UNIUNEA EUROPEANĂ
GUVERNUL ROMÂNIEI MINISTERUL MUNCII, FAMILIEI ŞI PROTECTIEI SOCIALE AMPOSDRU
Fondul Social European POSDRU 2007-2013
Instrumente Structurale 2007-2013
OIPOSDRU
UNIVERSITATEA TEHNICĂ "GHEORGHE ASACHI" DIN IAŞI

Universitatea Tehnică "Gheorghe Asachi" din Iaşi
Facultatea de Electronică, Telecomunicații şi Tehnologia Informației

Raport de cercetare II

CORPUSURILE ADNOTATE DE SEMNALE VOCALE DE UZ MEDICAL ŞI INSTRUMENTE AFERENTE

Doctorand:
Bioing. Alina Untu (Hulea)

Conducător ştiințific:
Prof.dr.ing. Horia Nicolai Teodorescu, m.c. A.R.

-Iaşi 2011-

Cuprins

1. Introducere

2. Corpusuri de semnale vocale şi instrumente aferente

2.1. Corpusuri comerciale existente pe site-ul LDC (Linguistic Data Consortium)

2.2. Metodologii utilizate la crearea de corpusuri de semnal vocal şi instrumente aferente

2.2.1. Metodologii pentru crearea de corpusuri destinate recunoaşterii vocale

2.2.2. Metodologii pentru crearea de corpusuri destinate sintezei de voce

2.2.3. Metodologie pentru crearea unui corpus de dialecte

2.2.4. Metodologie pentru crearea corpusului multilingv „PFSTAR”

2.2.5. Metodologie pentru crearea corpusului „CIAIR inCar”

3. Corpusuri de voci patologice şi instrumente aferente

3.1. Baza de date de voci patologice dezvoltată de MEEI (MEEI VDDb)

3.2. Metode de detecție de voci patologice care utilizează ca referință MEEI VDDb

3.3. Alte corpusuri de voci patologice şi instrumente aferente

4. Microcorpusuri şi metode de detecție a vocilor patologice

4.1. Aplicații în stomatologie

5. Cazul SRoL (Sunetele Limbii Române) şi extensie în limba franceză

5.1. Microcorpus de sunete gnatofonice în limba română

5.2. Microcorpus de sunete gnatofonice în limba franceză

6. Discuții şi concluzii

7. Direcții viitoare

1. Introducere

La ora actuală există o serie de consorţii care colaborează în vederea construirii unei baze de cunoştinţe şi instrumente de limbă care să fie disponibile cercetătărilor din domenii precum fonetică, acustică, informatică şi medicină. Câteva dintre cele mai importante sunt CLARIN, ELRA, ELDA, LDC şi TELRI.

CLARIN este o infrastuctură europeană care are ca scopuri principale: stocarea de corpusuri, instrumente şi resurse lexicale mono sau bilingve, pe domenii specifice sau generale; investigarea standardelor existente, adaptarea lor şi propunerea de sugestii pentru schimbări; crearea unei taxonomii şi investigarea formatului de codare a resurselor existente; investigarea caracteristicilor instrumentelor disponibile; stabilirea criteriului de evaluare a calităţii resurselor de limbă (http://www.clarin.eu/). Grupul CLARIN îşi propune să ofere servicii persistente sigure şi să furnizeze acces facil la resursele de procesare a limbajului.

ELRA este Asociaţia Europeană de Resurse pentru Limbaj incluzând ca parte operativă **ELDA** (Agenţia de evaluare şi distribuţie a resurselor de limbaj), care are rolul de a identifica, clasifica, colecta, valida şi produce resurse de limbaj. Se ocupă de asemenea de dezvoltarea ariei ştiinţifice din acest domeniu (http://www.elra.info/).

LDC (Linguistic Data Consortium) este un consorţiu alcătuit din universităţi, companii şi laboratoare de cercetare guvernamentale care crează, colectează şi distribuie baze de date de text şi voce, lexicoane şi alte resurse comerciale în scop de cercetare şi dezvoltare (http://www.ldc.upenn.edu/).

2. Corpusuri de semnale vocale şi instrumente aferente

2.1. Corpusuri comerciale existente pe site-ul LDC

Corpusul acustico-fonetic de vorbire continuă „**TIMIT**" este creat pentru a furniza înregistrări de voce pentru studii acustico-fonetice şi pentru dezvoltarea şi evaluarea sistemelor de recunoaştere automată. Este alcătuit din înregistrări realizate cu microfon provenite de la 630 vorbitori din opt dialecte majore de engleză americană, fiecare înregistrare conținând câte zece propoziții citite. Au fost colectate informații legate de vârstă, sex, etnie, nivel de educație, înălțime şi dialect corespunzătoare fiecărui vorbitor.

Corpusul TIMIT include ortografie aliniată în timp, transcripție la nivel de fonem şi cuvânt şi câte un fişier .wav pe 16 biți cu frecvența de eşantionare de 16 kHz pentru fiecare pronunție. A fost dezvoltat de către Massachusetts Institute of Technology (MIT), SRI International (SRI) şi Texas Instruments, Inc. (TI) şi este disponibil pe CD-ROM (J. S. Garofolo et al., 1986).

Corpusul „**Articulation Index Corpus**„ (J. Wright, 2005) a fost dezvoltat în scopul evaluării gradului de percepție corectă a silabelor în mediu zgomotos de către vorbitori. Alte aplicații sunt identificarea, modelarea şi procesarea de limbă şi modelarea pronunției. Corpusul conține înregistrări de limbă engleză la microfon, eşantionate cu 16 kHz în format pcm. Fiecare vorbitor a pronunțat silabe ale limbii engleze dintre care unele sunt cuvinte altele sunt silabe nonsens. Scopul a fost ca fiecare vorbitor să pronunțe un set de 2.000 de silabe comune tuturor vorbitorilor şi 20 de silabe unice fiecărui vorbitor. Corpusul este disponibil pe DVD contracost.

„**CSLU: Kids' Speech Version 1.1**" (K. Shobaki et al., 2007) dezvoltat de LDC este o colecție de înregistrări în limba engleză realizate pe vorbire spontană şi citită provenind de la 1100 copii de grădiniță şi până în 10 ani. Fiecare copil a citit aproximativ 60 dintr-o listă de 319 cuvinte, propoziții sau litere. Fiecare pronunție de vorbire spontană conține la început o recitare a alfabetului şi un monolog de un minut. Corpusul conține 1017 fişiere de 8-10 minute pentru fiecare vorbitor, digitizate pe 16 biți cu o frecvență de eşantionare de 16 kHz utilizând carduri audio Soundblaster 16 PnP. Pentru înregistrări au fost folosite căşti cu microfon. Acest corpus a fost dezvoltat pentru a facilita analiza caracteristicilor vocilor copiilor la diverse stadii ale vârstei şi la antrenarea şi evaluarea sistemenlor de recunoaştere utilizate în învațarea limbii şi alte sarcini care implică copii

incluzînd dezvoltarea limbii în cazul copiilor fără auz. Informații cu privire la vârstă, sex, limba vorbită şi condițiile fizice care influențează vorbirea au fost de asemenea incluse în corpus.

„**Emotional Prosody Speech and Transcripts**" (M. Liberman et al., 2002) este un corpus care conține înregistrări audio în limba engleză cu microfon şi transcripții corespunzătoare colectate pe o perioadă de opt luni între anii 2000-2001 pe două canale şi cu rată de eşantionare de 22050 Hz. Înregistrările aparțin unor actori profesionişti care au citit o serie de enunțuri neutre din punct de vedere semnatic, acoperind paisprezece categorii de emoții. Corpusul are aplicații în modelare de pronunție, prozodie şi recunoaştere vocală.

NTIMIT (Jankowski et al., 1990) este un alt corpus de limbă engleză cu aplicații în recunoaştere vocală care a fost dezvoltat de către Grupul NYNEX Science and Technology Speech Communication Group şi este disponibil pe DVD. Acesta a fost realizat pentru a furniza înregistrări telefonice asociate corpusului TIMIT. NTIMIT a fost creat prin transmitereea a 6300 înregistrări originale printr-un telefon pe diverse canale către rețeaua telefonică NYNEX care au fost redigitizate. Înregistrările au fost transmise prin zece arii de transport şi acces local dintre care pentru jumătate din acestea era necesară utilizarea purtătorilor de distanță mare. Formele de undă reînregistrate au fost aliniate în timp cu formele de undă originale ale corpusului TIMIT astfel încât transcripțiile TIMIT să poate fi utilizate concomitent cu cele NTIMIT.

Alte corpusuri existente în catalogul LDC care au fost create din înregistrări de conversații telefonice cu aplicații în recunoaştere vocală sunt:

- „**CALLHOME American English Lexicon (PRONLEX**)" (P. Kingsbury et al., 1997), conține 90,988 cuvinte selectate din texte aparținând jurnalului „Wall Street Journal" utilizate recent în corpusul ARPA de recunoaştere a vorbirii continue.
- „**CALLHOME German Lexicon**" (K. Karins et al., 1997), conține 318,807 cuvinte din care 315,503 sunt adaptate din lexiconul german CELEX produs de Centrul pentru informație lexicală, Max Planck Institute for Psycholinguistics in Nijmigen şi 3,304 sunt cuvinte adiționale provenite de la 80 de transcripții de antrenare şi 20 de dezvoltare a câte zece minute fiecare din cadrul corpusului LDC German CALLHOME. Lexiconul german conține câmpuri separate cu informații despre ortografie, morfologie, fonologie, accent, sursă şi frecvență pentru fiecare cuvânt.

- „**CALLHOME Spanish Lexicon**” (S. Garrett et al., 1996) conține 45,582 cuvinte în limba spaniolă şi câmpuri separate cu informații fonologice, morfologice şi despre frecvență pentru fiecare cuvânt.

În mod similar au fost dezvoltate corpusuri pentru limba japoneză (M. Kobayashi et al., 1996), coreană (N. Han et al., 2003), egipteană (Kilany et al., 2002) şi mandarină (S. Huang et al., 1996). Acestea sunt disponibile pe web contracost.

O altă categorie de corpusuri sunt cele create din înregistrări ale emisiunilor de radio, TV, care au ca scop detecția şi urmărirea de subiect (Topic Detection and Tracking (TDT)). Printre acestea se enumeră:

- „**TDT4 Multilingual Broadcast News Speech Corpus**” conține înregistrări radio cu o frecvență de eşantionare de 16 kHz, pe două canale, în limbile arabă, engleză şi mandarină şi este disponibil pe DVD contracost. A fost utilizat în anii 2002-2003 pentru detectarea şi urmărirea de subiect în ştiri. Adiacent la acest corpus au fost dezvoltate tehnici automate de găsire a materialelor cu subiect comun în şiruri de date provenind din ştiri difuzate la emisiuni radio. Tehnica de evaluare utilizată include segmentarea unor ştiri sursă în ştiri separate, urmărirea subiectelor cunoscute, detectarea subiectelor necunoscute, detectarea ştirilor inițiale cu subiect necunoscut şi detectarea de perechi de ştiri cu subiect comun (J. Kong şi D. Graff, 2005).
- „**TDT3 English Audio**” conține înregistrări audio ale emisiunilor de ştiri colectate zilnic de la şase surse de ştiri în engleză americană pe o perioadă de trei luni (1998). Acestea sunt înregistrări pe un singur canal de 30 – 60 minute care au fost eşantionate cu 16 kHz utilizând eşantioane pe 16 biți şi compresate cu algoritmul „shorten”. Acest corpus a fost creat ca suport pentru tehnologia TDT3 care include segmentare, detecție şi urmărire, detectarea primei ştiri şi a legăturii dintre ştiri (D. Graff, 2001a).
- „**TDT3 Mandarin Audio**” (D. Graff, 2001b) realizat după aceaşi metodologie cu „TDT3 English Audio” conține înregistrări de ştiri în limba mandarină preluate de pe postul de radio „Voice of America”.

O altă categorie de corpusuri sunt destinate recunoaşterii de voce în medii zgomotoase cum ar fi traficul aerian. Un astfel de corpus este „**Air Traffic Control**” (ATC0) disponibil în mai multe variante : „Air Traffic Control DFW” (J. J. Godfrey, 1994a), „Air Traffic Control BOS” (J. J. Godfrey, 1994b), „Air Traffic Control DCA” (J. J. Godfrey, 1994c), „Air Traffic Control Complete” (J. J. Godfrey, 1994d). ATC0 conține opt discuri de înregistrări de voce destinate utilizării în cercetare şi activităților de dezvoltare în

aria recunoaşterii vocale robuste, în domenii similare controlului traficului aerian (mai mulți vorbitori, canale cu zgomot, vocabular mic). Înregistrările se bazează pe comunicarea între diverse turnuri de control şi piloți. Fişierele audio sunt eşantionate cu 8 kHz pe 16 biți şi reprezintă monitorizarea continuă fără eliminarea zonelor cu linişte, pe o durată de la una la două ore. Sunt de asemenea disponibile fişiere care indică amplitudinea semnalului primit la intervale de 10 milisecunde. Setul complet conține 70 ore de transmisii colectate prin antene şi receptoare radio care sunt localizate în vecinătatea aeroporturilor.

Au fost create corpusuri de înregistrări audio în cadrul întâlnirilor în scopul analizei de discurs, extracției de metadate, identificare de vorbitor şi recunoaştere vocală. Dintre acestea se enumeră „**2004 Spring NIST Rich Transcription (RT-04S) Evaluation Data**" (J. Fiscus et al., 2007a) care conține material de test (înregistrări în cadrul întâlnirilor şi transcripții referință) utilizat în evaluarea RT-04S. Transcripția bogată este definită ca o fuziune între tehnologia speech-to-text şi tehnologii de extracție metadate creat pentru a furniza o bază pentru generarea celor mai utilizate transcripții în cadrul întâlnirilor dintre oameni. RT-04S include două tipuri de sarcini şi anume transcripție vorbire – text (STT) şi „diarization" (cine şi cînd vorbeşte) (SPKR). Sarcina STT include selectarea condițiilor în care se fac înregistrările şi condițiile de procesare. Se pot utiliza unul sau mai multe microfoane la distanță sau căşti individuale cu microfon, iar timpul de procesare poate fi nelimitat, mai mic sau egal cu 10, 20 de ori sau o dată din timpul real. În cadrul SPKR se poate utiliza unul sau mai multe microfoane la distanță, timpul de procesare este similar cu cel pentru STT şi trebuie luate în calcul condițiile de input şi anume se poate utiliza ca input doar voce sau voce plus o transcripție de referință. Aceiaşi autori au dezvoltat în mod similar şi un corpus de dezvoltare „**2004 Spring NIST Rich Transcription (RT-04S) Development Data**" (J. Fiscus et al., 2007b). Ambele sunt disponibile pe DVD contracost.

„**2006 NIST Spoken Term Detection Development Set**" (NIST Multimodal Information Group, 2011b) şi „2006 NIST Spoken Term Detection Evaluation Set" (NIST Multimodal Information Group, 2011a) sunt două corpusuri dezvoltate în scopul detecției de termen (expresie) în limba vorbită. Acestea conțin înregistrări ale emisiunilor radio de ştiri, întâlniri în săli de conferințe şi conversații telefonice. Fişierele provenite de la înregistrările emisiunilor de ştiri sunt pe un singur canal, codate pcm, cu frecvență de eşantionare de 16 kHz, în format SPHERE. Cele provenite de la conversațiile telefonice sunt pe două canale, eşantionate pe 8 kHz, iar cele provenite de la conferințe sunt pe un singur canal. Aceste corpusuri au fost dezvoltate în scopul găsirii tuturor ocurențelor unui

anumit termen (o secvență de unul sau mai multe cuvinte) într-un anumit corpus de semnal vocal. Evaluarea s-a realizat în scopul dezvoltării unei tehnologii pentru căutare rapidă în cantități mari de date audio. Sistemele au fost implementate în două faze: indexare şi căutare. În faza de indexare sistemul procesează datele audio fără cunoaşterea termenilor, iar în faza de căutare sistemul utilizează termenii, indexul şi opțional datele audio pentru a detecta ocurențele de termen. Cele două corpusuri sunt disponibile pe DVD contracost.

