Conversational UIs

Spoken Language Processing

Topic 4: Speaker Recognition

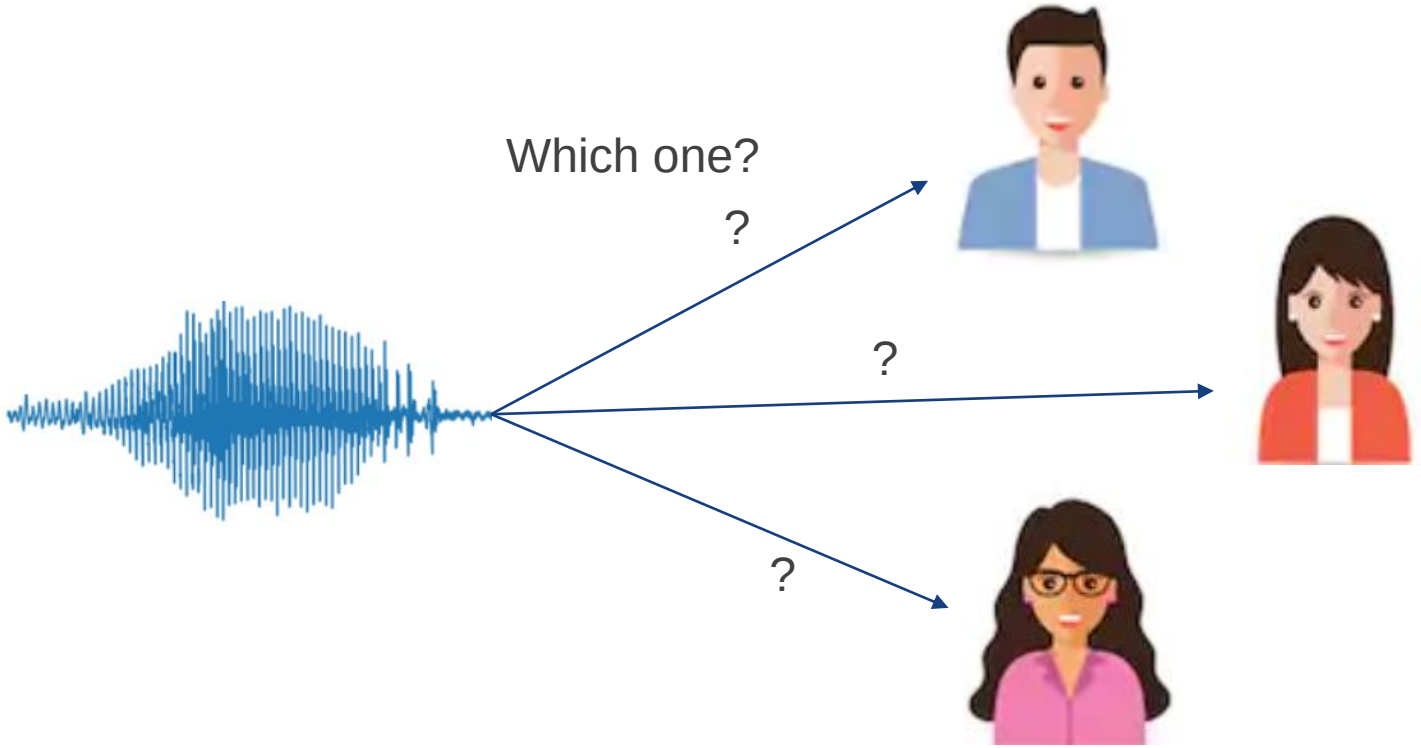
Dr. Dong Minghui
Institute for Infocomm Research
Agency for Science, Technology and Research (A-Star)
Email: mhdong@i2r.a-star.edu.sg

Speaker Recognition

- **To recognize persons from their voices**
- **Input modes**
  - Text dependent
  - Text independent
  - Text prompted
- **Decision modes**
  - Identification (one-to-many matching)
  - Verification/detection (one-to-one matching)

Speaker Identification

- **To find the speaker from a group of people.**
- **One to many matching.**



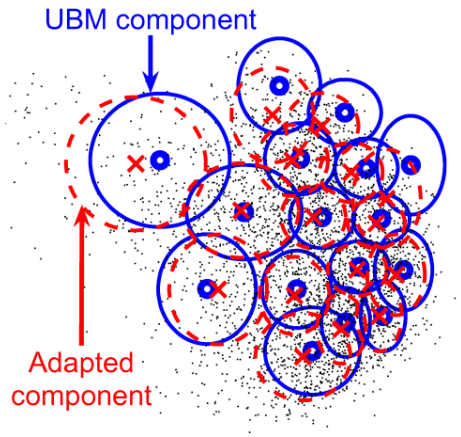
Speaker Verification

- **To detect whether the voice belongs to the claimed speaker.**
- **One-to-one matching.**

Yes or No?

