

## 하이브리드형 자율신경망을 이용한 임베디드 음성인식시스템의 설계 및 구현

류홍석\* · 강성인\* · 이상배\*\*

\*한국해양대학교 대학원, \*\*한국해양대학교 전자정보통신공학부

### Design and Implementation of Embedded Speech Recognition System Using the Autonomic Nervous System of Hybrid type

H. S. Ryu\* · S. I. Kang\* · S. B. Lee\*\*

\*Graduate school of National Korea Maritime University, Busan 606-791, Korea

\*\*Division of Radio and Information Communication on Engineering National Korea Maritime University

**요 약 :** 음성인식 기술은 기계가 인간의 언어를 받아 들여, 그 언어에 맞게 적절한 행동을 수행하는 것을 말한다. 이런 기술은 산업 전반에 사용되어지고 있으며, 특히 정보 산업분야, 디지털 통신 분야, 가전 분야, 멀티 미디어 등에 적용되어지고 있다. 이러한 기술을 통해서 본 연구실에서도 이동로봇 즉 전동휠체어 시스템에 적용, 손발이 불편한 장애인에게 좀 더 편리성을 주고자 이 시스템을 개발하게 되었다. 본 연구에서는 플랜트가 전동휠체어라는 것을 고려해서, 음성인식 중 화자 종속이면서 비교적 인식률이 높은 DTW를 사용하고 있다. 그러나 실시간이란 것을 고려해서 적은 메모리와 빠른 처리속도의 성능을 가져야 한다. 그래서 화자 독립의 압축 알고리즘으로 많이 사용되는 벡터 양자화를 도입해서 빠른 인식과 적은 메모리를 확보할 수 있었다. 하지만 벡터 양자화를 사용함으로써 인식률이 저하되는 것을 발견했다. 그래서 인식률 향상을 위해서 후처리 알고리즘으로 ART2 알고리즘을 적용, 약 5%정도의 인식률 향상을 가져왔다. ART2를 사용하기 위해서는 오차범위라는 것이 적용되어야 된다. 오차범위는 DTW를 적용하기 위해서 나온 각 distance에 대해, 2순위 distance에서 1순위 distance를 뺀 결과가 20이상이면 오차범위가 적용된다. 이렇게 ART2를 적용해서 빠른 처리와 높은 인식률을 가져올 수 있었다. 그리고 이동되는 물체이므로, 임베디드로 시스템을 구현해야 한다. 그래서 많은 계산량을 비교적 빠르게 처리하는 TMS320C32라는 칩을 선정해서 임베디드 시스템을 구현하고 있다. 메모리도 음성이라는 것을 고려해서 많은 음성 데이터를 저장할 수 있게 128kbyte의 RAM 메모리와 64kbyte의 ROM 메모리를 확보 시켰다. 음성의 입력은 16비트 스테레오 오디오 코덱을 사용, 높은 분해능을 통해서 비교적 정확한 데이터를 확보할 수 있다.

**핵심용어 :** 음성인식, 임베디드, ART2, VQ, DTW

**ABSTRACT :** *Speech recognition technology called that machine receives human's language and achieves suitable motions according to this language. The electric motion wheelchair system applied this technology is developed by this LAB for giving more convenience to physically handicapped people especially hand and feet uncomfortable disable people. This study consider that the plant is electric motion wheelchair. So this system is used DTW(Dynamic Time Warping). Because it used DTW being higher recognition. But it has to get small memory and fast processing speed considering real time. So it introduce VQ(Vector Quantization) used in data compression algorithm of speaker independent. Because It secure fast recognition and small memory. But it discovered that recognition rate was fallen by using VQ. For that reason, it is applied ART2 algorithm as after treatment algorithm. This result, it enhance recognition rate about 5%. To use ART2 must be applied error range. Error range is applied that result extracting 1th order distance in 2nd order distance is more than 20 by each distance to apply DTW. In this way it applying ART2 has fast processing and high correct recognition rate. And It must implement by embedded system because it is moving object. For that reason, this study implement embedded system choosing chip of TMS320C32 that can relatively process fast and a lot of computation complexity. RAM of 128Kbyte and ROM of 64Kbyte can store a lot of voice data to save memories. Input of speech use 16bits stereo audio codec to secure correct data through high resolution.*

**KEY WORD :** Speech Recognition, embedded, ART2, VQ, DTW

며, 그 중 한 방법인 음성을 이용한 인식이다. 음성 인식 기술은 기계가 인간의 언어를 받아들여, 그 언어에 맞게 적절한 행동을 수행하는 것을 말한다. 이런 기술은 각종 산업 전반에 사용되어지고 있으며, 특히 정보 산업 분야, 디지털 통신 분야, 가전분야, 멀티미디어 등에 적용되어지고 있다. 이러한 기술을 통해 본 연구실에서도 이동로봇(전동휠체어 시스템)에 적용, 손발이 불편한 장애인에게 좀더 편리성을 주고자 이 시스템을 개발하게 되었다.

