ĐẠI HỌC QUỐC GIA THÀNH PHỐ HỒ CHÍ MINH
ĐẠI HỌC KHOA HỌC TỰ NHIÊN
KHOA CÔNG NGHỆ THÔNG TIN - BỘ MÔN KHOA HỌC MÁY TÍNH

SINH TRẮC HỌC GIỌNG NÓI
VOICE BIOMETRICS

Nguyễn Hoàng Đức, Lê Nhựt Nam, Nguyễn Viết Dũng

GV Lý thuyết: **PGS. TS** Lê Hoàng Thái
GV Hướng dẫn: Nguyễn Ngọc Thảo, Lê Thanh Phong

Ngày 14 tháng 5 năm 2021

Nội dung trình bày

A. Trình bày nội dung tìm hiểu được từ Chapter 8 - Voice Biometrics

- Giới thiệu
- Xác định những thông tin trong tín hiệu giọng nói
- Rút trích đặc trưng và Phân tách dữ liệu
- Nhận dạng giọng nói phụ thuộc văn bản
- Nhận dạng giọng nói không phụ thuộc văn bản
- Ứng dụng

Nội dung trình bày

B. Trình bày các phương pháp STATE OF THE ART của Voice Recognition

- Mục đích
- Động lực nghiên cứu khoa học
- Phát biểu bài toán
- Các công trình liên quan
- Phương pháp giải thuật
- Demo
- Tài liệu tham khảo

Giới thiệu

- Giọng nói (Voice/Speech) là một đặc điểm sinh trắc học (nhân trắc học) dễ dàng tiếp cận nhất mà không cần phải có thêm thiết bị thu nhận và hệ thống truyền dẫn.
- Có lợi thế khi áp dụng vào các hệ thống điều khiển từ xa
- Giọng nói không chỉ liên quan đến các đặc trưng cá thể mà còn liên quan với môi trường xung quanh và vấn đề xã hội, do vậy việc sản sinh giọng nói là một kết quả của một quá trình hết sức phức tạp.

Những thông tin nhận dạng trong tín hiệu giọng nói

- Idiolectal characteristics: cách phát âm phản ánh khu vực bạn đang sống hoặc đã sống và các phong cách nói khác nhau thay đổi một cách tinh vi tùy thuộc vào người bạn đang nói đến.
- Phonotactics characteristics:
- Prosody characteristics:
- Short-term spectral characteristics

Phân tích theo từng đoạn ngắn (Short-term Analysis)

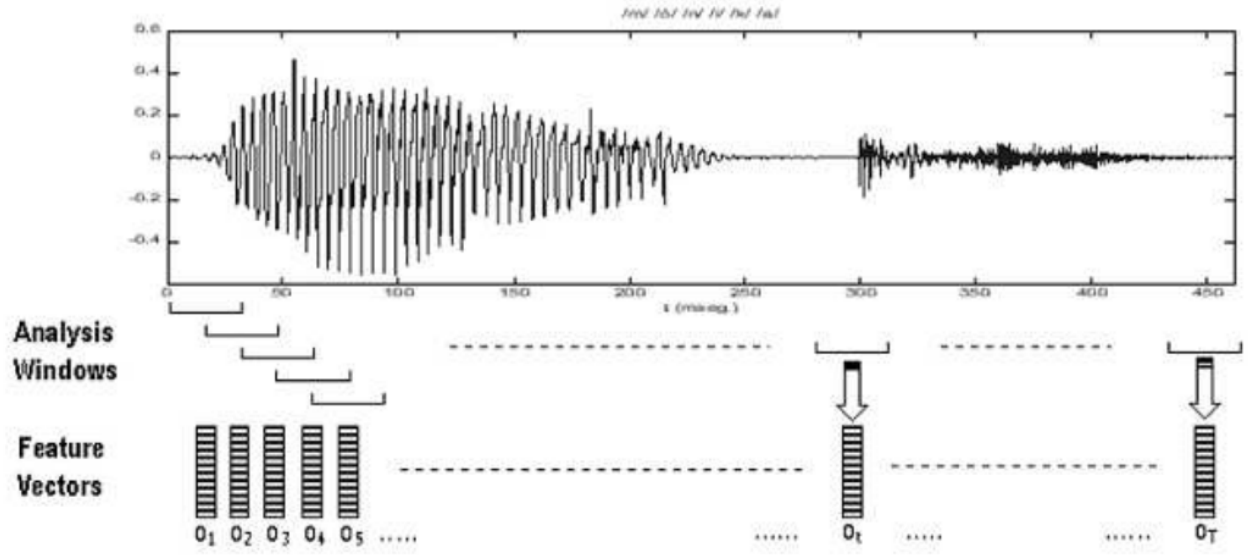
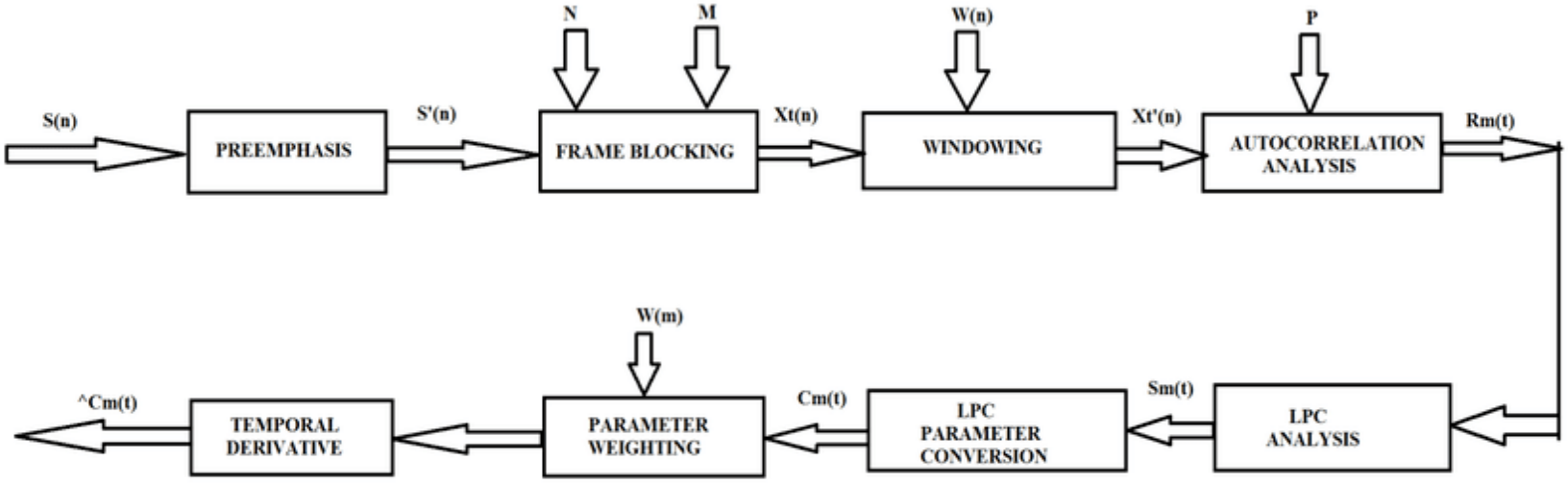


