한국음향학회지 제24권 제4호 pp. 210~216 (2005)

분리된 고유공간을 이용한 잡음환경에 강인한 특징 정규화 기법

Robust Feature Normalization Scheme Using Separated Eigenspace in Noisy Environments

이 윤 재*, 고 한 석*
(Yoonjae Lee*, Hanseok Ko*)

*고려대학교 전자컴퓨터 공학과

(접수일자: 2004년 12월 15일; 수정일자: 2005년 3월 11일; 채택일자: 2005년 4월 8일)

본 논문에서는 잡음에 강인한 음성인식을 위하여 고유공간에 기반을 둔 새로운 특징 정규화 기법을 제안한다. 일반적으로 평균과 분산의 정규화 (MVN)는 켑스트럼 상에서 수행된다. 그러나 최근에 고유공간을 이용한 MVN 기법이 소개되었고, 그 고유공간 정규화 기법에서는 하나의 고유공간을 이용하였다. 이 과정에는 켑스트럼 상의 특징 벡터를 선형 주성분 분석 (PCA)행렬을 통하여 고유공간으로 변환시킨 후 MVN을 수행하는 과정이 포함된다. 이 방법에서는 전체 39차의 특징분포를 하나의 고유공간으로 표현하였다. 그러나 이 기법의 경우 전체 특징 분포를 표현함에 세밀함이 떨어지기 때문에 더욱 세밀한 분포의 표현을 위해 본 논문에서는 static 특징, 1차 미분 계수, 2차 미분계수에 각각 유일하고 독립적인 분리된 고유공간을 적용하는 것을 제안하였다. 또한 고유공간에서 정규화된 훈련 데이터를 이용하여 모델을 만든다. 마지막으로 훈련 데이터의 분포와 잡음환경에서의 테스트 데이터의 분포 특성의 차이를 줄이기 위해 켑스트럼 상에서의 회전 기법을 적용시킨다. 그 결과, 기본적인 고유공간 정규화 기법보다 향상된 성능을 얻을 수 있었다.

핵심용어: 음성 인식, 평균 분산 정규화, 분리된 고유공간, 켑스트럼 상의 회전기법

투고분야: 음성처리 분야 (2.5)

We propose a new feature normalization scheme based on eigenspace, for achieving robust speech recognition. In general, mean and variance normalization (MVN) is performed in cepstral domain. However, another MVN approach using eigenspace was recently introduced, in that the eigenspace normalization procedure performs normalization in a single eigenspace. This procedure consists of linear PCA matrix feature transformation followed by mean and variance normalization of the transformed cepstral feature. In this method, 39 dimensional feature distribution is represented using only a single eigenspace. However it is observed to be insufficient to represent all data distribution using only a single eigenvector. For more specific representation, we apply unique and independent eigenspaces to cepstra, delta and delta-delta cepstra respectively in this paper. We also normalize training data in eigenspace and get the model from the normalized training data. Finally, a feature space rotation procedure is introduced to reduce the mismatch of training and test data distribution in noisy condition. As a result, we obtained a substantial recognition improvement over the basic eigenspace normalization.

Keywords: Speech recognition, Mean and variance normalization, Separated eigenspace, Feature space rotation

ASK subject classification: Speech Signal Processing (2.5)

책임저자: 고 한 석 (hsko@korea.ac.kr)
136-701 서울시 성북구 안암5가-1
고려대학교 전자컴퓨터 공학과
(전화: 02-3290-3239; 팩스: 02-3291-2450)

I. 서 론

음성 인식 시스템의 성능을 하락시키는 주요 원인은 음향 모델을 얻는 훈련 데이터의 환경과 실제 인식 데이터의 환경과의 불일치에 의한 것이다. 인식 환경에서의 배경잡음, 음성 신호의 전달과정에서 발생하는 채널왜곡 등이 성능 하락의 주요한 원인이며 두 환경의 차이를 최대한 줄이는 기술 개발은 음성 인식 분야에서의 가장 중요하고 필수적인 항목 중의 하나이다.

지금까지 훈련 환경과 인식 환경과의 차이를 줄이기 위한 전처리 단에서의 여러 기법들이 제안되어왔다. 그 중 켑스트럼 평균 정규화 (Cepstral Mean Normalization, CMN) 기법과 평균과 분산 정규화 (Mean and Variance Normalization, MVN) 기법은 다른 기법들에 비해 상대적으로 간단하면서도 효과적인 기법이다[1,2].

