

## 잡음음성 음향모델 적용에 기반한 잡음에 강인한 음성인식

### Noise Robust Speech Recognition Based on Noisy Speech Acoustic Model Adaptation

정 용 주<sup>1)</sup>

Chung, Yongjoo

#### ABSTRACT

In the Vector Taylor Series (VTS)-based noisy speech recognition methods, Hidden Markov Models (HMM) are usually trained with clean speech. However, better performance is expected by training the HMM with noisy speech. In a previous study, we could find that Minimum Mean Square Error (MMSE) estimation of the training noisy speech in the log-spectrum domain produce improved recognition results, but since the proposed algorithm was done in the log-spectrum domain, it could not be used for the HMM adaptation. In this paper, we modify the previous algorithm to derive a novel mathematical relation between test and training noisy speech in the cepstrum domain and the mean and covariance of the Multi-condition Training (MTR) trained noisy speech HMM are adapted. In the noisy speech recognition experiments on the Aurora 2 database, the proposed method produced 10.6% of relative improvement in Word Error Rates (WERs) over the MTR method while the previous MMSE estimation of the training noisy speech produced 4.3% of relative improvement, which shows the superiority of the proposed method.

**Keywords:** noisy speech recognition, model adaptation, VTS, HMM

#### 1. 서론

잡음환경에서의 음성인식은 많은 기술적인 발전에도 불구하고 여전히 어려운 문제로 남아 있다. HMM 파라미터들과 테스트 잡음음성간의 불일치를 줄이기 위한 방법은 크게 잡음에 강인한 특징추출, 음질향상 그리고 특징보상 및 모델 파라미터 적응 등으로 나누어진다[1]-[3].

최근에는 기존의 연구들과는 다른 관점에서 미리 잡음음성을 이용하여 HMM을 훈련하는 방법이 잡음음성 인식성능 향상을 위한 유용한 방법으로 간주되어 왔다. MTR 방식에서는 다양한 잡음환경의 잡음음성들을 모아서 HMM을 훈련함으로써 보다 잡음에 강인한 인식모델을 훈련과정 중에 구축할 수 있었다[4]. MMSR(Multi-Model based Speech Recognition) 방식에서는 잡음종류나 신호대잡음비(SNR: Signal to Noise Ratio)에 따라서 각각의 HMM을 훈련과정을 통해서 구성함으로써

테스트 잡음음성에 최적화된 인식모델을 사용할 수 있게 하였다[5]. 이러한 방법을 통해서 얻어지는 잡음음성 HMM은 깨끗한 음성 HMM에 비하여 테스트 잡음음성과의 불일치를 줄이는데 크게 기여하며 깨끗한 음성 HMM보다는 인식성능이 우수함이 알려져 있다.

한편, 잡음음성 HMM에 기존의 특징보상방법이나 모델적용 방법을 적용하려는 노력들이 있어 왔다. 이는 잡음음성 HMM이 깨끗한 음성 HMM에 비해서 성능향상에 유리하다는 생각에 기반 하였다. MTR 훈련된 HMM을 가상의 깨끗한 HMM으로 변환되도록 하여 환경의 변화에 보다 잘 적응하도록 함으로서 상당한 인식성능향상을 이룬 경우와[6], 잡음의 종류에 근거한 MMSE 특징 보상 방법이 MMSR 인식기에 적용되어 좋은 결과를 얻은 경우 등을 예로 들 수 있다[5].

우리는 기존의 연구에서, 훈련잡음음성과 테스트 잡음음성간의 관계를 로그-스펙트럼(log-spectrum) 영역에서 수학적으로 정립하였다[7]. 이 관계식을 VTS를 이용하여 간소화 한 후, 테스트 잡음음성을 이용하여 훈련잡음음성을 추정함으로써 MTR 방식에 비하여 향상된 인식성능을 얻을 수 있었다. 그러나 수학적 관계식이 로그-스펙트럼 영역에서만 성립되었으

1) 계명대학교 yjjung@kmu.ac.kr

접수일자: 2014년 3월 18일

수정일자: 2014년 5월 22일

게재결정: 2014년 6월 16일

로, 제안된 방식은 HMM 파라미터의 적응을 위해서는 사용될 수 없는 단점이 있었다. 따라서 본 논문에서는 기존의 알고리즘을 수정하여 이를 잡음음성 HMM의 파라미터 적응에 이용될 수 있도록 하였다. 제안된 방식은 Aurora 2 데이터베이스 [4]를 이용한 인식실험을 통하여 평가되었으며 기존의 방식과의 성능비교를 통해서 제안된 방식의 우수성을 보이고자 한다.

