The Journal of The Institute of Internet, Broadcasting and Communication (IIBC)
Vol. 17, No. 1, pp.199-206, Feb. 28, 2017. pISSN 2289-0238, eISSN 2289-0246

https://doi.org/10.7236/JIIBC.2017.17.1.199
JIIBC 2017-1-26

잡음 환경에서 음성인식을 위한 스펙트럼 기울기의 효과적인 보상 방법

Efficient Compensation of Spectral Tilt for Speech Recognition in Noisy Environment

조정호*

Jungho Cho*

요 약 환경 잡음은 음성인식 시스템의 성능을 떨어뜨릴 수 있다. 이 논문은 인식 시스템이 잡음에 강인하도록 만들기 위하여, 켑스트럼에 기초한 특징 보상을 수행하는 과정을 제시한다. 이 방법은 부가적인 잡음의 영향을 제거하기 위한 직접적인 스펙트럼 기울기 보상에 기초를 둔다. 잡음 보상 방법은 로그 전력 스펙트럼의 스펙트럼 기울기 계산에 의하여 캡스트럼 영역에서 동작한다. 스펙트럼 보상은 SNR에 의존하는 켑스트럼 평균 보상 방법과 함께 사용된다. 백색 가우스 잡음, 지하철 잡음 및 자동차 잡음에 있는 조건에서, 실험 결과는 제안한 보상 방법이 여러 SNR에서 인식률을 상당히 개선한다는 것을 보여준다.

Abstract Environmental noise can degrade the performance of speech recognition system. This paper presents a procedure for performing cepstrum based feature compensation to make recognition system robust to noise. The approach is based on direct compensation of spectral tilt to remove effects of additive noise. The noise compensation scheme operates in the cepstral domain by means of calculating spectral tilt of the log power spectrum. Spectral compensation is applied in combination with SNR-dependent cepstral mean compensation. Experimental results, in the presence of white Gaussian noise, subway noise and car noise, show that the proposed compensation method achieves substantial improvements in recognition accuracy at various SNR's.

Key Words : speech recognition, spectral tilt, cepstral mean, spectral compensation

Ⅰ. 서 론

음성인식 시스템의 성능을 떨어뜨리는 음향학적 왜곡(acoustical distortion)은 여러 가지가 있다. 그 중 가장 중요한 원인은 부가적인 잡음이다. 음성인식 시스템의 훈련과정이나 인식과정에서 잡음환경이 서로 같으면 음성 인식기는 가장 좋은 인식 성능을 가진다. 이동 중인 자동차나 잡음이 있는 사무실, 지하철 등의 다양한 잡음이 존재하는 환경에서 신호 대 잡음비(SNR: signal to noise ratio)가 낮아질수록 음성인식 시스템의 성능은 급격히 떨어진다.

잡음 음성인식은 잡음으로 인해 저하된 음성인식 시스템의 성능을 향상시키는 기술로서, 다양한 방법으로 잡음의 영향을 최소화하는 연구가 발표되고 있다[1-4]. 잡

*정회원, 동서울대학교 디지털전자과
접수일자: 2016년 10월 18일, 수정완료: 2016년 12월 18일
게재확정일자: 2017년 2월 3일

Received: 18 October, 2016 / Revised: 18 December, 2016
Accepted: 3 February, 2017
*Corresponding Author: jhcho@du.ac.kr
Dept. of Digital Electronics, Dongseoul University, Korea

음이 섞인 음성에서 잡음을 필터링하거나 깨끗한 음성의 파라미터를 추정하는 방법에는 Wiener 필터링[5], Kalman 필터링[6], 스펙트럼 차감(spectral subtraction)[7] 및 켑스트럼 평균 차감(CMS: cepstral mean subtraction)[8] 등의 방법이 있다. 특히 스펙트럼 차감은 배경 잡음의 차감에 효과적이며 음성인식, 음성 향상(speech enhancement) 및 화자 확인(speaker verification)에 널리 이용된다.

부가적인 잡음이 있는 음성신호는 깨끗한 음성신호의 음성신호에 비해, 스펙트럼 기울기(spectral tilt)와 켑스트럼 평균(cepstral mean)이 다르다. 본 논문은 잡음이 있는 음성인식 환경에서 스펙트럼 기울기의 효과적인 보상 방법을 제시한다. 이 방법은 잡음에 왜곡된 음성신호에 대해 켑스트럼에 기초한 특징 벡터 보상을 수행하는 것으로, 스펙트럼 기울기 보상(spectral tilt compensation)과 켑스트럼 평균 보상(cepstral mean compensation)으로 이루어져 있다. 스펙트럼 기울기 보상은 켑스트럼의 계수를 직접적으로 수정하여 스펙트럼의 기울기를 변화시킨다. 스펙트럼의 기울기는 음성 신호의 기울기와 주변 잡음의 기울기의 차이에 따라 결정된다. 켑스트럼 평균 보상은 음성 신호의 평균 스펙트럼을 주변 잡음의 평균 스펙트럼에 맞추는 과정이다. 제안한 방법은 잡음과 음성의 평균 로그 에너지에 따라 스펙트럼 기울기와 켑스트럼 평균 보상의 강도를 다르게 하므로, 효과적인 스펙트럼 보상이 가능하다. 제안한 방법의 타당성을 확인하기 위하여 백색 가우스 잡음(white Gaussian noise), 지하철 잡음 및 자동차 잡음에 대해 음성인식 실험을 수행하고 그 결과를 보인다. 본 논문의 구성은 다음과 같다: 2장에서 음성신호의 자기회귀 모델과 스펙트럼 기울기에 대해 소개하고, 3장에서는 스펙트럼 기울기 보상과 4장에서는 켑스트럼 평균 보상에 대해 각각 설명하고, 5장에서 실험 및 결과를 요약하여 6장에서 결론을 맺는다.

