Journal of Korea Multimedia Society Vol. 11, No. 10, October 2008(pp. 1325-1337)

ARMA 필터를 이용한 로그 에너지 특징의 정규화 방법

신광호[†], 정호열[††], 정현열[†††]

요 약

훈련과 인식의 환경적 차이가 음성 인식 성능 저하의 주요 요인이며, 이러한 환경적 불일치를 줄이기 위한 다양한 잡음 처리 방법들이 연구되고 있다. 이 가운데 로그 에너지 특징에 대한 ERN(log-Energy dynamic Range Normalization), SEN(Silence Energy Normalization) 등이 우수한 성능을 보이고 있다. 그러나 이들 방법은 상대적으로 큰 값을 갖는 로그 에너지 특징에 대해서는 처리가 불가능한 문제점이 있으며, 특히 SNR값이 작은 환경에서는 이러한 문제로 인하여 환경적 불일치가 더욱 크게 나타나고 있다. 이를 해결하기 위해서 본 논문은 자동 회귀 방식으로 이동 평균을 계산하여 로그 에너지 특징을 스무딩(smoothing)하는 ARMA(Auto-Regression and Moving Average) 필터를 후처리로 적용하는 방법을 제안한다. Aurora 2.0 DB를 이용한 인식 실험 결과, 제안 방법이 기존의 방법들에 비해 향상된 인식 결과를 얻을 수 있었다.

A Log-Energy Feature Normalization Method Using ARMA Filter

Guanghu Shen[†], Ho-Youl Jung[††], Hyun-Yeol Chung[†††]

ABSTRACT

The difference of environments between training and recognition is the major reason of degradation of speech recognition. To solve this mismatch of environments, various noise processing methods have been studied. Among them, ERN(log-Energy dynamic Range Normalization) and SEN(Silence Energy Normalization) for normalization of log energy features show better performance than others. However, these methods have a problem that they can hardly achieve normalization for the relatively higher values of log energy features and the environmental mismatch caused by this problem becomes bigger especially in low SNR environments. To solve these problems, we propose applying ARMA filter as post-processing for smoothing log energy features by calculating the moving average in auto-regression scheme. From the recognition results conducted on Aurora 2.0 DB, the proposed method shows improved recognition results comparing with conventional methods.

Key words: log energy normalization(로그 에너지 정규화), ARMA filter(ARMA 필터), speech recognition(음성인식)

1. 서 론

음성 인식 기술은 컴퓨터가 인간의 음성을 식별하는데 필요한 여러 가지 특징 정보를 분석, 추출한 다음 이를 인식을 위한 특징 파라미터의 열로서 표현한 후 패턴 인식 방법을 이용하여 언어 정보를 추출하는 기술이다. 현재의 음성 인식 기술은 잡음이 없는 환경에서는 상당히 좋은 결과를 보이고 있으나, 잡음이

※ 교신저자(Corresponding Author) : 정현열, 주소 : 경상북도 경산시 대동 214-1영남대학교(712-749), 전화 : 053)810-2496, FAX : 053)810-4742, E-mail : hychung@ynu.ac.kr
접수일 : 2008년 5월 21일, 완료일 : 2008년 8월 7일
[†] 정회원, 영남대학교 정보통신공학과 박사과정 (E-mail : guanghosin@ynu.ac.kr)
[††] 정회원, 영남대학교 전자정보공학부 부교수 (E-mail : hoyoul@ynu.ac.kr)
[†††] 정회원, 영남대학교 전자정보공학부 교수

존재하는 환경에서 보여주는 낮은 인식률 때문에 실제 환경에서의 응용에는 어려운 점이 있다. 이로 인하여 많은 연구자들이 잡음 환경 하에서의 음성 인식 시스템의 성능을 높이는 데 꾸준한 노력을 경주하고 있다.

음성 인식 시스템의 성능 저하는 주로 학습 환경과 인식 환경의 불일치(mismatch)에서 초래된다. 이러한 불일치를 줄이기 위하여 다양한 접근 방법들이 제안되고 있는 데 다음과 같이 크게 세 가지로 나눌 수 있다. 첫 번째는 음성 개선(speech enhancement) 방식, 두 번째는 특징 개선(feature enhancement) 방식 또는 강인한 특징 추출(robust feature extraction) 방식, 세 번째는 모델 보상(model compensation)방식이다[1]. 저자들은 특히 두 번째의 특징 개선 방식에 관심을 가지고 잡음에 강인한 특징을 얻기 위하여 특징 파라미터 영역에서의 정규화 방법에 대하여 지속적으로 검토해 오고 있다.

음성 인식을 위하여 추출된 특징은 주로 에너지 특징과 켑스트럼 특징으로 나눌 수 있다. 현재까지 이 두 특징의 정규화 방법에 대한 연구 결과는 상당히 많이 발표되고 있다. 먼저, 에너지 특징 개선에 관해 지금까지 발표된 연구에 대해 살펴보면, Hwang et al.,[2]은 프레임 에너지 특징을 HPF (high-pass filter)에 통과시켜 에너지의 delta 특징의 왜곡을 줄여주는 방법을 제안하였으며, Zhu et al.,[3]은 로그 에너지 특징의 동적 변화 범위를 정규화(ERN: log-Energy dynamic Range Normalization)하여 훈련과 인식의 환경적 불일치를 줄여주는 방법을 제안하였다. 그리고 Veth et al.,[4]은 음성 검출기(VAD: Voice Activity Detection)기반으로 프레임 벡터를 선택(FVS: Frame Vector Selection)하여 잡음으로 판별되는 프레임을 음성 인식 과정에서 배제하는 방법을 제안하였고 최근에는 이 방법을 개선한 묵음 에너지 정규화(SEN: Silence Energy Normalization)방법[5]도 소개되고 있다. SEN 는 음성 신호의 로그 에너지 특징을 IIR HPF(Infinite Impulse Response High-Pass Filter)를 통과시킨 후의 평균값을 문턱치 값으로 설정하고 원 로그 에너지 값이 문턱치 값보다 작은 경우에만 정규화하는 방법이다.

다음은 켑스트럼 특징에 대한 정규화 방법으로서 이 중 대표적인 방법으로는 통계적 보상 방법[6], RATZ(multi-vaRiate gAussian based cepsTrum normaliZation)[7], 켑스트럼 평균 정규화(CMN: Cepatral Mean Normalization)[8], 켑스트럼 분산 정규화(CVN: Cepatral Variance Normalization)[9], 켑스트럼 평균 및 분산 정규화(CMVN: Cepatral Mean and Variance Normalization)[9] 등이 있으며, 이 중에서 통계적 보상 방법과 RATZ 방법은 많은 적응 데이터가 필요하며 또한 잡음 환경이 변할 경우 그에 따른 보정이 필요하다는 단점이 있다[6,7]. 반면에 CMN, CVN 및 CMVN은 다른 기법들에 비해 상대적으로 간단하면서도 효과적인 방법으로 알려지고 있다[9].

이뿐만 아니라, ARMA(Auto-Regression and Moving Average) 필터를 CMVN 등의 후처리로 적용하는 CMVN-A[10-12], 또는 에너지와 켑스트럼 특징에 정규화 방법을 동시에 적용하는 ERN-CMN [3], ERN-CVN[3] 및 SEN-CMVN[5] 등도 등장하였으며, 이들을 적용한 음성 인식기의 인식 성능은 기존의 방법에 비해 향상된 결과를 가져온 것을 확인하였다. 이와 같은 연구결과를 종합해 보면 기존의 정규화 방법들에 대한 분석에 근거하여 ARMA 필터를 ERN 및 SEN의 후처리로 적용하는 경우와 여기에 다시 기존의 CMVN-A를 결합하는 경우 잡음 환경하의 음성 인식 성능을 향상시킬 수 있을 것으로 기대된다.

