잡음환경에서의 음성인식을 위한 모델 파라미터 변환 방식에 관한 연구

A Study on a Model Parameter Compensation Method for Noise-Robust Speech Recognition

장 육 현*, 정 용 주*, 박 성 현*, 은 종 관**
(Yuk-Hyeun Chang*, Yong-Joo Chung*, Sung-Hyun Park*, Chong-Kwan Un**)

요 약

본 논문에서는 잡음에 강한 음성 인식기를 위한 모델 파라미터 변환 방식에 관하여 살펴 보았다. 모델 파라미터 변환에 있어서 잡음에 대한 어떠한 통계 모델도 사용하지 않고 각 단어 단위로 수행되어 실시간 음성 인식이 가능하도록 하였다. Parallel model combination(PMC)은 본 논문에서 제안한 방법과의 성능 비교를 위하여 cepstrum 영역에서 구현되었다. 본 논문에서 제안한 PMC 방법은 modified PMC(MPMC)라 하며, 이 방법은 각 hidden Markov model(HMM)의 state별로 평균적인 가우시안 믹스처(Gaussian mixture)의 변화률과 개별적인 변화률간에 결합지수를 이용하여 평균을 재조정한다. 또한, vector Taylor series 근사화를 이용한 모델 파라미터 변환을 위하여 cepstrum 영역에서의 환경모델 예측을 위한 expectation-maximization(EM) 해를 유도하여 구현하였다. 본 논문에서 구현된 알고리즘들의 성능 위해 HMM 인식기를 이용한 화자독립 고립단어 인식을 수행하였다. 사용된 잡음은 가우시안 백색 잡음과 주행중에 녹음된 자동차 잡음이며, 각 잡음을 signal-to-noise ratio(SNR)별로 사용하였다. 잡음의 모델은 1 state HMM으로 단어시작 3 프레임(frame)을 이용하여 만들어졌다. 인식 결과는 VTS 접근방식을 이용하였을 경우 매우 우수한 인식률을 나타내었으며, MPMC의 경우도 기존의 PMC보다 인식률이 향상 되었다. 특히, 영차 VTS의 경우는 단순히 평균만을 조정하였음에도 불구하고 PMC와 MPMC보다 인식률이 우수하게 나타났다.

ABSTRACT

In this paper, we study a model parameter compensation method for noise-robust speech recognition. We study model parameter compensation on a sentence by sentence and no other informations are used. Parallel model combination(PMC), well known as a model parameter compensation algorithm, is implemented and used for a reference of performance comparision. We also propose a modified PMC method which tunes model parameter with an association factor that controls average variability of gaussian mixtures and variability of single gaussian mixture per state for more robust modeling. We obtain a re-estimation solution of environmental variables based on the expectation-maximization(EM) algorithm in the cepstral domain. To evaluate the performance of the model compensation methods, we perform experiments on speaker-independent isolated word recognition. Noise sources used are white gaussian and driving car noise. To get corrupted speech we added noise to clean speech at various signal-to-noise ratio(SNR). We use noise mean and variance modeled by 3 frame noise data. Experimental result of the VTS approach is superior to other methods. The scheme of the zero order VTS approch is similar to the modified PMC method in adapting mean vector only. But, the recognition rate of the zero order VTS approch is higher than PMC and modified PMC method based on log-normal approximation.

I. 서 론

음성인식기는 일반적으로 잡음환경에 노출될 경우 심각한 성능 저하를 보이게 되며 이의 해결이 실용화의 관건이다. 잡음에 강한 음성인식기를 만들기 위해 다양한 방법들이 연구되어 왔고, 지금도 많은 사람들이 노력을 경주하고 있지만 아직까지 만족할 만한 성능을 거두지 못하고 있다.

모델 파라미터 변환 방식은 특징벡터 영역에서의 변환이나, 청각 모델을 이용한 인식에 비하여 재학습이 필요하지 않아 실시간 계산 부담이 훨씬 적다. 또한, stereo 데이타베이스가 없는 상황에서도 인식대상의 단어를 이용

*LG 정보 통신(주) 중앙 연구소
**한국 과학 기술원 전기 및 전자 공학과
접수일자 : 1997년 5월 12일

하여 배경잡음을 예측한 후 모델 파라미터 compensation을 수행할 수 있다.

본 논문에서는 사전의 배경잡음에 대한 통계적 정보를 사용하지 않고 각 단어마다 배경잡음을 예측한 후, 주어진 모델 파라미터를 여러 compensation 알고리즘을 이용하여 결합하는 방식에 대하여 연구한다. 기존의 parallel model compensation(PMC)의 global한 모델 파라미터 compensation 방식의 문제점을 보완하기 위하여 세밀한 조정이 가능한 modified PMC를 제안하며 이에 대한 인식결과를 비교한다. PMC와 modified PMC는 모두 가산성 잡음만을 모델하여 spectral tilt와 같은 channel distortion을 모델내에 반영할 수 없다. 이러한 문제점을 해결하기 위한 방법으로 vector Taylor serise(VTS) 근사화 방법이 제기되었다. 그러나, 지금까지 발표된 논문은 VTS 근사화 방법을 높은 인식률을 얻을 수 있는 log-spectrum 영역에서 signal compensation에 적용하여 왔으나, 본 논문에서는 cepstrum 특징벡터를 모델 파라미터 보상 방법에 직접적으로 사용하기 위하여 EM해를 유도하였으며 실험을 통하여 결과를 살펴본다. VTS 근사화는 근사화되는 항을 증가할수록 정확한 환경을 나타낸다. 본 논문에서는 영차와 일차에 대하여 EM해를 유도하여 실험한다.

본 논문의 구성은 2장에서 기존의 모델 파라미터 변환 방식의 개요에 대하여 설명하고, 3장에서 본 논문에서 제안한 모델 파라미터 변환 방식을 설명한다. 4장에서는 실험결과를 분석하고, 마지막으로 5장에서 결론을 맺는다.

Ⅱ. 기존의 모델 파라미터 변환 방식

Hidden Markov model(HMM)을 기반으로 하는 인식 시스템을 배경잡음이 존재하는 곳에서 사용하기 위해서는 배경잡음에 의한 영향을 반영해 주어야 한다. 배경잡음에 의한 영향을 특징벡터 추출시에 반영하는 경우를 특징벡터 변환 방식이라하며, 모델 파라미터에 대하여 적용할 경우를 모델 파라미터 변환방식이라 한다.

모델 파라미터 변환 방식은 clean환경에서 훈련된 HMM 인식기를 기반으로 인식실험용 특징변수에서 얻은 모델 파라미터를 변환하여 인식을 한다. 따라서 인식기의 재학습이 필요하지 않으므로 인식 전단계의 과정이 줄어든다. 모델 파라미터 변환 방식은 변환을 위하여 사용되는 테스트 단어를 아느냐 모르느냐에따라 지도학습(supervised learning)과 자율학습(unsupervised learning)으로 나뉘어 진다. 또한, 특징벡터 변환 방식과 마찬가지로 테스트 단어를 noisy 특징벡터로 변환하는 방식과 clean 특징벡터와 유사하게 변환하는 방식이 있다. 본 논문을 위한 실험에서는 자율학습 방법으로 주어진 단어에서 배경잡음을 예측하여 잡음에 대한 모델을 만들고 clean 모델 파라미터와 함께 여러 방법을 적용하여 noisy 모델 파라미터를 만들어 인식실험을 한다.

