論文96-33B-11-11

대화형 음성인식 이동로봇에 관한 연구

(A Study on the interactive Speech Recognition Mobile Robot)

李 在 泳 *, 尹 晳 鉉 *, 洪 光 錫 *

(Jae-Young Lee, Seok-Hyun Yoon, and Kwang-Seok Hong)

요 약

본 논문은 대화형 음성인식 기술을 이동로봇에 적용한 연구이다. 음성명령은 문장단위의 연결단어로 발성되고 이동로봇의 무선마이크 시스템을 통하여 입력된다. 이 음성신호는 DSP 보드를 통하여 LPC-Cepstrum 계수와 shorttime Energy를 추출하여 노트북 PC로 전송한다. 노트북 PC에서는 DP matching 기술을 이용하여 입력된 명령어를 인식하며, 인식된 결과는 모터제어부로 전송되고 그에 따른 출력 펄스 신호를 Interface Card로 전송하여 스테핑 모터를 제어한다. 제작된 이동로봇은 인식속도를 줄이기 위해 Grammar Network을 적용하여 실시간 음성인식이 가능하도록 하였으며, Interface Revision를 적용하여 오인식된 부분을 로봇과의 대화를 통하여 정정하도록 하였다. 따라서 이동로봇은 사용자가 원하는 방향으로 안정적으로 이동할 수 있도록 설계되었다.

Abstract

This paper is a study on the implementation of speech recognition mobile robot to which the interactive speech recognition techniques is applied. The speech command uttered the sentential connected word and is asserted through the wireless Mic system. This speech signal transferred LPC-Cepstrum and shorttime Energy which are computed from the received signal on the DSP board to Notebook PC. In Notebook PC, DP matching technigue is used for recognizer and the recognition results are transferred to the motor control unit which output pulse signals corresponding to the recognized command and drive the stepping motor. Grammar Network applied to reduce the recognition speed of the recogniger, so that real time recognition is realized. The misrecognized command is revised by Interface Revision through the conversation with mobile robot. Therefore, user can move the mobile robot to the direction which user wants.

I. 서 론

움직이면서 사람의 말을 알아듣고, 명령에 따라 동작하며, 사람에게 필요한 정보를 제공하기도 하는 로봇은 마이크로컴퓨터 기술과 VLSI 기술의 진보로 인해 가까운 미래에는 상당한 정도로 실현 가능하게 될 것이라 예측되며, 이로 인한 사회, 문화, 생활, 산업, 경제적인 파급효과는 예측하기 힘들 정도로 다양하게 나타날 것이라 생각된다.

그러나 불특정 일반인의 음성인식이나 음성인식의 유연성을 제공하기에는 상당한 문제점을 갖고 있는 것이 현실이다. 게다가 대화를 나누기 위한 음성합성 기술 역시 사람처럼 어떠한 언어라도 마음대로 구사할 수 있는 수준은 아니다. 또한 지능형 로봇을 만든다는 것은 단순히 로봇을 만드는 기술 뿐만 아니라 음성인식, 시각인식 등의 제반기술이 복합적으로 결합되어야

* 正會員, 成均館大學校 電子工學科

(Dept. of Electronic Eng., Sung Kyun Kwan Univ.)

※ 이 논문은 성균관대학교의 1995년도 성균학술 연구비에 의하여 연구되었음

接受日字:1996年6月20日, 수정완료일:1996年10月25日

만 가능하게 된다. 이러한 제반 기술은 사용자 인터페이스, 인간-기계 인터페이스, 인간-컴퓨터 인터페이스의 영역에 대한 기술로써, 많은 나라의 연구소 및 학교에서 이에 대한 기술의 발전을 위하여 정진하고 있는 중이다. 또한 음성인식에 대하여는 각 나라마다 언어의 차이가 커서 자기 나라의 언어에 대하여는 해당하는 나라의 연구자가 개발을 해야하는 특성을 갖고 있다[1][2][4][8].

본 논문에서 구현한 이동로봇은 대화형 음성인식 이동로봇이다. 이동로봇의 실시간 특징파라메터 추출을 위하여 DSP 보드를 사용하였으며, 음성인식 부분은 시스템과 사용자에게 유연성을 부여하기 위하여 노트북 PC에서 수행토록 하였다.

음성명령어는 사용자의 자연스러운 대화형태인 문장 단위의 연결단어로서 무선마이크를 통해 입력을 받으므로서 각각의 명령어를 또박또박 발음해야하는 고립단어 방식보다 사용자의 편의를 높였으며, 노트북 PC에서 음성인식 알고리즘인 DTW(Dynamic Time Warping)를 사용하여 인식하였고, 인식기 알고리즘에 Grammar Network을 적용함으로서 이동로봇의 실시간 구현이 가능하였다. 또한 Interface Revision을 적용함으로서 사용자와 이동로봇 간의 대화를 통해서 오인식된 부분을 정정하여 이동할 수 있도록 설계하여 그 성능을 평가하였다.