음성인식 방법으로는 패턴의 매칭을 거리에 의해서 하는 DTW(Dynamic Time Warping)와 통계적으로 인식하는 방법인 HMM(Hidden Markov Model), 뇌의 구조를 모델링한 방법인 NN(Neural Network) 등이 있으며, 인식 방법에 따라 한 사람에게 적용되는 화자종속형과 여러 사람에게 적용되는 화자독립형으로 나눈다. 본 논문에서는 전동휠체어가 장애인들에게 적용된다는 것을 감안해서 화자종속형으로 설계했으며, 인식률이 비교적 좋은 DTW를 전반부 인식 알고리즘으로 사용하였으며, 오 인식을 보정하기 위해 이전에 학습되어진 패턴을 유지하면서 새로운 패턴을 학습하는데 필요한 유연성을 잃지 않도록 설계되어져 마치 분류기처럼 사용되어 지는 ART2(Adaptive Reason Theory 2) 알고리즘을 후반부 인식 알고리즘으로 적용했다.

본 논문의 구성은 5장으로 구성되어 있으며, 2장에서는 실음성 구간이 어떻게 검출되는지를 살펴보고, 많은 음성데이터의 특징부분만을 검출하는 방법인 MFCC를 다루며, 특징 추출된 벡터를 압축하기위해 VQ(Vector Quantization)를 사용하였다. 3장에서는 인식알고리즘인 DTW와 ART2에 대해 설계한 것을 서술한다. 4장에서는 이 알고리즘으로 만든 임베디드형 음성인식보드와 소프트웨어적으로 어떻게 처리되어지는지를 서술, 5장에서는 앞장을 근거로 실제 실험과 결과에 대해서 설명을 하고 6장에서 결론을 맺도록 하겠다.

## 2. 음성구간감지 및 특징 추출

### 2.1. 실음성 구간 검출

실음성 구간 검출이라는 것은 실제음성을 추출하는 것을 말한다. 즉, 실시간적으로 들어오는 음성을 음향신호를 음성과 잡음으로 구별하고, 음성이 검출된 부분에 대해서는 프레임으로 분할, 각 프레임을 계산하는 방식을 말한다. 본 논문에서는 절대 에너지 방식을 사용하고 있다. 절대 에너지방식은 다음 식(2.1)과 같이 계산되어 진다.

이 부분을 실제음성으로 사용하고 있다. 본 논문에서 사용한 실음성 구간 검출 절차는 그림 2.1과 같다.

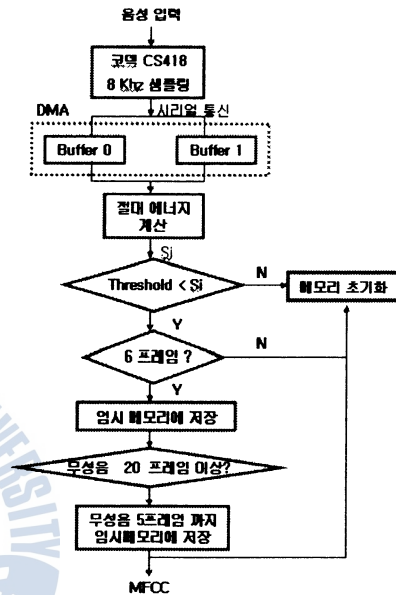


Fig 2.1 Region Detection of Real Speech

코덱 CS4218을 통해서 음성이 들어온다. 이 때 샘플링은 8Khz로 하고 있다. 들어온 음성은 DMA(Direct Memory Access)채널을 통해서 임시 저장된다. DMA채널은 두 개의 버퍼를 할당한다. 두 개의 버퍼작용은 다음과 같다. 한 개의 버퍼로 음성 데이터가 입력되면, 다른 버퍼는 절대 에너지를 계산하도록 설계되어 있다. 이 때 버퍼사이즈(프레임 사이즈)는 256으로 설정했다. 프레임 블록킹은 80으로 설정하였다. 그리고 음성/잡음 구별을 위해서 Threshold값을 설정하였다. Threshold값은 50000으로 설정하고 있다. 이 값을 넘은 데이터는 음성으로 간주되어지고 데이터는 6프레임 이상 되면 그 부분을 음성으로 간주하고 임시메모리에 저장한다. 이후 무성음 구간이 20프레임 이상 되면 끝점에 존재하는 무성음 구간을 고려하여 5프레임만 저장시키고, 나머지 15프레임은 버리는 방식을 사용하고 있다.

