

Betegségek automatikus szétválasztása időben eltolt akusztikai jellemzők korrelációs struktúrája alapján

Sztahó Dávid, Kiss Gábor, Tulics Miklós Gábor, Vicsi Klára

Budapesti Műszaki és Gazdaságtudományi Egyetem,
Távközlési és Médiainformatikai Tanszék
{sztaho, kiss, tulics, vicsi}@tmit.bme.hu

Kivonat: Egyes betegségtípusok különböző módon befolyásolhatják beszédkép-zésünk összetett mechanizmusait, patológiás beszédet eredményezve. Biomarkerek kinyerése a beszédből megbízható jelzői lehetnek a különböző betegségtípusoknak. A cikk célja egészséges és különböző betegségtípusokban szenvedő bemondók beszédmintáinak különválasztása. A vizsgált betegségtípusok a következők: depresszió, Parkinson-kór, hangképző szervek morfológiai elváltozása, a funkcionális diszfónia és a rekurrens paresis. Az osztályozó bemenetére formáns-frekvenciák (F1, F2, F3), a mel-szűrő sáv energia értékei, a mel-frekvencia kepsztrális együtthatók (MFCCs), az alaphfrekvencia (F0) és az intenzitás időben eltolt értékeinek korrelációs mátrixaiból származtatott értékei kerültek. Szupport vektor gépet, valamint k-legközelebbi szomszéd osztályozási eljárásokat használtunk az eredmények összehasonlítására. Hatosztályos osztályozás esetében a legjobb osztályozási pontosság 54.8%-nak adódott, míg négyosztályos esetben 77.6%. Az elért eredmények alapján kijelenthető, hogy egy beszédalapú rendszer létrehozható, amely segít a klinikai személyzetnek a korai diagnózis felállításában.

1 Bevezetés

A biomarkerek alkalmazása egyre népszerűbb, hiszen mérhető információt biztosítanak egy betegség súlyosságára vagy jelenlétére. A beszéd egyike azon biomarkereknek, amelyek számos betegséget jelezhetnek. Ez olcsó, nem invazív és hatékony módszerek fejlesztésére ad lehetőséget, amely segítheti a szakemberek munkáját.

A diszfónia a hangképzés komplex zavarát jelenti. Olyan patológiás állapot, melynek hátterében vagy a hangképző szerv organikus megbetegedése vagy idegrendszeri szabályozási zavar áll. A diszfónia a normálistól (euphonia) eltérő hangszínt, intenzitást, dallamot, hangmagasságot és a hangképző szerv csökkent terhelhetőségét eredményezi. A diszfóniás hang rendszerint rekedt, levegős, fátyolos [1][2]. A diszfóniát rendszerint két csoportra bontják. Az első akkor fordul elő, amikor az orvos hangbeli problémát észlel fiziológiai elváltozás hiányában, amelyet funkcionális diszfónia (FD - functional dysphonia) néven említenek, a második eset, amikor a hangproblémát a beszédképzés egyik alrendszerének fiziológiai torzulása kíséri, amelyet a vokális szervek morfológiai változásaként (MA - morphological alteration) illetnek. Az olyan betegségek, mint a hangszalagszomszék, a polipok, a gastrooesophagealis reflux betegség

(GERD), a ciszta és az egy vagy kétoldali hangszalagbénulás (RP - recurrent paresis, rekurrens paresis) mind a strukturális organikus rendellenességekbe sorolhatók, míg olyan betegségek, mint a stroke, Parkinson-kór (PD - Parkinson's disease) vagy sclerosis multiplex a neurológiai hangrendellenességek csoportjába sorolhatók.

A depresszió egy pszichiátriai betegség. A betegséget elsősorban a stressz vagy a kudarc okozhatja, amelynek érzelmi, kognitív, testi és motivációs tünetei lehetnek. A depresszió felismerési rátája alacsony, a páciensek emiatt nem kapnak megfelelő kezelést vagy félrekezelik őket. Azt jóslják, hogy 2020-ra a mentális fogyatékoság második legszignifikánsabb okozója lesz [6][7]. A beszéd a depresszió észlelésének jó objektív markere lehet, amit számos kutatás is alátámaszt [8][9][10][11][21].

