

병렬 결합된 혼합 모델 기반의 특징 보상 기술

Feature Compensation Method Based on Parallel Combined Mixture Model

김 우 일*, 이 흥 규**, 권 오 일***, 고 한 석*
(Wooil Kim*, Heungkyu Lee**, Ohil Kwon***, Hanseok Ko*)

* 고려대학교 전자컴퓨터공학과, ** 미디어젠, *** 현대오토넷
(접수일자: 2003년 6월 9일; 채택일자: 2003년 8월 28일)

본 논문에서는 잡음 환경에서 보다 강인한 성능을 얻기 위하여 음성 모델 기반의 효과적인 특징 보상 기법을 제안한다. 일반적인 모델 기반의 특징 보상 기법은 오염 음성 데이터베이스를 이용한 훈련 과정을 필요로 하므로 온라인 상에서의 적용 과정에 적합하지 않다. 제안한 방법에서는 보정 인자 추정 과정에서 병렬 모델 결합 기법을 도입함으로써 훈련 과정을 필요하지 않게 하였다. 모델의 결합 과정이 HMM 전체가 아닌 가우시안 혼합 (Mixture) 모델에만 적용이 되므로, 계산이 비교적 간단하게 되어 온라인 상에서의 모델 결합을 가능하게 하였다. 병렬적 모델 결합의 도입은 잡음 모델의 독립적인 이용을 가능하게 하였고, 본 논문에서는 MAP (Maximum A Posteriori) 적용을 통해 잡음 모델 갱신을 실시하였다. 또한 잡음 오염 과정에 대한 근사화를 통해 연속적 형태의 채널 정규화 기법을 유도하여 적용하였다. 보다 효율적인 구현을 위하여 선택적인 모델 결합 방식을 도입함으로써 연산량을 줄일 수 있는 방법을 제시하였다. 제안한 특징 보상 기법이 부가적인 배경 잡음과 채널 왜곡이 존재하는 잡음 환경에서 음성 인식 시스템의 성능을 향상시키는데 효과적임을 실험을 통해 확인할 수 있었다.

핵심용어: 음성 인식, 모델 기반 특징 보상, 병렬적 모델 결합, 모델 적응, 채널 정규화, 선택적 모델 결합

주요분야: 음성처리 분야 (2.5)

This paper proposes an effective feature compensation scheme based on speech model for achieving robust speech recognition. Conventional model-based method requires off-line training with noisy speech database and is not suitable for online adaptation. In the proposed scheme, we can relax the off-line training with noisy speech database by employing the parallel model combination technique for estimation of correction factors. Applying the model combination process over to the mixture model alone as opposed to entire HMM makes the online model combination possible. Exploiting the availability of noise model from off-line sources, we accomplish the online adaptation via MAP (Maximum A Posteriori) estimation. In addition, the online channel estimation procedure is induced within the proposed framework. For more efficient implementation, we propose a selective model combination which leads to reduction of the computational complexities. The representative experimental results indicate that the suggested algorithm is effective in realizing robust speech recognition under the combined adverse conditions of additive background noise and channel distortion.

Keywords: Speech recognition, Model-based feature compensation, Parallel model combination, Model adaptation, Channel normalization, Selective model combination

ASK subject classification: Speech signal processing (2.5)

I. 서론

음성 인식 시스템의 성능을 하락시키는 주요한 원인은 인식에 사용되는 음향 모델의 파라미터의 추정을 위해 사용되는 훈련 데이터와 실제 인식 환경에서 입력받는

음성 데이터의 음향학적 특성이 상이하다는 것이다. 음향학적 차이를 일으키는 요인으로 인식 시스템 주변의 배경 잡음에 의한 음성 신호의 오염, 마이크 및 전송 선로와 같이 음성신호의 전달과정에서 생기는 채널 왜곡, 화자 및 발음 방법의 변화, 인식 대상 (Task)의 변화 등을 들 수 있다. 이와 같은 요인에 의해 발생되는 인식 환경과 훈련 환경의 차이를 줄임으로써 인식 성능을 높이고자 하는 연구와 기술 개발이 근래 들어 집중되고 있다.

책임저자: 고한석 (hsko@korea.ac.kr)
1: 6-701 서울시 성북구 안암5가-1
고려대학교 전자컴퓨터공학과
(전화: 02-3290-3239; 팩스: 02-3291-2450)

이러한 차이를 줄이기 위한 방법은 음성 인식 처리 과정에서 어느 단계에 적용하는지에 따라 구분될 수 있다. 잡음 제거 기법과 같이 음성 입력 단계에서 적용할 수 있는 방법이 있고, 특징 추출 단계에서 잡음 혹은 변이 요소로 판단되는 부분을 제거하는 방법이 있다. 주파수 차감법 (Spectral subtraction)이나 Wiener 필터에 기반한 잡음 억압 및 음성 강화 기술, 캡스트럼 평균 정규화 (CMN, Cepstral Mean Normalization), 모델 기반의 특징 보상 (Feature compensation)과 같은 기법들이 전처리 단계에서 잡음 요소를 제거함으로써 깨끗한 환경에서 훈련된 음향 모델에 인식 환경을 일치시키고자 하는 시도이다 [1]. 인식에 사용되는 음향 모델 파라미터를 변화된 환경에 맞게 갱신시켜 주는 방법은 훈련이나 인식 단계에서 처리하는 기법이다. MAP (Maximum A Posteriori) 적응 (Adaptation)이나 MLLR (Maximum Likelihood Linear Regression) 적응 기법은 소량의 데이터로 변화된 환경에 맞게 음향 모델을 변화시키고자 하는 기술이며, 병렬적 모델 결합 (PMC, Parallel Model Combination) 기법은 깨끗한 음성 모델과 잡음 모델을 독립적으로 이용하여 잡음 환경에 맞는 오염 음성 모델을 생성해 내는 기술이다 [2-5].

본 논문에서는 전처리 단계에서 특징 보상 기술로 많이 사용되는 GMM (Gaussian Mixture Model) 기반의 특징 보상 기술에 대해 관심을 갖는다. GMM 기반의 특징 보상 기술에서는 깨끗한 음성 모델이 잡음 환경하에서 변화되는 통계적 요소를 추정하여, 오염 음성 입력에 보상해 주는 방식을 취한다 [6]. 구현이 간단하고 성능도 뛰어나지만, 인식 환경과 동일한 훈련 데이터베이스를 필요로 하는 단점을 가진다. 본 논문에서 제안하는 방법에서는 GMM 기반의 기술에서의 단점을 지적하고 이를 극복하기 위한 방법을 제안한다. 보정 인자의 추정 과정에서 PMC 기법을 도입함으로써 데이터베이스에 의한 훈련 과정이 더 이상 필요하지 않게 되었다. PMC 기법의 도입은 잡음 모델의 독립적인 이용을 가능하게 되어 잡음 모델의 적응적인 갱신을 가능하게 하였다. 또한 연속적인 채널 정규화 기법을 도입함으로써 채널 왜곡에 보다 강인한 특성을 갖도록 하였다. 효율적인 구현을 위하여 선택적인 모델 결합 방식을 도입함으로써 계산량을 줄일 수 있는 방법을 제시하였다.

