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

THÂN THẾ HUYẾN

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 cám 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 cám ơn quý thầy cô giảng dạy chương trình cao học "Khoa hoc 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 cám 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 cám ơn!

Thái Nguyên, ngày 22 tháng 06 năm 2018

Học viên

Thân Thế Huyến

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 Huyến

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
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 chuyển 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]