음성 인식을 위한 최적 가중 켑스트랄 거리 측정 방법

김원구*, 윤대희**
* 군산대학교 전기공학과, ** 연세대학교 전자공학과

Optimally Weighted Cepstral Distance Measure for Speech Recognition

Weon-Goo Kim*, Dae-Hee Youn**
* Dept. of Electrical Engineering, Kunsan National University
** Dept. of Electronics Engineering, Yonsei University

Abstract

In this paper, a method for designing an optimal weight function for the weighted cepstral distance measure is proposed. A conventional weight function or cepstral lifter is obtained experimentally depending on the spectral components to be emphasized. The proposed method minimizes the error between word reference patterns and the training data.

To compare the proposed optimal weight function with conventional functions, speech recognition systems based on Dpynamic Time Warping and Hidden Markov Models were constructed to conduct speaker independent isolated word recognition experiments. Results show that the proposed method gives better performance than conventional weight functions.

Ⅰ. 서론

현재까지 음성 인식에서는 여러가지 거리 측정 방법들이 제안되어 사용되어 왔다[1-4]. 특히 켑스트럼 계수를 사용하는 거리 측정 방법은 두 패턴의 대수 스펙트럼의 차이를 구하는 방법으로서, 계산이 간단하고 인식 성능도 우수하여 많이 사용되어 오고 있다.

또한 최근에는 음성 인식에 중요한 스펙트럼 성분들(포만트 등)은 강조하고 불필요한 성분들(스펙트럼 기울기, 채널 특성 등)은 억압하는 가중 함수를 켑스트럼 계수에 사용하여 인식 성능을 향상시키는 연구가 많이 진행되었다[5-10]. 이러한 과정을 켑스트랄 리프터링(cepstral liftering)이라고 하며, 음성 및 화자 인식에서 좋은 성능을 나타내었다[5-12].

음성 인식에 사용되는 가중 켑스트랄 거리 측정 방법의 가중 함수는 음성 인식에 유용한 켑스트럼 계수에는 큰 가중을 두고 나쁜 영향을 주는 켑스트럼 계수에는 작은 가중을 주는 것을 목적으로 한다. 그러나 인식율과 가중 함수의 관계를 정량적으로 나타낼 수 없기 때문에, 음성 인식에 사용되어온 기존 가중 함수들은 스펙트럼의 포만트 주파수를 강조하고 불필요한 성분들을 제거하거나 켑스트럼 계수의 분산을 일치시키는 등의 간접적인 방법을 사용하여 음성 인식 시스템 성능을 개선하려고 노력하였다. 따라서 가중 함수에 포함된 변수들은 많은 실험을 통하여 경험적으로 얻어져야만 하는 문제점이 있었다.

본 논문에서는 가중 켑스트랄 거리 측정 방법을 사용하는 음성 인식 시스템의 성능을 향상시키기 위하여 가중 켑스트랄 거리 측정 방법의 최적 가중 함수 설계 방법을 제안하였다. 제안된 방법의 특징은 각 단어의 기준 패턴과 학습 데이타 간 거리를 최소화하도록 가중 함수 또는 리프터(lifter)의 형태를 결정하는 것이다. 거리 측정을 사용하는 음성 인식 시스템에서는 최종 인식의 결정이 기준 패턴과 입력 패턴과의 거리에 의하여 결정되기 때문에, 이러한 거리를 최소화하는 가중 함수를 사용하여 인식 시스템의 성능을 향상시킬 수 있었다. 이러한 알고리즘은 가중 함수를 사용하지 않는 켑스트랄 거리 측정 방법을 사용하는 음성 인식 시스템을 사용하여 최적 가중 함수가 결정될 때까지 반복적으로 구현되었다.

또한 본 논문에서는 Dynamic Time Warping(DTW)[12]와 Hidden Markov model(HMM)[13]를 사용한 두가지 종류의 음성 인식 시스템에 최적 가중 함수를 적용한 가중 켑스트랄 거리 측정 방법을 사용하여, 본 논문에서 제안된 방법의 성능을 평가하였다.

