

ĐẠI HỌC THÁI NGUYÊN
TRƯỜNG ĐẠI HỌC CÔNG NGHỆ THÔNG TIN VÀ TRUYỀN
THÔNG
LUẬN VĂN THẠC SĨ KHOA HỌC MÁY TÍNH

HỌC VIÊN: Trần Thị Tuyết

Lớp: cao học k13a

Người hướng dẫn: Phùng Trung Nghĩa

Đề tài: NGHIÊN CỨU PHƯƠNG PHÁP NHẬN DẠNG
NGƯỜI NÓI SỬ DỤNG KỸ THUẬT PHA TRỘN
GAUSSIAN

Thái Nguyên, 2016

LỜI CẢM ƠN

Lời đầu tiên, em xin chân thành cảm ơn thầy giáo TS. Phùng Trung Nghĩa, người đã trực tiếp hướng dẫn em hoàn thành luận văn. Với những lời chỉ dẫn, những tài liệu, sự tận tình hướng dẫn và những lời động viên của thầy đã giúp em vượt qua nhiều khó khăn trong quá trình thực hiện luận văn này.

Em cũng xin cảm ơn quý thầy cô giảng dạy chương trình cao học chuyên ngành "Khoa học máy tính" tại trường ĐH Công nghệ thông tin và truyền thông đã truyền dạy những kiến thức quý báu, những kiến thức này rất hữu ích và giúp em nhiều khi thực hiện nghiên cứu.

Cuối cùng, em xin gửi lời cảm ơn tới gia đình và bạn bè đã luôn ủng hộ động viên giúp đỡ em trong suốt những năm học vừa qua.

Em xin chân thành cảm ơn!

Thái Nguyên, ngày 10 tháng 05 năm 2016

Học viên

Trần Thị Tuyết

LỜI CAM ĐOAN

Tên tôi là: Trần Thị Tuyết

Sinh ngày: 20/05/1987

Học viên lớp cao học K13A - Trường Đại học Công nghệ thông tin và Truyền thông - ĐHTN.

Em xin cam đoan: Luận văn này là công trình nghiên cứu thực sự của cá nhân, được thực hiện dưới sự hướng dẫn khoa học của thầy giáo TS. Phùng Trung Nghĩa.

Các số liệu, những kết luận nghiên cứu được trình bày trong luận văn này trung thực và chưa từng được công bố dưới bất cứ hình thức nào.

Em xin chịu trách nhiệm về nghiên cứu của mình.

Học viên

Trần Thị Tuyết

MỤC LỤC

LỜI CẢM ƠN	
LỜI CAM ĐOAN	
MỤC LỤC	i
DANH MỤC BẢNG.....	iii
DANH MỤC HÌNH.....	iv
DANH MỤC CHỮ VIẾT TẮT VÀ KÍ HIỆU	vi
MỞ ĐẦU	1
CHƯƠNG I: TỔNG QUAN VỀ TIẾNG NÓI VÀ NHẬN DẠNG NGƯỜI NÓI TRONG TIẾNG NÓI	4
1.1. Tổng quan về tiếng nói	4
1.2. Tổng quan về lý thuyết nhận dạng tiếng nói.....	6
1.3. Thông tin người nói trong tiếng nói.....	8
1.4. Vấn đề nhận dạng và xác minh người nói qua giọng nói.....	9
1.4.1. Phân loại nhận dạng và xác thực người nói dựa vào chức năng của bài toán.....	9
1.4.2. Phân loại nhận dạng và xác thực người nói dựa theo từ khóa	11
1.5. Đặc trưng tiếng nói liên quan đến thông tin người nói.....	13
1.5.1. Rút trích đặc trưng.....	13
1.5.2. Đặc trưng biên độ	14
1.5.3. Đặc trưng cao độ.....	15
1.5.4. Đặc trưng phổ	16
CHƯƠNG II: MỘT SỐ PHƯƠNG PHÁP PHÂN LỚP TRONG NHẬN DẠNG NGƯỜI NÓI QUA GIỌNG NÓI	20
2.1. Kỹ thuật so khớp mẫu trực tiếp	20
2.1.1. Phương pháp so sánh mẫu trực tiếp cổ điển dùng giải thuật thời gian động (Dynamic time warping - DTW)	20

