



**UNIVERSITATEA POLITEHNICA  
DIN BUCUREȘTI**



**Școala Doctorală de Electronică, Telecomunicații  
și Tehnologia Informației**

**Decizie nr. 971 din 08-12-2022**

# **REZUMAT TEZĂ DE DOCTORAT**

**Ing. Victor POPA**

---

**SEPARAREA OARBĂ A SURSELOR SONORE DIN  
MIXTURI BINAURALE**

**BLIND SOURCE SEPARATION FROM BINAURAL  
MIXTURES**

---

## **COMISIA DE DOCTORAT**

<b>Prof. Dr. Ing. Gheorghe BREZEANU</b> Univ. Politehnica din București	Președinte
<b>Prof. Dr. Ing. Ion MARGHESCU</b> Univ. Politehnica din București	Conducător de doctorat
<b>Prof. Dr. Ing. Daniela TĂRNICERIU</b> Univ. Tehnică "Gheorghe Asachi" din Iași	Referent
<b>Prof. Dr. Ing. Corneliu RUSU</b> Univ. Tehnică din Cluj-Napoca	Referent
<b>Prof. Dr. Ing. Cristian NEGRESCU</b> Univ. Politehnica din București	Referent

**BUCUREȘTI 2022**

---



# Cuprins

1. Introducere .....	1
1.1 Scopul tezei de doctorat .....	2
1.2 Conținutul tezei de doctorat .....	2
2. Generarea și analiza mixturilor binaurale .....	4
2.1 Analiza Fourier pe termen scurt cu reconstrucție perfectă.....	4
2.2 Măsurarea răspunsurilor la impuls ale capului.....	4
2.2.1 Sistemul de măsură .....	4
2.2.2 Egalizarea difuzorului .....	5
2.2.3 Extragerea răspunsului la impuls al capului și postprocesarea .....	5
2.2.4 Funcțiile de transfer ale capului măsurate .....	6
2.3 Analiza funcțiilor de transfer ale capului .....	6
2.4 Evaluarea performanțelor de separare a mixturilor.....	6
2.4.1 Evaluarea performanțelor pentru semnale cu câștig invariant în timp ....	7
2.4.2 Evaluarea performanțelor pentru semnale cu distorsiuni introduse de filtre invariante în timp .....	7
2.5 Concluzii .....	7
3. Determinarea direcției de sosire a unei surse sonore folosind descompunerea adaptivă a valorilor proprii.....	8
3.1 Metoda de descompunere a valorilor proprii .....	8
3.1.1 Sistemul de ecuații .....	8
3.1.2 Algoritmul adaptiv .....	8
3.2 Rezultate experimentale .....	9
3.3 Concluzii .....	10
4. Algoritmul maximizării mediei statistice pentru separarea surselor sonore din mixturi binaurale.....	11
4.1 Modelarea valorilor de pe spectrograme folosind indicii binaurali .....	11
4.2 Modelarea valorilor de pe spectrogramă folosind vectorii de mixare.....	12
4.3 Combinarea modelului binaural cu modelul vectorial într-o implementare a algoritmului maximizării medie statistice.....	12
4.3.1 Algoritmul maximizării mediei statistice.....	12
4.3.2 Alinierea permutării în frecvență folosind probabilitatea a posteriori pe baza vectorilor de mixare .....	13
4.3.3 Alinierea permutării în frecvență prin inițializarea corespunzătoare a parametrilor .....	14
4.4 Rezultate experimentale .....	14
4.5 Concluzii .....	15
5. Algoritmul Bayes variațional pentru separarea surselor sonore din mixturi binaurale.....	16
5.1 Formularea generală a inferenței variaționale.....	16
5.2 Modelarea statistică a setului de observații.....	17
5.2.1 Modelarea observațiilor extrase din spectrograme .....	17
5.2.2 Distribuțiile a priori.....	17

5.3	Procesul de optimizare .....	18
5.4	Algoritmul Bayes variațional .....	19
5.5	Rezultate experimentale .....	19
5.6	Concluzii .....	20
6.	Separare oarbă a surselor sonore folosind factorizarea matriceală nenegativă.....	21
6.1	Factorizarea matriceală nenegativă cu un singur canal.....	21
6.1.1	Proprietăți ale divergenței Itakura-Saito .....	21
6.1.2	Algoritmul de factorizare matriceală nenegativă.....	22
6.1.3	Utilizarea funcțiilor de cost auxiliare pentru optimizare .....	22
6.1.4	Rezultate experimentale pentru factorizarea matriceală nenegativă cu un singur canal.....	23
6.2	Factorizarea matriceală nenegativă multicanal .....	24
6.2.1	Modelarea Gaussiană .....	24
6.2.2	Modelarea spectrală .....	24
6.2.3	Combinarea modelului Gaussian cu cel spectral pentru NMF multicanal .....	25
6.2.4	Algoritmul NMF multicanal bazat pe minimizarea unei funcții de cost auxiliare .....	25
6.2.5	Extragerea estimatelor surselor din rezultatele NMF multicanal .....	26
6.2.6	Rezultate experimentale NMF multicanal .....	27
6.3	Concluzii .....	27
7.	Analiza matriceală independentă de rang scăzut .....	28
7.1	Analiza vectorilor independenți – IVA .....	28
7.2	Extinderea IVA cu NMF pentru a obține ILRMA .....	28
7.3	Restricționarea NMF multicanal pentru a obține ILRMA .....	29
7.4	Rezultate experimentale .....	30
7.5	Concluzii .....	30
8.	Concluzii.....	31
8.1	Rezultate obținute.....	31
8.2	Contribuții originale .....	36
8.3	Lista lucrărilor originale.....	37
8.4	Perspective de dezvoltare ulterioară.....	39
	Bibliografie .....	40

# Capitolul 1

## Introducere

Separarea oarbă a surselor din mixturi este un subiect abordat într-o gamă largă de domenii cum ar fi: procesarea semnalelor multimedia, comunicații, medicină, inginerie industrială, precum și altele. Conceptul de "separare oarbă" se referă la faptul că atât sistemul de mixare cât și sursele originale nu sunt cunoscute, cu toate că, în majoritatea aplicațiilor se fac o serie de presupuneri generale asupra tipului de surse implicate în sistemul de mixare. În scopul separării este accesibil doar rezultatul final al mixării ce poate avea una sau mai multe dimensiuni.

### 1.1 Prezentarea domeniului tezei de doctorat

Tehnica separării oarbe a surselor sonore (Blind Source Separation - BSS) este studiată de zeci de ani, iar cercetarea este încă în desfășurare. Există multe aplicații ce pot fi dezvoltate pe baza acestei tehnologii, însă accentul principal este pus pe prelucrarea semnalelor audio, cum ar fi rezolvarea efectului "cocktail party" folosind inteligență artificială, extragerea vorbirii țintă într-un mediu zgomotos pentru rezultate mai bune de recunoaștere a vorbirii, separarea fiecărei părți instrumentale muzicale a unei înregistrări audio pentru analiza muzicii. În tabelul 1.1 sunt prezentate principalele metode de separare oarbă a surselor sonore în funcție de o serie de criterii.

**Tabelul 1.1** Clasificarea principalelor metode de separare oarbă a surselor sonore

	Un singur canal $M = 1$	Multicanal $N$ - număr de surse $M$ - număr de canale/mixturi		
		Supradeterminată $N < M$	Determinată $N = M$	Subdeterminată $N > M$
Fără antrenare	NMF	PCA Toate după reducere dimensionalitate	DUET ICA, IVA MNMF ILRMA	GMM și alte metode de grupare statistică MNMF
Cu antrenare		Rețele neurale adânci (DNN)		

## 1.1 Scopul tezei de doctorat

Vom porni de la semnale de tip voce atât feminină cât și masculină și le vom combina într-o serie de mixturi realizate prin înregistrări binaurale reverberante ce simulează percepția spațială a sistemului auditiv uman în încăperi reale.

Un prim obiectiv este de a realiza și analiza o serie de înregistrări binaurale personalizate pentru a extrage funcțiile de transfer ale capului și răspunsurile la impuls corespunzătoare. Cu ajutorul acestor răspunsuri, precum și a altor răspunsuri din literatură, se vor genera mixturi prin simularea plasării simultane la diverse unghiuri a unui număr prestabilit de surse sonore într-un spațiu audio virtual.

Un al doilea obiectiv este determinarea unghiului de incidentă al unei surse din înregistrări cu două microfoane, aplicat pe înregistrări binaurale reverberante. Determinarea acestor unghiuri va ajuta la procesul de separare pentru metodele de separare a surselor sonore dezvoltate ulterior.

Al treilea obiectiv constă în implementarea și dezvoltarea de metode de separare a surselor sonore pentru mixturi determinate și subdeterminate bazate pe clasificarea punctelor de pe transformatele Fourier pe termen scurt ale mixturilor. Modelarea surselor va fi de tipul mixtură de Gaussiene și se vor dezvolta metode de optimizare a parametrilor distribuțiilor de probabilitate astfel încât să se realizeze o clasificare performantă a punctelor de pe transformatele Fourier pe termen scurt.

Ultimul obiectiv constă în implementarea unor metode de separare oarbă a surselor sonore bazate pe nenegativitatea datelor, testând în felul acesta performanța de separare pentru mixturi binaurale reverberante.

În final toate metodele vor fi comparate în scopul determinării metodelor optime pentru separarea mixturilor în funcție de cantitatea de reverberație din încăperile respective.

Metodele de separare oarbă a semnalelor vocale din mixturi binaurale pot avea aplicații pentru:

- Aprofundarea înțelegerii modului de separare a surselor sonore de către sistemul auditiv uman;
- Preprocesarea semnalelor vocale cu scopul creșterii inteligibilității înainte de a fi utilizate de sisteme de transcriere automată a vorbirii sau pentru sisteme de comandă vocală;
- Reducerea reverberației pentru semnale vocale înregistrate în încăperi reale;
- Accentuarea anumitor vorbitori (cum ar fi moderatorul) în aplicații de audio-conferință.

## 1.2 Conținutul tezei de doctorat

Capitolul al doilea va analiza metodele de prelucrare a mixturilor cu scopul pregătirii acestora pentru algoritmi de separare utilizați în teză. În primă fază se vor prezenta semnalele audio utilizate precum și modul de prelucrare al acestora folosind transformata Fourier pe termen scurt cu reconstrucție perfectă. Pentru a avea control asupra modului de creare al mixturilor se va dezvolta o metodă de măsurare a funcțiilor de transfer ale capului și a răspunsurilor la impuls corespunzătoare. Acestea se vor dovedi utile pentru plasarea surselor sonore într-un spațiu acustic virtual. Totodată, se va prezenta metoda de albire a semnalelor ca pas de preprocesare al mixturilor, metodă utilă pentru algoritmi bazați pe modelarea statistică a surselor. În

ultima parte a capitolului se vor discuta principalele mărimi utilizate pentru evaluarea performanței de separare a algoritmilor.

În capitolul trei se va investiga metoda de determinare a direcției de sosire a unei surse sonore într-un mediu reverberant folosind descompunerea adaptivă a valorilor proprii. Metoda se va dovedi utilă în determinarea informațiilor de localizare a surselor sonore ce vor îmbunătăți performanțele metodelor ulterior prezentate.

Capitolul patru va cerceta metoda maximizării mediei statistice pentru estimarea surselor sonore din mixturi ce are la bază premisa modelării spectrogramelor mixturilor observate sub formă de mixturi de distribuții Gaussiene. Observațiile analizate vor fi reprezentate de înregistrări binaurale simulate fie cu ajutorul unor răspunsuri la impuls binaurale măsurate în diverse încăperi, fie cu răspunsuri la impuls ale capului și înregistrări binaurale măsurate cu metoda descrisă în capitolul 2.

Capitolul cinci va explora și dezvolta o nouă metodă de estimare a surselor sonore din mixturi, similară cu cea prezentată în capitolul 4, cu diferența că modelarea se realizează tot cu mixturi de Gaussiene, însă se condiționează parametrii modelului cu o serie de distribuții a priori. Algoritmul rezultat se bazează pe modelul Bayes variațional și prezintă o serie de avantaje față de cel bazat pe maximizarea mediei statistice.

În capitolul șase se vor analiza și implementa metode bazate pe nenegativitatea mixturilor observate. În prima parte a capitolului se va prezenta metoda de factorizare matriceală nenegativă pentru cazul cu un singur canal, discutând despre aplicațiile acesteia și modul de implementare, iar în a doua parte se va cerceta extensia la cazul multicanal, discutând despre performanța acesteia atât pentru cazul mixturilor determinate cât și pentru cele subdeterminate, testând aceleași mixturi binaurale reverberante.

În capitolul șapte se va explora metoda de separare bazată pe analiza matriceală independentă de rang scăzut, ca o combinație între analiza vectorilor independenți și factorizarea matriceală nenegativă, discutând despre performanțele acesteia în cazul mixturilor determinate reverberante.

În final, capitolul de concluzii va prezenta rezultatele comparative ale metodelor precum și contribuțiile personale și dezvoltările viitoare.

# Capitolul 2

## Generarea și analiza mixturilor binaurale

### 2.1 Analiza Fourier pe termen scurt cu reconstrucție perfectă

O parte din cercetările în separarea surselor sonore se bazează pe descompunerea semnalelor corespunzătoare observațiilor într-o reprezentare în timp și frecvență. Pentru a realiza aceasta se utilizează transformata Fourier pe termen scurt a semnalelor discrete utilizând o fereastră care glisează în timp. Metoda de reconstrucție perfectă descrisă se bazează pe metoda de suprapunere și sumare.

Transformata Fourier pe termen scurt se obține ulterior prin aplicarea transformatei Fourier discrete (TFD) asupra fiecărui bloc de  $N$  elemente:

$$S(k, m) = TFD\{x(n) \cdot w_a(n - mp)\} \quad (1.1)$$

În cazul ferestrei Hann se poate realiza reconstrucția perfectă dacă pasul este  $N/4$ , iar  $N$  este multiplu de 4.

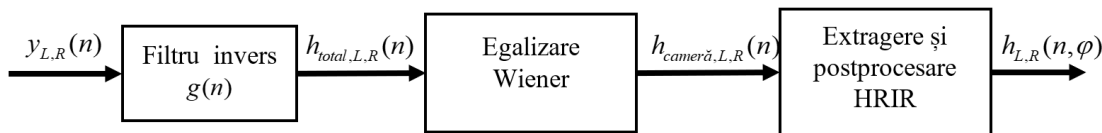
### 2.2 Măsurarea răspunsurilor la impuls ale capului

Funcțiile de transfer ale capului (*Head Related Transfer Function - HRTF*) și răspunsurile la impuls corespunzătoare în domeniul timp (*Head Related Impulse Response - HRIR*) au un rol decisiv în localizarea sunetului pentru oameni și, un set individualizat de măsurare, poate îmbunătăți acuratețea localizării. Măsurarea se dovedește a fi dificilă în absența unei camere anecoide.

#### 2.2.1 Sistemul de măsură

Sistemul de măsurare constă dintr-o platformă rotativă pe care este așezat ascultătorul, un set de microfoane binaurale calibrate purtate de ascultător și un difuzor plasat la o distanță fixă. Răspunsul la impuls acustic este măsurat cu metoda ESS [1], folosind implementarea îmbunătățită din [2]. Semnalul înregistrat, notat cu  $y_{L,R}(n)$  este apoi procesat conform schemei bloc reprezentată în figura 2.6.

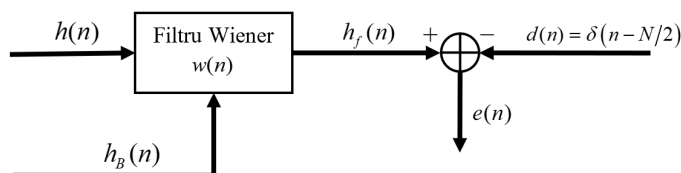




**Figura 2.1** Lanțul de măsurare și procesare HRIR

## 2.2.2 Egalizarea difuzorului

Din răspunsul la impuls măsurat a fost selectată doar porțiunea de undă directă și a fost aplicată o procedură de egalizare similară cu cea descrisă în [3] și îmbunătățită în [4]. Scopul este de a obține un răspuns în frecvență plat pentru partea liniară a sistemului acustic compus din amplificator audio, difuzor și microfon. Schema generală a filtrului Wiener este prezentată în figura 2.7, unde  $h(n)$  este răspunsul inițial al difuzorului,  $h_b(n)$  este răspunsul la impuls al unui filtru oprește bandă predeterminat,  $h_f(n)$  este răspunsul egalizat al difuzorului,  $d(n)$  este răspunsul dorit și  $e(n)$  este semnalul de eroare.