본 논문은 다음과 같이 구성된다. II장에서는 대표적인 GMM 기반의 특징 보상 기법인 RATZ에 관하여 설명하고, 이 기법이 갖는 문제점을 지적한다. III장에서는 제안하는 특징 보상 기술을 단계별로 상세히 기술한다. IV장

에서 제안하는 기법의 성능 평가를 위해 실시한 실험과 그 결과에 관해 고찰하고 V장에서 결론을 맺는다.

II. 가우시안 혼합 모델 기반의 특징 보상: RATZ

음성 모델을 기반으로 하는 특징 보상 기법은 Acero에 의해 처음으로 소개되었으며, 그 후 Moreno가 RATZ (Multivariate Gaussian Based Cepstral Normalization)라는 명칭으로 데이터 유도방식의 기법을 제안한 후 유사한 방식들의 근간이 되어 왔다 [6, 7]. RATZ에서는 잡음 환경 하에서의 음성 특징 즉, 캡스트럼 (Cepstrum)의 분포를 잡음에 오염된 훈련 데이터를 통해서 모델링한 뒤, 이를 이용하여 실제 잡음 환경에서 오염된 입력 음성 특징을 보상하는 과정으로 이루어진다.

깨끗한 음성 신호의 캡스트럼 분포는 다음과 같이 K 개의 가우시안 요소로 이루어진 혼합 모델 형태로 추정할 수 있다.

$$p(x) = \sum_{k=1}^K a_k N_x(\mu_{x,k}, \Sigma_{x,k}) \quad (1)$$

RATZ에서는 식 (1)과 같이 구성된 음성 분포가 잡음 환경에서 평균과 분산이 보정 인자 (Correction factors)에 의해 이동하는 형태로 변형되는 것으로 가정한다. 이러한 가정 하에서 오염된 음성의 분포는 다음과 같이 나타낼 수 있다.

$$p(y) = \sum_{k=1}^K a_k N_y(\mu_{x,k} + r_k, \Sigma_{x,k} + R_k) \quad (2)$$

여기에서 r_k 과 R_k 은 각각 평균과 분산에 대한 보정 인자로서 테스트 환경과 동일한 잡음으로 오염된 데이터베이스로부터 훈련 과정에 의해 얻어질 수 있다. 입력된 오염된 음성 특징 벡터는 MMSE (Minimum Mean Squared Error) 기반의 예측방법에 의해 깨끗한 음성 특징으로 복구된다.

$$\begin{aligned} \hat{x}_{MMSE} &= E\{x|y\} = \int x \cdot p(x|y) dx \\ &= y - \sum_{k=1}^K r_k p[k|y] \end{aligned} \quad (3)$$

이상과 같이 설명한 RATZ 방식의 특징 보상 기술은

비교적 계산량이 간단하고 처리 속도가 빠른 장점을 가지지만 실제 환경 적용시에 다음과 같은 문제점을 갖는다.

- 1) 보정 인자 추정을 위해 잡음 오염 음성 데이터를 필요로 한다.
- 2) 보정 인자 추정에 사용한 음성 데이터의 잡음 환경과 실제 인식 환경이 일치하지 않을 경우 오히려 성능의 저하를 가져온다.
- 3) 모델 적용을 하기 위해서는 상당한 음성 샘플을 필요로 하므로 온라인 적용에 적합하지 않다

비록 HMM (Hidden Markov Model)의 훈련과 같이 많은 데이터베이스를 요구하지는 않지만 잡음 환경에 맞는 오염 음성 데이터베이스를 구축하는 작업은 매우 어려운 일이다. 또한 실제 환경에서는 시간에 따라 그 잡음 환경이 변화하므로 이에 따라 특징 보상에 필요한 파라미터를 적절하게 갱신시켜 주는 과정이 반드시 요구되는데 기존의 RATS 기술에서 음성 모델 분포를 적용하기 위해서는 "sparse-data" 문제를 극복할 만큼의 오염 음성 샘플을 필요로 하므로 온라인 적용이 쉽지 않게 된다.

본 논문에서는 HMM 모델 변환 기법의 하나인 PMC 기술을 적용하여 오염 음성 데이터베이스에 의한 훈련 과정을 없애고자 한다. PMC 기술을 도입함으로써 잡음 모델에 대한 독립적인 처리가 가능해지므로 잡음 모델의 적용에 의한 성능 향상을 피하고자 한다. PMC는 기본적으로 부가잡음만을 가정하므로 채널 왜곡과 같은 전별루션 잡음의 예측 및 제거를 통한 성능 향상을 이루고자 한다. 또한 선택적인 모델 결합 방식을 통해서 계산량을 줄일 수 있는 방법을 제시한다.

III. 제안하는 방법

3.1. 병렬적 모델 결합 기법의 도입

일반적으로 훈련 환경과 인식 환경이 동일할 때 인식 시스템이 최적의 성능을 나타내는 것으로 알려져 있으며, PMC는 이러한 가정 하에 깨끗한 음성 HMM 모델을 인식 환경과 일치하는 잡음에 오염된 음성 모델로 변환하는 기술이다[5]. 오염된 음성 모델을 생성을 위해서는 깨끗한 음성 모델과 잡음 모델이 독립적으로 이용되어 결합된다. 잡음 오염 음성 데이터베이스에 의한 훈련 과정 없이 오염 음성 모델을 추정한다는 측면에서 상당한 장점을 갖는다.

