SỬ DỤNG ĐIỂM CẮT ZERO ĐỂ NHẬN DẠNG MỘT SỐ TỪ ĐƠN ÂM TRONG TIẾNG VIỆT

Trần Anh Tuấn[1] và Thái Quốc Thắng[2]

ABSTRACT

In recent years, many methods have been proposed for the identification problem of Vietnamese pronunciation. This paper introduces a different technique using the zero-crossing point and mathematical tools to extract the characteristics of the Vietnamese words by different voices and different speakers.

Keywords: Zero-crossing, identification, Vietnamese pronunciation

Title: The use of zero-crossing point for monosyllabic word identification of Vietnamese pronunciation

TÓM TẮT

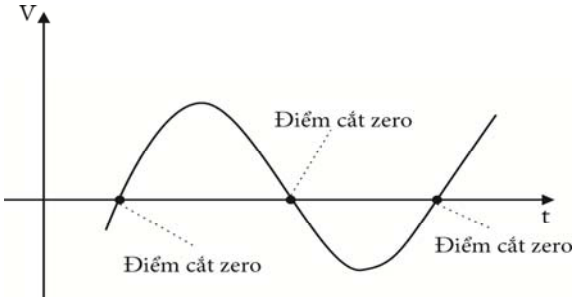
Trong những năm gần đây, nhiều phương pháp giải quyết bài toán nhận dạng từ ngữ trong phát âm tiếng Việt đã được đề xuất. Bài viết này giới thiệu một kỹ thuật khác sử dụng điểm cắt zero và các công cụ toán học để trích chọn các đặc trưng của những từ ngữ Việt được phát âm từ các giọng nói và các cá nhân thể hiện khác nhau.

Từ khóa: Điểm cắt zero, nhận dạng, phát âm tiếng Việt

1 ĐẶT VẤN ĐỀ

Để giải quyết bài toán nhận dạng tiếng nói, có ba phương pháp khá phổ biến hiện nay là: Phương pháp nhận dạng mẫu, phương pháp ứng dụng trí tuệ nhân tạo, phương pháp Âm học - Ngữ âm học. Tuy nhiên các phương pháp trên có nhược điểm là cần tìm xác suất của các mẫu và nó đòi hỏi số lượng mẫu quá lớn, và thường không tối ưu do khó sử dụng các công cụ toán học để phân tích. Vì vậy độ tin cậy và kết quả nhận dạng đạt được chưa cao.

Điểm cắt zero: Là một khái niệm được sử dụng phổ biến trong kỹ thuật điện, toán học và xử lý ảnh. Trong các khái niệm toán học, điểm cắt zero là điểm mà ở đó hàm số đổi dấu, ví dụ từ dương sang âm và được biểu diễn bằng điểm cắt trên hoành độ.

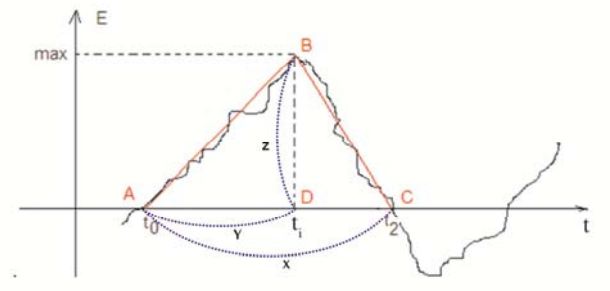


Hình 1: Điểm cắt zero biểu thị tương quan giữa điện áp và thời gian

[1] Phòng Khoa học và Hợp tác quốc tế,Trường Cao đẳng nghề Công nghiệp Thanh Hoá
[2] Phòng Đào tạo - Trường Cao đẳng nghề Công nghiệp Thanh Hoá

Trích chọn đặc trưng dựa vào điểm cắt zero:

Chúng ta xem đường cong tạo bởi tín hiệu của âm thanh là đường hình sin liên tục theo thời gian t, khi đó điểm cắt zero là điểm đường cong cắt trục thời gian (t). Thay cho việc lưu giữ các mẫu đo của tín hiệu trên cung ABC chúng ta chỉ lưu thông tin về tam giác ABC như mô tả ở hình 2.



