

BEKE ANDRÁS

MTA Nyelvtudományi Intézet

*beke.andras@gmail.com***AZ ALAPFREKVENCIA-ELOSZLÁS MODELLEZÉSE A
BESZÉLŐFELISMERÉSHEZ**

Fundamental frequency (F_0) is the acoustic measure of the vocal folds' vibration. It is considered to be one of the most important prosodic features characterizing speech and speaker. This study considers two aspects: describing F_0 and its modeling for speaker recognition. The F_0 was parsed using ACF (autocorrelation function). The article presents a model for short-term F_0 , combining parametric (referring to the statistical properties of the F_0 distribution) and non-parametric (analyzing the density function of the distribution by histograms) approaches. The database of the research consists of narrative and spontaneous speech samples of 20 male Hungarian speakers. The acoustic analysis was conducted using Praat 4.5. The statistical analysis was carried out using SPSS 13.0. The identification was performed using MATLAB 7.0. This study shows that F_0 may provide additional information for the speaker identification.

Bevezetés

Napjainkban a beszélő személyének azonosítása egyre nagyobb figyelmet kap. Kommunikáció közben a beszélő felek képesek felismerni, azonosítani partnereiket beszédük alapján akár közvetlen beszélgetés közben, akár telefonon keresztül, rádióban vagy egyéb kommunikációs csatornán keresztül. Ha a beszélő személy ismert a hallgató számára, akkor felismerhető hangjának hangszínezete, beszédstílusa alapján (Hautamäki, 2005). Ha nem, akkor számos becslést teszünk a beszélő testi jegyeire: korára, érzelmi állapotára, súlyára és magasságára (Gósy, 2001a). Az agy egy beszédrészlet észlelésekor többféle döntést hoz: dekódolja a beszédet úgy mint nyelvi üzenetet, azaz hogy milyen szavak, mondatok hangzottak el (beszédpercepció, megértés), valamint megítéli például a beszélő személy korát és hangulatát is (Gósy, 2001a, hivatkozva Fujimura-ra, 1972-ben; Gocsál, 1998). A beszélőfelismerés is két összefüggő, de nem azonos paraméterből tevődik össze: nyelvi és nem nyelvi információból, amelyet ugyanazon akusztikai jel hordozhat. Batliner, Steidl és Nöth szerint például az érdes hang jelezheti egy bemondás végét (nyelvi), kifejezhet unalmat (nem nyelvi) vagy lehet a beszélő személy hangszalagjainak a jellegzetessége (nem nyelvi) (Böhm, 2007 hivatkozva Batliner-Steidl-Nöth-re, 2007). A beszéd nem nyelvi jellemzői között megkülönböztetünk egy spektrális vetületet és a beszélőben kialakult neurális spektrális vetületet. A beszélő felismerése függ attól is, hogy létezik-e a beszélőtől származó minta, illetve attól, hogy ez a minta milyen akusztikai jeggyel rendelkezik (Gósy, 1999). A beszédhangból leképezett

neurális spektrális jegyek egy beszélőn belül kisebb varianciát mutatnak, mint a különböző beszélők között: ezért is mehet végbe a beszélőfelismerés. A spektrogramok mindig egyediek, a neurális spektrogramok pedig szükségszerűen valamiféle általánosított képek kell, hogy legyenek (Gósy, 2001). Alapvető kérdésként felmerül az, hogy ezek a neurális spektrogramok milyen összefüggésben vannak az egyén adott spektrogramjaival.

A beszélőfelismerésre különböző elméletek jöttek létre. Van Lancker, Kreiman és Emmorey (1985a) munkájukban alapul vették a jegymodellt, és továbbfejlesztették azt. *Gestalt*-modelljükben azt feltételezik, hogy a beszélő-hallgató párosnak a rendelkezésre álló jegyhalmaz redundáns, így elégséges azok közül néhányat felismerni az azonosításhoz (*Gestalt*-zárás) (Van Lancker, Kreiman és Emmorey-re 1985; később ebből alakult ki Kreiman és Papcun 1991 prototípus modellje: a multimodális személyfelismerés modell). A későbbi multimodális modellekben (Von Kriegstein és Giraud, 2006) összekapcsolták az arcfelismerést a beszédfelismeréssel. A kísérleteikkel azt bizonyították, hogy a hang- és arcfelismerő területek szorosan összekapcsolva, egymást kiegészítve működnek, így az arc és a hang mintegy multimodális *Gestalt*ként működik: csak a hangot érzékelve az agy befejezi a személységét, azaz melléteszi az arcot, hogy a későbbi személyfelismerést ezzel segítse (Von Kriegstein és Giraud, 2006). A prototípus modell szerint tehát elégséges egy vagy néhány paraméter vizsgálata, modellezése ahhoz, hogy a beszélő személyt azonosítani lehessen. A kérdés azonban az, hogy mely vagy melyek ezek a beszédakusztikai jegyek. Az akusztikai fonetikai kutatások kezdetén a beszéd akusztikai paraméterei közül a magánhangzók harmadik formánsát gondolták azonosító jegynek, majd később az alaphangmagasságot. A kilencvenes évektől a kutatások nagy része a spektrográfiai eljárásokra összpontosult (MFCC).