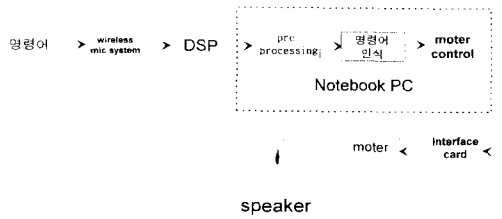


그림 1. 대화형 음성인식 이동로봇의 간략화된 블록도

Fig. 1. The simplified block diagram of the interactive speech recognition mobile robot.

Ⅱ. 이동로봇의 시스템

그림 1은 대화형 음성인식 이동로봇의 간략화된 블록도이다. 이동로봇은 사용자가 헤드셋 무선마이크를 통해 명령어를 발성하고, 이를 받아들이는 무선마이크 시스템 부분, 받아들여진 음성명령의 신호처리를 하는 DSP 부분, 음성의 인식을 처리하는 노트북 PC 부분, 모터의 구동을 위한 interface card 부분, motor 부분 등으로 나뉘어져있다.

사용자에 의해 발성되어진 음성명령은 무선마이크시스템을 통하여 입력되며, 무선마이크시스템을 사용함으로서 로봇 구동의 편의를 꾀할 수 있다. 이들 음성명령은 National 사의 TP3054 Codec을 사용하여 A/D 변환되며 Analog device 사의 ADSP2101 chip[6]을 사용하여 자체 제작한 신호처리용 DSP(Digital Signal Processing) 보드를 사용하여 특징 파라미터 즉 LPC-Cepstrum 계수와 에너지를 추출하였다. DSP 보드는 매 프레임마다 음성 파라미터를 계산하여 버퍼(FIFO)에 집어 넣는다. Notebook PC는 Polling을 통하여 새로운 data가 들어오면 FIFO로 부터 그 프래임의 음성 파라미터를 읽어 들인다. DSP 과정을 이동로봇에 추가시킨 것은 실시간 음성 파라미터의 추출을 위해서이며 구해진 계수들 및 에너지는 노트북 PC로 전송된다. 이의 과정을 그림 2에 나타내었다.

노트북 PC에서는 음성명령의 시작점 및 끝점 검출, keyword spotting을 통한 세그멘테이션, Grammar Network을 적용한 인식 알고리즘 수행, Interface Revision을 통한 오인식의 보완, 인식된 결과에 따른 모터제어부의 모터제어코드 발생을 처리하였다.

모터제어코드는 Interface card로 전해져서 인식된 결과에 따른 펄스를 발생시켜 모터를 제어한다. 사용된 모터는 제어하기 간단한 스테핑 모터 2 개를 사용하였으며 이것으로 이동로봇의 뒷바퀴를 회전시켰다[7].

그외 모터 로터리인코더, 모터기어, 밧데리 케이스, 프레임 등 좀더 복잡한 기계적 구성을 하고 있다.

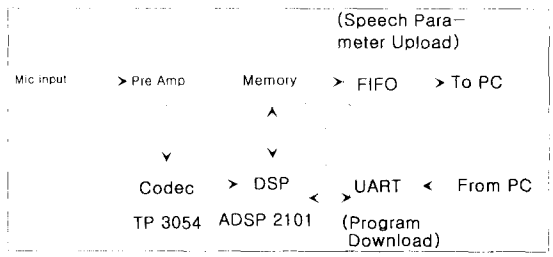


그림 2. DSP 보드의 블럭도

Fig. 2. The block diagram of DSP board.

Ⅲ. 음성인식 알고리즘

노트북 PC에서의 음성인식 과정을 그림 3에 나타내

었다. 음성인식 알고리즘은 노트북 PC에서 수행하도록 하여 시스템과 사용자에게 유연성을 발휘할 수 있도록 하였으며, 인식된 결과 및 Interface Revision을 통한 이동로봇과 사용자 간의 대화를 display 하고 speaker로 들려줌으로서 시각적, 청각적으로 이들 결과를 확인할 수 있도록 설계하였다.

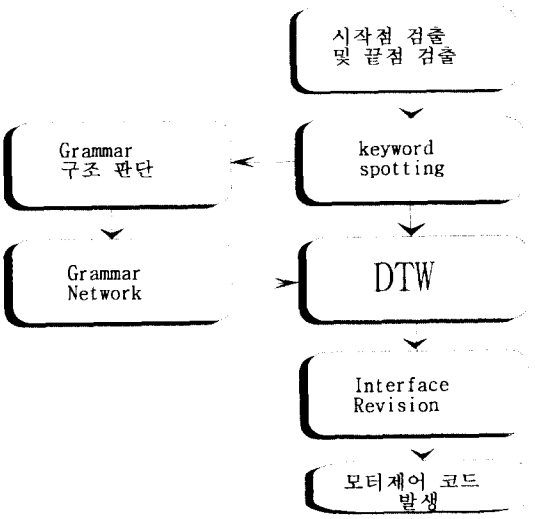


그림 3. 음성인식을 위한 노트북 PC의 역할
Fig. 3. The role of Notebook PC for the speech recognition.

