TIÊU ĐỀ BÁO CÁO

Trần Văn A, Phan Thị B, Đỗ C, Nguyễn D

Nhóm 9, lớp HP: xyz1615

|  |  |  |  |
| --- | --- | --- | --- |
| **Điểm**  (dành cho  GV ghi) | **Bảng phân chia nhiệm vụ**  (SV ghi càng cụ thể thì GV càng dễ đặt câu hỏi và cho điểm mỗi SV) | | **Chữ ký của SV**  (mỗi SV ký xác nhận trước khi nộp báo cáo) |
|  | Trần Văn A | Đọc tài liệu, cài đặt và viết báo cáo về thuật toán X (tr. 10-11) |  |
|  | Phan Thị B | Đọc tài liệu, cài đặt và viết báo cáo về thuật toán Y (tr. 6-9) |  |
|  | Đỗ C | Đọc tài liệu và viết báo cáo về thuật toán Z (tr. 3), thống kê sai số thuật toán T (tr. 20), làm slide và thuyết trình PowerPoint |  |
|  | Nguyễn D | Cài đặt thuật toán Z (tr. 12-15), viết báo cáo phần đặt vấn đề (tr. 1) và phần kết quả thực nghiệm (tr. 16-19) |  |

**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**— Tần số cơ bản của giọng nói con người (F0) là tần số giao động của dây thanh quản là đặc trưng cho mỗi giọng nói của mỗi người[1]. Có nhiều phương pháp khác nhau để tự động tính toán được tần số cơ bản (F0), trong báo cáo này chúng tôi sử dụng 2 cách để thu được tần số cơ bản của giọng nói 4 thành viên trong nhóm dựa trên áp dụng hàm tự tương quan (Autocorrelation Function)- xác định F0 trên miền thời gian và dùng thuật toán Harmonic Product Spectrum- xác định F0 trên miền tần số.Các thuật toán được tính toán trên các đoạn thu âm 5 nguyên âm cơ bản.Sau khi thu được kết quả F0 , sử dụng lọc trung vị(Median Filtering) để xử lý kết quả.Kết quả chúng tôi thu được kết quả F0 với sai số trung bình…. So với kết quả F0 đã tính toán thủ công từ trước.

**Từ khóa**—Autocorrelation, Fundamental Frequency, Time domain ,Median Filtering

# ĐẶT VẤN ĐỀ

Tần số cơ bản –Fundamental Frequency (F0 hoặc cao độ) của một tín hiệu tuần hoàn bằng nghịch đảo của chu kỳ của tín hiệu đó.Chu kỳ được xác định bằng khoảng thời gian ngắn nhất mà tín hiệu lặp lại trên miền thời gian.[2]Do đó chỉ có tín hiệu tuần hoàn mới xác định được tần số cơ bản. Đối với giọng nói của con người, tần số cơ bản là tần số giao động của dây thanh quản , mỗi giọng nói có một cao độ nhất định (pitch) đặc trưng cho mỗi giọng nói giúp người nghe phân biệt được các giọng nói khác nhau.Tần số cơ bản của đàn ông nằm trong khoảng 85-180Hz và phụ nữ nằm trong khoảng 165-255Hz[3].Trong báo cáo này chúng tôi cài đặt 2 thuật toán tìm tần số cơ bản là tự tương quan(Autocorrelation) và Harmonic Product Spectrum trên Matlab với dữ liệu đầu vào là các đoạn âm thanh của 5 nguyên âm cơ bản của các thành viên trong nhóm.

Về thuật toán tìm tần số cơ bản sử dụng hàm tự tương quan, đây là một trong những thuật toán phổ biến để tìm tần số cơ bản trên miền thời gian, xử dụng hàm xcorr() và findpeaks() trong Matlab để áp dụng tìm F0 của từng đoạn âm thanh dài 30ms nằm trong đoạn trung tâm của các nguyên âm để thu được kết quả chính xác nhất.

Về thuật toán tìm tần số cơ bản trên miền tần số…// Tài viết tiếp chỗ ni

Sau đó chúng tôi sử dụng phương pháp lọc trung vị để lọc kết quả F0 của những thuật toán trên….//Tấn viết tiếp chỗ ni

Bài viết có bố cục như sau .Phần II trình bày cơ sở lý thuyết của xử lý tín hiệu tiếng nói và các thuật toán, sơ đồ khối các bước triển khai từng thuật toán.Phần III Trình bày mã nguồn khi cài đặt từng thuật toán. Phần IV trình bày kết quả tính toán và hình vẽ, so sánh kết quả với kết quả đã tính toán bằng thủ công.Cuối cùng là phần V trình bày kết luận của báo cáo.

