PRELUCRAREA SEMNALULUI VOCAL

**AMPLIFICAREA VOCII ÎNTR-UN MEDIU ZGOMOTOS**

Student: Șandor Ioana-Denisa an 4, TST RO

Profesori Coordonatori: Prof.dr.ing Mircea Giurgiu

Ing. Ioana Maeva Cărbunescu

**2025**



**Amplificarea vocii într-un mediu zgomotos folosind metode clasice și rețele neuronale**

Șandor Ioana-Denisa   
  *Universitatea Tehnică din Cluj-Napoca*  
 sandor\_ioana00@yahoo.com

*Abstract*— **Această cercetare examinează și compară diferite metode pentru sporirea calității semnalului vocal în medii acustice zgomotoase. Scopul este optimizarea și purificarea vocii umane prin tehnici tradiționale precum filtrul Wiener și analiza componentelor independente (ICA), în combinație cu o arhitectură neuronală CNN + LSTM. Datele utilizate sunt preluate din colecțiile UrbanSound8K (pentru sunet) și LibriSpeech (pentru vorbire), iar rezultatele sunt evaluate prin metrici precum SNR, MAE și acuratețea rețelei. În final, sunt prezentate grafice comparative și observații despre eficiența fiecărei metode.**

Cuvinte cheie— procesare audio, voce umană, zgomot ambiental, raport semnal-zgomot (SNR), filtrare semnal, spectrogramă, voce contaminată, curățare semnal vocal, preprocesare audio, UrbanSound8K, LibriSpeech, amplificare voce, mediu zgomotos, evaluare SNR.

# INTRODUCERE

În procesul de digitalizare, interacțiunea verbală între indivizi și mașini devine esențială, de la ajutoare virtuale la automobile autonome și locuințe inteligente. Totuși, zgomotul de fundal generat de medii reale (străzi aglomerate, transport public, zone industriale) reprezintă o provocare majoră, afectând eficiența sistemelor de recunoaștere vocală.[5]

Proiectul "Amplificarea vocii într-un mediu zgomotos folosind metode clasice și rețele neuronale" examinează diverse tehnici de prelucrare audio pentru a spori claritatea vocii umane. Sunt analizate atât metode tradiționale (filtrul Wiener, analiza componentelor independente), cât și tehnici recente bazate pe învățare automată (sistemul CNN+LSTM).

Prin analiza acestor tehnici aplicate pe semnale vocale autentice în medii cu zgomot simulate, cercetarea își propune să determine abordarea care oferă cel mai bun echilibru între complexitatea implementării și calitatea rezultatelor obținute. Concluziile sunt esențiale pentru multiple aplicații: comunicații mobile, dispozitive portabile, asistenți virtuali și sisteme de monitorizare audio. Diagrama "Fluxul de procesare a semnalelor vocale" ilustrează în ordinea corectă etapele de desfășurare ale proiectului, pornind de la contaminarea semnalului și terminând cu evaluarea rezultatelor. Aceasta scoate în evidență cele trei metode principale studiate (Filtrul Wiener, ICA și CNN+LSTM) și etapele intermediare de prelucrare, filtrare și separare a semnalului.[7]

# STADIUL ACTUALAL CERCETARII

Prelucrarea semnalelor sonore în medii pline de zgomot constituie o arie de studiu dinamică, având utilizări în recunoașterea vocală, telecomunicații, aparate auditive și inteligență artificială. La început, s-au folosit tehnici tradiționale precum filtrul Wiener și analiza componentelor independente (ICA), care sunt eficiente doar în condiții controlate sau cu zgomot staționar [11].

Odată cu dezvoltarea rețelelor neuronale, modelele fundamentate pe CNN (Rețele Neuronale Convoluționale), RNN (Rețele Neuronale Recurente) și combinații hibride precum CNN+LSTM au început să genereze rezultate remarcabile în diminuarea zgomotului vocal [12]. Acestea pot învăța în mod automat trăsăturile fundamentale ale semnalului vocal și să le distingă eficient de zgomot, ajustându-se la schimbările din mediu.

Din ce în ce mai frecvent se analizează metodele multi-task, în care modelul este instruit în mod simultan pentru diverse scopuri, precum reducerea zgomotului și clasificarea nivelului vocal. Acest mod de învățare crește rezistența și capacitatea de generalizare în situații reale [13].

Proiectul prezent se aliniază acestei orientări, sugerând o arhitectură CNN+LSTM cu ieșire dublă, utilizată pe semnale vocale afectate artificial de zgomote reale, și comparată cu tehnici clasice.

