitáskülönbség is bonyolítja a kétféle jel együttes és összehasonlító feldolgozását. Nem utolsó sorban figyelemmel kell lennünk arra a tényre is, hogy ugyanaz a hangintenzitás két füllel érzékelve 1,3-1,7-szeres hangosságot jelent, ami még szintén intenzitás- és frekvenciafüggő.

Meggondolásainkat figyelembe véve egyelőre nehezen remélhető, hogy akár az egyfülű, akár a kétfülű hallás kísérleti adatai tisztáznák, hogy mennyiben készíti elő a külső fül az agy számára a teljes beszédmegértéshez szükséges alapadatokat.



1. ábra

A bal fülkagyló (pinna) áttekintő szerkezeti elemei. 1: fülsegély (helix), 2: belsegély (antihelix), 3: fülcsap (tragus), 4: fülcsoros (antitragus), 5: háromszögű árok, 6: a belsegély szára, 7: ovális árok, 8: a fülkagyló ürege (cavum conchae), 9: fülcimpa, 10: a külső hallójárat nyílása (meatus acusticus externus)



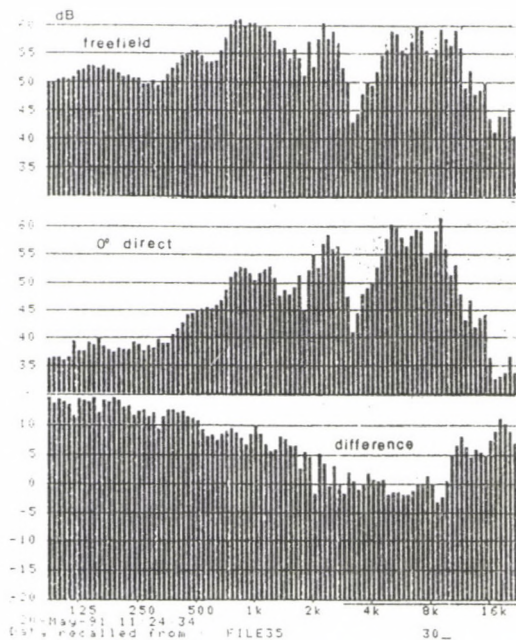
2. ábra

Íránymegjelölés:  $\varphi$  (Blaubert),  $\theta$  (Shaw),  $\emptyset$  (saját)

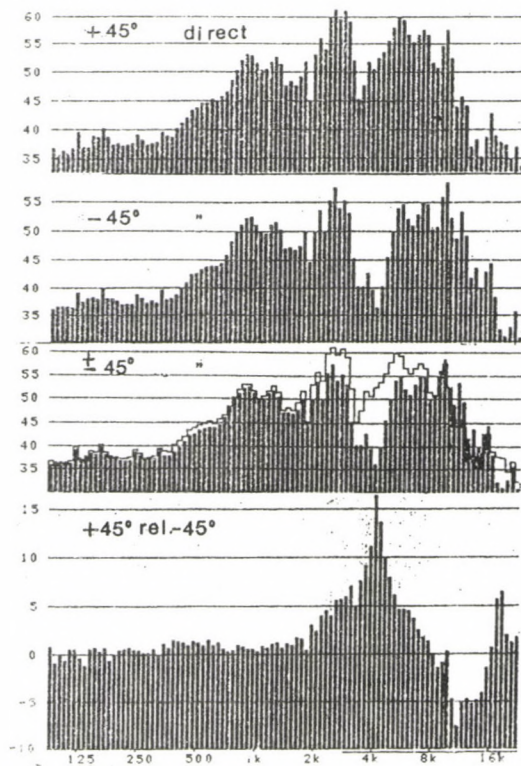


3. ábra

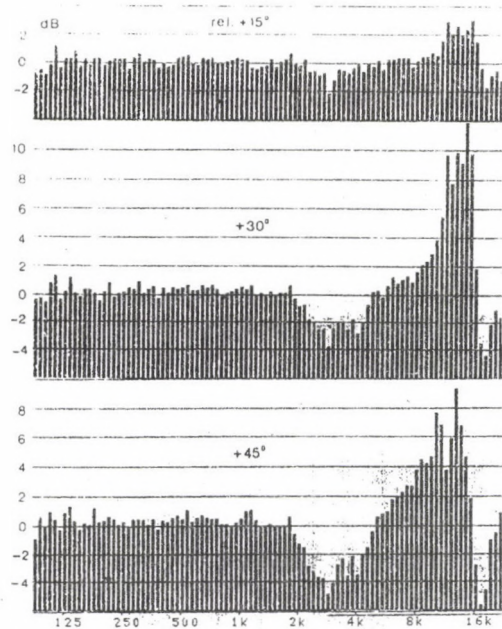
Felöltöttetett műfüles bábu a forgóasztalon



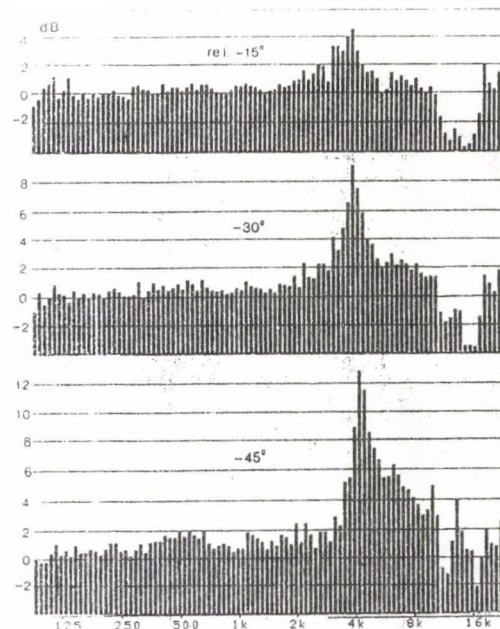
4. ábra: Három diagram az azonosidejű elemző szinképeiből. Felső kép: szabadtéri fölvétel a 3 m távolságból sugárzó hangforrás "rózsaszínű zörej" hangjáról. Középső kép: ugyanezen hang szinképe szemből érkező hangra a műfül dobhártyája helyén. Alsó kép: a két előbbi szinkép kivonásának eredménye. Magyarázat a szövegben



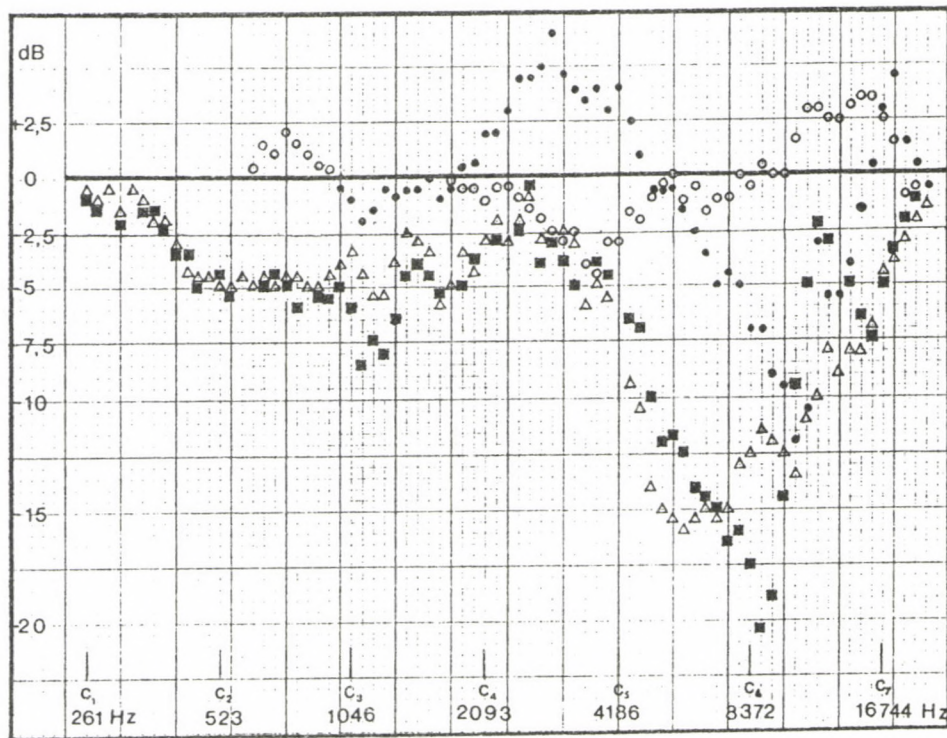
5. ábra: Néhány főirányból közvetlenül a műfüllel fölvett hangszinkép. Csak a szabadtéri adatból kivonva értelmezhetők, de egymással összehasonlíthatók, az összehasonlításuk lehetősége is látható. (Vigyázat, a dB-skálák nem mindenütt azonosak.)



6. ábra: Relatív (a  $0^\circ$  irányú színeképből kivont) színeképek pozitív ferdeszögű irányokból. Megjegyzendő, hogy a negatív dB-értékek intenzitás-gyarázkodást, a pozitívok intenzitás-gyengülést jelentenek az illető irányban

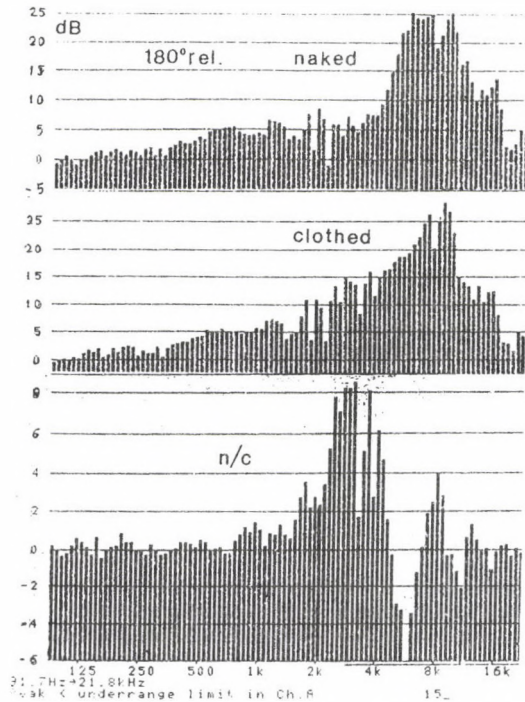


7. ábra: Mint a 8. ábra, de negatív, vagyis a fülkagylótól hátrafelé irányuló ferdeszögű irányokból

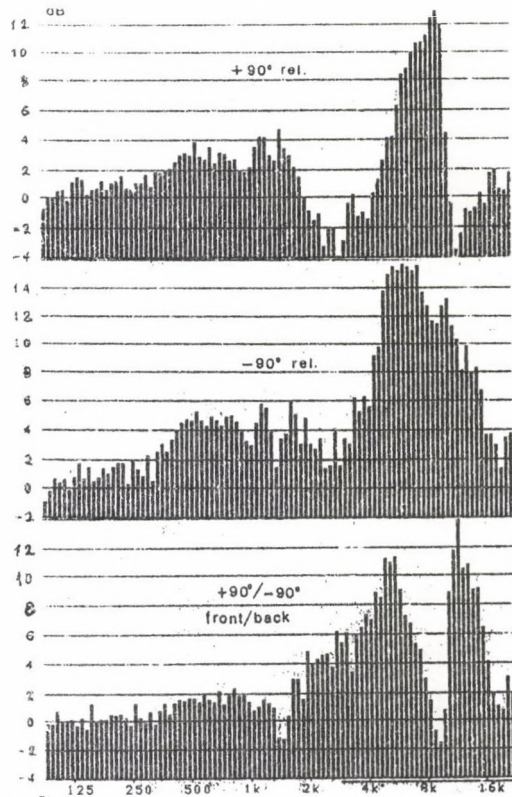


8. ábra

Néhány irányból érkező hangjel relatív szintjeinek alakulása a frekvencia és az irány mint paraméter függvényében. Telt kör: +60°, üres kör: -15°, üres háromszög: -90°, telt négyszög: +120°



9. ábra: Az öltözet hatása az öltözetlen bábuval szemben ( $180^\circ$  irányban)



10. ábra: Az elől-hátul érzékelés kimutatása  $+90^\circ$  és  $-90^\circ$  relatív szinképeinek kivonásával. Részletek a szövegben

**Irodalom**

Blauert, J.: Ein Beitrag zur Trägheit des Richtungshörens in der Horizontalebene. *Acustica* 20. 1968, 200-206.

Blauert, J.: Räumliches Hören. Hirzel Vg. Stuttgart 1974.

Blauert, J.: Räumliches Hören. Nachschrift. Hirzel Vg. Stuttgart 1985.

Jahn, G.-Vogelsang, S.: Das einohrige Richtcharakteristik des menschlichen Gehörs. *Hochfr. technic Elektroakust.* 68. 1959, 50-56.

Kietz, H.: Das räumliche Hören. *Acustica* 3. 1953, 73-86.

Schirmer, W.: Die Richtcharakteristik des Ohres. *Hochfr. technic Elektroakust.* 72. 1963, 39-48.

Shaw, E.A.G.: Earcanal pressure generated by a free sound field *JASA* 39. 1968, 465-470.

Shaw, E.A.G.-Terenishi, R.: Sound pressure generated in an external ear replica and real human ears by a nearby sound source. *JASA* 44. 1968, 240-249.

Shaw, E.A.G.: Hearing Threshold and ear canal pressure generated by a free sound field with varying acoustic field. *JASA* 46. 1969, 1502-1514.

Shaw, E.A.G.: Transformation of sound pressure level from the free field to the eardrum in the horizontal plane. *JASA* 56. 1974, 1848-1861.

Shaw, E.A.G.: The acoustics of the external ear. In: *Acoustical factors affecting hearing aid performance*. Ed: Studebaker, G.-Hochberg, J. Univ. Park Press, Baltimore 1980, 109-125.

Tarnóczy, T.: Über den Vorwärts-Rückwärts Eindruck. *Acustica* 8. 1958, 343.

Tarnóczy, T.: Über den Verstärkungs-Verminderungs-Effekt der Ohrmuschel und des Kopfes. *Proc. FASE-Congr. Zürich* 1992, 229-232.

Wiener, F.M.-Ross, D.A.: The pressure distribution in the auditory canal in a progressive sound field. *JASA* 18. 1946, 401-408.

Wiener, F.M.: On the diffraction of a progressive sound wave by the human head. *JASA* 19. 1947, 143-146.

## SZÜKSÉGES ÉS SZÜKSÉGTELEN HANGÁTMENETEK

Gósy Mária  
MTA Nyelvtudományi Intézete

A beszédhangok ún. tiszta fázisai között megjelenő hangrészletek, a hangátmenetek funkcióját vizsgálja a tanulmány a beszédészlelés szempontjából. A kísérleti személyek a hangátmeneteiktől megfosztott természetes ejtésű CV-hangkapcsolatok mássalhangzóit néhány kivételtől eltekintve helyesen azonosították. Az eredmények azt bizonyítják, hogy a hangátmenetek nem feltétlenül szükségesek a beszédhangok felismeréséhez, funkciójuk egyértelműen hangminőség- és kontextusfüggők.

### Bevezetés

A fonetikának, pontosabban magának a kísérleti fonetikának a forradalmi megújulását, illetőleg kiteljesedését jelentette az akusztikai fonetikának mint e tudomány szerves részének integrálódása az évszázad közepétől kezdődően. A beszéd akusztikai elemzésének eredményei nem csupán új megvilágításba helyezték a beszédről addig alkotott ismereteket, hanem egy sor olyan kérdés is megmagyarázhatóvá vált, amely a tradicionális elméleti kereteken belül és az addig alkalmazott módszerekkel nem volt lehetséges. Több, mint 200 évvel ezelőtt Kempelen Farkas – a beszélőgép építése során – tulajdonképpen "felfedezte" és leírta a hangátmeneti jelenséget. Nem elemezte azonban olyan tudományos igényvel, mint a beszédképzés más vonatkozásait (Kempelen 1791). Az elfogadható hangzás helyett a két létrehozott hang egybefolyását, együtthangzását, illetőleg az azok közötti szünet megjelenését tárgyalta szótag, szó létrehozásakor. Kempelen tehát az észlelés révén jut el ahhoz az egyértelműen nem megfogalmazott felismeréshez, hogy két hang nem illeszthető egymáshoz úgy, mint két betű az írásban.

A beszédképzés és a beszédmegértés folyamatosságának (analóg voltának) és a beszédhangok elkülöníthetőségének és kijelölhetőségének ("diszkrét" voltának) paradoxona megoldódni látszott a hangátmenetek "felfedezésével" (vö. pl.: Liberman 1957). A koartikuláció akusztikai következményei, az átmeneti hangrészek vagy hangátmenetek biztosítják a beszéd folyamatosságát (vö. az artikulációra Kantner és West 1933, ill. az akusztikumra pl.: Kent és Read 1992). A hangátmenetek akusztikai elemzése nem kevés nehézséget rejtett magában. Meg kellett határozni a hangkapcsolatok – hagyományos értelemben vett – magán- és mássalhangzói közötti eltérő időtartamú hangrészek hovatartozását (ti. a magánhangzóhoz vagy a mássalhangzóhoz történő "elvi" illesztést), funkcióját és paramétereit (pl. a "hub" és a "locus" elméletek: Lehiste és Peterson 1992). Tekintettel a metodológiai korlátokra, a kezdeti eredmények meglehetősen bizonytalanok voltak és szinte kizárólag a felpattanó zárhangokra vonatkoztak.

A mesterséges beszédelőállítás (beszédszintézis), majd az ennek nyomán lehetővé vált percepciós kísérletek eredményei kibővítették és pontosították az átmenetekről kialakított ismereteket. Lehiste és Peterson többféle definíciót is elfogad: (i) a hangátmenet olyan frekvenciamozgás, amely egy mássalhangzó-magánhangzó kapcsolatban a magánhangzó kezdetétől a magánhangzóra jellemző akusztikai pozíció bekövetkeztéig tart vagy (ii) a hangátmenet olyan formánsmozgás, amely egy tiszta fázistól egy következő tiszta fázisig tart avagy (iii) olyan mozgás, amely egy adott magánhangzó-mássalhangzó kapcsolatban a magánhangzó tiszta fázisától ugyanazon magánhangzó befejező pontjáig tart (i.m. 295).

Az eredmények évtizedek óta alátámasztják azt is, hogy a hangátmenetek nélkül a beszédhangok felismerése nem lehetséges avagy bizonytalan (elsődleges vagy másodlagos akusztikai kulcsok: Whalen 1991; a magyarra: Gósy 1989).

Mindezen eredmények és megállapítások ellenére sem a nemzetközi, sem a hazai szakirodalomban nem találunk olyan kísérletsorozatot, amely szisztematikusan vizsgálná meg a hangátmenetek beszédpercepciók funkcióját. Éppen ezért kísérletet terveztünk annak megállapítására, hogy a hangátmenetek a beszédészlelés szempontjából valóban szükséges avagy szükségtelen részei-e a beszédnek.

### **Anyag és módszer**

Előzetes kísérletek alapján 11 mássalhangzót és 3 magánhangzót választottunk a kísérlet hanganyagául: a 4 felpattanó zöngés zárhangot, a dentális és alveoláris zöngés és zöngétlen spiránsokat, valamint a laterális mássalhangzót, illetőleg az [a:, i:, ø:] magánhangzókat. Mindhárom magánhangzónak az összes mássalhangzóval alkotott CV kapcsolatot elemeztük természetes (női) ejtés alapján. A 33 hangkapcsolatot Kay 4300B digitális jelfeldolgozóval vizsgáltuk 10 kHz-es mintavételezési frekvenciával, 8000 Hz-es elemzési tartományban. Az optikai kép és az ismételt akusztikai információ összevetésével szubjektív döntés alapján kivágtuk a CV-kapcsolatok átmeneti részét. Az átmenetektől így módon megfosztott hangkapcsolatokat kétféle véletlenszerű sorrendben magnetofonszalagra rögzítettük és két kísérletsorozatot végeztünk éphalló, magyar anyanyelvű felnőtt kísérleti személyekkel, összesen 22 fővel (átlagéletkoruk 32 év volt).

Az első kísérletsorozat célja annak megállapítása volt, hogy vajon a részt vevő mássalhangzók felismerhetők-e vagy sem. A második kísérletsorozattal arra kerestünk választ, hogy vajon a mássalhangzók minősége torzul-e az átmenetek hiányakor, függetlenül a felismerés helyességétől.

### **Eredmények**

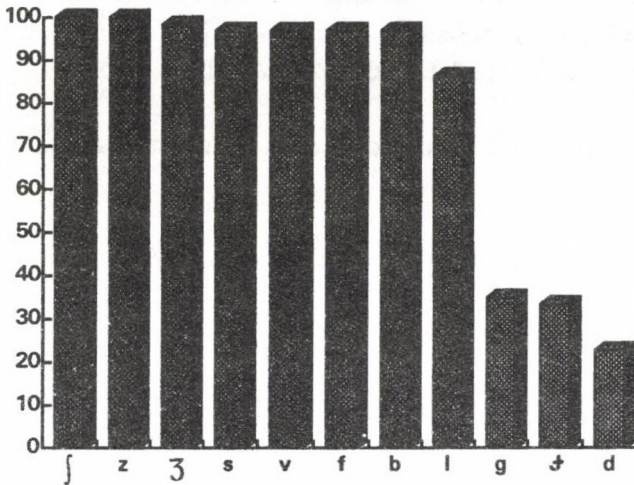
A hangátmenetek, mint az artikuláció akusztikus lenyomatai, az adott hangminőségekre jellemzőek; időtartamuk különböző. Az

eredeti hangkapcsolatokból kivágott hangátmenetek időtartamértéke az [a:]-val alkotott hangkapcsolatokban átlagosan 114,8 ms, az [ø:] -vel alkotottakban 86,2 ms, az [i:] -vel alkotottakban pedig 95,9 ms (a részletes adatok az 1. táblázatban).

1. táblázat: A CV hangkapcsolatok átmeneteinek időtartamértékei

Mássalhangzó- minőség	Az átmenet időtartama (ms)		
	V=a:	V=ø:	V=i:
b	131	109	94
d	161	136	138
ʃ	91	97	69
g	132	25	70
v	82	83	98
f	144	99	100
z	124	100	90
s	116	70	136
ʒ	105	87	98
j	106	71	74
l	71	72	88

A hangkapcsolatokra kapott azonosítási eredmények azt mutatják, hogy valóban vannak az észlelés számára "szükséges" és vannak "szükségtelen" átmenetek. A 11 vizsgált mássalhangzóból kettőnek a felismerése 100%-os, további öté pedig 90%-on felüli, vagyis ezen mássalhangzók esetében a hangátmenetek perceptíósan szükségtelenek (vö. 1. ábra). A laterális mássalhangzó felismerése is jó (86,3% a három magánhangzóval alkotott hangkapcsolat átlagát tekintve); a felpattanó zárhangok közül háromnak a felismerése azonban gyakorlatilag nem lehetséges az átmenet nélkül (korrekt azonosításuk 22,7% és 34,8% közötti az átlagokat tekintve), vö. 1. ábra.



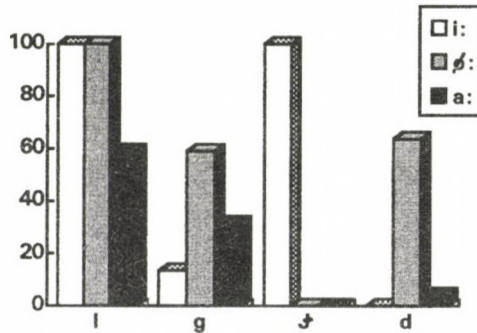
1. ábra

Mássalhangzók helyes felismerése hangátmenetek nélkül (a három magánhangzóval alkotott CV-kapcsolatok átlagában)

A percepció eredmények alapján felállítható a vizsgált mássalhangzók azonosíthatósági sorrendje, illetőleg – más megfogalmazásban – a hangátmenetek szükségessége szerinti sorrend, amely a következő (a leghelyesebben azonosíthatótól a nem azonosíthatóig): [j, z, ʒ, s, v, f, b, l, g, ʃ, d]. A 7 jól azonosítható mássalhangzó ([j, z, ʒ, s, v, f, b]) felismerése gyakorlatilag független a magánhangzó minőségétől. A 21 hangkapcsolatból 15-nél 100%-os a helyes azonosítás; 95,4% a *zső*, a *ff*, a *fő* és 90,9%-os a *szá*, a *vő* és a *bá* pontos észlelése.

A négy nem jól azonosítható mássalhangzó ([l, g, ʃ, d]) felismerése erősen kontextusfüggő; ez jelen esetben azt jelenti, hogy a helyes felismerés – az átmenet hiányában – a követő ma-

gánhangzó minőségének függvénye. A 2. ábra ezeket a különbségeket és függőségi viszonyokat szemlélteti.



2. ábra

Mássalhangzók helyes felismerése a követő magánhangzó függvényében

Az adatok azt mutatják, hogy a veláris magánhangzó előtt gyengébb a felismerés, mint a palatálisok előtt. A kontextushatás legerősebben tehát az [a:] és legkevésbé az [i:] magánhangzóval alkotott CV-kapcsolatban érvényesül. A hangátmenetet "nem igénylő" mássalhangzók felismerése az [a:]-val alkotott hangkapcsolatban 97,4%, az [ø:] -vel alkotott hangkapcsolatban 97,3%, az [i:] -vel alkotott hangkapcsolatban pedig 99,3%. A velárisal alkotott CV-kapcsolatban a hangátmenetet "igénylő" mássalhangzók felismerésének átlaga 23,5%; a palatális labiálissal alkotott kapcsolatban 55,6%; az illabiális palatálissal alkotott kapcsolatban pedig 53,4%. Az összes CV-kapcsolatban elemezve a V-től való függőséget, szignifikáns különbséget látunk a veláris és a két palatális magánhangzó hatásaként ( $p < 0.001$ ). A korrekt mássalhangzó-felismerés átlaga 82,6% az illabiális palatális, 82,2% a labiális palatális és 70,6% a veláris magánhangzó mellett.

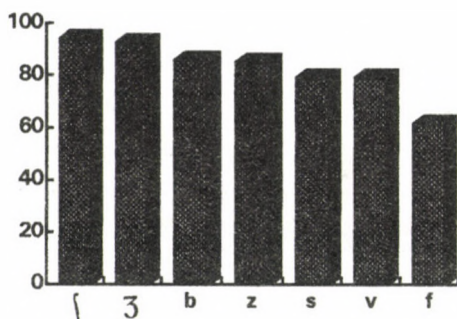
Elemeztük a tévedéseket, azaz hogy a kísérleti személyek milyen mássalhangzókkal tévesztették össze az eredeti beszédhangokat. A 2. táblázat a tévesztési mátrixot mutatja.

2. táblázat: A nem (jól) azonosítható mássalhangzók tévesztési mátrixa százalékban

Eredeti mássalhangzó	Azonosított mássalhangzó (%)				
	d	ʃ	g	b	0
d	22,7	-	-	75,7	1,5
ʃ	75,7	33,3	6,06	4,5	-
g	1,5	-	34,8	53	7,5

A zárhangokat – amint az várható volt – az ugyanazon képzésmódú osztályon belül tévesztik a kísérleti személyek. Legnagyobb mértékű a bilabiális zárhang helyettesítő megjelenése; a palatális zárhang helyett pedig perceptiósan elfogadható mértékben (!) azonosítanak dentális zárhangot. Meglepően alacsony a felismerhetetlennek minősített mássalhangzók aránya, vagyis a hangátmenet hiánya megváltoztathatja a mássalhangzó minőségét a felismerés számára, bizonytalanná teheti a döntést, de a beszédhang-észlelet megmarad. A laterális mássalhangzót ([l]) csupán a veláris magánhangzóval alkotott hangkapcsolatban nem lehetett egyértelműen azonosítani, itt a kísérleti személyek 41%-ban egyöntetűen bilabiális nazálist ([m]) ismertek fel.

A második kísérletsorozatban a kísérleti személyek feladata az elhangzott hangkapcsolatok minősítése volt a magyar iskolarendszerben használatos érdemjegyekkel (1-essel a legrosszabbnak, 5-szel a legjobbnak ítélt "ejtést" kellett minősíteniük, természetesen mind az öt osztályzati lehetőséget kihasználva). A 7 jól azonosított mássalhangzó esetében azt tapasztaltuk, hogy a kísérleti személyek a korrekt felismerés arányánál "alacsonyabbra" értékelték a hangok minőségét. A 3. ábra szemlélteti e minősítési eredményeket.

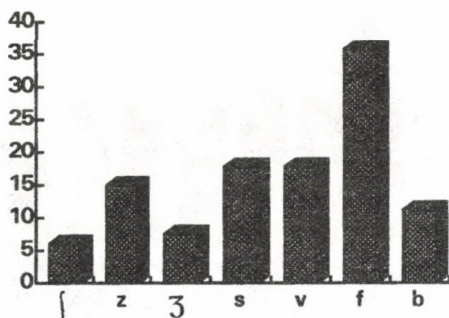


3. ábra

A jól azonosított hangkapcsolatok minősítésének grafikonja

Az adatok a hangátmenetek mintegy harmadlagos akusztikai kulcs funkcióját tükrözik: bár az adott mássalhangzókat a kísérleti személyek nehézség nélkül azonosították, a "minőségüket" azonban nem feltétlenül fogadták el. Egyértelműen jónak ítélték a két alveoláris réshangot, valamivel gyengébbnek találták a bilabiális felpattanó zárhangot és a dentális zöngés spiránst. Az alveoláris réshangokhoz képest lényegesen gyengébb a [s] és a [v] mássalhangzók minősítése. Legkevésbé elfogadhatónak, tehát a korrekt felismerhetőség mellett legtorzabbnak a [f] mássalhangzó bizonyult.

Elemeztük a felismerés és a minősítés matematikailag kifejezhető különbségét (4. ábra).



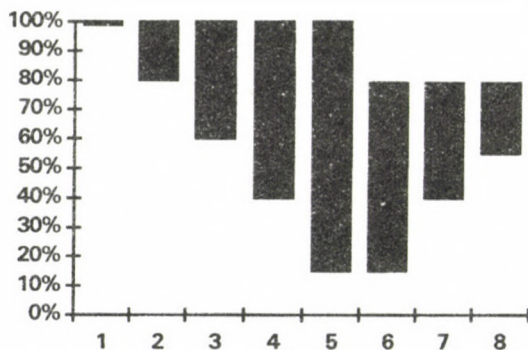
4. ábra

## A helyes azonosítás és a minősítés összefüggése

Az ábra szemléltette különbségeket jelzik, hogy az átmenetektől megfosztott hangkapcsolatok akusztikai megvalósulása nem egyezik az agyban tárolt neurális spektrogrammal. A négy nem (jól) azonosítható mássalhangzó esetében nehézkes a minősítések értékelése, hiszen a kísérleti személyek azokat a beszédhangokat minősítették, amelyeket észlelni véltek. A [ʃ] mássalhangzó [d]-ként történt felismerése igen jó minősítéssel járt együtt: 82,2% és 93,3%. A [ʃ] mássalhangzó csak az [i:] környezetében volt azonosítható, ekkor 100%-osan, ennek ellenére a *gyí* hangkapcsolat minősítése csupán 71,1%-os. Az eredeti *lá* hangkapcsolatot már csak 75,5%-ban tartották elfogadhatónak. A veláris zárhang minősítése [a:] mellett 31,1%-os, [ø:] mellett 77,7%-os és [i:] mellett 57,7%-os, vagyis nagyjából korrelál a felismerési adatokkal. Végül a dentális zárhang minősítése 66,6% mind az [a:], mind az [ø:] magánhangzóval alkotott kapcsolatban (a zárhang azonosítása nagy arányban [b]) és 82,2% az eredeti *dí* esetében (itt azonban 100%-os a *bí* felismerés).

A hangkapcsolatok mássalhangzó-azonosításait elemezve azt láttuk, hogy a kísérleti személyek ítéletei nagymértékben egye-

zőek; a minősítés azonban – tekintettel annak lényegesen szubjektívebb voltára – nagyobb szórást mutat. A legnagyobb egyértelműség a *lí* hangkapcsolat mássalhangzójának megítélésében volt; két osztályzatnyi eltérést tapasztaltunk az alveoláris réshangok valamennyi magánhangzóval alkotott kapcsolatában (százalékban kifejezve: 80-100%) és a *ví* esetében (százalékban: 60-80%). 10 mássalhangzó-kapcsolatnál – például *bá, szí, lő* – három osztályzat eltérést is találtunk (százalékban: 60-100%), illetve a *ří*-nél 40-80%. Négy osztályzatnyi különbséget mutatott a két dentális réshangra kapott ítéletek elemzése (40-100%). A legnagyobb szórást a *fő* (20-80%) és a *fá*, illetve a *gyí* (20-100%) hangkapcsolatok minősítésében tapasztaltuk. Ez utóbbiak egyértelműen a kísérleti személyek egyéni toleranciájával magyarázhatók; továbbá arra a metodológiai sajátosságra is utalnak, hogy nem egy esetben nehéz a minősítést elvégezni, ha az azonosítás nem okoz problémát (5. ábra).



