PHÂN ĐOẠN TÍN HIỆU THÀNH TIẾNG NÓI VÀ KHOẢNG LẶNG DÙNG HÀM NĂNG LƯỢNG NGẮN HẠN (STE), HÀM TỶ LỆ VƯỢT QUA KHÔNG (ZCR) VÀ HÀM TRUNG BÌNH BIÊN ĐỘ (MA) CỦA TÍN HIỆU TRÊN MIỀN THỜI GIAN

Nguyễn Đình Mẫn, Trần Thế Nam, Võ Trần Anh Khoa

Nhóm 4, lớp HP: 18.89

Điểm	Bảng phân công nhiệm vụ		Chữ ký của SV
	Nguyễn Đình Mẫn (nhóm trưởng)	• Phân công nhiệm vụ, đảm bảo tiến độ của các thành viên trong nhóm. • Đọc tài liệu, viết báo cáo các vấn đề về tiếng nói và xử lý tiếng nói (tr.3) • Đọc tài liệu, cài đặt thuật toán (tr.9, tr.13-14), viết báo cáo trình bày về thuật toán phân đoạn tín hiệu sử dụng hàm năng lượng ngắn hạn (STE) kết hợp với hàm tỷ lệ vượt qua không (ZCR). (tr.5-7) • Viết báo cáo phần kết quả thực nghiệm, kết luận (tr.14-21) • Tổng hợp, viết báo cáo hoàn chỉnh. • Làm slide, thuyết trình phần đặt vấn đề, cơ sở lý thuyết tiếng nói và xử lý tiếng nói, kết luận • Thuyết trình cơ sở lý thuyết, sơ đồ khối, cài đặt, kết quả của hàm kết hợp giữa năng lượng ngắn hạn (STE) và tỷ lệ vượt qua không (ZCR)	
	Trần Thế Nam	• Đọc tài liệu, cài đặt, viết báo cáo về thuật toán loại bỏ khoảng lặng có độ dài bé hơn 200ms.(tr.11) • Phân đoạn tiếng nói thủ công(tr.9) • Đọc tài liệu, cài đặt thuật toán(tr9, tr.14), viết báo cáo trình bày về thuật toán phân đoạn tín hiệu sử dụng hàm trung bình biên độ (MA) (tr.7-8) • Thống kê sai số của các thuật toán (tr.19-20) • Làm slide, thuyết trình phần cơ sở lý thuyết, sơ đồ khối,cài đặt và kết quả của hàm trung bình biên độ (MA), trình bày kết quả của các thuật toán.	
	Võ Trần Anh Khoa	• Đọc tài liệu, cài đặt thuật toán(tr.9), viết báo cáo trình bày về thuật toán phân đoạn tín hiệu sử dụng hàm năng lượng ngắn hạn (STE) (tr.4-5) • Thuyết trình về cơ sở lý thuyết, thuật toán, sơ đồ khối của hàm năng lượng ngắn hạn(STE).	

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 liệt kê ở 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— Bài báo cáo thể hiện ba thuật toán: Năng lượng ngắn hạn (STE), Năng lượng ngắn hạn (STE) kết hợp tỷ lệ vượt qua không (ZCR), hàm trung bình biên độ (MA) để phân đoạn tiếng nói (Voiced) và khoảng lặng (Silence). Các thuật toán trên được sử dụng cho 4 mẫu tín hiệu

được cung cấp. Từ đó có cái nhìn tổng quan về tính hiệu quả trong việc thực hiện yêu cầu của báo cáo, rút ra được phương pháp tốt nhất.

Từ khóa— Short-Time Energy, Zero Crossing Rate, Magnitude Average, Short Time Processing, Energy of Speech Signal

Mục lục

I. ĐẶT VẤN ĐỀ

Tiếng nói là một phương tiện cơ bản và quan trọng trong giao tiếp của con người. Hiện nay, tiếng nói còn được áp dụng trong việc giao tiếp giữa con người với máy móc nhằm mục đính thay thế các phương tiện giao tiếp truyền thống như bàn phím, chuột, màn hình,.. vì vậy, xử lý tiếng nói đóng vai trò rất quan trọng trong vấn đề này.

Phân đoạn tín hiệu tiếng nói là quá trình xác định ranh giới giữa các từ, âm tiết trong các ngôn ngữ tự nhiên.[1] Ở báo cáo này, chúng tôi sẽ trình bày ba phương pháp phân đoạn tiếng nói và khoảng lặng : Năng lượng ngắn hạn (Short-time energy), Năng lượng ngắn hạn (Short-Time Energy) kết hợp Tỷ lệ vượt qua không (Zero Crossing Rate) và hàm Trung bình biên độ (Magnitude Average).

Báo cáo được chia thành 6 phần, có bố cục như sau:

Phân I: Đặt vấn đề

Phần II: Trình bày cơ sở lý xử lý tín hiệu tiếng nói và các thuật toán

Phần III: Trình bày mã chương trình cài đặt các thuật toán

Phần IV: Trình bày kết quả thực nghiệm, đánh giá độ chính xác của thuật toán, đưa ra các đánh giá định tính và định lượng giữa xác thuật toán đã áp dụng.

Phần V: Trình bày và kết luận, đề xuất các phương hướng phát triển và cải thiện trong tương lai.

Phần VI: Tài liệu tham khảo

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

A. Lý thuyết xử lý tiếng nói

1. Tổng quan về tiếng nói và xử lý tiếng nói

Tiếng nói là một phương tiện giao tiếp cơ bản của con người nhằm trao đổi thông tin bằng ngôn ngữ cũng như tình cảm của người nói.

Xử lý tiếng nói là sự nghiên cứu tiếng nói của con người dưới dạng tín hiệu, và các phương pháp xử lý những tín hiệu này. Tín hiệu tiếng nói thường được thể hiện dưới dạng số, tức là được "số hóa", và do đó, xử lý tiếng nói có thể được coi là giao của "xử lý tín hiệu số" và "xử lý ngôn ngữ tự nhiên".[2]

Trong phạm vi nguyên cứu, chúng tôi tập trung việc xác định các biên của tiếng nói và khoảng lặng. Phương pháp cơ bản trong phân tích tín hiệu tiếng nói là phân tích thời gian ngắn hạn (phân đoạn tín hiệu). Hầu hết các tín hiệu âm thanh sẽ ổn định và biến đổi chậm trong những khoảng thời gian ngắn khoảng 10-30ms . Do đó, người ta thường chia tiếng nói thành nhiều đoạn có thời gian bằng nhau được gọi là khung (frame), mỗi khung có mỗi khung có độ dài từ 10 đến 30 ms.

Để phát hiện tiếng nói, phải xác định được điểm bắt đầu và điểm kết thúc của tiếng nói. Ở đây chúng tôi nêu ra ba phương pháp dựa trên ba đặc trưng năng lượng ngắn hạn, tỷ lệ vượt quá điểm không kết hợp năng lượng ngắn hạn và trung bình biên độ.

2. Các bước để xử lý tiếng nói

Bước 1: Phân tích tín hiệu thành các khung có độ dài 20ms
- Thành lập hàm phân khung tín hiệu với tham số đầu vào là tín hiệu cần phân tích, tần số lấy mẫu và độ dài khung tín hiệu (20ms)
- Xác định số mẫu trên một khung hình, và tổng số khung hình.

Bước 2: Tính năng lượng ngắn hạn của tín hiệu, tính tỷ lệ vượt qua không kết hợp năng lượng ngắn hạn và tính trung bình biên độ. Vẽ đồ thị của ba phương pháp trên.
- Thiết lập thuật toán năng lượng ngắn hạn (STE) để tính năng lượng mỗi khung
- Lưu dữ liệu của năng lượng ngắn hạn của mỗi khung vào một ma trận
- Thiết lập thuật toán tỷ lệ vượt qua không (ZCR) kết hợp hàm năng lượng ngắn hạn (STE)
- Lưu dữ liệu sau khi thực hiện thuật toán vào một ma trận
- Thiết lập thuật toán xác định trung bình biên độ(MA)
- Lưu dữ liệu sau khi thực hiện thuật toán MA vào một ma trận.
- Vẽ đồ thị của ba phương pháp đã nêu trên.

Bước 3: Xác định ngưỡng năng lượng chuẩn hóa để phân đoạn chính xác tiếng nói và khoảng lặng.
- Thiết lập thuật toán tìm ngưỡng cho mỗi phương pháp đã nêu trên.
- Xác định biên của tiếng nói và khoảng lặng dựa trên ngưỡng đã tìm ra.

Bước 4: Tìm biên chuẩn của các đoạn tín hiệu tiếng nói, khoảng lặng. Vẽ đồ thị thể hiện biên của tín hiệu
- Quan sát hình ảnh biểu diễn của tín hiệu, xác định thủ công biên của tiếng nói và khoảng lặng
- Thiết lập hàm tự động xác định biên của tiếng nói, khoảng lặng
- Vẽ đồ thị biên của tín hiệu để biểu diễn.