본 논문의 구성은 다음과 같다. 부가잡음과 채널잡음을 함께 고려한 VTS 기반의 잡음음성 HMM 모델 적용방법에 대한 설명이 2장에 소개된다. 3장에서는 제안된 방식을 Aurora 2 데이터베이스를 이용한 잡음음성 인식실험에 적용한 결과가 제시되며 4장에서는 결론을 맺는다.

## 2. VTS 기반의 잡음음성 HMM 모델 적용

이번 절에서는 VTS 축약을 통하여 잡음음성 HMM의 평균 벡터와 공분산 행렬에 대한 적응식을 유도하고자한다. 깨끗한 음성 HMM을 사용하는 기존의 VTS 방식과는 다르게 MTR 방식에 의해서 훈련된 잡음음성 HMM이 제안된 모델 적용 방식에서 사용된다. 캡스트럼(cepstrum) 영역에서 훈련잡음음성과 테스트 잡음음성간의 수학적 관계식이 먼저 유도되며, 비선형적인 이 관계식은 VTS에 의하여 축약된다. 또한 부가잡음과 채널잡음이 알려졌다고 가정함으로써 VTS 전개식으로부터 테스트 잡음음성에 대한 HMM의 평균벡터와 공분산 행렬이 추정된다. 잡음 파라미터에 대한 재추정을 위하여 반복적인 EM(Expectation Maximization) 알고리즘이 적용되며 이러한 반복적인 추정과정을 통해서 충분히 적응된 HMM 파라미터 값들을 이용하여 최종 인식시에 활용한다.

### 2.1 HMM 파라미터의 적응

캡스트럼 영역에서 깨끗한 음성  $x$ 와 부가잡음  $n$  및 채널잡음  $h$ 에 의해서 오염된 테스트 잡음음성  $y$ 간에는 다음과 같은 관계식이 일반적으로 성립한다 [3].

$$y = x + h + C \log(1 + \exp(C^{-1}(n - x - h))) \quad (1)$$

식 (1) 에서  $\mathbf{1}$ 는 단위벡터이며  $C$  와  $C^{-1}$  은 각각 DCT(Discrete Cosine Transform) 과 그 역을 나타낸다.

훈련잡음음성에는 채널잡음이 없다고 가정하면 훈련잡음음성  $y_{Tr}$ 는 아래와 같이 표현된다.

$$y_{Tr} = x + C \log(1 + \exp(C^{-1}(n_{Tr} - x))) \quad (2)$$

위 식에서  $n_{Tr}$ 은 훈련잡음음성에 포함된 부가잡음이며 훈

련과정 중에 추정된다.

식 (2)로부터,

$$\exp(C^{-1}x) = \exp(C^{-1}y_{Tr}) - \exp(C^{-1}n_{Tr}) \quad (3)$$

식 (3)을 식 (1)과 식 (2)에 삽입하면, 테스트 잡음음성  $y$ 와 훈련잡음음성  $y_{Tr}$  간에는 아래와 같은 관계식이 성립한다.

$$y = y_{Tr} + h + g(y_{Tr}, n, h, n_{Tr}) \quad (4)$$

$$g(y_{Tr}, n, h, n_{Tr}) = C \log(1 + \exp(C^{-1}(n - h - y_{Tr})) - \exp(C^{-1}(n_{Tr} - y_{Tr}))) \quad (5)$$

식 (4)는  $\{n, h, y_{Tr}\}$ 의 초기값인  $\{\mu_{n,0}, h_0, \mu_{y_{Tr}}\}$ 를 중심으로 하는 1차의 VTS 전개식을 통해서 아래식과 같이 표현된다.

