A GLOTTALIZÁLT MAGÁNHANGZÓK AUTOMATIKUS OSZTÁLYOZÁSA SPONTÁN MAGYAR BESZÉDBEN

Beke András — Heltovics Éva

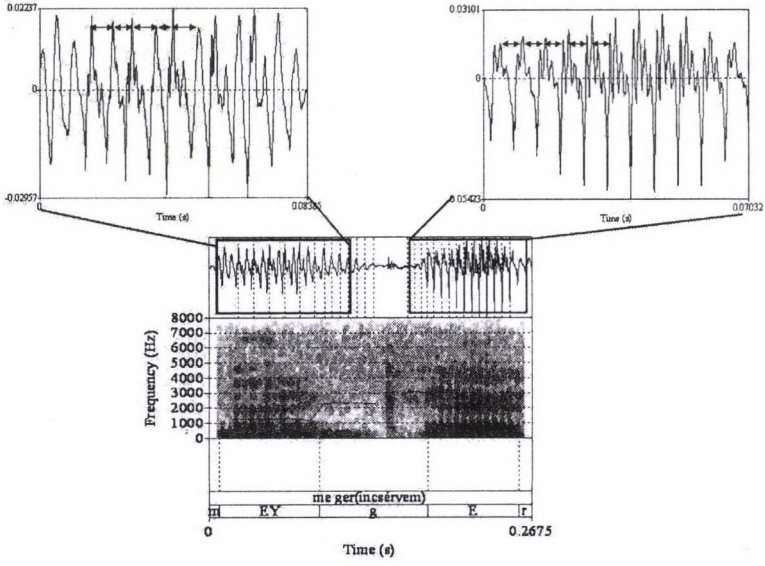
Bevezetés

A zöngeképzés során a hangszalagok általában közel regulárisan (kváziperiodikusan) rezegnek. A rezgés néha irregulárissá válik, ami azt jelenti, hogy a hangszalagok nem állandó időközönként, hanem periódusról periódusra változó szakaszokban csapódnak össze, egyes periódusok kimaradozhatnak (Bőhm–Ujváry 2008). Ezt a jelenséget nevezzük glottalizációnak. Az irregularitás megjelenhet a rezgés pillanatnyi frekvenciájában, amikor az alapfrekvencia hirtelen a beszélő jellemző hangterjedelme alá csökken, jelentkezhet a rezgés amplitúdójában, avagy a frekvenciájában és az amplitúdójában egyaránt (Bőhm 2006) (1. ábra).

A frekvenciában és az amplitúdóban történő ingadozás mértékének elég nagynak kell lennie ahhoz, hogy a hallgató számára észlelhetően eltérjen a reguláris zönge hangzásától. Olyan határértéket azonban nem tudunk adni az alapfrekvencia és az amplitúdó ingadozására, amely objektíven elválasztaná a reguláris hangszalagrezgést az irreguláristól. Ha ez az érték megadható lenne, akkor ez megoldaná a glottalizáció automatikus detektálását, és így a kézi címkézés szükségtelenné válna (Bőhm–Ujváry 2008).

A glottalizált beszédet a percepció érdes, rekedtes hangként azonosítja. A szakirodalomban számos hasonlattal írták le az általa kiváltott érzetet. Hasonlították például forró olaj sercegéséhez, kukorica pattogásához és egy fémkerítésen végighúzott rúd hangjához (Bőhm–Ujváry 2008). Éppen az általa kiváltott érzet miatt a glottalizációt egyes források *recsegő, érdes, rekedtes, nyikorgó zöngé*nek, mások *laringalizált, csikorgó beszéd*nek nevezik. Találkozhatunk olyan említésével is, hogy *irreguláris fonáció*. Az angol szaknyelv is számos különböző kifejezést használ a jelenségre, pl. *creaky voice, vocal fry, pulsed phonation, laryngealization, glottalization* (Slifka 2006). A glottalizáció produkcióját Laver (1980) a hangszalagok szoros összenyomásával magyarázza. A hangszalagok közötti erős zár a rezgést azért teszi instabillá, mert a tüdőből kipréselt levegő ezt csak ritkábban és rövidebb ideig tudja felfeszíteni, és akkor se a hangszalagok teljes hosszában. Így a reguláris zöngeképzéshez képest jelentősen kevesebb levegő áramlik át a hangrésen időegység alatt. A hangszalagok széthúzásával is lehet azonban irreguláris

alaphangot kelteni (Slifka 2006). Ebben az esetben az átlagos légáram nem csökken, hanem erősödik a reguláris fonációéhoz képest.



