

A genetikai összetevő a beszélő személy beszéd-sajátosságáiban

Bevezetés

A beszélő személy azonosítása tudományterület szempontjából az alkalmazott fonetika egyik legújabb ágához, a törvényszéki fonetikához tartozik. Célja: az emberi hang egyéni jellemzőinek kutatása; a beszélő személy egyértelmű felismerése vagy kizárása a beszéd akusztikai-fonetikai vizsgálata alapján. A nem kriminalisztikai célú alkalmazásokban, a biztonsági, beléptető vagy banki azonosító rendszerek esetében a beszélőfelismerés problémája megoldottnak tűnik; viszont a törvényszéki beszélőfelismerés, azaz a gyanúsított azonosítása a beszéde alapján, még sok megoldatlan problémát tartalmaz (Gósy–Nikléczy 1999: 2–3). A rövid hanganyag, a torzított beszéd, a zajos környezet, a szűk frekvenciatartomány vagy az átmeneti beszédzavar problémáin túl a beszéd genetikai tényezőjének kérdése tágabb értelemben szintén a beszélő személy beszéd alapján történő felismeréséhez kapcsolódik, hiszen minél többet tudunk a genetikai összetevőről, annál biztosabb az azonosítás. Ezen kívül olyan konkrét esetek is előfordulhatnak, amikor például azonos nemű testvéreket kell a hangjuk alapján megkülönböztetni. Általános probléma, hogy a tulajdonképpen beszédet eredményező akusztikai rezgések nem teljesen szabályosak, ezért az egy személy által létrehozott hangsorok sem ismételhetők meg teljesen azonosan (Gósy 1996: 66–75). Ez nehezíti a rokonok beszédének azonosítását is, hiszen egy személy két ejtése ugyanúgy csak egy bizonyos fokig hasonlít, mint például egy testvérpár két tagjának ejtései. A rokonsági kapcsolatot fonetikai szempontból, szakszerű ikervizsgálattal eddig csak Forrai és Gordos vizsgálták (1983: 315–22), pedig sok a felmerülő kérdés, amelyeket eddig még nem válaszoltak meg.

Tapasztalat, hogy az azonos nemű rokonok beszéde hasonló, összetéveszthető. Gósy (2001. 423) egy család tagjain végzett akusztikai-fonetikai vizsgálat során azt találta, hogy a rokoni kapcsolatban levők beszéde különböző mértékben ugyan, de objektív adatokkal alátámaszthatóan hasonló. A jelenséggel kapcsolatban a következő fonetikai szempontú kérdések merülnek fel: 1. A beszéd mely összetevőjéhez, összetevőihez köthetők ezek a hasonlóságok? 2. Megkülönböztethető-e akusztikai-fonetikai mérések alapján a testvérek, ikrek hangja? 3. Elkülöníthető-e a beszédben az öröklött és a tanult tényezők? 4. Meghatározható-e, hogy mekkora mértékben tartalmaz a beszéd öröklött komponenseket? Ez a vizsgálat sorozat arra irányult, hogy ezekre a kérdésekre próbáljon meg választ adni.

Anyag és módszer

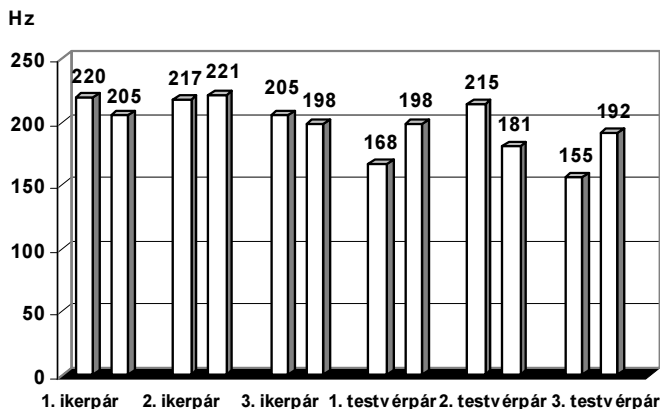
A genetikai kérdések megválaszolásához általában összehasonlító ikervizsgálatokat alkalmaznak (Atkinson 1999: 364–5). Ezeknek a lényege, hogy az egypetjű és a kétpetjű ikerpárok, illetve a sima testvérpárok összehasonlíthatók öröklött génjeik aránya alapján: az előbbieken minden génjükben osztoznak, az utóbbiak génállománya viszont csak körülbelül 50%-ban egyezik meg. Megfelelő összehasonlítással tehát elvileg elkülöníthetők az öröklött és a tanult tényezők, sőt ezek mértéke is