# CƠ SỞ LÝ THUYẾT XỬ LÝ TÍN HIỆU TIẾNG NÓI VÀ CÁC THUẬT TOÁN

Xử lý tín hiệu tiếng nói là việc thực hiện giải quyết các vấn đề để tìm ra một dạng thức tối ưu biểu diễn được tiếng nói một cách hiệu quả nhờ vào việc xác định các đặc trưng như nguồn kích thích(excitation), tần số cơ bản(fundamental frequency-pitch), biên độ (amplitude).Xử lý tín hiệu tiếng nói phải được tiến hành trong một khoảng thời gian giới hạn, được gọi là phân tích tiếng nói ngắn hạn[4]. Xử lý tiếng nói có thể thực hiện được trên miền thời gian (time domain) và miền tần số(frequency domain).

## Thuật toán tìm tần số cơ bản bằng sử dụng hàm tự tương quan:

Về cơ bản thuật toán tự tương quan chỉ ra rằng tín hiệu tuần hoàn,ngay cả không phải là sóng sin sẽ bao gồm những đoạn tín hiệu tương tự liên tiếp nhau.Điều này hoàn toàn đúng ngay cả khi biên độ tín hiệu liên tục thay đổi theo thời gian[5].Để thu được tần số cơ bản của tín hiệu ta lấy một đoạn tín hiệu âm thanh khoảng 30ms với số mẫu khoảng 1323 , sau đó thực hiện tìm hàm tự tương quan của đoạn tín hiệu đó.

Hàm tự tương quan của tín hiệu rời rạc được định nghĩa như sau:

trong đó 𝑥𝑗 là tín hiệu tại thời điểm j, 𝑟𝑡(𝜏) là hàm tự tương quan với độ trễ (lag) 𝜏 được tính tại thời điểm *t*, và *W* là độ dài của cửa sổ tín hiệu.[6]

Nếu tín hiệu là tuần hoàn thì hàm tự tương quan sẽ cho ra các đỉnh tại những thời điểm là bội số của chu kì tín hiệu. Phương thức tìm cao độ là dùng hàm tự tương quan chọn ra một đỉnh cao nhất với độ trễ lớn hơn 0 bằng cách xác định toàn bộ các giá trị độ trễ. Sau đó ta có công thức tính tần số cơ bản của tín hiệu:

F0=

Trong đó F0 là tần số cơ bản của tín hiệu, Fs là tần số lấy mẫu và tmax là độ trễ mà ở đó hàm tự tương quan có giá trị lớn nhất.[7]

Trong Matlab cung cấp hàm xcorr() để thực hiện việc tìm hàm tự tương quan của một tín hiệu, tuy nhiên hàm xcorr() thực hiện gấp đôi số mẫu của tín hiệu nên ta cần phải tạo một vector với số mẫu bằng với tín hiệu gốc với giá trị là giá trị của hàm tự tương quan.Hình 1 thể hiện hàm tự tương quan khi dùng hàm xcorr() với gốc tín hiệu lúc này nằm ở điểm có giá trị lag là 1024 với số mẫu gấp đôi tín hiệu ban đầu. Hình 2 thể hiện hàm tự tương quan điều chỉnh với số mẫu bằng số mẫu của tín hiệu ban đầu.

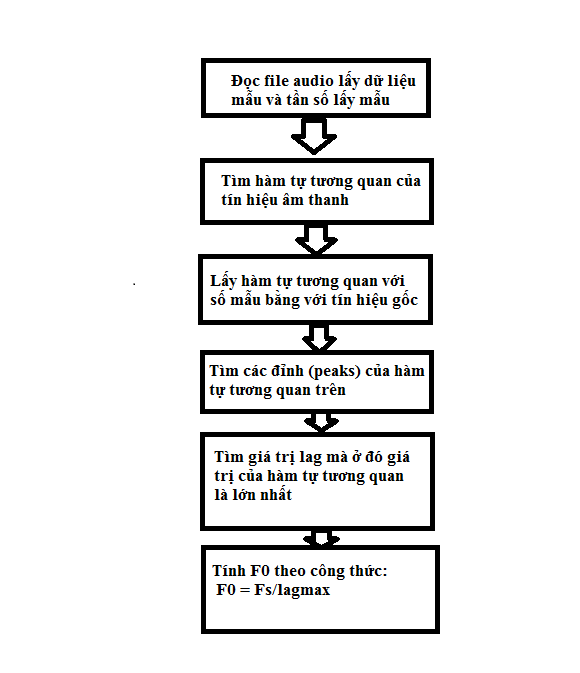


Hình .Tín hiệu gốc và hàm tự tương quan dùng hàm xcorr()



