




Welcome!!



PHÂN ĐOẠN TIẾNG NÓI VÀ KHOẢNG LẶNG

- Nguyễn Huy Cồn
 - Trần Trọng Bảo
 - Bùi Phan Minh Hưng
 - Phạm Ngọc Hiếu (nhóm trưởng)
 - Giảng viên hướng dẫn : Ninh Khánh Duy
- 

ĐẶT VẤN ĐỀ

01

Phân đoạn tiếng nói khoảng lặng là phân đoạn những đoạn có tín hiệu tiếng nói và những đoạn không có tín hiệu tiếng nói (nhưng có thể có rè, nhiễu)

02

Việc phân đoạn tiếng nói và khoảng lặng làm cơ sở cho việc trích xuất thông tin từ các tín hiệu giọng nói

03

Dựa trên các đặt trưng của tín hiệu không đổi trong thời gian ngắn hạn (10-30ms), ta có thể chia tín hiệu thành các khung để xử lý

LÝ THUYẾT CHUNG VỀ XỬ LÝ TÍN HIỆU VÀ TIẾNG NÓI

1

Xử lý tín hiệu

Là phân tích sửa đổi và tổng hợp các tín hiệu như âm thanh, hình ảnh

2

Tiếng nói

Là những gì do một cá nhân phát âm để bày tỏ và trao đổi tư tưởng

3

Xử lý tiếng nói

Là sự nghiên cứu tiếng nói con người dưới dạng tín hiệu, và các phương pháp xử lý những tín hiệu này



CƠ SỞ LÝ THUYẾT

I. THUẬT TOÁN CHUẨN HÓA NĂNG LƯỢNG 1

1. Chuẩn hóa phần dư

Là việc chuẩn hóa năng lượng về đoạn $[0,1]$ bằng phương pháp dựa vào giá trị min và max của năng lượng

Công thức toán học

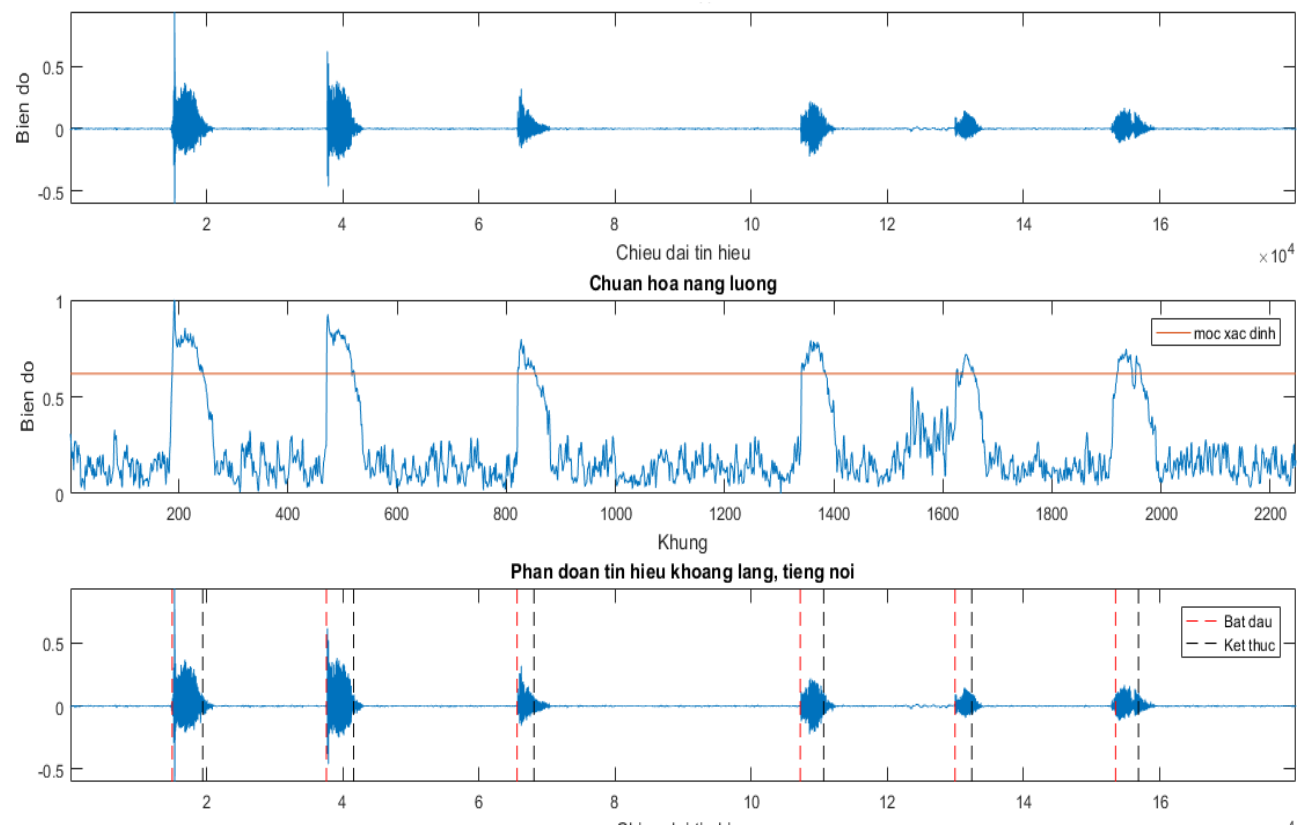
$$X_{norm} = \frac{x - x_{min}}{x_{max} - x_{min}}$$

Trong đó : X_{norm} là năng lượng sau khi chuẩn hóa

x là năng lượng đầu vào

x_{min} là giá trị năng lượng nhỏ nhất

x_{max} là giá trị năng lượng lớn nhất



CÁC THAM SỐ QUAN TRỌNG



Độ dài khung tín hiệu

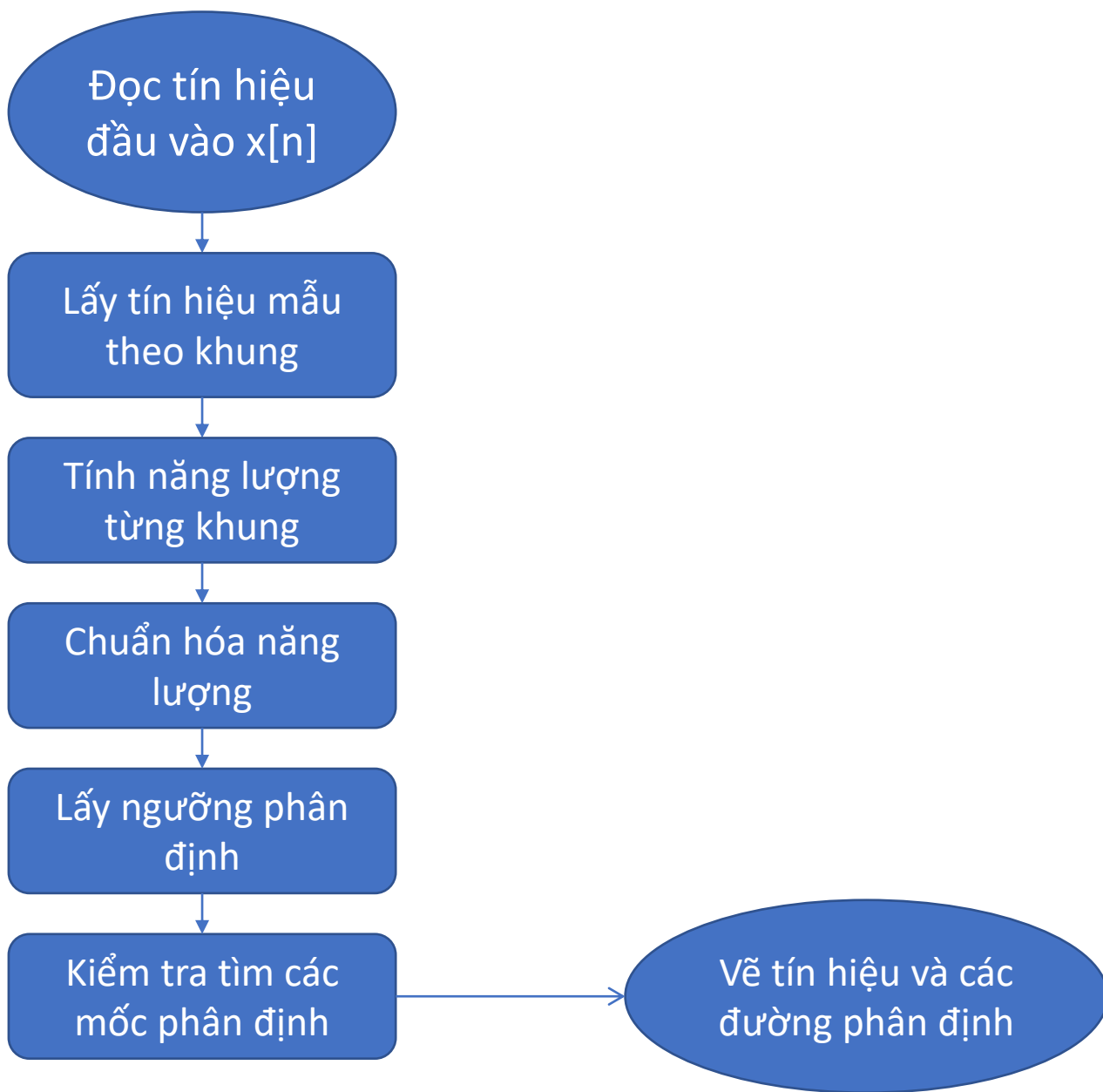
Độ dài khung tín hiệu không được quá ngắn không được quá dài

Tần số lấy mẫu

F_s càng lớn đặc trưng tín hiệu càng chính xác

Khoảng lặng tối thiểu

Lớn hơn 200 ms



VẤN ĐỀ PHÁT SINH VÀ CÁCH GIẢI QUYẾT



Vấn đề phát sinh

Chiều dài năng lượng khung khác chiều dài tín hiệu.
Nhiều có năng lượng lớn.

Cách giải quyết

Tìm mốc trên khung rồi tính ra mốc trên tín hiệu.
Lấy ngưỡng lớn để loại nhiễu.

II. THUẬT TOÁN CHUẨN HÓA NĂNG LƯỢNG 2

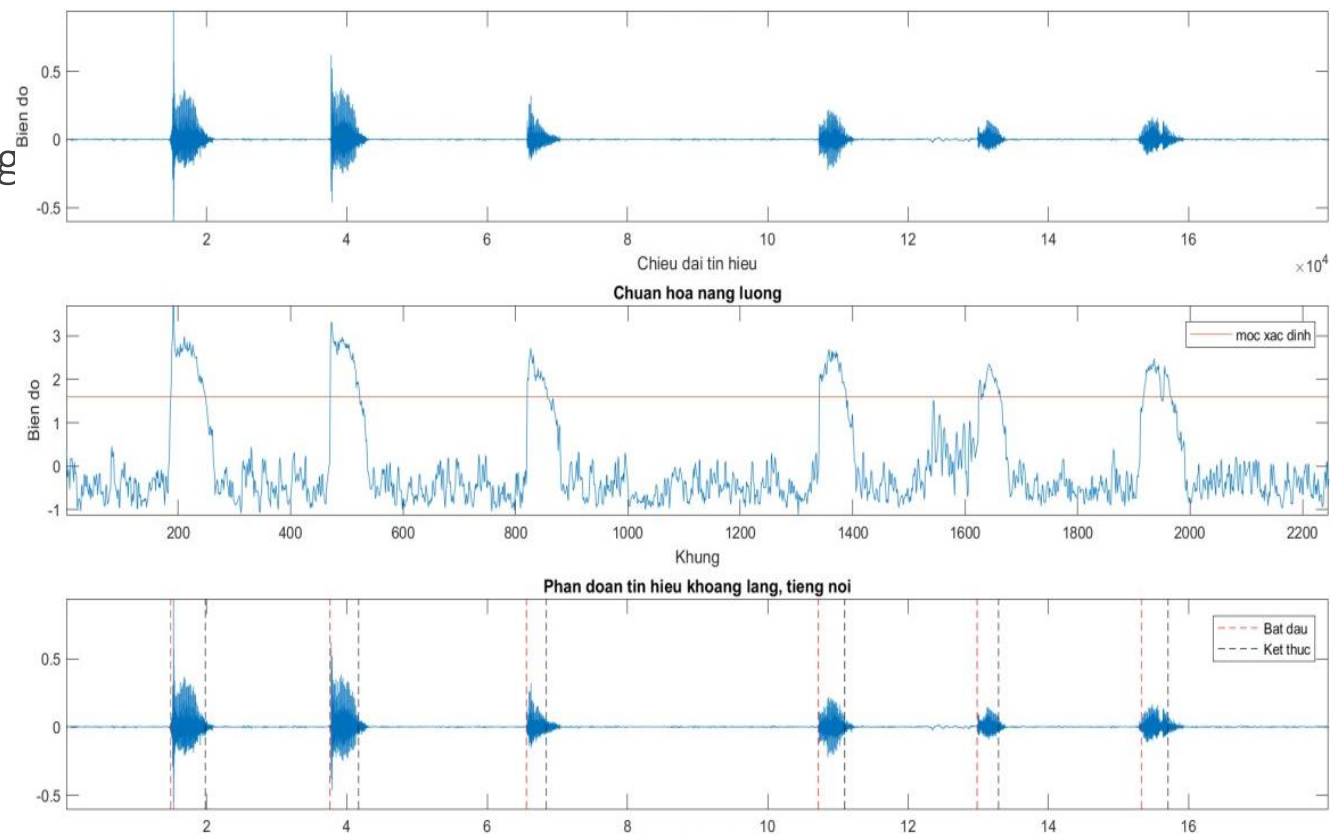
1. Chuẩn hóa bằng phân phối chuẩn

Là việc chuẩn hóa năng lượng về phân bố chuẩn bằng phương pháp dựa vào giá trị trung bình và độ lệch chuẩn