B. Các thuật toán

1. Thuật toán Short-Time Energy

a. Cở sở lý thuyết

Tính năng lượng của một tín hiệu là việc thường được sử dụng trong xử lý tín hiệu. Năng lượng của tín hiệu là một đặc trưng trên miền thời gian của tín hiệu rời rạc và được định nghĩa [2].

$$E(n) = \sum_{n=-\infty}^{\infty} (x[n-m])^2 \ [3]$$

Do tính chất của hầu hết tín hiệu âm thanh là ổn định biến đổi chậm trong khoảng thời gian ngắn, thông thường, ta thường sử dụng phương pháp phân tích ngắn hạn. Mặt khác năng lượng của tín hiệu là một trong những đặc trưng của tín hiệu âm thanh, vì vậy ta có thể áp dụng phương pháp phân tích ngắn hạn để xử lý tín hiệu này. Khi thực hiện, bằng ta chia tín hiệu thành nhiều khung có khoảng thời gian bằng nhau từ 10-30ms, thì công thức trên có thể viết lại thành:

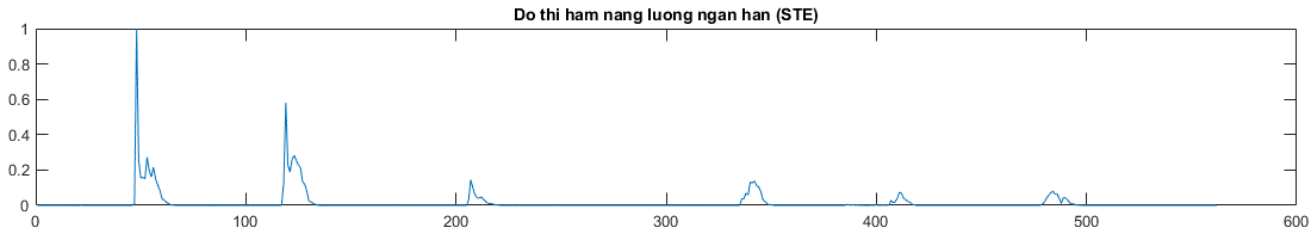
$$E_m(n) = \sum_{n=N_{1m}}^{N_{2m}} (x[n-m])^2$$

Trong đó: m là chỉ số khung thứ m và n ∈ $[N_{1m}; N_{2m}]$

N_1 và N_2 là chỉ số mẫu bắt đầu và kết thúc của khung thứ m.

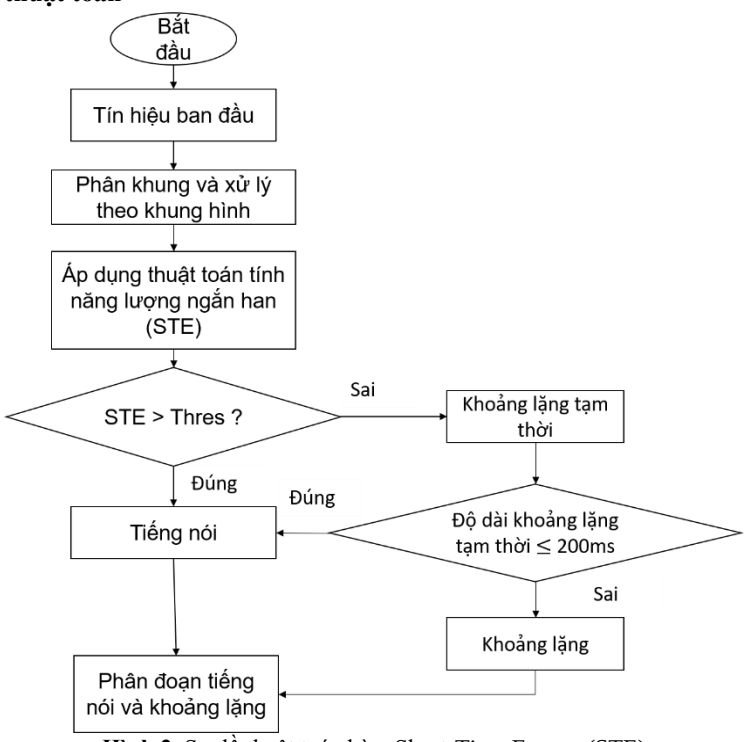
x[n] Biên độ tín hiệu tại ví trí thứ n

Sau đó chuẩn hóa về ngưỡng [0;1]: $E_{norm} = \frac{E - E_{min}}{E_{max}}$



Hình 1. Đồ thị hàm Short-Time Energy (STE)

b. Sơ đồ khối thuật toán



Hình 2. Sơ đồ thuật toán hàm Short-Time Energy (STE)

c. **Các tham số quan trọng trong thuật toán**

- Tần số lấy mẫu tín hiệu đầu vào
- Độ dài mỗi khung hình khoảng 20ms, do âm thanh sẽ ổn định và ít biến đổi trong khoảng thời gian ngắn.
- Độ lớn năng lượng của mỗi khung giúp nhận biết khung tín hiệu nào là tiếng nói hay khoảng lặng
- Ngưỡng lấy mẫu (Thres) giúp phân loại được tiếng nói và khoảng lặng, đây là tham số quan trọng nhất của thuật toán.

d. **Vấn đề và giải pháp khắc phục**

- Vấn đề: Do năng lượng của tín hiệu vào là khác nhau, và có ảnh hưởng khác như nhiễu tiếng ồn, tạp âm nên ngưỡng xác định có thể bị thay đổi.
- Giải pháp: Cần cải tiến thuật toán tìm ngưỡng để giảm sự ảnh hưởng của tạp âm.

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

a. Cơ sở lý thuyết

a.1 Thuật toán Short-Time Energy

Tính năng lượng của một tín hiệu là việc thường được sử dụng trong xử lý tín hiệu. Năng lượng của tín hiệu là một đặc trưng trên miền thời gian của tín hiệu rời rạc và được định nghĩa .

$$E[n] = \sum_{n=-\infty}^{\infty} (x[n-m])^2 \qquad [3]$$

Do tính chất của hầu hết tín hiệu âm thanh là ổn định biến đổi chậm trong khoảng thời gian ngắn, vì vậy ta thường sử dụng phương pháp phân tích ngắn hạn. Mặt khác năng lượng của tín hiệu là một trong những đặc trưng của tín hiệu âm thanh, vì vậy ta có thể áp dụng phương pháp phân tích ngắn hạn để xử lý tín hiệu này. Khi thực hiện, bằng ta chia tín hiệu thành nhiều khung có khoảng thời gian bằng nhau từ 10-30ms, thì công thức trên có thể viết lại thành

$$E_m[n] = \sum_{n=N_{1m}}^{N_{2m}} (x[n-m])^2$$

Trong đó: m là chỉ số khung thứ m và n ∈ $[N_{1m}; N_{2m}]$

N_1 và N_2 là chỉ số mẫu bắt đầu và kết thúc của khung thứ m.

x[n] Biên độ tín hiệu tại ví trí thứ n

Sau đó chuẩn hóa về ngưỡng [0;1]: $E_{norm} = \frac{E - E_{min}}{E_{max}}$

a.2 Thuật toán Zero Crossing rate

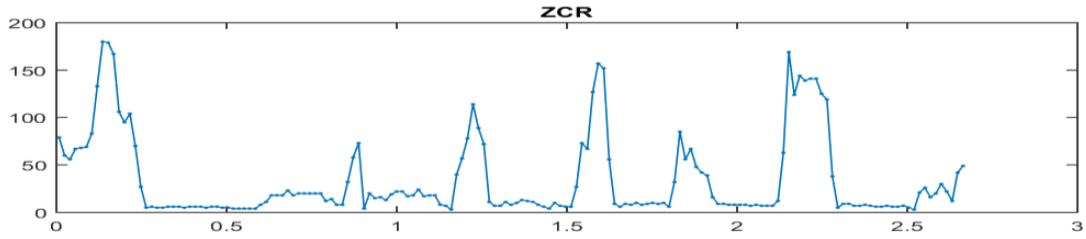
Tỷ lệ băng qua không (ZCR) là một trong những đặc trưng trên miền thời gian của tín hiệu rời rạc. Zero crossing rate là một thông số cho biết số lần mà biên độ băng qua điểm không trong một thời gian nhất định.

Theo đặc trưng của tín hiệu, ở những đoạn tín hiệu có zero crossing rate cao thì ở đó tín hiệu là khoảng lặng, ngược lại ở những vị trí có zero crossing rate thấp thì ở đó tín hiệu là có tiếng nói. Vì tín hiệu tiếng nói có tính chất biến đổi chậm theo thời gian, có nghĩa là đặc trưng của tín hiệu không thay đổi trong thời gian ngắn từ 10-30ms. Áp dụng vào xử lý tiếng nói ta sử dụng phương pháp ngắn hạn tức là chia tín hiệu ban đầu thành những khung bằng nhau, mỗi khung từ 10-30 ms.

Vì vậy ta có công thức hàm ZCR:

$$ZCR[n] = \sum_{m=0}^{N-1} |sgn[x(n-m)] - sgn[x(n-m-1)| \quad [3]$$

Với: $$sgn(x) = \begin{cases} 1 \text{ nếu } x > 0 \\ -1 \text{ nếu } x < 0 \end{cases} \qquad [3]$$



Hình 3. Đồ thị Thuật toán Zero Crossing Rate

a.3 Kết hợp thuật toán Short-Time Energy và Zero Crossing Rate

Từ những kết quả của hàm Short-Time Enery, ta suy ra được đại lượng:

$$P(n) = \frac{E(n)}{N}$$

Từ hàm Zero Crossing Rate, ta suy ra được đại lượng:

$$Z(n) = \frac{ZCR(n)}{N}$$

Trong đó: E(n) là giá trị năng lượng của từng khung
N là số tín mẫu trên mỗi khung
ZCR(n) là giá trị của hàm ZCR trên mỗi khung

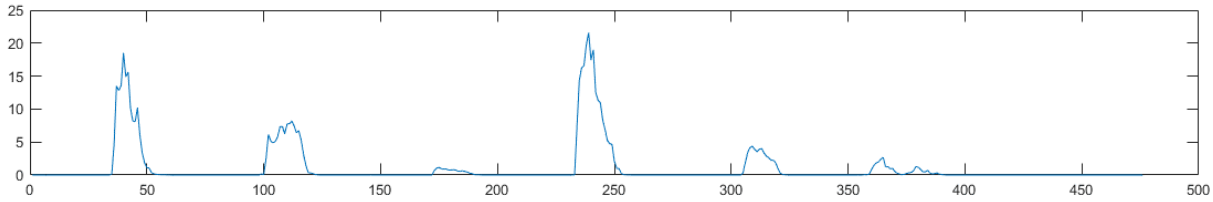
Từ các đại lượng trên, tính cho mỗi khung tín hiệu có độ dài trong khoảng từ 10-30ms. Các giá trị năng lượng ngắn hạn E sẽ lớn trong khoảng tín hiệu tiếng nói và nhỏ trong khoảng tín hiệu lặng. Trong khi đó giá trị của Z hay tần suất vượt điểm không đánh giá số lần chuyển đổi của tín hiệu qua giá trị 0 có xu hướng lớn trong khoảng tín hiệu lặng. Với giá thiết rằng 10 khung tín hiệu đầu tiên là khoảng lặng, giải thuật loại bỏ khoảng lặng được tiến hành như sau.[5]

$$W(n) = P(n) * \big(1 - Z(n)\big) * S_c \qquad [4]$$

Với giá trị của S_c vào khoảng 1000.

