한국생산제조시스템학회지 Vol.20 No.2 2011. 4. pp. 207~213
Korean Society of Manufacturing Technology Engineers

기술논문

음성명령에 의한 모바일로봇의 실시간 무선원격 제어 실현

심병균*, 한성현+

(논문접수일 2011. 1. 20, 심사완료일 2011. 3. 28)

Real-Time Implementation of Wireless Remote Control of Mobile Robot Based-on Speech Recognition Command

Byoung-Kyun Shim*, Sung-Hyun Han+

Abstract

In this paper, we present a study on the real-time implementation of mobile robot to which the interactive voice recognition technique is applied. The speech command utters the sentential connected word and asserted through the wireless remote control system. We implement an automatic distance speech command recognition system for voice-enabled services interactively. We construct a baseline automatic speech command recognition system, where acoustic models are trained from speech utterances spoken by a microphone. In order to improve the performance of the baseline automatic speech recognition system, the acoustic models are adapted to adjust the spectral characteristics of speech according to different microphones and the environmental mismatches between cross talking and distance speech. We illustrate the performance of the developed speech recognition system by experiments. As a result, it is illustrated that the average rates of proposed speech recognition system shows about 95% above.

Key Words : Voice Command(음성명령), Remote Control(원격제어), Wireless Communication(무선통신), Voice Recognition Algorithm (음성인식 알고리즘)

1. 서 론

지능형 이동로봇에 대한 연구가 활발히 진행됨에 따라 지능로봇은 향후 일상생활 용품의 하나로서 자리 잡을 가능성을 나타내면서, 관련 기술의 개발이 이루어지고 있으며 지능로봇 기술의 발전과 더불어 사람과 로봇의 상호작용을 위한 많은 방법들이 제안되어지고 있다[1~4]. 주로 음성, 영상동작 등을 통한 상호 작용들이 제안되어지고 있는데 로봇이 기존의 고정된 매니퓰레이터 형태에서 이동로봇의 형태로 진화함에 따라 사람과 로봇의 상호작용을 위한 기술인 음성명령에 의한 제어 기술의 필요성이 더욱 커지고 있다[5,6]. 따라서 자율주행 이동로봇은 공간 이동성과 특수 기능성을 기반으로 공장자동화, 사무실자동화, 설비관리, 무인탐사 등 다양한 분야에 응용이 가능하고, 특히 인간을 대신하여 작업을 할 수 있는 시스템의 개발에 자율주행 로봇의 활용이 다양한 형태로 연구되고 있고, 이를 통한 개발 후 실 적용 사례도 급증하고 있다[7,8].

음성인식 및 화자인식 현재의 기술에서 로봇 응용을 위한 원격 마이크 환경의 전처리 음질향상을 위해서는 어레이 신호처리 기술이 필수적이며, 현재 기술은 저주파 부분의 잡음제거 효과가 상대적으로 저조하고 방향 검지 성능에의 의존도가 큰

* 경남대학교 첨단공학과 (shimbk@kyungnam.ac.kr)
주소: 631-701 경남 창원시 마산합포구 월영동 449번지
+ 경남대학교 기계자동화공학부

문제가 있다[9]. 또한 음성언어 후처리 기술은 현재 초기 연구 단계로 아직 보편적으로 표준화된 연구결과가 없는 상태이며, 어휘 또는 간단한 구문/의미단계 지식만으로 수정하는 방식이 사용되고 있다[10].

대화 모델링을 위한 현재의 접근방식들은 지식 집약적인 접근 방식으로서 강건성 유지와 영역 확장이 어려운 문제점을 지니며, 대화모델의 개념적 부분 이외에 해당언어 특성에 대한 고려도 필요하다.

최근 로봇 시스템의 제어기는 조절하려는 시스템의 운동방정식을 알고 있을 때만 가능하다. 그러나 이동로봇의 구조는 로봇의 운동방정식이 비선형성을 나타냄으로써 대부분의 경우 운동방정식을 정확히 구하지 못하는 경우가 많다. 또한, 널리 사용되는 고전적 제어의 경우 이득 값이 고정되어 있으므로 이러한 고전적 제어방법은 외부환경의 변화에 대응할 수 없으므로 제어성능에 한계점을 드러내고 있다.

본 연구에서 설계되는 음성인식에 의해 제안된 로봇 시스템의 주요 특징을 요약하면 다음과 같다. 무선 네트워크를 통한 실시간 원격제어와 무인원격제어기능을 하며 Home Security 실험/실습 및 음성인식을 사용한 무인 로봇 원격제어가 가능하고 음성전달 방법은 PC상의 송신 마이크로폰을 이용하여 음원을 분석 후 로봇을 제어 할 수 있으며 음성인식 리모컨을 사용하여 로봇제어가 가능하도록 하였다.

