

잡음 환경에 효과적인 음성 인식을 위한 Gaussian mixture model deep neural network 하이브리드 기반의 특징 보상

A study on Gaussian mixture model deep neural network hybrid-based feature compensation for robust speech recognition in noisy environments

윤기무,¹ 김우일[†]

(Ki-mu Yoon¹ and Wooil Kim^{1†})

¹인천대학교 컴퓨터공학부

(Received September 27, 2018; accepted November 22, 2018)

초 록: 본 논문에서는 잡음 환경에서 효과적인 음성인식을 위하여 GMM(Gaussian Mixture Model)-DNN(Deep Neural Network) 하이브리드 기반의 특징 보상 기법을 제안한다. 기존의 GMM 기반의 특징 보상에서 필요로 하는 사후 확률을 DNN을 통해 계산한다. Aurora 2.0 데이터를 이용한 음성 인식 성능 평가에서 본 논문에서 제안한 GMM-DNN 하이브리드 기법이 기존의 GMM 기반 기법에 비해 Known, Unknown 잡음 환경에서 모두 평균적으로 우수한 성능을 나타낸다. 특히 Unknown 잡음 환경에서 평균 오류율이 9.13 %의 상대 향상률을 나타내고, 낮은 SNR(Signal to Noise Ratio) 잡음 환경에서 상당히 우수한 성능을 보인다.

핵심용어: 특징 보상, 심층 신경망, 가우시안 혼합 모델, 음성 인식, 잡음 환경

ABSTRACT: This paper proposes an GMM(Gaussian Mixture Model)-DNN(Deep Neural Network) hybrid-based feature compensation method for effective speech recognition in noisy environments. In the proposed algorithm, the posterior probability for the conventional GMM-based feature compensation method is calculated using DNN. The experimental results using the Aurora 2.0 framework and database demonstrate that the proposed GMM-DNN hybrid-based feature compensation method shows more effective in Known and Unknown noisy environments compared to the GMM-based method. In particular, the experiments of the Unknown environments show 9.13 % of relative improvement in the average of WER (Word Error Rate) and considerable improvements in lower SNR (Signal to Noise Ratio) conditions such as 0 and 5 dB SNR.

Keywords: Feature compensation, Deep neural network, Gaussian mixture model, Speech recognition, Noisy environments

PACS numbers: 43.72.Bs, 43.72.Ne

1. 서 론

지난 1980년대 이후로 음성 인식에 관한 연구가 활발히 진행되어 왔으며, 최근 Google의 Voice Search,

Apple의 Siri 등의 성공적인 출시와 더불어 일반 사용자의 관심이 증대되고 있다. 스마트폰 사용을 위한 소프트웨어, 자동차의 네비게이션과 내부 조작용을 위한 음성 명령 장치, 콜 센터에서 자동 응답 장치, 게임 및 오락을 위한 장치, 로봇 인터페이스 등에 음성 인식을 이용한 다양한 종류의 애플리케이션들이 상용화되어 등장하고 있다. 특히 최근에는 Google 글래스, 스마트 와치와 같은 웨어러블 컴퓨터의 상용화

[†]Corresponding author: Wooil Kim (wikim@inu.ac.kr)
Department of Computer Science & Engineering, Incheon National University, 119 Academy-ro, Yeonsu-gu, Incheon, Republic of Korea
(Tel: 82-32-835-8459, Fax: 82-32-835-0780)

“이 논문은 2018년도 한국음향학회 음성통신 및 신호처리 학술대회에서 발표하였던 논문임.”

에 따라 보다 정확한 음성 인식 기능이 급격히 요구되고 있다. 하지만 아직은 그 인식 성능이 일반 사용자의 기대에 미치지 못하는 실정이다.

인식 성능이 저하되는 가장 큰 원인 중 하나는 음성 인식 시스템에 장착되어지는 음향 모델을 훈련하는 환경과 실제 시스템을 적용하는 환경이 음향학적 측면에서 불일치 한다는 점이다. 일반적으로 음성 인식 시스템을 위한 음향 모델을 훈련하기 위해 사용하는 대용량 음성 데이터베이스는 잡음이 없는 깨끗한 환경에서 수집하게 된다. 따라서 훈련된 음향 모델은 깨끗한 발화 환경만을 표현하게 되어 실제 잡음 환경에서는 그 차이로 인한 오류가 불가피해지고 이는 음성 인식 성능 하락의 주요한 원인이 된다.