Fig. 8.1. Short-term analysis and parameterization of a speech signal.

Hình 1: Handbook of Biometrics, page 155

Tham số hóa

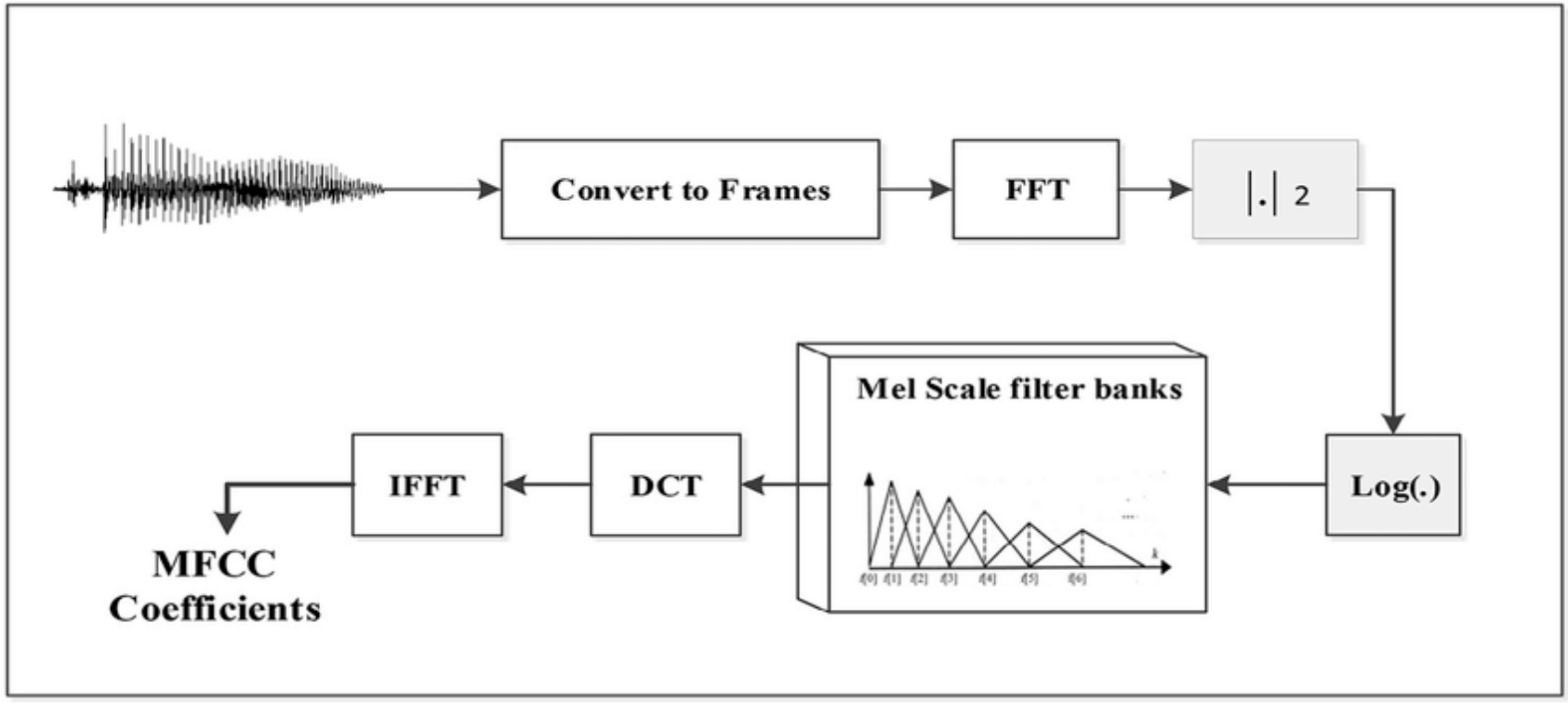
Tham số hóa bằng cách dùng Linear Predictive Coding (LPC)



