Universitatea Politehnica București

Facultatea de Automatică și Calculatoare

Proiect – Procesarea Semnalelor

Comparison of Speech Enhancement Algorithms - MBSS si PSS

Îndrumător științific: Dr. ing. Marios Choudary

Student: Mărgineanu Nicolae-Vlăduț

Seria și grupa: 341 C2

BUCUREŞTI

2020-2021

Cuprins

1. Introducere

In acest proiect am realizat o comparatie intre metodele Multi-Band Spectral Subtraction Method si Power Spectral Subtraction Method.

În lumea reală, zgomotul este în mare parte colorat (**power spectrum of a noise signal**), ceea ce afectează semnalul de vorbire diferit pe tot spectrul. Puține frecvențe sunt mai afectate decât altele, în funcție de caracteristicile spectrale ale zgomotului. Cele doua metode sunt folosite pentru îmbunătățirea vorbirii cu zgomot rezidual minimizat. Semnalul de vorbire degradat în prezența zgomotului de la sol este denumit ca vorbire zgomotoasă, care poate fi modelată ca model de zgomot aditiv, cu presupunerea de bază că atât semnalul de vorbire cât și cel de zgomot sunt necorelate**. Semnalul zgomotos poate fi modelat ca suma semnalului de vorbire curat și zgomotul:**

$$y(n) = x(n) + v(n),\ n \in (0,\ N-1);$$

Unde: y(n) – semnalul zgomotos, x(n) – semnalul de vorbire curat si v(n) – zgomotul.

In realizarea algoritmilor, trebuie sa tinem cont de urmatoarele notiuni teoretice:

1. Vorbirea și alte semnale audio sunt semnale variabile în timp. Dacă luăm spectrul pe întregul semnal, atunci obținem **spectrul mediu**, dar nu putem vedea modificări ale frecvențelor fundamentale.
2. În aplicațiile în timp real, trebuie **să împărțim semnalul în segmente** astfel încât să nu trebuiasca să așteptăm finalizarea integrității înainte de a începe procesarea. Dacă luăm spectrul din segmente mici (**windows**) apropiate unul de altul, putem observa apoi **evoluția spectrală a semnalului**. O astfel de reprezentare este cunoscută sub numele de **spectrogramă** a unui semnal.

3. Când spectrograma este calculată utilizând **windowing** și **transformata Fourier discretă** se numește **transformată Fourier de scurtă durată** (STFT). Acum STFT al semnalului zgomotos este reprezentat de: **Y(K) = X(K) + V(K) .**

2. Related Work – algoritmii folositi in Speech Enhancement

Power Spectral Substracion PSS - are la baza scăderea zgomotului (power noise spectrum) din semnalul de vorbire. Factorul de "over-substraction" - α este utilizat pentru a îmbunătăți rezultatul - α depinde de **SNR (signal to noise ratio)** segmentat. Metoda PSS este implementată **cu o singură bandă.**

Multi-Band Spectral Subtraction Method MBSS – este similar cu algoritmul PSS, in care se foloseste factorul de "**over-substraction"** – α. Una dintre diferentele dintre metode este ca aceasta metoda foloseste un **coeficient nou, δ band substraction** – care depinde de frecventa pentru fiecare banda. Metoda este implementata **pe mai multe benzi.**

3. Our Work – prezentarea algoritmilor PSS si MBSS (implementare)

A. Preprocesarea datelor de intrare

Inainte de aplicarea unei metode de Speech Enhancement, realizez urmatoarele operatii pe semnalul de intrare:

1. Se divizeaza vectorul audio in **frame**-uri **suprapuse** (50% overlapped frames)
2. Se aplica **Hanning window** (20 ms window) pe fiecare frame dupa care se salveaza fiecare frame intr-o matrice. Aplicam Hanning Window pentru a reduce artefactele de la marginile ferestrei.
3. Pentru fiecare frame din matrice, se aplica **functia FFT** pentru analiza in domeniul frecventa.

4. **Weighted Spectral Average [5]** – aceasta metoda ne ajuta sa netezim variatia zgomotului.

   Se realizeaza cu urmatoarea formula**:**

$$\bar{Y}_j(k) = \sum_{l=-M}^{M} W_l Y_{j-l}(k)$$

- unde j - reprezinta indexul frame-ului respectiv, M - numarul de frame-uri, W - filtrul weights, iar Y(k) semnalul zgomotos de intrare (obtinut prin FFT) (spectrul de magnitudine)

  Filtrul Weights este ales ca un vector de urmatoarele elemente care a fost determinat empiric si setat ca: (valorile sunt preluate din articolul stiintific [5]).

  W = [0.09, 0.25, 0.32, 0.25, 0.09]

  Media a fost limitata la 5 frame-uri succesive, dat de l din formula de mai sus, pentru a prevenii alterarea spectrului. Astfel, -2 <= l <= 2.

