

HMM을 기반으로 한 자율이동로봇의 음성명령 인식시스템의 개발

Development of Autonomous Mobile Robot with Speech Teaching Command Recognition System Based on Hidden Markov Model

조 현 수, 박 민 규, 이 현 정, 이 민 철*

(Hyeon Soo Cho, Min Gyu Park, Hyun-Jeong Lee, and Min Cheol Lee)

Abstract : Generally, a mobile robot is moved by original input programs. However, it is very hard for a non-expert to change the program generating the moving path of a mobile robot, because he doesn't know almost the teaching command and operating method for driving the robot. Therefore, the teaching method with speech command for a handicapped person without hands or a non-expert without an expert knowledge to generate the path is required gradually. In this study, for easily teaching the moving path of the autonomous mobile robot, the autonomous mobile robot with the function of speech recognition is developed. The use of human voice as the teaching method provides more convenient user-interface for mobile robot. To implement the teaching function, the designed robot system is composed of three separated control modules, which are speech preprocessing module, DC servo motor control module, and main control module. In this study, we design and implement a speaker dependent isolated word recognition system for creating moving path of an autonomous mobile robot in the unknown environment. The system uses word-level Hidden Markov Models(HMM) for designated command vocabularies to control a mobile robot, and it has postprocessing by neural network according to the condition based on confidence score. As the spectral analysis method, we use a filter-bank analysis model to extract of features of the voice. The proposed word recognition system is tested using 33 Korean words for control of the mobile robot navigation, and we also evaluate the performance of navigation of a mobile robot using only voice command.

Keywords : mobile robot, speech recognition, Hidden Markov Model, neural network

I. 서론

최근 산업계의 생산 체제가 자동화, 무인화 되어감에 따라서 건물 내에서의 제품 운반이나, 건물 외부로의 반출 등을 담당할 자율이동로봇(autonomous mobile robot)의 필요성이 높아지고 있으며, 그 응용에 관한 연구도 활발히 진행되고 있다[1,2].

음성은 인간의 가장 자연스러운 통신방법으로, 인간과 기계사이의 자연스러운 통신을 위해 음성인식에 관한 연구가 꾸준히 진행되어 왔으며 수십 단어 정도의 어휘를 대상으로 불특정 화자의 음성을 인식하거나, 화자적응을 통해 수만 단어를 인식할 수 있는 시스템들이 상용화되고 있다[3-5]. 지금까지의 음성인식 방법은 입력된 음성신호의 스펙트럼을 분석하여 학습과정을 통해 미리 준비된 기준 패턴과 대조시켜 가장 유사한 것을 찾아내는 패턴정합 방법이 주로 이용되는데, 패턴정합방법에는 DTW(Dynamic Time Warping), HMM(Hidden Markov Model), 신경회로망(neural network) 등이 있다[6-12]. DTW방식은 동적 프로그래밍에

의한 음성 패턴을 비선형 시간 정렬을 통하여 기준패턴과 인식하고자 하는 패턴을 비교하는 방식으로 소규모 단위의 음성 인식에는 좋은 성능을 보이지만, 어휘의 수가 증가하는 대용량 연속 음성인 경우에는 증가하는 어휘 수만큼 기준 패턴을 구성해야 하는 단점으로 인해 부적합하다[7,8]. HMM은 확률함수를 이용하여 음성신호의 시간에 따른 특징 변화를 통계적으로 모델링하는 것으로서 학습시에 많은 정보와 계산량을 요구하지만 인식시에는 상대적으로 계산량과 메모리 소요량이 적고 높은 인식 성능을 갖고 있어 가장 널리 사용되고 있다[6,9,10].

본 연구에서는 로봇에 음성에 의한 이동경로 및 이동거리, 방향의 교시기능을 부여하여 임의의 환경에서 경로를 생성하여 주행이 가능한 자율이동로봇을 개발하고자 한다.

로봇의 교시를 위한 음성인식시스템은 이산(discrete) HMM을 기반으로 하여 화자종속 고립단어 단위의 시스템으로 구축하였다. 그리고 인식을 향상시키기 위한 방안으로 HMM에 의한 인식을 수행한 후 그 신뢰도를 판단하여 신뢰도가 낮은 경우 신경회로망에 의한 재인식을 후처리과정에 도입하였다. 음성인식시스템을 구현함에 있어서 하드웨어적인 메모리 용량과 처리속도 등이 제한된 조건하에서 실시간 음성인식이 가능한 범위내에서 데이터 양과 계산량을 고려하였다. 또한 음성인식시스템을 실제 이동로봇에 적용하여 음성교시에 의한 주행 및 임의의 환경에서의 경로 생성 실험을 수행하여 성능을 확인하였다.

* 책임저자(Corresponding Author)

논문접수 : 2007. 5. 15., 채택확정 : 2007. 6. 25.

조현수, 이현정 : 부산대학교 지능기계공학과
(hscho@kia.co.kr/lhjeong@pusan.ac.kr)

박민규 : 부산대학교 기계기술연구소(pmg@pusan.ac.kr)

이민철 : 부산대학교 기계공학부(mclee@pusan.ac.kr)

※ 이 논문은 부산대학교 자유과제 학술연구비(2년)에 의하여 연구되었음.

II. 로봇의 주행경로 생성을 위한 음성인식시스템

본 연구의 고립단어 음성인식 시스템의 구성은 그림 1에 나타나 있으며, 크게 음성신호의 전처리 과정, 음성특징 추출 과정, 패턴인식 과정, 후처리 과정으로 나눌 수 있다.

1. 음성신호의 전처리

음성신호의 획득 및 필터링을 위한 음성 전처리부에는 12비트의 AD 변환기가 장착되어 있으며, 신호처리 과정은 그림 2와 같다. 60Hz~15kHz영역의 주파수 응답 특성을 가지는 마이크로폰(micro phone)을 통해 입력된 후 약 1000배 증폭된 음성신호는 음성 주파수 대역폭 신호만을 추출하기 위한 10Hz 고역통과필터와 8kHz 저역통과필터를 거치게 된다. 필터링을 거친 신호는 AD변환되며 이때의 샘플링 주파수는 6kHz로 설정하였다.