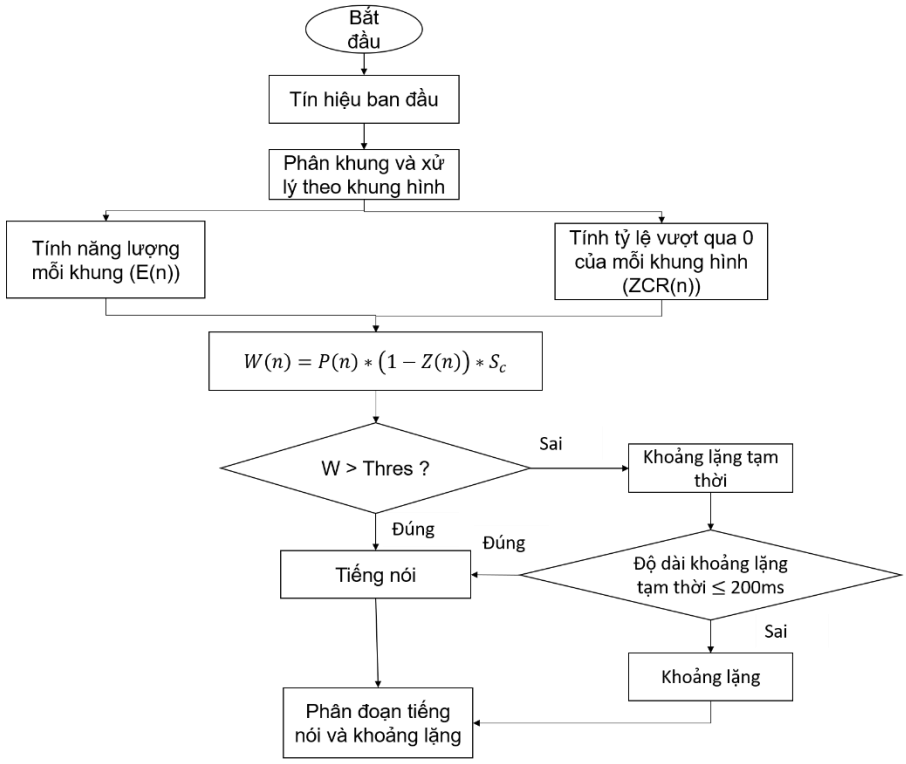
Ngưỡng kích hoạt cho hàm W được tính bởi công thức: Thres = K *(μ + αδ) [4]

Trong đó: μ là giá trị trung bình của tín hiệu
δ là giá trị phương sai của W cho 10 tín hiệu đầu
α là một hằng số được định nghĩa với $\alpha = 0.2\delta^{0.8}$
K là hằng số có thể thay đổi tùy thuộc vào thuật toán đang áp dụng.



Hình 4. Đồ thị thuật toán kết hợp giữa Short-Time Energy và Zero Crossing Rate

b. Sơ đồ khối thuật toán



Hình 5. Sơ đồ thuật toán kết hợp giữa Short-Time Energy và Zero Crossing Rate

c. Các tham số quan trọng trong thuật toán

- Tần số lấy mẫu tín hiệu đầu vào
- Độ dài mỗi khung hình khoảng 20ms, do âm thanh sẽ ổn định và ít biến đổi trong khoảng thời gian ngắn.
- Độ lớn năng lượng của mỗi khung giúp nhận biết khung tín hiệu nào là tiếng nói hay khoảng lặng
- Tỷ lệ băng qua không của mỗi khung giúp nhận biết khung tín hiệu nào là tiếng nói hay khoảng lặng
- Ngưỡng lấy mẫu (Thres), giúp phân loại được tiếng nói và khoảng lặng, đây là tham số quan trọng nhất của thuật toán.

d. Vấn đề và giải pháp khắc phục

- Vấn đề: Do năng lượng của tín hiệu vào là khác nhau, và có ảnh hưởng khác như nhiễu tiếng ồn, tạp âm nên ngưỡng xác định có thể bị thay đổi.
- Giải pháp: Cần cải tiến thuật toán tìm ngưỡng để giảm đi sự ảnh hưởng của tạp âm.

3. Thuật toán Magnitude Average

a. Cơ sở lý thuyết

Hàm trung bình biên độ dùng đặc trưng trên miền thời gian của tín hiệu âm thanh là cường độ để xử lý chính tín hiệu đó. Hàm Magnitude Aeverage có công thức như sau:

$$MA[n] = \sum_{m=0}^{N-1} |x[n-m]| \qquad [3]$$

Do tính chất của hầu hết tín hiệu âm thanh là ổn định biến đổi chậm trong khoảng thời gian ngắn, vì vậy ta thường sử dụng phương pháp phân tích ngắn hạn. Khi thực hiện, bằng cách ta chia tín hiệu thành nhiều khung có khoảng thời gian bằng nhau từ 10-30ms, thì công thức trên có thể viết lại thành.

$$MA[n] = \sum_{n=N_{1m}}^{N_{2m}} |x[n-m]|$$

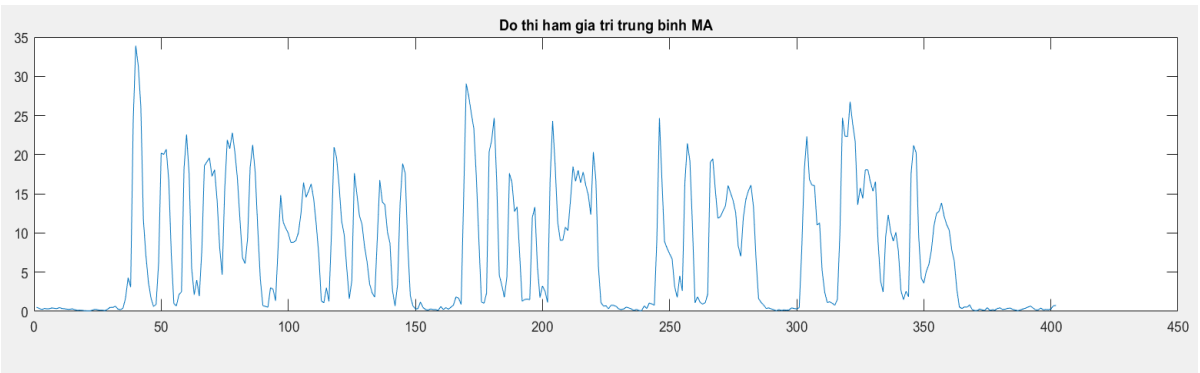
Trong đó:

m là chỉ số khung thứ m và n ∈ $[N_{1m}; N_{2m}]$

N_1 và N_2 là chỉ số mẫu bắt đầu và kết thúc của khung thứ m.

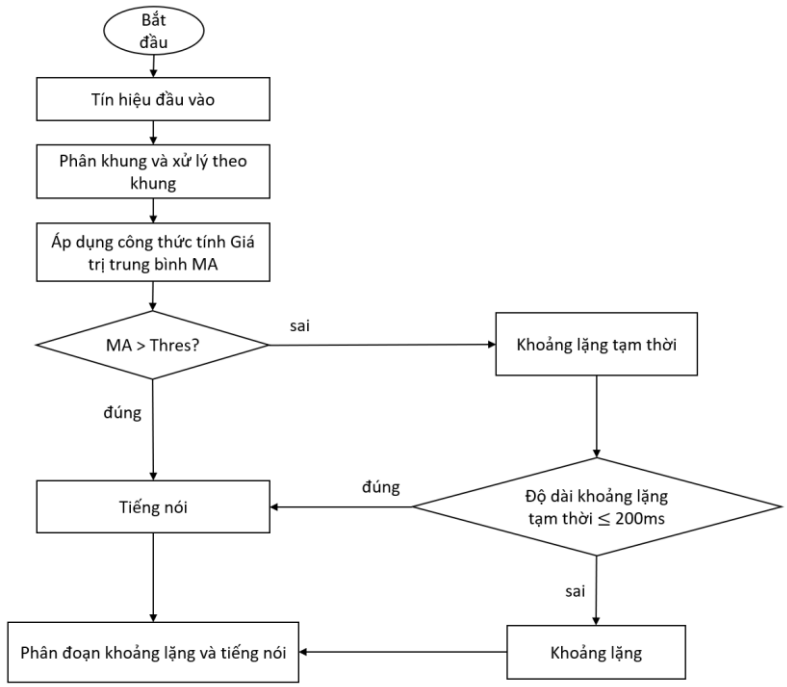
x[n] Biên độ tín hiệu tại ví trí thứ n

Sau đó chuẩn hóa về ngưỡng [0;1]: $MA_{norm} = \frac{MA - MA_{min}}{MA_{max}}$



Hình 6. Đồ thị hàm Magnitude Average

b. Sơ đồ khối thuật toán



Hình 7. Đồ thị hàm Magnitude Average

c. Các tham số quan trọng trong thuật toán

- Tần số lấy mẫu tín hiệu đầu vào
- Độ dài mỗi khung hình
- Độ lớn trung bình biên độ của mỗi khung giúp nhận biết khung tín hiệu nào là tiếng nói hay khoảng lặng
- Ngưỡng lấy mẫu (Thres), giúp phân loại được tiếng nói và khoảng lặng, đây là tham số quan trọng nhất của thuật toán.

d. Vấn đề và giải pháp khắc phục

- Vấn đề: Thuật toán xử lý tín hiệu dùng hàm MA gặp rắc rối với những tín hiệu đầu vào có chứa những âm thanh nhiễu có biên độ lớn.
- Nguyên nhân: Do hàm MA có thuật toán khá đơn giản, nên có độ chính xác không cao