B. Folosirea unei metode in procesarea semnalului de intrare

a) Multi-Band Spectral Subtraction Method

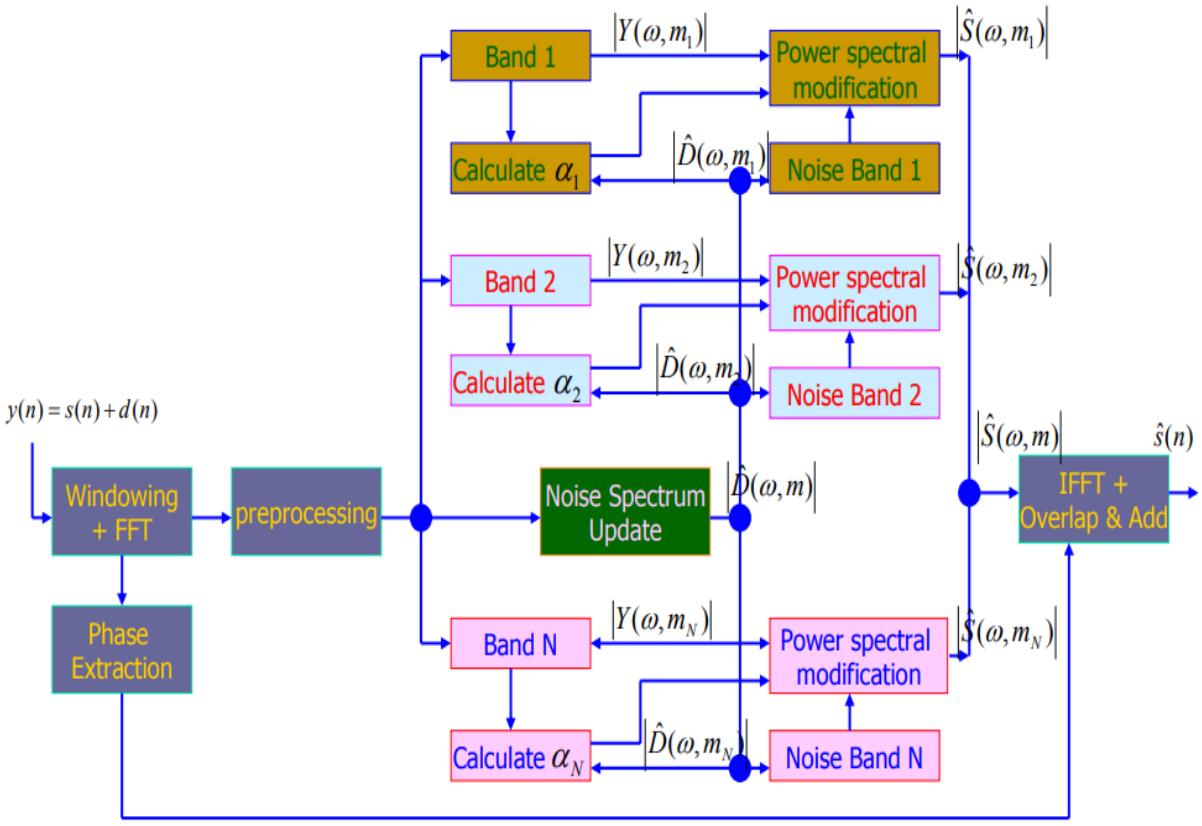
Aceasta metoda MBSS foloseste "**Linear Frequency Spacing**" [1], prin care trecem semnalul de intarare din timp, in frecvena, cu functia FFT, rezultand spectrul semnalului. FFT pentru semnalul de intrare, este dat de formula:

$$\mathrm{X(K)} = \sum_{n=0}^{N-1} x(n)\, e^{-jwnK}$$

X(K) – spectrul de magnitudine

|X(K)|^2 – power spectrum

Se trece prin urmatoarele etape in implementare:



1. În prima etapă, semnalul este "**windowed**" și **spectrul de magnitudine** este estimat folosind **functia FFT** (preprocesarea datelor de intrare).
2. În a doua etapa, spectrele de zgomot și vorbire sunt împărțite în **benzi de frecvență** diferite și se calculează **factorul de "over-substraction" – α (alfa)** pentru fiecare bandă.

3. A treia etapă include etapa de „„**band substraction**”” **(factorul δ delta)** în **benzi de frecvență** individuale **prin reajustarea factorului de “over-substraction”.**
4. În cele din urmă, **benzile de frecvență** modificate sunt îmbinate (**merged**) și semnalul in domeniul de timp este reconstruit cu ajutorul metodei de „**overlap-add**” și aplicand **functia IFFT**.

Prin urmare, estimarea spectrului de vorbire (clean speech spectrum) in banda a j - a este obtinuta prin urmatorul **algoritm**:

$$\left|\widehat{X_j}(\mathrm{K})\right|^2 = \begin{cases} \left|Y_j(K)\right|^2 - \alpha_\mathrm{j} \cdot \delta_\mathrm{j} \cdot \left|\widehat{V_j}(\mathrm{K})\right|^2, \\ \text{if } \left|\widehat{X_j}(\mathrm{K})\right|^2 > \beta \cdot \left|Y_j(\mathrm{K})\right|^2 \text{ else} \\ \beta \,.\left|Y_j(\mathrm{K})\right|^2 \end{cases}$$
$$\mathrm{K}_j < K < \mathrm{K}_{j+1}$$

Unde:

