Érzelmek automatikus osztályozása spontán beszédben

Sztahó Dávid, Imre Viktor, Vicsi Klára

Budapest Műszaki és Gazdaságtudományi Egyetem
Távközlési és Médiainformatikai Tanszék, Beszédakusztikai Laboratórium
1111 Budapest, Stoczek utca 2.
sztaho@tmit.bme.hu, imreviktor.bmevik@gmail.com,
vicsi@tmit.bme.hu

Kivonat: A Budapesti Műszaki és Gazdaságtudományi Egyetem Beszédakusztikai Laboratóriumában automatikus érzelemfelismerésre, valamint automatikus beszéddetekcióra, illetve beszédszegmentálásra irányuló vizsgálatok folynak. A cikk ismerteti az érzelem felismerése során felhasznált különböző akusztikai jellemzőkkel kapott eredményeket, valamint a szupport vektor gép alapú gépi tanulási eljáráshoz használt spontán beszédet tartalmazó adatbázisokat. A beszéddetektálás, illetve beszédszegmentálás eredményeinek bemutatása során ismertetjük a rejtett Markov-modelleken alapuló felismerési eljárást, valamint a felhasznált telefonos adatbázist. Célunk egy olyan detektáló eljárás kidolgozása, amelyet alkalmazva, a szegmentált beszéden a fentebb említett érzelmi osztályozást el tudjuk végezni.

1 Bevezetés

Az automatikus érzelemfelismerés összetett probléma. Ahhoz, hogy valós időben meg lehessen valósítani, magán az érzelemfelismerésen kívül a beszéd valós idejű detektálásával és szegmentálásával is szembe kell nézni. Ennek a problémának a megoldása szintén kritikus fontosságú, ugyanis az előre elkészített és megfelelő beszédegységekkel betanított érzelemfelismerő működése e nélkül nem megvalósítható.

Ezért a Budapesti Műszaki és Gazdaságtudományi Egyetem Beszédakusztikai Laboratóriumában automatikus érzelemfelismerésre, valamint automatikus beszéddetekcióra, illetve beszédszegmentálásra irányuló vizsgálatokat végzünk. Adatbázisokat hoztunk létre, szegmentáltunk, illetve annotáltunk, amelyekkel a fenti feladatok elvégzésére alkalmas rendszereket kísérleteztünk ki.

Az emberek érzelemfelismerési képessége nyolc érzelem esetén (hét érzelem + semleges) 60-65%-ra adódik abban az esetben, amikor a nyelvi tartalom a döntésben nem játszik közre [1]. Ennél jobb felismerési eredményt egy géptől sem várhatunk el. További kérdés, hogy a felismerésben milyen akusztikai jellemzők játszanak közre. A cikkben az irodalomban [2, 3] megtalálható alapvető jellemzőkön kívül egyéb spektrális jellemzőket is felhasználunk. A beszédfelismerésben leggyakrabban alkalmazott alapegység a szavak, illetve a mondatok szintje. Az általunk választott alapvető időtartam azonban a korábbi eredményeink alapján [4] a frázis. Ezen belül kívánjuk az

érzelmeket felismerni. Ennek megfelelően az automatikus beszéddetektáló, illetve -szegmentáló eljárásnál is ekkora egységet tekintünk a felismerés alapegységének.

2 Beszéddetektálás

A valós idejű érzelemfelismerés problémája több összetevőből áll. Az audiojelben a spontán beszéd detektálása, valamint annak tagolása kiemelt tényező. Az általunk használt felismerési egység a frázis. Ebben a fejezetben bemutatjuk az automatikus beszéddetektáló eljárását, valamint a felhasznált adatbázist.

2.1 Telefonsávú felvételek beszéddetektáláshoz

A beszéddetektálási rendszer betanításához, teszteléséhez olyan beszédadatbázisra volt szükség, amely a felhasználási körülményekhez hasonló hanganyagot tartalmaz. A felhasznált adatbázist a BME Távközlési és Médiainformatikai Tanszék Beszédtechnológiai Laboratóriumának dolgozói és hallgatói készítették mobiltelefonnal. A felvételeket három különböző zajszintre lehet osztani. Vannak tiszta beszédjelet tartalmazó, nagyjából zajmentes környezetben készült felvételek. A zajjal terhelt beszélgetések további két részre bonthatóak: közepesen zajos, ahol a beszéd még jól érthető, de különböző háttérzajok fordulnak elő (autózaj, utcai zajos, háttérbeszéd); az erősen zajos felvételekben a beszéd már nehezen érthető.

