

# Missing-Feature 복구를 위한 대역 독립 방식의 베이지안 분류기 기반 마스크 예측 기법

## Mask Estimation Based on Band-Independent Bayesian Classifier for Missing-Feature Reconstruction

김 우 일\*, Richard M. Stern\*\*, 고 한 석\*\*\*  
(Wooil Kim\*, Richard M. Stern\*\*, Hanseok Ko\*\*\*)

\*Dept. of Electrical Engineering, University of Texas at Dallas

\*\*Dept. of Electrical and Computer Engineering, Carnegie Mellon University

\*\*\*고려대학교 전자컴퓨터공학과

(접수일자: 2005년 11월 22일; 채택일자: 2005년 12월 13일)

본 논문에서는 알려지지 않은 잡음 환경에서 강인한 음성 인식 성능을 위하여 missing-feature 복구 기법을 다루며, 베이지안 분류기를 기반으로 하는 마스크 예측 기법의 성능을 향상시킬 수 있는 방법을 제안한다. 기존의 마스크 예측 기법에서는 배경 잡음 종류에 독립적인 성능을 위해 전 주파수 대역을 분할하여 발생시킨 유색 잡음을 마스크 예측기의 훈련에 이용하였으나, 제한된 양의 훈련 데이터베이스 조건에서는 성능의 한계가 불가피하다. 보다 다양한 잡음 스펙트럼을 반영하면서 마스크 예측의 성능을 향상시키기 위해, 서로 다른 주파수 대역에 독립적인 구조를 가지는 베이지안 분류기를 제안하며, 훈련에 사용하는 유색 잡음의 생성 방식을 이에 맞게 수정한다. 각각의 주파수 대역을 분할하여 유색 잡음을 생성함으로써 다양한 잡음 환경을 반영하는 동시에 훈련 데이터베이스 부족 문제를 줄일 수 있다. 제안하는 마스크 예측 기법을 클러스터 기반의 missing-feature 복구 기법과 결합하여 음성 인식기에 적용함으로써 성능을 평가한다. 실험 결과는 제안한 기법이 백색 잡음, 자동차 잡음, 배경 음악 환경에서 기존의 방법에 비해 향상된 성능을 가짐을 입증한다.

**핵심용어:** 음성 인식, missing-feature, 마스크 예측, 베이지안 분류기, 유색 잡음

**투고분야:** 음성처리 분야 (2,5)

In this paper, we propose an effective mask estimation scheme for missing-feature reconstruction in order to achieve robust speech recognition under unknown noise environments. In the previous work, colored noise is used for training the mask classifier, which is generated from the entire frequency partitioned signals. However, it gives a limited performance under the restricted number of training database. To reflect the spectral events of more various background noise and improve the performance simultaneously, a new Bayesian classifier for mask estimation is proposed, which works independent of other frequency bands. In the proposed method, we employ the colored noise which is obtained by combining colored noises generated from each frequency band in order to reflect more various noise environments and mitigate the "sparse" database problem. Combined with the cluster-based missing-feature reconstruction, the performance of the proposed method is evaluated on a task of noisy speech recognition. The results show that the proposed method has improved performance compared to the previous method under white noise, car noise and background music conditions.

**Keywords:** Speech Recognition, Missing-feature, Mask Estimation, Bayesian Classifier, Colored Noise

**ASK subject classification:** Speech Signal Processing (2,5)

## I. 서론

음성 인식 시스템의 성능을 하락시키는 주요한 원인은 음향 모델 훈련에 사용되는 음성 데이터베이스를 수집하는 환경과 실제 시스템이 사용되는 인식 환경의 음향학적 환경이 크게 달라지는데 있으며, 배경 잡음이 두 환경의 차이를 발생시키는 대표적인 예이다. 실제 환경에서의 음성 인식 성능을 최대화하기 위해 다양한 접근 방법을 통해 인식 환경과 훈련 환경의 차이를 줄이기 위한 연구가 최근 십수년간 집중되어 왔으며, stationary한 잡음 환경에서는 상당한 성과의 연구 결과를 나타냈다. 하지만, 배경 음악과 같이 시간에 따라 성질이 변하는 배경 잡음에 대해서는 여전히 취약한 성능을 나타낸다. 현재 존재하는 대부분의 기법들이 잡음 성분의 예측을 기초로 하며 얼마나 정확하게 잡음 성분을 예측하느냐가 그 성능을 좌우하게 되는데, 시간에 따라 변하는 잡음을 신뢰적으로 추정하는 것이 실제 환경에서는 매우 어렵기 때문이다. 오염된 음성에 섞인 상태의 잡음 요소를 예측하는 것은 문제를 더욱 어렵게 한다[1-2].

90년대 이후 연구되어 온 missing-feature 기반의 기법은 잡음 성분의 예측에 의존하기 보다는 잡음에 강인한 음성 정보를 인식 과정에 이용하려는 접근 방법을 가지므로 non-stationary한 잡음 환경에서 음성 인식 성능 향상에 보다 효과적인 것으로 알려지고 있다. 이론적으로는 오염시키는 잡음의 특성에 무관하게 음성 인식 성능 향상을 보장할 수 있으므로, 시간에 따라 변하는 잡음 환경에서도 우수한 성능을 나타낼 수 있다[4-5].

Missing-feature 기반 기법은 입력된 오염 음성으로부터 손실 (missing)된 스펙트럼 구역을 결정하는 마스크 (mask) 예측 단계와 손실되었다고 판단되는 구역을 어떻게 처리하는지에 관한 과정으로 이루어지며, 본 논문에서는 첫 번째 단계인 마스크 예측 기법에 초점을 맞춘다. Seltzer 등은 잡음 환경 종류에 독립적인 성능을 가지는 마스크 예측을 위해 베이시안 분류기 (Bayesian classifier) 기반의 기법을 제안하였으며, 백색 잡음 (white noise)을 모델 훈련에 이용하였다[6]. 이 연구의 후속 작업에서 백색 잡음만을 이용하는 것은 테스트 환경에 따라 성능이 달라질 수 있는 문제점이 제기되었으며, 주파수 대역을 분할하여 생성한 유색 (colored) 잡음을 훈련 기법에 이용하여 보다 향상된 성능을 가지는 마스크 예측 기법이 제안되었다[7].

본 논문에서는 기존에 제안된 마스크 예측 기법의 구

조에서는 주파수 전체를 분할하여 생성하는 유색 잡음을 훈련에 이용하는 방식이 성능 향상에 한계를 가지게 되는 문제점을 제기한다. 보다 다양한 잡음 환경의 특성을 효과적으로 반영하기 위해서는 베이시안 분류기의 구조가 근본적으로 바뀌어야 함을 기술하고, 그 대안으로서 다른 주파수 대역에 독립된 방식을 가지는 마스크 예측 기법을 제안한다. 대역 독립의 구조를 가지는 베이시안 분류기의 음향 모델에 다양한 잡음 특성을 반영하기 위하여 대역 내 분할된 유색 잡음을 훈련에 이용하는 기법을 제안하며, missing-feature 복구 기법과의 결합을 통해 다양한 잡음 환경에서 음성 인식 시스템의 전처리 과정으로서의 성능을 평가함으로써 기존의 방법들에 비해 우수한 성능을 가지는 것을 보이고자 한다.