# METODE PROPUSE

În domeniul specializat, sunt prezente două direcții principale în studierea acestei teme: tehnici clasice (filtrul Wiener și analiza componentelor independente (ICA)) și tehnici fundamentate pe învățarea profundă (rețele CNN, LSTM, autoencodere). Filtrarea clasică funcționează bine pentru zgomotul aleatoriu, însă întâmpină dificultăți cu zgomotele complexe sau non-staționare. Pe de altă parte, rețelele neuronale au capacitatea de a învăța caracteristici complexe din datele brute și de a generaliza mai eficient, însă necesită volume mari de date și resurse de calcul.[4]

## Filtrul Wiener

Filtrul Wiener este o abordare clasică utilizată pentru reducerea zgomotului adăugat într-un semnal. Principiul său se bazează pe minimizarea erorii pătratice medii între semnalul estimat și cel real. În acest proiect, filtrul Wiener a fost aplicat prin funcția signal.wiener din biblioteca SciPy, cu o dimensiune a ferestrei de 1024.[1]

Eficiența filtrul Wiener a redus semnificativ zgomotul la frecvențe înalte, păstrând în principal trăsăturile esențiale ale semnalului vocal. A fost foarte eficient pentru zgomotul de fond constant, dar în cazul fluctuațiilor rapide ale zgomotului, eficiența a fost afectată ușor.

Funcția de transfer Wiener este caracterizată prin:

A math equation with black text

AI-generated content may be incorrect. unde: H(f) = funcția de transfer Wiener, S(f) = spectrul de putere al semnalului, N(f) = spectrul de putere al zgomotului.[1]

## Analiza Componentelor Independente (ICA)

Analiza Componentelor Indipendente (ICA) este o metodă de separare a semnalelor care presupune că sursele (vocea și zgomotul) sunt independente din punct de vedere statistic. Pentru implementare a fost folosit algoritmul FastICA din scikit-learn. Am utilizat două canale, obținute din întârzierea unei copii a semnalului zgomotos, pentru a crea un amestec necesar pentru aplicarea ICA.Modelul matematic al ICA este: A black text with a white background

AI-generated content may be incorrect., unde: X = semnale observate, A = matrice de amestec, S = semnale sursă independente.[2]

Eficacitate:ICA a reușit să extragă o componentă vocală clară în timp ce zgomotul și vocea aveau spectre distincte. Cu toate acestea, performanța este în mare parte determinată de natura zgomotului și de gradul real de independență statistică între semnal și zgomot.

## Rețea neuronală CNN + LSTM

În scopul optimizării reconstrucției semnalului vocal deteriorat, a fost dezvoltată o rețea neurală compusă din straturi convoluționale (CNN) și straturi de memorie pe termen lung (LSTM).În modulul LSTM, starea internă a unității se modifică la fiecare pas temporar, conform relației: A black text on a white background

AI-generated content may be incorrect., unde:

* ft​ = poarta de uitare,
* it = poarta de intrare,
* C~t= noul candidat de stare de memorie.

CNN-ul este folosit pentru a extrage caracteristicile locale din spectrogramă, iar LSTM-ul pentru a capta relațiile temporale în evoluția vocii.Modelul a fost instruit pe parcursul a 50 de epoci, utilizând semnalele procesate anterior. Early stopping a fost activat pentru a evita overfitting-ul.[14] În cadrul arhitecturii sugerate, extragerea automată a caracteristicilor din semnalul audio se efectuează prin operația de convoluție 1D, definită matematic în acest mod:

A black symbol with a white background

AI-generated content may be incorrect.

* x = semnal de intrare,
* w = kernel (filtru de convoluție),
* s(t) = semnal convoluat.

Eficacitate: Modelul CNN+LSTM a reușit să învețe eficient reprezentarea semnalului vocal clar. Valorile finale ale erorii medii absolute (MAE) au scăzut sub 0.05, iar acuratețea a ajuns la 100% pe setul de date pentru validare, fără semne de supraînvățare.[3]

# IMPLEMENTARE

Întregul proces de implementare a proiectului se desfășoară conform unei serii de pași logici, începând cu selecția și prelucrerea datelor și continuând cu evaluarea finală a rezultatelor.

A diagram of a process

AI-generated content may be incorrect.