본 논문을 위한 실험에서 사용되는 모델 파라미터 변환 방법은 기본적인 PMC, modified PMC, VTS등이 있다. 아래에서는 모델 파라미터 변환 방식의 개요에 대하여 알아보고[1], 다음 장에서 본 논문에서 제안된 모델 파라미터 변환 방식을 설명한다.

HMM을 이용하여 signal decomposition 방법을 시작한 것은 1986년 Moore에 의해서이다[2]. 이 방법의 기본적인 idea는 동시에 두개의 신호가 존재할 때 인식을 효과적으로 하자는 데 있다. 병렬 HMM은 이러한 신호들을 모델하기 위하기 위하여 사용되며, 합성된 신호는 병렬 HMM의 각각의 출력을 어떤 함수에 의하여 모델된다. Varga와 Moore는 1990년에 이러한 기술을 사용하여 가산성 배경잡음을 처리하였다[3]. 음성과 잡음에 대한 HMM decomposition 방법의 주요한 장점을 살펴보면,

(가)각기 다른 HMM에 의해 잡음이 모델되어지는 경우 다양한 잡음-정상성 잡음(stationary noise), 비정상성 잡음(non-stationary noise)-을 용이하게 모델링할 수 있다.

(나)이 decomposition 기술은 특정한 형태의 신호결합 함수에 의해 만들어진 합성신호에 만족하는 것이 아니라 여러 다양한 함수에 의해 정의된 합성신호에 대하여 만족한다. 따라서 HMM을 이용한 signal decomposition 방법은 가산성 잡음뿐만 아니라 convolutional noise도 다룰 수 있다.

(다)잡음의 power를 표준적인 spectral subtraction 방식과 같이 반드시 분산이 '영'이라고 가정할 필요는 없다.

그러나 이러한 방법들은 계산량이 많은 단점이 있다. 그림 1에서 보는 바와 같이 함께 존재하는 두개의 신호를 인식하기 위해서는 3차원 상태공간-관측벡터, 잡음, 음성모델-에 대하여 검색이 수행되어야 한다. 따라서 고전적인 Viterbi decoder를 3차원 영역에서 디코딩하기 위한 확장이 필요하다. Noisy 음성 신호는 N×M상태 갯수를 갖는 HMM에 의해 모델된다. N은 clean 음성의 상태 갯수이며, M은 noise에 대한 상태갯수이다. 위와 유사한 방법이 1992년 Ephraim에 의해 사용되어졌는데, 여기에서는 speech enhancement가 목적이었다[4].

Varga와 Moore는 1990 additive stationary noise와 non-stationary noise(예:machine-gun)에 대하여 실험결과를 발

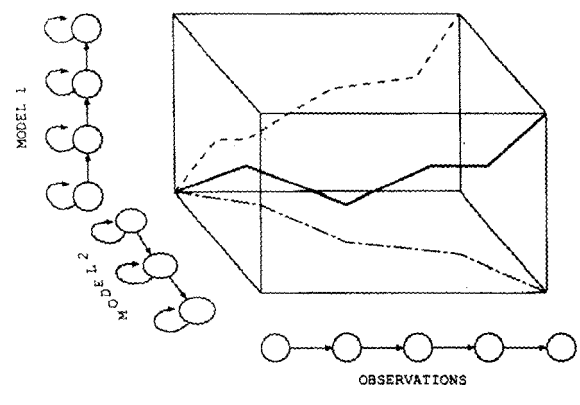

그림 1. 두 종류의 신호에 의해 나타난 상태공간

표하였다. 실험결과에 의하면 HMM *noise decomposition* 방법이 어떠한 잡음 보상법도 결합시키지 않는 baseline system보다 훨씬 뛰어난 인식률을 나타내었다. 또한, 이 실험에 대한 향상된 결과가 Kadirkamanathan과 Varga에 의해 1991년에 발표되었다[5]. 이 실험은 잡음이 섞인 데이타로부터 음성과 잡음 모델을 예측하기 위해 Baum-Welch re-estimation 방법을 일반화하여 사용할 것을 제안하였다. 그러나, 학습과정에서의 계산량이 단일 음성의 HMM의 경우보다 훨씬 많은 것으로 나타나는 단점이 있다. 잡음 환경이 변화할 때마다 잡음 모델을 바꾸어주는 부담을 덜기 위해, Gales와 Young(1992, 1993)이 작은 구간의 잡음 변화를 전체 잡음 모델로 사용할 것을 제안하였다[6][7]. 이러한 잡음 모델 방법은 정확한 끝점 검출의 필요성을 감소시킨다. Gales와 Young의 방법을 PMC 방법이라 부르며, 이 방법은 Varga와 Moore의 방법과 매우 깊은 연관성이 있다. 그러나 몇 가지 다른점이 존재하는데

㈀이 방법은 composition method이다. HMM decomposition 방법에서는 잡음이 섞인 관측벡터들을 다중차원의 공간에 대한 관측벡터들로 분해하여 인식이 수행된다. 반면에 HMM composition 방법에서는 인식하기 전에 잡음이 섞인 관측벡터를 모델링 하여 인식에 사용하며 따라서 계산부담이 줄어든다.

㈁기본적으로 cepstrum영역에서 수행된다.

㈂반드시 공분산(covariance)이 대각행렬일 필요는 없다.

㈃잡음의 변화성(variability)에 의존하므로 실행중에 부담이 크게 감소한다.