Ⅱ장에서는 가중 켑스트랄 거리 측정 방법의 최적 가중 함수 결정 방법을 제안하였고 Ⅲ장에서는 최적 가중 함수의 성능을 평가하기 위하여 실험 및 결과 고찰을 기술하였으며 Ⅳ장에서는 결론을 맺었다.

Ⅱ. 최적 가중 켑스트랄 거리 측정 방법

본 논문에서는 가중 켑스트랄 거리 측정 방법을 사용하는 음성 인식 시스템의 성능을 개선하기 위하여 최적 가중 함수 결정 알고리즘을 제안하였다. 제안된 방법은 각 단어의 학습 데이타와 기준 패턴간의 오차를 최소화하는 최적 가중 함수를 결정하는 것이다.

최적 가중 함수의 형태는 인식 대상과 인식 시스템의 형태에 따라서 달라진다. 본 장에서는 제안된 알고리즘을 기술하기 위하여 DTW를 이용한 인식 시스템을 대상으로 최적 가중 함수 설계 방법을 유도하였다.

우선, 가중 캡스트랄 거리 측정 방법을 다음과 같이 정의하였다.

$$d(a,b) = \sum_{k=1}^{P} w_k^2 (a_k - b_k)^2 = \sum_{k=1}^{P} w_k' (a_k - b_k)^2 \quad (1)$$

여기서 a 와 b 는 특징 벡터이고 $w = (w_1, w_2, \cdots, w_P)$는 가중 함수 또는 리프터이다. 또한 최적 가중 함수를 유도하기 위하여 $w' = (w_1', w_2', \cdots, w_P')$는 $w_k' = w_k^2$, $1 \le k \le P$로 정의하였다.

제안된 방법은 학습 데이타와 기준 패턴간 오차를 최소화하는 최적 가중 함수를 결정하는 것이다. 따라서 두 벡터간 자승 오차 벡터(squared error vector) $\varepsilon = (\varepsilon_1, \cdots, \varepsilon_P)$를 정의하면 (1)의 가중 캡스트랄 거리 측정 방법은 (2)와 같이 변형된다.

$$d(a,b) = \sum_{k=1}^{P} w_k' \varepsilon_k = \sum_{k=1}^{P} w_k' (a_k - b_k)^2 \quad (2)$$

즉, 두 벡터간의 가중 캡스트랄 거리는 오차 벡터 원소(element)의 합과 같다. 따라서 두 패턴간 평균 거리는 정합된 벡터들간에 발생하는 오차 벡터들의 평균으로 부터 구할 수 있다. 이러한 자승 평균 오차 벡터(mean squared error vector) $\bar{\varepsilon} = (\bar{\varepsilon}_1, \cdots, \bar{\varepsilon}_P)$는 각 단어마다 학습 데이타와 기준 패턴간 인식 실험을 수행하여 구한다. 이러한 자승 평균 오차 벡터는 기준 패턴과 학습 데이타간에 각 차수에서 발생하는 오차를 나타낸다. 따라서 최적 가중 함수는 이러한 자승 평균 오차 벡터의 크기에 역비례하게 결정된다.

2-1. DTW를 이용한 인식 시스템의 평균 거리와 오차 벡터

DTW를 이용한 단독음 인식 시스템은 V개 단어를 인식하기 위하여 각 단어마다 M개 학습 데이타를 집단화(clustering) 방법을 사용하여 만든 L개의 기준 패턴을 가지고 있다고 가정하였다. 이때 N개 특징 벡터로 구성된 입력 패턴을 $[a_1, \cdots, a_i, \cdots, a_N]$으로 표현하면 단어 v 의 기준 패턴과 입력 패턴간 최소 거리 d^v (L=1인 경우)는 와핑 경로를 사용하여 다음과 같이 표현할 수 있다.