따라서 본 논문에서는 ERN 및 SEN의 후처리로 ARMA 필터를 적용하는 ERN-A 및 SEN-A, 그리고 여기에 다시 기존의 CMVN을 추가하는 [ERN-CMVN]-A 및 [SEN-CMVN]-A를 제안한다. 이렇게 할 경우, ERN과 SEN방법이 주로 기준 값 이하의 로그 에너지 특징에 대한 잡음의 영향에 대해서만 효과적인 정규화 처리가 가능하고 그 이상의 로그 에너지 특징에 대해서는 불가능한 것을 ARMA 필터를 후처리 장치로 추가하여 얻을 수 있는 스무딩(smoothing) 효과를 이용하여 특징 파라미터간의 불일치를 줄여줄 수 있으므로 성능향상을 기대할 수 있다.

본 논문의 구성은 다음과 같다. 2장에서 로그 에너지에 대한 정규화 방법, 3장에서 켑스트럼 특징에 대한 정규화 방법, 4장에서 본 논문에서 제안하는 로그 에너지 특징의 정규화 방법을 소개하고, 5장에서 인식 실험 및 고찰을 한 후 마지막으로 6장에서 결론을

맺는다.

2. 로그 에너지 정규화 방법

음성 인식에서 사용하고 있는 특징 중에 에너지 특징은 유성음과 무성음을 구분시켜주는 등 음소간의 변별력을 향상시키기 때문에 켑스트럼 특징과 함께 널리 사용되고 있다. 그러나 에너지 특징은 사람의 목소리 크기와 잡음 환경의 변화에 따라 그 분포가 크게 달라지기 때문에, 로그 함수를 취한 로그 에너지를 보편적으로 사용하고 있다. 이외에 잡음 음성 인식을 위한 에너지 특징의 정규화 방법들에 대한 연구도 계속 진행되고 있다.

전통적인 로그 에너지 정규화 방법은 크게 두 종류로 분류할 수 있는 데 그 첫 번째는 HTK[14]에서 지원하는 모든 프레임으로부터 최대의 로그 에너지 값을 찾아서 각 프레임에 빼주므로 최대의 로그 에너지의 값을 0으로 만들어주는 방법이다. 그러나 이 방법은 화자간의 발성 크기의 차이 등은 보완할 수 있으나 잡음의 영향을 제거하는 데에는 효과적이지 못하다[14]. 두 번째는 로그 에너지를 켑스트럼 특징과 같이 취급하여 평균 및 분산 정규화하는 방법이다. 이 방법도 로그 에너지와 켑스트럼 특징이 지니고 있는 서로 다른 특성으로 인하여 잡음 환경에서 효과적이지 못하는 문제점을 가지고 있다[13].

이를 보완하는 방법으로 ERN과 SEN이 있다. 이 두 방법은 잡음이 음성 신호에 부가되기 전과 후를 관찰하여 잡음이 부가될 때 음성 신호가 잡음의 영향을 크게 받는 음성 구간을 찾아서 이를 정규화하여 로그 에너지 특징의 잡음에 대한 강인성을 향상시키는 방법이다. 이하 로그 에너지 특징의 정규화 방법인 ERN과 SEN에 대해 각각 살펴본다.

2.1 ERN(log-Energy dynamic Range Normalization)

잡음 환경에서 발성된 음성 신호는 큰 에너지 값을 가지는 음성 구간에서는 부가 잡음의 영향을 거의 받지 않으나 작은 에너지 값을 가지는 구간에서는 큰 영향을 받는다. 이 점에 착안하여 ERN은 음성 신호의 특징 파라미터 중에서 에너지 특징을 그림 1과 같이 처리한다. 즉, 작은 값을 가지는 로그 에너지 특징의 구간에서는 큰 에너지를 가지는 구간에

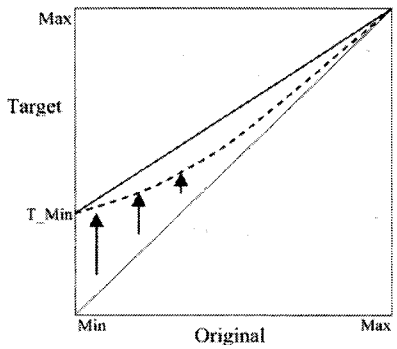


그림 1. ERN의 처리 과정

비해 상대적으로 로그 에너지 값을 더 많이 키워서 잡음이 포함한 음성 신호의 로그 에너지 특징의 크기와 비슷하게 하여 훈련과 인식 환경의 불일치를 줄여준다. 이를 위하여 각 음성 신호의 로그 에너지가 동일한 변화 범위를 갖기 위한 $D.R$(dynamic range)함수는 다음과 같이 정의한다.

$$D.R(dB) = 10 * \frac{Max(\log E_n)}{Min(\log E_n)},\ n = 1, \ldots, N \tag{1}$$

여기서 $Max(\log E_n)$는 N개 프레임에서 최대의 로그 에너지 값, $Min(\log E_n)$는 최소의 로그 에너지 값을 나타낸다. $D.R$은 각 음성 신호의 로그 에너지 특징에 대한 최대값과 최소값의 차이를 나타내는 척도이다.

이와 같이 $D.R$의 값이 정해지면 식 (2) 또는 식 (3)을 이용하여 각 음성 신호에 대한 로그 에너지 특징의 최소값 T_Min(Target Minimum)을 계산할 수 있다.

$$D.R(dB) = 10 * \frac{Max(\log E_n)}{T\_Min} \tag{2}$$

$$T\_Min = 10 * Max(\log E_n)/D.R(dB) \tag{3}$$

ERN의 처리 과정은 다음과 같다.

Step 1: N개의 프레임 중에서 로그 에너지의 최대값과 최소값을 찾는다.

$$Max = \max_{n=1,\ldots,N}\{\log E_n\},\quad Min = \min_{n=1,\ldots,N}\{\log E_n\} \tag{4}$$

Step 2: Target Minimum $T_Min = \alpha * Max$을 정한다.

Step 3: N개의 프레임 중에서 로그 에너지의 최소값이 T_Min 보다 작으면 식 (5) 또는 식 (6)을 이용하

여 정규화된 로그 에너지 $\overline{\log E_n}$를 얻는다.

이때 정규화된 로그 에너지 $\overline{\log E_n}$를 구하는 방법은 다음과 같은 두 가지가 있다.

Linear 방식:

$$\overline{\log E_n} = \log E_n + \frac{T\_Min - Min}{Max - Min} \times (Max - \log E_n) \quad (5)$$

Non-Linear 방식:

$$\overline{\log E_n} = \log E_n + \frac{T\_Min - Min}{\log(Max) - \log(Min)} \times (\log(Max) - \log(\log E_n)) \quad (6)$$

여기서 식 (5)의 Linear 방식은 그림 1의 실선 부분을, 식 (6)의 Non-Linear 방식은 그림 1의 점선 부분을 의미한다. 본 논문에서는 이 중에서 보다 우수한 성능을 나타내는 Non-Linear 방식의 ERN을 적용한다[3].