CMN은 시불변 채널왜곡 현상을 보상해주기 위해 특징 벡터의 평균을 정규화 해주는 기법이며 MVN은 특징 벡터의 평균뿐만 아니라 분산도 정규화 시켜 잡음 환경에 보다 강인하게 해주는 기법이다.

본 논문에서는 MVN에 대해 관심을 갖는다. 일반적인 MVN은 켑스트럼 영역에서 수행되나 본 논문에서는 고유공간에서의 MVN에 대해서 관심을 가지고 수행한다. MVN은 특징 벡터의 각각의 차원이 서로 상관관계 (correlation) 가 없다고 가정하고 각각의 차원에 대해서 독립적인 정규화 기법을 적용한다. 그러나 실제로 각 차원 간에는 적은 양이지만 서로 상관관계를 가지고 있다. 최근에 주성분 분석을 이용하여 켑스트럼 상의 특징을 고유공간으로 변환시키는 기법이 소개되었다[3]. 그 결과, 고유공간에서의 특징 벡터의 각 차원간 상관성은 더욱 줄어들게 되어 특징의 분산 행렬은 더욱 대각행렬에 가까워지게 된다[4]. 그러므로 고유공간에서의 MVN 은 켑스트럼 상에서의 MVN 보다 특징벡터의 차원간의 연관성이 더욱 줄어드는 환경에서 수행된다는 점에서 더욱 효과적이게 된다.

기존의 고유공간 정규화 기법은 전체 39차 특징벡터 (13차 static 특징, 1,2차 미분계수)에 39차의 하나의 고유공간을 인식 데이터에 적용하여 수행되었다. 그러나 고차원의 데이터 분포를 하나의 고유공간으로 표현하기에는 한계가 있으며 데이터 분포 표현의 세밀함이 떨어지게 된다. 또한 고유공간의 정규화를 인식 데이터에만 적용하는 것보다는 훈련 데이터에도 적용하여 음향 모델을 얻는 것이 성능 향상에 더욱 효과적일 것이라고 예상되었다. 이러한 관점에서 본 논문에서는 다음과 같은 과정의 실험이 수행되었다.

먼저, 하나의 그룹으로 이루어진 전체 특징 벡터를 세 개의 그룹 (1차 static 특징, 1차 미분 계수, 2차 미분계수) 로 분리하여 각각에 대해 세 개의 독립적인 고유공간 정규화 기법을 수행하고 더 높은 인식 성능 향상을 위해 훈련데이터에도 고유공간 정규화 기법을 적용한다. 마지막으로 추가적인 기술로 Siroko Molau가 제안한 특징 공간에서의 회전기법을 적용하였다[5]. 캡스트럼 상에서의 특징은 잡음환경에 의해 평균과 분산만 변하는 것이 아니라 전체적인 분포의 회전 현상도 발생하므로 이러한 현상을 보상해주기 위해 이 기법을 도입하였다.

본 논문은 다음과 같이 구성된다. II 장에서는 기본적인 고유공간에서의 정규화 기법에 대해 설명하고 III 장에서는 제안된 기법에 대해 설명한다. IV 장에서는 제안한 기법의 성능 평가를 위한 실험과 그 결과에 대해 고찰하고 V 장에서 결론을 맺는다.

II. 고유공간 정규화기법

평균이 0인 K 차원의 훈련 데이터의 특징 벡터를 고려할 때 켑스트럼 상의 특징, X_t^c 와 고유공간의 특징, Z_t 와의 관계는 Λ 에 의해 정의된다.

$$Z_t \sim N_K(0, I) \tag{1}$$

$$X_t^c = \Lambda Z_t \tag{2}$$

위 첨자 c 는 켑스트럼 영역을 의미하며 아래첨자 t 는 시간 색인을 의미한다. 고유공간 상에서의 훈련데이터의 분포는 K 차원의 표준 정규 분포를 따르며 변환 행렬, Λ 에 의해 고유공간에서 켑스트럼 상의 특징으로 변환된다.

Λ 는 PCA를 이용해 전체 훈련데이터로부터 구해지는 통계량이므로 켑스트럼 상의 훈련데이터는 $\Lambda^{-1}X_t^c$ 의 역변환에 의해 고유공간 상에서 평균 0, 분산 1인 분포를 가지게 된다. 하지만 부가잡음과 채널왜곡에 의해 오염된 음성 특징 $\hat{X}_t^c$ 은 역변환 $\Lambda^{-1}\hat{X}_t^c$ 을 수행하여도 평균 0, 분산 1의 분포를 가지지 않는다. 그러므로 각 차원간의 상관성이 더 줄어드는 역변환된 고유공간에서 MVN

을 수행하여 훈련데이터의 분포에 가깝게 만들어주게 된다. 자세한 과정은 아래와 같다.