meghatározható: amilyen mértékben hasonlóbbak az egyetétjű ikrek, mint a kétetétjűek vagy a testvérek egy vonás tekintetében, olyan nagyságú a vonás genetikai, azaz örökletes összetevője. A kísérletben három egyetétjű ikerpár és három testvérpár vett részt, közöttük a korkülönbség kettő–négy év volt, mindannyian húsz–huszonnégy éves budapesti egyetemista, illetve főiskolás nők. A hanganyagot, két-két perces irányított spontán beszédet, kép alapján, a Wave Studio nevű számítógépes programmal rögzítettem, a hangelemzés a Praat 4.2-es programmal (Boersma–Weenink 2004) történt. A mérésekre természetesen csak a megfelelő számú azonos szavak, hangkapcsolatok kigyűjtése után kerülhetett sor, hiszen csak ezek hasonlíthatók össze.

Eredmények

A beszélőazonosításhoz általában szükséges mérések közül a következő összetevők vizsgálatát végeztem el: alaphangadatok; magánhangzók, mássalhangzók, hangsorok időtartama, magánhangzók formánsszerkezete, alveoláris és posztalveoláris spiránsok frekvenciatartománya, hangsoron belüli intenzitásviszonyok, sajátos spektrográfias alakzat, felhangszerkezet, FFT, percepció ellenőrző teszt. Az egyes méréseknél átlagosan tíz-tíz előfordulást vizsgáltam személyenként, a továbbiakban ezeknek a méréseknek az eredményeit ismertetem.

Az első ábra Hz-ben szemlélteti az egyes, a kettes, a hármas számú ikerpár és az egyes, a kettes, valamint a hármas számú testvérpár tagjainak **átlagos alaphangmagasságát**. Míg az ikerpárok esetében az eltérés mindössze 4–15 Hz, addig a testvérpároknál ez körülbelül 30 Hz. Ezek az eltérések nagyfokú hasonlóságot mutatnak Forrai és Gordos eredményeivel (Forrai–Gordos 1983: 319–21).



1. ábra. A beszélők átlagos alaphangmagassága

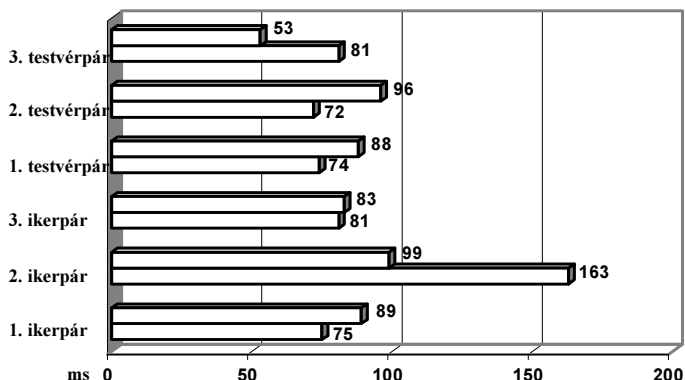
Az 1. táblázat a **zöngé minőségének** mérőszámait tartalmazza a három ikerpár és a három testvérpár esetében. Az emberi zöngéképzés nem teljesen szabályos, a zöngperiódusok szabálytalanok; ez megfigyelhető a hangrezgésgörbén, de számszerűsíthetjük is: a jitter a hangszalagrezgés frekvenciaingadozásának mértéke – minél kisebb az érték, annál állandóbb az egy másodpercre eső rezgésszám; a shimmer a hangszalagrezgés amplitúdóingadozását jelzi – minél kisebb az érték, annál állandóbb az intenzitás. A HNR, vagyis a jel-zaj viszony pedig a hang tisztaságát jelzi; minél kisebb az érték, annál kisebb zörejháttérrel mutat a hangképzés (Kovács 2003: 225). A három ikerpár és a három testvérpár tagjai között ugyanúgy felfedezhetők hasonlóságok, mint eltérések; a teljesen

megegyező értékek félkövérrel kiemelve láthatók az 1. és 3. ikerpár, illetve az 1. testvérpár esetében. Az alaphangon túl a jitter, a shimmer és a HNR is nagyobb korrelációt mutat az ikerpárok, mint a testvérpárok esetében.