본 논문은 다음과 같이 구성된다. II장에서 일반적인 missing-feature 기법을 구성하는 두 가지 단계에 관하여 설명하고, III장에서는 기존의 마스크 예측 기법이 가지고 있는 문제점을 제기함으로써 연구의 동기를 밝힌다. IV장에서는 본 논문에서 제안하는 대역 독립 방식의 마스크 예측 기법을 위한 베이시안 분류기의 설계 과정과 대역 내 분할된 유색 잡음의 제작 방법에 대해 상세히 기술한다. V장에서 제안하는 기법의 성능 평가를 위해 실시한 실험과 그 결과에 관해 고찰하고 VI장에서 결론을 맺는다.

## II. Missing-feature 복구 기법

잡음 환경에서 음성 인식 성능 향상을 위한 missing-feature 기반의 기법은 Lippmann 등에 의해 처음으로 소개되어 발전되어 왔다[3-4]. Missing-feature 기법은 오염된 입력 음성에서 잡음에 강인하다고 판단되는 음성 성분만을 인식 과정에 이용하려는 방법으로 잡음 종류에 무관한 성능을 얻기 위한 시도이다. Missing-feature 기법은 크게 두 가지 단계로 이루어진다. 첫 번째는 오염된 음성 입력의 스펙트럼으로부터 배경 잡음에 현저히 오염되어 음성 정보를 잃어버렸다고 판단되는 손실 (missing 또는 unreliable)된 부분과 여전히 음성 정보를 유지하여 음성 인식에 도움이 될 수 있다고 판단되는 신뢰적인 (reliable) 부분으로 구분하는 단계이다. 이와 같이 손실 구역과 신뢰 구역을 “마스킹 (masking)” 하는 단계를 마스크 예측 과정이라고 부른다. 두 번째

단계는 향후 음성 인식 처리를 위해 손실되었다고 판단되는 스펙트럼 구역을 어떻게 처리하는지에 관한 과정이다. 즉, 손실 구역을 그대로 두고 인식 (decoding) 단계의 확률 계산과정에서 특별히 고안된 계산 방식을 적용하거나, 손실 구역을 복구함으로써 일반적인 음성 인식 과정에 적용할 수 있다. 전자의 방법이 인식 시스템을 변화된 잡음 환경에 맞게 수정하는 “분류기 보상 (classifier compensation) 기법” 이라 하면, 후자는 일종의 전처리 기법으로서 일반적인 음성 인식 시스템에 적용이 가능한 “특징 보상 (feature compensation) 기법” 으로 볼 수 있다.

**2.1. 마스크 예측 (mask estimation)**

앞에서 설명한 것과 같이 마스크 예측은 입력 음성의 스펙트럼 중 음성 특징이 우세한 부분은 신뢰 구역, 잡음이 우세한 부분은 손실 구역으로 구분하는 과정이다. Missing-feature 기법이 소개된 이후 많은 연구에서는 다양한 방식으로 추정된 배경 잡음 정보와의 비교를 통해 신뢰 또는 손실 구역을 결정하는 방법을 채택하였으나, 이는 잡음 추정의 정확도에 따라 그 성능이 크게 좌우되며 특히, 시간에 따라 변하는 잡음 환경일 경우 성능이 크게 저하되는 것을 보였다. Seltzer 등은 배경 잡음의 종류에 무관한 성능을 가지는 마스크 예측을 위해 베이지안 분류기 기반의 기법을 제안하였으며, 이 기법에서는 잡음에 강인한 음성의 특징 (feature)을 이용하여 분류기를 설계하였다[6]. 본 논문에서는 그 후속 작업으로서 잡음 종류에 독립적으로 작동하면서 보다 향상된 성능을 가지는 마스크 예측 기법에 초점을 둔다.

**2.2. Missing-feature 복구 (reconstruction)**

Missing-feature 기법을 일반적인 음성 인식 시스템의 전처리 과정에 적용할 수 있도록 손실 구역으로 판단된 스펙트럼 부분을 복구하는 기법이 연구되었으며, 깨끗한 음성 특징의 분포와 MAP (Maximum A Posteriori) 기법을 이용하는 클러스터 (cluster) 기반의 복구 기법과 상관도 (correlation)를 이용하는 복구 기법이 제안되었다[5]. 전자는 특정 프레임에 대해 음성의 특징 분포 중 가장 유사한 클러스터를 선택하여 복구하는 방법이고, 후자는 인접 프레임간의 상관도를 이용한다. Marginal 계산 기법을 사용하는 분류기 보상 기법이 음성 특징 벡터로서 로그 스펙트럼을 사용해야하는 문제점을 감안하면, 캡스트럼으로 변환되어 전처리로서 적용할 수 있는

missing-feature 복구 기법이 상대적으로 우월한 성능을 나타낸다. 본 논문에서는 구현 과정이 비교적 간단하면서 좋은 성능을 나타내는 클러스터 기반의 복구 기법을 사용한다.

클러스터 기반의 missing-feature 복구 기법을 위해서는 깨끗한 음성의 로그 스펙트럼 특징 벡터의 확률 분포를 K개의 클러스터를 가지는 가우시안 혼합 모델 (Gaussian Mixture Model, GMM)로 모델링한다. 시간 t에서 오염된 입력 음성의 로그 스펙트럼 벡터 S(t)가 손실 구역 S<sub>m</sub>(t)를 가질 때, 입력 음성 S(t)에 대한 클러스터 인덱스  $\hat{k}_{s(t)}$ 는 식 (1)과 같이 사후 확률을 이용하여 결정할 수 있다. S(t)가 손실된 요소를 가지므로, 사후 확률은 다음과 같은 적분 식을 통해 계산한다.

$$\hat{k}_{s(t)} = \arg \max_k \{P(S(t) | k)P(k)\} \\ = \arg \max_k \left\{ P(k) \int_{-\infty}^{Y_m(t)} P(S(t) | k) dS_m(t) \right\} \quad (1)$$

식 (1)에서 Y<sub>m</sub>(t)는 손실 구역에서 실제 관찰된 스펙트럼 값을 나타낸다. 사후 확률을 가장 크게 하는 클러스터 인덱스  $\hat{k}_{s(t)}$ 가 결정되면 이에 해당하는 가우시안 분포를 이용하여 손실 구역을 복구한다. 식 (2)와 같이 신뢰 구역 S<sub>0</sub>(t), 인덱스  $\hat{k}_{s(t)}$ 에 해당하는 클러스터의 평균과 분산, 손실 구역에서의 관찰값 Y<sub>m</sub>(t)를 이용하여 bounded MAP 예측 기법을 통해 복구된 스펙트럼 S<sub>m</sub>(t)을 구할 수 있다.

