

## 인간과 로봇 협력작업을 위한 로봇 지능제어알고리즘 개발에 관한 연구

### A Study on Intelligent Control Algorithm Development for Cooperation Working of Human and Robot

이우송<sup>1\*</sup>, 정양근<sup>2</sup>, 박인만<sup>3</sup>, 정종교<sup>4</sup>, 김희진<sup>5</sup>, 김민성<sup>6</sup>, 한성현<sup>7</sup>

Woo-Song Lee<sup>1\*</sup>, Yang-Guen Jung<sup>2</sup>, In-Man Park<sup>3</sup>,  
Jong-Gyu Jung<sup>4</sup>, Hui-Jin Kim<sup>5</sup>, Min-Seong Kim<sup>6</sup>, Sung-Hyun Han<sup>7</sup>

#### 〈Abstract〉

This study proposed a new approach to develop an Intelligent control algorithm for cooperative working of human and robot based on voice recognition. In general case of speaker verification, Gaussian Mixture Model is used to model the feature vectors of reference speech signals. On the other hand, Dynamic Time Warping based template matching techniques were presented for the voice recognition about several years ago. We converge these two different concepts in a single method and then implement in a real time voice recognition enough to make reference model to satisfy 95% of recognition performance. In this paper it was illustrated the reliability of voice recognition by simulation and experiments for humanoid robot with 18 joints.

*Keywords : Isolated Word Recognition, Dynamic Time Warping(DTW), Gamma Distribution*

---

1\* 정회원, 교신저자, (주)선진기술 대표

E-mail: elflws@hanmail.net

2 신라정보기술(주) 대표

3 (주)인템 대표

4 (주)카바스 대표

5 경남대학교 산업대학원 석사과정

6 경남대학교 대학원 석사과정

7 경남대학교 기계공학부 교수

1\* SUNJIN TECHNOLOGY Co., Ltd.

E-mail: elflws@hanmail.net

2 Shinra Information Technology Co., Ltd.

3 Intem Co., Ltd.

4 KAVAS Co., Ltd.

5 M.S. Course, Manigerment & Indusy Graduate School  
Kyungnam University.

6 M.S. Course, Graduate School Kyungnam University.

7 Dept. of Mechanical Engineering, Kyungnam University.

## 1. 서론

최근 로봇산업의 성장과 함께 사람과 로봇간의 의사소통을 위해서 다양한 화자/음성 인식 알고리즘이 로봇 시스템에 사용되고 있다. 특히 인간과의 의사소통이 필수적인 실버메이트(Silver mate) 용 로봇의 경우, 로봇은 사용자의 음성을 분석하여 사용자가 누구인지 그리고 무엇을 말하는지를 파악해야한다. 본 논문에서는 음성인식분야 중에서도 고립단어인식에 대한 알고리즘을 소개한다. 이것은 로봇이 주인의 이름이나 명령어 등과 같은 단어 수준의 음성을 듣고 음성신호가 무엇인지, 누구에서부터의 것인지를 판단하는 것이 적용되고 있으며 특히 동적 시간 정합(DTW, Dymamic Time Warping) 알고리즘은 그 간단함과 적은 계산 량으로 인해 고립단어인식에 널리 사용되고 있다.

또 다른 음성 인식 기법에는 은닉 마르코브 모델(HMM, Hidden Markov Model)이나 신경회로망(ANN, Artificial Neural Network) 등과 같은 알고리즘이 있다. 이 두 가지 알고리즘의 경우, 연속단어인식(Continuous word recognition) 과 같은 고급 음성 인식에 사용되고 있다. 알고리즘도 고립단어인식에 사용될 수 있으나 간단한 계산을 필요로 하는 로봇 시스템에는 적용하기 힘들다.

음성 인식과 함께 로봇이 화자의 음성을 다른 것과 구별하기 위해서 화자 인식 알고리즘이 필요하다. 화자인식 알고리즘의 경우, 특정 화자의 특정 파라미터에 근거한 가우시안 혼합 모델(GMM, Gaussian Mixture Model)이 널리 사용되고 있다.

첨단산업사회로 변화하면서 사람들은 어렵고, 힘들고, 더러운 환경에서의 근무를 기피하는 현상이 일어나고 있다. 그래서 사람을 대신하여 다양한 작업을 수행하는 무인화 기술이 매우 요구되고 있는 실정이다. 즉, 기계장치들이 발전함에 따라 보다 편하게 조작하기 위한 연구가 진행되고 있으

며 주로 음성, 동작, 장애물회피 등을 통한 작업들이 많이 제안되고 있다. 특히 사람과 기계 사이의 편리한 상호 작용을 가능하게 하는 음성 인식 기능이 많이 제안되고 있다.

음성인식 기술은 사람의 언어를 해석하여 적절한 행동을 수행할 수 있는 지능시스템을 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 지속적으로 획기적인 제품들이 생산되고 있다[1-2].

사람의 목소리를 분석하고 응답하는 모바일 로봇은 컴퓨터가 발명된 이후로 많은 사람들이 공통적으로 추구해 온 꿈이며, 최근에는 음성인식기술은 비약적으로 발달하여 이를 다양한 분야에 적용시키려 꾸준히 노력하고 있다. 음성은 정보 전달 방법 중 가장 빠른 수단이며, 가장 친숙한 방법이므로 별도의 훈련이나 학습 없이도 사람들이 로봇을 손쉽게 사용할 수 있다. 그리고 산업현장과 과학 분야에서 이용되고 있는 로봇은 산업현장에서 뿐만 아니라 그 사용범위를 다양하게 하여 실생활에서도 수많은 작업을 수행할 수 있는 로봇의 필요성이 대두됨에 따라 모바일 로봇에 대한 연구가 활발히 진행되고 있다[3-7].