2. 음성구간의 검출

음성신호를 $s(n)$, 창 함수를 $w(n)$, 프레임 길이가 N 일 때, 단구간 진폭크기(short time magnitude)는 (1)과 같다.

$$E(m) = \sum_{n=0}^{N/2} |s(n)|w(m-n) \quad (1)$$

먼저 반복 실험을 통해 음성 구간이라 판단할 수 있는 단구간 진폭 크기의 임계값을 결정한다. 만약 단구간 진폭 크기가 임계값보다 크면 음성구간이라 간주한다. 실험으로 얻은 단구간 진폭 크기의 임계값은 약 1.7로 결정되었다.

3. 음성특징추출

본 연구에서 음성특징 파라미터로 사용한 MFCC(Mel Frequency Cepstral Coefficient)는 사람이 주관적으로 인지하는 주파수 특성을 반영하는 멜 스케일(mel scale)로 근사화한 임계대역(critical band) 필터뱅크를 씌워 추출하는데, 그림 3은 그 과정을 설명한다. 6kHz로 샘플링된 음성신호는 전달함수가 $1 - 0.97z^{-1}$ 인 1차 디지털 필터로 preemphasis를 행한다. Preemphasis는 음성신호의 저주파 성분을 약화시키고 고주파 성분을 강조하는데 이는 저주파 대역의 잡음을 감쇄시키기도 하고 동적 영역을 줄이는 역할을 한다. 그리고 길이가 약 42 msec(256 point)이고 10msec(64point)씩 중첩되는 프레임 단위로 나눈 다음, 각각의 프레임에 해밍창(hamming window)을 씌운다. 매 프레임을 FFT하여 주파수 영역에서 구한 파워스펙트럼(power spectrum)에 멜 스케일된 삼각 필터링을 수행하여 필터뱅크 출력을 얻는다. 표 1은 멜 스케일에 의해 임계적으로 결정된 필터뱅크의 주파수 대역을 나타낸다. 그리고 각각의 필터뱅크 출력의 log값을 입력으로 하여 Discrete Cosine Transform(DCT)을 수행하고, liftering을 수행하면 인식에 사용될 MFCC를 추출하게 된다. (2)는 본 연구에서 음성특징 추출에 사용된 DCT 관계식이다. (2)에서 $x[i]$ 는 필터뱅크의 출력 값이며, N 은 필터뱅크의 개수, 그리고 k 는 켈스트럼(cepstrum) 차수의 인덱스(index)이다.

$$MFCC_k = \sqrt{\frac{2}{N} \sum_{i=1}^N \left\{ \log(x[i]) \cos\left(\frac{2\pi k}{N}(i-0.5)\right) \right\}^2} \quad (2)$$

MFCC차수는 12차, 필터뱅크의 수는 26개로 설정하였다. 그림 4는 본 연구에서 사용한 멜 스케일된 삼각 필터뱅크

의 구성을 나타내며, 그림 5와 6은 위의 과정을 거쳐 추출된 음성 “앞으로”, “정지”의 MFCC 특성을 나타낸다.

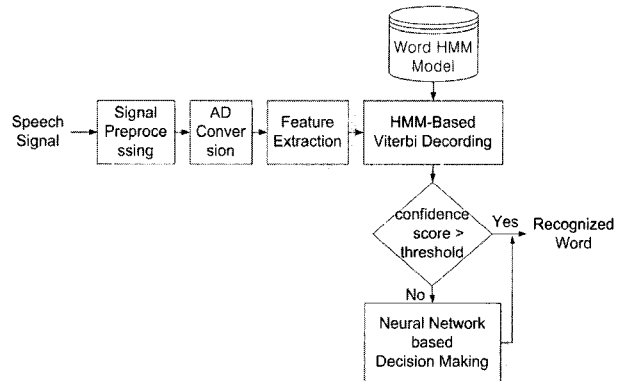


그림 1. Baseline 음성인식시스템.

Fig. 1. Baseline speech recognition system.

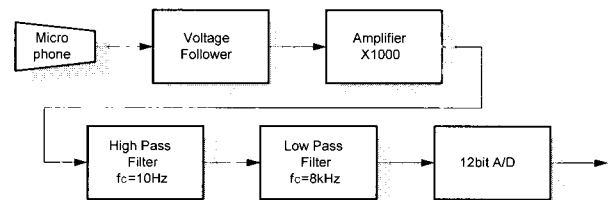


그림 2. 음성신호의 전처리 과정.

Fig. 2. Preprocessing of the speech signal.

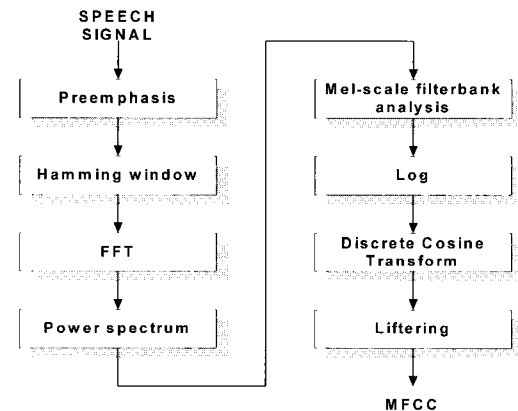


그림 3. MFCC 추출 과정.

Fig. 3. Process of MFCC extraction.

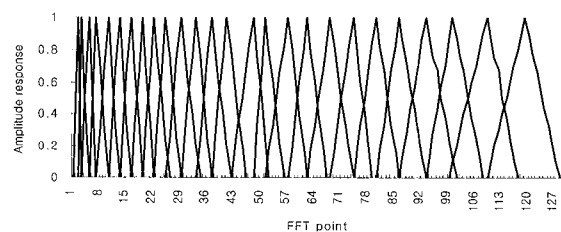


그림 4. 멜 스케일 삼각 필터뱅크.

Fig. 4. Mel-scale triangular filter bank.

표 1. Critical filter-bank의 구성.

Table 1. Composition of critical filter-bank.

