

시변 잡음에 강인한 음성 인식을 위한 PCA 기반의 Variational 모델 생성 기법

김우일*

PCA-based Variational Model Composition Method for Roust Speech Recognition with Time-Varying Background Noise

Wooil Kim*

School of Computer Science and Engineering, Incheon National University, Incheon 406-772, Korea

요 약

본 논문에서는 시간에 따라 변하는 잡음 환경에 강인한 음성 인식을 위해 효과적인 특징 보상 기법을 제안한다. 제안하는 기법에서는 기존의 Variational 모델 생성 기법의 모델 정확도를 향상시키고자 PCA를 도입한다. 제안된 기법은 다중 모델을 사용하는 PCGMM 기반의 특징 보상에 적용된다. 실험 결과는 제안한 PCA 기반의 Variational 모델 생성 기법이 배경 음악 환경의 다양한 SNR 조건에서 기존의 전처리 기법에 비하여 음성 인식 성능을 향상 시키는데 우수함을 입증한다. 제안한 모델 생성 기법이 기존의 Variational 모델 생성 방법에 비해 배경 음악 환경에서 평균 12.14%의 상대적 인식 성능 향상률을 나타낸다.

ABSTRACT

This paper proposes an effective feature compensation method to improve speech recognition performance in time-varying background noise condition. The proposed method employs principal component analysis to improve the variational model composition method. The proposed method is employed to generate multiple environmental models for the PCGMM-based feature compensation scheme. Experimental results prove that the proposed scheme is more effective at improving speech recognition accuracy in various SNR conditions of background music, compared to the conventional front-end methods. It shows 12.14% of average relative improvement in WER compared to the previous variational model composition method.

키워드 : PCA, Variational 모델 생성, 시변 잡음, 특징 보상, 음성 인식

Key word : Principal component Analysis, Variational model composition, Time-varying noise, Feature compensation, Speech recognition

접수일자 : 2013. 10. 21 심사완료일자 : 2013. 11. 08 게재확정일자 : 2013. 11. 20

* **Corresponding Author** Wooil Kim(wikim@incheon.ac.kr, Tel:+82-32-835-8459)

School of Computer Science and Engineering, Incheon National University, Incheon 406-772, Korea

Open Access <http://dx.doi.org/10.6109/jkice.2013.17.12.2793>

print ISSN: 2234-4772 online ISSN: 2288-4165

©This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License(<http://creativecommons.org/licenses/by-nc/3.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.
Copyright © The Korea Institute of Information and Communication Engineering.

I. 서 론

최근 Google의 Voice Search, Apple의 Siri 등의 성공적인 출시와 더불어 음성 인식 기술에 대한 일반 사용자의 관심이 증대되고 있다. 자동차의 네비게이션과 내부 조작용을 위한 음성 명령 장치, 콜 센터에서 자동 응답 장치, 게임 및 오락을 위한 장치 등에 음성 인식을 이용한 다양한 종류의 애플리케이션들이 상용화되어 등장하고 있다. 하지만, 아직은 그 인식 성능이 일반 사용자의 기대에 미치지 못하는 실정이다. 일반적으로 음성 인식 시스템을 위한 음향 모델을 훈련하기 위해 사용하는 대용량 음성 데이터베이스는 잡음이 없는 깨끗한 환경에 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[1-9]. 이러한 연구는 두 가지 측면으로 나눌 수 있는데, 하나는 음성 인식 시스템의 전처리 단계에서 음성 신호로부터 잡음을 제거하고 음성을 향상시키는 방법이다. 이러한 기법에는 주파수 차감법 (Spectral Subtraction)[1], 켈스트럼 평균 정규화 (Cepstrum Mean Normalization, CMN), 다양한 종류의 특징 보상 (Feature Compensation) 기법 [4][5] 등이 포함된다. 두 번째 접근 방법은 이미 훈련되어진 음향 모델을 새로운 잡음 환경과 일치하도록 적응 (Adaptation) 해주는 기법이다. 최대 사후 확률 (Maximum A Posteriori, MAP) 예측법[6], 최대 우도 선형 회귀 (Maximum Likelihood Linear Regression, MLLR) 기법[7], 병렬 모델 결합 기법 (Parallel Model Combination, PMC)[8] 등이 이 접근 방법에 속한다.

