DBpia

論文 95-7-7-1

# 음성 인식을 위한 최적 가중 켑스트랄 거리 측정 방법

正會員 金 元 九*, 尹 大 熙**

# Optimally Weighted Cepstral Distance Measure for Speech Recognition

Weon Goo Kim*, Dae Hee Youn** Regular Members

## 要 約

본 논문에서는 가중 켑스트랄 거리 측정(weighted cepstral distance measure) 방법의 최적 가중 함수(optimal weight function) 설계 방법을 제안하였다. 기존에 제안된 가중 함수 또는 켑스트랄 리프터(cepstral lifter)의 형태가 강조하고자 하는 스펙트럼 성분에 따라서 경험적으로 얻어져야만 했던 것과는 다르게, 본 논문에서는 각 단어의 기준 패턴과 학습 데이타간의 오차를 최소화하는 최적 가중 함수 또는 리프터(lifter) 설계 방법을 제안하였다.

거리 측정 방법을 사용하는 음성 인식 시스템에서는 최종 인식의 결정이 기준 패턴과 입력 데이타간의 오차에 의하여 결정되기 때문에, 이러한 오차를 최소화하는 가중 함수를 사용하여 인식 시스템 성능을 향상시킬 수 있었다.

제안된 최적 가중 함수의 성능을 평가하기 위하여 Dynamic Time Warping(DTW)과 Hidden Markov Models(HMM's)을 사용하는 두가지 음성 인식 시스템을 대상으로 화자 독립 단독음 인식 실험을 수행하여 기존에 제안된 가중 함수와 성능을 비교하였다. 실험 결과에서 최적 가중 함수는 기존에 제안된 여러가지 가중 함수보다 우수한 성능을 나타내었다.

## ABSTRACT

In this study, a method for designig 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. In this paper, a method which minimizes the error between word reference patterns and the tranining data is proposed. Speech recognition systems using distance measures depend on the error between reference patterns and the input data for recognition. Using a weight function that minimizes this error improves the recognition system performance.

*군산대학교 전기공학과
**연세대학교 전자공학과
論文番號 : 94133-0517
接受日字 : 1994年 5月 17日



www.dbpia.co.kr

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

## 1. 서 론

거리 측정(distance measure) 방법은 음성 신호의 차이를 측정하는 방법으로 음성 인식[1], 화자 인식[2]등에 적용되어 왔다. 상이성(dissimility) 또는 거리 측정은 최종 인식 결정이 이러한 거리를 기준으로 결정되기 때문에 매우 중요하다.

최근에는 음성 인식에 중요한 스펙트럼 성분들(포만트 등)은 강조하고 불필요한 성분들(스펙트럼 기울기, 채널 특성등)은 억압하는 가중 함수를 켑스트럼 계수에 사용하여 인식 성능을 향상시키는 연구가 많이 진행되었다[7-11]. 이러한 과정을 켑스트랄 리프터링(cepstral liftering)이라고 하며, 음성 및 화자 인식에서 좋은 성능을 나타내었다[8-12]. 이러한 가중 함수 또는 리프터(lifter)를 사용하는 가중 켑스트랄 거리 측정(weighted cepstral distance measure) 방법중의 한가지가 스펙트럼 기울기에 의한 거리 측정 방법(spectal slope distance measure)[9]으로서 선형 리프터(linear lifter)를 사용한다. 이러한 방법은 스펙트럼 기울기, 채널 특성과 같은 스펙트럼의 광대역 변화에 덜 민감하고 포만트 주파수 변화에 민감한 특징을 갖는다. 이러한 방법 이외에도 선형 리프터가 높은 차수 켑스트럼 계수를 지나치게 강조하는 것을 보완한 스무딩(smoothing)된 선형 리프터[7,8], 밴드 패스(band-pass) 리프터[10]등이 제안되어 좋은 성능을 나타내었다.

또다른 형태의 가중 함수로는 각 차수의 켑스트럼 계수 분산의 역을 가중 함수로 사용하여 분산을 균일하게 만드는 방법이 제안되었다[13-15].

이와 같이, 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 사용되어 온 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.

