

# Proiect ASRSV

## Analiza Semnalului Vocal

### Fundamentare teoretica

#### I. Energia medie pe timp scurt

Energia asociată cu vorbirea este variabilă în timp. Prin urmare, interesul pentru orice procesare automată a vorbirii este de a cunoaște modul în care energia variază în timp și pentru a fi mai specific, energia asociată cu regiunea de vorbire pe termen scurt. Semnalul de vorbire constă din regiuni vocale, nevocale și de tăcere. În plus, energia asociată cu regiunea vocală este mare în comparație cu regiunea nevocală și regiunea de tăcere nu va avea o energie minimă sau neglijabilă. Astfel, energia pe termen scurt poate fi utilizată pentru clasificarea vorbirii vocale, nevocate și a linistii.

Relația pentru găsirea energiei pe termen scurt poate fi derivată din relația de energie totală definită în procesarea semnalului.

Energia totală a unui semnal de energie este dată de:

$$E_T = \sum_{n=-\infty}^{\infty} s^2(n)$$

În cazul calculării energiei pe termen scurt, considerăm vorbirea în termeni de 10-30 msec. Fie ca eșantioanele dintr-un cadru de vorbire să fie date de „n=0 la n=N-1”, unde „N” este lungimea cadrului (eșantioane), apoi pentru calcularea energiei vorbirea va fi zero în afara lungimii cadrului. Atunci, pentru calcularea energiei, amplitudinea probelor de vorbire va fi zero în afara cadrului.

În consecință, putem scrie relația menționată mai sus ca:

$$E_T = \sum_{n=-\infty}^{-1} s^2(n) + \sum_{n=0}^{N-1} s^2(n) + \sum_{n=N}^{\infty} s^2(n)$$

$$E_T = \sum_{n=0}^{N-1} s^2(n)$$

Această relație va da energia totală prezentă în cadrul de vorbire de la „n=0 la n=N-1”. Pentru a reprezenta mai precis, un singur cadru de vorbire folosim relația:

$$s_w(n) = s(m).w(n-m)$$

unde "w(n)" reprezintă funcția de fereastră de durată finită. Există mai multe funcții de ferestre prezente în prelucrarea semnalelor. Cele mai utilizate includ dreptunghiulară, hanning și hamming. Pentru estimarea tuturor parametrilor din domeniul timi folosim fereastra dreptunghiulară pentru simplitatea ei.

Acum putem scrie relația de energie pe termen scurt după cum urmează:

$$e(n) = \sum_{m=-\infty}^{\infty} (s(m).w(n-m))^2$$

unde „n” este deplasarea/rata în numărul de mostre in care vrem să aflăm energia pe termen scurt. Deplasarea poate fi la fel de mică ca un eșantion sau la fel de mare ca dimensiunea cadrului. Energia pe termen scurt calculată pentru fiecare schimbare a eșantionului poate să nu fie necesară, deoarece variația energiei în cazul vorbirii este relativ lentă. Din acest motiv, deplasarea este menținută mult mai mare decât un eșantion. De obicei este aproximativ jumătate din dimensiunea cadrului.

## II. Numărul trecerilor prin zero

Numărul trecerilor prin zero (NTZ) reprezintă o aproximare a conținutului spectral al semnalului vocal.

Numărul trecerilor prin zero oferă informații despre numărul de treceri prin zero prezente într-un semnal dat. În mod intuitiv, dacă numărul de treceri cu zero este mai mare într-un semnal dat, atunci semnalul se schimbă rapid și, în consecință, semnalul poate conține informații de înaltă frecvență. Pe liniile similare, dacă numărul de treceri cu zero este mai mic, deci semnalul se schimbă lent și, în

consecință, semnalul poate conține informații de joasă frecvență. Astfel, ZCR oferă o informație indirectă despre conținutul de frecvență al semnalului.

Numărul de treceri prin zero (ZCC) pentru un bloc de N eșantioane de vorbire este definit ca:

$$ZCC_i = \sum_{k=1}^{N-1} 0.5 | \text{sign}(s[k]) - \text{sign}(s[k-1]) |$$

### III. Pitch Period Estimation

Deoarece frecvența înălțimii este de obicei mai mică de 600-700 Hz, semnalele de vorbire sunt mai întâi filtrate trece jos pentru a elimina componentele de peste acest interval de frecvență.