Hình 2.Tín hiệu gốc và hàm tự tương quan với số mẫu bằng tín hiệu gốc

Sơ đồ khối các bước thực hiện cài đặt thuật toán để tìm tần số cơ bản dùng hàm tự tương quan:



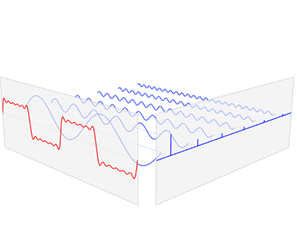
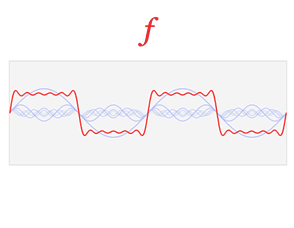
Hình 3.Sơ đối khối các bước cài đặt thuật toán

Đối với việc thực hiện tính toán tần số cơ bản bằng hàm tự tương quan có nhiều trường hợp kết quả không như mong đợi xảy ra do lỗi cao độ ảo

## Thuật toán tìm tần số cơ bản trên miền tần số:

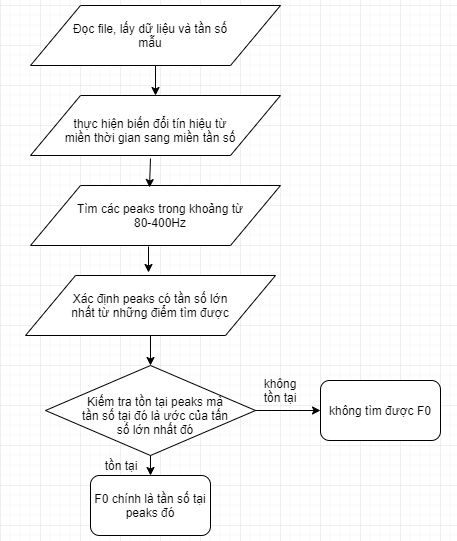
Trong phân tích tín hiệu tiếng nói, thay vì sử dụng trực tiếp tín hiệu tiếng nói trong miền thời gian, người ta thường hay sử dụng các đặc trưng phổ của tiếng nói. điều này xuất phát từ quan điểm rằng tín hiệu tiếng nói cũng giống như các tín hiệu xác định khác có thể xem như là tổng của các tín hiệu hình sin với biên độ và pha thay ñổi chậm.

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 biểu đồ spectrogram, thời gian và tần số tương ứng là các trục ngang và dọc, còn biên độ được biểu diễn bởi trục thẳng đứng



Hình 4.Mô phỏng tín hiệu qua 3 trục biên đọ, thời gian, tần số

Sơ đồ khối các bước thực hiện cài đặt thuật toán để tìm tần số cơ bản trên miền tần số:



Hình 5.Thuật toán tìm tấn số cơ bản trên miền tần số

# CÀI ĐẶT CÁC THUẬT TOÁN

***A.Thuật toán tìm tần số cơ bản bằng hàm tự tương quan:***

[y,Fs]=audioread('C:/Users/Laptop 88/Desktop/voice2/ami.wav');

temp=xcorr(y); %ham xcorr() tao vector ham tu tuong quan voi so mau gap doi tin hieu goc

len = length(y); %tim so mau cua tin hieu ban dau

autoCorrelation = zeros(1,len); %tao vector ham tu tuong quan voi so mau bang tin hieu goc

for n=1: len

autoCorrelation(n)=temp(n+len-1); %chuyen gia tri cua temp qua autoCorrelation

end;

subplot(2,1,1);

plot(y) ;

title('Sinwave');

xlabel('Time, [s]');

ylabel('Amplitude');

grid;

subplot(2,1,2);

plot(autoCorrelation) ;

title('Autocorrelation function of the sinewave');

xlabel('lags');

ylabel('Autocorrelation');

[autoValues,lagValues] = findpeaks(autoCorrelation); %tim cac dinh cua ham tu tuong quan

[M,index]=max(autoValues); %tim vi tri index de ham tu tuong quan dat gia tri Max

lagMax=lagValues(index); %tim gia tri lag tai vi tri ham tu tuong quan dat gia tri Max