본 연구에서는 음성인식을 이용하여 자율주행 로봇의 속도 및 방향 제어를 무선통신 원격제어에 의하여 수행되어 이동형 로봇 제어기의 성능이 검증된다.

2. 음성인식 알고리즘

2.1 알고리즘 개요

로봇이 화자인식(speaker recognition)이라는 것은 식별(identifying)되어야할 화자들 간의 발음으로부터 데이터를 모으는 작업인 훈련(training)과정과 임의의 발음을 식별하는 판별(testing)과정으로 이루어진다. 화자인식은 크게 두 가지 카테고리로 나누어지는데 closed-set problem과 open-set problem이다. Closed-set problem 이란 N명의 알려진 화자 중에서 어떤 사람인지를 식별해 내는 것이다. 이 경우 N이 커지면 커질수록 어려운 식별하는 것은 어려운 문제가 될 것이다. 이와 달리 open-set problem은 식별해 내고자 하는 화자가 N명의 알려진 화자의 그룹에 속하는가 하는 것을 판단해야 하는 문제이다. 이 두 문제에 대해서 전자는 화자식별(speaker identification) 후자는 화자검증(speaker verification)이라고도 부른다.

① 특징추출(Feature selection(MFCC))

화자인식과정에서 중요한 과정중 하나는 화자를 잘 식별할 수 있는 충분한 사전정보를 확보하는 일이다. 이러한 사전정보는 식별할 화자의 모델을 만드는데 사용된다. 음성은 시간이나 문맥에 따라서 변화가 심한 것이 특징이다. 하지만 근본적인 특징(essential characteristics)는 상대적으로 느리게 변한다. 화자인식은 개인의 음성 특징이 유일하다는 사실을 근거로 하고 있으며, 이는 사람마다 성도(vocal tract)의 크기/부피/길이 등이 다르기 때문에 가능하다. 즉, 화자인식을 위해서 화자의 특성을 잘 나타낼 수 있는 feature를 선택하는 것이 첫번째 문제이다.

음성은 성도에서 나오는 excitation신호가 성도, 입술 등과 같은 time-varying filter를 통해 나오는 신호라고 모델링 할 수 있으며, 음성특징의 변화는 비교적 느리므로 짧은 음성구간에서는 time-invariant filter라고 가정할 수 있다.

사람의 음성을 방대한 음성데이터를 중요한 정보로 표현하는 과정을 통해서 필요한 자료의 양을 줄일 수 있다. 만약에 16kHz 20ms의 음성 프레임을 12개의 특징벡터로 나타낼 수 있다면 320/12=26.7 의 데이터 reduction rate를 얻을 수 있다. 이렇게 음성의 특징을 나타내는 feature로는 cepstral feature가 있다.

$$\text{cepstrum(frame)}=\text{inverse FFT}(\log|\text{FFT(frame)}|)$$

여기서 FFT 는 fast Fourier transform을 의미한다. LPC기반 켑스트럼(LPCC)와 더불어 FFT기반 켑스트럼인 멜 켑스트럼(MFCC)은 음성 신호의 대표적 특징 파라미터이다. 이 MFCC는 인간의 귀가 저주파영역에서 민감하고, 고주파영역에서 둔감한 사실을 이용하여 filter bank를 통과시킨 것이다.

LPC기반 방법은 all-pole 모델에 기인한 스펙트럼 분석을 수행하며, 음성 특징 파라미터를 빠르고 정확하게 추정한다. 반면, FFT기반 켑스트럼 계수는 복수 켑스트럼이 Fourier 스펙트

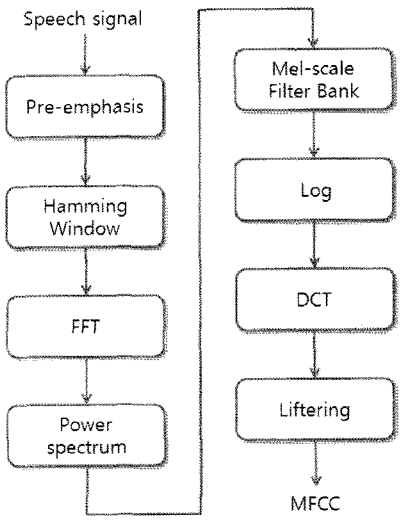


Fig. 1 Process of MFCC extraction

럼의 진폭만으로 계산될 수 있도록 최소 위상열로써 음성 신호를 가정함으로써 얻어진다. FFT기반 파라미터의 장점은 잡음에 강하며 비균일(bark, mel) 스케일로 주파수를 묶을 수 있다는 점이다.

MFCC 특징벡터를 얻기 위한 일련의 과정은 Fig. 1과 같다.

