

A MŰSZERES SZEMÉLYAZONOSÍTÁS LEHETŐSÉGEI RÖVID IDŐTARTAMÚ BESZÉDMINTÁK ALAPJÁN

Nikléczy Péter

Bevezetés

A Kempelen Farkas Beszédkutató Laboratóriumban már több éve folynak olyan kutatások amelyek célja a rögzített beszéd azonosítása a hanganyagot bemondó személlyel (vö. Nikléczy 1996, Gósy 1999). Ennek keretében – a hangfelvételek technikai minősége alapján – az alábbi vizsgálatokat végeztük el:

1. A beszélő személy felismerésének korlátai normál, torzítatlan beszéd esetében,
2. A beszélő személy felismerésének korlátai torzított beszéd esetében,
3. A minta időtartamának szerepe a személy azonosításában,
4. Az elemző berendezések technikai paramétereinek szerepe a személyazonosításban,
5. Az automatizált műszeres és az emberi közreműködéssel végzett (manuális) azonosítás alkalmazásának feltételei.

Az első feladat az volt, hogy a stúdióminőségű hanganyagon kijelöljük azokat a fogódzókat, amik lehetővé teszik, hogy nagy valószínűséggel azonosnak vagy nem azonosnak ítéljük meg a hanganyag alapján a kérdéses személyt. A műszeres személyazonosítás egyik problémája, hogy a különböző beszélők hangjáról készült regisztrátumok között csak ritkán találunk lényeges eltérést. Ez azt jelenti, hogy ha az ugyanazon nyelvet beszélő személyek hangja között nagyfokú hasonlóság van, akkor a személy azonosítását műszeres elemzéssel nem vagy csak hosszabb időtartamú hanganyag felhasználásával tudjuk meghatározni. Ezért minden esetben döntő szerepe volt a hangfelvétel időtartamának. A következőkben azokat az eredményeket, illetve elemzéses eljárásokat foglaljuk össze, amelyek lehetővé teszik a be-

szelő személy azonosítását manuális úton, rövidebb időtartamú hanganyagok esetén is a KAY CSL 4300B típusú digitális jelfeldolgozó segítségével.

A beszélő személy felismerése az ember számára a mindennapi életben nem okoz gondot, ha közeli ismerősünket halljuk, vagy ha az adott személy hangját egy-két alkalommal már agyunkban rögzítettük. Igaz ez akkor is, ha az a személy rádión keresztül hallható, és a hangjához nem kapcsolódik az illető vizuális megjelenése. Nem csak egy-két órán keresztül tudjuk azonosítani a személyt a hangja alapján, hanem ez a hangkép hosszú időre eltárolódik az agyunkban, és azt alkalomadtán elő tudjuk hívni. Tehát nemcsak arra vagyunk képesek, hogy a megfelelő, már tárolt mintát összehasonlítsuk egy hasonló akusztikai körülmények között kapott újabb mintával, hanem arra is, hogy egy korábban hallott minta szerint következtessünk a beszélő személy azonosságára. Ilyen eset az, ha az illető más hangsorokat ejt ki, ha a korábban kapott minta magyar hangsorokat tartalmazott, míg az újabb minta a kérdéses személytől például angol vagy német nyelven hallható, vagy az akusztikai körülmények teljesen mások (például szabad téren elhangzó beszéd). Hasonló a helyzet akkor is, ha a kérdéses személy hangját telefonon keresztül halljuk.

A modern technika már lehetővé teszi, hogy nagy mennyiségű adatot tároljunk, és ezt a tárolt adathalmazt az agyhoz hasonlóan elő tudjuk hívni. Felmerül az a kérdés, hogy a) az egyes személyeket fel lehet-e ismerni gépi úton, b) a gép hány ember között tud különbséget tenni, c) ez a különbségtétel milyen biztonsággal működik. Pillanatnyi tudásunk szerint kevés számú személy esetében, ha az adott hangminta is megfelelő műszeres elemzésre, akkor az azonosítás viszonylag nagy biztonsággal automatikusan is elvégezhető. Kérdéses azonban az, hogy még egy viszonylag kis létszámú csoport esetén is az összehasonlítandó paraméterek mennyire tekinthetők állandónak. Vajon az etalonnak számító minta milyen hosszú ideig tekinthető elfogadhatónak, hiszen az idő elteltével a beszélő személy hangja is változhat? Hozzájárul ezekhez a más akusztikai körülmények között rögzített hangfelvételek problémája is. Mindezek nemcsak az automatikus személyazonosítás megválaszolandó problémái, hanem a főleg a kri-

minalisztikában alkalmazott „manuális” meghatározásáé is. A gyakorlat azt mutatja (főleg a kriminalisztikai esetekben), hogy az összehasonlítható hangminták sem akusztikailag, sem időtartamban nem felelnek meg az automatikusan elvégezhető azonosítás követelményeinek. A célszemélyt a legtöbb esetben csak később, más technikai körülmények között rögzített hangfelvételek felhasználásával kell kiválasztani. Kijelenthetjük, hogy olyan eljárást kidolgozni, ami minden körülmények között nagy biztonsággal automatikusan meghatározza a személy azonosságát, jelenleg nem lehetséges. Ezért továbbra is elkerülhetetlenül szükség lesz az egyedi eseteket elbíráló személyazonosító eljárásra. A következőkben, példák felhasználásával, azokat a lehetőségeket fogjuk ismertetni, amelyek hozzájárulhatnak a személy azonosításához akkor is, ha az utólag készült hangfelvétel több tekintetben is jelentősen eltér az összehasonlítható hanganyagtól.

