
화자적응을 이용한 음성인식 제어시스템 개발

김용표* · 윤동한** · 최운하**

Development of Voice Activated Universal Remote Control System using the Speaker Adaptation

Yong-Pyo Kim* · Dong-han Yoon** · UnHa-Choi**

본 논문은 구미시청 산·학·연 기술개발자금 지원 받았음.

요 약

본 논문은 신경회로망을 이용한 화자적응 음성인식 제어시스템을 개발하였다. 화자종속시스템은 단일 화자의 음성만 등록하여 이용하므로 여러 화자의 음성을 인식하는 데는 문제가 있고, 화자독립시스템은 여러 화자를 인식한다. 본 연구 개발에서는 화자적응시스템을 구현하여 화자종속형의 단점을 보완하여 화자 독립과 화자 종속을 혼합하여 사용 할 수 있는 기능으로 화자 적용방법으로 구현하였고, 화자인증(Speaker Verification)도 가능하도록 프로그래밍 하였다.

ABSTRACT

In this paper, development of voice activated Universal Remote Control using the Neural Networks. A speaker dependent system is developed to operate for a single speaker. These systems are usually easier to develop, cheaper to buy and more accurate, but not as flexible as speaker adaptive or speaker independent systems. A speaker independent system is developed to operate for any speaker of a particular type (e.g. American English). These systems are the most difficult to develop, most expensive and accuracy is lower than speaker dependent systems. However, they are more flexible. A speaker adaptive system is developed to adapt its operation to the characteristics of new speakers. It's difficulty lies somewhere between speaker independent and speaker dependent systems. This paper is developed Speaker Adaptation using the Neural Networks.

키워드

Touch LCD, Remote Control, RF, Macro, Learn, IR

I. 서 론

문명이 발달하면서 음성을 이용하여 각종 기계나 도구들을 쉽게 조작하고자 하는 것은 인류의 오랜 꿈 중의 하나

였다. 음성인식은 사람의 억양과 음의 높낮이가 서로 다르다는 특성에 기인한 방식으로 마이크 등을 통해 전달된 음성의 특징을 분석한 후 신분을 확인할 수 있고, 음성인식을 사용하기 위한 별도의 교육이 필요하지 않으며 시

* 대구기능대학 전자과 교수

** 금오공과대학교 전자공학부 교수

스택 가격이 저렴하다는 장점이 있다. 하지만 감기나 기타 요인에 의해 목이 쉬었을 때나, 의도적으로 타인의 목소리를 흉내 내거나 주변 환경에 큰 소음이 있을 경우에는 인식 오류를 할 수 있다는 단점이 있다. 음성인식 방식에는 일반적으로 모든 사람의 음성을 인식하는 화자독립형(Speaker Independent)과 학습시킨 특정인의 목소리만을 인식하는 화자종속형(Speaker Dependent)의 두 가지 방식이 있다.

어떤 시스템은 인식기를 사용하기 전에 자신의 음성 샘플을 제공하여 화자를 등록하기 때문에 화자 종속(Speaker Dependent) 시스템이라 불리며, 다른 시스템은 등록을 필요로 하지 않기 때문에 화자 독립(Speaker Independent) 시스템으로 불린다. 화자에 대한 제한이 없으며 초기 학습 과정 이외에 별도의 학습이 필요없다. 화자적응(Speaker Adaptation)은 화자종속과 화자독립의 중간단계이다. 기본적으로 화자독립 기능을 제공하지만 준비된 훈련과정을 거쳐 특정화자에 대한 성능을 높인다. 거의 대부분의 음성 인식 시스템들은 이론적으로는 학습시키는 모드에 따라 화자 종속 또는 화자 독립 모드로 동작할 수 있다. 화자 종속 인식기는 음성 인식 시스템의 내부적인 음성 처리 과정의 모델을 특징짓는 매개변수(또는 모델)를 학습시키는데 단일 화자의 음성을 이용한다. 따라서 이런 음성 인식기는 훈련자의 음성만 인식하는데 사용되며, 여러 화자의 음성으로 훈련하여 다양한 화자의 음성을 인식하는 화자 독립인식기에 비하여 인식 성능이 좋다. 이처럼 화자 독립 시스템의 성능이 정확하다 하더라도 새로운 화자에 대해서는 다시 학습시켜야 한다는 단점이 존재한다. 따라서 화자 종속 또는 독립 모드의 인식기는 정확성과 편이성에 대한 균형이 따라야 한다.[1]