Hình 2: Hình mô tả cách biểu diễn đoạn tín hiệu giữa hai điểm cắt zero qua tam giác ABC

Thông tin về tam giác ABC gồm:

- Độ dài cạnh AC được đo bằng $x= t_2-t_0$
- Độ dài đến vị trí cực đại của đoạn tín hiệu ABC, đo bằng $y= t_i-t_0$
- Độ lớn cực đại (max) của tín hiệu trên đoạn ABC, kí hiệu là z

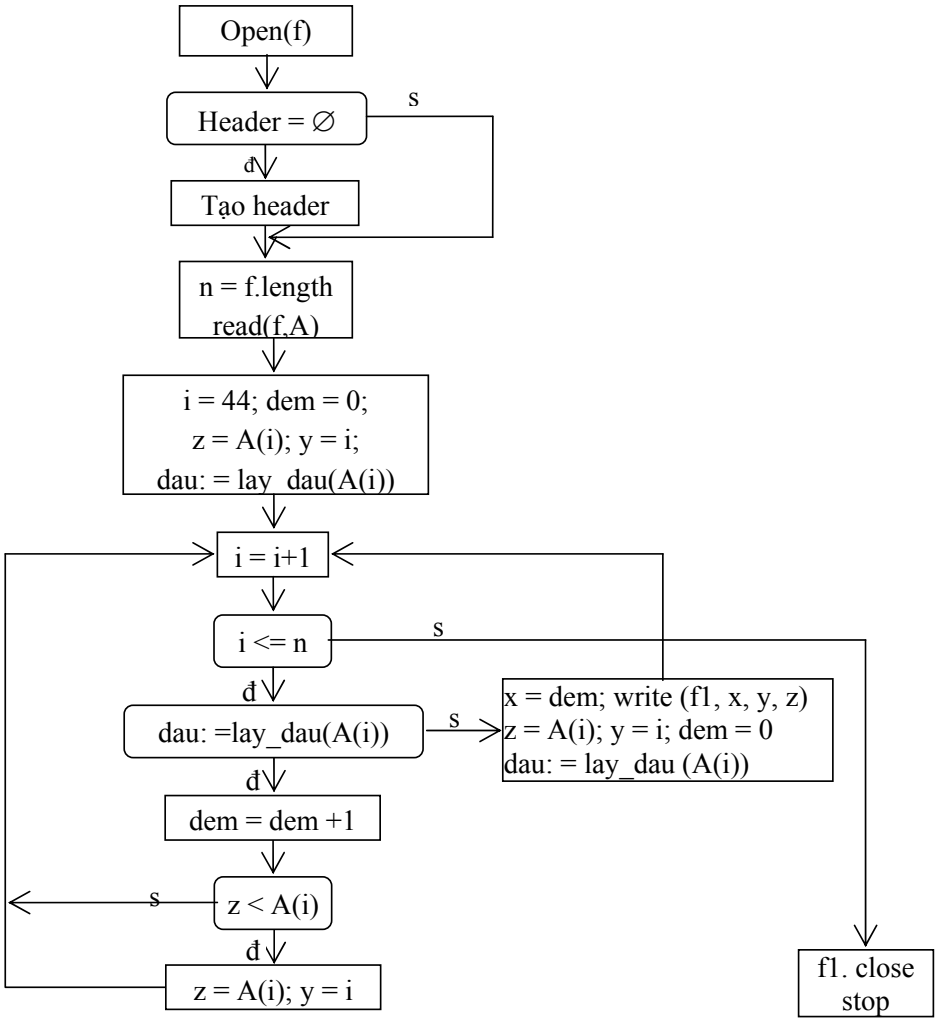
Khi đó kết quả thu được là tệp dữ liệu text mà mỗi đoạn nằm giữa của 2 điểm cắt zero liên tiếp ứng với bộ ba tham số (x,y,z).

