NHẬN DẠNG GIỌNG NÓI TIẾNG VIỆT BẰNG LOGIC MỜ

USING FUZZY LOGIC IN VIETNAMESE SPEECH RECOGNITION

Trần Đức Minh[1], Nguyễn Thiện Luận[2]

[1]Trường Đại học Thăng Long; Email: tdminh2110@yahoo.com
[2]Học viện Kỹ thuật Quân sự; Email: nthienluan@yahoo.com

Tóm tắt - Bài báo này giới thiệu phương pháp nhận dạng giọng nói Tiếng Việt bằng công cụ Logic mờ, cụ thể là nhận dạng phổ tín hiệu tiếng nói. Tiếng Việt là ngôn ngữ đơn âm, do đó với mỗi từ khi phát âm đều có một hình dạng phổ tín hiệu nhất định. Vì vậy, ta đưa bài toán nhận dạng giọng nói Tiếng Việt thành bài toán nhận dạng phổ tín hiệu âm thanh. Logic mờ là công cụ được áp dụng vào cả hai bài toán huấn luyện và nhận dạng tiếng nói. Đối với bài toán huấn luyện, thông tin đầu vào là các tín hiệu âm thanh được chuyển đổi thành dữ liệu mờ để lưu trữ nhằm phục vụ quá trình nhận dạng; đối với bài toán nhận dạng, phép hiệu đối xứng trên tập mờ giữa thông tin cần nhận dạng và dữ liệu mờ là công cụ quan trọng nhất hỗ trợ quá trình nhận dạng. Kết quả thực nghiệm cho thấy với lượng từ hữu hạn và phổ tín hiệu âm thanh có hình dạng tương đối khác nhau thì việc nhận dạng đạt được hiệu quả cao và đáng tin cậy.

Từ khóa - nhận dạng âm thanh; nhận dạng tiếng nói;nhận dạng giọng nói Tiếng Việt;Logic mờ; Logic mờ ứng dụng

Abstract - This paper proposes the Fuzzy logic based method to recognize Vietnamese speech, namely recognition of spectrum of voiced signal. As Vietnamese is a monosyllabus language, each word has a pronunciation with a specific spectral entity. The problem of recognizing Vietnamese speech can therefore be converted into the problem of recognizing spectral properties of voiced signals. In this paper we introduce how the fuzzy logic based method is applied to solving the problems of training and recognizing. The experiment results showed recognition of high efficiency and reliability when the data is small and spectra are diverse.

Key words - speech recognition; Vietnamese speech recognition; voice recognition; fuzzy logic; fuzzy logic application

1. Đặt vấn đề

Nhận dạng giọng nói là vấn đề đã được quan tâm từ nhiều năm trở lại đây do tính ứng dụng thực tiễn cao của lĩnh vực này trong cuộc sống. Trên thế giới hiện nay đã có khá nhiều ứng dụng nhận dạng giọng nói chạy trên cả máy tính và thiết bị cầm tay. Ở Việt Nam, hiện nay cũng đã xuất hiện một vài ứng dụng có sử dụng tính năng nhận dạng giọng nói, tuy nhiên mới chỉ áp dụng trong một số lĩnh vực cụ thể. Chính vì vậy tập từ nhận dạng thường hữu hạn, điển hình một vài ứng dụng được công bố gần đây nhất là: *Công cụ quản lý chi tiêu cá nhân điều khiển bằng tiếng nói* [1] và *Hệ thống tra cứu thông tin tuyển sinh bằng tiếng nói* [2] của Khoa CNTT, Đại học Huflit; ngoài ra cũng có một vài công trình nghiên cứu khác như *điều khiển ô tô từ xa bằng giọng nói* [3] trong lĩnh vực điều khiển hay *điều khiển cánh tay Robot bằng giọng nói Tiếng Việt* [4] trong lĩnh vực Robot, …

Đối với định hướng nghiên cứu, do về mặt ngữ âm, Tiếng Việt có đặc thù là ngôn ngữ đơn âm tiết, do đó ta không thể áp dụng các phương pháp nhận dạng của ngôn ngữ đa âm tiết. Trên thực tế, các nghiên cứu về nhận dạng giọng nói được công bố trên thế giới chủ yếu là các nghiên cứu dành cho ngôn ngữ đa âm tiết. Chính vì vậy, hệ thống nhận dạng giọng nói ở Việt Nam khó có thể kế thừa lại toàn bộ các nghiên cứu này. Do đó, hệ thống nhận dạng giọng nói ở Việt Nam cần phải được xây dựng theo hướng đi khác, đó là hướng đi dựa trên nền tảng ngữ âm Tiếng Việt.

Trong thời gian vừa qua đã có khá nhiều cá nhân, tổ chức đầu tư nghiên cứu một cách bài bản về vấn đề nhận dạng giọng nói Tiếng Việt và cũng đã đạt được một số kết quả nhất định. Điển hình như đề tài *Tổng hợp và nhận dạng tiếng nói ứng dụng vào vấn đề nhập đọc dữ liệu văn bản* [5], đề tài *Nghiên cứu các mô hình xử lý tín hiệu tiếng nói phục vụ cho việc nhận dạng Tiếng Việt nói liên tục* [6] và đề tài *Nghiên cứu kỹ thuật tổng hợp giọng nói ứng dụng trong đọc văn bản Tiếng Việt* [7]. Tuy nhiên, vẫn chưa có một chương trình ứng dụng nhận dạng giọng nói tổng thể nào dành cho Tiếng Việt được công bố.

Nhìn chung, bài toán nhận dạng giọng nói Tiếng Việt là một bài toán khó bởi để giải quyết bài toán này, trước tiên ta cần phải giải quyết khá nhiều bài toán phức tạp khác. Ví dụ giọng nói miền Bắc, miền Trung và miền Nam là tương đối khác nhau, để hỗ trợ cho bài toán nhận dạng giọng nói Tiếng Việt, trước tiên ta cần phải *nhận dạng phương ngữ của giọng nói*[8]; hay quan trọng trong vấn đề nhận dạng giọng nói mà bài toán cần phải giải quyết là *tách âm thanh của một câu nói Tiếng Việt thành âm thanh của từng từ riêng biệt* [9] hoặc *nhận dạng thanh điệu tiếng nói Tiếng Việt* [10].