Hình 2: Handbook of Biometrics, page 162

Tham số hóa

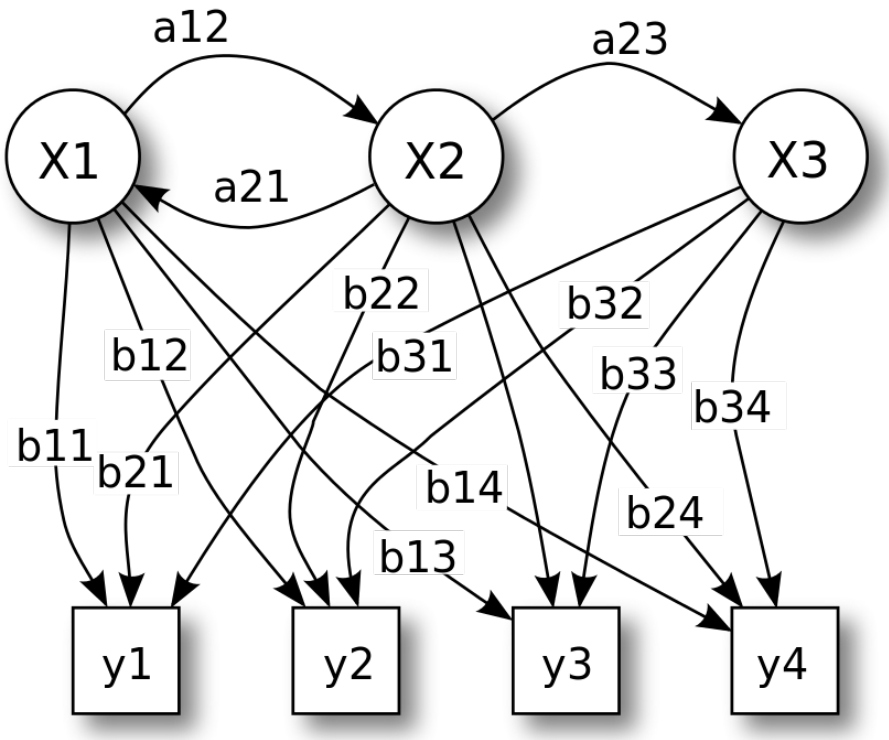
Tham số hóa bằng cách dùng Mel-Frequency based Cepstral Coefficients (MFCC)



Hình 3: Handbook of Biometrics, page 162

Phân tích ngữ âm và tách từ

Mô hình Hidden Markov



Hình 4: Sơ đồ mô hình Markov ẩn

Phân tách ngữ điệu

Dựa trên cơ sở là cao độ và năng lượng ở từng frame

- Cao độ: xác định bằng phương pháp tự động tương quan, phân rã cepstral dựa trên một số phương thức làm mịn bằng bộ lọc.
- Năng lượng: Năng lượng cửa sổ thu được rất dễ dàng thông qua định lý Parseval.

Giới thiệu

Hệ thống nhận dạng giọng nói phụ thuộc văn bản, sử dụng nội dung từ vựng của giọng nói phát ra để nhận dạng giọng nói, ứng dụng chính của hệ thống này trong các hệ thống tương tác, nơi cần có sự hợp tác từ người dùng để xác thực danh tính của họ.

Phân loại

- Hệ thống văn bản tĩnh: nội dung từ vựng trong ghi danh và các mẫu nhận dạng luôn giống nhau.
- Hệ thống văn bản động: tạo ra một lời nhắc mật khẩu được tạo ngẫu nhiên khác nhau mỗi khi người dùng được xác minh (hệ thống nhắc bằng văn bản)

Kho ngữ liệu

- YOHO Speaker Verification
- MIT Mobile Device Speaker Verification Corpus
- BIOSEC Baseline Corpus

Phương pháp thực hiện

- Phương pháp dựa trên khuôn mẫu: bao gồm một số chuỗi vectơ tương ứng với lời nói đăng ký và việc nhận dạng được thực hiện bằng cách so sánh lời nói xác minh với lời nói đăng ký.
- Phương pháp thống kê: Nổi bật nhất là mô hình Markov ẩn (HMM), cho phép chọn đơn vị tiếng nói từ đơn vị âm vị phụ đến từ và cho phép thiết kế hệ thống nhắc văn bản.