1. táblázat. A zöngé minőségének mérőszámai

	JITTER (%)	SHIMMER (dB)	HNR (dB)
1. IKERPÁR	1,421	1,88	9,208
	1,422	2,614	6,151
2. IKERPÁR	1,537	1,376	12,084
	1,435	1,16	13,247
3. IKERPÁR	2,134	1,709	9,805
	1,809	1,27	9,857
1. TESTVÉRPÁR	1,651	1,365	11,045
	1,616	1,595	11,552
2. TESTVÉRPÁR	1,37	1,277	13,948
	1,981	1,464	12,293
3. TESTVÉRPÁR	2,185	1,625	8,993
	2,069	1,378	12,056

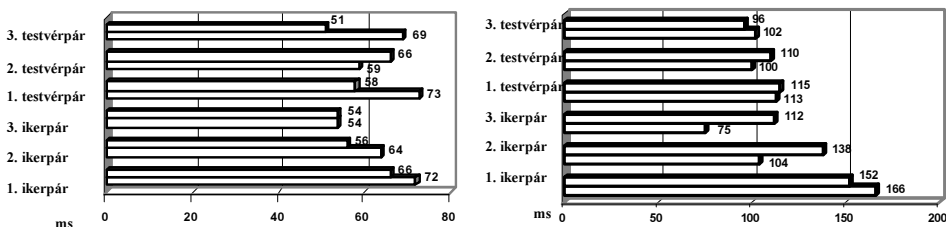
Kilenc **magánhangzó átlagos időtartamát** vizsgáltam azonos hangkörnyezetben, a 2. ábra ezek közül az [e:] magánhangzó átlagos időtartamát szemlélteti a tizenkét beszélő esetében. Látható, hogy leginkább a 3. ikerpár hasonlít 2 ms különbséggel, a testvérpárok közötti különbség 14–18 ms. Legeltérőbb a 2. ikerpár 64 ms különbséggel, ennek oka, hogy az „idősebbik” fél beszédére jellemző volt az a hezitációs jelenség, hogy megnyújtotta az ebben a hangkörnyezetben előforduló [e:] hangokat. Az időtartamokat tekintve a kilenc magánhangzó esetében az ikrek között jóval gyakrabban fordul elő erős pozitív korreláció, azaz +0,6 feletti érték, mint a testvérek között.



2. ábra. Az [e:] magánhangzó átlagos időtartama

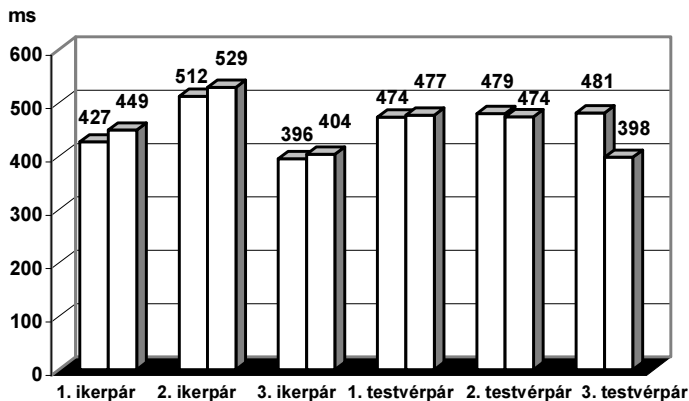
A [z] és a [j] **mássalhangzók átlagos időtartamát** tekintve (3. ábra) szintén az ikrek mutatnak nagyobb korrelációt, főleg a [z] mássalhangzónál. Az alveoláris, zöngétlen és a posztalveoláris, zöngés spiránsokra kapott adatokat is figyelembe véve arra lehet következtetni, hogy az ikrek közötti

erős pozitív korreláció valamivel többször fordul elő, mint a testvérek közötti; ez a különbség azonban nem annyira éles, mint a magánhangzók esetében.



