# 사상멤버쉽함수에 의한 화자적응 단어인식

# Speaker-adaptive Word Recognition Using Mapped Membership Function

이 기 영*, 최 갑 석**

(Ki Young Lee*, Kap Seok Choi**)

## 요 약

본논문에서는 불특정화자 음성인식의 문제점이 되는 개인차에 의한 변동을 흡수하기 위하여 사상멤버쉽함수에 의한 화자적응 단어인식 방법을 제안하였다. 이방법의 학습과정에서는 미지화자의 표준화자의 스펙트럼패턴 사이에서 작성된 사상코드북에 퍼지이론을 도입하여 사상멤버쉽함수를 작성하였으며, 인식과정에서는 미지화자의 음성패턴을 사상멤버쉽함수에 의해 표준화자의 음성패턴에 적응된 패턴으로 재구성하고 뉴럴-퍼지패턴매칭에 의해 단어를 인식하였다.

본방법의 타당성을 평가하기 위하여, 28개의 DDD 지역명을 대상으로 실험한 결과, 종래의 사상코드북에 의한 벡터양자화 화자적응방법에서는 64.9[%], 퍼지벡터양자화 화자적응방법에서는 76.1[%]의 인식율을 얻었으나, 사상멤버쉽함수에 의한 화자적응방법에서는 95.4[%]의 향상된 인식율을 얻으므로써 인식성능의 우수함을 확인하였다. 또한, 사상멤버쉽함수의 작성과정에서는 반복된 학습과정이 불필요하며, 기억용량과 계산량도 사상코드북에 의한 화자적응방법보다 각각 1/30, 1/500배 정도였다.

## ABSTRACT

In this paper, we propose the speaker-adaptive word recognition method using a mapped membership function, in order to absorb a fluctuation owing to personal difference which is a problem of speaker independent speech recognition. In the training procedure of this method, the mapped membership function is made with the fuzzy theory introducded into a mapped codebook, between an unknown speaker's spectrum pattern and a standard speaker's one. In the recognition procedure, an input pattern of an unknown speaker is reconstructed to the pattern which is adapted to that of a standard speaker by the mapped membership function.

To show the validity of this method, word recognition experiments are carried out using 28 DDD area names. The recognition rate of the conventional speaker-adaptive method using a mapped codebook by VQ is 64.9[%], and that made by a fuzzy VQ is 76.2[%]. Throughout the experiment using a mapped membership function, we can achieve 95.4[%] recognition rate. This shows that our proposed method is more excellent in recognition performance. Moreover, this method doesn't need an iterative training procedure to make the mapped membership function, and memory capacity and computation requirements for this method are reduced to 1/30 and 1/500 time of those for the conventional method using a mapped codebook, respectively.

* 명지대학교 전자공학과
** 명지대학교 정보통신공학과
접수일자 : 1992. 4. 16.

## 1. 서 론

인간과 인간 사이의 통신수단인 음성을 인간과 기계 사이의 통신수단으로 사용하고자 하는 음성인식에 관한 연구는 오래 전부터 음성연구자의 흥미를 끌어왔다. 음성을 사용하여 기계에 직접 정보전달을 하기 위해서는 기계에 직접 음성을 입력할 수 있는 음성인식장치가 요구된다. 1950년 벨연구소 개발 음성인식장치가 개발되면서 이분야의 연구가 활발히 진행되어 왔으며, 1980년 이후에는 성형예측분석[2]에 의한 음성분석법의 발전과 디지탈 신호처리기술의 진보, 및 음성인식기술의 급속한 발전에 힘입어 대량의 연산을 필요로 하는 음성인식기술이 실용적인 음성인식장치로 실현되었다.

음성인식은 특정화자용과 불특정화자용으로 나뉘어지며, 전자는 특정인의 음성을 대상으로 하는 데에 반하여 후자는 미지의 여러 화자의 음성을 대상으로 인식을 수행하기 때문에 화자의 개인차에 의한 변동을 흡수할 수 있는 방법[3]이어야 하므로 특정화자 음성인식에 비해 그 방법이 복잡하고 그 성능도 낮은 실정이다.

불특정화자를 대상으로 한 음성인식에 있어서 개인차에 의한 변동에 대처하는 가장기본적인 것은 화자에 따른 변동이 비교적 적은 음운성을 강조한 음향적특성을 추출하여 스펙트럼의 동적인 특성[4]을 이용하는 것이 유효한 것으로 확인되고 있다. 이러한 구체적인 방법으로 개인차의 변동에 의한 표준패턴의 변동의 폭을 감소시키거나, 입력음성의 인식을 수행하기 전에 소수의 단어 혹은 짧은 문장을 발성하여 그 사람의 음성에 자동적으로 적응할 수 있는 화자적응학습기능[5]을 인식시스템으로 하여금 갖게 할 필요가 있다.

음성인식 시스템에 화자적응기능을 갖게 하는 방법에 대한 연구는 성대파 스펙트럼과 성도의 길이를 정규화하는 방법[6], 일부의 음소로 부터 개인차에 적응하는 전음소의 스펙트럼을 추정하는 방법[7], 화자에 적응하는 표준패턴의 집합의 선택하는 방법[8], 벡터양자화에 의한 코드북의 스펙트럼 사상방법[5,9,10], 스펙트럼 사상을 확률모델이나 신경회로망에 도입하는 방법[11,12]등이 있다. 여기서 최근 활발히 연구되는 방법은 스펙트럼 사상에 의해 입력음성을 표준패턴의 음성 스펙트럼특성으로 적응시켜 화자적응 음성인식을 수행하는 방법이며 이들 방법에서 비교적 좋은 인식율을 얻고 있다. 국내에서는 한국어숫자음에 대해 스펙트럼 사상에 의해 작성된 사상코드북을 이용한 화자적응기법[9]이 시도되었으며, 화자적응을 위한 정규오차를 감소시키기위해 부분공간법을 퍼지벡터양자화에 적용한 화자적응기법[10]이 제시되었다.

그러나, 상기의 벡터양자화 화자적응기법에서의 사상코드북은 학습과정에서 미지화자와 표준화자의 각 코드북에 의해 벡터양자화된 동일단어에 대해 DTW를 수행하여 최적경로를 얻은 후, 최적경로에 따라 미지화자의 스펙트럼에 대응하는 표준화자의 스펙트럼들을 선형결합하여 생성하므로 이상적으로는 사상코드북에 의해 구성되는 미지화자의 패턴은 표준화자의 패턴과 동일해야 하지만 사상코드북의 스펙트럼은 미지화자의 스펙트럼에 대응하는 표준화자의 스펙트럼들을 선형결합하여 작성되기 때문에 학습회수가 적으면 표준화자의 스펙트럼과 동일하지 않을 뿐만 아니라, 음성신호의 주요스펙트럼 정보인 피크주파수 성분의 에너지가 작아지거나, 선형결합에 의해 피크주파수가 변위하거나 삭제되어 버리는 문제가 생긴다.

본연구에서는 이러한 문제를 보완하기 위하여 개인차 때문에 애매하게 대응하는 다수의 스펙트럼특성들을 유효하게 표현하고 그 특성들을 흡수할 수 있는 방법으로 사상 멤버쉽함수에 의한 화자적응방법을 제안하였으며, 인식방법으로 신경회로망에 의한 집단화로 주파수변동의 폭을 감소시키는 뉴럴-퍼지패턴매칭을 제시하였다. 이 화자적응방법에서 벡터양자화와 DTW를 이용한 사상코드북을 작성하는 대신에 미지화자의 스펙트럼에 대응하는 표준화자의 다수스펙트럼들을 각각 2진화 스펙트럼으로 변환하고 이들을 중첩하여 대응되는 스펙트럼의 피크주파수들을 모두 귀속등급으로 나타낸 사상멤버쉽함수를 작성하므로써 미지화자와 표준화자의 스펙트럼대응관계에서 개인차에 의한 애매성을 표현하였으며, 뉴럴-퍼지패턴매칭에 의한 인식 방법에서는 사상멤버쉽함수로 재구성된 미지화자의 시험패턴과 멤버쉽함수로 구성된 표준패턴을 퍼지이론에 의해 매칭시키므로써 종래의 사상코드북에 의한 화자적응기법보다 학습회수가 적고 인식성능이 우수한 화자적응 음성인식을 수행하고자 한다.

