

BÁO CÁO XỬ LÝ TÍN HIỆU SỐ

PHÂN ĐOẠN TÍN HIỆU THÀNH TIẾNG NÓI VÀ KHOẢNG LẶNG SỬ DỤNG HÀM NĂNG LƯỢNG VÀ HÀM ZERO – CROSSING – RATE

Lê Anh Huy, Nguyễn Anh Huy, Nguyễn Đức Quốc Huy, Lê Bá Hải Đăng

Nhóm 5 Lớp HP: 17N14

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)
	Lê Anh Huy	Đọc tài liệu, thiết kế thuật toán, viết chương trình, sửa lỗi chương trình, báo cáo tổng quát bài tập, phân công công việc nhóm.	
	Nguyễn Anh Huy	Đọc tài liệu, nghiên cứu, viết hàm năng lượng, kiểm – sửa lỗi chương trình, báo cáo về phần năng lượng.	
	Nguyễn Đức Quốc Huy	Đọc tài liệu, nghiên cứu, viết hàm Zero – crossing – rate (ZCR) viết báo cáo file word, báo cáo về phần ZCR.	
	Lê Bá Hải Đăng	Viết báo cáo file word, làm slide, Đọc tài liệu, nghiên cứu chuẩn hóa biểu đồ.	

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: Phân biệt ‘Tiếng nói’ và ‘Khoảng lặng’ của tín hiệu là bài toán cần thiết trong xử lý tín hiệu âm thanh. Bài báo cáo này thực hiện việc cài đặt thuật toán tìm năng lượng và tỉ lệ đi qua điểm không của tín hiệu tiếng nói trên miền thời gian. Tại các môi trường khác nhau thì khả năng phân biệt tiếng nói với khoảng lặng khác nhau. Và môi trường trong phòng thu ‘Studio’ là môi trường có thể phân biệt rõ tiếng nói và khoảng lặng nhất vì nó là môi kín và ít tạp âm nhất.

Từ khóa – Năng lượng tín hiệu, phân biệt tiếng nói và khoảng lặng, ZCR, matlab;

Đặt vấn đề.

Nhận dạng phân biệt tiếng nói và khoảng lặng là một phần trong hệ thống phân loại âm thanh ASC (Audio Signal Classifier) [1]. Có rất nhiều cách để phân biệt âm thanh và khoảng lặng: chuỗi thay thế, băng thông, tỉ lệ vượt điểm không (Zero – Crossing – Rate: ZCR), tần số cơ bản, năng lượng tín hiệu,... Trong bài báo cáo này, chúng tôi lựa chọn hai đặc trưng: Tỉ lệ vượt điểm không (ZCR) và Năng lượng tín hiệu để giải quyết vấn đề trên.

Các thuật toán tìm ZCR cũng như là năng lượng tín hiệu thường chỉ cho ra ma trận kết quả là những con số phức tạp, dài dòng, điều này dẫn đến khó khăn trong việc nhận biết, tổng quát các thông số trên miền thời gian rời rạc.

Để giải quyết vấn đề trên, chúng tôi xử lý biểu đồ về Năng lượng – ZCR rồi chuẩn hóa chúng về trên cùng một biểu đồ giúp người đọc dễ dàng phân biệt tiếng nói và khoảng lặng hơn.

I. Lý thuyết xử lý tín hiệu tiếng nói và các thuật toán.

Hàm tính năng lượng tín hiệu

1. Cơ sở lý thuyết.

Hàm tính năng lượng của tín hiệu rời rạc được định nghĩa như sau:

$$E = \sum_{n=-\infty}^{\infty} |x[n]|^2 \quad [2]$$

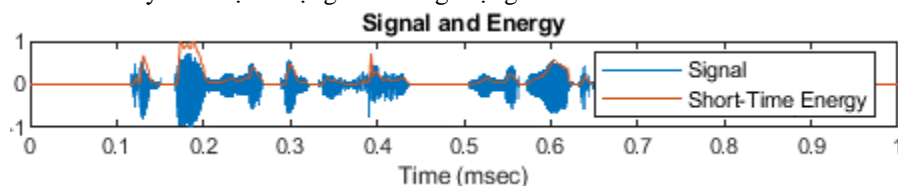
Trong đó: E là năng lượng tín hiệu trong miền thời gian của tín hiệu 'x[n]'.

Năng lượng của tín hiệu có độ dài theo độ dài frame:

$$E = \sum_{m=-\infty}^{\infty} [x(m)w(n-m)]^2 \quad [4]$$

Hàm tính năng lượng dùng để xác định năng lượng của cả tín hiệu tại từng thời điểm rời rạc. Nói rõ hơn: tại những thời điểm có tiếng nói thì có năng lượng cao hơn những thời điểm là khoảng lặng: $E_{tn} \gg E_{kl}$. [2]

