TÌM HIỂU KỸ THUẬT NHẬN DẠNG GIỌNG NÓI BẰNG MẠNG NƠ RON NHÂN TẠO

**Trần Thị Loan Thảo, Võ Hoàng Ân, Nguyễn Thị Mỹ Xuyến**

Khoa Công nghệ thông tin, Trường Đại học An Giang

Email: [ttlthao\_20th@student.agu.edu.vn](mailto:%20ttlthao_20th@student.agu.edu.vn)

Email: [vhan\_20th@student.agu.edu.vn](mailto:vhan_20th@student.agu.edu.vn)

Email: [ntmxuyen\_20th@student.agu.edu.vn](mailto:ntmxuyen_20th@student.agu.edu.vn)

**Tóm tắt:** *Nhận dạng tiếng nói là mong ước của khoa học và con người. Những người máy có thể hiểu được tiếng người nói và thực thi nhiệm vụ. Ứng dụng nhận dạng tiếng nói để điều khiển thiết bị là một lĩnh vực thiết thực trong cuộc sống. Có nhiều phương pháp tiếp cận đến nhận dạng tiếng nói, song do tính phức tạp vốn có của mỗi ngôn ngữ và mỗi chất giọng của từng dân tộc, mà lĩnh vực này luôn là một thách thức to lớn đối với những người đam mê. Hiện nay, nhận dạng tiếng nói chưa thực sự đáp ứng đầy đủ các yêu cầu thực tế, song những hệ thống nhận dạng tiếng nói đã có bước phát triển đáng kể. Bài báo đề cập đến một giải pháp ứng dụng mạng nơ-ron nhân tạo (Artificial neural networks) để điều khiển thiết bị bằng giọng nói tiếng Việt. Phép biến đổi Fourier thời gian ngắn - STFT (Shorttime Fourier Transform) được áp dụng để trích các đặc trưng cơ bản của tín hiệu tiếng nói. Một mạng nơ-ron nhân tạo được huấn luyện để nhận dạng giọng nói tiếng Việt của bất kỳ người nào, khi họ đọc một trong bốn từ lệnh ‘Trái’, ‘Phải’, ‘Tới’ và ‘Lui’ (áp dụng để điều khiển robot). Kết quả nghiên cứu được kiểm chứng thông qua việc điều khiển từ xa một xe vô tuyến. Ðộ chính xác được ước lượng xấp xỉ 90% và khả năng mở rộng tập lệnh điều khiển là rất cao. Xuất phát từ nhận thức trên, đề tài của nhóm em là tìm hiểu, đưa ra phương pháp và xây dựng một ứng dụng nhận dạng tiếng nói tiếng Việt. Với những khả năng của mạng nơron nhân tạo trong ứng dụng, đã cho em biết thêm một vài thông tin hữu ích. Vì vậy, nhóm em đã chọn mạng nơron làm cơ sở cho việc nghiên cứu nhận dạng tiếng nói tiếng Việt.*

Từ khóa: Nhận dạng giọng nói, mạng nơ ron nhân tạo, hệ thống điều khiển, vi điều khiển, chu kỳ cao độ

# 1. GIỚI THIỆU

- Nhận dạng tiếng nói là mong ước của khoa học và con người. Những người máy có thể hiểu được tiếng người nói và thực thi nhiệm vụ.

- Ứng dụng nhận dạng tiếng nói để điều khiển thiết bị là một lĩnh vực thiết thực trong cuộc sống. Có nhiều phương pháp tiếp cận đến nhận dạng tiếng nói, song do tính phức tạp vốn có của mỗi ngôn ngữ và mỗi chất giọng của từng dân tộc, mà lĩnh vực này luôn là một thách thức to lớn đối với những người đam mê.

- Hiện nay, nhận dạng tiếng nói chưa thực sự đáp ứng đầy đủ các yêu cầu thực tế, song những hệ thống nhận dạng tiếng nói đã có bước phát triển đáng kể.

- Trên thế giới, một số hệ thống nhận dạng tiếng nói cỡ lớn có độ chính xác tương đối cao. Các hệ thống này chủ yếu được phát triển trên nền công nghệ hiện đại với những máy tính lớn, những vi mạch xử lý tiếng nói chuyên dụng và sử dụng cơ sở dữ liệu tiếng nói khá hoàn chỉnh, nhưng phần lớn vẫn là xử lý cho tiếng Anh.

