

## 가중 투영 우도 측정 및 병렬 모델 결합을 이용한 잡음 환경에서의 음성 인식

### Speech Recognition in the Noisy Environment using Weighted Projection-Based Likelihood Measure and Parallel Model Combination

신 원 호\*, 양 태 영\*, 김 원 구\*\*, 윤 대 회\*, 차 일 환\*

(Won-Ho Shin\*, Tae-Young Yang\*, Weon-Goo Kim\*\*, Dae-Hee Youn\*, Il-Whan Cha\*)

※본 논문은 한국과학재단의 1997년도 수탁과제연구 자원에 의해 수행되었습니다.

#### 요 약

본 논문에서는 잡음이 존재하는 환경에 강인한 것으로 알려져 있는 투영 방법을 우도 측정에 가중 함수와 결합하여 사용하는 방법을 제안하였다. 반연속 HMM을 이용한 고립 단어의 인식 실험 결과, 제안한 방법이 실험에 사용된 잡음의 환경들에서 모두 좋은 성능을 나타내었다. 아울러 병렬 모델 결합 방법을 반연속 HMM에 적용하였는데 이는 코드북의 변환만으로 쉽게 잡음의 특성을 반영할 수 있다. 기중 투영 우도 측정 방법을 병렬 모델 결합 방법에 적용한 경우에도 우수한 성능을 거둘 수 있었다.

#### ABSTRACT

This paper proposes the method that combines projection technique known to be robust to noisy environments with weighting function in likelihood measure. From experimental results that use semi-continuous hidden Markov model in isolated word recognition, the proposed method shows good performance in the several noisy environments. And parallel model combination method is applied to semi-continuous hidden Markov model, which can easily compensate for noise characteristics through transforming codebook only. It also shows good performance when this method is applied with weighted projection-based likelihood measure.

#### I. 서 론

음성 인식 시스템이 점차로 실용화됨에 따라 환경 변화에 따른 인식 시스템의 성능에 관한 연구가 많이 이루어지고 있다[1]. 실제로 인식 시스템이 사용되는 환경은 여러 가지 형태의 잡음 및 주변 환경의 특성이 영향을 미치므로 먼저 인식 시스템이 사용되는 장소나 공간의 특성 및 그곳에 존재하는 잡음의 특성을 조사할 필요가 있다. 인식 시스템에 영향을 미치는 요소는 주변에 존재하는 잡음과 같은 부가적인 성분과 채널이나 마이크로폰을 통한 콘벌루션 특성을 갖는 것으로 구분할 수 있다. 부가적인 잡음이라고 하여도 여러 가지 종류의 잡음에 모두 강인한 인식 시스템을 구성하기는 실제로 매우 어렵다.

인식 환경에 존재하는 잡음 및 환경의 특성을 바탕으로 잡음의 영향을 최소로 할 수 있는 인식 시스템을 구성하여야 한다. 이러한 방법은 대체로 인식 시스템 자체가 잡음에 강인한 특성을 갖도록 하거나, 잡음의 특성을 제거 또는 이에 적응하는 형태로 구분할 수 있다. 먼저 첫 번째로 인식 시스템 자체가 잡음에 강인한 특성을 갖기 위하여 잡음에 강인한 특징 벡터나 거리 측정 방법을 이용한다. 두 번째로 포함된 잡음의 특성을 제거하거나 이에 적응하기 위해서는 인식 환경과 동일한 환경에서 인식 시스템을 학습하거나 인식 환경의 특성을 인식 시스템에 반영하도록 한다.

본 논문에서는 통계적인 특성을 이용하여 음성을 모델링하는 HMM(hidden Markov model)[2][3]에 대하여 위와 문제들을 해결하고자 하였다. 잡음의 경우에는 부가적인 주변 잡음의 영향을 고려하였으며 잡음의 영향을 최소화하기 위해서는 앞서 이야기한 두 가지 방법을 고려하였다. 먼저 인식 시스템이 잡음에 강인한 특성을 갖기 위해

\*연세대학교 전자공학과

\*\*군산대학교 전기공학과

접수일자: 1997년 10월 20일

잡음에 강인한 거리측정 방법으로 알려진 가중 함수와 투영 방법을 결합하는 가중 투영 거리 측정 방법을 제안하였다. 다음으로 시스템에 잡음의 특성을 반영하기 위해 병렬 모델 결합 방법(parallel model combination)[4]을 이용하였다. 본 논문에서 사용한 반연속 HMM[3]의 경우 시스템의 코드북을 이용하여 학습 모델이 구성되는데 잡음의 특성을 반영하기 위해서는 가우시안 확률 분포를 나타내는 코드북의 평균과 분산을 변화시킴으로써 쉽게 처리할 수 있다. 따라서 변환된 학습 모델에 대하여 별도의 계산량의 추가 없이 인식 과정을 수행할 수 있다.