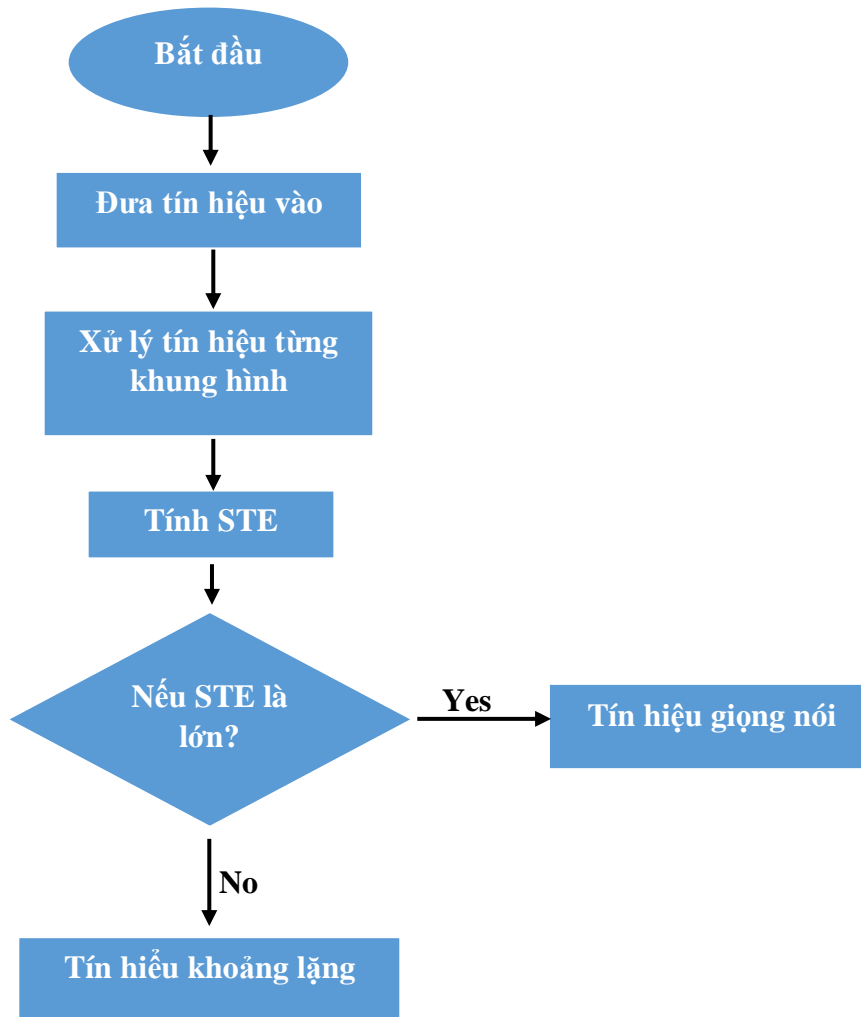
Dưới đây là ví dụ sử dụng hàm năng lượng:



Hình 1. Tín hiệu âm u và hàm tự tương quan của nó

2. Sơ đồ thuật toán.

Từ định nghĩa hàm năng lượng, ta có lưu đồ thuật toán như [Hình 2 - ở trang sau]:



Hình 2: Sơ đồ khối thuật toán hàm tính năng lượng STE

3. Đánh giá thuật toán.

Ưu điểm: Nhanh chóng xác định được giá trị năng lượng tại các thời điểm trong miền thời gian rời rạc, tổng quát hóa được chỉ số năng lượng của tín hiệu đầu vào.

Nhược điểm: Thuật toán còn chưa lấy được hết dữ liệu của tín hiệu. Trong trường hợp thời gian của một khung quá nhỏ thì sẽ không cung cấp đủ các giá trị trung bình từ đó việc biểu diễn hàm năng lượng còn hạn chế

4. Đề xuất giải pháp.

Sử dụng overlap và cửa sổ hình chữ nhật hoặc cửa sổ harming.

A. Zero – Crossing – Rate.

1. Cơ sở lý thuyết.

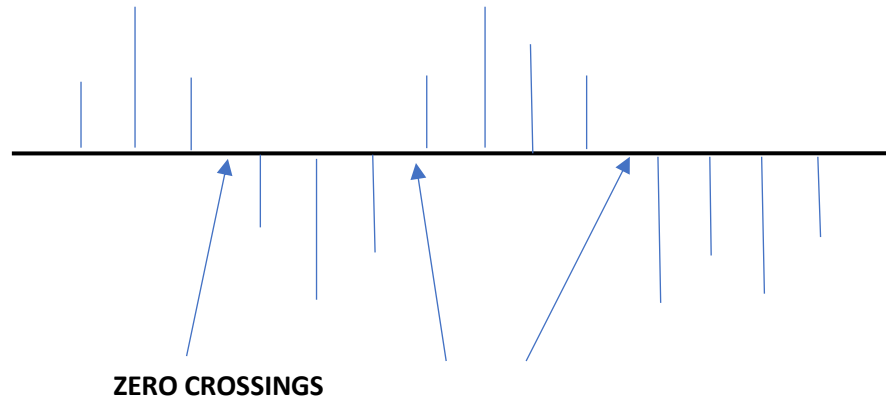
Zero – Crossing – Rate (ZCR) của 1 khung tín hiệu là tốc độ tín hiệu thay đổi dấu hiệu của nó trong khung. Nó biểu thị số lần tín hiệu thay đổi giá trị, từ dương sang âm và ngược lại. [3]

ZCR có thể được hiểu là thước đo “độ ồn” của tín hiệu. Ví dụ nó thường thể hiện các giá trị cao hơn trong trường hợp tín hiệu nhiễu. Nó cũng được dùng để phản ánh các đặc tính phổ

của tín hiệu, do đó ZCR dẫn đến việc áp dụng vào nhiều ứng dụng, bao gồm phân biệt âm nhạc, phát hiện giọng nói và phân loại thể loại âm nhạc. [3]