III. MÃ CHƯƠNG TRÌNH CÀI ĐẶT CÁC THUẬT TOÁN

```
%====================== PHAN CHUONG TRINH CHINH ============================

% Doc tin hieu tu file audio
% x: mang chua cac gia tri bien do
% Fs: Tan so lay mau
% [x,Fs] = audioread('lab_female.wav');
% [x,Fs] = audioread('lab_male.wav');
% [x,Fs] = audioread('studio_male.wav');
% [x,Fs] = audioread('studio_female.wav');
% [x,Fs] = audioread('LA001.wav');
% [x,Fs] = audioread('LA025.wav');
fr_time = 0.02;          %fr_time(s): Do dai moi khung hinh
fr_len = fr_time*Fs;     %fr_len: So mau tren moi khung hinh
N = length(x);           %N: tong so mau cua tin hieu
% Ham tinh tong so khung hinh(frame)
fr_num = NumberOfFrames(N,fr_len);
% Ham tinh nang luong cua moi khung hinh
Power = SumPower(x, fr_num, fr_len);
min_P = min(Power);
max_P = max(Power);
```

```
%------------------------ Chuong trinh thuat toan ham STE --------------------

% Ham chuan hoa ve nguong [0;1]
Standard = Standard_function(Power, fr_num, min_P, max_P);
% Ham xac dinh nguong thuat toan STE
Thres_STE = Ethres_function(Power,10);
% Ham phat hien tieng noi va khoang lang
Voiced_STE = DetecVoiced_function(Standard, fr_num, Thres_STE);
% Ham xac dinh vi tri bien dau-cuoi cua tieng noi
Position_STE = Position_function(Voiced_STE, fr_num, fr_len);
% Ham truoc khi xu ly nhung doan khoang lang < 200ms
Pre_Voiced_STE = Pre_DetecVoiced_function(Power, fr_num, Thres_STE);
% Ham xac dinh vi tri bien dau-cuoi cua tieng noi
Pre_Position_STE = Position_function(Pre_Voiced_STE, fr_num, fr_len);

%--------------------Chuong trinh thuat toan ham ZCR + STE -------------------

% Chuan hoa lai tin hieu de tim ham zcr
X_ZCR = X_ZCR_function(x);
% Dai luong can thiet de tinh ham zcr (-1,1)
sgn = sgn_function(X_ZCR, fr_num, fr_len);
% Ham zcr
ZCR = ZCR_function(sgn, fr_num, fr_len);
% Ham thuat toan ket hop giua ZCR va STE
W = W_function(Power, ZCR, fr_len, fr_num);
% Ham xac dinh nguong thuat toan ZCR ket hop STE
Thres_ZS = Ethres_function(W,30);
% Ham phat hien tieng noi va khoang lang
Voiced_ZS = DetecVoiced_function(W, fr_num, Thres_ZS);
% Ham xac dinh vi tri bien dau-cuoi cua tieng noi
Position_ZS = Position_function(Voiced_ZS, fr_num, fr_len);
% Ham truoc khi xu ly nhung doan khoang lang < 200ms
Pre_Voiced_ZS = Pre_DetecVoiced_function(W, fr_num, Thres_ZS);
% Ham xac dinh vi tri bien dau-cuoi cua tieng noi
Pre_Position_ZS = Position_function(Pre_Voiced_ZS, fr_num, fr_len);

%------------------------ Chuong trinh thuat toan MA ------------------------

MA = MA_function(x, fr_num, fr_len);
min_MA = min(MA);
max_MA = max(MA);
%Ham chuan hoa ve nguong [0;1]
Standard_MA = Standard_function(MA, fr_num, min_MA, max_MA);
%Ham xac dinh nguong thuat toan MA
Thres_MA = Ethres_function(Standard_MA,10);
%Ham phat hien tieng noi va khoang lang
Voiced_MA = DetecVoiced_function(Standard_MA, fr_num, Thres_MA);
%Ham xac dinh vi tri bien dau - cuoi cua tieng noi
Position_MA = Position_function(Voiced_MA, fr_num, fr_len);
% Ham truoc khi xu ly nhung doan khoang lang < 200ms
Pre_Voiced_MA = Pre_DetecVoiced_function(Standard_MA, fr_num, Thres_MA);
% Ham xac dinh vi tri bien dau-cuoi cua tieng noi
Pre_Position_MA = Position_function(Pre_Voiced_MA, fr_num, fr_len);
%------------------------------------------------------------------------------

% Bien tieng noi- khoang lang khi xac dinh thu cong cua 4 file
Lab_female = [11066, 19634, 31334, 39429, 55010, 61614, 74171, 82260,  97346,
103315, 113887, 125102];
Lab_male = [14519, 21359, 37435, 43339, 65582, 70570, 107175, 112666, 128879,
134030, 152803, 159493];
studio_male = [11201, 47415, 52431, 71211, 76935, 91042, 96296, 116301];
studio_female = [8996, 34395, 39570, 68135];
```

```
subplot(411), plot(x);
title('Do thi phan doan thu cong file Lab_female.wav');
hold on;
for i =1:length(Lab_female)/2
    xline(Lab_female(2*i-1), 'r','LineWidth',1);
    xline(Lab_female(2*i), 'g','LineWidth',1);
end
hold off;

subplot(412), plot(Standard);
title('Do thi ham nang luong ngan han (STE)');
hold on;
for i =1:length(Pre_Position_STE)/2
    xline(Pre_Position_STE(2*i-1)/fr_len, 'r','LineWidth',1);
    xline(Pre_Position_STE(2*i)/fr_len, 'g','LineWidth',1);
end
yline(Thres_STE,'-r','Thres');
hold off;

subplot(413), plot(x);
title('Phan doan tieng noi va khoang lang (STE)');
hold on;
for i =1:length(Position_STE)/2
    xline(Position_STE(2*i-1), 'r','LineWidth',1);
    xline(Position_STE(2*i), 'g','LineWidth',1);
end
hold off;
legend('Tin hieu mau','Vi tri bat dau co tieng noi','Vi tri ket thuc tieng
noi','location','northeast');
figure;
subplot(411);
plot(W);
title('Do thi ham W = P*(1-ZCR)*1000 (ZCR + STE)');
hold on;
for i =1:length(Pre_Position_ZS)/2
    xline(Pre_Position_ZS(2*i-1)/fr_len, 'r','LineWidth',1);
    xline(Pre_Position_ZS(2*i)/fr_len, 'g','LineWidth',1);
end
yline(Thres_ZS,'-r','Thres');
hold off;
subplot(412), plot(x);
title('Phan doan tieng noi va khoang lang (ZCR + STE)');
hold on;
for i =1:length(Position_ZS)/2
    xline(Position_ZS(2*i-1), 'r','LineWidth',1);
    xline(Position_ZS(2*i), 'g','LineWidth',1);
end
hold off;
legend('Tin hieu mau','Vi tri bat dau co tieng noi','Vi tri ket thuc
tiengnoi','location','northeast');
subplot(413);
plot(W);
title('Do thi ham trung binh co nghia (MA)');
hold on;
for i =1:length(Pre_Position_MA)/2
    xline(Pre_Position_MA(2*i-1)/fr_len, 'r','LineWidth',1);
    xline(Pre_Position_MA(2*i)/fr_len, 'g','LineWidth',1);
end
yline(Thres_ZS,'-r','Thres');
hold off;
subplot(414), plot(x);
title('Phan doan tieng noi va khoang lang (MA)');
hold on;
```

```
for i =1:length(Position_MA)/2
    xline(Position_MA(2*i-1), 'r','LineWidth',1);
    xline(Position_MA(2*i), 'g','LineWidth',1);
end
hold off;
legend('Tin hieu mau','Vi tri bat dau co tieng noi','Vi tri ket thuc
tiengnoi','location','northeast');

% % Xac dinh sai so giua xac dinh thu cong va thuat toan tren file lab_female
% RMSE_STE_lab_female = sqrt(mean((Position_STE - Lab_female).^2))
% RMSE_MA_lab_female = sqrt(mean((Position_MA - Lab_female).^2))
% RMSE_ZS_lab_female = sqrt(mean((Position_ZS - Lab_female).^2))

% % Xac dinh sai so giua xac dinh thu cong va thuat toan tren file lab_male
% RMSE_STE_lab_male = sqrt(mean((Position_STE - Lab_male).^2))
% RMSE_MA_lab_male = sqrt(mean((Position_MA - Lab_male).^2))
% RMSE_ZS_lab_male = sqrt(mean((Position_ZS - Lab_male).^2))

% % Xac dinh sai so giua xac dinh thu cong va thuat toan tren file studio_male
% RMSE_STE_studio_male = sqrt(mean((Position_STE - studio_male).^2))
% RMSE_MA_studio_male = sqrt(mean((Position_MA - studio_male).^2))
% RMSE_ZS_studio_male = sqrt(mean((Position_ZS - studio_male).^2))

% % Xac dinh sai so giua xac dinh thu cong va thuat toan tren file studio_male
% RMSE_STE_studio_female = sqrt(mean((Position_STE - studio_female).^2))
% RMSE_MA_studio_female = sqrt(mean((Position_MA - studio_female).^2))
% RMSE_ZS_studio_female = sqrt(mean((Position_ZS - studio_female).^2))

%============================PHAN DINH NGHIA HAM============================
% Ham tinh so khung hinh(frame) cua tin hieu
% Ham tra ve tong so frame cua tin hieu
% N: Tong so mau cua tin hieu
% fr_len: so mau cua tin hieu tren mot frame
% su dung ham floor de lay nguyen phep chia
function fr_num = NumberOfFrames(N,fr_len)
    fr_num = floor(N/fr_len);
end

% Ham tinh nang luong tren moi khung
% Ham tra ve: mang nang luong cua moi khung hinh
% x: mang gia tri bien do cua tin hieu am thanh
% fr_num: tong so frame cua tin hieu
% fr_len: so mau tren mot frame
function Power = SumPower(x, fr_num, fr_len)
    Power = zeros(1,fr_num);                    % Tao mang chua Power
    for k = 1:fr_num                            % Duyet tat ca cac khung
        tempPower = 0;
        for j =(k-1)*fr_len +1 : (fr_len*k -1)  % duyet cac mau co trong khung
            tempPower = tempPower + abs(x(j)^2);% tinh nang luong cua tung
                                                  khung
        end
        Power(k) = tempPower;  % Luu vao mang Power chua cac gia tri nang luong
    end
end

% Ham chuan hoa vector Power ve [0;1]
% fr_num: tong so frame cua tin hieu
% min, max la gia tri min va max cua vector Power
function Standard = Standard_function(Power, fr_num, min, max)
    Standard = zeros(1,fr_num);     % Tao mang chua cac gia tri sau khi chuan
                                      hoa ve nguong 0,1
    for k = 1:fr_num                % Duyet tat ca cac khung
```

```
        %cong thuc chuan hoa
        temp = (Power(k)-min)/(max-min);    % Chuan hoa bien do ve 0-1
        Standard(k) = temp;          % luu du lieu vao mang Standard sau khi
                                       chuan hoa
    end
end

% Ham dua tin hieu ve [0;1]
% Ham tra ve mot vector chua 0,1
% Voi 0 duoc hieu la khoang lang
%     1 duoc hieu la tieng noi
% Ham tu xac dinh khoang lang, neu phat hien khoang lang
% co thoi gian < 200ms thi bien doi thanh tieng noi
function [Voiced] = DetecVoiced_function(Standard, fr_num, Thres)
    Voiced = zeros(1,fr_num);
    for k = 1:fr_num                % Duyet het tat cac khung
        if (Standard(k) > Thres)    % So sanh voi nguong phan tach tieng noi
                                       khoang lang
            Voiced(k) = 1;          % Gan Voiced(i) = 1 neu khung lon hon gia
                                       tri nguong
        end
    end
    for k = 1:fr_num                % Duyet het tat cac khung
        if (Standard(k)< Thres)     % So sanh voi nguong phan tach tieng noi
                                       khoang lang
        Voiced(k) = 0;              % Gan Voiced(i) = 0 neu khung be hon gia
                                       tri nguong
        end
    end
    for k = 1: fr_num           % Duyet het tat cac khung
        if (Voiced(k) == 1)     % Kiem tra neu khung thu k co gia tri bang 1 ?
            for i = k:k+9       % Duyet tiep them 9 khung ke tu khung thu k
                if Voiced(i) == 1 % Kiem tra neu cac khung tiep theo co gia tri
                                     bang 1
                    for j=k:i          % Duyet cac khung tu k den k +9
                        Voiced(j) = 1; % Gan gia tri cac khung tu k den k+9= 1
                    end
                end
            end
        end
    end
end

% Ham xac dinh vi tri bien cua tieng noi-khoang lang
% Ham tra ve vi tri bien dau, bien cuoi cua tieng noi
% duoc luu vao vector Position
% fr_num: tong so frame cua tin hieu
% fr_len: so mau tren mot frame
function [Position] = Position_function(Voiced, fr_num, fr_len)
    Position = [];
    j = 1;
    for i = 2:fr_num      % Duyet tat ca cac khung
        if(Voiced(i) == 1 && Voiced(i-1)==0)
            Position(2*j - 1)= fr_len/2 + (i-1)*fr_len; % Luu vi tri bien dau
            j = j + 1;                                    cua tieng noi
        end
    end
    j = 1;
    for i = 2:fr_num
        if (Voiced(i) == 0 && Voiced(i-1) == 1)
            Position(2*j) = fr_len/2 + (i-1)*fr_len;% Luu vi tri bien cuoi cua
            j = j + 1;                                tieng noi
        end
```

```
    end
end

% Ham xac dinh nguong phan tach tieng noi - khoang lang
% Ham tra ve gia tri cua nguong phan tach tieng noi - khoang lang
function [Ethres] = Ethres_function(Power,S)
 avg = 0;
 for i= 1:10
    avg = avg + Power(i)/10; % Tim gia tri trung binh cua khung trong 10 khun
 end                           dau tien
 vari = 0;
 for i= 1:10
     vari = vari + ((Power(i) - (vari))^2)/10;  % Tim phuong sai cua 10 khung
end                                               dau tien
    Ethres = S*(avg + vari*0.2*vari^0.8);  % Ap dung cong thuc de tim nguong
end

% Ham chuyen doi dai luong can thiet de tim ham zcr
% Ham tra ve ma tran cac tin hieu (-1,1)
% fr_num: tong so frame cua tin hieu
% fr_len: so mau tren mot frame
function [sgn] = sgn_function(X_ZCR, fr_num, fr_len)
    sgn = zeros(fr_num,fr_len);
    k = 1;
    for i = 1:fr_num                % Duyet tat ca cac khung
        for j = 1:fr_len            % Duyet tat ca cac mau co trong khung
            x(i,j) = X_ZCR(k);
            if(X_ZCR(k)>0)          % Kiem tra ZCR cua khung thu k > 0
                sgn(i,j) = 1;       % Gan cho mang sgn gia tri 1
            else
                sgn(i,j) = -1;      % Gan cho mang sgn gia tri -1
            end
            k = k + 1 ;
        end
    end
end

% Ham tim ZCR
% Ham tra ve mang cac gia tri ZCR cua moi khung
% fr_num: tong so frame cua tin hieu
% fr_len: so mau tren mot frame
function [ZCR] = ZCR_function(sgn,fr_num,fr_len)
ZCR = zeros(1,fr_num);
    for i = 1 : fr_num                          % Duyet tat ca cac khung
        ZCR(1) = abs(sgn(1,1))/2;
        for j = 2:fr_len
            zcr = abs(sgn(i,j)- sgn(i,j-1))/2;  % Tinh toan zcr dua vao cong
                                                  thuc tính zcr
            ZCR(i) = ZCR(i) + zcr;              % Luu gia tri cua zcr vao mang
                                                ZCR
        end
    end
end

% Ham chuan hoa lai tin hieu ban dau de tim ham zcr
% Ham tra ve tin hieu sau khi da chuan hoa
function [X_ZCR] = X_ZCR_function(x)
avg = 0;
for i = 1:length(x)
    avg = avg + x(i);       % Tim tong tat cac cac gia tri bien do cua tin hieu
                              ban dau
end
avg = avg/length(x);        % Lay trung binh cong gan vao bien avg
```

