

THUẬT TOÁN MẠNG NƠON TRUYỀN THĂNG ĐA LỚP CHO PHÂN LOẠI TIẾNG NÓI TRONG ĐIỀU KHIỂN XE LĂN ĐIỆN CHO NGƯỜI TÀN TẬT

MLFNN ALGORITHM FOR SPEECH CLASSIFICATION IN ELECTRICAL WHEELCHAIR CONTROL FOR DISABLED PEOPLE

Nguyễn Thanh Hải

Trường Đại học Sư phạm Kỹ thuật Tp. Hồ Chí Minh; nthai@hcmute.edu.vn

Tóm tắt - Bài báo này đề xuất một thuật toán mạng nơon truyền thăng đa lớp (MLFNN) cho phân loại tiếng nói trong điều khiển xe lăn điện, trong đó trích dẫn đặc trưng của tín hiệu được thực hiện bằng phương pháp dựa vào hệ số phổ tần số Mel (MFCC). Những tín hiệu tiếng nói là những lệnh điều khiển được nhận biết như *Left*, *Right*, *Forward*, *Backward* và *Stop*. Hơn nữa, bộ lọc thông thấp Hamming được áp dụng để đặt khung cho mục đích giảm nhiễu trước khi trích dẫn đặc trưng. Vậy, những tín hiệu lệnh được phân biệt dùng mạng nơon sẽ được sử dụng để điều khiển xe lăn điện di chuyển. Kết quả thực nghiệm 30 lần của những lệnh tiếng nói trên xe lăn điện đã cho sự chính xác từ khoảng 85% đến 90%. Nghiên cứu này có thể hỗ trợ người tàn tật trong việc di chuyển xe lăn của mình một cách dễ dàng và thuận tiện hơn trong cuộc sống hàng ngày và cũng cho thấy tính hiệu quả của phương pháp đề xuất.

Từ khóa - trích đặc trưng MFCC; phân loại tiếng nói; 5 lệnh tiếng nói; bộ lọc thông thấp Hamming; mạng truyền thăng đa lớp.

Abstract - This paper proposes a Multilayer Feedforward Neural Network (MLFNN) algorithm for speech classification in electrical wheelchair control, in which the feature extraction of speech signals is performed using a Mel Frequency Cepstral Coefficients (MFCC) method. The speech signals are the control commands recognized such as *Left*, *Right*, *Forward*, *Backward* and *Stop*. In addition, a Hamming lowpass filter is applied to frame and reduce noise before feature extraction. Therefore, the recognized signals will be used to control the electrical wheelchair. Experimental results of 30 time experiments on an electrical wheelchair produced the accuracy from 85% to 90%. This study can support disabled people in moving their wheelchairs easily and more conveniently in everyday life and also shows the effectiveness of the proposed approach.

Key words - MFCC feature extraction; Speech classification; 5 speech commands; Hamming lowpass filter; Multiplayer Feedforward Neural Networks.

1. Giới thiệu

Người tàn tật ngày càng tăng trên khắp thế giới và họ có những khó khăn trong cuộc sống cộng đồng. Thiết bị trợ giúp luôn luôn hữu ích với họ trong các hoạt động hàng ngày. Trên thế giới, số lượng người tàn tật chiếm khoảng 15% dân số. Hơn nữa, người khuyết tật cảm thấy bị cô lập và không được tiếp cận với những cơ hội như nhau trong cộng đồng của họ. Đó là lý do tại sao một chiếc xe lăn thông minh là cần thiết cho họ.

Để thuận tiện hơn trong cuộc sống hiện đại, xe lăn điện được cải thiện từng ngày và một chiếc xe lăn thông minh là cần thiết cho người tàn tật [1]. Xe lăn thông minh được thiết kế để sử dụng cho môi trường trong nhà và người dùng có thể dễ dàng điều khiển nó bằng cách ra lệnh sử dụng giọng nói của mình [2].