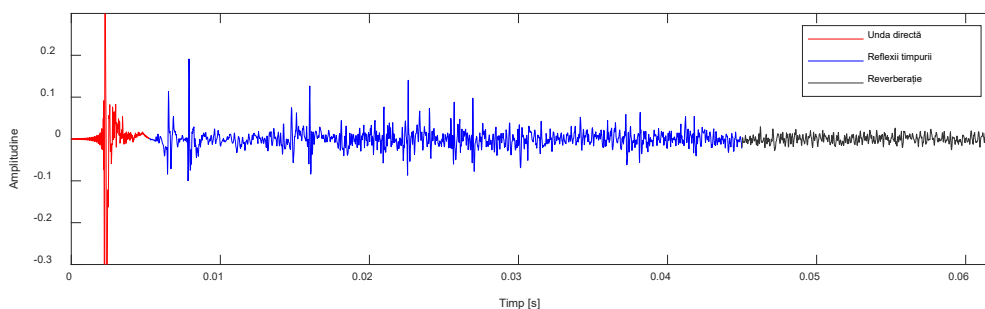


**Figura 2.2** Schemă bloc egalizare difuzor folosind filtrarea Wiener

## 2.2.3 Extragerea răspunsului la impuls al capului și postprocesarea

În figura 2.9 prezentăm principalele părți ale unui răspuns la impuls: unda directă, reflexiile timpurii și partea de reverberație. Prin urmare, am putea separa sunetul direct ce corespunde răspunsului la impuls al capului dacă primele reflexii sunt întârziate față de unda directă cu cel puțin  $t_{\min} = 4$  ms. Dacă distanța dintre sursa de sunet și microfon este  $d_{dir}$ , putem calcula lungimea minimă a căii de propagare pentru prima reflexie,  $d_{refl}$ , prin:

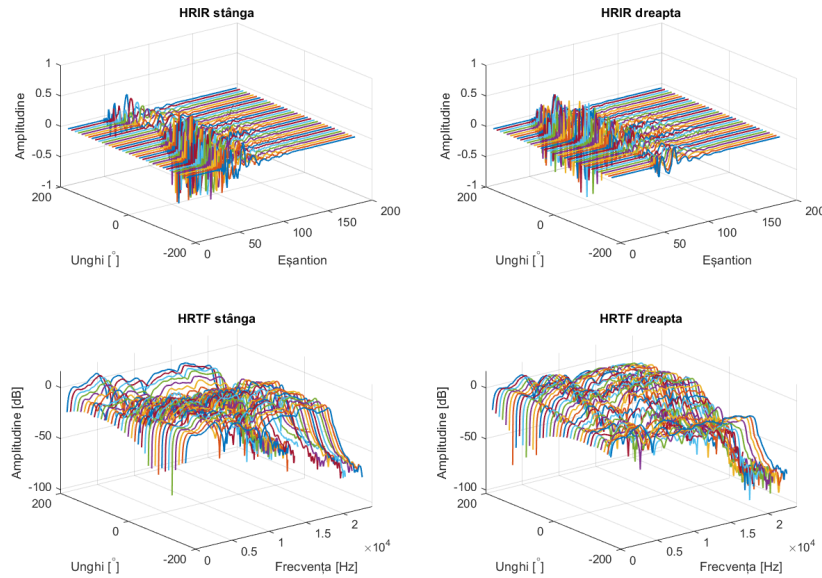
$$d_{refl} = d_{dir} + c \cdot t_{\min} \quad (1.2)$$



**Figura 2.3** Componenta unui răspuns la impuls acustic măsurat într-o încăpere

## 2.2.4 Funcțiile de transfer ale capului măsurate

Semnalul de test utilizat pentru măsurarea HRTF-urilor a fost generat cu o bandă de frecvență între 20 Hz și 20 kHz și o durată de 3 secunde, profitând de avantajele discutate în [5]. Semnalul de test se repetă de 36 de ori, o repetiție pentru fiecare increment de unghi de  $10^\circ$ , cu o pauză de 5 secunde între ele. Un set de funcții de transfer ale capului măsurate cu metoda descrisă poate fi observat în figura 2.11.



*Figura 2.4 Exemple de HRIR și HRTF măsurate*

## 2.3 Analiza funcțiilor de transfer ale capului

O primă analiză va fi prin transformata Fourier pe termen scurt aplicată întâi pe semnalele inițiale ce vor fi de tip voce. Semnalele sunt înregistrate la frecvența de eșantionare de 16 kHz, iar pentru analiza Fourier pe termen scurt s-a utilizat o fereastră Hann de 1024 de eșantioane și o suprapunere de 75%, garantând în felul acesta reconstruirea perfectă. Pentru a obține indicii interaurali vom exprima raportul dintre transformatele Fourier pe termen scurt a celor două canale:

$$\frac{X_L(\omega, t)}{X_R(\omega, t)} = 10^{\frac{\alpha(\omega, t)}{20}} \cdot e^{j\phi(\omega, t)} \quad (1.3)$$

unde  $\alpha(\omega, t)$  va fi indicele diferenței de nivel, ILD, exprimat în dB, iar  $\phi(\omega, t)$  este indicele de diferență de fază, IPD, exprimat în radiani.

## 2.4 Evaluarea performanțelor de separare a mixturilor

Evaluarea performanțelor algoritmilor de separare a surselor sonore este o problemă de sine stătătoare care este adresată în literatură în încercarea de a stabili o normă comună și unanim acceptată pentru a descrie performanțele diferiților algoritmi de separare a surselor sonore și pentru a oferi o modalitate de a îi compara [6].

### 2.4.1 Evaluarea performanțelor pentru semnale cu câștig invariant în timp

Raportul dintre sursă și distorsiune (Source to Distortion Ratio – SDR):

$$SDR := 10 \log_{10} \frac{\|s_{\text{șintă}}\|^2}{\|e_{\text{interf}} + e_{\text{zg}} + e_{\text{artef}}\|^2} \quad (1.4)$$

Raportul dintre sursă și interferențe (Source to Interference Ratio – SIR):

$$SIR := 10 \log_{10} \frac{\|s_{\text{șintă}}\|^2}{\|e_{\text{interf}}\|^2} \quad (1.5)$$

Raportul dintre sursă și zgomot (Source to Noise Ratio - SNR):

$$SNR := 10 \log_{10} \frac{\|s_{\text{șintă}} + e_{\text{interf}}\|^2}{\|e_{\text{zg}}\|^2} \quad (1.6)$$

Raportul dintre sursă și artefacte (Source to Artifacts Ratio - SAR):

$$SAR := 10 \log_{10} \frac{\|s_{\text{șintă}} + e_{\text{interf}} + e_{\text{zg}}\|^2}{\|e_{\text{artef}}\|^2} \quad (1.7)$$

### 2.4.2 Evaluarea performanțelor pentru semnale cu distorsiuni introduse de filtre invariante în timp

Dacă sunt permise ca distorsiuni filtrările cu filtre invariante în timp, atunci semnalul țintă  $s_{\text{șintă}}$  nu mai reprezintă o variantă amplificată (sau atenuată) a semnalului original  $s_i$ , ci rezultatul unei filtrări. Cu alte cuvinte,  $s_{\text{șintă}}$  se obține prin sumarea unui set de variante întârziate ale semnalului original cu anumite amplitudini exprimat prin:

$$s_{\text{șintă}}(n) = \sum_{k=0}^{L-1} h(k) \cdot s_i(n-k) \quad (1.8)$$

## 2.5 Concluzii

În urma discuțiilor referitoare la implementarea transformatei Fourier pe termen scurt cu reconstrucție perfectă am ales ca parametri pentru metoda suprapunerii și sumării fereastra de tip Hann cu o suprapunerea de 75% între cadre consecutive.

A fost prezentată o metodă de măsurare rapidă pentru extragerea HRIR-urilor și a HRTF-urilor asociate acestora din răspunsurile la impuls măsurate în camere reverberante. Din testele preliminare, indicii spațiali au fost păstrați, iar răspunsurile la impuls rezultate se dovedesc a fi foarte asemănătoare cu cele măsurate în condiții anecoice.

# Capitolul 3

## Determinarea direcției de sosire a unei surse sonore folosind descompunerea adaptivă a valorilor proprii

### 3.1 Metoda de descompunere a valorilor proprii

Această metodă face parte din categoria algoritmilor de identificare oarbă a canalelor pentru sisteme acustice SIMO, fiind un caz particular pentru o sursă și două microfoane.

#### 3.1.1 Sistemul de ecuații

Vom considera cazul ideal în care nu este prezent zgomotul și vom exploata proprietatea de comutativitate a convoluției [7]:

$$(y_1 * h_2)(n) = (s * h_1 * h_2)(n) = (y_2 * h_1)(n) \quad (1.9)$$

$$(y_1 * h_2)(n) - (y_2 * h_1)(n) = 0 \quad (1.10)$$

#### 3.1.2 Algoritm adaptiv

Schema bloc finală a algoritmului adaptiv se poate observa în figura 3.3.

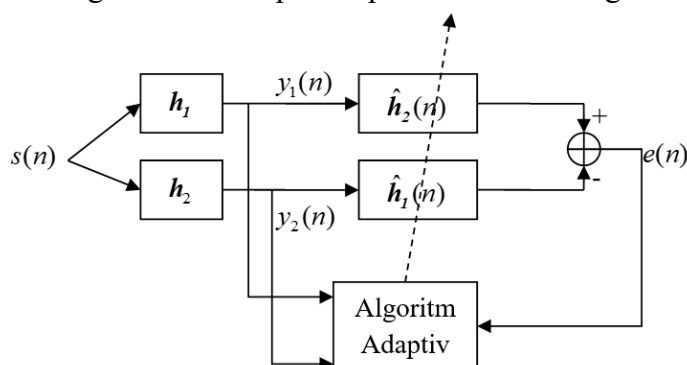


Figura 3.1 Schema bloc a algoritmului adaptiv

Algoritmul poate fi descris în felul următor:

**Inițializare:** $N$ 

$$\hat{\mathbf{h}}_2(0) = [0 \quad \dots \quad 0 \quad 1 \quad 0 \quad \dots \quad 0]^T, \text{ cu } 1 \text{ pe poziția } N/2$$

$$\hat{\mathbf{h}}_1(0) = [0 \quad 0 \quad \dots \quad 0]^T$$

$$\hat{\mathbf{h}}(0) = \begin{bmatrix} \hat{\mathbf{h}}_2(0)^T & -\hat{\mathbf{h}}_1(0)^T \end{bmatrix}^T$$

**Pentru**  $n = 0, \dots, N-1$ 

$$e(n) = \hat{\mathbf{h}}(n)^T \cdot \mathbf{y}(n)$$

$$\hat{\mathbf{h}}(n+1) = \frac{\hat{\mathbf{h}}(n) - \mu e(n) \mathbf{y}(n)}{\|\hat{\mathbf{h}}(n) - \mu e(n) \mathbf{y}(n)\|}$$

**După convergență:**

$$\tau_{12} = \tau_2 - \tau_1 = \left( N/2 - \arg \left\{ \min \left\{ \hat{\mathbf{h}}_1(n) \right\} \right\} \right) / F_s$$

## 3.2 Rezultate experimentale

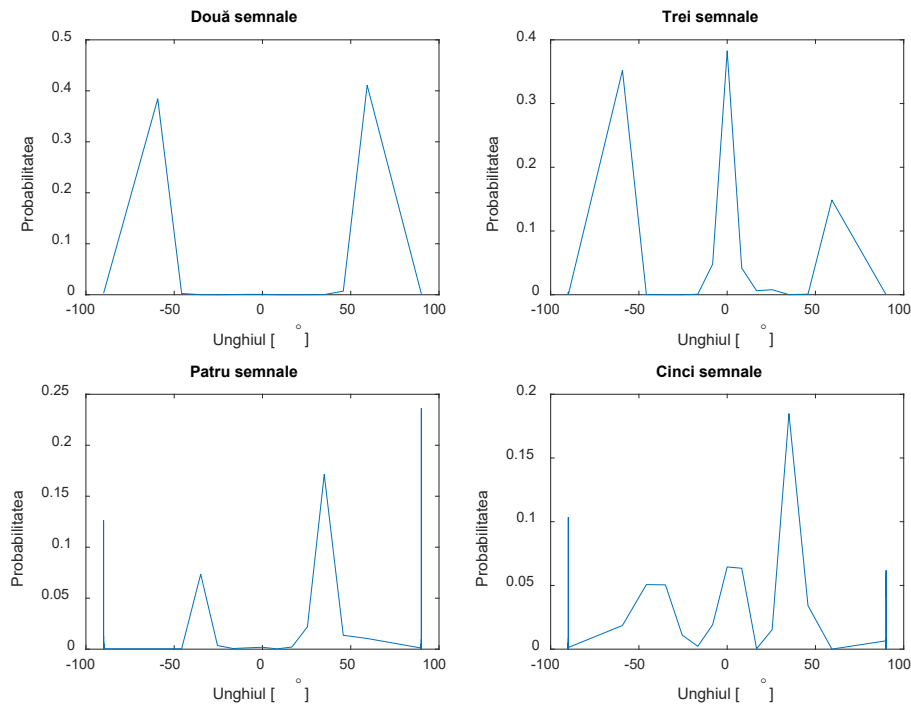
Pentru a testa algoritmul a fost folosită următoarea configurație:

- 2 microfoane cardioide plasate la distanța  $d = 0,1$  m ;
- 2 surse de sunet plasate la unghiurile  $\theta = 90^\circ$  și  $\theta = 135^\circ$ , iar distanța de la surse la microfoane  $r = 2,3$  m ;
- Camera reverberantă cu  $T_{60} \cong 0.53s$  ;
- A fost reprodus un semnal vocal, iar înregistrarea s-a făcut cu  $F_s = 48$  kHz .

Pentru a testa detecția de surse multiple vom aplica în continuare algoritmul pentru mixturi de 2, 3, 4 și 5 surse în mixturi binaurale realizate în camera descrisă în capitolul anterior. Unghiurile alese pentru fiecare mixtură sunt:

- Mixtură de 2 surse:  $-60^\circ$  și  $60^\circ$  ;
- Mixtură de 3 surse:  $-60^\circ$ ,  $0^\circ$  și  $60^\circ$  ;
- Mixtură de 4 surse:  $-90^\circ$ ,  $-30^\circ$ ,  $30^\circ$  și  $90^\circ$  ;
- Mixtură de 5 surse:  $-90^\circ$ ,  $-45^\circ$ ,  $0^\circ$ ,  $45^\circ$  și  $90^\circ$  ;

După aplicarea algoritmului s-a determinat histograma unghiurilor pentru a vizualiza mai bine unghiurile detectate. Rezultatele se pot observa în figura 3.15.



**Figura 3.2** Histograma unghiurilor detectate folosind algoritmul de descompunere adaptivă a valorilor proprii pentru mixturi binaurale de 2,3,4 și 5 surse

### 3.3 Concluzii

Algoritmul adaptiv bazat pe descompunerea valorilor proprii oferă rezultate satisfăcătoare atunci când la un moment de timp este prezentă o singură sursă și se dovedește robust chiar și în prezența reverberației puternice. Ca o comparație, s-a aplicat algoritmul GCC pentru același semnal, însă detecția corectă a direcției s-a putut face începând cu o mărime de fereastră de 5000 de eșantioane.

Algoritmul tinde să prefere semnalele de tip zgomot direcțional deoarece îndeplinesc mai bine condiția de varietate, ceea ce duce la concluzia că o soluție pentru detecția de semnale vocale ar fi pre-albirea acestora. În prezența a două surse acesta converge către cea care îndeplinește cel mai bine condițiile algoritmului, adică tinde către canalul cu valoarea proprie cea mai mică.

În situația aplicării algoritmului asupra mixturilor binaurale de două sau mai multe surse, am observat că algoritmul detectează corect toate unghiurile implicate și se dovedește a fi o măsură bună pentru determinarea numărului de surse și a unghiurilor acestora. Metoda va fi utilă în cadrul pasului de preprocesare și inițializare al algoritmilor de separare a surselor sonore din mixturi binaurale ce vor fi prezentați în capitolele următoare.

# Capitolul 4

## Algoritmul maximizării mediei statistice pentru separarea surselor sonore din mixturi binaurale

Ideea principală este de a modela diferența interaurală de fază (*Interaural Phase Difference - IPD*), diferența interaurală de nivel (*Interaural Level Difference – ILD*) ca în [8] și vectorii de mixare ca în [9] la nivel de unitate timp-frecvență cu modele de tip mixtură de Gaussiene pentru fiecare sursă și fiecare tip de observație, similar cu algoritmul din [10].