Azt, hogy milyen fizikai paraméterek alapján különítjük el az egyes hangokat, a fonetikával, beszédakusztikával foglalkozó szakemberek már az ötvenes évek végén kielemezték. Azonban arra a kérdésre, hogy milyen hangösszetevők segítik a hallgató személyt abban, hogy a már korábban kapott hangminták alapján a beszélő személy azonosságát megállapítsa csak az utóbbi és az elkövetkező évek kutatásai adtak, illetve adhatnak felvilágosítást. Már korábban kialakult az a nézet, hogy az azonosítás egyik fogódzója lehet a hangok formánstruktúrájának elhelyezkedése, a beszéd alaphangjának értéke, a beszéd időszerkezete, esetlegesen a hangsorok viszonyított intenzitásértékei, valamint a hangsorokban elkövetett kiejtési hibák stb. Ha ezeket a tényezőket külön-külön megvizsgáljuk, kiderül, hogy mindegyik hordoz magában olyan variációt is, ami nemcsak, hogy nem segíti elő a beszélő személy azonosítását, hanem tévesztésre ad lehetőséget. Ezek közé sorolhatjuk azokat az eseteket is, amikor az etalonnak számító mintában a vizsgált hangot, pillanatnyi kiejtési hiba következtében, az adott személy nem a tőle elvárható módon ejti. Így a többi, utólagosan rögzített hangmintát esetleg emiatt nem azonosítjuk az ő kiejtésével. Ilyen esetnek számít az is, amikor a fonéma nem a szokásos beszédhangban realizálódik. Sokszor előfordul, hogy [o] hang helyett az illető inkább [u]-t ejtett, ez azonban a hangsorban nem ve-

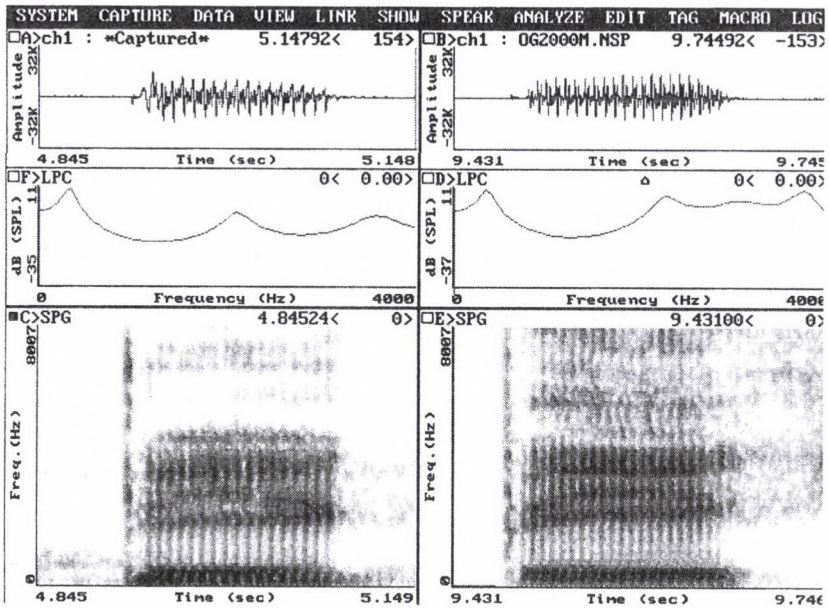
hető észre. Tévesztésre adhat okot, ha a beszéd átlag-alaphangértékét mérjük, és ennek alapján próbáljuk a személyazonosságot meghatározni. A beszéd dallamának vizsgálatánál bizonyos egyénre jellemző motorikus beidegződésekkel találkozhatunk, de a rendkívüli változottság miatt, ez sem tekinthető a megoldás alapjának.

A formánsok elhelyezkedése és sáv szélessége

A korábbi – már említett – kutatások és hangvizsgálatok megállapították, hogy az egyes beszédhangoknak jellemző formánsstruktúrájuk van, ami meghatározza a karakterüket. A hangra jellemző formáns szerkezet azonban nem egy-egy megadott értéken realizálható, hanem értékhatárok között (vö. Szalai 1995, Kovács 2001). Az [a:] például férfihang esetében F1 650-900 Hz-ig, F2 1250-1450 Hz-ig, F3 1950-2250 Hz-ig változhat. Ez azt jelenti, hogy ugyanaz a hang realizálódhat például 800 Hz (F1), 1400 Hz (F2), 2400 Hz-en (F3) és megjelenhet 900 Hz, 1350 Hz és 2350 Hz-es értékben is. A jelzett változás az F1-nél 11%, az F2-nél 4% az F3-nál 2%. Ez a változás nemcsak az egyes beszélők közötti különbséget jelentheti, hanem például az adott személy kiejtésében előforduló eltérésekre is utalhat.

