한국음향학회지 제23권 제5호 pp. 383-389 (2004)

바이어스 보상과 차원별 Eigenvoice 모델 평균을 이용한 고속화자적응의 성능향상

Performance Improvement of Rapid Speaker Adaptation Using Bias Compensation and Mean of Dimensional Eigenvoice Models

박 종 세*, 송 화 전*, 김 형 순*
(Jong Se Park*, Hwa Jeon Song*, Hyung Soon Kim*)

*부산대학교 전자공학과

(접수일자: 2003년 3월 5일; 수정일자: 2004년 1월 8일; 채택일자: 2004년 7월 13일)

본 논문에서는 훈련 및 인식 환경이 다른 상황에서 eigenvoice 기반 고속화자적응의 성능향상을 위하여 바이어스 보상을 적용한 eigenvoice 적응방식과 차원별 eigenvoice 모델 평균 가중합 방식을 제안하였다. PBW 452 DB를 사용한 어휘독립 단어인식 실험 결과에서 적은 양의 적응데이터를 사용했을 때 제안된 방식이 기존의 eigenvoice 방식에 비하여 많은 성능향상을 얻을 수 있었다. 적응단어수를 1개에서 50개로 변경시키면서 바이어스 보상을 적용한 eigenvoice 적응방식을 사용한 경우 기존 eigenvoice 방식보다 단어 오인식률이 약 22~30% 감소하였다. 또한 차원별 eigenvoice 모델 평균을 이용한 eigenvoice 적응방식에서는 1개의 단어를 적응데이터로 사용했을 경우에 기존 eigenvoice 방식보다 단어 오인식률이 최고 41%까지 감소하였다.

핵심용어: 음성인식, 화자적응, 고속화자적응, Eigenvoice

투고분야: 음성처리 분야 (2.5)

In this paper, we propose the bias compensation methods and the eigenvoice method using the mean of dimensional eigenvoice to improve the performance of rapid speaker adaptation based on eigenvoice under mismatch between training and test environment. Experimental results for vocabulary-independent word recognition task (using PBW 452 DB) show that the proposed methods yield improvements for small adaptation data. We obtained about 22~30% relative improvement by the bias compensation methods as amount of adaptation data varied from 1 to 50, and obtained 41% relative improvement in error rate by the eigenvoice method using the mean of dimensional eigenvoice with only single adaptation word.

Keywords: Speech recognition, Speaker adaptation, Rapid speaker adaptation, Eigenvoice

ASK subject classification: Speech signal processing (2.5)

I. 서 론

음성인식에 있어서 화자적응을 통하여 사용자의 특성을 더욱 잘 반영한 모델을 구성함으로써 인식성능 향상을 얻을 수 있다. 화자적응 방식은 크게 최대 사후 확률(Maximum A Posteriori, MAP)[1], 최대 우도 선형 회귀(Maximum Likelihood Linear Regression, MLLR)[2], 그리고 화자 군집화 (speaker clustering) 방식 등이 있다. 그 중에서 화자 군집화 방식의 하나인 eigenvoice 기법[3]이 고속화자적응에 유리한 것으로 알려져 있다.

Eigenvoice 기반 화자적응 방식에서는 새로운 화자의 적응데이터를 이용하여 각 차수별 eigenvoice의 가중치를 추정하고, 각 eigenvoice의 가중합으로 적응모델을 구성한다. Eigenvoice 적응방식은 기존의 다른 적응방식에 비해 적응 데이터로부터 추정할 파라메터 수가 적기 때문에 적응 데이터가 적은 경우에 성능이 우수하다. 그러나 적응 데이터가 많아져도 성능이 계속 향상되지 않고 빨리 수렴된다. 이러한 문제를 해결하기 위해 여러

책임저자 : 송 화 전 (hwajeon@pusan.ac.kr)
609-735 부산시 금정구 장전동 산30
부산대학교 전자공학과 음성통신연구실
(전화: 051-510-1704; 팩스: 051-515-5190)

가지 변형된 형태의 eigenvoice 적응 기법들이 연구되고 있다. 한 가지 방식은 eigenvoice 적응방식과 MAP 또는 MLLR 적응방식과 혼합된 형태[4]이고, 또 다른 방식으로는 segmental eigenvoice 적응방식[5], 차원별 eigenvoice 적응방식[6] 등의 eigenvoice를 더욱 세분화한 형태가 있다. 그 중에서 차원별 eigenvoice는 음성 특징벡터 차원별로 eigenvoice의 가중치를 추정하는 방식이며 적응데이터가 많은 경우에 기존 eigenvoice보다 향상된 인식성능을 얻을 수 있었다. 그러나, 이 방법의 경우 기존의 eigenvoice 방식에 비해 추정해야 하는 파라메터 수가 더 많기 때문에 적응데이터가 매우 적을 경우 오히려 성능이 급격히 떨어지는 문제를 나타낸다[6].