본 논문에서는 시간에 따라 변하는 잡음 환경에서 효과적인 음성 인식을 위한 전처리 기법을 제안한다. 일반적으로 배경 잡음 예측 기법은 입력된 음성 신호의 비음성 구간에서 잡음 모델을 예측하여 음성 구간에 적용하거나 최소 통계 기반의 예측 기법과 같은 방법으로 모델을 예측하여 사용한다[9]. 이러한 방법은 정상적 (Stationary)이거나 매우 천천히 변화하는 배경 잡음 환경에서는 효과적일 수 있으나, 배경 음악과 같이 시간에 따라 비교적 빠르게 변하는 잡음을 효과적으로 예측하는 데에는 한계가 있다. 본 논문에서는 시간에 따라

빠르게 변하는 배경 잡음에 강인한 음성 인식을 위하여 최근 개발된 Variational 모델 생성 기법[10]을 향상시키고자 하며, 성능 향상을 위해 PCA 기법 (Principal Component Analysis) 을 적용하는 방법을 제안한다.

II. Variational 모델 생성 기법

Variational Model Composition (VMC) 기법은 입력된 음성의 지속 기간 동안 변화하는 시변 잡음을 효과적으로 모델링하기 위해 제안된 기법으로 다음과 같은 과정에 의해 이루어진다[10].

단계 1: Basis 잡음 모델 예측

입력된 음성 신호의 비음성 구간의 켈스트럼 벡터로부터 단일 가우시안 (μ, σ^2)의 Basis 잡음 모델을 예측한다. 입력된 신호의 시작 부분과 끝 부분을 비음성 구간으로 가정하거나, 음성 활성 검출 (Voice Activity Detection, VAD) 기법을 사용하여 비음성 구간을 검출할 수 있다. 이와 같이 얻어진 Basis 모델은 전체 입력 음성 신호의 평균적인 잡음 모델로 가정할 수 있다.

단계 2: Variational 요소 결정

처음 단계에서 얻어진 단일 가우시안 모델의 분산 벡터로부터 V 개의 Variational 요소를 결정한다. 가우시안 모델의 분산 벡터를 이루는 각 요소를 식 (1)과 같이 내림차순으로 정리한 후 상위 V 개의 요소의 인덱스를 Variational 요소 $\{v_1, v_2, \dots, v_V\}$ 로 정의한다.

$$\sigma_{v_1}^2 \geq \sigma_{v_2}^2 \geq \dots \geq \sigma_{v_V}^2 \quad (1)$$

단계 3: Perturbation에 의한 모델 생성

단계 2에서 결정된 Variational 요소의 평균 파라미터에 Perturbation Factor f_p 를 다음과 같이 적용함으로써 복수개의 모델을 생성한다.

$$\tilde{\mu}_i = \begin{cases} \mu_i(1 + f_p), & \text{if } i \in \{v_1, v_2, \dots, v_V\} \\ \mu_i, & \text{otherwise} \end{cases} \quad (2)$$

위 식에서 $f_p = 0, -\alpha, \text{ or } +\alpha$ 이며, α 는 실험에

의해서 결정되는 값이다. 최종적으로 3^V 개의 모델을 가지는 Variational 모델 집합 $\{\tilde{\lambda} = (\tilde{\mu}, \sigma^2)\}$ 이 생성된다. 각 모델의 분산벡터는 Basis 모델과 동일한 분산 벡터를 가지는 것으로 가정한다.