Công trình tiêu biểu

- Introduction 1
- Introduction 2
- Introduction 3

Giới thiệu

Hệ thống phổ âm trong thời gian ngắn

Hệ thống Idiolectal

Hệ thống ngữ âm

Mô hình hệ thống ngữ âm

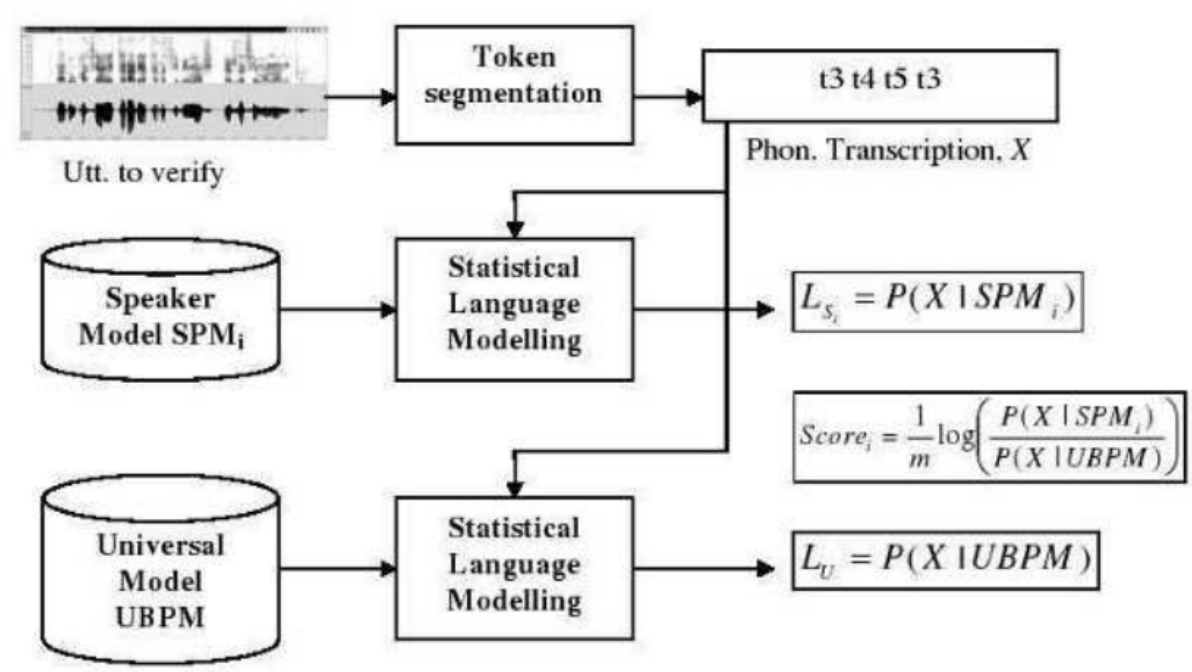


Fig. 8.3. Verification of an utterance against a speaker model in phonotactic speaker recognition

Hình 5: Handbook of Biometrics, page 162

Hệ thống ngữ điệu

Mô hình hệ thống ngữ điệu

TOKEN	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
FO	+F	+F	+S	+S	-F	-F	-S	-S	+F	+F	+S	+S	-F	-F	-S	-S	UV
E	+F	+S	+F	+S	-F	-S	-F	-S	-F	-S	-F	-F	+F	+S	+F	+S	"