- Y(K) – spectrul de magnitudine al semnalului zgomotos (obtinut prin FFT, trecut din timp in frecventa)
- X^(k) – spectrul de magnitudine **estimat** al vorbirii curate
- V^(n) – spectrul de magnitudine **estimat** al zgomotului calculat de-a lungul unei perioade de liniste
- |X^(K)|^2 – patratul din valoarea absoluta (abs) din spectrul de magnitudine. Se mai numeste si **power spectrum**

Capetele de frecvență de început și sfârșit ale benzii de frecvență j sunt date de **K*j* și K*j* + 1**. Banda specifică cu factorul de „**over-substraction**” este dată de αj în funcție de SNR-ul segmentat al benzii corespunzătoare. **SNR (signal to noise ratio)** segmentat al benzii j poate fi calculat astfel:

$$SNR_j(db) = 10log_{10}\left(\frac{\sum_{K=K_j}^{K_{j+1}}|Y_j(\mathrm{K})|^2}{\sum_{K=K_j}^{K_{j+1}}|\bar{V}_j(\mathrm{K})|^2}\right)$$

Factorul α poate fi calculat dupa formula urmatoare:

$$\alpha_j = \begin{cases} \alpha_{max, if SNR_j \leq SNR_{min}} \\ \alpha_{max} + (SNR_j - SNR_{min})\left(\frac{\alpha_{min} - \alpha_{max}}{SNR_{max} - SNR_{min}}\right) \quad if \quad SNR_{min} \leq SNR_{max} \\ \alpha_{min}, if SNR_j \geq SNR_{max}, \end{cases}$$

In cazul meu, factorul αmin = 1, αmax = 5, SNR min = −5dB, SNR max = 20 dB. [5]

$$\alpha_i = \begin{cases} 4.75, & SNR_i < -5 \\ 4 - \frac{3}{20}(SNR_i), & -5 \leq SNR_i \leq 20 \\ 1, & SNR_i > 20 \end{cases}$$

Factorul δj este un factor suplimentar de „**band substraction**”care oferă un grad suplimentar de control în cadrul fiecărei benzi. Valoarea utilizată în fiecare bandă este calculată empiric deoarece cea mai mare parte a energiei vorbirii este concentrată sub 1 kHz. (fiecare banda are cate un factor suplimentar)

$$\delta_j = \begin{cases} 1 & f_j \leq 1kHz \\ 2.5 & 1kHz < f_j \leq \frac{F_s}{2} - 2\ kHz \\ 1.5 & f_j > \frac{F_s}{2} - 2\ kHz \end{cases}$$

Primul frame corespunde tăcerii inițiale din speech. Se estimează pragul zgomotului. Aici, pentru estimarea exactă a zgomotului, am calculat în medie zgomotul atunci când a fost detectat cu zgomotul estimat anterior.

Valorile negative ale spectrului estimat sunt eliminate folosind parametrul „**spectral floor" β**, dupa formula:

$$|\widehat{S}_i(k)|^2 = \begin{cases} |\widehat{S}_i(k)|^2 & |\widehat{S}_i(k)|^2 > \beta|Y_i(k)|^2 \\ \beta|Y_i(k)|^2 & else \end{cases}$$

Parametrul **spectral floor** a fost setat la **β = 0.02.**

Pentru a obține o **distorsiune minimă** a vorbirii în regiunile cu **frecvență joasă** este de preferat sa folosim **valori mai mici ale δj** si **valori mai mari ale δj** în regiunile cu **frecvență înaltă**. Prin reducerea varianței de frecvență al vorbirii, zgomotul rezidual poate fi redus în vorbirea îmbunătățită. Prin urmare, în loc să se utilizeze spectrele de putere ale semnalului, se poate utiliza **o versiune netedă a spectrelor de putere**. Valoarea medie a magnitudinii ajută la îmbunătățirea calității vorbirii procesate.

b) Power Spectral Subtraction Method

Descriere succinta a metodei

- Etapele acestei metode sunt la fel ca in metoda MBSS, doar ca metoda este implementata **pe o singura banda**.
- Valoarea factorului αi „over-substraction" este calculat pe fiecare ecuatie in functie de SNR, unde i corespunde frame-ului cu indexul respectiv din formula PSS.
- Zgomotul este estimat de lungul perioadei de liniste si apoi „**spectral substraction**" este aplicat astfel: $|Y(k)|^2 = |S(k)|^2 + |D(k)|^2$.

- Spectral flooring (eliminarea valorilor negative), cu ajutorul parametrului „**spectral floor" β,** se face la fel ca in MBSS, astfel:

$$|\widehat{S}_i(k)|^2 = \begin{cases} |\widehat{S}_i(k)|^2 & |\widehat{S}_i(k)|^2 > \beta|Y_i(k)|^2 \\ \beta|Y_i(k)|^2 & else \end{cases}$$

Algoritmul:

1. Principiul de bază al acestei metode este de a scădea spectrul de magnitudine al zgomotului d (n) din vorbirea zgomotoasă y (n)

   y(n) = s(n) + d(n)

- unde s(n) este vorbirea curata "clean speech". Se presupune că zgomotul d(n) este necorelat și aditiv la semnalul de vorbire. Estimarea zgomotului d(n) este măsurată în timpul tăcerii sau al activității non-verbale în semnal.