Về phương pháp nhận dạng giọng nói, có rất nhiều phương pháp tính toán thông minh đã được áp dụng, một vài nghiên cứu gần đây như *Ứng dụng mô hình Markov ẩn để nhận dạng tiếng nói trên chip FPGA* [11]; *Mô hình nhận dạng giọng nói Tiếng Việt trong điều khiển theo góc độ từ riêng biệt* [12] đề xuất mô hình nhận dạng giọng nói Tiếng Việt dựa trên thuật toán quy hoạch động và mô hình Markov ẩn; *Nhận dạng tiếng nói bằng mạng Nơron nhân tạo* [13] đều cho ta kết quả nhận dạng khá chính xác với tập từ hữu hạn. Cũng không nằm ngoài xu hướng trên, bài báo này trình bày một cách tiếp cận nhận dạng giọng nói Tiếng Việt bằng công cụ Logic mờ. Dữ liệu được sử dụng trong bài báo là các tín hiệu tiếng nói được đưa vào hệ thống một cách rời rạc nhằm tăng độ chính xác và mỗi mẫu âm thanh cần xử lý chính là nội dung phổ của tín hiệu tiếng nói mà ta nhận được thông qua phép biến đổi Fourier nhanh [14].

Nội dung bài báo được chia thành 7 mục. Trong đó mục 2 giới thiệu phương pháp lấy phổ tín hiệu tiếng nói để phục vụ cho quá trình học và quá trình nhận dạng; mục 3 giới

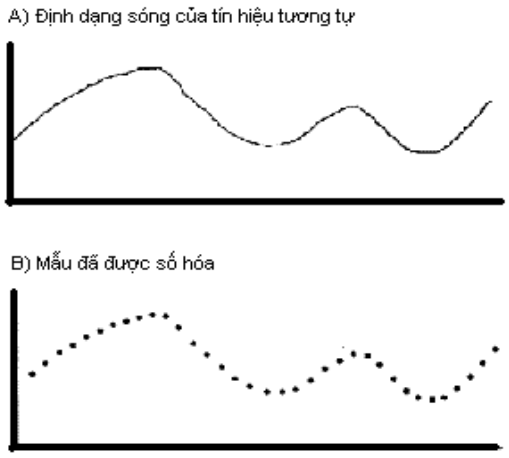
thiệu về Logic mờ; mục 4 trình bày tổng quan về phương pháp nhận dạng tiếng nói; mục 5 trình bày chi tiết phương pháp học và nhận dạng mẫu bằng Logic mờ; mục 6 đưa ra một số kết quả thử nghiệm của phương pháp đã đề xuất và cuối cùng là mục kết luận.

2. Phổ tín hiệu âm thanh

Để giải quyết bài toán nhận dạng giọng nói, bước đầu tiên cần phải xử lý là *số hóa tín hiệu âm thanh*, tức là ta cần phải chuyển đổi tín hiệu tương tự của âm thanh sang tín hiệu số.

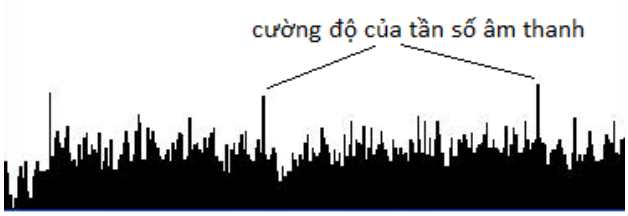
Trong quá trình lấy mẫu, phần cứng quan tâm chủ yếu đến một thiết bị ngoại vi chuyên dụng được gọi là thiết bị chuyển đổi tín hiệu tương tự sang tín hiệu số (Analog to Digital converter – viết tắt là ADC). Thiết bị này chịu trách nhiệm lấy tín hiệu tương tự của âm thanh rồi chuyển đổi nó thành những con số rời rạc để máy tính có thể dễ dàng xử lý.

Nhằm phục vụ quá trình nhận dạng, ta cần trích rút những thông tin cần thiết. Đối với phương pháp nhận dạng trong bài báo này, ta sử dụng phương pháp biến đổi Fourier nhanh (The Fast Fourier Transform - FFT) để trích rút thông tin *nội dung phổ của tín hiệu âm thanh*. Điều này đồng nghĩa với việc ta sẽ lấy được một dãy cường độ các tần số âm thanh khác nhau sau khi sử dụng phép biến đổi Fourier nhanh.



Hình 1. *Tín hiệu tương tự và tín hiệu số âm thanh*

Ta coi dãy cường độ của các tần số âm thanh khác nhau này là một véctơ $A = (a_1, a_2, \dots, a_n)$ chứa n phần tử số thực. Mỗi khi ta phát âm vào micro một từ Tiếng Việt bất kỳ, nội dung phổ của từ sẽ tạo ra một hình dạng nào đó, hình dạng này được tạo bởi giá trị của các phần tử a_i. Như vậy, với mỗi véctơ A ta có một **mẫu dữ liệu âm thanh**.



Hình 2. *Nội dung phổ tín hiệu âm thanh*

Bản chất của bài toán nhận dạng giọng nói Tiếng Việt trong bài báo này là nhận dạng phổ của tín hiệu âm thanh hay nhận dạng mẫu dữ liệu âm thanh.

3. Cơ sở lý thuyết Logic mờ

3.1. Định nghĩa tập mờ

Tập mờ A được xác định trên không gian nền kinh điển X là một tập mà mỗi phần tử của nó là một cặp $(x, \mu_A(x))$ trong đó $x \in X$ và $\mu_A(x)$ là ánh xạ:

$$\mu_A : X \rightarrow [0, 1]$$

Ánh xạ μ_A được gọi là hàm liên thuộc của tập mờ A.

3.2. Một số khái niệm của tập mờ

3.2.1. Định nghĩa 1

Độ cao của tập mờ A trên không gian nền X là giá trị:

$$h = \sup_{x \in X} \mu_A(x)$$

Ký hiệu $\sup_{x \in X} \mu_A(x)$ chỉ giá trị nhỏ nhất trong tất cả các giá trị chặn trên của hàm $\mu(x)$. Một tập mờ với ít nhất một phần tử có độ phụ thuộc bằng 1 được gọi là *tập mờ chính tắc* tức là $h = 1$, ngược lại một tập mờ A với $h < 1$ được gọi là *tập mờ không chính tắc*.