그리고, 훈련환경 및 인식환경의 불일치는 인식성능 하락의 또 다른 요소이다. Eigenvoice 적응방법에 기반하여 훈련 및 인식환경 불일치가 존재하는 화자공간에서 적은 적응데이터를 사용하여 새로운 화자에 대한 정확한 위치를 제대로 추정하기 어렵다.

본 논문에서는 훈련 및 인식환경의 불일치가 존재하는 상황에서 적응데이터가 어느 정도 많은 경우 뿐만 아니라 적응데이터가 적은 경우에도 기존의 eigenvoice 방법보다 성능을 향상시킬 수 있는 eigenvoice 기반의 두 가지 화자적응 방식을 제안하였다. 첫번째는 바이어스 보상 모델과 eigenvoice 적응 모델의 가중합 방식이며, 두번째는 eigenvoice 적응 모델과 차원별 eigenvoice 모델 평균을 이용한 방식이다.

본 논문의 구성은 다음과 같다. 2장에서는 eigenvoice 화자적응에 대하여 소개하고, 3장과 4장에서는 본 논문에서 제안한 eigenvoice 기반 화자적응의 성능향상 방식에 대하여 정리한다. 그리고 5장에서는 실험 및 성능평가를 정리하고, 마지막으로 6장에서 결론을 맺는다.

II. Eigenvoice 화자적응

Eigenvoice는 화자공간에서 각 화자들간의 변동을 가장 잘 대표하는 기저벡터를 설정하고 적응화자에 대하여 기저벡터 성분의 가중치를 추정하는 방식이다. 먼저 훈련환경의 T명의 SD모델 각각에 대해 모든 평균벡터를 연결하여 수퍼벡터로 만든 후 주성분 분석법 (principle component analysis, PCA)을 통해 eigenvoice들을 구성한다. Eigenvoice 기반 화자적응 방식에서 새로운 화자의 모델은 식 (1)과 같이 K개의 eigenvoice의 가중합으로 나타낼 수 있다.

$$\hat{\mu} = e(0) + \sum_{j=1}^{K} w(j)e(j) \tag{1}$$

여기서 e(0)는 T명의 화자종속 (SD) 모델의 평균이고, 또한 w(k)는 k차 eigenvoice, e(k)에 대한 가중치이며, K<T이다. 가중치 w(k)는 화자공간에서 새로운 화자에 대한 위치를 나타내며 이는 새로운 화자의 적응데이터로부터 최대 우도 고유치분석 (maximum likelihood eigen-decomposition, MLED)[3][4] 방법을 통해 추정된다. 그러나, 아주 적은 적응데이터를 이용하여 새로운 화자를 적응시키는 경우, 화자공간에서 적응된 모델의 위치는 새로운 화자의 화자공간에서의 실제위치와 큰 차이를 보일 수 있다. 특히 훈련환경과 인식환경이 다른 경우 더욱 더 심각해 질 것이므로 인식환경과 훈련환경의 불일치를 보상하는 것이 인식성능 향상에 유리하다고 판단된다.

III. 바이어스 보상을 적용한 Eigenvoice 적응방식

새로운 화자의 적응데이터가 주어지면 적응데이터와 모델간의 바이어스 성분을 추정하여 그림 1과 같이 음향공간에서 적응데이터의 환경 특성을 반영한 바이어스 보상 모델을 구하고 이를 eigenvoice 적응 시 사용하면 화자공간에서 새로운 화자의 위치를 보다 더 신뢰성 있게 추정할 수 있을 것이다.

본 논문에서는 화자특성 및 인식환경을 더욱 잘 반영할 수 있도록 eigenvoice 적응 방식과 함께 바이어스 보상 방식을 적용하였다. 그리고 바이어스 보상의 적용 형

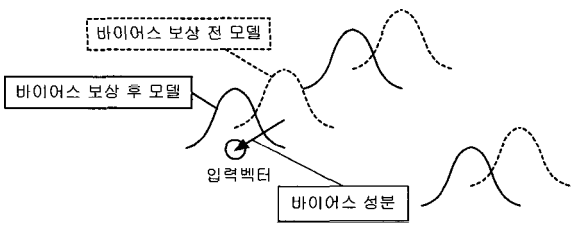

그림 1. 바이어스 성분 보상에 대한 예
Fig. 1. An example of compensation for the bias components.