2.2 SEN(Silence Energy Normalization)

앞서 설명한 ERN은 작은 값을 가지는 로그 에너지의 특징을 미리 정한 변화 범위 내로 키워 주는 방법인 반면, SEN은 작은 로그 에너지 값을 찾아서 작은 상수의 값으로 줄여주는 방법으로 이 두 방법의 원리와 목적은 거의 유사하다.

SEN은 먼저 각 프레임에 대하여 묵음/음성 판별을 수행한 후, 음성으로 판별되는 프레임의 로그 에너지는 원래의 값을 그대로 유지하는 반면에 묵음으로 판별되는 프레임의 로그 에너지 값은 작은 상수 값으로 정규화한다. 상세한 정규화 과정은 다음과 같다.

묵음/음성 판별의 기준값을 얻기 위하여 먼저 로그 에너지 특징을 IIR HPF를 통과시킨다. 필터의 입력과 출력의 상관관계는 식(7)과 같으며

$$y_n = \frac{1}{2}(\log E_{n+1} - y_{n-1}), \quad (7)$$

여기서 $\log E_n$은 n번째 프레임의 원 로그 에너지의 값을 나타내며, y_n은 필터의 출력값을 나타낸다. 다음에 식 (7)에서 얻은 y_n의 평균값을 묵음/음성 판별의 기준값 T(식(8))로 설정하고, 식 (9)를 이용하여 출력값 y_n이 기준값 T보다 크면 음성 구간으로 판별하여 원 로그 에너지 값을 그대로 유지하고, 작으면 묵음으로 판별하여 작은 상수 ε으로 정규화 한다.

$$T = \frac{1}{N}\sum_{n=1}^{N} y_n \quad (8)$$

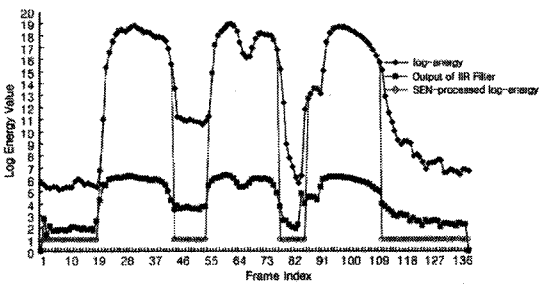


그림 2. SEN의 처리 과정

$$\overline{\log E_n} = \begin{cases} \log E_n & \text{if } y_n > T \\ \varepsilon & \text{if } y_n \le T \end{cases} \quad (9)$$

여기서 N은 프레임 수를 나타낸다. 본 논문에서 $\varepsilon = 1$로 설정하였다[5]. SEN의 처리 과정을 한 개의 깨끗한 음성 신호를 대상으로 실험한 결과를 그림 2에 나타내었다. 그림 2에서 음성 구간으로 판별된 큰 로그 에너지 값을 가진 프레임은 원 로그 에너지 값을 그대로 유지하는 반면, 묵음 구간으로 판별된 작은 로그 에너지 값을 가지는 프레임은 작은 상수 ε으로 정규화된 것을 알 수 있다.

3. 켑스트럼 특징의 정규화 방법

3.1 Cepstral Mean and Variance Normalization (CMVN)

켑스트럼 정규화는 음성신호로부터 추출한 켑스트럼 특징을 정규화하여 훈련 환경과 인식 환경의 차이를 감소시키는 방법이다. 대표적인 방법으로 켑스트럼 평균 정규화(CMN: Cepstral Mean Normalization)과 켑스트럼 평균 및 분산 정규화(CMVN: Cepstral Mean and Variance Normalization) 등이 있다.

CMN은 식 (10)과 같이 켑스트럼 특징의 각 차수에 대한 평균값을 구한 뒤 이를 차수별로 빼주는 방식이다. 여기서 켑스트럼 특징의 각 차수에 대한 평균값 μ는 식 (11)과 같이 구할 수 있다.

$$\overline{C_n}[d] = C_n[d] - \mu[d] \quad (10)$$

$$\mu[d] = \frac{1}{N}\sum_{n=1}^{N} C_n[d] \quad (11)$$

CMVN은 켑스트럼 특징의 분산을 감소시키기 위

해 제안된 방법으로 CMN으로부터 발전되었으며 유도 과정은 식 (12)와 같다. 즉, 전체 프레임에 대하여 각 차수별로 켑스트럼 특징의 분산값을 식 (13)과 같이 구한 뒤, 이를 켑스트럼 특징에 차수별로 나눠주어 평균값이 0, 분산값이 1이 되도록 한다.

$$\widehat{C}_n[d] = (\sigma^2[d])^{-1/2}\,\overline{C_n}[d] \tag{12}$$

$$\sigma^2[d] = \frac{1}{N}\sum_{n=1}^{N}(C_n[d]-\mu[d])^2 \tag{13}$$

여기서 $\sigma^2[d]$는 켑스트럼 특징에서 d번째 차수의 성분들에 대한 분산값을 나타낸다.

3.2 ARMA 필터

Auto-Regression Moving Average(ARMA) 필터는 일종의 LPF이다. 따라서 ARMA 필터를 켑스트럼 특징에 적용할 경우, 켑스트럼 특징에 포함된 고 주파수 성질을 갖는 잡음의 오염을 제거할 수 있으며 그리고 켑스트럼 특징의 분포를 스무딩하게 할 수 있다. Chen et al.,[10-12]는 ARMA 필터를 CMVN의 후처리로 적용하였으며, 이때 CMVN으로 정규화된 켑스트럼 특징에 대한 ARMA 필터의 입출력 관계는 식 (14)과 같다.

$$\widetilde{C}_n[d] = \begin{cases} \dfrac{\sum_{i=1}^{M}\widetilde{C_{n-i}}[d] + \sum_{j=0}^{M}\overline{C_{n+j}}[d]}{2M+1} & \text{if } M < n \le N-M \\ \overline{C_n}[d] & otherwise \end{cases}, \tag{14}$$

여기서 $\widetilde{C}$, $\overline{C}$ 및 M은 ARMA 필터의 입력 켑스트럼 특징의 값, 출력 켑스트럼 특징의 값 및 필터의 차수를 나타낸다.

ARMA 필터를 주파수 영역에서 고찰하기 위하여, ARMA 필터의 전달 함수를 식 (16)으로, 주파수 응답은 식 (17)으로 나타내었다.

$$(2M+1)\widetilde{C}_n[d] - \widetilde{C_{n-1}}[d] - \ldots - \widetilde{C_{n-M}}[d] = \overline{C_n}[d] + \ldots + \overline{C_{n+M}}[d] \tag{15}$$

$$H(z) = \frac{1+z+\ldots+z^M}{2M+1-z^{-1}-\ldots-z^{-M}} \tag{16}$$

$$H(e^{jw}) = \frac{1+e^{jw}+\ldots+e^{jMw}}{2M+1-e^{-jw}-\ldots-e^{-jMw}} = \frac{1-e^{j(M+1)w}}{2M+2-(2M+1)e^{jw}-e^{-jMw}} \tag{17}$$

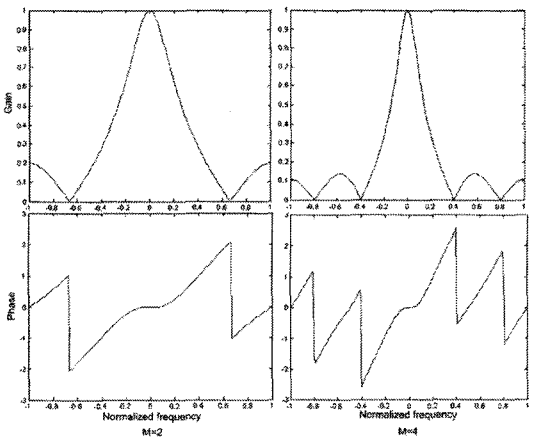


그림 3. ARMA 필터의 주파수 응답 특성

그림 3은 ARMA 필터의 주파수 응답을 나타낸다. 그림 3에서 좌측은 $M=2$인 경우의 ARMA 필터에 대한 각 주파수의 진폭과 위상 크기를, 우측은 $M=4$인 경우에 대한 것을 나타낸 것이다. 또한 그림 3의 주파수 진폭 응답 곡선으로부터 ARMA 필터는 LPF 특징을 나타냄을 관찰할 수 있으며, 주파수 $w=0$일 때 $H(e^{jw})=1$인 것을 알 수 있다.