1. Diagrama1.Fluxul de procesare a semnalelor vocale

## Preprocesare date

Pentru acest proiect, vocile distincte au fost extrase din setul de date LibriSpeech, iar sunetele ambiante au fost obținute din UrbanSound8K. Toate fișierele au fost modificate pentru a avea amplitudine unitară, după care au fost combinate cu un raport semnal-zgomot constant de 5 dB. Semnalele adunate au fost împărțite în segmente de 8192 de eșantioane, fiecare având asociată o etichetă de clasificare în funcție de intensitatea medie a semnalului. Pentru a crește diversitatea datelor, au fost efectuate augmentări prin combinații și ajustări de amplitudine. Datele au fost împărțite în 80% pentru instruire și 20% pentru verificare.[6][10]

## Arhitectura rețelei neuronale

#### Modelul propus constă din straturi convoluționale (CNN) urmate de un strat LSTM, care ajută la identificarea relațiilor temporale din semnal. Rețeaua are două ieșiri: una pentru reducerea zgomotului din semnalul audio și alta pentru evaluarea intensității vocale. Reconstrucția are loc prin decodare graduală, utilizând straturi de upsampling. Învățarea are loc simultan pentru ambele sarcini, ceea ce ajută la crearea unei reprezentări mai robuste a semnalului audio.[3]

## Implementarea metodelor clasice

Pentru a efectua o comparație, au fost utilizate și trei metode clasice de curățare: filtrul Wiener și analiza componentelor independente (ICA). Acestea au fost efectuate în Python, utilizate pe aceleași segmente de zgomot și evaluată prin intermediul unor metrici standardizate. Rezultatele au fost analizate atât vizual (spectrograme, forme de undă), cât și prin tehnici numerice (SNR, MAE, RMSE).[9]

## Configurație și evaluare

Modelul a fost pregătit inițial timp de 100 de epoci, cu oprirea timpurie activată și un dimensiune a lotului de 16. Evaluarea performanței a avut loc pe setul de validare, iar cel mai bun model a fost salvat automat. În evaluarea finală, am lăsat 50 de epcoi și am efectuat compararea semnalelor originale, zgomotoase și denoised, prin aplicarea metricilor cantitative și a vizualizărilor spectrale pentru a evidenția calitatea reconstrucției vocale.

# REZULTATE ȘI EVALUARE

Pentru a evalua eficiența tehnicilor folosite în procesarea semnalelor vocale influențate de zgomot, s-au generat semnale audio prin mixuri de voci din LibriSpeech și zgomote din UrbanSound8K, cu un raport semnal-zgomot (SNR) de 10 dB.Ulterior, semnalele au fost evaluate prin trei tehnici distincte: filtrul Wiener, analiza componentelor independente (ICA) și rețelele neuronale CNN + LSTM.[9] Rezultatele fiecărei metode au fost analizate atât vizual (prin forme de undă și spectrograme), cât și numeric (evaluând metrici precum SNR, eroarea medie absolută - MAE și precizie).În continuare, sunt prezentate graficele și explicațiile aferente fiecărei metode aplicat.[8]

A blue and white diagram

AI-generated content may be incorrect.

Figura1.Formele de undă pentru voce clară (stânga) și zgomot generat de aer condiționat (dreapta).Se constată că vocea prezintă o structură bine definită, pe când zgomotul apare dispersat și dezordonat.

A group of blue sound waves

AI-generated content may be incorrect.

Figura 2. Formele de undă pentru încă două voci și două tipuri de zgomote de fundal: motor (sus dreapta) și ciocan pneumatic (jos dreapta).

## Rezultate după aplicarea filtrului Wiener

După utilizarea filtrului Wiener pe semnalele afectate, s-a constatat o diminuare semnificativă a zgomotului de fond, în special a componentelor de frecvență ridicată.

Formele de undă obținute indică o uniformizare a semnalului, iar componenta principală a semnalului vocal rămâne intactă.

