

**UNIVERSITATEA TEHNICĂ "GHEORGHE ASACHI"  
DIN IAȘI**



**- Rezumat -**  
**CONTRIBUȚII LA ANALIZA CORELATIVĂ A SEMNALULUI  
VOCAL ÎN DIVERSE FORME DE DISCURS ÎN LIMBA  
ROMÂNĂ**

**Ștefan-Andrei Ghețu**

**Conducător de doctorat : Acad. prof. univ. dr. ing. Horia-Nicolai Teodorescu**

**UNIVERSITATEA TEHNICĂ "GHEORGHE ASACHI" DIN IAȘI**  
**RECTORATUL**

Către

---

---

Vă facem cunoscut că, în ziua de 10.07.2024 la ora 10:00 în Aula Magna "Carmen Sylva" a Universității Tehnice "Gheorghe Asachi" din Iași , va avea loc susținerea publică a tezei de doctorat intitulată:

"Contribuții la analiza corelativă a semnalului vocal în diverse forme de discurs în limba română"

elaborate de domnul Ghelțu Ștefan-Andrei, în vederea conferirii titlului științific de doctor.

Comisia de doctorat este alcătuită din:

1. Prof. dr. ing. Daniela Târniceriu, Facultatea de Electronică, Telecomunicații și Tehnologia Informației, Universitatea Tehnică „Gheorghe Asachi” din Iași - președinte
2. Acad. prof. univ. dr. ing. Horia-Nicolai Teodorescu, Facultatea de Electronică, Telecomunicații și Tehnologia Informației, Universitatea Tehnică „Gheorghe Asachi” din Iași - conducător de doctorat
3. Prof. dr. ing. Corneliu Burileanu, Universitatea Națională de Știință și Tehnologie „Politehnica” din București - referent oficial
4. Dr. ing. Dan. Cristea, CS I la Institutul de Informatică Teoretică al Academiei Române din Iași, membru corespondent al Academiei Române - referent oficial
5. Dr. ing. Vasile Apopei, CS II Director al Institutului de Informatică Teoretică al Academiei Române, Iași - referent oficial

Cu această ocazie vă invităm să participați la susținerea publică a tezei de doctorat.

**RECTOR,**

Prof.univ.dr.ing. Dan Cașcaval



**Secretar universitate,**

Ing. Cristina Nagiț

\*pentru susținerile online se va preciza link-ul și soluția de software

# **CONTRIBUȚII LA ANALIZA CORELATIVĂ A SEMNALULUI VOCAL ÎN DIVERSE FORME DE DISCURS ÎN LIMBA ROMÂNĂ**

Ștefan-Andrei Ghețu  
domeniul Inginerie Electronică, Telecomunicații și Tehnologii Informaționale

Președinte comisie doctorat:  
Conducător de doctorat:  
Referenți oficiali:

Prof.univ.dr.ing. Daniela Tărniceriu  
Acad. Horia-Nicolai Teodorescu  
Prof.univ.dr.ing. Corneliu Burileanu  
CS.I dr.ing. Dan Cristea  
CS.II dr.ing. Vasile Apopei

# Cuprins

<b>UNIVERSITATEA TEHNICĂ “GHEORGHE ASACHI” DIN IAȘI</b> .....	2
<b>RECTORATUL</b> .....	2
Capitolul 1. STADIUL ACTUAL ÎN TEHNOLOGIA VORBIRII – ELEMENTE INTRODUCATIVE PRIVIND ANALIZA VOCII EMOȚIONALE .....	6
1.1. Introducere .....	6
1.2. Semnalul vocal și sistemele de recunoaștere a vorbirii.....	7
1.3. Emoțiile la nivel de semnal vocal și recunoașterea lor .....	10
Capitolul 2. METODEDE ȘI TEHNICI FOLOSITE ÎN ANALIZA SEMNALULUI VOCAL, ÎN CURSUL CERCETĂRII .....	14
2.1. Introducere – elemente de statistică .....	14
2.2. Coeficienți de corelație .....	14
2.2.1. Coeficientul de corelație Spearman .....	14
2.2.2. Coeficientul de concordanță kendall.....	15
2.2.3. Coeficientul de corelație Pearson.....	16
2.3. Analiza de regresie.....	16
2.3.1. Regresia liniară simplă.....	17
2.3.2. Regresia liniară multiplă .....	19
Capitolul 3. APLICAREA METODELOR ȘI TEHNICILOR DE ANALIZĂ A SEMNALULUI VOCAL .....	24
3.1. Abordări în calcularea coeficientului de corelație – $r$ .....	24
3.2. Studiu preliminar asupra corelației formanților vocalelor din limba română .....	25
Capitolul 4. ANALIZA CORELATIVĂ A VORBIRII EMOTIVE / EMOTIILOR LA NIVELUL SEMNALULUI VOCAL – Contribuții .....	28
4.2. Rezultate .....	34
4.2.1. Vocale susținute .....	34
4.2.2. Vorbirea dinamică.....	39
Capitolul 5 – REZULTATE PRIVIND EFECTUL EMOȚIILOR ASUPRA NIVELULUI DE CORELAȚIE AL FORMANȚILOR .....	41
5.1. Calcularea coeficienților de regresie liniară.....	48
CAPITOLUL 6 – Concluzii, contribuții și direcții viitoare de cercetare .....	58
6.1. Sinteza analizei și contribuțiilor.....	58
6.2. Detalii privind structura capitolelor și contribuțiile descrise in capitole .....	59
6.3. Sinteza rezultatelor proprii.....	59

6.4. Lista lucrărilor publicate sau comunicate .....	60
6.4.1. Lucrări publicate în volumele conferințelor IEEE .....	60
6.4.2. Lucrare în revistă (GS, Sciendo).....	61
6.4.3. Lucrări comunicate .....	61
6.5. Direcții viitoare de cercetare deschise de cercetarea proprie (în colectiv).....	61

# Capitolul 1. STADIUL ACTUAL ÎN TEHNOLOGIA VORBIRII – ELEMENTE INTRODUCATIVE PRIVIND ANALIZA VOCII EMOȚIONALE

## 1.1. Introducere

În recunoașterea automată a vorbirii (ASR), deși în ultimii ani s-au realizat progrese semnificative în domeniul inteligenței artificiale și în ASR în particular, au rămas atât zone de cercetare neacoperite, cât și probleme de cercetare ce impun derularea unor studii empirice în scopul aprofundării unor direcții de analiză.

Performanța actuală a sistemelor de recunoaștere automată a vorbirii (ASR) este cu mult mai scăzută comparativ cu procesul de comunicare interumană (HSR-Human Speech Recognition) care include pe de-o parte, transmiterea informațiilor de la emițător la receptori sub forma unui mesaj vocal și pe de altă parte, include procesarea semantică precum și integrarea conținutului emoțional cuprins în mesaj. În cadrul procesului de comunicare interumană sunt procesate simultan trei tipuri de informații care țin de: *limbajul verbal*, reprezentat de structurile lingvistice ce alcătuiesc mesajul; *limbajul nonverbal*: reprezentat de elementele ce țin de mimică, gestică, menținerea contactului vizual, proximitatea față de interlocutor etc. și *limbajul paraverbal*: adică elementele de comunicare precum: tonalitatea, intonația, ritmul, accentele, pauzele din vorbire etc. Recunoașterea emoțiilor în discursul vorbitorilor este posibilă doar prin procesarea simultană a informațiilor provenite de la cele trei tipuri de limbaj. Totodată, se analizează concordanța dintre mesajul transmis la nivel verbal și informațiile exprimate la nivel non-verbal și para-verbal, pentru ca în final, interlocutorul să își adapteze răspunsul la specificul situației de comunicare.

Perfecționarea sistemelor de recunoaștere automată a vorbirii reprezintă o condiție indispensabilă pentru dezvoltarea unui sistem de dialog om-mașină (Human Computer Interaction) care să satisfacă exigențele procesului de comunicare. Primele astfel de sisteme au fost proiectate inițial să recunoască *cuvinte izolate*, ulterior, s-a trecut la recunoașterea *cuvintelor înlănțuite*. Progresele înregistrate până în acest punct, au dus la proiectarea sistemelor astfel încât să fie posibilă recunoașterea unor propoziții și fraze mai complexe. Scopul studiilor derulate în această direcție de cercetare vizează crearea unui sistem care să poată recunoaște automat *vorbirea curentă* exprimată de om și să mențină un dialog adecvat situației de comunicare.

Sistemele de recunoaștere automată a vorbirii (ASR) prezintă o limitare considerabilă în ceea ce privește recunoașterea conținutului emoțional al mesajului vocal. Astfel, principala problemă de cercetare în această direcție nu mai este reprezentată de simpla creare a unor sisteme automate de recunoaștere a vorbirii, ci extinderea capacității de lucru a acestora de la procesarea

semantică singulară a mesajului la o formă de procesare duală, adică procesarea mesajului vocal atât la nivel semantic, cât și la nivelul conținutului emoțional.

## 1.2. Semnalul vocal și sistemele de recunoaștere a vorbirii

Semnalului vocal poate fi analizat pe trei niveluri:

- nivelul *caracteristicilor prozodice*: rata vorbirii, pauze, sincronizarea vorbirii, frecvența fundamentală ( $F_0$ ), numită și înălțimea sunetului care exprimă numeric frecvența de vibrație a coardelor vocale;
- nivelul caracteristicilor specifice *tractului vocal* definit ca o succesiune de tuburi sau cavități acustice de lungimi și secțiuni diferite situate între glotă și buze. Tractul vocal include cavitatea faringiană, cavitatea bucală și cavitatea nazală;
- nivelul caracteristicilor *sursei semnalului* ce reprezintă proprietățile spectrale ale semnalului vocal care este prelucrat prin extragerea unor parametri dintre care cel mai des utilizați sunt: coeficienții mel-cepstrali (MFCC) și coeficienții perceptuali de predicție liniară (PLP) .

Proiectarea sistemelor (automate) de recunoaștere a vorbirii au la bază trei abordări:

- **abordarea fonetic-acustică** care presupune asocierea unor etichete specifice sunetelor pe baza fonemelor extrase din semnalul de intrare;
- **abordarea de recunoaștere a pattern-ului** care depinde de realizarea a două procese complementare *antrenarea* pattern-ului și *testarea* pattern-ului, clasificarea finalizându-se pe baza calculului unor algoritmi matematici;
- **abordarea inteligenței artificiale** care combină cele două abordări menționate anterior, ducând la obținerea unui mecanism de clasificare creat după modul în care sistemul psihic uman analizează, recunoaște și clasifică caracteristicile informaționale din mediu.

Deși toate cele trei abordări sunt utilizate în elaborarea unor sisteme de recunoaștere automată a vorbirii, cea de-a doua abordare este cel mai frecvent folosită atunci când se urmărește și recunoașterea conținutului emoțional al mesajului vocal.

În studiile de specialitate în care se utilizează sisteme de recunoaștere automată a vorbirii, autorii înregistrează un anumit număr de vorbitori, alcătuind o bază de date proprie sau folosesc o bază de date oficială ce aparține unor institute de cercetare. Realizarea înregistrărilor constă în faptul că autorii au ales un enunț și ca sarcină experimentală, au solicitat vorbitorilor antrenați, participanți la studiu, să-l pronunțe exprimând una dintre emoțiile vizate de cercetare. De asemenea, în câteva studii s-a folosit o bază de date alcătuită din înregistrări preluate din talk-

show-uri, aplicații call center sau emisiuni TV. Însă, comparând calitatea caracteristicilor semnalului vocal între înregistrările realizate într-un mediu controlat și cele similare limbajului natural, autorii au constatat că în primul caz, caracteristicile semnalului vocal sunt mai proeminente și mai ușor de distins [1].

În articolul [2], autorii realizează o sinteză teoretică a procesului de analiză a semnalului vocal, explicând principalele tehnici utilizate pentru fiecare etapă. Sistemele de recunoaștere automată a vorbirii parcurg următoarea etapizare succesivă: *Pre-procesarea*, *Extragerea Caracteristicilor*, *Selecția Caracteristicilor* și *Clasificarea*. Această etapizare reprezintă totodată schematizarea unui mecanism ce permite identificare emoțiilor în discursul vorbitorilor.

1. *Semnalul vocal* (Speech Signal)
2. *Pre-Procesarea* (*Pre-Processing*)
  - a. Eliminarea zgomotului de fundal (Background Noise Removal)
  - b. Detectarea cuvintelor vorbite (Speech Word Detection)
  - c. Pre-Accentuare (Pre-Emphasis)
  - d. Buffering
  - e. Ferestre de timp (Windowing)
3. *Extragerea Caracteristicilor* (*Feature Extraction*)
  - a. Analiza acustică a vorbirii (*Acoustic Front End of Speech Analysis*)
  - b. Compilarea vectorului de trăsături (*Compiling Extended Feature Vector*)
  - c. Transformare vectorului în trăsături compacte (*Transform Into Compact Feature Vector*)
4. *Selecția caracteristicilor* (Feature Selection)
  - a. Metode de filtrare (*Filter Method*)– LDA, ANOVA, Pearson Correlation, Chi-Square
  - b. Wrapper Method – SFS, Backward Elimination
  - c. Metode Hibrade (*Hybrid Method*) – Lasso Regression, Ridge Regression
5. *Clasificarea* (Classification)
  - SVM, HMM, ANN
6. Output

În etapa pre-procesării, semnalul vocal este transformat din analogic în semnal discret continuu în timp pentru a fi adecvat prelucrărilor digitale ulterioare. Acest proces începe cu eliminarea zgomotului de fundal, continuă cu identificarea cuvintelor, fazarea semnalului audio, stocarea acestuia într-o zonă de memorie (buffering) și segmentarea semnalului audio în ferestre de timp.



În etapa a treia (extragerea caracteristicilor), semnalul obținut după procesare prezintă o dimensionalitate ridicată ce ar putea influența negativ rata de eroare în recunoașterea cuvintelor. Pentru a obține rezultatele dorite, trebuie extrase din mulțimea de caracteristici ale semnalului vocal doar cele relevante pentru etapele ulterioare de selectare și clasificare. Așadar, scopul aplicării algoritmului este de a extrage vectorul caracteristică care să aibă o dimensionalitate adecvată, mai scăzută comparativ cu semnalul la intrare. Extragerea caracteristicilor se realizează în trei etape începând cu extragerea caracteristicilor în forma lor inițială (brută) și identificarea anvelopei puterii spectrale. Apoi, vectorului extins, alcătuit din caracteristici statice și dinamice ale semnalului, i se aplică un algoritm de comprimare, pentru ca în ultima etapă, să fie transformat într-un vector caracteristică compact și robust.

Cele mai uzuale tehnici de extragere a caracteristicilor semnalului vocal în domeniile timp și frecvență, în etapa de preprocesare, relevante pentru sinteza și recunoașterea vorbirii, sunt: transformata Wavelet discretă (DWT), coeficienții mel-cepstrali (MFCC), coeficienții perceptuali de predicție liniară (PLP), analiza componentelor principale (PCA), analiza discriminatorie liniară (LDA) și coeficientul cepstral de predicție liniară (LPCC).

Cea de-a treia etapă (selectarea caracteristicilor) urmărește atingerea a două obiective: convertirea setului de caracteristici astfel încât dimensionalitatea să fie cât mai scăzută și eficientizarea procesului computațional de predicție a celui mai bun model/set de caracteristici din totalitatea celor supuse analizei.

Selectarea caracteristicilor poate fi realizată utilizând: metode de filtrare, metode wrapper și metode hibrid/embedded (care le combină pe cele de filtrare și wrapper). Specificul metodelor de filtrare constă în faptul că variabilele sunt selectate indiferent de model. Printre tehnicile statistice reprezentative metodelor de filtrare sunt: analiza de varianță (ANOVA), Chi-Square ( $\chi^2$ ), corelația Pearson ș.a. Metodele wrapper se caracterizează prin crearea unor seturi de variabile (modele) și analiza interacțiunilor dintre acestea cu scopul identificării celui mai eficient model. Tehnicile statistice reprezentative pentru metodele hibrid sunt regresia Lasso și regresia Ridge.

În etapa finală se realizează clasificarea caracteristicilor și asocierea rezultatului clasificării cu emoția specifică. Cele mai des utilizate clasificatoare pentru recunoașterea emoțiilor din semnalul vocal sunt: mașini cu vectori suport (Support Vector Machine), Hidden Markov Model (HMM) și rețelele neuronale artificiale (Artificial Neural Networks) [3-5]. În studiul [11], utilizarea unei mașini cu vector suport a dus la o creștere a procentului de acuratețe în recunoașterea emoțiilor de la 64,1% la 90,7%.

Sinteza teoretică cuprinsă în acest subcapitol clarifică procesualitatea analizei semnalului vocal și prezintă tehnicile și procedurile statistice care permit decodare conținutului semantic și

emoțional din mesajul vocal, aducând o contribuție semnificativă în ceea ce privește clarificările conceptuale din această arie de cercetare.

### 1.3. Emoțiile la nivel de semnal vocal și recunoașterea lor

În literatura de specialitate și mai ales în studiile din psihologie, emoțiile sunt definite ca fiind impulsuri automate care direcționează comportamentul uman. Acestea au avut o contribuție majoră în procesul evolutiv, mobilizând resursele individului în situațiile de pericol sau amenințare.

Modelul circumplex al emoțiilor propus de Russell [12] clasifică stările emoționale ale individului în funcție de două dimensiuni: *polaritatea* (pozitivă sau negativă) și *intensitatea* emoției, adică măsura în care aceasta activează resursele individului și îi direcționează comportamentul în vederea luării unei decizii. Un exemplu reprezentativ îl constituie reacția de tip *luptă-sau-fugi* (fight-or-flight response) în care individul percepe un stimul amenințător și drept consecință, organismul său se pregătește la nivel fiziologic, cognitiv și afectiv pentru reacția ce va urma. Emoțiile de bază cu cel mai înalt nivel de intensitate și cea mai puternică orientare a comportamentului sunt frica și furia. În funcție de combinarea celor două dimensiuni (polaritatea și intensitatea), emoțiile pot fi plasate într-unul dintre cele patru cadrane ale diagramei lui Russell.

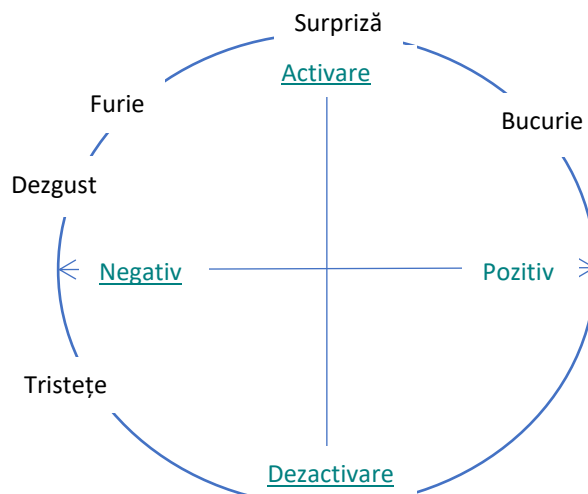


Figura 1. Modelul circumplex al emoțiilor propus de Russell [12]

În studiul [13], autorii au preluat modelul circumplex al emoțiilor și au creat un algoritm de recunoaștere a stărilor emoționale bazat pe un clasificator neuronal. Au combinat în analiză trei abordări: temporală, prozodică și frecvențială cu scopul de a extrage caracteristicile descriptive ale semnalului vocal. Aceste caracteristici au fost parametri pe baza cărora s-a creat un clasificator neuronal pentru patru din cele șase emoții de bază: bucuria, furia, frica și tristețea. Clasificatorul a fost antrenat să asocieze caracteristicile semnalului vocal cu nivelul scăzut/ridicat de activare și

polaritatea emoțiilor astfel încât să recunoască emoția și să o localizeze în cadranul corespunzător din diagrama lui Russell.

Semnalul vocal a fost segmentat în ferestre de câte 20 de ms, cu 50% grad de suprapunere, astfel încât să se obțină segmente de semnal cu caracteristici cvasi-staționare. S-a aplicat un algoritm de normalizare pentru semnalul audio, în urma căruia amplitudinea frecvenței semnalului a luat valori cuprinse între 0 și 1, minimizându-se în felul acesta rata erorilor date de niveluri foarte scăzute ale intensității semnalului vocal, cauzate de calitatea slabă a înregistrării.