그림 1은 Variational 모델 합성 기법에 의해 생성된 다양한 종류의 잡음 모델의 평균 벡터를 로그 스펙트럼 도메인에서 보여주고 있다. 각 그림에서 점선은 기본 잡음 모델이고 실선은 기본 모델로부터 합성된 Variational 모델을 나타낸다.

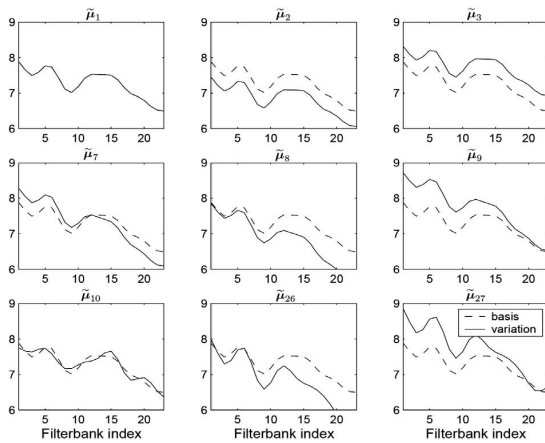


그림 1. Variational 모델 합성 기술에 의해 생성된 다양한 모델의 평균 파라미터[10]

Fig. 1 Mean parameters of variational models in log-spectral domain generated by the variational model composition method[10]

III. PCA 기반의 Variational 모델 생성 기법

기존의 Variational 모델 생성 기법에서는 Basis 모델의 분산 요소의 크기에 따라 Variational 요소를 결정했다. 본 논문에서는 보다 정확한 모델 생성을 위해 주성분 분석 (Principal Component Analysis, PCA)을 도입하여 Variational 요소를 결정하고 이를 기반으로 모델을 생성하는 새로운 기법을 제안한다. 제안한 기법은 다음과 같은 과정으로 이루어진다.

단계 1: Basis 잡음 모델 예측

기존의 방법과 같이 비음성 구간동안 캡스트럼 벡터

에 대한 단일 가우시안 모델 형태로 Basis 잡음 모델을 예측한다. 제안하는 기법에서는 PCA 적용을 위하여 완전 분산 (Full Covariance) 행렬 형태 (μ, Σ) 로 Basis 모델을 얻을 수 있다.

단계 2: PCA에 의한 Variational 요소 결정

단계 1에서 예측한 Basis 잡음 모델의 분산 행렬에 대해 고유 벡터 (Eigen Vector) $\{e_1, e_2, \dots, e_N\}$ 와 고유 값 (Eigen Value) $\{\lambda_1, \lambda_2, \dots, \lambda_N\}$ 을 구한다. 여기에서 N 은 캡스트럼 벡터의 크기를 말한다. 식 (3)와 같이 고유값을 내림차순으로 정렬하여 V 개의 주성분을 선택하고, 선택된 인덱스를 Variational 요소 $\{v_1, v_2, \dots, v_V\}$ 로 정의한다.

$$\lambda_{v_1} \geq \lambda_{v_2} \geq \dots \geq \lambda_{v_V} \tag{3}$$

단계 3: 최종 모델 생성

기존의 방법과 같이 평균 파라미터에 대해 Perturbation Factor를 적용하여 모델을 생성하는 방식은 유사하지만, 제안하는 기법에서는 식 (4)와 같이 Variational 요소의 고유벡터를 이용한다.

$$\tilde{\mu} = \mu + f_p e_{v_i}, \quad i = 1, 2, \dots, V \tag{4}$$

위 식에서 Perturbation Factor $f_p = 0, -\alpha, \text{ or } +\alpha$ 이며 α 는 기존 방법과 마찬가지로 실험에 의해 결정된다. 최종적으로 기존의 방법과 동일하게 3^V 개의 모델을 가지는 Variational 모델 집합 $\{\tilde{\lambda} = (\tilde{\mu}, \sigma^2)\}$ 이 생성된다. 제안하는 기법에서 각 모델의 분산벡터는 Basis 모델의 분산 행렬에서 대각 요소만 취한다.