태에 따라 그림 2와 같이 네 가지 방법으로 사용하였다.

방법 1: 바이어스 보상 가중합 방식

방법 1은 그림 2 (a)와 같이 나타낼 수 있다. 이 방법에서 최종 적응모델은 식 (2)와 같이 바이어스 성분을 보상한 모델과 eigenvoice로 적응한 모델과의 가중합으로 구성하였다.

$$\hat{\mu} = (1-\alpha)\hat{\mu}_{EV} + \alpha\,[\,\mu_{SI} + \hat{b}_1\,], \quad 0 \le \alpha \le 1 \tag{2}$$

여기서 $\hat{\mu}_{EV}$는 eigenvoice 방식으로 적응된 모델이고, $(\mu_{SI} + \hat{b}_1)$는 화자독립 모델에 대하여 바이어스 성분을 보상한 모델이다. 이때 바이어스 성분, $\hat{b}_1$는 SM(stochastic matching) 기법[7]에 기반하여 아래 식과 같이 추정하였다.

$$\hat{b}_1 = \frac{\sum_{t=1}^{T}\sum_{s}\sum_{m}\gamma_m^{(s)}(t)(\mathbf{o}_t - \mu_m^{(s)})}{\sum_{t=1}^{T}\sum_{s}\sum_{m}\gamma_m^{(s)}(t)} \tag{3}$$

여기서 T는 적응 데이터의 전체 프레임 수이고, $\mu_m^{(s)}$는 t 시간에서의 관측치 $\mathbf{o}_t$에 할당된 상태 s, 믹스쳐 m에 해당하는 화자독립 모델의 평균 벡터이다. 이 때 $\gamma_m^{(s)}(t)$는 시간 t 에서 상태 s, 믹스쳐 m에 머무를 확률(occupation probability)이다. 즉 추정된 바이어스 성분은 적응 데이터와 화자독립 모델간의 상태편차 평균이다.

식 (2)에서 α =0 인 경우에는 기존의 eigenvoice 적응 방식이 되고, α =1 인 경우에는 화자독립 모델에서 바이어스 성분만 보상한 방식이 된다. 성능향상을 위해서 화자 및 적응 데이터의 수에 따라 최적의 α 를 결정하는 것이 필요하다.

본 논문에서는 가중치 α 를 결정하는 기준으로 적응단어의 입력벡터와 두 모델간의 거리의 비율을 사용하였다. 먼저 각 프레임별로 입력벡터와 eigenvoice 모델과의 거리, 입력벡터와 바이어스 보상된 모델간의 거리의 비를 식 (4)와 같이 r_t로 정의하였다.

$$r_t = \frac{|\mathbf{o}_t - \hat{\mu}_{EV}|}{|\mathbf{o}_t - (\mu_{SI} + \hat{b}_1)|} \tag{4}$$

여기서 입력벡터가 eigenvoice 모델과 가까우면 r_t는 커지고 바이어스 보상 모델과 가까우면 EMBED Equation.3 r_t는 작아지게 된다. 그리고 r_t가 1보다 크면 양수, 1보다 작으면 음수가 되도록 로그 함수를 사용하여 식 (5)와 같이 x_t를 정의하고, 각 프레임별 가중치가 0 ~ 1 사이의 값을 가지도록 시그모이드 함수를 적용한 후, 식(2)의 가중치 α는 식 (6)과 같이 전체 발화 길이에 대한 평균값으로 나타나게 하였다.