2.2. Metodologii utilizate la crearea de corpusuri de semnal vocal şi instrumente aferente

2.2.1. Metodologii pentru crearea de corpusuri destinate recunoaşterii vocale

O. Salor et al. (O. Salor et al., 2007) a dezvoltat un corpus de semnal vocal care conține 193 înregistrări şi instrumente de recunoaştere vocală (un aliniator fonetic şi un sistem de recunoaştere de fonem) pentru limba turcă bazate pe motorul de recunoaştere vocală SONIC. Metoda utilizată de autori în dezvoltarea corpusului constă în analiza lingvistică în vederea determinării alfabetului, crearea corpusului de text şi ulterior crearea corpusului audio. Pentru obținerea unui set echilibrat de propoziții în limba turcă au fost considerate drept unități de bază trifonemele deoarece s-a raportat că sistemele de recunoaştere vocală care utilizează modelul Markov ascuns (HMM) modelează cu o acuratețe mai mare trifonemele în comparație cu fonemele, cuvintele sau silabele.

Metoda utilizată de autori în vederea creării unui corpus audio din trifoneme a constat în determinarea celor mai frecvente ocurențe dintr-un corpus de text suficient de mare pentru a modela limba în mod satisfăcător. Pe baza acestora a fost creat un set de propoziții care a stat la baza dezvoltării corpusului audio. Setul de propoziții a fost obținut prin traducerea în limba turcă a primelor 2000 de propoziții din corpusul TIMIT şi convertirea acestora în simboluri METUbet care reprezintă o mapare a simbolurilor IPA într-un nou set de simboluri ASCII. Ocurențele trifonemelor în setul de propoziții au fost obținute şi comparate cu cele găsite pentru un corpus alcătuit din 2.5 milioane simboluri de cuvinte. A fost adăugat un extra set de 462 propoziții pentru a acoperi cele mai frecvente 5000 de trifoneme existente în corpusul de text.

Corpusul audio a fost creat după metoda utilizată la dezvoltarea corpusului TIMIT. Pentru fiecare vorbitor au fost selectate în mod aleator 40 din 2462 propoziții care au fost citite o singură dată. Au fost colectate înregistrări provenind de la 193 vorbitori (89 de gen

feminin şi 104 de gen masculin) cu vârste cuprinse între 19 şi 50 ani, cu o medie de 23.9 ani care au fost realizate într-un mediu fără zgomot cu un set de căşti prevăzute cu microfon, pe un calculator, cu o frecvență de eşantionare de 16 kHz.

Sistemele de aliniere fonetică şi de recunoaştere de fonem dezvoltate de autori utilizează toolkit-ul SONIC care conține un set de intrumente de recunoaştere a vorbirii continue utilizat ca bază de testare a instrumentelor dezvoltate de specialişti în domeniu. Ambele sisteme utilizează o listă de foneme, un dicționar fonetic, modele fonetice şi modele acustice în limba turcă. Modelele acustice incluse în SONIC sunt modele Markov ascunse (HMM) bazate pe arbori de decizie la nivel de stare conținând funcții asociate densitate de probabilitate gamma pentru a modela duratele stării acestuia. Portul SONIC în limba turcă dezvoltat de autori este primul creat pentru o altă limbă decât limba engleză.

Corpusul dezvoltat de autori a fost acceptat de LDC în octombrie 2005 şi este disponibil pentru cercetători în domeniu. Sistemele de aliniere şi de recunoaştere fonetică au fost testate şi dau rezultate comparabile cu cele obținute pentru limba engleză. Rezultatele obținute pentru recunoaştereea fonetică au arătat că erorile de recunoaştere sunt mai mici la cuvintele finale din propoziții care în general sunt verbe. Matricilie de confuzie create pentru foneme au arătat că plozivele, fricativele şi africatele (P, T, B, D, YH, CH, S) din limba turcă necesită un algoritm de extragere de trăsături mai bun, iar vocala O este singura vocală care are o probabilitate de confuzie ridicată.

M. Boldea et al. (M. Boldea et al., 1998) a creat un corpus de semnal vocal adnotat în limba română destinat în principal dezvoltării de instrumente de recunoaştere vocală independente de vorbitor. Acesta este alcătuit din 100 vorbitori, iar durata totală de înregistrare este de 10 ore. Textele utilizate pentru înregistrări au fost construite prin traducerea versiunii din limba engleză a 40 pasaje din corpusul EUROM-1 care au fost împărțite în 10 clustere printr-o metodă euristică care calculează apriori numărul de ocurențe a fiecărui fonem în fiecare cluster. Pentru a creşte variația fonetică şi pentru a furniza modele de foneme dependente de context au fost adăugate un număr de 550 propoziții individuale, numere întregi, cuvinte în context CVC şi în contexte izolate sau controlate.

Înregistrările s-au realizat într-o cameră fără zgomot, cu softul EUROPEC pe un calculator compatibil şi echipat cu o placă de conversie OROS AU-21 A/D-D/A. Microfonul utilizat de tip electret a fost plasat la 25 cm față de gura vorbitorului, iar ieşirea lui a fost legată la placa OROS printr-un preamplificator cu câştig fix. Frecvența de

eşantionare setată a fost de 20.000 Hz, cu 16 biţi pe eşantion. Fiecare fişier rezultat a fost acompaniat de fişiere de configurare şi descriere.

Adnotarea fişierelor audio a fost realizată în mod similar cu cea pentru corpusul TIMIT şi anume transcripţie manuală, segmentare şi aliniere automată utilizînd HMM, verificare şi corectare manuală. Vorbitorii înregistraţi cu vârste cuprinse între 20 şi 50 de ani au fost împărţiţi pe cinci categorii de vârstă.

2.2.2. Metodologii pentru crearea de corpusuri destinate sintezei de voce

Stan et al. (Stan et al., 2011) au dezvoltat un corpus de limbă română numit RSS pe care l-au utilizat la crearea unui sistem de sinteză voce de înaltă calitate bazat pe HMM-uri, utilizând o frecvenţă de eşantionare înaltă. Metodologia constituirii corpusului constă în realizarea înregistrărilor într-o cameră fără ecou, cu trei microfoane: un microfon cu diafragmă mare (Neumann u89i), un microfon cu diafragmă mică şi bandă largă de frecvenţe (Sennheiser MKH 800) şi un set de căşti cu microfon (DPA 4035). Toate înregistrările au fost realizate cu o frecvenţă de eşantionare de 96 kHz pe 24 biţi şi reeşantionate cu 48 kHz. Această metodă numită supra-eşantionare care presupune eşantionare cu o frecvenţă de patru ori mai mare decât frecvenţă Nyquist ajută la eliminarea zgomotului, raportul semnal-zgomot îmbunătăţindu-se cu un factor egal cu 4. Pentru înregistrări şi reeşantionare autorii au utilizat instrumentele hardware şi software „Pro Tools HD”. Toate înregistrările aparţin unui vorbitor de gen feminin tânăr, vorbitor nativ de limbă română. S-au realizat opt sesiuni într-o lună, în care s-au înregistrat câte 500 de propoziţii.

Corpusul constituit este alcătuit din două seturi de propoziţii, unul de antrenare şi unul de testare. Durata totală de înregistrare pentru setul de antrenare a fost de 3.5 ore şi conţine 3500 propoziţii din care 1500 sunt alese aleator din propoziţii de ziar, 1000 de propoziţii sunt alese astfel încât să acopere cele mai frecvente difoneme şi 1000 provin din pasaje de basme. Timpul de înregistrare pentru setul de testare a fost de 0.5 ore şi conţine 200 respectiv 100 de propoziţii selectate aleator din ziare şi poveşti, iar 200 sunt propoziţii nepredictibile semantic. Rezultatele obţinute arată că, corpusul creat este potrivit pentru utilizarea în sisteme de sinteză voce bazate pe HMM-uri şi că sistemul de sinteză voce dezvoltat prezintă o bună inteligibilitate. Autorii au constatat că o reeşantionare la 32 kHz nu afectează rezultatele, iar reeşantionarea la 16 kHz degradează similaritatea vocii vorbitorului. Corpusul creat este disponibil gratuit online pentru cercetători în domeniu.

J. Matousek şi J. Romportl (J. Matousek şi J. Romportl, 2007) au creat un corpus de limbă cehă în vederea utilizării în sinteza de semnal vocal. În mod identic autorilor din (Stan et al., 2011), înregistrările s-au realizat pe o singură voce de gen feminin şi au inclus pasaje din ziare pe diverse domenii. Autorii au utilizat un algoritm de selecţie automat pentru selectarea propoziţiilor de interes. Din 524.472 propoziţii au fost selectate 5000 astfel încât setul rezultat să fie echilibrat din punct de vedere fonetic şi prozodic. Au mai fost adăugate o serie de propoziţii afirmative, interogative, exclamative şi imperative şi au fost eliminate cele care aveau un număr mai mic de trei sau mai mare de 30 cuvinte. Înregistrările s-au realizat într-un studio audio cu un microfon cu diafragmă mare prevăzut cu un filtru pentru a reduce forţa aerului expirat în timpul pronunţiei plozivelor bilabiale sau a altor consoane. Un card de captură cu fidelitate înaltă a fost utilizat pentru conversia AD de 48 kHz. Concomitent a fost măsurat semnalul glotal în vederea detectării momentelor de închidere a glotei utilizate la estimarea cu acurateţe a conturului frecvenţei fundamentale (F0), la sinteza de voce sincronizată cu F0 sau la detectarea precisă a zonelor vocalice/nevocalice.

Adnotarea corpusului creat s-a făcut în două faze cu ajutorul soft-ului Transcriber^{TM}. În prima fază un adnotator a realizat adnotarea având ca referinţă seturile de propoziţii selectate cu sistemul automat, iar în a doua fază un alt specialist a verificat adnotările facute de primul adnotator şi le-a corectat unde a fost necesar. În procesul de adnotare au fost luate în considerare o serie de reguli referitoare la simbolurile utilizate şi la porţiunile din înregistrări care trebuiau adnotate. Au fost etichetate fonemele, regiunile de linişte, lipsa segmentelor de linişte înaintea şi la sfârşitul înregistrărilor, zgomotele de tip respiraţie, cele produse de buze sau alte zgomote datorate sistemului de achiziţie. În urma adnotării finale au rezultat 62.332 cuvinte din care 7.6% evenimente fără semnal vocal şi 2.62% excepţii de tipul abrevierilor. Autorii concluzionează că deşi diferenţele dintre cele două faze ale adnotării nu sunt foarte mari şi anume în prima fază au fost adnotate corect un procent de 99.62 cuvinte, luând în considerare numărul total de cuvinte din corpus, 237 cuvinte au fost adnotate greşit ceea ce ar produce probleme în sistemul de sinteză voce în timpul selectării unităţilor.

2.2.3. Metodologie pentru crearea unui corpus de dialecte

C. G. Clopper şi D. B. Pisoni (C. G. Clopper şi D. B. Pisoni, 2006) au dezvoltat în cadrul proiectului „The Nationwide Speech Project” un corpus de limbă engleză în

scopul analizei variațiilor la nivel de dialect. Acesta este alcătuit din cuvinte, propoziții, pasaje şi interviuri citite de a câte cinci vorbitori de gen feminin şi cinci vorbitori de gen masculin din şase regiuni dialectale ale Statelor Unite. Înregistrările au fost realizate cu echipament de înregistrare de înaltă calitate într-un mediu fără zgomot. În constituirea corpusului autorii au ținut cont de următorii factori: informații demografice ale vorbitorilor, condiții de înregistrare şi tipul de material înregistrat. Au fost utilizate patru tipuri de material şi anume cuvinte izolate, propoziții, pasaje şi interviuri, timpul total de înregistrare per vorbitor fiind de aproximativ o oră. Cuvintele izolate au fost împărțite în seturi de cuvinte în context hVd, CVC şi cuvinte multisilabice. Cuvintele hVd conțin cinci repetiții a 10 vocale din engleza americană, iar cuvintele CVC sunt alcătuie din 76 cuvinte monosilabice. Fiecare din cele 14 vocale de tip monoftong şi diftong au fost incluse de cel puțin patru ori în lista CVC, iar contextul următor vocalei a fost variat astfel încât să includă consoane lichide, nazale, nesonore şi sonore obstructive.

Propozițiile înregistrate au fost împărțite în funcție de probabilitatea de ocurență în propoziții cu probabilitate de ocurență ridicată, mică şi propoziții greşite din punct de vedere semnatic. Au fost înregistrate două interviuri pentru fiecare vorbitor dintre care unul includea informații legate de rezidență, activități extracuriculare (5 min), iar al doilea a fost structurat astfel încât să conțină cuvinte țintă rostite de vorbitor în mod natural în timpul conversației (7-12 min).

Vorbitorii au fost recrutați din cadrul comunității Universității Indiana, au vârste cuprinse între 18 şi 25 ani, o singură limbă nativă, majoritatea nelicențiați şi fără probleme de auz sau patologii ale vocii la momentul înregistrării. Materialul înregistrat respectiv cuvinte izolate, propoziții, interviuri au fost separate pe blocuri şi afişate pe rând pe un ecran LCD legat la un laptop. Vorbitorul situat în fața ecranului a purtat un microfon ataşat la cap, situat la 2.5 cm de coltul stang al gurii. Ieşirea microfonului a fost legată la un preamplificator, iar câştigul acestuia a fost reglat de către experimentator în timpul pronunției unei propoziții test de către vorbitor. Ieşirea preamplificatorului a fost conectată printr-o o interfață audio USB Roland UA-30 la laptop. De această interfață a fost legat şi un set de căşti purtate de către experimentator pentru a auzi semnalul de input. Fiecare pronunție a fost înregistrată într-un fişier audio digital de tip AIFF pe 16 biți, la o frecvență de eşantionare de 44.1 kHz.

Corpusul de semnal vocal creat este organizat în funcție de vorbitor şi ulterior în funcție de tipul stimulului într-o structură ierarhică de fişiere numite printr-un identificator de trei caractere pentru vorbitor, un caracter pentru tipul stimulului şi patru caractere pentru

numărul stimulului. Pentru fiecare vorbitor s-a creat un fişier în care este definită ordinea în care i s-au prezentat blocurile de cuvinte şi un fişier log care conține ordinea prezentării stimulilor şi legături între numărul stimulului, transcripția ortografică a stimulului şi numele fişierului audio.

Propozițiile înregistrate care prezintă o probabilitate ridicată de ocurență au fost utilizate de către autori ca materiale de stimul într-o serie de sarcini perceptuale. Rezultatele obținute arată că ascultătorii de limbă nativă pot percepe şi reprezenta proprietăți foneto-acustice remarcabile ale dialectelor limbii engleze americane şi pot face judecăți explicite despre vorbitori bazate pe dialectul regional.

2.2.4. Metodologie pentru crearea corpusului multilingv „PFSTAR”

Corpusul de semnal vocal „**PF Star**” multilingv dezvoltat de A. Batliner et al. (A. Batliner et al., 2005) conține 60 de ore de înregistrări provenite de la 611 copii cu vârste cuprinse între 4-14 ani, vorbitori nativi de limbă engleză, germană, italiană şi suedeză. Înregistrările provin din citire, imitare, vorbire spontană sau emoțională. Corpusul este creat pentru a servi ca bază în dezvoltarea sistemelor de recunoaşte automată a vocilor de copii cu limbă nativă sau nu, a vorbirii spontane şi emoționale. Au fost utilizate materiale şi procedee de înregistrare comune pentru toate limbile, iar metodologia AIBO a fost utilizată pentru ambele site-uri de colectare a vorbirii spontane şi emoționale. Materialul utilizat pentru înregistrări constă în 220 cuvinte izolate în limba engleză, 50 de propoziții în limba engleză bogate din punct de vedere fonetic şi 400 propoziții generice. Toate acestea au fost utilizate la înregistrarea copiilor vorbitori nativi de limbă engleză, italiană şi suedeză. Pentru constituirea corpusului de voci spontane şi emoționale a fost utilizată metodologia „AIBO” care constă în utilizarea unui robot controlat căruia copii îi dau instrucțiuni crezând că vorbeşte cu ei. S-au realizat trei experimente şi anume cu robot ascultător care îndeplineşte sarcinile primite, cu robot neascultător care este utilizat pentru a elicita vorbirea emoțională şi cu robot care după instrucțiunile primite trebuie să localizeze obiecte.