A Parkinson-kór (PD) az egyik leggyakoribb neurológiai rendellenesség. A Parkinson-kórban szenvedő betegek hangjainak jellemzői közé tartozik a pontatlan és koordinálatlan artikuláció, csökkent hangosság, fokozott hangremegés, változó beszédsebesség és lélegzetvesztés, levegős és érdes hangminőség [12][13][14][15][16][19].

Az eddigi tanulmányok többnyire kétosztályos osztályozással foglalkoztak egészséges és patológiás beszéd szétválasztására. Korábbi munkáinkban kétosztályos osztályozási rendszereket fejlesztettünk ki, amely az egészséges beszédet a diszfóniásoktól [3], depresszióban szenvedő betegek hangjaitól [8], valamint Parkinson-kórban szenvedő betegek beszédétől [13] különböztetett meg. A gyakorlatban mindezen betegségek előfordulhatnak a páciensek körében. A jelenlegi kutatásban több (4 vagy 6) különböző betegség típusok szétválasztására fókuszálunk, többosztályos osztályozási módszer alkalmazásával. A vizsgált betegségtípusok a következők: depresszió, Parkinson-kór, vokális szervek morfológiai változása, funkcionális diszfónia és rekurrens paresis. Olyan akusztikai jellemzők, mint a jitter, shimmer, HNR (Harmonics-to-Noise Ratio) hasznosak az egészséges és diszfóniás hangok automatikus osztályozásában, folyamatos beszéd esetén [3][4][5].

Hipotézisünk, hogy ezek a betegségek befolyásolják a formánsfrekvenciákat (F1, F2, F3), a mel-szűrő sáv energia értékei, a mel-frekvencia kepsztrális együtthatók (MFCCs), az alapfrekvencia (F0) és az intenzitás időben eltolt értékeinek korrelációs mátrixait. (Korrelációs struktúra értékeket kétosztályos osztályozásra korábban is használtak [17][18][20][22].)

2 Adatbázisok

A kutatásban összesen négy adatbázist használtunk: hármat minden egyes betegségtípusra (a fonációs rendellenességek egy adatbázisban szerepelnek külön kategóriaként), valamint egy egészséges kontroll beszédadatbázist. Minden páciens Aiszóposz meséjét, „Az északi szél és a nap”-ot olvasta fel. Ezen népmese gyakran használt a foniatríai kutatásokban, a szöveganyagát úgy szerkesztették meg, hogy az adott nyelvben előforduló minden beszédhang, valamint a leggyakoribb hangkapcsolatok szerepeljen benne. Számos nyelvre elkészült ez a szöveg, köztük a jelen esetben is használt magyarra. A felvételek átlagosan 41 másodperc hosszúak voltak. Minden bemondó beleegezett a beszédének rögzítésébe, egy beleegező nyilatkozatot aláírva. Az adatbá-

zisok felvételeinek számát és leíró statisztikáit az 1. táblázatban foglaltuk össze. A felvételek minden esetben csendes orvosi rendelőben készültek, USB-s hangkártya segítségével.

2.1 Fonációs rendellenességek beszédadatbázisa (Phonation disorder Speech Database, PhoDb)

A felvételek az Országos Onkológiai Intézetben, foniáter szakorvos rendelésén lettek rögzítve a páciensek belegegyezésével. A szakrendelésre általában különböző hangpanaszokkal érkeznek a betegek. A beszédadatbázisban lévő betegségek a következők: morfológiai elváltozás (MA - morphological alteration), mint a hangképző szervrendszer különböző pontjain előforduló tumorok, gastroesophageal reflux (GERD), krónikus gégegyulladás, bulbar paresis (agyidegyulladás), amiotrófiás laterálszklerózis (ALS), leukoplakia, stb.); hangszalagbénulás (RP - recurrens paresis); funkcionális diszfónia (FD). A beszéd minőségét a diagnózist felállító orvos határozta meg az RBH-skála alapján [23]. A négy-fokozatú auditív rekedtségi skálán a 0 a normál hangminőségnek, míg a 3 a súlyos rekedtségnek felel meg. 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 jellemzik. A felvételek Monacor ECM-100 közel beszélő mikrofonnak készültek.