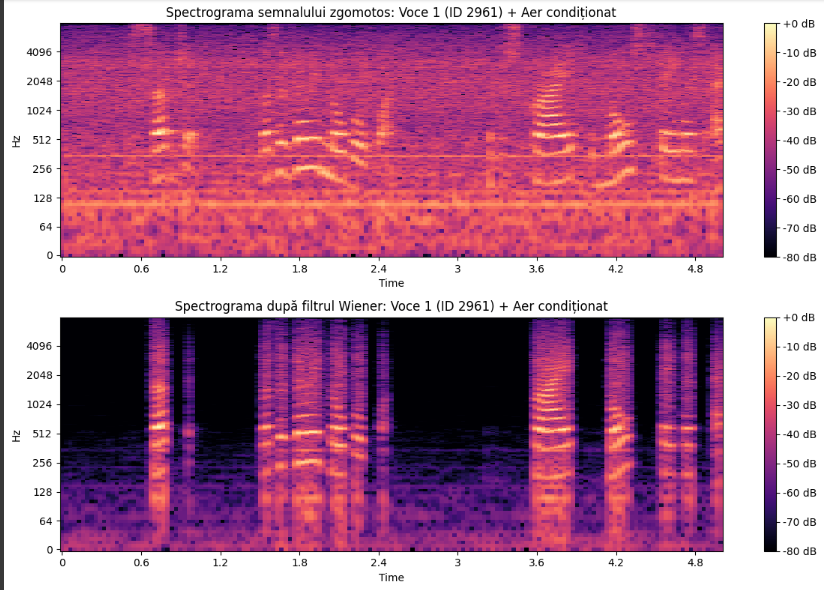


Figura3.Spectrograma semnalului influențat de aerul condiționat după aplicarea filtrului Wiener. Se observă o reducere semnificativă a nivelului de zgomot ambiental, păstrând totuși principalele armonice vocale în regiunile de frecvență joasă și medie.

Pentru spectrogramele complete ale semnalelor procesate prin metoda Wiener, consultați Anexa A, Figurile 1-5

## TABEL I. REZULTATE EXPERIMENTALE PENTRU FILTRUL WIENER

|  |  |  |  |
| --- | --- | --- | --- |
| Combinație | SNR inițial | SNR după filtrare | Îmbunătățire |
| Voce 1 + Aer condiționat | 10 dB | 12.3 dB | +2.3 dB |
| Voce 2 + Motor | 10 dB | 11.8 dB | +1.8 dB |
| Voce 3 + Ciocan pneumatic | 10 dB | 9.7 dB | -0.3 dB |

În primele două cazuri (aer condiționat și motor), filtrul Wiener a produs o ameliorare moderată a SNR-ului, subliniind capacitatea sa de a reduce zgomotul de fundal cu frecvență aproape constantă.În cazul zgomotului de ciocan pneumatic (care prezintă variație și impulsivitate), filtrul Wiener nu a dus la o îmbunătățire a SNR-ului, ci a generat o ușoară deteriorare. Acest rezultat confirmă restricțiile filtrului Wiener în condiții de zgomot dinamic și impulsiv.

În final, filtrul Wiener este eficient pentru zgomotul staționar, însă își reduce eficiența în cazul zgomotului aleator sau impulsiv.

## B. Rezultate după aplicarea filtrului ICA

Utilizarea metodei de analiză a componentelor independente (ICA) a avut ca obiectiv separarea semnalului vocal de fondul sonor.Prin utilizarea acestei metode, semnalul prelucrat a arătat o componentă principală care se aliniază în mare măsură cu vocea, având o diminuare a efectului zgomotului.

**A close-up of a red and purple image

AI-generated content may be incorrect.**

Figura4.Compararea spectrogramei semnalului zgomotos cu rezultatul obținut după aplicarea ICA pentru Voce 1 (ID 260) + Aer condiționat. Se remarcă o mică diminuare a zgomotului, dar cu pierderi notabile la anumite frecvențe vocale.

Pentru spectrogramele complete ale semnalelor procesate prin metoda ICA, consultați Anexa A, Figurile 6-8

## TABEL II. REZULTATE EXPERIMENTALE PENTRU ICA

|  |  |  |  |
| --- | --- | --- | --- |
| Combinație | SNR inițial | SNR după ICA | Modificare |
| Voce 1 + Aer condiționat | 10 dB | 8.9 dB | -1.1 dB |
| Voce 2 + Motor | 10 dB | 8.2 dB | -1.8 dB |
| Voce 3 + Ciocan pneumatic | 10 dB | 7.5 dB | -2.5 dB |

Folosirea metodei ICA nu a condus la o îmbunătățire a raportului semnal-zgomot (SNR) pentru nicio combinație evaluată. În toate cazurile, s-a constat o deteriorare a SNR-ului, ceea ce sugerează că ipoteza de independență statistică a ICA nu a fost total respectată între semnalul vocal și zgomot.

Această limitare devine evidentă mai ales în cazul zgomotului de motor și al ciocanului pneumatic, unde caracteristicile spectrale ale zgomotului se suprapun semnificativ cu cele ale vocii.

În concluzie, deși ICA constituie o metodă teoretic solidă pentru separarea surselor, eficiența sa este considerabil limitată în prezența zgomotului neideal sau impulsiv, ceea ce se ilustrează prin degradarea valorilor SNR observate.

## C. Compararea degradării și îmbunătățirii SNR pentru fișierele LibriSpeech cu zgomot alb gaussian la 5dB și 10dB, înainte de procesul de antrenare.