Công thức toán học

$$X_{norm} = \frac{x - \bar{x}}{\sigma x}$$

Trong đó : X_{norm} là năng lượng sau khi chuẩn hóa
 x là năng lượng đầu vào
 \bar{x} là giá trị trung bình
 σx là độ lệch chuẩn



CÁC THAM SỐ QUAN TRỌNG



Độ dài khung tín hiệu

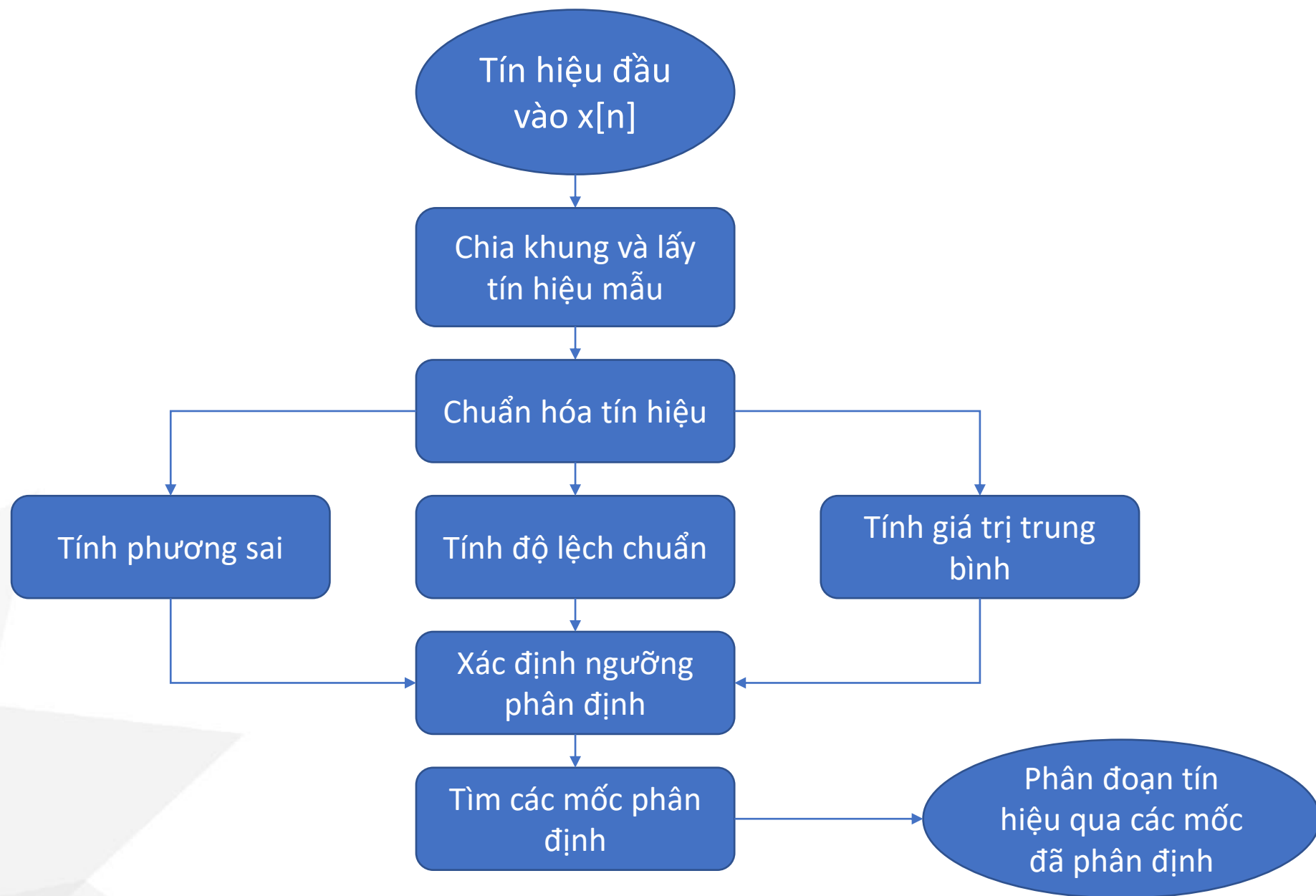
Độ dài khung tín hiệu không được quá ngắn không được quá dài

Tần số lấy mẫu

F_s càng lớn đặc trưng tín hiệu càng chính xác

Khoảng lặng tối thiểu

Lớn hơn 200 ms



VẤN ĐỀ PHÁT SINH VÀ CÁCH GIẢI QUYẾT



Vấn đề phát sinh

Miền năng lượng của các tín hiệu sau chuẩn hóa không giống nhau

Cách giải quyết

Đối với âm thanh có độ nhiễu lớn phải xác định ngưỡng bằng mắt

III. THUẬT TOÁN MEAN AVERAGE

1. Magnitude Average

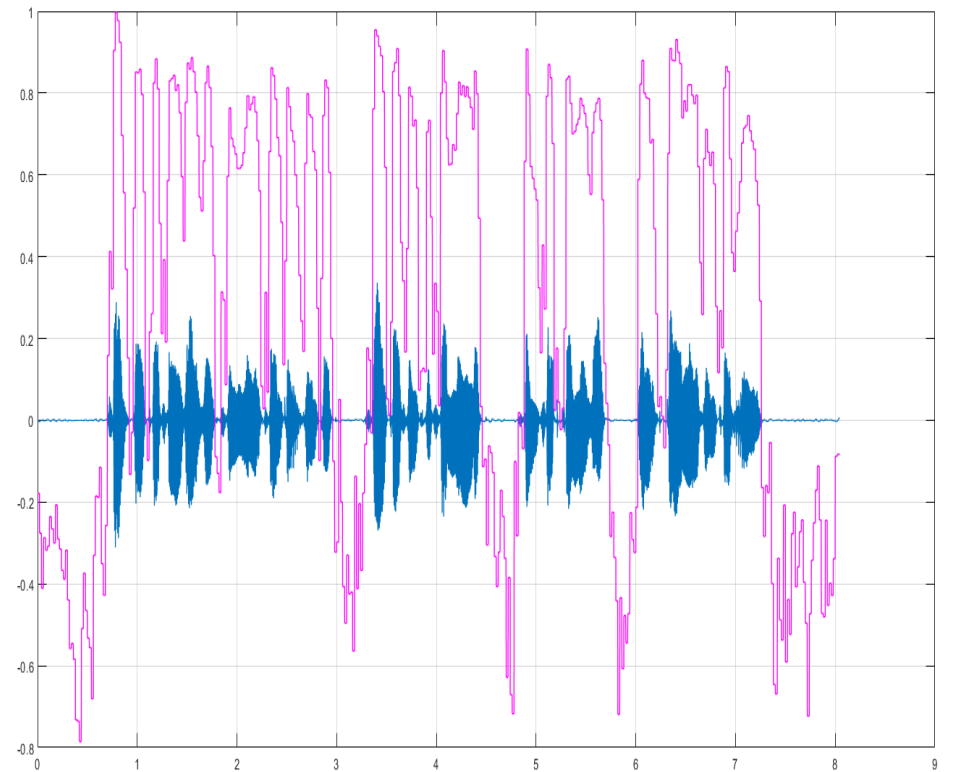
Magnitude Average là cường độ trung bình của tín hiệu trong khoảng thời gian ngắn

Công thức toán học

$$MA[n] = \sum_{m=-\infty}^{+\infty} |x[n-m]| \times w[m]$$

Trong đó : $w[n] = \begin{cases} 1, & 0 \leq n \leq N \\ x, & \text{còn lại} \end{cases}$

Các âm hữu thanh thường có cường độ trung bình lớn
còn âm vô thanh thường có cường độ trung bình nhỏ





2. Mean Average

Thuật toán này được xây dựng dựa trên việc nếu giá trị Magnitude Average của tín hiệu tiếng nói cao hơn nhiều so với khoảng lặng thì sẽ có một ngưỡng T và cách xây dựng thuật toán tìm T đã được ghi trong tài liệu. Bằng cách cân bằng khoảng nhằm lẫn giữa hai mảng f chứa giá trị MA ở khung thứ lẻ là mảng tiếng nói và g chứa giá trị MA ở khung thứ chẵn là mảng khoảng lặng

CÁC THAM SỐ QUAN TRỌNG



Độ dài khung tín hiệu

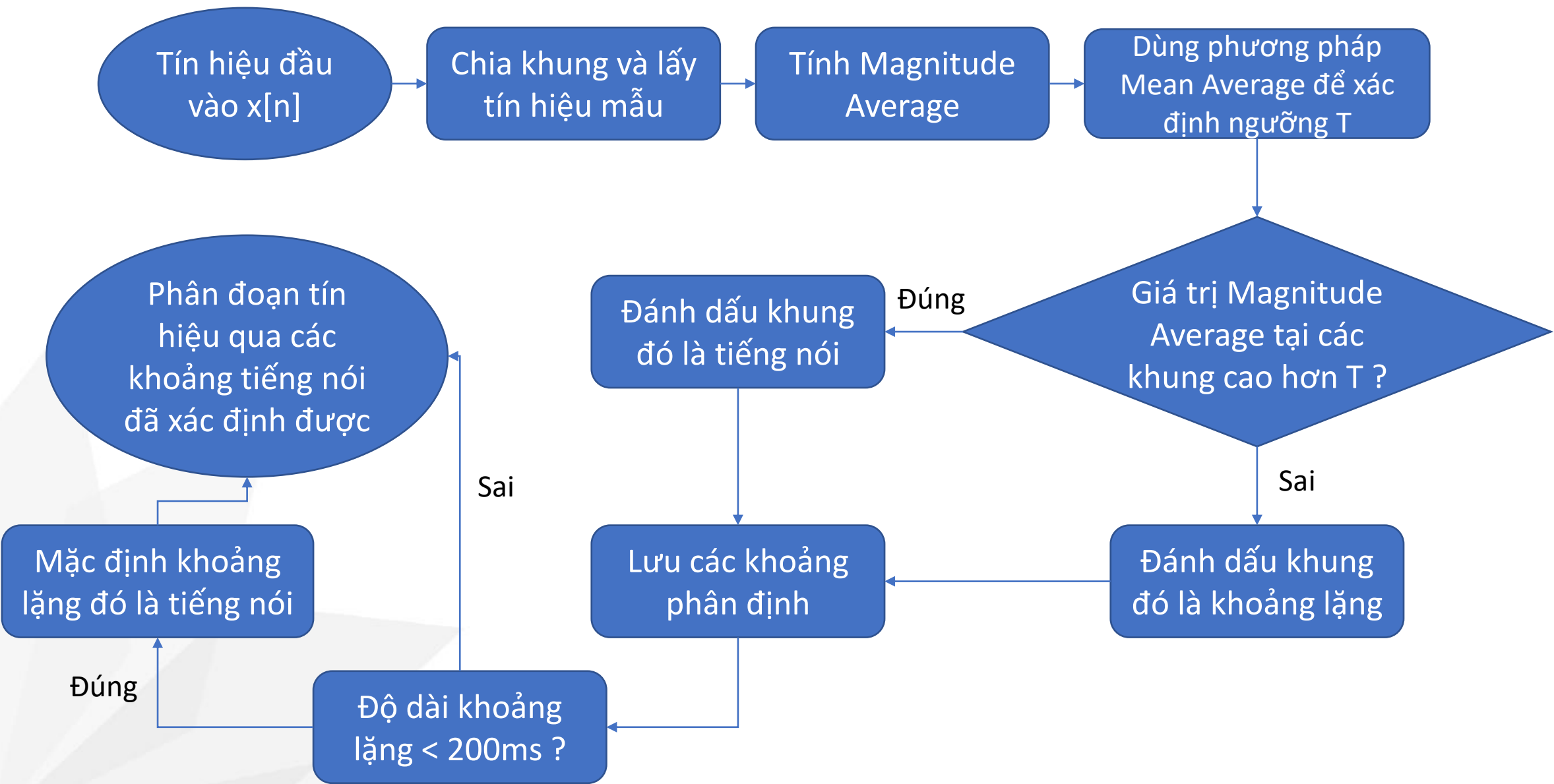
Độ dài khung tín hiệu không được quá ngắn không được quá dài