2.2 Depressziós beszédadatbázis (Depressed Speech Database, DSDb)

A depressziós (DE) adatbázis magyar anyanyelvű depresszióban szenvedő hangfelvételek gyűjteménye. A hangfelvételek a Semmelweis Egyetem Pszichiátriai és Pszichoterápiás Klinikával együttműködésben készültek. Az adatbázis az enyhe depressziótól a súlyos depresszióig terjedő páciensek hangfelvételeit tartalmazza, akiket neurológus szakember nem diagnosztizált más neurológiai betegséggel. A depresszió mérésére és a felvételek osztályozására a Beck Depression Inventory II (BDI) skálát alkalmaztuk [24]. A felvételek Audio-Technika ATR3350 csipetűs mikrofonnak készültek.

2.3 Parkinson-kór beszédadatbázis (Parkinson's Speech Database, PSDb)

Az adatbázis magyar anyanyelvű, Parkinson-kórban szenvedő páciensek beszédének gyűjteménye. A beszédmintákat két budapesti egészségügyi intézetben gyűjtöttük: a Virányos Klinikán és a Semmelweis Egyetemen. A Parkinson-kór súlyosságát a Hoehn & Yahr skála (H-Y) adja meg [25]. A felvételek Audio-Technika ATR3350 csipetűs mikrofonnak készültek.

2.4 Egészséges kontroll csoport (Healthy Control, HC)

Az egészséges kontroll csoport alanyai nem szenvedtek ismert betegségben és semmilyen orvosi kezelés alatt nem álltak. A felvételek ugyanannak a szövegnek a felolva-

sását tartalmazzák, mint a patológiás adatbázisok esetén, valamint a rögzítési körülmények is hasonlóak voltak. Az adatbázis 190 személy hangját tartalmazza: 85 férfi és 105 nő bemondóét. A felvételek Audio-Technika ATR3350 csiptetős mikrofonnak készültek.

1. Táblázat: Adatbázisok kor és betegség súlyosság szerinti leíró statisztikája

Adatbázis	Súlyossági mérték	Nem	Felvételek száma	Súlyosság	Kor
PhoDb - MA	RBH (0-3)	férfi	52	2.17(±0.88)	55.4(±12.8)
		nő	70	1.83(±0.82)	48.8(±15.3)
PhoDb - FD	RBH (0-3)	férfi	20	1.45(±0.69)	56.2(±14.5)
		nő	48	1.31(±0.59)	53.1(±17.3)
PhoDb - RP	RBH (0-3)	férfi	22	2.50(±0.80)	50.2(±15.4)
		nő	51	1.86(±0.83)	58.2(±10.6)
DSDb	BDI (0-61)	férfi	20	26.6(±8.9)	44.1(±14.3)
		nő	35	28.2(±10.2)	43.4(±13.5)
PSDb	H-Y (0-5)	férfi	40	2.74(±1.05)	64(±9.5)
		nő	36	2.74(±1.10)	65.4(±9.4)
HC	-	férfi	85	-	44.7(±18.7)
		nő	105	-	47.7(±13.8)

3. Módszerek

3.1 Akusztikai jellemzők

Számos akusztikai jellemzőt választottunk ki, amelyek patológiás esetekben követik a hangképzés változását. Ezeket az akusztikai jellemzőket alacsony szintű leíróknak neveztük, amelyekből a következő jellemző csoportokat alkottuk: formáns frekvenciák (F1, F2, F3), mel-sávós energia-értékek (27 sáv 60 Hz-től 8 kHz-ig), mel-frekvenciás kepsztrális együtthatók (MFCC-k, amelyek 12 koefficienssel rendelkeznek), valamint az alapfrekvenciát (F0) és intenzitást közösen tartalmazó csoportot. Minden akusztikai jellemzőt Praat [26] szoftverrel számítottunk 10 ms-os időközzel.