### 4.1 Modelarea valorilor de pe spectrograme folosind indicii binaurali

Spectrograma interaurală, adică raportul dintre spectrogramele corespunzătoare canalelor stânga și dreapta, se determină în felul următor:

$$X_L(\omega, t) = TFFTS \{x_L(n)\} \quad (1.11)$$

$$X_R(\omega, t) = TFFTS \{x_R(n)\} \quad (1.12)$$

$$\frac{X_L(\omega, t)}{X_R(\omega, t)} = 10^{\frac{\alpha(\omega, t)}{20}} \cdot e^{j\phi(\omega, t)}, \quad (1.13)$$

unde  $X_L(\omega, t)$  și  $X_R(\omega, t)$  sunt transformatele Fourier pe termen scurt pentru cele două semnale înregistrate la fiecare frecvență discretă  $\omega$  și moment de timp discret  $t$ .

Vom modela observațiile pentru ILD și IPD ca distribuții Gaussiene cu media  $\mu_i(\omega)$  și varianța  $\eta_i^2(\omega)$  pentru  $\alpha(\omega, t)$ , respectiv, media  $\xi_{i,\tau}(\omega)$  și varianța  $\sigma_{i,\tau}^2(\omega)$  pentru  $\hat{\phi}(\omega, t; \tau)$  [8] în felul următor:

$$p(\alpha(\omega, t) | \mu_i(\omega), \eta_i^2(\omega)) = \mathcal{N}(\alpha(\omega, t) | \mu_i(\omega), \eta_i^2(\omega))$$
$$\mathcal{N}(\alpha(\omega, t) | \mu_i(\omega), \eta_i^2(\omega)) = \frac{1}{\eta_i(\omega)\sqrt{2\pi}} \exp\left(-\frac{(\alpha(\omega, t) - \mu_i(\omega))^2}{2 \cdot \eta_i^2(\omega)}\right) \quad (1.14)$$

$$\begin{aligned}
p(\hat{\phi}(\omega, t; \tau) | \xi_{i,\tau}(\omega), \sigma_{i,\tau}^2(\omega)) &= \mathcal{N}(\hat{\phi}(\omega, t; \tau) | \xi_{i,\tau}(\omega), \sigma_{i,\tau}^2(\omega)) \\
\mathcal{N}(\hat{\phi}(\omega, t; \tau) | \xi_{i,\tau}(\omega), \sigma_{i,\tau}^2(\omega)) &= \frac{1}{\xi_{i,\tau}(\omega) \sqrt{2\pi}} \exp\left(-\frac{(\hat{\phi}(\omega, t; \tau) - \xi_{i,\tau}(\omega))^2}{2 \cdot \sigma_{i,\tau}^2(\omega)}\right)
\end{aligned}
\tag{1.15}$$

## 4.2 Modelarea valorilor de pe spectrogramă folosind vectorii de mixare

Pentru a modela astfel de vectori pentru fiecare sursă, urmăm ideea descrisă în [11] și folosim următoarea densitate de probabilitate de tipul complex Gaussiană multidimensională:

$$\begin{aligned}
p(\mathbf{X}(\omega, t) | \mathbf{a}_i(\omega), \gamma_i(\omega)) &= \mathcal{N}(\mathbf{X}(\omega, t) | \mathbf{a}_i(\omega), \gamma_i^2(\omega)) \\
&= \frac{1}{(\pi \cdot \gamma_i^2(\omega))^2} \exp\left(-\frac{\|\mathbf{X}(\omega, t) - (\mathbf{a}_i^H(\omega) \cdot \mathbf{X}(\omega, t)) \cdot \mathbf{a}_i(\omega)\|^2}{\gamma_i^2(\omega)}\right)
\end{aligned}
\tag{1.16}$$

unde  $\mathbf{a}_i$  are semnificație de medie bidimensional-complexă și este un centroid cu norma unitară  $\|\mathbf{a}_i\| = 1$ , iar  $\gamma_i^2$  pentru distribuția bidimensional-complexă Gaussiană.

## 4.3 Combinarea modelului binaural cu modelul vectorial într-o implementare a algoritmului maximizării medii statistice

Putem astfel exprima funcția de plauzibilitate în domeniul logaritmic al modelului compus:

$$\mathcal{L}(\Theta) = \sum_{\omega, t} \ln(p(\alpha(\omega, t), \phi(\omega, t), \mathbf{X}(\omega, t) | \Theta))
\tag{1.17}$$

### 4.3.1 Algoritmul maximizării mediei statistice

Algoritmul de maximizare a mediei statistice presupune determinarea iterativă a parametrilor optimi ce cresc la fiecare iterație plauzibilitatea logaritmică ca un punct timp-frecvență să aparțină unei surse și este împărțit în două etape:

1. **Pasul de mediere statistică (estimare - E)** în care se calculează media statistică a variabilei aleatoare ascunse  $z_{i,\tau}(\omega, t)$  pe baza observațiilor și a parametrilor estimați  $\hat{\Theta}$ , echivalent cu evaluarea modelului;



$$\begin{aligned}
q_{i,\tau}(\omega, t) &= p\left(z_{i,\tau}(\omega, t) \mid \alpha(\omega, t), \phi(\omega, t), \mathbf{X}(\omega, t), \hat{\Theta}\right) \\
&\propto p\left(z_{i,\tau}(\omega, t), \alpha(\omega, t), \phi(\omega, t), \mathbf{X}(\omega, t) \mid \hat{\Theta}\right) \\
&= \psi_{i,\tau}(\omega) \cdot \mathcal{N}\left(\alpha(\omega, t) \mid \mu_i(\omega), \eta_i^2(\omega)\right) \\
&\quad \cdot \mathcal{N}\left(\hat{\phi}(\omega, t; \tau) \mid \xi_{i,\tau}(\omega), \sigma_{i,\tau}^2(\omega)\right) \\
&\quad \cdot \mathcal{N}\left(\mathbf{X}(\omega, t) \mid \mathbf{a}_i(\omega), \gamma_i^2(\omega)\right).
\end{aligned} \tag{1.18}$$

2. **Pasul de maximizare (M)** în care  $Q$  este maximizat în funcție de  $\Theta$  pe baza valorii mediei statistice a lui  $z_{i,\tau}(\omega, t)$  determinată la pasul anterior. Pentru a simplifica notațiile vom defini operatorul de mediere ponderată:

$$\mathcal{M}_{a,b}\{x\} = \frac{\sum_{a,b} x \cdot q_{i,\tau}(\omega, t)}{\sum_{a,b} q_{i,\tau}(\omega, t)}, \tag{1.19}$$

Folosind acest operator putem deduce relațiile de actualizare pentru parametrii de model la **pasul de maximizare**:

$$\mu_i(\omega) = \mathcal{M}_{i,\tau}\{\alpha(\omega, t)\} \tag{1.20}$$

$$\eta_i^2(\omega) = \mathcal{M}_{i,\tau}\left\{\left(\alpha(\omega, t) - \mu_i(\omega)\right)^2\right\} \tag{1.21}$$

$$\xi_{i,\tau}(\omega) = \mathcal{M}_t\left\{\hat{\phi}(\omega, t; \tau)\right\} \tag{1.22}$$

$$\sigma_{i,\tau}^2(\omega) = \mathcal{M}_t\left\{\left(\hat{\phi}(\omega, t; \tau) - \xi_{i,\tau}(\omega)\right)^2\right\} \tag{1.23}$$

$$\mathbf{R}_i(\omega) = \sum_{i,\tau} q_{i,\tau}(\omega, t) \cdot \mathbf{X}(\omega, t) \cdot \mathbf{X}^H(\omega, t) \tag{1.24}$$

$$\gamma_i^2(\omega) = \mathcal{M}_{i,\tau}\left\{\left\|\mathbf{X}(\omega, t) - \left(\mathbf{a}_i^H(\omega) \cdot \mathbf{X}(\omega, t)\right) \cdot \mathbf{a}_i(\omega)\right\|^2\right\} \tag{1.25}$$

$$\psi_{i,\tau}(\omega) = \frac{1}{T} \sum_t q_{i,\tau}(\omega, t) \tag{1.26}$$

### 4.3.2 Alinierea permutării în frecvență folosind probabilitatea a posteriori pe baza vectorilor de mixare

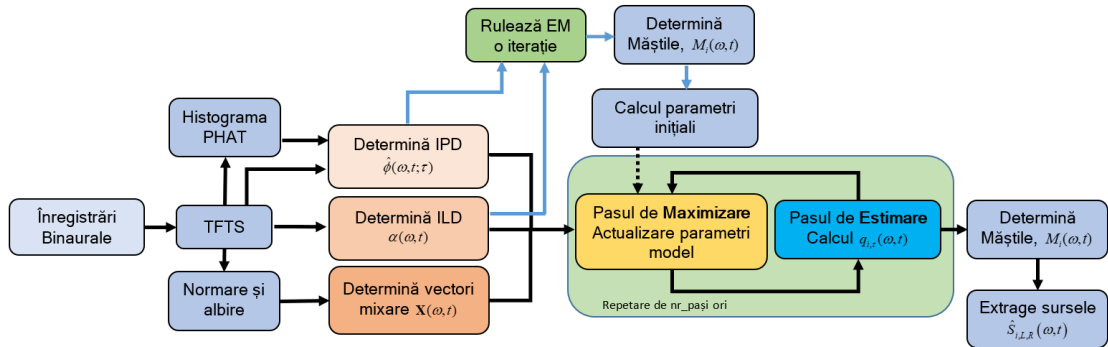
Luând în considerare modelarea dată de indicii binaurali și vectorii de mixare se poate realiza o aliniere a posteriori a ordinii surselor, conform descrierii din [9]. Aceasta presupune determinarea probabilității a posteriori

$$\begin{aligned}
q_{k,\tau}(\omega, t) &= p\left(z_{k,\tau}(\omega, t) \mid \alpha(\omega, t), \phi(\omega, t), \mathbf{X}(\omega, t), \hat{\Theta}\right) \\
&= p\left(C_k \mid \alpha(\omega, t), \phi(\omega, t), \mathbf{X}(\omega, t), \hat{\Theta}\right)
\end{aligned} \tag{1.27}$$

ce descrie faptul că sursa cu numărul  $k$ , descrisă de clasa  $C_k$ , este dominantă în cadrul setului de observații  $\Psi(\omega, t) = \overset{Not}{\{\alpha(\omega, t), \phi(\omega, t), \mathbf{X}(\omega, t)\}}$ .

### 4.3.3 Alinierea permutării în frecvență prin inițializarea corespunzătoare a parametrilor

Vom considera o abordare alternativă folosind informațiile din indicii binaurali, descrisă și implementată în [10] și [12]. Întrucât algoritmul maximizării mediei statistice poate fi inițializat fie din pasul de mediere statistică, fie din pasul de maximizare și, de asemenea, de obicei nu există informații prelabile despre sistemul de mixare, vom realiza întâi inițializarea măștii pentru spectrogramă și apoi vom estima valorile inițiale ale  $\mathbf{a}_i(\omega)$  și  $\gamma_i(\omega)$  pe baza spectrogramei mascate [10]. Schema bloc a algoritmului de maximizare a mediei statistice este prezentată în figura 4.6.

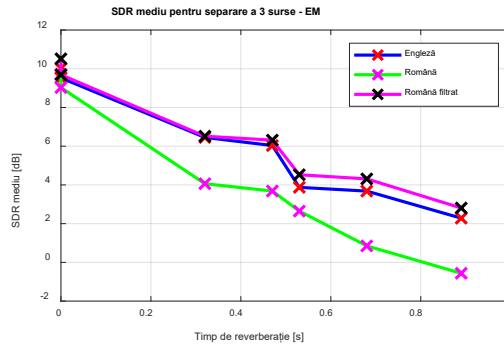


*Figura 4.1 Schema bloc a algoritmului de maximizare a expectației cu alinierea permutărilor*

## 4.4 Rezultate experimentale

Primul set de experimente s-a desfășurat folosind setul de înregistrări de tip voce din baza de date TIMIT [13]. Pentru a realiza setul de experimente s-au ales aleator 15 fraze rostite din baza de date și s-a micșorat durata acestora la 2,5 s. Pentru a putea investiga și independența de limbă a algoritmului s-au utilizat și 7 fraze rostite în limba română cu vorbitori de sex feminin și masculin. Pentru a obține mixturile binaurale s-a realizat apoi convoluția dintre răspunsurile la impuls binaurale și înregistrări. Primul set de răspunsurile la impuls binaurale măsurate în încăperi sunt cele determinate de Hummersone [14]. În plus față de acestea am adăugat și răspunsurile la impuls ale capului, notate ulterior cu HRIR, precum și o cameră din cadrul facultății de Electronică, Telecomunicații și Tehnologia Informației, UPB, notată cu E. Pentru experimentul cu 3 voci alese aleator în limba engleză sau în limba română, plasate la unghiurile  $-60^\circ$ ,  $0^\circ$  și  $60^\circ$ , folosind cele 7 tipuri de încăperi și mediind apoi SDR peste toate rezultatele obținem performanțele prezentate în figura

4.15. Semnalele în română au fost filtrate cu un banc de filtre parametrice astfel încât în medie spectrul să fie asemănător cu cel în engleză.



*Figura 4.2 Performanța de separare medie a trei surse în funcție de timpul de reverberație – algoritmul maximizării medie statistice (EM)*

## 4.5 Concluzii

În acest capitol a fost prezentată o implementare a algoritmului de maximizare a mediei statistice bazat pe modelarea transformatelor Fourier pe termen scurt a înregistrărilor binaurale. Pentru a evita ambiguitățile de fază ce duc la alierea spațială în frecvență s-a prezentat o metodă de asocierea a diferenței de timp de întârziere cu o serie discretă de valori. Observațiile au fost apoi modelate folosind distribuții Gaussiene și a fost dezvoltat un algoritm de optimizare bazat pe maximizarea medie statistice. Pentru a evita problema permutării ordinii surselor între frecvențe diferite s-a prezentat o metodă de inițializare a responsabilităților inițiale pe baza unei combinații dintre histograma PHAT și algoritmul adaptiv de descompunere a valorilor proprii prezentat în capitolul anterior.

# Capitolul 5

## Algoritmul Bayes variațional pentru separarea surselor sonore din mixturi binaurale

Spre deosebire de algoritmul maximizării mediei statistice descris în secțiunea anterioară, inferența Bayesiană își propune să modeleze în întregime distribuția de probabilitate a parametrilor a posteriori, descriși în cadrul algoritmului de maximizare a medie statistice de către mediile, varianțele și ponderile de mixare.

### 5.1 Formularea generală a inferenței variaționale

Având la dispoziție setul de date  $\mathbf{x}$ , setul de date ascunse  $\mathbf{z}$  și parametri modelului statistic ce caracterizează observațiile și datele ascunse,  $\Theta$ , vrem să estimăm probabilitatea  $p(\mathbf{x}, \mathbf{z}, \Theta)$ . Pentru datele disponibile, putem defini plauzibilitatea logaritmică marginală numită și dovada modelului ca:

$$\mathcal{L}_\Theta(\mathbf{x}) = \ln \iint p(\mathbf{x}, \mathbf{z}, \Theta) d\mathbf{z} d\Theta. \quad (1.28)$$

Pentru a introduce dependența parametrilor vom defini o distribuție de probabilitate variațională a posteriori  $q(\mathbf{z}, \Theta | \mathbf{x}, \theta)$  care aproximează adevăratul posterior  $p(\mathbf{z}, \Theta | \mathbf{x})$ , unde  $\theta$  se referă la un set de *hiperparametri* care modelează distribuțiile parametrilor. Putem astfel defini plauzibilitatea logaritmică variațională dependentă de setul de hiperparametri:

$$\begin{aligned} \mathcal{L}_\theta(\mathbf{x}) &= \iint q(\mathbf{z}, \Theta | \mathbf{x}, \theta) \ln \left( \frac{p(\mathbf{x}, \mathbf{z}, \Theta)}{q(\mathbf{z}, \Theta | \mathbf{x}, \theta)} \right) d\mathbf{z} d\Theta + \\ &+ \iint q(\mathbf{z}, \Theta | \mathbf{x}, \theta) \ln \left( \frac{q(\mathbf{z}, \Theta | \mathbf{x}, \theta)}{p(\mathbf{z}, \Theta | \mathbf{x})} \right) d\mathbf{z} d\Theta \\ &= L_{\text{inf}}(\mathbf{x}) + D_{KL}(q \| p) \end{aligned} \quad (1.29)$$

unde  $L_{\text{inf}}(\mathbf{x})$  este aceeași limită inferioară cu cea din (5.4), iar  $D_{KL}(q \| p)$  este divergența Kullback-Leibler dintre densitatea de probabilitate variațională a posteriori  $q(\mathbf{z}, \Theta | \mathbf{x}, \theta)$  și probabilitatea a posteriori adevărată  $p(\mathbf{z}, \Theta | \mathbf{x})$ .

