

**ĐẠI HỌC QUỐC GIA TP.HỒ CHÍ MINH**  
**TRƯỜNG ĐẠI HỌC BÁCH KHOA**  
**KHOA ĐIỆN-ĐIỆN TỬ**  
**BỘ MÔN KỸ THUẬT ĐIỆN TỬ**



# **LUẬN VĂN TỐT NGHIỆP**

**HỆ THỐNG NHẬN DẠNG NGƯỜI BẰNG GIỌNG NÓI VÀ KHUÔN MẶT**

**GVHD: ThS. Hồ Trung Mỹ**

**SV: Lê Tiến Đạt**

**MSSV: 1710948**

**SV: Võ Mai Trí Luận**

**MSSV: 1712083**

**TP.Hồ Chí Minh 12/08/2021**



# Nội dung

---

- 1. Giới thiệu tổng quan**
- 2. Nhiệm vụ đề tài**
- 3. Phát triển phần mềm**
- 4. Đặc tả hệ thống**
- 5. Thiết kế phần cứng**
- 6. Kết quả thực hiện**
- 7. Kết luận và hướng phát triển**



# 1. Giới thiệu tổng quan

## PART 1 GIỚI THIỆU TỔNG QUAN





# 1. Giới thiệu tổng quan

- Định nghĩa sinh trắc học: là công nghệ sử dụng những thuộc tính vật lý, đặc điểm sinh học riêng của mỗi cá nhân như vân tay, khuôn mặt, móng mắt,... để nhận diện, xác thực bảo mật.
- Định nghĩa nhận dạng giọng nói: là phân chia và đánh nhãn ngôn ngữ cho tín hiệu tiếng nói
- Định nghĩa nhận dạng khuôn mặt: là tự động xác định hoặc nhận dạng một người từ một bức ảnh kỹ thuật số hoặc một khung hình video



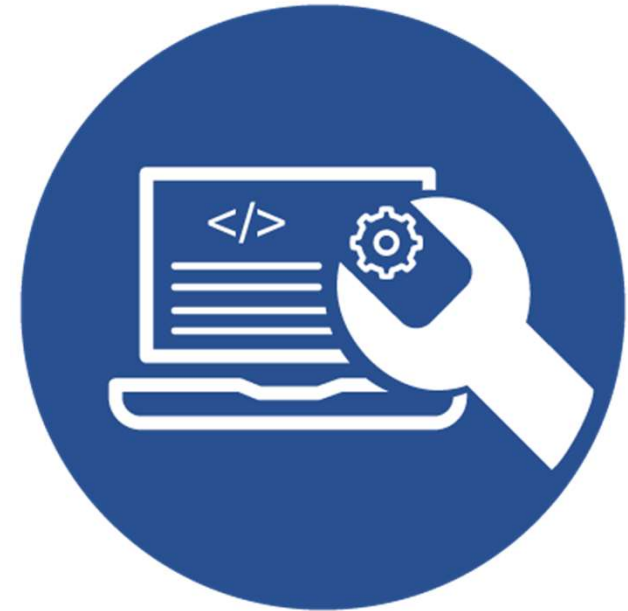


# 1. Giới thiệu tổng quan

---

## Ứng dụng:

- ☐ Bảo mật, ngân hàng
- ☐ Điều khiển và giao tiếp không dây
- ☐ Quảng cáo thông minh
- ☐ Pháp y, y tế
- ☐ ...





## 2. Nhiệm vụ đề tài

---

### Nhiệm vụ, công việc cần làm:

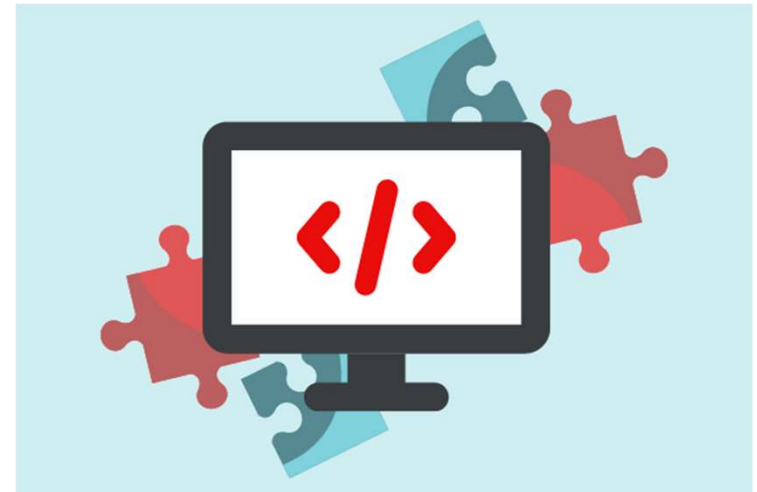
- Tìm tài liệu tham khảo về lý thuyết nhận dạng giọng nói và khuôn mặt
- Tìm hiểu các giải thuật, đánh giá, lựa chọn và thực hiện
- Thu thập dữ liệu cho luận văn, xây dựng bộ từ vựng nhỏ (10 từ).
- Tìm hiểu, thiết kế và thực hiện phần cứng
- Mô phỏng giải thuật trên MATLAB và thực thi giải thuật trên phần cứng



## 3. Phát triển phần mềm

---

# PART 3 PHÁT TRIỂN PHẦN MỀM





## 3. Phát triển phần mềm

---

NHẬN DẠNG GIỌNG NÓI:

- **TRÍCH XUẤT ĐẶC TRƯNG GIỌNG NÓI**
- **HUẤN LUYỆN CÁC VECTOR ĐẶC TRƯNG**
- **NHẬN DẠNG GIỌNG NÓI**





## 3.1 TRÍCH XUẤT ĐẶC TRƯNG

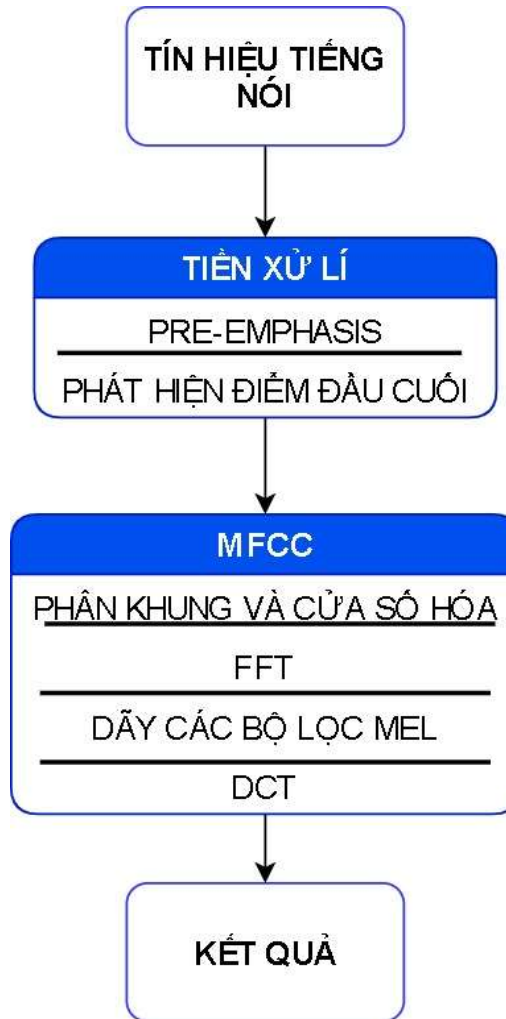
### TRÍCH XUẤT ĐẶC TRƯNG GIỌNG NÓI:

- **MFCC - Mel frequency cepstral coefficient**

Ưu điểm	Nhược điểm
<ul style="list-style-type: none"><li>• Loại bỏ những thông tin quá chi tiết tập trung vào những thông tin về cấu trúc âm thanh.</li><li>• Dễ hiểu và tính toán dễ dàng.</li><li>• Được sử dụng rộng rãi và hiệu quả đã được kiểm chứng.</li></ul>	<ul style="list-style-type: none"><li>• Nhạy cảm với nhiễu</li><li>• Sự lựa chọn các bộ lọc tam giác thường là tự ý và không dựa trên một lập luận vững chắc nào</li></ul>



## 3.1 TRÍCH XUẤT ĐẶC TRƯNG



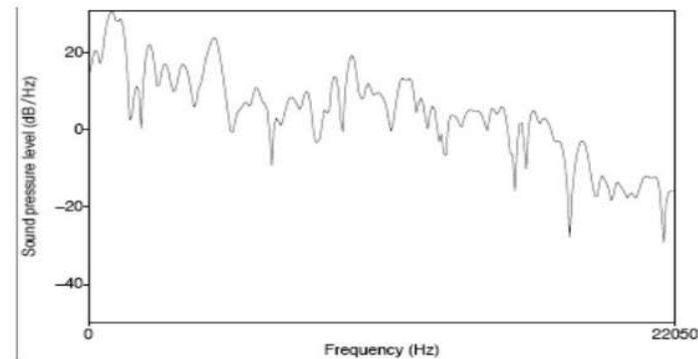
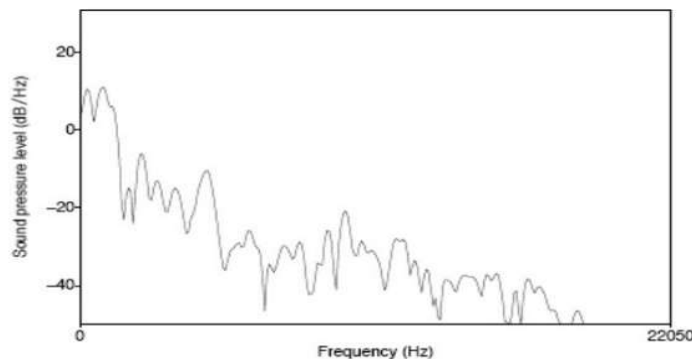
## 3.1 TRÍCH XUẤT ĐẶC TRƯNG

### B1: TIỀN NHẤN (PRE-EMPHASIS)

- Các âm ở tần số thấp có mức năng lượng cao.
- Các âm ở tần số cao lại có mức năng lượng khá thấp, chứa nhiều thông tin về âm vị.

$$x'[t_d] = x[t_d] - \alpha x[t_d - 1]$$

$$0.95 < \alpha < 0.99$$





## 3.1 TRÍCH XUẤT ĐẶC TRƯNG

---

- B2 : PHÁT HIỆN ĐIỂM ĐẦU CUỐI CỦA TỪ.
  - Tìm ra điểm bắt đầu và kết thúc của 1 từ trong tín hiệu.
  - Sử dụng phương pháp 2 ngưỡng dựa trên năng lượng trung bình ngắn hạn (short-term average energy) và trung bình tỷ lệ vượt qua điểm 0 (short-term average zero-crossing rate).



