

工學博士 學位論文

하이브리드 음성인식 기법을 이용한 인공지능형
이동로봇 시스템의 설계 및 구현

Design and Implementation of Artificial Intelligent Mobile
Robot System using Hybrid Speech Recognition Method

指導教授 李 尙 培

2004年 2月

韓國海洋大學校 大學院

電子通信工學科

姜 成 仁

本 論 文 을 姜 成 仁 의
工 學 博 士 學 位 論 文 으 로 認 准 함

委 員 長	梁 圭 植	印
委 員	林 宰 弘	印
委 員	卓 漢 浩	印
委 員	韓 雲 英	印
委 員	李 尙 培	印

2004年 2月

韓 國 海 洋 大 學 校 大 學 院

電 子 通 信 工 學 科

목 차

목 차	i
그림 목록	iii
표 목록	v
Abstract	vi
제 1 장 서 론	1
제 2 장 인공지능	3
2.1 퍼지이론	3
2.2 신경회로망	9
제 3 장 음성인식	17
3.1 잡음처리 방법	18
3.2 끝점 검출 알고리즘	21
3.3 특징 추출 알고리즘	23
3.4 음성인식 알고리즘	29
제 4 장 음성인식 시스템	41
4.1 하이브리드형 HMM-NN 음성인식 시스템	41
4.2 음성인식 시스템의 하드웨어 구성 및 특징	43

4.3 음성인식 시스템의 구현	49
제 5 장 제안된 인공지능형 제어 시스템의 설계 및 구현	55
5.1 인공지능형 제어 시스템의 하드웨어 구성	55
5.2 인공지능형 제어 시스템의 설계	65
제 6 장 실험 및 고찰	72
6.1 실시간 음성인식 시스템	72
6.2 인공지능형 제어 시스템	80
제 7 장 결 론	83
참 고 문 헌	85
부 록 A	92
부 록 B	90

그림 목록

그림 2.1 2개의 규칙집합에 대한 퍼지 추론	7
그림 2.2 저수준의 퍼지 제어기 구조	8
그림 2.3 고수준의 적응 퍼지 제어기	9
그림 2.4 방사 기저 함수 신경회로망의 구조	13
그림 2.5 2차 평면에 대한 오차 경사하강법	16
그림 3.1 스펙트럼 차감법 블록도	20
그림 3.2 위너 필터링 블록도	21
그림 3.3 음성인식용 선형예측법 및 켈스트럼 블록도	24
그림 3.4 Durbin 알고리즘 순서도	25
그림 3.5 삼각대역 필터	28
그림 3.6 MFCC 처리 절차	29
그림 3.7 벡터 양자화의 전체 블록도	30
그림 3.8 이진 트리 알고리즘 순서도	31
그림 3.9 HMM 학습 알고리즘	34
그림 3.10 HMM 인식 알고리즘	38
그림 4.1 하이브리드형 HMM-NN 음성인식 시스템의 구조	42
그림 4.2 음성인식 보드의 전체 하드웨어 블록도	45
그림 4.3 시리얼 오디오 코덱 CS4218의 구조	46
그림 4.4 메인 시스템 보드의 구성	47
그림 4.5 메인 시스템의 메모리 맵	48
그림 4.6 어드레스 디코더의 구성	48
그림 4.7 음성구간 감지에 대한 끝점 검출	50

그림 4.8 HMM 인식 및 훈련단계	52
그림 4.9 설계된 방사 기저 함수 신경회로망의 구조	53
그림 5.1 인공지능형 제어 시스템의 전체 구성도	56
그림 5.2 전원 시스템 계통도	57
그림 5.3 모터 제어를 위한 하드웨어 구조	58
그림 5.4 조이스틱 구조 및 메커니즘	59
그림 5.5 조이스틱의 위치에 따른 모터의 회전방향 및 속도 설계	60
그림 5.6 엔코더를 이용한 모터의 회전속도 측정법	64
그림 5.7 휠체어 시스템의 퍼지 제어기 구조	66
그림 5.8 휠체어 주 퍼지 제어기의 입·출력 소속함수	68
그림 5.9 설계된 퍼지 제어기의 추론과정	69
그림 5.10 퍼지 제어기의 제어평면	70
그림 5.11 부 제어기 FLC2와 FLC3의 입·출력 소속함수	71
그림 6.1 명령어 “앞으로”의 음성파형	75
그림 6.2 명령어 “앞으로”의 MFCC 특징벡터	75
그림 6.3 명령어 “앞으로”의 벡터 양자화 결과	76
그림 6.4 하이브리드형 음성인식 시스템의 학습 및 테스트 절차	77
그림 6.5 방사 기저 함수 신경회로망의 학습과정	77
그림 6.6 1200rpm 기준 입력 시 속도 응답 파형과 PWM 출력	80
그림 6.7 1200rpm 기준 입력 시 PWM 파형의 듀티비	81
그림 6.8 2000rpm 기준 입력 시 속도 응답 파형과 PWM 출력	81
그림 6.9 2000rpm 기준 입력 시 PWM 파형의 듀티비	82

표 목 록

표 4.1 TMS320C32의 특징	44
표 5.1 조이스틱 변화에 따른 모터의 회전 방향과 전압	60
표 5.2 조이스틱 위치에 따른 전압 값과 모터방향	61
표 5.3 음성 단어별 출력 데이터	62
표 5.4 엔코더의 정격 및 사양	64
표 5.4 휠체어 동작속도(V_w) 및 동작방향(Θ_w)의 범위	67
표 5.5. 왼쪽 모터의 속도와 방향 추출을 위한 퍼지 추론 규칙	68
표 5.6 오른쪽 모터의 속도와 방향 추출을 위한 퍼지 추론 규칙	69
표 5.7 FLC2와 FLC3의 퍼지 추론 규칙	71
표 6.1 여러 가지 형태의 음성인식 시스템과 파라미터	73
표 6.2 음성인식 시스템의 형태에 따른 인식율	74
표 6.3 HMM을 이용한 음성인식 시스템의 테스트 결과	79
표 6.4 하이브리드형 음성인식 시스템의 테스트 결과	79

Abstract

With the increase of social concern about the disables and elderly people, their participation in social activities is demanded. In this view, motorized wheelchair system, one of the mobile robot, is necessary for giving them better mobility and for saving them a considerable physical. However, it still proves to be difficult or impossible to use for some handicapped person. Therefore, the teaching method with speech command is necessary to a handicapped person without hands or a non-expert to generate the path.

This paper presents the design of intelligent mobile robot system using a real time speech recognition in implementation of motorized wheelchair. The proposed mobile robot system is composed of four separated module, which are main control module, speech recognition module, servo motor driving module and sensor module.

In main control module with microprocessor(80C196KC), one part of the artificial intelligences, fuzzy logic, was applied to the proposed intelligent control system. In order to improve the non-linear characteristic which depend on an user's weight and variable environment, encoder attached to the servo motors was used for feedback control.

Also, in this paper, we design and implement a speaker independent recognition system using TI's DSP(TMS320C32). This system uses Hidden Markov Models for the designated command vocabularies to

control a mobile robot, and it has postprocessed by RBFNet(Radical Basis Function Neural Network) to distinguish some fuzzy word command well. As the spectral analysis method, we use a MFCC(Mel Frequency Cepstral Coefficient) to extract the features of the voice.

The proposed recognition system is tested using 9 words for control of the mobile robot, and the performance of a mobile robot using voice and joystick command is also evaluated.

제 1 장 서 론

최근 이동로봇은 공장자동화, 빌딩감시 등의 일반적인 산업현장에서부터 해양 및 우주탐사 등의 극한작업 분야나 청소대행 혹은 의료용 보조기 등의 역할을 위한 서비스 분야에까지 그 필요성이 절실히 요구되고 있다. 특히 요즘은 산업기술 발달과 함께 교통사고 및 각종 산업재해로 인한 후천적 장애인구의 증가와 노령층의 증가가 사회문제로 대두되고 있다. 따라서 이들의 사회활동 참여의 기회를 넓히고 사회활동을 보조하기 위한 수단으로 사용되어온 일반 수동 휠체어가 다양한 신체장애의 형태에 따라 최소한의 조작으로 이동성이 보장될 수 있도록 이동로봇의 하나인 전동휠체어로 대체되고 있는 추세이다^{[1][2][3]}.

최근 기존의 조이스틱을 통한 조작을 기본으로 하는 전동휠체어의 형태에서 벗어나 조이스틱 조작이 불가능한 장애인들을 대상으로 다양한 형태의 휠체어가 개발되고 있으며, 이를 위해 VAHM(Vehicle Autonomous Handicaps Motors) 프로젝트 등과 같은 맨-머신 인터페이스에 관한 연구도 활발히 진행 중이다^{[4][5][7]}. 특히 인간의 언어를 해석하여 적절한 행동을 수행할 수 있도록 하는 음성인식 기법이 적용되고 있으나 음성인식의 특성상 오인식에 의한 휠체어의 안전성의 저하에 의해 많은 문제점이 대두되고 있는 실정이다^{[2][5]}.

또한, 실제 장애인들의 전동휠체어에 의한 이동시에 보이지 않는 장애물 또는 움직이는 장애물 등과 같은 주변 환경과 휠체어 제어기의 기능이 단순히 조이스틱 입력에 따른 휠의 회전비율에 따라 결정되기 때문에 장애인들의 이동시 안전에 많은 문제점이 있다^{[3][6]}. 그리고 현재 휠체어 제품들은 장애인들의 장애정도에 따라서 사용여부가 결정되기 때문에 같

은 전동휠체어라도 장애가 심한 사용자를 위해서 더욱 많은 기능들이 요구되고 있는 실정이나 국내에서 제품화된 휠체어들은 조이스틱을 이용한 운전방식을 사용하고 있고, 휠체어의 동작을 지시하는 가장 중요한 부분인 제어기를 수입에 의존하고 있다^{[7][8]}.

이에 본 논문은 이러한 단점들을 보완하기 위하여 센서에 의해 주변 환경을 인식하고, 동적환경에서 최적으로 휠체어의 동작을 제어 할 수 있는 인공지능형 이동로봇 제어기를 전동휠체어에 적용하였다. 또한 장애 정도에 따른 사용범위 제한을 최소화 할 수 있도록 하기 위해 통계적 패턴인식을 이용한 HMM(Hidden Markov Model) 음성인식 기법에 인식률 향상을 위한 방법으로 신경회로망(NN : Neural Networks)을 사용하여 화자 독립형 음성인식 모듈을 임베디드 시스템(embedded system)으로 구성하였다.

본 논문의 구성은 다음과 같다. 2장에서는 본 논문에서 활용하고자 하는 퍼지 및 신경회로망을 소개하였으며, 3장은 본 논문의 음성인식 시스템에 사용되는 음성인식 기법에 대해 소개하였다. 4장은 임베디드 시스템으로 구성된 음성인식 시스템의 구현과 제안된 음성인식 알고리즘이 제안되었고, 5장에서는 제안된 인공지능형 제어 시스템의 구성 및 제어기 설계 및 구현에 대해 소개하였다. 6장에서는 음성인식 및 실제 제어기의 성능평가를 위한 실험 및 고찰, 끝으로 7장에서는 결론으로 시스템의 성능평가에 관한 검토 및 향후 연구방향에 대해서 제시하였다.

제 2 장 인공지능

본 절에서는 제안된 시스템의 성능을 향상시키기 위하여 필요한 인공지능 기법을 살펴보기로 한다. 먼저 인공지능이 무엇인가를 정의해 보면 다음과 같다. 지능이라는 것은 문제를 해결하는 능력을 의미하며, 인공지능(AI : Artificial Intelligence)이라 함은 문제를 해결하는 능력을 가진 알고리즘이나 기계를 만든다는 것을 의미한다^{[9][10]}.

인공지능의 연구분야들은 크게 세 부분으로 나누어 볼 수 있다. 하나는 학습(learning)을 연구하는 영역이며, 그 다음은 추론(inference)을 연구하는 영역, 마지막으로 인식(recognition)을 연구하는 영역으로 나눌 수 있다. 인공지능을 구현하기 위해서는 인간의 학습 메커니즘과 의사결정 메커니즘을 이해해야 한다.

2.1 퍼지이론

추론영역에서의 퍼지이론은 전문가 시스템에서 중요한 기능으로서 인간이 기계에게 어떤 질문을 하였을 때 기계가 그 질문의 의미를 파악하여 자신이 알고 있는 사실들을 조합하여 적절한 대답을 생성해 내는 것을 연구하는 분야이다. 그러한 의미로서의 추론은 그 질문에 대한 직접적인 사실이 데이터베이스에 없어도 자신이 가지고 있는 기존의 몇 개의 사실들과 그 사실들을 조작하는 규칙들을 이용하여 새로운 사실을 생성해내어 질문에 대답하는 것을 말한다^[11].

현대 제어이론이 제안된 후 이론적인 연구는 많은 사람들에 의해 수행되어 확립되어 왔으나 그 실용적인 면에 대해서는 시스템 모델링, 시스템

의 온라인 계산에 필요한 컴퓨터 용량 및 속도에 대한 제한, 경제성 등으로 인해 기대에 미치지 못하였다. 이에 반해 퍼지제어는 실용화가 대단히 간편하여 학문적 이론보다 오히려 실용화 사례가 앞서가는 실정이다. 이와 같은 이점과 유용성 때문에 퍼지이론은 제어뿐만이 아니라 광범위한 공학 분야에 걸쳐 응용될 수 있는 진정으로 「인간의 지혜」를 살린 기술이라 말할 수 있다.

2.1.1 퍼지 제어기의 개요

전형적인 제어 시스템은 일반적으로 제어대상의 수학적 모델을 알 수 있고, 모델이 선형적인 시스템에 대해서는 기존의 PID 제어, 극배치 제어, 최적제어 등의 제어기 설계기법을 적용하여, 폐루프 시스템(close-loop system)이 원하는 성능을 발휘하도록 제어기를 설계하는 것이 일반적인 제어 시스템이다. 한편, 미지의 선형시스템 제어기법으로는 적응제어 기법(adaptive control method), 학습제어 기법(learning control method) 등이 제시되고 있다^{[11]-[14]}. 선형으로 근사화할 수 있는 비선형시스템은 선형시스템 설계기법을 적용하여 어느 정도 원하는 성능을 얻을 수 있으나, 비선형성이 강하면 선형으로 근사화할 수 없기 때문에 한정된 분야에서만 적용되는 비선형시스템 제어기법들이 소개되고 있다.

전문가 시스템(expert system)은 시스템이 크고 복잡하며 모델이 비선형이거나 모델을 얻을 수 없는 시스템은 기존의 제어기 설계기법으로는 자동화가 어려워 충분한 경험적 지식을 가지고 있는 숙련자가 시스템을 작동시킨다. 이러한 전문가 시스템의 특징은 숙련자가 현재의 조작량(제어량)을 결정하는 데 경험적 지식을 활용하는 것으로 지능(즉, 학습능력과 판단능력)을 가지고 있는 사람이 시스템에 포함되어 있다는 것을 의미한다.

다. 이와 같이 숙련자 중심 시스템을 자동화할 수 있는 방법의 하나는 제어기에 인공지능을 부여하여 숙련된 작업자를 흉내내도록 하는 것이다. 이러한 인공지능 제어의 하나가 퍼지논리 제어인 것이다^{[10]-[15][17]}.

전문가 시스템을 자동화할 수 있는 한 가지의 제어기법은 숙련자의 경험적 지식과 제어 전문가의 지식을 활용하여 시스템 제어에 필요한 제어 규칙을 기술할 수 있다는 것이다. 즉, 제어기에 판단능력을 보유하도록 하는 것이다. 제어기는 숙련작업자의 지식과 제어전문가의 지식을 어떠한 규칙 또는 데이터베이스를 기초로 하여 제어기가 주어진 편차에 대해 조작량을 판단하고 결정하는 기능을 가지게 하는 숙련자 중심 시스템과 같이 지능을 갖춘 시스템이 퍼지제어 시스템인 것이다.

특히, 퍼지 제어기는 본질적으로 비선형적인 시스템을 대상으로 하는 비선형 제어기라고도 한다. 그러므로 선형 시스템과 같은 개념 구성이 매우 어렵다. 그러나 퍼지 제어시스템은 넓은 범위의 비선형성을 대상으로 하고 있으며, 퍼지 제어시스템에서의 퍼지모델은 비선형시스템의 특성을 복수개의 선형시스템의 조합으로 근사화하는 넓은 범위의 비선형시스템을 대상으로 한다는 것이다. 그 외에도 모델의 기술방법 또한 종래의 것과 매우 다르다^{[11][12]}.

종래의 기술은 정확한 수치를 이용하는 하나의 수식에 의해 기술되었고, 전제조건으로는 입·출력량과 시스템 파라미터를 분명히 알 수 있다는 가정 하에서 전개된다. 뿐만 아니라 시스템이 그러한 파라미터나 수식에 의해 지배된다는 가정 하에서 전개된다. 시스템 측면에서 보면 확정적인 선형모델을 알 수 있고, 정보가 확실하다고 하는 가정 하에서의 이론인 것이다. 그러나 실제 시스템에서는 어떤 모델도 실제 모델에 대한 근사치에 지나지 않고, 확률론적 모델로부터 얻어진 기대치에 불과한 어떤 경향적인 모델인 것이다. 그러므로 시스템을 모델링할 때에는 근사치를

고려하여야 하며, 불확실하게 모델링된 시스템을 제어할 때에는 그 모델에 대한 적절한 제어기법이 필요하게 되는 것이다.

2.1.2 퍼지추론

퍼지이론은 시스템 파라미터들 사이의 상관관계를 정량적으로 명확하게 규명하기가 힘들고 전문가의 경험에 의존하여 어느 정도 밝혀져 있는 경우에 매우 유용한 제어 알고리즘이다.

본 논문에 사용할 퍼지추론법은 일반적인 제어에 쉽게 사용되는 “*Mamdani’s minimum fuzzy implication rule*”를 소개하고자 한다.

$$\begin{aligned} R^1 &: \text{IF } x \text{ is } A_1 \text{ AND } y \text{ is } B_1, \text{ THEN } z \text{ is } C_1, \\ R^2 &: \text{IF } x \text{ is } A_2 \text{ AND } y \text{ is } B_2, \text{ THEN } z \text{ is } C_2. \end{aligned} \quad (2.1)$$

식 (2.1)과 같은 퍼지제어 규칙에 있어서 첫 번째 규칙의 연결강도를 α_1 , 두 번째 규칙의 연결강도를 α_2 로 정의하면 식 (2.2)와 같이 표현된다.

$$\begin{aligned} \alpha_1 &= \mu_{A_1}(x_0) \wedge \mu_{B_1}(y_0) \\ \alpha_2 &= \mu_{A_2}(x_0) \wedge \mu_{B_2}(y_0) \end{aligned} \quad (2.2)$$

식 (2.2)에서 $\mu_{A_1}(x_0)$ 와 $\mu_{B_1}(y_0)$ 은 퍼지 규칙집합(rule-based)에서의 데이터와 입력으로 주어지는 데이터 사이의 소속되는 소속정도이다.

Mamdani 추론법에서 i -번째 퍼지규칙(fuzzy rule)은 식 (2.3)과 같이 정의된다.

$$\mu_{C_i}(w) = \alpha_i \wedge \mu_{C_i}(w) \quad (2.3)$$

최종적인 결론으로서의 $\mu_C(w)$ 는 식 (2.4)와 같다.

$$\mu_C(w) = \mu_{C_1} \vee \mu_{C_2}$$

$$= [a_1 \wedge \mu_{C_1}(w)] \vee [a_2 \wedge \mu_{C_2}(w)] \quad (2.4)$$

여기에서 근사추론의 결과인 $\mu_C(w)$ 는 퍼지값이므로 실제 제어대상의 입력으로는 사용할 수 없다. 그러므로 $\mu_C(w)$ 를 실제 제어값으로 사용하기 위해서는 실수의 값으로 비퍼지화(defuzzification)하여야 한다.

일반적으로 *Mamdani*의 추론법에서는 식 (2.5)와 같은 무게중심법(center of gravity method)을 주로 사용한다^{[10][15][16]}.

$$u = \frac{\sum_{j=1}^n \mu_C(z_j) z_j}{\sum_{j=1}^n \mu_C(z_j)} \quad (2.5)$$

이러한 퍼지추론 과정을 도식적으로 나타내면 그림 2.1과 같다.

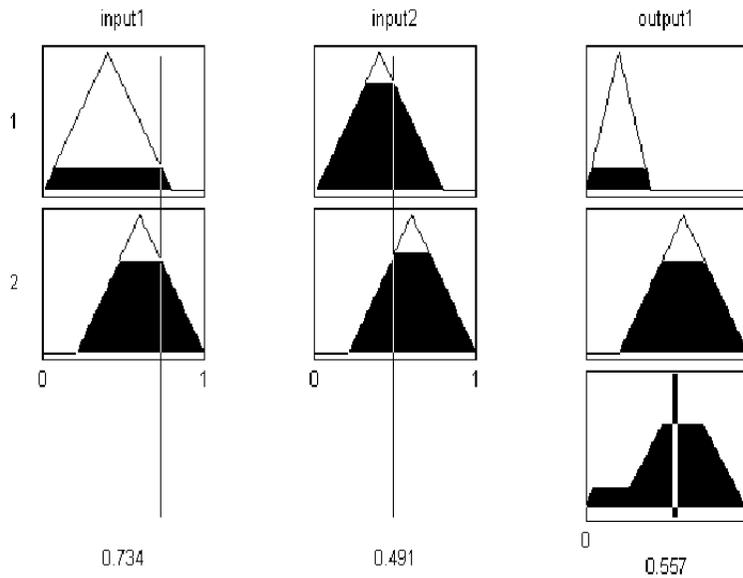


그림 2.1 2개의 규칙집합에 대한 퍼지추론

2.1.3 퍼지 제어기의 구조

퍼지제어기는 일반적으로 두 종류의 퍼지논리 제어구조를 가질 수 있다. 기본적으로 사용되는 제어구조는 대부분 그림 2.2와 같다. 이러한 제어기의 구조는 속도, 정확성, 그리고 감도에 관하여 몇 가지 결점을 가지고 있으며 아래의 조건들을 잘 고려해야 한다.

- ① 프로세서 시정수들이 상대적으로 높아야 한다(1ms 이상).
- ② 요구하는 제어 대역폭이 충분히 낮아야 한다.
- ③ 공정과정이 복잡하고 모호하게 정의되어 수학적 모델의 유도가 힘들고 실험에 의한 모델인식이 불가능 할 때 사용한다.
- ④ 숙련자가 제어 메커니즘에 포함되었고, 제어 전략을 퍼지집합과 같은 언어적 프로토콜의 집합으로서 표현될 수 있을 때 사용한다.

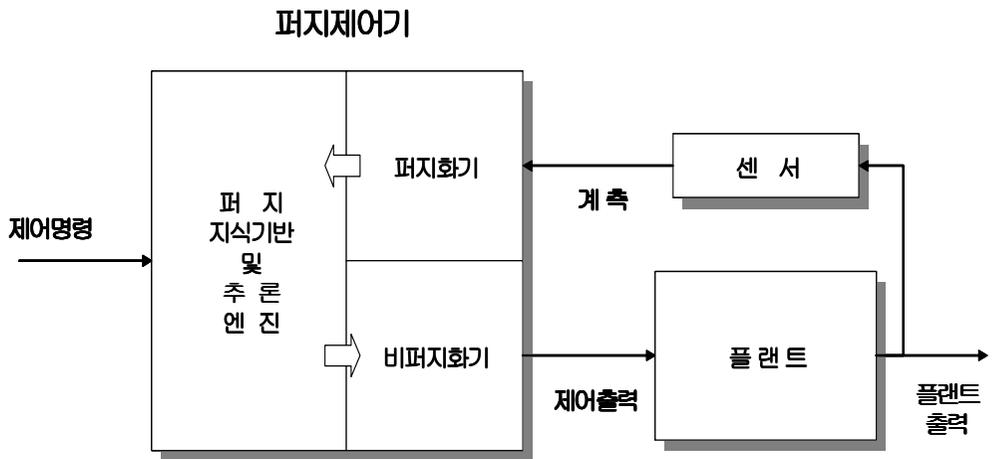


그림 2.2 저수준의 퍼지 제어기 구조

반면 동적인 시스템에서 요구되는 퍼지 제어기는 그림 2.3과 같이 계층

적인 구조를 가지고 있어야 효과적이다. 그림 2.3과 같은 적응 퍼지 제어기는 적응성 있게 동작되므로 일반적인 페루프 제어보다 가변성과 적응성이 뛰어나다. 전처리기에서 처리된 정보는 모니터링 시스템이 플랜트의 수행을 평가하기 위한 정보로 사용된다. 이러한 평가를 거쳐서 퍼지 의사결정 블록에서 시스템의 제어변수들을 조절하게 된다. 시스템 제어변수들을 온라인과 오프라인에서 조절하면 제어시스템은 자기조직 능력과 적응성을 가지게 되는 것이다.

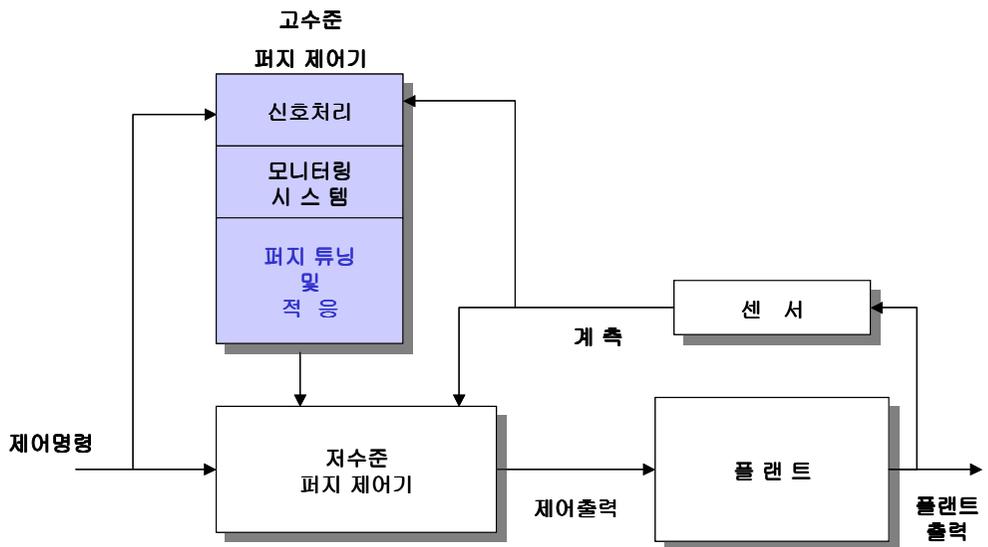


그림 2.3 고수준의 적응퍼지 제어기

2.2 신경 회로망

학습영역은 인간의 학습행위의 메커니즘을 모델링하고 개발하여서, 소프트웨어나 하드웨어가 자기 스스로 외부환경에 반응하고 학습하여 자신의 행위를 조절할 수 있도록 만드는 연구다. 인공지능에 있어서 학습이란,

기본적으로 새로운 지식의 획득에 의한 기계의 변화라고 할 수 있다. 이러한 점에서 학습이란 개별적인 것들로부터 공통적인 것, 즉 어떤 규칙이나 개념을 파악해내는 것을 의미한다.