$$d^v = \min_{1 \le l \le L} \left[\sum_{n=1}^{K_l^v} d(w(n)) \right]$$
$$= \min_{1 \le l \le L} \left[\sum_{n=1}^{K_l^v} d(a_{i(n)}, b_{j(n),l}^v) \right], \quad 1 \le v \le V \quad (3)$$

여기서 와핑 경로 $w(n) = (i(n), j(n))$, $1 \le n \le K_l$ 이고 K_l^v 은 단어 v 의 l 번째 기준 패턴과 입력 패턴간의 공통 시간축 길이를 나타낸다. 또한 $a_{i(n)}$은 $i(n)$번째 입력 벡터이고 $b_{j(n),l}^v$ 은 단어 v의 l 번째 기준 패턴에서 $j(n)$ 번째 벡터이다.

각 단어마다 학습 데이타와 기준 패턴간의 자승 평균 오차 벡터 $\bar{\varepsilon}^v = (\bar{\varepsilon}_1^v, \cdots, \bar{\varepsilon}_P^v)$를 구해야 한다. 우선, 단어 v 의 L개 기준 패턴과 M개 학습 데이타간 최소 평균 거리 D^v (L=1인 경우)는 다음과 같다.

$$D^v = \frac{1}{M} \sum_{m=1}^{M} \min_{1 \le l \le L} \left[\frac{1}{K_{l,m}^v} \sum_{n=1}^{K_{l,m}^v} d(a_{i(n),m}^v, b_{j(n),l}^v) \right], \quad 1 \le v \le V \quad \cdots\cdots(4)$$

여기서 $a_{i(n),m}^v$ 는 단어 v 의 m번째 학습 데이타의 $i(n)$번째 벡터이고 $K_{l,m}^v$ 은 단어 v 의 l 번째 기준 패턴과 m 번째 학습 데이타간의 공통 시간축 길이를 나타낸다.

각 단어 v 에 대한 학습 데이타와 기준 패턴간 자승 평균 오차 벡터 $\bar{\varepsilon}^v = (\bar{\varepsilon}_1^v, \cdots, \bar{\varepsilon}_P^v)$, $1 \le v \le V$를 구하기 위하여 (4)에 (1)을 적용하고 자승 평균 오차 벡터 $\bar{\varepsilon}^v$ 로 표현하면 다음과 같다.

$$D^v = \frac{1}{M} \sum_{m=1}^{M} \min_{1 \le l \le L} \left[\frac{1}{K_{l,m}^v} \sum_{n=1}^{K_{l,m}^v} \sum_{k=1}^{P} w_k' (a_{k,i(n),m}^v - b_{k,j(n),l}^v)^2 \right]$$
$$= \sum_{k=1}^{P} w_k' \frac{1}{M} \sum_{m=1}^{M} \min_{1 \le l \le L} \left[\frac{1}{K_{l,m}^v} \sum_{n=1}^{K_{l,m}^v} (a_{k,i(n),m}^v - b_{k,j(n),l}^v)^2 \right]$$
$$= \sum_{k=1}^{P} w_k' \bar{\varepsilon}_k^v, \quad 1 \le v \le V \quad (5)$$

2-2. 최적 가중 함수 결정 알고리즘

최적 가중 함수를 결정하기 위하여 V개 단어에 대한 총 평균 거리 D는 다음과 같이 정의한다.

$$D = \frac{1}{V} \sum_{v=1}^{V} D^v = \sum_{k=1}^{P} w_k' \frac{1}{V} \sum_{v=1}^{V} \bar{\varepsilon}_k^v = \sum_{k=1}^{P} w_k' \bar{\varepsilon}_k \quad (6)$$

여기서 $\bar{\varepsilon} = (\bar{\varepsilon}_1, \cdots, \bar{\varepsilon}_P)$는 학습 데이타의 총 자승 평균 오차 벡터이다. 따라서 가중 함수 w'는 (6)의 총 평균 거리 D를 최소화하도록 구해져야만 한다. 이때 (6)의 총 평균 거리를 최소화하는 가중 함수는 w'에 대한 제한 조건이 없다면 $w' = 0$ 이다. 본 논문에서는 가중 함수 w'에 다음과 같은 제한 조건을 사용하였다 [11].