즉, 제안된 방법에서는 각 차수의 켑스트럼 계수에서 발생하는 오차에 역비례하게 가중 함수를 결정하여 각각의 계수가 균일한 오차를 갖도록 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.

또한 본 논문에서는 Dynamic Time Warping (DTW)[16]와 Hidden Markov Model's(HMM's)[17]를 사용한 두가지 종류의 음성 인식 시스템에 최적 가중 함수를 적용한 가중 켑스트랄 거리 측정 방법을 사용하여 기존 가중 함수와 성능을 평가하였다.

2절에서는 가중 켑스트랄 거리 측정 방법의 최적 가중 함수 결정 방법을 제안하였고 3절에서는 최적 가중 함수의 성능을 평가하기 위한 실험 및 결과 고찰을 기술하였으며 4절에서는 결론을 맺었다.

## 2. 가중 켑스트랄 거리 측정 방법의 최적 가중 함수 결정 방법

본 연구에서는 가중 켑스트랄 거리 측정 방법을 사용하는 음성 인식 시스템의 성능을 개선하기 위하여 최적 가중 함수 결정 알고리즘을 제안하였다. 본 절에서는 제안된 알고리즘을 기술하기 위하여 DTW를 이용한 인식



www.dbpia.co.kr

시스템을 대상으로 최적 가중 함수 설계 방법을 유도하였다. 우선, 가중 켑스트랄 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 벡터, $w = (w_1, w_2, \cdots, w_p)$는 가중 함수 또는 리프터(lifter)이고 $P$는 켑스트럼 계수의 차수이다. 또한 최적 가중 함수를 유도하기 위하여 $w' = (w'_1, w'_2, \cdots w'_p)$는 $w'_k = w^2_k$, $1 \le k \le P$로 정의하였다.

제안된 방법은 학습 데이타와 기준 패턴간 오차를 최소화하는 최적 가중 함수를 결정하는 것이다. 따라서 두 벡터간 자승 오차 벡터(squared error vector) $\epsilon \equiv (\epsilon_1, \cdots, \epsilon_p)$를 $\epsilon = (a_k - b_k)^2$, $k=1, \cdots, P$이라 정의하면 (1)의 가중 켑스트랄 거리 측정 방법은 (2)와 같이 변형된다.

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

즉, 두 벡터간의 가중 켑스트랄 거리는 오차 벡터 원소(element)의 합과 같다. 따라서 두 패턴간 평균 거리는 정합된 벡터들간에 발생하는 오차 벡터들의 평균으로부터 구할 수 있다. 이러한 자승 평균 오차 벡터(mean squared error vector) $\bar{\epsilon} \equiv (\bar{\epsilon}_1, \cdots, \bar{\epsilon}_p)$는 각 단어마다 학습 데이타와 기준 패턴간 인식 실험을 수행하여 구한다.

### 2.1. DTW를 이용한 인식 시스템의 평균 거리와 오차 벡터

인식 시스템은 $V$개 단어를 인식하기 위하여 각 단어마다 $M$개 학습 데이타를 집단화(clustering) 방법을 사용하여 만든 $L$개의 기준 패턴을 가지고 있다고 가정하였다. 또한 결정 법칙은 Nearest Neighbor(NN) 법칙을 사용한다고 가정하였다. 이때 $N$개 특징 벡터로 구성된 입력 패턴을 $\{a_1, \cdots, a_i, \cdots, a_N\}$으로 표현하면 단어 $v$의 기준 패턴간 최서 거리 $d^v$는 와핑 경로를 사용하여 다음과 같이 표현할 수 있다.

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

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

각 단어 $v$에 대한 학습 데이타와 기준 패턴간의 자승 평균 오차 벡터 $\bar{\epsilon}^v \equiv (\bar{\epsilon}^v_1, \cdots, \bar{\epsilon}^v_p)$, $1 \le v \le V$를 구하기 위하여 단어 $v$의 $L$개 패턴과 $M$개 학습 데이타간 최소 평균 거리 $D^v$는 다음과 같다.

$$D^v = \frac{1}{M} \sum_{m=1}^{M} \min_{1 \le l \le L} \left[ \frac{1}{K^v_{l,m}} \sum_{n=1}^{K^v_{l,m}} d(a^v_{i(n),m}, b^v_{j(n),l}) \right], \quad 1 \le v \le V \quad (4)$$