이러한 음향학적 불일치를 줄이고 음성 인식 성능 향상을 위해 다양한 연구가 진행되어 왔다.^[1-8] 이러한 연구는 두 가지 측면으로 나눌 수 있는데, 하나는 음성 인식 시스템의 전처리 단계에서 음성 신호로부터 잡음을 제거하고 음성을 향상시키거나, 잡음에 강인한 음성 특징을 추출하거나, 또는 특징 영역에서 잡음을 제거하거나 보상하는 방법이다. 이러한 기법에는 주파수 차감법,^[1] 캡스트럼 평균 정규화, 다양한 종류의 특징 보상 기법^[4,5] 등이 포함된다. 두 번째 접근 방법은 이미 훈련되어진 음향 모델을 새로운 잡음 환경과 일치하도록 적응 해주는 기법이다. 최대 사후 확률 예측법,^[6] 최대 우도 선형 회귀 기법,^[7] 병렬 모델 결합 기법^[8] 등이 이 접근 방법에 속한다. 최근에는 심층 신경망을 활용한 기법들이 소개되었다.^[9,10]

본 논문에서는 잡음 환경에 강인한 음성 인식을 위한 효과적인 특징 보상 기법을 제안하고 그 성능을 평가한다. 기존의 가우시안 혼합 모델(Gaussian Mixture Model, GMM) 기반의 특징 보상 기법에서는 깨끗한 음성을 GMM 형태로 모델링한 후, 훈련 데이터, 입력 음성 등을 이용하여 오염 음성 GMM을 생성한다. 깨끗한 음성 GMM과 오염 음성 GMM의 통계적 차이를 이용하여 입력된 음성으로부터 깨끗한 음성을 복구하게 된다. 본 연구에서는 기존의 GMM 기반의 특징 보상 기법에서 보상 과정에서 필요한 사후확률 계산에 DNN을 사용함으로써 그 성능을 향상시키고자 한다. 성능 평가를 위해 Aurora 2.0 평가 프레임워크와 데이터베이스를 사용하였고, Aurora 2.0

에 포함되어 있는 테스트 데이터와 새로 생성한 테스트 데이터에 대해 성능 평가를 실시하여 known 잡음 환경과 unknown 잡음 환경에서의 인식 성능을 관찰하였다.

본 논문은 다음과 같이 구성된다. II장에서는 본 연구의 기반이 되는 기존의 데이터 유도 방식의 가우시안 혼합 모델 기반의 특징 보상 기법을 기술한다. III장에서는 본 논문에서 제안하는 GMM-DNN 하이브리드 기반의 특징 보상 기법을 설명한다. IV장에서는 Aurora 2.0 데이터베이스를 이용하여 음성 인식 성능 평가를 실시한 결과를 기술하고, V장에서 논문의 결론을 맺는다.

II. 데이터 유도 방식의 가우시안 혼합 모델(GMM) 기반의 특징 보상 기법

이 장에서는 본 논문에서 베이스라인 기법으로 채용한 GMM을 이용한 특징 보상 기법에 대하여 설명한다.^[4] GMM 기반의 특징 보상 기법에서는 GMM으로 표현되는 깨끗한 음성 모델과 오염된 음성 모델의 통계적인 차이를 이용하여 입력된 오염 음성 특징 벡터를 깨끗한 음성 신호로 복구한다. 오염 음성 모델은 깨끗한 음성 모델을 기반으로 하여 훈련 데이터를 이용하여 생성할 수도 있고,^[4] 입력된 음성 데이터로부터 추정할 수도 있으며,^[4] 모델 결합 법칙을 이용하여 잡음 데이터를 이용하여 생성할 수도 있다.^[5] 본 논문에서는 훈련 데이터를 이용하여 사전에 오염 음성 모델을 생성하는 방식인 데이터 유도 기반의 특징 보상 기법을 사용한다.

