

**PHÂN ĐOẠN TÍN HIỆU THÀNH TIẾNG NÓI VÀ KHOẢNG LẶNG DÙNG
MAGNITUDE AVERAGE, SHORT-TIME ENERGY VÀ ZERO CROSSING
RATE**

Bùi Duy Phương, Thái Văn Thọ, Nguyễn Thị Nga, Nguyễn Văn Hải

Nhóm 7, lớp HP: 1022470.2010.19.16

Điểm	Bảng phân công nhiệm vụ		Chữ ký của SV
	Bùi Duy Phương (nhóm trưởng)	Phân công nhiệm vụ và đảm bảo tiến độ của mỗi thành viên. Đọc tài liệu, cài đặt thuật toán Short Time Energy.	
	Thái Văn Thọ	Đọc tài liệu và cài đặt thuật toán về Magnitude Average.	
	Nguyễn Thị Nga	Đọc tài liệu và cài đặt thuật toán về Zero Crossing Rate kết hợp với Short Time Energy.	
	Nguyễn Văn Hải	Đặt vấn đề và vấn đề cần giải quyết. Đọc tài liệu và viết báo cáo và làm Slide.	

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 VI. 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— Phân đoạn tín hiệu thành tiếng nói và khoảng lặng là một bài toán cần thiết trong cuộc sống hiện đại. Bài báo này thực hiện việc phân đoạn tín hiệu dùng thuật toán Magnitude Average, Short Time Energy và thuật toán kết hợp với Zero Crossing Rate. Các thử nghiệm với 4 file tín hiệu với 2 môi trường khác nhau là lab_male.wav, lab_female.wav, studio_male.wav, studio_female.wav. Từ kết quả thực nghiệm cho ta thấy thuật toán Short Time Energy kết hợp Zero Crossing Rate cho kết quả ổn định và đúng hơn các phương pháp đã dùng.

Từ khóa— Short Time Energy(STE), Zero Crossing Rate(ZCR), Magnitude Average(MA).

Mục lục

I. ĐẶT VẤN ĐỀ..... 3

II. LÝ THUYẾT XỬ LÝ TÍN HIỆU TIẾNG NÓI VÀ CÁC THUẬT TOÁN 3

A. Vấn đề cần giải quyết 3

B. Thuật toán Short-Time Energy 3

1. Cơ sở lý thuyết..... 3

2. Sơ đồ thuật toán..... 4

C. Thuật toán kết hợp Zero Crossing Rate và Short-Time Energy 5

1. Cơ sở lý thuyết..... 5

2. Sơ đồ thuật toán..... 6

D. Thuật toán Magnitude Average..... 7

1. Cơ sở lý thuyết..... 7

2. Sơ đồ thuật toán..... 7

III. MÃ CHƯƠNG TRÌNH CÀI ĐẶT CÁC THUẬT TOÁN..... 8

A. Thuật toán Short Time Energy. 8

B. Thuật toán kết hợp Zero Crossing Rate và Short Time Energy 10

C. Thuật toán Magnitude Average..... 13

IV. KẾT QUẢ THỰC NGHIỆM 15

A. Kết quả định tính..... 15

1. Thuật toán Short – time Energy..... 15

2. Thuật toán Zero Crossing Rate kết hợp với Short Time Energy 17

3. Thuật toán Magnitude Average 19

B. Kết quả định lượng..... 21

V. KẾT LUẬN..... 22

VI. TÀI LIỆU THAM KHẢO 22

I. ĐẶT VẤN ĐỀ

Tiếng nói là một phương tiện giao tiếp cơ bản và quan trọng nhất giữa con người với con người. Hiện nay, tiếng nói còn được áp dụng để thực hiện việc giao tiếp giữa con người với máy móc nhằm hướng tới mục đích thay thế các phương pháp giao tiếp truyền thống như bàn phím, chuột, công tắc,... Vì vậy xử lý tiếng nói hay nói chính xác là nhận dạng tiếng nói đóng vai trò quan trọng trong vấn đề này. Tuy nhiên, cùng với sự phát triển của thế giới, ô nhiễm tiếng ồn ngày càng phức tạp hơn. Để làm được điều đó, việc phân đoạn tín hiệu tiếng nói là rất quan trọng.

Phân đoạn tín hiệu tiếng nói (speech signal segmentation) là quá trình xác định ranh giới giữa các từ, âm tiết hoặc âm vị trong các ngôn ngữ tự nhiên được nói [1]. Ở trong bài báo cáo này chúng tôi sẽ giới thiệu về ba phương pháp phân đoạn tín hiệu tiếng nói thường dùng là: Magnitude Average (MA), Short-Time Energy (STE) và Zero Crossing Rate (ZCR).

Báo cáo có bố cục như sau: Phần II trình bày về cơ sở lý của các thuật toán và các vấn đề liên quan đến việc phân đoạn tiếng nói dựa trên Short-time energy (Năng lượng ngắn hạn) và Zero Crossing Rate (Tốc độ băng qua 0). Phần III trình bày mã nguồn cài đặt các thuật toán. Phần IV trình bày kết quả thực nghiệm mô tả dữ liệu dùng để đánh giá độ chính xác của thuật toán, đưa ra các đánh giá định tính và định lượng, so sánh các thuật toán đã cài đặt với nhau. Phần V trình bày kết luận, tóm lại các kết quả đã đạt được và cải thiện trong tương lai.

II. LÝ THUYẾT XỬ LÝ TÍN HIỆU TIẾNG NÓI VÀ CÁC THUẬT TOÁN

A. Vấn đề cần giải quyết

Xử lý tiếng nói đang bùng nổ mạnh trong thời đại công nghệ hiện nay. Xử lý tiếng nói có thể ứng dụng trong nhiều lĩnh vực, đặc biệt trong việc nhận dạng tiếng nói. Trong quá trình nhận dạng tiếng nói, việc phân đoạn tín hiệu tiếng nói là vô cùng quan trọng trong bất cứ bài toán nào. Vậy phân đoạn tín hiệu tiếng nói là gì? Phân đoạn tín hiệu tiếng nói (speech signal segmentation) là quá trình xác định ranh giới giữa các từ, âm tiết hoặc âm vị trong các ngôn ngữ tự nhiên được nói hay nói một cách đơn giản là xác định ranh giới giữa tiếng nói và khoảng lặng.

Có nhiều phương pháp để có thể phân đoạn tín hiệu tiếng nói khác nhau như: Zero Crossing Rate (ZCR), Short-Time Energy (STE), Pre-Emphasized Energy Ratio (PEER),... Ở trong bài báo cáo này chúng tôi sẽ giới thiệu về ba phương pháp phân đoạn tín hiệu tiếng nói thường dùng là: Magnitude Average (MA), Zero Crossing Rate (ZCR) và Short-Time Energy (STE). Zero Crossing Rate là tốc độ thay đổi của tín hiệu dọc theo tín hiệu, nghĩa là tốc độ tín hiệu thay đổi từ dương sang 0 thành âm hoặc ngược lại. Short-Time Energy năng lượng của các mẫu thuộc một khung tín hiệu.

B. Thuật toán Short-Time Energy

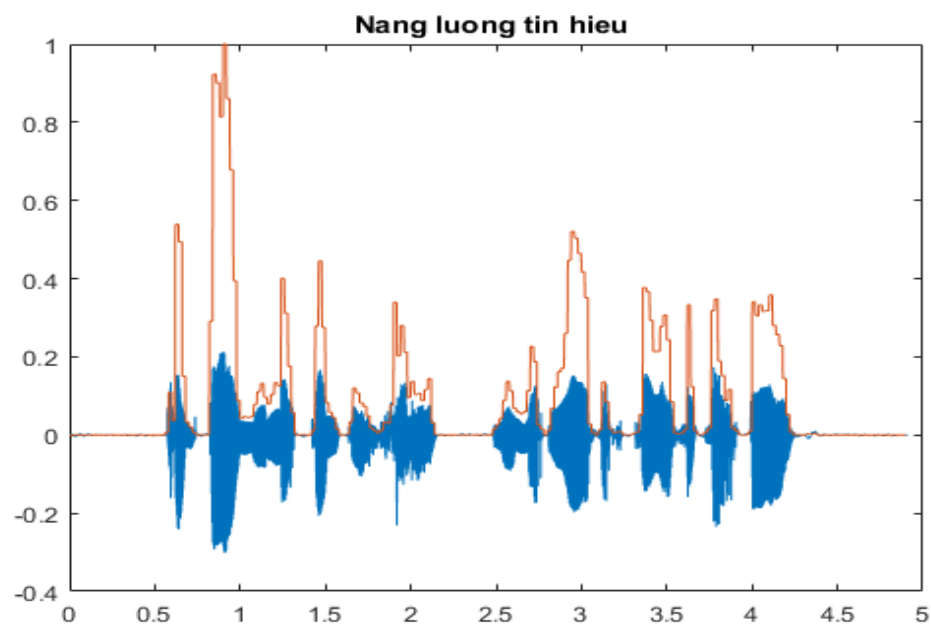
1. Cơ sở lý thuyết

Năng lượng là tổng bình phương biên độ của các mẫu tín hiệu:

$$E_x = \sum_{-\infty}^{+\infty} |x[n]|^2$$

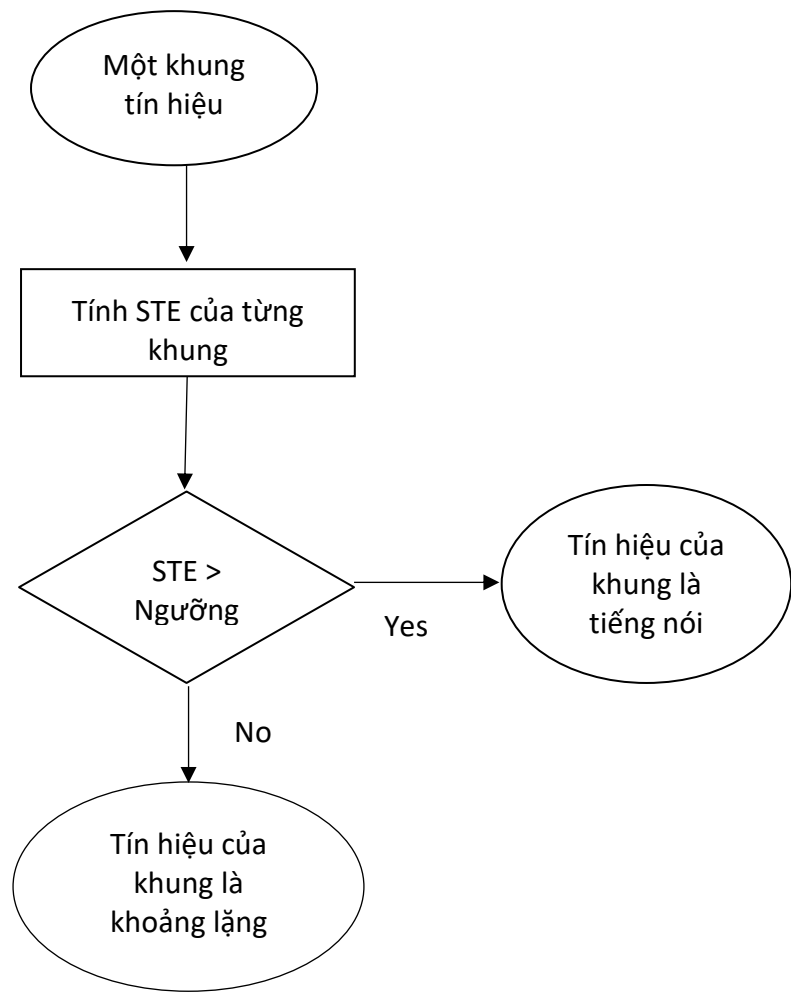
Bởi vì năng lượng của tín hiệu phụ thuộc vào biên độ của các mẫu tín hiệu, khi phân đoạn một tín hiệu âm thanh, có thể nhận xét được rằng:

- Các tín hiệu tiếng nói sẽ có giá trị năng lượng lớn.
- Các tín hiệu khoảng lặng sẽ có giá trị năng lượng nhỏ.



Hình 1. Ví dụ thuật toán Short-Time Energy trên file studio_female.wav

2. Sơ đồ thuật toán



- Chú ý:
 - Ngưỡng dựa trên dữ liệu STE của khung tín hiệu đã được chuẩn hóa trong đoạn [0;1].
 - Điều kiện ràng buộc bổ sung: Khoảng lặng là các khoảng tín hiệu có độ dài tối thiểu là 200ms để:
 - Loại trừ các trường hợp mà thuật toán dùng ngưỡng năng lượng xác định không đúng tiếng nói và khoảng lặng.
 - Loại trừ những khoảng tín hiệu có năng lượng nhỏ nhưng không phải là khoảng lặng.

C. Thuật toán kết hợp Zero Crossing Rate và Short-Time Energy

1. Cơ sở lý thuyết

- Cơ sở lý thuyết về thuật toán Zero Crossing Rate:[3]

+ Công thức hàm Zero-Crossing Rate có thể được biểu diễn như sau:

$$ZCR_n = \sum_{m=-\infty}^{\infty} |sgn[x(m)] - sgn[x(m-1)]|w(n-m)$$

➤ Với:

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

$$w(n) = \begin{cases} \frac{1}{2N} & 0 \leq n \leq N-1 \\ 0 & \text{ngược lại} \end{cases}$$

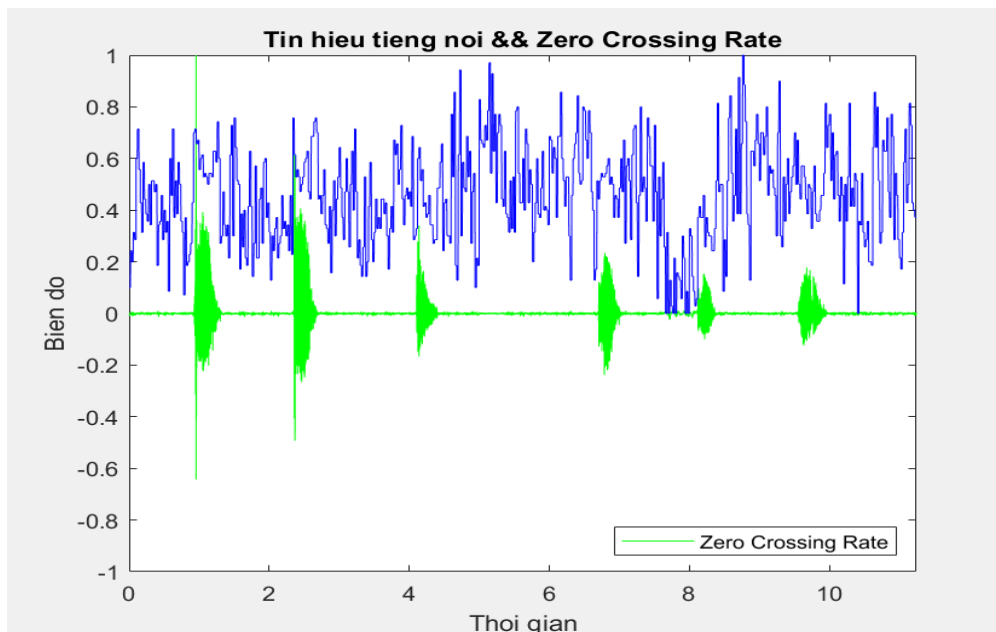
Sau khi chia tín hiệu thành các khung nhỏ, với mỗi khung m , ZCR_m là số lần biên độ tín hiệu băng qua 0 đối với các mẫu thuộc khung :

$$ZCR_m = \lim_{n \rightarrow N} \left(\frac{\text{số lần } (x[n] * x[n+1]) < 0}{2L+1} \right) \quad \text{với } N \in [-L, L]$$

- ZCR có những đặc điểm sau:

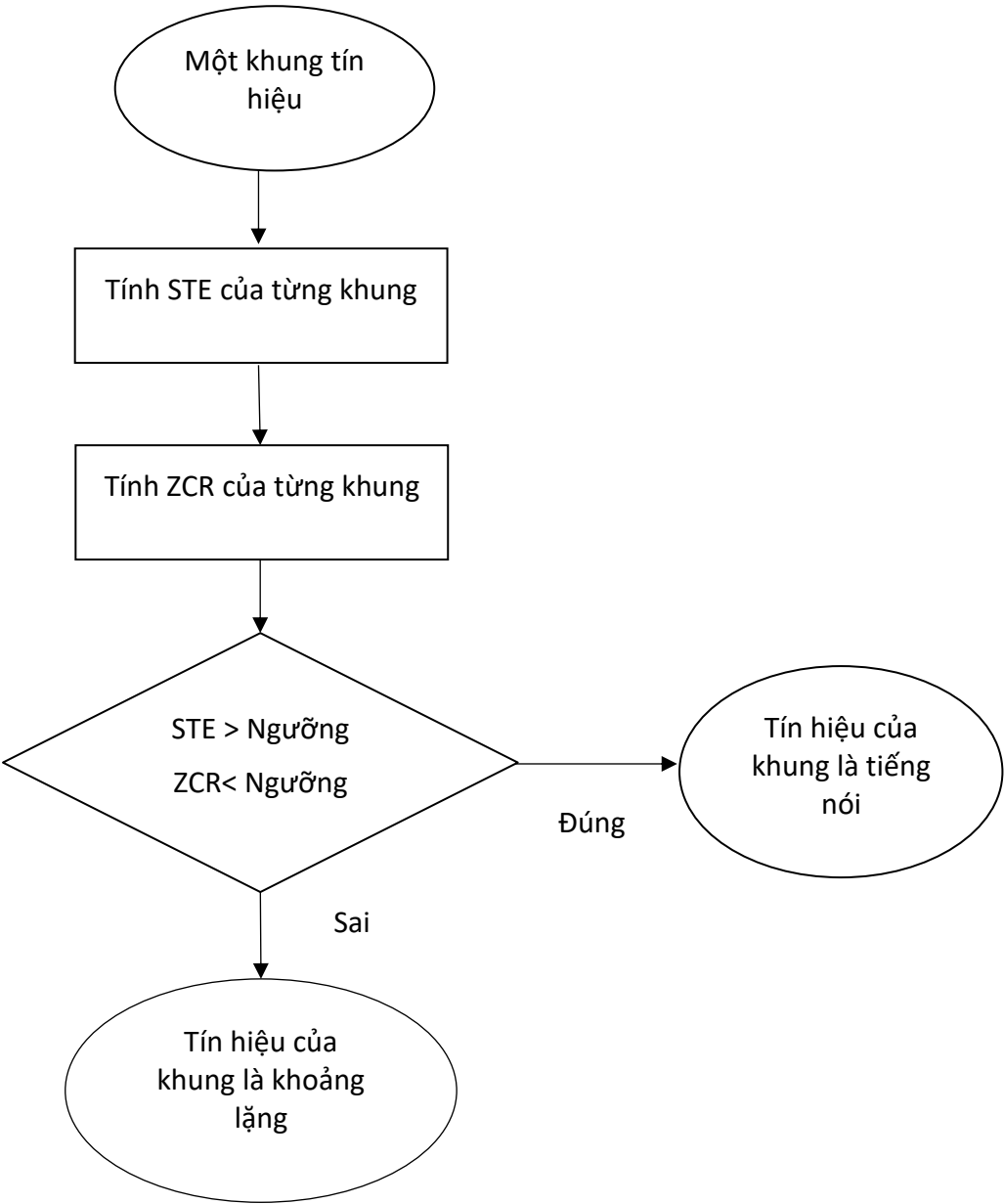
+ ZCR của khoảng lặng và tiếng ồn môi trường thường lớn hơn âm thanh phát ra.

+ Khó có thể phân biệt khoảng lặng từ tiếng ồn môi trường bằng cách sử dụng một mình ZCR, vì chúng có giá trị ZCR tương tự nhau.



Hình 2. Ví dụ thuật toán ZCR trên file lab_male.wav

2. Sơ đồ thuật toán



D. Thuật toán Magnitude Average

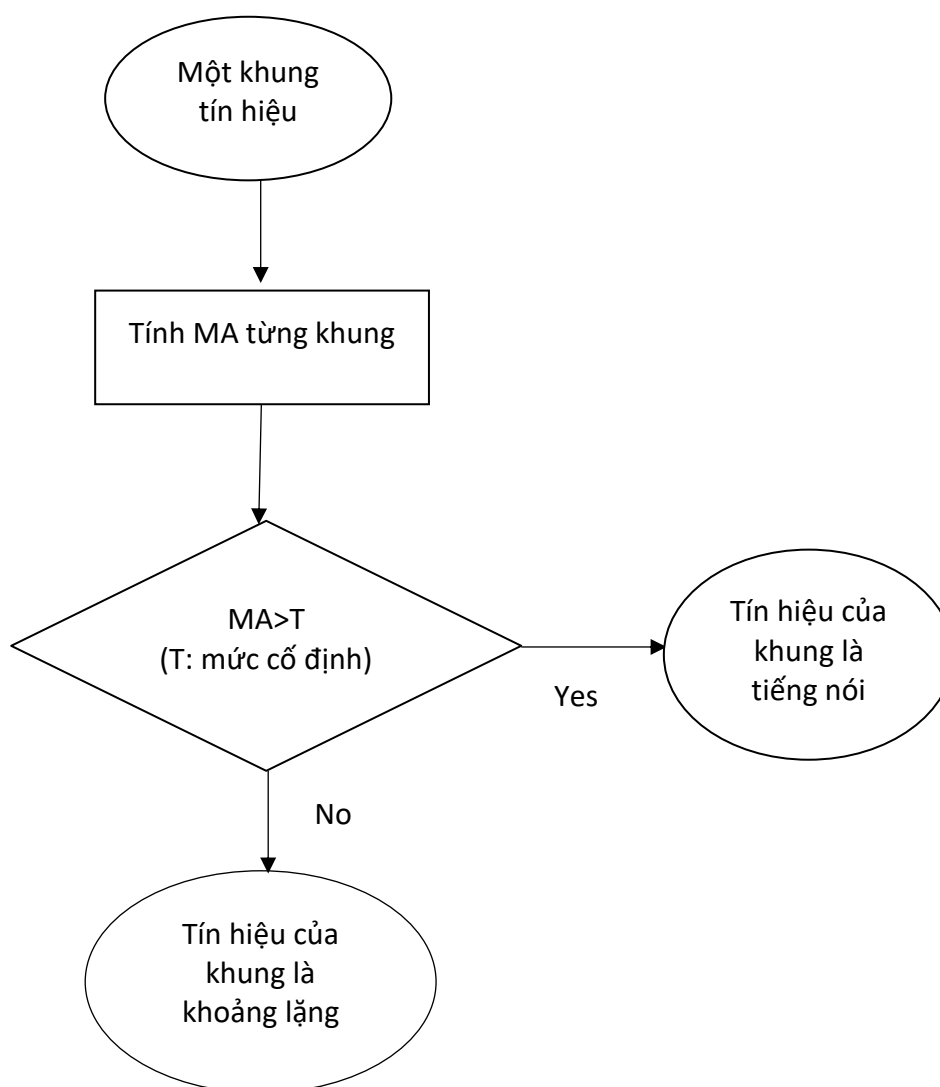
1. Thuật toán Magnitude Average được định nghĩa như sau

$$MA[n] = \sum_{m=0}^{N-1} |x[m - n]|$$

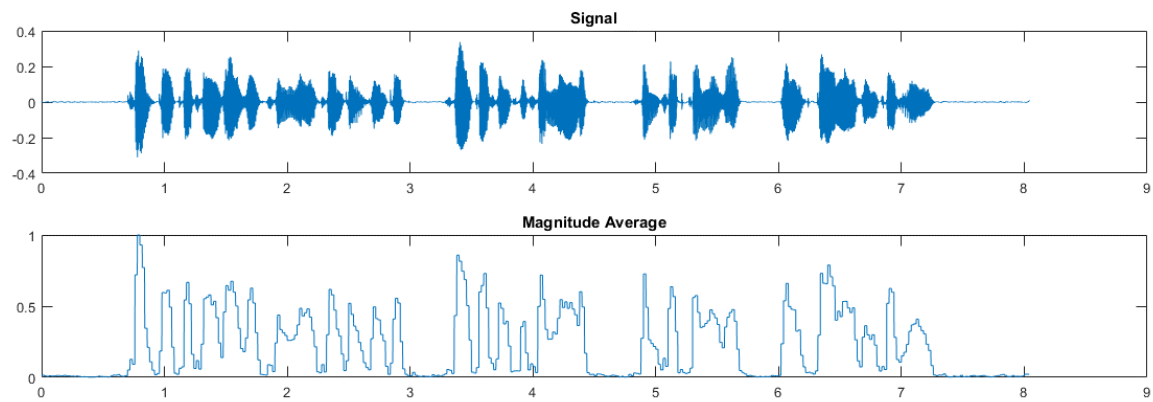
Hàm MA được tính bằng tổng giá trị tuyệt đối của những giá trị dạng sóng giống nhau (same waveform values)

Nếu tất cả các giá trị MA cho những phân đoạn tiếng nói đều lớn hơn so với phân đoạn khoảng lặng, sẽ có một mốc T mà $MA_{\text{speech}} > T$ và $MA_{\text{silence}} < T$. Việc tìm mức T là không quá quan trọng và T sẽ dùng để xác định chắc chắn mức cơ bản cho bất kì khoảng thời gian nào

2. Sơ đồ thuật toán



Ví dụ về sử dụng hàm Magnitude Average



Hình 3 .Ví dụ về sử dụng hàm MA(Magnitude Average)

Từ hình ta có thể thấy tín hiệu sau khi được xử lý theo hàm MA như trên. Ta có thể dễ dàng tìm được mốc để phân chia tiếng nói và khoảng lặng, nên trong chương trình, ta sẽ cho một mốc cố định để phân chia mà vẫn đảm bảo yêu cầu.