2. Spectrul de putere al semnalului zgomotos poate fi scris ca:

$$|Y(k)|^2 = |S(k)|^2 + |D(k)|^2$$

   D(k) – spectrul de magnitudine al zgomotului

   S(k) – spectrul de magnitudine al vorbirii curate

   Y(k) – spectrul de magnitudine al semnalului de intrare

3. Deoarece spectrul de zgomot D (k) nu poate fi obținut direct, "time-average" al spectrului de putere | Dˆ (k) | este calculat în timpul unei perioade de tăcere, o estimare a spectrului de vorbire modificat poate fi dată ca:

$$\left|\hat{S}(k)\right|^2 = \left|Y(k)\right|^2 - \alpha\left|\hat{D}(k)\right|^2$$

- α – factorul de „over-substraction” (functie cu SNR segmentat)
- S^(k) – spectrul de magnitudine **estimat** al vorbirii curate
- D^(k) – spectrul de magnitudine **estimat** al zgomotului calculat de-a lungul unei perioade de liniste

Metoda PSS este implementată cu o singură bandă.

SNR-ul si **Factorul α** sunt calculate ca in metoda MBSS, prezentate mai sus.

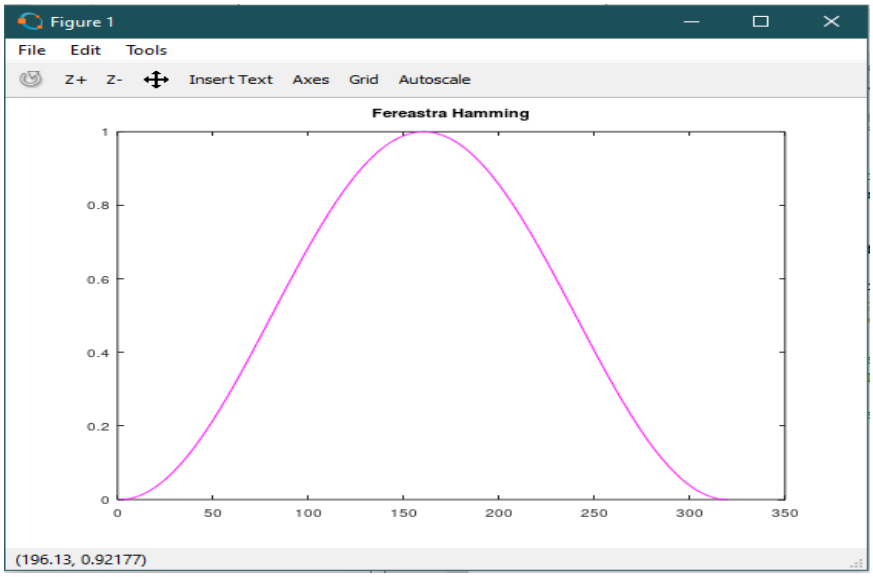
4. Spectrul modificat poate conține unele **valori negative** din cauza erorilor din spectrul de zgomot estimat. Aceste valori sunt rectificate folosind parametrul „**spectral floor” β.**

C. Reconstructia semnalului dupa aplicarea metodei PSS sau MBSS

Semnalul de vorbire obtinut din una din cele doua metode**, este reconstruit pentru a obtine semnalul de vorbire direct** si cu **aceeasi frecventa de esantionare** (**sampling frequency**). **Spectrul de vorbire** din fiecare cadru este transformat în **domeniu timp** folosind **functia IFFT** și apoi semnalul de vorbire complet este reconstruit folosind **Overlap Add**.

4. Rezultate si Experimente

1. MBSS method



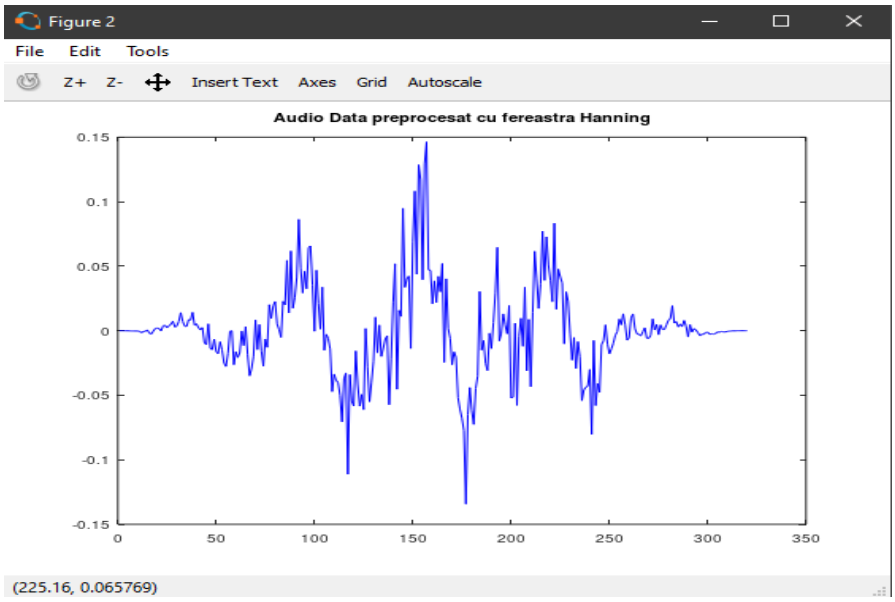



Figure 3
File Edit Tools
Z+ Z- Insert Text Axes Grid Autoscale
Weighted Spectral Average
red - before Weighted Spectral Average
blue - after Weighted Spectral Average
0.8
0.6
0.4
0.2
0
0 50 100 150 200 250 300 350

