말소리와 음성과학 제6권 제4호 (2014.12.31)
ISSN 2005-8063, pp. 127~132
www.speechsciences.or.kr
http://dx.doi.org/10.13064/KSSS.2014.6.4.127

가산잡음환경에서 강인음성인식을 위한 은닉 마르코프 모델 기반 손실 특징 복원

HMM-based missing feature reconstruction for robust speech recognition in additive noise environments

조 지 원[1] · 박 형 민[2]
Cho, Ji-Won · Park, Hyung-Min

ABSTRACT

This paper describes a robust speech recognition technique by reconstructing spectral components mismatched with a training environment. Although the cluster-based reconstruction method can compensate the unreliable components from reliable components in the same spectral vector by assuming an independent, identically distributed Gaussian-mixture process of training spectral vectors, the presented method exploits the temporal dependency of speech to reconstruct the components by introducing a hidden-Markov-model prior which incorporates an internal state transition plausible for an observed spectral vector sequence. The experimental results indicate that the described method can provide temporally consistent reconstruction and further improve recognition performance on average compared to the conventional method.

Keywords: missing feature reconstruction, robust speech recognition, cluster-based reconstruction, hidden Markov model

1. 서론

잡음이 현존하는 실제 환경에서 일반적인 음성 인식 시스템은 인식 시스템의 모델이 학습되는 데이터와 환경차이 때문에 그 성능이 크게 떨어진다[1]. 특히, 이러한 불일치는 학습 단계에서 고려하지 못한 음향 잡음에 기인한다. 손실 특징 복원 방법은 잡음이 섞인 입력 음성 데이터와 학습 음성 데이터 사이의 불일치를 보완하여 강인한 음성인식을 수행할 수 있게 한다[2]. 손실 특징 복원 기술 중 클러스터 기반 복원(cluster-based reconstruction) 방법은 인식 시스템을 수정하지 않고 켑스트럼 특징을 사용할 수 있어서 더 나은 인식 성능을 제공한다[2].

클러스터 기반 복원 방법은 학습 데이터의 스펙트럼 벡터를 시간에 따라 서로 독립인 정규혼합분포의 랜덤프로세스로 가정하여 동일 스펙트럼 벡터 중에 음성이 지배적인 신뢰 성분을 이용하여 잡음이 지배적인 비신뢰 성분을 추정하는 방법이다[2]. <그림 1>은 클러스터 기반 복원을 적용한 잡음에 강인한 음성인식 시스템을 나타낸다.

본 논문은 음성인식 시스템의 잡음에 대한 강인성을 향상시키기 위하여 새로운 사전 확률 밀도 함수를 도입한 클러스터 기반 손실 특징 복원 방법을 제안한다. 은닉 마르코프 모델(HMM)은 연속되는 관찰 신호의 정상 또는 과도 상태를 잘 표현하기 때문에, 신호처리 및 통신에서 광범위 하게 쓰인다. 또, 최근에 디지털 정보 송신 중 데이터 손실을 보상하기 위한 모델로 쓰이기도 한다[3, 4]. 음성은 시간에 따라 서로 의존적인 연속되는 관찰 신호이므로 은닉 마르코프 모델을 사전 확률 모델로 도입하여 기존의 방법이 주파수 성분간 상관관계만 고려하는 반면 제안하는 방법은 시간적인 상관관계까지 고려하였다. 은닉 마르코프 모델 기반 특징 복원 방법이 제안된 바 있으나[5], 이 방법은 주파수 성분간 상관관계를 고려하지 않기 때문에 제안 방법이 장점을 가질 수 있다.

<그림 1>에서 마스크 추정은 입력 스펙트럼 벡터의 신뢰도를 판단하는 손실 특징 복원의 성능을 결정하는 중요한 과정이다. 많은 손실 특징 복원 연구에서 미리 깨끗한 음성 특징을 안다고 가정하고 오라클 마스크를 사용하는데, 이는 오라클 마

1) 서강대학교, jiwonn85@sogang.ac.kr
2) 서강대학교, hpark@sogang.ac.kr, 교신저자

본 연구는 LG연암문화재단의 지원으로 수행되었음.