Mivel a formánsok a hangszalagok által keltett összetett periodikus rezgés felhangjainak a toldalékcsöben felerősödött részei, attól függően, hogy a személy a hangot milyen alaphangmagasságon ejtette, a rezonanciaértékek bizonyos mértékben változnak. Általában kimondható, hogy az alaphang változása a formánsok értékeit a maximum- vagy minimumérték felső pontjáig tolhatja el ugyanazon beszélő személy esetében is. A fentiekből következik, hogy a formánsértékek pontos meghatározása önmagában kevés, a rendelkezésre álló rövid minta esetén nem adhat a személy kilétéről pontos információt. Természetesen minden esetben nem lehet a formánsokat kizárni a személy azonosításánál. Ha egy hangot az adott személy konzekvensen, mindkét mintában tévesen ejt, vagy az artikulációja esetleg nyelvjárássra visszavezethető, akkor a hang vagy hangok mért formánsstruktúrája is bizonyító erőt képviselhet még viszonylag kevés számú minta esetében is. Ilyen lehet a nyílt és zárt [e] hang vagy a magyar hangrendszerrel jellegzetesen eltérő kiejtés.

Az érték szerinti meghatározás mellett a formánsok fontos velejárója a sávzélesség is. Ennek kimutatása és értékének pontos meghatározása csak a jó minőségű felvételek esetében lehetséges. Az adott formánshely sávzélességének értéke hozzásegíti az elemző személyt az azonosítás elvégzéséhez: az eddigi vizsgálatok szignifikáns összefüggést mutattak ki a személyre vonatkozóan. Az 1. ábrán két beszélőtől rögzített minta látható, amelyek formánshelyei kismértékben, a formáns-sávzélességei azonban nagymértékben eltérnek egymástól (vö.1. táblázat).



1. ábra

A formáns-sávzélesség mint az egyéni ejtőmód tényezője

I. táblázat: formáns sáv szélességek alakulása

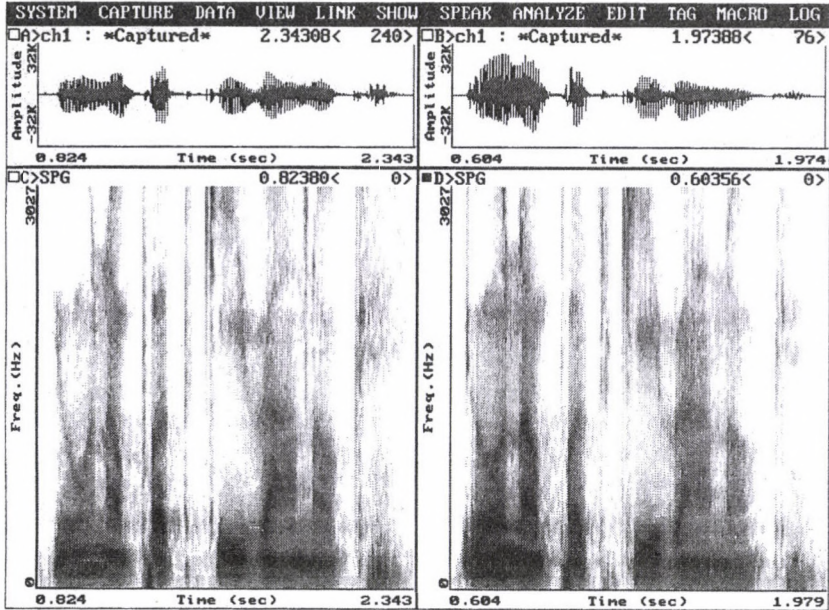
| Minta | I. minta | | II. minta | |
|---------|----------|---------------|-----------|---------------|
| | formáns | sáv szélesség | formáns | sáv szélesség |
| F1 (Hz) | 342 | 91 | 357 | 109 |
| F2 (Hz) | 2207 | 287 | 2274 | 217 |
| F3 (Hz) | 3167 | 1612 | 3042 | 490 |

A táblázatban numerikus adatok szemléltetik a különbségeket (csak az első három formáns értékét és sáv szélességét adtuk meg). Természetesen a jó minőségű hangfelvételek velejárója az értékelhető magasabb frekvenciatartomány is. A bemutatott ábra lényeges különbségeket a magasabb tartományokban is kimutat. Míg az I. mintán 3 kHz felett alig látunk a spektrogramon feketedést (tehát értékelhető frekvenciatartományt), addig a II. mintán ez sokkal intenzívebben jelenik meg.

A következőkben olyan eseteket mutatunk be, amikor a beszélő személy azonosságát spektrográfias eljárással a formánsstruktúra elemzésével próbáljuk meghatározni. Alig egy évtizede ezekhez a vizsgálatokhoz csak a nehezebben kezelhető VOICEPRINT 7000-as típusú hangspektrográf állt rendelkezésre. A spektrográffal készített hangszínekép értékelését megnehezítette, hogy csak két beépített sávszűrővel lehetett a formánsstruktúrát kiírni. Ez a technikai probléma eleve kizárta azt a lehetőséget, hogy az adott hang formánsokon belüli felhangszerkezetét elemezzük, hiszen annak meghatározásához a sávszűrő pontos értékének kiválasztása nagyon fontos. A formánsstruktúra egybevetését kizárólag a gép által hőérzékeny papírra megrajzolt adott sávszűrővel készített hangszínekép alapján végezhetjük el. A digitális jelfeldolgozóval azonban lehetőség nyílt az elemző berendezés paramétereinek tág határok közötti változtatására. A 2. ábrán különböző időpontban rögzített, ugyanazon beszélő személytől származó, ugyanazon szóról készült spektrogram látható. A manuálisan elvégzett mérések eredményeinek számszerű összehasonlításából következtethető ugyan azonosság, de ez mégsem mondható megalapozottnak. Azért nem tekinthetjük az előbbi hasonlóságot személyre vetíthető azonos-

ságnak, mert a mért formánsértékek más személynél is előfordulhatnak.