Cele două tehnici cel mai frecvent utilizate sunt:

- Funcție de autocorelare de scurtă durată
- Funcția de diferență de mărime medie (ADMF).

În timpul vorbirii vocale, semnalul de vorbire este „cvasi-periodic”. Fiecare tehnică încearcă să determine perioada (în eșantioane între „repetări” semnalului vocal vocal).

Funcția de autocorelare:

Corelația este o tehnică foarte frecvent utilizată în DSP pentru a determina „diferența de timp” dintre două semnale, unde unul este o versiune întârziată „aproape perfectă” a celuilalt. Autocorelația este aplicarea aceleiași tehnici pentru a determina „perioada” necunoscută a unui semnal cvasi-periodic, cum ar fi vorbirea.

Funcția de autocorelare pentru o valoare de întârziere a k eșantioane este:

$$\phi(k) = \frac{1}{N} \sum_{n=0}^{N-1} s[n]s[n-k].$$

#### IV. Average Magnitude Difference Function

AMDF este similar, dar opus cu funcția de autocorelare. Pentru o întârziere de  $k$  mostre, AMDF este definit ca:

$$D(k) = \frac{1}{N} \sum_{n=0}^{N-1} |s[n] - s[n-k]|$$

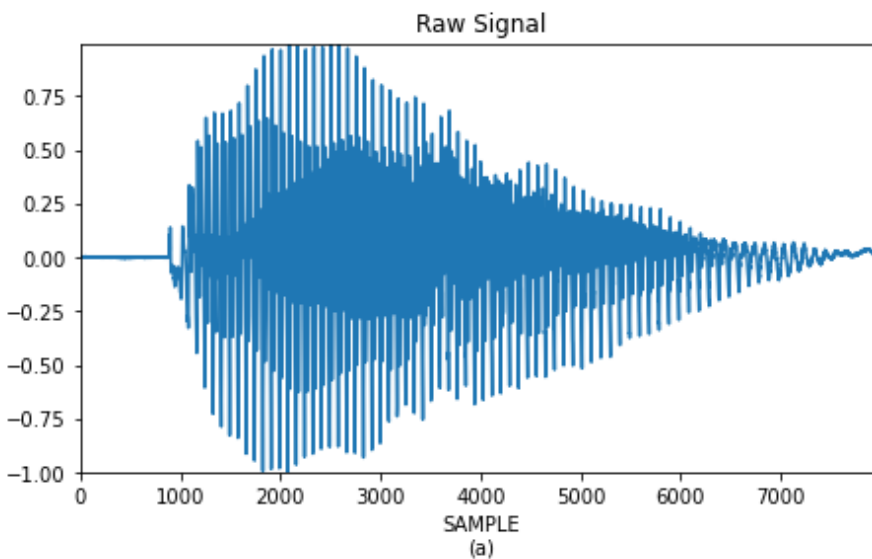
#### Implementare practica:

Pentru implementarea practica am folosit fisierul boy.wav.

#### Short-Time Energy

```
fs          ==> 16000 Hz  
len(signal) ==> 8000 samples
```

Diagrama semnalului:



```
samples/millisecond ==> 16  
samples/[20ms]frame ==> 320  
number of frames ==> 25
```

Vom afisa diagrama energie prin doua metode

Diagrama energiei pe termen scurt realizata prin metoda 1:

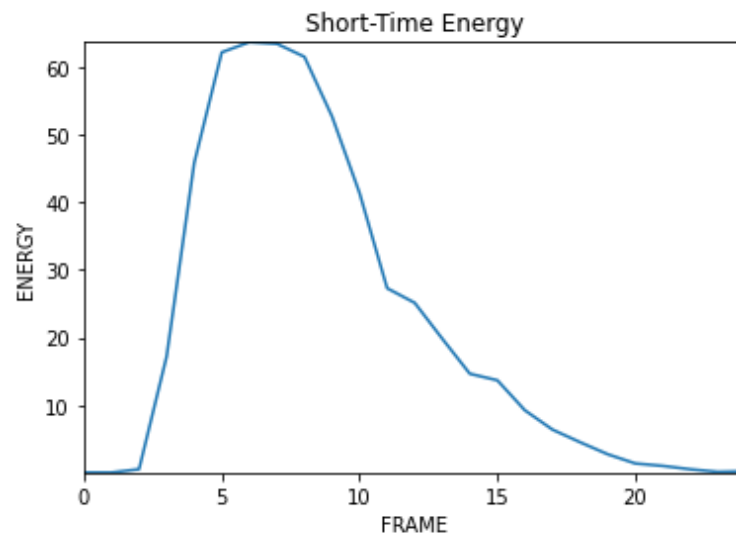


Diagrama energiei pe termen scurt realizata prin metoda 2:

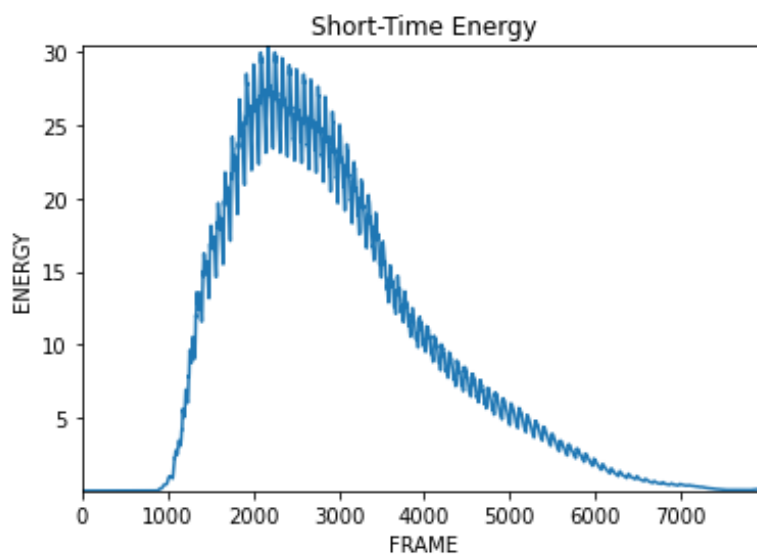


Diagrama realizata prin metoda 2 pare a avea mai mult zgomot dar este mai usor de calculat si se preteaza procesarii in timp real.

## Short-Time Zero Crossing Count

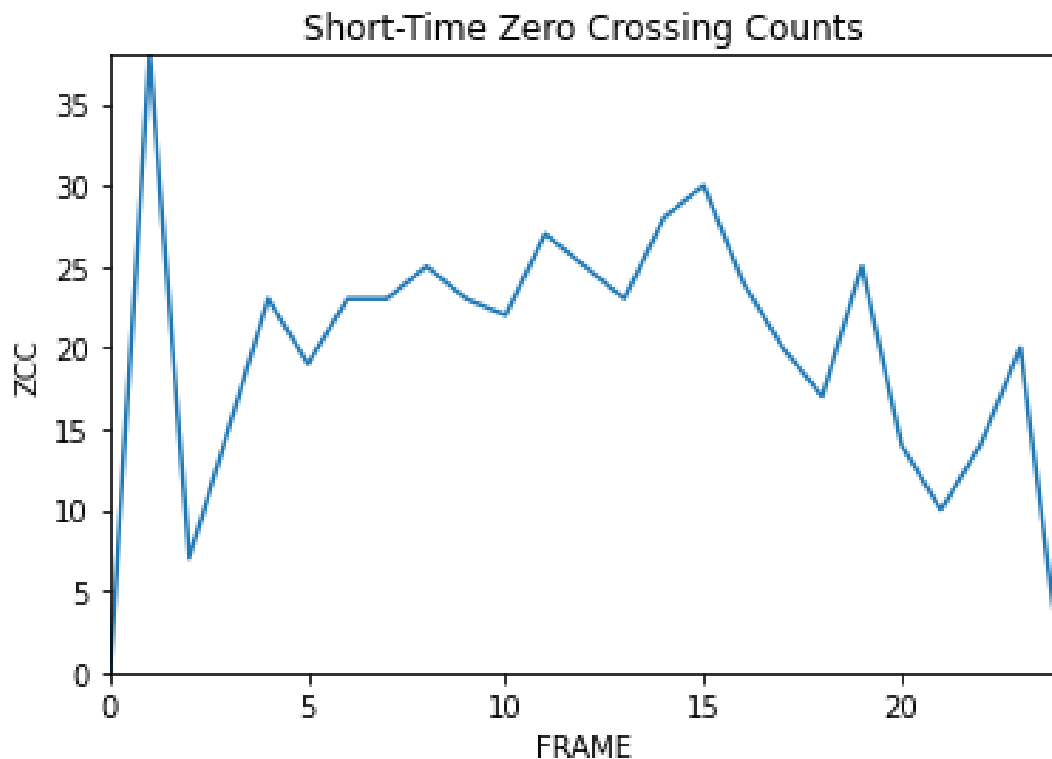
Primul pas este eliminarea offsetului DC:

```
DC ==> -0.0018053972545757073
mean(newSignal) ==> -2.0803671285651858e-18
```