## 5.2 Modelarea statistică a setului de observații

### 5.2.1 Modelarea observațiilor extrase din spectrograme

Cele trei seturi de observații,  $\mathbf{X}(\omega, t)$ ,  $\alpha(\omega, t)$  și  $\phi(\omega, t)$ , cu  $\phi(\omega, t)$  transformat în  $\hat{\phi}(\omega, t; \tau)$ , vor fi combinate într-o mixtură de probabilități, unde  $T$  este numărul total de cadre temporale,  $\Omega$  este numărul de canale de frecvență,  $I$  este numărul de surse, iar  $\Upsilon$  este numărul de diferențe interaurale de timp. Distribuția de probabilitate a variabilelor latente condiționată de coeficienții de mixaj se va exprima [15]:

$$p(\mathbf{Z} | \gamma_x) = \prod_{t=1}^T \prod_{i=1}^I \prod_{\tau=1}^{\Upsilon} \gamma_{x,i}^{z_{t,i,\tau}} \quad (1.30)$$

Distribuția setului de observații din spectrograme, condiționată de variabilele latente va fi exprimată sub formă de produs de distribuții Gaussiene complexe:

$$p(\mathbf{X} | \mathbf{Z}, \boldsymbol{\mu}_x, \boldsymbol{\lambda}_x) = \prod_{t=1}^T \prod_{i=1}^I \prod_{\tau=1}^{\Upsilon} \mathcal{N}_c(\mathbf{x}_t | \boldsymbol{\mu}_{x,i}, \boldsymbol{\lambda}_{x,i}^{-1})^{z_{t,i,\tau}}, \quad (1.31)$$

Modelarea diferenței interaurale de nivel  $\boldsymbol{\alpha} = \{\alpha_1, \dots, \alpha_T\}$  se face în mod asemănător, cu o mixtură de Gaussiene, cu media  $\boldsymbol{\mu}_\alpha = \{\mu_{\alpha,i}\}$  și precizia  $\boldsymbol{\lambda}_\alpha = \{\lambda_{\alpha,i}\}$ , având coeficienții de mixaj  $\gamma_\alpha = \{\gamma_{\alpha,i}\}$ :

$$p(\mathbf{Z} | \gamma_\alpha) = \prod_{t=1}^T \prod_{i=1}^I \prod_{\tau=1}^{\Upsilon} \gamma_{\alpha,i}^{z_{t,i,\tau}} \quad (1.32)$$

$$p(\boldsymbol{\alpha} | \mathbf{Z}, \boldsymbol{\mu}_\alpha, \boldsymbol{\lambda}_\alpha) = \prod_{t=1}^T \prod_{i=1}^I \prod_{\tau=1}^{\Upsilon} \mathcal{N}(\alpha_t | \mu_{\alpha,i}, \lambda_{\alpha,i}^{-1})^{z_{t,i,\tau}} \quad (1.33)$$

Observațiile corespunzătoare diferenței interaurale de fază vor fi modelate asemănător cu diferența interaurală de nivel folosind o mixtură de Gaussiene:

$$p(\mathbf{Z} | \gamma_\phi) = \prod_{t=1}^T \prod_{i=1}^I \prod_{\tau=1}^{\Upsilon} \gamma_{\phi,i,\tau}^{z_{t,i,\tau}} \quad (1.34)$$

$$p(\hat{\boldsymbol{\Phi}} | \mathbf{Z}, \boldsymbol{\mu}_\phi, \boldsymbol{\lambda}_\phi) = \prod_{t=1}^T \prod_{i=1}^I \prod_{\tau=1}^{\Upsilon} \mathcal{N}(\hat{\phi}_{t,\tau} | \mu_{\phi,i,\tau}, \lambda_{\phi,i,\tau}^{-1})^{z_{t,i,\tau}} \quad (1.35)$$

### 5.2.2 Distribuțiile a priori

Coeficienții de mixaj, având structura unei distribuții multinomiale vor avea ca distribuție a priori distribuția Dirichlet de forma:

$$p(\boldsymbol{\gamma}) = \text{Dir}(\boldsymbol{\gamma} | \mathbf{a}_0) = B(\mathbf{a}_0) \prod_{i=1}^I \prod_{\tau=1}^{\Upsilon} \gamma_{i,\tau}^{a_{i,\tau}-1}, \quad (1.36)$$

Pentru media și precizia distribuțiilor Gaussiene vom folosi distribuții a priori Gauss-Gamma de forma:

$$p(\boldsymbol{\mu}_x, \boldsymbol{\lambda}_x) = p(\boldsymbol{\mu}_x | \boldsymbol{\lambda}_x) p(\boldsymbol{\lambda}_x) = \prod_{i=1}^I \mathcal{N}_c(\boldsymbol{\mu}_{x,i} | \mathbf{m}_{x,0}, (\lambda_{x,i} \boldsymbol{\beta}_{x,0} \mathbf{I})^{-1}) \mathcal{G}(\lambda_{x,i} | b_{x,0}, c_{x,0}) \quad (1.37)$$

$$p(\boldsymbol{\mu}_\alpha, \boldsymbol{\lambda}_\alpha) = p(\boldsymbol{\mu}_\alpha | \boldsymbol{\lambda}_\alpha) p(\boldsymbol{\lambda}_\alpha) = \prod_{i=1}^I \mathcal{N}(\mu_{\alpha,i} | m_{\alpha,0}, (\lambda_{\alpha,i} \beta_{\alpha,0})^{-1}) \mathcal{G}(\lambda_{\alpha,i} | b_{\alpha,0}, c_{\alpha,0}) \quad (1.38)$$

$$\begin{aligned} p(\boldsymbol{\mu}_\phi, \boldsymbol{\lambda}_\phi) &= p(\boldsymbol{\mu}_\phi | \boldsymbol{\lambda}_\phi) p(\boldsymbol{\lambda}_\phi) \\ &= \prod_{i=1}^I \prod_{\tau=1}^Y \mathcal{N}(\mu_{\phi,i,\tau} | m_{\phi,0}, (\lambda_{\phi,i,\tau} \beta_{\phi,0})^{-1}) \mathcal{G}(\lambda_{\phi,i,\tau} | b_{\phi,0}, c_{\phi,0}) \end{aligned} \quad (1.39)$$

### 5.3 Procesul de optimizare

Optimizarea valorilor parametrilor de model presupune maximizarea plauzibilității logaritmice similar cu discuția din prima parte a capitolului, folosind (5.1).

#### *Pasul de actualizare a responsabilităților*

Prin responsabilități ne referim la probabilitatea ca un punct timp-frecvență sa aparțină unei anumite surse și să provină dintr-o anumită direcție. Din ecuația (5.42) observăm că distribuția a priori este mediată peste  $z_{t,i,\tau}$  și poate fi exprimată prin:

$$\ln q^{opt}(\mathbf{Z}) = \sum_{t=1}^T \sum_{i=1}^I \sum_{\tau=1}^Y z_{t,i,\tau} \ln \rho_{t,i,\tau} + const, \quad (1.40)$$

unde

$$\begin{aligned} \ln \rho_{t,i,\tau} &= \mathbb{E}_{\gamma_{i,\tau}} [\ln \gamma_{i,\tau}] - \ln \pi - \ln 2\pi + \mathbb{E}_{\lambda_{x,i}} [\ln \lambda_{x,i}] + \frac{1}{2} \mathbb{E}_{\lambda_{\alpha,i}} [\ln \lambda_{\alpha,i}] + \\ &+ \frac{1}{2} \mathbb{E}_{\lambda_{\phi,i,\tau}} [\ln \lambda_{\phi,i,\tau}] - \mathbb{E}_{\boldsymbol{\mu}_{x,i}, \lambda_{x,i}} [\lambda_{x,i} \|\mathbf{x}_t - (\boldsymbol{\mu}_{x,i}^H \mathbf{x}_t) \boldsymbol{\mu}_{x,i}\|^2] - \\ &- \frac{1}{2} \mathbb{E}_{\boldsymbol{\mu}_{\alpha,i}, \lambda_{\alpha,i}} [\lambda_{\alpha,i} (\alpha_t - \mu_{\alpha,i})^2] - \frac{1}{2} \mathbb{E}_{\boldsymbol{\mu}_{\phi,i,\tau}, \lambda_{\phi,i,\tau}} [\lambda_{\phi,i,\tau} (\hat{\phi}_{t,\tau} - \mu_{\phi,i,\tau})^2] \end{aligned} \quad (1.41)$$

#### *Pasul de actualizare a hiperparametrilor*

În consecință actualizarea hiperparametrilor pentru coeficienții de mixaj se realizează prin:

$$a_{i,\tau} = \sum_{t=1}^T r_{t,i,\tau} + a_0 \quad (1.42)$$

Pentru hiperparametrii distribuțiilor a priori a observațiilor vectorilor de mixaj se obține:

$$\begin{aligned} \boldsymbol{\beta}_{x,i} &= \beta_{x_0} \mathbf{I} - \sum_{t=1}^T \sum_{\tau=1}^Y r_{t,i,\tau} (\mathbf{x}_t \mathbf{x}_t^H) \\ \mathbf{m}_{x,i} &= \boldsymbol{\beta}_{x,i}^{-1} \beta_{x_0} \mathbf{I} \mathbf{m}_{x,0} \\ b_{x,i} &= b_{x,0} + \sum_{t=1}^T \sum_{\tau=1}^Y r_{t,i,\tau} \\ c_{x,i} &= c_{x,0} + \sum_{t=1}^T \sum_{\tau=1}^Y r_{t,i,\tau} (\mathbf{x}_t^H \mathbf{x}_t) + \mathbf{m}_{x,0}^H \beta_{x,0} \mathbf{I} \mathbf{m}_{x,0} - \mathbf{m}_{x,i}^H \boldsymbol{\beta}_{x,i} \mathbf{m}_{x,i} \end{aligned} \quad (1.43)$$

Asemănător determinăm hiperparametrii distribuțiilor pentru ILD și IPD:

$$\begin{aligned}
\beta_{\alpha,i} &= \beta_{\alpha,0} + \sum_{t=1}^T \sum_{\tau=1}^Y r_{t,i,\tau} \\
m_{\alpha,i} &= \beta_{\alpha,i}^{-1} \left( \sum_{t=1}^T \sum_{\tau=1}^Y r_{t,i,\tau} \alpha_t + m_{\alpha,0} \beta_{\alpha,0} \right) \\
b_{\alpha,i} &= b_{\alpha,0} + \frac{1}{2} \sum_{t=1}^T \sum_{\tau=1}^Y r_{t,i,\tau} \\
c_{\alpha,i} &= c_{\alpha,0} + \frac{1}{2} \sum_{t=1}^T \sum_{\tau=1}^Y r_{t,i,\tau} \alpha_t^2 + \frac{1}{2} m_{\alpha,0}^2 \beta_{\alpha,0} - \frac{1}{2} m_{\alpha,i}^2 \beta_{\alpha,i}
\end{aligned} \tag{1.44}$$

Calculul limitei inferioare  $L_{\text{inf}}(\mathbf{X}, \mathbf{a}, \hat{\Phi})$

$$\begin{aligned}
L_{\text{inf}}(\mathbf{X}, \mathbf{a}, \hat{\Phi}) &= \sum_{\mathbf{Z}} \int q(\mathbf{Z}, \Theta) \ln \left( \frac{p(\mathbf{X}, \mathbf{a}, \hat{\Phi}, \mathbf{Z}, \Theta)}{q(\mathbf{Z}, \Theta)} \right) d\Theta \\
&= \mathbb{E} \left[ \ln p(\mathbf{X}, \mathbf{a}, \hat{\Phi}, \mathbf{Z}, \Theta) \right] - \mathbb{E} \left[ \ln q(\mathbf{Z}, \Theta) \right]
\end{aligned} \tag{1.45}$$

## 5.4 Algoritmul Bayes variațional

Algoritmul este prezentat în figura 5.1 sub formă de schema bloc.

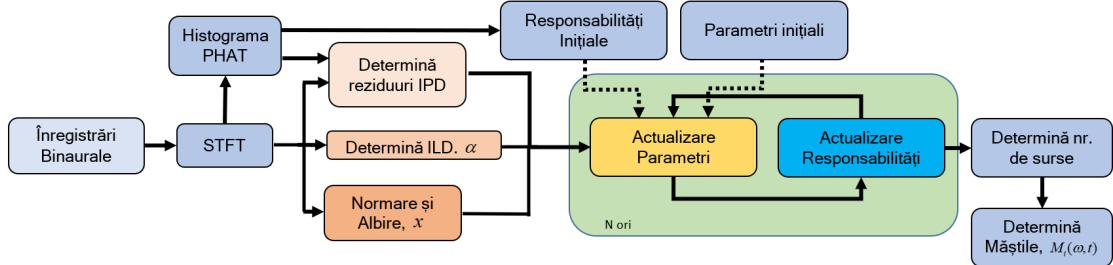


Figura 5.1. Algoritmul Bayes variațional

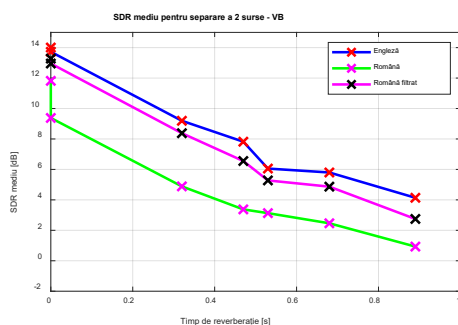
## 5.5 Rezultate experimentale

Vom rula algoritmul Bayes variațional pentru cele 7 tipuri de încăperi alegând aleator sursele din setul stabilit pentru un număr de voci inițiale de la 2 la 5 pentru a determina acuratețea identificării corecte a numărului de surse din mixturi pentru fiecare tip de încăpere. Rezultatele sunt prezentate în tabelul 5.1.

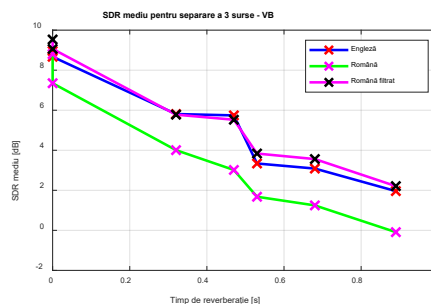
**Tabelul 5.1** Acuratețea [%] determinării numărului de surse din mixtură în funcție de cameră și număr de surse inițiale

Acuratețea [%]	Camera						
	Nr. surse	N	HRIR	A	B	C	D
2	100	100	98	97	92	85	95
3	100	100	95	93	86	76	91
4	93	96	81	76	69	74	73
5	62	68	53	45	38	35	42

Reluând experimentul asupra mixturilor de două surse pentru algoritmul Bayes variațional descris în rezultatele capitolului anterior obținem performanțele în funcție de timpul de reverberație din figura 5.6. Rezultatele pentru mixturi de trei surse sunt prezentate în figura 5.7.



**Figura 5.2** Performanța de separare medie a două surse în funcție de timpul de reverberație – algoritmul Bayes variațional (VB)



**Figura 5.3** Performanța de separare medie a trei surse în funcție de timpul de reverberație – algoritmul Bayes variațional (VB)

## 5.6 Concluzii

A fost prezentată o metodă de separare a vorbirii reverberante din mixturi binaurale bazată pe teoria Bayes variațională. Algoritmul propus beneficiază de faptul că inferența Bayesiană este mai puțin sensibilă la inițializări necorespunzătoare și, de asemenea, determină automat numărul de surse, ceea ce poate fi un avantaj în cazul înregistrărilor reale. A fost studiată convergența algoritmului prin determinarea limitei inferioare variaționale a plauzibilității logaritmice și s-a studiat influența unghiului dintre semnalele sursă asupra gradului de separare.



# Capitolul 6

## Separare oarbă a surselor sonore folosind factorizarea matriceală nenegativă

### 6.1 Factorizarea matriceală nenegativă cu un singur canal

Factorizarea matriceală nenegativă (*Nonnegative matrix factorization - NMF*) este o metodă des utilizată de reducere dimensională a unei cantități de date și este folosită pentru reprezentarea datelor nenegative [16]. Pornind de la o matrice  $\mathbf{V}$  de dimensiuni  $\Omega \times T$  ce va reprezenta o formă a spectrogramei semnalului cu  $\Omega$  numărul de puncte în frecvență și  $T$  numărul de puncte pe axa timpului discret, prin NMF se pune problema determinării factorizării:

$$\mathbf{V} \approx \mathbf{W}\mathbf{H}, \quad (1.46)$$

Rezolvarea ecuației (6.1) presupune determinarea matricelor  $\mathbf{W}$  și  $\mathbf{H}$  în urma minimizării:

$$\min_{\mathbf{W}, \mathbf{H} \geq 0} D(\mathbf{V}|\mathbf{W}\mathbf{H}), \quad (1.47)$$

unde  $D(\mathbf{V}|\mathbf{W}\mathbf{H})$  este o funcție de cost.