Critical band number	Center frequency F_{Hz}	lower cutoff frequency F_{Hz}	upper cutoff frequency F_{Hz}
1	27	-	54
2	83	54	112
3	143	112	174
4	205	174	236
5	271	236	306
6	341	306	376
7	415	376	454
8	492	454	530
9	574	530	618
10	660	618	702
11	751	702	800
12	847	800	894
13	948	894	1002
14	1056	1002	1110
15	1168	1110	1226
16	1287	1226	1348
17	1412	1348	1476
18	1545	1476	1614
19	1684	1614	1754
20	1831	1754	1908
21	1987	1908	2066
22	2150	2066	2234
23	2323	2234	2412
24	2505	2412	2598
25	2697	2598	2796
26	2900	2796	3000

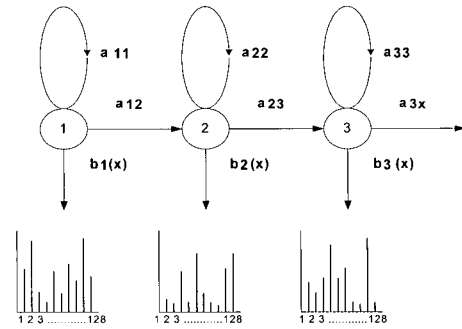


그림 7. Left-to-right HMM 구조.

Fig. 7. Structure of a left-to-right HMM.

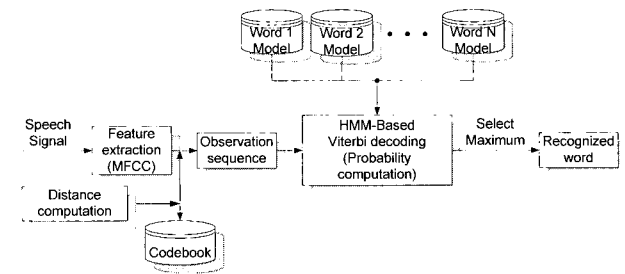


그림 8. 인식을 위한 시스템 구성.

Fig. 8. Configuration of system for recognition.

여기서 각 프레임은 MFCC 차수에 해당하는 벡터 값을 가지고 있다.

4. 벡터양자화(vector quantization)

음성인식에서 벡터양자화(VQ)는 원신호에서 추출된 음성특징 벡터들을 대표할 수 있는 한정된 코드북(codebook)을 생성하여 음성특징을 유한개의 벡터로 표현하는 것을 말한다. 본 연구에서는 벡터양자화에 필요한 초기 코드북을 생성하기 위해 반복 클러스터링 알고리즘(iterative clustering algorithm)의 하나로 분할 방법을 사용하여 코드북의 크기를 증가시켜 새로운 코드북을 생성하는 개선된 K-평균 클러스터링 알고리즘(modified K-means clustering algorithm)을 이용하였다[13,14]. 코드북 크기는 128로 설정하였으며 생성된 코드북은 각 코드워드마다 MFCC 차원별로 분할된 공간의 중심 값들을 가지고 있다.

5. HMM기반 음성인식시스템

본 연구에서는 훈련과정에서 패턴인식을 위한 음성의 기본 단위(단어, 반음절, 음소, 변이음 등)에 해당하는 패턴들의 통계적인 정보를 확률모델 형태로 저장하고, 인식과정에서 미지의 입력 패턴이 들어오면 각각의 모델에서 이 패턴이 나올 수 있는 확률을 계산함으로써 이 패턴에 가장 적합한 음성단위를 찾아내는 방법인 이산(discrete) HMM을 기반으로 하여 화자종속 고립단어 단위의 시스템을 구축하였다. 사용한 HMM의 구조는 음성신호와 같이 시간에 따른 상태 변화가 비가역적인 데이터를 모델링하는데 적합한 left-to-right 모델이며, 자신 및 다음 상태로만 천이를 허용하도록 하였다. HMM의 구조의 예는 그림 7과 같다. 예로 든 모델은 3개의 상태를 갖는 left-to-right HMM이며, 각 상

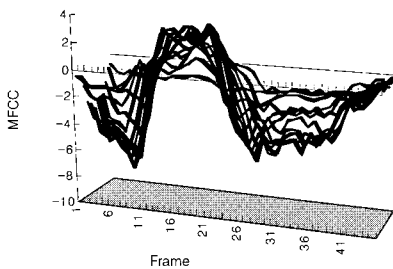


그림 5. 음성 “앞으로”의 MFCC.

Fig. 5. MFCC of voice “Apuro”.

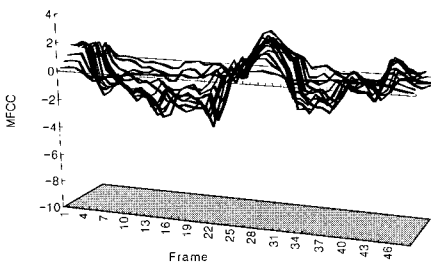


그림 6. 음성 “정지”의 MFCC.

Fig. 6. MFCC of voice “Jungji”.

태에서 관측벡터가 관측될 확률은 이산확률밀도로 주어진다. 일반적으로 대용량의 단어를 인식할 때에는 음소나 변이음 단위를 사용하는 것이 바람직하나 본 연구에서는 소량의 33개의 음성 명령어를 판별하기 때문에 단어 단위를 채택하였다. 그런데 이러한 경우 단어별로 할당되는 상태수는 음소수에 비해하도록 주는 것이 가장 알맞다고 알려져 있다[6,15]. 그러나 단어 내 음소수가 많다고 해서 반드시 프레임의 갯수도 정비례하게 길어지지는 않음을 실험적으로 확인하였다. 미리 저장된 음성데이터의 각 단어 당 평균 분석 프레임의 갯수를 근거로 하여 기준 상태수를 결정한다음 프레임 갯수에 비해하도록 대략 4프레임 당 1개의 상태수를 할당하였다. 그림 8은 인식을 위한 시스템의 구성을

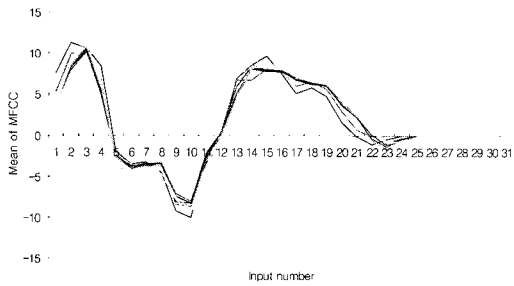


그림 9. 음성 “앞으로”의 MFCC 평균.
Fig. 9. Mean of MFCC about voice “Apuro”.