본 연구에서는 화자종속형의 단점을 보완하여 화자 독립과 화자 종속을 혼합하여 사용할 수 있는 기능으로 화자 적응(Speaker Adaptive)방법으로 구현하였고, 화자인증(Speaker Verification)도 가능하도록 프로그램 하였다. 본 연구개발에서는 신경회로망(Neural Networks) 이론을 이용하여 음성인식 모듈을 개발 제작하였다. 과학의 발달로 가전기기의 첨단화가 필수적인 시점에서 음성인식모듈을 개발하여 가전기기(TV, VCR, Audio etc)들을 음성인식을 통하여 제어함으로써, 가사자동화(Home Automation)에 기여할 것으로 판단된다.

II. 이론적 배경

음성인식 시스템은 1950년에 AT&T의 Davis, Biddulph, Balaskek 등에 의해서 음향-음성학적인 지식을 기반으로 한 화자종속 숫자음 인식실험이 그 시초이다. 이러한 연구에 뒤이어, 60년대에서 Sakoe와 Chilba 등에 의해서 제안된 동적 정합법, Reddy에 의해서 이루어진 동적인 음소 전이에 기반한 연속 음성인식 연구가 이루어졌다. 70년대에 들어서는 음성인식을 위한 특징 파라미터로써 널리 쓰였던 선형 예측코딩(LPC : Linear Predictive Coding)에 대한 연구가 이루어졌다. 또 다른 연구로는 IBM을 중심으로 한 대어휘 음성인식에 관한 연구가 이루어져서 단순한 데이터베이스 질의어나 간단한 사무실 메모를 위한 시스템 등이 선보였다. 마지막으로 AT&T 벨 실험실에서 진정한 의미의 화자독립 음성인식 시스템에 대한 일련의 실험실이 이루어졌다.

70년대의 연구가 독립발성된 단어인식에 초점을 맞추어 연구가 이루어졌다면, 80년대에는 연결 단어인식을 위해서 발성음으로부터 최적의 단어열을 찾는 방법들이 많이 연구되었다.

80년대의 가장 큰 특징은 기존의 패턴정합법(Template Matching)에 기반한 인식방법에서 확률적인 모델로의 전환이라고 말할 수 있다. 은닉 마르코프 모델로 대표되는 확률적 모델은 80년대 중반까지는 몇몇 실험실을 중심으로 연구되었고, 이때 한창 연구가 진행되었던 신경회로망에 대한 연구결과를 음성인식에 적용하는 연구도 활발히 진행되었다.

90년대에 들어서는 음성인식 연구에서 일본 전진하여 자연언어 처리의 전처리부를 결합하는 연구가 진행되었으며, 연구의 방향도 ATIS(Airline Traveling Information System)라는 항공기 여행을 위한 영역이나 WSJ(Wall Street Journal)등으로 인식대상 영역이 바뀌었다. 이러한 연구와 아울러, 전화상의 음성을 자동적으로 인식하거나, 교환원의 역할을 대체할 수 있는 전화서비스에 대한 많은 연구가 현재 이루어지고 있는 상황이다.[2]~[4]

음성 처리 분야에서 1975년에 패턴을 분류하기 위해 로젠 브러트가 제안한 신경망 모델인 퍼셉트론(perceptron)이 있다. 간단하게 신경망의 형태는 인간의 신경세포인 뉴런을 간략화해서 모델링한 것으로서 현재는 음성인식, 영상인식, 문자인식, 제어 및 예측 등의 분야에서 이용하고 있다.