Ca metodă generală de înregistrare a fost utilizat un set de căşti cu microfon prevăzut cu preamplificator, frecvența de eşantionate a fost de 44.1 kHz pe 16 biți şi semnalul a fost transmis prin intermediul placii audio la un calculator

Pentru înregistrările de voce spontană şi emoțională s-a utilizat un set de căşti cu microfon wireless, un DATrecorder – casetă audio digitală, o frecvență de eşantionare de

48 kHz şi cuantizare de 16 biți. Fiecare înregistrare a durat 30 de minute, iar timpul total al înregistrărilor eliminând secvențele de liniște datorate răspunsului robotului a fost de 9.2 ore. Toate înregistrările au fost adnotate cu transcripții fonetice marcându-se informațiile referitoare la zgomot sau sunete extralingvistice.

2.2.5. Metodologie pentru crearea corpusului „CIAIR inCar”

Un alt tip de metodologie a fost utilizat la crearea corpusului de semnal vocal „**CIAIR In Car**” (N. Kawaguchi et al., 2005) destinat analizei vorbirii în timpul conducerii unui autovehicul şi realizarea diverselor manevre în trafic. Autorii au dezvoltat un sistem numit „Data Collection Vehicle” (DCV) care suportă înregistrări sincrone de date multicanal provenite de la 12 microfoane plasate în interiorul vehiculului, înregistrări de la trei camere video multicanal, şi colectarea datelor referitoare la vehicul cum ar fi viteză, accelerare, frână, numărul de rotații pe minut ale motorului, etc. Corpusul este alcătuit din 812 subiecți şi conține 1960 sesiuni totalizând un număr de 187.5 ore. Autorii au studiat un corpus format din 1.06 milioane morfeme pentru a determina informații fundamentale specifice vocii şi au utilizat informațiile referitoare la autovehicul şi manevrele făcute pentru a studia modurile în care acestea influențează pronunțiiile vorbitorului. Au fost utilizate trei tipuri de sisteme dialog: un sistem bazat pe operator uman (HUM), un sistem de tip Wizard of Oz (WOZ) şi un sistem de dialog de limbă vorbită (SYS). Toți vorbitorii au completat un chestionar cu informații despre experiența în conducere, în utilizarea sistemelor electronice, de recunoaştere vocală şi de navigație. Corpusul a fost transcris fonetic la nivel de morfem în vederea analizei caracteristicilor specifice limbii vorbite în timpul conducerii de autovehicule.

În urma acestui studiu autorii au concluzionat că viteza de pronunție pentru şoferi a fost mai mică decât în cazul dialogului regular; dialogul cu sisteme de dialog a constat în pronunții mai scurte comparativ cu dialogul cu operator uman; condițiile de conducere (în timpul conducerii sau în pauză) nu influențează dialogul; operațiile efectuate de tip accelerare sau manevră volan influențează semnificativ pronunțiile.

3. Corpusuri de voci patologice şi instrumente aferente

În vederea dezvoltării de instrumente de evaluare, educare şi diagnostic automat a vocii patologice au fost dezvoltate corpusuri şi microcorpusuri de semnal vocal aparținînd persoanelor cu diverse afecțiuni care influențează producerea şi articularea corectă a sunetelor cum ar fi: afecțiuni ale laringelui, afecțiuni respiratorii, neurologice, stomatologice sau de comunicare.

Pentru constituirea acestor corpusuri este necesar un protocol şi o metodologie specială în funcție de aplicațiile urmărite. Suplimentar corpusurilor realizate pentru aplicații de sinteză şi recunoaştere voce este necesară colectarea de informații medicale referitoare la diagnostic, boli asociate şi simptome specifice care duc la modificarea vocii. În ceea ce privește materialul text de înregistrat trebuie ales în funcție de zona din canalul fonator afectată. Spre exemplu daca se dorește evaluarea funcției corzilor vocale sau a aparatului respirator se vor face înregistrări de vocale susținute.

În vederea identificării unor patologii de dentiție se va avea în vedere includerea în corpusul text de consoane a căror locuri de articulare sunt dental, labio-dental sau alveolar. Va fi de asemenea necesară semnarea unui protocol suplimentar de protecție a datelor pacientului. În ceea ce privește protocolul de înregistrare acesta nu diferă semnificativ de cele utilizate la crearea corpusurilor cu alte aplicații. Este necesară utilizarea unei camere cu zgomot redus şi instrumente de înregistrare performante cu respectarea protocoalelor standard legate de frecvența de eşantionare, cuantizare şi achiziție.

La metoda de înregistrare standard cu microfon se pot adăuga instrumente suplimentare cum am fi electroglotograf, stroboscop sau aparat de măsurare a fluxului de aer expirat în funcție de aplicație. După tipul analizei efectuate este necesară repetarea înregistrărilor înainte şi după tratament sau operație pentru a evalua îmbunătățirea sau înrăutățirea calității vocii. În cazul realizării corpusului de către un cercetător din domeniu tehnic-informatic este necesară colaborarea cu un medic specialist în vederea colectării informațiilor cu privire la diagnostic.

3.1. Baza de date de voci patologice dezvoltată de MEEI (MEEI VDDb)

O bază de date cu voci patologice în limba engleză comercială, disponibilă la ora actuală, este distribuită de Kay Elemetrics™ şi a fost dezvoltată de Massachusetts Eye and Ear Infirmary Voice and Speech Lab (**MEEI VDDb**) (Kay Elemetrics Corp, 1994). Corpusul conține 53 respectiv 657 fişiere audio de înregistrări cu voci normale, respectiv patologice ale vocalei susținute /ah/ şi 53 respectiv 661 fişiere de înregistrări cu voci normale respectiv patologice de vorbire continuă. Baza de date conține şi fişiere cu informații personale şi clinice pentru fiecare pacient şi rezultatele obținute în urma analizei acustice obținute cu programul MDVP dezvoltat de Kay Elemetrics™. Înregistrările au fost efectuate cu ajutorul programului CSL (Computerized Speech Lab.) cu frecvențe de eşantionare de 25 şi 50 kHz pe 16 biți. Fiecare înregistrare conține trei repetiții de vocală susținută /ah/ a câte 3 s fiecare. Un specialist pe voce patologică a ales din aceste trei pronunții cea mai bună variantă pentru a fi inclusă în corpus. Această bază de date este larg utilizată de către specialişti în domeniu la dezvoltarea de sisteme şi metode de detecție de voce patologică, însă conform observațiilor făcute de (N. S. Lechon et al., 2006) sunt câteva puncte cheie care trebuie luate în considerare în vederea utilizării acesteia în cercetare:

- nu toți pacienții cu voce patologică au înregistrări sau diagnostic asociat şi sunt unii pacienți care au mai multe înregistrări din diferite vizite la clinică;
- fişierele au frecvențe de eşantionare diferite. Cele provenite de la subiecții normali şi un mic procent din cele provenite de la subiecții patologici sunt eşantionate cu 50 Hz, iar celelalte cu 25 kHz;
- vocile normale şi patologice au fost înregistrate în locații diferite, iar subiecții normali nu au fost evaluați clinic;
- fişierele sunt deja editate astfel încât să includă doar partea stabilă din fonație şi astfel se pierd informațiile cu privire la raportul semnal-zgomot şi la părțile de onset şi offset care conform unor studii în domeniu conțin mai multă informație acustică de interes;
- fişierele cu vocale susținute şi cele cu vorbire continuă a subiecților normali au o durată de 3 s respectiv 12 s, iar cele aparținând subiecților cu voce patologică au durate de 1 s respectiv 9 s;
- există o singură fonație a unei singure vocale per pacient şi per vizită;

- există un număr heterogen de patologii în baza de date, în jur de 200 diagnostice diferite şi sunt multe fişiere etichetate cu mai multe diagnostice, aparținând la categorii diferite de patologii ale vocii;
- în comparație cu înregistrările de voci patologice există un număr mic de voci normale;
- înregistrările nu sunt evaluate din punct de vedere perceptiv cum ar fi indice GRABAS sau prin alte metode şi nu sunt înregistrări video asociate (stroboscopii, endoscopii).

3.2. Metode de detecție de voci patologice care utilizează ca referință MEEI VDDb

Corpusul de voci patologice dezvoltat de Massachusetts Eye and Ear Infirmary Voice and Speech Lab a fost larg utilizat ca referință în dezvoltarea şi testarea de metode şi sisteme de recunoaştere automată a vocilor disfonice, a dizartriei, a patologiei de laringe, a paraliziei şi edemului corzilor vocale, etc.

N. S. Lechon et al. (N. S. Lechon et al., 2006) prezintă o serie de preocupări metodologice care ar trebui luate în calcul la crearea sistemelor automate de detecție a vocii patologice în vederea comparării cu experimente precedente sau viitoare. Autorii subliniază că orice experiment trebuie să aibă o strategie de cros-validare, iar rezultatele trebuie să includă o matrice de confuzie, intervale de confidență pentru toate măsurătorile şi grafice de detectare a performanței sistemului cum ar fi curba DET (Detector Error Tradeoff = eroarea detectorului de compromis) şi curba ROC (receiver operating characteristic = caracteristica de operare a receptorului). Autorii prezintă un exemplu de metodologie şi un experiment bazat pe parametri pe termen scurt şi perceptroni multi-nivel. Experimentul este realizat pe un grup de 53 voci normale şi 173 voci patologice care au fost împărțite în două seturi: unul de antrenare şi unul de testare şi validare a rezultatelor, reprezentând 70% respectiv 30% din numărul total de voci, astfel încât sexul şi vârsta să fie uniform distribuite între cele două clase. După antrenarea sistemului sau calcularea modelelor, setul de test este utilizat pentru estimarea performanței detectorului. În acest scop sunt definite trei măsuri:

- (tp) rata adevărat-pozitiv (senzitivitatea) care reprezintă raportul dintre numărul vocilor patologice corect clasificate şi numărul total de voci patologice;
- (fn) rata fals-negativ care este raportul dintre numărul vocilor patologice greşit clasificate şi numărul total de voci patologice;

- (tn) rata adevărat-negativ (specificitatea) care reprezintă raportul dintre numărul vocilor normale clasificate corect şi numărul total de voci normale;
- (fp) rata fals-pozitiv care este raportul dintre numărul total de voci normale clasificate incorect şi numărul total de voci normale.

Acurateţea finală a sistemului se obţine prin calculul raportului dintre toate vocile clasificate corect de sistem şi numărul total de voci analizate.

Pentru evaluarea capacităţii de generalizare a sistemului trebuie adoptată o schemă de cros-validare. Cea mai simplă constă în a repeta fiecare experiment de N ori, cu un set de test diferit ales random din întreg setul de fişiere sau o altă metodă constă în împărţirea random a setului de date în K subseturi şi repetarea experimentului de K ori utilizând de fiecare dată un set diferit de testare a performanţei. Când numărul K este egal cu numărul F de fişiere disponibile metoda este cunoscută sub numele crosvalidare „leave-one-out". Experimentul este repetat de F ori, iar de fiecare dată sistemul este antrenat cu F-1 fişiere, lăsând fişierul rămas pentru testare. După crosvalidare rezultatele finale se mediază între repetiţii şi intervalele de confidenţă pot fi calculate utilizând deviaţiile standard ale măsurătorilor. O schemă a matricii de confuzie este dată în Tabelul 1.

Tabel 1. Ilustrarea unei matrici tipice de confuzie a sistemului de clasificare (N. S. Lechon et al., 2006)

Decizia detectorului	Diagnostic actual	
	Patologic	Normal
Patologic	tp	fp
Normal	fn	tn

Matricea de confuzie se calculează pe baza comparaţiei dintre scorul obţinut în urma testării şi o valoare prag. Dacă modificăm această valoare prag obţinem o serie de puncte posibile de operare ale sistemului care pot fi reprezentate printr-o curbă DET frecvent utilizată în verificarea vorbitorului. Această curbă se obţine prin reprezentarea grafică a valorilor fals pozitive în funcţie de cele fals negative la diferite valori prag. O altă curbă des utilizată în sistemele medicale de decizie este curba ROC care se obţine prin reprezentarea valorilor fals pozitive în funcţie de cele adevărat pozitive. Cea mai utilizată măsură este aria de sub curba ROC (AUC) care în cazul în care performanţa sistemului este slabă are valoare în jur de 0.5, iar daca performanţa este bună are valoarea 1. Eficienţa sistemului este ridicată când curba DET se apropie de colţul stâng de jos al graficului şi curba ROC se apropie de colţul stâng de sus al graficului.

Utilizând metodologia descrisă mai sus autorii au dezvoltat un sistem de recunoaştere a vocii patologice bazat pe o reţea neuronală care modelează cele două clase

de voci (normală şi patologică). Detectorul este un perceptron de tip „feedforward" cu mai multe nivele. Nivelul de intrare este alcătuit din mai multe intrări ale parametrilor MFCC, nivelul ascuns are 12 neuroni, iar nivelul de ieşire are două noduri. Cele două ieşiri sunt construite astfel încât să se obţină o rată a similitudinii sau un scor pentru fiecare pattern de intrare. Este utilizat un algoritm de învăţare supervizat bazat pe metoda backpropagation cu moment şi regulă delta. Funcţiile de activare ale tuturor nodurilor sunt de tip logic, iar pnderile de conectare sunt iniţializate cu valori random dintr-o distribuţie gauss cu medie zero şi deviaţie standard inversă numărului de ponderi al fiecărui neuron. Au fost realizate 40 de iteraţii ale algoritmului de antrenare, iar experimentul a fost repetat de 10 ori utilizînd seturi random diferite de antrenare şi testare. Acurateţea totală a sistemului este de 89.6% cu o deviaţie standard între experimente de ±2.49%, iar procentajul de voci clasificate corect este de 90.42%.

Un alt sistem de detectare automată a vocii normale şi patologice utilizând eşantioane de voce din corpusul MEEI VDDb a fost creat de **A. A. Dibazar şi S. Narayanan** (A. A. Dibazar şi S. Narayanan, 2002). Sistemul de detecţie este un clasificator de tip HMM care modelează coeficienţii Mel cepstrali şi dinamicile frecvenţei fundamentale prin mixuri gaussiene. Metoda a fost evaluată utilizând caracteristici ale vocalei susţinute /a/ provenind de la 700 subiecţi cu voci normale şi patologice şi a constat în asocierea fiecărui subiect a unui vector de trăsături care include 42 trăsături extrase din informaţia spectrală şi a F0. Pentru antrenarea HMM-ului a fost utilizat toolkit-ul HTK care a fost modificat pentru adaptarea la frecvenţa fundamentală. Autorii au comparat rezultatele obţinute cu această metodă cu cele obţinute prin clasificarea vocilor utilizând caracteristici în domeniul timp frecvenţă extrase cu programul MDVP (Multi Dimensional Voice Program) şi patru clasificatori: LDC (analiză liniară discriminantă), NCM (clasificatorul bazat pe cea mai apropiată medie), HMM (model cu trei mixturi gaussiene) şi o reţea neuronală cu 34 neuroni de intrare, un nivel ascuns şi doi neuroni de ieşire.