3. ábra. A [z] és a [j] mássalhangzó átlagos időtartama

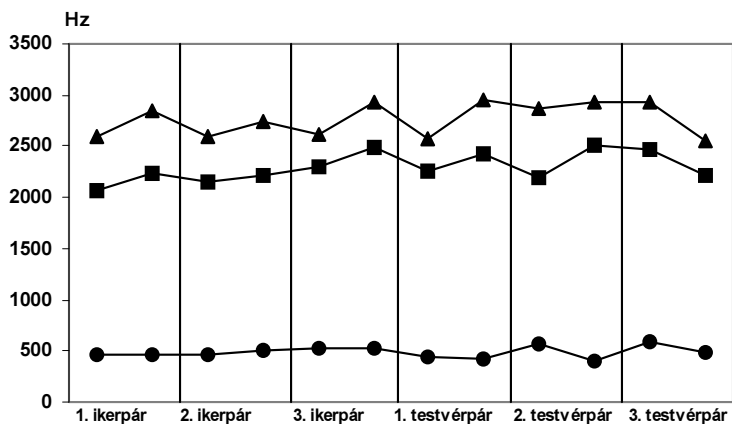
Összehasonlítottam az azonos hangsorok ejtésének időtartamait, személyenként átlagosan tizenegy hangsort vizsgáltam. Leghasonlóbb a 3. ikerpár, valamint az 1. és a 2. testvérpár; legkülönbözőbb a 3. testvérpár 83 ms eltéréssel (4. ábra).



4. ábra. Azonos hangsorok időtartamának átlaga

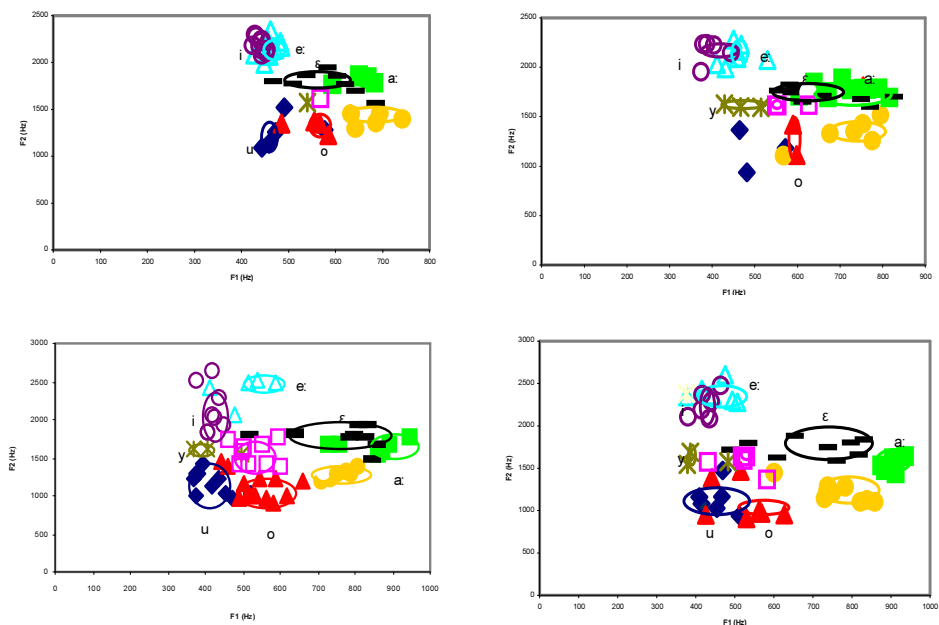
Kilenc **magánhangzó formánsszerkezetét** vizsgáltam azonos hangkörnyezetben, a 2. ábra ezek közül az [e:] magánhangzó első három formánsának átlagos frekvenciaértékeit szemlélteti a tizenkét beszélő esetében. Az első formáns egységes, kivéve a 2. és 3. testvérpárt, ahol 40–100 Hz-es eltérés látható. A második formáns is a testvérpároknál tér el inkább, a különbség 200–300 Hz. A harmadik formáns mutatja leginkább az egyéni különbségeket, az eltérések mértéke 300–400 Hz is lehet. A kilenc magánhangzót összességében tekintve az ikerpároknál valamivel gyakrabban fordul elő erős pozitív korreláció, főként a harmadik formánst tekintve. A formánssáv szélességei esetében viszont fordított volt a helyzet.

Ha a kilenc magánhangzó összes mért első és második formánsát egymáshoz képest ábrázoljuk egy koordináta-rendszerben, jellegzetes elhelyezkedésű ellipsziseket kapunk; ezek az úgynevezett **sűrűsödési ellipszisek**. Akár számadatok nélkül, vizuális benyomás alapján is dönthetünk a hasonlóság mértékéről; érdemes megfigyelni a jellegzetes átfedéseket. Az 1. ikerpár esetében, a 6. ábra két felső képén például megfigyelhető, hogy az 1. testvérpárhoz, tehát a lenti két képhez viszonyítva az [j], az [a:] és az [e] elhelyezkedése ugyanaz, és hogy az [i] és az [e:] átfedései is nagyon hasonlóak.