A hangszalagok kvázi-periodikus rezgésének eredménye a zöngé, amely az alaphangból és a felhangokból tevődik össze. Az alaphang frekvencia tehát jellemző akusztikai paramétere a beszédnek, amely a beszédhang legkisebb frekvenciájú összetevője (jele F_0). A beszéd során az alaphangból és felhangokból álló zöngé áthalad az üregrendszeren (garatüreg, szájüreg, orrüreg), amely rezonátorként működik. A zöngé saját frekvenciáiból az üreg a sajátfrekvenciájának megfelelő vagy ahhoz közel eső felhangokat (felharmonikusokat) felerősíti. Így bizonyos frekvenciákon energia-koncentrációk jönnek létre, amelyeket formánsoknak nevezünk (Gósy, 2004).

Korábbi vizsgálatokból tudjuk, hogy az alaphang frekvencia értéke megfelelő körülmények között elégséges jegy lehet a beszélő személy felismeréséhez (Rose, 2002). Ez lehet személyazonosító jegy a beszélők megkülönböztetésére nemcsak a nemek tekintetében, de azonos neműeken belül is, ha az alaphang frekvencia átlaga jól elkülöníthető egymástól (Rose, 2002; Ezzaidi, *et al.*, 2001; Asami, *et al.*, 2003). Az alaphang számos lehetőséget nyújt arra, hogy egyéb akusztikai paramétereket nyerjünk ki belőle a felismerés számára (például az érdes zöngét) (de Cheveigné & Kawahara, 2002; Böhm, 2006). Megállapították azt is, hogy az F_0

függetlenebb a kommunikációs csatornától, tehát nem vagy csak alig torzul a zaj által vagy speciális körülmények hatására (de Cheveigné & Kawahara, 2002). Korábban csupán az alapfrekvencia átlagát, vagyis az átlagos alapfrekvencia-értéket használták azonosító jegyként. Az alapfrekvencia értékeinek feldolgozására számos módszer alakult ki. Irena Chmielewska munkájában az alapfrekvenciát és az intenzitást használta fel beszélőazonosításra egy szövegfüggő rendszerben. Labutin, Koval és Raev (2007) az alapfrekvencia értékeinek statisztikai alapú elemzését hajtotta végre egy 126 beszélő telefonbeszélgetését tartalmazó adatbázisban (Labutin, *et al.*, 2007 hivatkozva RUSTEN adatbázisra, 2003). Svéd nyelvben Joan Lindh publikálta azt a 109 fiatal beszélőn végzett kísérletét (a kísérletben a svéd adatbázist használta, amelyben 17 éves fiatalok spontán beszédét rögzítették), amelyben az adatközlők alapfrekvenciájának értékét leíró módon jellemezte (Lindh, 2006). Később az alapfrekvencia mellett más akusztikai paramétert is felhasználnak a beszélő személyének azonosítására. Így jött létre a Hollien és munkatársai által létrehozott fonetikai alapú rendszer (SAUSI), amelyben az F0-t, a csendes szünetek számát és hosszát, a beszédtempót és a magánhangzók időtartamát használták a beszélőfelismeréshez (Gósy, 2001 hivatkozva Hollien-re, 1990). Hassan Ezzaidi és Jean Routat (2004) munkájukban az alaphang mellett MFCC-t és GMM-t is felhasznált a beszélőazonosításhoz (SPIDRE telefonos adatbázison). Olyan eredmények is megjelentek, amelyek az alapfrekvencia eloszlásával próbálták meg alkalmazható beszélőfelismerő rendszert készíteni (Rose, 2005; Labutin, *et al.*, 2007). Ezekben a kísérletekben az alapfrekvencia-eloszlást szövegfüggetlen anyagon vizsgálták, de a beszéd típusa azonos volt (telefonos beszélgetés). Felmerül a kérdés, hogy a különböző beszéd típusokban ezek a modellek hogyan alkalmazhatók.

A jelen kutatás célja az, hogy megvizsgáljuk, milyen eredménnyel lehet elérni magyar nyelven, szövegfüggetlen és beszédtypustól független anyagon a beszélő személyének azonosítását. Kísérletünkben azt vizsgáltuk, hogy az alapfrekvencia eloszlásgörbéjéből számolt parametrikus (*Euklidészi távolság*) és nem parametrikus (*Kullback-Leibler távolság*) távolságok közül a beszélőfelismerésben melyik milyen találati hibaarányal működik egy 20 beszélős adatbázisban. Elemeztük azt is, hogy milyen mértékben ront vagy javít az alapfrekvenciából mért érdes zöngé értéke a beszélő felismerésében. Feltevésünk az, hogy a beszélők beszélőn belüli megnyilatkozásainak F0 eloszlásfüggvényei közel azonosak lesznek a különböző beszéd típusokban.