그림 4(a)는 각 프레임에 대한 켑스트럼 특징 중에서 첫 번째 특징 $c[1]$에 대한 것을, 그림 4(b)는 CMVN을 수행한 후의 결과를, 그림 4(c)는 ARMA 필터를 적용했을 경우를 나타낸다. 그림에서 볼 수 있듯이 clean, 15dB 및 5dB인 경우, ARMA 필터를 적용한 후 켑스트럼 특징 $c[1]$의 포락선이 많이 스무딩 되었고 켑스트럼 특징들 간의 불일치가 많이 줄어든 것을 알 수 있다.

4. 제안하는 로그 에너지 정규화 방법

본 논문은 잡음 환경하의 음성 인식 성능을 향상하기 위하여 ARMA 필터를 기존의 ERN, SEN에 후처리로 적용하는 ERN-A 및 SEN-A, 그리고 기존의 켑스트럼 정규화 방법인 CMVN-A과 결합한 [ERN-CMVN]-A 및 [SEN-CMVN]-A를 제안한다. 이하 각 방법에 대하여 설명한다.

4.1 ERN-A

ERN은 로그 에너지 특징의 최대값과 최소값을 찾은 후, 미리 정한 D.R(dynamic range)에 의하여 상대

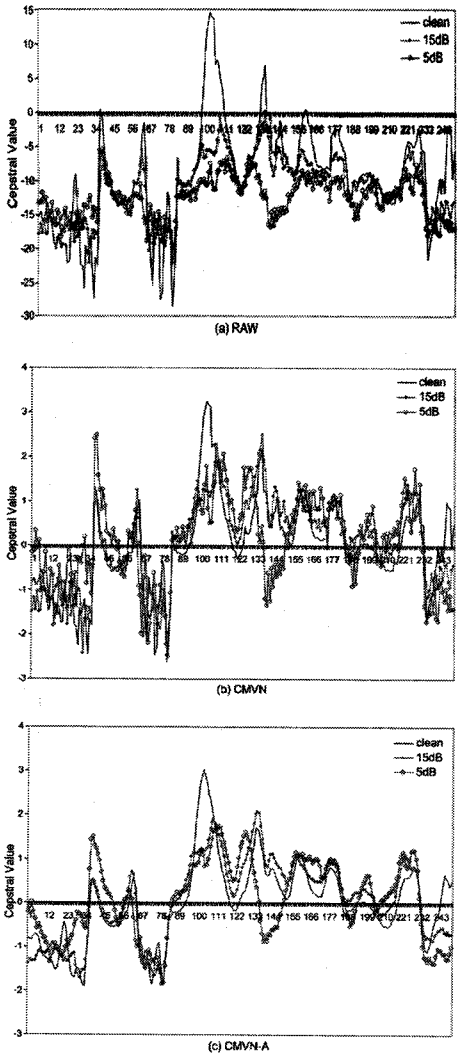


그림 4. 켑스트럼 특징에 CMVN, CMVN-A를 적용한 결과 비교

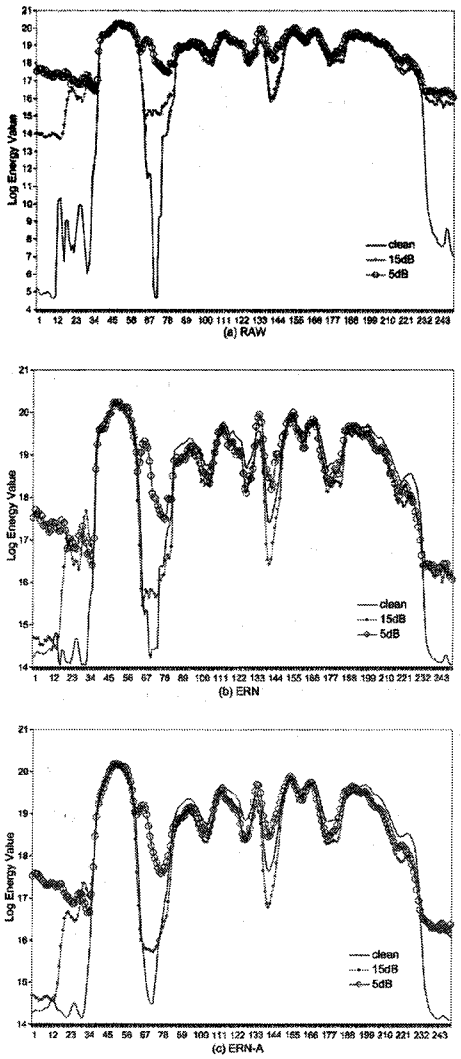


그림 5. 로그 에너지 특징에 ERN, ERN-A를 적용한 결과 비교($D.R$=14, M=2)

적으로 작은 로그 에너지를 더 많이 키워서 잡음에 의한 훈련과 인식의 환경적 불일치를 줄여주는 방법이다. 이는 잡음이 적게 포함된 음성 신호의 로그 에너지 특징의 전체적인 분포를 잡음이 많이 포함된 경우와 비슷하게 변환해 줄 수는 있으나, 그러나 각 로그 에너지 특징에 대한 잡음의 오염을 더욱 효과적으로 줄이기 위하여 본 논문에서는 자동 회귀(auto regression) 방식으로 이동 평균(moving average)값을 계산하여 스무딩하는 ARMA(Auto- Regression and Moving Average) 필터를 후처리로 적용하는 ERN-A를 제안한다.

SNR값이 서로 다른 잡음 환경에서 ARMA 필터의 스무딩 효과를 비교하기 위하여, 그림 5(a)에 어떤 처리도 하지 않은 경우, 그림 5(b)에 ERN으로 정규화한 경우, 그림 5(c)에 ERN에 ARMA 필터를 후처리로 적용한 경우의 로그 에너지 특징의 분포 그래프를 나타내었다. 비교한 결과, 그림 5(a)에서 clean, 15dB인 경우의 로그 에너지 특징이 각각 약 5~20, 13~20 사이에 분포하고 있으며, 그림 5(b)에서 ERN을 적용한 후 clean, 15dB인 경우의 로그 에너지 특징이 모두 약 14~20 사이에 분포됨을 알 수 있다. 또, 그림 5(c)에서 ARMA 필터를 적용한 후 로그 에너지

특징의 분포가 스무딩 되어서 전체적인 로그 에너지 특징의 분포가 더욱 유사하게 된 것을 확인 할 수 있다.

여기서 SNR값이 5dB인 경우를 살펴보면, 로그 에너지의 최소값 Min이 D.R(dynamic range)에 의하여 정해진 T_Min(Target Minimum)보다 크므로 ERN이 수행되지 못하는 문제가 발생한다. 따라서 이러한 경우에 ARMA 필터를 수행하므로 기존의 ERN의 문제점을 더욱 보완할 수 있다.