PMC에서 잡음은 간단한 HMM모델 - 한 개의 상태 - 에 의해 표현되므로 음성과 잡음의 결합이 직접적이고 단순화되어 더이상의 실시간 계산 부담이 생기지 않는다. 그러나 잡음이 비정상성 잡음일 경우는 더욱 복잡한 HMM 모델이 필요하다. 즉, 음성과 잡음을 가장 적절하게 합성하기 위해서는 디코딩 단계에서 수행되어져야 한다. 이를 위해 다양한 PMC변형이 제안되었는데, 대표적인예로 fixed grand variance를 사용한 PMC[6]와 state-based compensated variance를 사용한 PMC[7]등이 있다. Mel-frequency cepstral coefficient(MFCC)영역에서도 향상된 결과가 나타난다고 발표했으며, 특히 NOISEX-92 데이터베이스를 이용하여 가산성 잡음에 대해 좋은 결과를 보였다[7]. 또한 machine-gun과 같은 임펄시브 잡음(impulsive noise)에 대하여도 좋은 결과가 나타남을 발표하였다[8]. PMC의 기본적인 가정은 음성과 잡음이 linear 영역에서 더해진다는 것이다. 따라서 보상(compensation)은 정적 파라미터-cepstral coefficients-영역에서 주로 수행된다. 그러나, 동적 파라미터 영역에서의 보상을 위해 1993년 이후에 많은 방법들이 제안되었다. 또한, 잡음에 의해 변형된 음성모델의 파라미터들을 예측하기 위하여 gaussian 적분 대신에 직접 잡음이 섞인 표본데이터들을 가지고 음성 모델을 만들어 사용하는 data-driven PMC(DPMC)가 제안되었다[9]. 이러한 방법은 표준적인 PMC에 비해 계산량이 적지만, stereo 데이타베이스가 필요하게 된다. 즉, 일반적인 PMC는 잡음이 섞인 음성만을 사용하여 그것에서 잡음의 모델을 예측하여 전체 음성모델을 만들지만, DPMC의 경우는 clean 음성과 실제 잡음을 직접 합성하여 전체 음성모델을 만들어 사용한다. 그밖에 다양한 방법들이 결합되어 만들어진 model decomposition 방법들이 제안되었으며, 화자 인식에 사용된 방법과 signal compensation 방법에 사용된 기술의 변형 방법들이 모델 파라미터 변환에 적용되고 있다.

Ⅲ. 제안된 모델 파라미터 변환 방식

본 장에서는 기존의 PMC 방법의 문제점을 보완한 새로운 modified PMC방법과 cepstrum 영역에서 전개한 VTS 접근 방법에 대하여 간략히 살펴본다.

3.1 Modified PMC의 이론

이 절에서는 기본적인 PMC를 수정하여 SNR이 낮은 경우에 대하여 인식률을 향상시키는 방법에 대하여 토의한다. 기본적인 PMC의 경우는 배경잡음의 모델 파라미터를 예측한 후 이 모델 파라미터를 이용하여 전체 clean 음성 모델 파라미터를 변환시키게 된다. 즉, 일률적으로 모든 변수에 똑같은 값을 가감하여 주는 것이 된다. 그러나, 모든 모델 파라미터와 그 모델 파라미터의 gaussian mixture, 상태등이 모두 일률적으로 변화하지는 않는다. 어떤 모델 파라미터의 gaussian mixture는 예측하여 만들어진 잡음 모델 파라미터로의 변화성이 크지만, 어떤 모델 파라미터는 잡음모델의 변화성보다 작은 경우도 존재할 것이다. 따라서, 일률적인 변환이 아닌 모델 파라미터의 gaussian mixture별로 변환을 하여 더욱 정확한 변환이 되도록 한다. 이 방법을 modified PMC라 하고 다음과 같이 가정한다.

㈀HMM 모델의 각 상태에서 모델 파라미터의 gaussian mixture는 전체적으로는 비슷한 방향으로 변화한다.

㈁개개의 gaussian mixture의 공분산은 기본적인 PMC와 동일하다.

㈂HMM 모델의 각 상태에서 개개의 gaussian mixture의 평균벡터의 변화는 각 상태에 의존하여 결정된다.

위의 가정을 다음과 같은 식으로 표현할 수 있다.

$$b_1 = \hat{\mu} - \mu$$

$$b_2 = \frac{1}{M} \sum_{i=0}^{M-1} (\hat{\mu} - \mu) = E\{b_1\}$$

여기에서 $\hat{\mu}$과 μ은 PMC에 의해 변환된 평균벡터와 clean 모델 파라미터의 평균벡터이다. 따라서 vector b_1은 개별 평균벡터의 PMC에 의한 변화량을 나타낸다. 상수 M은 한 상태내의 gaussian mixture의 갯수이며, vector b_2는 개별적인 변화량의 평균벡터를 나타낸다. 이와같은 두 vector

를 이용하여 HMM모델내의 각 상태내의 gaussian mixture의 변화량을 결정하는 vector b를 결정할 수 있다. 이러한 vector를 바이어스 벡터(bias vector) b라 부르겠다.

$$b=(1-\lambda)b_1+\lambda b_2$$

여기에서 λ는 vector b_1과 vector b_2의 vector b를 만드는데 있어서의 기여도를 나타내며, 이를 결합지수(association factor)라 부르겠다. 따라서 vector b를 결정하는 문제는 λ를 결정하는 문제로 귀결된다. 여기에서는 많은 해법중 EM방법과 steepest-descent 방법을 사용하여 결합지수 λ를 구하도록 한다.

3.1.1 EM 알고리즘을 이용한 해법

지금부터 설명되는 모든 과정은 모두 cepstrum 영역에서 이루어지며 gaussian mixture의 평균벡터들에 대하여 이루어 진다. EM 알고리즘을 이용한 PMC의 순서도는 그림 2과 같다. 평균벡터의 변환은

$$\begin{aligned}\hat{\mu}_i &= \mu_i^o+\Delta\mu_i \\ &= \mu_i^o+(1-\lambda)b_{1i}+\lambda b_2\end{aligned}$$

이다. 여기에서, μ_i^o는 학습과정에서 얻은 i번째 mixture의 평균벡터이며, $\Delta\mu_i$는 i번째 mixture의 변화량을 나타낸다. 문제로 주어진 것은 clean 음성의 전체 모델 파라미터와 바이어스 벡터를 구성하는 vector b_1, b_2와 관측벡터 y이다. 초기의 결합지수 λ는 실험적인 방법에 의하여 결정되어 진다.

먼저 우리가 EM 알고리즘에 의하여 결합지수 λ를 결정하기 위해 다음과 같은 Q-function을 정의할 수 있다.

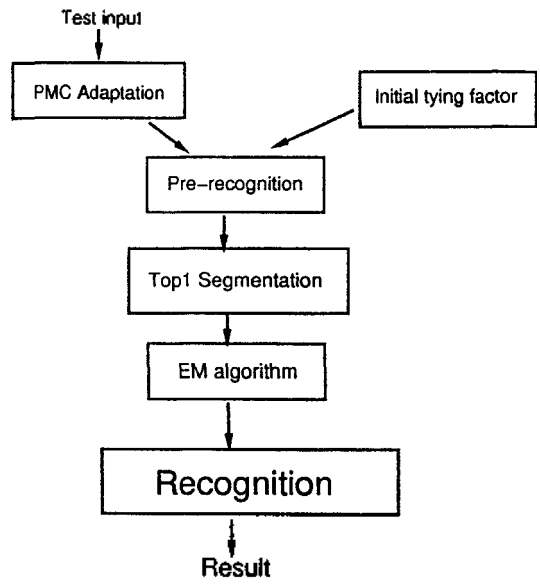

그림 2. EM 알고리즘을 이용한 PMC의 순서도

$$\begin{aligned}Q(\lambda,\bar{\lambda}) = &\sum_{t=0}^{T-1}\sum_{k=0}^{M-1} p(k|y_t,\lambda,\mu_k,\Sigma_k,b_{1k},b_2) \\ &\cdot\log p(y_t,k|\bar{\lambda},\mu_k,\Sigma_k,b_{1k},b_2)\end{aligned}$$

