



Universitatea POLITEHNICA din București

Școala Doctorală ETTI-B

TEZĂ DE DOCTORAT – REZUMAT –

**CONTRIBUȚII LA EXPERTIZA CRIMINALISTICĂ
A ÎNREGISTRĂRILOR AUDIO DIGITALE**

**CONTRIBUTIONS TO FORENSIC EXAMINATION
OF DIGITAL AUDIO RECORDINGS**

Doctorand: **MSc. Ing. Gheorghe POP**

Conducător de doctorat: **Prof. Dr. Ing. Corneliu BURILEANU**

BUCUREȘTI 2020

Mulțumiri

Această lucrare nu ar fi fost posibilă fără contribuția esențială a domnului profesor Corneliu Burileanu, atât în privința studiilor științifice aprofundate cât și a cercetărilor efectuate în cadrul programului de doctorat.

Îi sunt recunoscător în mod special domnului profesor Dragoș Burileanu, căruia doresc să-i mulțumesc și pe această cale, pentru contribuțiile sale la formarea mea științifică și umană, îndrumările, răbdarea și sprijinul moral oferite cu generozitate în ultimii zece ani și, mai ales, pe durata elaborării lucrării. Reușita acestei lucrări i se datorează într-o bună măsură.

Aprecieri profunde se cuvin și domnilor profesori Andi Buzo și Horia Cucu, pentru consolidarea pregătirii mele de specialitate în domeniul tehnologiilor vorbirii.

Le sunt îndatorat colegilor mei de cercetare doctorală, Dragoș Drăghicescu, Sorin Rusu, Alexandru Caranica și Șerban Mihalache, pentru susținerea, camaraderia și colaborarea lor, pe care mi le-au oferit fără ezitare.

Recunoștința autorului se îndreaptă în mod deosebit spre colegii experți criminaliști din cadrul Institutului Național de Expertize Criminalistice, cu precădere spre domnul Grigoraș Bețiu – directorul institutului, domnii Sorin Alămoreanu și Constantin Mirea – formatori în cadrul programului de masterat BIOSINF, și, nu în ultimul rând, spre domnii Cristian Dumitrescu și Florin Rușitoru, pentru permanenta încurajare și susținere în sensul ridicării continue a nivelului științific al cercetării și al lucrărilor de expertiză criminalistică pentru efectuarea cărora am fost desemnat.

Le mulțumesc de asemenea soției mele, Mona-Luiza, și fiului meu, Radu Cristian, pentru răbdarea și sprijinul moral oferite pe întreaga durată a elaborării acestei lucrări..

Cuprins

Mulțumiri	i	2
1. Introducere	1	4
1.1 Prezentarea domeniului tezei de doctorat	1	4
1.2 Scopul tezei de doctorat	2	4
1.3 Conținutul tezei de doctorat	2	4
2. Domeniul expertizei criminalistice a înregistrărilor audio	5	4
2.1 Locul expertizelor judiciare în contextul criminalisticii	5	4
2.2 Autentificarea înregistrărilor audio	9	5
2.3 Recunoașterea vorbitorului	10	5
2.4 Îmbunătățirea calității semnalului vocal	14	6
3. Metode de calcul în criminalistica înregistrărilor audio	19	6
3.1 Tehnici de calcul matematic în criminalistică	19	6
3.2 Învățarea automată. Metode și algoritmi	39	7
3.3 Algoritmi utilizați în recunoașterea vorbitorului	43	7
3.4 Învățarea automată a rețelelor neurale profunde	49	7
4. Autentificarea înregistrărilor audio digitale prin analiza ENF	53	8
4.1 Verificarea integrității înregistrărilor audio prin analiza ENF	54	8
4.2 Metodă propusă de construcție a referințelor ENF	56	8
4.3 Metodă propusă de căutare accelerată a potrivirii secvențelor-urmă cu baza de date ENF	63	9
4.4 Cadru evaluativ propus de autentificare a înregistrărilor audio bazată pe analiza ENF	66	9
4.5 Concluzii	79	12
5. Autentificarea înregistrărilor audio prin analiza compresiei	81	12
5.1 Trecere în revistă a literaturii de specialitate	82	12
5.2 Recunoașterea codecului AMR	84	12
5.3 Concluzii	94	14
6. Recunoașterea vorbitorului avizată de calitate	97	14
6.1 Problema recunoașterii vorbitorului	97	14
6.2 Sistem tipic de recunoaștere a vorbitorului	99	14
6.3 Sistem FASR avizat de calitatea semnalului vocal	102	14
6.4 Evaluarea sistemului FASR propus	105	15
6.5 Concluzii	111	17
7. Îmbunătățirea calității semnalelor vocale	113	17
7.1 Introducere	113	17
7.2 Probleme ale calității semnalului vocal înregistrat	114	17
7.3 Metode „clasice” de îmbunătățire a semnalului vocal	117	17
7.4 Recuperarea semnalului vocal cu rețele neurale profunde	121	18
7.5 Îmbunătățirea avizată de calitate a semnalelor vocale	123	18
7.6 Concluzii	128	19
8. Concluzii	129	19
8.1 Contribuții originale ale autorului	130	19
8.2 Rezultate obținute	132	20
8.3 Lista lucrărilor autorului în domeniul tezei	133	21
8.4 Perspective de dezvoltare	135	22
Bibliografie	137	22

1. Introducere

Înregistrările audio sau audio-video efectuate conform legii, de autorități ori de persoane fizice, sunt mijloace de probă în dreptul românesc. Acestea nu pot fi utilizate, însă, ca argument în motivarea hotărârilor judecătorești dacă nu corespund cu adevărul. Autenticitatea înregistrărilor audio depuse ca probe la dosarele organelor judiciare se verifică prin *expertiză (judiciară)*. Codul de Procedură Civilă al României caracterizează expertiza ca *opinie a unui specialist (sau a unei comisii de trei specialiști) asupra probelor depuse la dosar pentru lămurirea unor împrejurări de fapt*.

1.1 Prezentarea domeniului tezei de doctorat

Domeniul tezei de doctorat este problematica expertizei criminalistice a probelor sub formă de *înregistrări audio*, termen în care se includ și componentele audio cuprinse în înregistrări audio-video sau multimedia, cele mai importante subdomenii fiind:

- autentificarea înregistrărilor;
- recunoașterea vorbitorului; și
- îmbunătățirea vorbirii înregistrate.

1.2 Scopul tezei de doctorat

Teza de doctorat prezentată își propune să contribuie la rezolvarea unor probleme științifice și practice curente din domeniul expertizei criminalistice a înregistrărilor audio, evitând totuși să se constituie, prin furnizarea detaliilor necesare, într-un proiect de pregătire profesională pe cont propriu pentru persoanele interesate să devină experți criminaliști.

1.3 Conținutul tezei de doctorat

Cercetările prezentate în teza de față sunt concentrate pe patru tipuri de probleme ale expertizei criminalistice privind înregistrările cu conținut audio (Capitolele 4 – 7).

În capitolele 2 și 3 sunt prezentate considerente teoretice și noțiuni utile privind bazele științifice ale cercetărilor prezentate în capitolele următoare.

Capitolul 4 a fost dedicat contribuțiilor originale ale autorului privind autentificarea înregistrărilor audio prin analiza urmelor reziduale ale variațiilor frecvenței rețelei electrice (Electric Network Frequency – ENF), numită *analiză ENF*, prin prisma tuturor celor trei etape ale analizei ENF:

- construcția eficientă a bazelor de date de referință;
- căutarea accelerată a potrivirilor între urmele extrase și bazele de date de referință; și
- adaptarea criteriului ENF pentru utilizare într-un cadru evaluativ.

În capitolul 5 sunt prezentate contribuții la autentificarea înregistrărilor audio prin analiza urmelor de compresie, în particular recunoașterea urmelor codecului AMR (-NB).

Capitolul 6 introduce și evaluează un sistem automat de recunoaștere a vorbitorului în cadrul expertizei criminalistice, cu considerarea calității intrinseci a semnalului vocal.

În capitolul 7 este prezentată o metodă de îmbunătățire a calității semnalului vocal pentru expertiza criminalistică prin utilizarea unei rețele neurale profunde (Deep Neural Network – DNN) și a trei măsuri ale calității, utilizate în capitolul precedent.

În capitolul 8 sunt prezentate principalele contribuții ale autorului la expertiza criminalistică a înregistrărilor audio, rezultatele obținute și lucrările publicate de autor în domeniul tezei, precum și direcțiile de dezvoltare ulterioară avute în vedere.

2. Domeniul expertizei criminalistice a înregistrărilor audio

2.1 Locul expertizelor judiciare în contextul criminalisticii

Dintre reacțiile comunității criminalistice la constatări din diferite domenii de specialitate privind scăderea sau absența nivelului științific corespunzător al expertizelor se evidențiază reacția Comitetului pentru domeniul științific al probelor digitale/multimedia din cadrul OSAC (*Organization of Scientific Area Committees for Forensic Science*). Aceasta a înființat un grup de lucru incluzând academicieni, practicieni și reprezentanți ai organismelor internaționale de specialitate criminalistică, având misiunea de a delimita locul probelor digitale/multimedia în criminalistică. Cadrul formalizat definit în raportul grupului a condus la redefinirea unor concepte

fundamentale, printre care acela de *știință (studiul sistematic și coerent al fenomenelor de un anumit fel)* și *urmă (orice modificare, observabilă ulterior, creată ca rezultat al unui eveniment)*.

Criminalistica se evidențiază în acest context, ca *știința ce se ocupă cu studiul sistematic și coerent al urmelor în scopul autentificării, identificării, clasificării, reconstrucției și evaluării lor în context juridic* [Jac13].

Administrarea urmelor ca probe materiale atașate dosarelor organelor judiciare din România are loc în general în cadrul instanțelor de judecată ori în vederea viitoarei administrări într-o instanță de judecată, folosind *mijloace de probă*.

În Codul de Procedură Civilă este definită *expertiza*, stabilind detaliile prin legi speciale:

- Ordonanța Guvernului (OG) nr. 2/2000 privind expertiza tehnică judiciară și extrajudiciară;
- OG nr. 1/2000 privind activitatea și funcționarea instituțiilor de medicină legală; și
- OG nr. 75/2000 privind organizarea activității de expertiză criminalistică.

Expertiza criminalistică este o specie de expertiză judiciară care se distinge de expertiza tehnică, având ca obiect examinarea urmelor materiale în vederea autentificării lor, a identificării obiectelor creatoare ale urmelor (persoane, obiecte sau fenomene) ori a unor modificări ale acestora, precum și evaluarea forței probatorii a urmelor.

Expertiza criminalistică poate include activități de expertiză tehnică în domeniul înregistrărilor audio, în măsura în care acestea se impun în vederea realizării obiectivelor proprii. Expertizele criminalistice se efectuează în institute și laboratoare de specialitate, publice sau private, înființate potrivit legii, printre cele mai reprezentative institute numărându-se:

- Institutul Național de Criminalistică (INC) în cadrul Inspectoratului General al Poliției Române (IGPR), subordonat Ministerului Afacerilor Interne; și
- Institutul Național de Expertize Criminalistice (INEC), în subordinea Ministerului Justiției.

