

음성인식을 이용한 중증 장애인 생활환경 제어장치

The Environmental Control System for Quadriplegics using Speech Recognition

임재용*, 정혁준**, 이행세**
(Jaeyong Rim*, Hyuckjun Jung**, Hangsei Lee**)

*아주대학교 대학원 의용공학과, **아주대학교 전자공학부

* Dept of Biomedical Eng, Ajou Graduate School, **School of Electronics, Ajou Univ

E-mail : jaeyongrim@hotmail.com

요 약

1. 개 요

척수 손상에 의한 사지마비 중증 장애인이 일상 생활을 함에 있어서 가족이나 주변인의 도움을 받지 않고서는 가전제품의 작동이나 전화 통화 등과 같은 작은 일조차 스스로 할 수 없는 현실에서 국내 여건에 적합한 생활환경 제어장치의 개발이 필요하나 현재 이를 위한 마땅한 제어장치가 없는 실정이다. 각 가정에 널리 보급되어있는 PC를 활용할 수 있도록 PC기반의 음성인식기를 개발하고 이를 응용한 가전제품 제어기를 개발, 제작하여 사지마비 환자 스스로 간단한 동작만으로 환자와 가족의 신체적, 정신적 부담을 감소 하고자 한다. 이장치는 음성인식기 부분과 제어장치 부분 그리고 I/O 인터페이스 부분으로 구성 되어있다.

중증장애인은 이 장치를 사용하여 일상생활에서 기본적인 자유로운 생활의 영위를 도모하며 환자 가족의 부담을 줄일 수 있으며 환자와 가족 모두 심리적인 만족을 얻은 것으로 나타난다. 그러나 기기 사용을 위한 환자의 음성인식을 위한 언어학습과정의 불편함과 인식에러의 과제가 남아있다.

본 장치는 크게 세 부분으로 구분이 되어있다. 음성인식기 부분, 제어장치 부분, I/O 인터페이스 부분으로 나누어 볼 수 있는데, 음성인식기 부분은 HMM(Hidden Markov Model)을 이용하여 구현하였고, 13차 맥스 트럼 계수를 사용하여 음성의 특징을 추출하였으며 인식 단어의 수는 26가지로 하였다. 음성 구간 검출을 끝점 검출 알고리즘을 사용하였고 알고리즘의 입력 파라메타로 가중 단구간 영교차율을 이용하였다. 여기에 사용된 코드북은 K-means 집단화 알고리즘을 사용하여 백터 양자화 하여 64인 코드북을 생성시켜 음성인식에 사용하였고 인식률은 약 95%이다. 제어부분은 7225칩을 사용하였으며 릴레이를 이용하여 사용의 편의성과 업그레이드의 대비를 하였다. I/O 인터페이스 부분은 중증장애인의 사용 편리를 위하여 넥 마이크를 사용하였고 이를 PC의 사운드카드에 연결하였다. PC에서 인식한 결과를 37핀 D-타입 케이블을 통하여 제어보드에 명령을 전달하고 이를 적외선 리모트 컨트롤러를 이용한 장치에 전송하여 제어 대상에 전달하여 환자가 원하는 작동을 하게 된다.

2. 생활환경 제어장치의 구성

2-1 음성 인식기 부분

음성 인식기의 인식부는 Boland사의 C++빌더를 사용하여 프로그램화하였다. 프로그램은 음성의 끝점 검출, 음성의 특징 추출, 벡터 양자화 학습, 벡터 양자화 인식, HMM학습 그리고 HMM인식부분으로 나뉘어져있다. 벡터 양자화와 확률적 모델을 바탕으로 한 HMM을 이용하여 구성하였으며, 음성 데이터는 11.025kHz로 샘플링하고 16비트로 양자화 한 후 끝점 검출기를 통해서 음성 구간을 검출하게 된다. 검출된 음성에서 멜-캡스트럼 계수를 구하여 학습부와 인식부에서 사용하게 되며 학습부에서는 음성의 특징 추출 과정을 거친 음성 신호를 벡터 양자화 하여 코드북을 생성하고 여기서 생성된 코드북 인덱스열(관측 심벌)을 사용하여 각 단어 별 HMM의 확률 파라미터를 추정하게 된다. 인식부는 학습부에서 학습되어 구성되었던 HMM 참조 패턴들과 벡터 양자화에서 생성된 코드북, 그리고 임의의 입력 음성 데이터를 사용하여 코드북 인덱스열을 구한 후 코드북 인덱스열을 인식 알고리즘으로 패턴 비교하여 인식 결과를 얻게 된다.