3.2 Korrelációs struktúra jellemzők

A korreláció és kovariancia struktúrák (mátrixok) számítását és a belőlük származtatott jellemzőket a Williamson és társai eljárása [17][18] alapján végeztük. A korábban említett alacsony szintű leírójellemzőkből képezett idősorokat, mint csatorna használtuk (a [17] és [18] cikkek jelölései szerint) és a következő csoportokat hoztunk létre: „formánsok” (F1, F2, F3), „melsávok” (27 mel-sávós energia érték), „mfcc” (12 mfcc együttható), „enf0” (intenzitás és f0).

A korreláció és kovariancia struktúrákat minden egyes beszédmintára kiszámítottuk. Minden ilyen struktúra egy $(k*n) \times (k*n)$ mátrixot jelent, ahol a k a csatornák száma (például $k = 3$ a „formánsok” csoport esetében), míg n a késleltetések száma. Ez a mátrix felfogható úgyis, mint k^2 darab almátrix, amik elemei $n \times n$ méretű mátrixok. Minden ilyen almátrix adott két csatorna korrelációs vagy kovariancia együtthatóit tartalmazza. Adott két csatorna egyenként n féle különböző késleltetése mellett, ahol a csatorna az $i=1,2,\dots,k$ és $j=1,2,\dots,k$ lehetséges értékek között változik, míg az adott almátrixon belül pedig a késleltetés az egyik illetve a másik csatorna esetében $p=0,1,\dots,n-1$ illetve $q=0,1,\dots,n-1$ értékek között változik. A struktúra ilyenfajta felépítése következtében az átlón lévő almátrixokban az egyes csatornák autokorrelációs együtthatói találhatóak meg különböző késleltetések mellett. A mátrixok az átlóra szimmetrikusak, illetve a sajátértékeik pozitívak. A mátrixokat 4 különböző időskála esetén is kiszámítottuk, ahol az időskála értéke határozta meg, hogy az adott késleltetés mekkora időeltolást jelent a csatorna idősorában. A korreláció és kovariancia struktúra részletesebb leírása megtalálható a [22] irodalomban, ezek korábbi beszédjelre alkalmazott gyakorlati megvalósításai pedig [17][18] irodalmakban.

Az időkésleltetések száma 10 volt ($n = 10$) a „melsávok”, „mfcc” és „enf0” csoportok esetében, míg ez az érték 30 volt ($n = 30$) volt a „formánsok” csoport esetében. Ahogy korábban említettük, 4 különböző időskálát használtunk, amik értéke rendre 1,2,4 és 8 voltak minden csoport esetében, ami időben 10 ms, 20 ms, 40 ms és 80 ms időeltolást jelentett $n = 1$ időkésleltetés mellett. Az 1. ábrán példaként láthatóak az átlagos korreláció mátrixok a 4 különböző vizsgált osztály esetében, a „formánsok” csoport, 1-es időskála használata mellett. Összesen 16 korreláció és 16 kovariancia mátrix lett kiszámítva minden beszédminta esetében.

A korreláció és kovariancia struktúrákból a következő származtatott jellemzőket számítottuk ki és használtunk fel minden időskála esetében: korreláció mátrix sajátértékei, a korreláció mátrix sajátértékeinek entrópia értéke és a kovariancia mátrix sajátértékeinek négyzetes középben vett átlagát. Ezek a jellemzők voltak a bemenetei az osztályozásnak.