4.2 SEN-A

SEN은 로그 에너지 특징에서 상대적으로 작은 값을 가진 로그 에너지 특징만을 찾아서 작은 상수($\varepsilon=1$)로 정규화하는 방법으로 상대적으로 큰 값을 가지는 로그 에너지 특징은 변환시키지 않는다. 그러나 실제 잡음 환경에서 특히 SNR값이 작은 환경에서 상대적으로 큰 값을 가지는 로그 에너지 특징에서도 잡음의 오염이 많이 발생하고 있다.

따라서 잡음 환경하의 음성 인식 성능을 향상시키기 위하여 상대적으로 큰 값을 가지는 로그 에너지 특징에 대한 정규화처리도 필요하게 된다. 본 논문은 이러한 문제점을 고려하여 ARMA 필터를 후처리로 적용하는 SEN-A를 제안한다. 4.1 절에서 ERN-A와 유사하게 ARMA 필터를 적용하므로 SEN 처리 후의 로그 에너지 특징이 스무딩 되어 전체적인 로그 에너지 특징의 분포가 더욱 유사하게 된 것을 그림 6으로부터 확인 할 수 있다.

그림 6(a)는 어떤 처리도 하지 않은 경우, 그림 6(b)는 SEN으로 정규화한 경우, 그림 6(c)는 SEN에 ARMA 필터를 후처리로 적용한 경우의 로그 에너지 특징의 분포 그래프이다. 그림 6(b)에서 SEN에 의하여 묵음 구간으로 판별되는 프레임의 로그 에너지 특징은 작은 상수($\varepsilon=1$)로 정규화 되며, 그림 6(c)는 ARMA 필터의 스무딩 효과로부터 각 로그 에너지 특징의 분포 차이가 좀 더 줄여지었으며, 그리고 로그 에너지 특징의 전체적인 분포도 좀 더 자연스럽게 된 것을 확인할 수 있다.

4.3 [ERN-CMVN]-A 및 [SEN-CMVN]-A

로그 에너지와 켑스트럼 특징은 서로 독립적인 관계를 가지고 있으며, 일반적으로 이 두 특징에 대한

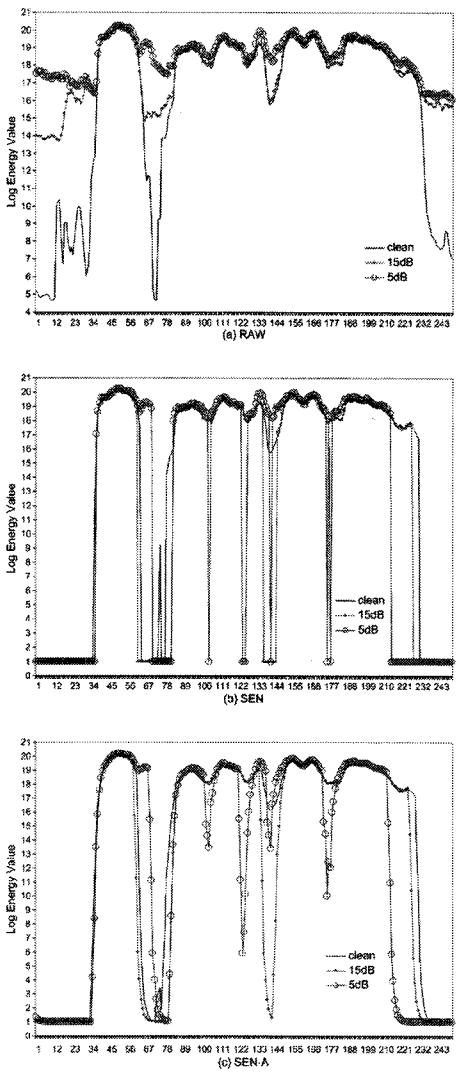


그림 6. 로그 에너지 특징에 SEN, SEN-A를 적용한 결과 비교(M=2)

정규화 방법을 동시에 적용하면 보다 향상된 인식 성능을 얻을 수 있다. 본 논문에서는 ERN-A 및 SEN-A를 각각 CMVN-A[10-12]와 결합한 [ERN-CMVN]-A 및 [SEN-CMVN]-A를 제안한다.

그림 7은 제안하는 방법의 처리 과정을 나타내는 블록 다이어그램이다. 먼저, 로그 에너지 특징에 대하여 ERN 또는 SEN으로, 켑스트럼 특징에 대하여 CMVN으로 정규화를 수행하고, 다음으로 ARMA 필터를 각 특징 파라미터에 후처리로 적용한다. 그러므로 ARMA 필터로부터 스무딩 효과를 에너지 특징과 켑스트럼 특징에 동시에 얻을 수 있어, 특징 파라

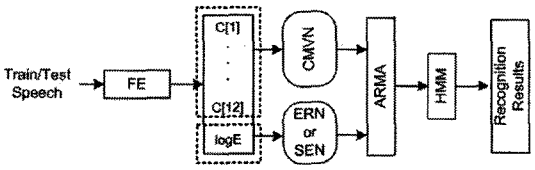


그림 7. (ERN-CMVN)-A 및 (SEN-CMVN)-A의 처리 과정

미터에 포함된 잡음의 오염을 더욱 줄일 수 있다.

5. 인식 실험 및 고찰

5.1 인식 실험 환경

인식 실험에는 Aurora 2.0 DB[15]를 사용한다. Aurora 2.0 DB에는 2 가지 훈련 환경 즉, 8440개의 clean 발성으로 구성된 clean-condition과 동일한 발성을 20개의 잡음 환경에 나누어 각 422개의 발성으로 구성된 multi-condition이 있다. 잡음 환경은 4 종류(subway, babble, car 및 exhibition)로 구성되어 있으며 이 잡음들은 5 단계의 SNR 레벨(clean, 20dB, 15dB, 10dB, 5dB)로 나누어져 있다.

인식 데이터는 3 가지의 subset으로 구성되어 있으며, 훈련에서 사용한 4 종류의 잡음을 포함한 Set A와 훈련에서 사용되지 않은 새로운 4 종류의 잡음을 포함한 Set B, 그리고 훈련과 다른 채널 특성을 가지고 Set A와 Set B에 나타난 2 종류 잡음(subway, street)을 포함한 Set C의 총 10 종류 잡음으로 -5dB에서 clean까지의 7 단계의 잡음 레벨로 구성되어 있다.

기본 인식기는 Aurora2-HTK를 사용한다. 단어 모델은 one, two, three, four, five, six, seven, eight, nine, zero, oh의 11개로 정의되고, 각 단어 모델은 3 혼합수(mixture), 16 상태(state)를 갖는 CHMM (Continuous Hidden Markov Models)으로 구성한다. 인식 시스템에는 11개의 단어 모델 외에 2개의 묵음 모델(silence model)이 포함되어 있는데, 각각 3 상태와 1 상태의 CHMM으로 구성한다. 특징 파라미터는 12차 MFCC와 1차 로그 에너지, 그리고 각각의 delta 및 delta-delta 계수를 포함한 총 39차로 구성한다. 분석 프레임의 크기는 25ms이며, 10ms씩 이동하면서 특징 파라미터를 추출한다.

인식 성능의 평가에는 단어 정확률(WA: word accuracy)을 사용하며 식 (18)과 같이 정의 된다.

$$WA = \frac{H-I}{N} \times 100 \tag{18}$$

$$H = N-D-S \tag{19}$$

여기서 H는 정확하게 인식된 단어의 수, I는 삽입된 단어의 수, D는 삭제된 단어의 수, S는 대치된 단어의 수, N은 전체 인식 단어의 수를 나타낸다.

