

PCMM 기반 특징 보상 기법에서 변별력 향상을 위한 Minimum Classification Error 훈련의 적용

Minimum Classification Error Training to Improve Discriminability of PCMM-Based Feature Compensation

김 우 일*, 고 한 석**
(Wooil Kim*, Hanseok Ko**)

Dept. of Electrical and Computer Engineering, Carnegie Mellon University*, 고려대 전자컴퓨터공학과**
(접수일자: 2004년 11월 22일; 수정일자: 2004년 12월 16일; 채택일자: 2004년 12월 20일)

본 논문에서는 잡음 환경에서 강인한 음성 인식을 위하여 특징 보상 기법의 성능을 향상시킬 수 있는 방법을 제안한다. 기존의 음성 모델 기반의 특징 보상 기법에서 이용되는 오염 음성 모델 추정 방식은 입력 음성에 대한 변별력 있는 사후 확률 예측을 보장하지 못하며, 부정확하게 계산된 사후 확률은 복구된 음성에서 명료도 하락의 문제를 일으킨다. 제안하는 기법에서는 오염 음성 모델 추정 과정에 분별적 훈련 방식의 하나인 최소 분류 오류 (MCE) 훈련 기법을 도입한다. MCE 훈련 기법을 적용하기 위해 변별력 하락의 가능성을 가지는 '경쟁 요소'를 결정하는 기법을 제안한다. 병렬 결합된 혼합 모델 (PCMM) 기반의 특징 보상에 MCE 훈련 기법을 적용하는 과정을 제안하고 변별력 향상의 영향을 관찰한다. Aurora 2.0 데이터베이스와 실제 자동차 주행 환경에서 수집된 음성 데이터베이스에 대한 성능 평가를 실시한다. 실험 결과는 제안한 기법이 음성 인식 성능 향상에 도움이 되는 것을 입증한다.

핵심용어: 음성 인식, 모델 기반 특징 보상, 변별력, 최소 분류 오류 훈련, 경쟁 요소, 병렬 결합 혼합 모델

투고분야: 음성처리 분야 (2.5)

In this paper, we propose a scheme to improve discriminative property in the feature compensation method for robust speech recognition under noisy environments. The estimation of noisy speech model used in existing feature compensation methods do not guarantee the computation of posterior probabilities which discriminate reliably among the Gaussian components. Estimation of posterior probabilities is a crucial step in determining the discriminative factor of the Gaussian models, which in turn determines the intelligibility of the restored speech signals. The proposed scheme employs minimum classification error (MCE) training for estimating the parameters of the noisy speech model. For applying the MCE training, we propose to identify and determine the 'competing components' that are expected to affect the discriminative ability. The proposed method is applied to feature compensation based on parallel combined mixture model (PCMM). The performance is examined over Aurora 2.0 database and over the speech recorded inside a car during real driving conditions. The experimental results show improved recognition performance in both simulated environments and real-life conditions. The result verifies the effectiveness of the proposed scheme for increasing the performance of robust speech recognition systems.

Keywords: Speech recognition, Model-based feature compensation, Discriminative property, Minimum classification error training, Competing components, Parallel combined mixture model

ASK subject classification: Speech Signal Processing (2.5)

I. 서론

음성 인식 시스템의 보급을 가로 막는 가장 큰 요인 중 하나는 자동차 주행과 같은 실제 환경에서 인식 성능이 크게 떨어진다는 것이다. 반복된 인식 명령의 실패는 편의를 목적으로 설치된 장치가 오히려 불편을 초래하게 되어 사용자에게 사양 장치로 간주하게 되는 부정적인 영향을 끼치게 된다. 인식 성능을 하락시키는 주요한 원인은 음향 모델 파라미터의 훈련을 위해 사용되는 음성 데이터베이스와 비교하여 실제 인식 환경에서 입력되는 음성 데이터의 음향학적 특성이 달라질 수 있다는 것이다. 주행하는 차량 내부를 예로 들면, 엔진 소음, 공기 또는 지면과의 마찰, 주변 차량, 에어컨, 음악 소리 등에 의한 배경 잡음이 입력 음성을 오염시킨다. 차량 내부에 의한 반향 잡음, 마이크와 같이 음성 신호의 전달과정에서 생기는 채널의 변화 등도 주요한 잡음 요소가 된다. 실제 환경에서의 음성 인식 성능을 최대화하려는 노력의 하나로서 이와 같은 요인에 의해 발생하는 인식 환경과 훈련 환경의 차이를 줄이기 위한 연구와 기술 개발이 최근 십수년간 집중되어 왔다.

훈련 환경과 인식 환경의 음향학적 차이를 줄이기 위한 방법은 음성 인식 과정에서 어느 단계에 적용하는지에 따라 구분될 수 있다. 본 논문에서는 음성 입력 단계에서 입력된 신호로부터 직접 잡음을 제거하거나 특징 추출 단계에서 잡음 요소로 판단되는 부분을 제거 혹은 보상에 주는 전처리 기법을 다루고자 한다. 주파수 차감법 (spectral subtraction)이나 다양한 필터를 기반으로 하는 잡음 억압 및 음성 강화 기술, 켈스트럼 평균 정규화 (cepstral mean normalization, CMN), 각종 특징 보상 (feature compensation)과 같은 기법들이 전처리 단계에서 잡음 요소를 제거함으로써 깨끗한 환경에서 훈련된 음향 모델에 인식 환경의 음향학적 특성을 일치시키고자 하는 시도이다[1]. 음성 인식 성능을 향상시키기 위한 전처리 기법으로 특징 추출 단계에 관한 다양한 연구가 진행되어 왔고, 그 중에서도 음성의 음향학적 분포를 기반으로 하는 특징 보상 기법이 우수한 성능으로 주목 받고 있다.

본 논문에서는 음성의 모델을 기반으로 하는 특징 보상 기법에 관심을 갖는다. 모델 기반의 특징 보상 기법에서는 깨끗한 음성 특징의 분포를 기반으로 잡음 환경에서 어떠한 분포를 갖게 되는지를 추정하여 이를 특징 보상 과정에 이용하게 된다. 본 논문에서는 기존의 오염

음성 모델의 추정 결과로 나타나는 변별력 하락의 문제를 지적하고자 한다. 최대 사후 확률 (maximum a posteriori, MAP) 기법을 기반으로 하는 기존의 일반적인 모델 추정 방식은 분류 오류를 최소화 하는 것을 보장하지 않으므로, 오염 음성 입력에 대한 변별력 있는 사후 확률을 기대할 수 없다. 특징 보상에 이용되는 MMSE (minimum mean squared error) 예측 기법에서는 사후 확률이 중요한 역할을 하므로, 부정확한 사후 확률의 계산은 복구된 특징에서 변별력을 떨어뜨리고, 그 결과 음성 인식 성능 하락을 가져온다. 본 논문에서는 최소 분류 오류 (minimum classification error, MCE) 훈련 기법을 오염 음성 모델 추정 과정에 적용하는 것을 제안한다. 특징 보상 기술의 하나인 병렬 결합된 혼합 모델 (parallel combined mixture model, PCMM) 기반의 보상 기법에 MCE 훈련 과정을 적용함으로써 복구된 특징의 변별력을 향상시키고자 한다[2,3].