## II. 뉴럴-퍼지패턴매칭 단어인식

### 2-1. 특징파라미터 추출

음성이 입력되면 1.0[kHz] 차단주파수의 저역통과필터를 통해 10[kHz]의 샘플링 주파수로 A/D 변환

하며, 특징파라미터로 구성된 패턴을 작성하기 위해 선형예측분석에 의해 스펙트럼을 추정한다. 스펙트럼 상에서 구해지는 음성의 피크주파수는 음원특성과 성도의 전달특성을 나타내기 때문에 음성인식에서 중요한 특징파라미터의 역할을 한다. 따라서 본 연구의 음성패턴을 구성하는 특징파라미터는 스펙트럼상의 각 피크주파수를 청각특성[13]에 근사한 1/3 옥타브 대역으로 변환하고 2진화 처리한 2진화스펙트럼으로 한다.

표 1에 1/3 옥타브 대역의 채널별 중심주파수를 보이고 있으며, 그림 1에 선형예측분석에 의하여 특징파라미터를 추출하는 과정을 나타내었다. 그림 1에서 음성신호가 입력되면 선형예측분석하여 스펙트럼을 추정하고 이 스펙트럼의 피크주파수를 피키피이킹 알고리즘[14]에 의하여 검출한다. 검출된 피크주파수들을 청각특성에 근사시키기 위해 다음 식에 의해 피크주파수를 1/3 옥타브 대역에 해당하는 채널로 변환한다.

$$k=\begin{cases} 1, & k=k_j \\ 0, & k\neq k_j \end{cases} \quad (k=1,\cdots,18) \tag{1}$$

$$k_j=\mathrm{INT}\left[3\times\log_2\frac{f_{pj}}{f_0}\right]\ (f_0=100[\mathrm{Hz}])$$

여기서 $k_j$는 1/3 옥타브 대역의 18개 대역 중에서 j번째 피크주파수 $f_{pj}$에 해당하는 대역의 채널을 나타내며 $f_0$는 1채널의 중심주파수 100[Hz]이고, INT[ · ]는 정수화함수이다.

식(1)은 선형예측분석에 의한 스펙트럼의 주파수축을 1/3옥타브 대역으로 나타내며 피크주파수에 해당하는 대역을 1로, 그 외의 대역을 0으로 표시하여 각 스펙트럼을 2진화 스펙트럼으로 변환한다.

### 2-2. 신경회로망에 의한 패턴작성

음성의 동일한 음운은 이상적으로 성도의 전달특성과 음원특성이 동일해야 하지만 동일한 화자일지라도 화자의 상태에 따른 성도의 순간적인 위치 변화와 유연성 때문에 동일한 음운의 스펙트럼이 변화하여 주파수변동을 일으킨다. 따라서, 음성으로 부터 추출된 1/3 옥타브 2진화스펙트럼은 주파수변동의 영향을 받으며, 인식성능을 저하시키는 요인이 된다.

본연구에서는 주파수 변동에 따른 인식성능의 저하를 방지하기 위해 신경회로망 특징 중의 하나인 연상기억능력을 이용하기로 한다. 즉, 동일한 음운의 유사한 2진화스펙트럼들을 입력으로 하고 그들의 대표 2진화스펙트럼을 교사응답으로 하여 학습시키므로써 신경회로망은 입력음성의 유사한 2진화스펙트럼들을 각각의 대표 2진화스펙트럼으로 변화시켜 동일한 음운에서 발생하는 주파수변동의 폭을 감소시킨다.

패턴작성을 위한 신경회로망은 빠른 학습능력을 갖는 역전파학습알고리즘[15]을 적용하였으며 이를 학습시키기 위해 교사응답으로는 종래의 벡터양자화에서 사용하는 집단화 알고리즘을 사용하여 얻은 대표

표 1. 1/3 옥타브대역의 채널별중심주파수

Table 1. Center frequency of a channel of 1/3 octave bands.

| CHANNEL | CENTER FREQUENCY [Hz] |
|---|---|
| 1 | 100 |
| 2 | 125 |
| 3 | 160 |
| 4 | 200 |
| 5 | 250 |
| 6 | 317 |
| 7 | 400 |
| 8 | 504 |
| 9 | 635 |
| 10 | 800 |
| 11 | 1,008 |
| 12 | 1,270 |
| 13 | 1,600 |
| 14 | 2,016 |
| 15 | 2,540 |
| 16 | 3,200 |
| 17 | 4,032 |
| 18 | 5,000 |

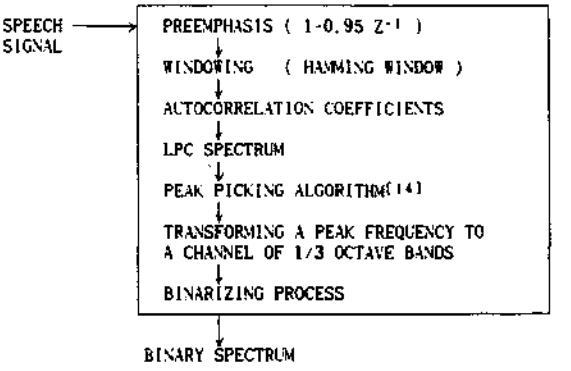


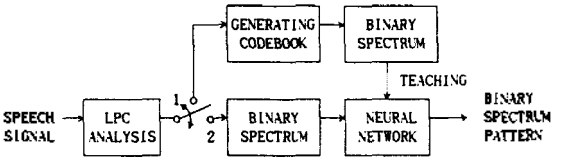

그림 1. 선형예측분석에 의한 1/3 옥타브 2진화스펙트럼의 추출과정

Fig. 1. Procedure for extracting 1/3 octave binary spectrum using linear prediction analysis.

스펙트럼을 2진화스펙트럼으로 변환하여 적용하였다. 그림 2는 신경회로망에 의해 2진화 스펙트럼패턴을 작성하는 전처리 과정이며, 그 학습과정은 다음과 같다.

(i) 음성을 선형예측분석한 후, 집단화 알고리즘에 의해 음성의 코드북을 생성하며 코드북의 각 집단의 대표 스펙트럼을 2진화스펙트럼으로 변환한다.

(ii) 각 집단의 대표 스펙트럼의 2진화스펙트럼을 교사응답으로 하고 그 스펙트럼 집단의 2진화 스펙트럼들을 학습데이타로 하여 신경회로망을 훈련시킨다.




1 : TRAINING 2 : EXECUTING

**그림 2.** 신경회로망에 의한 패턴작성의 전처리
**Fig. 2.** Preprocessiong for making pattern using a neural network.

이상의 학습과정에 의해 신경회로망이 훈련되면 다음과 같은 과정에 의해 패턴을 작성한다.

(i) 입력된 음성의 각 프레임으로 부터 선형예측분석에 의해 2진화스펙트럼을 작성하여 신경회로망에 입력하면 해당하는 집단의 대표 2진화 스펙트럼을 출력한다.