인식 결과는 그림 1 과 같이 모니터에 표시된다.

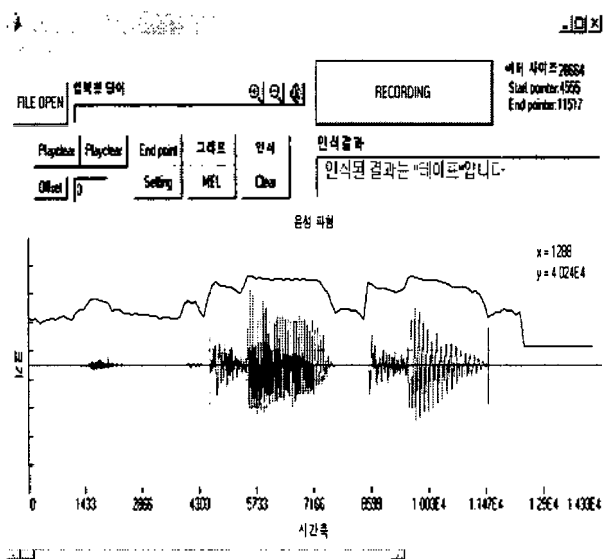


그림 1 " 테이프 " 발음의 인식 결과.

2-2 제어 장치 부분

제어장치 부분에는 7255 범용 보드를 사용하였고 이 보드에는 ACL-7255 칩과 16채널 릴레이가 사용되었다. 본 7255 범용보드는 IBM PC 상의 ISA 슬롯에 장착하여 사용하게 되며 37 핀 D-타입 케이블을 통하여 제어 신호를 적외선 리모트 컨트롤러에 전달하게 된다. 이 제어장치에서는 음성인식기에서 인식한 음성 인식 결과를 받아 이를 적외선 리모트 컨트롤러에 전송하는 역할을 하는데, 포트 신호를 이용하여 SSR(solid state relay)을 컨트롤하여 적외선 리모컨을 제어하거나 모니터링용 발광 다이오드로 신호를 디스플레이 한다.

ACL-7255 제어보드의 프로그램은 크게 두 가지로 구성 되어있다. 인식 결과를 받아들이는 인식 처리부와 각각의 하부 기기를 컨트롤하는 제어부로 구성되어는데, 인식 처리부는 음성인식기로부터 인식결과를 수신하며 제어 부는 해당하는 기기를 선택하여 전기신호를 전송하며 기타 입, 출력 장치 컨트롤, 인식부와의 인터럽트를 이용한 통신, 표시등의 기기와의 하부 인터페이스를 지원한다.

2-3 I/O 인터페이싱 부분

I/O 인터페이싱 부분에서는 중증 장애인 환자가 일상적으로 가정에서 사용하는 가정용 전자제품의 제어를 위한 부분으로 대부분의 가전제품의 원격 제어에 사용되는 적외선 리모트 컨트롤 시스템을 이용하였다. 적외선 리모트 컨트롤을 이용할 수 없는 전등의 경우 외부에 릴레이를 설치하여 제어하도록 하였다.

3. 음성 인식과 임상 적용 결과

3-1 음성인식 결과

실험 결과는 아래 표 1 과 같이 화자A는 94.81%, 화자 B는 95.77%, 화자C는 96.15%, 화자D는 95.58%의 인식률을 나타내었다. 표 3-1은 '오 인식 횟수 / 실험 횟수'로 나타내었다. 실험 결과를 살펴보면 유사한 단어간의 오 인식 횟수가 많은 것을 알 수 있으며, 오 인식을 하는 단어를 살펴보면 다음과 같이 3가지로 원인을 분석

할 수 가 있었다. 먼저 발음한 파형이 비슷한 경우, 그리고 입력 파형의 크기가 HMM 학습에 참여한 데이터 크기와 크게 차이가 나는 경우 마지막으로 HMM 학습 데이터의 부족의 경우로 볼 수 가 있다.