② 백터 양자화(Vector Quantization)

Speech data의 Cepstrum을 구함으로서 그 bit수를 줄이는데 성공함과 동시에 발음상의 특징을 나타내는 value를 찾는데 성공하였다. 이들 데이터의 분포를 나타내기 위한 손쉽고 효과적인 방법의 하나로 Vector quantization을 들 수 있다. 몇 개의 code vector들로서 많은 데이터를 나타내는 작업을 의미한다. 즉 입력 vector들과 code vector들 간의 match는 distortion measure 를 통해서 이루어지며 결과적으로 입력 벡터들은 이들 distortion measure 값이 최소화 되는 방향으로 이루어진다. 즉 vector quantization은 1. the distortion measure 2. the generation of optimal code words in the 코드북을 포함하는 일련의 과정으로 요약될 수 있다.

x를 각 원소가 real value 인 d-dimensional vector라고 하자. Quantization operator를 $q()$라고 하면 $z=q(x)$ 라고 code word를 나타낼 수 있다. $Z=\{z_j, 1 \le j \le M\}$ 의 원소 z_j는 j 번째 code word 이다. 이때 코드북 크기가 M이며 M개의 code word를 구하기 위해서 d-dimensional space를 M개의 Cell로 나누어야 한다.

③ Gaussian Mixture Model(GMM)

GMM은 출력확률밀도함수가 Gaussian density mixture인 1개의 상태만으로 구성된 CHMM(Continuous HMM)의 한 형태이다.

화자인식에 GMM을 사용하는 이유로 두 가지를 들 수 있다. 첫째로, GMM은 음향학적클래스(acoustic class)의 집합을 모델링할 수 있다는 것이다. 화자의 목소리에 대응되는 음향 공간은 모음이나 비음, 파찰음과 같은 음소를 표현하는 음향학적클래스의 집합으로 표현될 수 있는데, 이러한 음향학적클래스는 화자를 구별하는데 이용되는 화자의 성도에 대한 정보를 가지고 있다. i^{th} component 밀도의 평균 μ_i으로 표현되고, 평균 스펙트럼형태의 변화는 공분산행렬 $\sum_i$로 표현된다. 모든 학습 및 테스트의 음성은 레이블 되지 않기 때문에, 음향학적클래스는 hidden으로 볼 수 있다. 독립특징벡터를 가정하면, 이러한 hidden 음향학적클래스로부터 추출된 특징벡터의 관측밀도가 Gaussian mixture이다.

두 번째로, Gaussian basis 함수의 선형조합은 샘플분포(sample distribution)의 클래스를 표현할 수 있다는 것이다. GMM의 성질 중 하나가 임의의 형태를 가지는 밀도를 부드러운 형태로 근사시키는 것이다. Unimodal 가우시안 화자모델은 평균벡터(mean vector)와 공분산(covariance)으로 화자의 특징분포를 표현하고, VQ distortion 모델은 특징벡터의 이산집합으로 화자분포를 표현한다. 이와 같은 점을 고려하여 구성된 GMM은 가우시안 함수의 이산집합을 사용하고, 각각의 평균과 공분산을 가지게 함으로써 이들 두 모델의 특징을 혼합한 형태이다.

따라서 주요 명령어만을 인식하거나, 명령어 전후에 임의의 음성이 발성되어도 명령어만을 추출하여 인식하거나, 또는 정형화된 문장 형태로 발음된 음성을 인식하여 명령을 수행하는 기술을 말하며 여기서는 회자독립방식에 의한 음성인식 알고리즘을 이용한 로봇 원격 제어를 수행하였다.

2.2 음성인식 방법

① 인식대상단어

Automatic Speech recognizer(ASR, 음성인식기)는 미리 정해져 있는 인식대상단어나 문장을 인식하도록 구성되어 있다. 인식기는 수백 msec정도의 utterance를 입력으로 받아서 발성음에서 noise와 음성을 구분해서 인식대상단어 중 utterance와 주파수특성이 가장 유사한 단어를 추정하게 된다.

② Hidden Markov Models(HMM)

Hidden Markov Models(HMM)는 음성데이터를 통계적으로 모델링한 것으로 단어(word)나 보조단어(sub-word) (e.g. phoneme)를 구성하는데 사용된다. 각 모델은 음성데이터를 사용하여 통계적으로 음성데이터를 가장 잘 표현할 수 있도록 계산된 것으로, 이 과정을 '훈련(training)'이라고 한다. 트레이닝에 사용되는 음성데이터는 인식에 사용될 수 있는 음성데이터들의 대표집단으로 구성되어야 하지만, 일반적으로 그렇지 못하기 때문에 모델을 인식에 사용되는 음성데이터와 유사하게 수정하는 '각색(adaptation)'과정이 필요하게 된다. 트레이닝에 사용되는 음성데이터를 한 사람에게서 모두 받으면 'speaker dependent system'이라고 하며, 다양한 사람들로부터 많은 음성데이터를 받아서 훈련(training)하는 경우 'speaker-independent system'기법을 적용하였다.