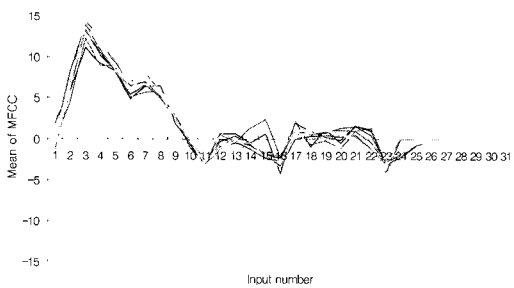


그림 10. 음성 “정지”의 MFCC 평균.
Fig. 10. Mean of MFCC about voice “Jungji”.

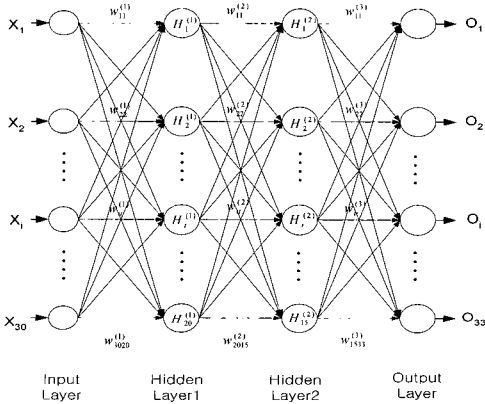


그림 11. 신경회로망의 구조.
Fig. 11. Structure of neural network.

표 2. 인식음성에 대한 신경회로망의 출력 패턴.
Table2. Output pattern of neural network.

출력뉴런 입력음성	1	2	3	4	5	6	7	...	15	16	...	32	33
“앞으로”	1	0	0	0	0	0	0	0	0	0	0	0	0
“뒤로”	0	1	0	0	0	0	0	0	0	0	0	0	0
“좌로”	0	0	1	0	0	0	0	0	0	0	0	0	0
“우로”	0	0	0	1	0	0	0	0	0	0	0	0	0
“정지”	0	0	0	0	1	0	0	0	0	0	0	0	0
“1m”	0	0	0	0	0	1	0	0	0	0	0	0	0
“2m”	0	0	0	0	0	0	1	0	0	0	0	0	0
⋮	0	0	0	0	0	0	0	1	0	0	0	0	0
“10도”	0	0	0	0	0	0	0	0	1	0	0	0	0
“20도”	0	0	0	0	0	0	0	0	0	1	0	0	0
⋮	0	0	0	0	0	0	0	0	0	0	1	0	0
“170도”	0	0	0	0	0	0	0	0	0	0	0	1	0
“180도”	0	0	0	0	0	0	0	0	0	0	0	0	1

나타낸다. 그 과정을 살펴보면 음성특징추출 과정에서 입력 신호의 음성 특징이 검출되면 미리 만들어진 코드북의 각 코드워드들과 거리를 계산하여 가장 작은 값을 가지는 코드북의 인덱스열 형태로 된 관측열을 생성한다. 생성된 관측열을 기반으로 한 모든 단어의 HMM에 대한 확률 값을 Viterbi 디코딩에 의해 계산하여 인식이 이루어지게 된다[16].

6. 신경회로망에 의한 후처리 방안

HMM에 의한 인식실험을 통해 오인식이 발생하는 경우 인식된 단어 후보들 중에서 상위 두 후보의 누적 확률 값의 차가 미소하다는 것을 확인할 수 있었다. 따라서 본 연구에서는 HMM에 의한 인식을 먼저 수행한 후 누적 확률 값이 가장 큰 후보와 두 번째로 큰 후보의 누적 확률 값의 차를 구하여, 이 값이 미리 정해둔 임계값 이하이면 신경회로망에 의한 재인식을 수행하여 인식결과를 출력하도록 하였다. 반복 실험을 통해 임계값을 10으로 설정하였다. 여기서 패턴인식을 위한 신경회로망의 입력으로는 음성특징 추출과정에서 얻어진 각 분석 프레임마다 각 차원별 MFCC 값들을 합한 후 평균값을 구하고, 다시 전체 분석 프레임에 대한 평균값을 구하여 사용하였다. 그림 9와 10은 신경회로망의 학습 시 입력으로 작용하는 단어 당 14개씩의 “앞으로”, “정지”의 각 분석프레임의 MFCC 평균값을 나타낸다. 본 연구에서 사용한 신경회로망은 비선형 문제를 해결하는데 우수한 특성을 지닌 Rumelhart가 제안한 오차 역전파 (error back propagation) 다층 신경회로망을 사용하였다[17, 18]. 신경회로망의 구조는 그림 11과 같으며, 은닉층을 구성하는 뉴런의 갯수는 우수한 출력을 갖도록 시행착오적으로 설정하였다. 표 2는 로봇의 명령교시를 위한 음성 및 인식음성에 대응되는 신경회로망의 출력 패턴을 나타낸다. 각 출력 뉴런은 대응하는 인식음성의 입력에 대하여 1이 출력되고 그 외의 경우에는 0이 출력되도록 학습하였다. 뉴런의 출력 함수는 역치함수와 선형함수의 특성을 가지며 미분이 가능한 S형태의 함수인 시그모이드(sigmoid) 함수를 사용하였다. 신경회로망의 출력에 대한 음성의 판별 조건은 (3)과 같다.

$$(O_i / \sum_{i=1}^N O_i) \times 100\% \geq 80\% \quad (3)$$

O_i 는 신경회로망의 i 번째 뉴런의 출력이며, N 은 인식음성의 수를 나타낸다. 즉, 특정 뉴런의 출력이 전체 출력의 80% 이상이면 그 출력 뉴런에 해당하는 음성으로 인식하게 된다. 판별조건을 이와 같이 설정한 것은 신경회로망의 경우 HMM기반 음성인식시스템의 인식을 향상을 위한 후처리 방안에 해당하므로 특정 음성에 해당하는 뉴런의 출력이 그 외의 뉴런의 출력에 비해 현저히 높을 때만 음성인식이 이루어지도록 하여 안정성을 고려한 것이다.