(ii) 위와 같은 과정을 음성의 전 프레임의 2진화스펙트럼에 적용하여 각 집단의 대표 2진화스펙트럼으로 구성된 패턴을 작성한다.

그림 3은 신경회로망에 의해 패턴을 작성하는 예를 보이고 있다. 여기서 (a)는 음성신호, (b)는 음성신호의 스펙트럼 시계열, (c)는 식(1)에 의한 2진화스펙트럼 패턴, (d)는 신경회로망에 의해 작성된 2진화스펙트럼 패턴이다.

### 2-3. 뉴럴-퍼지패턴매칭[16, 17]에 의한 단어인식

뉴럴-퍼지패턴매칭에 의한 단어인식은 신경회로망의 연상기억능력에 의해 패턴을 작성하여 주파수변동을 흡수하며, 선형신축방법에 의한 유사도측정을 이용하여 시간변동의 문제를 보완한다.

뉴럴-퍼지패턴매칭에 의한 음성인식에서 표준패턴으로 등록된 단어들을 $I=\{i_1, i_2, \cdots, i_n\}$, 단어명 $i_n$를 발성했을 때 1/3 옥타브 2진화 스펙트럼로 구성된 패턴을 $X=\{x_1, x_2, \cdots, x_n\}$, 각 단어패턴의 귀속성을 나타내는 멤버쉽함수를 $M=\{m_1, m_2, \cdots, m_n\}$이라 하자. 여기서 멤버쉽함수 $m_i$는 단어패턴 $x_i$의 귀속도를 나타내야 하므로 여러번 발성한 각 단어의 패턴들을 중첩하여 작성한다. 그림 4는 신경회로망에 의해 작성되는 각 단어의 패턴을 선형신축방법[16]으로 중첩한 표준패턴을 보인다.

그림 5는 신경회로망을 전처리과정에서 사용하여 주파수변동을 흡수하며 표준패턴을 멤버쉽함수로 작성하여 시간변동 문제를 보완하는 뉴럴-퍼지패턴매칭과 최대유사도결정규칙에 의해 단어를 인식하는 구성도이다. 그림 5에서 신경회로망은 2-2절의 학습과정에 의해 훈련된 것이며, 음성이 입력되면 신경회로망을 거쳐 2진화스펙트럼으로 구성된 음성의 패턴을 작성한다. 또한, 같은 음성의 패턴들을 중첩한 멤버쉽함수로 표준패턴을 작성하며 시험패턴과 표준패턴 사이의 유사성을 측정하여 최대유사도결정규칙에 의해 음성을 인식한다. 또한, 같은 음성의 패턴들을 중첩한 멤버쉽함수로 표준패턴을 작성하며 시험패턴과 표준패턴 사이의 유사성을 측정하여 최대유사도결정규칙에 의한 음성을 인식한다. 시험패턴을 $y(y \in X)$, j번째 등록된 단어의 표준패턴인 멤버쉽함수를 $m_j$라 할때 이들 사이의 유사도를 측정하는 식은 다음과 같다.

$$S_{jy} = \frac{P_{jy}}{P_{jd} - P_{jy}} \cdot \frac{P'_{jy}}{P_y - P'_{jy}} \tag{2}$$

여기서, $P_{jd} = \sum_f \sum_t m_j(f,t)$

$$P_y = \sum_f \sum_t y(f,t)$$

$$P_{jy} = \sum_f \sum_t m_j(f,t) \cdot y(f,t)$$

$$P'_{jy} = \sum_f \sum_t m_j(f,t) * y(f,t) \tag{3}$$

단, 여기서 $y$와 멤버쉽함수 $m_j$는 주파수 $f$와 시간 $t$의 2차원 패턴이며 $\cdot$는 두 패턴 사이의 논리곱이고 $*$는 $y$와 $m_j$의 $\alpha$-cut사이의 논리곱이다. 즉,

$$m_j(f,t) \cdot y(f,t) = \begin{cases} 1, & m_j(f,t) \geq 1 \text{ and } y(f,t) = 1 \\ 0, & \text{otherwise} \end{cases} \tag{4}$$

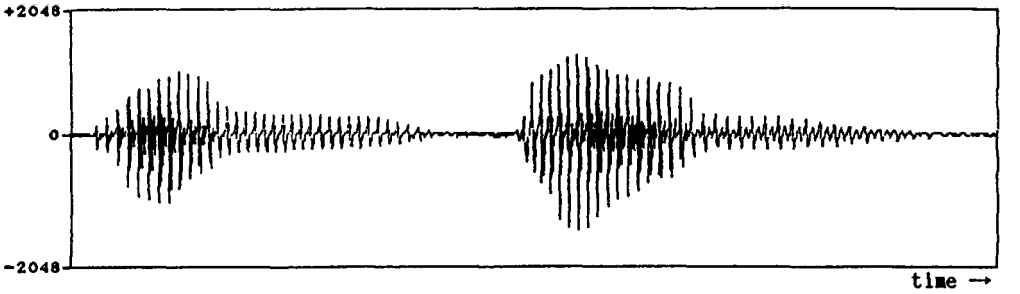
$$m_i(f,t) * y(f,t) = \begin{cases} 1, & m_i(f,t) \geq \alpha \text{ and } y(f,t) = 1 \\ 0, & \text{otherwise} \end{cases} \quad (5)$$

여기서 $\alpha$는 1부터 맴버쉽함수의 최대치 사이에서 선택한다.

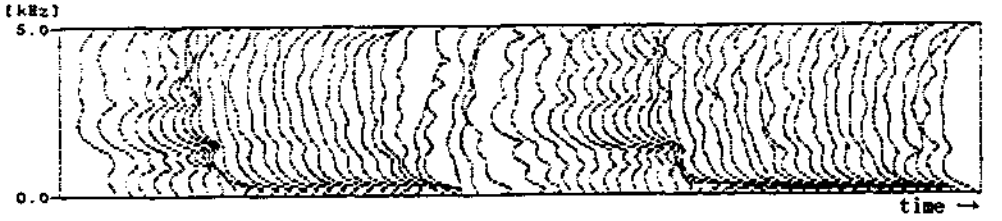
다음 식은 결정규칙에서 최대유사도를 갖는 표준패턴의 단어명을 선택하여 미지단어를 인식하는 식이다.

$$j^* = \arg\max_{1 \leq j \leq n} [S_{ij}] \quad (6)$$

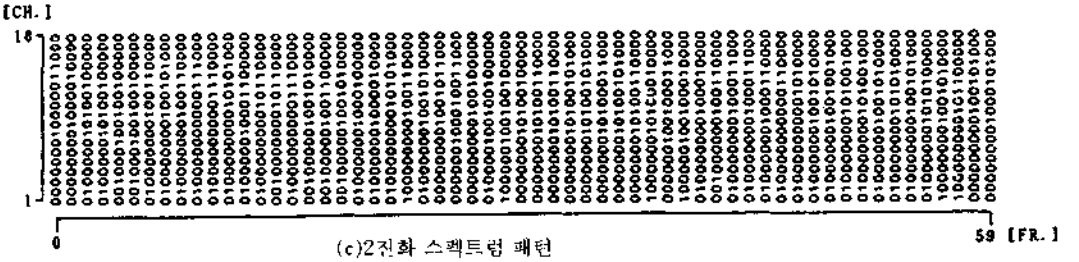
여기서, n은 등록 단어의 수이며, $S_{ij}$는 유사도 측정식이다.