## 3.1 TRÍCH XUẤT ĐẶC TRƯNG

B2 : PHÁT HIỆN ĐIỂM ĐẦU CUỐI CỦA TỪ.

Năng lượng trung bình ngắn hạn:

- $x(n)$  : tín hiệu tiếng nói trong miền thời gian.
- $w(n)$ : hàm cửa sổ.
- $inc$  : độ chồng lấn khung.
- $fn$  : tổng số khung của tín hiệu.

$$y_i(n) = w(n) * x((i-1) * inc + n), 1 \leq n \leq L, 1 \leq i \leq fn$$

$$E(i) = \sum_{n=0}^{L-1} y_i^2(n)$$



## 3.1 TRÍCH XUẤT ĐẶC TRƯNG

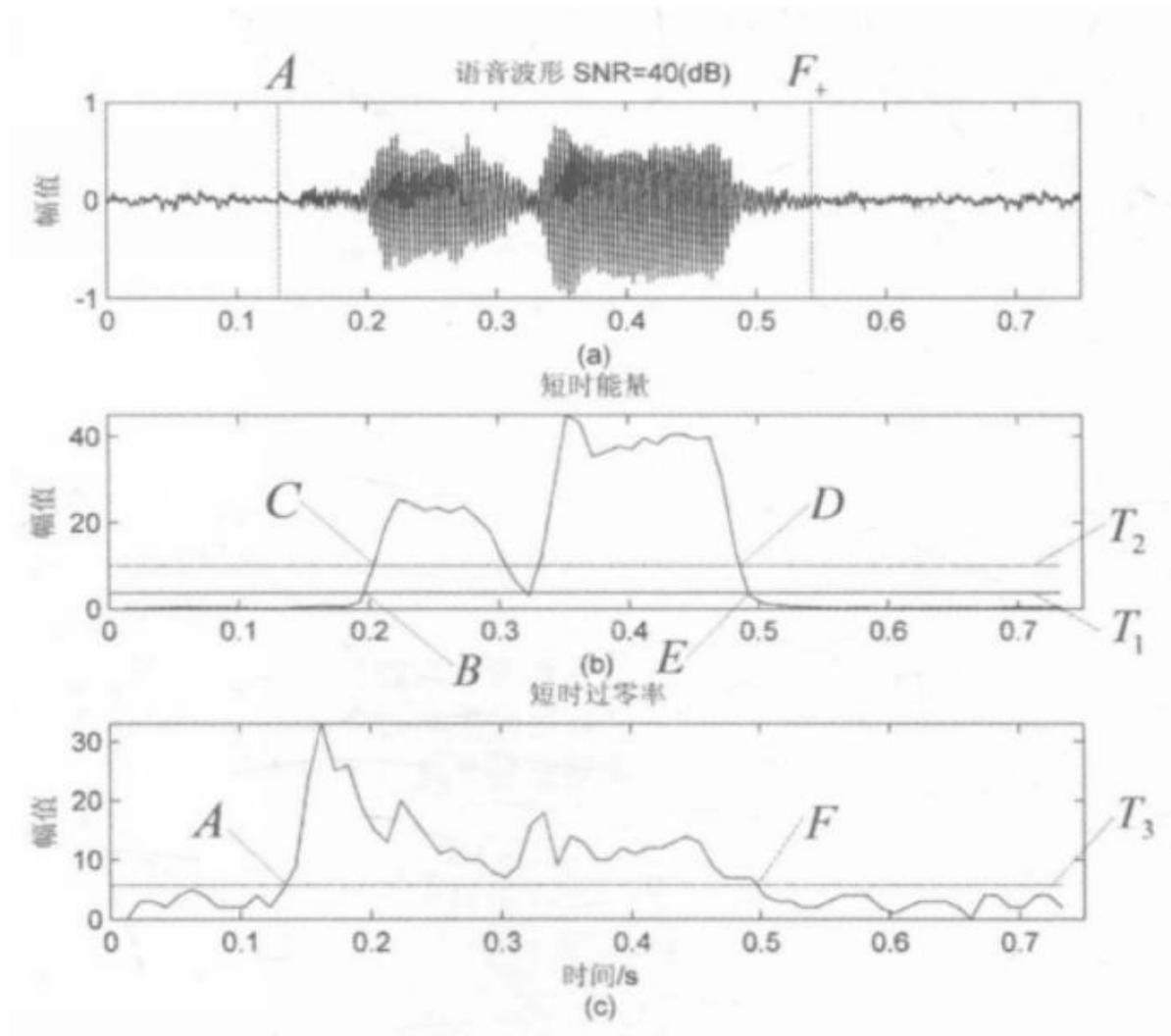
Trung bình tỉ lệ vượt qua điểm 0 :

- Tỷ lệ vượt mức 0 trung bình ngắn hạn cho biết số lần dạng sóng tín hiệu tiếng nói vượt qua trục hoành (điểm 0) trong khung tín hiệu tiếng nói

$$z(i) = \frac{1}{2} \sum_{n=0}^{L-1} |sgn[yi(n)] - sgn[yi(n-1)]|$$

$$sgn[x] = \begin{cases} 1 & x \geq 0 \\ -1 & x < 0 \end{cases}$$

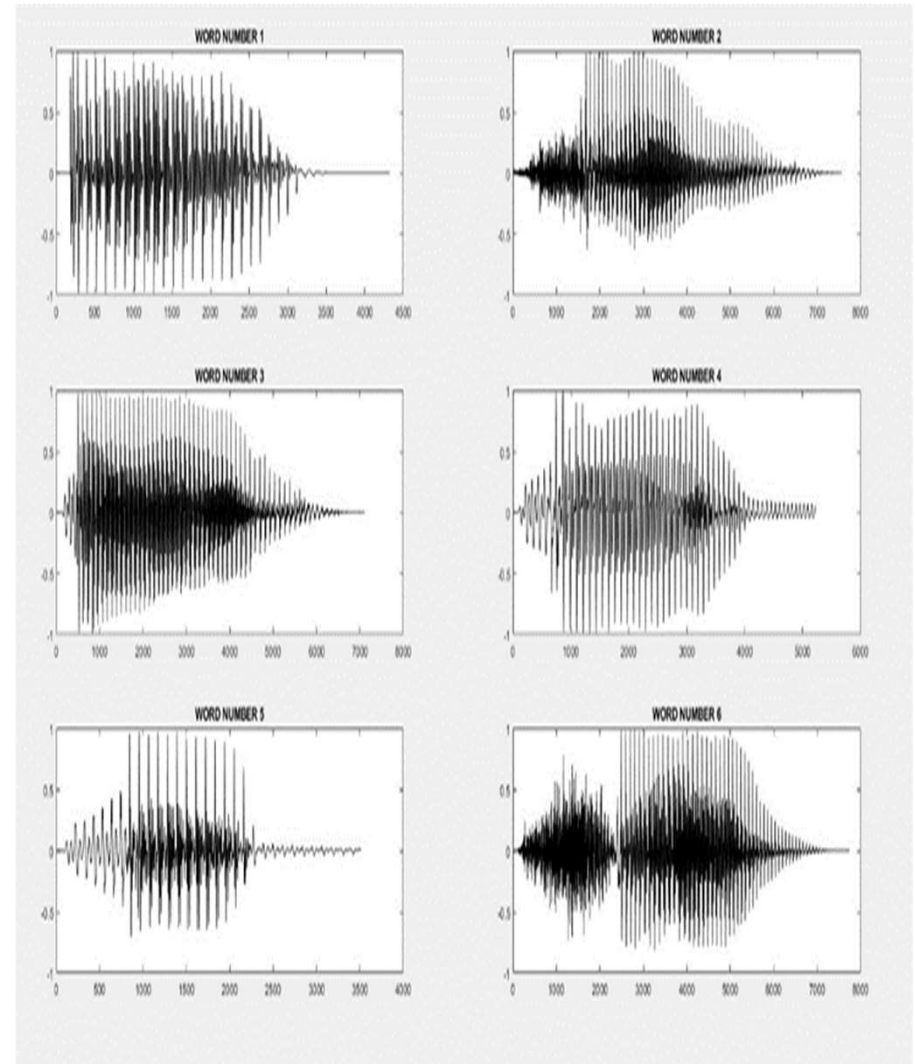
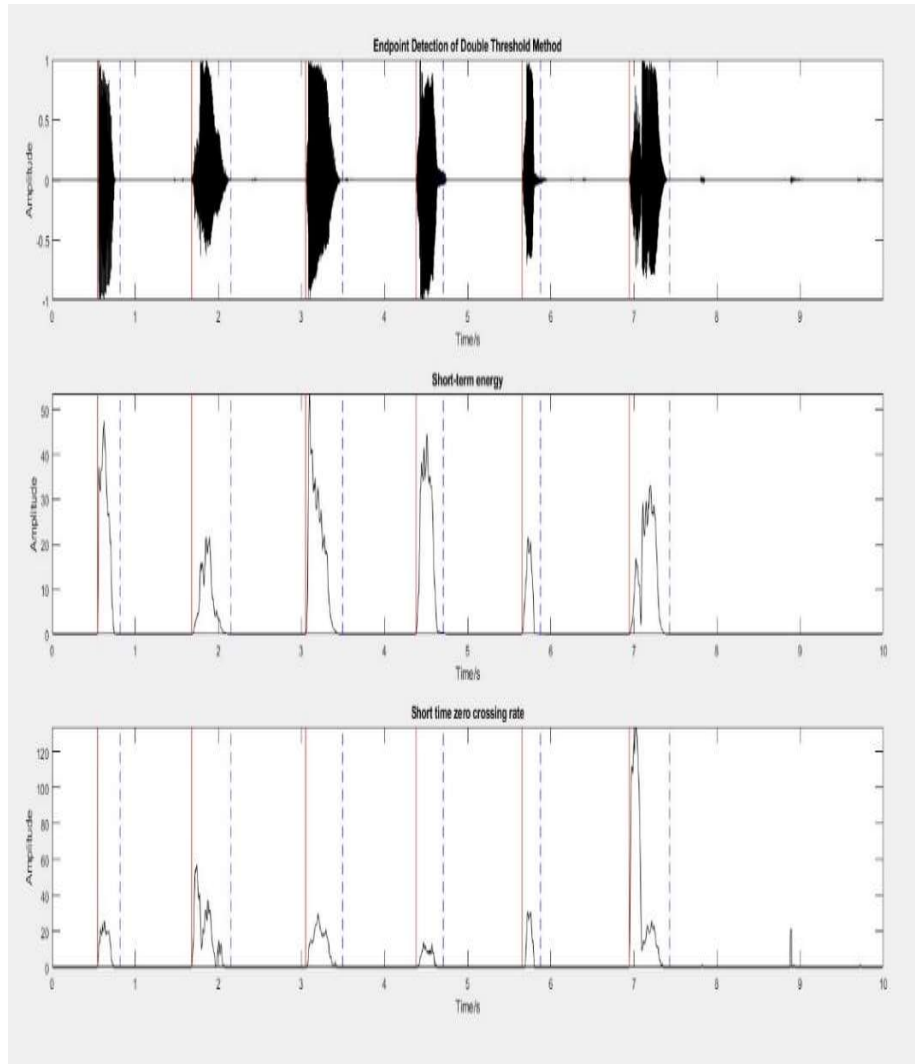
# 3.1 TRÍCH XUẤT ĐẶC TRƯNG



