

ASPECTE METODOLOGICE DE ORGANIZARE A DATELOR ȘI DE ANALIZĂ STATISTICĂ A VOCILOR EMOȚIONALE

HORIA-NICOLAI TEODORESCU^{1, 2}, IOAN PĂVĂLOI¹, MONICA FERARU¹

¹*Institutul de Informatică Teoretică al Academiei Române,*

Filiala Iași a Academiei Române

²*Universitatea Tehnică Gheorghe Asachi din Iași*

{hteodor}@etc.tuiasi.ro

Rezumat

Prezentăm o metodologie și un program pentru analiza statistică a vocilor emoționale. Prezentăm în principal unele rezultate preliminare privind modul de organizare a datelor în cadrul unei aplicații de analiză a adnotărilor pentru fișierele de semnal vocal. Programul permite analiza statistică a caracteristicilor formantice ale vocalelor și semivocalelor pe subclase de fișiere selectate conform unor caracteristici date de utilizator. Aplicația permite un grad de rafinare a analizei emotivității în voce la nivelul sunetelor vocalice.

1. Introducere

Analiza expresivității emotive (Teodorescu & Feraru, 2007), (Teodorescu & Feraru, 2009) stabilirea unor metodologii teoretice și a unor algoritmi pentru identificarea stărilor emoționale (Ververidis & Kotropoulos, 2006), (Nakatsu et al., 1999), (Mcgilloway et al., 2000) au numeroase aplicații practice (Scherer, 2000), (Kienast & Sendlmeier, 2008) și științifice în domeniul informaticii, lingvisticii, psihologiei etc. Aceste analize sunt însă laborioase și necesită resurse considerabile, atunci când privesc corpusuri mari de înregistrări de vorbire. În această lucrare prezentăm realizarea unui program care efectuează o statistică generală a formanților, cu aplicație la unele vocalele („a”, „e”, „i”, „u”, „ă”) din limba română cu selecția fișierelor de voce în funcție de starea emoțională.

Scopul cercetării raportate a fost stabilirea metodologiei de organizare și de procesare a datelor în vederea prelucrării statistice. Aceasta permite flexibilitate în accesarea informațiilor, validarea datelor existente, rapiditate în prelucrarea datelor, precum și dezvoltarea de noi aplicații în domeniul procesării limbajului vorbit. Tratarea unitară a volumului mare de date ne permite o abordare eficientă în vederea realizării unor prelucrări statistice complexe. Lucrarea este o continuare a cercetărilor autorilor privitor la analiza și caracterizarea emotivității în voce. Lucrarea este structurată după cum urmează: secțiunea a doua și a treia sunt dedicate metodologiei de organizare și de procesare a datelor în aplicația dezvoltată de noi (și numită WinCollection); în secțiunea a patra sunt prezentate rezultatele preliminare obținute, iar în ultima secțiune sunt descrise concluziile și direcțiile viitoare.

2. Metodologia de organizare a datelor

2.1 Datele primare

Propozițiile înregistrate sunt: „Vine mama”, „Cine a făcut asta”, „Ai venit iar la mine” și „Aseară”. Stările emoționale supuse analizei sunt: bucurie, tristețe, furie și tonul neutru. Exemplificăm analiza pentru un număr de cinci vorbitori (2 feminini și 3 masculini) pentru care avem 63 de înregistrări cu o medie de trei rostiri pe fiecare înregistrare. Există trei înregistrări masculine (bucurie, tristețe și neutru) care au un grad de exprimare a emoției redus comparativ cu ceilalți vorbitori. Toate înregistrările fac parte din corpusul adnotat și documentat SRoL (Teodorescu et al., 2005) și au fost realizate cu o frecvență de eșantionare de 22kHz, cu o rezoluție de 16 biți.

Aplicația este reprezentată de un program realizat în Visual C++. Datele de intrare sunt conținute în patru fișiere: un fișier .wav (înregistrarea propriu-zisă) și trei fișiere .txt (un fișier de adnotare TextGrid și două fișiere cu valori ale frecvenței fundamentale, F0 și ale formanților F1-F3). Adnotarea fișierelor s-a realizat cu utilitarul Praat (Boersma & Weenink, 2006), la nivel de propoziție, cuvânt, silabă și fonem. În procesul adnotării am definit următoarele tipuri de pauze: i) pauzele intravorbire (în rostirea de silabă, cuvânt), care au fost marcate cu simbolul „\$”; ii) pauzele intervorbire (între rostiri de propoziții) care au fost codate prin blank; iii) pauzele care nu sunt percepute și care sunt detectate la nivelul unei analize mai fine; acestea au fost notate cu simbolul „%”.