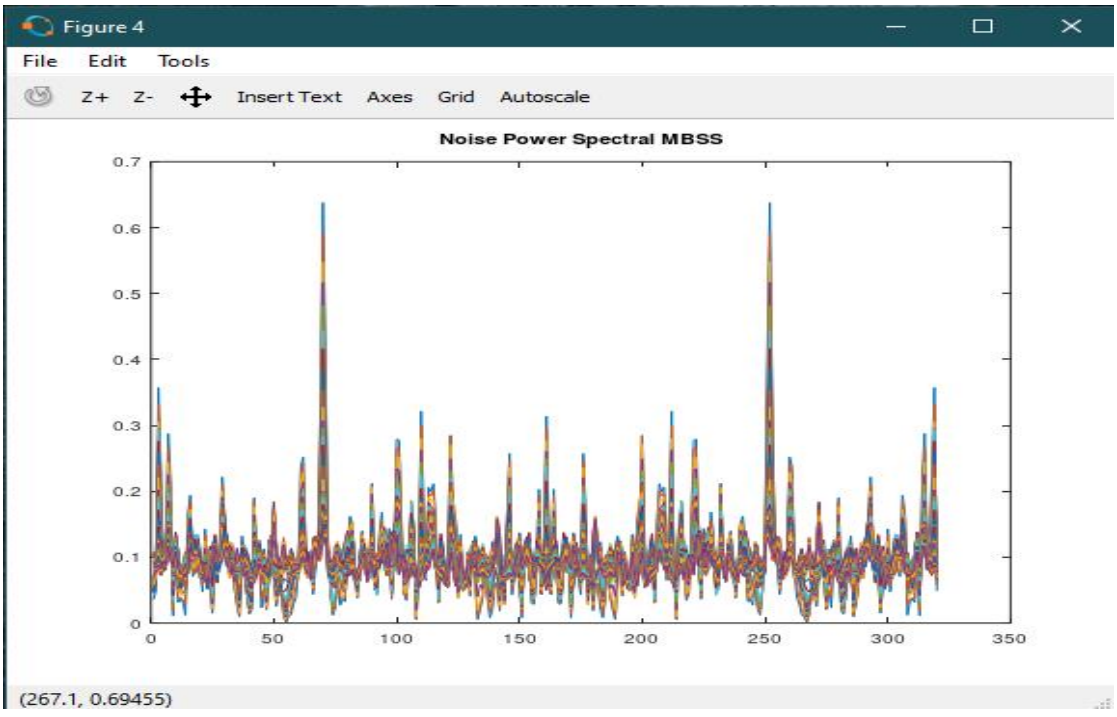


Figure 4
File Edit Tools
Z+ Z- Insert Text Axes Grid Autoscale
Noise Power Spectral MBSS
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0 50 100 150 200 250 300 350
(267.1, 0.69455)