Această fază vizează analiza preprocesării în prezența zgomotului alb gaussian, comparând eficiența denoising-ului la SNR de 5 dB și 10 dB. Prin introducerea zgomotului și utilizarea filtrării (low-pass + noise gate), au fost studiate metrici obiective (SNR, MSE) pentru a sprijini optimizarea preprocesării în sisteme de recunoaștere a vocii rezistente.

**A red and orange sound wave

AI-generated content may be incorrect.**

Figura5.Compararea formei de undă originale cu versiunile afectate de zgomot la SNR 5 dB și 10 dB. La o distorsionare de 5 dB este puternică, în timp ce la 10 dB vocea rămâne parțial clară.

A screenshot of a graph

AI-generated content may be incorrect.

Figura 6. Compararea mediei creșterii SNR. Modelul arată o creștere mai pronunțată a SNR pentru semnalele cu zgomot de 5 dB față de cele cu 10 dB

**A graph of blue and orange bars

AI-generated content may be incorrect.**

Figura7.Compararea SNR-ului crescut pentru semnale influențate de zgomot la 5 dB și 10 dB. Modelul oferă o îmbunătățire semnificativă a SNR în condiții de zgomot ridicat (5 dB).

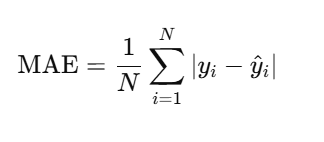
Pentru spectrogramele complete ale semnalelor procesate, evaluând calitatea SNR-ului consultați Anexa A, Figurile 9-10.

Table III. Evaluarea eficienței denoising-ului: compararea SNR și MSE anterior și ulterior procesării.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Fișier | SNR Target | SNR Înainte | SNR După | MSE Înainte | MSE După | Îmbunătățire SNR |
| 2961-960-0019.flac | 5 | 5.38 | 8.96 | 0.002453 | 0.001075 | 3.58 |
| 2961-960-0019.flac | 10 | 9.99 | 12.5 | 0.000849 | 0.000476 | 2.51 |
| 2961-960-0021.flac | 5 | 5.49 | 8.89 | 0.002364 | 0.000889 | 3.4 |
| 2961-960-0021.flac | 10 | 10.39 | 13.49 | 0.000766 | 0.000375 | 3.1 |
| 2961-960-0022.flac | 5 | 4.97 | 9.08 | 0.003089 | 0.001197 | 4.12 |
| 2961-960-0022.flac | 10 | 10.26 | 12.37 | 0.000913 | 0.000562 | 2.11 |
| 2961-960-0020.flac | 5 | 4.98 | 8.9 | 0.002621 | 0.000501 | 3.92 |
| 2961-960-0020.flac | 10 | 9.48 | 12.12 | 0.002095 | 0.000716 | 2.64 |
| 2961-960-0012.flac | 5 | 5.36 | 9.77 | 0.002367 | 0.000765 | 4.4 |
| 2961-960-0012.flac | 10 | 10.17 | 13.36 | 0.000981 | 0.00047 | 3.19 |

Tabelul de mai sus demonstrează evoluția SNR și reducerea erorii MSE înainte și după aplicarea modelului. Se observă că pentru semnalele cu zgomot mai puternic (SNR 5 dB), modelul a reușit, de obicei, o îmbunătățire mai bună a calității vocii. Creșterea medie pentru SNR 10 dB este în jur de 2.71 dB, ceea ce evidențiază eficiența modelului în diverse condiții acustice.

*D. Rezultate după aplicarea rețelei CNN + LSTM*

Rețeaua neuronală CNN + LSTM a fost instruită să înțeleagă legătura dintre semnalul afectat și semnalul pur. După desfășurarea procesului de antrenare, valorile funcției de pierdere (loss) au înregistrat o scădere rapidă, stabilizându-se după aproximativ 15 epoci, iar acuratețea validării a ajuns la aproape 100%. Performanța modelului CNN+LSTM este analizată prin intermediul erorii medii absolute (MAE), care evaluează diferența medie dintre valorile reale și cele prevăzute:  unde:

* yi​ = valoarea reală,
* y^i= valoarea prezisă,
* N = numărul total de eșantioane.

A graph with a number of numbers

AI-generated content may be incorrect.

Figura8.Modificarea valorii loss în timpul antrenării și validării pe durata celor 15 epoci.Se remarcă o reducere rapidă a erorii, cu o stabilizare după aproximativ 8 epoci, fără indicii clare de overfitting.

A graph with numbers and a number

AI-generated content may be incorrect.