GMM 기반 특징 보상 기법 과정의 첫 번째 단계로 깨끗한 훈련용 음성 데이터를 이용하여 다음과 같이 GMM 모델을 얻는다.

$$p(\mathbf{x}) = \sum_{k=1}^K \omega_k N(\mathbf{x}; \boldsymbol{\mu}_{\mathbf{x}, k}, \boldsymbol{\Sigma}_{\mathbf{x}, k}). \quad (1)$$

위 식에서 \mathbf{x} 는 깨끗한 음성의 캡스트럼 특징 벡터를 나타내며, K 개의 가우시안 요소로 이루어진 혼합 모델임을 의미한다. 오염된 음성 모델 역시 Eq. (2)와 같은 GMM으로 가정하며, 깨끗한 음성 모델과 오염된

음성 모델의 각 가우시안 요소의 평균 벡터와 분산 행렬 사이에 Eqs. (3)과 (4)와 같은 관계가 있음을 가정한다.

$$p(\mathbf{y}) = \sum_{k=1}^K \omega_k N(\mathbf{y}; \boldsymbol{\mu}_{\mathbf{y},k}, \boldsymbol{\Sigma}_{\mathbf{y},k}). \quad (2)$$

$$\boldsymbol{\mu}_{\mathbf{y},k} = \boldsymbol{\mu}_{\mathbf{x},k} + \mathbf{r}_k. \quad (3)$$

$$\boldsymbol{\Sigma}_{\mathbf{y},k} = \boldsymbol{\Sigma}_{\mathbf{x},k} + \mathbf{R}_k. \quad (4)$$

평균 벡터와 분산 행렬에 대한 보정 요소인 \mathbf{r}_k 과 \mathbf{R}_k 는 음성 인식이 이루어지는 잡음 환경과 동일하다고 가정된 환경에서 취득된 훈련용 음성 데이터베이스를 이용하여 얻을 수 있다. 보정 요소의 예측에는 반복적인 EM(Expectation Maximization) 알고리즘이 이용되며 평균 벡터와 분산 행렬 보정 요소에 대해 각각 다음과 같은 식을 사용하여 추정할 수 있다.

$$\tilde{\mathbf{r}}_k = \frac{\sum_{t=1}^T P(k|\mathbf{y}_t, \lambda) \mathbf{y}_t}{\sum_{t=1}^T P(k|\mathbf{y}_t, \lambda)} - \boldsymbol{\mu}_{\mathbf{x},k}. \quad (5)$$

$$\tilde{\mathbf{R}}_k = \frac{\sum_{t=1}^T P(k|\mathbf{y}_t, \lambda) [\mathbf{y}_t - \boldsymbol{\mu}_{\mathbf{x},k} - \tilde{\mathbf{r}}_k][\mathbf{y}_t - \boldsymbol{\mu}_{\mathbf{x},k} - \tilde{\mathbf{r}}_k]^T}{\sum_{t=1}^T P(k|\mathbf{y}_t, \lambda)} - \boldsymbol{\Sigma}_{\mathbf{x},k}. \quad (6)$$

Eqs. (5)와 (6)에서 $P(k|\mathbf{y}_t, \lambda)$ 는 이전 반복 단계에서 추정된 보정 요소를 이용한 오염 음성 GMM의 모델 파라미터 λ 가 주어졌을 때, 오염 음성 \mathbf{y}_t 에 대한 k 번째 가우시안 요소의 사후 확률을 나타낸다.

지금까지 설명한 과정에 의해 얻어진 깨끗한 음성 모델과 잡음에 오염된 음성 모델을 이용하여 Eq. (7)과 같이 최소 평균 제곱 오차 예측 기법에 의해 입력된 음성 특징 벡터로부터 깨끗한 음성 특징 벡터로 복구한다.

$$\tilde{\mathbf{x}}_{t,MMSE} = \mathbf{y}_t - \sum_{k=1}^K \mathbf{r}_k P(k|\mathbf{y}_t). \quad (7)$$

위 식에서 $p(k|\mathbf{y}_t)$ 는 입력 음성 \mathbf{y}_t 이 주어졌을 때 k 번째 가우시안 요소의 사후 확률을 나타낸다.