본 논문은 다음과 같이 구성된다. II장에서 일반적인 음성 모델 기반의 특징 보상 기법에 관하여 설명하고, III장에서는 발생할 수 있는 문제점을 지적한다. IV장과 V장에서는 본 논문에서 제안하는 변별력 향상을 위한 경쟁 요소의 결정 방법과 MCE 훈련 기법에 대해서 설명하고, VI장에서는 기반 기술인 PCMM 기법에 MCE 훈련 과정을 어떻게 적용하는지 상세히 기술한다. VII장에서 제안하는 기법의 성능 평가를 위해 실시한 실험과 그 결과에 관해 고찰하고 VIII장에서 결론을 맺는다.

II. 가우시안 혼합 모델 기반의 특징 보상 기법

음성 모델을 기반으로 하는 특징 보상 기법은 Acero에 의해 처음으로 소개되었으며, 그 후 Moreno가 RATS (multivariate Gaussian based cepstral normalization)라는 명칭으로 데이터 유도방식의 기법을 제안한 후 유사한 방식들의 근간이 되어 왔다[4,5]. 특징 보상의 과정은 로그 스펙트럼이나 켈스트럼 (cepstrum)과 같은 음성 특징 벡터 상에서 이루어지며, 특징 벡터의 확률 분포는 가우시안의 혼합 모델 (Gaussian mixture model, GMM) 형태로 표현된다. 잡음 환경 하에서 특징 벡터의 분포가 통계적으로 어떻게 변화하는지를 추정한 뒤, 변화된 정보를 이용하여 실제 잡음 환경에서 오염된 입력 음성 특징을 보상하는 과정으로 이루어진다. 오염된 음성의 확률 분포를 신뢰성 있게 효율적으로 추정하는 방법이 주요 문제로 다루어져 왔고, 이와 관련한 다

양한 연구가 이루어졌다. 데이터베이스를 이용하여 훈련을 통해 얻는 방법, 입력되는 음성 샘플에 대해 적응적으로 예측하는 방법, 잡음 모델과의 병렬적 결합을 통해 추정하는 방법 등이 있다.

깨끗한 음성 특징 벡터 x 의 분포는 다음과 같이 K 개의 가우시안 요소로 이루어진 혼합 모델 형태로 추정할 수 있다.

$$p(x) = \sum_{k=1}^K \omega_k N(x; \mu_{x,k}, \Sigma_{x,k}) \quad (1)$$

오염된 음성 특징 벡터 y 은 음성의 오염 과정을 수학적으로 나타내는 비선형 함수로 표현되며, 캡스트럼 특징 벡터의 경우 다음과 같이 나타낼 수 있다.

$$y = f(x, n, h) = x + h + C \log(1 + \exp(C^{-1}(n - x - h))) \quad (2)$$

식 (2)에서 h 와 n 은 각각 채널 선형 필터 응답과 부가적인 배경 잡음의 캡스트럼 벡터를 나타내며, C 는 이산 코사인 변환(DCT, discrete cosine transform) 행렬이다. 이와 같은 관계를 기초로 하여 오염된 음성 특징의 분포 역시 식 (3)과 같이 가우시안 혼합 모델 형태로 추정한다.

$$p(y) = \sum_{k=1}^K \omega_k N(y; \mu_{y,k}, \Sigma_{y,k}) \quad (3)$$

추정한 음성 모델과 MMSE 기반의 예측방법을 이용하여 오염된 입력 음성 특징 벡터로부터 깨끗한 음성 특징을 복구한다.

$$\hat{x}_{MMSE} = E\{x|y\} = \int x p(x|y) dx \cong y - \sum_{k=1}^K (\mu_{y,k} - \mu_{x,k}) p[k|y] \quad (4)$$

III. GMM 기반 특징 보상에서의 문제점

앞장의 식 (4)에서 볼 수 있듯이, MMSE 예측 기법을 사용하는 GMM 기반 특징 보상에서는 각 가우시안 요소

에 대한 보상 벡터 $(\mu_{y,k} - \mu_{x,k})$ 가 사후 확률 $p[k|y]$ 에 의해 비율이 정해진다. 즉, 각 가우시안 요소에 대한 사후 확률 값이 특징 보상의 정도를 결정하게 된다.

GMM 기반 특징 보상 기법에서는 MAP 예측 기법을 기준으로 하여 오염 음성 모델을 추정하며 일반적으로 EM(expectation maximization) 알고리즘을 적용한다. MAP 기반의 알고리즘은 각 가우시안 요소에 대한 사후 확률을 최대로 하는 것을 목표로 할 뿐, 각 요소를 변별력 있게 분류할 수 있는 성능은 보장하지 않는다. 따라서 깨끗한 환경에서는 통계적으로 달랐던 가우시안 요소가 잡음 환경에서는 MAP 예측을 통해 서로 유사하게 추정될 수 있다. 이와 같이 유사하게 추정된 가우시안 요소는 입력된 오염 음성에 대해 유사한 사후 확률을 발생시키고, 결과적으로 비슷한 비율로 보상이 이루어지게 된다. 상황에 따라서는 이와 같은 과정이 복구된 음성 특징의 변별력을 떨어뜨려 음성 인식 성능을 하락시키는 원인이 된다.

그림 1을 통해 앞에서 설명한 보상 과정에서의 문제점을 설명하고자 한다. 2차원의 특징 벡터 평면에서 8개의 가우시안 요소를 가지는 음성 모델이 잡음에 오염되어 그림과 같이 변화되었다고 가정한다. 가우시안 요소 1과 2의 경우, 깨끗한 환경에서는 상당한 거리에 위치하여 통계적으로 서로 다르지만, 잡음에 오염되었을 경우 매우 유사한 위치로 이동한다. 만일 두 요소가 같은 분산 행렬을 가진다고 가정하면, 동일한 입력 음성에 대해 유사한 사후 확률을 가지며 MMSE에 의해 두 요소에 의한 보상 벡터는 대략 0.5 $(r_1 + r_2)$ 에 사후 확률 값을 곱한 결과가 된다. 실제 환경에서 입력되는 오염 음성은 'incomplete' 데이터이므로 본래 깨끗한 환경에서 어느 가우시안 요소로부터 발생했는지 알 수 없다. 따라서 유사한 사후 확률 값을 부여하는 과정이 타당하며, 실제로

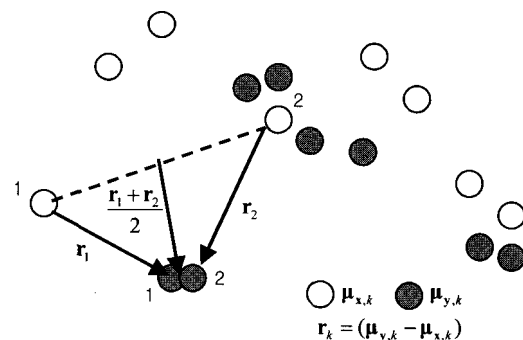


그림 1. GMM 기반 특징 보상의 예
Fig. 1. Illustration of GMM-based feature compensation.

