

ĐẠI HỌC THÁI NGUYÊN
TRƯỜNG ĐẠI HỌC CÔNG NGHỆ THÔNG TIN VÀ TRUYỀN THÔNG

NGUYỄN THỊ THU HUYỀN

**MÔ HÌNH MARKOV ẨN VÀ ỨNG DỤNG
XÂY DỰNG HỆ THỐNG NHẬN DẠNG TIẾNG NÓI**

LUẬN VĂN THẠC SĨ KHOA HỌC MÁY TÍNH

THÁI NGUYÊN - 2018

ĐẠI HỌC THÁI NGUYÊN
TRƯỜNG ĐẠI HỌC CÔNG NGHỆ THÔNG TIN VÀ TRUYỀN THÔNG

NGUYỄN THỊ THU HUYỀN

**MÔ HÌNH MARKOV ẨN VÀ ỨNG DỤNG
XÂY DỰNG HỆ THỐNG NHẬN DẠNG TIẾNG NÓI**

Chuyên ngành: Khoa học máy tính

Mã số: 8480101

LUẬN VĂN THẠC SĨ KHOA HỌC MÁY TÍNH

Người hướng dẫn khoa học: TS. Vũ Vinh Quang

THÁI NGUYÊN - 2018

LỜI CAM ĐOAN

Tôi xin cam đoan rằng đây là công trình nghiên cứu của tôi, có sự hỗ trợ từ Giáo viên hướng dẫn là TS. Vũ Vinh Quang . Các nội dung nghiên cứu và kết quả trong đề tài này là trung thực và chưa từng được ai công bố trong bất cứ công trình nghiên cứu nào trước đây. Những số liệu trong các hình phục vụ cho việc phân tích, nhận xét, đánh giá được chính tác giả thu thập từ các nguồn khác nhau có ghi trong phần tài liệu tham khảo. Ngoài ra, đề tài còn sử dụng một số nhận xét, đánh giá cũng như số liệu của các tác giả, cơ quan tổ chức khác, và cũng được thể hiện trong phần tài liệu tham khảo.

Nếu phát hiện có bất kỳ sự gian lận nào tôi xin hoàn toàn chịu trách nhiệm trước Hội đồng, cũng như kết quả luận văn của mình.

Thái nguyên, ngày tháng năm 2018

Tác giả

NGUYỄN THỊ THU HUYỀN

LỜI CẢM ƠN

Để hoàn thành luận văn này, em xin tỏ lòng biết ơn sâu sắc đến thầy **TS. Vũ Vinh Quang**, đã tận tình hướng dẫn trong suốt quá trình viết luận văn tốt nghiệp.

Em chân thành cảm ơn quý thầy, cô trong trường Đại Học Công nghệ Thông tin và Truyền thông đã tận tình truyền đạt kiến thức trong hai năm học tập. Với vốn kiến thức được tiếp thu trong quá trình học là nền tảng cho quá trình nghiên cứu để em hoàn thành luận văn.

Thái nguyên, ngày tháng năm 2018

Tác giả

NGUYỄN THỊ THU HUYỀN

MỤC LỤC

LỜI CAM ĐOAN	i
LỜI CẢM ƠN	ii
DANH MỤC CÁC HÌNH ẢNH	v
DANH MỤC CÁC BẢNG BIỂU	vi
DANH MỤC CÁC TỪ VIẾT TẮT	vii
LỜI MỞ ĐẦU	1
CHƯƠNG 1. MỘT SỐ KIẾN THỨC CƠ BẢN VỀ NHẬN DẠNG TIẾNG NÓI	3
1.1. Khái niệm chung	3
1.1.1. Khái niệm nhận dạng tiếng nói	3
1.1.2. Phân loại nhận dạng tiếng nói	4
1.1.3. Hệ thống nhận dạng tiếng nói tự động	7
1.2. Các nghiên cứu hiện thời về nhận dạng tiếng nói	8
1.2.1. Các nghiên cứu về nhận dạng tiếng nói ở nước ngoài	8
1.2.2. Các nghiên cứu về nhận dạng tiếng nói tiếng Việt	9
1.3. Các hướng tiếp cận nhận dạng tiếng nói	10
1.4. Những khó khăn trong nhận dạng tiếng nói	11
1.5. Xử lý tiếng nói	13
1.5.1. Khái niệm xử lý tiếng nói	13
1.5.2. Kỹ thuật lấy mẫu tín hiệu	13
1.5.3. Cấu trúc bộ lọc tín hiệu	13
1.5.4. Kỹ thuật dò tìm điểm cuối	15
1.5.5. Kỹ thuật rút trích đặc trưng	16
1.5.6. Phương pháp tính hệ số MFCC	18
1.6. Sơ lược về ngôn ngữ tiếng Việt	23
1.6.1. Giới thiệu về âm tiết	23
1.6.2. Một số đặc điểm của âm tiết tiếng Việt	23
1.6.3. Lý thuyết về âm vị	25
CHƯƠNG 2. TỔNG QUAN VỀ MÔ HÌNH MẠNG MARKOV ẨN	27