2 THUẬT TOÁN XÁC ĐỊNH DÃY

Ngõ vào: Tín hiệu tiếng nói, là chuỗi các biên độ tương ứng các thời điểm đó.

Ngõ ra: Dữ liệu là một chuỗi của các bộ 3 tham số (x,y,z) tương ứng tín hiệu tại mỗi đoạn giữa của 2 điểm cắt zero liên tiếp.

Đặt n là độ dài tệp dữ liệu được gọi tên là f.wave, dùng mảng A để đọc dữ liệu tiếng nói từ tệp dữ liệu f. Duyệt từ byte thứ 44 cho đến cuối mảng A (do cấu trúc tệp dữ liệu dạng wave, 44 byte đầu tiên lưu thông tin Header của tệp dữ liệu), xét dấu từng mẫu trong tín hiệu, nếu có sự đổi dấu của tín hiệu ở mẫu liền kề tức là có tồn tại một điểm cắt zero. Trong đoạn giữa hai điểm cắt zero liên tiếp này, tính z bằng với max{|A(i)|}, y là vị trí tính z và x là độ dài đoạn tín hiệu đang khảo sát, nếu chọn bước lấy mẫu là đơn vị thì x cũng là số mẫu được lấy trên đoạn tín hiệu trên. Lưu bộ 3 giá trị này vào tệp dữ liệu f1. Tiếp tục thực hiện như trên cho đến khi hết tệp dữ liệu f, tệp dữ liệu có tên f1.txt nhận được, chỉ chứa các bộ ba (x,y,z) của tệp dữ liệu ban đầu f.wave.



Hình 3: Sơ đồ mô tả thuật toán xác định tệp (x,y,z)

Các biến được sử dụng trong thuật toán nén điểm cắt zero được mô tả trên hình 3:

dau: *nhận giá trị - hoặc +; để nhận biết vị trí dãy tín hiệu đổi dấu có nghĩa là tín hiệu có cắt trục 0 (tồn tại điểm cắt zero).*

A: *lưu giá trị tín hiệu*

x: *lưu số mẫu của một bước sóng.*

y: *vị trí mẫu có biên độ cực đại.*

z: *giá trị biên độ cực đại.*

n: *số mẫu thuộc toàn bộ tín hiệu khảo sát.*

dem: *biến trung gian đếm số mẫu trong một bước sóng (giới hạn bởi 2 điểm cắt zero)*

File f: *chứa dữ liệu tiếng nói ngõ vào.*

File f1: *chứa dữ liệu nén ngõ ra.*

3 XÁC ĐỊNH CÁC ĐẶC TRƯNG

Từ tệp dữ liệu f.wave ta được dãy $\{x_i, y_i, z_i\}$, trong đó i = 1,2...n. Vấn đề là cần phải tìm các đặc trưng cho dãy $\{x_i, y_i, z_i\}$.

Dựa vào tính tuần hoàn của sóng âm thanh ta suy ra $\{x_i, y_i, z_i\}$ phải chứa các dãy con lặp lại.

Thuật toán phát hiện ra các tương quan:

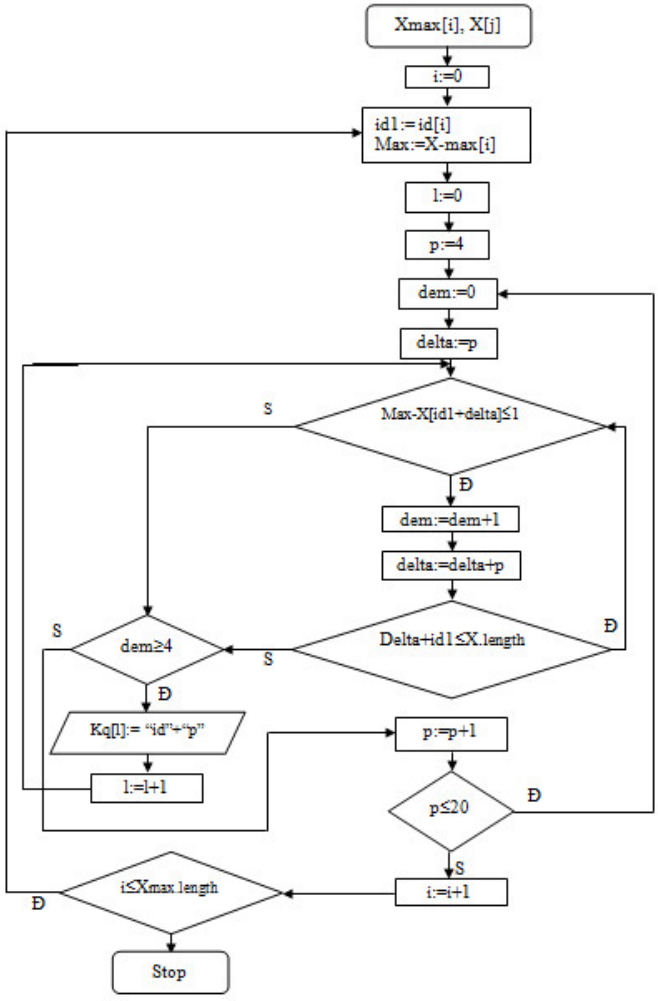
Trong dãy (x,y,z) có giá trị z là độ lớn của tiếng nói, có thể được thay đổi để làm cho biên độ âm thanh lớn lên hay nhỏ đi. Vậy dãy lặp cần tìm có tính đặc trưng chứa ở các thông tin là x và y (tệp <x,y>).

Để xác định tệp <x,y>, từ tệp <x,y,z> ta tách ra tệp <x>, sau đó từ tệp <x> ta lọc ra các phần tử tệp $<x_{max}>$ có giá trị từ n trở lên đến m.

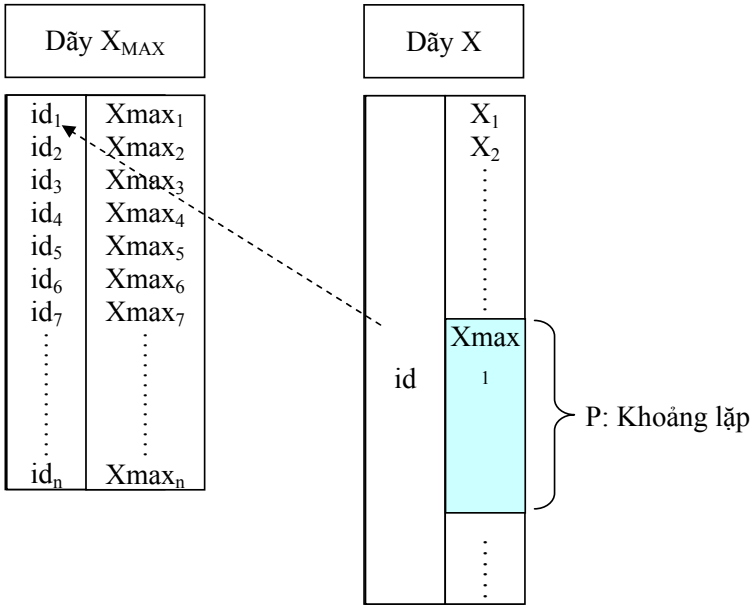
Ngõ vào: Dãy các giá trị $\{X_{max}\}$, $\{X\}$

Ngõ ra: Tập các dãy lặp $\{ V_1, V_2, ...V_k\}$. Trong đó: $V_i = \{x_{i1}, x_{i2}, ... x_{ip}\}$ và p là khoảng lặp

Ta có sơ đồ khối mô tả thuật toán như sau:



Hình 4: Thuật toán phát hiện ra các dãy lặp (x, y)