5. ábra

A hangkapcsolatok minősítésének szórása százalékban  
 (1=*lí*, 2=*sá, ző, zső, sí, zsí, ső, zsá*, 3=*bá, bí, bő, vá, vő, szí, sző, zí, lő*, 4=*zá, szá*, 5=*fá, gyí*, 6=*fő*, 7=*ří*, 8=*ví*)

### Következtetések

E tanulmány címe előjelezte annak lehetőségét, hogy a hangátmenetek esetleg szükségtelenek. A kísérleti eredmények alapján egyértelműen kimondható, hogy 1. a hangátmenetek percepciósi funkciója változó, valamint hogy 2. ez a funkció hangminőség- és kontextusfüggő. Az első megállapítás annak tömör megfogalmazása, hogy vannak olyan hangkapcsolatok, amelyek biztos felismeréséhez a hangátmenet nem szükséges, vagyis a beszédhangok tiszta fázisai közötti hangrészlet nem elsődleges felismerési kulcs. Az a szakirodalmi állítás, hogy a felpattanó zöngés zárhangok felismerése nem lehetséges az átmenetek nélkül, a jelen kísérlet eredményei szerint szintén **revízióra szorul**. A magyarra vonatkozóan legalábbis "szükségtelen" a hangátmenet a bilabiális felpattanó zöngés zárhang, részben a palatális és a veláris zöngés zárhang azonosításához. A **hangátmenetek tehát részben valóban elsődleges akusztikai kulcsok** a percepcióban, nem egy esetben azonban **másodlagos felismerési kulcsok** (például a labiodentális réshangoknál), sőt több esetben csak **kiegészítő szerepük** van (például a dentális és alveoláris réshangoknál). Ez utóbbi esetekre kimondható, hogy a **hangátmenetek szükségtelenek**.

A neurális spektrogramokkal történt minőségi összevetés egyértelműen jelzi, hogy a hangátmenetek "**esztétikai**" többletet jelentenek a beszédhangok felismerésében. Mindezek alapján felvetődik a **hangátmenetek definíciójának újrafogalmazása** is. Hangátmenetnek nevezünk a koartikulációs jelenségek akusztikai következményeit, amelyek a formánsmozgások révén a mássalhangzók minőségétől és kontextusától függően vagy invariáns jegyként, vagy másodlagos akusztikai kulcsként, nem egy esetben azonban redundáns tényezőként funkcionálnak az észlelésben.

**Irodalom**

Gósy Mária: Beszédészlelés. MTA Nyelvtudományi Intézete. Budapest 1989.

Kantner, C.E.–West, R.: Phonetics. Harper and Brothers. New York 1933.

Kempelen, W. von: Mechanismus der menschlichen Sprache nebst Beschreibung seiner sprechenden Maschine. Wien 1791.

Kent, R.D.–Read, Ch. (eds.): The Acoustic Analysis of Speech. Singular Publishing Group. San Diego 1992.

Lehiste, I.–Peterson, G.E.: Transitions, glides, and diphthongs. In: Readings in Clinical Spectrography of Speech. Eds.: Baken, R.J.–Daniloff, R.G. Singular Publishing Group and Kay Elemetrics Corp. San Diego 1992, 286-295.

Lieberman, A.M.: Some results of research in speech perception. JASA 29. 1957, 117-123.

Whalen, D.H.: Perception of the English /s/ – /ʃ/ distinction relies on fricative noises and transitions, not on brief spectral slices. JASA 90. 1991, 1776-1785.

## MONDATHANGSÚLY A MAGYARBAN

Hunyadi László

KLTE Általános és Alkalmazott Nyelvészeti Tanszék

Több fonológiai és szintaktikai nyelvelmélet is alapvetően támaszkodik a hangsúly fogalmára. A hangsúlyra olyan feltevésekkel hivatkoznak, amelyeknek lényegi következményei vannak az adott elmélet szempontjából. Bár nyilvánvaló, hogy a hangsúly fontos akusztikai tulajdonságokkal rendelkezik, ezen tulajdonságok összetett volta miatt a hangsúlyra való hivatkozásainkban sokkal inkább a percepcióra támaszkodunk, mint a konkrét, parametrizált mérési eredményekre. A jelen dolgozatban egy olyan megközelítést mutatunk be, ami lehetővé teszi a magyar mondathangsúly helyének és intenzitásának műszeres kimutatását. Ez egyben a hangsúly fogalmára támaszkodó elméleti megállapításaink verifikálhatóságát is lehetővé teszi.

### Bevezetés

A hangsúly fogalmára gyakorta történik hivatkozás mind az elméleti nyelvészet, mind az alkalmazott nyelvészet számos területén. Az olyan nyelvekben, mint az angol vagy az orosz, komoly jelentősége van olyan szabályrendszer kidolgozásának, amely leírja azokat az elsősorban fonológiai és morfológiai feltételeket, amelyek mellett az egyes szavakon belül a hangsúly helye meghatározható. A magyarban, ahol a szavakat lényegében az első szótagjukon hangsúlyozzuk, a szóhangsúlynak kisebb elméleti jelentősége van. Ezzel szemben lényegi kérdések merülnek fel a magyar mondatszerkezet leírása során azzal kapcsolatban, hogy mely szavak milyen hangsúlyozottsággal rendelkeznek. A magyarban tehát, amikor hangsúlyról beszélünk, vizsgálatunk tárgyának elsősorban a mondathangsúlyt tekintjük.

A magyar mondatszerkezet leírásában (vö. É. Kiss 1978, 1994) a szintaktikai pozíciók meghatározásánál alapvető hivatkozás történik az egyes szavak hangsúlyosságára. Így, egy nem-neutrális mondatban az F-pozícióban álló fókuszált szó mindig hangsú-

lyos és a mondat főhangsúlyát is viseli, ha az előtte levő Q-pozíció nincs kitöltve (vö. 1a). Ha a Q-pozíció ki van töltve, akkor az ebben a pozícióban álló kvantor viseli a mondat főhangsúlyát (vö. 1b). A T-pozícióban álló topikalizált elem pedig mindig kisebb hangsúllyal rendelkezik, mint a Q- vagy az F-pozícióban álló elemek (vö. 1c):

(1a) [FPéter olvasta el a könyvet.]

(1b) [QMinden könyvet [FPéter olvasott el.]]

(1c) [TA könyvet [FPéter olvasta el.]]

A hangsúlyra való hivatkozásnak nem csupán a szintaktikai pozíciók meghatározásánál van jelentősége, hanem a mondat logikai szerkezetének a leírásánál is. Míg természetesnek tűnik az az elv, miszerint a felszíni sorrend egyben logikai hatóköri viszonyokat is kifejez, ezen elvnel általánosabb a logikai hatókörök kifejezése a hangsúlyok megfelelő elrendezésével (vö. Hunyadi 1981). Ezt mutatja az, hogy míg (1b) példa a lineáris sorrend szerinti hatókör kifejezésére, addig az ezzel logikailag ekvivalens (1d) ennek ellentmond:

(1d) 'Péter olvasott el 'minden könyvet.

Azaz a felszíni sorrendtől függetlenül mind (1b)-ben, mind (1d)-ben az univerzális kvantor-kifejezésnek (*minden könyvet*) nagyobb a hatóköre, mint a *Péter* fókusznak. Kitűnik, hogy a két mondatban valójában az az azonos, hogy a nagy hatókörű univerzális kvantor-kifejezés mindkettőben főhangsúlyos, szintaktikai pozíciótól függetlenül. A hangsúlynak a lineáris sorrenddel szembeni meghatározó szerepére utal az is, hogy azonos felszíni sorrend, de különböző hangsúlyozás mellett a hatóköri viszonyok különböznek. Ezt látjuk a fenti (1d) és (1e) összevetéséből:

(1e) 'Péter olvasott el minden könyvet.

Az utóbbi (1e) mondatban a főhangsúlyt viselő fókusznak (*Péter*) nagyobb a hatóköre, mint a hangsúlytalan univerzális kvantor-kifejezésnek (*minden könyvet*). Ez arra utal, hogy a szin-

taktikai sorrendnél is nagyobb jelentősége van a mondaton belüli hangsúly-eloszlásnak.

Egy másik nyelvelméletnek, a metrikus fonológiának kifejezetten alapfogalma a hangsúly. Ezen elmélet alapvető feladata az, hogy leírja azokat a szabályokat, amelyek segítségével meghatározható a hangsúly eloszlása szótagok és szavak között. Mivel a magyar szavak hangsúlya kötött és gyakorlatilag minden szónak egy hangsúlya van, lehetővé válik, hogy a szótagokra épülő metrikus fonológia alapelveit a magyar mondat szavai közötti hangsúlyelosztásra alkalmazzuk. A magyar mondatok egy ilyen elméleti leírásában (vö. Hunyadi 1995a) a hangsúlynak kitüntetett szerepe van. Az elmélet feladata azon szabályrendszer kidolgozása, amelynek a segítségével egy adott mondat összes lehetséges hangsúlyozási változata leírható, így pl. az olyan változatoké, mint (2a-c):

(2a) 'Péter olvasott el 'minden 'könyvet.

(2b) 'Péter olvasott el 'minden könyvet.

(2c) 'Péter olvasott el minden könyvet.

Mint a fenti elméleti megközelítések példáiból is kitűnik, a hangsúly fogalmának fontos elméleti státusza van. Mivel a hangsúly kézenfekvő akusztikai-fonetikai tulajdonságokkal is rendelkezik (a hangsúlyra vonatkozó állításainkat mindig a percepcióra való hivatkozással támasztjuk alá), kézenfekvőnek tűnik annak a vizsgálata, vajon melyek lehetnek a hangsúly azon akusztikai-fizikai tulajdonságai, amelyekre percepciószámításunkat alapozzuk.

A fenti elméleti kívánalmak mellett bizonyos alkalmazott nyelvészeti megközelítések, elsősorban a digitális beszéd-feldolgozás számára lehet különösen fontos a hangsúly akusztikai megfelelőinek a feltárása és leírása. Az itt következőkben azon törekvéseinkről kívánunk beszámolni, amelyeknek éppen a hangsúly akusztikai jellemzése a célja.

### **A hangsúlyról alkotott néhány felfogás**

A hangsúlyról számos felfogás alakult ki, melyek lényegében annak vagy a fiziológiai, vagy a percepciószámítás vonatkozásait helyezik

figyelmük középpontjába. Mivel jelenleg elsősorban a percepció akusztikai ‘fogódzóit’ kívánjuk megismerni, néhány ezzel kapcsolatos nézetet tekintünk át. (Egy szélesebb körű áttekintésre vö. Hunyadi 1995b.)

A hangsúly mibenlétének a kutatása különösen e század első harmadától kapott friss lendületet azáltal, hogy elérhetővé váltak az első műszeres vizsgálatok. A műszerek lehetővé tették a legfontosabbnak ítélt három paraméter, az alaphfrekvencia, az intenzitás és az időtartam vizsgálatát. Ezen paraméterek közül többnyire az alaphfrekvenciát tekintik a leginkább meghatározónak. Ladd (Ladd 1990) szerint a legtöbb európai nyelvben a dallamkontúrának a dallammal kifejezett mondathangsúly (a továbbiakban: dallam-hangsúly [*pitch accent*]) a legfontosabb összetevője. Beckman és Edwards (Beckman és Edwards 1990) szerint a dallam-hangsúly tesz egy szótagot a leghangsúlyosabbá. Hasonlóképpen, az alaphfrekvencia szerepét emeli ki Ladd, Verhoeven és Jacobs (Ladd, Verhoeven és Jacobs 1994) a dallam-hangsúly kialakulásában.

Míg valóban természetesnek tűnik az, hogy a dallam-változásnak jelentős szerepe van a hangsúly-érzet kialakulásában, a kép mégsem teljesen egyértelmű. Fónagy (Fónagy 1958) szerint a hangsúlyt nem lehet valamelyik paraméter kizárólagos funkciójának tekinteni. Varga (Varga 1983) ugyancsak azon a véleményen van, hogy a mondathangsúly a magyarban egy összetett jelenség, amelynek a kialakulásában mind az intenzitásnak, mind az intonációnak szerepe van.

Saját intuícióink is alátámasztják, hogy a dallam-hangsúly nem lehet a kizárólagos eszköze a hangsúly jelölésének, hiszen előfordulhat, hogy a hangsúlyos szótag jelölése nem jár együtt az alaphfrekvencia emelkedésével, időnként a csökkenésével sem.

Mivel a hangsúly akusztikai jelölőit kívánjuk közelebbről feltárni úgy, hogy egyben a korábban megfigyelt ellentmondásokra is választ keresünk, feltesszük, hogy minimálisan a fő paraméterek

(az alapfrekvencia, az intenzitás és az időtartam) együttesen képezik a hangsúly-percepció akusztikai alapját. Célunk egy olyan eljárás kidolgozása, ami e paraméterek komplex összefüggését tükrözi és megbízható módon segítséget nyújt a hangsúly akusztikai feltárásához.

### **A hangsúly mint relatív, akusztikai tényező**

Abból az alapfeltevésből indulunk ki, hogy a hangsúly észlelése során vannak jól leírható akusztikai "fogódzóink" és célunk ezek feltárása. Ugyancsak feltesszük, hogy a hanghullám legáltalánosabb paraméterei (alapfrekvencia, intenzitás és időtartam) együttesen egy összetett paramétert alkotnak, egy olyan paramétert, ami azok relatív viszonyát fejezi ki. E paramétert PET-nek fogjuk nevezni és arra lesz hivatott, hogy az alapfrekvencia és az intenzitás időbeli relatív relációját kifejezze (PET= "*pitch and energy over time*").

E paraméter segítségével arra a kérdésre kívánunk választ adni, hogy vajon az egyes mintavételi pontokban milyen mértékben vesz részt a hangsúlyos pozíció jelölésében az alapfrekvencia és az intenzitás. Nem e két paraméter abszolút értékei lesznek érdekesek számunkra, hanem az, hogy milyen mértékben változik meg az alapfrekvencia és az intenzitás saját összértékük viszonylatában. A PET-érték ezt a százalékban kifejezett relatív változást mutatja meg dinamikusan, a mintavételi pontok láncolatán keresztül.

Így a hangsúly fogalmát az alábbiakban határozzuk meg:

(3) A hangsúly az alapfrekvenciának és az intenzitásnak az idő függvényében kifejezett relatív relációja, melyet a PET-értékekkel jelölünk.

Mivel a PET-érték kiszámítása az alapfrekvencia és az intenzitás aktuális, kHz-ben és dB-ben kifejezett értékeinek viszonyításán alapszik, szükséges a két paraméter normalizálása. Ezt az alábbi módon érjük el:

(4) Az alapfrekvencia értékeinek a normalizálása

$$P_i = p_i \cdot 100 / \sum p_n, \text{ ahol}$$

$p$  = az alapfrekvencia aktuális értéke az  $i$  mérési pontban

$P$  = az alapfrekvencia százalékos értéke az  $i$  mérési pontban az összes alapfrekvencia-értékhez viszonyítva;

(5) Az intenzitás értékeinek a normalizálása

$$E_i = e_i \cdot 100 / \sum p_n, \text{ ahol}$$

$e$  = az intenzitás aktuális értéke az  $i$  mérési pontban

$E$  = az intenzitás százalékos értéke az  $i$  mérési pontban az összes intenzitás-értékhez viszonyítva;

(6) Az alapfrekvencia és az intenzitás dinamikus relációja (PET)

$$PET = P - E.$$

Az alapfrekvencia és az intenzitás dinamikus relációja tehát egy olyan érték, ami minden egyes mintavételi pontban megadja az alapfrekvencia és az intenzitás relatív viszonyát egyrészt saját összértékükhöz, másrészt egymáshoz viszonyítva. Ennek alapján egyrészt a  $P$  és az  $E$  értékből megtudhatjuk, hogy az adott mintavételi pontban az összes alapfrekvencia-értéknek, illetve az összes intenzitás-értéknek mekkora része (hány százaléka) vesz részt az adott pont akusztikai jellegének a kialakításában, másrészt, különbségük azt mutatja meg, hogy melyik paraméternek a relatív részesedése a nagyobb az adott mintavételi pontban.

Így, ha  $PET = 0$ , akkor ez azt jelzi, hogy az adott mintavételi pont akusztikai jellemzőinek a kialakításában a két paraméter azonos mértékben vesz részt. Ha  $PET > 0$ , akkor az alapfrekvenciának, ha  $PET < 0$ , akkor az intenzitásnak van nagyobb relatív jelentősége.

Mivel a  $PET$ -érték az alapfrekvencia és az intenzitás dinamikus relációját jelenti, ennek segítségével a két paraméter egymáshoz való viszonyát tudjuk az idő függvényében folyamatosan végigkísérni. A  $PET$  egy olyan változó, ami az idő függvényében tudja bemutatni a két paraméter relatív változását, így e változás folyamatosan nyomon követhető és egy koordináta-rendszerben

ábrázolható is. Ez utóbbi pedig lehetővé teszi, hogy – hasonlóan az alapprofrekvencia és az intenzitás ábrázolásához – vizuálisan is követhessük a hangsúly akusztikai paramétereinek a relatív változását.

### A kísérlet ismertetése

14 bemondóval készítettünk felvételeket 26 mondat minimális párpain úgy, hogy azok sorrendje véletlenszerűen volt megválasztva. Minden mondatához megadtunk egy rövid szövegösszefüggést, aminek az volt a célja, hogy egyértelművé tegyük a lehetséges hangsúlyozási mintát. A kísérlet során minden mondatot kétszeri bemondással vettünk fel. A felvételeket a KAY Elemetrics CSL 4300-as hangdigitalizálójával rögzítettük és dolgoztuk fel úgy, hogy előszűrést nem alkalmaztunk. Alapprofrekvencia- és intenzitásadataink a nyers adatokat tartalmazzák azzal a megszorítással, hogy ott, ahol a hanghullámnak nem volt numerikus alapprofrekvencia-értéke, ott az intenzitás-értékeket is eltávolítottuk.

A kísérletből itt csupán egyetlen mondatpár elemzésének az eredményeit mutatjuk be. Azon mondatpárét, amelyiknek a bemondása a mondatpár kézenfekvő logikai különbsége miatt minden kétséget kizáró módon egyértelmű volt. Az alábbi (7) és (8) mondatokban a felvételi sorszámozásban 19-esnek és 24-esnek jelölt mondatokat mutatjuk be a megadott szöveg-összefüggéssel együtt:

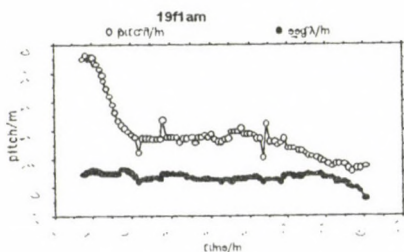
(7) Valamennyien eljöttek. (mindnyájan)

(8) Valamennyien eljöttek. (bár néhányan nem)

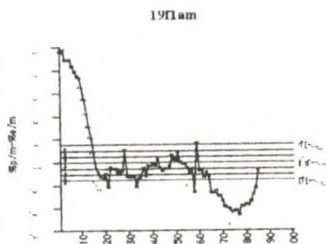
A kétféle és jól elkülöníthető értelmezésnek megfelelően azt vártuk, hogy (7)-ben a főhangsúly az univerzális kvantorra (*valamennyien*), míg (8)-ban az igére (*eljöttek*) fog esni és a kvantor kevésbé lesz hangsúlyos. Az eredményeket két bemondó alapján mutatjuk be.

## (9) 'Valamennyien eljöttek

(9a)



(9b)

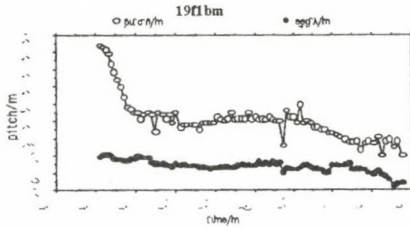


Közelebbről megvizsgálva az ábrákat a következőket látjuk:

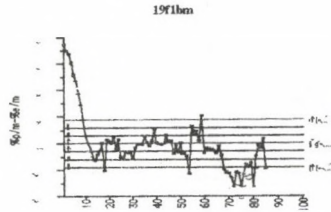
(9a) az ismert módon mutatja be az alapprofrekvencia- és az intenzitás-értékeket. A hangsúlyos *valamennyien* szón egy világos dallam-hangsúly van. A relatív reláció (PET) ábrájából (9b)-ben azt olvashatjuk ki, hogy a hangsúlyos szó egy relatíve magas alapprofrekvenciával kezdődik és az hirtelen átvált egy egyre fokozódó relatív intenzitás-növekedésbe. Itt és a továbbiakban a 3-as számok sorozata azt jelöli, hogy minden egyes 3-as szám előtt a görbe hat olyan pontból áll, aminek az értéke folyamatosan csökken. Ez egy statisztikailag releváns tendenciának felel meg. Mivel a hangsúly nem egy pillanatnyi, hanem egy fokozatosan végbemenő változással jár, ez a jelölés a hangsúlyra mint a relatív változás jól megragadható tendenciájára utal.

(9b)-ben egy olyan tendenciát is láthatunk, amit közvetlenül nem tudunk megfigyelni (9a) alapadataiból. A 60-as és a 90-es beosztások közötti szakaszban, az *eljöttek* utolsó szótagjában jelöl egy másodlagos hangsúlyt, amit a felpattanó zárhang intenzitásának tulajdoníthatunk.

(10) 'Valamennyien eljöttek  
(10a)

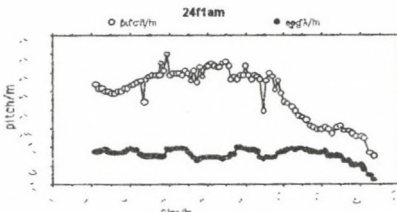


(10b)

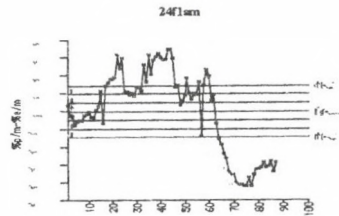


(10a) és (10b) összevetéséből az derül ki, hogy egyrészt itt is a PET alapján tudjuk egyértelműen kimutatni a másodlagos hangsúlyt, másrészt láthatjuk, hogy a kezdő hangsúly prominenciáját az intenzitás meredek relatív növekedése okozza, sokkal inkább, mint a (10a)-ból kézenfekvőnek tűnő hirtelen alapprofven-  
cia-csökkenés önmagában.

(11) Valamennyien 'eljöttek.  
(11a)



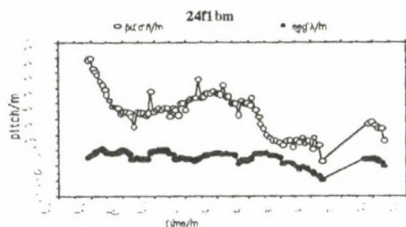
(11b)



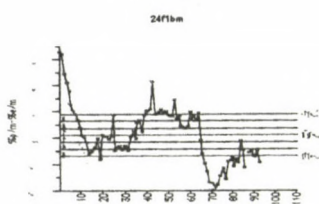
(11a) természetesen egy másfajta intonációs kontúrt mutat, ahol a hangsúlyos második szót (*eljöttek*) egy fokozatos alapprofven-  
cia-növekedés előz meg. Az ugyanakkor csak (11b) PET-ábrázolásából tűnik ki, hogy az alapprofven-  
cia abszolút értékét jóval meghaladó mértékben csökken az alapprofven-  
cia relatív prominenciája és jelentősen nő az intenzitása. Az is megfigyelhető, hogy a hangsúly jelölését egy kisebb mértékű relatív alapprofven-  
cia-prominencia emelkedés előzi meg.

(12) Valamennyien 'eljöttek.

(12a)



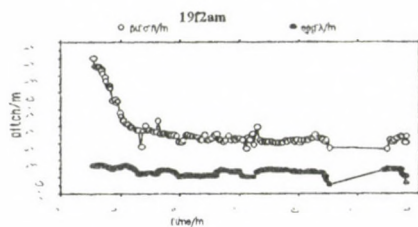
(12b)



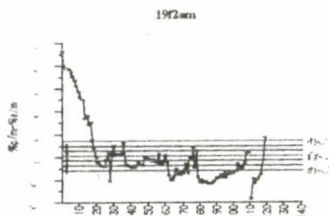
Míg (12a) egyértelműen jelez egy kezdeti dallam-hangsúlyt, valójában (12b) mutatja ki azt, hogy egy jelentős hangsúly van a második szón is. Míg az első hangsúlyt (12a) alapján az alapfrekvencia hirtelen esésével hozhatjuk összefüggésbe, a másodikat már csak a PET-reláció egyértelműsíti. Mindkét hangsúly-tendenciát egyébként a 3-as szám is jelzi.

(13) 'Valamennyien eljöttek.

(13a)



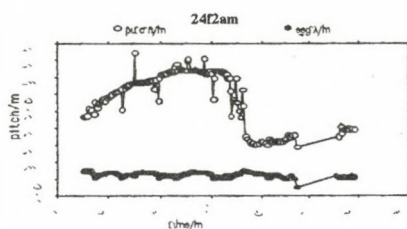
(13b)



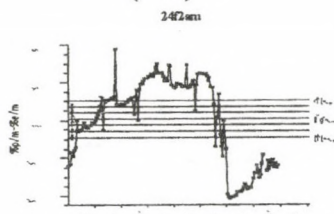
(13a) és (13b) összevetése alapján megerősíthetjük, hogy a mondat egy hangsúlyos szóval kezdődik (a szótag erős prominenciáját egyébként (13b) határozottan jelzi az intenzitás meredek relatív növekedésével). (13b) egyben lehetőséget nyújt a másodlagos hangsúly kimutatására (a felpattanó zárhangéra), amire egyébként nem lenne egyértelmű utalás (13a) alapadataiban.

(14) Valamennyien 'eljöttek.

(14a)



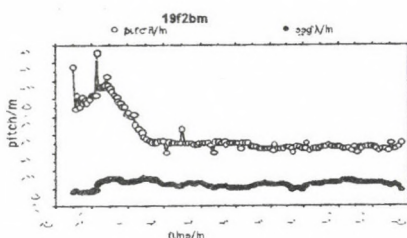
(14b)



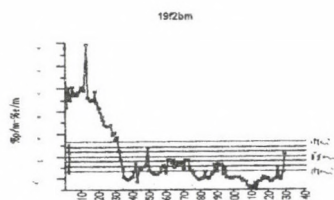
(14a) és (14b) nagyon hasonló alakzatot mutat. Azzal a különbséggel, hogy míg (14a) csupán a dallam-hangsúlyra tud utalni, addig (14b) kimutatja az intenzitás jelentős relatív növekedését is. És ugyanúgy, mint korábbi példáinkban, itt is a PET-adatokból tűnik ki, hogy a hangsúly kezdetét egy rövid relatív alapprofrekvencia-emelkedés jelöli.

(15) 'Valamennyien eljöttek.

(15a)



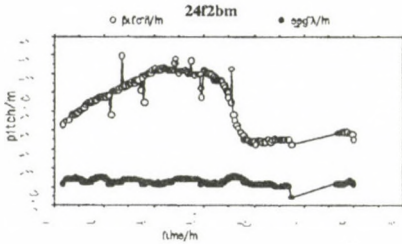
(15b)



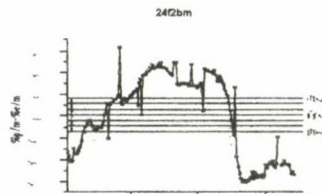
(15a) és (15b) alapján ismét azt állapíthatjuk meg, hogy míg a dallam-hangsúlyra (15a) alapprofrekvencia-kontúrjából is következtethetünk, a mellékhangsúlyokat csak a dinamikus reláció teszi érzékelhetővé - láthatóvá.

(16a) Valamennyien 'eljöttek.

(16a)



(16b)



(16a) és (16b) összehasonlítása újrag hangsúlyozza a dinamikus reláció-számítás azon előnyét, hogy ennek segítségével a kisebb, a dallam-hangsúlyból kevésbé kikövetkeztethető relatív mozgások is tetten érhetők. (16b)-ben jól látható, hogy a második szón (*eljöttek*) levő hangsúly egy relatív alapprofrekvencia-többszörrel indul, és csak ezt követi a relatív intenzitás-növekedés. A felfelé ívelő dallamgörbe kisebb relatív egyenetlenségei (amire (16b)-ből következtethetünk) ugyancsak a kisebb (szótag-) hangsúlyok helyeire utalnak. Végül, a hangsúly helyén igen jelentős relatív intenzitás-növekedés figyelhető meg, amit az erőteljes tendenciára utaló 3-asok is világosan utalnak.

### Kiértékelés

Példáink alapján láthatóvá vált, hogy az alapprofrekvencia és az intenzitás dinamikus relációjának a vizsgálata egyrészt alátámasztja azokat a hangsúly-értelmezéseket, amelyek a dallam-hangsúllyal jellemezhetők, másrészt hangsúly-ismereteinket jelentősen növelik azzal, hogy olyan hangsúlyos szótagokat is kimutat, amelyet sem az alapprofrekvencia, sem az intenzitás alapján egyedül nem tehetnénk meg. Különösen fontosnak tűnik a másodlagos hangsúlyok kimutatása.

A hangsúly mint a dinamikus reláció tendenciájának az értelmezése további lehetőségeket tár fel. Azzal, hogy külön jelen-

tőséget tulajdonítunk annak, hogy milyen tartós maga az adott egyirányú változás, egyben a változás és így a hangsúly jelöltségére is következtethetünk. Az itt bemutatott eljárással az időtartamot nem a hagyományos értelemben kapcsoltuk a vizsgált paraméterek közé, hiszen jelenlegi eszközeinkkel nem szóltunk arról, hogy egy magán- vagy mássalhangzó valós időtartama mennyire kimutatható módon befolyásolja a hangsúly azonosítását. A tendencia meghatározásával azonban mégis jelentős szerepre tesz szert maga az idő azzal, hogy a hangsúly-jelölés határozottságát a változás időbeli lefolyásához viszonyítjuk.

### Összefoglalás

A dolgozatban a hangsúlyt mint akusztikai paraméterek komplex relációját mutattuk be. Feltettük, hogy a hangsúly akusztikai jegyei paraméterek dinamikus relációjában mutathatók ki. Ezt tanúsították kísérleteink, amelyekben az alapfrekvencia és az intenzitás dinamikus, az idő függvényében végbemenő változását számítottuk ki. A PET-adatokra épül az alapfrekvenciának és az intenzitásnak az idő függvényében kiszámítható dinamikus relációja, ami – hasonlóan az alapfrekvencia és az intenzitás koordináta-rendszerben történő ábrázolásához – kézenfekvően ábrázolható és kiértékelhető. Ennek alapján fény derül arra, hogy a dallam-hangsúlyt csak mint a hangsúly egy lehetséges változatát tekinthetjük, míg általánosságban – és a mellékhangsúlyok esetében elsődlegesen – a dinamikus reláció alapján azonosíthatjuk a hangsúlyt. Az így értelmezett hangsúlyra az jellemző, hogy általában egy relatív alapfrekvencia-többslettel indul, ami ezt követően igen meredeken átvált egy relatív intenzitás-többsletbe. A változás egy statisztikailag releváns tendenciát követ. Az akusztikai paraméterek dinamikus relációján alapuló, itt bemutatott eljárás lehetővé teszi a hangsúly automatikus azonosítását, és az így megfoghatóvá váló hangsúly empirikus kiindulásul szolgálhat mind a hangsúly fogalmára épülő elméletek, mind a mérnöki alkalmazások számára.

### Irodalom

Beckman, M.E.–Edwards, J.: Lengthenings and shortenings and the nature of prosodic constituency. In: Kingston, J.–Beckman, M.E. (eds.) 1990, 152-178.

É. Kiss Katalin: A magyar mondat generatív leírása. Nyelvtudományi Értekezések 136. Budapest 1994.

É. Kiss, K.: Sentence structure and word order. In: Kiefer, F.–É. Kiss, K. (eds.): Syntax and Semantics 27. 1994, 1-90.

Fónagy Iván: A hangsúlyról. Akadémiai Kiadó. Budapest 1958.

Hunyadi, L.: Hungarian syntactic structure and metrical prosody. In: Proceedings of the First International Conference on Semantics and Pragmatics. Brighton 1995a.

Hunyadi, L.: Acoustic cues to sentential stress in Hungarian and their measurement. In: Hunyadi, L.–Gósy, M.–Olaszy, G. (eds.): Studies in Applied Linguistics Debrecen 1995b.

Kingston, J.–Beckman, M.E.: Papers in laboratory phonology II. Between the grammar and physics of speech. Cambridge Univ. Press. Cambridge 1990.

Ladd, D.R.: An introduction to intonational phonology. In: Kingston, J.–Beckman, M.E. (eds.) 1990, 321-334.

Ladd, D.R.–Verhoven, J.–Jacobs, K.: Influence of adjacent pitch accents on each other's perceived prominence: two contradictory effects. Journal of Phonetics 22. 1994, 889-899.

Varga László: A magyar intonáció – funkcionális szempontból. Nyelvtudományi Közlemények LXXXV, Budapest 1983/2, 313-339.

# A KÉRÉS, A FIGYELMEZTETÉS, A FELSZÓLÍTÁS ÉS A KÉRDÉS PROZÓDIÁJA A KIJELENTŐ MONDAT TÜKRÉBEN

Olaszy Gábor  
MTA Nyelvtudományi Intézete

A verbális információközlés folyamatában a nyelvi tartalom és a közlési szándék kifejezését változatos és jellemző prozódiai megoldások segítik. A dolgozatban a magyar prozódiai szerkezetek néhány jellemző fajtáját vizsgáljuk, különös tekintettel a szerkezetek egymás közötti kapcsolatrendszerére, úgy, hogy azokat a kijelentő mondat szerkezeteihez viszonyítva jellemezzük.