1. 훈련 데이터로부터 Λ를 얻는다.

전체 훈련데이터로부터 공분산 행렬 $E[X_t^c X_t^{cT}]$을 구한다. Λ는 $UV^{1/2}$로부터 얻어진다. $E[X_t^c X_t^{cT}] = UVU^T$이며 U와 V는 각각 훈련 데이터의 공분산 행렬의 고유벡터와 고유값이다.

2. 켑스트럼 상의 인식 특징 벡터 $\hat{X}_t^c$을 고유공간 상의 벡터로 변환시킨다.

$$\hat{Z}_t = \Lambda^{-1}\hat{X}_t^c = V^{-1/2}U^T\hat{X}_t^c$$

3. 고유공간에서의 MVN을 수행한다.

$$\overline{Z}_t = \frac{\hat{Z}_t - mean(\hat{Z}_t)}{std(\hat{Z}_t)}$$

$mean(\cdot)$와 $std(\cdot)$은 각각 한 샘플 $\hat{Z}_t$에 대한 표본 평균 및 표준편차를 나타낸다.

4. 고유공간에서 정규화된 특징을 다시 켑스트럼 상의 특징으로 변환시킨다.

$$\overline{X}_t^c = \Lambda\overline{Z}_t$$

고유공간 상에서는 각 차원간의 상관관계가 켑스트럼 상에서 보다 더 줄어들기 때문에 MVN이 더욱 효과적이다. 그러나 하나의 고유공간으로 고차원의 전체 분포를 효율적으로 표현하기는 부족하다. 따라서 본 논문에서는 분리된 고유공간들로 전체 데이터의 분포를 표현하고 훈련데이터에도 고유공간에서의 정규화 기법을 적용하므로써 더 높은 성능 향상을 꾀하려고 한다. 마지막으로 켑스트럼 상에서의 회전 기법을 도입하여 더욱 강인한 시스템을 구현하고자 하였다.

III. 제안한 기법

3.1. 훈련데이터의 고유공간 정규화 기법

일반적으로 MVN은 두 가지 다른 방법으로 수행할 수 있다. 첫 번째 방법은 오직 인식 데이터에만 적용하는 것이다. 이것은 인식 데이터의 분포를 훈련 데이터의 전체 분포와 같게 만들어주는 방법이다. 두 번째 방법은 MVN을 훈련데이터와 인식 데이터 모두 적용시키는 방법이다. 이것은 훈련 데이터의 분포를 모두 평균 0, 분산 1인 분포를 만들어 이 데이터로부터 음향 모델을 얻은 후, 인식 데이터에도 똑같은 방법을 적용하여 인식하게 하는 방식이다. 일반적으로 후자의 방법에서 조금 더 높은 성능을 기대할 수 있다. 그 이유는 인식 데이터에 적용되는 기법과 같은 기법이 적용된 훈련데이터로 훈련시킨 음향 모델이 더욱더 효과적이기 때문이며 또한 실제 구현에 있어 다양한 채널의 훈련데이터, 테스트데이터를 이용하기 때문에 훈련과 인식 데이터 모두에 정규화 기법을 적용할 필요가 있는 것이다.

이러한 관점에서 훈련데이터에도 고유공간에서의 정규화를 적용할 수 있다. II 장에서 전체 훈련 데이터로부터 고유값과 고유벡터를 구하였다. 구한 고유벡터를 이용하여 켑스트럼 상의 특징 벡터를 고유공간 상의 특징으로 변환시킬 수 있다. 그림 1과 같이 고유공간 상에서의 특징 분포는 켑스트럼 상에 비해 더욱 상관관계가 없어지게 된다. 이러한 과정에서는 II장에서 구한 것처럼 고유값을 구할 필요가 없게 된다. 고유값은 단순히 분산과 관계가 있는, 특징값의 크기를 조절해주는 값이다. 그러나 이 고유값은 전체 훈련 데이터에 대한 통계량이므로 각각의 훈련 데이터에 적용을 시켜도 훈련데이터 하나의 샘플에 대해서는 평균 0, 분산 1인 정확한 분포를 가지게 만들지 못하므로 다시 분산을 1로 만드는 과정을 수행하여야 한다. 그러므로 고유벡터만을 이용하여 영역간의 변환을 시킨 후, MVN 과정을 수행하는 것이 중복된 과정 없이 효과적으로 수행하는 것이 된다. 이와 같은 훈련데이터의 정규화는 고유값을 몰라도 되는 장점을 가진다.