본 논문에서는 음성 인식 시스템이 많이 사용될 수 있는 전시장이나 컴퓨터실의 잡음 및 백색 잡음을 이용하여 구성된 인식 시스템의 성능을 평가하였다. II에서는 가중 투영 거리 측정 및 병렬 모델 결합 방법을 나타내었고 III에서 이를 이용한 실험 및 인식 결과를 나타내었다. 그리고 IV에서 인식 결과에 대한 결론을 맺었다.

II. 가중 투영 우도 측정 및 병렬 모델 결합

2.1 가중 투영 우도 측정 (weighted projection-based likelihood measure)

2.1.1 투영 거리 측정 (projection-based distance measure)

D.MANSURE와 B.H. JUANG[5] 등은 켈스트럼 계수를 특징 벡터로 사용하여 특징 벡터간의 거리 측정 방법을 변화시켜서 주위 환경의 변화에 대한 영향을 덜 받을 수 있는 거리 측정 방법을 제안하였다. 이들은 실험적으로 백색 잡음이 섞였을 때의 켈스트럼 계수는 다음과 같은 특징이 있음을 파악하였다.

- 1) 켈스트럼 계수의 놈(norm)이 줄어든다.
- 2) 놈(norm)이 작은 켈스트럼 계수가 놈(norm)이 큰 것보다 더 많이 영향을 받는다.
- 3) 벡터 놈(norm)은 소음에 의하여 영향을 받는 것보다 켈스트럼 벡터의 방향(orientation or direction)에 의해 적게 영향을 받는다.
- 4) 낮은 차수보다 높은 차수의 켈스트럼 계수가 놈(norm)이 적게 줄어든다.

이와 같은 실험을 바탕으로 제안한 거리 측정 방법은 다음과 같다.

$$d_{CP1} = \left| \left( \frac{c_t}{|c_t|} \right) - \left( \frac{c_r}{|c_r|} \right) \right|^2 = 2(1 - \cos \beta) \tag{1}$$

$$d_{CP2} = |c_t|(1 - \cos \beta) = |c_t| - \frac{c_t \cdot c_r}{|c_r|} \tag{2}$$

여기서  $\cos \beta = \frac{c_t \cdot c_r}{|c_t||c_r|}$  으로서 비교되는 두 벡터들 간의 방향 코사인(directional cosine)이고  $|c|$ 는 벡터의 norm 을 의미한다.

식 (2)에서  $|c_r|$ 는 상수이므로 인식 결과에 영향을 미치

지 못하고  $\hat{c}_r$ 을 정규화된(normalized) 기준 켈스트럼 벡터라 하면 다음과 같이 쓸 수 있다.

$$d_{CP2} = -c_t \cdot \hat{c}_r \tag{3}$$

2.1.2 투영 기반 우도 측정 (projection based likelihood measure)

연속 및 반연속 밀도 함수를 적용한 HMM 음성 인식 시스템의 경우 가우시안 확률 밀도 함수를 이용하게 된다. B. Carlson과 M. Clements[6]는 투영 거리 측정을 확률적인 방법에 적용한 우도 측정 방법을 제시하였다. 가우시안 확률 밀도 함수  $b_i(c_t)$ 는 다음과 같이 나타낼 수 있다.

$$b_i(c_t) = N(c_t, u_i, C_i) = (2\pi)^{-\frac{N}{2}} |C_i|^{-\frac{1}{2}} \exp \left( -\frac{1}{2} (c_t, u_i)^T C_i^{-1} (c_t, u_i) \right) \tag{4}$$

$c_t$  : 관측 벡터(observation vector)

$u_i$  : 모델의  $i$ 상태 켈스트럼 평균(mean cepstral vector of  $i$ th state in the model)