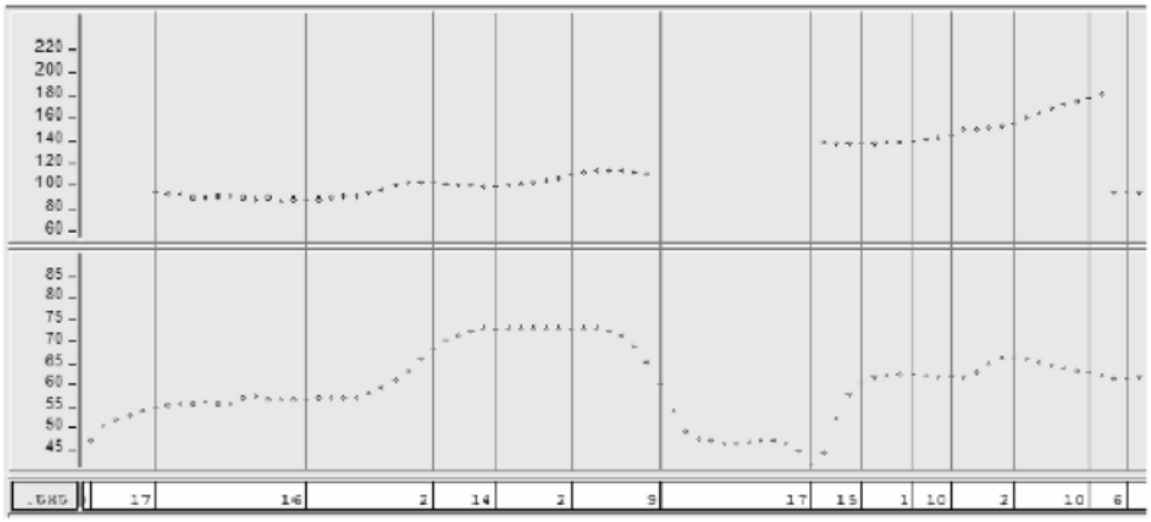


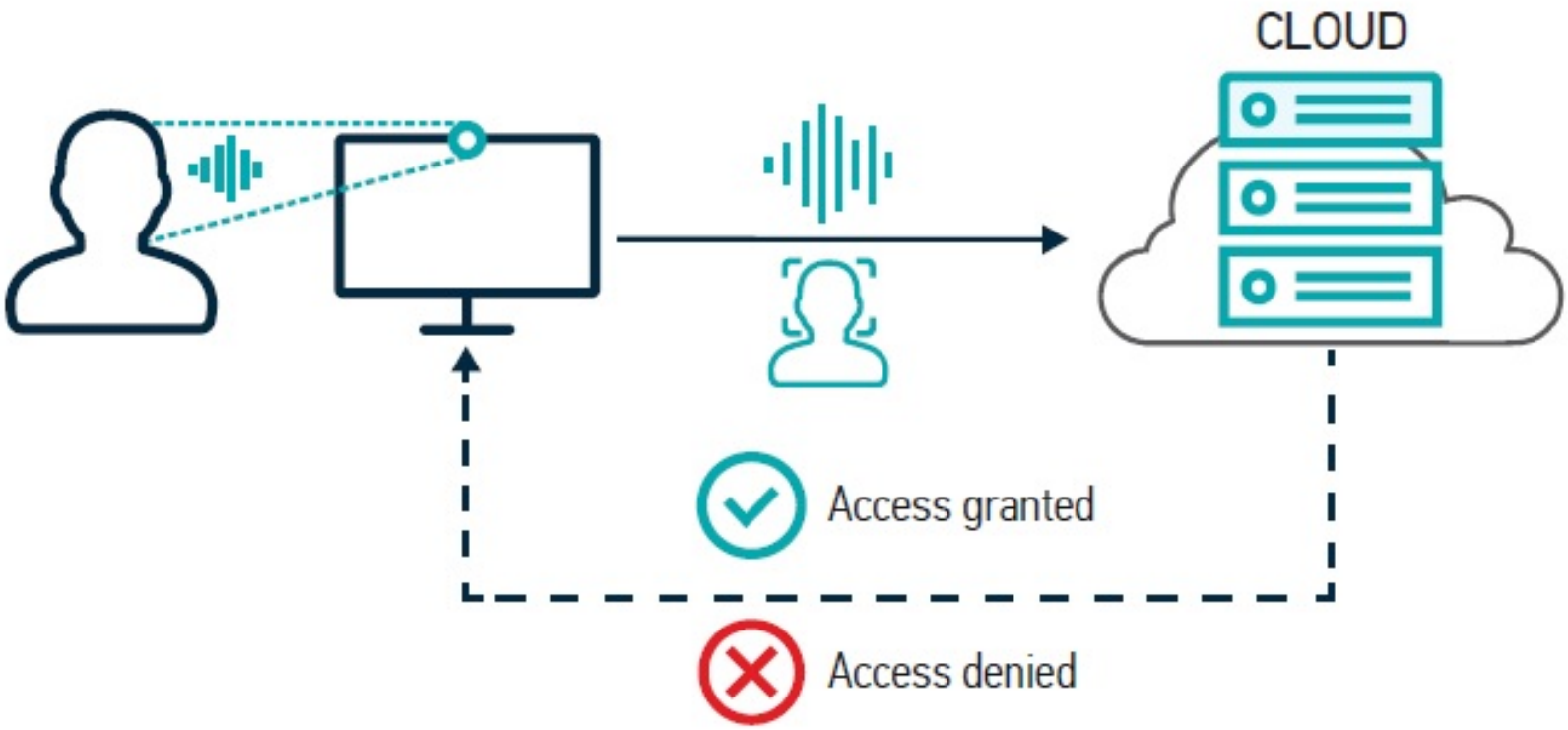
Fig. 8.4. Prosodic token alphabet (top table) and sample tokenization of pitch and energy contours (bottom figure).

Hình 6: Handbook of Biometrics, page 163

Công trình tiêu biểu

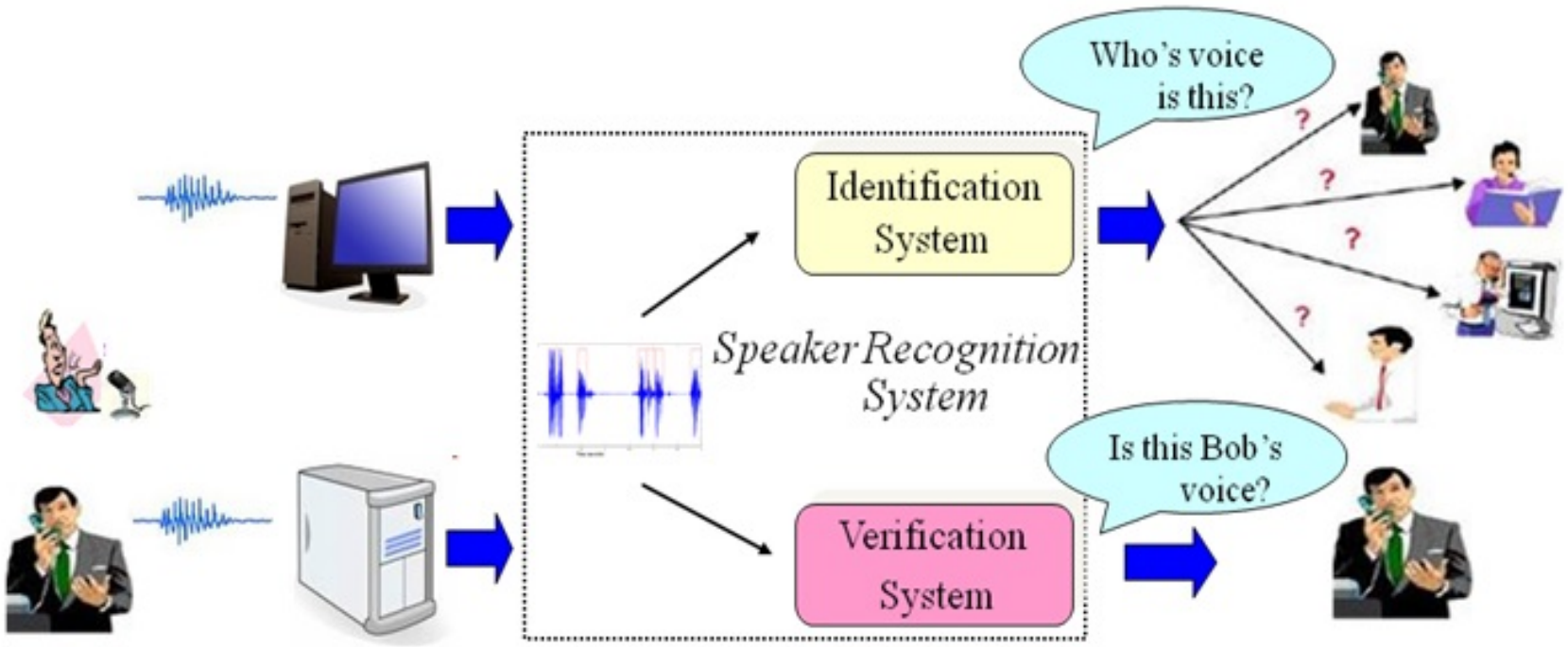
- Introduction 1
- Introduction 2
- Introduction 3