$$\prod_{k=1}^{P} w_k = \prod_{k=1}^{P} w_k^2 = 1 \quad (7)$$

즉, 가중 함수의 기하 평균(geometric mean)은 1이다. 그러므로 총 평균 거리 D를 최소화하는 최적 가중 함수 w'는 (6)과 (7)에 Lagrange multiplier 방법을 사용하여 구하면 다음과 같다.

$$w_k' = \frac{\sqrt[P]{\prod_{k=1}^{P} \bar{\varepsilon}_k}}{\bar{\varepsilon}_k}, \quad 1 \le k \le P \quad (8)$$

(8)에서 알 수 있듯이 가중 함수 w_k' 는 기준 패턴과 학습 데이타 간 오차 분산의 역에 비례하고 오차 분산의 기하학적(geometric)인 평균에 의하여 정규화(normalized)되었다.

(8)의 가중 함수는 모든 캡스트럼 계수에 공통적으로 사용되는 가중 함수이다. 그러나 실제로 오차 벡터의 모양은 각 단어마다 다른 것이 보통이다. 그림 1에서 알 수 있듯이 각 단어마다 발생하는 오차 형태는 각 단어마다 조금씩 다르다. 따라서

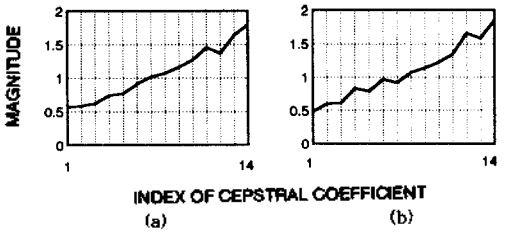

그림 1. 단어 '일'(a)과 '오'(b)의 자승 평균 오차 벡터.

각 단어마다 고유의 가중 함수를 갖는 것이 필요하다. 이러한 가중 함수 $w'_v = (w'_{v1}, w'_{v2}, \cdots, w'_{vP})$는 $w'_{vk} = w^2_{vk}$, $1 \le k \le P$, $1 \le v \le V$로 정의한다. 따라서 각 단어의 가중 함수는 다음과 같은 형태의 제한 조건을 사용하여 구한다.

$$\prod_{k=1}^{P} w'_{vk} = \prod_{k=1}^{P} w^2_{vk} = 1 , \quad 1 \le v \le V \tag{9}$$

따라서 총 평균 거리 D를 최소화하는 단어 v 에 대한 최적 가중 함수 w'_{vk} 는 Lagrange multiplier 방법을 사용하여 해를 구하면 다음과 같다.

$$w'_{vk} = \frac{\sqrt[P]{\prod_{k=1}^{P} \bar{\varepsilon}^{v}_{k}}}{\bar{\varepsilon}^{v}_{k}} , \quad 1 \le k \le P, \ 1 \le v \le V \tag{10}$$

(8)와 (10)의 최적 가중 함수를 구하기 위해서는 자승 평균 오차 벡터 $\bar{\varepsilon}$ 와 $\bar{\varepsilon}^{v}$ 를 구해야만 한다. 그러나 $\bar{\varepsilon}$ 와 $\bar{\varepsilon}^{v}$ 는 기준 패턴과 학습 데이타간 정합되는 오차 벡터로 부터 구하여 진다. 이때 정합되는 경로는 가중 캡스트럼 거리 측정 방법의 가중 함수에 따라서 달라지므로 w'_k 와 w'_{vk} 는 순환적으로 구해야 한다. 따라서 가중 함수의 초기값은 기존 캡스트럼 거리 측정 방법인 $w'_k = 1$, $1 \le k \le P$ 또는 $w'_{vk} = 1$, $1 \le k \le P$, $1 \le v \le V$로 한다. 이러한 가중 함수를 초기값으로하여 각 단어 v 마다 기준 패턴과 학습 데이타간 자승 평균 오차 벡터 $\bar{\varepsilon}^{v}$ 와 총 자승 평균 오차 벡터 $\bar{\varepsilon}$ 를 구한다. 다음은 이러한 오차 벡터를 사용하여 최적 가중 함수 w'_k 와 w'_{vk} 를 구한다. 이렇게 최적 가중 함수가 추정되면 학습 데이타와 기준 패턴간 오차 벡터의 형태가 바뀌므로 오차 벡터와 최적 가중 함수가 수렴할 때까지 위의 과정을 반복한다.