1. ábra
Példa az irreguláris (balra) és a reguláris (jobbra) fonációra

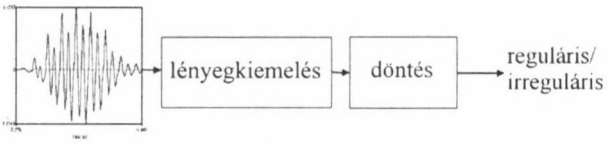
Minthogy a glottalizáció érzékelhető a hallgatók számára, ezért szerepet játszik a beszélő személy felismerésében; a glottalizáció gyakorisága jellemző továbbá az egyes beszélőkre, mégpedig az átlagos alapfrekvenciánál kisebb, de a hangterjedelemhez hasonló mértékben (Bőhm 2006).

A felolvasott beszéd különféle típusaira (például újságfelolvasás, hírbemondás, időjárás-jelentés) már léteznek olyan felismerő (speech-to-text) rendszerek, amelyek sokszor 90%-os szótalálási eredménnyel fordítják át a beszédet szöveggé (Mihajlik et al. 2006). A beszédfelismerő rendszerek eredménye azonban a spontán beszéden drasztikusan romlik (Furui 2007). Az eredmények romlását az okozza, hogy az akusztikai és nyelvi modelleket általában az írott nyelvtan szabályaiból és a felolvasott szövegek nyelvéből építik ki, a spontán beszéd és a felolvasás pedig jelentősen különbözik mind akusztikailag, mind nyelvtanilag (Furui 2005). Az akusztikai és a nyelvtani jelenségek sokkal kevésbé szabálykövetően, ill. kevésbé előre jelezhetően jelennek meg a spontán beszéd közben, mint az előre megfogalmazott beszédben. A spontán beszéd felismerésében ezért olyan fonetikai variációkkal és a

spontán beszéd jellegéből adódó egyéb sajátosságokkal is számolni kell a nyelvi modellben, amelyek mint elkülönült egységek vesznek részt a felismerésben (Beke–Szaszák 2009). Ilyen jelenség lehet a glottalizáció is.

Bár már régóta ismert a glottalizáció a szakirodalomban, a jelenleg alkalmazott beszédtechnológiák általában nem integrálják rendszereikbe. Ennek valószínűleg az az oka, hogy korábban úgy tekintettek rá, mint a beszédben igen ritka, elhanyagolható jelenségre (Bőhm 2009). Mára már azonban egyre több eredmény utal arra, hogy ez a zöngeképzési mód viszonylag gyakori és nem tekinthetjük marginális jelenségnek (Markó 2005). Számos szakirodalmi előzmény úgy szól a glottalizációról, mint a számítógépes beszédelemzést megnehezítő jelenségről: a dallammenetek kiszámításakor megakadályozza a hanglejtésformák kinyerését, és kényszerített beszédfelismerésnél a hanghatárok hibás bejelölését okozza (Bőhm–Ujváry 2008). Megfelelően kezelve azonban a beszédtechnológia számos területén hasznosítható lehet. A mesterséges beszédre ültetett prozódia természetesebb hangzású lehet, ha figyelembe vesszük a glottalizációt, illetve hitelesebben kifejezhetők az egyes érzelmek, mivel a glottalizáció hozzájárul az egyes érzelmi töltetek akusztikai jelöléséhez (Gobl–Ní Chasaide 2003).

A beszédhangok zöngeminőségének gépi osztályozása esetén a reguláris, irreguláris és zöngétlen beszédhangokról hoz döntést az algoritmus. Az osztályozás alapvetően két részre osztható: lényegkiemelés és döntés. Az osztályozó algoritmus a bemenő akusztikai jelből kiemeli a lényeges akusztikai jellemzőket, és a tanító halmazon elsajátított szabályok révén döntést hoz a zönge minőségéről (2. ábra).