### 2.2. 전처리 단계 및 벡터 양자화 [1][3][5][10]

2.1에서 검출된 음성은 MFCC 특징벡터로 변환된다. 본 논문에서는 12차원의 MFCC벡터를 이용하였다. MFCC는 그림 2.2와 같이 얻어진다.

먼저 프리엡과시스는 음성의 저주파 성분을 약화시키고, 고주파 성분만을 강조시켜 음성신호의 DC

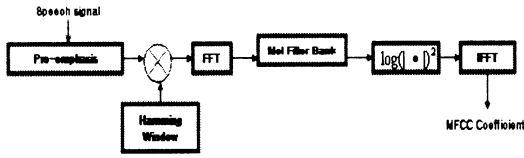


Fig 2.2 Process of MFCC

본 논문에서는 a의 계수를 0.95로 설정하고 있다. 프리엡파시스와 프레임 블록킹을 통해서 프레임별로 잘라내고 중첩되어진 음성데이터는 각 프레임의 시작점과 끝점에서 신호 불연속을 최소화 시켜야 된다. 그래서 양 끝에 비교적 노이즈가 적은 해밍 윈도우를 사용하고 있다. 처리되어진 음성은 멜 캡스트럼을 사용해서 특징벡터를 만들고 있다. 멜 캡스트럼은 인간의 청각이 피치를 선형적으로 이해하지 못한다는 데 착안을 두고 설계된 것이다. 그래서 1000Hz에 1000mel이 되도록 설정하고, 인지된 피치가 기준 주파수의 두 배가 되면 2000mel로 표기했다. 이 대응 관계를 통해서 실제 물리적인 주파수와 인지된 주파수 사이에 관계를 구하면 식 (2.3)과 같다. 1Khz이하에서는 선형적으로 1Khz이상에서는 대수적 즉 로그스케일로 대응되는 관계를 사용하고 있다.

$$F_{mel} = 2595 \log_{10} \left( 1 + \frac{F_{Hz}}{700} \right) \quad (2.3)$$

( $F_{mel}$  : 각각 근사식에 의해 구해진 인지된 주파수  
 $F_{Hz}$  : 실제 주파수)

멜 캡스트럼은 DFT또는 FFT 크기를 멜과 주파수의 대응관계에 따라 주파수 축에서 와핑(warping)하여 이의 대수 값을 역 DFT/FFT하여 8에서 14차 사이에 계수를 구한다.

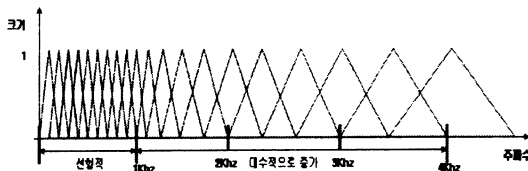


Fig 2.3 Bandwidth of Triangle filter

그림 2.3에서 나타난 20개의 삼각 대역 통과필터를 이용하여 임계 대역 필터를 통과한 로그 에너지 출력을  $X_k$  라 하면 M개의 캡스트럼 계수는 다음 식 (2.4)와 같이 표현한다.

Transform)를 통해서 한 프레임 당 12개의 계수 값을 생성한다.

프레임 당 12개의 값을 생성하게 된 음성은 VQ(Vector Quantization)를 통해서 차원 수를 줄이게 된다. 본 논문에서는 MFCC 계수 값을 이용해서 양자화 테이블을 생성하게 된다. 양자화 테이블 생성 알고리즘은 LGB 알고리즘을 사용하고 있다. 벡터 양자화의 처리는 다음 그림 2.4와 같다

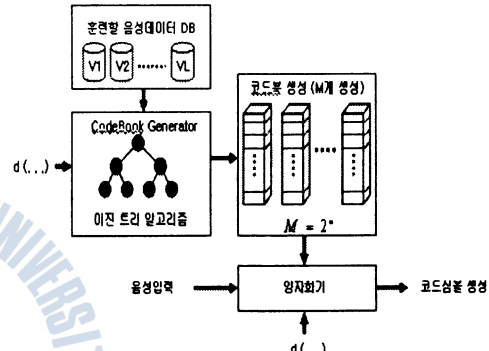


Fig 2.4 Processing of Vector Quantization

VQ의 순서는 그림 2.5와 같다.