③ 끝점 탐색(End Point Detection)

인식대상단어는 단어가 될 수도 있고, 문장이 될 수도 있다. 그래서 사용자는 정해지지 않은 임의의 기간 동안 발성하게 되므로, 인식시스템도 음성이 시작되는 부분과 끝나는 부분을 추정하여 음성구간동안 인식을 하고 음성이 끝나면 인식결과를 사용자에게 알려주어야 한다. 이와 같이 음성이 시작되는 구간과, 단어사이의 정지시간(pause)는 무시하고 사용자의 전체 발성이 끝나는 곳을 검출하는 모듈(module)을 끝점 탐색 기법으로서 화자독립방식의 실현에 적용하였다.

④ Front End

인식률과 효용성을 높이기 위해 HMM은 음성데이터를 바로 모델링 하지 않고, 음성데이터를 Mel-frequency cepstral coefficients(MFCC) 같은 음성특징벡터로 변환시켜서 모델링 한다. 이 변환과정을 'front end'부분에서 수행되어지는 'feature extraction'기법을 적용하여 실행하였다.

⑤ 파서(Parser)

인식기가 실제로 인식하기 위해서는 sub-word model들로 구성된 인식 네트워크와 utterance가 가장 매치(match)가 잘되는 경로(path)를 찾아야 하는데, 인식기에서 경로를 찾는 기능인 '파서(parser)'라고 하는 기능을 구비하고 있다.

⑥ 거절(Rejection)

음성인식시스템이 오인인식을 일으키게 하는 많은 이유들이 있기 때문에 방언을 사용한다던지, 인식기의 훈련(training)시에 사용되지 않은 억양으로 발성을 한다든지, 배경잡음이 높거나, 인식기의 인식대상단어로 등록되어 있지 않은 단어나 문장을 발성하는 경우 등으로 인해 오인식이 발생한다. 이러한 이유로 인해 인식기에서는 거절기능인 '거절(rejection)'기능을 가지고 있는데, 거절(rejection)기능은 인식기에서 인식결과의 신뢰도를 추정하여 신뢰도가 낮은 경우 사용자에게 인식결과를 검증할 수 있도록 하는 기능을 구비하고 있다.

2.3 음성인식 구조

Fig. 2는 인식시스템의 개념적인 모델이다. Endpoint detection에서 음성의 시작점을 검출하면, 검출된 음성에 대해서 front end feature extraction에서 MFCC로 변환하고, MFCC를 이용하여 acoustic matcher에서 speech model과 acoustic score를 구한다. 그리고 score를 이용하여 인식 network을 parsing해서 인식된 결과가 나오면 rejection에서 음성의 신뢰도를 평가하게 된다.

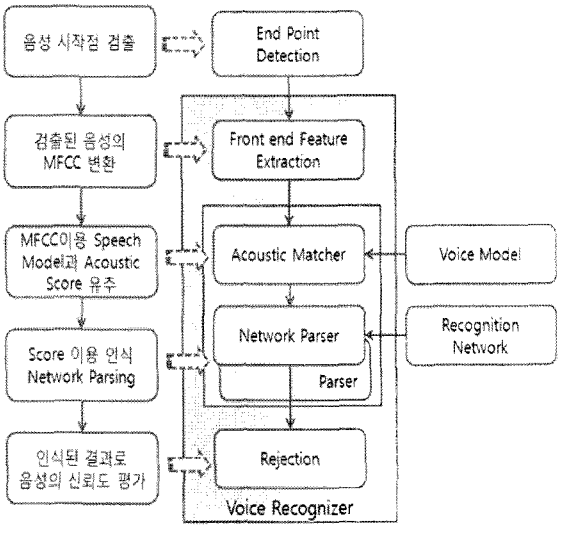


Fig. 2 Schematic diagram of voice recognition algorithm

2.4 화자인식 방법

화자인식은 화자 식별(Speaker Identification)기술과 화자검증(Speaker aerification)기술로 나눌 수 있다. 화자식별 기술은 고립 단어인식과 개념이 비슷하다. 고립단어인식은 발화된 음성과 가장 가까운 단어를 등록된 인식 대상 단어 중에서 찾아낸다. 마찬가지로 화자식별도 등록된 화자 중에서 가장 유사한 화자를 골라내는 것이다. 화자 검증 기술은 핵심어 인식처럼 승인(Acceptance)과 거절(Rejection) 과정을 거치게 된다. 이 과정은 기준 패턴과 입력 패턴을 서로 비교해 미리 정해 놓은 발생확률 값을 넘으면 승인하고, 그렇지 않으면 거절하는 것이다. 이를 응용하면 음성 자물쇠로 이용할 수 있다.

