

Laboratorul 1: Primii pași în analiza audio cu Python

Obiective:

- Dezvoltarea abilității de a manipula și analiza fișiere audio folosind limbajul Python.
- Înțelegerea conceptelor fundamentale de semnal audio digital: formă de undă, rată de eșantionare, spectrogramă.
- Familiarizarea cu biblioteca librosa pentru operații de bază precum încărcarea, vizualizarea și redarea semnalelor audio.
- Aplicarea practică a Transformatei Fourier pe Termen Scurt (STFT) pentru a analiza conținutul spectral al vorbirii.

Resurse Necesare:

- Un mediu de programare Python (ex: Jupyter Notebook, Google Colab).
- Bibliotecile Python: librosa, matplotlib, numpy.
- Un fișier audio în format .wav pentru analiză (se poate folosi un exemplu din librosa sau o înregistrare proprie).

Sarcini:

1. **Configurarea mediului de lucru:**
 - Instalați bibliotecile necesare folosind comanda: *pip install librosa matplotlib numpy scipy*.
2. **Încărcarea și inspectarea fișierului audio:**
 - Utilizați funcția *librosa.load()* pentru a încărca un fișier audio. Aceasta va returna seria temporală (un vector NumPy) și rata de eșantionare.
 - Analizați și afișați forma (.shape) seriei temporale și valoarea ratei de eșantionare.
3. **Vizualizarea formei de undă (Waveform):**
 - Folosiți *matplotlib.pyplot* și *librosa.display.waveshow()* pentru a plota forma de undă a semnalului.
 - Observați variațiile de amplitudine în timp și corelați-le cu segmentele de vorbire sau liniște.
4. **Calculul și afișarea spectrogramei:**
 - Calculați Transformata Fourier pe Termen Scurt (STFT) a semnalului folosind *librosa.stft()*.
 - Converteți magnitudinea STFT la scară logaritmică (decibeli) cu *librosa.amplitude_to_db()*.
 - Afișați spectrograma rezultată folosind *librosa.display.specshow()*, cu axa timpului și o axă a frecvențelor logaritmică.

Întrebări de Reflecție:

- Ce informații oferă forma de undă despre semnalul vocal, și ce informații *nu* oferă?
- Cum se manifestă vizual pe spectrogramă sunetele sonore (precum vocalele) față de cele surde sau de pauze?
- Care este rolul ratei de eșantionare și ce se întâmplă dacă Teorema Nyquist-Shannon nu este respectată?

Secțiunea I – Definiții și Concepte

1. Definiți conceptele de **Recunoaștere Automată a Vorbirii (ASR)** și **Sinteză Text-Vorbire (TTS)**.
2. Care este scopul principal al digitalizării unui semnal audio?
 - a) Creșterea amplitudinii semnalului
 - b) Conversia unui semnal analogic continuu într-o secvență discretă de valori numerice
 - c) Eliminarea zgomotului de fond
 - d) Reducerea duratei fișierului audio

Secțiunea II – Analiză Tehnică

3. Menționați cele două etape fundamentale ale conversiei analog-digitale.
4. Asociați fiecare reprezentare cu domeniul corespunzător:

Reprezentare	Domeniu
1. Forma de Undă (Waveform)	a) Domeniul Frecvență
2. Spectrograma	b) Domeniul Timp

Secțiunea III – Aplicații Practice și Studii de Caz

5. Ce reprezintă benzile orizontale de energie concentrată pe o spectrogramă a unei vocale?
6. Sistemul "Audrey" de la Bell Labs (1952) este un exemplu de:
 - a) Sistem de sinteză vocală modern
 - b) Sistem de recunoaștere a cifrelor rostite, dependent de vorbitor
 - c) Model end-to-end bazat pe Transformer
 - d) Sistem de traducere automată

Secțiunea IV – Laborator și Aplicații Practice

7. Ce funcție din biblioteca librosa este utilizată pentru a încărca un fișier audio?
8. De ce este utilă conversia magnitudinii spectrului în decibeli (dB) înainte de vizualizare?

Secțiunea V – Reflecție Scurtă

9. În opinia dumneavoastră, care este cea mai mare provocare în recunoașterea automată a vorbirii (variabilitatea vorbitorilor, zgomotul de fond, ambiguitatea limbajului etc.) și de ce?