1. táblázat: Felvételek száma osztályok szerint.

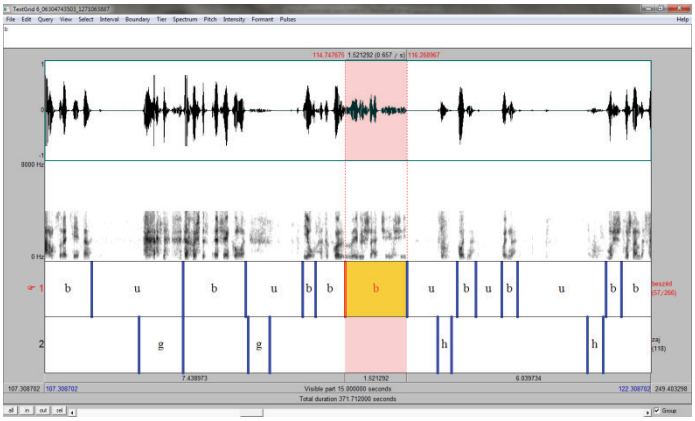
Zajszint	Felvételek száma
Alacsony	9
Közepes	16
Magas	6

2. táblázat: Alkalmazott jelölések az adatbázis annotálása során.

Sor neve	Hangtípus	Jelölés
beszéd	beszéd	b
	nem beszéd	u
zaj	gépjárműzaj	a
	gesztusok	g
	beszéd a háttérben	k
	szélzaj	s
	telefonhang	t
	recsegés	r
	sziréna	i
	ütés	h
	papírzörej	p
	levegővétel	l

A felvételek a felhasználás alapján is két csoportra oszthatóak: a kötött beszédet tartalmazó felvételek időben jól elkülönülő különálló mondatokat, míg a célzottan beszéddetektálásra készült felvételek egybefüggő, spontán beszédet tartalmaznak.

A felvételek annotálása során a frázisszintű címkézést a Praat szoftver felhasználásával végeztük el [5], amelyre egy mintát az 1. ábrán mutatunk be. A címkefájl két sort tartalmaz, a „beszéd” és „zaj” sávot. A beszédsávban a beszéd-nem beszéd részeket, és azok határait jelöltük. A zajsávban a különböző háttérzajokat és azok határait adtuk meg. A megkülönböztetett zajtípusokat a 2. táblázat tartalmazza.



1. ábra. Példa a kézi szegmentálásra.

2.2 Beszéddetektálási eljárás

Az automatikus felismerés rejtett Markov-modellek segítségével történt. Ehhez a HTK Toolkit-et [6] alkalmaztuk, amely egy beszédfelismerő keretrendszer, rejtett Markov-modell megvalósítással.

Az eljárás lényege, hogy a különböző zajtípusokra, valamint a beszéd (frázis) szakaszokra külön Markov-modelleket építünk, a 2.1. részben bemutatott adatbázis segítségével, amelyhez először egy akusztikai előfeldolgozást kell végezni. A beszéddetekció során, szintén egy akusztikai előfeldolgozás után, az egymás utáni időszakaszokra kapott legvalószínűbb Markov-modellek alapján lehetséges a beszéd szakaszok határainak bejelölése. Az eljárás erőssége az, hogy a felismert időszakasz hossza nem előre meghatározott, hanem változó hosszúságú lehet.

Az akusztikai előfeldolgozás során a következő jellemzőket használtuk fel a 3. táblázatban megadott számítási paraméterekkel. Ezután a kiszámított jellemzőket 50 ms-os ablakot alkalmazva kétszer deriváltuk. A végső tanítóvektorba az alapjellemzők, valamint az első, illetve második deriváltak kerültek.

A Markov-modellek építése során különböző hosszúságú (állapotszámú) modelleket alkalmaztunk a beszédre, valamint a zajokra. Előkísérletek alapján beszéd esetén 11 állapotú Markov-modelleket, zaj esetén 5 állapotú Markov-modelleket, valamint

csend esetén 3 állapotú Markov-modellek lettek elkészítve. Így a beszédrészeket az automatikus felismerő nem darabolja fel apró részekre, valamint a kevesebb állapotszámú zajmodellek segítségével a rövidebb időtartamú zajok is detektálhatóak.