2.2 Autentificarea înregistrărilor audio

Urmele sub formă de evenimente acustice înregistrate, fie acestea înregistrări audio sau audio-video, fac obiectul *expertizei vocii și a vorbirii*, având ca obiective *autentificarea înregistrărilor audio, recunoașterea vorbitorului și îmbunătățirea calității semnalului vocal* [Pop14b].

Sub sancțiunea nulității, hotărârile pe care le pronunță instanțele de judecată nu se pot întemeia pe probe neautentice, motiv pentru care probele ce nu pot fi interpretate direct de magistrați se *verifică* (respectiv se *autentifică*) prin expertiză.

Metodele de autentificare sunt în general metode „oarbe”, care urmăresc stabilirea unor caracteristici fără a se cunoaște a priori modalitatea de realizare a înregistrărilor.

Expertiza de recunoaștere a vorbitorului a fost prima metodă de autentificare a înregistrărilor convorbirilor, evoluând ulterior ca expertiză de sine stătătoare.

2.3 Recunoașterea vorbitorului

Tehnic vorbind, *identificarea* vorbitorului dintr-o mulțime *închisă* (în care toți vorbitorii sursă potențială a probei sunt cunoscuți) sau *deschisă* (nu se exclude ipoteza unui vorbitor necunoscut), *detecția* vorbitorului (stabilirea prezenței acestuia după vorbire) și *verificarea* vorbitorului (există un singur vorbitor bănuț) reprezintă scenarii ale aceleiași abordări generale în domeniu, numită *recunoașterea vorbitorului*.

Baza științifică a expertizei de recunoaștere a vorbitorului este setul de proprietăți anatomice și comportamentale ale sistemelor de producere și percepție a vorbirii, reflectate în vorbire.

Vorbirea poate fi definită ca o succesiune de evenimente acustice care transportă o încărcătură lingvistică. La nivelul cortexului cerebral are loc atât înțelegerea vorbirii auzite (girusul Heschl, zona Vernike etc.), cât și elaborarea mesajelor ce urmează a fi comunicate (zona Broca etc.) și producerea semnalului acustic vocal, care ajunge până la urechile interlocutorilor. În contextul zgomotului perceput, oamenii vorbesc mai tare, în mod reflex (efectul Lombard).

În timpul producerii vorbirii, tractul vocal procesează fluxul de aer expulzat de sistemul respirator, îl filtrează și îl eliberează atât pe gură cât și pe nas. Căile aeriene supraglotice dispun de structuri anatomice prin care se efectuează *fonația* și *articularea*. În timp ce fonația se realizează în laringe, articularea se realizează simultan și coordonat prin toate elementele mobile ale tractului vocal, numite din acest motiv *articulatori*.

Vorbirea în limba maternă se învață la vârsta imitației. Copilul repetă ceea ce aude, iar apoi utilizează vorbirea în mod reflex. Automatismele dobândite la învățarea limbii materne, da și a altor limbi, se reflectă în vorbire, contribuind astfel la recunoașterea vorbitorului.

Cunoașterea parametrilor filtrului realizat de tractul vocal al unei persoane bănuite, estimat pe baza unor eșantioane de vorbire de referință, permite a se verifica dacă un semnal vocal suspus expertizării, constând în anumite cuvinte înregistrate, a putut fi generat de sistemul de producere a vorbirii propriu persoanei respective.

2.4 Îmbunătățirea semnalului vocal

Calitatea scăzută a vorbirii înregistrate este cauzată în general de condițiile dificile de înregistrare sau transmitere, îmbunătățirea având ca obiectiv asigurarea înțelegerii convorbirilor.

Semnalul vocal prezintă redundanță informațională, care poate fi redusă prin *compresie*, bazată pe proprietățile percepției auditive umane. Componentele anatomice ale analizatorului auditiv uman (urechea) sunt grupate în *urechea externă*, *urechea medie* și *urechea internă*, iar senzația de auz se formează în cortexul cerebral, stimulată de impulsurile culese de nervul auditiv.

Transmiterea pe cale neuronală a informației, care presupune eliberarea în sinapsele nervului auditiv a unor neurotransmițători, introduce un *timp de relaxare* al urechii, necesar pentru refacerea cantităților de substanțe eliberate. Evenimentele acustice mai apropiate decât timpul de relaxare, de ordinul a 50 ms – 100 ms, nu pot fi distinse.

Semnalul vocal natural are o bandă de frecvențe ce se poate extinde aproximativ de la 70 Hz până la 7 kHz, deși circa 95% din energia sa se găsește într-o plajă de frecvențe limitată între 100 Hz – 5 kHz. Vorbirea înregistrată prin canale telefonice are o bandă de frecvențe între 300 Hz și 3.4 kHz. Aceste cifre evidențiază o deosebire considerabilă între benzile de frecvență ale semnalelor vocale înregistrate și cea a semnalelor ce pot fi auzite de oameni, situată între 20 Hz și 20 kHz.

3. Metode de calcul în criminalistica înregistrărilor audio

Acest capitol își propune să concentreze principalele metode de calcul matematic utilizate în cadrul programului de cercetare – specifice domeniului înregistrărilor audio și aplicației criminalistice a procesării semnalelor – astfel încât prezentarea cercetărilor efectuate să fie cât mai simplă și clară.

3.1 Tehnici de calcul matematic în criminalistică

În criminalistică sunt necesare raționamente în condiții de incertitudine, pentru asigurarea conexiunilor etiologice corecte între urme și concluziile expertizei. Învățarea automată se utilizează pentru a rezolva probleme ce nu pot fi rezolvate analitic.

În cadrul expertizei vocii și a vorbirii se învață modele ale vorbirii și vorbitorului care se utilizează în sarcinile de recunoaștere a vorbitorului în diferite scenarii (identificare, verificare, detecție, screening etc.).

Se evidențiază câteva modalități uzuale de raționament:

- *Raționamentul prin analogie*, centrat pe compararea cunoscut/necunoscut.
- *Raționamentul deductiv*.
- *Raționamentul inductiv*.
- *Raționamentul probabilistic*.

Organizația Internațională pentru Standardizare (*International Organization for Standardization* – ISO) a publicat un ghid pentru tratarea incertitudinii de măsurare, aplicabil tuturor determinărilor cu rezultat numeric.

Incetitudinea se comunică numeric și prin expresii verbale. *European Network of Forensic Science Institutes* (ENFSI) recomandă utilizarea *raportului de plauzibilități* (Likelihood Ratio – LR), iar echivalența valorilor acestuia cu expresii verbale este stabilită convențional.

Când expertul criminalist trebuie să stabilească răspunsul la întrebarea pusă de organul judiciar, se află în situația de a cerceta lucruri pe care nu le-a perceput în direct, la momentul producerii lor. Investigarea lor completă este nefezabilă, întrucât nu pot fi verificate toate ipotezele posibile. În acest context se folosește *metoda testelor negative* (numai testele care pot exclude autenticitatea sunt relevante. Orice *studiu* (sau *metodă* sau *experiment*) proiectat numai să confirme o ipoteză sau teorie este, prin această definiție, *neștiințific* [Pop63].

Analiza semnalelor vocale în vederea schimbării reprezentării sale pe termen scurt presupune prelucrarea pe fragmente de semnal cu durată redusă, de 10-35 ms, cu sau fără suprapunerii. Prin sinteză se poate restaura reprezentarea inițială.

În practica de expertiză, semnalele digitale sunt secvențe de numere, cărora li se aplică prin analiză reprezentări pe termen scurt precum: spectrul discret de putere, spectrograma, descompuneri cu bancuri de filtre (în special pentru calculul cepstrului), seturi de coeficienți (LPC, MFCC etc.) sau măsuri ale calității semnalului vocal.

Trăsăturile reprezentării pe termen scurt se descriu prin distribuțiile lor de probabilitate.

3.2 Învățarea automată. Metode și algoritmi

Obiectivul principal al învățării automate este să ajute sistemele de calcul în îndeplinirea de sarcini fără a necesita intervenția umană.

Sistemele inteligente utilizează mai multe moduri de învățare pentru a achiziționa informații privind stările ascunse ale mediului:

- *nesupravegheată*, pentru autoasociere sau reducerea dimensionalității datelor;
- *supravegheată*, pentru modelare generativă sau discriminativă;
- *semi-supravegheată*, pentru cazul seturilor de date etichetate parțial;
- *susținută*, pentru a stabili, în colaborare cu un expert în domeniu, ce acțiuni urmează a fi întreprinse în domeniul de aplicație.

3.3 Algoritmi utilizați în recunoașterea vorbitorului

Vorbitorii comparabili se examinează adecvat scenariului de recunoaștere stabilit și modelelor de vorbitor și de populație, care se obțin prin învățare automată. În baza modelelor antrenate se stabilește ipoteza criminalistică susținută de fiecare comparație realizată, precum și gradul de susținere oferit ipotezei respective.

Vorbitorii pot fi modelați spre exemplu printr-o densitate de probabilitate a *coeficienților cepstrali pe scara Mel* (Mel-Frequency Cepstral Coefficients – MFCC), numiți și *coeficienții Mel-cepstrali*, sau *trăsături MFCC*. În contextul variațiilor vocii sub acțiunea unui număr mare de factori, modelul *densitate de probabilitate a trăsăturilor MFCC* tinde spre un *model mixtură de Gaussiene* (GMM).

Modelele mixtură de Gaussiene se estimează prin algoritmul EM (Expectation Maximization), dacă se dispune de o cantitate suficientă de date. Este cazul modelelor de fond universal (UBM) pentru populații. În cazul uzual, nu sunt îndeplinite condițiile pentru EM pentru fiecare vorbitor, fiind preferat algoritmul MAP (Maximum A Posteriori). În sisteme automate, se calculează scoruri de similitudine sau de plauzibilitate. Scorurile de similitudine pot crea probleme din cauza similitudinilor întâmplătoare.

Astfel se impune necesitatea unui instrument de discriminare între trăsăturile tipice și cele specifice, oferit în acest caz chiar de algoritmul MAP. Scorul de plauzibilitate tipic este

$$LR_k = \frac{p(X | H_{0k})}{p(X | H_1)}, \quad (3.67)$$

unde H_{0k} și H_1 sunt ipotezele criminalistice utilizate în raționamentul evaluativ:

H_{0k} – vorbitorul necunoscut este vorbitorul k ; respectiv

H_1 – vorbitorul necunoscut este un membru oarecare al populației relevante.

Prin intermediul LR este livrată la ieșire atât decizia de recunoaștere cât și tăria probei. Corpul de profesioniști din domeniul juridic nu poate opera, însă, cu valori în formă numerică ale forței probatorii, motiv pentru care se preferă utilizarea în paralel a unor scale verbale.

3.4 Învățarea automată a rețelelor neurale profunde

Rămase la nivel de idee interesantă, din cauza lipsei în epocă a puterii de calcul, rețelele neuronale artificiale au fost reactivate în a doua jumătate a anilor '80. În prezent este utilizată pe scară largă ajustarea automată a parametrilor rețelei neurale chiar și pentru arhitecturi profunde, numită acum *antrenare profundă* (Deep Learning – DL).