5. ábra. Az [e:] magánhangzó formánsszerkezete

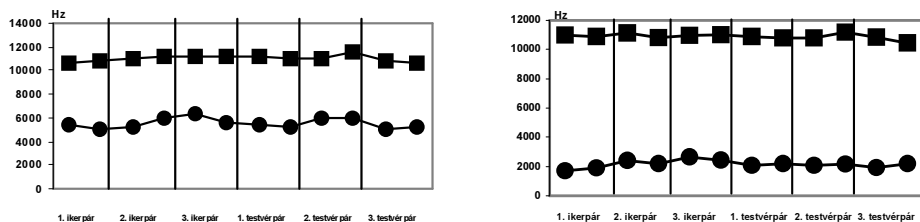
A 2. ikerpár esetében az [ɔ], az [a:] és az [ɛ] szintén hasonlóan helyezkedett el – mindkettőjükön egymáshoz jellegzetesen közel; valamint az [o] és az [u] elhelyezkedése is hasonló volt. Az összes ábra esetén tapasztalható volt az egyes párok hasonlósága, ugyanakkor mindegyik sajátosan jellemezte az egyént.



6. ábra. Sűrűsödési ellipszisek az 1. ikerpár (fent) és az 1. testvérpár (lent) magánhangzóiról

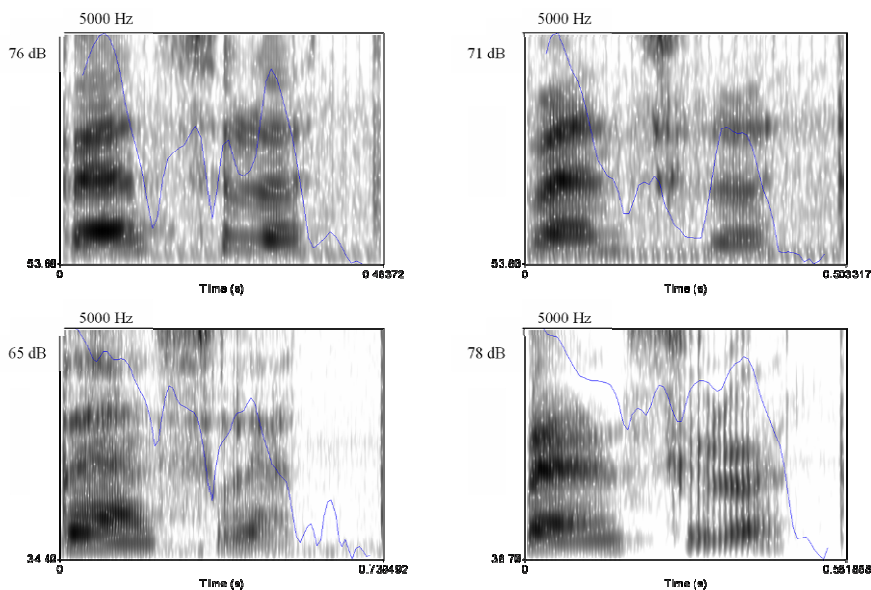
Az **alveoláris és posztalveoláris spiránsok frekvenciatartományára** elvileg alig hat a hangkörnyezet (Gósy 2004: 130), ezért a tíz-tíz spiráns, amelyekből átlagoltam, nem azonos hangkörnyezetből valók. Ezért is lehetséges az, hogy az intenzív zörejek megjelenését tekintve mégis ellentmon-

dásosak az adatok. A [j] mássalhangzónál az ikrek, a [z] mássalhangzónál a testvérek korrelálnak jobban (7. ábra). A legmagasabb frekvenciartomány, ahol az intenzív zörejegők még jelen vannak, csak a [z] mássalhangzó esetében mutat pozitív korrelációt, és ez az ikreknél magasabb.