1. 명령어

대화형 음성인식 이동로봇에 사용되는 명령어는 구동명령어와 대화명령어로 구별하여 사용하였다. 구동명령어는 이동 방향을 정해주는 "앞으로, 뒤로, 좌로, 우로, 반대로"의 방향명령어 5개, 이동 거리를 정해주는 "1 미터, 2 미터, 3 미터, 4 미터, 5 미터, 6 미터, 7 미터, 8 미터, 9 미터, 10 미터"의 거리명령어 10개, 회전각도를 정해주는 "10 도, 20 도, 30 도, 40 도, 50 도, 60 도, 70 도, 80 도, 90 도"의 각도 명령어 9개, 속도를 정해주는 "빨리, 천천히"의 속도명령어 2개, 실행을 정해주는 "가, 서, 돌아, 좀더"의 실행명령어 4개며 총 30개의 명령어로 구성되어 있다. 구동명령어는 연결단어의 형태로 한국어 문법 구조에 맞게 문장 단위로 발성되어지며 6개의 stage까지 구성될 수 있도록 설계하였다. 따라서 각각의 stage에 사용된 단어의 조합에 의해 총 603종류의 문장 표현이 가능하며 사용자가 원하는 위치로 음성명령어의해 이동로봇을 이동시킬 수 있다.

대화명령어는 명령어의 확인을 물어보는 "예, 아니오"의 확인명령어 2개, 틀린 명령어의 개수를 확인하는 "한 개, 두 개, 세 개, 네 개, 다섯 개, 여섯 개"의 개수명령어 6개, 틀린 명령어의 위치를 물어보는 "첫번째, 두번째, 세번째, 네번째, 다섯번째, 여섯번째"의 위치명령어 6개의 총 14개의 명령어로 구성되어있다. 표 1은 이동로봇에 사용된 명령어를 나타내었다.

표 1. 이동로봇에 사용된 명령어
Table 1. The used command to the mobile robot.

구동명령어	방향명령어	앞으로, 뒤로, 좌로, 우로, 반대로
	거리명령어	1 미터, 2 미터, 3 미터, 4 미터, 5 미터, 6 미터, 7 미터, 8 미터, 9 미터, 10 미터
	각도명령어	10 도, 20 도, 30 도, 40 도, 50 도, 60 도, 70 도, 80 도, 90 도
	속도명령어	빨리, 천천히
	실행명령어	가, 서, 돌아, 좀더
대화명령어	확인명령어	예, 아니오
	개수명령어	한 개, 두 개, 세 개, 네 개, 다섯 개, 여섯 개
	위치명령어	첫 번째, 두 번째, 세 번째, 네 번째, 다섯 번째, 여섯 번째

2. 시작점 검출 및 끝점 검출

노트북 PC로 전송된 특징 파라미터 중에서 에너지는 음성명령어의 시작점 및 끝점을 검출하는데 사용한다. 여기서, 에너지는 한 프레임의 에너지를 나타내는 것이며 DSP 보드에서 발생하는 DC Bias 때문에 어느 정도 이상의 값을 나타낸다. 일상 주변이나 연구실 환경하에서 noise와 DSP 보드의 DC Bias가 합쳐진 값은 마이크를 통한 명령어 입력이 없을 때 어느 정도 임계치(threshold)보다 낮은 값을 나타내고 이들 임계치는 실험적으로 얻어진다. 따라서 임계치보다 낮은 Energy 값이 PC로 전송되면 명령어가 입력되지 않은 것으로 판단하여 버리게 되고 임계치 이상일 때는 명령어가 입력되었는지의 유무를 판단하여 음성명령의 시작점을 검출하고 끝점은 이의 반대과정으로 검출하게 된다. 즉, 다음과 같은 방법을 따르게 된다.

① if, $ET(Energy\ Threshold) < E_i$ (1)

$H_i = 1$

else, $H_i = 0$

$i=1,2,3,\cdots\cdots,N$

② $\underset{i}{\arg\min}\left(\sum_{i=j}^{j+10} H_i=10\right)=startpoint$ (2)

$j=1,2,\cdots\cdots,N-10$

③ $\underset{j}{\arg\min}\left(\sum_{i=j}^{j+10} H_i=0\right)=endpoint$ (3)

$j=startpoint, startpoint+1,\cdots\cdots,N-10$

여기서, E_i : i번째 프레임의 에너지

H_i : i번째 프레임의 History

10 : stable level(실험적 수치)

식(1)은 실험적으로 구해진 일상 주변이나 연구실 환경하에서 noise와 DSP 보드의 DC Bias가 합쳐진 값과 입력된 명령어와의 비교를 통해서 입력명령어들의 프레임의 History를 정하게된다. 식(2)는 입력된 명령어의 시작점을 찾기위한 과정으로 10 프레임 동안 안정적으로 입력이 이루어진다면 명령어가 입력되었다고 판단하며, 식(3)은 입력된 명령어의 끝점을 찾기위한 과정으로 10 프레임 동안의 무음구간을 확인하기 위해서이다.