Paul Werbos a propus în 1990 unul dintre cei mai buni algoritmi de antrenare a rețelelor neurale, numit *propagare inversă* (în limba engleză *backpropagation*) [Wer90].

Ca reacție împotriva așa-numitelor *efecte de ordin superior* a fost propusă în [Iof15] *normalizarea pe loturi* (în limba engleză *batch normalization*). Măsurile de normalizare pe loturi fac gradientii mai relevanți statistic pentru antrenarea rețelei, însă nu elimină *efectele de ordin superior* care, stimulate de datele de intrare, pot deturna învățarea. O soluție mult mai cuprinzătoare este *normalizarea pe straturi* (în limba engleză *layer normalization*). Diferența principală între cele două moduri de normalizare este raza de acțiune a statisticilor de normalizare.

În primul caz statisticile sunt comune tuturor neuronilor aflați pe aceeași poziție în fiecare vector de date de intrare din lot, în timp ce în a doua abordare, statisticile sunt aceleași pentru toți neuronii din fiecare vector de intrare din lot, probabil diferite pentru loturi diferite.

O metodă de prevenire a învățarea co-dependențelor dintre activările neuronilor, este *omiterea probabilistică* (în limba engleză *dropout*).

4. Autentificarea înregistrărilor audio prin analiza ENF

Rețeaua de distribuție a energiei electrice cu acoperire națională, numită și rețea electrică, lucrează în interconectare și sincronism cu rețeaua europeană.

4.1 Verificarea integrității înregistrărilor prin analiza ENF

Sarcina rețelelor electrice este variabilă din motive evidente, iar la dezechilibrele temporare se declanșează mecanisme de control ale rețelei, proiectate să mențină parametrii standard ai acesteia. Puterea generată este controlată pentru a urmări puterea consumată, conducând la variații ale frecvenței sub forma unor modificări rapide, impredictibile, urmate de reveniri asimptotice spre nivelul de echilibru. Primele cercetări în domeniu, efectuate de Cătălin Grigoraș [Gri03], au definit criteriul ENF, care presupune trei etape:

- colectarea bazei de date de referință;
- detectarea și recuperarea urmelor semnalului rețelei electrice; și
- potrivirea urmelor în raport cu referința.

Urmele detectate în materialul înregistrat sunt extrase pentru analiză sub forma unei serii de valori ale frecvenței, care se compară cu secvența de referință de la ora vizată. Bazele de date de referință sunt gestionate de operatorii de transport al energiei, dar pot fi construite de orice laborator.

4.2 Metodă propusă de construcție a referințelor ENF

În figura 4.1 este ilustrată schema bloc a modulului de colectare implementat de autor [Pop17]. Semnalele în antifază sunt conectate la circuite analogice de axare, care translatează complet formele de undă în plaja tensiunilor pozitive. La intrarea periodmetrelor semnalul este format dreptunghiular și acționează o poartă pentru impulsurile unui ceas local, care sunt numărate.

Variațiile frecvenței rețelei sunt reduse,

$$\varepsilon = \frac{|\Delta f|}{f_G} < \frac{0.15}{50} = 0.3\%. \quad (4.2)$$

Având în vedere că $\varepsilon^2 < 0.003^2 = 0.000009$, pot fi omiși termenii cu exponent mai mare decât 1 din dezvoltarea în serie infinită MacLaurin, obținându-se

$$1 + \varepsilon \approx \frac{1}{1 - \varepsilon}. \quad (4.4)$$

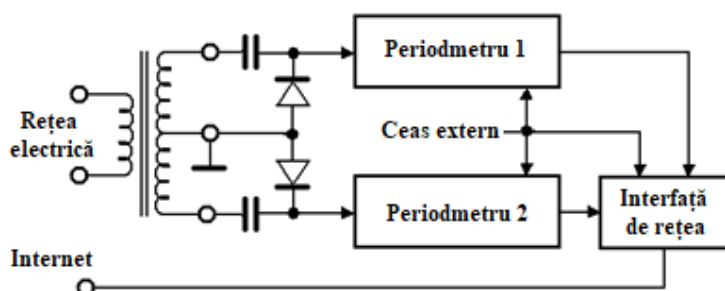


Figura 4.1 Schema bloc a modulului colector propus [Pop17].

Ecuția (4.4) are un impact dublu. Ea permite o conversie foarte rapidă a perioadelor relative în frecvențe relative. Pe de altă parte, dacă seria $1+\varepsilon$ este Gaussiană, și seria $1-\varepsilon$ este Gaussiană. Valorile așteptate ale seriilor $1+\varepsilon$ și $1-\varepsilon$, sunt 1, în timp ce abaterile lor standard sunt identice.

Succesul metodei este asigurat însă prin exprimarea pe 15 biți doar a părții rapid variabile a duratei, situată în intervalul 19.9 – 20.1 ms pentru 99.99% din timp. Pentru o frecvență a ceasului intern de 30 MHz, metoda asigură o rezoluție temporală de 33 ns, respectiv o rezoluție relativă de $33 \cdot 10^{-9} / 20 \cdot 10^{-3} = 1.65 \cdot 10^{-6}$.

Comparativ cu abordarea originală din [Gri03], în care se stochează 120 de eșantioane audio pe secundă, ocupând 16 biți fiecare, metoda propusă în această lucrare folosește spațiul de stocare de 120 de ori mai eficient, iar datele anuale încap în **121 MB**, adică mai puțin de 20% din capacitatea unui disc optic compact (CD) pe an.

Precizia relativă de măsurare este de $1.65 \cdot 10^{-6}$, respectiv de **0.00008 Hz** în termeni absoluți. Precizia metodei se compară în tabelul 4.1 cu cea a altor metode.

Tabelul 4.1 Precizia și durata cadrului de analiză pentru metodele comparate

Metoda de extragere	Precizia [Hz]	Durata cadrului [ms]
FDR [Zha10]	0.0005	20
ESPRIT [Haj12]	0.0001	2000
MUSIC [Haj12]	0.0002	2000
Spectrografică (QIFFT) [Coo09]	0.0006	4000
Metoda propusă	0.00008	1000

4.3 Metodă de căutare accelerată a potrivirilor

Comportamentul pe termen scurt al unei secvențe-referință ENF poate fi evaluat prin comparație cu *media sa mobilă* (Moving Average – MA) potrivit exemplului din figura 4.4. Domeniul de căutare este condensat prin eliminarea pozițiilor în care factorul de corelație binar,

$$C(k) = 1 - \frac{1}{N} \sum_{n=1}^N \text{XOR}(bq_n, br_{n+k}), \quad (4.5)$$

ia o valoare mai mică decât un prag stabilit experimental.

Decizia finală cu privire la potrivirea secvențelor este luată în pasul al doilea, bazat pe utilizarea de măsuri ale similitudinii formelor de variație. Raportului de condensare este cel puțin dublu pentru secvențe-urmă de 240 de eșantioane, și **de cel puțin 7 ori** pentru secvențe de 600 de eșantioane.

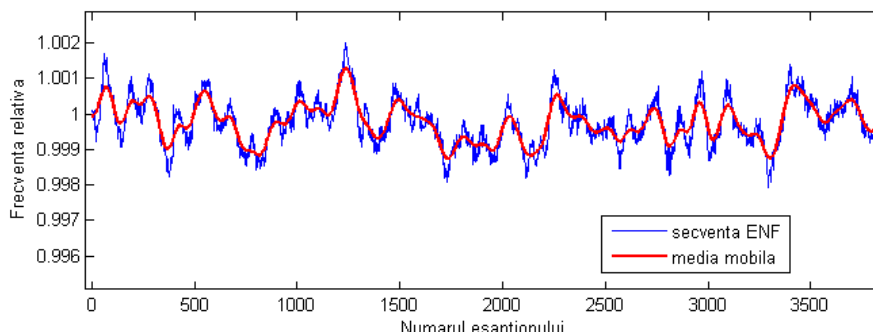


Figura 4.4 Exemplu de secvență de referință și relația ei cu media mobilă.

4.4 Cadru evaluativ de autentificare a înregistrărilor audio

Cadrele de lucru evaluative consideră toate interpretările posibile ale probelor. Fondul problemei decurge din descrierea în [Gri03] a cadrului de referință și din necesitatea criminalistică a formulării concluziilor într-un cadru evaluativ. Bazele de date cu fișiere purtătoare de ENF din cazuri reale sunt încă rare, astfel încât cadrele de autentificare discutate în acest capitol se compară pe corpusuri preparate prin mixarea semnalului rețelei electrice cu perturbații staționare și variabile.

Cadrul de referință utilizează ca indicatori de similitudine între secvența recuperată și baza de date de referință, atât coeficientul de corelație cât și eroarea pătratică medie (CC și MSE). Pentru o pereche de secvențe de aceeași lungime și cu același pas, referința este etichetată fie ca „potrivire”, fie ca „nepotrivire”. Tăria probei nu se evaluează.

Cadrul de lucru propus este o **implementare evaluativă a criteriului ENF**, în care, alături de decizia de potrivire se stabilește și forța probatorie a acesteia. Anterior aplicării se stabilește existența și unicitatea în fișier a urmei. Pornind de la indiciile disponibile în speță, domeniul de căutare poate fi condensat folosind metoda propusă în acest sens, descrisă în secțiunea 4.3. Fiecare secvență de referință cu potențial de potrivire este verificată folosind **măsura propusă de autor pentru similitudine**, respectiv *similitudinea relevantă* (Relevant Similarity – RS), definită ca *numărul de eșantioane din secvența-urmă ENF ce diferă de eșantioanele de referință cu mai puțin decât un anumit prag*.

Pentru testarea cadrului evaluativ propus, descris detaliat în subsecțiunile 4.4.2 și 4.4.3, am utilizat o bază de 4464 fișiere conținând înregistrări audio cu semnalul rețelei electrice, fiecare cu durată de 10 minute, pe care le-am înregistrat consecutiv timp de 31 de zile, într-un aranjament fix. Din această bază s-au obținut ulterior două corpusuri: unul format din fișiere cu zgomot alb, staționar, cu patru valori fixe ale SNR, și un corpus de fișiere având nivel variabil al SNR, zgomotul constând în muzică rock.

Primul corpus a fost utilizat pentru evaluarea experimentală a performanțelor cadrului de autentificare propus cu urme ale semnalului rețelei electrice pe fond de zgomot staționar. Pentru evaluarea cadrului propus cu urme de calitate variabilă au fost utilizate ambele corpusuri pregătite. Au fost realizate în total patru experimente.

- Primul experiment a dotat cadrul de referință cu capabilitate evaluativă prin relația

$$LR = \frac{\text{pdf}(CC|H_0, D)}{\text{pdf}(CC|H_1, D)}, \quad (4.11)$$

unde $\text{pdf}(\cdot)$ este funcția densitate de probabilitate, CC este coeficientul de corelație, iar D reprezintă datele contextuale, precum valoarea SNR, durată urmei etc.

Distribuțiile de probabilitate ale CC sunt redată în figura 4.9 (a), (b), (c) și (d). Din figură reiese faptul că pragul de decizie optim depinde atât de SNR, cât și de factori necunoscuți, transformarea în cadru evaluativ nefiind fezabilă.