Tần số lấy mẫu

F_s càng lớn đặc trưng tín hiệu càng chính xác

Khoảng lặng tối thiểu

Lớn hơn 200 ms



VẤN ĐỀ PHÁT SINH VÀ CÁCH GIẢI QUYẾT

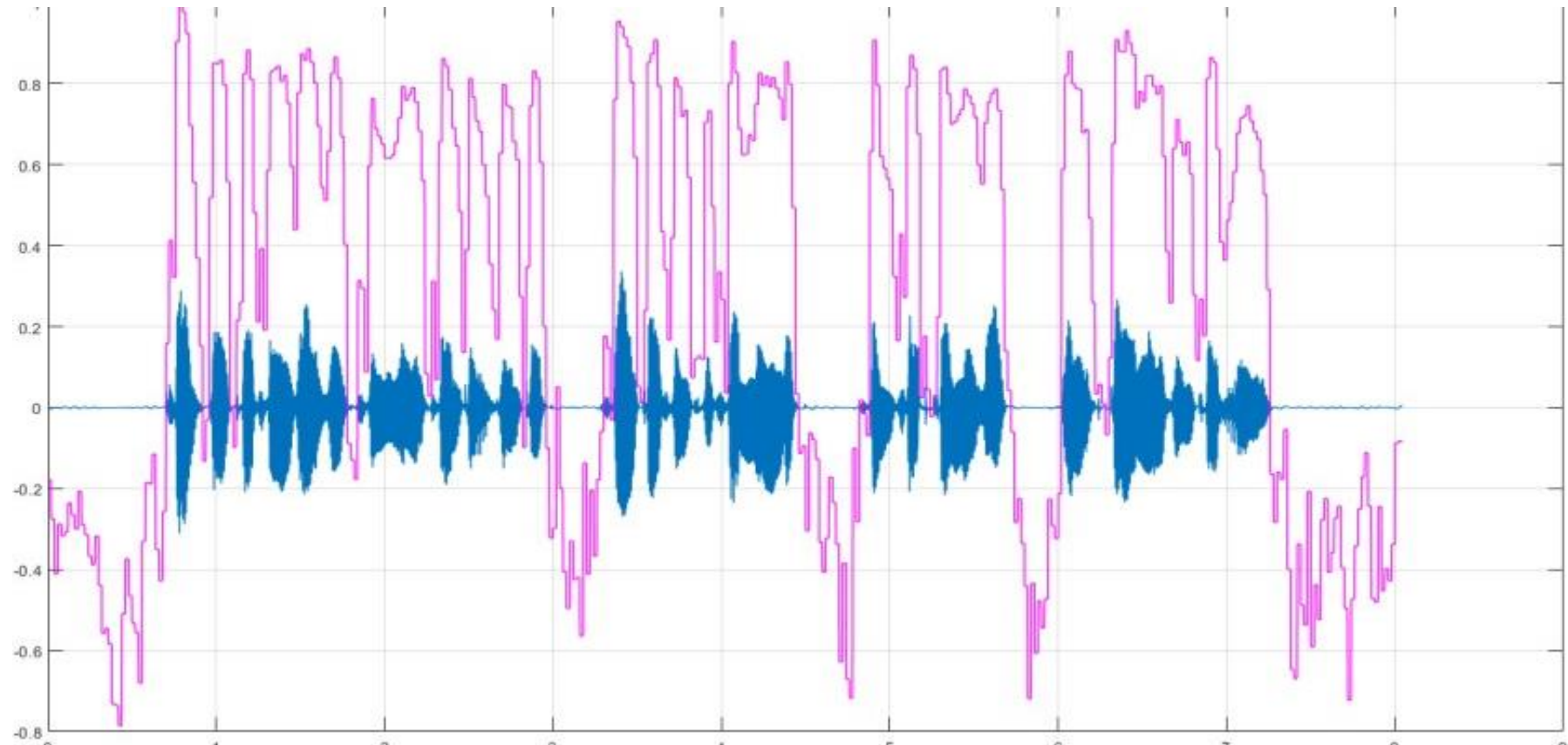


Vấn đề phát sinh

Muốn vẽ được tín hiệu, cường độ trung bình lên cùng 1 đồ thị thì yêu cầu độ dài các vector phải bằng nhau

Cách giải quyết

Tạo một vector mới có độ dài bằng độ dài vector tín hiệu



IV. THUẬT TOÁN KẾT HỢP GIỮA STE & ZCR

1. Short-Time Energy

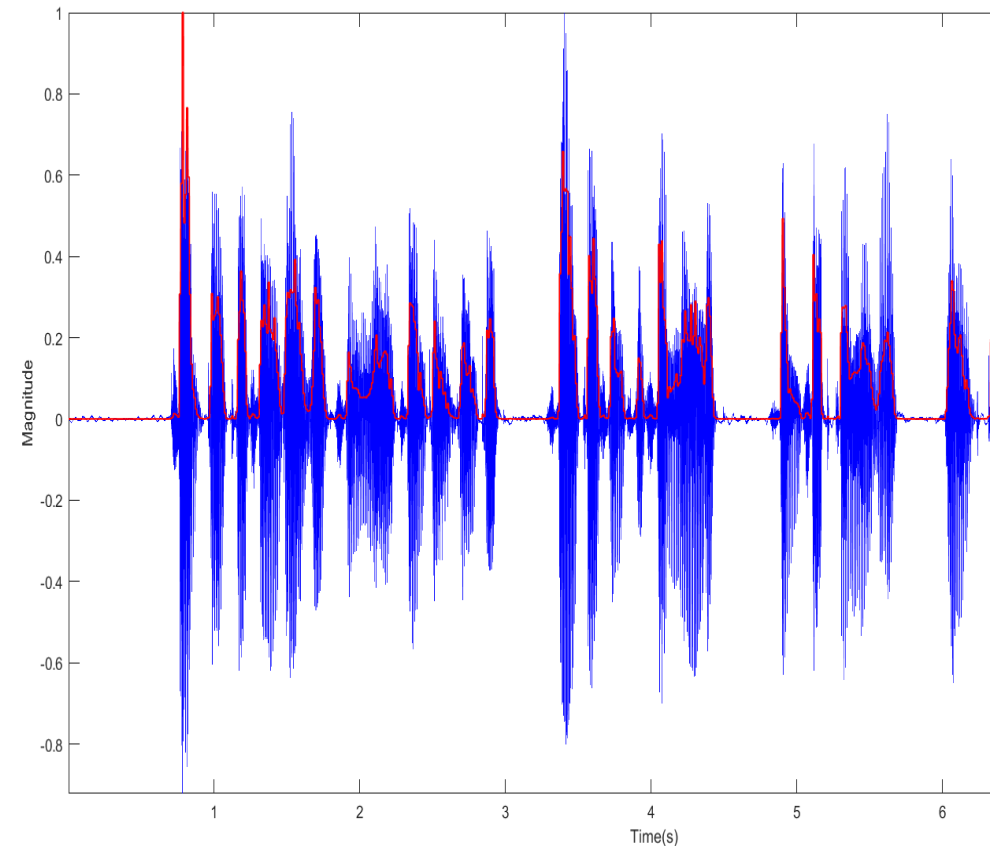
Short-Time Energy (năng lượng ngắn hạn) là năng lượng của tín hiệu khi xét trong một khoảng thời gian ngắn

Công thức toán học

$$E_n = \sum_{m=-\infty}^{+\infty} ((x[m] \times w[n-m]))^2$$

Trong đó : $w[n] = \begin{cases} \frac{1}{2N}, & 0 \leq x \leq N-1 \\ x, & \text{còn lại} \end{cases}$

Các âm hữu thanh thường có năng lượng lớn còn âm vô thanh thường có năng lượng nhỏ



2. Zero-Crossing Rate

Zero-crossing Rate(tốc độ băng qua 0) là tốc độ vượt qua điểm 0 của tín hiệu(từ dương sang 0 sang âm hoặc từ âm sang 0 sang dương)

Công thức toán học

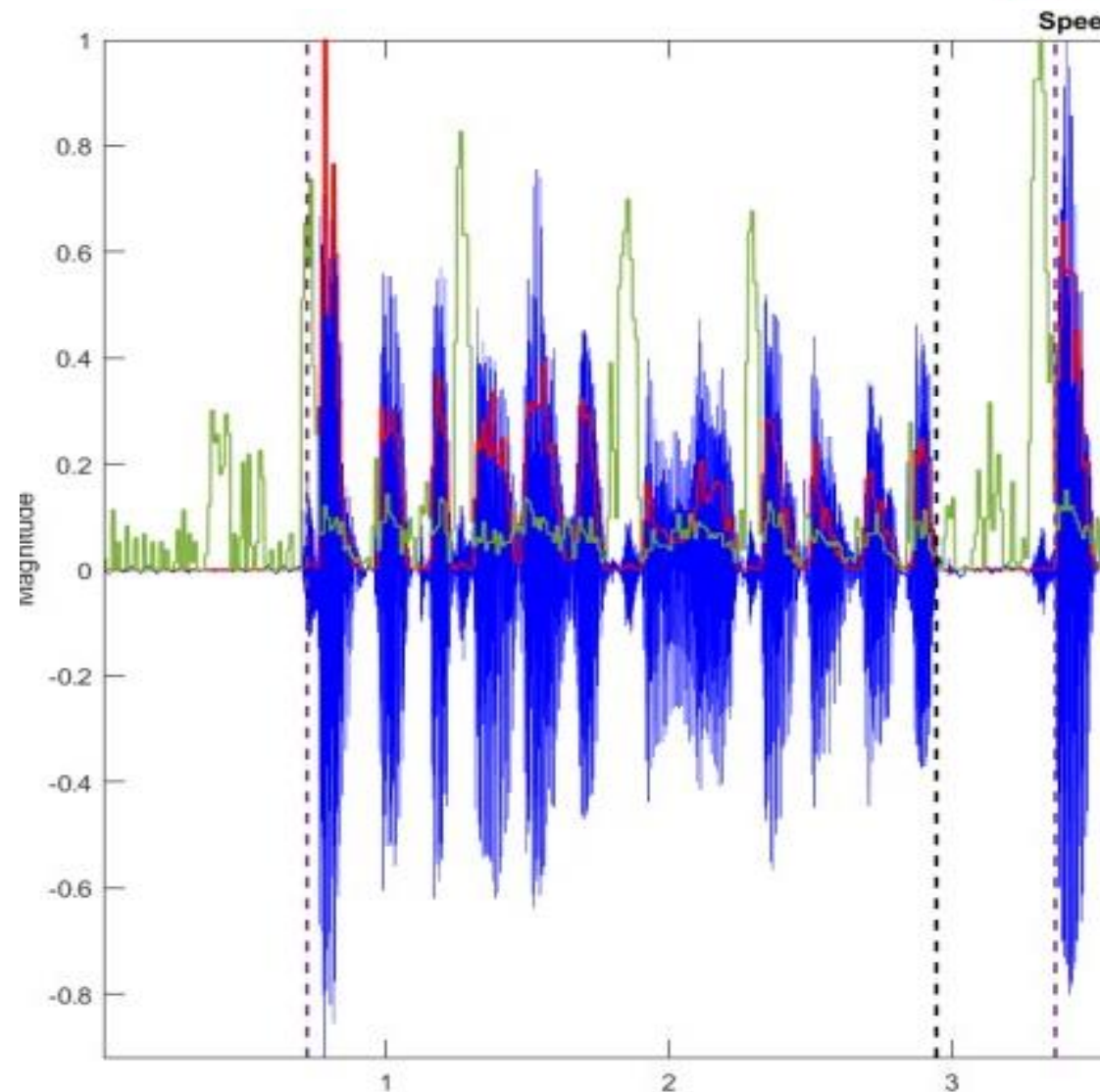
$$Z = \sum_{m=-\infty}^{+\infty} |\text{sgn}(x[m] - \text{sgn}(x[m-1])) \times w[n-m]|$$

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

$$w[n] = \begin{cases} \frac{1}{2N}, & 0 \leq x \leq N-1 \\ x, & \text{còn lại} \end{cases}$$

N: độ dài khung tín hiệu (mẫu)

Các âm hữu thanh có tốc độ băng qua 0 thấp còn các âm vô thanh có tốc độ băng qua 0 cao



