Bài toán: Nhận dạng nguyên âm không phụ thuộc người nói.

Input:

tín hiệu tiếng nói (chứa 01 nguyên âm và khoảng lặng) của tập kiểm thử (gồm 21 người, 105 file test).

Output:

- Kết quả nhận dạng (dự đoán) nhãn nguyên âm của mỗi file test (/a/, …,/u/), Đúng/Sai (dựa vào nhãn tên file).
- Xuất 05 vector đặc trưng của 05 nguyên âm.
- Kết quả độ chính xác nhận dạng tổng hợp (%) theo số chiều của vector đặc trưng N (=13, 26, 39).
- Ma trận nhầm lẫn (confusion matrix) theo số chiều của vector đặc trưng N: thống kê số lần nhận dạng đúng/sai của mỗi cặp nguyên âm (có highlight nguyên âm đc nhận dạng đúng và bị nhận dạng sai nhiều nhất).

		Nhãn dự đoán				
		/a/	/e/	/i/	/o/	/u/
Nhãn đúng	/a/	Số lần /a/ nhận dạng đúng là /a/	Số lần /a/ nhận dạng sai thành /e/	Số lần /a/ nhận dạng sai thành /i/	Số lần /a/ nhận dạng sai thành /o/	Số lần /a/ nhận dạng sai thành /u/
	/e/	Số lần /e/ nhận dạng sai thành /a/	Số lần /e/ nhận dạng đúng là /e/			
	/i/	Số lần /i/ nhận dạng sai thành /a/		Số lần /i/ nhận dạng đúng là /i/		
	/o/				Số lần /o/ nhận dạng đúng là /o/	
	/u/					Số lần /u/ nhận dạng đúng là /u/

Được sử dụng: tín hiệu tiếng nói (chứa 01 nguyên âm và khoảng lặng) của tập huấn luyện (gồm 21 người, 105 file huấn luyện).

Yêu cầu:

Cài đặt BT nhận dạng theo mô hình tương tự như BT tìm kiếm âm thanh (trong TLTK [4]) gồm 3 thuật toán:

1. Phân đoạn tín hiệu thành nguyên âm và khoảng lặng (đã làm thi GK và CK): chọn thuật toán nhóm đã cài đặt có ĐCX cao nhất.
2. Trích xuất vector đặc trưng đường bao phổ (spectral envelope) của 05 nguyên âm dựa trên tập huấn luyện (gồm 21 người, 105 file huấn luyện):
    a. Đánh dấu vùng có đặc trưng phổ ổn định đặc trưng cho nguyên âm: chia vùng nguyên âm thành 3 phần bằng nhau và lấy đoạn nằm giữa (gồm M khung).
    b. Trích xuất vector MFCC (mel-frequency cepstral coefficients) của 1 khung tín hiệu với số chiều (chính là số lượng hệ số MFCC) là N (=13, 26, 39), dùng các hàm của thư viện Voicebox (Matlab) hoặc libbrosa (python).
    c. Tính vector đặc trưng cho 1 nguyên âm của 1 người nói = Trung bình cộng của M vector MFCC của M khung thuộc vùng ổn định.
    d. Tính vector đặc trưng cho 1 nguyên âm của nhiều người nói = Trung bình cộng của các vector đặc trưng cho 1 nguyên âm của 21 người nói (trong tập huấn luyện).

3. So khớp vector MFCC của tín hiệu nguyên âm đầu vào (thuộc tập kiểm thử) với 5 vector đặc trưng đã trích xuất của 5 nguyên âm (dựa trên tập huấn luyện) để đưa ra kết quả nhận dạng nguyên âm: tính 5 khoảng cách Euclidean giữa 2 vector và đưa ra quyết định nhận dạng dựa trên k/c nhỏ nhất (SV tự cài đặt).

Phần nâng cao (dành cho các nhóm SV muốn làm thêm để cải thiện độ chính xác nhận dạng và nhận điểm tối đa):

- Mục 2c và 2d: Nếu chỉ tính 1 vector đặc trưng cho 1 nguyên âm của nhiều người nói thì độ chính xác biểu diễn không cao do các người nói có chất giọng ít/nhiều khác nhau → làm giảm độ chính xác nhận dạng. Do đó, có thể tăng độ chính xác biểu diễn bằng cách tính K vector đặc trưng cho 1 nguyên âm của nhiều người nói dùng thuật toán phân cụm K-trung bình (K-mean clustering) với K=2,3,4,5. Chạy K-mean clustering trên tất cả các vector MFCC của các khung nằm trong phần ổn định của 1 nguyên âm của 21 người trong tập huấn luyện để thu được K vector trung bình làm K vector đặc trưng cho 1 nguyên âm.
- Mục 3: So khớp vector MFCC của tín hiệu nguyên âm đầu vào (thuộc tập kiểm thử) với 5xK vector đặc trưng đã trích xuất của 5 nguyên âm (dựa trên tập huấn luyện) để đưa ra kết quả nhận dạng nguyên âm: tính 5xK khoảng cách Euclidean giữa 2 vector và đưa ra quyết định nhận dạng dựa trên k/c nhỏ nhất (SV tự cài đặt).
- Thuật toán phân cụm K-trung bình: SV có thể tự cài đặt hoặc dùng hàm thư viện có sẵn.
- Lập bảng báo cáo kết quả độ chính xác nhận dạng tổng hợp (%) theo số chiều của vector đặc trưng N và số cụm K.

TLTK:

[1] Bài TH 1-Frequency-domain processing.pdf (phần 1 về spectrogram).
[2] Spectrogram-Cepstrum-and Mel-Frequency Analysis_CMU_PPT (slide 16-49 về đường bao phổ và các hệ số MFCC).
[3] Phân tích formant các nguyên âm của nhiều người nói_Luận văn_PĐThiện_2021 (phần 2.2.1 về khái niệm formant, phần 3.5 và 3.6.2 về một số kết quả đo để tham khảo về dải giá trị của F1,F2 và F3 của các nguyên âm).
[4]CITA_SoSanhPhuongPhapDuongBaoPhovaPhuongPhapAnhPhoTrongTimKiemAmNhac_2021 (phần 2 và 3.1 về mô hình tìm kiếm/nhận dạng và thuật toán phân cụm K-trung bình).

Kiểm tra và đánh giá (trọng số điểm 60%):

GV kiểm tra chương trình và báo cáo của mỗi nhóm trong tuần sau. Các SV trong nhóm tự phân công nhiệm vụ. Khi GV chấm bài, mỗi SV trình bày riêng phần mình làm trên slide CHUNG của cả nhóm.

Chú ý:

- SV chỉ báo cáo cách làm step-by-step và kết quả (bằng hình ảnh và số liệu cụ thể), KHÔNG báo cáo lý thuyết.
- Nộp Slide và Code CHUNG của cả nhóm 10 phút đầu buổi thi.
- Trình bày slide: 10 phút
- Demo và hỏi/đáp: 5 phút

Thời gian kiểm tra trên Team riêng theo lớp HP:

- 1910A: 13h30 thứ bảy 22/1.
- 1911A: 07h30 thứ bảy 22/1.
- 1912A: 13h30 thứ sáu 21/1.
- 1913A: 07h30 thứ sáu 21/1.
- 1914A: 13h30 thứ năm 20/1.
- 1915A: 07h30 thứ năm 20/1.