인간의 두뇌는 수많은 뉴론(neuron)들이 서로 거미줄처럼 연결되어 있는 신경회로망 구조를 이루고 있다고 알려져 있다. 이와 같이 인간 두뇌의 신경회로망 조직을 모델링 한 것이 신경회로망 모델이며, 뇌(brain)에 존재하는 생물학적 신경세포와 그것들의 연결관계를 단순화시키고 수학적 으로 모델링 함으로서 뇌가 나타내는 지능적 형태를 구현하고자 하는 것이 신경회로망이다^{[16]-[20]}.

2.2.1 신경회로망의 구조

신경회로망이란 생명체의 신경조직을 모방하여 모델화한 정보처리 시스템으로 단순 소자들의 병렬, 분산된 연결 구조를 가지고 있으며, 외부로부터 들어오는 입력에 의하여 동적상태 응답(dynamic state response)을 일으킴으로써 필요한 출력을 형성시키는 것이다^{[17][20]}.

일반적으로 신경회로망은 지식의 내부적인 표현이 신경회로망 전체에 걸쳐 분산되어 있다. 기존의 기억장치의 경우 기억장치 일부의 파손은 그곳에 기억된 자료의 완전한 손실을 가져오지만 신경회로망의 경우는 신경회로망의 일부가 파손되더라도 그 정도가 약해질 뿐 작동은 가능하며 자료의 일부에 오류가 있더라도 그것으로 인해 최종 결과에 미치는 영향은 그리 크지 않다. 또한 신경회로망에서 학습능력이라고 하는 것은 신경회로망의 정보 저장에 있어서 특정한 응용목적에 따라 뉴론 간의 연결강도(connection weight)를 변화시키는 과정을 학습이라 한다^{[16]-[19]}. 이것은 동일한 정보를 기억하는데 있어서도 사람의 IQ 또는 학습방법에 따라 학습

시간이 다르듯이 신경회로망의 학습에 있어서도 어떠한 신경회로망 모델 또는 학습방법을 선택하느냐에 따라 학습시간이 빨라지기도 하고 느려지기도 하며, 심지어는 영원히 학습이 되지 않을 수도 있다는 사실을 염두에 두어야 한다. 이러한 특징을 가진 신경회로망은 최근 들어 컴퓨터의 발달로 인하여 공학, 자연과학, 의학 등 매우 다양한 분야에서 심도 있게 연구되고 있다.

신경회로망의 구조는 입력층(input layer)과 출력층(output layer) 그리고 은닉층(hidden layer)의 세 부분으로 구성된다. 입력층은 시스템의 외부로부터 입력을 받아들이는 층으로 사람의 감각기관에 존재하는 신경세포들에 해당한다. 출력층은 시스템의 외부로 출력을 내보내는 층으로 뇌의 명령을 운동기관에 전달하는 신경세포에 해당한다. 은닉층은 감각기관들로부터 전달된 자료들을 통해 반응을 결정하는 뇌에 해당된다. 각층의 신경세포는 정보전달정도를 나타내는 연결강도라는 시냅스(synapse)로 연결되어 있다^{[20][21]}. 그리고 신경회로망을 응용하는데 있어서 고려할 사항은 다음과 같다.

- ① 어떤 신경회로망 모델을 이용할 것인가?
- ② 신경회로망의 구조는 어떻게 구성할 것인가?
- ③ 신경회로망을 어떻게 효과적으로 학습시킬 것인가?
- ④ 학습패턴의 특징은 어떤 방법으로 추출할 것인가?
- ⑤ 실제구현은 어떻게 할 것인가?

신경회로망의 응용에 있어서 가장 먼저 고려해야 할 사항은 다양한 신경회로망 모델 중에서 응용목적에 적합한 신경회로망 모델을 선택하는 것이다. 왜냐하면 신경회로망 모델의 선택이 잘못되면 원하는 결과를 얻지 못할 수도 있기 때문이다.

신경회로망 모델을 선택한 다음에는 그 구조를 설계하여야 한다. 구조

의 설계라 함은 뉴런 수는 몇 개로 하며, 뉴런들을 어떤 형태로 배치하고 어떻게 연결시킬 것인지, 입력과 출력은 어떻게 할 것이지를 결정하는 것이다^{[26][29][30]}.

신경회로망의 구성이 완료되면 원하는 일을 할 수 있도록 신경회로망을 학습시켜야 한다. 사람이 새로운 학문을 습득하는 방법에는 학교에 가서 선생님께서 배우거나 자습서를 이용하여 혼자 독학하는 방법이 있듯이 신경회로망의 학습방법에도 지도학습(supervised learning)과 자율학습(unsupervised learning) 방법이 있다. 이들 학습방법은 신경회로망 모델에 따라 적합한 방법을 선택하여야 한다^{[17][18]}.

학습방법이 결정되면 신경회로망을 학습시킬 학습패턴을 선정하여야 한다. 여기서 학습패턴이라 함은 지도학습의 경우에는 입력패턴과 원하는 출력패턴의 짝을 의미하며, 자율학습의 경우에는 신경회로망이 자율적으로 결과를 출력하므로 입력패턴만을 의미한다. 또한 학습패턴의 특징추출이라 함은 패턴 중 특히 응용목적에 크게 영향을 미치는 중요한 부분만을 학습패턴으로 사용함으로써 정보량을 줄이는 것이다.

최종적으로는 실제 신경회로망을 어떻게 구현할 것인가를 결정하여야 한다. 신경회로망을 구현하는 방법에는 VLSI로 구현하는 방법, 광학적으로 구현하는 방법, 디지털 컴퓨터로 구현하는 방법 등이 있다.

2.2.2 방사 기저함수 신경회로망

방사 기저함수 신경회로망(RBF NN : Radial Basis Function Neural Network)은 다층 신경회로망을 대신해서 함수 근사화나 제어응용에 인기를 끌고 있다. 그 이유는 구조가 간단하고, 바라는 목표점에 대한 수렴속도가 빠르며 수학적인 배경이 잘 이루어져 있기 때문이다. 방사 기저함수

신경회로망의 특징으로는 지역적인 특성의 학습과 새로운 패턴 입력이 추가되어도 추가된 부분만 학습이 가능하다.

방사 기저함수 신경회로망의 학습 알고리즘은 지도학습으로서 패턴 군들에 대하여 가우시안(gaussian) 함수의 중심과 폭을 결정할 수 있게 학습한다^{[15]-[19]}.

I개의 입력, J개의 노드(node), K개의 출력을 가진 가우시안 방사기저 신경회로망은 그림 2.4에 나타내었다.

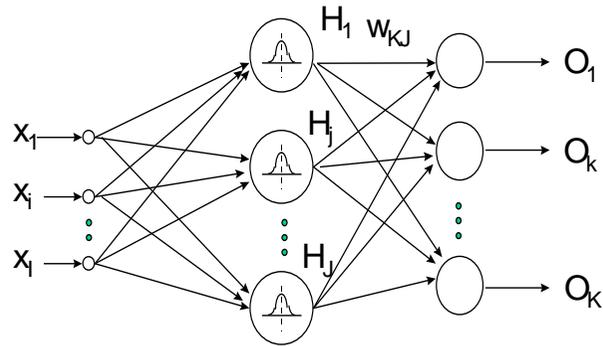


그림 2.4 방사 기저함수 신경회로망의 구조

그림 2.4에서 입력은 식 (2.6)이며,

$$\mathbf{x} = [x_1, x_2, x_3, \dots, x_i, \dots, x_I] \quad i=1,2,\dots,I \quad (2.6)$$

입력에 대한 은닉 층의 출력은 식 (2.7)에 나타내었다.

$$H_j = \exp\left[-\sum_{i=1}^I \frac{(x_i - m_{ji})^2}{\sigma_{ji}^2}\right] \quad j=1,2,\dots,J \quad (2.7)$$

여기서, m_{ji} 는 i -번째 입력에 대한 j -번째 방사 기저함수 함수의 중심값, σ_{ji}^2 는 i -번째 입력에 대한 j -번째 방사 기저함수 함수의 분산, H_j 는 은닉 층에 있는 j -번째 뉴런의 출력이다.

그리고 출력 층의 출력은 식 (2.8)과 같으며, O_k 는 출력층에 있는 K -번

째 뉴런의 출력이다.

$$O_k = \sum_{j=1}^J w_{kj} H_j \quad k = 1, 2, \dots, K \quad (2.8)$$

가우시안 함수의 중심과 표준편차의 선택은 양호한 근사화를 하기 위해 중요하다. 이들을 결정하기 위해 일반적으로 2가지 방법이 있다.

$x \ni [x_1, x_2, x_3, \dots, x_N]^T$ 가 R^N 에 완전하게 속한다고 가정하고, 첫째 방법은 입력영역의 모든 차원에 일정한 간격으로 방사 기저함수의 중심을 배치하는 것이고, 표준편차는 상수로 주어진다. 이 방법은 간단하지만 차원으로부터 손실이 있으며, 정확한 근사화를 보증할 수 있는 방사 기저함수의 노드 수를 결정하기가 어렵다. 두 번째 방법은 입력 데이터 집단에 대해 코호넨(Kohonen)의 자기-조직화 학습을 이용하여 자기-조직화된 집단의 중심에 방사 기저함수의 중심을 배치시키는 것이다. 표준편차는 어떤 한 중심과 그에 이웃한 중심사이의 거리를 짧게 설정한다. 이 방법은 패턴 분류의 문제에서 매우 실용적이며, 함수 근사화 문제에 대해서도 입력 데이터는 주어진 집합 x 에 의해 균등하게 분산되었다고 가정된다.

만약 가우시안 함수들의 중심이 극점과 변곡점 그리고 도함수의 절대값에 일치하는 표준편차에 위치한다면, 정밀한 근사화는 소수의 신경회로망 노드와 짧은 학습 시간으로도 접근할 수 있다. 즉 가우시안 함수의 균등하게 분산된 중심이 최상의 해일 수 없다. 따라서 함수 근사화를 위해서 가우시안 함수의 최적 중심과 폭을 결정하여야 한다. 출력오차를 최소화 하는 최적의 m_{ji} , σ_{ji} , w_{kj} 를 구하기 위해 신경회로망의 적응 학습 알고리즘은 다음과 같다.

출력오차는 식 (2.9)와 같다.

$$E = \frac{1}{2} \sum_{k=1}^K [d_k - O_k]^2 \quad (2.9)$$

여기서 d_k 는 바라는 출력이다. 신경회로망의 출력 오차가 최소가 되기 위한 학습 알고리즘은 아래의 식들로 이루어지며, 신경회로망의 출력층은 델타학습규칙과 같다.

$$w_{kj} \leftarrow w_{kj} + \Delta w_{kj} \quad (2.10)$$

$$\Delta w_{kj} = n_w \frac{-\partial E}{\partial w_{kj}} = n_w (d_k - O_k) H_j \quad (2.11)$$

출력 오차에 대한 출력층의 연결강도 변화량은 식 (2.11)과 같고, 출력 오차에 대한 가우시안 함수의 중심 변화량 Δm_{ji} 를 구하면 다음과 같다.

$$m_{ji} \leftarrow m_{ji} + \Delta m_{ji} \quad (2.12)$$

$$\Delta m_{ji} = -n_m \frac{\partial E}{\partial m_{ji}} = 2 \sum_{k=1}^K n_m (d_k - O_k) \cdot H_j \cdot \frac{(x_i - m_{ji})}{\sigma_{ji}^2} \quad (2.13)$$

그리고 출력 오차에 대한 가우시안 함수의 폭 변화량 $\Delta \sigma_{ji}$ 는 다음 식과 같다.

$$\sigma_{ji} \leftarrow \sigma_{ji} + \Delta \sigma_{ji} \quad (2.14)$$

$$\Delta \sigma_{ji} = -n_\sigma \frac{\partial E}{\partial \sigma_{ji}} = 2n_\sigma \sum_{k=1}^K (d_k - O_k) w_{kj} H_j \frac{(x_i - m_{ji})^2}{\sigma_{ji}^3} \quad (2.15)$$

여기서 n_w , n_σ , n_m 는 학습 상수이다. 또한 연결강도의 갱신은 학습의 가속화를 위해 모멘텀항(momentum coefficient)을 첨가한 경사하강법에 의해 학습되어진다.

$$\Delta w_{ij}(t+1) = a \Delta w_{ij}(t) - n \frac{\partial E}{\partial w_{ij}} \quad (2.16)$$

여기서 a 는 모멘텀항, n 는 학습률이다.

위 식을 그림으로 나타내면 그림 2.5의 오른쪽 부분과 같다.

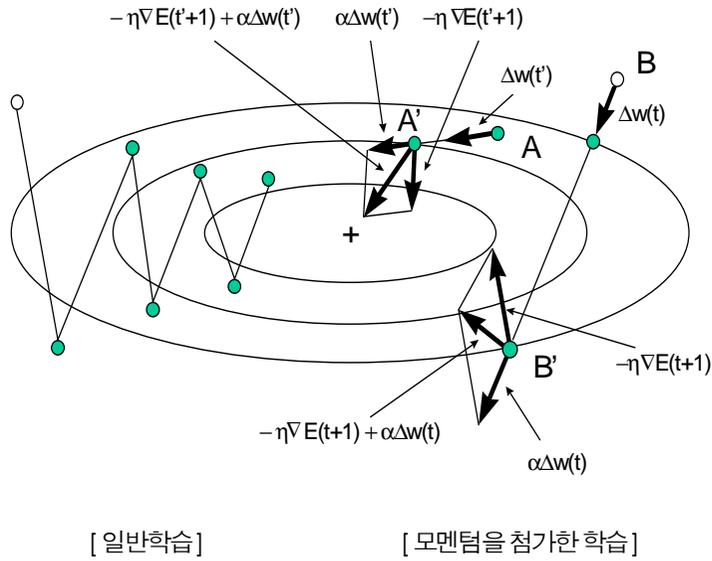


그림 2.5 2차 평면에 대한 오차 경사하강법

그림 2.5에서 왼쪽 부분은 모멘텀항이 없을 때의 경로를 나타내는 것이고, 오른쪽 부분은 모멘텀항을 추가했을 때에 나타나는 경로를 나타낸 그림이다. 위에서 보는 바와 같이 모멘텀항이 있을 때의 학습속도가 상대적으로 빠르다는 것을 알 수 있다^{[18][19][20]}.

일반적으로 모멘텀을 첨가한 학습 알고리즘은 학습이 느려지는 문제점을 해결하기 위하여 학습 단계에서 연결강도를 변화시킬 때 이전의 학습 단계의 연결강도도 변화량을 보조적으로 활용하는 방법이다

제 3 장 음성인식

음성처리 기술은 최근 HCI(Human Computer Interface)의 발전에 의해 많이 향상되고 있으며, 특히 음성인식 제품은 계속적으로 수요가 증가 추세에 있다.

음성인식 기술은 인간의 언어를 해석하여 적절한 행동을 수행할 수 있는 기계를 만드는 것을 목적으로 한다. 이러한 기술은 정보 산업 분야, 디지털 통신분야, 가전분야, 멀티미디어 등에 적용 범위가 크게 확대되고 있으며, PCS, 장난감, 전화기, 인터넷, 컴퓨터, 자동차 주변 장치(navigation system) 등 계속적으로 획기적인 제품들이 출시되고 있다.

여기서 음성인식 방법을 간단히 살펴보면, 패턴매칭의 방법을 이용하는 DTW(Dynamic Time Warping)와 통계적 패턴인식을 이용한 HMM, 뇌의 구조를 모델링한 학습패턴 방법인 NN이 대표적으로 많이 이용되고 있다. 이 3가지 방법은 각각 산업현장에서 적용이 활발히 진행 중에 있다 [32][36][37].

인식 방법에도 크게 2가지로 분류하면 화자 종속형(speaker dependent type)과 화자 독립형(speaker independent type)으로 나눌 수 있는데, 전자는 한 사람에게만 적용이 가능한 방법이며, 후자는 모든 사람에게 적용되는 것이다. 이 2가지 방법 중 화자 종속형으로 DTW 인식 알고리즘이 많이 사용되며, 때때로 NN도 사용되고 있다. 화자 독립형은 HMM 인식 알고리즘이 많이 사용되고 있으며, 음소별 음성인식에는 NN도 사용되고 있다 [33]-[37].

DTW를 이용한 인식방법은 훈련과정이 필요 없고, 고립단어 같은 작은 어휘 등에서 좋은 인식성능을 나타내므로 소규모나 DSP(Digital

Signal Processor) 응용 분야 등에서 널리 이용되고 있다^[41]. 또한 NN이 나 HMM도 지속적인 하드웨어 발전에 의해 임베디드 시스템에 적용되고 있는 실정이다.

따라서 본 장에서는 음성인식을 위한 여러 전처리 과정과 인식알고리즘으로 사용한 HMM 음성인식 과정에 대한 내용을 소개한다.

3.1 잡음처리 방법

잡음이 존재하는 환경에서는 음성인식 시스템의 인식률은 현저히 차이가 난다. 이처럼 잡음이 음성인식 시스템의 성능을 저하시키는 가장 큰 원인은 끝점 검출오차에 의한 영향과 잡음에 의한 특징 벡터의 왜곡이다. 본 음성 모듈에서는 잡음환경에 강한 시스템을 설계 및 구현을 하기 위해 첫 번째로 전제되어야 할 부분인 잡음을 처리하여 깨끗한 음성을 사용하기 위함이다.

본 절에서는 잡음이 포함된 음성신호 중 실음성만을 추출하는 방법으로 스펙트럼 차감법 및 위너 필터링 방법을 살펴본다.

3.1.1 스펙트럼 차감법

스펙트럼 차감법(spectral subtraction method)은 정적인 배경 잡음이 음성에 첨가되어 있고, 잡음이 새로운 상태로 바뀔 때는 새롭게 배경 잡음의 스펙트럼 크기를 추정할 충분한 시간이 필요하다^{[35][38]}. 만약 서서히 변화하는 비정적인 잡음 환경에서 음성구간과 잡음구간을 검출할 수 있는 알고리즘이 필요하다. 그리고 대부분의 잡음제거는 스펙트럼의 크기만으로 제거하는 것이 가능하다.

만약 잡음신호 $d(k)$ 가 음성신호 $s(k)$ 에 더해졌을 때 잡음이 섞인 음성신호 $y(k)$ 는 식 (3.1)과 같다.

$$y(k) = s(k)+d(k) \quad (3.1)$$

식 (3.1)을 주파수 영역으로 변환하면 식 (3.2)로 표현된다.

$$Y(w) = S(w)+D(w) \quad (3.2)$$

여기서, 식 (3.3)과 같이 정의한다.

$$Y(w) = \sum_{k=0}^{N-1} y(k)e^{-jwk}$$

$$y(k) = \frac{1}{2\pi} \int_{-\pi}^{\pi} Y(w)e^{jwk} dw \quad (3.3)$$

식 (3.2)를 이용하여 파워 스펙트럼 추정 방법을 적용해서 전개하면 다음과 같이 된다.

$$|Y_m(w)|^2 = |S_m(w)|^2 + |D_m(w)|^2 + 2S_m(w)D_m(w) \quad (3.4)$$

$$\approx |S_m(w)|^2 + |D_m(w)|^2 \quad (3.5)$$

여기서, 식 (3.6)과 같이 가정한다.

$$2S_m(w)D_m(w) = 0 \quad (3.6)$$

이후 식 (3.5)를 식 (3.7)으로 다시 나타내어, 이를 스펙트럼 차감법이라 한다.

$$|S_m(w)|^2 \approx |Y_m(w)|^2 - P_d(w) \quad (3.7)$$

$$P_d(w) = |D_m(w)|^2$$

식 (3.8)에서는 위상에 대한 정보를 나타낸 식이다.

$$S_m(w) = |S_m(w)| \angle Y_m(w) \Rightarrow S_m(n) = F^{-1}\{|S_m(w)|\} \quad (3.8)$$

이 때 스펙트럼 차감 필터의 $H(w)$ 는 미리 구한 잡음의 스펙트럼

$D_m(w)$ 의 평균을 사용한다. 그림 3.1은 스펙트럼 파워 차감법에 대한 간단한 블록도이다.

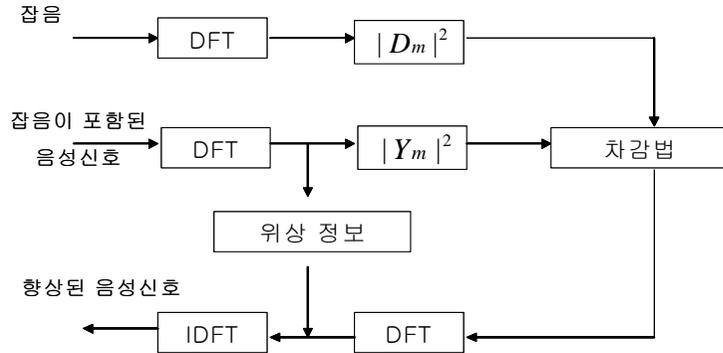


그림 3.1 스펙트럼 차감법 블록도

3.1.2 위너 필터링 방법

스펙트럼 차감법은 소음이 섞인 신호의 순시 스펙트럼과 소음의 시간 평균 스펙트럼을 이용하는 반면, 위너 필터링 방법(wiener filtering method)의 경우 다음 식 (3.9)와 같이 신호와 소음의 스펙트럼의 앙상블(ensemble) 평균을 이용한다^{[37][38]}. 여기서 $H(w)$ 는 위너 필터의 전달함수로 정의한다.

$$H(w) = \frac{P_s(w)}{P_y(w)} = \frac{P_s(w)}{P_s(w) + P_d(w)} \approx \begin{cases} 1, & P_s(w) \gg P_d(w) \\ 0, & P_s(w) \ll P_d(w) \end{cases} \quad (3.9)$$

그리고, $P_s(w)$ 는 깨끗한 음성신호 세그먼트이고, $P_y(w)$ 는 잡음이 섞인 음성신호 세그먼트이다.

만약 식 (3.9)에서 $P_s(w)$ 가 $P_d(w)$ 보다 훨씬 크다고 가정하면 1로 전달

되고, 반대로 $P_d(w)$ 가 $P_s(w)$ 보다 크다고 한다면 0으로 전달된다. 그림 3.2는 위너 필터링에 대한 블록도이다.

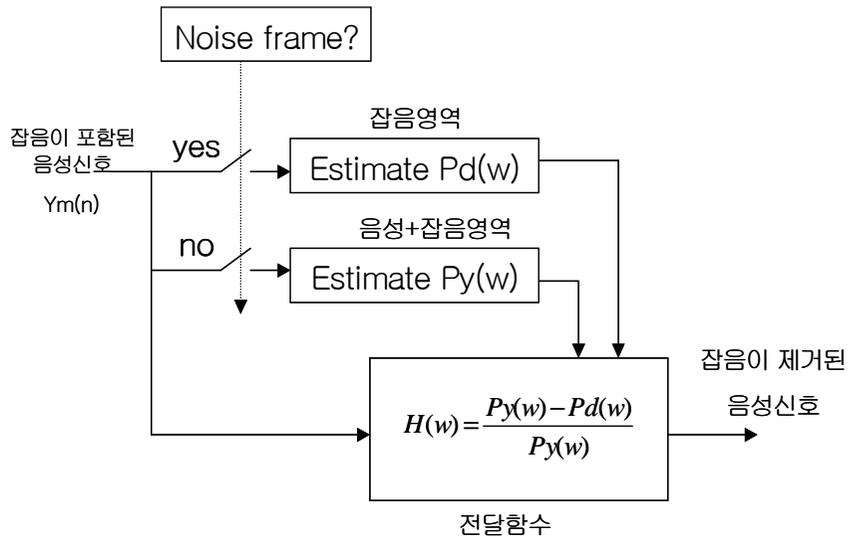


그림 3.2 위너 필터링 블록도

3.2 끝점 검출 알고리즘

음성신호를 처리하는 기본단위를 프레임이라 하고, 한 프레임에 포함되어 있는 신호의 샘플수를 프레임 길이라 한다^{[32][33]}. 프레임의 길이는 샘플링 주파수와 관계를 갖고 결정되는 것이 보편적이며, 본 논문에서는 한 프레임의 길이를 256 샘플로 테스트를 시도하였다.

프레임 사이의 간격은 가급적 작게 설정하여 처리를 시도하였으며, 프레임을 서로 중복시켜 프레임 사이에서 데이터의 손실을 방지할 수 있도록 하기 위해 프레임 중첩(frame blocking)을 적용했다^{[32][33]}. 본 실험에서는 80샘플을 중복시켜 적용해 보았다.

3.2.1 영교차율

영교차율(ZCR : Zero Crossing Rate)은 분석구간 프레임 내에서 파형이 영점(zero)측과 교차하는 횟수를 말하며 이산 신호에서 연속 샘플링 값이 서로 다른 대수적 부호를 가질 때 일어나는데 이는 음성의 분할, 분석, 인식에 있어 매우 유용하게 사용되어 왔다^{[32]-[35]}.

영교차율은 음성음과 무성음을 판별하는데 유용한 파라미터가 되며 음성신호 중 영교차율이 높으면 무성음으로, 영교차율이 낮으면 음성음으로 판단한다. 영교차율에 대한 일반식은 식 (3.10)과 같다.

$$Z_i = \sum_{n=0}^{N-1} |\text{sgn}[s_i(n)] - \text{sgn}[s_i(n-1)]| \quad (3.10)$$

$$\text{sgn}[x] = \begin{cases} 1, & x \geq 0 \\ -1, & x < 0 \end{cases}$$

영교차율은 실시간 음성인식 시스템에서는 성능이 떨어지는 단점이 있어 본 논문에서는 적용하지 않는다.

3.2.2 절대 에너지 방식

음성신호에서 절대 에너지는 무성음보다 음성음 부분이 크다^[35]. 절대 에너지 방식은 크게 Short-Time Energy, Short-time Magnitude 2가지 방식이 있다. 에너지 E/M은 이산계(discrete system)에서 다음 식 (3.11), 식 (3.12)로 주어진다.

$$E_i = \sum_{n=0}^{N-1} [s_i(n)]^2 \quad (3.11)$$

$$M_i = \sum_{n=0}^{N-1} |s_i(n)| \quad (3.12)$$

본 논문에서는 가장 간단한 방식인 절대 에너지, 식 (3.11)을 사용하여 음성 감지(detection)에 사용하였다. 여기서 말하는 음성의 감지는 유성음이나 무성음을 구별하는 방법이다.

3.3 특징 추출 알고리즘

3.3.1 선형예측법 및 켈스트럼

음성인식을 위한 음성 특징 분석방법에는 여러 가지가 있다. 그중에서 사람의 성도를 모델링한 선형예측법(LPC : Linear Predictive Coefficient)은 시간영역에서의 분석방법을 사용하여 계산량이 적고, 조용한 환경에서는 매우 우수한 인식 성능을 나타낸다. 그림 3.3은 음성인식을 위한 선형예측법 및 켈스트럼(LPC & Cepstrum) 처리 절차이다^{[36][37][38]}.

1) 프리엠피시스(pre-emphasis)

음성신호 $s(n)$ 은 스펙트럼적으로 신호를 평탄하게 하기 위해 저차의 디지털 필터를 통과시킨다.

$$\bar{s}(n) = s(n) - as(n-1), \quad (0.9 < a < 1.0) \quad (3.13)$$

2) 프레임 중첩(frame blocking)

프리엠피시스 과정을 거친 신호 $\bar{s}(n)$ 는 N개의 샘플들로 구성된 프레임

들로 인접한 프레임과 $N-M$ 샘플씩 중첩하여 처리되며, 여기서 N 은 프레임의 크기이고, M 은 프레임의 중첩 크기이다.