$$y = \mu_{y_{Tr}} + h_0 + g(\mu_{y_{Tr}}, \mu_{n,0}, h_0, n_{Tr}) + \quad (6)$$

$$G_{y_{Tr}}(y_{Tr} - \mu_{y_{Tr}}) + G_n(n - \mu_{n,0}) + G_h(h - h_0)$$

$$[G_{y_{Tr}}]_{il} = \sum_k C_{ik} \frac{C_{kl}^{-1}}{1 + B_k - A_k}, \quad (7)$$

$$[G_h]_{il} = \sum_k C_{ik} \frac{C_{kl}^{-1}(1 - A_k)}{1 + B_k - A_k},$$

$$[G_n]_{il} = \sum_k C_{ik} \frac{C_{kl}^{-1}B_k}{1 + B_k - A_k}$$

$$A_k = \exp\left(\sum_j C_{kj}^{-1}(n_{Tr,j} - \mu_{y_{Tr},j})\right), \quad (8)$$

$$B_k = \exp\left(\sum_j C_{kj}^{-1}(n_{0,j} - h_{0,j} - \mu_{y_{Tr},j})\right)$$

위식에서  $\mu_{y_{Tr}}$ 는 훈련잡음음성 HMM의 각 혼합성분의 평균벡터를 의미하며  $n_{Tr}$ 는 고정된 값으로 간주한다.  $[\cdot]_{il}$ 는 행렬의  $i$ 번째 행과  $l$ 번째 열에 있는 원소를 의미한다. 식 (6)을 이용하면, 적응된 HMM의 각 혼합성분의 평균  $\mu_y$ 와 분산  $\Sigma_y$ 를 훈련잡음음성과 부가잡음신호  $n$ 의 통계정보를 이용하여 아래식과 같이 구할 수 있다. 이때 채널잡음  $h$ 는 단순 파라미터로 취급되며 랜덤변수로는 고려되지 않는다.

$$\mu_y = \mu_{y_{Tr}} + h_0 + g(\mu_{y_{Tr}}, \mu_{n,0}, h_0, n_{Tr}) + G_n(\mu_n - \mu_{n,0}) + G_h(h - h_0) \quad (9)$$

$$\Sigma_y = G_{y_{Tr}} \Sigma_{y_{Tr}} G_{y_{Tr}}^T + G_n \Sigma_n G_n^T \quad (10)$$

위식에서 부가잡음신호  $n$ 에 대한 공분산 행렬  $\Sigma_n$ 은 테스트 잡음음성의 묵음구간을 이용하여 추정되며 본 논문에서는 알고리즘의 단순화를 위하여 재추정 하지는 않았다.

차분과 차차분 특징벡터에 대한 평균  $(\mu_{\Delta,y}, \mu_{\Delta\Delta,y})$ 과 공분산 $(\Sigma_{\Delta,y}, \Sigma_{\Delta\Delta,y})$ 은 연속-시간 축약[8]을 이용하여 아래식과 같이 구하였다.

$$\begin{aligned} \mu_{\Delta,y} &= G_{y_{Tr}} \mu_{\Delta,y_{Tr}} \\ \Sigma_{\Delta,y} &= G_{y_{Tr}} \Sigma_{\Delta,y_{Tr}} G_{y_{Tr}}^T + G_n \Sigma_{\Delta,n} G_n^T \\ \mu_{\Delta\Delta,y} &= G_{y_{Tr}} \mu_{\Delta\Delta,y_{Tr}} \\ \Sigma_{\Delta\Delta,y} &= G_{y_{Tr}} \Sigma_{\Delta\Delta,y_{Tr}} G_{y_{Tr}}^T + G_n \Sigma_{\Delta\Delta,n} G_n^T \end{aligned} \quad (11)$$

## 2.2 HMM 파라미터의 적응

반복적인 EM 과정을 통하여 잡음파라미터  $h$ 와  $\mu_n$ 에 대한 재추정을 실시하였다. 부가잡음의 평균과 공분산은 테스트 잡음음성의 묵음구간의 샘플을 이용하여 초기화되었으며 채널 잡음  $h$ 은 0으로 초기화 되었다.