2. ábra
A reguláris/irreguláris fonáció osztályozójának sematikus rajza

A nemzetközi és a hazai kutatásokban számos különböző akusztikai jellemzőt és osztályozó algoritmust alkalmaztak már. Surana és Slifka (2006) rendszerében az alapfrekvencia, a normalizált RMS intenzitás (normalized RMS intensity, NRMS), a simított energiakülönbség (Smoothed Energy Difference, SED) és az eltoláskülönbség-amplitúdó (Shift-Difference amplitude, SD) alapján történik az osztályozás. A döntést szupport vektor géppel (Support Vector Machine, SVM) végezték, az osztályozó tanításához és teszteléséhez a TIMIT beszédkorpuszt alkalmazták. Az általuk elért pontosság: 91,25% találati arány és 4,98% téves riasztás. Ishi és munkatársai (2008) osz-

tályozója glottális impulzusokról reguláris/irreguláris/bizonytalan döntést hozott a következő akusztikai jellemzők alapján: energiacsúcs-emelkedés és -ereszkedés (PoWer Peak falling and rising, PWP), kereten belüli periodicitás (IntraFrame Periodicity, IFP), impulzusok közötti hasonlóság (InterPulse Similarity, IPS). A döntést küszöbértékekre alapozták. Vizsgált anyaguk a JSP/CREST ESP korpuszból került ki. Osztályozójuk találati aránya 74%, téves riasztási aránya pedig 13%. Vishnubhotla és Espy-Wilson (2007) az átlagos magnitúdókülönbség függvényt (Average Magnitude Difference Function, AMDF), a nullátmenetek számát (Zero-Crossing Rate, ZCR), a spektrum lejtését (spectral slope) és az autokorreláció-függvény csúcsértékét (F_0 autocorrelation peak value) vették figyelembe. A döntés küszöbértekeken alapszik. A tanító és a tesztelő halmazok a TIMIT és a NIST 98 korpuszokból kerültek ki. Eredményeik a következők: a TIMIT-korpuszon 91,8% találat és 17,4% téves riasztás; a NIST 98 korpuszon pedig 91,5% találat és 12,8% téves riasztás. Yoon és munkatársai (2006) szintén az autokorrelációs függvény csúcsértékét, valamint az első és második harmonikus közötti amplitúdókülönbséget vették figyelembe, és reguláris/irreguláris döntéssel dolgoztak. A döntést küszöbértékekre alapozták. Korpuszukat a Switchboardból válogatták össze. Osztályozójuk 69,23%-os pontosságot ért el (a szerzők sem találati, sem tévesztési arányt nem közöltek). Kiessling és munkatársai (1995) kétféle osztályozót hoztak létre. Az egyik egy kevert Gauss-modell (GMM: Gaussian Mixture Model) osztályozó, amely az egyes hangokról reguláris/irreguláris/zöngétlen döntést hozott a beszédjel kepsztrális jellemzői alapján. Ennek eredménye 80%-os, a téves riasztás 8%-os volt. A másikban neurális hálózatot (ANN: Artificial Neural Network) és LPC-alapú lényegkiemelést alkalmazott, amelynek eredménye 65%-os volt 12%-os téves riasztás mellett.

Bőhm (2009) egy rendszerbe integrálta a Surana, valamint Ishi és munkatársai osztályozója által használt akusztikai jellemzőket. Az akusztikus jellemzők értékelésére ROC-görbéket alkalmazott. Ez az elemzés átfogóbb képet ad a megelőző tanulmányokban alkalmazott eljárásoknál, és lehetővé teszi az egyes jellemzők reguláris-irreguláris szeparációs képességének a választott küszöbtől független vizsgálatát és számszerűsítését, a jellemzők eloszlásával kapcsolatos előfeltevések nélkül. Az osztályozó magánhangzókon hoz reguláris/irreguláris döntést SVM segítségével. A teszthalmaz és a tanítóhalmaz a TIMIT-korpuszból származik. Az általa elért pontosság: 98,85%-os találati arány és 3,47%-os téves riasztási arány. Bőhm osztályozó algoritmusa angol nyelvű korpuszt elemez, a hangsúlyos helyzetben lévő magánhangzókról hoz döntést. A döntés hangszintű, így a magánhangzók határainak beállítása nehézkes.