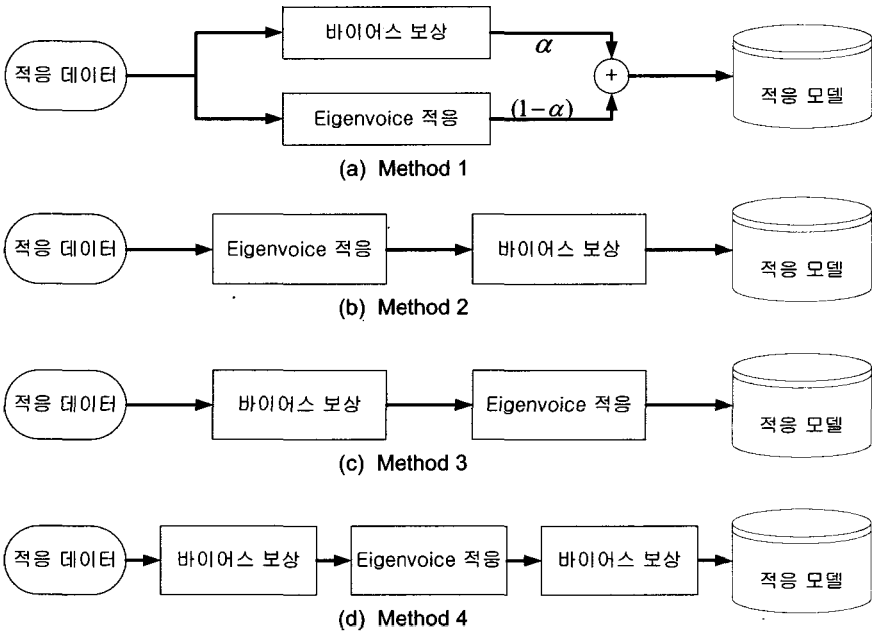

그림 2. 바이어스 보상을 적용한 eigenvoice 적응방식

Fig. 2. Eigenvoice adaptation methods applying the bias compensation.

$$x_t = \log(r_t) \tag{5}$$

$$\alpha = \frac{\sum_{t=1}^{T}\sum_{s}\sum_{m}\gamma_m^{(s)}(t)\left(\frac{1}{1+e^{-x_t}}\right)}{\sum_{t=1}^{T}\sum_{s}\sum_{m}\gamma_m^{(s)}(t)} = \frac{\sum_{t=1}^{T}\sum_{s}\sum_{m}\gamma_m^{(s)}(t)\left(\frac{1}{1+r_t^{-1}}\right)}{\sum_{t=1}^{T}\sum_{s}\sum_{m}\gamma_m^{(s)}(t)} \tag{6}$$

실험결과에서 가중치 α를 고정시킨 경우보다 식 (6)과 같이 화자 및 적응데이터의 수에 따라 자동으로 결정한 경우에 인식성능이 좋게 나타났다.

방법 2: Eigenvoice 적응 후 바이어스 보상 방식

방법 2는 그림 2 (b)와 같이 eigenvoice 방식으로 적응 후에 입력벡터에 대한 바이어스 성분을 보상하는 방식이다. 먼저 적응데이터로부터 eigenvoice 적응모델을 구성하고 다시 바이어스 성분을 추정하여 식 (7)과 같이 최종 적응모델을 구성하였다.

$$\hat{\mu} = \hat{\mu}_{EV} + \hat{b}_2 \tag{7}$$

이 때 바이어스 성분 3 $\hat{b}_2$은 적응데이터의 입력벡터와 eigenvoice 적응모델로부터 식 (8)과 같이 추정할 수 있다.

$$\hat{b}_2 = \frac{\sum_{t=1}^{T}\sum_{s}\sum_{m}\gamma_m^{(s)}(t)(\mathbf{o}_t - \tilde{\mu}_m^{(s)})}{\sum_{t=1}^{T}\sum_{s}\sum_{m}\gamma_m^{(s)}(t)} \tag{8}$$

여기서 $\tilde{\mu}_m^{(s)}$는 t 시간에서의 관측치 $\mathbf{o}_t$에 할당된 상태 s, 믹스쳐 m에 해당하는 eigenvoice 적응모델의 평균벡터이다. 위의 식 (8)은 식(3)과 비슷한 형태이지만 식 (3)에서는 $\mu_m^{(s)}$가 화자독립(SI) 모델의 평균벡터이고, 식 (8)에서는 $\tilde{\mu}_m^{(s)}$이 eigenvoice 적응모델의 평균벡터이다.

방법 3: 바이어스 보상 후 eigenvoice 적응 방식

방법 3은 그림 2 (c)와 같이 바이어스 보상 후에 eigenvoice 적응모델을 구성한다. 앞의 방법과 반대로 먼저 적응데이터의 입력벡터로부터 바이어스 성분을 보상한 후, eigenvoice 방식으로 적응 모델을 구성한다. 이 방법에서는 바이어스 보상된 결과를 MLED 과정에 적용시키기 위해서 식 (9)와 같이 추정된 바이어스 성분을 $e(0)$에 보상하여 MLED 과정에서는 보상된 eigenvoice, $\hat{e}(0)$을 $e(0)$ 대신 사용한다.

$$\hat{e}(0) = e(0) + \hat{b}_3 \tag{9}$$

이 때 바이어스 성분 $\hat{b}_3$은 적응데이터의 입력벡터와 eigenvoice 부벡터(subvector), $e(0)$로부터 식 (10)과 같이 추정할 수 있다.

$$\hat{b}_3 = \frac{\sum_{t=1}^{T}\sum_{s}\sum_{m}\gamma_m^{(s)}(t)(\mathbf{o}_t - e_m^{(s)}(0))}{\sum_{t=1}^{T}\sum_{s}\sum_{m}\gamma_m^{(s)}(t)} \tag{10}$$

여기서 $e_m^{(s)}$는 t 시간에서의 관측치 ion.3 $\mathbf{o}_t$에 할당된 상태 s, 믹스쳐 m에 해당하는 eigenvoice 부벡터, $e(0)$이다.