Atât în fișierele Praat cât și în fișierele de intrare ale programului am codat: â cu a-, ă cu a+, ș cu sh și ț cu tz.

Fișierele text cu valori instantanee ale formanților au fost obținute cu ferestre alunecătoare a căror dimensiune este de 0.025s, iar pasul ferestrei este de 0.01s. Fișierele rulate de program au o durată de maxim 10s, ceea ce corespunde în medie la trei rostiri a propoziției înregistrate. Programul a rulat doar peste fișierele date de utilitarul Praat, dar poate prelucra și alte tipuri de fișiere.

În figura 1, exemplificăm rezultatul rostirii pentru vorbitorul 11861, pentru care avem testate 3 stări emoționale: bucurie, furie și ton neutru. Frazele ale căror pronunții sunt supuse analizei sunt: „Ai venit iar la mine”, „Cine a făcut asta”, „Vine mama” și „Aseară”, iar fonemele analizate sunt „a”, „e”, „i”, „e”, „ă”.

```

speaker : 11861
Emotion : bucurie
Phrase : ai venit iar la mine
          3 rostiri
Phrase : aseara
          3 rostiri
Phrase : cine a facut asta
          3 rostiri
Phrase : vine mama
          3 rostiri
Emotion : furie
Phrase : ai venit iar la mine
          3 rostiri
Phrase : aseara
          3 rostiri
Phrase : cine a facut asta
          3 rostiri
Phrase : vine mama
          3 rostiri
    
```

Figura 1: Screenshot referitor la sinteza analizei pentru vorbitorul 11861

2.2 Preprocesare de fișiere

Valorile lui F0 și F1-F4 au fost calculate automat folosind utilitarul Praat și salvate în fișierele corespunzătoare vorbitorului, emoției și propoziției respective. Există unele situații în care F0 apare ca fiind nedefinit. Aceste situații sunt următoarele:

- între rostiri - nu există F0 și nici F1-F4. În tabelul 1 exemplificăm un segment de valori ale lui F0 care este nedefinit și altul definit pentru propoziția „Ai venit iar la mine”, diftong „ai”, prima rostire. Primele valori nedefinite sunt cele care corespund valorilor inexistente din intervalul dinaintea rostirii. Programul ignoră la prelucrare zonele în care valorile sunt nedefinite de utilitarul Praat.
- zone unde nu este sunet vocalic;
- zone unde este sunet vocalic, dar utilitarul Praat nu detectează F0 întotdeauna – eroare de detecție;
- în cazul în care pe un sunet vocalic se obțin cu utilitarul Praat valori aberante – mult mai mică sau mai mare decât o valoare în limite normale (de exemplu. 70-500Hz, la bărbați).

Validarea se realizează prin parcurgerea tuturor datelor de intrare din fișierele de adnotare TextGrid create de Praat, iar pe baza unor criterii, sunt eliminate valori eronate furnizate de același utilitar. Fiecărui fișier TextGrid (fișier de adnotare asociat unei înregistrări) îi corespunde un fișier cu valorile formantului F0 precum și un fișier cu valorile formanților F1-F4. Corelarea fișierelor se face în mod automat, absența unuia dintre ele fiind semnalată ca eroare. Aceste criterii sunt stabilite de către utilizator și ele pot fi modificate luând în considerare intervalele de variație ale fundamentalei și a formanților. Segmentele existente în fișierul de adnotare cu markerii respectivi de timp permit validarea sau invalidarea existenței formanților. De exemplu, dacă o pauză este corespunzătoare unui segment între t1 și t2 (markeri de timp), în fișierul de formanți pe același segment se invalidează F0 și formanții F1-F3.

Adăugarea de noi informații se realizează prin reprocesarea datelor. Fișierul binar conține informații despre data creării, versiunea programului WinCollection utilizat, versiunea fișierului binar, despre persoana care a creat colecția, precum și toate informațiile primare. Un alt avantaj al organizării datelor îl constituie posibilitatea reproductibilității rezultatelor obținute.

