다기능 전동휠체어의 음성인식 모듈에 관한 연구

류 홍 석, 김 정 훈, 강 성 인, 강 재 명, 이 상 배
한국해양대학교 전자통신공학과
Tel) 051-410-4907

Voice Recognition Module for Multi-functional Electric Wheelchair

Department of electronic & communication Korea maritime University
Ryu Hong Suk, Kim Jung Hoon, Kang Sung In, Kang Jae myung, Lee Sang Bae
mail : rhs1109@kmaritime.ac.kr

Abstract

This paper intends to provide convenience to the disabled, losing the use of their limbs, through voice recognition technology. The voice recognition part of this system recognizes voice by DTW (Dynamic Time Warping) which is most widely used in speaker dependent system. Specially, S/N rate was improved through Wiener filter in the pre-treatment phase while considering real environmental conditions; the result values of 12th order feature pattern per frame are extracted by DTW algorithm using LPC and Cepsturm in feature extraction process. Furthermore, miniaturization is pursued using TMS320C32, TI's the floating-point DSP, for the hardware part. Currently, 90% of hardware porting has been completed, but we can confirm that the recognition rate was 96% as a result of performing the DTW algorithm in PC.

I. 서론

우리가 일상생활에서 가장 보편적이고, 편리한 정보 수단을 얘기한다면, 그건 아마 음성일 것이다. 이 음성을 통해 기계와 대화를 한다면, 즉 기계 및 사용 장치에 인식 시켜서 동작을 하게 되면, 더욱더 편리하게 이용할 수 있을 것이다. 최근 음성인식 분야는 각종 가전제품, 자동차등 여러 제어 분야에 활발히 연구되어지고 있다. [6]

본 논문에서는 이 음성인식 기술을 이용, 휠체어에 적용시킴으로써 수족이 불편한 장애인에게 편리성을 제공하고자 하는데 그 목적이 있다. [7]

음성은 인식의 형태에 따라 화자종속과 화자독립으로 나눌 수 있고, 알고리즘의 형태에 따라 DTW(Dynamic Time Warping), HMM(Hidden Markov Model), TDNN (Time Delay Neural Network) 등 여러 가지가 있다. 본 논문에서는 화자종속방식에서 비교적 많이 사용되는 DTW방식을 적용해서 음성인식을 시키고 있다. 현재 휠체어에 적용되는 음성은 7개의 단어를 적용하기 때문에 그렇게 큰 메모리 용량을 요구하지 않아 DTW방식을 채용했다.

음성인식 방법은 대용량의 메모리 요구로 인해 기존에는 메모리 용량에 비교적 제한을 받지 않는 PC에서 그 연구가 활발히 이루어졌다. 하지만 여기서는 휠체어라는 점을 감안한다면, PC를 직접 탑재 할 수 없기 때문에 TI사의 부동소수점 범용 DSP인 TMS320C32를 사용하여 소형화를 시도하였다. [1][2][5]

본 논문의 구성은 우선 전체적인 휠체어의 시스템을 알아보고, 그리고 음성인식의 절차와 그리고 하드웨어의 구성을 알아본 후 본 논문에서 실험과 고찰을 통해서 결론을 짓도록 하겠다.

II. 휠체어 시스템의 전체적인 구성

휠체어의 전체적인 시스템은 그림 1과 같다. 현재 휠체어의 구성 상태는 제어부, 모터부, 전원부, 음성인식부로 나누어져 있다.