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

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

$$\begin{aligned} D^v &= \frac{1}{M} \sum_{m=1}^{M} \min_{1 \le l \le L} \left[ \frac{1}{K^v_{l,m}} \sum_{n=1}^{K^v_{l,m}} \sum_{k=1}^{P} w_k' (a^v_{k,i(n),m} - b^v_{k,j(n),l})^2 \right] \\ &= \sum_{k=1}^{P} w_k' \frac{1}{M} \sum_{m=1}^{M} \min_{1 \le l \le L} \left[ \frac{1}{K^v_{l,m}} \sum_{n=1}^{K^v_{l,m}} (a^v_{k,i(n),m} - b^v_{k,j(n),l})^2 \right] \\ &= \sum_{k=1}^{P} w_k' \bar{\epsilon}^v_k, \qquad 1 \le v \le V \end{aligned} \quad (5)$$

### 2.2 최적 가중 함수 결정 알고리즘

최적 가중 함수를 결정하기 위하여 $V$개 단어에 대한 총 평균 거리 $D$는 다음과 같이 정의한다.

$$D = \frac{1}{V} \sum_{v=1}^{V} D^v = \sum_{k=1}^{P} w_k' \frac{1}{V} \sum_{v=1}^{V} \bar{\epsilon}^v_k = \sum_{k=1}^{P} w_k' \bar{\epsilon}_k = \sum_{k=1}^{P} w_k^2 \bar{\epsilon}_k \quad (6)$$



www.dbpia.co.kr

여기서 $\bar{\epsilon} \equiv (\bar{\epsilon}_1, \cdots, \bar{\epsilon}_p)$는 학습 데이타의 총 자승 평균 오차 벡터이다.

따라서 가중 함수 w'는 (6)의 총 평균 거리 $D$를 최소화도록 구해져야만 한다. 이때 (6)의 총 평균 거리 또는 오차를 최소화하는 가중 함수는 w'에 대한 제한 조건이 없다면 w' = 0이다. 본 논문에서는 가중 함수 w'에 다음과 같은 제한 조건을 사용하였다[14].

$$\prod_{k=1}^{P} w'_k = \prod_{k=1}^{P} w'^2_k = 1 \tag{7}$$

즉, 가중 함수의 기하 평균(geometric mean)은 1이다.

그러므로 총 평균 거리 $D$를 최소화하는 최적 가중 함수 w'는 (6)과 (7)에 Lagrange multiplier 방법을 사용하여 해를 구하면 다음과 같다.

$$w'_k = \frac{\sqrt[P]{\prod_{k=1}^{P} \bar{\epsilon}_k}}{\bar{\epsilon}_k}, \quad 1 \le k \le P \tag{8}$$

(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. 가중 함수에 대한 다른 형태의 제한 조건으로는 $\sum_{k=1}^{P} w'^{\alpha}_k = C$을 고려할 수 있다[14]. 여기서 $\alpha$는 실수이고 C는 상수이다. 예로, $\alpha$=2인 경우, 최적 가중 함수 $w'_k$는 $\bar{\epsilon}_k$, $1 \le k \le P$ 중에서 최소값을 갖는 차수 $j$에 대해서만 $w_j = \sqrt{C}$의 값을 갖고 나머지 차수 $k(k \neq j)$에 대해서는 0의 값을 갖는다. 이러한 형태의 가중 함수는 켑스트럼 계수중에서 가장 작은 오차를 갖는 한개의 차수만을 사용하는 것이므로 비정상적인 형태이다. 또한, $\alpha \neq 2$인 경우, 여러가지 형태의 제한 조건이 가능하지만 최대의 인식 성능을 나타내는 $\alpha$의 값은 실험적으로 구해져야만하는 문제점이 있다.

