

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



NGUYỄN TRUNG THÀNH

**NGHIÊN CỨU MỘT SỐ PHƯƠNG PHÁP BIẾN ĐỔI THÔNG
TIN NGƯỜI NÓI TRONG TIẾNG NÓI**

Chuyên ngành: **Khoa học máy tính**
Mã số: **60.48.01.01**

LUẬN VĂN THẠC SĨ KHOA HỌC MÁY TÍNH

Người hướng dẫn khoa học: **TS. PHÙNG TRUNG NGHĨA**

Thái Nguyên năm 2015

LỜI CẢM ƠN

Lời đầu tiên, em xin chân thành cảm ơn **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 "Khoa học máy tính" đã truyền dạy những kiến thức quý báu, những kiến thức này rất hữu ích và giúp em nhiều khi thực hiện nghiên cứu. Đặc biệt, em xin cảm ơn thầy Đỗ Huy Khôi về những góp ý có ý nghĩa rất lớn khi em thực hiện luận văn 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 09 tháng 07 năm 2015

Học viên

Nguyễn Trung Thành

LỜI CAM ĐOAN

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 **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

Nguyễn Trung Thành

MỤC LỤC

LỜI CẢM ƠN	ii
LỜI CAM ĐOAN	iv
MỤC LỤC	v
DANH MỤC BẢNG	viii
DANH MỤC HÌNH	ix
DANH MỤC CHỮ VIẾT TẮT	xi
MỞ ĐẦU	1
1. Lý do chọn đề tài	1
2. Mục tiêu của đề tài	2
3. Đối tượng và phạm vi nghiên cứu	3
4. Phương pháp nghiên cứu	3
5. Ý nghĩa khoa học và thực tiễn	3
CHƯƠNG 1. TỔNG QUAN VỀ TIẾNG NÓI VÀ VẤN ĐỀ BIẾN ĐỔI THÔNG TIN NGƯỜI NÓI TRONG TIẾNG NÓI	4
1.1. Thông tin tiếng nói	4
1.2. Tín hiệu tiếng nói	4
1.3. Quá trình tạo tiếng nói	6
1.4. Cơ quan thính giác	9
1.5. Xử lý tiếng nói	12
1.6. Thông tin người nói trong tiếng nói	13
1.7. Biến đổi thông tin người nói trong tiếng nói và ứng dụng	14
CHƯƠNG 2. MỘT SỐ PHƯƠNG PHÁP BIẾN ĐỔI THÔNG TIN NGƯỜI NÓI TRONG TIẾNG NÓI	17
2.1. Phương pháp biến đổi tham số trực tiếp	17
2.1.1. Phân khung	17
2.1.2. Đặc trưng biên độ	18

2.1.3. Đặc trưng cao độ	19
2.1.4. Đặc trưng phổ.....	20
2.1.5. Mô hình nguồn âm / bộ lọc	24
2.1.6. Ưu nhược điểm của phương pháp.....	25
2.2. Phương pháp thay thế khung tiếng nói	25
2.2.1. Phương pháp tìm kiếm mẫu tiếng nói thay thế	26
2.2.2. Thuật toán tìm kiếm khung tiếng nói thay thế	28
2.2.3. Làm trơn điểm ghép nói.....	28
2.2.4. Ưu nhược điểm của phương pháp.....	29
2.3. Phương pháp sử dụng học máy thống kê GMM.....	29
2.3.1. Phân bố Gauss	29
2.3.2. Mô hình Gaussian hỗn hợp	30
2.3.3. Mô hình hóa người nói bằng mô hình GMM	31
2.3.4. Huấn luyện	32
2.3.5. Biến đổi	33

CHƯƠNG 3. ĐÁNH GIÁ THỰC NGHIỆM CÁC PHƯƠNG PHÁP

BIẾN ĐỔI THÔNG TIN NGƯỜI NÓI TRONG TIẾNG NÓI.....	35
3.1. Ngữ âm tiếng Việt.....	35
3.2. Cơ sở dữ liệu tiếng nói tiếng Việt.....	37
3.3. Tổng hợp tiếng nói tiếng Việt	40
3.4. Lựa chọn cơ sở dữ liệu.....	40
3.5. Cài đặt các phương pháp biến đổi thông tin người nói.....	41
3.5.1. Phương pháp thay đổi tham số trực tiếp	41
3.5.2. Phương pháp thay thế tiếng nói theo mẫu.....	44
3.5.3. Phương pháp biến đổi sử dụng học máy thống kê.....	47
3.6. Đánh giá các phương pháp.....	49
3.6.1. Tiêu chí đánh giá.....	49
3.6.2. Đánh giá thực nghiệm	50