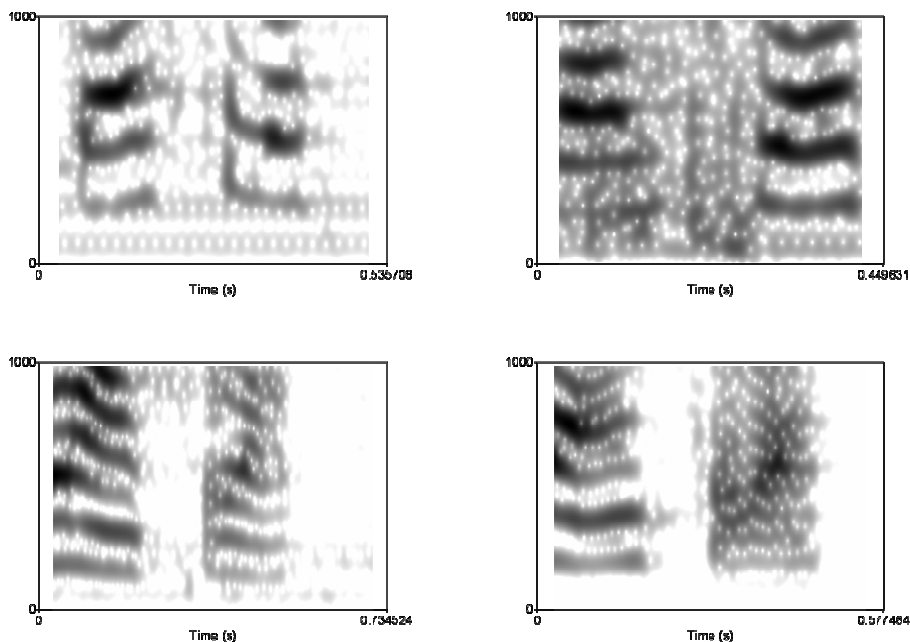
7. ábra. A [z] és a [j] mássalhangzó átlagos frekvenciartománya

A 8. ábra két felső képén az 1. ikerpár, két lenti képén a 3. testvérpár két tagjának ejtéséről készült széles sávú **spektrogramok** láthatók 5000 Hz-ig, a kék vonal pedig a **hangsoron belüli intenzitásvizonyokat** mutatja dB-ben. A [bensylöt] hangsort ejtették ki mindannyian. Az ikrek hangszínepein a formánsok elhelyezkedése szinte teljesen azonos, az intenzitásgörbe lefutása már eltérő; de még mindig nem akkora a különbség, mint a testvérpár esetében.



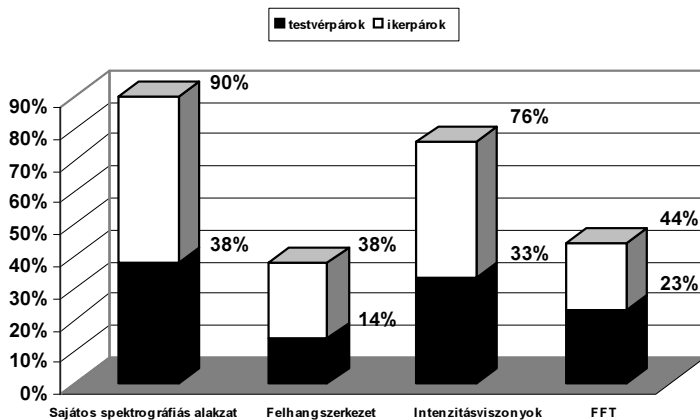
8. ábra. A [bensylöt] hangsor spektrogramja és intenzitásvizonyai

A 9. ábrán az előbbi spektrogramok keskeny sávú változatai láthatók 1000 Hz-ig, tehát az előbbi beszélők ugyanazon ejtéseinek **felhangszerkezetét** figyelhetjük most meg: felül az 1. ikerpár, alul a 3. testvérpár ejtéseiről készült képeket. A felhangok elhelyezkedését vizsgálva nemcsak a testvérpárok, hanem az ikerpárok között is különbséget tudunk tenni.