(8)의 가중 함수는 모든 켑스트럼 계수에 공통적으로 사용되는 가중 함수이다. 그러나 실제로 오차 벡터의 모양은 각 단어마다 다른 것이 보통이다. 따라서 각 단어마다 고유의 가중 함수를 갖는 것이 필요하다. 이러한 가중 함수 $w' = (w'_{v1}, w'_{v2}, \cdots, w'_{vp})$는 $w'_{vk} = w'^2_{vk}$, $1 \le k \le P$, $1 \le v \le V$로 정의한다. (5)의 단어 v와 기준패턴간의 최소 평균 거리 $D^v$를 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 단어 $v$의 평균 거리 Dv를 최소화하는 최적 가중 함수 $w'_{vk}$는 Lagrange multiplier 방법을 사용하여 해를 구하면 다음과 같다.

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

(8)와 (10)의 최적 가중 함수를 구하기 위해서는 자승 평균 오차 벡터 $\bar{\epsilon}$ 와 $\bar{\epsilon}^v$를 구해야만 한다. 이러한 $\bar{\epsilon}$와 $\bar{\epsilon}^v$는 기준 패턴과 학습 데이타간 정합되는 오차 벡터로 부터 구하여 진다. 이때 정합되는 경로는 가중 켑스트랄 거리 측정 방법의 가중 함수에 따라서 달라지므로 $w'_k$와 $w'_{vk}$는 순환적으로 구해야 한다. 가중 함수의 초기값은 $w'_k$=1, $1 \le k \le P$ 또는 $w'_{vk}$=1, $1 \le k \le P$, $1 \le v \le V$로 한다. 이러한 가중 함수는 기존 켑스트랄 거리 측정 방법인 유클리디안(Euclideam) 거리 측정 방법이다. 이러한 가중 함수를 초기값으로하여 각 단어 v마다 기준 패턴과 학습 데이타간 자승 평균 오차 벡터 $\bar{\epsilon}^v$와 총 자승 평균 오차 벡터 $\bar{\epsilon}$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 패턴간의 정합되는 경로가 달라져서 오차 벡터의 형태가 바뀌므로, 오차 벡터와 최적 가중 함수가 수렴할 때까지 위의 과정을 반복한다.

## 3. 실험 및 결과 고찰

실험에서는 기존에 제안된 가중 함수들과 본 논문에서 제안된 최적 가중 함수의 성능을 비교하기 위하여 DTW 이용한 음성 인식 시스템을 사용하여 화자 독립 단독음 인식 실험을 수행하였다. 또한 벡터 양자화기를 사용하는 이산 관찰 HMM에 최적 가중 함수를 적용하여 기존에 제안된 가중 함수와의 성능을 비교하였다.

### 3.1 데이타 베이스(data base)

음성 인식에 사용되는 데이타 베이스는 11개 숫자음(0, 1, ⋯, 9, 공)과 3개 명령어(걸어, 최소, 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 2



www.dbpia.co.kr

회씩 발음한(14 * 20명 * 2회=560개) 음성으로 구성되었다. 각 음성은 비교적 조용한 연구실에서 지향성 마이크(AT831b)를 사용하여 DAT(Digital Audio Tape)에 녹음되었다.

### 3.2 음성 인식 시스템 구성

#### 3.2.1 음성 분석

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 프리엠파시스(pre-emphasis) 필터를 사용하여 고주파 성분을 강조한다. 이러한 음성 신호는 끝점검출 과정에서 묵음(silence)과 음성으로 구분된다. 검출된 음성 신호는 20ms(200 샘플)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 켑스트럼 계수를 LPC 계수와 동일한 차수까지 구한다.

#### 3.2.2 기존 가중 함수

표 1은 실험에 사용된 기존 가중 함수이다. 여기서 $w_k$, $1 \le k \le P$는 (1)의 가중 함수이다.. 표에서, 분산을 구하기 위하여 50명의 20-30대 남성이 발음한 한국어 숫자음(0,1,…,9,공)과 명령어(걸어, 최소, 다음)을 사용하여 구한 켑스트럼 계수를 사용하여 가중 함수를 구하였다.