그림 2는 기존의 Variational 모델 생성 방식과 본 논문에서 제안하는 PCA 기반의 모델 생성 방식을 나타낸 것이다. 그림의 예에서는 2차원의 특징 벡터를 가정하였고 회색으로 나타낸 부분은 배경 잡음의 실제 분포이다. Variational 요소는 1개를 가정하였고 이에 대한 방향을 점선으로 나타내었다. 그림 2 (a)와 같이 기존의 방법에서는 대각 행렬로 얻어지는 분산 벡터에서 크기가 큰 요소를 선택하여 모델을 생성하게 되므로 요소 2의 축 방향으로 Perturbation Factor를 적용하게

된다. 그림 2 (b)에서와 같이 제안한 기법에서는 PCA 기법을 통해 주성분 벡터를 찾게 되고 주성분의 벡터 방향으로 Perturbation Factor를 적용한다.

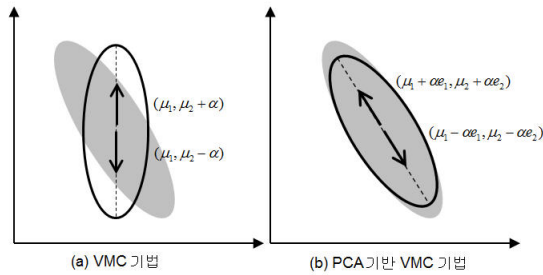


그림 2. Variational 모델 생성 예시 (a) 기존 방법, (b) 제안하는 방법

Fig. 2 Illustration of variational model composition; (a) the previous method, (b) the proposed method

IV. Variational 모델을 채용한 특징 보상 기법

본 논문에서 제안하는 기법을 통해 얻어진 복수개의 잡음 모델은 특징 보상 기법에 적용되어 음성 인식 성능을 향상시킨다. 본 논문에서는 기존에 개발된 병렬 결합된 가우시안 혼합 모델 (Parallel Combined Gaussian Mixture Model, PCGMM)을 이용한 특징 보상 기법을 이용한다[5]. PCGMM 기반의 특징 보상 기법은 음성 모델을 이용하는 특징 보상 기법의 하나로서, 가우시안 혼합 모델 (GMM)을 이용하여 표현되는 깨끗한 음성 모델과 오염된 음성 모델의 통계적인 차이를 이용하여 입력된 오염 음성 특징 벡터를 깨끗한 음성 신호로 복구한다.

PCGMM 기반의 특징 보상에서는 병렬 모델 결합 (Parallel Model Combination, PMC) 기법을 이용하여 깨끗한 음성 모델과 잡음 모델을 수학적으로 결합함으로써 잡음에 오염된 음성 모델을 생성한다[5][8]. PCGMM 기법에서는 로그-노말 가정법 (Log-normal Approximation)을 채용하여 잡음에 오염된 음성 모델을 생성한다. 이와 같은 병렬 결합 과정에 의해 얻어진 오염된 음성 모델과 깨끗한 음성 모델의 평균 벡터 사이에는 다음과 같은 Bias 변환 관계가 있음을 가정한다.

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

깨끗한 음성 모델과 잡음에 오염된 음성 모델을 이용하여 최소 평균 제곱 오차 (Minimum Mean Squared Error, MMSE) 예측 기법에 의해 입력된 오염 음성 벡터로부터 깨끗한 음성으로 복구한다.