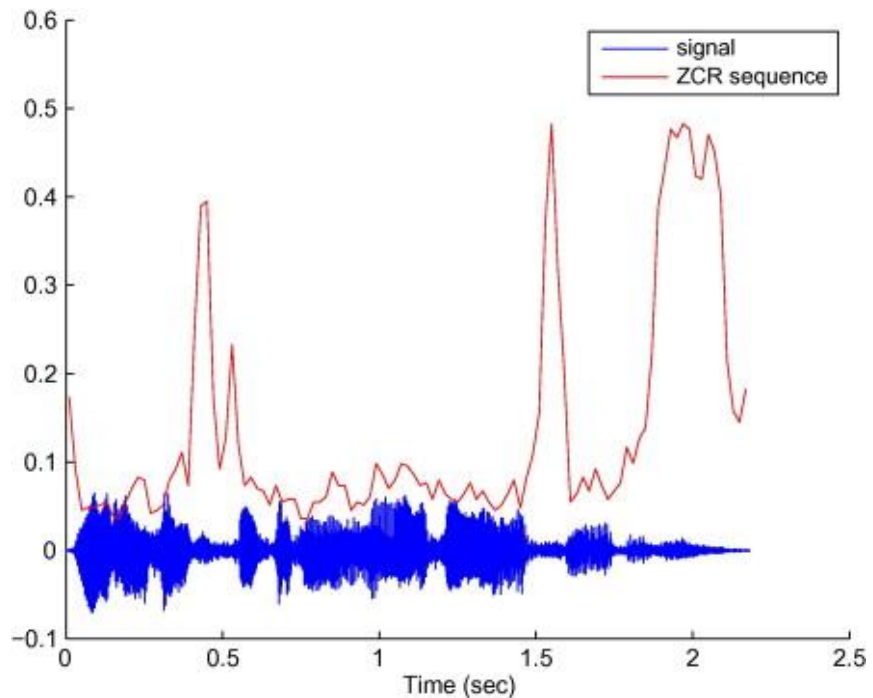
Nó là bước trung gian đơn giản để có được biểu diễn miền tần số của lời nói. Cách tiếp cận này liên quan đến việc lọc băng thông của tín hiệu giọng nói trong một số dải tần số liên kế nhau. Sau đó thu được cho các đầu ra bộ lọc. Các biểu diễn này cùng nhau đưa ra một biểu diễn phản ánh chính xác các tính chất phổ của tín hiệu. [4]

Hình ảnh biểu diễn ZCR:



Hình 3: Hình ảnh minh họa cho các đường ZCR.

Ví dụ: Hình bên trình bày tín hiệu giọng nói cùng với chuỗi ZCR tương ứng. Ta thấy rằng các giá trị của ZCR cao hơn đối với các phần nhiều của tín hiệu, trong khi trong các khung giọng nói, các giá trị ZCR tương ứng thường thấp hơn.



Hình 4: Đường biểu diễn của ZCR và tín hiệu. (Nguồn Internet)

ZCR được định nghĩa theo phương trình sau:

$$Z_n = \sum_{m=-\infty}^{\infty} |\text{sgn}[x(m)] - \text{sgn}[x(m-1)]|w(n-m) \quad (2.1)$$

Trong đó: $\text{Sgn}[x(n)] = 1 \quad x(n) \geq 0$

$$= -1 \quad x(n) < 0$$

Và $w(n)$ là hàm window với kích thước window là N mẫu:

$$w(n) = \frac{1}{2N} \quad 0 \leq n \leq N-1$$
$$= 0$$

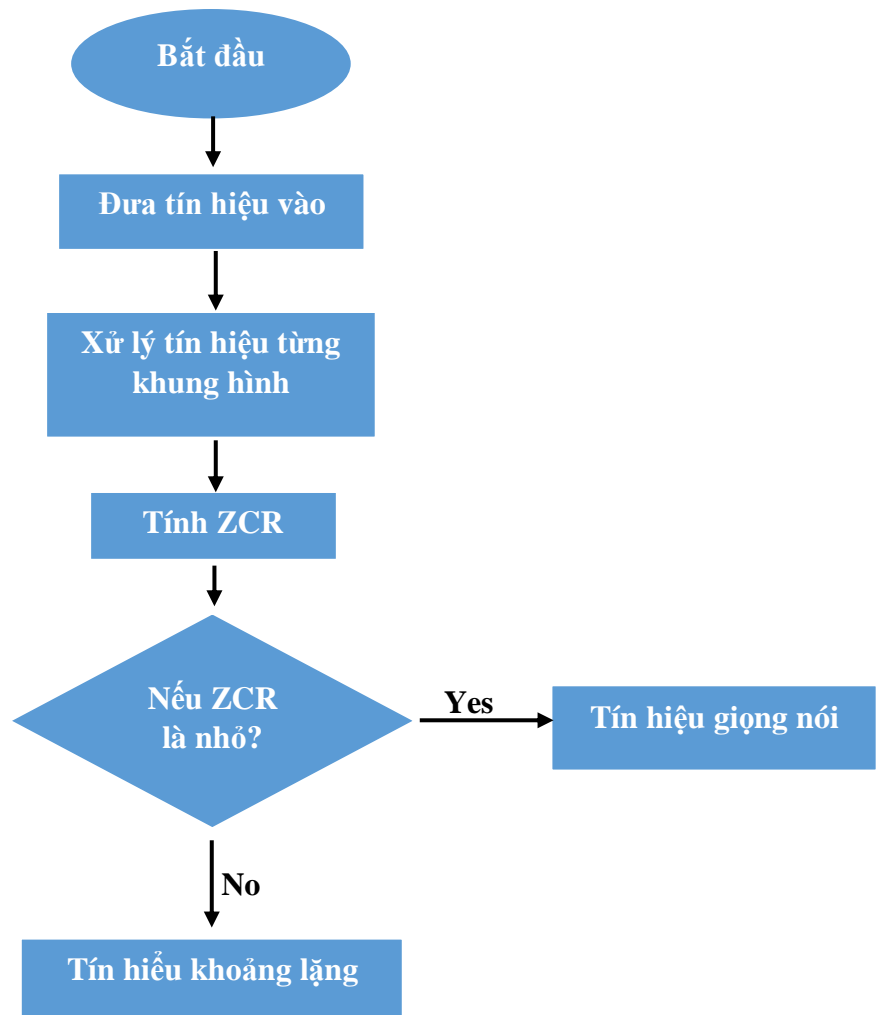
2. Ý tưởng thuật toán.

