

다양한 잡음 환경하에서 환경 군집화를 통한 화자 및 환경 동시 적응

Simultaneous Speaker and Environment Adaptation by Environment Clustering in Various Noise Environments

김 영 국*, 송 화 전**, 김 형 순**
(Young Kuk Kim*, Hwa Jeon Song**, Hyung Soon Kim**)

*LG전자기술원, **부산대학교 전자전기공학부
(접수일자: 2009년 7월 14일; 채택일자: 2009년 8월 4일)

본 논문에서는 eigenvoice 방식에 기반하여 다양한 잡음 환경에 강인한 고속 화자 적응 방법을 제안하였다. 제안된 방법은 잡음 제거 기술과 환경 군집화 방법을 기반으로 한다. 그러나, 잡음 제거 기술을 통해 잡음을 제거한 후에도 여전히 잔여 잡음이 존재하므로 비음성 구간의 cepstrum 평균을 사용하여 잡음 환경별로 화자 적응 데이터를 분류한 후 각각의 환경별로 환경 모델을 구성한다. 이러한 환경 군집화를 적응데이터에 대해 구성한 후 테스트 음성이 입력되면 군집화된 모델 중에서 인식 데이터와 가장 유사한 복수의 환경별 군집화된 화자 적응 모델을 구한 후 이들의 가중합을 통해 화자 적응을 수행하는 방법이다. 제안된 방법은 적응 및 평가를 통해 화자 독립 모델을 사용한 경우에 비해 40~59% 인식 오류 감소율을 얻었다.

핵심용어: 음성인식, 화자적응, 환경 군집화, Eigenvoice, Eigen-environment

투고분야: 음성처리 분야 (2,5)

This paper proposes noise-robust fast speaker adaptation method based on the eigenvoice framework in various noisy environments. The proposed method is focused on de-noising and environment clustering. Since the de-noised adaptation DB still has residual noise in itself, environment clustering divides the noisy adaptation data into similar environments by a clustering method using the cepstral mean of non-speech segments as a feature vector. Then each adaptation data in the same cluster is used to build an environment-clustered speaker adapted (SA) model. After selecting multiple environmentally clustered SA models which are similar to test environment, the speaker adaptation based on an appropriate linear combination of clustered SA models is conducted. According to our experiments, we observe that the proposed method provides error rate reduction of 40~59% over baseline with speaker independent model.

Keywords: Speech recognition, Speaker adaptation, Environment clustering, Eigenvoice, Eigen-environment

ASK subject classification: Speech Signal Processing (2,5)

I. 서론

훈련 환경과 테스트 환경이 다른 경우 음성인식 시스템에서 심각한 성능 저하가 발생한다. 이러한 환경 불일치는 발성 환경의 차이, 화자간의 차이 등에 의해 발생하며, 화자 적응 방식은 이를 보상하기 위한 대표적 방법이다. 특히 화자 적응 방법 중 eigenvoice 방법 [1]은 기저

모델 (basis model) 수를 조절할 수 있으므로 maximum a posteriori (MAP) [2] 또는 maximum likelihood linear regression (MLLR) [3] 보다 추정할 파라미터 수가 적어 고속 화자적응에 유리한 것으로 알려져 있다. 그러나 eigenvoice 기법에 의해 적응된 모델은 화자 공간 (speaker space)에서 환경이 다른 새로운 화자에 대한 정확한 위치를 제공하지 못하며 환경이 달라진 경우에 이를 보상할 수 있는 적응 방식이 필요하다.

본 논문에서는 인식시 훈련 및 화자 적응시 관찰되지 않은 다양한 잡음이 존재하는 환경에서 군집화에 기반을

책임저자: 송 화 전 (hwajeon@pusan.ac.kr)
609-735 부산시 금정구 잠전동 산30 부산대학교 전자전기공학부
음성통신실험실
(전화: 051-510-1704; 팩스: 051-516-1946)

둔 잡음에 강인한 고속 화자적응 방식을 제안하였다. 우선 적응 데이터와 테스트 데이터의 불일치를 감소시키기 위해 특징 공간 (feature space)에서의 잡음 제거 방법으로 eigenvoice 방법을 응용한 아이겐 환경 (eigen-environment) 방법 [4] (또는 eigen-noise방법) 을 사용하였다. 다음으로 본 논문에서 제안한 모델 공간 (model space)에서의 환경 군집화에 기반을 둔 고속 화자적응 방식을 수행하였다. 특히 훈련이나 적응시 전혀 관측되지 않은 인식환경의 음성에 대해서 환경 군집화를 통해 구축된 복수의 환경 모델을 선택하여 이들의 가중합으로 인식 환경을 적절하게 표현함으로써 다양한 잡음 환경에서의 인식성능 향상을 도모한다.