2.1. Mô hình Markov ẩn.....	27
2.2. Các bài toán cơ bản của mô hình Markov ẩn.....	30
2.2.1. Bài toán 1	30
2.2.2. Bài toán 2	30
2.2.3. Bài toán 3	31
2.3. Các thuật toán cơ bản	31
2.3.1. Thuật toán tiến (forward)	31
2.3.2. Thuật toán lùi (Backward)	32
2.3.3. Các giải pháp giải quyết các bài toán cơ bản	32
2.4. So sánh các loại mô hình Markov ẩn	33
CHƯƠNG 3. XÂY DỰNG HỆ THỐNG NHẬN DẠNG TIẾNG VIỆT.....	35
3.1. Tổng quan về HTK (HMM Tool Kit)	35
3.1.1. Giới thiệu hệ thống.....	35
3.1.2. Mô hình các cấu trúc tập tin cơ bản khai báo HMM và MFCC cho nhận dạng tiếng nói	36
3.1.3. Các bước xây dựng một mô hình nhận dạng tiếng nói sử dụng HTK	40
3.1.4. Một số Modul sử dụng trong quá trình xây dựng hệ thống nhận dạng tiếng nói tiếng việt trong bộ công cụ HTK.....	44
3.2. Xây dựng hệ thống nhận dạng chữ số Tiếng việt.....	47
3.2.1. Xây dựng cơ sở dữ liệu chữ số tiếng việt.....	47
3.2.2. Bảng phiên âm 10 chữ số tiếng Việt	47
3.2.3. Phương pháp xây dựng hệ thống nhận dạng chữ số tiếng việt	47
3.2.4. Các kết quả thực nghiệm.....	48
KẾT LUẬN	51
TÀI LIỆU THAM KHẢO	52

DANH MỤC CÁC HÌNH ẢNH

Hình 1.1: Quá trình phát âm sẽ khác nhau tùy theo người nói	6
Hình 1.2: Mô hình nhận dạng tiếng nói bán độc lập người nói	6
Hình 1.3: Cấu trúc cơ bản của hệ thống ASR	7
Hình 1.4: Cấu trúc hệ thống nhận dạng ngôn ngữ có thanh điệu	9
Hình 1.5. Ví dụ về lấy mẫu tín hiệu $F(t)$ trên miền thời gian	13
Hình 1.6. Minh họa hoạt động của bộ lọc FIR	14
Hình 1.7. Minh họa hoạt động của bộ lọc IIR	15
Hình 1.8. Đồ tìm điểm cuối dựa trên mức năng lượng	16
Hình 1.9. Sơ đồ rút trích vevtor đặc trưng tổng quát	17
Hình 1.10. Các bước xử lý tính hệ số MFCC	18
Hình 1.11. Sóng âm của chữ “ANH” trước khi làm phẳng (bên trái) và sau khi làm phẳng (Bên phải)	19
Hình 1.12. Phân đoạn tín hiệu tiếng nói thành các khung có sự chồng lấp	20
Hình 1.13. Cửa sổ Hamming	20
Hình 1.14. Cửa sổ Hanning	21
Hình 1.15. Cửa sổ hình chữ nhật	21
Hình 2.1: Minh họa mô hình Markov trong nhận dạng tiếng nói với 5 trạng thái và chèn sp	28
Hình 2.2: Mô hình Markov ẩn với sáu trạng thái	30
Hình 3.1: Minh họa sự giảm dần của năng lượng. Error! Bookmark not defined.	
Hình 3.2. Các Module và các chức năng trong HTK	35
Hình 3.3. Các công cụ và chức năng trong HTK	36
Hình 3.4. Phân bố các tham số trong một số vector đặc trưng của HTK	38
Hình 3.5. Các bước cơ bản xây dựng hệ thống nhận dạng tiếng nói	40
Hình 3.6: Mô hình HMM	42