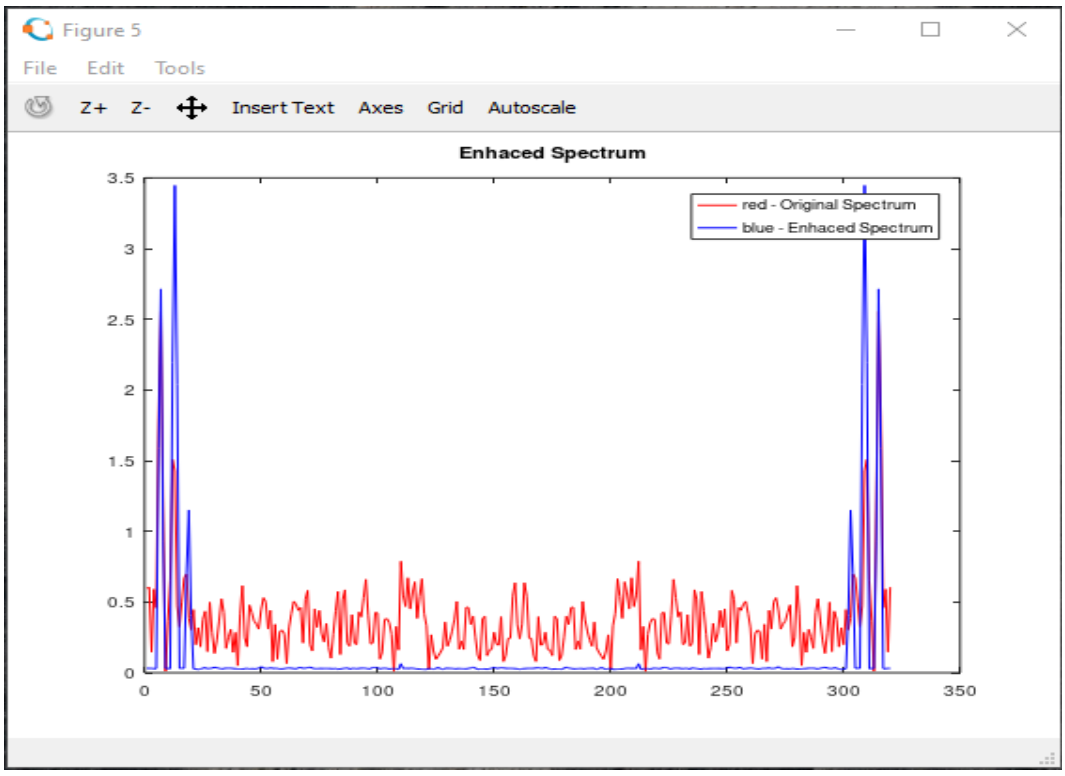


Figure 5
File Edit Tools
Z+ Z- Insert Text Axes Grid Autoscale
Enhaced Spectrum
red - Original Spectrum
blue - Enhaced Spectrum
3.5
3
2.5
2
1.5
1
0.5
0
0 50 100 150 200 250 300 350

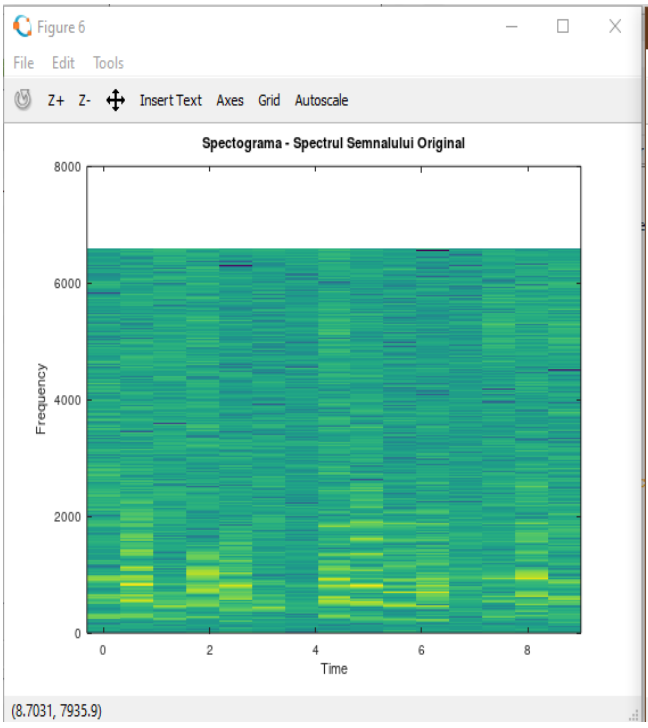


Figure 6
File Edit Tools
Z+ Z- Insert Text Axes Grid Autoscale
Spectograma - Spectrul Semnalului Original
8000
6000
4000
2000
0
Frequency
0 2 4 6 8
Time
(8.7031, 7935.9)

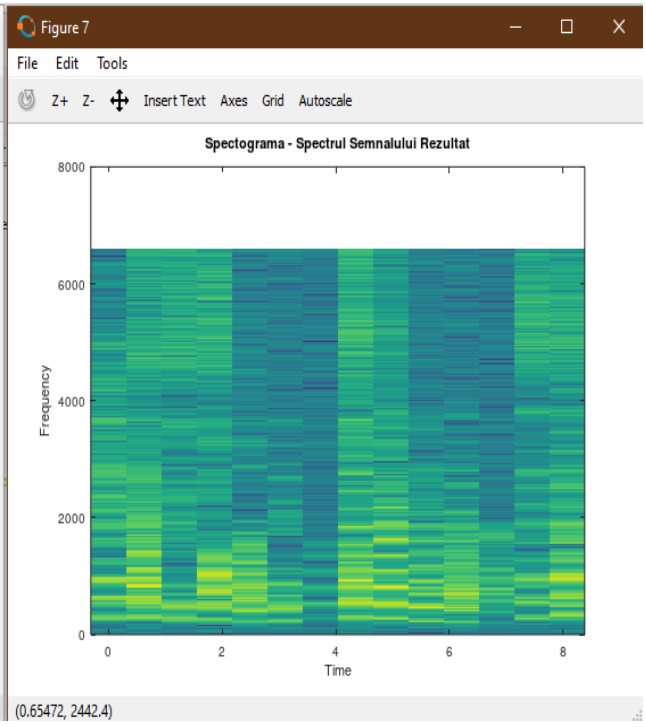


Figure 7
File Edit Tools
Z+ Z- Insert Text Axes Grid Autoscale
Spectograma - Spectrul Semnalului Rezultat
8000
6000
4000
2000
0
Frequency
0 2 4 6 8
Time
(0.65472, 2442.4)