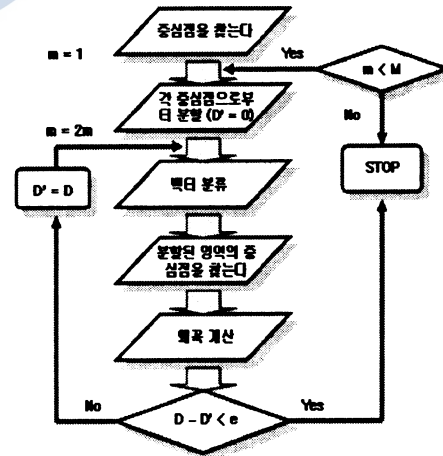


Fig 2.5 Procedure of Binary Tree Algorithm

- Step1) 초기화 하나의 centroid 생성
- Step2) 분할 Centroid의 수가 두 배가 되도록 분할

$$Y_n = Y_n(1 + e), Y_n = Y_n(1 - e) \quad (1)$$

( $e$  : splitting parameter ( $0.01 < e < 0.05$ ),  $Y_n$  : 벡터)

$$C_n^{(i+1)} = \frac{\sum Q(X_m) = C_n^{(i)} X}{\sum Q(X_m) = C_n^{(i)} 1} \quad (2)$$

( $X_m$ : 훈련할 데이터,  $Q(X_m)$ : 현재 소속 코드북)

**Step4) 갱신** 각 부공간을 학습자료를 통해서 갱신

$$D_{ave}^{(i)} = \frac{1}{M} \sum_{m=1}^M \|X_m - Q(X_m)\|^2 \quad (3)$$

( $D_{ave}^{(i)}$ : 평균 왜곡,  $X_m$ : 훈련할 데이,  $Q(X_m)$ : 현재 소속된 코드북,  $M$ : 소속된 코드북 총 수)

**step5) 종료**

조건 1) 현재의 전체 왜곡이 이전의 전체 왜곡과의 차이가 허용범위보다 작으면 종료, 아니면 step3)에서 계속

조건 2) 원하는 코드북보다 작으면 step2), 아니면 종료

생성된 코드북은 새로운 입력벡터들과의 사이에 최적의 코드북의 심볼(index)을 출력한다. 최적의 코드북 심볼은 식 (2.5)와 같다.

$$M^* = \arg \min_{1 < m < M} d(v, y_m) \quad (2.5)$$

$M^*$ : 최상의 벡터 인덱스 값 행렬

$Y_m$ :  $M$ 차원의 벡터 코드북 ( $1 \leq m \leq M$ ),

$v$ : 특징추출된 벡터 입력 값)

본 논문에서는 원하는 코드북의 크기를 512로 설정하고 있으며, 에러 비율은 0.01,  $e$  (splitting parameter)는 0.001로 하고 있다. 생성된 코드북은 새로운 입력벡터들과의 사이에 최적의 코드북의 심볼(index)을 출력한다. 최적의 코드북 심볼은 식 (2.5)와 같다.

$$M^* = \arg \min_{1 < m < M} d(v, y_m) \quad (2.5)$$

$M^*$ : 최상의 벡터 인덱스 값 행렬

$Y_m$ :  $M$ 차원의 벡터 코드북 ( $1 \leq m \leq M$ ),

$v$ : 특징추출된 벡터 입력 값)

본 논문에서는 원하는 코드북의 크기를 512로 설정하고 있으며, 에러 비율은 0.01,  $e$  (splitting parameter)는 0.001로 하고 있다.

### 3. 음성 인식

**Step1) 초기화**:  $D_A(1,1) = d(1,1)m(1)$

**Step2) 반복**:  $1 \leq i_x \leq T_x, 1 \leq i_y \leq T_y$  인

$i_x, i_y$ 에 대해

$$D_A(i_x, i_y) = \min_{(i'_x, i'_y)} [D_A(i'_x, i'_y) + \zeta((i'_x, i'_y), (i_x, i_y))] ]$$

**Step3) 실행**:  $d(x, y) = \frac{D_A(T_x, T_y)}{M^*}$

본 연구에서는 지역거리는 유클리드 거리법을 사용하였고, 지역 제약 방법은 ITAKURA 방식을 사용하고 있다.