Dựa vào định nghĩa trên, chúng ta có thể suy ra được 02 cách tính ZCR bằng Matlab:

- Chuyển toàn bộ mọi giá trị sample sang 1 (nếu sample có giá trị dương) và -1 nếu sample có giá trị âm), từ đó đưa vào mảng SIGNAL, áp dụng công thức (2.1) ta tính được ZCR.
- So sánh dấu của tích 2 sample kế tiếp nhau của tín hiệu, nếu trái dấu là một điểm ZCR, nếu cùng dấu thì không phải. Lưu vào trong Frame, gộp tất cả Frame lại dùng hàm Plot ta có đồ thị ZCR.

(Đã chú thích trong file code Matlab)

3. Sơ đồ thuật toán.



Hình 5: Sơ đồ khối thuật toán hàm tính ZCR

4. Đánh giá thuật toán:

Ưu điểm : Nhanh chóng xác định được giá trị ZCR tại các thời điểm trong miền thời gian rời rạc, tổng quát hóa được chỉ số ZCR của tín hiệu đầu vào.

Nhược điểm: Thuật toán còn chưa lấy được hết dữ liệu của tín hiệu. Trong trường hợp thời gian của một khung quá nhỏ thì sẽ không cung cấp đủ các giá trị trung bình từ đó việc biểu diễn hàm năng lượng còn hạn chế.

B. Framing with Overlap.

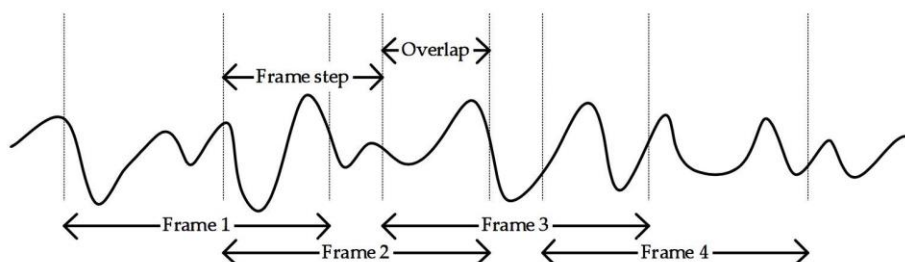
1. Vấn đề cần giải quyết.

Thông thường tín hiệu lời nói là một tín hiệu không cố định (non-stationary signal). Nhưng nếu ta nhìn nhận tín hiệu đó trong 1 khoảng thời gian rất ngắn (short time), kết quả từ phân tích ta có thể chấp nhận là nó đứng yên. Do đó, có thể áp dụng các kỹ thuật phân tích các miền (khung) cho âm thanh. Thời gian cho một khung hình được chọn trong báo cáo này sẽ là 10 đến 30 ms. Và từ đó lựa chọn ra khung nào là hợp lý nhất. Trong một khung, ta coi tín hiệu đó là đứng yên, và áp dụng các kỹ thuật xác định năng lượng và tốc độ băng qua 0 trong từ khung đó.

Việc chia khung ảnh hưởng rất nhiều đến tính toán các điểm năng lượng và tốc độ băng qua không. Và với cách chia khung đơn thuần là chia đều độ dài tín hiệu thành từng khung bằng nhau sẽ gây ra nhiều sai số. Nguyên nhân trong quá trình phân đoạn tín hiệu, ta thu thập số liệu của năng lượng và tốc độ băng qua 0 trong từng khung và đồ thị vẽ từ các số liệu đó là các điểm riêng lẻ nằm trên miền thời gian được nối thẳng lại với nhau. Bởi vậy, giả sử số lượng khung ta có thể tăng lên để xử lý, nên sẽ có thêm nhiều điểm rời rạc để nối và làm cho đồ thị trở nên chính xác hóa và lý tưởng hơn.

2. Ý tưởng thuật toán.