방법 4: 방법 2 + 방법 3

방법 4는 그림 2 (d)와 같이 eigenvoice 적응단계의 전후에 바이어스 보상을 적용하는 방식이다. 따라서 방법 4는 앞에서 설명했던 방법 2와 방법 3을 함께 사용하는 방법이다.

IV. 차원별 Eigenvoice 모델 평균을 이용한 Eigenvoice 적응방식

기존 연구에서 음성 특징벡터의 차원별로 eigenvoice 가중치를 추정하는 방식인 차원별 eigenvoice 방식을 사용하여 발화수가 많은 경우에 높은 인식률을 얻을 수 있었다[6]. 또한 인식환경과 동일한 환경의 다른 화자들의 데이터가 주어진 경우에 다른 화자들의 정보를 사용하여 성능을 향상시킬 수 있을 것이다. 본 논문에서는 인식환경의 다른 화자들로부터 구성한 차원별 eigenvoice 모델 평균을 이용한 eigenvoice 기반 화자적응의 성능향상 방식을 살펴본다. 차원별 eigenvoice 모델 평균은 미리 여러 화자들로부터 차원별 eigenvoice 가중치를 추정한 후 평균 가중치를 계산하여 얻을 수 있다.

방법 A: 차원별 eigenvoice 모델 평균 가중합 방식

본 논문에서는 차원별 eigenvoice 모델 평균을 이용한 한가지 방법으로 차원별 eigenvoice 모델 평균 가중합 방식을 적용하였다. 이 방식에서는 새로운 화자의 모델을 식 (11)과 같이 적응 데이터로부터 구한 eigenvoice 모델과 여러 화자로부터 얻은 차원별 eigenvoice 모델 평균과의 가중합으로 구성하였다.

$$\hat{\mu} = (1-\beta)\mu_{EV} + \beta\bar{\mu}_{EV\_DIM}, \quad 0 \le \beta \le 1 \tag{11}$$

이 때 μ_{EV} 는 기존 eigenvoice 모델이고, $\bar{\mu}_{EV\_DIM}$ 은 여러 화자로부터 구성한 차원별 eigenvoice 모델의 평균이다. 여기서 β =0 인 경우는 기존의 eigenvoice 방식이며, β =1 인 경우는 화자의 적응 데이터를 사용하지 않고 다른 화자들로부터 얻은 차원별 eigenvoice 모델의 평균을 사용하는 방식이다. 따라서 적응화자의 최종 모델은 eigenvoice로 구성한 모델과 미리 구성된 차원별 eigenvoice 모델 평균 사이에 위치하게 된다. 본 논문에서는 β 를 실험적으로 변화시켜가면서 성능평가를 하였다.

방법 B: 차원별 eigenvoice 모델평균 적용 후 eigenvoice 적응

본 논문에서는 차원별 eigenvoice 모델 평균을 이용한 다른 방식으로 차원별 eigenvoice 모델 평균을 새로운 SD 모델의 평균으로 사용하는 방식을 적용하였다. 이 방식에서는 eigenvoice의 부벡터(subvector), $e(0)$ 성분을 식 (12)와 같이 차원별 eigenvoice 모델 평균으로 대체하였다. 그리고 MLED 과정에서는 변경된 eigenvoice, $\hat{e}(0)$ 를 사용하여 eigenvoice의 가중치를 추정하게 된다.

$$\hat{e}(0) = \bar{\mu}_{EV\_DIM} \tag{12}$$

여기서 $\bar{\mu}_{EV\_DIM}$ 은 여러 화자로부터 구성한 차원별 eigenvoice 모델의 평균이다.

V. 실험 및 결과

5.1. 실험환경[6]

본 논문에서는 교사 방식(supervised mode) 고속화자 적응을 적용하여 가변어휘독립 음성인식시스템의 성능을 향상시키고자 하였다. 먼저, 화자독립 (SI) 모델을 구성하기 위해 ETRI에서 구축한 3음소열 최적화 단어 (Phonetically Optimized Words, POW)[8] 음성 데이터베이스 중에서 남성화자 40명의 음성 데이터베이스를 사용하였다.