2.1.2. Phương pháp phân lớp dùng lượng tử hóa vector (Vector Quantization - VQ).....	23
2.2. Phương pháp sử dụng mô hình pha trộn Gaussian.....	30
2.2.1. Đặc tả mô hình.....	30
2.2.2 Ước lượng tham số mô hình GMM.....	33
2.2.3. Mô hình hóa người nói không phụ thuộc văn bản với mô hình Gaussian Mixture Model - GMM.....	34
2.2.4. Huấn luyện với mô hình Gaussian Mixture Model - GMM.....	35
2.2.5. Nhận dạng với mô hình Gaussian Mixture Model - GMM.....	36
2.3. Phân lớp bằng mô hình GMM-HMM.....	37
2.3.1. Giới thiệu.....	37
2.3.2. Đặc tả mô hình GMM-HMM.....	39
2.3.3. GMM-HMM và bài toán định danh người nói.....	40
CHƯƠNG III: ĐÁNH GIÁ THỰC NGHIỆM PHƯƠNG PHÁP NHẬN DẠNG NGƯỜI NÓI DÙNG VQ VÀ MÔ HÌNH GMM.....	44
3.1. Lựa chọn cơ sở dữ liệu.....	44
3.1.1. Phạm vi của các cơ sở dữ liệu ATR.....	44
3.1.2. Thu thập dữ liệu tiếng nói trong ATR.....	46
3.1.3. Gán nhãn trong ATR.....	48
3.2. Cài đặt các phương pháp trên MATLAB.....	51
3.2.1. Cài đặt phương pháp VQ.....	51
3.2.2. Cài đặt phương pháp GMM.....	53
3.3. Kết quả của các phương pháp.....	56
3.4. Đánh giá các kết quả.....	56
KẾT LUẬN.....	57
TÀI LIỆU THAM KHẢO.....	58

DANH MỤC BẢNG

Bảng 1.1: Một số giá trị của tần số cơ bản ứng với giới tính và độ tuổi	15
Bảng 3.1: Thống kê các thông số của cơ sở dữ liệu	45
Bảng 3.2: Các lớp phiên âm	48
Bảng 3.3: Các ký hiệu âm thanh – âm cho lớp thứ 2.....	49

DANH MỤC HÌNH

Hình 1.1: Các ứng dụng xử lý tiếng nói.....	6
Hình 1.2: Sơ đồ nhận dạng tổng quát.....	7
Hình 1.4: Đặc trưng phổ formant đặc trưng cho cơ quan phát âm.....	9
Hình 1.5: Mô hình chung nhận dạng người nói.....	10
Hình 1.6: Bài toán định danh người nói.....	10
Hình 1.7: Bài toán xác thực người nói.....	11
Hình 1.8: Phân loại bài toán nhận dạng người nói theo từ khóa.....	12
Hình 1.9: Sơ đồ rút trích vector đặc trưng tổng quát.....	13
Hình 1.10: Sơ đồ rút trích đặc trưng chi tiết	14
Hình 1.11: Đặc trưng cao độ	16
Hình 1.12: Đặc trưng phổ và đường bao phổ đặc trưng cho cơ quan phát âm.....	17
Hình 1.13: Đồ thị biểu diễn mối quan hệ giữa Mel và Hz.....	18
Hình 1.14: Các bước trích chọn đặc trưng	18
Hình 1.15: Bộ lọc trên thang Mel.....	19
Hình 1.16: Bộ lọc trên tần số thật.....	19
Hình 1.17: Minh họa các bước biến đổi MFCC	19
Hình 2.1: Hai chuỗi dữ liệu trong DTW theo thời gian.....	21
Hình 2.2: Giãn tín hiệu có độ dài khác nhau: tín hiệu màu đỏ đã được giãn để có độ dài tương ứng với tín hiệu màu xanh.....	22
Hình 2.3: Khoảng cách Euclidean tính cho 2 mẫu tiếng nói đã giãn để có độ dài bằng nhau	22
Hình 2.4a: Huấn luyện.....	24
Hình 2.4b: Nhận dạng	25
Hình 2.5: Hàm mật độ Gauss.....	30
Hình 2.6: Mô hình GMM.	31
Hình 2.7: Hàm mật độ của GMM có 3 phân phối Gauss.....	32