*3.2.2. Định nghĩa 2*Miền xác định của tập mờ A trên không gian nền X được ký hiệu bởi S là tập con của X thỏa mãn:

$$S = Supp\ \mu_A(x) = \{ x \in X \mid \mu_A(x) > 0 \}$$

Ký hiệu *Supp* chỉ rõ tập con trong X với các phần tử x mà tại đó hàm $\mu_A(x)$ có giá trị dương.

3.2.3. Định nghĩa 3

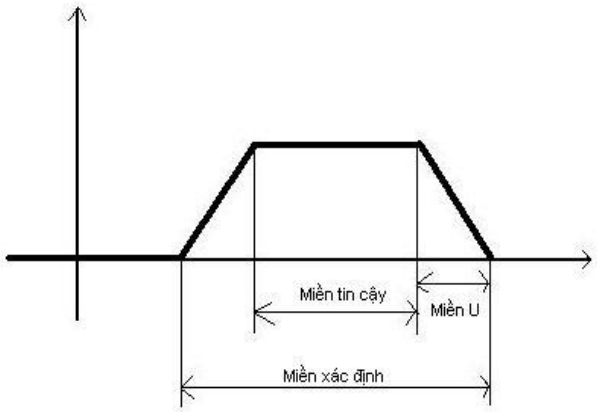
Miền tin cậy của tập mờ A trên không gian nền X được ký hiệu bởi T là tập con của X thỏa mãn:

$$T = \{ x \in X \mid \mu_A(x) = 1 \}$$

3.2.4. Định nghĩa 4

Miền biên của tập mờ A trên không gian nền X được ký hiệu bởi U là tập con của X thỏa mãn:

$$U = \{ x \in X \mid 0 < \mu_A(x) < 1 \}$$



Hình 3. *Biểu diễn các miền của một tập mờ*

3.2.5. Định nghĩa 5

Lực lượng của tập mờ A trên không gian nền X được biểu diễn như sau:

$$N(A, \mu_A(x)) = \sum_{x \in A} \mu_A(x)$$

3.3. Các phép toán trên tập mờ

Trong các định nghĩa sau các tập A, B, C thuộc cùng thuộc không gian nền X.

3.3.1. Phép hợp hai tập mờ

Định nghĩa: Hợp của hai tập mờ *A* và *B* là một tập mờ $A \cup B$ cùng xác định trong không gian nền *X* có hàm liên thuộc $\mu_{A\cup B}(x)$ thỏa mãn các tiên đề sau:

a. Chỉ phụ thuộc vào $\mu_A(x)$ và $\mu_B(x)$

b. Nếu $\mu_B(x) = 0$ với mọi *x* thì $\mu_{A\cup B}(x) = \mu_A(x)$

c. Có tính giao hoán $\mu_{A\cup B}(x) = \mu_{B\cup A}(x)$

d. Có tính kết hợp $\mu_{(A\cup B)\cup C}(x) = \mu_{A\cup(B\cup C)}(x)$

e. Có tính không giảm (đồng biến). Nếu $A_1 \subseteq A_2$ thì $A_1 \cup B \subseteq A_2 \cup B$:

$$\mu_{A1}(x) \le \mu_{A2}(x) => \mu_{A1\cup B}(x) \le \mu_{A2\cup B}(x)$$

Một số công thức định nghĩa hàm liên thuộc $\mu_{A\cup B}(x)$ cho hợp của hai tập mờ:

i. $\mu_{A\cup B}(x) = Max\{\mu_A(x), \mu_B(x)\}$

ii. $\mu_{A\cup B}(x) =$
$\begin{cases} Max\{\mu_A(x), \mu_B(x)\} \text{ nếu } Min\{\mu_A(x), \mu_B(x)\} = 0 \\ Min\{\mu_A(x), \mu_B(x)\} \text{ nếu } Min\{\mu_A(x), \mu_B(x)\} \neq 0 \end{cases}$

iii. *Phép hợp theo Lukasiewicz*

$$\mu_{A\cup B}(x) = Min\{1, \mu_A(x) + \mu_B(x)\}$$

iv. *Tổng Einstein*

$$\mu_{A\cup B}(x) = \frac{\mu_A(x) + \mu_B(x)}{1 + \mu_A(x)\mu_B(x)}$$

v. *Tổng trực tiếp*

$$\mu_{A\cup B}(x) = \mu_A(x) + \mu_B(x) - \mu_A(x)\,\mu_B(x)$$

3.3.2. Phép giao hai tập mờ

Định nghĩa: Giao của hai tập mờ *A* và *B* là một tập mờ $A \cap B$ cùng xác định trong không gian nền *X* có hàm liên thuộc $\mu_{A\cap B}(x)$ thỏa mãn các tiên đề sau:

a. Chỉ phụ thuộc vào $\mu_A(x)$ và $\mu_B(x)$

b. Nếu $\mu_B(x) = 1$ với mọi *x* thì $\mu_{A\cap B}(x) = \mu_A(x)$

c. Có tính giao hoán $\mu_{A\cap B}(x) = \mu_{B\cap A}(x)$

d. Có tính kết hợp $\mu_{(A\cap B)\cap C}(x) = \mu_{A\cap(B\cap C)}(x)$

e. Có tính không giảm (đồng biến). Nếu $A_1 \subseteq A_2$ thì $A_1 \cap B \subseteq A_2 \cap B$:

$$\mu_{A1}(x) \le \mu_{A2}(x) => \mu_{A1\cap B}(x) \le \mu_{A2\cap B}(x)$$

Một số công thức định nghĩa hàm liên thuộc $\mu_{A\cap B}(x)$ cho giao của hai tập mờ:

i. $\mu_{A\cap B}(x) = Min\{\mu_A(x), \mu_B(x)\}$

ii. $\mu_{A\cap B}(x) =$
$\begin{cases} Min\{\mu_A(x), \mu_B(x)\} \text{ nếu } Max\{\mu_A(x), \mu_B(x)\} = 1 \\ 0 \qquad\qquad\quad \text{nếu } Maxs\{\mu_A(x), \mu_B(x)\} \neq 0 \end{cases}$

iii. $\mu_{A\cap B}(x) = Max\{0, \mu_A(x) + \mu_B(x) - 1\}$

iv. $\mu_{A\cap B}(x) = \frac{\mu_A(x)\mu_B(x)}{2 + \mu_A(x)\mu_B(x) - (\mu_A(x) + \mu_B(x))}$