Anyag, módszer és kísérleti személyek

A kutatásban a BEA-ban (magyar spontánbeszéd-adatbázis) rögzített hanganyagon dolgoztunk (Gósy, 2008). Az elemzéseket 20 beszélőn végeztük el: 10 férfi és 10 nő: életkoruk 25-60 év, ép hallásúak, nem beszédhibások. A beszéd-adatbázisban lehetőség van arra, hogy különböző beszéd típusokat válasszunk ki, amelyek nagy mennyiségben állnak rendelkezésünkre. Két beszéd típusban részt vevő adatközlők beszédét vizsgáltuk: felolvasás és spontán beszéd. Az elsődleges minta a felolvasásból kinyert adatok, míg a referenciaminta a spontán beszédből származó adatok voltak. A hanganyagokból minden beszéd típusban 2 percet használtunk fel.

Az alapfrekvencia elemzéséhez *autokorrelációs* algoritmust alkalmaztunk. Az eloszlásgörbéből kiszámoltuk a középérték, a szórás, a ferdeség és a laposság értékeit. Az elemzéseket a Praat 4.5 szoftverrel végeztük el. A statisztikai eredményeket az SPSS 13.0 verziójú szoftverrel készítettük. A statisztikából származó adatokat a MATLAB 7.0 matematikai szoftverben dolgoztuk fel. Ebben a programban számoltuk ki az eloszlásgörbék között fellépő távolság értéket: az egyik esetben az eloszlásgörbéből számolt paraméterekből számoltuk ki a különbözőség mértékét, amit az euklidészi távolsággal végeztünk el. A második esetben az eloszlásgörbe egészén végeztük el az összehasonlítást, ami egy nem parametrikus módszert kívánt meg, amit a *Kullback-Leibler távolsággal* számoltunk. A beszédminták azonosságának vagy különbözőségének a megállapítására is a MATLAB-ot használtunk. Ebben a programban futtattuk a felismerésre alkalmazott algoritmusokat, és összegeztük az eredményeket, megállapítva a felismerési hibaaarányt.

A rövid idejű alapfrekvencia analízishez használt algoritmusok közül a következőket kell megemlíteni: *autokorrelációs algoritmus* (ACF), *átlagos magnitúdókülönbség-függvény* (AMDF) és a *kepstrum analízis*. Az elemzések közül a jelen kutatásban az *autokorrelációs* algoritmust alkalmaztuk.

Az *autokorrelációs algoritmus* az egyik leggyakrabban használt időtartománybeli eljárás, azt mutatja meg, hogy a jel minden egyes pontjában milyen mértékben hasonlít önmagára. Az egyenlete:

ahol N a jel hossza mintaszámban kifejezve, $r[i]$ az autokorrelációs függvény értéke az i -től függően. Az alapfrekvenciához tartozó periódusidő egyenlő az $r[i]$ függvény maximumának helyével az $i=0$ indextől nagyobb értékekre. Az ACF még hatékonyabban számolható a frekvenciatartományban a gyors Fourier-transzformáció segítségével (FFT). Először frekvenciatartományba transzformáljuk a jelet, majd az amplitúdóspektrumot megszorozzuk komplex konjugáltjával, végül visszatranszformáljuk újra időtartományba:

Általában ezt a megoldást szokták alkalmazni az időtartományban történő konvolúció helyett, mivel ez gyorsabb. Az autokorreláció kiemeli a harmonikus frekvenciákon elhelyezkedő amplitúdókat:

Ez az egyenlet kifejezi a pontos működést; amikor az i egyenlő a jelalak periódusával, akkor az amplitúdóspektrum négyzete maximálisan súlyozódik. Az ACF alapú hangmagasságot számító eljárások érzéketlenek a zajokra, viszont érzékenyen reagálnak a spektrumban megjelenő kiugró értékekre. Az amplitúdóspektrum négyzetre emelésével nő a zaj, de a kiugró értékek szerepe is megnő.

Eredmények

Az alapprofrekvenciából autokorrelációs analízissel kapott eredményeket eloszlásfüggvényként kezeltük. A vizsgálat során az F0 eloszlására koncentráltunk, míg az F0-értékek időbeli paraméterét elhagytuk. Az alapprofrekvencia időbeli lefutása ugyanis sokkal inkább a pszichikai állapotnak vagy a szövegnek a sajátossága, mint a beszélő hangjának (Hautamäki, 2005). Az F0 eloszlásfüggvényéhez hisztogramot használtunk mint a gyakran használt területek grafikus reprezentációját. Az eloszlásfüggvény megvalósítása előtt a kapott adatokat úgy normalizáltuk, hogy az eloszlásgörbe teljes térrre vett integrálja 1 legyen. Az eloszlásfüggvényt kétféleképpen kezeltük. Az első eljárásban az eloszlásfüggvényből négy olyan paramétert számoltunk, amelyek jól jellemzik ezt a függvényt. Majd ezekből a paraméterekből számoltunk *euklidészi* távolságot. A második esetben a két eloszlásfüggvény között fellépő különbséget írtuk le a *Kullback-Leibler* távolság módszerével.