F0=Fs/lagMax %tinh F0 = Fs/ lagMax

***B.Thuật toán tìm tần số cơ bản trên miền tần số:***

**close all;**

**clear all;**

**[yo,Fo]=audioread('E:\XLTHS\BT\main/main\_o.wav'); %doc file**

**t = linspace(0,length(yo)/Fo,length(yo));%tao vecto thoi gian**

**subplot(2,1,1);**

**plot(t,yo);**

**findpeaks(yo,t);%tim cac peaks**

**xlabel('time(s)')**

**title('Time domain');**

**nfft = 1024;%chieu dai bien doi fourier**

**f = linspace(0, Fo, nfft);%tao vecto tan so**

**g = abs(fft(yo, nfft));%thuc hien bien doi fourier (g la fft cua y0 trong 1024 mau)**

**subplot(2,1,2);**

**plot(f(1:nfft/2), g(1:nfft/2));**

**xlabel('frequency (Hz)')**

**title('Frequency domain');**

**[pks,locs] = findpeaks(g(1:nfft/2),f(1:nfft/2));%tim cac peaks**

**findpeaks(g(1:nfft/2),f(1:nfft/2));**

**%tim tan so lon nhat trong khoang (80,400)**

**f\_max = 0;**

**for i=1:1:length(locs)**

**if((locs(i) >= 80) && (locs(i) <= 400) && (locs(i) > f\_max))**

**f\_max = locs(i);**

**end**

**end**

**%tim tan so co ban trong khoang (80,400)**

**foo = 0;**

**for i=1:1:length(locs)**

**if((locs(i) >= 80) && (locs(i) <= 400))**

**if ((mod(f\_max, locs(i))/locs(i)) == 0)**

**foo = locs(i);**

**break;**

**end**

**end**

**end**

# KẾT QUẢ THỰC NGHIỆM

## Hình vẽ

## Từ hình 4 tới hình 10 biểu diễn hình vẽ tín hiệu của 5 nguyên âm cơ bản và hàm tự tương quan của nó với số mẫu là 1323 tương ứng với file audio 30ms