인식에 사용되는 인식 모델은 clean 환경에서 작성하여 인식 실험에 이용하며, 성능 평가는 Set A, Set B 및 Set C 의 각 잡음의 종류에 대해서 clean 부터 0dB까지의 평균 WA를 구하여 인식률로 한다.

5.2 인식 실험 및 결과

5.2.1 ERN-A에 대한 인식 실험

4.1 절에서 설명한 바와 같이 ERN-A는 ERN과 ARMA 필터를 결합하는 방법으로서 ERN 의 성능은 $D.R$의 값에 의하여 결정되며, ARMA 필터의 성능은 스무딩 정도를 결정하는 필터의 차수 M에 의하여 결정된다. 따라서 최대 인식 성능을 발휘하게 할 수 있는 ERN의 $D.R$과 필터의 차수 M을 찾는 것이 중요하다. 이를 위하여 $D.R$를 10~14, M을 0, 2, 4로 변화시키면서 ERN-A에 대한 인식 실험을 수행 한 결과를 표 1에 나타내었다. 표 1에서 보는 바와 같이 $D.R$=12, M=2으로 설정하였을 때 가장 우수한 인식 성능을 나타냄을 알 수 있었다.

5.2.2 SEN-A에 대한 인식 실험

4.2 절에서 설명한 바와 같이 SEN-A는 SEN과 ARMA 필터를 결합한 방식이다. 여기서 SEN은 각 음성 신호의 로그 에너지 특징으로부터 그 음성 신호에 적합한 기준값을 찾아서 정규화하므로, 결합 방식의 인식 성능은 오직 ARMA 필터의 차수 M에 의하여 결정된다. 최적 ARMA 필터의 차수 M을 결정하기 위하여 M을 0, 2, 4로 변경하면서 인식 실험을 수행한 결과, M=2인 경우 가장 우수한 인식 성능을 나타냄을 알 수 있었다(표 2 참고).

표 1과 표 2에서 보는 바와 같이 기존의 ERN 및 SEN에 ARMA 필터를 후처리로 적용하였을 경우, 인식 성능이 향상된 것을 확인 할 수 있으며, 특히 SNR값이 낮아짐에 따라 더욱 큰 향상효과를 얻을 수 있음을 알 수 있다.

이는 ERN의 경우, SNR이 작을 때 4.1 절의 step

표 1. $D.R$의 값과 ARMA 필터의 차수 M의 값에 따른 ERN-A의 인식 성능 비교 (예: ERN12-A2는 $D.R=12$, $M=2$를 의미)

Set A	ERN10	ERN11	ERN12	ERN13	ERN14	ERN10 -A2	ERN11 -A2	ERN12 -A2	ERN13 -A2	ERN14 -A2	ERN10 -A4	ERN11 -A4	ERN12 -A4	ERN13 -A4	ERN14 -A4
clean	98.76	98.69	98.69	98.78	98.83	98.68	98.67	98.72	98.78	98.82	98.68	98.67	98.71	98.70	98.74
20dB	90.67	95.59	96.39	96.96	97.09	89.47	95.72	96.85	97.17	97.24	88.81	95.58	96.96	97.17	97.23
15dB	80.30	90.92	92.78	93.66	93.94	78.16	91.28	93.47	93.87	94.49	77.00	90.61	93.50	94.01	94.24
10dB	58.34	77.18	83.33	84.48	84.19	56.46	77.54	84.61	85.33	85.35	55.78	76.70	84.73	85.65	85.54
5dB	33.89	53.10	63.15	62.03	58.71	32.75	53.49	65.39	63.22	61.07	32.54	51.76	64.63	63.81	60.47
0dB	20.07	27.72	31.52	28.28	25.55	19.85	28.67	34.06	28.75	27.53	19.23	27.70	33.82	31.10	27.20
Avg	63.67	73.87	77.64	77.36	76.39	62.56	74.23	**78.85**	77.85	77.42	62.01	73.50	78.73	78.41	77.24

Set B	ERN10	ERN11	ERN12	ERN13	ERN14	ERN10 -A2	ERN11 -A2	ERN12 -A2	ERN13 -A2	ERN14 -A2	ERN10 -A4	ERN11 -A4	ERN12 -A4	ERN13 -A4	ERN14 -A4
clean	98.76	98.69	98.69	98.78	98.83	98.68	98.67	98.72	98.78	98.82	98.68	98.67	98.71	98.70	98.74
20dB	90.08	94.75	94.79	95.75	96.63	89.69	94.91	95.83	96.15	96.80	88.84	94.96	95.95	96.06	96.61
15dB	80.58	90.02	90.64	91.97	93.30	79.32	90.27	92.19	92.75	93.97	78.36	90.22	92.43	92.57	93.35
10dB	62.67	77.87	80.36	82.24	82.59	61.36	78.20	82.41	83.10	83.72	60.64	78.10	82.60	82.89	83.02
5dB	40.37	54.61	58.50	57.73	54.44	39.74	55.48	62.42	60.26	57.46	39.24	55.34	61.41	59.38	55.12
0dB	23.34	28.67	25.74	22.16	20.02	23.24	29.04	29.76	24.83	21.77	22.47	28.99	28.19	23.07	19.25
Avg	65.96	74.10	74.79	74.77	74.30	65.34	74.43	**76.89**	75.98	75.42	64.70	74.38	76.55	75.44	74.35

Set C	ERN10	ERN11	ERN12	ERN13	ERN14	ERN10 -A2	ERN11 -A2	ERN12 -A2	ERN13 -A2	ERN14 -A2	ERN10 -A4	ERN11 -A4	ERN12 -A4	ERN13 -A4	ERN14 -A4
clean	98.60	98.48	98.68	98.86	98.80	98.51	98.59	98.80	98.68	98.83	98.53	98.62	98.77	98.77	98.78
20dB	91.35	93.95	95.37	95.47	95.14	90.43	93.91	95.72	95.64	95.63	89.59	93.31	95.38	95.67	95.52
15dB	77.57	85.13	88.27	88.51	87.62	75.30	84.63	89.30	89.34	88.80	74.29	84.00	88.93	89.45	89.21
10dB	51.40	64.74	72.35	72.72	71.45	48.79	64.26	73.61	74.32	73.80	48.14	62.30	73.32	74.73	74.46
5dB	28.21	39.18	48.52	49.10	48.75	27.39	38.58	49.14	50.33	50.49	27.29	36.86	50.01	52.34	52.90
0dB	18.15	23.00	26.97	25.95	24.45	18.29	23.09	26.01	24.76	25.01	18.43	22.48	27.58	27.31	27.08
Avg	60.88	67.41	71.69	71.77	71.03	59.78	67.17	**72.10**	72.18	72.09	59.38	66.26	72.33	73.04	72.99

표 2. ARMA 필터의 차수 M의 값에 따른 SEN-A의 인식 성능 비교 (예: SEN-A2는 $M=2$를 의미)

Method / SNR	Set A			Set B			Set C		
	SEN	SEN-A2	SEN-A4	wSEN	SEN-A2	SEN-A4	SEN	SEN-A2	SEN-A4
clean	98.70	98.73	98.56	98.70	98.73	98.56	98.74	98.74	98.51
20dB	96.73	96.60	96.57	96.81	96.91	96.64	94.00	94.49	93.71
15dB	93.26	93.75	92.91	94.06	94.52	93.75	86.75	87.83	84.90
10dB	83.65	85.39	83.30	85.46	86.37	84.36	69.44	71.55	65.82
5dB	61.17	64.70	60.37	64.07	67.10	63.71	43.55	45.30	41.01
0dB	29.01	32.77	28.17	31.41	34.38	30.22	20.51	22.43	21.03
Avg	77.09	**78.66**	76.65	78.42	**79.67**	77.87	68.83	**70.05**	67.50

3에서 나타낸 바와 같이 로그 에너지의 최소값 Min이 D.R(dynamic range)에 의해 정해진 T_Min (Target Minimum) 보다 크게 되어 ERN이 수행되지 못한 결과이기 때문이며, 이를 극복하기 위하여 ARMA 필터를 후처리 적용하므로서 로그 에너지 특징에 포함된 변화가 심한 잡음의 영향을 줄일 수 있어 향상된 인식률을 나타낸 것으로 생각된다.