II. Eigenvoice 기반의 잡음 환경에서의 고속화자적응

화자적응의 한 방법으로 고속 화자적응에 유리한 화자 군집화 방식의 하나인 eigenvoice 적응방법이 널리 사용되고 있다. Eigenvoice에서는 새로운 화자는 다음 식과 같이 K 개의 eigenvoice로 나타낼 수 있다.

$$\hat{\mu} = \mathbf{e}(0) + \sum_{k=1}^K w(k)\mathbf{e}(k) \quad (1)$$

여기서 $\mathbf{e}(0)$ 는 T 명의 SD 모델의 평균을 나타내고, 가중치 $w(k)$ 는 새로운 화자의 적응 데이터를 이용하여 MLED (Maximum Likelihood Eigen-Decomposition) 방법에 의해 추정할 수 있다 [1].

새로운 화자의 모델이 잡음 등의 이유로 인해 화자 공간에서 훈련환경과 전혀 다른 위치에 있는 경우 eigenvoice에 의해 생성된 K 차원 공간으로의 투영에 의해 적응 모델을 생성할 때 큰 투영 오류가 발생한다. 새로운 화자에 대한 음성은 화자의 특성을 적응하기 위해 eigenvoice 및 환경을 보상하기 위해 보상 벡터의 기저 벡터 (basis vector)의 가중 합으로 나타낼 수 있으며 식 (1)에서 상태 s , 믹스처 m 인 평균 부벡터는 다음과 같이 확장될 수 있다 [4].

$$\hat{\mu}_m^{(s)} = \mathbf{e}(0)_m^{(s)} + \sum_{k=1}^K w(k)\mathbf{e}(k)_m^{(s)} + \sum_{d=1}^D b(d)\mathbf{I}(d) \quad (2)$$

여기서, $\mathbf{I}(d) = [\delta(d-1), \dots, \delta(d-D)]^T$ 이며, $\delta(x)$ 는 delta 함수를 뜻한다. 따라서 eigenvoice 수가 D 개만큼 증가하

는 형태가 되며, 이 가중치도 MLED를 통하여 동시에 추정할 수 있다.

III. 다양한 잡음 환경하에서의 고속화자적응

2절의 잡음 보상방식의 경우는 적응 및 인식 데이터의 잡음 환경이 일정한 경우에 효과적이다. 즉, 배경 잡음이 계속해서 변하는 환경에서 화자 적응을 수행하는 경우는 2절의 잡음 보상방식으로는 효과적으로 대처하기가 어렵다. 따라서 유사한 환경별로 적응 데이터를 분류하고 각 환경별로 적응모델을 구성한 후, 테스트시 유사한 환경 모델을 선택해서 인식하는 것이 성능 향상을 위한 한가지 해결책으로 제시될 수 있다. 먼저 그림 1에 다양한 잡음 환경 하에서 화자적응의 성능을 향상시키기 위한 방법으로 본 논문에서 제안한 환경 군집화에 기반을 둔 화자적응 시스템의 전체 프로세스를 나타내었다. 그림 1에서 보면 우선 크게 오프-라인 (off-line)과 온-라인 (on-line) 두 가지 단계로 나눌 수 있다. 오프-라인 단계에서는 깨끗한 DB를 이용해 화자 독립 (Speaker Independent; SI) 모델을 만들며, 적응용 잡음 음성 DB를 이용해 먼저 잡음 제거 기술을 적용한 후 화자적응을 수행하여 SA (Speaker Adapted) 모델을 만들게 된다. 여기서 적응용 잡음 음성 DB를 "Environment Classification" 모듈을 통해 유사한 잡음 데이터들로 분류할 수 있고, 분류된 적응용 잡음 음성 DB를 이용해 잡음 환경별 SA 모델을 만든다. 그리고 나서 온-라인 단계에서 실제 테스트 음성이 입력되면 잡음 제거 후 가장 유사한 환경을 선택하여 이와 관련된 모델을 이용하여 인식을 수행한다. 다음에 각각의 모듈에 대해 상세히 설명한다.