3. 세그멘테이션

입력된 음성명령의 시작점과 끝점을 추출한 후, 연결단어로 발성된 음성명령을 규칙에 의거한 세그멘테이션을 한다. 이와같은 과정을 수행하는 것은 입력된 명령어가 연결단어 형태의 명령어이기 때문에 명령어를 분리하여 각각을 인식기의 pattern으로 입력을 하게된다.

세그멘테이션을 위해서 keyword spotting을 이용하였다. 이는 연속적으로 발성된 명령의 꼭 필요한 정보(keyword)를 찾아냄으로서 명령어간의 경계지역을 찾을 수 있다. 문장단위로 발성된 연결단어 명령어는 공통적인 keyword들의 조합으로 이루어져있으며 이 keyword들은 방향명령어에 대한 '로', 거리명령어에 대한 '미터', 각도명령어에 대한 '십도' 그리고 '돌아' 이며 이들을 reference로 저장을 해놓고 test pattern의 가장 유사한 프레임을 찾아서 각각의 음성 명령어간의 경계로 사용하게 된다. 그 과정은 다음과 같은 방법을 따르게 하였다.

while($k \le number\ of\ keyword$) {

① $\sum_{j=0}^{P}\left(T_{i,j}-R_{i,j,k}\right)^2=D_{i,k}$ (4)

$i=startpoint, startpoint+1,\cdots\cdots,endpoint$

② $\underset{i}{\arg\min}\left(D_{i,k}\right)=segment(k)$ (5)

}

여기서, k : keyword의 번호

P : LPC Cepstrum의 order

$T_{i,j}$: i번째 프레임의 Test pattern

$R_{i,j}$: i번째 프레임의 Reference pattern

식(4)는 입력 pattern과 각각의 keyword를 비교를 하여 그 유사성을 구하게되며 식(5)는 가장 유사한 곳의 위치를 찾아냄으로서 각각의 keyword의 위치를 입력 pattern에서 찾아내게 된다.

4. Grammar Network

Grammar Network은 이동로봇의 구현에 있어 실시간 음성인식을 위해서 사용하였다. 즉, 사용자는 문법구조에 맞게 문장단위로 발성을 하고, 발성된 문법구조에 의해 인식기의 비교 pattern인 reference pattern의 종류를 제한함으로서, 인식속도를 감소 시킬 수 있다. 이동로봇에 적용된 Grammar Network은 구동명령어들의 조합에의해 문장으로 구성되며 이동로봇에 적용된 Grammar Network 및 keywords을 표 2에 나타내었다.

keyword spotting으로 keyword를 찾은 후, 각각의 keyword들의 순서를 알수 있다. 이 순서를 Grammar Network과 비교함으로서 입력된 명령어의 개수와 문법구조를 알수 있으며 각각의 stage에서 입력 pattern과 비교될 reference pattern의 종류를 알수 있다. 예를들어 "좌로 3십도 돌아 앞으로 5미터 가"는 keywords의 순서가 '로', '십도', '돌아', '로', '미터' 이며 표 2의 최하단에 적용됨을 알 수 있다. 따라서 문법구조에 의해 첫 번째 stage의 명령어는 "좌로, 우로, 반대로"의 3가지 명령어만이 입력 가능하며 인식속도는 Grammar Network이 적용되지않았을 때와 비교하면 비교될 reference의 개수가 30개에서 3개로 줄어들므로 첫 번째 stage에서 인식속도가 0.1배로 줄어들게된다. 이와같이하여 두 번째 stage는 0.3배, 세 번째 stage는 문법구조에 의해 인식기를 거치지않고 인식되며, 네 번째 stage는 0.17배, 다섯 번째 stage는 0.33배, 여섯 번째 stage는 문법구조에 의해 인식기를 거치지않고 인식되어 실시간 음성인식이 가능하였다. 인식

기를 거치지않고 인식되는 경우는 전단의 명령어가 각도명령어이면 다음단은 반드시 '돌아' 명령어만이 나오며, 거리명령어이면 반드시 '가' 명령어가 나오기 때문에 인식단을 거치지 않고 인식된다.