- $T_1 = 0.5$
- $T_2 = 1.5$
- $T_3 = 0.7$



# 3.1 TRÍCH XUẤT ĐẶC TRƯNG





## 3.1 TRÍCH XUẤT ĐẶC TRƯNG

- B3 : Phân khung và cửa sổ hóa.
  - Tín hiệu được phân thành những khung, mỗi khung N mẫu, độ chồng lấp overlap có M mẫu: M thường xấp xỉ  $1/3N$ , với  $N = 256$  để dễ cho việc tính toán FFT.





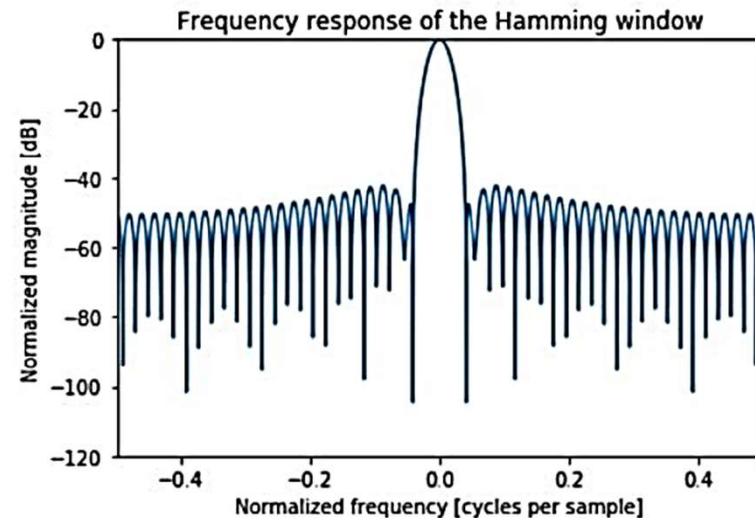
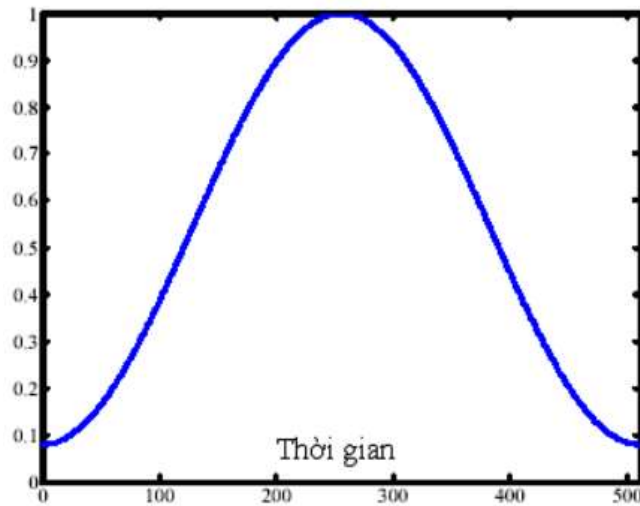
## 3.1 TRÍCH XUẤT ĐẶC TRƯNG

- Cửa sổ hóa :
  - Giảm sự không liên tục giữa điểm đầu và điểm cuối của mỗi khung frame.
  - Gọi cửa sổ là  $w(n)$ ,  $0 \leq n \leq N - 1$ , với  $N$  là số phần tử trong mỗi frame. Tín hiệu sau khi cửa sổ hóa như sau:

$$y_i(n) = x_i(n) * w(n) \quad (0 \leq n \leq N - 1)$$

## 3.1 TRÍCH XUẤT ĐẶC TRƯNG

- Cửa sổ Hamming :



$$w(n) = 0.54 - 0.46 \times \cos\left(\frac{2\pi n}{N}\right) ; n = 0..N - 1$$



## 3.1 TRÍCH XUẤT ĐẶC TRƯNG

- B4 : Fast Fourier Transform
  - Chuyển đổi tín hiệu miền thời gian sang miền tần số

$$X[k] = \sum_{n=0}^{N-1} x[n] \exp\left(-j \frac{2\pi}{N} kn\right)$$



## 3.1 TRÍCH XUẤT ĐẶC TRƯNG

B5 : Dãy bộ lọc Mel

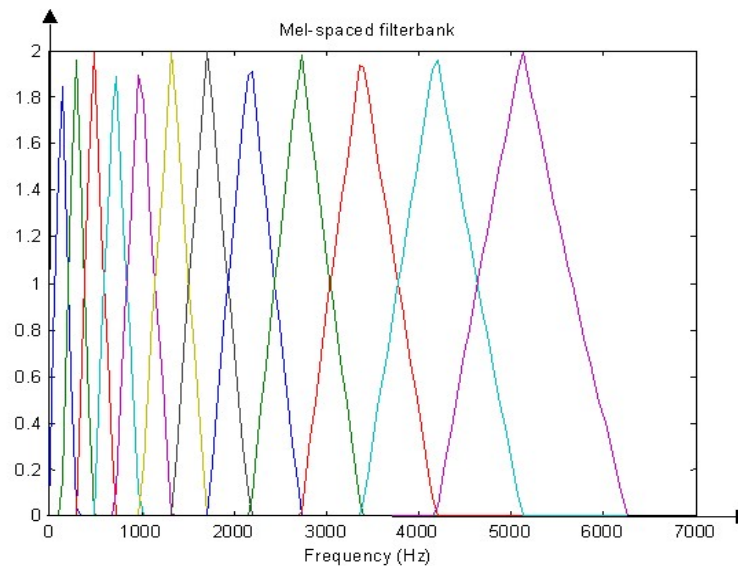
- chuyển giữa Hert ( $f$ ) và Mel ( $m$ ):
  - $\text{mel}(f) = 2595 * \log(1 + f / 700)$
  - $f = 700(10^{m/2595} - 1)$

$$\begin{aligned} H_m(k) &= 0 && \text{iff } k < f(m-1) \\ &= \frac{k - f(m-1)}{f(m) - f(m-1)} && \text{iff } f(m-1) \leq k < f(m) \\ &= 1 && \text{iff } k = f(m) \\ &= \frac{f(m+1) - k}{f(m+1) - f(m)} && \text{iff } f(m) < k \leq f(m+1) \\ &= 0 && \text{iff } k > f(m+1) \end{aligned}$$