음성은 사람의 가장 자연스러운 통신방법으로, 사람과 기계사이의 자연스러운 통신을 위해 음성 인식에 관한 연구가 꾸준히 진행되어 왔으며 수십 단어 정도의 어휘를 대상으로 불특정 화자의 음성을 인식하거나, 화자적응을 통해 수만 단어를 인식할 수 있는 시스템들이 상용화되기 시작했다[8]. 따라서 로봇의 교시에 음성인식 기술을 활용한다면 사람에게 보다 친숙하면서 전문가가 아니더라도 로봇의 교시를 가능하게 하여, 전문 엔지니어 양성 및 프로그램 재개발에 따른 비용과 시간의

절감 효과를 가져 올 수 있다. 특히 열악한 환경에서의 로봇의 원격 조작 및 신체장애자를 위한 서비스를 제공할 수 있는 로봇으로도 이용될 수 있는 등 응용분야가 다양하다[9-10].

지금까지의 음성인식 방법을 크게 두 가지로 나누면, 음성·언어학적인 지식을 이용하는 지식 기반 접근방식과 입력된 음성신호의 스펙트럼을 분석하여 학습과정을 통해 미리 준비된 기준 패턴과 대조시켜 가장 유사한 것을 찾아내는 패턴정합 방법으로 나눌 수 있다[11].

인식하고자하는 음성 파형에 대한 학습과정이 필요한 패턴정합 방법에는 DTW(Dynamic Time Warping), HMM(Hidden Markov Model), 신경회로망(Neural Network) 등이 있다. DTW방식은 동적 프로그래밍에 의한 음성 패턴을 비선형 시간 정렬을 통하여 기준패턴과 인식하고자 하는 패턴을 비교하는 방식으로서 소규모 단위의 음성 인식에는 좋은 성능을 보이지만, 어휘의 수가 증가하는 대용량 연속 음성인 경우에는 증가하는 어휘수만큼 기준 패턴을 구성해야 하는 단점이 있다. HMM은 확률함수를 이용하여 음성신호의 시간에 따른 특징 변화를 통계적으로 모델링하는 것으로서 학습 시에 많은 정보와 계산량을 요구하지만 다수 화자의 음성특성을 모델링하는데 유리하고 높은 인식 성능을 나타낸다. 신경회로망은 인간 두뇌의 생물학적 신경 계통을 모방한 인공 신경망을 이용하여 여러 연결 마디들에 음성특징들을 분산 분포시켜 음성인식을 실현하고자 하는 방식으로 현재 소규모의 고립단어 인식에서는 우수한 성능 나타내고 있으나 대용량 또는 연속음성인식에 있어서는 아직 많은 연구를 필요로 한다[12-14].

본 연구에서는 음성인식 기법을 이용하여 모바일 로봇의 실시간 작업 동작제어에 대한 성능실험을 수행하고자 한다. 먼저 음성인식알고리즘개발을 위한 연구를 수행한다.

마지막으로 음성인식알고리즘 성능 실험을 수행하고 모바일 로봇의 동작제어를 음성 인식을 통하여 작업명령을 위한 작업동작제어 실현에 대한 성능실험을 검증하였다.

## 2. 모바일 로봇 명령제어 음성인식 알고리즘 개발

음성인식 방법에는 패턴매칭의 방법을 이용하는 DTW(Dynamic Time Warping)와 통계적 패턴인식을 이용한 HMM(Hidden Markov Model), 뇌의 구조를 모델링한 학습패턴 방법인 신경회로망(Neural Networks)이 있다.

인식 방법에는 2가지로 나누면 화자 종속형(speaker dependent type)과 화자 독립형(speaker independent type)으로 나눌 수 있는데, 전자는 한 사람에게만 적용이 가능한 방법이며, 후자는 모든 사람에게 적용하는 방법이다. 이 2가지 방법 중 화자 종속형은 DTW 인식 알고리즘이 사용되고 있다. 화자 독립형은 HMM 인식 알고리즘이 많이 사용되고 있으며, 음소별 음성인식에는 신경회로망도 사용되고 있다[15-16].

DTW를 이용한 인식방법은 훈련과정이 필요 없고, 고립단어 같은 작은 어휘 등에서 좋은 인식 성능을 나타내고 있다.

### 2.1 벡터 양자화

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. 양자화 테이블의 생성에는 K-means

알고리즘이나 이진트리 알고리즘이 주로 사용되며, 본 논문에서는 양자화 테이블을 생성하기 위해서 이진트리 알고리즘을 사용하였다. 그리고 벡터 양자화 크기인 M을 512로 설정하였으며, 오차 비율은 0.000001, e(splitting parameter)값은 0.001로 각각 설정했다.

이진트리 알고리즘은 1개의 벡터 코드북으로부터 분할(splitting)을 이용하여 M개의 코드북을 작성하는 방법이다[17-18].