III. GMM-DNN 하이브리드 기반의 특징 보상 기법

본 논문에서는 Eq. (7)의 특징 보상 과정에서 사용되는 사후 확률을 DNN을 이용하여 계산하는 방법을 제안한다. L 개의 층을 갖는 DNN에서 l 번째 층은 다음과 같이 표현할 수 있다.

$$\mathbf{v}^l = f(\mathbf{W}^l \mathbf{v}^{l-1} + \mathbf{b}^l). \quad (8)$$

위 식에서 \mathbf{W}^l 과 \mathbf{b}^l 는 각각 l 번째 층에서의 가중치 행렬과 바이어스 벡터를 나타내며, 함수 $f(\cdot)$ 는 사용된 활성화 함수를 나타낸다. 따라서 \mathbf{v}^0 는 입력 관찰 벡터를 나타내고 \mathbf{v}^L 은 해당 심층 신경망의 최종 출력을 나타낸다. 본 논문에서 제안하는 방법에서는 입력 데이터로 기존의 GMM 기반 특징 보상 기법과 동일한 쉐프트럼 특징 벡터를 사용하고, 출력으로 오염된 음성 모델의 각 가우시안 요소의 사후 확률 값을 생성하는 DNN을 학습을 통해 얻고자 한다. 따라서 입력 층의 뉴런의 개수는 음성 특징 벡터의 사이즈와 동일하고, 출력 층의 뉴런의 개수는 가우시안 혼합 모델의 요소 개수와 동일하다.

제안하는 GMM-DNN 하이브리드 기반 특징 보상 기법의 구현을 위한 첫 단계로, II장에서 설명한 GMM 기반의 특징 보상을 위한 깨끗한 음성 GMM과 오염된 음성 GMM을 Eqs. (1)~(6)을 통하여 구한다. 심층 신경망 학습 단계에서는 각 훈련 데이터의 특징 벡터에 대해 Eq. (2)로 얻어진 오염 음성 GMM을 이용하여 사후 확률을 계산하고, 이 중 최대값을 갖는 가우시안 인덱스의 요소를 1로 하고 나머지는 0을 갖는 레이블 데이터를 생성한다. 생성된 레이블 데이터를 목표로 하여 신경망을 학습한다. 제안하는 GMM-DNN 하이브리드 기반 특징 보상 기법에서는 학습을 통해 얻어진 신경망의 가중치 행렬과 바이어스 벡터를 이용하여 다음과 같이 특징을 복구할 수 있다.

$$\tilde{\mathbf{x}}_t = \mathbf{y}_t - \sum_{k=1}^K \mathbf{r}_k f(\mathbf{y}_t; \mathbf{W}; \mathbf{b}). \quad (9)$$

IV. 실험 및 결과

객관적인 성능 평가를 위해서 Aurora 2.0에서 제공하는 평가 방식을 따랐다. Aurora 2.0의 평가 방식의 주요 특징은 다음과 같다.^[11]

- 영어 음성, 연속숫자음 인식, 11단어+목음 구간 (Silence)+짧은 휴지(Short pause)
- ETSI(European Telecommunications Standards Institute) 표준 방식의 특징 추출^[12]
- 13차 static 특징(c1~c12+c0) 추출 후 인식 단에서 미분계수 추출(총 39차)
- 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. 잡음 환경에서의 성능 향상을 위해 캡스트럼 평균 정규화 기법을 공통으로 적용하였다.

객관적인 성능 평가를 위하여 기존의 대표적인 전처리 알고리즘과 성능을 비교하였다. 대표적 전처리 알고리즘으로 가장 일반적으로 사용되는 주파수 차감법 기법을 선택하였으며 배경 잡음을 추정하기 위해 250 msec의 시간 지연을 갖는 최소 통계 기법을 적용하였다.^[13] 성능 비교를 위해 기존의 대표적인 특징 보상 기법인 VTS(Vector Taylor Series) 기반 알고리즘을 평가하였다.^[4] VTS 기법에서는 EM 기법을 이용하여 적응적으로 잡음 성분을 추정하는 것으로 알려져 있다. 본 논문에서는 음성 인식 성능의 지표로 단어 오인식율(Word Error Rate, WER)을 사용하였다.