CÁC THAM SỐ QUAN TRỌNG



Độ dài khung tín hiệu

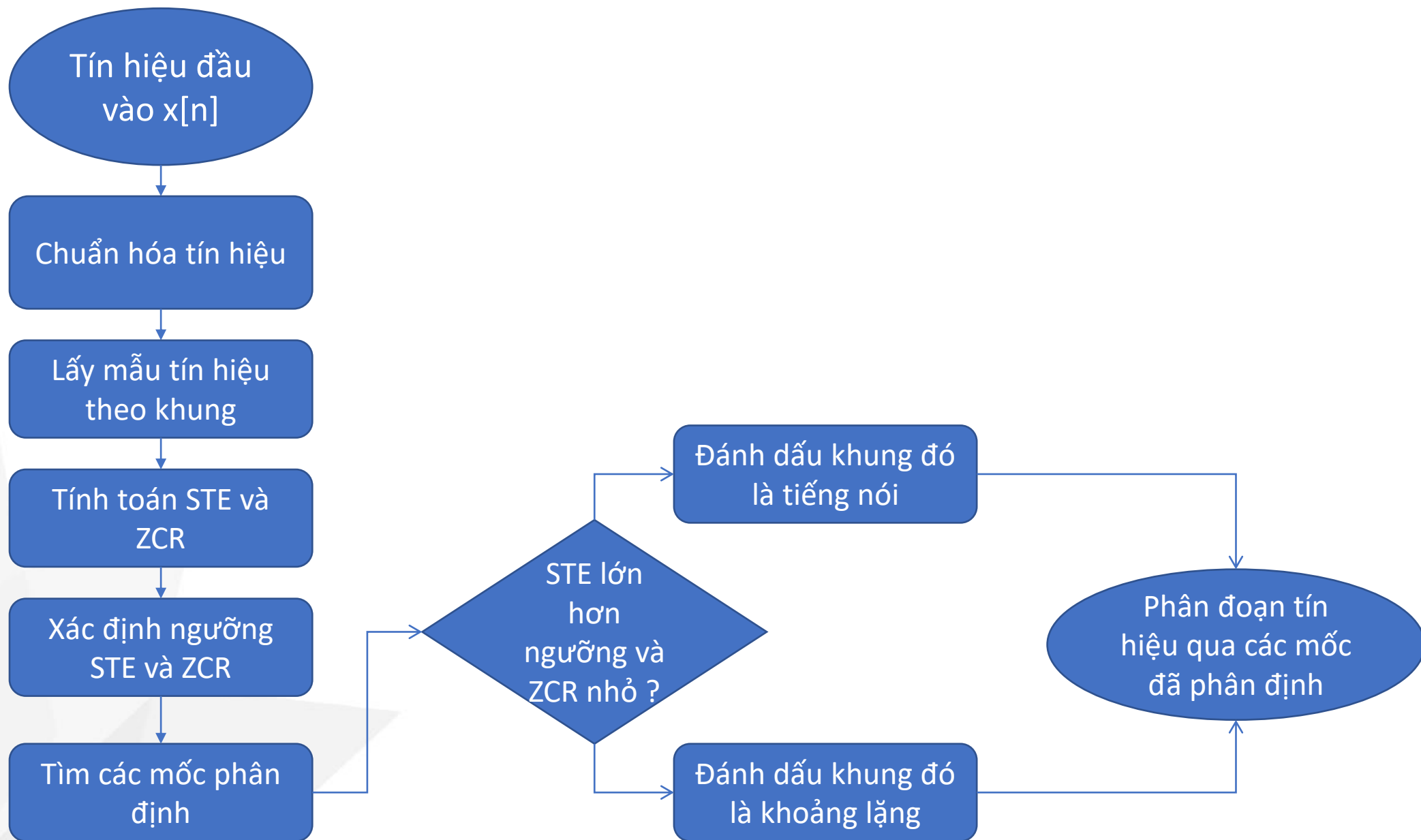
Độ dài khung tín hiệu không được quá ngắn không được quá dài

Tần số lấy mẫu

F_s càng lớn đặc trưng tín hiệu càng chính xác

Khoảng lặng tối thiểu

Lớn hơn 200 ms



VẤN ĐỀ PHÁT SINH VÀ CÁCH GIẢI QUYẾT

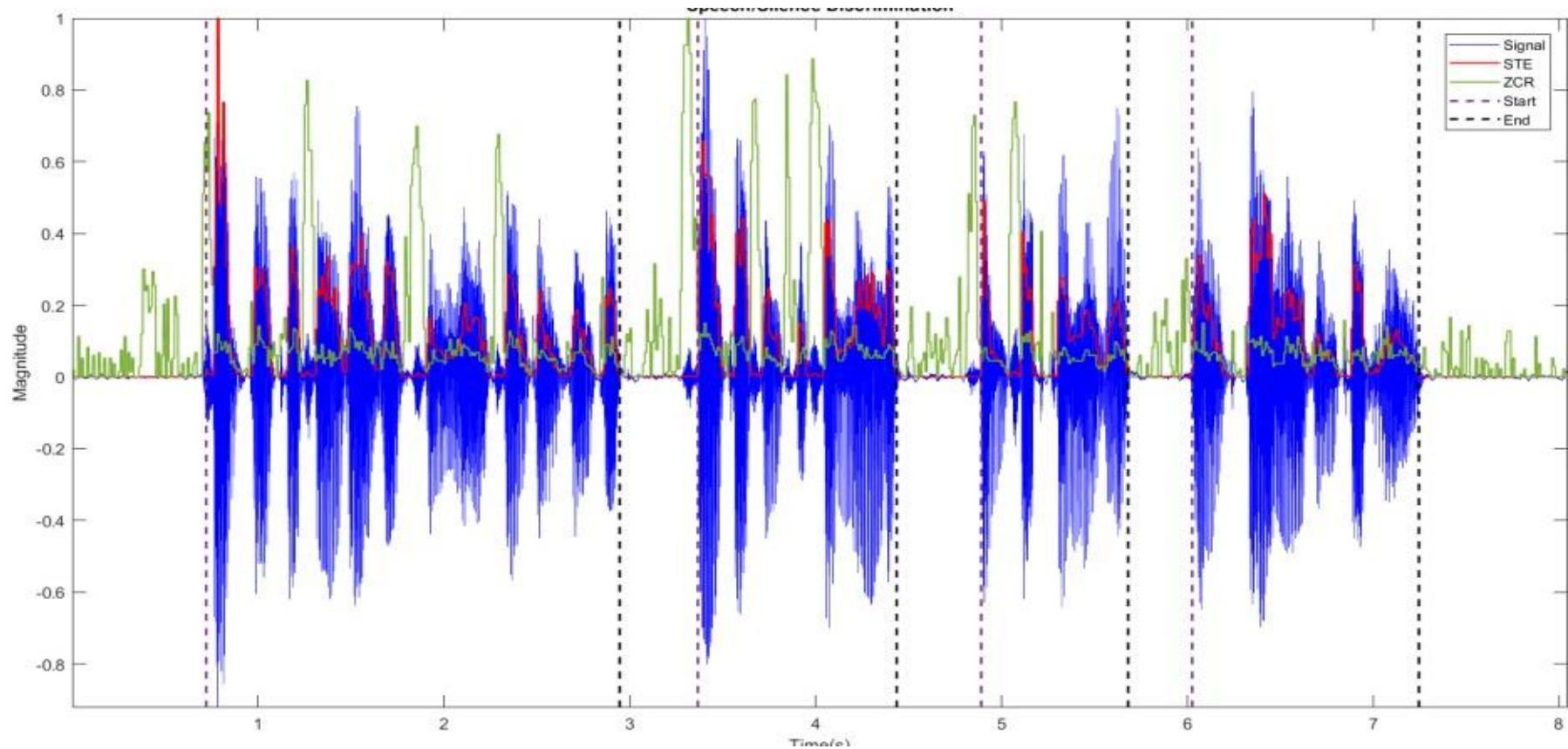


Vấn đề phát sinh

Muốn vẽ được tín hiệu, tốc độ băng qua 0, năng lượng lên cùng 1 đồ thị thì yêu cầu độ dài các vector phải bằng nhau

Cách giải quyết

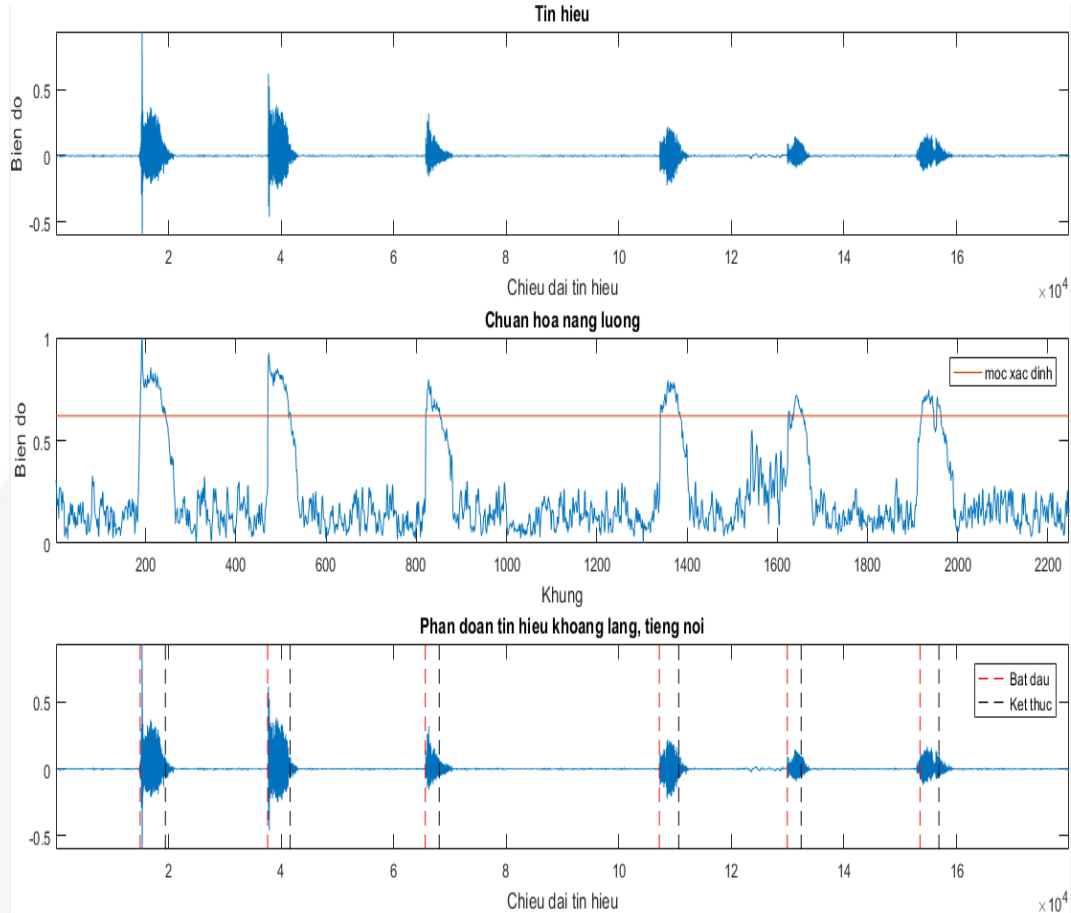
Tạo một vector mới có độ dài bằng độ dài vector tín hiệu