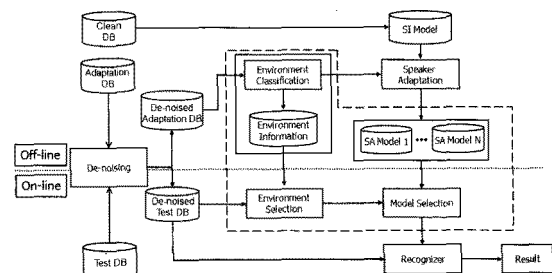


그림 1. 잡음 처리 및 환경 군집화에 기반한 고속 화자적응
Fig. 1. Fast speaker adaptation using de-noising and environment clustering.

3.1. 잡음 제거

우선 적응 데이터와 테스트 데이터의 불일치를 감소시키기 위한 잡음제거 단계로 기존의 다양한 잡음 제거 방식들이 "de-noising" 모듈에 사용될 수 있다. 본 논문에서는 eigenvoicc 방법을 응용한 아이겐 환경방법 [6]을 사용하였다. 이 방식은 eigenvoicc에서 사용한 방식과 유사하게 각각의 잡음 환경에 대해 clean 환경과의 차이를 K 개의 eigenvector의 선형 결합으로 표현하고, 새로운 환경이 입력으로 들어올 경우 기존의 환경들로 새로운 환경을 표현할 수 있는 방식이다.

3.2. 환경 분류 및 선택

환경 군집화를 위한 전제 조건은 화자 적응을 위한 적응 데이터가 유사한 환경 별로 잘 군집화 되어 있어야 한다. 따라서 좋은 군집화 결과를 얻기 위해 환경 특성 변별력 있는 특징 벡터에 대한 연구가 필요하다. 본 논문에서는 환경 군집화에 사용하기 위한 특징 벡터로 캡스트럼 평균을 다음과 같이 두 가지 형태로 나누어 사용하였

다. 우선 전체 음성의 캡스트럼 평균을 사용하여 환경 별 군집화 성능을 비교하는데 사용하였고, 두 번째로 비음성 구간의 캡스트럼 평균을 이용하여 환경 군집화 성능을 비교하는데 사용하였다. 비음성 캡스트럼은 음성의 시작과 끝 부분의 10 프레임을 사용하였다.

환경 군집화를 위한 이들 두 가지 특징벡터의 성능을 비교하기 위해서 적응 데이터로 사용된 SNR 20 dB와 SNR 10 dB를 가지는 자동차, 배틀, 로봇 잡음에 대해서 군집화 결과를 비교해 보았다. 그림 2에 자동차, 배틀, 로봇 잡음이 부가된 적응 데이터의 캡스트럼 평균의 분포를 그렸다. 그림에서 기호 (o, *), (□, ◇), (⊙, ⊕) 등은 각각 자동차, 배틀 그리고 로봇의 SNR 20 dB와 SNR 10 dB의 캡스트럼 평균을 나타낸다. 그림 2에서 1차에서 4차까지의 캡스트럼 평균의 분포를 보면, 음성 구간보다 비음성 구간의 캡스트럼 평균을 이용하는 것이 동일한 환경 별로 군집화가 잘 되는 것을 볼 수 있다. 따라서 이후 실험에서는 비음성 구간의 캡스트럼 평균을 사용해서 화자 군집화를 수행 하였다.

