잡음하의 음성인식을 위한 스펙트럴 보상과 주파수 가중 HMM

이광석*

A Frequency Weighted HMM with Spectral Compensation for Noisy Speech Recognition

Kwang-Seok Lee

요 약

잡음환경에서의 음성인식은 실제의 환경에서의 음성인식에서 매우 중요한 애로기술로써 이를 해결하기 위한 연구는 꾸준히 연구되고 있다. 따라서 본 연구는 음성인식분야에서 가장 많이 사용하고 있는 HMM처리 시 잡음처리의 문제점을 주파수 가중치 부가 HMM으로 해결하는 방법을 제안하고 그 성능을 인식실험을 통하여 검토하였다. 그 결과 SS처리를 함께 사용하는 MCE μ, MCE-ρ가 가장 잡음에 강한 방식임을 알 수 있었다.

ABSTRACT

This paper is simulation research to improve speech recognition rates under the noisy environment. We examines recognition ratio based on frequency-weighted HMM together with spectral subtraction. As results, frequency-weighted HMM with scaling coefficients is trained as a minimum error classification criterion, and is presents a higher recognition rates in noisy condition than a conventional method. Furthermore, spectral subtraction method gives 11 to 28% improvements for this frequency-weighted HMM in low SNR, and gives recognition rates of 81.7% at 6dB SNR of noisy speech.

* 진주산업대학교 전자공학과

Ⅰ.서 론

HMM은 화자의 개인차등에 따른 음성패턴의 변동을 통계적으로 처리한 후, 그 통계량을 확률적인 형태의 모델에 반영하여 인식하는 방법으로서 확률모델을 사용하기 때문에, 개인차나 조음결합 등의 영향으로 음성패턴의 변동이 보다 정확히 반영되며 음소나 음절단위의 모델을 단어 및 문장 등의 단위로 쉽게 확장할 수 있는 장점을 가지고 있다. 따라서, HMM은 음성 인식분야의 실용화 모델에서 가장 널리 사용되고 있다.

한편 실제 환경에서의 음성인식은 주위의 환경변화에 의해 학습 시와는 달리 인식률이 저하되고 있다.[1]-[4] 그러므로 이러한 잡음음성의 인식성능 개선을 위한 연구는 꾸준히 진행되고 있지만, 이는 대체로 (1)front-end에서의 잡음제거 (2)잡음에 강한 특징파라미터 혹은 HMM이용 (3)인식부의 잡음에의 적응방법 등으로 구현되고 있다.

(1)의 대표적 방법으로 추정된 잡음 스펙트럴을 잡음중첩 스펙트럴로 부터 감소되는 것에 의해 가법성 잡음성분을 추출하는 스펙트럴 절단(SS: Spectral Subtraction)법[3]이 있으나, 이 방법은 정상적 잡음에는 효과가 있으나 잡음의 완전제거에는 문제가 있으며, 특히 비정상적인 잡음에는 효과가 떨어진다는 단점을 가지고 있다. (2)방법은 가중치를 부가한 켑스트럼이나 평활화군 지연 스펙트럴 등의 잡음에 강한 특징파라미터를 이용하는 방법이 있다. 또한 이를 HMM으로 처리하고 역분산을 주파수 가중치로 치환하는 HMM이 있으며, 이는 잡음에 의한 변화가 비교적 작은 경우에 효과가 있지만 변화가 큰 경우에는 대체로 효과가 적은 것으로 알려져 있다. 그리고 (3)의 대표적 방법은 잡음 HMM과 무잡음 HMM으로부터 잡음중첩 HMM을 합성하는 PMC와 NOVO법 등이 제안되고 있으나 이들 역시 정상잡음에 대해서는 레벨이나 스펙트럴이 변하는 잡음에 대해서는 빠른 적응이 곤란하다는 문제가 있다. 따라서, 본 연구에서는 front-end에서 스펙트럴 절단법을 이용하여 잡음에 의한 스펙트럴 변동을 감소시키고 잡음 스펙트럴의 급격한 변동으로 인해 제거되지 않았던 잡음성분과 스펙트럴 절단법에서 발생하는 문제점에 대해 주파수 가중치 부가 HMM으로 처리하는 방법[3]-[4]을 제안 검토하였으며, 또한 이 주파수 가중치부가 HMM의 가중치 특성을 개개의 성분분포 마다 독립적으로 식별에러최소기준(MCE) 학습을 행하는 방법을 검토하였다. 위의 두가지 방법을 함께 사용함으로써 스펙트럴 변동이 주파수 가중치부가 HMM의 허용범위 내에 수렴하며 광범위의 잡음환경에 인식적응이 가능하리라 생각한다. 이를 위하여, 우선 주파수 가중치부가 HMM과 MCE학습에 대해 논하며 스펙트럴 절단법 및 그 조합방법에 대하여 불특정화자의 숫자음성인식실험으로 그 유효성을 검토하였다.[1]-[6]

