**TIÊU ĐỀ BÁO CÁO**

**Phan Văn Vũ, Võ Minh Đức, Lê Đức Minh, Nguyễn Sỹ Tuấn Thành**

**Nhóm 11, lớp HP: 17Nh11**

|  |  |  |  |
| --- | --- | --- | --- |
| **Điểm**  (dành cho  GV ghi) | **Bảng phân công nhiệm vụ** | | **Chữ ký của SV**  (mỗi SV ký xác nhận trước khi nộp báo cáo) |
|  | Lê Đức Minh | Đọc tài liệu, viết báo cáo phần cơ sở lý thuyết, viết báo cáo phần kết quả thực nghiệm |  |
|  | Nguyễn Sỹ Tuấn Thành | Đọc tài liệu, viết báo cáo phần đặt vấn đề, sơ đồ khối thuật toán. làm slide thuyết trình PowerPoint |  |
|  | Phan Văn Vũ (Nhóm Trưởng) | Đọc tài liệu và viết báo cáo phần kết luận, tìm các yếu tố ảnh hưởng đến độ chính xác thuật toán, kết quả thực nghiệm,kết luận |  |
|  | Võ Minh Đức | Đọc tài liệu, cài đặt và viết báo cáo về thuật toán STFT, làm slide thuyết trình |  |

Lời cam đoan: Chúng tôi, gồm các sinh viên có chữ ký ở trên, cam đoan rằng báo cáo này là do chúng tôi tự viết dựa trên các tài liệu tham khảo ghi rõ trong phần VII. Các số liệu thực nghiệm và mã nguồn chương trình nếu không chỉ dẫn nguồn tham khảo đều do chúng tôi tự làm. Nếu vi phạm thì chúng tôi xin chịu trách nhiệm và tuân theo xử lý của giáo viên hướng dẫn.

TÓM TẮT— Tìm các đặc trưng tần số của tín hiệu là bài toán cần thiết trong xử lý tín hiệu âm thanh, đặc biệt là tín hiệu tiếng nói. Bài báo cáo này thực hiện việc tìm các đặc trưng tần số của tín hiệu tiếng nói trên miền thời gian và tần số thông qua các biên độ của ảnh phổ. Các thử nghiệm với tín hiệu của 5 nguyên âm (/a/, /e/, /i/, /o/ và /u/) cho thấy bảng thống kê 3 tần số formant. Kết quả thực nghiệm cũng cho thấy ta có thể sử dụng các đặc trưng tần số trên miền thời gian và miền tần số để phân biệt tín hiệu tín hiệu tiếng nói.

Từ khóa— Formant, STFT(Short time Fourier Transform), DFT, FFT.

Mục lục

[I. ĐẶT VẤN ĐỀ 3](#_Toc26101323)

[II. LÝ THUYẾT XỬ LÝ TÍN HIỆU TIẾNG NÓI VÀ CÁC THUẬT TOÁN 3](#_Toc26101324)

[A. Lý thuyết 3](#_Toc26101325)

[*1.* Sơ đồ khối thuật toán 3](#_Toc26101326)

[*2.* STFT ( Short-time Fourier Transform ) 3](#_Toc26101327)

[*3.* The Spectrograph & Spectral Analysis 4](#_Toc26101328)

[B. Kích thước khác 4](#_Toc26101330)

[*1.* Phần tiêu đề 4](#_Toc26101331)

[*2.* Tác giả 4](#_Toc26101332)

[III. MÃ CHƯƠNG TRÌNH CÀI ĐẶT CÁC THUẬT TOÁN 4](#_Toc26101333)

[IV. KẾT QUẢ THỰC NGHIỆM 4](#_Toc26101334)

[A. Hình vẽ 4](#_Toc26101335)

[B. Bảng biểu 5](#_Toc26101336)

[V. KẾT LUẬN 5](#_Toc26101337)

[VI. TÀI LIỆU THAM KHẢO 5](#_Toc26101338)

# ĐẶT VẤN ĐỀ

Xử lý tiếng nói từ khi xuất hiện đã có một vai trò rất quan trọng trong cuộc sống của chúng ta. Cùng với sự phát triển của khoa học kỹ thuật, nhu cầu xử lý tiếng nói của con người ngày càng tăng cao. Xử lý tiếng nói có ứng dụng về nhiều mặt, về cơ bản có ứng dụng như nhận dạng tiếng nói, người nói, tăng chất lượng giọng nói và tổng hợp tiếng nói. Để làm được điều đó, việc nghiên cứu biên phổ của tín hiệu là rất quan trọng. Nhiều nghiên cứu về xử lý và nhận dạng tiếng nói đã chỉ ra rằng các tham số formant là ứng cử viên tốt nhất cho việc biểu diễn phổ. Các phương pháp xác định formant liên quan đến việc tìm kiếm các đỉnh trong các biểu diễn phổ, thường là từ kết quả phân tích phổ theo phương pháp STFT hoặc mã hóa dự đoán tuyến tính (LPC).