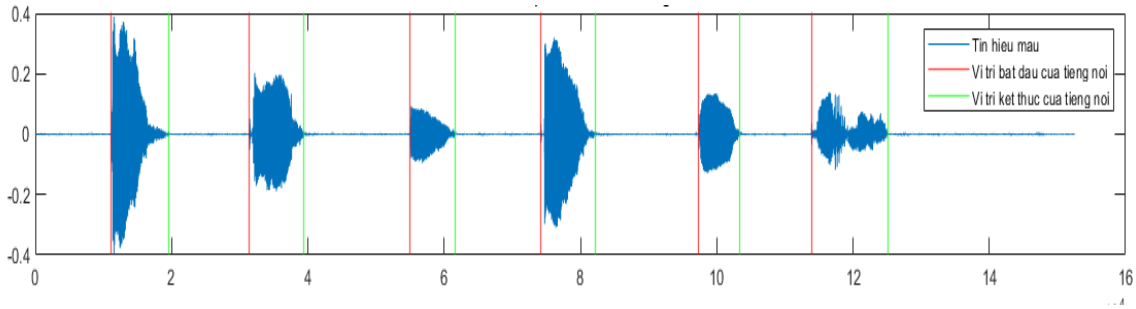
```
x1 = zeros(1,length(x));
for i = 1:length(x)
    x1(i) = x(i) - avg;       % Dich chuyen vi tri cua x avg don vi
end
X_ZCR = x1;
end
% Ham xac dinh W ket hop giua ham STE va ZCR
% Ham tra ve mang gia tri cua W khi ket hop hai ham STE va ZCR
function [W] = W_function(Power, ZCR, fr_len, fr_num)
    P = Power/(fr_len);
    ZCR = ZCR/(fr_len);
    W = zeros(1,fr_num);
    for i = 1:fr_num
        W(i) = P(i)*(1-ZCR(i))*1000;
    end
end
% Ham truoc khi xu ly nhung doan khoang lang < 200ms
function [Pre_Voiced] = Pre_DetecVoiced_function(Standard, fr_num, Thres)
    Pre_Voiced = zeros(1,fr_num);
    for k = 1:fr_num
        if (Standard(k) > Thres)
            Pre_Voiced(k) = 1;
        end
    end
    for k = 1:fr_num
        if (Standard(k)< Thres)
        Pre_Voiced(k) = 0;
        end
    end
end

% Ham tinh trung binh do lon tren moi khung
% Tra ve: Gia tri trung binh tren moi khung
% x: mang gia tri bien do cua tin hieu am thanh
% fr_num: tong so frame cua tin hieu
% fr_len: so mau tren mot frame
function [MA] = MA_function(x, fr_num, fr_len)
    MA = zeros(1, fr_num); % Tao mang de chua cac gia tri trung binh
    for k = 1:fr_num       % Duyet tat ca cac khung
        temp = 0;          % Bien tam dung luu gia tri trung binh cua moi khung
                             khi chay vong lap
                           % Duyet tat ca cac mau cua tin hieu trong 1 khung
        for j = (k-1)*fr_len + 1 : k*fr_len - 1
            temp = temp + abs(x(j));
        end
        MA(k) = temp;  % Gia tri trung binh tren moi khung duoc dua vao mang MA
    end
end
```

IV. KẾT QUẢ THỰC NGHIỆM

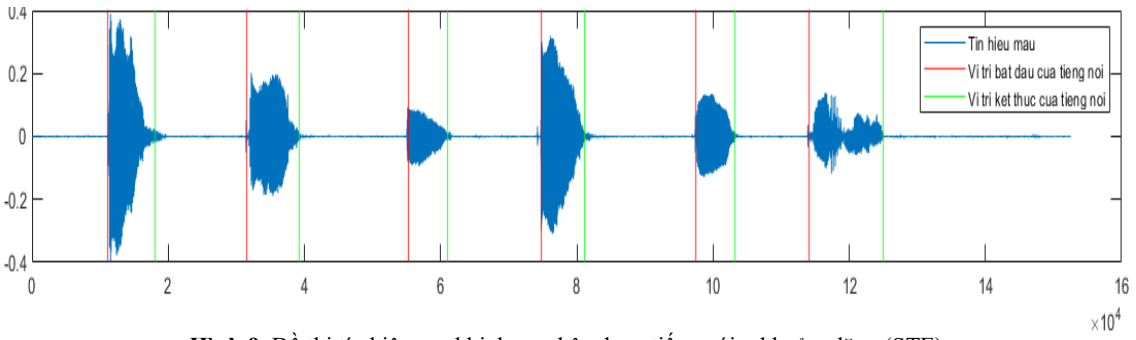
A. Kết quả định tính

Dưới đây là các hình vẽ biên thời gian được xác định thủ công và biên thời gian được xác định bằng thuật toán:

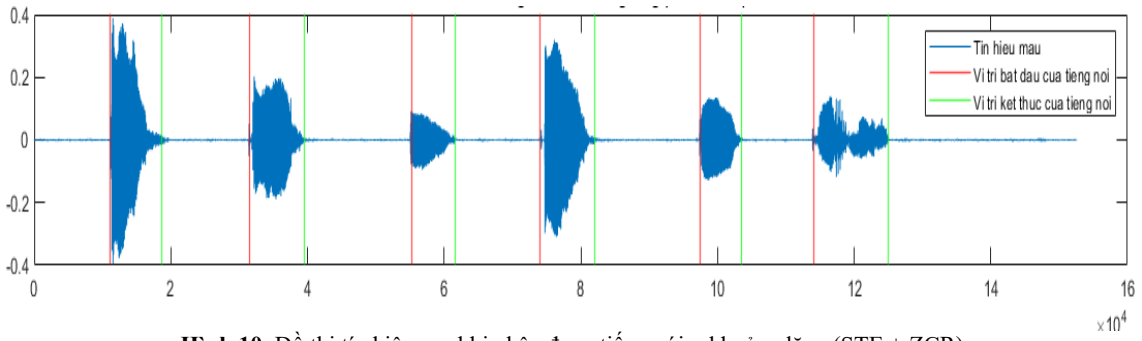
1. Kết quả đối với File lab_female.wav



Hình 8. Biên thời gian được xác định thủ công



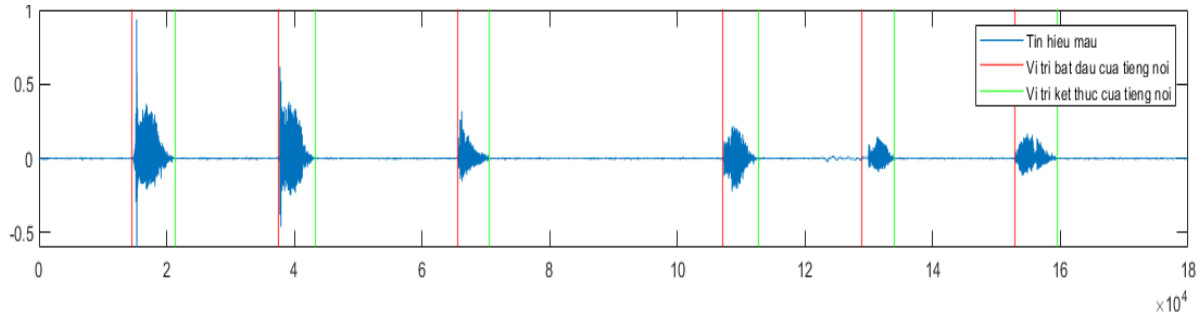
Hình 9. Đồ thị tín hiệu sau khi được phân đoạn tiếng nói – khoảng lặng (STE)



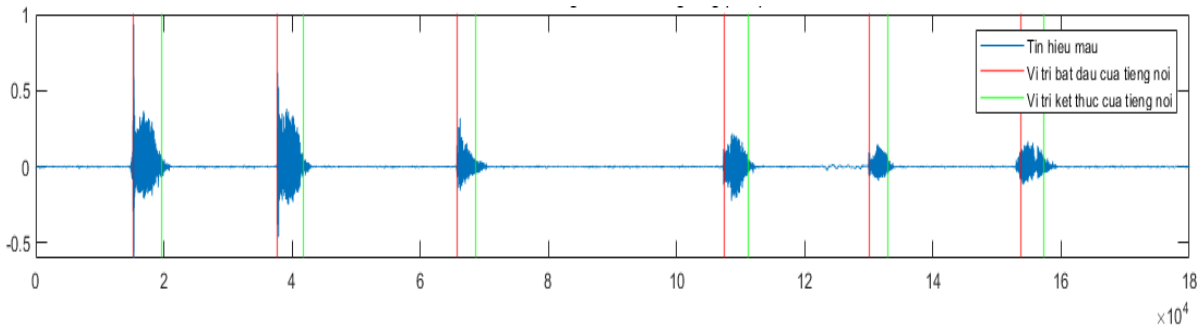
Hình 10. Đồ thị tín hiệu sau khi phân đoạn tiếng nói – khoảng lặng (STE + ZCR)

Hình 11. Đồ thị tín hiệu sau khi phân đoạn tiếng nói – khoảng lặng (MA)

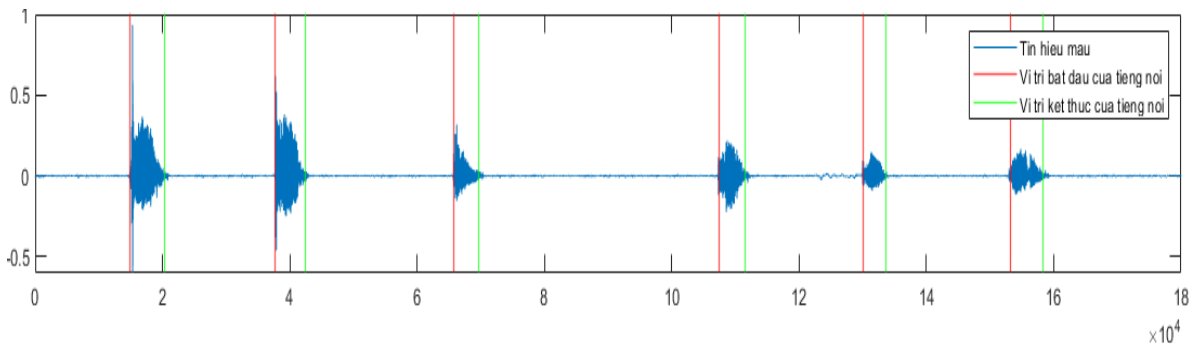
2. Kết quả đối với File lab_male.wav



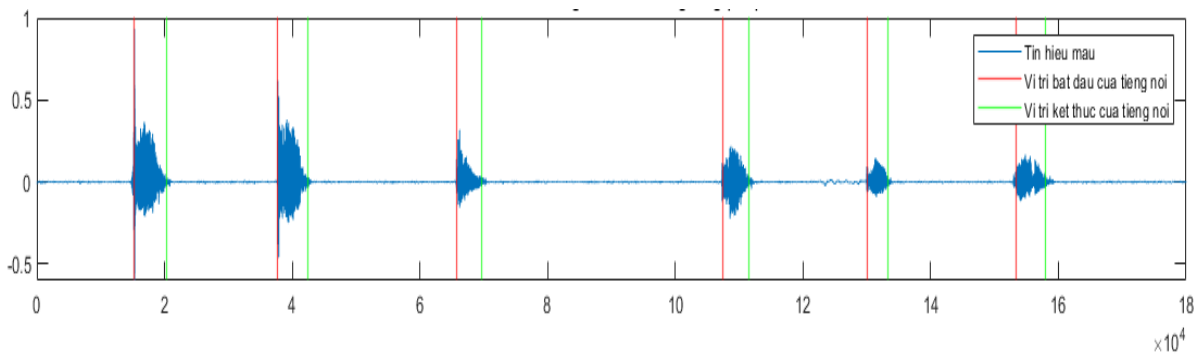
Hình 12. Biên thời gian được xác định thủ công



Hình 13. Đồ thị tín hiệu sau khi được phân đoạn tiếng nói – khoảng lặng (STE)

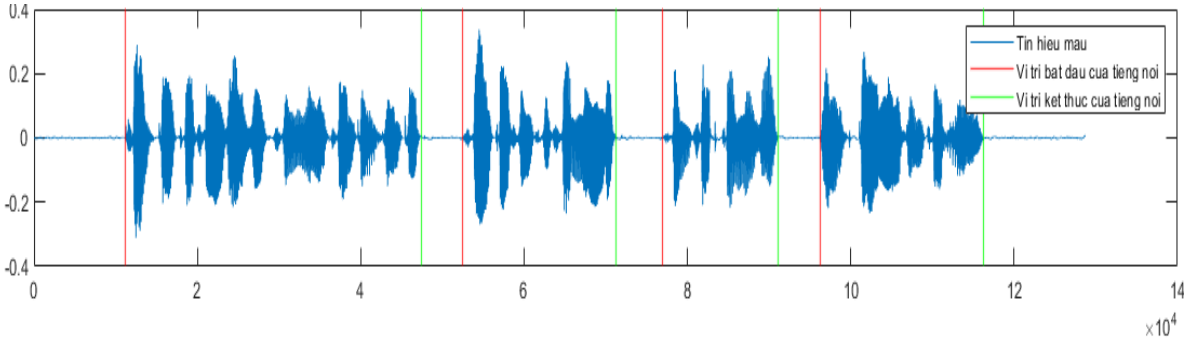


Hình 14. Đồ thị tín hiệu sau khi phân đoạn tiếng nói – khoảng lặng (STE + ZCR)

