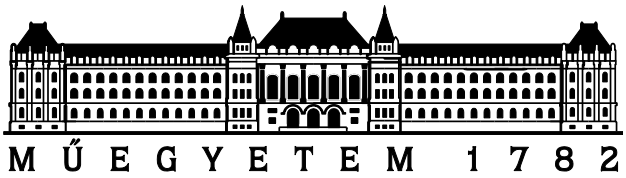


Kóros beszédhangok statisztikai analízise, és osztályozása

TDK dolgozat
Budapesti Műszaki és Gazdaságtudományi Egyetem
Távközlési és Médiainformatikai Tanszék

Készítették: Imre Viktor és Barlangi Renáta
Konzulens: Dr. Vicsi Klára

Tartalomjegyzék

1 Bevezetés

A beszédhang kutatókat és a foniáter szakorvosokat már évtizedek óta foglalkoztatja az a kérdés, hogy a beszédhangok akusztikai paramétereiből milyen módon állapítható meg a hangképzési betegségek mibenléte.

Ma már a technológiai fejlettség lehetővé teszi, hogy nagy számú mintán statisztikai vizsgálatokat, valamint osztályozást tudjuk végezni. Ehhez a kóros beszédhangokat tartalmazó, jól megtervezett és címkézett adatbázisra van szükség. Ez ideig a kutatók széles körben csak kitartott hangokat tartalmazó adatbázisokat használtak a vizsgálatokhoz. A valóságban azonban a kommunikáció során a folyamatos beszédet ítéljük meg, továbbá a tapasztalt audiológus, fül-orr-gégész és foniáter szakorvosok is a folyamatos beszéd alapján döntenek és sorolják be a páciens hangját egy szubjektív minőségi skála szerinti csoportokba.

Ezért határoztuk el, hogy mi a folyamatos beszéd hangjait vizsgáljuk, és jelentősen támaszkodunk a foniátriai gyakorlatban alkalmazott szubjektív minőség szerinti osztályozásra is. [1, 2] Továbbá a kutatásnál felhasználjuk a gépi beszédfelismerés eredményeit is. Az adatbázis hanganyagának gyűjtését három éve kezdtük el és azóta is folyamatosan bővítjük az Országos Onkológiai Intézet Fej- és Nyaksebészeti Osztályának III. számú ambulanciáján.

Elsődleges célunk az egészséges és kóros minták statisztikus elemzése és automatikus elkülönítése. Az adatbázisban, a folyamatos beszéd magánhangzói kvázi stacioner szakaszainak statisztikai vizsgálata alapján meghatároztuk azokat az akusztikai paramétereket, amelyek a leglényegesebbek az egészséges és kóros minták elkülönítésében. Majd ezek után az elemzési eredményekre támaszkodva SVM (Support Vector Machine) alapú osztályozót hoztunk létre. Alkalmasan megválasztott tesztelési sorozattal az osztályozást optimalizáltuk. Az eredmények azt mutatják, hogy a folyamatos beszéd alapján az egészséges és kóros minták szétválaszthatók.

Szeretnénk kiemelni, hogy a közösen bemutatott eredményeket közös munkával értük el, ennek ellenére egyes munkafolyamatok jól elkülöníthetők a szerzők között. Az adatbázis gyűjtésében és felcímkézésében mindketten részt vettünk, megközelítőleg azonos arányban, mint ahogy a felismerő és előfeldolgozó rendszer kiépítésében is. A statisztikai analízist Imre Viktor végezte el, a tanítási/tesztelési munkákat Barlangi Renáta.

1.1 Beszédhangok képzése

Az emberi kommunikáció során a beszéd segítségével akusztikus gondolatátvitelt valósítunk meg. Ennek során először megfogalmazódik egy gondolat, amelyet a nyelv segítségével megfogalmazunk, és az idegi vezérlés parancsai hatására akusztikus jellé alakítjuk. A beszéd során a tüdőt az izmok összenyomják és levegőt préselnek a hangszalagoknak, melyek a fokozódó nyomásnak nem tudnak ellenállni, és kinyílnak. Így a levegő útja szabad lesz, kiáramlik a légcsőből, és a nyomáscsökkenés hatására a hangszalagok lezárnak. Az állandó tüdőből érkező gerjesztés hatására periodikus rezgés jön létre, ami nyomáshullámot generál. Ez a hullám végighalad a hangképző csatornán, melynek részei a szájüreg, nyelv, orrüreg, fogak… A szájnyílásnál a nyomáshullám kicsatolódik a külvilágba, mikrofonnal mérhető jelet kapunk, amelyet a mikrofon feszültséggé konvertál. [3]

A hangok szubjektív minőségét tehát erősen befolyásolhatja az előzőekben felsorolt rendszer bármely eleme. Ha nem megfelelő az agyi vezérlés, ha a hangszalag tumoros elváltozást mutat, vagy a nyelv egy része el van távolítva műtétileg, az mind az akusztikai kép változásához vezet. Ha képesek vagyunk a hang abnormalitását a fülünkkel detektálni, akkor egészséges és beteg hangminták között is képesek lehetünk automatikusan döntést hozni.

2 Irodalmi áttekintés

Az orvosi hang alapú diagnosztika kutatásának jelentős nemzetközi múltja és jelene van, ellentétben a magyar kutatásokkal. Ez idáig Magyarországon nem találtunk olyan cikket, amely a témával foglalkozott volna. A külföldi kutatócsoportok egymástól jellegzetesen elkülönülő irányokba fejlesztenek, azonban vannak olyan technológiák és eszközök, melyek sokszor felbukkannak. Alapvetően 4 kérdésben vizsgáltuk a rendelkezésre álló irodalmakat:

1. Milyen adatbázison került sor az elemzésre, tanításra, tesztelésre?
2. Milyen előfeldolgozási lépéseket végeztek?
3. Talán a legfontosabb: milyen csoportokat különítettek el?
4. Milyen módszert alkalmaztak az automatikus osztályozásra?

Az elemzést olyan formában közöljük, amely a levont konklúziókat tartalmazza, és nem az egyes cikkek egyenkénti bemutatását.

Először nézzük, hogy milyen adatbázisokat, milyen hangmintákat használtak fel a kutatók. Mint azt már évekkel ezelőtt is lehetett látni, a kutatások nagy része azonos hangadatbázist használ. Ez a Key Elemetrics által a MEEI (Massachusetts Eye and Ear Infirmary) laboratóriumaiban, süketszobai környezet biztosításával készült hangfelvételeket tartalmazó adatbázis [5, 7, 9, 12, 13, 15]. Minden felvétel csak kitartott hangokat tartalmaz. Tovább vizsgálva az adatbázisokat, jól látható, hogy lényegében minden más kutató saját maga által készített kórházi felvételeket használt [4, 6, 8, 11, 14]. Ezekből nyilvánvalóvá válik, hogy nemzetközileg jelentős adatbázis lényegében egy létezik, ez is angol nyelvkörnyezetben került rögzítésre, valamint nem ad lehetőséget folyamatos beszéd vizsgálatára.

A folyamatos beszéd vizsgálatának, és felismerésben való felhasználásának az ötlete számunkra a foniátriai gyakorlatból eredt [1]. A cikkek alapvetően kitartott hang elemzésével foglakoznak, de van amelyik említi, hogy érdemes lenne folyamatos beszéd hangjainak használatával is kísérletezni [15]. Alapvetően megerősítettnek látjuk azt, hogy igen előremutató fejlesztési irányt határoztunk meg, amikor egy valós környezetben működőképes osztályozó megalkotásán fáradozunk, mivel a süketszobai körülmények nem biztosíthatóak egy háziorvosi vagy otthoni környezetben, de még a kórházakban sem gyakran.

Másik érdekes kérdés, hogy milyen előfeldolgozási, lényegkiemelési eljárásokat alkalmaznak az egyes kutatásokban. Több esetben is használnak ingadozásértékeket, mint a

jitter és a shimmer , valamint ezen paraméterek enyhe módosításait is (kisebb-nagyobb ablak) [4, 8]. Egyes munkákban előfordul a lényegkiemelési eljárások között klasszikus MFCC (Mel-frekvenciás kepsztrális komponensek) paraméter alkalmazása is [6]. Találkozunk egyéb megoldásokkal is, de a legérdekesebb új irány a wavelet alapú feldolgozás lehet, amely jellegzetes hullámformák összegére való felbontást jelent [10, 14]. Ezt nem találjuk egyértelműen vizsgálandó iránynak, mivel a wavelet transzformáció lényegi részeit tekintve hasonlít az általánosan elterjedt Fourier vagy diszkrét koszinusz transzformációkhoz.

Kérdésként merült fel, hogy milyen csoportok közötti döntések automatizálásán dolgoznak a külföldi kutatók. Azt tapasztaltuk, hogy a legtöbb esetben nem volt előrelépés a beteg és egészséges elkülönítésen fölül, nem finomították a csoportokat [4, 8, 12, 13]. A legelőremutatóbb próbálkozások azok voltak, amelyekben az egészséges mellett két betegségcsoportot igyekeztek elkülöníteni, de ezek elég jelentősen speciális csoportok voltak minden esetben. A legtöbb ilyen cikkben csomókat, tumoros eseteket választottak el ödémás esetektől [10, 14]. Jól látszik, hogy nem foglalkoztak bénulásos, vagy egyéb neurális eredetű betegségekkel, vagy a funkcionális diszphonyával amelyek például a mi adatbázisunkban jelentős számban szerepelnek. Ezek alapján egyértelmű, hogy az egészséges és beteg minták elkülönítése sem megoldott probléma, így vizsgálata mindenképpen aktuális téma.

Utolsóként arra fordítottunk figyelmet, hogy milyen osztályozási eljárásokat használnak az egyes kutatások. Alkalmaztak klasszikus neurális hálózatot, mint az MLP (Multi Layer Perceptron) [8]. Jelentős számú rendszert építettek fel SVM alapon [5, 10, 14, 15]. Ezek mellett viszonylag sokan használnak GMM-et (Gaussian Mixture Model) amely hasonlatos a bázisfüggvényes neurális hálózatokhoz [12, 13]. Az osztályozási eredmények alapján levonható következtetés, hogy egy megfelelően használt SVM is ugyanolyan osztályozási pontosságot képes adni, mint az egyéb megoldások, így más típusú, vagy bonyolultabb automatikus osztályozó rendszer felépítése nem indokolt.

Az irodalmak feldolgozása alapján világossá vált, hogy azok az irányok, melyeket már korábbi munkáink során is követtünk [2, 16], alapvetően helyesek, jó eredményekre vezethetnek, és mindenekelőtt előremutatóak. Országos szinten egyedülálló kutatási területen dolgozunk, és nemzetközi szinten is aktuális témával foglalkozunk. A későbbiekben bemutatott eredményeinkkel nemzetközi szinten is értékes információkhoz jutottunk, mivel folyamatos beszéd kutatása jelenleg csak jövőbeni tervekben szerepel.

3 A hangadatbázis

3 év leforgása alatt sikerült összeállítanunk egy olyan magyar nyelvű, folyamatos beszédet és kitartott hangokat is tartalmazó hangadatbázist, amely egészséges és kóros hangmintákat egyaránt tartalmaz. Erre a hangadatbázisra azért volt nagy szükség, mert az országban még senki nem állított össze, még csak kitartott hangokat tartalmazó olyan adathalmazt sem, amelyen kóros és egészséges minták elkülönítését lehetett volna vizsgálni. Folyamatos beszéden történő vizsgálat, vagy adatbázis még csak szóba sem került.

Összességében látható, hogy az általunk készített, és folyamatosan fejlesztés alatt álló hangadatbázis nem csak az országban egyedi, hanem nemzetközi szinten is.

3.1 Adatgyűjtés

A felvételek az Országos Onkológiai intézet járóbeteg rendelésén, a Fej-nyak Sebészeti Osztály III. számú Ambulanciáján lettek rögzítve dr. Mészáros Krisztina foniáter szakorvos rendelésén. A szakrendelésre általában különböző hangpanaszokkal érkeznek a betegek. A szükséges vizsgálatok, vagy hangterápia elvégzése után megkérdeztük a pácienseket, hogy hozzájárulnak-e beszédhangjuk rögzítéséhez. Amennyiben beleegyeztek a következő feladatot kellett végrehajtaniuk:

1. 3 darab kitartott „o” hang kimondását, mindegyik előtt mély levegő vételével.
2. Egy a foniátriai gyakorlatban gyakran használt magyar nyelvre fonetikailag reprezentatív népmese felolvasását, a címe nélkül.

A felolvasandó szöveg:

Népmese: Az északi szél, és a nap

Az északi szél nagy vitában volt a nappal, hogy kettejük közül melyiknek van több ereje. Egyszer csak egy utast pillantottak meg, amint köpenybe burkolódzva közeledett. Elhatározták, hogy a vitát az nyeri meg, amelyik előbb veszi rá az utast, hogy kabátját levegye. Az északi szél összeszedte egész erejét, és fújni kezdett, de minél erősebben fújt, az utas annál szorosabbra fogta össze a kabátját. Az északi szél végre feladta a harcot. Ekkor a nap küldte meleg sugarait az utasra, aki rövidesen levette a kabátját. Az északi szélnek tehát el kellett ismernie, hogy kettejük közül a nap az erősebb.

A betegek adatai egy Excel táblában kerültek rögzítésre, amelyben feljegyeztük a felvétel sorszámát, dátumát, valamint a páciens nemét, betegségét és az RBH szubjektív skála szerinti kódját. Ezen kívül bekerült még egy megjegyzés oszlop is, ahova a felvétel készítője írhatta le tapasztalatait. Az adatbázist leíró táblázat egy részletét mutatja az első ábra.