În continuare ne vom concentra în mod specific pe divergența Itakura-Saito (IS) ca și funcție de cost pentru NMF dată de:

$$d_{IS}(x|y) = \frac{x}{y} - \ln \frac{x}{y} - 1 \quad (1.48)$$

#### 6.1.1 Proprietăți ale divergenței Itakura-Saito

*Legătura cu divergența  $\beta$*

Divergența IS este un caz limită al divergenței  $\beta$  pe care o definim ca:

$$d_{\beta}(x|y) \stackrel{def}{=} \begin{cases} \frac{1}{\beta(\beta-1)} [x^{\beta} + (\beta-1)y^{\beta} - \beta xy^{\beta-1}] & , \beta \in \mathbb{R} \setminus \{0,1\} \\ x \ln \frac{x}{y} + (y-x) & , \beta = 1 \\ \frac{x}{y} - \ln \frac{x}{y} - 1 & , \beta = 0 \end{cases} \quad (1.49)$$

*Proprietatea de invarianță la scală*

Pentru orice valoare a lui  $\beta$  are loc următoarea proprietate:

$$d_{\beta}(\gamma x|\gamma y) = \gamma^{\beta} d_{\beta}(x|y) \quad (1.50)$$

Aceasta implică faptul că divergența IS, pentru care  $\beta = 0$ , este invariantă la scală deoarece  $d_{\beta}(\gamma x|\gamma y) = d_{\beta}(x|y)$ . Invarianța la scală înseamnă că se acordă aceeași pondere relativă coeficienților mici și mari ai lui  $\mathbf{V}$  în funcția de cost în sensul că o potrivire slabă a factorizării pentru un coeficient de putere redusă  $\mathbf{V}_{\omega,t}$  va costa la fel de mult ca o potrivire slabă pentru un coeficient de putere mai mare  $\mathbf{V}_{\omega',t}$ .

### 6.1.2 Algoritm de factorizare matriceală nenegativă

Regulile de actualizare rezultate conduc la algoritmul de mai jos.

#### **Algoritm IS-NMF cu actualizări multiplicative**

**Intrare:** matricea nenegativă  $\mathbf{V}$

**Ieșire:** matricele  $\mathbf{W}$  și  $\mathbf{H}$  optime astfel încât  $\mathbf{V} \approx \mathbf{WH}$

1. **Inițializare**  $\mathbf{W}$  și  $\mathbf{H}$  cu valori nenegative aleatoare

2. **Pentru**  $i = 1 : nr\_iter$

$$\mathbf{H} \leftarrow \mathbf{H} \circ \frac{\mathbf{W}^T \left( (\mathbf{WH})^{(-2)} \circ \mathbf{V} \right)}{\mathbf{W}^T (\mathbf{WH})^{(-1)}}$$

$$\mathbf{W} \leftarrow \mathbf{W} \circ \frac{\left( (\mathbf{WH})^{(-2)} \circ \mathbf{V} \right) \mathbf{H}^T}{(\mathbf{WH})^{(-1)} \mathbf{H}^T}$$

Normalizare  $\mathbf{W}$  și  $\mathbf{H}$

**Sfârșit pentru**

### 6.1.3 Utilizarea funcțiilor de cost auxiliare pentru optimizare

O metodă de a obține valorile optime pentru  $\mathbf{W}$  și  $\mathbf{H}$  prin minimizarea funcției de cost este abordarea sub numele de algoritm de majorizare-minimizare [17], [18].

Fie  $\theta$  un set de variabile obiectiv. Pentru o funcție de cost, numită în continuare funcție obiectiv  $\mathcal{C}(\theta)$ , vom crea o funcție auxiliară  $\mathcal{C}^+(\theta, \tilde{\theta})$  cu un set de variabile auxiliare  $\tilde{\theta}$  ce satisface următoarele două condiții:

1. Funcția auxiliară este mai mare sau egală cu funcția obiectiv:

$$\mathcal{C}^+(\theta, \tilde{\theta}) \geq \mathcal{C}(\theta) \quad (1.51)$$

2. Minimizarea funcției auxiliare după variabilele auxiliare duce la funcția obiectiv:

$$\min_{\tilde{\theta}} \mathcal{C}^+(\theta, \tilde{\theta}) = \mathcal{C}(\theta) \quad (1.52)$$

Folosind aceste două condiții se poate dezvolta un algoritm de minimizare al funcției obiectiv  $\mathcal{C}(\theta)$  într-un mod indirect prin minimizarea funcției auxiliare  $\mathcal{C}^+(\theta, \tilde{\theta})$  iterativ. Vom obține ecuațiile de actualizare multiplicative similare cu cele din algoritmul general din secțiunea anterioară:

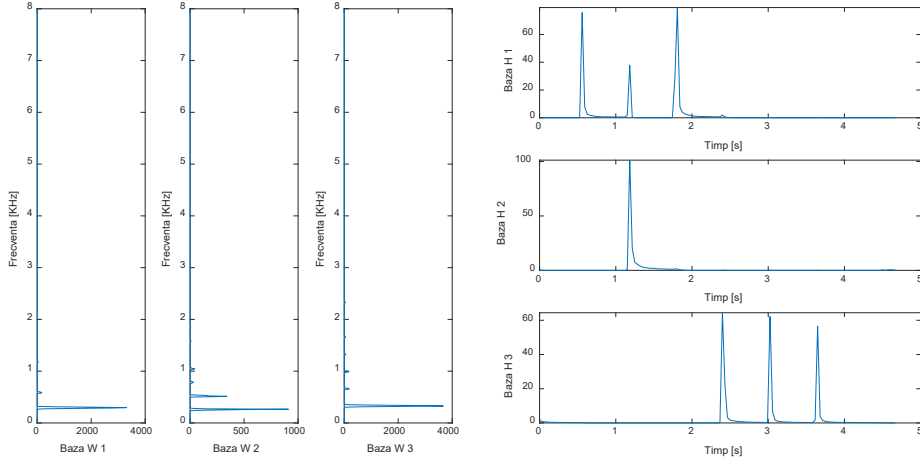
$$w_{\omega,n} \leftarrow w_{\omega,n} \sqrt{\frac{\sum_{t=1}^T h_{n,t} |x_{\omega,t}|^2}{\hat{x}_{\omega,t}^2} \frac{\sum_{t=1}^T h_{n,t}}{\hat{x}_{\omega,t}}} \quad (1.53)$$

$$h_{n,t} \leftarrow h_{n,t} \sqrt{\frac{\sum_{\omega=1}^{\Omega} w_{\omega,n} |x_{\omega,t}|^2}{\hat{x}_{\omega,t}^2} \frac{\sum_{\omega=1}^{\Omega} w_{\omega,n}}{\hat{x}_{\omega,t}}} \quad (1.54)$$

#### 6.1.4 Rezultate experimentale pentru factorizarea matriceală nenegativă cu un singur canal

Testarea gradului de separare s-a realizat pe un semnal audio de test cu frecvența de eșantionare de  $F_s = 16000$  Hz și de durată 4.7 s. Deoarece teoria din spatele factorizării matriceale nenegative are la bază presupunerea că observațiile (în cazul nostru punctele timp-frecvență de pe spectrogramă) sunt distribuite conform unei distribuții super-Gaussiene, sau Gaussiene, am ales în prima fază semnalul de tip sunet de pian, mai exact secvența de note muzicale Mi, Re, Do, Re, Mi, Mi, Mi.

Utilizând algoritmul NMF și regulile de actualizare multiplicative (6.40) și (6.41) se determină iterativ în 100 de pași bazele  $\mathbf{W}$  și  $\mathbf{H}$  (figura 6.3). Matricele ce compun bazele au fost inițializate aleator cu valori pozitive între 0 și 1.



**Figura 6.1** Bazele NMF pentru semnalul de tip pian

În urma factorizării matriceale nenegative pentru mixtura de semnale de tip voce s-au determinat bazele și s-au extras semnalele. Extragerea nu se realizează corect deoarece punctele de pe spectrograma de voce nu se clasifică ca distribuție de tip super-Gaussiană.

## 6.2 Factorizarea matriceală nenegativă multicanal

Având în vedere succesul NMF pentru separarea surselor cu un singur canal, au existat mai multe încercări de extindere pentru mixturi multicanal [19].

### 6.2.1 Modelarea Gaussiană

Având în vedere ecuația de mixaj multicanal și ipotezele de independență, coeficienții TFTS de mixare pot fi modelați cu o distribuție de forma:

$$\mathbf{x}_{\omega,t} \sim \mathcal{N}_c \left( 0, \sum_{n=1}^N \mathbf{R}_{n,\omega,t} v_{n,\omega,t} \right) \quad (1.55)$$

### 6.2.2 Modelarea spectrală

Modelarea spectrală se poate realiza prin modelarea NMF a fiecărei surse în parte și constă în structurarea varianțelor surselor  $v_{n,\omega,t}$ :

$$v_{n,\omega,t} = \sum_{k=1}^{K_n} w_{n,\omega,k} h_{n,k,t}, \quad (1.56)$$

unde  $K_n$  depinde de sursa  $n$  și este mai mic decât  $\Omega$  și  $T$ , iar  $w_{n,\omega,k}$  și  $h_{n,k,t}$  sunt nenegative. În plus, pentru a realiza o asociere între cele  $K$  componente NMF și cele  $N$  surse se va crea o nouă matrice nenegativă de dimensiuni  $N \times K$ , notată cu  $\mathbf{G} = [\mathbf{g}_{n,k}] \in \mathbb{R}_+^{N \times K}$  [20], iar varianțele surselor sunt reprezentate sub forma:

$$v_{n,\omega,t} = \sum_{k=1}^K w_{\omega,k} h_{k,t} \mathbf{g}_{n,k} \quad (1.57)$$

### 6.2.3 Combinarea modelului Gaussian cu cel spectral pentru NMF multicanal

Pentru a putea implementa NMF multicanal vom considera transformatele Fourier pe termen scurt ale canalelor, notate cu  $[x_{m,\omega,t}]$  și vom structura fiecare punct timp-frecvență într-un vector  $M$ -dimensional de forma  $\mathbf{x}_{\omega,t} = [x_{1,\omega,t}, \dots, x_{M,\omega,t}]^T \in \mathbb{C}^M$  pe baza (6.42). Vom determina:

$$\mathbf{X}_{\omega,t} = \mathbf{x}_{\omega,t} \mathbf{x}_{\omega,t}^H = \begin{bmatrix} |x_{1,\omega,t}|^2 & \cdots & x_{1,\omega,t} x_{M,\omega,t}^* \\ \vdots & \ddots & \vdots \\ x_{M,\omega,t} x_{1,\omega,t}^* & \cdots & |x_{M,\omega,t}|^2 \end{bmatrix} \quad (1.58)$$

Astfel vom putea aproxima produsele exterioare cu o structură de rang  $K$ :

$$\mathbf{X}_{\omega,t} \approx \sum_{k=1}^K \mathbf{G}_{\omega,k} w_{\omega,k} h_{k,t} = \hat{\mathbf{X}}_{\omega,t} \quad (1.59)$$

Similar cu prezentarea din secțiunea 6.1.3 vom defini funcția obiectiv prin logaritmare densității de probabilitate totale:

$$\mathcal{C}(\mathbf{W}, \mathbf{H}, \mathbf{G}) = -\log p(\mathbf{X} | \mathbf{W}, \mathbf{H}, \mathbf{G}) = \sum_{\omega=1}^{\Omega} \sum_{t=1}^T \left[ \mathbf{x}_{\omega,t}^H \hat{\mathbf{X}}_{\omega,t}^{-1} \mathbf{x}_{\omega,t} + \log \det \hat{\mathbf{X}}_{\omega,t} \right] \quad (1.60)$$

Ca urmare, modelul (6.49) și funcția obiectiv (6.52) devin [21]:

$$\hat{\mathbf{X}}_{\omega,t} = \sum_{k=1}^K \left( \sum_{n=1}^N z_{k,n} \mathbf{G}_{\omega,n} \right) w_{\omega,k} h_{k,t} \quad (1.61)$$

$$\mathcal{C}(\mathbf{W}, \mathbf{H}, \mathbf{G}, \mathbf{Z}) = \sum_{\omega=1}^{\Omega} \sum_{t=1}^T \left[ \mathbf{x}_{\omega,t}^H \hat{\mathbf{X}}_{\omega,t}^{-1} \mathbf{x}_{\omega,t} + \log \det \hat{\mathbf{X}}_{\omega,t} \right], \quad (1.62)$$

cu  $\mathbf{Z} = [z_{k,n}]$  și matricea multidimensională  $\mathbf{G} = \left\{ \mathbf{G}_{\omega,n} \right\}_{\substack{\omega=1..\\ \Omega \\ n=1..N}}$ .

### 6.2.4 Algoritmul NMF multicanal bazat pe minimizarea unei funcții de cost auxiliare

Definim funcția de cost auxiliară pentru NMF multicanal relativ la funcția de cost obiectiv (6.52) în felul următor:

$$\begin{aligned} \mathcal{C}^+(\mathbf{W}, \mathbf{H}, \mathbf{G}, \mathbf{R}, \mathbf{Q}) = & \sum_{\omega=1}^{\Omega} \sum_{t=1}^T \sum_{k=1}^K \frac{\mathbf{x}_{\omega,t}^H \mathbf{R}_{\omega,t,k} \mathbf{G}_{\omega,k}^{-1} \mathbf{R}_{\omega,t,k} \mathbf{x}_{\omega,t}}{w_{\omega,k} h_{k,t}} + \\ & + \sum_{\omega=1}^{\Omega} \sum_{t=1}^T \left[ \text{tr}(\hat{\mathbf{X}}_{\omega,t} \mathbf{Q}_{\omega,t}^{-1}) + \log \det \mathbf{Q}_{\omega,t} - M \right] \end{aligned} \quad (1.63)$$

Actualizările pentru variabilele funcției obiectiv se determină prin egalarea cu zero a derivatelor parțiale ale funcției auxiliare  $\mathcal{C}^+(\mathbf{W}, \mathbf{H}, \mathbf{G}, \mathbf{R}, \mathbf{Q})$  în funcție de  $\mathbf{W}$ ,  $\mathbf{H}$  și  $\mathbf{G}$ , obținându-se:

$$\begin{aligned}
w_{\omega,k} &\leftarrow w_{\omega,k} \sqrt{\frac{\sum_{t=1}^T h_{k,t} \mathbf{x}_{\omega,t}^H \hat{\mathbf{X}}_{\omega,t}^{-1} \mathbf{G}_{\omega,k} \hat{\mathbf{X}}_{\omega,t}^{-1} \mathbf{x}_{\omega,t}}{\sum_{t=1}^T h_{k,t} \text{tr}(\hat{\mathbf{X}}_{\omega,t}^{-1} \mathbf{G}_{\omega,k})}} \\
h_{k,t} &\leftarrow h_{k,t} \sqrt{\frac{\sum_{\omega=1}^{\Omega} w_{\omega,k} \mathbf{x}_{\omega,t}^H \hat{\mathbf{X}}_{\omega,t}^{-1} \mathbf{G}_{\omega,k} \hat{\mathbf{X}}_{\omega,t}^{-1} \mathbf{x}_{\omega,t}}{\sum_{\omega=1}^{\Omega} w_{\omega,k} \text{tr}(\hat{\mathbf{X}}_{\omega,t}^{-1} \mathbf{G}_{\omega,k})}} \\
\mathbf{G}_{\omega,k} &\leftarrow \mathbf{A}^{-1} \# (\mathbf{G}_{\omega,k} \mathbf{B} \mathbf{G}_{\omega,k})
\end{aligned} \tag{1.64}$$

În (6.60)  $\mathbf{A} = \sum_{t=1}^T h_{k,t} \hat{\mathbf{X}}_{\omega,t}^{-1}$  și  $\mathbf{B} = \sum_{t=1}^T h_{k,t} \hat{\mathbf{X}}_{\omega,t}^{-1} \mathbf{X}_{\omega,t} \hat{\mathbf{X}}_{\omega,t}^{-1}$  iar operatorul  $\mathbf{X} \# \mathbf{Y}$  este definit ca medie geometrică a două matrice pozitiv semidefinite:

$$\mathbf{A} \# \mathbf{B} = \mathbf{A} (\mathbf{A}^{-1} \mathbf{B})^{\frac{1}{2}} \tag{1.65}$$