A jelen kutatás célja, hogy MFCC-vel előfeldolgozva (mel-frekvenciás kepsztrális komponenseket) HMM-ekkel (rejtett Markov-modellel) reguláris/irreguláris/zöngétlen beszédhang automatikus osztályozót hozzunk létre elsőként magyar nyelvű spontán beszédre.

Anyag, módszer, kísérleti személyek

A BEA adatbázisból (Gósy 2008) 11 fiatal beszélő (5 férfi és 6 nő) spontán beszédét vizsgáltuk; életkoruk átlaga 25 év (a legfiatalabb 22, a legidősebb 30 éves). A spontán narratíváik átlagos időtartama 1,8 perc volt (szórása 0,9 perc), összesen több mint 40 percnyi anyagot elemeztünk. A hanganyagokat először szakaszszinten címkéztük fel a Praat szoftver alkalmazásával (Boersma–Weenink 2009); majd automatikus szegmentálóval (MAUS szoftver: ftp://ftp.bas.uni-muenchen.de/pub/BAS/SOFTW/MAUS) hangszinten annotáltuk. Összesen 16 659 beszédhangot címkéztünk fel. A gépi annotálást – ahol szükséges volt – manuálisan korrigáltuk.

Az irreguláris magánhangzókat kézzel jelöltük be. A címkézést a nemzetközi és a hazai kritériumok mentén végeztük akusztikai és auditív információk alapján. Az akusztikai kritérium szerint egy magánhangzó irreguláris képzésű, ha az alapperiódusok időtartama vagy amplitúdója hirtelen, periódusról periódusra jelentős változásokat mutat. Ugyancsak teljesíti az akusztikai kritériumot az a magánhangzó, amelyben az alapfrekvencia hirtelen jelentősen lecsökken. Az auditív kritérium szerint akkor irreguláris egy magánhangzó, ha a hangszínezet érzékelhetően megváltozik, a beszéd érdessé, rekedtessé válik. A következetes címkézés érdekében a korpusz összes felvételén mindkét szerző egy rögzített kritériumrendszer alapján megjelölte az irreguláris beszédszakaszokat, majd a párhuzamos címkeállományokat összehasonlították, és kialakítottuk a végleges címkéket.

Az irreguláris magánhangzók osztályozásához 13 együtthatós MFCC-t és azok első és második deriváltját használtuk (39 jellemzővektort). Az irreguláris és a reguláris magánhangzók osztályozásához rejtett Markov-modellt alkalmaztunk. Az osztályozó tanításához összesen 7104 magánhangzót használtunk, ebből 5111 reguláris fonációjú, és 1993 irreguláris fonációjú. A tanításhoz ennek kétharmadát, a teszteléshez egyharmadát használtuk. Minden modellt 2, 4, 8, 16 Gauss kibocsátási valószínűséget leíró függvénnyel tanítottuk.

Az eredmények kiértékelésére több értéket is megadtunk. Az adott elemeket két osztályba soroljuk: pozitív és negatív. Ezeket az elemeket valamilyen osztályozóval besoroltatjuk. A kézi és az automatikus címkézés azonosságából és különbségeiből áll elő a klasszifikációs mátrix (1. táblázat).

1. táblázat: Klasszifikációs mátrix a zöngeminőségre értelmezve

		Kézi címkézés	
		Irreguláris	**Reguláris**
Automatikus címkézés	**Irreguláris**	TP helyes (irreguláris)	TF elsőfajú hiba
	Reguláris	FN másodfajú hiba	TN helyes besorolás (reguláris)

A felismerési arányt a következőképpen számoltuk ki,

$$\text{Helyes felismerési arány} = \frac{N-(TF+FN)}{N},$$

ahol az N az elemek száma, a TF az elsőfajú hiba és az FN a másodfajú hiba elemszáma.

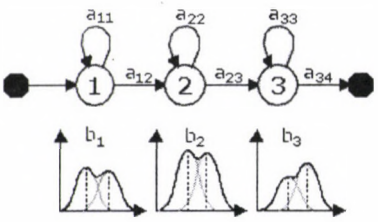
Az elsőfajú hibának azt tekintjük, amikor egy hangot irregulárisnak minősít az osztályozó, pedig az reguláris. Másodfajú hiba alatt azt értjük, amikor egy hangot regulárisnak minősít az osztályozó, pedig az irreguláris.