v. $\mu_{A\cap B}(x) = \mu_A(x)\mu_B(x)$

3.3.3. Phép bù của một tập mờ

Định nghĩa: Tập bù của tập mờ A trên nền X là một tập mờ $(\bar{A}, \mu_{\bar{A}})$ xác định trên không gian nền X với hàm liên thuộc $\mu(\mu_A) : [0, 1] -> [0, 1]$ thỏa mãn các điều kiện sau:

a. $\mu(1) = 0$

b. $\mu(0) = 1$

c. $\mu_A \le \mu_B => \mu(\mu_A) \ge \mu(\mu_B)$

Nếu hàm một biến $\mu(\mu_A)$ còn thỏa mãn

d. *Liên tục* và

e. $\mu_A < \mu_B => \mu(\mu_A) > \mu(\mu_B)$

thì phép bù trên còn gọi là phép *bù mờ chặt*. Một phép bù mờ chặt được gọi là phép *bù mờ mạnh* nếu:

f. $\mu(\mu(\mu_A)) = \mu_A$ *tức là* $\bar{\bar{A}} = A$

Hàm liên thuộc $\mu(\mu_A)$ của phép bù mờ mạnh được gọi là hàm **phủ định mạnh**.

Một số công thức định nghĩa hàm liên thuộc cho phép lấy phần bù của tập mờ:

i. $\mu_{\bar{A}}(x) = 1 - \mu_A(x)$

ii. *Hàm bù ngưỡng* λ

$$\mu_\lambda(t) = \begin{cases} 1 \text{ nếu } t \le \lambda \\ 0 \text{ nếu } t > \lambda \end{cases} \text{ với } \lambda \in [0, 1]$$

iii. *Hàm bù Cosin*

$$\mu(t) = \frac{1 + \cos(\pi t)}{2}$$

iv. *Hàm bù Sugeno*

$$\mu_\lambda(t) = \frac{1 - t}{1 + \lambda t} \text{ với } \lambda \in [-1, \infty]$$

3.3.4. Phép hiệu đối xứng

Mở rộng công thức cho phép hiệu đối xứng các tập kinh điển:

$$A \nabla B = (A \cup B)(A \cap B) = (A \cap \overline{B}) \cup (\overline{A} \cap B)$$

ta có thể xây dựng phép hiệu đối xứng cho các tập mờ. Ngoài việc có thể áp dụng hàm liên thuộc cho các phép toán trên tập hợp, ta cũng có thể xây dựng hàm liên thuộc cho phép hiệu đối xứng của hai tập mờ phụ thuộc vào việc lựa chọn các công thức cho phép hợp và phép giao của các tập mờ.

4. Tổng quan về phương pháp nhận dạng

Một hệ thống nhận dạng nhìn chung phải trải qua ba bước cơ bản: Bước học, bước lưu trữ và bước nhận dạng.

4.1. Bước học

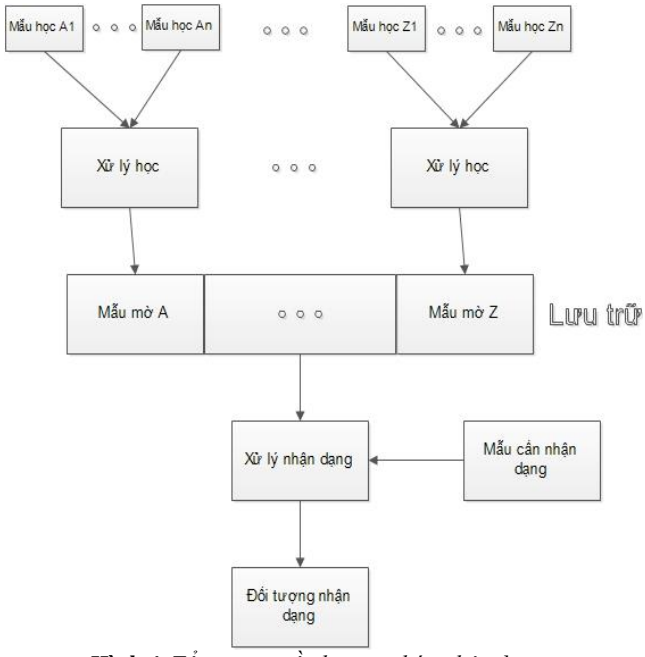
Hay còn gọi là bước huấn luyện. Ở bước này, với mỗi từ hoặc âm cần học, hệ thống được cung cấp một tập hợp các mẫu dữ liệu âm thanh chuẩn của từ hoặc âm đó. Ta sẽ xử lý các mẫu này cùng với nhau theo một quy tắc xác định để nhận được một mẫu dữ liệu âm thanh "mờ" đại diện cho từ cần huấn luyện. Tập hợp nhiều mẫu dữ liệu âm thanh "mờ" sẽ tạo nên cơ sở dữ liệu từ vựng của hệ thống.

4.2. Bước lưu trữ

Tập các mẫu dữ liệu âm thanh "mờ" sẽ được lưu giữ lại để sử dụng cho quá trình nhận dạng. Việc lưu trữ này có thể sử dụng một hệ quản trị cơ sở dữ liệu hay một file nhị phân có cấu trúc do hệ thống tự định nghĩa.

4.3. Bước nhận dạng

Đây là bước ra quyết định xem mẫu được đưa vào hệ thống giống với từ hay âm nào nhất căn cứ vào cơ sở dữ liệu từ vựng của hệ thống.