3. táblázat: Felhasznált akusztikai jellemzők.

Jellemző	Időablak	Lépésköz
Alaphang	75 ms	10 ms
Intenzitás	250 ms	10 ms
Mel-frekvenciás kepsztrális együtthatók (MFCC)	500 és 250 ms	10 ms

A tanításra és tesztelésre következetesen elkülönített minták kerültek felhasználásra. Ez azt jelenti, hogy minden tesztet ugyanazon mintacsoporton végeztünk el, amelynek mintáit véletlenszerűen, de a változatosságot figyelembe véve válogattuk ki. Így extrém zajos, valamint normál minőségű, enyhén zajos (felhúzott ablak, kocsiban, nem kihangosítóval készült) minták is szerepeltek a tanító adatbázisban, valamint a tesztelő adatbázisban is.

A minőség kiértékelésére egy egyszerű, a döntést meggyorsító indexet használtunk. Két mátrixot számoltunk, melyekben beszúrási és tévesztési statisztikák szerepelnek. A beszúrási mátrix sorai azt mondják meg, hogy az eredetileg adott akusztikai osztálynak jelölt időintervallumok alatt hány darab jelölés található meg, tehát egy eredeti szakaszhoz mennyi felismert szakasz tartozik. A tévesztési mátrix sorai ehhez hasonlóan: az eredeti akusztikai osztály egyes intervallumaihoz mint (változó hosszúságú) időegységhez vesszük az ezen intervallumok alatt lévő jelölések időtartamát, tehát az eredeti szakaszokhoz időarányosan mennyi felismert időintervallum tartozik.

Ezek a mátrixok azonban bizonyos esetekben elég nagyok lehetnek, például sok címketípus esetén. Ez azzal a következménnyel jár, hogy nehezen átláthatóak, sok ideig tart, míg megállapítja valaki, hogy első közelítésben mennyire jó a felismerés. Ennek a kiküszöbölésére, az átláthatóság kedvéért egy egyszerű indexszámítást vezettünk be. Ez két részből áll: egyrészt az úgynevezett beszédindex, másrészt a zajindex. Ezeknek súlyozott összegéből adódik az összesített index, melyben a zajindex csak negyed súllyal szerepel. Ennek értelme az, hogy a végső felismerés céljából elhanyagolható, hogy a zajt milyen arányban találjuk el helyesen, ha a beszédet viszont annál jobban, mivel az automatikus felismerés végső célja a beszéd detektálása.

A beszédindex két összetevőből áll össze: beszúrási arány, valamint a tévesztési arány.

$$\text{beszúrási arány} = \frac{\text{az osztályban jól beszúrt intervallumok száma}}{\text{az osztály eredeti intervallumainak száma}}$$

$$\text{tévesztési arány} = \frac{\text{lefedett időintervallum száma}}{\text{az osztály eredeti intervallumainak száma}}$$

Látható, hogy a tévesztési arány maximuma 1, míg a beszúrási arány lényegében akármekkora lehet, így a beszédindexnek sem 100 a maximuma. Ahhoz, hogy legyen maximum, 100-nál törést kellett bevezetni, vagyis ha a beszúrási arány 1-nél nagyobb, akkor a beszédindexet maximalizáljuk. Az eredmények értékelésekor látható, hogy ez a változtatás a kiértékelhetőséget nem rontja. 80-as beszédindex körül már elfogadható felismerés adódik.

A későbbiekben egy, a zajos beszéd jelölésére szolgáló osztály ezt a számítási módot a következőképpen módosította: nem számít, hogy zajos beszéd és beszéd között mit döntünk, így ezeket ezután egyben kezeltük.

A zajindex az előzőekben elmondottakkal azonosan kerül kiszámításra az egyes zajokra, majd a végső index pedig ezeknek az átlaga. Az összesített index pedig:

$$\text{összindex} = \frac{3}{4} * \text{beszédindex} + \frac{1}{4} * \text{zajindex}$$

2.3 Eredmények

A tesztsorozat megkezdésekor a következő osztályok voltak felvéve tanításra: b (beszéd), u (csend/szünet), a (autózaj), g (gesztus), k (háttérbeszéd), s (szélzaj), t (telefonos jelzés), r (recsegés), i (sziréna).