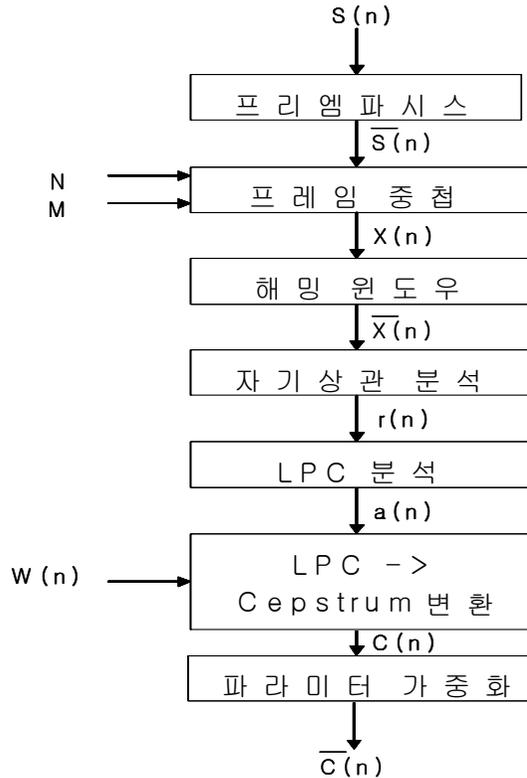


그림 3.3 음성인식용 선형예측법 및 켈스트럼 블록도

3) 윈도우

각 프레임의 시작과 끝에서의 신호 불연속을 최소화시키기 위해서 사용하며, 주로 해밍(Hamming) 윈도우가 많이 사용된다. 해밍 윈도우는 다음 식 (3.14)로 표현된다.

$$w(n) = 0.54 - 0.46 \cos\left(\frac{2\pi n}{N-1}\right), \quad (0 \leq n \leq N-1) \quad (3.14)$$

4) 자기상관 분석(autocorrelation analysis)

윈도우를 이용하여 추출한 각 신호 프레임으로부터 자기상관계수를 구하는 과정으로 식 (3.15)와 같다. 여기서 m 은 차수를 나타낸다.

$$r_1(m) = \sum_{n=0}^{N-1-m} x(n)x(n+m), \quad (m=0,1,\dots,p) \quad (3.15)$$

5) 선형예측법 분석

자기상관 분석과정을 거쳐 얻어진 $p+1$ 개의 자기상관 계수로부터 선형예측법 파라미터 세트로 변형하는 과정이 선형예측법 분석과정이다. Durbin 알고리즘은 그림 3.4에서 표현되어 있다.

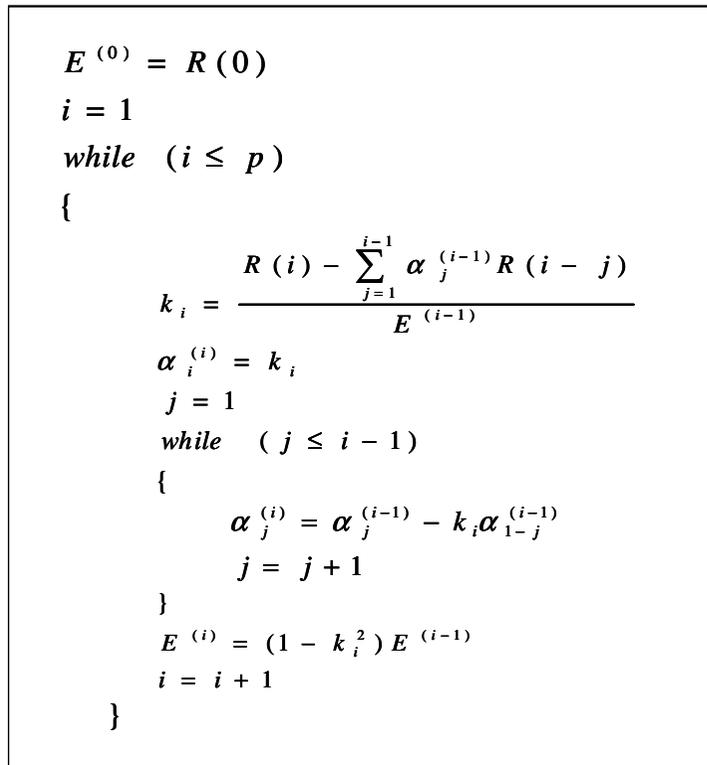


그림 3.4 Durbin 알고리즘 순서도

Durbin 알고리즘에서 계산되어 나온 출력 해는 식 (3.16), 식 (3.17), 식 (3.18)과 같이 된다.

$$a_m = LPC_coefficients = a_m^{(p)}, \quad (1 < m < p) \quad (3.16)$$

$$k_m = PARCOR_coefficients \quad (3.17)$$

$$g_m = \log_area_ratio_coefficients = \log\left(\frac{1-k_m}{1+k_m}\right) \quad (3.18)$$

6) 선형예측법 계수의 켈스트럼 계수로의 변환

켈스트럼은 진폭을 로그변환 후 푸리에 변환한 것으로 선형예측법 계수, PARCOR 계수, 대수영역비(log area ratio)계수보다 음성인식에 더욱 강건하고 신뢰할 만한 특징을 가지고 있다. 선형예측법을 켈스트럼으로 변환하는 식은 (3.19), (3.20), (3.21)와 같이 된다.

$$c_0 = \ln \sigma^2 \quad (3.19)$$

$$c_m = a_m + \sum_{k=1}^{m-1} \left(\frac{k}{m}\right) c_k a_{m-k}, \quad (1 < m < p) \quad (3.20)$$

$$c_m = \sum_{k=1}^{m-1} \left(\frac{k}{m}\right) c_k a_{m-k}, \quad (m > p) \quad (3.21)$$

7) 파라미터 가중화(parameter weighting)

저차 켈스트럼 계수는 전체 스펙트럼 기울기에 대해 민감하고, 고차 켈스트럼 계수는 잡음에 민감하므로 이를 완화하기 위해 가중치를 이용한다. 가중치 윈도우는 식 (3.22)과 같다.

$$w_m = \left[1 + \frac{Q}{2} \sin\left(\frac{\pi m}{Q}\right) \right], \quad (1 < m < Q) \quad (3.22)$$

여기서 Q는 캡스트럼 차수(12차 사용)이다.

3.3.2 멜 캡스트럼

잡음에 강인한 특징 벡터로는 멜 캡스트럼 계수(MFCC : Mel-Frequency Cepstrum Coefficient)와 루트 캡스트럼 계수 등이 있다^{[32]-[36]}. 이 중 본 논문에서는 음성인식 시스템에서 가장 많이 적용되고 있는 멜 캡스트럼에 대해서 설명하도록 하겠다.

1) 멜 캡스트럼 계수

캡스트럼 분석방법 중 인간의 청각 특성을 이용한 것으로 멜 캡스트럼이 있다. 멜은 톤 신호의 인지된 피치 또는 주파수 측정치를 나타내는 단위로서, 인간의 청각 시스템이 피치를 선형적으로 인지하지 않는 것처럼 톤 신호의 물리적인 주파수에 선형적으로 대응하지 않는다. Stevens와 Volkman은 임의로 1KHz를 1000mel로 선정하고, 사람들에게 인지된 피치가 기준 주파수의 두 배가 되도록 변화시켜 이를 2000mel로 표기하였다. 이러한 방법으로 실제 물리적인 주파수와 인지된 주파수 사이의 대응관계를 구하였다. 이 대응관계를 보면 1KHz 이하에서는 선형적으로 1KHz 이상에서는 대수적으로 대응됨을 볼 수 있다. 멜과 주파수사이의 대응 관계를 근사적으로 나타내면 식 (3.23)과 같다^{[36][37]}.

$$F_{mel} = 2595 \log_{10} \left(1 + \frac{F_{Hz}}{700} \right) \quad (3.23)$$

여기서, F_{mel} 은 각각 근사식에 의해 구해진 인지된 주파수이고, F_{Hz} 는 실제의 주파수를 나타낸다.

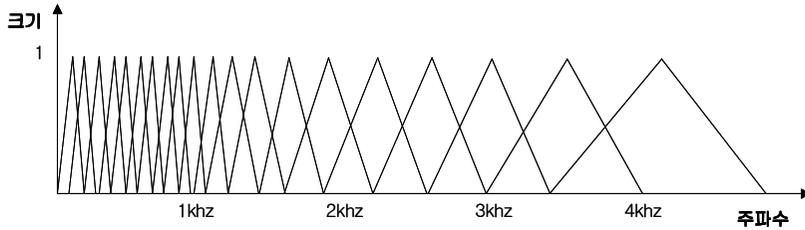


그림 3.5 삼각대역 필터

멜 캡스트럼은 DFT(Discrete Fourier Transform) 또는 FFT(Fast Fourier Transform) 크기를 멜과 주파수 사이의 대응 관계에 따라 주파수 축에서 와핑(warping)하여 이의 대수 값을 역 DFT/FFT하여 8에서 14차 정도의 계수를 구한다. 그림 3.5에 나타난 20개의 삼각 대역통과필터를 이용하여 임계 대역필터를 통과한 로그 에너지 출력을 X_k 라 하면 M개의 캡스트럼 계수는 식 (3.24)와 같이 나타내어진다^{[33][34][38]}.

$$c_n = \frac{1}{20} \sum_{k=1}^{20} X_k \cos \left[n \left(k - \frac{1}{2} \right) \frac{\pi}{20} \right] \quad (3.24)$$

여기서 C_0 는 음성 프레임의 평균 에너지로 초기화 값이다.

2) MFCC 처리 절차

그림 3.6에서는 MFCC의 처리 절차를 블록도로 표현했으며, 프리엠파시스 처리 후, 해밍 윈도우를 씌워주고, 시간영역을 주파수영역으로 변환하는 과정을 수행한다. 본 논문에서는 FFT를 사용했으며, 이후 20개의 로그형 멜 필터 뱅크를 거치는데 이 때 사용되는 필터는 그림 3.5에 나타내었다. 다음으로는 대수화 과정, DCT(Discrete Cosine Transform)를 거치게 되면 최종 1 프레임당 12 계수값이 나오게 된다. 이 때 DCT는 식

(3.24)에 나타나 있다.

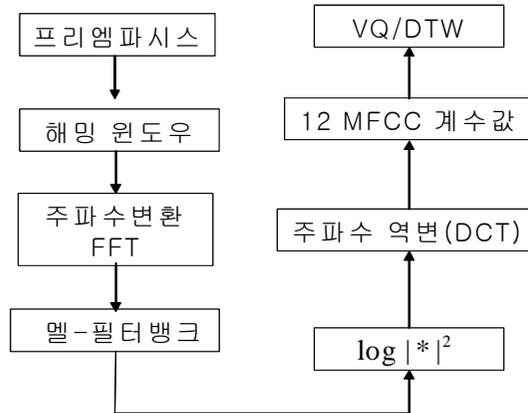


그림 3.6 MFCC 처리 절차

3.4 음성인식 알고리즘

3.4.1 벡터 양자화의 개요

벡터 양자화는 선형예측법 및 켈스트럼계수나 MFCC 계수 등을 이용하여 만들어진 양자화 테이블과 입력되는 음성신호로부터 추출한 계수와의 거리가 가장 가까운 양자화 테이블의 심볼을 얻는 것을 말한다^{[42]-[45]}. 양자화 테이블의 생성에는 K-means 알고리즘이나 이진 트리 알고리즘이 주로 사용되며, 본 논문에서는 양자화 테이블을 생성하기 위해서 이진 트리 알고리즘을 사용하였다^{[39][40][41]}. 그리고 벡터 양자화 크기인 M을 512로 설정하였으며, 오차비율은 0.000001, e(splitting parameter)값은 0.001로 각각 설정했다. 그림 3.7은 벡터 양자화의 개념도이다.

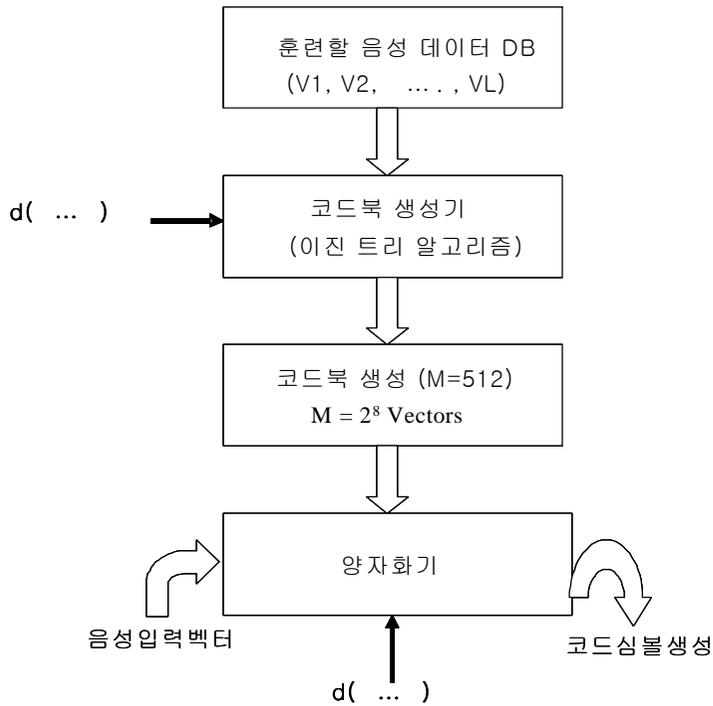


그림 3.7 벡터 양자화의 전체 블록도

벡터 양자화 프로그램의 순서는 다음과 같이 3가지로 구분할 수 있으며, 본 논문에서는 ②, ③을 집중적으로 분석하여 설명한다.

- ①특징추출이 저장되어 있는 수십-수백개의 파일을 한 개로 파일로 만드는 과정 - Vector Gather
- ②훈련할 데이터를 이용해서 코드북 크기(M=256-1024)에 따라 벡터 양자화 코드북 생성 - Clustering the Training Vector
- ③입력되는 음성신호로부터 추출한 계수와의 거리가 가장 가까운 양자화 테이블의 심볼 생성 - Vector Classification

1) 훈련 벡터의 분류

이진 트리 알고리즘은 1개의 벡터 코드북으로부터 분할(splitting)을 이용하여 M 개의 코드북을 작성하는 방법이다.

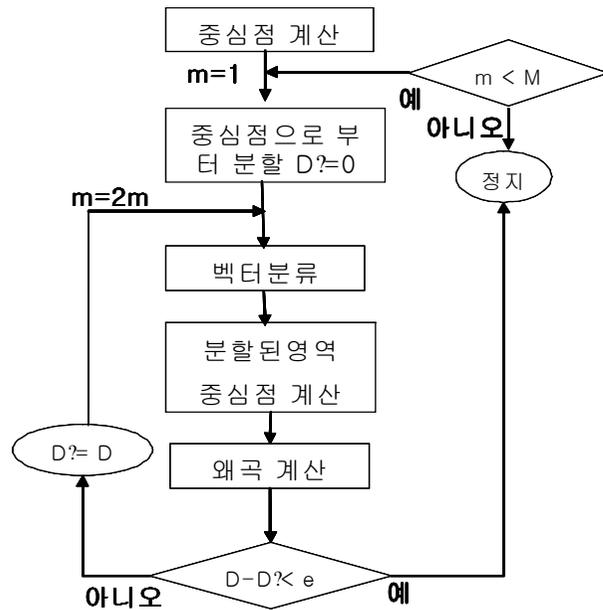


그림 3.8 이진 트리 알고리즘 순서도

이진 트리 알고리즘 순서는 다음과 같이 다섯 단계로 나눌 수 있고 그림 3.8에서 도식적으로 나타내었다.

- (1) 전체의 훈련데이터로부터 중심값으로 구성된 1개의 벡터 코드북을 생성한다. 즉, 전체 벡터들을 차수별로 합산 및 평균을 구하여 새로 생성된 코드북에 넣는다.
- (2) 현재의 코드북을 각각 분할하는데 아래의 식 (3.25), 식 (3.26)과 같

이 분할한다. 여기에서 e 는 분할 파라미터(splitting parameter)이고, Y_n 은 현재 벡터 코드북의 대표벡터 값이다.

$$Y_n = Y_n(1+e) \quad (3.25)$$

$$Y_n = Y_n(1-e) \quad (3.26)$$

- (3) 각 분할된 코드북에 대한 K-mean 알고리즘을 이용하여 최적 Centroid 벡터를 구한다. 즉 훈련할 벡터들을 분할된 코드북에 가까운 곳을 찾아 그곳에 소속시키고, 소속된 벡터들끼리 다시 Centroid 을 시켜 최상의 코드북을 생성한다.

$$C_n^{(i+1)} = \frac{\sum Q(X_m) = C_n^{(i)} X_m}{\sum Q(X_m) = C_n^{(i)} 1} \quad (3.27)$$

여기서 X_m 은 훈련할 데이터, $Q(X_m)$ 은 현재 소속 코드북이다.

- (4) 왜곡을 구한다. 왜곡은 코드북과 벡터사이의 거리들을 합산시켜서 평균화한 값을 왜곡이라 한다.

$$D_{ave}(i) = \frac{1}{M_k} \sum_{m=1}^M \| X_m - Q(X_m) \|^2 \quad (3.28)$$

여기서 X_m 은 훈련할 데이터, $Q(X_m)$ 는 현재 소속 코드북, M_k 는 소속된 코드북의 총 수, D_{ave} 는 평균 왜곡이다.

- (5) 코드북 크기가 $M(512)$ 에 도달할 때까지 (2)-(4)단계를 반복 수행한다.

2) 벡터 심볼 추출 절차

생성된 코드북과 새로운 입력 벡터들 사이의 최적의 코드북 심볼(index)을 출력하는 과정이다. 총 벡터북은 크기(M)는 512로 설정하였으

며, 최적의 코드북 심볼은 식 (3.29)와 같이 표현된다.

$$M^* = \arg \min_{1 \leq m \leq M} d(v, y_m) \quad (3.29)$$

여기서 y_m 은 M 차원 벡터 코드북($1 \leq m \leq M$)이고, v 는 특징추출된 벡터 입력값, m^* 은 최상의 벡터 인덱스값 행렬이다.

3.4.2 HMM 음성인식 알고리즘

HMM은 크게 이산 HMM(DHMM : Discrete HMM), 반연속 HMM(SCHMM : Semi-Continuous HMM), 연속 HMM(CHMM : Continuous HMM)으로 나눌 수 있다^{[32][33][46][47]}. 이 중 본 논문에서는 화자 독립형 독립단어를 구현하기 적합한 DHMM을 사용하였다. DHMM은 특징 벡터인 코드북을 만든 후 벡터 양자화를 통하여 코드북 인덱스로 관찰열을 만든다. 생성된 관찰열은 HMM 학습 알고리즘을 통해서 학습 뒤, 인식 알고리즘에서 새로운 단어가 입력이 되면 학습된 단어들과 각각 확률을 구해서 가장 큰 확률을 가지는 것을 인식하게 된다^{[47][48][49]}.

1) HMM에서 사용되는 구성요소

HMM은 상태 천이 확률(state transition probability), 상태 관찰 확률(state observation probability), 초기 상태 확률(initial state probability)인 세가지 파라미터로 구성된다. 이들을 간단히 모델 λ 로 표시하면 식 (3.30)과 같이 표현된다.

$$\lambda = (I, A, B) \quad (3.30)$$

① 초기상태 확률 $\Pi = \{\pi_i\}$

1부터 N까지 각각의 상태에서 t=1인 첫 번째 관찰이 일어날 확률을 나타낸다.

$$\pi_i = P(s_1 = i) \quad , \quad 1 \leq i \leq N \quad (3.31)$$

② 상태전이 확률 행렬 $A = \{a_{ij}\}$

각 상태간의 전이할 확률을 나타낸다. 시간 t에서의 상태 S_t 가 i일 때, 시간 t+1에서의 상태 S_{t+1} 이 j로 전이할 확률을 a_{ij} 로 나타내며, 이것은 행렬 A의 원소를 이룬다. 상태가 N개인 모델의 경우 N*N행렬이 된다.

$$a_{ij} = \Pr(S_{t+1}=j \mid S_t=i) \quad (1 \leq i, j \leq N) \quad (3.32)$$

③ 상태관찰 확률 $B = \{b_i(O_t)\}$

각 상태에서 특정한 관찰이 일어날 확률들의 집합이다. 시간 t에서 상태 j일 때, O_t 가 관찰될 확률을 $b_i(O_t)$ 로 나타낸다.

$$b_i(O_t) = \Pr(O_t|S_t=j) \quad (1 \leq j \leq N) \quad (3.33)$$

2) HMM 학습 알고리즘

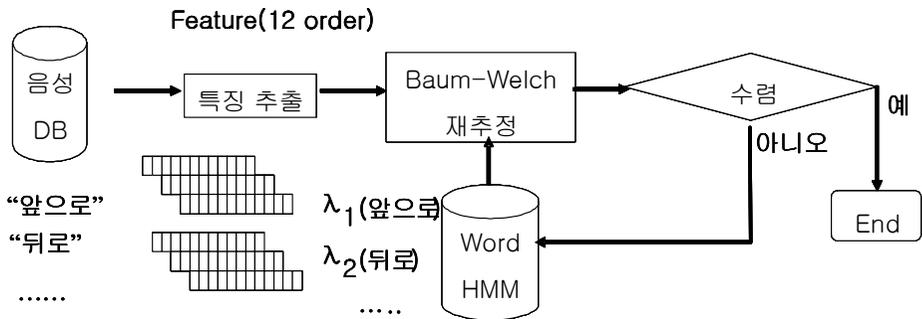


그림 3.9 HMM 학습 알고리즘

주어진 HMM에 대하여 학습은 해당 음성 데이터가 발생할 확률을 최대로 하는 모델 파라미터 λ 를 추정하는 것이다. 이러한 문제는 가장 널리 사용되고 있는 Baum-Welch 알고리즘을 이용하여 해결할 수 있다.

그림 3.9에서는 HMM을 이용하여 고립단어의 학습단계를 표현한 것이다. 여기서 $\lambda_1 \lambda_2 \lambda_3 \dots \lambda_n$ 에서 n 은 단어의 개수를 나타낸다. Baum-Welch 알고리즘을 사용하기 위하여, 우선 모델 λ 가 주어졌을 때 관찰열 $O=\{O_1, O_2, O_3, \dots, O_T\}$ 가 발생할 확률을 계산할 수 있어야 하며, 이는 다음과 같이 구할 수 있다. 시간 $t=1, 2, \dots, T$ 에서 HMM의 상태를 S 로 나타내면, $S=S_1, S_2, \dots, S_T$ 라고 표현된다.

$$\Pr(O | S, \lambda) = b_{S_1}(o_1) b_{S_2}(o_2) \dots b_{S_T}(o_T) \quad (3.34)$$

식 (3.34)는 관찰열의 상태가 주어졌을 때의 확률이 나오는 것을 나타낸다. 상태열 S 가 나올 확률은 다음 식 (3.35)와 같다.

$$\begin{aligned} \Pr(S | \lambda) &= \pi_{S_1} a_{S_1 S_2} a_{S_2 S_3} \dots a_{S_{T-1} S_T} \\ &= a_{S_0 S_1} a_{S_1 S_2} a_{S_2 S_3} \dots a_{S_{T-1} S_T} \end{aligned} \quad (3.35)$$

따라서 관찰열의 확률은 다음 식 (3.36), 식 (3.37)과 같이 주어진다.

$$\Pr(O, S | \lambda) = \Pr(O | S, \lambda) \Pr(S | \lambda) \quad (3.36)$$

$$\begin{aligned} \Pr(O | \lambda) &= \sum_{all(s)} \Pr(O | S, \lambda) \Pr(S | \lambda) \\ &= \sum_{all(s)} \prod_{t=1}^T a_{S_0 S_1} b_{S_1}(O_t) \end{aligned} \quad (3.37)$$

주어진 관찰열에서 가능한 모든 상태열에 대하여 합하여야 되기 때문에, 위 식 (3.36), 식 (3.37)의 계산량은 $O(N^T)$ 이다. 이를 해결하기 위해서 직접적으로 계산하지 않고 중간 변수를 도입하여 반복적으로 계산하는 전-후방 알고리즘(forward backward algorithm)이 제안되었다. 전방중간 변

수 $\alpha_t(i)$ 는 다음 식 (3.38)과 같이 정의된다.

$$\alpha_t(i) = \Pr(o_1, o_2, o_3, \dots, o_t, s_t = i | \lambda) \quad (3.38)$$

$\alpha_t(i)$ 는 모델 λ 에 대하여 시간 t 에서 상태 S_t 에 있을 때, 처음부터 t 까지의 관찰열 $O_1, O_2, O_3, \dots, O_t$ 을 관찰할 확률이다. 주어진 관찰열 O 의 확률 $\Pr(O|\lambda)$ 은 다음 식(3.39)와 같은 반복적인 과정에 의해서 구할 수 있다.

$$\begin{aligned} \alpha_t(i) &= \pi_i b(o_i) \quad , \quad (1 \leq i \leq N) \\ \alpha_{t+1} &= \left[\sum_{i=1}^N \alpha_t(i) a_{ij} \right] b_j(o_j) \quad , \quad (1 \leq t \leq T-1, 1 \leq j \leq N) \\ \Pr(O | \lambda) &= \sum_{i=1}^N \alpha_T(i) \end{aligned} \quad (3.39)$$

후방 중간 변수 $\beta_t(i)$ 도 비슷한 방법으로 식 (3.40)과 같이 정의되며, 시간 t 에서 상태 S_t 에 있을 때, 시간 $t+1$ 에서 끝까지의 관찰열 $O_{t+1}, O_{t+2}, O_{t+3}, \dots, O_T$ 를 관찰할 확률이다.

$$\beta_t(i) = \Pr(o_{t+1}, o_{t+2}, o_{t+3}, \dots, o_T | s_t = i, \lambda) \quad (3.40)$$

$\beta_t(i)$ 도 반복적인 방법으로 구할 수 있다.

$$\beta_t(i) = 1 \quad , \quad (1 \leq i \leq N) \quad (3.41)$$

$$\beta_{t+1}(j) = \sum_{i=1}^N a_{ij} b_i(o_{t+1}) \beta_{t+1}(i) \quad , \quad (1 \leq t \leq T-1, 1 \leq j \leq N) \quad (3.42)$$

$$\Pr(O | \lambda) = \sum_{i=1}^N \pi_i b_i(o_1) \beta_1(i) \quad (3.43)$$

이와 같이 구해진 $\Pr(O|\lambda)$ 를 ML(Maximum Likelihood)에 의해 최대화 하기 위해서 몇 가지 중간변수를 정의하고, 이들을 전방변수와 후방변수를 이용하여 추정한다.

$$\begin{aligned}
\varepsilon_t(i, j) &= \Pr(s_t = i, s_{t+1} = j | O, \lambda) \\
&= \frac{\alpha_t(i) a_{ij} b_j(o_{t+1}) \beta_{t+1}(j)}{\Pr(O | \lambda)} \\
&= \frac{\alpha_t(i) a_{ij} b_j(o_{t+1}) \beta_{t+1}(j)}{\sum_{i=1}^N \sum_{j=1}^N \alpha_t(i) a_{ij} b_j(o_{t+1}) \beta_{t+1}(j)}
\end{aligned} \tag{3.44}$$

$$\begin{aligned}
\gamma_t(i) &= \Pr(s_t = i | O, \lambda) \\
&= \frac{\alpha_t(i) \beta_t(i)}{\sum_{i=1}^N \sum_{j=1}^N \alpha_t(i) a_{ij} b_j(o_{t+1}) \beta_{t+1}(j)}
\end{aligned} \tag{3.45}$$

정의된 중간변수를 사용하여 HMM의 파라미터를 계산한다. 재 추정된 파라미터 $\overline{\pi}_i, \overline{a}_{ij}, \overline{b}_j$ 로 기존의 파라미터를 대치하고 재 추정 과정을 반복한다. 이 과정은 계속 반복되다가 $\Pr(O|\lambda)$ 가 수렴하면 멈춘다. 다음에 중간 변수를 사용한 재 추정은 식 (3.46), (3.47), (3.48)과 같다.

$$\overline{\pi}_i = \gamma_1(i) \tag{3.46}$$

$$\overline{a}_{ij} = \frac{\sum_{t=1}^{T-1} \varepsilon_t(i, j)}{\sum_{t=1}^{T-1} \gamma_t(i)}, \quad (1 \leq t \leq T-1, 1 \leq j \leq N) \tag{3.47}$$

$$\overline{b}_i(k) = \frac{\sum_{t \in \{t|o_t=k\}} \gamma_t(i)}{\sum_{t=1}^T \gamma_t(i)} \tag{3.48}$$

3) HMM의 인식 알고리즘

그림 3.10에서는 HMM의 인식 단계를 표현하는 것으로 새로운 단어가 입력되면 특징 추출단계 이후 훈련된 각 단어별 λ 와의 확률계산을 통해서 가장 확률 값이 높은 것을 인식한다.