Nhận dạng tiếng nói là một phương pháp tích cực và phổ biến, nó giúp chuyển dịch những tiếng nói của con người sang những lệnh cho điều khiển. Mô hình thường được sử dụng trong nhận biết như: mô hình Markov ẩn (Hidden Markov Model_HMM) [3], lượng tử vector (Vector Quantization), trích đặc trưng dùng MFCC-DTW và mạng nơon [4]. Nhận dạng được sử dụng trong điều khiển xe lăn điện, nhận biết ký tự [5], hoặc đếm số [6]. Hơn nữa, mô hình có độ chính xác, tuy nhiên nó rất phức tạp và mất thời gian huấn luyện. Trong khi mô hình MFCC và chuỗi xoắn thời gian động (Dynamic Time Warping_DTW) là đơn giản [7], không mất nhiều thời gian huấn luyện, nhưng độ chính xác thì thấp hơn mô hình HMM. Hệ thống nhận dạng giọng nói được sử dụng để phát hiện tiếng nói trong nền của nhiều tiếng ồn. Điểm ban đầu và cuối của một từ được phát hiện cho việc nhận dạng từ. Khó khăn chính của

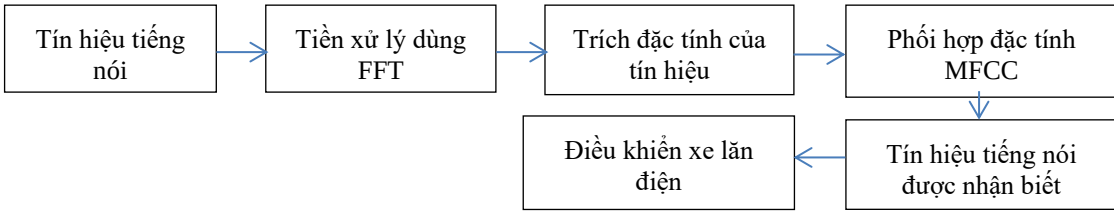
nhận dạng giọng nói là cùng một từ được nói bởi những người nói khác nhau tùy thuộc vào những cách nói, giọng điệu, khu vực, giới tính và các mẫu nói. Ngoài ra, nhiễu và sự thay đổi của tín hiệu theo thời gian là vấn đề được xem xét trong nhận dạng giọng nói.

Nhận dạng tiếng nói đóng vai trò quan trọng trong hệ thống xe lăn điện sử dụng microphone. Thuật toán MFCC và DTW được áp dụng cho trích đặc trưng và nhận biết tiếng nói [8]. Những lệnh tiếng nói như *Left*, *Right*, *Forward*, *Backward* và *Stop* sẽ được nhận biết cho điều khiển xe lăn điện. Những thí nghiệm với lệnh tiếng nói được xác định bằng phương pháp đề xuất có thể được người sử dụng thực hiện trong điều khiển xe lăn. Trong thực tế, khi người sử dụng muốn di chuyển xe lăn thì có thể sử dụng lệnh tiếng nói bằng việc nói vào microphone [9] thông qua máy tính để điều khiển xe lăn.

Mạng nơon được áp dụng cho xác định các tín hiệu dựa vào những đặc trưng của chúng sử dụng MFCC. Trong nghiên cứu này, các tác giả đã sử dụng mạng nơon để nhận biết tiếng nói [11-12]. Trong bài báo này, thuật toán nhận dạng dùng mạng nơon ALFNN sẽ được áp dụng trong xe lăn điện để phân loại những tín hiệu tiếng nói tương ứng với những lệnh điều khiển như là *Left*, *Right*, *Forward*, *Backward* và *Stop*. Kết quả thí nghiệm trên xe lăn sẽ được đưa ra để minh họa cho tính hiệu quả của phương pháp được đề xuất.