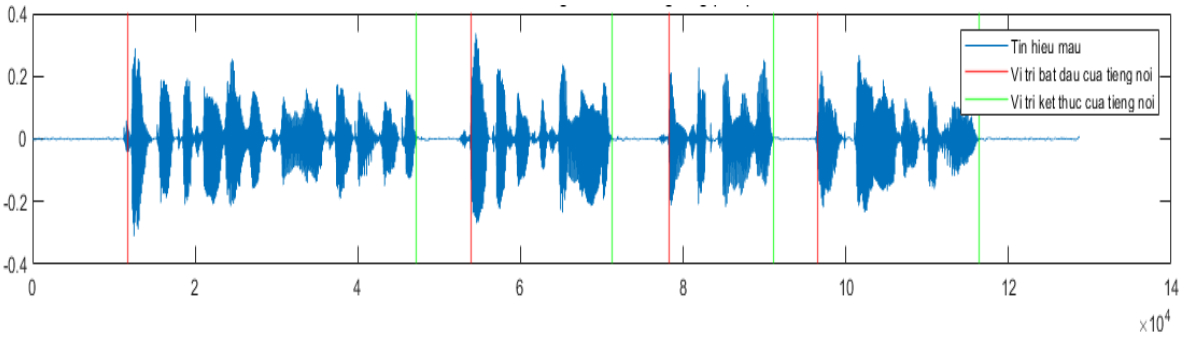


Hình 15. Đồ thị tín hiệu sau khi phân đoạn tiếng nói – khoảng lặng (MA)

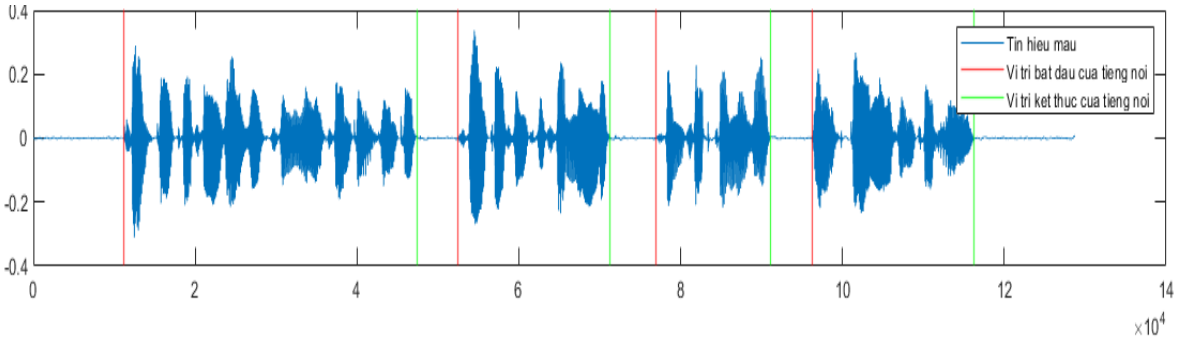
3. Kết quả đối với File studio_male.wav



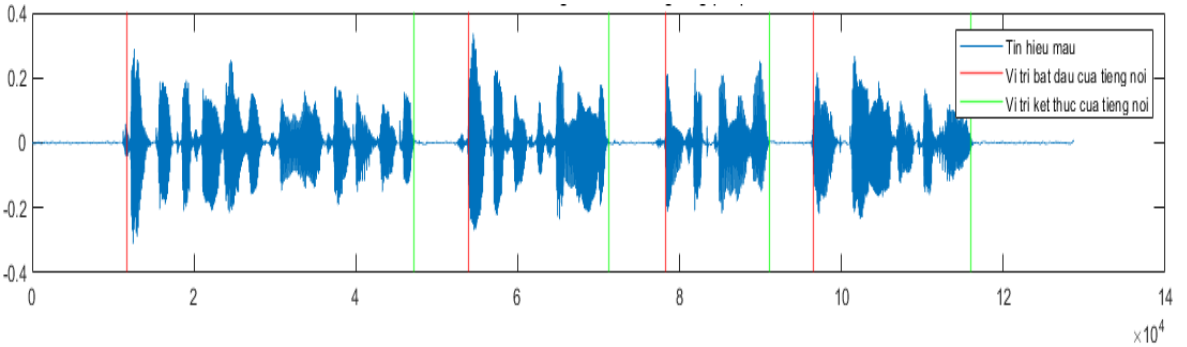
Hình 16. Biên thời gian được xác định thủ công



Hình 17. Đồ thị tín hiệu sau khi được phân đoạn tiếng nói – khoảng lặng (STE)

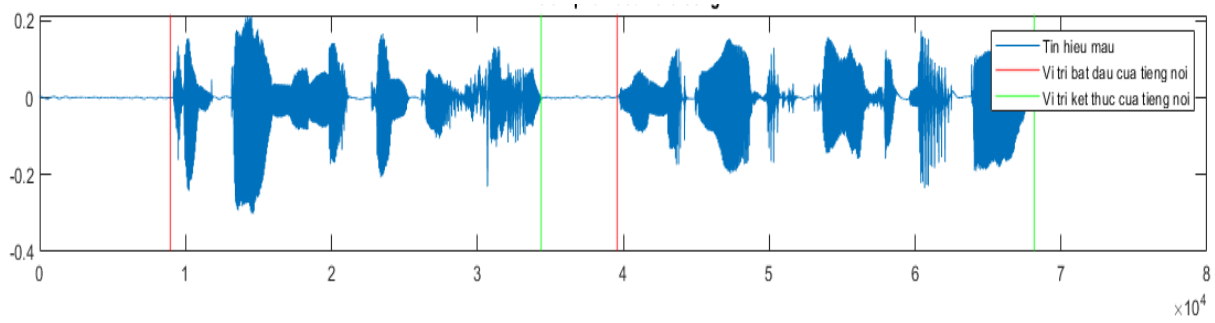


Hình 18. Đồ thị tín hiệu sau khi phân đoạn tiếng nói – khoảng lặng (STE + ZCR)

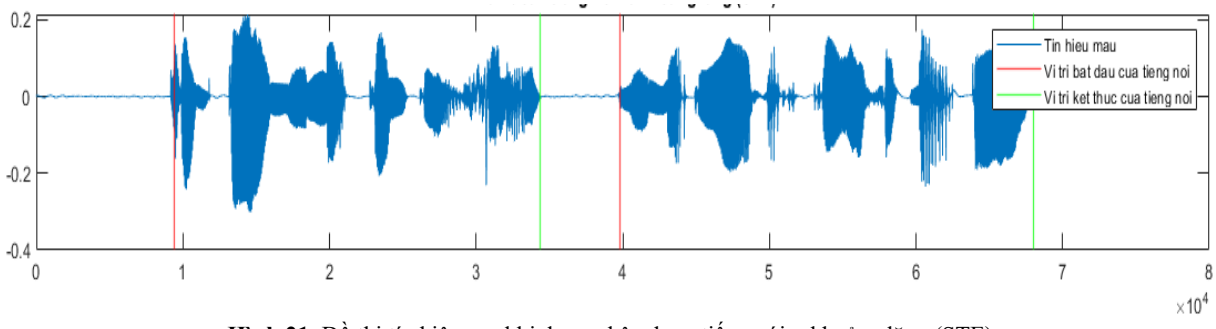


Hình 19. Đồ thị tín hiệu sau khi phân đoạn tiếng nói – khoảng lặng (MA)

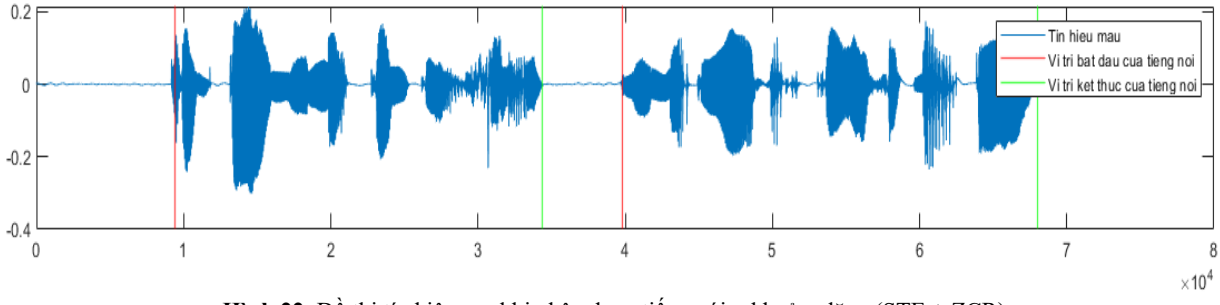
4. Kết quả đối với File studio_female.wav



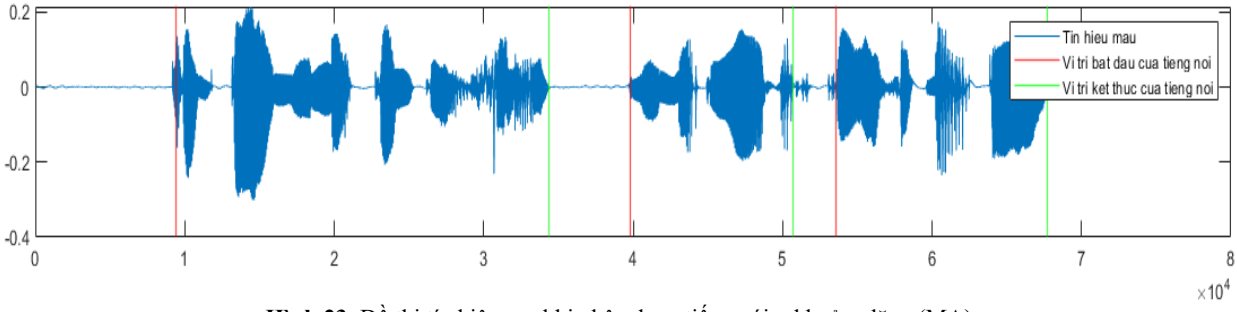
Hình 20. Biên thời gian được xác định thủ công



Hình 21. Đồ thị tín hiệu sau khi được phân đoạn tiếng nói – khoảng lặng (STE)



Hình 22. Đồ thị tín hiệu sau khi phân đoạn tiếng nói – khoảng lặng (STE + ZCR)



Hình 23. Đồ thị tín hiệu sau khi phân đoạn tiếng nói – khoảng lặng (MA)

5. Nhận xét kết quả:

Sau khi các thuật toán được thực hiện trên 4 file mẫu, các biên thời gian tiếng nói và khoảng lặng đã được biểu diễn trên các đồ thị, sau khi quan sát đồ thị chúng tôi nhận thấy rằng:

- Thuật toán xác định biên nhờ vào sự kết hợp giữa hàm năng lượng ngắn hạn và tỷ lệ vượt qua không là hiệu quả nhất. Các biên được xác định chênh lệch rất ít so với biên được xác định thủ công và điều này đúng trên tất cả các file tín hiệu. Tuy nhiên vẫn còn một số khung tiếng nói, hoặc khoảng lặng ở xung quanh biên vẫn bị bỏ qua.
- Thuật toán xác định biên nhờ hàm biên độ trung bình cho hiệu quả kém nhất. Các biên được xác định có sự chênh lệch nhiều so với biên được xác định thủ công, bỏ qua rất nhiều khung hình chứa tiếng nói. Bên cạnh đó, ở file tín hiệu studio_female.wav, thuật toán đã xác định sai biên, cụ thể thêm 1 khoảng tiếng nói so với biên được xác định thủ công.
- Thuật toán xác định biên nhờ hàm năng lượng ngắn hạn cho hiểu quả trung bình, khả năng phân loại tiếng nói khoảng lặng thấp hơn thuật toán năng lượng ngắn hạn kết hợp tỷ lệ vượt qua không, nhưng cao hơn thuật toán tìm biên độ trung bình. Tuy nhiên vẫn một số khung khhung tiếng nói hoặc khoảng lặng xung quanh biên đã bị bỏ qua.