	A	B	C	D	E	F
1	sorszám	felvétel napja	Páciens neme	betegsége	megjegyzés	RBH kód
38	37	2009.06.15	nő	funkcionális dysphonia, asztma	1 éve rekedt ezért orsó alakú zárási elégtelenség	R1B1H1
39	38	2009.06.15	nő	rekurrens parízis		R2B1H2
40	39	2009.06.15	nő	rekurrens parízis		R1B1H1
41	40	2009.06.15	nő	hangszalagbénulás, jobboldali	baloldali jól mozog	R1B0H1
42	41	2009.06.15	nő	rekurrens parízis		R3B1H3
43	42	2009.06.15	ffi	hangszalag tumor?	megy szövettani elemzésre	R3B0H3
44	43	2009.06.16	nő	rekurrens parízis	nem tudta felolvasni, egy mondat	R3B1H3
45	44	2009.06.16	nő	funkcionális dysphonia		R0B0H0
46	45	2009.06.16	ffi	nyelvtumor volt, műtött	szerinte mélyült, ránézésre szép	R0B0H0
47	46	2009.06.16	ffi	bal oldali chordectomia után		R3B1H3
48	47	2009.06.16	ffi	funkcionális dysphonia		R0B0H0
49	48	2009.06.16	ffi	funkcionális dysphonia		R2B0H2
50	49	2009.06.16	ffi	nyelvtumor	artikuláció zavar	R0B0H0
51	50	2009.06.16	ffi	nyelvtumor		R0B0H0
52	51	2009.06.16	nő	funkcionális dysphonia		R1B0H1
53	52	2009.06.16	ffi	funkcionális dysphonia		R1B0H1
54	53	2009.06.18	nő	laringitis cronica, krónikus gégegyulladás		R2B0H2
55	54	2009.06.18	nő	rekurrens parézis	már volt, de most saját hangján igyekezett, ismétlés 31	R2B0H2
56	55	2009.06.18	ffi	hangszalag tumor, sugár után	o-nál R1B0H1, egyébként R0B0H0	R0B0H0
57	56	2009.06.18	nő	funkcionális dysphonia		R0B0H0
58	57	2009.06.18	ffi	ép hang	színész	R0B0H0
59	58	2009.06.19	nő	spazmodikus diszfónia		R3B0H3
60	59	2009.06.19	nő	ALS	nazalitás	R1B0H1
61	60	2009.06.19	ffi	rec par		R3B1H3
62	61	2009.06.19	nő	ellenőrzés		R0B0H0
63	62	2009.06.19	ffi	ép hang		R0B0H0

lista nevek

1. ábra. Az adatbázist leíró táblázat egy részlete

Az RBH kód egy négy-fokozatú auditív rekedtségi skála, ahol a 0 a normál hangminőség, 3 a súlyos rekedtség. Az R (Rauhingkeit) a hangszalagok rezgési irregularitásából adódó érdességet, a B (Bechauchtkeit) a hangszalagok zárási elégtelenségéből adódó levegő-turbulenciát, a H (Heiserkeit) a rekedtséget általában jellemzi. Ezen rekedtségi skála segítségével a beteg aktuális állapota számszerűsíthető, valamint könnyű figyelemmel kísérni a hangminőség javulást a gyógyulás során. [1]

Ezt az RBH kódot minden egyes páciensnél a vizsgálatot végző szakorvos állapította meg. A betegek nevét is eltároltuk az Excel táblában azért, hogy később mi is figyelemmel tudjuk kísérni a többször is visszatérő betegek hangminőségének változását.

Az adatgyűjtéshez a következő eszközöket használtuk:

- Monacor ECM-100: Elektrolit kondenzátor közeltéri mikrofon. (A mikrofon karakterisztikája olyan, hogy nagy érzékenységgel csak az előtte 20-30 centiméterre lévő teret veszi)
- Creative Soundblaster Audigy 2 NX: külső, USB-s, hangkártya, 24 bites A/D álalakítóval, ezen keresztül történik a felvétel.

- Waversurfer: freeware hangfelvevő és analizáló program, minden felvétel 44100 Hz-es mintavételezési frekvenciával készült, 16 bites, lineáris kódolási eljárással, Windows PCM wav fájlba mentve.

Minden hangfájlból készíteni kellett egy 100%-ra normalizált változatot, amely során a program a globális maximumot egyenlővé teszi a wav formátum által biztosított elméleti feszültségérték maximumával (16 bites számábrázolás) és a jel többi részét pedig arányosan felnagyítja. Az adatbázis minőségi jellemzőit az első táblázat tartalmazza.

1. táblázat. Az adatbázis minőségi jellemzői

Forrás	Formátum (KHz)	Rögzítési környezet	Bemondás módja
Mikrofon	44,1	Kórházi környezet	Kitartott hang és felolvasott szöveg.

3.2 Jelenlegi állapot

Az adatgyűjtés 2009-ben kezdődött el, tehát már közel három éve folyamatosan bővítjük hangadatbázisunkat. Jelenleg 327 darab saját készítésű, kóros és egészséges hangminta áll rendelkezésünkre. A felvételek készítésekor számos problémába ütköztünk. A korházi környezet olykor nagyon zajos is lehet, ami nagyban befolyásolhatja a felvételek minőségét. Ezen kívül számos esetben nem sikerült a beszédhangrögzítést tökéletesen végrehajtani: túl közel, vagy távol került a mikrofon a beszélőtől, háttérben egyéb beszéd is hallható, valamint egyes páciensek betegsége annyira súlyos, hogy értékelhető hangfelvételt nem sikerült velük készíteni. Ezért a rendelkezésünkre álló hanganyagot többször újrahallgattuk és kiválasztottuk azt a 169 darab hangfájlt (97 női, 72 férfi), amelyekkel a későbbiekben dolgozni tudtunk. Mivel jelen dolgozat szempontjából ez a 169 darab hangminta a fontos, ezért az alábbi statisztikai adatok csak erre a kisebb csoportra vonatkoznak.

A felvételeken hallható páciensek betegség típusai széles skálán mozognak. Van olyan betegségtípus, amiből csak 1-2 felvétel van, míg akad 50-es nagyságrendű betegségcsoport is. A leggyakrabban előforduló betegségek a következők:

- Funkcionális dysphonia: olyan hangképzési zavar, amely akkor lép fel, ha a hangszalagok rezgése és az artikuláció nincs összhangban. Kialakulásának okai: hang túlerőltetése, nem megfelelő énekhangregiszter, gyenge beszélőszervek.
- Recurrens paresis: egy- vagy kétoldali hangszalagbénulás, amit agyi rendellenesség, vagy a gégéhez vezető idegek károsodása okozhat.
- Kóros szövetburjánzások: különböző tumorok és daganatok.
- GERD (gastrooesophagealis refluxbetegség): az első, nyelőcsőt záró izom nem záródik rendesen és a gyomor tartalma visszaáramlik a nyelőcsőbe. A sav irritációt, égető érzést kelt. A hangszalagok irritációjával hangelváltozást, rekedtséget válthat ki.

A betegségek típusonkénti eloszlását a 2. táblázat mutatja.

2. táblázat. Betegségek típusonkénti eloszlása

Betegség csoport	Nő	Férfi	Összesen
Funkcionális dysphonia	41	11	52
Recurrens paresis	21	10	31
Tumor	3	12	15
GERD	2	4	6
Daganat	0	4	4
Gégegyulladás	3	1	4
Egyéb	27	30	57

A betegségek eloszlásán kívül fontos adat még számunkra, hogy az RBH kód szerint hogyan lettek besorolva a páciensek. A beteg és egészséges csoportok kialakítása az RBH kód H értéke alapján történt meg, amely a hang rekedtségét általánosan jellemzi. A H0 kóddal rendelkezőket egészségesnek, míg H1, H2, vagy H3-as páciensek betegnek tekintendő. A páciensek H kód alapú eloszlását az alábbi táblázat mutatja:

3. táblázat. H paraméter nemenkénti eloszlása

Nem	H0	H1	H2	H3
Nő	40	30	17	10
Férfi	19	26	11	16
Összesen	59	56	28	26

A később ismertetett osztályozási tesztekhez 59 darab egészséges és 110 darab kóros hangminta állt rendelkezésre. Ezek alapján a felhasznált felvételek mennyiségi jellemzőit a 4. táblázat tartalmazza.

4. táblázat. Felhasznált felvételek mennyiségi jellemzői

Adatbázis	Felvételi idő (óra:perc:mp)	Méret	Beszélők száma	Adathordozó
Egészséges hangminták	1:10:44	351 MB	59	1 DVD
Kóros hangminták	2:20:15	702 MB	110	

3.3 A hanganyag feldolgozása

Az adatbázis legfontosabb része a szegmentált, címkézett hanganyag. A szegmentálás és címkézés a beszédfolyamat lineáris tagolása, vagyis a hangtest hangegységekre történő bontása. Ennek során a beszéd időfüggvényében bejelöljük a fizikailag megfigyelhető beszédhangokat és azok határait. A szegmentálás célja, hogy az adatbázis gépi feldolgozásához megadjuk a beszédjel és a fonetikai átirat közti időbeli kapcsolatot; azt, hogy melyik szimbólum a beszédjel mely időintervallumának felel meg. A beszédhang szintű szegmentálást és címkézést csakis nagy figyelmet igénylő és hosszadalmas kézi munkával lehet elvégezni. Megkönnyítheti, és meggyorsíthatja viszont a munkát egy megfelelő, a Beszédakusztikai Laborban fejlesztett automatikus beszédhang felismerőre alapozott, speciálisan erre a célra kialakított algoritmus, amely megkísérli automatikusan elhelyezni a beszédhang határokat. Olyan program készíthető, amely a határok jó részét elég pontosan helyezi el. A szegmentáló személy feladata ilyenkor csak az automata szegmentáló javaslatainak ellenőrzése és korrekciója, ami által a szegmentálás felgyorsítható. [17]

Ennek megfelelően az első feladatunk az volt, hogy a rendelkezésre álló hangmintákat szegmentáljuk és felcímkézzük.

3.3.1 A SAMPA jelölésrendszerről

A beszédhangok kiejtését tükröző nemzetközi jelölésrendszer az IPA (International Phonetic Alphabet). Egy adott IPA szimbólum – nyelvtől függetlenül – egy meghatározott hangot szimbolizál. Ez a jelölésrendszer azonban nem illeszkedik a számítógép billentyűzetére. Ezért nemzetközi szinten egy új jelölésrendszer bevezetésére volt szükség,

amely a SAMPA (Speech Assessment Methods Phonetic Alphabet) lett. Ez a szimbólumcsalád illeszkedik a számítógép billentyűzetére és nyelvfüggetlen is egyben. Az 5. táblázatban láthatóak a SAMPA jelek, melyek egy részét a dolgozatban is használni fogjuk. A hosszú mássalhangzókat kettősponttal jelöljük.[18]

5. táblázat. A magyar nyelv fonémáinak SAMPA jelölésrendszere

Magánhangzók	SAMPA	Mássalhangzók	SAMPA
a	O	p	p
á	A:	b	b
e	E	t	t
é	e:	d	d
i	i	k	k
í	i:	g	g
o	o	c	ts
ó	o:	dz	dz
ö	2	cs	tS
ő	2:	dzs	dZ
u	u	ty	t'
ú	u:	gy	d'
ü	y	f	f
ű	y:	v	v
		sz	s
		z	z
		sz	S
		zs	Z
		h	h
		r	r
		l	l
		j	j
		m	m
		n	n
		ny	J

3.3.2 Szegmentálási folyamat

A hanganyagok feldolgozásához a Beszédakusztikai Laborban fejlesztett automata szegmentálót használt, amelynek egy HTK (Hidden Markov Model Toolkit)[1] programcsomaggal fejlesztett beszédhang felismerő az alapja. A Markov láncokat gyakran használják valamilyen fizikai folyamat modellezésére, ahol különböző megfigyelések alapján

[1] http://htk.eng.cam.ac.uk/

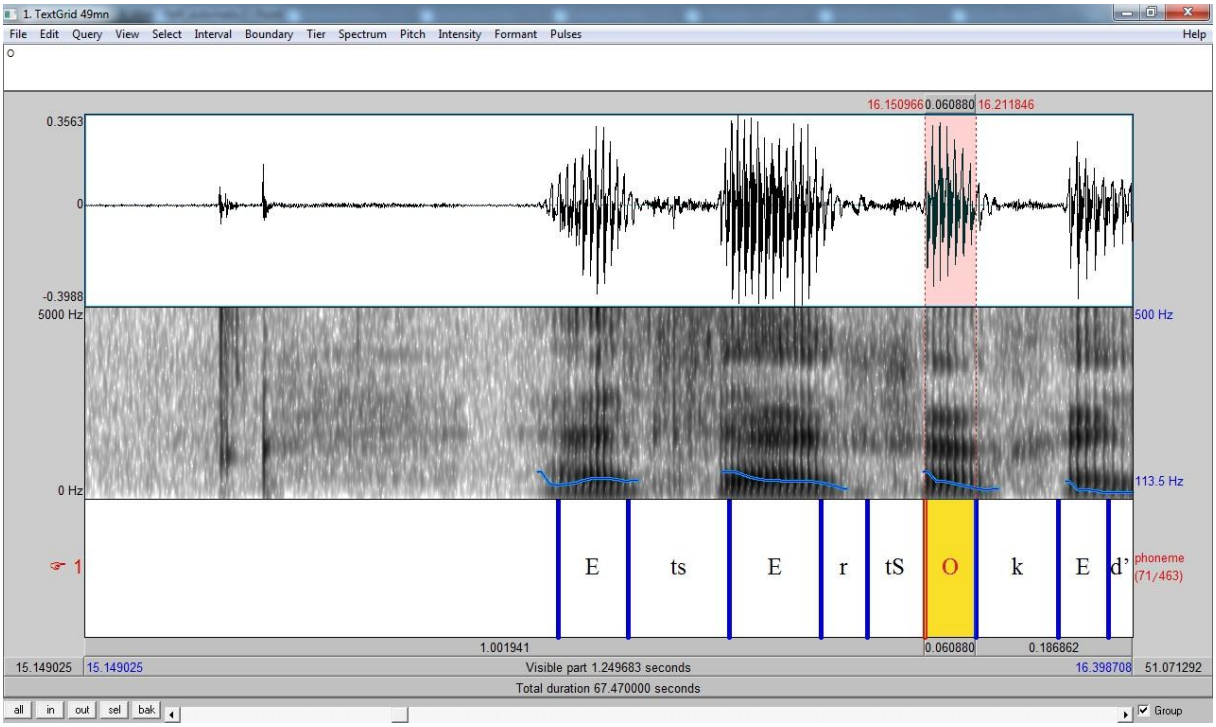
kell szimulálni – modellezni – a folyamatot. Ha a megfigyelés egyértelműen azonosítja, hogy a folyamat milyen állapotban van, akkor a használt modellt megfigyelhető Markov-modellnek vagy egyszerűen Markov-láncnak nevezzük. Azonban léteznek olyan folyamatok, amelyekre a megfigyelések alapján nem lehet következtetni, hogy épp melyik állapotban vagyunk. Az ilyen esetekben használatos a rejtett Markov-modell, amely esetén a megfigyelő nem látja az állapotokat. Egy beszédfelismerési feladat megoldását a lehetséges állapotsorozatok közül a legnagyobb valószínűségű állapotsorozat megtalálása jelenti. [19]