### 3.2. ART2(Adaptive Reason Theory2) [7][8][9]

ART2는 이진 입력패턴 뿐만 아니라 아날로그 입력 패턴에 대해서도 학습이 가능한 알고리즘이다. ART2는 F1층을 여러 개의 서브계층(W, X, U, V, P, Q)로 나누어서 피드백과 피드포워드 처리를 한다. ART2는 그림 3.1과 같이 하부계층과 이득제어부분으로 이루어져 있다.

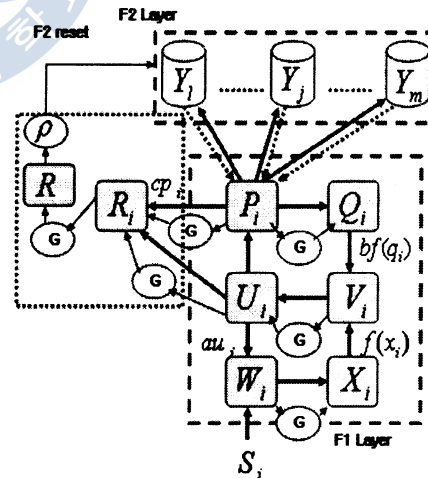


Fig 3.1 Configuration of ART2

ART2는 다음과 같이 총 11단계로 이루어 진다.

**step1)** 파라미터들을 초기화 한다.

$$a > 0, b > 0, 0 < c \leq 1, 0 \leq d \leq 1, 0 \leq \theta \leq 1, 0 \leq \rho \leq 1, e \ll 1, \frac{cd}{1-d} \leq 1$$

하향 연결 가중치:  $Z_{ij}(0) = 0$

Step4) F1층의 유닛의 활성화 및 재활성화

F1층 유닛의 활성화	F1층 유닛의 재활성화
$u_i = 0$	$u_i = \frac{s_i}{e + \ v\ }$
$x_i = \frac{s_i}{e + \ s_i\ }$	$x_i = \frac{w_i}{e + \ w_i\ }$
$w_i = s_i$	$w_i = s_i + \alpha w_i$
$q_i = 0$	$q_i = \frac{p_i}{e + \ p\ }$
$p_i = 0$	$p_i = u_i$
$v_i = f(x_i)$	$v_i = f(x_i) + bf(q_i)$

$w_i = s_i + \alpha u_i$
$p_i = u_i + dz_{ij}$
$x_i = \frac{w_i}{e + \ w\ }$
$q_i = \frac{p_i}{e + \ p\ }$
$v_i = f(x_i) + bf(q_i)$

그 순서도는 다음 그림 3.2와 같다.

Step5) F2층에서 각 노드에 대해 식 (1)에 의해 계산

$$y_j = \sum_{i=1}^n p_i z_{ji} \quad (1)$$

Step6) reset이 설정되면 Step7)~Step 8)을 반복한다.

Step7) F2의 노드 중 최대 출력 값을 갖는 J를 선택한다. 그리고 reset의 여부를 확인하기 위해서 r을 계산한다.

Step8) reset의 조건을 검사한다.

조건	조건식
	$u_i = \frac{v_i}{e + \ v\ }, p_i = u_i + dz_{ij}, r_i = \frac{u_i + \alpha p_i}{\ u\  + \alpha \ p\ }$
	If, $\ r\  < \rho - e$ then $y_j = -1$ reset is true ; repeat step 5
	If $\ r\  < \rho - e$ , then
	$w_i = s_i + \alpha u_i, x_i = \frac{w_i}{e + \ w_i\ }, q_i = \frac{p_i}{e + \ p\ }, v_i = f(x_i) + bf(q_i)$

Step9) reset이 되면 Step5)으로 가고, 그렇지 않으면 Step10)~ Step11)을 수행한다.

Step10) F2 계층의 승자노드의 상, 하향 가중치를 조절한다.