2. Phương pháp nhận biết tín hiệu tiếng nói

Một người sử dụng sẽ thực hiện việc điều khiển xe lăn điện sử dụng lệnh tiếng nói như: *Left*, *Right*, *Forward*, *Backward* và *Stop* cho sự di chuyển. Một tín hiệu tiếng



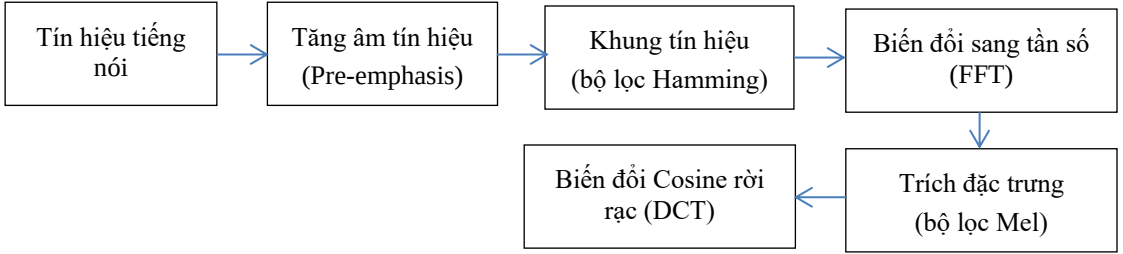
Hình 1. Sơ đồ khối của quá trình nhận dạng tiếng nói

nói được ghi trong khoảng thời gian 3s và một người đã tham gia thực hiện mỗi lệnh 30 lần. Tất cả các tín hiệu lệnh được tiền xử lý tại tần số lấy mẫu 16KHz trước khi trích đặc trưng cho việc nhận biết mỗi lệnh tiếng như Hình 1. Trong tiền xử lý, tín hiệu tiếng nói sử dụng sự kết

hợp của các đặc trưng dựa vào MFCC và dò tìm giọng nói. Cuối cùng là mạng nơron được sử dụng cho nhận dạng trong điều khiển xe lăn.

2.1. Trích đặc trưng và biến đổi thời gian

Hình 2 là biểu diễn trích đặc trưng sử dụng MFCC.



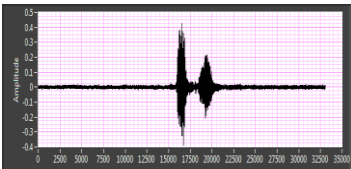
Hình 2. Sơ đồ khối của quá trình trích đặc trưng

2.1.1. Tăng âm tín hiệu

Bước này xử lý với mục đích chỉnh thành phần tần số cao. Cụ thể, tín hiệu tiếng nói được xử lý sử dụng một bộ lọc để làm tăng lên những tín hiệu tần số cao hơn. Nó có nghĩa, quá trình này sẽ làm tăng năng lượng của tín hiệu ở tần số cao hơn. Vậy, tín hiệu ngõ ra của bộ tăng âm được tính toán như sau:

$$H[n] = u(n) - au(n-1) \quad (1)$$

Ở đây, $H[n]$ là tín hiệu đầu ra của quá trình làm tăng, $u(n)$ là tín hiệu tiếng, giá trị của $a=0,95$ (độ lợi lớn hơn 20dB cho tần số cao). Kết quả của quá trình này như biểu diễn trong Hình 3.



Hình 3. Tín hiệu sau khi tăng âm

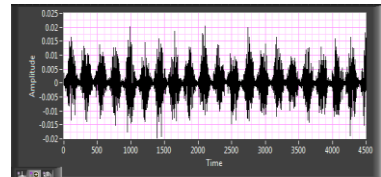
2.1.2. Đặt khung tín hiệu

Các tín hiệu sau khi tăng âm được phân đoạn do tín hiệu tiếng nói là liên tục với thời gian. Độ tin cậy giọng nói có thể được đảm bảo trong một khoảng thời gian ngắn. Khung quá trình không thể chờ đợi cho mẫu trước đó, mà nó phân chia sự gián đoạn tín hiệu ở đầu và cuối của mỗi khung tín hiệu, trong đó chiều dài khung từ 10 đến 30ms. Tín hiệu tiếng nói được chia thành các khung của N mẫu. Quá trình này là rất quan trọng để giữ những đặc trưng thời gian ngắn. Phân tích thời gian ngắn được thực hiện bởi việc phân chia cửa sổ tín hiệu như trong Hình 4. Cửa sổ Hamming với $(W(n), 0 \leq n \leq N-1)$ được sử dụng in phương trình như sau:

$$Y(n) = H(n) \times W(n) \quad (2)$$

Trong đó, N là số mẫu trong mỗi khung hình, $Y(n)$ mô tả tín hiệu đầu ra và $H(n)$ là tín hiệu đầu vào. Cụ thể, kết quả của hàm Hamming được biểu diễn như sau:

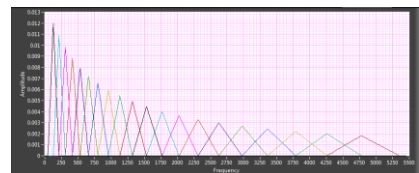
$$W(n) = 0,54 - 0,46 \cos\left(\frac{2\pi n}{N-1}\right) \quad (3)$$



Hình 4. Tín hiệu tiếng nói sau khi đặt khung dùng bộ lọc Hamming

2.1.3. Dây bộ lọc Mel

Việc nghe âm thanh của người thì không nhạy cảm như nhau với tất cả các dải tần số. Nó thường kém nhạy cảm hơn với những tín hiệu ở tần số cao hơn khoảng 1kHz và cảm nhận tần số giọng nói của người là phi tuyến tính. Phổ Mel là tổng của phổ tín hiệu sau khi biến đổi Fourier rời rạc nhân với trọng số của bộ lọc Mel. Dây của bộ lọc Mel là chuỗi của bộ lọc tam giác với nhiều tần số trung tâm và giảm tuyến tính đến không [9] như trong Hình 5.



Hình 5. Dây bộ lọc Mel với dây tần số từ 50 đến 5400 Hz, số dây bộ lọc là 20

Mỗi đầu ra của bộ lọc là tổng của các thành phần phổ được lọc và phương trình được sử dụng để tính cho thành phần Mel với tần số f :

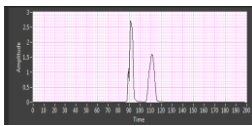
$$f(\text{mel}) = 2595 \log_{10} \left(1 + \frac{f}{100} \right) \quad (4)$$

2.1.4. Biến đổi cosine rời rạc

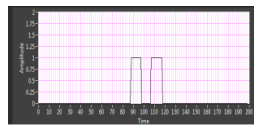
Trong bước cuối cùng này, phổ Mel (mel) với mỗi tần số được lấy logarit trước khi được biến đổi sang thời gian. Bởi vì những hệ số phổ Mel là số thực, chúng ta có thể biến đổi chúng sang miền thời gian bằng cách sử dụng DCT. Tập hợp các hệ số được gọi là vectơ có chứa những giá trị âm học. Vậy thì mỗi lời nói vào được biến đổi thành một chuỗi các vectơ râm.

DCT được thực hiện trên những hệ số phổ Mel[10] của mỗi khung và người ta thu được các MFCC. Hai hệ số đầu tiên của tập MFCC thu được được loại bỏ khi chúng thay đổi đáng kể giữa những phát âm khác nhau của cùng một từ. Hệ số đầu tiên của MFCC của mỗi khung được thay thế bằng năng lượng log của khung đó. Những hệ số Delta được tìm thấy từ MFCC để tăng miền vectơ đặc trưng của các khung tín hiệu cho việc tăng độ chính xác. Năng lượng trong một khung cho một tín hiệu của một cửa sổ được lưu trữ trong một mảng. Các giá trị này được sử dụng để phát hiện năng lượng ngưỡng của tín hiệu tiếng nói và loại bỏ nhiễu như được biểu diễn trong Hình 6a và 6b. Năng lượng được biểu diễn bằng phương trình sau:

$$E = \sum_{n=0}^{N-1} x^2(n) \quad (5)$$



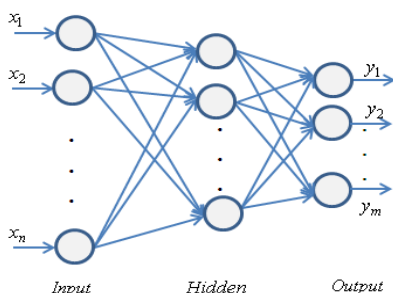
Hình 6a. Năng lượng tín hiệu



Hình 6b. Năng lượng tín hiệu được làm trơn

2.2. Thuật toán AFLNN

Việc nhận biết lệnh tiếng nói, một mô hình nơron truyền thẳng đa lớp (ALFNN) với thuật toán huấn luyện lan truyền ngược được áp dụng cho việc huấn luyện như Hình 7. Mạng lan truyền ngược là làm tối thiểu hàm sai lệch trong không gian trọng số sử dụng phương pháp làm giảm gradient. Bởi phương pháp này tính toán gradient của hàm sai lệch liên tục ở mỗi lần lặp cho đến khi hàm sai lệch không thể chia được nữa [11].



Hình 7. Sơ đồ mạng nơron truyền thẳng đa lớp

Nói rõ hơn, thuật toán lan truyền ngược được áp dụng để tìm cực tiểu cục bộ của hàm sai lệch. Vì vậy, gradient của hàm sai lệch được tính toán để thay đổi các giá trị trọng số ban đầu cho mạng. Các trọng số này là các thông số được thay đổi để giảm đi sai lệch. Trong nghiên cứu này, số lượng của nơron lớp ẩn được tính phụ thuộc vào nhiều thông số như số lượng các nút đầu vào-ra và tập mẫu huấn luyện. Vậy thì, số nơron lớp ẩn được xác định bằng phương trình sau đây:

$$H = \frac{Q}{5(I + O)} \quad (6)$$

trong đó, I là số nút đầu vào, O mô tả kích thước ngõ ra và Q là độ dài của tập huấn luyện.

3. Kết quả thực hiện

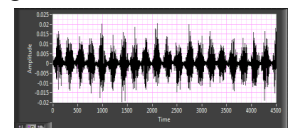
3.1. Trích đặc trưng cho tín hiệu tiếng nói

Sau khi thực hiện thuật toán kiến nghị, kết quả thực nghiệm những tín hiệu tiếng nói được thu với hiệu suất khả quang. Cụ thể, một người sử dụng xe lăn nói vào microphone những từ như *Left*, *Right*, *Forward*, *Backward* và *Stop*, và hệ thống máy tính sẽ lưu những từ này. Những từ ngõ vào này sẽ được nhận biết tương ứng với các mẫu có điểm phối hợp thấp nhất. Thuật toán ALFNN được sử dụng cho tính toán và nhận biết khoảng cách giữa tiếng nói đã được kiểm tra và những tập dữ liệu từ tiếng nói tham chiếu. Trong xử lý tín hiệu tiếng nói này, các tín hiệu được xử lý bằng bộ lọc thông thấp với khoảng tần số cắt [80, 1200] Hz. Hơn nữa, các tín hiệu có tốc độ lấy mẫu 11025Hz [12] và tiếng nói được ghi trong khoảng thời gian 3s cho mỗi từ.

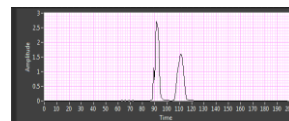
Hình 8a là tín hiệu đầu vào ban đầu đã được xử lý tăng âm và Hình 8b biểu diễn việc lấy khung của tín hiệu, trong đó khung này là để phân chia các tín hiệu không liên tục của đầu và cuối mỗi khung có độ dài từ 10ms đến 30ms. Hình 8c là tín hiệu năng lượng được làm nhẵn bằng một ngưỡng và Hình 8d biểu diễn tín hiệu được nhận biết dùng mạng nơron. Tất cả các tín hiệu tiếng nói sau khi được tiền xử lý và trích đặc trưng được đưa vào mạng nơron cho phân biệt lệnh tiếng.