여기에서, t는 시간을 나타내며 T는 단어의 전체시간을 나타낸다. M은 각 상태에서의 mixture의 개수이다. Q-function을 최대화할 수 있는 결합지수 λ가 구하고자 하는 해가 되므로, 이 Q-function을 미분하고 정리하면,

$$\begin{aligned}\frac{\partial Q(\lambda,\bar{\lambda})}{\partial\bar{\lambda}} &= \sum_{t=0}^{T-1}\sum_{k=0}^{M-1} p(k|y_t,\lambda,\mu_k,\Sigma_k,b_{1k},b_2) \\ &\quad\frac{\partial}{\partial\bar{\lambda}}\log p(y_t,k|\bar{\lambda},\mu_k,\Sigma_k,b_{1k},b_2) \\ &= \sum_{t=0}^{T-1}\sum_{k=0}^{M-1}\gamma_t(k)[(b_2-b_{1k})^t\Sigma^{-1}(y_t-\mu_k^o-\lambda(b_2-b_{1k})-b_{1k})]\end{aligned} \quad (1)$$

위에서, $\gamma_t(k)$는 *posteriori* 확률이라 하며 다음과 같이 정의할 수 있다.

$$\gamma_t(k)=\frac{w_kN_k(y_t,M)}{\sum_{l=0}^{M-1}w_lN_l(y_t,M)}$$

여기에서 M은 HMM의 모델을 나타내며, y_t는 시간 t에서의 관측벡터를 나타내며, $N_l(\cdot)$은 l번째 가우시안 믹스처의 output 확률을 나타낸다.

3.1.2 Steepest-descent 알고리즘을 이용한 해법

본 절에서는 EM 알고리즘을 사용하여 결합지수를 예측하는 대신에 steepest-descent 알고리즘을 이용하는 방법을 살펴본다. 우리가 원하는 것은 모델이 주어져 있을때 관측벡터가 나타날 확률이 최대가 되도록 결합지수를 바꾸어 주는 것이므로 그와 유사한 log-likelihood가 최대가 되도록 하는 것이다. 모델이 주어져 있을때 관측벡터 y_t가 될 확률 P는

$$\begin{aligned}P &= \prod_{t=0}^{T-1} p(y_t|M=(\lambda,\mu,\Sigma,b_1,b_2)) \\ &= \prod_{t=0}^{T-1}\sum_{k=0}^{M-1} w_kN(y_t|M=(\lambda,\mu_k,\Sigma_k,b_{1k},b_2))\end{aligned}$$

이고 양변에 log를 취한 후 결합지수 λ에 대하여 미분을 하여 정리하면

$$\log_{10}P=\sum_{t=0}^{T-1}\log_{10}[\sum_{k=0}^{M-1}w_kN(y_t|M=(\lambda,\mu_k,\Sigma_k,b_{1k},b_2))]$$

$$\frac{\partial\log P}{\partial\lambda}=\sum_{t=0}^{T-1}\frac{\sum_{k=0}^{M-1}w_kN(y_t|M)(b_2-b_{1k})^t\Sigma_k^{-1}(y_t-\mu_k)}{\sum_{k=0}^{M-1}w_kN(y_t|M)}$$

$$= \sum_{t=0}^{T-1} \sum_{k=0}^{M-1} \{ \frac{w_k N_k(y_t | M)}{\sum_{k=0}^{M-1} w_l N(y_t | M)} \} \cdot (b_2, b_{1k})' \Sigma_k^{-1} (y_t - \mu_k)$$

이 된다. 여기에서 평균벡터 μ_k는

$$\mu_k = \bar{\mu} + \lambda(b_2 - b_{1k}) + b_{1k}$$

이고 $\bar{\mu}$는 학습에서 얻은 모델 파라미터중 평균벡터를 나타낸다. 따라서, 새로이 예측된 결합지수 λ는

$$\bar{\lambda} = \lambda - \beta \cdot \frac{\partial \log P}{\partial \lambda}$$

이 된다. 여기에서 λ는 이전에 예측된 결합지수이며, β는 수렴속도를 결정하는 factor로 학습률이다. 이와같은 과정을 수렴할 때까지 반복할 수도 있으나, 본 논문에서는 실시간 인식을 위하여iteration을 한번만 하며 실험적으로 학습률을 정하였다. 위의 log-likelihood 미분한 것을 다시 한번 미분을 하면

$$\frac{\partial^2 \log P}{\partial \lambda^2} = - \sum_{t=0}^{T-1} \sum_{k=0}^{M-1} \gamma_t(k) (b_2, b_{1k})' \Sigma_k^{-1} (y_t - \mu_k)$$

이다. 여기에서, 공분산이 대각행렬이고 모두 양의 값을 가지며, 두 바이어스 벡터의 차이의 부호에 상관 없으므로 언제나 음의 값을 가지게 된다. 따라서, log-likelihood는 최대값을 가지게 되어 해가 존재함을 알 수 있다.

3.2 Cepstrum 영역에서의 VTS 접근방법

본 절에서는 cepstrum영역에서의 환경 파라미터 예측 방법에 대하여 알아 보겠다. 여기에서 사용되는 환경 모델은 그림 3과 같다. 환경 모델을 수식으로 나타내면

$$Y = |H(\omega)| X + N$$

이고, Y, X, N은 각각 noisy 음성벡터와 clean 음성벡터, 잡음 vector의 집합을 나타낸다. 전달함수 $H(\cdot)$은 채널 왜곡(channel distortion)과 같은 spectral tilt를 나타낸다. Noisy vector n과 spectral tilt vector는 미지의 vector인 랜덤변수로 간주하며, 서로 통계학적으로 독립(statistically independent)이며, clean 음성벡터와도 각각 통계학적으로 독립이라고 가정한다.

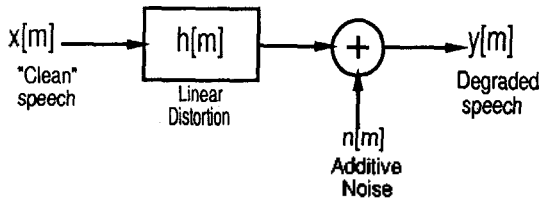

그림 3. 가산성 잡음과 linear 채널을 포함한 환경모델

각각의 양변에 대수로그(log)를 취하고 DCT를 취하면,

$$y = n + C \cdot [\log\{1 + 10^{C^{-1}(x_0 + h_0 - n_0)}\}] \quad (2)$$

이 된다. 여기에서 vector y, n, h, x은 각각 cepstral 영역에서의 noisy 음성벡터, 잡음 vector, spectral tilt, clean 음성벡터를 나타낸다. 먼저, 환경함수에 대한 영차 근사화를 이용한 해를 유도하고 다음에 일차 근사화를 이용한 해를 유도한다.