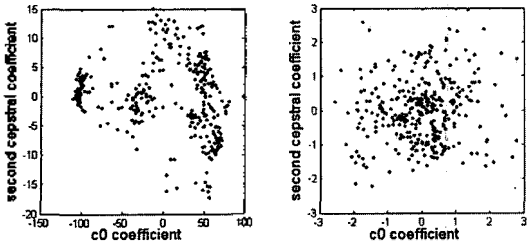
$$\tilde{X}_t = U^T X_t^c \tag{3}$$

U는 $E[X_t^c X_t^{cT}] = UVU^T$에서 나온 값이며 $\tilde{X}_t$는 고유벡터에 의해 변환된 고유공간 상에서의 특징이며 다음의 정규화 과정을 거치게 된다.

$$\tilde{Z}_t = \frac{\tilde{X}_t - mean(\tilde{X}_t)}{std(\tilde{X}_t)} \tag{4}$$

정규화된 벡터 $\tilde{Z}_t$는 고유벡터에 의해 다시 켑스트럼 상으로 변환된다.

$$X_t^c = U\tilde{Z}_t \tag{5}$$



(a) (b)

그림 1. Aurora2/testA/clean3/FAK_48Z66ZZA.08 데이타의 2차원 분포 (c0 와 c2) (a) 켑스트럼영역 (b) 고유공간

Fig. 1. Plot of 2-dimension (c0 and c2) features of testA's clean3 FAK_48Z66ZZA.08 in Aurora2. (a) cepstral domain (b) eigenspace.

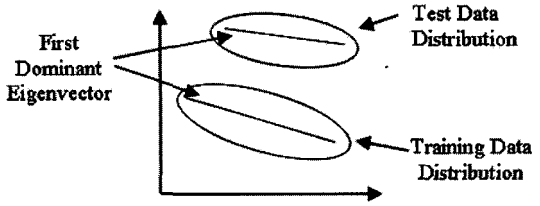


그림 2. 훈련데이터의 분포와 잡음에 의해 오염된 테스트 데이터의 도식적인 2차원 분포

Fig. 2. The schematic distribution of training and test data in a two-dimensional feature space.

마지막으로 이와 같은 과정을 인식 데이터에도 같이 적용시킨다.

3.2. 분리된 고유공간에서의 정규화 기법

기존의 고유공간에서의 정규화 기법에서는 오직 하나의 고유공간만을 이용하였다. 39차의 고차원의 데이터 분포를 39차의 하나의 고유공간을 이용하여 분포를 표현하고자 하였다. 그러나 고차원의 분포를 하나의 고유공간으로 표현하기에는 분포 표현의 세밀함에서 떨어지게 된다. 또한 39차 특징 벡터는 13차 static 특징, 1차 미분계수, 2차 미분계수의 3개의 그룹으로 분리가 가능하고 같은 그룹 내의 계수들은 같은 정의에 의해 나온 값이며 비슷한 특성을 지닌다. 이러한 관점으로 3개의 그룹을 독립적으로 처리를 하는 분리된 고유공간 (Separated Eigenspace Normalization, SEN) 을 제안하였다. 39차의 분포를 3개의 독립적인 13차 분포로 분리함으로써 전체 데이터 분포를 더욱 효율적으로 명확히 표현할 수 있다.

먼저 훈련데이터로부터 3개의 Λ 즉, Λ_{cep}, Λ_{del}, $\Lambda_{del-del}$ 를 구한다. 각각의 Λ 는 $E[X_i^{cep}X_i^{cep^T}]$, $E[X_i^{del}X_i^{del^T}]$, $E[X_i^{de-del}X_i^{del-del^T}]$로부터 구해지며 각각 static 특징, 1차, 2차 공분산 행렬을 나타낸다.

39차의 벡터를 3개의 Λ 를 이용하여 각각 다른 공간의 고유공간으로 변환 시킨 후 MVN을 수행한 다음, 다시 3개의 Λ 를 이용하여 켑스트럼 상으로 변환 시킨다.

3.3. 켑스트럼 상에서의 회전 기법

켑스트럼 상의 특징이 잡음환경에 의해 왜곡될 경우, 특징 분포의 평균과 분산만 변하는 것이 아니라 그림 2와 같이 전체적인 분포의 방향성이 변하게 된다. 즉, 약간의 회전이 일어나게 된다[5].