$$\hat{S}_m(t) = \arg \max_{S_m} \{P(S_m(t) | S_0(t), \mu_{\hat{k}_{s(t)}}, \Sigma_{\hat{k}_{s(t)}}, S_m(t) \leq Y_m(t))\} \quad (2)$$

**III. 기존 베이지안 분류기 기반의 마스크 예측 기법에서의 문제점**

**3.1. 베이지안 분류기 기반의 마스크 예측 기법 [6]**

기존의 연구로서 잡음 종류에 무관한 missing-feature 기법을 구현하기 위해 베이지안 분류기를 기반으로 하는 마스크 예측 기법이 제안되었다. 이 기법에서는 배경 잡음에 강인한 음성 특징 (feature)을 이용하여 마스크 예측을 위한 베이지안 분류기를 설계한다. 이를 위한 음성 특징으로 Comb-filter Ratio (CFR), 서브밴드 에너지와

전체 에너지 비율, 서브밴드 에너지와 잡음의 최저 에너지와의 비율, 스펙트럼의 편평도 (flatness) 등을 사용한다. 음성 특징 벡터는 각 로그 스펙트럼 요소에 대응하는 주파수 대역의 음성 신호 성분으로부터 계산하는데, 각 주파수 대역은 로그 스펙트럼 추출에 사용되는 멜 필터 뱅크(Mel-filter bank)의 대역과 동일하다. 신뢰 부분과 손실 부분이 표기된 훈련 데이터를 이용하여 각 멜 필터 뱅크 대역에 대해 베이지안 분류기에 사용되는 음향 모델을 훈련한다. 음향 모델은 우도 (likelihood)를 계산하기 위한 GMM 모델과 사전 확률 (a priori) 정보로 구성된다. CFR 값을 얻을 수 있는 유성음과 그렇지 않은 무성음이 서로 다른 모델을 가지며, 유성음과 무성음에 대해 각각 신뢰 모델과 손실 모델을 구성한다. 음성이 입력되면 각 멜 필터 뱅크의 주파수 대역으로부터 마스크 예측을 위한 특징을 추출하고, 이미 훈련된 분류기를 적용하여 사후 확률이 높은 모델을 선택함으로써 해당 대역이 신뢰 또는 손실 구역임을 결정한다.

기존 연구에서는 백색 잡음에 오염된 음성 데이터를 이용하여 베이지안 분류기를 위한 음향 모델을 훈련하였는데, 이는 다양한 SNR (신호 대 잡음 비)의 백색 잡음을 오염시킴으로써 각 로그 스펙트럼 요소가 다양한 잡음에 오염된 효과를 적용하기 위한 시도이다. 하지만, 마스크 예측에 사용하는 음성 특징들이 각 대역에 독립적이기 보다는 다른 대역에 의해 영향 받을 수 있는 성질을 가지기 때문에, 백색 잡음에 의한 훈련 방식은 잡음 환경에 따라 성능이 많이 달라지는 문제점이 있음을 후속 연구에서 실험적으로 입증하였으며, 다음 절에서 설명한다.

### 3.2. 유색 잡음 (colored noise)을 훈련에 이용하는 마스크 예측 기법 [7]

본 논문의 선행 연구로서 마스크 예측을 위한 음향 모델의 훈련 과정에 유색 잡음을 이용하는 기법이 제안되었다. 이 연구에서는 마스크 예측을 위해 추출되는 특정 주파수 대역의 음성 특징이 다른 주파수 대역 뿐 아니라 인접 시간의 신호 값에 영향 받을 수 있음을 지적하고, 기존 연구에서 사용한 백색 잡음만으로는 다양한 잡음 환경을 효과적으로 반영하지 못하는 문제점을 실험적으로 입증하였다. 이 연구에서는 모든 주파수 대역에 걸쳐 다양한 스펙트럼 모양을 나타내는 유색 잡음을 인공적으로 발생시켜 이를 훈련을 위한 데이터베이스 제작에 이용하였다.

유색 잡음을 훈련에 이용한 마스크 예측 기법에서는 전체 주파수 대역을  $N$ 개로 분할하여 각 주파수 대역에 대해 대역통과 (bandpass) 필터를 설계한 후 백색 잡음을 통과시켜  $N$ 종류의 협역 (narrow band) 유색 잡음을 발생시켰다.  $N$ 종류의 잡음을 일정 시간을 두고 랜덤하게 조합함으로써  $2^N$  종류의 유색 잡음을 발생시킬 수 있으며, 생성된 유색 잡음의 조합이 실제 환경에서 관찰할 수 있는 다양한 잡음 환경을 반영하는 것으로 가정하였다. 이와 같이 발생된 유색 잡음을 훈련에 사용되는 깨끗한 음성 샘플에 다양한 SNR로 오염시킴으로써 최종적인 오염된 훈련 데이터베이스를 제작하였다. 그림 1은 전체 주파수 대역을 8개로 분할하여 생성한 유색 잡음의 예이며 (a)는 stationary한 특장을, (b)는 시변적인 (time-varying) 특장을 가지도록 발생시킨 것이다.

이와 같은 방법에서는 주파수 대역의 분할 개수  $N$ 을 증가시킴으로써 보다 다양한 잡음 종류를 발생시킬 수 있으므로, 이론적으로는  $N$ 을 DFT (Discrete Fourier Transform) 포인트의 절반만큼 증가시키고 다양한 SNR을 적용하면 자연 현상에서 발생하는 모든 종류의 잡음을 표현할 수 있다. 하지만, 잡음의 종류를 다양하게 할수록 제한된 양의 훈련 데이터베이스 안에서는 각각의 유색 잡음 종류의 발생 빈도가 낮아지므로, 훈련 데이터베이스에 효과적으로 반영하기 위해서는 데이터베이스의 양이 이에 맞게 증가되어야 하는데, 이는 현실적인 방법이 아니다. 후속 연구의 실험에서도  $N$ 을 증가시켜도 훈련 데이터의 양이 제한되어 있으므로 그 성능이 동일하게 증가되지 않음을 보였다.

본 논문에서는 이러한 문제점을 근본적으로 해결하기 위한 접근 방법을 제안한다. 즉, 마스크 예측을 위한 베이지안 분류기에 사용되는 음성 특징 벡터가 서로 다른 주파수 대역으로부터 영향을 받는 기존의 구조에서는, 전체 주파수 상에서 다양한 스펙트럼 종류를 반영하기 위해 기존 방식의 유색 잡음을 이용하는 것은 불가피하게 훈련 데이터베이스의 제한 현상을 가져온다. 이에 따라 본 논문에서는 서로 다른 주파수 대역에 완전히 독립적인 베이지안 분류기의 설계를 제안한다. 다른 주파수 대역에 독립적인 분류기가 설계되면, 다양한 잡음 현상을 반영하기 위해서는 같은 대역 내에서만 다양한 훈련 데이터를 적용해주면 되므로, 상대적으로 작은 숫자의 잡음 종류만으로도 전체적으로는 다양한 잡음 환경을 반영할 수 있는 효과를 가진다.