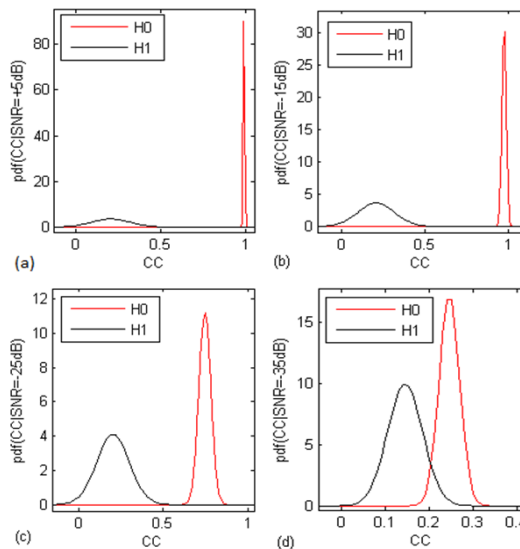


Figura 4.9 Distribuții de probabilitate CC pentru secvențe asociate cu valori SNR de (a) +5 dB, (b) –15 dB, (c) –25 dB și (d) –35 dB.

- În cadrul celui de-al doilea experiment, au fost construite distribuțiile de probabilitate ale RS, redată în figura 4.10 (a), (b), (c) și (d). Din figura 4.10 se deduce că înlocuirea directă a măsurii CC cu măsura RS nu produce un avantaj aparent, pentru aceleași valori ale SNR.

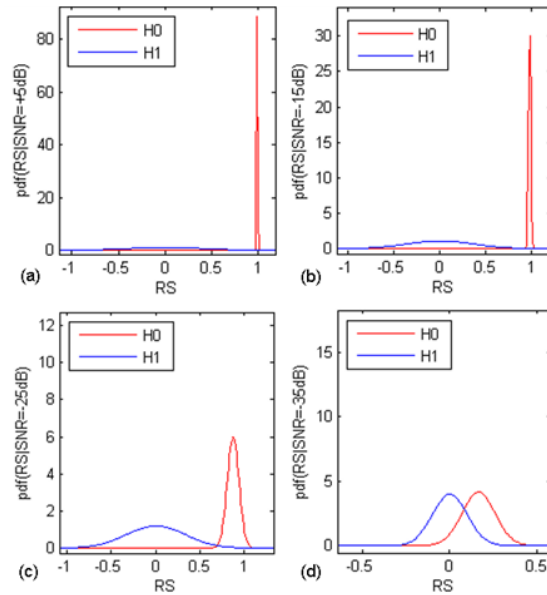


Figura 4.10 Distribuții de probabilitate RS, la valori SNR de (a) +5 dB, (b) -15 dB, (c) -25 dB și (d) -35 dB.

- În cadrul celui de-al treilea experiment au fost construite distribuțiile de probabilitate ale STRS pe primul corpus, redate în figura 4.11 (a), (b), (c) și (d).

Consecința metodei de potrivire propuse, vizibilă în figura 4.11, este că pentru valori SNR de cel puțin -25 dB, distribuțiile de probabilitate a potrivirii secvențelor în cele două ipoteze criminalistice sunt *fixe*. În aceste condiții, forța probatorie a potrivirii între două secvențe se stabilește calculând raportul plauzibilități, cu relația

$$LR = \frac{\text{pdf}(E|H_0, D)}{\text{pdf}(E|H_1, D)}, \quad (4.12)$$

unde H_0 și H_1 sunt ipotezele criminalistice, iar D este setul de date contextuale.

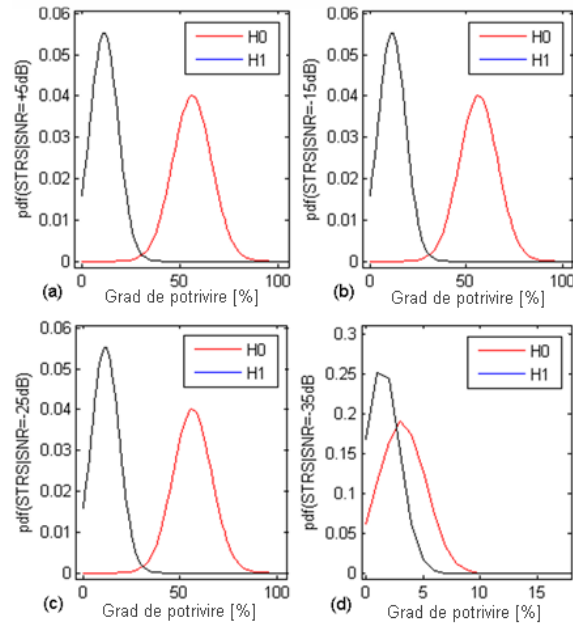


Figura 4.11 Distribuții ale STRS în ambele ipoteze criminalistice, pentru valori SNR de (a) 5 dB, (b) -15 dB, (c) -25 dB și (d) -35 dB.

- În cadrul experimentului 4 au fost aplicate ambele cadre pe corpusul cu zgomot variabil, și s-a efectuat comparația bazată pe curbele ROC din figura 4.13. Se constată că în cazul urmelor ENF cu calitate variabilă, cadrul propus depășește în acuratețe cadrul de referință, permițând valorificarea acestor tipuri de urme cu o rată minimă de erori de detecție fals pozitivă a potrivirii.

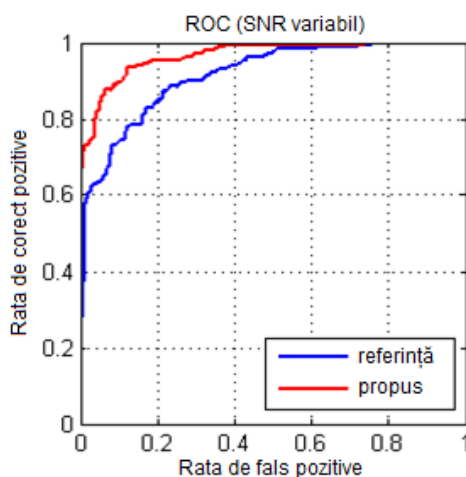


Figura 4.13 Curbele ROC pentru SNR variabil între -45 dB și -25 dB.

4.5 Concluzii

În acest capitol au fost prezentate contribuții originale la analiza ENF privind toate cele trei etape pe care le presupune aceasta:

- o metodă de construcție a bazelor de date ENF de referință;
- o metodă de căutare accelerată a potrivirilor; și
- un cadru de lucru evaluativ de aplicare a criteriului.

În vederea realizării acestor contribuții, au mai fost propuse:

- o măsură a similitudinii formelor de variație a secvențelor ENF – *similitudinea relevantă* (RS) – în general sau pe termen scurt; și
- o metodă de determinare a valorii probatorii a potrivirii secvențelor, bazată pe aplicarea măsurii RS pe intervale scurte (STRS).

5. Autentificarea înregistrărilor audio prin analiza compresiei

Compresia cu pierderi în echipamentele de înregistrare se obține de regulă cu *codoare*, printr-o procesare pe cadre de analiză. La restaurarea semnalelor codate se folosesc *decodoare*. Diferite codoare, printre care și codorul *Adaptive Multi-Rate* (AMR), au capacitate de adaptare la lărgimea canalului de transmisie ori la spațiul de stocare limitat.

5.1 Trecere în revistă a literaturii de specialitate

Cercetările prezentate în literatură privind compresia AMR simplă și dublă au abordat atât analiza formei de undă cât și formatul comprimat al pachetelor de date.

În [Luo16] se descrie un sistem de detecție a compresiei care examinează *ratele de repetiție a eşantioanelor* (Sample Repetition Rate – SRR) pentru a stabili dacă semnalul a fost comprimat AMR.

Autorul a propus în [Dra14a] și [Dra15] un algoritm holistic de recunoaștere a codecului după anumite caracteristici, aplicabil semnalelor decomprimate. A fost utilizată paradigma GMM-UBM pentru a se detecta la examinarea în format necomprimat, urmele unor codec-uri precum AMR, G.729, MP3 sau WMA.

O metodă de detecție a codării AMR-NB și a ratei de bit pentru semnalele vocale, bazată pe DNN, a fost propusă în [Shi18]. Pentru fiecare cadru de semnal s-au extras 184 de trăsături, constând în parametri sensibili la distorsiunile vorbirii precum coeficienții LP de ordinul 10, 12 coeficienți mel-cepstrali și *rata trecerilor prin zero* (ZCR).

5.2 Recunoașterea codecului AMR

Scopul general în expertiza criminalistică al analizei compresiei AMR este autentificarea înregistrării, prin:

- detectarea și identificarea artefactelor de compresie;
- verificarea ratelor de bit; și
- verificarea integrității fluxului de date audio după decompresie.

Restaurarea semnalului vocal la recepție se realizează prin *modelul de sinteză ACELP*, ilustrat în figura 5.1. Erorile de aproximare a sursei și filtrului cresc odată cu scăderea numărului de biți disponibili, respectiv cu creșterea gradului de compresie.

Pentru fiecare *cadru* (20 ms) se decodifică din pachetele de date recepționate *modul de compresie* (exprimat printr-un număr întreg între 0 și 7) și *indicele setului curent de coeficienți LSP* *cuantizați*, care se interpolează pentru fiecare subcadru.

Pentru fiecare *subcadru* (5 ms) se decodifică indecșii rămași și se sintetizează un subcadru de semnal. Anterior livrării la ieșire, semnalul vocal sintetizat este *postfiltrat*, pentru accentuarea formanțelor, compensarea înclinării spectrului, și înlăturarea componentelor sub 60 Hz. Recunoașterea codecului AMR în lucrarea de față se bazează pe utilizarea unor rețele neurale, care realizează clasificarea unor vectori de trăsături extrași din fragmente de semnal cu durată de 20 ms, cu pasul de 5 ms. Pentru antrenarea rețelelor și testarea metodei propuse au fost utilizate fișiere cu semnal vocal de calitate ridicată, atât nemodificate cât și comprimate AMR.

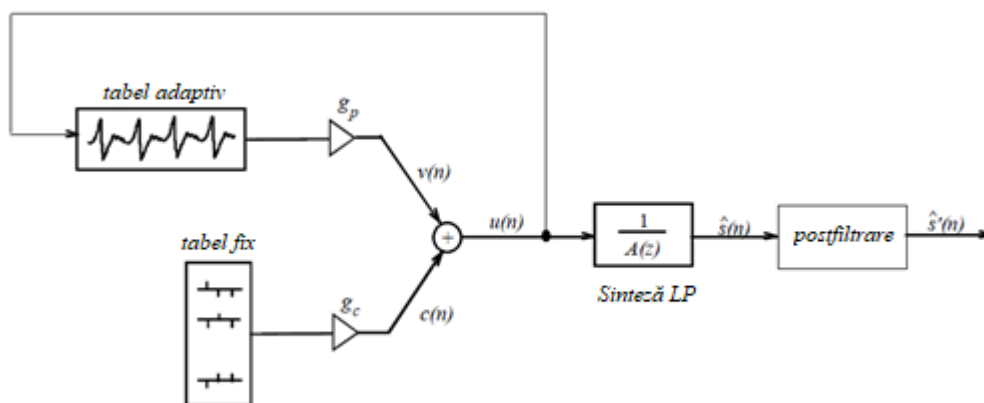


Figura 5.1 Schema bloc de sinteză ACELP a vorbirii în decodorul AMR [3GP02].