Până în acest punct s-a realizat minimizarea funcției de cost (6.52). Pentru a beneficia și de gruparea componentelor determinate în surse corelate pe baza structurii spațiale vom continua cu optimizarea funcției de cost obiectiv (6.54) în urma căreia vom obține următoarele actualizări pentru parametri:

$$\begin{aligned}
w_{\omega,k} &\leftarrow w_{\omega,k} \sqrt{\frac{\sum_{l=1}^L z_{l,k} \sum_{t=1}^T h_{k,t} \text{tr}(\hat{\mathbf{X}}_{\omega,t}^{-1} \mathbf{X}_{\omega,t} \hat{\mathbf{X}}_{\omega,t}^{-1} \mathbf{G}_{\omega,l})}{\sum_{l=1}^L z_{l,k} \sum_{t=1}^T h_{k,t} \text{tr}(\hat{\mathbf{X}}_{\omega,t}^{-1} \mathbf{G}_{\omega,l})}} \\
h_{k,t} &\leftarrow h_{k,t} \sqrt{\frac{\sum_{l=1}^L z_{l,k} \sum_{\omega=1}^{\Omega} w_{\omega,k} \text{tr}(\hat{\mathbf{X}}_{\omega,t}^{-1} \mathbf{X}_{\omega,t} \hat{\mathbf{X}}_{\omega,t}^{-1} \mathbf{G}_{\omega,l})}{\sum_{l=1}^L z_{l,k} \sum_{\omega=1}^{\Omega} w_{\omega,k} \text{tr}(\hat{\mathbf{X}}_{\omega,t}^{-1} \mathbf{G}_{\omega,l})}} \\
z_{l,k} &\leftarrow z_{l,k} \sqrt{\frac{\sum_{\omega=1}^{\Omega} \sum_{t=1}^T w_{\omega,k} h_{k,t} \text{tr}(\hat{\mathbf{X}}_{\omega,t}^{-1} \mathbf{X}_{\omega,t} \hat{\mathbf{X}}_{\omega,t}^{-1} \mathbf{G}_{\omega,l})}{\sum_{\omega=1}^{\Omega} \sum_{t=1}^T w_{\omega,k} h_{k,t} \text{tr}(\hat{\mathbf{X}}_{\omega,t}^{-1} \mathbf{G}_{\omega,l})}} \\
\mathbf{G}_{\omega,l} &\leftarrow \mathbf{A}^{-1} (\mathbf{A} \mathbf{G}_{\omega,l} \mathbf{B} \mathbf{G}_{\omega,l})^{\frac{1}{2}}
\end{aligned} \tag{1.66}$$

### 6.2.5 Extragerea estimatelor surselor din rezultatele NMF multicanal

Aceasta se poate realiza folosind un filtru Wiener cu un singur canal pentru al  $m$ -lea canal din mixtura în felul următor:

$$\{\mathbf{y}_{\omega,t}^{(l)}\}_m = \frac{\{\mathbf{G}_{\omega,t}\}_{mm} \sum_{k=1}^K z_{l,k} w_{\omega,k} h_{k,t}}{\sum_{l=1}^L \{\mathbf{G}_{\omega,t}\}_{mm} \sum_{k=1}^K z_{l,k} w_{\omega,k} h_{k,t}} \{\mathbf{x}_{\omega,t}\}_m \quad (1.67)$$

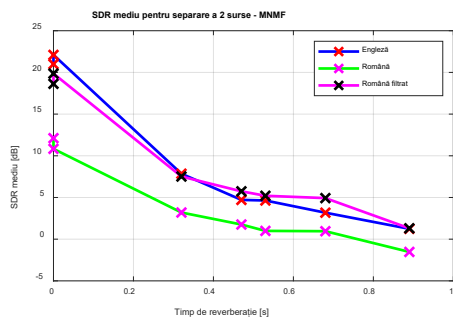
Alternativ se poate folosi un filtru Wiener multicanal de forma:

$$\mathbf{y}_{\omega,t}^{(l)} = \left( \sum_{k=1}^K z_{l,k} w_{\omega,k} h_{k,t} \right) \mathbf{G}_{\omega,t} \hat{\mathbf{X}}_{\omega,t}^{-1} \mathbf{x}_{\omega,t}, \quad (1.68)$$

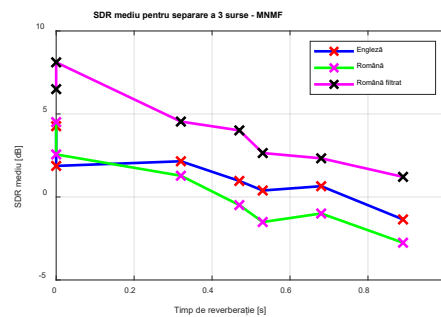
## 6.2.6 Rezultate experimentale NMF multicanal

Folosind aceleași mixturi ca în capitolele precedente s-a testat separarea a două surse din înregistrări binaurale. Numărul de baze NMF alese a fost de 20 pentru fiecare sursă în parte și a fost ales în urma testării mai multor valori.

În urma testării performanțelor NMF multicanal asupra celor 20 de mixturi a câte 2 surse de tip voce alese aleator obținem rezultatele din figura 6.13. Pentru a verifica și capacitatea de a separa 3 surse a fost refăcut testul și se observă în figura 6.14 că performanța este mult mai slabă față de rezultatele metodelor anterioare, în special în cazul mixturilor reverberante.



**Figura 6.2** Performanța de separare medie a două surse în funcție de timpul de reverberație – algoritmul NMF multicanal (MNMF)



**Figura 6.3** Performanța de separare medie a trei surse în funcție de timpul de reverberație – algoritmul NMF multicanal (MNMF)

## 6.3 Concluzii

În cazul celei cu un singur canal se observă o separare bună în cazul semnalelor armonice (cum ar fi cele generate de instrumente muzicale), cu o spectrogramă caracterizată de mixturi de distribuții super-Gaussiene, însă pentru semnale de tip voce, a căror conținut spectral este mai complex, performanțele sunt foarte scăzute.

Pentru cazul multicanal se observă o separare foarte bună în condiții anecoice, iar pentru camere reverberante performanța este foarte asemănătoare cu cele din capitolele anterioare.

# Capitolul 7

## Analiza matriceală independentă de rang scăzut

Analiza matriceală independentă de rang scăzut (*Independent Low-Rank Matrix Analysis - ILRMA*) se dovedește a fi o soluție la separarea mixturilor multicanal și ia naștere în urma combinării analizei vectorilor independenți (*Independent Vector Analysis - IVA*) și a factorizării matriceale nenegative.

### 7.1 Analiza vectorilor independenți – IVA

Analiza vectorilor independenți este o extindere a metodei de analiză a componentelor independente (*Independent Component Analysis - ICA*) și se aplică pentru mixturi determinate unde numărul microfoanelor este egal cu numărul surselor din mixtură, în cazul nostru  $N = M$  [22].

Metoda de separare folosind ICA presupune determinarea matricei pătratică  $\mathbf{W}_\omega$  de dimensiune  $M$  ce transformă liniar mixturile  $\mathbf{x}_{\omega,t}$  în estimate ale surselor originale  $\mathbf{y}_{\omega,t} = [y_{1,\omega,t}, \dots, y_{N,\omega,t}]^T \in \mathbb{C}^N$  în felul următor:

$$\mathbf{y}_{\omega,t} = \mathbf{W}_\omega \mathbf{x}_{\omega,t} \quad (1.69)$$

În cazul IVA noțiunea de independență este extinsă la variabile de tip vector. Acești vectori sunt definiți peste toate frecvențele corespunzătoare unui moment de timp sub forma  $\mathbf{y}_{n,t} = [y_{n,1,t}, \dots, y_{n,\Omega,t}]^T$ .

### 7.2 Extinderea IVA cu NMF pentru a obține ILRMA

O primă metodă de a implementa ILRMA este extinderea analizei vectorilor independenți prin introducerea factorizării matriceale nenegative [23].

Folosind notațiile pentru matricele de factorizare  $\mathbf{W}_n = [w_{n,\omega,k}]$  și  $\mathbf{H}_n = [h_{n,k,t}]$  putem defini funcția obiectiv:



$$\mathcal{C}(\mathcal{T}, \{\mathbf{W}_n\}, \{\mathbf{H}_n\}) = \sum_{n=1}^N \sum_{\omega=1}^{\Omega} \sum_{t=1}^T \left[ \frac{|y_{n,\omega,t}|^2}{\hat{y}_{n,\omega,t}} + \log(\hat{y}_{n,\omega,t}) \right] - 2T \sum_{\omega=1}^{\Omega} \log |\det(\mathbf{T}_{\omega})| \quad (1.70)$$

Minimizarea funcției obiectiv (7.23) se poate realiza prin alternarea actualizărilor NMF similare cu (6.40) și (6.41) cu rezolvarea problemei HEAD [24]. Întâi ne vom concentra pe primul termen din (7.23) și vom determina actualizările pentru  $\mathbf{W}_n = [w_{n,\omega,k}]$  și  $\mathbf{H}_n = [h_{n,k,t}]$ . Vom observa că pentru fiecare  $n$  în parte putem obține actualizările prin înlocuirea  $|x_{\omega,t}|^2$  și  $\hat{x}_{\omega,t}$  cu  $|y_{n,\omega,t}|^2$ , respectiv  $\hat{y}_{n,\omega,t}$  în (6.40) și (6.41), obținând actualizările:

$$w_{n,\omega,k} \leftarrow w_{n,\omega,k} \sqrt{\frac{\sum_{t=1}^T h_{n,k,t} |y_{n,\omega,t}|^2}{\hat{y}_{n,\omega,t}^2}} \quad (1.71)$$

$$h_{n,k,t} \leftarrow h_{n,k,t} \sqrt{\frac{\sum_{\omega=1}^{\Omega} w_{n,\omega,k} |y_{n,\omega,t}|^2}{\hat{y}_{n,\omega,t}^2}} \quad (1.72)$$

### 7.3 Restricționarea NMF multicanal pentru a obține ILRMA

O a doua metodă de a obține metoda ILRMA este prin a restricționa la NMF multicanal proprietatea spațială definită prin matricea  $\mathbf{G}_{n,\omega} \in \mathbb{C}^{M \times M}$  să fie de rang 1. Pornind apoi de la modelul MNMF (6.53), îl putem simplifica astfel încât:

$$\hat{\mathbf{X}}_{\omega,t} = \mathbf{G}_{\omega} \mathbf{D}_{\omega,t} \mathbf{G}_{\omega}^H, \quad (1.73)$$

unde  $\mathbf{G}_{\omega} = [\mathbf{g}_{1,\omega}, \dots, \mathbf{g}_{N,\omega}]$  și  $\mathbf{D}_{\omega,t}$  este o matrice pătratică diagonală de dimensiuni  $N \times N$  al cărei al  $n$ -lea element de pe diagonală are valoarea:

$$\hat{y}_{n,\omega,t} = \sum_{k=1}^K z_{k,n} w_{\omega,k} h_{k,t} \quad (1.74)$$

Mai departe se va restricționa sistemul de mixare la unul determinat prin  $N = M$  pentru a obține matricea de separare  $\mathbf{T}_{\omega}$  din matricea de mixare  $\mathbf{G}_{\omega}$ :

$$\mathcal{C}(\mathcal{T}, \mathbf{W}, \mathbf{H}, \mathbf{Z}) = \sum_{\omega=1}^{\Omega} \sum_{t=1}^T \sum_{n=1}^N \left[ \frac{|y_{n,\omega,t}|^2}{\hat{y}_{n,\omega,t}} + \log(\hat{y}_{n,\omega,t}) \right] - 2T \sum_{\omega=1}^{\Omega} \log |\det(\mathbf{T}_{\omega})| \quad (1.75)$$

În urma minimizării funcției obiectiv fără a utiliza funcții auxiliare vom obține actualizarea pentru variabila de partiționare a surselor  $z_{k,n}$ :

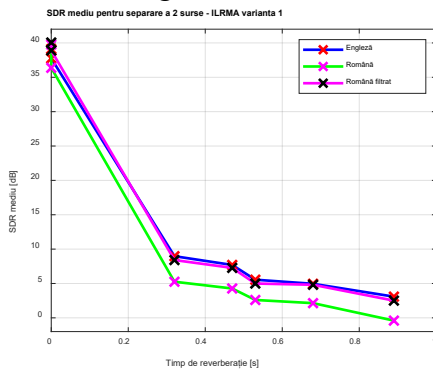
$$z_{k,n} \leftarrow z_{k,n} \sqrt{\frac{\sum_{\omega=1}^{\Omega} \sum_{t=1}^T \frac{w_{\omega,k} h_{k,t} |y_{n,\omega,t}|^2}{\hat{y}_{n,\omega,t}^2}}{\sum_{\omega=1}^{\Omega} \sum_{t=1}^T \frac{w_{\omega,k} h_{k,t}}{\hat{y}_{n,\omega,t}}}} \quad (1.76)$$

$$w_{\omega,k} \leftarrow w_{\omega,k} \sqrt{\frac{\sum_{n=1}^N \sum_{t=1}^T \frac{z_{k,n} h_{k,t} |y_{n,\omega,t}|^2}{\hat{y}_{n,\omega,t}^2}}{\sum_{n=1}^N \sum_{t=1}^T \frac{z_{k,n} h_{k,t}}{\hat{y}_{n,\omega,t}}}} \quad (1.77)$$

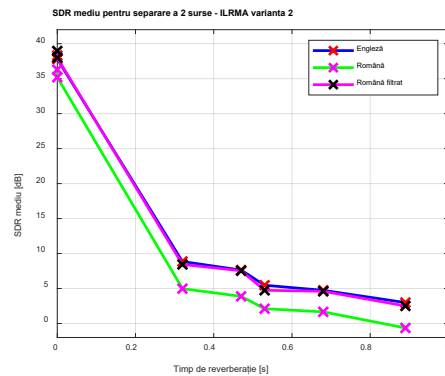
$$h_{k,t} \leftarrow h_{k,t} \sqrt{\frac{\sum_{\omega=1}^{\Omega} \sum_{n=1}^N \frac{w_{\omega,k} z_{k,n} |y_{n,\omega,t}|^2}{\hat{y}_{n,\omega,t}^2}}{\sum_{\omega=1}^{\Omega} \sum_{n=1}^N \frac{w_{\omega,k} z_{k,n}}{\hat{y}_{n,\omega,t}}}} \quad (1.78)$$

## 7.4 Rezultate experimentale

Asemănător cu testele din capitolele precedente vom testa performanța de separare folosind cele 20 de mixturi în limba engleză și cele 10 mixturi în limba română a câte două propoziții rostite. Rezultatele pentru algoritmul ILRMA fără funcția de partiționare se pot observa în figura 7.4, pentru cea cu funcție de partiționare le putem vizualiza în figura 7.5.



**Figura 7.1** Performanța de separare medie a două surse în funcție de timpul de reverberație – algoritmul ILRMA varianta 1



**Figura 7.2** Performanța de separare medie a două surse în funcție de timpul de reverberație – algoritmul ILRMA varianta 2

## 7.5 Concluzii

Performanțele de separare ale metodei sunt foarte bune pentru mixturi în condiții anecoice și moderate pentru mixturi reverberante. Cu toate acestea o analiză a timpului de execuție pentru metodă va permite luarea unei decizii în privința performanței totale a metodei.

# Capitolul 8

## Concluzii

### 8.1 Rezultate obținute

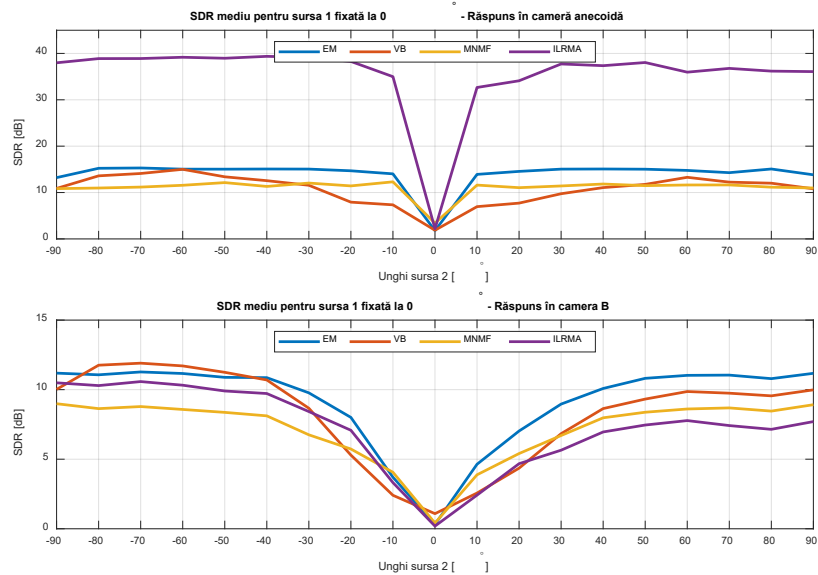
În lucrarea de față au fost prezentate patru metode de separare a surselor sonore de tip voce din mixturi binaurale reverberante.