$$Z_{ji} = \alpha d u_i + \{1 + \alpha d(d-1)\} Z_{ji}$$

$$Z_{uj} = \alpha d u_i + \{1 + \alpha d(d-1)\} Z_{uj}$$

Step11) 입력 벡터들을 제거하고, 비활성화된 F2계

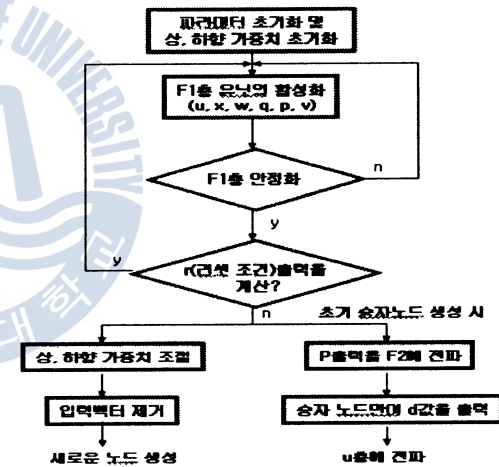
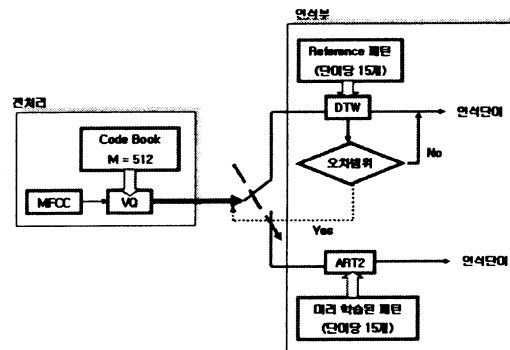


Fig 3.2 ART2 flowchart

3.3. DTW+ART2인식 알고리즘

본 연구는 DTW와 ART2를 사용한 인식방법을 사용하고 있다. 그구조는 그림 3.3과 같다.



범위를 적용해서 오차범위와 일치되면 ART2로 전환하는 방식을 취하고 있다. 일단 이 범위가 존재하려면 1순위와 2순위 distance값이 일치하지 않는다는 전제조건을 가져야 된다. 또 두 distance 값을 빼서 나온 것이 20이상일 경우는 ART2를 적용한다. 적용된 데이터는 어떤 단어의 데이터인지를 결정한다. 만약 레퍼런스에 없는 단어가 생성되면 인식을 중지한다. 오차범위는 그림 3.4와 같다.

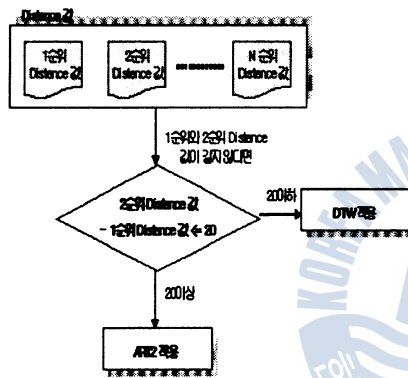


Fig 3.4 Error Range

#### 4. 임베디드 음성인식보드 [4]

음성은 그 특성상 많은 데이터를 가지고 있다. 그리고 많은 데이터에 대해 압축, 또 특징 값을 찾아내기 위해서는 빠른 산술 처리능력을 가진 칩이 필요하다. 본 논문에서는 이런 점을 감안해서 그림 4.1과 같은 음성인식 보드를 설계했다.

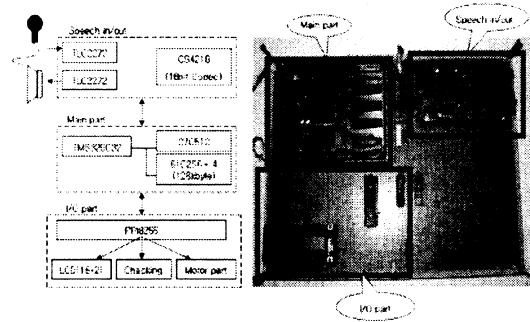
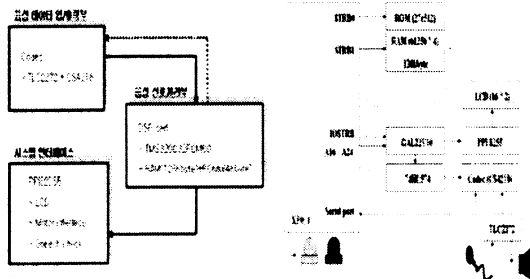


Fig 4.2 Real Speech Recognition Board

보드의 대략적인 구성을 보면, 16비트 분해능을 가지고 있는 CS4218을 통해서 8Khz로 시리얼로 DSP인 TMS320C32에 전달된다. 전달된 음성 데이터는 DSP내부의 산술연산 처리에 의해서 처리를 수행한다. 음성은 그 특성상 많은 메모리를 확보해야 된다는 점을 감안해서 RAM은 128Kbyte (61256\*4)를 ROM은 64Kbyte(27C512)를 사용하고 있다. 인식되어진 최종 값은 I/O 인터페이스 (PPI8255)를 통해서 LCD와 이동로봇으로 전달되어진다.