A rejtett Markov-modell

Az automatikus beszédfelismerésben az egyes beszédhangok modellezésére elterjedten használják a rejtett Markov-modelleket. A beszédfelismerő fontos tudásbázisa a beszédhangok modelljein kívül még a szótár és a nyelvi modell (vö. Rabiner 1989; Mihajlik et al. 2006). A beszédhangok modelljeinek szerepe az akusztikai beszédjel megfeleltetése, leképezése az egyes beszédhangoknak(/ra). A jó megfeleltethetőséghez az akusztikai beszédjelet előzetesen feldolgozzák (előfeldolgozás), például az emberi hallást is modellező Mel-frekvenciás kepsztrális együtthatókra (angolul: Mel Frequency Cepstral Coefficients) való átalakítás, de más eljárások is ismeretesek (pl. PLP perceptual linear prediction). A MFC-ket az alábbi módon számíthatjuk ki: elsőként gyors Fourier-transzformáció történik (Fast Fourier Transform, röviden FFT). Ekkor a beszédből egy rövid időtartamú részt (jellemzően 25 ms hosszú darabot) kivágunk, majd egy ún. ablakfüggvénnyel súlyozzuk, és Fourier-transzformációval meghatározzuk a spektrumát. Ezután 10 ms-ot továbblépünk, és ugyanezt ismételjük mindaddig, amíg el nem érünk a feldolgozandó szakasz végéig. Második lépésként a spektrumokat a Mel-skála szerint, ún. kritikus frekvenciasávok szerint bontjuk fel. A szűrősor általában 20 sávátereszto szűrőből áll, amelyek kimenetén egy, a sávba eső intenzitással arányos számszerű érték jelenik meg, azaz tulajdonképpen 10 ms-onként egy 20 dimenziós vektort kapunk. Mivel ezek az értékek egymással korrelálnak, ezért a dimenziószám csökkenthető, mégpedig egészen 12-re, az ún. diszkrét koszinusztranszformáció segítségével (Discrete Cosine Transform, DCT). Ezután általában hozzáveszik a vektorhoz a teljes beszédjelszelet átlagos energiáját, majd az így összesen 13 érték első és második deriváltjait, így összesen egy 39 dimenziós vektort, ún. jellemzővektort kapunk.

A beszédhangok rejtett Markov-modelljei lényegében a hangra jellemző vektorok eloszlását adják meg. Figyelembe véve a jellemzővektorok spektrális származtatását, ez tehát frekvenciatartománybeli modellezést jelent. A Markov-modellek leggyakrabban 3 állapotú, balról jobbra felépítésű modellek – az utóbbi azt jelenti, hogy a Markov-modell egyes állapotai között átmenet csak balról jobbra lehetséges (3. ábra). Minden állapothoz tartozik tehát egy valószínűségi eloszlást megadó függvény, amelyet statisztikai eszköztárral, leg-

gyakrabban normális eloszlások szuperponálásával becslünk a modell ún. betanítása során.



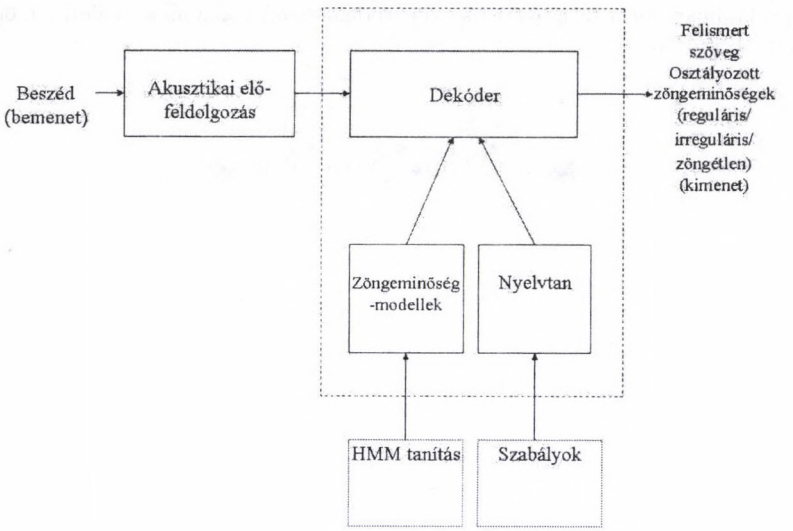
3. ábra
3 állapotú, balról jobbra felépítésű Markov-modell

A modell tanítás során becsülendő paraméterei az állapotátmeneti valószínűségek (a_{ij}), valamint a normális eloszlások súlya, várható érték- és szórásvektora, amelyek együttesen megadják az eloszlást [$b_j(t)$].