## IV. 다른 주파수 대역에 독립적인 마스크 예측을 위한 베이시안 분류기의 설계

### 4.1. 다른 주파수 대역에 독립적인 음성 특징 추출

#### 4.1.1. Comb Filter Ratio (CFR)

이전 연구에서도 사용했던 음성 특징으로서 유성음 구간에서 기본 주파수 (fundamental frequency)의 하모닉 (harmonic)이 형성되는 스펙트럼의 봉우리 (peak) 부분과 골짜기 (valley) 부분의 에너지 비율을 측정함으로써 잡음에 얼마나 오염되었는지를 나타낼 수 있는 수치이다. 잡음에 오염될수록 골짜기 영역의 에너지가 증가하게 되어 그 비율이 작아지게 된다. 하모닉 영역에서의 에너지는 다음과 같은 전달함수를 가지는 comb 필터를 통과시킴으로써 계산할 수 있다.

$$H_{comb}(z) = z^{-p} / (1 - gz^{-p}) \quad (3)$$

식 (3)에서  $p$ 는 기본 주파수에 대응하는 피치 (pitch)의 주기를 나타내며,  $g$ 는 필터의 기울기를 결정하는 파라미터이다. 스펙트럼의 골짜기 부분의 에너지는 위 전달 함수를 shift 시킨 또 다른 comb 필터를 통과시킴으로써 계산할 수 있다. 입력된 신호로부터 피치를 검출하여 comb 필터를 설계하고, 두 필터를 통과시킨 신호에 멜 필터 बैं크를 적용함으로써 각각의 주파수 대역에 대해 CFR 값을 얻을 수 있다. 본 논문에서는 피치 검출을 위하여 histogram 기반의 피치 검출 기법을 사용하였다[8].

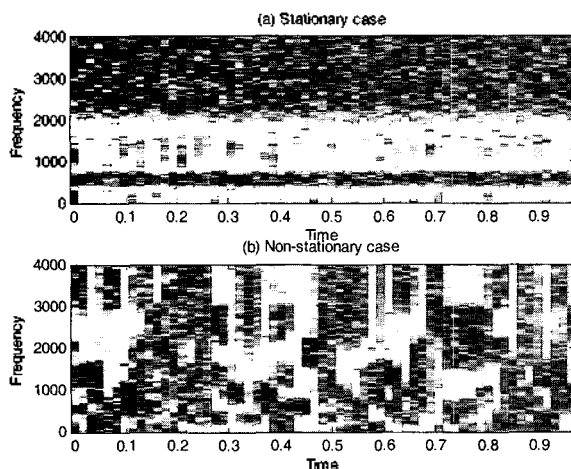


그림 1. 주파수 전체 분할 방식으로 생성된 유색 잡음의 스펙트로그램 (7).

Fig. 1. Spectrogram of the colored-noise obtained by entire frequency partition method (7).

#### 4.1.2. 서브밴드 켈스트럼 (cepstrum) 계수

음성 인식 과정에서와 같이 스펙트럼의 모양 정보를 이용하기 위해 켈스트럼 계수를 특징으로 사용한다. 단, 다른 주파수 밴드에 독립적인 특징을 위해 멜 필터 बैं크에 의해 선택된 스펙트럼 영역에 로그를 취한 후 DCT 변환을 함으로써 서브밴드 켈스트럼을 계산한다. 멜 필터 बैं크에 의해 선택되는 주파수 요소 개수와 동일한 DCT 변환을 적용하여 상위  $N$ 개의 켈스트럼 계수를 얻으며, 본 논문에서는 5차의 계수를 사용한다. 시간에 따른 스펙트럼의 변화 정도를 반영하기 위해 1차 미분값을 함께 사용한다.

#### 4.1.3. Spectral Flatness Measure (SFM)

SFM은 분석하고자 하는 신호가 잡음과 같은 랜덤한 스펙트럼 형태를 가지는지 톤 (tone) 성분이 우세한지에 대한 정보를 나타내며 음성 처리 분야에서는 주로 유/무성음 구간을 판별하는 척도로 이용된다[9]. SFM은 식 (4)과 같이  $N$ 개의 스펙트럼 요소  $X[n]$ 의 기하 평균과 산술 평균의 비율로 구할 수 있다.

$$SFM = \frac{\left( \prod_{n=1}^N X[n] \right)^{\frac{1}{N}}}{\frac{1}{N} \sum_{n=1}^N X[n]} \quad (4)$$

본 논문에서는 SFM 정보가 잡음의 오염 정도를 나타낼 수 있는 정보로 기대하며 멜 필터 बैं크로 선택되는 각각의 로그 스펙트럼의 영역으로부터 이를 계산하여 마스크 예측을 위한 음성 특징으로 사용한다.

### 4.2. 베이시안 분류기의 설계

이전 연구에서와 마찬가지로 마스크 예측을 위한 베이시안 분류기를 설계한다. 앞 절에서 설명한 음성 특징을 이용하며, 모든 멜 필터 बैं크의 주파수 대역에 대해 유성음은 12차, 무성음은 11차 특징 벡터를 가지는 음향 모델을 신뢰 영역과 손실 영역에 대해 각각 구성한다. 모든 음향 모델은 음성 특징 벡터의 확률 분포를 나타내는  $K$ 개의 가우시안 요소로 구성되는 GMM과 사전 확률 정보로 구성되어 마스크 예측 과정에서 입력 음성에 대한 사후 확률을 구할 수 있도록 한다.

#### 4.3. 모델 훈련을 위한 대역 내 분할된 유색 잡음제작

본 논문에서는 다른 주파수 대역에 독립적인 마스크

예측을 위해 설계된 베이지안 분류기의 훈련을 위해 이전 연구에서 제안된 유색 잡음 이용 기법을 대역 독립의 구조에 맞도록 수정한다. 기존 연구에서 이용된 유색 잡음은 전 주파수 대역에 걸쳐 다양한 스펙트럼 모양을 얻기 위해 주파수 전체를 N개로 분할하였다. 본 논문에서 제안하는 방법에서는 다른 주파수 대역의 영향을 고려하지 않아도 되므로 특정 주파수 대역 내에서의 다양한 스펙트럼 모양만을 반영해 주면 된다. 이에 따라 제안하는 방법에서는 전체 주파수 대역이 아닌 각각의 주파수 대역을 N개로 분할하여 다양한 잡음 환경을 표현하고자 하며 본 논문에서는 이를 대역 내 분할된 유색 잡음이라 부른다.