각 단어의 HMM을 구성할 때 동일 단어 무리의 통계를 바탕으로 학습하게 되고 학습 데이터가 부족할 경우 그만큼 단어 모델의 신뢰도가 떨어지게 된다. 그러므로 오 인식 될 확률이 높아지게 되는 것이다.

화자 명령어	화자A	화자B	화자C	화자D
텔레비전	0/20	1/20	0/20	0/20
라디오	1/20	0/20	1/20	0/20
테이프	2/20	0/20	1/20	1/20
씨디	0/20	0/20	0/20	0/20
비디오	1/20	2/20	0/20	0/20
채널유	0/20	0/20	2/20	1/20
채널칠	0/20	0/20	0/20	1/20
채널구	0/20	2/20	1/20	0/20
채널십일	1/20	2/20	0/20	0/20
일 변	4/20	3/20	3/20	1/20
이 변	2/20	1/20	2/20	1/20
삼 변	0/20	1/20	0/20	0/20
사 변	1/20	0/20	1/20	2/20
플레이	2/20	0/20	1/20	2/20
정 지	0/20	0/20	0/20	0/20
앞으로	1/20	0/20	1/20	1/20
뒤 로	2/20	3/20	2/20	0/20
소리크게	0/20	0/20	0/20	1/20
소리작게	1/20	2/20	1/20	3/20
조용히	1/20	0/20	0/20	1/20
전 등	2/20	1/20	1/20	0/20
에어컨	0/20	0/20	0/20	1/20
강하게	3/20	1/20	1/20	2/20
약하게	1/20	0/20	0/20	1/20
튜닝업	2/20	3/20	0/20	3/20
튜닝다운	0/20	0/20	2/20	1/20
인식률	94.81%	95.77%	96.15%	95.58%

표 1 인식 실험 결과

3-2 임상 적용 실험 결과

제 4 경수 손상에 의한 사지마비 장애인을 대상으로 임상 적용 실험을 하였고 환자가 거주하는 방에 생활환경 제어장치를 설치하였다. 텔레비전과 오디오는 환자가 평소에 사용한 것을 그대로 이용하였으며 전등과 비디

오는 별도로 제공하였고, 에어컨은 실험할 수 없었다. 실험을 통하여 환자의 행동 변화를 보면 TV의 시청의 경우 채널의 변경과 볼륨의 조작을 위한 보호자의 도움이 불필요하게 되어 장치를 적용하기에는, 하루 약 6시간을 시청하였는데 적용 후 12시간 이상으로 증가하였고, 오디오의 경우 장치 적용 전 거의 사용을 하지 않았으나 TV가 방영되지 않는 시간동안의 청취를 위한 조작이 많아졌다.

장치 적용 후 환자는 보호자의 도움을 위한 보호자 호출 횟수가 하루평균 절반 이하로 줄었으나 장치의 제한적인 사용 한계로 인하여 보호자의 외출시간은 1시간 이내에 머물 수밖에 없었다. 장치의 사용 전과 후의 Beck 우울 척도의 비교에서 대상 장애인의 경우 42점에서 38점으로 보호자의 경우 23점에서 17점으로 감소하였다. 그러나 16점 이상은 우울증 증상으로 보는 Beck 우울 척도의 기준에 따라 우울증으로 볼 수 있다.

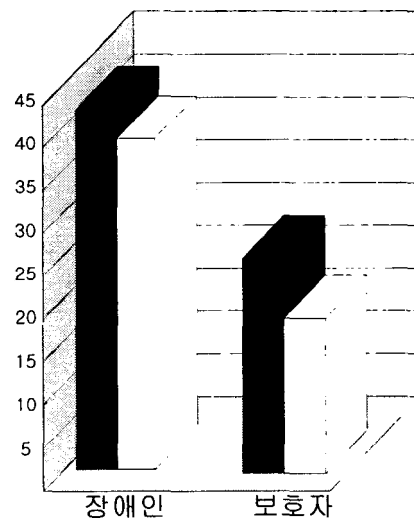


그림 2. 제어장치 적용 후 변화 비교도

4. 결 론

본 논문에서 음성인식을 이용한 중증 장애인의 생활 환경 제어장치를 구현해 보았다. 음성인식기의 구현과 실험을 위해 사용한 장치는 일반적으로 사용하는 IBM PC와 사운드 보드였기 때문에 전문적인 음성 입