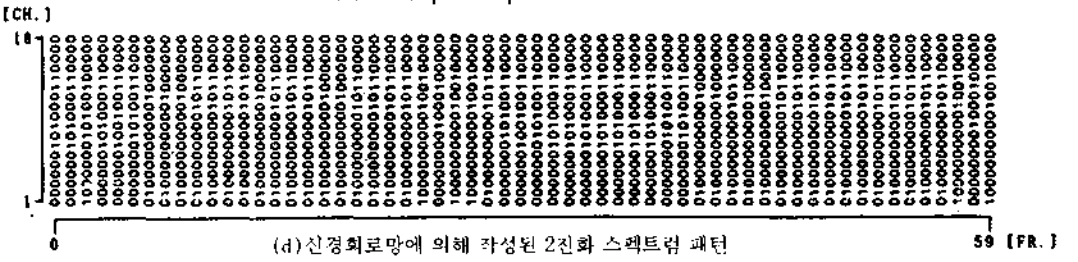

(a)음성신호 "안산"
(a)Speech signal "ansan"




(b)스펙트럼 시계열
(b)Spectrum sequence



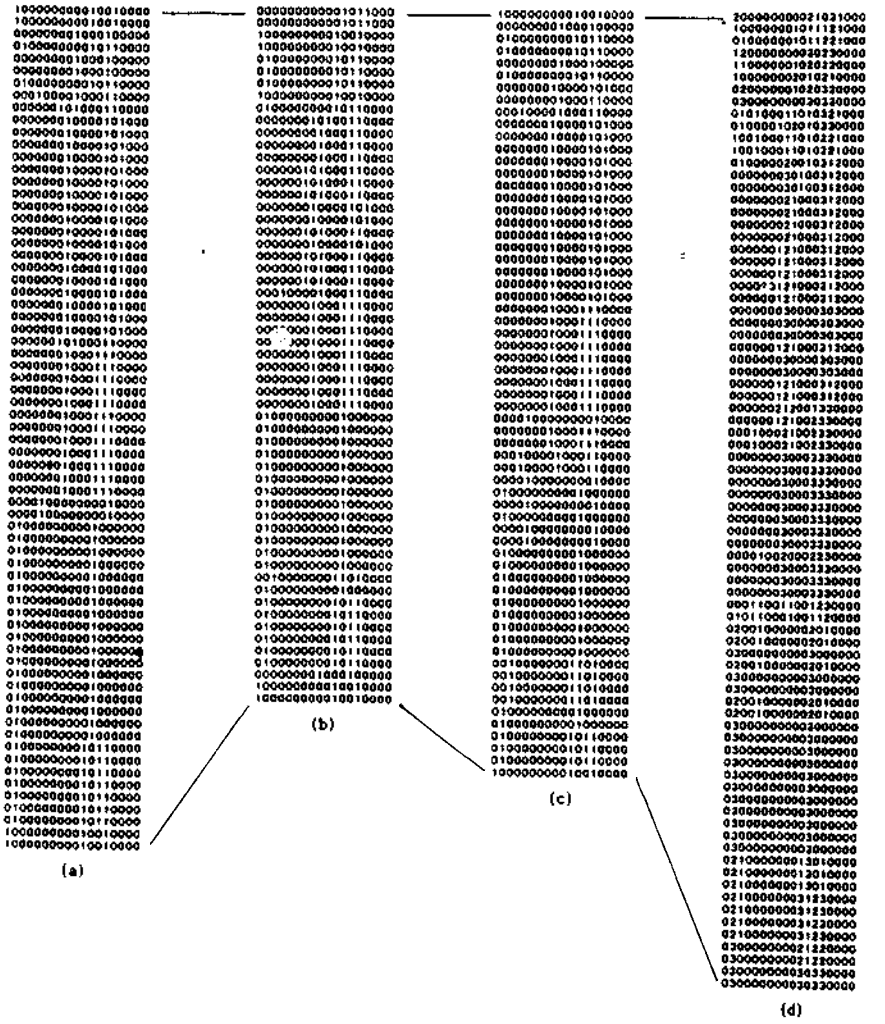

(c)2진화 스펙트럼 패턴
(c)Binary spectrum pattern




(d)신경회로망에 의해 작성된 2진화 스펙트럼 패턴
(d)Binary spectrum pattern made by a neural network

[CH.] : 1/3 octave channel, [FR.] : Frame number

그림 3. 신경회로망에 의한 2진화스펙트럼 패턴
Fig. 3. Binary spectrum pattern made by a neural network.




(a)(b)(c) : "서울"의 2진화스펙트럼패턴
(d) : 멤버쉽함수(표준패턴)
(a)(b)(c) : Binary spectrum pattern of "Seoul".
(d) : Membership function-reference pattern

**그림 4.** 멤버쉽함수(표준패턴)
**Fig. 4.** Membership function (Reference pattern).

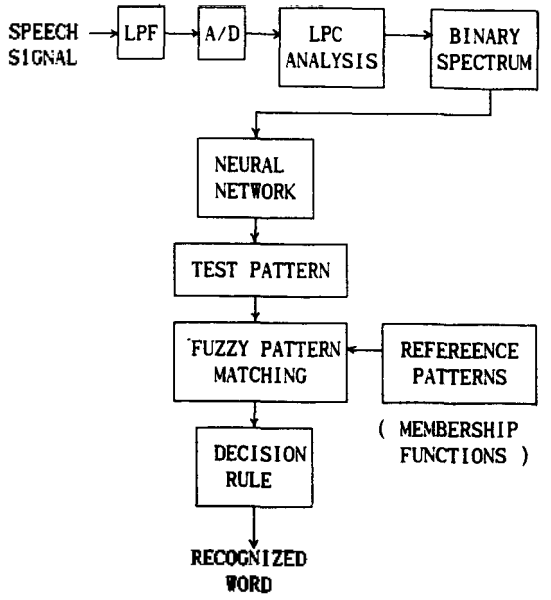



그림 5. 뉴럴-퍼지패턴매칭에 의한 단어인식 구성도
Fig. 5. Block diagram of word recognition using neural-fuzzy pattern matching.

## Ⅲ. 사상멤버쉽함수에 의한 화자적응 단어인식

미지화자 음성과 표준화자 음성사이의 개인차에 의한 변동을 흡수하기 위하여 본연구에서는 사상멤버쉽함수에 의해 화자적응적 패턴을 작성하고 뉴럴 퍼지패턴의 표준패턴을 이용하여 단어인식을 수행한다.

### 3-1. 학습과정

종래의 벡터양자화 화자적응기법에서는 화자적응을 위해 미지화자 음성의 스펙트럼공간을 표준화자 음성의 스펙트럼 공간으로 정규화하는 방법으로 두 화자의 스펙트럼사이의 대응관계에 따라 사상코드북을 작성하여 스펙트럼 정규화와 화자적응 단어인식에 사용하였다. 사상코드북은 학습과정에서 미지화자와 표준화자의 동일단어에 대한 DTW를 수행하여 최적경로를 얻은 후, 최적경로에 따라 미지화자의 스펙트럼에 대응하는 표준화자의 스펙트럼들을 선형결합하여 생성하게 된다. 이상적으로는 사상코드북의 각 스펙트럼은 표준화자의 스펙트럼과 동일해야 하지만 두 화자의 두 스펙트럼 공간은 개인차에 의해 서로 다르기 때문에 미지화자의 스펙트럼에 대응하는 표준화자의 스펙트럼이 복수로 존재하여 선형결합에 의해 작성된 스펙트럼은 표준화자의 스펙트럼과 일치하지 않을 뿐만 아니라, 음성신호의 주요 스펙트럼 특성인 피크주파수 성분의 에너지가 작거나 대응 횟수가 적으면 선형결합에 의해 피크주파수가 변위하거나 없어져 버리는 문제가 생긴다.