EM을 적용하기 위하여 다음과 같은 Q-function을 사용하면

$$Q(n, \bar{n}) = \sum_t \sum_k p(k | y_t, n, M) \log p(k | y_t, n, M)$$

여기에서 $M = (\mu_x, \Sigma_x, h)$이다. Q-function의 양변을 $\bar{n}$에 대하여 미분하여 정리하면 다음과 같이 예측된 잡음 vector $\bar{n}$을 구할 수 있다.

$$\bar{n} = \frac{\sum_t \sum_k \gamma_t(k) \Sigma_k^{-1} (y_t - b_k)}{\sum_t \sum_k \gamma_t(k) \Sigma_k^{-1}}$$

$$b_k = C \cdot \log_{10}(1 + 10^{C^{-1}(\mu_{x,k} + h_0 - n_0)})$$

$$\gamma_t(k) = \frac{w_k N_k(y_t, n, M)}{\sum_{l=0}^{M-1} w_l N_l(y_t, n, M)}$$

이다. 윗식에서 h_0, n_0은 초기값을 나타내며, $\mu_{x,k}$은 clean vector x의 k번째 gaussian mixture의 평균벡터이다. Spectral tilt h를 구하는 방법 역시 잡음 vector n을 구하는 방법과 비슷하며, 이 경우 잡음 vector n을 아는 변수로 spectral tilt h를 미지의 변수로 생각하여 미지의 변수 h에 대하여 식을 전개하여 해를 구하면 된다. 예측된 $\bar{h}$는

$$\bar{h} = \frac{\sum_{t=0}^{T-1} \sum_{k=0}^{M-1} \gamma_t(k) \Sigma_{x,k}^{-1} (y_t - a_k)}{\sum_{t=0}^{T-1} \sum_{k=0}^{M-1} \gamma_t(k) \Sigma_{x,k}^{-1}}$$

$$a_k = \mu_{x,k} + C \cdot \log_{10}|I + 10^{C(-(n_0 - h_0 - \mu_x))}|$$

과 같다. 영차 VTS 근사화를 적용하여 noisy 음성의 모델 파라미터를 구하면

$$\mu_y = \mu_x + h + g(\mu_x, n_0, h_0)$$

$$\Sigma_y \simeq \Sigma_x$$

윗식에서 알 수 있듯이 영차 VTS 근사화의 경우 noisy 음성의 gaussian mixture중 공분산은 clean음성의 gaussian

mixture의 공분산으로 근사화됨을 알 수 있다. 이상의 과정은 그림 4와 같다.

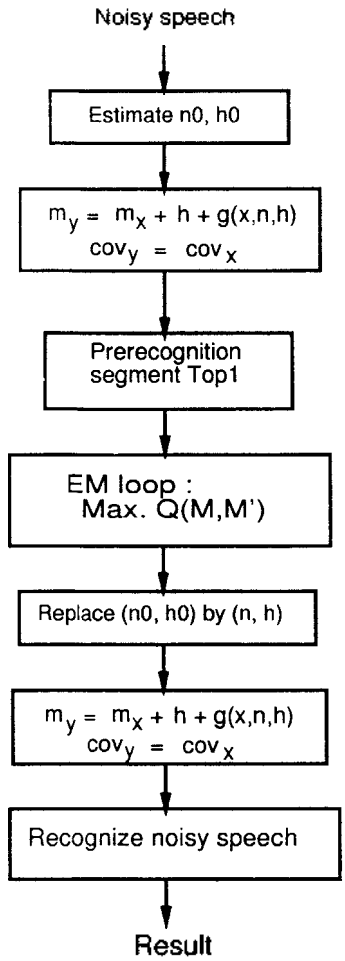

그림 4. Cepstrum 영역에서 EM방식의 영차 VTS 순서도

다음에는 일차 VTS 근사화를 통한 환경 파라미터 예측과 모델 파라미터 보상 과정을 살펴보겠다. Noisy 음성 벡터 y를 일차 근사화를 이용하여 나타내면,

$$y \simeq x + f'(x, n, q)$$
$$= x + (\nabla_x f)' x + (\nabla_n f)' n + (\nabla_q f)' q + g(x_0, n_0, q_0)$$

여기서, $g(x_0, n_0, q_0)$는

$$g(x_0, n_0, q_0) = f(x_0, n_0, q_0) - (\nabla_x f)' x_0 - (\nabla_n f)' n_0 - (\nabla_q f)' q_0$$

이다. 일차 근사화를 적용할 경우 noisy 음성 벡터 y의 모델 파라미터인 평균과 공분산은 다음과 같다.

$$\mu_y(n_t, k, \bar{q}) = (I + \nabla_x f)' \mu_k + (\nabla_n f)' n_t + \nabla_q f)' \bar{q} + q + g(x_0, n_0, q_0)$$

$$\Sigma_y(n_t, k, \bar{q}) = (I + \nabla_x f)' \Sigma_k + (I + \nabla_x f)$$

윗 식에서 "'"는 행렬의 전치를 나타내고, f는 잡음에 의해 영향을 받은 정도를 나타내는 환경 함수이다. 새로운 모델 파라미터를 구하기 위하여 참고 문헌 [10]에서 사용한 Q-function을 사용하여 잡음 벡터와 모델 파라미터와 spectral tilt를 예측한다.[9] 먼저, 잡음의 평균벡터와 공분산 행렬은

$$\hat{\mu}_n = \frac{\sum_{t=0}^{T-1} \sum_{n=0}^{N-1} \sum_{m=0}^{M-1} \gamma_t(n, m) \mu_n(y_t, n, m, \lambda)}{\sum_{t=0}^{T-1} \sum_{n=0}^{N-1} \sum_{m=0}^{M-1} \gamma_t(n, m)}$$

$$\hat{\Sigma}_n = \frac{\sum_{t=0}^{T-1} \sum_{n=0}^{N-1} \sum_{m=0}^{M-1} \gamma_t(n, m)[\Sigma_n(y_t, n, m, \lambda) + \mu_n(y_t, n, m, \lambda)\mu_n'(y_t, n, m, \lambda)]}{\sum_{t=0}^{T-1} \sum_{n=0}^{N-1} \sum_{m=0}^{M-1} \gamma_t(n, m)} - \hat{\mu}_n \hat{\mu}_n'$$

여기에서

$$\mu_n(y_t, n, m, \lambda) = E[n \mid y_t, n, m, \lambda]$$
$$\Sigma_n(y_t, n, m, \lambda) = E[(n_t - \mu_n(y_t, n, m, \lambda)) (n_t - \mu_n(y_t, n, m, \lambda))' \mid y_t, n, m, \lambda]$$

이며, λ는 주어진 모델의 집합을 나타낸다. 그리고, $\gamma_t = p(Y, s_t = n, c_t = m \mid \lambda)$은 모델 파라미터 λ가 주어질 때 관측 벡터열 Y와 n번째 상태의 m번째 gaussian mixture 사이의 joint likelihood이다.