Hình 6.Kết quả đo nguyên âm a với hình vẽ tín hiệu gốc và hàm tự tương quan



Hình 7.Kết quả đo nguyên âm e với hình vẽ tín hiệu gốc và hàm tự tương quan



Hình 8.Kết quả đo nguyên âm i với hình vẽ tín hiệu gốc và hàm tự tương quan

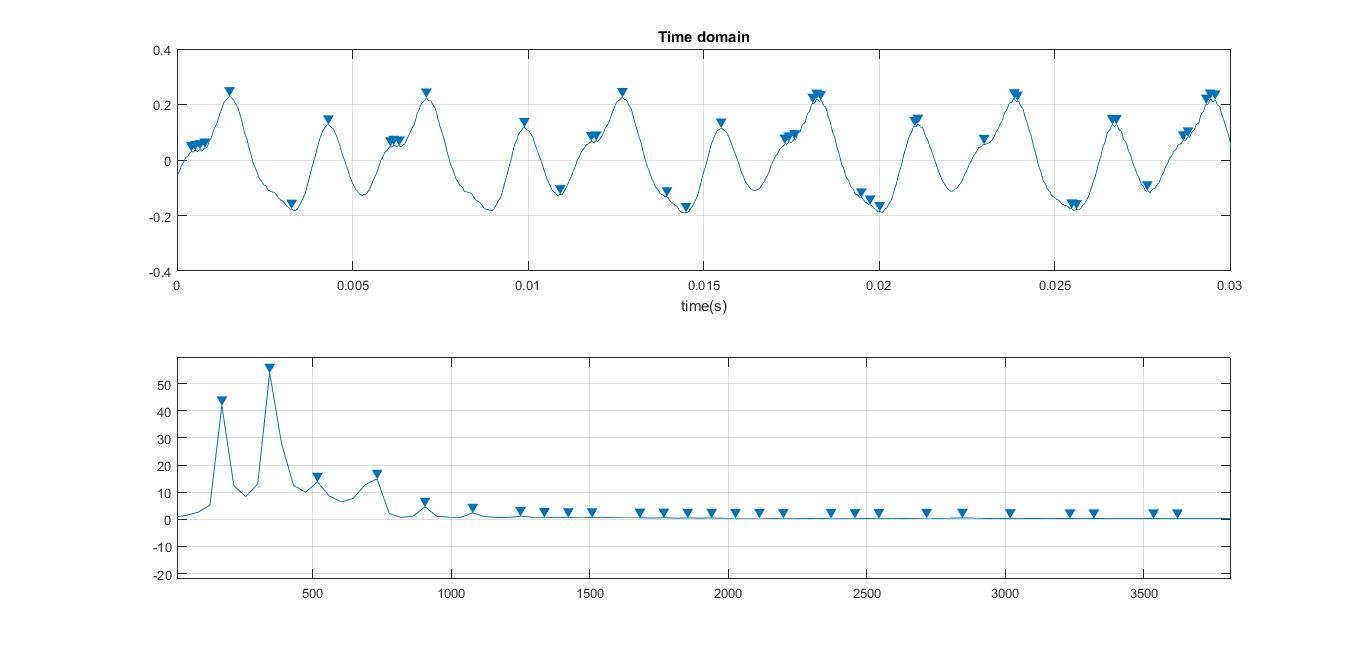


Hình 9.Kết quả đo nguyên âm o với hình vẽ tín hiệu gốc và hàm tự tương quan

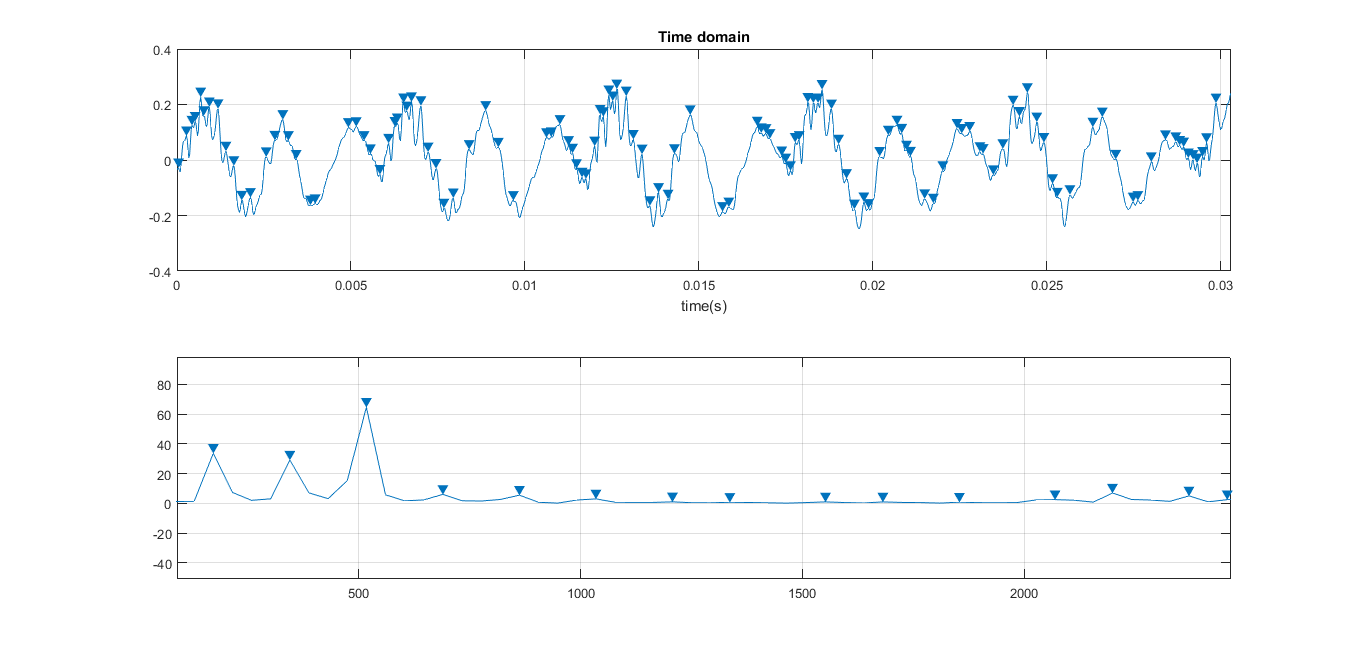