# LÝ THUYẾT XÁC ĐỊNH ẢNH PHỔ BA CHIỀU

Phần này trình bày các lý thuyết có liên quan đến vấn đề cần giải quyết, cơ sở lý thuyết của các thuật toán, sơ đồ khối và các tham số quan trọng của mỗi thuật toán, phân tích các vấn đề của mỗi thuật toán và đề ra giải pháp khắc phục (nếu có). Nên dùng hình vẽ để minh hoạ ý tưởng.

## Lý thuyết

### Sơ đồ khối thuật toán

Màu sắc (Biên độ phổ)

Thời gian

Phân khung tín hiệu

Tín hiệu vào x(n)

Lấy DTFT trên từng khung

STFT của x(n)

Tần số

### STFT ( Short-time Fourier Transform )

#### Định nghĩa

* Biến đổi fourier thời gian ngắn là sự phân chia chuỗi thời gian các khối chồng nhau (overlaping blocks) có chiều dài bằng nhau và áp dụng biến đổi Fourier (DFT) dự trên thuật toán FFT cho mỗi khối một cách tuần tự.

Biến đổi DFT của tín hiệu x[n]:

Trong đó:

L : chiều dài cửa sổ

k : số điểm DFT của mỗi cửa sổ ()

w[m]: hàm cửa sổ

#### Tính chất

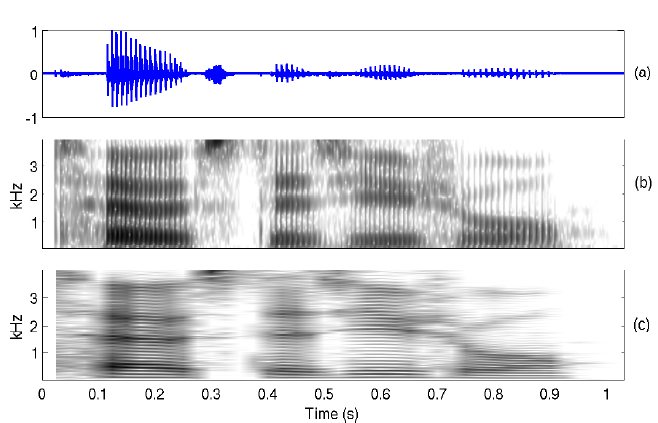
STFT không có tính chất bảo toàn năng lượng

Để thực hiện phương pháp này một cách tốt nhất thì yêu cầu phải chọn khoảng thời gian của các đoạn để phân chia sao cho tín hiệu ở mỗi khoảng thời gian đó có thể coi là tĩnh. Vì STFT chỉ xử lí số liệu tĩnh trên mỗi đoạn nên nó chỉ tính một cặp giá trị biên độ và pha

STFT là một phương pháp phổ biến và tính toán hiệu quả. Nhược điểm lớn nhất của phương pháp này là khi tín hiệu có một dải động lớn thì cụm tần số thấp. Trong trường hợp đó hướng tạp âm tần số cao có thể che cấu trúc tín hiệu tần số cao

### Narrowband và wideband:

Băng thông là dãy tần số tập trung hầu hết năng lượng (công suất) của tín hiệu. Nếu năng lượng tín hiệu tập trung vào một dãy tần số nào đó giữa tần số thấp và tần số cao thì đó là tin hiệu thông tải. Lúc này khái niệm băng hẹp (narrowband) được dùng để chỉ tín hiệu có băng thông F2 - F1 rất nhỏ (khoảng 10% hoặc nhỏ hơn) so với tần số trung tâm (F1+F2)/2. Ngược lại gọi là băng thông rộng.[4]



Hình 1: Narrowband(c) và wideband(b)