## Bevezetés

A magyar prozódiai szerkezetek komplex leírására és szintézissel ellenőrzött fonetikai elemzésére még nem került sor. Noha a magyar beszéd intonációjának vizsgálatára már századeleji munkákban is van példa (Csüri 1919), átfogóbb tanulmányok e témában csak a század második felében születtek. Az eddigi legrészletesebb vizsgálat eredményeit Fónagy–Magdics munkája (1967) foglalja össze. A magyar beszéd dallamának vizsgálatára több száz mondat hanglejtését jegyezték le és mutatják be ötvonalas kottairásos módszerrel. E bemutatási módot más kutatók bírálják, mivel a hallás után lejegyzett, kottaképpel jellemzett dallam nem ad elég információt a konkrét frekvenciaviszonyokról. A magyar beszéd intonációjának elemzésére az elmúlt tíz évben számos kutató végzett elemzéseket (Gósy 1989, Olaszy 1989, Varga 1993, Bolla 1995). Gósy az intonáció működését percepciós kísérletekkel vizsgálja, aminek lényege, hogy az adott frekvencia-meneteket milyen intonációnak ítéli a hallgató. Varga a magyar intonáció fonológiai rendszerét írja le, azonban konkrét frekvencia- adatokat nem közöl, hanem vonalas rajzokon mutatja be az

intonációs rendszer elemeit. Bolla pedig – noha több száz mondatot dolgoz fel, egyeseket szöveggörnyezettel – csak intonációs görbéket és egyéb regisztrátumokat mutat be, de azokat nem elemzi, verbálisan nem vizsgálja sajátosságait, és nem helyezi rendszerbe őket. Fontos megjegyezni e munka kapcsán, hogy (1) a kutatók eddig leginkább csak izoláltan ejtett egyedi mondatokat (leginkább kijelentéseket és kérdéseket) elemeztek, a szöveggörnyezeti elemeket nem vizsgálták, (2) a prozódiahoz tartozó intenzitás- és tempóváltozásokat sem elemezték. Így a prozódia működését adott szöveggörnyezetben az izolált mondatokra tett megállapítások alapján nem lehet meghatározni.

Ha a beszélt nyelvi kommunikáció prozódiai formáinak gyakoriságát vizsgáljuk, azt mondhatjuk, hogy a leggyakrabban előforduló kifejezési forma a kijelentés, majd utána a kérdés következik. Mindennapi életünkben azonban szituációk sorozatában használjuk a kérés, az utasítás, valamint a figyelmeztetés közlésformáit is. Ezek prozódijának vizsgálata napjainkban előtérbe került, mivel a gépi beszéddel működtetett ún. dialógus-rendszerekben a párbeszédes forma tartalmazhatja ezeket a kifejezési formákat, így ezek generálásához szabályokat kell kialakítani. A verbális kommunikáció komplex és folyamatos hangzásformája ugyanis megköveteli, hogy az egymás után ugyanazon beszélőtől elhangzó közlések egymáshoz hangzásilag kapcsolódjanak, prozódiai struktúrájuk egymásból következzen, vagyis a beszéd hangzása folyamatos és kifejező legyen és érzékeltesse a szituáció nyelvi tartalmát.

A jelen tanulmány arról a kísérletsorozatról számol be, amelynek során megpróbáltunk általános szabályokat megfogalmazni a fenti kifejezési formák prozódiai szerkezetének rendszerszerű leírására. A dolgozatban a hangmagasságot, a hangerőt és beszédtempót, illetve ezek változásait elemztük. E sorrend fontossági sorrend is, vagyis a legfontosabb ezek közül a hangmagasság változása.

Feltételezésünk a következő. A fent említett kifejezési formák prozódája, elsősorban a dallamszerkezetük az adott szituációs egységen belül szoros összefüggésben van egymással. Ezen összefüggéseket a nyelv általános intonációs rendszere határozza meg, és tudatunktól függetlenül használjuk őket az egyes kifejezési formák megszólaltatására.

A fenti munkával az volt a célunk, hogy kísérleti-fonetikai eszközökkel olyan szabályrendszert állítsunk össze, amely konkrét értékekkel írja le a fenti közlésformák legfontosabb akusztikai adatait, és azt, hogy ezek az adatok milyen kapcsolatban vannak egymással egy zárt szöveggörnyezeti rendszeren belül (pl.: dialógus). További célkitűzés, hogy a szabályrendszer olyan legyen, amely lehetőséget ad arra, hogy a fenti közlésformák prozódiaját elő tudjuk állítani a leggyakoribb közlésforma, a kijelentés struktúráiból. E célkitűzés teljesítése biztosíthatja, hogy szituációk, dialógusok prozódiajának előállításához az adatokat előre meg tudjuk tervezni és előállításuk egy beszédgeneráló rendszerben egyszerű eszközökkel megvalósítható legyen.

### **Anyag és módszer**

A kísérlethez olyan nyelvi anyagot állítottunk össze, amelyben előfordulnak a fenti közlésformák. A történet szövege a következő:

Kati, a titkárnő a számítógép előtt ült és írta a levelet. Véletlenül hibázott, ekkor a gép megszólalt:

– **Hibás parancsot kaptam! Nyomja meg az Enter gombot!**

Kati megnyomta a kért billentyűt. Ekkor a gép ismét megszólalt:

– **Mikor indul a repülő? – Kérem írja be a repülőgép indulásának időpontját!**

Kati beírta a kért időpontot, de idegességében megint elrontott valamit. A gép visszaszólt:

– **Nem értem! Mikor?**

Kati mérgesen rászólt a gépre: – **Ne türelmetlenkedj! Azonnal beírom az időpontot!**

Azt hitte, hogy a gép megérti, amit mond és várta a választ, de a gép néma maradt. Ekkor megint rákiabált: – **Szólj már meg te mindentudó!** A gépben azonban nem volt beszédfelismerő egység és így nem reagált Kati szavaira. Mire erre rájött, megint eltelt egy perc. A gép türelmesen várt, de aztán kis idő után megszólalt kérlelő hangon:

– **Tessék már beírni azt az időpontot!**

Katit meghatotta ez a kérlelő hangnem és beírta a kért számot. A gép udvariasan reagált:

– **Köszönöm. Folytatjuk?** és a képernyőn megjelent az igen és nem választási lehetőség.

– **Hát nem is tudom, hogy folytatjuk-e!** – mormolta Kati félhangosan magának.

– **Nekem még sokat kell tanulnom, hogy veled dolgozni tudjak.** – mormolta és rákattintott a nem utasításra.

A gép udvariasan közölte – **VÉGE.**

A történetet 3 férfi és 3 női bemondóval felolvastattuk, úgy, mintha egy novellát olvasnának fel. A felolvasást magnetofonra rögzítettük. Fontos, hogy kiejtési instrukciót, illetve hangzó példát nem adtunk a felolvasáshoz, tehát a közlésformák megformálása mindig a felolvasó által spontán megalkotott intonációt, hangerőt, beszédtempót tartalmazta. A felolvasók nem hallották egymás produkcióit a hangfelvételkor. Az elemzések során a vastag betűkkel jelzett mondatok prozódiai szerkezetét elemeztük és hasonlítottuk össze. A teljes vizsgálati anyag  $14 \times 6 = 84$  mondat.

Az analízishez Kay 4300B CSL (Computerised Speech Laboratory) beszédanalizátor szoftvert használtuk. A közlések hangmagasságát, intenzitásstruktúráját és időtartamviszonyait egymás alatti három regisztrátumon rögzítettük és elemeztük. E három paraméter változásának együttes vizsgálatával olyan háromváltozós prozódiai függvényt  $Pr(f)$  alkottunk, amely a kijelentő mondat hangmagasság-, intenzitás- és beszédtempó változásainak mint viszonyítási alapadatoknak, a segítségével kifejezi az egyes további közlésformák jellemző prozódiai struktúráját. Tehát például a kérést kifejező mondatot a kijelentő mondatból a következő összefüggés alkalmazásával származtatjuk.

$$\text{Pr (kérés)} = \text{Pr}(\text{kijelentés}) \times (\text{P,I,T})$$

ahol P a hangmagasság, I az intenzitás és T a beszédtempó változását tartalmazó függvény. Fontos megjegyezni, hogy a P és T esetében relatív %-ban, az I esetében dB-ben adjuk meg a változás mértékét. A P esetében a %-os kifejezésre azért van szükség, hogy a hangfekvés transzponálásának a lehetőségét is biztosítsuk. A P esetében külön figyelmet szenteltünk a közlés általános tendenciagörbéjének és a szóhangsúly megvalósulásának összefüggésére is. A beszédtempó változását hang/s adatokból számítottuk, azonban tudnunk kell, hogy a tempó széles határok között változhat a beszélő és a szituáció függvényében.

A fenti összefüggés segítségével kiszámíthatók az egyes kifejezési formák alapvető intonációs, intenzitás és beszédsebesség jellemzői a kijelentő mondat jellemzőiből. Így a közlésformák prozódiaja illeszthető lesz egymáshoz. Az analízis során figyelemmel voltunk arra, hogy az egyes közlések digitalizálásakor a hangerőt ne változtassuk. Ezzel biztosítottuk, hogy az egyes bemondók által produkált hangerőadatok a történet összes mondatában – ugyanazon beszélőnél – egymással összemérhetők, ami az objektív mérés lehetőségét biztosította.

### **A kijelentés prozódiai szerkezete**

A kijelentő mondat mint viszonyítási alap szerepelt a vizsgálatban. Az alaphang- és intenzitásgörbék indulási értékét 100%-nak tekintjük, és ehhez viszonyítva fejezzük ki a többi közlésforma ugyanezen görbéinek jellemző pontjait. Ez azt jelenti, hogy ha például a kijelentő mondat indulási alaphangfrekvenciája 120 Hz, akkor ez felel meg 100%-nak. Ha egy másik közlésben az indulási alaphangfrekvencia 80%-nak van megadva, akkor az  $120 \times 0,8 = 96$  Hz lesz.

A kijelentés általános mondatintonációs tendenciagörbéje ebben a vizsgálatban is igazolta számos kutató korábbi

megállapítását, mely szerint ez "eső" forma. Az esés mértéke méréseink szerint 25-35%. Gósy (1989) percepciósi méréseiben hasonló értékre kapta meg a kijelentő hangzás érzetét. A három férfi ejtésében az eső tendenciagörbe induló alaphangfrekvenciája átlagosan 125 Hz, a mondatvégi érték pedig 78 Hz volt. Női ejtésben ugyanezek az értékek a következők: 260 Hz és 175 Hz.

A magyarra jellemző első szótagi hangsúlyozási szabály megvalósul ebben a közlési formában, ami azt jelenti, hogy tendenciagörbéből a szavak első szótagján kiemelkedések mérhetők.

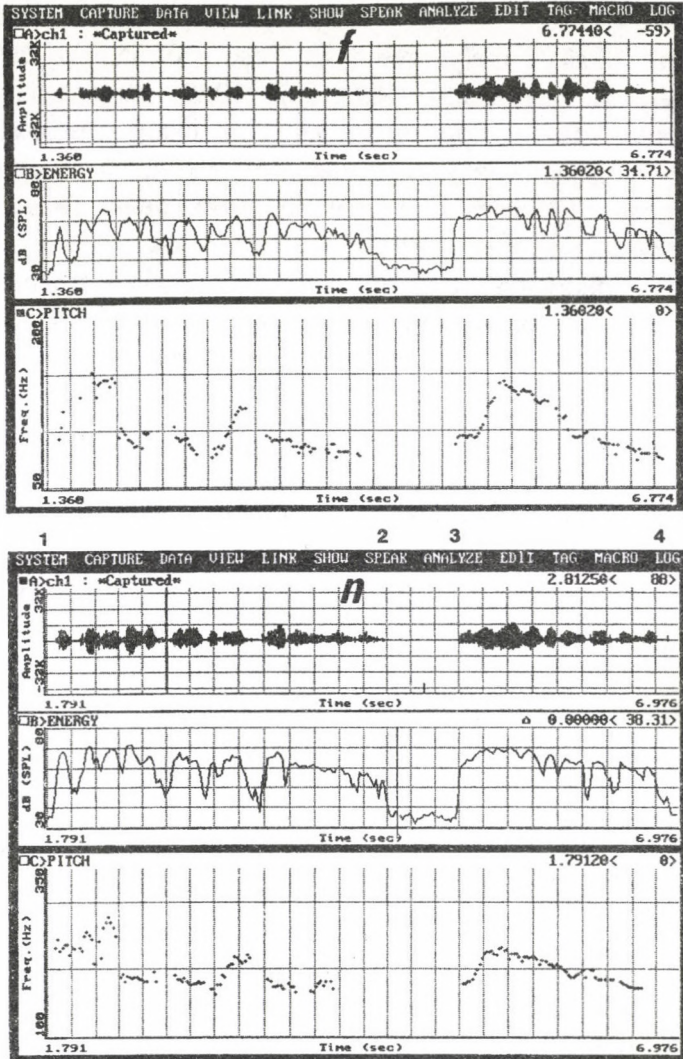
A kijelentő mondatok intenzitás tendenciagörbéje hasonló alakú, mint az intonációs görbe. Az esés mértéke kb. 35 dB. Mind női, mind pedig férfi ejtésre hasonló értéktartományt mértünk.

A beszédsebesség átlagértékei: 13-15 hang/s. Ezt a beszédsebességet vettük normál tempónak. Ehhez viszonyítottuk a többi közlésforma beszédsebességét.

### **A kérés prozódiai szerkezete**

A kérés több fajtája közül azt a változatot vizsgáltuk, amelyben egy kérlelő hangnem realizálódik kis türelmetlenséggel párosítva (1. ábra).

Az intonációs tendenciagörbe indítási alaphangfrekvenciája itt kb. 30 %-kal mélyebb, mint a kijelentésnél, vagyis ez az indulási érték megközelíti a kijelentés befejező részében kialakuló hangmagasságot. Ez az érték mérhető az első szótag egész tartamában. Az alaphangfrekvencia a második szótagban kezd emelkedni és a szótag végére 30-40 %-kal magasabb lesz, mint az induló érték. Ezután lassan, egyenletesen ereszkedik a közlés végéig néhány százalékkal a kiindulási érték alá. Ez a "csúcsos" intonációs görbe jellemző a kérlelés megformálására. Az eltérés a kijelentő mondat intonációs görbétől tehát jelentős.



1. ábra

A kérérlő hangnemben ejtett és az azt megelőző kijelentő mondat  
 P, I, T struktúrája férfi és női ejtésben.

Az ejtett mondatok: 1-2 *Ekkor újból megszólalt a gép kérérlő hangon.* 3-4 *Nyomja már meg az ENTER gombot!*

Ebben a közlési formában nem mérhetőek szóhangsúlyra jellemző alaphangmagasság kiemelkedések és csökkenések a szavakon. Ez azt jelenti, hogy a kérlelés megformálásánál a hangsúlyozási szabálynak nincs számottevő szerepe.

A kérlelés megformálásánál az indítási intenzitás értéke kb. 6 dB-lel alacsonyabb, mint a kijelentésé. A további tendencia az intonációs görbében mérthez hasonlít, vagyis a második szótagtól kezdve fokozatosan emelkedik, annak végére eléri a maximumot, majd ezen a maximumértéken marad a 3., 4. szótagban. Ezután a hangerő fokozatosan csökken a kifejezés végéig, ahol kb. -30 dB-t ér el. Látható, hogy a kérlelésre egy "púpos" intenzitásgörbe a jellemző, amely főleg indításában tér el a kijelentésétől.

A beszédsebesség igen változó lehet ebben a közlési formában. Beszélt nyelvi megfigyeléseink szerint a kérlelési forma inkább a lassabb tempót valósítja meg. Az általunk vizsgált mondatokban a beszédtempóban minimális lassulást mértünk a kérlelést megelőző mondat beszédtempójához viszonyítva. Az eltérés mértéke 1 hang/s volt.

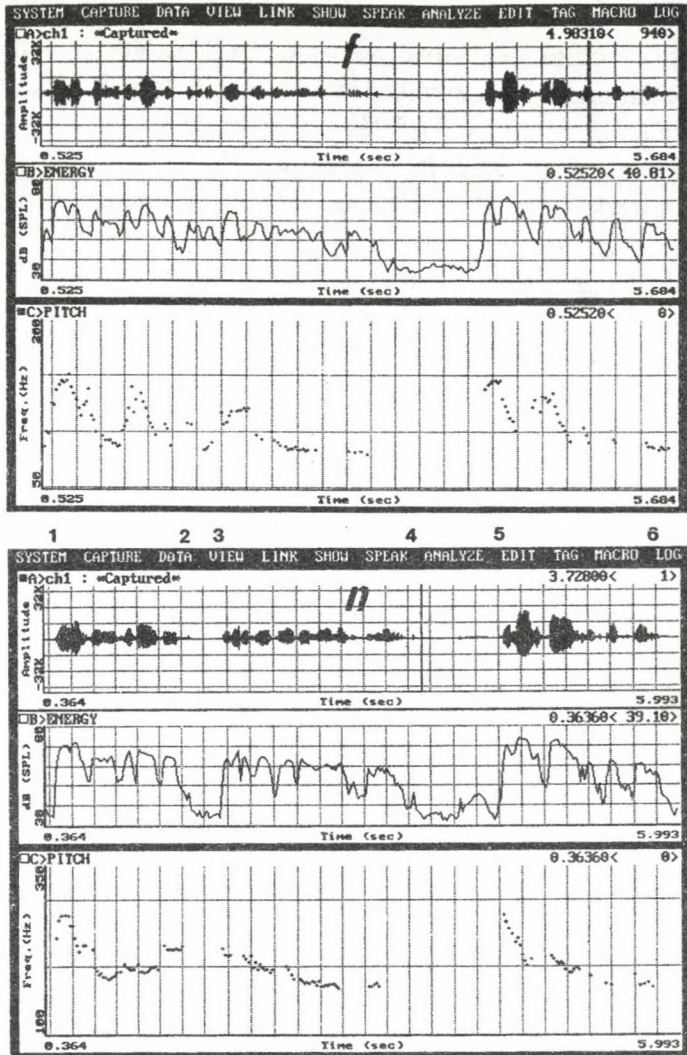
### **A figyelmeztetés prozódiai szerkezete**

A figyelmeztetés sokféle megvalósulási formájából a tévedésre való felhívást vizsgáltuk (2. ábra).

Az intonációs tendenciagörbe indítási alaphangmagassága hasonló lineáris csökkenést mutat, mint a kijelentésé, csak nem süllyed olyan mélyre, vagyis a kb. 70%-os végértékkel szemben itt csak kb. 80%-ra csökken az alaphangmagasság. A kijelentő mondatához képesti eltérés tehát nem a struktúrában van, hanem a végértékben.

Az első szótagi hangsúlyozási szabály méréseink szerint megvalósul a szavakon ebben a közlési formában.

Az intenzitásgörbében is hasonló tendencia figyelhető meg, mint amit az alaphangmagasság görbében láttunk, vagyis a görbe menete hasonló a kijelentéséhez. Két ponton van különbség. A kezdő intenzitás szint kb. +5, +10 dB-lel magasabb (az ejtéstől és a kívánt



2. ábra

A figyelmeztetés és az azt megelőző kijelentő mondat P, I, T struktúrája férfi és női ejtésben  
 Az ejtett mondatok: 0-1 Véletlenül hibázott 2-1 és a gép azonnal megszólalt. 4-5 Hibás parancsot kaptam!

intenzitásszint-kifejezéstől függően), továbbá a görbe lefutásának meredeksége minimális a kifejezés első felében, és csak a második felében lesz a csökkenés meredekebb. Ez a meredekség sem éri el azonban a kijelentő mondatét, vagyis a befejezési érték mintegy 5 dB-lel magasabb, mint a kijelentő mondatnál. A jelentős változás a kijelentő mondatához képest a kezdeti hangerőszint megemelkedése.

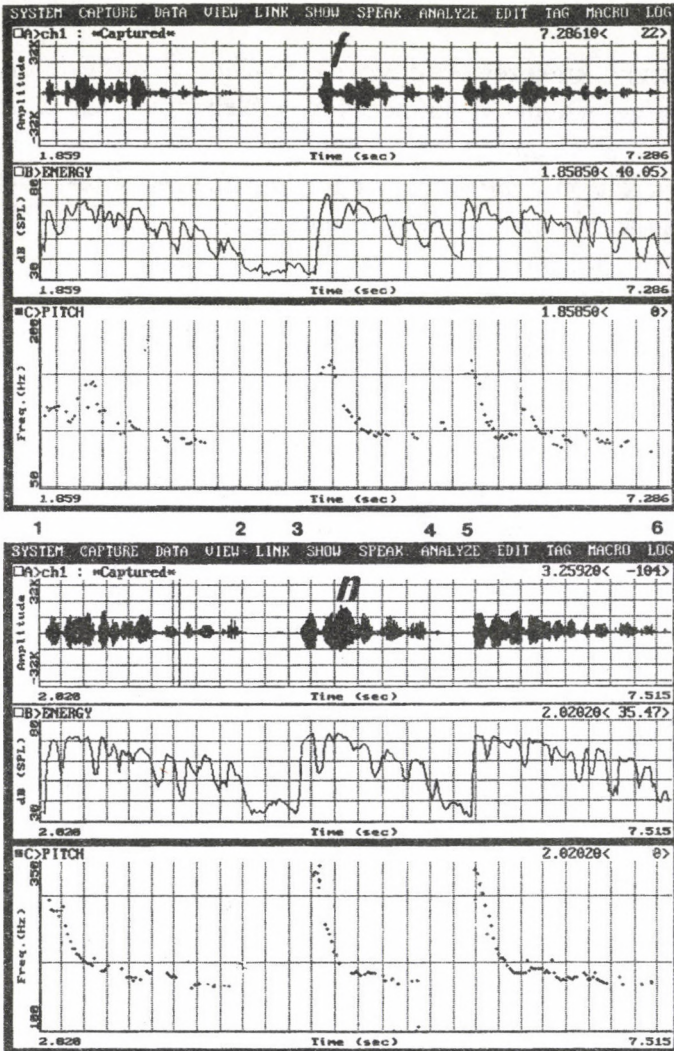
A figyelmeztetés eme formájánál ugyanolyan beszédtempót állapítottunk meg, mint amilyen a kijelentésé.

### **A felszólítás prozódiai szerkezete**

A kísérletben vizsgált felszólításokban (3. ábra) előfordult indulatosan és kevésbé indulatosan megformált közlésforma. Az indulati fokozatok inkább az intenzitásgörbében jelentkeznek, az intonáció indulattól függetlenül ugyanolyan struktúrát mutat.

A felszólításban az alaphang indítása a kijelentéshez viszonyítva magasabb, kb. 140%-os értékről kezdődik. Már az első szótagban meredek esés figyelhető meg, és ez az esés adja a felszólítási kifejezési forma lényegét. E meredek szakasz az első szótag végére kb. 90%-ra csökken, majd a második és harmadik szótagban további kb. 10%-os esés tapasztalható, majd a görbe fokozatosan a kijelentésnél mért végértékre csökken a kifejezés végére. A felszólításban tehát három különböző meredekséggel csökkenő dallammenet mutatható ki. A legmeredekebb a legrövidebb, a legkevesbé meredek a leghosszabb (több szavas mondatnál). Látható tehát, hogy mind értékben, mind pedig struktúrában ismét lényeges eltérés van a kijelentő mondat intonációs tendenciagörbéjéhez képest.

Az első szótagi hangsúlyozási szabály megvalósulhat ebben a közlési formában a beszélő szándékától függően. Az ábrán bemutatott esetben a férfi ejtésnél (5-6 jelzésű mondat) nemcsak az első szó első szótagján, de a *beírom* szó első szótagján is



3. ábra

A felszólításként ejtett és az azt megelőző kijelentő mondatok

P, I, T struktúrája férfi és női ejtésben

Az ejtett mondatok: 1-2 *Kati mérgesen rászólt a gépre.* 3-4 *Ne türelmetlenkedj!* 5-6 *Azonnal beírom az időpontot.*

láthatunk alaphangfrekvencia kiemelkedést, míg a női ejtésnél ugyanezen a szón ilyen változás nem látható, itt csak az első szó első szótagját ejtette igen nyomatékosan a beszélő.

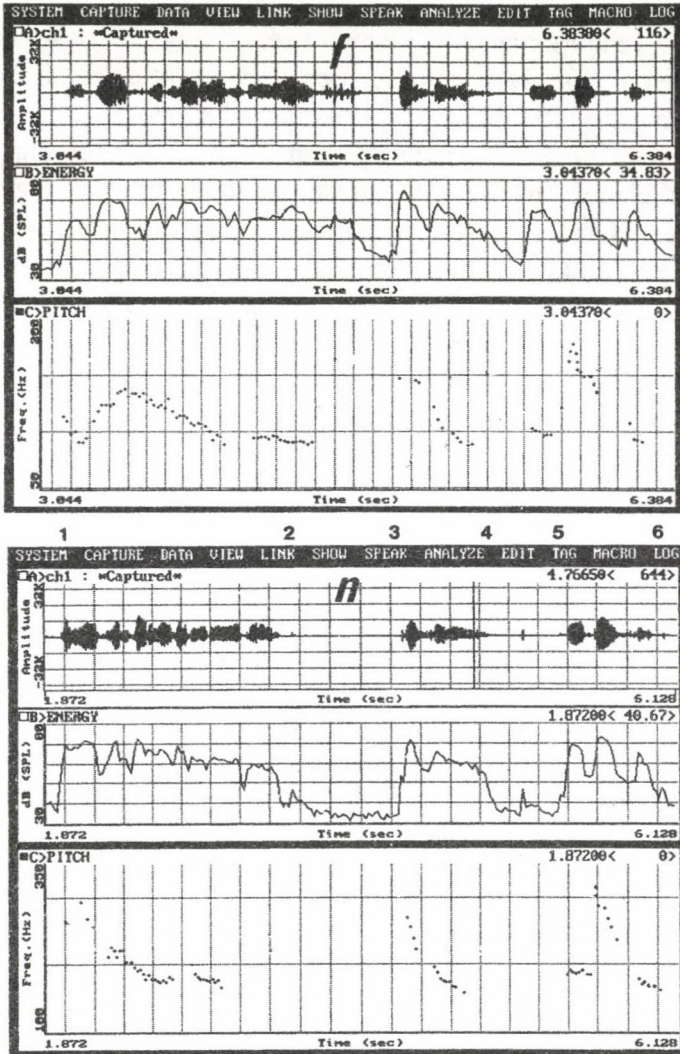
Az intenzitásgörbe indulási pontja kb. 5 dB-lel magasabb, mint a kijelentés indítási intenzitása, de ez az érték – érzelemtől függően – akár 15 dB-lel is magasabb lehet. A jellemző intenzitásgörbe – hasonlóan a figyelmeztetéséhez – csekély meredekséggel csökken kb. 15 dB-t a közlés utolsó negyedéig, majd ebben a negyedben meredeken csökken további 20 dB-t. A figyelmeztetésben a hangerő átfogása tehát 50 dB is lehet. A lényeges különbség a kijelentő mondat intenzitásgörbéjéhez képest az indítás nagyobb hangerejében, majd az átlagos hangerő lassú csökkenésében jelentkezik.

A felszólításban a beszédtempóra ugyanazt az értéket kaptuk, mint amit a kijelentő mondatnál mértünk.

### **A kérdés prozódiai szerkezete**

Három kérdésfajtára, az eldöntendő, a kiegészítendő kérdésre, valamint a visszakérdezésre végeztük el a vizsgálatot. A visszakérdezés tulajdonképpen a kiegészítendő kérdés kategóriájába tartozik és kezelése csak akkor okoz problémát, ha a kérdőszavak visszakérdezett formáját kell megvalósítani. Ekkor ugyanis nem a megszokott eső görbét, hanem az emelkedőt kell alkalmaznunk. E szavakra tehát kétféle intonációs görbe is ráültethető, attól függően, hogy mit akarunk kifejezni.

Az eldöntendő kérdés (és a visszakérdezés) (4. ábra) intonációja három alkategóriában határozható meg egy, kettő, illetve többszótagú változat. Itt most csak a három szótagnál hosszabb kérdő mondatét tárgyaljuk. Az alaphangmagasság görbe indulási pontja alacsonyabb, mint a kijelentő mondaté, annak kb. a 80%-a. Utána enyhe (max. 10%-os) fokozatos emelkedés tapasztalható az utolsó előtti szótag elejéig, ahol az alaphangmagasság hirtelen emelkedni kezd kb. 120%-ra, majd e



4. ábra

Az eldöntendő kérdés és az azt megelőző két kijelentő mondat P,  
I, T struktúrája férfi és női ejtésben  
Az ejtett mondatok: 1-2 *A gép udvariasan reagált.* 3-4 *Köszönöm.*  
5-6 *Folytatjuk?*

szótag végétől ismét meredeken csökken, és a mondat végén a kezdeti érték alá esik, kb. 70%-ot ér el. A kiegészítendő kérdés indulási alaphangmagassága magasabban van, mint a kijelentő mondaté. A kezdő érték kb. 120% és ezt tartja is az első szótagban. A második szótag elejére ugrásszerűen, hirtelen kb. 80%-ra csökken, majd innen kezdve fokozatosan csökken a mondat végéig 70%-ra.

Az első szótagi hangsúlyozási szabály megvalósítására utaló alaphangmagasság-kiemelkedéseket nem találtunk a szavakban sem a kiegészítendő, sem az eldöntendő kérdésnél. Ez azt jelenti, hogy ezeknél a kérdésfajtáknál a hangsúlyozási szabálynak nincs számottevő szerepe a prozódia kialakításában.

Mindkét esetben lényeges eltérés mutatkozik a kérdés és a kijelentés intonációs görbéi között.

A kérdések intenzitásképe is különbözik. A kiegészítendő kérdésnél a kijelentéséhez hasonló, egyenletesen eső intenzitásgörbe látható. Az eldöntendő kérdésnél más a görbe alakulása, az intenzitás indulási értéke mintegy -6 dB-lel alacsonyabb, mint a kijelentés indulási értéke, és alig változik egészen az utolsó előtti szótagig. Itt ezen a szótagon esetleg egy 3-6 dB-es kiemelkedés tapasztalható, majd az intenzitás rohamosan csökken a mondat végéig.

A beszédsebesség tekintetében nem mértünk különbséget a két kérdésfajta között, a normál beszédtempóra mért adatokat kaptuk.

### **Összefoglalás**

A fenti kísérletsorozattal megállapítottuk a vizsgált kifejezési formák jellemző alaphangmagasság-, intenzitás- és beszédsebesség-változásait a kijelentő mondat ugyanezen adataira vetítve. Az eredményeket az első táblázatban szemléletesen foglaljuk össze.

1. táblázat: A különböző kifejezési formák prozódiai jellemzői a kijelentő mondatéhoz viszonyítva

Közlés- forma	Jellemző intonáció	Jellemző hangerő	Jellemző tempó	Első szótagi hangsúly
Kijelentés			14 hang/s	van
Kérlelés			13 hang/s	nincs
Figyel- meztetés			14 hang/s	van
Felszólítás			14 hang/s	van
Kérdés (eldöntendő) (visszakérd.)			14 hang/s	nincs
Kérdés (kiegész.)			14 hang/s	nincs

Ezen adatok segítségével a kijelentő mondat – mint kiindulási egység – adataiból előállíthatjuk a fenti közlésformákat úgy, hogy azok szöveggörnyezetben is helyes hangzást eredményezzenek.

További vizsgálatokat tervezünk arra vonatkozóan, hogy a fenti P, I, T kategóriák milyen határok között változtathatók meg anélkül, hogy az elhangzó szituáció, dialógus természetes hangzása, a mondatok egymásból való hangzásbeli következése csorbát szenvedne. További vizsgálatokat tervezünk arra vonatkozóan is, hogy a fenti szabályrendszert szintetizált beszéden alkalmazzuk és percepciók tesztekkel állapítsuk meg a szabályok működésének eredményét és korlátait.

#### Irodalom

Bolla Kálmán: Szupraszegmentális elemzések. Egyetemi Fonetikai Füzetek 7. Budapest 1992.

Csúri Bálint: Hanglejtés. Nyr 78. Budapest 1919, 71-78.

Fónagy Iván–Magdics Klára: A magyar beszéd dallama. Akadémiai Kiadó. Budapest 1969.

Gósy Mária: Beszédészlelés. MTA Nyelvtudományi Intézete. Budapest 1989.

Olaszy Gábor: Elektronikus beszédelőállítás. Műszaki Könyvkiadó. Budapest 1989.

Varga László: A magyar beszéddallamok fonológiai, szemantikai és szintaktikai vonatkozásai. Nyelvtudományi Értekezések 135. Budapest 1993.

# HANGTÖRTÉNETI VÁLTOZÁSOK PERCEPCIÓS VONATKOZÁSAI (Miért nincsenek /ji/ és /vu/ szókezdetek finnugor nyelvekben?)

**Bakró-Nagy Marianne**  
MTA Nyelvtudományi Intézete

Az előadás kettős célkitűzésű. Egyrészt megvizsgálja a /ji/ és /vu/ szekvenciára vonatkozó fonotaktikai szabályokat két, egymástól történetileg jelentősen elkülönült finnugor nyelvben, a finnben és a magyarban, s kettejük közös előzményében, a finnugor alapnyelvben. Másrészt érveket keres arra nézvést, hogy e szekvenciákra vonatkozó szabályoktól leírt jelenségek miért és milyen mértékben köthetők artikulációs vagy percepciós indítékokhoz. Főbb megállapításai a következők: 1. a vizsgált szekvenciák szinte teljes hiánya tulajdonítható a finnugor alapnyelvi sajátosság továbbélésének is, 2. az alapnyelvtől a mai nyelvállapotig vezető változások a szabály megerősödésének irányában hatnak, 3. a szabály maga s a változások iránya percepciós indítékúak. Az elemzés végeredményben a percepciós szempontjainak nyelvtörténetbeli fontosságát emeli ki.