A szirénahangot az első teszteléskor rögtön eltávolítottuk a tanított osztályok közül, mivel összesen egyetlen hangfájlban szerepelt, és abban is rövid ideig. A p (papírzörgés) és h (ütés/ütődés) hangokat a recsegéshez vontuk, elégtelen mennyiségű minta miatt, valamint a hangok akusztikai hasonlósága miatt. A tesztek során bevezettünk egy légzés címkét is, amely a telefonban jól hallhatóan a beszélőtől származó belégzési zörejeket fogja össze. Az 1. tesztsorozatban 100, 250, 500 és 750 milliszekundumos ablakokkal számolt mel-frekvenciás kepsztrális együtthatók, az intenzitás és az alaphang értékek szerepeltek, valamint ezek első, illetve második deriváltja. A legjobb eredményeket az 500 ms-os ablakmérettel számolt MFCC paraméterek esetén kaptuk (5. táblázat).

4. táblázat: Osztályokhoz rendelt Markov-modellek hossza.

Állapotszám	Címkék (osztályok)
11 állapotú modell	*b*, *k*
5 állapotú modell	*a*, *g*, *s*, *t*, *r*, *u*, *l*

A legrosszabb minőségű hangfájlok esetében (autóban, kihangosítóval) az osztályozási eredmények is rossz minőségűek lettek. Szinte egyáltalán nem ismert fel beszédet a rendszer ezekben a fájlokban. Ennek javítása érdekében bevezettünk egy zajos beszéd osztályt ("z" címkével jelölve). Az így kapott eredmények és az eredeti modellekkel kapott eredmények az 5. táblázatban láthatóak.

Az osztályozás további javításának érdekében többféle megközelítés szerint igyekeztünk módosítani a modelleket. A vélelmezett bonyolultság (akusztikai osztály összetettsége), az osztályozás alapján hibásnak vélt címkék, valamint az egyes hangminták átlagos hossza alapján hoztuk létre a modellek különböző csoportjait, ame-

lyekhez ezután különböző állapotszámú Markov-modelleket rendeltünk. Az így kapott osztálycsoportokat, valamint a hozzájuk tartozó felismerés eredményét a 6. és 7. táblázat mutatja.

5. táblázat: A legjobb, 500 ms-os időablakkal kapott osztályozási eredmények a különböző indexek szerint [%]-ban.

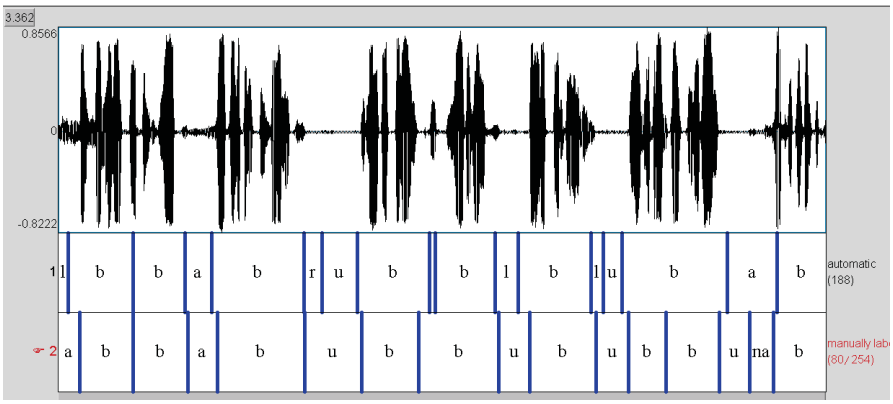
	Eredeti modellek esetén			Zajos beszédmodell bevezetése után		
Hangfelvétel-azonosító	Beszédindex	Zajindex	Összindex	Beszédindex	Zajindex	Összindex
01	0,69	63,95	16,51	46,81	63,3	50,93
02	11,36	24,29	14,59	32,74	24,29	30,6
03	100	33,7	83,42	100	35,58	83,89
04	83,62	29,39	70,07	68,43	29,07	58,59
05	100	15,34	78,84	82,64	9,8	64,43
06	98,75	22,9	79,79	98,88	23,34	79,99
07	67,22	33,4	58,76	76,8	33,28	65,92
08	83,61	33,1	70,98	84,22	32,71	71,34
09	76,31	0,46	57,35	80,06	0,58	60,19
10	84,55	36,79	72,61	88,82	38,24	76,17

6. táblázat: A módosított osztálycsoportosítás eredménye.