Hình 2.8: HMM với 3 trạng thái và trọng số chuyển trạng thái.....	37
Hình 2.9: Nhận dạng người nói dùng HMM.....	38
Hình 2.10: Mô hình GMM-HMM 3 trạng thái.....	39
Hình 3.1: Sơ đồ khối hệ thống thu thập dữ liệu	45
Hình 3.2: Một ví dụ về kết quả phiên âm đa tầng.	50
Hình 3.3: Thuật toán huấn luyện VQ.....	52
Hình 3.4: Thuật toán nhận dạng VQ.....	53
Hình 3.5: Thuật toán huấn luyện GMM.....	54
Hình 3.6: Thuật toán nhận dạng GMM.....	55

DANH MỤC CHỮ VIẾT TẮT VÀ KÍ HIỆU

Ký tự	Ý nghĩa
F0	Tần số dao động cơ bản
MFCC	Hệ số Cepstral tần số Mel
IDFT	Phép biến đổi Fourier ngược
DCT	Phép biến đổi cosin rời rạc
GMM	Mô hình Gaussian hỗn hợp
VQ	Kỹ thuật lượng tử hóa vector
FFT	Phép biến đổi Fourier nhanh

MỞ ĐẦU

1. Lý do chọn đề tài

Tiếng nói là phương tiện giao tiếp cơ bản của con người. Vì vậy tiếng nói cũng là loại hình thông tin cơ bản và phổ biến nhất trong các hệ thống truyền thông. Tín hiệu tiếng nói mang nhiều thông tin, như thông tin ngôn ngữ, thông tin về người nói, thông tin về sắc thái tình cảm khi nói,...

Hầu hết các hệ thống xử lý và nhận dạng tiếng nói truyền thống tập trung vào xử lý các thông tin ngôn ngữ để đảm bảo nhận dạng được nội dung ngôn ngữ hay ngữ nghĩa được nói [5], [11]. Tuy nhiên để các ứng dụng xử lý tiếng nói trong máy tính có thể được áp dụng rộng rãi trong thực tế, một trong những vấn đề quan trọng cần đảm bảo là khả năng nhận dạng và xác minh người nói [2], [12].

Trên thế giới đã có nhiều nghiên cứu về nhận dạng người nói qua giọng nói [12], [14]. Tại Việt Nam cũng có một số nghiên cứu ban đầu, đặc biệt là một số nghiên cứu tại Viện Công nghệ thông tin [3] và Viện nghiên cứu MICA – Đại học Bách Khoa Hà Nội [1], [2]. Tuy nhiên ở Việt Nam vẫn chưa có nhiều các nghiên cứu đánh giá một cách tổng hợp các phương pháp nhận dạng người nói phổ biến. Đặc biệt, hai phương pháp nhận dạng người nói hiện đại dùng phép lượng tử hóa vector – VQ và mô hình pha trộn Gaussian - GMM [10], [12], [13] lại chưa được nghiên cứu nhiều tại Việt Nam. Vì vậy, luận văn này nghiên cứu một số phương pháp nhận dạng người nói bằng giọng nói, tập trung vào hai phương pháp dùng phép lượng tử hóa vector và mô hình pha trộn Gaussian, đánh giá thực nghiệm các phương pháp, và đưa ra những khuyến nghị.