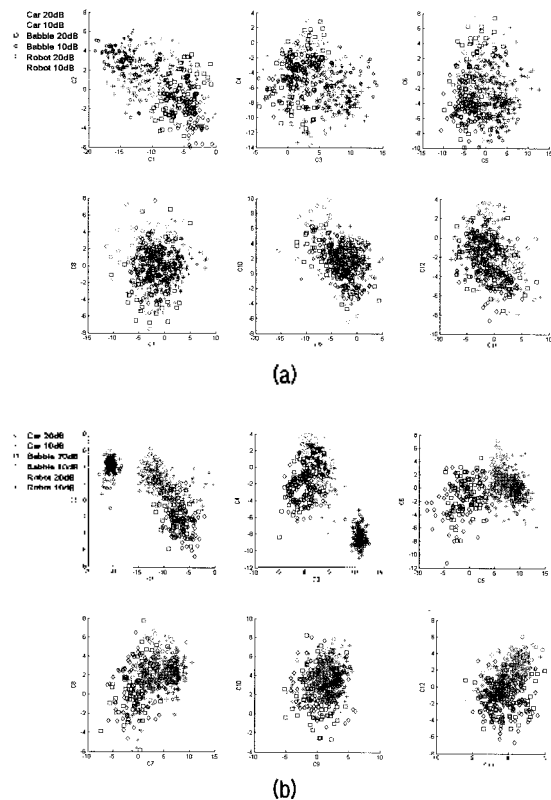


그림 2. 잡음제거 후의 자동차, 배틀, 로봇 적응 데이터의 음성/비음성 구간의 캡스트럼 평균 분포 (a) 음성 구간 (b) 비음성 구간

Fig. 2. Cepstral mean distributions of adaptation data with car, babble and robot noise after de-noising (a) speech segments (b) non-speech segments.

3.3. 군집수 결정 방법

환경 군집화에서는 적응 데이터를 그룹화 하기 위해서 오프-라인 단계에서 벡터 양자화 (Vector Quantizer; VQ) 분류기를 사용하여 유사한 환경 별로 적응 데이터를 미리 분류해서 환경별 SA 모델을 만들어 둔다. 물론 VQ 분류기 외에 다양한 방법의 군집화 기법을 도입할 수 있으나, 적응 데이터의 수가 소량이며 또한 계산량 등을 고려하여 간단하지만 좋은 성능을 보이는 VQ 분류기를 선택하였다. 하지만 실제 환경에서는 환경 군집화를 위한 최적의 군집 수를 알 수 없기 때문에 적절한 VQ 코드북 사이즈에 대한 결정이 필요하다. 이를 위해 본 논문에서는 실험에 사용된 다양한 잡음 적응 데이터에 대하여 코드북 사이즈를 2개에서 5개까지 바꾸어 가면서 인식 실험을 한 결과 3개일 때가 가장 좋은 성능을 보였다.

3.4. 가중합 방식의 SA 모델을 이용한 방법

인식시 군집화 모델을 이용하는 한가지 방법으로 입력 음성의 환경과 가장 유사한 환경을 나타내는 환경 모델을 선택하여 인식시 사용하는 방식이다. 그러나, 이 방식의 경우는 인식환경이 적응시의 환경과 전혀 다른 경우에는 인식 성능이 하락될 수 있다. 따라서, 본 논문에서는 테스트 데이터와 적응 데이터와의 유사도 정보를 이용해서 군집화된 적응 데이터 별로 생성된 SA 모델들의 가중 합

형식으로 인식에 사용하는 방식을 제안한다.

그림 1의 오프-라인 과정에서 다양한 잡음이 존재하는 적응 데이터를 VQ 군집화를 통해 N 개의 군집을 생성한다. 그런 다음 테스트 데이터가 입력으로 들어오면 군집화된 적응 데이터의 VQ 코드북과의 유사도를 계산하여, 테스트 데이터와 유사한 K ($K \leq N$)개의 적응 데이터로부터 생성된 SA 모델들의 가중 합 형식으로 인식에 사용한다. 자세한 방법은 다음과 같다.

환경 군집화를 통해 유사한 환경 별로 분류된 각각의 적응 데이터를 이용해 eigenvoice 화자적응을 한다. 그리고 이를 통해 업데이트 된 모델의 평균을 $\hat{\mu}_1, \dots, \hat{\mu}_k, \dots, \hat{\mu}_N$ 이라 둔다. 그런 다음 테스트 데이터와 군집화된 환경과의 거리를 구해서 다음과 같이 유사한 정도에 따라서 SA 모델을 가중 합의 형태로 사용한다.

$$\hat{\mu}_{new} = \sum_{k=1}^K \hat{\mu}_k w_{new}(k) \tag{3}$$

여기서 $\hat{\mu}_k$ 는 군집화를 통해 분류된 k 번째 적응 데이터 군집을 이용해서 업데이트된 SA 모델의 평균이고, $w_{new}(k)$ 는 환경별 가중치 (weight)를 의미한다. 그리고 $\hat{\mu}_{new}$ 는 SA 모델들의 가중 합으로 업데이트 된 모델의 평균을 나타낸다. 본 논문에서는 fuzzy k-means 알고리즘에서 거리기반 가중치 추정함수인 membership 함수 [9]를 사용하여 유사한 K 개 그룹의 가중치를 선정하였다. Membership 함수는 다음과 같이 정의된다.