Ⅱ.주파수 가중치 부가 HMM

1. 특징 파라미터

특징 파라미터로써 잡음에 강한 평활화군 지연 스펙트럴을 사용하였다. 우선 p차의 켑스트럼을 식(1)과 같이 두며, 평활화군 지연 스펙트럴은 식(2)와 같이 가중치 부가 켑스트럼의 변환으로 얻을 수 있다.

$$x = [c_1,\ c_2,\ \cdots,\ c_p] \quad \cdots\cdots (1)$$

$$T(\lambda_i) = 2\sum_{n=1}^{p} nc_n \cdot \cos(n\lambda) \quad \cdots\cdots (2)$$

$T(\lambda)$를 아래와 같이 표본화한 P점의 이산 평활화군 지연 스펙트럴 $T(\lambda_i)$를 성분으로 하는 특징벡터 y를 이용하였으며 식(3)에 나타내었다.

여기서, C는 (p×p)차원의 여현변환 행렬이다.

$$\lambda_i = 2\pi \frac{i}{2P+1}, \quad (i = 1,2,\cdots,p)$$

$$y = [T(\lambda_1),\cdots,T(\lambda_p)]^T = Cx \quad \cdots\cdots (3)$$

평활화군 지연 스펙트럴은 잡음중첩에 의해 스펙트럴의 변동에 강하고 SNR이 좋은 스펙트럴 피크에 대하여 감도가 좋기 때문에 잡음에 우수하다고 알려져 있다.

2. 주파수 가중치 부가 HMM

주파수 가중치 부가 HMM에서 이산 평활화군 지연 스펙트럴 **y**의 내잡음성 효과를 유효 적절하게 얻기 위하여 잡음중첩에서 스펙트럴 변동과 스펙트럴의 청각도를 고찰하고 모델 c, 상태 s 및 분기 m의 성분분포 $P_{csm}(y)$의 공분산 행렬을 다음

과 같이 주파수가중 행렬 $\rho_{csm}W_{csm}^{-1}$으로 치환한다.

$$P_{csm}(y) = \frac{1}{\sqrt{(2\pi)^p \mid \rho_{csm}W_{csm}^{-1} \mid}}$$

$$\cdot \exp\left\{-\frac{(y-\mu_{csm})^T W_{csm}(y-\mu_{csm})}{2\rho_{csm}}\right\}$$

여기서, ρ_{csm}는 W_{csm}^{-1}을 공분산으로 변환하기 위한 크기 성분이다. 주파수 가중행렬은 잡음중첩에서 스펙트럴의 각 주파수성분의 통계적 변동이 power에 반비례한다고 가정하고 청각감도를 고려하여 구한 주파수 가중함수을 기초로 분포의 분산을 미리 주기 위한 것이다.

3. 주파수 가중치 부가 함수

주파수 가중치 부가 함수로는 평균 스펙트럴과 저역 가중 함수로부터 구한 다음의 함수를 이용한다.

$$\omega_{csmi} = \mid 1+\alpha_{csm}\exp(j\lambda_i) \mid^2 \cdot \exp\{\beta_{csm}l_{csmi}\} \quad \cdots\cdots (5)$$

$$[l_{csm1}, \cdots, l_{csmp}]^T =$$

$$C\ diag\left[\frac{1}{1}, \cdots, \frac{1}{q}, 0, \cdots, 0\right]C^{-1} \cdot \mu_{csm} \quad \cdots\cdots (6)$$

여기서, 가중치 함수의 특성은 q: 평활도, α: 고역감쇠계수, β: 압축계수 등의 파라미터로 결정하였다.