Parametrikus modell

Az eloszlásfüggvény jellemzésére különböző paramétereket mértünk, amelyeket a továbbiakban a beszélőre jellemző vektorokként kezeltünk. Ezek a következők: középérték, szórás, ferdeség és laposság. Az átlag az elemek értékeinek összege, amelyet elosztunk az elemek számával:

ahol az N a hangból vett mintákat jelenti.

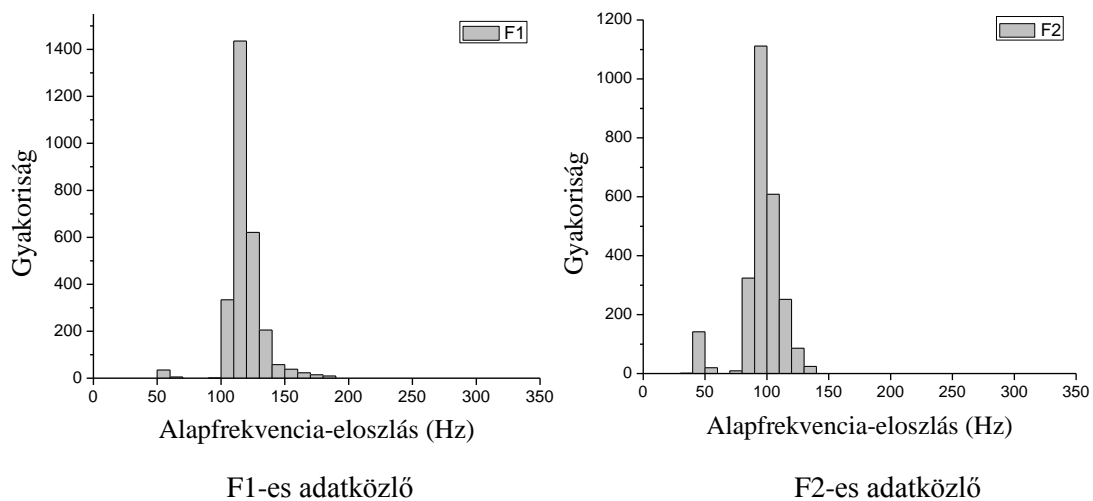
A szórás az adatok középértéktől számított abszolút értékű eltéréseinek átlagolása:

A ferdeség egy eloszlás ferdeségét határozza meg. A ferdeség az eloszlás középérték körüli aszimmetriájának mértékét jelzi. A pozitív ferdeség a pozitív értékek irányába nyúló aszimmetrikus eloszlást jelez, míg a negatív ferdeség a negatív értékek irányába:

A görbe ferdesége negatív, ha $\gamma_1 < 0$. Normál eloszlású, ha $\gamma_1 = 0$, és pozitív, ha $\gamma_1 > 0$.

A laposság az eloszlásban azt mutatja meg, hogy a görbe szélességére és magasságára jellemző-e a lapultság:

Ezekből az értékekből származtatjuk az F0 vektorait: $P = (\mu, \sigma, \gamma_1, \gamma_2)$. Az 1. ábrán látható két beszélő eloszlásgörbéje és az 1. táblázatban az ebből számolt vektorértékek.



1. ábra

Két adatközlő beszédének F0-eloszlása

1. táblázat

Példa az alapfrekvencia-eloszlásból számolt vektorokra két beszélőnél

Beszélők	Paraméterek	Középérték	Átlagos eltérés	Ferdeség	Laposság
F1	<i>Felolvasás</i>	118,5789	43,51239	3,431574	12,0359
	<i>Spontánbeszéd</i>	121,7631	54,86523	3,542065	12,82492
F2	<i>Felolvasás</i>	96,29334	35,47517	1,766735	2,27129
	<i>Spontánbeszéd</i>	86,98211	39,1772	2,01486	3,158022