Spectral tilt q도 유사한 방법으로 구할 수 있다.

$$\hat{q} = [\sum_{t=0}^{T-1} \sum_{n=0}^{N-1} \sum_{m=0}^{M-1} \gamma_t(n, m)(\nabla_q f)(I + \nabla_x f)^{-1} \Sigma_{n,m}^{-1} (I + \nabla_x f)^{-t} (\nabla_q f)']^{-1}$$
$$[\sum_{t=0}^{T-1} \sum_{n=0}^{N-1} \sum_{m=0}^{M-1} \gamma_t(n, m)(\nabla_q f)(I + \nabla_x f)^{-1} \Sigma_{n,m}^{-1} (I + \nabla_x f)^{-t} (\nabla_q f)' \cdot q(y_t, n, m, \lambda)]$$

윗식에서, $q(y_t, n, m, \lambda)$은

$$q(y_t, n, m, \lambda) = (\nabla_q f)^{-t} (y_t - (I + \nabla_x f)' \mu_{n,m} - (\nabla_n f)' \mu_n (z_t, n, m, \lambda) - g(n_0, x_0, q_0))$$

이다. 여기에서 $(\cdot)^{-t} = [(\cdot)]^{-1}$이고, $\mu_{n,m}$과 $\Sigma_{n,m}$은 n번째 상태의 m번째 gaussian mixture를 나타낸다.

Ⅳ. 실험 결과

4.1 데이타 베이스의 구성과 특징 파라미터

본 실험에 사용된 인식기는 continuous HMM(CHMM)[11]이며, 인식 실험을 위해 사용된 데이터베이스는 음운학적으로 균형을 이룬 75개의 고립 단어[12]로 이루어져 있다. 녹음은 조용한 사무실 환경에서 이루어졌고 발음한 음성 신호는 16kHz, 16bit로 A/D 변환되었다[13]. 학습 데이터는 15명의 화자가 한번씩 발음한 것으로 구성되었고, 인식 실험에는 학습에 참가하지 않은 5명의 화자가 한 번씩 발음한 것을 사용하였다. 잡음은 가산성 백색 잡음을 컴퓨터에서 만들어 사용하고, 자동차 잡음은 속도 90-120 [km/h]로 고속도로에서 주행시 창문을 닫고 녹음되었다.[1] 본 논문에서 사용된 특징벡터는 MFCC이다.[2] HMM 인식기는 음소단위로 모델링하였으며, 발음사전의 구성은 32개의 phone like unit(PLU)[14]을 이용하였다.

4.2 Baseline 인식기 및 학습과 인식환경이 같은 경우의 인식실험

본 절에서는 환경에 대한 영향을 반영하지 않은 기본적인 HMM인식기를 이용하여 인식 실험 결과를 토의한다. 무잡음 환경 및 가산성 백색잡음과 자동차잡음에 대하여 각각 SNR별로 실험을 하였으며, 결과는 표 1와 같다.

표 1. Baseline 인식기의 결과

	clean	30dB	20dB	10dB	0dB
AWG	93.9	87.2	54.9	24.3	4.7
CAR	93.9	94.4	89.6	63.2	26.4

위의 결과를 살펴보면, 가산성 백색잡음의 경우 20 dB 이하와 자동차잡음의 경우 10 dB이하에서 인식률이 급격히 떨어지는 것을 볼 수 있다. 특히, 0 dB의 가산성 백색 잡음의 경우 인식률이 거의 '영'에 가깝도록 인식기의 성능저하가 나타난다.

그러면, 인식환경과 동일한 환경으로 학습한 HMM인식기의 경우는 인식률이 어느 정도일까? 동일한 환경일 때 인식결과는 표 2와 같다.

표 2. 학습과 인식환경이 같은 경우의 결과

	30db	20db	10db	0db
AWG	94.4	91.5	85.3	56.3
CAR	93.9	93.9	90.1	89.6

인식결과를 살펴보면, 앞의 baseline인식기의 경우보다 매우 높은 인식률을 나타낸다. Baseline인식기에서 저조한 인식률을 나타낸 20 dB이하(가산성 백색 잡음)와 10 dB(자동차잡음) 이하의 인식률이 크게 증가하였다. 특히, 0 dB의 가산성 백색잡음의 경우 인식률이 10배가량 증가하였음을 알 수 있다. 본 실험은 대략적으로 인식률의 상한선을 나타낸다. 이와같은 결과는 학습과정에서 배경잡음에 대한 영향이 충분히 반영되었기 때문이다. 그러나, 이러한 방식은 동일한 인식기를 가지고 새로운 환경에 적용하기 위해서는 변화된 배경잡음을 가진 음성이 학습을 위하여 필요하며 인식기의 재학습이 필요하다. 따라서, 잡음환경의 변화에 상관없이 일정 수준 이상의 인식률을 얻을 수 있는 인식 알고리즘이 필요하다. 이후에는 clean 음성의 파라미터로 학습된 인식기를 가지고 앞에서 설명한 모델 보상 알고리즘을 적용하여 실험을 한다.

4.2.1 각 알고리즘별 인식실험

모델 파라미터 변환 알고리즘을 이용한 인식기의 실험결과를 살펴본다. 본 실험에서는 인식 실험에 더욱 많은 화자를 포함시키기 위하여 전체 화자를 5명씩 4팀으로 나눈다. 학습과 인식 실험용으로 3팀과 1팀으로 구성되도록 만들어 4가지 인식실험을 수행하며 각 실험 결과를 평균을 내어 최종 인식 결과로 사용한다. MPMC-EM과 MPMC-STP는 각각 PMC를 변형시킨 경우로 환경 파라미터 예측시에 EM 알고리즘과 steepest descent 방법을 사용한 경우이며, VTS-0는 영차 VTS 알고리즘을 나타낸다. 각 알고리즘별 인식결과는 표 3, 4와 같다.

표 3. AWG 잡음환경에서의 인식결과

	clean	30dB	20dB	10dB	0dB
PMC	91.4	90.5	81.9	65.0	29.6
MPMC-EM	91.7	90.9	87.0	70.3	31.7
MPMC-STP	91.6	91.1	86.7	70.8	31.5
VTS-0	92.6	91.0	87.0	68.8	27.6

표 4. 자동차 잡음환경에서의 인식결과

	clean	30dB	20dB	10dB	0dB
PMC	91.4	91.3	89.8	85.9	72.5
MPMC-EM	91.7	92.1	90.9	87.8	78.0
MPMC-STP	91.6	91.7	90.7	86.8	73.2
VTS-0	92.6	92.2	91.3	88.9	77.6

인식결과를 살펴보면, 각 알고리즘의 인식결과가 baseline인식기보다 더욱 높은 인식률을 보이며, 학습과 인식환경이 동일할 경우의 인식률에 대략적으로 접근하게 된다. 인식률의 대략적인 순위는 VTS-1, VTS-0, MPMC-EM, MPMC-STP, PMC 순이다. MPMC-EM과 MPMC-STP는

1. 자동차 잡음은 삼성 종합 기술원에서 녹음하여 제공하였다.
2. 본 실험은 energy를 포함한 13차 cepstrum(MFCC)을 사용하였다.