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, $\sigma$=5-25 |
| BPL | $1+0.5L\sin(\pi k/L)$ | $L$=15-30 |
| GEL | $k^s$ | $0 \le s \le 1$ |
| IVL | $\frac{1}{\sigma_k}$ | $\sigma^2_k$는 켑스트럼 계수 분산 |

#### 3.2.3 음성 인식 시스템

DTW를 이용한 화자 독립 단독음 인식 시스템의 기준패턴은 단어당 100개의 학습 데이타를 Modified K-Means (MKM) 알고리즘을 사용하여 집단화하여 각 단어당 최대 12개의 기준 패턴을 생성하였다. 이때, 거리 측정 방법은 가중 함수를 사용하지 않았다(CEP). 인식 과정은 입력 음성에 대하여 기준 패턴과의 거리를 구한후 K-Nearest Neighbor(KNN) 법칙을 사용하여 최소 오차를 갖는 단어를 입력 음성 단어로 결정하였다.

또한 HMM을 이용한 화자 독립 단독음 인식 시스템에 사용된 HMM은 이산 관찰 HMM으로서 모델 형태는 Lef-to-riht 모델을[17] 사용하였고 Baum-Welch 알고리즘을 사용하여 HMM을 학습시켜서 단어당 1개의 모델을 생성하였다. 사용된 코드북은 모든 단어의 학습 데이타와 LBG 알고리즘을 사용하여 만는 256개로 구성하였고, 코드북을 만들때도 가중 함수를 사용하지 않는 거리 측정 방법(CEP)를 사용하였다. HMM을 이용한 음성 인식에서는 Viterbi 알고리즘으로 입력 관찰열과 모델간의 유사도를 측정하여 최대 확률을 갖는 모델을 입력 단어로 결정한다.

### 3.3 DTW를 이요한 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 표 1과 같다. 각 가중 함수의 파라메터를 구하기 위하여 DTW를 사용한 단독음 인식 실험에서 파라메터 값을 변화시키면서 인식 실험을 수행한 결과, SLL은 $s$ = 1.0, $\sigma$ = 10, BPL은 $L$ = 21, GEL은 $s$=0.5일 때 각각 최대 인식율을 얻었다. 이러한 파라메터를 갖는 가중 함수의 형태는 그림 1(a)와 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 구하였다. 이때 가중함수는 반복 과정에서 3~4회 정도에서 수렴하지만 수렴을 관찰하기 위하여 10회 반복시켰다. 여기에서는 최적 가중 함수 $w_k$와 $w_{vk}$를 각각 OPT-I과 OPT-Ⅱ로 정의하였다.

그림 1(b)는 각 단어당 4개의 기준 패턴을 사용했을



www.dbpia.co.kr

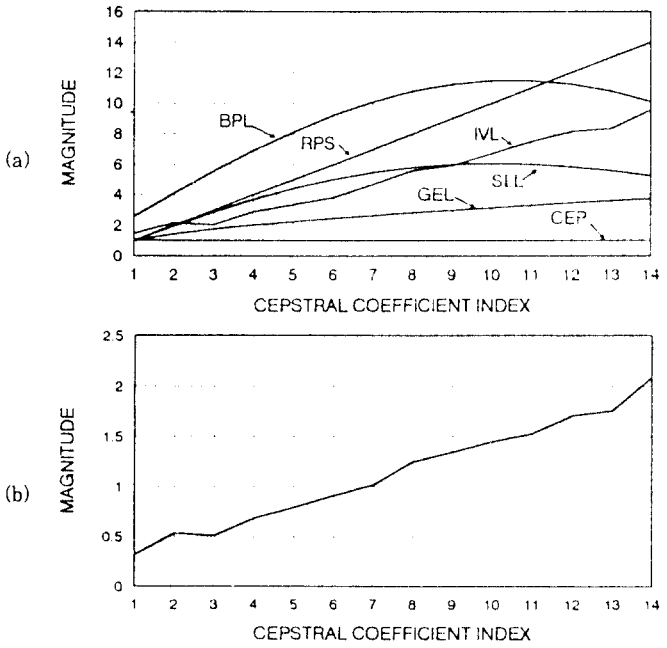


그림 1. DTW를 사용한 인식 시스템에 대한 기존 가중 함수들(a)과 최적 가중 함수(OPT-I)(b).
Fig. 1. Conventional weight function(a) and optimal weight function(OPT-I)(b) for DTW based word recognizer

때, 최적 가중 함수 OPT-I을 나타낸다. OPT-I의 모양은 기존 가중 함수와 마찬가지로 낮은 차수의 켑스트럼 계수에는 작은 가중을 두고 높은 차수에는 큰 가중을 둔다.