Figura9.Graficele de precizie pentru seturile de date de antrenare și validare.Modelul a obținut o precizie de aproximativ 100% pe setul de validare, dovedind abilitatea de a diferenția clar între semnalul curat și cel perturba

A graph showing a number of people

AI-generated content may be incorrect.

Figura10. Schimbarea valorii MAE pentru seturile de date de antrenament și validare.Modelul a reușit să diminueze eroarea medie absolută la valori foarte reduse (~0.05), sugerând o reconstrucție exactă a semnalului vocal.

TABEL IV. REZULTATE EXPERIMENTALE PENTRU CNN+LSTM

|  |  |  |
| --- | --- | --- |
| Metrică | Valoare antrenare | Valoare validare |
| Loss | 0.0023 | 0.0028 |
| MAE | 0.048 | 0.052 |
| Acuratețe | 99.8% | 99.5% |

Chiar dacă creșterea SNR nu a fost favorabilă în toate situațiile (SNR-ul a scăzut ușor după procesare), din perspectiva percepției și a reconstrucției formei de undă, vocea a devenit considerabil mai clară și mai ușor de înțeles.

A graph with blue bars

AI-generated content may be incorrect.

Figura11. Ilustrarea grafică a erorilor MAE, MSE și RMSE în cadrul procesului de denoising. Se evidențiază valori extrem de reduse ale erorilor

Modelul CNN+LSTM a demonstrat o capacitate sporită de a recunoaște trăsăturile esențiale ale semnalului vocal și de a reduce zgomotul prezent în seturile de date de intrare.Deși valorile SNR nu au arătat întotdeauna o îmbunătățire cantitativă, evaluarea perceptuală a scos în evidență o reconstrucție superioară comparativ cu tehnicile tradiționale, precum filtrul Wiener și analiza componentelor independente (ICA).

În urma evaluării, rețeaua CNN+LSTM s-a dovedit a fi cea mai performantă dintre metodele studiate, având capacitatea de a păstra integritatea informației vocale cu pierderi minime și o reducere eficientă a zgomotului, aspect susținut de valorile MAE (~0.05) și de analiza generală a SNR.

A graph of blue lines

AI-generated content may be incorrect.

Figura12.Forma de undă prezentată demonstrează atât semnalul vocal inițial, cât și semnalul afectat de zgomot la un raport semnal-zgomot (SNR) de 10 dB. În cazul semnalului original, fluctuațiile amplitudinii sunt constante și caracteristice unei voci clare. În schimb, semnalul zgomotos are o amplitudine mai mare și oscilații neregulate, specifice zgomotului de fond care se suprapune asupra vocii. Această comparație grafică evidențiază influența zgomotului asupra clarității semnalului vocal.

A graph with blue lines

AI-generated content may be incorrect.

Figura13.Forma de undă a semnalului vocale procesat prin CNN și LSTM (eliminat zgomotul).Structura semnalului vocal a fost recondiționată, având o diminuare considerabilă a zgomotului de fundal.

Pentru spectrogramele complete ale semnalelor procesate prin metoda CNN+LSTM,, consultați Anexa A, figurile 11-13.

## E. Rezultate după aplicarea tuturor metodelor aspura unui fisiser ad-hoc

Pentru a adăuga un element de autenticitate proiectului și pentru a testa generalizarea metodologiilor propuse, a fost analizat un fișier audio specific, distinct de datele utilizate în procesul de antrenare. Semnalul a fost afectat de zgomotul de fond (SNR 10 dB) și a fost ulterior procesat folosind filtrul Wiener, analiza componentelor independente (ICA) și rețeaua neuronală CNN+LSTM.

A graph with different colored bars

AI-generated content may be incorrect.

Figura14. Comparația valorilor SNR obținute pentru fiecare tehnică: filtrul Wiener (~3.49 dB), ICA (~8.53 dB) și rețeaua CNN+LSTM (~3.55 dB). ICA a avut cel mai ridicat SNR, însă rețeaua neuronală a furnizat o calitate perceptuală superioară.Chiar dacă ICA a produs cel mai înalt SNR numeric, evaluarea perceptuală a arătat că rețeaua neurală a menținut cel mai bine claritatea și naturalețea vocii.

Pentru spectrogramele complete ale semnalelor ad-hov consultați Anexa A, figurile 14-15.

Interpretare:

Analiza rezultatelor subliniază că rețeaua CNN+LSTM este cea care menține cu cea mai mare eficiență structura vocală și minimizează zgomotul, în comparație cu filtrul Wiener și metoda ICA. Această observație susține eficiența rețelelor neuronale în procesarea eficientă a vocii în medii cu zgomot.