III. MÃ CHƯƠNG TRÌNH CÀI ĐẶT CÁC THUẬT TOÁN

A. Thuật toán Short Time Energy.

3. Function tính năng lượng tín hiệu

```
%# Ham tính song Nang luong. Nhan vao la Frames.
%# Doi tuong ra ve la STE cua tung frame da qua chuan hoa
function [ste] = STEFunc(frames)
    [frameNum, frameWidth] = size(frames); %so khung la hang va do dai khung la cot
    ste = 0; %#Short-Time Energy
    for i = 1 : frameNum %vong lap chay den tung hang cua so khung
        ste(i) = sum(frames(i,:).^2); %#Ap dung cong thuc!
    end
    %#Chuan hoa STE
    ste = (ste - min(ste))/(max(ste)-min(ste));
end
```

4. Hàm chính

```
%[y,fs]=audioread(studio_female.wav');
[y,fs]=audioread(studio_male.wav');
[y,fs]=audioread(lab_female.wav');
[y,fs]=audioread('lab_male.wav');
sound(y,fs);

nguongchung = 0.001; %# Nguong nang luong, > tieng noi, < khoang lang.
```



```
%%%%%%%%%
```

```
%chia khung tin hieu theo thoi gian
```

```
ThoiLuongKhung= 0.02; %s=20ms;
```

```
DoDaiKhung=ThoiLuongKhung * fs; %so mau trong 1 khung
```

```
SoLuongKhung= floor(length(y)/DoDaiKhung); %so luong khung trong tin hieu y
```

```
Khung=zeros(SoLuongKhung,DoDaiKhung); %tao ma tran cac khung co SoLuongKhung cot va DoDaiKhung hang
```

```
for k=1:SoLuongKhung
```

```
    Khung(k,:)=y(DoDaiKhung*(k-1)+1 :DoDaiKhung*k);%chia tung khung cua tin hieu y vao tung khung
```

```
end
```

```
%%%%%%%%%
```

```
%tinh nang luong cua tung khung
```

```
ste = STEFunc(Khung);
```

```
%ghep cac nang luong tung khung thanh tin hieu song
```

```
ste_wave = 0; %tin hieu nang luong song
```

```
for j = 1 : length(ste)
```

```
    l = length(ste_wave);
```

```
    ste_wave(l : l + DoDaiKhung) = ste(j);
```

```
end
```

```
t = linspace(0,length(y)/fs,length(y)); %tao vecto t gom length(y) diem cach deu nhau tu 0 den length(y)/fs
```

```
t1 = linspace(0,length(ste_wave)/fs,length(ste_wave));%tao vecto t1
```

```
subplot(3,1,1);
```

```
plot(t,y); %ve tin hieu vao
```

```
title('tin hieu vao');
```

```
subplot(3,1,2);
```

```
plot(t1,ste_wave); %ve nang luong cua tin hieu
```

```
title('Nang luong tin hieu ');
```

```
a = [];
```

```
%# Danh dau cac diem > nguong = 1, <nguong = 0
```

```
for diem = 1 : length(ste) % vong lap cac diem tu 1 den do dai cua nang luong
```

```
    if (ste(diem) > nguongchung) %kiem tra dieu kien
```

```
        a = [a 1]; %dung thi gan bang 1 va dua vao vector a
```

```
    else
```

```
        a = [a 0]; %sai thi gan bang 0 va dua vao vector a
```

```
end
```

```

end
%# Kiểm tra điều kiện : không tồn tại khoảng lặng nhỏ hơn 200ms
for i = 1 : length(a)-10 % 200ms = 10 khung tín hiệu
    if (a(i) == 1 && a(i+10) == 1) %nếu điều kiện trên đúng thì các phần tử i đến i+10 sẽ gán bằng 1
        a(i : i+10) = 1;
    end
end
%vẽ đường phân chia tiếng nói khoảng lặng

for i=1:length(a)-1
    if (a(i) == 0 && a(i+1) == 1) || (a(i+1) == 0 && a(i) == 1)
        subplot(3,1,3);
        plot(t,y); hold on; %vẽ y
        plot([i*ThoiLuongKhung i*ThoiLuongKhung], [-1 1], '--k');
    end
end
end

```

B. Thuật toán kết hợp Zero Crossing Rate và Short Time Energy .

1. Hàm function Short Time Energy:

```

function [ste, ste_mau] = STE(khung, DoDaiKhung)
    [frameNum, frameWidth] = size(khung);
    ste = 0;
    for i = 1 : frameNum
        ste(i) = sum(khung(i,:).^2);
    end

    ste = ste./max(ste); %chuan hoa
    ste_mau = 0;
    for j = 1 : length(ste)
        l = length(ste_mau);
        ste_mau(l : l + DoDaiKhung) = ste(j);
    end
end
end

```

2. Hàm function Zero Crossing Rate:

```

function [zcr, zcr_mau] = ZCR(khung, N, DoDaiKhung)
[r, c] = size(khung);
for i = 1 : r
    x = khung(i, :);
    %dem so lan bang qua 0
    zc(i) = 0;
    for k = 1:length(x) - 1
        if (x(k) * x(k + 1) < 0)
            zc(i) = zc(i) + 1;
        end
    end
end
end

zcr = zc / N; %tox do bang qua 0 ZCR
zcr = zcr/max(zcr); %chuan hoa
zcr_mau = 0;
for j = 1 : length(zcr)
    l = length(zcr_mau);
    zcr_mau(l + 1 + DoDaiKhung) = zcr(j);
end

```

3. Hàm chính:

```

%[data, Fs] = audioread('lab_male.wav');
%[data, Fs] = audioread('lab_female.wav');
%[data, Fs] = audioread('studio_male.wav');
[data, Fs] = audioread('studio_female.wav');

% Chuan hoa du lieu
data = data / abs(max(data));

ThoiLuongKhung = 0.02; %10-30ms
N = length(data);
DoDaiKhung = ThoiLuongKhung * Fs;

%% Chia khung
n = floor(N / DoDaiKhung); %Tong so frame
tmp = 0;
for i = 1 : n
    khung(i,:) = data(tmp + 1 : tmp + DoDaiKhung);

```

```

tmp = tmp + DoDaiKhung;
end

%Zero Crossing Rate
[zcr, zcr_mau] = ZCR(khung, N, DoDaiKhung);
%Short-Time Energy(STE)
[ste, ste_mau] = STE(khung, DoDaiKhung);

%% Ve do thi ZCR,STE,data
t = (0 : 1/Fs : length(data) / Fs);
t = t(1:end - 1);
t1 = (0 : 1/Fs : length(zcr_mau) / Fs);
t1 = t1(1:end - 1);
t2 = (0 : 1/Fs : length(ste_mau) / Fs);
t2 = t2(1:end - 1);
subplot(3,1,1);
plot(t, data, 'g');
axis([0 n*ThoiLuongKhung -1 1]);
legend({'Tin hieu tieng noi', 'Location','southeast'});
title('Tin hieu tieng noi');
xlabel('Thoi gian');
ylabel('Bien do');
subplot(3,1,2);
plot(t1, zcr_mau,'b', t2, ste_mau,'r');
axis([0 n*ThoiLuongKhung -1 1]);
legend({'Zero Crossing Rate', 'Short time energy'}, 'Location','southeast');
title('Short-time Energy - Zero-crossing Rate');
xlabel('Thoi gian');
ylabel('Bien do');

%% Xac dinh vi tri cho khoang lang va tieng noi
ZCRct = 0.8; Ect = 0.001;%thay doi nguoi de co duoc ket qua chinh xac
id_zcr = find(zcr < ZCRct);
id_ste = find(ste > Ect);
id = [];
for i = 1:n
    j = 1;
    while j <= length(id_zcr)
        if(i == id_zcr(j))
            k = 1;