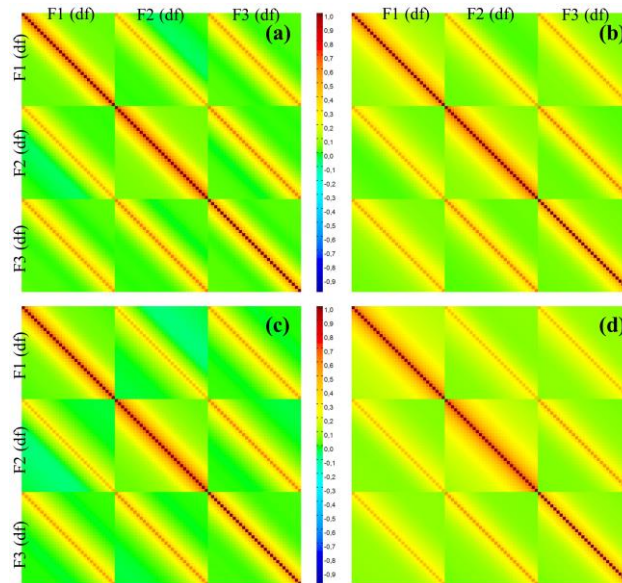
3.3 Osztályozás

A RapidMiner Studio 7.5 [27] szoftvert használtuk a gépi tanulási kísérletekhez. Az osztályozási módszerek paraméterei az adott eljárás során bevett alapértelmezett értékei voltak. A kutatás során a k -legközelebbi szomszédok (k -NN, k paramétert 9-re állítva) és szupport vektor gépeket használtunk, c -SVC lineáris ($C = 1$ paraméterrel) és radiális bázis alapú kernelfüggvénnyel (ahol C -nek az akusztikai jellemzők számát választottuk és $\gamma = \frac{1}{\text{akusztikai paraméterek száma}}$). Minden vizsgálatot 10-szeres keresztvalidációval végeztünk, ahol az egyes osztályok eloszlása egyenletes volt.

Először hat csoportot külön kíséreltünk meg osztályozni: HC, DE, PD és az MA, FD és RP osztályokat a Fonációs rendellenességek beszédadatbázisából. A Fonációs rendellenességek beszédadatbázis három csoportját azért is kezeltük külön, mert korábbi munkánkban azt találtuk, hogy az MA és RP csoportok elkülöníthetők lehetnek egymástól [5].

Ezek után a Fonációs rendellenességek beszédatbázisa három csoportját egybevonva (Fonációs rendellenességek, továbbá FR) négy osztályos osztályozást végeztünk az HC, DE, PD és FR csoportok között.

Optimális akusztikai jellemzők megtalálása érdekében Forward Selection jellemző-kiválasztó eljárást használtunk. Költségfüggvényként pontosságot (accuracy) választottunk, a maximálisan kiválasztott jellemzők számára 20-at választottunk.



1. ábra. Formáns frekvencia csoport korrelációs mátrixa, 1-es skálát használva, (a)-egészséges, (b)-depresszió, (c)-morfológiai elváltozás, (d)-Parkinson-kór

4. Eredmények

A hat, illetve a négyosztályos osztályozás pontosság ($\frac{\text{helyesen felismert minták száma}}{\text{összes minta száma}}$) eredményeit a 2. táblázatban foglaltuk össze. A táblázatban megtalálhatók minden akusztikai jellemző csoporttal külön végzett, valamint együttesen használva kapott eredmények. Különböző akusztikai jellemzők csoportjai különböző elkülönítési teljesítménnyel rendelkeznek. Általánosan az 'enf0' csoport teljesített a legrosszabbul, ebből az következik, hogy az intenzitás és az alaphang auto- és keresztkorrelációs értékei nem rendelkeznek magas elkülönítési képességgel. A további három jellemző csoport mind magasabb osztályozási pontosság értéket ért el, ezek közül is a 'melsávok' csoport teljesített a legjobban.