음성 모델과 잡음 모델의 결합은 다음과 같은 Mismatch 함수에 의해 이루어지게 된다.

$$Y'_i(\tau) = F(X'_i(\tau), N'_i(\tau)) = \log(g \exp(X'_i(\tau)) + \exp(N'_i(\tau))) \quad (4)$$

식 (4)에서 X'_i , N'_i , Y'_i 는 각각 로그 스펙트럼 상에서의 깨끗한 음성, 잡음, 오염된 음성 신호를 나타낸다. 로그 스펙트럼 상에서 가우시안 분포를 갖는 랜덤 변수는 지수 (Exponential) 변환에 의해 선형 스펙트럼 상에서 로그-정규 (Log-normal) 분포를 갖게 된다. PMC의 모델 결합 기법 중의 하나인 로그-정규 가정법에서는 두개의 로그-정규 분포의 합도 같은 로그-정규 분포를 따르게 되는 것을 가정한다. 이러한 가정 하에 선형 스펙트럼 상에서의 오염된 음성 모델의 평균과 공분산은 다음과 같은 수식에 의하여 계산할 수 있다.

$$\hat{\mu} = g\mu + \tilde{\mu} \\ \hat{\Sigma} = g^2 \Sigma + \tilde{\Sigma} \quad (5)$$

식 (5)에서 μ , $\tilde{\mu}$, $\hat{\mu}$ 는 각각 깨끗한 음성, 잡음, 오염된 음성 모델의 평균을 말하며, Σ , $\tilde{\Sigma}$, $\hat{\Sigma}$ 는 그들의 공분산을 의미한다. 로그-정규 분포를 갖는 선형 스펙트럼 상에서의 평균과 공분산은 다음 식에 의하여 로그 스펙트럼의 모델 파라미터로부터 구한다.

$$\mu_i = \exp(\mu'_i + \sum_{ii} / 2) \\ \Sigma_{ij} = \mu_i \mu_j [\exp(\sum'_{ij}) - 1] \quad (6)$$

식 (5)에 의해 구해지는 선형 스펙트럼 상에서의 오염된 음성의 평균과 공분산은 다시 로그 스펙트럼 상으로 변환한다. 다음의 식에 의하여 선형 스펙트럼의 모델 파라미터로부터 로그 스펙트럼 상에서의 평균과 공분산을 구할 수 있다.

$$\hat{\mu}'_i \approx \log(\hat{\mu}_i) - \frac{1}{2} \log \left(\frac{\hat{\Sigma}_{ii}}{\hat{\mu}_i^2} + 1 \right) \\ \hat{\Sigma}'_{ij} \approx \log \left(\frac{\hat{\Sigma}_{ij}}{\hat{\mu}_i \hat{\mu}_j} + 1 \right) \quad (7)$$

3.2. 병렬 결합된 혼합 모델 기반의 특징 보상

RATS에서의 같이 음성 분포 전체를 나타내는 혼합 모델 형태의 가우시안 분포 기반의 특징 보상 기술을 근간으로 하지만, 잡음 환경 하에서 모델 변화를 나타내는 보정 인자를 추정하기 위해 오염 음성 데이터베이스를 이용하여 훈련하는 것이 아니라 앞에서 소개한 PMC 기법을 적용하여 깨끗한 음성 모델과 잡음 모델을 이용하여 오염

된 음성의 분포를 얻게 된다. 캡스트럼 상에서의 평균과 공분산은 다음의 식과 같이 역 (Inverse) 이산 코사인 변환 (DCT, Discrete Cosine Transform)에 의해 로그 스펙트럼 상으로 변환할 수 있다.

$$\begin{aligned} \mu' &= C^{-1}\mu^c \\ \Sigma' &= C^{-1}\Sigma^c(C^{-1})^T \end{aligned} \quad (8)$$

로그 스펙트럼 상으로 변환된 모델 파라미터는 PMC 과정에 의해 잡음 모델과 결합되어 로그 스펙트럼 상에서의 잡음 오염 음성 모델을 생성하게 되고 이는 다시 식 (8)의 반대 과정인 DCT 변환에 의해 캡스트럼 상의 평균과 공분산으로 변환하게 된다. RATZ에서 가정했던 잡음 환경에서의 평균과 분산의 이동을 근거로 하여 깨끗한 음성 분포와 오염 음성 분포를 이용하여 보정 인자를 다음과 같이 계산한다.

$$\begin{aligned} r &= \mu_y - \mu_x \\ R &= \Sigma_y - \Sigma_x \end{aligned} \quad (9)$$

깨끗한 특징으로의 복구는 RATZ에서와 같이 MMSE 예측법을 그대로 이용한다. 이상과 같이 본 논문에서 제안하는 병렬 결합된 혼합 모델 (PCMM, Parallel Combined Mixture Model) 기반의 특징 보상 기술은 다음과 같은 장점을 갖는다.

- 1) 보정 인자 추정을 위해 오염 음성 데이터베이스에 의한 훈련 과정이 불필요하다. 잡음 데이터베이스를 이용하여 잡음 모델을 취득하게 되면 병렬적 결합 과정에 의해 가우시안 혼합 모델로 구성된 오염된 음성의 분포를 얻을 수 있다.
- 2) PMC 기술에서와 같이 모든 HMM에 대해 모델의 결합 과정이 이루어지는 것이 아니고 음성 분포를 나타내는 가우시안 혼합 모델에 대해서 이루어진다. 또한 특징 보상과정에서는 정적 (Static) 파라미터에 대해서만 특징 보상이 이루어지게 되면, 이로부터 1차, 2차 미분 계수인 동적 (Dynamic) 특징 벡터의 보상된 결과를 얻을 수 있다. 따라서 정적 파라미터에 대한 모델 결합만 필요하므로 계산량의 대폭적인 감소를 얻을 수 있다. 이러한 계산량의 감소에 따라 음성 모델과 잡음 모델의 결합 과정이 온라인 상에서도 이루어질 수 있으며 이는 잡음 모델의 적응적인 갱신에 따른 오염 음성 모델의 온라인 적응을 가능하게 한다.
- 3) 잡음 모델을 독립적으로 이용할 수 있으므로 음성 모델이 아닌 잡음 모델에 대한 적응 (Adaptation) 과정이

가능하다. 비 음성 구간에서 잡음에 대한 적응을 통해 새로운 잡음 모델을 추정할 수 있으며, 잡음 모델은 단일 가우시안과 같이 비교적 적은 수의 파라미터로 구성되므로 비 음성 구간과 같이 적은 수의 샘플로도 효과적인 모델 추정을 할 수 있다. 이와 같이 잡음 모델만의 갱신을 바탕으로 병렬적 모델 결합 과정을 통한 오염 음성 모델의 적응은 짧은 음성 구간에 대해서도 적응에 필요한 "sparse-data" 문제를 해결할 수 있게 한다.