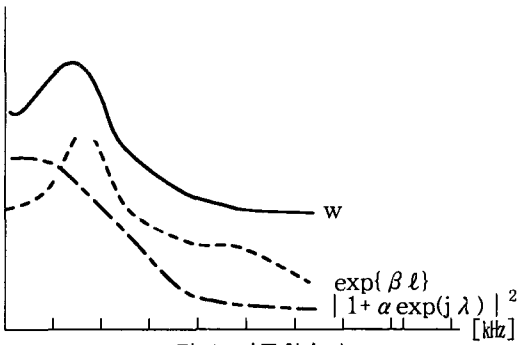

그림 1. 가중함수치

Fig. 1 Value of weighting function

식(5)의 주파수 가중계수는 α=0, β=0 로 가중치가 없는 것이고 공분산 행렬은 대각 등분산이다. α가 1에 가까워질 수록 고역주파수 성분의 가중치가 작게 되고 잡음중첩에 의한 고역의 주파수 부분의 변동으로의 영향이 작게 된다. 또한, 식(5)의 지수함수부는 평균 스펙트럴로부터 유도된 power 스펙트럴을 β승으로 압축한 것으로 β가 크게될수록 주파수 가중함수의 피크부분과 곡부분의 차가 크게되고 피크부분의 변동에 대한 감도가 증대한다. q는 식(6)에서 알 수 있듯이 평균 스펙트럴을 대수 스펙트럴로 역여현 변환할 때의 스펙트럴 평활도를 조정하는 차수이다. 또한 스케일링 계수 ρ_{csm}의 초기치로 다음의 정규화 스케일링을 이용하였다.

$$\rho_{csm} = \frac{1}{\sum_i \omega_{csmi}} \quad \cdots\cdots (7)$$

앞에서 제안한 주파수 가중치 부가 HMM에서는 가중 파라미터를 전모델의 전분포에 대해 공통으로 하고 실험적으로 결정하였다. 그러나, 여기서는 내잡음성을 간단히 개선하기 위하여 이러한 파라미터를 각 모델, 각 성분분포마다 독립으로 하고 MCE학습으로 학습하였다.

4. MCE에 의한 가중치 특성 학습

MCE에 대하여 k카테고리의 n학습단어의 관측벡터열 Y_{kn}에 대한 오차함수 $d(Y_{kn}, \Theta)$와 손실함수 $l(d(Y_{kn}, \Theta))$을 다음과 같이 정의한다.

$$d(Y_{kn}, \Theta) = -\log(P(Y_{kn}, x_c \mid \theta_c))$$

$$+ \log\left(\left[\frac{1}{h}\sum_{c,c\neq k}^{h} P(Y_{kn}, x_c \mid \theta_c)^{\eta}\right]^{\frac{1}{\eta}}\right) \quad \cdots\cdots (8)$$

$$l(d(Y_{kn}, \Theta)) = \begin{cases} d(Y_{kn}, \Theta) & d>0 \\ 0 & d\leq 0 \end{cases} \quad \cdots\cdots (9)$$

여기서, x_c는 Y_{kn}에 대한 모델 c의 Viterbi 최적상태 천이계열이며 θ_c는 c번째의 HMM 파라미터를 나타낸다. 오차함수는 정해단어 k모델에 대한 오차와 부정해단어에 대한 오차와의 평균치(높은 쪽부터 순서대로 h개를 대상으로 한다.)의

차이다. 이 값이 양일 때는 부정해 단어의 오차가 크기 때문에 식별오차가 있음을 표시하고, 음의 경우에는 식별이 바르다는 것을 표시한다. 이것을 이용하여 총 손실을 다음과 같이 정의한다.

$$L(Y,\Theta) = \sum_{k=1}^{K}\sum_{n=1}^{N_k} l(d(Y_{kn},\Theta)) \quad \cdots\cdots (10)$$