표 2. Grammar Network 및 keywords
Table 2. Grammar Network and the keywords. (keyword)

1 stage	2	3	4	5	6	개수 (개)
실행						3
방향(로)	실행					15
속도	실행					4
방향(로)	거리(미터)	실행				40
방향(로)	각도(십도)	실행(돌아)				27
방향(로)	각도(십도)	실행(돌아)	실행			18
방향(로)	실행(돌아)	방향(로)	실행			4
방향(로)	실행(돌아)	방향(로)	거리(미터)	실행		60
방향(로)	각도(십도)	실행(돌아)	방향(로)	실행		36
방향(로)	각도(십도)	실행(돌아)	속도	실행		36
방향(로)	각도(십도)	실행(돌아)	방향(로)	거리(미터)	실행	360

5. 인식 알고리즘

음성인식 방법은 DTW(Dynamic Time Warping), HMM(Hidden Markov Model), ANN(Artificial Newral Network) 등이 있는데 여기에서는 단어 단위에서 비교적 높은 인식률을 보이는 DTW를 사용한다.

DTW는 패턴 매칭(Pattern Matching) 기법의 일종으로 dynamic programming 이라는 보다 일반적인 계산 기법에 기초를 두며, 수행 방법은 reference pattern으로 저장해 놓은 명령어들과 무선마이크 시스템을 통하여 입력된 명령어와 프레임 단위로 거리값을 비교한다. 이때 비교되는 프레임의 가장 유사한 프레임을 비교함으로서 사용자의 발성 시간의 변화에대한 왜0곡을 줄일 수 있게 된다. 수행과정 중 불필요한 부분과의 비교는 인식시간의 지연을 초래하며 비슷한 명령어는 reference pattern과 test pattern의 기울기와 유사한 지역에서 pattern matching이 이루어 지므로 전체 경로제약과 소구간 경로제약을 주게 되었다[1] [2] [3] [5].

무선마이크 시스템에 큰 noise가 입력되든지 또는 로봇 구현 방법을 모르는 사용자가 엉뚱한 명령어를 내리던지 하였을 경우에 즉, 임계치 이상의 거리값을 나타내었을 때 인식된 결과는 rejection 되며 스피커를 통해서 "다시 명령해 주세요" 라는 sound를 출력시켜서 로봇을 안정화시키고 또한 사용자를 환기시킨다. 인식되어진 모든 명령어는 임계치 이하를 나타낸다면 모터 제어부를 거쳐 로봇을 구동시킴과 동시에 그에 알맞은 음성응답을 스피커를 통해서 출력함으로서 인식된 결과를 시각적 뿐 아니라 청각적으로도 확인할 수 있다.

인식기에 keyword spotting를 이용하여 각각의 명령어의 경계지역을 찾으며, Grammar Network을 이용하여 reference pattern을 제한하고, Interface Revision을 적용하여 오인식 결과를 정정하는 방법을 그림 4에 나타내었다.

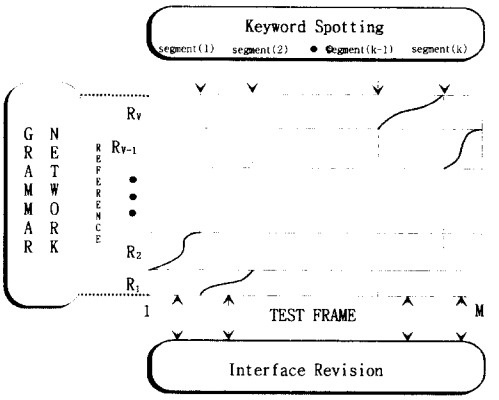


그림 4. 제안된 인식기
Fig. 4. The proposed recognizer.

6. Interface Revision

Interfae Revision은 명령어 인식결과를 확인 및 정정 함으로서 이동로봇이 안정적으로 이동하도록 하기 위함이다. 즉 확인명령어, 개수명령어, 위치명령어 등 대화명령어의 물어보는 말을 노트북 PC의 speaker로 출력을 하여 사용자와 대화를 통한 오인식 부분의 정정을 위함이며 다음과 같이 수행된다.

① 인식결과의 확인 단계
② if, "reject"이면 명령어의 재발성
if else, "예"이면 이동로봇 구동
else, "아니오"③의 과정 수행
③ 오인식된 개수 확인 단계
④ 오인식된 위치의 확인 단계
⑤ 오인식된 부분을 각각 다시 발성

⑥ 다시발성된 부분의 인식결과 확인 단계
⑦ if, "아니오"이면 ⑤,⑥를 다시 수행
else, 이동로봇 구동

7. 모터제어부, Interface card, 모터

모터제어부는 인식된 코딩 명령에 따라서 Interface Card의 8254, 8255의 출력 포트 출력값을 제어함으로서 모터의 회전 방향을 제어하기 위한 부분이다. 즉 Interface Card의 H/W를 제어하는 S/W이다. 인식의 출력 코딩 명령에 따라 Interface Card의 8254, 8255 각각에서 모터 부분의 로터리 인코더(Rotary Encoder)의 각각에 일정한 입력을 주도록 프로그래밍된다.