Ⅲ. 실험 및 결과 고찰

실험에서는 기존에 제안된 가중 함수들과 본 논문에서 제안된 최적 가중 함수의 성능을 비교하기 위하여 DTW 이용한 음성 인식 시스템를 사용하여 화자 독립 단독음 인식 실험을 수행하였다. 또한 벡터 양자화기를 사용하는 이산 관찰 HMM에 최적 가중 함수를 적용하여 기존에 제안된 가중 함수와의 성능을 비교하였다. 표 1은 실험에 사용된 기존 가중 함수이다. 여기서 w_k($w_k = \sqrt{w_k}$) 는 (1)의 가중 함수이다.

표 1. 기존 가중 함수
Table 1. Conventional weight functions

가중 함수	w_k , $k=1,\cdots,P$	사용된 변수값
CEP	1	
RPS	k	
SLL	$k^s e^{-k^2/2\sigma^2}$	s = 1.0, σ = 5 - 25
BPL	$1. + 0.5L\sin(\pi k/L)$	L = 15 - 30
GEL	k^s	$0 \le s \le 1$
IVL	$\frac{1}{\sigma_k}$	σ_k^2는 캡스트럼 계수 분산

3-1. 데이타 베이스

음성 인식에 사용되는 데이타 베이스는 11개 숫자음(0,1, ···,9,공)과 3개 명령어(걸어, 취소, 다음)의 14개로 구성되었다. 학습 데이타는 20-30대 남성 화자 50명이 각 단어를 2회씩 발음한(14단어*50명*2회=1400개) 음성으로 구성되었고, 시험 데이타는 학습 데이타에 포함되지 않은 20-30대 남성 화자 20명이 각 단어를 2회씩 발음한(14단어*20명*2회=560개) 음성으로 구성되었다. 각 음성은 비교적 조용한 연구실에서 지향성 마이크(AT831b)를 사용하여 DAT(Digital Audio Tape)에 녹음되었다.

3-2. 음성 분석 및 인식 시스템 구성

음성 분석 과정은 다음과 같다. 4.5kHz의 차단 주파수(cutoff frequency)를 갖는 저역 통과 필터(low pass filter)를 통과한 음성 신호는 10kHz, 16비트로 표본화된다. 표본화된 음성 신호는 $1-0.95z^{-1}$ 의 전달 함수를 갖는 프리엠퍼시스(pre-emphasis) 필터를 사용하여 고주파 성분을 강조한다. 이러한 음성 신호는 끝점 검출[55] 과정에서 묵음(silence)과 음성으로 구분된다. 검출된 음성 신호는 20ms(200샘플)의 크기를 갖는 해밍 창을 사용하여 10ms씩 이동하면서 차수 P=14인 선형 예측 계수를 구하는 LPC 분석 과정을 거친다. 이러한 LPC 계수로부터 인식 과정에 사용될 LPC 캡스트럼 계수를 LPC 계수와 동일한 차수까지 구한다.

DTW를 이용한 화자 독립 단독음 인식 시스템의 구성은 다음과 같다. 각 단어의 기준 패턴은 단어당 100개의 학습 데이타를 Modified K-Means(MKM) 알고리즘을 사용하여 군집화하여 각 단어당 최대 12개의 기준 패턴을 생성하였다. 인식 과정은 입력 음성에 대하여 기준 패턴과의 거리를 구한후 KNN 법칙을 사용하여 최소 오차를 갖는 단어를 입력 음성 단어로 결정하였다.