이진트리 알고리즘 순서는 다음과 같이 다섯 단계로 나눌 수 있고 Fig. 1에서 도식적으로 나타내었다.[19]

- 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 (1), 식 (2)과 같이 분할한다. 여기에는 e는 분할 파라미터(splitting parameter)이고,  $Y_n$ 은 현재 벡터 코드북의 대표벡터 값이다.

$$Y_n = Y_n(1+e) \tag{1}$$

$$Y_n = Y_n(1-e) \tag{2}$$

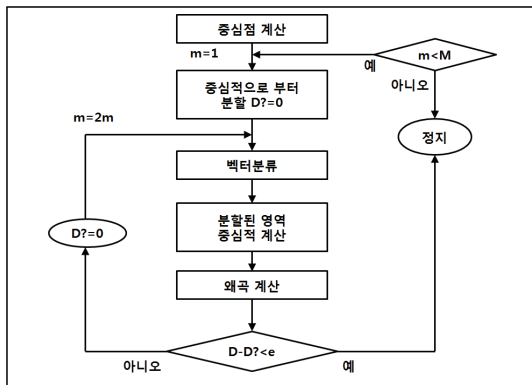


Fig. 1 Flowchart of binary tree algorithm

## 2.2 잡음처리 방법

잡음이 존재하는 환경에서는 음성인식시스템의 인식률은 현저히 차이가 난다. 이처럼 잡음이 음성인식시스템의 성능을 저하시키는 가장 큰 원인은 끝점 검출오차에 의한 영향과 잡음에 의한 특징 벡터의 왜곡이다. 본 연구에서는 잡음환경에 강한 시스템을 설계 및 구현을 하기 위해 잡음이 포함된 음성신호 중 실음성만을 추출하는 방법으로 스펙트럼 차감법 및 위너 필터링 방법을 분석 활용한다[20-21].

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. 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 (2.6)과 같다.[22]

$$y(k) = s(k) + d(k) \tag{3}$$

식 (3)을 주파수 영역으로 변환하면 식 (4)로 표현된다.

$$Y(w) = S(w) + D(w) \tag{4}$$

여기서, 식 (5)과 같이 정의한다.

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

$$Y(k) = \frac{1}{2\pi} \int_{-\pi}^{\pi} Y(w)e^{jwk} dw \tag{5}$$

식 (4)를 이용하여 파워 스펙트럼 추정 방법을 적용해서 전개하면 다음과 같이 된다.

$$|Y(w)|^2 = |S_m(w)|^2 + |D_m(w)|^2 + 2S_m(w)D_m(w) \quad (6)$$

$$\approx |S_m(w)|^2 + |D_m(w)|^2 \quad (7)$$

여기서, 식(8)과 같이 가정한다.

$$2S_m(w)D_m(w) = 0 \quad (8)$$

이후 식 (7)을 식 (9)으로 다시 나타내어, 이를 스펙트럼 차감법이라 한다.

$$|S_m(w)|^2 \approx |Y_m(w)|^2 - P_d(w) \quad (9)$$

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

식 (10)에서는 위상에 대한 정보를 나타낸 식이다.

$$S_m(w) = |S_m(w)| \angle Y_m(w) \Rightarrow S_m(n) = F^{-1}S_m(w) \quad (10)$$

이 때 스펙트럼 차감 필터의 H(w)는 미리 구한 잡음의 스펙트럼  $D_m(w)$ 의 평균을 사용한다. Fig. 2은 스펙트럼 파워 차감법에 대한 간단한 블록도이다.[23]

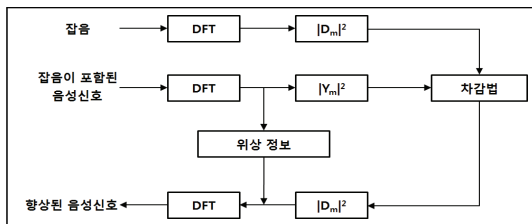


Fig. 2 Block diagram of spectral subtraction method

### 2.3 끝점 검출 알고리즘

음성신호를 처리하는 기본단위를 프레임이라고 하고, 한 프레임에 포함되어 있는 신호의 샘플수를 프레임 길이라 한다. 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[24].

#### 1) 영교차율

영교차율(ZCR : Zero Crossing Rate)은 분석구간 프레임 내에서 파형이 영점(zero)축과 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.[25]

영교차율은 유성음과 무성음을 판별하는데 유용한 파라미터가 되며 음성신호 중 영교차율이 높으면 무성음으로, 영교차율이 낮으면 유성음으로 판단한다. 영교차율에 대한 일반식은 식 (11)과 같다.

$$Z_i = \sum_{n=0}^{N-1} |sgn[s_i(n)] - sgn[s_i(n-1)]| \quad (11)$$

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

#### 2) 절대 에너지 방식

음성신호에서 절대 에너지는 무성음보다 유성음 부분이 크다. 절대 에너지 방식은 크게 Short-Time Energy, Short-time Magnitude 2가지 방식이 있다. 에너지 E/M은 이산계(discrete system)에서

다음 식 (12), 식 (13)로 주어진다.[26]

$$E_i = \sum_{n=0}^{N-1} [s_i(n)]^2 \quad (12)$$

$$M_i = \sum_{n=0}^{N-1} [s_i(n)] \quad (13)$$

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

### 2.4 특징 추출 알고리즘

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. Fig. 3는 음성 인식을 위한 선형예측법 및 켈스트럼(LPC & Cepstrum) 처리 절차이다[27].