4.1 Known 잡음 환경 성능 평가

Table 1과 Fig. 1은 “Known” 잡음 환경에서 기존의 GMM 기반의 특징 보상 기법과 본 논문에서 제안하는 GMM-DNN 하이브리드 기반의 특징 보상 기법의 성능 평가 비교를 나타낸다. 이 실험에서는 성능 평가를 위해 Aurora 2.0에 포함되어 있는 SetA 테스트 데이터를 사용하였다. SetA 테스트 데이터에는 Subway,

Babble, Car, Exhibition의 4 종류의 잡음을 0 dB, 5 dB, 10 dB, 15 dB, 20 dB SNR에 따라 부가적으로 오염시켜 생성한 잡음 오염 음성이 포함되어 있다. Table 1의 결과는 각 테스트 잡음 환경에서 모든 SNR 환경에 대한 WER의 평균을 구한 결과이다.

GMM 기반 특징 보상 기법과 GMM-DNN 하이브리드 특징 보상 기법의 모델 훈련을 위하여 Aurora 2.0에서 제공하는 Multi-condition training 데이터를 사용하였다. Multi-condition training 데이터는 8,840개의 깨끗한 훈련용 음성 데이터를 SetA 테스트 데이터와 동일한 잡음 데이터를 이용하여 clean 환경, 0 dB, 5 dB, 10 dB, 15 dB, 20 dB SNR에 맞게 생성한 훈련 데이터로서 테스트 환경을 포함하고 있다. 따라서 테스트용 잡음 환경에 대해 사전에 알고 있는 “Known” 조건으로 간주할 수 있다.

GMM 기반 특징 보상 기법에서는 128개의 가우시안 요소를 사용하여 깨끗한 음성 GMM을 생성하고

Table 1. Recognition performance in “Known” noisy environments as average over all SNRs: 0 dB, 5 dB, 10 dB, 15 dB, and 20 dB (WER, %).

	Subway	Babble	Car	Exhib.	Avg.
No processing	36.12	31.25	37.08	37.60	35.51
SS	24.50	25.38	24.49	27.92	25.57
VTS	21.09	19.81	25.98	24.29	22.79
GMM	18.38	23.07	25.28	19.41	21.53
GMM-DNN	17.67	23.26	23.77	19.68	21.10

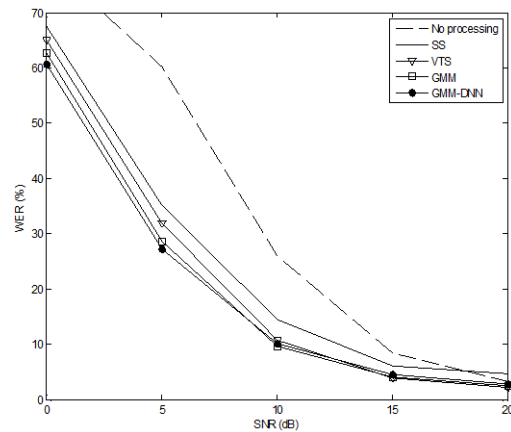


Fig. 1. Recognition performance in “Known” noisy environments at different SNRs as average over all environments: Subway, Babble, Car and Exhibition (WER, %).

II장에서 설명한 EM 기반의 훈련 과정을 통하여 오염 음성 GMM을 생성하였다. GMM-DNN 하이브리드 기반 특징 보상 기법에서는 2개의 은닉 층을 갖는 심층 신경망을 사용하였고 다양한 실험을 통해 각각 200개와 400개의 뉴런을 갖는 은닉층을 최종적으로 사용했다. 활성화 함수로는 ReLU 함수를 사용하고, 출력 층에서는 사후 확률 형태의 값을 얻기 위하여 Softmax 함수를 사용하였다. 학습 데이터 양의 10% 크기를 배치 사이즈로 하였고 0.01의 학습률을 사용하였다. Adam 최적화 알고리즘을 사용하여 교차 엔트로피를 기준으로 학습하였다.