프레임 전체 길이가  $T$ 인 테스트 잡음음성  $Y = \{y_1, y_2, \dots, y_T\}$ 가 주어지면 잡음파라미터  $\phi = \{\mu_n, h\}$ 는 아래의 보조함수  $Q(\phi|\bar{\phi})$ 를 최대화함으로써 재추정된다.

$$Q(\phi|\bar{\phi}) = \sum_{t=1}^T \sum_{s \in \Omega_s} \sum_{m \in \Omega_m} \gamma_{tsm} \log p(y_t | s_t = s, m_t = m, \phi) \quad (12)$$

식 (11) 에서  $\Omega_s$ 와  $\Omega_m$ 은 HMM의 전체 상태의 집합과 각 상태에서의 혼합성분의 집합을 나타낸다.  $s_t$ 와  $m_t$ 는 시간  $t$ 에서의 상태와 혼합성분을 각각 의미한다.  $\gamma_{tsm}$ 은 사후 확률  $p(s_t = s, m_t = m | Y, \bar{\phi})$ 을 나타내며  $p(y_t | s_t = s, m_t = m, \phi)$ 는 평균과 분산이 식 (9), (10)에서 구해지는 가우시안 확률밀도함수  $N(y_t; \mu_{y,sm}, \Sigma_{y,sm})$ 가 된다. 식 (12)의 보조함수를 최대화하기 위해서 식(12)의  $\phi = \{\mu_n, h\}$ 에 대한 미분을 각각 구하고 그 값을 0으로 두면 아래와 같은 잡음파라미터에 대한 재추정 수식이 얻어지는 것을 알 수 있다. 유도과정의 보다 자세한 사항은 생략한다.

$$\begin{aligned} \mu_n &= \mu_{n,0} + \left\{ \sum_{t,s,m} \gamma_{tsm} G_n^T \Sigma_{y,sm}^{-1} G_n \right\}^{-1} \cdot \\ &\quad \left\{ \sum_{t,s,m} \gamma_{tsm} G_n^T \Sigma_{y,sm}^{-1} (y_t - \mu_{y_{Tr}} - h_0 - g_0) \right\} \end{aligned} \quad (13)$$

$$h = h_0 + \left\{ \sum_{t,s,m} \gamma_{tsm} G_h^T \Sigma_{y,sm}^{-1} G_h \right\}^{-1} \cdot \quad (14)$$

$$\left\{ \sum_{t,s,m} \gamma_{tsm} G_h^T \Sigma_{y,sm}^{-1} (y_t - \mu_{y_{Tr}} - h_0 - g_0) \right\}$$

$$g_0 = g(\mu_{y_{Tr}}, \mu_{n,0}, h_0, n_{Tr}) \quad (15)$$

## 2.3 알고리즘의 구현

위에서 언급된 것처럼 제안된 HMM 모델적용 방식은 주어진 MTR 훈련된 HMM에 대해서 HMM 파라미터값의 적응과 잡음파라미터의 재추정이 반복적인 EM 과정에 의해 이루어지며 전체적인 알고리즘의 구성은 아래의 순서로 이루어진다.

1. 테스트 잡음음성을 불러온다.
2. 채널잡음의 초기값을 0 으로 한다.
3. 부가잡음의 평균과 분산에 대한 초기값은 테스트 잡음음성의 처음과 끝부분의 묵음구간을 이용하여 구한다.
4. 식 (9)~(11)을 이용하여 테스트 잡음음성에 해당하는 HMM 파라미터 값을 구한다.
5. 적응된 HMM 파라미터값을 이용하여 테스트 잡음음성에 대한 인식을 수행한다.
6. 식 (13)~(15)을 이용하여 부가잡음의 평균과 채널잡음을 재추정한다.
7. 재추정된 부가잡음의 평균과 채널잡음을 이용하여 식 (9)~(11)을 이용하여 테스트 잡음음성에 해당하는 HMM 파라미터값을 재추정한다.
8. 7에서 구한 HMM 파라미터 값을 이용하여 주어진 테스트 잡음음성에 대해서 최종인식실험을 수행한다.
9. 더 많은 테스트 잡음음성이 있다면 1 로 돌아가고 그렇지 않으면 끝낸다.

위의 과정에서는 EM 과정을 1회만 하는 것으로 가정하였으나 반복적인 EM 과정을 수행하고자 한다면 6 과 7을 반복적으로 수행하게 할 수 도 있다. 또한 여러 번의 디코딩 (multi-pass decoding)과정을 거치고 싶다면 5 와 7 사이를 반복적으로 수행할 수 도 있다. 본 연구에서는 EM과정은 주어진 테스트 잡음음성에 대한 로그 유사도가 수렴할 때 까지 반복적으로 수행하였으나 multi-pass decoding 과정은 1회로 한정하였다.