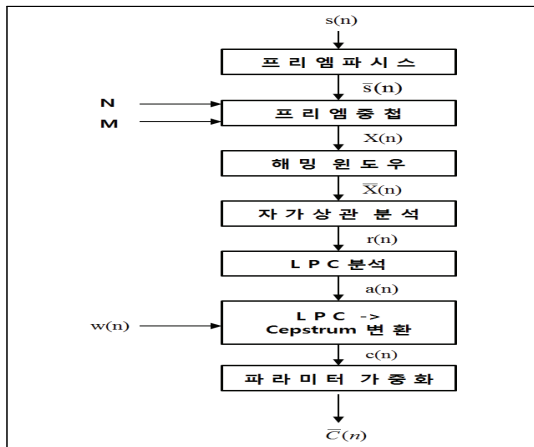


Fig. 3 Block diagram of LPC & Cepstrum

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

$$\bar{s}(n) = s(n) - as(n-1), \quad (0.9 < a < 1.0) \quad (14)$$

프리엠파시스 과정을 거친 신호  $\bar{s}(n)$ 는 N개의 샘플들로 구성된 프레임들로 인접한 프레임과 N-M 샘플씩 중첩하여 처리되며, 여기서 N은 프레임의 크기이고, M은 프레임의 중첩 크기이다.

각 프레임의 시작과 끝에서의 신호 불연속을 최소화시키기 위해서 사용하며, 주로 해밍(Hamming) 윈도우가 많이 사용된다. 해밍 윈도우는 다음 식 (15)로 표현된다.

$$w(n) = 0.54 - 0.46\cos\left(\frac{2\pi n}{N-1}\right), \quad (0 < n < N-1) \quad (15)$$

윈도우를 이용하여 추출한 각 신호 프레임으로부터 자기관계수를 구하는 과정으로 식 (16)과 같다. 여기서 m은 차수를 나타낸다.

$$r_x(m) = \sum_{n=0}^{N-1-m} \bar{x}(n)\bar{x}(n+m), \quad (m = 0, 1, \dots, p) \quad (16)$$

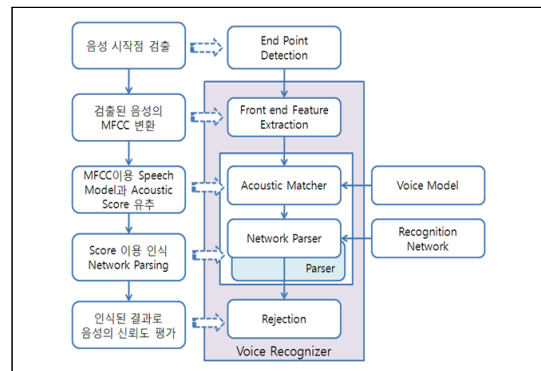


Fig. 4 Voice recognition model



## 2.5 음성인식장치의 하드웨어 구성

음성인식시스템을 실시간으로 처리하고 구성할 수 있는 임베디드 시스템으로 구성하는 것이 매우 중요한 특징이라 할 수 있다. 음성신호처리의 특성상 많은 양의 데이터를 고속으로 처리해야 하고 실시간 동작되는 시스템을 구성하기 위해서 충분히 속도가 빠른 고성능 32비트 부동소수점 연산 방식의 범용 DSPs인 TMS320C31를 사용하였다. TMS320C31의 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[29-30].

먼저 메모리로는 RAM에 SRAM 61C256\*4 (128Kbyte)을 사용하였고 ROM에 64Kbyte EPROM인 27C512를 사용하고 있다. STRB0, 1과 IOSTRB는 TMS320C31에서 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 코텍인 CS4218의 인터페이스에 사용하였다.

보드의 구성

- High Performance Floating-Point TMS320C31
- Analog data acquisition via the CS4218 analog interface circuit
- Expand memory : 64K×32 Bit High Speed CMOS SRAM
- EPROM

- XDS-510 (MPSD) emulator connecto
- Digital I/O 82C55 Interface eight 7-Segmen

실시간 처리를 위해서는 무엇보다도 시리얼 포트로 들어온 Data를 실시간 읽어 들일 수 있어야 한다. 그러기 위해서는 DMA는 CPU의 동작과는 독립적으로 Data를 읽을 수 있는 작용을 하기 때문에 실시간 처리를 위해서는 DMA의 기능은 중요하다. CS4218 16-Bit Stereo Audio Codec은 Sampling frequencies가 4kHz에서 50kHz까지 가능하며 본 논문에서는 8kHz로 Sampling이 되도록 하였다. 메모리는 SRAM의 64K×32bit를 사용하여 Data 크기를 구현하는데 사용하였다.

## 3. 성능 실험

### 3.1 음성인식모델의 구조 및 특징

Endpoint detection에서 음성의 시작점을 검출하면, 검출된 음성에 대해서 front end feature extraction에서 MFCC로 변환하고, MFCC를 이용하여 acoustic matcher에서 voice model과 acoustic score를 구한다. 그리고 그 score를 이용하여 인식 network을 parsing해서 인식된 결과가 나오면 rejection에서 음성의 신뢰도를 평가하게 된다.

### 3.2 음성인식 순서

음성인식은 등록되어있는 음성DB와 비교하여 결과를 추출한다.