Hat osztályos osztályozás esetében az MA, FD és RP osztályok esetében sok esetben fordult elő az egymásra tévesztés. A minden akusztikai jellemzőt felhasználó SVM-RBF osztályozás tévesztési mátrixát a 3. táblázat foglalja össze. Az egymásra tévesztés jelensége miatt vontuk össze egy osztállyá az MA, FD és RP osztályokat, így a négy

2. Táblázat: Osztályozási eredmények (pontosság, accuracy) 6, illetve 4 osztályos esetben

Jellemző csoport	skála	k-nn	svm-linear	svm-rbf
enfő	1	37,85 / 51,94	41,20 / 54,05	41,73 / 56,51
	2	38,73 / 54,23	42,43 / 54,05	41,20 / 54,93
	4	35,92 / 49,65	40,32 / 53,00	42,08 / 57,75
	8	32,92 / 46,30	36,17 / 48,06	35,21 / 47,71
	összes	38,73 / 53,87	34,51 / 55,89	35,21 / 56,34
formánsok	1	38,03 / 55,11	46,13 / 64,26	44,72 / 65,49
	2	36,27 / 53,87	43,31 / 61,80	43,31 / 62,15
	4	37,50 / 57,75	45,95 / 62,68	43,49 / 63,56
	8	38,38 / 57,39	47,71 / 63,56	45,25 / 65,85
	összes	38,38 / 58,10	42,78 / 64,96	42,08 / 64,61
melsávok	1	35,21 / 51,58	44,54 / 60,56	45,95 / 63,91
	2	38,56 / 51,76	48,06 / 63,73	49,12 / 69,54
	4	39,44 / 55,28	49,82 / 65,32	47,54 / 66,73
	8	42,43 / 52,28	50,53 / 67,08	49,47 / 70,25
	összes	41,55 / 56,34	51,06 / 72,36	50,35 / 74,12
mfcc	0	36,97 / 50,53	41,55 / 57,22	40,32 / 57,57
	1	36,97 / 53,87	42,78 / 60,21	42,25 / 63,03
	2	39,44 / 54,93	41,78 / 59,15	39,61 / 57,75
	3	40,14 / 59,68	41,20 / 64,79	41,55 / 65,49
	összes	42,08 / 60,74	43,84 / 68,66	44,89 / 69,37
Összes jellemző	0	39,26 / 57,22	45,42 / 72,01	45,42 / 71,48
	1	44,54 / 59,68	50,00 / 74,30	46,30 / 74,47
	2	45,42 / 62,68	46,65 / 67,25	47,01 / 68,84
	3	46,65 / 63,03	48,94 / 75,00	47,01 / 73,42
	összes	47,54 / 63,20	48,77 / 76,23	48,42 / 77,64
Összes jellemző jellemző-kiválasztással	0	43,13 / 65,49	46,48 / 72,76	53,87 / 72,18
	1	44,89 / 61,27	54,93 / 72,40	52,64 / 72,36
	2	48,06 / 66,55	53,52 / 69,24	51,94 / 69,72
	3	48,77 / 68,31	52,46 / 72,15	52,64 / 71,83
	összes	51,41 / 71,30	53,32 / 76,17	54,75 / 77,59

osztályos osztályozás eredményeképpen 77,64%-os pontosságot értünk el SVM-RBF-et használva.

Általánosságban elmondható, hogy az összes időskála felhasználása javított az osztályozási eredményeken. A legmagasabb pontosságot akkor értük el, amikor az összes akusztikai jellemző csoport felhasználásra került. A jellemző-kiválasztásos kísérletek során a legnagyobb pontosság 54,75%-nak adódott hatosztályos esetben, valamint 77,64% négyosztályos esetben. A jellemző-kiválasztás növelte az osztályozási pontosságot k-NN esetben is. Említésre méltó, hogy jellemző-kiválasztással egy olyan egyszerű algoritmus, mint amilyen a k-NN, összemérhető eredményeket produkált egy sokkal komplexebb osztályozóval, mint amilyen a szupervektor gép.