총 손실 $L(Y,\Theta)$를 최소화하기 위하여 최소자승오차(LMS)법으로 학습하며 학습의 반복에 의하여 HMM 파라미터 Θ는 다음과 같이 변경시킨다.

$$\Theta(n) = \Theta(n-1) - \varepsilon(n)\nabla L(Y,\Theta) \quad \cdots\cdots (11)$$

$\nabla L(Y,\Theta)$에 관하여 모델 c의 상태 s, 분기 m의 파라미터 θ_{csm}에 관한 성분은 식(12)와 같이 주어지며 여기서, 손실함수의 도함수는 식(13)과 같이 주어진다.

$$\frac{\partial L(Y,\theta)}{\partial \theta_{csm}} = \sum_{k=1}^{K}\sum_{n=1}^{N_k} \frac{\partial l(d(Y_{kn},\Theta))}{\partial d(Y_{kn},\Theta)} \frac{\partial d(Y_{kn},\Theta)}{\partial \theta_{csm}} \quad \cdots\cdots (12)$$

$$\frac{\partial l(d(Y_{k,n},\Theta))}{\partial d(Y_{k,n},\Theta)} = \cdot \begin{cases} 1 & d>0 \\ 0 & d\le 0 \end{cases} \quad \cdots\cdots (13)$$

또한, 오차함수의 도함수는 각각 식(14), 식(15)와 같다.

(1) c=k의 경우

$$\frac{\partial d(Y_{kn},\Theta)}{\partial \theta_{csm}} = -\frac{\partial g(Y_{kn},\theta_c)}{\partial \theta_{csm}} = -\sum_{t=1}^{T_{kn}} \delta(x_{ct}-s)\frac{\partial \log(b_{cs}(y_{knt}))}{\partial \theta_{csm}} \quad \cdots\cdots (14)$$

(2) c≠k의 경우

$$\frac{\partial d(Y_{kn},\Theta)}{\partial \theta_{csm}} = \frac{\partial G(Y_{kn},\theta)}{\partial \theta_{csm}} = \frac{\exp(\eta g(Y_{kn},\theta_c))}{\sum_{j=1}^{k}\exp(\eta g(Y_{kn},\theta_j))} \cdot \sum_{t=1}^{T_{kn}}(x_{ct}-s)\frac{\partial \log(b_{cs}(y_{knt}))}{\partial \theta_{csm}} \quad \cdots\cdots (15)$$

여기서, x_{ct}는 관측치열 Y_{kn}에 대한 모델 c에 관한 시간 t의 상태이다. 또한 MSE에 의한 제n번째 학습반복에 의해 스텝크기 $\varepsilon(n)$는 $\nabla L(Y,\Theta)$의 전회와의 방향여현 $\cos\phi_{n,n-1}$를 이용하여 다음과 같이 조정하였다.

$$\varepsilon(n) = \varepsilon(n-1)\cdot 2^{\cos\phi_{n,n-1}} \quad \cdots\cdots (16)$$

III. 스펙트럴 절단

일반적인 스펙트럴 절단법에서 추정음성의 power스펙트럴 $\hat{X}_k$는 다음과 같이 주어진다.

$$\hat{X}_k = \max(Y_k - \alpha\hat{N}_k,\ \beta\hat{N}_k) \quad \cdots\cdots (17)$$

여기서, 첨자 k : FFT분석시 k번째의 스펙트럴 성분, Y_k : 잡음부가음성의 power 스펙트럴, $\hat{N}_k$: 잡음의 평균 power 스펙트럴, α : 절단계수 등을 표시한다. 식(17)에서의 처리는 입력음성으로부터 추정잡음의 α배의 차를 나타내고 있으며 계수 β를 작게 하면, 절단후의 스펙트럴의 곡이 크게 되고 값을 크게 하면, power가 작은 성분이 추정잡음의 β배로 바뀌며 잡음저감효과는 작아지게 된다. 따라서, 본 연구에서는 실험적 최적치로 α와 β를 각각 2.0, 1.0으로 설정하였으며 $\hat{N}_k$는 무음구간 30프레임의 평균 스펙트럴로 추정하였다.