III. 음성교시에 의한 경로 주행실험 및 고찰

1. 자율이동로봇의 구성

본 연구에서 개발한 이동 로봇의 외관과 구조가 그림 12에 보여진다. 상판에는 로봇 제어기와 음성인식을 위한 마이크, 자이로 센서, 초음파 센서 그리고 모터 구동용 배터리가 장착되며 하판은 구동부로서 2개의 DC 기어드 모터 (geared motor)와 2개의 엔코더가 장착되어 있다. 그리고 음성인식 및 제어기 역할을 하는 펜티엄급 산업용 PC가 상판에 탑재되어 있다. 상판에 장착된 초음파 센서는 모두 5조로 구성하였으며 로봇 전면의 1조를 중심으로 좌우 양쪽에 각각 20°와 45°의 각도로 4조를 배치하였다. 자이로 센서는 제어기 기판 지지대 옆면에 부착되어 있다.

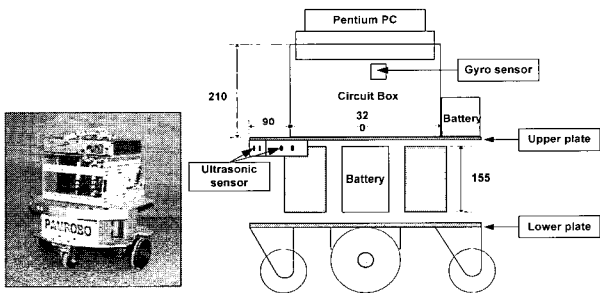


그림 12. 자율이동로봇의 측면도.
Fig. 12. Side view of mobile robot.

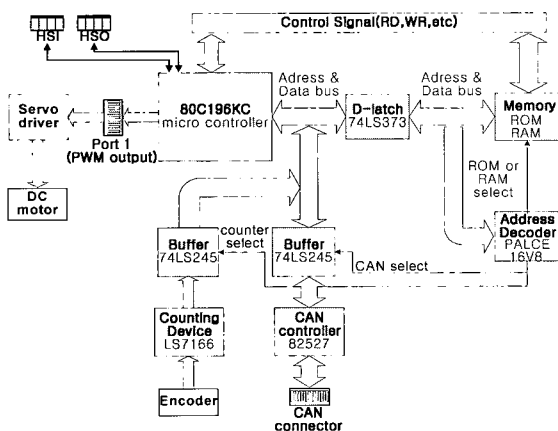


그림 13. 서보 모터 제어기의 구성.
Fig. 13. Block diagram of servo motor control module.

2. 자율이동로봇의 제어기 구성 및 통신

개발된 이동로봇의 제어 시스템은 로봇의 구동을 위한 서보 모터 제어부, AD보드가 장착되어 음성신호의 획득 및 필터링을 수행한 후 이를 주제어부로 넘겨주는 음성 전처리부, 음성인식의 수행 및 주제어부 역할을 하는 펜티엄급 PC, 장애물 회피용 환경인식을 위한 초음파 센서부를 포함하여 총 4개 모듈로 구성되어 있다.

서보 모터 제어부는 80C196KC 마이크로 컨트롤러를 사용하는데, 원하는 속도로 모터를 구동하고 엔코더에 의한 속도 정보를 검출하여 주제어부로 전달하는 역할을 하며 그 구성도는 그림 13과 같다. 음성 전처리부는 마이크에 의해 음성 교시명령에 대한 음성신호를 획득하고 아날로그적인 필터링을 거친 후, 그 결과를 주제어부인 PC로 전달하는 역할을 한다. 음성 전처리부의 구성도는 그림 14와 같다. 메인 프로세서(main processor)로 TI(Texas Instrument)사의 DSP(TMS320C31)를 사용하였으며, 12bit AD와 RAM 등이 장착되어 있다. 주제어기는 음성인식을 수행하고 인식음성에 해당하는 모터속도를 서보 모터 제어부로 전송한다. 자이로 센서의 신호를 입력받기 위한 AD 컨버터(converter)가 포함되어 있으며, 자이로 센서는 -80~80°/s의 각속도를 측정할 수 있는 MURATA사의 ENV-05DB를 사용하였다. 그리고 필터링된 음성신호의 획득을 위해 음성 전처리부와 주제어부 사이엔 EPP(Enhanced Parallel Protocol) mode를 이용하였고, 음성인식을 수행한 후 인식된 결과를 서보 모터 제어부로 전송하기 위해 RS232C 시리얼통신을 사용하였다. 그림 15는 각 제어기 모듈간 신호흐름을 나타낸다.

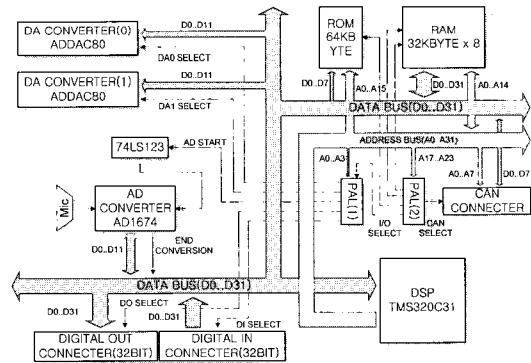


그림 14. 음성전처리부의 구성도.
Fig. 14. Block diagram of speech preprocessing module.

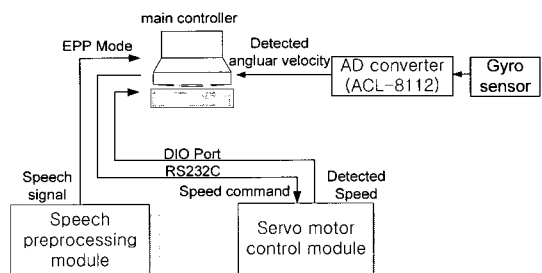


그림 15. 제어기간의 신호흐름도.
Fig. 15. Signal flow among control modules.

3. 음성 데이터베이스

음성 데이터베이스는 임의 환경에서 이동로봇의 경로생성을 위한 33개의 고립단어로 구성되어있고 음성구간의 검출이 완료된 것이다. 고립단어의 목록은 표 3과 같으며, 이동로봇의 이동경로 및 거리, 방향교시를 위한 단어들로 구성되어 있다. 음성신호의 획득은 비교적 소음이 적은 실험실 환경에서 이루어졌다. 음성 데이터는 남성 화자 1명이 각 인식단어 당 100번씩 발성하였다. 표 4는 이동로봇에 적용한 인식단어 목록이다.