3.3. 비 음성 구간에서의 잡음 모델 적응

잡음 모델의 적응은 잡음 신호만이 존재한다고 가정되는 비음성 구간에서 MAP (Maximum A Posteriori) 예측 기법에 의한 적응을 적용한다[2,3]. 일반적으로 잡음 모델은 단일 가우시안 분포로 모델링되므로 짧은 비음성 구간에서도 MAP 예측에 의한 적응의 효과가 있을 것으로 판단된다.

MAP 예측은 일반적으로 모델 예측에 사용되는 최대 우도 (ML, Maximum Likelihood) 예측 기법에 구하고자 하는 파라미터에 대한 사전 확률 (Priori) 정보를 이용하여 사후 확률 (Posteriori)을 구성하여 이를 최대로 하는 파라미터를 예측하는 과정으로 이루어진다.

$$\begin{aligned} \theta_{MAP} &= \arg \max_{\theta} f(\theta | \mathbf{x}) \\ &= \arg \max_{\theta} f(\theta) f(\mathbf{x} | \theta) \end{aligned} \quad (10)$$

관찰값에만 의존하게 되는 ML 예측에서 관찰값이 충분하지 못할 경우, 관찰값의 전체 영역을 신뢰성 있게 반영하지 못하는 파라미터를 추정하게 된다. 이러한 오류를 막기 위하여 파라미터에 대한 사전 확률을 적용함으로써 보다 안정적인 범위에서 파라미터를 추정하는 효과를 갖는다. 반복적인 EM (Expectation Maximization) 알고리즘을 적용하여 ML 예측을 통해 불완전한 (Incomplete) 훈련 데이터로부터 모델 파라미터를 추정하는 방식에 모델에 대한 사전 확률을 결합한 MAP 예측법이 알려져 있으며 단일 가우시안 분포의 평균에 대한 추정식은 다음과 같이 간략화할 수 있다.

$$\hat{\mu} = \frac{\tau\mu + 1 \sum_{t=1}^T x_t}{\tau + 1} \quad (11)$$

즉, 사전 정보로 이용되는 초기 잡음 모델의 평균과 입력된 잡음 구간의 평균이 적응율을 결정하는 값에 의해 적절한 비율로 합쳐져서 새롭게 적응된 잡음의 평균으로 사용된다.

3.4. 채널 왜곡의 예측 및 정규화

오염 음성 모델 생성에 사용되는 PMC 기술은 일반적으로 부가적인 잡음만을 가정하게 된다. 따라서 채널 왜곡과 같은 컨벌루션에 의한 잡음을 제거해 줄 필요가 있다. 본 논문에서는 음성의 오염과정을 근사화한 수식을 가정하고 이로부터 연속적 (Sequential) 형태의 EM 알고리즘을 통해 실시간 온라인 채널 왜곡 예측을 수행한다. 부가 잡음에 비해 채널 왜곡의 영향이 덜하다는 가정 하에 음성 신호의 오염 과정은 다음과 같이 근사화할 수 있다.

$$\begin{aligned} \mathbf{y} &= \mathbf{x} + \mathbf{h} + \mathbf{C} \log(\mathbf{1} + \exp(\mathbf{C}^{-1}(\mathbf{n} - \mathbf{x} - \mathbf{h}))) \\ &\approx \mathbf{x} + \mathbf{h} + \mathbf{C} \log(\mathbf{1} + \exp(\mathbf{C}^{-1}(\mathbf{n} - \mathbf{x}))) \\ &= \mathbf{y}_n + \mathbf{h} \end{aligned} \quad (12)$$

식 (12)과 같이 근사화를 통해 배경 잡음 \mathbf{n} 에 오염된 특징 벡터 \mathbf{y}_n 에 채널 왜곡 벡터 \mathbf{h} 가 부가적인 바이어스 요소로 작용하는 것으로 가정하면 최종적인 오염 음성 \mathbf{y} 는 다음과 같은 분포를 따른다.

$$\mathbf{y} \sim N(E[\mathbf{y}_n] + \mathbf{h}, Cov[\mathbf{y}_n]) \quad (13)$$

식 (13)에서 배경 잡음에 오염된 특징 벡터 \mathbf{y}_n 에 대한 확률 분포 파라미터는 앞에서 제안한 PCMM 과정 즉, 깨끗한 음성 분포와 잡음 모델에 대한 병렬 결합 과정을 적용하여 얻을 수 있다. 음성 모델은 K 개의 가우시안의 혼합 형태로 모델링되며 입력되는 오염 음성 특징 벡터 \mathbf{y} 는 각각의 가우시안 요소에 대해 불완전한 성질을 가지므로 EM 알고리즘을 통해 채널 왜곡 벡터 \mathbf{h} 를 예측한다. EM 알고리즘 전개를 위한 보조 (Auxiliary) 함수는 다음과 같다.

$$\begin{aligned} \zeta(\mathbf{h}, \bar{\mathbf{h}}) &= E[L(\mathbf{Y}, \mathbf{S}) | \bar{\mathbf{h}}] | \mathbf{Y}, \mathbf{h} \\ &= \sum_{t=1}^T \sum_{k=1}^K \frac{P(\mathbf{y}_t, k | \mathbf{h})}{p(\mathbf{y}_t | \mathbf{h})} \log(p(\mathbf{y}_t, k | \bar{\mathbf{h}})) \end{aligned} \quad (14)$$

식 (14)에서 (\mathbf{Y}, \mathbf{S}) 는 관찰 벡터가 어떤 가우시안 요소에서 발생했는지를 알 수 있는 완전한 (Complete) 데이터이다. 보조 함수를 최대화 하는 $\bar{\mathbf{h}}$ 를 찾기 위해 $\nabla_{\bar{\mathbf{h}}} \zeta(\mathbf{h}, \bar{\mathbf{h}}) = 0$ 인 $\bar{\mathbf{h}}$ 에 대해 정리하면 최종적인 예측은 다음과 같다.

$$\bar{\mathbf{h}} = \frac{\sum_{t=1}^T \sum_{k=1}^K P(k | \mathbf{y}_t, \mathbf{h}) \sum_{y_{t,k}}^{-1} (\mathbf{y}_t - \boldsymbol{\mu}_{y_{t,k}})}{\sum_{t=1}^T \sum_{k=1}^K P(k | \mathbf{y}_t, \mathbf{h}) \sum_{y_{t,k}}^{-1}} \quad (15)$$

본 논문에서는 실시간으로 채널 예측을 위해 다음과

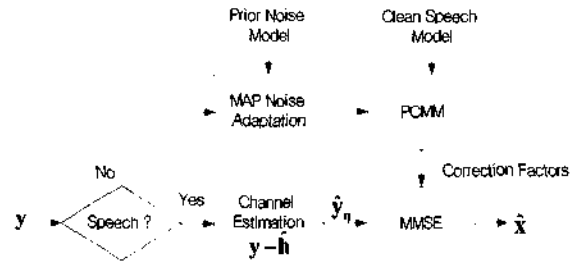


그림 1. 제안한 특징 보상 기법의 블록선도
Fig. 1. Block diagram of the proposed scheme.