접수일자: 2014년 11월 4일
수정일자: 2014년 12월 3일
게재결정: 2014년 12월 13일

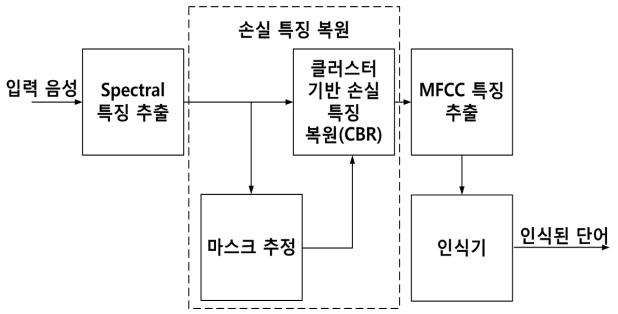

그림 1. 클러스터 복원 기반 잡음에 강인한 음성인식 시스템

스크가 손실 특징 복원 성능의 상한을 결정하기 때문이다. 본 논문은 오라클 마스크를 사용하여 제안한 방법의 상한을 기존 방법과 비교하는 것으로 한정하고, 마스크 추정 방법은 추후 연구로 미루고자 한다.

본 논문은 다음과 같은 순서로 구성되어 있다. 2장에서 기존의 클러스터 기반 손실 특징 복원 알고리즘에 대해 살펴보고, 3장에서 음성의 시간적 특성을 부과하는 방법에 대하여 설명한다. 그리고 4장에서는 기존 방법과 3장에서 제안한 방법을 비교하기 위한 실험 방법 및 결과를 기술한다. 마지막으로 5장에서 결론을 맺는다.

2. 클러스터 기반 손실 특징 복원 알고리즘

클러스터 기반 손실 특징 복원 방법은 학습 음성 데이터의 스펙트럼 벡터가 서로 독립이고 동일한 분포(i.i.d.)를 따른다는 가정을 두고 신뢰성 있는 스펙트럼 성분들을 이용해서 신뢰성이 없는 성분들을 복원하는 알고리즘이다. 스펙트럼 벡터의 선험적 확률 밀도 함수를 다음과 같이 정규혼합분포이라고 가정한다.

$$p(\boldsymbol{x})=\sum_{l}\frac{P(l)}{\sqrt{(2\pi)^{d}|\boldsymbol{\Theta}_{l}|}}exp\left(-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu}_{l})^{T}\boldsymbol{\Theta}_{l}^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_{l})\right). \quad (1)$$

여기서 $\boldsymbol{x}$는 깨끗한 학습 음성 데이터의 스펙트럼 벡터이고 d는 벡터의 차원수이다. 또, $P(l)$, $\boldsymbol{\mu}_{l}$, $\boldsymbol{\Theta}_{l}$은 각각 l번째 가우시안 성분의 선험적 확률, 평균 벡터, 그리고 공분산 행렬이다. 이 분포 파라미터들은 expectation maximization (EM) 알고리즘을 이용해 충분한 양의 깨끗한 음성 데이터로 학습된다[6].

$\boldsymbol{y}$를 잡음에 의해 $\boldsymbol{x}$가 왜곡된 입력 스펙트로그램 벡터라고 할 때, $\boldsymbol{y}_{r}$과 $\boldsymbol{y}_{u}$는 각각 $\boldsymbol{y}$의 신뢰, 비신뢰 성분이라고 한다. 마찬가지로 $\boldsymbol{x}_{r}$, $\boldsymbol{x}_{u}$를 각각 $\boldsymbol{x}$의 신뢰, 비신뢰 성분이라고 한다. 여기서 $\boldsymbol{x}_{r}$은 $\boldsymbol{y}_{r}$와 근사적으로 같다고 할 수 있고, $\boldsymbol{x}_{u}$는 $\boldsymbol{y}_{u}$를 상한으로 갖는 사후확률 최대화(bounded maximum a posteriori: BMAP)를 통해 다음과 같이 추정할 수 있다[2].

$$\widehat{\boldsymbol{x}_{u}}=\sum_{l}P(l|\boldsymbol{x}_{r},\boldsymbol{x}_{u}\le\boldsymbol{y}_{u})\operatorname{argmax}_{\boldsymbol{x}_{u}}p(\boldsymbol{x}_{u}|l,\boldsymbol{x}_{r},\boldsymbol{x}_{u}\le\boldsymbol{y}_{u}). \quad (2)$$