```

```

while k <= length(id_ste)
    if(i == id_ste(k))
        id = [id,i];
    end
    k=k+1;
end
end
j=j+1;
end
end

%% Tim bien thoi gian giao nhau giua tieng noi va khoang lang
id_speech=id(1)-1;
m=2;
for i = 2:length(id)
    if((ThoiLuongKhung*id(i)-ThoiLuongKhung*id(i-1)) > 0.2)
        id_speech(m)=id(i-1);%bat dau
        id_speech(m+1)=id(i)-1;%ket thuc
        m=m+2;
    end
end
id_speech(m)=id(i);
local_speech=ThoiLuongKhung*id_speech;
% ve do thi phan doan tieng noi va khoang lang
y = (-1: 1);
subplot(3,1,3);
plot(t, data, 'g');
axis([0 n*ThoiLuongKhung -1 1]);
hold on;
for i = 1:length(local_speech)
    plot(local_speech(i)*ones(size(y)), y,'m --');
end
hold off;
legend('Tin hieu tieng noi','Vi tri bat dau va ket thuc tieng noi','Location','southeast');
title('Phan doan tieng noi va khoang lang');
xlabel('Thoi gian');
ylabel('Bien do');

```

C. Thuật toán Magnitude Average

1. Hàm function tính Magnitude Average

```

%ham tinh Magnitude Average. Nhap vao la la frames
%doi tuong ve ra la MA cua tung khung da chuan hoa
function [ma] = MA(frames)
    [frameNum, frameWidth] = size(frames);
    ma = 0;%magnitude average
    for i = 1:frameNum
        ma(i) = sum(abs(frames(i,:)));
        %ma(i) = abs(log10(ma(i)));
    end
    %chuan hoa Ma
    ma = ma/max(ma);
end

```

2. Hàm chính

```

%[data, Fs] = audioread('lab_male.wav');
%[data, Fs] = audioread('lab_female.wav');
%[data, Fs] = audioread('studio_female.wav');
[y,fs]=audioread('studio_male.wav');
nguonchung = 0.08;
%# Threshold, > Voice, < Silence

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

%Chia khung tin hieu theo thoi gian
ThoiLuongKhung= 0.02; %s=20ms;
DoDaiKhung=ThoiLuongKhung * fs;
SoLuongKhung= floor(length(y)/DoDaiKhung); %So luong khung trong tin hieu y
Khung=zeros(SoLuongKhung,DoDaiKhung); %Tao ma tran cac khung co SoLuongKhung cot va DoDaiKhung hang
for k=1:SoLuongKhung
    Khung(k,:)=y(DoDaiKhung*(k-1)+1 :DoDaiKhung*k);%Chia tung khung
end

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

%tinh do lon trung binh cua tung khung
ma = MA(Khung);
%ghep cac do lon trung binh tung khung thanh tin hieu song
ma_wave = 0; %tin hieu song
for j = 1 : length(ma)

```

```

l = length(ma_wave);
ma_wave(l : l + DoDaiKhung) = ma(j);
end
t = linspace(0,length(y)/fs,length(y)); %tao vecto t gom length(y) diem cach deu nhau tu 0 den length(y)/fs
t1 = linspace(0,length(ma_wave)/fs,length(ma_wave));%tao vecto t1
subplot(3,1,1);
plot(t,y); %ve tin hieu vao
title('Signal');
subplot(3,1,2);
plot(t1,ma_wave); %ve nang luong cua tin hieu
title('Magnitude Average');
id = [];
%# Danh dau cac diem > nguong = 1, <nguong = 0
for sam = 1 : length(ma)
    if (ma(sam) > nguongchung)
        id = [id 1];
    else
        id = [id 0];
    end
end
%# Kiem tra dieu kien : khong ton tai khoang lang nho hon 200ms
for i = 1 : length(id)-10 % 200ms = 10 khung tin hieu
    if (id(i) == 1 && id(i+10)==1)
        id(i : i+10) = 1;
    end
end
%# Bien cac id hop le theo don vi thoi gian
id2 = [];
for i = 1 : length(id)-1
    if (id(i) + id(i+1)== 1)
        id2 = [id2 i*ThoiLuongKhung];
    end
end
%# Ke cac duong phan chia tin hieu

subplot(3,1,3);
plot(t,y,'g'); hold on;
plot([id2;id2], [-1 1], '--k');
title('Voice/ Silience Discriminate')

```

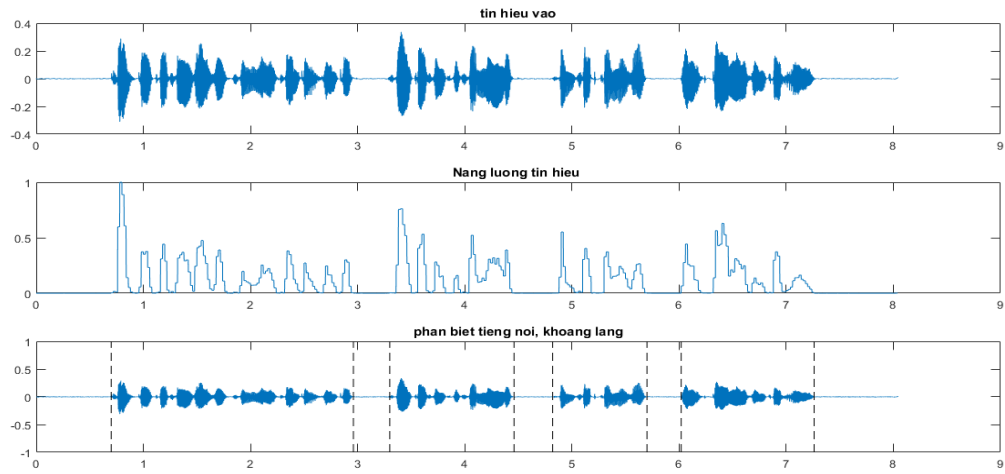
IV. KẾT QUẢ THỰC NGHIỆM

A. Kết quả định tính

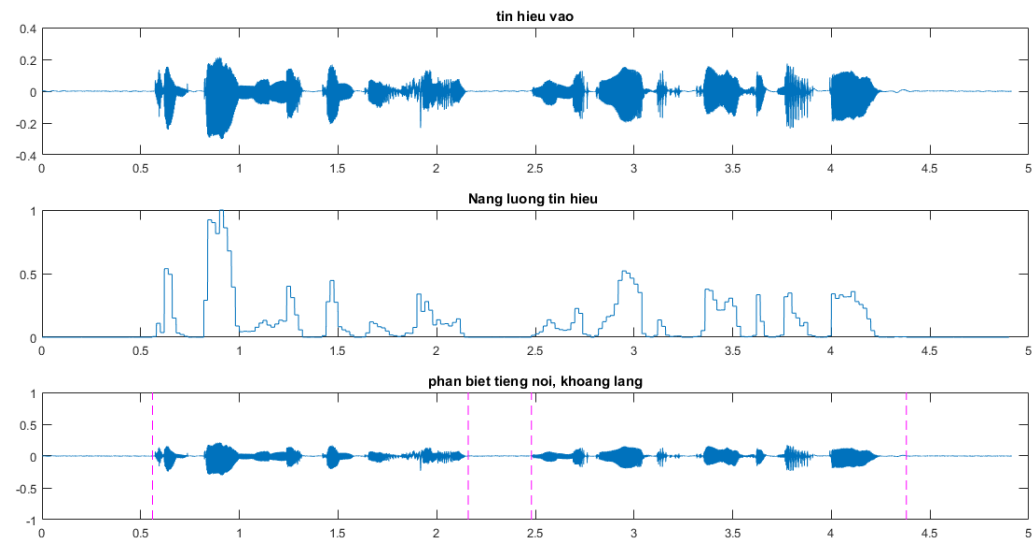
Dữ liệu dùng để đánh giá thuật toán là các 4 file tín hiệu lab_male.wav, lab_female.wav, phone_male.wav, phone_female.wav, studio_male, studio_male.wav.