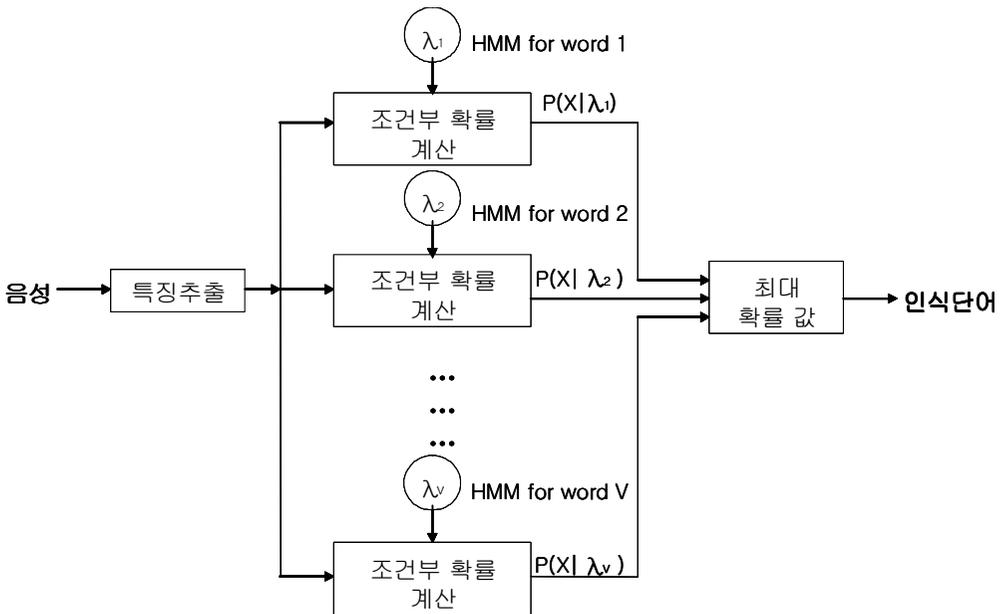


그림 3.10 HMM 인식 알고리즘

미지의 상태열을 구하기 위해 HMM을 이용한 인식과정은 학습된 모델 λ 와 관찰열 $O=\{O_1, O_2, O_3, \dots, O_T\}$ 가 주어졌을 때 $\Pr(O, S|\lambda)$ 가 최대가 되도록 하는 상태열 $S^*=\{q_1, q_2, q_3, \dots, q_T\}$ 를 찾는 것이다. 이를 비터비 알고리즘(viterbi algorithm)이라 하며, 이 방법으로 앞의 훈련된 각 단어별 λ 와의 확률계산을 통해서 가장 확률 값이 높은 것을 인식하는 문제를 해결할 수 있다.

전후방 알고리즘에서 HMM 모델 λ 에 대한 주어진 관찰열 O 의 확률 $\Pr(O|\lambda)$ 는 모든 가능한 상태 열들에서 얻어진 확률들이 누적되어 있는 값이다. 이에 반해, 비터비 알고리즘은 이 누적 확률 중에서 관찰열이 가질 수 있는 최적의 상태열에서 얻어지는 확률을 구하는 것이다. 이것은 시간 t 에서 N 까지의 최적 상태 열로부터 얻어지는 확률 $\delta_t(i)$ 와 최적의 상태열 $\phi_t(i)$ 의 두 변수를 이용하여 관찰열 O 에 대한 최적의 확률 P^* 를 구하는 방법이다. 최적의 상태열 s^* 은 최종 시간으로부터 가장 가능성이 큰 모든 상태를 역으로 추적하여 구한다. 비터비 알고리즘의 과정은 다음과 같다.

① 초기화

$$\delta_1(i) = \pi_i b_i(o_1), \quad (1 \leq i \leq N) \quad (3.49)$$

$$\phi_1(i) = 0 \quad (3.50)$$

② 반복

$$\delta_t(j) = \max_i [\delta_{t-1}(i) a_{ij}] b_j(o_t), \quad (2 \leq t \leq T) \quad (3.51)$$

$$\phi_t(j) = \arg \max_i [\delta_{t-1}(i) a_{ij}] \quad 1 \leq j \leq N \quad (3.52)$$

③ 종료

$$P^* = \max_i [\delta_T(i)] \quad (3.53)$$

$$q_s = \arg \max_i [\delta_T(i)] \quad (3.54)$$

④ 최적의 상태열 역 추적

$$q_s = \phi_{t+1}(q_{t+1}) \quad (t=T-1, T-2, \dots, 1) \quad (3.55)$$

이와 같은 방법을 이용하여 얻어진 각 단어들의 P^* 를 후처리 절차를 통해서 인식 결과를 얻어내게 된다.

4) HMM 구현에서 필요한 사항

① 좌우모델

음성신호는 시간에 따라 변하므로 HMM의 상태간의 천이 모델은 한쪽 방향으로 흘러가는 즉, 좌우모델(left to right model)이 음성신호 특성에 가장 적합한 방식이다. 이 모델은 현재 상태보다 앞 상태로 천이할 수 없도록 하며, 다음 상태로 천이하는 간격도 최대 천이값 p 로 정해준다. 이 경우 상태 천이확률 A 는 상위 삼각 행렬(upper triangular matrix)이 된다.

② 스케일링(scaling)

HMM을 실제로 구현할 때 전후방 알고리즘을 이용하여 $\alpha_t(i)$ 와 $\beta_t(i)$ 를 구할 경우 0과 1사이의 값을 계속 곱하면 언더플로우 현상이 생겨 결국 0의 값을 가지게 된다. 이를 없애기 위해 스케일링을 시켜 주면 값을 그대로 유지 시켜줄 수 있다. $\alpha_t(i)$ 와 $\beta_t(i)$ 의 정규화된 c_t 는 다음 식 (3.56), 식 (3.57), 식 (3.58)과 같다.

$$c_t = \frac{1}{\sum_{i=1}^N \alpha_t(i)} \quad (3.56)$$

$$\overline{\alpha}_t(i) = c_t \alpha_t(i) \quad (3.57)$$

$$\overline{\beta}_t(i) = c_t \beta_t(i) \quad (3.58)$$

제 4 장 음성인식 시스템

본 논문에서 구현된 음성인식 시스템은 HMM을 이용하여 인식을 수행할 경우의 오인식을 최소화하기 위한 후처리 과정으로 신경회로망의 일종인 방사기저함수 신경회로망을 결합하여 보다 높은 인식률을 얻을 수 있는 HMM과 방사 기저함수 신경회로망의 하이브리드 시스템으로 구현하였다.

또한 제안된 화자 독립형 음성인식 시스템을 휠체어에 탑재할 수 있도록 소형으로 구현하기 위해서 일반 PC 기반이 아닌 Texas Instrument (TI)사의 DSP TMS320C32를 사용한 임베디드 시스템기반으로 구현하였다. 이에 4.1절에서는 본 논문에서 제안한 HMM과 방사 기저함수 신경회로망의 하이브리드 음성인식 시스템의 구조에 대하여 살펴보고 4.2절에서는 음성인식 시스템의 구현을 위해 설계된 임베디드 시스템을 소개한다.

4.1 하이브리드형 HMM-NN 음성인식 시스템

HMM은 음성의 변동을 통계적으로 처리하고 이 통계량을 확률형태의 모델에 반영하여 음성을 인식하는 방법으로 음성을 모델링하는 효율적인 방법이며 실제로 좋은 결과를 얻을 수 있다. 그럼에도 불구하고 HMM은 음성데이터를 모델링하는 확률, 통계적인 가정에서 출발하는 한계를 가지는 단점을 갖고 있다^{[50]-[53]}. 따라서 입출력 간의 매핑에 의해서 패턴을 분류하는 신경회로망을 사용하여 HMM의 오인식을 보상하고자 한다. 특히 신경회로망 중에서도 구조가 간단하고 계산속도가 다른 신경회로망 모델에 비하여 빠른 방사 기저함수 신경회로망을 사용하였다.

따라서 본 논문에서는 화자 독립형 고립단어 인식에 적합한 DHMM을 사용한다. 각 분류목록에 대응하는 HMM 학습 테이블을 만든 다음 HMM 학습에 참여하지 않은 데이터를 학습 완료된 HMM 테이블을 사용하여 각각의 출력확률을 계산한다. 이러한 과정으로 각 입력 단어에 대응하는 하나의 출력확률 벡터를 만든다. 이렇게 생성된 출력확률 벡터열을 방사 기저함수 신경회로망의 입력층으로 인가하여 방사 기저함수 신경회로망을 학습하는 하이브리드 음성인식 시스템을 구성한다. 그림 4.1은 제안된 하이브리드형 HMM-NN 음성인식 시스템을 나타낸다.

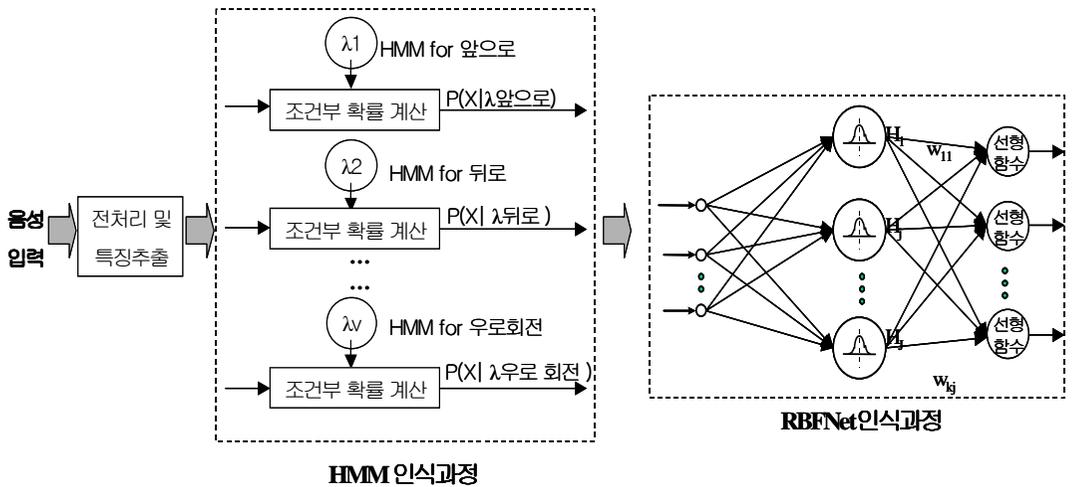


그림 4.1 하이브리드형 HMM-NN 음성인식 시스템의 구조

HMM인식과정 중에 오인식 되어지는 데이터의 특성을 반영하기 위해서 HMM 학습에 참여하지 않은 데이터를 사용하여 출력확률 벡터열을 구하고 방사 기저함수 신경회로망의 입력으로 사용하여 학습시켰다. 또한 구성된 하이브리드형 HMM-NN 음성인식 시스템의 성능을 테스트하기 위한 방법으로 HMM과 NN에의 학습에 사용되지 않은 제 3의 데이터를

사용하여 HMM으로만 학습한 음성인식 시스템과 인식결과를 비교하였다.

HMM의 확률, 통계적인 가정에서 출발하는 한계를 가지는 단점을 활성화 함수에 비선형적인 가우시안 함수를 사용하여 입출력간의 비선형 특성을 학습할 수 있는 방사 기저함수 신경회로망에 의하여 후처리함으로써 HMM의 인식률을 향상시킬 수 있었다.

또한 제안된 하이브리드형 HMM-NN 시스템은 먼저 PC상에서 학습 및 성능평가과정을 거쳐서 HMM의 학습데이터인 모델 파라미터 λ 와 신경회로망의 학습 파라미터인 가우시안 함수의 중심(m) 및 폭을 나타내는 표준편차(σ)를 임베디드 시스템인 TMS320C32로 구성된 음성인식 모듈에 사용하여 실시간으로 음성인식을 실행하였다.

4.2 음성인식 시스템의 하드웨어 구성 및 특징

4.1절에서 제안한 하이브리드형 HMM-NN 음성인식 시스템을 실시간으로 처리하고 또한 월체어 탑재용으로 제작하기 위해서는 모듈형으로 구성할 수 있는 임베디드 시스템으로 구성해야하는 필요성이 있다. 음성신호 처리의 특성상 많은 양의 데이터를 고속으로 처리해야 하고 실시간으로 동작되는 시스템을 구성하기 위해서 충분히 속도가 빠른 TI사에서 개발한 고성능 32비트 부동소수점 연산방식의 범용 DSP인 TMS320C32를 사용하였다. TMS320C32의 내부를 살펴보면 실수 연산에 적합하도록 설계된 내부 산술연산 장치(ALU : Arithmetic Logic Unit) 뿐만 아니라, 32비트 데이터 버스와 24비트 어드레스 버스, 2개의 타이머, 2채널 DMA, 그리고 직렬통신 포트 등이 갖추어져 있음을 알 수 있다.

이러한 TMS320C32의 특징 중 중요한 것들을 정리하면 표 4.1과 같다.

표 4.1 TMS320C32의 특징

특징	C32
외부 버스	One bus: ■ 32비트 데이터, 24비트 주소 0h-7FFFFFFh와 880000f-8FFFFFFh에서 $\overline{STRB0}$ 동작 ■ 8, 16, 32비트 와이드메모리에 8, 16, 32비트 데이터, 900000h-FFFFFFh에서 $\overline{STRB1}$ 동작 ■ 8, 16, 32비트 와이드메모리에 8, 16, 32비트 데이터, 81000h-82FFFFh에서 \overline{IOSTRB} 동작
ROM	없음
Boot loader	있음
On-chip RAM	512 주소: 810000h-82FFFFh
DMA	2채널 순위지정가능
시리얼 포트	1
타이머	2
인터럽트	레벨트리거 또는 에지, 레벨트리거 혼합
인터럽트 벡터 테이블	이동 지정가능 (relocatable)
패키지	144 PQFP
전압	5V
온도	0° ~ 85°C (상업용) -40° ~ 125°C (확장판) -55° ~ 125°C (군사용)

그리고 그림 4.2는 TMS320C32를 사용해 구현된 음성인식보드의 전체 블록도를 나타낸 것이다.

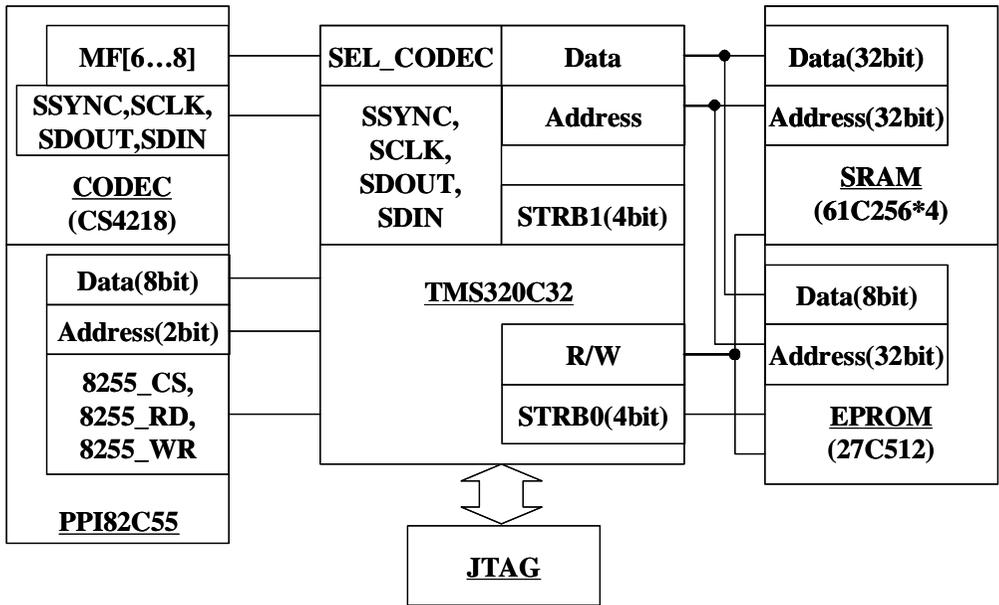


그림 4.2 음성인식 보드의 전체 하드웨어 블록도

4.2.1 음성입력부(CS4218)

마이크를 통해서 입력된 음성신호는 아날로그 신호이다. 따라서 먼저 TLC2272라는 OP-AMP를 사용하여 입력신호인 음성신호를 증폭시킨다. 또한 사람의 음성은 약 100Hz에서 8KHz의 범위를 가지므로 증폭된 음성신호는 16비트 분해능을 가진 시리얼 오디오 코덱인 CS4218에 의해서 최대 주파수 8KHz로 샘플링되어 입력된다. CS4218은 내부에 AD와 DA 변환처리를 할 수 있는 기능이 내장되어 있고, 속음(decimation) 필터와 평활(smoothing) 필터를 내장하고 있어서, 별도의 필터를 필요로 하지 않는다. 그리고 레벨 세팅을 수행하며 디지털 컴퓨터 시스템을 위해 4개의 오디오 입력과 2개의 오디오 출력을 가진다. 데이터는 SM4 모드에 SMOD3

의 세팅에 의해서 처음 16비트는 좌측오디오 채널로 다음 16비트는 우측 오디오 채널로 보내어 진다. 그림 4.3은 CS4218의 구조를 나타내고 있다.

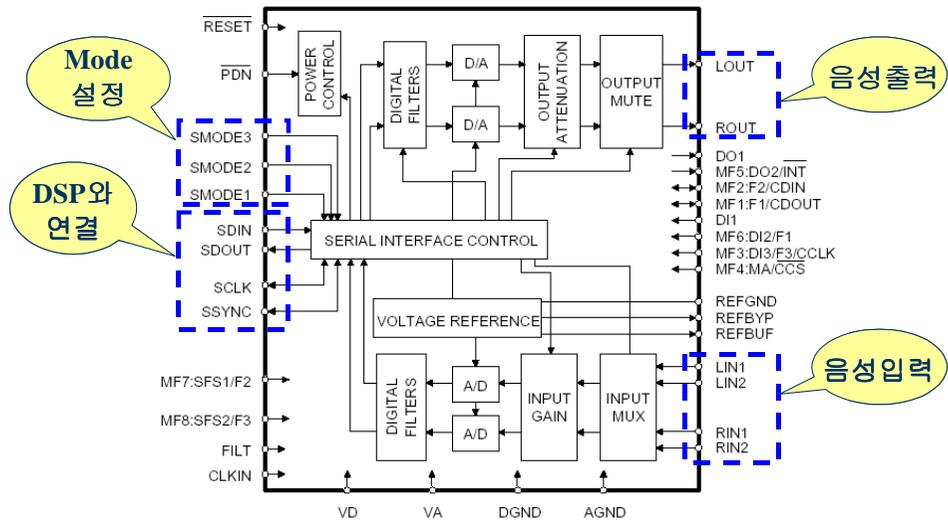


그림 4.3 시리얼 오디오 코덱 CS4218의 구조

4.2.2 메인 시스템의 구성

TMS320C32를 사용한 메인 시스템 보드의 구성을 그림 4.4에 나타내었다. 먼저 메모리로는 RAM에 SRAM 61C256*4(128Kbyte)을 사용하였고 ROM에 64Kbyte EPROM인 27C512를 사용하고 있다. STRB0, 1과 IOSTRB는 TMS320C32에서 제공하는 메모리 인터페이스를 위한 포트이다. RAM은 32비트 액세스를 위해 4개의 STRB1단자에 SRAM의 CS단자를 각각 연결해서 4개의 SRAM을 동시에 액세스하여 32비트로 액세스할 수 있도록 인터페이스를 구성하였고, ROM은 1개의 STRB0 단자에 EPROM의 CS단자를 연결해서 8비트로 액세스하는 방식을 취하고 있다.

그리고 IOSTRB는 PPI82C55와 코덱인 CS4218의 인터페이스에 사용하였다. 이와 같이 인터페이스된 메모리와 주변 장치들의 어드레스 맵은 그림 4.5와 같다. STRB0는 EPROM의 액세스 영역으로 0x100H이며, STRB1은 SRAM의 액세스 영역으로 0x900000H이다. 그리고 IOSTRB는 PPI82C55는 0x810000H로 코덱인 CS4218은 0x818000H로 맵핑하여 사용하였다. 그림 4.6에서는 외부장치인 PPI82C55와 CS4218의 어드레스 맵핑을 위하여 사용된 어드레스 디코더의 구성을 보여주고 있다. PPI82C55에서는 포트 A는 LCD의 데이터 버스로, 포트 B는 5비트로 구성된 음성인식결과를 이동로봇으로 출력하는 포트, 포트 C는 LCD의 제어버스로 사용된다. 그 외의 자세한 시스템의 각 부분별 실제 회로도 및 실제 구현된 보드의 사진은 부록에 첨가하였다.

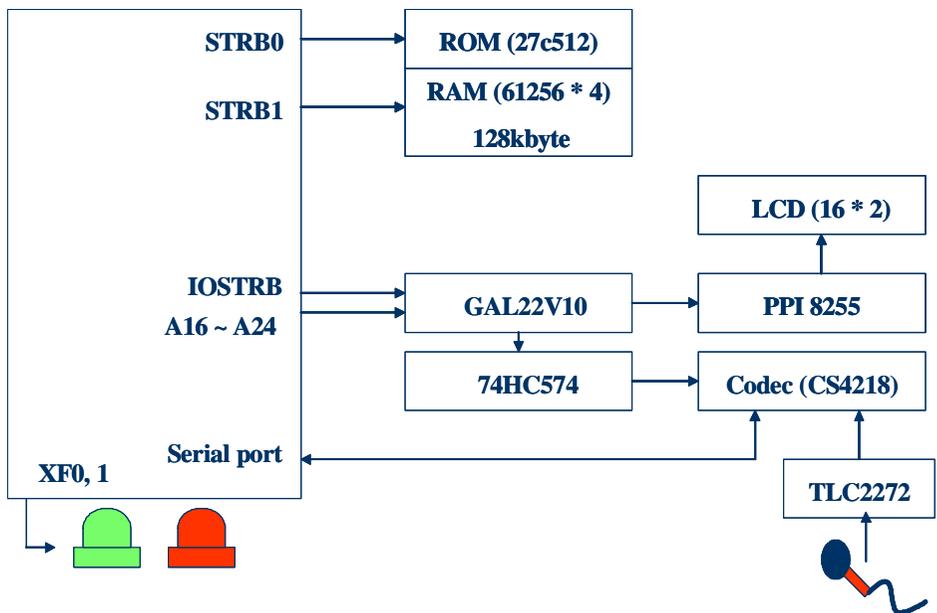


그림 4.4 메인 시스템 보드의 구성

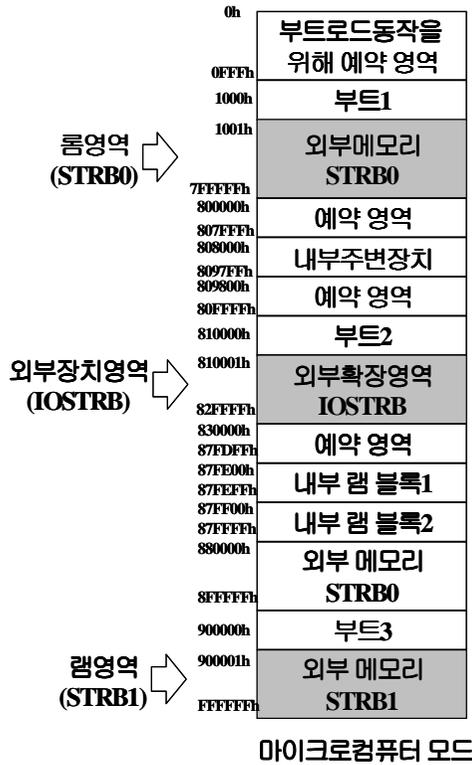


그림 4.5 메인 시스템의 메모리 맵

어드레스 디코더의 구성 IOSTRB의 영역 분할

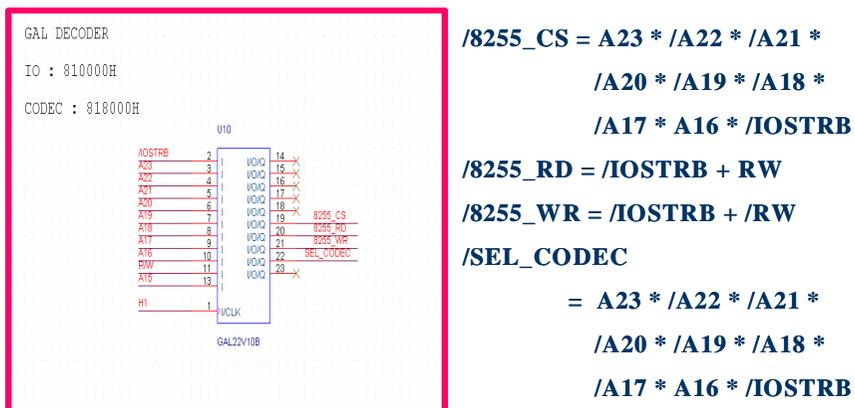


그림 4.6 어드레스 디코더의 구성

4.3 음성인식 시스템의 구현

4.3.1 음성 DB

다기능 전동휠체어에서 작동이 가능한 명령어로 “앞으로”, “뒤로”, “왼쪽으로”, “오른쪽으로”, “정지”, “빨리”, “천천히”, “좌로 회전”, “우로 회전”의 9개 단어를 사용하였다.

인식 시스템의 학습과정을 위해 음성 수집용 소프트웨어 ‘Cooledit2000’을 이용하여 PC에서 사운드카드를 음성데이터를 수집하였다. 그리고 음성데이터 수집 시 샘플링 주파수는 8KHz, 분해능은 16비트로 설정하였다. 음성 DB는 총 45명(남성 30명, 여성 15명)이 3번 발성하여 단어 당 135개의 음성파일로 구성되고 휠체어에 적용한 명령어는 9개이므로, 총 단어 수인 1215개의 음성데이터로 구성된다. 수집할 때 단어의 끝점 검출과정을 거치게 된다. 여기서 총 45명이 9개 단어를 3번 발성하여 수집한 음성파일을 15명(남성 10명, 여성 5명)씩 세 부류로 나누어 첫 번째는 HMM 학습을 위한 데이터로, 두 번째는 방사 기저함수 신경회로망을 학습하는 데이터로, 마지막 세 번째는 성능평가를 위한 테스트용으로 사용하였다.