Pentru concentrarea relevanței datelor de intrare într-un număr mai mic de trăsături, am elaborat vectori de *trăsături îmbunătățite*, cu 176 de componente per cadru.

În cadrul cercetării descrise în acest capitol au fost desfășurate două experimente de recunoaștere a codecului AMR, bazate pe utilizarea unor rețele neurale cu patru straturi, ilustrate schematic în figura 5.4, cu decizii per cadru de semnal.

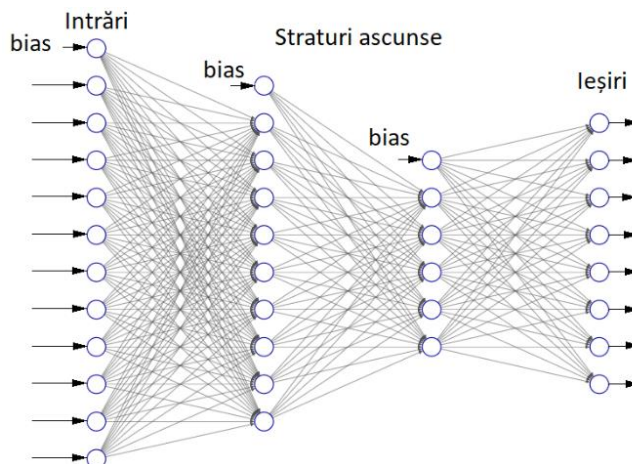


Figura 5.4 Schemă a rețelelor neurale utilizate la recunoașterea codecului AMR.

- Pentru primul experiment, rețeaua neurală a avut două unități în stratul de ieșire („comprimat”/„necomprimat”). Analizând subsecvențe de 20 până la 500 de etichete (0.4 s – 10 s) s-a realizat o acuratețe de 100%, începând cu lungimea de 80 de cadre (1.6 s).

Metoda propusă a depășit metodele descrise în [Luo14] (bazată pe trăsături MDCT) și în [Shi18] (bazată pe DNN), **întrucât lungimea mai mică a seriei de etichete necesară pentru detecție îi permite să acționeze pe intervale mai scurte față de aceste metode.**

- Rețeaua neurală antrenată pentru al doilea experiment are 8 unități în stratul de ieșire („0”, „1”, ..., „7”), semnalul fiind deja cunoscut drept comprimat AMR.

Prin modificarea duratei fragmentelor analizate, **acuratețea de recunoaștere a ratei de bit** suferă variații, evidențiate în tabelul 5.1.

Tabelul 5.1 Variația acurateței clasificării ratei de bit după durata considerată.

F [cadre] (Durata [s])	20 (0.4)	200 (4)	500 (10)
Acuratețea [%]	23.2	72.1	78.0

Acuratețea maximă obținută la detecția ratei de bit a compresiei a fost de **78%**, pentru serii de 500 de vectori de trăsături îmbunătățite (sau 10 s). Din cauza faptului că metodele de referință nu explorează dependența acurateței de durata fragmentului, performanțele sunt comparate în tabelul 5.2 numai pentru fragmente de câte 4 s.

Tabelul 5.2 Comparația clasificatorilor de rată de bit, pe fragmente de 4 s.

Metode comparate	Acuratețe [%]
Metoda descrisă în [Luo14] (bazată pe MDCT)	32.4
Metoda descrisă în [Shi18] (bazată pe DNN)	69.7
Metoda propusă	72.1

5.3 Concluzii

Metoda propusă a detectat cu acuratețe maximă – 100% corect – compresia AMR. De asemenea, prin acuratețea de **72.1%**, metoda propusă pentru detecția ratei de bit depășește metodele de referință pentru această sarcină.

6. Recunoașterea vorbitorului avizată de calitate

6.1 Problema recunoașterii vorbitorului

Încă de la apariția primelor metode tehnice de recunoaștere a vorbitorului, s-a constatat posibilitatea de eroare cauzată de similitudinile întâmplătoare între vocile unor persoane. Experimentele pe scară largă făcute de unele instituții academice au evidențiat amploarea erorilor acestor metode și au determinat luarea unor decizii de către instanțele de judecată privind admiterea probei.

O problemă importantă a semnalelor vocale utilizate în *sistemele automate de recunoaștere a vorbitorului pentru expertiza criminalistică* (FASR) este îndeplinirea unor criterii calitative și cantitative, cu rolul de a se evita obținerea de rezultate imposibil de interpretat.

6.2 Sistem tipic de recunoaștere a vorbitorului

Un tip important de sisteme FASR este cel bazat pe GMM-UBM. Acestea includ o etapă de *preprocesare*, în cadrul căreia se extrag trăsături MFCC, din cadre de semnal cu durata de 15-30 ms și pasul de 10-15 ms. Algoritmul EM (Expectation Maximization), detaliat în subsecțiunea 3.3.2, examinează datele multi-dimensionale și identifică aglomerările formate de acestea.

Algoritmul MAP (Maximum A Posteriori), de asemenea detaliat în subsecțiunea 3.3.2, exploatează relevanța statistică a modelului de fond, adaptându-l la un set de date specific unui singur vorbitor.

În figura 6.2 este ilustrată schema bloc a unui sistem FASR tipic, bazat pe paradigma GMM-UBM. După inițializare (dotare cu UBM), sistemul FASR poate fi utilizat prin parcurgerea etapelor de *înrolare* a fiecărui vorbitor. Pentru inferența Bayesiană ulterioară, se consideră valoarea LR determinată la testare.

6.3 Sistem FASR avizat de calitatea semnalului vocal

În această secțiune este descris sistemul propus pentru expertiza criminalistică de recunoaștere a vorbitorului pe baza modelelor GMM. Un sistem de recunoaștere poate deveni *avizat de calitate* **dacă este dotat cu o schemă de asistență a deciziilor prin calitatea semnalului vocal de la intrare**. Dacă în telecomunicații se consideră calitatea semnalului vocal în raport cu ascultătorul mediu, experții criminaliști o definesc în raport cu sarcina de rezolvat, iar în biometrie se definește o *măsură a calității* ca „informație care contribuie la stabilirea probabilității ca o decizie biometrică să fie corectă”.

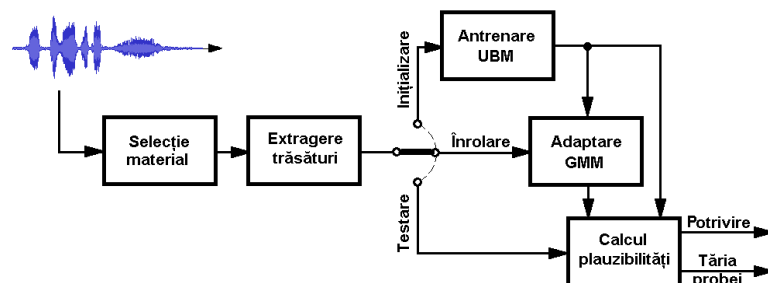


Figura 6.2 Schema bloc a unui sistem FASR bazat pe paradigma GMM-UBM.

Într-un sistem FASR cu un UBM și un model de vorbitor cu M densități Gaussiene d -dimensionale, se creează condiții pentru determinarea contribuțiilor fiecărui cadru de semnal la raportul de plauzibilități, liniar sau logaritm.

6.4 Evaluarea sistemului FASR propus

Sistemul FASR propus, avizat cu măsurile calității prezentate în cadrul tezei, are schema bloc din figura 6.4. Sistemul propus se deosebește de sistemul tipic, descris în secțiunea 6.2, prin extinderi reprezentând trei canale de transmitere a informației de avizare:

1. Modul de selecție a materialului;
2. Modul de calcul al scorului de potrivire; și
3. Structura vectorilor de trăsături.

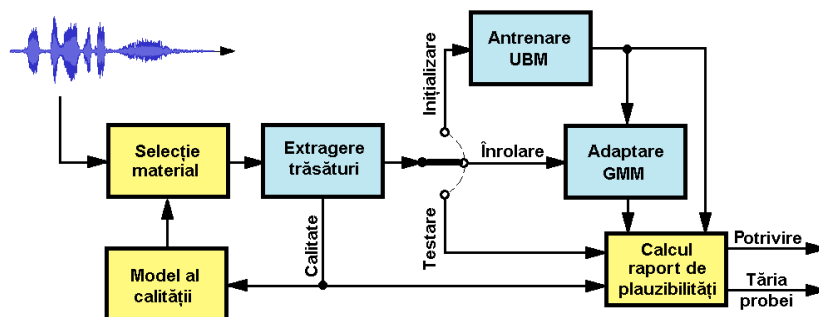


Figura 6.4 Schema bloc a sistemului FASR propus, avizat de calitate.

Cele trei configurații au fost comparate pe baza *acurateței* și *ratei de eroare egală* (EER). Pentru evaluare sintetică, s-au utilizat *matricea de confuzie* și curbele DET.

Fișierele audio cu vorbire, utilizate la antrenarea modelului de fond, a modelelor individuale și pentru evaluare provin din secțiunea *test* a bazei de date NIST2008 SRE [NIS11]. Dintre cele 942 de ore de convorbiri, în care este colectată vorbire provenind de la 1460 de persoane, de ambele sexe, 368 de ore reprezintă înregistrări telefonice, celelalte 574 de ore fiind înregistrări efectuate în studio.

Vorbirea a 420 de persoane a fost utilizată la antrenarea unui UBM, iar câte două fișierele pentru fiecare din ceilalți 39 s-au utilizat pentru înrolarea și testarea acestora.

- Primul experiment a evaluat efectul selecției celui mai bun material asupra recunoașterii vorbitorului. Criteriul de selecție a materialului a fost stabilit impunând un set de praguri măsurilor calității, descrise în lucrare. Matricea de confuzie, în forma nestandard, este redată în figura 6.5. Acuratețea realizată de sistemul propus a fost de 93.1%, cu o rată de eroare egală de 4.7%.

Îmbunătățirea ratei de eroare egală, față de cazul neutilizării informației privind calitatea a fost de **0.9%**, depășind metoda descrisă în [Gar06] și [God93].

- Al doilea experiment a fost dedicat ponderării pe baza calității a contribuțiilor vectorilor de trăsături la calcularea scorului de potrivire. Această configurație a sistemului propus a fost evaluată de autor și în cadrul lucrării [Dra14b], pentru vorbire suprapusă.

Matricea de confuzie pentru configurația 2 este ilustrată în figura 6.6.

Acuratețea pentru configurația 2 a fost de **99.08%**, rata de eroare egală fiind de **2.0%**.