1. Thuật toán Short – Time Energy

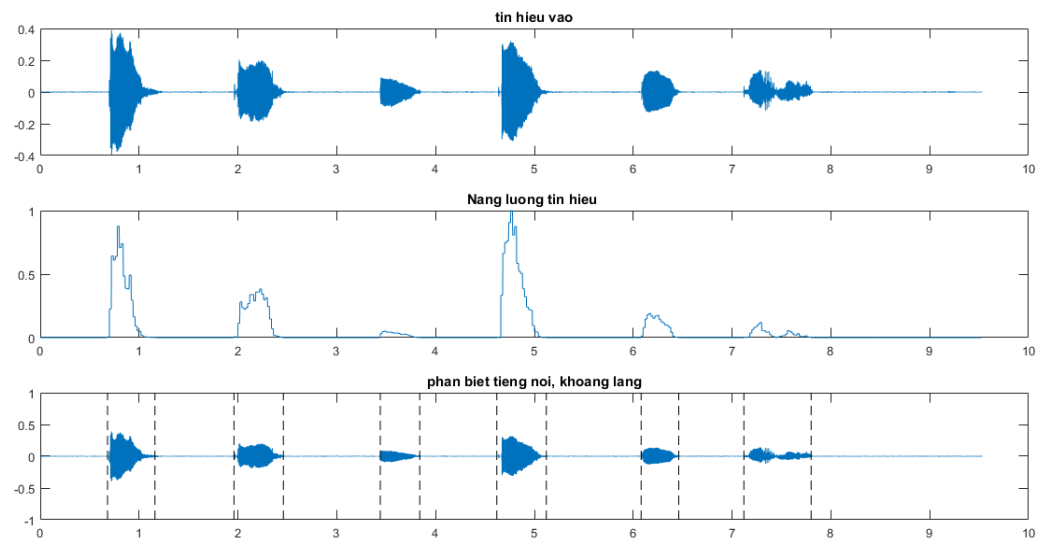
Ngưỡng = 0.001.



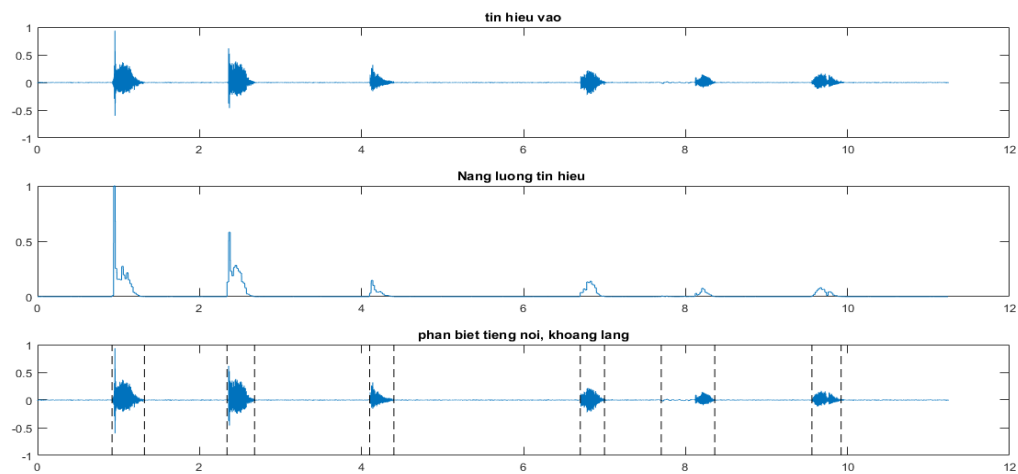
Hình 4. Studio_male



Hình 5. Studio_female



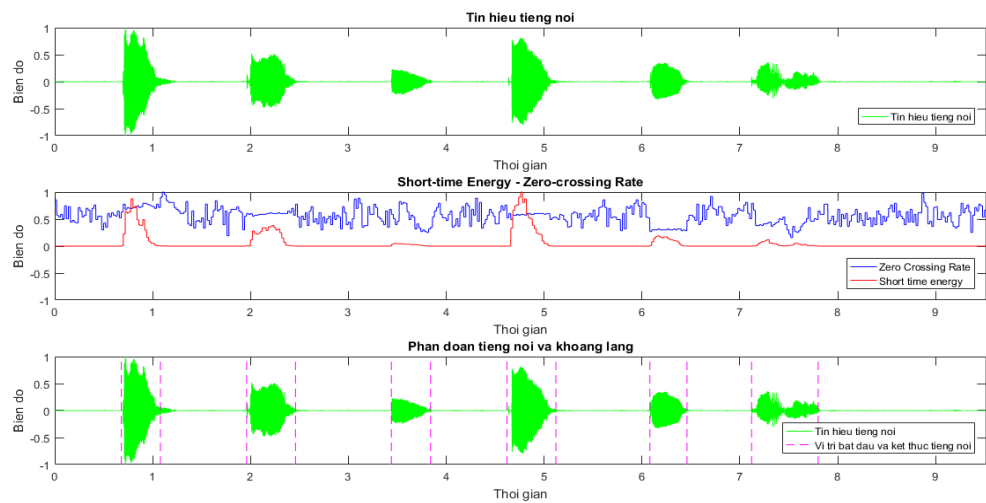
Hình 6. Lab_female



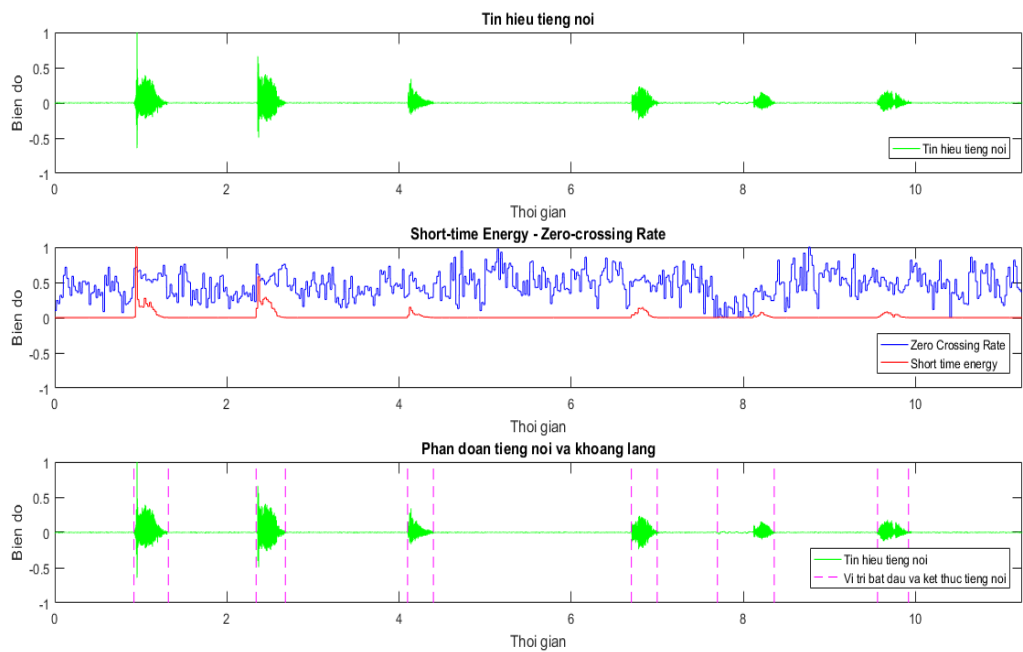
Hình 7. Lab_male

2. Thuật toán Zero Crossing Rate kết hợp với Short Time Energy

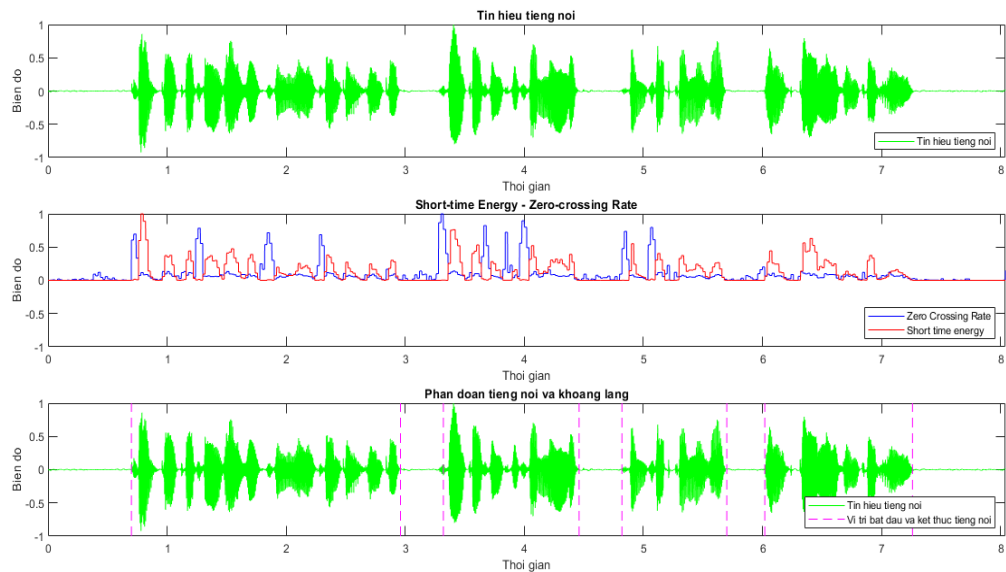
Với ngưỡng ZCRct = 0.8 và Ect = 0.001.