## Célkitűzés

Előadásomban egy – jelen esetben csupán a finn és a magyar nyelv köznyelvi változatának szempontjából jellemzendő – fonotaktikai szabályról (1), a szabállyal leírt jelenség kialakulásának történeti előzményeiről (2), s a történeti előzmény lehetséges fonetikai magyarázatáról (3) szándékozom számot adni. Mondandóm tehát alapvetően nyelvtörténeti aspektusú, háttérét pedig az az elvi és módszertani megalapozottságú állítás adja, mely szerint a percepciós és akusztikai vonatkozások figyelembevétele nélkül a hangváltozások egy sor meghatározott esetben egyszerűen magyarázat nélkül maradnak. Ez az állítás azért hangsúlyozandó, mert – nagy általánosságban – a hangtörténeti változások végső okaként inkább jönnek számításba artikulációs

tényezők, semmint percepciósak vagy akusztikaiak. Noha megfigyelés természetesen következhetik abból a tényből, hogy a beszédhangok sorozatának a képzésekor is az artikulációs szempont játssza az elsődleges szerepet, a nyelvi változások létrejöttében (sem) elhanyagolható szempont, hogy mi az, amit a hallgató hall (ti. szemben azzal, amit a beszélő mond).

### 1. A szabály és hatóköre a finnben és a magyarban

Általános megfigyelések szerint (Ohala–Kawasaki 1984, 113–127, Maddieson–Precoda 1992, 55–56) a nyelvek egy részére érvényes az alábbi megszorítás:

$$\sigma \rightarrow /j_i C_{\sigma} / \\ /w_u C_{\sigma} / \quad /$$

azaz vagy egyáltalán nincsenek avagy csak igen ritkák a palatális, illetőleg bilabiális félmagánhangzóból és az elöl képzett illabiális felső, illetőleg hátulképzett labiális felső nyelvvállású elemekből álló szótagok, helyezkedjenek el bárhol is a morfémában vagy annak a határán. Érdemes hangsúlyozni, hogy általános megfigyelés szerint indifferens, vajon a mondott szegmentumokból építkező szótagok pusztán a /j/, /w/ szótagkezdetből s az /i/, /u/ szótagmagból állnak-e, avagy a szótagot kóda zárja-e, mint ahogyan közömbös az is, hogy a rájuk következő szótag milyen szerkezetű. Ezt azért kell külön is kiemelni, mert a szabálynak a finnre és a magyarra való alkalmazásakor egyebek mellett tisztázandó a fentiek érvényessége.

A szabályban foglaltak hatóköre legalábbis a következők szerint mérlegelendő: a) milyen (morféma, szó) határok között érvényesül a két nyelvben, b) milyen szótagszerkezetbe épülve érvényesül a két nyelvben, c) milyen elemekre terjed ki a szabály. (A forrásokat a két nyelv sztenderd értelmező és etimológiai szótárain kívül Karlsson 1982, Papp 1969 és Szende 1973 jelenti.)

a) Mindkét nyelvben csak morfémahatárok között érvényes, a morfémaak illeszkedésének a határán már nem (l. pl. a ragozott finn *kantajia*, *suvun*, magyar *altaji*, *lovunk* alakokat). A finnben a morfémahatárokon belül az első szótag határán túl ugyancsak nem érvényes (pl. *tavu*), s ugyanez vonatkozik a magyarra is. A magyar /ji/ szekvenciával már más a helyzet: a /j/-nek hiátustöltőként való jelentkezése következtében olykor találkozunk vele további szótagokban is (hiszen ebbéli szerepében nem kötődik kizárólag morfémahatárhoz, pl. *la[ji]kus*, l. még Nádasy-Siptár 1994, 174). E relatíve ritka előfordulástól eltekintve azonban jellemző, hogy e két szekvencia kezdő szótagként nem jelentkezik e nyelvekben, illetőleg csak elhanyagolható számú kivétellel, amelyek pedig a következők: a finnben a /ji/ kezdetűek száma nem éri el a tizet (pl. *jiddis*, *jigi*, *jiikata*, *jiina* stb.), s ugyanez vonatkozik a /vu/ kezdetűekre is (pl. *vulfinetti*, *vulgaari*, *vulkaani* stb.); mindkét csoportra jellemző, hogy nemzetközi vagy jövevényszók, illetőleg fonostilisztikailag jelöltek. A magyarban a kivételek még ritkábbak: /ji/ kezdetű a köznyelvben csupán a *jiddis*, /vu/ kezdetű meg csak a *vulgáris*, *vulkán*, *vurstli*, mind a négy forma átvétel. Mindezek figyelembe vételével tehát a finnben és a magyarban a szabály a következőképp specifikálható:

Finn–magyar:

$$\sigma \rightarrow /jiC\text{ } / \# \underline{\quad}$$

Finn:

$$\sigma \rightarrow /vuC\text{ } / \# \underline{\quad}$$

Magyar:

$$\sigma \rightarrow /vuC\text{ } / \# \underline{\quad}$$

b) Míg a magyarban nincs megkötés arra nézvést, hogy a szekvenciák milyen szerkezetű (tehát pl. nyílt vagy zárt) szótagnak a szerkezeti elemei, addig a finnben ez csak részben igaz. A finn esetében a /j/ úgy viselkedik, mint a magyarban, a /v/ esetében a szabály viszont csak akkor működik, ha a szótag csupán egy /v/ kezdetből és /u/ magból áll (azaz a kóda kitöltetlen) és a rákövetkező szótag mássalhangzóval kezdődik, illetőleg ha az /u/ magot kóda is követi, s a rákövetkező szótag ugyancsak mássalhangzóval kezdődik (azaz mediális mássalhangzó-kapcsolattal van dolgunk). Amikor azonban az /u/ magot (kóda-zárás nélkül) a következő szótag magja követi (tehát mediális diftongussal van dolgunk), a /vu/ morfémakezdet nincs kizárva. Lényeges megjegyezni, hogy e diftongus második eleme csak /o/ lehet (pl. *vuori*). E szerkezet létének magyarázata majd a nyelvtörténetben keresendő.

c) A szabály mindkét nyelvben érvényes a /j/-re, azzal a különbséggel, hogy e hang a finnben félhangzó, a magyarban viszont approximáns. A bilabiális félhangzó mindkét nyelvből hiányzik, ellenben mindkettőben megvan a /v/ – a finnben félhangzóként, a magyarban teljes mássalhangzóként, azaz réshangként –, s a szabály rá érvényesül. Magyarázata nyelvtörténeti szempontból – mint látni fogjuk – több mint egyszerű.

Az elmondottak eredményeként megtudhattuk tehát, hogy – a megfelelő megszorításokkal – a finn és a magyar is beilleszthető a mondott szekvenciákat kerülő nyelvek csoportjába. Kérdés azonban, hogy miként értékeljük az általános szabálytól eltérő megszorításokat. Ezek tekintetében ugyanis például a következő kérdések tehetők fel: Alapvetően miért az első szótagot érinti a szabály? Miért a /v/-re vonatkozik a szabály? A finnben miért nem érvényes a szabály a /vuo/ morfémakezdetekre? Az e kérdésekre adható válaszok könnyen megfogalmazhatók egy természetszerűen felmerülő lehetőség végig gondolásának a mentén: mivel rokon

nyelvekről van szó, kérdéses, hogy egy elterjedtnek mondható fonetikai jelenség egymástól független realizálódásáról van-e szó avagy olyasvalamiről, amit akár örökségként is hozhattak magukkal egy közös előzményből.

## 2. A szabálytól leírt jelenség lehetséges történeti előzményei

Elsőként is tisztázandó, hogy a szabály tekintetében milyen volt a rekonstruálható finnugor alapnyelvi (PFU) előzmény. (Ezúttal csupán a finn és a magyar múltjának szempontjából legrégebbnek mondható nyelvi szinteket mérem fel, s nem térek ki arra, hogy a PFU, a proto-finn-permi (PFP) és proto-ugor (PUG) szintek után következő korszakok milyen szerepet játszhattak a jelenlegi állapot kialakulásában.)

A PFU-ban (továbbá a finn szempontjából releváns PFP-ben, s a magyar szempontjából fontos PUG-ban) a mai /j/ és /v/ előzményeként rekonstruálható \*j, illetőleg \*w félhangzó volt. Nyelvtörténeti és egyéb (rekonstrukciós) okokból az alapnyelvben csupán a tömorfémák első szótagjai vehetők figyelembe, mert

- a PFU tömorfémák - elenyészőn kevés kivételtől eltekintve - kétszótagúak,
- második szótagjuk nyílt,
- az uralkodó magánhangzó-harmónia értelmében a második szótagi, tövégi vokális hangszínét az első szótag vokálisa határozza meg. Ennek következtében az csak \*a, \*ā vagy \*e magánhangzóként realizálódhat, azaz a szempontunkból lényeges \*i és \*u itt nem fordulhat elő. Az összevetési lehetőséget e tény meglehetősen leszűkíti, mert csak az első szótagra korlátozza. Mivel azonban azt tapasztaltuk, hogy a magyar hiátustöltő /j/-től eltekintve a szabály a két nyelvben csak az első, morfémakezdő szótagokat érinti, az előzménnyel való összevetés e ponton szolid alapokon nyugszik. Ugyanakkor van még egy további hozadéka is: mivel a PFU az első szótagnál hátrább nem engedí meg a \*ji/\*wu

szekvenciát, megállapítható, hogy azoknak a mai nyelvi állapotokban is megfigyelhető hiánya lehet akár az alapnyelvből öröklött tulajdonság is.

A részletező nyelvtörténeti elemzéseket elhagyva, summázatként a következők állapíthatók meg: a. a *\*ji/\*wu* kezdetű morfémák az alapnyelvi előzményekben is csak igen korlátozott számban (a PUG-ban egyáltalán nem) fordulhatnak elő, s kérdéses, hogy a vokálisok előtt álló félhangzók valóban szerepeltek-e bennük, avagy pusztán vokálissal kezdődtek-e a morfémák (pl. magyar *év* ~ finn *ikä* < PFU *\*jikä* (?*\*ikä*) 'Alter, Jahr'), b. a biztosra mondható *\*ji/\*wu* kezdetű rekonstrukciók sorsa viszont igen csak egyformán alakult a finnben és a magyarban: PFU *\*ji* > finn-magyar *\*i*, *\*wu* > finn-magyar *\*u*, azaz a félhangzók *\*i* és *\*u* előtti helyzetben eltűntek (l. a fenti példát s még: magyar *új* ~ finn *uusi* < PFU *\*wud'e* 'neu').

A szabály tehát, ha nem is teljes kizárólagossággal (amennyiben csak a kezdő szótagokra érvényes, s néhány kivétellel) működni látszik az előzményekben is, s még csak megerősítést nyer a hangváltozások révén, hiszen azok éppen hogy a szabály státuszának a megerősödése irányában hatnak. Mégpedig olyannyira, hogy végeredményében a szabály kiterjed a változás eredményeként létrejövő (ti. nem *\*u* előtti helyzetű) finn és magyar /v/-re (< *\*w*) is.

A szabály működésének időbeli határát a két nyelvből vett kivételek (jövényszavak) megjelenése jelzi, s a finnből regisztrált /*vo*/ szekvencia magyarozatára is most kerülhet sor: a finnben az /*uo*/ diftongus *\*o* (< *\*o*) előzményre megy vissza, azaz a szabály hatása a diftongálódás szakaszában már gyöngülhetett.

Összegezve tehát kézenfekvőnek látszik, hogy e nyelvek történeti előzményeiket tükrözik a vizsgált jelenséget illetően, ami természetesen a későbbiek folyamán is párosulhatott egy, az előzményektől akár függetlenül is érvényesülő, percepciósi indíttatású fonetikai hatással. A hátralevőkben ezt a – már a

történeti előzményekben működni látszó – percepciós szempontot veszem szemügyre.

### **3. A szabályok történeti előzményeinek lehetséges fonetikai okai**

A vizsgált szótagok hiányának magyarázata a szakirodalmi előzmények nyomán abban keresendő, hogy nohár e szekvenciák képzése artikulációs szempontból rendkívül gazdaságos (hiszen a hangképző szerveknek relatíve kis távolságot kell áthidalniuk a produkció során), a percepció szempontjából mégsem lehet kedvelt. Mivel a percepció számára az a hangsor minősül optimálisnak, amelyben a szegmentumok minél inkább eltérnek egymástól, azaz minél jobban megkülönböztethetők, az artikulációs szempontból oly kényelmes /ji/, /vi/ szekvencia kerülendőnek bizonyul. (Ugyanakkor azt is megfigyelték, hogy e félhangzó+vokális megszorítás nem illeszkedik valamely általánosabb tendenciába, azaz például a labiális mássalhangzókat szívesen követik hátul képzett labiális vokálisok, mint pl. vö. Maddieson–Precoda 1992, 55-56).

Első pillantásra is úgy látszik, hogy e percepciós indítatású tendencia már az alapnyelvben érvényesülhetett. Ez azonban csak akkor igazolható, ha az egyéb morfémakezdő CV-szekvenciákkal, ti. azok artikulációs, illetőleg percepciós tulajdonságaival szembesítjük.

A pillanatnyilag rendelkezésre álló teljes PFU szóanyag feldolgozásának eredményeként előálló alábbi sorrend azt mutatja, hogy a (képzéshelyi osztályokba rendezett) PFU mássalhangzók elöl-, illetőleg hátul képzett vokálisokkal állnak-e együtt szívesebben (ti. szívesebben, mint ahogyan pusztán a hangok gyakorisági mutatói alapján elvárható lenne – Bakró 1993).

	PAL (*i *e *ä)	VEL (*u *o *a)
LAB (*p *m)	+	-
(*w)	-	+
KAK (*c *s)	+	-
DENT(*t *n *s)	+	-
ALV (*l *r)	=	=
ALVP(*s' *c' *n' δ')	=	=
PAL (*j)	+	-
VEL (*k)	-	+

A preferencia sorrend a következőket mutatja: a labiális, kakuminális és dentális mássalhangzókat szívesebben követik elől képzett magánhangzók; az alveoláris és alveopalatális mássalhangzók nem tüntetik ki sem egyik, sem másik magánhangzó osztályt; a veláris konzonáns szívesebben áll együtt hátul képzett magánhangzókkal. E tendencia alól, amelyik egyértelműen az artikulációs szempont érvényesülésére vall mindezen esetekben, csupán két kivétel mutatkozik, nevezetesen a labiálisok osztályától elkülönülően viselkedő \*w, s az alveopalatálisokhoz affinitást nem mutató, egyedüli palatális, a \*j: ezek ketten éppen hogy ellenkező tendenciáról tanúskodnak. Ha pedig így van, akkor kénytelenek vagyunk megállapítani, hogy esetükben a percepció szempont a döntő, illetőleg kevésbé kategórikusan fogalmazva: döntőbb, mint a többi szekvencia esetében.

A percepció szempont érvényesülése mellett még további érv is fölhozható. Tudjuk, hogy a nyelvi változások során a hangok egy része stabilabb, ellenállóbbnak bizonyul. A félhangzók a kevésbé stabilak közé tartoznak: vokalizálódásuk, teljes kiesésük megszokott jelenség. Kérdés azonban, hogy oka a fonetikán belül mikor köthető az artikulációhoz s mikor a percepcióhoz. Ha e változások gyöngülési folyamatok realizálódásai, akkor nyilvánvaló az artikulációs indíték. Mivel azonban a gyöngülések tipikus (bár nem kizárólagos) helye a szekvenciában az

intervokalikus, illetőleg a szó- vagy szótagvégi helyzet, a gyöngülés a mi esetünkben nem jöhet számításba. (Messzire, s ide nem illő nyelvtörténeti fejtegetésekhez vezetne annak végig gondolása, hogy a PFU-ra feltett, igen csekély számú *\*ji/\*wu* szókezdetet mennyire indokolt valóban félhangzós kezdetűnek rekonstruálni. Annyit azonban minden esetre érdemes megjegyeznünk, hogy esetleges elvetésük még csak erősítené a szabály alapnyelvi kori érvényességét.) Mondandóm végül is e distinkció megtételének a lehetőségét célozta.

### Irodalom

Bakró-Nagy Marianne.: PFU #CV. In: Bakró-Nagy M.–Szij E. (szerk.): Hajdú Péter 70 éves. *Linguistica. Series A. Studia et dissertationes*, 15. MTA Nyelvtudományi Intézete. Budapest 1993, 33-40.

Karlsson, F.: Suomen kielen aane- ja muotorakenne. Porvoo–Helsinki 1982.

Maddieson, I.–Precoda, K.: Syllable structure and phonetic models. *Phonology* 9. 1992, 45-60.

Nádasdy Ádám–Siptár Péter: A magánhangzók. In: Kiefer Ferenc (szerk.): *Strukturális magyar nyelvtan 2. Fonológia*. Akadémiai Kiadó. Budapest 1994, 42-182.

Ohala, J.J.–Kawasaki, H.: Prosodic phonology and phonetics. *Phonology Yearbook* 1. 1984, 113-127.

Papp, F.: *Reverse Alphabetized Dictionary of the Hungarian Language*. Budapest 1969.

Szende Tamás: Spontán beszédanyag gyakorisági mutatói. *Nyelvtudományi Értekezések* 81. Budapest 1973.

# IDŐTARTAM-ÍTÉLETEK FÜGGŐSÉGI VISZONYAI

**Kovács Magdolna**

**KLTE Általános és Alkalmazott Nyelvészeti Tanszék**

A magyar magánhangzó-rendszer jellemzésének kevésbé feldolgozott kérdésköre a fonológiai releváns rövid/hosszú tartamtípusok nyelvi tudatbeli reprezentációja. A tanulmány a fizikai, az észlelt és a nyelvi időtartam viszonyrendszerét, ezen belül a fizikai időtartamnak és a mássalhangzó környezetnek a percepció felismerésben játszott szerepét vizsgálja.

## **Bevezetés**

A magyar beszédhangok, s ezen belül a magánhangzók időtartamviszonyainak eszközfonetikai vizsgálata nagy hagyományokkal rendelkezik. Gombocz, Magdics, Tarnóczy, Kassai, Olasz és mások kutatási eredményei képet adnak a hangok fizikai megvalósulásának időtartamát befolyásoló tényezők (úgy mint a hangnak a hangsorban elfoglalt helye, a hangsor terjedelme, hangkörnyezet, beszédtempó stb.) hatásmechanizmusáról. E tényezők mint változó paraméterek függvényében változik a fonológiai rövid/hosszú magánhangzó párok időtartamaránya; folyamatos beszédben a két nyelvi tartamtípus időtartamértékeinek szóródásában jelentős az átfedés: Kassai mérési eredményei alapján a hangsúlyos rövid magánhangzók átlag időtartama 100 ms, szóródási zónája 55-195 ms, míg a hosszúakra 195-210 ms közötti átlagidőtartam és 90-385 ms-os szóródási zóna adódott (Kassai 1982, 131).

Kevesebbet tudunk azonban a két tartamtípus nyelvi tudatbeli reprezentációjáról. Az időtartamnak a magyar magánhangzók percepciójában játszott szerepével kapcsolatban eddig Gósy végzett kísérleteket (Gósy 1989). Egyik vizsgálata, amelyben a frekvencia- és az időtartam-változás összefüggését izolált magánhangzók első két formánsának és időtartamának lépésenkénti

változtatása alapján elemzi, az időtartam-ítéletek bizonytalanságáról tanúskodik: a 100, 200, 300 és 400 ms időtartam értékek közül csak a 300-400 ms fizikai időtartamnál, azaz csak a megfigyelt produkciós átlagot jóval meghaladó értékeknél fordultak elő nagyobb arányban a nyelvileg hosszú magánhangzó-felismerések. Az izolált tiszta fázisok alapján különösen bizonytalan volt a hosszú magánhangzó jelölése az [i, i:] és [u, u:] hangpároknál.

Felvetődik a kérdés, hogy mi okozta a relatív időtartam megítélésének bizonytalanságát. Elvárható-e egyáltalán rövidebb időtartamok esetén is nyelvileg hosszú magánhangzó azonosítás vagy a két tartamtípus szóródási zónájának említett átfedése miatt csak a hosszú magánhangzók megvalósulásának felső határértékei biztosítják a percepcióban a hosszú tartamtípussal való azonosítást? Befolyásolta-e az időtartam-azonosítást, hogy az adatközlők izolált tiszta fázisokat hallottak?

A szakirodalomban ugyanis eltérő annak megítélése, hogy az izolált magánhangzók, különösképpen a tiszta fázisok elegendő akusztika információt tartalmaznak-e a biztos azonosításhoz. Strange et al. (1976) és Verbrugge et al. (1976) statisztikailag szignifikáns különbséget találtak ugyanazon angol magánhangzók felismerési hibaszázalékában az izolált, ill. a mássalhangzó környezetet is tartalmazó stimulusoknál: az izolált magánhangzók esetében 31,2%-os hibaszázalék /p\_p/ mássalhangzó környezetben (azaz *pip*, *peep*, *pup* stb. szavakban) 9,5%-ra csökkent. A "mássalhangzó-környezet effektus" hatása azonban nem érvényesült például Diehl et al. (1981) és Assmann et al. (1982) kísérleti eredményeiben: a természetes ejtésű izolált magánhangzókat a kísérleti személyek igen magas arányban azonosították helyesen. Assmann kísérletében például a teljes magánhangzók (szótagmag) felismerésének a hibaszázaléka mindössze 4,09% volt, s a tévesztések aránya az azonos időtartamú tiszta fázisok percepciójában is csupán 9,5%-ra emelkedett.

Az alábbi kísérlet célja az volt, hogy természetes hangsorból szegmentált palatális magánhangzók felismerésére kapott eredmények alapján képet alkossak a fizikai időtartamnak és a mássalhangzó környezetnek a magyar magánhangzók felismerésében játszott szerepéről.

### Anyag és módszer

A percepciós kísérlethez egy művelt, igényes köznyelvet beszélő férfi bemondóval ugyanazon hordozó mondatba (*Nem tudta, hogy van a \_\_\_ angolul.*) foglalt 1 és 2 szótagú szavakat rögzítettem DAT magnetofonra. Ezt a hanganyagot a KAY Elemetrics 4300 beszédelemző berendezéssel 10000 Hz-es mintavételezéssel digitalizáltam és feldolgoztam. A percepciós kísérlet nyersanyagául választott – a produkciós anyagban változó mássalhangzó környezetben, hangsúlyos VC, CVC (pl. *üt, fűzér, fűzés, tűz* stb.) hangkapcsolatokban ejtett – magánhangzók akusztikai jellemzőit az 1. ábra tartalmazza.

Az oszcillogram, a spektrogram és az intenzitásgörbe vizuális elemzésére (ld. 2. ábra) és állandó auditív ellenőrzésre támaszkodva, a hanganyagból összesen 110 stimulust szegmentáltam. A szegmentált szakaszok az alábbi típusokat képviselték:

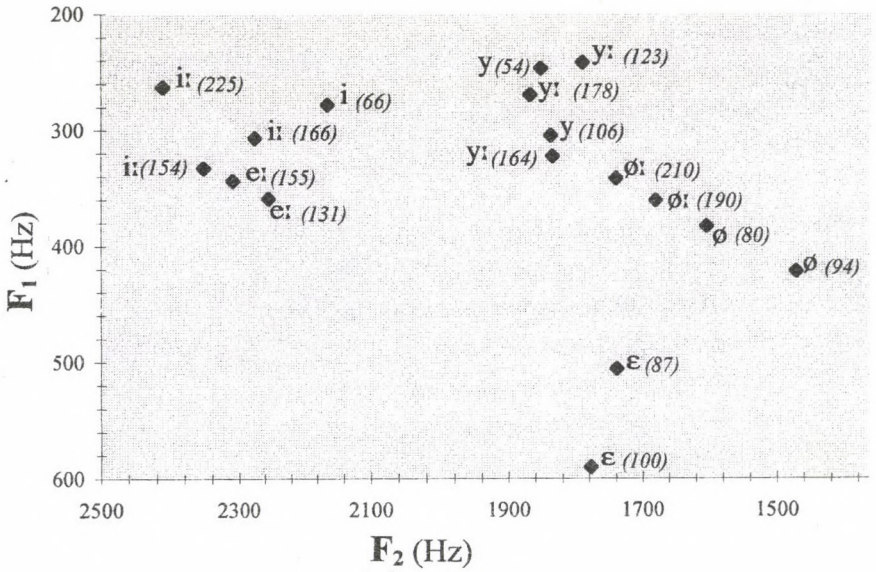
1. a hangsúlyos magánhangzó tiszta fázisa:  $V_{\text{tisza}}$ ;

2. a tiszta fázis a megelőző és/vagy a magánhangzót követő mássalhangzóhoz való átmenettel:  $\text{átm}_1 + V_{\text{tisza}}$

$$\begin{aligned} &V_{\text{tisza}} + \text{átm}_2 \\ &\text{átm}_1 + V_{\text{tisza}} + \text{átm}_2; \end{aligned}$$

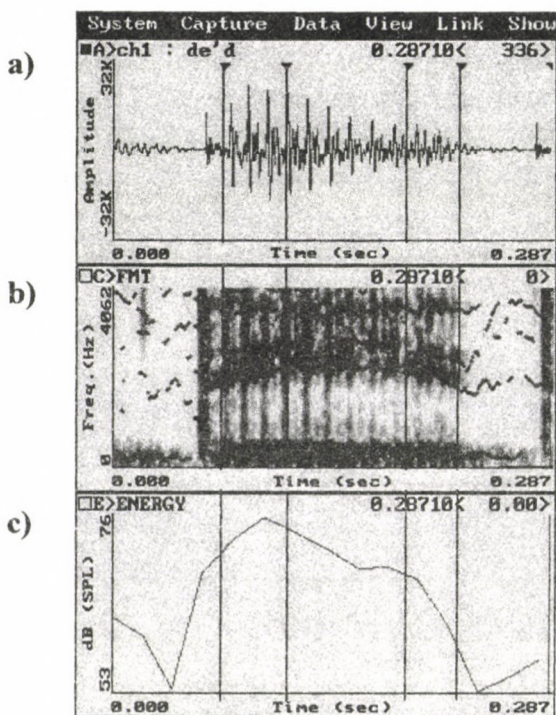
3. az átmeneteket tartalmazó szakaszok a megelőző és/vagy a követő mássalhangzóval együtt:  $C + \text{átm}_1 + V_{\text{tisza}}$

$$\begin{aligned} &C + \text{átm}_1 + V_{\text{tisza}} + \text{átm}_2 \\ &V_{\text{tisza}} + \text{átm}_2 + C \\ &\text{átm}_1 + V_{\text{tisza}} + \text{átm}_2 + C \\ &\text{CVC}. \end{aligned}$$



1. ábra

A percepciós kísérlet nyersanyagául használt hangsúlyos magánhangzók akusztikai jellemzői (A formánsadatok az LPC elemzéssel nyert formáns-középvértékeknek a hangzók tiszta fázisára számított átlagai, a zárójelben megadott értékek pedig a magánhangzók (szótagmag= $\text{átm}_1 + V_{\text{tisztá}} + \text{átm}_2$ ) időtartam adatai ezredmásodpercben.)



2. ábra

A *déd* hangsor szegmentálása az a) oscillogram, b) az LPC formánkövetéssel kiegészített spektrogram és c) az intenzitásgörbe alapján

A stimulusokat véletlen sorrendben tartalmazó hanganyagot 15 kísérleti személy (23 és 50 év közötti értelmiségiek) hallotta. Az adatközlők a tesztlapon előre megadott magánhangzó betűjelek közül választották ki azonosított hangnak megfelelőt.

## Eredmények, következtetések

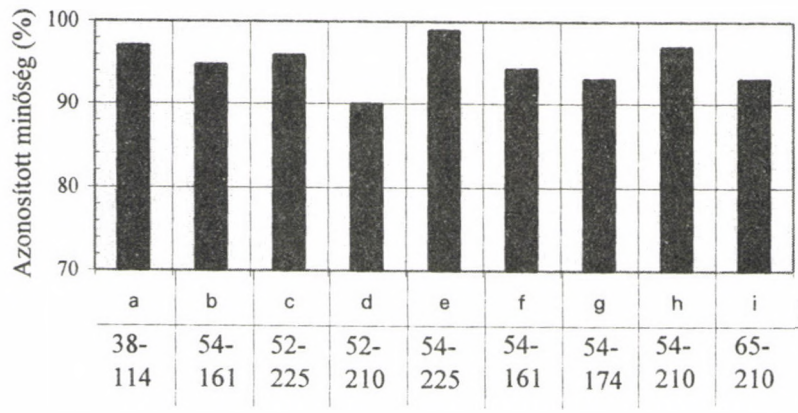
### *I. A hangminőség felismerése*

A kísérleti személyek a hangminőséget valamennyi palatális magánhangzónál igen nagy biztonsággal ismerték fel: a teljes kísérleti korpuszra számított helyes azonosítási átlag 94,87%. 30%-nál magasabb hibaszázalék mindössze 3 stimulusnál adódott (l. 1. táblázat).

Megvizsgáltam, hogy az előfordult hangminőség tévesztések kötődnek-e tipikusan valamely hangminőséghez, stimulus típushoz, vagy a magánhangzók relatíve hosszú, ill. rövid időtartamához.

A magánhangzók között a minőség felismerés csökkenő arányában a következő sorrend alakult ki: [ë, ë:, y, E, i, y:, e:, i:]; az [ë, ë:] hangok azonosítása csaknem 100%-os, de a legbizonytalanabban azonosított [i:] -t is 87,1%-ban felismerték a kísérleti személyek. Mind a 70%-nál rosszabbul azonosított magánhangzókat, mind a kísérleti anyag egészében előforduló további szórványos hibákat tekintve is megállapítható, hogy a tévesztések leggyakrabban az [i:] hang [e:] -ként, ritkábban az [e:] hang [i - i:] -ként, és az [y:] hang [ë - ë:] -ként való azonosításában jelentkeztek, azaz a korábbi megfigyelésekkel egybehangzóan azokat a magánhangzókat érintették, amelyekben nagy az első és második formáns távolsága.

A 3. ábráról leolvasható, hogy a hangminőség tévesztések arányában nincs kiugró eltérés a stimulus típusok függvényében, valamennyi esetben magas, 90%-os vagy a fölötti volt a helyes azonosítások aránya. Figyelemre méltó, hogy mind a 38-114 ms között szóródó tiszta fázisok, mind a mássalhangzó környezetüktől megfosztott hangátmeneti fázis(ok)at is tartalmazó 90 ms alatti stimulusok felismerési aránya 97%-os volt. Az adatok egyértelműen azt mutatják, hogy a vizsgált magánhangzók formánsszerkezete igen rövid időtartamok mellett és a mássalhangzó környezet nélkül is biztos támpontot nyújtott a minőség helyes azonosításához.



**Jelmagyarázat:**

- a =  $V_{tisza}$
- b =  $\acute{a}tm_1 + V_{tisza}$
- c =  $V_{tisza} + \acute{a}tm_2$
- d =  $\acute{a}tm_1 + V_{tisza} + \acute{a}tm_2$
- e =  $C + \acute{a}tm_1 + V_{tisza}$
- f =  $V_{tisza} + \acute{a}tm_2 + C$
- g =  $\acute{a}tm_1 + V_{tisza} + \acute{a}tm_2 + C$
- h =  $\acute{a}tm_1 + V_{tisza} + \acute{a}tm_2 + C$
- i = CVC

Stimulus típus	a	b	c	d	e	f	g	h	i
Időtartam (ms)	38-114	54-161	52-225	52-210	54-225	54-161	54-174	54-210	65-210

3. ábra  
A palatális magánhangzók hangminőségének azonosítása

Ezzel kapcsolatban érdekes adalékul szolgált a 3 legbizonytalanabbul azonosított stimulus elemzése.

1. táblázat

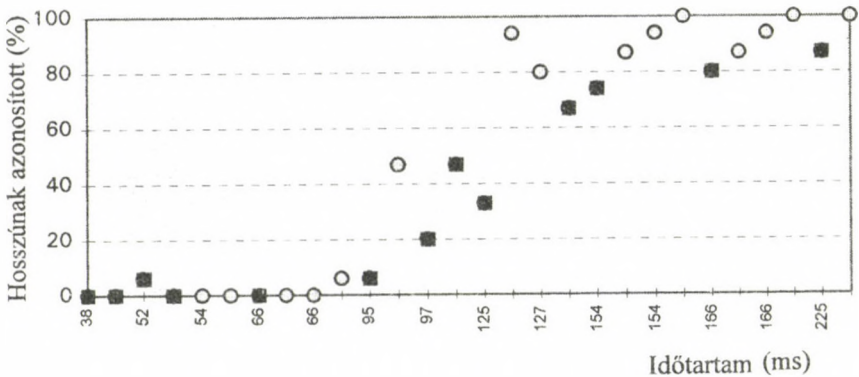
Hang	Időtartam (ms)	Stimulus típus	Hangminőség azonosítása (%)
i:	166	$\text{átm}_1 + V_{\text{tisza}} + \text{átm}_2$	67
i:	97	$V_{\text{tisza}}$	94
i:	154	$\text{átm}_1 + V_{\text{tisza}} + \text{átm}_2 + C$	60
i:	66	$V_{\text{tisza}}$	100
e:	131	$\text{átm}_1 + V_{\text{tisza}} + \text{átm}_2$	26
e:	75	$V_{\text{tisza}}$	94

Az 1. táblázat adatai azt mutatják, hogy az ugyanazon szavakból szegmentált szótagmagok, ill. tiszta fázisok azonosítása között jelentős volt a különbség. Azt várhatnánk, hogy minél több akusztikai információt tartalmaz a hangszelet, annál biztosabb a felismerés. A kísérlet eredményei azonban éppen az ellenkezőjét tükrözik: a rövidebb időtartamú tiszta fázisok felismerése mindhárom esetben lényegesen jobb, mint az átmeneti fázisokat is tartalmazó és hosszabb szegmentumoké. Úgy tűnik, ebben az esetben a szakirodalomban már néhány percepció vizsgálatnál tapasztalt paradoxonnal van dolgunk: ha nem jelentésszerű egységekről kell döntenünk, a több akusztikai információ rosszabb felismerést eredményez.