$P(l|\boldsymbol{x}_{r},\boldsymbol{x}_{u}\le\boldsymbol{y}_{u})$를 계산하기 위한 l번째 가우시안 성분의 공헌확률은 베이지안 법칙에 의해 다음과 같이 기술된다.

$$P(l|\boldsymbol{x}_{r},\boldsymbol{x}_{u}\le\boldsymbol{y}_{u})=\frac{P(l)p(\boldsymbol{x}_{r},\boldsymbol{x}_{u}\le\boldsymbol{y}_{u}|l)}{\sum_{l'}p(\boldsymbol{x}_{r},\boldsymbol{x}_{u}\le\boldsymbol{y}_{u}|l')}. \quad (3)$$

여기서, $p(\boldsymbol{x}_{r},\boldsymbol{x}_{u}\le\boldsymbol{y}_{u}|l)$은 다음과 같이 계산할 수 있다.

$$p(\boldsymbol{x}_{r},\boldsymbol{x}_{u}\le\boldsymbol{y}_{u}|l)=\int_{-\infty}^{\boldsymbol{y}_{u}}p(\boldsymbol{x}_{r},\boldsymbol{x}_{u}|l)d\boldsymbol{x}_{u}. \quad (4)$$

클러스터 기반 복원 방법은 깨끗한 음성 스펙트로그램 벡터가 시간에 따라 서로 독립이고 동일한 분포를 따르는 랜덤프로세스라고 가정하기 때문에, 비신뢰 성분들을 복원할 때, 같은 시간 프레임 안의 스펙트로그램 신뢰 성분들을 이용하여 복원한다. 따라서, 추정된 비신뢰 성분 값은 입력 스펙트로그램 벡터의 신뢰 성분이 적을 때, 해당 신뢰 성분에 그 값이 큰 영향을 받으므로 자주 실제 값과 차이가 크게 나타난다. 특히, 식 (3)에서 l번째 가우시안 성분의 입력 스펙트로그램 벡터 $\boldsymbol{y}$에 대한 공헌도 $P(l|\boldsymbol{x}_{r},\boldsymbol{x}_{u}\le\boldsymbol{y}_{u})$는 $P(l)$과 $p(\boldsymbol{x}_{r},\boldsymbol{x}_{u}\le\boldsymbol{y}_{u}|l)$로 계산한다. 선험적 확률 $P(l)$은 $\boldsymbol{y}$에 상관없이 고정되어 있기 때문에, 현재 입력 스펙트로그램 벡터 $\boldsymbol{y}$를 확률밀도함수 $p(\boldsymbol{x}_{r},\boldsymbol{x}_{u}\le\boldsymbol{y}_{u}|l)$에 대입하여 $P(l|\boldsymbol{x}_{r},\boldsymbol{x}_{u}\le\boldsymbol{y}_{u})$를 계산한다. 불행히도, $p(\boldsymbol{x}_{r},\boldsymbol{x}_{u}\le\boldsymbol{y}_{u}|l)$의 값을 얻기 위해서 식 (3)과 같이 주변화가 실행될 때 비신뢰 성분의 개수가 많을수록, 더 많은 오차를 발생시킨다.

만약, 음성 스펙트로그램에 대한 선험적인 정보가 있다면 적은 개수의 비신뢰 성분에도 정확한 추정을 위한 중요한 역할을 할 수 있다. 음성은 고유의 시간적 의존성을 가지고 있는데, 이 특성을 반영한 상태 천이를 갖고 있는 은닉 마르코프 모델을 이용하여 음성인식에 사용한다. 본 논문에서는 은닉 마르코프 모델을 깨끗한 음성 스펙트로그램 벡터의 선험적 확률 밀도 함수로 도입하여 음성의 시간적 의존성을 이용한다. 따라서, 특정 시간 프레임 안에서 적은 수의 신뢰 성분을 갖더라도 l번째 상태의 공헌도는 연속되는 입력 스펙트로그램 벡터에 대한 상태 천이 확률이 반영되어, 더 정확하게 비신뢰 성분 복원이 수행된다.