3.6.3. Kết quả đánh giá	51
3.6.4. Nhận xét chung về kết quả đánh giá	51
KẾT LUẬN	53
TÀI LIỆU THAM KHẢO	54

DANH MỤC BẢNG

Bảng 3.1 Cấu trúc âm tiết tiếng Việt.....	37
Bảng 3.2: Sáu thanh điệu tiếng Việt	37
Bảng 3.3. Các tham số thực nghiệm	50
Bảng 3.4: Kết quả đánh giá khách quan.....	51
Bảng 3.5: Kết quả đánh giá chủ quan ABX.....	51

DANH MỤC HÌNH

Hình 1.1: Dạng sóng tiếng nói một câu tiếng Việt	5
Hình 1.2: Tiếng nói hữu thanh	5
Hình 1.3: Bộ phận cung cấp làn hơi.....	6
Hình 1.4: Dây thanh âm	7
Hình 1.5: Cấu trúc cơ quan phát âm	8
Hình 1.6: Hình dáng cơ quan phát âm thay đổi trong quá trình phát âm.....	8
Hình 1.7: Mô hình hóa cơ quan phát âm.....	9
Hình 1.8: Biểu diễn mô hình hóa cơ quan phát âm đầy đủ bằng máy tính.....	9
Hình 1.9: Mô hình cơ quan thính giác	10
Hình 1.10: Thang tần số Bark.....	10
Hình 1.11: Ngưỡng nghe	11
Hình 1.12: Mặt nạ thời gian	11
Hình 1.13: Mặt nạ tần số.....	11
Hình 1.14: Một số ứng dụng của xử lý tiếng nói	12
Hình 1.15: Hệ thống nhận dạng người nói.....	13
Hình 1.16: Người nói khác nhau có cơ quan phát âm và cách phát âm khác nhau dẫn tới tiếng nói khác nhau	14
Hình 1.17: Đặc trưng phổ formant đặc trưng cho cơ quan phát âm	14
Hình 1.18: Tính toán vector đặc trưng.....	15
Hình 1.19: Mô hình học máy thống kê GMM	16
Hình 2.1: Phân đoạn tiếng nói thành các khung chồng lấp.....	18
Hình 2.2: Đặc trưng trường độ.....	19
Hình 2.3: Đặc trưng phổ và đường bao phổ.....	21
Hình 2.4: Đồ thị biểu diễn mối quan hệ giữa Mel và Hz.....	22
Hình 2.5: Các bước trích đặc trưng MFCC	22
Hình 2.6: Bộ lọc trên thang Mel	22
Hình 2.7: Bộ lọc trên tần số thật	23

Hình 2.8: Minh họa các bước biến đổi MFCC	23
Hình 2.9: Mô hình nguồn âm / bộ lọc	24
Hình 2.10: Đặc trưng F0, Gain và LSF	27
Hình 2.11: Quá trình tìm kiếm và thay thế mẫu có khoảng cách ngắn nhất...	28
Hình 2.12: Quá trình làm trơn biên ghép nối	29
Hình 2.13: Hàm mật độ xác suất Gauss	30
Hình 2.14: Mô hình GMM Biến đổi người nói theo mô hình GMM	32
Hình 3.1: Đường F0 sáu thanh điệu tiếng Việt.	36
Hình 3.2: Phân tích phổ, F0 và tái tạo bằng STRAIGHT	41
Hình 3.3: Thuật toán điều chỉnh trực tiếp tham số tiếng nói	43
Hình 3.4: Giao diện chương trình điều chỉnh trực tiếp tham số tiếng nói	44
Hình 3.5: Lưu đồ thuật toán tính khoảng cách giữa 2 khung tiếng nói	45
Hình 3.6: Thuật toán tìm kiếm kết hợp	46
Hình 3.7: Huấn luyện mô hình GMM cho tham số phổ LSF	47
Hình 3.8: Chuyển đổi mô hình GMM cho tham số phổ LSF	48

DANH MỤC CHỮ VIẾT TẮT VÀ KÍ HIỆU

Ký tự	Ý nghĩa
F0	Tần số dao động cơ bản
AMDF	Hàm hiệu biên độ trung bình
LP	Phương pháp dự đoán tuyến tính
PCM	Kỹ thuật điều chế xung mã
WAV	Dữ liệu âm thanh không nén
PSTN	Mạng điện thoại chuyển mạch công cộng
LSF	Tham số phổ đường
ABX	Thang điểm đánh giá theo cặp
GMM	Mô hình Gaussian hỗn hợp
PI	Chỉ số hiệu năng
MOS	Thang điểm đánh giá chủ quan trung bình