## II. Az [i, i:], [y, y:], [ë, ë:] hangpárok nyelvi időtartamának megítélése

Az [i, i:], [y, y:], [ë, ë:] hangpárok nyelvi hosszúságának megítélésében az a nyilvánvaló elvárás igazolódott, hogy minél

hosszabb a hang, annál nagyobb a valószínűsége a nyelvileg hosszú magánhangzó-azonosításnak. A 4. ábra a fizikai időtartam és a nyelvileg hosszú magánhangzó-azonosítások összefüggését mutatja az [i, i:] hangpár példáján. Azt látjuk, hogy a 38-66 ms közötti stimulusokat a kísérleti személyek egyöntetűen rövid magánhangzókkal azonosították, a 154-225 ms közöttiek pedig nagy arányban hosszúnak ítélték. E két időhatár között a nyelvi hosszúság megítélése bizonytalan. A grafiknról az is leolvasható, hogy a hosszú magánhangzók azonosítása nagyobb arányban jelentkezik a mássalhangzó-környezetet tartalmazó stimulusoknál akkor is, ha fizikai időtartamuk rövidebb. Izolált magánhangzóknál a nyelvi döntés jobban követi a fizikai időtartamok alakulását.

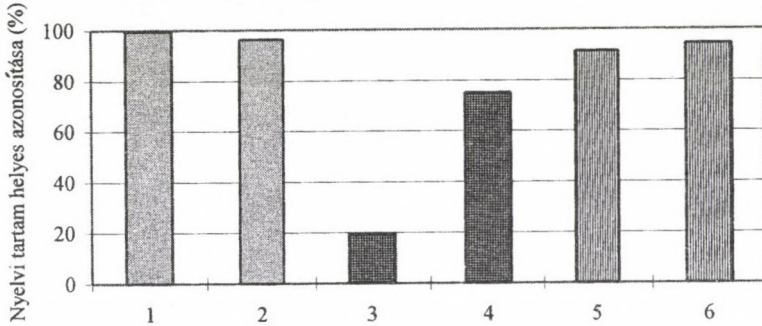


4. ábra

Az [i] - [i:] hangok hosszúságának megítélése (A négyzet az izolált magánhangzókra, a kör pedig a mássalhangzó környezetet is tartalmazó stimulusokra kapott adatokat jelöli.)

A produkcióban szándékolt hosszúság felismerésére kapott eredmények alapján (l. 5. ábra) a következők állapíthatók meg:

A választott kísérleti feltételek mellett az 54-106 ms között szóródó nyelvileg rövid és a 123-225 ms között szóródó nyelvileg hosszú magánhangzók ( $\text{átm}_1 + V_{\text{tisza}} + \text{átm}_2$ ) közötti időtartamkülönbség elegendő támpontot biztosított a kísérleti személyeknek mind a két tartamtípus helyes azonosításához.



#### Jelmagyarázat:

1. [i, y, ø] -C (54-106 ms)	3. [i, y:, ø:] -C (90-130 ms)	5. [i:, y:, ø:] -C (150-225 ms)
2. [i, y, ø] +C (54-106 ms)	4. [i:, y:, ø:] +C (90-130 ms)	6. [i:, y:, ø:] +C (150-225 ms)

#### 5. ábra

#### A nyelvi tartam helyes azonosítása

Az izolált hosszú szótagmagok felismerése csak a *fűzés* szóból szegmentált 123 ms-os [y:] hang esetében nem éri el a 75%-ot. A hosszú magánhangzók azonosításának bizonytalansága általában jellemzi – a viszonyítási alapul választott produkciós szórás legalsó sávját jelentő – 90-130 ms-os izolált magánhangzók felismerését. Ezekben az esetekben érvényesült a legerőteljesebben a "mássalhangzó-környezet effektus" hatása (ld. 3. táblázat): a CV és VC hangsorokban, s különösképpen az értelmes szavakban

ugyanazon az izoláltan rövidnek minősített magánhangzót a kísérleti személyek hosszú beszédhanggal azonosították. Feltételezhetnénk, hogy például a [fy:] hangsorban az azonosítást kizárólag az a tény javítja, hogy jelentésszerűről van szó, s a kodifikált helyesírás hatott az adatközlők percepciójára. Megfigyelhető azonban, hogy a hosszú magánhangzók felismerésének arányát ezekben az esetekben is befolyásolja az abszolút fizikai időtartam: nagyobb időtartam esetén a hosszú magánhangzó azonosítás aránya is nagyobb. Nagyon valószínűnek tűnik továbbá, hogy a szótag felépítése is hatással van a nyelvi hosszúság megítélésére.

2. táblázat

Eredeti hangkapcsolat	Időtartam (ms)	-C		+C	
		Stimulus típus	Nyelvileg hosszú azonosítás (%)	Stimulus típus	Nyelvileg hosszú azonosítás (%)
[si:]	95	ál+t	6	C+a1+t	47
[fy:]	93	ál+t	13	C+ál+t	67
[fy:]	123	ál+t+a2	47	C+ál+t+a2	87
[hë:]	121	ál+t	40	C+ál+t	74
[kë:]	161	ál+t	94	C+ál+t	100
[y:z]	104	t+a2	0	t+a2+C	47
[y:z]	123	ál+t+a2	47	a1+t+a2+C	87
[i:t]	125	t+a2	33	t+a2+C	94

A jelen kísérlet igyekezett támpontokat, kiinduló adatokat találni annak feltérképezéséhez, hogy mely tényezők a meghatározók a magyar magánhangzók tartamtípusainak nyelvi tudatbeli reprezentációjában. A korábbi megfigyelések alapján is ismert volt, hogy a fizikai, az észlelt és a nyelvi időtartam nem

azonos; azonban ezek viszonyrendszere csak kísérletekből nyert adatok – a teljes magyar magánhangzó-rendszer további szisztematikus vizsgálata – alapján írható le.

### Irodalom

Assmann, P.F.–Nearey, T.M.–Hogan, J.T.: Vowel identification: Orthographic, perceptual, and acoustic aspects. *JASA* 71. 1982, 975-989.

Diehl, R.L.–McCusker, S.B.–Chapman, L.S.: Perceiving vowels in isolation and in consonantal context. *JASA* 69. 1981, 239-248.

Gombocz Zoltán: A magyar hangok időtartamáról. *Nyelvtudomány II.*, Budapest 1909, 93-100.

Gósy Mária: Beszédészlelés. MTA Nyelvtudományi Intézete. Budapest 1989.

Kassai Ilona: A magyar beszédhangok időtartamviszonyai. In: *Fejezetek a magyar leíró hangtanból*. Szerk.: Bolla Kálmán. Budapest 1982. 115-154.

Magdics Klára: A magyar beszédhangok időtartama. *NyK LXVIII*, 1966, 125-139.

Olaszy Gábor: Hangidőtartam-módosító kísérletek a gépi beszéd ritmusának javítására. In: *Beszéd kutatás '94*. Szerk.: Gósy Mária. MTA Nyelvtudományi Intézete. Budapest 1994, 140-152.

Strange, W.–Verbrugge, R.–Shankweiler, D.–Edman, T.: Consonant context specifies vowel identity. *JASA* 60. 1976, 213-224.

Tarnóczy Tamás: A magánhangzók akusztikai vizsgálatának problémái. *ÁltNyT X.*, 1974, 181-196.

Verbrugge, R.–Strange, W.–Shankweiler, D.–Edman, T.: What information enables a listener to map a talker's vowel space? *JASA* 60. 1976, 198-212.

# AZ [u:], [a:] ÉS [i] HANGOK KOARTIKULÁCIÓS MEZŐIRŐL

Szalai Enikő  
MTA Nyelvtudományi Intézete

A jelen tanulmányban arra kerestünk választ, hogy a mássalhangzó-környezet mennyire befolyásolja a magánhangzók tiszta fázisainak formánsszerkezetét. Az általános formánsérték-meghatározásokon túl definiáljuk és elemezzük a három kardinális magyar magánhangzó, az [u:], az [a:] és az [i] ún. koartikulációs mezőit.

## Bevezetés

Az akusztikai fonetika kezdeti vizsgálatai közé tartozott a beszédhangok akusztikai jellemzőinek meghatározása. A kutatások kiindulása az a megállapítás volt, amely szerint a beszédhangok egyértelműen meghatározható artikulációs konfigurációval és ennek következtében, pontosan definiálható akusztikai paraméterekkel rendelkeznek.

Az artikulációs folyamatot tekintve a magánhangzók a legegyszerűbben kivitelezhető beszédhangok. Akusztikai szerkezetüket alapvetően az időtartam, az alaphfrekvencia, a formánsok frekvenciaértéke, azok amplitúdója, valamint a formánsok sáv szélessége határozza meg. E jegyek közül – a kísérletek tanúbizonysága szerint – a hangminőség, a beszéd-percepció szempontjából döntően a spektrális eloszlás, ezen belül is az első két formáns értéke a meghatározó (Flanagan, 1965).

A magyar magánhangzók formánsértékeinek meghatározására már a 40-es évek elejétől találunk példát (Tarnóczy-féle Fourier elemzés, 1941), majd az eszközfonetika további fejlődése során Magdics Klára (1963), Molnár József (1969), Bolla Kálmán (1980) és legutóbb Olaszgy Gábor (1989) közölnek adatokat spektrografikus elemzések alapján.

Az első táblázat a három kardinális magánhangzó az [u:] az [a:] és az [i] első, második és harmadik formánsának határértékeit szemlélteti a fent említett kutatók vizsgálatainak alapján. (A táblázat a hangsúlyos és hangsúlytalan magánhangzók adatait összevontan tartalmazza.)

1. táblázat: Az [u:, a:, i] hangok F1, F2 és F3 határértékei férfi ejtésben

	[u:]			[a:]			[i]		
(Hz)	F1	F2	F3	F1	F2	F3	F1	F2	F3
Magdics	250- 350	600- 800	1950- 2200	700- 900	1300- 1500	2000- 2300	250- 400	1950- 2200	2600- 3000
Molnár	250- 370	600- 730	2000- 2250	700- 900	1300- 1500	2000- 2400	250- 400	1950- 2200	2600- 3000
Bolla	200- 350	600- 800	2000- 2200	650- 850	1250- 1450	1950- 2250	250- 400	2100- 2300	2800- 3000
Olaszy	300- 400	600- 800	1900- 2000	700- 900	1300- 1500	1900- 2300	300- 400	2100- 2300	2800- 3000

Annak ellenére, hogy mindegyik kísérlet ugyanazon nyelv, ugyanazon magánhangzóinak formánsértékeit határozta meg, az adatok nem fedik egymást minden esetben. A különbségek adódhatnak a vizsgált nyelvi korpusz eltéréseiből, a kísérleti személyek egyéni, ejtésbeli különbségeiből és nem hagyható figyelmen kívül a vizsgálatok során használt módszerek, illetve eszközök különbözősége sem.

Az egymástól eltérő adatok nyomán felmerül a kérdés, hogy csupán a formánsok közéértékei, illetve a szélsőértékek által meghatározott különböző frekvenciatartományok elegendő és pontos információval szolgálnak-e a vizsgált magánhangzó formáns-szerkezetéről? Valóban a folyamatos beszédben megvalósuló akusztikai paramétereket tükrözik-e?

A szakirodalmi vizsgálatok mindegyike evidenciaként kezeli, hogy az egyes beszédhangok egymástól nem izolált egységekként, hanem egymásra kölcsönösen hatást gyakorló alkotóelemekként, hangkapcsolatok formájában hozzák létre a beszéd folyamatát. Így az eredmények elemzése során figyelembe kell vennünk a Lindblom (1963) által "tökéletlen megvalósításnak" ("target undershoot") nevezett jelenséget, amelynek értelmében CVC kapcsolatokban még a magánhangzók tiszta fázisainak formánsértékei sem egyeznek meg egyértelműen az izoláltan ejtett magánhangzók formánsértékeivel.

Az átlag- és a szélsőérték-adatok tehát elfedik az adott magánhangzó formánsértékeiben jelentkező különbségeket, és nem szolgálnak információval arra a kérdésre vonatkozóan, hogy hogyan alakulnak az egyes magánhangzók formánsértékei CVC hangkapcsolatokban, illetve milyen az F1 és az F2 viszonyrendszere. Éppen ezért a nemzetközi fonetikai szakirodalom a magánhangzók formánsstruktúrájának meghatározására a kétdimenziós ábrázolási módot használja, amely lehetőséget nyújt az egyes magánhangzók frekvenciaértékeiben jelentkező "átfedések", így a magánhangzók megbízható diszkriminációjához elégséges fonetikai kontraszt megállapítására.

A magyar fonetikai szakirodalomban a teljes magyar magánhangzóképzlet e szempontok szerint történő spektrografikus vizsgálatát kizárólag Tarnóczy (1965), majd gyermeknyelvi anyagon az [i], [e] és [y] magánhangzókra vonatkozóan Gósy (1995) végezte el.

Az eddig említett, nagyszámú, részletes vizsgálatok vitathatatlan eredményei ellenére is számos kérdés merül fel arra vonatkozóan, hogy a valóságnak megfelelően milyen frekvenciatartományokba esik a percepció szempontjából alapvetően meghatározó első három formáns értéke.

Éppen ezért, jelen kísérletünk céljaul a három kardinális magánhangzó, az [u:], az [a:] és az [i] formánsszerkezetének vizsgálatát tűztük ki.

A vizsgálat során arra kerestünk választ, hogy mennyire befolyásolja a mássalhangzókönyezet a magánhangzók formánsszerkezetét; az eddigi vizsgálatokból jól ismert jellemző frekvenciatartományokon belül vagy túl, milyen ún. **koartikulációs mezőkkel** jellemezhetők az adott magánhangzók.

Koartikulációs mezőnek neveztük azt az első és második formánsok által meghatározott frekvenciatartományt, amely a magánhangzóknak a különböző mássalhangzó-környezetekben megjelenő formánsérték-változatait tartalmazza.

### **Anyag és módszer**

A három magánhangzót CVC hangkapcsolatokba ágyztuk; ahol hangkörnyezetként az összes magyar mássalhangzót (a [h], [dz] és [dʒ] kivételével) az összes lehetséges variációban felhasználtuk. A 3×22×22, azaz 1452 szótagot egyetlen férfi ejtésében laboratóriumi körülmények között STUDER A-80-as típusú magnetofonnal rögzítettük, majd CSL 4300-as jelfeldolgozó segítségével a magánhangzók tiszta fázisairól spektrális elemzéseket készítettünk.

Az eredményeket befolyásoló lehetséges változó faktorok közül ezúttal tehát a beszélőt tekintettük állandónak, egyetlen férfi bemondó személyében. A jelen kísérletben nem foglalkoztunk az átmenetek kérdéskörével, kizárólag a magánhangzók tiszta fázisait vizsgáltuk. A tiszta fázisok formánsértékei alapján állapítottuk meg a három vizsgált magánhangzó koartikulációs mezőjét.

### **Eredmények**

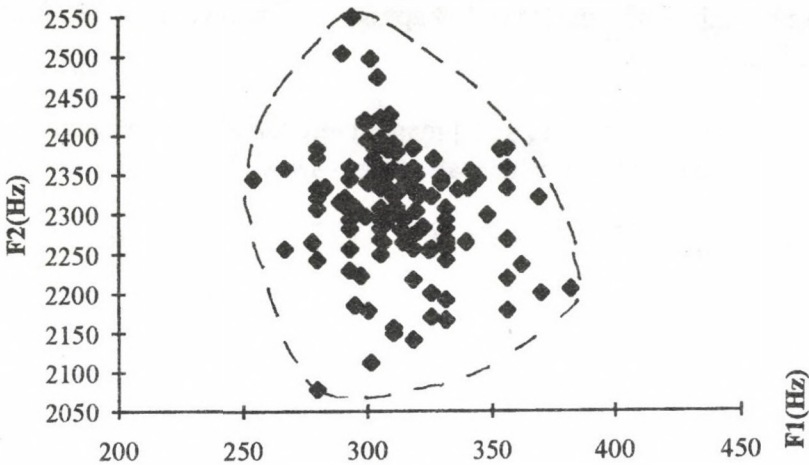
Az összesen 4356 adat alapján meghatározott átlag- és szélsőértékeket a 2. táblázat szemlélteti. (A középértékeket az esetenként kapott értékek összeadásával, majd az összegnek az

esetszámmal való osztásával kaptuk, ez magyarázza a nem kerekített értékeket.)

2. táblázat: Az [u:, a:, i] magánhangzók első, második és harmadik formánsának átlag- és szélsőértékei

	(Hz)	[u:]	[a:]	[i]
<b>F1</b>	szélső- értékek	257 – 414	665 – 892	252–382
	átlagok	<b>346</b>	<b>794</b>	<b>310</b>
<b>F2</b>	szélső- értékek	637 – 803	1300 – 1496	2179 – 2559
	átlagok	<b>725</b>	<b>1408</b>	<b>2335</b>
<b>F3</b>	szélső- értékek	2228 – 2705	2340 – 2971	2798 – 2951
	átlagok	<b>2461</b>	<b>2514</b>	<b>2866</b>

Az átlag- és szélsőértékek meghatározásával a korábbi kísérletek korszerűbb technikai feltételek mellett végrehajtott reprodukálása lett volna csupán jelen vizsgálatunk. Ez önmagában is újabb adalékkal szolgált volna a magánhangzók formánsstruktúrájának meghatározásához, de mivel jelen kísérletünk célja a magánhangzót érő koartikulációs folyamatok hatásának vizsgálata, ezért az első két formáns elhelyezkedését kétdimenziós ábrázolásban szemléltettük, ahol a horizontális tengelyen az F1, a vertikális tengelyen pedig az F2 értékeit tüntettük fel. Az ábrákon a szótagkezdő mássalhangzók szerinti formánsérték-eltéréseket szemléltetjük, így lehetőség nyílik az F1 és F2 viszonyrendszerének, valamint a hangkörnyezet hatásának tanulmányozására.



1. ábra

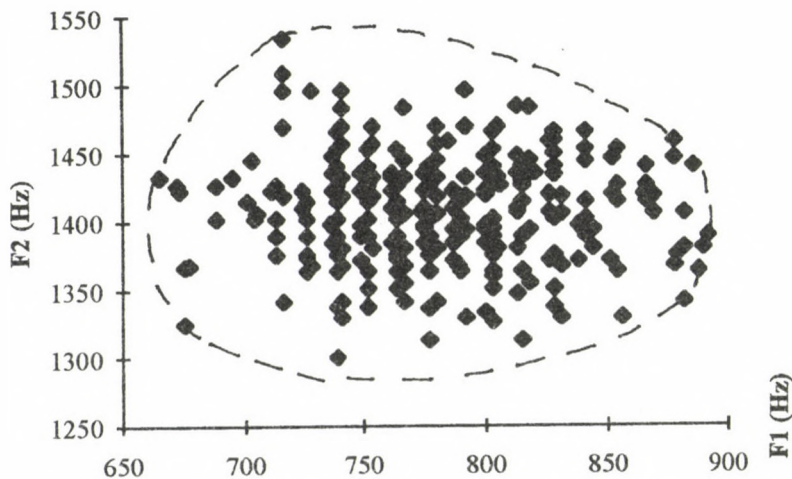
Az [i] hang koartikulációs mezője CVC hangkapcsolatok alapján, férfi ejtésben

Már az [i] hang koartikulációs mezőjének ábrázolásából (1. ábra) kitűnik, hogy a mező kifejezés többet árul el a formánsértékek elhelyezkedéséről, mint csupán a szélsőértékek által meghatározott frekvenciasávok. Egyrészt szemléltetni képes a koartikuláció akusztikai következményeit, másrészt lehetőséget teremt a különböző hangkörnyezetek hatására végbemenő változások összevetésére.

Az elvárásoknak megfelelően, az [i] hang mezőstruktúrájával összehasonlítva egészen más alakzatú mezőben helyezkednek el az [a:] hang formánsai (2. ábra).

Míg az [i] hang esetében függőleges irányban, az F2 tengely mentén elnyúlt mezőt kapunk, az [a:] esetében az F2 irányú kiterjedés jóval szűkebb, a nagyobb szórás az első formáns értékeiben jelentkezik. Az eredmények alátámasztják azt a megállapítást, hogy a hangkörnyezet hatása elsősorban a második formánst mozditja el (vö. 1. ábra).

Az első formáns azoknál a magánhangzóknál mozdul el nagyobb mértékben, amelyeknél az első és a második formánst kis távolság választja el egymástól (vö. 2. és 3. ábra).



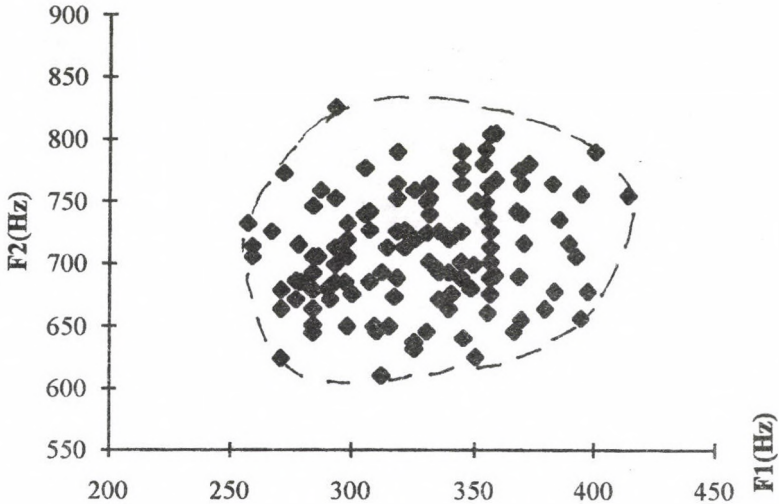
2. ábra

Az [a:] hang koartikulációs mezője CVC hangkapcsolatok alapján, férfi ejtésben

A legszimmetrikusabb alakzatú mezőt az [u:] esetben kaptuk, ahol a formánsértékek egyenletes eloszlásban, az átlagértékek köré, a mező centrumában csoportosulva, extrém szélsőértékek nélkül jellemzik a magánhangzók tiszta fázisait a CVC hangkapcsolatokban.

Az elemzés során lehetőség nyílt a mássalhangzó-környezet módosító hatásának összevetésére is a három vizsgált mássalhangzó esetében. A megállapítások természetesen nem tekinthetjük tézis-értékűeknek, hiszen csak a kardinális magánhangzókat és ezek lehetséges mássalhangzókapcsolódásaink egy kiválasztott csoportját (CVC) figyelembe vevő kísérletsorozatról volt szó, mégis a formánsértékek

lokalizációs tendenciái további vizsgálatokat inspiráló kísérleti eredmények.



3. ábra

Az [u:] hang koartikulációs mezője CVC hangkapcsolatok alapján, férfi ejtésben

A mássalhangzók által okozott formánsérték-változások tendenciája annál szembetűnőbb, minél elnyújtottabb a koartikulációs mező az F1 vagy az F2 irányában (vö. 1., 2. ábra).

A mindkét tengely irányában közel egyforma kiterjedésű mező esetében ([u:] hang) találunk legkevésbé körülírható mássalhangzó-csoportosulásokat.

Az egyes mássalhangzók formánsérték-befolyásoló hatása jellemzően mutatkozik meg az [a:] és az [i] hang esetében. Az egyes formáns csoportok kiterjedésének iránya e két hang esetében követi a magánhangzó teljes koartikulációs mezőjének kiterjedési alakját és irányát. Az [a:] hang esetében horizontális, míg az [i] hang esetében vertikális irányú eltolódást tapasztalunk.

Legjellegzetesebben az [i] koartikulációs mezőjében tömörülnek csoportokba az egyes mássalhangzók által befolyásolt formánsértékek. Minden vizsgált magánhangzó koartikulációs mezőjére jellemző, hogy a sziszegő hangok hatására változnak legkevésbé a formánsértékek az átlagadatokhoz viszonyítva; a [s, z, t̂s] hangok által meghatározott formánsértékek a mezők centrumában lokalizálódnak.

A zöngés-zöngételen mássalhangzópárok összevetése során azt a megállapítást tehetjük, hogy a mássalhangzópárok egymás közelében helyezkednek el, a zöngés mássalhangzók azonban minden esetben nagyobb területet foglalnak el, mint zöngétlen párjaik.

Jellegzetes kép, hogy a [n] hang alakította formánsértékek szóródási területei a legnagyobbak, mind az [a:], mind pedig az [i] hang esetében. Míg az [i] esetében az F2 tengely mentén találunk több, mint 100 Hz-es eltérést (2260-2380 Hz) a formánsértékekben, az [a:] hang esetében ugyanez a szórás (670 Hz - 790 Hz) az F1 tengely mentén jelentkezik.

### Összegzés

A szakirodalmi mérési eredményeknek jelen kísérleti eredményeinkkel való összevetése azt mutatja, hogy az általunk meghatározott koartikulációs mezők a főbb területeken fedik az eddig megállapított formánstartományokat. A vizsgálat azonban bebizonyította, hogy a formánsértékek koartikulációs mezőkben történő ábrázolása, elemzése jóval több információval szolgál a magánhangzók frekvencia-paramétereiről, mint a kizárólag szélsőértékek által definiált frekvenciatartományok.

Az ábrázolási módszer lehetőséget teremt a különböző mássalhangzó-környezetben eltérő akusztikai sajátságokat mutató magánhangzók frekvenciaértékek összevetésére, valamint az F1, az F2 – és háromdimenziós ábrázolásmódban – az F3 viszonyrendszerének elemzésére.

Jelen vizsgálatunk első fázisa egy átfogó vizsgálat-sorozatnak, amelynek célja, hogy – a más nyelvek magánhangzóinak már meghatározott koartikulációs mezőrendszeré után (pl.: Peterson–Barney 1952) – definiálja a teljes magyar magánhangzó-állomány koartikulációs mezőrendszerét.

A vizsgálat-sorozat nemcsak az eddig a magyar szakirodalomból hiányzó, teljes mezőrendszer leírását pótolja – lehetőség teremtve a magyar magánhangzóknak a különböző nyelvek magánhangzóival való összevetésére –, hanem a magánhangzók formánsértékei közötti "átfedések" kijelölésével további percepciók kísérletek kiindulási alapjául szolgálhat.

#### Irodalom

Bolla Kálmán: Magyar hangalbum. Magyar Fonetikai Füzetek 6. Budapest 1980.

Flanagan, J. L.: Speech Analysis, Synthesis and Perception. Berlin, Heidelberg, New York 1965.

Gósy, M.: Funny language: Mother tongue awareness of a Hungarian child. Eurasian Studies Yearbook 67. 1995, 127-139.

Lindblom, B.E.F.: Spectrographic study of vowel reduction. JASA 35. 1963, 1773-1781.

Magdics Klára: A magyar beszédhangok akusztikai szerkezete. Nyelvtudományi Értekezések 49. Akadémiai Kiadó. Budapest 1965.

Molnár József: A magyar beszédhangok atlasza. Tankönyvkiadó. Budapest 1973.

Olaszy Gábor: Elektronikus beszédelőállítás. Műszaki Könyvkiadó. Budapest 1989.

Peterson, G.–Barney, H.: Control methods used in a study of the vowels. JASA 24. 1952, 175-184.

Tarnóczy Tamás: A magyar magánhangzók akusztikai szerkezete. Kir. Magy. Pázmány Péter Tudományegyetem Ált. Nyelvészeti és Fonetikai Intézete. Budapest 1941.

Tarnóczy, T.: Acoustic Analysis of Hungarian Vowels. Quaterly Progress and Status Report, Januar 1965. Speech Transmission Laboratory, KTH. Stockholm 1965, 8-12.

## A NYELVI KOMPETENCIÁRÓL

Bahczerowski Janusz

ELTE, Lengyel Filológiai Tanszék, Budapest

Előadásom a következő kérdéseket tárgyalja: (1) a vizuális és az auditív információ feldolgozása, osztályozása és tárolása; (2) a nyelvi kommunikációs modell funkcionálása; (3) nyelvi kompetencia; (4) nyelvi pragmatikai kompetencia; (5) metanyelvi kompetencia; (6) metanyelvi pragmatikai kompetencia.

Az objektív valóság bizonyos energiamennyiség leadása árán információt szolgáltat az objektív rendszerekről és viselkedési módjaikról. Ennek eredményeképpen az emberi tudatban kialakul a valóság modellje. Az információszerzés és az információközlés szempontjából különösen két analizátor játszik fontos szerepet, nevezetesen a vizuális (látó) és az auditív (halló) rendszer, amely egy meghatározott típusú információ felvételére képes, azaz a meghatározott hullámhosszra reagál. Az emberi agy a felvett információ legcélszerűbb és leggazdaságosabb feldolgozására törekszik, amelynek eredménye az osztályozás. Az osztályozás feltehetőleg két matematikai összefüggésen alapul:

- (1) az inklúzió (halmazelméletből), valamint
- (2) a feltételes valószínűségen.

Ez a két összefüggés az osztályozás hierarchikus struktúrájára utal. Meg kell azonban jegyezni, hogy az objektumok klaszifikációs folyamata és annak hierarchikus struktúrája az osztályozandó objektumok közös tulajdonságainak köszönhető. A klaszifikációs mechanizmus többlépcsős jellegű, és különböző absztrakciósinteket tételez fel. Minél nagyobb az elvonatkoztatás, annál magasabb az absztrakció foka. A vizuális és az auditív analizátorok egymással kapcsolatban vannak, és az idegrendszer bemeneteit képezik, amelyek a kimenethez kapcsolódnak, mégpedig az artikulációs és a grafémikus rendszerhez. Ha az in-

formációt vizuális rendszer dolgozza fel, akkor vizuéma és optikai kép (vagy akusztikai kép) jön létre. Viszont az auditív rendszer információ-feldolgozásakor audéma és akusztikai kép (vagy optikai kép) keletkezik. Ily módon vizuális és auditív invariánsok, tehát vizuémák és audémák jönnek létre. (Az invariáns fogalma azonos az osztály fogalmával.) A vizuémák, illetve az audémák a valóság egy osztályához tartozó objektumok modelljei; képzésük csak a releváns, azaz csak az objektumok egész osztályára jellemző tulajdonságok figyelembe vételével történik. Tehát a vizuális és az auditív invariánsok alapján az ember képes megalkotni az adott osztályhoz tartozó objektumok optikai ill. akusztikai modelljét. Hozzá kell tennünk, hogy ahhoz, hogy az ember emlékezetében az adott objektumnak nyoma maradjon, elengedhetelen bizonyos információs minimum. Egyébként egyáltalán nem ritkák az olyan esetek, amikor bizonyos invariánsoknak nincsenek reprezentánsai a valóságban. Az ilyen invariánsok a logikai osztályozás eredményeképpen jönnek létre, és az esetek többségében a legmagasabb absztrakciós szintű osztályokhoz tartoznak. Ezek egyébként létrejöhetnek a valós tárgyak modelljei alapján is. Érdeemes hozzátenni, hogy a logikai osztályozás másodlagos a vizuális és az auditív osztályozáshoz képest. A fentiek tükrében azt mondhatjuk, hogy az ember információs univerzumát (belső tárolóját) a vizuémák, audémák és a logikai klasszifikáció következtében létrejött osztályok képezik. A nyelvi kommunikáció folyamatában a megfelelő invariáns kiválasztása és a megfelelő grafémák, illetve artikulémák hozzárendelése (kódolás) megy végbe. Ennek eredményeképpen a reprezentáció síkján megjelennek a vizuális, illetve az auditív jelek. A vizuális jelek a valóságos objektumoknak, vizuémáknak és grafémáknak az optikai modelljei. Az auditív jelek viszont a valóságos objektumoknak, az audémáknak és az artikulémáknak az akusztikai modelljei. A nyelvi információ befogadásakor a megfelelő audémák, vizuémák, magasabb szintű invariánsok,

artikulémák és grafémák azonosítása történik. Tehát a képfelismerés a mi értelmezésünkben felöleli az analízatornak megfelelő információk befogadásának folyamatát, a kapott információk besorolását a nekik megfelelő, különböző absztrakciós szintű osztályokba és a besorolás kontrollját, ami áll

- (1) az információ befogadásából,
- (2) az információnak a meghatározott osztályba való besorolásából,
- (3) a besorolás kontrolljából.