신경망 모델에는 단층 퍼셉트론(Single Layer Perceptron),

다층 퍼셉트론(MultiLayer Perceptron), Kohonen's feature map 알고리즘 등과 같은 형태로 나눌 수 있다. 보통 음성 인식에는 다층 퍼셉트론 형태의 신경망을 이용하고 있는데 이 방법은 다층 구조를 가지는 신경회로망의 synapse의 결합계수(weight)를 목적에 맞도록 바꾸어 가는 학습 방법이다. 다시 말하면 입력층에 주어진 입력 패턴이 출력층에 전파되면서 변환 출력 패턴과 비교하여 차이를 감소시키는 방향으로 결합계수를 조절하는 학습 방법을 나타낸 것이다. 학습 목표는 원하는 목표 출력과 출력값을 비교하기 위해 에러가 최소가 되도록 gradient decent를 사용하고 있다. 그리고 출력 값과 목표 출력 값과의 차이를 최소로 하면 되는데 일반적으로 최소자승법(LMS)을 사용하며 이것에 미분을 취하여 최소치를 구하여 결합계수를 조절하고 있다. 입력과 출력 노드 사이에 하나 또는 그 이상의 노드 층을 가지는 feed-forward에서 다층 퍼셉트론이 많이 사용되고 있는데 결합계수에 대한 초기 파라미터에 대해 학습 과정은 cost function이 변하지 않고 남거나 받아들일 수 있는 값이 될 때까지 결합계수를 갱신하기 위해 반복된다. Back propagation으로 학습한 다층 퍼셉트론을 사용한 음성 인식에는 고립 단어인식과 고립 음소인식이 있는데, 그 이유는 신경망을 이용한 음성인식은 음성을 분할하는 segmentation과정이 먼저 이루어 져야 하기 때문이다.[5][6]

III. 음성인식 모듈의 구현

본 연구 개발은 사용자가 화자속속형으로 음성인식을 구현하도록 설계하였고, 독립된 사용자를 20명까지 입력 할 수 있도록 하여 화자독립형으로 구현 했을 때 보다 제품 재료가 월등히 저렴하고, 주위 소음 및 인식률에 의한 오동작을 원천 제거하여 설계하였다. 하드웨어 구조는 그림 1과 같다.

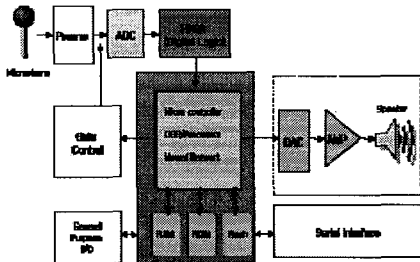


그림 1. 하드웨어 구성
Fig. 1. Hardware Configuration

3.1 Architecture 설계

시스템 설계는 구동 시스템을 분석하여 시스템에 필요한 시그널을 In-system으로 다양하게 변경시켜 공급할 수 있는 FPGA를 이용한 시스템 Architecture를 설계한다.

3.2. Timing Generator 회로 설계

다양한 Timing Pattern을 Hardware적으로 발생시킬 수 있는 Finite State Machine의 구조와 회로를 설계하였다.

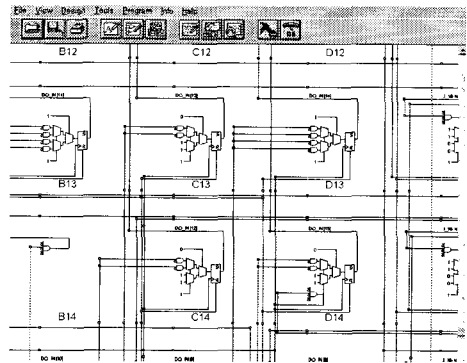


그림 2. FPGA를 이용한 구조 설계
Fig. 2. Architecture Design Using FPGA

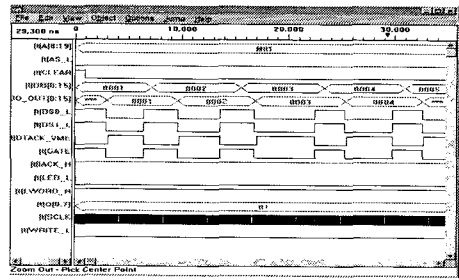


그림 3. Timing Generator 회로설계
Fig. 3. System Architecture for Timing Generator

3.3 FPGA 설계 및 Simulation

하드웨어를 FPGA로 구현하기 위한 최적의 FPGA Chip을 선정하고 하드웨어를 설계하여 프로그램한 후 Simulation하여 봄으로서 하드웨어 설계를 검증한다. 본 과제에서는 FPGA 기법으로 설계하여 간단, 단순화 및 Noise, 신뢰성을 향상시키고 경제성이 확보되는 제품을 개발하고자 한다.