- Ở Việt Nam, việc tìm hiểu, nghiên cứu và phát triển các hệ thống nhận dạng tiếng nói còn đang bước đầu có kết quả. Do có những đặc thù riêng của tiếng Việt, nên việc chọn lựa phương pháp tiếp cận bài toán nhận dạng phù hợp với tiếng Việt là một vấn đề tương đối khó khăn.

- Những năm gần đây, cũng có khá nhiều đề tài nghiên cứu về nhận dạng tiếng nói tiếng Việt. Các hệ thống nhận dạng tiếng nói thành công nhất chủ yếu dựa trên khuynh hướng nhận dạng mẫu. Các kỹ thuật nhận dạng mẫu đơn giản như lượng tử hoá véctơ, hiệu chỉnh thời gian động…, đã được áp dụng khá thành công vào các chương trình nhận dạng tiếng nói tiếng Việt phát âm rời rạc với số lượng từ vựng hạn chế.

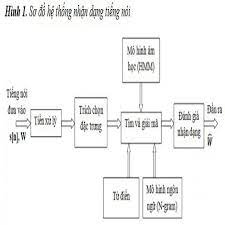
- Tuy nhiên, mục tiêu của nhận dạng tiếng nói tự động bằng máy là phải tiến tới hệ thống nhận dạng tiếng nói liên tục, kích thước từ điển lớn, không phụ thuộc vào người nói. Vì vậy, các hệ thống nhận dạng tiếng nói hiện nay thường xây dựng trên cơ sở áp dụng các kỹ thuật nhận dạng mẫu phức tạp hơn, đó là mô hình Markov ẩn và mạng nơron nhân tạo đã cho một số thành công nhất định.

- Xuất phát từ nhận thức trên, đề tài của nhóm em là tìm hiểu, đưa ra phương pháp và xây dựng một ứng dụng nhận dạng tiếng nói tiếng Việt. Với những khả năng của mạng nơron nhân tạo trong ứng dụng, đã cho nhóm em biết thêm một vài thông tin hữu ích. Vì vậy, nhóm em đã chọn mạng nơron làm cơ sở cho việc nghiên cứu nhận dạng tiếng nói tiếng Việt.

# 2. PHƯƠNG PHÁP

**2.1 Mô hình tổng quan về kỹ thuật nhận dạng giọng nói bằng mạng nơ ron nhân tạo**

Chúng ta có thể thấy được một cách trực quan bài toán nhận dạng tiếng nói qua hình 1



**Hình 1. Sơ đồ về nhận dạng giọng nói**

## 2.2 Thuật toán phân loại văn bản và text

• Input: Tiếng nói đưa vào

• Quá trình xử lý: Tiền xử lý dữ liệu -> Trích đặc trưng tín hiệu tiếng nói -> Sau khi trải qua các bước như Mô hình âm học (HMM), Từ điển, Mô hình ngôn ngữ sẽ Tìm và giải mã

• Output: Đánh giá nhận dạng -> Đầu ra

# 3. Ứng dụng minh họa

Trong phần này chúng em giới thiệu ứng dụng thực nghiệm về kỹ thuật nhận dạng giọng nói bằng mạng nơ ron nhân tạo của nhóm tác giá Nguyễn Chí Ngôn và cộng sự Trần Thanh Hùng, Trương Thị Thanh Tuyền và Nguyễn Thái Nghe ([1]).