Pentru banda de frecvență, s-a aplicat un filtru FIR în sens Chebyshev (minimax) (pentru a se minimiza valoarea maximă față de răspunsul dorit) având frecvență de tăiere de 20Hz și o frecvență de comutare ridicată, de 6800 Hz. Criteriul folosit pentru stabilirea filtrului a fost CMU Sphinx [14]. Apoi, s-a utilizat un detector vocal pentru identificarea și eliminarea pauzelor în vorbire. Ulterior, s-a calculat [13] numărul de ferestre de timp care a fost echivalent cu numărul de segmente de voce activă detectate. S-a calculat rata trecerilor prin zero care indică variația ratei de energie la frecvențe scăzute și respectiv, ridicate. În cazul unui semnal eșantionat, se calculează rata trecerilor prin zero atunci când cel puțin două eșantioane succesive au semne opuse.

$$Zn = \sum_{m=-\infty}^{\infty} \frac{|sgn(x(m)) - sgn(x(m-1))| \cdot w(n-m)}{2} \quad (1)$$

unde:

$$sgn(x) = \begin{cases} 1, \forall x \geq 0 \\ 0, \forall x \leq 0 \end{cases}$$

$x(m)$  = semnalul la momentul  $m$ ;

$x(m-1)$  = semnalul la momentul anterior  $m-1$ ;

$w(n-m)$  = fereastra de timp.

În [13] s-a calculat indicele de aplatizare kurtosis, numit și coeficient de boltire. Acesta măsoară gradul de deformare al distribuției prin raportare la o distribuție normală (gaussiană). Atunci când coeficientul kurtosis are o valoare mai mare decât 3, distribuția datelor este leptocurtică (vârful distribuției este mai ascuțit, iar datele sunt concentrate în jurul unui interval), iar atunci când are o valoare mai mică decât 3, distribuția datelor este platicurtică (datele sunt ”împrăștiate”).

$$k = \frac{\frac{1}{n} \sum (x_i - \bar{x})^4 \cdot n_i}{\left( \frac{1}{n} \sum (x_i - \bar{x})^2 \cdot n_i \right)^2} \quad (2)$$

unde:

$k$  = coeficientul Kurtosis;

$x_1$  = valoarea fiecărui eșantion de date;  
 $\bar{x}$  = media eșantionului;  
 $n_i$  = frecvența corespunzătoare fiecărei valori.

Autorii [13] au stabilit caracteristicile prozodice ale semnalului vocal, numite și caracteristici suprasegmentale prin evaluarea nivelului de energie din voce. În acest sens, ei au calculat RMS (Root Mean Square - rădăcina pătrată medie) pentru a estima energia semnalului vocal pe intervale.

$$En = \sum_{m=-\infty}^{\infty} (x(m)w(n-m))^2 \quad (3)$$

unde:

$En$  = rata de energie a segmentului de semnal;  
 $x_m$  = valoarea semnalului la momentul  $m$ ;  
 $w(n-m)$  = fereastra de timp.

Din perspectiva abordării frecvențiale, autorii [13] au extras parametrii vocali utilizând metoda mel-cepstrală. În acest sens, a fost calculat coeficientul MFCC (Mel Frequency Cepstral Coefficients) definit ca fiind cepstrum semnalului segmentat în ferestre de timp pe o scală non-liniară de frecvență. Astfel, se urmărește compararea nivelului de energie pentru fiecare bandă de frecvență a semnalului și cuantificarea acestuia într-o valoare numerică prin aplicarea formulei:

$$mfcc(n) = \frac{1}{M} \sum_{m=1}^M Sm(m) \cos \left[ \frac{2\pi}{M} (m + 0,5)n \right] \quad (4)$$

unde:

$Sm(m)$  = valoarea logaritmică a nivelului de energie la ieșire pentru fiecare filtru Mel;  
 $M$  = numărul total de filtre (40 de filtre);  
 $n$  = reprezintă valoarea calculată a coeficientului MFCC.

Caracteristicile spectrale ale semnalului au fost extrase în [13] și prin aplicarea unor tehnici nonparametrice. Autorii au aplicat metoda lui Welch care reprezintă o tehnică de estimare a densității spectrale (PSD-Power Spectral Density) prin analizarea puterii semnalului la diferite frecvențe,

$$P_{Welch} = \frac{1}{p} \sum_{p=1}^p P_{xx}(k) \quad (5)$$

unde:

$P_{xx}(k)$  = densitatea spectrală pentru fiecare segment de semnal;  
 $P$  = numărul total de segmente;

Pentru  $0 \leq k \leq N - 1$ , unde  $N$  reprezintă dimensiunea transformatei Fourier.

Baza de date a fost alcătuită din 120 de înregistrări divizate în patru grupuri de câte 30 de înregistrări corespunzătoare fiecărei emoții analizate [13]. Antrenarea rețelei neuronale s-a efectuat pentru patru emoții: bucurie, tristețe, furie și frică. Autorii au utilizat metoda Levenberg-Marquard backpropagation. Baza de date a fost divizată randomizat în trei condiții experimentale: 70% - antrenare, 15% - validare și 15% - testare. Rezultatul clasificării a fost proiectat să creeze o asociere între parametrii analizați și una dintre cele patru etichete: furie=3; bucurie=2; frică=1; tristețe=0.

Rezultatele pe care autorii [13] le-au obținut indică un nivel de performanță al clasificatorului pe toate emoțiile de 95%. Clasificatorul a obținut cel mai ridicat nivel de acuratețe în detectarea emoției *tristețe*, prezentând un randament de 96,67%. Pe de altă parte, pentru emoția *frică* clasificatorul a prezentat cel mai scăzut nivel de randament, identificând corect emoția în procent de 70%. Autorii au constatat că frica a fost confundată cu bucuria, ceea ce semnifică faptul că decizia de clasificare a fost în mai mare măsură determinată de intensitatea emoției decât de polaritatea acesteia (conform modelului circumplex al lui Russell [12], furia și bucuria au în comun nivelul ridicat de intensitate, însă valențele lor sunt opuse). Procentajul de acuratețe în ceea ce privește recunoașterea emoției *bucurie* a fost de 93,44%, iar pentru emoția *furie*, 86,67%.

Autorii [13] au inclus în analiză încă două emoții: *plictiseala* și *dezgustul*. Clasificatorul nu a fost antrenat să recunoască aceste două emoții, însă s-a urmărit corectitudinea asocierii acestora pe baza similarității caracteristicilor semnalului vocal cu cele patru emoțiile antrenate. Cu alte cuvinte, autorii au pornit de la ipoteza că *plictiseala* va fi asociată de clasificator cu *tristețea* (deoarece ambele prezintă un nivel scăzut de intensitate și polaritate negativă), iar *dezgustul* va fi asociat în mod similar cu *furia* sau *frica* (fiindcă emoțiile au în comun un nivel ridicat de intensitate și polaritate negativă). Rezultatele obținute confirmă parțial ipoteza cercetării în [13]. Astfel, în ceea ce privește *plictiseala* s-a observat o asociere cu *tristețea* în procent de 66,66%, procent semnificativ mai ridicat comparativ cu celelalte trei emoții antrenate (bucurie - 9,8%; furie - 0%; frică - 23,52%). Așadar, clasificatorul a avut o performanță ridicată în recunoașterea unei emoții neantrenate (*plictiseala*) prin asocierea acesteia cu o emoție antrenată (*tristețea*). Autorii se așteptau să observe o asociere între *dezgust* cu *furia* sau *frica*, însă au observat o asociere în procent de 35,48% cu bucuria, 29,03% cu *tristețea* și doar 35,48% cu *furia* și *frica* (emoții cu valență negativă și intensitate ridicată, similare *dezgustului*).

Ca direcție viitoare de cercetare, autorii propun testarea clasificatorului pe alte emoții și reluarea analizei comparative a rezultatelor pentru a identifica eventuale modificări ale procentului de acuratețe în recunoașterea emoțiilor. Până în acest punct, performanța globală a clasificatorului rămâne de 95%, iar *dezgustul* continuă să fie emoția cu cel mai scăzut nivel de acuratețe în recunoaștere, conform [13].

# Capitolul 2. METODE ȘI TEHNICI FOLOSITE ÎN ANALIZA SEMNALULUI VOCAL, ÎN CURSUL CERCETĂRII

## 2.1. Introducere – elemente de statistică

Statistica este latura științei care utilizează un ansamblu de metode ce presupun colectarea, sistematizarea, prelucrarea și interpretarea unor seturi de date despre diverse fenomene [45-46].

Statistica descriptivă [46] folosește tehnici ce permit descrierea unor grupuri de date și are ca scop organizarea acestora într-o formă clară și utilizabilă. Datele astfel obținute sunt exprimate prin valori numerice însoțite de reprezentări grafice. Statistica inferențială [46] se ocupă cu realizarea unor inferențe, inducții și emiterea unor concluzii. În funcție de specificul datelor investigate, statistica poate fi: *parametrică* (variabilele sunt cantitative și exprimate prin date numerice, măsurate pe scale de interval și de raport) sau *neparametrică* (variabilele sunt calitative și măsurate pe scale nominale sau ordinale).

## 2.2. Coeficienți de corelație

Gradul de asociere este exprimat prin coeficienți standardizați care permit realizarea unor comparații directe.

**Coeficientul de corelație** [47] este o valoare cantitativă ce descrie relația dintre două sau mai multe variabile. Poate lua valori cuprinse între (-1 și +1), unde valoarea +1 indică existența unei corelații pozitive perfecte între cele două variabile, valoarea 0 indică absența totală a corelației, iar valoarea -1 semnifică existența unei corelații negative, adică asocierea dintre variabile este inversă (raport invers proporțional).

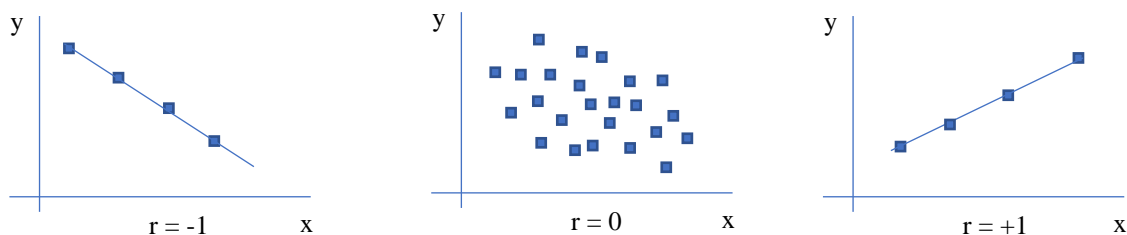


Figura 1. Exemple de corelație

**2.2.1. Coeficientul de corelație Spearman** [45],[47-48] numit și coeficientul de corelație al rangurilor, se folosește atunci când variabilele sunt de tip ordinal sau atunci când una/ambele variabile se măsoară pe scale de interval sau de raport, însă există valori extreme în distribuția datelor care nu pot fi eliminate. Normalizarea se realizează prin transformarea valorilor variabilei în ranguri și ulterior, se aplică coeficientul de corelație Spearman ( $\rho_{SP}$ ) după formula:

$$\rho_{SP} = 1 - \frac{6 \sum_{i=1}^n d_i^2}{n(n-1)} \quad (2)$$

unde:

$n$  – numărul de perechi de valori ordonate crescător;

$d_i$ - diferența rangurilor celor două variabile:

$$d_i = rangx_i - rangy_i \quad (3)$$

unde:

$rangx_i$  – rangul valorii  $x_i$  în sistemul ordonat crescător;

$rangy_i$  – rangul valorii  $y_i$  în sistemul ordonat crescător.

Testarea semnificației coeficientului se bazează pe calcularea valorii lui  $t$ :

$$t = \frac{\rho \sqrt{n-2}}{\sqrt{1-\rho^2}} \quad (4)$$

### 2.2.2. Coeficientul de concordanță kendall

Atunci când se lucrează cu două sau mai multe serii de date și se dorește identificarea gradului de similitudine dintre acestea, se aplică **coeficientul** de concordanță **kendall** ( $\tau_K$ )[47],[49]. Acest coeficient este un indicator al apropierii dintre date, poate lua valori cuprinse între +1 (când există concordanță deplină între cele două seturi de date) și -1 (când seriile de valori sunt asociate invers).

$$\tau_K = \frac{2S}{n(n-1)} \quad (5)$$

$S$  – suma algebrică a numerelor de pe coloanele corespunzătoare seriilor de date;

$n$  – numărul de perechi de valori ordonate.

Pentru datele continue, măsurate pe scale de interval și de raport, se aplică analiza corelațională, respectând particularitățile specifice datelor prelucrate. Calculele aferente pot urmări două direcții:

- Calcularea unui indicator al asocierii (corelației);
- Estimarea unui model funcțional care să reprezinte asocierea dintre variabile (regresie).

**2.2.3. Coeficientul de corelație Pearson** [47],[49] reprezintă un indicator sintetic, direcțional al relației de asociere dintre două sau mai multe variabile.

Pentru două variabile normal distribuite (x, y), coeficientul corelației este definit astfel:

$$r_{x,y} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{(\sum (x_i - \bar{x})^2)(\sum (y_i - \bar{y})^2)}} \quad (6)$$

Abateră de la distribuția (repartiția) normală a valorilor corespunzătoare variabilelor x și y determină modificări ale interpretării coeficientului de corelație. Valoarea coeficientului de corelație este în dependență directă cu distribuția perechilor de valori  $(x_i, y_i)$  într-un sistem rectangular de referință XOY.

Valoarea minimă a coeficientului Pearson ( $r_{xy} = 0$ ) nu este un indicator al independenței celor două caracteristici, ci doar indică absența totală a corelației dintre cele două variabile. Acestea pot fi corelate printr-un alt tip de relație funcțională (logistică, binară). Pentru interpretarea valorilor nenule ale coeficienților de corelație, o ilustrare grafică este mai sugestivă – astfel, dacă apare o aliniere perfectă a punctelor de-a lungul unei drepte ascendente, atunci  $r_{xy} = +1$ , iar dacă sensul dreptei este descendent, atunci  $r_{xy} = -1$ .

## 2.3. Analiza de regresie

**Regresia statistică** [50],[51] ca procedură de analiză a datelor, poate fi utilizată:

- în scop *predictiv* → se folosește pentru a realiza prognoze cu privire la evoluția unor serii de date pe baza unuia sau a mai multor predictorii;
- în scop *explicativ* → prezintă influența unui set de factori asupra unei variabile-criteriu. Regresia explicativă izolează (controlează) influența factorilor pentru a vedea în ce măsură aceștia contribuie la predicția criteriului.

De asemenea, există și o diferență ce ține de specificitatea terminologiei. În regresia de tip predictiv, se folosesc termenii predictor și criteriu, în timp ce în regresia de tip explicativ, se utilizează predilect termenii de variabilă independentă și variabilă dependentă. În cazul utilizării regresiei în scop predictiv, se urmărește dezvoltarea unei ecuații care să estimeze cât mai bine evoluția criteriului.

În funcție de numărul de predictorii, regresia se clasifică în:

- regresie liniară simplă (se folosește un singur predictor pentru estimarea criteriului);



- regresia liniară multiplă, numită și regresie multiliniară (mai mulți predictorii sunt incluși în modelele de predicție).

Pentru a putea aplica regresia liniară simplă pe o serie de date este necesar să fie îndeplinite următoarele condiții:

- se verifică normalitatea distribuției (repartiția normală a valorilor) în cazul variabilei independente și a celei dependente;
- relația dintre variabila dependentă și cea independentă trebuie să fie liniară;
- trebuie evitate cazurile extreme și cazurile influente;
- regresia liniară este semnificativă doar atunci când predictorul X și criteriul Y sunt corelați (când se aplică regresia liniară simplă, trebuie îndeplinită condiția existenței unui singur predictor);
- trebuie să fie găsită o funcție liniară care să descrie raportul dintre X și Y.

Pașii generali ai analizei de regresie [50]:

- se identifică predictorul/predictorii și criteriul;
- se concepe modelul liniar de regresie;
- se estimează funcția de regresie pentru seria de date;
- se aplică funcția de regresie;
- se interpretează rezultatele obținute.

Legătura dintre analiza de corelație și de regresie pornește de la premisa că între Y și X există o relație liniară perfectă, în sensul că  $Y = \alpha X + \beta$ , iar coeficientul de corelație se calculează astfel:

$$\rho = \frac{a}{|a|} \quad (7)$$

unde:  $\rho = +1$ , dacă  $\alpha > 0$  și  $\rho = -1$ , dacă  $\alpha < 0$

### 2.3.1. Regresia liniară simplă

Atunci când se analizează modul în care schimbările lui Y sunt asociate cu schimbările lui X, se testează ipoteza asocierii celor două variabile și ulterior, în funcție de confirmarea sau infirmarea ipotezei, se aplică regresia liniară simplă după formula matematică [47]:

$$Y_e = a + bX \quad (8)$$

$Y_e$  – valoarea estimată a variabilei dependente;

$a$  – reprezintă *constantă* - termenul liber al dreptei de regresie (valoarea lui Y pentru  $X=0$ );

$b$  – coeficientul de regresie (adică cantitatea cu care se modifică  $Y$  atunci când  $X$  crește cu o unitate);

$X$  – valoarea variabilei independente.

Valorile pentru  $a$  și  $b$  se obțin prin aplicarea următoarelor formule:

$$b = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sum(x_i - \bar{x})^2} \quad (9)$$

$$a = \bar{Y} - b\bar{x} \quad (10)$$

Valorile obținute pentru  $a$  și  $b$  sunt totuși estimative, acuratețea acestora depinde de gradul de potrivire între tendința norului de puncte (datele reale) și dreapta de regresie (linia celei mai bune ajustări).

În analiza de regresie liniară [50] se acceptă că fiecare variabilă independentă (predictor) se află în relație liniară cu variabila dependentă (criteriu). În ceea ce privește dreapta de regresie,  $b$  indică panta dreptei, iar  $a$  indică punctul intersecție cu axa  $OY$ .

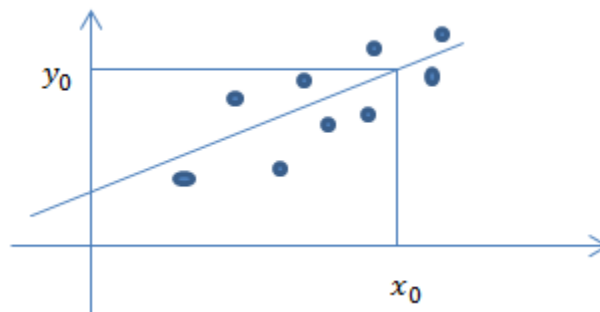


Figura 2. Dreapta de regresie

Pentru a evalua gradul de potrivire, se calculează *eroarea standard a estimației*, definită ca abaterea standard a erorilor de estimare și calculată după formula:

$$s = \sqrt{\frac{\sum(y_i - y_{ei})^2}{n-2}} \quad (11)$$

unde:

$y_{ei}$  – valoarea estimată prin ecuația de regresie a lui  $x_i$

Cu cât eroarea standard are o valoare mai mare, cu atât valorile observate sunt poziționate la o distanță mai mare față de dreapta de regresie [51]. Cu alte cuvinte, valorile observate sunt foarte diferite de valorile adevărate, iar efectul erorilor de estimare asupra datelor este semnificativ mai ridicat.

Ecuția regresiei liniare simple [47],[50-52] determinată prin calcularea parametrilor menționați anterior, este doar o estimare particulară bazată pe caracteristicile seturilor de date incluse în analiză. Prin urmare, trebuie aplicat un test de semnificație care să demonstreze în ce măsură ecuația de regresie calculată  $Y_e = a + bX$  este o estimare a modelului general al regresiei  $Y = \alpha X + \beta$ . Această procedură este numită analiza distribuțională a ecuației de regresie simplă. Ipoteza de la care se pornește este că pentru orice valoare fixată X (variabila independentă), Y (variabila dependentă) este normal distribuită. Adică, se testează dacă  $Exp(a) = \alpha$  și  $Exp(b) = \beta$ .

Apoi, se aplică coeficientului X un test de semnificație (testul t pentru eșantioane independente). Ipoteza de cercetare ( $H_1$ ) presupune că  $\beta \neq 0$ , iar ipoteza de nul ( $H_0$ ) presupune  $\beta = 0$ . Dacă se infirmă ipoteza de cercetare și se confirmă ipoteza de nul, atunci legătura liniară dintre Y și X nu este semnificativă statistic.

**Valorile estimate** de dreapta de regresie sunt de fapt, medii ale valorilor lui Y asociate cu o valoare particulară lui X. Se calculează apoi *abaterea standard* a estimației [47],[50].

$$s(y_0) = s \sqrt{\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum(x_i - \bar{x})^2}} \quad (12)$$

unde:

s – abaterea standard a estimației;

$y_0$  – valoare estimată corespunzătoare lui  $x_0$ .

Formula pentru stabilirea unui prag de semnificație și a intervalului de încredere este:

$$y_0 - t_{1-\frac{\alpha}{2}; n-2} s(y_0), y_0 + t_{1-\frac{\alpha}{2}; n-2} s(y_0) \quad (13)$$

Atunci când valorile lui  $x_0$  se depărtează mai mult de media lui X, intervalul de încredere se mărește, iar predicțiile realizate pe baza modelului sunt însoțite de erori mai mari. În sens invers, predicțiile pentru valorile lui Y vor fi mai corecte atunci când valorile lui  $x_0$  vor fi mai apropiate de media lui X.