3. Táblázat: Tévesztési mátrix minden jellemzőt felhasználva (összes skála) SVM-RBF esetén. A cella értékei százalékok.

Prediktált\Valós	HC	DE	PD	FD	MA	RP
HC	73.16	25.45	16.88	26.87	11.61	1.49
DE	8.42	56.36	9.09	1.49	1.79	0.00
PD	5.79	10.91	72.73	0.00	1.79	1.49
FD	5.26	0.00	0.00	32.84	17.86	8.96
MA	6.84	5.45	1.30	28.36	19.64	80.60
RP	0.53	1.82	0.00	10.45	47.32	7.46

5. Következtetések

Ebben a kutatásban kísérletet tettünk különböző típusú patológiás rendellenességek automatikus szétválasztására formánsfrekvenciák (F1, F2, F3), mel-sávós energia értékek, mel-frekvencia kepsztrális együtthatók (MFCC), az alapfrekvencia (F0) és az intenzitás időben eltoló értékeinek korrelációs mátrixai alapján. Többi eltolási skálát és különböző osztályozási eljárást használva a legjobb osztályozási pontosságként 77,64%-ot értünk el négyosztályos osztályozás esetében. Ez ígéretes eredménynek számít, hiszen az adatbázisokban szereplő hangfelvételek száma limitált. Ez az eredmény azt sugallja, hogy valóban vannak korrelációs különbségek a mért időtartománybeli jellemzőkben a négy vizsgált betegség típus esetén. Az eredmények alapján a korrelációs struktúrák integrálhatók egy automatikus komplex diagnosztikai rendszerbe.

A hatosztályos osztályozási kísérletekben a tévesztési mátrixok alapján elmondható, hogy a fonációs rendellenesség betegség típusokat (MA, FD és RP csoportokat) a rendszer sok esetben összekeverte. Ahhoz, hogy ezek a betegség típusok egymástól automatikusan megkülönböztethetők legyenek, további akusztikai jellemzőket kell bevonni.

Bibliográfia

1. Tulics, M.G., Kazinczi, F., Vicsi, K., "Statistical analysis of acoustical parameters in the voice of children with juvenile dysphonia," in: International Conference on Speech and Computer, Springer, 2016, pp. 667–674.
2. Ruotsalainen, J., Sellman, J., Lehto, L., Verbeek, J., "Systematic review of the treatment of functional dysphonia and prevention of voice disorders," *Otolaryngology-Head and Neck Surgery* 138, 2008, pp. 557–565.
3. Kazinczi, F., Mészáros, K., Vicsi, K., "Automatic detection of voice disorders," in: International Conference on Statistical Language and Speech Processing, Springer, 2015, pp. 143–152.
4. Grygiel J. and Strumillo P., "Application of Mel Cepstral Representation of Voice Recordings for Diagnosing Vocal Disorders," *Przegląd Elektrotechniczny (Electrical Review)*, 2012.