두 분포의 특징을 가장 잘 나타내주는 가장 큰 고유값에 해당하는 고유벡터를 회전시켜 서로의 방향을 같게 하여 두 분포의 차이를 줄일 수 있다. 이 방법은 Molau에 의해 제안되었으며 본 논문에서는 보다 더 잡음에 강인한 특징을 얻기 위해 이 이론을 이용하였다. Molau는 가장 큰 고유값에 해당하는 하나의 고유벡터부터 다수의 고유벡터까지 이용하는 이론을 제시하였다. 본 논문에서는 그 중에서 가장 영향력이 큰 하나의 고유벡터만을 이용하였다. 과정을 간단히 소개하면 다음과 같다.

먼저 II장에서처럼 전체 훈련 데이터의 고유값과 고유벡터를 얻는다. $\tilde{v}$ 를 훈련데이터 분포의 가장 큰 고유값에 해당하는 고유벡터라고 하고 v 를 인식할 하나의 발음에 해당하는 특징 분포의 가장 큰 고유값에 해당하는 고유벡터라 한다. 두 고유벡터사이의 회전각도 α 는 내적에 의해 다음과 같이 계산된다.

$$\alpha = \arccos(\tilde{v} \cdot v) \tag{6}$$

$$R = \begin{pmatrix} \cos(\alpha) & \sin(\alpha) \\ -\sin(\alpha) & \cos(\alpha) \end{pmatrix} \tag{7}$$

R 은 회전 행렬을 나타내며 회전은 2차원 평면상에서 이루어져야 한다. 두 고유벡터는 서로 직교하지 않기 때문에 Gram-Schmidt 알고리즘을 $\tilde{v}$ 에 적용시켜 회전할 평면에 위에 놓여있는 단위직교 기저 벡터 $\hat{v}$ 를 얻는다.

$$\hat{v} = \frac{v - (\tilde{v} \cdot v) \cdot \tilde{v}}{\| v - (\tilde{v} \cdot v) \cdot \tilde{v} \|} \tag{8}$$

$\hat{v}$ 을 구한 후, 인식 데이터 벡터를 $\hat{v}$ 와 $\tilde{v}$ 로 펼쳐지는 평면에 투영시킨다. 투영 행렬은 다음 식과 같이 두 벡터로 구성된다.

$$J = (\tilde{v}, \hat{v}) \tag{9}$$

마지막으로 단위행렬 I 가 포함된 보정행렬 $I - JJ^T$ 로 투영 시 잃어버린 차원들을 복구 시켜 준다. 전체 과정을 포함한 최종적으로 얻어지는 행렬 Q 는 다음 식과 같다.

$$Q = JRJ^T + I - JJ^T \tag{10}$$

켑스트럼 상에서 회전된 특징 $\hat{X}$ 는 식 11과 같이 얻어진다.

$$\hat{X}_t^c = QX_t^c \tag{11}$$

그림 3은 제안한 방식의 전체 구성도를 나타낸다.

3.3.1. 실험 및 결과

본 논문에서는 객관적인 성능 평가를 위해 ELRA (European Language Resources Association)의 Aurora2.0에서 제공하는 평가 방식을 그대로 따랐으며 특징 추출은 ETSI (European Telecommunications Standards Instituted) 표준의 DSR (Distributed Speech Recognition) 방식을 따랐다[6,7].

Aurora2.0의 데이터베이스는 연속 영어 숫자음으로 이루어져 있고 각각 4종류의 배경잡음 환경으로 이루어진 SetA와 SetB가 있으며 채널 왜곡이 반영된 SetC 가 있다. 각각의 Set 에는 신호 대 잡음비 (SNR)에 따라 1,001개의 샘플이 있으며 회전 기법을 이용하기 위해 Aurora2 에서 제공하는 SetA 와 같은 종류의 잡음원이 들어간 훈련데이터를 이용하였다. 이 훈련데이터는 각각의 신호 대 잡음비에 따라 422개의 샘플이 있다. 각 샘플은 일정하지는 않으나 20~50ms 의 앞뒤 묵음 구간을 가진다. 고유공간은 큰 분산을 가지는 요소가 있을 때 일관성 있게 잘 정의 되므로 본 실험에서는 13차 static 특징 중 로그에너지 대신 c0 계수를 사용하였다.

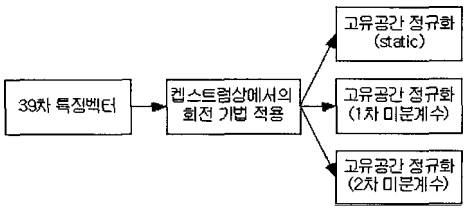


그림 3. 제안한 기법의 전체 흐름도
Fig. 3. Block diagram of proposed scheme.