2. PSS method

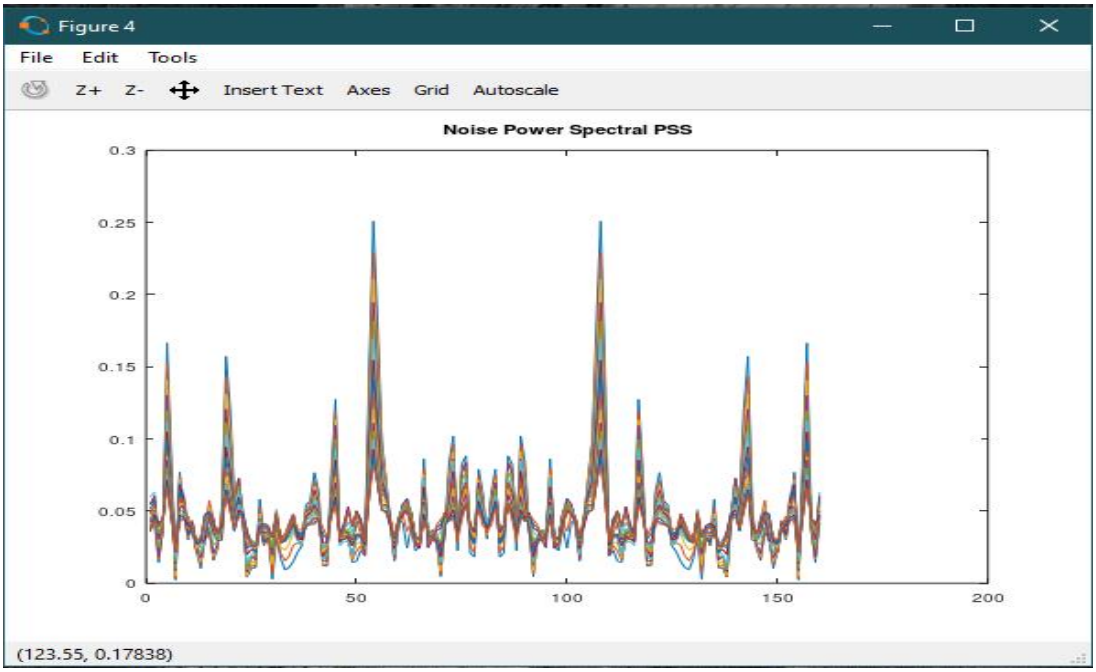


Figure 4
File Edit Tools
Z+ Z- Insert Text Axes Grid Autoscale
Noise Power Spectral PSS
0.3
0.25
0.2
0.15
0.1
0.05
0
0 50 100 150 200
(123.55, 0.17838)

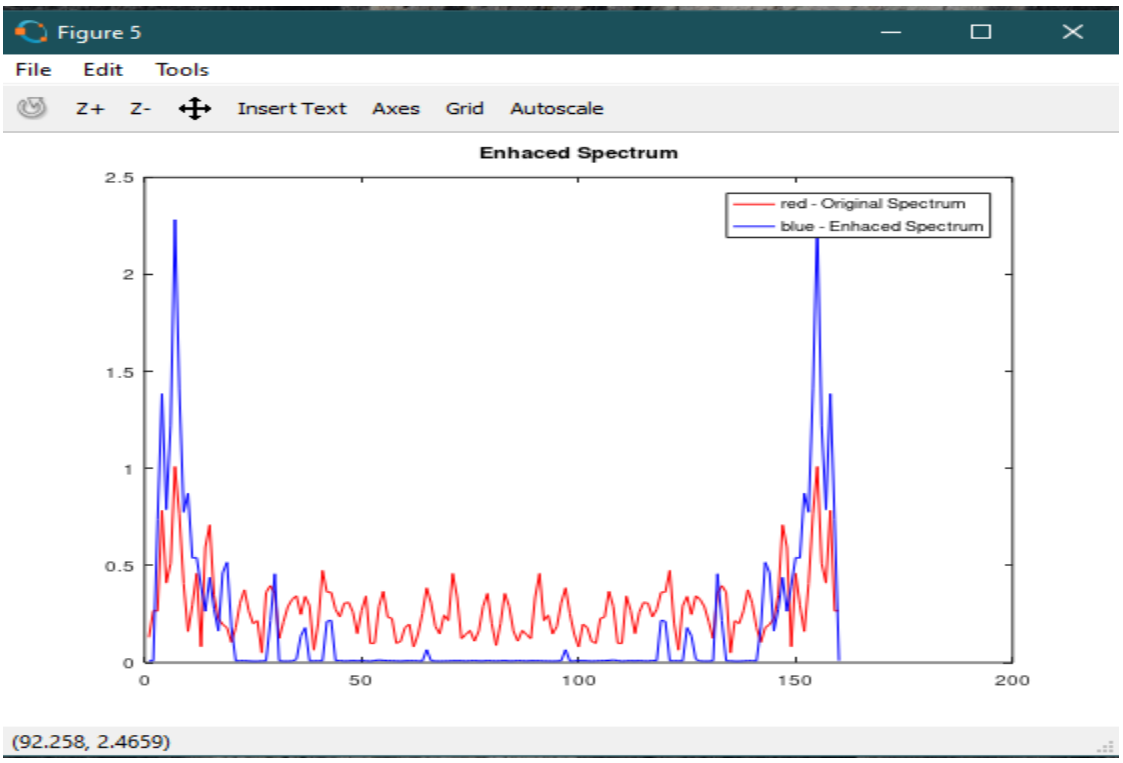
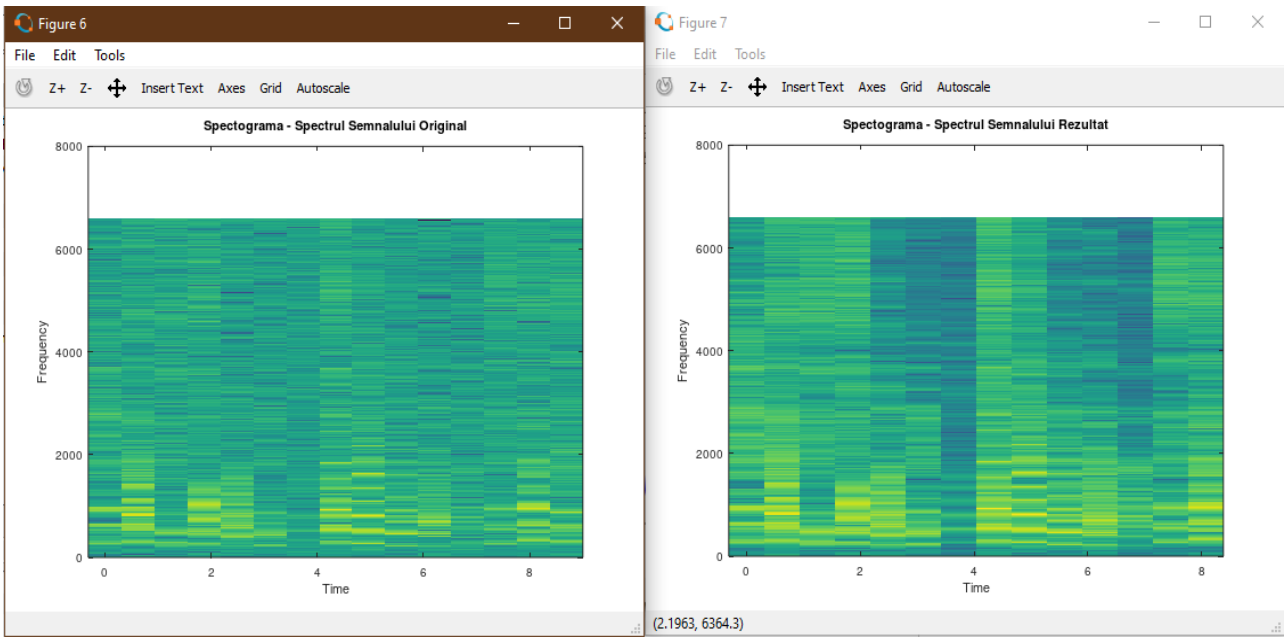