깨끗한 음성 특징이 두 요소의 유사한 기여에 의해 발생된 것이라고 한다면, 앞에서와 같은 보상 과정은 이치에 맞다. 하지만, 요소 1 또는 요소 2 단독에 의해 발생한 것이라면 보상 벡터 r_1 또는 r_2 에 의해서 보상되어야 한다. 이런 경우에 $0.5(r_1+r_2)$ 에 의한 보상은 복구된 특징 벡터의 변별력을 감소시키는 영향을 주며, 결과적으로 음성 인식의 성능을 하락시킨다.

본 논문에서는 그림 1의 가우시안 요소1, 2와 같이 깨끗한 환경에서는 통계적으로 서로 다르지만 잡음 환경에서는 유사하게 변화하여 비슷한 사후 확률 값을 갖게 되는 요소를 '경쟁 요소 (competing components)'라 정의한다. 경쟁 요소는 복구된 특징의 변별력을 떨어뜨림으로써 음성 인식 성능을 하락시키는 가능성을 가진다. 본 논문의 목표는 변별력을 떨어뜨리는 원인이 되는 경쟁 요소를 결정하고, 변별력을 높이기 위한 방법으로 특징 보상 과정에 사용되는 오염 음성 모델 추정에 MCE 훈련 기법을 적용함으로써 음성 인식의 성능을 향상시키는 것이다. 다음 장에서는 경쟁 요소를 결정하는 방법을 다룬다.

IV. 경쟁 요소의 결정

수개의 가우시안 요소로 이루어진 음성 모델을 가정하면, 통계적 유사성의 기준을 어떻게 두느냐에 따라 다양한 조합의 경쟁 요소를 결정할 수 있다. 본 논문에서는 모델의 군집화 (clustering)를 이용한 경쟁 요소 결정 기법을 제안하고자 한다.

우선 다음과 같은 잡음 환경에서의 가우시안 요소의 변화를 가정한다.

$$\mu_{y,k} = \mu_{x,k} + r_k, \quad \Sigma_{y,k} = \Sigma_{x,k} + R_k \quad (5)$$

식 (1)과 같이 표현하는 깨끗한 환경에서의 음성 모델에서 각 가우시안 요소의 평균 벡터가 이동하고 분산 행렬이 축소 혹은 팽창하는 것으로 오염 음성 모델을 가정한다. 식 (5)와 같은 가정 하에 예측된 오염된 음성 모델은 식 (3)과 같이 나타낼 수 있다. K개의 가우시안으로 표현되는 깨끗한 음성 모델과 오염된 음성 모델을 각각 M_{clean} , M_{noisy} 개의 클러스터로 군집화 한다.

$$\begin{aligned} k \in C_i^{clean} &\leftrightarrow \text{dist}(m_k^{clean}, c_i^{clean}) < \text{dist}(m_k^{clean}, c_j^{clean}), \forall j \neq i \\ k \in C_i^{noisy} &\leftrightarrow \text{dist}(m_k^{noisy}, c_i^{noisy}) < \text{dist}(m_k^{noisy}, c_j^{noisy}), \forall j \neq i \end{aligned} \quad (6)$$

식 (6)에서 $\text{dist}()$ 는 두 가우시안 분포 사이의 통계학적 유사도를 측정하는 척도이며, m_k^{clean} 와 m_k^{noisy} 는 각각 깨끗한 음성 모델과 오염 음성 모델의 k번째 가우시안 요소를 뜻한다. c_i^{clean} 와 c_i^{noisy} 은 각각 깨끗한 음성 모델과 오염된 음성 모델의 i번째 클러스터를 나타내며, C_i^{clean} 와 C_i^{noisy} 은 i번째 클러스터에 속하는 가우시안 요소의 집합을 말한다. 군집화가 완료되면 각 가우시안 요소가 깨끗한 환경에서는 어느 클러스터에 속하는지, 오염된 환경에서는 어느 클러스터에 속하게 되는지 여부를 집합 C_i^{clean} 와 C_i^{noisy} 을 통해 알 수 있다. 집합 C_i^{clean} 와 C_i^{noisy} 을 이용하여 다음 식에 의해 경쟁 요소를 결정한다.

$$CS_{ij} = \{k | k \in C_i^{noisy} \wedge k \in C_j^{clean}\} \quad (7)$$

식 (7)에서 정의한 집합 CS_{ij} 에는 오염된 환경에서는 i번째 클러스터에 속하고 깨끗한 환경에서는 j번째 클러스터에 속하는 요소들이 포함된다. 집합 CS_{ij} 에 속하는 요소들은 깨끗한 환경에서와 오염된 환경에서 항상 유사한 통계학적 거리를 가진다. 집합 CS_{ij} 의 요소와 CS_{ik} 의 요소는 깨끗한 환경에서는 서로 다른 클러스터에 속하고 잡음 환경에서는 동일한 클러스터에 속하게 되는 요소이다. 즉, 두 집합의 요소들은 잡음 환경의 i번째 클러스터 안에서 서로 경쟁 관계를 가지게 되는 요소가 된다.

그림 2는 앞에서 설명한 경쟁 요소의 결정 과정을 보여준다. 깨끗한 환경에서 3개의 클러스터 형태로 상당한 통계적 거리를 유지하면서 존재했던 가우시안 요소들이 잡음에 오염되면서 그림의 점선과 같은 형태로 이동하게 된다. C_k^{noisy} 에 포함되는 요소 즉, 잡음 환경에서 유사하게 위치하는 요소들 중에서 C_1^{clean} , C_2^{clean} , C_3^{clean} 에 속했던 요소들을 각각 집합 CS_{k1} , CS_{k2} , CS_{k3} 으로 결정한다. 결정된 집합은 클러스터 C_k^{noisy} 안에서 사후 확률의 변별력을 떨어뜨리는 '경쟁 요소'로 작용한다.

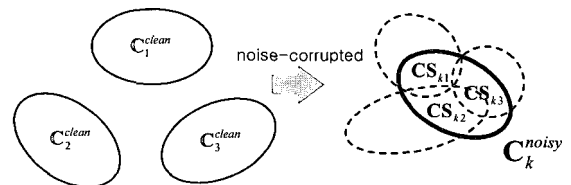


그림 2. 경쟁 요소 집합의 결정
Fig. 2. Determination of 'Competing Components' set.