대역 내 분할된 유색 잡음을 발생시키기 위해서는 우선 특정 주파수 대역을 N개의 구역으로 분할한 후, 이에 따른 협역의 유색 잡음을 발생시킨다. 기존의 방법에서는 백색 잡음을 대역 통과 필터를 적용하여 협역 유색

잡음을 발생시켰으나, 본 연구에서의 '협역'은 하나의 주파수 대역을 N개로 분할한 구역이므로 매우 좁은 주파수 대역을 가지게 되며, 이에 따라 필터로 구현하기에 현실적인 어려움이 따른다. 본 논문에서는 대역 통과 필터를 사용하지 않고, 백색 잡음을 푸리에 변환하여 스펙트럼 영역에서 분할된 주파수 대역만 남기고 푸리에 역변환을 통해 신호를 얻는 방식을 사용하였다.

앞에서 설명한 과정을 거친 결과, 대역 내 분할된 매우 좁은 주파수 대역을 가지는 N종류의 유색 잡음을 얻게 되며 이를 랜덤하게 조합함으로써  $2^N$  종류의 유색 잡음을 얻게 된다. 기존의 방법과 마찬가지로 일정 시간마다 랜덤의 조합 선택함으로써 유색 잡음의 stationary 특성을 반영하였다. 본 논문에서는 30, 60, 300ms 마다 랜덤하게 조합을 선택함으로써 시간에 따라 변하는 잡음을 생성하였고, 잡음 전체 구간동안 같은 조합을 사용하여 stationary한 잡음을 발생시켰다. 이와 같이 얻어진 대역 내 분할된 유색 잡음을 훈련에 사용되는 깨끗한 음성 데이터베이스에 다양한 종류의 SNR에 맞추어 오염시킴으로써 마스크 예측을 위한 베이지안 분류기의 음향 모델의 훈련 데이터베이스를 제작한다. 본 논문에서는 20, 15, 10, 5, 0 dB의 SNR을 사용하였다.

그림 2는 기존의 유색 잡음 생성 방법과 본 논문에서 제안하는 방법과의 차이를 그림으로 나타낸 것이다. 그림에서는 기존의 방법은 주파수 전 대역을 4개로 분할하고, 제안하는 방법에서는 멜 필터 बैं크의 주파수 대역을 각각 4개로 분할하는 예를 보여준다. 기존의 방법에서는 조합을 통해  $2^4=16$  종류의 잡음을 발생시킬 수 있다. 제안하는 방법에서는 각각의 주파수 대역에서는 마찬가지로 16 조합이 생성되지만, 각 주파수 대역에 대한 음향 모델이 독립적으로 훈련되므로 주파수 전체를  $4 \times$  필터 बैं크 개수로 분할하여 조합한 효과, 즉  $2^{(4 \times \text{필터 बैं크 개수})}$  종류의 효과를 가진다. 이러한 효과는 제한된 양의 데이터베이스 조건에서 제안한 기법이 보다 다양한 잡음 스펙트럼을 반영할 수 있음을 나타낸다. 그림 3은 제안하는 방법에 따라 각 주파수 대역을 4개로 분할하여 생성한 유색 잡음의 예이다.

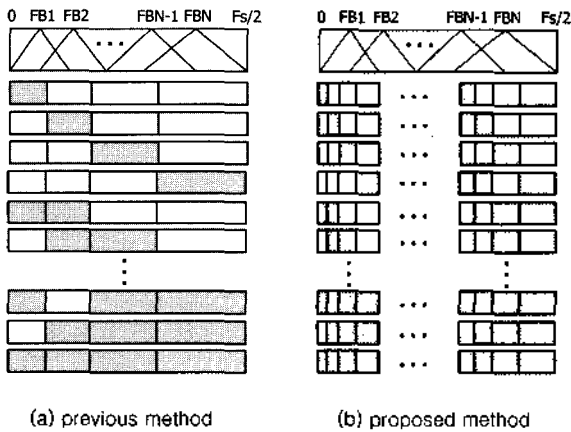


그림 2. 유색 잡음 생성을 위한 주파수 전체 분할 방식과 대역 내 분할 방식의 비교

Fig. 2. Comparison of entire frequency partition method and intra-band partition method.

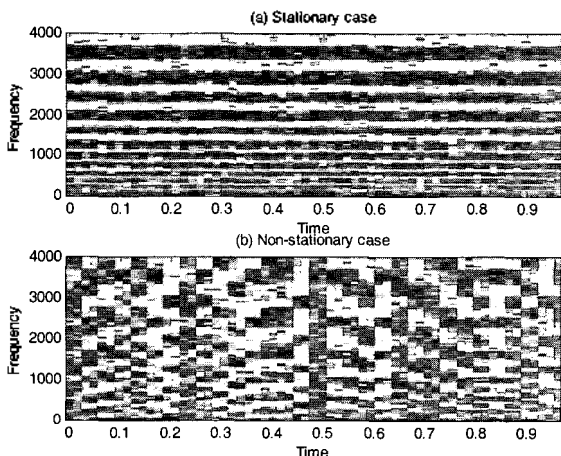


그림 3. 대역 내 분할 방식으로 생성된 유색 잡음의 스펙트로그램

Fig. 3. Spectrogram of the colored noise obtained by intra-band partition method.

## V. 실험 및 결과

### 5.1. 실험 환경 및 베이스라인 성능 평가

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

- 1) 영어 음성, 연속 숫자음 인식, 11단어+묵음 구간 (silence)+짧은 휴지 (short pause)
- 2) ETSI (European Telecommunications Standards Institute) 표준의 MFCC 특징 추출[11], 23개로 이루어진 멜 필터 बैं크 처리
- 3) 3-mixture, 16-state의 단어 모델, 2종류의 묵음 모델

Aurora 2.0에서 제공하는 clean-condition training, multi-condition testing 방식에 따라 깨끗한 환경에서 수집된 8,840개의 음성 데이터를 이용하여 음성 인식을 위한 HMM 및 클러스터 기반의 missing-feature 복구를 위한 GMM 모델을 훈련하였다. 본 논문에서는 보다 다양한 특성을 나타내는 잡음 환경에 대한 성능 평가를 위해 Aurora 2.0에서 제공하는 테스트용 데이터를 그대로 사용하지 않고 잡음 샘플을 이용하여 직접 제작하였다. Multi-condition testing 용으로 제공되는 Set A에 포함되어 있는 깨끗한 음성 데이터를 4개의 배경 잡음 (백색 잡음, 자동차 주행 잡음, speech babble, 배경 음악)에 5가지의 신호 대 잡음 비 (SNR, 20, 15, 10, 5, 0dB)에 따라 부가적으로 (additively) 오염시켜 새로운 multi-condition testing을 위한 데이터베이스를 제작하였다. 백색 잡음과 자동차 주행 잡음은 stationary 특성을 가지는 잡음 종류를 대표하며, speech babble과 배경 음악은 시간에 따라 변하는 non-stationary 특성을 반영하는 것으로 가정하였다. 백색 잡음은 NOISEX92 샘플로부터, 자동차 잡음과 speech babble은 Aurora2.0의 잡음 샘플로부터, 배경 음악은 비트와 빠르기가 다양한 유명 한국 가요 10 곡의 전주 부분에서 샘플링 하였

다. 최종적으로 잡음 환경 테스트 데이터는 5가지의 SNR에 따라 각각 1,001개의 샘플로 구성되어 전체 수는 총 20, 020 샘플이다.