화자인식 시스템을 어떤 형태로 구현할 것인가의 관점에서 보면, 문맥 종속과 문맥 독립으로 나눌 수 있다. 문맥 종속은 정해진 말, 즉 미리 정해 놓은 단어나 문장 등을 뜻 한다. 일반적으로 음성은 음향학적, 음성학적으로 변화가 많기 때문에 문맥독립은 문맥종속에 비해 많은 훈련 데이터를 필요로 한다. 따라서 문맥종속 시스템의 경우에는 그 특성 때문에 DTW (Dynamic Time Wrapping) Algorithm을 사용해 성능이 좋은 반면 다른 사람이 정해진 말을 엿듣고 흉내 낼 우려가 있다. 문맥 독립은 미리 정한 말이 없이 아무 말이나 하는 것이다. 문맥 독립 시스템의 경우 HMM Algorithm을 많이 사용해 문맥 종속 시스템의 단점을 감소시킬 수 있다. 참고로 미국의 경우 1,000명의 다른 사람이 시험한 결과 1명 이하의 사람을 잘못 승인하고, 100번 발성해 1번 이하의 잘못된 거절(False Rejection)을 화자인식 시스템의 최소 규격으로 삼고 있다[13].

DTW 기법은 기준이 되는 음성신호의 패턴과 입력된 음성신호간의 유사도(distance)를 동적 프로그래밍(dynamic programming)을 이용해 구하는 방법이다.

본 연구에서는 Fig. 3 과 같은 화자인식시스템을 사용한다. 입력된 음성으로부터 추출된 특징파라미터들은 등록된 화자의 표준패턴 또는 모델과 비교되는데 이때 판단척도로서는 두 패턴간의 거리값(혹은 유사도값)이 많이 이용된다. 화자 검증에서는 위에서도 언급한 바와 같이 입력음성이 등록된 화자의 발

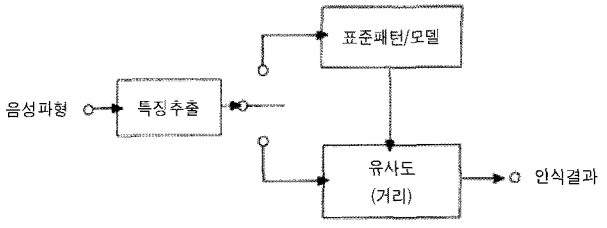


Fig. 3 Speaker recognition system

성패턴과 비교되어 두 패턴간의 거리가 특정 임계값 이하인 경우 승인되나 그 이상인 경우는 거절된다. 화자식별의 경우는 입력음성이 등록된 여러 화자의 발성패턴과의 거리를 측정하여 가장 작은 거리를 가진 등록화자가 선택된다.

3. 음성인식기반 명령제어

음성인식 자율주행 로봇에서 음성원격제어가 다른 인터페이스보다 뛰어난 점은 첫째는 편의성이고 둘째 병렬성이다. 음성으로 대화하며 키보드 치는 등의 동시적 처리가 가능하며, 정보를 사용하거나 입력하는데 자유로이 움직이면서 정보의 입출력이 가능하다는 장점이 있다. 셋째 자료입력의 고속화와 원거리 입력이 가능하다.

3.1 음성인식 프로세서

음성인식 프로세서 프로그램은 음성인식 Library를 이용하여 사용자가 쉽게 음성인식 결과를 확인 할 수 있도록 개발된 프로그램으로서 사용자 Application에 포함이 되어 있지 않고 윈도우 Tray 프로그램으로 동작된다.

이 프로그램이 동작하고 있는 동안 마이크로를 통하여 입력되는 정보는 사용자가 입력한 음성단어와 분석 작업을 통하여 대칭되는 단어를 사용자 Application에 매칭단어 및 인식률을 메시지로 전달하는 방법으로 구성되었다.

3.2 음성인식 알고리즘

① HMM(Hidden Markov model) 디코더[11,12]

HMM 모델은 연속 HMM과 이산 HMM으로 나뉘는데, 연속 HMM이 음성인식, 화자 인증 또는 인식 등에서 우수한 성능을 보인다. 연속 HMM은 연속 관측 확률밀도함수를 사용하는데 아래의 식(1), (2)로 표현된다.

$$b_i(x_t)=\sum_{m=1}^{N_m} g_m \Phi_{im}(x_t;\mu_{im},\sum_{im}),\ 1\le i\le N_s \tag{1}$$

$$\sum_{m=1}^{N_m} c_{im}=1,\ 1\le i\le N_s \tag{2}$$

$$c_{im}\ge 0,\ 1\le i\le N_s,\ 1\le m\le N_s$$

위 식에서 x_t는 관측벡터이며 Φ_{im}는 커널함수로 i번째 상태의 가중치가 c_{im}, 평균벡터 μ_{im}, 공분산 $\sum_{im}$과 N_m은 총 개수이다.