같이 연속적 형태로 구현하였다.

$$\hat{\mathbf{h}} = \frac{\sum_{t=1}^{t-1} \sum_{k=1}^K P(k | \mathbf{y}_t, \hat{\mathbf{h}}_{t-1}) \sum_{y_{t,k}}^{-1} (\mathbf{y}_t - \boldsymbol{\mu}_{y_{t,k}})}{\sum_{t=1}^{t-1} \sum_{k=1}^K P(k | \mathbf{y}_t, \hat{\mathbf{h}}_{t-1}) \sum_{y_{t,k}}^{-1}} \quad (16)$$

그림 1은 본 논문에서 제안하는 PCMM 기반의 특징 보상 기법의 전체적인 처리 과정을 나타낸다.

3.5. 선택적 모델 결합에 의한 연산 속도 향상

앞 절에서 설명한 바와 같이, 본 논문에서 제안하는 PCMM 기반의 특징 보상 기법은 잡음 모델의 갱신에 따라 적응적인 잡음 오염 음성 모델의 생성을 가능하게 한다. 모델의 결합 과정이 혼합 모델에 대해서만 이루어지지만 병렬 모델 결합 (PMC) 과정에 따라 선행 스펙트럼 상에서의 모델 결합 후, 로그 스펙트럼을 거쳐 캡스트럼 상으로 변환해 주어야 하는 복잡한 연산을 행하여야 한다. PMC 기법이 등장한 뒤, 이러한 계산의 복잡성을 줄이기 위한 다양한 방법들이 소개되었으며[8,9], 본 논문에서는 가우시안 요소를 선택적으로 모델 결합에 참여시킴으로써 근본적으로 계산량을 단축하고자 한다.

선행 스펙트럼 상에서 부가적인 관계에 있는 잡음에 오염된 음성 신호의 로그 스펙트럼 상에서의 분포는 로그 스펙트럼 상에서의 깨끗한 음성 모델과 잡음 모델의 평균의 거리 즉, SNR과 분산의 비율에 크게 의존한다는 사실이 이미 논문을 통해 알려져 있다[10]. 본 논문에서는 이와 같은 사실에 입각하여 로그 스펙트럼 상에서 음성 모델과 잡음 모델 사이의 평균의 차와 분산의 비율이 일정 문턱값보다 크게 변하는 가우시안 요소에 대해서만 선택적으로 병렬 모델 결합을 취함으로써 잡음에 강인한 성능을 유지하면서 계산량을 단축시키는 방안을 제시한다.

식 (17)과 같이 로그 스펙트럼 상에서의 음성 모델과 잡음 모델의 평균의 거리와 분산의 비율로 모델 결합에 참여할 가우시안 요소를 선택하는 기준을 정의하였다.

$$d_{mean} = (\mu'_x - \mu'_y)^T (\mu'_x - \mu'_y), d_{var} = \sum_i \frac{\sum_{ii}}{\sum_{ii}} \quad (17)$$

식 (18)은 선택 조건을 나타내며, $d_{mean}^{(m)}$ 과 $d_{var}^{(m)}$ 은 적응 과정에 의해 갱신된 새로운 잡음 모델과의 거리를 말하며, $d_{mean}^{(m-1)}$ 과 $d_{var}^{(m-1)}$ 는 이전 단계에서의 잡음 모델과의 거리를 말한다. α 와 β 의 문턱값에 따라 식 (17)과 같이 정의한 기준값의 변화가 큰 가우시안 요소를 모델 결합에 참여시키도록 선택하였다.

$$\left| \frac{d_{mean}^{(m)}}{d_{mean}^{(m-1)}} - 1.0 \right| > \alpha, \left| d_{var}^{(m)} - d_{var}^{(m-1)} \right| > \beta \quad (18)$$

IV. 실험 및 결과 고찰

4.1. 실험 환경 및 베이스라인 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[11].

- 1) 영어 음성, 연속 숫자음 인식, 11단어+묵음 구간 (Silence)+짧은 휴지 (Short pause)
- 2) ETSI (European Telecommunications Standards Institute) 표준의 DSR (Distributed Speech Recognition) 방식의 특징 추출[12]
- 3) 13차 정적 특징 (c1~c12+로그 에너지) 추출 후 인식 단에서 미분계수 추출 (총 39차): 본 논문의 실험에서는 병렬적인 모델 결합 과정 (PMC) 구현의 편의를 위하여 로그 에너지 대신 캡스트럼의 0차 계수를 사용하였다.
- 4) 3-mixture, 16-state의 단어 모델, 2종류의 묵음 모델

Aurora 2.0에서 제공하는 Clean-condition Training, Multi-condition Testing 방식에 따라 음향 모델은 깨끗한 환경에서 수집된 8,840개의 음성 데이터를 이용하여 훈련하였으며, 인식 성능은 set A의 자동차 주행 잡음 (Car noise) 환경과 채널 왜곡이 반영된 set C의 지하철 잡음 (Subway, MIRS)에 대하여 실시하였다. 각 테스트 데이터는 7가지의 신호 대 잡음비 (SNR)에 따라 각각 1,001개의 샘플로 구성되어 있다.

비교 평가를 위해 GMM 기반의 특징 보상 기법의 하나인 RATZ에 대한 성능 평가를 실시하였고, 잡음 제거 기법 중 가장 일반적인 형태인 주파수 차감법 (SS)과 채널 왜곡 요소 제거를 위한 캡스트럼 정규화 (CMN) 기법을 적용하

여 기본적인 성능을 확인하였다. RATZ와 PCMM에서 사용하게 되는 깨끗한 음성 분포는 HMM의 훈련과 동일한 음성 데이터를 이용하여 128개의 가우시안 요소로 이루어진 혼합 모델로 구성하였다. RATZ 적용을 위한 보정 인자의 추정은 Aurora 2.0에 포함되어 있는 Multi-condition Training 데이터 중에서 각 SNR 환경과 일치하는 데이터를 이용하여 non-stereo 방식의 EM 기법을 적용하여 추정하였다. 0 dB와 -5 dB의 경우 Aurora 2.0 내에 이에 해당하는 잡음 음성 훈련 데이터가 없기 때문에 이와 일치하는 환경에 해당하는 RATZ 결과를 얻을 수 없었다. 주파수 차감법은 차감 계수와 최소 계수 (flooring)를 각각 4.0과 0.2로 두었으며, 약 250 msec의 시간 지연을 이용하여 최소 통계 (Minimum Statistics)에 기반을 둔 배경 잡음 예측 방법을 적용하였다[13].