그림 2은 여러가지 가중 함수를 사용하여 화자 독립 단독음 인식 실험을 수행한 결과이다. 각 단어당 1, 4, 9, 12개의 기준 패턴을 갖는 인식 시스템에 대한 8가지 가중 함수의 성능을 비교하였다. 최적 가중 함수는 각 단어당 기준 패턴이 1개 일때 KNN=1, 4개 일때 KNN=1, 9개 일때 KNN=4, 12개 일때 KNN=4의 결정 법칙을 사용하는 인식 시스템을 대상으로 구하여진다.

그림에서 알 수 있듯이 최적 가중 함수 OPT-I, OPT-Ⅱ는 기존 가중 함수보다 단어당 기준 패턴의 갯수가 적은 경우에 기존 가중 함수보다 큰 성능 차이를 보이며, 기준 패턴의 갯수가 증가함에 따라서 CEP와 RPS를 제외한 모든 가중 함수는 시스템의 최대 인식율로 생각되는 98.8%에 도달한다.

최적 가중 함수 OPT-I은 각 단어당 12개 기준 패턴을 사용한 경우, BPL, SLL, GEL, IVL과 같은 98.8%의 인식율을 보였고 OPT-Ⅱ는 98.9%로 인식율을 나타내었다. 특히 4개의 기준 패턴을 사용한 경우 OPT-Ⅱ는 12개의 기준 패턴을 사용한 경우와 같은 98.8%의 인식율을 보였고 OPT-I도 98.6%의 인식율을 나타내었다. 이러한 것은 최적 가중 함수를 사용하면 인식 시스템의 성능 저하없이 기준 패턴의 갯수를 1/3로 크게 줄일 수 있고 인식 시간도 1/3로 감소된다는 것을 의미한다.

### 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



www.dbpia.co.kr

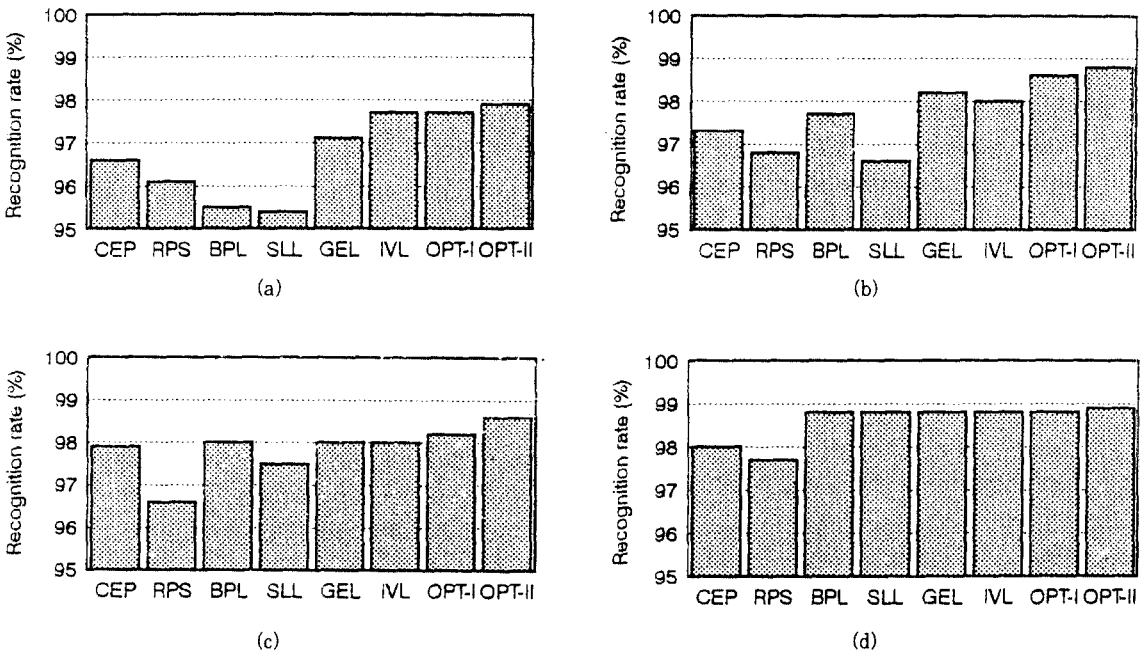


그림 2. DTW를 사용한 음성 인식 시스템에 대한 가중 함수에 따른 화자 독립 단독음 인식 결과.
(단어당 기준 패턴 갯수 (a)1 (b)4 (c)9 (d)12)