따라서 본연구에서는 대응스펙트럼들의 피크주파수 정보를 손실하지 않기 위하여 스펙트럼 대신에 피크주파수를 1로 하고 그 외의 성분을 0으로 하는 2진화스펙트럼을 사용하였으며, DTW의 최적경로에 따라 대응하는 스펙트럼의 2진화스펙트럼들을 가중하여 사상멤버쉽함수를 작성하므로써 미지화자의 스펙트럼에 대응하는 표준화자의 스펙트럼 특성을 지닌 멤버쉽함수를 생성하였다. 즉, DTW의 최적경로에 따라 대응하는 표준화자 스펙트럼의 2진화 스펙트럼들을 중첩하여 미지화자의 스펙트럼에 대응하는 표준화자의 스펙트럼 특성을 포함하는 사상멤버쉽함수를 작성하였으며, 대응하는 스펙트럼들의 피크주파수 위치를 멤버쉽함수의 귀속등급으로 표현하므로써 개인성 변동에 의한 애매성을 흡수할 수 있도록 하였다. 그림 6은 사상멤버쉽함수의 기본개념도이다.

그림 6에서 사상멤버쉽함수는 미지화자 A와 표준화자 B사이의 스펙트럼 대응관계에서 미지화자 A의 스펙트럼에 1대 1의 대응하는 스펙트럼을 생성하는 것이 아니라, 미지화자 A의 스펙트럼에 대응하는 표준화자 B 스펙트럼들의 피크주파수 정보들이 멤버쉽함수로 작성되어 개인차에 의한 변동의 폭을 귀속등급으로 표현한다. 사상멤버쉽함수를 작성하기 위한 학습과정은 다음과 같으며, 그림 7은 사상멤버쉽함수에 의한 화자적응 방법에서 학습과정의 구성되이다.

(ⅰ)미지화자의 학습단어에 대한 코드북을 생성한다.

(ⅱ)미지화자 코드북과 표준화자 코드북사이의 거리 행렬을 구한다.

(ⅲ)표준패턴 미지패턴의 DTW를 행하여 동일 단어 사이의 최적경로를 구한다.

(ⅳ)최적경로에 따라 대응하는 코드벡터들의 히스토그램을 구한다.

(ⅴ)다음 식에 의해 사상멤버쉽함수를 구한다.

$$m_i(A \rightarrow B) = \frac{\sum_{j=1}^{l} h_{ij}\, x_j^{(B)}}{\sum_{j=1}^{l} h_{ij}} \qquad (7)$$

여기서, $x_i^{(A)}$는 미지화자 A의 i번째 코드벡터의 2진화 스펙트럼

$x_i^{(B)}$는 표준화자 B의 i번째 코드벡터의 2진화 스펙트럼

$m_i(A \rightarrow B)$는 i번째 사상멤버쉽함수

$h_{ij}$는 대응하는 $x_i^{(A)}$와 $x_j^{(B)}$의 히스토그램

L은 코드북사이즈 이다.

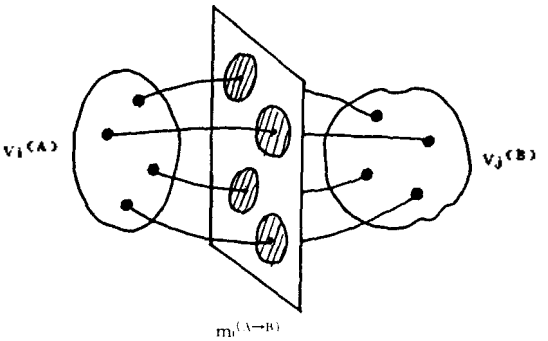



그림 6. 사상멤버쉽함수의 기본개념도

Fig. 6. Basic concept of mapped membership function.

### 3-2. 인식과정

3-1절과 같이 사상멤버쉽함수가 학습과정에 의해 생성되면 미지화자 음성의 코드북의 각 스펙트럼을 사상멤버쉽함수로 대치하여 멤버쉽함수로 구성된 패턴을 작성하므로써 화자정규화를 수행한다. 사상멤버쉽함수는 미지화자의 스펙트럼 패턴과 대응하는 스펙트럼 패턴의 구성요서가 되며, 미지화자의 스펙트럼에 대응하는 스펙트럼들의 피크주파수 정보를 귀속등급으로 표현하므로써 벡터양자화에 의한 화자적응 방법보다 효과적인 화자적응 단어인식을 수행할 수 있다.

사상멤버쉽함수에 의한 화자적응 단어인식은 미지화자의 학습단어가 입력되면 먼저 식(7)에서 설명한 바와 같이 DTW의 최적경로에 따라 히스토그램을 작성하고 사상멤버쉽함수를 생성하며 미지화자의 미지단어의 벡터들을 사상멤버쉽함수로 대치하여 화자정규화한 후, 뉴럴-퍼지패턴매칭 단어인식 방법에 의해 미지단어를 단어를 인식한다. 여기서, 사상멤버쉽함수에 의한 화자적응적패턴 $m_i^{(A \rightarrow B)}(f,t)$와 표준패턴 $m_i(f,t)$ 사이의 유사도측정식은 다음과 같다.

$$S_i = \frac{Q_{jd}}{Q_{jd} - Q_{jo}} \cdot \frac{\tilde{Q}_{jo}}{Q_{jd} - \tilde{Q}_{jo}} \qquad (8)$$

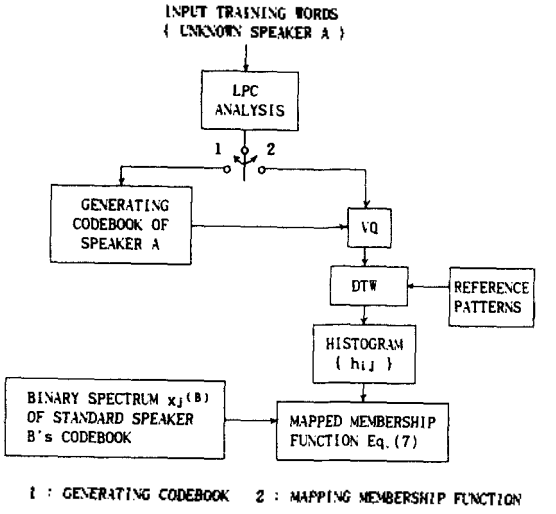



그림 7. 사상멤버쉽함수를 작성하는 학습과정

Fig. 7. Training Procedure for making a mapped membership function.

여기서,

$$Q_{jd} = \sum_i \sum_t m_i(f,t)$$
$$Q_{jo} = \sum_i \sum_t m_i^{(A \rightarrow B)}(f,t)$$
$$Q_{jo} = \sum_i \sum_t m_i(f,t) \cdot m_i^{(A \rightarrow B)}(f,t)$$
$$\tilde{Q}_{jo} = \sum_i \sum_t m_i(f,t) \ast m_i^{(A \rightarrow B)}(f,t) \qquad (9)$$

이며,

$$m_i(f,t) \cdot m_i^{(A \rightarrow B)}(f,t) = \begin{cases} 1, & m_i^{(A \rightarrow B)}(f,t) \geq 1 \text{ and } m_i(f,t) \geq 1 \\ 0, & \text{otherwise} \end{cases} \qquad (10)$$

$$m_i(f,t) \ast m_i^{(A \rightarrow B)}(f,t) = \begin{cases} 1, & m_i^{(A \rightarrow B)}(f,t) \geq \alpha_1 \text{ and } m_i(f,t) \geq \alpha_2 \\ 0, & \text{otherwise} \end{cases} \qquad (11)$$

이다. 여기서 $\alpha_1$는 1부터 사상멤버쉽함수 $m_i^{(A \rightarrow B)}(f,t)$의 최대치 사이에서 선택하며, $\alpha_2$는 1부터 멤버쉽함수 $m_i(f,t)$의 최대치 사이에서 선택한다. 또한, 결정규칙은 식(6)과 같다. 다음은 사상멤버쉽함수에 의해 화자정규화 및 미지화자의 미지단어를 화자적응적으로 인식하는 과정이다.