### 2.3.2. Regresia liniară multiplă

Regresia liniară multiplă [47],[52] evidențiază relația dintre o variabilă dependentă (explicată, rezultativă) și două sau mai multe variabile independente (explicative, factoriale, predictorii). Scopul regresiei liniare multiple este de a demonstra care dintre predictorii are cea mai mare putere de predicție a criteriului.

Conform modelului liniar general [47],[53] care analizează relația dintre un anumit număr de predictorii ( $V_{I1}, V_{I2}, \dots, V_{In}$ ) și două sau mai multe criterii ( $VD_1, VD_2, \dots, VD_n$ ) ecuația de regresie liniară multiplă se calculează cu ajutorul formulei:

$$y = x\alpha + \varepsilon \quad (14)$$

unde:

$y$  – este variabila dependentă (criteriul);

$x$  – este vectorul variabilelor independente, iar dimensiunea lui este  $1 \times p$ ;

$\alpha$  - este vectorul coeficientului, iar dimensiunea lui este  $p \times 1$ ;

$\varepsilon$  – reprezintă eroarea de măsurare.

Extinzând formula generală a regresiei multiple, obținem următoarea expresie matematică care permite calcularea relației liniare dintre  $y$  și  $x$ :

$$y = a_1x_1 + a_2x_2 + \dots + a_px_p + \varepsilon \quad (15)$$

unde:

$X_p$  – este variabila independentă ( $p$ =valorile lui  $x$  din seria de date);

$a_p$  – valorile coeficientul de regresie pentru variabila independentă.

În primul rând, liniaritatea relației se referă la coeficienți și nu la variabilele propriu-zise. În al doilea rând, atunci când  $x_1$  are constant valoarea egală cu 1 ( $x_1 = 1$ )  $\rightarrow$  rezultă un model liniar care are o constantă (termenul liber al modelului). În al treilea rând, atunci când  $\rho = 2$  și  $x_1 = 1$  se obține modelul liniar simplu, adică dreapta de regresie. În ultimul rând, scopul modelului liniar este de a prezice valorile lui  $Y$  din valorile cunoscute ale predictorilor  $X$ .

În analiză, se pornește de la următoarele date:

- Se notează cu  $y$  un vector de tip  $n \times 1$  care este compus din valorile variabilei  $y$ ;
- Se notează cu  $x$  matrice de tipul  $n \times p$  care este alcătuită din valorile măsurate ale variabilei  $x$ ;
- Se notează cu  $\varepsilon$  un vectori de tip  $n \times 1$  compus din valorile erorilor.

Se testează următoarele ipoteze:

- Matricea de experiență nu trebuie să fie stohastică, ci trebuie să fie fixă:  $X_{n \times p}$ , cu  $n \gg p$ ;
- $X$  este de rang  $p$  (coloanele sunt liniar independente);
- Vectorul de perturbații ( $n$  dimensional)  $\varepsilon$ , este alcătuit din  $n$  variabile aleatorii cu media 0 și aceeași valoare a dispersiei.

$$Exp(\varepsilon) = 0 \quad (16)$$

$$\text{Var}(\varepsilon) = \text{Exp}(\varepsilon\varepsilon') = \sigma^2 I_n \quad (17)$$

unde:

$\sigma^2$  – parametru necunoscut.

O estimație (ajustare) [49],[50] a modelului este reprezentată orice soluție  $\{a, \varepsilon\}$  a ecuației de regresie multiplă. Se consideră soluția  $\mathbf{a}$ , valoarea care minimizează suma pătratelor erorilor. Apoi, se obține valoarea lui  $\mathbf{a}$  aplicând formula:

$$\mathbf{a} = (\hat{X}X)^{-1} \hat{X}y \quad (18)$$

Dacă luăm un coeficient  $a_i$ , calculat după formula prezentată anterior, o interpretare a acestei valori ar indica că modificarea cu 1 a valorii variabilei  $x_i$ , determină o modificare cu  $a_i$  unități în cazul valorii lui  $y$ .

Se estimează *dispersia erorilor* ( $\sigma^2$ ) prin calcularea erorii de ajustare conform formulei:

$$e_i = y_i - \hat{y}_i, i = 1, \dots, n \quad (19)$$

Erorile de ajustare sunt numite **reziduuri**. În ceea ce privește precizia ajustării, se pornește de la principiul că modelul regresiei va fi cu atât mai bun, cu cât variația valorilor estimate va fi mai apropiată de variația valorilor observate. Așadar, se calculează **coeficientul de determinare** ( $R^2$ ) [47],[50] care reprezintă pătratul coeficientul de corelație, se exprimă procentual și se obține aplicând formula:

$$R^2 = \frac{\sum(y_{ei} - \bar{y})^2}{\sum(y_i - \bar{y})^2} \quad (20)$$

unde:

$y_{ei}$  – reprezintă valorile estimate de regresie liniară;

$y_i$  – valorile observate.

Așadar, calcularea coeficientului de determinare stabilește cât din variația datelor inițiale este explicată de dreapta de regresie (adică variabila  $x$  – variabila independentă). Cu alte cuvinte, coeficientul de determinare exprimă cât din variația variabilei dependente este explicată de ecuația estimată.

$$R^2 = \frac{\sum_i(\hat{y}_i - \bar{y})^2}{\sum_i(y_i - \hat{y}_i)^2} \quad (21)$$

unde:

$\hat{y}_i$  – valoarea ajustată a variabilei  $y$ .

Coeficientul de determinare ( $R^2$ ) crește atunci când se adaugă mai mulți predictorii în model. Astfel, apare un efect de supraestimare în modelele extinse, cele cu mai mulți predictorii, și din acest motiv, apare necesitatea de ajustare a acestui coeficient.

$$\overline{R^2} = R^2 - \frac{p-1}{n-p} (1 - R^2) \quad (22)$$

Pentru a se testa gradul de asociere dintre  $y$  și ansamblul valorilor lui  $x$ , se aplică coeficientul de corelație Pearson între  $y$  și o combinație de variabile  $x$ . Dacă valoarea este apropiată de 0, atunci regresia este ne semnificativă statistic, iar valorile prezise de modelul de regresie nu diferă de cele care ar fi fost obținute prin ghicire aleatorie.

Pentru a testa ipotezele, se calculează:

- *suma pătratelor globale:*

$$SP_g = \sum_i (y_i - \bar{y})^2 \quad (23)$$

cu gradele de libertate:  $v_g = n - 1$

- *suma pătratelor regresiei:*

$$SP_{reg} = \sum_i (\hat{y}_i - \bar{y})^2 \quad (24)$$

cu gradele de libertate:  $v_{reg} = p - 1$

- *suma pătratelor reziduale:*

$$SP_{rez} = \sum_i (y_i - \hat{y}_i)^2 \quad (25)$$

cu gradele de libertate:  $v_{rez} = n - p$

Apoi, se aplică *testul F* de semnificație globală în care ipotezele testate sunt [50]:

$$H_0: \alpha_1 = \alpha_2 = \dots = \alpha_p = 0$$

$$H_1: \exists i, \text{ astfel încât } \alpha_i \neq 0$$

Se urmărește confirmarea ipotezei de cercetare și respingerea ipotezei de nul. Astfel, respingerea ipotezei de cercetare demonstrează că datele observate nu permit identificarea unui model liniar valid, astfel că regresia nu poate fi folosită cu scopul predicției cum s-a propus inițial.

Conform [47] intervalul de încredere pentru parametrul  $\alpha$  este exprimat prin relația:

$$\alpha_i - t_{1-\frac{\alpha}{2}; n-2} s(\alpha_i) \leq \alpha_i < \alpha_i + t_{1-\frac{\alpha}{2}; n-2} s(\alpha_i) \quad (26)$$

Observăm că perechile de valori se distanțează în mai mare sau mai mică măsură de dreapta de regresie (numită și linia celei mai bune ajustări). Aceste deviații se datorează faptului că există și alte surse de influență în ceea ce privește schimbările variabilei dependente, surse care nu sunt

cuprinse în ecuația de regresie. Deviațiile pot fi explicate și de existența unor erori de măsurare numite *reziduuri* ( $\varepsilon$ ). Acestea sunt „deviații care reprezintă diferența dintre o valoare observată și valoarea prognozată de model” [54]. Se trece apoi la etapa de *analiză a reziduurilor*, în care se testează ipoteza normalității repartiției erorilor. Se analizează histograma reziduurilor sau diagrame ale reziduurilor în raport cu valorile estimate ale variabilei independente. Se identifică valorile aberante și se exclud din seriile de date.

O situație particulară în regresia multiplă este *multicoliniaritatea* [45],[50] adică existența unui grup de variabile independente cu corelații foarte ridicate (peste 0,80 sau 0,90). Dacă se include în model una dintre variabilele din grup, restul variabilelor nu mai aduc o informație semnificativă. Se produce astfel o supraevaluare a coeficientului de determinare, ceea ce poate denatura acuratețea modelului de predicție.

Pentru a detecta existența multicoliniarității [45], se analizează matricea de corelație dintre variabilele  $x$ . În felul acesta, se pot identifica perechile de variabile independente puternic corelate între ele. Eliminarea multicoliniarității se realizează prin reținerea în model doar a uneia dintre variabilele independente puternic corelate.

Până în acest punct al analizei, s-au realizat majoritatea prelucrărilor matematice. În funcție de valorile coeficienților calculați, se poate identifica mulțimea de variabile independente (asocierea dintre predictorii) care permite estimarea unui model liniar semnificativ și adecvat valorilor observate ale lui  $Y$ .

Etapile selectării celui mai bun model de regresie [45],[50]:

1. Identificarea tuturor variabilelor independente posibile;
2. Specificarea criteriului de selecție a celei mai bune regresii;
3. Specificare strategiei pentru selectarea variabilelor independente;
4. Realizarea estimării și analiza modelului;
5. Evaluarea gradului de precizie a estimării pentru modelul ales.

În general, se estimează toate modelele posibile de regresie și apoi, se analizează valorile lui  $R^2$  corespunzătoare acestora și se extrage submulțimea de variabile (adică modelul) pentru care există un nivel ridicat de concordanță între numărul de variabile și mărimea coeficientului de determinare.

## Capitolul 3. APLICAREA METODELOR ȘI TEHNICILOR DE ANALIZĂ A SEMNALULUI VOCAL

### 3.1. Abordări în calcularea coeficientului de corelație – $r$

Formula de calcul (1) propusă de Pearson în 1895, exprimă un coeficient de corelație  $r$  invariabil transformărilor liniare ale oricăreia dintre variabile. În această formulă, la numărător scorurile brute sunt centrate prin extragerea mediei fiecărei variabile și adunarea sumelor produselor încrucișate. Numitorul, pe de altă parte, ajustează variabilele astfel încât să fie exprimate în unități egale. Utilizând inegalitatea Cauchy-Schwartz se poate demonstra că valoarea absolută a numărătorului este mai mică sau egală decât cea a numitorului astfel încât, pentru  $r$ , limitele  $\pm 1$  sunt stabilite. Joseph Lee Rodgers și W. Alan Nicewander [60] prezintă treisprezece modalități de abordare a calculului coeficientului de corelație. În prima abordare, corelația este reprezentată ca funcție a scorurilor și mediilor brute.

Această modalitate de calcul a coeficientului de corelație am utilizat-o într-unul dintre studiile [38] pe care le-am realizat și care a avut ca scop analiza corelațională a formațiilor vocalelor limbii române, în timpul pronunției izolate a vocalelor. Aplicabilitatea rezultatelor a fost dată de posibilitate extrapolării acestora la nivelul de analiză a pronunției, în sinteza vorbirii, în detectarea vorbirii patologice și în recunoașterea emoțiilor din discursul vorbitorilor.

Baza de date [61] ce a fost utilizată era alcătuită din înregistrări ale vocalelor limbii române: /a/, /e/, /i/, /o/, /u/, /ă/ și /â/. Fiecare vocală a fost înregistrată de trei ori și a avut durată medie a intervalului de timp 3s. Acest interval de timp nu a fost ales aleatoriu, ci s-a urmărit evitarea situației în care pronunția susținută a vocalelor ar putea fi afectată de fluctuațiile din voce. Înregistrările au fost realizate prin participarea voluntară a unor vorbitori de gen masculin și feminin, cu vârsta cuprinsă între 25-30 de ani, din zona Iași-Vaslui.

Rezultatele obținute indică faptul că nu apar diferențe semnificative statistic între nivelul de corelație al formațiilor vocalelor pentru vorbitorii de gen masculin comparativ cu vorbitorii de gen feminin. De asemenea, atât în cazul vorbitorilor de gen masculin, cât și în cazul celor de gen feminin, formații vocalelor /a/, /e/, /o/, /u/, /ă/ și /â/ prezintă un nivel de corelație semnificativ mai ridicat comparativ cu cel al vocalei /i/.

Ulterior, studiul [60] a fost continuat, în plus față de cele șapte vocale ale limbii române, au fost incluse în analiză două propoziții simple: /vina mama/ și /cine a făcut asta/. În procesul de analiză a datelor, s-a ținut cont de două aspecte: 1 – poziția vocalei în interiorul cuvântului și 2 – gradul de accentuare al vocalei (accentuată sau neaccentuată). S-a comparat nivelul de corelație al



vocalelor în timpul pronunției susținute cu nivelul de corelație pe care formanții vocalelor l-au avut atunci când pronunția s-a realizat în interiorul cuvintelor.

Rezultatele au indicat că în cazul vocalei /a/, apar diferențe semnificative între nivelul de corelație al formanților în timpul pronunției susținute comparativ cu pronunția în interiorul cuvintelor, în sensul că în timpul pronunției susținute, vocala /a/ a prezentat un nivel de corelație semnificativ mai ridicat.

Pentru structura /vine mama/, vocala accentuată (a1) are un nivel de corelație al formațiilor semnificativ mai scăzut decât vocala neaccentuată (a2). Astfel, constatăm că nivelul de corelație al formanților este puternic influențat de contextul în care apare vocala (vocala susținută, izolată, accentuată sau neaccentuată).

Cea de-a doua abordare din [62] exprimă calculul coeficientul de corelație ca versiune normalizată a covarianței dintre X și Y,

$$r = \frac{s_{XY}}{s_X s_Y} (1)$$

unde:

$s_{XY}$  – covarianța eșantionului;

$s_X s_Y$  – abaterile standard pentru variabilele X și Y.

**Corelația și covarianța** reprezintă modalități de estimare a gradului de asociere liniară dintre variabile. Covarianța se calculează ca fiind *suma produselor încrucișate* ale variabilelor aleatorii X și Y, adică exprimă măsura de variație comună dintre acestea. Dacă două variabile aleatorii au comportamente similare, covarianța este pozitivă, iar atunci când valorile ridicate ale unei variabile corespund valorilor scăzute ale celeilalte variabile și invers (adică comportamentele sunt opuse), covarianța este negativă.

Covarianța reprezintă variația unei variabile cu ea însăși [62]. Însă, covarianța a două variabile nu reprezintă un indicator statistic util pentru descrierea gradului de asociere, deoarece este direct influențată de scalele de măsurare specifice pentru X și Y. Din acest motiv, se divizează covarianța cu produsul abaterilor standard corespunzătoare variabilelor, pentru a se obține o valoare ce aparține intervalului  $\pm 1$ .

### 3.2. Studiu preliminar asupra corelației formanților vocalelor din limba română

În analiză, am utilizat propoziția /vine mama/, frecvent folosită în discursul cotidian al vorbitorilor și am urmărit variațiile corelației vocalelor în funcție de emoția exprimată (bucurie, tristețe, furie și starea de neutralitate emoțională).

Concluziile la care am ajuns în urma demersului de cercetare pentru acest studiu au indicat că:

- I. Emoțiile constituie o variabilă cu o influență puternică asupra nivelului de corelație al formanților vocalelor din vorbirea curentă a indivizilor.
- II. În cazul emoțiilor cu valență negativă (*tristețea* și *furia*), formanții de ordin superior (F3 și F4) prezintă un nivel de corelație semnificativ mai ridicat comparativ cu formanții de ordin inferior (F1 și F2).
- III. Pentru tonul neutru, variația nivelului de corelație al perechilor de formanți este mai omogenă, în sensul că s-au observat mai puține oscilații de la valori extrem de scăzute pentru o anumită vocală, la valori extrem de ridicate.

Pornind de la studiul [40] am preluat designul de cercetare și am continuat analiza corelațională a formanților utilizând baza de date din studiul anterior. În articolul anterior menționat, s-a investigat măsura în care apar modificări în variația formanților atunci când discursul vorbitorilor exprimă bucurie, tristețe și furie comparativ cu situația în care vorbirea este neutră din punct de vedere emoțional. Rezultatele obținute în urma aplicării funcției *Corr* și calculării coeficientului de regresie liniară au dus la formularea următoarelor concluzii:

1. Formantul F4 nu a avut modificări semnificative ale variației în funcție de emoțiile exprimate de vorbitori.
2. F0 (pitch) a avut valori semnificativ mai ridicate pentru bucurie în cazul tuturor vocalelor.
3. Pentru tristețe, s-a observat o scădere ușoară a valorilor frecvenței fundamentale.

Am continuat să folosesc baza de date SRoL [61] și am calculat *media* pentru fiecare formant F1-F4 și pentru frecvența fundamentală (F0) corespunzătoare fiecărei vocale, în cazul fiecărui vorbitor, pentru toate cele patru emoții.

Formații au fost definiți ca fiind:

$$F_{k,v}(P_x v_y) \quad (15)$$

$$F_{k,v}(P_1 v_1, P_2 v_1, \dots, P_n v_1) \quad (16)$$

unde:

k – reprezintă ordinul formantului,  $k \in \{0, \dots, 4\}$ ;

v – este vocala (*i, e, a1, a2*);

$P_x v_y$  – reprezintă pronunția *x* pentru vorbitorul *y*.

Calcularea *mediei* s-a realizat prin aplicarea formulei:

$$\overline{F_{k,v}(P_x v_y)} = \frac{1}{n} \sum_{n=1}^N F_{k,v}(P_n v_y) \quad (17)$$

unde:

$P_n v_y$  – este numărul pronunției ( $n=1, \dots, N$ ) pentru un vorbitor  $v_y$ ;

$n$  – reprezintă numărul total de valori.

Rezultatele pe care le-am obținut au dus la formularea următoarelor concluzii:

(1) Coeficienții de corelație dintre frecvența fundamentală și formanți au avut valori semnificativ mai scăzute comparativ cu valorile de corelație obținute pentru cuplurile de formanți, indiferent de tipul emoției analizate;

(2) Valorile de corelație dintre formanții vocalei /i/, pentru emoția *furie*, sunt semnificativ mai scăzute comparativ cu valorile obținute pentru celelalte vocale (explicațiile sugerate de studiile anterioare [71],[72] susțin că mecanismul specific de articulare a sunetului /i/ și de pronunție a vocalei ar determina un grad de corelație al formanților mai slab);

(3) Pentru starea de *neutralitate emoțională*, nivelul de corelație dintre  $F_0$  și formanți este mai ridicat decât cel care a fost obținut pentru emoțiile *bucurie*, *tristețe* și *furie*. În plus, am constatat că formanții de ordin superior sunt mai puternic asociați comparativ formanții de ordin inferior. De asemenea, pentru starea de neutralitate emoțională, variația nivelului de corelație al perechilor de formanți este mai omogenă, în sensul că au existat mai puține oscilații de la valori extrem de scăzute la valori extrem de ridicate, de la o vocală la alta;

(4) Valorile coeficientului de determinare au fost scăzute atât pentru /a1/, cât și pentru /a2/ indiferent de tipul emoției. Cea mai mare valoare a lui  $R^2$ , a fost raportată pentru vocala /a2/ asociată emoției *furie*. Astfel, pentru vocala /a2/, frecvența fundamentală explică 27% din variația lui  $F_1$ ;

(5) În ceea ce privește vocala /a1/, frecvența fundamentală este un bun predictor pentru  $F_1$ , doar în cazul emoțiilor *tristețe* și *furie*, iar pentru  $F_2$  este un predictor eficient atunci când se asociază cu *furia* și tonul *neutru*;

(6) În privința vocalei neaccentuate /a2/, frecvența fundamentală este un predictor eficient și pentru  $F_1$ , și pentru  $F_2$  numai în cazul emoției *furie*.