20ms Hamming window를 10ms씩 이동시키면서 얻은 12차의 MFCC (Mel-Frequency Cepstral Coefficient)와 1,2차 미분치를 구하여 총 36차의 음성 특징벡터를 사용하였다. 그리고 유성음화 자음 및 묵음을 포함한 46 유사음소 (PLU) set[9]을 기본으로 상태 수준에서의 결정 트리 기반 군집화 기법 (tree-based clustering)을 적용한 트라이폰을 기본모델로 사용하였으며 모델 당 상태수는 3개이며, 사용된 전체 공유 상태수는 4050개이다.

POW DB 중에서 남성화자 40명의 음성 데이터베이스를 사용하여 SI 모델을 구성한 후, 각 화자에 대하여 MLLR 을 적용한 후 MAP 적응방식으로 40개의 화자종속 (SD) 모델을 구성하였다. 그리고 40개의 화자종속 모델의 평균모델들을 연결하여 수퍼벡터 (supervector)를 만들고 40개의 평균 벡터 e(0)를 구하였다. 그리고, 각각의 SD 모델의 수퍼벡터에 e(0)를 차감한 후 PCA를 적용하여 40차의 eigenvoice를 구성하고 그 중 30차의 eigenvoice를 사용하였다. 본 논문에서 사용된 수퍼벡터의 차원은 상태당 믹스쳐 수가 1개일 경우 4050 * 36 * 1 =145800 이고, 믹스쳐 수가 2개일 경우는 4050 * 36 * 2 = 291600 이다.

화자적응 및 성능평가를 위해서는 훈련에 사용한 DB와는 참가인원 및 발성 내용, 녹음 환경이 완전히 틀린 452 균일 음소 분포 단어 (Phonetically Balanced Words, PBW)[10] 데이터베이스의 일부인 남성화자 10명을 사용하였으며, 각 화자별로 1개부터 50개까지 적응 단어 수를 늘려가며 적응에 사용하였고, 나머지 중 400개의 단어를 성능평가에 사용하였다.

5.2. 실험결과

기존의 여러 가지 화자적응 방식 및 환경보상을 적용한 인식성능을 비교하였다. Baseline 시스템은 40명의 POW DB로부터 구성한 화자독립 (SI) 모델을 사용하였다. 그림 3은 믹스쳐 수가 1개인 경우 기존에 사용한 baseline 시스템과 MAP, MLLR 그리고 eigenvoice 방식으로 화자적응을 수행한 결과[6]와 환경보상을 위한 대표적인 방법 중에 하나인 켑스트럴 평균 차감법 (CMS)을 적용한 경우의 인식결과를 나타내었다.

본 논문에서 사용된 훈련 및 인식 DB 사이에는 부가 잡음보다는 채널에 대한 영향이 있으므로 CMS를 통해 인식성능이 향상되는 것을 알 수 있다.

그림 4와 그림 5는 본 논문에서 제안한 바이어스 보상을 적용한 eigenvoice 적응방식의 성능을 나타낸 것이다. 그림 4는 상태 당 믹스쳐 수가 1개일 때, 그림 5는 상태 당 믹스쳐 수가 2개일 때의 결과이다.

실험결과에 나타난 바와 같이 바이어스 보상 방식을 적용한 eigenvoice 적응방식은 기존 eigenvoice 적응방식 (그림에서 EV)에 비하여 많은 성능향상이 있음을 알 수 있다. 그리고 바이어스 보상 방식들 중에서 방법 4 (그림에서 method 4)가 가장 성능이 우수하였다. 이는 방법 4가 방법 2와 방법 3에 의한 성능향상을 함께 가질 수 있기 때문이다. 기존 eigenvoice 방법에 비해 방법4를 사용하여 적응 단어가 1개일 경우 단어 오인식률이 2.70%에서 2.12%로 21.5%가 감소하였으며, 적응 단어가 50개일 경우에는 단어 오인식률이 1.82%에서 1.27%로 30% 감소하였다. 그리고, 방법 2와 3을 비교하면 먼저 환경에 대한 불일치를 보상한 후에 화자적응을 하는 것이 좀 더 유리하다고 판단된다.

적응화자에 대하여 차원별 eigenvoice 모델 평균을 이용한 eigenvoice 적응방식에 대한 성능평가를 하였으며 그림 6은 실험결과를 보여준다. 차원별 eigenvoice 모델 평균은 PBW DB 중에서 인식실험에 참가하지 않은 남성화자 28명에 대하여 각 화자별로 50개의 단어를 사용하여 차원별 eigenvoice 모델을 추정하고 평균을 구하였다.