3. 은닉 마르코프 모델 기반 손실 특징 복원 알고리즘

기존의 클러스터 기반 손실 특징 복원은 음성의 선험적 확률 밀도 함수로 정규혼합분포를 사용하는데 반하여, 본 논문에서는 음성의 시간적 의존성을 이용하기 위해 은닉 마르코프

모델을 사용한다. 본 논문에서는 다방향 은닉 마르코프 모델을 이용하였고 상태당 관찰확률분포로 다음과 같은 하나의 가우시안 분포를 가정한다[7].

$$p(\boldsymbol{x}^{m}|\gamma^{m}=l) = \sum_{l}\frac{P(l)}{\sqrt{(2\pi)^{d}|\boldsymbol{\Theta}_{l}|}}exp\Big(-\frac{1}{2}(\boldsymbol{x}^{m}-\boldsymbol{\mu}_{l})^{T}\boldsymbol{\Theta}_{l}^{-1}(\boldsymbol{x}^{m}-\boldsymbol{\mu}_{l})\Big). \quad (5)$$

여기서 $\boldsymbol{x}^{m}$, γ^{m}은 각각 m번째 시간 프레임에서 스펙트로그램 벡터와 은닉 상태 색인을 나타낸다. 다방향 은닉 마르코프 모델 파라미터 중 l번째 상태의 초기 확률값 $P(\gamma^{1}=l)$, 평균 벡터 $\boldsymbol{\mu}_{l}$, 공분산 행렬 $\boldsymbol{\Theta}_{l}$, l'번째 상태에서 l번째 상태로의 상태 천이 확률 $P(\gamma^{m}=l|\gamma^{m-1}=l')$은 깨끗한 음성 데이터로 EM(expectation and maximization) 알고리즘을 이용해 추정한다.

$\boldsymbol{y}^{1:m}=(\boldsymbol{y}^{1},\cdots\boldsymbol{y}^{m})$으로 표현할 때, 음성의 선험적 확률밀도 함수를 은닉 마르코프 모델로 대체하면서 식 (2)의 사후확률 최대화 방법이 다음과 같이 수정된다.

$$\widehat{\boldsymbol{x}_{u}^{m}} = \sum_{l}P(\gamma^{m}=l|\boldsymbol{x}_{r}^{1:m},\boldsymbol{x}_{u}^{1:m}\le\boldsymbol{y}_{u}^{1:m}) \quad (6)$$
$$\arg\max_{\boldsymbol{x}_{u}^{m}} p(\boldsymbol{x}_{u}^{m}|\gamma^{m}=l,\boldsymbol{x}_{r}^{1:m},\boldsymbol{x}_{u}^{1:m}\le\boldsymbol{y}_{u}^{1:m}).$$

여기서, m번째 시간 프레임에서 상태 색인이 주어졌기에 상한이 있는 사후확률 최대화를 위한 확률 밀도 함수는 $p(\boldsymbol{x}_{u}^{m}|\gamma^{m}=l,\boldsymbol{x}_{r}^{1:m},\boldsymbol{x}_{u}^{1:m}\le\boldsymbol{y}_{u}^{1:m})=p(\boldsymbol{x}_{u}^{m}|\gamma^{m}=l,\boldsymbol{x}_{r}^{m},\boldsymbol{x}_{u}^{m}\le\boldsymbol{y}_{u}^{m})$로 근사화할 수 있다. $\boldsymbol{y}^{1:m}$에 대한 m번째 시간 프레임에서 l번째 상태의 공헌도, $P(\gamma^{m}=l|\boldsymbol{x}_{r}^{1:m},\boldsymbol{x}_{u}^{1:m}\le\boldsymbol{y}_{u}^{1:m})$는 베이지안 법칙에 의해 다음과 같이 쓸 수 있다.

$$P(\gamma^{m}=l|\boldsymbol{x}_{r}^{1:m},\boldsymbol{x}_{u}^{1:m}\le\boldsymbol{y}_{u}^{1:m}) = \frac{1}{Z}P(\gamma^{m}=l|\boldsymbol{x}_{r}^{1:m-1},\boldsymbol{x}_{u}^{1:m-1}\le\boldsymbol{y}_{u}^{1:m-1}) \quad (7)$$
$$p(\boldsymbol{x}_{r}^{m},\boldsymbol{x}_{u}^{m}\le\boldsymbol{y}_{u}^{m}|\gamma^{m}=l,\boldsymbol{x}_{r}^{1:m-1},\boldsymbol{x}_{u}^{1:m-1}\le\boldsymbol{y}_{u}^{1:m-1}).$$