## **Capitolul 4. ANALIZA CORELATIVĂ A VORBIRII EMOTIVE / EMOTIILOR LA NIVELUL SEMNALULUI VOCAL – Contribuții**

Conform Dicționarului Merriam-Webster (2002) [73], emoțiile sunt „reacții mentale conștiente experimentate subiectiv ca sentimente puternice, de obicei îndreptate către un obiect specific și însoțite de obicei de schimbări fiziologice și comportamentale în corp”. Klaus R. Scherer [74] în 2003, a propus o definiție mai cuprinzătoare, afirmând că emoțiile sunt „episoade relativ scurte de răspuns sincronizat al tuturor sau al majorității subsistemelor organismului ca răspuns la evaluarea unui eveniment extern sau intern de importanță majoră”.

Emoțiile rezultă din evaluările cognitive ale evenimentului implicat în declanșarea reacției emoționale. În ceea ce privește efectul lor, emoțiile au un puternic impact comportamental, generând schimbări specifice adaptate contextului. Prin urmare, emoțiile: a) evidențiază semnificația sau relevanța unui stimul; b) pregătesc organismul pentru acțiune prin mobilizarea resurselor fiziologice disponibile; c) indică starea emoțională a individului, precum și intențiile sale față de ceilalți, facilitând astfel comunicarea (Scherer, 2003)[74].

Rezumând, procesualitatea emoțiilor implică: (1) existența unui stimul/eveniment declanșator; (2) evaluarea cognitivă a acestuia; (3) apariția unor modificări fiziologice (puls, tensiune arterială, tremur, modificări ale frecvenței respiratorii etc.); (4) manifestarea unor modificări comportamentale asociate (reacția de luptă, atac sau îngheț) și (5) evaluarea stării percepute sau a experienței subiective.

Literatura anterioară a prezentat două direcții principale de cercetare care integrează majoritatea studiilor privind emoțiile, oferind date științifice valide și permițând stabilirea a două modele teoretice explicative: 1) teoriile discrete și 2) teoriile dimensionale ale emoțiilor. În cadrul primei, primele contribuții au fost observațiile empirice ale lui Darwin privind rolul evolutiv al emoțiilor și caracterul lor universal. Paul Ekman [75] a propus modelul celor șase emoții de bază, înnăscute și universale: bucuria, tristețea, furia, frica, surpriza și dezgustul. Robert Plutchik (1997)[76] a adăugat acceptarea și anticiparea în categoria emoțiilor fundamentale, subliniind rolul lor primordial în procesul de mobilizare a resurselor de adaptare. Pe de altă parte, teoriile dimensionale explică emoțiile analizând modularea lor pe următoarele trei dimensiuni: (1) valența; (2) activare; (3) intensitate. Autorii care susțin această paradigmă au analizat un număr relativ redus de emoții caracterizate prin modele specifice de răspunsuri la nivel fiziologic și în semnalul vocal al vorbitorului.

Privind exprimarea vocală a emoțiilor, pentru a explica acest proces, Klaus R. Scherer (2003)[74] a adaptat un model de lentilă brunswikiană. Modelul sintetizează procesul urmând trei

etape: codificarea, transmiterea și decodificarea. În timpul primei etape, emoția este încorporată în semnalul vocal, în următoarea etapă are loc transmiterea conținutului semantic și emoțional, iar a treia și ultima etapă constă în decodificarea și evaluarea informației. Atunci când mesajul ajunge la receptor, acesta întreprinde un dublu proces de evaluare: pe de o parte, este analizat conținutul informațional, iar pe de altă parte, se realizează recunoașterea stării emoționale.

În cadrul acestui proces, un factor de relevanță majoră este reprezentat de indicii indicatori, care permit modelarea semnalului vocal, ținând cont de modelele specifice ale emoției asociate. În funcție de gradul de obiectivitate al măsurătorii, Klaus R. Scherer (2003)[74] clasifică parametrii vocali în: parametri acustico-fonici (rata de vorbire, F0, intensitatea, panta spectrală, raportul armonic/zgomot, formanți și lățimea de bandă a formanților) și parametri acustici (intensitatea percepută, înălțimea percepută, ritmul perceput).

În conformitate cu acești parametri, tristețea se caracterizează prin: o rată de vorbire mai mică (mai puține silabe pe secundă), valori mai mici în comparație cu alte emoții pentru media F0, deviația F0 și intervalul F0, un nivel mai scăzut de amplitudine și armonice, o lățime de bandă mai mare și un timp mai lung de pronunție a silabelor (Scherer, 2003)[74].

Pentru limba română, primele studii privind influența emoțiilor la nivelul corelației formanților vocalelor în timpul pronunției dinamice au fost realizate de (Teodorescu & Feraru, 2007[71]; Teodorescu, Feraru, & Zbancioc, 2010[77]; Teodorescu, Feraru, Zbancioc & Trandabăț, 2014[78]; Apopei & Păduraru, 2015[36]; Ghelțu & Teodorescu, 2018[40]; Teodorescu, 2019[79]; Ghelțu, 2020[80]).

Unul dintre primele studii care a abordat problema în cauză a fost „*A comparative study of speech rate estimation techniques*” [81]. Autorii din [81] au efectuat o revizuire comparativă a tehnicilor de estimare a ritmului de vorbire, în urma căreia au constatat că poziția vocalelor în cuvinte conduce la un nivel semnificativ ridicat de intercorelație a formanților (F1-F3).

Într-un alt studiu, „*A fuzzy information space approach to speech signal non-linear analysis*” [82], o metodologie fundamentată (bazată pe corelația lui Pearson) a evidențiat necesitatea utilizării unor sisteme de analiză adecvate pentru caracteristicile de bază ale semnalului audio ales. Manifestând interes pentru comportamentul dinamic al semnalului audio, autorii au optat pentru utilizarea metodei fuzzy. Setul de date al semnalului vocal utilizat este un sistem complex și neliniar, care se caracterizează, de asemenea, prin dinamism, deoarece variază în timp. Această variabilitate a datelor reduce valoarea predictivă a analizei. Cu toate acestea, incertitudinea datelor se diminuează datorită specificității metodei aplicate.

Printre lucrările relevante se numără și [83], care a descris probleme și metode de rezolvare a problemelor de recunoaștere a vorbirii. Autorii au făcut o analiză a aproximativ 15.000 de cuvinte. Metoda utilizată pentru recunoaștere a fost modelul Markov; prin urmare, ei s-au

concentrat pe îmbunătățirea performanțelor sistemului. Prima metodă a fost crearea unui hibrid neuro-statistic prin integrarea unui perceptron multistrat (MLP) ca estimator de probabilitate. A doua metodă a fost construirea unui hibrid fuzzy static, FHMM, bazat pe o măsură de similaritate fuzzy în locul măsurii probabilistice specifice modelului statistic obișnuit. Prin utilizarea celor două metode, s-a observat o rată de recunoaștere cuprinsă între 92% și 98%.

„Există multe motive pentru a studia intercorelația formaților. În primul rând, aceasta poate dezvălui informații despre mecanismele vorbirii, utile pentru înțelegerea fiziologiei vorbirii și pentru posibilele aplicații în sinteza și recunoașterea vorbirii [84] și în diagnosticarea patologiei limbajului [85]. În al doilea rând, corelațiile încrucișate (cross-correlations) pot contribui la îmbunătățirea analizei discursului emoțional”[84].

În mod surprinzător, o căutare detaliată a literaturii de specialitate privind caracteristicile vorbirii a găsit puține informații privind interdependențele dintre formați și corelațiile încrucișate ale formaților în timpul vorbirii. Unele dintre articolele care abordează un subiect similar cu cel discutat în lucrarea de față și care prezintă unele date relevante sunt [86-90]. Gomez-Vilda et al. [86] au studiat corelațiile și dependențele dintre activitatea mioelectrică facială și formați. Lucrarea lui Inbar și Eden [87] este, din punctul de vedere al studiului nostru, una dintre cele mai relevante contribuții, deoarece relevă o legătură directă între activitatea EMG în mușchii implicați în producerea vorbirii și corelația acesteia cu parametrii vocii, inclusiv cu formații. Williamson et al. [88-90], folosind o tehnică pe care au dezvoltat-o pentru prima dată pentru analiza semnalelor EEG [89], pornesc de la principiul că „formați sunt corelați între ei atât în domeniul frecvenței, cât și în cel al timpului” [88]; prin urmare, studiul lor este, de asemenea, paralel cu al nostru. Acești autori au făcut remarcă importantă că "autocorelațiile și corelațiile încrucișate ale semnalelor măsurate pot dezvălui parametrii ascunși în sistemele stocastic-dinamice" legate de "coordonarea articulatorie" [88]. În cercetările din [88-90] s-a optat pentru analiza markerilor ce construiesc vorbirea.

Există mai multe motive pentru a studia dependențele și intercorelațiile dintre formați. În primul rând, acesta poate dezvălui informații despre mecanismele de vorbire, care sunt utile pentru înțelegerea fiziologiei vorbirii și pentru posibile aplicații în sinteza vorbirii și diagnosticarea tulburărilor de limbaj. Corelațiile dintre formați dezvăluie corelații în mișcarea și modificările de formă ale laringelui. În al doilea rând, corelațiile încrucișate între formați pot contribui la îmbunătățirea analizei vorbirii emoționale.

Există, de asemenea, multe lucrări privind utilizarea PCA în vorbire, legate de formați, de exemplu [91, 92]; aceste studii au arătat în mod clar că seriile temporale ale formaților vorbirii nu sunt independente; totuși, nu am găsit niciun studiu care să obțină regresii și să tragă concluzii privind mecanismele de producție a vorbirii.

### 4.1.1. Baza de date

Formantul este o concentrare de energie acustică în jurul unei anumite frecvențe din unda de vorbire. Frecvențele de rezonanță ale tractului vocal al vorbitorului sunt estimate prin predicție liniară la intervale regulate în semnal, iar formații sunt identificați din vârfuri. O utilizare tipică constă în citirea valorilor urmelor formațiilor ca alternativă la măsurarea frecvențelor formațiilor pe spectrograma propriu-zisă.

Baza de date conține înregistrări ale vocalelor limbii române: /a/, /e/, /o/, /ă/ și /â/, toate fiind înregistrate de trei ori pentru fiecare dintre subiecți. După faza inițială de înregistrare, datele au fost analizate pentru a determina nivelul de zgomot. Acesta se referă la semnalul perturbator care însoțește implicit semnalul util. Nivelul mediu de zgomot observat a fost de 25 dB.

Baza de date utilizată în această cercetare a fost realizată din două părți: o parte conține baza de date SRoL [61] și o parte este alcătuită din înregistrări proprii. Au rezultat un număr de 450 de înregistrări..

Fișierele au fost segmentate manual pentru determinarea segmentelor vocalice (detaliate în tabelul din Anexa 2). Pentru vocalele din propoziția "*vine mama*", numărul mediu de eşantioane per segment a fost de 9,4-25, iar răspândirea a fost de 4,5-7.

Procedura de prelucrare a datelor a implicat extragerea vocalelor /i/, /e/, /a1/ și /a2/. Segmentarea a fost realizată manual (Figura 1), folosind pachetul software Praat [94], [95].

„Baza de date a fost organizată în funcție de cele patru emoții incluse în analiză, iar pentru fiecare vocală au fost extrase formații F1-F4 și frecvența fundamentală (F0). Am constatat că mărimea frecvenței fundamentale nu corespundea cu cea a formațiilor. Deși fereastra de timp a fost aceeași (aproximativ 6ms), numărul de puncte extrase pentru F0 a fost diferit de cel al formațiilor. Din acest motiv, am ales să calculăm media pentru fiecare formant F1-F4 și pentru frecvența fundamentală (F0) corespunzătoare fiecărei vocale, în cazul fiecărui vorbitor, pentru toate cele patru emoții” [80].

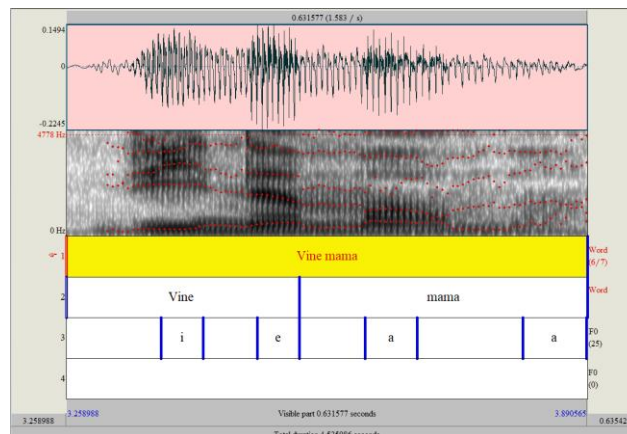


Figura 1. Segmentare manuală a semnalului vocal prelucrat

Pentru formanți, am folosit următoarele formule:

$$F_{k,v}(P_x v_y) \quad (1)$$

$$F_{k,v}(P_1 v_1, P_2 v_1, \dots, P_n v_1) \quad (2)$$

unde:

k – numeral de ordine al formantului,  $k \in \{0, \dots, 4\}$ ;

v – vocala (i, e, a1, a2);

$P_x v_y$  – pronunția x pentru vorbitorul y.

Media a fost calculată prin aplicarea formulei:

$$\overline{F_{k,v}(P_x v_y)} = \frac{1}{n} \sum_{n=1}^N F_{k,v}(P_n v_y) \quad (3)$$

unde:

$P_n v_y$  – este numărul pronunției ( $n=1, \dots, N$ ) pentru vorbitorii  $v_y$ ;

n – este numărul total de valori.

Formanții F1, F2, F3, F4 reprezintă valorile inițiale care au fost corelate. Valorile formanților sunt notate cu x și y, reamintim formula pentru calculul coeficientului de corelație Pearson:

$$C_{xy} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 (y_i - \bar{y})^2}} \quad (4)$$

n reprezentând numărul total de valori observate.

După ce s-a realizat concatenarea, s-a obținut un fișier de serii temporale pentru fiecare vocală și fiecare emoție corespunzătoare. Toți vorbitorii (participanții) au fost incluși în fișier în ordinea numerotării în care au fost procesate înregistrările lor. Astfel, în urma asocierii formanților (F1-F4) ale vocalelor /i/, /e/, /a1/ și /a2/ din structura /vine mama/ au fost obținute 64 de serii temporale, câte una pentru fiecare dintre cele patru emoții utilizate în analiză. Acest rezultat a fost obținut prin aplicarea formulei (5)

$$F_k(v(E = N)) \quad (5)$$





## 4.2. Rezultate

### 4.2.1. Vocale susținute

Au fost determinate valorile de corelație pentru cuplurile de formanți F1, F2, F3 și F4, F1 cu F2, F1 cu F3, F1 cu F4, F2 cu F3, F2 cu F4 și F3 cu F4. Tabelele I și II prezintă valorile absolute ale formanților în cadrul corelației; se observă că semnul pozitiv sau negativ poate fi irelevant pentru scopuri specifice.

TABEL I. VALORILE ABSOLUTE ALE DATELOR DE CORELAȚIE OBTINUTE DE SUBIECȚII DE SEX MASCULIN (preluat din [38])

	F1,F2	F1,F3	F1,F4	F2,F3	F2,F4	F3,F4
<i>A</i>	0.748	0.664	0.626	0.792	0.785	0.746
<i>E</i>	0.505	0.730	0.688	0.546	0.416	0.703
<i>I</i>	0.342	0.576	0.743	0.487	0.289	0.598
<i>O</i>	0.761	0.700	0.664	0.786	0.806	0.788
<i>U</i>	0.662	0.601	0.602	0.797	0.753	0.755
<i>Ă</i>	0.806	0.722	0.34	0.801	0.789	0.780
<i>Î</i>	0.746	0.725	0.670	0.738	0.688	0.787

TABEL II. VALORILE ABSOLUTE ALE DATELOR DE CORELAȚIE OBTINUTE DE SUBIECȚII DE SEX FEMININ (preluat din [38])

	F1,F2	F1,F3	F1,F4	F2,F3	F2,F4	F3,F4
<i>A</i>	0.644	0.524	0.420	0.667	0.493	0.667
<i>E</i>	0.473	0.613	0.502	0.674	0.561	0.763
<i>I</i>	0.398	0.364	0.476	0.547	0.470	0.652
<i>O</i>	0.821	0.618	0.542	0.622	0.542	0.659
<i>U</i>	0.835	0.636	0.556	0.715	0.632	0.698
<i>Ă</i>	0.724	0.471	0.363	0.521	0.408	0.570
<i>Î</i>	0.747	0.634	0.469	0.657	0.474	0.595

Valorile de corelație ridicate au fost observate pentru vocale: /a/, /e/, /o/, /ă/ și /â/. Valori scăzute de corelație au fost observate mai ales pentru vocala /i/.

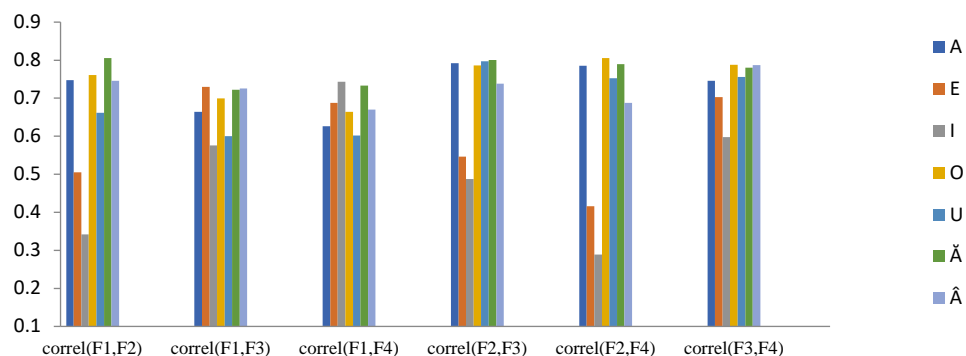


Figura 2. Valorile de corelației absolute pentru vocalele subiecților de sex masculin. preluat din [38]

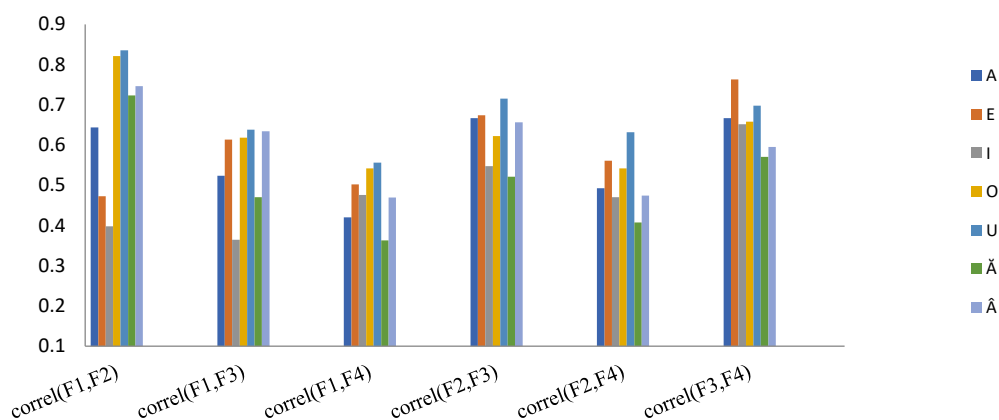


Figura 3. Valorile de corelație pentru formații vocalelor, subiecți de sex feminin. preluat din [38]

În figurile 2 și 3, valorile de corelație pentru vocalele susținute /i/ și /î/ au un interval mai larg decât în cazul vocalelor /a/, /e/, /o/, /u/. Mediile valorilor semnate ale unora dintre corelațiile formațiilor /i/ și /î/ sunt mai apropiate de zero, dar acest lucru se datorează faptului că unii vorbitori au valori de corelație pozitive, alții valori de corelație negative, în timp ce alții au atât valori de corelație negative, cât și pozitive, în funcție de pronunția lor. Tabelele III și IV conțin valorile relative ale datelor analizate, rezultând valori scăzute ale corelației.

TABEL III. VALORILE RELATIVE ALE DATELOR DE CORELAȚIE OBTINUTE DE LA SUBIECȚII DE SEX MASCULIN (preluat din [38])

	F1,F2	F1,F3	F1,F4	F2,F3	F2,F4	F3,F4
<b>A</b>	0.748	0.664	0.626	0.792	0.785	0.746
<b>E</b>	0.477	0.730	0.688	0.533	0.387	0.703
<b>I</b>	0.174	0.574	0.743	0.450	0.114	0.591
<b>O</b>	0.757	0.700	0.664	0.786	0.806	0.788
<b>U</b>	0.662	0.601	0.602	0.797	0.753	0.744
<b>Ă</b>	0.806	0.722	0.734	0.801	0.789	0.780
<b>Î</b>	0.746	0.725	0.670	0.738	0.688	0.787

**TABEL IV. VALORILE RELATIVE ALE DATELOR DE CORELAȚIE OBTINUTE DE LA SUBIECȚII DE SEX FEMININ**  
(preluat din [38])

	<b>F1,F2</b>	<b>F1,F3</b>	<b>F1,F4</b>	<b>F2,F3</b>	<b>F2,F4</b>	<b>F3,F4</b>
<b>A</b>	0.644	0.524	0.420	0.667	0.493	0.667
<b>E</b>	0.444	0.613	0.502	0.674	0.553	0.763
<b>I</b>	0.048	0.295	0.469	0.544	0.364	0.652
<b>O</b>	0.821	0.602	0.524	0.622	0.534	0.659
<b>U</b>	0.835	0.638	0.544	0.715	0.616	0.698
<b>Ă</b>	0.723	0.431	0.346	0.385	0.380	0.544
<b>Î</b>	0.714	0.621	0.454	0.653	0.474	0.595

Valorilor din tabelele V și VI, se observă diferențele relative care rezultă din variația valorilor formanților pentru /i/.