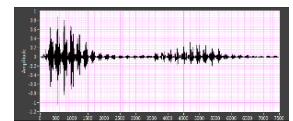
Hình 8a. Tín hiệu vào đã được tăng âm



Hình 8b. Tín hiệu sau khi đặt khung



Hình 8c. Tín hiệu năng lượng

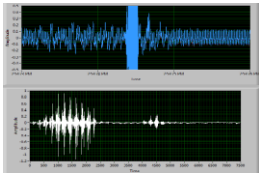


Hình 8d. Tín hiệu được nhận biết

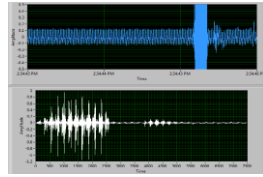
3.2. Nhận dạng tiếng nói dùng mạng nơron

Sau khi trích các đặc trưng tiếng nói sử dụng thuật toán MFCC, người ta thu được 13 hệ số đặc trưng, 13 phần tử năng lượng và 13 hệ số delta ngữ âm. Tuy nhiên,

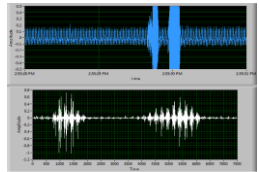
hai hệ số đặc trưng đầu tiên bị loại bỏ do phát âm khác nhau của cùng một từ. Những vector đặc trưng được đưa vào mạng của mạng nơron cho nhận biết.



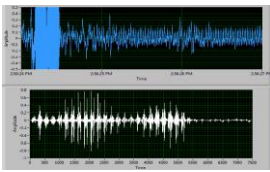
Hình 9a. Tín hiệu Trái trước và sau khi được nhận biết



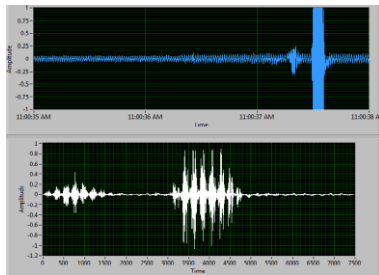
Hình 9b. Tín hiệu Phải trước và sau khi được nhận biết



Hình 9c. Tín hiệu Lui trước và sau khi được nhận biết



Hình 9d. Tín hiệu Tới trước và sau khi được nhận biết



Hình 9e. Tín hiệu Dừng trước và sau khi được nhận biết

Loại mạng nơron được sử dụng trong nghiên cứu này là thuật toán truyền thẳng (feedforward) đa lớp được huấn luyện dùng lan truyền ngược (back-propagation), bao gồm một lớp đầu vào, một lớp ẩn và một lớp ngõ ra. Số lượng nơron đầu vào được thiết lập với 35 giá trị của vector đặc trưng và số nơron đầu ra là 5 tương ứng với các lệnh tiếng nói như *Left*, *Right*, *Forward*, *Backward* và *Stop*.

Trong thí nghiệm này, có 5 đầu ra, 5000 mẫu huấn luyện cho mỗi từ tiếng nói. Vậy thì, số lượng nơron ẩn lớp tối đa là 25 và mỗi nơron là một hàm tuyến tính. Hơn nữa, hiệu suất của mạng phụ thuộc vào chất lượng của tín hiệu liên quan đến tiền xử lý và trích đặc trưng. Kết quả của nhận dạng từ sử dụng ALFNN được thể hiện trong Hình 9a đến 9e.

Những từ tiếng nói đã được nhận biết sử dụng thuật toán ALFNN cho điều khiển xe lăn điện. Độ chính xác của phân loại từ tiếng nói được thực hiện trong 30 lần cho điều khiển xe lăn với hiệu suất như Bảng 1.