4. 인식시스템의 구성

음성인식 시스템은 로봇의 경로생성을 위한 33개 각각의 단어모델로 구성하였다. 신경회로망을 통한 인식을 수행하기 위해 표본 데이터를 각 인식 음성당 14개씩을 미리 획득하여 학습오차가 0.03 이하가 될 때까지 학습시켰다. 신경회로망의 경우 HMM기반 음성인식 시스템의 인식을 향상 위한 후처리 방안이 해당하므로 특정 뉴런의 출력이 전체 출력의 80% 이상이 되는 확실한 경우만 출력하도록 하여 안정성을 꾀하였다. 즉 HMM에 의한 인식결과의 신뢰도가 낮고, 신경회로망에 의한 인식결과인 특정뉴런의 출력이 전체 출력에 비해 현저히 높지 않으면 인식결과를 출력하지 않도록 하였다. 또한 검출된 분석 프레임의 수가 인식대상 단어의 평균 분석 프레임수보다 너무 적거나 많을 경우, 음성구간의 검출이 제대로 되지 않았거나 주위 소음의 영향이 크다고 판단하여 음성의 재입력을 요구하도록 하였다.

5. 로봇의 구동을 위한 동작설정

이동로봇의 구동을 통한 성능평가를 위해 14개 인식단어 각각에 해당하는 좌우 서보 모터의 속도를 미리 설정하였다. 로봇은 전진 또는 후진에 의해 위치 이동이 이루어지며, 방향전환은 제자리 회전에 의해서만 이루어지도록 하였다. 그리고 이동거리 제어 시 해당음성이 출력된 시점을 기준으로 해당 거리만큼 이동한 후 정지하도록 하였고, 이동 방향은 이동로봇 상판에 부착된 자이로 센서에 의해 검출된 각속도를 이용하여 해당각도만큼 회전하도록 하였다.

표 3. 인식대상 어휘목록.

Table 3. List of isolated word.

	교시단어
이동 경로	앞으로, 뒤로, 좌로, 우로, 정지
이동 거리	1m, 2m, 3m, 4m, 5m, 6m, 7m, 8m, 9m, 10m
이동 방향	10도, 20도, 30도, 40도, 50도, 60도, 70도, 80도, 90도, 100도, 110도, 120도, 130도, 140도, 150도, 160도, 170도, 180도

표 4. 이동로봇 적용 어휘목록.

Table 4. List of word for a mobile robot.

	교시단어
이동 경로	앞으로, 뒤로, 좌로, 우로, 정지
이동 거리	1m, 2m, 3m
이동 방향	30도, 60도, 90도, 120도, 150도, 180도

6. 실험결과 및 고찰

표 3과 표 4에 나타난 전체 인식단어 및 로봇적용 단어를 대상으로 화자중속 형태의 인식실험을 수행하였다. 그리고 로봇적용 단어를 실제 이동로봇에 적용하여 음성인식을 통한 경로생성 및 주행실험을 수행하였다. 비교적 조용한 실험실 환경에서 각 단어 당 20번씩 발성하여 전체 인식단어 및 로봇적용 단어에 대해 실시간으로 인식률을 측정하였다. 인식을 측정시 상위 두 후보단어의 확률값 차의 임계값을 10으로 설정했는데, 후처리과정으로 넘어가는 경우는 약 8% 정도였다. 그리고 후처리과정인 신경회로망에 의한 재인식과정에서 특정뉴런의 출력이 전체 출력의 80%를 넘지 않아 음성의 재입력을 요구하는 경우는 인식을 측정에서 제외하였다. 인식결과는 그림 16과 17에 나타난 바와 같이 전체 인식단어의 경우 후처리과정을 도입하지 않았을 때 91.1%, 후처리과정을 도입한 경우는 91.9%, 그리고 로봇적용 단어의 경우 후처리과정을 도입하지 않았을 때 94.2%, 후처리과정을 도입한 경우는 95.7%의 인식률을 보였다. 인식결과에서 일부 단어에 대해 후처리과정을 거친 후 인식률이 낮아지는 현상도 보였으나 전체적으로 인식률이 향상되었음을 알 수 있었다. 그리고 주로 숫자음 관련단어에서 인식률이 저조하였는데, 이는 숫자음의 경우 음성특징이 유사한 경우가 많아 오인식 확률도 높았기 때문으로 판단된다.

비교적 소음이 적고 굴곡이 없는 건물내의 복도에서 로봇적용 단어에 대해 로봇의 주행 및 경로생성 실험을 수행하였다. 먼저 로봇의 주행특성을 알아보기 위해 이동거리 및 방향교시에 따른 로봇의 동작특성을 확인하였고, 2개의 경유점과 목표점을 가진 환경에서 음성교시를 통해 로봇이 경유점을 거쳐 목표점까지 주행하는 실험을 행하였다. 음성

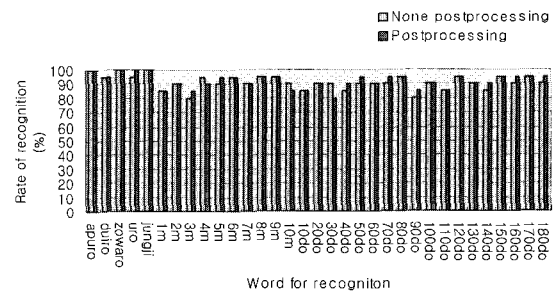


그림 16. 전체 인식단어의 인식률.

Fig. 16. Rate of recognition for whole word.

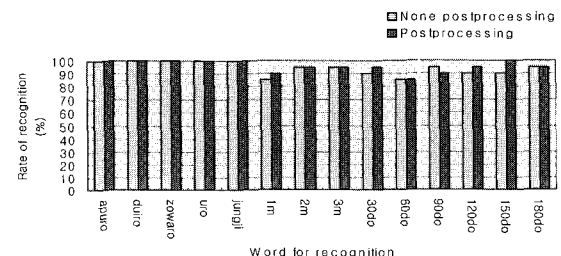


그림 17. 로봇적용 단어의 인식률.

Fig. 17. Rate of recognition of word for a mobile robot.