- ① 음성DB Open  
 등록된 음성DB에는 각 단어의 Block Size, Zero Crossing Count, Average Energy Data 들이 들어있으며 음성DB 등록 프로그램을 이용하여 미리 음성 DB를 작성한다.
- ② Mic(마이크로폰)로 들어온 음성의 길이 추출  
 들어온 단어의 앞 & 뒤 의 잡음 길이를 제외한 순수 단어의 길이를 추출한다.
- ③ Block의 Zero Crossing횟수와 평균 Energy 계산  
 음성의 Block Size(일정한파형의 개수)와 Zero Crossing 횟수 Energy 값을 계산한다.
- ④ DB의 음성과 길이와 Zero Crossing횟수, 평균 Energy 비교하여 추출  
 DB에 있는 단어들의 Data를 현재 들어온 음성 Data와 비교하여 인식단어들을 축소시킨다.
- ⑤ Mic(마이크로폰)로 들어온 음성의 특징 추출(2진화)  
 현재 들어온 단어에 대한 특징점을 이진화하여 추출한다.
- ⑥ 축소된 DB의 결과에 대한 특징을 비교
- ⑦ 인식실험 및 성능분석

### 3.3 인식 단어 추출

마이크로폰으로부터 들어온 음성은 잡음 길이를 제외한 순수 음성의 단어만을 추출하고 Block의 Zero 교차 횟수와 평균 Energy를 계산하여 기존 등록된 단어와 비교하여 인식단어를 축소시키며

현재 들어온 단어에 대한 특징점을 이진화하여 추출하여 70%이상 일치되었을 때 인식하게 된다. Fig. 5은 “준비”에 대한 등록단어 및 인식단어의 파형이며, Fig. 6은 “앞으로”에 대한 등록단어 및 인식단어의 파형으로 나타낸다.

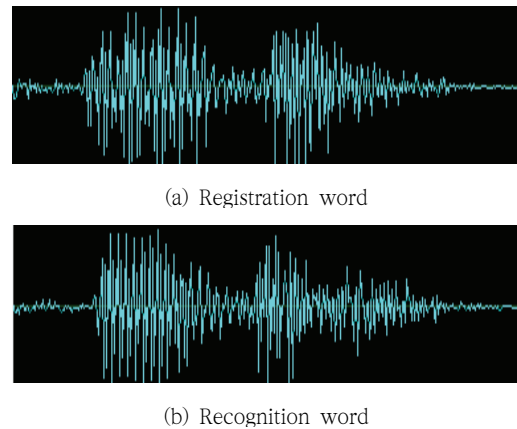


Fig. 5 Registration word and realization word about “Junbi”

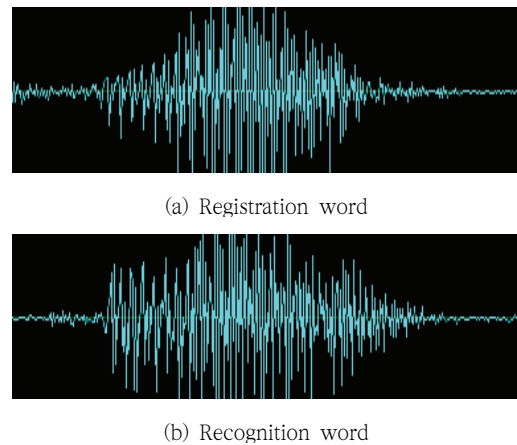


Fig. 6 Registration word and realization word about “Apuro”

다음의 Fig. 7은 등록된 단어의 Wave 파형을 나타낸다.



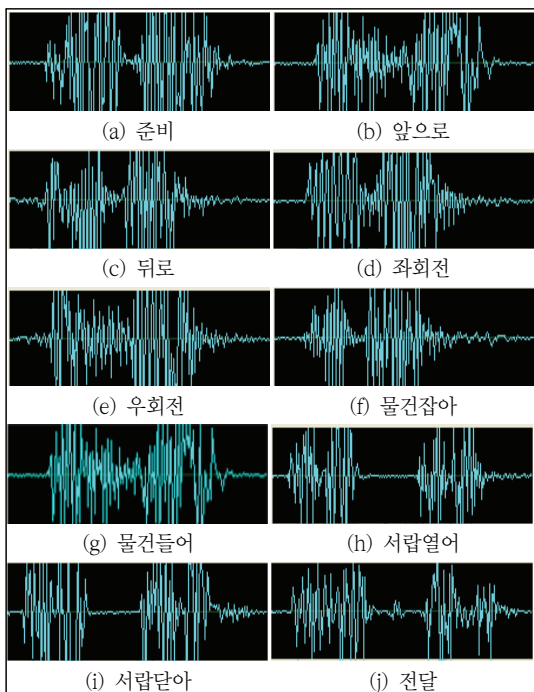


Fig. 7 Wave of Registered word

### 3.4 음성인식 실험방법

Table. 1. Voice recognition words.