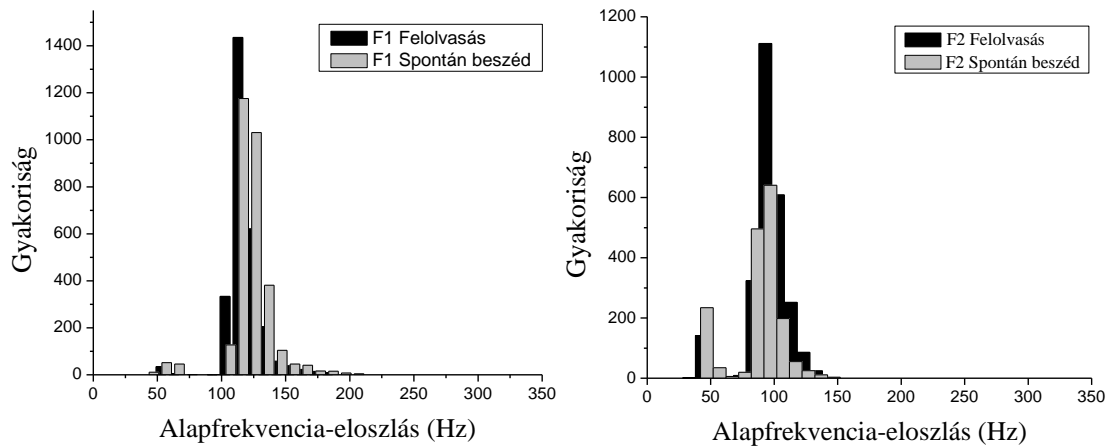
A parametrikus eljárás során ezekkel a vektorokkal számoltunk tovább, megállapítva a két beszélő között fellépő távolságot (d), amit az *euklidészi* távolsággal kaptunk meg (*Euclidean távolság*):

ahol az X és Y két beszélőnek a vektorát jelenti ($X_{F1}=P=(\mu, \sigma, \gamma1, \gamma2)$). Az *euklidészi* távolság két eloszlásból számított vektor között lévő négyzetes különbséget írja le.

Nem parametrikus modell

A parametrikus modell mellett nem parametrikus modellt is felhasználtunk a biztosabb azonosítás érdekében. Ez az *euklidészi* távolság vektorok között határozta meg a négyzetes távolságot. Az F0-eloszlás pontosabb modellezésére a *Kullback-Leibler* távolságot alkalmaztuk, amely két eloszlás között fellépő távolságot ír le (Cambell, 1997; Cover és Thomas, 1991; Jauquet, Verlinde és Vloeberghs, 1997; Christopher Bishop, 1995). Ezt a következőképpen számoljuk:

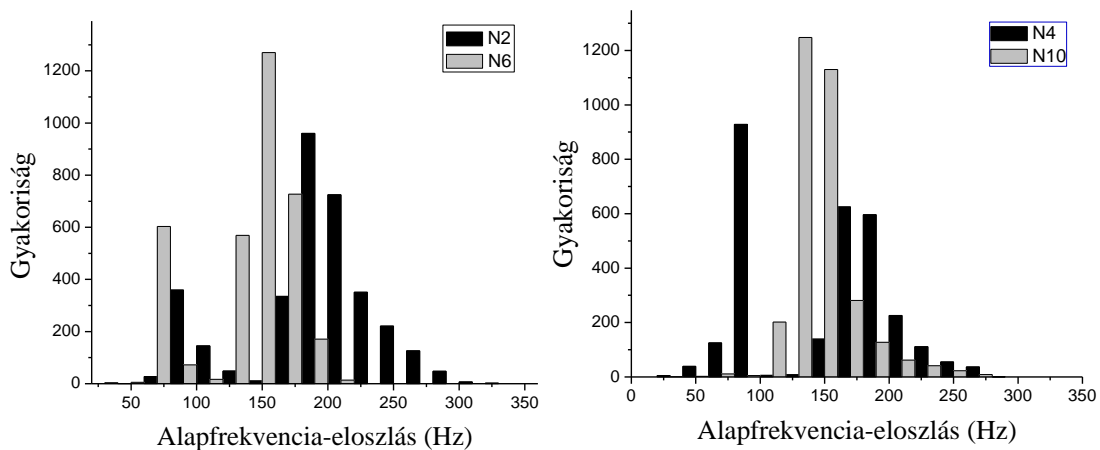
ahol p_k és q_k az eloszlásban lévő oszlopok értékét jelenti, és ahol általában ($d(p, q) \neq d(q, p)$). A távolság értékeivel jó közelítéssel leírtuk az F0-eloszlások között fellépő különbséget (a 2. és 3. ábrán látható az eloszlásfüggvények közötti hasonlóság és különbség), így egy alapfrekvenciára épülő rendszert hoztunk létre.



2. ábra

Hisztogramok azonos beszélőkre

A 2. ábrán látható, hogy a két beszéd típusban mért alapfrekvencia-értékek eloszlása közel azonos módon realizálódik azonos beszélők esetében, vagyis fedésben vannak. Ez azt jelenti, hogy a köztük fellépő *Kullback-Leibler* távolság is kisebb értéket vesz fel. A 3. ábrán látható, hogy két különböző beszélő esetében nincsenek fedésben a hisztogramok, ezért a köztük fellépő távolság is nagyobb lesz.