Acurateţea sistemului bazat pe mixturi gaussiene cu utilizarea coeficienţilor Mel cepstrali şi a frecvenţei fundamentale este de 99.40%, iar cea obţinută în cazul utilizării parametrilor extraşi cu MDVP (F0 minim, F0 maxim, F0 mediu, jitter, shimmer, etc.) este de 97.97%. Autorii au realizat şi o clasificare la nivel de patologie, împărţind vocile patologice în patru clase şi anume patologie minoră, uşoară, moderată şi severă. La evaluarea patologiei utilizând parametrii MDVP pentru vocala susţinută /a/ şi în cazul utilizării parametrilor spectrali şi F0 s-au obţinut rate de clasificare de 57.99%, respectiv 72.55%. Atât la discriminarea de voce normală-patologică cât şi la identificarea de tip de

patologie sistemul HMM cu mixturi gaussiene utilizând trăsături spectrale şi F0 a arătat o îmbunătățire semnificativă a rezultatelor de clasificare.

R. J. Moran et al. (R. J. Moran et al., 2006) au construit un sistem de detectare a patologiilor corzilor vocale prin înregistrări telefonice. Sistemul are la bază un clasificator liniar, măsurători de procesare a perturbațiilor frecvenței fundamentale, a amplitudinii şi a raportului semnal-zgomot (HNR). Pentru dezvoltarea şi validarea sistemului au fost utilizate eşantioane de voce din MEEI VDDb .

Plecând de la baza de date originală au fost create cinci corpusuri. Patru din ele alcătuite din 631 înregistrări de voce au fost create prin reeşantionarea la 10 kHz şi adăugarea a diverse distorsiuni care pot fi transmise prin telefon. Acestea au fost utilizate la evaluarea acurateții de discriminare între vocile normale şi patologice cu diferite seturi de trăsături. Al cincilea corpus a fost construit prin transmiterea tuturor înregistrărilor prin intermediul unui canal telefonic la distanță mare şi înregistrarea fiecărui semnal primit la receptor.

Trăsăturile utilizate în construirea clasificatorului au fost perturbațiile de F0 şi de amplitudine şi HNR. Dintre trăsăturile comune ale F0 şi amplitudinii sunt parametrii statistici medie ($F0\_av, Amp\_av$), minim ($F0\_lo, A\_lo$), maxim ($F0\_hi, A\_hi$) şi deviație standard ($F0\_sd, A\_sd$). Alte trăsături derivate din F0 sunt [ ]:

- PFR (Phonatory frecvency range):

$$PFR = \frac{\log(\frac{F0\_hi}{F0\_lo})}{\log 2} \times 12, \tag{1}$$

- MAJ (Media absolută a jitter-ului):

$$MAJ = \frac{1}{n-1}\sum_{i=1}^{n-1} | F_{i+1} - F_i |, \tag{2}$$

- Jitter (%):

$$JITT = \frac{MAJ}{F0\_av}, \tag{3}$$

- RAP (Media perturbării relative regulată pe 3 perioade):

$$RAP = \frac{\frac{1}{n-2}\sum_{i=2}^{n-1}\left|\frac{F_{i+1}+F_i+F_{i-1}}{3} - F_i\right|}{F0\_av} \times 100, \tag{4}$$

- PPQ_5 (Coeficientul de perturbație a F0 regulat pe 5 perioade):

$$PPQ\_5 = \frac{\frac{1}{n-4}\sum_{i=3}^{n-2}\left|\frac{\sum_{n=i-2}^{i+2}F(k)}{5} - F_i\right|}{F0\_av} \times 100, \quad (5)$$

- PPQ_55 (Coeficientul de perturbație a F0 regulat pe 55 perioade):

$$PPQ\_55 = \frac{\frac{1}{n-54}\sum_{i=28}^{n-27}\left|\frac{\sum_{n=i-27}^{i+27}F(k)}{55} - F_i\right|}{F0\_av} \times 100, \quad (6)$$

- PPF (factorul de perturbație a F0):

$$PPF = \frac{N_{p\geq threshold}}{N_{voice}}, \quad (7)$$

unde N_p este perturbația în magnitudine a perioadei în timp mai mare decît 0.5 msec.

- DPF (factorul direcțional de perturbație):

$$DPF = \frac{N_{\Delta\pm}}{N_{voice}} \times 100, \quad (8)$$

unde $N_{\Delta\pm}$ este perturbația epocii în timp pentru care există o schimbare se semn.

Trăsăturile derivate din perturbațiile amplitudinii sunt:

- Shimmer (SHIM)(%):

$$SHIM\% = \frac{MAS}{Amp\_av}, \quad (9)$$

unde *MAS* este media absolută a shimmer-ului care se calculează similar cu *MAJ* .

- Shimmer (SHIM)(db):

$$SHIM = \frac{1}{n-1}\sum_{i=1}^{n-1} 20\log\frac{A_i}{A_{i+1}}, \quad (10)$$

- ARP_3 (media perturbării relative a aplitudinii regulată pe 3 perioade), APQ_5, APQ_55, APF şi ADPF care se calculează similar cu RAP, PPQ_5, PPQ_55, PPF şi DPF.

Histogramele aferente unui număr de 35 trăsături utilizate realizate pe grupul de voci normale şi cel de voci patologice au arătat că acestea prezintă o distribuție aproximativ gaussiană ceea ce a înlesnit construirea unui clasificator de discriminare liniară (LDC).

Metoda generală de antrenare a clasificatorului utilizată de către autori a constat în următorii paşi: (a) crearea unui vector coloană *x* care conține *d* trăsături care trebuie asignat la una din cele două clase; se consideră că există N_1 vectori de trăsături pentru antrenarea clasificatorului din clasa 1 şi N_2 trăsături din clasa 2; al *n*-lea vector de trăsături pentru

antrenare în clasa *k* este notat cu $x_n^{(k)}$;(b) determinarea vectorilor medie μ_1, μ_2 care condiționează clasele cu formulele [ ]:

$$\mu_1 = \frac{1}{N_1}\sum_{i=1}^{N_1} x_n^{(1)}, \mu_2 = \frac{1}{N_2}\sum_{i=1}^{N_2} x_n^{(2)}. \quad (11)$$

şi a matricii comune de covarianță:

$$\sum = \frac{1}{N_1 + N_2 - 2}\sum_{k=1}^{2}\sum_{n=1}^{N_k}(x_n^{(k)} - \mu_k)(x_n^{(k)} - \mu_k)^T. \quad (12)$$

(c) clasificarea unui vector de trăsături *x* prin setarea unei probabilități a priori π_1 ($\pi_1=\pi_2$-$\pi_1=0.5$) la clasa 1 şi calculul valorii discriminante *y*:

$$y = (\mu_1 - \mu_2)^T \sum^{-1} x - \frac{1}{2}(\mu_1 - \mu_2)^T \sum^{-1}(\mu_1 + \mu_2) + \log(\frac{\pi_1}{1-\pi_1}). \quad (13)$$

(d) calcularea probabilităților posterioare:

$$P(1 \mid x) = \frac{\exp(y)}{\exp(y)+1}, P(2 \mid x) = 1 - P(1 \mid x). \quad (14)$$

(e) alegerea probabilității posterioare maxime.

Clasificatorul a fost testat prin metoda cros-validării. Rezultatele arată că o fonație susținută înregistrată într-un mediu controlat poate fi clasificată drept normală sau patologică cu o acuratețe de 89.1 %, în timp ce vocea telefonică poate fi clasificată drept normală sau patologică utilizând aceeaşi metodă cu o acuratețe de 74.2%. Caracteristicile referitoare la perturbațiile de amplitudine s-au dovedit a fi cele mai robuste în clasificare. Înregistrările au fost împărțite pe patru categorii şi anume voci normale, cu patologie neuromusculară, fizică şi mixtă (neuromusculare şi fizice). A fost dezvoltat un clasificator separat pentru discriminarea grupului normal de fiecare din cele trei patologii. Rezultatele au arătat că patologiile neuromusculare pot fi detectate de la distanță cu o acuratețe de 87%, cele fizice cu 78%, iar cele mixte cu 61%.

3.3. Alte corpusuri de voci patologice şi instrumente aferente

Un corpus de voce patologică şi normală în limba germană numit **COPAS** (Dutch Corpus of Pathological and Normal Speech) a fost dezvoltat de către G. Van Nuffelen et al. şi a fost făcut public prin intermediul Uniunii Limbii Germane (http://www.inl.nl/tst-centrale/images/stories/producten/documentatie/copas_manual.pdf). Acesta a fost dezvoltat pentru a constitui baza unui instrument de evaluare a vocii patologice bazat pe diverse teste de inteligibilitate. Corpusul este alcătuit din opt categorii distincte de diagnostic şi anume

voce normală, voce patologică, laringectomii, disartrii, glosectomii, defecte de auz, fisuri de palat şi buză şi defecte de articulare, totalizând un număr de 319 cazuri din care 122 normale.

Informaţiile cu privire la vorbitor, de tip cod, vârstă, sex, numărul înregistrării, perioada de la prima înregistrare, tipul microfonului utilizat, subseturile de cuvinte şi indicele de inteligibilitate aferent şi tipul patologiei sunt stocate într-un fişier Excel™. Au fost înregistrate o serie de 11 teste, însă nu toţi subiecţii au participat la fiecare test şi nu toate materialele text utilizate pentru teste au fost adnotate. Primul test a constat în înregistrarea a 50 de cuvinte în context CVC şi calcularea inteligibilităţii la nivel de fonem, al doile test a constat în citirea a 11 pasaje de text diferite, iar testele teste au constat în citirea unui text care conţine un număr echilibrat de foneme în limba germană, evaluarea articulării prin numirea de desene, măsurarea ratei diadochokinetică care constă în repetarea fonemului /p/ timp de opt secunde, evaluarea tranziţiilor formantice prin repetarea alternativă timp de 6 secunde a vocalelor /i/ şi /u/, în vorbire spontană şi semispontană, şi susţinerea vocalei /a/ timp de 5 s.

Toate înregistrările au fost realizate într-o cameră clinică liniştită, cu două microfoane, unul de masă şi unul cu cască şi au fost salvate pe minidiscuri şi apoi transferate pe un notebook. Transferul s-a făcut cu un editor audio gratuit Audacity®. Fiecare eşantion audio a fost salvat în format .wav pe 16 biţi, cu o frecvenţă de eşantionare de 16 kHz. Toate eşantioanele audio înregistrate pentru un anumit test sunt salvate în acelaşi document. Adnotările au fost realizate cu programul Praat™ şi au fost stocate în fişiere .TextGrid. Fiecare fişier audio şi de adnotare a aceluiaşi vorbitor a fost numit cu un caracter de început semnificând patologia, un cod vorbitor şi un string referitor la subtestul înregistrat.

J. F. Bonastre et al. (J. F. Bonastre et al., 2007) a dezvoltat tehnici complementare de evaluare a vocii patologice utilizând o selecţie de înregistrări din baza de date a Departamentului ENT al Centrului Universitar Spitalicesc Timone din Marseille. Primul studiu s-a bazat pe un număr de 449 subiecţi din care 391 pacienţi cu voci patologice (308 femei şi 20 bărbaţi) şi 58 subiecţi de control cu voce normală. Pacienţii prezentau o serie de afecţiuni ale vocii întâlnite frecvent în practica clinică şi anume 96 pacienţi cu noduli vocali, 91 cu polipi, 65 cu disfonie paralitică, 55 cu edem Reinke, 27 cu chişti, 24 cu disfonie funcţională, 19 cu displazie şi 14 cu „Sulcus Glotidis".

S-a realizat o evaluare perceptuală de către patru ascultători cu experienţă după scala GRBAS, însă doar componenta G a fost utilizată pentru acest studiu. Analiza

obiectivă s-a efectuat cu ajutorul staţiei de lucru EVA®. Acest sistem permite realizarea de măsurători simultane ale parametrilor acustici şi aerodinamici utilizând o piesă specială situată în dreptul gurii prevăzute cu microfon şi pneumotahograf. Presiunea intraorală este măsurată cu ajutorul unui senzor de presiune montat în interiorul piesei.

Metoda constă în pronunţia consecutivă a trei vocale susţinute (/a/) care sunt analizate uterior la nivel de F0, intensitate, jitter, shimmer, SNR (raportul semnal – zgomot), fluxul de aer oral şi coeficientul Lyapunov. Au fost realizate trei sesiuni de măsurători pentru fiecare parametru şi datele corespunzătoare pe fiecare parametru sunt o mediere a celor trei măsurători. S-au măsurat de asemenea presiunea subglotică în timp ce subiectul a pronunţat opt grupuri CV (/pa/) consecutive cu nivel normal de pich şi claritate, minimul şi maximul F0 şi durata maximă a fonaţiei.

Rezultatele au arătat că o combinare neliniară de şapte parametri a permis clasificarea identică cu cea a juriului a 82% eşantioane de voce. Al doilea studiu realizat de autori se referă la adaptarea tehnicilor de recunoaştere automată de vorbitor la evaluarea vocii patologice. Sistemul construit pentru această sarcină are la bază modele de mixturi gaussiene şi este testat pe 80 de voci de gen feminin din care 20 sunt normale. Rezultatele obţinute în acest studiu arată că gradul cel mai mare de confuzie apare între grade adiacente de patologie.

J. B. Tomblin (J. B. Tomblin, 2010) a dezvoltat corpusul EpiSLI care conţine înregistrări de semnal vocal provenite de la copii de grădiniţă cu defecte de limbaj. Acesta este creat în scopul utilizării pentru detecţia automată de defecte de limbaj şi voce patologică. Grupul target a fost un cluster stratificat în funcţie de locul de rezidenţă şi şcoală format din 6000 de copii vorbitori nativi de limbă engleză. Acesta a fost împărţit în trei categorii urban, suburban şi rural. Părinţii au semnat un consimţământ cu privire la participarea copiilor lor la acest studiu. Un procent de 2% (161) au refuzat să participe.

Copii au fost evaluaţi mai întâi în ceea ce priveşte performanţa limbajului cu ajutorul unui test din 40 de puncte dezvoltat de autori, care a durat în jur de 10 minute. Copii din grupul de control au fost selectaţi din aceaşi şcoală cu cei care nu au trecut primul test. Din 3877 copii care au fost selectaţi pentru faza de diagnostic doar 2084 au primit permisiunea părinţilor. 70 dintre aceştia sunt vorbitori de o a doua limbă şi au fost excluşi. Grupul final a fost alcătuit din 2009 copii vorbitori nativi de limbă engleză.

Scopul diagnosticării a constat în identificarea copiilor cu defecte de limbaj şi a celor cu limbaj normal pentru grupul de control. Faza de diagnostic a inclus evaluarea auzului, a limbajului, a vorbirii, a înţelegerii şi observaţii motorii şi a durat aproximativ

două ore la fiecare copil. Baza de date creată este disponibilă la cerere fără cost, făcând parte din inițiativa Institutelor Naționale de împărțire a datelor.

4. Microcorpusuri şi metode de detecție a vocilor patologice

G. Schlotthauer şi M.E. Torres (G. Schlotthauer şi M.E. Torres, 2006) a realizat un sistem de clasificare automat a vocilor disfonice şi normale, care discriminează şi între două tipuri diferite de disfonie şi anume disfonia spasmodică (SD) de natură neurologică şi disfonia datorată tensiunii musculare (MTD) de natură fiziologică, care poate fi tratată. Clasificatorul automat utilizează măsurători acustice ale vocalei susținute /a/ şi instrumente de recunoaştere de pattern-uri bazate pe rețele neuronale.

Corpusul utilizat este alcătuit din 89 vorbitori împărțiți în două grupuri: 36 cu disfonie (15 cu MTD şi 21 cu SD) şi 53 cu voci normale care au fost instruiți să susțină vocala /a/ cel puțin 3 s. Înregistrările au fost realizate cu o frecvență de eşantionare de 22 kHz pe 16 biți. În acest studiu autorii au utilizat un singur eşantion de voce per pacient din care au extras opt parametri acustici care formează un pattern, incluzând perturbația pe termen scurt a frecvenței fundamentale şi a intensității şi zgomotul glotal. Toate pattern-urile au fost clasificate în trei categorii: normale SD şi MTD utilizînd o rețea neuronală cu percepron multistrat. Pentru a reduce numărul de intrări ale rețelei neuronale autorii au realizat o analiză a componentelor principale (PCA) în urma căreia a rezultat că un număr de cinci componente a contribuit cu 99.5% la varianța setului de date. Astfel vectorii de intrare ai rețelei neuronale au fost reduşi de la opt la şase. Dimensiunea nivelului ascuns a fost variată între 8 şi 34 neuroni şi s-au rulat 100 experimente pentru fiecare caz în vederea selecției celei mai bune topologii. Nivelul de ieşire are trei neuroni, câte unul pentru fiecare clasă (SD, MTD şi normal), iar cel câştigător este cel cu valoare maximă.