Állapotszám	Osztályok
14	b, z, k
11	s, a, u
5	g, r
4	l, t

7. táblázat: A módosított osztálycsoportokkal kapott felismerési eredmény [%]-ban.

Hangfelvétel azonosító	Beszédindex	Zajindex	Összindex
01	49,65	57,96	51,73
02	16,75	28,95	19,79
03	100	38,34	84,58
04	87,23	17,75	69,86
05	82,64	8,61	64,13
06	100	29,2	82,3
07	65,2	30,09	56,42
08	86,91	37,24	74,49
09	83,24	0,58	62,57
10	88,1	36,89	75,3



2. ábra. Példa az automatikus osztályozás eredményére.

3 Érzelemfelismerés

3.1 Érzelmi adatbázis

Az érzelemfelismerés megvalósításához folyamatos beszélgetéseket tartalmazó spontán telefonos felvételek, különböző talkshow-k felvételei kerültek összegyűjtésre, valamint annotálásra. A folyamatos beszéd frázisegységekre lett feltagolva, a frázisok pedig érzelem szerint lettek annotálva, mely során a legjellemzőbb érzelmi minták kerültek bejelölésre. A folyamatos feldolgozás során az derült ki, hogy a szövegkörnyezet figyelembevétele nélkül a frázis egységek érzelmi osztályozása számos esetben nem egyértelmű. Ezért a bejelölést végző személyeknek ezután csupán az érzelemmel töltött részeket kellett megjelölni, azok osztályozását külön szubjektív tesztsorozat során több lehallgató végezte el. Így végül 2540 érzelmes szakasz szubjektív lehallgatását 30 személy végezte el, amelyek után végül 43 beszélőtől, összesen 985 érzelmes szakasz lett kiválasztva, 6 érzelem szerint. A kiválasztás során csupán azokat a hangmintákat válogattuk ki, amelyeknél a szubjektív lehallgatás során 70%-os egyezés volt a döntésekben. Az érzelmek az alábbiak voltak: semleges, szomorú, meglepett, dühös/ideges, nevetés beszéd közben, valamint boldog. A kategóriák közötti eloszlást a 8. táblázat mutatja.

8. táblázat: A 30 lehallgató személy által kiválasztott érzelmes minták száma.

Érzelemtípus	**Frázisok száma (a lehallgatók döntéseinek 70%-os egyezése)**
Semleges	517
Dühös/ideges	290
Boldog	39
Nevetve beszél	42
Szomorú	54
Meglepett	43

3.2 Érzelemfelismerési eljárás

Az érzelemfelismerési kísérletek során végül 4 érzelmet használtunk fel, mivel ezekhez volt elegendő hangminta, amellyel tanítani lehetett. A 10. táblázat alapján ezek a következők: semleges, harag/ideges, öröm és nevetve beszél együtt, szomorú. Az automatikus osztályozáshoz szupport vektor gépeket alkalmaztunk, amelyhez az SVMLib [7] szabadon letölthető C# programozási nyelvű könyvtárcsomagját használtuk. A kísérletek célja az volt, hogy megvizsgáljuk, milyen akusztikai jellemzők szükségesek az érzelem felismeréséhez.

A következő jellemzőket vizsgáltuk meg:

- az alaphang átlaga, maximuma, tartománya és szórása (jelölés: F0)
- az alaphang deriváltjának átlaga, maximuma, tartománya és szórása (jelölés: ΔF0)
- az intenzitás átlaga, maximuma, tartománya és szórása (jelölés: EN)
- az intenzitás deriváltjának átlaga, maximuma, tartománya és szórása (jelölés: ΔEN)
- 12 mel-frekvenciás kepsztrális együttható átlaga, maximuma, tartománya és szórása (jelölés: $MFCC_i$)
- harmonicity értékek átlaga, maximuma, tartománya és szórása (jelölés: HARM)

Minden jellemzőt 10 ms-os lépésközzel nyertünk ki, majd frázisonként számoltuk ki a megfelelő statisztikai jellemzőt. Így egy frázisra egy ilyen érték adódott, ezekből állt végül elő a hangmintához tartozó jellemzővektor.