V. 최소 분류 오류 기법의 적용

최소 분류 오류 (MCE) 기법은 최대 사후 확률 기법에 의한 훈련 과정이 분류 성능의 지표가 되는 분류 오류를 최소화하는 것을 보장하지 못하는 단점을 극복하기 위해 제안된 것으로 음성 인식 분야에서는 HMM의 변별력 (discriminative)을 향상시키기 위해 다양한 시도가 이루어져 왔다[6]. 본 논문에서는 경쟁 관계에 있는 요소간의 변별력 향상을 위해 오염 음성 모델의 추정 과정에 MCE 훈련 기법의 적용을 제안한다. MCE의 효율적인 적용을 위해 식 (7)에서 정의한 각 경쟁 요소 집합을 대표하는 가우시안 분포를 다음과 같이 결정한다.

$$\begin{aligned} \mu_{ij}^{CS} &= \frac{1}{N_{CS_j}} \sum_{y,k \in CS_j} \mu_{y,k \in CS_j} \\ \Sigma_{ij}^{CS} &= \frac{1}{N_{CS_j}} \sum_{y,k \in CS_j} \left\{ \sum_{y,k \in CS_j} (\mu_{ij}^{CS} - \mu_{y,k \in CS_j})(\mu_{ij}^{CS} - \mu_{y,k \in CS_j})^T \right\} \end{aligned} \quad (8)$$

식 (8)은 집합 CS_j 에 속하는 오염 음성 모델의 가우시안 요소의 중심(center)을 각 집합을 대표하는 확률 분포로 결정함을 나타내며, N_{CS_j} 는 집합 CS_j 에 속하는 원소의 개수를 의미한다. MCE 기법을 적용하기 위해서 다음과 같이 분별 함수 (discriminative function)를 정의한다.

$$g_{ij}(\mathbf{y}) = \log(p_{CS_j}(\mathbf{y} | i, j)) = \log(N(\mathbf{y} | \mu_{ij}^{CS}, \Sigma_{ij}^{CS})) \quad (9)$$

식 (9)에서 정의한 분별 함수는 오염된 음성 모델의 클러스터 i 안에서 깨끗한 음성 모델 클러스터 j 에 대한 관찰 값의 로그 우도(log-likelihood) 함수를 의미한다. 분별 함수로 정의한 우도 함수는 집합 CS_j 을 대표하므로, CS_j 에 속해 있는 가우시안 요소의 우도 함수와 유사한 경향을 가지게 된다. MMSE 과정에서 계산되는 사후 확률이 각 요소의 우도 함수에 대응되므로, 정의한 대표 우도 함수의 파라미터에 MCE 기법을 적용함으로써 각 요소의 사후 확률의 변별력을 향상시킬 수 있을 것으로 판단된다. 정의한 분별 함수를 이용하여 다음과 같이 클래스 오류분류 척도(class misclassification measure)를 정의할 수 있다.

$$d_{ij}(\mathbf{Y}) = -g_{ij}(\mathbf{Y}) + \log \left[\frac{1}{M-1} \sum_{k, k \neq j} \exp\{g_{ik}(\mathbf{Y})\} \right]^{1/n} \quad (10)$$

식 (10)에서 M 은 Class의 개수를 뜻하며, 여기에서는 i 번째 오염 음성 클러스터에 존재하는 경쟁 요소 집합의 개수를 뜻한다. 예를 들어 그림 2와 같은 경우, k 번째 오염 음성 클러스터는 3개의 경쟁 요소 집합을 갖는다. n 은 양의 상수로서 경쟁 관계에 있는 분별 함수와의 관계를 설정한다. 본 논문에서는 매우 큰 값으로 설정하여 가장 큰 분별 함수를 갖는 경쟁 요소와의 차를 클래스 오류분류 척도로 사용하였다. 다음의 식 (11)은 손실 함수 (loss function)를 뜻하며, 일반적인 sigmoid 함수를 사용하였다. γ 는 sigmoid 함수의 기울기를 결정하는 파라미터이고, θ 는 일반적으로 0으로 설정한다.

$$\lambda_{ij}(\mathbf{Y}) = \lambda(d_{ij}(\mathbf{Y})), \quad \lambda(d) = \frac{1}{1 + \exp(-\gamma d + \theta)} \quad (11)$$

MCE 알고리즘에서는 식 (11)의 손실 함수가 최소가 되는 방향으로 분류 함수의 파라미터를 갱신하는 과정이 이루어진다. 손실 함수를 최소로 하는 것은 식 (10)에서 훈련 데이터가 속하는 class의 분류 함수와 경쟁 요소의 분류 함수의 차를 증가시키는 방향을 의미하므로 본 논문에서 목적으로 하고 있는 사후 확률의 변별력 향상에 기여하게 된다. 이미 알려져 있는 HMM 파라미터의 적용 식을 바탕으로 본 논문에서 정의한 분류 함수를 표현하는 경쟁 요소 집합의 대표 파라미터는 다음과 같이 갱신할 수 있다[6].

$$\begin{aligned} \tilde{\mu}_{ij}^{CS}(n+1) &= \tilde{\mu}_{ij}^{CS}(n) - \epsilon \frac{\partial \lambda_{ij}(\mathbf{y})}{\partial \tilde{\mu}_{ij}^{CS}}, \quad \tilde{\mu}_{ij}^{CS} = \frac{\mu_{ij}^{CS}}{\sigma_{ij}^{CS}} \\ \tilde{\sigma}_{ij}^{CS}(n+1) &= \tilde{\sigma}_{ij}^{CS}(n) - \epsilon \frac{\partial \lambda_{ij}(\mathbf{y})}{\partial \tilde{\sigma}_{ij}^{CS}}, \quad \tilde{\sigma}_{ij}^{CS} = \log(\sigma_{ij}^{CS}) \end{aligned} \quad (12)$$

식 (12)에서 $\mu_{ij}^{CS}, \sigma_{ij}^{CS}$ 는 각각 경쟁 요소 집합을 대표하는 파라미터인 $\mu_{ij}^{CS}, \Sigma_{ij}^{CS}$ 의 d 번째 성분(element)을 나타낸다. 본 논문에서 가우시안 분포의 분산행렬은 대각 성분만을 가지는 것을 가정한다. 분별적 갱신 과정이 음성 모델의 파라미터 제한 조건하에서 이루어질 수 있도록 $\tilde{\mu}_{ij}^{CS}, \tilde{\sigma}_{ij}^{CS}$ 와 같은 파라미터의 변환이 필요하다. 손실 함수의 미분식은 부록에 자세하게 다루었다. 최종적인 음성 모델 파라미터의 갱신은 다음과 같이 이루어진다.

$$\begin{aligned} \hat{\mu}_{y,kd} &= \mu_{y,kd} + (\mu_{ijd}^{CS}(N_i) - \mu_{ijd}^{CS}(0)) \\ \hat{\sigma}_{y,kd} &= \sigma_{y,kd} + (\sigma_{ijd}^{CS}(N) - \sigma_{ijd}^{CS}(0)), \text{ if } k \in \text{CS}_{ij} \end{aligned} \quad (13)$$

NI은 최종 반복 횟수를 나타낸다. 식 (13)은 경쟁 요소 집합의 대표 파라미터가 MCE 알고리즘에 의해 갱신된 양만큼, 해당 집합에 속하는 오염 음성 모델의 가우시안 요소의 파라미터를 최종적으로 갱신하는 것을 의미한다. 다음 장에서는 본 논문에서 GMM 기반의 특징 보상 기법의 대상으로 고려하는 병렬 결합된 혼합 모델 (PCMM) 기반의 특징 보상 기법을 간략하게 소개하고, MCE 기법의 구체적인 적용 과정을 설명한다.