Voice authentication



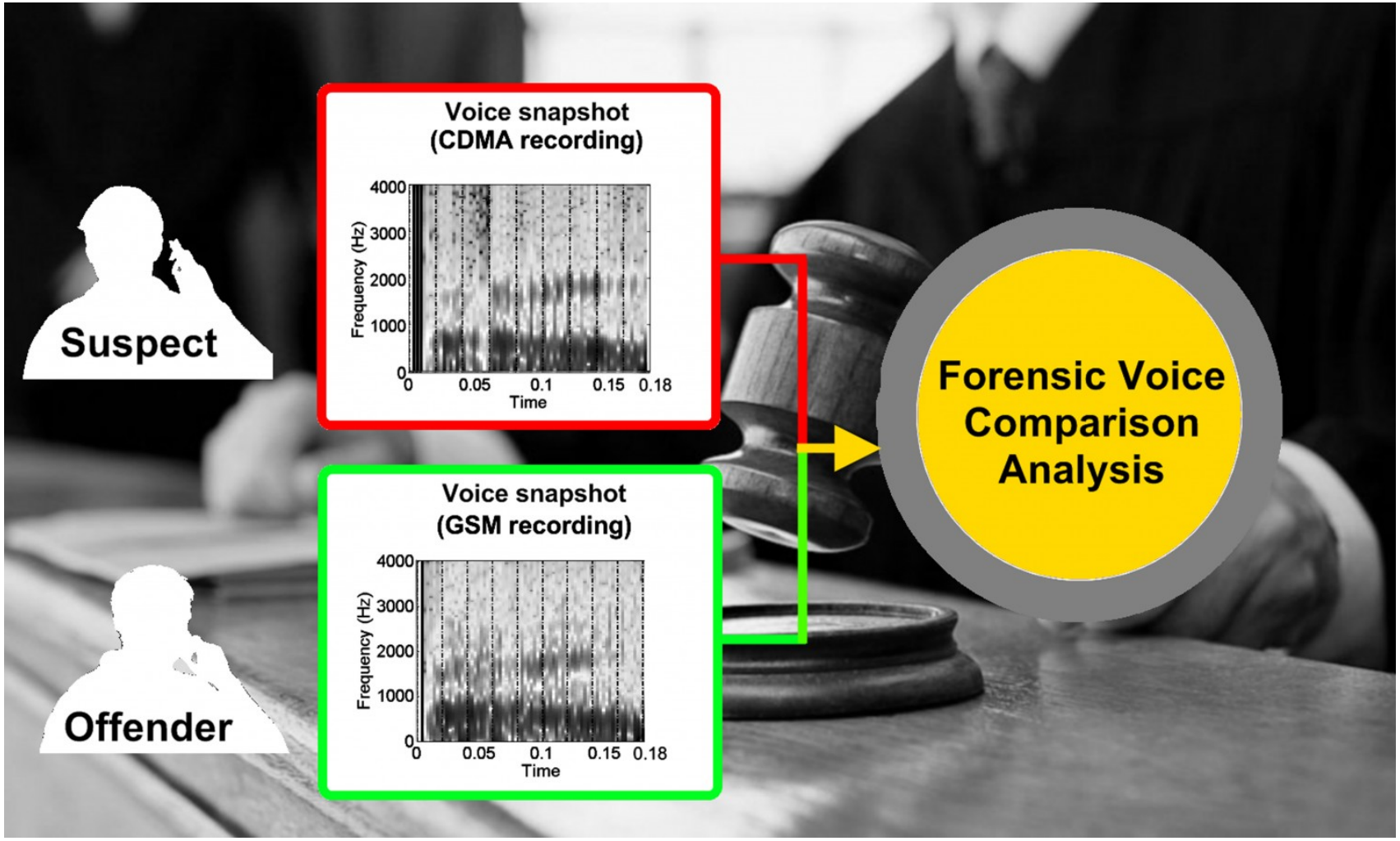
Hình 7: Ví dụ Voice authentication/ Verification

Speaker Identification and Verification



Hình 8: Ví dụ Speaker Recognition Systems

Forensic speaker recognition



Hình 9: Ví dụ Speaker Recognition Systems

Động lực nghiên cứu khoa học

- Những phương pháp đã tìm hiểu từ sách Handbook of Biometrics: Voice Biometrics đã cho chúng ta cái nhìn tổng quan về lĩnh vực Nhận dạng giọng nói và những phương pháp truyền thống (tạm gọi là thời kỳ trước Deep Learning) cùng với những thông tin các công trình nghiên cứu nổi bật.
- Các phương pháp SOTA dựa trên việc biểu diễn i-vectors của những đoạn giọng nói, cải thiện đáng kể so với mô hình Gaussian Mixture Model-Universal Background Models
- Sự phát triển của Deep Learning