3-3. DTW를 사용한 인식 시스템의 최적 가중 함수 성능 평가

본 절에서는 DTW를 사용한 음성 인식 시스템의 성능을 향상시키기 위한 최적 가중 함수와 기존에 제안된 여러가지 가중 함수의 성능을 비교하였다. 사용된 기존 가중 함수는 1과 같다.

각 가중 함수의 파라메터는 단어당 4개의 기준 패턴을 갖는 DTW를 사용한 단독음 인식 실험에서 파라메터 값을 변화시키면서 인식 실험을 수행한 결과, SLL은 s = 1.0, a = 10, BPL은 L = 21, GEL은 s = 0.5일때 각각 최대 인식율을 얻었다. 이러한 파라메터를 갖는 가중 함수의 형태는 그림 2와 같다.

또한 DTW를 사용한 음성 인식 시스템의 최적 가중 함수를 구하기 위하여 (8)과 (10)의 가중 함수 w_k 와 w_{vk} , $1 \le k \le P$, $1 \le v \le V$를 반복적으로 구하였다. 이때 가중 함수는 2-3회 정도 반복하면 수렴하지만 수렴을 관찰하기 위하여 10회 반복시켰다. 여기에서는 최적 가중 함수 w_k와 w_{vk} 를 각각 OPT-I과 OPT-II로 정의하였다.

그림 2는 각 단어당 4개의 기준 패턴을 사용하고 KNN=1의 결정 법칙을 사용했을 때, 최적 가중 함수 OPT-I을 나타낸다. OPT-I의 모양은 기존 가중 함수와 마찬가지로 낮은 차수의 캡스트럼 계수에는 작은 가중을 두고 높은 차수에는 큰 가중을 둔다.

그림 3에는 여러가지 가중 함수를 사용하여 화자 독립 단독음 인식 실험을 수행한 결과이다. 각 단어당 1, 4, 12개의 기준 패턴을 갖는 인식 시스템에 대한 8가지 가중 함수의 성능을 비교하였다. 이때 최적 가중 함수 OPT-I과 OPT-II는 음성 인식 시스템의 형태에 따라서 달라지므로 단어당 기준 패턴 갯수에 따라서 각각 구하였다.

그림 3에서 알 수 있듯이 제안된 최적 가중 함수 가중 함수 OPT-I, OPT-II C는 기준 패턴 갯수의 갯수에 상관없이 항상 기존 가중 함수보다 높은 인식 성능을 나타내었다. 특히 기준 패턴의 갯수가 증가함에 따라서 각 거리 측정 방법은 비슷한 인식 성능을 나타내었다. 즉, 제안된 가중 함수는 인식 성능을 최대로 하는 최적의 가중 함수로 생각할 수 있다.

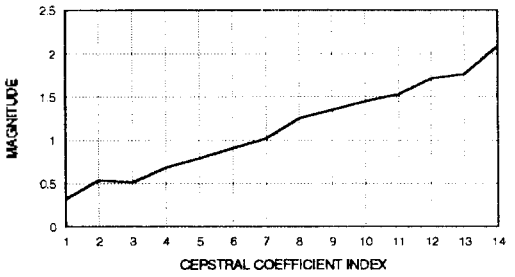

그림 2. DTW를 사용한 음성 인식 시스템에 대한 최적 가중 함수

3-4. HMM을 이용한 음성 인식 시스템의 최적 가중 함수 성능 평가

본 절에서는 DTW를 사용한 음성 인식 시스템을 이용하여 구한 최적 가중 함수들을 HMM에 적용하였다. 이때 사용한 HMM은 벡터 양자화기를 사용하는 이산 관찰 HMM으로서, 최적 가중 함수를 사용한 가중 캡스트랄 거리 측정 또는 최적 가중 캡스트랄 거리 측정 방법과 벡터 양자화기를 사용하여 관찰열을 만들었다. 이산 관찰 HMM의 성능은 관찰열에 따라 큰 영향을 받으므로 안정된 코드열을 만들 수 있는 거리 측정 방법