각 부분을 자세히 살펴보면 제어부는 Intel사의 80C 196KC를 사용하였고, 현재 조이스틱의 제어부분을 담당하고 있다. 모터부는 80C196KC의 HSO를 통해 전달받은 제어신호가 H-브릿지 회로를 통해 제어하는 방식을 취하고 있다. 전원부는 모터부에 사용하는 24V와 각 프로세서들이 사용하는 5V를 고려해서 설계를 했으며, 그리고 음성인식부분은 현재 TI사의 부동소수점 DSP인 TMS320C32를 통해 제어를 하고 있다. [3]

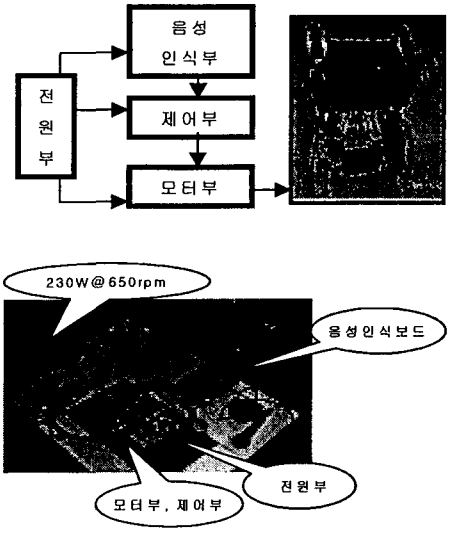

그림 1 휠체어의 전체적인 구성도

III. 음성인식의 절차

음성인식의 절차를 블록도로 표현하면 그림 2와 같으며, 그 과정을 살펴보면 다음과 같다.

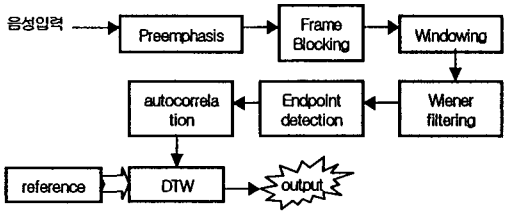

그림 2 음성인식처리 절차

음성은 그 특성상 많은 데이터로 구성되어 있다. 그래서 이 많은 데이터를 컴퓨터 시스템에서 알아 볼 수 있는 데이터로 바꾸어 주어야 된다. 그 과정은 그림 2와 같다. 먼저 Preemphasis는 음성신호의 저주파 성분을 약화시키고 고주파성분을 강조 시켜 음성신호의 DC성분을 제거한다. DC성분은 주로 마이크에서 많이 발생하는 신호이다. 이 Preemphasis는 다음과 같다.

$$H(z) = (1 - az^{-1}) \qquad 0.9 \leqq a \leqq 1.0$$

노이즈(DC)성분을 제거한 후 Frame Blocking을 통해 N개의 샘플로 샘플링을 한 다음 윈도우를 통해 계산을 하는데, 여기에서는 끝단 부분에 노이즈가 비교적 적은 Hamming window를 주로 사용했다. 이 윈도우는 30msec의 길이를 이용했고, Hamming window는 다음과 같이 적용한다. [3][7]

$$W(n) = 0.54 - 0.46\cos(\frac{2\pi}{N-1}) \qquad 0 \leqq n \leqq N-1$$

이 후에 음성은 많은 잡음을 포함하고 있다. 그래서 본 논문에서는 여러 가지 필터링 중에서 응용분야에 가장 많이 적용되고 있는 Winer filtering을 통해 S/N 비를 향상시키고 있다. 이 필터링은 잡음이 포함된 신호 Ym(n)는 전달 함수 H(w)를 통해 잡음이 제거된 원신호가 추출되게 하는 방식이다. [8]