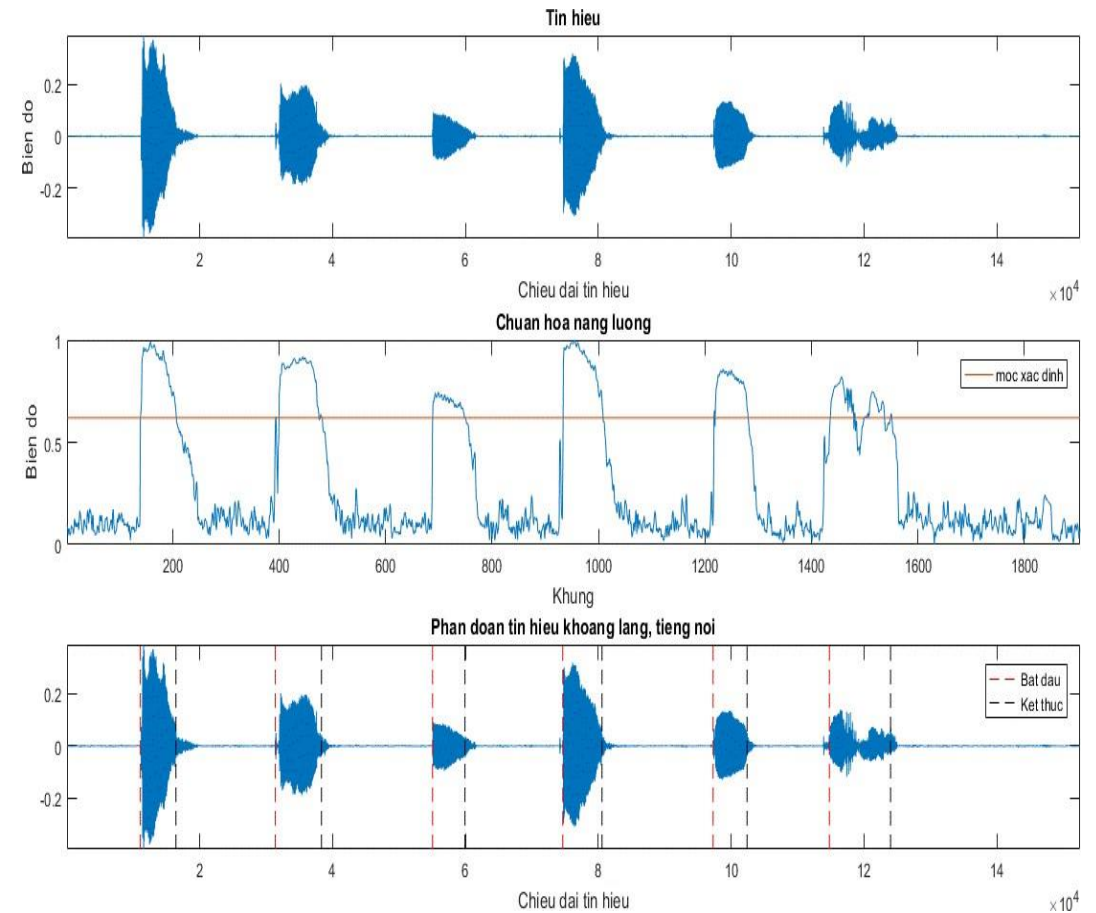


KẾT QUẢ THỰC NGHIỆM

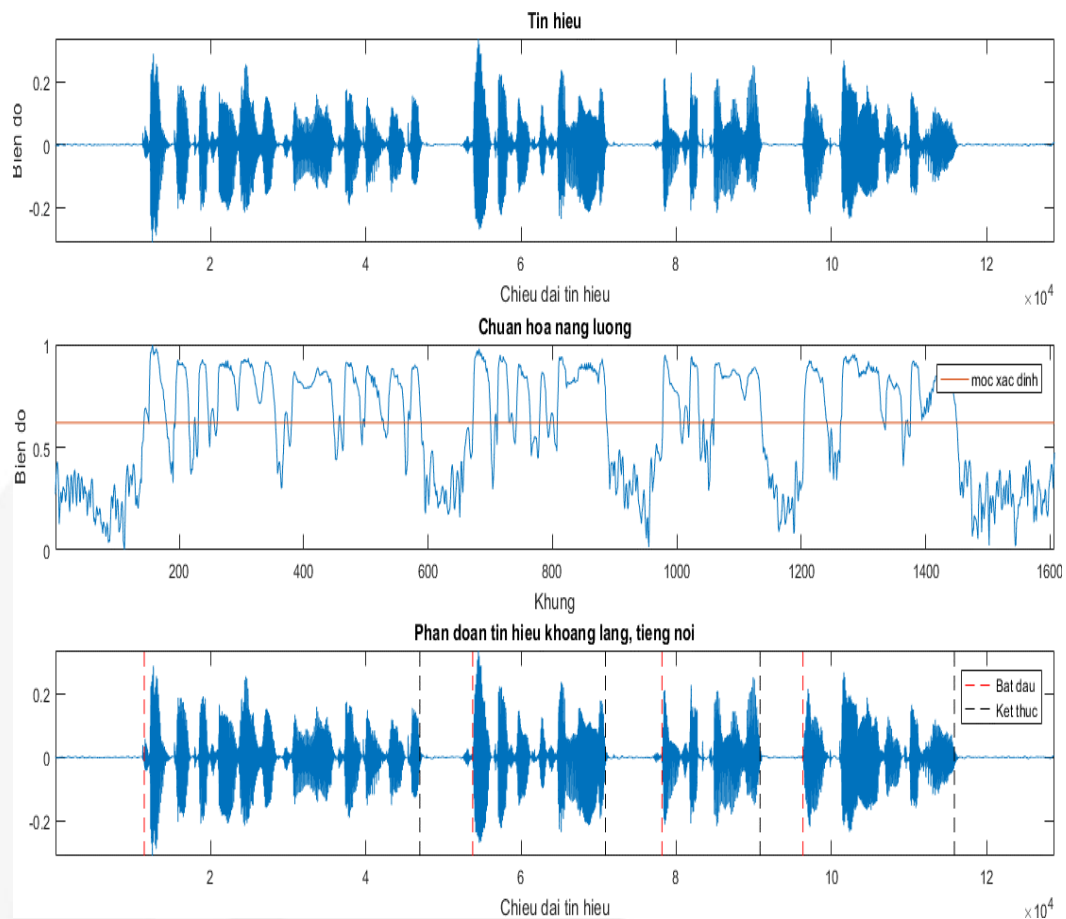
I. PHƯƠNG PHÁP CHUẨN HÓA NĂNG LƯỢNG 1



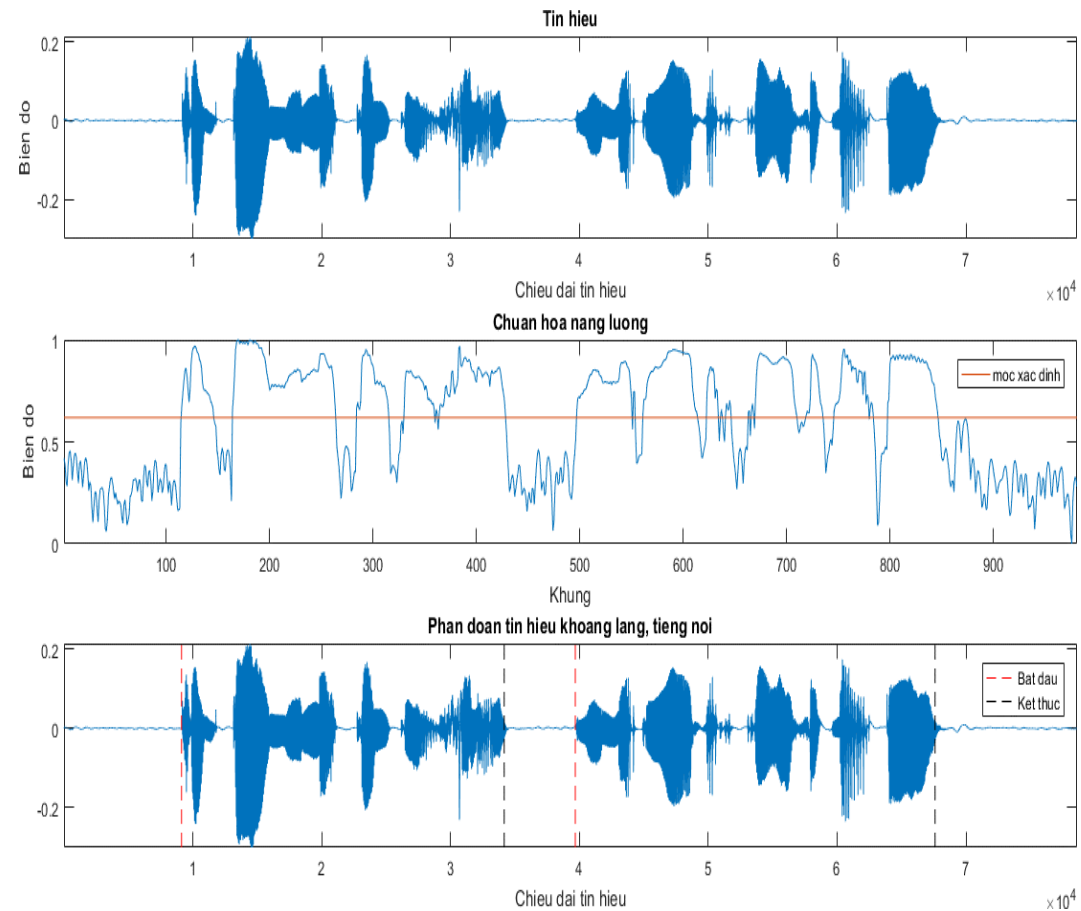
“Lab_male.wav” chia tự động



“Lab_female.wav” chia tự động



“Studio_male.wav” chia tự động

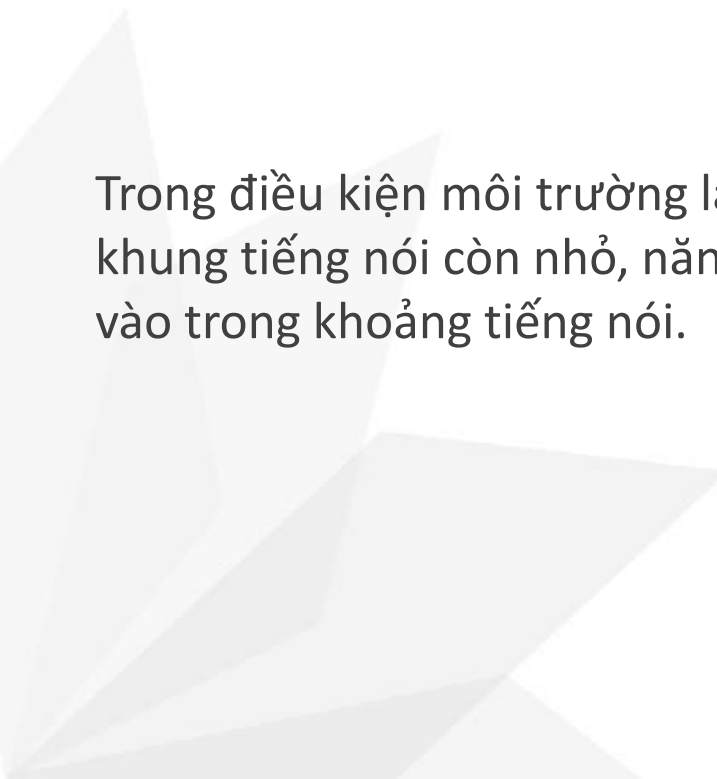


“Studio_female.wav” chia tự động



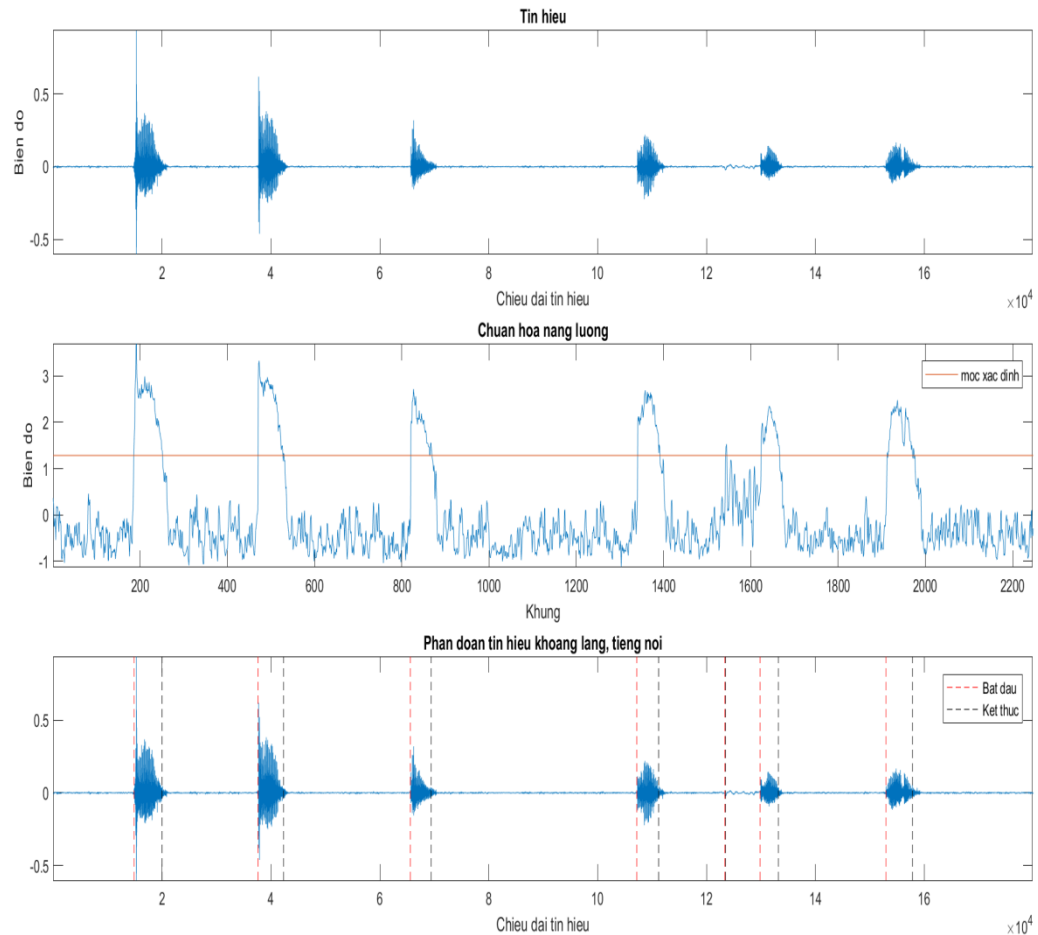
Nhận xét:

Trong điều kiện môi trường studio, thuật toán hoạt động hiệu quả, vì các tín hiệu ở môi trường này, năng lượng các khung tiếng nói có giá trị đồng đều gần sát cực đại, lớn hơn năng lượng nhiễu. Các biên xác định tương đối chính xác.