A két utolsó művelet egymástól elválaszthatatlan és egyenként nem alkalmazható. Eközben két törvényszerűség érvényesül:

(1) Minél nagyobb a kapott információ mennyisége, azaz a kép komponenseinek kontrasztfoka szemmel láthatóbb (az ilyen rendszer nagyobb mértékben rendezett, redundanciája 1-hez tart –  $R > 1$ ), akkor a kontrollinformáció a minimumra korlátozódik. Ez a minimum belső artikuláció impulzusainak erős redukációjában jut kifejezésre. Ez az eset a felismerésnek olyan folyamatához tartozik, amelynek jellemzője a nagyobb automatizálás. Ez azt jelenti, hogy a kép által közvetített információ valószínűsége kölcsönösen összefügg az adott invariánssal. Az entrópia ilyen esetekben nullával egyenlő. Az ilyen típusú folyamatok alig kívánnak kontrollt és majdnem függetlenek az időtől (automatikus egybevetés).

(2) Ha a kép bonyolult, akkor közöttük nem található élesen elhatárolt belső függőség, vagyis olyan függőség, mint az előző esetekben, mivel a képben foglalt információ mennyisége  $H_{max}$  irányában növekszik. A kép és egy meghatározott osztály egybevetésének folyamata kevésbé automatikus, ezért nagyobb hatásfokú kontrollinformáció felhasználását teszi szükségessé. Ez abban nyilvánul meg, hogy a belső artikulációnak erőteljesebbek az impulzusai. Ez a képfelismerés különbözik az előzőtől és időfüggő. A belső beszéd figyelemmel kísérhető (ellenőrizhető) nemcsak a szóbeli megértetésnél, hanem a szemléletes gondolkodásnál

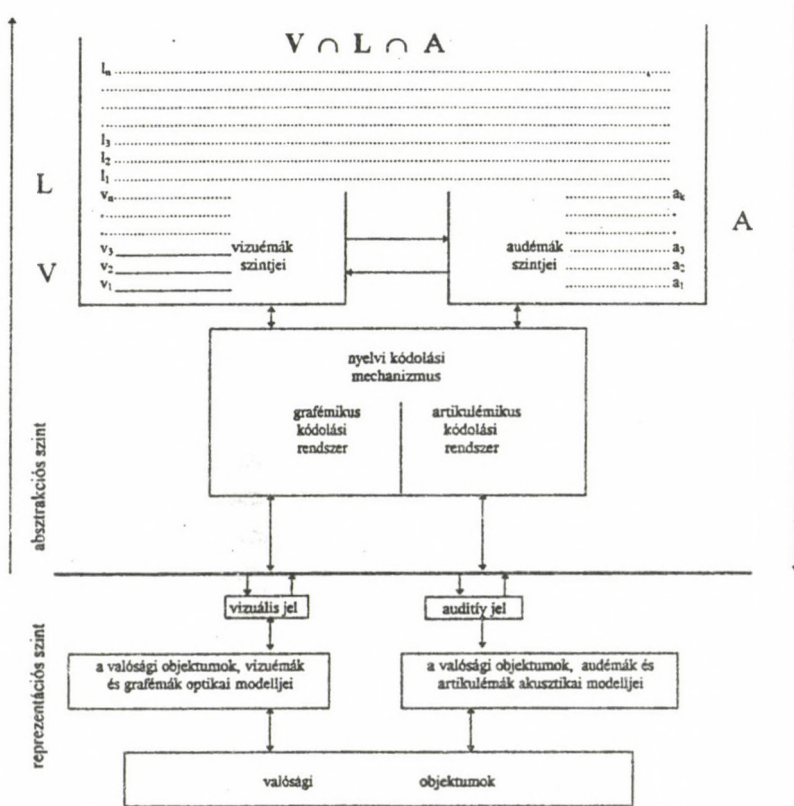
is, valamint a szóbeli és a szemléltető anyag érzékelésének, emlékezetbe vésésének és felidézésének folyamatánál. Az artikuláció (mind a belső, mind pedig a külső) mint kontrollinformáció láthatóan különleges szerepet játszik az emberi gondolkodás folyamatában.

A tanulási (információszerzési) és a kommunikációs folyamat valószínű modellje a következőképpen fest. (Lásd a köv. oldalon). Valamilyen invariáns kiválasztása, amely az absztrakció meghatározott szintjén lép fel, a fentebb említett bemenetek egyikének vagy másikának alapján vagy akár mindkettően egyidejűleg realizálható. Modellünk egyik rész-szisztémájának aktivizációja maga után vonja a modell minden rész-szisztémájának aktivizálódását. Az artikuláció (belső és külső) minden egyes esetben kontroll információként jelentkezik. Ha az információt az auditív rendszer fogadta be, akkor feldolgozása a következő azonosítási műveletek formájában történik:

- (a) az információ és a megfelelő audémák azonosítása,
- (b) az információ és a megfelelő magasabb szintű invariánsok azonosítása,
- (c) az információ és a megfelelő vizuémák azonosítása,
- (d) az információ és a megfelelő artikulémák azonosítása,
- (e) az információ és a megfelelő grafémák azonosítása.

Mindezek az azonosítási műveletek a genetikailag adott feltételes reflexeken alapulnak. Csaknem egyidejűleg működnek akkor, amikor a kölcsönös valószínűség 1-gyel azonos.

Ha a vizuális rendszer fogja fel az információt, akkor az előbb felsorolt műveletek analóg módon folynak le. Az információt ekkor a megfelelő (a) vizuémákkal, (b) magasabb szintű osztályokkal, (c) artikulémákkal, (d) audémákkal, (e) grafémákkal azonosítjuk. Az általunk elemzett modell funkcionálásának alapja az egyes osztályok azonosítási lehetőségében rejlik. Ez azt jelenti, hogy az audémák bizonyos láncszemei összefüggésben vannak az artikulémák, vizuémák és magasabb szintű osztályok bizonyos

**Jelmagyarázat:**

L = absztrakció magasabb szintjei, amelyek a logikai klasszifikáción alapulnak (a logikai klasszifikáció viszonylag független az objektív valóságtól - alkotó gondolat)

$V \cap L \cap A = A \cap V, L, A$  metszete

$v_1, v_2, v_3, \dots, v_n$  = a vizuémák absztrakciós szintjei

$a_1, a_2, a_3, \dots, a_k$  = az audémák absztrakciós szintjei

$l_1, l_2, l_3, \dots, l_m$  = a logikai klasszifikáció szintjei

↑ = mélységirány, az absztrakció fokának növekedése (= információfeldolgozás, tanulás, információátvitel)

↓ = felszínirány (= kommunikációs folyamat)

láncszemeivel. Egy osztály (invariáns) kiválasztása jelentős mértékben függ egy másik osztály kiválasztásának valószínűségétől. Ennek a valószínűségnek a nagyságától függ, mint tudjuk, az egyes osztályoknak más osztályokkal való korrelációjának az ideje, ez a valószínűség határozza meg az azonosítás automatizálásának fokát.

Az artikulémák legnagyobb egységei a szóartikulémák. Belőlük képződnek a mondat- és szövegartikulémák, amelyek a szóartikulémák kiválasztásával és kapcsolódásával függenek össze.

Mindezek a vizsgáldások a következő hipotézisek felállítását teszik lehetővé:

(1) Az audémák és az artikulémák szintje között egyértelmű megfelelés van (a szóartikulémák és szóaudémák halmaza véges számú). A vizuémák és a magasabb szintű osztályok halmaza nagyobb halmaz a szóartikulémák és a szóaudémák halmazánál.

Ebből a hipotézisből következik az alábbi:

(2) A szóartikulémák és szóaudémák halmaza, valamint a vizuémák és magasabb szintű osztályok halmaza között nincs egyértelmű megfelelés. Más szóval, a szóartikulémák halmaza nem elegendő ahhoz, hogy világosan kifejezze a vizuémák szintjét, sem pedig az absztrakció magasabb osztályainak szintjét.

Az emberi agy megtalálta az optimálisan gazdaságos kiutat ezekből a nehézségekből. Ez következőképpen történik: egyes vizuémákhoz és magasabb szintű osztályokhoz a szóartikulémák és szóaudémák, másokhoz viszont a mondatok és a szövegek artikulémái és audémái kötődnek. A szóaudémák és szóartikulémák az információ nagyobb mennyiségét hordozzák, mint a mondat- és szövegartikulémák és audémák, mivel előbbieket az absztrakció magasabb szintű osztályaihoz, utóbbiak pedig az alacsonyabb szintűekhez tartoznak.

Nyelvészeti értekezésekben gyakran használják fel a paradigmatis és szintagmatikus síkok fogalmát. A mi modellünkben a paradigmatis tárolás nem csupán a szóaudémák

mennyiségére, hanem az absztrakció minden szintjére kiterjed. Szintagmatikus tárolással csak a reprezentáció szintjén, vagyis a kommunikáció és a tanulás folyamatában találkozunk. Az anyanyelv paradigmaticus tárolója a szintagmatikus tároló alapján keletkezik (szintagmatikus információ). Az anyanyelvi kommunikációs modellben az információ átadásánál és az információ befogadásánál átlépünk a paradigmaticus tárolóból a szintagmatikus tárolóba. Az anyanyelvi tanulás modelljében mélyítjük (fejlesztjük) a paradigmaticus tárolót, úgyszintén szélesítjük (bővítjük) a szintagmatikus tárolót. A paradigmaticus tároló mélyítése új (a) vizuémák, (b) magasabb szintű osztályok, (c) szóaudémák, (d) szóartikulémák keletkezése útján történik. Változatlanok maradnak a paradigmaticus tárolóban (a) az audémák szintjei, amelyeknek morfológiai, fonológiai és fonetikai egységek (ha ilyenek valóban léteznek) felelnek meg, (b) az audémáknak megfelelő artikulémák szintjei, (c) a paradigmaticus tárolóból a szintagmatikus tárolóba történő (és fordítva) átmenet szabályai.

A természetes nyelv reálisan csak a konkrét emberi lények tulajdonságaként létezhet. A nyelvi megnyilatkozások nem mások, mint a nyelvi mechanizmus kimenetén megjelenő, végső, kész produktumok, más "termelési" rendszerekhez hasonlóan. Tehát, ahogy már fentebb említettük a nyelven egy bizonyos kódolási rendszert, kódolási mechanizmust értünk, amelynek bemenetét az ember belső információs tárolója, információs univerzuma, azaz "mentális világa" alkotja. Ezt a mechanizmust mindenekelőtt egy bonyolult szabályrendszer képezi, amelynek az a feladata, hogy az információs univerzum elemeihez a nyelvi kódolási ábécé elemeit hozzárendelje. Ennek eredményeképpen olyan idegszignálsorok (neurémák) jönnek létre, amelyek az artikulációs (illetve grafémikus) rendszer bemenetén megjelenve a kimenetén átkódolás következtében fonikus (illetve grafikus) sorok formájában jelenítődnek meg. Az ember nemcsak továbbítani, percipiálni, feldolgozni és megérteni képes a nyelvi közléseket, hanem ezeket

használni is tudja. Ahogy már említettük, a valóságban csak az egyes személyek nyelvei, azaz az ún. idiolektusok léteznek. A valóságban soha nem találkozunk két egyforma idiolektussal. Természetesen az egyes idiolektusok kisebb-nagyobb mértékben hasonlóak egymáshoz, ami azt jelenti, hogy felfedezhetünk bennük olyan közös részt, amely az adott nyelvi közösséghez tartozó egyének minden idiolektusában állandóan ismétlődik. Így a nyelvet mint társadalmi, azaz egyén fölötti jelenséget kétféle módon interpretálhatjuk: (1) az összes idiolektus halmazaként, valamint (2) egy bizonyos absztrakciós lingvisztikai konstruktként. Az utóbbi értelmezés egy önálló absztrakciós létet feltételez, amely az idiolektusokhoz képest elsődleges. Ilyenfajta megállapításokat gyakran összefüggésbe hoznak a Saussure-féle "langue", illetve a Chomsky-féle "ideális beszélő-hallgató" kategóriákkal. A nyelv Chomsky-féle, egy bizonyos tudásfajtaként történő értelmezésének arra kell emlékeztetnie bennünket, hogy a tudás fogalma egyidejűleg a tudás hordozóját is feltételezi, ami azt jelenti, hogy a tudás nem más, mint valakinek a mindenkori tudása. Az egyén fölötti értelemben vett tudás viszont egy tetszőleges gyűjtő (kollektív) alany tudását jelenti.

A természetes nyelv kommunikációs eszközként való értelmezése több szempontból is helytelennek tűnik. Többek között azért is, mert a nyelv egyben a világ átélésének, kategorizálásának, a róla való gondolkodásnak stb. eszköze is. A fenti reflexiók tükrében a nyelv nem más, mint kizárólag a megnyilatkozások (szövegek) generálását szolgáló eszköz, és csak az általa létrehozott közlések képesek betölteni a kommunikációs eszköz funkcióit ugyanúgy, ahogy a gépjárműgyárat sem lehet közlekedési eszköznek tekinteni, csak az általa gyártott járműveket.

Azokból a megfigyelésekből, amelyek szerint nem mindenki képes azonos szinten használni a nyelvet, azt a következtetést vonhatjuk le, hogy a nyelv és a nyelvhasználat képessége

különböző dolgok. A nyelvhasználat lényege a megnyilatkozások generálásában rejlik, viszont a megnyilatkozások funkcionális használatának a képessége már a pragmatika területét érinti. Különböség van tehát: (1) a nyelv mint egy bizonyos, a nyelvi mechanizmusra (szabályrendszerre) vonatkozó tudásfajta, és a róla szóló tudás között, (2) a nyelv használatának képessége és a róla szóló tudás között, (3) a nyelvi megnyilatkozások és a róluk szóló tudás között, valamint (4) a nyelvi megnyilatkozások használatának képessége és a róla szóló tudás között. A itt felsoroltakról természetesen minden nyelvhasználó rendelkezik egy bizonyos tudással, bár ez a tudás tudományosan nem okvetlenül megindokolt.

Mint ismeretes a nyelvi kompetencia fogalmat a nyelvészetbe N. Chomsky vezette be. Felfogása szerint a nyelvi kompetencia egy bizonyos tudásfajta, amely az ideális beszélő-hallgatót jellemzi. Ez a definíció azonban, úgy vélem, több vonatkozásban is pontosításra szorul. Nem tudhatjuk például azt, hogy milyen konkrét tudásról van szó: a nyelvről, mint a megnyilatkozások (mondatok) generálását szolgáló szabályrendszerrel, e szabályok használatának képességéről, vagy a megnyilatkozásokról, illetve a megnyilatkozások használatának képességéről szóló tudásról. Chomsky inkább a mondatokat generáló szabályrendszerre és nem a megnyilatkozások gyakorlati használatának szabályaira gondolt.

Bármilyen kompetenciáról, így a nyelviről is csak akkor beszélhetünk értelmesen, ha azt arra az alanyra vonatkoztatjuk, amely ennek a kompetenciának birtokában van. Az egyéni alanyok nyelvi kompetenciájának esetében a probléma sokkal egyszerűbb, mint a kollektív nyelvi kompetencia fogalmánál. Felmerülhet a kérdés, hogyan lehet meghatározni a gyűjtőalanyt az azt a nyelvi kompetenciáját, amely az egyéni alanyok halmazára vonatkozik. Chomsky ezt a kérdést oly módon oldotta meg, hogy bevezette "az ideális beszélő-hallgató" fogalmát és az "ideális beszélő-hallgató nyelvi kompetenciáját" is. A nyelv valódi használóinak nyelvi kompetenciája nagyon differenciált, és változó képet mutat. A

nyelvi kompetencia változhat mind mennyiségi, mind pedig minőségi tekintetben. Egyetlen egy nyelvhasználó sem rendelkezik teljes nyelvi kompetenciával, nem is beszélve az ideálisról. Chomsky ideális beszélő-hallgatója egy tökéletes automatával egyenlő, amely kizárólag csak a generatív szintaktikai műveleteket hajtja végre. Világosan kell látnunk viszont, hogy az ideális beszélő-hallgató csak ideális nyelvet használhat, és csak ideális mondatokat generálhat. Chomsky a nyelvi kompetenciát a nyelvészeti kutatások középpontjába helyezte, és ennek következtében a mai nyelvészet súlypontja áthelyeződött a konkrét megnyilatkozások, idiolektusok tanulmányozásáról a spekulatív-formális lingvisztika irányába. A Chomsky-féle nyelvi kompetencia további fejlődése viszont a fokozatos konkretizálás irányában folytatódott. A kompetencia kutatási tárgyát folyamatosan kibővítették a valódi beszélő-hallgató, a valódi megnyilatkozások és a belső, valamint külső valódi kontextusokban való használatuk képességének tanulmányozásával. Ennek eredményeképpen a nyelvtudomány, amely eddig tulajdonképpen csak a nyelvvel foglalkozott, olyan tudománnyá vált, amelynek tárgya az emberi kommunikáció sok más aspektusával is jelentősen bővült. Itt például a nyelvi műveletek intencionális-teleologikus aspektusaira, a nyelvi kommunikáció kognitív tényezőire, kultúrdeterminánsaira stb. gondolok. Így a hangsúly az egész emberi kommunikációs tevékenységre és a bármely kontextusban megjelenő megnyilatkozások valamennyi fajtájára helyeződött át. Erről legjobban Hymes és mások munkái tanúskodnak.

A kommunikációs kompetencia komponensét alkotó nyelv tudásához hasonlóan, a nyelvhasználat képessége mind minőségi, mind pedig a mennyiségi aspektus tekintetében is különböző lehet. Tehát meg kell világosan különböztetnünk, és el kell különítenünk egymástól a szűkebb értelemben vett nyelvhasználat képességét és a nyelvi megnyilatkozások használatának képességét. Eddig nem tettek különbséget e két fogalom között. Ez a megkülönböztetés

azért is szükséges, mert, ahogy már fentebb szó volt róla, a nyelvi megnyilatkozás a nyelv mint kódolási mechanizmus működésének eredményeképpen létrejött végső produktum. A nyelvhasználat képességét a megfelelő nyelvi alrendszerekhez kellene viszonyítani. Így például beszélhetnénk egyrészt az artikulációs, fonetikai, szóképzési, lexikális stb. képességről, másrészt pedig a beszédre, az információ-továbbítási, interpretációs, perszváziós stb. aspektusára vonatkozó képességről is. A fentieket összegezve a nyelvi tudást és használatának a képességét nyelvi kompetenciának, viszont a kommunikatív céllal generált nyelvi megnyilatkozások használati szabályrendszerének tudását (ismeretét) nyelvi pragmatikai kompetenciának fogjuk nevezni.

A nyelvi közlések (szövegek) által továbbított információkat elemezve láthatjuk, hogy azok a tárgyról szóló információn kívül olyan kifejezéseket is tartalmaznak, amelyeknek célja, hogy a nyelvi információátadás organizációját biztosítsák, vagy az egyes szövegfragmentumokra vonatkozó állításokat tartalmazzanak azokról, mint egy sajátos tárgyról. Tehát azt is mondhatjuk, hogy ezek a kifejezések metanyelvi (metainformációs) operátorok szerepét töltik be. Ezek az elemek metainformációt szolgáltatnak például a beszédaktusról, a kommunikációs csatorna fajtájáról, a megnyilatkozás fragmentumai közötti kapcsolatokról, a szövegen belüli ekvivalenciáról, ill. kváziekvivalenciáról, a dialógusban betölthetik a fatikus funkciót, szignáljai lehetnek a kontaktus elkezdésének, fenntartásának, a végéhez való közeledésének vagy megszakításának, a szövegbe bizonyos magyarázatokat vezethetnek be, hierarchizálhatják a szöveget az érdekes, ill. kevésbé érdekes, fontos, ill. kevésbé fontos, emóciókat ébresztő, ill. emóciókat nem ébresztő stb. információk tekintetében is, szignáljai lehetnek a beszélt megnyilatkozások beszédközbeni korrekciójának stb. Ezen operátoroknak az elemzése olyan következtetésre jogsít fel, hogy ezek a struktúrák a szöveg síkján

viszonylag különálló, autonóm és nyelvspecifikus metainformációs (metanyelvi) lexikont alkotnak. A metainformációs operátorok egy bizonyos metaszöveget alkotnak, amely ráakódik az alapszövegre, és vele együtt egy teljes kommunikációs egységet képez. Ebből az következik, hogy a nyelvi kód (nyelv) létezése implikálja a metanyelvi kód (metanyelv) létezését is. Ennek tükrében a nyelvi és a nyelvi pragmatikai kompetencia mintájára megkülönböztethetjük a metanyelvi és a metanyelvi pragmatikai kompetenciát is. A metanyelvi rendszer a nyelvi rendszerhez hasonlóan reálisan csak az idiometalektusok formájában létezhet. A metanyelvi kompetencia hiányát tapasztaljuk az idegen nyelveket beszélő egyéneknél, és ez végsősoron komolyan megnehezítheti az információ cseréjét, és több problémát okozhat az interlokutorok közötti kommunikációban.

### Irodalom

Bańczerowski Janusz: A nyelvi kommunikáció és az információ néhány kérdése. Nyelvtud. Ért. 99. Akadémiai Kiadó, Budapest 1979.

Bańczerowski Janusz: Metainformációs struktúrák a nyelvi szöveg síkján. Magyar Nyelv XC, 1994, 30-41.

Bańczerowski Janusz: A kommunikációs kompetencia és összetevői. Magyar Nyelvőr, 118/3. 1994, 277-286.

Chomsky, N.: Aspects of the Theory of Syntax. The M.I.T Press, Cambridge, Mass. 1965.

Chomsky, N.: Language and Mind. Harcourt, Brace World, New York 1968.

# A DIGITALIZÁLT BESZÉD MINŐSÉGE

## A nemzetközi ajánlások változása

Lajtha György

Magyar Távközlési Részvénytársaság, Budapest

A beszédátvitel és -kezelés az utolsó évtizedben döntő mértékben digitális formában történik. A digitális átvitel során az átviteli út zaja, lineáris és nonlinearis torzítása nincs közvetlenül hatással a minőségre. Ugyancsak megváltozott a beszédátvitel, -feldolgozás, és -kezelés módszere, mert a digitális technikát alkalmazva az információ az eszközök hibáira kevésbé érzékeny. A kezelés során viszont felléphetnek érthetőséget rontó hatások. A digitalizálás hatására a beszédminőségre vonatkozó nemzetközi ajánlások is megváltoztak és jelenleg is átalakulóban vannak. A cikkben a digitális jelkezelés hatása és szubjektív beszéd értékelés kapcsolatát keressük.

### 1. Elsődleges jellemzők

A beszédet döntő mértékben jelenleg is emberek közötti kapcsolat céljára használjuk. Bár előfordul, hogy gép érzékeli a beszédet, vagy gép beszél, ezen esetekben is a szubjektív ítélet mérvadó. Tehát, eszköztől függetlenül a véleményvizsgálat vagy ezzel egyenértékű beszélgetési kísérletek határozzák meg elsősorban az átvitel és a beszédkezelés minőségromtó hatását. Ugyancsak változatlan, hogy a szubjektív vélemények statisztikus középértékét a *mean opinion score*-t csak igen sok kísérlettel lehet meghatározni. Ezért a minősítésre előnyösebb az *érthetőség* és a *felismerhetőség*, amely pontosabb, számszerű értékkel jellemzi a felhasznált eszközöket. Bár, a logatom érthetőség véges számú kísérlettel, és néhány személlyel egy napon belül meghatározható és hasonlóképpen a felismerhetőségi vizsgálatok is elfogadható időn belül eredményre vezetnek, üzemszerű vizsgálatokra mégis nehezen használhatók.

Az analóg technikában, a szubjektív sztochasztikus vizsgálatokon alapuló követelmények – mint ahogyan arról az 1989.

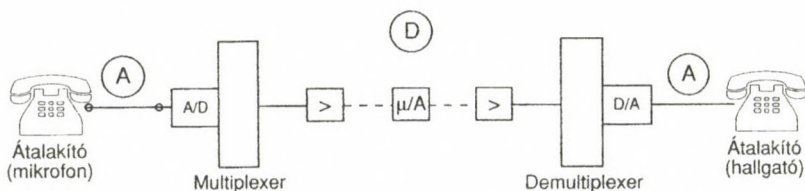
évi Beszéd kutatás Konferencián beszámoltam – Fletcher kutatásaira alapozva stacionárius jellemzőkre vezethetők vissza. Ilyenek a zaj, az amplitúdó, a fázis, és a frekvencia-torzítás, valamint a non-lineáris torzítások. Ezek a jellemzők az átvitel és beszédkezelés során alkalmazott eszközök csillapításának, zajának, csillapítás-fázis-frekvencia karakterisztikáinak mérésével meghatározhatók. Diagramok segítségével a mérési eredmények is az érthetőségi index összerendelhető.

## 2. Digitális jelkezelés

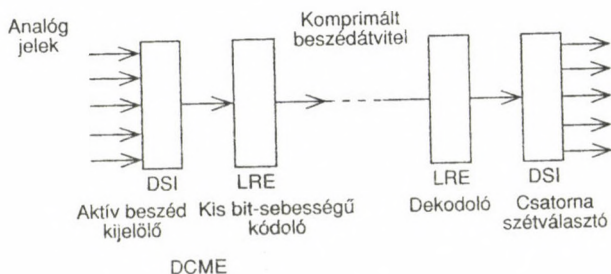
Az 1. ábrán bemutatott általános beszédátviteli és beszédkezelő rendszerek bemenetén analóg/digitális átalakító helyezkedik el. Ennek hatására a vonalon az elektromos jelek nem közvetlenül arányosak a hangnyomással, tehát az elektromos jel torzulásából a Fletcher-féle módszerrel nem lehet következtetni a vett beszéd minőségére. A minőséget befolyásoló tényezőket részben az 1/a ábrán látható távközlési összeköttetés, 1/b ábrán látható beszédkompresziós rendszer, valamint az 1/c ábrán látható tárolt és visszaadott beszédre vonatkozó egységek vizsgálatából igyekszünk meghatározni.

A bemutatott példákön az elektroakusztikus átalakító és az A/D konverter között jelöltünk egy analóg szakaszt. Ez általában rövid és csak különleges esetekben van befolyással a minőségre. Előfordulhat azonban, hogy ezen kéthuzalos szakaszon reflexiók miatt jelentős jel-visszafordulással kell számolnunk, ami áthallás, önhang vagy visszhang formájában zavaró lehet. Amennyiben a kéthuzalos szakasz csillapítása is számottevő, akkor az A/D konverterre érkező szint eltérhet az optimálistól, ami a kvantálási torzítás megnövekedését eredményezi. Vételi oldalon a dekódoló és D/A konverter határozza meg a minőséget.

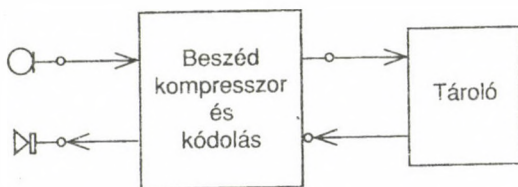
A digitális szakaszok minőségét a mintavételezési hiba, a kvantálási torzítás, a vonali bithibaarány, a terjedési idő és a reflexiók befolyásolják. A terjedési idő részben a nagy távolságú,



1/a ábra



1/b ábra



1/c ábra

szatellit áramkör alkalmazásából, részben az esetleg a beszédkezelésre, a redundancia csökkentésre és bitsebesség korlátozásra használt eszközök működési idejéből vezethető le. Ugyanakkor az áramkörök és tárolók jobb kihasználását biztosító redundancia csökkentő (LBR) és beszédszünet hasznosító (DSI) rendszerek kezdőhang veszteséget vagy torlódásból adódó digitális bitsomag veszteséget eredményezhetnek, amely véletlenszerű minőségcsökkenést okoz. A következőkben az ITU-T

vizsgálatai alapján a minőségromlás szubjektív hatást befolyásoló tényezőket tekintjük át.

### 3. A jelkezelés minőségrontó hatásai

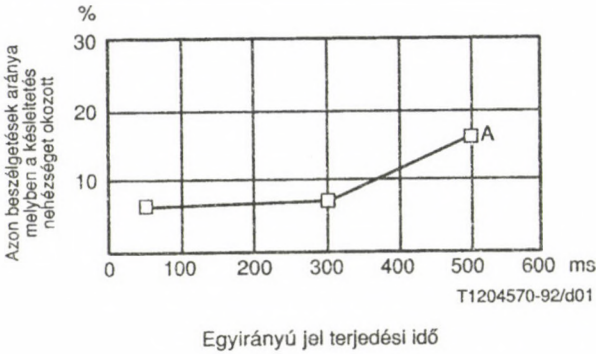
A jelkezelés célja lehet az átvitel során a

- *csatorna többszörös kihasználása digitális módszerekkel (DCME)*
- *csomagkapcsolt beszédátvitel*
- *beszédszünetek kihasználása más digitális információval (DSI)*
- *kis bitsebességű átvitel (LBRC)*
- *változó bitsebességű átvitel (DLC), (VBR)*

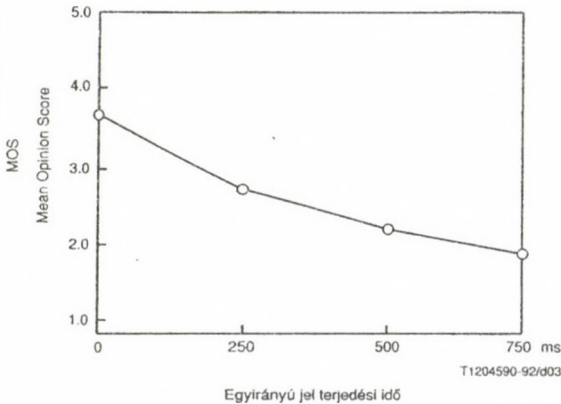
Ezen eljárások figyelik a beszédben rejlő redundanciákat az egymást követő minták hasonlóságát és ennek segítségével csökkentik az átviteli sebességet. Felhasználhatják a beszédszint figyelést is és ezzel észlelik a beszéd szüneteket, melynek helyén más beszélgetés vagy más információ kerülhet a vonalra. Lehet különböző könyvtárak és táblázatok segítségével a legvalószínűbb kódszavakat rövidebbel helyettesíteni. Új kódolási eljárások, mint például adaptív differenciál (ADPCM) vagy vektor kvantálás is segíthet az információhoz rendelő bitek csökkentésében.

Figyelembe kell venni, hogy a minőség megítélése függ az átvitt információ jellegétől, mennyiségétől és ezen keresztül a torlódásoktól, ezért nem vizsgálható a minőség kizárólag objektív eszközökkel. Más szóval ez azt jelenti, hogy sztochasztikus folyamatok által keltett zavarok, sztochasztikus összegezése adja az eredő minőséget. Amennyiben egy összeköttetésben analóg és digitális szakaszok is vannak, akkor ezek összegezése már csak a *MOS (Mean Opinion Score)* alapján lehetséges. Analóg összeköttetésben a beszédszünetekben lévő zajok és áthallások a legzavaróbbak, ugyanakkor a digitális összeköttetésekben a beszéd és a beszédszint szoros kapcsolatban van a minőségrontó tényezőkkel.

A jelkezelés minőségrontó hatásai közül ki kell emelni a visszhangból és a késleltetési időből származó zavarokat. Ezek ugyanis analóg rendszereknél is előfordulnak, azonban a nagyobb szakasz csillapítások és a kisebb késleltetési idők miatt nem játszanak jelentős szerepet. A reflexió és a késleltetési idő minőségrontó hatását a 2. ábra mutatja. Az a) ábra a késleltetés közvetlen hatását veti össze a beszélgetés nehézségeivel, a b) ábra a MOS és az egyirányú terjedési idő kapcsolatát mutatja.

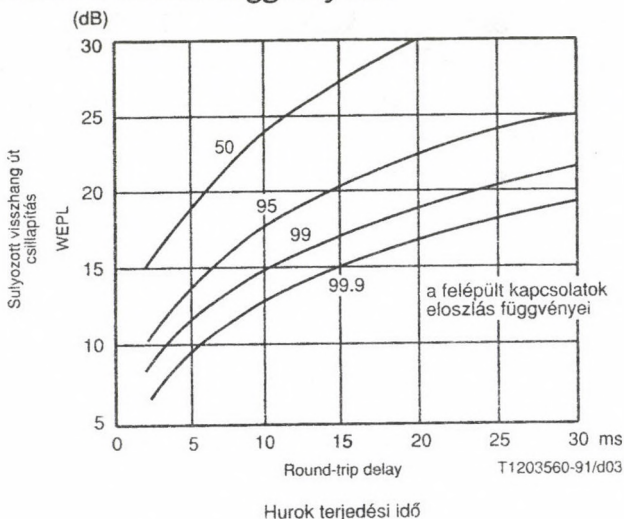


2/a ábra



2/b ábra

Végül a c) ábrán az ITU-T a terjedési idő és a súlyozott visszhangút csillapítás kapcsolatát vizsgálja a felépült összeköttetések eloszlás függvényével.



2/c ábra

Itt látszik, hogy a digitális szakaszok kisebb csillapítása következtében vagy visszhangtörő, vagy visszahangzár kapcsolások szükségesek, mert már 20-30 ms késleltetés jelentős megértési zavarokat okoz. Ebből származó zavarok tehát analóg jellegűek és nem közvetlenül kapcsolódnak a jelkezelés hibáihoz.