3. ábra

Hisztogramok különböző beszélőkre

A beszélőfelismerésben az egyik legfontosabb annak az algoritmusnak vagy statisztikai eljárásnak a megválasztása, amely az adathalmazban csoportosítani tudja az azonos beszélőket, és szét tudja választani a különbözőket. A jelen ku-

tatásban az *euklideszi* távolságból és a *Kullback-Leibler* távolságból származó adatokat mátrixtáblázatba rendeztük. A mátrixban minden beszélő alapfrekvencia eloszlásából az algoritmus kiszámolja az *euklideszi* és *Kullback-Leibler* távolságot. A klasszifikációhoz használt algoritmus a mátrix oszlopaiban megkeresi a legkisebb értéket, és kiírja annak a cellaindexét. Az algoritmus azt a mintát fogja azonosnak mutatni, amely a vizsgált mintához a legkisebb távolságra van. Majd ezek után összehasonlítja az általunk megadott helyes mintával, és ahol eltér az algoritmus találat a helyes megoldástól, ott hibát jelöl. Ezek után összeszámolja a tévesztéseket és kiszámítja a felismerési hibát $((N_{\text{téves}}/N_{\text{összes}})*100)$.

Az alapfrekvencia-eloszlásra alkalmazott parametrikus és nem parametrikus módszerekkel 90%-os felismerési hibát ért el Hautamáki (2005), ami egy részeredmény. A dolgozatban ezen a részeredményen tudunk javítani. Ezért feltételezzük azt, hogy a későbbiekben a Hautamáki által megadott végső eredményeken is tudunk majd javítani (Hautamáki MFCC paraméterrel kombinált rendszerrel csökkenteni tudta a felismerési hibát).

Hautamáki eredményén, ami 90%-os felismerési hiba, 38,5%-ot tudunk javítani, vagyis az általunk létrehozott rendszer legjobb teljesítménye 52,5%-os felismerési hiba. Abban az esetben, amikor az eloszlásgörbéből nem parametrikus úton számoltunk *Kullback-Leibler* módszerrel, kisebb felismerési hibát kaptunk, mint amikor parametrikus úton *euklideszi* módszerrel, amelyet az alapfrekvencia-eloszlás 4 vektorából számoltunk (One-way ANOVA $F(2, 7) = 8,737$; $p=0,021$). A nem parametrikus módszerek közül pedig a *Kullback-Leibler* távolsággal végzett számítás működik jobban. Minden esetben a nők hangja azonosítható kisebb felismerési hibaarányal (One-way ANOVA $F(2, 6)=32$; $p=0,030$). Ez azért lehetséges, mert a női beszélők között nagyobb a variancia (Beke, 2008). Mindezen számításokat az érdes zöngéből adódó értékek beszámításával kapjuk (2. táblázat).

2. táblázat

Felismerési hiba érdes zöngével

	Módszer	Adatközlők	Euklideszi	Kullback-Leibler
F0	Nem parametrikus	<i>Férfiak</i>	85%	70%
		<i>Nők</i>	30%	20%
		Összesen	60%	52,5
	Parametrikus	<i>Férfiak</i>	100%	-
		<i>Nők</i>	90%	-
		Összesen	99,5%	-

A felismerési hiba nő akkor, ha az eloszlásból kihagyjuk az érdes zöngéből adódó értékeket. Ez a jelenség jelentősen növeli a *Kullback-Leibler* távolsággal számított felismerési hibát (átlagosan 26,66%-kal). Az *euklidészi* távolsággal számított felismerési hiba értéke a férfiaknál 10%-kal nő. Az összes beszélőre nézve ez az érték ugyanez (az eltérés átlagosan 10%-kal nő). A nőknél ez az érték 30%-kal nő, tehát nőknél a személyfelismerés az érdes zöngé nélkül nehezebb (3. táblázat).

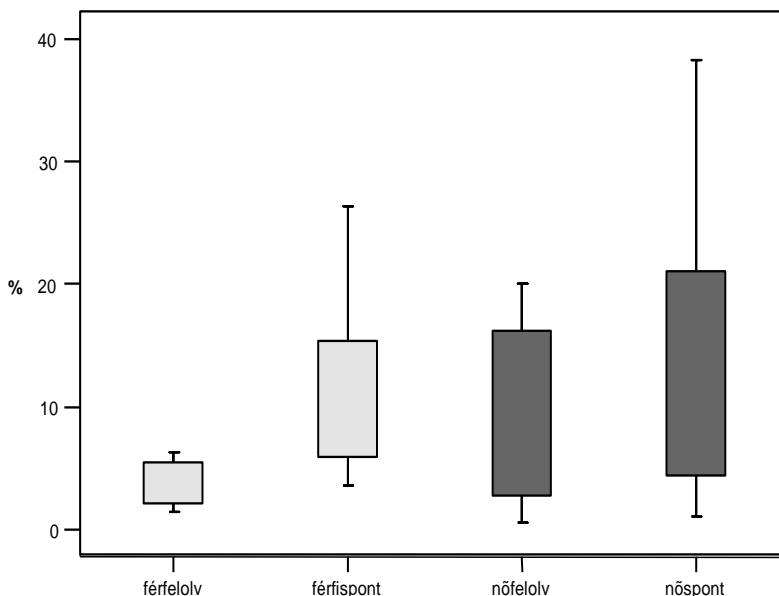
3. táblázat

Felismerési hiba (error rate ER%) érdes zöngé nélkül

	Módszer	Adatközlők	Euklidészi	Kullback-Leibler
F0	Nem parametrikus	<i>Férfiak</i>	85%	90%
		<i>Nők</i>	60%	45%
		<i>Összesen</i>	62,5%	87,5%
	Parametrikus	<i>Férfiak</i>	99,5	-
		<i>Nők</i>	100%	-
		<i>Összesen</i>	99,25%	-