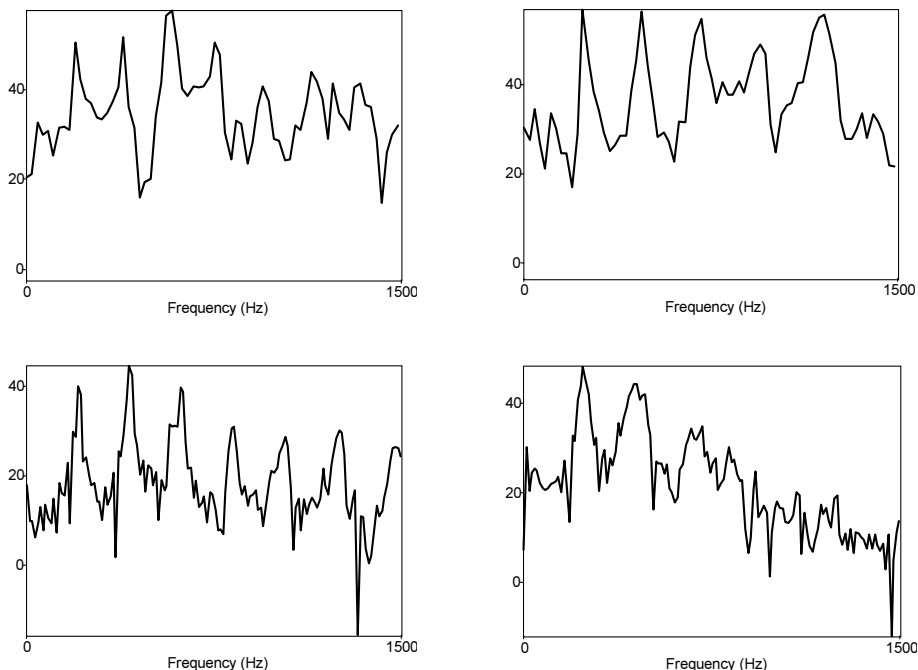
9. ábra. A [bensylot] hangsor felhangszerkezete

Az imént ismertetett háromféle összehasonlítást mind a hat párnál elvégeztem személyenként hét hangsort vizsgálva. A 10. ábra a nagy hasonlóságot mutató képek arányát szemlélteti az ikerpárok és a testvérpárok esetében. Az adatok azt mutatják, hogy a legnehezebben a hangspektrogram, majd az intenzitásgörbe alapján különíthetők el az egyes párok, a legkönnyebben pedig a felhangszerkezet alapján. Az ikerpárok tagjainak elkülönítése mindhárom esetben nehezebb volt, mint a testvérpároké.



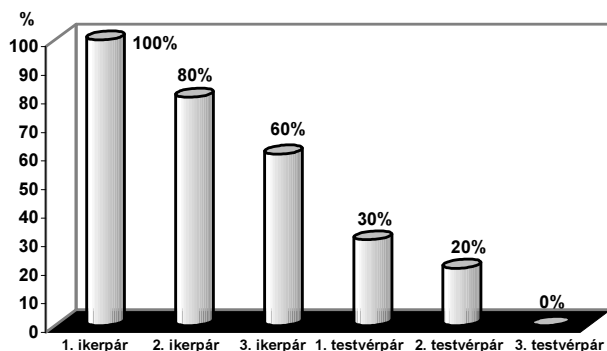
10. ábra. Azonos hangsorok adatainak hasonlósága

Az FFT a Fourier-féle sorfejtés alapján azokat a szinuszrezgéseket számítja ki, amelyekből összeáll a beszédhang. A számítás eredményének grafikai ábrázolása az **FFT-energiاسpektrum**, amelyről leolvashatók a felhangok intenzitásviszonyai és frekvenciaértékei. Az FFT-vel elsősorban a veláris magánhangzókat vizsgáltam 1500 Hz-ig. A 11. ábra két fenti képen az „u” magánhangzó FFT-energiاسpektruma látható a 3. ikerpár tagjainak ejtésében – a hasonlóság igazolható, de nem olyan nagy mértékben, mintha egyetlen személy két ejtését vizsgálnánk. A 2. testvérpár két tagjának ejtése (11. ábra lenti képei) abban különbözik, hogy a fiatalabb testvér energiaspektrumán (jobb oldali kép) a lefutás csökkenő tendenciája nagyobb mértékű. Az FFT segítségével közel olyan jól elkülöníthetők az egyes párok, mint a keskeny sávú spektrogramok alapján.



11. ábra. Az [u] magánhangzó FFT energiaspektruma

Az akusztikai-fonetikai méréseket **percepcióس tesztelés** egészítette ki. A percepcióس tesztet tíz fővel végeztettem el, mindannyian budapesti egyetemisták vagy egyetemet végzett fiatalok, a férfi-nő arány 50–50%. A résztvevők tíz alkalommal hallották a [bensylöt] hangsor két-két ejtését, a feladatuk az volt, hogy eldöntsék, vajon egyetlen személy két ejtését vagy két különböző személy ejtéseit hallották-e. A felvett szavak összepárosítása és a lejátszás sorrendje is véletlenszerű volt. A 12. ábra az ikerpárok és a testvérpárok „moduljainál” ejtett hibás tippek arányát mutatja, ez jelen esetben hasonlósági fokmérőként szolgál. Legnehezebben tehát az 1. ikerpár, legkönnyebben a 3. testvérpár tagjait tudták elkülöníteni. Gósy (2001. 428) vizsgálati eredményei szerint egyébként a testvérek beszédének ugyanezzel a módszerrel kiszámított hasonlósági foka 58,4%.