B. Kết quả định lượng

1. Kết quả đối với File Lab_female.wav

File Lab_female.wav							
Thông tin của file							
Fs(Hz)	Tổng số mẫu	Độ dài khung hình(ms)	Tổng số khung hình	Tổng số mẫu/khung hình			
16000	152529	20	476	320			
Vị trí các biên bắt đầu và kết thúc của tiếng nói							RMSE (Thủ công-Thuật toán)
Thủ công	11066-19634	31334-39429	55010-61614	74171-82260	97346-103315	113887-125102	
Thuật toán STE	11040-18080	31520-39200	55200-60960	74720-81120	97440-103200	114080-124960	622.289
Thuật toán STE +ZCR	11040-18720	31520-39520	55200-61600	74080-82080	97440-103520	114080-124960	297.798
Thuật toán MA	11360-18720	31520-39200	55200-60960	74720-81440	97440-103200	114080-124960	269.9938

2. Kết quả đối với File Lab_male.wav

File Lab_male.wav							
Thông tin của file							
Fs(Hz)	Tổng số mẫu	Độ dài khung hình(ms)	Tổng số khung hình	Tổng số mẫu/khung hình			
16000	179985	20	562	320			
Vị trí các biên bắt đầu và kết thúc của tiếng nói							RMSE (Thủ công - Thuật toán)
Thủ công	14519-21359	37435-43339	65582-70570	107175-112666	128876-134030	152803-159493	
Thuật toán STE	15200-19680	37600-41760	65760-68640	107360-111200	130080-132960	153760-157280	1295.90
Thuật toán STE + ZCR	14880-20320	37600-42400	65760-69600	107360-111520	130080-133600	153120-158240	803.3226
Thuật toán MA	15200-20320	37600-42400	65760-69600	107360-111520	130080-133280	153440-157920	897.43

3. Kết quả đối với File studio_male.wav

File studio_male.wav					
Thông tin của file					
Fs(Hz)	Tổng số mẫu	Độ dài khung hình(ms)	Tổng số khung hình	Tổng số mẫu/khung hình	
16000	128736	20	402	320	
Vị trí các biên bắt đầu và kết thúc của tiếng nói					RMSE (Thủ công – Thuật toán)
Thủ công	11201-47415	52431-71211	76935-91042	96296-116301	
Thuật toán STE	11680-47200	53920-71200	78240-91040	96480-116320	727.1652
Thuật toán STE + ZCR	11680-47520	53920-71200	78240-91040	96480-116320	724.1334
Thuật toán MA	11680-47200	53920-71200	78240-91040	96480-116000	734.8804

4. Kết quả đối với File studio_female.wav

File studio_female.wav				
Thông tin của file				
Fs(Hz)	Tổng số mẫu	Độ dài khung hình(ms)	Tổng số khung hình	Tổng số mẫu/khung hình
16000	78618	20	245	320
Vị trí các biên bắt đầu và kết thúc của tiếng nói				RMSE (Thủ công - Thuật toán)
Thủ công	8996-34395	39570-68135		
Thuật toán STE	9440-34400	39840-68000		268.4614
Thuật toán STE + ZCR	9440-34400	39840-68000		268.4614
Thuật toán MA	9440-34400	39840-50720	53600-67680	Lỗi do xác định sai các biên.

5. Nhận xét kết quả:

Sau khi các thuật toán được thực hiện ở tất cả các file, kết quả định lượng được mô tả ở trên bảng thống kê cho thấy rằng:

- Việc xác định biên tiếng nói và khoảng lặng ảnh hưởng rất nhiều từ thuật toán chọn ngưỡng.
- Từ việc tính toán sự sai khác giữa giá trị biên thời gian được xác định thủ công và biên thời gian được xác định bằng thuật toán thể hiên qua chỉ số RMSE. Thì thuật toán kết hợp giữa Năng lượng ngắn hạn (STE) và hàm tỷ lệ vượt qua điểm không (ZCR) cho thấy thuật toán này hiệu quả hơn hai thuật toán còn lại. Điều này thể hiện ở chỉ số RMSE thấp nhất ở hầu hết các file, trừ file lab_female.wav, nhỉnh hơn một ít so với thuật toán MA.
- Thuật toán dùng hàm trung bình biên độ (MA) cho hiệu quả trung bình, đối với các file lab, chỉ số RMSE thấp hơn so với thuật toán năng lượng ngắn hạn, nhưng chỉ số này lại cao hơn nếu thu âm trong điều kiện studio. Bên cạnh đó, với file studio_female.wav, thuật toán đã xác định sai biên, cụ thể là thêm một khoảng tiếng nói, vì vậy không thể tính chỉ số RMSE ở tín hiệu này.
- Thuật toán dung hàm năng lượng ngắn hạn (STE) cho hiệu quả trung bình, trái ngược với thuật toán MA, các file này hoạt động hiệu quả hơn ở môi trường studio, và kém hiệu quả hơn so với file thu âm ở môi trường lab. Điều này thể hiện rõ ở chỉ số RMSE trên hai môi trường thu âm.
- Từ đây có thể thấy thuật toán kết hợp giữa năng lượng ngắn hạn và tỷ lệ vượt qua không là hoạt động ổn định nhất.

Từ kết quả thực nghiệm, chúng tôi đưa ra những kết luận như sau:

Ưu điểm:

- Hàm năng lượng ngắn hạng, hàm năng lượng ngắn hạn kết hợp hàm tỷ lệ vượt qua không và hàm trung bình biên độ là các phương pháp đơn giản và khá hiệu quả trong việc giải bài toán tìm biên tiếng nói và khoảng lặng của các tín hiệu âm thanh.
- Các thuật toán hoạt động với hiệu suất cao, ổn định đối với những tín hiệu âm thanh ít lẫn tạp âm

Nhược điểm:

- Các thuật toán khá nhạy cảm đối với các tín hiệu âm thanh lẫn nhiều tạp âm, dễ dẫn đến sự sai lệch khi xác định biên thời gian tiếng nói và khoảng lặng.
- Phụ thuộc lớn vào thuật toán tìm ngưỡng để phân tách tiếng nói và khoảng lặng.

V. KẾT LUẬN

A. Kết quả đạt được

Báo cáo này thực hiện việc xử lý và phân đoạn tín hiệu để xác định biên thời gian giữa tiếng nói và khoảng lặng. Được thực hiện dựa vào ba thuật toán năng lượng ngắn hạn (STE), năng lượng ngắn hạn (STE) kết hợp với tỷ lệ vượt qua không (ZCR) và thuật toán trung bình biên độ (MA).

Từ kết quả qua quá trình thực nghiệm trên 4 file tín hiệu cho trước đã cho thấy rằng các thuật toán có thể hoạt động hiệu quả, đã phân biệt được các khoảng tiếng nói và khoảng lặng. Trong các thuật toán đã thực hiện, thuật toán kết hợp giữa năng lượng ngắn hạn (STE) và tỷ lệ vượt qua không (ZCR) hoạt động hiệu quả nhất. Tuy nhiên vẫn còn một số phân đoạn tiếng nói nhỏ xung quanh các biên vẫn chưa được xác định. Bên cạnh đó, việc thuật toán có hoạt động hiệu quả hay không phụ thuộc lớn vào môi trường thu âm, nếu môi trường có quá nhiều tạp âm, hay tiếng ồn, điều ảnh ảnh hưởng rất nhiều đến kết quả của thuật toán.

B. Phương hướng phát triển

Trong tương lai, để xử lý các vấn đề còn xảy ra đối với thuật toán này, chúng tôi sẽ tìm hiểu thêm về những đặc trưng khác ví dụ như tần số cơ bản, dạng sóng trong chu kỳ cơ bản,… của tín hiệu âm thanh để có thể cải thiện hiệu suất của thuật toán, hoặc có thể tạo ra những thuật toán khác giúp giải quyết những vấn đề này nhằm nâng cao hiệu suất xử lý tín hiệu âm thanh. Cuối cùng, những kết quả đạt được sẽ được ứng dụng cho trong việc phân tích và xử lý nhận dạng giọng nói.

VI. TÀI LIỆU THAM KHẢO

[1] Speech segementation, from https://en.wikipedia.org/wiki/Speech_segmentation
[2] Xử lý tiếng nói, from https://vi.wikipedia.org/wiki/X%E1%BB%AD_l%C3%BD_ti%E1%BA%BFng_n%C3%B3i
[3] Matthieu Hodgkinson, “CS425 Audio and Speech Processing_Hodgkinson_2012.pdf”,vol.2,no.1, p.33-34,April 25, 2012
[4] Loại bỏ khoảng lặng, from https://tieuluan.info/li-ni-u-phn-i-gii-thiu-t-engine-sh7760-4.html?page=5&fbclid=IwAR2KYsEVSOAcgBIVqtWTKemA_7xYeK6RofcSq-QIcixcxHrGw9eWbrO6F0g