Để giải quyết vấn đề trên, ta sử dụng thêm một cách lấy khung khác là chồng khung: “*Framing with overlap*”. Chồng khung có nghĩa là các khung giờ đây không còn nằm liên tiếp cách biệt nhau nữa mà thay vào đó, nó sẽ chồng lên nhau. Điểm mở đầu của một khung sẽ bắt đầu trước điểm kết thúc của khung trước và dừng tại sau điểm bắt đầu của khung sau. Bằng cách đó sẽ làm tăng thêm số liệu cho năng lượng và tốc độ băng qua không, bởi số lượng khung đã được tăng trên 1 miền thời gian tín hiệu không đổi. Tăng số lượng điểm nối hay tức là tăng số lượng khung sẽ làm cho đồ thị của năng lượng và tốc độ băng qua không chính xác hơn so các ban đầu.



C. Chuẩn hóa số liệu.

1. Vấn đề cần giải quyết.

Sau khi sử dụng hai hàm: Hàm tính năng lượng và Hàm tính ZCR của tín hiệu, kết quả thu được chỉ mới là những ma trận số, những biểu đồ thể hiện số liệu chúng. Tuy nhiên, biểu đồ thu được là những biểu đồ riêng lẻ và bất đồng đơn vị. Nó khiến cho việc kết luận tiếng nói và khoảng lặng gặp nhiều khó khăn vì chưa đủ cơ sở kết luận. Và để khắc phục những vấn đề trên, trong bài báo cáo này, chúng tôi đề cập tới thuật toán về việc chuẩn hóa biểu đồ.

2. Ý tưởng chuẩn hóa.

Các biểu đồ trong thuật toán đều có hai trục là trục đơn vị X (đơn vị của năng lượng hoặc đơn vị của ZCR) và trục thời gian T. Để các biểu đồ không chung đơn vị đo lường có thể biểu diễn trong cùng một sơ đồ, ta cần khử đi đơn vị hiện tại của chúng.

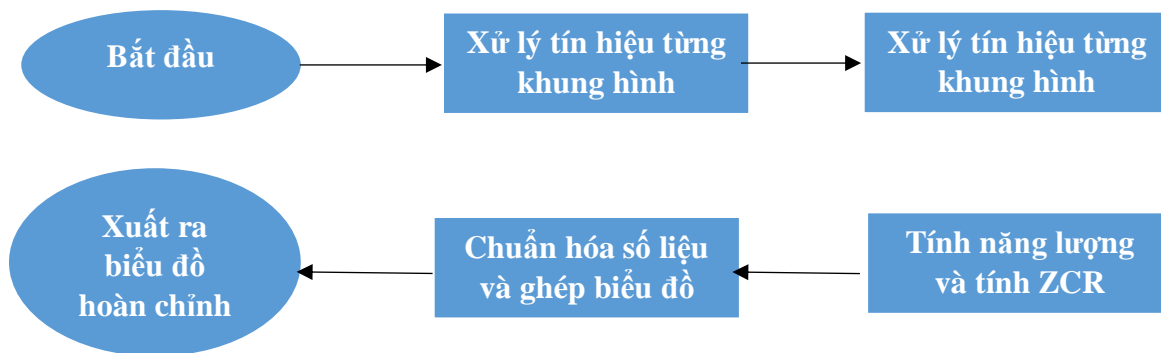
Trong mỗi biểu đồ luôn luôn tồn tại một giá trị cực đại, nếu lấy tất cả giá trị trong biểu đồ đó và chia cho cực đại giá trị, ta có thể thu được một biểu đồ mới với hai trục là trục Giá trị đã được chuẩn hóa phần trăm X và trục thời gian T. Từ đó chúng ta có thể gộp hai biểu đồ trên vào chung một biểu đồ. Ta gọi đó là chuẩn hóa số liệu trên miền thời gian.

3. Công thức chuẩn hóa số liệu.

Dựa theo ý tưởng chuẩn hóa, ta suy ra được công thức chuẩn hóa số liệu của năng lượng và ZCR trên miền thời gian như sau:

$$Dau ra = \frac{Bien}{Max(|Bien|)}$$

4. Sơ đồ khối.



Hình 7: Sơ đồ khối thuật toán chuẩn hóa số liệu.

II. Cài đặt thuật toán.

1. Chia khung hình (Framing) và tạo các ma trận cho mẫu.

Từ file: solution1.m

Dưới đây là hàm main với việc đọc tín hiệu âm thanh, chia khung và tạo ma trận (Tín hiệu, năng lượng, tín hiệu) cho mẫu tín hiệu cần xử lý.

```

clear;
clc;
[Y,Fs] = audioread('Studio_Female.wav'); %doc file am thanh

% do dai cai 1 frame theo giay
fr_du = 0.025;

% do dai cua 1 frame theo sample
fr_le = fr_du * Fs;

% do dai cua file am thanh theo sample
N = length(Y);

% tao ma tran thoi gian cua file am thanh
time = (0:N-1)/Fs;

% tim so frame va lay gia tri nguyen
nu_fr = floor(N/fr_le);
%tao ma tran cot doc signal
  