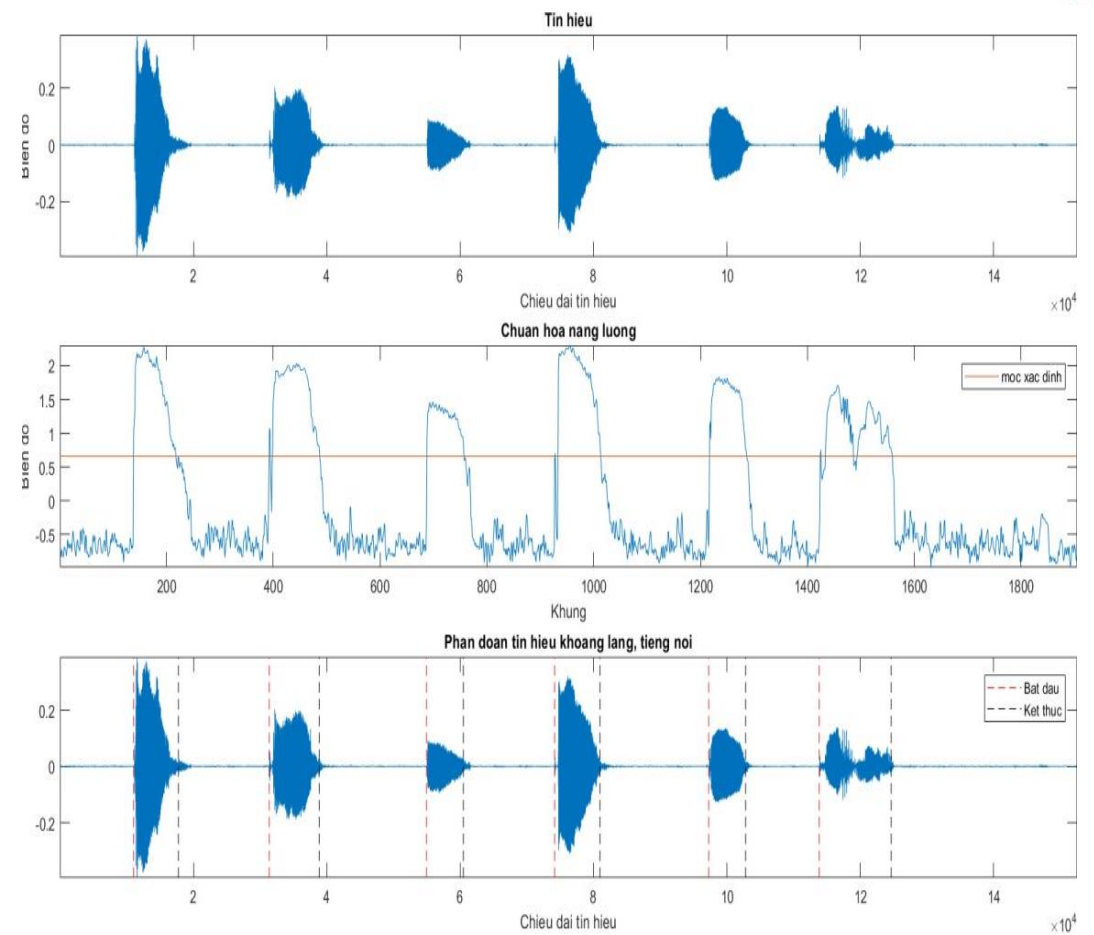


Trong điều kiện môi trường lab, thuật toán hoạt động chưa hiệu quả, vì năng lượng sau chuẩn hóa của các khung tiếng nói còn nhỏ, năng lượng 2 đầu biên các khoảng tiếng nói bé hơn ngưỡng. 2 đầu biên bị cắt sâu vào trong khoảng tiếng nói.

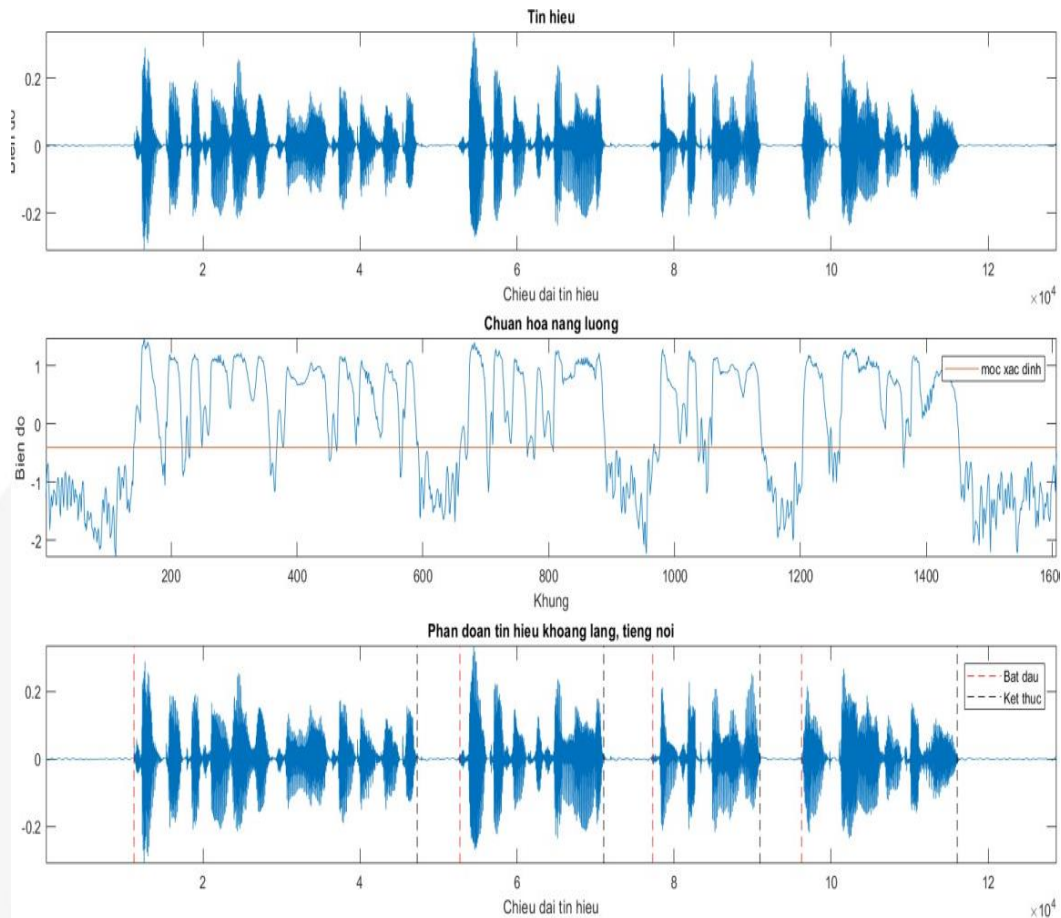
II. PHƯƠNG PHÁP CHUẨN HÓA NĂNG LƯỢNG 2



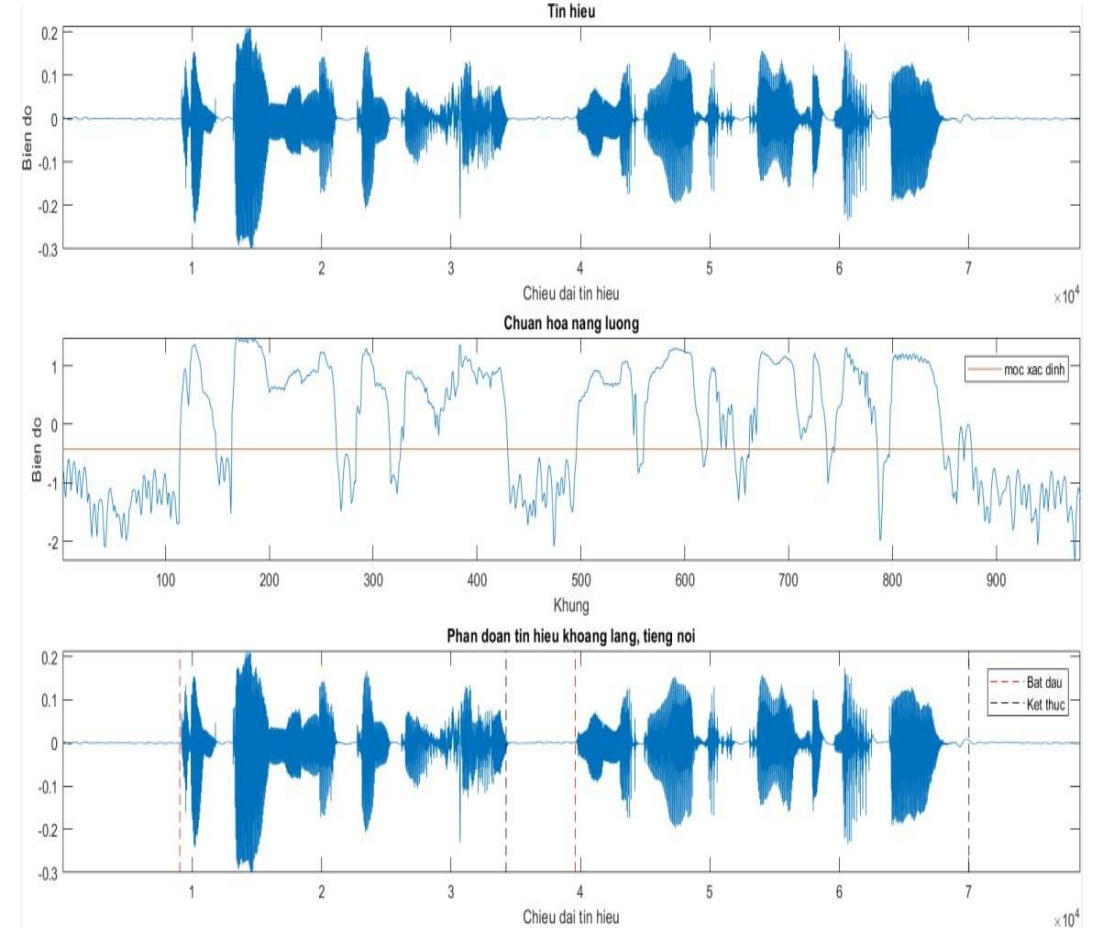
“Lab_male.wav” chia tự động



“Lab_female.wav” chia tự động



“Studio_male.wav” chia tự động



“Studio_female.wav” chia tự động

Nhận xét:

Tín hiệu trong môi trường "lab"

- Sau khi chuẩn hóa tín hiệu ta thấy ở tín hiệu "lab_male" độ nhiễu ở âm thứ 4 độ lớn xấp xỉ tiếng nói.
- Đối với tín hiệu lab_female độ lớn tiếng nói và khoảng lặng phân biệt rõ rệt

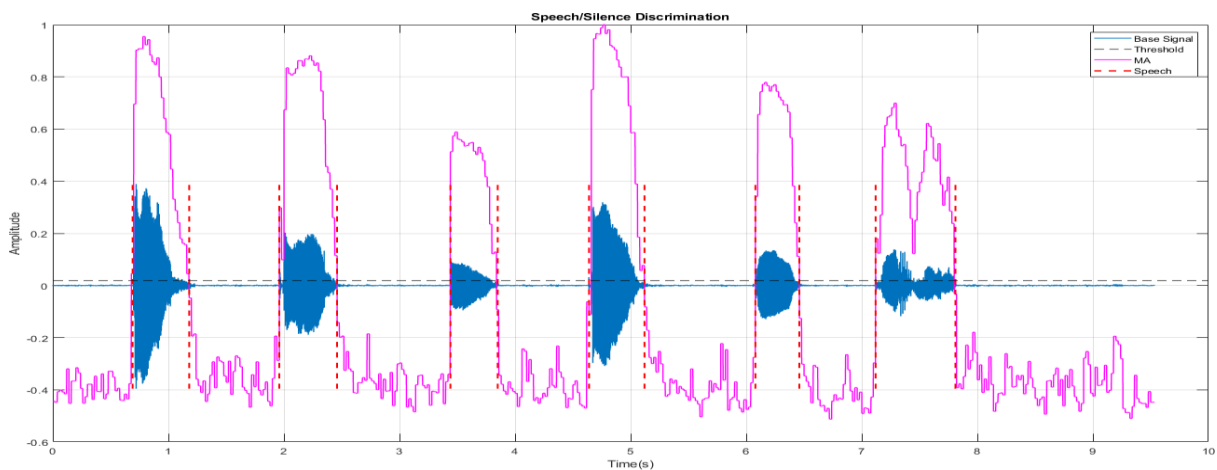
=> Độ chính xác thấp

Tín hiệu trong môi trường "studio"