그림 18. 이동거리 및 방향교시에 따른 로봇의 동작실험.
 Fig. 18. Experiment of action of a mobile robot according to moving distance and direction of teaching.

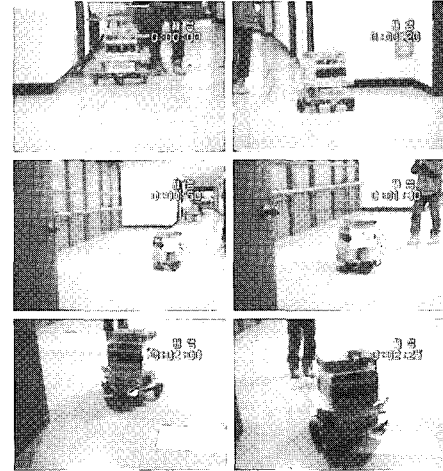


그림 20. 경로생성을 위한 로봇의 주행실험.
 Fig. 20. Experiment of traveling for generating the moving path of a mobile robot.

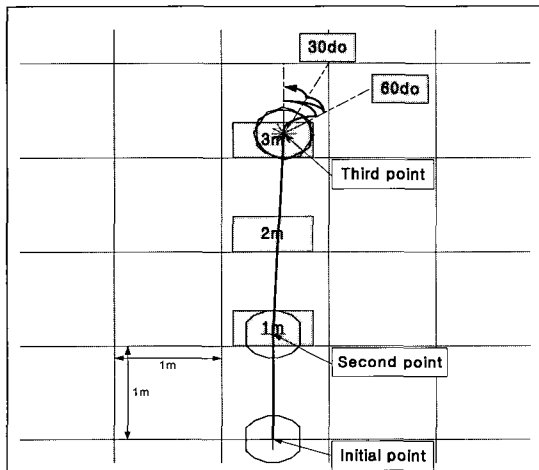


그림 19. 이동거리 및 방향교시에 따른 로봇의 동작.
 Fig. 19. Action of mobile robot according to moving distance and direction of teaching.

신호는 실험자가 긴 유선으로 로봇 제어기에 연결된 마이크를 들고 로봇과 함께 이동하며 전달하였다. 그림 18과 20은 실험결과를 나타낸 것이며 그림 19와 21의 로봇 이동 경로는 실제 실험결과를 비디오 촬영한 후 개략적으로 도시한 것이다. 그림 19에서 로봇은 초기위치에서 “1m”라는 명령을 인식한 후 1m 지점에 해당하는 두 번째 위치로 이동한다. 그리고 다시 “2m”라는 명령을 내려 3m 지점에 해당하는 세 번째 위치로 이동시킨 후 “30도”와 “60도”에 해당하는 명령을 내려 해당각도만큼 제자리에서 회전하도록 하였다. 주행실험 결과 로봇은 만족할만한 오차범위 내에서 동작함을 확인할 수 있었다. 그리고 경로생성 실험시 로봇의 주행에 따라 상황을 판단하여 로봇적용 어휘목록의 단어를 적절히 선정하고 이를 음성으로 교시하였다. 교시 명령은 단어를 합성하여 연속단어 형태로 교시하지는 않았고 한 단어씩 교시하였다. 그림 21에서 로봇은 주어진 음성명

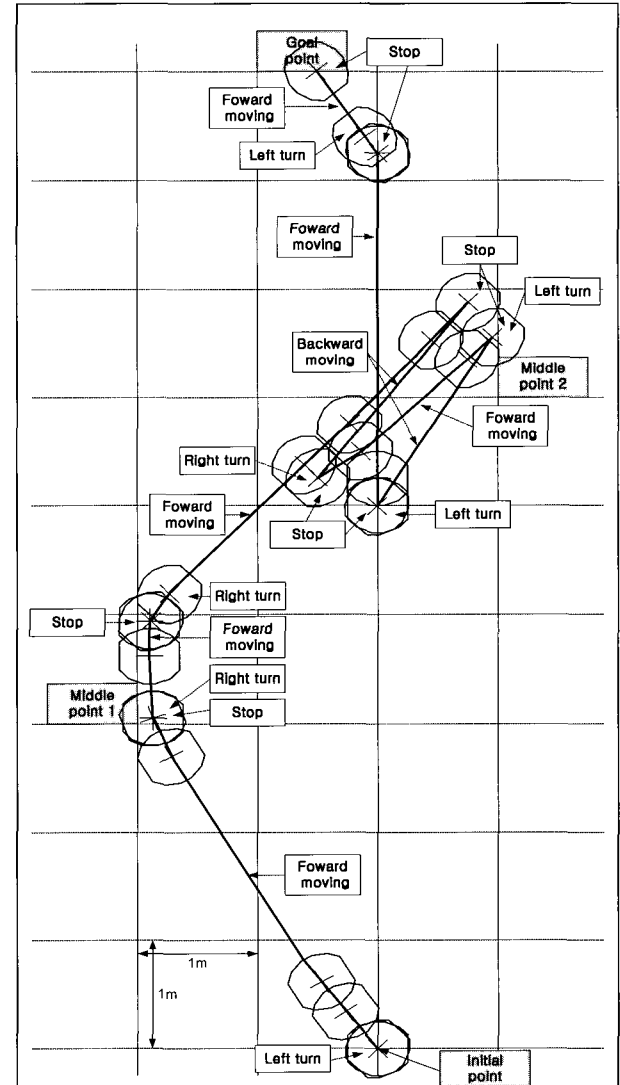


그림 21. 로봇의 주행 경로.
 Fig. 21. Moving path of a mobile robot.

령에 따라 전진 및 후진, 정지 그리고 해당 각도만큼 회전하여 주어진 임의의 환경에서 음성의 교시만으로 안정한 경로를 생성하여 목표점에 도달하였다. 다만 하드웨어적인 처리속도의 한계 등으로 인한 음성인식 시 소요되는 시간이 다소 길어 좁은 공간에서의 로봇의 정밀 위치제어에 문제점이 있었다. 또한 로봇 구동 시 발생하는 모터 구동음과 같은 소음의 영향으로 인해 다소 인식률이 떨어짐을 알 수 있었다.