$$w_{new}(k) = \frac{1}{\sum_{i=1}^K \left[\frac{\|x - C_k\|}{\|x - C_i\|} \right]^\alpha} \tag{4}$$

여기서 K 는 클러스터 개수를 의미하며, C_i 는 VQ 코드북에서 i 번째 중심값을 의미하고, x 는 환경 군집화에 사용된 특징벡터 (캡스트럼 평균)을 의미한다. α 는 가중치 변수로 0에서 ∞ 사이의 값을 가진다. α 가 0인 경우 K 개의 SA 모델은 모두 $1/K$ 의 값으로 동일한 가중치를 가지며, α 가 ∞ 인 경우 가장 큰 가중치를 가지는 하나의 SA 모델만을 사용하게 된다. 그리고 모든 가중치의 합은 다음과 같이 1이 된다.

$$\sum_{k=1}^K w_{new}(k) = 1 \tag{5}$$

IV. 실험 및 결과

4.1. 실험 환경

본 논문에서 사용한 음성 특징 파라미터로는 20 ms 해밍창 (Hamming window)를 10 ms씩 이동시키면서 12차 MFCC, delta 및 delta-delta를 구하여 총 36차의 파라미터를 사용하였다. 그리고 46 유사음소 (PLU) set을 기본으로 tree-based clustering를 사용한 triphone 은닉 마르코프 모델 (hidden Markov model, HMM)을 기본 모델로 사용하였으며 모델 당 상태수는 3개 및 상태당 믹스처 수는 1개로 하였다. 본 논문에서 사용한 tied state 수는 4,050개이다. 가변어휘 음성 인식기의 훈련을 위해서는 ETRI에서 구축한 음소연 최적화 단어 (Phonetically Optimized Words, POW) 음성 데이터베이스 중에서 남성 40명분의 음성 데이터베이스를 이용해서 모델을 훈련시킨 후 40명의 화자 모델을 SI로부터 MAP을 사용하여 구성한 후 eigenvoice를 구성하였다. 그리고 화자적응 및 인식 실험을 위해서는 훈련용 POW 3,848 음성 데이터베이스와는 어휘 내용이 다른 SiTEC에서 구축한 452 균일 음소 분포 단어 (Phonetically Balanced Words, PBW) 데이터베이스의 일부를 사용하였다. 남성화자 10명의 1회 발성분에 대해서 처음 50개 단어 수를 1, 5, 10, 20, 30, 40, 50개로 늘려가면서 적응에 사용하였고, 나머지 중 400개 단어를 성능 평가에 사용하였다.

잡음 음성 데이터를 생성하기 위해 Aurora 2 DB 제작에서 사용한 방법과 동일한 방식 [5]으로 배블 (babble), 자동차, 모뎀 그리고 기차역 잡음을 이용하였다. 배블 잡음의 경우 NOISEX 92 DB를 사용하였으며, 자동차 잡음은 SiTEC 자동차 잡음 DB (CarNoiseDB01)의 일부로 콘크리트 포장된 고속도로를 100 km/h로 주행하는 자동차에서 녹음한 것이다. 그리고 로봇 잡음은 실험실에서 테모용 로봇에서 직접 녹음한 것을 사용하였다. 기차역 잡음은 'FindSounds' 사이트에서 구한 음원으로 기차역에서의 기차 소리와 사람 소리등이 섞여 있는 것을 사용하였다. 적응 데이터의 경우 자동차, 배블 그리고 로봇 잡음을 신호 대 잡음비 (Signal-to-noise ration, SNR)이 20 dB와 10 dB가 되도록 부가시켜 사용하였다. 그리고 테스트 데이터의 경우 적응 환경과 동일한 환경을 위해서 자동차, 배블 그리고 로봇 잡음을 이용한 SNR 20dB와 10dB의 데이터와 동일한 잡음에 SNR 15dB와 5dB를 추가한 것도 사용하였으며, 적응환경과 다른 환경을 만들기 위해 기차역 잡음을 사용한 SNR 20 dB, 15 dB, 10 dB 및

5 dB의 데이터를 사용하였다. 아이젠 환경을 얻기 위해 stereo DB를 이용하는 방식을 사용하고, 사용된 DB는 clean DB와 자동차, 배블 그리고 로봇 잡음이 부가된 20 dB와 10 dB의 SNR을 가지는 잡음 음성 DB를 사용하였다.