```

```

SIGNAL = zeros(N,1);
for h=1:N
    if (Y(h) >= 0)
        SIGNAL(h) = 1;
    else
        SIGNAL(h) = -1;
    end
end
% tao ma tran cot doc
E = zeros(nu_fr,1); % Nang luong
ZCR = zeros(nu_fr,1); %
Z = zeros(nu_fr,1); %

```

2. Chia khung hình (Framing with overlap) và tạo các ma trận cho mẫu.

Từ file: solution2.m

Dưới đây là hàm main với việc đọc tín hiệu âm thanh, chia khung và tạo ma trận (Tín hiệu, năng lượng, tín hiệu) cho mẫu tín hiệu cần xử lý.

```

clear;
clc;
[Y,Fs] = audioread('Studio_Female.wav'); %doc file am thanh

% do dai cai 1 frame theo giay
fr_du = 0.01;

% do dai cua 1 frame theo sample
fr_le = fr_du * Fs;

% do dai cua file am thanh theo sample
N = length(Y);

% tao ma tran thoi gian cua file am thanh
time = (0:N-1)/Fs;

lap = 0.003;
% do theo sample cua 0.03
Overlap = lap * Fs;

% So giay cua tin hieu = nu_fr*fr_le -0.003*(nu_fr - 1)
% nu_fr = (sample_length - 0.03)/(fr_le - 0.03)
% tim so frame va lay gia tri nguyen
nu_fr = floor((N - lap)/(fr_le - lap));

%tao ma tran cot doc signal
SIGNAL = zeros(N,1);
for h=1:N
    if (Y(h) >= 0)
        SIGNAL(h) = 1;
    else
        SIGNAL(h) = -1;
    end
end

% tao ma tran cot doc
E = zeros(nu_fr,1); % Nang luong

```



```
ZCR = zeros(nu_fr,1); %  
Z = zeros(nu_fr,1); %
```

3. Các hàm sử dụng.

- a. Hàm Energy: hàm tính năng lượng của cả tín hiệu.
Đầu vào: dãy tín hiệu cần tính năng lượng.
Đầu ra: ma trận giá trị năng lượng của cả dãy tín hiệu.

```
function OutputE = Energy(inputE)  
OutputE = sum(abs(inputE).^2);  
end
```

- b. Hàm ZCRsolution2: hàm tính ZCR của cả tín hiệu
Đầu vào: dãy tín hiệu cần tính ZCR.
Đầu ra: ma trận giá trị ZCR của cả dãy tín hiệu.

```
function OutZCR = ZCRsolution2(fr_le,k,DoDai,SIGNAL)  
Z2 = 0;  
for p = (k-1)*fr_le+1:DoDai+fr_le*(k-1)-1  
    Z1 = abs(SIGNAL(p)-SIGNAL(p+1));  
    Z2 = Z1 + Z2;  
end  
OutZCR = Z2;  
end
```

- c. Hàm ChuanHoa: Hàm chuẩn hóa số liệu năng lượng và ZCR.
Đầu vào: số liệu từ hàm tính năng lượng và hàm ZCR.
Đầu ra: số liệu đã được chuẩn hóa.

```
function Daura = ChuanHoa(Bien)  
Daura = Bien/max(abs(Bien));  
end
```

- d. Hàm GiaoDiem: Hàm phân đoạn giữa tiếng nói và khoảng lặng trong tín hiệu.
Đầu vào: ZCR, năng lượng, và số lượng khung của tín hiệu.
Đầu ra: vị trí của tiếng nói.

```
function Inter = TimGiaoDiem(z,e,nu_fr)  
TamThoi = zeros(nu_fr,1);  
for i=1:nu_fr  
    if (e(i) - z(i) > 0.005)  
        TamThoi(i) = 1;  
    end  
end  
Inter = TamThoi;  
end
```

- e. Hàm PLOT4: Hàm vẽ đồ thị.
Đầu vào: các ma trận của năng lượng, ZCR, tín hiệu, thời gian.
Đầu ra: đồ thị của năng lượng – ZCR – tín hiệu, giao điểm của các vùng chứa tiếng nói.

```

function PLOT4 (MATRIX, TIME, z, e, zcr, GiaoDiem, y)
subplot(4,1,1);
plot(MATRIX,z);
title('Zero Crossing Rate');
xlabel('Time (msec)');

subplot(4,1,3);
plot(TIME,y);
title('Signal and Energy');
%xlims = get(gca,'Xlim');
hold on;
plot(MATRIX,e);
%xlim(xlims);
xlabel('Time (msec)');
legend({'Signal','Short-Time Energy'});
hold off;

subplot(4,1,2)
plot(MATRIX,z);
xlabel('Time (msec)');
hold on;
plot(MATRIX,e);
title('Energy + Zero Crossing Rate');
legend({'Short-time ZCR','Short-Time Energy'});
hold off;

subplot(4,1,4);
stem(MATRIX,GiaoDiem);
xlabel('Time (msec)');
end

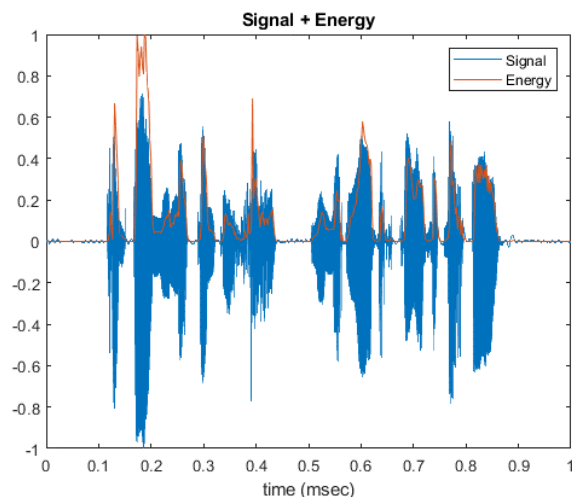
```