향후 실험에서 사용되는 베이스라인 시스템의 성능과 본 논문에서 missing-feature의 복구를 위해 사용하는 클러스터 기반 기법의 성능을 각 잡음 환경의 5dB SNR 조건에서 평가하여 그 결과를 그림 4에 나타내었다. 베이스라인은 어떤 전처리 기법도 적용하지 않은 음성 인식기의 기본 성능을 말하며 성능 비교를 위해 전처리 기법으로 가장 일반적으로 사용되는 주파수 차감법 (spectral subtraction, SS)을 적용하여 평가하였다. 주파수 차감법에서는 배경 잡음을 추정하기 위해 250msec의 시간 지연을 갖는 최소 통계 (minimum statistics) 기법을 적용하였다[12]. 클러스터 기반의 복구 기법 자체의 성능 평가는 마스크 정보가 주어진 상태에서 이루어졌으며, 깨끗한 음성의 로그 스펙트럼 분포는 16개의 클러스터로 이루어진 GMM으로 추정하였다. 실험에 사용된 마스크 정보는 테스트 대상이 되는 오염 음성과 이에 사용된 깨끗한 음성의 스펙트럼 상에서 에너지 크기를 비교함으로써 얻을 수 있으며 이를 "오라클 (Oracle) 마스크"라 부른다. 특징 추출 과정에서 23개의 멜 필터 बैं크를 사용하므로 missing-feature 복구의 대상이 되는 로그 스펙트럼은 23차원 벡터이며, 이에 따라 각 프레임에서 예측되는 마스크 정보는 바이너리 값으로 구성되는 23차원 벡터이다. 따라서 향후에 논의되는 마스크 예측을 위한 베이스안 분류기 역시 23개의 요소에 대해 각각 독립적으로 구성된다.

그림 4의 결과로부터 본 논문에 사용된 음성 인식 시스템은 Aurora 2.0에서 추출된 자동차 잡음과 speech babble 잡음에 오염된 음성에 특별히 취약한 성능을 나타내는 것을 알 수 있으며, 주파수 차감법을 적용함으로써 모든 잡음 환경에 대해 어느 정도의 음성 인식 성능 향상이 있음을 알 수 있다. 클러스터 기반의 missing-feature 복구 기법의 성능에 주목할 필요가 있는데, 비록 오라클 마스크 정보를 이용했지만 매우 높은 성능 향상을 가져옴을 확인할 수 있었다. 특히 백색 잡음과 같은 광대역 잡음 보다는 자동차 잡음이나 speech babble과 같이 스펙트럼의 일부를 오염시키는 협대역 잡음에 대해 큰 폭의 성능 향상이 있음을 알 수 있으며, 이로부터 클러스터 기반의 missing-feature 복구 기법이 배경 잡음으로부터 오염되어 일부 손실된 음성 스펙트럼을 복구하는데 매우 효과적임을 확인할 수 있다. 또한 배경

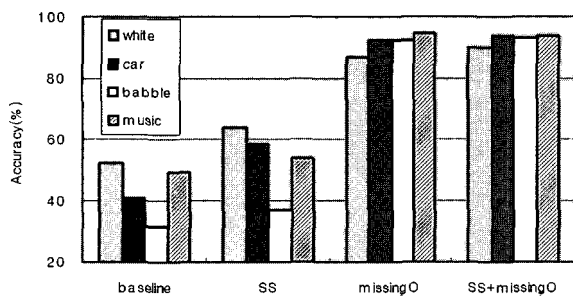


그림 4. 5dB SNR 환경에서 베이스 라인의 단어 인식률 (%)  
 Fig. 4. Word accuracy of baseline system under 5dB SNR conditions (%).

음악과 같이 시간에 따라 변화가 큰 잡음 환경에 대해서도 큰 성능 향상이 있었다. 주파수 차감법을 적용한 뒤 클러스터 기반 복구 기법을 적용하여 모든 잡음 환경에 대해 보다 향상된 성능을 얻을 수 있었으며 이는 주파수 차감법을 적용함으로써 SNR이 향상되어 그만큼 신뢰적인 음성 정보를 더 이용할 수 있음을 말한다. 본 논문에서는 향후 모든 실험에서 missing-feature 복구 기법을 주파수 차감법과 결합된 형태로 사용한다. 본 절에서 다른 오라클 마스크를 사용한 missing-feature의 복구 결과는 향후 동일한 클러스터 기반 기법을 적용하는 실험에서 최고 성능의 한계를 나타내는 지표가 된다.

5.2. 제안하는 마스크 예측 기법 성능 평가

그림 5와 그 이후의 결과에서는 본 논문에서 제안하는 마스크 예측 기법의 성능 평가를 나타낸다. 마스크 예측 기법의 평가는 위해 오라클 마스크와의 비교를 이용할 수도 있으나, 마스크 패턴에 따라 missing-feature의 복구 결과가 달라질 수 있으므로 오라클 마스크와 유사할수록 향상된 음성 인식 성능을 보장하는 것은 아니다

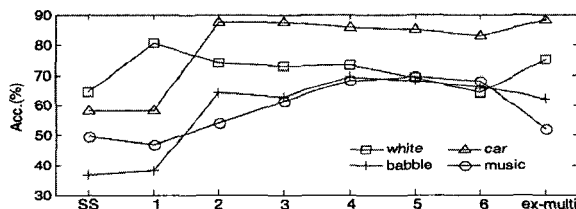


그림 5. 5dB SNR 환경에서 대역 유색 잡음 생성에 사용된 대역 내 분할 개수에 따른 성능 비교 (%)  
 Fig. 5. Performance as the number of intra-band partition for the colored noise under 5dB SNR conditions (%).