커널함수 Φ_{im}는 가우시안분포이며, HMM 모델은 위 관측모델을 포함하여 다음과 같이 표기된다.

$$\lambda=(A,B,\pi) \tag{3}$$

이때 A는 상태천이행렬, B는 방사행렬이고, π는 초기 확률벡터이다.

관측패턴 $\{X_t\}(t=1,2,...,T)$와 HMM이 $\lambda=(A,B,\pi)$로 이루어진 확률$P(X|\lambda)$를 구하기 위해 순방향 과정을 사용하면 일반적인 상태열은 $Q=[q_1,q_2,...,q_t,...,q_T]$로 나타낸다. 순방향 변수 $\alpha_t(i)$는 식(4)로 표현된다.

$$\alpha_t(i)=P(x_1,x_2,...,x_t,\ q_t=S_i|\lambda) \tag{4}$$

위 식은 모델 λ이 주어지고 시간 t에서 상태변수가 S_i일때 $t=1$에서 t까지 관측패턴의 확률이다. 다음 식은 $\alpha_t(i)$를 구하기 위한 반복적인 과정이다.

i) $\alpha_1(i)=\pi_i b_i(x_1),\ 1\le i\le N_s$ (5)

ii) $\alpha_{t+1}(j)=\left(\sum_{i=1}^{N_s}\alpha_t(i)a_{ij}\right)b_j(x_{t+1})$ (6)

iii) $\alpha_T(i)=P(x_1,x_2,...,x_t,...,x_T,q_T=S_i|\lambda)$ (7)

또한 $P(X|\lambda)$은 $\alpha_t(i)$의 전체 합이므로 다음과 같다.

$$P(X|\lambda)=\sum_{i=1}^{N_s}\alpha_T(i) \tag{8}$$

HMM 분류기는 출력값을 로그함수의 형태를 취하므로 $ll(X|\lambda_n)$으로 간략화 하며 화자인식에 대한 결정규칙은 $\arg Max_i ll(X|\lambda_i)$로 나타낸다.

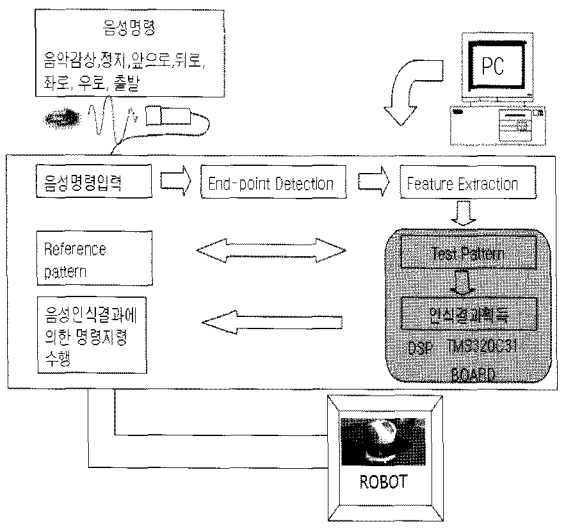


Fig. 4 Schematic diagram of voice command robot system

② 변형된 HMM 디코딩

모든 측정은 주변의 잡음 또는 관측 오차가 존재하므로, 정확한 측정을 불가능하게 만든다. 그러므로 HMM 디코딩 과정에 관측신뢰도를 사용하는 것은 충분히 타당한 접근 방법이다. 각각의 관측된 벡터에 대한 적당한 신뢰도 값을 가지고 있을 때, 이를 ρ_t라고 표현하자. ρ_1는 t번째 관측벡터의 신뢰도를 나타내는 멤버쉽 값으로 0과 1사이의 값을 갖는다. 따라서 본 연구에서 관측패턴은 $\{x_t, \rho_t\}$, $t=1,2,...,T$와 같으며, 모델은 HMM $\lambda=(A,B,\pi)$와 같다.

제안된 HMM 디코딩 방법은 기본적으로 식(5)에서 식(8)의 과정을 반복한다. 단, 식(6)에서 사용되는 관측확률$b_j(x_{t+1})$를 관측신뢰도 ρ_{t+1}에 의하여 가중함으로써 변형되게 된다. 따라서 본 연구에서 제안하는 변형된 디코딩(순방향과정)은 식(6)을 식(9)와 같이 변형함으로써 이루어진다.

$$\alpha_{t+1}(j)=\left(\sum_{i=1}^{N_s}\alpha_t(i)a_{ij}\right)\{b_j(x_{t+1})\rho_{t+1}\} \quad (9)$$

$1 \le t \le T-1,\ 1 \le j \le N_s$

3.3 음성인식 성능실험

Table 1에서 제어방법 및 조건은 PC기반 음성 S/W 명령어로 지령을 하였고 화자독립방식에 의해 음성전달을 하였다 그리고 무선통신 리모컨 제어방식 동시 적용하여 RF무선 통신 및 무선랜 통신가능하게 하였다. 인터페이스로는 RS232C를 사용하였으며 마이크로폰 송신장치를 이용하여 음성명령을 입력하도록 하였다. Table 2는 로봇의 시스템에 저장해놓은 음성단어이다.