시간에 따라 변하는 배경 잡음에 효과적으로 대응하기 위해 복수개의 잡음 모델을 이용하는 방법을 제안하였으며, 선행 연구에서 Variational 모델 생성 기법을 적용하여 잡음 환경에서 음성 인식 성능 향상에 효과적임을 입증하였다[5][10]. 다중 모델 채용 기법은 입력되는 오염 음성에 대해 각 잡음 환경 모델에 대한 사후 확률을 계산하여 이를 MMSE 예측 단계에 적용함으로써 구현된다. 각 잡음 환경 모델에 대한 d 시간 동안의 입력 오염 음성 $\mathbf{Y}_t = [\mathbf{y}_{t-d+1}, \mathbf{y}_{t-d+2}, \dots, \mathbf{y}_t]$ 의 사후 확률은 다음과 같이 계산할 수 있다.

$$p(i | \mathbf{Y}_t) = \frac{p(i)p(\mathbf{Y}_{t-1}|i)p(\mathbf{y}_t|i)}{\sum_{e=1}^E p(e)p(\mathbf{Y}_{t-1}|e)p(\mathbf{y}_t|e)} \quad (6)$$

위 식에서 $p(\mathbf{Y}_{t-1}|i) = \prod_{\tau=t-d+1}^{t-1} p(\mathbf{y}_\tau|i)$ 이고, $p(i)$ 는 각 잡음 환경 모델에 대한 사전 확률을 나타낸다. 이렇게 구한 사후 확률을 MMSE 기법에 적용하면 다음과 같이 깨끗한 음성 특징으로 복구한다.

$$\tilde{\mathbf{x}}_{t,MMSE} = \mathbf{y}_t - \sum_{e=1}^E p(e | \mathbf{Y}_t) \sum_{k=1}^K \mathbf{r}_{e,k} p(k|e, \mathbf{y}_t) \quad (7)$$

위 식에서 $\mathbf{r}_{e,k}$ 는 e 번째 환경 모델의 k 번째 가우시안 요소로부터 계산한 Bias 요소이며, $p(k|e, \mathbf{y}_t)$ 는 이에 해당하는 사후 확률을 가리킨다.

Variational 모델 생성 기법에서 생성되는 복수개의 잡음 모델과 깨끗한 음성 모델을 병렬 결합함으로써 다중의 오염 음성 모델을 생성할 수 있고, 생성된 다중의 오염 음성 모델을 본 장에서 설명한 환경 모델로 사용할 수 있다. 따라서 V 개의 Variational 요소를 가정한다면 환경 모델의 개수 E 는 3^V 가 된다. 그림 3은

Variational 모델 생성 기법을 채용한 PCGMM 기반의 특징 보상 기법을 나타낸다.

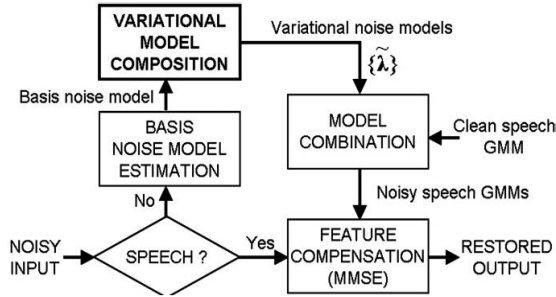


그림 3. Variational 모델 생성 기법을 채용한 PCGMM 기반의 특징 보상 기법의 블록 다이어그램[10]
 Fig. 3 Block diagram of the PCGMM-based feature compensation scheme employing the variational model composition method[10]

V. 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, 연속 숫자음 인식, 11단어 + 묵음 구간 (Silence) + 짧은 휴지(Short pause)
- ETSI(European Telecommunications Standards Institute) 표준 방식의 특징 추출[12]
- 13차 static 특징(c1~c12+로그 에너지) 추출 후 인식 단에서 미분계수 추출(총 39차) : 본 논문의 실험에서는 PCMM 구현의 편의를 위하여 로그 에너지 대신 켈프스트럼의 0차 계수를 사용하였다.
- 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. 본 논문에서는 시간에 따라 변하는 실생활 잡음 환경을 반영하기 위해 음악 오디오 샘플에 5가지의 신호 대 잡음 비 (SNR, 20, 15, 10, 5, 0dB)에

따라 부가적으로 오염시켜 새로운 테스트 데이터베이스를 제작하였다. 배경 음악은 비트와 빠르기가 다양한 유명 한국 가요 10곡의 전주 부분에서 샘플링 하였다.