Figure 5
File Edit Tools
Z+ Z- Insert Text Axes Grid Autoscale
Enhaced Spectrum
red - Original Spectrum
blue - Enhaced Spectrum
2.5
2
1.5
1
0.5
0
0 50 100 150 200
(92.258, 2.4659)



5. Concluzii

1. Pentru metoda **MBSS**, **zgomotul este updatat pentru fiecare frame**. Se verifica pentru fiecare frame in parte daca este predominant speech sau zgomot si se actualizeaza vectorul noise.
2. Pentru metoda PSS, **vectorul noise se initializeaza la inceput**.
3. Metoda MBSS este mai buna **pentru speech-urile de mai lunga durata**, deoarece se **actualizeaza mreu noise vector.**
4. Factorul **substraction-band** (delta) ne ajuta sa avem o estimare mai buna a zgomotului prezent in MBSS.

Avantajul metodei MBSS – avantajul pre-procesarii în această metodă este de a reduce variatia estimării spectrale și, ulterior, de a reduce zgomotul rezidual.

Acest algoritm reajustează factorul de „**over-subtraction”(alfa)** din fiecare bandă.

6. Bibliografie

[1] P. Sunitha, Dr. K. Satya Prasad „Multi Band Spectral Subtraction for Speech Enhancement with Different Frequency Spacing Methods and their Effect on Objective Quality Measures” http://j.mecs-press.net/ijigsp/ijigsp-v11-n5/IJIGSP-V11-N5-6.pdf

[2] D. Deepa, A. Shanmugam „SPECTRAL SUBTRACTION METHOD OF SPEECH ENHANCEMENT USING ADAPTIVE ESTIMATION OF NOISE WITH PDE METHOD AS A PREPROCESSING TECHNIQUE” https://www.researchgate.net/publication/271770420_SPECTRAL_SUBTRACTION_METHOD_OF_SPEECH_ENHANCEMENT_USING_ADAPTIVE_ESTIMATION_OF_NOISE_WITH_PDE_METHOD_AS_A_PREPROCESSING_TECHNIQUE

[3] Siddala Vihari, A. Sreenivasa Murthy, Priyanka Soni and D. C. Naik “Comparison of Speech Enhancement Algorithms” https://www.researchgate.net/publication/306362918_Comparison_of_Speech_Enhancement_Algorithms

[4] Yasser Ghanbari, Mohammad Reza Karami-Mollaei, Behnam Amelifard “IMPROVED MULTI-BAND SPECTRAL SUBTRACTION METHOD FOR SPEECH ENHANCEMENT”

https://www.researchgate.net/profile/Yasser_Ghanbari/publication/228609880_Improved_multi-band_spectral_subtraction_method_for_speech_enhancement/links/00463517956744dd92000000.pdf

[5] Sunil D. Kamath and Philipos C. Loizou “A MULTI-BAND SPECTRAL SUBTRACTION METHOD FOR ENHANCING SPEECH CORRUPTED BY COLORED NOISE”

http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.7.4102&rep=rep1&type=pdf