Vom calcula numărul de trecere cu zero:

```
samples/millisecond ==> 16
samples/[20ms]frame ==> 320
number of frames ==> 25
```

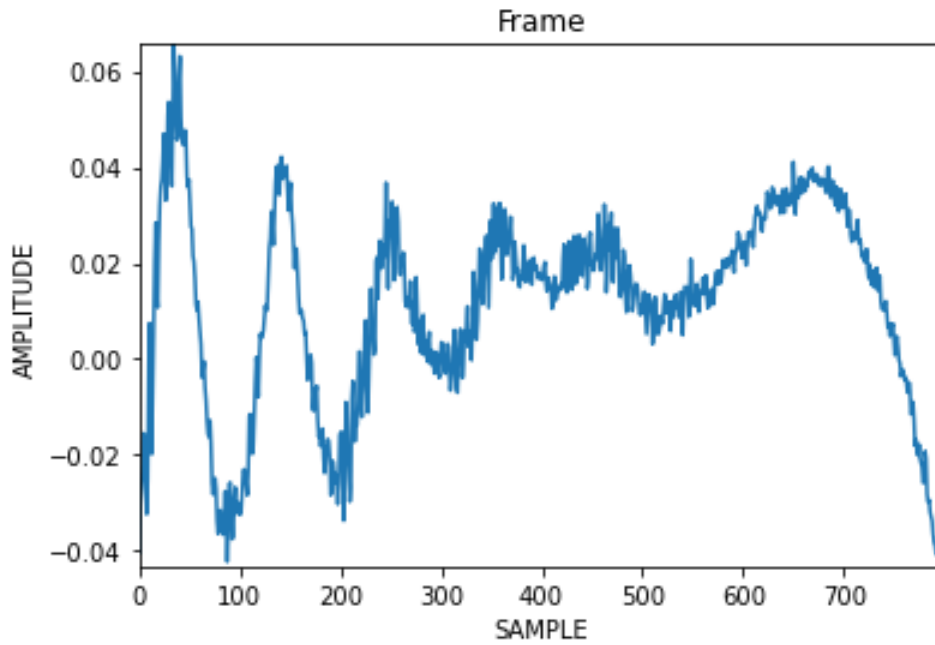
Mai jos avem graficul rezultatelor. Zonele figurii cu un număr foarte mare de trecere cu zero corespund regiunilor nevocale ale semnalului. Această metodă are de suferit într-un mediu zgomotos.