두 방법 모두가 기본적인 PMC방법보다 인식률이 높게 나타났다. 이는 미미한 성능향상이지만, 기본적인 PMC와 달리 공분산을 변화시키지 않고 각 상태별로 믹스처와 관측벡터의 정보를 이용하여 평균벡터만을 변환시키는 것이 global하게 일률적으로 변환하는 것에 비하여 인식률을 높이는 방법이 될 수 있음을 보여주며, 이러한 방법을 분산에도 적용할 경우 인식률 향상이 기대된다.

자동차 잡음에 대한 인식률이 가산성 백색잡음에 대한 인식률보다 높은 이유는 자동차 잡음의 스펙트럼이 저주파 영역에 집중되어 있기 때문이다. 영차 VTS의 경우는 대략적인 알고리즘의 순서가 잡음을 예측한 후 이를 토대로 pre-recognition을 하여, top1을 알아낸 후 이 정보를 이용하여 평균벡터만을 변환시키는 점에서는 modified PMC와 유사하다. 그러나, 분산의 경우 modified PMC 방법은 top1을 찾기전에 기본적인 PMC 방법으로 global하게 분산이 변환된 반면에 영차 VTS 방법은 clean 음성의 분산을 그대로 사용할 수 있다. 이렇게 modified PMC 방법이 영차 VTS 방법에 비하여 많은 모델 파라미터를 변환하였음에도 불구하고 인식 결과는 영차 VTS 방법이 더욱 우수하게 나타났다. 즉, 많은 모델 파라미터를 예측된 배경잡음을 이용하여 변환시켰음에도 불구하고, 그렇지 못한 경우보다도 인식률이 떨어져 나타나는 것은 무슨 이유일까?[3] 이유는 간단하다. 두 방법을 만들어 낸 사용된 가정이 다름이 그 원인이다. Modified PMC 방법은 PMC 방법에 근간을 이루므로 그 가정 또한 비슷하다. 즉, linear 영역에서 clean 음성과 noise가 합해져서 noisy 음성을 만든다. 영차 VTS 방법은 기본적인 가정이 noisy 음성은 환경 파라미터로 나타내지는 여러 요소에 의하여 복합적으로 나타나며 본 실험에서는 환경 파라미터로써 가산성 잡음과 spectral tilt를 사용했다. 두 가정에서 뚜렷이 다른 것은 영차 VTS 방법에서는 spectral tilt 환경 파라미터를 사용한 것이다. Modified PMC 방법은 단순히 배경잡음이 더하여 진다는 사실만을 모델링하였으나, 실제 환경에서의 noisy 음성의 spectrum을 보면, noise와 clean 음성이 단순히 더해져 만들어 내는 spectrum같이 smooth한 곡선을 나타내지 않는다. 따라서, 주어진 관측벡터를 이용하여 spectral tilt로써 spectrum상의 변화를 모델할 수 있는 영차 VTS 방법이 인식률이 더 높게 나타나는 것은 당연한 귀결이다.

위에서의 실험결과를 통하여 다음과 같은 사실을 알 수 있다. Baseline 인식기가 SNR이 감소함에 따라 인식률이 급격히 감소됨을 보았다. 이는 배경잡음의 영향을 baseline 인식기는 전혀 수용할 수 없도록 만들어 졌기 때문이다. 기본적인 PMC의 경우는 적은 데이타에서 얻은 잡음 모델을 이용하여 다른 많은 clean 음성 모델을 변환시킨다. 이러한 결과, 원하지 않던 모델 파라미터까지도 변환이 이루어지며, 이로 인하여 인식률이 떨어지게 된다. Modified PMC는 이러한 global한 모델 파라미터 변환으로 인한 인식률 저하를 해결하기 위해 state-dependent한 결합지수를 사용하여 더욱 세밀한 adaptation이 되도록 하였다. 이와 같은 실험의 결과는 눈에 띌 만큼의 큰 향상은 아니나, 평균에 대한 미세조정만으로도 인식률이 향상되었으므로 분산에 대한 미세조정을 통한 인식률 향상을 기대할만하다. Modified PMC 역시 인식환경과 훈련환경이 같은 경우에 대한 인식률보다는 낮다. 이는 PMC가 가산성 잡음만을 모델하기 때문에 정확한 환경을 모델하지 못하며 1회의 반복만을 할 경우 global minimum으로 수렴한다고 보장할 수 없기 때문이다. 이러한 약점을 해결하려고 한 것이 VTS이다. 특히, 영차 VTS의 경우 분산은 clean 음성 모델의 분산과 동일하고 modified PMC와 동일하게 평균만을 변환시킴에도 불구하고 두 종류의 PMC보다 인식률이 높다. 이는 단순히 잡음만의 환경 파라미터에 의해 모델하는 것보다는 여러 다양한 환경 파라미터에 의한 환경 모델이 필요함을 보인다.

4.2.2 VTS-1의 인식실험

1차 approximation을 이용한 VTS는 잡음의 통계 모델에 따라 매우 큰 성능 차이를 나타내었는데, 우선 앞의 실험에서와 같이 음성 시작 직전의 3 frame을 이용하여 잡음모델을 정하여 실험하였다. 실험결과는 표 5와 같다. 두 번째 VTS-1의 구현은 data-driven 방식의 초기 잡음 통계 방식을 이용하였는데, 묵음구간으로부터 잡음의 평균벡터를 구하고 이 평균벡터와 VTS-0 approximation을 이용하여 인식대상이 되는 음성들로부터 잡음을 추정한다. 추정된 잡음들을 sample로 하여 다시 평균과 분산을 구하여 이를 VTS-1의 잡음 초기 모델로 하여 실험하였다. 이 방법의 특징은 인식대상의 음성만으로도 신뢰도 있는 잡음의 분산모델을 구할 수 있다는데 있으며, 특히 몰리적인 잡음의 모델뿐 아니라 실제 음성안에 잡음벡터들이 끼치는 영향까지 고려된 초기치를 구할 수 있다는 장점이 있다. 표 6에 새로운 방법을 이용한 인식결과를 나타내었는데 다른 방식에 비해 매우 높은 성능을 보임을 알 수 있다.

표 5. VTS-1을 이용한 잡음환경에서의 인식결과

	clean	30dB	20dB	10dB	0dB
AWG	93.6	91.2	86.1	69.6	37.9
CAR	93.6	93.6	91.7	84.8	69.1

표 6. 새로운 초기치 설정 방식에 기초한 VTS-1을 이용한 잡음환경에서의 인식결과

	clean	30dB	20dB	10dB	0dB
AWG	93.6	94.9	92.0	78.7	47.2
CAR	93.6	94.1	94.1	93.3	82.4

첫 번째 방식이 두 번째 방식에 비하여 성능이 떨어지는 이유를 분석해 본 결과 작은 sample로부터 얻은 잡음

의 통계모델에 문제가 있음을 발견할 수 있었다. 즉, 잡음 평균벡터는 어느 정도 신뢰도가 있었지만, 분산의 경우는 부족한 sample 갯수에 의해 매우 작은 값으로 추정되거나 차수의 일반적 성질에 위배되는 경우가 많이 발생하였고, 결과적으로 noisy 모델의 분산을 구하는 과정에 좋지 못한 영향을 끼치게 된다. 따라서, data-driven 방식으로 잡음을 예측한 후 이를 이용하여 잡음의 분산을 예측한 두 번째 방식이 더욱 정확한 분산을 얻을 수 있었으며 인식 결과 또한 더욱 높을 수 밖에 없다.