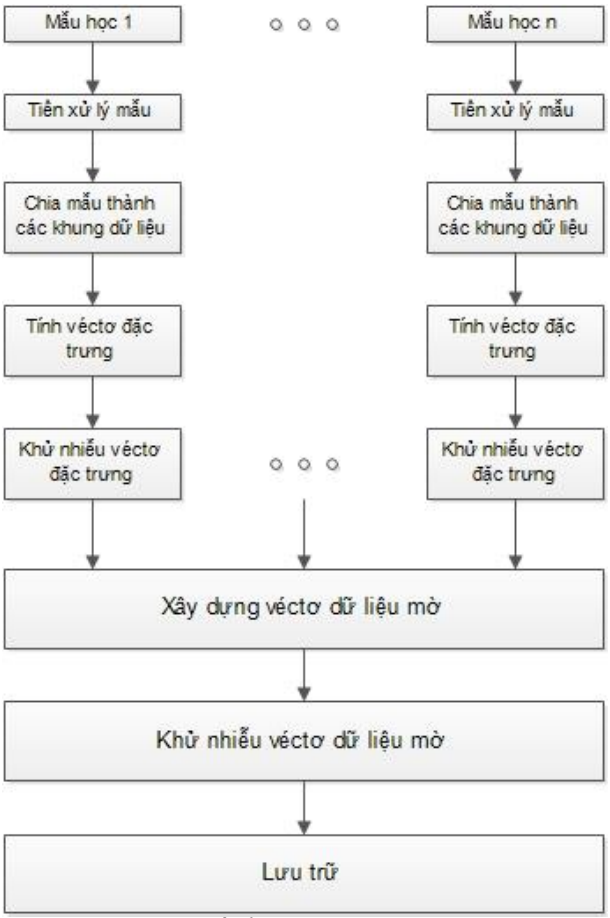


Hình 4. *Tổng quan về phương pháp nhận dạng*

5. Phương pháp nhận dạng dựa trên Logic mờ

5.1. Bước xử lý học

Giả sử hệ thống được cung cấp một tập hợp các mẫu dữ liệu âm thanh chuẩn của một từ hoặc âm nào đó. Các bước từ một cho đến bốn dưới đây là các bước bắt buộc phải xử lý đối với mỗi mẫu dữ liệu âm thanh đưa vào hệ thống.



Hình 5. *Sơ đồ tổng quan quá trình xử lý học*

Bước một (bước tiền xử lý mẫu): Như đã biết, với mỗi mẫu âm thanh hệ thống nhận được, ta coi mẫu đó như một véctơ $A = (a_1, a_2, \dots, a_n)$ với a_i là số thực.

Mục đích của bước tiền xử lý mẫu là đưa các giá trị a_i nằm trong mẫu về khoảng *[0,1]*. Để giải quyết vấn đề này, ta lấy a_i chia cho một số m (số m này được đưa ra bằng việc nghiên cứu thực nghiệm về cường độ của tần số âm thanh mà ta lấy được thông qua phép biến đổi Fourier nhanh, ở hệ thống này tác giả chọn $m = 800$. Bản chất của việc này là ta cần phải chọn số m sao cho không có quá nhiều $\frac{a_i}{m} \geq 1$). Nếu sau khi chia ta nhận được $\frac{a_i}{m} > 1$ thì ta coi $\frac{a_i}{m} = 1$.

Sau bước tiền xử lý mẫu, ta nhận được véctơ A^* có giá trị như sau:

$$A^* = (\frac{a_1}{m}, \frac{a_2}{m}, \dots, \frac{a_n}{m})$$

Bước hai: Chia mẫu A^* thành T khung dữ liệu. Số T này do ta quy định (cần xác định số T không quá lớn cũng không quá nhỏ).

Chú ý: Tất cả các mẫu đưa vào đều phải chia thành đúng T khung dữ liệu.

Bước ba: Tính véctơ đặc trưng $A_{ĐT}$. Số phần tử của véctơ đặc trưng bằng T (bằng với số khung dữ liệu ở *Bước hai*). Giá trị của mỗi phần tử trong véctơ đặc trưng $A_{ĐT}$ bằng trung bình cộng các phần tử nằm trong cùng một khung T_i của véctơ A^*. Sau khi tính toán ta có:

$$A_{ĐT} = (a_1^{ĐT}, a_2^{ĐT}, \dots, a_T^{ĐT})$$

Bước bốn: Khử nhiễu cho véctơ đặc trưng bằng công thức sau:

$$\begin{cases} A_{KN} = 2A_{ĐT}^2 \text{ nếu } 0 \leq A_{ĐT} < 0.5 \\ A_{KN} = 1 - 2(1 - A_{ĐT})^2 \text{ nếu } 0.5 \leq A_{ĐT} \leq 1 \end{cases}$$

Quá trình khử nhiễu này mục đích chính là để những giá trị nào nằm trong khoảng [0, 0.5) sẽ dần về 0 và những giá trị nằm trong khoảng [0.5, 1] sẽ dần đến 1. Quá trình này có thể lặp đi lặp lại nhiều lần. Ở hệ thống này tác giả chọn khử nhiễu một lần.

Kết quả thu được sau khi khử nhiễu, ta sẽ thu được véctơ khử nhiễu có số phần tử bằng số phần tử của véctơ đặc trưng và giá trị mỗi phần tử trong véctơ khử nhiễu được tính bởi công thức trên. Sau khi tính toán ta có:

$$A_{KN} = (a_1^{KN}, a_2^{KN}, \dots, a_T^{KN})$$

Chú ý:

- Với bất kỳ mẫu âm thanh nào ta cũng đều phải xử lý qua tất cả bốn bước trên để nhận được véctơ khử nhiễu.
- Sau khi xây dựng xong toàn bộ véctơ khử nhiễu đối với các mẫu âm thanh chuẩn của một từ hoặc âm đưa vào, ta sẽ xây dựng "dữ liệu mờ" cho từ hoặc âm đó thông qua toàn bộ véctơ khử nhiễu này.
- Quá trình xây dựng "dữ liệu mờ" bắt đầu từ *Bước năm* trở đi.

Bước năm: Ta xây dựng véctơ dữ liệu mờ của một từ hoặc âm dựa trên các véctơ đặc trưng đã khử nhiễu (véctơ A_{KN}) ở trên. Véctơ dữ liệu mờ được ký hiệu là $\tilde{A}$ *(A ngã)*. $\tilde{A}$ là véctơ dữ liệu mờ của một từ hoặc âm.

Việc xây dựng $\tilde{A}$ tương đương với việc xây dựng tập

mờ *F* với mỗi phần tử của *F* là một cặp $(x, \mu_{F(x)})$.Ở đây, *x* được xác định là chỉ số của các khung chia *T*. Ta ký hiệu $x_1, x_2, \ldots x_T$ là các chỉ số.

$\mu_F$là hàm liên thuộc của tập mờ *F*. $\mu_{F(x_i)}$được tính bằng trung bình cộng của các giá trị khung chiaT_{x_i} trong các véctơ khử nhiễu. Sau khi tính toán ta có:

$$\tilde{A} = (\widetilde{a_1}, \widetilde{a_2}, \ldots, \widetilde{a_T})$$

Bước sáu: Khử nhiễu véctơ dữ liệu mờ $\tilde{A}$ thông qua công thức gần giống ở *Bước 4*:

$$\begin{cases} \mu' = 2\mu^2 \text{ nếu } 0 \le \mu < 0.5 \\ \mu' = 1 - 2(1-\mu)^2 \text{ nếu } 0.5 \le \mu \le 1 \end{cases}$$

Việc khử nhiễu này có thể được lặp đi lặp lại nhiều lần. Về ý nghĩa của điều này tương tự như *Bước 4*.