Az automata szegmentáló 16 kHz-es mintavételezési felvételeket vár, ezért az eredeti hangfájlokat át kell transzformálni, mert azok 44,1 kHz-es mintavételezéssel készültek. Ezek után többször meg kell hallgatni a felvétel, amely alapján el kell készíteni egy szövegfájlt. Ez a fájl a páciens által kimondott szavakat tartalmazza, sorrendhelyesen. Oda kell figyelni arra, hogy az elhangzott szavak kerüljenek a fájlba, különben a szegmentáló program elcsúszhat az illesztéssel. Tipikus kiejtési hibák a névelő elhagyása, tagmondatok ismételt felolvasása, vagy a félreolvasás, például: a felolvasandó szövegben „kettejük” szerepel, míg a legtöbben „kettőjüket” mondanak, vagy „veszi” helyett „viszi” olvasása.

Az általunk szegmentálni kívánt hang- és szövegfájlt a program gyökérkönyvtárába kell átmásolni. Fontos, hogy a két állomány neve megegyezzen. A program a bemeneti szövegfájlból egy olyan szövegfájlt készít, amely minden szót csak egyszer tartalmaz, abc sorrendben. Ezek után az eredeti szavak és a hozzájuk tartozó SAMPA karakterek társítása történik meg. Végül a bemeneti SAMPA átírás és a hangfájl segítségével „forced alignment” történik, azaz kényszerített illesztés, amely során az adott átirat és a jel közötti beszédhanghatár-összerendelés valósul meg. Az automata kimenete egy „TextGrid” típusú fájl, ami a hanghatárokat és a hozzájuk tartozó SAMPA karaktereket tartalmazza.

Az automata jó kiindulási alapot adott a további munkához, ám kézzel történő javításra minden esetben szükség volt. Számos alkalommal a betegség súlyossága miatt a program nehezen tudott megfelelő illesztést végezni, ezért akár egész mondatrészek is elcsúszhattak. A kézzel történő javítás során ezeket a hibákat kellett feltárni, illetve a kisebb elcsúszásokat helyreigazítani. Erre a feladatra a Praat program a legalkalmasabb, ami hangelemzések egyszerű elvégzésére ad lehetőséget. Az alkalmazás segítségével nyomon tudjuk követni a beszédjel hullámformáját, spektogramját, vagyis a frekvenciaösszetevők teljesítményszint-eloszlásának időbeli eloszlását, valamint a hanghatárokat egyszerűen bejelölhetjük, módosíthatjuk és felcímkézhetjük.

Az alábbi ábrán egy férfi pacienssel - akit nyelv tumorral diagnosztizáltak - készült felvétel annotálási folyamata látható. A betegség súlyossága miatt, a beszéd nehezen érthetővé vált, az automata az illesztés során elcsúszott. Az „e” és „c” hangokat kellett színkép alapján jól elhelyezni, amelyek helyét az „r” hang foglalta el. A javítás utáni állapotot mutatja a 2. ábra.



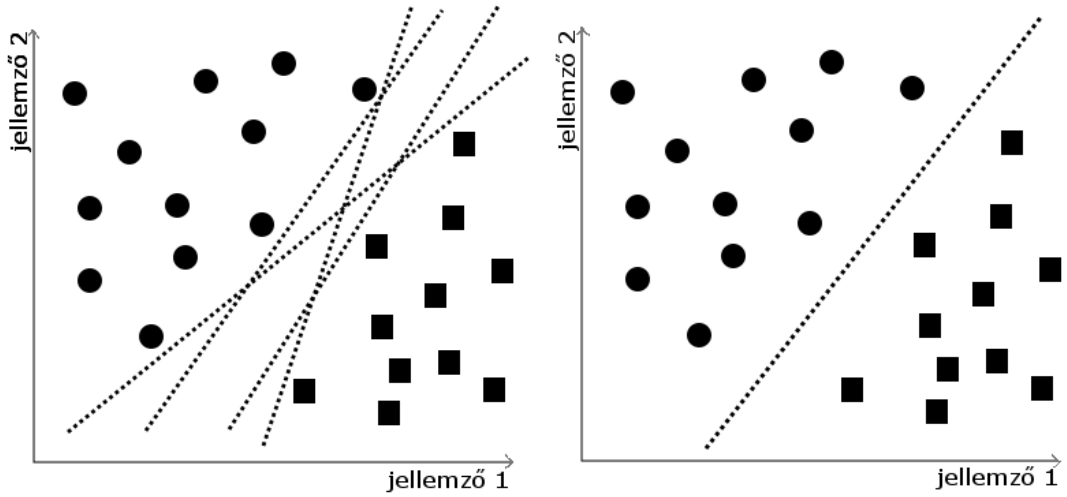
2. ábra. Hanghatárok kézi javítása

4 A felismerő motorja

A kóros és az egészséges minták elkülönítésére a statisztikai elemzés eredményei alapján statisztikai osztályozó eljárást alkalmaztunk. Az osztályozási feladat megoldására speciális kernelgépet használtunk. Előnyük, hogy a tanulási folyamat során nem csak hiba (pl. négyzetes eltérés) szerint optimalizálnak, hanem a valódi kockázat szerint is. A valódi kockázat mértéke kezeli a szükségtelenül nagy számú szabad paraméterből következő bizonytalanságot.

A kernelmódszer a bemenő adatokat elvi szinten egy bázisfüggvény segítségével a jellemző térbe transzformálja, amely az adatok egy jobb reprezentációját biztosítja az eredeti térnél. A jellemzőtérből a következő transzformációs lépésben a kerneltérbe jutunk. A kerneltérben lévő szabad paraméterek száma független a jellemzőtér szabad paramétereinek számától. Ez a kezünkbe adja annak a lehetőségét, hogy végtelen dimenziós jellemzőtérrel dolgozzunk, miközben a kerneltérbeli reprezentációnk véges dimenziójú. Ilyen eset az egyik legtöbbször használt radiális bázisfüggvényű kernelgép.

A rendszerünk felépítése során a kernelgépek egy speciális fajtáját alkalmaztuk. Ez az SVM , vagyis Support Vector Machine (szupport vektor gép). Az optimalizálás egyrészt hibát minimalizál (az általunk használt LS-SVM, Least Square SVM, négyzetes hibát) másrészt a mintahalmazok egy olyan sokdimenziós térbeli elválasztását teszi lehetővé, amelynél az elválasztó hipersík szintén optimalizált helyzetben van a mintahalmazokhoz képest. A harmadik ábrán jól látható a különbség a lehetséges elválasztások között. [20]



3. ábra. Halmazok elválasztási lehetőségei, és az optimális megoldás

A hálózatok tanítására általában ellenőrzött tanítási módszert használnak, és mi is egy ilyen megoldást alkalmaztunk. Az LS-SVM (a továbbiakban csak SVM) általános esetben kétosztályos osztályozóként működik. Ez azt jelenti, hogy a hálózat két osztály megkülönböztetésére képes. Ha több osztály között szeretnénk dönteni, akkor több SVM-et kell megtanítanunk úgy, hogy mindegyik egy adott osztály, és az összes többi között döntsön. A matematikai modellből, és a tanítási algoritmusból következik, hogy egyszerre nem fog több osztály mellett dönteni a rendszer.

Mivel kóros és egészséges minták között döntünk, ezért lényegében két osztályunk van. Nézzük meg, hogy miként tudjuk megtanítani az SVM számára, hogy melyik minta melyik osztályba tartozik! A folyamat során van egy tanítómintákat tartalmazó halmazunk. Minden tanítóminta egy x vektornak felel meg, és minden ilyen x vektorhoz tartozik egy d elvárt válasz, amely lehet -1 és +1 (a matematikai modellből származnak, és kötöttek). A hálózat az aktuális x-re ad egy y választ. Adottak tehát az $\{x_i, d_i\}_{i=1}^{P}$ mintavektorok, és elvárt válaszok.

Az összes pontra felírható a következő egyenlőtlenség:

$$d_i \left(\underline{w}^T * \underline{\varphi}(\underline{x}_i) + b \right) \geq 1 - e_i \tag{1}$$

Amelyben w az x vektorokhoz tartozó (mindhez ugyanaz) súlyok vektora, míg b egy skalárral történő eltolás. Az egyenlőtlenség jobb oldalán látható e_i az x_i minta távolsága az elválasztó hipersíktól (2D esetben egyenestől). A $\underline{\varphi}(\underline{x}_i)$ függvény a bázisfüggvény, amely a jellemzőtérbe történő transzformációt írja le. Az SVM-eknél is alkalmazott kernel-trükk azonban sosem számol ezzel a függvénnyel, így ezt nem is kell meghatároznunk. A trükk lényege, hogy a kernelfüggvény felírható

$$\underline{\underline{K}}(\underline{x}_i, \underline{x}_j) = \underline{\varphi}(\underline{x}_i)^T \underline{\varphi}(\underline{x}_j) \tag{2}$$

formában is, így nem kell a jellemzőtérben számolnunk. A kernelfüggvényt azonban csak formailag írjuk így fel, megválasztása ettől függetlenül, lényegében tapasztalati úton működik. Általában a következő függvényeket szokták használni ilyen célra:

lineáris: $$\underline{\underline{K}}(\underline{x}_i, \underline{x}_j) = \underline{x}_i^T \underline{x}_j \tag{3}$$

polinomiális: $$\underline{\underline{K}}(\underline{x}_i, \underline{x}_j) = (\underline{x}_i^T \underline{x}_j + a)^l \tag{4}$$

radiális: $$\underline{\underline{K}}(\underline{x}_i, \underline{x}_j) = \exp(-\gamma \|\underline{x}_i - \underline{x}_j\|^2) \tag{5}$$

Mivel a gyakorlatban a radiális kernelfüggvény terjedt el (amelynek a jellemzőtere végtelen dimenziójú), mi is ezt a megoldást alkalmaztuk. A fenti gondolatmenetből az is következik, hogy a jellemzőteret a legtöbb esetben nem is ismerjük.

A tanítás során egy optimalizációs problémára keressük a megoldást. A kifejezés a következőképpen néz ki:

$$\mathrm{J(w)} = \frac{1}{2} * \underline{w}^T * \underline{w} + C * \sum_{i=1}^{p} e_i \tag{6}$$

A minimalizálandó kifejezést tehát a J(w) függvény jelöli. C egy korrekciós tényező, amely a mintahalmazokból kilógó minták miatt került bevezetésre. Segítségével bizonyos mértékig ki tudunk küszöbölni kiugró minták által okozott torzulásokat, mert hagy egy biztonsági sávot az elválasztandó halmazok között. A minimalizálást ezen kifejezés alapján Lagrange multiplikátoros eljárással oldhatjuk meg. Fel kell írnunk a következő egyenletet:

$$L\left(\underline{w}, \underline{e}, \underline{\alpha}, \underline{\gamma}, b\right) = \frac{1}{2} \cdot \underline{w}^T \cdot \underline{w} + c \cdot \sum_{i=1}^{P} e_i - \sum_{i=1}^{P} \alpha_i \cdot \left\{d_i \cdot \left(\underline{w}^T \cdot \underline{x_i} + b\right) - 1 + e_i\right\} - \sum_{i=1}^{P} \gamma_i \cdot e_i$$

(7)

Ezt w,b,e_i-k szerint minimalizálni kell, és egyúttal α_i-k és γ_i-k szerint maximalizálni. A megoldás további lépéseit csak vázlatosan felsorolnám:

- w,b, és e szerinti deriváltakat kell felírni,
- a gradiensek értékének 0-vá tételével kapott összefüggéseket beírjuk a Lagrange egyenletbe,
- kapunk egy duális feladatot, amit kvadratikus programozással megoldhatunk,
- az eredményben lesznek 0 értékű α_i-k. Ezek jelölik azon x_i vektorokat, amelyek nem vesznek részt a hipersík meghatározásában. Minden olyan x_i, amelynek 0-tól különbözik az α_i paramétere, egy szupport vektor.

Az SVM megvalósításához a LIBSVM nevű C# alapú toolboxot használtuk, amelyre alapulva felépítettük a felismerőt.