$C_i$  : 공분산 행렬(corresponding covariance matrix)

$N$  : 모수 벡터 차수(order of the parameter vectors)

이에 대하여 대수를 취하면 다음과 같다.

$$\log b_i(c_t) = (c_t - u_i)^T C_i^{-1} (c_t - u_i) + \log |C_i| + N \log 2\pi \tag{5}$$

(5)식의 뒤 두 항은 관측 값과 무관하므로 이를 생략하고 가중 우도 측정값을  $d_{ew}$ (euclidean)라고 나타내면 다음과 같다.

$$d_{ew}(c_t, b_i) = (c_t - u_i)^T C_i^{-1} (c_t - u_i) \tag{6}$$

여기에서 백색 잡음에 의한 놈(norm) 감소를 보상하기 위한 스케일 요소를 고려하여 (4), (5)식을 나타내면 다음과 같이 나타낼 수 있다.

$$\tilde{b}_i(c_t) = N(c_t, \lambda_{i,t} u_i, C_i) = (2\pi)^{-\frac{N}{2}} |C_i|^{-\frac{1}{2}} \exp \left( -\frac{1}{2} (c_t, \lambda_{i,t} u_i)^T C_i^{-1} (c_t - \lambda_{i,t} u_i) \right) \tag{7}$$

$$\log \tilde{b}_i(c_t) = (c_t - \lambda_{i,t} u_i)^T C_i^{-1} (c_t - \lambda_{i,t} u_i) + \log |C_i| + N \log 2\pi \tag{8}$$

직교 이론(orthogonality principle)[5]으로부터 최적의  $\lambda_{i,t}$ 는 공분산 행렬이 포함된 경우 다음과 같이 나타낸다.

$$\lambda_{i,t} = \frac{c_t^T C_i^{-1} u_i}{u_i^T C_i^{-1} u_i} \tag{9}$$

이를 이용하여 투영 우도 측정 방법을  $d_{CP3}$ 라고 나타내면 (10)과 같이 된다.

$$d_{CP3}(c_i, b_i) = \left( c_i - \frac{c_i^T C_i^{-1} u_i}{u_i^T C_i^{-1} u_i} u_i \right)^T C_i^{-1} \left( c_i - \frac{c_i^T C_i^{-1} u_i}{u_i^T C_i^{-1} u_i} u_i \right) \quad (10)$$

이를 식(2)와 같이 나타내면 다음과 같고  $\cos \beta$ 는 (12)와 같이 나타낸다.

$$d_{CP3}(c_i, b_i) = \left| c_i C_i^{-\frac{1}{2}} \right|^2 (1 - \cos^2 \beta) \quad (11)$$

$$\cos^2 \beta = \frac{c_i^T C_i^{-1} u_i}{\left| c_i C_i^{-\frac{1}{2}} \right| \left| u_i C_i^{-\frac{1}{2}} \right|} \quad (12)$$

### 2.1.3 가중 투영 우도 측정 (weighted projection-based likelihood measure)

기존의 가중 함수 및 켈스트랄 리프터링 방법은 잡음에 의하여 영향을 덜 받는 부분을 강조하기 위하여 이용되어 왔다. 따라서 이를 투영 거리 및 우도 측정 방법과 결합하면 좋은 성능을 나타낼 것으로 기대할 수 있다.

따라서 우도 측정에 이를 적용하기 위해 대각 행렬인 가중 함수  $w$ 를 곱하면 식(10)은 다음과 같은 형태가 된다.

$$\begin{aligned} d_{wCP3}(wc_i, bi) &= \left( wc_i - \frac{(wc_i)^T w^{-1} C_i^{-1} w^{-1} (wu_i)}{(wu_i)^T w^{-1} C_i^{-1} w^{-1} (wu_i)} wu_i \right)^T \\ &\quad w^{-1} C_i^{-1} w^{-1} \\ &= \left( wc_i - \frac{(wc_i)^T w^{-1} C_i^{-1} w^{-1} (wu_i)}{(wu_i)^T w^{-1} C_i^{-1} w^{-1} (wu_i)} wu_i \right)^T \\ &= \left( wc_i - \frac{c_i^T C_i^{-1} u_i}{u_i^T C_i^{-1} u_i} wu_i \right)^T w^{-1} C_i^{-1} w^{-1} \\ &= \left( wc_i - \frac{c_i^T C_i^{-1} u_i}{u_i^T C_i^{-1} u_i} wu_i \right) \\ &= \left( c_i - \frac{c_i^T C_i^{-1} u_i}{u_i^T C_i^{-1} u_i} u_i \right)^T C_i^{-1} \left( c_i - \frac{c_i^T C_i^{-1} u_i}{u_i^T C_i^{-1} u_i} u_i \right) \end{aligned} \quad (13)$$