3.4 FPGA In-System Programming을 위한 Hardware 회로설계

FPGA를 Program 할 수 있는 PC Interface 회로를 설계하여 제작한다.

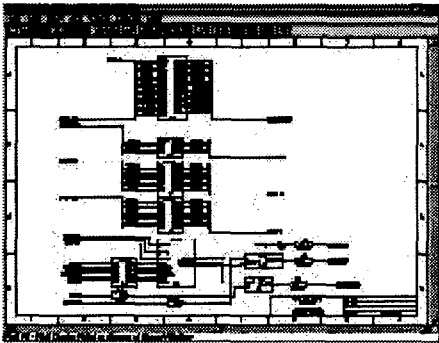


그림 4. FPGA 설계 Simulation 결과
Fig 4. Simulation Result of FPGA Architecture

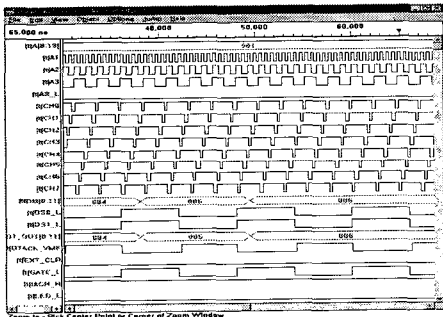


그림5. Timing Generator Simulation 결과
Fig 5. Simulation Result for Timing Generator

IV. 실험 및 고찰

4.1 시스템 Test 및 평가

시스템 Test 및 평가항목은 화자독립, 화자 종속, 화자 인증, 음성합성, 음향효과, DTMF Ton Generation, 기억/재생, 단어 선별 인식, 연속 인식 기능 등 음성인식을 이용하여 제품을 개발할 수 있는 개발 구현 및 테스트한다. 아래의 그래프는 신경회로망 알고리즘으로 음성 단어 학습 시 음성인식률이 개선되는 것을 Simulation하였다.

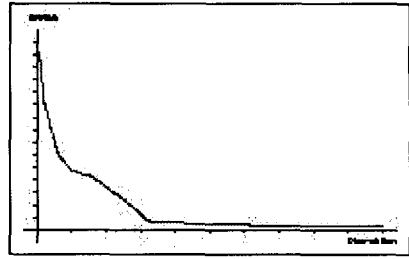


그림 6. 신경회로망의 학습결과표
Fig 6. Learning Result Table of Neuron Network

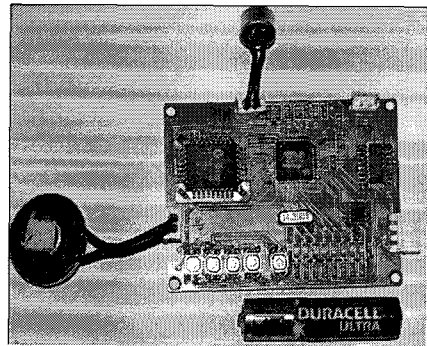


그림 7. 연구개발된 음성인식 모듈
Fig 7. Developed Voice Activate Module

4.2 기대효과

음성인식 시스템 개발을 통하여 제작된 모듈을 리모컨에 사용함으로써 회로개발 시간을 단축할 수 있으며 다양한 실험을 통한 회로설계의 최적 파라미터를 찾아냄으로써 음성인식 제품의 고품질화에 기여할 수 있다. 또한 음성인식 모듈의 설계를 위한 회로분석을 하므로 회로설계를 위한 기술자료 수집과 최적의 설계를 할 수 있는 기술 축적 등 음성인식에 의한 동작으로 인한 동작 오류 영향과 이것의 이론적 분석으로 노이즈에 의한 간섭을 최소화시킬 수 있는 방안에 대한 연구를 병행시킴으로써 안정적인 회로 동작을 위한 기술축적을 얻을 수 있었다.

향후 개발 연구로는 움직이는 사람의 얼굴, 자동차 등의 실시간 추적 및 인식, 컴퓨터 시각에 의한 사람의 동작 인식(수화 시스템), 골프 자세 수정 시스템, 암세포 인식, 얼굴 탐지 및 인식의 자동화에 의한 인물 데이터 베이스에서 특정 인물 찾기, 보안 지역에 대한 무인 감시, 음성 분석 및 인식에 의한 컴퓨터 인터페이스 구현 등에 활발한 연구 활동이 기대된다.