Cumularea informațiilor într-un singur fișier asigură flexibilitate în accesarea lor, selectarea rapidă a lor pe baza unor criterii de intrare stabilite de utilizator. Nu există date primare, chiar și valorile eronate furnizate de Praat, care să nu se regăsească în colecție.

Structurarea ne permite analize statistice complexe pe un set de date provenit dintr-o selecție a lor și obținerea rapidă a informațiilor asupra variației parametrilor analizați. În această lucrare raportăm numai statistici pentru vocale. Sunetele semivocalice (despre care se știe că au F0 și formanți superiori) nu sunt luate în considerare în rulările pentru această lucrare, rulări în care s-au setat ca zone de interes doar vocalele.

3. Metodologia de procesare a datelor

Programul permite realizarea unei „statistici tomografice”, în sensul că permite determinarea unor „secțiuni de interes” în statisticile realizate. Programul asigură flexibilitate pentru o analiză statistică pe orizontală cât și pe verticală. De exemplu, programul permite analiza unui singur parametru (F0) pentru toți vorbitorii și pentru toate stările emoționale, sau analiza rezultatelor obținute de la un vorbitor în comparație cu toți ceilalți vorbitori. Analiza este organizată la nivelul stărilor emoționale, la nivelul vocalelor și la nivelul valorilor formanților din cadrul vocalelor.

Programul WinCollection are ca date de intrare fișierele .wav, fișierele de adnotare TextGrid, fișierele cu valorile formanților F0 și F1-F3, fișierele Codes.txt, Phrases.txt și Emotions.txt. Fișierul Codes.txt conține date despre toți vorbitorii ale căror înregistrări trebuie să fie în directorul respectiv. Fișierul Phrases.txt este necesar pentru a stoca frazele permise în fișierele de adnotare. Fișierul e utilizat în depistarea erorilor de adnotare. Prin compararea stringurilor din fișierul Phrases.txt cu stringurile corespunzătoare adnotărilor, sunt depistate lipsa de litere, inversiuni de litere sau alte erori de acest tip. F0 din figura 2 notează toate fișierele generate de Praat conținând valorile lui F0 pentru fișierele .wav din directorul respectiv. Similar, fișierele F1-F3 din figura 2 desemnează ansamblul fișierelor generate de Praat conținând acești formanți. Fișierul Emotions.txt include stringul care desemnează emoțiile din înregistrări. Acest fișier este utilizat la verificarea corectitudinii denumirilor directoarelor destinate să conțină fișierele înregistrărilor cu emoții, respectiv și fișierele anexe corespunzătoare. Se observă că mai mult din fișierele de intrare folosite sunt destinate doar etapei de verificare automată a corectitudinii denumirilor și adnotărilor. Fișierele de înregistrare au extensia .wav, fișierele de adnotare au extensia TextGrid. Erorile depistate sunt înscrise într-un fișier special numit Dc.log. Același fișier este folosit pentru înscrierea și a altor tipuri de erori depistate în fazele următoare ale programului. Fișierul va fi util ulterior în corectarea corpusului respectiv.

Programul tratează (blocul tratarea erorilor conform figura 2) și alte tipuri de erori existente în corpus, de exemplu incompletitudinea corpusului (lipsa unor înregistrări cu o emoție, sau existența unui fișier cu valorile F0, dar nu și cu valorile F1-F3, etc.). Erorile depistate sunt incluse în fișierul Dc.log.

Fișierul binar Dc.bin conține toate datele de intrare (mai puțin .wav) cu structuri organizate în forma unui array cu elemente de dimensiuni egale. Principalele elemente sunt structuri - de tipuri diferite. De exemplu, structura vorbitorilor conține informațiile din fișa vorbitorului. Prin această structurare, informația este accesibilă ușor pentru prelucrări statistice. Rezultatele programului sunt fișierele text Dump.txt și Statistics.txt (vezi figura 2).