표 1은 자동차 잡음에 대한 베이스라인 및 기존 기술들에 대한 성능 평가를 나타낸다. 실험 결과에서 알 수 있듯이 인식 환경과 동일한 잡음 상태의 훈련 데이터를 이용한 RATZ의 경우 주파수 차감법과 CMN의 단독 혹은 조합된 적용보다 월등한 성능을 나타내었다. 이는 GMM으로 표현되는 음성 분포 모델에 기반한 특징 보상 기법이 잡음 환경에서 뛰어난 성능을 나타내는 것을 입증하는 결과이다. 표 2의 실험에 사용된 지하철 잡음 환경 데이터는 지하철 잡음이 부가적인 형태로 오염되고, 이동 통신 환경을 시뮬레이션한 필터를 사용하여 낮은 주파수 영역이 감쇄되는 채널 왜곡을 반영한 테스트 데이터베이스이다. Aurora 2.0의 Multi-condition Training 데이터는 이러한 채널 왜곡이 적용되지 않고 오직 지하철 잡음만이 부가적으로 오염된 훈련 데이터가 제공된다. 표 2에서 RATZ의 전체적인 성능이 낮고, 높은 SNR 환경에서 오히려 베이스라인에 비해 하락하는 현상은 채널 왜곡이 적용되지 않은 훈련 데이터로 보정 인자를 추정하여 적용함

표 1. Aurora2.0의 자동차 잡음에 대한 베이스라인의 단어 인식을 (%)

Table 1. Word accuracy for baseline system to car noise condition in Aurora2.0 (%).

	Baseline	RATZ	SS	CMN	SS-CMN
Clean	98.84	98.84	98.63	98.87	98.81
20 dB	96.42	97.79	97.38	97.44	97.58
15 dB	87.62	96.33	93.98	92.01	95.53
10 dB	61.71	91.95	81.42	71.43	87.89
5 dB	26.87	82.17	50.16	34.12	64.75
0 dB	10.38		17.66	19.62	28.36
-5 dB	8.41		5.99	11.66	14.76
Average	56.60		68.12	62.92	74.82

표 2. Aurora2.0의 지하철 (MIRS) 잡음에 대한 베이스라인의 단어 인식률 (%)

Table 2. Word accuracy for baseline system to Subway (MIRS) condition in Aurora2.0 (%).

	Baseline	RATZ	SS	CMN	SS-CMN
Clean	99.08	99.08	98.74	98.93	98.74
20 dB	95.86	94.32	96.62	96.22	96.41
15 dB	91.31	89.50	94.44	90.42	94.17
10 dB	78.29	79.18	87.17	72.46	86.92
5 dB	43.63	62.79	63.83	39.02	66.38
0 dB	16.70		24.38	21.37	33.13
-5 dB	8.93		10.16	12.65	15.04
Average	65.16		73.29	63.90	75.40

로써, 훈련 환경과 인식 환경의 불일치로 인해 나타나는 결과이다. CMN을 단독 적용할 경우에도 베이스라인보다 낮은 성능을 나타내었는데, 이것은 배경 잡음 제거없이 채널 정규화만을 시도할 때 오히려 음성 왜곡이 증가하여 인식 성능이 하락하는 것으로 판단할 수 있다.

4.2. 제안한 PCMM 특징 보상의 성능 평가

본 논문에서 제안하는 PCMM 기반의 특징 보상 기술에 대한 성능 평가를 베이스라인의 성능평가와 동일한 환경에서 실시하였다. RATZ의 훈련에 사용한 Aurora2.0의 Multi-condition Training 샘플들의 시작 부분 100 msec을 묵음구간으로 가정하고 이로부터 잡음 샘플을 추출하여 각 SNR에 맞는 잡음 모델을 생성하였다. 잡음 모델은 단일 가우시안 분포로 추정하였으며 0 dB와 -5 dB의 경우 이에 해당하는 훈련 데이터베이스가 없기 때문에 부득이하게 테스트에 사용하는 음성 샘플에서 같은 방법으로 묵음 구간을 추출하여 잡음 모델 생성에 사용하였다. 구현한 각 기능들에 대한 객관적 비교를 위해 다음과 같이 기술별 조합에 의한 성능 평가를 실시하였다.

- 1) PCMM: 훈련으로 얻은 잡음 모델을 이용한 PCMM 기반의 특징 보상
- 2) PCMM-NA: 테스트 음성 입력 시작 부분 10프레임(약 100 msec)에서 잡음 모델에 대한 MAP 적용 후 PCMM 기반의 특징 보상
- 3) PCMM-CN: 매 프레임마다 실시간 채널 예측 및 정규화.
- 4) PCMM-NA-CN: 잡음 모델 적용 및 채널 정규화

표 3에서 알 수 있듯이 본 논문에서 제안하는 PCMM 방식의 특징 보상 기술 적용 결과, 잡음 환경에서 성능이 향상되는 것을 확인할 수 있었으며 이 수치는 베이스라인 성능 평가에서 살펴본 주파수 차감법보다도 높은 인식 성능이다. 이것은 모델의 결합 과정이 부가 잡음의 오염

표 3. Aurora2.0의 자동차 잡음에 대한 제안한 기법의 단어 인식률 (%)

Table 3. Word accuracy for the proposed scheme to car noise condition in Aurora2.0 (%).