4.3.2 음성구간 감지 단계

음성구간 감지는 구현된 음성인식 모듈에서 실시간으로 들어오는 음성을 유성음 및 무성음을 구별하는 단계로써, 만약 유성음이 감지되면 그 유성음 각 프레임의 데이터 값을 절대에너지 방식으로 계산하여 설정된 크기보다 크면 음성으로 간주하여 이 구간을 실제 음성으로 사용한다.

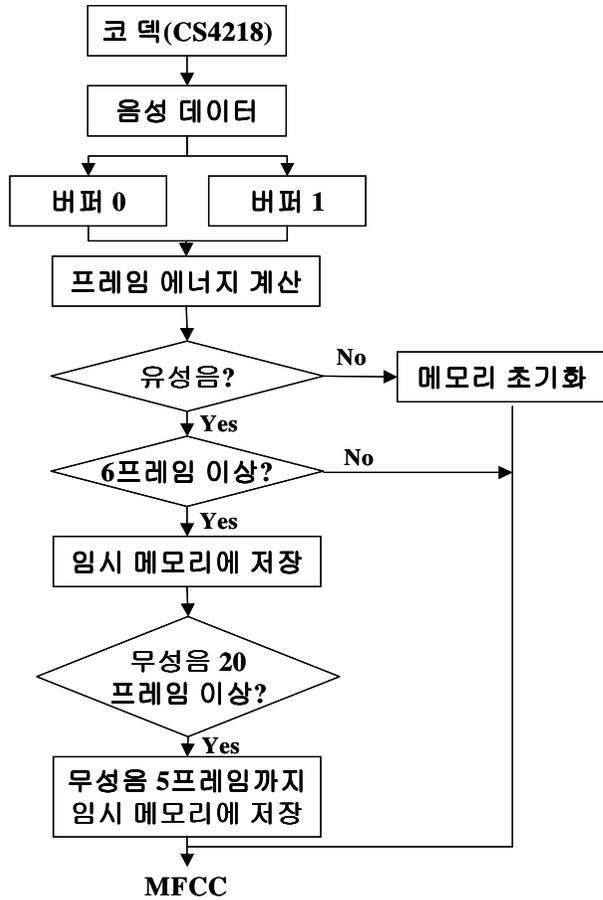


그림 4.7 음성구간 감지에 대한 끝점 검출

다시 말해 음성의 처리단위는 단구간 분석을 하는데, DMA(Direct Memory Access) 채널로 입력된 음성을 임시 저장하기 위해 버퍼 2개를 할당하고, 첫 번째 버퍼로 음성을 입력하는 중에 다른 버퍼에서는 에너지 계산을 하도록 하여 실시간으로 음성의 구간을 검출하도록 구성하였다.

본 실험에서는 사용된 프레임의 크기는 256샘플로 설정하였으며, 프레임 중첩 간격은 80샘플로 설정하였다. 여기에서 음성인지 아닌지를 판단

하는 절대 에너지 방식의 임계값은 실험 데이터를 통하여 획득된 값에 의해서 에너지 스펙트럼 값을 50000으로 설정하여 음성음 무성음을 판단하였다.

그림 4.7은 음성구간 감지에 대한 끝점 검출의 플로우 차트를 나타낸 것으로 음성 프레임이 6프레임 이상이 되면, 그 부분은 음성으로 간주하여 메모리에 저장되고, 이후 무성음 구간이 20 프레임 이상 나타나면 끝점에 존재하는 무성음 구간을 고려하여 5 프레임만을 저장시키고 나머지 15프레임은 삭제하였다.

4.3.3 전처리 단계 및 HMM 훈련 및 인식 단계

음성신호는 특징추출 단계인 MFCC 과정을 거치게 되면, 1 프레임 당 12 계수(order) 값으로 줄어들게 된다. 이 후 VQ/HMM 훈련 알고리즘에 의해서 명령별로 학습을 시킨다. 여기서 코드북 크기는 512로 이진 트리 알고리즘으로 학습 및 심볼을 추출하여 생성시켰다.

명령어별로 벡터 양자화된 심볼값들은 DHMM의 학습 알고리즘인 Baum-Welch 재 추정으로 명령어별 λ 값을 재 추정한다. 즉 명령어가 9개 이므로 λ 가 9개가 생성된다. 이렇게 생성된 명령어별의 λ 를 가지고 음성 DB로 사용하게 된다.

또한 인식과정은 테스트할 단어가 입력되면 특징추출 단계이후, 훈련된 각 단어별 λ 와의 확률계산을 통해서 가장 확률 값이 큰 것을 인식한다. 학습과정과 인식과정에서 사용된 음성데이터는 총 45명이 9개의 월체어 명령어 단어를 3번씩 발성한 데이터 중에서 15명씩 세 부류로 나누어 첫 번째 부류와 마지막 부류를 HMM 학습과 테스트용 인식과정에 사용하였다. 그리고 HMM 인식 및 훈련 단계를 그림 4.8에서 나타내었다.

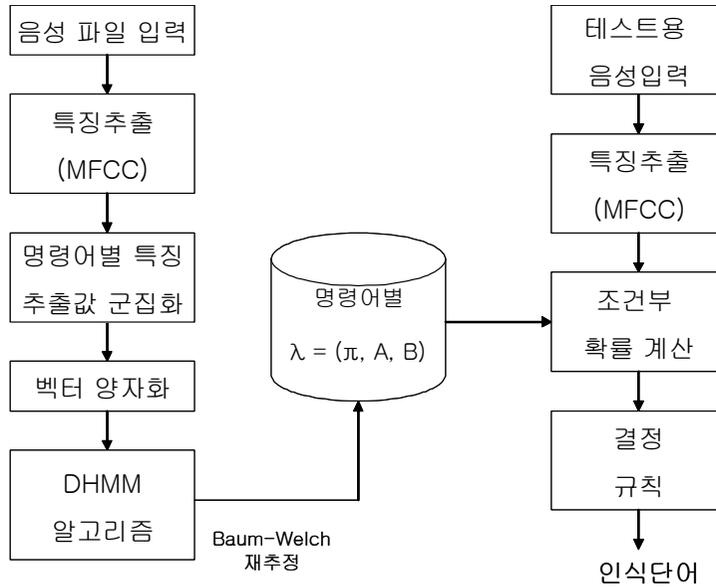


그림 4.8 HMM 인식 및 훈련 단계

4.3.4 방사 기저함수 신경회로망의 학습 및 최종 인식 단계

4.1절에서 소개된 바와 같이 HMM의 오인식을 보상하기 위하여 사용된 방사 기저함수 신경회로망은 일반 MLP(MultiLayer Perceptron)에 비하여 구조가 간단하고 계산량이 적어 실시간으로 처리하는 시스템에 더욱 효과적이라 할 수 있다. 따라서 HMM의 인식단계의 후처리 과정에 학습된 방사 기저함수 신경회로망을 사용하여 인식률을 향상시킬 수 있다.

먼저 수집된 45명의 음성데이터 중에서 15명의 음성데이터를 사용하여 4.3.3절의 HMM의 모델 파라미터 λ 를 학습시킨 후, HMM 학습에 참여하지 않은 15명의 음성데이터를 사용하여 학습된 HMM λ 학습 테이블에 의해서 계산된 각 음성명령의 출력확률을 구하고, 구해진 출력확률 벡터

열을 신경망의 입력패턴으로, 각 단어를 분류하기 위한 9개의 출력 유닛 조합은 출력패턴으로 설정하여 방사 기저함수 신경회로망을 학습시킨다.

설계된 방사 기저함수 신경회로망은 3층구조를 갖는다. 입력층은 출력 확률벡터를 나타내는 9개의 뉴런으로 구성되어 있고, 은닉층은 가우시안 함수를 활성화 함수로 갖는 20개의 뉴런으로 구성되었으며, 출력층은 선형함수를 활성화 함수로 갖는 9개의 뉴런으로 구성되었고, 출력층엔 바이어스를 추가시켜 설계하였다. 설계된 방사 기저함수 신경회로망의 구조는 그림 4.9에 나타내었다.

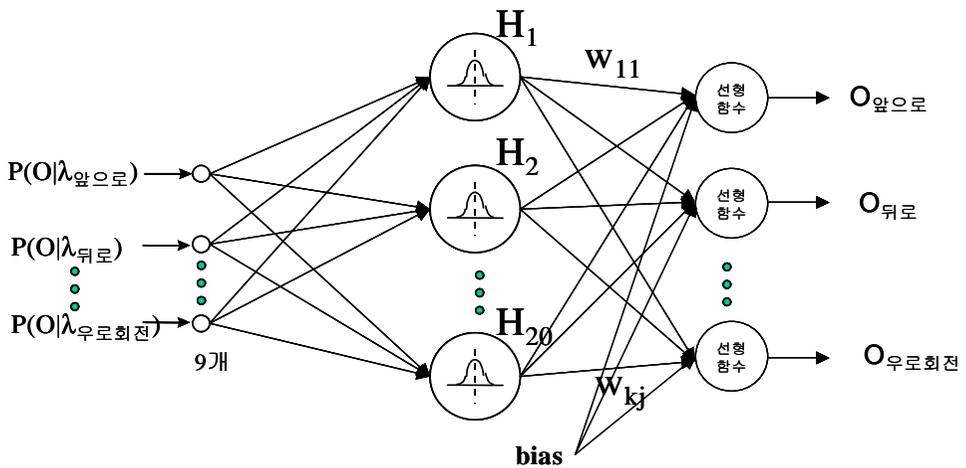


그림 4.9 설계된 방사 기저함수 신경회로망의 구조

주어진 방사 기저함수 신경회로망의 파라미터는 다음과 같다.

- ① 학습 상수 : 0.8(m), 0.3(σ), 0.75(선형함수)
- ② 모멘텀 항 : 0.2
- ③ 입력노드 수 : 9
- ④ 은닉노드 수: 20

- ⑤ 출력노드 수 : 3
- ⑥ 바이어스 : 출력층에서 사용
- ⑦ 활성화 함수 : 가우시안 함수(은닉노드),
선형 함수(출력노드)

제 5장 제안된 인공지능형 제어 시스템의 설계 및 구현

본 논문에서 적용한 이동로봇인 휠체어의 인공지능형 제어 시스템은 장애인들의 보다 안전하고 편안한 휠체어 조작을 위해 지능 제어 알고리즘인 퍼지 추론에 의한 퍼지 제어기를 설계하여 적용하였다. 또한 일반 조이스틱 입력 외에도 3장에서 구현된 음성인식 모듈에 의한 음성입력에 의해서도 휠체어를 조작할 수 있는 지능형 제어 시스템이 소개된다.

본 장에선 지능형 제어 시스템이 적용된 이동로봇인 전동휠체어의 구성, 지능형 제어 시스템의 하드웨어 구성, 그리고 설계된 지능형 제어 시스템에 대하여 논의한다.

5.1 인공지능형 제어 시스템의 하드웨어 구성

본 논문에서 제안한 이동로봇의 인공지능형 제어 시스템의 하드웨어 구성은 그림 5.1과 같다.

인공지능형 이동로봇 시스템은 일반 휠체어의 조향장치로 많이 사용되는 조이스틱과 3장에서 구현된 음성인식모듈을 사용한 음성입력의 두 가지 입력모드로 동작되도록 설계되었으며. 전원부 및 모터구동부, 조이스틱 및 음성입력부, 제어부, 센서부로 나눌 수 있다.

5.1.1 전원부

전원부에는 DC 24V로 구동하는 DC 모터 전압과 전체 시스템을 제어

하는 제어부 전원인 5V전압, 엔코더 및 음성인식 모듈로 공급되는 12V로 구성되어진다. 각 시스템의 안정성을 위해서 모터에 공급되는 전원과 시스템 제어에 사용되는 전원을 각각 분리하여 공급하였다.

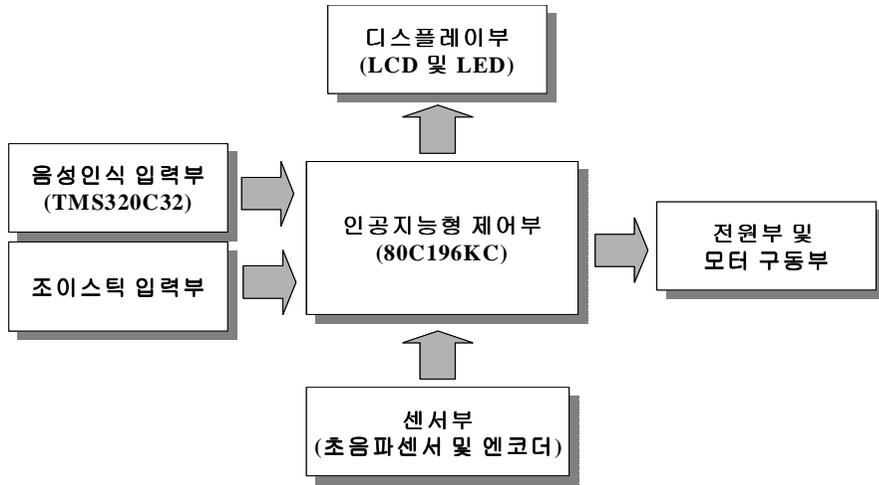


그림 5.1 인공지능형 제어 시스템 전체 구성도

실제 전동휠체어에 사용되는 전원으로써 배터리를 이용하였는데 배터리는 12V로 최대 40A로 사용 할 수 있으며, 두 개의 배터리를 직렬로 연결하여 24V로 만들었으며, 각각의 12V 배터리를 보호하기 위하여 직렬로 연결되는 부분에 배터리 보호회로를 구성하여서 과부하가 걸릴 경우에 배터리 보호회로가 동작하여 배터리를 보호하게 만들었다.

전원부 구성소자 중 중요한 소자인 전압을 다운시키는 리니어 레귤레이터를 이용함에 있어 열이 많이 발생하므로 방열판을 부착하여 열의 방출을 도우며 입출력 단자 사이에 플라이휠 다이오드를 부착하여 역 전류가 흐르는 것을 방지하는 회로를 채택하였다. 그림 5.2는 전원부의 전원 시스템 계통도이다.

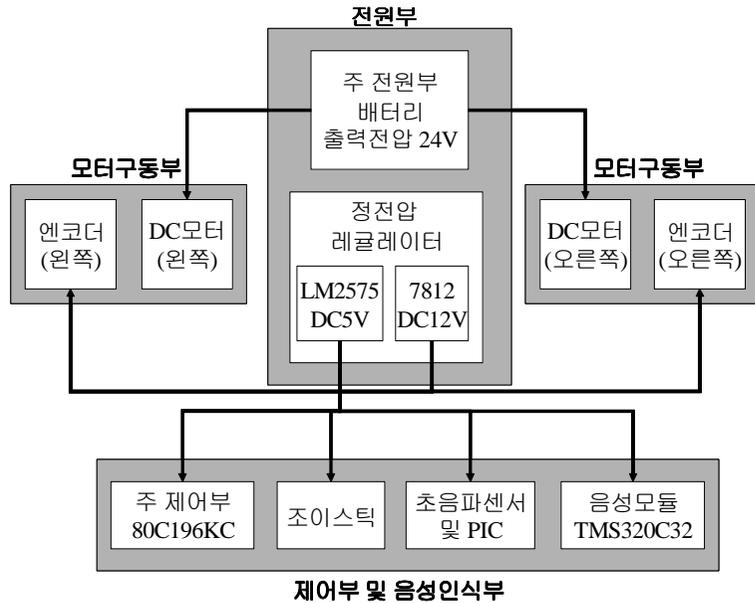


그림 5.2 전원 시스템 계통도

그리고 모터 뒤에 장착된 전자 브레이크는 모터에 혹은 제어기에 전원 공급이 차단 될 경우에 전동휠체어의 안전을 위하여서, 전동휠체어의 왼쪽과 오른쪽 모터가 동작되지 않게 하고, 다시금 24V 전압을 전자 브레이크에 공급을 하여야 모터가 정상적으로 동작하게 하였다.

5.1.2 모터 구동부

휠체어 구동에 사용된 모터인 DC 서보모터를 제어하고, 구동하기 위한 모터 구동회로로는 모터의 빠른 정역변환과 정확한 속도제어가 가능한 H-Bridge 회로로 구성하였으며, H-Bridge 회로에 사용된 스위칭 소자로는 전압 구동소자이며, 빠른 스위칭 속도와 매우 작은 온(on) 저항을 갖고, 휠체어와 같은 대 전력용에서 많이 사용되는 POWER MOSFET인

IRFP064를 사용하였다. 부하시 모터 구동에는 전류량이 증가하여 FET에 열이 발생하므로 각 FET에는 방열판을 부착하여 열적 안정성을 도모하였다. 회로의 안전성을 확보하기 위하여 실제 제어부의 마이크로프로세서인 80C196KC와 포토커플러에 의하여 모터 드라이브단과 전기적으로 절연하여 제어부의 회로를 보호하였다. 모터의 정역변환 및 속도제어는 80C196KC의 HSO 0~3핀에 PWM(Pulse Width Modulation)신호를 인가하여 구현하였다. 그림 5.3에서 모터제어를 위한 하드웨어 설계를 나타내었다.

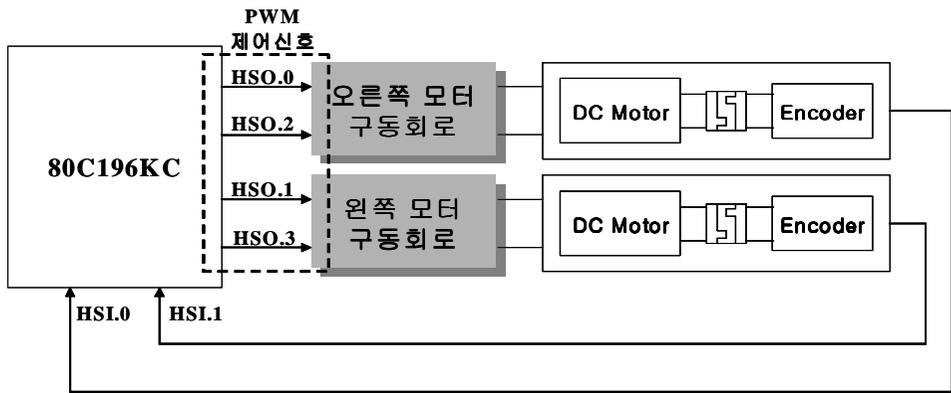


그림 5.3 모터 제어를 위한 하드웨어 구조

5.1.3 조이스틱 및 음성입력부

본 논문에서는 조이스틱 입력과 음성인식모듈에 의해서 인식된 음성 입력에 의한 휠체어의 제어가 주된 연구목적 중의 하나이다. 따라서 사용자는 조이스틱과 음성 중에 하나를 선택하여 휠체어를 동작시킬 수 있게 된다.

먼저 일반적으로 많이 사용되어지는 조이스틱의 물리적인 구조와 동작

을 살펴보면 조이스틱의 좌우에 2개의 포텐시오미터(potentiometer)를 부착하여 조이스틱을 상하 혹은 좌우로 움직였을 경우 포텐시오미터의 저항값이 변화되어져서 조이스틱에 공급된 5V 전압이 저항의 변화된 양만큼 아날로그 전압으로 출력하게 된다. 따라서 두 아날로그 전압이 80C196KC의 AD채널인 포트0의 1,2번으로 입력되고 입력된 값들은 10비트 AD로 변환되어 휠체어의 제어 입력으로 사용된다. 그림 5.4는 조이스틱의 구조 및 메커니즘을 나타내고 있다.

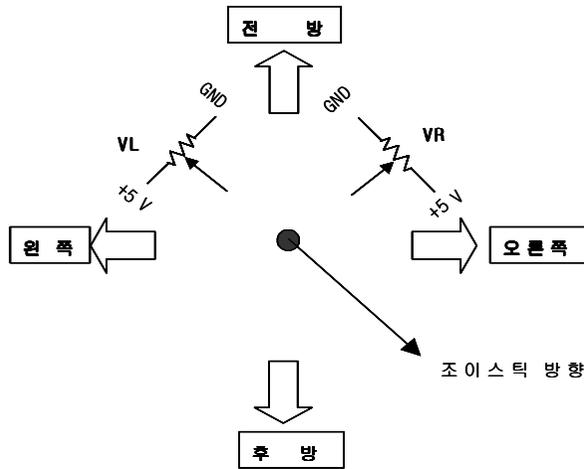


그림 5.4 조이스틱 구조 및 메커니즘

그림 5.5는 실제 휠체어가 동작 할 경우 모터의 방향을 제어하기 위하여 조이스틱을 움직였을 경우 휠체어의 모터 회전방향 및 속도를 선정한 것을 나타내고 있다. 그림에서 사각형은 조이스틱의 위치에 따른 왼쪽과 오른쪽의 모터의 방향을 표시하며, 화살표의 길이는 왼쪽, 오른쪽 모터의 속도 크기를 화살표 길이로 표시한 것이다. 그리고 조이스틱 방향의 위치에 따라서 +0도에서 +180도까지는 1사분면과 4사분면으로 일정 사이의 값으로 생각하고, -0도에서 -180도까지 역시 2사분면과 3사분면 사이의

표 5.2 조이스틱 위치에 따른 전압 값과 모터방향

조이스틱 위치	모터(왼쪽)		모터(오른쪽)	
	전 압	회전방향	전 압	회전방향
전 진	0.00~2.00	시계방향	0.00~2.00	시계방향
전진-오른쪽	0.00~2.00	시계방향	2.02~2.98	정 지
오른쪽	0.00~2.00	시계방향	3.00~5.00	반시계방향
오른쪽-후진	2.02~2.98	정 지	3.00~5.00	반시계방향
후 진	3.00~5.00	반시계방향	3.00~5.00	반시계방향
후진-왼쪽	3.00~5.00	반시계방향	2.02~2.98	정 지
왼 쪽	3.00~5.00	반시계방향	0.00~2.00	시계방향
왼쪽-전진	2.02~2.98	정 지	0.00~2.00	시계방향

또한 음성인식모듈에 의해 인식된 음성입력은 80C196KC의 포트1의 0~4핀에 5비트 데이터가 제어부에 입력된다. 표 5.3은 음성인식 모듈에서의 음성 단어별 데이터 값을 표로 나타내었다. 특히 최상위 비트(2^4)는 음성의 입력 여부를 알 수 있는 기준이 된다. 음성입력 중 “천천히”와 “빨리”명령어는 입력 횟수를 카운팅하여 속도를 가변할 수 있는 기능을 추가하였고, “왼쪽으로”와 “오른쪽으로”는 입력 횟수에 따라서 방향을 가변할 수 있다. 따라서 마이크를 통해서 음성이 입력되면 바로 최상위 비트가 반전되어 음성입력 여부를 판단할 수 있도록 음성모듈에서 구성하였다. 주 제어부인 80C196KC는 음성입력의 최상위 비트(2^4)가 입력되는 포트1.4를 확인하고 이전 상태값과 비교하여 상태값이 반전될 경우 사용자 음성이 입력되었음을 알 수 있다. 따라서 음성입력 횟수를 카운팅할 수 있게 되어 속도 가변이나 방향 가변 시에 사용하게 된다.

표 5.3 음성 단어별 출력 데이터

명령어	2^4	2^3	2^2	2^1	2^0
앞으로	음성 입력 여부 판단	0	0	0	1
뒤로		0	0	1	0
왼쪽으로		0	0	1	1
오른쪽으로		0	1	0	0
정지		0	1	0	1
천천히		0	1	1	0
빨 리		0	1	1	1
좌로회전		1	0	0	0
우로회전		1	0	0	1
거부반응		0	0	0	0

5.1.4 제어부

본 논문에서는 인텔사의 16비트 마이크로프로세서인 80C196KC를 사용한 임베디드 시스템으로 제어부를 구성하였다. 휠체어의 모든 동작은 제어부인 80C196KC에서 처리된다. 먼저 제어부에서 조이스틱 입력은 80C196KC의 AD채널인 포트0의 1, 2번으로 입력되고 입력된 값들은 10비트 AD로 변환되어 휠체어의 제어입력으로 사용된다. 또한 음성인식 모듈에 의해 인식된 음성입력은 포트1의 0~4핀에 5비트 데이터로 입력된다. 이와 같이 입력된 조이스틱 입력과 음성으로 인식된 음성입력은 각각 두 개의 외부 인터럽트 EXINT(포트0.7), EXINT1(포트2.2)에 의해서 사용자가 선택할 수 있도록 하였다. 즉 EXINT가 발생하면 조이스틱 모드로 전

환되고, EXTINT1이 발생하면 음성인식모드로 전환되어 휠체어가 동작된다.

휠체어 제어량이 입력되면 내부 제어 프로그램에 의해서 HSO(High Speed Output) 0~3포트를 통해서 PWM 제어신호를 발생시키면서 HSO 0, 2는 왼쪽 모터 구동부의 제어 출력 신호로 나가게 되고, HSO 1, 3은 오른쪽 모터 구동부의 제어 출력신호로 각각 출력하게 된다. 이 때 HSO 포트 선택에 의하여 모터의 정역방향의 구동을 선택할 수 있고, 출력되는 HSO의 PWM신호 주파수에 의해서 모터의 속도를 제어할 수 있게 된다. 그리고 모터가 회전 할 때마다 모터 뒤에 부착된 엔코더를 통하여 제어기의 HSI(High Speed Input)로 모터의 회전수를 보내준다. 여기서 HSI.0은 왼쪽 모터의 RPM값의 정보를 HSI.1은 오른쪽 모터의 RPM값을 각각 입력받아 각 모터의 보상량을 계산하는데 사용한다.

5.1.5 센서부

센서부로는 크게 모터의 속도를 측정하는 엔코더와 모터의 휠체어 시스템의 안정성을 위해 후진시와 장애물 근접시 경보음을 내주는 초음파 센서모듈 부분으로 나눌 수 있다.

먼저 모터 측정을 위한 엔코더로는 오토닉사의 축형 인크리멘탈 로타리 엔코더 E40S-1024-3-2를 사용하였으며, 표 5.4에 사용한 엔코더의 정격을 나타내었다.

엔코더의 출력을 80C196KC에서 받아들이기 위해 고속 입력(HSI : High Speed Input) 0번과 1번을 통하여 각각 왼쪽 모터와 오른쪽모터의 속도를 측정한다. 측정시에는 모터의 고속 동작시 HSI의 타이머가 입력사건을 놓쳐버리는 경우가 발생하므로 HSI_Mode를 8분주 카운터를 사용하

여 8개의 상승에지가 입력될 때마다 사건을 인식하는 모드를 사용하여 고속 동작시에도 엔코더의 펄스를 충분히 인식할 수 있도록 설계하였다. 그리고 기준입력 값이 RPM값으로 주어지므로 엔코더 펄스의 수를 측정하여 RPM으로 변환하여 비교하였다.