A 3. ábra képein látható, hogy az előbbi azonosság formánsszinten is kimutatható, pedig a beszélő személy nem azonos az előbbivel. A példa azt szemlélteti, hogy a legtöbb esetben sokkal finomabb elemzés szükséges a pontos azonosításhoz, mint a formánsszerkezeti adatolás.

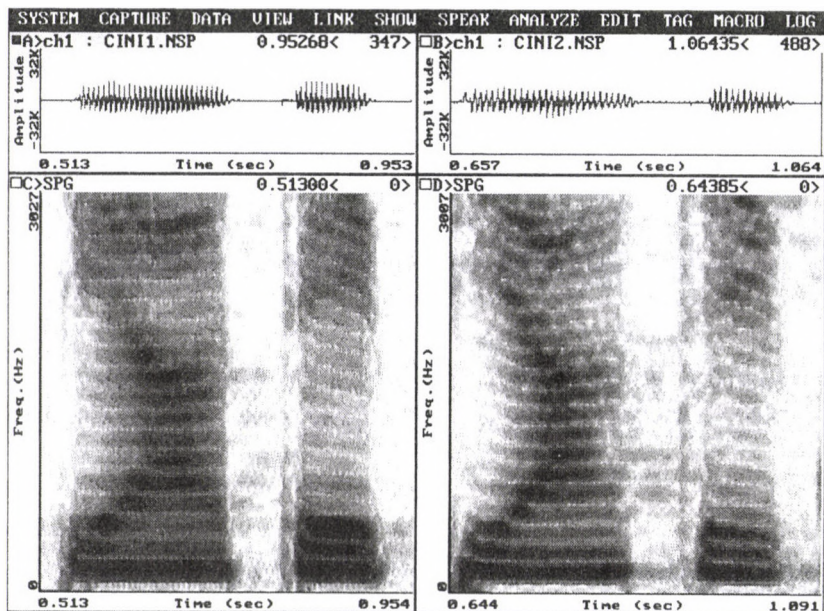


2. ábra

Ugyanazon beszélő két különböző időpontban rögzített ejtéséről készült hangszinkép

A zöngés beszédhangok formánsszerkezetét nemcsak a már korábban említett három formáns alkotja. A technikailag jó minőségű felvételeken kimutatható a jelen lévő többi formáns is. Ezeknek a formánsoknak a megjelenítése azonban csak ritka esetben lehetséges anélkül, hogy a magasabb tartomány intenzitását ne emeljük meg. A

modern technikai elemző berendezések lehetőséget adnak arra, hogy a felsőbb formánsstruktúrát szűréses eljárással felerősítsük, illetve azt vizuálisan megjeleníthessük. Ilyen esetekben azonban már figyelemmel kell lennünk a hangfelvételeken lévő zaj hatására is, mert a felerősített zaj elfedheti vagy felerősítheti a formánst.



3. ábra

Két különböző beszélő ugyanazon hangsor ejtéséről készült hangszíneképe is nagy hasonlóságot mutat

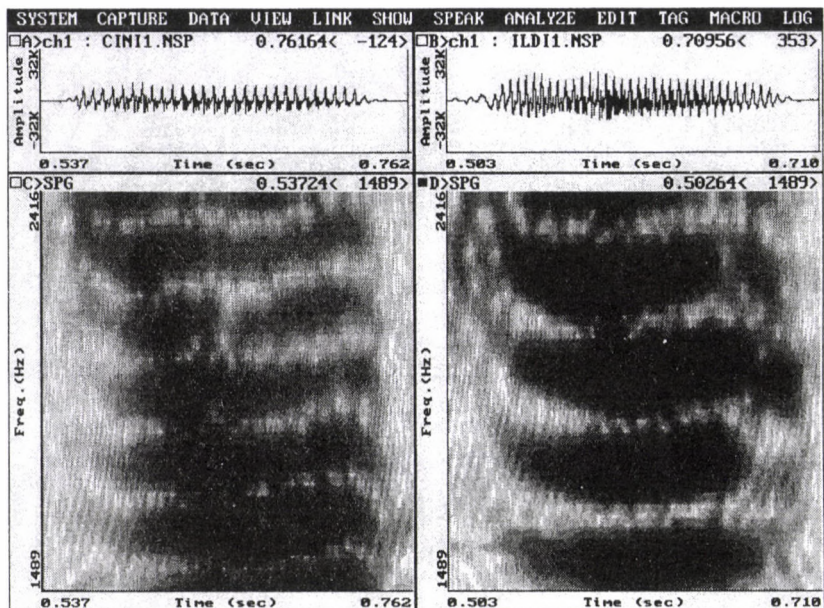
A beszéd felhangstruktúrája

Az előzőekben már utaltunk arra, hogy a vizsgálatoknak ki kell terjedniük a felhangokra is. Miért érdekes a beszéd felhangstruktúrája? Azért, mert a felhangstruktúra változása nagymértékben függ az egyén artikulációs csatornájának beszéd közbeni változásától, és ez a változás jobban tükrözi az egyéni jellemzőket, mint a formánsstruktúra. Természetesen nemcsak a frekvenciaérték szerinti elhelyezkedésük