여기서 $Z=\sum_{l'}P(\gamma^{m}=l'|\boldsymbol{x}_{r}^{1:m-1},\boldsymbol{x}_{u}^{1:m-1}\le\boldsymbol{y}_{u}^{1:m-1})$

$p(\boldsymbol{x}_{r}^{m},\boldsymbol{x}_{u}^{m}\le\boldsymbol{y}_{u}^{m}|\gamma^{m}=l',\boldsymbol{x}_{r}^{1:m-1},\boldsymbol{x}_{u}^{1:m-1}\le\boldsymbol{y}_{u}^{1:m-1})$이고, 상태 천이 확률 $P(\gamma^{m}=l|\gamma^{m-1}=l')$을 이용하여, 식 (7)의 첫 번째 항 $P(\gamma^{m}=l|\boldsymbol{x}_{r}^{1:m-1},\boldsymbol{x}_{u}^{1:m-1}\le\boldsymbol{y}_{u}^{1:m-1})$을 다음과 같이 재귀적으로 계산할 수 있다.

$$P(\gamma^{m}=l|\boldsymbol{x}_{r}^{1:m-1},\boldsymbol{x}_{u}^{1:m-1}\le\boldsymbol{y}_{u}^{1:m-1}) = \begin{cases} P(\gamma^{m}=l), & m=1, \\ \sum_{l'}P(\gamma^{m}=l|\gamma^{m-1}=l') & \\ \quad P(\gamma^{m-1}=l'|\boldsymbol{x}_{r}^{1:m-1},\boldsymbol{x}_{u}^{1:m-1}\le\boldsymbol{y}_{u}^{1:m-1}), & m\ge 2. \end{cases} \quad (8)$$

또, 식 (7)의 두 번째 항을 $p(\boldsymbol{x}_{r}^{m},\boldsymbol{x}_{u}^{m}\le\boldsymbol{y}_{u}^{m}|\gamma^{m}=l)$로 근사화하여 사용한다.

4. 실험 결과

제안한 손실 특징 복원 방법의 성능을 평가하기 위해 16kHz로 표본화된 DARPA Resource Management (RM) 데이터베이스[8]와 HMM Toolkit[9]을 사용하여 인식 실험을 하였다. 복원은 로그 멜-주파수 에너지 특징 영역에서 수행하였다. 25-ms 길이의 해밍(Hamming) 윈도우를 사용하여 매 10-ms마다 고속 푸리에 변환을 하고 출력 계수 크기에 제곱을 하여 삼각 멜 필터 뱅크를 적용하여 멜-주파수 에너지 특징을 추출한 뒤 로그 연산을 한다. 복원된 로그 멜-주파수 에너지 특징은 이산 코사인 변환(discrete cosine transform: DCT)을 통해 13차 켑스트럼 특징으로 변환하였다. 변환된 특징의 속도(delta), 가속도(acceleration)를 계산해 39차원 최종 특징 벡터를 추출한다. 깨끗한 음성으로 추출된 39차원 특징 벡터들을 이용하여 인식 성능을 평가할 음향 모델을 학습한다. RM 데이터베이스의 "tri-phone" 음향 모델은 단방향 은닉 마르코프 모델을 사용하여 한 상태당 관찰확률분포는 7개의 가우시안으로 이루어진 정규혼합분포를 가정하였고, 총 3개의 상태로 모델링하였다. 또, 묵음(silence)과 중지(short pause)는 한 상태당 6개의 가우시안으로 이루어져 있고 각각 3개, 1개 상태로 모델링하였다.

깨끗한 3,990개의 학습 데이터를 이용해 기존과 제안 방식의 손실 특징 복원을 위한 정규혼합분포, 다방향 은닉 마르코프 선험적 모델을 128개의 성분 또는 상태를 갖도록 각각 학습하였다. <그림 2.>는 다방향 은닉 마르코프 모델의 상태 천이 확률로 이루어진 행렬을 보여준다. 상태의 개수는 128개이며 특히, 임의의 한 상태에서 같은 상태로의 상태 천이 확률은 다른 상태로의 천이 확률보다 큰 것을 보여주는데 이것은 음성 고유의 시간적 의존성을 의미한다.