Fig. 2. The results of speaker independent isolated word recognition experiments according to the weight functions using DTW based recognizer.
(number neference pattern per word is (a)1 (b)4 (c)9 (d)12)

관찰 HMM으로서, 기존 가중함수들과 제안된 최적 가중 함수를 사용하는 가중 켑스트랄 거리 측정 방법과 벡터 양자화기를 사용하여 관찰열을 만들었다. 사용된 최적 가중 함수는 표 1의 가중함수와 각 단어당 4개의 기준 패턴을 갖는 DTW를 사용한 음성 인식 시스템에서 구한 OPT-I(그림 1(b))을 사용하였다.

표 2는 여러가지 가중 함수를 사용하였을 경우의 화자 독립 단독음 인식율이다. 표에서 알 수 있듯이 최적 가중 함수를 제외하고는 켑스트랄 거리 측정 방법의 가중 함수인 CEP에 비하여 인식율이 다소 감소하는 것을 알 수 있다. 이러한 것은 코드북 생성시 CEP를 사용하는 켑스트랄 거리 측정 방법을 사용하였기 때문이라고 생각된다. 최적 가중 함수 OPT-I은 CEP의 최대 인식율 97.3%보다 1.1% 향상된 98.4%의 인식율을 보였다. 이러한 이유는 DTW를 사용한 음성 인식실험에서 구한 최적 가중 함수가 각 단어의 오차를 최소화하도록

표 2. 가중 켑스트랄 거리 측정 방법과 이산 관찰 HMM을 이용한 화자 독립 단독음 인식율(%)

Table 2. Error rates of speaker independent isolated word recognition using weighted cepstral distance measure and discrete HMM(%)

| 가중 함수 | number of state | | | | | |
|---|---|---|---|---|---|---|
| | 5 | 6 | 7 | 8 | 9 | 10 |
| CEP | 94.1 | 95.4 | 96.1 | 96.2 | 97.3 | 97.3 |
| RPS | 93.7 | 93.9 | 94.6 | 94.8 | 94.5 | 95.4 |
| BPL | 92.9 | 95.2 | 96.8 | 96.2 | 96.8 | 96.1 |
| SLL | 92.7 | 95.2 | 94.1 | 95.2 | 95.0 | 95.2 |
| GEL | 94.8 | 94.6 | 95.5 | 95.5 | 95.5 | 96.4 |
| IVL | 94.6 | 95.9 | 96.2 | 95.9 | 96.8 | 97.1 |
| OPT-I | 96.6 | 97.1 | 97.7 | 98.4 | 98.2 | 97.1 |



www.dbpia.co.kr

구하여졌기 때문에 HMM의 벡터 양자화기에서도 각 단어마다 구별되는 안정된 관찰열을 구하여졌기 때문에 HMM의 벡터 양자화기에서도 각 단어마다 구별되는 안정된 관찰열을 발생시켰기 때문이다. 이와 같이 최적가중 함수는 HMM과 같은 통계적인 방법을 사용하는 음성 인식 시스템에 직접 적용하여 설계할수는 없지만, 위와 같이 벡터 양자화기에 최적 가중 켑스트랄 거리 측정 방법을 적용하여 HMM을 이용한 음성 인식 시스템의 성능을 개선시킬 수 있었다.

## 4. 결 론

본 절에서는 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. 이러한 알고리즘은 가중 함수를 사용하지 않는 켑스트랄 거리 측정 방법을 사용하는 음성 인식 시스템을 사용하여 최적 가중 함수가 결정될 때까지 반복적으로 구현되었다.