#### 4. Ekvivalens analóg zaj

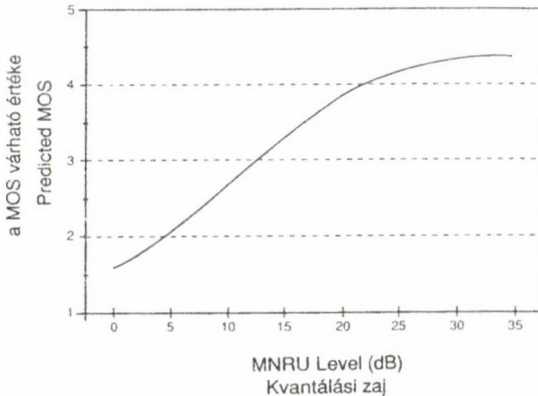
A korábbiakban említett elvi különbség, miszerint az analóg rendszerekben a zaj a beszédszünetekben a legzavaróbb. Ezzel szemben a digitális zaj a mintavételezés, valamint a kvantálás hibájából és a vonali bithibaarányból következik, vagyis nagysága közel arányos a hasznos jellel. Ezért az ITU-T bevezette a QDU-t, a kvantálási torzítás egységét, amely annyi zajt termel, mint egy ideális kódoló/dekódoló pár. Megadta, hogy különböző jelkezelő

egységek hány QDU zajt termelhetnek. Egy 8 bites kodek definíció szerint egy egységet termelhet, míg egy 7 bites kodek hármat és egy 16 kbit/s-os LD-CELP kodek már 3,5 egység. A Supplement G29 kísérleti tapasztalatokkal összeállított egy táblázatot, amely megadja az ugyanolyan szubjektív minőségrontást jelentő QDU és jel/zaj viszony (S/N dB) értékeket, mely alább látható.

m (qdu)	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
S/No (dB)	59,8	49,6	44,0	39,7	36,6	33,8	31,5	29,5	27,8	26,4	25,0	23,9	22,7	21,6	20,5

A QDU mérésére az ITU-T ajánl egy referencia műszert. Ennek megnevezése átlagos zaj mérőegység (MNRU), amely decibelben kalibrálva összehasonlító mérést tesz lehetővé. A zaj szintje ugyanis a beszéd átlag szintjéhez van viszonyítva.

A 3. ábrán bemutatjuk, hogy számos kísérlet alapján a relatív kvantálási zajszint és a MOS milyen kapcsolatban van. Az értékek 95%-os konfidencia szinttel érvényesek.



3. ábra

## 5. A jelkezelés vizsgálata

A jelkezelés alapja az analóg/digitális átalakítás során alkalmazott kódolók működési elve. A 64 kbit/s-os kodek 1 QDU torzítása az alapegység. Amennyiben az átviteli sávval való takarékoság vagy a tároló kapacitásának korlátjai miatt kisebb bitsebességet kell alkalmazni, akkor elsősorban a ADPCM jön számításba. Ezzel 32 és 16 kbit-es sebességgel körülbelül 3 QDU torzítással lehet digitális bitfolyamokat előállítani. Az erre vonatkozó műszaki ismertetés a G726 Ajánlásban található. 16 vagy 12 kbit sebességgel működő rendszereket LD-CELP (kis késleltetésű kódvezérelt lineáris predikció) eljárással lehet a legjobb minőségben előállítani (G728).\*

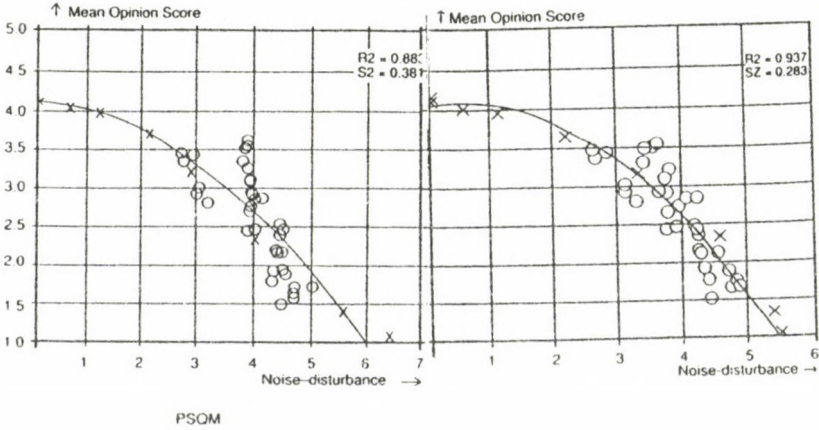
Az ITU munkája során több módszert hasonlítottak össze a szubjektív értékeléssel. Keresték azt a műszeres vizsgálati eljárást, mely a legszorosabb korrelációt adja a MOS-sal. Az alkalmazott vizsgálati eljárások a következők:

i) EPR (Expert Petten Recognition), vagyis szakértői jelalak vizsgálati módszer, melyre nézve a japán NTT végzett vizsgálatok, de kipróbálta a gyakorlatban a francia CNET, a német, az angol és az olasz távközlési vállalat is. Bár az eredmények, mint a 4/b ábrából láthatóan, mutattak korrelációt a szubjektív értékkel, mégis tovább fejlesztették.

A vizsgáló algoritmust az EPR-t, a Cepstrum távolsággal (CD), a koherencia függvényvel (CHF) és a LALOU által kidolgozott és a francia távközlés által előterjesztett információs indexszel (II) kombinálták. Az eredmények számos esetben már jól megközelítették a szubjektív ítéletet (4/d ábra).

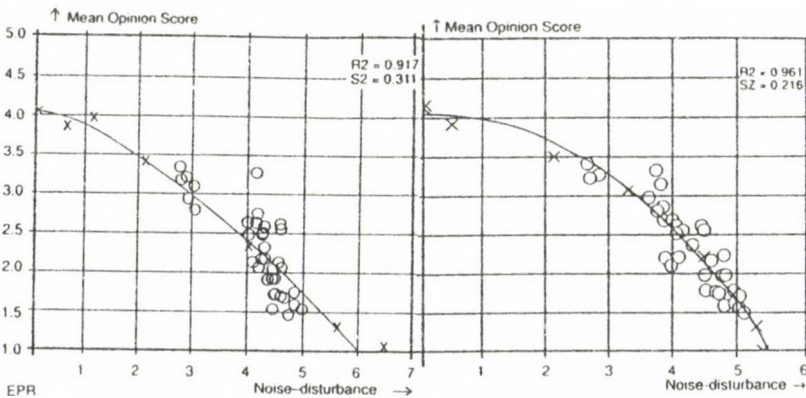
ii) Megfigyelhető (észlelhető) szubjektív minőségi mérték (PSQM), melynél a bemenet és kimenet különbségének összehasonlításán alapul a mérés. Ezt az angol és holland távközlési vállalatok tovább javították, annak érdekében, hogy a beszédből hiányzó részek és a torzításból adódó többletek ne kompenzálhassák egymást, külön vizsgálták a plusz és mínusz

hibákat. Mindkettőhöz meghatározták a PSQM értéket és ezek összegezéséből következtek a szubjektív ítéletekre. Az eredmények a 4/a és 4/c ábrán láthatók.



4/a ábra

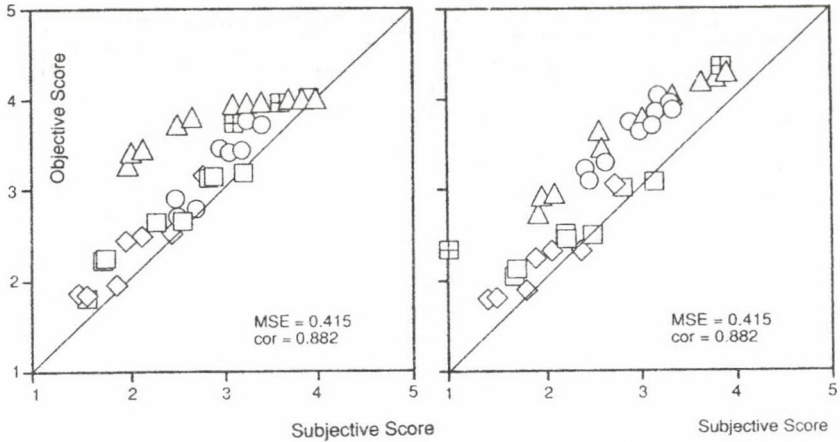
4/c ábra



4/b ábra

4/d ábra

iii) A kétféle rendszer összehasonlítása beszélőtől és rendszertől függő eredményt adott. A statisztikus összehasonlítás az 5. ábrán látható.



A szubjektív és az objektív értékelés kapcsolata az EPR módszer alkalmazásával

A szubjektív és az objektív értékelés kapcsolata a PSQM módszerek alkalmazásával

- ◇ CODEC b
- G.726
- △ G711

- ⊞ MNR & Direct
- CODEC a

5/a ábra

5/b ábra

## 6. Összefoglalás

A jelkezelés hibái csak a jelalakok precíz összehasonlítását elvégző és a különbségeket szakszerűen összegező algoritmus segítségével lehetséges. Szerepük jelentős, mert már fejlesztési eredmények vannak 4,8 kb/s és 2,4 kb/s digitális beszéd-átvitellel. Ugyannakkor a vizsgáló eljárások fejlesztése még folyamatban van. A különböző technikák előnyeit és hátrányait figyelembe vevő még további algoritmusok kidolgozása várható. Ennek ellenére már most megkezdődött a beszéd és a zene digitális átvitelét

minősítő műszerek kidolgozása. Az ebből adódó eredmény az analóg szakaszok torzításaival csak oly módon adható össze, hogy megadjuk mindkét esetben az elméleti felső határt, rögzítjük, hogy ennek hány százalékát használták el a vizsgált szakaszok és ezeket a százalékos értékeket összegezzük.

Mindezen vizsgálatok főként új kódoló, jelkezelő és ismétlő berendezések fejlesztésénél játszanak szerepet, mert a digitális hálózatban, ahol már az átkódolás és a kompressziós pontok száma minimális, ott az átvitel minősége a megengedett felső korlát 30%-a alatt marad. Ezért a digitális-analóg kombinált rendszereknél az analóg szakaszok vagy a digitális szakaszokon fellépő analóg jellegű minőségromlás a meghatározó.

\*A MOS és a QDU beszédátviteli jellemzők. Amennyiben más hírányag (kép, adat, szöveg) digitális kezelést kell minősíteni ezen jellemzők kritikával használandók.

### Irodalom

Fletcher, H.: Speech and Hearing in Communication. D van Nostrand, New York 1953.

ITU-T (CCITT) Ajánlások: (Genf 1993.) G113, G114, G122, G126, G726, G728, P11, P80, P81

ITU-T Speech Quality Expert Group Munkaanyagai 8.94, 15.94, 17.94, 18.94, 22.94, 25.94

Lajtha, G.: International recommendations on speech quality. In: Speech Research '89 Abstracts, Budapest 1989.

Richards, D. L.: Telecommunication by speech. The transmission performance of telephone networks. Butlerworth London 1973.

# A GÉPI BESZÉDDEL KIEGÉSZÍTETT SZÁMÍTÓGÉPEK ALKALMAZÁSI LEHETŐSÉGEI VAK ÉS CSÖKKENTLÁTÓ EMBEREK SZÁMÁRA

**Ecsedi Csaba**  
**PROMPT 92 KFT**

A vak és csökkentlátó magyar anyanyelvű tanulók, munkájuknál számítógépet használó szakemberek túlnyomó többsége IBM PC számítógépet használ DOS-os környezetben. Vakügyi felhasználásra három magyar fejlesztésű, magyar nyelvű beszédszintetizátor terjedt el, a PC Voice, Brailab PC és Multivox PC ROBOT. Szükséges lenne egy olyan céleszköz kifejlesztése, amely magyar szövegekörnyezetben a programok használata közben megjelenő angol nyelvű üzeneteket helyes angol kiejtéssel mondaná el. Új, felbecsülhetetlen értékű lehetőség a vak emberek számára az optikai karakterfelismerés, mely visszaadja nekik az olvasás örömét.

## 1. Bevezetés

Egy Angliában kiadott katalógus (Equipment, 1993) szerint mintegy 200 db vakok és gyengénlátók számára használható számítógépes céleszköz van a világon. Miközben a piacon új géptípusok terjednek el, a DOS-t fejlettebb operációs rendszerek váltják fel, a látássérültek túlnyomó többsége kénytelen megmaradni IBM PC használatánál és DOS-os környezetnél. A DOS-os alkalmazásokon belül is inkább a szöveges jellegűekhez lehet hozzáférni különböző képernyőolvasók segítségével. Arról le kell mondani, hogy a grafikát élvezzük, holott az új programoknál (pl.: a Windows-os alkalmazások esetében) az eredetileg nem-grafikus feladatokat is grafikusán ábrázolják. A grafika a szemnek készül és nem lehet igazán elmondani az információ tartalmát.

A Windows és alkalmazásainak használata előbb-utóbb megoldódik a vakok számára is, hisz jelenleg már létezik hanggal, ill. Braille írással kijelzett Windows kezelőfelület és szövegfeldolgozó, melyeket az IBM és a BAUM cég fejlesztett ki

vakok számára. Amíg ezen termékek árai nem csökkennek, meg kell elégedni az IBM PC DOS-os felülettel.

A jelen leírás csak a DOS alatt futó alkalmazásokat tárgyalja.

## **2. Helyzetkép**

Az országban élő látássérültek mintegy 1/3-a, a budapestiek 70-80%-a rendelkezik valamilyen magyarul beszélő számítógéppel. Beszédszintetizátorral kiegészített számítógépet használnak a Vakok Általános Iskolájában (200 tanuló), a Magyar Vakok és Gyengénlátók Szövetségében, ahol számítógépes alapozó tanfolyam van, a Neuman János Számítástechnikai Szakközépiskolában (9 tanuló), a Vakok Elemi Rehabilitációs Csoportjánál.

Látássérültek körében, ahol beszélő gépeket használnak, mindenhol tapasztalható az a tragikus nyelvérzék romboló hatás, mely abból adódik, hogy az angol szöveg magyar kiejtéssel (fonémarendszer szerint) hangzik el! A jelenleg rendelkezésre álló magyar "Text to Speech" természetszerűen nem alkalmas arra a célra, hogy különbséget tegyen magyar vagy idegen szavak kiejtése között.

Tehát valós igényt oldana meg egy olyan céleszköz (kötött szókészlettel is) amely magyar szövegkörnyezetben a használat közben megjelenő angol nyelvű üzeneteket helyes angol kiejtéssel mondaná el. Különösen fontos ez főleg az angol nyelvet nem eléggé ismerő látássérültek számára, akik számítógépet használnak tanulásuk során.

## **3. Magyar nyelvű beszédszintetizátorok vakügyi felhasználásra**

### **3.1. PC-Voice**

Az első vakügyi célra kifejlesztett képernyő olvasó, melyet már csak kevesen használnak.

### **3.2. Brailab PC (Brailab, 1993)**

A magyar vakok körében ez a legelterjedtebb céleszköz, a számítógéppel rendelkezők mintegy 80%-a használja. Konceptiója túl egyszerű és zárt! Minden egyes alkalmazáshoz külön-külön kell megírni egy kiegészítő hang-kimenetet, melyet kizárólag a szerzők készíthetnek el (pl.: NE-hoz, NC-hez, Micro Star-hoz léteznek kísérleti jelleggel).

A Brailab PC adapter egy hangszóróból áll beépített beszéd szintetizátorral, egy tápegységből és két mágneslemezről a felhasználói programokkal. Hordozhatósága és a különböző géptípushoz való alkalmazhatósága jó. A számítógépbe hardware-es beavatkozást nem igényel, egy szabad párhuzamos printer csatlakozóra dugaszolható, a software installálása egyszerű. A program kezelhetősége kiváló, ami annak is köszönhető, hogy a szerzők egyike maga is vak és nagyon sok célszerű kezeléstechnikai szempontot figyelembe vett a fejlesztéskor. A BrailabPC zártsága abban áll, hogy más képernyőolvasóval nem működik együtt más hardware-rel sem. Egy képernyőolvasót úgy célszerű megalkotni, hogy bármelyik felhasználói programhoz annak igényeitől függően online beállítható legyen.

### 3.3. PC-Robot (Rc-Robot, 1994)

A PC-Robot egy MultiVox kompatibilis interaktív alkotó rendszer, magyar beszéd és ének szintetizátor grafikai megjelenítővel.

A rendszer elemei:

- PC-Robot beszéd szintetizátor kártya egy hangszóróval,
- egy felhasználói és programozási kézikönyv,
- egy mágneslemez a programokkal.

A beszéd szintetizáló programrendszer, mind software, mind hardware vonatkozásában egyedülálló képességekkel rendelkezik. A speciális célokra kifejlesztett programcsomagok biztosítják, hogy a társadalom széles rétegei eredményesen tudják használni ezt a rendszert. A beszéd és ének előállítás szabadon programozható, bármilyen tartalmú szöveg felolvastatható, ének

szerkeszthető. A rendszer a kimondandó szöveget magyar nyelven tiszta hangzású, dallamos, ritmusos köznapi férfi hanggal mondja el. Számos programozási opció biztosítja, hogy a beszéd programozása során a legváltozatosabb jellemzőkkel rendelkező beszédjelet állítsuk elő.

A PC-Robottal együtt használhatók mindazok a képernyőolvasók, amelyek standard kommunikációs PORT-ra csatlakoznak. Erre egy példa a magyar fejlesztésű PSR (Professional Screen Reader). Számos, hazánkban egyre gyakrabban fellelhető képernyő olvasó használható a PC ROBOT-tal együtt (pl.: "Vocal Eyes", "HAL").

#### **4. Részletes teszteredmények**

A Brailab PC és MultiVox PC ROBOT + HAL összehasonlítása

A beszéd érthetősége kiváló mindkét eszköznél. A Multivox PC ROBOT (továbbiakban MV) nyolc, a Brailab PC (a továbbiakban Br) hat beszéd-paraméter beállítási opcióval rendelkezik. A Br-nál hanghossz és tagoltsági fokozat állítására nincs mód, a magánhangzók túl hosszúak (*e, á*), némely mássalhangzó szinte nem hallható (*t, k, p*), ezek megváltoztatására nincs mód.

A MV +Hal esetén a legnagyobb beszédsebesség a Br-hoz képest kétszerese is lehet. A Br-ban a leggyorsabb sebesség meglehetősen lassú a gyakorlott használók számára, 80-90 szó percenként.

Brailab PC esetén online üzemmódban (nem képernyőolvasó mód) használat közben csak az a szöveg hangzik el, mely a BIOS 10H megszakításon keresztül kerül kiírásra a képernyőre, vagy a virtuális COM4-re van kiirányítva. Csak ebben az üzemmódban intonál. Abban az esetben, ha egy adott alkalmazás közvetlen a VideoRAM-ba ír, azt a Br és a felhasználó nem veszi észre, elolvasása offline módban lehetséges. Nincs lehetőség arra, hogy a

képernyő tartalma (vagy részlete) egy gyorsgomb nyomásra elhangozzék. (Nagyon hiányzik!)

Offline üzemmódban (képernyő olvasó mód) a Br nem függeszti fel az adott program futását, tehát ha nem áll a kép, ily módon sem lehet hozzáférni a tartalmához (pl.: BBS-nél). Aktivizálni az INS és a DEL egyszerre történő lenyomásával lehet. Ez a lehetőség a billentyűzet használat szempontjából nagyon kellemes, ezen billentyű-kombinációt kevés alkalmazás használja. De vannak alkalmazások, melyek nem engedik meg az INS és DEL egyszerre történő lenyomását, azaz nem lehet átmenni olvasó módba. Ezek gyakorlatilag használhatatlanok ily módon a vakok számára, például Kartotek, Telix. Ezen alkalmazások a Hal-lal kellemesen használhatóak, olvasó módba sem kell átmenni, a nyilak mozgására a megfelelő menüsor hangzik el és kényelmesen lehet kiválasztani és aktivizálni őket.

On és offline beszéd paraméterezése: a Br egy file-ban tárolja a kivételszótárt, amely a felhasználó által editálható, bővíthető ASCII text file. Ennek a használatával megkülönböztethető a szavak leírt és elhangzott alakja (pl.: voice=vojsz) kizárólag online módban. A Hal tartalmaz egy "prompt.hal" fájlt. Ez egy ASCII text, melynek átszerkesztésével mindkét üzemmódban azt lehet hallani, amit a felhasználó kíván.

Billentyűzet: A Br-nak onlineban nincs gyorsgomb készlete, csak olvasó módban. A Hal az alábbi gyorsgombokkal rendelkezik online-ban: előző, aktuális, következő betű, szó, sor, paragrafus olvasása, ill. betűzés, olvasás a kurzorig (többnyire összeegyeztethető az alkalmazással).

Billentyű echo: A Br-nál csak megszakításos üzemmód van, amely a leghasználhatóbb gépeléskor. A Hal-nál nem megszakításos is létezik, előnye, hogy billentyű-leütésre a program tovább fut, de a beszédpufferben lévő szöveg még elhangzik. A Br-nál Keyboard echo megváltoztatására nincs mód, hiányos. Nem tanítható, tehát például a "§" paragrafus jel sohasem hangzik el.

Magyar nyelvű billentyűzet használata: a Br tartalmaz magyar billentyűzet kiosztást, változtatható magyarra offline-ban is (nem tartalmaz "/" per jelet). Semmilyen más billentyű-átdefiniáló program nem használható a Br-bal együtt, más karaktert echo-z és más jelenik meg. A Br a nagybetűket 2-3-szor hangosabban mondja, a Hal nem változtat hangerőt, kimondja hogy "nagy", és/vagy sípol hozzá.

Némítás: a Br esetén a beszéd ki-be kapcsolására csak a DOS promptnál van lehetőség. Ideiglenes némításra a Shift, CTRL, vagy Alt nyomvatartásával van mód. A Hal esetén mindez gyorsgombokkal érhető el.

Helyigény a memóriában: a Br 3.0 verziója 61 KB. A MV és Hal öt db rezidens részei 264 KB-ot foglalnak le a memóriából. Ez utóbbi kissé megterhelő annak ellenére, hogy a felső memóriába is tölthetőek. Néhány memóriaigényes program használhatatlan emiatt, úgymint NDD, Speeddisk, Terminat ...

Szövegszerkesztés: a legszélesebb körű felhasználási terület vakok számára. A tesztelés néhány közismert szövegszerkesztővel történt mindkét beszélőt használva, úgymint NE, kedit, word55, WP51 és a Frame Work programcsomag. A részletes leírás meghaladja ezen kereteket.

Az adatkezelők már bonyolultabbak. A Br-bal, gyakorlatilag csak offline módban lehet elolvasni a táblázatokat. Az összefüggések megértése is elég időigényes, mivel hogy sor-oszlop pozíciókat kell lekérdezni, majd olvasó ablakokat definiálni rájuk. Ezzel szemben a Hal vagy Vocal Eyes fel tudja osztani a képernyőt színattributum, oszlop-, sor- és ablakszám szerint, és ezután egyszerű képernyőn való navigálással online módon a kiemelt adathalmaz hangzik el. Erre jó példa a FrameWork.

A Multivoznak van egy igen jó beszédszolgáltatása. Sok szöveges állományban előfordulnak magyar ékezetes karakterek a következőképpen jelölve: é=e', á=a', ú=u', ö=o:, ü=u:, ű=u", mindezek érthetetlenek Br-al olvastatva fel. A MV képes arra,

hogy ezeket a betűkapcsolatokat, stringeket magyar ékezetes betűknek megfelelően ejti ki.

A Br olvasó módba belépés után nem függeszti fel az adott program futását. Ha tehát nem áll a képernyő mozdulatlan, de az alkalmazás dolgozik, (pl.: írás - olvasás a lemezre, memória műveletek), akkor késleltetve lehet elolvasni a képernyő tartalmát. Például a DOS Smartdrv használatakor a késleltetett lemezre írás miatt ki kell várni a lemezre mentés idejét, vagy ismételten ugyanazt a sorolvasást kell kérni. A Hal esetén nincs ilyen probléma, de az alkalmazás futásának felfüggesztése, vagy leállítása men biztos, hogy mindig megtehető.

A Br-nál csak offline-ban lehet hozzáférni a képernyőn lévő karakterek attribútumaihoz, a Halnál majdnem mindkét állapotban. Ez alól van néhány kivétel, de a Br azokra eleve nem képes. Például a Br-nál nincs billentyű és képernyő karakter tanítási lehetőség, a színeket nem tudja elmondani, (előtér - háttérszín) csak a hexadecimális színek kódját. A Hal, Vocal Eyes és PSR esetében ASCII text-file tartalmazza a színek és a karakter képének leírását, mind a 255-öt és ezen állományok átírhatók akár más nyelvre is.

Olvasó ablakok: a teljes képernyő négyzethálószerűen felosztható ablakokra úgy, hogy a területek értelmileg egy egységbe tartoznak és ezen kijelölt részeket összefüggően lehet felolvasatni. Mindkét képernyőolvasó koncepciója ugyanaz, csak a hozzáférési és definiálási módban vannak különbségek. A Br-ban hattal több (16 db) ablakot lehet definiálni egy adott alkalmazásra, melyek induláskor a DOS promptnál file-paraméterrel adhatók meg. Ezeket a fix ablakokat az offline módban a funkcióbillentyűkkel lehet aktivizálni.

A Hal esetében online-ban gyorsgombokkal lehet felolvasatni a kiválasztott ablak teljes tartalmát. Az olvasó ablakok koordinátáit és a hangkimenetre vonatkozó beállításokat \*.SPC

file-ok tartalmazzák, melyeknek cseréje, letöltése és definiálása az alkalmazás futása közben is online lehetséges.

### **5. Optikai karakterfelismerés Recognita programmal**

Az optikai karakterfelismeréssel nyomtatott szöveges állományok, könyvek hozzáférhetőek, elolvashatóak a látássérültek számára. Ez a lehetőség a vakok számára szinte felmérhetetlen előny. A papíron lévő szöveget szkennelés után a Recognita a kívánt ASCII text file formátumban készíti el, mely ezután felolvastatható, átszerkeszthető, javítható a beszélő géppel. Sajnos a folyóiratok, napilapok optikai karakterfelismerése nem a legtökéletesebb, elég sok a nem felismert szimbólum (betű), a karaktertévesztésből adódó bizonytalanság a szöveg megértését rontja. Ez leginkább olyan esetekben fordul elő, amikor nyomdatechnikailag hibás vagy gyenge minőségű az írott szöveg. Ezzel szemben szakirodalmi, szépirodalmi s olyan nyomtatványok esetén, melyek laser vagy megközelítően hasonló jó minőségűek, a karakterfelismerés 1-2 ezrelék jósági azonosítási arányt is elér, gyakorlatilag tökéletesen használható felolvasásra.

A Recognita programnak létezik egy "Auge" elnevezésű kezelőfelülete csökkentlátóknak. A fő és almenük mind a Brailab PC, mind a Multivox + HAL képernyőolvasónál egyformán online-ban beszélnek, jól kezelhetőek. Folyamatos szkennelés esetén van lehetőség mindkét beszélő rendszerben a beszéd elnémitására a gyorsabb feldolgozás céljából.

### **Irodalom**

- Brailab PC felhasználói kézikönyv. KFKI. Budapest 1993.
- Equipment for Visually Disabled People: An International Guide. Royal National Institute for the Blind. London 1993.
- PC-Robot felhasználói kézikönyv. Nikol. Budapest 1994.

# ADDITÍV ZAJ HATÁSA A BESZÉD LÉNYEGKIEMELT PARAMÉTEREIRE

Ast László–Ács György–Kalotay Béla–Marosi Gyula–  
Tatai Péter

**Budapesti Műszaki Egyetem,  
Távközlési és Telematikai Tanszék**

Jelen cikkben a szerzők ismertetik a beszédfeldolgozás 6. széleskörben használt, lényegkiemelt paraméterének változását 3 különböző típusú additív zaj jelenlétében. Két, egymástól alapvetően eltérő eljárás kerül bemutatásra a zaj által okozott nem kívánt hatások csökkentésére. Előbb a jól ismert spektrális kivonás, amely nem bizonyult kielégítőnek, majd a referencia paraméterek módosításának módszere, ami nagy javulást eredményez a beszédfelismerésben.

## 1. Bevezetés

Általában az automatikus beszédfelismerő rendszereket és az alacsony bitsebességű beszédkódolókat zajmentes munkahelyen, jó minőségű audio bemenetekkel vizsgálják, azonban a gyakorlatban a környezet ritkán zajtalan, és a legtöbb alkalmazás esetén alkalmazkodni kell a változó akusztikai környezethez. A napjainkban működő rendszerek minősége általában látványosan romlik, ha más akusztikai környezetben használják, mint amiben tervezték. Megvizsgálva a jelentős minőségromlás alapvető okait kiderül, hogy a probléma a lényegkiemelt paraméterek szintjén keresendő. Ezek a paraméterek szorosan kötődnek a beszéd becsült, rövid idejű spektrumához, így additív zaj esetén egyértelműen tükrözik a környezeti zaj által okozott spektrumváltozást. Mivel a háttérzaj jellege tipikusan időben változik és általában előre nem ismert módon, ezért a megoldást egy olyan adaptív eljárás jelentheti, amely lehetőleg zajérzéketlen lényegkiemelt paramétereket képes szolgáltatni.

Jelen cikk a beszédfeldolgozás 6, széleskörben használt, lényegkiemelt paraméterének változását vizsgálja meg 3 különböző típusú additív zaj jelenlétében, 2 különböző mintavételi frekvencia (8 illetve 16 kHz) esetén. A zajérzékenység csökkentésére 2 eljárást vizsgálunk meg: előbb az irodalomban is javasolt spektrális kivonást a rövid idejű autokorrelációs együtthetők tartományában, majd a referencia paraméterek módosításának módszerét.

## 2. Az analízis általános körülményei

Egy paraméterjavító eljárás ismeretéhez először is magát a javítandó paramétert és a javulás mérésének módszerét kell egyértelműen definiálni. A vizsgálat során 6 különböző spektrális paramétert használtunk, amelyek értékelésére egy távolság mértéket választottunk. A cikk ábrái relatív távolságot mutatnak az egyes paraméterek saját terében. A  $K$  szegmenses  $\xi$  jel  $P$  dimenziós  $F$  lényegkiemelt paramétervektora esetén,  $\mu$  additív zaj jelenlétében a távolság definíciója:

$$d = \frac{\sum_{i=0}^{K-1} \|F_{\xi+\mu}^{(i)} - F_{\xi}^{(i)}\|}{\sum_{i=0}^{K-1} \|F_{\xi}^{(i)}\|}$$

– normaként az euklideszi norma négyzetét használva – a következő kifejezéssé alakul:

$$d = \frac{\sum_{i=0}^{K-1} \sum_{p=0}^{P-1} (F_{\xi+\mu}^{(i)}[p] - F_{\xi}^{(i)}[p])^2}{\sum_{i=0}^{K-1} \sum_{p=0}^{P-1} (F_{\xi}^{(i)}[p])^2}$$

A vizsgálat során használt adatbázis az 1, 2, 3, ... 10 számok angol nyelven, anyanyelvi beszélők által 200-szor kiejtett hanganyagából állt. Három jellegzetes zajtípust használtunk: fehérzajt, egy nagy sebességgel haladó autó belsejében rögzített autózajt és háttérbeszéd zajt. A 6 vizsgált lényegkiemelt paraméter a következő:

- rövid idejű autokorrelációs együtthetők (ACC),
- lineáris predikciós együtthetők (LPC),
- reflexiós (avagy parciális korrelációs) együtthetők (PCC),
- logaritmikus területarány együtthetők (LAR),
- kepsztrális együtthetők (CEP),
- spektrum vonal párok (LSP).

### 3. A zaj hatása az autokorrelációs tartományban

#### 3.1. A zaj additív tulajdonsága az AC tartományban

A legtöbb alkalmazott beszédfelismerő rendszer valamely szintjén megjelennek a bemeneti jel rövid idejű autokorrelációs (AC) együtthetói, mivel a lényegkiemelő algoritmusok többsége a lényegkiemelt paramétereket az AC tartományon keresztül számítja.<sup>1</sup> Vizsgáljuk meg az autokorrelációs együtthetők viselkedését additív jelek esetén! Természetesen a beszédfeldolgozó

<sup>1</sup> Például a lineáris predikciós, reflexiós, logaritmikus területarány és a kepsztrum együtthetőit is általában a rövid idejű autokorrelációs együtthetőkből számítják, azaz megjelennek az AC együtthetők, bár csak mint közbülső eredmények.

rendszerek a gyakorlatban  $R_{\xi}$  ( $\xi$  autokorrelációs függvénye) rövid idejű megfelelőjét használják a beszéd feldolgozásának szegmentáltsága miatt, amit a továbbiakban  $AC_{\xi}$ -vel jelölünk. Így a beszéd  $i$ . szegmensében  $R_{\xi}$  helyett:

$$AC_{\xi}^{(i)}[n] = \frac{1}{N-n} \cdot \sum_{k=n}^{N-1} \xi_{i:N+k} \cdot \xi_{i:N+k-n} \quad n \in [0 \dots N-1]$$

írunk, ahol  $N$  a szegmensek mintáinak száma. Bár a nulla várható értékű ergodikusan jelek  $R_{\xi}$  függvényei additívak, a rövid idejű autokorrelációs függvényre ez a definícióból is jól láthatóan már nem igaz:

$$AC_{\xi+\mu}^{(i)}[n] = \frac{1}{N-n} \cdot \sum_{k=n}^{N-1} (\xi_{i:N+k} + \mu_{i:N+k}) (\xi_{i:N+k-n} + \mu_{i:N+k-n}).$$

Az összegben belüli szorzásokat kifejtve a következőt kapjuk:

$$\begin{aligned} AC_{\xi+\mu}^{(i)}[n] &= AC_{\xi}^{(i)}[n] + AC_{\mu}^{(i)}[n] + \frac{1}{N-n} \cdot \sum_{k=n}^{N-1} \xi_{i:N+k} \cdot \mu_{i:N+k-n} + \\ &+ \frac{1}{N-n} \cdot \sum_{k=n}^{N-1} \xi_{i:N+k-n} \cdot \mu_{i:N+k} \end{aligned}$$

Vegyük észre, hogy az iménti eredmény eltér az általunk remélttől:

$$AC_{\xi+\mu}^{(i)}[n] = AC_{\xi}^{(i)}[n] + AC_{\mu}^{(i)}[n].$$

Milyen körülmények között számít mégis jó becslésnek a rövid idejű autokorrelációs együtthatók additivitását feltételezni? Mivel az egzakt összeadhatóság nyilvánvalóan nem érhető el, ezért

ennek a közelítése kisebb-nagyobb pontossággal lesz csak igaz. A legnagyobb pontosság a következő négy feltétel egyidejű teljesülése mellett érhető el:  $E\{\xi\}=E\{\mu\}=0$ ;  $\xi$  és  $\mu$  független;  $\xi$ ,  $\mu$  és  $\xi\mu$  ergodikusak, és az összegek hossza ( $N$ ) a végtelenhez tart.