5 Előfeldolgozás

A gépi felismerés egyik alapvető eleme, hogy a mintáinkból (amely esetünkben elsődlegesen egy nyomáshullám digitális ábrázolása) valamilyen eljárással kiemeljük a lényeget. Az ilyen eljárásokat előfeldolgozásnak is hívjuk. A lényegkiemelés során olyan információkhoz jutunk, amely sokkal jobban jellemzi a problémánkat, döntéshelyzetünket, mint az eredeti mintáink.

A hangminták előfeldolgozását a beszédkutatásokban széles körben használt Praat fonetikai elemző szoftverrel végeztük el. A Praat program sok különböző számítási algoritmust implementál, melyek számunkra különösen hasznosnak bizonyultak. Szerencsére a szoftver rendelkezik egy szkriptnyelvvel, amelyen keresztül automatizálható volt az előfeldolgozás. A következő részben szeretnénk bemutatni a statisztikai analízis, illetve a tanítás során használt paramétereket, illetve azok számítási módját. [21, 22]

A kiszámított paraméterek a következőek:

- lokális jitter
- ddp jitter
- lokális shimmer
- dda shimmer
- HNR (Harmonics to Noise Ratio)
- zöngés-zöngétlen arány

5.1 Harmonikus-zaj arány

A HNR paraméter összehasonlítja a harmonikus komponensek energiáját a zaj jellegű komponensek energiájával. A definíció alapján a következő összefüggés írható fel:

$$HNR = 10 * log \frac{E_H}{E_Z} \tag{8}$$

Ehhez azonban tudnunk kell a jel harmonikus komponenseinek energiáját (E_H), és a zajkomponensek energiáját (E_Z), amihez viszont szét kell választani a jelet ezen két összetevő összegére.

Alapvetően két számítási elvről beszélhetünk, melyek különböző típusú jelek periodikusságának meghatározására használatosak. A spektrumban, vagy az abból számított kepsztrumban (a későbbiekben részletesebben lesz erről szó) keresett csúcsok, melyek helye

közvetlenül az alaphang frekvenciáját adják meg lényegesen kevésbé robosztusak, mint a Praat által használt korrelációs módszerek. Ilyenek az egymástól lényegesen nem eltérő autó-, és keresztkorrelációs módszerek.

A jel komponensekre bontását is autókorrelációs függvény segítségével tudjuk megtenni, mely a következő alakban írható fel:

$$r_x(\tau) = \int x(t) * x(t+\tau)dt \tag{9}$$

ahol x(t) a hangminta időtartománybeli reprezentációja. A fenti összefüggés τ = 0 esetben pontosan a jel energiáját adja, és itt globális maximuma is van. Amennyiben másik globális maximumot is detektálhatunk τ != 0 érték esetén, akkor periodikus a jelünk. Természetesen ezt ki fogjuk használni az alapperiódus meghatározására, azonban a gyakorlati esetekben lényegében mindig csak lokális maximumokat találhatunk. Az egyes lokális maximumokat súlyozzuk, mégpedig a szakasz végén bemutatott R értékekkel. Az autókorrelációs függvénynek ott van lokális maximuma normál zajos esetben, ahol periodikus összetevő van. Az autókorrelációs függvényt felírhatjuk a következő alakban is:

$$r_x(\tau) = r_H(\tau) + r_N(\tau) \tag{10}$$

$$r_x(\tau_{max}) = r_H(T_0) = r_H(0) \tag{11}$$

Ahol T_0 a periódusidő. Ennek a felírási formának a jel energiájával normalizált alakja:

$$r'_x(\tau_{max}) = \frac{r_H(0)}{r_x(0)} \tag{12}$$

A következő lépés, hogy mindent zajnak tekintünk, amely nem periodikus összetevő:

$$1 - r'_x(\tau_{\max}) = \frac{r_N(0)}{r_x(0)} \tag{13}$$

Láttuk, hogy a jelkomponensek energiája megegyezik az autókorrelációs függvényeik értékével a 0 helyen véve. Ezek alapján felírható a HNR paraméter sokkal használhatóbb megfogalmazása:

$$HNR = 10 * log \frac{r'_x(\tau_{max})}{1-r'_x(\tau_{max})} \tag{14}$$

A program működéséhez természetesen ez így nem elegendő. Először is beszélnünk kell arról, hogy a jelet ablakozzuk, amikor egy adott szakaszát vizsgáljuk, és ezzel korlátozzuk a maximális mérhető HNR értéket. Legoptimálisabb értéket Hanning ablakkal, és az ablakban lévő megfelelő számú periódussal érhetünk el, méghozzá maximum 80 dB-es érték körül. Hamming ablakkal 50 dB-t, míg szögletes megoldás esetén a 40-et sem érhetük el. Ez

azonban egészséges beszéd esetén sem kerül megközelítésre, ugyanis a legnagyobb értékek sem mennek 30 dB fölé (általában 20 dB alatt vannak a mért értékek).

Az ablakozott jel autókorrelációjából könnyen megkaphatjuk az adott részhez tartozó eredeti jel autókorrelációját, hogyha elosztjuk az ablakozó függvény autókorrelációjával. [21] Ennek segítségével visszavezethetjük az ablakozást eredményét az eredeti jel HNR értékére:

$$r_x(\tau) = \frac{r_a(\tau)}{r_w(\tau)} \tag{15}$$

Fontos még beszélnünk az algoritmus paramétereiről, melyek beállításától függnek a felismert csúcsok, a lokális maximumok súlyozása, ezeken keresztül pedig a mért eredmények. A használt ablak méretét a minimálisnak beállított alaphang frekvenciája határozza meg. Az ablakozás mérete ezen frekvencia reciprokának (a periódusidőnek) a háromszorosa. A számítási módszer pontossága természetesen függ az ablak méretétől. Minél többször fér bele a mért hang periódusideje az ablakba, annál pontosabb a mérés. Maximális pontosság eléréséhez körülbelül 60 alaphang periódusra lenne szükség, ami azonban folyamatos beszéd esetén lényegében sosem áll elő.

Azt, hogy az adott ablak hanghoz tartozik-e, és értékelhető-e az autókorrelációs függvény kimenete lokális maximumként, két paraméter megadásával tudjuk befolyásolni. Ezek a paraméterek a voicing és a silence threshold (hangképzési és csend küszöbértékek). A hangképzési küszöb alapértelmezett értéke a számítások során 0.45, ami azt jelenti, hogy az ezen érték fölötti mért autókorrelációs értékek fognak számítani a lokális maximumkeresésben.

Patológiás esetekben akár 0.2-ig le kell menni ezen értékkel, mert annyira rossz minőségű a hang, hogy különben nem ismeri fel a program a zöngét, erre azonban az automatikus feldolgozás során nincsen lehetőség. A számítások idejét sokszorosára növelné ugyanis, ha ezen határokat adaptívan szeretnénk megvalósítani. Ezen korlát miatt jelentős nehézségekkel kellett szembenéznünk a fejlesztések során. A csönd küszöb azt adja meg, hogy milyen érték alatt tekintjük a jelet csöndnek.

A két értékből, valamint a lokális és globális csúcsértékekből számítódik az adott ablak súlya:

$$R = VoicingThreshold + \max(0.2 - \frac{\frac{local\ abs\ peak}{global\ abs\ peak}}{\frac{SilenceThreshold}{1 + VoicingThreshold}})$$

(16)

Az R paraméter segítségével súlyozva tudjuk eldönteni, hogy a lehetséges jelöltek közül melyik lokális maximumot fogadjuk el az alapperiódust meghatározónak. Jól látható, hogy a mérési elv nagy pontosságot tesz lehetővé megfelelően nagy ablakok alkalmazásával, ám ezt mi nem tudjuk kihasználni, mivel folyamatos beszédet dolgozunk fel, ahol egy-egy hanghoz kevés periódus tartozik.

5.2 Jitter és shimmer

A jitter értékek a beszélő alaphangjának ingadozását, annak kitéréseit jellemzik. A shimmer paraméter a mért nyomáshullám ingadozását mutatja meg nekünk.

A HNR számításánál autókorrelációt alkalmaztunk a periódusok komponensek meghatározására. A periódusok pontos helyére hasonlóképpen szükségünk van a különböző jitter és shimmer paraméterek számítása esetén is. A bemutatott perióduskereső eljárást, és a beállítható küszöbértékeket ezen számítások során is használjuk.

Az algoritmus meghatározza a súlyozott korrelációs csúcsok segítségével az alaphang periódusidejét, majd egy vizsgálati ablakkal végighaladunk a hangmintán. Minden vizsgálati ablakban amplitúdó maximumot keresünk, melynek környezetét nevezzünk bázisnak. A periódusidő 0.8-szeresén belül keressük a bázishoz leghasonlóbb hullámformát korrelációszámítás alapján. Így meghatározzuk a periódusok helyét a maximumokhoz kapcsolva. [22] Ezek a csúcsok valójában a glottális impulzusokat jelentik, vagyis a hangszalagok nyitását.

Amint megvannak a periódusok az adott szakaszra (N darab), a következő összefüggésekkel számítja a program az egyes értékeket:

$$jitter_{local} = \frac{N * \sum_{i=1}^{N-1} |T_i - T_{i+1}|}{(N-1) * \sum_{i=1}^{N} T_i} \quad (17)$$

$$jitter_{ddp} = \frac{N * \sum_{i=2}^{N-1} |2 * T_i - T_{i-1} - T_{i+1}|}{(N-1) * \sum_{i=2}^{N-1} T_i} \quad (18)$$

Ahol T-k az egymás utáni periódusok periódusideje. Ehhez hasonlóan az egymás utáni periódusokhoz tartozó A amplitúdókkal a shimmer értékek:

$$shimmer_{local} = \frac{N*\sum_{i=1}^{N-1}|A_i - A_{i+1}|}{(N-1)*\sum_{i=1}^{N} A_i} \tag{19}$$

$$shimmer_{dda} = \frac{N*\sum_{i=2}^{N-1}|2*A_i - A_{i-1} - A_{i+1}|}{(N-1)*\sum_{i=2}^{N-1} A_i} \tag{20}$$

Természetesen a korrelációs algoritmusoknak is vannak hátrányaik, melyek jelentős problémákat vetnek fel a későbbiekben. Az alaphangok viszonylag alacsonyak is lehetnek, így kilóghatnak a mérési tartományunkból. Mindemellett a hangképzési és csönd küszöb paraméterek erősen tudják befolyásolni a mérés eredményét, ha rossz minőségű hangot használunk.

5.3 Zöngés-zöngétlen arány

Az előfeldolgozás során kiszámítunk egy olyan paramétert, amelynek alapja saját ötlet, célja pedig azon hangminták megkeresése, amelyekre a használt Praat algoritmusok nem képesek értékelhető eredményt adni. A nehézséget természetesen az okozza, hogy miként tudjuk detektálni, hogy a hang nem jól mérhető, lehetőleg azokra a mérési eljárásokra támaszkodva, amelyeket egyébként is használunk.

A felismerő által hozott rossz döntések, és problémás minták vizsgálata során egyértelművé vált, hogy a legtöbb probléma a nagyon beteg páciensek hangmintáival van. A bemutatott mérési eljárások alapján nyilvánvaló, hogy ha a beteg nem, vagy csak nagyon korlátozott mértékben képes zönge képzésére (nem rezeg a hangszalag), akkor az autókorrelációs perióduskeresési eljárás nem fog értékelhető maximumokat találni.

Természetesen, ha nincsenek periódusaink, mert a mért hangokra nem teljesül az alapvetően feltételezett kvázistacioner jelleg, akkor a mérések eredménye lényegében véletlenszerű lesz. A probléma enyhítésére találtuk a zöngés-zöngétlen arány mérését, amely reményeink szerint segít az ilyen esetek kiválogatásában az automatikus osztályozás során.

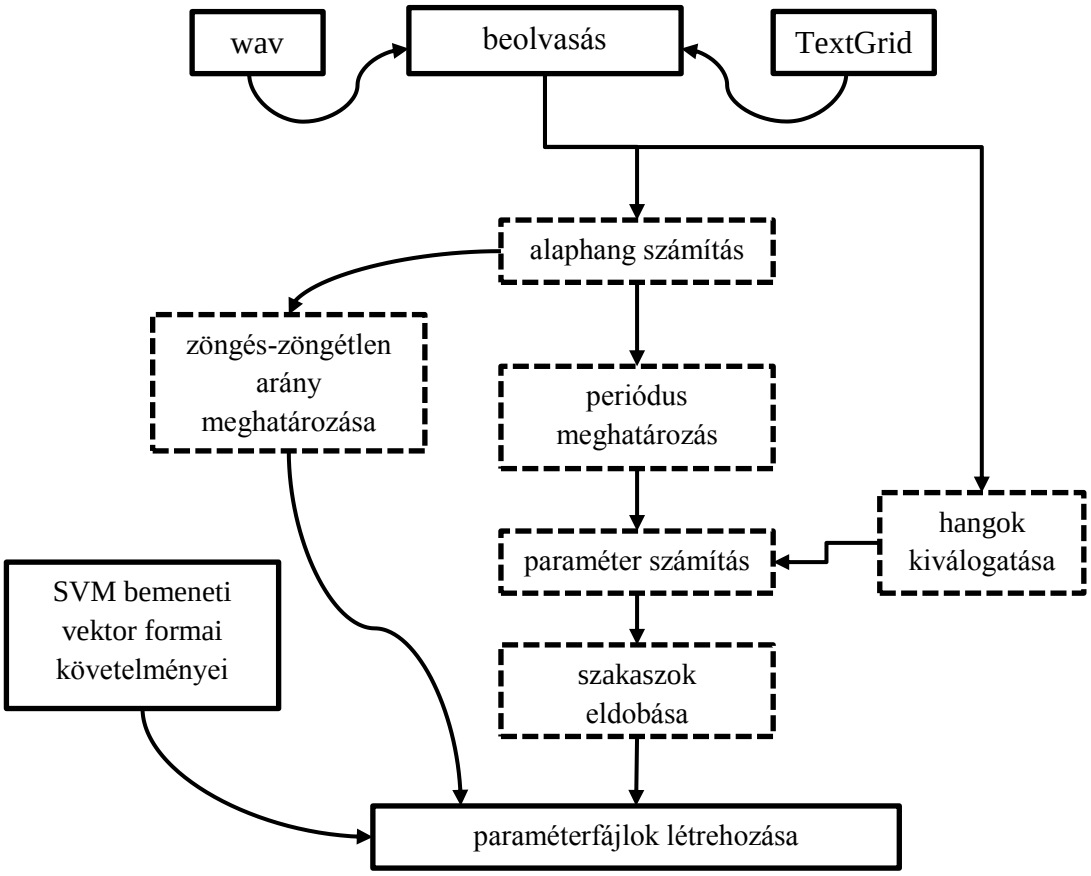
Az arány meghatározásának alapja az alaphangszámítás előzőekben bemutatott autókorrelációs módszerrel történő meghatározása. A program az alaphangot keretenként számítja, és a kereteknek van egy lépésköze. Ebből következik, hogy hány keret van egy időtartam alatt. A metódus során a vizsgált személy minden egyes hangjához tartozó minden

keretet megvizsgálunk, hogy tartalmaz-e alaphangot, és feljegyezzük az eredményeket. Végül az összes zöngés és zöngétlen kereteket elosztjuk egymással.