Hình 5: Cách lấy dãy X_{max} từ dãy X và lấy các giá trị khoảng lặp p từ dãy X

Các biến được sử dụng trong thuật toán phát hiện ra các dãy lặp được mô tả trên Hình 4:

i: Chỉ số các phần tử của dãy X_{max}
l: chỉ số của mảng kq[l] = id,p
Xmax[i]: Dãy các các số lớn hơn hoặc bằng m và nhỏ hơn hoặc bằng n.
id: Chỉ số các phần tử X_{max} trong với dãy các phần tử thuộc X.
p: Khoảng lặp lại.
dem: Biến đếm trong quá trình phát hiện các dãy lặp lại.
delta: Biến gán ứng giá trị p.
X.length: Độ dài của dãy các phần tử thuộc X.
Xmax.length: Độ dài của dãy các phần tử thuộc Xmax.
Kq[l]: Kết quả của thuật toán.

Cách trích đặc trưng khi loại dữ liệu dãy lặp nhờ tương quan:

Từ tập các đoạn lặp thu được từ thuật toán trên ta loại bỏ bớt các đoạn lặp nhờ việc tính tương quan giữa các đoạn và chọn được dãy dữ liệu có tính đặc trưng.

4 THUẬT TOÁN NHẬN DẠNG

Ngõ vào: Từ tiếng Việt dạng file wave.

Ngõ ra: Kết quả nhận dạng, phát âm ra từ đó.

Mô tả các biến được sử dụng trong thuật toán:

- **Mau:** Các dãy đặc trưng của từ cần nhận dạng.
- **Vi:** Khối tập đặc trưng thứ i trong bộ dữ liệu.
- **i:** Chỉ số của khối tập đặc trưng trong bộ dữ liệu.
- **k:** Số lượng khối đặc trưng trong bộ dữ liệu.

- **Vitri:** Vị trí của khối tập đặc trưng trong bộ dữ liệu mà từ nhận dạng được.
- **c:** Hệ số tin cậy để đạt mức nhận dạng được giữa V_i và Mau.

Thuật toán gồm các bước như sau:

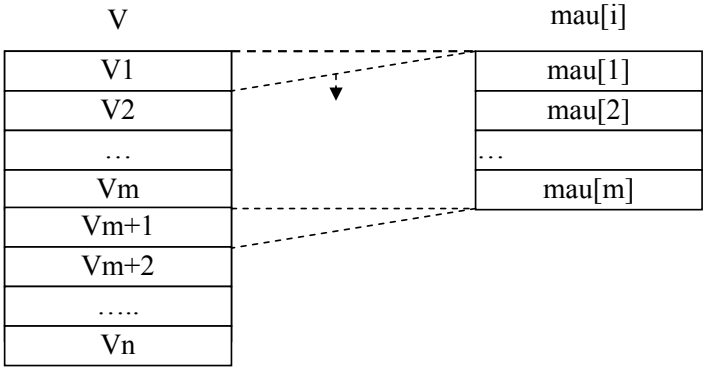
- Tạo lập được bộ dữ liệu đặc trưng hay gọi là "Không gian mẫu".
- Nhập từ cần nhận dạng vào thu được các dãy đặc trưng của mẫu cần nhận dạng
- Chọn hệ số tin cậy "c".
- Sử dụng tính tương quan để đối sánh giữa "Không gian mẫu" và "Mau", khi giá trị tương quan đạt lớn hơn bằng "c" thì nhận dạng được.