3.3. Điều khiển xe lăn điện

Tín hiệu ngõ ra của bộ nhận dạng truyền đến xe lăn điện cho điều khiển qua bộ chuyển đổi Số-Tương tự USB-6008 và đến động cơ của chiếc xe lăn điện để điều khiển chuyển động xe với những lệnh như *Left*, *Right*, *Forward*, *Backward* và *Stop*. Mô hình của xe lăn điện được biểu diễn trong Hình 10 dùng những lệnh tiếng nói để di chuyển xe.



Hình 10. Người sử dụng đang điều khiển xe lăn sử dụng tiếng nói

Bảng 1. Mức độ chính xác của những lệnh tín hiệu với số nơron của lớp ẩn khác nhau.

Từ tiếng nói	Nơron lớp ẩn	Chính xác (%)
Left	10	86,25
	15	98,75
Right	10	88,25
	15	94,50
Backward	10	82,00
	15	95,50
Forward	10	89,00
	15	95,25
Stop	10	88,25
	15	94,50

Kết quả nhận dạng dùng mạng nơron ALFNN tương ứng với sự thay đổi lớp ẩn với các nơron lần lượt là 10 và 15, ta có được các kết quả xác định chính xác khác nhau. Với 10 nơron lớp ẩn, kết quả nhận dạng là thấp hơn so với mạng lớp ẩn có 15 nơron, nhưng thời gian huấn luyện là nhanh hơn. Với thuật toán ALFNN này, độ chính xác của điều khiển xe lăn được thể hiện trong Bảng 1.

Nghiên cứu này còn đưa ra một sự so sánh nhỏ giữa hai thuật toán nhận dạng ALFNN và mô hình Marko vốn đã được thực hiện để nhận dạng từ tiếng nói của nghiên cứu trước đó [14]. Trong khi đó, phương pháp logic mờ được sử dụng trong việc nhận biết tiếng Việt với đơn âm dựa vào hình dạng phổ của mỗi âm [15]. Trong việc điều khiển xe lăn điện dùng tiếng nói [16], các tác giả đã thiết kế bộ điều khiển dùng card DSP cho việc lọc nhiễu và bộ vi điều khiển cho điều khiển với 10 lệnh [17]. Tuy nhiên hệ thống khá phức tạp.

Trong nghiên cứu này, phương pháp mạng nơron truyền thẳng MLFNN với 13 hệ số đặc trưng dùng MFCC, 13 hệ số năng lượng và 13 hệ số delta âm ngữ được sử dụng. Trong cụ thể, phép biến đổi Fourier nhanh (FFT) với 256 điểm được sử dụng để tìm phổ năng lượng của tín hiệu sử dụng bộ lọc Mel với 24 dãy. Cuối cùng, độ chính xác là khoảng 89% được so sánh với 82% dùng mô hình Markov trong một môi trường thông thoáng. Nghiên

cứu này cho điều khiển xe lăn điện với phương pháp trích đặc trưng MFCC với những hệ số rõ ràng cùng với bộ lọc Mel cho 5 lệnh điều khiển. Hơn nữa, mạng nơron ẩn được chọn trong nghiên cứu này là 15 đã cho một hiệu suất ổn định với độ chính xác khoảng 90%.

4. Kết luận

Trong bài báo này, tín hiệu tiếng nói đã được lọc bởi bộ lọc thông thấp Hamming và đặc trưng được trích cho việc phân loại từ tiếng nói sử dụng những hệ số Mel bằng phương pháp MFCC. Từ các hệ số MFCC, mô hình mạng nơron truyền thẳng MLFNN đã được sử dụng để phân loại những tín hiệu tiếng nói. Từ những lệnh tiếng nói như: *Left, Right, Forward, Backward* và *Stop* được người khuyết tật sử dụng để điều khiển xe lăn điện di chuyển trong môi trường trong nhà. Những kết quả thực nghiệm thu được cho thấy tính hiệu quả với thuật toán kiến nghị ALFNN với trích đặc trưng dùng MFCC.