Semnalele audio de tip voce feminină sau masculină utilizate au fost extrase din setul de date TIMIT [13], ce cuprinde propoziții rostite limba engleză. Setul de voci feminine și masculine în limba română au fost reprezentate de 7 propoziții înregistrate în condiții de studio. Toate semnalele audio au fost reeșantionate la frecvența de eșantionare  $F_s = 16$  kHz. Compunerea mixturilor prin convoluție s-a realizat cu un set de înregistrări binaurale realizate de Hummersone [14] cu un cap binaural plasat în cameră anecoidă și într-un set de încăperi tipice cu timpi de reverberație diferiți. Am notat răspunsul la impuls al capului binaural cu termenul de "Anecoidă" pentru a putea face distincție între acesta și alte răspunsuri la impuls ale capului. Camerele au fost notate cu A, B, C și D în ordine crescătoare a timpului de reverberație. La acest set de răspunsuri la impuls au mai fost adăugate răspunsurile la impuls ale capului personalizate (ale unei persoane reale) măsurate și extrase cu metoda descrisă în capitolul 2 și în [25] și setul de răspunsuri binaurale asociat încăperii în care s-au măsurat. Răspunsul la impuls al capului a fost notat în rezultate cu HRIR, iar camera respectivă fost notată cu E. Proprietățile acustice ale camerelor, împreună cu notațiile utilizate pentru fiecare se regăsesc în tabelul 8.1.

**Tabelul 8.1** Proprietățile acustice ale încăperilor exprimate prin întârzierea inițială (Initial Time Delay Gap – ITDG), raportul dintre unda directă și reverberație (Direct-to-Reverberant Ratio – DRR) și timpul de reverberație  $T_{60}$

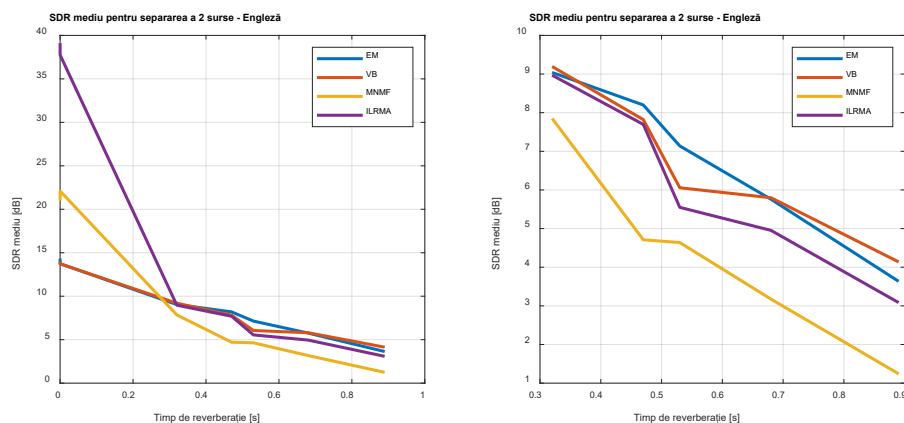
	ITDG [ms]	DRR [dB]	$T_{60}$ [s]
Anecoidă	2	-	0
HRIR	2	-	0
Camera A	8,72	6,09	0,32
Camera B	9,66	5,31	0,47
Camera C	11,9	8,82	0,68
Camera D	21,6	6,12	0,89
Camera E	9,32	5,78	0,53

Pentru că toate metodele prezentate se bazează pe o separare spațială a surselor sonore în mixtura finală am analizat în primă fază performanța de separare în funcție de unghiul dintre două surse. Performanța a fost evaluată prin determinarea raportului sursă distorsiune (SDR) deoarece acesta înglobează toate artefactele, zgomotele și interferențele ce pot apărea în procesul de separare. După aplicarea metodelor o parte dintre rezultate au fost reprezentate în figura 8.1.

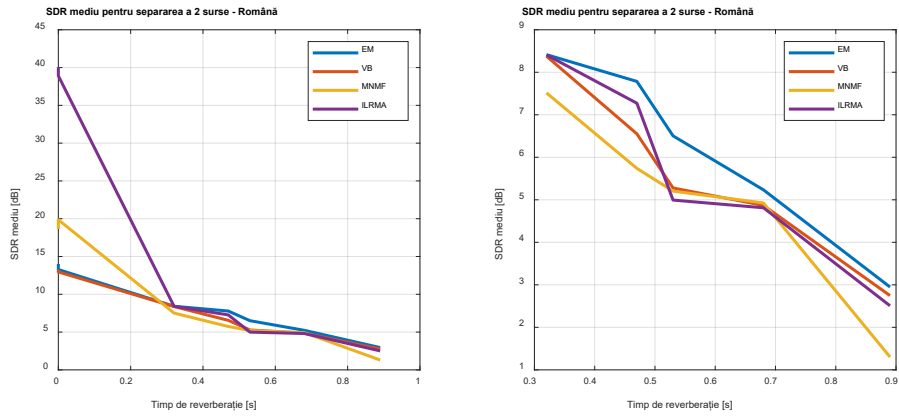


**Figura 8.1** Performanța de separare în funcție de unghiul dintre surse în SDR pentru sursa fixată; Sus: în camera anecoidă; Jos: în camera B

Se observă că în condiții anecoice unghiul minim dintre surse poate fi de  $10^{\circ} - 20^{\circ}$ , iar în condiții reverberante unghiul minim ar trebui să fie de cel puțin  $30^{\circ}$ . Următoarea analiză a fost realizată asupra unei mixturi de două surse sonore plasate la unghiurile de  $-30^{\circ}$  și  $30^{\circ}$  în cele 7 tipuri de încăperi. Rezultatele pentru separarea a două surse în limba engleză se pot observa în figura 8.2, iar pentru limba română în figura 8.3. Se observa că metodele MNMF și ILRMA au o performanță foarte bună de separare în cazul mixturilor determinate nereverberante. În condiții reverberante metoda EM este cea mai performantă, iar cea VB urmează la egalitate cu ILRMA.

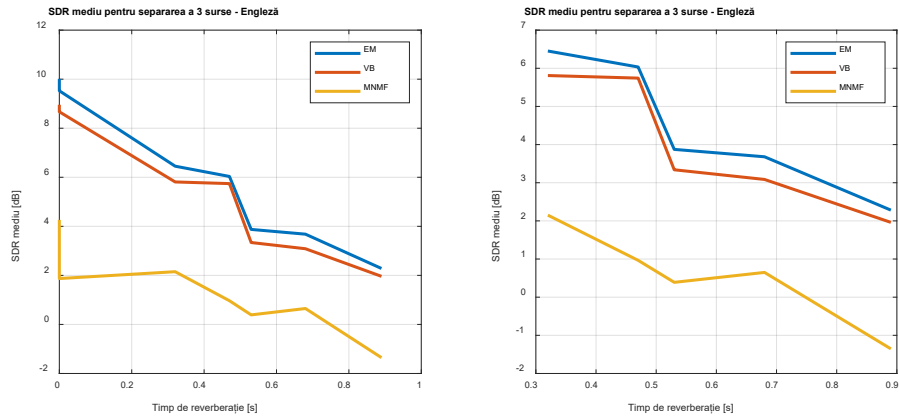


**Figura 8.2** Performanța de separare pentru mixturi de două surse în limba engleză. SDR [dB] mediu peste sursele recuperate și peste mixturi; Stânga: pentru toate încăperile; Dreapta: Pentru încăperile reverberante

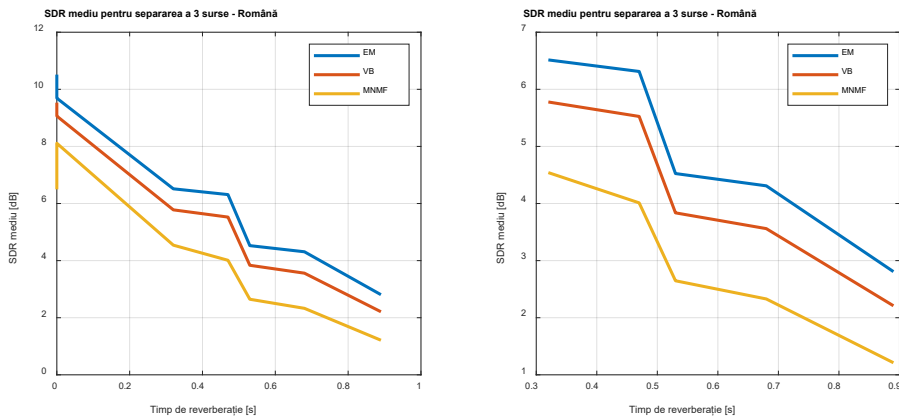


**Figura 8.3** Performanța de separare pentru mixturi de două surse în limba română. SDR [dB] mediu peste sursele recuperate și peste mixturi; Stânga: pentru toate încăperile; Dreapta: Pentru încăperile reverberante

Rezultatele pentru mixturi de trei surse se pot observa în figura 8.4 pentru limba engleză și în figura 8.5 pentru limba română.



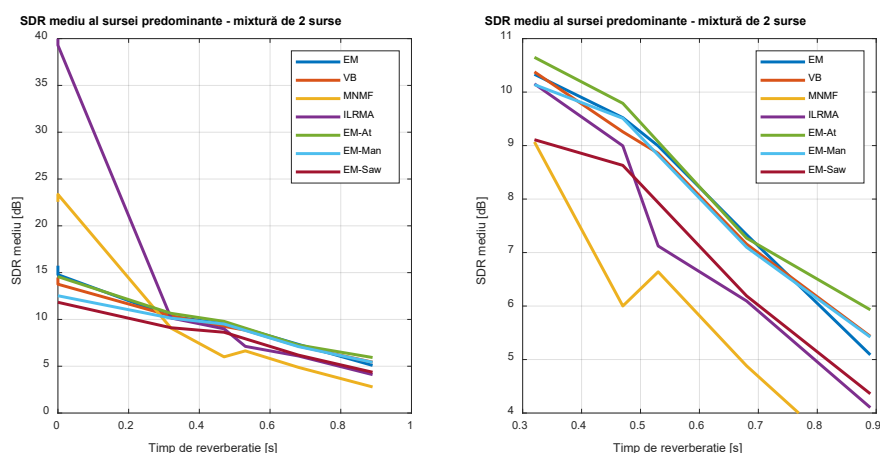
**Figura 8.4** Performanța de separare pentru mixturi de trei surse în limba engleză. SDR [dB] mediu peste sursele recuperate și peste mixturi; Stânga: pentru toate încăperile; Dreapta: Pentru încăperile reverberante



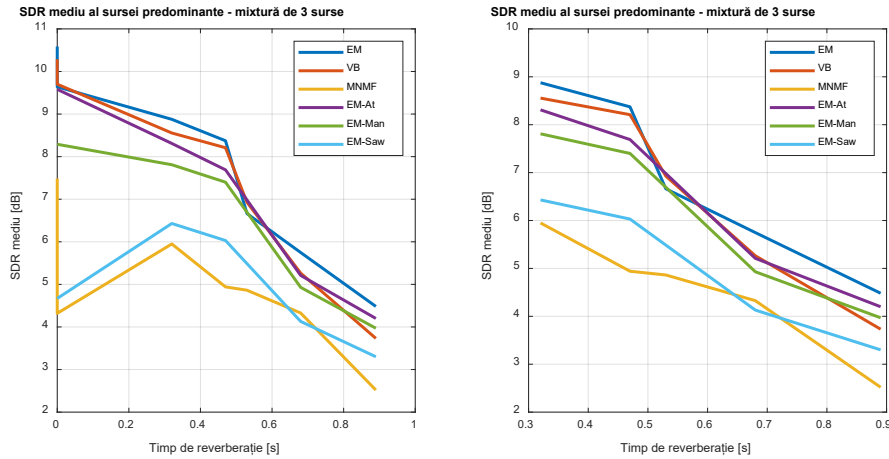
**Figura 8.5** Performanța de separare pentru mixturi de trei surse în limba română. SDR [dB] mediu peste sursele recuperate și peste mixturi; Stânga: pentru toate încăperile; Dreapta: Pentru încăperile reverberante

În urma analizei separării surselor sonore din mixturi subdeterminate observăm că metoda MNMF are performanțe mai scăzute față de metodele EM și VB, diferențele de SDR fiind între 2 și 5 dB. O concluzie ar fi că metoda MNMF nu funcționează corespunzător pentru mixturi subdeterminate, fiind mult mai performantă pentru cele determinate. Ca și comparație între metodele EM și VB, modelarea statistică între cele două este foarte similară, precum și inițializarea, diferența fiind că metoda VB folosește distribuții a priori pentru parametrii distribuțiilor Gaussiene. Observăm o diferență de performanță sistematică între cele două de aproximativ 0,5 dB, cea mai performantă fiind metoda EM. Cu toate acestea avantajul metodei VB este că nu necesită cunoașterea numărului de surse sonore, acesta fiind determinat de către algoritm.

În continuare vom realiza o comparație cu rezultate similare din literatură. O serie de rezultate se regăsesc în [10] și îmbunătățite în [26], realizate de Atiyeh Alinaghi folosind o metodă similară cu cea de maximizare a mediei statistice prezentată în această lucrare, cu diferența că în această teză se folosește o inițializare a responsabilităților incluzând și rezultatele descompunerii adaptive a valorilor proprii, iar metoda implementată de Atiyeh folosește o sursă suplimentară pentru a prelua doar reverberația. Rezultatele acestea au fost notate pe graficul mixturilor în limba engleză cu EM-At. Modul de testare a fost schimbat deoarece în articolele menționate valoarea SDR determinată se referă la o sursă țintă, ce am presupus că este sursa cu cel mai mare SDR, celelalte fiind considerate semnale de tip interferență. În mod asemănător au fost incluse rezultatele metodei dezvoltate de Mandel în [8] bazată tot pe maximizarea mediei statistice folosind un model similar cu al lui Atiyeh, însă doar cu indicii binaurali, și rezultatele lui Sawada pe baza metodei descrise în [27], bazată tot pe maximizarea mediei statistice, însă pentru vectorii de mixare. Metoda lui Madel a fost notată pe grafice cu EM-Man, iar ce a lui Sawada cu EM-Saw. Performanța de separare pentru mixturi de două surse se poate observa în figura 8.6, iar performanța de separare pentru mixturi de 3 surse în figura 8.7.



**Figura 8.6** Performanța de separare pentru mixturi de două surse în limba engleză. SDR [dB] mediu pentru sursa predominantă; Stânga: pentru toate încăperile; Dreapta: Pentru încăperile reverberante



**Figura 8.7** Performanța de separare pentru mixturi de două surse în limba engleză. SDR [dB] mediu pentru sursa predominantă; Stânga: pentru toate încăperile; Dreapta: Pentru încăperile reverberante

Din rezultatele comparative prezentate observăm că metoda EM implementată în această lucrare, metoda VB, metoda EM-At și EM-Man au rezultate similare în ce privește performanța de separare atât pentru mixturi determinate cât și pentru cele subdeterminate. Dintre acestea metoda EM pare să aibă rezultate mai bune în cazul subdeterminat față de celelalte. Trebuie să facem distincția între cazul anecoid și cel reverberant deoarece metodele ILRMA și MNMF au rezultate foarte bune, mult peste celelalte, în cazul anecoid determinat, însă performanța în medii reverberante este mediocră, sub celelalte metode. Metoda EM-Saw pare să aibă rezultate mai slabe decât celelalte metode, fiind peste MNMF în cazul determinat și subdeterminat reverberant. Privind rezultatele medii pentru încăperile reverberante observăm că metodele EM și VB au performanțe medii foarte apropiate, diferența fiind de  $\pm 0,2$  dB, cu VB fiind marginal mai performant în cazul subdeterminat. Diferențele față de rezultatele cele mai performante din literatură sunt de aproximativ 0,3 dB, cele din lucrare fiind mai performante.

Pentru a oferi un grad suplimentar de comparație între metodele implementate în această lucrare am structurat în tabelul 8.3 timpii medii de execuție pentru fiecare metodă în parte în urma separării unei mixturi de două sau trei surse.

**Tabelul 8.2** Timp de execuție medii ai fiecărei metode pentru separarea unei mixturi

Metoda	Nr. Surse	Iterații metodă	Timp rulare [s]
<b>EM</b>	2	20	9,95
<b>EM</b>	3	20	13,39
<b>VB</b>	2	40	22,44
<b>VB</b>	3	40	34,06
<b>MNMF</b>	2	200	46,18
<b>MNMF</b>	3	200	64,49
<b>ILRMA v1</b>	2	50	3,82
<b>ILRMA v2</b>	2	50	4,36

În urma analizei rezultatelor obținute putem stabili o serie de criterii în alegerea metodei potrivite în funcție de mixtura binaurală pe care se aplică:

- Pentru mixturi anecoice, sau cu timpi de reverberație mici în cazul determinat cea mai bună metodă este ILRMA, având atât rezultate foarte bune cât și un timp de execuție redus;
- Pentru mixturi determinate sau subdeterminate reverberante în care este cunoscut numărul de surse se poate aplica atât EM cât și VB, însă metoda EM este mai rapidă și, în anumite situații mai performantă;
- Pentru mixturi supradeterminate, determinate sau subdeterminate în care nu este cunoscut numărul de surse se poate aplica metoda VB ce are rezultate bune de separare și nu necesită cunoașterea numărului de surse din mixtură;
- Metodele funcționează similar indiferent de limba în care sunt rostite frazele, o influență mai mare având-o conținutul spectral de joasă frecvență în procesul de separare.