먼저 깨끗한 환경에서 훈련시킨 음향 모델을 이용하여 베이스라인 실험을 하였다. 일반적으로 static 특징뿐만 아니라 전체 특징에 대해 정규화를 하는 것이 더욱 효과적인 것으로 알려져있지만 확실한 검증을 위해 본 실험에 앞서 정규화된 static으로부터 미분계수들을 구하는 것과 static, delta, delta-delta 전체에 정규화를 한 것의 비교실험을 Aurora2.0 car noise 에 대해서만 하였다. 그리고 본 실험으로 켑스트럼 상에서의 MVN 과 고유공간 상에서의 MVN을 비교하기 위하여 인식 데이터에만 MVN을 적용하는 실험을 한 후, 훈련데이터, 인식 데이터 모두 MVN을 적용하는 실험을 하고 같은 방식으로 분리된 고유공간에서의 MVN을 실험하였다. 마지막으로 분리된 고유공간에서의 MVN 과 함께 켑스트럼 상에서의 회전 기법을 적용하여 실험하였다. 실험 결과표에 나열한 약어들은 다음과 같다.

1) EIG_static : static 으로부터 고유정규화 한 후 미분 계수 구하는 기법 (13차 정규화)

2) MVN : 켑스트럼 상에서의 평균과 분산 정규화 기법

3) EIG : 고유공간 상에서의 평균과 분산 정규화 기법 (39차 정규화)

4) SEN : 분리된 고유공간 상에서의 평균과 분산 정규화 기법

5) SEN_ro : 켑스트럼 상에서의 회전 기법 + 분리된 공간상에서의 정규화 기법 (각각의 인식 발음에서 얻어진 고유벡터 이용)

6) SEN_ro_20 : 켑스트럼 상에서의 회전 기법 + 분리된 공간상에서의 정규화 기법 (인식 set 과 같은 set 의 20dB 훈련 데이터로부터 고유벡터를 얻음)

먼저 실험 결과표 1을 통해 예상대로 모든 특징에 대해 정규화를 해주는 방식 (EIG) 이 static 에만 정규화를 해준 후 미분 계수들을 구하는 방식 (EIG_static) 보다

표 1. Aurora2.0의 자동차 잡음에 대한 정규화방식에 따른 단어 인식률 (%) (인식 데이터의 분포를 훈련데이터의 분포와 같게함)
Table 1. Word accuracy for different normalization scheme of the car noise condition in Aurora2.0(%) (normalization of the test data to the distribution of training data).

	Baseline	EIG_static	EIG
Clean	98.84	98.60	98.90
20dB	96.42	95.50	97.08
15dB	87.62	90.52	95.20
10dB	61.71	81.03	89.71
5dB	26.87	60.87	75.69
0dB	10.38	28.72	46.26
-5dB	8.41	10.50	18.40
Avg	56.60	71.33	80.79

표 2. Aurora2.0의 자동차 잡음에 대한 제안한 방법의 단어 인식률(%) (인식 데이터의 분포를 훈련데이터의 분포와 같게함)
Table 2. Word accuracy for the proposed scheme of the car noise condition in Aurora2.0(%) (normalization of the test data to the distribution of training data).

	Baseline	MVN	EIG	SEN	SEN_ro	SEN_ro_20
Clean	98.84	98.81	98.90	98.96	98.30	98.90
20dB	96.42	96.93	97.08	97.08	95.85	97.23
15dB	87.62	94.72	95.20	95.08	93.62	95.35
10dB	61.71	88.70	89.71	90.13	87.21	90.46
5dB	26.87	73.34	75.69	77.93	73.46	78.44
0dB	10.38	43.51	46.26	49.12	44.83	49.90
-5dB	8.41	15.63	18.40	19.65	16.76	19.80
Avg	56.60	79.44	80.79	81.87	78.99	82.28

표 3. Aurora2.0의 자동차 잡음에 대한 제안한 방법의 단어 인식률(%) (훈련데이터와 인식데이터 모두의 정규화)
Table 3. Word accuracy for the proposed scheme of the car noise condition in Aurora2.0(%) (normalization of both training and test data).