5. Tulics, M.G., and Vicsi, K., "Phonetic-class based correlation analysis for severity of dysphonia," in: *Cognitive Infocommunications (CogInfoCom)*, 2017 8th IEEE Conference on, IEEE, 2017, pp. 21-26.
6. Kessler, R.C., Bromet, E.J., "The epidemiology of depression across cultures," *Annual review of public health* 34, 2013, pp. 119–138.
7. Lépine, J.P., Briley, M., "The increasing burden of depression," *Neuropsychiatric disease and treatment* 7, 2011, pp 3.
8. Kiss, G., Vicsi, K., "Mono-and multi-lingual depression prediction based on speech processing," *International Journal of Speech Technology*, 2017, pp. 1–17.
9. Cummins, N., Scherer, S., Krajewski, J., Schnieder, S., Epps, J., Quatieri, T.F., "A review of depression and suicide risk assessment using speech analysis," *Speech Communication* 71, 2015, pp. 10–49.
10. Asgari, M., Shafran, I., "Improvements to harmonic model for extracting better speech features in clinical applications," *Computer Speech & Language* 47, 2018, pp. 298–313.
11. Valstar, M., Schuller, B., Smith, K., Eyben, F., Jiang, B., Bilakhia, S., Schnieder, S., Cowie, R., Pantic, M., "Avec 2013: the continuous audio/visual emotion and depression recognition challenge," in: *Proceedings of the 3rd ACM international workshop on Audio/visual emotion challenge*, ACM. 2013., pp. 3–10.
12. Sztahó D, Vicsi, K., "Estimating the severity of Parkinson's disease using voiced ratio and nonlinear parameters," in: Pavel Král, Carlos Martín-Vide, *Statistical Language and Speech Processing: 4th International Conference, SLSP 2016, Proceedings*. Springer International Publishing, 2016. pp. 96-107.
13. An, G., Brizan, D. G., Ma, M., Morales, M., Syed, A. R., & Rosenberg, A., "Automatic Recognition of Unified Parkinson's Disease Rating from Speech with Acoustic, i-Vector and Phonotactic Features," *Sixteenth Annual Conference of the International Speech Communication Association*. 2015.
14. Naranjo, L., Pérez, C.J., Campos-Roca, Y., Martín, J., "Addressing voice recording replications for parkinson's disease detection," *Expert Systems with Applications* 46, 2016, pp. 286–292.
15. Mekyska, J., Smekal, Z., Galaz, Z., Mzourek, Z., Rektorova, I., Faundez-Zanuy, M., López-de Ipiña, K., "Perceptual features as markers of parkinson's disease: the issue of clinical interpretability," in: *Recent Advances in Nonlinear Speech Processing*. Springer, 2016, pp. 83–91.
16. Pompili, A., Abad, A., Romano, P., Martins, I.P., Cardoso, R., Santos, H., Carvalho, J., Guimaraes, I., Ferreira, J.J., "Automatic detection of parkinson's disease: An experimental analysis of common speech production tasks used for diagnosis," in: *International Conference on Text, Speech, and Dialogue*, Springer, 2017, pp. 411–419.
17. J. R. Williamson, T. F. Quatieri, B. S. Helfer, G. Ciccarelli, and D. D. Mehta, "Vocal and facial biomarkers of depression based on motor incoordination and timing," in *Proceedings of the 4th ACM International Workshop on Audio/Visual Emotion Challenge (AVEC)*, 2014, pp. 65–72.
18. J. R. Williamson, T. F. Quatieri, B. S. Helfer, R. Horwitz, B. Yu, and D. D. Mehta, "Vocal biomarkers of depression based on motor incoordination," in *Proceedings of the 3rd ACM International Workshop on Audio/Visual Emotion Challenge*, 2013, pp. 41–48.
19. Williamson, James R., et al. "Segment-dependent dynamics in predicting Parkinson's disease." *Sixteenth Annual Conference of the International Speech Communication Association*. 2015.
20. B. Yu, T. F. Quatieri, J. W. Williamson, and J. Mundt, "Prediction of cognitive performance in an animal fluency task based on rate and articulatory markers," in *15th Annual Conference of the International Speech Communication Association*, September 9–13, Portland, Oregon, Proceedings, 2014.

21. B. S. Helfer, T. F. Quatieri, J. R. Williamson, L. Keyes, B. Evans, W. N. Greene, J. Palmer, and K. Heaton, “Articulatory dynamics and coordination in classifying cognitive change with preclinical mTBI,” in 15th Annual Conference of the International Speech Communication Association, September 9–13, Portland, Oregon, Proceedings, 2014.
22. J. R. Williamson, D. Bliss, D. W. Browne, and J. T. Narayanan, “Seizure prediction using EEG spatiotemporal correlation structure,” *Epilepsy and Behavior*, vol. 25, no. 2, 2012, pp. 230–238.
23. Wendler, J., Rauhut, A., Kruger, H., “Classification of voice qualities,” *Journal of Phonetics* 14, 1986, pp. 483–488.
24. Beck, A.T., Steer, R.A., Ball, R., Ranieri, W.F., “Comparison of beck depression inventories-ia and-ii in psychiatric outpatients,” *Journal of personality assessment* 67, 1996, pp. 588–597.
25. Hoehn, M.M., Yahr, M.D., “Parkinsonism onset, progression, and mortality,” *Neurology* 17, 1967, pp. 427–427.
26. Boersma, Paul & Weenink, David (2018). Praat: doing phonetics by computer [Computer program]. Version 6.0.39, retrieved 3 April 2018 from <http://www.praat.org/>
27. Hofmann, M. & Klinkenberg, R. “RapidMiner: Data Mining Use Cases and Business Analytics Applications”. 2013