3.3 Eredmények

A tesztek során a következő osztályjelölések szerepelnek: harag/ideges: A, boldog: J, semleges: N, szomorú: S. A 9. táblázat(csoport) négy kísérleti összeállítás eredményeit tartalmazza.

9. táblázat: Automatikus felismerési eredmények [%]-ban négy jellemzővektor-összeállítás esetén.

jellemzővektor: F0, ΔF0, EN, ΔEN				
	A	**J**	**N**	**S**
A	51	15	5	4
J	18	32	17	2
N	6	9	57	3
S	15	4	13	7
			Felismerési eredmény: 56,98	

jellemzővektor: F0, ΔF0, EN, ΔEN, HARM,				
	A	**J**	**N**	**S**
A	46	13	10	6
J	17	30	16	6
N	7	8	56	4
S	12	7	12	8
			Felismerési eredmény: 54,26	

jellemzővektor: F0, ΔF0, EN, ΔEN, $MFCC_i$				
	A	**J**	**N**	**S**
A	57	13	4	1
J	12	37	13	7
N	4	12	55	4
S	5	17	5	12
			Felismerési eredmény: 62,40	

jellemzővektor: F0, ΔF0, EN, ΔEN, HARM, $MFCC_i$				
	A	**J**	**N**	**S**
A	61	9	4	1
J	11	41	11	6
N	3	12	56	4
S	5	16	5	13
			Felismerési eredmény: 66,27	

A felismerési eredmények azt mutatják, hogy az alapjellemzőkön kívül (alaphang, intenzitás) a mel-frekvenciás mel-kepsztrum jellemzők nagy szerepet játszanak az automatikus felismerésben. A harmonicity értékek ezt még javítani tudják. Ám mivel a minták száma jelenleg még nem kielégítő, ezért ahhoz, hogy ezeket az eredményeket megbízhatóbbá tegyük, folyamatos adatbázisgyűjtés és -feldolgozás szükséges.

Annak ellenére, hogy a tesztek során az alaphang és intenzitás értékek normáltan szerepeltek a jellemzővektorban, érdemes megnézni az eredményeket akkor, ha a hangmintákat külön válogatjuk női, illetve férfi mintákra. Ennek eredménye látható a 10. táblázatban. Habár a felismerés enyhe javulást mutat, a hangminták nem kielégítő száma miatt ez csupán pár hangmintaeltérést jelent.

10. táblázat: Automatikus felismerés eredménye [%]-ban női és férfi hangminták esetén a legjobb felismerési teljesítményt adott jellemzővektor esetén.

férfi beszélők

	A	**J**	**N**	**S**
A	17	0	4	1
J	1	7	2	7
N	2	2	18	0
S	1	5	0	14

Felismerési eredmény: 69,14

női beszélők

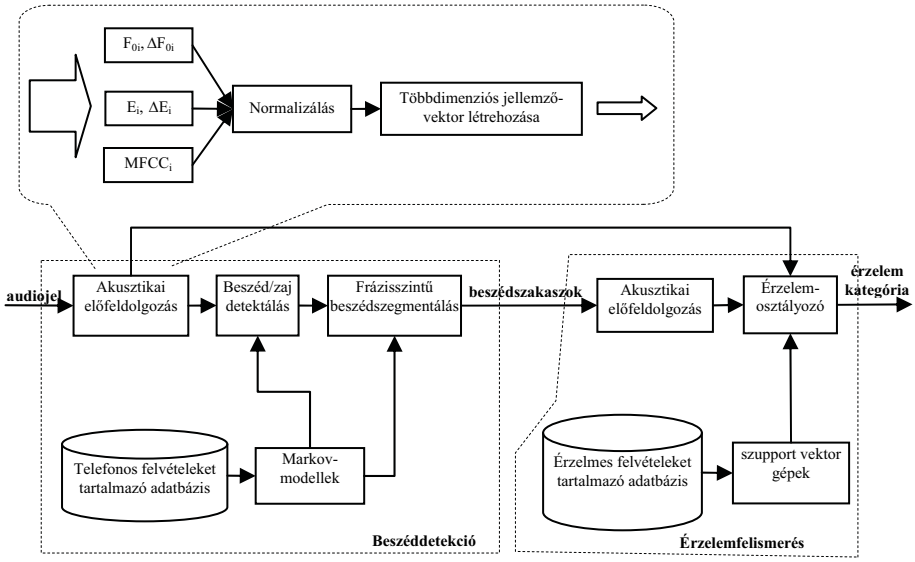
	A	**J**	**N**	**S**
A	46	6	1	0
J	9	31	11	1
N	1	9	40	3
S	3	8	6	2