Table 1에서 확인할 수 있는 것과 같이 본 논문에서 제안하는 GMM-DNN 하이브리드 기반의 특징 보상 기법이 기존의 GMM 기반의 특징 보상과 비교하여 Subway 잡음과 Car 잡음 환경에서 낮은 WER을 보이고, 4종류 환경에 대한 전체 평균 WER도 다소 앞서는 것으로 관찰되었다. Fig. 1은 Table 1과 동일한 실험 결과를 각 SNR 별로 나타낸 결과이다. Fig. 1에서는 낮은 SNR (0 dB, 5 dB)에서 제안한 GMM-DNN 하이브리드 기법이 기존의 GMM 기반의 기법에 비하여 우수한 성능을 보이는 것을 확인할 수 있다.

4.2 Unknown 잡음 환경 성능 평가

Table 2와 Fig. 2는 특징 보상 기법 모델 훈련을 위해 사용한 훈련용 오염 음성 데이터에 테스트 환경의 잡음이 포함되지 않은 “Unknown” 잡음 환경에 대한 성능 평가를 나타낸다. Unknown 잡음 환경 성능 평가를 위하여 Aurora 2.0에 SetA 테스트 데이터에 포함된 깨끗한 음성 데이터와 다른 리소스의 잡음 샘플을 이용하여 새로운 잡음 오염 음성을 생성하였다. 잡음 데이터로는 NOISEX92에 포함되어 있는 Factory, Car, Babble 잡음을 사용하였고, 한국 가요의 전주부분에서 취득한 오디오 데이터를 이용하여 배경음악(Music) 잡음으로 사용하였다. 이와 같은 잡음을 이용하여 0 dB, 5 dB, 10 dB, 15 dB, 20 dB SNR에 맞게 부가하여 인공적으로 잡음 오염 음성을 생성하였다.

Table 2에서 확인할 수 있는 것과 같이 Music 잡음 환경을 제외한 3가지 잡음 환경에서 제안한 GMM-DNN 하이브리드 기반 특징 보상 기법이 기존의 GMM 기

Table 2, Recognition performance in “Unknown” noisy environments as average over all SNRs: 0 dB, 5 dB, 10 dB, 15 dB, and 20 dB (WER, %).

	Factory	Babble2	Car2	Music	Avg.
No processing	24.24	26.75	26.82	22.92	25.18
SS	13.82	20.18	15.41	24.62	18.51
VTS	14.16	16.26	16.73	18.30	16.36
GMM	14.85	19.94	17.17	16.85	17.20
GMM-DNN	12.23	18.95	14.11	17.24	15.63

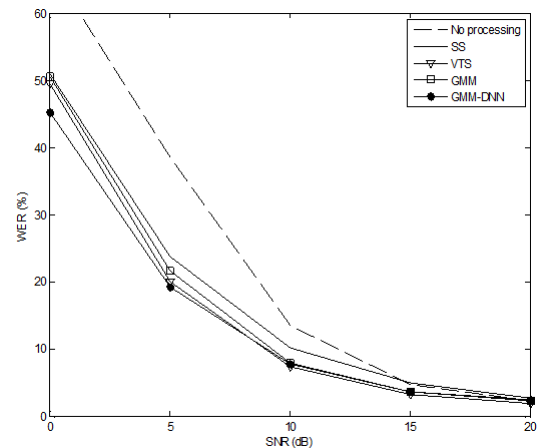


Fig. 2. Recognition performance in “Unknown” noisy environments at different SNRs as average over all environments: Factory, Babble2, Car2 and Music (WER, %).

반의 특징 보상 기법에 비해 상당히 우수한 성능을 나타내는 것을 알 수 있다. 전체 잡음 환경에 대한 평균 WER로 기존의 GMM 기법은 17.20%, 제안한 GMM-DNN 기법은 15.63%를 나타내어 9.13%의 상대 향상률을 나타냈다. Fig. 2에서는 각 SNR에서 4종류 잡음의 평균 성능을 비교한다. 앞의 실험에서와 유사하게 낮은 SNR 환경(0 dB, 5 dB)에서 제안한 GMM-DNN 기반의 특징 보상 기법이 기존의 GMM 기반 특징 보상에 비해 상당히 우수한 성능을 나타낸다.