VI. MCE를 적용한 PCMM 특징 보상

PCMM 기반의 특징 보상 기법은 음성 모델을 기반으로 하는 특징 보상 기법의 하나로서, 오염 음성의 가우시안 혼합 모델 예측에 병렬 모델 결합 (parallel model combination, PMC) 기법이 적용된다[2,3]. PMC 기법은 깨끗한 음성 모델과 잡음 모델을 독립적으로 이용하여 오염된 음성 모델을 생성하는 방법으로 오염된 음성 데이터베이스를 이용한 부가적인 훈련 과정이 필요하지 않는 장점을 가진다[7]. 혼합 모델의 결합은 입력되는 음성에서 잡음을 예측하여 적용적으로 이루어질 수 있으며, 오프라인에서 미리 결합된 모델을 이용할 수도 있다. 후자의 경우에는 변하는 잡음 환경에 대처하기 위해 다중의 모델을 이용하는 기법이 제안되었다[8,9].

본 논문에서 제안하는 MCE 기법의 적용은 특징 보상에 필요한 잡음 음성 모델의 추정 과정에서 이루어진다. 우선 기존의 PCMM 과정에 의해 깨끗한 음성의 가우시안 혼합 모델과 잡음 모델을 병렬적으로 결합함으로써 오염 음성의 가우시안 혼합 모델을 추정한다. 모델의 결합은 로그 정규 가정법 (log-normal approximation)이나 데이터 유도 (data-driven) 기법을 적용할 수 있으며, 본 논문에서는 오프라인에서의 훈련 과정을 고려하므로 보다 신뢰성 있는 모델을 얻기 위해 데이터 유도 방식의 모델 결합 방식을 선택하였다. MCE 기법은 추정된 오염 음성 모델을 기본 모델로 하여, 일종의 적응 (adaptation) 형태로 모델 파라미터를 갱신하게 된다.

MCE 훈련 기법을 적용하기 위해서는 어떤 class에 속하는지 알 수 있는 레이블 (label)된 훈련 데이터가 필요

하다. 본 논문에서 제안하는 기법에서는 데이터 유도 방식의 병렬 모델 결합에서 사용되는 특징 벡터 생성 과정을 이용함으로써 MCE 훈련에 필요한 훈련 데이터를 발생시킬 수 있다. 배경 잡음에 의한 오염만을 가정할 때, 분별 함수 $g_{ij}(Y)$ 의 갱신을 위해서 발생하는 특징 벡터 Y 는 다음을 만족해야 한다.

$$\begin{aligned} y_i &= \mathbf{f}(\mathbf{x}_i, \mathbf{n}), y_i \in \mathbf{C}_i^{\text{noisy}} \wedge \mathbf{x}_i \in \mathbf{C}_j^{\text{clean}} \\ \mathbf{Y} &= \{y_1, y_2, \dots, y_T\}, \mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_T\} \end{aligned} \quad (14)$$

즉, 깨끗한 음성 모델의 j 번째 클러스터에서 발생하여 잡음 환경에서는 i 번째 클러스터에 속하게 되는 오염 음성 특징 벡터 Y 를 이용하여 분별 함수 $g_{ij}(Y)$ 를 갱신할 수 있다. 배경 잡음에 의한 오염만을 가정할 때, 식 (15)와 같이 집합 CF_{ij} 에 속하는 깨끗한 음성의 가우시안 요소로부터 발생시킨 x_k 와 잡음 모델에서 발생시킨 n 을 이용하여 오염 음성 특징 벡터 y_i 를 발생시킬 수 있다.

$$\begin{aligned} y_i &= \mathbf{x}_i + \mathbf{C}(1 + \exp(\mathbf{C}^{-1}(\mathbf{n}_i - \mathbf{x}_i))) \\ \mathbf{x}_i &\sim N(\boldsymbol{\mu}_{x,k}, \boldsymbol{\Sigma}_{x,k}), k \in \text{CF}_{ij}, \mathbf{n}_i \sim N(\boldsymbol{\mu}_n, \boldsymbol{\Sigma}_n) \end{aligned} \quad (15)$$

식 (14)와 (15)에 의해 발생시킨 훈련 데이터 Y 를 이용하여 MCE 훈련 과정을 적용하고 식 (13)에 의해 최종적인 오염 음성 모델의 파라미터를 얻을 수 있다. 이와 같이 얻어진 파라미터를 기존의 MMSE 특징 복구 과정에 적용한다.

$$\begin{aligned} \hat{\mathbf{x}}_{\text{MMSE}} &\equiv \mathbf{y} - \sum_{k=1}^K (\hat{\boldsymbol{\mu}}_{y,k} - \boldsymbol{\mu}_{x,k}) p[k | \mathbf{y}] \\ p[k | \mathbf{y}] &= \frac{\omega_k N(\mathbf{y}; \hat{\boldsymbol{\mu}}_{y,k}, \hat{\boldsymbol{\Sigma}}_{y,k})}{\sum_j \omega_j N(\mathbf{y}; \hat{\boldsymbol{\mu}}_{y,j}, \hat{\boldsymbol{\Sigma}}_{y,j})} \end{aligned} \quad (16)$$

그림 3은 본 논문에서 제안하는 MCE 훈련을 적용한 PCMM 기반의 오염 음성 모델 추정 과정을 다이어그램으로 나타낸 것이다.

VII. 실험 및 결과

7.1. Aurora 2.0에 대한 성능 평가

객관적인 성능 평가를 위해서 ELRA (European Language Resources Association)의 Aurora 2.0에서 제공하는 평가 방식을 따랐다. Aurora 2.0에서의 평가 방식의 주요 특징은 다음과 같다[10].

1) 영어 음성, 연속 숫자를 인식, 11단어+묵음 구간 (silence)+짧은 휴지 (short pause)

2) ETSI (European Telecommunications Standards Institute) 표준의 DSR(distributed speech recognition) 방식의 MFCC 특징 추출[11]

3) 13차 static 특징 (c1~c12+로그 에너지) 추출 후 인식 단계에서 미분계수 추출 (총 39차) : 본 논문의 실험에서는 PCMM 구현의 편의를 위하여 로그 에너지 대신 캡스트럼의 0차 계수를 사용하였다.