우선 기존의 대표적인 전처리 알고리즘에 대해 잡음 환경에서 베이스라인 시스템의 성능을 비교하여 그 결과를 표 1에 나타내었다. 대표적 전처리 알고리즘으로 가장 일반적으로 사용되는 주파수 차감법 (Spectral Subtraction, SS)과 켈프스트럼 정규화 (Cepstral Mean Normalization, CMN) 기법을 선택하였으며 주파수 차감법에서는 배경 잡음을 추정하기 위해 250msec의 시간 지연을 갖는 최소 통계 (Minimum statistics) 기법을 적용하였다[9]. 성능 비교를 위해 기존의 대표적인 특징 보상 기법인 Vector Taylor Series (VTS) 기반 알고리즘을 평가하였다[4]. VTS 기법에서는 EM (Expectation Maximization) 기법을 이용하여 적응적으로 잡음 성분을 추정하는 것으로 알려져있다. 또한 ETSI에서 개발한 Advanced Front-End (AFE) 알고리즘도 최신 기법의 하나로 평가하였다[13]. AFE에서는 반복적인 Wiener 필터와 Blind Equalization 기법을 채용하여 잡음 환경에서 음성 인식 성능을 높이는데 월등한 성능을 가지는 것으로 알려져 있다. 본 논문에서는 음성 인식 성능의 지표로 WER (Word Error Rate, 단어 오인식율)을 사용하였다. 표 1에서 보는 것과 같이 SNR이 감소함에 따라 즉, 배경 잡음으로 작용하는 배경 음악의 강도가 커짐에 따라 오인식률이 증가하는 것을 알 수 있다. 기존의 전처리 알고리즘을 적용한 결과 ETSI AFE 알고리즘이 다른 전처리 기법에 비해 가장 효과적임을 알 수 있다.

표 1. 배경 음악 잡음 환경에서 기존의 전처리 기법을 채용한 음성 인식 성능 평가 (WER, %)

Table. 1 Speech recognition performance with the conventional methods in background music noise conditions (WER, %)

	20dB	15dB	10dB	5dB	0dB	평균
No processing	4.84	12.19	28.11	51.34	74.27	34.15
SS+CMN	3.39	7.28	15.24	29.93	54.52	22.07
VTS	4.69	8.92	16.08	31.60	54.67	23.19
AFE	2.99	6.76	11.72	25.55	44.43	18.29

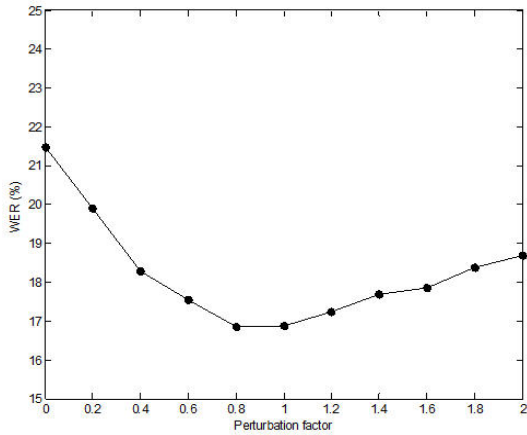


그림 4. Perturbation factor의 변화에 따른 인식 성능의 변화 (WER, %)

Fig. 4 Recognition performance as change of the perturbation factor (WER, %)

표 2. 배경 음악 잡음 환경에서 모델 생성 기법을 채용한 PCGMM 특징 보상 기법의 음성 인식 성능 평가 (WER, %)

Table. 2 Speech recognition performance with PCGMM schemes employing the model composition methods in background music noise conditions (WER, %)