## 3. 실험결과

본 연구에서는 잡음음성인식 실험을 위하여 Aurora 2 데이터베이스를 사용하였다. Aurora 2 데이터베이스는 깨끗한 연속 숫자음 음성신호에 인공적으로 부가잡음을 더해주고 채널

왜곡을 인가하여 만들어진 잡음음성으로 구성되어 있으며 국제적으로 공인되어 가장 많이 사용되는 음성데이터 중의 하나이다[4]. 훈련방법으로는 CLEAN과 MTR 두 가지 종류가 있으며 CLEAN 방식은 HMM의 훈련을 위하여 깨끗한 음성데이터만을 이용하며 MTR 방식은 깨끗한 음성과 더불어 여러 가지 종류의 잡음과 다양한 SNR 값으로 구성된 잡음음성을 이용하여 HMM을 훈련하는 방식이다. 인식실험에는 3가지 종류의 음성데이터 Set이 사용된다. Set A는 MTR 훈련시 알려진 잡음신호(Subway, Babble, Car, Exhibition)들로부터 생성된 잡음음성으로 구성되며 Set B는 훈련시 알려지지 않은 잡음신호(Restaurant, Street, Airport, Train-Station)들로부터 생성된 잡음음성신호로 구성되며 마지막으로 Set C는 부가 잡음 외에도 채널 왜곡이 인위적으로 조성된 잡음음성신호들로 구성된 경우이다.

음성특징 추출을 위해서는 음성개선 알고리즘이 적용된 ETSI AFE를 사용하였다[9]. 0차의 cepstral 계수를 포함한 13차의 MFCC(Mel-frequency Cepstral Coefficient) 특징벡터를 추출하고 이들의 차분과 차차분 계수를 추가하여 전체 39차의 특징벡터를 사용하였다.

음향모델들은 Aurora 2 데이터베이스에서 제공된 숫자음에 대한 Simple Back End (SBE) 스크립트와 Complex Back End (CBE) 스크립트를 함께 사용하였다 [4]. SBE 스크립트의 경우에는 음향모델을 위해서 각 숫자에 대한 HMM은 3개의 Gaussian 성분을 가지는 16개의 상태로 이루어지나 묵음에 관한 HMM은 6개의 Gaussian 성분을 가지는 3개의 상태로 구성된다. 또한 1개의 상태를 가지는 짧은 묵음에 관한 HMM도 구성되며 이는 묵음 HMM의 가운데 상태와 동일하다고 간주된다. 반면에 CBE 스크립트의 경우에는 각 상태의 혼합성분의 개수가 숫자음에 대해서는 20개로 증가하고 묵음에 대해서는 36개로 증가하게 된다. 또한 음향모델의 훈련과 인식을 위해서는 HTK(Hidden Markov Toolkit)를 사용하였다[10].

<표 1>에는 SBE 스크립트를 사용한 경우에 제안된 모델적응 알고리즘의 성능을 기존의 연구결과와 비교 하였다. MTR 훈련방식을 사용하여 잡음음성 HMM을 구성하였으며 Baseline은 깨끗한 음성으로 훈련된 HMM을 사용하는 것을 의미한다.

<표 1>의 결과를 통해서 보면 제안된 모델적응 방식은 기존의 특징보상 방식에 비해서 상당한 인식성능의 향상을 보임을 알 수 있다. 특징보상의 경우에는 Set A의 경우에는 MTR에 비해서도 성능이 오히려 저조하였으나 모델적응을 적용함으로써 Set A에 대해서도 성능이 향상됨을 알 수 있다.

<그림 1>에서 보듯이 제안된 모델적응방식은 기존의 특징보상방식에[7] 비해서 MTR 방식대비 성능에서 우수한 결과를 나타냄을 알 수 있다. 특히, Set C에서 특징보상방식은 MTR 방식에 비해서 약 5% 정도의 상대적 단어오인식율 향상을

을 보였으나 모델적응방식은 MTR에 비해서 약 18% 정도의 상대적 단어오인식율 향상을 보여서 제안된 모델적응 방식이 채널잡음의 영향을 줄이는데 크게 기여하는 것으로 나타났다. Set A에서는 특징보상방식이 성능향상을 보이지 못하는데 비해서 모델적응방식은 5% 정도의 상대적 단어오인식율 향상을 보임을 알 수 있으며 Set B에서는 특징보상과 모델적응방식의 차이가 가장 크지 않은 것으로 나타난다.