A mérés továbbfejlesztéseként elkészítettünk egy olyan módosított verziót is, amely csak az egyébként is zöngés hangokra vizsgálja az arányt, de ez nem okozott szignifikáns különbséget az eredményekben (mivel a páciensek jó közelítéssel azonos szöveget mondanak be).

5.4 Előfeldolgozás a Praatban

Az előzőekben felsorolt paramétereket a hangfájlokból egy Praat szkript segítségével nyerjük ki. A feldolgozó algoritmus bemenete az adatbázisban található hang és címkefájlokból áll. Természetesen csak olyan fájlokat tudunk felhasználni, amelyhez az előzőekben már elkészült a címkézés. A szkript különböző részein mindketten dolgoztunk, végül összeállítottuk egységes rendszerré. Az alábbi ábrán az előzőekben felsorolt eljárások egymásra épülése látható.



4. ábra. Az előfeldolgozás vázlatos működése

Az előfeldolgozás során először beolvassuk a hangfájlokat, és a címkefájlokat is. A lokális és ddp jitter, a lokális és dda shimmer valamint a HNR értékek kiszámításához szükséges meghatározni az alaphangot, és a periódusok pontos helyét. A Praat megkötései miatt a teljes hangfájlra számítjuk ezeket.

A legfontosabb eddig nem tárgyalt lépés a hangok kiválogatása. Ennek során a meghatározott SAMPA karakterre keresünk, például „E”, és minden találatra kiszámítjuk a paramétereket. Fontos, hogy csak a folyamatos beszéd hangjai között keresünk, kitartott hangok között nem. A témához kapcsolódó publikációk általában kitartott hangokkal dolgoznak, mint azt az adatbázisokról szóló részben már kifejtettük, azonban több publikáció kikacsint a folyamatos hangok felé is. Mivel a kérdés nincsen lezárva, ezért korábbi TDK munka keretében már foglalkoztunk a problémával. Kimutattuk, hogy az általunk alkalmazott szubjektív megközelítésre sokkal jobban alkalmazhatóak a folyamatos beszéd magánhangzói.

Szeretnék kitérni azon művelet szükségességére, amely során eldobunk egyes szakaszokat. A bemutatott mérési eljárások feltételezik, hogy az aktuálisan mért jelszakasz elégséges számú periódust tartalmaz. Amennyiben ez nem teljesül, mert túl gyors a beszéd, túl mély az alaphang, vagy nincsen zönge, akkor lehetséges, hogy az egyes függvények visszatérési értéke undefined (meghatározatlan) lesz. Az ilyen esetek detektálására a periódus meghatározása után lenne először lehetőség, de mivel a paraméterek kiszámítása ettől a ponttól már viszonylag csekély erőforrást vesz igénybe, ezért a szakasz értékelhetetlenségét a kimenetek ellenőrzésével detektáljuk. Amennyiben bármely paraméter meghatározatlan értéket vesz fel, akkor a szakaszhoz tartozó minden paramétert eldobunk, mert a mérések jelentős pontatlanságot mutathatnak.

A most bemutatott előfeldolgozási eljárást alkalmaztuk mind a statisztikai kiértékelésnél, mind pedig az automatikus osztályozási eljárásnál.

6 Statisztikai analízis

6.1 Az elemzés feladata és módszere

A kóros és egészséges mintákat tartalmazó felcímkézett adatbázist automatikus osztályozási feladatokra szeretnénk használni. Keressük a minták egy olyan reprezentációját, amely a lehető legtöbb redundanciát kiszűri, és a lehető legjobb elkülönítést teszi lehetővé. Ehhez szükséges egy olyan átfogó elemzést elvégezni, amely támpontot nyújthat nekünk az aktuális, és lehetséges jövőbeli feladataink megoldásához. Ezt az elemzést, és a hozzá szükséges programozási feladatok megoldását én, Imre Viktor hajtottam végre. A statisztikai analízis a következő kérdésekre keresi a választ:

- Elégséges a saját adatbázisunk használata, vagy érdemes felhasználnunk a Babel magyar nyelvű hangadatbázis egészséges mintáit is?
- Várhatóan mely paraméterek lesznek a legmegfelelőbbek?
- Melyik hangokat használjuk fel az adatbázisból (E, O, i, u)?
- Milyen további lehetőségeket hordoz magában a módszer?

Az elemzéshez használt paraméterek számítási módját, és a felmerülő problémákat már az előfeldolgozásról szóló részben bemutattuk. A statisztikai analízis során olyan általános vizsgálatokat folytattam, amelyek értékes információt szolgáltatnak ezen dolgozat keretei között tárgyalt osztályozáshoz, valamint további fontos tanulságok levonását is lehetővé teszik.

Négyféle magánhangzót használtam fel a vizsgálatokhoz. A magánhangzók csoportja a magyar fonémák között az egyetlen, amelynek gerjesztése zöngés, és az akusztikai produktumot nem befolyásolják zörej jellegű komponensek, így ez a legalkalmasabb fonématípus a vizsgálataimhoz. Az „E” hangot azért választottam ki, mert ez van a legjobban reprezentálva az adatbázisban. Jól reprezentált még az „O” hang is, azonban az „O”, „i” és „u” kiválasztásának szempontja az volt, hogy megjelenítsem a hangképzés változatos eseteit.

A statisztikai analízis során vizsgáltam a különböző mintacsoportok szignifikáns elkülöníthetőségét a következő esetekben, melyekben a H osztályok az RBH kód általános minőségi jellemzőjét jelentik (H0 egészséges, H1-től H3-ig pedig kóros osztályok).

- H0 és H1 minőségi osztályok elkülöníthetősége
- H1 és H2 minőségi osztályok elkülöníthetősége
- H2 és H3 minőségi osztályok elkülöníthetősége
- H0 és (H1+H2+H3) minőségi osztályok elkülöníthetősége, ahol (H1+H2+H3) a kóros, míg H0 az egészséges minták összességét jelenti
- B és H1 osztályok elkülöníthetősége, ahol B a Babel általános célú magyar beszédadatbázisból vett egészséges minták halmazát jelöli
- B és (H1+H2+H3) osztályok elkülöníthetősége
- (H0+B) és (H1+H2+H3) összes egészséges, és összes kóros minták elkülöníthetősége a Babellel kiegészíttet adatbázison

Az SVM számára bemenő paraméterként használt jellemzővektorok felépítéséről még nem esett szó. Az osztályozás során szeretnénk két osztály között dönteni, vagyis egészségesnek vagy kórosnak nyilvánítani a páciens hangját. Mivel az SVM matematikai modellje nem teszi lehetővé a változó számú minta használatát, ezért az egyes személyek adataira valamilyen aggregációt kell végrehajtanunk. Ez azt jelenti, hogyha egy ember minden „E” hangjára megmérjük a lokális jitter értékeket, akkor ebből szeretnénk tetszőleges 1-nél nagyobb számú „E” hangminta esetén fix számú paramétert kapni.

Természetszerűleg adódott, hogy számítsuk ki a minták átlagát. Ez egy olyan értéket ad nekünk, amely a személy átlagos hangminőségét jellemzi, azonban szerettünk volna az értékek ingadozására jellemző statisztikát is használni. A legegyszerűbb lehetőség a mért paraméterek szórásának a felhasználása. Így tulajdonképpen minden mért paraméterhez személyenként kettő adatunk áll rendelkezésre: egy átlag és egy szórás. A statisztikai analízis során azonban nem végeztem személyenkénti aggregációt.

Szeretném hangsúlyozni, hogy a statisztikai kiértékelés során az egyes emberek hatását csökkentettem azáltal, hogy nem a személyenkénti átlagokat használtam fel. Az egyes döntési csoportokba tartozó minden ember minden mintáját összevontam, így nagy halmazokat alkottam. Az összehasonlítást ilyen módon képzett halmazok között végeztem el. Ez a módszer megfelel a korábban említett elveknek, mivel értékes információkat szolgáltat a statisztikai alapú automatikus osztályozás számára, azonban ennél sokkal általánosabban következtetések levonását is lehetővé teszi, ami más jellegű (pl. fuzzy) osztályozók, vagy távolabbi célok alapjául szolgálhat.

6.2 A szignifikancia vizsgálatról röviden

A statisztikai analízis során megvizsgáltam, hogy szignifikánsan megkülönböztethetőek-e egymástól az egyes osztályok. Ennek során feltételeztem, hogy az előzőek szerint kialakított csoportok normális eloszlást követnek, és kétmintás t-próbát hajtottam végre 95%-os szignifikancia szint mellett. A következő módon számíthatjuk a t próbastatisztika értékét.

$$t = \frac{\overline{X}-\overline{Y}}{\sqrt{(n-1)*s_X^2+(m-1)*s_Y^2}} * \sqrt{\frac{n*m*(n+m-2)}{n+m}} \tag{21}$$

A (22) kifejezésben X és Y valószínűségi változók átlagának különbsége szerepel az első tört számlálójában. Az s_X és s_Y ugyanezen valószínűségi változók szórása, n és m pedig a minták száma, melyekből az átlagokat és szórásokat számoltuk. A t próbastatisztikát összehasonlítjuk a Student-t eloszlás táblázatából kapott t_t értékkel, amely a szignifikancia szint és a szabadsági fokok számának (n+m-2) függvénye. Amennyiben

$$|t| \geq t_t \tag{22}$$

akkor adott szignifikancia szint mellett a két mintahalmaz átlaga szignifikánsan eltér egymástól. Ezek alapján tehát keressük azokat az eseteket, amikor t értéke a lehető legnagyobb t_t-hez képest, mert ezek jelentik a legjobban elkülönülő eloszlásokat. Mindezt Excel táblázatok és függvények segítségével valósítottam meg minden kiszámított paraméterre, és minden egyes vizsgált hangra.

6.3 Adatbázisok vizsgálata

Az analízis során először meg kellett állapítani, hogy szükségünk van-e a saját felvételeinken kívül plusz hangmintákra egészséges személyektől. Rendelkezésünkre áll a Babel általános célú magyar beszédadatbázis. Mivel a Babel készítésének célja az volt, hogy nemben és korban változatos személyekkel készült felvételeket tartalmazzon, ezért ezekből válogatnunk kellett. Ezen válogatás után 34 egészséges személy bemondásait használtuk fel a saját egészséges felvételeink kiegészítésére.

Minden paraméterre (lokális jitter, ddp jitter, lokális shimmer…) megvizsgáltam, hogy okoz-e javulást a kétmintás t-próba esetén, de most csak néhány lényeges esetet emelnék ki ezek közül. A vizsgálatok során RBH kód elemei közül a H általános érzeti jellemző szerinti

besorolást alkalmaztam. A H = 0 érték szubjektíve egészséges, míg a H=1, 2 vagy 3 a hang minőségének (érdesség, rekedtség, levegősség... együttesen) fokozatos romlását jelölik.

Azt, hogy javít-e az eredményeinken a Babel használata, a következő esetben tudjuk legjobban eldönteni. Először összehasonlítottam a saját adatbázisunkban lévő egészséges (H0-ra értékelt) és a szubjektíven lehető legkevésbé rossz állapotú (H1) páciensek hangjait. A második esetben a Babel mintáit H0-ás besorolással láttam el, így a saját egészséges mintákat kiegészítettem. Kerestem azokat a határeseteket, amikor csak az egyik esetben van szignifikáns eltérés az eloszlások között.

A keresés eredményeként azt tapasztaltam, hogy minden ilyen határesetben a Babellel kiegészített adatbázisunkon tudtunk szignifikánsan elkülöníteni.

6. táblázat. A Babellel történő kiegészítés hatása a határesetekben

	osztályok	szabadsági fokok	Student-t	próbastatisztika	t > tt
"u" shimmer dda	H0 vs H1	799	1,96	1,61	nem
	H0+B vs H1	1039	1,96	4,65	igen
"u" shimmer local	H0 vs H1	799	1,96	0,11	nem
	H0+B vs H1	1039	1,96	2,48	igen
"u" HNR	H0 vs H1	799	1,96	0,5	nem
	H0+B vs H1	1039	1,96	4,85	igen

A fenti táblázatban szereplő értékeknél látszik, hogy minden esetben lényegesen javult a próbastatisztika értéke a kiegészített adatbázissal (ezt H0+B-vel jelöltem). További érdekes tanulság, amelyet a következő alfejezetben fogok tárgyalni, hogy az összes ilyen jellegű határeset „u” hangokon történt mérésekből származik.