각각의 코딩 명령어에 대해서 모터가 급격히 회전 방향을 바꾼다면 이동 중의 로봇 자체의 가속에 의해서 모터에 무리한 힘을 주게되어 모터의 수명에 큰 저하가 있게 된다. 따라서 그림 5와같이 어느 방향의 코딩 명령이 입력되어도 가/감속의 단계를 수행하여 모터에 무리한 힘을 주지 않게 하였으며, 로봇의 미끄러짐 또한 방지하였다.

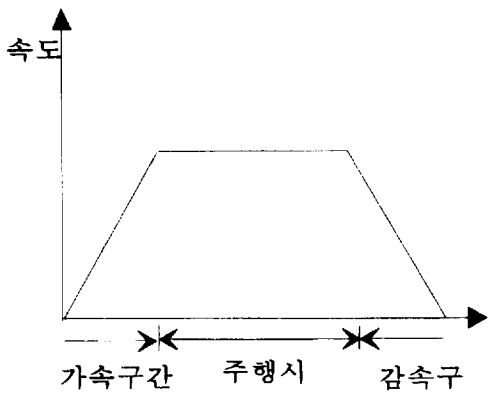


그림 5. 가/감속 단계
Fig. 5. Acceleration/Deceleration.

또한, 그림 6와같이 곡률 주행시 모터에 무리한 힘을 주지않으며 최소한의 반경을 그리면서 회전할 수 있도록 설계하였다. 최소한의 반경은 로봇의 회전 시 주위의 장애물에 부딪치지 않을 최소한의 공간을 의미한다.

Interface Card는 8254, 8255 등으로 구성된 H/W로서 코딩 명령어 따라 일정한 pulse를 로터리 인코더 부분에 보내주게 된다. 즉, 8255에서는 모터의 hold를 on/off 해주는 pulse를 8254에서는 모터의 속도를 제어해주는 pulse를 출력하게 된다. 모터는 1 pulse 당 일정한 회전을 하는 스테핑 모터 2개를 사용하였다[7]. 이동로봇 시스템은 그림 7과 같다.

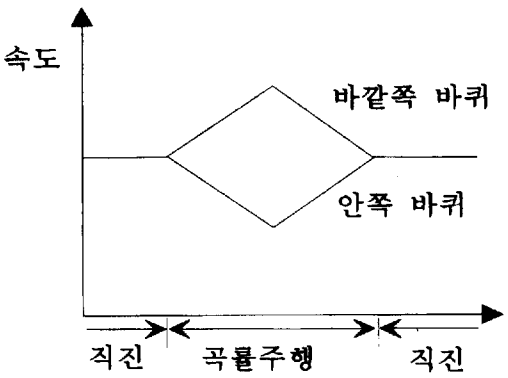


그림 6. 이동로봇의 곡률 주행
Fig. 6. The curvature moving of the mobile robot.

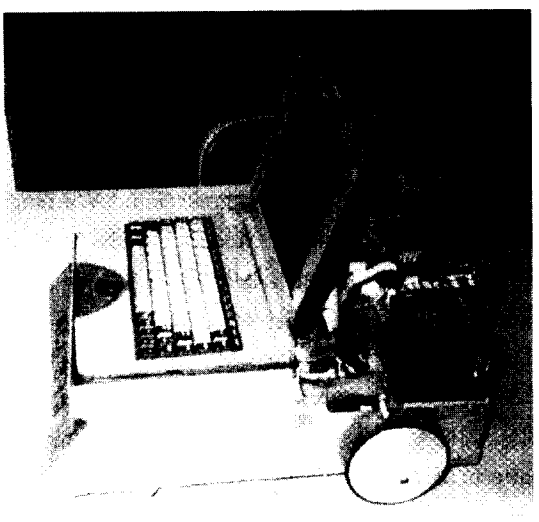


그림 7. 대화형 음성인식 이동로봇
Fig. 7. The Interactive Speech Recognition Mobile Robot.

Ⅳ. 인식실험 및 고찰

실험의 조건은 20대 남성화자 1명이 발성하였으며, 8 kHz의 sampling rate, Hamming window, 25mSec의 프레임 길이, LPC-Cepstrum 계수 14차, Reference template는 1개, 2개를 사용한 화자종속실험을 하였다. 실험조건을 표 3에 나타내었다.

표 3. 실험조건
Table 3. The experiment condition.