SEN 방법의 경우, SNR이 작을 때는 잡음의 영향

으로 묵음/음성을 판별에 대한 기준값의 정확도가 떨어지며 이로 인해 음성으로 잘 못 판별된 구간의 로그 에너지 특징은 변화가 심한 잡음의 영향을 그대로 나타내기 때문에 낮은 인식률로 이어질 수 있게 되는 데, 이 경우 ARMA 필터를 적용하면 이와 같이 변화가 큰 잡음을 스무딩하므로서 인식률 저하를 방지한 결과로 분석된다.

ARMA 필터의 인식 성능은 필터의 차수 M에 의하여 차이를 보인다. 표 1과 표 2로부터 $M=2$일 경우 인식 성능이 가장 우수함을 알 수 있었다. 여기서 ARMA 필터의 차수 M과 특징 파라미터에 포함한 잡음 오염의 제거간의 trade-off 관계가 존재하는 것을 알 수 있었는데 이는 ARMA 필터의 차수 M을 증가시킴으로서 스무딩 효과를 더욱 크게 할 수 있으나 그 결과 특징 파라미터에 포함된 음성 신호의 특징이 손실되기 때문으로 생각된다. 따라서 적절한 ARMA 필터의 차수 M을 찾아서 사용하는 것이 매우 중요하다.

5.2.3 [ERN-CMVN]-A 및 [SEN-CMVN]-A에 대한 인식 실험

본 절에서는 4.3 절에서 설명한 ERN-A 및 SEN-A를 CMVN-A와 결합한 경우에 대한 인식 실험을 수행하였다. 표 3에 ERN에 관한 인식 실험 결과를, 표 4에 SEN에 관한 인식 실험 결과를 나타내었다. 이 때 ERN, SEN 및 ARMA 필터의 실험 조건은 선행 실험에서 가장 우수한 인식 성능을 나타내었던 $D.R=12$, $M=2$로 설정하였다.

표 3과 표 4로부터 알 수 있는 바와 같이 ARMA 필터를 로그 에너지 정규화 방법에 후처리로 적용하였을 경우 인식 성능의 향상은 전반적으로 켑스트럼 특징 파라미터의 경우보다 미미함을 확인할 수 있다. 이는 ARMA 필터가 갖는 low-pass filter 특성이 켑스트럼 특징에 포함된 고주파 성질을 갖는 잡음 성분

표 3. ERN에 관한 인식 성능 비교

Set A	Baseline	CMVN	CMVN-A	ERN-CMVN	ERN	ERN-A	[ERN-CMVN]-A
clean	99.02	98.98	98.95	98.82	98.69	98.72	98.76
20dB	95.25	95.98	96.73	97.10	96.39	96.85	97.01
15dB	87.33	91.65	93.25	94.57	92.78	93.47	94.33
10dB	67.71	80.43	84.47	86.88	83.33	84.61	88.36
5dB	39.48	57.47	66.47	65.50	63.15	65.39	74.74
0dB	16.95	26.62	38.03	30.89	31.52	34.06	44.85
Avg	67.62	75.19	79.65	78.96	77.64	78.85	83.01

Set B	Baseline	CMVN	CMVN-A	ERN-CMVN	ERN	ERN-A	[ERN-CMVN]-A
clean	99.02	98.98	98.95	98.82	98.69	98.72	98.76
20dB	92.77	96.37	96.89	97.14	94.79	95.83	96.37
15dB	81.34	92.07	93.74	94.78	90.64	92.19	93.94
10dB	59.01	81.90	85.11	87.12	80.36	82.41	87.69
5dB	31.93	58.64	66.67	65.12	58.50	62.42	73.30
0dB	13.70	26.72	37.66	32.36	25.74	29.76	44.30
Avg	62.96	75.78	79.84	79.22	74.79	76.89	82.39

Set C	Baseline	CMVN	CMVN-A	ERN-CMVN	ERN	ERN-A	[ERN-CMVN]-A
clean	99.06	99.10	99.01	98.89	98.68	98.80	98.78
20dB	94.30	95.45	96.12	96.68	95.37	95.72	96.79
15dB	87.84	88.69	91.57	93.60	88.27	89.30	94.42
10dB	74.15	74.23	80.01	85.95	72.35	73.61	88.34
5dB	50.24	51.14	60.44	67.56	48.52	49.14	74.62
0dB	24.17	24.23	34.30	35.71	26.97	26.01	48.06
Avg	71.62	72.14	76.91	79.73	71.69	72.10	83.50

표 4. SEN에 관한 인식 성능 비교

Set A	Baseline	CMVN	CMVN-A	SEN-CMVN	SEN	SEN-A	[SEN-CMVN]-A
clean	99.02	98.98	98.95	98.79	98.70	98.73	98.76
20dB	95.25	95.98	96.73	96.25	96.73	96.60	96.76
15dB	87.33	91.65	93.25	93.13	93.26	93.75	94.31
10dB	67.71	80.43	84.47	86.45	83.65	85.39	88.54
5dB	39.48	57.47	66.47	71.42	61.17	64.70	74.54
0dB	16.95	26.62	38.03	42.43	29.01	32.77	46.06
Avg	67.62	75.19	79.65	81.41	77.09	78.66	83.16

Set B	Baseline	CMVN	CMVN-A	SEN-CMVN	SEN	SEN-A	[SEN-CMVN]-A
clean	99.02	98.98	98.95	98.79	98.70	98.73	98.76
20dB	92.77	96.37	96.89	96.67	96.81	96.91	96.93
15dB	81.34	92.07	93.74	93.99	94.06	94.52	94.92
10dB	59.01	81.90	85.11	87.08	85.46	86.37	88.77
5dB	31.93	58.64	66.67	71.30	64.07	67.10	74.03
0dB	13.70	26.72	37.66	42.32	31.41	34.38	46.56
Avg	62.96	75.78	79.84	81.69	78.42	79.67	83.33

Set C	Baseline	CMVN	CMVN-A	SEN-CMVN	SEN	SEN-A	[SEN-CMVN]-A
clean	99.06	99.10	99.01	98.83	98.74	98.74	98.80
20dB	94.30	95.45	96.12	95.31	94.00	94.49	96.07
15dB	87.84	88.69	91.57	91.05	86.75	87.83	92.67
10dB	74.15	74.23	80.01	81.51	69.44	71.55	83.48
5dB	50.24	51.14	60.44	63.03	43.55	45.30	63.98
0dB	24.17	24.23	34.30	35.46	20.51	22.43	35.99
Avg	71.62	72.14	76.91	77.53	68.83	70.05	78.50

을 함께 제거하기 때문으로 분석된다.