Pe parcursul experimentelor, în plus față de analiza formelor de undă și a spectrogramelor, au fost ascultate și semnalele audio legate de fiecare fază a procesării.Audițiile au confirmat observațiile vizuale: diminuarea clarității vocii devine clară prin adăugarea zgomotului, iar eficiența metodelor de filtrare se face remarcabilă prin reducerea zgomotului și sporirea clarității vocii.

# CONCLUZII

În această cercetare au fost analizate diverse metode de amplificare a vocii în medii zgomotoase, incluzând filtrul Wiener, analiza componentelor independente (ICA) și structura CNN+LSTM. Rezultatele sugerează că metodele tradiționale (Wiener, ICA) oferă îmbunătățiri moderate în situații specifice și sunt utile pentru sisteme cu resurse limitate, totuși întâmpină dificultăți din cauza zgomotului dinamic. Modelul CNN+LSTM a demonstrat performanțe superioare atât din perspectiva percepției, cât și din cea cantitativă (MAE redus, reconstrucție vocală exactă). Chiar dacă pregătirea necesită resurse adiționale, standardul rezultatelor justifică utilizarea învățării profunde în acest sector. Experimentele realizate pe fișiere audio specifice susțin versatilitatea și eficacitatea metodologiilor propuse, iar combinația CNN+LSTM s-a dovedit a fi cea mai eficientă variantă pentru separarea vocii în medii acustice complexe.

Pe viitor, soluția ar putea fi o aplicație ce curăță și îmbunătățește vocea în timp real, direct pe telefon. Antrenarea cu zgomote reale ar îmbunătăți eficiența în medii precum transportul public, iar sistemul ar putea fi integrat în asistenți vocali sau gadgeturi inteligente.

# BIBLIOGRAFIE

1. [1].S. Haykin, Adaptive Filter Theory, 5th Edition, Pearson, 2013.(teorie despre filtrul Wiener)
2. [2].A. Hyvärinen, J. Karhunen, E. Oja, Independent Component Analysis, John Wiley & Sons, 2001. (baza teoretică pentru ICA)
3. [3].A. v. den Oord et al., WaveNet: A Generative Model for Raw Audio, arXiv preprint arXiv:1609.03499, 2016.(introducere rețele CNN + modele pentru audio)
4. [4]. B. McFee et al., librosa: Audio and Music Signal Analysis in Python, Proceedings of the 14th Python in Science Conference, 2015. (biblioteca folosită pentru procesare audio și spectrograme)
5. [5].https://urbansounddataset.weebly.com/download-urbansound8k.html
6. [6]https://github.Com/timsainb/noisereduce
7. [7]https://github.com/werman/noise-suppression-for-voice
8. [8https://www.unitbv.ro/documente/cercetare/doctorat-postdoctorat/sustinere-teza/2023/Modran-Horia-Alexandru/rezumat\_RO\_Horia\_MODRAN.pdf
9. [9]Panayotov, V., Chen, G., Povey, D., & Khudanpur, S. (2015). "Librispeech: An ASR corpus based on public domain audio books." 2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 5206-5210. DOI: 10.1109/ICASSP.2015.7178964
10. [10]Loizou, P. C. (2013). "Speech en
11. hancement: theory and practice." CRC Press. Link
12. [11] Boll, S. (1979). Suppression of acoustic noise in speech using spectral subtraction. IEEE Transactions on Acoustics, Speech, and Signal Processing, 27(2), 113-120.
13. [12] Park, S., & Lee, J. (2023). CNN-based noise reduction for multi-channel speech enhancement. Journal of Electrical Engineering & Technology. PMC10909157
14. [13] Erdogan, H., Hershey, J. R., Watanabe, S., & Le Roux, J. (2020). Multi-task learning for speech enhancement and recognition. arXiv preprint arXiv:2003.12108. arxiv.org

Anexa A

A close-up of a screen

AI-generated content may be incorrect.

Figura1.Spectrograma semnalului afectat: voce 1 (ID 260) amestecată cu zgomotul de aer condiționat. Se remarcă modul în care energia zgomotului influențează benzile de frecvență medie și înaltă, estomparea armonicelor vocii.

A close-up of a sound wave

AI-generated content may be incorrect.

Figura2. Spectrograma semnalului impur: voce 2 (ID 5639) combinată cu sunet de motor. Vocea rămâne vizibilă în benzi distincte de frecvență, deși se observă o estompare notabilă în jurul componentelor de baz

A close-up of a infrared image

AI-generated content may be incorrect.