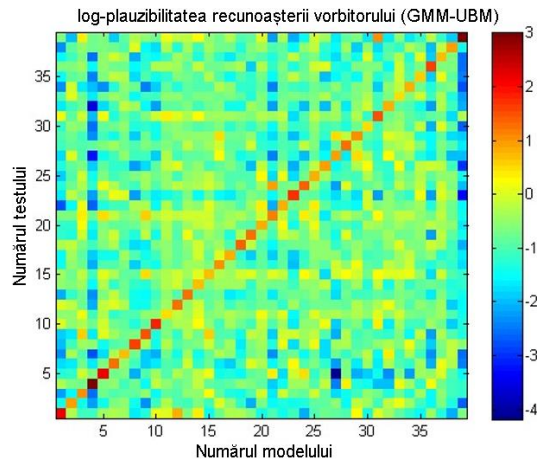


Figura 6.5 Matricea de confuzie a sistemului în configurația 1.

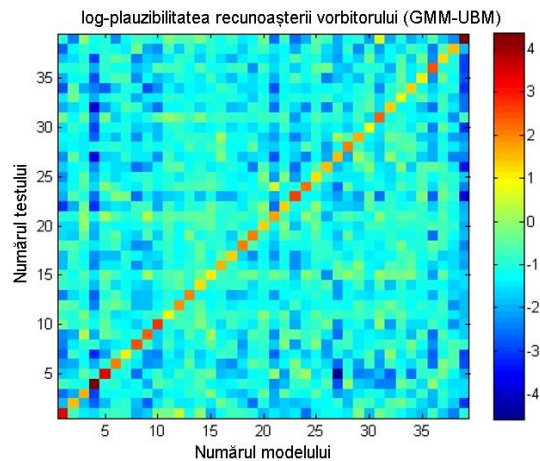


Figura 6.6 Matricea de confuzie pentru configurația cu scor ponderat.

- Extinderea vectorilor de trăsături a impus antrenarea, pornind de la același corpus de fișiere cu semnal vocal, a unui UBM nou, avizat de calitate, și obținerea de GMM-uri individuale pentru vorbitorii cunoscuți, matricea de confuzie a sistemului FASR propus, configurat pentru al treilea experiment, fiind cea ilustrată în figura 6.7.

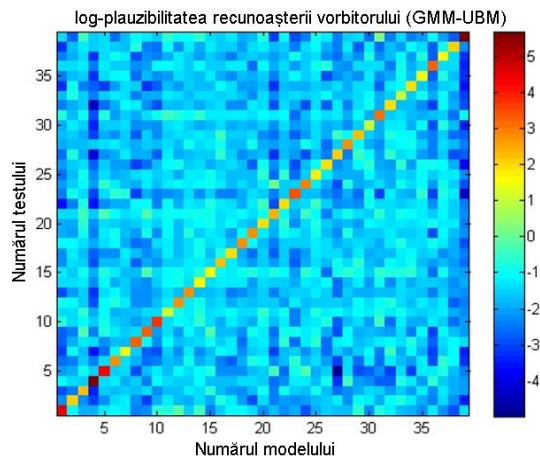


Figura 6.7 Matricea de confuzie pentru configurația cu vectori extinși.

Acuratețea realizată de noua configurație a sistemului propus a fost de **99.8%**, rata de eroare egală fiind de **0.6%**. Această valoare depășește cele mai bune rezultate ale sistemului de referință, obținute pe corpusul TIMIT (EER de 1.01%, după [Sad13]). În figura 6.8 sunt prezentate comparativ curbele *compromisului erorilor de detecție* (Detection Error Tradeoff – DET) pentru cele trei configurații ale sistemului propus.

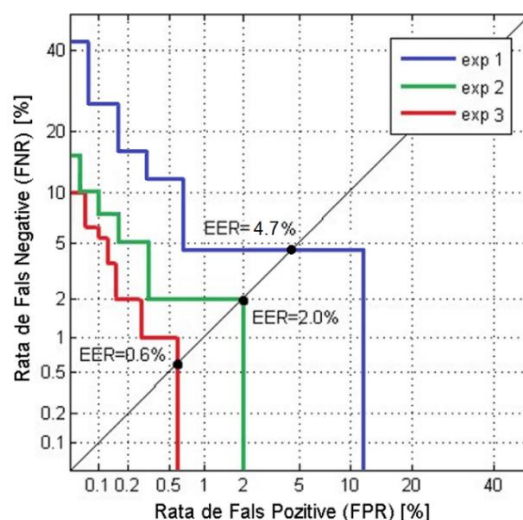


Figura 6.8 Curbele DET ale celor trei configurații experimentale.

6.5 Concluzii

Avizarea cu măsuri ale calității la nivelul calculului scorurilor de potrivire a permis sistemului de recunoaștere a vorbitorului obținerea unei îmbunătățiri considerabile a acurateței, de la 93.1% la **99.08%**, valoare considerată acceptabilă pentru utilizare în cadrul expertizei criminalistice. De asemenea, în configurația în care măsurile calității au fost utilizate în același timp și la selecția materialului de intrare și la calculul scorului de potrivire, sistemul de recunoaștere a atins o performanță la nivelul stării artei, constând într-o acuratețe de **99.8%** și o valoare a EER de **0.6%**.

7. Îmbunătățirea semnalelor vocale

7.1 Introducere

În domeniul expertizei criminalistice îmbunătățirea semnalului se definește ca „procesul de restaurare a semnalului vocal în vederea înțelegerii sale de către oameni”. Acest proces are ca scop obținerea unui mijloc de probă, prin redarea ulterioară în scris a convorbirilor, iar criteriul calitativ principal este măsura în care vorbirea înregistrată poate fi înțeleasă.

Îmbunătățirea calității semnalelor vocale înregistrate a parcurs un drum lung de la egalizoarele grafice analogice până la utilizarea de rețele neurale profunde (DNN).

7.2 Probleme ale calității semnalului vocal înregistrat

De la vorbitor până la stocarea într-un fișier, evenimentele acustice supuse înregistrării parcurg o etapă acustică, o etapă electrică analogică și o etapă digitală.

Calitatea semnalului vocal depinde de:

- *calitatea intrinsecă a semnalului vocal;*
- *canalul acustic sursă – microfon;*
- *superpoziția semnalului de interes cu alte semnale acustice;*
- *neliniaritatea curbei de sensibilitate a microfonului;*
- *zgomotul de conversie analog-digitală; și*
- *zgomotul de compresie cu pierderi.*

Metodele de îmbunătățire a calității semnalului vocal sunt ținute de obicei spre una sau mai multe dintre problemele de calitate enumerate, iar rezultatele lor se măsoară în funcție de obiectivul propus, în cazul ales, prin intermediul *inteligibilității* vorbirii.

7.3 Metode „clasice” de îmbunătățire a semnalului vocal

Una dintre primele tehnici de îmbunătățire a semnalului vocal a fost utilizarea unui aparat analogic denumit *egalizor grafic de spectru* („grafic” pentru ca operatorul să poată aprecia efectul introdus de acesta la un moment dat privind elementele de comandă).

Cele mai cunoscute metode „clasice” de îmbunătățire sunt:

- scăderea spectrală (și diferite versiuni ale acesteia);
- filtrarea optimală (Wiener);

- tehnici bazate pe descompuneri wavelet;
- analiza pe componente independente (Independent Component Analysis – ICA);
- utilizarea unei măști binare ideale (IBM) în domeniul timp-frecvență.

7.4 Recuperarea semnalului vocal cu rețele neurale profunde

Rețelele neurale au demonstrat că pot să învețe corespondențe complexe între datele de intrare și de ieșire, sau chiar reprezentări proprii ale datelor. Cu o antrenare adecvată, o *rețea neurală profundă unidirecțională*, a cărei structură și învățare au fost discutate în secțiunea 3.4, poate să recunoască ori să sintetizeze vorbire, sau să îndeplinească diferite sarcini, printre care și îmbunătățirea calității semnalului vocal.

7.5 Îmbunătățirea semnalelor vocale avizată de calitate

Metoda propusă în acest capitol pentru utilizare în expertiza criminalistică se bazează pe îmbunătățirea semnalului vocal reprezentat ca secvență de vectori de trăsături pe termen scurt.

În figura 7.1 este ilustrată schema bloc utilizată.

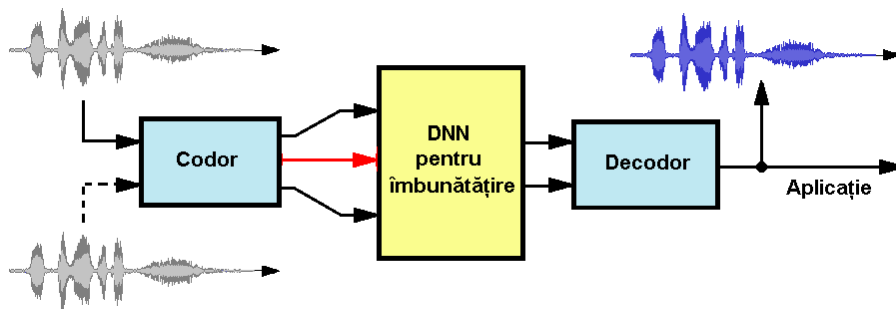


Figura 7.1 Sistem de îmbunătățire a semnalului vocal bazat pe metoda propusă.

- Modulul codor aplică transformata Fourier rapidă, în 512 puncte pe fiecare canal, pe cadre de 32 ms, cu pas de 16 ms. Procesând în continuare spectrul de amplitudini, se obțin câte 256 de *trăsături log-spectrale*, 93 de *trăsături cepstrale* (31 de coeficienți MFCC și câte 31 de variații ale acestora de ordinele unu și doi) și trei *trăsături calitative*.
- *Rețeaua neurală profundă* (DNN) propusă pentru îmbunătățire are o structură unidirecțională, ilustrată schematic în figura 7.2 și este antrenată cu *învățare prin transfer* (în limba engleză *transfer learning*), de la o rețea neurală deja antrenată [Xu15].

Rezultatele testelor efectuate sunt ilustrate în Tabelul 7.1. Valorile WER obținute pe corpusurile NSDTSEA și SSC-eval, descrise în cadrul tezei, au fost determinate folosind sistemul de *recunoaștere automată a vorbirii* (ASR), descris în [Geo18], bazat pe kitul de instrumente Kaldi. Pentru expertiza criminalistică, cel mai important efect al aplicării unei metode de îmbunătățire a semnalului vocal sunt creșterea inteligibilității pe termen scurt (STOI) și distanța Itakura-Saito (DIS) între semnalul vocal recuperat și cel supus înregistrării.

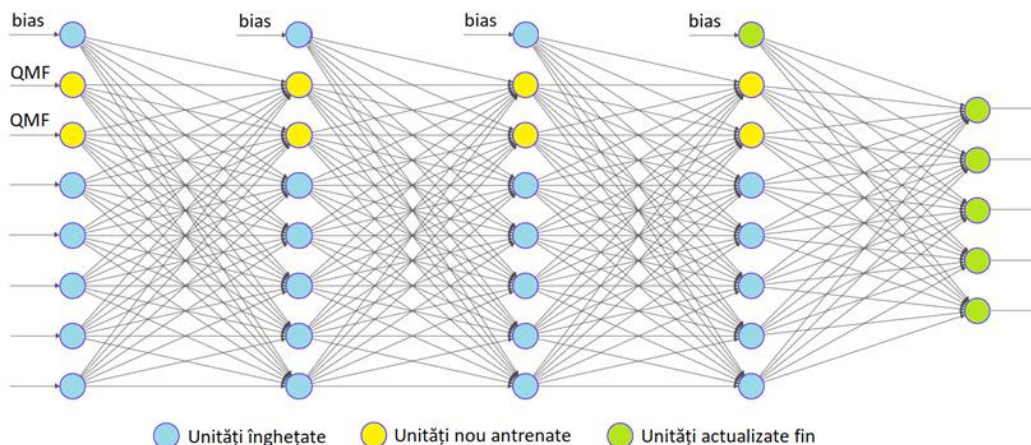


Figura 7.2 Schema structurii rețelei neurale profunde utilizate pentru îmbunătățire.