	Baseline	PCMM	PCMM-NA	PCMM-CN	PCMM-NA-CN
Clean	98.84	98.84	98.81	98.66	98.72
20 dB	96.42	97.61	97.97	98.21	98.30
15 dB	87.62	96.57	96.75	97.29	97.41
10 dB	61.71	91.32	92.81	93.59	94.36
5 dB	26.87	77.51	80.79	79.66	82.22
0 dB	10.38	47.84	51.57	49.39	53.36
-5 dB	8.41	18.73	20.25	18.16	21.12
Average	56.60	82.17	83.98	83.63	85.13

과정을 잘 표현하여 잡음 오염 음성 모델의 생성에 효과적이었음을 말해주는 결과이다. 잡음 모델의 적용을 통해 평균 1.81%의 인식 성능 향상을 보였으며 이것은 MAP 방식의 예측 방법이 입력 때마다 달라지는 잡음 환경의 적용에 효과적임을 보여주는 것이다. 본 평가에 사용한 잡음 환경이 훈련 및 테스트 환경에서 그 변화 정도가 작기 때문에 성능 향상이 크지 않았던 것으로 보인다. 채널 정규화를 통해 PCMM만을 단독으로 사용한 경우보다 평균 1.46%의 성능 향상을 확인했다. 평가에 사용한 자동차 잡음 환경의 데이터베이스에는 채널 왜곡이 반영되지 않았지만, 컨벌루션 잡음 요소를 예측하고 제거하는 과정은 화자간의 변이요소를 정규화하는 효과를 가지는 것으로 해석할 수 있다. 잡음 모델에 대한 적용과 채널 예측을 동시에 적용함으로써 각 기술들을 단독으로 적용했을 때 뿐 아니라, 표 1의 기존의 기술보다 높은 성능을 나타내었으며 이것은 제안하는 특징 보상 기술이 부가적인 배경 잡음과 채널 왜곡과 같은 컨벌루션 잡음에 강인한 성능을 가지는 것을 입증하는 결과이다.

표 4는 채널 왜곡이 적용된 지하철 잡음 환경에 대해 PCMM 기반 특징 보상 기술의 성능 평가를 나타낸 결과이다. 표 2의 RATZ 실험에서와 마찬가지로 채널 왜곡이 적용된 훈련 데이터가 제공되지 않으므로, 부가적인 요소로만 작용하는 지하철 잡음 모델을 초기 모델로 하여 잡음 모델의 적용 효과를 관찰하고 실시간 채널 예측에 의한 성능 향상을 확인하였다. 잡음 모델의 적용 결과 초기 모델을 그대로 사용한 경우보다 3.90%의 상당한 성능 향상을 가져왔으며 이것은 적용 과정에서 배경 잡음 요소 뿐 아니라 채널 왜곡 요소 등이 잡음 모델에 효과적으로 반영되는 것을 의미한다. 채널 왜곡이 반영된 테스트 샘플이므로 채널 예측 및 제거를 통해 급격한 성능 향상을

나타내었으며 표 3의 결과와 일관성이 있는 경향을 확인할 수 있었다.

4.3. 선택적 모델 결합에 의한 연산 속도 향상

본 논문에서 제안한 선택적 모델 결합에 의한 연산 속도의 향상을 실험을 통해 확인하였다. 표 3의 실험과 같은 자동차 잡음 환경에 대해 실시하였으며 PCMM-NA의 실험에서 잡음 모델의 적응 후 오염 음성 모델의 생성을 위한 모델 결합시에 논문에서 제안한 기준을 이용하여 모델 결합에 참여시킬 가우시안 요소를 선택하였다.

표 5는 문턱값 α 와 β 를 각각 0.1과 3.0으로 두었을 때의 실험 결과이다. 첫째 열은 선택적 모델 결합에 의한 단어 인식률이며 둘째 열은 PCMM-NA 즉, 모든 가우시안 요소를 모델 결합에 참여시켰을 때 단독 PCMM 기법에 대한 인식률 향상을 100.0이라 두었을 때 선택적 모델 결합에 의한 인식률 향상의 상대적인 비율이며 다음과 같이 구할 수 있다.

표 4. Aurora2.0의 지하철 (MIRS) 잡음에 대한 제안한 기법의 단어 인식률 (%)
Table 4. Word accuracy for the proposed scheme to Subway (MIRS) condition in Aurora2.0 (%).

	Baseline	PCMM	PCMM-NA	PCMM-CN	PCMM-NA-CN
Clean	99.08	99.08	99.11	98.93	98.89
20 dB	95.86	96.44	97.33	97.91	97.91
15 dB	91.31	93.58	94.84	96.13	96.16
10 dB	78.29	85.94	89.04	92.23	92.26
5 dB	43.63	63.62	71.51	84.71	84.92
0 dB	16.70	31.96	38.32	63.89	65.34
-5 dB	8.93	14.34	16.52	31.93	34.76
Average	65.16	74.31	78.21	86.97	87.32

표 5. $\alpha=0.1, \beta=3.0$ 일 때 선택적 PCMM 기법의 성능 평가
Table 5. Performance evaluation of selective PCMM with $\alpha=0.1, \beta=3.0$.

	Word Accuracy (%)	Relative Improve. (%)	Updated Mixture Number	Updated Mixture Rate (%)
Clean	98.81	100.00	99.45	77.70
20 dB	97.94	91.87	86.16	67.31
15 dB	96.84	150.00	90.03	70.34
10 dB	92.69	91.95	85.97	67.16
5 dB	79.84	71.04	82.14	64.17
0 dB	49.57	46.38	79.54	62.14
-5 dB	19.45	47.37	74.75	58.40
Average	83.38	90.21	84.77	66.23

표 6. $\alpha=0.125, \beta=3.5$ 일 때 선택적 PCMM 기법의 성능 평가
Table 6. Performance evaluation of selective PCMM with $\alpha=0.125, \beta=3.5$.

	Word Accuracy (%)	Relative Improve. (%)	Updated Mixture Number	Updated Mixture Rate (%)
Clean	98.81	100.00	87.92	68.69
20 dB	97.85	66.67	70.85	55.35
15 dB	96.78	116.67	72.37	56.54
10 dB	92.57	83.89	68.73	53.70
5 dB	79.27	53.66	65.89	51.48
0 dB	48.76	24.86	63.28	49.44
-5 dB	19.50	50.66	59.16	46.22
Average	83.05	69.11	68.22	53.30

$$\frac{\text{Selective PCMM} - \text{PCMM}}{\text{PCMM_NA} - \text{PCMM}} \times 100 \quad (19)$$

셋째 열은 선택 조건에 의해 모델 결합에 참여한 가우시안 요소의 개수이고, 넷째 열은 전체 가우시안 요소 128개에 대한 선택된 요소의 비율을 말한다. 이와 같은 수치를 통하여 선택적 기법에 의한 계산량의 단축이 특징 보상 성능의 하락에 어떠한 영향을 미치는지를 관찰할 수 있다. 예를 들어 10 dB의 경우 선택적 기법을 사용하여 92.69%의 인식률을 보였으며 이는 모든 잡음 요소를 결합하여 적용한 PCMM-NA의 결과인 92.81% (표 3)에 비하여 91.95%의 상대적인 인식 향상 성능을 나타내었다. 이러한 인식 성능은 전체 128개 중 평균 85.97개의 가우시안 요소 즉, 67.16%를 모델 결합에 참여시켜 얻은 결과이다. 이와 같은 결과는 비록 8.05% (=100-91.95)의 상대적인 성능 하락은 있었지만, 32.84% (=100.0-67.16)의 계산량 감소를 얻을 수 있었음을 의미하는 것이다. 평균적으로 전체 계산량의 감소에 비해 성능 하락은 크지 않았으며, 특히 5 dB 이상의 잡음 환경에서 선택적 기법을 사용하여 의미있는 결과를 얻을 수 있었다. 표 6은 문턱값을 0.125와 0.35로 둬으로써 계산량을 보다 크게 줄였을 때 상대적인 성능 하락 현상을 관찰하기 위한 실험으로 계산량의 감소에 비해 성능 하락이 크지 않음을 확인할 수 있었다.