## Pitch Period Estimation

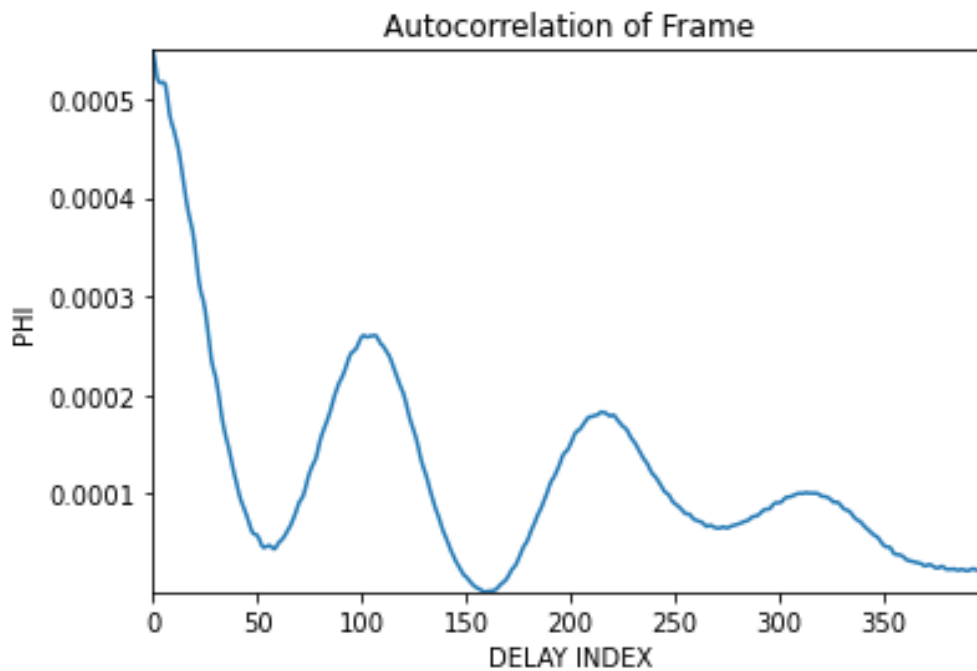
Perioada de pas a cadrului este determinată prin măsurarea distanței (în eșantioane) între vârfurile graficului de autocorelație.

---



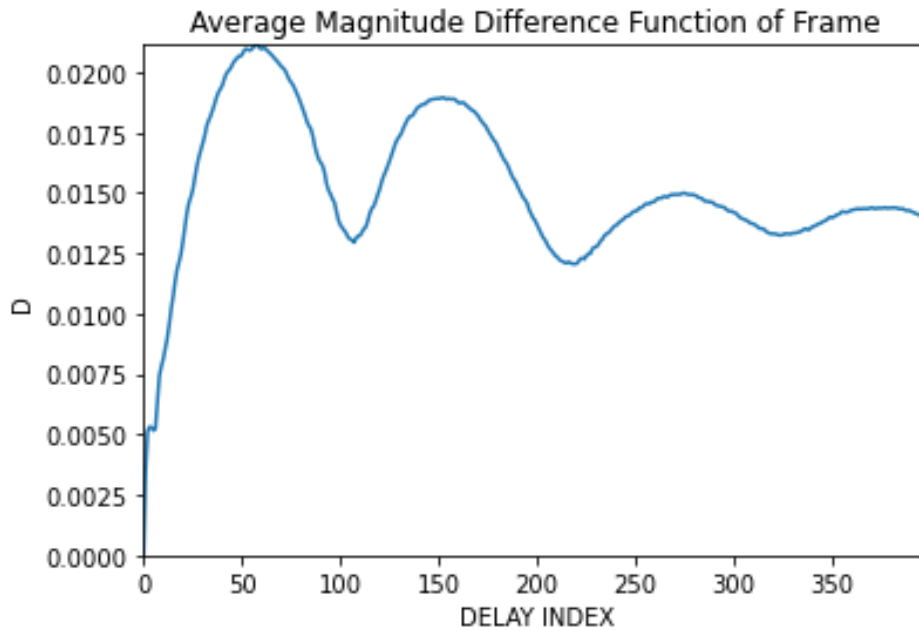
Autocorelatia aratand in felul urmator:

---



## Average Magnitude Difference Function

---



---

Obsevam ca figura arata ca o versiune inversata a autocorelatiei

AMDF-ul este de fapt mai ușor de calculat decât autocorelația și se pare că canalele ar putea fi de fapt mai ușor de selectat automat.