	Baseline	MVN	EIG	SEN	SEN_ro	SEN_ro_20
Clean	98.84	98.90	98.90	99.11	98.99	99.05
20dB	96.42	97.82	97.79	98.03	97.35	98.09
15dB	87.62	95.94	95.94	96.09	95.14	96.21
10dB	61.71	90.46	90.67	90.22	88.73	90.61
5dB	26.87	75.93	76.35	76.02	73.07	76.68
0dB	10.38	46.08	49.21	49.30	43.99	49.87
-5dB	8.41	17.21	19.68	19.42	16.40	19.86
Avg	56.60	81.25	81.99	81.93	79.66	82.29

좋은 성능이 나오는 것을 확인 할 수 있었다.

실험 결과표 2와 3을 통해 고유공간에서의 정규화 기법이 켑스트럼 상에서의 정규화 기법보다 더욱 효과적임을 알 수 있다. 이것은 벡터의 각 차원간의 상관관계가 줄어들 때, 독립적으로 시행하는 분산 정규화 기법이 더욱 효과적임을 입증한다. 표 2에서 인식 데이터의 분포를 훈련데이터의 분포와 같게 해줄 때, 제안한 SEN 기법이 EIG 기법 보다 1.08% 의 단어 인식률이 증가함을 확인할 수 있다. 표 4의 모든 set 의 평균 단어 인식률 결과에서 보듯이 제안한 방법, SEN이 기존의 방법, EIG 보다 효과적임을 알 수 있다. 특히 낮은 신호 대 잡음비에서 더욱 인식률 향상이 큰 것을 확인할 수 있다.

표 3과 2, 표 5와 4를 비교할 때, 훈련데이터와 인식데이터 모두 정규화해주는 기법을 이용하는 것이 인식데이터를 훈련데이터의 분포와 같게 해주는 정규화 기법을 이용하는 것보다 더 좋은 성능을 얻을 수 있음을 확인할 수 있다. 그러나 setC의 SEN의 결과만 약간 성능이 하락되었다. 또한 SEN기법을 훈련데이터와 인식 데이터 모두에 적용할 때, 그 인식률 증가는 다른 기법들에 비해 상대적으로 작게 나왔다. 이 결과에 대한 정확한 원인을 현재 연구 중이다.

켑스트럼 상에서의 회전기법의 경우, 각각의 인식 발음으로부터 가장 영향력이 큰 고유벡터를 구하여 적용하였을 때 (SEN_ro), 오히려 성능이 하락되었다. 성능 하락의 요인은 하나의 샘플 데이터로 공분산을 추정하기에는 데이터양이 너무 작기 때문이라고 분석된다. 그 결과, 공분산으로부터 구한 고유벡터도 안정적으로 얻기 불가능하다. 일반적으로 PCA를 이용한 데이터 분포 분석 및 영역 변환은 충분한 양의 데이터를 필요로 한다. 이러한 관점에서 Aurora2.0에서 제공하는 각각의 인식 set과 같은 잡음 환경의 훈련데이터로부터 각 신호 대 잡음비에 따라 고유벡터를 얻었다

처음에는 각 인식 환경의 신호 대 잡음비와 같은 훈련데이터의 고유벡터를 이용하였을 때 인식률 향상을 기대했으나 성능 향상을 보장하지 못하는 것을 관찰하였다. 신호 대 잡음비가 감소할수록 성능 향상률이 줄어들고 약간의 성능하락까지 발생하기도 하였다. 반면, 20dB의 훈련데이터로부터 구한 고유벡터를 모든 신호 대 잡음비의 인식 데이터에 적용하였을 때 가장 높은 성능 향상을 얻었다. 낮은 신호 대 잡음비에서는, 즉 깨끗한 신호가 잡음에 의해 많이 오염될수록 켑스트럼 상에서의 특징의 분포는 분산이 줄어들어 압축된다. 결과적으로 큰 분산에 의한 분포의 식별성이 줄어들어 안정적인 고유벡터를 얻기 힘들기 때문이라고 분석된다. 이 같은 이유로 20dB의 고유벡터를 이용한 결과가 가장 높은 성능을 내는 것이다. Aurora2.0에서 제공하는 20dB의 훈련 데이터로부터 잡음의 특성, 즉 회전 각도를 분석하여 안정적으로 회전현상을 보상해 줄 수 있었다. Aurora2.0에서 SetA에 대해서만 훈련 데이터를 제공하는 이유로 켑스트럼 상에서의 회전 기법은 SetA에 대해서만 수행되었다. 표

표 4. Aurora2.0의 모든 set 에 대한 제안한 방법의 평균 단어 인식률(%) (인식 데이터의 분포를 훈련데이터의 분포와 같게함)

Table 4. Average word accuracy for the proposed scheme of all data set in Aurora2.0(%) (normalization of the test data to the distribution of training data).