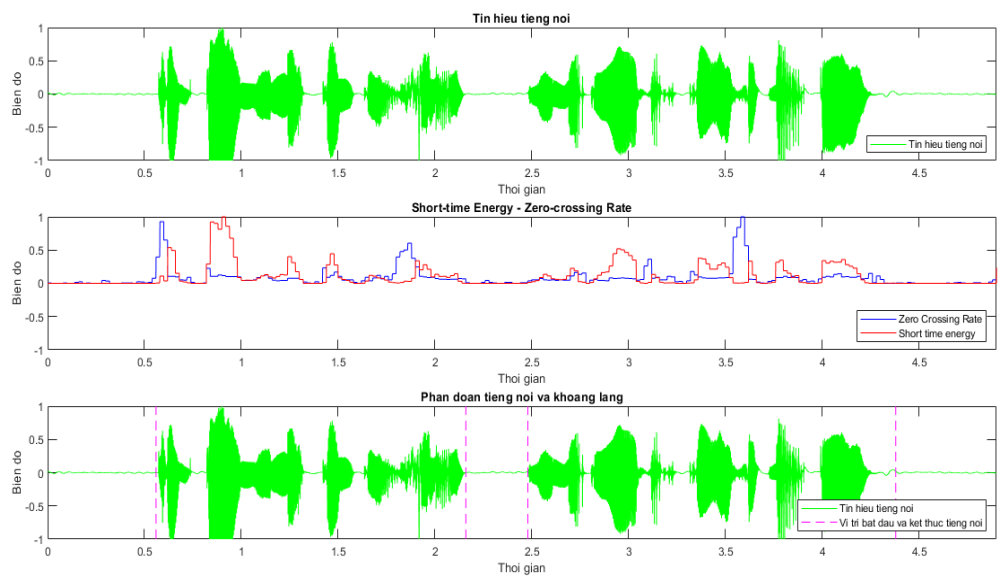
Hình 8. Lab_female



Hình 9. Lab_male

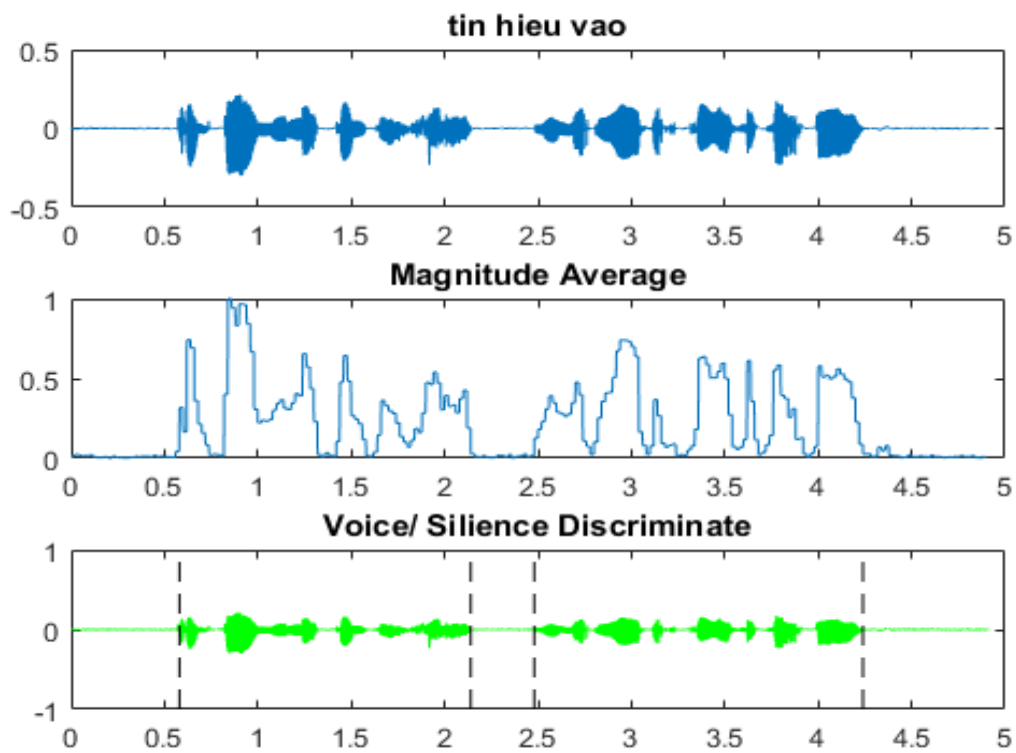


Hình 10. Studio_male

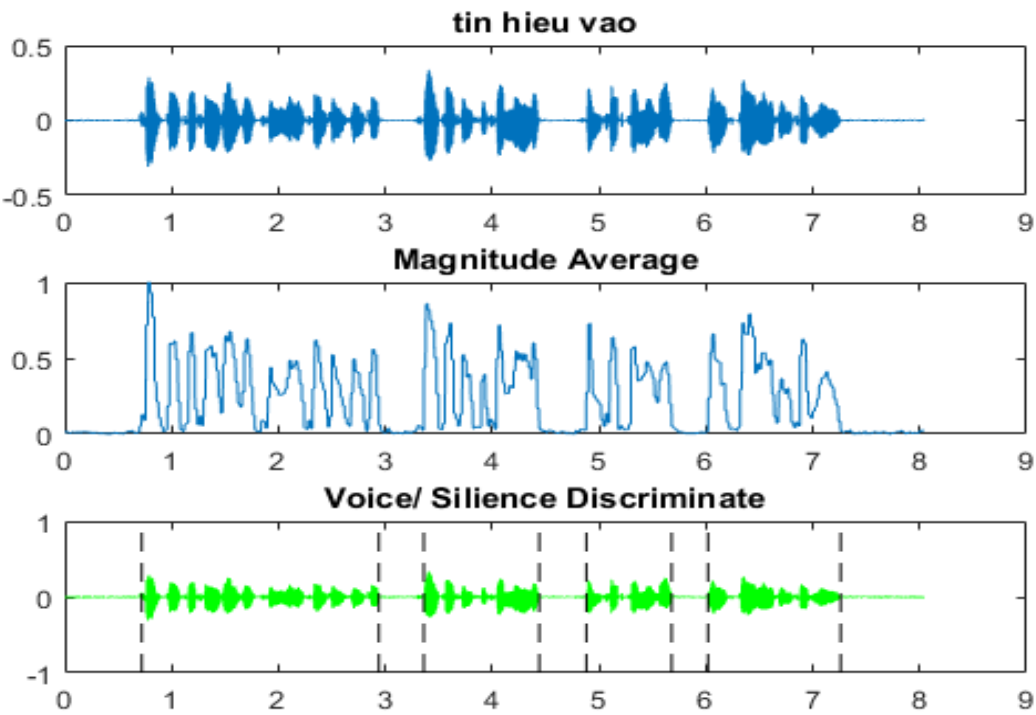


Hình 11. Studio_female

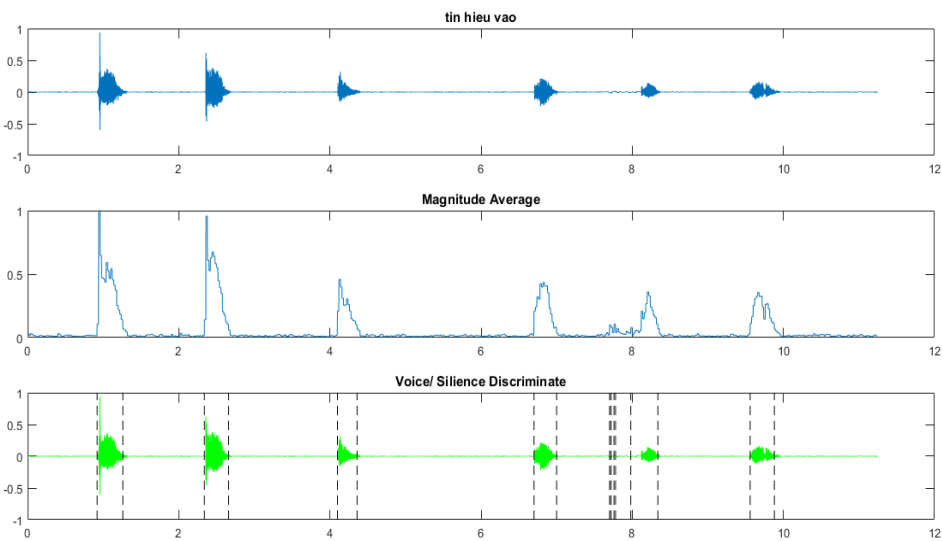
3. Thuật toán Magnitude Average



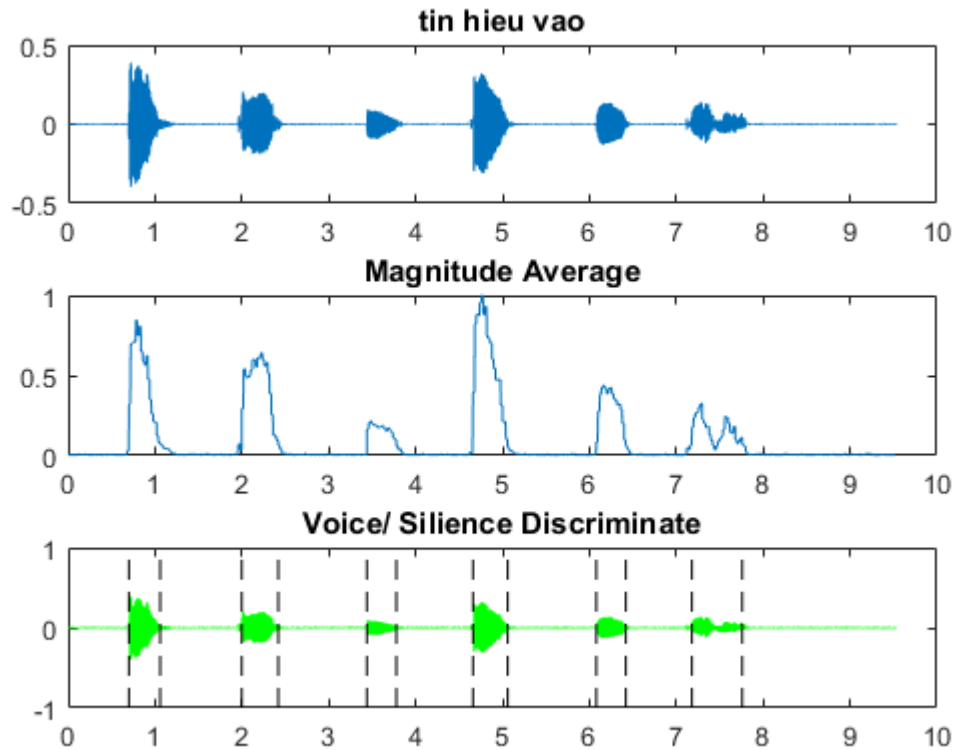
Hình 12. Studio_female



Hình 13. Studio_male



Hình 14. Lab_male



Hình 15. Lab_female

B. Kết quả định lượng

1. Thuật toán Magnitude Average

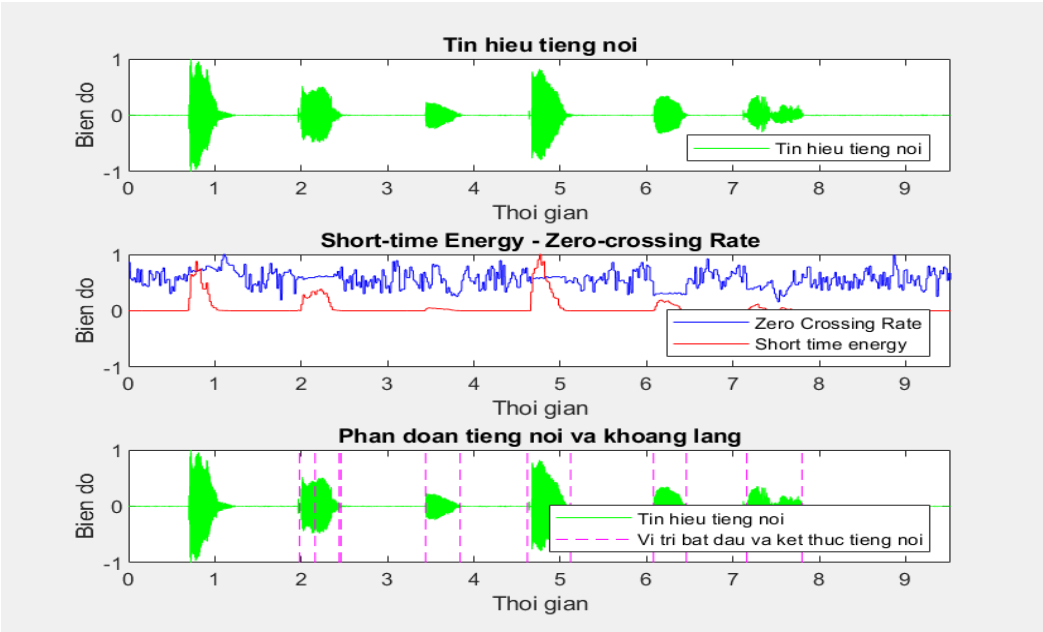
- Mặc dù những chi tiết và quá trình thực hiện bước đầu có vẻ khá phức tạp, phương pháp này là một ý tưởng rất đơn giản và dễ nhìn thấy: Phân đoạn thủ công một tín hiệu gồm cả tiếng nói và khoảng lặng, đo MA của cả 2; đặt một mức vào vùng chồng lên nhau; đo MA của bất cứ tín hiệu tiếng nói nào trong cùng một điều kiện; phân loại tất cả các chỉ số dưới mức thì là khoảng lặng.
- So sánh với thuật toán STE: Trong quá trình tính toán thì việc bình phương trong STE tạo ra những giá trị rất lớn, điều này khiến việc sử dụng MA phản ánh tốt hơn đường bao dạng sóng và được sử dụng nhiều hơn trong việc phân biệt dựa theo cấp độ (level-based discrimination)

2. Thuật toán Short Time Energy

- Tín hiệu đầu vào file lab_male ta thấy năng lượng tỉ lệ thuận với biên độ tiếng nói, mặt khác file được thu trong môi trường nhiễu lớn nên trong quá trình xử lý thuật toán có thể hiểu nhầm nhiều là tiếng nói.
- Tín hiệu đầu vào file lab_female ta thấy rằng file được thu trong môi trường nhiễu lớn nhưng do giọng nữ cao hơn giọng nam nên thu trong cùng một điều kiện file lab_female có kết quả chính xác hơn do có sự chênh lệch lớn về giá trị của năng lượng của tiếng nói và khoảng lặng.