ASPECTE METODOLOGICE DE ORGANIZARE A DATELOR ȘI DE ANALIZĂ STATISTICĂ A VOCILOR EMOȚIONALE

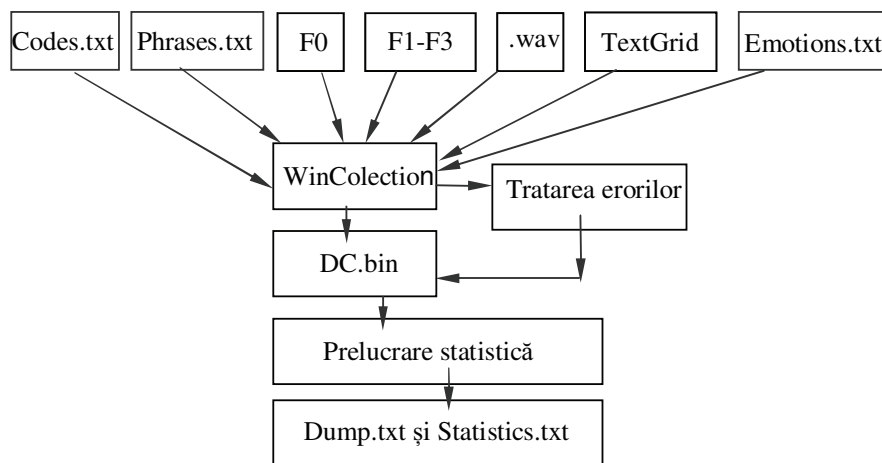


Figura 2: Organigrama programului WinCollection

Programul WinCollection creează colecția de date, fișierul binar DC.bin, care se va obține după parcurgerea și procesarea tuturor fișierelor de intrare. Acest fișier este creat în directorul colecției de date, fiind ulterior folosit de aplicație în analiza statistică a vocilor emoționale.

3.1 Etapele programului

Etapele programului sunt:

1. încărcarea informațiilor despre vorbitori din fișierul Codes.txt;
2. parcurgerea fișierului Phrases.txt și încărcarea informațiilor despre propozițiile pentru care există înregistrări;
3. încărcarea informațiilor despre stările emoționale pentru care există înregistrări, din fișierul Emotions.txt;
4. procesarea datelor de intrare în vederea detectării și semnalării eventualelor erori pe baza informațiilor din fișierele Codes.txt, Phrases.txt și Emotions.txt;
 - 4.1. încărcarea în memorie a datelor de intrare din fișierul de adnotare și a fișierelor cu valorile formașilor F0 și F1-F3;
 - 4.2. semnalarea în fișierul Dc.log a pașilor încărcării datelor și semnalarea erorilor;
5. crearea fișierul binar Dc.bin, care va conține întreaga colecție de date;
6. crearea fișierelor text, Dump.txt și Statistics.txt, descrise ulterior.

Programul parcurge toate sub-directoarele care conțin înregistrări despre o anumită stare emoțională, fiecare stare emoțională având în corespondență un sub-director care poartă numele respectivei stări. În fiecare sub-director, programul identifică toate seturile de câte patru fișiere (fișierul .wav, adnotarea corespunzătoare împreună cu fișierele ce conțin valorile formașilor F0, F1-F4). Identificarea se face pe baza codului vorbitorului și a propozițiilor pentru care există înregistrări.

3.2 Fișiere de ieșire

Fișierul binar conține și următoarele informații suplimentare: data și ora creării, versiunea de program utilizată, informații despre persoana care a realizat prelucrarea, etc.

Fișierul text Dump.txt conține:

- informații generale despre colecție (data creării, versiunea colecției, versiunea programului WinCollection, etc.);
- un sumar al bazei de date (număr vorbitori, fraze, stări, număr înregistrări pentru fiecare stare și frază);
- datele de intrare;
- un set de rezultate statistice pentru fiecare vocală, inclusiv precizarea numărului de apariții pentru fiecare stare emoțională (rezultatele se refera la minimul, maximum, valoarea medie și dispersia, calculate pentru valorile F0 respectiv F1-F3, pentru fiecare vocală, pentru fiecare apariție a vocalei într-o înregistrare).

Fișierul Statistics.txt generat de către programul WinCollection conține următoarele informații:

- pentru fiecare vocală, numărul de apariții pentru fiecare stare emoțională;
- pentru fiecare vocală, valoarea medie pentru populație, pentru formantul F0, respectiv F1-F3;

3.3 Tratarea erorilor

Programul semnalează, în urma prelucrării datelor de intrare, erorile găsite. Principalele erori semnalate de către program în urma prelucrării datelor sunt:

- lipsa adnotării pentru o înregistrare existentă în directorul analizat;
- erori în adnotare (apariția la un anumit nivel a unui fonem care nu apare în propoziția înregistrată);
- absența fișierului care conține informațiile despre formantul F0 pentru o anumită înregistrare;
- absența fișierului care conține informațiile despre formanții F1-F4 pentru o anumită înregistrare.