4.2. 실험 결과

그림 3에서는 적응 데이터 및 테스트 데이터가 동일한 잡음을 사용하였지만 적응 데이터는 SNR 20 dB와 10 dB 인 배블, 자동차, 그리고 로봇 잡음을 동일 비율로 사용하였고, 테스트 데이터에 15 dB와 5 dB가 추가된 경우에 대한 각각의 SNR별 실험 결과들에 평균 성능을 나타낸다. 이 경우 SI모델을 사용한 baseline의 인식률은 60.90%이다. 그리고 그림 4에는 적응 데이터로 자동차, 배블 그리고 로봇 잡음을 SNR 20 dB와 10 dB를 사용하였고, 테스트 데이터에는 기차역 잡음을 SNR 20 dB, 15 dB, 10 dB 그리고 5 dB를 사용한 실험 결과를 나타낸다. 이 경우 SI모델을 사용한 baseline의 인식률은 42.3%이다. 각각의 그림에서 "EV"는 eigenvoice 적응 방식을 이용한 화자 적응 방식이다. 그리고 "EV+Bias"는 바이어스 보상을 통한 환경 화자 동시 적응 방식이며, "EV+Bias+Clustering (1-best)"은 식 (4)에서 $\alpha = \infty$ 인 경우 (즉, 가장 유사한 1개의 모델을 사용한 경우)를 뜻하고, "EV+Bias+Clustering (W.S)"는 다중 모델 가중합을 사용한 경우를 뜻한다. 여기서, α 는 실험적으로 결정하였다. 그리고 "EN" 잡음 제거를 위한 전처리 과정으로 아이젠 환경 방식을 적용하는 것을 의미한다.

적응 환경과 테스트 환경에 동일한 잡음을 사용한 그림 3을 보면 적응 데이터가 20개 미만의 경우 "EN + EV + Bias + Clustering (1-best)" 방식에 비해 "EN + EV + Bias + Clustering (W.S)" 방식이 우수한 성능을 나타내는 반면 적응 데이터가 좀 더 증가하는 경우 성능이 약간 저하되는 것을 볼 수 있다. 이것은 적응 데이터가 증가할 수록 테스트 환경과 유사한 환경의 적응 데이터도 늘어나기 때문에 테스트 환경을 더 잘 표현하는 한 가지 모델을 사용하는 것이 여러 환경과의 가중합 모델보다 좀 더 유리하다고 볼 수 있다. 그러나 평균 성능으로 볼 때 본 절에서 제안한 "EN + EV + Bias + Clustering (W.S)" 방식이 가장 우수한 성능을 보였다.

그림 4에서는 적응 환경과 테스트 환경에서 서로 다른 잡음을 사용하는 경우에 대한 각각의 SNR별 실험 결과들에 평균 성능을 나타내었다. 그림에서 보면 기차역 잡음을 사용하는 모든 경우에 테스트 데이터와 가장 유사한 적응 데이터를 이용해 인식 실험을 하는 "EN + EV + Bias

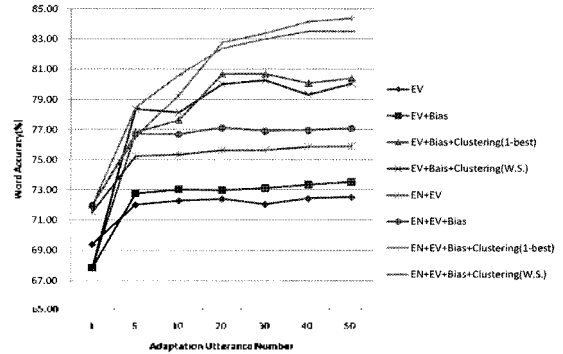


그림 3. 배블, 자동차, 로봇 잡음에 대한 화자적응 실험결과 (적응: SNR 20 dB, 10 dB의 배블, 자동차, 로봇 잡음 / 테스트: SNR 20 dB, 15 dB, 10 dB, 5 dB의 배블, 자동차, 로봇 잡음)

Fig. 3. Results of speaker adaptation by the proposed method with babble, car and robot noise (adaptation : babble, car, robot noise with SNR 20 dB, 10 dB / test : babble, car, robot noise with SNR 20 dB, 15 dB, 10 dB, 5 dB).