본 실험을 통해 논문에서 제안한 기준에 의한 선택적인 모델 결합이 비교적 성능을 떨어뜨리지 않으면서도 연산량을 감소시키는데 도움이 되는 것을 확인하였으며, 특히 높은 SNR 환경에서는 선택적 기법을 적용하는 것이 시스템 구현에 유리한 것으로 판단할 수 있었다.

V. 결론

본 논문에서는 잡음 환경에 강인한 음성 인식 시스템을

위한 전처리 방법으로 음성 모델에 기반한 효과적인 특징 보상 기법을 제안하였다. 병렬적 모델 결합 기법을 이용하여 깨끗한 음성 모델과 잡음 모델로부터 오염 음성 모델을 생성함으로써 잡음 오염 음성 데이터베이스에 의한 훈련 과정을 덜 수 있었다. 모델의 결합 과정이 HMM 전체가 아닌 가우시안 혼합 모델에 대해서 이루어지므로 계산량이 비교적 작아지게 되고 온라인 상에서의 모델 생성이 가능해졌다. 이것은 묵음 구간에서의 잡음 모델에 대한 적용 후 오염 음성 모델의 갱신이 각각의 발음에 대해 실시간 처리가 가능하므로 환경 변화에 보다 적응적으로 대처할 수 있음을 의미한다. 채널 왜곡에 강인한 특징 보상을 위해 EM 알고리즘에 기반을 둔 연속적 형태의 채널 정규화 과정을 도입하였다. 효율적 구현을 위해 모델 거리의 변화율을 고려한 선택적인 모델 결합 방식을 도입함으로써 계산량을 줄이는 방법을 제안하였다. Aurora2.0의 평가 방식을 통하여 제안한 특징 보상 기법이 부가적인 잡음과 채널 왜곡 환경에서 음성 인식 성능 향상에 효과적임을 확인하였다.

감사의 글

본 논문은 한국과학재단에서 지원한 국제공동연구 (과제번호: 20006-302-04-2) 사업을 통해 수행된 연구 결과의 일부입니다.

참고 문헌

1. X. Huang, A. Acero and H. Hon, *Spoken Language Processing*, Prentice Hall PTR, 2001.
2. C. H. Lee, C. H. Lin, and B. H. Juang, "A Study on speaker adaptation of the parameters of continuous density hidden markov models," *IEEE Trans. on Signal Processing*, 39 (4), 806-814, April 1991.
3. J. L. Gauvain and C. H. Lee, "Maximum a posteriori estimation for multivariate gaussian mixture observations of markov chains," *IEEE Trans. on Speech and Audio Processing*, 2 (2), 291-298, April 1994.
4. C. J. Leggetter and P. C. Woodland, "Maximum likelihood linear regression for speaker adaptation of continuous density HMMs," *Computer Speech and Language*, 9, 171-185, 1995.
5. 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.
6. 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.
7. P. J. Moreno, *Speech Recognition in Noisy Environments*, PhD Thesis, Carnegie Mellon University, 1996.

8. M. J. F. Gales, and S. J. Young, "A fast and flexible implementation of parallel model combination," *Proc. of ICASSP-95*, 133-136, 1995.
9. R. Sarikaya and J. H. L. Hansen, "PCA-PMC: a novel use of a priori knowledge for fast parallel model combination," *Proc. of ICASSP-2000*, 1113-1116, 2000.
10. P. J. Moreno, B. Raj, E. Gouvea and R. M. Stern, "Multivariate-gaussian-based cepstral normalization for robust speech recognition," *Proc. of ICASSP-95*, 137-140, 1995.
11. H. G. Hirsch, and D. Pearce, "The AURORA experimental framework for the performance evaluations of speech recognition systems under noisy conditions," *ISCA ITRW ASR2000*, 2000.
12. 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.
13. R. Martin, "Spectral subtraction based on minimum statistics," *EUSIPCO-94*, 1182-1185, 1994.

저자 약력

● 김 우 일 (Wooil Kim)



1996년 2월: 고려대학교 전자공학과 (공학사)
 1998년 8월: 고려대학교 전자공학과 (공학석사)
 2003년 8월: 고려대학교 전자공학과 (공학박사)
 2003년 9월~현재: 고려대학교 BK21 정보기술사업단 연구교원 (박사후 과정)
 ※ 주관심분야: 신호처리, 음성인식, 잡음처리

● 이 흥 규 (Heungkyu Lee)



1995년 2월: 서경대학교 컴퓨터과학과 학사
 1999년 2월: 서경대학교 컴퓨터과학과 석사
 1999년 3월~현재: 고려대학교 영상정보처리학과 박사과정
 2001년 2월~2002년3월: (주)인티 망관리연구소 책임연구원
 2002년 3월~현재: 미디어젠 음성사업본부 책임연구원
 ※ 주관심분야: 영상 및 음성 신호처리, 패턴 인식, 데이터 융합

● 권 오 일 (Ohil Kwon)



1991년 2월: 고려대학교 전자공학과 학사
 1993년 2월: 고려대학교 전자공학과 석사
 1996년 8월: 고려대학교 전자공학과 박사
 1996년 8월~2000년 3월: 현대전자산업주식회사 차장
 2000년 3월~현재: 현대오트넷주식회사 차장
 ※ 주관심분야: 음성인식/합성

● 고 한 석 (Hanseok Ko)



1982년 5월: 미국 카네기 멜론 대학교 전기공학 (공학사)
 1986년 5월: 미국 메릴랜드 대학교 시스템공학 (공학석사)
 1988년 5월: 미국 존스 홉킨스 대학교 전기공학 (공학석사)
 1992년 5월: 미국 카톨릭 대학교 전기공학 (공학박사)
 1995년 3월~현재: 고려대학교 전자컴퓨터공학과 교수
 ※ 주관심분야: 영상 및 음성 신호처리, 패턴 인식, 데이터 융합