Hình 10.Kết quả đo nguyên âm u với hình vẽ tín hiệu gốc và hàm tự tương quan

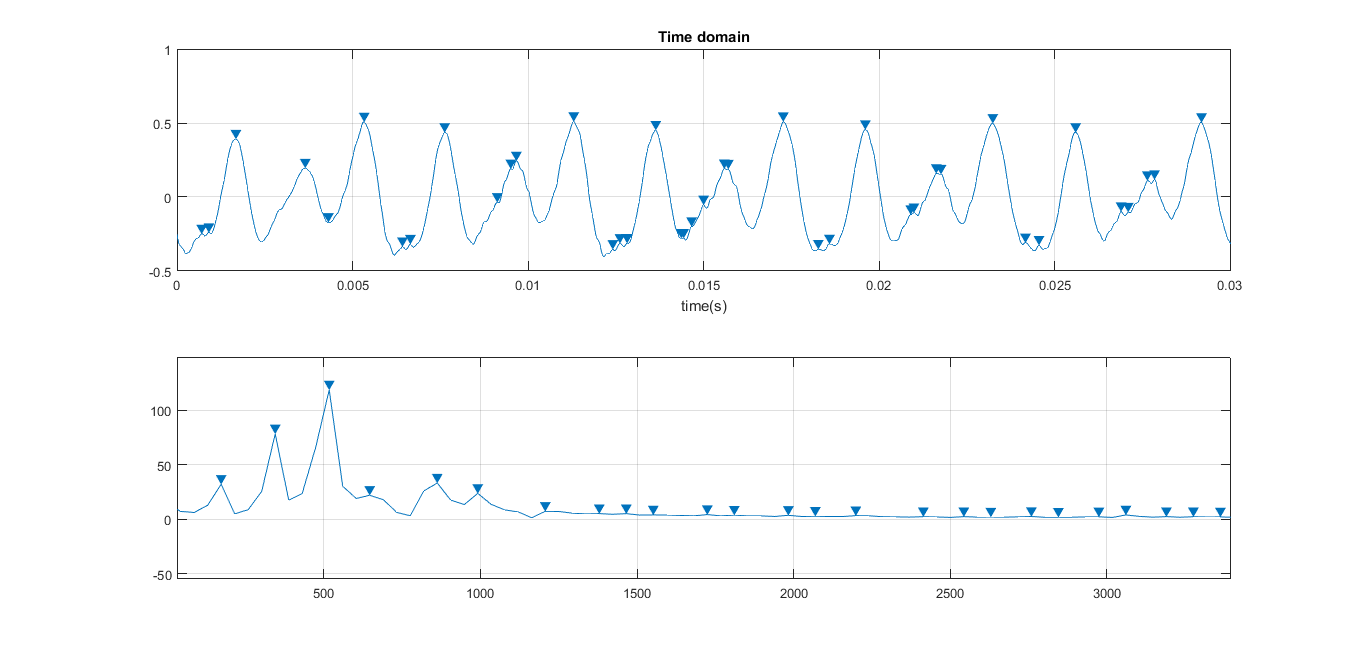
Từ ***hình 11 tới hình 15*** biểu diễn hình vẽ tín hiệu của 5 nguyên âm cơ bản và hàm tự tương quan của nó với số mẫu là 1323 tương ứng với file audio 30ms



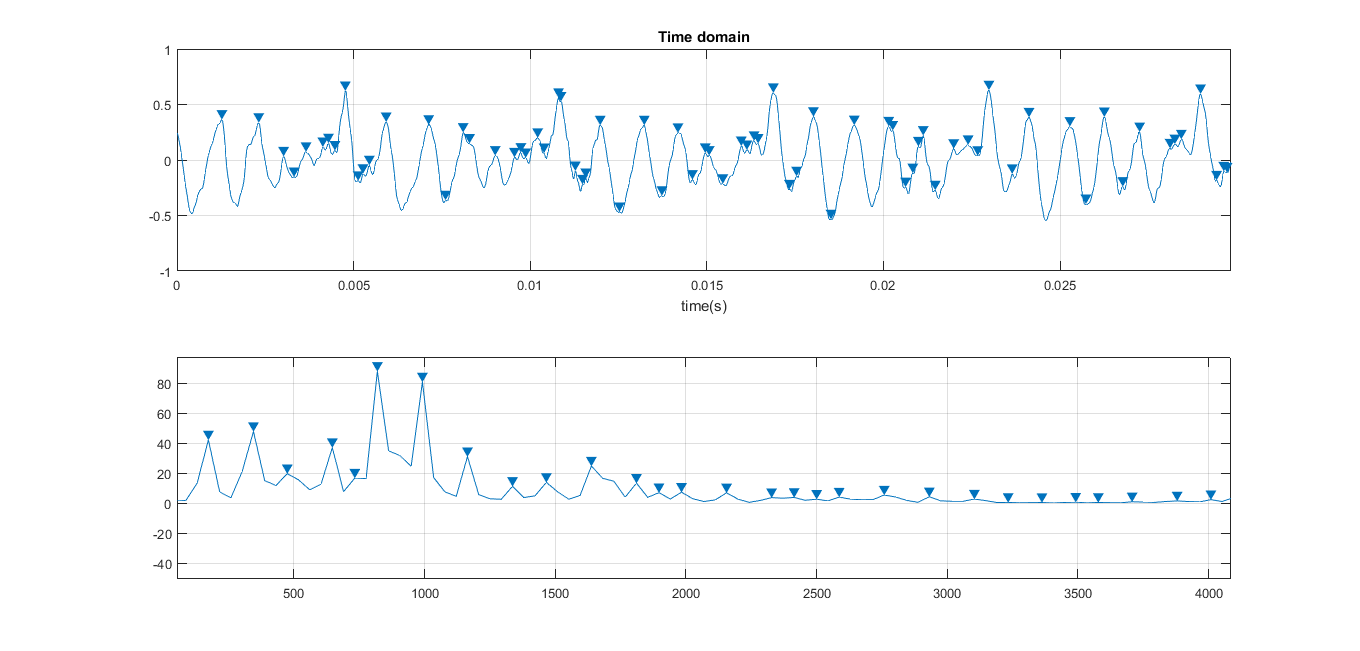
**Hình 11.Kết quả đo nguyên âm u trên miền thời gian và tần số**