V. 결 론

Clean 환경에서의 음성인식은 높은 인식률을 보이고 있으나, 인식 시스템 실용화의 관건이 되는 잡음환경에서 음성 인식은 아직은 많은 연구가 필요하다. 잡음환경에서 음성 인식을 위하여 이제까지 많은 방법들이 제안되었으며, 어떤 방법은 실용 시스템에 적용되기도 하였다. 그러나, 기존의 방법은 인식 과정에서 많은 계산이 필요하거나, 환경이 변화될 때마다 새로이 학습을 해야만 하는 부담이 있었으며, 재학습이 필요 없는 경우는 높은 인식률을 얻기 위하여 calibration 음성이 필요하다. 본 논문에서는 재학습이나 calibration 음성을 사용하지 않고 또한 어떠한 잡음에 대한 통계적 자료도 사용하지 않으면서 단지 인식할 단어만을 가지고 모델 파라미터를 변환시켜 좋은 인식 결과를 얻었다.

본 논문에서는 기존의 log 영역에서의 접근 방법보다 구현이 간편한 cepstrum 영역에서의 방법을 제안하였다. Modified PMC의 경우 분산을 변환시키지 않으면서 분산을 변환시킨 경우의 PMC 보다 인식률이 높게 나타났다. 또한, VTS는 보다 정확한 환경을 모델할 수 있으며, 보다 정확한 환경 파라미터를 예측할 수 있으므로 인식률도 다른 모델 보상 방법보다 높게 나타났다. 특히, 영차 VTS의 경우 분산은 clean 음성 모델의 분산과 동일하고 PMC와 유사하게 평균만을 변환시켰으나 다른 방법보다 인식률이 높다. 이와같은 이유는 실제 spectrum상에 나타나는 spectrum 왜곡등을 spectral tilt 환경 파라미터를 이용하여 나타내었기 때문이다. 따라서, 잡음환경에서의 인식을 위해서는 무엇보다 환경에 대한 영향을 정확하게 반영할 수 있는 환경 파라미터를 구하고 이를 정확히 예측할 수 있는 방법을 찾는 것이 중요하며, 분산을 변환시킬 경우 예측된 분산에 대한 신뢰도를 고려하여야만 한다.

참 고 문 헌

1. J. Junqua and J. Haton, Robustness in auto-matic speech recognition. Kluwer academic pub-lishers, 1996.
2. R. Moore, "Signal decomposition using mar-kov model techniques," Tech. Rep.3931, Royal Signal & Radar Establishment Memo. 1986.
3. A. Varga and R. Moore, "Hidden Markov model decomposition of speech and noise," in Proc. ICASSP, pp. 845-848, 1990.
4. Y. Ephraim, "A bayesian estimation approach for speech enhancement using hidden markov models," IEEE Trans. on SP, vol. 40, no. 6, pp. 1303-1316, 1992.
5. M. Kadirkamanathan and A. Varga, "Simulta-neous model re-estimation from contaminated data by composed hidden markov modelling," in Proc. ICASSP, pp. 897-900, 1991.
6. M. Gales and S. Young, "An improved approa-ch to the hidden Markov model decomposition of speech and noise," in Proc. ICASSP, pp. 233-236, 1992.
7. M. Gales and S. Young, "Cepstral parameter compensation for HMM recognition," Speech Communication, no. 3, pp. 233-236, 1993.
8. S. Vaseghi and B. Milner, "Speech recogni- tion in impulsive noise," in Proc. ICASSP, pp. 437-440, 1995.
9. M. Gales and S. Young, "A fast and flexible implementation of parallel model combination," in Proc. ICASSP, pp. 133-136, 1995.
10. 은종관 외, 음성인식을 위한 잡음처리기술 연구, 최종 보고서, 한국 과학 기술원, 12. 1996.
11. L. R. Rabiner and B. H. Juang, "An Intro- duction to Hidden Markov Models," IEEE ASSP Magazine, no. 1, pp. 4-16, Jan. 1986.
12. 정호영, 청각 구조를 이용한 잡음 환경에서의 음성 특징 추출에 관한 연구, 석사 학위 논문, 한국 과학 기술원, 1995.
13. 최인정, 권오욱, 박종렬, 김도영, 정호영, 은종관, "자동 통역용 한국어 음성 데이터베이스", 음성 통신 및 신호처리 워크샵 논문집, pp. 287-290, 1994.
14. 김도영, Hidden Markov model을 이용한 음성인식 시스템의 codebook 최적화에 관한 연구, 석사 학위 논문, 한국 과학 기술원, 1993.
15. M. Gales and S. Young, "Parallel model combination for speech recognition in noise," Tech. Rep. 135, Cambridge university, June 1993.
16. M. Gales, Model-based techniques for noise robust speech recognition. PhD thesis, Gonville and Caius College, Sep. 1995.
17. P. Moreno, Speech Recognition in Noisy Environments. PhD thesis, Carnegie Mellon Univ., April 1996.

▲장 육 현(Yuk Hyeun Chang)

1995년 2월 : 건국대학교 전자공학과 졸업(공학사)

1997년 2월 : 한국과학기술원 전기 및 전자공학과 졸업(공학석사)

현재 : LG 정보통신(주) 중앙 연구소 재직

※주관심분야 : 잡음환경에서의 음성인식, Speech Enhancement, 음성 신호 처리

▲정 용 주(Yong Joo Chung)

1988년 2월 : 서울대학교 전자공학과 졸업(공학사)
1995년 8월 : 한국과학기술원 전기 및 전자공학과 졸업(공학박사)
현재 : LG 정보통신(주) 중앙연구소 재직(선임연구원)
※주관심분야 : 음성인식, 신호처리, 신경망 응용

▲박 성 현(Sung Hyun Park) 1952년 3월 24일생

1978년 2월 : 한양대학교 전자공학과 (공학사)
1990년 2월 : 한국과학기술원 전기 및 전자공학과 졸업(공학석사)
1978년 6월~1987년 6월 : 금성통신연구소(선임연구원)
1987년 7월~1996년 12월 : LG정보통신 중앙연구소(책임/수석연구원)
1997년 1월~현재 : LG정보통신 중앙연구소(연구위원)
※주관심분야 : 음성신호처리, 음성다이얼링, 부가통신서비스

▲은 종 관(Chong Kwan Un) : 음향학회 15권 4E호 참조