Patologiile SD şi MTD au fost recunoscute cu o acuratețe de 95.24%, respectiv 86.67%. În ceea ce priveşte vocile normale, acestea au fost recunoscute în procent de 100%. Rezultatele obținute de către autori utilizând o rețea neuronală cu perceptron multistrat cu nivelul ascuns alcătuit din opt neuroni indică o acuratețe de recunoaştere de 98.94%, depăşind alte raportări din literatura de specialitate bazate pe clasificator de tip SVM (Suppor Vector Machine) (96.5%).

R. F. Pozo et al. (R. F. Pozo et al., 2009) au creat un microcorpus de voci patologice şi normale în vederea detectării automate a pacienților cu apnee obstructivă în

somn (OSA), utilizând analiza semnalului vocal şi tehnici de recunoaştere automată de voce şi vorbitor. Acesta este alcătuit din 80 de subiecţi de gen masculin dintre care 50% suferă de apnee severă şi 50% sunt fie subiecţi sănătoşi fie cu forme uşoare de apnee. Subiecţii din ambele grupuri au caracteristici fizice cum ar fi vârsta şi indicele de masă corporal similare. Pentru grupul subiecţilor cu apnee s-au realiat două sesiuni de înregistrări, una înainte de a fi diagnosticaţi şi una după câteva luni de tratament.

Corpusul conţine înregistrări a patru propoziţii în limba spaniolă care au fost repetate de trei ori de fiecare vorbitor. Acestea au fost selectate astfel încât să poată fi studiate anomaliile de rezonanţă, de fonaţie şi de articulare. S-a urmărit gradul de nazalizare a vocalelor cu şi fără context nazal, a sunetelor sonore continue în vederea măsurării pattern-urilor de voce neregulată cauzate de oboseala muşchilor şi a sunetelor sonore afectate de câteva foneme precedente care au locul de articulare velar (/g/).

Înregistrările au fost efectuate la o frecvenţă de eşantionare de 48 kHz într-o încăpere izolată acustic, iar aparatura utilizată a constat dintr-un laptop prevăzut cu placă audio şi un microfon de tip cască prevăzut cu conversie A/D şi transmiterea datelor digitale prin USB. În mod adiţional pentru fiecare subiect au fost colectate şi două imagini faciale pe fond alb cu iluminare controlată cu ajutorul unei camere digitale întrucât inspecţia vizuală reprezintă un prim pas în diagnoticul pacienţilor suspectaţi de OSA.

În ceea ce priveşte anomaliile de articulare autorii au constatat că în cazul subiecţilor cu apnee distanţa dintre formanţii F2 şi F3 ai vocalei /i/ este mai mare decât în cazul pacienţilor sănătoşi. Pentru detectarea anomaliilor de fonaţie au fost calculaţi parametrii HNR (Raportul armonici – zgomot) şi disperiodicitatea care se referă la anomaliile semnalului generat de excitaţia glotală. O voce normală ar trebui să aibă un HNR mare şi o disperiodicitate mai mică decât o voce patologică. Aceşti parametri au fost calculaţi pentru a patra frază din baza de date deoarece conţine mai multe sunete sonore. Se cunoaşte faptul că pacienţii cu apnee prezintă o nazalitate anormală prin prezenţa în voce unei extracomponente de frecvenţă joasă. Autorii au găsit o variaţie a nazalizării mai mică la subiecţii cu OSA decât la cei normali. O ipoteză emisă de autorilor în urma acestui rezultat este aceea că pacienţii cu apnee au un nivel de hipernazalitate global mai mare datorat disfuncţiei velofaringeale şi astfel diferenţa între vocalele orale şi nazale este mai mică deoarece şi cele orale sunt nazalizate.

Luând în calcul ca trăsătură de bază nazalitatea autorii au testat un model de mixturi gaussiene (GMM) în vederea discriminării între pacienţii cu OSA şi cei sănătoşi. Rezultatele au arătat că există diferenţe semnificative între cele două grupuri în ceea ce

priveşte nivelele relative de nazalizare între diferite contexte lingvistice. Autorii au testat de asemenea puterea discriminativă a modelului GMM bazat pe tehnici de recunoaştere de vorbitor adaptate la detecţia apneei severe obţinând o rată de clasificare corectă de 81%

L. Salhi et al. (L. Salhi et al., 2010) au creat o nouă metodă de detectare a vocii patologice bazată pe o reţea neuronală multistrat (MNN). Algoritmul de procesare are la bază o tehnică hibridă care utilizează ca intrare energia coeficienţilor wavelet. Primul pas a constat în antrenarea supervizată a sistemului în vederea discriminării între voci normale şi patologice şi al doilea pas a constat în clasificarea patologiilor neurale şi vocale de tip Parkinson, laringeale, Alzeiheimer etc. A fost utilizată o bază de date care conţine voci normale şi patologice colectate de la spitalul „Rabta-Tunis". Iniţial autorii au testat o metodă de determinare a vocii patologice şi normale prin analiză formantică şi de F0. Algoritmul a constat în extragerea F0 şi a formanţilor prin metoda cepstrumului respectiv LPC şi compararea cu valorile normale şi împarţirea vocilor în două clase o clasă cu voci normale şi o clasă cu voci patologice împărţită la rândul ei în două subclase cu patologie neurală şi organică.

Algoritmul utilizat la metoda hibridă de clasificare constă în preaccentuarea şi segmentarea semnalului cu o fereastra Hamming şi aplicarea transformatei wavelet în vederea extragerii energiei coeficienţilor care vor constitui vectorul de intrare în reţeaua MNN. Sistemul propus de autori a fost dezvoltat în Matlab şi este alcătuit din trei nivele: un nivel de intrare alcătuit dintr-un număr de neuroni egal cu numărul componentelor vectorului de trăsături, un nivel ascuns alcătuit din 15 neuroni şi un nivel de ieşire alcătuit dintr-un singur neuron corespunzător deciziei patologic sau normal. Fiecare neuron din stratul ascuns este conectat la neuronii de intrare şi nu există nici o conexiune între celulele aceluiaşi strat. Funcţia de activare utilizată este cea de tip sigmoid. Pentru extragerea coeficienţilor wavelet a fost utilizat un banc de filtre şi ulterior energia fiecărui nivel a fost normalizată în raport cu energia totală a semnalului:

$$E_N(i) = \frac{E_i}{\sum_i E_i}, \tag{15}$$

unde E_i este energia la nivelul i.

Autorii au testat sistemul creat cu trei, cinci şi şapte coeficienţi de energie wavelet pe un total de 60 cuvinte pronunţate de vorbitori diferiţi, din care 30 cu voce normală şi restul cu voce patologică. Pentru antrenare au fost utilizate câte 20 de cuvinte din cele două

seturi. Rata de clasificare corectă cu sistemul dezvoltat a fost între 80% şi 100%, cea maximă fiind obținută în cazul utilizării a şapte coeficienți de energie.

J. Lee şi M. Hahn (J. Lee şi M. Hahn, 2009) propun o nouă metodă de detectare a vocii patologice prin analiză statistică de ordin înalt (HOS) în domeniu timp, bazată pe coeficienți de predicție liniară reziduali (LPC). Ca trăsături de bază au fost utilizați parametrii statistici skewness şi kurtosis normalizați. Studiul se bazează pe un corpus format din 83 eşantioane de voce care conțin vocala susținută /a/, furnizate de către Societatea Logopedică şi Foniatrică Japoneză. Acestea au fost evaluate de către terapeuți de limbă conform scalei GRBAS şi împărțite pe categorii în funcție de severitatea disfoniei. Drept clasificator a fost utilizat un arbore de decizie regresiv, iar cele mai bune rezultate cu o acuratețe de clasificare de 92.9% au fost obținute prin combinarea celor doi parametri statistici.

4.1. Aplicații în stomatologie

La ora actuală nu este disponibilă o bază de date de referință cu înregistrări de voci patologice cauzate de diverse patologii ale aparatului stomatognat care să poată fi utilizată în evaluarea calității vocii în stomatologie. S-au realizat studii pe grupuri mici de pacienți, iar bazele de date sunt utilizate doar de grupul de cercetare care le-au dezvoltat. Evaluarea eşantioanelor de voce s-a realizat fie prin metode subiective şi anume percepție auditivă de către specialişti (K. Stevens et al., 2011), fie prin utilizarea de software-uri de analiză vocală (P. Jindra et al., 2002), fie prin dezvoltarea de metode automate proprii (Bocklet et al., 2010).

Tehnici de recunoaştere vocală automată au fost utilizate în evaluarea şi diagnosticul automat al pacienților edentați, cu dantură insuficient fixată. Astfel **Bocklet et al.** (Bocklet et al., 2010) au testat trei sisteme de clasificare diferite, în vederea determinării dacă dantura unui pacient edentat este fixată corect. Baza de date utilizată în acest studiu este alcătuită din 13 pacienți edentați cu vârste cuprinse între 54 şi 74 ani. Aceştia au fost diagnosticați ca având dantură insuficientă şi li s-au construit danturi adiționale. Tehnica a constat în înregistrarea pacienților fără dantură, cu dantură insuficientă şi ulterior cu dantură suficientă. Materialul de înregistrat este un text standard în limba germană, echilibrat fonetic, care conține 108 cuvinte. Înregistrările au fost realizate cu o frecvență de eşantionare de 16 kHz şi cuantizate pe 16 biți. Autorii au creat un vector de trăsături de dimensiune 24 care conține energia pe termen scurt, 11 coeficienți mel cepstrali şi

derivatele delta de ordin 1 ale acestora, pe care l-au utilizat la ASR şi la două din cele trei sisteme de clasificare.

Sistemul de clasificare de bază are un vector de trăsături unidimensional reprezentat de acuratețea de recunoaştere a cuvintelor (WA) cu un sistem de recunoaştere vocală automată (ASR). În urma testării ASR pe cele trei seturi de înregistrări (fără dantură, cu dantură insuficientă şi suficientă) s-au obținut rezultate diferite pentru WA (60.06%, 64.35%, respectiv 70.91%). Din acest motiv autorii au utilizat WA ca parametru pentru discriminarea celor trei clase de vorbitori cu un sistem de clasificare de tip SVM (Support Vector Machine).

Al doilea sistem de clasificare utilizează tehnica DTW (Dinamic Time Warping) pentru a extrage vectorul de trăsături care codează distanțele în spațiul MFCC dintre înregistrarea unui vorbitor de test cu dantură şi un vorbitor de referință fără patologie. Vectorul de trăsături de dimensiune 2314 este ulterior clasificat cu un SVM. Al treilea sistem de clasificare utilizează ca trăsături media vectorilor unui model de mixturi gaussiene (GMM) de dimensiune 128. Supervectorul GMM de dimensiune 3072 a fost ulterior clasificat cu un SVM.

Rezultatele obținute în urma clasificării celor trei seturi de înregistrări relevă o acuratețe a clasificării de 61.5%, 73.1% respectiv 80.1% cu sistemele de clasificare bazate pe WA, DTW, respectiv GMM.

K. T. Bressmann et al. (K. S. Bressmann et al., 2011) au studiat impactul expanderilor palatali rapizi (RPE) asupra articulării sunetelor. Aceştia sunt prevăzuți cu atașamente cimentate la dinți şi au o porțiune care acoperă palatul. Datorită poziției pe care o ocupă şi dimensiunii pot afecta vorbirea. Scopul studiului a constat în evaluarea perturbației şi adaptării vocii de-a lungul timpului, relaționate la aplicarea expanderilor. Au fost înregistrați 22 de pacienți cu vârste cuprinse între 9 şi 19 ani, care urmau să fie tratați cu RPE. Înregistrările s-au realizat în şase etape: înainte de plasarea expanderilor, după plasare, în timpul expansiunii, în timpul retenției, după eliminarea expanderului şi după patru săptămâni de la eliminare.

Materialul înregistrat a fost alcătuit din 35 de propoziții din care trei au fost alese pentru analiză. Eşantioanele de voce au fost evaluate perceptiv de către 10 ascultători specialişti, iar formanții vocalei /i/ şi spectrul fricativelor /s/ şi /ʃ/ au fost măsurate cu software-ul de analiză vocală Wavesurfer^{TM}.

Rezultatele au arătat că după plasarea expanderului vocea s-a degradat, în timp s-a îmbunătăţit şi după eliminarea expanderului a revenit la normal. Primul formant al vocalei /i/ a crescut şi al doilea formant a scăzut în frecvenţă indicând centralizarea vocalei. Formanţii au revenit la valorile dinaintea aplicării expanderului în timpul tratamentului. Pentru fricativele /s/ şi /ʃ/ rapoartele frecvenţelor joase şi înalte au arătat prezenţa unei distorsiuni care a dispărut după îndepărtarea expanderului.

Rezultatele au arătat că media spectrală a scăzut iar skewness-ul a devenit pozitiv. Analiza repetată a arătat o varianţă semnificativă a măsurătorilor acustice. Vocea a fost alterată când expanderul a fost aplicat prima dată, s-a îmbunătăţit pe parcurs şi a revenit la normal după îndepărtarea lui.

Gradul de îmbunătăţire a articulării sunetelor la pacienţi edentaţi cărora li s-a aplicat dantură parţială sau completă a fost studiat şi de P. Jindra et al. (P. Jindra et al., 2002). Au fost investigaţi parametrii semnalului vocal pentru un set de silabe test pronunţate de câteva ori de către un grup de zece pacienţi dintr-o clinică stomatologică dintre care cinci bărbaţi şi cinci femei cu vârste cuprinse între 58 şi 81 ani. Silabele analizate au fost : ”si, že, ša, ta, da,ře, ra, fa, vi, pa, bo”, pronunţate de către pacienţi cu şi fără proteză parţială sau completă aplicată pe maxilar sau mandibulă. Odată cu înregistrările a fost realizată şi o anamneză a pacienţilor care conţinea probleme subiective întâmpinate de pacienţi în vorbirea fără proteză. Analiza Fourier a înregistrărilor realizată cu programul ”SoundForge 5.0” a arătat că există diferenţe în ceea ce priveşte parametrii acustici ai consoanelor dar şi a vocalelor pronunţate de pacienţi fără proteze, în special la silabele care conţin consoane alveolare.

Analiza silabei ”si” a arătat o scădere în limita frecvenţelor superioare şi o descreştere a maximului spre frecvenţele joase când articularea s-a realizat fără proteză la toţi pacienţii examinaţi. În mod similar s-a constatat o schimbare în pronunţia consoanei /f/ din silaba ”fa” la şase pacienţi. La articularea fără proteză au dispărut total frecvenţele mai mari de 3.5 kHz, iar la articularea cu proteză a apărut o creştere în zona frecvenţelor de peste 13 kHz. La pronunţarea cu proteză formanţii sunt mai clar delimitaţi şi frecvenţele în jur de 3.5 kHz sunt mai puţin frecvente. În cazul consoanei /r/ pronunţată fără proteză s-a constatat de asemenea schimbări în spectru la toţi pacienţii. În cazul câtorva pacienţi vocalele în special /i/ şi /u/ îşi schimbă frecvenţa fundamentală şi arată apariţia unui zgomot manifestat ca o delimitare neclară a formanţilor.

Autorii concluzionează că factorii cei mai importanţi în construirea danturii din punct de vedere fonetic sunt overbite, overjet, înălţimea palatului, grosimea materialului din

care este confecționat palatul, poziția incizorilor şi modelarea palatului dur. Acest studiu arată că silabele care conțin consoane alveolare frontale, labiodentale şi laterale /s/, /f/, /v/, /r/ şi vocale înalte /i/, /u/ sunt indicatori în evaluarea articulării sunetelor în stomatologie.