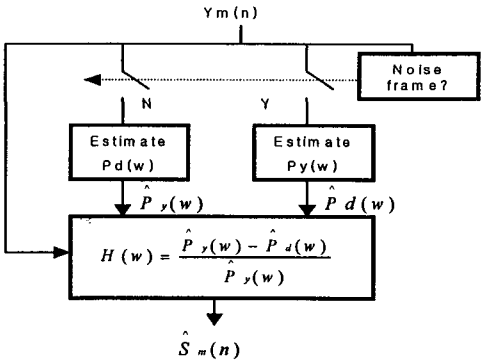

그림 3 Wiener filtering

이제 이렇게 된 후 음성의 구간 검출을 실시해야 된다. 이 과정을 End point detection 이라 하며, 음성인 구간만을 검출해서 음성인식에 적용시키기 위함이다. 여기에는 음성신호의 교차율을 통해 검출하는 Zero Crossing rate와 음성 에너지의 절대적인 크기에 따라 검출하는 Short time energy가 주로 사용되는데, 본 논문에서는 Short time energy를 통해 음성 구간을 검출했다. Short time energy는 음성신호의 유무를 판단하기 위해서 가장 최근에 입력된 음성 샘플들의 평균 에너지를 실시간으로 계산하여 음성신호가 있는 동안만 필요한 처리를 하는 방법으로 그 방법은 다음과 같

이 처리한다. [3][7]

$$Ei = \sum_{n=0}^{N-1} [\, Wi(n) Si(n)]^2$$

현재 본 논문에서는 두 개의 버퍼를 사용하여 계산을 하는데, 한 개의 버퍼는 에너지의 구간을 계산하고, 다른 한 개는 음성을 저장하는 역할을 수행한다. 수행 중에 음성 샘플이 에너지를 구하고자 하는 프레임 수가 되면 두 버퍼의 역할을 교대하여 수행 하게끔 했다. 음성구간을 검출한 후에는 LPC(Linear Prediction Coefficient)를 취하기 이전에 Autocorrelation을 취해 주는데 이것은 LPC분석을 안정되게 하기 위해서이다. LPC는 과거의 음성 샘플을 가지고 현재의 음성샘플을 예측하는 방법으로 사람의 성도를 모델로 한 필터로서 여기에서 발생된 계수를 음성의 특징 계수로 사용하는 방식이다. LPC 알고리즘은 Durbin method를 사용 했다. 여기에 관련된 알고리즘은 다음과 같다. [7]

$$r(m) = \sum_{n=0}^{(N-1)-m} x(n)x(n+m) \qquad m = 0, 1, 2, 3....P$$

$$E^{(0)} = r(0)$$

$$ki = \{r(i) - \sum_{j=1}^{L-1} a_j^{(i-1)} r(|\, i-j\,|\,)\}/E^{(i-1)} \quad 1 \leqq i \leqq P$$

$$a_i^{(i)} = ki$$

$$a_j^{(i)} = a_j^{(i-1)} - kia^{(i-1)}{}_{i-j}$$

$$E^{(i)} = (1 - ki^2) E^{(i-1)}$$

다음은 Cepstrum과정인데, LPC과정까지 통과한 음성은 비선형적으로 구성되어있다, 이것을 선형적으로 바꾸어 주는 과정이 Cepstrum이다. 본 논문에서는 LPC처리된 음성을 12차 Cepstrum으로 바꾸었는데, 그 과정을 보면 다음과 같다.

$$C_1 = -a_1$$

$$Cn = -a_n + \sum_{m=1}^{n-1} \frac{m}{n} a_m C_{n-m} \qquad (1 < n \leqq P)$$

$$D_n = \sum_{m=1}^{n-1} \frac{m}{n} a_n C_{n-m} \qquad (P < n)$$

위의 전 처리 과정을 거친 음성은 미리 저장된 reference와 비교해서 output을 생성하는데, 이 때 사용되는 방식이 DTW이다. 음성인식 처리루틴은 그림 4와 같이 설계했다. DTW라는 것은 입력 패턴과 참조 패턴 사이의 거리를 계산해서 그 유사도를 측정하는 방식으로 즉 제한된 경로 내에서 단조 증가를 해서 가장 가까운 거리를 판별해서 유사도를 측정한다. 그 식은 다음과 같다.

$$a_c(x_i y_i) = g_d(x_i y_i) + min\{a_d(x_{i-1}, y_i), a_d(x_{i-1}, y_{i-1}), a_d(x_{i-1}, y_{i-2})\}$$