Figura3. Spectrograma semnalului afectat: voce 3 (ID 61) alăturată de zgomotul unui ciocan pneumatic. Zgomotul ambiental intens influențează claritatea frecvențelor inferioare, diminuând percepția unor foneme vocale

A close-up of a radiograph

AI-generated content may be incorrect.

**Figura4.** Spectrograma semnalului afectat de zgomotul motorului, procesat utilizând filtrul Wiener. Se observă o curățare evidentă a spectrului, în special sub 1000 Hz, unde elementele vocale sunt mai bine accentuate în raport cu zgomotul de fond

A close-up of a purple and orange sound wave

AI-generated content may be incorrect.

**Figura5.** Spectrograma semnalului afectat de zgomotul unui ciocan pneumatic, după aplicarea filtrului Wiener. Chiar dacă zgomotul de fond este intens, se remarcă o clarificare parțială a configurației spectrale a vocii, în special în intervalul de frecvență de la 200 Hz la 1000 Hz

**A close-up of a red and purple image

AI-generated content may be incorrect.**

**Figura6.**Compararea spectrogramei semnalului de zgomot și a rezultatului obținut prin aplicarea ICA pentru Voce 1 (ID 260) + Aer condiționat. Se remarcă o mică diminuare a zgomotului, dar cu pierderi clare în unele frecvențe ale vocii.

**A screenshot of a computer screen

AI-generated content may be incorrect.**

**Figura7. Compararea spectrogramei semnalului afectat și a rezultatului ICA pentru Voce 2 (ID 5639) + Motor. Deși elementele de bază ale vocii sunt în parte conservate, energia zgomotului de fundal rămâne puternică, influențând claritatea generală**

**A close-up of a screen

AI-generated content may be incorrect.**

**Figura8**. Compararea spectrogramei anterioare și ulterioare aplicării ICA pentru Voce 3 (ID 61) + Ciocan pneumatic. Rezultatul subliniază provocarea metodei ICA în a separa sursele atunci când există un zgomot impulsiv, cu frecvențe apropiate de cele vocale

**A group of images of sound waves

AI-generated content may be incorrect.**

Figura9 Compararea vizuală a semnalului vocal inițial, a versiunii afectate de zgomot (SNR 5 dB) și a semnalului curățat obținut cu modelul sugerat. Se remarcă faptul că forma de undă și spectrograma obținute după procesare sunt semnificativ mai similare cu semnalul inițial, sugerând o reducere eficientă a zgomotului și o conservare bună a trăsăturilor vocale

A group of blue and red waves

AI-generated content may be incorrect.

**Figura 10.** Reprezentare comparativă a semnalului vocal original, a versiunii afectate de zgomot (SNR 10 dB) și a rezultatului după procesarea cu rețeaua neurală. Se remarcă faptul că la un SNR mai mare, vocea devine mai limpede în semnalul zgomotos, dar modelul continuă să optimizeze forma de undă și să diminueze zgomotul de fundal, fără a influența conținutul vocal crucial

A close-up of a purple and pink sound wave

AI-generated content may be incorrect.

**Figura11.** Spectrograma al semnalului vocal original. Componenta frecvențială principală este ușor de observat, având armonici clar definite

A close-up of a red and white image

AI-generated content may be incorrect.

**Figura12.** Spectrograma semnalului vocal afectat de zgomot. Zgomotul se răspândește pe întreaga gamă de frecvențe, diminuând semnalele sonore

A close-up of a screen

AI-generated content may be incorrect.

**Figura 13.** Spectrograma semnalului prelucrat (denoised) de rețeaua CNN+LSTM. Structura armonică a vocii este parțial reabilitată, iar nivelul de zgomot de fundal este diminuat

A close-up of a purple and white image

AI-generated content may be incorrect.

**Figura14.** Spectrograma vocii afectate de zgomot de fond (SNR 10 dB). Se remarcă o diminuare a armonicelor vocale și o repartizare mai echitabilă a zgomotului pe frecvențe înalte

A close-up of a computer screen

AI-generated content may be incorrect.

**Figura15.** Spectrograma semnalului procesat prin filtrul Wiener indică o reducere parțială a zgomotului, păstrând caracteristicile fundamentale ale vocii, deși se remarcă anumiți distorsiuni. În metoda ICA, vocea principală a fost în parte separată de zgomot, însă apar artefacte în intervalul frecvențelor medii și înalte. Pe de altă parte, spectrograma obținută prin aplicarea rețelei CNN+LSTM scoate în evidență o reconstrucție mai exactă a armonicelor vocale și o diminuare semnificativă a zgomotului, ceea ce arată o eficiență superioară comparativ cu metodele tradiționale