그러나 표 3으로부터 알 수 있는 바와 같이 ARMA 필터를 기존의 로그 에너지 정규화 방법과 켑스트럼 정규화 방법에 동시에 적용하였을 경우, [ERN-CMVN]-A는 기존의 CMVN-A, ERN-CMVN에 비해 각각 Set A에서 4.0%, 4.1%, Set B에서 2.6%, 3.2%, Set C에서 6.6%, 3.8%의 향상된 인식 결과를 나타냄을 알 수 있으며, 표 4로부터 [SEN-CMVN]-A는 기존의 CMVN-A, SEN-CMVN에 비해 각각 Set A에서 3.5%, 1.8%, Set B에서 3.5%, 1.6%, Set C에서 1.6%, 1.0%의 향상된 인식 결과를 나타냄을 알 수 있다.

이는 ARMA 필터를 기존의 로그 에너지 정규화 방법과 켑스트럼 정규화 방법에 동시에 적용하여, 서로 독립적인 특징 파라미터에 적절한 스무딩 처리를 수행하므로 잡음의 영향을 더욱 줄일 수 있기 때문으로 분석된다.

6. 결 론

본 논문에서는 잡음 환경 하에서의 음성 인식 성능 저하의 주요 원인이 되고 있는 훈련과 인식의 환경적 차이를 줄이기 위한 로그 에너지 특징의 스무딩 방법으로 ARMA 필터를 후처리로 적용하는 정규화 방법을 제안하였다. Aurora 2.0 DB를 이용한 인식 실험 결과, 제안하는 [ERN-CMVN]-A 방법은 기존의 CMVN-A, ERN-CMVN 방법에 비해 각각 Set A에서 4.0%, 4.1%, Set B에서 2.6%, 3.2%, Set C에서 6.6%, 3.8%의 향상된 인식 결과를, [SEN-CMVN]-A 방법은 기존의 CMVN-A, SEN-CMVN 방법에 비해 각각 Set A에서 3.5%, 1.8%, Set B에서 3.5%, 1.6%, Set C에서 1.6%, 1.0%의 향상된 인식 결과를 얻을 수 있었다. 이는 ARMA 필터를 기존의 로그 에너지 정규화 방법과 켑스트럼 정규화 방법에 동시에 적용하여, 서로 독립적인 특징 파라미터에 적절한

스무딩을 수행함이 음성 인식에서의 인식과 훈련의 환경적 차이를 효과적으로 줄일 수 있는 것으로 분석된다.

참 고 문 헌

[1] K. S. Yao, E. Visser, O. W. Kwon, and T. W. Lee, "A Speech Processing Front-End with Eigenspace Normalization for Robust Speech Recognition in Noisy Automobile Environments," Proc. Eurospeech, pp. 9-12, Sep. 2003.

[2] T. H. Hwang, and S. C. Chang, "Energy Contour Enhancement for Noisy Speech Recognition," International Symposium on Chinese Spoken Language Processing, pp. 249-252, 2004.

[3] W. Z. Zhu, and D. O. Shaughnessy, "Log energy Dynamic Range Normalization for Robust for Robust Speech Recognition," Proc. ICASSP, Vol.1, pp. 245-248, 2005.

[4] J. D. Veth, L. Manuuary, B. Noe, F. D. Wet, J. Sienel, L. Boves, and D. Jovet, "Feature Vector Selection to Improve ASR Robustness in Noisy Conditions," Proc. Eurospeech, pp. 201-204, 2001.

[5] C. F. Tai, and J. W. Hung, "Silence Energy Normalization for Robust Speech Recognition in Additive Noise Environments," Proc. ICSLP, pp. 2558-2561, 2006.

[6] A. Sankar, and C. H. Lee, "A Maximum Likelihood Approach to Stochastic Matching for Robust Speech Recognition," *IEEE Trans. on Speech and Audio Processing*, Vol.4, No.3, pp. 190-202, May 1996.

[7] P. Moreno, B. Raj, E. Gouvea, and R. Stern, "Multivariate Gaussian Based Cepstral Normalization for Robust Speech Recognition," Proc. ICASSP, pp. 137-140, May 1995.

[8] S. Furui, "Cepstral Analysis Technique for Automatic Speaker Verification," *IEEE Trans. On Acoustics, Speech, and Signal Processing*, Vol.ASSP-29, No.2, pp. 254-272, 1981.

[9] O. Viikki, and K. Laurila, "Cepstral Domain Segmental Feature Vector Normalization for Noise Robust Speech Recognition," *Speech Communication*, Vol.25, pp. 133-147, 1998.

[10] C. P. Chen, J. Bilmes, and K. Kirchhoff, "Low-Resource Noise-Robust Feature Post-Processing on Aurora 2.0," Proc. ICSLP, pp.2445-2448, Sep. 2002.

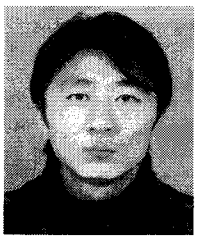
[11] C. P. Chen, K. Filali, and J. Bilmes, "Frontend Post-Processing and Backend Model Enhancement on the Aurora 2.0/3.0 Databases," Proc. ICSLP, pp.241-244, Sep. 2002.

[12] C. P. Chen, and J. Bilmes, "MVA Processing of Speech Features," *IEEE Trans. on Audio, Speech and Language Processing*, Vol.15, No.1, pp.257-270, 2007.

[13] S. M. Ahidi, H. Sheikhzadeh, R. L. Brennan, and G. H. Freeman, "An Energy Normalization Scheme for Improved Robustness in Speech Recognition," Proc. ICSLP, Vol.3, pp. 1649-1652, 2004.

[14] S. Young, D. Kershaw, J. Odell, D. Ollason, V. Valtchev, and P. Woodland, The HTK Book version 3.0, 2000.

[15] H. G. Hirsch, and D. Pearce, "The AURORA Experimental Framework for The Performance Evaluation of Speech Recognition Systems Under Noisy Conditions," ISCA ITRW ASR, France, Sep. 2000.



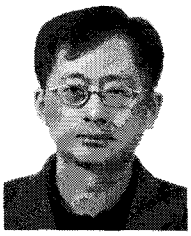
신 광 호

1998년 9월~2002년 8월 중국 연변대학교 사범대학 수학교육학과 이학사

2003년 9월~2005년 8월 영남대학교 대학원 정보통신공학과 공학석사

2005년 9월~현재 영남대학교 대학원 정보통신공학과 박사과정

관심분야 : 잡음처리, 잡음음성인식, 디지털 신호처리



정 호 열

1988년 2월 아주대학교 전자공학과 공학사
1990년 2월 아주대학교 전자공학과 공학석사
1998년 (프)리옹국립응용과학원 (INSA de Lyon) 전자 공학전공 공학박사
1998년 4월~1998년 12월 (프)CREATIS 박사후 과정
1999년 3월~현재 영남대학교 전자정보공학부 교수
관심분야 : 음성처리, 영상처리, 디지털 워터마킹

정 현 열

1975년 영남대학교 전자공학과 공학사
1989년 일본 동북대학교 정보공학과 공학박사
1989년 3월~현재 영남대학교 전자정보공학부 교수
1992년 7월~1993년 7월 미국 CMU Robotics 연구소 객원연구원
1994년 12월~1995년 2월 일본 토요하시기술과학대학 외국인 연구자
2000년 6월~2000년 8월 미국 Qualcomm Inc. 수석 엔지니어
관심분야 : 음성인식, 화자인식, 음성합성 및 DSP 응용분야