#### 5. 실험 및 고찰

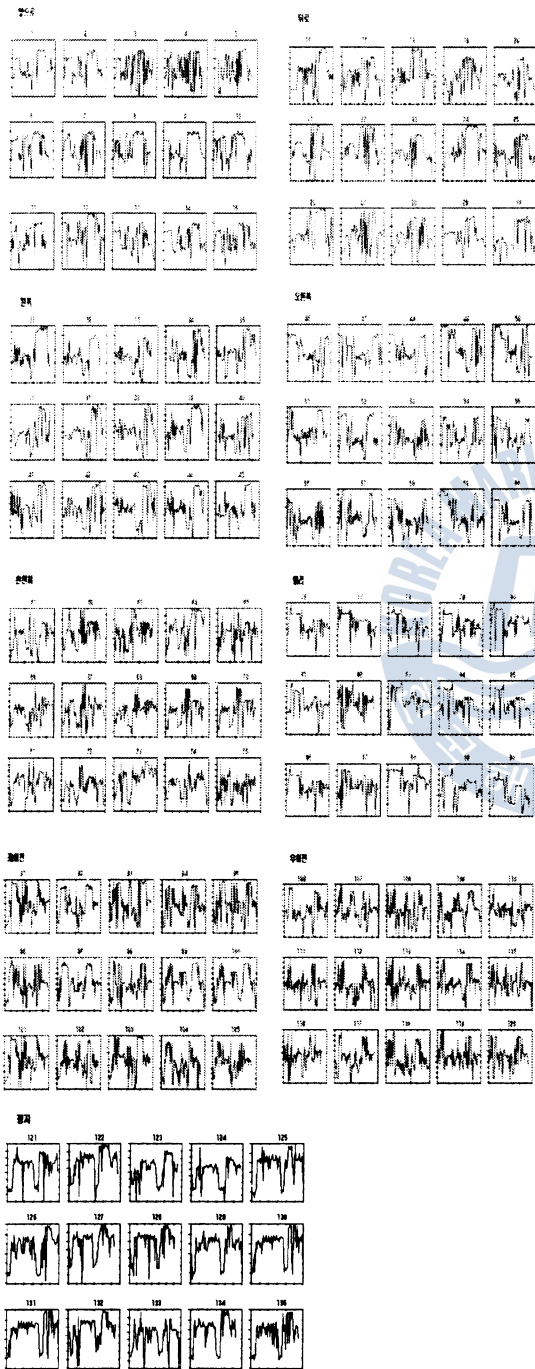
본 장은 음성인식 시 변화되는 과정과 인식률에 대해 측정했다. 본 측정은 실시간 상에서는 빠른 데이터 처리와 시뮬레이션이 불가피해서, 오프라인에서 측정한 것을 토대로 했으며, 인식부분에서는 실시간으로 처리된 것을 중심으로 DTW와 ART2사용시를 비교해서 인식률 차이에 대해서 실험해 보았다.

##### 5.1. 실험 조건

이동로봇의 진행 방향에 따른 발성 총 135번 발성.

앞으로	1 ~ 15	뒤로	16 ~ 30
왼쪽	31 ~ 45	오른쪽	46 ~ 60
천천히	61 ~ 75	빨리	76 ~ 90
좌회전	91 ~ 105	우회전	106 ~ 120
정지	121 ~ 135		

하이브리드형 자율신경망을 이용한 임베디드 음성인식시스템의 설계 및 구현



앞으로	(1)	1, 2, 3, 6, 7
	(2)	4, 8, 13
	(3)	5, 9, 10, 15
	(4)	11, 12, 14
뒤로	(5)	16, 17, 27
	(6)	18, 19, 20, 21, 26, 29
	(7)	22, 23, 24, 25, 28
	(8)	30
왼쪽	(9)	31, 32, 33, 35, 37, 39, 44, 45
	(10)	34, 36, 38, 41
오른쪽	(11)	40, 42, 43
	(12)	46, 47, 48, 49, 50, 56, 57
천천히	(13)	51, 52, 54
	(14)	53, 55
	(15)	58, 59, 60
	(16)	61, 62, 68, 69, 71, 75
빨리	(17)	63, 64, 66, 67, 72
	(18)	65, 70, 73, 74
	(19)	76, 77, 78, 79, 85, 86, 88, 90
좌회전	(20)	80, 81, 82, 87, 89
	(21)	83, 84
	(22)	91, 92, 93
	(23)	94, 95, 96
	(24)	97, 98, 99
	(25)	100, 101, 104
	(26)	102
	(27)	103, 105
우회전	(28)	106, 108
	(29)	107, 119, 120
	(30)	109, 110
	(31)	111, 114, 118
	(32)	112, 117
	(33)	113
	(34)	115, 116
정지	(35)	121, 122, 123, 124, 125, 126
	(36)	127, 131, 132, 134
	(37)	128, 129, 130, 133, 135

( ): 승자노드

ART2 파라미터 초기화는 다음과 같다.