300문장의 테스트 데이터는 각 실험조건별로 RM 데이터베이스의 해당 깨끗한 음성을 NOISEX[10]와 Sound Jay[11]의 지하철, 배블, 자동차, 전시회 잡음과 입력 신호 대 잡음비(SNR)에 맞추어 혼합하였다. 혼합 테스트 데이터에 대응하는 깨끗한 음성 데이터를 사용하여 각 프레임-주파수 영역별로 SNR을 계산하고 임계값 이하의 비신뢰 성분을 다음과 같이 정의할 수 있다.

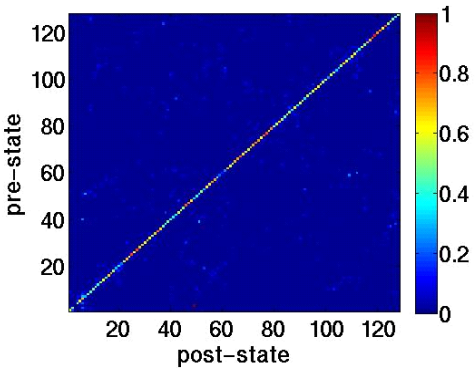

그림 2. 깨끗한 음성 데이터를 이용해 학습된 다방향 은닉 마르코프 모델의 상태 천이 확률 행렬.

$$b_k^m = \begin{cases} 1, & \text{if } 20\log\dfrac{x_k^m}{y_k^m - x_k^m} \geq Th, \\ 0, & \text{otherwise.} \end{cases} \quad (9)$$

여기서 b_k^m, y_k^m, 그리고 x_k^m는 각각 k번째 멜-주파수, m번째 프레임에서 0 또는 1의 값을 갖는 이진 마스크, 혼합 테스트 데이터와 그에 대응하는 깨끗한 음성 데이터의 로그 멜-주파수 에너지 스펙트로그램이다. Th는 입력 성분의 신뢰를 결정하는 임계값인데 각 잡음 조건에 대해 입력 SNR이 0 dB에서 최고의 성능을 보여주는 값을 실험적으로 찾아 다른 입력 SNR을 갖는 혼합 테스트 데이터에 적용하였으며, 본 실험에서는 -3 ~ -5 범위에 있었다. 이진 마스크에서 0 값을 갖는 성분을 비신뢰 성분으로 한다.

<그림 3>은 입력 신호 대 잡음비 -5 dB의 배블 잡음 환경에서 'most' 단어에 대해 각각 기존의 클러스터 기반 복원 방법과 제안한 방법에 의해 복원된 로그 멜-주파수 에너지 계수들을 나타낸다. 비교를 위해서, 대응하는 깨끗한 음성의 로그 멜-주파수 에너지 스펙트로그램과 이진 마스크를 같이 보였다. (특히 60 ~ 70번째 프레임 사이에서 명확히 드러나는 것처럼) 기존의 클러스터 기반 손실 특징 복원 방법이 비신뢰 성분에 대해 잘못된 값을 추정하고 있을 때, 제안한 방법은 시간적으로 일관적이고 깨끗한 데이터와 더욱 유사한 것을 볼 수 있다. 같은 마스크를 사용했지만 은닉 마르코프 선험적 모델의 시간적 의존성을 이용하여 원래의 스펙트로그램과 더 유사한 복원 결과를 얻을 수 있음을 알 수 있다. 실제로, 기존 방법에 의해 복원된 스펙트로그램은 'vessel'로 잘못 인식되는 결과를 얻었다.

실험 데이터에 대한 단어 인식률을 <표 1>에 요약하였다. 기존 방법과 제안한 방법 모두 상당한 단어 인식률 개선이 있었다. 또, 평균적으로 제안한 방법은 입력 SNR이 높은 10 ~ 20 dB에서 기존 방법과 유사한 단어 인식률을 나타낸 반면에, 입력 SNR이 낮은 -5 ~ 5 dB에서 기존 방법보다 단어 인식률이 높음을 알 수 있다.