	Baseline	MVN	EIG	SEN	SEN_ro_20
SetA	59.58	77.90	79.81	80.27	80.87
SetB	57.18	79.49	81.21	81.77	-
SetC	66.81	77.90	78.96	79.32	-

표 5. Aurora2.0의 모든 set 에 대한 제안한 방법의 평균 단어 인식률(%) (훈련데이터와 인식데이터 모두의 정규화)

Table 5. Average word accuracy for the proposed scheme of all data set in Aurora2.0(%) (normalization of both training and test data).

	Baseline	MVN	EIG	SEN	SEN_ro_20
SetA	59.58	80.27	80.43	80.51	81.08
SetB	57.18	81.84	82.87	82.49	-
SetC	66.81	79.32	79.23	79.10	-

4, 5에서 보듯이 세 가지 기법들을 조합하여 최고의 성능향상을 얻을 수 있었다.

IV. 결 론

본 논문에서는 잡음에 강인한 음성 인식 성능을 얻기 위해 고유공간에 기반을 둔 새로운 정규화 기법을 제안하였다. 하나의 고유공간을 세 개의 분리된, 독립적인 고유공간으로 나눔으로써 향상된 인식 결과를 얻었으며 이것은 제안한 방식으로 데이터의 분포를 더욱 효과적으로 표현할 수 있음을 의미한다. 또한 인식데이터 뿐만 아니라 훈련데이터도 고유공간에서 정규화가 가능함을 보였으며 이 방법으로 조금 더 향상된 결과를 얻을 수 있었다. 마지막으로 켑스트럼 상에서의 회전기법을 이용하여 훈련데이터와 인식데이터의 불일치를 보상하고자 하였다. 비록 원하는 고유벡터를 얻기 위해 제공된 훈련데이터를 이용하였지만 잡음에 대한 소량의 통계량만 알고 있으면 회전에 의한 왜곡을 보상할 수 있음을 확인하였다. 이러한 전체적인 알고리즘을 통하여 Aurora2.0의 자동차 잡음 환경에 대해 82.29%까지의 단어인식률을 얻을 수 있었다.

참 고 문 헌

1. X. Huang, A. Acero and H. Hon, *Spoken Language Processing*, (Prentice Hall PTR, 2001).
2. P. Jain and H. Hermansky, "Improved Mean and Variance Normalization for Robust Speech Recognition", Proc.of ICASSP, 2001.
3. Kaisheng Yao, Erik Visser, Oh-Wook Kwon, and Te-Won Lee, "A Speech Processing Front-End with Eigenspace Normalization for Robust Speech Reconition in Noisy Automobile Environments", Eurospeech 2003, 9-12, 2003.
4. A. Vinciarelli and S. Bengio "Offline Cursive Word Recognition using Continuous Density Hidden Markov Models trained with PCA or ICA Features", Proc.of 16th International Conference on Pattern Recognition, 3, 81-84, 2002.
5. Sirko Molau, Daniel Keysers and Hermann Ney, "Matching Training and Test data Distributions for Robust Speech Reconnition", Speech Communication, 41 (4), 579-601, 2003.
6. H. G. Hirsch and D.Pearce, "The AURORA Experimental Framework for the Performance Evaluations of Speech Recognition Systems under Noisy Conditions", ISCA ITRW ASR2000, 2000.
7. ETSI standard document, *Speech Processing, Transmission and Quality aspects (STQ); Distributed speech recognition; Front-end feature extraction algorithm; Compression algorithms*, ETSI ES 201 108 v1.1.3 (2000-04), 2000.

저자 약력

• 이 윤 재 (Yoonjae Lee)

2003년 2월:고려대학교 전기전자전파공학부 (공학사)
2003년 2월~현재:고려대학교 전자컴퓨터학과 석사과정 재학중
*주관심 분야: 신호처리, 음성 인식, 잡음 처리

• 고 한 석 (Hanseok Ko)

1982년 5월:미국 카네기 멜론 대학교 전기공학 (공학사)
1986년 5월:미국 메릴랜드 대학교 시스템 공학(공학석사)
1988년 5월:미국 존스 홉킨스 대학교 전기공학 (공학석사)
1992년 5월:미국 카톨릭 대학교 전기공학 (공학박사)
1995년 3월 ~ 현재:고려대학교 전자컴퓨터공학과 교수
*주관심 분야 : 영상 및 음성 신호처리, 패턴 인식, 데이터 융합