Sampling rate	8kHz
window 종류	Hamming
프레임 길이	25mS
LPC-Cepstrum 차수	14차
Reference template 갯수	1개, 2개

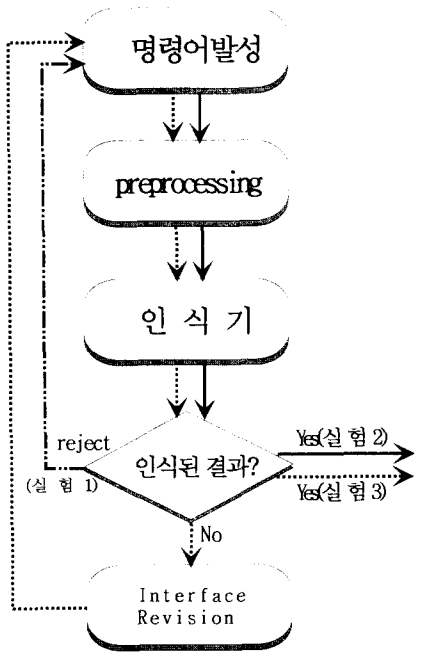


그림 8. 인식실험 블록도
Fig. 8. The recognition experiment block diagram.

실험은 그림 8과같이 세단계의 인식실험을 하였으며 실험1은 keyword spotting을 통한 세그멘테이션의 실패로 인하여 Grammar Network에서의 Grammar 구조가 오적용되어 입력된 명령어를 reject 하는 경우이다. 실험2는 연결단어로 발성된 명령어가 reject 되지않는다면, Interface Revision을 적용하지 않고 인식기를 통한 인식된 결과만을 나타내는 인식실험을 하였다. 즉 사용자가 이동로봇에게 무선마이크를 통해서 음성명령을 내리는 즉시 인식기를 통한 인식을 한 후, 이동을 할 수 있는 인식실험이다. 실험3은 3.6 Interface Revision의 과정 중에서 ⑥의 과정까지만 거친 인식실험을 하였다. 즉 이동로봇의 음성명령에대한 오인식 고려하여 이동로봇과의 대화를 통하여 오인식된 부분의 정정을 정정하며 빠른 구현 시간이 고려된 안정적으로 이동할 수 있는 인식실험이다. 단 ⑦의 과정까지 거친다면 좀더 안정적인 인식결과를 얻을 수 있을 것이다.

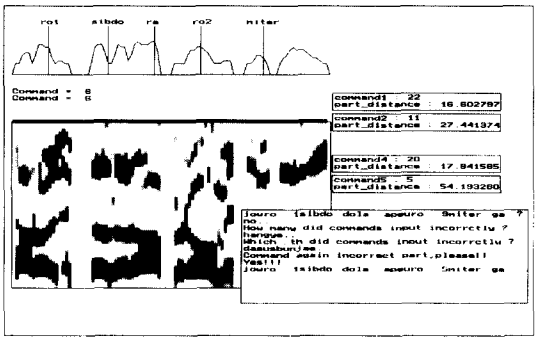


그림 9. 음성 인식 화면
Fig. 9. The speech recognition display.

그림 9은 이동로봇 시스템 출력을 나타내었다. 입력된 음성 명령어에 대한 프레임 에너지를 왼쪽 상단에 출력하였으며 이부분에 keyword인 '로', '십도', '돌아', '미터'의 위치를 나타내었다. 방향명령어가 두 개 나올 경우, ro1은 첫 번째 '로', ro2는 두 번째 '로'이며, sibdo는 '십도', miter는 '미터', ra는 '돌아'이다. 이들 keywords의 순서에 의해 명령어의 개수 6을 찾게되었으며 Grammar Network을 적용하여 문법구조인 6를 찾게된 것을 나타내었다. 또한 인식된 결과와 문법구조에 의해 인식단을 거치지않은 결과를 오른쪽 중앙에 나타내었으며 Interface Revision의 적용에 의해 사용자와 이동로봇 간의 대화를 통해서 오인식된 부분이 정정되고 있음을 오른쪽 하단에 나타내었다. 즉 인식된 결과인 "좌로 1십도 돌아 앞으로 9미터 가?"라는 이동로봇의 물음에 "아니오"라는 사용자의 대답, "얼마나 많은 명령어가 옳지 않게 입력되었는가?"라는 이동로봇의 물음, "한개"라는 사용자의 대답, "몇 번째 명령어가 잘못 입력되어 있는가? 라는 이동로봇의 물음, "다섯번째"라는 사용자의 대답, "잘못된 부분을 다시 명령해 주세요!" 라는 이동로봇의 물음, 사용자가 잘못된 부분을 다시 발성한 후, 오인식된 부분만을 정정하여 "예!, 좌로 1십도 돌아 앞으로 5미터 가겠습니다"라는 이동로봇의 대답 후에 이동로봇은 이동하게 된다.

실험의 결과는 603가지 문장에 대한 각각 stage의 전체 평균 인식률이며 이에 대한 결과를 표 4에 나타내었다.