즉 공분산 행렬에도 동일한 가중이 가해지므로 가중 함수의 항이 소거되고 식(10)과 동일한 형태가 된다. 따라서 가중 함수에 의한 효과를 얻기 위하여 공분산 행렬을 고려하지 않고 (1)식 및 (2)식과 유사한 형태의 우도 측정 방법을 고려하면 다음과 같다.

$$\begin{aligned} d_{wCP1}(wc_i, bi) &= \left( \frac{wc_i}{|wc_i|} - \frac{wu_i}{|wu_i|} \right)^T C_i^{-1} \\ &\quad \left( \frac{wc_i}{|wc_i|} - \frac{wu_i}{|wu_i|} \right) \end{aligned} \quad (14)$$

$$\begin{aligned} d_{wCP2}(wc_i, bi) &= \left( wc_i - \frac{|wc_i|}{|wu_i|} wu_i \right)^T C_i^{-1} \\ &\quad \left( wc_i - \frac{|wc_i|}{|wu_i|} wu_i \right) \end{aligned} \quad (15)$$

식(14)의 경우 정규화된 관측값을 의미하므로  $c_i^{-1}$ 는 정규화된 관측값으로부터 구한 공분산 행렬을 나타낸다.

### 2.2 병렬 모델 결합 (parallel model combination)

병렬 모델 결합 방법[4]은 잡음이 없는 환경에서 학습한 모델과 배경 잡음 구간으로부터 구한 모델을 결합하여 잡음의 특성을 반영하도록 한다. 각 모델의 결합은 이들이 가지고 있는 켈스트럼 계수를 선형 주파수 영역으로 변환하여 이들을 더한 후에 다시 켈스트럼 영역으로 변환한다. 이 때 학습과 인식에 사용된 신호와의 크기 및 잡음의 정도를 비교하여 더하는 비율을 결정한다. 그림 1은 병렬 모델 결합 방법을 개략적으로 나타낸 것이다.

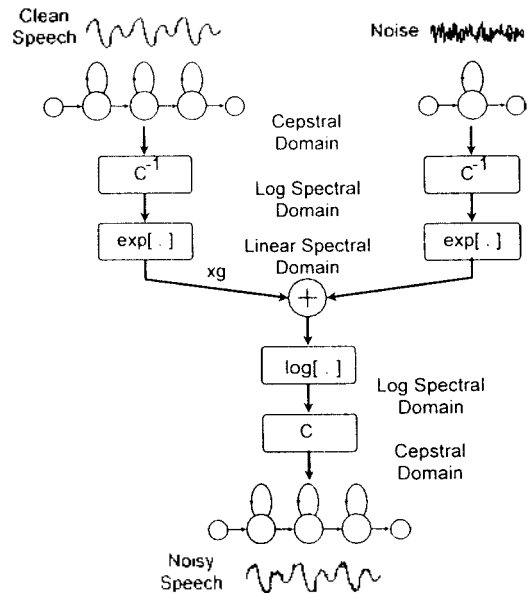


그림 1. 병렬 모델 결합 방법  
Fig. 1 Parallel model combination method

이들 과정을 보다 자세히 살펴보면 다음과 같다. 먼저 음성과 잡음이 선형 주파수 영역에서 부가적(additive)이며 독립적이라고 가정한다. 학습된 음성 및 잡음의 모델은 가우시안 분포의 평균과 분산으로 표현된다.