차원별 eigenvoice 모델 평균을 이용한 eigenvoice 적응방식을 사용하여 기존 eigenvoice 방식보다 성능이 많이 향상되었으며 적응단어수가 적은 경우에 성능향상 폭이 크다. 그리고 $e(0)$를 차원별 eigenvoice 모델 평균으로 변경하는 방식 (방법 B)이 차원별 eigenvoice 모델 평균 가중합 방식 (방법 A)보다 전체적으로 성능이 우수하였다. 이는 화자공간에서 훈련환경의 기준점을 인식환경의 기준점으로 미리 이동시킨 후 eigenvoice를 사용하여 화자의 변화량을 고려하기 때문이다.

차원별 eigenvoice 모델 평균 가중합 방식에서는 β는 0에서 1까지 0.1 단위로 변경시켜나가면서 실험을 수행하였다. 0.3에서 0.5까지는 인식성능이 거의 동일하였고 그 이상의 경우 다시 인식성능이 하락하였다. β =1.0

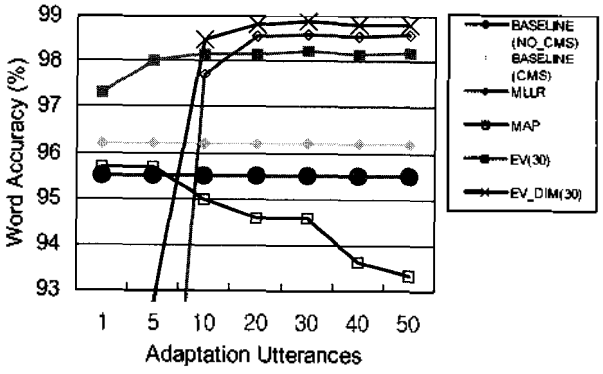

그림 3 CMS 적용 및 여러 가지 적응방식들의 성능비교 (1-mixture)
Fig. 3. Performance of application of CMS and several adaptation methods. (1-mixture)

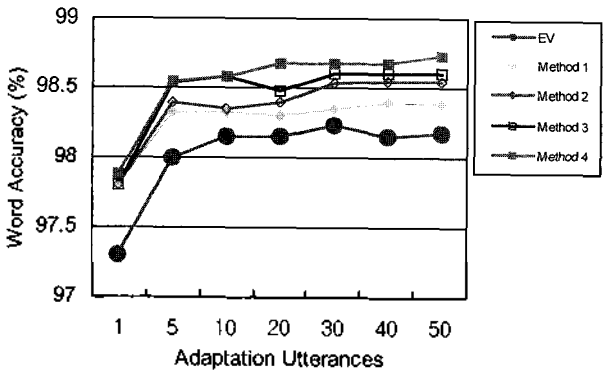

그림 4. 바이어스 보상을 적용한 eigenvoice 적응방식 실험결과 (1-mixture)
Fig. 4. Results of Eigenvoice adaptation methods applying the bias compensation. (1-mixture)

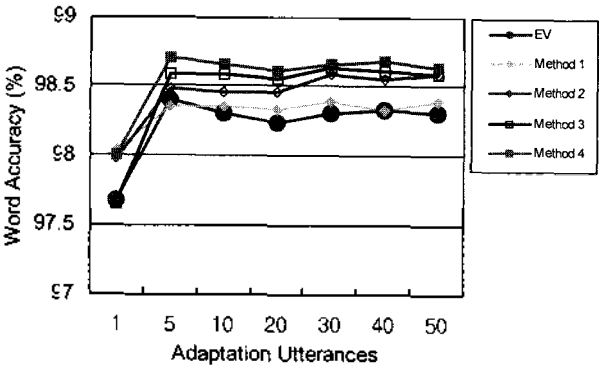

그림 5. 바이어스 보상을 적용한 eigenvoice 적응방식 실험결과 (2-mixture)
Fig. 5 Results of Eigenvoice adaptation methods applying the bias compensation. (2-mixture)

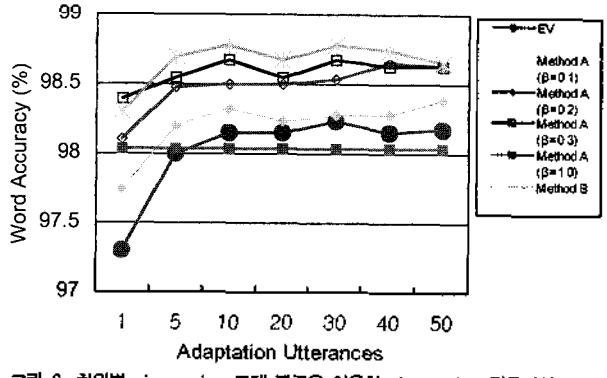

그림 6. 차원별 eigenvoice 모델 평균을 이용한 eigenvoice 적응방식 실험결과 (1-mixture)
Fig. 6. Results of eigenvoice adaptation using mean of dimensional eigenvoice models. (1-mixture)

일때는 차원별 모델 평균만을 사용하는 경우이며, β가 1.0 근처로 갈수록 새로운 화자의 적응데이터에 대한 정보가 반영되지 못하므로 적응데이터 수가 증가함에 따라 인식성능이 향상되지 않음을 알 수 있다. D Equation.3 β=0.3 일 때 가장 좋은 성능을 나타냈으며, 특히 발화수가 1개인 경우에 기존의 eigenvoice 방식에 비해 단어 오인식률이 2.7%에서 1.6%로 41%가 감소하였다.

