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 KĨ 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 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

DANH MỤC BẢNG

DANH MỤC HÌNH

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.