이상과 같은 결과는 본 논문에서 제안하는 GMM-DNN 하이브리드 기반 특징 보상 기법에 사용된 심층 신경망의 비선형 구조가 사후 확률 계산을 일반화하는 능력이 GMM 기반 방식보다 우수하기 때문에 낮은 SNR과 Unknown 잡음 환경과 같이 훈련 데이터베이스가 적게 반영된 테스트 환경에 보다 효과적인 임을 입증하는 것으로 판단된다.

V. 결 론

본 논문에서는 잡음 환경에서 효과적인 음성인식을 위하여 GMM-DNN 하이브리드 기반의 특징 보상 기법을 제안하였다. 기존의 GMM 기반의 특징 보상에서 필요로 하는 사후 확률을 DNN 을 통해 계산하였다. Aurora 2.0 데이터를 이용한 음성 인식 성능 평가에서 본 논문에서 제안한 GMM-DNN 하이브리드 기법이 기존의 GMM 기반 기법에 비해 Known, Unknown 잡음 환경에서 모두 평균적으로 우수한 성능을 나타냈다. 특히 Unknown 잡음 환경에서 평균 오류율이 9.13%의 상대 향상률을 나타냈고, 낮은 SNR 잡음 환경에서 현저히 우수한 성능을 보임을 확인하였다.

감사의 글

본 논문은 인천대학교 2013년 자체연구비 지원에 의하여 연구되었음.

References

1. S. F. Boll, "Suppression of acoustic noise in speech using spectral subtraction," Proc. IEEE Trans. on Acoustics, Speech and Signal, **27**, 113-120 (1979).
2. Y. Ephraim and D. Malah, "Speech enhancement using minimum mean square error short time spectral amplitude estimator," Proc. IEEE Trans. on Acoustics, Speech and Signal, **32**, 109-1121 (1984).
3. J. H. L. Hansen and M. Clements, "Constrained iterative speech enhancement with application to speech recognition," Proc. IEEE Trans. on Signal, **39**, 795-805 (1991).
4. P. J. Moreno, B. Raj, and R. M. Stern, "Data-driven environmental compensation for speech recognition: a unified approach," Speech Communication, **24**, 267-285 (1998).
5. W. Kim and J. H. L. Hansen, "Feature compensation in the cepstral domain employing model combination," Speech Communication, **51**, 83-96 (2009).
6. J. L. Gauvain and C. H. Lee, "Maximum a posteriori estimation for multivariate gaussian mixture observations of markov chains," Proc. IEEE Trans. on Speech and Audio, **2**, 291-298 (1994).
7. 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).
8. M. J. F. Gales and S. J. Young, "Robust continuous speech recognition using parallel model combination," Proc. IEEE Trans. on Speech and Audio, **4**, 352-359 (1996).
9. J. Du, L.-R. Dai, and Q. Huo, "Synthesized stereo mapping via deep neural networks for noisy speech recognition," ICASSP 2014, 1764-1768 (2014).
10. K. Han, Y. He, D. Bagchi, E. Fosler-Lussier, and D. Wang, "Deep neural network based spectral feature mapping for robust speech recognition," Interspeech 2015, 2484-2488 (2015).
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, Sep. 2000.
12. ETSI ES 201 108, ETSI standard document, v1.1.2(2000-04), Feb. 2000.
13. R. Martin, "Spectral Subtraction Based on Minimum Statistics," EUSIPCO-94, 1182-1185 (1994).

저자 약력

▶ 윤 기 무 (Ki-mu Yoon)



2018년 2월: 인천대학교 수학과 학사
2018년 3월 ~ 현재: 인천대학교 컴퓨터공학부 석사과정

▶ 김 우 일 (Wooil Kim)



1996년 2월: 고려대학교 전자공학과 학사
1998년 8월: 고려대학교 전자공학과 석사
2003년 8월: 고려대학교 전자공학과 박사
2012년 8월 ~ 현재: 인천대학교 컴퓨터공학부 조교수, 부교수