Ezzel az eredménnyel azonban még nem elégedtem meg, hanem megvizsgáltam, hogy más hangok és paraméterek esetén is mutatnak-e javulást a próbastatisztika értékei. Az eddig bemutatott összehasonlítás során a Babellel történő kiegészítés hatását vizsgáltam a H0 és H1 csoportok elkülöníthetőségén. A következőkben az osztályozási feladatok szempontjából relevánsabb egészséges (H0,) és kóros (H1+H2+H3 = H123 jelöléssel) csoportok elkülöníthetőségére vett hatást tekintem.

7. táblázat. A Babellel történő kiegészítés hatása az osztályok közötti szignifikáns eltérések esetén

	osztályok	szabadsági fokok	Student-t	próbastatisztika	$t > t_t$
"E" jitter ddp	H0 vs H123	8543	1,96	20,85	igen
	H0+B vs H123	10524	1,96	24,37	igen
"O" jitter ddp	H0 vs H123	7179	1,96	18,67	igen
	H0+B vs H123	9020	1,96	20,27	igen
"O" shimmer dda	H0 vs H123	7179	1,96	21,97	igen
	H0+B vs H123	9020	1,96	29,06	igen
"i" HNR	H0 vs H123	2718	1,96	13,16	igen
	H0+B vs H123	3552	1,96	21,40	igen

A 7. táblázatban látható néhány példa majdnem minden hang-paraméter párosra hasonlóan alakul. Ha a saját adatbázisunkon is szignifikánsak a különbségek, akkor ez még tovább nő a Babellel kiegészített esetben, bár a javulás mértéke erősen ingadozik. Néhány itt be nem mutatott esetben, ha a t nem nagyobb, mint a Student-t értéke, akkor egy egészen kicsit (egy-két századdal) kisebb értékek is kijöttek eredményül a kiegészített esetre. Ezek azonban nem releváns eredmények, mivel azokat a hang-paraméter kombinációkat, ahol nem szignifikáns a különbség, nem használjuk osztályozó eljárás során.

Az eredmények alapján egyértelműen kimondható, hogy az egészséges mintáinkhoz érdemes hozzávenni a Babel kiválogatott hangmintáit, mert a statisztikai analízis alapján lényegesen növelhetik az osztályozást eredményességét.

6.4 Paraméterek és hangok vizsgálata

Ebben az alfejezetben kifejtem, hogy a korábban bemutatott paraméterek és hangok közül melyeket, és miért javasolom használni a legjobb osztályozási pontosság elérése érdekében. Az elemzés módszere természetesen az előzőekben is megfogalmazottakkal egyezik. Szeretném kiemelni, hogy az előző fejezet következtetései alapján már csak a Babel hangadatbázissal kiegészített saját adatbázis használatával kapott eredmények bemutatását tartom hasznosnak.

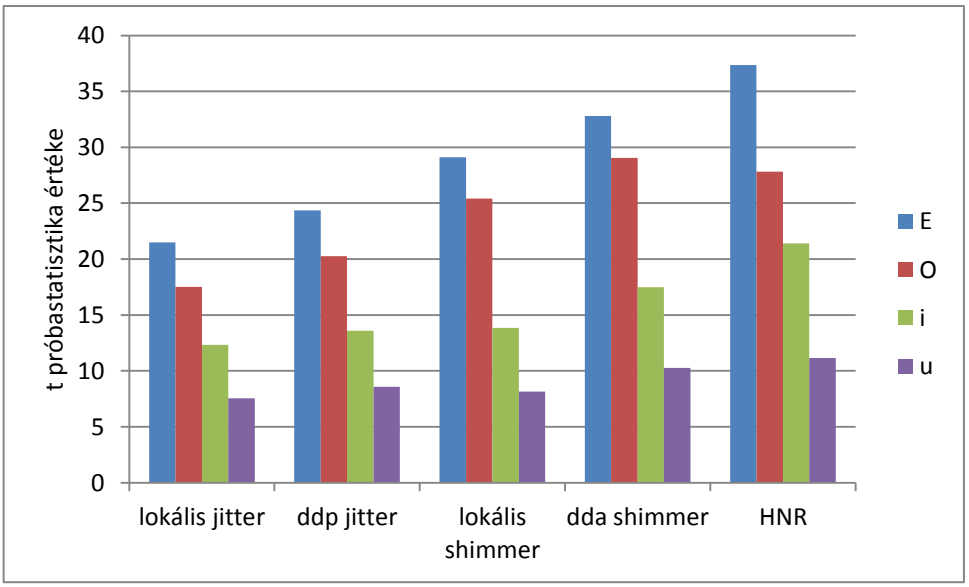
A mért hangok vizsgálatával folytatom az analízis eredményeinek bemutatását. A saját adatbázisban lévő hanganyagok mindegyike egy felolvasott népmesét tartalmaz, amelyben viszonylag kötött számú magánhangzó van (eltéréseket a félreolvasások okozhatnak). A Babelből átemelt felvételek többféle szöveget tartalmaznak, amelynek eredménye, hogy az

ezekben található magánhangzók száma átlagosan is eltér egymástól. A végleges darabszámokat ezen felül befolyásolják a mérési eljárások is, mivel azon hangok, amelyeken nem tudunk méréseket végezni nem is értékesek számunkra. Az összesített darabszámok a nyolcadik táblázatban láthatóak.

8. táblázat. A mérhető magánhangzók darabszámai

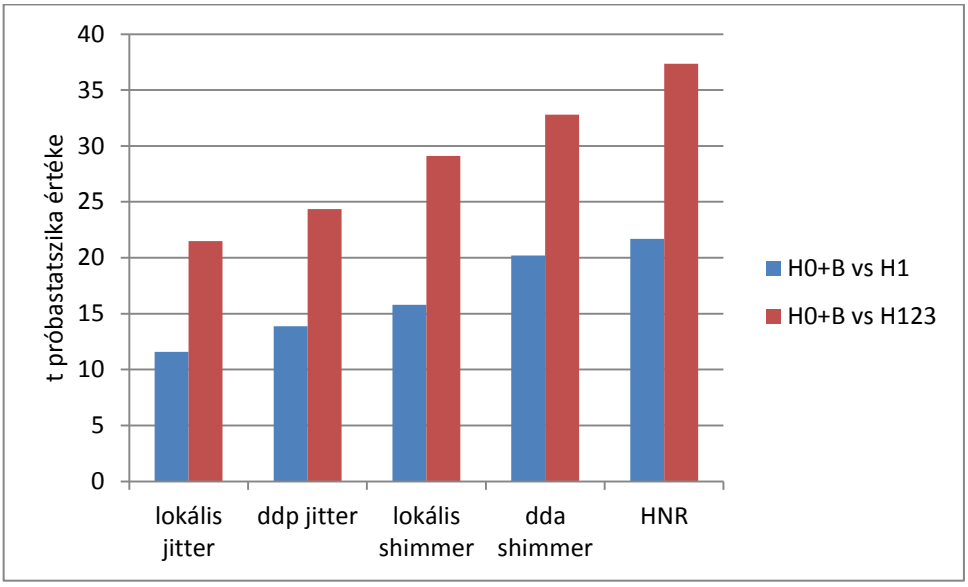
	H0 + B	H1	H2	H3	Szumma
E	5298	3109	1399	720	10526
O	4631	2636	1175	580	9022
i	1862	1015	452	225	3554
u	636	405	179	104	1324

A H0+B továbbra is a Babellel kiegészített egészséges halmazt jelenti, míg a H1, H2 és H3 halmazok szubjektív minőségi romlást fejeznek ki. Nyilvánvaló, hogy statisztikai értelemben a nagyobb összes mintaszám jobb eredményekre vezethet a kétmintás t-próba képlete alapján (21. egyenlet). A magánhangzók használhatóságát automatikus elkülönítési célokra jelen esetben elsődlegesen a darabszámuk határozza meg. Az „u” és „i” hangok alulreprezentáltak, így nem hasonlíthatjuk össze az „E” és „O” magánhangzókkal. A következő ábrán látható hangok közötti eltérések alapján tehát nem következtethetünk fizikai jellegű okokra a darabszámok drasztikus eltérése miatt. Az ábrán a t próbastatisztika értékei minden esetben szignifikáns eltérést jelentenek (0,05-ös szignifikancia szint mellett Student-t = 1,96 a legkisebb mintaszámokra is) az egészséges és kóros minták között.



5. ábra. A t próbastatisztika értékei minden paraméterre és minden hangra H0+B és H123 közötti szignifikancia vizsgálat esetén

Annak érdekében, hogy a különböző paraméterek alapján történő elkülöníthetőséget megfelelően tudjam értékelni, hangonként külön kell kezelnem a mérési eredményeket. A próbastatisztikák tüzetes átvizsgálásával egyértelművé válik, hogy a különböző hangok esetén nagyon hasonlóan viselkednek az egyes paraméterek, amint ez az előző ábrán is látható. A következőkben már csak a két nagy darabszámú magánhangzó mintáin végrehajtott mérések eredményeit mutatom be.



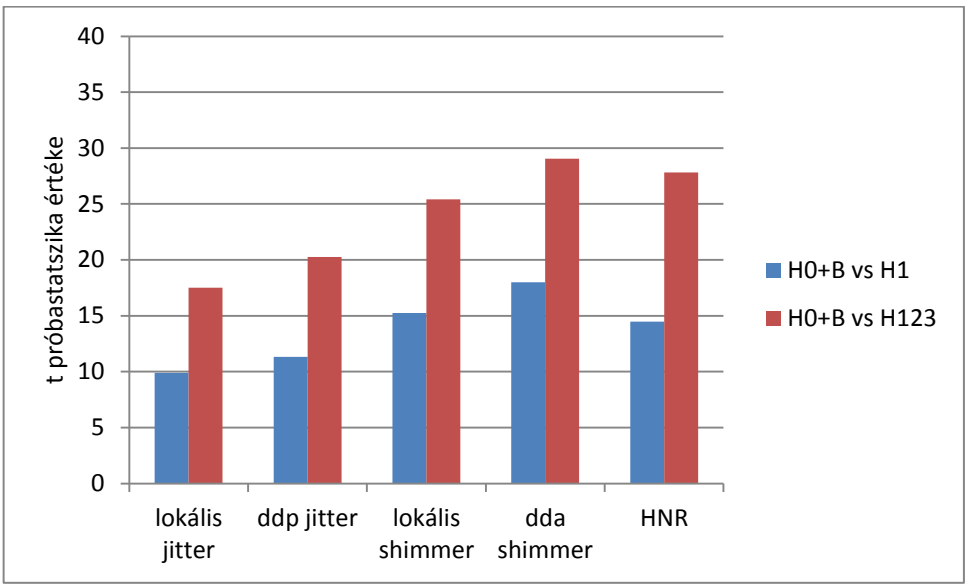
6. ábra. Fontosabb t próbastatisztika eredmények „E” hang esetén

A 6. ábrához tartozó Student-t értékek 0,05-ös szignifikancia szint mellett 1,96-ot vesznek fel, melyeknél minden ábrázolt esetben nagyobb próbastatisztika értéket kaptunk, ami minden esetben szignifikáns eltérést jelent.

Fontosnak tartottam mind az egészséges és kóros, mind pedig az egészséges és legkevésbé rossz állapotú mintahalmazok elkülöníthetőségének ábrázolását. Nyilvánvaló, hogy a H1 csoport sokkal kevésbé szignifikánsan különül el az egészséges mintáktól, mint az összevont beteg minták csoportja. Érdemes ezt vizsgálni, mert az automatikus döntési módszerek használatakor a határesetek megkülönböztetése lehet a legkritikusabb pont. Az ábra alapján kijelenthetem, hogy mind a két vizsgált megközelítés esetén minden paraméter alkalmas lehet az elkülönítésre, mivel az átlagok szignifikánsan eltérnek egymástól.

A legjobb eredményeket a harmonikus-zaj arány produkálta, amely a hang zajtartalmát jellemzi. Ennél valamivel gyengébb a két shimmer érték, amelyek az amplitúdó ingadozását jellemzik, és a leggyengébbek az alaphang ingadozását jellemző jitter paraméterek. A dda és ddp paraméterek jobban teljesítenek a lokális változatokhoz képest. Míg a lokális megoldás az aktuális periódust az következővel hasonlítja, addig a dda és ddp megoldások az aktuálisan

vizsgált periódust az előző és következő periódussal is. A nagyobb keretben végzett ingadozásvizsgálat tehát jobbnak bizonyul E hang esetén a kisebbhez képest.