érdekes – hiszen az az alaphang értékének egész számú többszöröse –, hanem az is, hogy az adott helyen milyen intenzitásértékben vagy az intenzitástól függetlenül milyen képi forma szerint változnak a hangsorokban. Ez utóbbi főleg a közepes frekvenciatartományban (1500-3500 Hz között) tartalmaz értékelhetően egyéni jellemzőket. A hangok képzésekor a hangszalagok által gerjesztett összetett hangrezgésből a hangszalagok előtt és után következő rész mint rezonátorüreg kiemeli, illetve felerősíti a felhangokat, tehát a felhangstruktúra az egyéni hangszínezet kialakításában lényeges szerepet tölt be. A hangszalagok utáni rész mint rezonátorüreg nem tekinthető állandónak, hiszen a hangok kialakítása közben változik: szinte biztosra vehető, hogy egy megelőző állapot aligha realizálódik újra. Ez már a formánsok esetében látható is volt, hiszen azonos hangok esetében sem kapunk kétszer megegyező értéket. Az viszont biztosra vehető, hogy a beszélő, önmagához képest is eltérő kiejtése esetén is, felismerhető. Ez arra vezethető vissza, hogy az artikulációs folyamatokkal nem változtatjuk meg lényegesen a rezonátorüreg egyénre jellemző tulajdonságait, ugyanis az nemcsak a toldalékcsofótól, hanem a hangszalagok előtti üregrendszerrel és a tudatosan nem változtatható orrüregtől is függ. Ennek következtében a személyre jellemző orgánum a felhangstruktúrában jelen kell, hogy legyen. Ennek a kimutatása bizonyos fokig lehetséges is. A spektrografikus eljárás lehetővé teszi a felhangok kiíratását akár 20 Hz-es szűrő-sávszélességgel is. Így jól tudjuk vizsgálni a felhangstruktúra változását mind intenzitásértékében, mind a frekvenciaértékben.

A 4. ábrán két keskeny sávú spektrogram látható. Mind a két esetben 35 Hz-es szűrőbeállítást alkalmaztunk. Az elemzett frekvenciasáv 1489 Hz-től 2416 Hz-ig terjedt. A hangszínképeken látható a két személy hangja közötti különbség, annak ellenére, hogy ugyanazon hangsor azonos hangját elemeztük.

Az elemzéseknél figyelembe kell venni, hogy a nagyobb sávszélességgel megjelenő formánsok esetében csak megfelelő digitális szűrők alkalmazásával lehet a kívánt eredményt elérni.

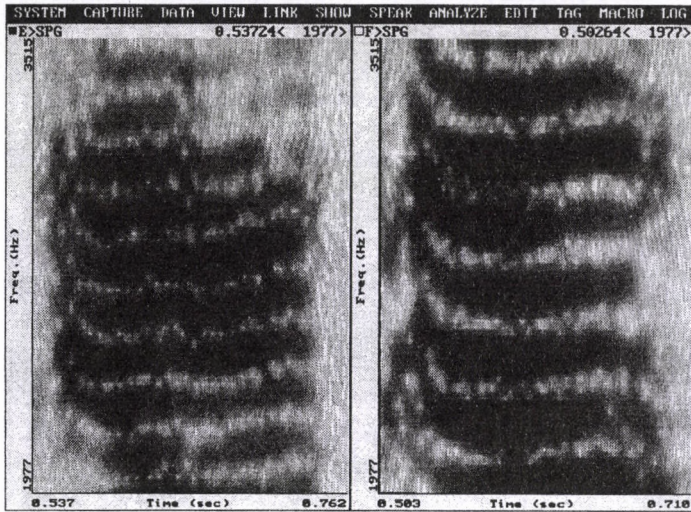


4. ábra

Két beszélő ugyanazon hangsorából izolált beszédhangjának felhangszerkezete

Mint már korábban említettük, a felhangok az alaphang frekvenciaértékének egész számú többszörösénél jelennek meg. A megjelenési forma azonban oly mértékben változhat, hogy bizonyos frekvenciasávokban intenzívebben, másutt kevésbé intenzíven jelennek meg az adott felhangok. (Ezek az intenzívebb részek természetesen nem a formánshelyek.) Az intenzív, illetve kevésbé intenzív megjelenés mellett még az is különböző, hogy a felhangon belül milyen intenzitású góccok rajzolódnak ki. Az adott hang kiejtésénél a felhangon belül intenzív vagy kevésbé intenzív góccok rajzolódhatnak ki. Az 5. ábrán látható a két beszélőtől rögzített ugyanazon hangsorból szegmentált hang felhangszerkezetének lényeges különbsége az 1977-3515 Hz-es frekvenciatartományban. Míg a bal oldali spektrogramon a góccok

elmosódottabbak, addig a jobb oldalin intenzívebbek, egyes felhangoknál dupla maggal jelennek meg.