켈스트럼 영역에서의 평균 및 분산을  $\mu^c, \Sigma^c$ 로 나타내며 로그 스펙트럼 영역에서 이산 코사인 변환에 의해 얻어지는 켈스트럼 계수는 행렬  $C$ 로 나타낸다. 따라서 대수 스펙트럼 영역에서의 평균 및 분산을  $\mu^l, \Sigma^l$ 이라고 한

다면 이는 식(16), (17)과 같이 주어진다.

$$\hat{\mu}^t = C^{-1} \mu^c \tag{16}$$

$$\hat{\Sigma}^t = C^{-1} \Sigma^c (C^{-1})^T \tag{17}$$

켈스트럼 및 대수 스펙트럼 영역에서의 분포가 가우시안이라면 선형 영역에서의 분포는 대수 정규(log normal) 분포가 된다. 따라서 이를 이용하여 선형 영역에서  $\mu$  및  $\Sigma$ 의  $i$ 번째 성분은 다음과 같이 나타내어 질 수 있다.

$$\mu_i = e^{\mu_i^t + \Sigma_i^t/2} \tag{18}$$

$$\Sigma_{ij} = \mu_i \mu_j | e^{\Sigma_{ij}^t - 1} | \tag{19}$$

앞에서 언급한 바와 같이 선형 주파수 영역에서 음성과 잡음이 부가적이며 독립인 경우 잡음의 특성을 결합한 선형 주파수 영역의 평균과 분산은 다음과 같다.

$$\hat{\mu} = g\mu + \tilde{\mu} \tag{20}$$

$$\hat{\Sigma} = g^2 \Sigma + \tilde{\Sigma} \tag{21}$$

이 때  $\tilde{\mu}$  및  $\tilde{\Sigma}$ 는 잡음 모델의 선형 주파수 영역 모수를 나타낸 것이다. 또한  $g$ 는 학습에 이용된 신호의 크기와 잡음이 섞인 신호 크기 사이의 보상을 위하여 다음과 같이 결정된다.

$$g = \frac{E_{ns} - E_n}{E_s} \tag{22}$$

$E_s$ : 학습 신호의 평균 에너지

$E_n$ : 잡음 신호의 평균 에너지

$E_{ns}$ : 잡음이 섞인 신호의 평균 에너지

이들을 다시 로그 스펙트럼 영역으로 변환하면 (23), (24)과 같이 나타낼 수 있다.

$$\hat{\mu}_i^t = \log(\hat{\mu}_i) - \frac{1}{2} \log \left| \frac{\hat{\Sigma}_{ii}^t}{\hat{\mu}_i^2} + 1 \right| \tag{23}$$

$$\hat{\Sigma}_{ij}^t = \log \left| \frac{\hat{\Sigma}_{ij}^t}{\hat{\mu}_i \hat{\mu}_j} \right| \tag{24}$$

마찬가지로 이를 행렬  $C$ 를 이용하여 켈스트럼 영역으로 변환하여 새로운 코드북의 모수를 얻는다.

$$\hat{\mu}^c = C \hat{\mu}^t \tag{25}$$

$$\hat{\Sigma}^c = C \hat{\Sigma}^t C^T \tag{26}$$

본 논문에서는 이러한 병렬 모델 결합 방법을 반연속 HMM에 적용하였다. 반연속 HMM의 경우 코드북이 가지고 있는 가우시안 확률분포의 조합에 의해 각 상태의 확률 분포를 나타내게 된다. 따라서 코드북의 확률 분포를 잡음의 특성에 맞게 변환하면 인식단에서는 부가적인

계산량의 추가 없이 잡음이 존재하지 않는 경우와 동일하게 인식을 수행할 수 있다.

### III. 실험 및 결과

#### 3.1 인식 과정

그림 2는 위에서 설명한 가중 투영 우도 측정 및 병렬 모델 결합을 통한 반연속 HMM의 인식 과정을 나타낸 것이다. 학습 모델을 학습된 코드북에 결합함으로써 잡음 환경에서 학습한 것과 같은 확률 분포를 얻도록 한다.