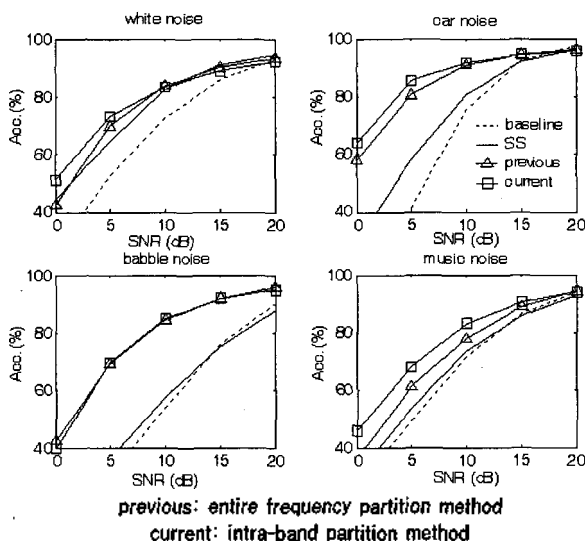


그림 6. 전체 테스트 세트에 대한 제안한 기법과 기존의 기법의 성능 비교 (%)  
 Fig. 6. Performance comparison on the entire testing set (%).

[6]. 따라서 본 논문에서는 missing-feature 복구와의 결합을 통해 음성 인식을 위한 전처리 기법으로서의 성능을 평가하고자한다. 그림 5의 실험을 위해서 4.1절에서 설명한 다른 주파수 밴드에 독립적인 특징 추출 기법을 사용하여 베이시안 분류기를 설계하였으며, 4.2절에서 제안한 대역 내 분할된 유색 잡음을 생성하여 마스크 예측을 위한 분류기의 음향 모델 훈련에 사용하였다. 대역 내에서의 다양한 잡음 종류를 생성하기 위해 대역 내 분할 개수  $N$ 을 1에서 6까지 증가시키며 성능을 비교하였다. 분할 개수 1은 대역을 분할하지 않은 것이므로 백색 잡음을 의미하며 기존의 제안된 실험 방법과 동일하다. 성능 비교를 위해서 실험에 사용되는 4가지 잡음 중 테스트 환경과 일치하는 잡음을 제외한 나머지 3가지를 혼합하여 훈련한 exclusive multi-style 훈련 기법 (ex-multi)의 결과를 같이 나타내었다. 즉, 백색 잡음 환경을 테스트 할 경우에는 백색 잡음을 제외한 자동차, speech babble, 배경 음악을 모델 훈련에 참여시킨 것을 말하며, 이와 같은 방법은 테스트 환경에 대한 정보가 없을 때, 현재 취득 가능한 모든 오염 음성을 이용하여 대안적으로 사용할 수 있는 multi-style 훈련 방식이다.

유색 잡음의 대역 내 분할 개수가 1일 경우는 분할하지 않은 경우이므로 이를 논외로 하면, 백색 잡음이나 자동차 잡음과 같이 stationary한 특징을 보이는 잡음 환경에서는 분할 개수를 증가시키며 따라 완만하게 성능이 하락하는 것을 볼 수 있다. Speech babble이나 배경 음악과 같이 시간에 따라 성질이 변하면서 다양한 스펙트럼 분포를 가지는 잡음의 경우에는 분할 개수를 늘일수록 점점 성능이 향상되다가 어느 지점에서 정체되어 결국에는 조금 하락하는 것을 확인할 수 있다.

이와 같이 잡음 환경의 종류에 따라 그 성능의 추이가 다른 것은 대역 내 분할된 유색 잡음에서 나타날 수 있는 스펙트럼 종류의 빈도와 각각의 잡음이 나타내는 스펙트럼 분포와 관련이 있다. 분할 개수가 적을수록 생성된 유색 잡음에서 발생하는 스펙트럼의 종류는 다양하지 못하지만, 제한된 개수의 훈련 데이터베이스 안에서는 나타날 수 있는 빈도수가 높아진다. 반대로 분할 개수가 많아질수록 점점 다양한 스펙트럼을 가지는 잡음 종류를 표현할 수 있지만 그만큼 나타나는 빈도수는 적어진다. 예를 들어 밴드를 2개로 분할했을 경우 발생할 수 있는 스펙트럼의 종류는  $2^2=4$  조합이 가능하며, 각각의 스펙트럼은 1/4의 발생 빈도를 가진다. 만일 밴드를 5개로 분할할 경우에는  $2^5=32$  조합의 종류를 표현할 수 있지만



표 1. 전체 테스트 세트에 대한 평균 단어 인식률과 상대 향상률 (%)  
Table 1. Average word accuracy and relative improvement ratio on the entire testing set (%).

	베이스 라인	주파수 차감법	기존 기법	제안한 기법	상대향상률 (%)*
백색잡음	66.24	75.39	75.96	77.87	4.59
자동차 잡음	64.35	71.67	84.20	86.44	3.38
speech babble	53.02	55.07	76.80	76.14	-1.36
배경 음악	65.78	67.40	71.64	76.43	9.52

\*상대 향상률=100\*(제안한기법-기존기법/기존기법)

1/32로 발생 빈도가 줄어들게 된다. 제한된 데이터베이스 내에서 특정 스펙트럼의 종류의 발생 빈도가 낮다는 것은 해당 스펙트럼의 특성을 반영하는 음향 모델을 훈련하는데 그만큼 충분하지 못함을 의미한다.

백색 잡음 환경의 경우에는 분할하지 않은 유색 잡음, 즉 백색 잡음을 그대로 훈련에 이용할 경우가 훈련 환경과 테스트 환경이 일치하는 경우이므로 가장 성능이 좋을 것으로 예상할 수 있으며 실험 결과에서도 이와 같이 나타났다. 유색 잡음의 분할 개수를 증가시킬수록 훈련 데이터베이스에서 나타나는 순수한 백색 잡음의 빈도수는 점점 감소되어 그 성능이 하락하는 것으로 생각할 수 있다. 마찬가지로 자동차 잡음도 스펙트럼 종류가 다양하지 않으면서 stationary한 특성을 가지므로 작은 분할 개수의 유색 잡음에서 그 발생 빈도가 높을 것으로 예상할 수 있으며 이에 따라 분할 개수를 늘임에 따라 성능이 하락함을 추론할 수 있다. Speech babble이나 배경 음악의 경우에는 상대적으로 다양한 스펙트럼 종류가 나타나므로 다양한 스펙트럼의 표현을 위해서는 훈련에 사용되는 유색 잡음의 대역 내 분할 개수가 많아져야 한다. 하지만, 어느 정도 증가한 후에는 마찬가지로 제한된 데이터베이스 안에서의 빈도수 하락의 이유로 성능이 정체되어 결국 하락하는 것으로 생각할 수 있다. 모든 잡음 환경에서의 성능 추이를 감안하면 전체적으로 각각의 대역을 4개로 분할했을 때 가장 좋은 성능을 보였다. Ex-multi 훈련 방식과 비교하여 자동차 잡음 환경을 제외하고는 백색 잡음에서는 대등한 성능을, speech babble과 배경 음악에서는 우수한 성능을 나타내었다. 이는 제안하는 훈련 기법이 미지 (unknown)의 테스트 환경에서 효과적으로 사용될 수 있음을 나타낸다.