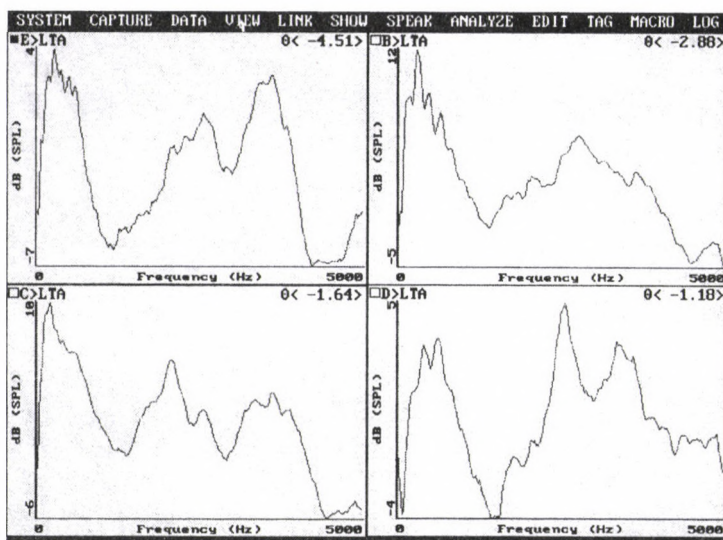


5. ábra

A mondatból izolált azonos hangok felhangszerkezetének egyéntől függő különbsége

A fenti ábra a zöngés beszédhang felhangstruktúráját érzékelteti, lehetőséget adva az eltérések képi vizsgálatára. Az elemzések során azonban szükség van arra is, hogy egy adott hang bizonyos időtartamú részletének átlagos energiaspektrumát ellenőrizzük. Ez megmutatja, hogy egy hang az adott intervallumban milyen energiamaximumokat, illetve energiaminimumokat tartalmaz a frekvencia függvényében. A kapott görbe visszavezethető a toldalékcso átviteli karakterisztikájára, ami egyéni jellemzőket mutat. A fent említett időintervallum több hangot is magában foglalhat. Ebben az esetben a az elemzett hangok átlagának energiaspektrumát kapjuk. A 6. ábrán az izoláltan ejtett magyar magánhangzók FFT-analízisen alapuló átlag-energiaspektruma látható négy személy ejtésében. A görbék jól mutatják, hogy az

egyedek ejtése között lényeges eltérés van. Mivel minden magánhangzó tartalmazza az egyéni tulajdonságokat, feltételezésünk szerint az átlagnak is tartalmaznia kell azokat. Kérdés az, hogy a kapott görbevonulat mennyire mondható az egyénre nézve állandónak, illetve jobban tükrözi-e az egyéni hangsjátosságokat, mint a hangonként készített energiagörbe. Elsőként azt vizsgáltuk meg, hogy ezek a görbék mennyire azonosak azokkal, amelyeket az illető személytől rögzített szövegből izolált magánhangzókról kaptunk. Lényeges szempont az, hogy az összehasonlítandó hanganyag is azonos számú és azonos karakterű hangokat tartalmazzon.

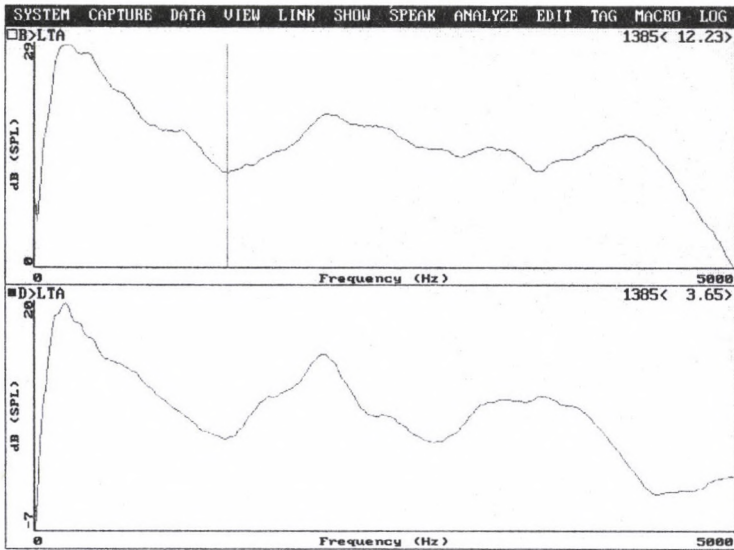


6. ábra

A magyar magánhangzók átlagspektruma négy személy ejtésében

A 7. ábra felső része az izoláltan ejtett magánhangzók teljesítményspektrumát mutatja, míg az alsó rész az ugyanazon személy ejtésében elhangzó spontán beszédből izolált magánhangzók átlag-energiáspektrumát mutatják. A görbe vonulatából kitűnik, hogy körülbelül 3 kHz-ig a minimum és maximumpontok megegyeznek, és

csak a fölött van lényegesebb eltérés. Az azonban itt is látható, hogy a 3 kHz feletti rész is lényegesen eltér az 5. ábrán látható másik három spektrum hasonló szakaszától. Az elemzéseket 0-5 kHz-es tartományban végeztük.