구분	단어	구분	단어
1	준비	6	물건잡아
2	앞으로	7	물건들어
3	뒤로	8	서랍열어
4	좌회전	9	서랍닫아
5	우회전	10	전달

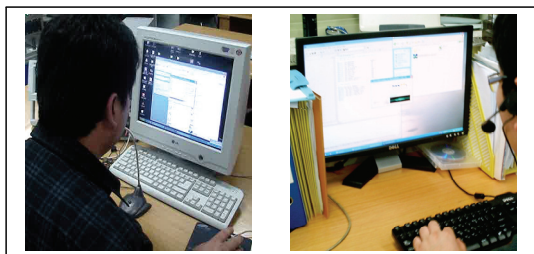


Fig. 8 Performance experiment scene on PC

본 연구에서는 고성능 음성인식시스템을 구성하기 위해 HMM 인식 알고리즘을 이용하여 Table. 1에 제시된 10개의 단어에 대해 인식률을 테스트 하였다. 또한 후처리 과정으로 방사 기저함수 신경회로망의 학습 후 음성인식을 구성하여 인식률을 HMM 인식알고리즘과 비교하여 성능평가를 하였고 인식시간을 테스트하기 위해 오실리스코프를 사용하였으며, 마이크의 잡음신호를 제거하기 위해 로우 패스 필터를 사용하였다. Fig. 8는 PC에서 성능 실험한 장면을 나타낸다.

Fig. 9~Fig. 14는 음성인식 단어 “앞으로”, “준비”, “우회전”의 음성파형, 벡터양자화 결과를 나타낸다.

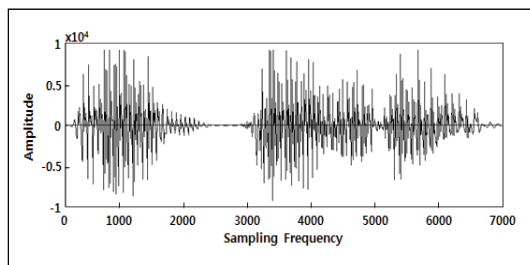


Fig. 9 Voice waveform of voice “Apuro”

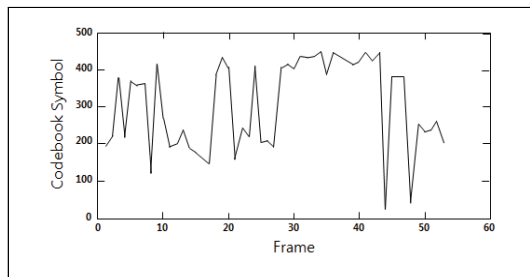


Fig. 10 Vector quantization result of the command “Apuro”

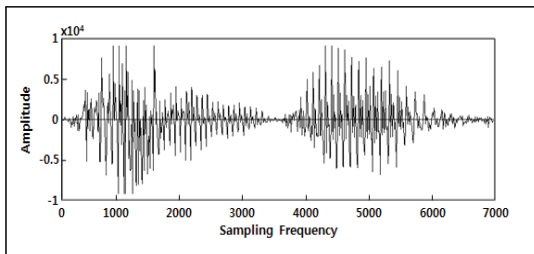


Fig. 11 Voice waveform of voice "Junbi"

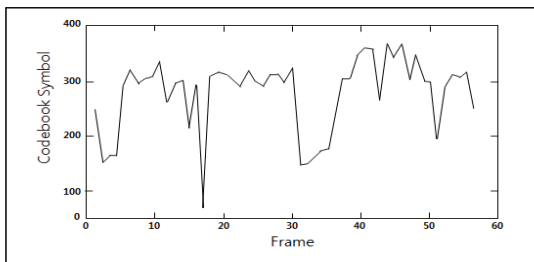


Fig. 12 Vector quantization result of the command "Junbi"

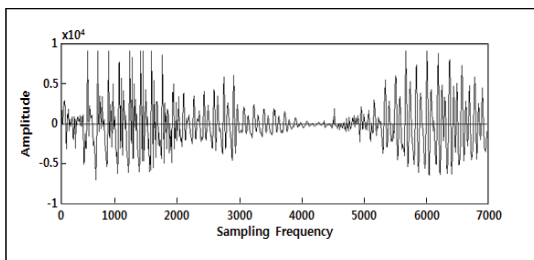


Fig. 13 Voice waveform of voice "Woohoejeon"

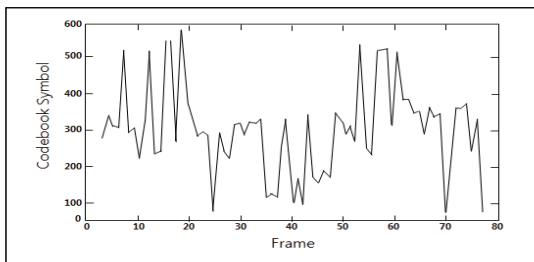


Fig. 14 Vector quantization result of the command "Woohoejeon"

### 3.5 로봇 동작제어 구현 성능 실험

#### 3.5.1 로봇의 제어시스템 구조 및 특징



Fig. 15 Remote control schematic diagram of voice recognition robot

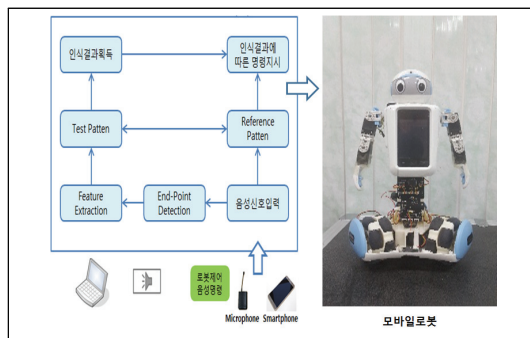


Fig. 16 voice recognition system

#### 3.5.2 음성인식 실험결과

음성인식 실험 프로그램 신뢰성 검증하였다.