- 상수 :  $a = 10, b = 10, c = 0.1, d = 0.995,$
- 학습인자 :  $\alpha = 0.129, \beta = 0.2, M = 50,$
- 에러율 :  $e = 0.01$

Fig 5.1 Preprocessing of Input Speech

### 5.4. 음성인식

인식 실험은 실 시간적인 환경에서 테스트를 하였으며, 표 2에서는 ART2만을 이용하여 DSP 보드에서 부류론 한 결과로 오프라인 상태보다는 부류론

Table 2. Recognition of ART2

	1	2	3	4	5
앞으로	(4)	(1)	(3)	(3)	(x)
뒤로	(6)	(5)	(7)	(5)	(3)
왼쪽	(11)	(9)	(9)	(10)	(9)
오른쪽	(x)	(x)	(13)	(14)	(13)
천천히	(16)	(17)	(16)	(18)	(18)
빨리	(19)	(19)	(20)	(21)	(19)
좌회전	(25)	(22)	(26)	(24)	(x)
우회전	(33)	(34)	(x)	(31)	(x)
정지	(35)	(36)	(37)	(37)	(35)

(x): 오인식

Table 3. Result of Speech Recognition Test

명령어	횟수	a) DTW		b) DTW+ART2	
		인식수	인식률	인식수	인식률
앞으로	50	46	92%	49	98%
뒤로	50	45	90%	49	98%
왼쪽	50	46	92%	48	96%
오른쪽	50	47	94%	49	98%
천천히	50	45	90%	48	96%
빨리	50	45	90%	48	96%
좌회전	50	44	88%	46	92%
우회전	50	44	88%	47	94%
정지	50	47	94%	49	98%

위의 결과로 DTW 단독으로 사용할 때보다 Hybrid 형태로 사용 시 전체 평균의 약 5% 정도 상승된 것을 확인할 수 있다. 위의 실험들로 볼 때 후처리 인식 알고리즘을 사용하게 됨으로써, 인식률 향상을 가져온다는 것을 확인할 수 있었다.

하였다.

또한 본 논문에서는 실 시간이라는 개념을 도입하여 시간에 따라 연속적으로 입력되는 음성을 유성음만을 감지해서 이를 빠르게 처리하고, 인식시켜서 전동 휠체어 동작에 적용시키는 것이 목적이다. 그래서 벡터 양자화 개념을 도입하여 음성 벡터들을 1/12로 압축하여 인식 처리에 걸리는 시간을 현저히 줄였다.

하지만 벡터 양자화 사용 시 인식률이 다소 저하되는 것이 확인되어 이를 보상하기 위해 본 논문에서는 DTW + ART2의 Hybrid 형태를 적용하여 인식률을 향상시켰다. 그 결과 약 5%라는 인식 향상을 가져올 수 있었다.

### 참 고 문 헌

- [1] Lawrence Rabiner and Biing Hwang Juang, "Fundamentals of Speech Recognition", Prentice Hall, 1993
- [2] Steven L. Gay, Jacob Benesty, "Acoustic Signal Processing for Telecommunication", Kluwer Academic Publishers, 2000.
- [3] Lawrence Rabiner, "On the Application of Vector Quantization and Hidden Markov Models to Speaker Independent Isolated Word Recognition", Bell System Technical Journal, Vol. 62, No.4, April 1983.
- [4] 이지홍, 서일 DSP기술연구소공저, "DSP chip의 활용", 서일 DSP 주식회사
- [5] 유강주, "DHMM을 이용한 숫자음 인식의 Data Fusion에 관한 연구", 한국해양대학교 공학석사학위 논문, 1998.
- [6] 김정훈, 류홍석, 강성인, "음성인식을 적용한 다기능 휠체어 시스템의 설계 및 구현", 한국퍼지 및 지능시스템학회집 제12권 제1호, pp1-5, 2002
- [7] Chin-Der Wann, Stelios C. A. Thomopoulos, "A comparative study of self-organizing clustering algorithms dignet and ART2", Neural Networks, vol.10. No. 4, pp. 737 ~ 753, 1997
- [8] 김태영, "signal Processing using Fuzzy Logic and Neural Network for Welding Gap Detection", 한국해양대학교 석사학위 논문, 1999.
- [9] Carpenter, G.A., Grossberg S., "The ART of adaptive pattern recognition by a self-organizing neural network", IEEE Vol. 21, No. 3, pp. 77-88, 1988