Un alt studiu referitor la evaluarea vocii pacienților trataţi cu proteză totală care poate fi scoasă a fost realizat de **F. Stelzle et al.** (F. Stelzle et al., 2010). Autorii au utilizat un ASR pentru discriminarea între un grup de 28 de pacienți şi un grup de control alcătuit din 40 de subiecți fără patologii de dentiție prin utilizarea parametrului WA:

$$WA = \frac{C - W}{R} \times 100, \tag{16}$$

unde C este numărul de cuvinte recunoscute corect, W numărul de cuvinte recunoscute greşit şi R numărul total de cuvinte analizate. S-au realizat două înregistrări per pacient cu proteză şi fără proteză, iar subiecții examinați au fost împărțiți în funcție de unele simptome specifice în 18 pacienți cu proteză completă suficientă şi 10 pacienți cu proteză completă insuficientă.

Rezultatele evaluării vorbirii de către ascultători specialişti şi cu ASR au arătat o corelație de 0.71. Acuratețea de recunoaştere a cuvintelor a fost semnificativ redusă la pacienții edentați (55.42%) comparativ cu grupul de control (69.79%). În urma acestui studiu s-a constatat că după pierderea completă a danturii calitatea vocii este semnificativ alterată, iar ASR s-a dovedit a fi un instrument util şi uşor de aplicat în evaluarea automată a vorbirii într-un mod standardizat.

5. Cazul SRoL (Sunetele Limbii Române) şi extensie în limba franceză

Site-ul web intitulat ”**Sunetele limbii române**” este o colecție care include mii de înregistrări de vocale, consoane, diftongi, propoziții cu voci emoționale şi dialectale şi o arhivă de sunete gnatofonice şi gnatosonice dezvoltată de (H.N. Teodorescu et al., 2005-2007). Înregistrările sunt adnotate şi documentate conform metodologiei şi protocoalelor concepute de autori. Site-ul include de asemenea documentație despre limba română, tehnologia vorbirii şi instrumente dezvoltate de echipa SRoL pentru analiză vocală. Resursele şi instrumentele sunt dedicate învățării virtuale a limbii române, aplicațiilor fonetice, de tehnologie a vorbirii, medicale şi de reabilitare a vorbirii şi fac parte din Rețeaua Europeană de resurse pentru limbaj CLARIN. SRoL este singura bază de date disponibilă pe internet, de voci emoționale în limba română care conține peste 1500 de înregistrări adnotate în diverse formate (.wav., .ogg, .txt) eşantionate cu 22 kHz şi precizie

de 16 sau 24 biți. Conform (S.M. Feraru et al., 2010) site-ul conține șase secțiuni principale:

- Pronunții standard de vocale susținute, consoane, diftongi, cuvinte şi propoziții scurte în limba română destinate învățării corecte a pronunției limbii române şi cercetării statistice în fonetică;
- Construcții sinctactice speciale cum ar fi propoziții cu subiect dublu şi apoziție destinate cercetării;
- Voci emoționale (M. Feraru şi H.N. Teodorescu, 2008);
- Comparație între vocea normală şi cea sintetică (H.N. Teodorescu şi M. Feraru, 2008);
- Pronunții dialectale;
- Microcorpus de sunete gnatofonice şi gnatosonice (H.N. Teodorescu şi M. Feraru, 2007), (H. N. Teodorescu şi A. Untu, 2010), (A. Untu şi H.N.Teodorescu, 2011).

Instrumentele dezvoltate de echipa SRoL se referă la extragerea de pattern-uri din semnalul vocal şi de calcul a traseelor frecvenței fundamentale şi a formanților F1, F2 şi F3 (M. Zbancioc, 2006). Pe lângă programe executabile autorii pun la dispoziție şi o descriere a acestora în vederea înțelegerii funcționării lor.

5.1. Microcorpus de sunete gnatofonice în limba română

Microcorpusul de sunete gnatofonice dezvoltat de echipa SRoL, prezentat în (H.N. Teodorescu şi M. Feraru, 2007), şi apoi dezvoltat în (H. N. Teodorescu şi A. Untu, 2010) şi (A. Untu şi H.N. Teodorescu, 2011) conține la ora actuală 29 înregistrări provenind de la 24 vorbitori din care 10 de gen feminin şi 14 de gen masculin. Pe SRoL sunt disponibile doar 19 din cele 29 înregistrări restul fiind protejate. Din setul total de vorbitori 16 sunt fără patologii de dentiție şi 8 cu diverse patologii. În continuare sunt detaliate statistica corpusului la momentul respectiv şi metodologiile utilizate şi prezentate în (H. N. Teodorescu şi A. Untu, 2010).

Metodologia de culegere a datelor

Metodologia, protocolul de înregistrare şi protocolul de documentare sunt cele prezentate pe sit-ul ”Sunetele limbii române”. Subiecții au fost informați anterior înregistrărilor de obiectivele proiectului, fiind asigurați de confidențialitatea datelor personale. Subiecții au semnat un consimțământ informat în conformitate cu protocolul de protecție a subiecților umani şi cu principiile etice ale cercetărilor care implică ființa umană

existente la nivel național şi internațional. Cercetarea pe subiecți umani privind analiza sunetelor gnatofonice şi gnatosonice a fost aprobată de către consiliul Facultății de Electronică, Telecomunicații şi Tehnologia Informației din cadrul Universității Tehnice ”Gheorghe Asachi” din Iaşi (H.N. Teodorescu et al., 2005-2007).

Metodologia de înregistrare

Sunetele gnatofonice au fost înregistrate cu ajutorul unui microfon prevăzut cu căşti, A4 Tech Stereo HS-60, având caracteristicile: frecvență de răspuns 20 Hz-20 kHz, impedanță: U=3V, R=1,5kΩ, sensibilitate: -58dB±2. Placa de bază a calculatorului pe care au fost efectuate înregistrările este Sony Vaio MBX-189 având încorporată o placă de sunet Intel® High Definition Audio compatible 3D audio (Direct Sound 3D support) cu următoarele caracteristici: procesor de semnal 44-kHz / 16-bit stereo CD quality, mod de ieşire a sunetului 8 canale, 192 kHz / 32 bit, standard Intel HD audio.

Pentru înregistrări s-a utilizat programul GoldWave™, versiunea 5.54, la o frecvență de eşantionare de 22050 Hz cu atributele PCM signed (16-24 bits mono). Culegerea de semnal vocal se realizează în condiții de zgomot redus (amplitudinea zgomotului trebuie să fie mai mică cu cel puțin 20 dB decât amplitudinea frecvenței fundamentale). Conform metodologiei de înregistrare de pe situl SRoL, se recomandă ca poziția microfonului să fie mai jos de gură, aproximativ în dreptul bărbiei (la câțiva centimetri de aceasta), iar distanța de la bărbie să fie aproximativ egală cu distanța până la buze (Teodorescu et al., 2005-2007).

Metodologia de documentare

Fişele subiecților ”Profil vorbitor”, ”Chestionar Patologie Vocală şi Factori Obiectivi” şi ”Fişă dinți subiect” au fost completate de către al doilea autor. Pentru “Fişă dinți subiect” s-au luat în considerare informații ce pot fi cunoscute de către subiectul înregistrat sau vizualizate de către persoana care completează fişa, de tip prezență / absență dinte, plombă, implant, coroană, punte. Am considerat elemente de interes şi plombele întrucât acestea dacă nu sunt realizate corect pot modifica structura dintelui având repercursiuni asupra spațiului interdentar, în consecință şi asupra vorbirii (spațierile dintre incisivi pot produce în vorbire efectul de “fonfăit”). Pentru a valida aceste idei este necesară o analiză statistică pertinentă (cel puțin 20 de subiecți). Un exemplu de fişă ce conține informații referitoare la patologiile arcadelor superioară şi inferioară este dat în

figura 1. Subiectul prezintă opt plombe simple la nivelul maxilarului şi opt plombe simple la nivelul mandibulei, care afectează primii molari de pe fiecare hemiarcadă.

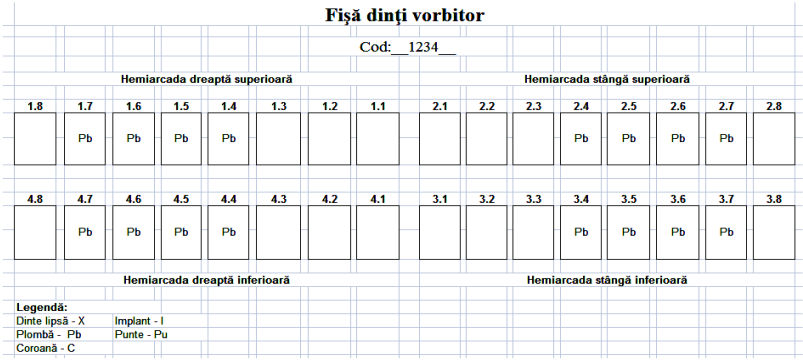

Figura 1: Exemplu de fişă dinţi subiect.

Listă cuvinte

Pentru realizarea corpusului de sunete gnatofonice s-a utilizat setul de cuvinte stabilit anterior de primul autor (©), cuvinte ce conţin consoanele *s*, *f*, *ş*, *v*, *z*, *j*. Analiza comparativă a pronunţiei consoanelor *s*, *f*, *ş*, şi a celor semi-vocalice *v*, *z*, *j*, furnizează informaţii despre afectarea dentiţiei.

Cuvintele utilizate pentru înregistrările gnatofonice sunt: vată / fată, var / far, vuiet (pronunţat vvvvvuiet) / vuiet (pronunţat normal, scurt, vuiet) / fffffui / fui, vvvvvaiet / vaiet / fffffaieton / faieton, vecin / fecior, vvvvvânt / vânt / fffffân / fân, vvvvvine / vvvvvine / vvvine / fffffine, vine / fine, vehement / ferment, vierme / fierbe, vâjâit / gâjâit / sâsâit / fâsâit, bâzâie / zâzâie, bâzzzzzâie / zâzzzzzâie, fâşâit / fâlfâit, vâjâie / fâlfâie / sâsâie / fâşâie, vâjjjjjâie / fâlfffffâie / sâsssssâie / fâşşşşşâie, vâjjjjjâit / sâssssâit / fâssssâit / fâşşşşşâit / fâlfffffâit, vorbit / fortuit / sortit, suit / vuit.

Scopul cercetării noastre este de a identifica şi analiza modificările ce apar în dinamica formanţilor / pseudo-formanţilor consoanelor fricative pronunţate de către două clase de vorbitori: o clasă martor formată din vorbitori cu dentiţie normală (neafectată de patologii ale aparatului stomatognat) şi o clasă de vorbitori cu defecte de dentiţie. Setul de cuvinte întocmit în acest scop include cuvinte ce conţin consoanele *v*, *f*, *s*, *ş*, *j*, *z* în context consoană-vocală (CV) sau vocală-consoană-vocală (VCV).

Pronunţiile celor 56 de cuvinte pentru fiecare din cei 10 vorbitori sunt înregistrate într-un singur fişier .wav care este salvat cu un nume mnemonic, în formatul ”cod_subiect_sex” (ex. 1234_m). Fişierele au fost filtrate şi re-salvate în format .wav şi în format .ogg (16 şi 24 biţi). Înregistrările au fost ascultate pentru o analiză perceptuală a zgomotului, iar pronunţiile cu trunchiere (saturaţie) în amplitudine au fost eliminate.

Discuția corpusului

Corpusul de sunete gnatofonice este disponibil pe sit-ul ”Sunetele limbii române”, în cadrul arhivei pentru aplicații de gnatofonie şi gnatosonie (Figura 2). În paralel am realizat şi un corpus de sunete gnatosonice pe care îl vom prezenta într-o lucrare ulterioară.

Arhiva pentru aplicații de gnatofonie şi gnatosonie

Metodologia de culegere a semnalelor gnatofonice este identică cu cea de culegere de semnal vocal.

Cuvintele utilizate pentru înregistrările gnatofonice sunt alese astfel încât să se poată analiza comparativ modificările de siflante, fricative şi de consoane semi-vocalice.

Cuvintele utilizate pentru înregistrările gnatofonice sunt: vată / fată; var / far; vuiet (pronunțat wwwuiet) / vuiet (pronunțat normal, scurt, vuiet) / fui / vaiet (pronunțat wwwaiet) / vaiet (pronunțat vaiet) / faieton / vecin / fecior / vânt (pronunțat wwwânt) / vânt (pronunțat vânt) / fân / wwwwine, wwine, wwine / vine / fine / vehement / ferment / vierme / fierbe / vâjât / wwwwâjjjjjât / wwâjjjjjâie / ffffâşşşşşşâie / ffffâşşşşşâit / fâşâit / sâsâit / sssssâsssssâie / gâjât / zâzâie / bââzzzzâââie / bâz&226;ie.

Un numar de 10 înregistrări gnatofonice provenite de la 5 subiecți de gen feminin şi 5 subiecți de gen masculin sunt adnotate, iar fişierele TextGrid şi wav corespunzătoare se pot vizualiza (accesa) la adresa (http://www.etc.tuiasi.ro /sibm/romanian_spoken_language/ro/sunete_gnatofonice.htm) sau aici.

Fişă vorbitor	Gnatofonie		Gnatofonie - filtrate		Gnatosonie		Gnatosonie - filtrate	
	wav	ogg	wav	ogg	wav	ogg	wav	ogg
Fişă vorbitor #394715	394715_f 394715_f_v1 394715_f_v2	394715_f 394715_f_v1 394715_f_v2	394715_f 394715_f_v1 394715_f_v2	394715_f 394715_f_v1 394715_f_v2				
Fişă vorbitor #231515	231515_m	231515_m	231515_m	231515_m				
Fişă vorbitor #280269	280269_m	280269_m	280269_m	280269_m				
Fişă vorbitor #20048					20048_f_v01	20048_f_v01	20048_f_v01	20048_f_v01
Fişă vorbitor #55555	55555_f 55555_f_v1	55555_f 55555_f_v1	55555_f 55555_f_v1	55555_f 55555_f_v1	55555_f	55555_f	55555_f	55555_f
Fişă vorbitor #1234	1234_m	1234_m	1234_m	1234_m	1234_m	1234_m	1234_m	1234_m
Fişă vorbitor #1357	1357_m	1357_m	1357_m	1357_m	1357_m	1357_m	1357_m	1357_m
Fişă vorbitor #2001	2001_f	2001_f	2001_f	2001_f	2001_f	2001_f	2001_f	2001_f

Figura 2: Arhiva pentru aplicații de gantofonie şi gnatosonie.

Statistica înregistrărilor şi patologiilor

Am realizat înregistrări de sunete gnatofonice provenind de la cinci subiecți de gen masculin şi cinci subiecți de gen feminin, cu vârste cuprinse între 21 şi 46 ani, majoritatea cu studii superioare, fără afecțiuni respiratorii, laringeale, neurologice, sau psihologice.

Corpusul realizat are la bază înregistrări provenind de la nouă subiecți fără defecte majore de dentiție şi fără deficiențe de vorbire detectabile prin percepție auditivă şi un subiect cu edentație majoră (13 molari lipsă). Starea dentiției subiecților înregistrați este ilustrată în Tabelul 2.