## 3.1 TRÍCH XUẤT ĐẶC TRƯNG

- Xác định mel – spectrum, cho biên độ phổ tần số sau bước FFT ở trên qua bộ lọc mel, với công thức tính như sau:

$$\tilde{S}(l) = \sum_{k=0}^{N/2} X(k) * M_i(k)$$





## 3.1 TRÍCH XUẤT ĐẶC TRƯNG

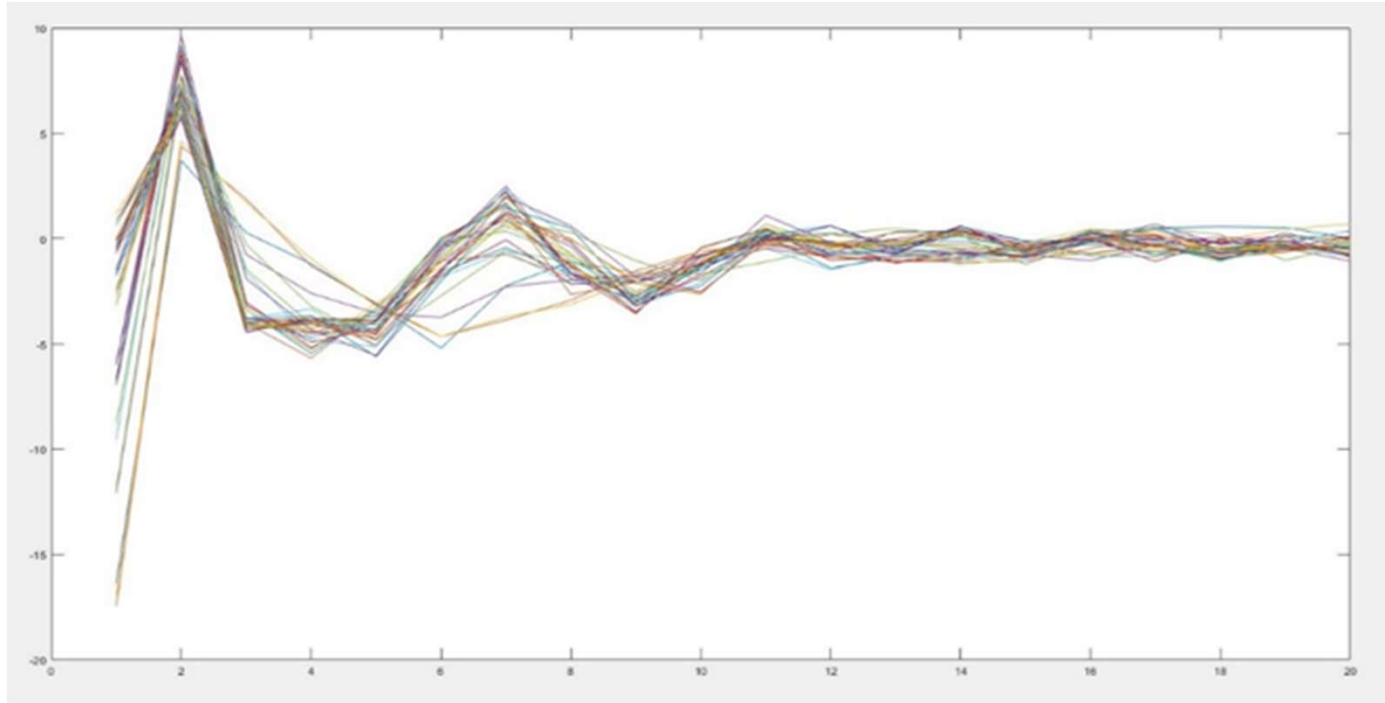
B6 :Cepstrum:

- Chuyển đổi logarit của mel spectrum về miền thời gian, kết quả được gọi là mel frequency cepstrum coefficients (MFCC)
- Ở bước này dùng DCT (Discrete cosin transform – biến đổi cosin rời rạc)

$$\hat{C} = \sum_{k=1}^K \log(\hat{S}_k) * \cos\left[n * \left(k - \frac{1}{2}\right) \frac{\pi}{K}\right]$$

$$n=0,1,\dots,K-1$$

## 3.1 TRÍCH XUẤT ĐẶC TRƯNG



Acoustic vectors của từ “Một”



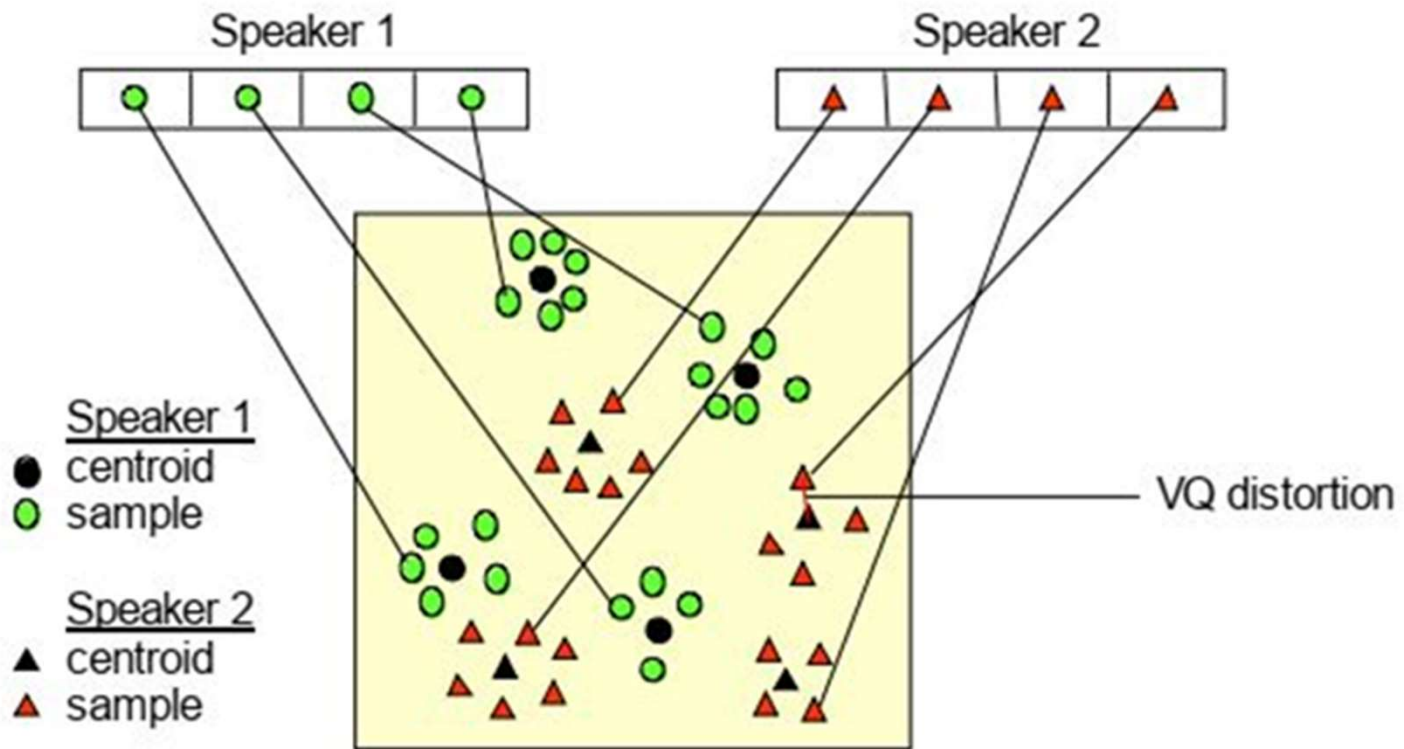
## 3.2 HUẤN LUYỆN

- **Lượng tử vector ( VQ-Vector Quantization).**
  - VQ là phương pháp ánh xạ những vector trong một không gian lớn thành một số lượng hữu hạn các vector.

Ưu điểm	Khuyết điểm
<ul style="list-style-type: none"><li>• Giảm thiểu lượng dữ liệu lưu trữ</li><li>• Giảm thời gian tính toán độ giống nhau giữa các vector phổ</li><li>• Biểu diễn rời rạc về mặt âm học của tiếng nói</li></ul>	<ul style="list-style-type: none"><li>• Việc lượng tử vector chắc chắn dẫn đến sai số lượng tử hóa</li><li>• Việc chọn kích thước codebook cho VQ không đơn giản</li></ul>

## 3.2 HUẤN LUYỆN

- VQ – Vector Quantization:

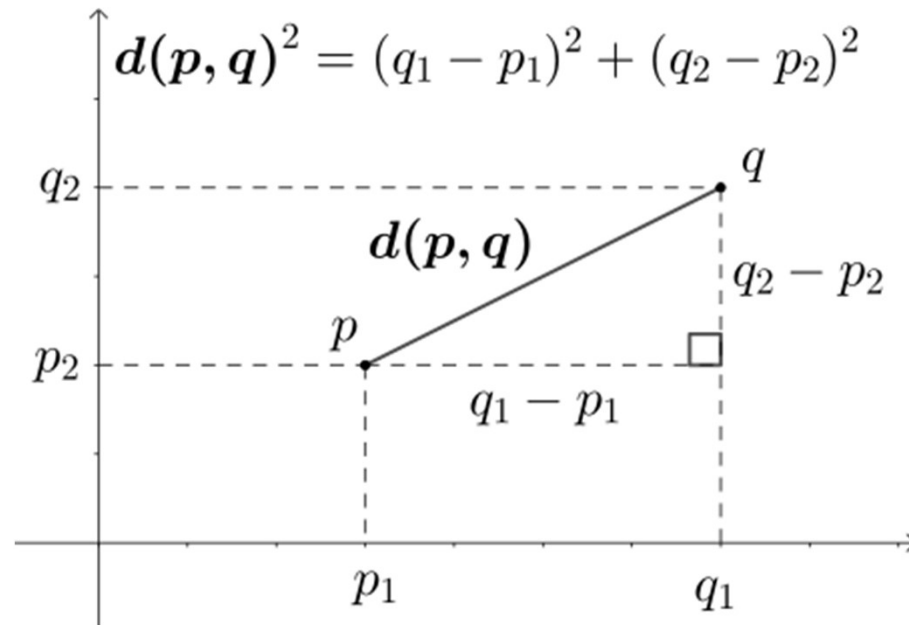


## 3.2 HUẤN LUYỆN

- Khoảng cách Euclid :