După corecția erorilor, este necesară re-procesarea tuturor datelor de intrare în vederea obținerii fișierului binar. Nu se fac adăugiri la colecție; adăugarea unei noi înregistrări la o colecție deja existentă se face prin re-procesarea datelor de intrare și obținerea unui nou fișier binar.

În prezent analiza statistică este implementată pe o vocală specifică, pentru un același vorbitor, în cadrul aceleiași propoziții specifice, pentru o stare specificată.

3.4 Estimarea variabilității

Exemplificăm utilizarea programului în vederea obținerii unui coeficient de variabilitate inter și intra-stare pentru valorile formanților precum și o statistică generală a formanților pentru vocale în limba română („a”, „e”, „i”, „u”, „ă”) în funcție de starea emoțională. În cadrul analizei statistice, primul autor a introdus un coeficient de variabilitate (asimetrie) definit conform formulei:

$$\eta = \frac{1}{4} \sum_{k=0}^3 \frac{1}{N} \sum_{\varphi=1}^N \frac{1}{M} \sum_{\psi=1}^M \frac{|F_k[j; \varphi, \psi] - F_k[\psi]|}{F_k[\psi]}, \quad (1)$$

unde: M- este numărul de apariții ale fonemului ψ în fraza φ respectivă, iar k – numărul formantului curent. $F_k[\psi]$ reprezintă valoarea medie a formantului k , pentru fonemul ψ , așa cum este determinată pentru un mare număr de vorbitori, în diverse contexte de pronunție.

Se recunoaște imediat că se realizează media diferențelor absolute ale valorilor formanților respectivi, în pronunțiile date ale fonemului ψ , diferențe normalizate cu valoarea medie a formantului respectiv pentru fonemul respectiv, $F_k[\psi]$:

$$\frac{1}{M} \sum_{\psi=1}^M \frac{|F_k[j; \varphi, \psi] - F_k[\psi]|}{F_k[\psi]}.$$

Deoarece în sumă apar diferențele unui singur vorbitor față de toți ceilalți, semnificația este aceea a diferenței vorbitorului față de media vorbitorilor (specificitatea vorbitorului în pronunția fonemului respectiv, în cadrul frazei date, pentru formantul ales). Aceste valori medii devin elemente în următoarea sumă, care, prin multiplicare cu $1/N$ devine media pentru toate frazele (mediile pe frază sunt mediate în sumă după φ). În fine, se face prin ultima sumă media, pentru toți formanții, a diferențelor vorbitor- media vorbitorilor. De remarcat că dacă se calculează numai sumele interioare -care reprezintă rezultate parțiale în program – se determină specificitatea vorbitorului la nivel de formant al unui fonem în fraza specificată, respectiv la nivel de formant al fonemului specificat pe toate frazele. Desigur, pentru fiecare fonem (aici, vocală) și pentru fiecare vorbitor, se obține câte un coeficient de variabilitate.

„Coeficientul de asimetrie” în pronunție a vorbitorului v_j față de media generală la aceeași stare și aceeași vocală, pentru formantul k este dat de formula:

$$\alpha_{B,v,k} = \frac{\overline{F}_{\#k}(\text{starea "B"; vocala "a"; vorbitor "v"}) - \overline{F}_{\#k}^0(B, a, \text{toti vorbitorii})}{\overline{F}_{\#k}^0(B, a, \text{toti vorbitorii})}, \quad (2)$$

4. Exemple de rezultate preliminare

Exemplele prezentate nu reprezintă rezultate definitive.

Numărul de ocurențe pentru fiecare vocală în fișierele care au fost analizate sunt date în tabelul 1. Menționăm că vocalele analizate din cele patru propoziții sunt opt „a”, patru „e”, patru „i”, doi „ă” și un „u”, pentru cinci vorbitori, patru stări emoționale cu o medie

de trei rostiri ale fiecărei propoziții. În cazul stării de tristețe am supus analizei doar două propoziții pentru trei vorbitori, iar pentru neutru trei propoziții pentru doi vorbitori.