7. ábra. Fontosabb t próbastatisztika eredmények „O” hang esetén

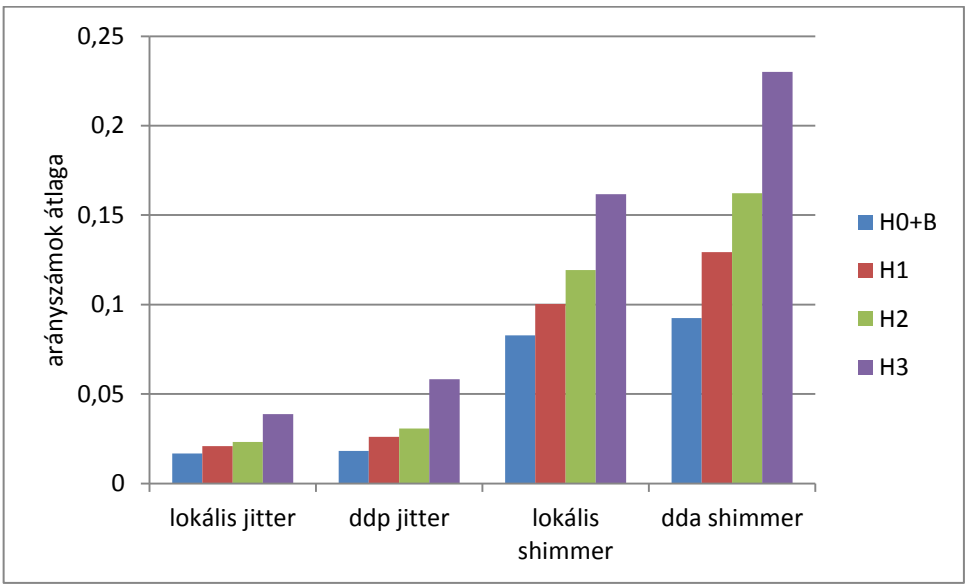
Az ábra az előzőhöz hasonlóan minden vizsgált paraméter próbastatisztika értékeit bemutatja két összehasonlítási esetben. Az eredmények nagyon hasonlatosak az „E” hangnál kapottakhoz, de két különbség van. Az elsőről már volt szó, de a H0+B és H1 összehasonlítása esetén is előkerül. Mivel az „O” magánhangzóból kevesebb darab van a hangfájlokban, mint az „E”-ből, így kisebb próbastatisztika értékek fognak hozzátartozni. A második észrevehető különbség, hogy a HNR a többi paraméterhez képest is jobban visszaesett, bár még így is a második legszignifikánsabban eltérést mutatja egészséges és kóros mintahalmazok között. Ennek oka valószínűleg a szövegekben található hangkapcsolatokban keresendő, mivel jelentősen csökkenhetnek a HNR értékek a hangok egymásra hatása miatt is.

6.5 A statisztikai analízis összefoglalása

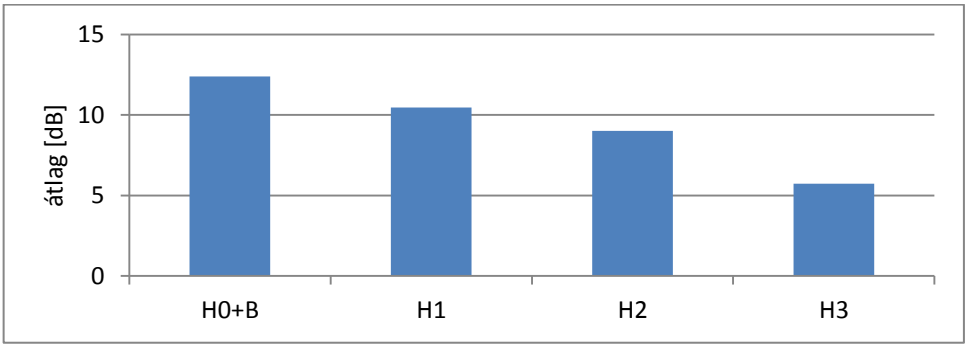
Az elemzés alapján tehát érdemes felhasználni a Babel kiválogatott egészséges hangmintáit a saját adatbázisunk egészséges hangmintáinak kiegészítésére, mivel jelentős javulás várható általuk az automatikus osztályozásban. A beszédhangok közül az „E” és az „O” tesztelése jó eredményeket hozhat, mivel mind az öt mért paraméter esetén szignifikáns különbségek mutathatóak ki az egészséges és beteg halmazok között, valamint az egészséges és a legkevésbé rossz hangminőségű beteg halmaza között is. Ezzel összhangban az összes

vizsgált paraméter alkalmas lehet a mintacsoportok elkülönítésére, így ezek tesztelését is érdemes lehet elvégezni.

Az elemzések során erősen hagyatkoztam az egészséges és beteg minták osztályozhatóságának vizsgálatára, azonban az értékelés részeként szeretnék kitérni a tovább mutató eredményekre. Csak az „E” hangra, mint legnagyobb számban reprezentált magánhangzóra mutatnám be, hogy az eddig felsorolt paraméterek szignifikáns eltérést mutattak minden minőségi osztály között.



8. ábra. A mért ingadozás paraméterek átlaga az egyes hangminőség osztályokban „E” hang esetén



9. ábra. A mért HNR paraméterek átlaga az egyes hangminőség osztályokban „E” hang esetén

A 8. és 9. ábrákon az egyes osztályokba tartozó minták mérési eredményeinek átlagát láthatjuk, melyek a 8. ábrán szigorú növekedést mutatnak. Ez logikusan elvárt, mivel a rosszabb hangminőséghez nagyobb ingadozás értékeket várunk. A 9. ábrán lévő HNR paraméterek osztályonkénti átlaga csökkenő tendenciát mutat, amely szintén az elvárt viselkedés. Ez előrevetíti annak lehetőségét, hogy az eddig is használt és bemutatott

lényegkiemelési eljárások segítségével képesek lehetünk szubjektív minőségi skála szerint osztályozni a páciensek hangminőségét. Egy ilyen osztályozó segítségével a hangterápiák során lehetőség adódna a betegek javulásának automatikus mérésére.

7 Automatikus osztályozás

A statisztikai analízis eredményire támaszkodva az osztályozási kísérleteket én, Barlangi Renáta végeztem el. Az osztályozáshoz a 4. fejezetben ismertetett LS-SVM motort használtam. Az osztályozási kísérleteket két nagy csoportra bontottam. Az első csoportba tartoznak az általunk összeállított és felcímkézett adatbázison végzett tesztek. Az egészséges és kóros minták meghatározása az RBH kód H paramétere alapján történt. 169 minta állt rendelkezésemre, 59 egészséges és 110 kóros hangfájllal végeztem az osztályozást.

A második csoportba azok az elkülönítési kísérletek tartoznak, amelyek a Babel adatbázissal kiegészített adathalmazon történtek. Így második esetben 203 mintával végeztem el az osztályozást: 93 egészséges és 110 beteg hangfelvétel.

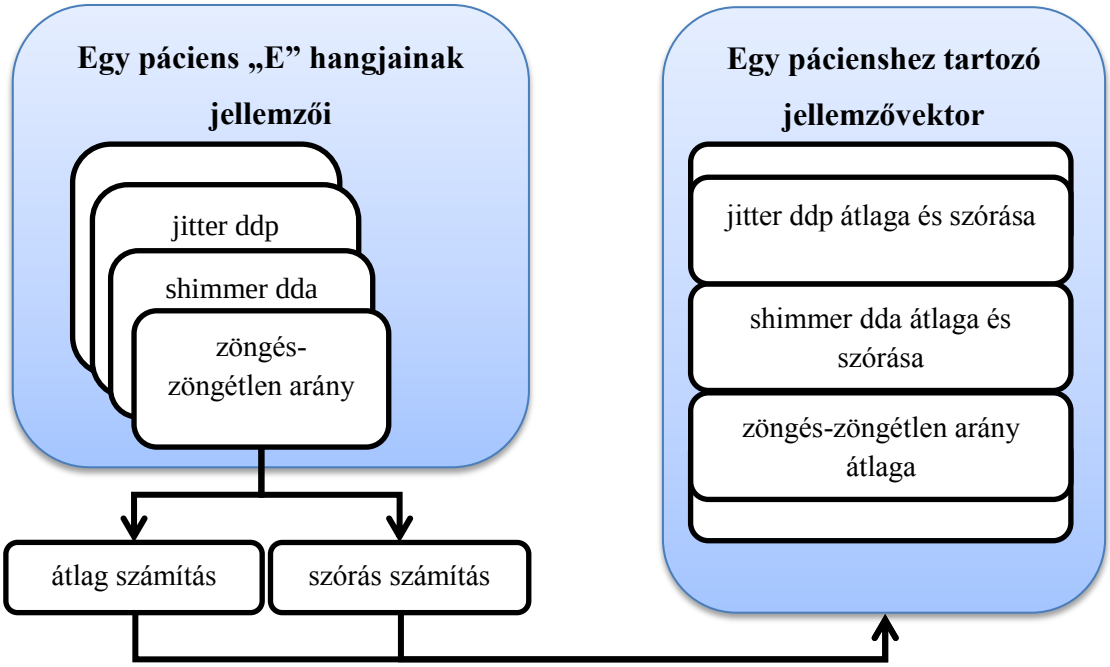
A csoportbontásra azért volt szükség, hogy megvizsgálhassam a Babel adatbázis mintáinak hatását az osztályozási pontosságra.

7.1 Bemeneti vektorok előállítása

Az osztályozásra a Beszédakusztikai Laborban fejlesztett SVM osztályozót használtam. Az osztályozó bemenetére vektorok kerülnek, amelyeket az előfeldolgozás kimeneti fájljaiból állítottam össze. Az lényegkiemelési eljárásokat az 5. fejezetben bemutatott módon hajtottam végre.

Először a lényegkiemelt paraméterek statisztikai paramétereit kellett kiszámíttatni. Ehhez egy külön programot használtam, amely a bemenetére kapott összes hangfájlhoz tartozó lényegkiemelt jellemzőhöz meghatározza a statisztikai paramétereket. Jellemzők alatt a jitter ddp, shimmer dda és HNR értékeket értem. Az általam használt statisztikai paraméterek a következőek voltak: átlag és szórás. A két elkülönítendő osztály pedig az egészséges és a beteg.

A fent említett program a lényegkiemelt jellemzők 28 darab, különböző statisztikai paramétereit számolja ki. Ezért minden egyes jellemzőre ki kellett választani, hogy mely statisztikai paraméterei kerüljenek a vektorokba. Legvégül beállítottam, hogy mely jellemzők és osztályok alapján szeretnénk a tesztelést elvégezni. Ezt követően a program elő tudta állítani a betanításhoz szükséges többdimenziós vektorokat. Fontos megemlíteni, hogy ezzel a konstrukcióval minden személyhez egy darab jellemzővektor tartozik és minden vektor azonos hosszú. Ebben a vektorban található, az összes kiválasztott jellemzőre számított statisztika érték, vagy értékek. Egy bemeneti vektor előállatását szemlélteti a 10. ábra.



10. ábra. SVM bemeneti vektorainak előállítása

A jellemzővektorok segítségével el lehet végezni az SVM betanítását. A tesztelések során teljes kereszt validációt alkalmaztam. Ilyenkor a program automatikusan alakítja ki a betanításhoz, valamint a teszteléshez szükséges halmazokat. Ez a folyamat a következőképpen áll össze:

1. A program a mintahalmazból kivesz egy darab vektort, ami egy pácienshez tartozik.
2. A maradék mintával elvégzi a betanítást.
3. A kivett vektorral elvégzi az osztályozási kísérletet.
4. A fenti lépéseket az összes vektorra megismétli és végül ad egy százalékos eredményt és egy tévesztési mátrixot az elkülönítésre. A tévesztési mátrix értelmezését a következő táblázat szemlélteti:

9. táblázat. A tévesztési mátrix felépítése

	B	E	Eredmény
B	Azon betegek száma, amelyeket a felismerő betegnek ítélt. (Helyes döntés)	Azon beteg száma, amelyeket a felismerő egészségesnek ítélt. (Helytelen döntés)	Az osztályozás találati aránya százalékban megadva.
E	Azon egészségesek száma, amelyeket a felismerő betegnek ítélt. (Helytelen döntés)	Azon egészségesek száma, amelyeket a felismerő egészségesnek ítélt. (Helyes döntés)	Az osztályozás találati aránya százalékban megadva.

Az osztályozás minden esetben RBF kernellel történt. Ilyenkor lehetőség van a gamma és C paraméterek változtatására. A C paraméter a biztonsági sáv szélességét határozza meg. Minél nagyobb az értéke annál nagyobb a megengedett hibahatár. A gamma pedig, az egyes minták hipertéri környezetükre gyakorolt hatását befolyásolja. A következőkben a különböző tesztesetek eredményeit mutatom be.

7.2 Osztályozási kísérletek a saját adatbázison

Először a saját adatbázisunkon végeztem elkülönítési teszteket, 110 kóros és 59 egészséges mintával. Különböző összetételű és színképű magánhangzókat választottam ki az osztályozás elvégzésére: „E”, „O”, „u” és „i”. Minden magánhangzó esetén négy különböző bemeneti vektorral végeztem el az osztályozást. A bemeneti vektorok felépítése:

1. jitter ddp átlaga és szórása, shimmer dda átlaga és szórása
2. jitter ddp átlaga és szórása, shimmer dda átlaga és szórása, zöngés-zöngétlen arány (továbbiakban: zz arány) átlaga
3. jitter ddp átlaga és szórása, shimmer dda átlaga és szórása, csak magánhangzókra vett zöngés-zöngétlen arány (továbbiakban: zzm arány) átlaga
4. jitter ddp átlaga és szórása, shimmer dda átlaga és szórása, zz arány átlaga, zzm arány átlaga

A jitter ddp és shimmer dda paraméterekre azért esett a választásom, mert a statisztikai analízis során jó eredményeket mutattak, valamint korábbi tesztjeim során velük lehetett elérni a legnagyobb pontosságú osztályozást. [16] Ezen kívül megvizsgáltam, hogy javulás

érhető-e el a zz, vagy zzm arányok felhasználásával. Az eredmények az 10. táblázatban láthatóak.