**TABEL V. DIFERENȚELE DINTRE VALORILE RELATIVE ALE DATELOR DE CORELAȚIE OBTINUTE DE LA SUBIECȚII DE SEX MASCULIN** (preluat din [38])

	<b>F1,F2</b>	<b>F1,F3</b>	<b>F1,F4</b>	<b>F2,F3</b>	<b>F2,F4</b>	<b>F3,F4</b>
<b>A</b>	0.000	0.000	0.000	0.000	0.000	0.000
<b>E</b>	-0.028	0.000	0.000	-0.013	-0.029	0.000
<b>I</b>	-0.168	-0.002	0.000	-0.037	-0.175	-0.007
<b>O</b>	-0.004	0.000	0.000	0.000	0.000	0.000
<b>U</b>	0.000	0.000	0.000	0.000	0.000	-0.011
<b>Ă</b>	0.000	0.000	0.000	0.000	0.000	0.000
<b>Î</b>	0.000	0.000	0.000	0.000	0.000	0.000

**TABEL VI. DIFERENȚELE DINTRE VALORILE RELATIVE ALE DATELOR DE CORELAȚIE OBTINUTE DE LA SUBIECȚII DE SEX FEMININ** (preluat din [38])

	<b>F1,F2</b>	<b>F1,F3</b>	<b>F1,F4</b>	<b>F2,F3</b>	<b>F2,F4</b>	<b>F3,F4</b>
<b>A</b>	0.000	0.000	0.000	0.000	0.000	0.000
<b>E</b>	-0.029	0.000	0.000	0.000	-0.008	0.000
<b>I</b>	-0.350	-0.069	-0.007	-0.003	-0.106	0.000
<b>O</b>	0.000	-0.016	-0.019	0.000	-0.008	0.000
<b>U</b>	0.000	0.000	-0.012	0.000	-0.016	0.000
<b>Ă</b>	-0.001	-0.039	-0.018	-0.136	-0.028	-0.026
<b>Î</b>	-0.033	-0.013	-0.015	0.000	0.000	0.000

În tabelele V și VI sunt exemplificate diferențele de corelație dintre valorile formanților. Valori semnificative au fost găsite pentru subiecții de sex feminin și pentru vocalele /i/, /ă/ și /â/, datorită diferențelor de nuanță a pronunției și de tonalitate.

Pentru următoarea fază a analizei noastre, am concatenat seriile temporale ale formanților specifice vocalelor. Acestea au fost corelate ulterior în referință cu întreaga bază de date, iar rezultatele sunt prezentate în tabelul VII.

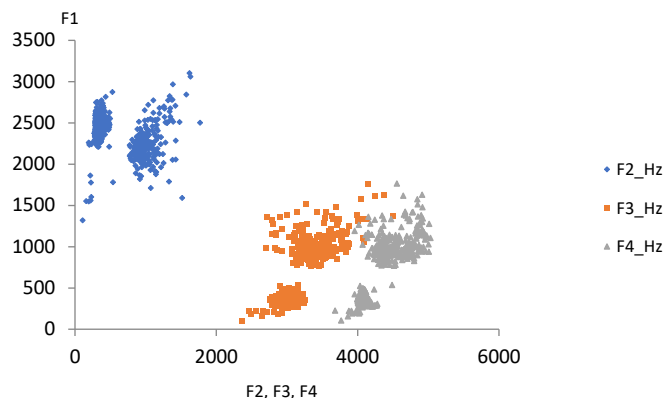


Figura 4. Exemplu de valori ale formantului F1 în raport cu F2, F1 cu F3 și F1 cu F4 pentru vocala /i/ (man\_i3\_M\_p3) (preluat din [38])

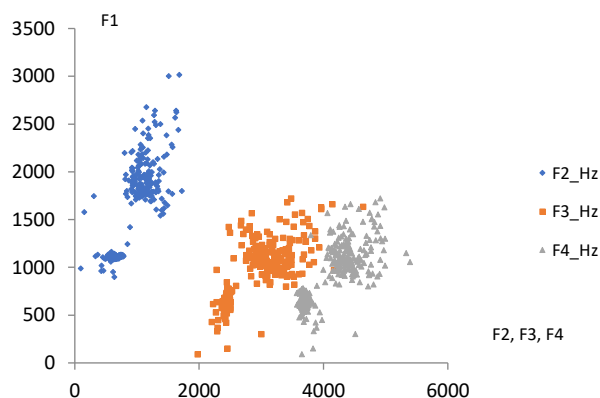


Figura 5. Exemplu de reprezentare a valorilor relative ale formantului F1 în raport cu F2, F1 cu F3 și F1 cu F4 pentru vocala /a/ (man\_a2\_p3) (preluat din [38])

TABEL VII. VALORILE DE CORELAȚIE ALE FORMANȚILOR LA SUBIECȚII DE GEN MASCULIN PENTRU ÎNTREAGA BAZĂ DE DATE (preluat din [38])

	F1,F2	F1,F3	F1,F4	F2,F3	F2,F4	F3,F4
A	0.595	0.602	0.505	0.776	0.629	0.673
E	0.473	0.713	0.640	0.639	0.397	0.753
I	0.088	0.449	0.678	0.416	0.150	0.662
O	0.770	0.697	0.669	0.756	0.741	0.716
U	0.581	0.552	0.502	0.836	0.693	0.752
Ă	0.765	0.646	0.674	0.797	0.774	0.792
Î	0.658	0.672	0.563	0.739	0.714	0.789

Se observă gradul de corelație între formanții vocalelor din limba română se aplică atât la nivel general, cât și la nivel particular, cu excepția vocalei /i/, pentru care nivelul de corelație observat a fost mai scăzut, în special datorită structurii fonetice a acesteia.

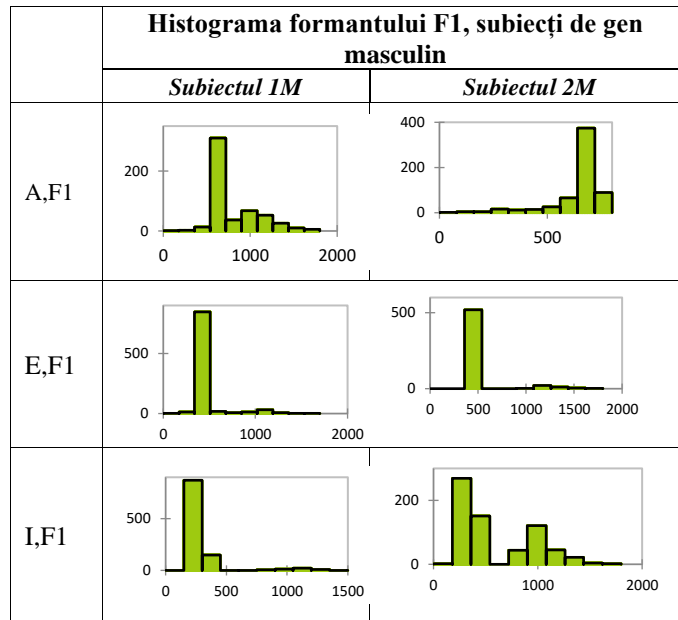


Figura 6. Exemple de histograme ale formanților F1 pentru subiecții de sex masculin preluat din [38]

Figura 6 ilustrează histogramele formanților F1 pentru vocalele /a/, /e/ și /i/ din înregistrările subiecților de sex masculin. În plus, analiza histogramei prezentate în figurile 6 și 7 relevă faptul că variabilitatea distribuției datelor nu este influențată de sexul subiecților.

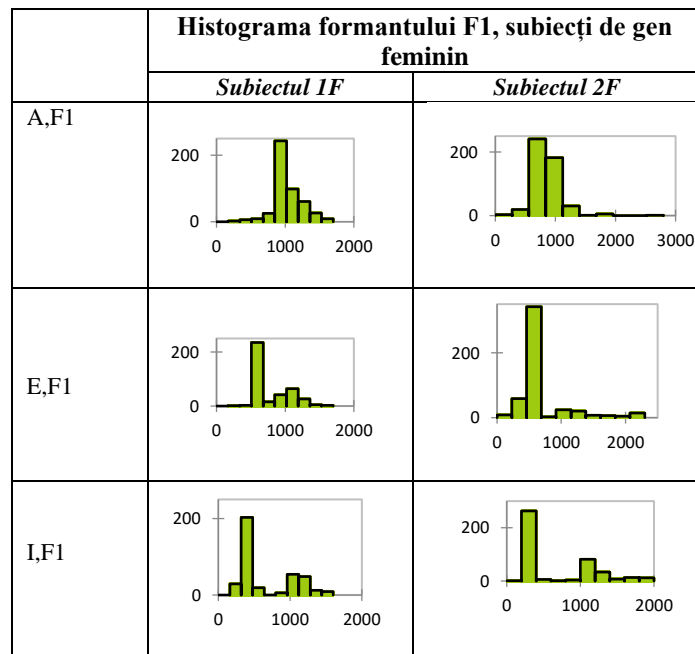


Figura 7. Exemple de histograme ale formanților F1 pentru subiecții de sex feminin (preluat din [38])

În figura 8 sunt prezentate toate histogramele care cuprind datele analizate pentru formanților înregistrate la subiecții de sex masculin.

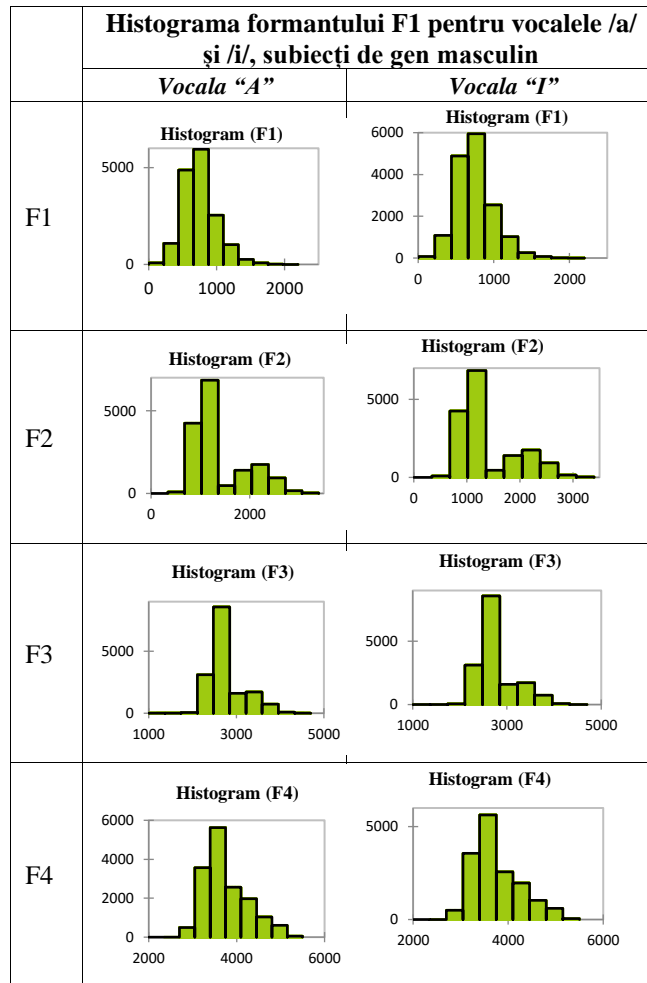


Figura 8. Histograme ale formanților F1, F2, F3 și F4 pentru subiecții de sex masculin, pentru toate vocalele /a/ și /i/ (preluat din [38])

Distribuțiile valorilor formanților sunt asimetrice și se compară bine cu distribuțiile raportate în [38] pentru limba română. În cadrul histogramei, o valoare modală este identificată ca un punct maxim relativ. Ținând cont de faptul că histogramele anterioare înregistrează două valori semnificativ mai mari în comparație cu întregul set de date, distribuția formanților F1, F2, F3 și F4 la subiecții de sex masculin pentru vocalele /a/ și /i/ este bi-modală.

#### 4.2.2. Vorbirea dinamică

Corelațiile formanților pentru vocalele din propoziția "vine mama" pentru toți vorbitorii sunt cele din tabelul VIII și figura 9, arătând că pentru același vorbitor, cele mai mari corelații apar pentru vocale susținute.

TABEL VIII. CORELAȚII PE SERIILE CONCATENATE ALE SEGMENTELOR CENTRALE (preluat din [40])

	/a1/ 'vine mama'	/a2/ 'vine mama'	/e/ 'vine mama'	/i/ 'vine mama'
Correl(F1,F2)	0.228	<b>0.704</b>	<b>0.698</b>	0.141
Correl(F1,F3)	0.546	<b>0.725</b>	0.607	0.301
Correl(F1,F4)	0.536	0.541	<b>0.750</b>	0.222
Correl(F2,F3)	0.419	<b>0.749</b>	<b>0.728</b>	<b>0.704</b>
Correl(F2,F4)	0.371	0.481	<b>0.813</b>	<b>0.674</b>
Correl(F3,F4)	<b>0.719</b>	0.527	<b>0.730</b>	0.666

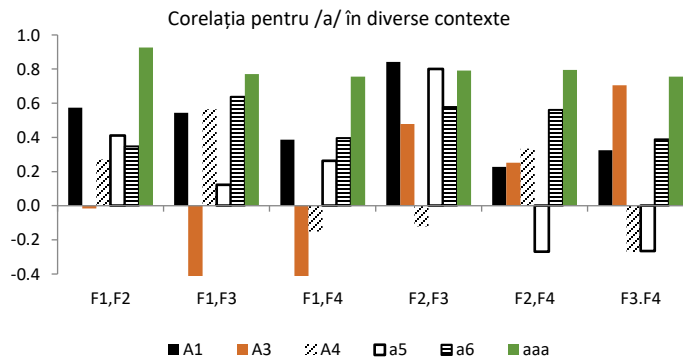


Figura 9. Vorbitorul #1(feminin), A1-A4 din "cine a făcut asta", A5, A6 din "vine mama", "aaa" denotă vocala susținută (preluat din [40])

Comparativ, pentru /e/ în "vine mama", corelațiile dintre formanții F2 și F4 și F3 și F4 sunt cele mai mari, iar toate valorile corelațiilor dintre formanți sunt mari pentru /e/. Variația corelațiilor pentru /e/ este prezentată în figura 10.

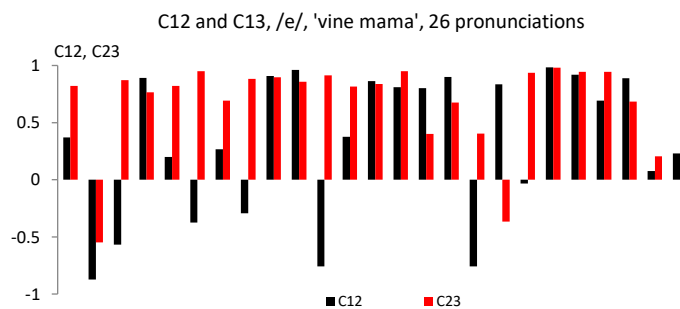


Figura 10. Variația corelațiilor pentru /e/ în același cuvânt și propunere. Fiecare două linii reprezintă un vorbitor (preluat din [40])

„S-a constatat că vocala accentuată /a/, adică /a1/, în "vine mama" are corelații mai mici decât /a/ în silabele neaccentuate (adică /a2/), cu excepția F3-F4, unde este mai mare, și F1-F4, care sunt aproape egale. Concluzionăm că, pentru seriile cu mai mulți vorbitori, corelațiile, chiar și pentru un singur subiect, depind mai mult de contextul vocalei (vocala susținută, vocala izolată, accentul, poziția cuvântului în propoziție și fonemele anterioare și posterioare) decât de vocală “ [40].



## Capitolul 5 – REZULTATE PRIVIND EFECTUL EMOȚIILOR ASUPRA NIVELULUI DE CORELAȚIE AL FORMANȚILOR

În tabelele de mai jos sunt prezentate valorile coeficienților de corelație pentru perechile de formanți ai vocalei /i/ în funcție de cele trei emoții analizate și de starea de neutralitate emoțională. Am constatat că, pentru vocala /i/, nivelul de corelație al perechilor de formanți F1-F2, F1-F3, F1-F4 rămâne scăzut, indiferent de tipul de emoție. În același timp, gradul de corelație rezultat este similar cu cel raportat în studiul pe care l-am realizat anterior [38]. Prin urmare, în ceea ce privește vocala /i/, tipul de emoție nu influențează gradul de corelație al perechilor de formanți.

TABEL I. VALORILE RELATIVE ALE PERECHILOR DE FORMANȚI PENTRU VOCALA /i/ (preluat din [70])

Vocala /i/				
	<i>Bucurie</i>	<i>Furie</i>	<i>Tristețe</i>	<i>Neutral</i>
Correl(F1,F2)	0.163	0.250	0.073	-0.168
Correl(F1,F3)	0.224	0.200	-0.088	0.298
Correl(F1,F4)	0.073	0.186	-0.013	0.245
Correl(F2,F3)	<b>0.804</b>	<b>0.596</b>	<b>0.788</b>	<b>0.627</b>
Correl(F2,F4)	0.639	0.348	0.330	0.476
Correl(F3,F4)	0.675	0.675	0.332	0.593

După cum se arată în tabelul II, în cazul vocalei /e/, perechile de formanți au valori de corelație mult mai mici pentru bucurie, furie și tristețe decât pentru starea de neutralitate emoțională. Gradul de corelație al perechilor de formanți F1-F2, F1-F3 și F1-F4 este semnificativ mai mic decât cel raportat pentru perechile F2-F3, F2-F4 și F3-F4.

TABEL II. VALORILE RELATIVE ALE PERECHILOR DE FORMANȚI PENTRU VOCALA /e/ (preluat din [70])

Vocala /e/				
	<i>Bucurie</i>	<i>Furie</i>	<i>Tristețe</i>	<i>Neutral</i>
Correl(F1,F2)	0.240	-0.183	0.158	<b>0.568</b>
Correl(F1,F3)	0.341	0.128	0.000	<b>0.748</b>
Correl(F1,F4)	0.323	0.220	0.315	<b>0.771</b>
Correl(F2,F3)	0.729	0.395	0.541	<b>0.706</b>
Correl(F2,F4)	0.601	0.261	0.701	0.733
Correl(F3,F4)	0.763	0.709	0.498	0.908

„Rezultatele unui studiu realizat pe aceeași bază de date și cu o temă de cercetare similară [40] au arătat că atunci când vocea vorbitorului este neutră, vocala accentuată din cuvântul /mama/, respectiv /a1/, are un nivel de corelație a formanților mai scăzut, în comparație cu vocala neaccentuată /a2/. Tabelele III și IV sintetizează rezultatele obținute în urma analizei corelaționale a formanților, analiză care a fost dezvoltată în cadrul cercetării din acest studiu. Putem observa că am obținut rezultate similare, în sensul că vocala neaccentuată prezintă un grad mai mare de

corelație a formanților, comparativ cu vocala accentuată, dar numai în cazul tristeții și al neutralității emoționale”[70].

**TABEL III. VALORILE RELATIVE ALE PERECHILOR DE FORMANȚI PENTRU VOCALA /a1/, ACCENTUATĂ (preluat din [70])**

<b>Vocala /a1/</b>				
	<i>Bucurie</i>	<i>Furie</i>	<i>Tristețe</i>	<i>Neutru</i>
Correl(F1,F2)	0.444	<b>0.514</b>	0.510	0.592
Correl(F1,F3)	0.439	<b>0.772</b>	0.709	0.404
Correl(F1,F4)	0.307	<b>0.522</b>	0.696	0.479
Correl(F2,F3)	0.546	<b>0.529</b>	0.489	0.325
Correl(F2,F4)	0.286	<b>0.630</b>	0.412	0.295
Correl(F3,F4)	0.675	0.567	0.697	0.752

**TABEL IV. VALORILE RELATIVE ALE PERECHILOR DE FORMANȚI PENTRU VOCALA /a2/, NEACCENTUATĂ (preluat din [70])**

<b>Vocala /a2/</b>				
	<i>Bucurie</i>	<i>Furie</i>	<i>Tristețe</i>	<i>Neutru</i>
Correl(F1,F2)	0.458	0.254	0.689	0.661
Correl(F1,F3)	0.667	0.476	0.736	0.679
Correl(F1,F4)	0.397	0.286	0.421	0.657
Correl(F2,F3)	0.454	0.336	0.782	0.820
Correl(F2,F4)	0.398	0.524	0.584	0.685
Correl(F3,F4)	0.293	0.587	0.584	0.806

În ceea ce privește emoția bucuriei, Tabelul V și figura 1 evidențiază faptul că nu există diferențe semnificative din punct de vedere statistic între gradul de corelație al perechii de formanți /a1/ și /a2/, cu excepția perechii F3-F4, pentru care coeficientul de corelație este mai mic pentru vocala neaccentuată /a2/.