표 5.4 엔코더의 정격 및 사양

모델명	E40S-1024-3-2
출력형태	NPN 오픈 콜렉터 출력
전원전압	12VDC
1회전당 펄스 수	1024
출력상	A, B, Z 상
샤프트 외경	8mm

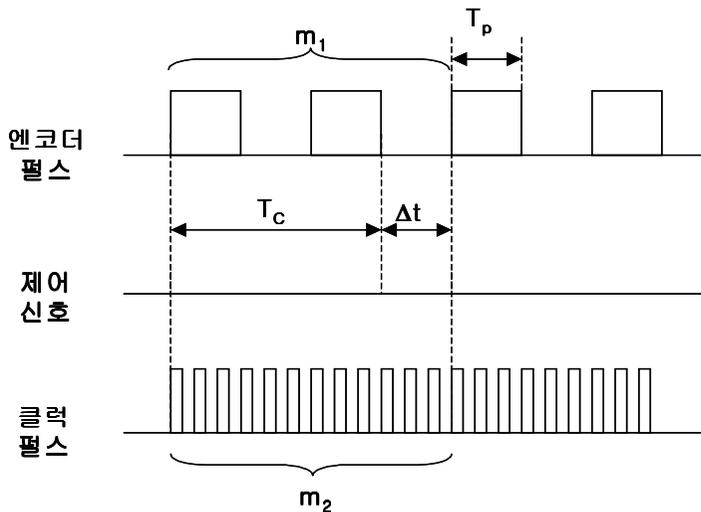


그림 5.6 엔코더를 이용한 모터의 회전속도 측정법

$$\omega_r = \frac{60 \times m1}{ppr \times T_s} [rpm] \quad (5.1)$$

식 (5.1)은 일정한 측정 주기 T_s 동안 엔코더로부터 출력된 펄스 수 $m1$ 을 HSI로 계수하고 이로부터 회전수를 계산하는 방법이다. ppr은 사용한 엔코더의 분해능으로 1024이다.

또한 초음파 센서는 PWM 방식의 초음파 모듈을 사용하였으며, 신호가 감지되면 위험 신호(‘뺨’)를 발생하게 제작하였고, 구성 프로세서는 PIC16F84를 사용하였다. 이렇게 구성된 모듈은 장애물의 후방의 장애물의 위치가 80cm이하로 근접하게 되면 경보음을 내줌으로써 후방 운전 시 보이지 않는 장애물로부터 안전하게 할 수 있다.

5.2 인공지능형 제어 시스템의 설계

대부분의 휠체어 제어기로 기능이 단순히 조이스틱 입력에 따른 휠의 회전비율에 따라 결정되기 때문에 정확한 휠체어의 동작 제어가 이루어지지 않아 장애인들의 이동시 안전에 많은 문제점이 있다. 또한 본 논문에서는 음성인식기능을 추가함으로써 조이스틱 조작에 의한 동작보다 음성입력을 통한 휠체어의 동작 시는 훨씬 많은 위험요소가 있기 때문에 더욱 높은 정밀도의 휠체어 동작제어가 필요하게 된다. 그림 5.7은 설계된 지능형 퍼지 제어기의 구조를 나타낸다.

먼저 조이스틱 동작시에는 전압값(V_x, V_y)이 입력되면 입력변환표에 의해서 퍼지 제어기의 입력으로 사용될 휠체어의 동작 속도(V_w)와 휠체어의 동작방향(θ_w)으로 변경 된다. 또한 음성입력의 경우도 마찬가지로 음성입력이 휠체어의 동작속도(V_w)와 휠체어의 동작방향(θ_w)으로 변경 된다.

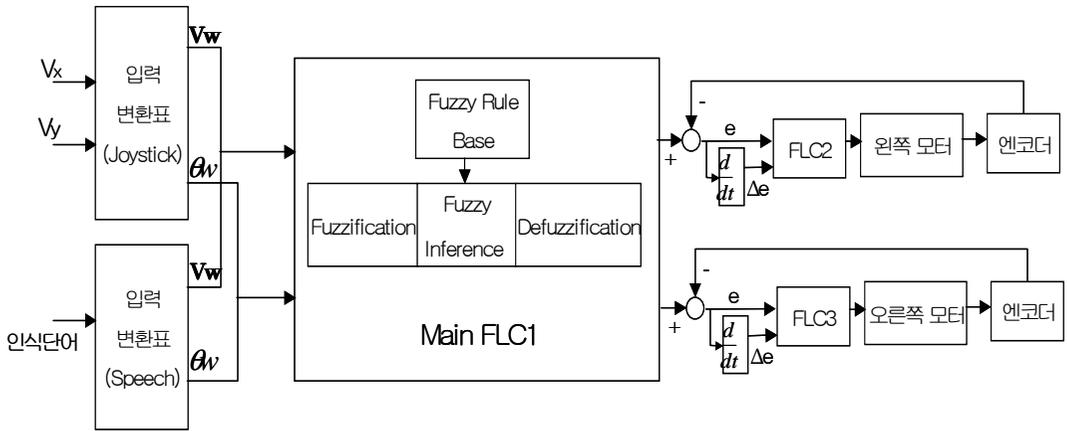


그림 5.7 휠 제어 시스템의 퍼지 제어기 구조

입력된 V_w , θ_w 는 주 퍼지 제어기 퍼지화 과정을 거쳐 추론과정을 실행하고 비퍼지화를 거쳐서 크리스프(crisp)한 값으로 출력을 보내게 된다. 여기서 주 퍼지 제어기는 두 개의 입력에 두 개의 출력을 내주는 다변수 퍼지 제어기의 구조를 갖추고 있다. 주 퍼지 제어기에서의 출력은 전동휠 제어의 왼쪽과 오른쪽 모터 구동부의 구동 입력으로 PWM 출력 신호의 듀티비를 나타낸다.

본 논문에서 주 퍼지 제어기의 구성을 살펴보면, 먼저 주 퍼지 제어기로 들어가는 퍼지 입력변수의 선정은 본 시스템의 환경에 적합한 요소들로 구성되었다. 실제 주 퍼지 제어기에 입력되는 요소는 휠 제어의 동작을 결정짓는 동작속도 V_w 와 동작방향 θ_w 로 설정하였다. 그리고 퍼지 입력 변수의 입력범위는 표 5.4과 같다.

표 5.4 휠제어 동작속도(V_w) 및 동작방향(Θ_w)의 범위

변 수	범 위
V_w (속도)	0 ~ 9 Km
Θ_w (각도)	1,4 사분면 (+0°~+180°)
	2,3 사분면 (-0°~-180°)

입력변수 중 전동휠제어의 속도는 사용자의 안전성을 보장하기 위해서 최고 9Km까지로 정하여 그 범위 안에서 제어 동작하도록 하였다.

각도는 0°~180°각도로 표시하여 나타내었으며, 퍼지 출력변수는 엔코더의 모터 측정속도와 비교를 위한 두 모터의 기준속도 값이다. 출력 변수는 왼쪽과 오른쪽 모터의 기준 속도로 각각 2개로 구성하였다. 그림 5.8은 동작속도(V_w)와 동작방향(Θ_w)인 입력변수의 소속함수와 왼쪽모터와 오른쪽모터 속도인 출력변수의 소속함수를 나타내었다.

또한 추론을 위한 퍼지 규칙을 표 5.5와 표 5.6에 나타내었다. 주 퍼지 제어기는 두 개의 입력과 두 개의 출력을 갖는 다변수 시스템이므로 왼쪽 모터와 오른쪽 모터 각각 24개의 퍼지 규칙을 작성하였다. 이와 같이 구성된 퍼지규칙에 의해서 “Mamdani”의 Max-Min 연산을 사용하는 추론 방법을 이용하였고, 비퍼지화에서는 무게중심법을 사용하여 각 모터의 기준 속도값을 아날로그출력 값으로 얻을 수 있었다.

그림 5.9와 그림 5.10은 퍼지 제어기의 추론과정과 퍼지 제어기의 제어평면을 Matlab을 이용하여 시뮬레이션한 결과이다. Matlab를 이용한 시뮬레이션에 의해서 시스템 제어에 대한 타당성을 검토한 후 실제 구성된 시스템에 적용하였다.

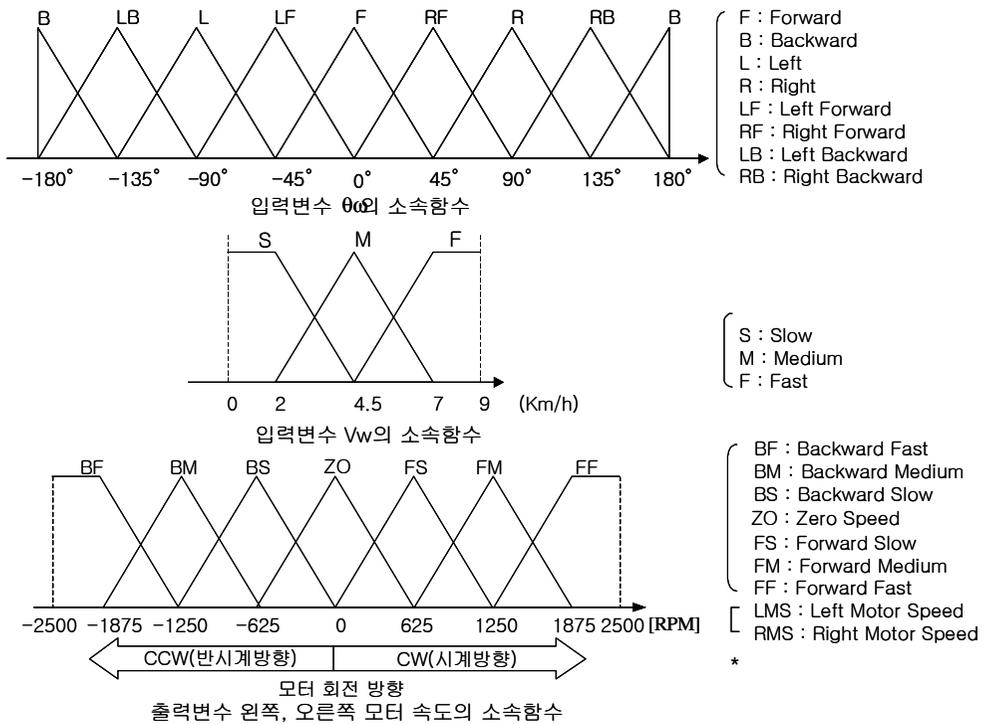


그림 5.8 휠체어 주 펌퍼 제어기의 입·출력 소속함수

표 5.5. 왼쪽 모터의 속도와 방향 추출을 위한 펌퍼 추론 규칙

θ V	B	LB	L	LF	F	RF	R	RB	B
S	BS	BS	BS	ZO	FS	FS	FS	ZO	BS
M	BM	BM	BM	ZO	FM	FM	FM	ZO	BM
F	BF	BF	BF	ZO	FF	FF	FF	ZO	BF

표 5.6 오른쪽 모터의 속도와 방향 추출을 위한 퍼지 추론 규칙

θ V	B	LB	L	LF	F	RF	R	RB	B
S	BS	ZO	FS	FS	FS	ZO	BS	BS	BS
M	BM	ZO	FM	FM	FM	ZO	BM	BM	BM
F	BF	ZO	FF	FF	FF	ZO	BF	BF	BF

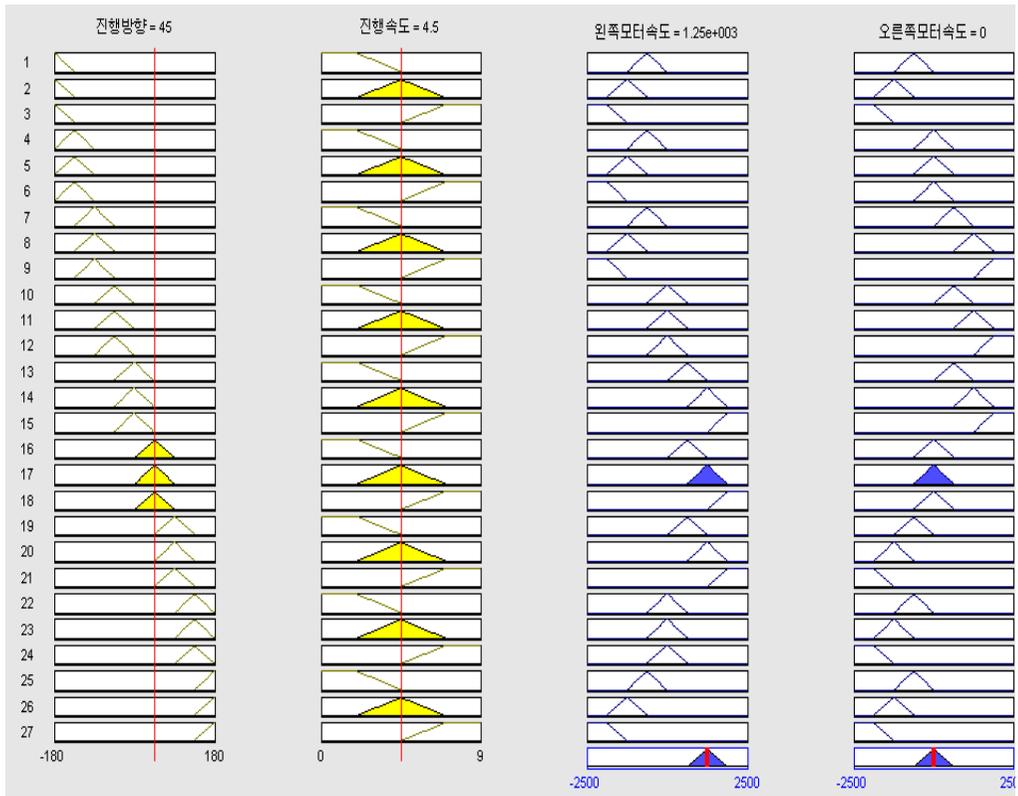


그림 5.9 설계된 퍼지 제어기의 추론과정

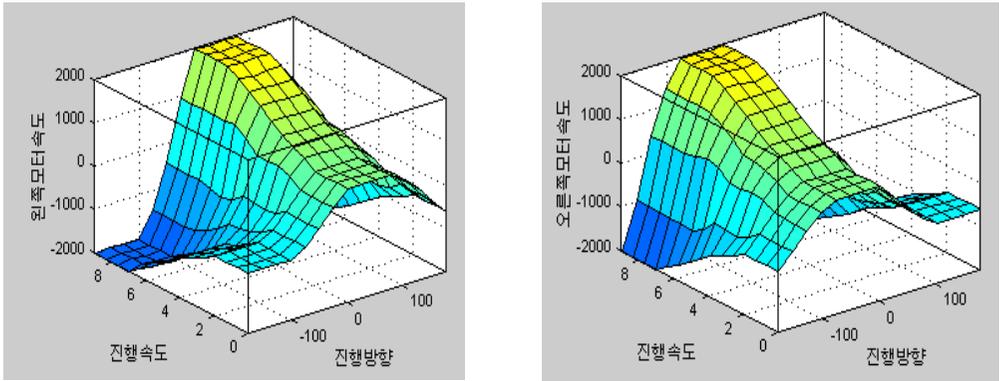


그림 5.10 퍼지 제어기의 제어 평면

부 제어기로 사용한 FLC(Fuzzy Logic Controller)2와 FLC3는 좌우 모터의 엔코더(encoder)값에 의한 피드백 제어를 통해 실제 모터구동부에 입력될 PWM 신호의 듀티비를 출력하도록 설계되었다. FLC2와 FLC3의 입력변수는 엔코더 출력과 기준입력으로부터 계산되는 현재 오차와 오차의 변화량을 입력 변수로 선정하였다. 오차의 계산은 주 제어기인 FLC1에서 출력되는 모터속도의 기준 입력과 엔코더로부터 측정된 모터속도를 비교하여 현재의 오차를 구할 수 있다.

그림 5.11은 현재 오차와 오차의 변화량에 대한 입력 소속함수 및 제어량인 출력 소속함수를 나타낸 것이다. 그리고 FLC2와 FLC3의 추론에서 사용된 추론규칙 13개를 표 5.7에 나타내었다. 최대 가능한 수의 제어규칙은 49개가 될 수 있으나 시스템의 특성상 그리고 계산량을 줄이기 위해서 선정된 13개의 제어규칙을 제안하였다.

이와 같이 주 퍼지 제어기는 전향보상형식(feedforward)의 제어기로 다변수 제어 시스템의 정확한 제어 출력을 얻을 수 있었고, 주행 도로의 상태와 사용자의 몸무게 등과 같은 여러 주행 환경에 의해 발생할 수 있는

모터의 속도오차를 피드백형식의 두 개의 퍼지 제어기로 오차를 보상하여 정확한 제어가 될 수 있도록 인공지능형 제어 시스템을 구성 하였다.

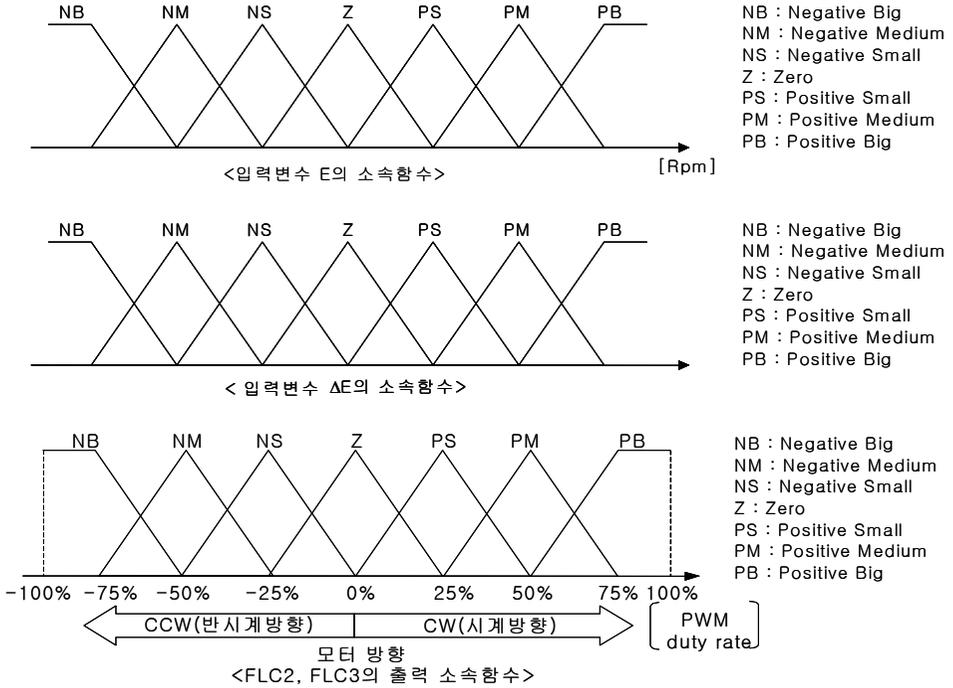


그림 5.11 부 제어기 FLC2와 FLC3의 입·출력 소속함수

표 5.7 FLC2와 FLC3의 퍼지 추론 규칙

E \ ΔE	NB	NM	NS	ZO	PS	PM	PB
NB				NB	NM		
NM				NM			
NS				NS	ZO		PM
ZO	NB	NM	NS	ZO	PS	PM	PB
PS	NM		ZO	PS			
PM				PM			
PB			PM	PB			

제 6 장 실험 및 고찰

본 논문에서 제안된 음성인식 기법을 이용한 인공지능형 이동로봇인 휠체어 시스템의 성능평가를 위해서, 본 장에서는 실시간 음성인식을 위한 음성인식 기법 및 모듈에 관한 실험과 조이스틱과 음성입력에 따른 인공지능형 이동로봇 제어기의 실험 및 고찰로 구성하였다.

6.1 실시간 음성인식 시스템

제안된 하이브리드형 음성인식 알고리즘과 DSP인 TMS320C32를 사용하여 설계된 임베디드형 음성인식 모듈의 성능을 평가하기 위해 크게 세 가지 실험 과정을 거쳐 제안된 시스템의 성능을 평가 할 수 있었다. 먼저 잡음환경에 강한 음성인식의 전처리 과정의 알고리즘 선정을 위한 실험을 수행하고, 두 번째로는 음성인식 알고리즘으로 사용된 HMM과 하이브리드형 HMM-NN의 성능 비교를 위한 실험과 마지막으로 실제음성인식 모듈에서의 인식을 평가를 수행하였다.

6.1.1 음성인식 전처리 과정의 알고리즘의 선정

본 실험에서는 여러 가지 잡음환경에서 음성인식을 수행하기 위하여 음성인식 시스템의 전반에 걸쳐서 소음에 대한 대책을 강구하였다. 본 절에서는 이러한 방법들의 유효성을 평가하기 위하여 각각의 방법을 사용하는 음성인식 시스템의 성능을 비교하였다. 소음환경에서 음성인식을 수행하기 위하여 본 실험에서 사용하고 있는 방법을 정리하면 다음과 같다.

- (1) 역 필터링과 고역 통과 필터 사용한 음성 구간 검출
- (2) 스펙트럼 차감법을 사용하여 잡음 제거
- (3) 위너 필터링을 사용한 잡음 제거
- (4) 끝점 검출을 사용한 음성 구간 검출

위와 같은 방법들의 유효성을 평가하고 가장 우수한 성능을 나타내는 방법을 선정하기 위하여 표 6.1과 같은 5가지 형태의 음성인식 시스템을 구현하였다.

표 6.1 여러 가지 형태의 음성인식 시스템과 파라미터

방법 \ 형태	Case1	Case2	Case3	Case4	Case5
역필터링과 고역 통과 필터		☑	☑	☑	☑
스펙트럼 차감법		☑		☑	
위너 필터링			☑		☑
끝점 검출을 사용한 음성 구간 검출	☑			☑	☑

Case 1은 필터링을 사용하지 않고 끝점 검출만 사용한 방법이고, Case 2는 끝점 검출을 하지 않는 방법으로 필터와 스펙트럼 차감법을 사용하였다. Case 3은 Case 2에서 스펙트럼 차감법 대신에 위너 필터링을 사용한 방법이며, Case 4는 필터와 스펙트럼 차감법과 끝점 검출을 사용한 방법이다. 마지막으로 Case 5는 Case 4에서 스펙트럼 차감법 대신에 위너 필터링을 사용하였다.

표 6.2 음성인식 시스템의 형태에 따른 인식율

대상\형태	Case 1	Case 2	Case 3	Case 4	Case 5
테스트 A	93.00	83.75	85.00	93.00	94.00
테스트 B	94.75	85.50	87.50	93.50	94.50
테스트 C	93.25	84.00	86.00	93.75	95.25
평균	93.67	84.42	86.17	93.42	94.58

표 6.2는 5가지 경우에 대한 인식실험 결과를 보여준다. 인식실험은 각각의 방법에 의해서 전처리된 데이터를 사용하여 특징 추출후 HMM으로 학습한 음성인식기를 구성하여 학습데이터 이외의 3사람의 음성(테스트A, B, C)에 대하여만 테스트하였다. 표 6.2에서 인식률이 가장 좋은 결과는 Case 5와 Case 4로 유사하게 인식 결과가 나왔다. Case 2와 Case 3에서는 끝점 검출을 하지 않을 경우에는 현저히 인식률이 저조하게 나왔다. 따라서 본 논문에서 제안한 음성인식 시스템에서는 Case 5와 같은 전처리 과정을 거쳐 특징추출을 하였다.

6.1.2 하이브리드형 HMM-NN 음성인식 시스템의 성능 평가

본 실험에서는 화자 독립형 시스템을 구성하기 위해 고립 단어용 DHMM 인식 알고리즘을 이용하여 9개의 단어에 대해 인식률을 테스트하였다. 또한 후처리 과정으로 방사 기저함수 신경회로망의 학습 후 하이브리드형 음성인식기를 구성하여 인식률을 HMM 인식기와 비교하여 성능평가를 하였다. 본 실험은 먼저 PC상에서 학습 및 테스트과정을 거친 후, 실제 완성된 시스템은 4.2절에서 DSP TMS320C32를 사용하여 구현

한 음성인식 보드에서 마지막 테스트를 하였다. 그림 6.1, 그림 6.2, 그림 6.3은 한 사용자의 명령어 “앞으로”의 음성파형, MFCC 특징벡터, 벡터 양자화 결과를 보여준다.

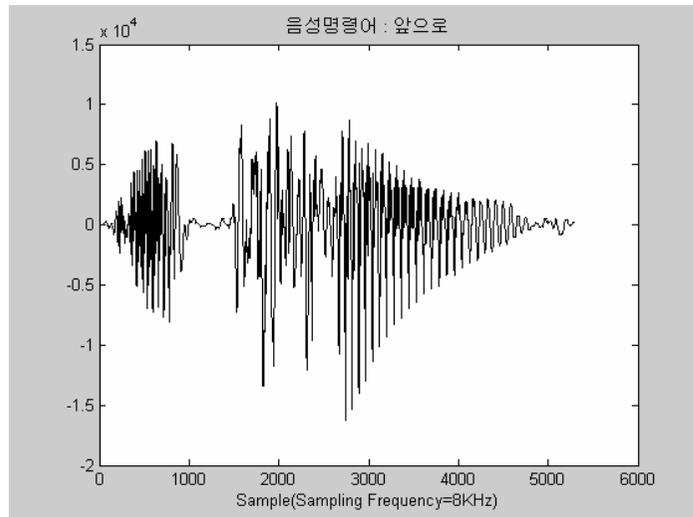


그림 6.1 명령어 “앞으로”의 음성파형

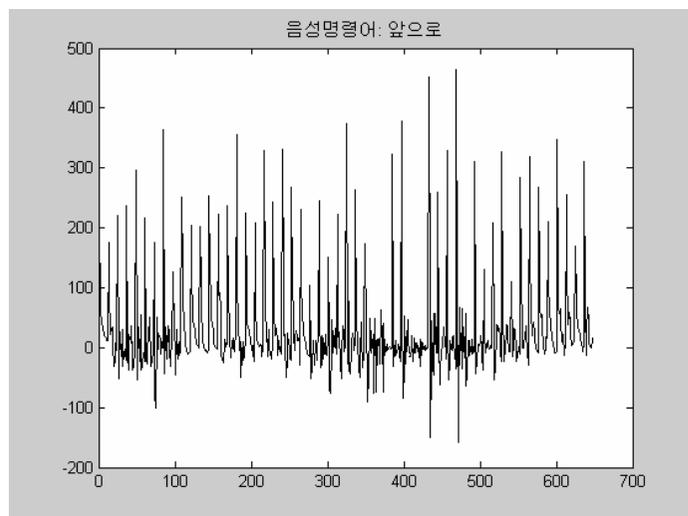


그림 6.2 명령어 “앞으로”의 MFCC 특징벡터

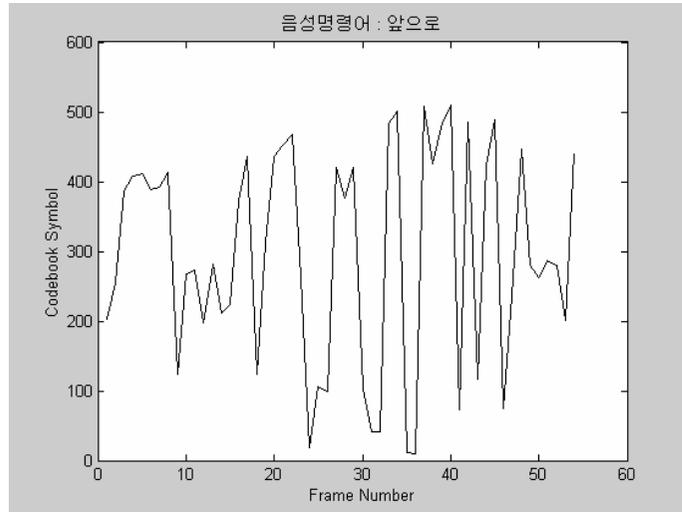


그림 6.3 명령어 “앞으로”의 벡터 양자화 결과

여기서 음성 DB의 구성은 총 45명이 9개 단어를 3번 발성하여 수집한 음성파일을 15명(남성 10명, 여성 5명)씩 세 부류로 나누어 첫 번째 음성 DB는 HMM 학습을 위한 데이터로(음성 DB①), 두 번째 음성 DB는 방사 기저함수 신경회로망을 학습하는 데이터로(음성 DB②), 마지막 세 번째 음성 DB는 성능평가를 위한 테스트용으로(음성 DB③) 사용하였다. 그리고 하이브리드형 HMM-NN과 비교를 위하여 실제 HMM인식기를 사용한 인식시스템에서는 학습 데이터를 음성 DB①과 음성 DB②를 합한 30명의 음성 DB를 사용하였다. 4.3.4절에서 제안한 방사 기저함수 신경회로망을 사용한 HMM 후처리 시에서는 음성 DB①을 HMM 학습에 사용하였고, 음성 DB②는 방사 기저함수 신경회로망의 학습에 마지막으로 음성 DB③ 테스트용으로 사용 실험을 하였다. 그림 6.4에는 제안된 하이브리드형 HMM-NN 음성인식 시스템의 학습 및 테스트 절차를 보여준다.