## 3.1 Môi trường thực nghiệm

## 3.2 Một số kết quả thực nghiệm

- Hiện tại chúng em chỉ xây dựng mạng với số lượng từ nhận dạng còn ít (4 từ) dựa trên giọng đọc của 10 người (gồm 9 nam và 1 nữ, chất giọng miền tây nam bộ). Đây cũng là điểm hạn chế của đề tài, vì cơ sở dữ liệu không đủ tổng quát để áp dụng cho những miền khác của Việt nam. Hơn nữa, quá trình huấn luyện, mạng chỉ phân loại một lần tín hiệu đầu vào và chia ra thành 4 nhóm, đặc trưng cho 4 từ: ‘Tới’, ‘Lui’, ‘Trái’ và ‘Phải’. Tuy nhiên, kết quả nghiên cứu cho thấy khả năng tăng số lượng từ nhận dạng là khả thi. Trong trường hợp đó, chúng tôi dự kiến sẽ phân lớp dữ liệu trước khi nhận dạng; sao cho mỗi lớp chứa các từ có đặc trưng gần giống nhau. Chẳng hạn, chúng em dùng một mạng tổng quát để phân biệt từ vừa đọc thuộc nhóm “thanh bằng” hay “thanh trắc”, sau đó đưa vào mạng chuyên biệt để nhận dạng chính xác từ vừa đọc. Ngoài ra chúng em thấy rằng, kiểm tra trên nhóm người có giọng đọc đã được sử dụng để huấn luyện, mạng nơ-ron làm việc khá hiệu quả. Vì thế, chúng em đề xuất 1 phương án

**Bảng 1: Kết quả nhận dạng trên nhóm người có giọng nói đã dùng để rèn luyện**

| **Từ lệnh** | **Số lần đọc** | **Số lần nhận dạng đúng** | **Tỉ lệ xấp xỉ** |
| --- | --- | --- | --- |
| Tới | 1000 | 961 | 96,10% |
| Lui | 800 | 755 | 94,37% |
| Trái | 1150 | 1037 | 90,17% |
| Phải | 800 | 794 | 99,25% |

Bảng 2: Kết quả nhận dạng trên nhóm người có giọng nói chưa dùng để huấn luyện

| **Từ lệnh** | **Số lần đọc** | **Số lần nhận dạng đúng** | **Tỉ lệ xấp xỉ** |
| --- | --- | --- | --- |
| Tới | 750 | 652 | 86,93% |
| Lui | 600 | 493 | 82,17% |
| Trái | 800 | 612 | 76,65% |
| Phải | 750 | 681 | 90,08% |

# KẾt luẬn

Kết quả: Sau khi huấn luyện, mạng được áp dụng để điều khiển từ xa một xe vô tuyến. Một chương trình tự động ghi nhận kết quả nhận dạng (sau khi được xác nhận của người điều khiển), được áp dụng. Thống kê trên 1000 lần đọc các lệnh ‘Tới’, ‘Lui’, ‘Trái’,’Phải’ của nhiều người, chúng tôi ước lượng được độ chính xác như sau (Bảng 1 và Bảng 2):  
- 95% đối với nhóm người có giọng đọc đã được sử dụng để huấn luyện mạng.  
- 84% đối với nhóm người có giọng đọc chưa được sử dụng để huấn luyện mạng. Tuy nhiên, chúng ta có thể ghi âm giọng nói của những người này và huấn luyện tiếp để cải thiện tỉ lệ lỗi.

Phương pháp: Bài báo đề cập đến một giải pháp ứng dụng mạng nơ-ron nhân tạo (Artificial neural networks) để điều khiển thiết bị bằng giọng nói tiếng Việt. Phép biến đổi Fourier thời gian ngắn - STFT (Shorttime Fourier Transform) được áp dụng để trích các đặc trưng cơ bản của tín hiệu tiếng nói. Một mạng nơ-ron nhân tạo được huấn luyện để nhận dạng giọng nói tiếng Việt của bất kỳ người nào, khi họ đọc một trong bốn từ lệnh ‘Trái’, ‘Phải’, ‘Tới’ và ‘Lui’ (áp dụng để điều khiển robot)

Công việc trong tương lai: chưa nghĩ ra

**LỜI CẢM ƠN**

Phần này xin được gởi lời cảm ơn chân thành đến 2 thành viên đã giúp đỡ trong quá trình hoàn thành bài báo cáo.

**TÀI LIỆU THAM KHẢO**

[1] Ngôn N. C., Hùng T. T., and Tuyền T. T. T., “ỨNG DỤNG MẠNG NƠ-RON NHÂN TẠO ÐỂ ÐIỀU KHIỂN THIẾT BỊ BẰNG GIỌNG NÓI TIẾNG VIỆT,” p. 8.

[2] Hoan N. Q., “Nhận dạng tiếng nói trên cơ sở mạng Nơron nhân tạo”.