DANH MỤC CÁC BẢNG BIỂU

Bảng 1.1: So sánh tỷ lệ lỗi nhận dạng của con người và máy tính trong một số hệ thống nhận dạng.....	8
Bảng 3.1: Bảng phiên âm 10 chữ số tiếng Việt	47
Bảng 3.2. Kết quả thử nghiệm hệ thống nhận dạng với bộ từ điển có chèn sp và không chèn sp	49
Bảng 3.3. Kết quả thử nghiệm hệ thống nhận dạng với dữ liệu test trùng với dữ liệu huấn luyện.....	50

DANH MỤC CÁC TỪ VIẾT TẮT

Từ viết tắt	Tiếng Anh	Tiếng Việt
AI	Artificial intelligence	Trí tuệ nhân tạo
AR	Augmented reality	Tương tác thực tại ảo
CSDL	Database	Cơ sở dữ liệu
DBMS	Database Management System	Hệ quản trị cơ sở dữ liệu
IoT	Internet of things	Internet vạn vật
IR	Information Retrieval	Hệ thống tự động truy tìm thông tin
MIRS	Multimedia Indexing & Retrieval System	Hệ thống chỉ mục và truy tìm thông tin đa phương tiện
MMDBMS	Multimedia Database Management System	Hệ thống quản trị cơ sở dữ liệu đa phương tiện
MRI	Magnetic Resonance Imaging	
VR	Virtual reality	Thực tế ảo
WWW	World Wide Web	

LỜI MỞ ĐẦU

Tiếng nói là một phương tiện giao tiếp cơ bản nhất của loài người, nó hình thành và phát triển cùng với quá trình tiến hóa của loài người. Ngay từ khi hình thành tiếng nói đã thể hiện được sức mạnh như: Khả năng truyền đạt thông tin, tốc độ...

Ngày nay, nhờ sự phát triển mạnh mẽ của khoa học kỹ thuật, máy móc dần dần thay thế các lao động bằng tay chân. Tuy nhiên để điều khiển máy móc, con người phải làm khá nhiều thao tác vừa tốn thời gian vừa phức tạp và cần phải được đào tạo. Chính vì lẽ đó, nhận dạng tiếng nói sẽ đóng vai trò rất quan trọng trong giao tiếp giữa người và máy. Nó giúp con người có thể điều khiển máy móc, thiết bị bằng giọng nói của mình đặc biệt là các thiết bị như: máy tính, điện thoại di động, rô bốt...

Hiện nay, cùng với sự phát triển vượt bậc của ngành khoa học nhận dạng tiếng nói, những chiếc điện thoại quay số bằng giọng nói đã ra đời, máy tính cá nhân với phần mềm nhận dạng tiếng nói có thể trợ giúp cho những người khuyết tật về tay, chân hay khiếm thính. . Tuy vậy, những thành tựu trên lại là những nghiên cứu về nhận dạng tiếng nói ngôn ngữ nước ngoài. Với tiếng Việt, số lượng các nghiên cứu còn khá ít và dường như ít được quan tâm. Có thể nói rằng những nghiên cứu nhận dạng tiếng nói tiếng Việt chủ yếu tập trung vào nhận dạng từ rời rạc, hay hệ thống nhận dạng liên tục nhưng với kích thước nhỏ.

Tuy nhiên, để ngành khoa học nhận dạng tiếng nói ở Việt Nam thực sự phát triển và đạt được những thành tựu to lớn để áp dụng vào thực tế thì lĩnh vực nhận dạng tiếng nói cần phát triển hơn nữa. Chính vì vậy, việc nghiên cứu nhận dạng tiếng nói đặc biệt là nhận dạng tiếng nói tiếng Việt là rất cần thiết để xây dựng các hệ thống nhận dạng, các ứng dụng nhận dạng tiếng nói thực sự hữu ích của riêng tiếng Việt.

Đề tài này được xây dựng với mục đích tìm hiểu, nghiên cứu, so sánh và đánh giá các phương pháp, các hướng tiếp cận nhận dạng tiếng nói hiện nay trên thế giới như: Tiếp cận âm học, tiếp cận trí tuệ nhân tạo, nhận dạng mẫu thống kê (dựa vào