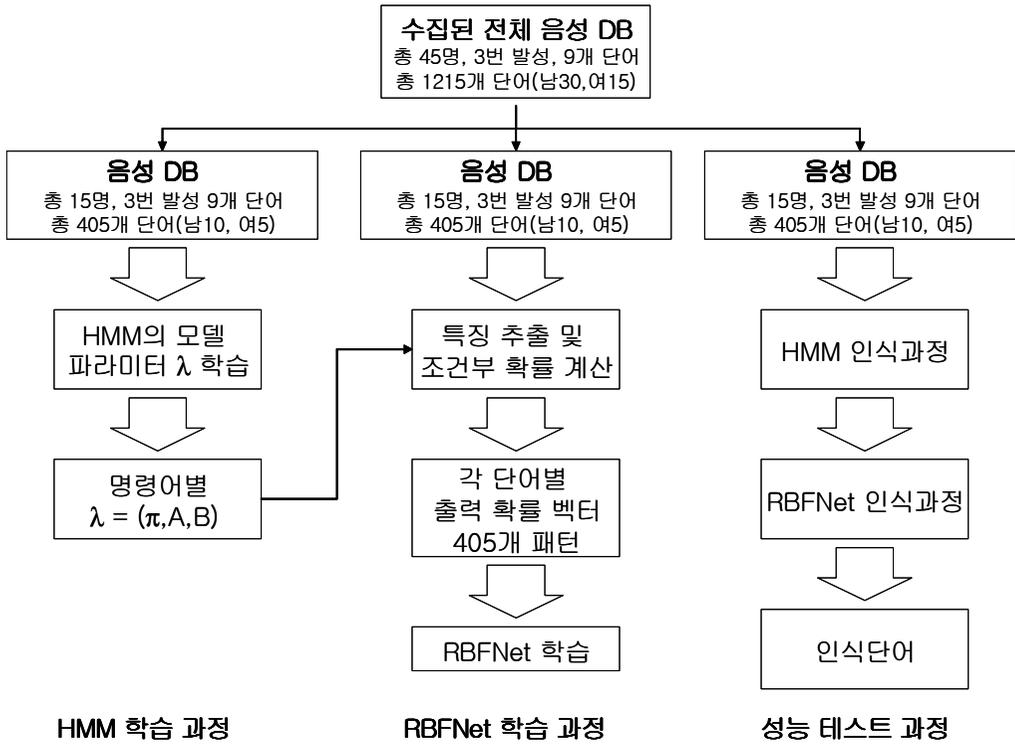


그림 6.4 하이브리드형 음성인식 시스템의 학습 및 테스트 절차

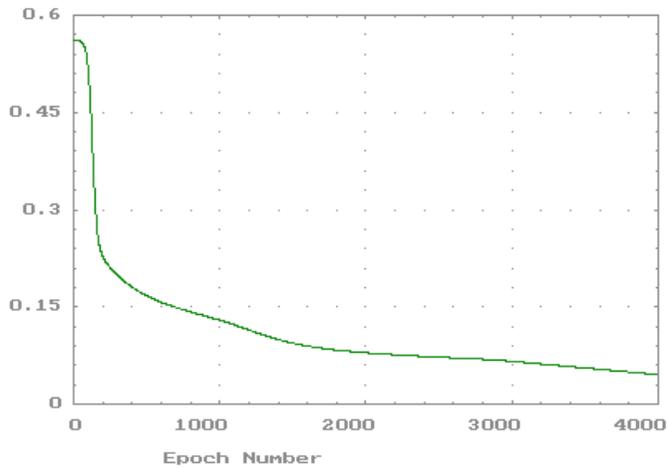


그림 6.5 방사 기저함수 신경회로망의 학습과정

그림 6.5에서는 방사 기저함수 신경회로망을 학습한 과정을 보여 주고 있다. 신경망의 학습은 오차가 1%가 될 때 까지 학습하였다.

표 6.3은 30명이 9개 단어를 3번씩 발성한 음성 DB를 사용한(음성 DB① + 음성 DB②) HMM 학습 후 학습에 사용하지 않은 15명이 9개 단어를 3번씩 발성한 음성 DB(음성 DB③)를 사용하여 테스트 한 결과이다.

실제 인식실험에서 적용되는 시스템의 특성상 방사 기저함수 신경회로망의 최종 인식단어의 결정은 각 출력뉴런의 최대값이 아닌 전체 출력 뉴런 총합의 80%를 넘는 출력 뉴런을 최종 인식단어로 결정하였다. 따라서 그 이외의 경우는 오인식으로 처리하여 시스템의 신뢰성이 보장되도록 테스트를 하였다.

표 6.4에서는 하이브리드형 HMM-NN 음성인식 시스템의 테스트 결과를 보여주고 있다. 실제 HMM만으로 학습 후 구성된 인식시스템에 비하여 약 1.7% 인식률이 향상되었음을 알 수 있었다. 이를 기반으로 하여 HMM의 학습데이터인 모델 파라미터 λ 와 신경회로망의 학습 파라미터인 가우시안 함수의 중심(m)과 폭을 나타내는 표준편차(σ), 출력층의 가중치를 사용하여 실시간 음성인식 보드를 구성할 때는 인식과정 만을 실행하도록 프로그래밍 하였다. 음성인식 보드에서의 인식실험은 소음 환경인 적은 실험실내에서 무작위로 10명이 각 단어 당 10번씩 발성하여 음성인식실험을 실행하였고, 전체 평균 인식률은 95.7%로 실제 PC의 실험결과보다 약간 저하된 결과를 얻을 수 있었다. 이는 실제 PC와 보드의 테스트 데이터가 같지 않고, 실시간 처리 시 잡음의 영향이 있는 것으로 예상할 수 있었다. 또한 제안한 하이브리드형 인식시스템에 의해 인식률이 향상됨을 알 수 있었고, 휠체어의 구동에도 적합할 정도의 인식률을 얻을 수 있었다.

표 6.3 HMM을 이용한 음성인식 시스템의 테스트 결과

명령어	테스트 횟수	인식 횟수	인식률
앞으로	45	45	100%
뒤로	45	44	97.8%
왼쪽으로	45	40	88.8%
오른쪽으로	45	41	91.1%
정지	45	45	100%
빨리	45	43	95.5%
천천히	45	44	97.8%
좌로회전	45	40	88.8%
우로회전	45	42	93.3%
합계	405	384	94.8%

표 6.4 하이브리드형 음성인식 시스템의 테스트 결과

명령어	테스트 횟수	인식 횟수	인식률
앞으로	45	45	100%
뒤로	45	45	100%
왼쪽으로	45	42	93.3%
오른쪽으로	45	42	93.3%
정지	45	45	100%
빨리	45	43	95.5%
천천히	45	44	97.8%
좌로회전	45	42	93.3%
우로회전	45	43	95.5%
합계	405	391	96.5%

6.2 인공지능형 제어 시스템

5.2절에서 제안한 인공지능형 제어 시스템의 성능평가를 위해서 두 가지 실험을 하였다. 먼저 FLC2과 FLC3의 성능평가를 위해 기준 입력을 1200rpm과 2000rpm으로 스텝변화 시켰을 때 퍼지 제어기의 응답특성에 대한 실험을 하였고, 조이스틱 입력이나 음성입력을 사용하여 실제 휠체어의 모터 속도의 선형성을 평가하는 실험을 하였다.

그림 6.6은 1200rpm의 기준입력을 주었을 때 모터의 속도 응답특성을 나타내고 있다. 프로그램에 의한 모터 속도 제어 주기가 5ms이다. 용량과 관성이 큰 DC 모터이지만 50ms 즉 10주기 만에 기준입력으로 수렴되어 지는 것을 확인할 수 있었다. 그림 6.6에서는 실제 전압파형을 출력으로 15V의 전압값으로 수렴되어 지며, 그 때의 모터속도 값은 엔코더를 통해 1200rpm이 측정됨을 확인할 수 있었다.

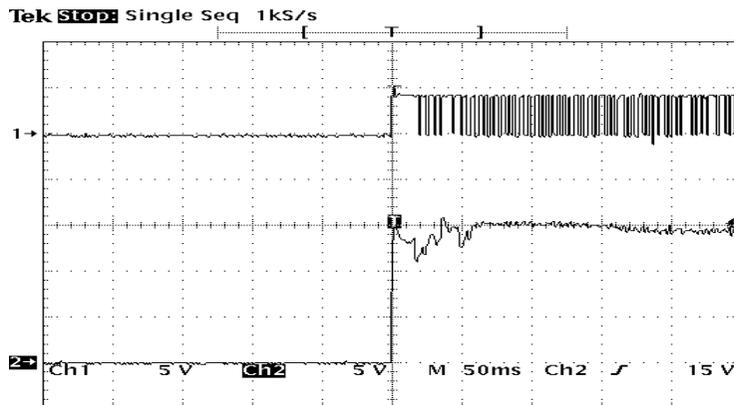


그림 6.6 1200rpm 기준 입력 시 속도 응답 파형과 PWM 출력

그림 6.7에서는 실제 PWM 파형의 듀티비를 측정된 것으로 주파수가 6.25KHz로 주기가 0.16ms이다.

따라서 PWM 듀티비는 $0.11\text{ms}/0.16\text{ms}=68.7\%$ 가 측정되었다.

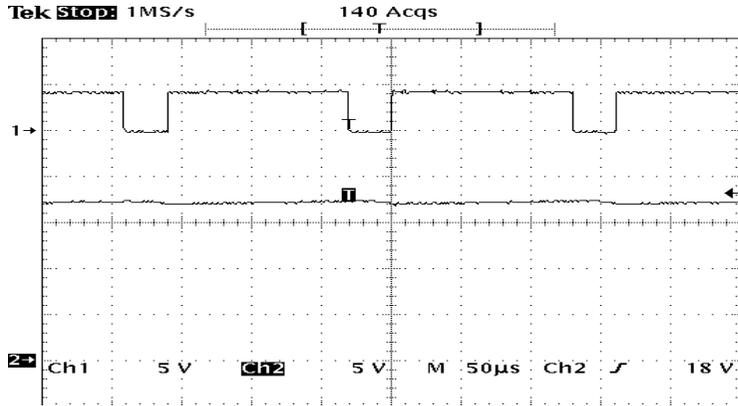


그림 6.7 1200rpm 기준 입력 시 PWM 파형의 듀티비

그림 6.8에서는 2000rpm 기준 입력시 모터의 속도 응답 특성을 나타내고, 그림 6.9는 PWM 파형의 듀티비를 나타내고 있다.

그림 6.8에서는 2000rpm 기준 입력을 사용하여 70ms에 수렴하는 것을 확인하였고, 그림 6.9에서는 PWM 파형의 듀티비가 $0.13\text{ms}/0.16\text{ms}=81.2\%$ 가 측정됨을 알 수 있다.

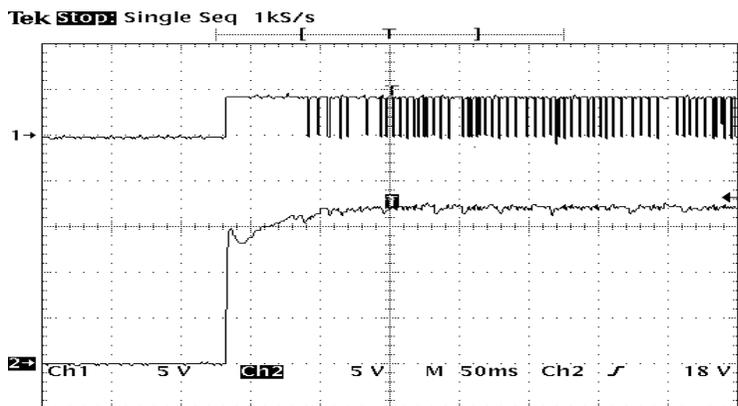


그림 6.8 2000rpm 기준 입력 시 속도 응답 파형과 PWM 출력

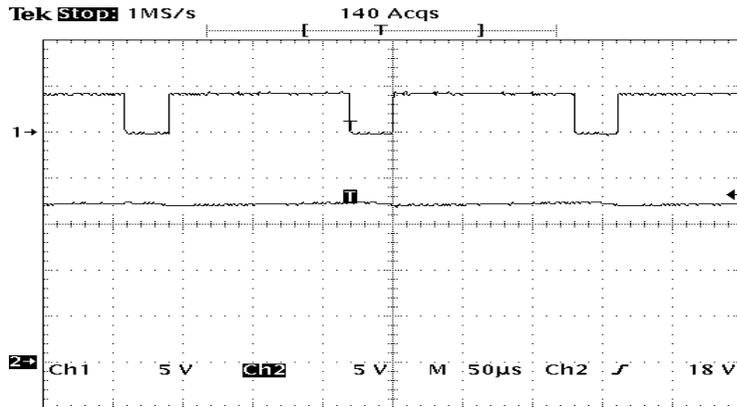


그림 6.9 2000rpm 기준 입력 시 PWM 파형의 듀티비

따라서 실제 설계된 FLC2과 FLC3가 모터의 속도제어 시 정확한 보상량을 내어주고 있음을 확인할 수 있고, 수렴속도 50ms~80ms로 일반 고전제어기 보다 성능이 우수하다는 것을 확인할 수 있었다^{[7][8]}.

이것은 피드백 제어기의 유무에 따라서 부하가 있을 경우와 부하가 없을 경우의 모터 응답의 선형성이 향상된 결과를 나타내고 있다.

따라서, FLC1에 의해서 휠체어 동작을 위한 정확한 두 모터 속도의 추정과 피드백 제어기로 사용된 FLC2, FLC3에 의한 오차 보정에 의해서 설계된 인공지능형 제어 시스템은 음성이나 조이스틱에 의한 정확한 제어 동작을 구현할 수 있도록 하여 안전성을 향상 시킬 수 있었다. 그리고 인공지능형 제어 시스템의 정확한 제어 동작에 의해서 제어 명령이 9개로 한정되어 있는 음성인식에 의한 휠체어 동작 방식에도 충분히 적용 가능하여 특히 장애정도가 심해 조이스틱 조작이 불가능한 장애인들도 휠체어를 사용한 이동이 가능할 것으로 예상된다.

제 7 장 결 론

본 논문에서 제시한 인공지능형 이동로봇 시스템은 조이스틱 및 음성모듈로부터 입력받은 사용자 정보에 의해서 이동로봇인 전동휠체어시스템의 동작을 최적으로 제어할 수 있는 방법을 제안하였다.

또한 실시간으로 이동로봇인 전동휠체어의 음성명령을 인식하기 위하여 DSP TMS320C32를 사용하여 임베디드 시스템을 구성하였고, 인공지능형 이동로봇 시스템의 주 제어부 역시 80C196KC를 사용하여 임베디드 시스템으로 구성하였다.

구성된 시스템에 의하여 음성인식 시스템의 인식실험과 실제 전동휠체어에 적용한 제어 성능 평가를 통해 다음과 같은 결과를 얻을 수 있었다.

1) 화자독립 형태의 고립단어 음성인식 시스템 구현을 위해 HMM을 사용하였고, HMM의 오인식을 보상하기 위한 방법으로 지능형 알고리즘인 신경회로망 중에서 처리속도가 빠르고 구조가 간단한 방사 기저함수 신경회로망을 사용하여 후처리하는 방법을 제안하였다. 제안된 방법에 의해 휠체어 음성 명령어에 대한 인식실험 결과 96.5%의 인식률을 얻을 수 있었고 이는 후처리 과정을 통해 약 1.7%의 인식률 향상을 보인 것이다.

2) 이동로봇의 제어부에서는 퍼지 추론을 사용한 퍼지 제어를 제안하여 이동로봇인 전동휠체어를 최적으로 제어 할 수 있도록 시스템을 구성하였다. 구성된 시스템의 성능 평가를 위해 각각의 모터 응답 특성을 테스트 한 결과 일반 제어기에 비해 수렴시간의 단축 및 정밀성에서 향상된 결과를 얻을 수 있었고 실제 조이스틱 및 음성 명령을 통한 휠체어 제어시 사용자의 입력에 따른 원활한 주행이 이루어짐을 확인할 수 있었다.

이동로봇 적용 대상인 전동휠체어시스템은 장애인들의 이동 수단이기

때문에, 음성 명령을 통한 휠체어의 주행 시에 안전에 중요한 요인인 인식률의 향상과 제안된 인공지능형 시스템의 휠체어 동작의 최적 제어에 의해서 안전성이 향상되어 실제 주행에도 충분히 적용될 수 있으리라 예상된다.

이로 인하여, 전동휠체어의 제어 입력을 조이스틱 이외에 음성으로 제어할 수 있어 장애 정도에 따른 사용범위 제한을 최소화 할 수 있을 것으로 예상된다. 또한 국내 업체들의 영세성 때문에 기술개발보다는 수입 의존도가 높은 전동휠체어 시장의 수입대체효과로 인하여 국내 휠체어제 작업계의 경쟁력 향상과 기술력증대 효과가 클 것이고 더 나아가 어려운 장애인들에게도 전동휠체어의 보급에 큰 기여를 할 것으로 사료된다.

앞으로의 연구과제로는 구현된 센서부에 의한 정밀 위치제어를 통해 감지된 장애물을 회피하면서 주행할 수 있는 알고리즘의 개발 및 보완과 나아가 인터넷 기반에 의해서 환자의 상태를 점검할 수 있는 시스템을 개발하고자 한다.

참 고 문 헌

- [1] Ulrich Borgolte, Helmut Hoyer, Christian Buhler, Helmut Heck, "Architectural Concepts of a Semi-autonomous Wheelchair", Journal of Intelligent and Robotic System, pp. 22:233-253, 1998.
- [2] S. Fioretti, T. Leo, S. Longhi, "A Navigation System for Increasing the Autonomy and the Security of Powered Wheelchairs", IEEE Transactions on rehabilitation engineering. Vol. 8, No. 4, pp. 490-498, December 2000.
- [3] Rory A. Cooper, Daniel K. Jones, Shirley Fitzgerald, Michael L. Boninger, Steven J. Albright, "Analysis of Position and Isometric Joysticks for Powered Wheelchair Driving", IEEE Transactions on Biomedical engineering. Vol. 47, No. 7, July 2000.
- [4] Karim A. Tahboub and Hary H. Asada, "A Semi-Autonomous Control Architecture Applied to Robotic Wheelchairs", Proceedings of the 1999 IEEE/RSJ International Conference on Intelligent Robots and Systems. pp. 906-911, 1999
- [5] Ren C. Luo, Tse Min Chen, Meng Hsien Lin, "Automatic Guided Intelligent Wheelchair System Using Hierarchial Grey-Fuzzy Motion Decision-Making Algorithms" Proceedings of the 1999 IEEE/RSJ International Conference on Intelligent Robots and Systems. pp. 900-905, 1999
- [6] F. Matia, R. Sanz and E. A. Puente, "Increasing Intelligence in Autonomous Wheelchairs", Journal of Intelligent and Robotic

- System, pp. 22:211-232, 1998.
- [7] 강성인, 강재명, 김정훈, 류홍석, 김관형, 이상배, “지능형 전동휠체어의 설계 및 구현”, 한국퍼지 및 지능시스템학회 2002년도 춘계학술대회, 12권, 1호, pp. 10-13.
- [8] 김정훈, 류홍석, 강재명, 강성인, 이상배, “HMM 음성인식 알고리즘을 이용한 효율적인 음성인식 모듈 개발 설계 관한 연구”, 한국퍼지 및 지능시스템학회 2002년도 추계학술대회, 12권, 2호, pp. 337-340.
- [9] 강성인, 김관형, 이주상, 이상배, 김일, “뉴로-퍼지 알고리즘을 이용한 자율 이동로봇의 충돌회피”, 한국퍼지 및 지능시스템학회 1999년도 추계학술대회, 9권, 2호, pp. 64-68.
- [10] Lefteri H. Tsoukalas and Robert E. Uhrig. “Fuzzy and Neural Approaches in Engineering”, John Wiley & Sons Inc., 1992.
- [11] Mohammad Jamshidi, Nader Vadiiee, Timothy J. Ross, Fuzzy Logic and Control, 1993.
- [12] Mohammad Jamshidi, Lotfi Zadeh, Andre Titli, Serge Boverie, Application of Fuzzy Logic, Prentice Hall, 1997.
- [13] C.W. de Silva, “ Intelligent Control-fuzzy logic applications”, CRC press, 1995.
- [14] Lin, Lee, “Neural Fuzzy systems-A Neuro-Fuzzy Synergism to Intelligent Systems”, Prentice Hall, 1996.
- [15] J.-S. R Jang, C.-T. Sun, E. Mizutani, “Neuro-Fuzzy and Soft Computing”, Prentice Hall, 1997.
- [16] Chin-Teng Lin and C.S. George Lee, “Neural fuzzy systems”, Prentice Hall, 1996.
- [17] Sigeru Omtsu, Marzuki Khalid and Rubiyah Yusof, “Neuro-Control

- and its Applications”, 1995.
- [18] Jacek M. Zurada, “Introduction to Artificial Neural System”, West Publishing Company, 1992.
- [19] Simon Haykin, “Neural Networks”, Macmillan Company, 1994
- [20] M. Minsky and S. Paper, “Perceptrons”, MIT Press, 1969.
- [21] Yan-Qing Yang and Abraham Kandel, “Compensatory Neuro-fuzzy Systems with Fast Learning Algorithm”, IEEE Trans. on Neural Networks, Vol. 9, No. 1, pp. 83-105, 1998
- [22] Y. Chen and F. Bastani, “ANN with Two Dendrite Neurons and Weight Initialization” , Proc. IJCNN, Baltimore, Vol.III, pp. 139-146, 1992.
- [23] Jae Hyun Lee, Sung Joo Kim, Sang Bae Lee, “The Study on the Intelligent Control of Robot System using Neural Network”, Proceedings of Asian Control Conference, Vol. III, No. III, pp. 67-70, 1997
- [24] K. S. Narendra and K. Parthasarathy, “A diagrammatic representation of back propagation” , Center for Syst. Sci., Dept. of Electrical Eng., Yale University, New Haven, CT, tech. rep. pp. 78-85, Aug. 1988.
- [25] P. G. Gallman and K. S. Narendra, “Identification of nonlinear systems using a Uryson model”, Becton Center, Yale University, New Haven, CT, tech. rep., Apr. 1976.
- [26] Alexander Waibel, Toshiyuki Hanazawa, Geoffrey Hinmton, Kiyohiro Shikano, Kevin J. Lang, “Phone Recognition Using Time-Delay Neural Networks”, IEEE Transactions on Acoustics,

- Speech, and Signal Processing, Vol. 37, No. 3, pp.328-339, March 1989.
- [27] C. Koch, T. Poggio, "Multiplying with Synapses and Neurons" , in Single Neuron Computation, T. Mckenna, J.Davis, and S. F. Zonnetzer[eds.], pp. 3165-3455, 1992.
- [28] A. G. Barto, R. S. Sutton, and C. W. Anderson, "Neuronlike adaptive elements that can solve difficult learning control problems", IEEE Trans. Syst. Man. Cybern, Vol. SMC-13, No.5 pp. 834-847, 1983.
- [29] Hopfield, J.J. and D.W.Tank, "Neural computation of decision in Optimization problems", Biological Cybernetics Vol.52. pp. 141-155 , 1985
- [30] Hopfield, J.J, "Neural networks and physical systems with emergent computational abilities", Proc. of National Academy of Secinces", Vol.79, pp. 2554-2558, 1982
- [31] S. Chen et al, "Non-linear System Identification Using Neural Networks", Int.J.control, vol 6, pp, 1191-1214, 1990
- [32] L.R. Rabiner and R.W. Schafer, Digital processing of speech signal, Prentice-Hall, Inc., 1978
- [33] Lawrence Rabiner and Biing Hwang Juang, "Fundamentals of Speech Recognition", Prentice Hall, 1993
- [34] Claudio Becchetti, "Speech Recognition Theory and C++ Implementation ", John Wiley & Sons Ltd, 1999.
- [35] Steven L.Gay, Jacob Benesty, "Acoustic Signal Processing For Telecommunication", Kluwer Academic Publishers, 2000.