Xét các trường hợp khi tính hàm tương quan: n = | V | ; m = | mau[i] |

Trường hợp 1: n = m với r_i = tuongquan(V_i, mau[i][j]) ≥ 0,9 thì V ≡ mau[i]

Trường hợp 2: n > m

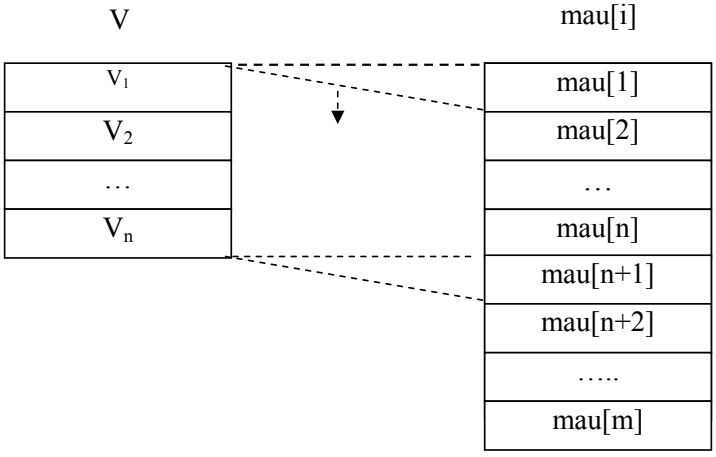
Ta thực hiện cách đối sánh theo mô hình sau:



Hình 6: Xét sự tương quan giữa hai mảng trường hợp n > m

Trường hợp 3: n < m

Ta thực hiện cách đối sánh theo mô hình sau:



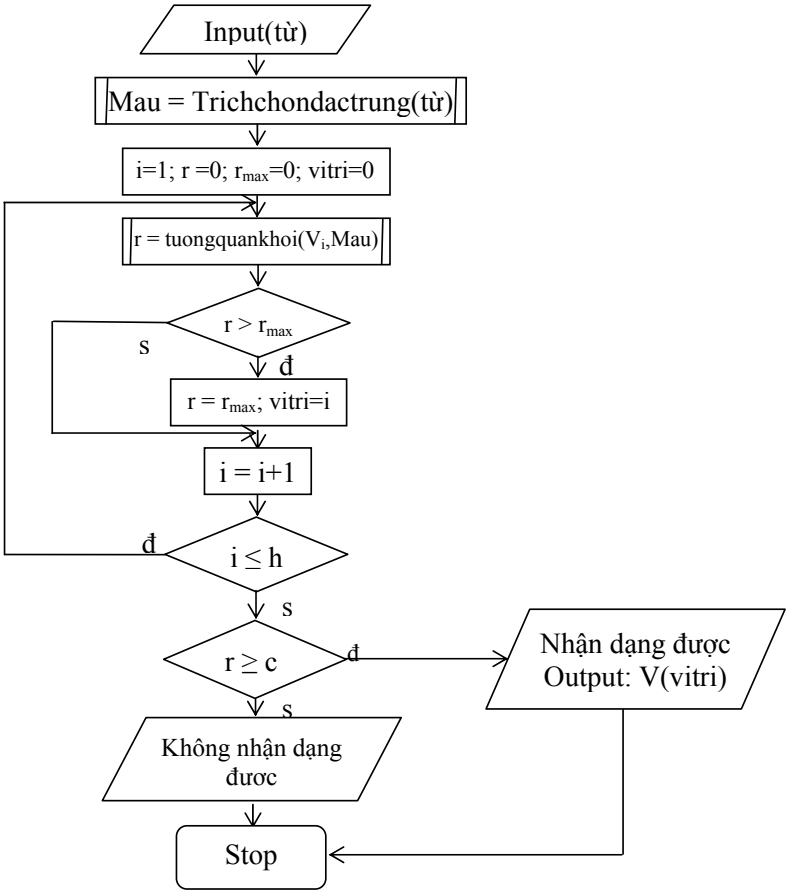
Hình 7: Xét sự tương quan giữa hai mảng trường hợp n < m

Tính hệ số tương quan cho cả ba trường hợp trên:

$$r_i = \text{tuongquan}(V_i, \text{mau}[i][j]) = \frac{\text{sodongtuongquan}}{\max(n, m)}$$

với sodongtuongquan: số dòng tương quan giữa hai mảng.