Az érdes zöngé nélkül is a paraméteres eljárással számított távolságnak nagyobb a felismerési hibaaránya ($F(2, 7)=6,433; p=0,039$). Minden esetben tehát, akár érdes zöngével akár a nélkül a parametrikus módszer bizonyult a legrosszabbnak ($F(2, 17)=14,081; p=0,002$). A nem parametrikus módszerek közül a *Kullback-Leibler*-féle távolsággal számolt felismerés hozta a legkisebb felismerési hibát abban az esetben, amikor az érdes zöngé által adott értékek is részét képezték az eloszlásgörbének. Ennek a módszernek azonban az érdes zöngé eloszlásgörbéje nélkül romlik a felismerés biztonsága. Az *euklidészi* távolsággal számolt felismerés érdes zöngé esetében rosszabb, mint a *Kullback-Leibler*-rel számolt, de stabilabb is, hiszen nem érinti olyan mértékben a felismerési hibaarányt az, ha az érdes zöngé nélkül számolunk. Az *euklidészi* távolsággal tehát 60-62,5%-os felismerési hibát tudunk elérni.

Az érdes zöngé szerepe tehát fontos a beszélő személyazonosításában, hiszen a glottalizáció mértéke és értékei méréseink szerint is személyfüggő jegyek (Böhm, 2006). Ugyanakkor számos más jelzéssel rendelkezhet: jelezheti egy bemondás végét (Böhm, 2006 hivatkozva Slifka-ra, 2000). A kísérletünk során megfigyeltük azt is, hogy az érdes zöngé megjelenésének mértéke nemcsak a beszélő személytől függ, hanem a beszéd típusától is. Ez a jelenség a férfiaknál jelenik meg nagyobb mértékben (4. ábra).



4. ábra

A glottalizáció mértéke a felolvasásban és spontán beszédben

A két beszéd típusban létrejött glottalizáció mértéke a férfiaknál szignifikáns (kétmintás t -próba: $t=-3,633$; $p=0,008$). A nők esetében is megvan ez a különbség, de a jelenség nem szignifikáns ($t=-1,548$; $p=0,156$). Ez is alátámasztja a felismerésnél kapott eredményt, hiszen ez azt jelenti, hogy bár a nők közötti variancia nagy, de az egyes beszélőkön belül kicsi. Ezért a nők könnyebben azonosíthatók beszédük alapján.

Összefoglalás

Megállapítható, hogy az érdes zöngével mért adatokkal, amelyet *Kullback-Leibler* távolságon alapuló rendszerben hoztunk létre, eredményesebb felismerést tudunk elérni. A Hautamáki személyazonosító rendszer egyik módszerének részeredményén 38,5%-ot tudunk javítani. Ezért feltételezzük azt, hogy az eredményeken az MFCC eljárás alapján alapuló analízissel és rejtett Markov-modellt (HMM) alkalmazó rendszerrel lehetne tovább javítani, és ezzel egy több dimenziós felismerő egységet létrehozni. Eredményeink egy lehetséges további paraméterül szolgálhatnak a beszélő személy azonosításában, és hozzájárulhatnak az érdes zöng szerepének megítéléséhez a beszédben.

