

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

**THÂN THẾ HUYỀN**

**NGHIÊN CỨU PHƯƠNG PHÁP  
BIẾN ĐỔI THÔNG TIN NGƯỜI NÓI TRONG TIẾNG NÓI DÙNG KỸ  
THUẬT PHÂN RÃ THEO THỜI GIAN**

**LUẬN VĂN THẠC SĨ KHOA HỌC MÁY TÍNH**

**THÁI NGUYÊN, 2018**

## 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.

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 22 tháng 06 năm 2018*

**Học viên**

**Thân Thế Huyền**

## **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**

**Thân Thế Huyền**

## MỤC LỤC

<b>LỜI CẢM ƠN</b> .....	<b>1</b>
<b>LỜI CAM ĐOAN</b> .....	<b>ii</b>
<b>MỤC LỤC</b> .....	<b>iii</b>
<b>DANH MỤC BẢNG</b> .....	<b>v</b>
<b>DANH MỤC CHỮ VIẾT TẮT VÀ KÍ HIỆU</b> .....	<b>viii</b>
<b>MỞ ĐẦU</b> .....	<b>1</b>
1. Lý do chọn đề tài.....	1
2. Đối tượng và phạm vi nghiên cứu.....	2
3. Hướng nghiên cứu của luận văn .....	3
4. Những nội dung nghiên cứu chính.....	3
5. Phương pháp nghiên cứu.....	4
6. Ý nghĩa khoa học của luận văn: .....	4
<b>CHƯƠNG 1: TỔNG QUAN VỀ TIẾNG NÓI VÀ VẤN ĐỀ BIẾN ĐỔI THÔNG TIN NGƯỜI NÓI TRONG TIẾNG NÓI</b> .....	<b>5</b>
1.1. Thông tin tiếng nói .....	5
1.2. Tín hiệu tiếng nói .....	5
1.3. Quá trình tạo tiếng nói .....	7
1.4. Cơ quan thính giác .....	10
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 .....	15
1.8. Phương pháp biến đổi thay đổi tham số trực tiếp .....	16
<b>CHƯƠNG 2: KỸ THUẬT PHÂN RÃ THEO THỜI GIAN TD VÀ ỨNG DỤNG TRONG BIẾN ĐỔI THÔNG TIN NGƯỜI NÓI</b> .....	<b>21</b>
2.1. Kỹ thuật phân rã tiếng nói theo thời gian.....	21
2.1.1. Phương pháp TD nguyên thủy .....	21

2.1.2. Phương pháp phân rã tiếng nói theo thời gian giới hạn RTD .....	24
2.1.3. Phương pháp MRTD .....	27
2.2. Một số kỹ thuật biến đổi thông tin người nói dùng TD .....	32
2.2.1. Biến đổi thông tin người nói bằng TD-GMM.....	32
2.2.2. Biến đổi thông tin người nói bằng HTD [12] .....	34
<b>CHƯƠNG 3: ĐÁNH GIÁ THỰC NGHIỆM CÁC PHƯƠNG PHÁP</b>	
<b>BIẾN ĐỔI THÔNG TIN NGƯỜI NÓI TRONG TIẾNG NÓI.....</b>	<b>42</b>
3.1. Ngữ âm tiếng Việt.....	42
3.2. Cơ sở dữ liệu tiếng nói tiếng Việt .....	44
3.3. Tổng hợp tiếng nói tiếng Việt .....	47
3.4. Lựa chọn cơ sở dữ liệu.....	47
3.5. Đánh giá các phương pháp.....	48
3.5.1. Tiêu chí đánh giá.....	48
3.5.2. Thực nghiệm các phương pháp.....	49
3.5.3. Kết quả đánh giá.....	50
3.5.4. Thảo luận.....	51
<b>KẾT LUẬN .....</b>	<b>53</b>
<b>TÀI LIỆU THAM KHẢO .....</b>	<b>54</b>

## DANH MỤC BẢNG

Bảng 3.1: Cấu trúc âm tiết tiếng Việt.....	44
Bảng 3.2: Sáu thanh điệu tiếng Việt .....	44
Bảng 3.3. Các tham số thực nghiệm .....	49
Bảng 3.4. Kết quả đánh giá khách quan.....	50
Bảng 3.5. Kết quả đánh giá chủ quan ABX.....	50