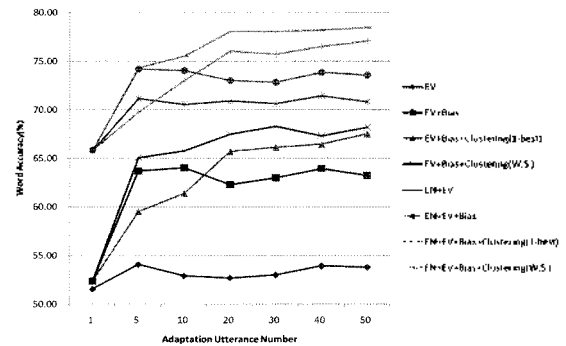


그림 4. 기차역 잡음에 대한 화자적응 실험결과 (적응: SNR 20 dB, 10 dB의 배블, 자동차, 로봇 잡음 / 테스트: SNR 20 dB, 15 dB, 10 dB, 5 dB의 기차역 잡음)

Fig. 4. Results of speaker adaptation by the proposed method with train station noise (adaptation : babble, car, robot noise with SNR 20 dB, 10 dB / test : train station noise with SNR 20 dB, 15 dB, 10 dB, 5 dB).

+ Clustering (1-best)" 방식에 비해 테스트 데이터와 적응 데이터 사이의 유사도 정보를 이용해 SA 모델의 가중합의 형태로 인식 실험에 사용한 "EN + EV + Bias + Clustering (W,S)" 방식이 보다 우수한 결과를 얻었다. 이것은 다양한 환경 그룹의 적응 모델들의 가중합을 통해 적응 데이터에 전혀 포함되지 않은 새로운 환경을 적절하게 표현함으로써 성능 향상을 얻을 수 있다고 볼 수 있다.

V. 결론

본 논문에서는 다양한 잡음 환경 하에서 화자적응을 통해 인식성능 향상시키기 위해 방법을 제안하였다. 먼

서 다양한 잡음 환경을 분리하기 위한 방법으로 환경 군집화 방법을 이용해 유사한 환경별로 군집화를 하였고, 군집화된 환경별로 SA 모델을 생성한 후 적응시 전혀 관측되지 않은 새로운 테스트 환경에 대처하기 위해 적응시 생성된 환경들의 가중합을 통해 인식 환경을 적절하게 표현할 수 있도록 하였다. 실험결과 환경 군집화시 테스트 데이터와 유사도가 높은 여러 군집들의 가중합을 이용한 방식이 가장 우수한 성능을 보였으며, SI모델을 사용한 baseline 시스템에 비해 적응 데이터 수에 따라서 40%에서 59%의 오류 감소율을 얻었다.

감사의 글

이 논문은 부산대학교 자유과제 학술연구비 (2년)에 의하여 연구되었음.

참고 문헌

1. 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*, pp.1771-1774, Nov. 30-Dec. 4, 1998.
2. 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*, vol.39, no.4, pp.806-814, 1991.
3. C. J. Leggetter and P. C. Woodland, "Maximum likelihood linear regression for speaker adaptation of continuous density hidden Markov models," *Computer Speech and Language*, vol.9, no.1, pp.171-185, 1995.
4. H. J. Song and H. S. Kim, "Simultaneous estimation of weights of eigenvoices and bias compensation vector for rapid speaker adaptation," in *Proc. ICSLP*, pp.2945-2948, Oct. 2004.
5. ITU recommendation P.56, *Objective measurement of active speech level*, Mar., 1993.
6. H. J. Song and H. S. Kim, "Eigen-environment based noise compensation method for robust speech recognition," In *Proc. Eurospeech*, pp.981-984, Sep. 2005.

저자 약력

• **김 영 국 (Young Kuk Kim)**

2008년 2월: 부산대학교 공과대학 전자공학과 (공학박사)
현재: LG전자기술원

• **송 화 전 (Hwa Jeon Song)**

현재: 부산대학교 컴퓨터 및 정보통신연구소 연구원
한국음향학회지 제22권 제1호 참조

• **김 형 순 (Hyung Soon Kim)**

현재: 부산대학교 전자공학과 교수
한국음향학회지 제22권 제1호 참조