단계1) 미지화자의 미지단어가 입력되면 미리 작성한 코드북에 의해 벡터양자화한다.

단계2) 학습과정에서 작성한 사상멤버쉽함수를 벡터양자화된 미지단어의 각 벡터와 대치하여 화자적응화된 패턴을 작성한다.

단계3)화자적응화된 미지단어의 패턴을 퍼지패턴 매칭하여 최대유사도를 갖는 표준패턴을 선택하므로써 미지단어를 인식한다.

다음 그림8은 사상멤버쉽함수에 의한 화자적응 단어인식의 구성도이다.

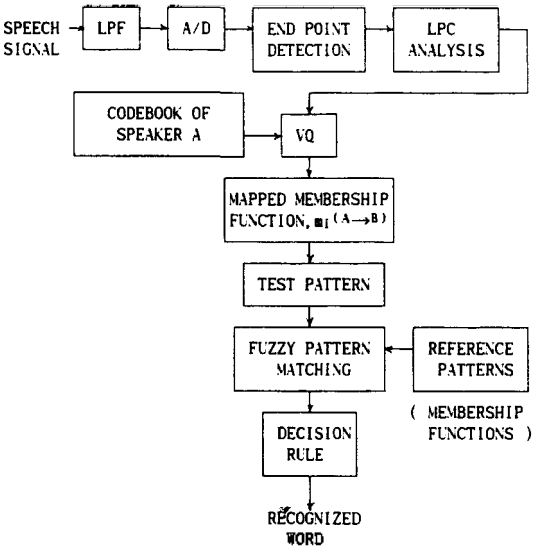



**그림 8.** 사상멤버쉽함수에 의한 화자적응 단어인식 구성도

**Fig. 8.** Block diagram of speaker adaptive word recognition using a mapped membership function.

## Ⅳ. 실험 및 결과 고찰

### 4-1. 데이타베이소의 구성

본 연구의 음성인식시스템에서 사용할 음성 데이타베이스는 음성을 인식하기 위한 학습에 사용될 학습데이타와 인식시 사용될 시험데이타로 구성된다. 데이타베이스는 방음처리가 되지 않은 실내에서 남성 2인과 여성 1인의 화자에 의해 DDD전화 지역명 28개(직할시 이상 6개와 경기도 지역명 22개)를 3인×28개×5회=420개의 음성데이타로 구성된다. 이중에서 각 화자의 코드북을 생성하기 위한 학습데이타는 각 화자가 3회씩 발성한 지역명 28개×3회=84개를 사용하였으며, 화자적응을 위한 학습단어는 각 화자가 1회씩 발성한 지역명 28개를 사용하였으며, 표준패턴을 작성하기 위한 음성데이타는 각 단어별로 3회씩 발성한 것을 이용하였으며 나머지 2회씩 발성한 것은 시험패턴으로 하였다.

이러한 음성들을 차단주파수가 4.0[kHz]인 저역통과필터에 통과시켜 고주파 성분을 제거하였으며, 샘플링 주파수가 10[kHz]인 12비트 분해도의 AD변환기로 샘플링하여 PC/AT의 하드 디스크에 저장하였다. 위의 과정을 거쳐 PC/AT에 저장된 데이타는 16비트의 정수로 변환되어 일시 저장되며 이 음성데이타로부터 자기상관계수와 14차 선형예측계수를 추출하여 음성데이타와 함께 영구 저장을 위하여 플로피 디스크에 저장한다. 여기서 음성신호의 프레임 구간은 20.0[ms](200샘플)로 하였으며, 이동구간은 10.0[ms](100샘플)로 하여 50[%]가 중첩되게 하였다. 데이타베이스의 생성을 위한 해석조건은 표 2에 나타내었다.

**표 2.** 분석조건

**Table 2.** Analysis conditions

| | |
|---|---|
| Sampling frequency | : 10.0 [kHz] |
| Cut-off frequency of LPF | : 4.0 [kHz] |
| A / D resolution | : 12 [bit] |
| Analysis window | : 20.0 [ms] |
| Overlapping interval | : 10.0 [ms] |
| Window function | : Hamming window |
| LPC order | : 14 |

### 4-2. 화자적응 인식결과 및 고찰

본절에서는 본연구에서 제안한 사상멤버쉽함수에 의한 화자적응 인식결과와 벡터양자화및 퍼지벡터양자화의 사상코드북에 의한 화자적응 인식결과를 비교 고찰한다.

#### 4-2-1. 사상코드북에 의한 인식결과

사상코드북에 의하여 미지화자의 음성을 화자적응적으로 인식하기 위해 벡터양자화와 DTW에 의한 스펙트럼사상[9][10]을 이용하였다. 즉, 사상코드북을 이용한 화자적응 음성인식의 첫 번째 방법은 문헌[9]에 기술된 벡터양자화에 의한 화자적응 음성인식으로 미지화자와 표준화자가 발성한 동일한 학습단어를 대상으로 학습과정을 거쳐 히스토그램에 의해 사상코드북을 생성하며 미지화자의 코드북을 사상코드북으로 대치해 줌으로써 스펙트럼을 정규화한 후, DTW와 KNN 결정규칙에 의해 단어인식을 수행한다. 여기서 사상코드북의 스펙트럼은 표준화자의 스펙트럼과 정규화 오차가 작아야 하기 때문에 DTW에 의한 거리가 수렴할 때 까지 학습과정을 반복해야 한

다.

두 번째 방법은 문헌[10]에 기술된 퍼지벡터양자화와 퍼지히스토그램을 이용하여 양자화 왜곡의 감소와 대응확률의 정밀성으로 화자적응을 보다 고정도화할 수 있는 사상코드북을 생성한다. 단어인식 방식은 첫 번째 방법과 같이 DTW와 KNN 결정규칙을 이용한다. 여기에서도 사상된 스펙트럼과 표준화자의 스펙트럼 사이의 정규화 오차를 줄이기 위하여 학습과정을 반복하여 사상코드북을 생성한다. 표 3에는 각 방법에 따른 인식율을 나타내었다.

**표 3.** 사상코드북에 의한 화자적응 인식율의 비교
**Table 3.** Comparison of speaker-adaptive recognition rates using a mapped codebook.

| METHODS | RECOGNITION RATES [%] | | |
|---|---|---|---|
| | MA→MB | FC→MB | AVERAGE |
| WA | 58.3 | 43.3 | 50.8 |
| VQ+H | 702. | 59.6 | 64.9 |
| FVQ+FH | 81.1 | 70.2 | 76.1 |
| SD | 97.8 | 95.4 | 96.6 |

여기서,

WA(without adaptation)은 화자적응 학습을 하지 않은 표준화자의 코드북으로 민지화자의 단어를 인식하는 방법,

VQ+H(histogram)은 히스토그램에 의한 학습과정으로 사상코드북을 작성하고 벡터양자화에 의해 화자적응 단어인식하는 방법,

FVQ+FH(fuzzy histogram)은 퍼지 히스토그램에 의한 학습과정으로 사상코드북을 작성하고 퍼지벡터양자화에 의해 화자적응 단어인식하는 방법,

SD(Speaker dependent)는 표준화자의 특정화자 단어인식 방법이다.

화자적응을 하지 않는 WA 방법에서는 남성간 58.9[%]의 인식율을 얻었으며, 남녀간에는 44.6[%]의 인식율을 얻었다. 여기서, 남녀간의 인식율이 남성간의 인식율보다 낮은 것은 화자정규화 하지 않은 경우 같은 음성이라도 남녀간의 피치주기나 공진주파수의 차이가 크기 때문에 생기는 결과로 생각된다.