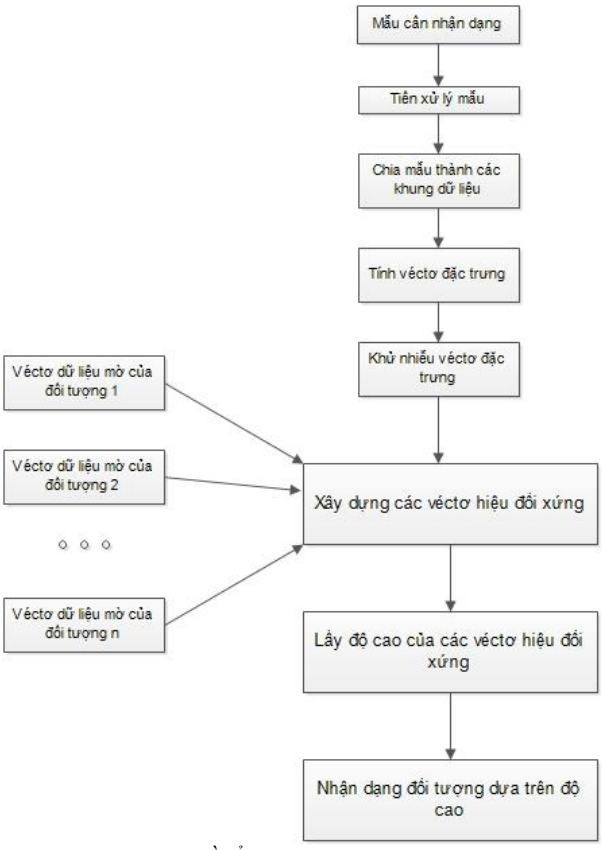
5.2. Bước lưu trữ

Bước này lưu giữ lại véctơ $\tilde{A}$ sau khi đã khử nhiễu. Đây chính là "dữ liệu mờ" của một từ hoặc âm sau khi đã được học.

Chú ý: Mỗi từ hoặc âmđược huấn luyện sẽ có 01 véctơ $\widetilde{A_{KN}}$ làm đại diện cho từ hoặc âm đó. Tức là, tương ứng với *n* từ hoặc âm cần nhận dạng, ta sẽ có *n*véctơ$\widetilde{A_{KN}}$ được lưu trữ để phục vụ quá trình nhận dạng.

5.3. Bước nhận dạng

Đầu vào của bước nhận dạng là một mẫu âm thanh bất kỳ. Nhiệm vụ của của bước này là tìm trong tập mẫu dữ liệu âm thanh "mờ" đã được lưu trữ (cơ sở dữ liệu từ vựng của hệ thống), mẫu nào giống với mẫu âm thanh được đưa vào nhất thì ta kết luận từ hoặc âm đại diện cho mẫu đó là từ hoặc âm cần nhận dạng.



Hình 6. *Sơ đồ tổng quan quá trình nhận dạng*

Bốn bước xử lý đầu tiên đối với mẫu âm thanh cần nhận dạng giống hệt với bốn bước xử lý đầu tiên của quá trình học. Ta cần chú ý giá trị *m* ở *Bước một* và giá trị *T* ở *Bước ba* trong quá trình nhận dạng cũng phải bằng với giá trị *m* ở *Bước một* và giá trị *T* ở *Bước ba* trong quá trình học.

Bước năm: Trước tiên, ta lấy toàn bộ dữ liệu đã được học để đem ra phục vụ quá trình xử lý nhận dạng. Dữ liệu này chính là các véctơ $\widetilde{A_{KN}}$ đại diện cho các từ hoặc âm đã được huấn luyện ở quá trình học.

Xây dựng các véctơ hiệu đối xứng của từng véctơ $\widetilde{A_{KN}}$ với véctơ đặc trưng của mẫu cần nhận dạng sau khi đã được khử nhiễu. Cụ thể như sau: Giả sử véctơ *A* đại diện cho véctơ $\widetilde{A_{KN_i}}$ nào đó, véctơ đặc trưng *B* sau khi đã khử nhiễu là đại diện cho mẫu cần nhận dạng.

Ta có thể mở rộng công thức cho phép hiệu đối xứng các tập kinh điển:

$$A \nabla \mathrm{B} = (\mathrm{A} \cup \mathrm{B})(\mathrm{A} \cap \mathrm{B}) = (\mathrm{A} \cap \overline{B}) \cup (\overline{A} \cap \mathrm{B})$$

để xây dựng phép hiệu đối xứng cho các tập mờ. Trong đó các phép tính hợp ta chọn *1* trong *5* công thức định nghĩa hàm liên thuộc $\mu_{A\cup B}(x)$ cho hợp của hai tập mờ ở *mục 3.3.1*; phép tính giao ta chọn *1* trong *5* công thức định nghĩa hàm liên thuộc $\mu_{A\cap B}(x)$ cho giao của hai tập mờ ở *mục 3.3.2*.

Ví dụ 1: Ta áp dụng công thức *(i)* của phép tính hợp và công thức *(i)* của phép tính giao để tính véctơ hiệu đối xứng.

$$\mu_{A \nabla B}(x) = Max\left(Min(\mu_A(x), \mu_{\bar{B}}(x)), Min(\mu_{\bar{A}}(x), \mu_B(x))\right)$$
$$= Max\left(Min(\mu_A(x), 1 - \mu_B(x)), Min(1 - \mu_A(x), \mu_B(x))\right)$$
$$= \begin{cases} Max\{\mu_A(x), \mu_B(x)\} \text{ nếu } \mu_A(x) + \mu_B(x) \le 1 \\ Max\{1-\mu_A(x), 1-\mu_B(x)\} \text{ nếu } \mu_A(x) + \mu_B(x) > 1 \end{cases}$$

Ví dụ 2: Ta áp dụng công thức *(iii)* của phép tính hợp và công thức *(i)* của phép tính giao để tính véctơ hiệu đối xứng.