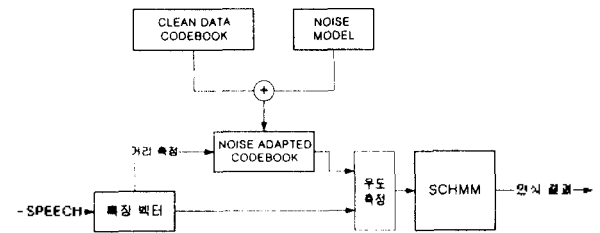


그림 2. 가중 우도 측정 및 병렬 모델 결합을 이용한 인식 시스템의 블럭다이어그램

Fig. 2 Blockdiagram of speech recognition system using weighted projection-based likelihood measure and parallel model combination

#### 3.2 데이터베이스 및 인식 시스템 구성

음성 인식 실험에 사용된 데이터 베이스는 14개 숫자와 17개 명령어로 구성된 단독음을 이용하였다. 50명이 각 단어를 2회씩 발음한 음성과 학습 데이터에 포함되지 않은 남성 화자 20명이 각 단어를 2회씩 발음한 음성을 학습 및 테스트 데이터로 이용하였다. 잡음 데이터베이스는 JEIDA (Japan Electronic Industry Development Association)의 잡음 데이터베이스, 진시장 및 컴퓨터실에서 얻은 잡음과 백색 잡음을 20, 10, 0dB의 SNR에 대하여 이용하였다.

이들 음성 신호로부터 멜 켈스트럼(MFCC: Mel Frequency Cepstral Coefficient)의 특징 벡터를 추출하였으며 이와 함께 차등 켈스트럼, 에너지와 차등 에너지를 특징 벡터로 갖는 반연속 HMM 모델을 구성하였다. 각각의 코드북은 128개(M=128)의 가우시안 분포를 갖도록 하였다. 학습과 인식 과정은 Baum-Welch 알고리즘[2][3] 및 Viterbi 디코딩 방법[2][3]을 이용하였다.

#### 3.3 실험 결과

입력 음성의 특징 벡터와 가장 유사한 코드북의 코드 벡터를 선택하기 위하여 켈스트럼간의 거리 측정을 이용하는 데, 이에는 여러 가지 가중 함수를 이용할 수 있으나 [7] 본 실험에서는 잡음이 섞여있는 경우에 좋은 성능을 나타내는 것으로 알려진 RPS(root power sum)[8]을 대표로 사용하였다. 투영 방법을 이용한 경우에도 이의 두 가지를 이용하였다. 특히 관측값을 정규화하여 투영하는

경우 정규화된 데이터들을 이용하여 초기 코드북을 생성하도록 하였다.

병렬 모델을 결합하기 위한 잡음 모델을 구성할 때, 잡음의 특성에 따라 몇 가지의 상태와 확률 분포를 갖는 모델을 구성할 수 있으나 그러한 경우 잡음의 특성을 반영한 코드북은 잡음의 확률 분포 배 수 만큼 증가하게 된다. 즉  $N$ 개의 분포 수를 갖는 잡음 모델을 이용한  $L$ 개의 크기를 갖는 코드북은  $(N \times L)$ 의 크기를 갖게 된다. 본 실험에서는 1개의 상태와 확률 분포를 갖는 잡음 모델을 가정하여 고려하였다. 그러한 경우 원래 학습된 모델과 동일한 크기의 코드북을 갖게 된다.

다음으로 잡음 모델을 구성할 경우, 각 입력 신호의 음성이 포함되지 않은 북음 구간으로부터 얻은 잡음 신호를 이용하거나 비교적 충분한 잡음 구간의 신호로부터 얻은 잡음을 이용할 수 있다. 본 실험에서는 음성이 입력되지 않은 약 200 프레임의 잡음 신호로부터 얻은 캡스트럼의 평균 및 분산을 잡음 모델로 이용하였다. 이 때 고려해야 할 사항으로는 (20), (21), (22)에 나타난  $g$ 값의 결정이다. 이  $g$ 값은 학습에 사용된 신호와 인식 실험에 이용된 신호가 대체로 비슷한 레벨로 입력되었다고 가정하여 1.0의 값을 설정하였다. 학습된 모델의 코드북은 캡스트럼의 코드북만을 변환하였고 차등 캡스트럼 및 에너지와 차등 에너지의 코드북은 변환하지 않고 학습된 것을 그대로 이용하였다.

먼저 표 1은 가중 함수 및 투영 방법을 적용한 인식 결과이다. 표에서 EUC 및 RPS는 입력 음성의 특징 벡터와 가장 유사한 코드 벡터를 선정하기 위하여 사용되는 가중 함수를 나타낸다. 그러나 이 가중 함수는 우도 측정에는

영향을 미치지 못하게 된다. 따라서 WCP1(·) 및 WCP2(·)는 (14), (15)와 같이 가중 함수를 고려한 우도 측정 방법을 나타낸다.