IV. 음성 인식 실험

1. 음성 데이터 및 분석조건

인식평가 실험에서는 ETRI 표준 음성DB 중, 남녀 각 30명이 5회 발성한 한국어 숫자음성(1~9)을 사용하였으며 이중 20명분을 불특정화자의

HMM작성용 학습데이터로 하고 나머지 10명분은 인식실험에 사용하였으며 부가잡음으로는 백색가우스 잡음, 자동차 내 잡음, 음성잡음 및 공장내 잡음 등을 사용하였다. 분석조건은 표 1과 같다.

표 1. 음성분석조건
Table 1. Conditions of speech analysis

-표본화주파수	16kHz
-분석창	20 ms 해밍창
-프레임주기	5 ms
-LPC 분석차수	고정 $1-0.98z^{-1}$
-특징파라미터	16차 멜-켑스트럼
-단어 HMM	혼합무상관정규분포형
-상태수	16상태 15 loop
-분기수	분기수 : 1~8

2. 음성 인식 실험

(1) 주파수가중 HMM의 MCE학습효과

주파수가중 HMM의 각 파라미터를 MCE학습한 효과를 검토하기 위하여 (1)무잡음 음성에서 학습한 단일 무상관 HMM를 초기 모델로 정규화형 스케일링을 이용하여 가중특성을 실험적으로 결정(α=0.42, β=0.62, q=2.0)한 주파수 가중치 부가 HMM(HMM-a) (2)HMM-a를 초기모델로 무잡음 및 잡음부가음성(백색잡음 0dB 또는 자동차 잡음 0dB)을 이용하고 ρ_{csm}, β_{csm}, α_{csm}들 중 1파라미터만을 MCE 학습(η=2, n=5)한 HMM(MCE-ρ, MCE-β, MCE-α)에 관하여 백색잡음 및 자동차 잡음하의 인식실험을 행하고 그 결과를 그림3~4에 각각 나타내었다.

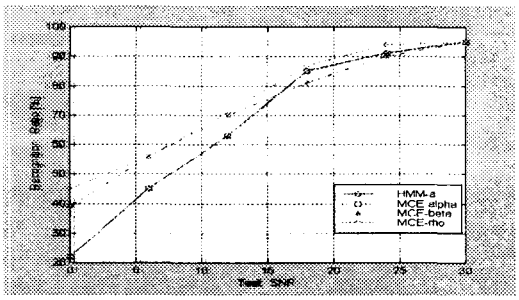

그림 3.백색잡음에서의 인식률
Fig. 3 Recognition rates with white noise

그 결과, ρ 및 β의 MCE학습에서는 거의 같은 정도의 인식률을 얻었으며 종래의 HMM-a에 비해 0~6dB의 백색잡음에서 12~22%, 또한 -12~-6dB의 자동차 잡음에서는 약 10~8%의 인식률이 개선되었다. 또한, α에 관해서는 효과가 거의 없었으며, 복수의 파라미터로 MCE학습한 HMM은 단일 파라미터의 학습에서 손실이 0에 근접하기 때문에 거의 효과가 없었다. 위의 결과로부터 다음절의 스펙트럴 절단을 함께 사용한 실험에서는 MCE-ρ를 사용하였다.

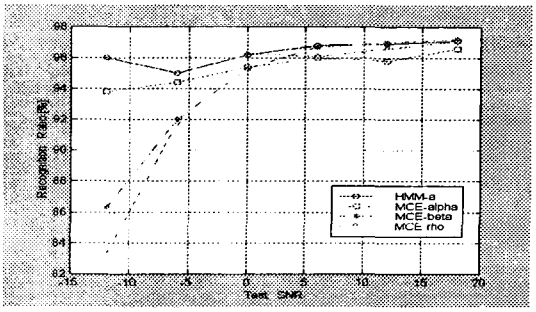

그림 4. 자동차잡음에서의 인식률
Fig. 4 Recognition rates with car's noise

(2) SS의 병용 사용 효과

다음의 조건에 의해 HMM을 작성하였다.