$$\mu_{A \nabla B}(x) = Min(1, \mu_{A\cap\bar{B}}(x) + \mu_{\bar{A}\cap B}(x))$$
$$= Min\left(1, Min(\mu_A(x), \mu_{\bar{B}}(x)) + Min(\mu_{\bar{A}}(x), \mu_B(x))\right)$$
$$= Min\left(1, Min(\mu_A(x), 1-\mu_B(x)) + Min(1-\mu_A(x), \mu_B(x))\right)$$
$$= \begin{cases} Min\{1, \mu_A(x) + \mu_B(x)\} \text{ nếu } \mu_A(x) + \mu_B(x) \le 1 \\ 2 - (\mu_A(x) + \mu_B(x)) \text{ nếu } \mu_A(x) + \mu_B(x) > 1 \end{cases}$$

Sau khi tính theo công thức hiệu đối xứng của *n* véctơ $\widetilde{A_{KN}}$với véctơ đặc trưng đã khử nhiễu của mẫu cần nhận dạng, ta sẽ nhận được *n*véctơ hiệu đối xứng tương ứng. Sau khi tính toán mỗi véctơ hiệu đối xứng đều có dạng:

$$A_{HĐX} = (a_1^{HĐX}, a_2^{HĐX}, \ldots, a_T^{HĐX})$$

Bước sáu: Tính độ cao δ dựa trên véctơ hiệu đối xứng. Độ cao δ này có thể được tính theo một trong các cách sau:

- $\delta = Min(a_i^{HĐX})$: tức là chọn giá trị của δ bằng với giá trị của phần tử nhỏ nhất trong véctơ hiệu đối xứng.
- $\delta = \frac{\sum a_i^{HĐX}}{T}$:tức là chọn giá trị trung bình cộng của các phần tử trong véctơ hiệu đối xứng.

Như vậy mỗi véctơ hiệu đối xứng đều có một độ cao δ_i xác định.

Bước bảy: Xác định từ hoặc âm giống với mẫu âm thanh được đưa vào nhận dạng nhất bằng cách:

$$Tính\ \delta_{min} = Min(\delta_i)$$

Chọn ra từ hoặc âm có giá trị $\delta = \delta_{min}$. Đây chính là từ hoặc âm giống với mẫu âm thanh được đưa vào nhất.

6. Thử nghiệm và đánh giá

6.1. Thử nghiệm

Chương trình thực nghiệm nhận dạng giọng nói Tiếng Việt sử dụng công cụ Logic mờ được xây dựng và chạy thử nghiệm trên máy tính cá nhân. Các mẫu âm thanh học và nhận dạng được truyền trực tiếp từ micro vào máy tính. Chương trình thực nghiệm gồm 3 module chính: Module học, module lưu trữ và module nhận dạng.

Module học: Giải quyết bước xử lý học. Người sử dụng đọc lặp đi lặp lại liên tiếp nhưng rời rạc một từ Tiếng Việt với số lần tùy ý vào micro. Kết thúc quá trình này, hệ thống sẽ có được dữ liệu mờ của từ mà người sử dụng vừa đọc. Để cho hệ thống học tiếp từ khác người sử dụng lại lặp lại quá trình trên.

Module lưu trữ: Dữ liệu mờ của tất cả những từ hay âm sau khi được học hệ thống sẽ tự động lưu trữ ra một file nhị phân ở bộ nhớ ngoài của máy tính.

Module nhận dạng: Giải quyết bước nhận dạng. Người sử dụng đọc từ muốn nhận dạng vào micro. Hệ thống sẽ tự động tính toán, so sánh và đánh giá từ muốn nhận dạng với tập dữ liệu từ đã được mờ hóa để đưa ra đối tượng giống với từ người sử dụng muốn nhận dạng nhất lên màn hình.

6.2. Đánh giá

Tác giả xây dựng 3 tập mẫu: Tập mẫu thứ nhất chỉ gồm 5 từ có phổ âm thanh tương đối khác nhau; Tập mẫu thứ hai gồm 10 từ cũng có phổ âm thanh tương đối khác nhau; Tập mẫu thứ 3 gồm 4 từ có phổ âm thanh gần giống nhau.

- *Ở bước học*: mỗi từ trong các tập mẫu được đọc lặp đi lặp lại 10 lần.
- *Ở bước nhận dạng*: mỗi từ cũng được nói vào micro 10 lần để xác định tần suất nhận dạng chính xác.

6.2.1. Tập mẫu 1

Tác giả chọn 5 từ: *"Chữ", "Số", "Một", "Hai", "Năm"*. Đặc điểm của 5 từ này là có phổ âm thanh tương đối khác nhau. Kết quả nhận dạng như sau:

Bảng 1. *Kết quả nhận dạng Tập mẫu 1*

Từ	Số lượng mẫu huấn luyện	Số lượng mẫu nhận dạng	Kết quả nhận dạng
Chữ	10	10	100%
Số	10	10	100%
Một	10	10	100%
Hai	10	10	100%
Năm	10	10	100%

6.2.2. Tập mẫu 2

Tác giả chọn 10 từ: *"Chữ", "Số", "Một", "Hai", "Năm", "Sáu", "Bảy", "Tám", "Chín", "Mười"*. Đặc điểm của 10 từ này là cũng có phổ âm thanh tương đối khác nhau. Kết quả nhận dạng như sau:

Bảng 2. *Kết quả nhận dạng Tập mẫu 2*

Từ	Số lượng mẫu huấn luyện	Số lượng mẫu nhận dạng	Kết quả nhận dạng
Chữ	10	10	90%
Số	10	10	100%
Một	10	10	90%
Hai	10	10	90%
Năm	10	10	80%
Sáu	10	10	90%
Bảy	10	10	100%
Tám	10	10	80%
Chín	10	10	100%
Mười	10	10	90%

6.2.3. Tập mẫu 3

Tác giả chọn 4 từ: *"Một", "Bốn", "Cột", "Trốn"*. Đặc điểm của 4 từ này là cặp từ "Một", "Bốn" và "Cột", "Trốn" có phổ âm thanh tương đối giống nhau. Kết quả nhận dạng như sau:

Bảng 3. *Kết quả nhận dạng Tập mẫu 3*

Từ	Số lượng mẫu huấn luyện	Số lượng mẫu nhận dạng	Kết quả nhận dạng
Một	10	10	50%
Bốn	10	10	50%
Cột	10	10	40%
Trốn	10	10	30%

7. Kết luận và đề xuất

Như vậy với tập từ Tiếng Việt hữu hạn và có phổ tín hiệu âm thanh tương đối khác nhau, hệ thống có khả năng nhận dạng lên đến 90%. Với tập từ có phổ của tín hiệu âm thanh gần giống nhau, khả năng nhận dạng có kém hơn. Do đó trong quá trình huấn luyện ta phải đọc đi đọc lại từ vựng đó nhiều lần hơn và cần có những phương pháp khử nhiễu, làm nổi rõ tín hiệu chính của phổ âm thanh tốt hơn nữa.