	20dB	15dB	10dB	5dB	0dB	평균
PCGMM	2.34	4.54	9.50	21.47	38.01	15.17
VMC-PCGMM	2.25	3.86	8.36	20.39	35.76	14.12
PVMC-PCGMM (Relative)	2.01 (10.67)	3.24 (16.06)	7.13 (14.71)	16.87 (17.26)	32.80 (8.28)	12.41 (12.14)

본 논문에서 제안하는 PCA기반의 Variational 모델 생성 기법 (PVMC)의 Perturbation Factor를 실험으로 결정하였다. 생성된 잡음 모델은 PCGMM 기반의 특징보상 기법의 잡음 모델로 이용되어 음성 인식 시스템의 전처리로 적용된 후 음성 인식 평가를 하였다. 그림 4에서 보는 것과 같이 Perturbation Factor를 0에서 2까지 0.2씩 증가해 가며 SNR이 5dB인 잡음 환경에 대한 테스트를 하였다. Perturbation Factor가 0인 조건은 단일 가우시안 모델을 채용한 기본 PCGMM 기법과 동일한 실험을 의미한다. 실험 결과 Perturbation Factor가 0.8과 1.0에서 가장 좋은 성능을 나타내었고, 본 논문에서는 0.8을 모든 SNR 환경에도 공통적으로 적용하였다.

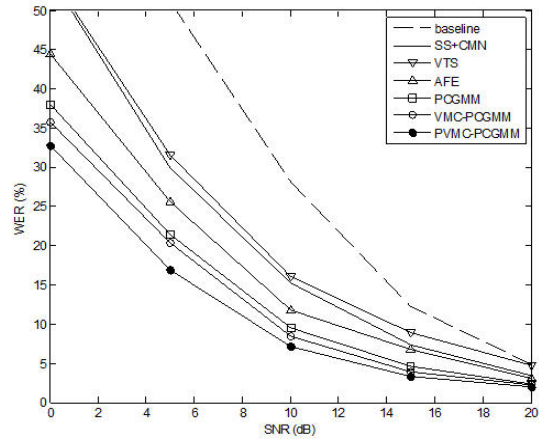


그림 5. 배경 음악 환경에서 SNR에 따른 각 전처리 기법의 인식 성능 비교 (WER, %)

Fig. 5 Recognition performance comparison at different SNR conditions of the background music over the different front-end methods (WER, %)

표 2는 PCGMM 기법을 전처리 기법으로 적용하여 성능 평가 결과를 나타낸 것이다. PCGMM은 단일 가우시안으로 배경 잡음을 모델링하여 사용한 기본적인 PCGMM 기법을 나타내고, VMC-PCGMM은 잡음 모델 생성을 위해 기존의 방법인 Variational 모델 생성 기법을 적용한 방법이다. PVMC-PCGMM 기법은 본 논문에서 제안하는 PCA 기반의 모델 생성 기법을 잡음 모델 예측을 위해 적용한 것을 말한다. 실험 결과와 같이 다중 모델을 채용한 VMC-PCGMM과 PVMC-PCGMM 기법 모두 단일 가우시안 모델을 사용하는 기본적인 PCGMM 보다 낮은 오인식률을 보이는 것을 알 수 있다. PVMC-PCGMM 기법이 기존의 VMC-PCGMM 기법보다 모든 SNR 조건에서 항상 우월한 성능을 나타내는 것을 알 수 있다. 괄호 안의 값은 VMC-PCGMM 기법에 비하여 PVMC-PCGMM의 상대적인 성능 향상율을 나타낸 것이다. 제안한 PVMC-PCGMM 기법이 기존의 VMC-PCGMM 기법에 비해 배경 음악을 잡음 환경에서, 절대 WER 값에서 평균 1.71% 향상하였고, 평균 12.14%의 상대적 성능 향상률을 보였다. 그림 5는 표 1과 표 2의 결과를 그래프로 나타낸 것이다. 이러한 결과를 바탕으로 본 논문에서 제안하는 PCA를 기반으로 하는 잡음 모델 생성 기법이 시간에 따라 변화하는 배경 음악 잡음 환경에서 음성 인식 성능을 향상시키는데 효과적임을 입증할 수 있다.