**TABEL V. VALORILE RELATIVE ALE PERECHILOR DE FORMANȚI PENTRU EMOȚIA /bucurie/ (preluat din [70])**

<b>Bucurie</b>				
	<i>i</i>	<i>e</i>	<i>a1</i>	<i>a2</i>
Correl(F1,F2)	0.163	0.240	0.444	0.458
Correl(F1,F3)	0.224	0.341	0.439	0.667
Correl(F1,F4)	0.073	0.323	0.307	0.397
Correl(F2,F3)	0.804	0.729	0.546	0.454
Correl(F2,F4)	0.675	0.763	0.675	0.486
Correl(F3,F4)	0.675	0.763	0.675	0.293

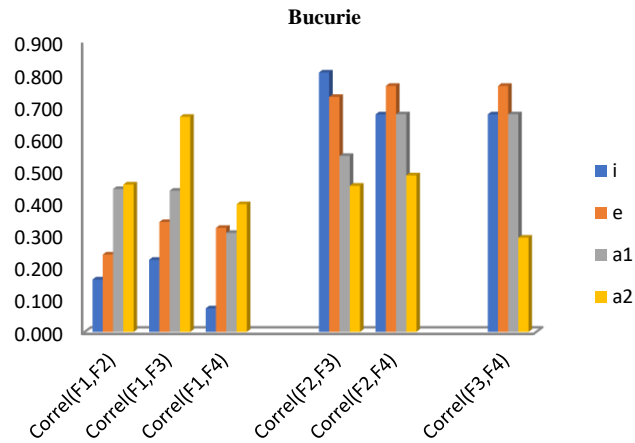


Figura 1. Reprezentarea grafică a formațiilor pentru bucurie (preluat din [70])

În ceea ce privește furia, ilustrată în tabelul VI și în figura 2, vocala accentuată /a1/ din cuvântul /mama/ are un nivel de corelație al perechilor de formații semnificativ mai mare decât cel al vocalei neaccentuate /a2/, acesta este un aspect surprinzător, deoarece atât în studiile anterioare [78], cât și în cazul celorlalte trei emoții, formații vocalelor neaccentuate au fost cei care au corelat mai puternic. Prin urmare, putem observa că specificitățile semnalului vocal pentru furie au condus la inversarea raportului de corelație între vocala accentuată și cea neaccentuată.

TABEL VI. VALORILE RELATIVE ALE PERECHILOR DE FORMAȚII PENTRU FURIE (preluat din [70])

Furie				
	<i>i</i>	<i>e</i>	<i>a1</i>	<i>a2</i>
Correl(F1,F2)	0.250	-0.183	0.514	0.254
Correl(F1,F3)	0.200	0.128	0.772	0.476
Correl(F1,F4)	0.186	0.220	0.522	0.286
Correl(F2,F3)	0.596	0.395	0.529	0.336
Correl(F2,F4)	0.675	0.709	0.567	0.587
Correl(F3,F4)	0.675	0.709	0.567	0.587

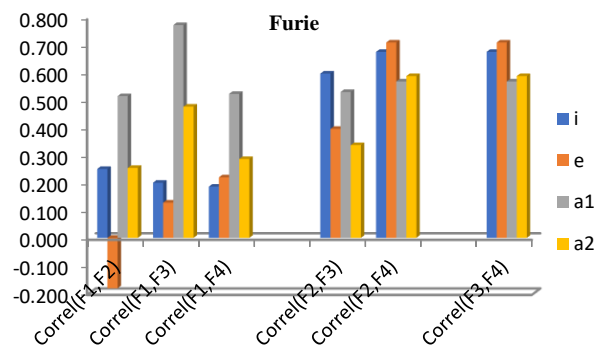


Figura 2. Reprezentarea grafică a formațiilor pentru furie (preluat din [70])

În ceea ce privește tristețea, menționată în tabelul VII și în figura 3, s-a observat că vocala /a/, fie ea accentuată sau neaccentuată, prezintă un nivel de corelație semnificativ mai mare al

perechilor de formați, comparativ cu nivelul de corelație al perechilor de formați ai vocalelor /i/ sau /e/. De exemplu, pentru perechea (F1, F3), coeficientul de corelație calculat este de 0,73 pentru vocala /a2/ și este evident mai mare decât 0,08, care este valoarea coeficientului de corelație a perechii de formați pentru vocala /i/.

TABEL VII. VALORILE RELATIVE ALE PERECHILOR DE FORMAȚI PENTRU TRISTEȚE (preluat din [70])

Tristețe				
	<i>i</i>	<i>e</i>	<i>a1</i>	<i>a2</i>
Correl(F1,F2)	0.073	0.158	0.510	0.689
Correl(F1,F3)	-0.088	0.000	0.709	0.736
Correl(F1,F4)	-0.013	0.315	0.696	0.421
Correl(F2,F3)	0.788	0.541	0.489	0.782
Correl(F2,F4)	0.332	0.498	0.697	0.584
Correl(F3,F4)	0.332	0.498	0.697	0.584

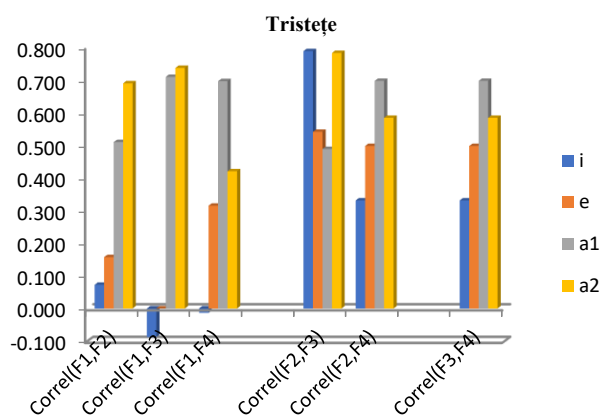


Figura 3. Reprezentarea grafică a formațiilor pentru tristețe (preluat din [70])

Rezultatele pe care le-am obținut și care sunt sintetizate în tabelul VIII și ilustrate în figura 4 arată că, în ceea ce privește starea de neutralitate emoțională, nivelul de corelație al perechilor de formați corespunzători celor patru vocale este ridicat (în medie există valori ale coeficientului de corelație mai mari de 0,5). De asemenea, am descoperit că, spre deosebire de nivelul de corelație al formațiilor raportat pentru vocalele asociate emoțiilor, în cazul tonului neutru, corelația formațiilor este mai omogenă, fără oscilații extreme de la o vocală la alta.

TABEL VIII. VALORILE RELATIVE ALE PERECHILOR DE FORMAȚI PENTRU NEUTRALITATE (preluat din [70])

Neutru				
	<i>i</i>	<i>e</i>	<i>a1</i>	<i>a2</i>
Correl(F1,F2)	-0.168	0.568	0.592	0.661
Correl(F1,F3)	0.298	0.748	0.404	0.679
Correl(F1,F4)	0.245	0.771	0.479	0.657

Neutru				
	<i>i</i>	<i>e</i>	<i>a1</i>	<i>a2</i>
Correl(F2,F3)	0.627	0.706	0.325	0.820
Correl(F2,F4)	0.593	0.908	0.752	0.806
Correl(F3,F4)	0.593	0.908	0.752	0.806

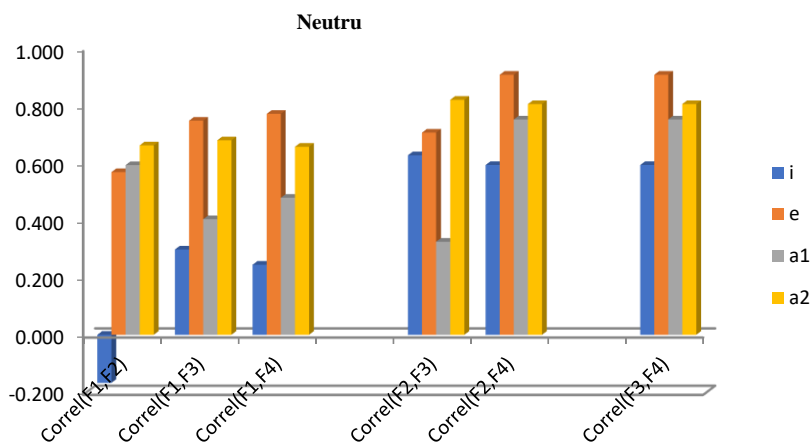


Figura 4. Reprezentarea grafică a formațiilor pentru neutralitate (preluat din [70])

Emoțiile influențează intensitatea, tonalitatea, accentul și ritmul vocii, determinând astfel fluctuații calitative în caracteristicile semnalului vocal. Prin urmare, este de așteptat ca valorile intercorelațiilor dintre formații vocalice să varieze în funcție de specificul emoțiilor vorbitorului (intensitate, polaritate, excitație, durată) și invers, în cazul tonusului neutru, valorile vor fi mai omogene.

Tabelul IX prezintă valorile coeficienților de corelație pentru perechile de formații ai celor patru vocale asociate cu bucuria. Observăm că gradul de corelație dintre F0 și formații prezintă valori semnificativ mai mici în comparație cu cele calculate pentru perechile de formații (F1-F2, ..., F3-F4). În același timp, am constatat că perechile de formații de ordin înalt prezintă un nivel de corelație mai mare decât cel raportat pentru formații de ordin inferior.

TABEL IX. VALORILE COEFICIENȚILOR DE CORELAȚIE PENTRU EMOȚIA BUCURIE (preluat din [80])

Corelații inter-vocalice	Bucurie			
	<i>I</i>	<i>E</i>	<i>AI</i>	<i>A2</i>
<b>F0,F1</b>	0.10	0.14	-0.03	0.23
<b>F0,F2</b>	-0.41	-0.24	0.23	0.23
<b>F0,F3</b>	-0.36	-0.29	-0.38	0.05
<b>F0,F4</b>	-0.57	-0.40	-0.14	0.14
<b>F1,F2</b>	0.16	0.06	0.34	0.41
<b>F1,F3</b>	0.14	0.36	0.42	0.63
<b>F1,F4</b>	-0.04	0.36	0.35	0.14
<b>F2,F3</b>	0.70	0.71	0.52	0.66
<b>F2,F4</b>	0.57	0.53	0.28	0.29
<b>F3,F4</b>	0.75	0.68	0.62	0.29

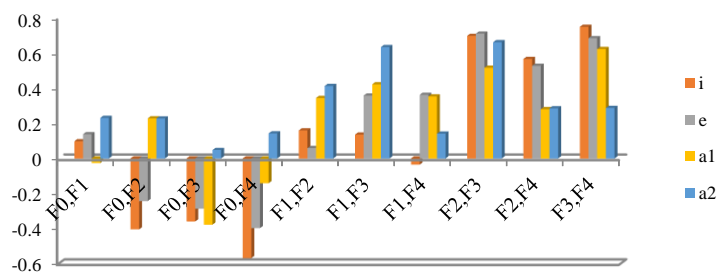


Figura 5. Reprezentarea grafică a formațiilor pentru bucurie (preluat din [80])

În ceea ce privește tristețea, rezultatele prezentate în tabelul X susțin constatarea că nivelul de corelație dintre frecvența fundamentală și formații este mai mic decât cel al perechilor de formații. De asemenea, pentru vocala neaccentuată /a2/ observăm că valorile coeficientului de corelație sunt mai mari în comparație cu cele ale vocalei accentuate /a1/. Aceste rezultate sunt similare cu cele raportate în [70], [38], deși studiile respective au fost realizate prin utilizarea valorilor reale ale formațiilor, iar frecvența fundamentală nu a fost inclusă în analiza corelațională.

TABEL X. VALORILE COEFICIENȚILOR DE CORELAȚIE PENTRU EMOȚIA TRISTEȚE (preluat din [80])

Corelații inter-vocalice	Tristețe			
	I	E	AI	A2
$\overline{F0,F1}$	0.33	-0.17	0.43	0.17
$\overline{F0,F2}$	0.33	0.28	0.01	0.32
$\overline{F0,F3}$	0.44	0.25	0.22	0.26
$\overline{F0,F4}$	0.34	0.17	0.41	0.34
$\overline{F1,F2}$	0.20	-0.21	0.30	0.53
$\overline{F1,F3}$	0.17	-0.21	0.62	0.43
$\overline{F1,F4}$	0.11	0.13	0.78	0.36
$\overline{F2,F3}$	0.79	0.67	0.39	0.80
$\overline{F2,F4}$	0.57	0.84	0.51	0.66
$\overline{F3,F4}$	0.53	0.61	0.80	0.85

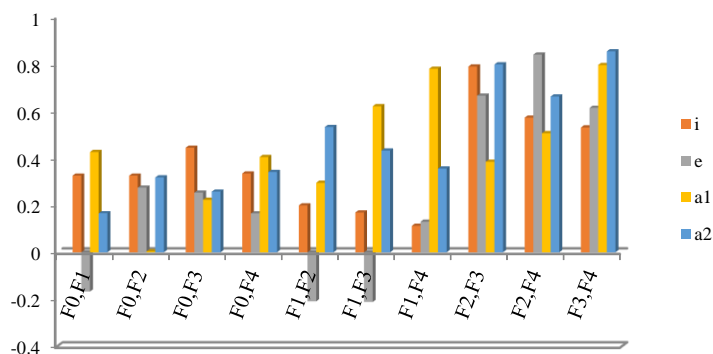


Figura 6. Reprezentarea grafică a formațiilor pentru emoția tristețe (preluat din [80])

TABEL XI. VALORILE COEFICIENȚILOR DE CORELAȚIE PENTRU EMOȚIA FURIE (preluat din [80])

Corelații inter-vocalice	Furie			
	<i>I</i>	<i>E</i>	<i>AI</i>	<i>A2</i>
$\overline{F0,F1}$	0.15	0.77	-0.12	-0.56
$\overline{F0,F2}$	-0.25	-0.66	0.51	0.47
$\overline{F0,F3}$	-0.15	-0.09	-0.10	0.09
$\overline{F0,F4}$	-0.08	0.17	0.66	0.76
$\overline{F1,F2}$	0.32	-0.58	0.63	0.05
$\overline{F1,F3}$	0.01	-0.07	0.78	0.58
$\overline{F1,F4}$	-0.22	0.16	0.54	0.03
$\overline{F2,F3}$	0.41	0.30	0.55	0.30
$\overline{F2,F4}$	0.10	0.25	0.70	0.71
$\overline{F3,F4}$	0.76	0.73	0.54	0.55

În tabelul XI, observăm că valorile de corelație ale formanților pentru vocala /i/, în cazul furiei, sunt mult mai mici în comparație cu celelalte vocale. Rezultate similare au fost raportate în [71], [72], una dintre explicațiile sugerate fiind aceea că mecanismul specific de articulare a sunetului și de pronunție a vocalei /i/ ar determina un grad mai mic de corelație a formanților.

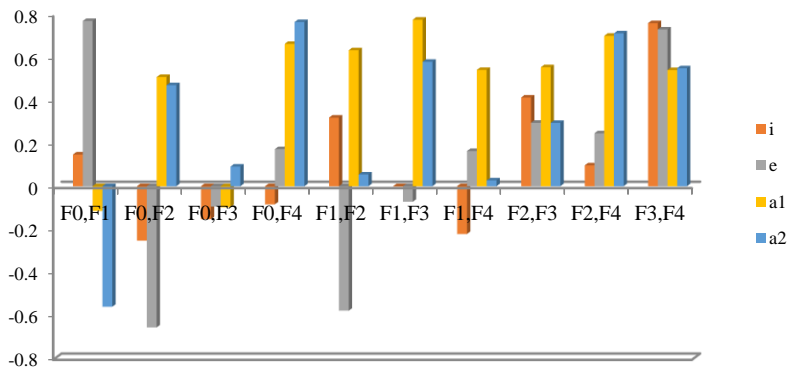


Figura 7. Reprezentarea grafică a formanților pentru emoția furie (preluat din [80])

“Pentru starea de neutralitate emoțională, constatăm că nivelul de corelație dintre frecvența fundamentală și formanți este mai mare decât rezultatele obținute pentru bucurie, furie și tristețe (corelația este moderată, cu valori medii peste 0,3). Am constatat, de asemenea, că formanții de ordin înalt sunt mai puternic asociați în comparație cu cei de ordin scăzut. Spre deosebire de rezultatele obținute pentru emoții, valorile corelației pentru neutralitatea emoțională sunt mai omogene, în sensul că nu există atât de multe oscilații extreme de la o vocală la alta. În plus, ca indicator statistic, media reprezintă tendința centrală a unei serii de valori și arată zona în care datele tind să se concentreze. Faptul că am efectuat analiza statistică folosind valorile medii ale formanților este un argument care ar putea justifica variația scăzută a valorilor coeficienților de corelație calculați pentru starea de neutralitate emoțională. Pentru a susține această observație, includem figura 8.” [80]

TABEL XII. VALORILE COEFICIENȚILOR DE CORELAȚIE PENTRU NEUTRALITATEA EMOȚIONALĂ (preluat din [80])

Corelații inter-vocalice	Neutru			
	<i>I</i>	<i>E</i>	<i>A1</i>	<i>A2</i>
$\overline{F0, F1}$	0.29	0.35	0.24	0.41
$\overline{F0, F2}$	0.51	0.47	0.58	0.34
$\overline{F0, F3}$	0.53	0.69	-0.04	0.31
$\overline{F0, F4}$	0.55	0.66	-0.20	0.17
$\overline{F1, F2}$	0.04	0.57	0.39	0.65
$\overline{F1, F3}$	0.29	0.65	0.47	0.63
$\overline{F1, F4}$	0.31	0.71	0.26	0.67
$\overline{F2, F3}$	0.79	0.59	0.04	0.84
$\overline{F2, F4}$	0.62	0.73	-0.32	0.75
$\overline{F3, F4}$	0.71	0.90	0.63	0.88

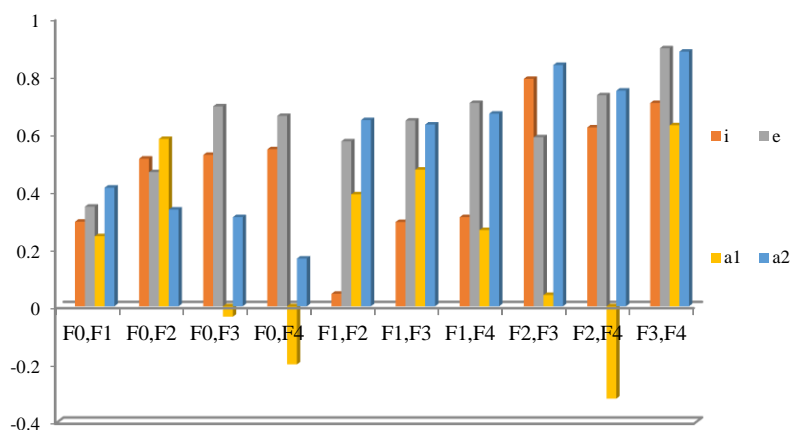


Figura 8. Reprezentarea grafică a formanților pentru neutralitate emoțională (preluat din [80])

### 5.1. Calcularea coeficienților de regresie liniară

Pentru vocala accentuată /a1/ și vocala neaccentuată /a2/, din enunțul /vine mama/, am extras formanții și am aplicat regresia liniară simplă, considerând F0 ca variabilă independentă și F1 și F2 ca variabile dependente. Cu alte cuvinte, am verificat în ce măsură frecvența fundamentală este un predictor eficient pentru estimarea schimbărilor în seriile de date asociate criteriului (F1 sau F2). Rezultatele obținute sunt rezumate în tabelele XIII și XIV.

