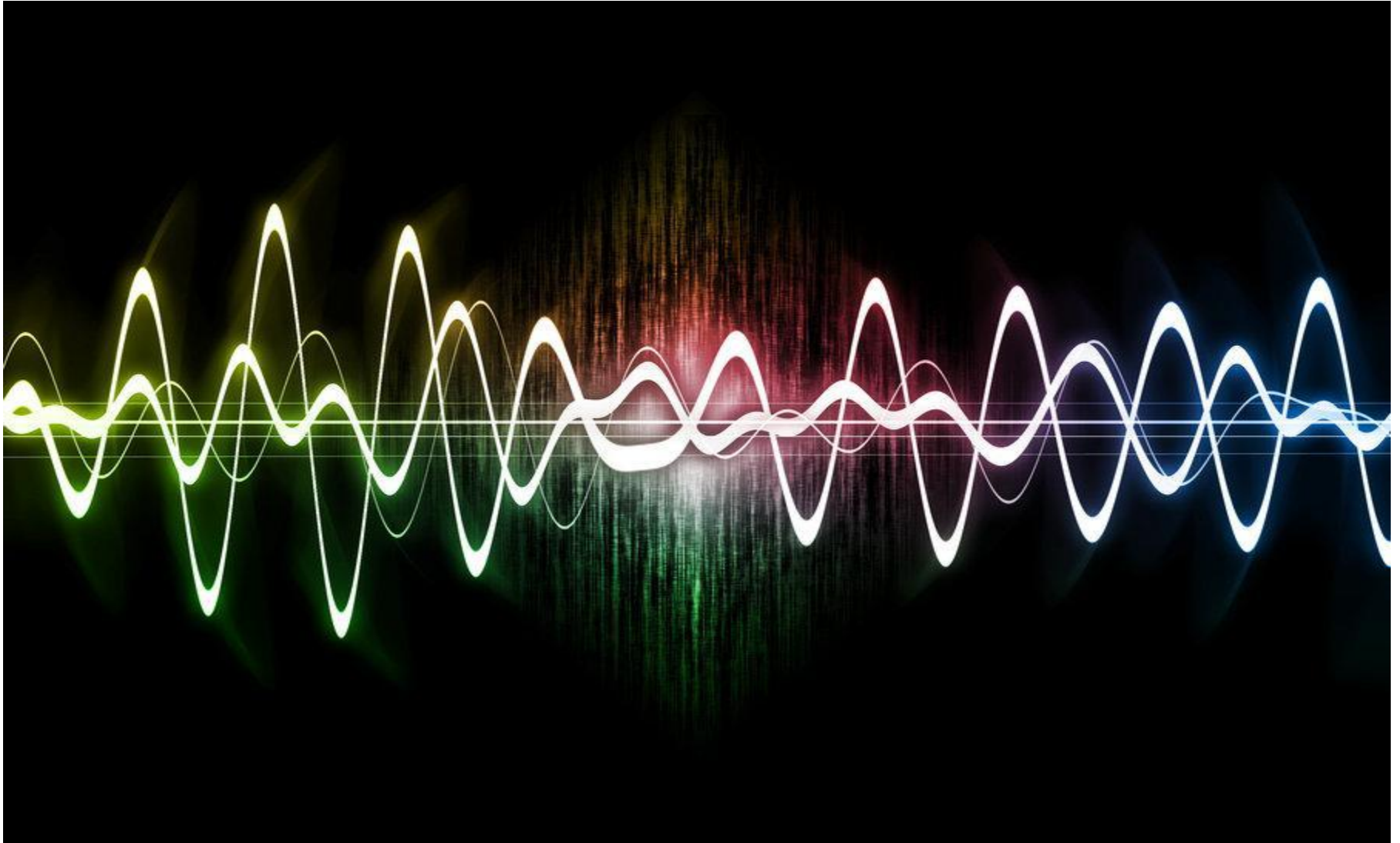

CMUSphinx CDL Project

Speaker Diarization

Vorbirea



CMUSphinx

- Open Source inceput de Carnegie Mellon University
 - IRC (#cmusphinx pe freenode)
 - Mailing Lists
 - Comunitate
-

Vorbirea(1)

- fenomen complex
 - proces dinamic fara parti bine definite
 - depinde de vorbitor, context, stil si altele
 - sistemele de recunoastere nu au niciodata acuratete de 100%
-

Vorbirea(2)

- este un stream audio continuu
 - este alcatuita din foneme (senone - depind de context)
 - se considera unitati sub-fonetice (in general 3):
 - prima parte depinde de fonemul anterior
 - a doua parte este stabila
 - a treia parte depinde de fonemul urmator
-

Sisteme de recunoastere automata a vorbirii (1)

- metoda generala: se ia unda, se imparte in **utterances** in functie de momentele de liniste si se incearca recunoasterea pentru fiecare utterance.
 - din semnal se extrage un **feature vector** (39 de numere care reprezinta sunetul)
 - **MMA** / retele neuronale
-

Sisteme de recunoastere automata a vorbirii (2) - Modele

- Recunoasterea se bazeaza pe 3 modele:
 - Modelul de limba
 - Modelul acustic
 - Dictionar fonetic
-

Speaker Diarization (1)

- "Who spoke when?"
 - Presupune paritionarea inputului audio in segmente omogene in functie de vorbitor.
 - Imbunatateste lizibilatea textului si poate avea multe alte aplicatii.
 - Vom implementa in Java
-

Speaker Diarization (2)

- exista mai multe metode de implementare descrise in diferite paper-uri
 - LIUM SpkDiarization tool
 - clusterizare de Gaussian Mixture Models (GMMs)
-