V. 결 론

본 연구개발에서는 신경회로망 이론을 이용하여 음성 인식 모듈을 개발을 연구하였다. 과학의 발달로 가전기기의 첨단화가 필수적인 시점에서 음성인식모듈을 개발하여 가전기기(TV, VCR, AUDIO etc)들을 음성인식을 통하여 제어함으로써 사용자의 편리성 제공과 제어기기의 고급화에 크게 기여할 수 있었다.

본 연구개발에서는 신경회로망 이론을 이용해 음성을 검출하여 가전기기 및 음성제어를 필요로 하는 통제시스템에 모니터링 시스템으로 사용가능 하도록 연구하였다.

사용 방법으로는 음성 마이크에 사람의 음성에서 나온 신호를 A/D변환기를 이용하여 디지털 신호로 변환하였으며 A/D 변환기에 의하여 읽혀진 데이터는 MICOM에 의한 파형의 분석을 통하여 검사를 수행하였다. 또한 사용자의 편의를 위하여 window상에서 Button에 의하여 시스템이 구동되도록 소프트웨어를 구성하였다. 본 음성인식 모듈은 on-line 실시간으로 입력 음성을 검사 및 보상이 수행되므로 노약자등 관련지식이 없는 사용자에 큰 도움이 될 것으로 사료된다.

본 연구개발에서 구성된 음성인식 모듈은 응용기기에 적용하여 사용할 수 있다. 또한, Window상에서 그래픽 프로그램으로 시스템을 구성할 수 있게 함으로서 개발환경이 열악한 중소기업에서도 개발하기에 적당하다. 즉 본 음성인식 개발모듈을 응용하기 위해서 숙련되지 않은 사용자(기술자)도 쉽게 사용할 수 있도록 구성하였다. 본 개발품은 on-line real time으로 수행되므로 사용자의 대기시간 없이 처리 가능함으로 사용자의 편리성을 극대화하였고, 개발업체에서 상품화 진행시 응용개발이 가능하게 처리했다. 본 연구 개발품을 응용 기기에 규정된 인터페이스만 구성하면 모든 기기를 연결할 수 있도록 개발 완료하였다. 향후 추가적인 개발은 각각의 응용 기기에 맞게 회로설계를 변경하여 보다 많은 응용 기기에 사용되기를 바란다. 앞으로 개발인력이 열악한 중소기업 기술지원에 보다 많은 노력을 하여 중소기업육성에 이바지하겠습니다.

참고문헌

[1] J. Nagumo and S. Sato, "On a Response Characteristics of a Mathematical Neuron Model", Kybernetik, 10,

pp. 155-164, 1972.

- [2] K. Aihara, T. Takabe and M. Toyoda, "Chaotic Neural Networks", Phys. Lett A144, pp 333-340, 1990.
- [3] Sang-Hee Kim, H.G. Choi, C.H. Chai, "Trajectory Control of Robotic Manipulator using Chaotic Neural Networks", IEEE ICNN, vol.3, pp.1685-1688, 1997
- [4] C.C. KU and K.Y.Lee. "System Identification and Control using Diagonal Recurrent Neural Networks", ACC/W M2, pp.545-549, 1992
- [5] 최운하, 김상희, Pregilter 형태의 카오틱 신경망 속도 보상기를 이용한 로봇 제어기 설계, 전기학회 논문지, 50권 4호, pp184-191, 2001
- [6] 김용표, 최운하, 신경회로망을 이용한 음성인식 모듈 개발, 구미기능대논문집, 2003.

저자소개

김 용 표(Yong-Pyo Kim)



금오공과대학교 공학석사
금오공과대학교 박사과정
대구기능대학 전자과 교수

※관심분야: 멀티미디어 및 신호처리 계측제어

윤 동 환(Dong-Han Yoon)



1968년 광운대학교 공학사
1980년 명지대학교 공학석사
1987년 명지대학교 공학박사
1979~현재 금오공과대학교 전자공학부교수

※관심분야: 멀티미디어 및 신호처리 센서공학

최 운 하(Choi-UnHa)



1996년 금오공과대학교 공학사
1998년 금오공과대학교 공학석사

※관심분야: 신경망, 제어계측 홈네트워크, 유비쿼터스