표 2. HMM의 학습조건
Table 2. Training condition of HMM

구 분	초 기 모 델	학습파라미터	학습음성	비 고
HMM-a	-	평균스펙트럴,분산	무잡음 음성	
HMM-b	-	평균스펙트럴,분산	SS처리후의 3종류 음성	※3종류 음성 -무잡음 음성 -백색잡음: 0dB -자동차잡음: -6dB
MCE-ρ	HMM-a	스케일링 계수	SS처리후의 3종류 음성	
MCE-μ	HMM-a	평균스펙트럴	SS처리후의 3종류 음성	

(1) 무잡음 음성으로 학습한 통상의 무상관 HMM (HMM-b) (2) 무잡음 음성, 백색잡음: 0dB, 자동차잡음: -6dB의 3종류의 음성을 SS처리한 데이터를 사용하여 학습한 무상관HMM(HMM-c)

(3) HMM-a를 초기모델로 하고 무잡음 및 잡음 부가음성(백색 SNR: 0dB, 자동차 SNR: -6dB)을 사용하고 ρ_{csm}을 MCE학습한 주파수가중치 부가 HMM(HMM-ρ) (4) HMM-a를 초기 모델로 하고 (2)와 같은 3종류의 데이터로 평균 스펙트럴만을 MCE학습한 주파수가중치 부가HMM(HMM-μ). 이러한 HMM의 명칭, 초기모델 등에 대해 표 2에 나타내었으며 인식실험 결과를 그림 5~8에 각각 나타내었다.

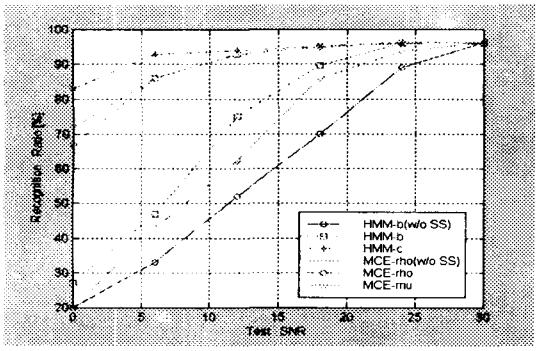

그림 5. 학습내 잡음에 대한 인식률(백색잡음)
Fig. 5 Recognition rates with training speech(white noise)

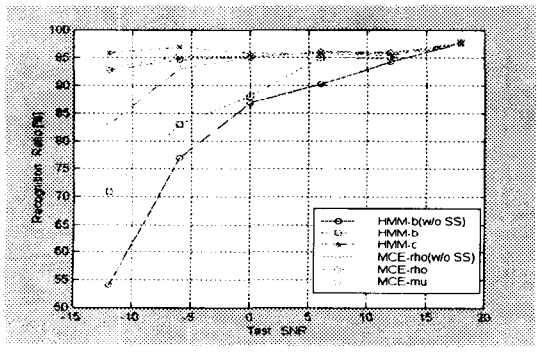

그림 6 학습내 잡음에 대한 인식률(자동차 잡음)
Fig. 6 Recognition rates with training speech(car's noise)

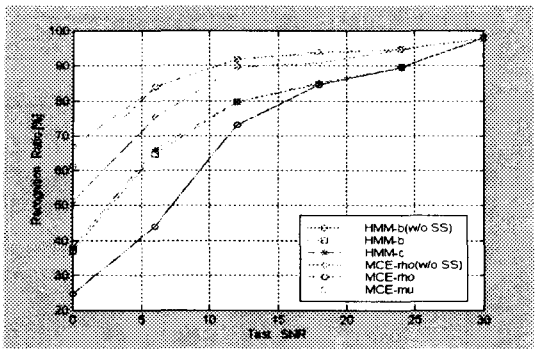

그림 7 학습외 잡음에 대한 인식률(음성잡음)
Fig. 7 Recognition rates (speech noise)

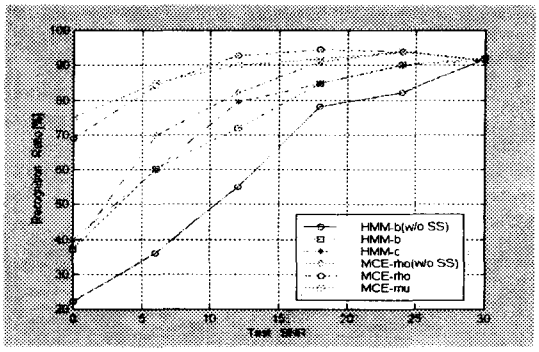

그림 8 학습외 잡음에 대한 인식률(공장잡음)
Fig. 8 Recognition rates (factory's noise)