4) 3-mixture, 16-state의 단어 모델, 2종류의 silence 모델

Aurora 2.0에서 제공하는 clean-condition training, multi-condition testing 방식에 따라 음향 모델은 깨끗한 환경에서 수집된 8,840개의 음성 데이터를 이용하여 훈련하였다. Multi-condition Testing을 위한 데이터베이스 중본 논문에서는 Set A의 자동차 환경 잡음에 대한

성능 평가를 실시하였다. 잡음 환경 테스트 데이터는 7가지의 신호 대 잡음비 (SNR)에 따라 각각 1,001개의 샘플로 구성되어 있다.

기존의 대표적인 전처리 알고리즘에 대해 잡음 환경에서 베이스라인 시스템의 성능을 비교하여 그 결과를 표 1에 나타내었다. 베이스라인 성능이 Aurora 2.0에서 제공하는 베이스라인 성능 결과와 다소 다른 것은 로그에너지 대신 캡스트럼 0차 계수를 사용하였기 때문으로 판단된다. 대표적 전처리 알고리즘으로 가장 일반적으로 사용되는 주파수 차감법 (spectral subtraction, SS)을 선택하였다. 주파수 차감법에서는 배경 잡음을 추정하기 위해 250msec의 시간 지연을 갖는 최소 통계 (minimum statistics) 기법을 적용하였다[12]. PCMM은 테스트 환경과 동일한 SNR을 가지도록 잡음 모델을 선택하여 오염 음성 모델의 결합과정을 적용한 병렬 결합된 혼합 모델 기반의 특징 보상 기법을 나타낸다. 신뢰성 있는 음향 모델을 얻기 위해 데이터 유도에 의한 결합 방식을 택했다. 깨끗한 음성 특징 분포는 128개의 가우시안 요소를 가지는 GMM으로 추정하였으며, 추정 과정에서 베이스라인 시스템의 HMM 훈련과 동일한 데이터베이스를 사용하였다. 각각의 잡음 모델은 하나의 요소를 가지는 가우시안 분포로 추정하였다. MCE-PCMM은 PCMM에 의해 얻어진 오염 음성 모델에 본 논문에서 제안하는 MCE 훈련 기법을 적용한 것을 말한다.

표 1로부터 테스트 환경과 일치하는 SNR을 갖도록 결합 모델을 생성하여 적용한 PCMM 기반의 특징 보상 기법이 주파수 차감법 보다 잡음 환경에서 우수한 성능을 가지는 것을 알 수 있다. 이러한 결과는 본 논문에서 기초로 하고 있는 음성 모델 기반의 특징 보상 기법이 음성 인식 과정의 직접적인 대상이 되는 캡스트럼 영역을 깨끗한 음성 특징의 분포로 이동시킴으로써 일반적인 전처리 기법보다 음성 인식의 성능 향상 측면에서 우수함을 입증한다. 본 논문에서 제안한 MCE 기법을 적용한 경우 평균 85.88%의 단어 인식률을 얻었다. 단어 오류율 (WER, word error rate)를 기준으로 기존의 PCMM 기법에 비하여 상대적으로 약 5.74% 향상된 결과이다. 이러한 수치는 MCE 기법을 적용함으로써 다양한 SNR 환경에서 오염된 음성의 변별력을 향상시킴으로써 음성 인식 성능 향상에 도움이 되는 것을 보여주는 것이다.

표 2는 다중 (Multiple) 모델의 보간을 이용한 PCMM (interpolated PCMM, IPCMM) 기반 특징 보상 기법에 MCE 기법을 적용한 성능 평가 결과이다. 다중 모델

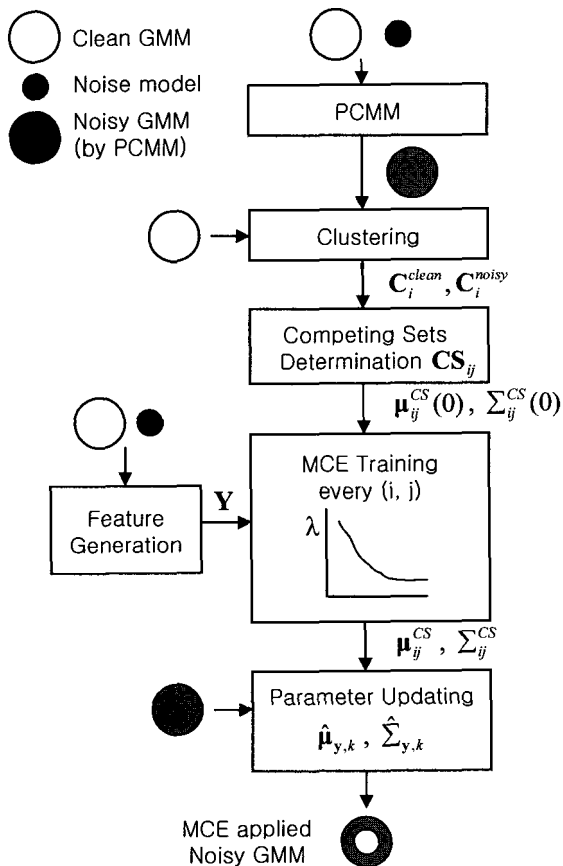


그림 3. 본 논문에서 제안하는 MCE 훈련의 적용 과정
Fig.3. Applying MCE training in the proposed scheme.

표 1. Aurora2.0의 자동차 잡음에 대한 베이스라인의 단어 인식률(%)
Table. 1. Word accuracy for baseline system to car noise condition in Aurora2.0(%).

	Baseline	SS	PCMM	MCE-PCMM
Clean	98.84	98.84	98.84	98.84
20dB	96.42	97.61	97.91	97.82
15dB	87.62	95.82	97.02	97.39
10dB	61.71	83.45	93.89	94.39
5dB	26.87	50.64	82.37	83.20
0dB	10.38	17.78	53.92	56.61
-5dB	8.41	7.52	22.16	22.00
평균	56.60	69.06	85.02	85.88

표 3. 실제 주행 환경 데이터베이스 Car01의 채널 4에 대한 베이스 라인 시스템의 단어 인식률 (%)
Table. 3. Word accuracy for the baseline system to the real car-driving condition, channel 4 of Car01 database (%).

Clean (ch1)	Noisy (ch4)	SS (ch4)	PCMM	MCE-PCMM
93.70	63.59	87.14	88.69	88.87

PCMM 기법은 복수 개의 오염 음성 모델을 사용하여 알려지지 않은 잡음 환경이나 시간에 따라 변하는 잡음 환경에 적응적으로 대처할 수 있다는 측면에서 실제 application 환경에서 PCMM 기법을 구현할 수 있도록 하는 기술이다. 알려지지 않은 SNR 환경에 적응적으로 대처할 수 있는 특징을 가지므로, 주파수 차감법과의 결합을 통해 향상된 SNR 음성에 PCMM을 적용함으로써 성능 향상을 얻을 수 있는 장점을 가진다. 다중 모델의 보간을 위해서 17dB, 7dB, -2dB의 SNR을 가지는 3개의 오염 음성 혼합 모델을 생성하였다.