IV. 결론

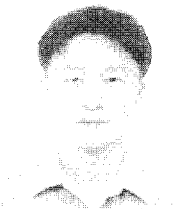
본 연구에서는 로봇의 경로를 생성하는데 있어 음성에 의한 이동경로 및 이동거리, 방향의 교시기능을 부여하여 임의의 환경에서 쉽게 경로를 생성하여 주행을 가능하게 하는 자율이동로봇을 개발하였으며 음성인식시스템의 인식 실험 및 실제 로봇에 적용하여 성능평가를 통해 다음과 같은 결과를 얻었다.

1) 임의의 환경에서 음성에 의한 이동로봇의 경로생성이 가능한 이산분포 HMM 기반 화자중속 형태의 고립단어 음성인식시스템을 구축하였으며, 인식률 향상을 위한 방안으로 HMM에 의한 인식을 수행한 후 그 신뢰도를 적절히 판단하여 신뢰도 낮은 경우 신경회로망에 의한 재인식 과정을 후처리에 도입하였다. 인식실험 결과 전체 인식단어와 로봇적용 단어에 대해 각각 91.9%와 95.7%의 인식률을 얻을 수 있었다. 이는 후처리과정을 통해 각각 0.8%와 1.5%의 인식률 향상을 보인 것이다.

2) 로봇적용 단어를 대상으로 실제 이동로봇에 적용하여 음성교시에 의한 주행 및 경로생성 실험을 수행하여 성능평가를 하였다. 설정된 환경에서 로봇은 음성교시만으로 원활한 주행을 이루어졌으며 안정한 경로를 생성하였다. 다만 하드웨어적인 처리속도 등의 한계로 인한 인식에 소요되는 시간이 다소 길어 좁은 공간에서 로봇의 정밀 위치제어에 문제점이 있었다.

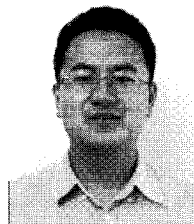
참고문헌

- [1] 성경학, 김진오, 김성권, “공장자동화를 위한 지능 로봇 시스템,” 제어·자동화·시스템 공학회지, 제 2 권 제 3 호, pp. 16-24, 1997.
- [2] 윤지섭, “원자력용 매니플레이터의 세계적 개발 현황,” 제어·자동화·시스템 공학회지, 제 2 권 제 5 호, pp. 17-25, 1997.
- [3] P. J. McKerrow, *Introduction to Robotics*, Addison Wesley, 1993.
- [4] 구명완, “음성인식기술 현황과 전망,” 전자공학회지, 제 20 권 제 5 호, pp. 548-555, 1993.
- [5] D. B. Roe and J. G. Wilpon, “Whither speech recognition: The next 25 Years,” *IEEE Communications Magazine*, vol. 31, no. 11, pp. 54-62, November 1993.
- [6] X. D. Huang, Y. Ariki, and M. A. Jack, *hidden markov models for speech Recognition*, Edinburgh University Press, Edinburgh, England, 1990.
- [7] C. S. Myers and L. R. Rabiner, “A comparative study of several dynamic time-warping algorithms for connected-word recognition,” *Bell system Tech. J.*, vol. 60, no. 7, pp. 1389-1409, September 1981.
- [8] H. F. Siverman and D. P. Morgan, “The application of dynamic programming to connected speech recognition,” *IEEE Transactions on Acoustics, Speech, and Signal Processing Magazine*, vol. 7, pp. 6-25, July 1990.
- [9] 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.
- [10] L. R. Rabiner and B. H. Juang, “An introduction to hidden markov models,” *IEEE ASSP Magazine*, vol. 3, no. 1, pp. 4-16, January 1986.
- [11] R. P. Lippmann, “An introduction to computing with neural nets,” *IEEE Transactions on Acoustics, Speech, and Signal processing Magazine*, vol. 4, pp. 4-22, Apr. 1987.
- [12] Jang, Sun and Mizutani, *Neuro-Fuzzy and Soft Computing*, Prentice-Hall, 1997.
- [13] J. G. Wilpon and L. R. Rabiner, “A modified K-means clustering algorithm for use in isolated word recognition,” *IEEE Trans. on ASSP*, vol. 33, no. 3, pp. 587-594, 1985.
- [14] J. Makhoul, S. Roucos, and H. Gish, “Vector quantization in speech coding,” *Proc. of IEEE*, vol. 73, no. 11, pp. 1551-1558, 1985.
- [15] K. F. Lee, *Automatic Speech Recognition*, Kluwer Academic Publishers, Norwell, M.A. U.S.A, 1976.
- [16] L. Rabiner and B.-H. Juang, *Fundamentals of Speech Recognition*, New Jersey: Prentice Hall, 1993.
- [17] J. A. Freeman and D. M. Skapura, *Neural Networks: Algorithms, Applications, and Programming Techniques*, Addison Wesley, 1991.
- [18] K. J. Lang and Waibel A., “A time-delay neural network architecture for isolated work recognition,” *Neural Networks*, vol. 3, pp. 23-43, 1990.



조 현 수

1999년 부산대학교 기계공학부 졸업.
2001년 부산대학교 지능기계공학과 대학원 공학석사. 2001년~현재 기아자동차 근무. 관심분야는 신호처리, 로봇제어, 설비 자동화.



박 민 규

1998년 부산대학교 정밀기계공학과 졸업. 2000년 부산대학교 지능기계공학과 대학원 공학석사. 2004년 동 대학 박사. 2004년~현재 기계기술연구소 연구원. 관심분야는 이동로봇, 경로계획, 지능제어, 이족보행로봇, 메카트로닉스.



이 현 정

2002년 인제대학교 전자공학과 졸업.
2004년 부산대학교 지능기계공학과 대학원 공학석사. 2004년~현재 부산대학교 지능기계공학과 박사과정. 관심분야는 이동로봇, 머신비전, 지능제어, 메카트로닉스.



이 민 철

1983년 부산대학교 기계공학과 졸업.
1988년 일본 쓰쿠바 대학교 이공학 연구과 공학석사. 1991년 동대학 물리공학 연구과 공학박사. 2000년 8월~2001년 8월 노스캐롤라이나 주립대학(NCSU) 방문교수. 1991년~현재 부산대학교 기계공학부 교수. 관심분야는 시스템 규명, 로봇제어, 지능형 서비스로봇, 이족보행로봇, 메카트로닉스.