위의 식을 보면 최종 누적거리는 현재의 거리와 이전거리의 합으로 계산되어지는 것을 알 수 있다. 인식되어진 값은 각 패턴으로 나누어져 모터 루틴의 신호로 사용되어 진다. [3][7]

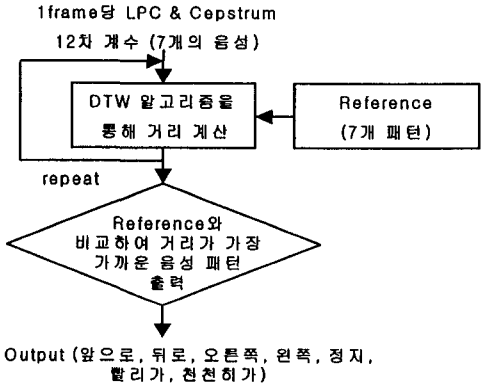

그림4 DTW 처리 절차

IV 하드웨어의 구성

하드웨어처리의 블록도는 그림 5를 보면 잘 나타나 있다.

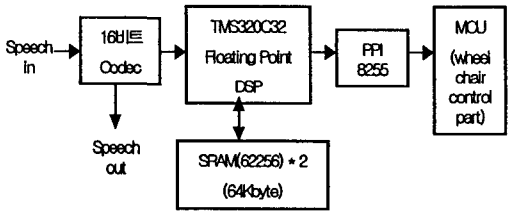

그림 5 음성인식의 하드웨어 구성

그 구조는 16비트 코덱과 TMS320C32 floating point DSP, SRAM(32K * 2), PPI 8255로 구성되어진다. 16비트 코덱은 아날로그신호 즉 음성을 16비트로 샘플링 시켜서 serial통신으로 DSP에 전달한다. 음성은 비교적 많은 데이터를 보유하므로, 빠른 처리 능력을 요하는 프로세서가 필요하다. 여기에 PC가 가장 적합

하지만 휠체어라는 점을 감안해서 빠른 처리(30MIPS) 능력을 요하면서, 대용량의 메모리(최대 16M) 확장이 가능한 TMS320C32를 통해 음성 신호를 처리한다.

TMS320C32에는 총 2개의 DMA가 있다. 이 DMA를 통해 하나는 음성의 저장을, 다른 하나는 음성의 재생을 담당 시키고, 내부 인터럽트를 통해 음성을 메모리에 적재한다. 음성의 저장은 short time energy를 이용하여 음성의 구간이 검출된 부분만 저장을 시킨다. 이때 음성인식보드에 있는 스위치에 따라 인식모드와 학습모드로 나누어지는데, 우선 보드에 있는 S2(S1은 리셋)를 누르면 학습모드로 동작하여, 7개의 패턴을 만든다. 다음에 S3를 눌러 인식모드로 동작 시켜 음성을 저장하면, 이미 저장된 7개의 패턴과 비교하여 가장 가까운 인식 값을 출력시킨다. 이 값을 PPI 8255를 통해 병렬로 휠체어의 제어기 Intel 80C196KC에 전달되어진다. 전달 받은 데이터를 통해 모터의 각 루틴을 실행한다.

V. 실험 및 고찰

현재 하드웨어의 포팅 작업은 90%정도 진행 중이며, 본 실험은 PC에서 구현한 음성인식을 통해 나타난 결과 값을 출력했다. 이 실험은 제안된 모델 7개의 명령어에 대한 인식 실험을 실시했다. 여기서 1사람의 목소리로 각 단어를 3번씩 발음 했고, 이 때 음성은 8Khz샘플링, 분해능은 16bit로 저장 했다. 다음 보는 표는 결과 값으로, 각 단어의 distance값을 나타내었다.

표에서 보는 바와 같이 숫자가 낮은 부분이 인식된 결과 값을 나타내며, 10으로 표기 된 부분은 입력된 음성이 reference음성과 전혀 맞지 않는 부분을 나타낸다.