표 4. 실험 결과
Table 4. The experiment results.

	template 1개	template 2개
실험 1 (성공율) (Grammar Network)	78.87%	79.88%
실험 2 (인식률) (Interface Revision 無)	87.5%	89.6%
실험 3 (인식률) (Interface Revision 有)	90.3%	96.3%

실험1은 연결단어로 발성된 명령어에대한 세그멘테이션의 성공율을 나타내었다. 이 성공률은 Grammar Network 적용의 성공률과도 같은 의미이며 3.6 Interface Revision의 과정 중 ③, ④, ⑤, ⑥, ⑦을 적용할 수 있는 경우이다. 따라서 실험3은 실험1이 성공하였을 때만 행하여지게 되었다. 실험2는 실험1이 성공하였을 경우, 각각 stage의 평균인식률로서 Interface Revision을 적용하지않는 이동로봇의 빠른 구현을 위한 자료로서 쓰일 수 있다. 실험3은 Interface Revision의 과정에서 사용된 대화명령어와 구동명령어에 대한 오인식률이 포함되어있으며 이의 적용으로 이동로봇의 안정적인 이동을 위한 자료로서 쓰일 수 있음을 나타낸다. 단 이동로봇를 이동시키기 위한 구현시간을 고려하여 3.6 Interface Revision의 과정 중 ⑦의 과정은 생략하게 되었다.

실험결과에 의하면 reference template의 수를 증가하면 당연히 인식률이 좋아지지만 두 개까지가 실시간 인식처리에 적당하다고 생각된다.

새로운 사용자가 이동로봇을 화자종속으로 사용할 경우는 총 44개의 명령어를 각 단어에 대해 template의 수만큼 발성하여 reference를 만들후 이를 화자정보와 같이 등록한 후, 이동로봇을 구동시킬 수 있다.

V. 결 론

본 연구는 음성인식 기술을 대화형 이동로봇에 적용하였으며, 사람의 자연스러운 발성 형태인 연결단어로 입력을 받으므로서 이동로봇 사용자의 발성의 편의가 고려되었고 무선마이크시스템을 사용함으로서 이동로봇 구동의 편의가 고려되어 설계되었다.

음성 명령어의 특징파라미터 추출에 DSP 보드를 사용함과 인식기에 Grammar Network을 적용함으로서 이동로봇의 실시간 처리가 가능하게 하여 음성인식 이동로봇의 활용가능성을 높였다. 또한 이동명령어의 인식된 결과에 Interface Revision을 적용함으로서 사용자와 이동로봇 간에 대화를 통해서 오인식률을 줄여 이동로봇이 안정적으로 이동하도록 하였다.

참 고 문 헌

[1] Cory S. Myers, Lawrence R. Rabiner "Connected Digit Recognition Using a Level-Building DTW Algorithm", *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. ASSP-29, no 3, 351-363, 1981.

[2] Cory S. Myers, Lawrence R. Rabiner "A Level Building Dynamic Time Warping Algorithm for Connected Word Recognition", *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. ASSP-29, no 2, 284-297, 1981.

[3] Hermann Ney "The Use of a One-Stage Dynamic Programming Algorithm for Connected Word Recognition", *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. ASSP-32, no 2, 263-271, 1984.

[4] David B. Roe and Jay G. Wilpon, *"Voice Communication Between Humans and Machines"*, National Academy of Sciences, 1994.

[5] Rabiner L. and B. H. Juang, *"Fundamental of Speech Recognition"*, Prentice-Hall International, 1993.

[6] The Applications Engineering staff of Analog Device, DSP Division, *"Digital Signal Processing Applications Using the Adsp-2101 Family"*, Prentice Hall, vol 1, 1990.

[7] 전윤호 외10, *"마이크로 지능 로보트"*, ohm 사, 1991

[8] 동역 메카트로닉스 연구소, *"음성합성과 음성인식 시스템"*, 영진출판사, 1990

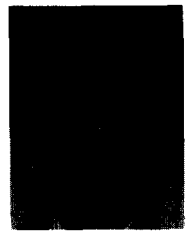
저 자 소 개

李 在 泳(正會員)

1970年 1月 1日生. 1995年 2月 성균관대학교 전자공학과 공학사. 1995年 3月 ~ 현재 성균관대학교 대학원 전자공학과 석사 과정. 주관심 분야는 음성인식 및 음성신호처리 등임.

洪 光 錫(正會員)

1959年 2月 7日生. 1985年 2月 성균관대학교 전자공학과 공학사. 1988年 2月 성균관대학교 전자공학과 공학석사. 1992年 2月 성균관대학교 전자공학과 공학박사. 1990年 3月 ~ 1993年 2月 서울보건전문대학 전산정보처리학과 전임강사. 1993年 3月 ~ 1995年 2月 현재 제주대학교 정보공학과 전임강사. 1995年 3月 ~ 현재 성균관대학교 전자공학과 조교수. 주관심 분야는 음성 및 신호처리, 패턴인식 등임.



尹 晳 鉉(正會員)

1966年 12月 1日生. 1992年 2月 성균관대학교 전자공학과 공학사. 1996年 2月 성균관대학교 전자공학과 공학석사. 1996年 3月 ~ 현재 성균관대학교 대학원 전자공학과 박사 과정. 주관심 분야는 음성 및 신호처리, 신경회로망 등임.