표 1. Baseline, MTR, MTR+특징보상과 제안된 MTR+모델적응 기법의 단어오인식율(Word Error Rate) (%)의 비교 (SBE 스크립트 사용시)

Table 1. Comparison in word error rates (%) of the proposed MTR+Model Adaptation method with the Baseline, MTR and MTR+Feature Compensation methods (SBE script)

	Set A	Set B	Set C	평균
Baseline	12.25	12.90	14.56	12.97
MTR	8.20	8.88	9.42	8.72
MTR+특징보상	8.31	8.03	8.97	8.33
MTR+모델적응	7.75	7.80	7.69	7.76

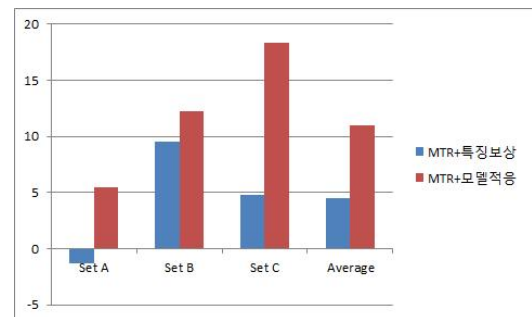


그림 1. 제안된 모델적응방식과 기존의 특징보상방식의 MTR방식 대비 상대적 단어오인식 향상율(%) (SBE 스크립트 사용시)

Figure 1. Relative improvement in word error rates (%) over the MTR method of the proposed model adaptation method and the conventional feature compensation method(SBE script).

HMM의 모델링 복잡도에 따라서 인식알고리즘의 성능이 변하는 현상은 자주 일어난다. 따라서 제안된 알고리즘의 성능의 강인성을 확인하기 위해서 SBE 스크립트 보다 HMM의 상태의 혼합성분의 수가 훨씬 많이 설정된 CBE 스크립트에 대해서도 제안된 알고리즘의 성능분석을 실시하였다.

<표 2>의 결과를 통해서 제안된 모델적응 방식이 CBE 스크립트를 사용한 경우에도 SBE 스크립트를 사용한 경우와 마찬가지로 기존의 특징보상 방식에 비해서 우수한 성능을 보임을 알 수 있다.

<그림 2>에는 CBE 스크립트를 사용한 경우 제안된 모델적응방식과 기존의 특징보상방식이 MTR 방식에 비해서 성취한 상대적 오인식 향상률을 보여준다. <그림 1>에서 보여준 SBE 경우와 비슷하게 제안된 모델적응 방식은 Set C에서 가장 큰 상대적 오인식 향상을 보여주며 Set A에 대해서도 MTR 방식에 비해서 더 나은 성능을 보임을 알 수 있다. 반면 Set B에서는 특징보상 방식과 큰 차이가 나지 않는 것을 알 수 있었다. <그림 1>과 <그림 2>의 결과를 비교해보면 제안된 모델적응방식은 HMM의 복잡성과 관계없이 유사한 성능향상을 보임을 알 수 있으며, 이는 제안된 알고리즘의 강인성을 보여준다.

표 2. Baseline, MTR, MTR+특징보상과 제안된 MTR+모델적응 기법의 단어오인식율(Word Error Rate) (%)의 비교 (CBE 스크립트 사용시)

Table 1. Comparison in word error rates (%) of the proposed MTR+Model Adaptation method with the Baseline, MTR and MTR+Feature Compensation methods (CBE script)

	Set A	Set B	Set C	평균
Baseline	11.58	12.10	13.68	12.20
MTR	6.31	7.0	7.31	6.79
MTR+특징보상	6.47	6.41	6.75	6.5
MTR+모델적응	6.0	6.26	5.81	6.07

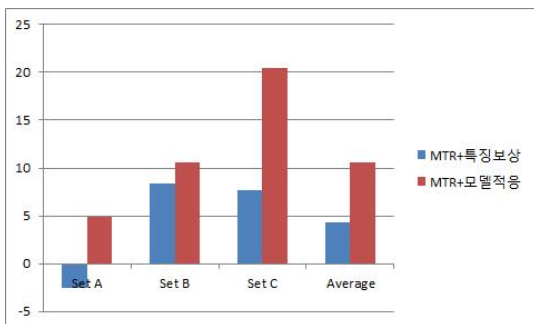


그림 2. 제안된 모델적응방식과 기존의 특징보상방식의 MTR방식 대비 상대적 오인식 향상률(%) (CBE 스크립트 사용시)

Figure 2. Relative improvement in word error rates (%) over the MTR method of the proposed model adaptation method and the conventional feature compensation method(CBE script).