Table. 2. Test result of voice recognition executive program.

준비	앞으로
<p>The screenshot shows the '준비' (Prepare) window of the 'Speech Recognition for FA SOLUTION' software. It includes a list of commands on the left and buttons for '음성연진 초기화' (Initialize voice recognition), '음성인식 시작' (Start voice recognition), and '음성인식 중지' (Stop voice recognition). The '준비' (Prepare) button is highlighted.</p>	<p>The screenshot shows the '앞으로' (Next) window of the 'Speech Recognition for FA SOLUTION' software. It includes a list of commands on the left and buttons for '음성연진 초기화' (Initialize voice recognition), '음성인식 시작' (Start voice recognition), and '음성인식 중지' (Stop voice recognition). The '앞으로' (Next) button is highlighted.</p>

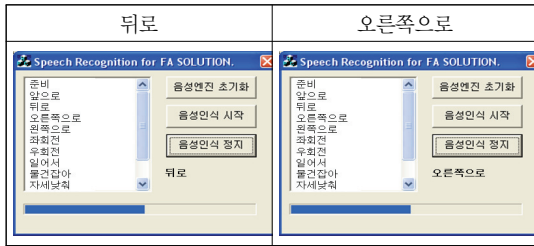


Table. 3은 20명이 10개 단어를 10번씩 발성한 음성 DB를 사용한(음성 DB①+음성 DB②) 음성을 학습에 사용하지 않은 10명이 10개 단어를 5번씩 발성한 음성 DB(음성 DB③)를 사용하여 테스트 한 결과이다.

Table 3. Result of voice recognition algorithm test for robot control

명명어	테스트 횟수	인식 횟수	인식률
전진	100	96	96%
후진	100	95	95%
좌로	100	94	94%
우로	100	96	96%
준비	100	94	94%
물건잡아	100	95	95%
물건놓아	100	94	94%
문열어	100	95	95%
문닫아	100	96	96%
전달	100	95	95%
합계	1000	잘못된 계산식	95.0%

#### 4. 결론

본 연구에서는 음성인식 기법을 이용한 휴모바일 로봇의 실시간 작업 동작제어에 관한 연구를 수행하였다. 구현된 음성인식시스템은 통계적 확률방법을 이용하여 인식을 수행 할 경우의 오인식을 최소화하기 위한 후처리 과정으로 신경회로망

의 일종인 방사 기저함수 신경회로망을 결합하여 보다 높은 인식률을 얻을 수 있는 통계적 확률방법과 방사 기저함수 신경회로망의 시스템을 적용하였다.

음성인식알고리즘의 인식실험 및 실제 로봇에 적용하여 성능평가를 통해 다음과 같은 결과를 얻었다.

첫 번째로 임의의 환경에서 음성에 의한 모바일 로봇의 경로생성이 가능한 통계적 확률방법을 이용한 음성인식알고리즘을 이용하여, 인식을 향상 위한 방안으로 통계적 확률방법에 의한 인식을 수행한 후 그 신뢰도를 적절히 판단하여 신뢰도가 낮은 경우 신경회로망에 의한 재인식 과정을 후처리에 도입하였다.

음성인식 실험결과는 10명이 각 단어 당 5번씩 발성하여 음성인식실험을 실행하였고, 또한 제안한 고성능 음성인식알고리즘의 인식률이 87.4%로 향상됨을 알 수 있었다.

두 번째로 로봇적용 단어를 대상으로 실제 모바일 로봇에 적용하여 음성교시에 의한 경로생성 실험을 수행하여 성능평가를 하였다. 설정된 환경에서 로봇은 전체 인식단어와 로봇적용 단어에 대해 80%이상의 인식률을 얻을 수 있었다.

본 연구에서는 고성능음성인식알고리즘을 통해 음성인식률 향상을 보였으나, 로봇의 작업동작 실험 수행 시에 외부의 잡음이 없는 상태에서 실험을 수행하여 외부 상황에서 발생 시 음성인식에 대하여 향후 연구가 계속 필요할 것으로 생각된다.

#### 참고문헌

[1] L.R. Rabiner, "A Tutorial on Hidden Markov Models and Selected Application in Speech Recognition," Proc. of the IEEE, Vol. 77, No.