### Spectrogram:

Spectrogram là một trong những công cụ cơ bản của phân tích phổ tín hiệu tiếng nói, trong đó nó chuyển đổi dạng sóng tín hiệu tiếng nói hai chiều thanh cấu trúc ba chiều (biên độ/tần số/thời gian). Trong đồ thị spectrogram, thời gian và tần số tương ứng là các trục ngang và dọc, còn biên độ được biểu diễn bởi độ đậm nhạt của màu biểu diễn. Các đỉnh của phổ tín hiệu xuất hiện là các dải nằm ngang màu đậm. Tần số trung tâm của các dải thường được coi là tần số formant. Các âm hữu thanh tạo ra các mảng dọc trong biểu đồ spectrogram bởi vì có một sự tăng cường biên độ tín hiệu tiếng nói mỗi khi thanh quản đóng lại. Nhiễu trong các âm vô thanh tạo ra các cấu trúc đậm hình chữ nhật và kết thúc ngẫu nhiên với nhiều đốm nhạt do sự thay đổi tức thì của năng lượng tín hiệu. Lược đồ spectrogram chỉ diễn tả biên độ phổ của tín hiệu mà bỏ qua các thông tin về pha bởi vì các thông tin về pha được cho rằng không có vai trò quan trọng trong hầu hết các ứng dụng liên quan đến tiếng nói.[5]

## Thuật toán

INPUT : x[n], w[m] , H, K, fs

1: Xác định độ dài của tín hiệu vào N

2: Xác định độ dài của hàm cửa sổ window M

3: Xác định số khung frame K

4: for l = 0, 1, …, K-1

5: step 1: Tính w[m]\* x[ l\*H + 1 ,….M + l\*H] ⋅

6: step 2: Lấy FFT của NFFT điểm trên từng khung trả về 1 vector đã tính DFT

7: step 3: Lưu vào ma trận S( Ma trận tính STFT )

8: end for

9: Tính vector thời gian và tần số

t = (M/2,M/2 + H ,…, M/2 + (K-1)\*H);

w = (0,…,NFFT/2)\*fs/K;

OUTPUT : S,w,t

# MÃ CHƯƠNG TRÌNH CÀI ĐẶT CÁC THUẬT TOÁN

### File stft.m để tính STFT của tín hiệu vào

function [S, f, t] = stft(x, win, step, NFFT, fs)

%

% Input:

% x - Tín hieu tren mien thoi gian

% win - Ham cua so window

% step - Buoc nhay cua so

% NFFT - So diem lay FFT

% fs - Tan so lay mau, Hz

%

% Output:

% STFT - STFT-Ma tran (only unique points, time

% across columns, frequency across rows)

% f - Vector tan so, Hz

% t - Vector thoi gian, s

N = length(x);

L = length(win);

K = fix((N-L + step)/step); % Tính so luong khung tin hieu

N2 = NFFT/2 + 1; % Tinh so diem fft duy nhat(không lap lai)

S =zeros(K,N2); % Khai bao ma tran STFT

for k=0:K-1

% windowing

xw=x(1+k\*step:L+k\*step).\*win;

%FFT

X=fft(xw,NFFT);

X1=X(1:N2)';

S(k+1,1:N2)=X1.\*conj(X1);

end

% Tính Vector thoi gian và tan so

t=(0:K-1)'\*step/fs;

f=(0:NFFT/2)'\*fs/NFFT;

end

### File SpectrogramTest.m để vẽ ảnh phổ 3 chiều của tín hiệu vào.

clear, clc, close all

% load an audio file

[x, fs] = audioread('aeiuo\_lon\_data.wav');

% Tham sô dau vào

L = 256; % Chieu dai cua so

step = L/4; % Buoc nhay giua cac cua so

NFFT = 2048;% So diem lay FFT

win = hanning(L);

[S, f, t] = stft(x,win,step,NFFT,fs);

S=(S)';

S=S/max(max(S)); % Chuan hoa bien do

S = 20\*log10(S);

figure(1)

surf(t, f, S)

shading interp

axis tight

view(0, 90)

set(gca, 'FontName', 'Times New Roman', 'FontSize', 14)

xlabel('Thoi gian, s')

ylabel('Tan so, Hz')

title('Pho bien do cua tin hieu')

hcol = colorbar;

set(hcol, 'FontName', 'Times New Roman', 'FontSize', 14)

ylabel(hcol, 'Cuong do am, dB')