Phát biểu bài toán

Tác vụ: Định danh người nói

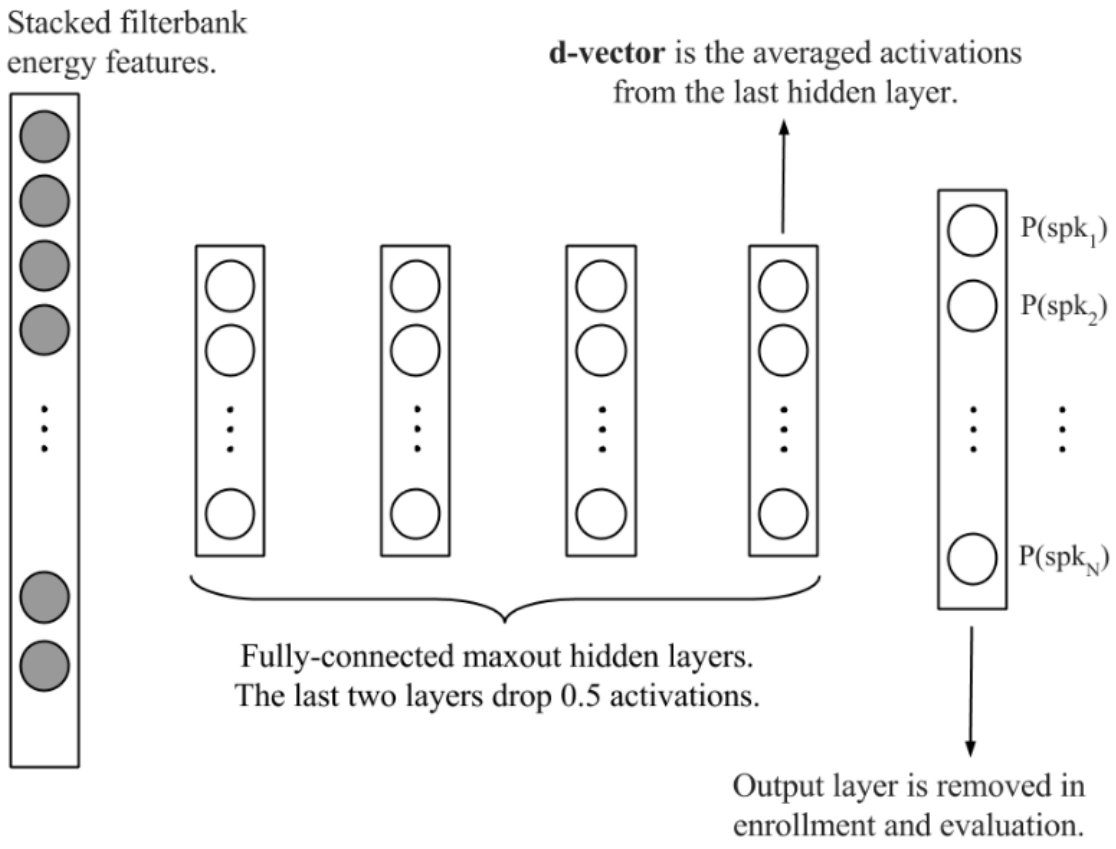
- Đầu vào (Input): Dữ liệu âm thanh giọng nói
- Đầu ra (Output): Danh tính của người nói

