

Laboratorul 2: Extragerea de trăsături acustice: MFCC și LPC

Obiective:

- Înțelegerea modelului Sursă-Filtru al producției vorbirii și a rolului său în extracția de trăsături.
- Dezvoltarea abilității de a calcula și interpreta Coeficienții Cepstrali pe Scara Mel (MFCC).
- Familiarizarea cu conceptul de Codare Liniar Predictivă (LPC) și vizualizarea anvelopei spectrale.
- Aplicarea tehnicilor de normalizare, precum CMVN, pentru a crește robustețea trăsăturilor.

Resurse necesare:

- Mediul Python și bibliotecile *librosa*, *numpy*, *matplotlib*.
- Opțional: o bibliotecă pentru calculul LPC, precum *scipy.signal* sau *turbolpc*.
- Fișiere audio .wav conținând cuvinte sau vocale rostite de diferiți vorbitori.

Sarcini:

1. Extragerea coeficienților MFCC:

- Încărcați un fișier audio.
- Utilizați funcția *librosa.feature.mfcc()* pentru a extrage trăsăturile.
- Experimentați cu parametrii cheie: *n_mfcc* (numărul de coeficienți), *n_fft* (dimensiunea ferestrei FFT), *hop_length* (pasul de suprapunere).
- Vizualizați matricea de MFCC-uri rezultată folosind *librosa.display.specshow()*.

2. Analiza prin Codare Liniar Predictivă (LPC):

- Folosind un segment scurt de semnal (o vocală), calculați coeficienții LPC (de ex., cu *librosa.lpc()*).
- Calculați și plotați răspunsul în frecvență al filtrului LPC pentru a vizualiza anelopa spectrală și a identifica formanții.

3. Normalizarea trăsăturilor prin CMVN:

- Calculați MFCC-urile pentru un set de fișiere audio.
- Implementați o funcție simplă care calculează media și deviația standard pentru fiecare coeficient pe întregul set de date.
- Normalizați trăsăturile scăzând media și împărțind la deviația standard. Comparați vizual trăsăturile normalizate cu cele originale.

Întrebări de Reflecție:

- De ce este necesară extracția de trăsături în loc de a folosi direct forma de undă brută?
- Ce legătură există între scara Mel (folosită în MFCC) și percepția auditivă umană?
- Ce componentă a modelului Sursă-Filtru este modelată în principal de LPC?
- Ce tip de variabilitate este redusă prin aplicarea normalizării CMVN?

Test Formativ 2

Secțiunea I – Definiții și Concepte

1. Descrieți pe scurt cele două componente ale **modelului Sursă-Filtru** în producția vorbirii.
2. Care este scopul principal al aplicării unei ferestre (ex: Hamming) în procesul de extracție a trăsăturilor?
 - a) Creșterea frecvenței de eșantionare
 - b) Reducerea efectelor de margine și a discontinuităților la analiza pe cadre
 - c) Amplificarea frecvențelor joase
 - d) Eliminarea completă a zgomotului

Secțiunea II – Tehnologii și Metode

3. Menționați doi pași din pipeline-ul de calcul al **MFCC**.
4. Asociați fiecare tehnică cu descrierea sa principală:

Tehnică	Descriere
1. MFCC	a) Modelează tractul vocal prezicând un eșantion ca o combinație liniară a celor anterioare.
2. LPC	b) Standardizează trăsăturile pentru a avea medie zero și varianță unitară.
3. CMVN	c) Utilizează o bancă de filtre pe scara Mel și Transformata Cosinus Discretă.

Secțiunea III – Aplicații Practice

5. Ce trăsătură acustică este considerată standardul *de facto* în sistemele ASR moderne?
6. Anvelopa spectrală calculată prin LPC este utilă în special pentru a vizualiza:
 - a) Perioada pitch-ului
 - b) Formanții (rezonanțele tractului vocal)
 - c) Energia totală a semnalului
 - d) Rata de trecere prin zero

Secțiunea IV – Laborator și Aplicații Practice

7. Ce funcție din librosa este folosită pentru a calcula MFCC-urile?
8. De ce este importantă Transformata Cosinus Discretă (DCT) ca pas final în calculul MFCC?

Secțiunea V – Reflecție Scurtă

Comparați pe scurt avantajele și dezavantajele MFCC față de LPC în contextul recunoașterii vorbirii.