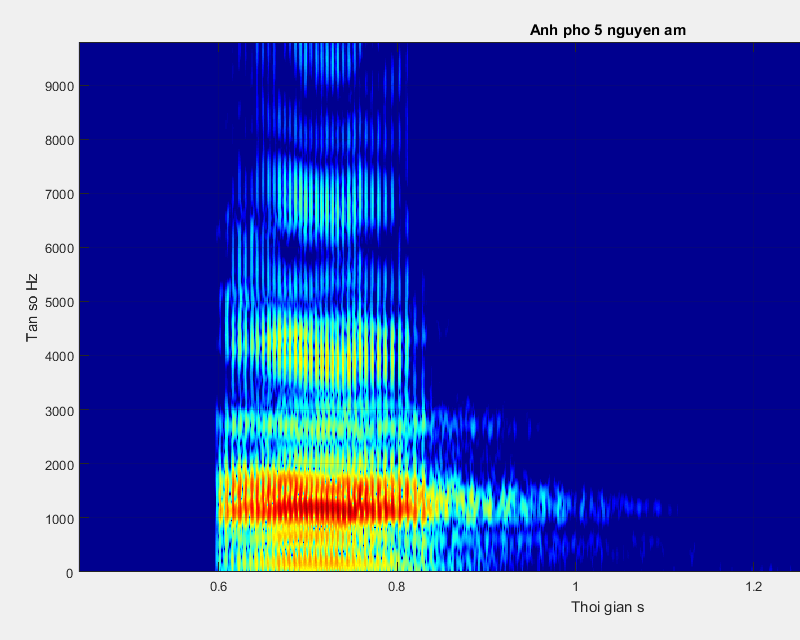
# KẾT QUẢ THỰC NGHIỆM

Dưới đây là kết quả thực nghiệm sau khi thực hiện hàm tính phổ 3 chiều (Spectrogram) của file tín hiệu 5 nguyên âm /a/, /e/, /i/, /u/, /o/ bằng phương pháp Short-Time-Fourier-Transform (STFT) đồng thời xác định được các tần số Formant để phân biệt giữa các nguyên âm với nhau.

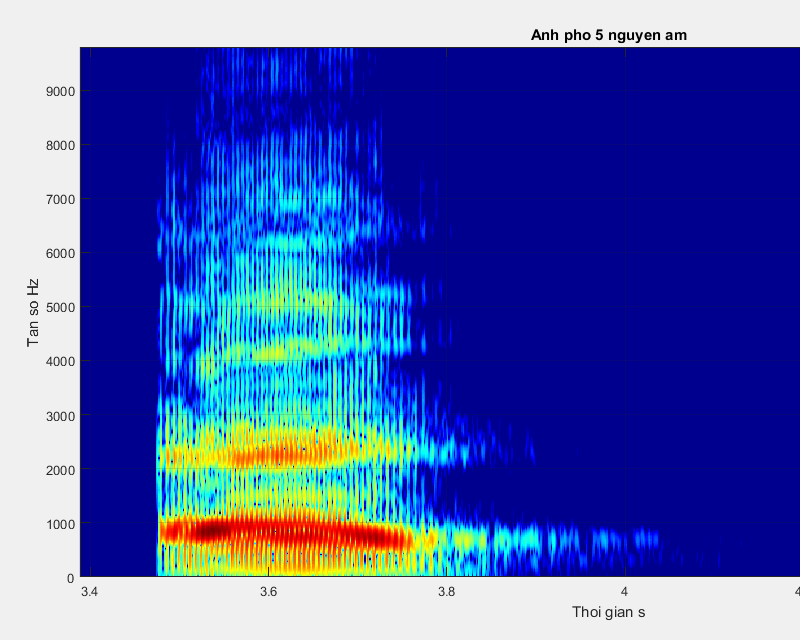
## Hình vẽ



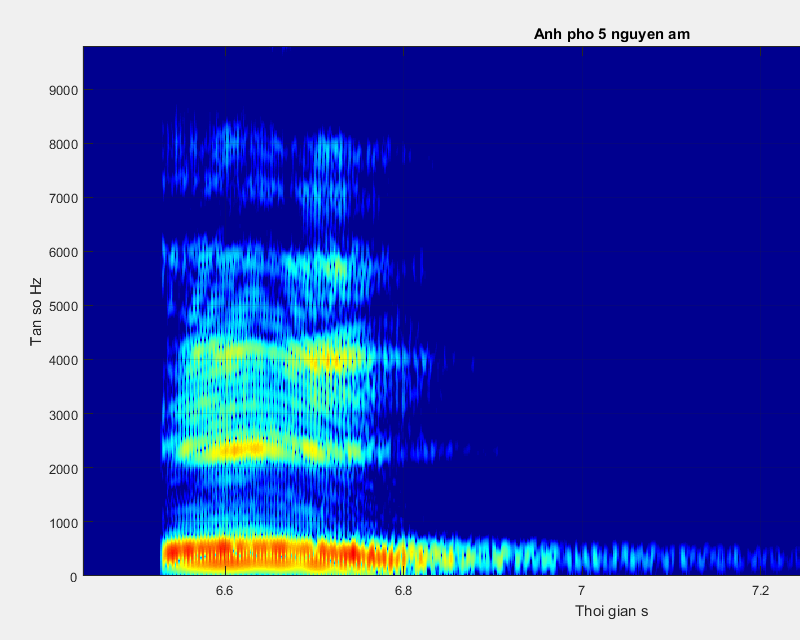
1. Ảnh phổ 3 chiều (Spectrogram) của 5 nguyên âm /a/, /e/, /i/, /u/, /o/