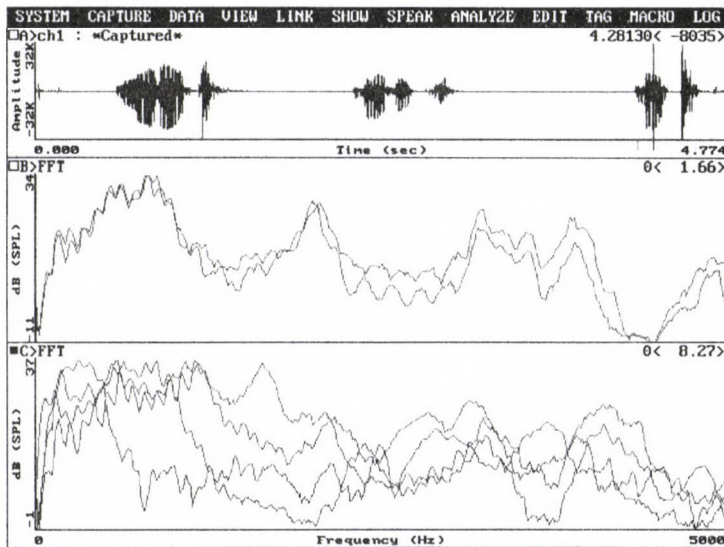


7. ábra.

Izoláltan ejtett magánhangzók (felső ábra), és az ugyanazon személy ejtésében elhangzó hangsorból szegmentált magánhangzók átlag-energiáspektruma

Megnéztük, hogy milyen hasonlóság mutatható ki, ha ugyanezt a vizsgálatot izoláltan ejtett magánhangzóval végezzük el, és ezt hasonlítjuk össze a spontán szövegben elhangzó ugyanolyan magánhangzóval. Az eredményt a 8. ábra mutatja. Az ábra felső részében az [ɔ] magánhangzó tiszta fázisának két 10 ms-os átlag-energiáspektrumát rajzoltattuk ki. A két görbe vonulata láthatóan eltér egymástól. Az ábra alsó része pedig a hangsor azonos magánhangzóinak tiszta fázisá-

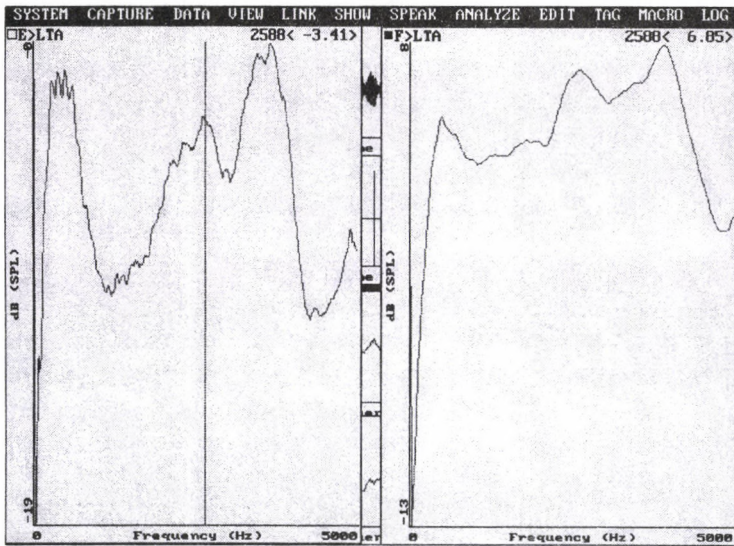
ból vett energiaspektrumokat mutatja. Szembetűnik a görbék maximum- és minimumpontjainak lényeges különbsége. Ez annak tulajdonítható, hogy a már korábban említett formánshely-változások egy hang vizsgálata esetén jobban befolyásolhatják az energiaspektrum eloszlását. Előfordulhat olyan eset, amikor az adott hanganyagban nincs meg minden magánhangzó. Kísérleteink során azt tapasztaltuk, hogy egy vagy két hang hiányában számolnunk kell az átlagértékek bizonyos fokú romlásával. (Ebben az esetben természetesen az izoláltan ejtett hangorból is ki kell emelni a hiányzó hangot.) Ez a romlás azonban lényegesen nagyobb azoknál a domináns hangoknál, amelyek tapasztalataink szerint talán a legjobban viselik magukon az egyéni hangszínezetet (például [a:], [ɔ], [e:]).



8. ábra

Azonos magánhangzók energiaspektrumának szórása izolált ejtésben (felső ábrarész) és hangorból szegmentálva (alsó ábrarész)

Azt is megnéztük, hogy mi a helyzet abban az esetben, ha a folyamatos beszédből nem szegmentáljuk a magánhangzókat, hanem az elhangzó szöveg teljes időtartamára elvégezzük az analízist. Meg kell jegyezni azonban, hogy olyan mondatot választottunk, ahol az egyes magánhangzók csak egyszer fordulnak elő. Az eredményt a 9. ábra mutatja.