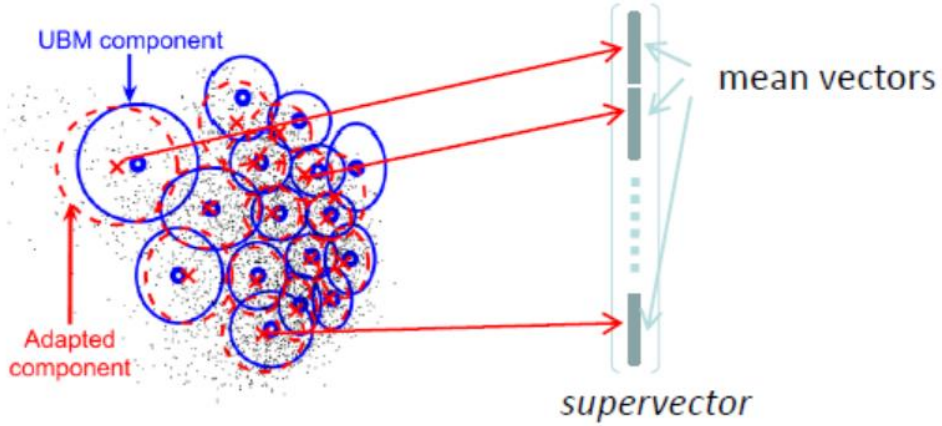
Speaker Recognition Methods

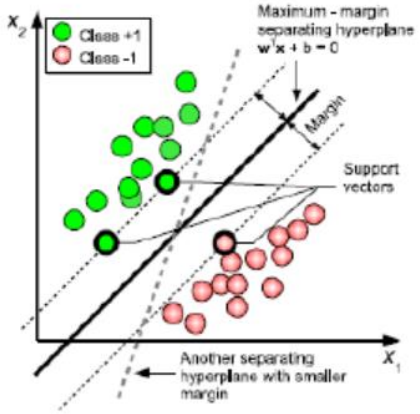
- **GMM-UBM method**
  - Build a Universal Background Model (UBM) with many people's speech data.
  - Use sample voice of the target speaker to adapt the GMM model to the target speaker's model.
  - The adapted model is used for recognition.



Speaker Recognition Methods

- **SuperVector + SVM**
  - Build UBM model first
  - Build adapted speaker model.
  - Create supervector
  - Use support vector machine (SVM) to do classification





Speaker Feature: i-Vector

- An i-Vector is a fixed-length low-dimensional representation of a variable-length speech utterance.
- i-Vecor = speaker + language + recording device + transmission channel + acoustic environment.
- A universal background model (UBM) that is used to collect sufficient statistics, a large projection matrix to extract i-vectors
- A probabilistic linear discriminant analysis (PLDA) backend to compute a similarity score between i-vectors.

Speaker Feature：x-Vector

- Embeddings are extracted from a feedforward deep neural network
- Long-term speaker characteristics are captured in the network by a temporal pooling layer that aggregates over the input speech.
- After training, utterances are mapped directly to fixed-dimensional speaker embeddings.
- Pairs of embeddings are scored using a PLDA-based backend.

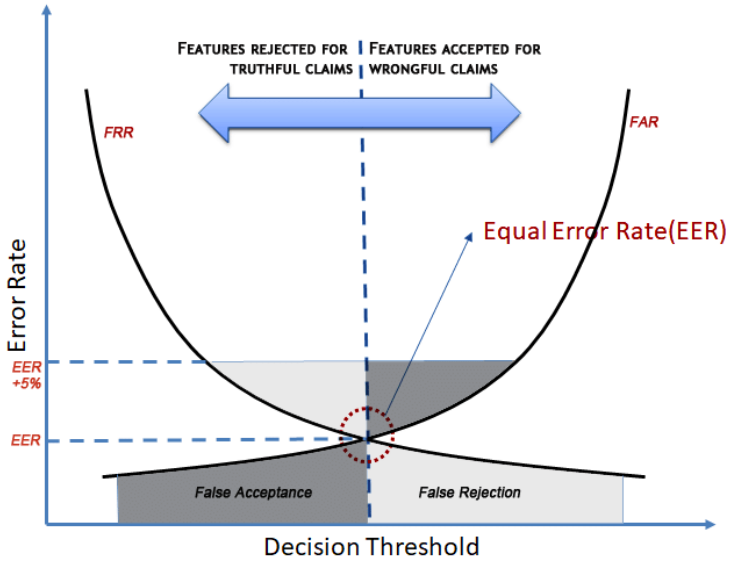
Accuracy Measure

Measures for speaker verification

- False Acceptance Rate (FAR)
- False Rejection Rate (FRR)
- Equal Error Rate (EER)

How to set threshold:

- FAR and FRR depend on the threshold of similarity measure for a decision.
- High false rejection means a higher security level towards imposters.
- High false acceptance means more convenience for true users.
- EER is a balanced choice. Lower EER means better accuracy of the system.



True identify	Recognition result	
	Accepted	Rejected
True user	correct	False rejection
Imposter	False acceptance	correct

Use Speaker Recognition

- **Enrollment**
  - To register the user in the system by supplying sample speech.
  - Normally, longer recording to get enough information. e.g 30 seconds, or read a few sentences.
- **Threshold**
  - For verification task, a threshold should be set based on application.
  - A development dataset is normally used to set a threshold.
- **Recognition**
  - The speaker recognition system will be able to recognize speech. A few seconds voice for fast applications, or the longer speech for better accuracy.

Pros and Cons

- **Pros:**
  - Identity can be recognized over telephone line. No special devices are needed.
  - Background recognition can be done while talking.
  - Difficult to break if text is prompted by system.
- **Cons:**
  - Long speech recordings are needed for high accuracy applications, e.g. bank application.
  - Users may be required to read certain text for better accuracy. Not a very natural way.

Applications Scenarios

- **Authentication**
  - Second and third factor authentication for remote banking
  - Access control
- **Personalized application**
  - Detect speaker automatically

Speaker Recognition Resource

- **ALIZÉ**
  - An opensource platform for speaker recognition.
  - Provides low-level and high-level frameworks
  - Supports verification/identification, segmenting, etc.
- **Kaldi**
  - Complete speaker recognition support

Thank you!

mhdong@i2r.a-star.edu.sg