표 2의 결과를 보면 다중 모델을 이용하는 PCMM 기법이 표 1의 동일한 SNR을 사용하는 단일 모델 PCMM과 평균적으로 동일한 성능을 나타내는 것을 알 수 있다. 이러한 결과는 복수개의 모델을 이용하여 다양한 SNR 환경에 효과적으로 대처할 수 있음을 보여주는 것으로, 본 논문에서 기초로 하는 PCMM 기반 특징 보상 기법이 알려지지 않은 잡음 환경에서 음성 인식 성능 향상을 위해 유용할 수 있음을 말해준다. 단일 모델 경우와 비교하여 큰 폭은 아니지만 MCE 기법을 적용한 결과에서 향상된 결과를 얻었다. 주파수 차감법을 결합한 실험에서는 주파수 차감으로 향상된 SNR 음성 입력에 PCMM 기반의 특징 보상 기법이 효과적으로 동작함을 알 수 있으며, MCE 기법을 적용한 결과에서도 향상된 결과를 얻었다.

본 실험에서는 MCE 훈련을 적용하기 위해 깨끗한 음성 모델과 오염 음성 모델 모두 8개의 클러스터로 군집

표 2. Aurora2.0의 자동차 잡음에 대한 제안한 기법의 단어 인식률(%)
Table. 2. Word accuracy for the proposed schemes to car noise condition in Aurora2.0(%).

	IPCMM	MCE-IPCMM	SS+IPCMM	SS+MCE-IPCMM
Clean	98.84	98.84	98.84	98.84
20dB	97.97	97.88	97.73	97.67
15dB	97.41	97.41	97.17	97.17
10dB	93.26	93.77	94.69	94.90
5dB	81.84	82.76	86.19	87.03
0dB	54.61	55.14	63.64	64.54
-5dB	20.73	21.32	28.93	29.11
평균	85.02	85.39	87.88	88.26

표 4. Car01의 채널 4에 대한 제안한 기법의 단어 인식률 (%)
Table. 4. Word accuracy for the proposed schemes to the channel 4 of Car01 database (%).

IPCMM	MCE-IPCMM	SS+IPCMM	SS+MCE-PCMM
87.96	88.50	89.60	89.87

화 하였다. 클러스터의 개수는 각 음성 모델이 가지는 가우시안 요소의 개수와 경쟁 요소 집합의 형태를 고려하여 결정하였다. 클러스터의 개수가 너무 적으면 특히 오염 음성 모델의 경우 일부 클러스터로 집중되는 결과를 낳게 되며, 경쟁 요소 집합을 결정하기가 어려워짐을 관찰할 수 있었다. 클러스터의 개수가 너무 많을 경우, 경쟁 관계의 집합이 너무 많아지게 되어 MCE 훈련을 적용해도 변별력 향상에 큰 도움을 주지 못하는 것을 확인하였다. MCE 기법을 적용함으로써 향상된 성능을 얻을 수 있었지만, 예상에 비해 큰 성능 향상은 아니다. 이러한 이유 중 하나는 경쟁 관계에 있는 요소들의 변별력을 향상시키는 과정에서, 선택된 경쟁 요소 집합 이외의 요소들에 대해서는 변별력을 떨어뜨릴 가능성이 있기 때문으로 판단된다. 이러한 현상은 MMSE 예측 과정에서 사후 확률을 부정확하게 추정하게 하여, 복원된 음성 특징의 명료도 향상에 부정적인 영향을 줄 수 있다. 향후 연구에서 이를 개선하는 방법 등을 다루고자 한다.

이상과 같은 실험을 통해 본 논문에서 제안하는 MCE 훈련을 적용한 PCMM 기반의 오염 음성 모델 추정 방식이 Aurora 2.0에서 제공하는 자동차 잡음의 다양한 SNR 환경에서 음성 인식 향상에 도움을 주는 것을 확인할 수 있다. 이와 같은 결과는 MCE 훈련 기법을 적용하여 추정된 오염 음성 모델이 입력되는 오염 음성에 대해 변별력 있는 사후 확률의 계산에 도움을 주며, 결과적으로 복구된 음성 특징이 음성 인식 성과와 관계된 향상된

명료도를 가지는 것을 입증한다.

7.2. 실제 자동차 주행 환경에 대한 성능 평가

제안한 기법이 실제 상황에서 효과적으로 동작하는지 알아보기 위하여 실제 자동차 주행 중에 녹음한 음성 데이터베이스에 대한 성능 평가를 실시하였다. 데이터베이스는 SITEC (Speech Information Technology & Industry Center)에서 배포한 Car01과 CarNoise01을 사용하였다[13]. Car01은 80km/h의 속도로 달리는 승용차 안에서 한국어 고립 단어 음성을 수집한 것이고, CarNoise01은 다양한 주행 조건에서 잡음 신호만을 녹음한 것이다.

Car01 데이터베이스에 포함되어 있는 548 어휘의 한국어 단어를 대상으로 인식 평가를 실시하였다. 548 어휘는 자동차 안에서 사용될 수 있는 음성 명령어로 이루어져있다. Headset 마이크로 녹음된 4,348 음성 샘플을 깨끗한 음성으로 가정하여 HMM 훈련에 사용하였고, 운전자의 햇빛 가리개에 설치된 방향성 마이크로 수집된 1,096 음성 샘플을 테스트 용 오염 음성 샘플로 사용하였다.

표 3은 Car01 데이터베이스에 대해 베이스라인 시스템과 주파수 차감법에 대한 단일 모델 PCMM 기법의 성능 비교를 나타낸 것이며, 표 4는 다중 모델 보간을 이용한 PCMM 기법에 대한 성능 평가 결과이다. PCMM은 평가 환경인 Car01과 동일한 단일 잡음 환경 모델을 사용한 것으로, CarNoise01 데이터베이스에 포함된 80km/h의 주행 속도에서 수집된 잡음 샘플로 잡음 모델을 훈련한 것이다. 다중 모델 보간을 이용한 PCMM (IPCMM)의 구현을 위해서는 CarNoise01에 있는 50km/h, 80km/h, 100km/h의 주행 조건에서 수집된 잡음 샘플을 이용하여 3가지 오염 음성 모델을 생성하여 사용하였다.