종래의 VQ+H에 의한 화자적응 단어인식 방법에서는 남성간 71.4[%]의 인식율과 남녀간 60.7[%]의 인식율을 얻으므로써 WA 방법보다 평균 인식율이 14.1[%] 개선되었으며, 본 연구에서 제안한 FVQ+FH에 의한 방법에서는 남성간 80.4[%]의 인식율과 남녀간 69.6[%]의 인식율을 얻으므로써 WA 방법보다 평균 23.2[%], VQ+H에 의한 방법보다 8.9[%] 향상된 인식율을 얻었다. FVQ+FH에 의한 방법에서 VQ+H 방법보다 인식율이 높은 것은, 학습과정에서 퍼지 히스토그램에 의한 대응확률을 히스토그램에 의한 것보다 정확하게 구할 수 있으며, 인식과정의 DTW와 KNN 결정규칙에서는 평균 왜곡이 최소인 것을 선택하여 인식하기 때문에 VQ에 의해 양자화하여 인식하면 귀속등급에 의한 복호화에 의해 양자화 왜곡이 저감된 퍼지벡터양자화에 의해 복호화할 때보다 오인식이 많은 것으로 생각된다.

### 4-2-2. 사상멤버쉽함수에 의한 화자적응 인식결과

사상멤버쉽함수에 의한 화자적응 단어인식을 하기 위해 학습과정에서는 3-1절에 기술한 바와 같이 벡터양자화와 DTW 및 표준화자 코드북의 2진화스펙트럼을 이용하여 사상멤버쉽함수를 작성한다. 사상멤버쉽함수는 사상코드북의 스펙트럼과 달리 스펙트럼의 피크주파수 정보를 귀속등급으로 표현하기 때문에 정규화오차를 줄이기 위해 학습과정을 반복해 줄 필요가 없다. 이 방법의 인식방법은 뉴럴-퍼지패턴매칭과 같기때문에 코드워드의 수는 음운의 수와 밀접한 관계를 갖는다.

본실험에서는 코드워드의 수를 변수로 하여 화자적응 인식율을 비교하였으며 표준패턴은 표준화자가 뉴럴-퍼지패턴매칭에서 사용하였던 멤버쉽함수이고, 시험패턴도 사상 멤버쉽함수에 의한 멤버쉽함수이므로 각 코드워드의 수에 따라 시험패턴의 $\alpha$-cut 값 $\alpha_1$과 표준패턴의 $\alpha$-cut 값 $\alpha_2$를 변수로 하여 인식율을 비교하였다. 여기서 $\alpha_1$은 귀속등급이 제일 큰 것을 100으로 정규화하여 사용하였다. 표 4에는 코드워드의 수와 $\alpha_1$, $\alpha_2$를 변수로 한 화자적응 인식율을 나타내었다.

표 4(a)에서 코드워드의 수가 4인 경우 $\alpha_1=50$이고 $\alpha_2=2$일 때 남성간 87.9[%]의 인식율과 남녀간 80.7[%]의 인식율을 얻어 84.3[%]의 평균 인식율을 얻었으며 표 3의 WA 방법보다 평균 인식율이 32.5[%], VQ+H 방법보다 평균 인식율이 18.2[%], FVQ+FH 방법보다 9.3[%] 개선되었음을 알 수 있었다.

코드워드의 수가 8인 경우 $\alpha_1=10$이고 $\alpha_2=2$일 때 남성간 100.0[%]의 인식율과 남녀간 94.3[%]의 인식율을 얻어 97.2[%]의 평균인식율을 얻었으며 표 3

의 WA 방법보다 평균 인식율이 45.4[%], VQ+H 방법 보다 평균 인식율이 31.1[%], FVQ+FH 방법보다 22.2[%] 개선되었음을 알 수 있었다. 그런데 남녀간의 인식율은 $\alpha_1=50$이고 $\alpha_1=3$일 때, 95.0[%]로 가장 높게 나타남을 알 수 있었다.

**표 4.** 사상멤버쉽함수에 의한 화자적응 단어인식율
(MA : 남자 1, MB : 남자 2, FC : 여자 3)

**Table 4.** Speaker-adaptive word recognition rates using a mapped membership function.
(MA : Male 1, MB : Male 2, FC : Female 3)

(a)Codewords=4 [%]

| $\alpha_1$ / $\alpha_2$ | 10 | | | 50 | | | 100 | | |
|---|---|---|---|---|---|---|---|---|---|
| RELATION | 1 | 2 | 3 | 1 | 2 | 3 | 1 | 2 | 3 |
| MA→MB | 84.3 | 87.9 | 85.7 | 87.9 | 87.9 | 87.9 | 80.7 | 84.3 | 84.3 |
| FC→MB | 77.1 | 80.0 | 80.0 | 80.0 | 80.7 | 82.9 | 77.9 | 78.6 | 75.7 |
| AVERAGE | 80.7 | 83.9 | 82.9 | 83.9 | 84.3 | 85.4 | 79.3 | 81.5 | 80.0 |

(b)Codewords=8 [%]

| $\alpha_1$ / $\alpha_2$ | 10 | | | 50 | | | 100 | | |
|---|---|---|---|---|---|---|---|---|---|
| RELATION | 1 | 2 | 3 | 1 | 2 | 3 | 1 | 2 | 3 |
| MA→MB | 100.0 | 100.0 | 98.6 | 98.6 | 97.9 | 97.9 | 95.7 | 100.0 | 100.0 |
| FC→MB | 80.7 | 94.3 | 93.6 | 89.3 | 95.0 | 93.6 | 81.4 | 91.4 | 92.1 |
| AVERAGE | 90.4 | 97.2 | 96.1 | 93.9 | 96.5 | 95.8 | 88.6 | 95.7 | 96.1 |

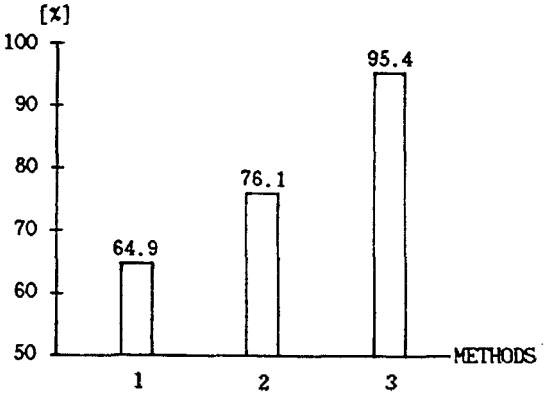
(c)Codewords=16 [%]

| $\alpha_1$ / $\alpha_2$ | 10 | | | 50 | | | 100 | | |
|---|---|---|---|---|---|---|---|---|---|
| RELATION | 1 | 2 | 3 | 1 | 2 | 3 | 1 | 2 | 3 |
| MA→MB | 75.0 | 82.9 | 85.0 | 77.9 | 82.9 | 85.7 | 74.3 | 80.7 | 83.6 |
| FC→MB | 58.6 | 72.1 | 78.6 | 62.9 | 73.6 | 76.4 | 57.9 | 67.1 | 71.4 |
| AVERAGE | 66.8 | 77.5 | 81.8 | 70.4 | 78.3 | 81.1 | 66.1 | 73.9 | 77.5 |

코드워드의 수가 16인 경우 $\alpha_1=50$이고 $\alpha_2=3$일 때 남성간 85.7[%]의 인식율과 남녀간 76.4[%]의 인식율을 얻어 81.1[%]의 평균 인식율을 얻었으며 표 3의 WA 방법보다 평균 인식율이 29.3[%], VQ+H 방법 보다 평균 인식율이 15.0[%], FVQ+FH 방법보다 6.1[%] 개선되었음을 알 수 있었다. 그림 10에는 사상코드북에 의한 화자적응 방법과 사상멤버쉽함수에 의한 화자적응방법을 비교하기 위하여 벡터양자화 화자적응 방법과 퍼지벡터양자화 화자적응 방법에 의한 인식율을 나타내었다.