5. 결론

본 논문에서는 은닉 마르코프 모델을 로그 멜-주파수 에너지 스펙트로그램의 선험적 모델로 하여 기존의 주파수 성분간 상관 뿐 아니라 음성의 시간적 의존성을 이용한 손실 특징 복

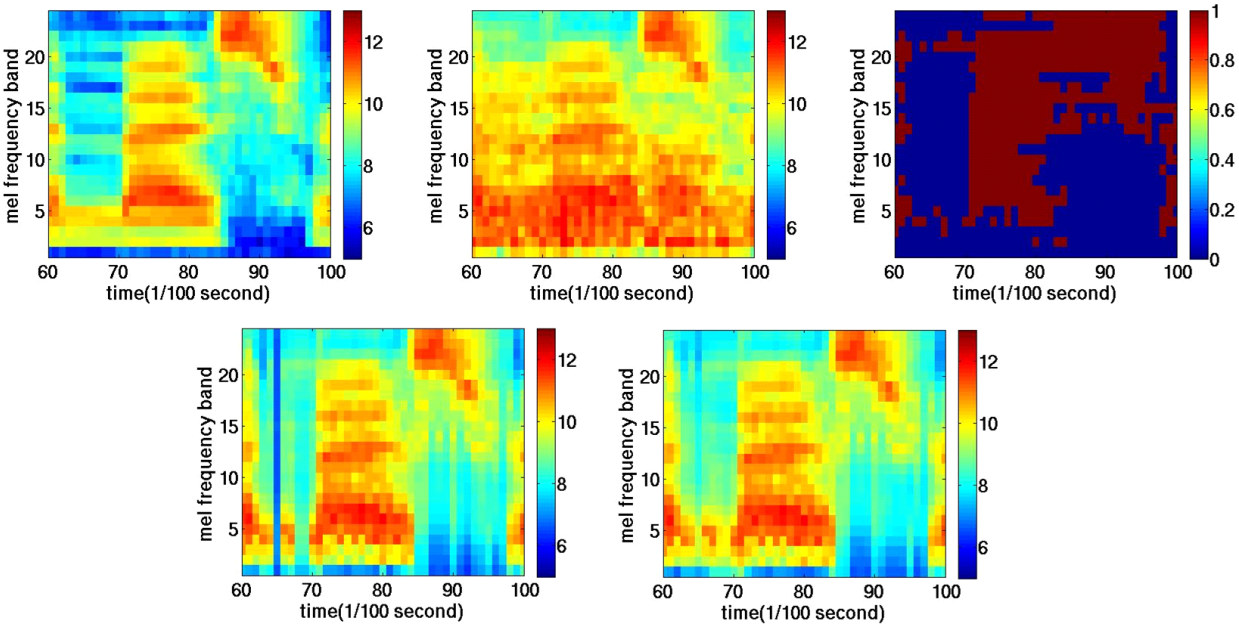

그림 3. -5 dB 입력 신호 대 잡음비, 배블 잡음 환경에서 'most' 단어에 대해 로그 멜-주파수 에너지 스펙트로그램, 또는 이진 마스크. 왼쪽 위부터 시계방향으로 각각 깨끗한 음성, 잡음에 의해 왜곡된 혼합 음성, 이진 마스크, 기존 정규혼합분포를 이용하여 복원, 은닉 마르코프 모델을 이용하여 복원한 멜-주파수 에너지 스펙트로그램이다.

표 1. 지하철, 배블, 자동차, 전시회 잡음과 혼합된 테스트 데이터에 대한 단어 인식률 (%)