Tabelul 7.1 Comparația metodei propuse cu alte metode

Metoda	WER [%] pe SSC-eval	WER [%] pe NSDTSEA	PESQ	STOI	DIS
Estimare oarbă a zgomotului, cu filtrare Bayesiană [Doi16]	21.73	9.42	2.47	0.88	53.1
CNN dilatăta (WaveNet) [Qia17]	47.32	10.23	1.15	0.06	57.0
DNN unidirecțională în domeniul spectral [Xu15]	20.68	8.47	2.26	0.87	19.4
Semnal curat	20.02	8.60	–	–	–
Metoda propusă	20.34	8.21	2.49	0.90	16.6

7.6 Concluzii

În baza comparațiilor prezentate în tabelul 7.1 se deduce că antrenarea rețelei neurale cu informație suplimentară constând în măsuri ale calității, precum și cu semnale de intrare corupte cu zgomot aleatoriu, contribuie la o capacitate mai mare de generalizare la medii acustice necunoscute, la vorbitori necunoscuți și la zgomote neîntâlnite anterior, care se situează la nivelul stării artei.

Rezultatele comparațiilor arată de asemenea că metoda propusă realizează cea mai mică pierdere la nivelul WER, în raport cu vorbirea necontaminată, depășind la toate criteriile metoda de referință, descrisă în [Xu15].

8. Concluzii

În lucrarea de față sunt prezentate o serie de contribuții în principalele subdomenii aparținând domeniului expertizei criminalistice a înregistrărilor audio, și anume autentificarea înregistrărilor, recunoașterea vorbitorului și îmbunătățirea semnalului vocal.

Partea introductivă a tezei descrie atât organizarea activităților de expertiză judiciară și extrajudiciară, cât și baza științifică a expertizelor criminalistice, cuprinzând noțiuni de anatomie și fiziologie a producerii și percepției vorbirii, noțiuni de statistică și probabilități, precum și de metodologie și raționament criminalistic.

Capitolele 4 – 7 urmăresc principale obiective ale tezei:

a. punerea la punct a unor algoritmi de creștere a performanței analizei ENF, în toate cele trei etape ale acesteia:

1. colectarea mai eficientă bazelor de date de referință;
2. căutarea accelerată a potrivirilor urmelor cu baza de date; și
3. raportarea evaluativă a gradului de potrivire a secvențelor.

b. elaborarea unei metode de recunoaștere a urmelor codecului AMR folosind un set de trăsături îmbunătățite;

c. implementarea unui sistem de recunoaștere automată a vorbitorului cu utilizarea pe termen scurt a informației privind calitatea semnalului vocal; și

d. elaborarea unei metode de îmbunătățire a semnalului vocal înregistrat pe unul sau două canale prin utilizarea unei rețele neurale în domeniul trăsăturilor spectrale, cepstrale și calitative ale semnalului.

8.1 Contribuții originale ale autorului

Capitolul 4, dedicat aplicațiilor în expertiza criminalistică ale analizei urmelor reziduale ale frecvenței semnalului rețelei de distribuție a energiei electrice, găzduiește prezentarea a trei contribuții originale principale.

1. Elaborarea unei metode de construcție a bazelor de date de referință pentru frecvența rețelei electrice, folosind un modul colector versatil în arhitectură client-server [Pop17].
2. Elaborarea unei metode de căutare accelerată a potrivirilor secvențelor de valori ale frecvenței unei componente ce constituie urma ENF și secvența de valori de referință pentru variațiile reale ale frecvenței rețelei electrice [Pop17].

3. Definirea unei măsuri a similitudinii secvențelor de valori ale frecvenței rețelei electrice, de natură să permită raportarea evaluativă a potrivirilor, indiferent dacă intervalul de timp corespunzător înregistrării a fost cunoscut în prealabil sau a fost stabilit în urma căutării într-o bază de date [Pop18b].

Capitolul 5 a fost dedicat cercetărilor legate de autentificarea înregistrărilor audio prin analiza urmelor de compresie, în particular prin analiza urmelor compresiei AMR, cu accent pe versiunea sa de bandă îngustă. Contribuțiile originale principale ale autorului, prezentate în acest capitol, sunt:

4. Punerea la punct a unei metode de recunoaștere a urmelor de compresie AMR și stabilire a modului de compresie suferit anterior, folosind o rețea neurală profundă [Pop19a].
5. Conceperea și utilizarea unui tip nou de date, cu trăsături primare obținute prin analiza pe termen scurt a semnalului, la nivel de subcadru de semnal, și vectori de trăsături obținuți prin îmbunătățirea acestora pe termen mediu (0.2 s – 4 s) [Pop19a].

Capitolul 6 introduce și evaluează prin prisma calității semnalului vocal un sistem automat de recunoaștere a vorbitorului în cadrul expertizei criminalistice. Contribuțiile originale ale autorului, prezentate în acest capitol, constau în:

6. Demonstrarea caracterului individual al măsurilor calității, prin utilizarea acestora la extinderea vectorilor de trăsături în FASR bazat pe paradigma GMM-UBM [Pop18a].
7. Evidențierea superiorității selecției materialului audio de la intrare bazată pe măsurile calității față de selecția realizată doar prin aplicarea unui VAD [Pop18a].

În capitolul 7 este prezentată o metodă de îmbunătățire a calității semnalului vocal pentru expertiza criminalistică. Contribuțiile originale ale autorului în acest capitol sunt:

8. Îmbunătățirea semnalului vocal bazată pe extinderea datelor de la intrarea unei rețele neurale și transferul învățării dobândite de o altă rețea [Pop19b].
9. Conceperea unui sistem de mixare-demixare care extinde sistemul de îmbunătățire monocanal la semnale cu unul și două canale [Pop19b].

8.2 Rezultate obținute

Capitolul 4:

- Precizia relativă de determinare a frecvenței rețelei: $1.65 \cdot 10^{-6}$ [Pop17].
- Precizia absolută de determinare a frecvenței rețelei: $8 \cdot 10^{-5}$ Hz [Pop17].
- Stocarea secvențelor în termeni relativi, astfel încât dualitatea durată-frecvență se concentrează în semnul abaterii relative [Pop17].
- Colectarea eşantioanelor pe 15 biți, care necesită numai 121 MB pe an [Pop17].
- Stocarea este imună la erori de scriere pe suport [Pop17].
- Sporul de viteză la căutarea în baza de date ENF a secvențelor de cel puțin 240 de secunde: de 2 până la 7 ori [Pop17].
- Raportare evaluativă a potrivirilor secvențelor ENF [Pop18b].
- Capacitate de lucru cu urme ENF cu SNR variabil în înregistrări [Pop18b].
- Aprecierea performanțelor cadrelor de lucru prin curbe ROC [Pop18b].

Capitolul 5:

- Lungimea minimă a fragmentului pentru detecția 100% compresiei AMR: 1.6 s [Pop19a].
- Lungimea minimă a fragmentului pentru recunoașterea ratei de bit: 40 ms [Pop19a].
- Acuratețea de recunoaștere a ratei de bit pentru fragmente cu durata de 10s: 78.0% [Pop19a].

Capitolul 6:

- Sistem de recunoaștere a vorbitorului pentru expertiza criminalistică, cu considerarea calității semnalului vocal în vectorii de trăsături extinși, având acuratețea de 99.8% [Pop18a].
- EER pe NIST 2008 SRE: 0.6%, state-of-the-art la momentul publicării [Pop18a].
- Implementarea unui VAD a posteriori bazat pe măsurile calității, care îmbunătățește modelul vorbirii neutilizabile [Pop18a].

Capitolul 7:

- Îmbunătățirea semnalului peste nivelul de calitate intrinsecă al semnalului curat [Pop19b].
- Aplicabilitate în context criminalistic a metodei de îmbunătățire propuse, justificată de îmbunătățirea inteligibilității și a altor parametri obiectivi ai semnalului vocal [Pop19b].
- Aplicabilitate fără modificări pentru semnale cu unul sau două canale [Pop19b].

8.3 Lista lucrărilor autorului în domeniul tezei

- **Gh. Pop**, Ș. Mihalache, D. Burileanu, „Forensic Recognition of Narrowband AMR Signals”, *Proceedings of the 10th International Conference on Speech Technology and Human-Computer Dialogue (SpeD)*, Timișoara, 6 p., Octombrie 2019, IEEE NY, ISBN: 978-1-7281-0984-8 Indexare IEEE Xplore DOI:10.1109/SPED.2019.8906279
- **Gh. Pop**, D. Burileanu, „Speech Enhancement for Forensic Purposes”, *UPB Scientific Bulletin, Series C – Electrical Engineering and Computer Science*, Vol. 81, Issue 3, Ed. Politehnica Press, București, pp. 41-52, 2019, ISSN: 2286-3540. **ISI WOS:000477996400004**.
- **Gh. Pop**, D. Burileanu, Ș. Mihalache, „An Evaluative ENF-Based Framework for Forensic Authentication of Digital Audio Recordings”, *Proceedings of the Romanian Academy Series A – Mathematics, Physics, Technical Sciences, Information Science*, Ed. Academiei Române, București, Vol. 19, Issue 4, pp. 605-612, Dec. 2018, ISSN: 1454-9069. **ISI WOS:000454140900012 (Q2, IF: 1.752 – Dec. 2018)**.
- **Gh. Pop**, Ș. Mihalache, D. Burileanu, „Forensic Speaker Identification Using Speech Quality Data”, *Proceedings of the 12th International Conference on Communications (COMM)*, București, pp. 509-512, 2018, ISBN: 978-1-5386-2350-3. Indexare IEEE Xplore (DOI:10.1109/ICComm.2018.8484766). **ISI WOS:000449526000096**.
- **Gh. Pop**, D. Drăghicescu, D. Burileanu, H. Cucu, C. Burileanu, „Fast Method for ENF Database Build and Search”, *Proceedings of the 9th International Conference “Speech Technology and Human-Computer Dialogue (SpeD)”*, București, 6 p., 2017, IEEE NY, ISBN: 978-1-5090-6497-7. Indexare IEEE Xplore (DOI:10.1109/SPED.2017.7990447). **ISI WOS:000425849600022**.
- **Gh. Pop**, „Digital Forensics”, Comunicare științifică invitată (*keynote speaker*) la *The International Conference “MODERN FORENSICS TECHNIQUES – possibilities of access via international mutual legal assistance procedures”*, Poiana Brașov, Nov. 16-18, 2016.
- D. Drăghicescu, **Gh. Pop**, D. Burileanu, C. Burileanu, „GMM-based Audio Codec Detection with Application in Forensics”, *Proceedings of the 37th International Conference on Telecommunications and Signal Processing – TSP 2014*, Berlin, pp. 442-446, 2014, ISBN: 978-1-4799-8498-5 (publicat în 2015, împreună cu *TSP 2015 Proceedings*). Indexare IEEE Xplore (DOI:10.1109/TSP.2015.7296421). **ISI WOS:000375231000047**.
- **Gh. Pop**, D. Drăghicescu, D. Burileanu, „A Quality-Aware Forensic Speaker Recognition System”, *Romanian Journal of Information Science and Technology*, vol. 17, Issue 2, Ed. Academiei Române, pp. 134-149, 2014, ISSN: 1453-8245. **ISI WOS:000350281600002**.
- D. Drăghicescu, **Gh. Pop**, C. Burileanu, „Codec Recognition from Decoded Audio”, *Proceedings of the 10th International Conference on Communications (COMM)*, București, pp. 37-40, May 29-31, 2014, ISBN: 987-1-4799-2385-4. Indexare IEEE Xplore (DOI: 10.1109/ICComm.2014.6866684). **ISI WOS:000345844600028**.
- **Gh. Pop**, D. Drăghicescu, D. Burileanu, „On Forensic Speaker Recognition Case Pre-Assessment”, *Proceedings of the 7th International Conference „Speech Technology and Human-Computer Dialogue (SpeD)”*, Cluj-Napoca, pp. 169-176, 2013, ISBN: 978-1-4799-1065-6. Indexare IEEE Xplore (DOI:10.1109/SpeD.2013.6682668). **ISI WOS:000330672700025**.