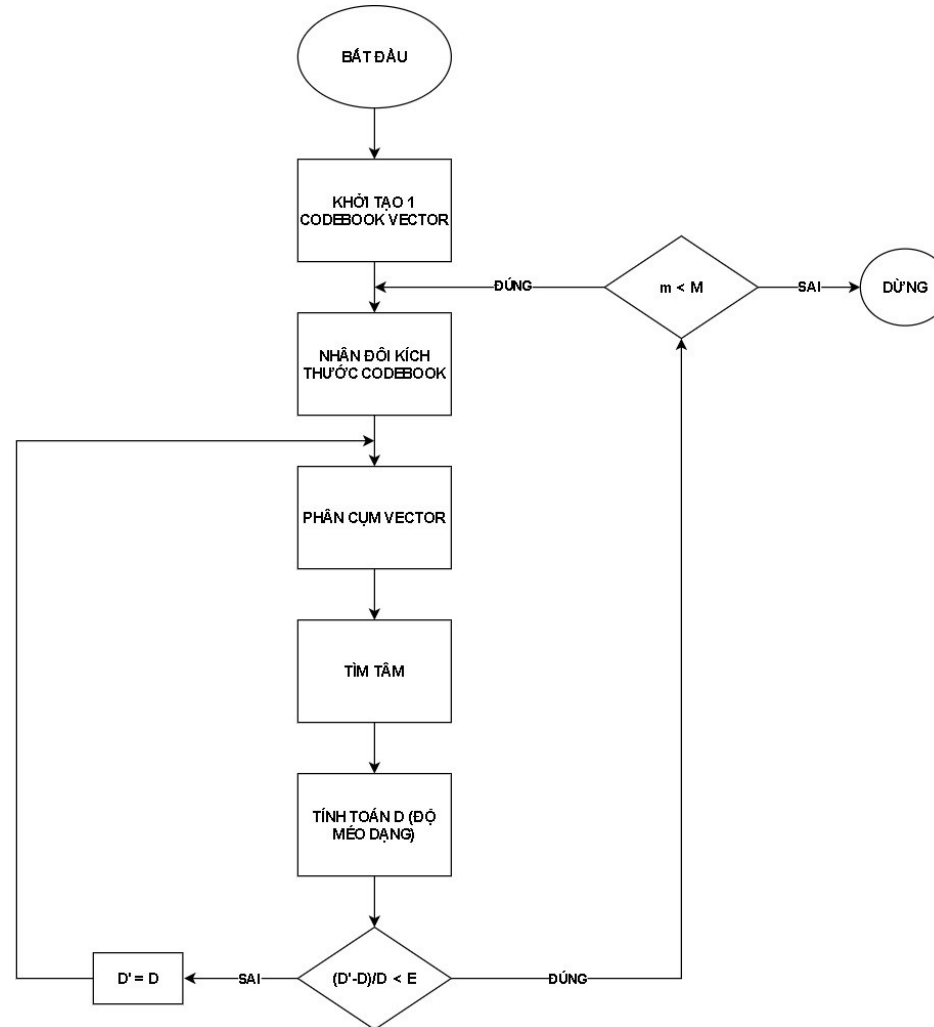
$$l = \sqrt{(b_1 - a_1)^2 + (b_2 - a_2)^2 + \dots + (b_n - a_n)^2}$$

- VD trong không gian 2D:



## 3.2 HUẤN LUYỆN

- LBG (Linde, Buzo & Gray, 1980)



## 3.3 NHẬN DẠNG

- Tính trung bình khoảng cách đến từng codeword trong codebook:

$$D_Q(X, C) = \frac{1}{T} \sum_{i=1}^T d_q(x_i, C)$$

**Trong đó :  $d_q(x_i, C) = \min d(x_i, c_j)$ .**

- Tìm ra từ có khoảng cách trung bình nhỏ nhất.



## 3.3 NHẬN DẠNG

---

- Thực hiện so sánh khoảng cách nhỏ nhất với ngưỡng để kiểm tra xem người nói có phải chính chủ hay không.
- Phương pháp đặt ngưỡng cho mỗi số:

## 3.3 NHẬN DẠNG

	Người 1	Người 2	Người 3	Người 4	Người 5	Người 6	Người 7	Người 8	Người 9
<b>Số 0</b>	15.4147 4266	18.9229 5762	15.8128 6256	15.4156 5584	15.6370 9608	19.7145 5184	19.3576 5198	16.0500 6931	16.7429 2398
<b>Số 1</b>	15.5605 7074	15.5228 6313	13.0628 5997	15.2301 0278	16.2204 102	14.0026 4675	12.8321 8126	13.7268 1987	11.4517 5156
<b>Số 2</b>	14.4388 6039	23.7796 97	15.8376 6771	15.5285 1754	18.4963 7392	15.9899 2353	16.8372 7752	18.0326 0528	16.0888 7528
<b>Số 3</b>	9.66744 9517	16.0929 8267	12.0866 1326	13.4684 1954	17.3324 422	12.1943 0592	16.0245 1582	13.5451 5841	15.1026 7093
<b>Số 4</b>	9.76098 4447	11.5950 6482	14.4779 3232	14.8771 702	11.6202 9577	12.0511 4167	13.0307 8879	15.9760 6923	13.9294 8427
<b>Số 5</b>	15.0185 5099	18.5316 2763	16.6080 4455	12.7980 4334	14.9944 4397	18.2034 6573	18.4876 6671	16.4459 2178	15.2925 5209
<b>Số 6</b>	12.4740 4805	16.9183 0886	13.9271 9869	13.3679 4655	15.0162 2485	14.2346 3601	14.6738 8807	13.1669 1818	14.9628 1554
<b>Số 7</b>	15.0136 2958	15.5681 4945	13.9472 5694	13.9434 1132	19.7734 7563	12.0151 1598	12.1484 879	13.8789 5161	13.4468 1764
<b>Số 8</b>	11.1714 9841	13.7390 8183	13.5267 6961	12.6129 1265	10.8478 784	16.0597 4478	12.4895 3103	14.7148 9099	13.3708 9957
<b>Số 9</b>	16.1901 7603	14.1475 3185	14.7700 6076	11.1798 5802	14.5319 9862	15.3404 8951	11.9661 6509	13.6537 0812	11.2168 2549

## 3.3 NHẬN DẠNG

Người 10(chính chủ)	Người 11	Người 12	Người 13	Người 14	Người 15	Người 16	Người 17	Người 18	Người 19	Người 20
8.2410796 47	12.013 69928	12.112 7066	10.866 34537	11.209 32506	10.056 04623	11.201 41602	11.343 32472	10.646 59749	9.6090 93362	13.131 32071
8.2445561 99	8.6005 82972	8.7374 55526	9.8001 07544	11.684 89786	10.508 60291	10.322 37134	9.3655 24374	11.815 37648	12.442 87071	9.2149 59728
7.4881997 42	11.592 68409	11.330 8309	9.3293 81851	9.8658 72959	9.8651 2916	11.585 36235	11.081 08461	11.758 80688	10.039 87224	12.157 49359
8.3086287 72	11.402 07637	9.3893 65364	10.757 80865	9.8914 82842	9.6979 96923	9.8828 85776	8.6717 83587	10.473 43352	8.3788 35429	12.190 51654
7.4736496 26	9.5439 65003	8.0102 37153	12.696 5971	8.9982 6301	7.5548 02562	8.0367 04988	8.4973 1339	9.7740 84302	8.4961 82224	9.1560 51422
8.6197216 83	12.025 96163	10.508 95676	10.821 36028	8.7396 34	10.964 32466	10.579 9341	9.1178 81206	10.149 46689	8.9154 3092	12.365 08833
8.5372540 45	13.322 52186	12.779 37849	12.445 68348	11.598 07541	9.8186 06102	10.020 60224	12.243 24851	12.844 68344	11.986 08441	12.331 59053
8.1398757 5	10.464 10112	10.150 13374	12.154 43704	10.054 86513	11.249 29545	12.306 11	9.2692 9868	10.237 47511	11.074 73463	11.832 48932
7.1868237	7.4206 98899	10.092 14174	9.9339 71901	9.0478 89999	7.4473 21164	9.3719 61173	8.9148 04973	7.5850 92332	7.4603 71394	10.421 79596
7.2264676 09	8.2749 52822	9.0793 19249	9.4704 95072	8.8519 80573	10.122 58168	8.8388 885	9.3128 85935	8.3345 21924	8.1559 27249	10.078 53864