본 논문에서는 두가지 형태의 최적 가중 함수를 제안하였다. 첫째, 기존 가중 함수처럼 모든 단어의 켑스트럼 계수에 공통적으로 적용되는 한개의 최적 가중 함수(OPT-I) 설계 방법이다. 또 다른 방법은 각 단어마다 고유의 최적 가중 함수(OPT-Ⅱ)를 갖는 방법을 제안하였다. 이러한 방법은 한개의 가중 함수를 사용할 때보다 좋은 성능을 얻을 수 있었다. 이러한 이유는 각 단어마다 발생하는 오차 벡터의 형태가 다르기 때문에 각 단어마다 고유의 가중 함수가 필요하기 때문이다.

제안된 최적 가중 함수의 성능을 평가하기 위한 DRW를 이용한 단독음 인식 실험에서, 제안된 방법은 기존 가중 함수보다 좋은 성능을 나타냈다.

또한, 본 논문에서는 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, "Minimum Prediction Residual Priniciple Applied to Speech Recognition," *IEEE Trans. Acoust., Speech, Signal Processing*, Vol. ASSP-23, No. 1, pp.67-72, Feb., 1975.
2. B. S. Atal, "Effectivencess of Linear Prediction Characteristics of the Speech Wave for Automatic Speaker Identification and Verification", *J. Acoust. Soc. Am*, Vol. 55, pp.1304-1312, June, 1974.
3. 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.
4. A. H. Gray and Jr., J. D. Markel, "Distance Measrues for Speech Processing," *IEEE Trans. Acoust., Speech, Signal Processing*, Vol. ASSP-24, No. 5, pp.380-391, Oct., 1976.
5. 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.
6. Y. T. Lee and D. Kahn, "Information-Theoretic Distance Measure Measures for Speech Recognition : Theoritical Considerations and Experimental Results," in *Proc. JCASSP*, pp.785-788, April, 1990.
7. 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.
8. 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.
9. B. A. Hanson and H. Wakita, "Spectral Slope Distance Measure with Linear Prediction Analysis for Word Recogntion in Noise," *IEEE Trans. Acoust., Speech, Signal Processing*,




www.dbpia.co.kr

Vol. ASSP-35, No. 7, pp.968-973, July, 1987.
10. B. H. Juang, L. R. Rabiner and J. G. Wilpon, "On the Use of Bandpass Liftering in Speech Recognition," *IEEE Trans. Acoust., Signal Processing*, Vol. ASSP-35, No. 7, pp.947-954, July, 1987.
11. 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.
12. 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.
13. 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.
14. R. H. Wang, L. S. He and H. Fujiski, "A Weighted Distance measure on the Fine Structure of Feature Space : Application to Speaker Recognition," in *Proc. ICASSP*, pp.273-276, April, 1990.
15. P. Meyer, J. Schroeter and M. M. Sondhi, "Design and Evaluation of Optimal Cepstral Lifters for Accessing Articulatory Codebooks," *IEEE Trans. Signal Processing*, Vol. ASSP-39, No. 7, pp.1493-1502, July, 1991.
16. C. Myers, L. R. Rabiner and A. E. Rosenberg, "Performance Tradeoffs in Dynamic Time Warping Algorithm for Isolated Word Recognition," *IEEE Trans. Acoust., Speech, Signal Processing*, Vol. ASSP-28, No. 6, pp.623-634, Dec., 1980.
17. L. R. Rabiner, "A Tutorial on Hidden Markov Models and Seleted Applications in Speech Recognition," *Proc. IEEE*, Vol. 77, No. 2, pp.257-286, Feb., 1989.

**金 元 九**(Weon Goo Kim) 정회원

1983년 3월~1987년 2월 : 연세대학교 전자공학과 학사
1987년 9월~1989년 8월 : 연세대학교 전자공학과 석사
1989년 9월~1994년 2월 연세대학교 전자공학과 박사
1994년 9월~현재 군산대학교 전기공학과 전임강사
※주관심 분야 : 음성 및 디지탈 신호처리, 음성 인식, 음성 통신 등임

**尹 大 熙**(Dae Hee Youn) 정회원

현재 : 연세대학교 전자공학과 교수
한국 통신학회 논문지, Vol 20. No. 2, Feb. 1995. 참조.



www.dbpia.co.kr