Table 1 Specification of voice recognition module

Function	Specification
운영방식	- 하드웨어적인 음성인식 모듈 (리모콘식 제어) - 소프트웨어적 음성인식 모듈 (PC기반 음성인식제어)
음성전달 방 식	- 송수신 음원분석 후 무선통신 제어 명령 전달 - PC에서 음원 분석 후 무선 랜으로 제어명령
음성인식 알고리즘 (S/W, H/W)	- S/W적 음성인식 알고리즘 제공 - PC기반 화자독립방식 음성인식 (300단어이상 음성인식 기능)
제어/ 통신방식 (S/W, H/W)	- PC기반 원격제어방식(S/W방식) - 무선통신(RF방법) 원격제어 방식 - 마이크로폰기반 리모콘 무선원격 제어방식 - RS-232 인터페이스 - 윈도우 XP 인터페이스(S/W)

음성인식 실험은 실험의 신뢰성을 위하여 3명의 서로 다른 목소리를 대상으로 위의 조건에서 제시한 조건에서 명령어를 각각 100번씩 발성하게 하고, 5번의 발성 음은 학습용에 사용하고 나머지 음은 인식용으로 사용하여 실험 결과를 분석 평가하였다.

Table 2 Voice command words of experiment

단어ID	단어이름	단어ID	단어이름
1	로봇전진	9	감상
2	로봇후진	10	전진
3	로봇정지	11	후진
4	로봇우로	12	출발
5	로봇좌로	13	가속
6	좌회전	14	감속
7	우회전	15	멈춤
8	음악	16	시작

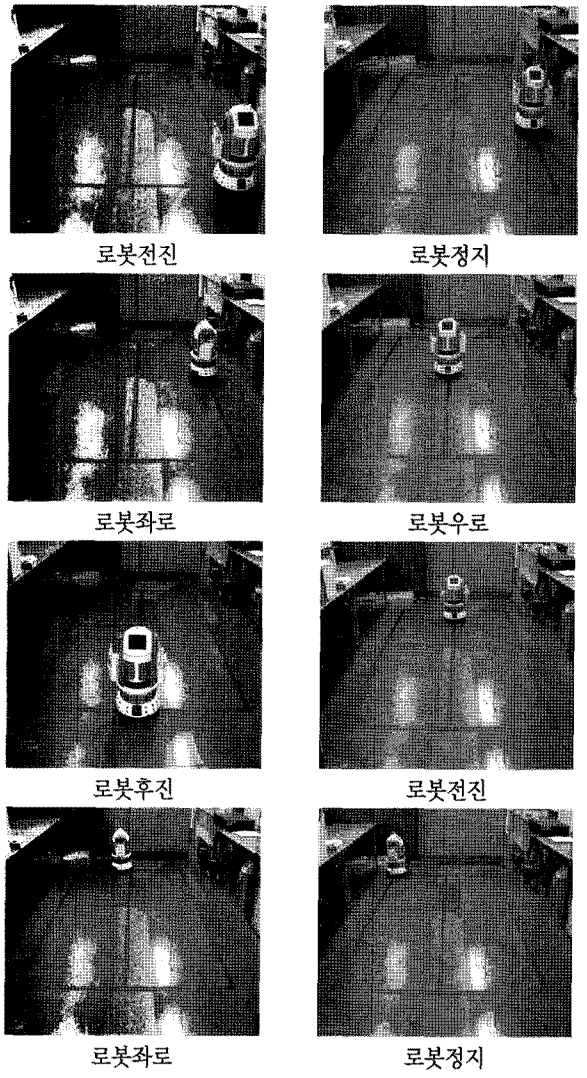
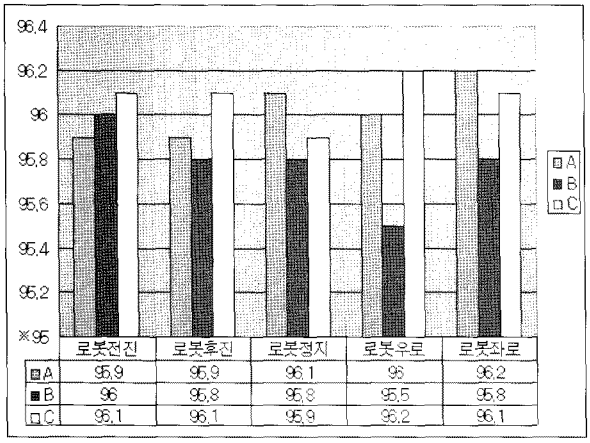