따라서, 본 논문에서 제안한 바이어스 보상방식을 이용한 eigenvoice 화자적응 방법이 훈련환경 및 인식환경의 불일치가 존재하는 경우에 특히 적응 데이터가 매우 적은 고속화자 적응에도 인식성능 향상에 효과적임을 알 수 있다.

VI. 결 론

본 논문에서는 훈련 및 인식환경이 다른 상황에 대해 두가지 eigenvoice 기반 고속화자적응의 성능향상 방식을 제안하였다. 먼저 적응단어수를 1개에서 50개로 변경시키면서 바이어스 보상을 적용한 eigenvoice 적응방식을 사용하여 기존 eigenvoice 적응방식에 비하여 22~30%의 성능향상을 얻을 수 있었다. 또한 차원별 eigenvoice 모델 평균을 이용한 eigenvoice 적응방식에서도 기존의 eigenvoice 방식에 비하여 26~41%의 성능향상을 얻을 수 있었다. 그러나, 차원별 eigenvoice 모델 평균을 얻기 위해서 미리 인식환경과 동일한 다른 화자의 정보가 요구된다. 앞으로 이러한 문제를 제거하는 방법과 또한 발화수가 증가함에 따라 차원별 eigenvoice의 성능을 유지하는 방법을 개발하는 것이 요구된다.

참 고 문 헌

1. 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. Signal Processing, 39(4), pp.806-814, April, 1991.
2. C. J. Leggetter and P. C. Woodland, "Maximum likelihood linear regression for speaker adaptation of continuous density hidden Markov models," Computer Speech and Language, 9(1), pp.171-185, Sep, 1995.
3. R. Kuhn, P. Nguyen, J. C. Jungua, L. Goldwasser, N. Niedzielski, S. Finche, K. Field and M. Contolini, "Eigenvoices for speaker adaptation," in Proc. ICSLP, 5, pp.1771-1774, 1998.
4. R. Kuhn, J. C. Jungua, P. Nguyen, and N. Niedzielski, "Rapid speaker adaptation in eigenvoice space," IEEE Trans. Speech and Audio Processing, 8(6), pp.695-707, Nov. 2000.
5. Y. Tsao, S. M. Lee, F. C. Chou and L. S. Lee, "Segmental eigenvoice for rapid speaker adaptation," in Proc. Eurospeech, 2, pp.1269-1272, 2001.
6. 송화전, 이윤근, 김형순, "차원별 Eigenvoice와 화자적응 모드 선택에 기반한 고속화자적응 성능 향상," 한국음향학회지, 22(1), pp.48-53, 2003년 1월.
7. A. Sanker, "A maximum-likelihood approach to stochastic matching for robust speech recognition," IEEE Trans. Speech and Audio Processing, 4(3), pp.190 -202, May, 1996.
8. Y. Lim and Y. Lee, "Implementation of the POW (Phonetically Optimized Words) algorithm for speech database," In Proc. ICASSP, 1, pp.89-91, 1995.
9. 유재원, 연속음성인식을 위한 음성 단위 발음사전 구성방법 연구, 위탁과제 최종연구보고서, 한국전자통신연구소, 1995.
10. 이용주, 김봉완, 김종진, 양옥렬, 임선영, "음성 DB용 PBW에 관한 검토," 제 12회 음성통신 및 신호처리워크샵 논문집, pp.310-314, 1995년 6월.

저자 약력

• 박 종 세 (Jong Se Park)

2000년 2월: 부산대학교 공과대학 전자공학과 (공학사)
2003년 2월: 부산대학교 대학원 전자공학과 (공학석사)
※ 주관심분야: 음성인식, 음성합성, 음성신호처리

• 송 화 전 (Hwa Jeon Song)

현재: 부산대학교 전자공학과 박사과정
한국음향학회지 22권 1호 참조

• 김 형 순 (Hyung Soon Kim)

현재: 부산대학교 전자공학과 부교수
한국음향학회지 22권 1호 참조