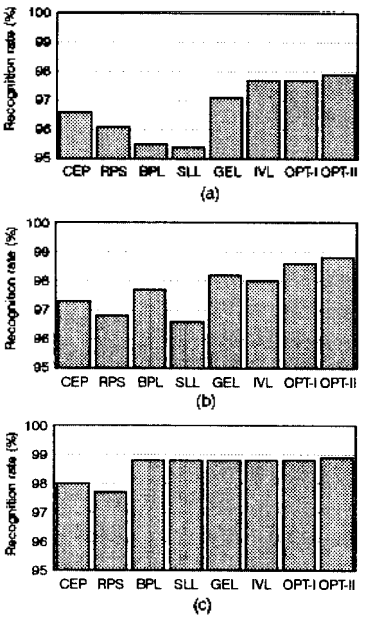

그림 3. DTW를 사용한 음성 인식 시스템에 대한 가중 함수에 따른 화자 독립 단독음 인식 결과 - 단어당 기준 패턴 갯수
(a) 1 (b) 4 (c) 12

을 사용하는 벡터 양자화기는 HMM을 이용한 음성 인식 시스템의 성능을 향상 시킬 수 있다. 기존 가중 함수와 성능을 비교하기 위하여 표 1에 정리된 가중 함수들을 사용한 HMM과 성능을 비교하였다. 사용된 최적 가중 함수는 표 1의 가중 함수와 각 단어당 4개의 기준 패턴을 갖는 DTW를 사용한 음성 인식 시스템에서 구한 OPT-I(그림 2)를 사용하였다.

표 2는 여러가지 가중 함수를 사용하였을 경우의 화자 독립 단독음 인식 오차이다. 표에서 알 수 있듯이 최적 가중 함수를 제외하고는 캡스트랄 거리 측정 방법인 CEP에 비하여 오차가 증가하는 것을 알 수 있었다. 이러한 것은 캡스트랄 거리 측정 방법이 다른 가중 함수를 사용하는 방법보다 안정된 관찰열을 발생 시켰다는 것을 나타낸다.

표 2. 가중 함수를 사용한 벡터 양자화기와 이산 관찰 HMM을 이용한 화자 독립 단독음 인식 오차(%)

가중 함수	number of state							
	3	4	5	6	7	8	9	10
CEP	4.1	4.6	5.9	4.6	3.9	3.8	2.7	2.7
RPS	7.0	7.3	6.3	6.1	5.4	5.2	5.5	4.6
BPL	6.1	5.9	7.1	4.8	4.2	3.8	3.2	3.9
SLL	7.1	7.7	7.3	4.8	5.9	4.8	5.0	4.8
GEL	6.3	5.9	5.2	5.4	4.5	4.5	4.5	3.6
IVL	5.7	6.9	5.4	4.1	3.8	4.1	3.2	2.9
OPT-I	5.0	3.9	3.4	2.9	2.3	1.6	1.8	2.9

Ⅳ 결 론

본 논문에서는 가중 켑스트랄 거리 측정 방법을 사용하는 음성 인식 시스템의 성능을 향상시키기 위하여 가중 켑스트랄 거리 측정 방법의 최적 가중 함수 설계 방법을 제안하였다. 가중 함수 또는 리프터는 인식 결과에 직접적으로 영향을 미치는 기준 패턴과 학습 데이타간 오차를 최소화하도록 결정되었다. 즉, 제안된 방법에서는 각 차수의 켑스트럼 계수에서 발생하는 오차에 역비례하게 가중함수를 결정하여 각각의 계수가 균일한 오차를 갖게 하였다. 거리 측정을 사용하는 음성 인식 시스템에서는 최종 인식의 결정이 기준 패턴과 입력 패턴과의 거리에 의하여 결정되기 때문에, 이러한 거리를 최소화하는 가중 함수를 사용하여 인식 시스템의 성능을 향상시킬 수 있었다. 이러한 알고리즘은 가중 함수를 사용하지 않는 켑스트랄 거리 측정 방법을 사용하는 음성 인식 시스템을 사용하여 최적 가중 함수가 결정될 때까지 반복적으로 구현되었다.