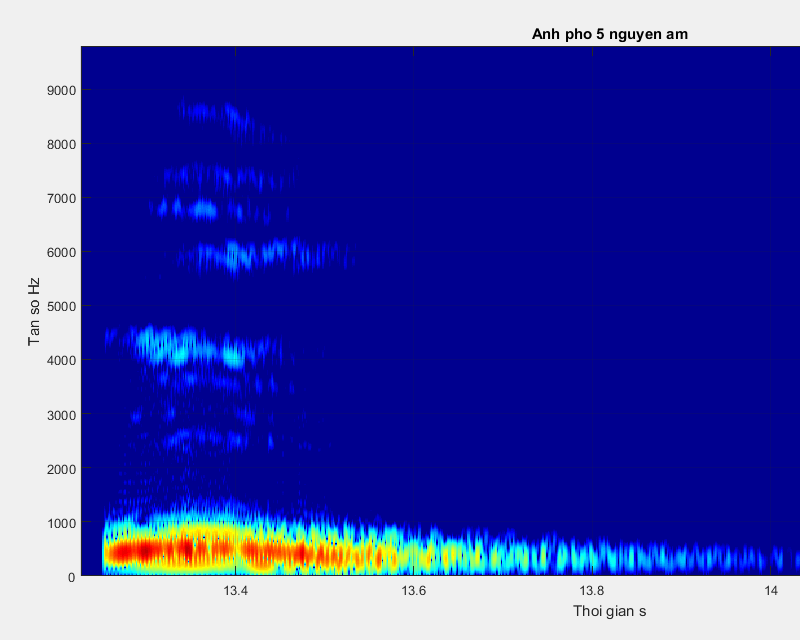
1. Ảnh phổ tần số của nguyên âm /a/



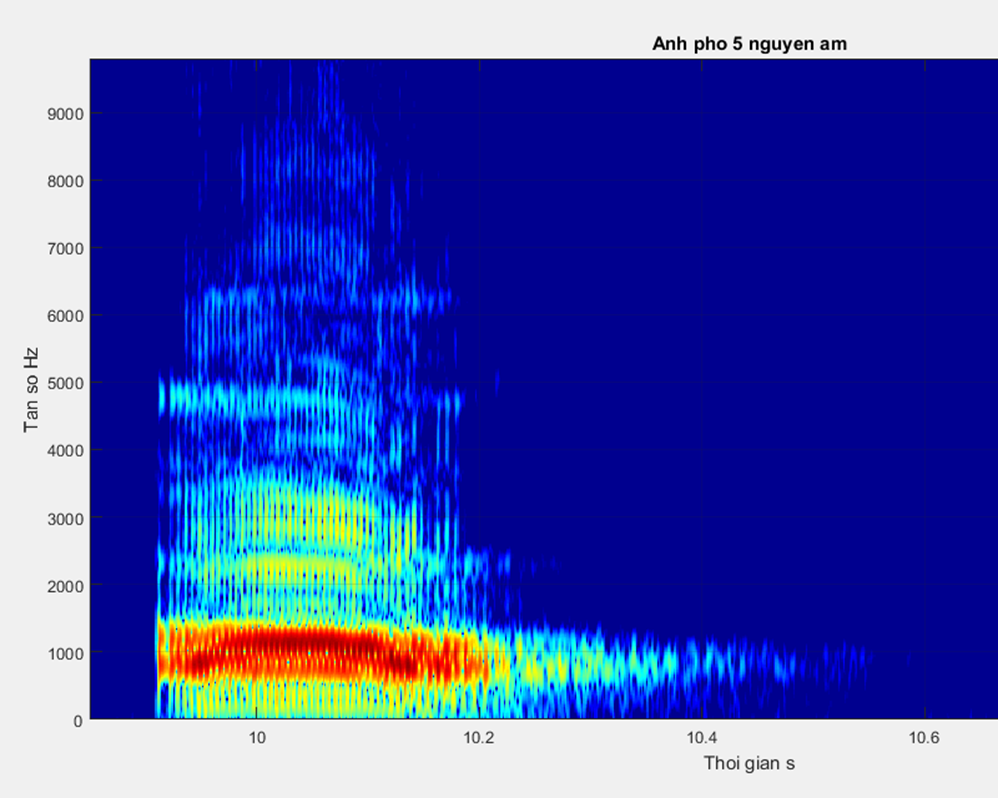
1. Ảnh phổ tần số của nguyên âm /e/



1. Ảnh phổ tần số của nguyên âm /i/



1. Ảnh phổ tần số của nguyên âm /o/



1. Ảnh phổ tần số của nguyên âm /u/

## Thống kê tần số Formant của 5 nguyên âm /a/, /e/, /i/, /u/, /o/

## 

Hình 7. Ảnh phổ nguyên âm /a/ để xác định tần số Formant F1, F2, F3

## 

Hình 8. Tần số Formant của nguyên âm /a/ sau khi sử dụng phần mềm Praat.

### Bảng thống kê tần số Formant của 5 nguyên âm và giữa 2 người nói khác nhau.

1. Tần số Formant (Hz) của 5 nguyên âm của người 1

|  |  |  |  |
| --- | --- | --- | --- |
|  | F1 | F2 | F3 |
| Âm /a/ | 1132 | 1548 | 2709 |
| Âm /e/ | 870 | 2227 | 2796 |
| Âm /i/ | 518 | 2293 | 3191 |
| Âm /u/ | 519 | 780 | 2830 |
| Âm /o/ | 847 | 1220 | 2446 |

1. Tần số Formant (Hz) của 5 nguyên âm của người 2

|  |  |  |  |
| --- | --- | --- | --- |
|  | F1 | F2 | F3 |
| Âm /a/ | 935 | 1329 | 2577 |
| Âm /e/ | 694 | 1680 | 2380 |
| Âm /i/ | 409 | 2468 | 2804 |
| Âm /u/ | 388 | 716 | 2709 |
| Âm /o/ | 672 | 1023 | 2731 |

**Nhận xét:**

* Qua ảnh phổ và kết quả thực nghiệm tần số Formant trên 5 nguyên âm, ta thấy rằng tần số Formant đặc trưng cho mỗi âm khác nhau. Tần số F1 của âm /a/ là cao nhất (d/c), âm /i/ có tần số F2 và F3 cao nhất (d/c). Với mỗi âm, ta có một bộ ba tần số Formant (F1, F2, F3) thường nằm trong một khoảng nhất đinh