### 3.2. Módosítás az autokorrelációs tartományban

A következő megválaszolendő kérdés: hogyan használható a fenti állítás amennyiben ismeretünk  $\xi$  és  $\mu$ -ról  $M\{AC_{\mu}[n]\}$ -ra korlátozódik, ahol  $M\{\cdot\}$ -t az átlagolás operátora az összes szegmensre. Úgy tűnik,  $M\{AC_{\mu}[n]\}$  megfelelő közelítést ad  $R_{\mu}[n]$ -re. Mivel

$$R_{\xi+\mu}[n] = R_{\xi}[n] + R_{\mu}[n] \approx R_{\xi}[n] + M\{AC_{\mu}[n]\},$$

és  $\xi$  és  $\xi+\mu$  ergodicitása miatt:

$$AC_{\xi+\mu}[n] \approx AC_{\xi}[n] + M\{AC_{\mu}[n]\}$$

így a jel egy szegmensének rövid idejű autokorrelációs együtthatói a következőképpen közelíthetők:

$$AC_{\xi}[n] \approx AC_{\xi+\mu}[n] - M\{AC_{\mu}[n]\}$$

egy bizonyos pontossággal.

Felmerül egy lényeges kérdés: lehet egyáltalán  $AC'$  egy valós jel rövid idejű autokorrelációs függvénye? Sajnos, az általános válasz *nem*, ami egyben azt is jelenti, hogy az így kapott együtthatók nem használhatók további számolásra (mint például a reflexiók vagy spektrális együtthatók számítására).

Empirikusan is belátható, hogy  $AC[0]$  (ami a jel adott szegmensének az energiája és így nem lehet negatív) negatív lesz a

jel bizonyos szegmenseire az átlagának kivonása miatt.<sup>2</sup> Matematikailag, az  $R_{N \times N}$  mátrix pozitív szemidefinit kell legyen,<sup>3</sup> ám két pozitív szemidefinit mátrix különbségére ez nem biztosított.

#### 4. Referencia paraméterek módosításának módszere

Egy másfajta zajérzékenység csökkentő eljárási lehetőség a referencia paraméterek módosítása. A módszer egy alapvetően ötletes megoldáson alapszik: ne a zajjal terhelt jel paramétereit próbáljuk javítani, hanem módosítsuk (az eredetileg zajmentesnek feltételezett) referencia jeleinket adaptívan a zajosság irányába, és így a felismerési eljárásban a felismerendő jelet – pontosabban annak lényegkiemelt paramétereit – egy hasonló zajkörülmenyek közé transzformált referencia adatbázis elemeivel vethetjük össze. A módszer előnye, hogy megkerüli a nem pozitív szemidefinit rövid idejű autokorrelációs mátrixok megjelenésének problémáját, hisz amennyiben például az AC tartományban végezzük a referencia adatbázis adaptálását, akkor az  $R_{N \times N}$  mátrixokat összeadjuk, amely művelet megőrzi az adott szemidefinitási tulajdonságot. Nyilvánvaló hátrányként kell viszont megemlíteni, hogy ekkor nem *egy* jelet (a felismerendőt) kell valamely algoritmus szerint transzformálni, hanem az *összes* többbit.

Az 1-6. ábrák a beszédfelismerésben széles körben használt lényegkiemelt paraméterek változását mutatják additív zaj hatására 3 különböző esetben: a paraméterek korrekciója nélkül, korrekcióval a rövid idejű autokorrelációs tartományban,

<sup>2</sup> A spektrális tartományban gondolkodva (mivel a Fourier transzformáció megőrzi az additívítást) a spektrumból egy konstans vektort kivonva negatívvá válhat bizonyos időtartamokra. Abból, hogy

$$R_{\xi} = \xi[n] * \xi[-n] \text{ látható, hogy } \mathcal{F}\{R_{\xi}\} \geq 0, \text{ és } \text{Im}\{\mathcal{F}\{R_{\xi}\}\} = 0 \text{ -nek}$$

pedig fenn kell állnia, mert

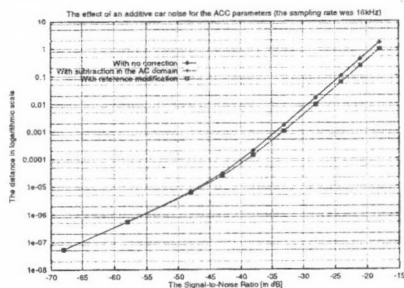
$$\mathcal{F}\{R_{\xi}\} = \mathcal{F}\{\xi[n] * \xi[-n]\} = \mathcal{F}\{\xi\} \cdot \mathcal{F}^*\{\xi\} = |\mathcal{F}\{\xi\}|^2.$$

<sup>3</sup> A vektorok átlagából képzett  $R_{\xi+\mu}$  és Toeplitz mátrix.

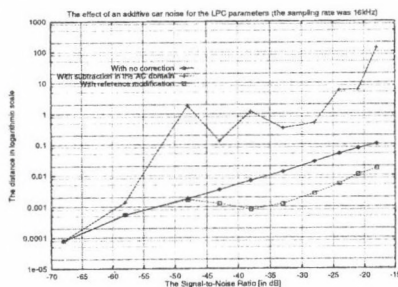
illetve a referencia adatbázis adaptív transzformálásával; a vízszintes tengelyeken a jel-zaj viszony dB-ben kifejezett értéke, a függőleges tengelyeken a paraméterek egymáshoz viszonyított távolsága található logaritmikusan léptékben ábrázolva. Az ábrák 16 kHz-es mintavételi frekvenciával, autózaj hatására végzett kísérletek eredményeit mutatják.

## 5. Következtetések

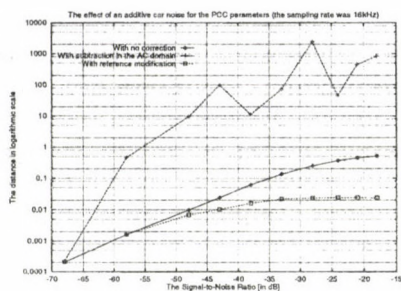
Két módszert mutattunk be, melyek bizonyos paraméterekre képesek egy automatikus beszédfelismerő rendszer megbízhatóságát javítani. Azokat a paramétereket mindenképpen célszerű egy felismerő osztályozási algoritmusának figyelembe venni, melyek a leginkább zajtűrőek. A tapasztalatok azt mutatják, hogy az LPC alapú paraméterek az autokorrelációs tartományi kivonás után megbízhatóbb eredményt szolgáltatnak, ez a transzformációs módszer javulást csak az autokorrelációs paraméterekre alapozott távolságnál okoz, viszont pusztán ezeket a paramétereket beszédfelismerésre a gyakorlatban nem használják. A reflexiók, a kepsztrális, és a LAR paraméter tartományban a referencia mósosítással végzett kísérletek viszont jó eredményeket szolgáltatnak, az additív zaj által okozott paramétertorzulás jelentősen csökken.



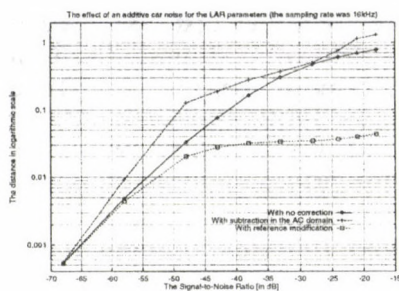
1. ábra



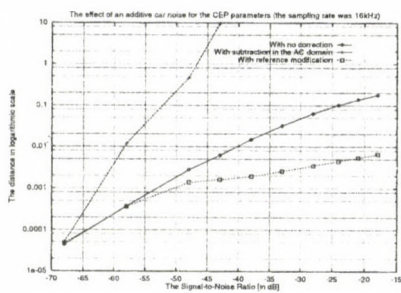
2. ábra



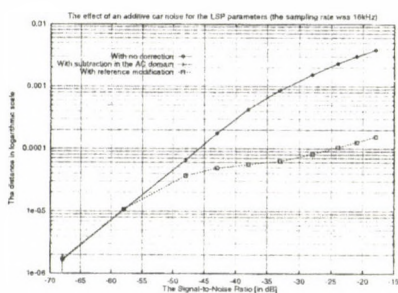
3. ábra



4. ábra



5. ábra



6. ábra

### Irodalom

Ahmed, M. S.: Comparison of noisy speech enhancement algorithms in terms of LPC perturbation. IEEE Trans. on Acoustics, Speech, and Signal Processing, Vol.37, No.1 January 1989, 121-125.

Hermansky, H. et al.: Compensation for the effect of the communication channel in auditory-like analysis of speech (RASTA-PLP). Proc. EUROSPEECH '91. Genoa 1991, 1367-1370.

Hermansky, H.-Morgan, N.-Hirsch, H-G.: Recognition of speech in additive and convolutional noise based on RASTA spectral processing. Proc. ICASSP '93 (preprint)

Gordos Géza–Takács György: Digitális beszédfeldolgozás. Műszaki Könyvkiadó. Budapest 1983, 154-170.

Kalotay, B.–Marosi, Gy.–Tatai, P.: Automatic speech recognition. OTKA report. Department of Telecommunications and Telematics, TUB. Budapest 1994.

Markel, J.D.–Gray, Jr. A.H.: Linear prediction of speech. Springer-Verlag. 1976.

Rabiner. L.R.– Schafer, R.W.: Digital processing of speech signals. Prentice-Hall. Inc. Englewood Cliffs. New Jersey 1978.

Tatai, P.: Front-end processing for speech recognition over distorted channels. Proc. of the Czech-Polish-Hungarian workshop on circuits theory and applications. Prenet 1993, 120-124.

## HAZAI ÉS NEMZETKÖZI TÁMOGATÁSSAL VÉGZETT BESZÉDKUTATÁSOK

A beszédkutatás fontosságát, valamint a hazai eredmények elismerését mutatja az a tény, hogy az utóbbi években mind hazai, mind pedig nemzetközi támogatásokat is elnyertünk. Hazai viszonylatban az OTKA és az OMFB támogatások segítették a munkát, nemzetközi viszonylatban pedig az Európai Közösség COST 219 projektje, valamint egy SPEAK! Copernicus projekt (német–magyar közös kutatás és fejlesztés, amelyet az Európai Közösség finanszíroz). Az alábbiakban ezekről a munkákról számolunk be.

### 1. A magyar beszéd ritmikai és dallamszerkezetének kutatása

Támogató: OTKA T013565, az 1994–97 időszakra

Témafelelős: Olasz Gábor, MTA Nyelvtudományi Intézet, Fonetikai Osztály

A kutatás **célja**: megismerni és szabályokkal jellemezni a magyar beszéd ritmikai és dallamszerkezetének főbb építőelemeit.

A kutatás **részletezése**: a magyar beszéd mondatszintű elemeinek akusztikai vizsgálata időszerkezeti és dallamszerkezeti szempontból. Az időszerkezet tekintetében annak feltárása, hogy a kiejtett mondatban milyen időszerkezeti struktúrák realizálódnak, azok mennyire köthetők szabályokhoz. A dallamszerkezet szempontjából az invariáns dallamformák meghatározása, valamint az érzelmek kifejeződésének tükröződése a dallamszerkezetben. A kutatás második szakaszában a két vizsgálatot összekapcsolva próbáljuk meghatározni a párbeszédes közlésformákra jellemző prozódiai (dallam, idő- és intenzitás szerkezet együttese) struktúrákat. A fenti vizsgálatokat kiegészítik percepció tesztek, akusztikai vizsgálatok a férfi és női beszéd belső dallamtörvényeinek meghatározására, valamint az érzelem kifejezésének akusztikai szintű kutatásai. Az egész kutatás jellemzője a

szintézissel történő visszacsatolás, vagyis az eredmények, szabályok ellenőrzése a beszéd-szintézis módszerével.

## 2. Szófonetikai vizsgálatok

Támogató: OTKA 13863 az 1994–97 időszakra

Témafelelős: Gósy Mária, MTA Nyelvtudományi Intézet, Fonetikai Osztály

A kutatás **célja**: azoknak a kommunikatív értékű paramétereknek a meghatározása és elemzése, amelyek a szavakat és a szavak kapcsolódásait jellemzik az izolált ejtéstől a folyamatos beszédig.

A kutatás **részletezése**: választ akarunk kapni arra, hogy az izolált ejtésű, illetve a mondatban, szövegben megjelenő szavak artikulációjában melyek az állandó és mik a változó paraméterek. Ennek következtében mely akusztikai jegyek tekinthetők invariánsnak, melyek redundánsnak és melyek irrelevánsnak mind a hullámforma-elemzés, mind a beszédészlelés szempontjából. Magyarázatot igyekszünk megfogalmazni az egyedi beszéd artikulációs/akusztikus eltéréseire vonatkozóan. Elemzéseket terveztünk annak leírására, hogy a koartikulációs jelenségek milyen mértékben kontextusfüggőek, illetőleg miként és mennyiben érvényesül a szupraszegmentális szerkezet (pl. hangsúly, intonáció, tempó) módosító hatása az adott szóképzésben, szó szerkezetben izolált és szöveges megjelenéskor. Vizsgáljuk a beszédészlelési és beszédmegértési stratégiák módosulását a fentiekben vázolt különbségek esetében; valamint a dekódolás különféle szintjeinek együttműködését a kidolgozandó szófelismerési folyamaton belül.

A szófonetikai kutatások döntően kísérleti vizsgálódásokra épülnek, amelyeket – az elméleti keretnek megfelelően – történeti fonotaktikai elemzések egészítenek ki. Minthogy cél a szövegfonetikához történő kapcsolódás járható útjának kidolgozása, a tervezett munkában a beszédelemzések mellett a nyelvelsajátítás és a kóros szóformálás vizsgálata szükségszerűen

helyet kap. Ez utóbbi három aspektus (ti. a finnugrisztikai alapok fonetikai szemléletű megközelítése, a nyelvelsajátítás adathalmazából nyerhető tanúságtétel és a hibás kódolás/artikuláció eredményezte alakzatok létrejöttének elemzése) nem kiegészítő jellegű adat- és ismeretanyagot képvisel, hanem szerves tartozéka a központi problémakörnek. A történetet, az "ontogenezist" és a társadalmilag "nem elfogadhatót" is tekintetbe vevő koncepció az összefüggések pontosabb feltárásához, bizonyos kérdések egyértelműbb magyarázatához vezet.

Az eredményeket mind a további elméleti fonetikai kutatások, a fonológia, egyéb nyelvészeti diszciplínák (pl. szemantikakutatás, grammatikaelmélet), mind az alkalmazott kutatások (pl. beszédpatológia, idegennyelv-oktatás), valamint – az interdiszciplináris jellegből adódóan – a társtudományok (pl. pszicholingvisztika, neurolingvisztika) használhatják fel. Sok esetben az alkalmazásokhoz (pl. beszélőazonosítás) olyannyira szükséges alapkutatások tekintetében hiányt pótlóak a szófonetikai kutatások eredményei.

### **3. Future telecommunication and teleinformatics facilities for disabled and elderly people (COST 219, 27 európai részt vevő ország)**

Támogató: Európai Közösség, (Cooperation in Science and Technology, 1993-tól)

Magyarországi koordinátor: Németh Géza, BME Távközlési és Telematikai Tanszék

Nyelvészeti kutatási partner: MTA Nyelvtudományi Intézet, Fonetikai Osztály

Témafelelős: Gósy Mária, MTA Nyelvtudományi Intézet, Fonetikai Osztály

A kutatás **célja**: koraszülöttek és rizikó újszülöttek iskolába lépést megelőző vizsgálata annak megállapítására, hogy beszédészlelési és beszédmegértési szintjük megfelel-e az életkoruknak.

A kutatás **részletezése**: több, mint száz koraszülöttet és rizikó újszülöttet vizsgáltunk a GMP-diagnosztikával. Tizennégy-féle alteszt eredményeinek osztályozása és elemzése alapján határoztuk meg az egyes gyermekek beszédpercepció teljesítményét. A vizsgálatok a következő területekre terjedtek ki: hallás, az akusztikai kulcsok felismerése, a beszédészlelés három szintje, a szeriális észlelés, beszédhang-diszkrimináció, vizuális észlelés, ritmusészlelés, szókincsaktivizálás, a rövid idejű vizuális és verbális memória, kézdominancia, irányfelismerés és a szövegértés. A kapott eredmények az azonos életkorú és ép fejlődésű gyermekek szokásos teljesítményéhez viszonyíthatók.

#### **4. Improvement, evaluation and testing of a low cost multilingual portable speaking aid for the speech impaired, which can be used in a European context**

Támogató: OMFB (ACCORD project No. H 9112-0507, az 1993–95-ös időszakra)

Külföldi kutatási partnerek:

Department of Electrotechnical and Computer Engineering,  
Faculty of Engineering of the University of Porto, Portugal

Department of Signals, Systems and Radio Communication,  
School of Telecommunications, Technical University of Madrid,  
Spain

Magyarországi koordinátor: Németh Géza, BME Távközlési és Telematikai Tanszék

Nyelvészeti kutatási partner: MTA Nyelvtudományi Intézet, Fonetikai Osztály

Témafelelős: Olasz Gábor

A kutatás **célja**: soknyelvű (magyar, spanyol, portugál) beszédsszintézis technológiájának kidolgozása és megvalósítása. Beszédsérültek által verbális kommunikációra (telefonálásra is) használható beszédkommunikátor készülék nyelvi moduljának

kidolgozása a fenti három nyelvre a külföldi partnerek közreműködésével és támogatásával.

A kutatás **részletezése**: A kutatás során a gépi beszédelőállítás tudományos és gyakorlati tapasztalatait felhasználva továbbfejlesztettük a magyar és spanyol beszédszintézist, valamint megterveztük és létrehoztuk az első portugál nyelvű, PC-n használható beszédszintetizátort. A fejlesztések során új hangadatbázisokat dolgoztunk ki a hangszínezet természetesebbé tételére, új szabályrendszereket fejlesztettünk a mondat-prozódia javítására és ezeket beintegráltuk a beszédprotézis eszközbe. A kész mintakészülékeket mindhárom országban orvosi szakintézményekben tesztelték és használják.

## 5. Speech Generation in Multimodal Information Systems

Támogató: Európai Közösség (Copernicus, Project No. 10393, az 1994–96 időszakra)

Témafelelős: Olasz Gábor, MTA Nyelvtudományi Intézet, Fonetikai Osztály (NyI) és BME Távközlési és Telematikai Tanszék

A kutatási partner: Technische Hochschule Darmstadt (THD)

A kutatás **célja**: jó minőségű, a beszélt nyelvi prozódiahoz közelálló német szintetikus beszéd létrehozásához szükséges prozódiai szabályrendszer kikísérletezése és megvalósítása információs rendszerekben való felhasználáshoz.

A kutatás **részletezése**: a modern automatikus információs rendszerekben egyre nagyobb probléma, hogy a nagy tömegű információból a keresett témához való hozzáférés felhasználóbarát és rugalmas legyen. A képernyő korlátozott mérete szükségszerűvé teszi, hogy bizonyos információkat ne szöveges formában, hanem beszéddel jelenítsünk meg. Ehhez – a jelen projekt keretében – a MULTIVOX beszédszintetizátor (BME fejlesztés) által előállított szintetizált beszédet használjuk, amelyhez új prozódiai modul kísérleteztünk ki. Ezt a prozódiát a szövegenerálásból származtatott jelentés-értelmező egység (THD

fejlesztése) jósolja meg és integrálja a szöveg szavai közé prozódiai "jelzések" formájában. A szintetizátor részére kialakított szöveg tehát tartalmazza a frázishatárokat, azon belül a fókuszt, valamint a frázisokra jellemző intonációs formát meghatározó jelzéseket. Ezek képezik a mondatintonáció alapját, amire, mint vázra ráültetjük a szószintű prozodiát (BME-NyI fejlesztés). Az így kialakított szintetikus beszéddel – a kísérletek szerint – jól ki lehet fejezni az egyes kifejezési, jelentési formákat (kérés, utasítás, közlés, figyelmeztetés stb.) és így a beszélt nyelvi információközlés segíti a felhasználót az információs adatbankban való keresésben. Jelenleg német nyelvre ez a rendszer képezi az egyik legfejlettebb beszéd-szintézis formát.

## VISSZANÉZŐ

Kötetünk az 1993-ban megindított *Beszéd kutatás*-sorozat harmadik tagja. Az 1994-ben megjelent második kötet végén hagyományteremtő szándékkal megjelent egy rövid összefoglaló az 1993-as (első) kötet tanulmányairól, **Szalai Enikő** tollából. Az alábbiakban az 1994-es (második) kötet tartalmát tekintjük át.

A magyar magánhangzók fonológiájának viszonylag kevésbé feldolgozott részlete a kerek ségi harmónia kérdésköre. A kötet első tanulmányában (*A kerek ségi harmóniáról*, 1–23) **Siptár Péter** ezt elemezte, felhasználva és ismertetve a modern fonológiaelmélet elvei és eljárásai közül a szigorú ciklus elvét, az alulszabottság (underspecification) és a foltozósabályok (default rules) fogalmát, valamint az autoszegmentális fonológia jólformáltsági feltételeit és az ezeket kielégítő (automatikus) terjedési konvenciók, illetve (nyelvsajátos) terjedési szabályok működését.

A csökevényes, hiányos hangképzés igen sokféle szerepet képes betölteni. Lehet a gondolkodás támogatója, az írni-olvasni tanuló gyerek segítője, a varázsló tevékenységének része; de e funkciók egyik legfontosabbika a beszéd észlelésének szolgálata. **Vértes O. András** írása (*Csőkevényes, hiányos hangképzések funkciói*, 24–32) egyebek között ez utóbbi funkció, valamint a Carpenter-féle hatás leírásával rámutatott a percepció és a motorikus tényezők általánosabb (nemcsak a beszédészlelést illető) összefüggéseire is.

A beszélőtől a hallgatóhoz érkező akusztikai jelsorozat tartalmazza mindazokat a paramétereket, amelyek a jelenté ses egységek szegmentálását és azonosítását lehetővé teszik. A beszédészlelési mechanizmus azonban olykor mégsem kap elég fogódzót a biztos működéshez: ennek jellegzetes tévedések, "félrehallások" a következményei. **Gósy Mária** tanulmányában (*A szegmentálás működése a szófelismerésben*, 33–50) arra keresett választ, hogy milyen akusztikai jegyek alapján szegmentálja a hallgató az adott hang-

sort, ha az szintaktikailag és szemantikailag kétféleképpen is azonosítható (pl. *jó zsilett/Józsi lett*). A bemutatott kísérleti eredmények a percepciók feldolgozás kontextusfüggő voltára utaltak.

A beszédészlelés és a beszédmegértés modellálása több tudományágnak, így a neurokibernetikának, a neuropszichológiának, a pszicholingvisztikának, a neurolingvisztikának is a tárgya. Ennek eredményeképpen a kutatók számos modellt alkottak róla. **Banczerowski Janusz** (*Néhány megjegyzés a beszédmegértés modellálásáról*, 51–68) áttekintette a legismertebb modelleket, figyelembe véve a köztük meglévő különbségeket és hasonlóságokat, továbbá felhívta a figyelmet e modellekkel kapcsolatos különféle problémákra. Szót ejtett az invariancia fogalmáról, a vizuális percepcióról, a fonetikai szimbolizmusról, a nyelvi redundanciáról is. Végül felvázolta a nyelv auditív rendszerének felépítésére vonatkozó elgondolását.

**Vértés Edit** írása (*Pápay József és a beszélt osztják nyelv*, 69–80) a neves finnugrista hagyatékában található, jövés-menésről, rénszarvasokról, vásárlásokról, az időjárásról, emberek hogylétéről, családjukról stb. szóló rövid beszélgetéseket, mondatokat tekintette át. Ezek a cári időkből csak tőle fennmaradt köznapi feljegyzések arra szolgálhattak, hogy segítségükkel Pápay jól megtanulhasson osztjákul.

*Megjegyzések a magyar magánhangzó-állományról* (81–93) című cikkében **Siptár Péter** az alábbi kérdésekre próbált választ találni: Hány elemből áll a magyar nyelv mögöttes magánhangzó-rendszere? Milyen fonológiai kategóriák (jegyek) szükségesek és elégségesek a magyar magánhangzók leírásához? Releváns-e fonológiaiilag a felszíni [ɔ] enyhén kerekített volta, valamint a hosszú-rövid magánhangzó párok tagjai között fennálló többi kisebb-nagyobb minőségi (nyelvállásbeli és egyéb) különbség? Milyen rendezési paradoxonba botlunk, ha feltételezzük, hogy az [ɔ:] és az [ɛ:] levezetett (nem kontrasztív) szegmentumok? Hogyan jellemezhetőek a magyar magánhangzók egyértékű jegyek segítségével?

Száz hatéves óvodással végzett kísérleteinek eredményeit összegezte **Gósy Mária** *A mondatértés és a szövegértés összefüggései* című tanulmánya (94–120). Arra a kérdésre, hogy az anyanyelv-elsajátításnak ebben a szakaszában milyen összefüggések mutathatók ki a kisgyermekek mondatértése és szövegértése között, a következő választ találta: a mondatértés lényegesen jobb, mint a szövegértés, s az előbbi a gyermekek legnagyobb részénél megfelel az életkoruk alapján tőlük elvárt teljesítmény szintjének. A szövegértés tekintetében a vizsgált hatévesek egy részénél jelentős elmaradást tapasztalt. Úgy találta, hogy az adatok alátámasztják a beszédmegértési tevékenység egyes szintjeinek viszonylagos önállóságával kapcsolatos feltevéseit.

**Osmanné Sági Judit** tanulmányában (*A beszédmegértés vizsgálata afáziás betegeknél*, 121–139) két, különböző diagnosztikai kategóriába tartozó, spontán beszédük jellegzetességei alapján élesen eltérő afáziás betegnél végzett esettanulmányainak eredményeit foglalta össze. A bemeneti lexikon önmagában mind a két betegnél épnek bizonyult, míg a beszédhangok feldolgozása különböző mértékben volt károsodott. A Wernicke-afáziás betegnél a mássalhangzó-megkülönböztetés zavara kihatott ugyan a lexikális azonosítás folyamatára, ám ezt a deficitet sikeresen kompenzálta a szójelentések épen maradt ismerete. A szavak jelentésmezéjében végzett kognitív műveletek viszont a másik, Broca-afáziás betegnél bizonyultak károsodottnak. Az adatok azt mutatták, hogy a funkcionális beszéd-megértés, a beszédhang-percepció, valamint a szójelentések ismeretének épsége az afáziás betegeknél nem feltétlenül korrelál.

A folyamatos beszéd időszerkezetének szegmentális szintű vizsgálatáról szólt **Olaszy Gábor** tanulmánya (*Hangidőtartam-módosító kísérletek a gépi beszéd ritmusának javítására*, 140–151), amelyben a szerző az egyes beszédhangok adott tempóhoz tartozó specifikus időtartamaira vetítve jellemezte az egyszerű kijelentő mondatok beszédhangjainak időtartamát.

A számítógépes beszédelőállítás során a grafémák fonémákká történő leképezéséhez, amely a tényleges beszéd szintézis kiindulópontja, rendszerint nem elegendők a fonetikai és fonológiai szabályok, mert bizonyos esetekben a magasabb nyelvi szintek: a morfológiai, a szintaktikai vagy a szemantikai szerkezet tulajdonságai határozzák meg a helyes kiejtést, a megfelelő prozódia kialakításához pedig a mondatok szerkezetének felismerésén túl sokszor szemantikai és pragmatikai ismeretek is szükségesek. **Koutny Iona** írása (*Nyelvelemzés a számítógépes beszédelőállításban*, 152–163) a szótár nélküli szintézisre összpontosítva körvonalazta, hogy hol a helye a nyelvi elemzésnek a beszédelőállításban és milyen feladatokat kell teljesítenie.

**Vicsi Klára, Vig Attila és Berényi Péter** (*Magyar nyelvű folyamatos beszéd gépi felismerése akusztikai–fonetikai–fonológiai szinten*, 164–182) bemutatta szövegtől független, folyamatos beszédet felismerő számítógépes rendszerét, valamint az ennek létrehozásához szükséges előzetes statisztikai vizsgálatokat. Ez utóbbiak eredményeként a felismerés alapegységül a félszótagot választották. Rendszerük több szintből épül fel; a feldolgozási módszer szintenként más és más. Egészében véve különböző szerkezetű neurális hálózatoknak és a szabály-alapú feldolgozásnak mindig az adott szinten elvégzendő feladathoz illeszkedő kombinációja. A rendszer kimenete egy félszótaglánc-sorozat, amely a felismerési valószínűségektől függően első-, másod- és harmadrendű jelölteket tüntet fel.

A modern információs rendszerek nyelvi moduljaiban egyre fontosabb szerepet kapnak az egyes nyelvek fonetikai alakjának leírásával kapcsolatban végzett kutatások. Az információt ma legnagyobbbrészt írott formában tárolják, de sok esetben beszéd formájában közvetítik a felhasználónak. E két forma összekapcsoló eleme az írott szöveget fonetikai reprezentációvá (átírássá) alakító eljárás. Az átalakítás nehézségi foka nyelvenként különböző, attól függően, hogy az adott nyelv helyesírása milyen mérték-

ben valósítja meg a betűkép–hangkép megfeleltetést. Ilyen szempontból az angol az egyik legkevésbé fonetikus helyesírású nyelv. **Olaszi Péter** dolgozata (*Számítógépes algoritmus angol szöveg fonetikus átalakítására*, 183–197) egy egyedi fejlesztésű automatikus átírási eljárást mutatott be a brit angolra.

**Lengyel Zsolt** tanulmánya (*Az írott szöveg ontogenezisének vizsgálata: szegmentálási és pragmatikai kérdések*, 198–210) tíz éves magyar gyerekek írásos nyelvhasználatának néhány, a szöveg tagolását érintő sajátosságát tekintette át. A szövegek mondatokra osztását, majd a mondatokon belül a szavak egymástól való elhatárolását, végül pedig az írni tudás egyedi fejlődésének egyes pragmatikai kérdéseit vizsgálta.

A beszédbeli kommunikációra nem képes fogyatékos gyermekek oktatásában, rehabilitációjában nagy szerepet játszik a több évtizede kialakított Bliss-jelképrendszer használata. **Olaszy Gábor**, **Kálmán Zsófia** és **Olaszi Péter** tanulmánya (*A BLISSVOX beszélő kommunikációs rendszer*, 228–236) egy új számítógépes segédeszközt mutatott be, amelynek felhasználásával a jelképrendszert ismerő személy a Bliss-jelképekből összeállított üzenetet beszédszintetizátor segítségével meg is tudja szólaltatni. Így az ilyen fogyatékos emberek számára is lehetségessé válik a verbális kommunikáció.

A kötetet **Tarnóczy Tamás** ismertetése (211–227) egészíti ki *Peter B. Denes és Elliot N. Pinson: The Speech Chain. The Physics and Biology of Spoken Language* [A beszédlánc. A beszélt nyelv fizikája és biológiája] című könyvének nemrég megjelent második kiadásáról (New York, 1993).

**Összeállította: Siptár Péter**





## MEGRENDELÉS

Megrendelem az alábbi példányokat a Beszédkutatás sorozatból:

Beszédkutatás 1995 kötetet ..... példányban .....x 1500 Ft , azaz..... Ft értékben  
Beszédkutatás 1994 kötetet ..... példányban .....x 350 Ft , azaz .....Ft értékben  
Beszédkutatás 1993 kötetet ..... példányban .....x 350 Ft, azaz .....Ft értékben

A fizetés módja: átutalás  postai utánvét  egyéb:

A kézbesítés módja: postai  személyes átvétel  egyéb:

Postai kézbesítés esetén a postaköltséget a vételárral egyidőben kiegyenlítem.

A megrendelő neve:.....

A-megrendelő címe:..... Ir. szám:.....

Az ügyintéző neve:..... Telefon:.....

Ezt a megrendelést a következő címre kérjük elküldeni:

MTA Nyelvtudományi Intézet, Fonetikai Laboratórium

Dr. Gósy Mária

Budapest, 1250 Pf. 19.



Postacím:

MTA Nyelvtudományi Intézete  
Budapest, 1250 Pf. 19.  
T: 1758-011

Információ: dr. Gósy Mária  
MTA Fonetikai Laboratórium  
1014 Budapest,  
Szentháromság tér 6. mfszt. 61.  
T: 1557-122 / 9218-as mellék  
Fax: 212-2050  
e-mail: [gosity@nytud.hu](mailto:gosity@nytud.hu)