Ta có sơ đồ khối mô tả thuật toán nhận dạng như sau:



Hình 8: Thuật toán nhận dạng

5 KẾT QUẢ

Trong thực nghiệm này, cơ sở dữ liệu tiếng nói được sử dụng là bộ file wave chứa các từ: (có, không, mở, đóng) với 40 người nói khác nhau bao gồm 25 nam, 15 nữ có độ tuổi từ 16-50.

Nạp bộ dữ liệu đặc trưng là 160 từ (có, không, mở, đóng) của 40 người nói khác nhau.

Sử dụng 160 từ (có, không, mở, đóng) khác của 40 người nói trên và 20 từ (có), 20 từ (không) của 20 người mới đưa vào để nhận dạng.

Input \ Output	Số lần nhận dạng đúng	Số lần nhận dạng sai
160 từ *(có, không, mở, đóng)* khác của 40 người đầu	160	0
20 từ *(có)* của 20 người mới	14	6
20 từ *(không)* của 20 người mới	18	2

Nhận xét: Mặc dù trong tất cả các thử nghiệm trên với kết quả nhận dạng chưa đạt được như mong muốn. Nhưng trên thực tế cho thấy quá trình lấy mẫu và nhận dạng đều được thực hiện trong môi trường ồn ào, không cần phòng kín, cách âm, số mẫu được lấy còn giới hạn. Vì vậy để nâng cao tỷ lệ nhận dạng, cần phải có số lượng mẫu lớn hơn nhiều so với số mẫu đã được sử dụng trong nghiên cứu này.

Mặt khác, phương pháp sử dụng điểm cắt zero dùng hệ số tương quan để đánh giá mối quan hệ giữa mẫu nhận dạng với không gian mẫu (đã huấn luyện). Khi hệ số tương quan càng gần 1 thì độ tin cậy nhận dạng càng cao nhưng tỷ lệ nhận dạng lại giảm và khi hệ số tương quan càng gần 0 thì độ tin cậy nhận dạng càng giảm nhưng tỷ lệ nhận dạng lại tăng dẫn đến việc nhận dạng sai. Do đó trong các nghiên cứu tiếp theo có thể tính tương quan bằng phương pháp Hồi quy tuyến tính để khắc phục.

6 KẾT LUẬN

Bài toán nhận dạng từ tiếng Việt là một trong những bài toán khó, đã có nhiều phương pháp nhận dạng khác nhau, nhưng kết quả đạt được còn hạn chế. Phương pháp sử dụng điểm cắt zero để nhận dạng tiếng nói hiện nay đang được thế giới quan tâm. Chúng tôi đã sử dụng điểm cắt zero và các công cụ toán học để thử nghiệm tách chọn các đặc trưng của bốn từ đơn (có, không, mở, đóng) để kiểm chứng chương trình nhận dạng do chúng tôi đề xuất. Để có kết quả tốt hơn chúng tôi cần thực hiện các nghiên cứu tiếp theo như: Cực tiểu hóa tập các đặc trưng và thử nghiệm thêm nhiều người nói ở vùng miền và lứa tuổi khác nhau để chọn ra tập đặc trưng thích hợp nhất.

TÀI LIỆU THAM KHẢO

Trần Anh Tuấn, Sử dụng điểm cắt zero để nén và giải nén dữ liệu âm thanh *(ý tưởng điểm cắt zero, thuật toán xác định dãy)*, tạp chí khoa học Đại học Cần Thơ (6/2012).

Hồ Tú Bảo, Lương Chi Mai (2008), Về xử lý tiếng Việt trong Công nghệ thông tin *(nhận dạng tiếng nói)*.

David Salomon (2004), Data Compression The Complete Reference, 3ed (Springer).

Wiley (2003), Speech Coding Algorithms Foundation and Evolution of Standardized Coders, Ebooks.

Ian H.Witten, Radford M. Neal and John G, (1987), Arithmetic coding for data compression, Clearyin *Communications of the ACM*.