#### 4. 결론

본 논문에서는 강인한 잡음음성인식을 위한 HMM 파라미터 적응방식을 제안하였다. HMM이 깨끗한 음성에 의해서 혼

련되는 기존의 적응방식에 비해서 제안된 방식은 잡음음성 HMM의 파라미터를 적응시킴으로서 인식성능의 향상을 이루었다. 테스트 잡음음성과 훈련잡음음성간의 새로운 관계식이 캡스트럼 영역에서 도출되었으며, VTS 기반의 축약을 이용하여 MTR 방식에 의해서 훈련된 HMM의 파라미터들이 적응되도록 하였다. 또한 잡음과 채널잡음의 파라미터들에 대한 재추정이 반복적인 EM 알고리즘을 통하여 일어난다. 제안된 방식은 기존의 특징보상 방식에 비해서 뛰어난 성능을 보임을 Aurora 2 데이터베이스를 이용한 인식실험 결과 확인할 수 있었다.

#### 감사의 글

이 논문은 2013년도 정부(교육과학기술부)의 재원으로 한국연구재단의 기초연구사업 지원을 받아 수행된 것임. (NRF-2011-0006994)

#### 참고문헌

- [1] Gales, M. (1995). Model based techniques for noise-robust speech recognition. Ph.D. Dissertation, University of Cambridge, United Kingdom.
- [2] Ball, S. F. (1979). Suppression of acoustic noise in speech using spectral subtraction, *IEEE Trans. Acoust., Speech, Signal Process.* Vol. 27, No. 2, 113-122.
- [3] Moreno, P. J. (1996). Speech recognition in noisy environments. Ph.D. Dissertation, Carnegie Mellon University, United States of America.
- [4] Hirsch, H. G. & Pearce, D. (2000). The Aurora experimental framework for the performance evaluation of speech recognition systems under noisy conditions, *in Proc. ICSLP.* 18-20.
- [5] Xu, H., Tan, Z. -H., Dalsgaard, P. & Lindberg, B. (2007). Noise condition-dependent training based on noise classification and SNR estimation. *IEEE Trans. Audio, Speech and Language Process.* Vol. 15, No. 8, 2431-2443.
- [6] Kalinli, O., Seltzer, M. L., Droppo, J., & Acero, A. (2010). Noise adaptive training for robust automatic speech recognition. *IEEE Trans. Audio, Speech and Language Process.* Vol. 18, No. 8, 1889-1901.
- [7] Chung, Y. & Hansen, J.H.L. (2013). Compensation of SNR and noise type mismatch using an environmental sniffing based speech recognition solution. *EURASIP Journal on Audio, Speech, and Music Processing*, 2013:12, 1-14.
- [8] Gopinath, R. A., Gales, M., Gopalakrishnan, P. S.,

- Balakrishnan-Aiyer, S. & Pocheny M. A. (1995). Robust speech recognition in Noise : Performance of the IBM continuous speech recognizer on the ARPA noise spoke task. *in Proc. ARPA Spoken Language System Technology*. 127-130.
- [9] ETSI draft standard doc., Speech processing, Transmission and Quality aspects (STQ); Distributed speech recognition; Front-end feature extraction algorithm; Compression algorithm. ETSI Standard ES 202 050, 2002.
- [10] Young, S. (1993). *HTK: Hidden Markov Model Toolkit V3.4.1*. Cambridge University, Engineering Department, Speech Group.

• **정용주 (Chung, Yong Joo)**

계명대학교 전자공학과

주소 대구광역시 달서구 신당동 1000번지

Tel: 053-580-5925

Email: yjjung@kmu.ac.kr

관심분야: 음성인식, 패턴인식, 통계신호처리