VI. 결 론

본 논문에서는 시간에 따라 변하는 잡음 환경에 강인한 음성 인식을 위해 효과적인 특징 보상 기법을 제안하였다. 제안하는 기법에서는 기존의 Variational 모델 생성 기법의 모델 정확도를 향상시키고자 PCA 기법을 사용하여 Variational 요소를 선택하였다. 제안된 기법은 다중 모델을 사용하는 PCGMM 기반의 특징 보상에 적용되었다. 제안된 기법을 Aurora 2.0의 음성 인식 평가 방식에 따라 성능을 평가하였고, 시간에 따라 변하는 잡음으로 배경 음악을 사용하였다. 실험 결과는 제안한 PCA 기반의 Variational 모델 생성 기법이 배경 음악 환경의 다양한 SNR 조건에서 기존의 전처리 기법 뿐 아니라 기존의 Variational 모델 생성 기법 보다 음성 인식 성능을 향상 시키는데 보다 우수함을 입증하였다.

감사의 글

이 논문은 인천대학교 2012년도 자체연구비 (신임교수연구과제) 지원에 의하여 연구되었음.

REFERENCES

- [1] S. F. Boll, "Suppression of Acoustic Noise in Speech Using Spectral Subtraction," *IEEE Trans. on Acoustics, Speech and Signal Proc.*, vol.27, pp.113-120, 1979.
- [2] Y. Ephraim and D. Malah, "Speech Enhancement Using Minimum Mean Square Error Short Time Spectral Amplitude Estimator," *IEEE Trans. on Acoustics, Speech and Signal Proc.*, vol.32, no.6, pp.1109-1121, 1984.
- [3] J. H. L. Hansen and M. Clements, "Constrained Iterative Speech Enhancement with Application to Speech Recognition," *IEEE Trans. on Signal Proc.*, vol.39, no.4, pp.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*, vol.24, no4, pp.267- 285, 1998.
- [5] W. Kim and J. H. L. Hansen, "Feature Compensation in the Cepstral Domain Employing Model Combination," *Speech Communication*, vol.51, no.2, pp.83-96, 2009.
- [6] 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 Proc.*, vol.2, no.2, pp.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, pp.171-185, 1995.
- [8] M. J. F. Gales and S. J. Young, "Robust Continuous Speech Recognition Using Parallel Model Combination," *IEEE Trans. on Speech and Audio Proc.*, vol.4, no.5, pp.352-359, 1996.
- [9] R. Martin, "Spectral Subtraction Based on Minimum Statistics," *EUSIPCO-94*, pp.1182-1185, Sep. 1994.
- [10] W. Kim and J. H. L. Hansen, "Variational Noise Model Composition Through Model Perturbation for Robust Speech Recognition with Time-Varying Background Noise," *Speech Communication*, vol.53, no4, pp.451-464, April 2011.
- [11] 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.
- [12] ETSI standard document, ETSI ES 201 108 v1.1.2 (2000-04), Feb. 2000.
- [13] ETSI Standard Document, ETSI ES 202 050 v1.1.1 (2002-10), 2002.



김우일(Wooil Kim)

2003년 고려대학교 전자공학과 공학박사
 2004년 ~ 2005년 미국 카네기 멜론 대학교 박사후 연구원
 2005년 ~ 2012년 미국 텍사스 주립대 (UT Dallas) 연구원 및 연구교수
 2012년 ~ 현재 인천대학교 컴퓨터공학부 조교수
 ※ 관심분야 : 신호처리, 패턴인식, 음성인식, 휴먼 컴퓨터 인터페이스