Tác vụ: Xác nhận người nói

- Đầu vào (Input): Dữ liệu âm thanh giọng nói
- Đầu ra (Output): Đồng ý/ Từ chối

Công trình tiêu biểu sử dụng d-vectors

Mô hình d-vectors



Hình 10: Mô hình DNN với d-vectors

Công trình tiêu biểu sử dụng d-vectors

Giới thiệu về công trình: Speaker Recognition from raw waveform with SincNet

Công trình tiêu biểu sử dụng d-vectors

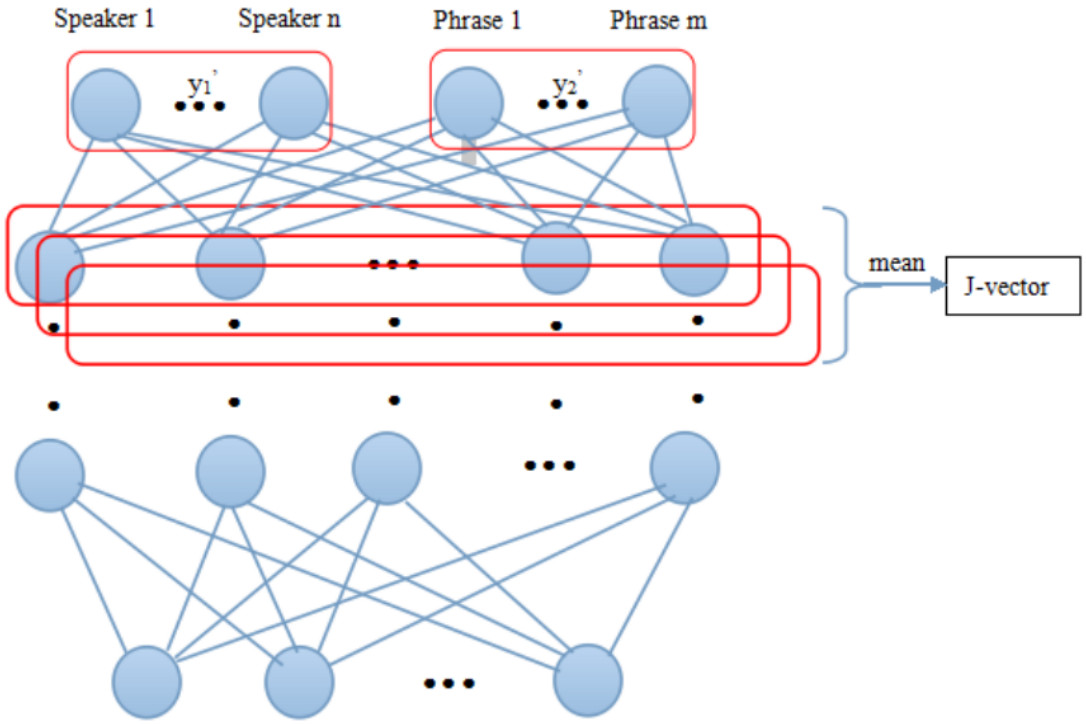
Phương pháp tiếp cận

Công trình tiêu biểu sử dụng d-vectors

Kết quả đạt được

Công trình tiêu biểu sử dụng j-vectors

Mô hình j-vectors



Hình 11: Mô hình DNN với j-vectors

Công trình tiêu biểu sử dụng j-vectors

Giới thiệu về công trình

Công trình tiêu biểu sử dụng j-vectors

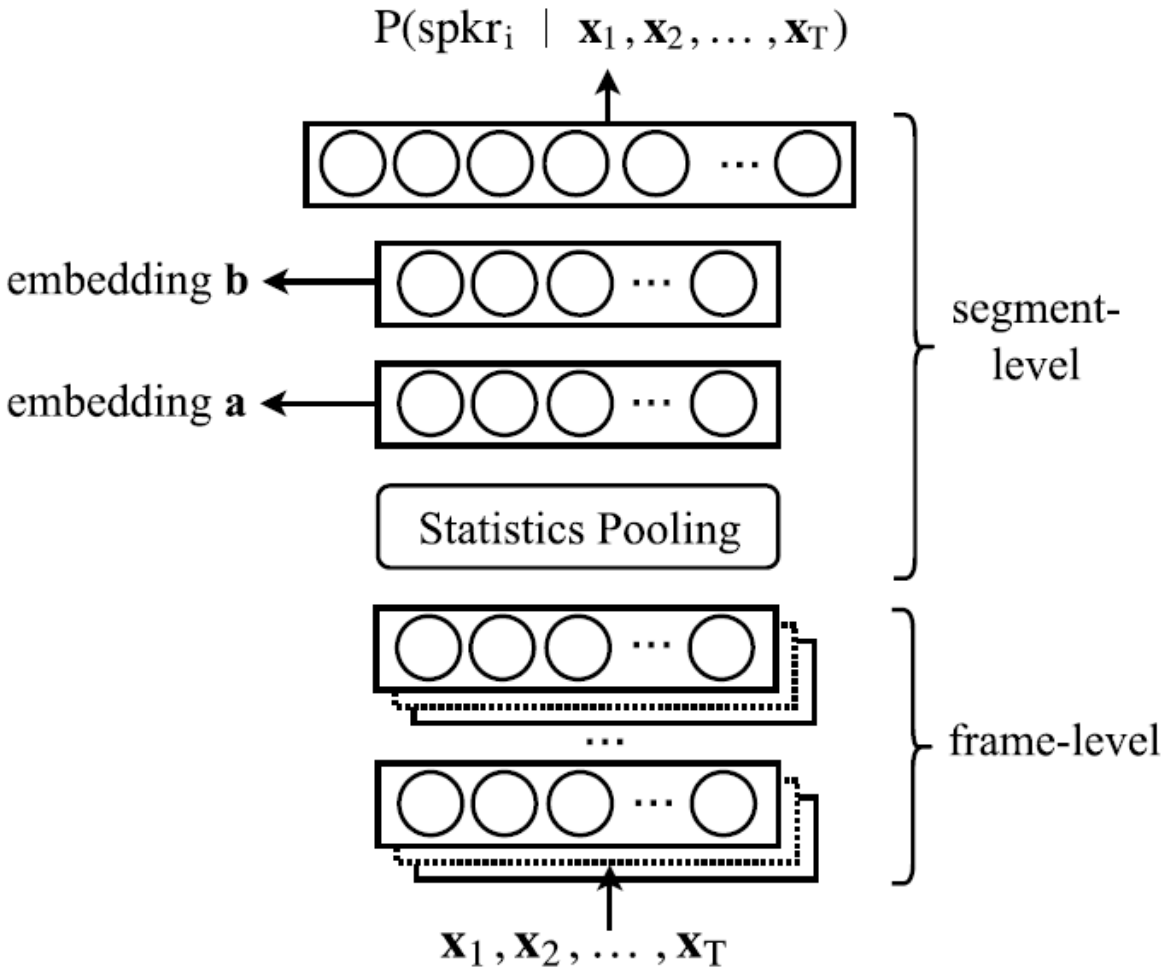
Phương pháp tiếp cận

Công trình tiêu biểu sử dụng j-vectors

Các kết quả đạt được

Công trình tiêu biểu sử dụng x-vectors

Mô hình x-vectors



Hình 12: Mô hình DNN với x-vectors

Công trình tiêu biểu sử dụng x-vectors

Giới thiệu về công trình

Công trình tiêu biểu sử dụng x-vectors

Phương pháp

Công trình tiêu biểu sử dụng x-vectors

Các kết quả đạt được

So sánh d-vectors, j-vectors và x-vectors

Thực nghiệm


Shervin Minaee, Amirali Abdolrashidi, Hang Su, Mohammed Bennamoun, and David Zhang.
Biometrics recognition using deep learning: A survey, 2021.


Mirco Ravanelli and Yoshua Bengio.
Speaker recognition from raw waveform with sincnet, 2019.


Dávid Sztahó, György Szaszák, and András Beke.
Deep learning methods in speaker recognition: a review, 2019.