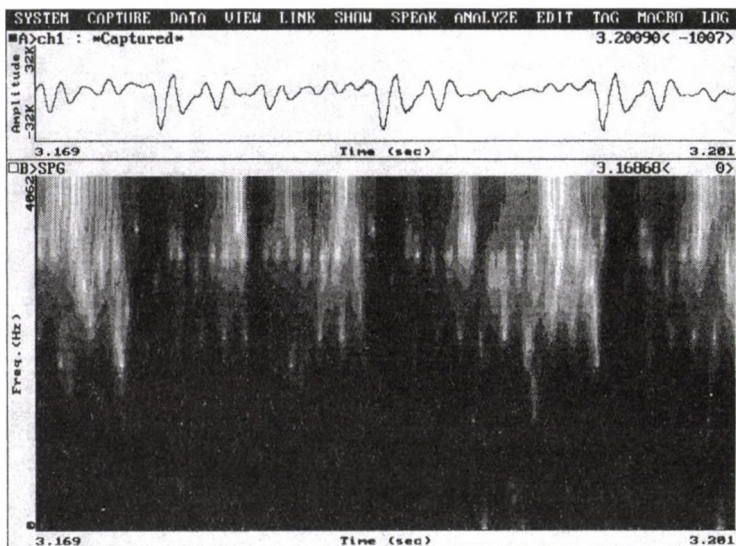


9. ábra

Az azonos beszélőtől rögzített izoláltan ejtett magánhangzók és egy mondat átlag energiaspektruma

A középső frekvenciatartományban a 2588 Hz-nél lévő maximumpont azonos, de az alacsonyabb és a magasabb tartományban lényeges az eltérés. Minden olyan esetben, amikor a hangfelvétel csak egy szűk frekvenciasávban analizálható (például telefonvonalról készített hangfelvétel) vagy zajjal erősen fedett, akkor ezt az eredmények kiértékelésénél figyelembe kell venni.

A részletes vizsgálathoz hozzátartozik még a felhangszerkezet periódusonkénti vizsgálata is. Ez azt jelenti, hogy a zöngés hangok periódusainak belső eloszlását vizsgáljuk 0-5 kHz-es tartományban. A digitális jelfeldolgozó lehetővé teszi, hogy akár 850 Hz-es sávszélességű szűrővel analizáljuk az adott hang egy vagy két periódusát. Az így kirajzolt spektrogram különböző színárnyalatokra, illetve színekre lebontva, képet ad a perióduson belül lezajló energia eloszlásról. A 10. ábra az [a:] hang három teljes periódusának rajzolata. Az alkalmazott sávszűrő 850 Hz volt.



10. ábra

Az [a:] hang három periódusának spektrogramja 850 Hz-es sávszélességű analízálási szűrőállással

Ezek a periódusonkénti energia eloszlások tapasztalataink szerint a beszélő személytől függően más-más értékeket mutatnak.

A fent leírt vizsgálati lehetőségek egy átfogó, minden részletre kiterjedő elemző eljárás részei. A beszélő személy azonosításában meg-

határozó a vizsgálat pontos megtervezése az adott hanganyagoknak és az adott célnak megfelelően. Fontos, hogy az elemzést az illető személy hangjának legmegfelelőbb beállítási értékekkel végezzük. A mindenkori lehetőségek figyelembevételével az alábbi sorrend (protokoll) ajánlott:

1. A hanganyag akusztikus meghallgatása:
 - a) szubjektív minősítés,
 - b) szükség szerinti átírása (rögzítése),
2. A szükséges bevételi mintavételezési érték kiválasztása.
3. A megfelelő hangrészletek kiválasztása.
4. Az elemző eljárás(ok) meghatározása:
 - a) hangszínekép-elemzés,
 - b) a formánsok sáv szélességének mérése,
 - c) LPC-analízis,
 - d) FFT-analízis,
 - e) CEPSTRUM-analízis,
 - f) időszerkezeti elemzések,
 - g) a beszéd dallamának vizsgálata,
 - h) a beszéd átlagos alaphangértékének meghatározása,
 - i) a beszéd intenzitás szerkezetének vizsgálata,
5. Esetleges percepciók tesztelés.

A leírtak természetesen mutatják a problémákat. A beszélő személy azonosításának növekvő valószínűségét a protokoll betartása és a fennálló problémák megoldása jelentheti.

Irodalom

- Gósy Mária (1999): A beszédprodukciónak tudatos változtatása: a beszélő személy utánzása. In: Beszédkutatás '99. Szerk.: Gósy Mária. MTA Nyelvtudományi Intézet. Budapest, 53-68.
- Nikléczy Péter (1996): A beszélő személy azonosítása szűk frekvenciás szavak alapján. In: Beszédkutatás '96. Szerk.: Gósy Mária. MTA Nyelvtudományi Intézet. Budapest, 20-31.

Kovács Magdolna (2001): Tendenciák és szabályszerűségek a magánhangzó-időtartamok produkciójában és percepciójában. Debreceni Egyetem. Debrecen, (PhD-disszertáció).

Szalai Enikő (1995): Az [u:], [a:], [i] koartikulációs mezőiről. In: Beszédkutatás '95. Szerk.: Gósy Mária. MTA Nyelvtudományi Intézet. Budapest, 83-93.

A kutatás a T025965 sz. OTKA-munkálat keretében folyt.