훈련에 사용된 데이터베이스와 동일한 환경인 헤드셋 마이크 입력 (ch1)에서 93.70%의 인식 성능을 나타내었고, 본 실험의 대상이 되는 햇빛 가리개에 위치한 마이크 (ch4) 입력 데이터에 대해서는 63.59%의 성능을 보였다. 주파수 차감법 (SS), 단일 모델 PCMM, 다중모델 보간 PCMM (IPCMM)의 적용결과 Aurora 데이터베이스와 동일한 경향을 나타내었으며, 주파수 차감법과 IPCMM의 결합에 있어서도 각각의 기법보다 우수한 결과를 보였다. 본 논문에서 제안한 MCE 훈련 기법의 적용을 통해서 단일 PCMM, 다중모델 PCMM, 주파수차감

법과 결합한 IPCMM 모두 MCE 기법을 적용하지 않은 경우 보다 향상된 인식 성능을 나타내었으며, 이는 Aurora 데이터베이스 실험 결과와 일치하는 경향이다. 이러한 결과는 본 논문에서 제안하는 MCE 훈련 기법을 적용한 PCMM 기반의 특징 보상 기법이 실제 잡음 환경에서도 특징 복구 과정에서 변별력을 향상시키는데 도움이 되며, 결과적으로 제안하는 전처리 기법을 음성 인식 성능 향상을 위해 활용할 수 있음을 보여준다.

VIII. 결론

본 논문에서는 음성 모델 기반의 특징 보상 기법에서 변별력을 향상시키기 위한 기법을 제안하였다. 모델 기반 특징 보상에서 요구되는 오염 음성 모델 추정 과정에 MCE 훈련 기법을 적용함으로써 입력된 오염 음성에 대해 사후 확률 계산을 위한 모델의 변별력을 높일 수 있도록 하였다. 기존의 PCMM 특징 보상 기법에 MCE 훈련 과정을 적용하여 변별력 향상을 관찰하였다. Aurora2.0과 실제 자동차 주행 환경에서 수집된 데이터베이스에 대해 성능 평가를 실시하여 기존의 기법에 비해 향상된 결과를 얻었으며, 이를 통해 제안한 MCE 훈련 과정의 적용이 변별력 향상에 효과적임을 확인하였다.

부 록

식 (12)과 관련된 손실 함수의 미분식은 다음과 같다.

$$\frac{\partial \lambda_{ij}(\mathbf{Y})}{\partial \tilde{\mu}_{ijd}^{CS}} = \frac{\partial \lambda_{ij}}{\partial d_{ij}} \frac{\partial d_{ij}}{\partial \tilde{\mu}_{ijd}^{CS}} \quad (17)$$

$$\frac{\partial \lambda_{ij}}{\partial d_{ij}} = \lambda_{ij}(d_{ij})(1 - \lambda_{ij}(d_{ij})) \quad (18)$$

$$\begin{aligned} \frac{\partial d_{ij}(\mathbf{Y})}{\partial \tilde{\mu}_{ijd}^{CS}} &= -\sum_{i=1}^T \frac{\partial \log(p_{CS}(\mathbf{y}_i | i, j))}{\partial \tilde{\mu}_{ijd}^{CS}} \\ &= -\sum_{i=1}^T \left(\frac{y_{id}}{\sigma_{ijd}^{CS}} - \tilde{\mu}_{ijd}^{CS} \right) \end{aligned} \quad (19)$$

$$\frac{\partial \lambda_{ij}(\mathbf{Y})}{\partial \tilde{\sigma}_{ijd}^{CS}} = \frac{\partial \lambda_{ij}}{\partial d_{ij}} \frac{\partial d_{ij}}{\partial \tilde{\sigma}_{ijd}^{CS}} \quad (20)$$

$$\frac{\partial d_{ij}(\mathbf{Y})}{\partial \hat{\sigma}_{ijd}^{CS}} = -\sum_{t=1}^T \left[\left(\frac{y_{td} - \mu_{ijd}^{CS}}{\sigma_{ijd}^{CS}} \right)^2 - 1 \right] \quad (21)$$

감사의 글

본 논문은 두뇌한국21사업과 한국과학재단의 해외 Post-doc, 연수지원을 통해 수행된 연구 결과의 일부입니다.

참고 문헌

1. X. Huang, A. Acero and H. Hon, *Spoken Language Processing*, Prentice Hall PTR, 2001.
2. W. Kim, S. Ahn and H. Ko, "Feature Compensation Scheme Based on Parallel Combined Mixture Model," Proc. of Eurospeech2003, pp.677-680, Sep. 2003.
3. 김우일, 이용규, 권오일, 고한석, "병렬 결합된 혼합 모델 기반의 특징 보상 기술," *한국음향학회지*, 22 (7), pp.603-611, Oct., 2003.
4. P. J. Moreno, *Speech Recognition in Noisy Environments*, PhD Thesis, Carnegie Mellon University, 1996.
5. P. J. Moreno, B. Raj and R. M. Stern, "Data-driven Environmental Compensation for Speech Recognition: A Unified Approach," *Speech Communication*, 24 (4), 267-285, July 1998.
6. B. Juang, W. Chou and C. Lee, "Minimum Classification Error Rate Methods for Speech Recognition," *IEEE Trans. On Speech and Audio Processing*, 5 (3), 257-265, May 1997.
7. M. J. F. Gales and S. J. Young, "Robust Continuous Speech Recognition Using Parallel Model Combination," *IEEE Trans. on Speech and Audio Processing*, 4 (5), 352-359, Sep. 1996.
8. W. Kim, O. Kwon, H. Ko, "PCMM-based Feature Compensation Schemes Using Model Interpolation and Mixture Sharing," Proc. ICASSP2004, 989-992, May. 2004.
9. 김우일, 고한석, "시변 잡음에 대처하기 위한 다중 모델을 이용한 PCMM 기반 특징 보상 기법," *한국음향학회지*, 23 (6), 473-480, Oct., 2004.
10. H. G. Hirsch & D. Pearce, "The AURORA Experimental Framework for the Performance Evaluations of Speech Recognition Systems under Noisy Conditions", ISCA ITRW ASR2000, Sep. 2000.
11. ETSI standard document, "Speech Processing, Transmission and Quality aspects (STQ): Distributed speech recognition: Front-end feature extraction algorithm: Compression algorithms," ETSI ES 201 108 v1.1.2 (2000-04), Feb. 2000.
12. R. Martin, "Spectral Subtraction Based on Minimum Statistics." *EUSIPCO-94*, pp.1182-1185, Sep. 1994.
13. <http://www.sitec.or.kr>

저자 약력

• 김우일 (Wooil Kim)

1996년 2월 : 고려대학교 전자공학과 (공학사)
 1998년 8월 : 고려대학교 전자공학과 (공학석사)
 2003년 8월 : 고려대학교 전자공학과 (공학박사)
 2003년 9월~2004년 8월 : 고려대학교 BK21 정보기술사업단 박사후 연구원
 2004년 8월~현재 : 미국 카네기멜론 대학교 전기컴퓨터공학과 박사후 연구원
 주관심분야: 신호처리, 음성인식, 잡음처리

• 고한석 (Hanseok Ko)

1982년 5월 : 미국 카네기 멜론 대학교 전기공학 (공학사)
 1988년 5월 : 미국 존스 홉킨스 대학교 전기공학 (공학석사)
 1992년 5월 : 미국 카톨릭 대학교 전기공학 (공학박사)
 1995년 3월 현재 : 고려대학교 전기컴퓨터공학과 교수
 주관심분야: 영상 및 음성 신호처리, 패턴 인식