## DANH MỤC HÌNH

Hình 1.1: Dạng sóng tiếng nói một câu tiếng Việt .....	6
Hình 1.2: Tiếng nói hữu thanh .....	6
Hình 1.3: Bộ phận cung cấp làn hơi.....	7
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.....	9
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 ....	10
Hình 1.9: Mô hình cơ quan thính giác .....	10
Hình 1.10: Thang tần số Bark.....	11
Hình 1.11: Ngưỡng nghe .....	11
Hình 1.12: Mặt nạ thời gian (che âm thanh liền trước và liền sau) .....	12
Hình 1.13: Mặt nạ tần số (che âm thanh có tần số khác nhau phát cùng thời điểm).....	12
Hình 1.14: Một số ứng dụng của xử lý tiếng nói .....	13
Hình 1.15: Hệ thống nhận dạng người nói – một trong các ứng dụng xử lý thông tin 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 2.1: Ví dụ về hai hàm sự kiện liền kề.....	25
Hình 2.2: Hàm sự kiện có tính chất “hình học chuẩn” và “hình học không chuẩn”.....	27
Hình 2.3: Thuật toán chuẩn hóa vector sự kiện trong MRTD .....	31
Hình 2.4: Hình vẽ các hàm sự kiện nhận được khi MRTD phân tích một câu tiếng Nhật, chỉ số trên miền thời gian là số khung. ....	32
Hình 2.5: Phương pháp biến đổi TD-GMM.....	34

Hình 2.6: Mô hình biến đổi giọng người nói HTD .....	35
Hình 2.7: Ví dụ phân tích / tái tạo tiếng nói bằng MRTD với N khung và K điểm sự kiện .....	37
Hình 3.1: Đường F0 sáu thanh điệu tiếng Việt theo, dấu ? ở thanh ngã chỉ ra rằng đường F0 của thanh ngã không thống nhất giữa các mẫu ở vùng giữa. .	43



## DANH MỤC CHỮ VIẾT TẮT VÀ KÍ HIỆU

Ký tự	Ý nghĩa
F0	Tần số dao động cơ bản
TD	Phân rã theo thời gian
RTD	Giới hạn
LSF	Tham số đường phổ
DLSF	Các ràng buộc
MRTD	PP Phân rã tiếng nói theo thời gian giới hạn cải tiến
GMM	Mô hình Gaussian hỗn hợp
TD- GMM	Mô hình pha trộn Gausce
HTD	Kỹ thuật phân rã (kết hợp HTT+TD)
PI	Chỉ số hiệu năng
PI-LSF	Hiệu năng phổ
MOS	Thang điểm đánh giá chủ quan trung bình
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 chuyên mạch công cộng
LSF	Tham số phổ đường
ABX	Thang điểm đánh giá theo cặp

## 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 viễn thông. Tín hiệu tiếng nói mang nhiều thông tin, như thông tin ngôn ngữ, thông tin về người nói, thông tin về cảm xúc khi nói,...

Hầu hết các hệ thống xử lý tiếng nói truyền thông tập trung vào xử lý các thông tin ngôn ngữ để đảm bảo tiếng nói sau xử lý có thể hiểu được [1]. Tuy nhiên để các ứng dụng xử lý tiếng nói trong máy tính có thể được áp dụng rộng rãi trong thực tế, tính tự nhiên của tiếng nói được xử lý cũng cần được quan tâm và cũng đã được quan tâm nghiên cứu trong thời gian gần đây [2]. Để đảm bảo tiếng nói sau xử lý (như tiếng nói được tổng hợp) được tự nhiên, một trong những vấn đề quan trọng cần đảm bảo là thông tin về người nói, bao gồm cả các thông tin chung về người nói như giới tính, độ tuổi,... đến các thông tin chi tiết như thông tin nhận danh chính xác người nói [5,6,7,9,10,11].

Các hệ thống tổng hợp tiếng nói nhân tạo thường chỉ có thể tổng hợp ra tiếng nói của một số giọng nói đã được thu sẵn và huấn luyện trước cho máy tính. Trong nhiều ứng dụng truyền thông đa phương tiện hiện đại, việc biến đổi thông tin người nói trong tín hiệu tiếng nói có vai trò quan trọng. Một số ví dụ điển hình như:

- Trong các bộ phim lịch sử cần diễn viên nói với giọng giống với giọng của nhân vật lịch sử [6].

- Trong các clips quảng cáo, âm nhạc cần biến đổi giọng nói, giọng hát của diễn viên theo các tiêu chí cụ thể khác nhau như cao hơn, trầm hơn, giống với nhân vật thật hơn,... [6]