10. táblázat. Saját adatbázison elért elkülönítési pontosságok

Figyelembe vett értékek	„E”	„O”	„u”	„i”
jitter ddp, shimmer dda átlaga és szórása	**83.4%**	82.8%	68%	77.5%
jitter ddp, shimmer dda átlaga és szórása, zz átlaga	82.2%	**83.4%**	68%	77.5%
jitter ddp, shimmer dda átlaga és szórása, zzm átlaga	80.5%	81.1%	70.4%	75.7%
jitter ddp, shimmer dda átlaga és szórása, zz és zzm átlaga	77,50%	79.9%	68.6%	76.3%

Az „E” és az „O” hanggal lehetett elérni a legnagyobb pontosságot, mindkét esetben 83.4%-ot. A legjobb eredményt „E” hangra akkor kaptam, amikor a bemeneti vektorban a jitter ddp és shimmer dda jellemzők átlaga és szórása volt. Az „O” hang esetén még szükség volt a zz arány átlagára a hasonló eredmény elérésének érdekében. Az is látható, hogy a zz és zzm arány alkalmazása az SVM bemeneti vektoriban nem hozott szignifikáns javulást az osztályozásban.

7.3 Osztályozási kísérletek a Babel adatbázissal kiegészített adathalmazon

Miután kiértékeltem, hogy a jelenleg rendelkezésemre álló felvételeken mekkora pontosságú elkülönítést lehet elérni, nekiláttam a Babellel kiegészített adathalmaz vizsgálatának (110 kóros és 93 egészséges minta). A vizsgálat magánhangzók és a betanításhoz használt vektorok ugyanazok voltak, mint az előző esetben. Ezen tesztek eredményeit mutatja az alábbi táblázat:

11. táblázat. Babellel kiegészített adathalmazon elért elkülönítési pontosság

Figyelembe vett értékek	„E”	„O”	„u”	„i”
jitter ddp, shimmer dda átlaga és szórása	**85.2%**	82.8%	70.4%	81.3%
jitter ddp, shimmer dda átlaga és szórása, zz átlaga	83.3%	84.7%	73.4%	78.8%
jitter ddp, shimmer dda átlaga és szórása, zzm átlaga	82.3%	80.8%	70.9%	80.3%
jitter ddp, shimmer dda átlaga és szórása, zz és zzm átlaga	80.8%	81.8%	73.9%	80.8%

A saját adatbázison történt tesztekhez képest minden esetben javulás következett be. Ez egybevág a statisztikai analízis során kapott eredményekkel. Tehát a továbbiakban is érdemes a Babelből kiválasztott mintákat használni, hogy az egészséges minták számát megnöveljük.

Akárcsak a saját adatbázison történt osztályozások esetén, itt is az „E” és az „O” hang esetén kaptam a legjobb eredményeket. Ez azért lehetséges, mert a vizsgált hanganyagban kevesebb az „i” és „u” magánhangzók darabszáma (ld. 8. táblázat).

A legjobb elkülönítési eredmény 85.2%, amit „E” hanggal lehetett elérni, ha a bemeneti vektorokban a jitter ddp és shimmer dda jellemzők átlagát és szórását használtam fel. Itt is jól látszik, hogy a zz és zzm arányok figyelembe vétele nem okozott szignifikáns javulást.

7.4 A statisztikai analízis és az osztályozás eredményeinek összehasonlítása

A statisztikai analízis kimutatta, hogy „E” és „O” hang esetén mindkét fajta jitter, shimmer, valamint a HNR értékek mind alkalmasak lehetnek az egészséges és kóros minták szétválasztására. A lokális jitter és shimmer értékeket nem használtam fel, mert a korábbi vizsgálataink alapján elég az alaphangra és az amplitúdóra egyetlen ingadozás értéket felhasználni. [23]

Megvizsgáltam, hogy mi történik akkor, ha a HNR értéket is felhasználom. Az osztályozáshoz a Babellel kiegészített mintahalmazt használtam, hiszen erre jobb eredményeket kaptam korábbi tesztjeim során. A tanító vektorokba „E” és „O” hang esetén is a következő jellemzőket ültettem: jitter ddp átlaga és szórása, shimmer dda átlaga és szórása, HNR átlaga. Az eredmények az alábbi táblázatban találhatóak.

12. táblázat. Statisztikai analízisen alapuló osztályozási eredmények

Figyelembe vett értékek	„E”	„O”
jitter ddp, shimmer dda átlaga és szórása	**85.2%**	82.8%
jitter ddp, shimmer dda átlaga és szórása, HNR átlaga	83.3%	84.7%

A HNR paraméter figyelembe vétele az „E” hang esetén kisebb romlást, míg „O” hang esetén kisebb javulást eredményezett, de egyik esetben sem jelentős a változás. Ez némileg eltér az analízis alapján várt eredményektől, ugyanis a HNR paraméter alkalmazásával nem kellene az osztályozási pontosságának romlania. Azonban nem szabad elfeledkeznünk arról, hogy az osztályozó bemeneti vektorainak összeállításakor személyenkénti statisztikákat számítunk. Ez a módszer önmagában is okozhat ilyen eltéréseket az analízis és az osztályozás eredményei között.

7.5 Tesztelés összegzése

A rendelkezésemre álló hangfelvételekkel osztályozási feladatokat végeztem el, ahol a két elkülönítendő csoport az egészséges és kóros minták voltak. A teszteléshez SVM osztályozót használtam. Külön végeztem teszteket a saját adatbázisunkon, valamit a Babel általános célú beszédadatbázis hangfájljaival kiegészített mintahalmazon. A vizsgálatok egyértelműen kimutatták, hogy jobb eredmények érhetőek el akkor, ha a Babeles mintákat hozzávesszük az egészséges mintahalmazhoz. Ekkor 85.2%-os elkülönítési pontosságot tudtam elérni.

A statisztikai analízis eredményei összhangban vannak az osztályozási eredményekkel, és az apróbb eltérések magyarázhatóak az osztályozó bemenetei vektorainak összeállításakor történő személyenkénti statisztikaszámításokkal.

A továbbiakban ajánlott a Babellel kiegészített adathalmaz, valamint az „E” vagy „O” hangok használata, a bemeneti vektorba pedig a shimmer dda és jitter ddp értékek szórását és átlagát érdemes tenni.

Végül összehasonlításképpen, az egy évvel ezelőtt készített szakdolgozatomban 41 kóros és 32 egészséges mintával tudtam osztályozási kísérleteket végezni. Ekkor jitter ddp és shimmer dda paramétereket esetén 65% és 78% között változtak a pontosságok. Tehát jól

látszik, hogy az adathalmaz megnövelésével az osztályozó egyre pontosabb elkülönítésre képes. Ennek egyértelmű oka, hogy sokféle betegségtípus mintái találhatóak az adatbázisban, így a mintaszám növelésével egyre jobban reprezentáltak az különböző betegségtípusok is. A jövőben, szándékunkban áll a hangadatbázist tovább fejleszteni, ezért az eredményekben további javulás várható.

8 Összefoglalás

A dolgozat a kóros hangminták statisztikai analízisét és osztályozási lehetőségeit tárgyalja. Először egy kóros és egészséges mintákat tartalmazó beszédadatbázist kellett összeállítani, amelyben folyamatos beszéd található. A hangmintákat fonéma szinten kellett szegmentálni és SAMPA karakterekkel felcímkézni. A beszédadatbázis gyűjtése három évvel ezelőtt kezdődött el. Jelenleg 327 hangfájl áll rendelkezésünkre, melyből 169-et választottunk ki a kitűzött feladat végrehajtásához.

A statisztikai analízis szépen mutatja, hogy az általunk használt lényegkiemelési eljárások alapján lehetséges egy olyan eljárás kidolgozása, amelynek segítségével a hangterápia során lehetőség adódna a betegek javulásának automatikus mérése.

A statisztikai analízis eredményei és az LS-SVM típusú automatikus osztályozás, tesztelés eredményei szinkronban vannak. A statisztikai analízis és az osztályozási kísérletek egyaránt kimutatták, hogy a saját adatbázisunk egészséges hangmintáit célszerű kiegészíteni a Babelből kiválasztott egészséges felvételekkel, ugyanis ekkor jelentős javulás érhető el az automatikus osztályozásban. Ezen felül mindkét vizsgálat kimutatta, hogy a legnagyobb számban reprezentált „E” és „O” hangokat érdemes használni a kóros és egészséges minták elkülönítésére. A két vizsgálat eredményei alapján az „E” és „O” magánhangzó shimmer dda és jitter ddp értékeket ajánlott használni az osztályozáshoz, valamint érdemes lesz figyelemmel követni a HNR paraméterrel kiegészített felismerések pontosságát is.

Az osztályozáshoz használt optimális bemenő vektorral - amelybe a jitter ddp és shimmer dda értékek átlaga és szórása lett beültetve – 85.2%-os felismerési biztonság érhető el, annak ellenére, hogy a vizsgált betegségek típusa igen széles skálán mozgott (ld. 2. táblázat). Az automatikus osztályozás és tesztelés kimutatta, hogy a korábbi elkülönítési eredményekhez képest sikerült 10%-os javulást elérni.

Jövőbeni terveink között szerepel az adatbázis további növelése annak érdekében, hogy az osztályozási biztonság tovább növekedjen. Továbbá, az adatbázis növelésével elérhető, hogy az egyes betegségtípusokból is megfelelően nagyszámú minta álljon rendelkezésünkre, és ekkor ezzel a megnövelt adathalmazzal már lehetőségünk lesz a betegségek egymás közötti elkülöníthetőségének vizsgálatára is.

Irodalomjegyzék

[1] dr. Mészáros Krisztina, Bánó Zsuzsanna: A funkcionális dysphonia kezelésének értékelése az RBH-szisztéma segítségével, Fül-, Orr-, Gégegyógyászati Folyóirat 2005. 51. évfolyam 4. szám, 233-234. oldal

[2] Imre Viktor: Hangképzés zavarainak akusztikai vizsgálata, az egészséges és kóros minták automatikus elkülönítése, Budapest: BME szakdolgozat, 2010

[3] Olaszy Gábor: A beszédképzés folyamata In: Németh Géza, Olaszy Gábor (szerk.) A magyar beszéd: Beszédkutatás, beszédtechnológia, beszédinformációs rendszerek. Budapest: Akadémiai Kiadó, 2010. pp. 19-26.

[4] Jangling Wang, Cheolwoo Jo(2006). Performacne of Gaussian Mixture Models as a Classifier for Pathological Voice, Proceedings of the 11th Australian Conference on Speech Science & Technology, 2006

[5] Maria Markaki, Yannis Stylianou: Using Modulation Spectra for Voice Pathology Detection and Classification

[6] M. Sarria-Paja, G. Daza-Santacoloma, J.I. Godino-Llorente, G. Casellanos-Dominquez, N. Sáenz-Lechón: Selection in Pathological Voice Classification Using Dinamyc of Component Analysis, 2006

[7] Ji-Yeoun Lee, Sangbae Jeong, Minsoo Hahn Pathological Voice Detection Using Efficient Combination of Heterogeneous Features, IEICE Trans. Inf. & Syst., 2008, Vol.E91-D, No.2

[8] Tao Li et al. : Classification of Pathological Voice including Severely Noisy Cases, Interspeech 2004

[9] Behnaz Ghoraani and Sridhar Krishnan : A Joint Time-Frequency and Matrix Decomposition Feature Extraction Methodology for Pathological Voice Classification, EURASIP Journal on Advances in Signal Processing, Volume 2009, Article ID 928974

[10] Náthalee Cavalcanti et al. Comparative Analysis between Wavelets for the Identification of Pathalogical Voices, CIARP 2010, LNCS 6419

[11] Gastón Schlotthauer, María E. Torres and Hugo L. Rufiner: Pathological Voice Analysis and Classification Based on Empirical Mode Decomposition, Development of Multimodal Interfaces: Active Listening and Synchrony, 2010, Vol. 5967, pp. 364-381,

[12] Xiang Wang, Jianping Zhang, Yonghony Yan: Automatic Detection of Pathological Voices Using GMM-SVM Method

[13] Xiang Wang, Jianping Zhang, Yonghony Yan: Automatic Detection of Pathological Voices Using GMM-MLLR Approach

[14] Everthon S. Fonseca, José C. Pereira: Normal Versus Pathological Voice Signals, IEEE Engineering in Medicine and Biology Magazine, 2009, September/october

[15] Maria Markaki, Yannis Stylianou: Voice Pathology Detection and Discrimination based on Modulation Spectral Features, IEEE 2010

[16] Barlangi Renáta: Hangképzési rendellenességek osztályozási lehetőségei akusztikai paraméterek alapján, Budapest: BME szakdolgozat, 2010

[17] Zigri Gyula, Tóth László, Kocsor András, Sejtes Györgyi: Az automata és kézi szegmentálás ejtésvariációk okozta problémái, http://www.inf.u-szeged.hu/~kocsor/publications/Papers/2004/Conf-2004-MSZNY-TL/Web/ZTK04.pdf

[18] Vicsi Klára: A beszéd fizikai jellemzése. In: Németh Géza, Olaszy Gábor (szerk.) A magyar beszéd: Beszédkutatás, beszédtechnológia, beszédinformációs rendszerek.Budapest: Akadémiai Kiadó, 2010. pp. 38-56.

[19] Tóth Szabolcs Levente: Beszédalapú szolgáltatások silabusz, 2007. 23-24. oldal

[20] Horváth Gábor, Altrichter Márta, Pataki Béla, Strausz György, Takács Gábor, Valyon József: Neurális hálózatok, Hungarian Edition, Budapest: Panem Könyvkiadó Kft., 2006

[21] Paul Boersma: Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound, Proceedings 17, 1993, 97-110.

[22] Boersma, Paul: Praat, a system for doing phonetics by computer. Glot International 2001 5:9/10, 341-345.

[23] Vicsi Klára, Imre Viktor, and Mészáros Krisztina: Voice Disorder Detection on the Basis of Continuous Speech, IFMBE 2011 Proceedings 37, p. 86 ff.