Irodalom

- Asami, A. G., Michaescu, R., Reynolds, D. A. and Godfrey, J. J.** (2003) Modeling prosodic dynamics for speaker recognition. *Proceedings IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, volume IV, Hong Kong, April 2003. pp. 788-791.
- Batliner, A., Steidl, S. and Nöth E.** (2007) Laryngealizations and emotions: how many babushkas? *Proceedings of The International Workshop on Paralinguistic Speech*, August 3, Saarbrücken. pp. 17-22.
- Beke A.** (2008) A felolvasás és a spontán beszéd alaphang-szerkezetek vizsgálata. In: Gósy M. (szerk.) *Beszédkutatás 2008*. Budapest: MTA Nyelvtudományi Intézet. 93-108.
- Bishop, Ch. M.** (1995) *Neural Networks for Pattern Recognition*. Oxford: Clarendon Press.
- Bőhm T.** (2006) A glottalizáció szerepe a beszélő személy felismerésében. In: Gósy M. (szerk.) *Beszédkutatás 2006*. Budapest: MTA Nyelvtudományi Intézet. 197-208.
- Bőhm T.** (2007) Beszélőfelismerés – neurológiai háttér és pszichológiai modellek. *Magyar Pszichológiai Szemle* 2007/62. 541-563.
- Campbell, J. P.** (1997) Speaker recognition: a tutorial. *Proceedings of the IEEE*, volume 85, September 1997. pp. 1437-1462.
- Cheveigné, A. de and Kawahara, H.** (2002) Yin, a fundamental frequency estimator for speech and music. *Journal Acoustical Society of America* 111/4. pp. 1917-1930.
- Chmielewska, I.** (2004) *Prosody-based text-independent speaker identification method*. Poznan, Poland: Institute of Control and System Engineering, Poznan University of Technology.
- Cover, T. M. and Thomas, J. A.** (1991) *Elements of Information theory*. New York: John Wiley & Sons.
- Ezzaidi, H., Pouat, J. and O’Saughnessy, D.** (2001) Towards combining pitch and mfcc for speaker identification system. *Proceedings of European conference on speech communication and technology (EUROSPEECH)*, September 2001. pp. 2825-2828.
- Ezzaidi, H. and Rouat, J.** (2004) Pitch and MFCC dependent GMM models for speaker identification systems. *Electrical and Computer Engineering*, 2004 1/2-5 pp. 43-46.
- Gocsál Á.** (1998) Életkorbecslés a beszélő hangja alapján. In: Gósy M. (szerk.) *Beszédkutatás 1998*. Budapest: MTA Nyelvtudományi Intézet. 122-134.
- Gósy M.** (1999) Az egyéni hangszínezet és a beszélő felismerésének kísérleti-fonetikai megközelítése. *Magyar Nyelvőr*, 123/4. 424-438.
- Gósy M.** (2001) A testalkat és az életkor becslése a beszéd alapján. *Magyar Nyelvőr*, 125/4. 478-487.
- Gósy M.** (2004) *Fonetika, a beszéd tudománya*. Budapest: Osiris Kiadó.
- Gósy M.** (2008) Magyar spontánbeszéd-adatbázis – BEA. In: Gósy M. (szerk.) *Beszédkutatás 2008*. Budapest: MTA Nyelvtudományi Intézet. 194-208.
- Gósy M. és Nikléczy P.** (1999) A beszélő felismerése a beszéde alapján: elméleti háttér és módszertani megközelítések. In: Gósy M. (szerk.) *Beszédkutatás 1999*. Budapest: MTA Nyelvtudományi Intézet. 1-19.
- Hautamäki, R. E. G.** (2005) *Fundamental Frequency Estimation and Modeling for Speaker Recognition*. Master’s thesis. University of Joensuu, Department of Computer Science.
- Jauquet, F., Verlinde, P. and Vloeberghs, C.** (1997) Histogram classifiers using vocal tract and pitch information for text-independent speaker identification. *ProRISC 9th Annual Workshop on circuits, systems and signal processing*, 1997. pp. 213-218.
- Labutin, B., Koval, S. and Raev, A.** (2007) Speaker identification based on the statistical analysis of f0. *Proceeding, IAFPA 2007*, UK, Plymouth: The College of Mark & St John.
- Lindh, J.** (2006) Preliminary Descriptive F0-statistics for Young Male Speakers. *Working Papers* 52. Lund University. 89-92.
- Markó A.** (2006) Nonverbális vokális jelek a társalgásban. In: Gósy M. (szerk.) *Beszédkutatás 2006*. Budapest: MTA Nyelvtudományi Intézet. 57-69.
- Rose, P.** (2002) *Forensic Speaker Identification*. London: Taylor and Francis.
- Von Kriegstein, G.** (2006) Implicit multisensory associations influence voice recognition. *PLoS Biology*, 4/10. pp. 1809-1820.

Van Lancker, D., Kreiman, J., and Emmorey, K. (1985) Familiar voice recognition: patterns and parameters; Part I. *Journal of Phonetics*, 13. pp. 19-38.

A Modern Filológiai Szemle szerkesztőinek felhívása

2009. február 18.

Tisztelettel felhívjuk a figyelmüket egy újonnan induló folyóiratra: a *Modern Filológiai Szemlére*. A folyóirat kiadója az MTA Modern Filológiai Társasága, és nyit mind a nyelvészet, mind pedig az irodalomtudomány felé. Két formában fog megjelenni: online módon (már készül a honlapja), illetve évente négyszer hagyományos, papíros folyóiratként is. A szerkesztőségi bizottság neves tudósokból áll – ami garancia a minőségre és a színvonalra is.

Örömmel fogadjuk tanulmányaikat, cikkeiket, recenzióikat. E-mailt a szerkesztőségi címre írhatnak: mfszemle@gmail.com.

Minden felmerülő kérdésre szívesen válaszolunk. Kérjük, továbbítsák e felhívást ismerőseiknek, tanítványaiknak is, akiket érdekelhet ez a lehetőség.

Tisztelettel,

Pál Dániel Levente, Veszelszki Ágnes
szerkesztők