Tabel 1: Numărul de apariții ale vocalelor analizate în funcție de starea emoțională

Vocale	Starea de bucurie	Starea de furie	Starea de tristețe	Ton neutru
A	84	103	52	102
E	40	50	21	54
I	39	48	24	60
Ă	22	27	14	24
U	11	14	0	9

Exemplificăm, în tabelele 3 și 4, utilitatea metodei expuse pentru caracterizarea unui vorbitor. Dintre cei cinci vorbitori analizați în această lucrare, unul (11861) se distinge de ceilalți prin valorile mult diferite pentru vocalele „u”, „e”, „i”, pentru valorile formanților F0 și F1, iar alt vorbitor (05392) se distinge prin valorile mult mai mici pe vocalele „e” și „a”, pentru valorile F0 și F3. Această concluzie este justificată de valorile prezentate în tabelele 3, 4 și 5. Această distanțare nu se manifestă la formanții F1 și F2 în general.

Tabel 2: Valorile coeficientului de variabilitate pentru toate fonemele („a”, „e”, „i”, „u”, „ă”) analizate, pentru toți vorbitorii analizați în funcție de starea emoțională

Codul Vorbitorului	Starea de bucurie	Starea de furie	Starea de tristețe	Tonul neutru
11861	0.156	0.150	0.110	0.178
12121	0.157	0.172	0.102	0.126
83714	0.206	0.166	0.116	0.219
05392	0.093	0.091	0.104	0.059
32167	-	0.102	0.049	0.103

Tabel 3: Valorile coeficientului de asimetrie pentru valorile formanților F0-F3 în funcție de starea emoțională, pentru vorbitor 11861m, vocala „a”

11861/vocala A	Coef. de asimetrie pentru valoarea lui F0	Coef. de asimetrie pentru valoarea lui F1	Coef. de asimetrie pentru valoarea lui F2	Coef. de asimetrie pentru valoarea lui F3
Starea de bucurie	0.33	0.01	0.03	0.003
Starea de tristețe	0.39	0.18	0.28	0.10
Starea de furie	0.27	0.03	0.12	0.01
Ton neutru	0.33	0.10	0.13	0.07

Tabel 4: Valorile coeficientului de asimetrie pentru valorile formanților F0-F3 în funcție de starea emoțională, pentru vorbitor 05392m, vocala „a”

05392/vocala A	Coef. de asimetrie pentru valoarea lui F0	Coef. de asimetrie pentru valoarea lui F1	Coef. de asimetrie pentru valoarea lui F2	Coef. de asimetrie pentru valoarea lui F3
Starea de bucurie	0.11	0.11	0.18	0.07
Starea de tristețe	0.11	0.16	0.12	0.03

ASPECTE METODOLOGICE DE ORGANIZARE A DATELOR ȘI DE ANALIZĂ STATISTICĂ A
VOCILOR EMOȚIONALE

05392/vocala A	Coef. de asimetrie pentru valoarea lui F0	Coef. de asimetrie pentru valoarea lui F1	Coef. de asimetrie pentru valoarea lui F2	Coef. de asimetrie pentru valoarea lui F3
Starea de furie	0.09	0.05	0.12	0.04
Ton neutru	0.08	0.13	0.10	0.02

Tabel 5: Valorile coeficientului de asimetrie pentru valorile formațiilor F0-F3 în starea de bucurie, pentru patru vorbitori, vocala „u”

Vorbitori/vocala la U/Starea de bucurie	Coef. de asimetrie pentru valoarea lui F0	Coef. de asimetrie pentru valoarea lui F1	Coef. de asimetrie pentru valoarea lui F2	Coef. de asimetrie pentru valoarea lui F3
11861	0.38	0.32	0.31	0.11
12121	-	0.47	0.19	0.16
83714	0.23	0.43	0.29	0.15
05392	0.13	0.24	0.12	0.07

Rezultatele din tabelul 2 arată că valorile coeficienților de variabilitate sunt mai mari în cazul stării de bucurie și sunt mai scăzute în cazul stării de tristețe. În tabelele 3, 4 și 5 valorile coeficientului de asimetrie sunt cele mai mari în cazul frecvenței fundamentale, iar cele mai mici pentru formantul de ordin trei.

S-a urmărit diferențierea unui vorbitor față de ceilalți în funcție de stările emoționale (tabel 2) precum și analiza parametrilor F0-F3 pentru toți vorbitorii, în funcție de vocale (tabel 5). În tabelele 3 și 5 am exemplificat diferențele ce apar între doi vorbitori pentru aceeași vocală „a” pe baza parametrilor F0-F3 în funcție de stările emoționale.