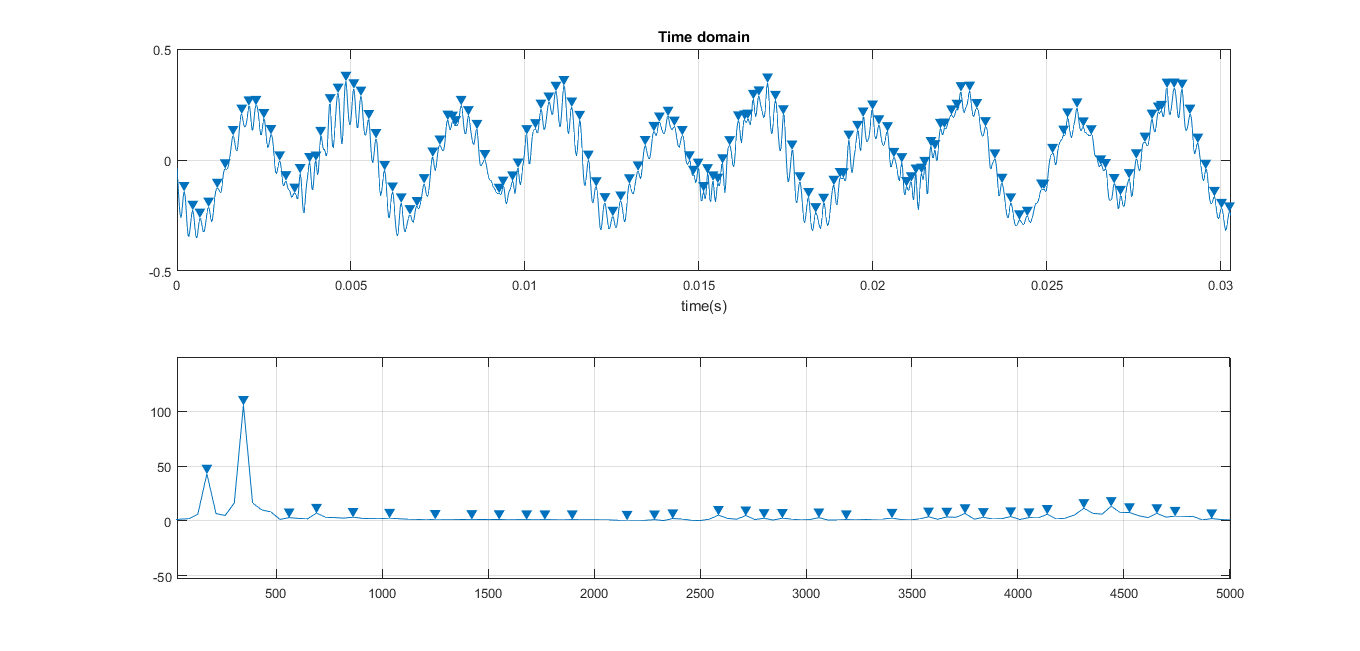
**Hình 12.Kết quả đo nguyên âm e trên miền thời gian và tần số**



**Hình 13.Kết quả đo nguyên âm o trên miền thời gian và tần số**



**Hình 14.Kết quả đo nguyên âm a trên miền thời gian và tần số**



**Hình 15.Kết quả đo nguyên âm i trên miền thời gian và tần số**

**B.Bảng biểu:**

Bảng 1.Bảng kết quả tính toán bằng hàm tự tương quan

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Âm a | Âm e | Âm i | Âm o | Âm u |
| F0(Hz) | 126 | 120.16 | 139.11 | 114.55 | 130.86 |
| F0 trung bình | 126.136 | | | | |
| Độ lệch chuẩn | 7.08 | | | | |

Bảng 2.Bảng so sánh các thông số kết quả đo tự động và thủ công

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | | Âm a | Âm e | Âm i | Âm o | Âm u |
| F0(Hz) | Đo tự động | 126 | 120.16 | 139.11 | 114.55 | 130.86 |
| Đo thủ công | 125.28 | 121.15 | 139.56 | 115.14 | 132.03 |
| F0 Trung bình | Đo tự động | 126.136 | | | | |
| Đo thủ công | 126.632 | | | | |
| Độ lệch chuẩn | Đo tự động | 7.08 | | | | |
| Đo thủ công | 7.33 | | | | |