위의 인식결과에서 인식시의 잡음환경이 학습시와 같은 경우에는 SS처리 후의 데이터로 학습한 HMM이 가장 높은 인식률을 얻지만, SS처리와 주파수 가중치 부가 HMM과의 조합하는 경우와 비슷한 인식정도를 가진다. 한편, 학습외 잡음환경에서는 통상의 HMM의 인식정도는 매우 낮으며, SS처리 없는 MCE-ρ보다도 인식률이 낮다는 것을 알 수 있었으며 결국, SS처리를 함께 사용하는 MCE-μ, MCE-ρ가 가장 잡음에 강한 방식임을 알 수 있었다. 또한, 통상의 HMM-a에서는 SS에 의해 10~20%의 인식률을 개선시키지만, 백색잡음을 제외한 SS처리 없는 MCE-ρ에는 영향이 없음을 알았으며 주파수 가중치 부가

HMM의 스케일링 계수를 MCE 학습한 MCE-ρ와 SS처리를 함께 사용하는 것보다 학습내·외를 막론하고 인식정도가 개선되며 특히, SNR부에서의 SS처리 효과가 크다. 예를 들면, 백색잡음에서의 0dB에서 21%에서 약 72%까지 개선되며 MCE-μ는 평균 스펙트럴이 SS후의 잡음음성에 가깝기 때문에 저 SNR부에서 MCE-ρ보다도 약간 인식률이 개선됨을 알 수 있었다.

V. 결 론

스펙트럴 보상법에 대해서 SS법을 사용하고 제거시킨 잡음 및 SS처리와 함께 주파수 가중치 부가HMM에서 처리하는 것에 의해 보다 광범위한 잡음환경, 특히 학습외 잡음에 대해서 높은 인식정도가 얻어지는 것을 알 수 있었다. 향후, 주파수 가중치 부가함수의 개선 및 다양한 잡음환경에서의 인식실험을 계속 연구해 나갈 예정이다.

참고 문헌

[1] F. Martin et al., "Recognition of noisy speech by hidden marcov models", 信學技報, pp. 9-16, 1992

[2] 妹背, 松本, "雜音下音聲認識における周波數重み付けHMMの改良と評價", 信學技報, pp.25-32, SP93-107, 1993

[3] P. Lockwood and J. Bondy, "Experiments with a nonlinear spectral subtractor(NSS), hidden marcov models and the protection for robust speech in car", Speech Communication 11, pp. 215-228, 1992

[4] 小野, 宋本, "誤り最小基準によるHMMの學習", 日本音響學會講演論文集, 3-5-10, 1996

[5] M.Berouti, R. Schwartz, and J. Makhoul, "Enhancement of speech corrupted by acoustic noise", Proc. of ICASSP '79, pp. 208-211

[6] Gales MJF and Young SJ, "Parallel model combination for speech recognition in noise", Technical ReportCUED/FINFENG/TR135, 1993

[7] Rabiner L. R., Wilpon J. G., Soong F. K., "High performance connected digit recognition using hidden markov models", IEEE Trans. on Acoustics, Speech, Signal Processing, vol.4 ASSP-37, pp 1214-1225, Aug. 1989.

[8] 이광석, 심장엽, 이영재, 고시영, 허강인, "HMM에 의한 연속음성인식 시스템의 구현", 제13회 한국음향학회 음성통신 및 신호처리워크샵 논문집 제13권1호, pp. 325-330, 1996.

이 광 석(Kwang-Seok Lee)

1983년 2월 동아대학교 전자공학과 졸업(공학사)

1985년 2월 동아대학교 대학원 전자공학과 졸업(공학석사)

1991년 2월 동아대학교 대학원 전자공학과 졸업(공학박사)

1995년~현재 진주산업대학교 전자공학과 조교수

※관심분야 : 음성신호처리 및 인식, 지능시스템 통신 시스템