Tabel 2. Starea dentiției subiecților înregistrați

Cod subiect	Sex	Vârs-tă (ani)	Nr. Pb (plombă) / localizare		Nr. X (lipsă dinte) / localizare		Nr. C (coroană) / localizare		Nr. Pu (punte) / localizare		Tip voce / patologii
2404	M	30	1	primul incisiv HSS	-	-	-	-	-	-	Profesio-nală, fără patologii
2001	F	26	7	primul molar HDI, al 2-lea şi al 3-lea molar HDS şi HSS, al 2-lea molar HSI, al 3-lea molar HDI	-	-	1	al 2-lea molar HDI	-	-	Neprofe-sională, fără patologii
4312	F	29	3	primii incisivi HSS şi HDS, al 3-lea molar HDS	1	al 3-lea molar HDI	2	al 2-lea molar, HSS, al 3-lea molar HSI	-	-	Profesio-nală, fără patologii
1357	M	30	2	canin HSS, al 2-lea molar HSS	2	ultimul molar HDI şi HSI	2	al doilea molar HDI şi HSI	1	molarii 1, 3, HDS	Neprofe-sională, fără patologii
2202	F	26	3	al 2-lea molar HSS, al 3-lea molar HDS, al 4-lea molar HSI	1	al 3-lea molar HSI	-	-	-	-	Neprofe-sională, fără patologii
1234	M	30	16	primii 4 molari de pe fiecare hemiarcadă	-	-	-	-	-	-	Neprofe-sională, fără patologii
3371	F	21	4	primul incisiv HSS, al 2-lea molar HSS, al 4-lea molar HDI, al 3-lea molar HSI	2	al 3-lea molar HDI şi HSI	-	-	-	-	Neprofe-sională, fără patologii
3298	M	30	3	2 pe primii incisivi de pe HSS şi HDS, al 3-lea molar HDS	-	-	2	al 2-lea molar HDI şi HSI	-	-	Profesio-nală, fără patologii
01321	F	30	2	canin şi primul molar HDI	-	-	1	primul molar HSS	-	-	Neprofe-sională, fără patologii
5343	M	46	1	ultimul molar HSI	13	incisivii 1, 2 HSS, primii 3 molari HDS, ultimii 3 molari HSS, molarii 2,3,4 HSI, molarii 3, 4 HDI	-	-	-	-	Neprofe-sională, fără patologii

Metodologia de analiză

Metoda de prefiltrare

Toate fişierele au fost prefiltrate cu un filtru trece bandă cu frecvenţele de tăiere de cca. 70 Hz şi 7 kHz şi cu atenuare de 100 dB la frecvenţele de 50 Hz şi 12 kHz (în afara benzii de trecere), setat în utilitarul GoldWave^{TM} (Figura 3). Filtrul se poate aplica o dată sau, repetat, de două ori (echivalent, două filtre înseriate) astfel ca nivelul final de zgomot la 50 Hz să fie cu cel puţin 20 dB mai mic decât nivelul la frecvenţa fundamentală (F0). În cazul înregistrărilor curente, filtrul s-a aplicat o singură dată.

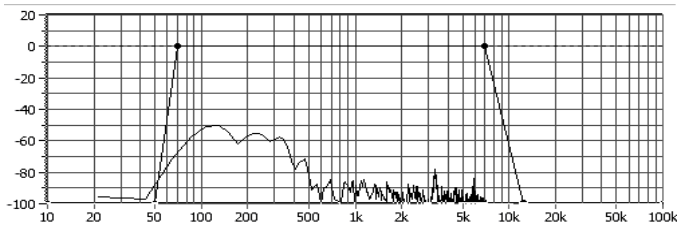

Figura 3: Filtru trece bandă de la 70 Hz la 7 kHz setat în utilitarul GoldWave™ .

Metodologia de adnotare

Pentru o analiză ulterioară a diferenţelor / similitudinilor ce apar la nivelul consoanelor fricative, am segmentat şi adnotat cele 10 fişiere la nivel de fonem / silabă / cuvânt cu ajutorul utilitarului Praat™. Am realizat împărţirea fişierelor în subfişiere ce conţin consoanele *f, s, ş, v, z, j*, în context CV şi VCV. Au rezultat câte nouă sau 11 subfişiere la fiecare subiect, numite astfel: vf_CV(a)_cod_subiect_sex – consoanele *v* şi *f* în context CV, V=vocala *a*, jsfshz_VCV(a-)_cod_subiect_sex – consoanele *j, s, f, ş, z* în context VCV, V=vocala *â*, vfs_CV(o)_cod_subiect_sex – consoanele *v, f, s* în context CV, V=vocala *o*, etc. Pentru *ş*, *â* şi *ă* am utilizat notaţiile *sh*, *a-* şi *a+*.

În cadrul procesului de adnotare am luat în considerare şi pauzele intravorbire (ce apar în rostirea de silabă, cuvânt) notate cu simbolul $, pauzele intervorbire (dintre cuvinte) fără notaţie (blanc) şi pauzele care nu se aud, notate cu simbolul %. Adnotarea s-a realizat ţinând cont de percepţia auditivă, prezenţa / absenţa frecvenţei fundamentale (ce evidenţiază caracterul vocalic, sonor, cu activarea corzilor vocale), forma de undă, energia şi modificările ce apar pe spectrogramă de la un fonem la altul.

Metodologia de analiză formantică şi temporală

Cu ajutorul utilitarului Praat™ am extras F0, formanţii / pseudo-formanţii F1, F2, F3, F4 şi timpii corespunzători fiecărui fonem şi fiecărui cuvânt în parte.

Analiza a constat din următoarele etape: i) Segmentarea, în vederea analizei modificărilor duratelor fonemelor. Segmentarea este efectuată la nivelurile fonem, silabă şi cuvânt, pentru a se determina momentele de timp corespunzătoare graniţelor dintre foneme şi cuvinte. ii) Generarea fişierelor .txt cu instrumentul Praat, cu valorile formanţilor în evoluţia lor; iii) Importarea fişierelor într-un utilitar (de ex., Excel™) şi separarea manuală a fonemelor în funcţie de limitele temporale corespunzătoare fiecărui fonem; iv) Analiza dinamicii şi analiza statistică a formanţilor pe segmentele de interes, comparativ între subiecţii cu afecţiuni şi tratamente ale aparatului stomatognat şi subiecţii sănătoşi. Am analizat dinamica formanţilor / pseudo-formanţilor consoanelor *v, f, s* din cuvintele *vată*,

fată, sâsâit, fâsâit, sâsssâit, fâsssâit. Acestea s-au aplicat la patru vorbitori (doi de gen feminin şi doi de gen masculin).

5.2. Microcorpus de sunete gnatofonice în limba franceză

Complementar microcorpusului de sunete gnatofonice în limba română am dezvoltat un microcorpus de sunete gnatofonice în limba franceză raportat în (Untu (Hulea), A., et al., 2011) care conține 51 de înregistrări de cuvinte care încep cu fricativele /v/ şi /f/ în context CV (consoană-vocală) provenind de la 17 vorbitori de gen feminin şi 34 vorbitori de gen masculin. Subiecții sunt studenți ai Institutului Universitar de Tehnologie din Angouleme şi au vârste cuprinse între 18 şi 24 ani, fără patologii de tip respirator, laringeal, neurologic sau psihologic care ar influența vocea. Microcorpusul este alcătuit din studenți cu dantură normală şi subiecți cu patologii dentare. Patru respectiv opt subiecți de gen feminin au malocluzie clasa I respectiv II şi 15 subiecți de gen masculin au malocluzie clasa I.

Pentru crearea microcorpusului în limba franceză am urmat aceaşi metodologie ca şi în cazul microcorpusului în limba română. Am creat o fişă de dentiție în limba franceză care include arcadele dentare, litere simbol pentru patologii dentare posibile şi ilustrarea tipurilor de malocluzie posibile. Studenții au marcat pe arcade în dreptul fiecărui dinte/molar şi pe pozele ilustrative tipul de patologie respectiv tipul de malocluzie prezentă. Câteva exemple de cuvinte din lista creată pentru limba franceză sunt: “varier” (a varia) / “fatiguer” (a obosi), “vol” (zbor) / “folle” (nebun).

Înregistrările le-am realizat cu sistemul de achiziție ”Harmonie” produs de ”01db” utilizând software-ul dBBATI32. Am utilizat un microfon de tip G.R.A.S. 40 AC, cu un răspuns în frecvență de 3.15 Hz - 45 kHz at ±2dB şi senzitivitate de 12.5 mV/Pa. Înregistrările au fost realizate în timpul unui stagiu de cercetare în cadrul Laboratorului de electroacustică şi acustică a mediului al IUT ANGOULEME.

6. Discuții şi concluzii

Corpusurile de semnal vocal sunt indispensabile în vederea dezvoltării şi testării instrumentelor de recunoaştere vocală, sinteză vocală, a sistemelor de recunoaştere automată de voce patologică, de detecție a patologiilor vocale induse de dentiție sau de evaluare a calității aplicării unui tratament în stomatologie.

Dezvoltarea unui corpus de semnal vocal presupune respectarea unei metodologii bine puse la punct şi adaptată în funcţie de aplicaţie. De o mare importanţă sunt parametrii utilizaţi la realizarea înregistrărilor şi anume frecvenţa de eşantionare şi precizia. Semnalul vocal se întinde pe o bandă de la câteva zeci de hertzi la 20 kHz. Deşi se consideră că limita superioară din semnalul vocal a intervalului în care este conţinută cea mai importantă informaţie este la 8 kHz, o eşantionare cu 16 kHz ar altera consoanele fricative care au un spectru bogat în frecvenţe înalte de peste 8 kHz. Conform teoremei Shannon care impune utilizarea unei frecvenţe de eşantionare de cel puţin de două ori mai mari decât frecvenţa maximă a semnalului, pentru înregistrări de calitate este necesară o frecvenţă de eşantionare de peste 20 kHz, fapt ce nu este respectat la zece din corpusurile prezentate şi sintetizate în Anexă cum ar fi (J.S. Gorofolo et. al., 1986), (K. Shobaki et al., 2007), (O. Salor et al., 2007) etc. De asemenea înregistrările trebuie filtrate pentru eliminarea zgomotelor introduse de aparatură, alimentare, cum ar fi zgomotul de 50 Hz. În nici unul din materialele referitoare la corpusurile prezentate cu excepţia SRoL nu este specificat dacă înregistrările au fost filtrate. Din punct de vedere statistic, corpusurile trebuie să fie compuse dintr-un număr cât mai mare de vorbitori de ambele genuri şi un set diversificat de foneme, cuvinte, propoziţii. Corpusuri limitate din punct de vedere al numărului de vorbitori sunt (A. Stan et al. 2011) şi (J. Matousek şi J. Romportl, 2007) care au un singur vorbitor. Deşi aplicaţia de bază este sinteză de voce, sistemele dezvoltate ar trebui testate pe mai multe voci.

Informaţiile demografice cu privire la subiecţii înregistraţi şi anamneza în cazul constituirii corpusurilor de voci patologice sunt de asemeni importante. Un îndeajuns a câtorva corpusuri şi instrumente dezvoltate din Anexa este acela că în articolele aferente nu sunt specificate întotdeauna parametrii de înregistrare (frecvenţă de eşantionare, precizie), număr de vorbitori, informaţii despre vorbitori, adnotare şi limba utilizată cum ar fi în (J. F. Bonastre et al., 2007), (J. B. Tomblin, 2010). În cazul dezvoltării unui corpus de voci patologice este necesară dezvoltarea complementară a unui corpus de voci normale de control pentru utilizarea la sisteme de clasificare de voce sau detecţie de voce patologică. (F. Bocklet et al., 2010) şi (K. T. Bressmann et al., 2011) au evaluat articularea sunetelor la pacienţi edentaţi înainte de tratament şi după tratament fără a compara cu persoane fără patologii de dentiţie.

La ora actuală nu există un corpus de referinţă statistic valabil cu voci provenite de la pacienţi cu diverse patologii de dentiţie care să poată fi utilizat de specialişti în domeniu în dezvoltarea de sisteme de diagnostic automat, de evaluare a vocii înainte, în timpul şi după

tratament sau de detectare a protezelor incorect fixate sau nepotrivite. Câteva limitări ale SRoL (Teodorescu et al. 2005-2007) sunt un număr redus de vorbitori în special la înregistrările de vocale susținute şi consoane, necesitatea utilizării unei frecvențe de eşantionare mai mari în cazul sunetelor gnatofonice pentru a nu tăia componentele spectrale ale consoanelor fricative de interes şi un număr insuficient de înregistrări cu voci patologice. Extinderea corpusului de vocale susținute ar lărgi aplicabilitatea site-ului în domeniul testării sistemelor de recunoaştere automată a vocilor disfonice, putând fi utilizat ca material de control.

7. Direcții viitoare

Ca direcții viitoare ne propunem dezvoltarea microcorpusului de sunete gnatofonice în limba română atât cu înregistrări de voci normale cât şi cu înregistrări de voci patologice induse de defectelor de dentiție, în vederea constituirii unei baze de cunoştințe care să înglobeze caracteristicile acusto-fonetice ale sunetelor gnatofonice.

Alte scopuri sunt identificarea de trăsături specifice vocilor patologice utile în dezvoltarea de sisteme de clasificare de voce, design-ul teoretic şi dezvoltarea unui sistem de detectare automată a patologiilor de dentiție sau de evaluare a vocii persoanelor protezate.

Autori	Nr. Vb.	Limba	Material text	Adnotare/ Transcripție	Frecv. eşant.	Precizie	Tip înregistrări	Instrum.	Aplicații	Comercial
J.S. Gorofolo et. al., 1986-TIMIT	630	6 dialecte lb. engleză	10 propoziții citite/subiect	DA	16 kHz	16 biți	microfon	---	-recunoaştere vocală; studii acustico-fonetice	DA
J. Wright, 2005	---	lb. engleză	Silabe-2000 comune, 20 unice/subiect	---	16 kHz	---	microfon	---	-modelare pronunție;identificarea, modelarea şi procesarea de limbă	DA
K. Shobaki et al., 2007	1100	lb. engleză copii	Vorbire spontană şi citită	---	16 kHz	16 biți	microfon	---	-recunoaştere vocală, dezvoltarea limbii copiilor fără auz, învățării limbii	DA
M. Liberman et al., 2002	---	lb. engleză, actori	Enunțuri neutre, 14 emoții	DA	22 kHz	---	microfon	---	-recunoaştere vocală, modelare pronunție, prozodie	DA
Jankowski et al., 1990 - NTIMIT	630	Lb. engleză	Text TIMIT	DA	---	---	telefon	---	Recunoaştere vocală	DA
P. Kingsburg	---	Lb.	90988	---	---	---	telefon	---	Recunoaştere vocală	DA

et al., 1997		engleză	cuvinte de jurnal							
K. Karins et al., 1997	---	Lb. germană	318807 cuvinte	---	---	---	telefon	---	Recunoaştere vocală	DA
S. Garrett et al., 1996	---	Lb. spaniolă	45582 cuvinte	---	---	---	telefon	---	Recunoaştere vocală	DA
J. Kong şi D. Graff, 2005	---	Lb. engleză, arabă, mandarină	ştiri	---	16 kHz / 2 canale	---	radio	DA	Detectare subiect comun în ştiri, detectare perechi de ştiri cu subiect comun etc.	DA
D. Graff, 2001a	---	Lb engleză	ştiri	---	16 kHz/1 canal	16 biți	radio	DA	Detectare subiect comun în ştiri, detectare perechi de ştiri cu subiect comun etc.	DA
D. Graff, 2001b	---	Lb. mandarină	ştiri	---	16 kHz/1 canal	16 biți	radio	DA	Detectare subiect comun în ştiri, detectare perechi de ştiri cu subiect comun	DA

									etc.	
J.J. Gofrey, 1994a,b,c,d	---		Comunicare control aerian	---	8 kHz	---	Transmisie aeriană	---	Recunoaştere vocală în medii zgomotoase	DA
J. Fiscus et al., 2007a,b	---	---	Discursuri în cadrul întâlnirilor	---	---	---	microfon	---	-analiză discurs, recunoaştere vocală, extragere metadate	DA
NIST Multimodal Information Group,2011a,b	---	---	ştiri, convorbiri telefonice, comunicări	---	16 kHz-ştiri radio; 8 kHz/2 canale-telefon	---	Radio, Telefon, microfon	---	-detectare subiect, recunoaştere vocală	DA
O. Salor et al., 2007	193	Lb. turcă	Traducere 2000 propoziţii TIMIT	---	16 kHz	---	microfon	DA	Recunoaştere vocală	DA
M. Boldea et	100	Lb.	Traducere 40	DA	20 kHz	16 biţi	microfon	NU	Dezvoltarea de	---

al., 1998		română	pasaje lb. engleză corpus EUROM-1, 550 propoziţii						instrumente de recunoaştere vocală automată independentă de vorbitor	
A. Stan et al. 2011	1	Lb. română	3800 propoziţii din ziare, poveşti	DA	96 kHz/ 44 kHz	24 biţi	microfon	DA	Sinteza voce	---
J. Matousek şi J. Romportl, 2007	1	Lb. cehă	5000 pasaje ziare	DA	-	-	microfon	NU	Sinteza voce	---
C.G. Clopper şi D.B. Pisoni, 2006	60	6 dialecte lb. engleză	Cuvinte, propoziţii, pasaje	DA	44 kHz	16 biţi	microfon	DA	Determinare dialecte	---
A. Batliner et al., 2005	611	Lb. engleză, suedeză, germană, italiană	220 cuvinte izolate, 50 propoziţii engleză, 400 propoziţii	DA	44 kHz	16 biţi	microfon	NU	Recunoaşterea automată a vocii spontane şi emoţionale de copii	---

		copii	generice, voce spontană, citită							
N. Kawaguchi et al., 2005	812	---	1 milion morfeme	DA	---	---	microfon, video	DA	Analiza vocii în timpul conducerii de autovehicule	---
Kay Elemetrics Corp, 1994	53 N, 657 P	Lb. engleză	Vocala susținută /a/	---	25 kHz/50 kHz	16 biți	microfon	---	Detecție voce patologică	DA
G. Nuffelen et al.,	122 N, 197 P	Lb. germană	50 cuvinte context CVC, vocala susținută /a/, vorbire spontană și semispontană	DA	16 kHz	16 biți	microfon	NU	Instrumente de evaluare a vocii patologice, teste de inteligibilitate	---
J. F. Bonastre et al., 2007	58 N,	---	Vocala susținută /a/	---	---	---	microfon	DA	Evaluare voce patologică	---