5. Concluzii și direcții viitoare

Scopul principal al acestei lucrări a fost de a prezenta metodologia de organizare a datelor în vederea procesării lor cu o nouă aplicație informatică dezvoltată de colectivul nostru și de a prezenta sumar capabilitățile aplicației. Am evidențiat principalele obiective care au stat la baza metodologiei de organizare și avantajele aplicației realizate. S-au prezentat etapele obținerii colecției de date precum și posibilitatea de manipulare pentru prelucrări statistice. Analiza sumară realizată exemplifică flexibilitatea în utilizare.

În viitor, pe baza informațiilor privind sexul vorbitorului din fișierul Codes.txt se vor adapta limitele acceptabile pentru F0 ce se vor aplica în validarea valorilor din fișierul cu F0, notat F0 în figura 2. În etapa următoare, analiza va fi extinsă la toți vorbitorii din cadrul corpusului SRoL.

Mulțumiri. Cercetarea a fost realizată cu sprijinul Academiei Române, în cadrul temei interne a Institutului de Informatică Teoretică din Iași. Autorii mulțumesc celorlalți co-autori ai sitului Sunetele Limbii Române pentru includerea altor înregistrări noi care au fost analizate în lucrare precum și referențelor care au realizat observații pertinente.

Contribuția autorilor: Primul autor a inițiat tema cercetării și structura lucrării, a elaborat metodologia generală de lucru, conceptul și structura generală a aplicației informatice, a propus formulele menționate

în lucrare și a coordonat întreaga cercetare. Al doilea autor a implementat în C++ programul WinCollections. Al treilea autor a realizat înregistrările, a efectuat manual adnotările și a creat cu utilitarul Praat fișierele necesare utilizării programului WinCollections. Toți autorii au contribuit la redactarea lucrării.

Referințe bibliografice

- Boersma, P., Weenink, D., Institute of Phonetic Science, University of Amsterdam, Praat: doing phonetics by computer, www.praat.org.
- Feraru, M., Teodorescu, H.N. (2009). Classification of the Emotional States in Speech using the SRoL Database – Preliminary Results, *Proc. 3rd Int. Conf. Electronics, Computers and Artificial Intelligence*, Pitesti, România, ISBN 1843-2115, 27-32.
- Kienast, M., Sendlmeier, W. F. (2008). Acoustical Analysis of Spectral and Temporal Changes in Emotional Speech, An Acoustic Framework for Detecting Fatigue in Speech Based Human-Computer-Interaction, *Lecture Notes in Computer Science*, ISBN 978-3-540-70539-0, 5105/2008, 54-61.
- Mcgilloway, S., Cowie, R., Douglas-Cowie, E. et al. (2000). Approaching Automatic Recognition of Emotion from Voice: A rough Benchmark. *Proc. ISCA Workshop Speech and Emotion, Newcastle*, 207-212.
- Nakatsu, R., Solomides, A., Tosa, N. (1999). Emotion Recognition and its Application to Computer Agents with Spontaneous Interactive Capabilities. *Proc. IEEE Int. Conf. Multimedia Computing and Systems*, Florence, Italy, 2, 804-808.
- Scherer, K., A. (2000). Cross-Cultural Investigation of Emotion Inferences from Voice and Speech: Implications for Speech Technology. *Proc. Conf. Spoken Language Processing (ICSLP)*, China, http://www.unige.ch/fapse/emotion/publications/pdf/ic脾l00_crosscul.pdf.
- Teodorescu, H.N., Feraru, M. (2007). A study on Speech with Manifest Emotions, 10th International Conference on Text, Speech and Dialogue, *Lecture Notes in Computer Science, Springer Verlag*, ISBN 978-3-540-74627-0, 4629/2007, 254-262.
- Teodorescu, H.N., Feraru, M., Trandabăț, D., Zbancioc, M., Luca, R., Verbuță, A., M. Ganea, R. Voroneanu, O. Pistol, L. (2005). Proiectul Sunetele Limbii Române, www.etc.tuiasi.ro/sibm/romanian_spoken_language/index.htm.
- Ververidis, D., Kotropoulos C. (2006). Emotional speech recognition: Resources, features, and methods, *Speech Communications*, 48: 9, 1162-1181.