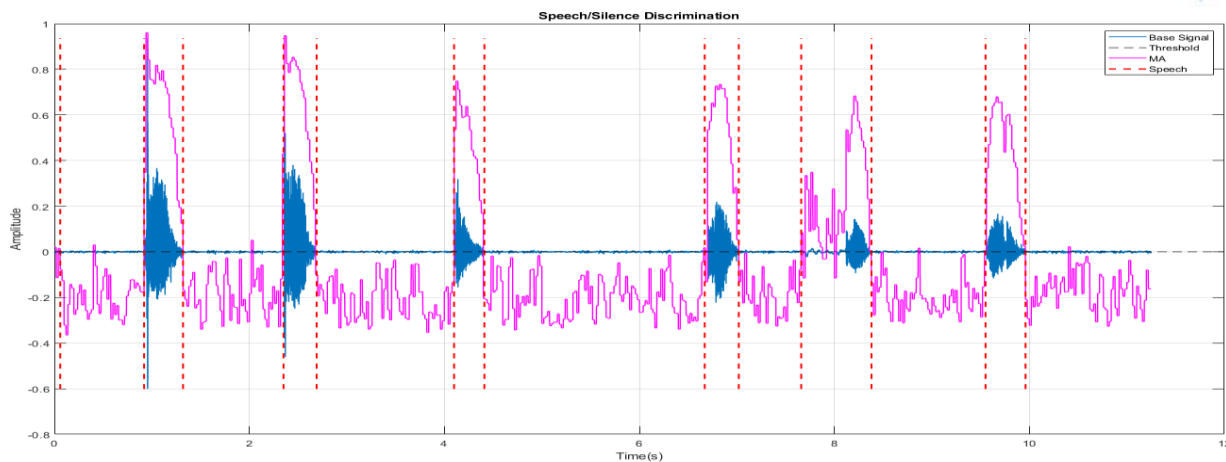
- Sau khi chuẩn hóa tín hiệu ta thấy các tín hiệu studio có độ phân biệt rõ rệt

=> Độ chính xác tương đối chính xác

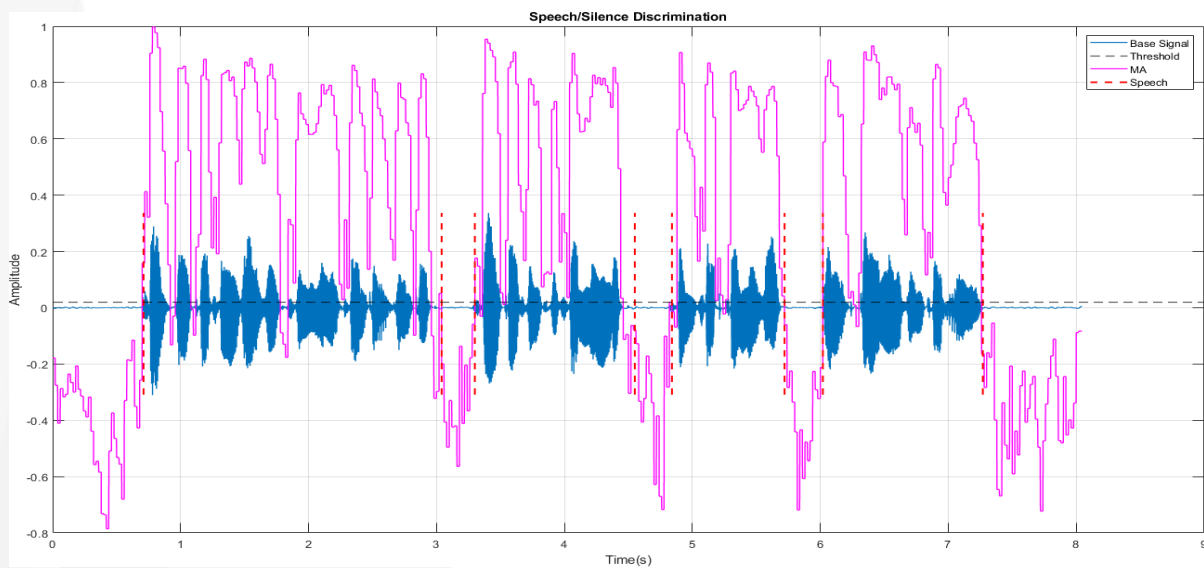
III. THUẬT TOÁN MEAN AVERAGE



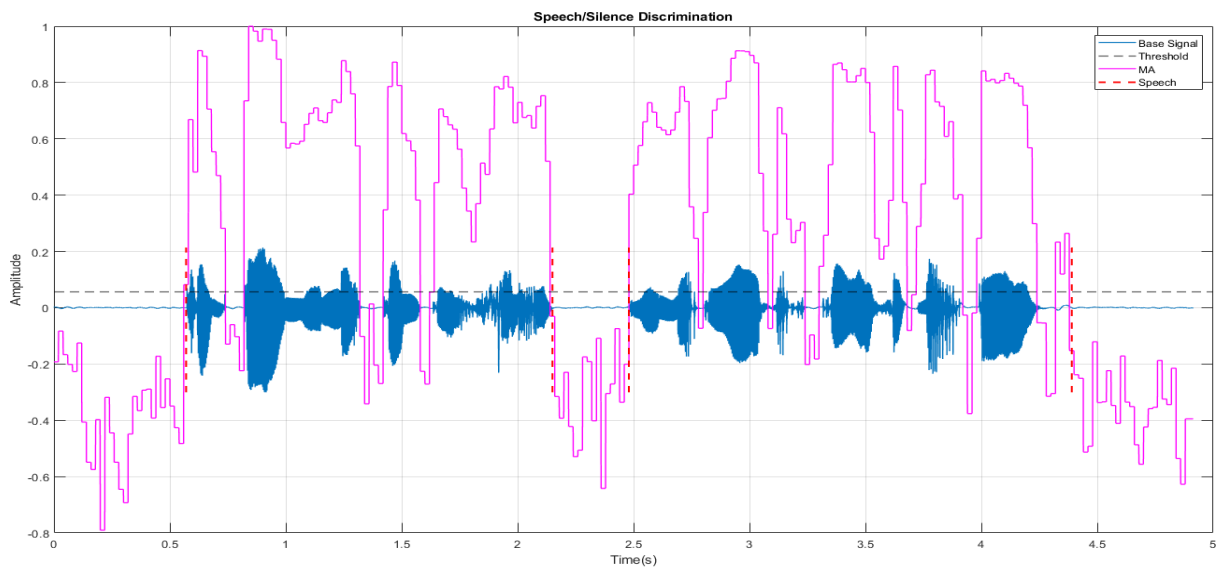
“Lab_male.wav” chia tự động



“Lab_female.wav” chia tự động



“Studio_male.wav” chia tự động



“Studio_female.wav” chia tự động

Nhận xét:

Ở 2 file âm thanh được thu ở môi trường “lab”

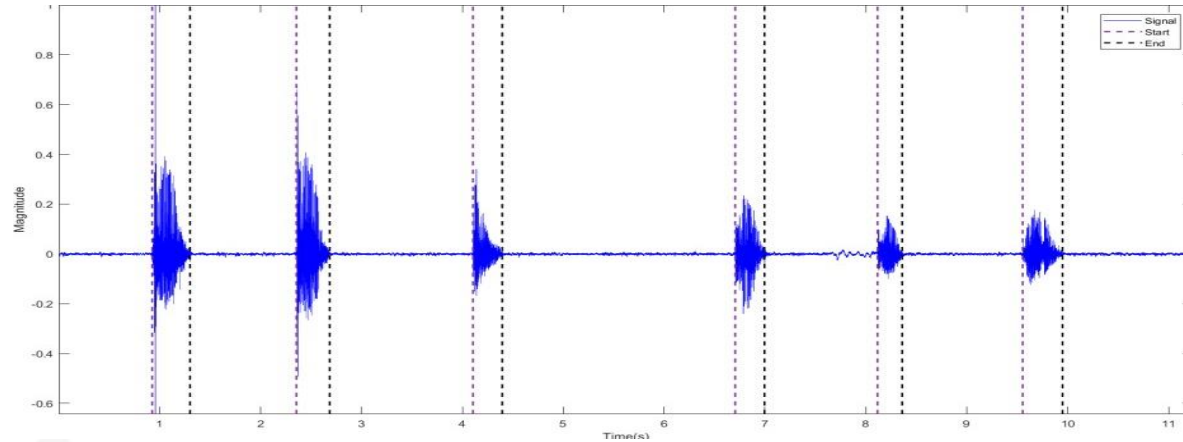
- Ở file “lab_male.wav”, tín hiệu tiếng nói được phân đoạn khá rõ ràng vì cường độ trung bình của nhiễu có giá trị phân biệt so với giọng nói
- Ở file “lab_female.wav”, trong tín hiệu tiếng nói được phân đoạn có lẫn cả tín hiệu của nhiễu ở bên trong vì cường độ trung bình của nhiễu có giá trị gần giống như giá trị của giọng nói

Ở 2 file âm thanh được thu ở môi trường “studio”

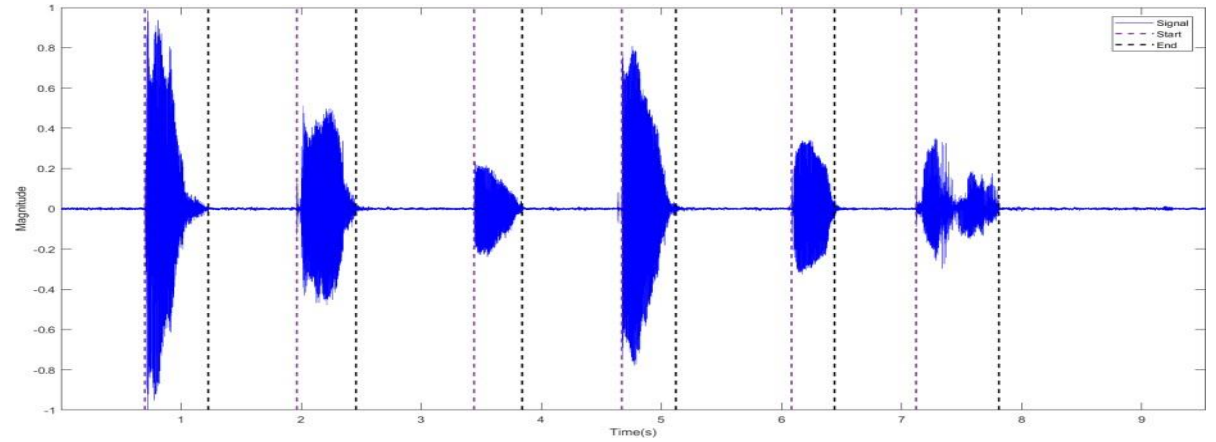
- Ở file “studio_male.wav”, tín hiệu tiếng nói được phân đoạn khá rõ ràng vì cường độ trung bình của nhiễu có giá trị phân biệt so với giọng nói
- Ở file “studio_female.wav” trong tín hiệu tiếng nói chỉ nhầm lẫn 1 chút phân đoạn nhiễu phần cuối của giọng nói vì cường độ trung bình của nhiễu khá lớn

=> Tín hiệu được thu ở môi trường “studio” thì thuật toán hoạt động chính xác hơn nhiều so với tín hiệu được thu ở môi trường “lab”