III. Kết quả thực nghiệm.

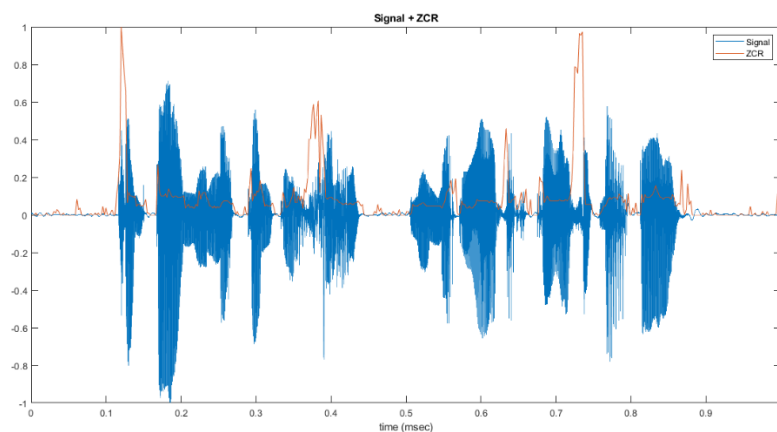
Thông qua chương trình, kết quả thu được là biểu đồ có đầy đủ thông tin về năng lượng, ZCR, các vị trí giao nhau của hai đường biểu diễn cho năng lượng và cho ZCR. Từ đó dễ dàng xác định các khoảng tiếng nói và các khoảng lặng.

A. Hình vẽ.

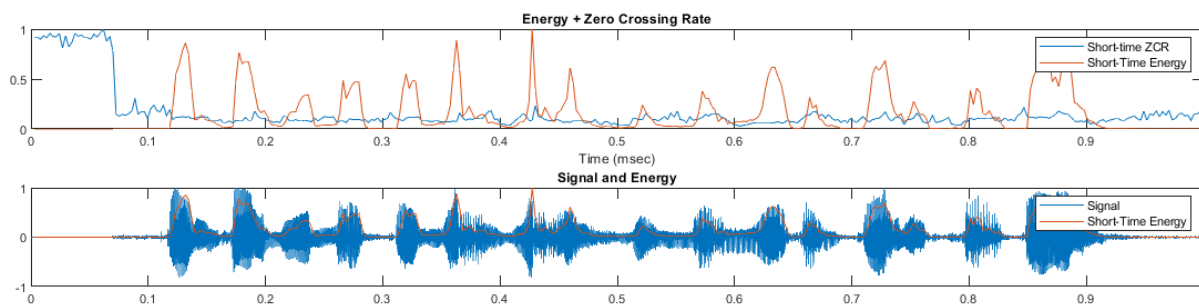
1. Kết quả của ZCR và năng lượng của tín hiệu.