- [36] Tomas F. Quatieri, "Discrete-Time Speech Signal Processing", Prentice Hall, 2002.
- [37] Emmanuel C. Ifeachor, Barrie W. Jervis, "Digital Signal Processing A Practical Approach", Addison-Wesley, 1993.
- [38] Alan V. Oppenheim, Ronald W.Schafer, John R. Buck, "Discrete-Time Signal Processing", Prentice Hall, 1998.
- [39] Abhijit S. Pandya, Robert B. Macy, "Pattern Recognition with Neural Networks in C++", CRC Press, 1995.
- [40] Julius T. Tou, Rafael C. Gonzalez, "Pattern Recognition Principles", Addison-Wesley Publishing Company, 1974.
- [41] J.G. Wilpon and L.R. Rabiner, "A Modified K-Means clustering Algorithm for Uae in Isolated Wrod Recognition", IEEE Trans, on ASSP, Vol33, No3,1985
- [42] John Makhoul, Salim Roucos and Herrert Gish, "Vector Quantization in Speech Coding", Proc. IEEE, Vol. 73, No. 11, pp 1551-1588, november 1985.
- [43] Myers C.S. and Rabiner L.R., "A comparative study of several Dynamic Time-Warping Algorithms for Connected Word Recognition", Bell system Tech. J., 60(7) pp.1389-1409, September 1981
- [44] J. Makhoul, S. Roucos, and H. Gish, "Vector Quantization in Speech Coding", Proc. of IEEE, Vol. 73, No.11, pp.1551-1558, 1985
- [45] Y. Linde, A. Buzo, and R.M. Gray, "An Algorithm for Vector Quantization Design", IEEE Trans. on Communication, Vol. 28, No.1, pp84-95, 1980

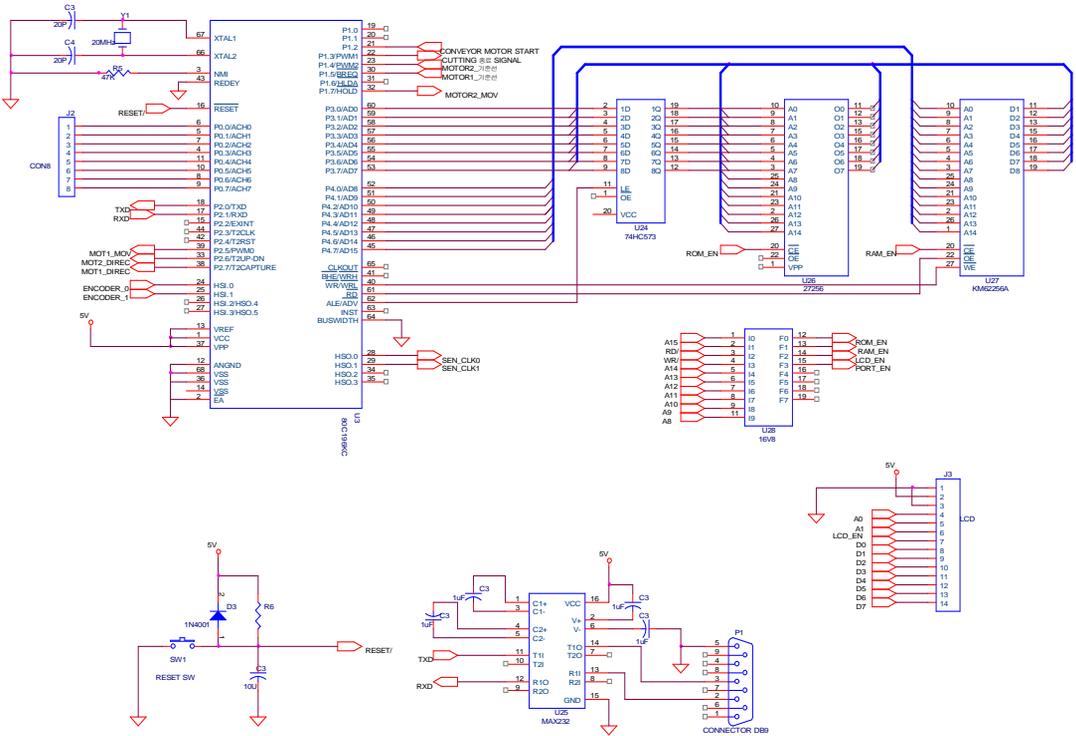
- [46] Lawrence Rabiner, "A Tutorial on Hidden Markov Models and Selected Application in Speech Recognition", Proc. IEEE, Vol 77, No. 2, february 1989.
- [47] L.R. Rabiner, S.E. Levinson and M.M.Sondhi, "On the Application of Vector Quantization and Hidden Markov Models to Speaker Independent Isolated Word Recognition", Bell System Technical Journal, Vol. 62, No.4, april 1983.
- [48] L.R. Rabiner, B.H. Juang, S.E.Levinson, and M.M. Sondhi, "Recognition of Isolated Digits using Hidden Markov Models with Continuous Mixture Densities", AT&T Technical Journal, Vol. 64, No.6, pp. 1211-1234, 1985
- [49] X.D. Huang and M.A. Jack, "Semi-continuous Hidden Markov Models for speech Signal", Computer Speech and Language, Vol. 3 No. 3, pp. 239-251, 1989
- [50] Christoph Neukirchen, Jorg Rottland, Daniel Willett, Gerhard Rigoll, "A Continuous Density Interpretation of Discrete HMM Systems and MMI-Neural Networks", IEEE Transactions on Speech and Audio Processing. Vol. 9, No. 4, pp. 367-377, May 2001
- [51] Yuan-Fu Liao, Sin-Horng Chenm "A Modular RNN-Based Method for Continuous Mandarin Speech Recognition", IEEE Transactions on Speech and Audio Processing, Vol. 9, No. 3, pp. 252-263, March 2001.
- [52] Dario Albesano, Roberto Gemello, Franco Mana, "Hybrid HMM-NN modeling of stationary transitional units for continuous

speech recognition", Information Science 123, pp 3-11, 2000.

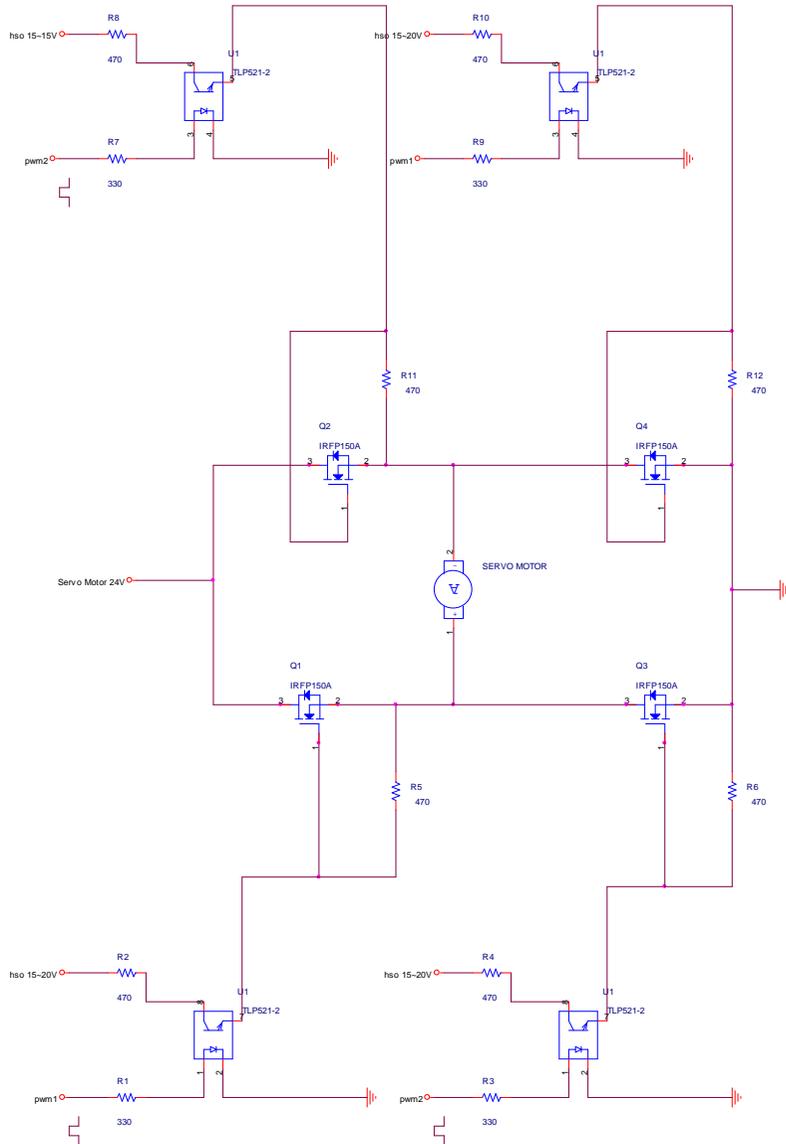
- [53] Ingrid Kirschning, Jun-Ichi, "Neural networks and th Time-Sliced Paradigm for Speech Recognition", IEEE Trans. Ine. & Syst., Vol. E79 D, No. 12, pp. 1690-1692, December 1996.

부록 A. 회로도

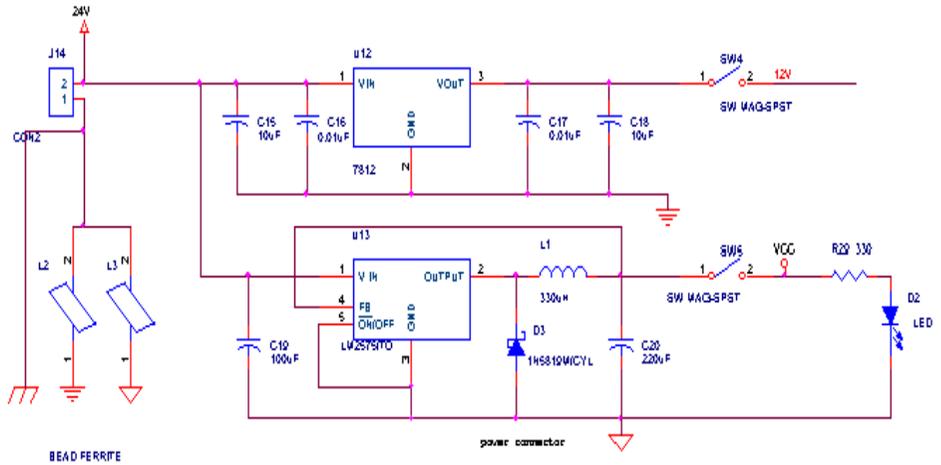
부록 A.1 80C196KC Main System



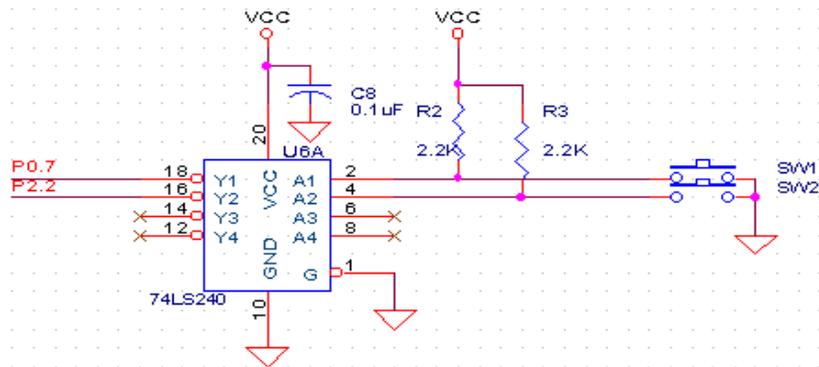
부록A.2 모터 구동부



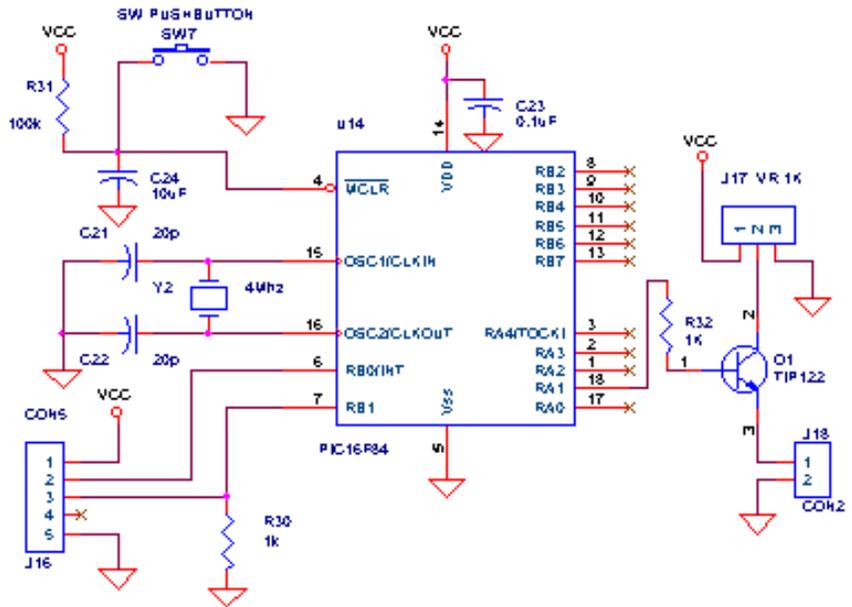
부록A.3 전원부



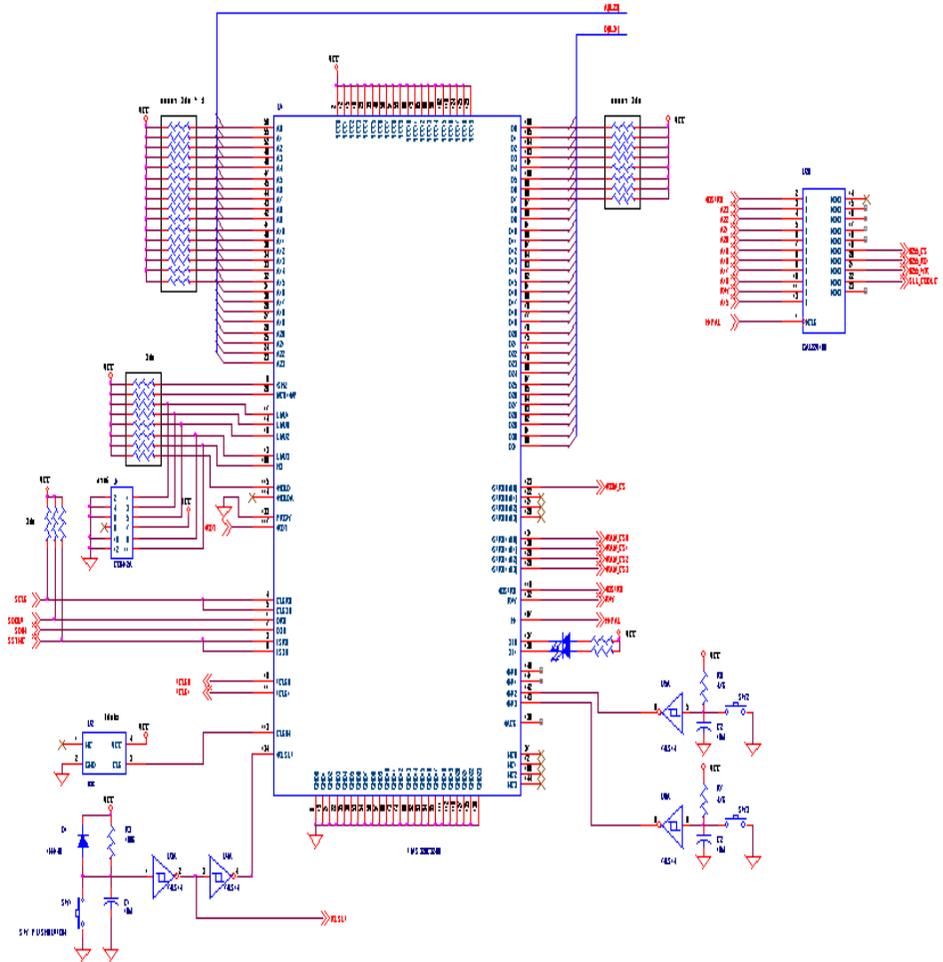
부록A.4 외부 인터럽트



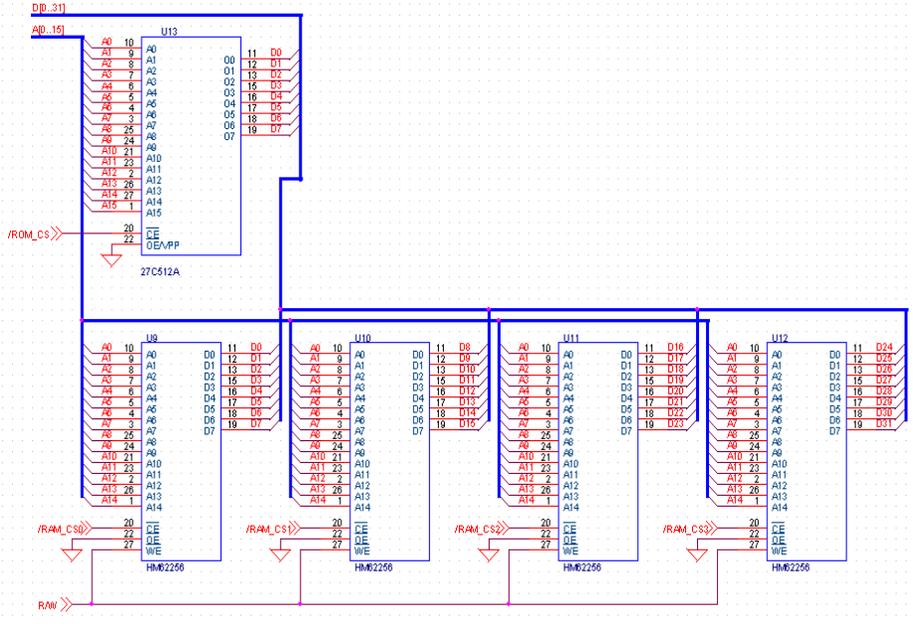
부록A.5 초음파 센서



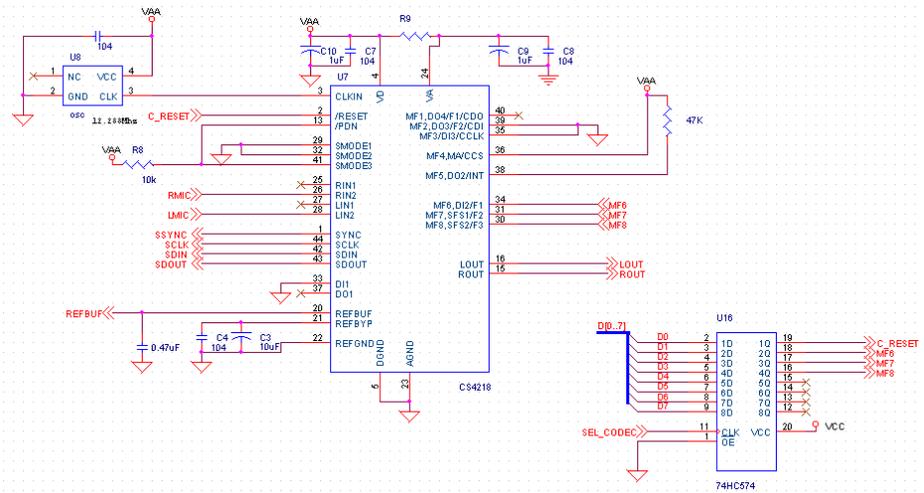
부록A.6 TMS320C32의 구조



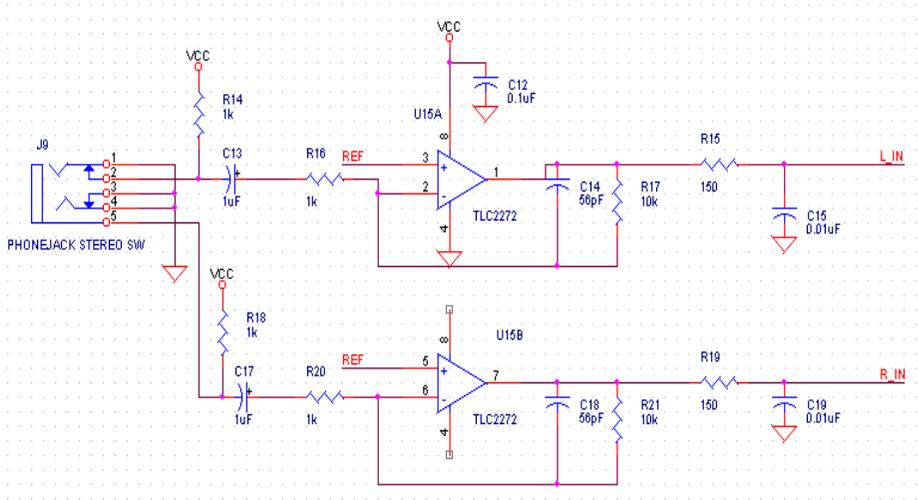
부록A.7 메모리 인터페이스 구조



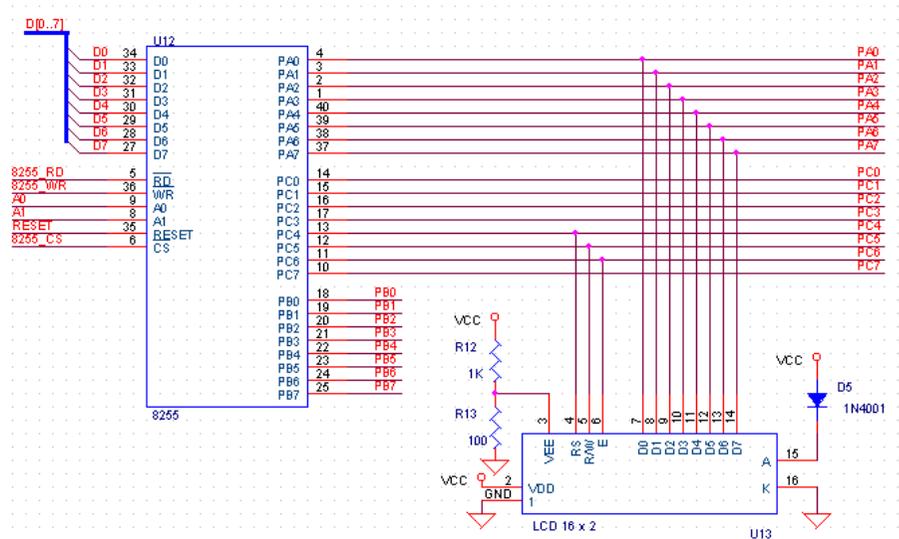
부록A.8 Codec CS4218 인터페이스 구조



부록A.9 음성입력부



부록A.10 LCD 인터페이스



부록 B. 사진

부록B.1 80C196KC Main Board



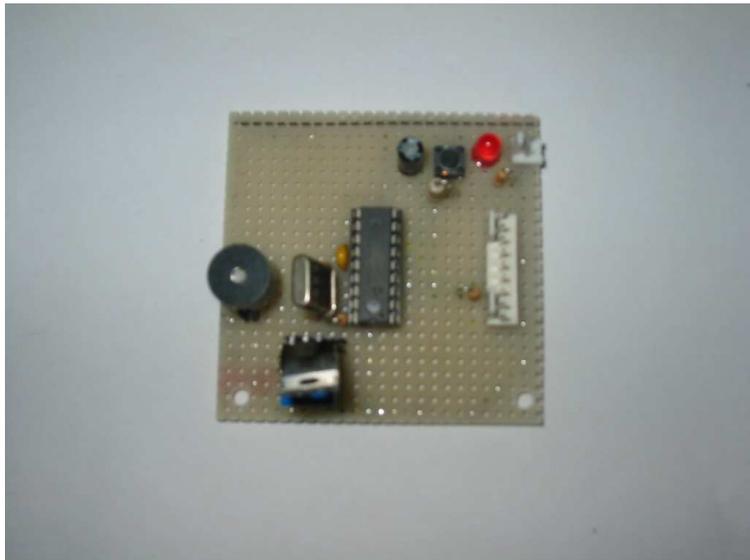
부록B.2 모터 구동부



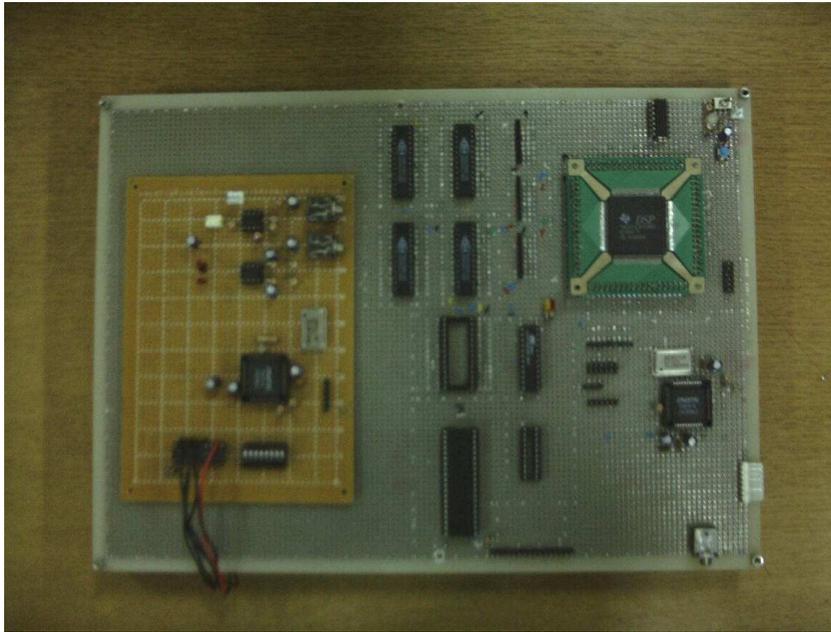
부록B.3 전원 레귤레이터



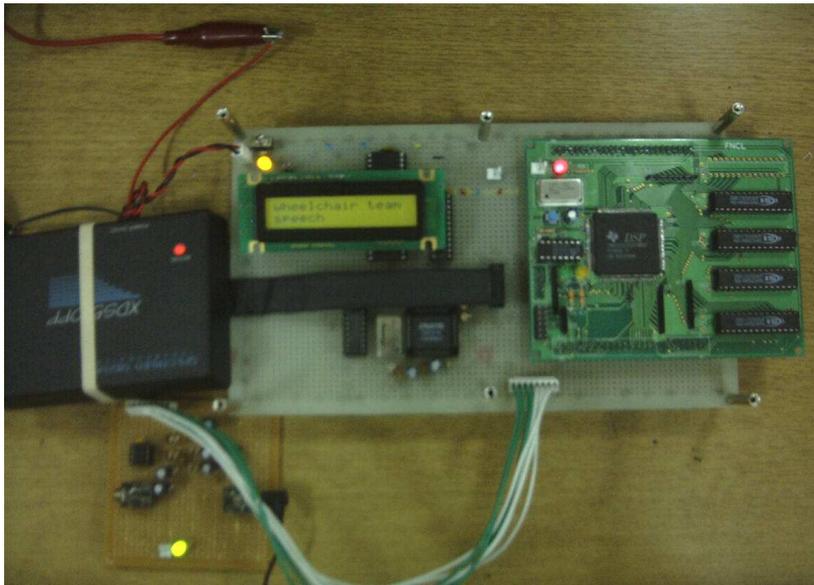
부록B.4 초음파센서 구동부



부록B.5 음성인식 보드(PCB전)



부록B.6 음성인식 보드 테스트 과정



부록B.7 설계된 음성인식 보드(PCB후)



감사의 글

본 논문이 완성되기까지 학부 때부터 세심한 배려와 관심으로 지도해주신 이상배 지도교수님께 머리 숙여 진심으로 감사의 말씀을 올립니다. 또한 논문 심사과정을 통해 부족한 점들을 조언해 주신 양규식 교수님, 임재홍 교수님, 진주산업대학교 탁한호 교수님과 메가트랜드 한운영 박사님께 깊은 감사를 드립니다.

항상 저희 연구실의 든든한 버팀목이 되어주신 동부산전문대학의 김일 교수님, 부족한 저를 항상 친형처럼 돌봐주고 격려해주신 최성욱, 채명기, 김관형, 이계현 선배님들과 졸업 후에도 언제나 따뜻한 관심을 보내주신 이영섭, 강동우, 김성주, 문주영 선배님들, 그리고 사랑하는 제 동기들 김태영, 이주원, 방은오 형님들과 박승현님, 또한 부족한 선배와 동고동락한 조병일, 이용수, 손창우, 공석민, 조동민, 이주상, 강제명, 김동한, 김영탁, 이창규, 문희근, 김정훈, 김수정, 강종윤, 류홍석, 장원일, 정성훈, 박주원님 들께도 깊은 감사를 드립니다. 그리고 항상 좋은 형님으로 많은 도움을 주신 이병도, 신영암님께도 감사드리며, 여러모로 같이한 한주섭, 양승국, 김정호님께도 앞으로 하시는 일에 항상 축복이 따르길 기원합니다. 또한 제가 부산에서 생활하는 동안 저에게 너무나 많은 도움을 주신 모든 분들에게 죄송한 마음과 함께 깊은 감사를 드립니다.

항상 멋있고 터프한 모습을 잃지 않고 변함없는 우정으로 뭉친 전자통신공학과 93학번 동기들과, 선후배님들께도 깊은 감사를 드리며 항상 고향의 따뜻함을 전해주는 승훈, 남의, 명구, 국현과 광주에 있는 여러 친구들에게도 감사의 말을 전하고 싶습니다.

언제나 부족한 제가 공부할 수 있도록 항상 어려움을 몰래 감추시고 한 평생 희생하신 아버님과, 항상 따뜻한 사랑으로 감싸주신 어머님, 그리고 항상 곁에서 돌봐주신 장인어른, 장모님, 언제나 부족한 저를 믿고 버팀목이 되어준 사랑하는 나의 예쁜 아내와 동생들 은미, 은경, 성도를 비롯한 모든 가족과 나의 보물들 민석, 동현에게 아빠로서 이 작은 결실을 바칩니다.