## 8.2 Contribuții originale

Pentru a realiza măsurătorile din capitolul 2 am creat o bibliotecă software, publicată în articolul de conferință [C1], ce implementează metoda de determinare a răspunsului acustic la impuls folosind semnalul de test sinusoidal cu varierea exponențială în frecvență ce prezintă o serie de avantaje în ceea ce privește separarea răspunsului liniar de armonici de ordin superior. Totodată am prezentat realizarea de răspunsuri la impuls acustice sincronizate. Acest software a fost utilizat pentru a determina parametrii acustici ai încăperii în care s-au realizat o parte din teste.

În articolul de conferință [C2] am dezvoltat o metodă de reducere a timpului de măsură pentru un răspuns la impuls acustic folosind metoda ESS prin secționarea unui răspuns la impuls și suprapunerea în timp a celor două părți ce a eficientizat măsurătorile pentru răspunsul la impuls al capului din capitolul 2.

În articolul de revistă [R1] am realizat o metodă de egalizare a răspunsului la impuls pentru un difuzor și sunt analizate performanțele acesteia. Egalizarea a fost utilizată pentru difuzorul folosit la măsurarea răspunsurilor la impuls ale capului în cadrul capitolului 2.

În articolul de conferință [C7] am dezvoltat o metodă de extragere a răspunsului la impuls al capului din răspunsuri la impuls acustice binaurale realizate în camere reverberante, avantajul major fiind eliminarea necesității unei camere anecoide cu păstrarea unei calități bune ale răspunsurilor măsurate. Această metodă a fost prezentată în capitolul 2.

În capitolul 3 am cercetat o metodă de determinare a direcției de sosire a surselor sonore folosind descompunerea adaptivă a valorilor proprii și am analizat experimental performanțele de detecție a acesteia. În plus, prin realizarea unei histograme pentru unghiurile detectate într-o mixtură binaurală am dovedit că metoda poate detecta unghiuri multiple cât timp mixturile de semnale de tip voce respectă condiția de raritate.



În capitolul 4 am prezentat și implementat o metodă de separare a surselor sonore din mixturi binaurale reverberante determinate și subdeterminate din literatură, bazată pe maximizarea mediei statistice a unei mixturi de Gaussiene ce modelează atât vectorii de mixare cât și indicii binaurali de localizare. Un plus de performanță al acestei metode a fost adus prin combinarea inițializării bazate pe histograma PHAT cu metoda de determinare a direcției de sosire a surselor sonore bazată pe descompunerea adaptivă a valorilor proprii prezentată în capitolul 3.

În capitolul 5 am conceput, dezvoltat și implementat o nouă metodă de separare a surselor sonore din mixturi binaurale subdeterminate bazată pe un algoritm Bayes variațional în care sunt, în primă fază, modelate observațiile vectorilor de mixare și indicilor binaurali ca o mixtură de distribuții Gaussiene. În faza a doua se asignează distribuții a priori mediilor și varianțelor acestor distribuții Gaussiene creând astfel modelul variațional. Am determinat și implementat apoi regulile de actualizare a hiperparametrilor pentru a atinge convergența. Modelul precum și rezultatele de separare obținute cu această metodă au fost publicate în articolul de conferință [C3].

În capitolul 6 am dezvoltat și implementat metoda cunoscută de separare a surselor sonore din mixturi binaurale prin factorizarea matriceală nenegativă multicanal ce se bazează pe descompunerea matricei de covarianță a transformatelor Fourier pe termen scurt a observațiilor în matrice de dimensiuni mai mici ce descriu activitățile în timp și frecvență a unor componente de bază.

În capitolul 7 am prezentat și implementat analiza matriceală independentă de rang scăzut ce este o metodă de separare oarbă a mixturilor determinate, bazată fie pe combinarea analizei vectorilor independenți cu factorizarea matriceală nenegativă, fie pe reducerea la rang 1 a matricei ce descrie proprietatea spațială în cadrul metodei de factorizare matriceală nenegativă multicanal.

Toate metodele au fost implementate și a fost realizată o comparație relativ la performanțele de separare a mixturilor binaurale determinate sau subdeterminate în condiții reverberante sau anecoice.

### 8.3 Lista lucrărilor originale

*Cărți:*

- [B1] **V. Popa**, C. Negrescu, "Măsurarea și caracterizarea sistemelor acustice : îndrumar de laborator", Editura Politehnica Press, București, 2013, ISBN 978-606-515-505-3
- [B2] C. Negrescu, A. Ciobanu, **V. Popa**, „Inginerie audio – Îndrumar de laborator”, Editura Politehnica Press, București, 2013, ISBN 978-606-515-527-5
- [B3] R. M. Udrea, C. A. Cojocariu, **V. Popa**, "Servere de conținut, procesoare de flux și terminale multimedia : îndrumar de laborator", Editura Politehnica Press, București, 2013, ISBN 978-606-515-504-6.

*Articole publicate în reviste științifice:*

- [R1] T. M. Culda, **V. Popa**, C. Cojocariu, D. Stanomir, C. Negrescu, "The Influence of Loudspeaker Performance in Loudspeaker Equalization Using Wiener Approach", U.P.B. Sci. Bull., Series C, Vol. 74, Iss. 4, P. 219-228, 2012, ISSN 1454-234x

*Articole de conferință:*

- [C1] **V. Popa**, T. M. Culda, C. Negrescu, "Software Library for Managing a Complete Audio Chain Intended for Determining Acoustic Room Parameters and Spatial Impression", The Annual Symposium Of The Institute Of Solid Mechanics (SISOM), Mai 2012 (**BDI**: IndexCopernicus, MathSciNet, Reaserchgate)
- [C2] T. M. Culda, **V. Popa**, D. Stanomir, C. Negrescu, "Reducing Time in Acoustic Impulse Response Measurements Using Exponential Sine Sweeps", International Symposium On Signals, Circuits And Systems (ISSCS), P. 1-4, Iași, Iulie 2013 (**ISI, IEEE**)
- [C3] **V. Popa**, W. Wang, A. Alinaghi, "Underdetermined Model-Based Blind Source Separation of Reverberant Speech Mixtures Using Spatial Cues in a Variational Bayesian Framework", Intelligent Signal Processing (ISP), Londra, Decembrie 2013, DOI: 10.1049/Cp.2013.2074 (**IEEE, Scopus**)
- [C4] A. Ciobanu, V. A. Niță, **V. Popa**, "Forgery Detection Based on Reverberation Time Estimation in Multiple Bands", 13<sup>th</sup> International Symposium on Electronics and Telecommunications (ISETC), Timișoara, Noiembrie 2018, WOS:000463031500068 (**ISI, IEEE**)
- [C5] V. A. Niță, **V. Popa**, "A Framework for Privacy Assurance in a Public Video-Surveillance System", 2019, International Symposium on Signals, Circuits and Systems (ISSCS), Iasi, Romania, 2019, Pp. 1-4, WOS:000503459500066 (**ISI, IEEE**)
- [C6] V. A. Niță, **V. Popa**, "Bringing Technology into Fencing Training. the Art of Counterattacking", 2019 11th International Symposium on Advanced Topics in Electrical Engineering, ATEE 2019, WOS:000475904500146 (**ISI, IEEE**)
- [C7] **V. Popa**, B. Moroșanu, C. Negrescu, "Head related transfer function measurement in reverberant rooms", 11th edition of the International Conference "Advanced Topics in Optoelectronics, Microelectronics and Nanotechnologies" (ATOM-N), Constanța, România, 2022, în curs de indexare WOS
- [C8] B. Moroșanu, **V. Popa**, C. Negrescu, Ionuț Fîciu, "Control Room Design for Subjective Audio Critical Listening", 11th edition of the International Conference "Advanced Topics in Optoelectronics, Microelectronics and Nanotechnologies" (ATOM-N), Constanța, România, 2022, în curs de indexare WOS

*Rapoarte de doctorat:*

- [RC1] **V. Popa**, “Determinarea direcției de sosire a unei surse sonore într-un mediu reverberant folosind descompunerea adaptivă a valorii proprii” - Primul raport de cercetare științifică, coordonator științific Prof.dr. ing. Ion Marghescu.
- [RC2] **V. Popa**, “Pregătirea și evaluarea mixturilor de semnale audio folosind transformata Fourier pe termen scurt. Evaluarea performanțelor de separare” – Al doilea raport de cercetare științifică, coordonator științific Prof.dr. ing. Ion Marghescu.
- [RC3] **V. Popa**, “Algoritmul maximizării mediei statistice pentru separarea surselor sonore din mixturi subdeterminate” – Al treilea raport de cercetare științifică, coordonator științific Prof.dr. ing. Ion Marghescu.
- [RC4] **V. Popa**, “Algoritmul Bayes variațional pentru separarea mixturilor binaurale” – Al patrulea raport de cercetare științifică, coordonator științific Prof.dr. ing. Ion Marghescu.
- [RC5] **V. Popa**, “Tehnici de separare oarbă a surselor sonore folosind factorizarea matriceală nenegativă” – Al cincilea raport de cercetare științifică, coordonator științific Prof.dr. ing. Ion Marghescu.

## **8.4 Perspective de dezvoltare ulterioară**

O primă dezvoltare ulterioară ar putea fi extinderea modelului Bayes variațional la mixturi subdeterminate reverberante cu multiple microfoane plasate în încăperi. Modelarea ar putea fi realizată prin înlocuirea indicilor binaurali cu un set matrice ce exprimă caracteristicile spațiale ca diferență de nivel cât și de fază pentru observațiile respective. Având în vedere modul de separare al metodei s-ar putea investiga și posibilitatea utilizării acesteia ca metodă de reducere a reverberației semnalelor înregistrate pentru a îmbunătăți inteligibilitatea acestora.

Pornind tot de la metoda variațională s-ar putea investiga utilizarea acesteia în acustica încăperilor, prin extragerea parametrilor acustici de bază folosind fie înregistrări de instrumente sau voci în încăperile respective, fie utilizând răspunsurile acustice la impuls măsurate cu multiple microfoane. Determinarea acestor parametri acustici ar fi utilă în mod special pentru reprezentații live deoarece ar permite egalizarea răspunsului sistemului de reproducere audio în timp real în funcție de condițiile întâlnite în locația respectivă.

De asemenea, s-ar putea investiga și utilizarea celorlalte metode în domeniul acusticii pentru a dezvolta algoritmi robusți de măsurare și interpretare a caracteristicilor acustice ale încăperilor.

O altă direcție de dezvoltare s-ar putea axa pe combinarea metodelor prezentate cu rețele neurale adânci, fie ca parte de preprocesare pentru acestea fie ca măsuri de întărire sau de optimizare a procesului de învățare.

# Bibliografie

- [1] A. Farina, “Advancements in Impulse Response Measurements by Sine Sweeps,” *Journal of The Audio Engineering Society*, 2007.
- [2] V. Popa, T. M. Culda, and C. Negrescu, “Software Library for Managing a Complete Audio Chain Intended for Determining Acoustic Room Parameters and Spatial Impression,” presented at the The Annual Symposium Of The Institute Of Solid Mechanics (SISOM), Bucuresti, 2012.
- [3] O. Kirkeby and P. Nelson, “Digital Filter Design for Inversion Problems in Sound Reproduction,” *Journal of the Audio Engineering Society*, vol. 47, Jan. 1999.
- [4] T. M. Culda, V. Popa, C. Cojocariu, D. Stanomir, and C. Negrescu, “The influence of loudspeaker performance in loudspeaker equalization using wiener approach,” *UPB Scientific Bulletin, Series C: Electrical Engineering*, vol. 74, pp. 219–228, Jan. 2012.
- [5] T. M. Culda, V. Popa, D. Stanomir, and C. Negrescu, “Reducing time in acoustic impulse response measurements using exponential sine sweeps,” in *International Symposium on Signals, Circuits and Systems ISSCS2013*, 2013, pp. 1–4.
- [6] E. Vincent, R. Gribonval, and C. Fevotte, “Performance measurement in blind audio source separation,” *IEEE Transactions on Audio, Speech and Language Processing*, vol. 14, no. 4, pp. 1462–1469, Jul. 2006.
- [7] Y. Huang, J. Benesty, and G. W. Elko, “Adaptive eigenvalue decomposition algorithm for real time acoustic source localization system,” 1999, pp. 937–940 vol.2.
- [8] M. I. Mandel, R. J. Weiss, and D. P. W. Ellis, “Model-Based Expectation-Maximization Source Separation and Localization,” *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 18, no. 2, pp. 382–394, 2010.
- [9] H. Sawada, S. Araki, and S. Makino, “A Two-Stage Frequency-Domain Blind Source Separation Method for Underdetermined Convolutional Mixtures,” in *2007 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, 2007, pp. 139–142.
- [10] A. Alinaghi, W. Wang, and P. J. Jackson, “Integrating binaural cues and blind source separation method for separating reverberant speech mixtures,” in *2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2011, pp. 209–212.
- [11] S. Araki, H. Sawada, R. Mukai, and S. Makino, “Underdetermined blind sparse source separation for arbitrarily arranged multiple sensors,” *Signal Processing*, vol. 87, no. 8, pp. 1833–1847, 2007.
- [12] V. Popa, W. Wang, and A. Alinaghi, “Underdetermined model-based blind source separation of reverberant speech mixtures using spatial cues in a variational Bayesian framework,” in *IET Intelligent Signal Processing Conference 2013 (ISP 2013)*, 2013, pp. 1–6.
- [13] J. S. Garofolo, L. F. Lamel, W. M. Fisher, J. G. Fiscus, D. S. Pallett, and N. L. Dahlgren, “DARPA TIMIT Acoustic Phonetic Continuous Speech Corpus CDROM.” NIST, 1993.

- [14] C. Hummersone, “A psychoacoustic engineering approach to machine sound source separation in reverberant environments,” Ph.D. thesis, Music and Sound Recording, University of Surrey, UK, 2011.
- [15] J. Taghia, N. Mohammadiha, and A. Leijon, “A variational Bayes approach to the underdetermined blind source separation with automatic determination of the number of sources,” in *2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2012, pp. 253–256.
- [16] C. Févotte, N. Bertin, and J.-L. Durrieu, “Nonnegative Matrix Factorization with the Itakura-Saito Divergence: With Application to Music Analysis,” *Neural computation*, vol. 21, pp. 793–830, Oct. 2008.
- [17] K. Lange, D. R. Hunter, and I. Yang, “Optimization Transfer Using Surrogate Objective Functions,” *Journal of Computational and Graphical Statistics*, vol. 9, no. 1, pp. 1–20, 2000.
- [18] H. Sawada, N. Ono, H. Kameoka, D. Kitamura, and H. Saruwatari, “A review of blind source separation methods: two converging routes to ILRMA originating from ICA and NMF,” *APSIPA Transactions on Signal and Information Processing*, vol. 8, p. e12, 2019.
- [19] H. Sawada, H. Kameoka, S. Araki, and N. Ueda, “Multichannel Extensions of Non-Negative Matrix Factorization With Complex-Valued Data,” *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 21, no. 5, pp. 971–982, 2013.
- [20] J. Nikunen and T. Virtanen, “Direction of Arrival Based Spatial Covariance Model for Blind Sound Source Separation,” *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 22, no. 3, pp. 727–739, 2014.
- [21] A. Ozerov, C. Févotte, R. Blouet, and J.-L. Durrieu, “Multichannel nonnegative tensor factorization with structured constraints for user-guided audio source separation,” in *2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2011, pp. 257–260.
- [22] T. Kim, “Real-Time Independent Vector Analysis for Convolutional Blind Source Separation,” *Circuits and Systems I: Regular Papers, IEEE Transactions on*, vol. 57, pp. 1431–1438, Aug. 2010.
- [23] D. Kitamura, N. Ono, H. Sawada, H. Kameoka, and H. Saruwatari, “Determined Blind Source Separation Unifying Independent Vector Analysis and Nonnegative Matrix Factorization,” *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 24, no. 9, pp. 1626–1641, 2016.
- [24] N. Ono, “Stable and fast update rules for independent vector analysis based on auxiliary function technique,” in *2011 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, 2011, pp. 189–192.
- [25] V. Popa, B. Moroşanu, and C. Negrescu, “Head related transfer function measurement in reverberant rooms,” presented at the ATOM-N, Constanţa, 2022.
- [26] A. Alinaghi, P. J. Jackson, Q. Liu, and W. Wang, “Joint Mixing Vector and Binaural Model Based Stereo Source Separation,” *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 22, no. 9, pp. 1434–1448, 2014.
- [27] H. Sawada, S. Araki, and S. Makino, “Underdetermined Convolutional Blind Source Separation via Frequency Bin-Wise Clustering and Permutation Alignment,” *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 19, no. 3, pp. 516–527, Mar. 2011.