SNR(dB)	-5	0	5	10	15	20
지하철 환경						
잡음 신호(%)	1.17	9.76	33.62	61.07	79.46	88.17
정규혼합모델 복원 신호(%)	74.97	86.72	90.32	92.54	**94.14**	94.26
은닉 마르코프 모델 복원 신호(%)	**78.68**	**87.31**	**91.02**	**92.7**	93.83	**94.46**
배블 환경						
잡음 신호(%)	0	0.62	10.39	39.2	71.73	85.47
정규혼합모델 복원 신호(%)	53.26	75.28	85.16	90.24	92.39	**93.60**
은닉 마르코프 모델 복원 신호(%)	**58.02**	**78.29**	**86.41**	**90.67**	**92.7**	93.24
자동차 환경						
잡음 신호(%)	9.84	21.91	45.76	66.73	82.7	89.85
정규혼합모델 복원 신호(%)	75.87	85.83	**90.55**	**92.46**	**93.32**	**94.26**
은닉 마르코프 모델 복원 신호(%)	**80.83**	**86.53**	**90.55**	92.11	93.05	94.06
전시회 환경						
잡음 신호(%)	2.89	3.94	16.17	48.5	76.61	88.01
정규혼합모델 복원 신호(%)	54.24	75.13	**86.8**	89.26	92.15	**92.85**
은닉 마르코프 모델 복원 신호(%)	**62.2**	**79.66**	85.86	**89.57**	**91.68**	92.58
평균						
잡음 신호(%)	3.10	7.52	21.88	45.19	68.61	84.15
정규혼합모델 복원 신호(%)	57.24	76.53	85.83	**89.89**	**92.10**	**93.14**
은닉 마르코프 모델 복원 신호(%)	**63.05**	**79.22**	**86.32**	89.67	91.89	93.08

원 방법을 제안하였다. 제안한 방법은 기존 방법으로 추정된 비신뢰 성분 값에 비해 시간적으로 일관된 스펙트로그램을 복원하였고 대체적으로 더 높은 인식 성능을 나타내었다.

참고문헌

[1] Acero, A. (1990). Acoustic and Environmental Robustness in Automatic Speech Recognition, PhD. thesis, Dept. of Electrical and Computer Engineering, Carnegie Mellon University, PA.

[2] Raj, B. & Stern, R. M. (2005). Missing feature approaches in speech recognition, *IEEE Signal Processing Magazine,* vol. 22, 101-116.

[3] Peinado, A. M., Sanchez, V., Segura, J. C., & Perez-Cordoba, J. L. (2001). MMSE-based Channel Mitigation for Distributed Speech Recognition, *Proc. EUROSPEECH,* 2707-2710

[4] Peinado, A. M., Sanchez, V., Perez-Cordoba, J. L., Segura, J. C., & Rubio, J. (2002). HMM-Based Methods for Channel Error Mitigation in Distributed Speech Recognition, *Proc. ICSLP02*, 2205-2208.

[5] Borgström, B. J. & Alwan A. (2010). HMM-based reconstruction of unreliable spectrographic data for noise robust speech recognition, *IEEE Transactions on Audio, Speech, and Language Processing,* vol. 18, 1612-1623.

[6] Huang, X., Acero, A., & Hon, H.-W. (2001). *Spoken language processing: a guide to theory, algorithm, and system development,* NJ: Prentice-Hall.

[7] Cho, J.-W. & Park, H.-M. (2013). An efficient HMM-based feature enhancement method with filter esimation for reverberant speech recognition, *IEEE Signal Processing Letter,* vol. 20, 1199-1202.

[8] Price, P., Fisher, W.M., Bernstein, J., Pallet, D.S.(1988). The DARPA 1000-Word Resource Management Database for Continuous Speech Recognition, *Proc. IEEE ICASSP*, 651-654

[9] Young, S., Evermann, G., Gales, M., Hain, T., Kershaw, D., Moore, G., Odell, J., Ollason, D., Povey, D., Valtchev, V., & Woodland, P. (2006). *The HTK book*, Cambridge, UK: Cambridge University Press.

[10] Varga, A., Steeneken, H.J. (1993) Assessment for automatic speech recognition: 2. In: NOISEX 1992: A Database and anExpeiment to Study the Effect of Additive Noise on Speech Recognition Systems. *Speech Comm.*, vol. 12, 247-251.

[11] Sound Jay. www.soundjay.com.

• **조지원 (Cho, Ji-Won)**
서강대학교 전자공학과
서울시 마포구 백범로 35 (신수동)
Tel: 02-711-8916
Email: jiwonn85@sogang.ac.kr
관심분야: 음성 전처리, 음성인식
현재 전자공학과 대학원 박사과정 재학 중

• **박형민 (Park, Hyung-Min)** 교신저자
서강대학교 전자공학과
서울시 마포구 백범로 35 (신수동)
Tel: 02-705-8916
Email: hpark@sogang.ac.kr
관심분야: 음성신호처리, 음성인식, 잡음제거 등
2007~현재 전자공학과 부교수