* Đối với giọng của mỗi người khác nhau, thì bộ ba tần số Formant cũng là khác nhau. Theo như kết quả thu được, bộ tần số Formant của người 1 luôn có xu hướng cao hơn người 2.

# KẾT LUẬN

Bài báo cáo này thực hiện việc cài đặt thuật toán để tìm ra ảnh phổ 3 chiều Spectrogram của 5 nguyên âm /e/, /i/, /u/, /o/ bằng cách sử dụng phương pháp Short-Time-Fourier-Transform. Ta thấy độ phân giải tần số (frequency resolution) tỉ lệ thuận với độ dài của hàm cửa sổ window. Đồng thời qua đây ta tìm ra được các mức tần số Formant để phân biệt từng nguyên âm với nhau cũng như phân biệt giữa mỗi người nói với nhau. Trong tương lai chúng tôi sẽ cải tiến thuật toán để cải thiện chất lượng của ảnh phổ tín hiệu và tăng độ chính xác của các tần số Formant để phân biệt tín hiệu giọng nói.

# TÀI LIỆU THAM KHẢO

[1] Author1\_Name, Author2\_Name,“Paper Title ACASH: An Adaptive Web Caching method based on the Heterogeneity of Reference Characteristics”, Journal of AICIT, AICIT(Publication\_Name), vol. 10, no. 4, pp.169-711, 2015.

[2] Author1\_Name, Author2\_Name, Web Caching and Replication, Addison-Wesley(Publication\_ Name), USA, 2014

[3] Link: https://en.wikipedia.org/wiki/Window\_function

[4] 4.4.4 băng thông bài giảng xử lí tín hiệu số giảng viên cung cấp

[5] 3.5.2 spectrogram bài giảng xử lí tiếng nói Phạm văn sự lê xuân thành học viện bưu chính viễn thông