8.4 Perspective de dezvoltare ulterioară

Sarcina expertizei criminalistice având ca obiect înregistrările audio crește continuu în complexitate, în principal din cauza accesului din ce în ce mai larg al populației la mijloace tehnice cu performanțe de vârf în domeniul înregistrării și procesării datelor audio. Această evoluție a societății impune în criminalistică minimizarea timpului necesar de la descoperirea unui mod de falsificare nou până la validarea științifică a soluției.

Efortul de cercetare trebuie să continue, atât pe direcțiile de examinare prezentate în cadrul lucrării, cât și după următoarele direcții:

1. Utilizarea în recunoașterea vorbitorului a patologiilor vocale.
2. Identificarea de noi estimatori ai calității vorbirii.
3. Punerea la punct a unor sisteme expert dedicate, pe specialități, prin învățare automată.
4. Îmbunătățirea semnalului vocal prin sinteză pornind de la parametrii extrași din semnalul audio contaminat.
5. Autentificare activă prin inserarea în înregistrări a imaginilor faciale ale vorbitorului, filmate în timpul rostirilor.

Bibliografie selectivă

- [Coo09] A.J. Cooper, „An automated approach to the electric network frequency (ENF) criterion – theory and practice”, *International Journal of Speech Language and the Law*, vol. **16**, nr. 2, pp. 193-218, 2009.
- [Doi16] C.S.J. Doire, *Single-channel enhancement of speech corrupted by reverberation and noise*, teză de doctorat, Colegiul Imperial din Londra (Marea Britanie), 2016.
- [Dra14a] D. Drăghicescu, **G. Pop**, C. Burileanu, „Codec Recognition From Decoded Audio”, *Proc of the 10th International Conference on Communications (COMM)*, București, 29-31 Mai 2014, pp. 1-4.
- [Dra14b] D. Drăghicescu, *Analiza înregistrărilor audio și recunoașterea vorbitorului în criminalistică*, teză de doctorat, Facultatea de Electronică, Telecomunicații și Tehnologia Informației, Universitatea „Politehnica”, București, 2014.
- [Dra15] D. Drăghicescu, **G. Pop**, D. Burileanu, C. Burileanu, „GMM-based Audio Codec Detection with Application in Forensics”, *The 38th International Conference on Telecommunications and Signal Processing (TSP)*, Berlin (Germania), 1-3 Iul. 2015, pp. 1-5.
- [Gar06] D. Garcia-Romero, J. Fierrez-Aguillar, J. Gonzalez-Rodriguez, J. Ortega-Garcia, „Using Quality Measures for Multi-Level Speaker Recognition”, *Computer Speech and Lang.*, vol. **20**, nr. 2-3, pp. 192-209, 2006.
- [Geo18] A.L. Georgescu, H. Cucu, „Automatic Annotation of Speech Corpora Using Complementary GMM and DNN Acoustic Models”, *Proceedings of the 41-st International Conference on Telecommunications and Signal Processing (TSP)*, Atena (Grecia), 4-6 Iul. 2018, pp. 1-4.
- [God93] J.J. Godfrey, E. Holliman, *Switchboard-1, Release 2, Telephone Speech Corpus, LDC97S62*, Philadelphia, PA (SUA), Linguistic Data Consortium, 1993.
- [Gri03] C. Grigoraș, „Digital audio recording analysis – The electric network frequency criterion”, *Application note AN-4*, Diamond Cut Productions Inc., Oct. 2003.
- [Haj12] A. Hajj-Ahmad, R. Garg, M. Wu, „Instantaneous frequency estimation and localization for ENF signals”, *Proceedings of The Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA)*, Hollywood, CA (SUA), 3-6 Dec. 2012, pp. 1-10.
- [Iof15] S. Ioffe, C. Szegedy, „Batch normalization: Accelerating deep network training by reducing internal covariate shift”, *Proceedings of the 32nd International Conference on Machine Learning (ICML)*, Lille (Franța), 7-9 Iul. 2015, vol. **37**, pp. 448-456.

- [Jac13] D.-O. Jaquet-Chiffelle, *Traces: New Definition*, Curs introductiv în criminalistica probelor digitale, Școala de Justiție Penală, Universitatea din Lausanne (Elveția), https://serval.unil.ch/resource/serval:BIB_DFE9D125DECA.P001/REF, 2019.
- [Luo14] D. Luo, W. Luo, R. Yang, J. Huang, „Identifying compression history of wave audio and its applications”, *ACM Trans. Multimedia Computing, Communications and Applications*, vol. **10**, nr. 3, pp. 1-19, 2014.
- [NIS11] NIST Multimodal Information Group, *2008 NIST Speaker Recognition Evaluation Test Set LDC2011S08*, <https://catalog.ldc.upenn.edu/LDC2011S08>, Philadelphia, PA (SUA), Linguistic Data Consortium, 2011.
- [Pop13a] **G. Pop**, D. Drăghicescu, D. Burileanu, „On Forensic Speaker Recognition Case Pre-Assessment”, *Proceedings of the 7th Conference on Speech Technology and Human-Computer Dialogue (SpeD)*, Cluj-Napoca, pp. 169-176, 2013.
- [Pop13b] **G. Pop**, *Recunoașterea vorbitorului în criminalistică*, Lucrare de disertație, programul de masterat BIOSINF („Tehnologii Multimedia în Aplicații de Biometrie și Securitatea Informației”), Universitatea POLITEHNICA din București, Facultatea de Electronică, Telecomunicații și Tehnologia Informației, Iun. 2013.
- [Pop14a] **G. Pop**, D. Drăghicescu, D. Burileanu, „A Quality-Aware Forensic Speaker Recognition System”, în *Romanian Journal of Information Science and Technology (ROMJIST)*, Editura Academiei Române, București, vol. **17**, nr. 2, pp. 134–149, 2014.
- [Pop14b] **G. Pop**, „Posibilități și limite ale expertizei criminalistice a vocii și a vorbirii”, *Revista de Criminologie, Criminalistică și Penologie – Studii*, vol. **2**, nr. 1-2, pp. 178-183, 2014.
- [Pop16] **G. Pop**, *Digital Forensics*, Comunicare științifică invitată (keynote speaker), The International Conference “Modern Forensics Techniques – Possibilities of Access Via International Mutual Legal Assistance Procedures”, Poiana Brașov, Nov. 16-18, 2016.
- [Pop17] **G. Pop**, D. Drăghicescu, D. Burileanu, H. Cucu, C. Burileanu, „Fast Method for ENF Database Build and Search”, *Proceedings of the 9-th International Conference on Speech Technology and Human-Computer Dialogue (SpeD)*, București, 6-9 Iul. 2017, pp. 1-6.
- [Pop18a] **G. Pop**, Ș. Mihalache, D. Burileanu, „Forensic Speaker Identification Using Speech Quality Data”, *Proceedings of the 12th International Conference on Communications (COMM)*, București, 14-16 Iun. 2018, pp. 509-512.
- [Pop18b] **G. Pop**, D. Burileanu, Ș. Mihalache, „An Evaluative Framework for ENF-based Authentication of Audio Recordings”, *Proceedings of the Romanian Academy, Series A – Mathematics, Physics, Technical Sciences, Information Science*, Ed. Academiei Române, București, vol. **19**, nr. 4, pp. 605-612, Dec. 2018.
- [Pop19a] **G. Pop**, Ș. Mihalache, D. Burileanu, „Forensic Recognition of Narrowband AMR Signals”, *Proceedings of the 11-th International Conference On Human-Computer Dialogue (SpeD)*, Timișoara, 18-20 Nov. 2019, pp. 1-6.
- [Pop19b] **G. Pop**, D. Burileanu, „Speech Enhancement for Forensic Purposes”, *UPB Scientific Bulletin, Series C*, Ed. Politehnica Press, București, vol. **81**, nr. 3, pp. 41-52, 2019.
- [Pop63] K. Popper, *Conjectures and Refutations: The growth of scientific knowledge*, Routledge and Keagan Paul, Londra (Marea Britanie), pp. 33–39, 1963.
- [Qia17] K. Qian, Y. Zhang, S. Chang, X. Yang, D. Florencio, M. Hasegawa-Johnson, „Speech Enhancement Using Bayesian WavNets”, *Proceedings of the 18th Annual Conference of the International Speech Communication Association (INTERSPEECH)*, Stokholm (Suedia), 20-24 Aug. 2017, pp. 2013-2017.
- [Sad13] S.O. Sadjadi, M. Slaney, L. Heck, *MSR Identity Toolbox: A MATLAB Toolbox for Speaker Recognition Research, v1.0*, în Microsoft Research Technical Report, 17 Oct. 2013.

- [Shi18] S.-H. Shin, W.-J. Jang, H.-W. Yun H. Park, „Encoding Detection and Bit Rate Classification of AMR-Coded Speech Based on Deep Neural Network”, *IEICE Transaction on Information and Systems*, vol. **E101 D**, nr. 1, Jan. 2018, pp. 269-272.
- [Xu15] Y. Xu, J. Du, L.-R. Dai, C.-H. Lee, „A Regression Approach to Speech Enhancement Based on Deep Neural Networks”, *IEEE/ACM Transactions on Audio, Speech and Language Processing*, vol. **23**, nr. 1, pp. 7-19, 2015.
- [Zha10] Y. Zhang, P. Markham, T. Xia, L. Chen, Y. Ye, Z. Wu, Z. Yuan, L. Wang, J. Bank, J. Burgett, R.W. Conners, Y. Liu, „Wide-area frequency monitoring network (FNET) architecture and applications”, *IEEE Transactions on Smart Grid*, vol. **1**, nr. 2, pp. 159-167, 2010.