	앞으로	뒤로	오른쪽	왼쪽	정지	빨라가	천천히가
앞으로1	1.24	2.37	2.55	2.37	3.66	2.81	2.95
앞으로2	1.86	2.72	2.5	2.62	4.08	3.03	3.28
앞으로3	2.01	2.64	2.44	2.52	3.94	2.98	2.91
뒤로1	2.37	1.45	3.01	2.12	4.18	2.89	10
뒤로2	3.01	1.84	3.48	2.71	4.4	3.11	10
뒤로3	2.62	1.96	3.17	2.63	3.78	3.16	10
오른쪽1	2.55	3.01	0.99	2.74	3.48	3.29	3.15
오른쪽2	2.65	3.11	2.21	2.73	3.2	3	2.86
오른쪽3	2.68	3.18	2.01	2.74	3.07	3.23	3.19
왼쪽1	2.37	2.12	2.74	1.34	3.11	2.92	3.03
왼쪽2	2.35	2.15	2.57	1.77	2.94	2.79	2.96
왼쪽3	2.74	2.25	2.61	1.7	3.63	3.05	2.71
정지1	3.66	4.18	3.48	3.11	1.47	3.62	3.86
정지2	3.7	4.24	3.49	3.09	1.68	3.71	3.93
정지3	3.82	4.09	3.39	3.07	1.86	3.39	3.57
빨라가1	2.81	2.89	3.29	2.92	3.36	1.08	2.91
빨라가2	2.94	3.45	3.32	3.1	3.64	1.79	2.77
빨라가3	2.99	3.3	3.19	3.1	3.78	1.83	2.73
천천히가1	2.95	10	3.15	3.03	3.86	2.91	1.27
천천히가2	2.93	10	3.19	3.12	3.53	2.67	1.78
천천히가3	2.93	10	3.04	3.14	3.68	2.84	1.64

<표> 입력된 음성과 출력된 음성의 비교

VI. 결론 및 향후과제

본 연구는 화자종속에서 가장 많이 사용되는 DTW 알고리즘을 통해 휠체어 시스템에 접목시킴으로써 수족이 불편한 장애인에 대해 편리성을 제공하고자 하는 것이었다.

특히 실제 환경에서 잡음을 고려해서 Wiener filter를 사용함으로써 약간은 개선된 인식률을 볼 수 있었다.

이제 향후과제로는 현재 DSP에 포팅이 완벽하게 이루어지지 않아 완벽한 포팅을 할 예정이며, HMM을 통해서 화자독립으로도 구현해 볼 예정이다. 그리고 좀 더 실제 환경에서 강한 필터도 적용해 볼 예정이다.

참고 문헌

[1] 김창근, 한학용, "TMS320C32를 이용한 실시간 음성인식 무선자동차의 구현", 대한시스템 학회, 2001

[2] 정익주, 정훈 "TMS320C32 DSP를 이용한 실시간 화자 종속 음성인식하드웨어모듈(VR32)의 구현", 한국음향 학회 Vol.17, No.4.14-22, 1998

[3] 박준혁, "음성인식이 가능한 이동 로봇에 관한 연구", 부산대학교 석사학위 논문, 1998

[4] 손영선, 추명경, "DTW방식을 이용한 음성 명령에 의한 커서 조작", 퍼지 및 지능시스템학회 논문지 2001, Vol. 11, No 1, pp 3-8

[5] 이지홍, "DSP chip의 활용", 서일 DSP기술연구서 공저, 2000

[6] 오영환, 음성언어정보처리, 홍릉과학출판사, 1998

[7] Satoru Nakanishi, Yoshinori Kuno, "Robotic Wheel chair Based on Observations of Both User and Environment", IEEE/RSJ International Conference on Intelligent Robots and System, 1999.

[8] Lawrence Rabiner, Biing-Hwang Juang, "Fundamentals of speech recognition", Prentice Hall International Inc. 1993

[9] Steven L.Gray, "Acoustic signal processing for Telecommunication", Kluwer Academic Publishers, 2000