A betanítás során ezeknek a függvényeknek a paramétereit becsüljük meg. A beszédfelismerés során pedig a beszédből előállított jellemzővektorokat hasonlítjuk az egyes beszédhangok állapotainak megfelelő eloszlásokhoz. Minél inkább illeszkedik a jellemzővektor egy adott állapot eloszlásához, annál nagyobb súlyt rendel a hozzá kapcsolódó útvonalhoz a dekóder, azaz a tulajdonképpeni beszéd-szöveg átalakító.

A Markov-modellek a beszédfelismerőben valójában kettős feladatot látnak el, a jellemzővektorok osztályozása mellett a beszédjelet illesztik is a neki megfelelő beszédhangsorozatra, azaz meghatározzák az egyes beszédhangok kezdő és végidőpontjait. A Markov-modellek úgy is használhatók, hogy előre elkülönített osztályokra betanított modellek alapján osztályozzanak. Ilyen esetben a beszédfelismerőben használatos szótár szerepét az egyes osztályok listája, a nyelvtan szerepét pedig az osztályozás szabályai veszik át. A beszédhangok osztályozása esetén a lista az osztályozni kívánt beszédhangokból áll, az osztályozás szabályai pedig megadják, milyen beszédhangokat milyen sorrendben lehet illeszteni.

A szerzők *a)* a magánhangzók zöngeminőségének osztályozását (reguláris vagy irreguláris magánhangzó), illetve *b)* a magánhangzón belül irreguláris/reguláris és zöngétlen mássalhangzó osztályozását valósították meg rejtett Markov-modellekkel. A HTK környezetben megvalósított osztályozó felépítése a 4. ábrán látható.



4. ábra
A zöngeminőséget osztályozó gépi rendszer

Eredmények

Az általunk létrehozott rendszert 400 irreguláris magánhangzót tartalmazó korpuszon teszteltük.

a) Az első osztályozó modelljeit a magánhangzókra építettük. A nem magánhangzókra külön úgynevezett „szemétmodellt" készítettünk, amelyet a felismeréskor kiszorítottunk és a tesztelésnél a modellek közül kihagytuk. A magánhangzókra két modellt építettünk a zöngeminőségüknek megfelelően: reguláris/irreguláris. A legjobb eredményt a 16 Gauss kibocsátási valószínűséget leíró függvénnyel értük el. Az összes elem helyes felismerési aránya 95,7%-os. Ezen belül a reguláris magánhangzókat 95%-ban, míg az irreguláris magánhangzókat 99,2%-ban osztályozta helyesen az algoritmus (2. táblázat).

2. táblázat: A reguláris/irreguláris zöngeminőséget osztályozó eredménye (%)

		Kézi címkézés	
		reguláris	**irreguláris**
Automatikus címkézés	**reguláris**	**95,0**	5,0
	irreguláris	0,8	**99,2**

Az osztályozó tehát a 400 irreguláris beszédhangból 3,2-t téveszt el. A téves riasztás 5%, ekkor az algoritmus a beszédben reguláris magánhangzót irregulárisnak osztályozza. Az irreguláris mellett a reguláris beszédhangokat a megfelelő zöngeminőségbe sorolja az algoritmus, ennek az eredménye 95%.

b) A második osztályozó modelljeit az összes beszédhangra készítettük el a felismerés folyamán. A beszédhangokra három modellt építettünk a zöngeminőségüknek megfelelően: reguláris/irreguláris/zöngétlen. A legjobb eredményt a 2 Gauss kibocsátási valószínűséget leíró függvénnyel értük el, amely az összes elemre kapott eredménye 88,0%. A reguláris magánhangzókat 90,0%-os, az irreguláris beszédhangokat 92,9%-os, míg a zöngétlen beszédhangokat 82,7%-os eredménnyel osztályozta az algoritmus (3. táblázat).