Hình 8: mối quan hệ của năng lượng với tiếng nói và khoảng lặng trong tín hiệu

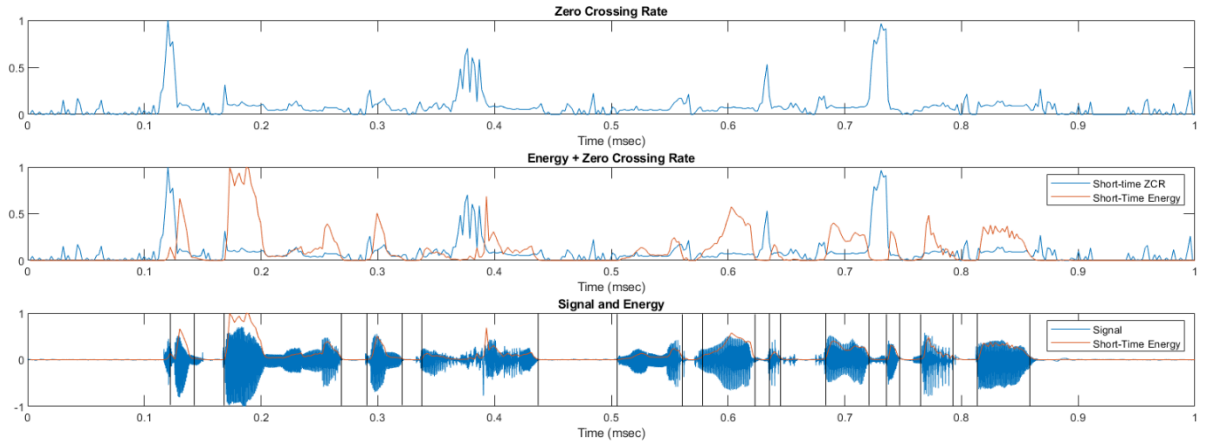


Hình 9: mối quan hệ của ZCR với tiếng nói và khoảng lặng trong tín hiệu

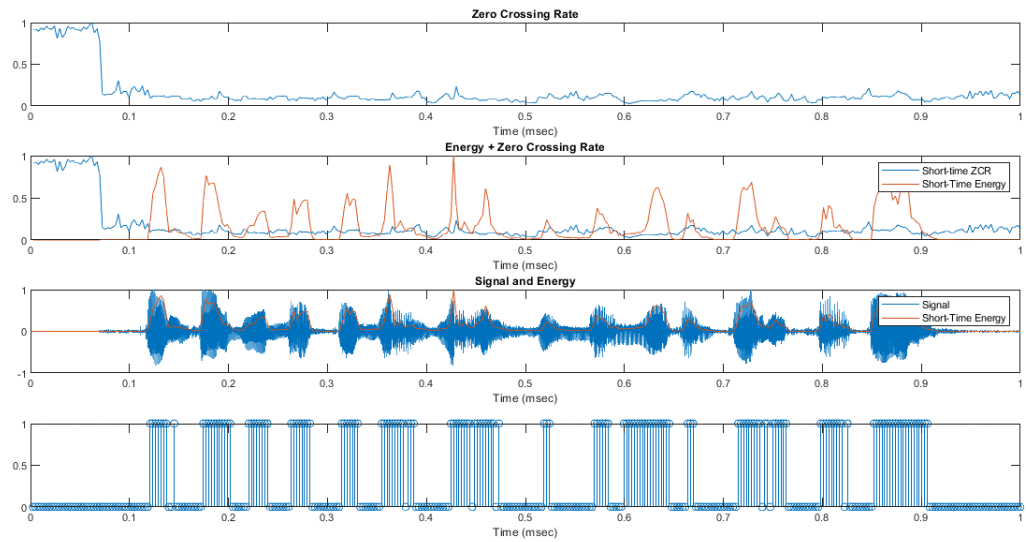


Hình 10: Mối quan hệ của năng lượng và ZCR của tín hiệu

2. Kết quả của việc phân đoạn khoảng lặng và tín hiệu.



Hình 11: Phân vùng tiếng nói và khoảng lặng bằng thủ công



Hình 12: Phân vùng tiếng nói và khoảng lặng bằng thuật toán

B. So sánh khả năng của thuật toán trong các môi trường khác nhau theo từng cách chia khung tín hiệu.

Đối với mẫu 'Female' trong ba môi trường: phone – lab – studio.

Bảng 1: Thống kê sai số theo cách chia khung Framing.

Frame – time (ms)	Phone	Lab	Studio
10	0	49	58
20	0	209	218
30	0	369	378

Đối với mẫu 'Female' trong ba môi trường: phone – lab – studio.

Bảng 2: Thống kê sai số theo cách chia khung tín hiệu Framing with Overlap.

Frame – time (ms)	Phone	Lab	Studio
10	0	0	0
20	0	0	0
30	0	0	0

Từ hai bảng trên ta thấy việc sử dụng thuật toán chia khung tín hiệu Framing with Overlap cho ra kết quả lý tưởng, không bị mất mẫu cũng như là sai số trong quá trình thực thi chương trình. Bên cạnh đó trong các môi trường khác nhau sẽ cho ra kết quả sai số khác nhau.

IV. KẾT LUẬN.

Trong bài báo cáo này ta đi xử lý phân biệt các khoảng tiếng nói và khoảng lặng. Với việc nhiều công thức, cách tính toán cũng như là nhiều thuật toán khác nhau, rút ra được kết luận với chất lượng mẫu khác nhau thì kết quả thí nghiệm cũng phụ thuộc vào nó. Cụ thể với chất lượng mẫu là ‘Studio’ (môi trường ít nhiễu nhất) và cách chia khung, thông qua chương trình, cho được kết quả chính xác nhất.

V. NHỮNG ĐIỀU HỌC ĐƯỢC

Qua báo cáo này, chúng tôi hiểu được hơn về những đặc tính trong tín hiệu âm thanh. Biết được rằng môi trường thu thập tín hiệu có ảnh hưởng rõ rệt đến chất lượng tín hiệu. Môi trường có độ nhiễu, tạp âm càng ít là môi trường đang dần tiến tới lý tưởng, giảm thiểu được việc nhận thông tin tín hiệu bị sai trong quá trình thu thập tín hiệu. Nắm được cách xử lý cũng như nhận xét số liệu và kết quả thực nghiệm.

Bên cạnh đó, trong quá trình làm báo cáo, chúng tôi học được cách làm việc nhóm, kỹ năng thuyết trình và xây dựng một bài báo cáo hoàn chỉnh.

VI. LỜI CẢM ƠN

Chúng tôi xin chân thành cảm ơn sự hướng dẫn của thầy Ninh Khánh Duy – Giảng Viên bộ môn Hệ thống nhúng, Trường Đại học Bách Khoa – Đại học Đà Nẵng.

VIII. TÀI LIỆU THAM KHẢO

[1] “Nghiên cứu phương pháp nhận dạng phân biệt tiếng nói – âm nhạc (Tạp chí khoa học công nghệ)” Đỗ Thị Loan – ĐH CNTT và Truyền thông – ĐH Thái Nguyên.

[2] “BaiGiangXLTHS” – Ninh Khánh Duy

[3] Theo Wikipedia.

[4] Prentice Hall - Digital Processing Of Speech Signals_1978 (phần 4.4).