- 2, pp. 257-285, February (1989).
- [2] Siverman H. F., and Morgan D. P, "The application of dynamic programming to connected speech recognition," IEEE Transaction on Acoustics, Speech, and Signal Processing Magazine, Vol. 7, pp. 6-25, July (1990).
- [3] Myers C. S. and Rabiner L. R., "A comparative study of several Dynamic Time-Warping Algorithm for Connected-Word Recognition," Bell system Tech. J., 60(7) pp 1389-1409, September (1981).
- [4] 박진영, "HMM을 이용한 화자인증 시스템의 구현", 동아대학교 공학석사 학위논문, (2004).
- [5] 정양근, 조상영, 양준석, 한성현, "화자독립방식에 의한 음성인식 알고리즘 개발 및 실시간 실현에 관한 연구", 한국산업융합학회 논문집 제 18권 3호, pp.250-259, (2015).
- [6] L.R. Rabiner, "A Tutorial on Hidden Markov Models and Selected Application in Speech Recognition," Proc. of the IEEE, Vol. 77, No.2, pp. 257-285, February (1989).
- [7] 이우송, 하연태, 정양근, 박인만, "스마트팩토리를 위한 듀얼암을 갖는 모바일 로봇의 유연제어에 관한 연구", 한국산업융합학회 논문집 제 19권 1호, pp.69-74, (2016).
- [8] 심병균, 이우송, 한성현, "대화기능에 의한 모바일 로봇의 실시간 주행제어", 한국산업융합학회 논문집 제 16권 4호, pp.127-132, (2013).
- [9] S. Yamamoto, K. Nakadai, H. Tsujino, and H. Okuno, "Assessment of general applicability of robot audition system by recognizing three simultaneous speeches." in Proc. of IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS-2004). IEEE and RSJ, p. to appear, (2001).
- [10] Seong-Su, Lee, "Hybrid Neural Network Controller Design of DSP for 3-Phase Induction Motor of Real-time Control," Wonkwang University, Master's thesis, pp.20-27, (2004).
- [11] M.T Jhonson, "Capacity and Complexity of HMM Duration Modeling Techniques," IEEE Signal Processing Letters, Vol. 12, 2, pp. 407-410, (2005).
- [12] P. Renevey, R. Vetter, and J. Kraus, "Robust speech recognition using missing feature theory and vector quantization," in Proc. Of Eurospeech-2001. ESCA, pp. 1107-1110, (2001).
- [13] 이우송, "음성명령에 의한 모바일 로봇의 기능제어에 관한 연구", 경남대학교 공학박사 학위논문, (2010).
- [14] 이윤근, 박준, 김상훈, "음성인터페이스 기술," 전자통신 동향분석, 제 20권 4호, pp. 38-48, (2005).
- [15] Lawrence Rabiner, "A Tutorial on Hidden Markov Models and Selected Application in Speech Recognition", Proc. IEEE, Vol 77, No. 2, february (1989).
- [16] 심병균, "음성명령 기반 자율주행 이동로봇의 기능 제어", 경남대학교 공학석사 학위논문, (2010).
- [17] 김 상덕, 이 극, "연속분포 HMM을 이용한 음성인식 시스템에 관한 연구", 한국멀티미디어학회 추계학술발표논문집 pp. 221 - 225, (1998).
- [18] D.Psalitis, A. sideris, and A. Yamamura, "AMultilayered Netral Network Controller", IEEE Control Systems Magazine, pp17-21, (1988).
- [19] 정양근, 백승하, "2월 구동 모바일 로봇의 정밀 위치제어", 한국산업융합학회 논문집 제 18권 2호, pp. 67-74, (2015).
- [20] L. R. Rabinar and B. H. Juang, "An Introduction to Hidden Markov Models," IEEE ASSP Magazine, Vol. 3, No. 1, pp 4-16, (1986).
- [21] L.-X. Wang, A Course in Fuzzy Systems and Control. Upper Saddle River, NJ: Prentice-Hall, (1997).
- [22] Jang, Sun, and Mizutani, Neuro-Fuzzy and Soft Computing, Prentice-Hali, (1997).
- [23] J. T. Spooner and K. M. Passino, "Stable adaptive control using fuzzy systems and neural networks," IEEE Trans. Fuzzy Syst., vol. 4, pp. 339-359, Aug. (1996).
- [24] T. Matsuoka "Large-vocabulary continuous-speech recognition using Japanes business newspaper (Nikkei)" DARPA Speech Recognition Workshop, pp. 137 - 142, (1997).

- [25] H. J. M. Steeneken and D. A. van Leenwen, "MultiLingual Assessment of speaker independent large vocabulary speech-recognition systems: SQUALE Project," Proc. EUROSPEECH-95, pp. 1271-1274.
- [26] 노연 후 쿡, 김기복, 조상영, "모바일 로봇의 충돌회피 알고리즘 개발", 한국응용융합학회 논문집 제 18권 2호, pp. 99-109, (2015).
- [27] Won-Jun Hwang, In-Man Park, Un-Wook Kang, Sung-Hyun Han, "A Study on Real-Time Autonomous Travelling Control of Two-wheel Driving Robot Based Ultrasonic Sensors", 한국산업융합학회 논문집 제 17권 3호, pp. 151-169, (2014).
- [28] 이우승, 심현석, 하연태, 김종수, "이동로봇의 자율주행제어에 관한 연구", 한국산업융합학회 논문집 제 18권 1호, pp. 10-17, (2015).
- [29] 심현석, 김태관, "비전기술에 의한 모바일 로봇의 정밀 자세 제어 알고리즘 개발에 관한 연구", 한국산업융합학회 논문집 제 18권 2호, pp. 129-138, (2015).
- [30] 김현근, 심현석, 황원준, "무인FA를 위한 자율주행 로봇의 경로계획 및 실시간 궤적제어에 관한 연구", 한국산업융합학회 논문집 제 19권 2호, pp. 75-80, (2016).
- [31] 이우승, 심현석, 하연태, 김종수, "이동로봇의 자율주행제어에 관한 연구", 한국산업융합학회 논문집 제 18권 1호, pp. 10-17, (2015).

(접수: 2017.07.28. 수정: 2017.08.12. 게재확정: 2017.08.26.)