Bảng 4.Bảng kết quả tính toán trên miền tần số

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Âm a | Âm e | Âm i | Âm o | Âm u |
| F0(Hz) | 172.434 | 172.434 | 172.434 | 172.434 | 172.434 |
| F0 trung bình | 172.434 | | | | |
| Độ lệch chuẩn | 0 | | | | |

Bảng 2.Bảng so sánh các thông số kết quả đo tự động và thủ công

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | | Âm a | Âm e | Âm i | Âm o | Âm u |
| F0(Hz) | Đo tự động | 172.434 | 172.434 | 172.434 | 172.434 | 172.434 |
| Đo thủ công | 164.474 | 170.155 | 169.642 | 166.251 | 178.987 |
| F0 Trung bình | Đo tự động | 172.434 | | | | |
| Đo thủ công | 169.902 | | | | |
| Độ lệch chuẩn | Đo tự động | 0 | | | | |
| Đo thủ công | 3.735 | | | | |

Từ kết quả ở Bảng 1 cho thấy F0 trung bình có giá trị là 126.136 Hz với sai số trung bình là 7.08 Hz , mức sai số tương đối thấp nhưng cho thấy thuật toán tính tần số cơ bản dùng hàm tự tương quan vẫn chưa đạt được độ chính xác cao.

Từ kết quả Bảng 2 cho thấy mức độ sai lệch của đo tự động và đo thủ công là không đáng kể ở từng nguyên âm, và sai lệch ở kết quả trung bình là 126.632 Hz của đo thủ công so với 126.136 của đo tự động tức lệch 126.632-126.136=0.496 Hz , mức sai lệch không đáng kể. Kết quả cho thấy việc tính toán bằng hàm tự tương quan và tính toán bằng thủ công mang lại kết quả gần như nhau tuy nhiên sai số ở 2 phép đo này tương đối đáng kể do chất lượng tín hiệu chọn làm mẫu đo.

# KẾT LUẬN

Bài báo cáo này thực hiện việc đo tần số cơ bản của giọng nói trên miền thời gian dùng hàm tự tương quan và trên miền tần số dùng thuật toán Harmonic Product Spectrum sau đó lọc kết quả với lọc trung vị. Các phép đo thực hiện với tín hiệu là các đoạn ghi âm các nguyên âm cơ bản của giọng nói đem lại kết quả gần chính xác với việc đo thủ công.Với sai số là 7.08 Hz và lệch 0.496Hz so với đo thủ công của việc đo trên miền thời gian bằng hàm tự tương quan.Với sai số là …….//Tài viết tiếp chỗ ni giống như trên.Trong tương lại chúng tôi dự định dùng các thuật toán này áp dụng phát triển các ứng dụng nhận diện người dùng bằng giọng nói.Vì mỗi người có một tần số cơ bản riêng nên xây dựng 1 phần mềm tính toán so khớp tần số cơ bản để nhận diện giọng nói là ý tưởng chúng tôi hướng tới trong tương lai.

# NHỮNG ĐIỀU ĐÃ HỌC ĐƯỢC

Qua bài tập lớn về chủ đề đo tự động tần số cơ bản của giọng nói, chúng tôi học được và trau dồi được thêm những kĩ năng về đọc và nghiên cứu tài liệu, đặc biệt là tài liệu tiếng Anh , học được thêm kĩ năng viết một bài báo cáo theo tiêu chuẩn của một bài báo cáo nghiên cứu khoa học. Và các kĩ năng quan trọng khác được trau dồi là việc làm việc nhóm để phân công công việc, tổng hợp công việc của các thành viên để làm thành 1 bài báo cáo hoàn chỉnh. Qua bài tập lớn này chúng tôi xin gởi lời cảm ơn chân thành tới các thành viên của nhóm đã làm việc tích cực và đặc biệt là giáo viên hướng dẫn – Thầy Ninh Khánh Duy đã hướng dẫn , tạo điều kiện để chúng tôi thực hiện được bài báo cáo này.

# TÀI LIỆU THAM KHẢO

1. The frequency range of the voice fundamental in the speech of male and female adults .Institutionen för lingvistik, Stockholms universitet, S-106 91 Stockholm, Sweden.
2. ,[6],[7] Nguyễn Bình Thiên, Ninh Khánh Duy, CẢI TIẾN THUẬT TOÁN TỰ TƯƠNG QUAN TÌM CAO ĐỘ CỦA TÍN HIỆU ĐÀN GHI-TA TRÊN VI XỬ LÝ ARM CORTEX-M4
3. Wikipedia - Voice frequency
4. Phạm Văn Sự, Lê Xuân Thành, Bai giang XLTN\_HV CNBCVT
5. Charlet ReedStrom ,Pitch Detection Algorithms

**(chú ý chỉ đưa vào các tài liệu có trích dẫn [1], [2], … trong báo cáo)**