사상멤버쉽함수에 의한 화자적응 단어인식 방법이 사상코드북에 의한 방법보다 인식율이 높은 것은 사상코드북이 미지화자의 스펙트럼에 대응하는 표준화자의 다수 스펙트럼들을 선형결합하여 하나의 스펙트럼을 작성하지만 사상멤버쉽함수에 의한 화자적응 방법에서는 미지화자의 스펙트럼에 대응하는 표준화자의 다수 스펙트럼에서 피크주파수 정보를 이용하여 화자사이의 개인차에 의한 주파수 변동의 폭을 귀속등급으로 표현해 주며, 이 귀속등급으로 구성된 사상멤버쉽함수에 의한 미지화자의 음성을 패턴으로 작성하여 인식하기 때문에 사상멤버쉽함수에 의한 화자적응 방법이 보다 더 효과적인 방법이라고 생각

된다.




1 : 벡터양자화 화자적응
(Vector quantization speaker adaptation)
2 : 퍼지벡터양자화 화자적응
(Fuzzy vector quantization speaker adaptation)
3 : 사상멤버쉽에 의한 화자적응
(Speaker adaptation using a mapped membership function)

**그림 9.** 화자적응 인식율을 비교
**Fig. 9.** Comparison of speaker-adaptive recognition rates.

## V. 결 론

본연구에서는 불특정화자 음성인식의 문제점으로 대두되고 있는 개인차에 의한 변동을 흡수하기 위하여 사상멤버쉽함수에 의한 화자적응 단어인식 방법을 제안하였다. 이 방법에서 사상코드북에 퍼지 이론을 도입하여 사상멤버쉽함수를 작성하였으며, 이를 이용하여 화자적응적 패턴을 작성하였다. 또한, 주파수변동 문제를 개선하고 인식율을 향상시키기 위하여 신경회로망과 퍼지 이론을 결합한 뉴럴-퍼지패턴매칭에 의해 화자적응적 패턴을 인식하였다. 이 방법의 타당성을 평가하기 위하여 남자 2인과 여성 1인이 발성한 28개의 DDD 전화 지역명을 대상으로 실험한 결과, 다음과 같은 결론을 얻을 수 있었다.

1. 사상멤버쉽함수의 사상과정에서는 조화의 학습과정을 요구하기 때문에 반복된 학습과정을 필요로 하는 사상코드북보다 학습시간이 짧다.
2. 사상멤버쉽함수에 의한 화자적응 단어인식과정에서 요구되는 기억용량과 계산량은 DTW에 의한 단어인식방법에 비해 각각 1/30, 1/500 배 정도로 감소되었다.
3. 종래의 사상코드북에 의한 벡터양자화 화자적응 방법에서는 64.9[%], 퍼지벡터양자화 화자적응 방법에서는 76.1[%]의 인식율을 얻었으나, 사상멤버쉽함수에 의한 화자적응 단어인식방법에서는 95.4[%]의 인식율을 얻었으므로 인식율이 향상되었음을 확인하였다.

따라서, 뉴럴-퍼지패턴매칭 방법은 주파수변동 문제를 개선하여 인식율을 향상시킬수 있는 효과적인 특정화자 단어인식 방법이며, 사상멤버쉽함수에 의한 화자적응 단어 인식방법은 음성의 피크주파수를 멤버쉽함수로 표현하여 사상하므로써 종래의 사상코드북에 의한 화자적응 단어인식방법보다 더욱 효과적으로 개인차의 변동을 흡수할 수 있는 방법임을 확인할 수 있었다.

## 참 고 문 헌

1. W.A.Lea, Tends in speech recognition, Prentice-Hall, Inc. 1980.
2. L.R.Rabiner and R.W.Shafer, Digital processing of speech signal, Prentice Hall, 1978.
3. K.Shikano, et al., "Recent trends on speech recognition," The Journal of the Institute of Electronics, Information and Communication Eng., Vol.71, No.11, pp.1170, 1988.
4. S. Furui, "Speaker-independent isolated word recognition using dynamic features of speech spectrum," IEEE Trans., Acoust.Speech, Signal Processing, Vol.ASSP-34, No.1, pp. 52, 1986.
5. K.Shikano, K.F.Lee, R.Reddy, "Speaker adaptation through vector quantization," Proc. ICASSP 86, 49.5, 1986.
6. H.Matsumoto, H.Wakita, "Vowel normalization by frequency warped spectral Matching," Speech Comm., Vol.5, No.2, pp.239, 1986.
7. S. Furui, "A training procedure for isolated word recognition systems," IEEE Trans., Acoust. Speech, Signal Processing, Vol.ASSP-28, No.2, pp.128, 1980.
8. 溝口, 木下, 角所, "セット化音韻テンプレトに基づく不特定話者單語音聲認識システム," 新學論 J67-A, 6, Jun. 1984.
9. 최갑석, 이기영, "사상코드북을 이용한 화자적응 한국어 숫자음 인식에 관한 연구," 한국음향학회지 제9권 5호 1990.
10. 최갑석, "벡터 양자화 화자적응기법을 사용한 한국어 단어인식," 한국음향학회지 제10권 4호 1991.

11. S.Nakamura, K.Shikano, "VQ-based speaker adaptation applied to HMM phoneme recognition," The Journal of the Acoustic Society of Japan, Vol.45, No.12, pp.942, 1989.

12. H.Tseng, M.J.Sabin, E.A.Lee, "Fuzzy vector quantization applied to hidden Markov modeling," Proc. ICASSP 87, 15.5, 1987.

13. R.A.Cole, Perceptron and production of fluent speech, Lawrence Erlbaum Associates, Inc., 1980.

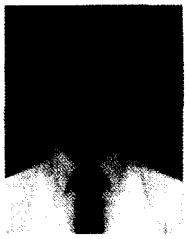
14. 崔元奎, 李義一, 秋月影雄, "韓國語音聲認識のための改良形ホルマントトラッキングと數子音聲認識への應用," 電學論 C, Vol. 109-C, No.10, 1988.

15. Yoh-Han Pao, Adaptive pattern recognition and neural networks, Addison-Wesley Pub. Com., Inc., 1989.

16. J.Fuzimoto, T.Nakatani, M.Yoneyama, "Speaker-independent word recognition using fuzzy pattern matching," Fuzzy Sets and Systems 32, North-Holland, pp.181, 1989.

17. 이기영, 김한재, 최갑석, "퍼지패턴매칭에 의한 음성인식에 관한 연구," 한국음향학회 학술발표회 논문집, Vol.10, No.1,(s), 1991.

▲이 기 영



1960년 5월 7일생.
1980년 3월~1984년 2월 : 명지대학교 전자공학과 졸업
1984년 3월~1984년 2월 : 명지대학교 대학원 전자공학과 졸업
1987년 9월~1992년 2월 : 명지대학교 대학원 전자공학과 공학박사

▲최 갑 석

1930년 9월 30일생.
1951년 9월~1955년 3월 : 서울대학교 공과대학 통신공학과 졸업
1975년 2월 25일 : 명지대학교 공학박사
1980년 8월~1981년 8월 : 일본 동경대학 공학부 전자공학과 연수
1972년 3월~현재 : 명지대학교 전자공학과 교수