12. ábra. Percepció ellenőrző teszt

Következtetések

A cikk elején megfogalmazott kérdésekre a következő válaszok adhatók: 1. Az ikerpárok hangja nagyobb mértékben tartalmaz hasonló akusztikai-fonetikai adatokat, mint a testvérpároké – ez a genetikai tényező elkülöníthetőségére utal. A párok hasonlósági fokát tekintve a három ikerpár egyszerre végzett az első helyen, utánuk következik az első, a második, végül a harmadik testvérpár. 2. A hasonlóságok minden mért beszédparaméterben fellelhetők, a párok hangja leginkább a felhangszerkezet figyelembevételével különböztethető meg. A felhangszerkezet tehát olyannyira jellemző az egyénre, hogy még az egyetjű ikerpároknál is kimutatható a különbség; kriminalisztikai elemzés esetén tehát ezt kell a leginkább figyelembe venni. 3. A kutatás több személyre való kiterjesztésével pedig statisztikai szempontból is vizsgálhatóvá válna az örökletesség mértéke.

IRODALOM

- Atkinson, Richard C.–Atkinson, Rita L. 1999. *Pszichológia*. Osiris Kiadó, Budapest, 65–8; 364–70; 561–70.
- Boersma, Paul–Weenink, David 2004. *Praat: doing phonetics by computer 4.2*. Institute of Phonetic Sciences, University of Amsterdam. <http://www.praat.org>
- Forrai G.–Gordos G. 1983. A new acoustic method for the discrimination of monozygotic and dizygotic twins. *Acta Paediatrica Hungarica*. 24: 315–22.
- Gósy Mária 1996. A beszéd akusztikai szerkezetének állandóságáról. In: Terts István (szerk.): *Nyelv, nyelvész, társadalom. II.* Keraban Könyvkiadó, Pécs, 66–75.
- Gósy Mária 2001. A genetikai tényező a beszélő személy felismerésében. In: Andor József–Szűcs Tibor–Terts István (szerk.): *Színes eszmék nem alszanak. Szépe György 70. születésnapjára I.* Lingua Franca Csoport, Pécs.
- Gósy Mária 2004. *Fonetika, a beszéd tudománya*. Osiris Kiadó, Budapest.
- Gósy Mária–Nikléczy Péter 1999. A beszélő felismerése a beszéde alapján: elméleti háttér és módszertani megközelítések. In: Gósy Mária (szerk.): *Beszédkutatás '99*. MTA Nyelvtudományi Intézet, Budapest, 1–19.
- Kovács Magdolna 2003. A beszéd vizsgálata az elemektől a folyamatig a Praat program tükrében. In: Hunyadi László (szerk.): *Kísérleti fonetika – laboratóriumi fonológia 2003*. Debreceni Egyetem Kossuth Kiadója, Debrecen, 213–39.

SUMMARY

*Pápay, Kinga***Hereditary phonetic parameters of the human voice**

Voice recognition or speaker identification has been studied from several points of view. The close similarity of the voices of consanguineous persons has suggested that there must exist some hereditary phonetic parameters. To test this hypothesis, a method for characterizing the similarity or dissimilarity of voices of sisters and identical twins in numerical terms was developed. If some parameters of human voice are genetically determined, monozygotic twins must have a higher intra-pair similarity of voice than dizygotic like-sexed twins or brothers/sisters. If agreement within monozygotic twins significantly exceeds that observed in dizygotic twins of the same gender, it may be anticipated that the condition is more under genetic than environmental control. In the present work, the voices of both members of three female 21–22-year-old pairs of monozygotic twins and three female 20–24-year-old pairs of sisters speaking about the same picture for a duration of about 2 minutes were recorded by the computer program Wave Studio. The material was analyzed by the Praat 4.2 voice analyzing program: average pitch, first three formants and formant bandwidths of nine vowels, the duration of words, vowels and alveolar fricatives, word intensity, FFT spectra of vowels, broad band and narrow band spectrograms of words, and broad band spectrograms of alveolar fricatives were analyzed and numerically characterized. The results, in accordance with our expectations, show that intra-pair differences between monozygotic twins were lower than those between sisters. It may be concluded that a single parameter is not capable of complete discrimination; there is no definite answer as to which parameter is the single most characteristic one of the heritability of human voice.