TÀI LIỆU THAM KHẢO

- [1] Gunda Gautam, Gunda Sumanth, Karthikeyan K C, Shyam Sundar, D.Venkataraman "Eye Movement Based Electronic Wheel Chair For Physically Challenged Persons" *International Journal of Scientific & Technology Research* Vol. 3, No. 2, 2014.
- [2] Mohamed Fezari and Ibrahim M. El-Emary "Speech and Sensors in Improving the Guidance of an Electric Wheelchair by Handicapped Persons" *Proceedings of the 2014 International Conference on Circuits, Systems and Control*, 2014.
- [3] Bhupinder Singh, Neha Kapur, Puneet Kaur, "Speech Recognition with Hidden Markov Model: A Review", *International Journal of Advanced in Computer Science and Soft Engineering*, Vol. 2, pp. 400-403, 2012.
- [4] Nguyen Tu ha, Pham Van Tuan, "Design of Television Remote Control by Vietnamese Speech", *Tạp chí Khoa học và Công nghệ, Đại học Đà Nẵng*, 2014.
- [5] Dipmoy Gupta, Radha Mounima C. Navya Manjunath, Manoj PB, "Isolated Word Speech Recognition Using Vector Quantization", *International Journal of Advanced in Computer Science and Soft Engineering*, Vol. 2, pp. 164-168, 2012.
- [6] Talal Bin Amin, Iftekhar Mahmood, "Speech Recognition Using Dynamic Time Warping", *ICAST*, 2008.
- [7] A.Revathi, R.Ganapathy and Y.Venkataramani, "Text Independent Speaker Recognition and Speaker Independent Speech Recognition Using Iterative Clustering Approach", *International Journal of Computer science & Information Technology*, Vol. 1, No 2, pp. 30-42, 2009.
- [8] Lindasalwa Muda, Mumtaj Begam and I. Elamvazuthi, "Voice Recognition Algorithms using Mel Frequency Cepstral Coefficient (MFCC) and Dynamic Time Warping (DTW) Techniques", *Journal Of Computing*, Vol. 2, Issue 3, pp. 138-143, 2010.
- [9] Kohei Arai, Ronny Mardiyanto, Eyes Based Electric Wheel Chair Control System, *International Journal of Advanced Computer Science and Applications*, 2011.
- [10] S. M. Azam, Z.A. Mansoor, M. Shahzad Mughal, S. Mohsin, "Urdu Spoken Digits Recognition Using Classified MFCC and Backpropagation Neural Network", *Computer Graphics, Imaging and Visualisation IEEE*, Vol.1, pp. 7695-2928, 2007.
- [11] Chin Luh Tan and Adznan Jantan, "Digit Recognition Using Neural Networks", *Malaysian Journal of Computer Science*, Vol. 17, No. 2, pp. 40-54, 2004.
- [12] R. Rojas, *Neural Networks*. Berlin: Springer-Verlag, 1996.
- [13] Fu-Hua Liu; Richard M. Stern; Xuedong Huang; Alejandro Acero, "Efficient cepstral normalization for robust speech recognition, human language technology", *Proceedings of a Workshop Held at Plainsboro*, pp. 21-24, 1993.
- [14] Z. Hachkar et al., "A Comparison of DHMM and DTW for Isolated Digits Recognition System of Arabic Language", *International J. on Computer Science and Engineering*, 2011.
- [15] Tran Duc Minh, Nguyen Thien Luan, "Using Fuzzy Logic in Vietnamese Speech Recognition", *Tạp chí Khoa học và Công nghệ, Đại học Đà Nẵng*, 2014.
- [16] Anoop.K.J, Inbaezhilan, Sathish raj, Ramaseenivasan, CholaPandian, "Designing and Modeling of Voice Controlled Wheel Chair Incorporated with Home Automation" *International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering*, Vol. 17, No. 2, pp. 53-59, 2014.
- [17] Mohamed Fezari and Ibrahim M. El-Emary, "Speech and Sensors in Improving the Guidance of an Electric Wheelchair by Handicapped Persons", *Proceedings of the International Conference on Circuits, Systems and Control*, 2014.

(BBT nhận bài: 10/03/2015, phản biện xong: 25/08/2015)