TABEL XIII. PARAMETRII DE REGRESIE LINIARĂ PENTRU VOCALA /A1/ ÎN ENUNȚUL /VINE MAMA/ (preluat din [80])

	Vocala accentuată /a1/	
	$F1(F0)$	$F2(F0)$
Bucurie	$R^2=0,0005$ $Y = 818.864 + 0.079F0$	$R^2=0.0559$ $Y = 1286.87 + 0.5219F0$
Tristețe	$R^2= 0.1829$ $Y = 552.18 + 1.480F0$	$R^2= 2.256E-11$ $Y=1302.05 + (1.84E-05)F0$
Furie	$R^2= 0.0201$ $Y = 950.16 + 0.513F0$	$R^2=0.2305$ $Y = 1101.77 + 0.871F0$
Neutru	$R^2= 0.0637$ $Y = 599.19 + 0.730F0$	$R^2= 0.2191$ $Y = 883.90 + 2.690F0$

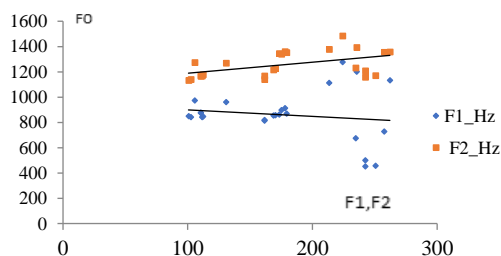


**TABEL XIV. PARAMETRII DE REGRESIE LINIARĂ PENTRU VOCALA /A2/ ÎN ENUNȚUL /VINE MAMA/ (preluat din [80])**

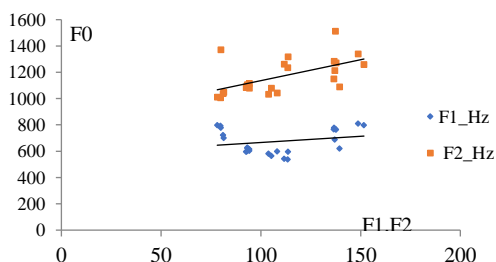
	Vocala neaccentuată /a2/	
	F1(F0)	F2(F0)
Bucurie	$R^2=0.0572$ $Y = 773.29 + 0.5612F0$	$R^2= 0.0590$ $Y = 1457.86 + 0.5285F0$
Tristețe	$R^2= 0.0359$ $Y = 785.33 + 0.270F0$	$R^2=0.1025$ $Y = 1534.81 + 0.910F0$
Furie	$R^2= 0.2763$ $Y = 113.56 + 2.328F0$	$R^2= 0.2263$ $Y = 1079.68 + 1.375F0$
Neutru	$R^2= 0.1696$ $Y = 732.80 + 0.540F0$	$R^2=0.1127$ $Y = 1574.79 + 0.940F0$

Observăm că valorile coeficientului de determinare ( $R^2$ ) – care se calculează ca pătrat al coeficientului de corelație multiplă – sunt scăzute atât pentru vocala /a1/, cât și pentru /a2/. De exemplu, pentru vocala /a1/, asociată cu bucuria, frecvența F0 explică 5% din variația variabilei dependente – F2 ( $R^2_{F_2(F_0)} = 0.0559$ ). Pentru vocala /a2/, asociată cu tristețea, frecvența F0 explică 3% din variația variabilei dependente – F1 ( $R^2_{F_1(F_0)} = 0.0359$ ). În privința vocalei /a1/, valori moderate ale coeficientului de determinare au fost obținute pentru furie (unde variația F0 prezice 23% din variația variabilei criteriu – F2) și neutralitate (unde variația F0 prezice 21% din variația F2). Pentru vocala /a2/, valori moderate ale  $R^2$  au fost obținute pentru furie (unde frecvența F0 explică 27% din variația lui F1) și pentru tonul neutru (16%). În același timp, frecvența F0 explică 22% din variația F2, în cazul furiei și 11% în cazul tonului neutru

Graficele de mai jos prezintă liniile de regresie pentru frecvența fundamentală (F0) și formanții F1 și F2 ai vocalelor /a1/ și /a2/, pentru furie și neutralitate emoțională.



**Figura 9. Parametrii de regresie liniară ai vocalei /a1/ pentru emoția furie (preluat din [80])**



**Figura 10. Parametrii de regresie liniară ai vocalei /a1/ pentru starea de neutralitate (preluat din [80])**

În tabelele de mai jos sunt prezentate valorile coeficienților calculați în analiza de varianță (ANOVA). “Testul F verifică dacă linia de regresie este semnificativ diferită de valoarea zero, adică testează dacă estimarea criteriului în funcție de predictor este mai bună decât predicția bazată pe șansă sau accident. Testul F este utilizat pentru a verifica ipoteza nulă conform căreia  $b$ , în ecuația  $Y = a + bX$ , este zero (situație în care linia de regresie ar fi perfect orizontală). Astfel, în cazul acestui studiu, dacă  $F_0$  este un bun predictor pentru formanții  $F_1$  și  $F_2$ ” [80].

Rezultatele obținute sunt prezentate în următoarele tabele.

**TABEL XV. PARAMETRII ANALIZEI DE VARIANȚĂ ANOVA PENTRU VOCALA /A1/ ÎN ENUNȚUL /VINE MAMA/ (preluat din [80])**

<b>A1</b>	<b>F1(F0)</b>	<b>F2(F0)</b>
Bucurie	F(1.31) = 0.0173 p=0.8960 >0.05	F(1.31) = 1.8377 p=0.1850 >0.05
Tristețe*	F(1.27) = 6.0438 p=0.020 <0.05	F(1.27) = 6.09E-10 p= 0.999 >0.05
Furie**	F(1.23) = 0.4741 p=0.0497 <0.05	F(1.23) = 6.8916 p= 0.015 <0.05
Neutru*	F(1.25) = 0.9539 p=0.338 >0.05	F(1.25) = 7.016 p= 0.013 <0.05

**TABEL XVI. PARAMETRII ANALIZEI DE VARIANȚĂ ANOVA PENTRU VOCALA /A2/ ÎN ENUNȚUL /VINE MAMA/ (preluat din [80])**

<b>A2</b>	<b>F1(F0)</b>	<b>F2(F0)</b>
Bucurie	F(1.28) = 1.6987 p=0.2030 >0.05	F(1.28) = 1.7581 p=0.1955 >0.05
Tristețe	F(1.25) = 0.9335 p=0.343 >0.05	F(1.25) = 2.8560 p= 0.103 >0.05
Furie**	F(1.23) = 8.7853 p=0.006 <0.05	F(1.23) = 6.7281 p=0.016 <0.05
Neutru	F(1.18) = 3.6771 p=0.071 >0.05	F(1.18) = 2.2865 p= 0.147 >0.05

Constatăm că, în ceea ce privește vocala accentuată /a1/, frecvența fundamentală este un bun predictor pentru  $F_1$  doar atunci când este asociată cu tristețea și furia, iar pentru  $F_2$  este un predictor semnificativ doar în asociere cu furia și tonul neutru. Pe de altă parte, în ceea ce privește vocala neaccentuată /a2/, frecvența fundamentală este un bun predictor doar în cazul furiei, pentru a explica atât varianța lui  $F_1$ , cât și pe cea a lui  $F_2$ .

În continuare, am testat în ce măsură formanții de ordin superior sunt predictorii semnificativi pentru cei de ordin inferior și am obținut următoarele rezultate:

TABEL XVII. PARAMETRI DE REGRESIE LINIARĂ PENTRU VOCALA ACCENTUATĂ /a2/ - /Aseară/

Vocala accentuată /a2/				
VD	Bărbați		Femei	
	Tristețe	Neutru	Tristețe	Neutru
<b>F1</b>	Y=619.06+0.05F2+0.10F3-0.06F4 R <sup>2</sup> =0.11 F(3,546)=23.96 p<0.001	Y=287.20+0.36F2+0.10F3-0.10F4 R <sup>2</sup> =0.31 F(3,449)=69.70 p<0.001	Y=277.24+0.09F2+0.13F3-0.00061F4 R <sup>2</sup> =0.15 F(3,1451)=91.06 p<0.001	Y=286.09+0.15F2+0.02F3+0.001F4 R <sup>2</sup> =0.16 F(3,557)=36.79 p<0.001
<b>F2</b>	Y=382.25+0.37F3+0.04F4 R <sup>2</sup> =0.44 F(2,547)=221.56 p<0.001	Y=551.30+0.35F3+0.001F4 R <sup>2</sup> =0.24 F(2,450)=74.14 p<0.001	Y=916.04+0.20F3+0.001F4 R <sup>2</sup> =0.18 F(2,1452)=165.07 p<0.001	Y=463.15+0.26F3+0.09F4 R <sup>2</sup> =0.30 F(2,558)=121.80 p<0.001
<b>F3</b>	Y=630.90+0.46F4 R <sup>2</sup> =0.37 F(1,548)=324.79 p<0.001	Y=973.22+0.40F4 R <sup>2</sup> =0.30 F(1,451)=194.32 p<0.001	Y=-543.06+0.79F4 R <sup>2</sup> =0.44 F(1,1453)=1159.24 p<0.001	Y=908.43+0.43F4 R <sup>2</sup> =0.41 F(1,559)=291.37 p<0.001

“Atunci când vorbitorii de gen feminin exprimă atât tristețe, cât și ton neutru, am observat o tendință de creștere a puterii de predicție pentru modelul de regresie, datorită excluderii formanțelor anteriori (F1, F2, F3). În cazul vorbitorilor de gen masculin care exprimă tristețe, cea mai mare valoare pentru coeficientul de determinare (R<sup>2</sup>) este cea corespunzătoare lui F2" [96].

TABEL XVIII. PARAMETRI DE REGRESIE LINIARĂ PENTRU VOCALA ACCENTUATĂ /a2/ - /Cine a făcut asta/

Vocala accentuată /a2/				
VD	Bărbați		Femei	
	Tristețe	Neutru	Tristețe	Neutru
<b>F1</b>	Y=-130.58+0.03F2+0.13F3+0.08F4 R <sup>2</sup> =0.35 F(3,564)=102.26 p<0.001	Y=-324.61+0.32F2-0.005F3+0.14F4 R <sup>2</sup> =0.54 F(3,610)=245.81 p<0.001	Y=66.82+0.39F2+0.05F3-0.03F4 R <sup>2</sup> =0.28 F(3,1084)=146.40 p<0.001	Y=434.50+0.025F2+0.09F3+0.005F4 R <sup>2</sup> =0.12 F(3,758)=34.69 p<0.001
<b>F2</b>	Y=-528.44+0.44F3+0.18F4 R <sup>2</sup> =0.49 F(2,565)=280.28 p<0.001	Y=179.03+0.42F3+0.03F4 R <sup>2</sup> =0.42 F(2,611)=225.86 p<0.001	Y=-117.66+0.43F3+0.09F4 R <sup>2</sup> =0.60 F(2,1082)=825.61 p<0.001	Y=606.722+0.21F3+0.05F4 R <sup>2</sup> =0.26 F(2,759)=135.79 p<0.001
<b>F3</b>	Y=258.17+0.61F4 R <sup>2</sup> =0.38 F(1,566)=348.97 p<0.001	Y=-681.67+0.85F4 R <sup>2</sup> =0.50 F(1,612)=623.17 p<0.001	Y=-812.90+0.84F4 R <sup>2</sup> =0.46 F(1,1083)=947.46 p<0.001	Y=-219.91+0.71F4 R <sup>2</sup> =0.34 F(1,760)=399.91 p<0.001

„În cazul în care vorbitorii de gen masculin folosesc tonul neutru, există valori semnificativ mai mari pentru coeficientul de determinare (R<sup>2</sup>) în comparație cu valorile obținute pentru aceeași propoziție, aceeași vocală, ton neutru, dar vorbitori de gen feminin. De asemenea, pentru

vorbitorii de gen masculin, ton neutru, predictorii F2, F3 și F4 au explicat 54% din varianța variabilei dependente, care, în acest caz, a fost F1.” [96]

Pe de altă parte, pentru vorbitorii de gen feminin, ton neutru, aceiași predictorii au explicat doar 12% din varianța variabilei criteriu, F2. Mai mult, atunci când vorbitorii de gen feminin au exprimat tristețe, formanții de ordin superior (F3-F4) au prezis 60% din varianța variabilei criteriu, F2.

TABEL XIX. PARAMETRI DE REGRESIE LINIARĂ PENTRU VOCALA ACCENTUATĂ /a1/ - /Vine mama!

Vocala accentuată /a1/				
VD	Bărbați		Femei	
	Tristețe	Neutru	Tristețe	Neutru
<b>F1</b>	Y=451.78+0.35F2-0.004F3-0.04F4 R <sup>2</sup> =0.47 F(3,544)=160.81 p<0.001	Y=296.17+0.14F2+0.15F3-0.05F4 R <sup>2</sup> =0.45 F(3,585)=161.32 p<0.001	Y=-245.01+0.18F2+0.14F3+0.10F4 R <sup>2</sup> =0.37 F(3,742)=147.32 p<0.001	Y=-59.12+0.25F2+0.06F3+0.08F4 R <sup>2</sup> =0.35 F(3,601)=112.45 p<0.001
<b>F2</b>	Y=615.62+0.22F3+0.001F4 R <sup>2</sup> =0.16 F(2,545)=54.48 p<0.001	Y=-523.02+0.66F3+0.03F4 R <sup>2</sup> =0.44 F(2,586)=238.62 p<0.001	Y=497.42+0.28F3+0.001F4 R <sup>2</sup> =0.29 F(2,743)=157.33 p<0.001	Y=488.92+0.14F3+0.11F4 R <sup>2</sup> =0.31 F(2,602)=135.25 p<0.001
<b>F3</b>	Y=588.34+0.51F4 R <sup>2</sup> =0.14 F(1,546)=92.96 p<0.001	Y=-328.48+0.76F4 R <sup>2</sup> =0.55 F(1,587)=717.96 p<0.001	Y=-335.27+0.78F4 R <sup>2</sup> =0.45 F(1,744)=627.33 p<0.001	Y=-247.21+0.75F4 R <sup>2</sup> =0.48 F(1,603)=573.12 p<0.001

„În mod similar cu cazul anterior (a doua vocală /a/ din propoziția /cine a făcut asta/), pentru tonul neutru, s-a observat că valorile R<sup>2</sup> tind să fie mai mari pentru vorbitorii de gen masculin în comparație cu cele calculate pentru vorbitorii de gen feminin. De asemenea, la bărbați, F4 a crescut semnificativ puterea de predicție a varianței variabilei dependente F3, cu excepția cazului tristeții” [96].

În mod surprinzător, posibilitatea existenței unor corelații între formanți în timpul vorbirii nu a atras atenția până de curând. Deși au fost identificate câteva studii legate de acest subiect, nu a fost găsit niciunul care să abordeze noțiunea de corelații între formanți. Spre exemplu, un scurt rezumat datând din 1986 (Wood) [98] folosește cuvintele “corelația formanților în vorbire”, însă această exprimare este improprie, deoarece se face referire la “Line spectral pairs [to] formant correlation”, și nu la corelații între formanți. Diverse măsurători bazate pe formanți, cum ar fi aria spațiului vocalic, raportul de centralizare al formanților, indicele de articulare a vocalelor (Caverlé & Vogel, 2020)[99], (Fletcher et al., 2017) [100], (Sapir et al., 2010) [101] au la bază medii ale valorilor formanților și sunt adesea derivate din pronunții susținute, fără nicio legătură cu noțiunea

de corelație (Pearson). Despre acest aspect, Rowe et al. (2020) [102] spunea “*corelația formanților poate corespunde unui cuplaj mai mic al mișcărilor limbii*”.

O mențiune neechivocă a corelației funcționale dintre elementele anatomice ale sistemului fonator în timpul vorbirii datează cel puțin din 1981: “*Aceste elemente structurale ale laringelui sunt puternic corelate din punct de vedere funcțional*” (Teodorescu, p. 107) [103]. Noțiunea de corelație între formanți datează din anii 1980, a fost menționată neambiguu în (Teodorescu, p. 42, de asemenea figura. 2.24) [104]. Conceptul a fost susținut de rezultatele raportate în câteva studii anterioare din [40], [105]. Noțiunea de corelație între formanți bazată pe considerente fiziologice a fost analizată preliminar în [100],[106-108], dar din căutările efectuate în literatura de specialitate, acesta este primul studiu care aduce dovezi statistice privind corelațiile între formanții vocalelor limbii române.

Rezultatele unei astfel de cercetări au aplicabilitate în domeniile foneticii, fonologiei și tehnologiei vorbirii. În domeniul foneticii și al fonologiei, aceste valori pot fi utilizate în dezvoltarea unor instrumente pentru analiza vorbirii. În medicină, metodele bazate pe analiza corelativă pot ajuta la diagnosticarea unor tulburări de limbaj din sfera patologiei vorbirii. În tehnologia vorbirii, analiza corelativă poate contribui la îmbunătățirea calității acustice a comunicării vocale cu lățime de bandă limitată, prin reducerea datelor transmise privind formanții și înlocuirea parțială a acestora cu date privind corelațiile formanților.

Având în vedere că focusul cercetării a fost mutat de la valorile formanților spre valorile coeficienților de corelație între perechi de formanți, putem spune că populația analizată este o populație de coeficienți de corelație (PCC) și de aceea trebuie acordată o atenție deosebită transformatei Fisher. Din cauza incertitudinii cu privire la proprietățile populației, adoptăm alternativ supozițiile că valorile populației fie sunt normale distribuite (cel puțin după transformata Fisher), fie non-normal distribuite; în cazul încălcării asumției de normalitate a distribuției, ipoteza de nul  $H_0$  este pur și simplu că perechile de populații sunt identice și se aplică teste statistice neparametrice, cum ar fi spre exemplu, testul U Mann-Whitney și testul medianei lui Mood. Vom fi foarte conservatori și vom fi mulțumiți de rezultatul confirmării  $H_1$  doar atunci când  $H_1$  este susținută atât de testele parametrice, cât și de cele neparametrice.

La prima întrebare de cercetate, (i. Corelează semnificativ statistic formanții?), se răspunde prin verificarea ipotezei  $H_0$ : Pentru orice alegere de propoziție (S), vocală (V) și emoție (Em), populația valorilor de corelație notată  $Ch_j(Em, V, S)$  are media 0 (nicio corelație).

Ipoteza  $H_1$  va fi că corelațiile nu sunt egale cu zero, este necesar un test statistic two-tailed. Dacă  $H_0$  nu este verificată pentru un set substanțial de combinații (h, j, Em, vocală, S), suntem îndreptățiți să concluzionăm că cel puțin unii formanți corelează în anumite circumstanțe (Em, vocală, S).

O problemă cu această metodă este decizia privind alegerea testului statistic adecvat. În cazul nostru, nu putem presupune o populație specifică pentru a o utiliza în ipoteză – populațiile de corelație comparate sunt în esență necunoscute. Astfel, un test one-tail este inadecvat și avem nevoie de un test two-tail. Abaterile standard ale populațiilor sunt necunoscute, astfel că un test Z nu ar fi o alegere bună. Așadar, aplicăm un test t deoarece comparăm doar mediile și ne permite asumții mai puțin rigide cu privire la distribuția valorilor. Testul t al lui Welch se prezintă ca o alegere bună.

În mod specific pentru astfel de cazuri, Fisher a propus transformata care îi poartă numele; transformata Fisher transformă  $[-1,1]$  în  $(-\infty, \infty)$  și reduce asimetria, producând, în mod ideal, o distribuție a valorilor apropiată de o distribuție gaussiană. Cu toate acestea, testăm ipoteza de nul conform căreia valorile de corelație sunt, în medie, nule. Aceasta induce ideea că valorile cele mai frecvente ale corelațiilor sunt apropiate de zero sau, cel puțin, că valorile sunt distribuite simetric în jurul valorii zero. În primul caz, transformata Fisher este inutilă, deoarece pentru valori situate aproximativ în intervalul  $(-0,5, 0,5)$  transformata Fisher nu are practic niciun efect. În cel de-al doilea caz, cu valori de corelație distribuite asimetric, media nu va fi zero, astfel încât transformata Fisher nu este cu adevărat necesară. Doar atunci când multe valori sunt apropiate de -1 sau de 1 sau când distribuțiile sunt foarte înclinate, este recomandabil să se aplice mai întâi transformata Fisher, pentru a obține populații mai apropiate de normalitate și pentru a putea obține valori p semnificative.

Utilizarea controversată a transformatei lui Fisher și utilitatea discutabilă a acesteia necesită prudență în derivarea concluziilor. Exercițarea prudenței poate însemna să se verifice dacă concluziile sunt la fel de valabile cu o probabilitate ridicată înainte și după aplicarea transformatei lui Fisher.

Tabelul XX exemplifică rezultatele obținute pentru corelații, pentru mai mulți vorbitori și mai multe pronunții per vorbitor.