력 보드들과는 달리 스위칭 잡음, 시간이나 게인(gain) 조종에 따른 옵셋 변화 등에 의한 잡음이 입력되었다. 또한 일반 잡음과 생체 잡음 등에 의한 잡음도 문제가 되었다. 이러한 문제가 끝점 검출에 미치는 영향을 막기 위해 개선된 끝점검출 알고리즘을 사용하였다. 또한 사람이 발음함에 있어서 학습할 때와 인식할 때 발음의 차이가 많이 나는 경향이 있다. 학습을 할 때는 편안한 마음을 가지고 하나 인식 실험을 할 경우에는 긴장을 하기 때문에 두 발음의 차이가 많아 오 인식하는 경우가 있었으며, 더욱이 적용 대상이 중증 장애인이었기 때문에 인식 학습에 사용되는 음성의 발체에도 어려움이 있었다.

제어부에서는 대부분의 기존 가전 제품에서 사용하는 적외선 리모트 컨트롤러를 사용하였기 때문에 별다른 어려움은 없었으나 제어기의 단어수가 한정되어 보다 많은 단어를 인식시킬 수 없었지만 기본적으로 장애인 가정에서 필요한 단어의 수가 한정되어 있기 때문에 큰 불편은 없었다. 본 논문에서 가전제품의 리모트 컨트롤러 명령어 중 26개의 명령어를 인식 단어로 구성하여 화자 종속 실험을 하여 95.6%의 인식률을 나타내었고 장애인에 직접 적용하였을 때 장애인의 경우 간호 의존도와 우울 경향 감소 및 사회 관계의 증진을 볼 수 있었으며 간병인 보호자의 간호 부담과 우울 경향의 감소를 관찰해 볼 수 있었다. 본 연구를 통하여 중증 장애인을 위한 생활 환경 제어장치의 필요성을 확인 할 수 있었고 가장 편리한 의사소통 수단인 음성을 이용한 한국형 음성인식 생활 환경 제어기의 상용화를 위한 여러 과제들과 그 가능성을 모색하는 단초가 된 것으로 사료된다.

5. 참고 문헌

[1] L. R. Rabinar and B. H. Juang, *Fundamentals of Speech Recognition*, Prentice Hall, 1993.
 [2] S. Saito and K. Nakata, *Fundamentals of Speech Signal Processing*, Academic Press, 1985.
 [3] L. R. Rabiner, J. G. Wilpon and F. K. Soong, *High Performance Connected Digit Recognition Using Hidden*

Markov Models, *IEEE Trans. Acoust., Speech, Signal Processing*, Vol. 37, No. 8, 1989.
 [4] P. D. Wasserman, *Neural Computing : Theory and Practice*, Van. Nostrand Reinhold, 1993.
 [5] John R. Deller, Jr., John G. Proakis and John H. L. Hansen, *Discrete-Time Processing of Speech Signals*, Macmillan Publishing Company, 1993.
 [6] Khalid Sayood, *Introduction to Data Compression*, Morgan Kaufmann Publishers, Inc, 1996.
 [7] Valtcho Valtchev, "Discriminate Methods in HMM-Based Speech Recognition," Univ. of Cambridge, 1995.
 [8] R. M. Gray, "Vector Quantization," *IEEE Trans. Acoust., Speech, Signal Processing Magazine*, April 1984.
 [9] T. W. Parsons, *Voice and Speech Processing*, McGraw Hill, 1986.
 [10] L. R. Rabinar and R. W. Schafer, *Digital Processing of Speech Signals*, Prentice Hall, 1978.
 [11] H. Harmanskey, Kazuhiro Tsuga, Shozo Makino, and Wakita, "Perceptually Based Processing In Automatic Speech Recognition," *ICASSP*, pp.1971~1162, 1986.
 [12] H. Hamansky, "Perceptual Linear Predictive(PLP) analysis of speech," *Journal of Acoustic Society of Am.*, 87(4):1738~1752, 1990.
 [13] G. D. Forney, Jr., "The Viterbi algorithm," *Proc. IEEE*, Vol. 61, No. 3, Mar. 1973.
 [14] B.A. Hanson, T.H. Applebaum, "Subband or Cepstral domain Filtering for Recognition of LOMBARD and Channel-Distorted Speech," *Proc. ICASSP II*, 79 ~ II.82 1993
 [15] 이 행세, *음성 인식*, 1999, 청문각.