## 3.3 NHẬN DẠNG

SỐ	NGŨỖNG
KHÔNG	8.6
MỘT	8.1
HAI	7.9
BA	6.5
BỐN	8.7
NĂM	9.0
SÁU	8.3
BẢY	8.6
TÁM	8.0
CHÍN	9.0



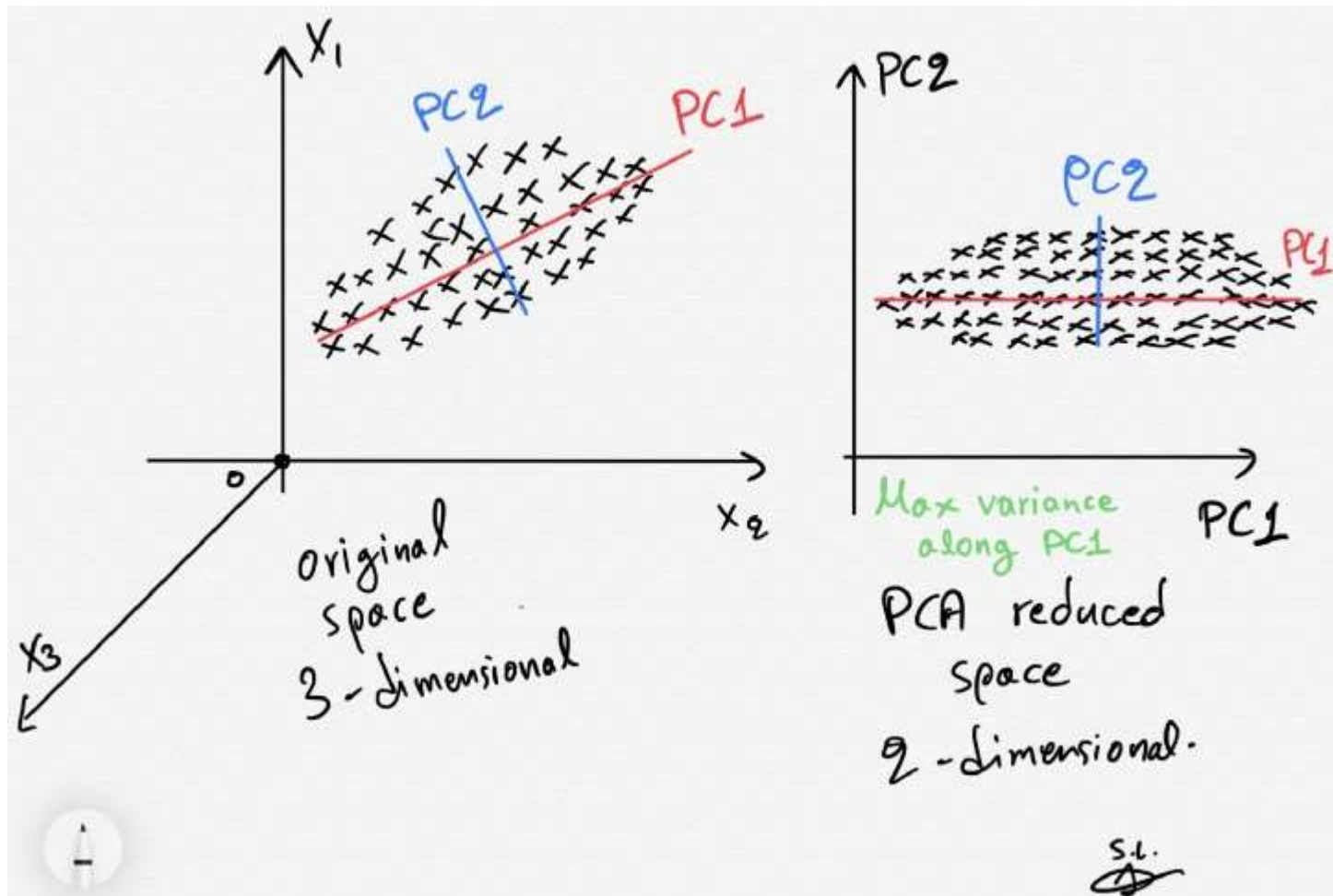
# 3. Thuật toán PCA

## Thuật toán PCA:

- Principal Component Analysis về cơ bản là một phương pháp giảm kích thước đơn giản, biến đổi các cột của bộ dữ liệu thành một tập các đặc trưng mới

Ưu điểm	Nhược điểm
<ul style="list-style-type: none"><li>▪ Dễ cài đặt</li><li>▪ Dễ tìm đặc trưng tiêu biểu của đối tượng</li><li>▪ Thực hiện tốt với độ phân giải cao</li><li>▪ Đơn giản và dễ áp dụng với các thuật toán khác (SVM, NN,...)</li></ul>	<ul style="list-style-type: none"><li>▪ Phụ thuộc vào tập huấn luyện (giống về góc chụp, kích thước, tư thế, độ sáng,...)</li><li>▪ Nhạy cảm với nhiễu</li><li>▪ Độ chính xác không quá cao</li></ul>

# 3. Thuật toán PCA



# 3. Thuật toán PCA

## Phần huấn luyện:

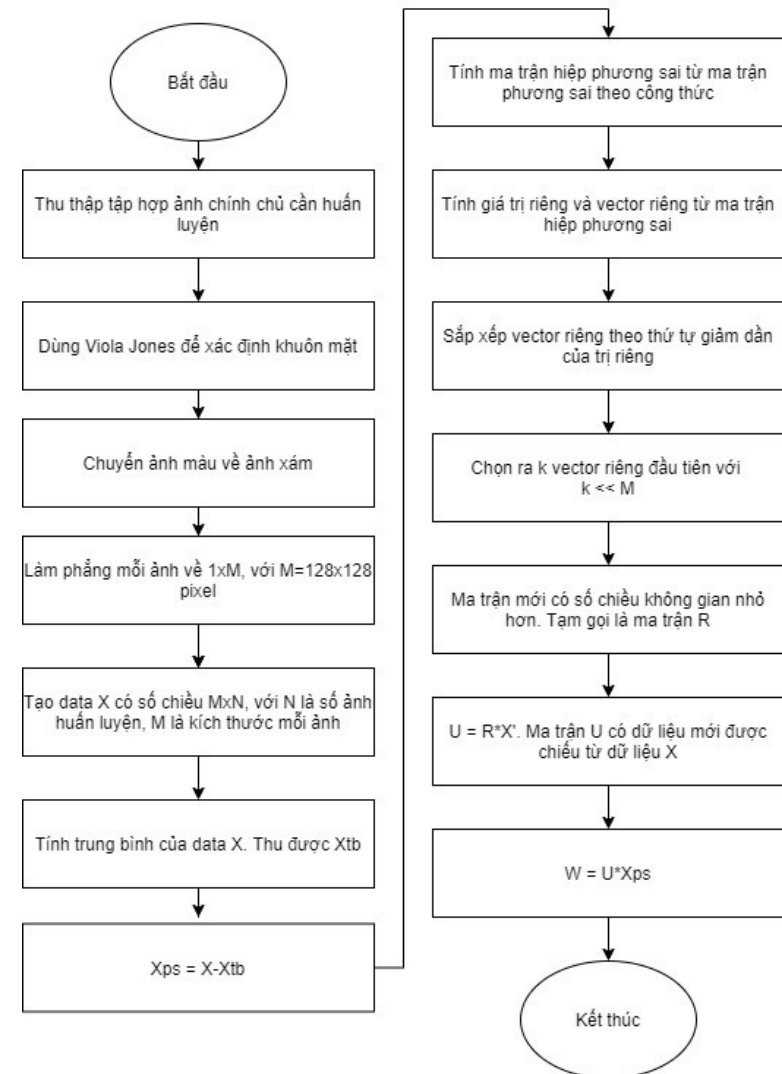
Ma trận phương sai:  $X = X - X_{tb}$

Ma trận hiệp phương sai:

$$Q = \frac{1}{N} \sum_{i=0}^n X_i * X_i'$$

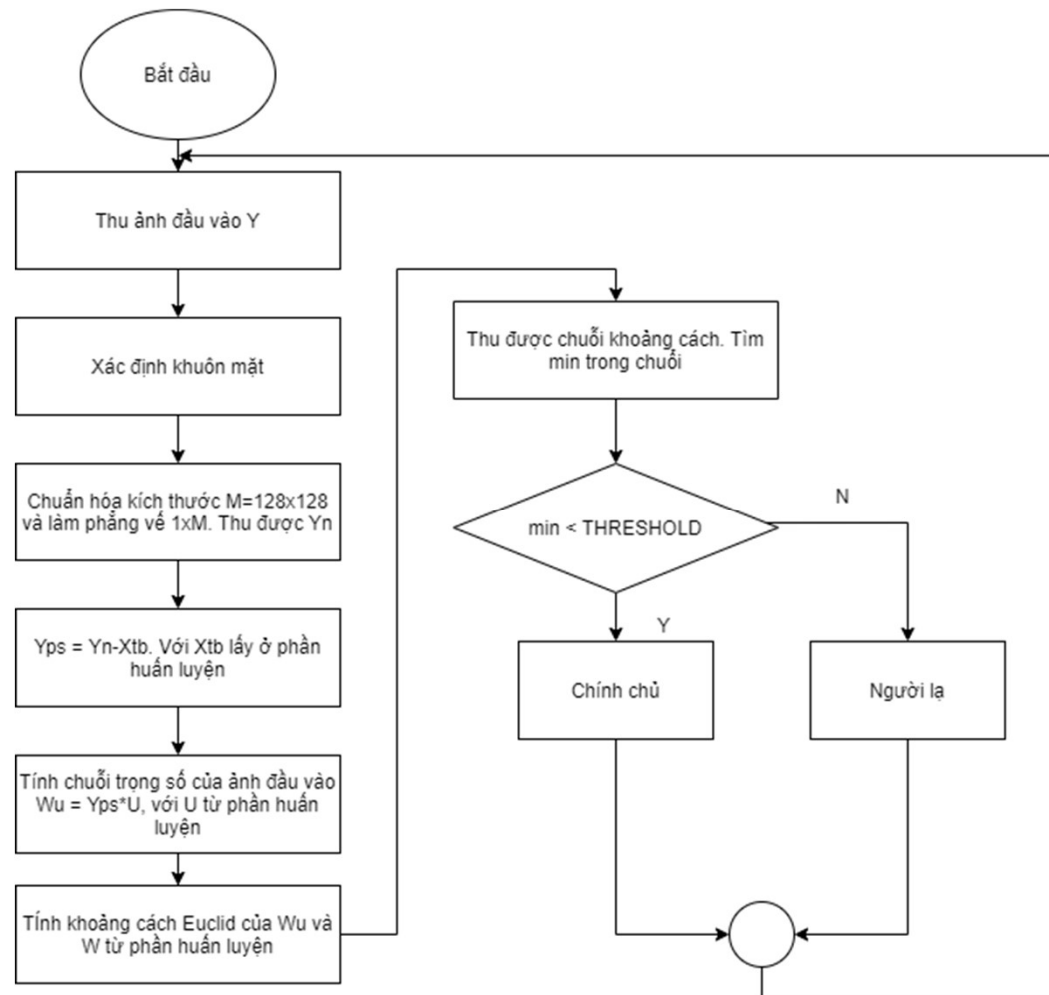
Trị riêng và vector riêng:

$$A \in \mathbb{R}^{n \times n} \quad AX = \lambda X$$



# 3. Thuật toán PCA

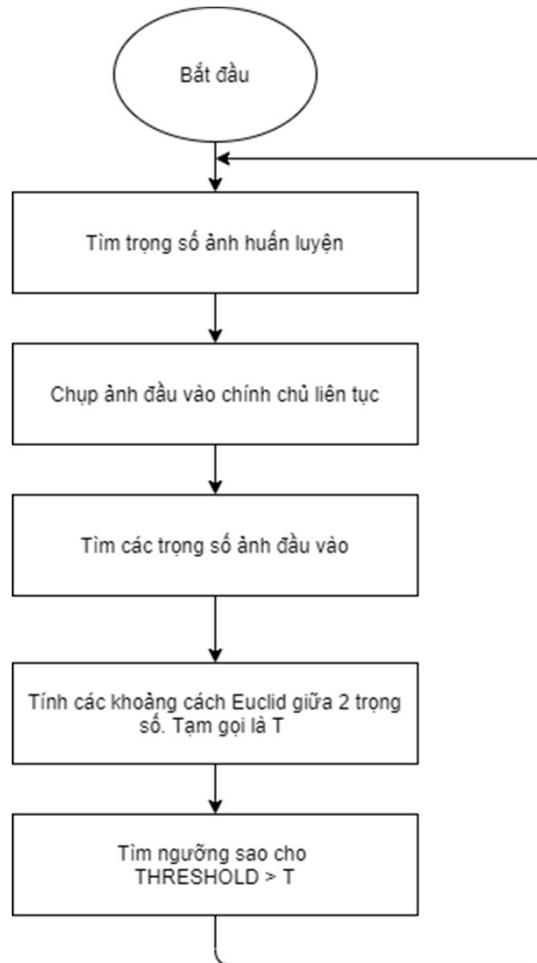
## Phần nhận dạng:





### 3. Thuật toán PCA

Phương pháp tìm ngưỡng cho nhận dạng khuôn mặt:





## 4. Đặc tả hệ thống

---

### PART 4 ĐẶC TẢ HỆ THỐNG





## 4. Đặc tả hệ thống

### Yêu cầu thiết kế:

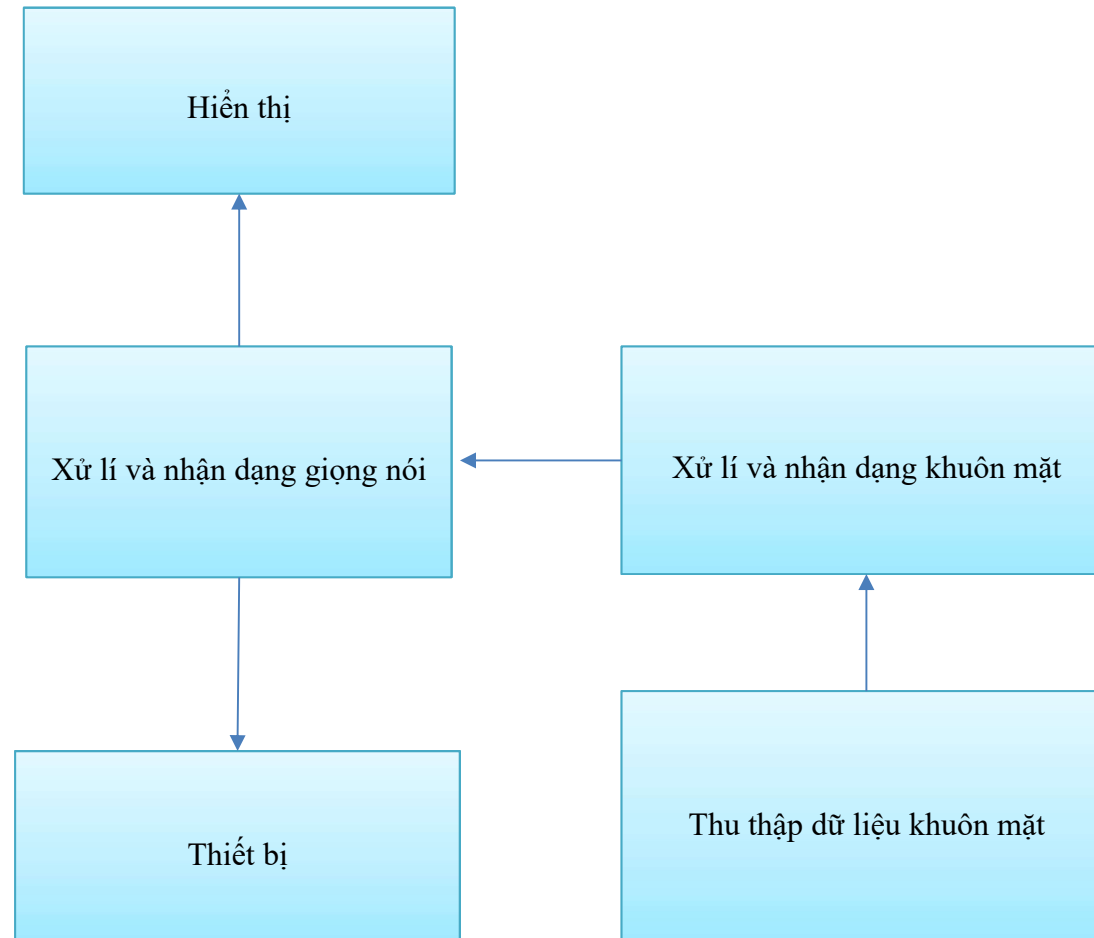
- Bộ từ vựng: “Không”, “Một”, “Hai”, “Ba”, “Bốn”, “Năm”, “Sáu”, “Bảy”, “Tám”, “Chín”
- Thời gian xử lý và nhận dạng dưới 0.5s
- Thời gian trễ không đáng kể
- Độ chính xác của nhận dạng trên 80%
- Có thể hoạt động trong thời gian dài
- Điện năng tiêu thụ thấp





## 4. Đặc tả hệ thống

- Sơ đồ khối:





## 5. Thiết kế phần cứng

---

### PART 5 THIẾT KẾ PHẦN CỨNG





# 5. Thiết kế phần cứng

## STM32F4 Discovery là gì?

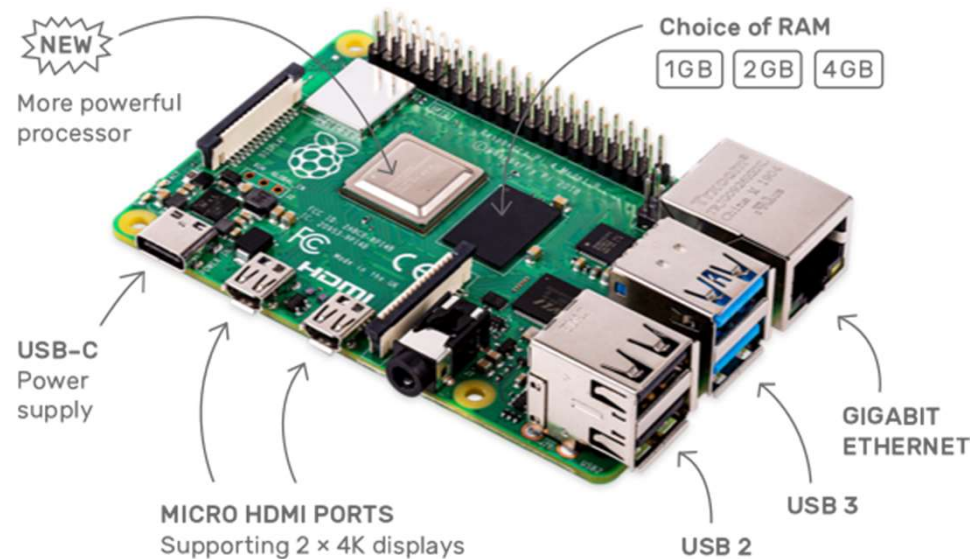
- 32-bit ARM Cortex-M4, 1 MB Flash, 192 KB RAM
- Bộ xử lý âm thanh: MP45DT02, ST MEMS audio sensor, omnidirectional digital microphone.
- Thư viện DSP