3. táblázat: A reguláris/irreguláris/zöngétlen zöngeminőséget osztályozó eredménye (%)

		Kézi címkézés		
		reguláris	**irreguláris**	**zöngétlen**
Automatikus címkézés	**reguláris**	**90,0**	8,2	1,8
	irreguláris	6,3	**92,9**	0,8
	zöngétlen	7,3	10,0	**82,7**

Ha az összes beszédhangot modellezzük a zöngeminőségük függvényében, akkor az osztályozás helyessége csökken, de még mindig igen magas marad. A helyesen felismert irreguláris beszédhangok aránya csökkent; 6,3%-ban a reguláris beszédhangokkal téveszti össze az osztályozó. Az irreguláris zöngeminőséggel képzett hangok 10%-át a zöngétlen elemekkel téveszti össze az algoritmus, ami az akusztikai jegyek hasonlóságának tulajdonítható.

c) A harmadik osztályozóval modelleztük az egyes magánhangzó-minőségeket a reguláris és az irreguláris zöngéjű hangok esetében. Az elemzésben csak azokat a magánhangzókat modelleztük, amelyeknek az elemszáma elegendő volt. A modellek a következők voltak: a reguláris fonáció esetében: [ɔ, ɛ, i, o, u]; az irreguláris fonáció esetében: ɔY, ɛY, iY, oY, uY. A legjobb osztályozási eredményt a 4 Gauss kibocsátási valószínűséget leíró függvénnyel értük el (4. táblázat).

Mind a reguláris, mind az irreguláris fonáción belüli magánhangzó-minőségek osztályozási eredménye átlagosan 62%. Az osztályozási eredmények tehát annak ellenére, hogy irreguláris fonációval realizálódott a magánhangzó, megközelítették a reguláris fonációval létrejött magánhangzók eredményét.

4. táblázat: A magánhangzó-minőséget és a zöngeminőséget osztályozó algoritmus eredménye (%)

		Kézi címkézés									
		oY	ɔY	ɛY	iY	uY	ɔ	ɛ	o	u	i
Automatikus címkézés	oY	**52,6**	9,6	2,6	3,5	2,6	7,9	1,8	5,3	1,8	12,3
	ɔY	0,0	**90,7**	1,9	1,9	1,9	0,0	3,7	0,0	0,0	0,0
	ɛY	1,8	9,6	**60,5**	0,9	3,5	7,0	4,4	1,8	2,6	7,9
	iY	3,7	3,7	0,0	**63,0**	3,7	7,4	3,7	7,4	0,0	7,4
	uY	14,3	42,9	0,0	0,0	**28,6**	0,0	14,3	0,0	0,0	0,0
	ɔ	3,0	7,2	2,7	2,1	1,2	**73,4**	2,1	2,1	1,8	4,5
	ɛ	4,4	13,1	4,4	2,6	1,5	8,4	**51,5**	3,5	2,3	8,4
	o	2,4	12,6	3,9	2,9	2,9	5,8	6,3	**51,2**	3,4	8,7
	u	6,9	12,1	6,9	3,4	1,7	6,9	0,0	3,4	**51,7**	6,9
	i	1,7	3,4	3,9	1,1	0,0	2,8	1,7	1,1	1,7	**82,7**

Következtetések

Tanulmányunkban az irreguláris fonációval képzett magánhangzókat MFCC jellemzőik alapján HMM-ekkel modelleztük magyar nyelvű spontán beszédben. Az irreguláris fonációval képzett magánhangzókat MFCC-vel előfeldolgozva HMM-ekkel 99%-os találati aránnyal és 5% téves riasztási aránnyal tudtuk automatikusan osztályozni. A bemutatott MFCC-vel előfeldolgozott HMM-ekkel történő zöngeminőség osztályozó előnyei a következők.

a) Képes minden magánhangzót (hangsúlyhelyzettől függetlenül) és irreguláris beszédrészletet osztályozni.

b) A bemenő beszédjelhez a tanítás után nem szükséges hangszintű címke, mivel a HMM az osztályozás során megkeresi a beszédhangok határait is.

c) Az irreguláris fonációt nemcsak a reguláris zöngétől, hanem a zöngétlen beszédtől is el tudja különíteni.

d) A magánhangzó-minőséget is felismerve képes dönteni a zöngeképzési módról.