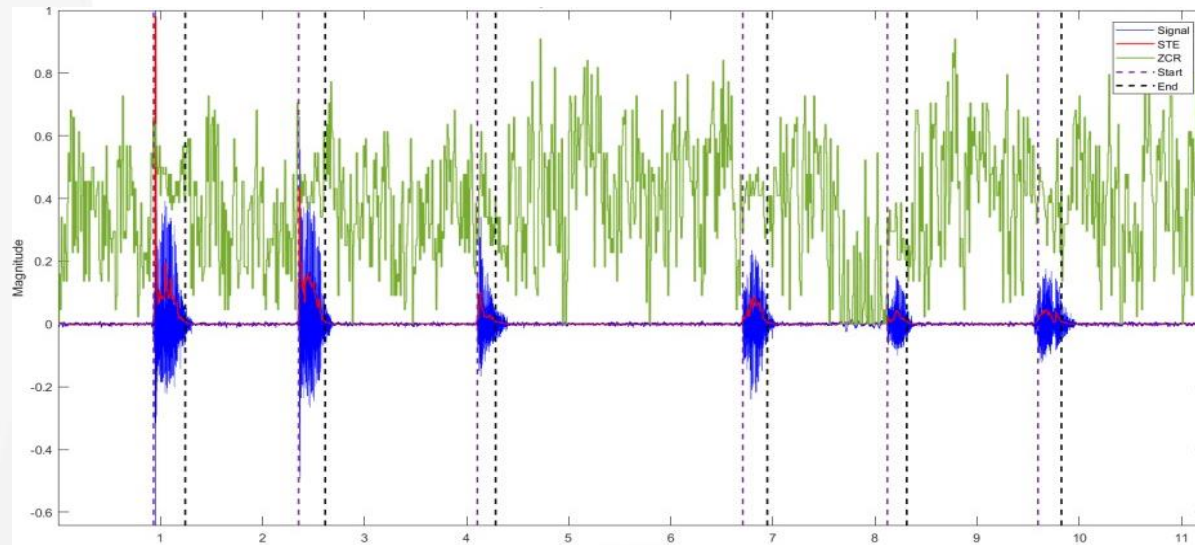
IV. THUẬT TOÁN STE & ZCR



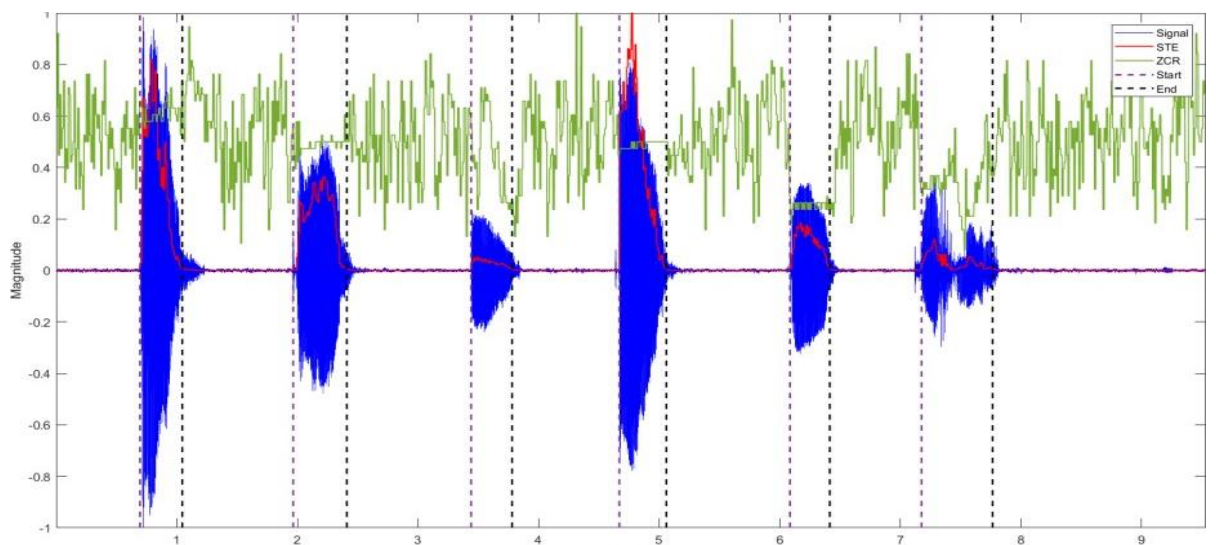
"Lab_male.wav" chia thủ công



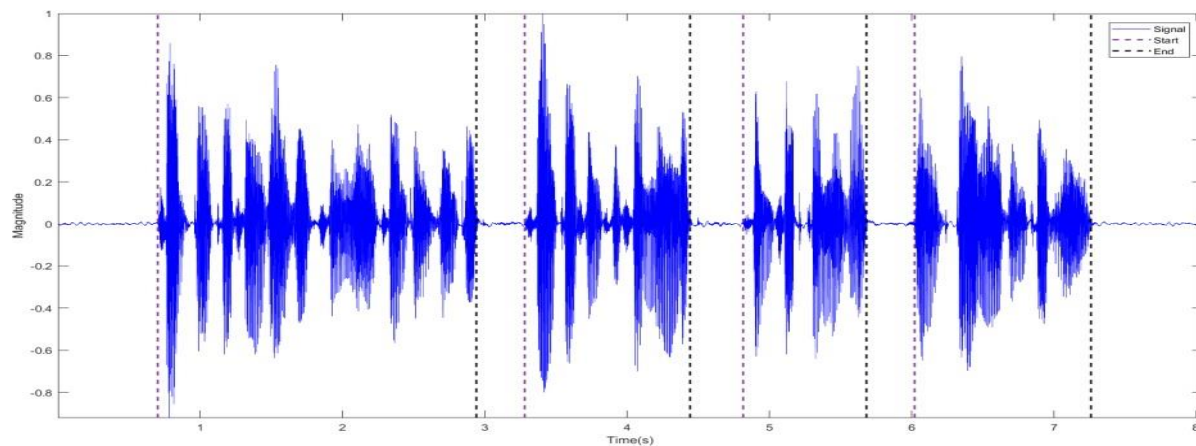
"Lab_female.wav" chia thủ công



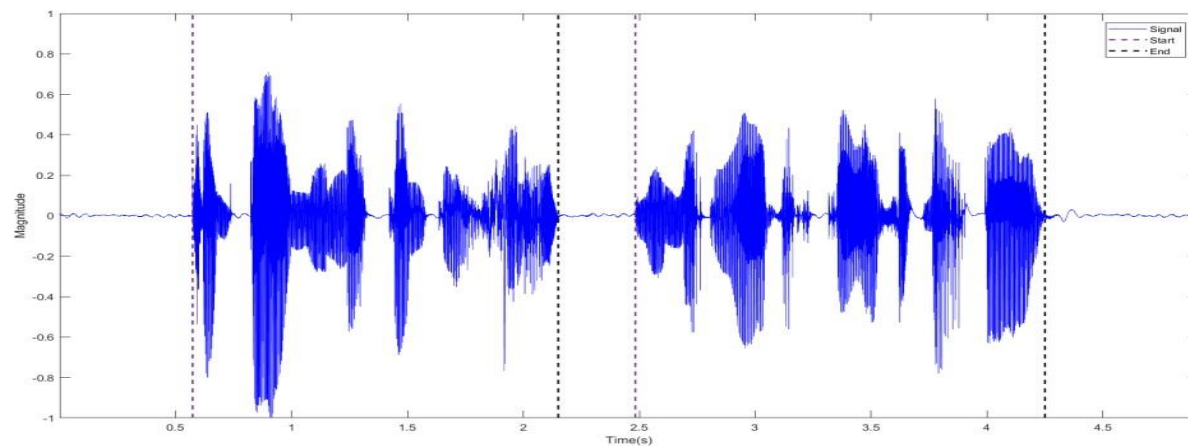
"Studio_male.wav" chia tự động



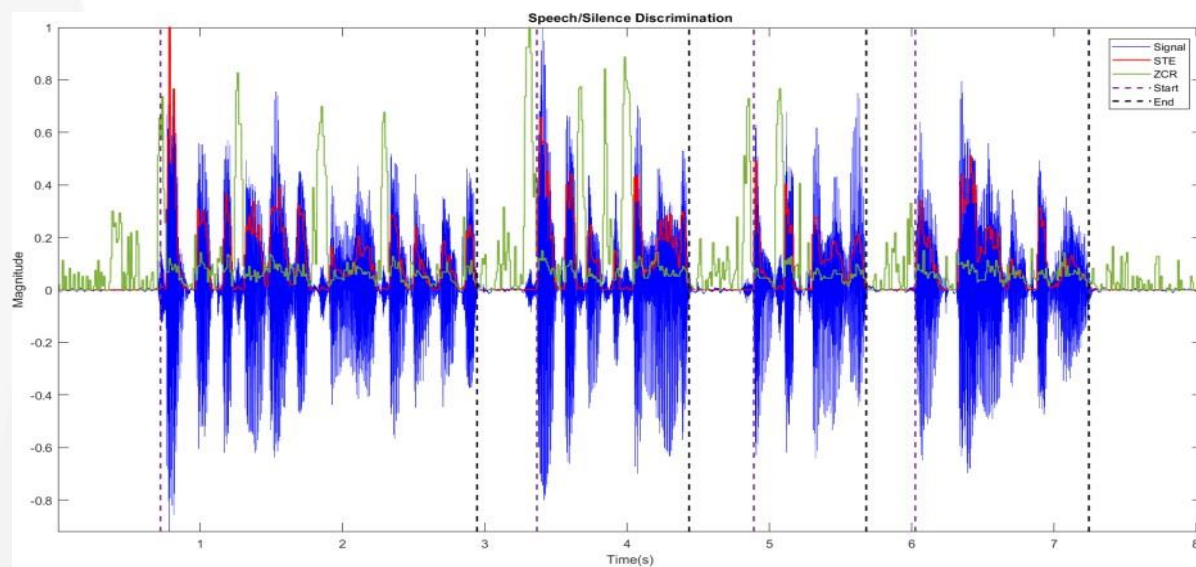
"Studio_female.wav" chia tự động



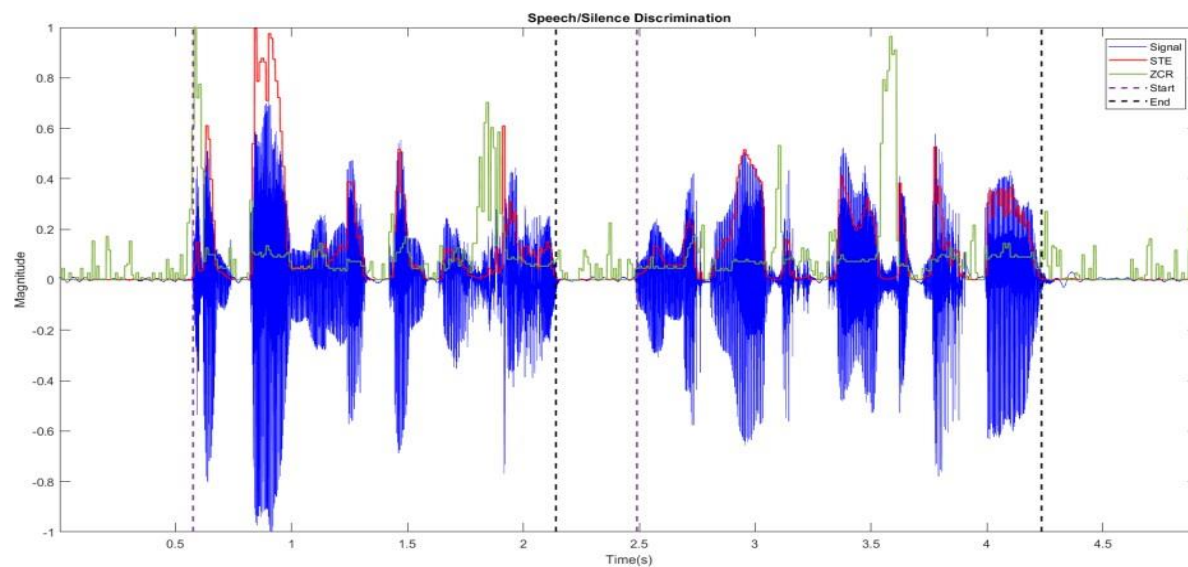
“Studio_male.wav” chia thủ công



“Studio_female.wav” chia thủ công



“Studio_male.wav” chia tự động



“Studio_female.wav” chia tự động



Nhận xét:

Đối với hai file “lab_male.wav ” và “lab_female.wav”:

- Năng lượng ở đoạn đầu tiếng nói cao hơn hẳn so với khoảng lặng nhưng ở vị trí cuối tiếng nói chuẩn bị chuyển qua khoảng lặng thì năng lượng khá thấp
- Tốc độ băng qua 0 không có sự khác nhau rõ rệt giữa tiếng nói và khoảng lặng

Đối với hai file “studio_male.wav” và “studio_female.wav”:

- Năng lượng ở đoạn tiếng nói vẫn cao hơn hẳn khoảng lặng, nhưng ở vị trí bắt đầu tiếng nói thứ 2 và 3 của file “studio_male.wav” năng lượng lại khá nhỏ.
- Tốc độ băng qua 0 giữa tiếng nói và khoảng lặng có sự khác nhau rõ ràng hơn so với hai file trên.

=>Ta thấy trong môi trường studio thuật toán hoạt động hiệu quả hơn.



KẾT LUẬN



Thông qua bài tập nhóm này, chúng em đã có được những kiến thức cơ bản trong lĩnh vực xử lý tín hiệu tiếng nói và âm thanh. Có thể phân đoạn tiếng nói và khoảng lặng tương ứng tương đối chính xác thông qua các phương pháp khác nhau dựa trên việc phân tích các đặc trưng của tín hiệu âm thanh và đã phát triển được các kỹ năng hữu ích như làm việc nhóm, tìm hiểu tài liệu, lập trình Matlab,... Song, chúng em vẫn còn thiếu nhiều kinh nghiệm nên vẫn còn 1 số sai sót nhưng chúng em sẽ cố gắng cải thiện trong tương lai



Thank you