Felismerési eredmény: 67,23

4 Kvázi valós idejű beszédfelismerési eljárás terve spontán beszédben

Beszédkommunikáció közben, főként hosszú beszélgetés esetén, a beszélő személy érzelmi állapota folyamatosan változik. Annak érdekében, hogy a beszélő mentális állapotát követni tudjuk, a folyamatos beszélgetést szakaszokra kell tagolnunk. Jelen esetünkben a frázist választottuk a szegmentálás alapegységének.

Az automatikus frázisszintű szegmentálást a megvalósítandó valós idejű felismerőben a már fentebb bemutatott beszéddetektáló végzi. Az egybeépített automatikus felismerő blokkvázlata a 3. ábrán látható. Az ábrán a fentebb bemutatott két különálló felismerő akusztikai feldolgozása külön szerepel, mivel azokat két különálló modul végzi. A végső szoftverben azonban sebességoptimalizálási célból ezt egyetlen modul fogja végezni.



3. ábra. Az automatikus érzelemfelismerő blokk vázlata spontán beszéd esetén.

5 Összefoglalás

A cikkben bemutatásra került egy olyan automatikus érzelemfelismerési eljárás, amely spontán zajos környezetű beszédben, valós időben képes érzelmek felismerésére kizárólag a beszéd prozódiai jellemzői alapján.

Ehhez kifejlesztettünk egy olyan rejtett Markov-modelleken alapuló eljárást, amely a hanganyagot frázisegységekre szegmentálja, és osztályozza beszédosztályra, valamint egyéb akusztikai környezeti zajosztályokra. Így oldva meg a beszéd-nem beszéd detektálást és a frázisszintű szegmentálást.

A beszéddetektálási eredmények kiértékelése során megállapítható, hogy a detektáló eljárás alkalmazható spontán beszédre. A kapott beszédindex-eredmény nem kiemelkedően zajos felvételek esetén eléri a 80 %-ot, ami, ahogy az eredményeket bemutató ábrán is látható, elfogadható teljesítmény.

A detektáló, fázisszegmentáló eljárást követi az érzelemfelismerő eljárás. Négy érzelemre szubjektív lehallgatással kiválogatott hangminták betanítása esetén a szupport vektor gép alapú automatikus felismerő 66%-ban osztályozta megfelelően az érzelmes hangmintákat.

Köszönetnyilvánítás

Ez a kutatás a Jedlik OM-00102/2007 számú "TELEAUTO" projekt és a TÁMOP-4.2.2-08/1/KMR-2008-0007 projekt keretein belül készült.

Bibliográfia

1. Tóth, Sz. L., Sztahó, D., Vicsi, K.: Speech Emotion Perception by Human and Machine. In: Proceedings of COST Action 2102 International Conference. Patras, Greece, October 29-31, 2007. Revised Papers in Verbal and Nonverbal Features of Human-Human and Human-Machine Interaction 2008. ISBN: 978-3-540-70871-1. Springer LNCS (2008) 213–224
2. Hozjan, V., Kacic, Z.: A rule-based emotion-dependent feature extraction method for emotion analysis from speech. The Journal of the Acoustical Society of America. Vol. 119 No. 5 (2006) 3109–3120
3. Navas, E., Hernáez, I., Luengo, I.: An Objective and Subjective Study of the Role of Semantics and Prosodic Features in Building Corpora for Emotional TTS. IEEE Transactions on Audio, Speech and Language Processing Vol. 14 No.4 (2006)
4. Vicsi K., Sztahó D.: Ügyfél érzelmi állapotának detektálása telefonos ügyfélszolgálati dialógusban. In: VI. Magyar Számítógépes Nyelvészeti Konferencia. Szeged (2009) 217–225
5. Boersma, P., Weenink, D.: Praat: doing phonetics by computer [Computer program]. Retrieved from http://www.praat.org
6. The Hidden Markov Model Toolkit (HTK). http://htk.eng.cam.ac.uk/
7. Chang, C.C., Lin, C-J.: LIBSVM : a library for support vector machines. Software available at http://www.csie.ntu.edu.tw/~cjlin/libsvm (2001)