Eredményeink csak nagyon általánosan hasonlíthatóak más eddig elkészített osztályozóval, mivel sem a korpusz, sem az akusztikai jellemzők, sem az osztályozó algoritmus nem azonos. A létrehozott automatikus irreguláris osztályozó eredményei még kezdetiek, amelyeken nagyobb korpusz, trifónos modellezés, illetve más akusztikai jellemzők bevonásával valószínűleg javítani lehet.

Irodalom

Boersma, Paul – Weenink, David 2009. *Praat: Doing phonetics by computer* (Version 5.1). http://www.fon.hum.uva.nl/praat/

Bőhm Tamás 2006. A glottalizáció szerepe a beszélő személy felismerésében. *Beszédkutatás* 2006. 197–207.

Bőhm Tamás – Ujváry István 2008. Az irreguláris fonáció mint egyéni hangjellemző a magyar beszédben. *Beszédkutatás* 2008. 108–120.

Bőhm Tamás 2009. *Analysis and modeling of speech produced with irregular phonation*. PhD-értekezés. BME, Budapest.

Furui, Sadaoki 2005. Recent progress in corpus-based spontaneous speech recognition. In *IEICE-Transactions on Information and Systems E88-D*. 366–375.

Furui, Sadaoki 2007. Recent advances in automatic speech summarization. Evans, David – Furui, Sadaoki – Soul, Chantal (eds.): *Proceeding of the IEEE/ACL Workshop on Spoken Language Technology 2007*. Los Alamitos, 115–122.

Gobl, Christer – Ní Chasaide, Ailbhe 2003. The role of voice quality in communicating emotion, mood and attitude. *Speech Communication* 40. 189–212.

Gósy Mária 2008. Magyar spontán beszéd adatbázis – BEA. *Beszédkutatás* 2008. 194–207.

Ishi, Carlos T. – Sakakibara, Ken I. – Ishiguro, Hiroshi – Hagita, Norihiro 2008. A method for automatic detection of vocal fry. *Audio, Speech, and Language Processing. IEEE Transactions 16/1*. 47–56.

Kiessling, Andreas – Kompe, Ralf – Niemann, Heinrich – Nöth, Elmar – Batliner, Anton 1995. Voice source state as a source of information in speech recognition: detection of laryngealizations. In Rubio Ayuso, Antonio J. – Lopez Soler, Juan M. (eds.): *Speech recognition and coding – New advances and trends*. Springer-Verlag, Berlin, 329–332.

Markó Alexandra 2005. *A spontán beszéd néhány szupraszegmentális jellegzetessége*. PhD-értekezés. ELTE, Budapest.

Mihajlik Péter – Fegyó Tibor – Tatai Péter 2006. Új eljárás a gépi beszédfelismerés környezetfüggő beszédhangmodelljeinek kialakítására. *Beszédkutatás 2006*. 218–230.

Rabiner, Lawrence R. 1989. A tutorial on hidden Markov models and selected applications in speech recognition. *Proceedings of the IEEE* 77/2. 257–286.

Slifka, Janet 2006. Some physiological correlates to regular and irregular phonation at the end of an utterance. *Journal of Voice* 20/2. 171–186.

Surana, Kushan – Slifka, Janet 2006. Acoustic cues for the classification of regular and irregular phonation. In *Interspeech 2006*. 693–696.

Vishnubhotla, Srikanth – Espy-Wilson, Carol 2007. Detection of irregular phonation in speech. In Trouvain, Jürgen–Barry, William J. (eds.): *16th International Congress of Phonetic Sciences. Proceedings*. Pirrot GmbH., Dudweiler, 2053–2056.

Yoon, Tea-Jin – Zhuang, Xiaodan – Cole, Jennifer – Hasegawa-Johnson, Mark 2006. Voice quality dependent speech recognition. In Tseng, S. (ed.): *International Symposium on Linguistic Patterns in Spontaneous Speech*. Academia Sinica, Taipei, Taiwan. https://netfiles.uiuc.edu/tyoon/www/docs/Yoon-etal-LPSS.pdf