인식 결과를 살펴보면 알려져있는 바와 같이 가중 함수를 적용한 경우에, 그렇지 않은 경우 보다 좋은 성능을 나타내고 있음을 볼 수 있다. 다음으로 가중 우도 측정 방법의 적용에 따른 성능을 평가하였는데 이를 적용하지 않은 경우에 비하여 상대적으로 좋은 성능을 나타내고 있음을 볼 수 있다. 전체적으로 가장 좋은 성능을 나타낸 경우는 제안된  $d_{wcp2}$ 를 이용한 WCP2(RPS)의 경우인데 실험에 적용한 3가지의 잡음에 대하여 모두 좋은 성능을 보여주고 있다. WCP3의 방법도 좋은 성능을 나타내고 있으나 가중 함수의 적용이 불가능하므로 백색 잡음의 경우에서 제안된 WCP1이나 WCP2의 성능에 미치지 못하고 있다. 따라서 가중 함수를 투영 우도 측정에 적용한 제안된 방법이 여러 가지 경우의 잡음 환경에서 좋은 성능을 보이고 있다.

다음으로는 이를 병렬 모델 결합 방법에 적용하여 이의 성능을 비교하였다. 성능의 평가를 위하여 스펙트럼 차감법[9]을 이용한 인식 실험을 함께 수행하였다. 실험에서는 위의 실험에서 좋은 성능을 나타낸 WCP2(RPS)만을 병렬 모델 결합 방법에 적용하였다. 다음 표 2는 이를 이용한 인식 결과이다.

표 2에는 비교를 위하여 표 1에서 얻은 RPS 및 WCP2(RPS)의 인식 결과를 함께 나타내었다. 인식 결과를 보면 스펙트럼 차감법 및 병렬 모델 결합 방법을 적용한 경우, 예상한 대로 성능의 향상을 나타내고 있다. 전시장 잡음의 경우에는 다소의 성능 저하가 있으나 이는 잡음 제거 및 보상에 이용된 잡음의 특성이 안정적이지 않은 데서 오는 결과로 보여진다. 앞에서 좋은 성능을 나타낸 WCP2(RPS)에 병렬 모델 결합을 적용한 경우 성능의 향

표 1. 가중 투영 우도 측정을 이용한 인식 결과  
Table 1. Recognition result using weighted projection likelihood measure

우도 측정 잡음, SNR		EUC	RPS	WCP1 (EUC)	WCP1 (RPS)	WCP2 (EUC)	WCP2 (RPS)	WCP3
CLEAN		99.82	98.93	99.29	99.11	99.11	99.46	99.46
전시장	20dB	91.25	98.39	94.64	96.43	95.71	97.50	96.43
	10dB	63.04	91.07	73.57	88.75	77.68	90.00	85.18
	0dB	15.89	49.64	19.82	50.18	22.32	50.71	44.46
컴퓨터실	20dB	95.36	96.07	96.79	96.61	97.50	96.79	97.14
	10dB	80.89	86.96	90.18	88.75	92.86	90.36	90.00
	0dB	35.18	51.07	47.57	50.18	49.82	60.00	60.54
백색 잡음	20dB	67.50	92.32	74.82	95.00	84.64	96.43	87.86
	10dB	38.04	75.89	39.29	87.86	45.00	92.32	59.11
	0dB	4.46	34.29	7.86	48.21	8.04	53.75	16.61

EUC: euclidean distance measure  
 WCP1(EUC):  $d_{wcp1}^{EUC}$   
 WCP2(EUC):  $d_{wcp2}^{EUC}$   
 WCP3:  $d_{wcp3}$   
 RPS: root power sum  
 WCP1(RPS):  $d_{wcp1}^{RPS}$   
 WCP2(RPS):  $d_{wcp2}^{RPS}$

표 2. 병렬 모델 결합 방법을 이용한 인식 결과  
Table 2. Recognition result using parallel model combination