- Tín hiệu đầu vào file studio_male được thu ở điều kiện lý tưởng, ít tồn tại nhiễu. Kết quả phân đoạn tiếng nói thu được có độ chính xác cao.
- Tín hiệu đầu vào file studio_female được thu ở điều kiện lý tưởng, ít tồn tại nhiễu. Kết quả phân đoạn tiếng nói thu được có độ chính xác cao.

3. Thuật toán kết hợp Short Time Energy và Zero Crossing Rate:

- Có nhiều yếu tố ảnh hưởng đến sự chính xác của thuật toán như:
 - + Tín hiệu đầu vào có nhiễu hay không có nhiễu
 - + Sự xác định các giá trị ngưỡng phù hợp cho từng trường hợp để mang lại các kết quả khác nhau :



Hình 16. Kết quả của lab_female khi ZCRct = 0.6 và Ect = 0.001

- Ta so sánh kết quả ở 2 hình 9 và 17 thì ta sẽ thấy với ZCRct = 0.8 thì sẽ cho ra kết quả chính xác hơn.

IV. KẾT LUẬN

- Tổng quan về bài báo cáo đã nêu được lý thuyết về đặc trưng của tín hiệu là năng lượng thời gian ngắn hạn, ZCR và magnitude average, tín hiệu tiếng nói, phân đoạn tín hiệu. Qua đó thấy được rằng thuật toán dùng STE cho kết quả chính xác hơn thuật toán STE kết hợp với ZCR và thuật toán Magnitude Average.
- Trong tương lai chúng tôi sẽ thử nghiệm với các thuật toán phân đoạn tiếng nói khác để tìm ra thuật toán tối ưu nhất. Bên cạnh đó sẽ ứng dụng các thuật toán này vào trong các chương trình phân tích, nhận dạng giọng nói.
- Môi trường thu âm là yếu tố ảnh hưởng nhiều nhất đến kết quả thu được. Muốn có độ chính xác cao cần thu âm trong môi trường yên tĩnh, không có nhiễu. Tiếng ồn và tiếng nhiễu,... đã ảnh hưởng đến độ chính xác của thuật toán đạt kết quả không cao.

V. TÀI LIỆU THAM KHẢO

[1] CS425 Audio and Speech Processing

[2] Link: https://en.wikipedia.org/wiki/Speech_segmentation

[3] Prentice Hall - Digital Processing Of Speech Signals_1978.pdf