**TABEL XX. VALORILE DE CORELAȚIE A FORMANȚILOR / Ə /, “ASEARĂ”, TON NEUTRU GEN FEMININ FĂRĂ FISHER TRANSFORM (preluată din [96])**

	<b>C12</b>	<b>C13</b>	<b>C14</b>	<b>C23</b>	<b>C24</b>	<b>C34</b>
<b>0745_aa_1-</b>	-0.163	-0.022	0.257	-0.130	-0.362	0.570
0745_aa_2-	0.573	0.347	0.439	0.560	0.855	0.660
0745_aa_3-	0.345	0.470	0.372	0.712	0.485	0.401
2703_aa_1-	0.573	0.846	0.725	0.512	0.561	0.502
2703_aa_2-	0.640	0.568	0.098	0.588	-0.028	0.440
2703_aa_3-	0.370	0.385	0.242	0.432	0.338	0.766
2718_aa_1-	0.542	0.478	0.080	0.635	0.591	0.342
2718_aa_2-	0.317	0.688	0.491	0.506	0.193	0.497
2718_aa_3-	0.878	0.802	0.518	0.787	0.494	0.792

29125_	0.691	0.911	0.825	0.865	0.796	0.908
29125_	0.231	-0.310	0.290	0.452	-0.041	0.132
29125_	0.665	0.710	0.450	0.434	0.008	0.510
697123	0.884	0.777	0.732	0.869	0.807	0.704
697123	0.441	0.291	0.458	0.814	0.734	0.596
697123	0.201	-0.263	0.126	0.583	0.399	0.495
<b>Average</b>	<b>0.479</b>	<b>0.445</b>	<b>0.407</b>	<b>0.575</b>	<b>0.389</b>	<b>0.554</b>
<b>STDE</b>	<b>0.275</b>	<b>0.386</b>	<b>0.232</b>	<b>0.247</b>	<b>0.366</b>	<b>0.195</b>
<b>t-score</b>	6.75	4.47	6.80	9.00	4.11	10.99
<b>t-dist(2t)</b>	0.00001	0.00053	0.00001	0.00000	0.00106	0.00000

Observăm în tabelul XX că majoritatea valorilor în valoare absolută sunt sub 0,6 și că nicio valoare nu este mai mare de 0,92. Acest lucru justifică validarea lui H0 fără aplicarea transformării Fisher. Cu toate acestea, deoarece există și valori mari, pentru a fi siguri de invalidarea lui H0, este necesară aplicarea transformării Fisher. Tabelul XXI prezintă rezultatele.

**TABEL XXI. VALORILE DE CORELAȚIE A FORMANȚILOR / Ə /, “ASEARĂ”, TON NEUTRU GEN FEMININ FARA FISHER CU TRANSFORMATA FISHER (preluată din [96])**

	<b>C12</b>	<b>C13</b>	<b>C14</b>	<b>C23</b>	<b>C24</b>	<b>C34</b>
<b>0745_aa_1-</b>	-0.16	-0.02	0.26	-0.13	-0.38	0.65
0745_aa_2-	0.65	0.36	0.47	0.63	1.27	0.79
0745_aa_3-	0.36	0.51	0.39	0.89	0.53	0.42
2703_aa_1-	0.65	1.24	0.92	0.57	0.63	0.55
2703_aa_2-	0.76	0.64	0.10	0.67	-0.03	0.47
2703_aa_3-	0.39	0.41	0.25	0.46	0.35	1.01
2718_aa_1-	0.61	0.52	0.08	0.75	0.68	0.36
2718_aa_2-	0.33	0.84	0.54	0.56	0.20	0.55
2718_aa_3-	1.37	1.10	0.57	1.06	0.54	1.08
29125_	0.85	1.53	1.17	1.31	1.09	1.52
29125_	0.24	-0.32	0.30	0.49	-0.04	0.13
29125_	0.80	0.89	0.48	0.47	0.01	0.56
697123	1.39	1.04	0.93	1.33	1.12	0.87
697123	0.47	0.30	0.49	1.14	0.94	0.69
697123	0.20	-0.27	0.13	0.67	0.42	0.54
<b>Average</b>	<b>0.594</b>	<b>0.585</b>	<b>0.473</b>	<b>0.724</b>	<b>0.489</b>	<b>0.679</b>

<b>STDEV</b>	<b>0.415</b>	<b>0.538</b>	<b>0.323</b>	<b>0.380</b>	<b>0.484</b>	<b>0.338</b>
<b>Fisher inverse of average</b>	<b>0.53</b>	<b>0.53</b>	<b>0.44</b>	<b>0.62</b>	<b>0.45</b>	<b>0.59</b>
<b>Fisher inverse of STDEV</b>	<b>0.39</b>	<b>0.49</b>	<b>0.31</b>	<b>0.36</b>	<b>0.45</b>	<b>0.33</b>
t-score	5.25	4.15	5.46	6.62	3.91	7.03
t.dist.2d	0.00012	0.00099	0.00008	0.00001	0.00158	0.00001

În tabelul XXI, scorul t și distanța t-distance two-tail dublă au fost obținute pentru media și STDEV produse de transformarea inversă Fisher. Cu excepția C24, toate corelațiile sunt nenule cu o valoare p mai mică de 0,1%. Pentru o verificare suplimentară, au fost calculate corelațiile pentru seria temporală compusă din concatenările tuturor valorilor formașilor (pentru toți vorbitorii și pronunțiile), pentru /ə/ în “aseară”, ton neutru; rezultatele sunt în tabelul XXII.

**TABEL XXII. CORELAȚIILE FORMAȘILOR PENTRU ÎNTREAGA SERIE COMPUSĂ PRIN CONCATENĂRI PE TOȚI VORBITORII (preluată din [96])**

<b>C12</b>	<b>C13</b>	<b>C14</b>	<b>C23</b>	<b>C24</b>	<b>C34</b>
0.448	0.418	0.343	0.584	0.563	0.596

**TABEL XXIII. VALORI DE CORELAȚIE A FORMAȘILOR, /A1/, “ASEARĂ”, TON NEUTRU, VORBITORI FEMININI, FĂRĂ TRANSFORMARE FISHER ȘI CU TRANSFORMATA FISHER (preluată din [96])**

	<b>C12</b>	<b>C13</b>	<b>C14</b>	<b>C23</b>	<b>C24</b>	<b>C34</b>
<b>average</b>	<b>0.130</b>	<b>0.157</b>	<b>0.048</b>	<b>0.555</b>	<b>0.453</b>	<b>0.495</b>
<b>STDE</b>	<b>0.574</b>	<b>0.529</b>	<b>0.503</b>	<b>0.383</b>	<b>0.348</b>	<b>0.387</b>
<b>t-score</b>	1.22	1.60	0.52	7.80	7.01	6.89
<b>t-dist(2t)</b>	0.23247	0.11989	0.60943	0.00000	0.00000	0.00000
<b>After Fisher Transform</b>						
<b>average</b>	0.228	0.248	0.067	0.791	0.597	0.683
<b>STDE</b>	0.825	0.840	0.693	0.624	0.522	0.610
<b>t-score</b>	1.49	1.59	0.52	6.83	6.15	6.04
<b>t-dist(2t)</b>	0.1485	0.1230	0.6061	0.0000	0.0000	0.0000



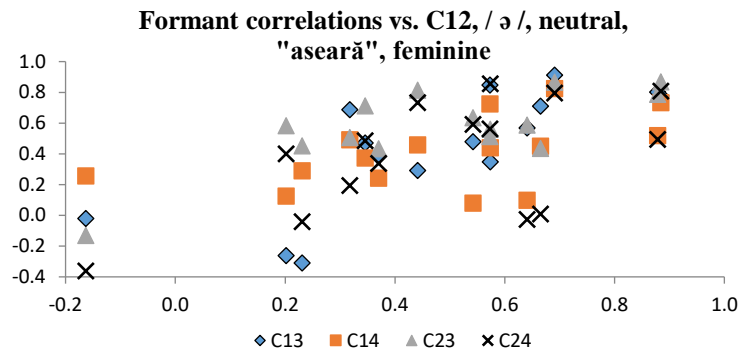


Figura 11. Corelații formanților C12. Ce mai mare panta pentru regresia liniară este de 1.07 pentru C13(C12), cu  $R^2 = 0.576$ ; coeficientul de determinare este mai mic ( $R^2 = 0.25$ , C14(C12)) (preluată din [96])

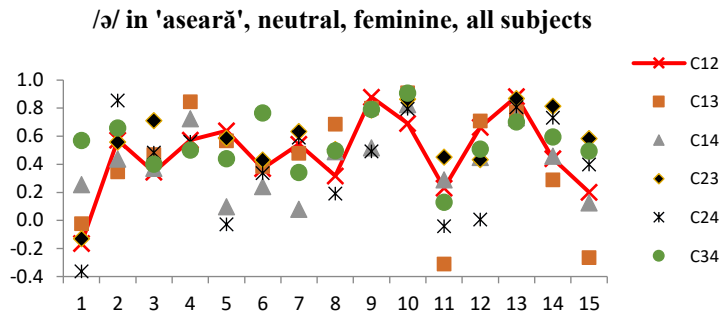


Figura 12. Corelațiile formanților pentru toți subiecții pentru vocala /ə/ din “aseară”, ton neutru, gen feminin (preluată din [96])

„Din punct de vedere fonetic, aceste corelații de formanți pentru /ə/ (ă) sunt surprinzătoare, deoarece fonemul /ă/ se află în ultima silabă, neaccentuată; în plus, se știe că acest fonem este produs cu limba în poziție aproape de echilibru (Avram, 2014) [108]. De reținut că pentru ambele cazuri (cu/fără transformare Fisher), formanții rămân corelați semnificativ. Cu toate acestea, acest lucru nu este valabil pentru toate vocalele. Primul /a/ din "aseară" nu prezintă nicio corelație, a se vedea Tabelul XIX. Pentru /a1/ în "aseară" în stare neutră, feminină, doar C23, C24 și C34 sunt corelate” [97].

# CAPITOLUL 6 – Concluzii, contribuții și direcții viitoare de cercetare

## 6.1. Sinteza analizei și contribuțiilor

Cercetarea a fost orientată spre crearea și testarea unor modele de recunoaștere și interpretare a emoțiilor prin aplicarea unor analize statistice unor parametri caracteristici semnalului vocal. A fost colectată o bază de date conținând atât pronunții susținute, cât și vorbire dinamică. Baza de date include pronunții pentru patru emoții neutru, furie, bucurie și supărare/tristețe, atât voci feminine, cât și masculine, distribuție în funcție de gen relativ echilibrată. Vocile sunt ale unor vorbitori tineri. Pentru fiecare vorbitor s-a întocmit o fișă de caracterizare a vorbitorului (vârsta, ultimul nivel de educație absolvit, eventuale patologii ale vorbirii). Baza de date este anonimată.

Cercetarea desfășurată a analizat efectul emoțiilor vorbitorului asupra formațiilor vocalelor limbii române, precum și efectul produs de poziția vocalei și accentuarea acesteia în structura cuvântului. Nu au apărut diferențe semnificative în ceea ce privește nivelul de corelație al formațiilor vocalelor limbii în funcție de genul participanților, nici în pronunția susținută, nici în vorbirea dinamică.

Studiul a sintetizat informații teoretice din mai multe cercetări privind stadiul actual în recunoașterea vorbirii și a raportat mai multe serii de rezultate statistice privind analize corelaționale ale formațiilor vocalelor limbii române. Analiza corelațională a formațiilor a fost principalul instrument de lucru. Rezultatele analizelor statistice au permis formularea mai multor concluzii: emoțiile au o influență semnificativă asupra nivelului de corelație al formațiilor vocalelor limbii române în timpul pronunției dinamice. De asemenea, s-a constatat că în cazul emoțiilor cu valență negativă (tristețe și furie), formații de ordin superior (F3 și F4) au un nivel de corelație mai ridicat decât formații de ordin inferior (F1 și F2).

Au fost obținute dovezi statistic semnificative privind corelarea formațiilor pentru anumite circumstanțe de pronunție a vocalelor, acest lucru sporind înțelegerea mecanismelor vorbirii și deschizând calea spre aplicații. Pentru starea de neutralitate emoțională variația nivelului de corelație a perechilor de formații a fost mai omogenă de la o vocală la alta. Totodată, corelația dintre F0 și formații a fost mai ridicată decât pentru celelalte emoții (bucurie, tristețe și furie). S-a constatat că vocala neaccentuată prezintă un nivel semnificativ mai ridicat de corelație a formațiilor comparativ cu vocala neaccentuată.

## 6.2. Detalii privind structura capitolelor și contribuțiile descrise în capitole

Capitolele 1 și 2 reprezintă o trecere în revista a unor rezultate din literatură. Sunt prezentate și citate un număr de paisprezece lucrări, majoritatea din literatura ultimului deceniu.

Capitolul 3 prezintă o serie de rezultate proprii, tehnici de analiză a semnalului vocal și aplicații ale acestora. Metodele folosite se referă în principal la coeficienții de regresie Pearson între formanții principali și cei secundari, pentru vocale în vorbire continuă, cu diverse emoții.

Capitolele 4 și 5 sintetizează raportarea unor rezultate proprii privind analiza corelațională atât pentru vocale susținute, cât și pentru vorbirea dinamică. Metoda folosită se referă în principal la analiza de regresie între formanții principali și cei secundari, pentru vocale în vorbire continuă, cu diverse emoții. Aceste rezultate au fost prezentate în lucrările publicate.

Capitolul 6 prezintă pe scurt unele concluzii, contribuțiile aduse și sugestii asupra unor direcții viitoare de cercetare.

## 6.3. Sinteza rezultatelor proprii

Principalele rezultate ale cercetării sunt:

a) Analizarea elementelor de corelație în timpul pronunției susținute a vocalelor limbii române în propoziții scurte. Potrivit rezultatelor, nivelul de corelare a formanților este ridicat pentru toate vocalele cu excepția vocalei /i/.

b) Analizarea dependențelor și a corelației formanților în timpul vorbirii dinamice și compararea, împreună cu îndrumătorul, a valorilor de corelație cu cele obținute în cazul pronunției susținute a vocalelor.

c) Am investigat, împreună cu îndrumătorul, influența emoțiilor asupra nivelului de corelare a formanților vocalelor limbii române în timpul pronunției dinamice. De remarcat a fost că emoțiile sunt o variabilă cu o influență puternică asupra nivelului de corelare a formanților. Un alt element identificat a fost faptul că în cazul emoțiilor cu valență negativă (tristețea și furia), formanții de ordin superior (F3 și F4) prezintă un nivel de corelație semnificativ mai ridicat comparativ cu formanții de ordin inferior (F1 și F2).

d) Am remarcat atât pentru vorbitorii de gen masculin, cât și pentru cei de gen feminin, că nu au existat diferențe semnificative între valorile de corelație ale formanților atunci când au folosit un ton neutru comparativ cu situația în care au exprimat tristețe;

e) Pentru vocala /a1/, în propoziția */vine mama/*, frecvența fundamentală (F0) este un bun predictor al formantului F1 pentru furie și tristețe, iar pentru F2 este un predictor semnificativ doar în cazul furiei și al tonului neutru. Pentru vocala /a2/, F0 este un bun predictor atât pentru F1, cât și pentru F2, doar pentru furie, rezultate obținute și analizate împreună cu îndrumătorul.

f) Corelațiile formanțelor apar în anumite circumstanțe (accentuare) și variază în funcție de poziția vocalei în structura cuvântului, precum și în funcție de emoția vorbitorului.

## 6.4. Lista lucrărilor publicate sau comunicate

### 6.4.1. Lucrări publicate în volumele conferințelor IEEE

1. **Ș.A. Ghelțu**, "*Correlation of the formants of the Romanian Vowels*", 10th International Conference on Electronics, Computers and Artificial Intelligence (ECAI), Iasi, Romania, June 2018, pp. 1-5, doi: 10.1109/ECAI.2018.8679075 (conferință IEEE)

2. **S.A. Ghelțu** and H.N. Teodorescu, "*Correlations and dependencies of the formants during speech in Romanian*," 10th International Conference on Electronics, Computers and Artificial Intelligence (ECAI), Iasi, Romania, 2018, pp. 1-6, doi: 10.1109/ECAI.2018.8678995 (conferință IEEE)

**Contribuții proprii, conform lucrării (în traducere):** "Contribuțiile autorilor. A conceput și proiectat experimente și studiul: HNT. Prelucrate datele primare: SG, HNT. Au analizat corelațiile pe serii concatenate: SG; A analizat corelațiile pe serii individuale: HNT; regresie analiza: HNT. A contribuit la colectarea datelor suplimentare SG. A scris lucrarea: HNT. Lucrare în forma finală: HNT, SG."

3. **Ș.A. Ghelțu**, "*Correlations of the vowel formants in Romanian and emotional effects on formants*". 2019 Int. Conference on Applied Electronics (AE), Pilsen, Czech R., 2019, Sept, doi: 10.23919/AE.2019.8867028 (**conferință IEEE externă**)

4. **Ș.A. Ghelțu**, "*A correlational and predictive analysis of the relationship between F0 and the formants of Romanian vowels*". 12th Int. Conf Electronics, Computers and AI (ECAI). 2020, June, doi: 10.1109/ECAI50035.2020.9223249. (conferință IEEE)

5. Teodorescu, H. N., Andrian, S., Ghiorghe, C. A., **S.A. Ghelțu**, S., & Tărăboanță, I., "*Speech under combined Schizophrenia and Salivary Flow Alterations—Preliminary Data and Results*", 2021 International Conference on Speech Technology and Human-Computer Dialogue (SpeD) 2021, October ,(pp. 25-30), doi: 10.1109/SpeD53181.2021.9587347 (conferință IEEE).

**Contribuții proprii, conform lucrării (în traducere):** „Contribuția autorilor. AS a sugerat subiectul, pe baza prezentării de către HNT a unui cadru general. HNT a propus părțile legate de vorbire și tehnică, în timp ce SA și CAG au propus aspectele medicale ale protocolului. CAG, IT și SA au efectuat înregistrările. HNT a efectuat

cea mai mare parte a procesării vorbirii și a scris lucrarea. SG a ajutat la procesarea parțială pentru doi subiecți.”

6. Teodorescu, H. N., & **Ghelțu S. A.**, „*Analysis of Formant Correlations in Emotional Speech—Do Formants Correlate?*”, 2023 International Conference on Speech Technology and Human-Computer Dialogue (SpeD), 2023, October, pp. 83-88, doi: 10.1109/SpeD59241.2023.10314902 (conferință IEEE)

*Contribuții proprii, conform lucrării (în traducere):* "Contribuțiile și responsabilitățile autorilor. HNT a conceput problema, a proiectat experimentul, a realizat întreaga prelucrare a datelor brute - parte din ele de la SG. HNT a scris lucrarea. SG a colectat parțial baza de date și a extras serii temporale de formați folosind Praat.”

#### **6.4.2. Lucrare în revistă (GS, Sciendo)**

1. **Ghelțu, Ș. A.** „*The Use of Formants’ Correlation in Assessing the Tristete State of the Speakers*”. Bulletin of the Polytechnic Institute of Iași. Electrical Engineering, Power Engineering, Electronics Section, 68(3), 21-34, doi: 10.2478/bipie-2022-0014.

\*

Conform Google Scholar, lucrările de mai sus au 15 citări.

#### **6.4.3. Lucrări comunicate**

1. H.N. Teodorescu, **Ștefan Ghelțu**, Cecilia Bolea, „*Correlations of the formants in speech – Impact of the dentition state*”, Congres Internațional al Universității „Apollonia” din Iași, Iași, România, 28 februarie – 3 martie 2019.

2. Horia-Nicolai TEODORESCU, C.A. GHIORGHE, S. ANDRIAN, **Ștefan GHELȚU** – “*Analiza F0 și a formațiilor în cuvinte cu fricative, la subiecți xerostomici*”, ”Zilele Academice Ieșene” Ediția A XXXVI-A, Simpozionul „Sisteme Inteligente Și Aplicații” Academia Română – Filiala Iași, Institutul de Informatică Teoretică (On-line în 22 octombrie 2021)

#### **6.5. Direcții viitoare de cercetare deschise de cercetarea proprie (în colectiv)**

În ceea ce privește viitorul cercetării acesteia, câteva aspecte trebuie luate în considerare. În primul rând, au fost utilizate în analize date transversale, care limitează înțelegerea proceselor investigate. În al doilea rând, rezultatele au fost obținute pornind de la un număr destul de redus de înregistrări, astfel că rezultatele obținute nu pot fi extrapolate la nivelul populației generale ci trebuie interpretate specific în contextul lotului de participanți investigat. Studii realizate la scară

mai mare, incluzând participanți din diverse subseturi de populație (vârste, regiuni diferite) ar putea indica gradul în care rezultatele noastre ar putea fi generalizate. Pentru cercetările ulterioare este de preferat realizarea unor studii longitudinale, precum și utilizarea unor manipulări experimentale, cu controlul riguros al potențialelor variabile parazite. O altă direcție ar fi reprezentată de extinderea bazei de date către un număr mai mare de propoziții. Îndrumătorul de doctorat a sugerat și realizarea unei comparații între formații vocalelor unor cuvinte comune limbilor romanice (franceză, spaniolă, italiană, etc) și investigarea măsurii în care apar diferențe generate de efectele limbii asupra formațiilor.