처리 방법 잡음, SNR		RPS	SUB_RPS	PMC_RPS	WCP2 (RPS)	WCP2(RPS) _PMC
전시장	20dB	98.39	94.11	97.50	97.50	97.32
	10dB	91.07	85.71	88.39	90.00	92.50
	0dB	49.64	49.11	38.39	50.71	56.79
컴퓨터실	20dB	96.07	97.14	98.04	96.79	97.50
	10dB	86.96	91.79	92.68	90.36	94.11
	0dB	51.07	74.46	57.68	60.00	66.61
백색 잡음	20dB	92.32	97.32	97.50	96.43	96.68
	10dB	75.89	90.89	93.04	92.32	93.04
	0dB	34.29	66.43	69.11	53.75	69.29

SUB\_RPS: spectral subtraction + RPS  
 PMC\_RPS: parallel model combination + RPS  
 WCP2(RPS):  $d_{wcp2}^{RPS}$   
 WCP2(RPS)\_PMC:  $d_{wcp2}^{RPS} + PMC$

상을 나타냄을 볼 수 있다. 따라서 제안된 가중 투영 우도 측정 방법은 다양한 잡음 환경 및 SNR과 병렬 모델 결합 시 모두 좋은 성능을 나타냄을 볼 수 있다.

#### IV. 결 론

본 논문에서는 투영 방법을 이용한 우도 측정이 잡음에 강인한 특성을 이용하여 이를 가중 함수와 함께 적용하는 방법을 제안하였다. 아울러 병렬 모델 결합 방법을 반연속 HMM에 적용하여 잡음의 영향을 제거하고자 하였다. 반연속 HMM에서는 코드북의 변환만으로 이루어지므로 병렬 모델 결합 방법을 적용하기 가장 적합한 구조를 가지고 있다. 실험 결과, 제안한 가중 투영 우도 측정을 이용한 경우, 기존의 방법에 비하여 향상된 인식 성능을 나타냄을 확인할 수 있었고 병렬 모델 결합 방법에 적용 시에도 향상된 우수한 성능을 나타내었다.

#### 참 고 문 헌

1. Y. Gong, "Speech Recognition in Noisy Environments: A Survey", *Speech Communication*, vol. 16, pp. 261-291, 1995.
2. L. R. Rabiner, "A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition," *Proc. IEEE*, Vol. 77, No. 2, pp. 257-285, 1989.
3. X. D. Huang, Y. Ariki and M. A. Jack, *Hidden Markov Models for Speech Recognition*, Edinburgh University Press, 1990.
4. M. J. F. Gales, S. J. Young, "Cepstral Parameter Compensation for HMM Recognition in Noise," *Speech Communication*, vol. 12, no. 3, pp. 231-239, 1993.
5. D. Mansour and B. H. Juang, "A Family of Distortion Measure Based upon Projection Operation for Robust Speech Recognition," *IEEE Trans. Acoust., Speech, Signal Processing*, Vol. ASSP-37, No. 11, pp. 1659-1671, Nov. 1989.
6. B. A. Carlson, M. A. Clements, "A Projection-Based Likelihood Measure for Speech Recognition in Noise," *IEEE Trans. Corr. Speech and Audio Processing* Vol. 2, No. 1, pp. 97-102, Jan. 1994.
7. 신 원호, 양 태영, 서 동권, 김 원구, 윤 대회, 차 일환, "부가 잡음처리를 이용한 잡음 환경에서의 음성 인식," 제13회 음성통신 및 신호처리 워크샵 1996.
8. 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.
9. S. F. Boll, "Suppression of Acoustic Noise in Speech Using Spectral Subtraction," *IEEE Trans. Acoust., Speech, Signal Processing*, Vol. ASSP-27, No. 2, pp. 113-120, April 1979.

#### ▲신 원 호(Won Ho Shin)

한국음향학회지 1996년 15권 5호 참조

현재: 연세대학교 대학원 전자공학과 박사과정

#### ▲양 태 영(Tae Young Yang)

한국음향학회지 1996년 15권 5호 참조

현재: 연세대학교 대학원 전자공학과 박사과정

#### ▲김 원 구(Weon Goo Kim)

한국음향학회지 1996년 13권 1호 참조

현재: 군산대학교 전기공학과 조교수

#### ▲윤 대 회(Dae Hee Youn)

한국음향학회지 1996년 13권 1호 참조

현재: 연세대학교 전자공학과 교수

#### ▲차 일 환(Il-Whan Cha)

한국음향학회지 제14권 4호 참조