그림 6에서는 본 논문에서 제안하는 마스크 예측 기법을 적용한 missing-feature 복구 기법을 각각의 잡음 환경에서 다양한 SNR 조건에 대해 기존의 방법과의 성능 비교를 나타낸다. 기존의 방법은 III장에서 설명한 기존의

마스크 예측기와 주파수 전체 분할 방식에 의해 생성된 유색 잡음을 예측기 훈련에 이용한 기법을 말한다. 제안하는 방법에서는 그림 5의 결과에 따라 대역 내 분할 개수를 4로 했으며, 기존 방법의 전체 주파수 분할 개수는 이전 연구를 참조하여 12로 하였다. 결과로부터 알 수 있듯이 speech babble 환경을 제외하고 본 논문에서 제안하는 기법이 나머지 세 가지 환경에서 주파수 전체 분할 유색 잡음 기법보다 우수한 성능을 나타내었다. 특히 백색 잡음과 자동차 잡음 환경에서는 5dB와 0dB와 같은 낮은 SNR 조건에서, 배경 음악 환경에서는 모든 SNR 조건에서 상당한 폭의 성능 향상을 보였다. 표 1에서는 전체 SNR에 대해 평균 성능과 제안한 기법의 기존 방법에 대한 상대적인 성능 향상률을 나타낸다. 특히, 배경 음악과 같이 다양한 종류의 스펙트럼 특성을 나타내는 잡음 환경에서 기존의 방법에 비해 상당한 성능 향상이 있었던 것은 다양한 잡음 스펙트럼을 반영하고자 했던 본 논문의 목적에 부합하는 결과로 볼 수 있다. Speech babble 환경에서는 기존의 방법보다 오히려 소폭 하락한 성능을 보여주는데, 상당한 삽입 오류 (insertion error)에 의한 영향으로 분석되었다. 마스크 예측 과정에서 피치 검출에 오류가 있을 경우 유성음과 무성음 모델이 서로 뒤바뀌어 선택되고, 이에 따라 마스크 예측 결과에 오류를 가져오는 것을 확인하였다. 특히 speech babble 환경에서 제안하는 방법이 기존의 방법보다 이러한 현상을 많이 발생시켰다. 향후 연구에서 피치 검출 오류에 의한 영향을 줄일 수 있는 후처리 기법에 관해 다루고자 한다.

## VI. 결론

본 논문에서는 잡음 환경에서 향상된 음성 인식 성능을 얻기 위해 missing-feature 기법의 전처리 과정으로서 효과적인 마스크 예측 기법을 제안하였다. 주파수 전체를 분할하여 유색 잡음을 생성하는 기존의 방식은 제한된 양의 데이터베이스 조건에서는 훈련 데이터베이스 부족 문제가 있음을 제기하였다. 제안하는 방법에서는 서로 다른 주파수 대역에 독립적으로 동작하는 베이지안 분류기를 설계하였으며, 대역 내 분할된 유색 잡음을 생성하여 분류기의 훈련에 사용하였다. 클러스터 기반의 missing-feature 복구 기법과 결합하여 음성 인식기의 전처리로서의 성능 평가를 실시하였으며, 실험 결과는 제안한 기법이 기존의 방

법에 비교하여 백색 잡음과 자동차 잡음 환경에서는 낮은 SNR 조건, 배경 음악 환경에서는 모든 SNR 조건에서 음성 인식 성능 향상에 효과적임을 나타내었다.

## 감사의 글

본 논문은 한국학술진흥재단의 해외 Post-doc, 연수 지원 (구 한국과학재단 지원, 과제번호: KRF-2004-214-D00142)을 통해 수행된 연구 결과의 일부입니다.

## 참고 문헌

1. R. Singha, R. M. Stern, and B. Raj, "Signal and Feature Compensation Methods for Robust Speech Recognition," Chapter in CRC Handbook on Noise Reduction in Speech Applications, CRC Press, 2002.
2. R. Singha, R. M. Stern, and B. Raj, "Model Compensation and Matched Condition Methods for Robust Speech Recognition," Chapter in CRC Handbook on Noise Reduction in Speech Applications, CRC Press, 2002.
3. R.P. Lippmann and B.A. Carlson, "Using missing feature theory to actively select features for robust speech recognition with interruptions, filtering, and noise," Eurospeech, 37-40, 1997.
4. M. Cooke, P. Green, L. Josifovski, and A. Vizinho, "Robust automatic speech recognition with missing and unreliable acoustic data," Speech Communication, 34 (3): 267-285, 2001.
5. B. Raj, M. L. Seltzer, and R. M. Stern, "Reconstruction of missing features for robust speech recognition," Speech Communication, 43 (4): 275-296, 2004.
6. M. L. Seltzer, B. Raj, and R. M. Stern, "A Bayesian classifier for spectrographic mask estimation for missing-feature speech recognition," Speech Communication, 43 (4): 379-393, 2004.
7. W. Kim, R. M. Stern, and H. Ko, "Environment-Independent Mask Estimation for Missing-Feature Reconstruction," Interspeech 2005, .2637-2640, Sep. 2005.
8. M. L. Seltzer, Automatic Detection of Corrupted Speech Features for Robust Speech Recognition, M.S. thesis, Carnegie Mellon University, 2000.
9. J. D. Johnston, "Transform coding of audio signals using perceptual noise criteria," IEEE Journal on Selected Areas in Communications, 6, 314-323, Feb. 1988.
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, 1182-1185, Sep. 1994.

---

## 저자 약력

---

### • 김 우 일 (Wooil Kim)

1996년 2월: 고려대학교 전자공학과 (공학사)  
 1998년 8월: 고려대학교 전자공학과 (공학석사)  
 2003년 8월: 고려대학교 전자공학과 (공학박사)  
 2004년 8월~2005년 8월: 미국 Carnegie Mellon University 전기컴퓨터공학과 박사후 연구원  
 2005년 9월~현재: 미국 University of Texas at Dallas 전기공학과 연구원  
 \*주관심분야: 신호처리, 음성인식, 잡음처리

### • Richard M. Stern (Richard M. Stern)

1970년 5월: 미국 MIT 전기공학 (공학사)  
 1972년 5월: 미국 University of California at Berkeley 전기공학 (공학석사)  
 1976년 5월: 미국 MIT 전기공학 (공학박사)  
 1977년 3월~현재: 미국 Carnegie Mellon University 전기컴퓨터 공학과 및 컴퓨터 과학과 교수  
 \*주관심분야: 신호처리, 철각 음향

### • 고 한 석 (Hanseok Ko)

1982년 5월: 미국 Carnegie Mellon University 전기공학 (공학사)  
 1988년 5월: 미국 Johns Hopkins University 전기공학 (공학석사)  
 1992년 5월: 미국 Catholic University of America 전기공학 (공학박사)  
 1995년 3월~현재: 고려대학교 전기컴퓨터공학과 교수  
 \*주관심분야: 영상 및 음성 신호처리, 패턴 인식