Fig. 5 Performance experiment scene of voice recognition

Table 3 Voice recognition experiment result



	로봇전진	로봇후진	로봇정지	로봇우로	로봇좌로
A	95.9	95.9	96.1	96	96.2
B	96	95.8	95.8	95.5	95.8
C	96.1	96.1	95.9	96.2	96.1

Table 3은 Table 2에 저장된 단어의 일부 단어에 대한 음성인식에 의한 로봇자율주행 시험의 인식률을 나타내고 있다. 또한 자동 음성명령 기준점의 인식률은 ※점인 95%로 설정하였다. A, B, C의 서로 다른 목소리로 각각의 명령어를 100번씩 발성했을때 나타나는 인식률로써 평균적으로 95% 이상의 높은 음성인식 결과가 나왔다. 하지만 주변 노이즈가 거의 없는 실험실에서 실험을 진행하였기 때문에 노이즈가 있는 경우에 향후 연구가 필요할 것으로 생각된다.

4. 결 론

본 연구에서는 음성인식에 의한 무인원격 모바일 로봇의 주행제어에 관한 연구를 수행하였다. 제안된 이동로봇의 음성인식 성능을 확인하기 위하여 여러번의 반복 성능 실험을 통해 예증하였다. 또한 본 연구에서 오프라인 무선원격제어 자율주행기술 및 자율주행 시뮬레이터의 개발로 보다 다양한 방법의 자율주행시스템을 음성인식에 의한 무인원격 제어기법을 이용한 모바일 로봇의 주행제어를 통하여 무인 공장자동화실현 및 원거리 실시간 무인 원격제어 성능을 확인하였다.

후 기

본 연구는 (경남대학교 로봇지능기술연구센터를 통한) 지식경제부/한국산업기술진흥원 융복합형로봇 전문인력양성사업의 지원으로 수행되었음

참 고 문 헌

(1) Rumelhart, D. E., and Mcclelland, J. L., 1987, Parallel Distributed Processing: Explorations in the Microstructure of Cognition, Vol. 1, MIT Press, Mass.

(2) Mamdani, E. H., 1974, "Application of fuzzy algorithms for control of simple dynamic plant," Proc. IEEE, Vol. 121, No. 12, pp. 1585~1588.

(3) Psalitis, D., Sideris, A., and Yamamura, A., 1988, "A Multilayered Neural Network Controller," IEEE Control Systems Magazine, April, pp.17~21.

(4) Horikawa, S., Furuhashi, T., Okuma, S., And Uchikawa, Y., 1991, "A Learning Fuzzy Controller Using a Neural Network," Trans. SICE, Vol. 27, No. 2, pp. 208~215.

(5) Horikawa, S., et al., 1990, "A Fuzzy Controller Using a Neural Network and Its Capability to Learn Expert's Control Rules," IIZUKA'90, pp. 103~106.

(6) Gauvain, J. L., Lamel, L., Adda, G., And Matrouf, D., 1996, "Developments in continuous speech dictation using the 1995 ARPA WSJ News task," Proc ICASSP '96. pp. 73~76.

(7) Woodlard, P. C., Legetter, C. H., Odell, J. J., Valtchev, V., and Young, S. J., 1995, "The Developments of the 1994 HTK Large Vocabulary Speech Recognition Rystem," Proc. ARPA Spoken Language Systems Workshop, Austin, Taxas.

(8) Gopalakrishnan, P., Bahi, L., and Mercer, R., 1995, "A Tree Search Strategy for Large Vocabulary Continuous Speech Recognition," Proc ICASSP '95, pp 572~575.

(9) Yu, H. J., Kim, H., Hong, J. M., Kim, M. S., and Lee, J. S., 2000, "Large Vocaburary Korean Continuous Speech Recognition using a One-Pass Algorithm," Proc. ICSLP2000. Beijing, China.

(10) Rabiner, L. R., and Juang. B. H., 1986, " An Introduction to Hidden Markov Models," IEEE ASSP MAGAZINE.

(11) Rabiner, L. R., Levinson, S. E., and Sondhi, M. M., 1983, "On the Application of Vector Quantization and Hidden Markov Models to Speaker-independent, Isolated Word Recognition," Bell System Technical Journal, Vol. 62, No. 4.

(12) Myers, C. S., and Rabiner, R. R., 1981, "Connected Digit Recognition Using a Level-Building DTW Algorithm," IEEE Trans. Acoust Speech, Signal Processing, Vol. ASSP-29, pp. 351~363

(13) Pallett, D. S., 1991, "DARPA Resource Management and ATIS Bench Mark Test Poster Session," Proceedings of the DARPA speech and Natural language Workshop, pp. 49~58.