	391 P									
J. B. Tomblin, 2010	2009	Lb. engleză, copii	---	---	---	---	microfon	DA	Evaluare voce patologică şi detecție defecte de limbaj	---
G. Schlottauer şi M.E. Torres, 2006	53 N, 36 P	---	Vocala /a/ susținută	---	22 kHz	16 biti	microfon	DA	Clasificarea automată a vocii disfonice	---
F. Pozo et al., 2009	40 N, 40 P	Lb. spaniolă	Patru propoziții	---	48 kHz	---	microfon	DA	Detectarea automată a apneei obstructive	---
L. Salhi et al., 2010	30 N, 30 P	---	cuvinte	---	---	---	microfon	DA	Detecție voce patologică	---
J. Lee şi M. Hahn, 2009	83 P	---	Vocala susținută /a/	---	---	---	---	---	Detecție voce patologică	---
F. Bocklet et al., 2010	13 P	Lb. germană	Text standard din 108 cuvinte	---	16 kHz	16 biți	microfon	DA	Determinarea automată a eficienței tratamentului	---
K. T.	22 P	---	35 propoziții	---	---	---	microfon	NU	Determinarea	---

Bressmann et al., 2011									modificării vocii după aplicarea de expanderi	
P. Jindra et al., 2002	10 P	---	Silabe care conţin consoane bilabial, alveolare, labiodentale	---	---	---	microfon	DA	Detectarea aplicării corecte de proteză parţială şi completă	---
F. Stelzle et al., 2010	40 N, 28 P	---	---	---	---	---	microfon	DA	Recunoaştere automată a vocii patologice	---
H.N. Teodorescu şi S.M. Feraru, 2007; H.N. Teodorescu şi A. Untu, 2010; A. Untu, H.N. Teodorescu,	16 N, 8P	Lb. română	Cuvinte care conţin consoane fricative	DA	22 kHz	16 biţi	microfon	DA	Diagnosticarea automată a vocilor patologice de cauză stomatologică	DA

2011										
A. Untu et al., 2011	24 N, 27 P	Lb. franceză	Cuvinte care conțin consoane fricative	DA	22 kHz	16 biți	microfon	DA	Diagnosticarea automată a vocilor patologice de cauză stomatologică	NU

--- = nespecificat

Referințe

1. Batliner, A., Blomberg, M., D'Arcy, S., Elenius, D., Giuliani, D., Gerosa, M., Hacker, C., Russell, M., Steidl, S. and Wong, M., *The PF-STAR Children's Speech Corpus*, in Proc. Eurospeech, pp. 2761–2764, 2005.
2. Bocklet T., Hönig, F., Haderlein, T., Stelzle, F., Knipfer C. and Nöth E., *Automatic Detection and Evaluation of Edentulous Speakers with Insufficient Dentures*, Lecture Notes in Computer Science, , vol. 6231, pp. 243-250, 2010.
3. Boldea, M., Munteanu, C. and Doroga, A., *Design, Collection, and Annotation of a Romanian Speech Database*, In Proceedings LREC Workshop on Speech Database Development for Central and Eastern European Languages, Granada, Spain, 1998.
4. Bonastre, J.F., Fredouille, C., Ghio, A., Giovanni, A., Pouchoulin, Revis, G. J., Teston, B. and Yu, P., *Complementary Approaches for Voice Disorder Assessment*, in Proc. Interspeech, pp. 1194–1197, 2007.
5. Bressmann, K.T., Gong, S. and Tompson, B.D., *Impact of a Rapid Palatal Expander on Speech Articulation*, American Journal of Orthodontics and Dentofacial Orthopedics, vol. 140, nr. 2, pp. 67-75, 2011.
6. Clopper, C.G., Pisoni, D.B., *The Nationwide Speech Project: A new corpus of American English dialects*, Speech Communication., vol.48, pp. 633–644, 2006.
7. Dibazar, A.A. and Narayanan, S.S., *A System for Automatic Detection of Pathological Speech*, in 36th Asilomar Conf. Signal, Systems, and Computers, Asilomar, CA, USA, 2002.
8. Feraru S.M., Teodorescu H.N., Zbancioc M.D., *SRoL-Web-based Resources for Languages and Language Technology e-Learning*, International Journal of Computers Communications & Control, vol. 5, nr. 3, pp. 301-313, 2010.
9. Feraru, M., Teodorescu, H.N., *The Emotional Speech Section of the Romanian Spoken Language, Archive*, Conf. On Intelligent Systems and Technologies, Proc. 5th European, Iasi, Romania, 2008.
10. Fernández-Pozo R, Murillo J.L.B., Gómez L.H., Gonzalo E.L., Ramírez J.A., Toledano D.T., *Assessment of Severe Apnoea through Voice Analysis, Automatic Speech, and Speaker Recognition Techniques*, EURASIP Journal on Advances in Signal Processing - Special issue on recent advances in biometric systems: a signal processing perspective, 2009.
11. Fiscus, J. et al., *2004 Spring NIST Rich Transcription (RT-04S) Development Data*, Linguistic Data Consortium, Philadelphia, 2007a.
12. Fiscus, J. et al., *2004 Spring NIST Rich Transcription (RT-04S) Evaluation Data*, Linguistic Data Consortium, Philadelphia, 2007b.
13. Garofolo, J. S., Lamel, L. F., Fisher, W. M., Fiscus, J. G., Pallett, D. S. and Dahlgren N. L., *The DARPA TIMIT Acoustic-Phonetic Continuous Speech Corpus CDROM*, NIST, 1986 [www.ldc.upenn.edu/lol/docs/TIMIT.html].
14. Garrett, S., Morton, T. and McLemore, C., *CALLHOME Spanish Lexicon*, Linguistic Data Consortium, Philadelphia, 1996.
15. Godfrey, J. J., *Air Traffic Control BOS*, Linguistic Data Consortium, Philadelphia, 1994a.
16. Godfrey, J. J., *Air Traffic Control Complete*, Linguistic Data Consortium, Philadelphia, 1994b.
17. Godfrey, J. J., *Air Traffic Control DCA*, Linguistic Data Consortium, Philadelphia, 1994c.

18. Godfrey, J. J., *Air Traffic Control DFW*, Linguistic Data Consortium, Philadelphia, 1994d.
19. Graff, D., *TDT3 English Audio*, Linguistic Data Consortium, Philadelphia, 2001a.
20. Graff, D., *TDT3 Mandarin Audio*, Linguistic Data Consortium, Philadelphia, 2001b.
21. Han, N., Graff, D. and Kim, M., *Korean Telephone Conversations Lexicon*, Linguistic Data Consortium, Philadelphia, 2003.
22. http://www.inl.nl/tst-centrale/images/stories/producten/documentatie/copas_manual.pdf
23. http://www.clarin.eu/
24. http://www.elra.info/
25. http://www.ldc.upenn.edu/
26. Huang, S., Bian, X., Wu, G. and McLemore, C., *CALLHOME Mandarin Chinese Lexicon*, Linguistic Data Consortium, Philadelphia, 1996.
27. Jankowski et al., *NTIMIT: A Phonetically Balanced, Continuous Speech, Telephone Bandwidth Speech Database*, Proc. ICASSP-90, April 1990.
28. Karins, K. et al., *CALLHOME German Lexicon*, Linguistic Data Consortium, Philadelphia, 1997.
29. Kawaguchi, N., Matsubara, S., Takeda, K. and Itakura, F., *CIAIR In-Car Speech Corpus -Influence of Driving Status*, in IEICE Transactions on Information and Systems, vol. E88-D, nr. 3, pp. 578–582, 2005.
30. Kay Elemetrics Corp, *Disordered Voice Database*, Version 1.03, 1994.
31. Kilany, et al., *Egyptian Colloquial Arabic Lexicon*, Linguistic Data Consortium, Philadelphia, 2002.
32. Kingsbury, P. et al., *CALLHOME American English Lexicon (PRONLEX)*, Linguistic Data Consortium, Philadelphia, 1997.
33. Kobayashi, M., Crist, S., Kaneko, M. and McLemore, C., *CALLHOME Japanese Lexicon*, Linguistic Data Consortium, Philadelphia, 1996.
34. Kong, J. and Graff, D., *TDT4 Multilingual Broadcast News Speech Corpus*, Linguistic Data Consortium, Philadelphia, 2005.
35. Lee, J., Hahn, M., *Automatic Assessment of Pathological Voice Quality Using Higher-Order Statistics in the LPC Residual Domain*, EURASIP Journal on Advances in Signal Processing, vol. 2009, 2009.
36. Liberman, M. et al., *Emotional Prosody Speech and Transcripts*, Linguistic Data Consortium, Philadelphia, 2002.
37. Matousek, J., Romportl, J., *Recording and Annotation of Speech Corpus for Czech Unit Selection Speech Synthesis*, Lecture Notes in Artificial Intelligence, vol. 4629, Springer, Berlin-Heidelberg, pp. 326–333, 2007.
38. Moran, R. J., Reilly, R. B., de Chazal, P., and Lacy, P. D., *Telephony-based Voice Pathology Assessment using Automated Speech Analysis*, IEEE Trans. Biomed. Eng., vol. 53, nr. 3, pp. 468–477, 2006.
39. NIST Multimodal Information Group, *2006 NIST Spoken Term Detection Evaluation Set*, Linguistic Data Consortium, Philadelphia, 2011.
40. NIST Multimodal Information Group, 2011, *2006 NIST Spoken Term Detection Development Set*, Linguistic Data Consortium, Philadelphia
41. P. Jindra, M. Eber, J. Pešák, *The Spectral Analysis of Syllables in Patients using Dentures*, Biomed. Papers, vol. 146, nr. 2, pp. 91–94, 2002.
42. Saenz-Lechon, N., Godino-Llorente, J.I., Osma-Ruiz, V., Gomez-Vilda, P., *Methodological Issues in the Development of Automatic Systems for Voice Pathology Detection*, Biomed. Signal Process. Control 1, pp.120–128, 2006.

43. Salhi, L., Mourad, T. and Cherif, A., *Voice Disorders Identification Using Multilayer Neural Network*, The International Arab Journal of Information Technology, vol. 7, nr. 2, 2010.
44. Salor, O., Pellom, B. L., Iloglu, T., Demirekler, C. M., *Turkish Speech Corpora and Recognition Tools Developed by Porting Sonic: Towards Multilingual Speech Recognition*, Comput. Speech. Lang., vol. 21, nr. 4, pp. 580–593, 2007.
45. Schlotthauer, G., Torres, M.E., *Automatic Diagnosis of Pathological Voices*, Proceeding of the 6th WSEAS International Conference on Signal, Speech and Image Processing, Lisbon, Portugal, pp. 150-155, 2006.
46. Shobaki, K., Hosom, J. P. and Cole, R., *CSLU: Kids` Speech Version 1.1*, Linguistic Data Consortium, Philadelphia, 2007.
47. Stan, A., Yamagishi, J., King, S., and Aylett, M., *The Romanian Speech Synthesis (RSS) Corpus: Building a High Quality HMM-based Speech Synthesis System using a High Sampling Rate*, Speech Communication, vol. 53, pp. 442–450, 2011.
48. Stelzle, F., Ugrinovic, B., Knipfer, C., Bocklet, T., Noth, E., Schuster, M., Eitner, S., Seiss, M. şi Nkenke, E., *Automatic, Computer-Based Speech Assessment on Edentulous Patients with and without Complete Dentures – Preliminary Results*, Journal of Oral Rehabilitation, vol. 37, pp. 209–216, 2010.
49. Teodorescu, H.N., Feraru, S.M., Trandabăț, D., Zbancioc, M., Luca, R., Verbuță, A., Hnatiuc, M., Ganea, R., Voroneanu, O., Pistol, L., Șcheianu, D., *Situl Web Sunetele Limbii Române*, http://www.etc.tuiasi.ro/sibm/romanian_spoken_language/index.html, 2005-2007
50. Teodorescu, H.N., Feraru, M., *De ce nu Place Vocea Sintetizată - Câteva Elemente de Comparație cu Vocea Umană*, Trandabăț, Cristea, Tufiş (eds.), ConsILR 2008, Resurse lingvistice şi instrumente pentru prelucrarea limbii române, Editura Universității "Al. I. Cuza", 19-21 noiembrie, pp. 21-30, Iaşi, România, 2008.
51. Teodorescu, H.N., Feraru, M., *Micro-corpus de Sunete Gnatosonice si Gnatofonice*, Pistol, Cristea, Tufis (Eds.), Resurse lingvistice si instrumente pentru prelucrarea limbii romane, Ed. Universitatii "Al.I.Cuza" Iasi, pp.21-30, 2007.
52. Teodorescu, H.N., **Untu, A**., *Corpus pentru Gnatofonie: Protocol, Metodologie, Adnotare*, ConsILR 2010, Bucureşti, 6-7 mai, Editori: Adrian Iftene, Horia-Nicolai Teodorescu, Dan Cristea, Dan Tufiş, Editura Univesității "Al. I. Cuza" Iaşi, pp. 13-22, 2010.
53. Tomblin, J. B., *The EpiSLI Database: A Publicly Available Database on Speech and Language*, Language, Speech, and Hearing Services in Schools, vol. 41, nr. 1, pp. 108-117, 2010.
54. **Untu (Hulea), A**., Duvanaud, C., Teodorescu, H.N., *A Study of the Relationship between the Statistical Acoustical Features of the /v/ and /f/ French Fricative Consonants and the Dentistry Pathologies*, (acceptată, IEEE International Conference on E-Health and Bioengineering, 24-26 nov., 2011, Iasi, Romania).
55. **Untu, A**., Teodorescu, H.N., *Pattern Analysis of /v/ Pronunciations in _/v/V Contexts*, The Eighth IASTED Int. Conf. on Signal Processing, Pattern Recognition, and Applications SPPRA 2011, Innsbruck, 2011.
56. Wright, J., *Articulation Index*, Linguistic Data Consortium, Philadelphia, 2005.
57. Zbancioc, M., *Tools for the Archive of the Romanian Language Sounds Project*, 4th European Conf. On Intelligent Systems and Technologies, Iasi, Romania, 2006.