# 5. Thiết kế phần cứng

## Raspberry Pi là gì?

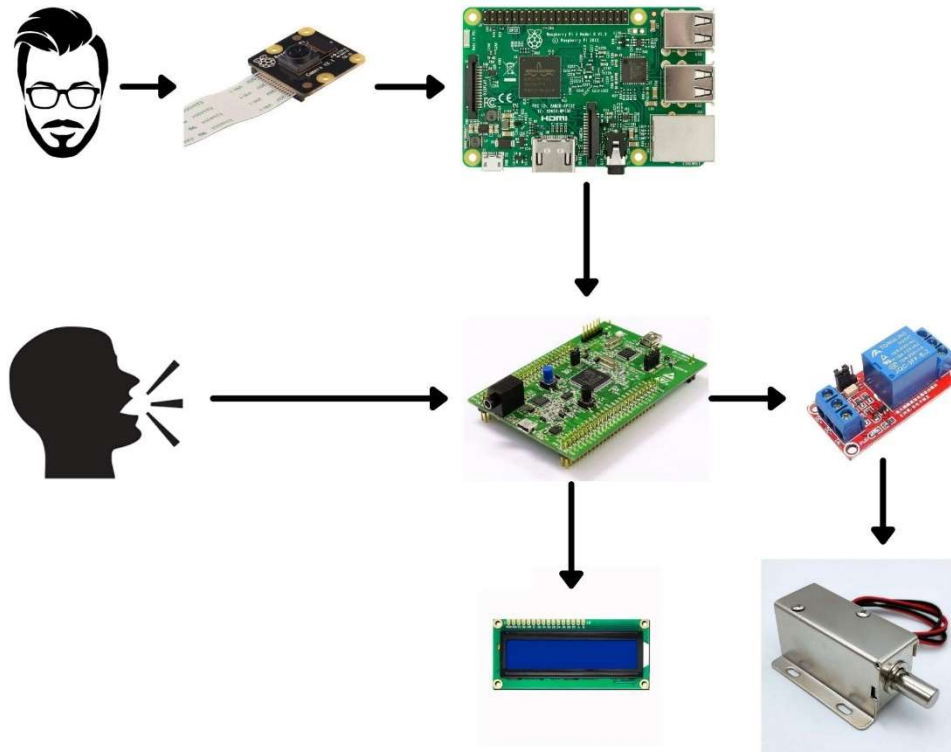
- Raspberry Pi là một máy tính rất nhỏ gọn, kích thước hai cạnh như bằng khoảng một cái thẻ ATM và chạy hệ điều hành Linux.
- Có thể sử dụng Raspberry Pi như một máy vi tính bởi người ta đã tích hợp mọi thứ cần thiết trong đó.





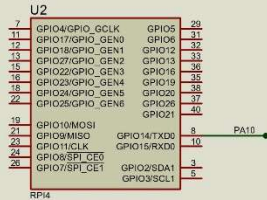
# 5. Thiết kế phần cứng

Cách thức hoạt động:

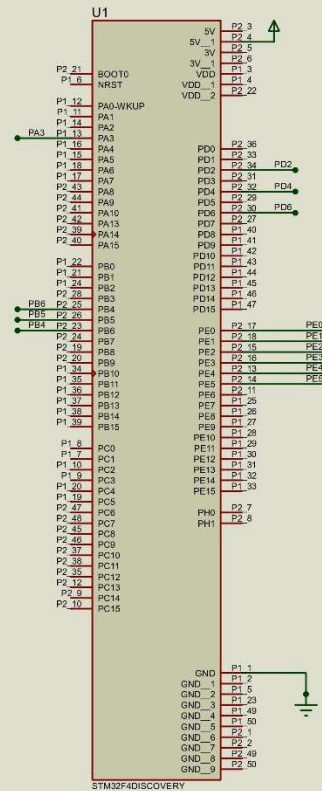


# 5. Thiết kế phần cứng

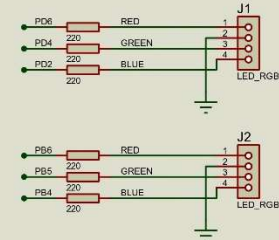
## Sơ đồ chi tiết:



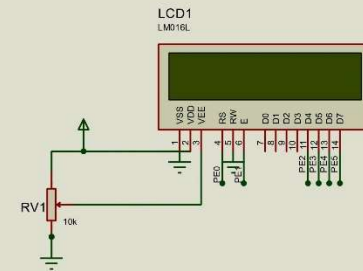
<RASPERRY PI 4>



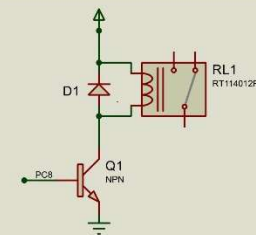
<STM32F4DISCOVERY Discovery>



<LED RGB>



<LCD 16X2>



<RELAY>



## 6. Kết quả

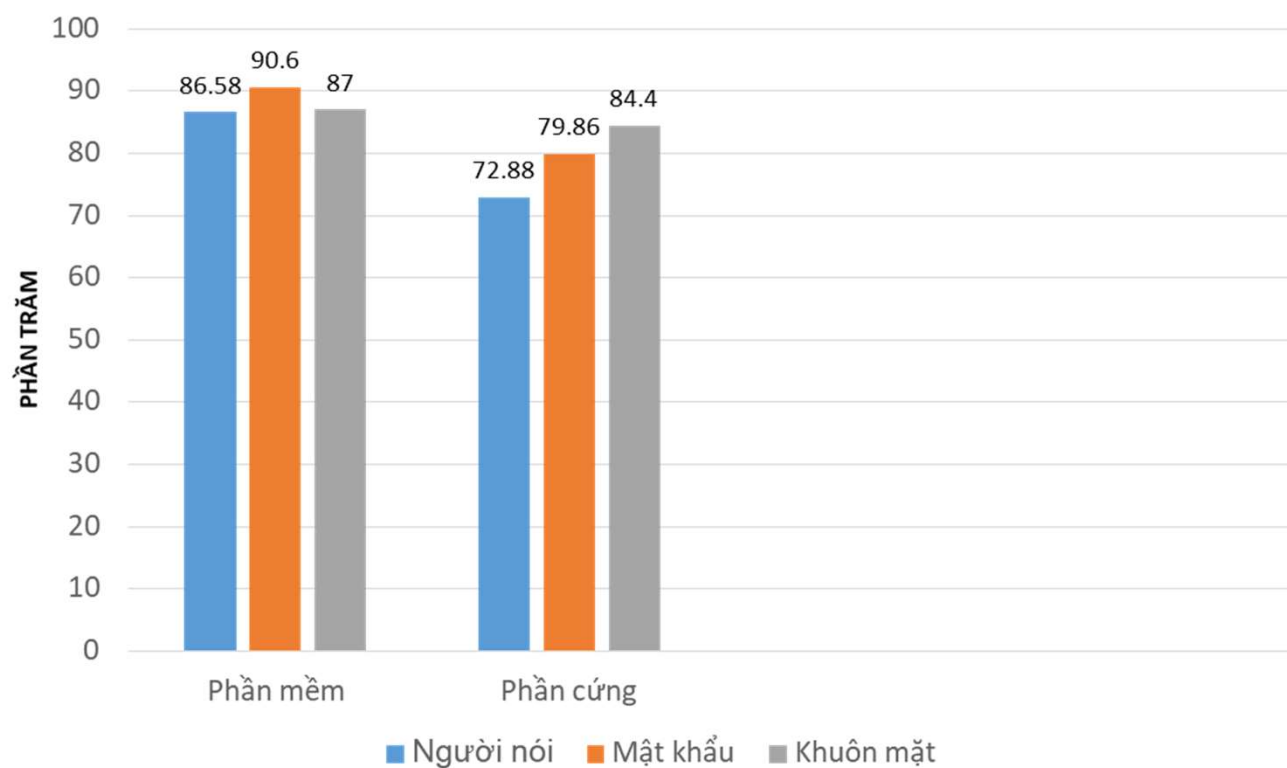
### PART 6 KẾT QUẢ





## 6. Kết quả

BIỂU ĐỒ SO SÁNH HIỆU QUẢ NHẬN DẠNG  
TRUNG BÌNH







# 7. Kết luận, hướng phát triển

## PART 7 KẾT LUẬN & HƯỚNG PHÁT TRIỂN





# 7. Kết luận, hướng phát triển

## Kết luận

- Nghiên cứu và áp dụng thành công bài toán nhận dạng giọng nói và khuôn mặt
- Nghiên cứu các phương pháp trích chọn đặc trưng và huấn luyện: MFCC, VQ, PCA,...
- Nắm được thêm kiến thức về xử lý tín hiệu số, vi điều khiển, máy tính nhúng
- Cải thiện kỹ năng lập trình
- Còn sai sót trong khâu nhận dạng và làm phần cứng
- Cần cải thiện thuật toán



## 7. Kết luận, hướng phát triển

### Hướng phát triển:

- Nghiên cứu tối ưu hóa phần cứng bằng vi điều khiển mạnh hơn
- Phát triển thuật toán phức tạp hơn: LBG, Neural Network, HMM,...
- Mở rộng bộ dữ liệu
- Làm mô hình cho ứng dụng mở khóa cửa





# CẢM ƠN THẦY CÔ VÀ CÁC BẠN ĐÃ LẮNG NGHE

*Thank you!*