제안된 최적 가중 함수의 성능을 평가하기 위한 DTW를 이용한 단독음 인식 실험에서 제안된 방법은 기존 가중 함수보다 좋은 성능을 나타냈다. 또한, HMM을 이용한 음성 인식 시스템에 최적 가중 함수를 적용하기 위하여 이산 관찰 HMM에 DTW를 이용한 음성 인식 실험에서 구한 최적 가중 함수를 적용하여 기존 가중 함수와 성능을 비교하였다. 실험 결과에서 최적 가중 함수를 사용한 HMM이 가장 우수한 성능을 얻을 수 있었다.

참고문헌

[1] F. Itakura and S. Saito, "A Statistical Method for Estimation of Speech Spectral Density and Formant Frequencies," *Electron Commun. Japan*, Vol. 53-A, pp 36-43, 1970.

[2] A. H. Gray and Jr., J. D. Markel, "Distance Measures for Speech Processing," *IEEE Trans. Acoust., Speech, Signal Processing*, Vol. ASSP-24, No. 5, pp. 380-391, Oct. 1976.

[3] N. Nocerino, F. K. Soong, L. R. Rabiner and D. H. Klatt, "Comparative Study of Several Distance Measures for Speech Recognition," in *Proc. ICASSP*, pp. 25-28, March 1985.

[4] Y. T. Lee and D. Kahn, "Information-Theoretic Distance Measures for Speech Recognition : Theoritical Considerations and Experimental Results," in *Proc . ICASSP*, pp 785-788, April 1990.

[4] F. Itakura and T. Umezaki, "Distance Measure for Speech Recognition based on the Smoothed Group Delay Spectrum," in *Proc. ICASSP*, pp. 1257-1260, April 1987.

[5] J. Junqua and H. Wakita, "A Comparative Study of Cepstral Lifters and Distance Measures for All Pole Models of Speech in Noise," in *Proc. ICASSP*, pp. 476-479, May 1989.

[6] B. A. Hanson and H. Wakita, "Spectral Slope Distance Measure with Linear Prediction Analysis for Word Recognition in Noise," *IEEE Trans. Acoust., Speech, Signal Processing*, Vol. ASSP-35, No. 7, pp. 968-973, July 1987.

[7] B. H. Juang, L. R. Rabiner and J. G. Wilpon, "On the Use of Bandpass Liftering in Speech Recognition," *IEEE Trans. Acoust., Speech, Signal Processing*, Vol. ASSP-35, No. 7, pp. 947-954, July 1987.

[8] K. Shikano and M. Sugiyama, "Evaluation of LPC Spectral Matching Measures for Spoken Word Recognition," Trans. IECE, Vol. J65-D, No. 5, pp. 535-541, May 1982.

[9] Y. Tohkura, "A Weighted Cepstral Distance Measure for Speech Recognition," *IEEE Trans. Acoust., Speech, Signal Processing*, Vol. ASSP-35, No. 10, pp. 1414-1422, Oct. 1987.

[10] F. K. Soong and A. E. Rosenberg, "On the Use of Instantaneous and Transitional Spectral Information in Speaker Recognition," *IEEE Trans. Acoust., Speech, Signal Processing*, Vol. ASSP-36, No. 6, pp. 871-879, June 1988.

[11] R. H. Wang, L. S. He and H. Fujisaki,"A Weighted Distance measure on the Fine Structure of Feature Space : Application to Speaker Recognition," in *Proc . ICASSP*, pp 273-276, April 1990.

[12] H. Sakoe and S. Chiba, "Dynamic Programming Algorithm Optimization for Spoken Word Recognition," *IEEE Trans. Acoust., Speech, Signal Processing*, Vol. ASSP-26, No. 1, pp. 43-49, Feb. 1978.

[13] L. R. Rabiner, "A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition," *Proc. IEEE*, Vol. 77, No. 2, pp. 257-286, Feb. 1989.