Bài toán nhận dạng tiếng nói là một bài toán nghiên cứu lớn và khó nhưng lại có rất nhiều ứng dụng trong thực tiễn. Phương pháp được đề xuất trong bài báo này là ứng dụng Logic mờ trong bài toán nhận dạng giọng nói Tiếng Việt. Tuy còn nhiều hạn chế nhưng đã đáp ứng được phần nào mục tiêu ban đầu đưa ra. Ngoài việc tìm hiểu cơ sở lý thuyết cho bài toán nhận dạng giọng nói Tiếng Việt, tác giả đã xây dựng được một chương trình thực nghiệm cụ thể nhằm chứng minh tính đúng đắn của mô hình lý thuyết trong thực tiễn. Điều này cho thấy việc ứng dụng Logic mờ trong bài toán nhận dạng giọng nói Tiếng Việt là một hướng đi mở và chính xác. Vì lý do đó, ta có thể đề xuất thêm các nghiên cứu ở mức sâu hơn nữa trong vấn đề này.

Do nắm được yếu điểm của phương pháp nhận dạng giọng nói nêu trên đó là nếu phổ tín hiệu âm thanh của các mẫu tương đối giống nhau sẽ dẫn đến việc nhận dạng có độ chính xác không cao. Tác giả đề xuất nghiên cứu thêm một số phương pháp lọc nhiễu và làm nổi rõ các tín hiệu chính của phổ âm thanh, nhằm phục vụ quá trình nhận dạng được tốt hơn. Đây cũng là những vấn đề cần phải nghiên cứu trong lĩnh vực nhận dạng tiếng nói.

TÀI LIỆU THAM KHẢO

[1] Trần Khải Thiện, Văn Thế Quốc, Nguyễn Phạm Bảo Nguyên, Nguyễn Vũ Kiều Anh, Vũ Thanh Hiền, *Xây dựng công cụ quản lý chi tiêu cá nhân điều khiển bằng tiếng nói*, Khoa CNTT, Đại học Huflit, Hội nghị khoa học Quốc gia lần thứ VII, 19-20 tháng 6/2014.

[2] Trần Khải Thiện, Văn Thế Quốc, Nguyễn Phạm Bảo Nguyên, Nguyễn Vũ Kiều Anh, Vũ Thanh Hiền, *Hệ thống tra cứu thông tin tuyển sinh Đại học HUFLIT bằng tiếng nói*, Khoa CNTT, Đại học Huflit, Hội nghị khoa học Quốc gia lần thứ VII, 19-20 tháng 6/2014.

[3] Nguyễn Văn Giáp, Trần Việt Hồng, *Kỹ thuật nhận dạng tiếng nói và ứng dụng trong điều khiển*, Đại học Bách khoa TP Hồ Chí Minh.

[4] Lê Tiến Thường, Hoàng Đình Chiến, *Vietnamese Speech Recognition Applied to Robot Communications*, Au Journal of Technology, Volume 7 No. 3 January 2004.

[5] Chủ nhiệm: Hoàng Văn Kiếm, *Tổng hợp và nhận dạng tiếng nói ứng dụng vào nhập đọc dữ liệu văn bản kiểm soát bảo vệ điều khiển các hệ thống thông tin máy tính, hỗ trợ xây dựng các sản phẩm multi media dạy học trên cơ sở Tiếng Việt*, Đại học Khoa học tự nhiên, Đại học quốc gia TP Hồ Chí Minh, Giải thưởng Vifotech 1999.

[6] *Nghiên cứu các mô hình xử lý tín hiệu tiếng nói phục vụ cho việc nhận dạng Tiếng Việt nói liên tục*, mã số 203806, Trường Đại học Công nghệ, Đai học Quốc Gia Hà Nội, 2006-2008.

[7] Chủ nhiệm: Phạm Ngọc Hưng, *Nghiên cứu kỹ thuật tổng hợp giọng nói ứng dụng trong đọc văn bản Tiếng Việt*, Đại học Sư phạm kỹ thuật Hưng Yên, 2008.

[8] Phạm Ngọc Hưng, Trịnh Văn Loan, Nguyễn Hồng Quang, Phạm Quốc Hùng, *Nhận dạng phương ngữ Tiếng Việt sử dụng mô hình Gauss hỗn hợp*, Đại học sư phạm kỹ thuật Hưng Yên, Viện CNTT & TT Đại học Bách Khoa Hà Nội, 2014.

[9] Vũ Đức Lung, Nguyễn Thái Ân, Đào Anh Nguyên, *Tổng hợp các phương pháp tách âm thanh của một từ Tiếng Việt và đề xuất phương pháp cải tiến*, Đại học Công nghệ Thông tin, Đại học Quốc Gia HCM, Hội nghị khoa học Quốc gia lần thứ VII, 19-20 tháng 6/2014.

[10] Lê Tiến Thường, *Nhận dạng thanh điệu tiếng nói Tiếng Việt bằng mạng Nơron phân tầng*, Tạp chí tin học và điều khiển học, 2005.

[11] Nguyễn Cao Quý, *Ứng dụng mô hình Markov ẩn để nhận dạng tiếng nói trên FPGA*, Tạp chí khoa học, Đại học Cần Thơ, 2013

[12] Đào Anh Nguyên, Vũ Đức Lung, Nguyễn Thái Ân, *Mô hình nhận dạng giọng nói Tiếng Việt trong điều khiển theo góc độ từ riêng biệt*, Đại học Công nghệ Thông tin, Đại học Quốc Gia HCM, Hội nghị khoa học Quốc gia lần thứ VII, 19-20 tháng 6/2014.s

[13] Phùng Chí Dũng, Nhận dạng tiếng nói bằng mạng Nơron nhân tạo, *Tạp chí bưu chính viễn thông*, 2003.

[14] PGS.TS. Nguyễn Hữu Phương, *Xử lý tín hiệu số*, Nhà xuất bản Giao thông vận tải, 2000

(BBT nhận bài: 28/07/2014, phản biện xong: 07/08/2014)