II. 자기회귀 모델과 스펙트럼 기울기

음성신호 $s(n)$에 대한 자기회귀 선형예측(autoregressive linear prediction) 모델은 다음의 차분방정식으로 표현된다[9].

$$s(n)=\sum_{i=1}^{P} a_i s(n-i)+e(n) \tag{1}$$

여기서 $e(n)$과 a_i는 각각 예측오차와 예측계수(predictor coefficient)를 나타낸다. 전극 전달함수(all-pole transfer function)는

$$H(z)=\frac{S(z)}{E(z)}=\frac{1}{A(z)}=\frac{1}{1-\sum_{i=1}^{P} a_i z^{-i}} \tag{2}$$

이다. 여기서 $S(z)$와 $E(z)$는 각각 $s(n)$과 $e(n)$의 z-변환이다. 예측계수 a_i는 짧은 구간의 프레임에서 계산된다. 이 구간에서 성도 구조는 안정된 상태(stationary)로 가정한다. LP 켑스트럼 계수(linear prediction cepstral coefficient) $c(k)$ $(k \geq 1)$은 다음과 같이 예측계수 a_i에서 직접 구할 수 있다.

$$c(k)=a_k+\sum_{i=1}^{k-1}\left(\frac{i}{k}\right) c(i)\, a_{k-i} \tag{3}$$

선형 스펙트럼 기울기(linear spectral tilt)는 신호의 전력 분포(power distribution)와 주파수의 관계에 대한 대략적인 척도이다. 전달함수 $H(z)$에 대하여 주파수 범위 $0 \leq \omega \leq \pi$ $rads/\sec$ 에서 로그 전력 스펙트럼(log power spectrum) $S(\omega)=\ln|H(e^{j\omega})|^2$ 에 가장 잘 맞는 직선을

$$y(\omega)=m_{opt}\omega+b_{opt} \tag{4}$$

로 표현할 때, 스펙트럼 기울기는 이 직선의 기울기 m_{opt}를 의미한다. 여기서 m_{opt}와 b_{opt}는 상수이다. 최소 자승법(least square method)의 오차 E는

$$E=\int_0^{\pi}\{m\omega+b-\ln|H(e^{jw})^2|\}^2 d\omega \tag{5}$$

이다. 계수 $b=b_{opt}$이고 $m=m_{opt}$이면 E는 최소가 된다. 로그 전력 스펙트럼 $S(\omega)$은 켑스트럼 계수를 곱한 코사인 함수의 합으로 표현할 수 있다[10].

$$S(\omega)=\ln|H(e^{j\omega})|^2=2\sum_{k=1}^{\infty} c_k\cos(k\omega) \tag{6}$$

스펙트럼 기울기(spectral tilt) m을 구하면

$$m = Tilt\{S(\omega)\} \quad (7)$$
$$= \sum_{k=1}^{\infty} Tilt\{2c_k \cos(k\omega)\}$$
$$= -\frac{48}{\pi^3} \sum_{k=1,3,\cdots}^{\infty} c_k \frac{c_k}{k^2}$$

이다[11]. 여기서 $Tilt\{S(\omega)\}$는 스펙트럼 기울기를 나타낸다. 주파수 범위 $0 \le \omega \le \pi$ $rads/s$ 에서 켑스트럼 계수의 짝수 항은 주기의 정수배를 가지는 코사인 항이므로 스펙트럼 기울기에 포함되지 않는다.

Ⅲ. 스펙트럼 기울기 보상

잡음이 섞인 음성신호 $y(n)$은 그림 1에 보인 바와 같이, 깨끗한 음성신호 $s(n)$과 환경 잡음 $w(n)$의 부가적인 합으로 표현된다.

$$y(n) = s(n) + w(n) \quad (8)$$

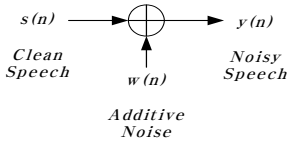

그림 1. 부가 잡음 모델
Fig. 1. Additive noise model

그림 2는 성인 남성이 발음한 모음'아'의 특정한 프레임에서 깨끗한 켑스트럼의 스펙트럼(실선)과 백색 가우스 잡음이 섞인 경우의 스펙트럼(점선)을 비교한 것이다. 잡음이 부가된 켑스트럼의 스펙트럼은 포먼트 주파수(formant frequency)의 대역폭이 상대적으로 넓어지고 스펙트럼 기울기가 변화된다.

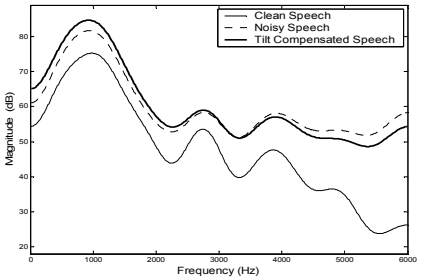

그림 2. 모음 '아' 의 깨끗한 켑스트럼과 잡음이 섞인 켑스트럼의 스펙트럼 예시
Fig. 2. Illustration of spectra of clean cepstrum and noisy cepstrum of a vowel /ɑ/

환경 잡음에 의해 왜곡된 스펙트럼의 기울기는 다음과 같이 보상할 수 있다. 잡음이 섞인 음성신호 $y(n)$에 대한 전극모델을 $H_y(z)$로 표기하면, 이의 로그 전력 스펙트럼

$$S_y(\omega) = \ln|H_y(e^{j\omega})|^2 \quad (9)$$

이다. 여기에 주파수 영역에서 기울기가 a인 1차 함수를 추가하여 스펙트럼 기울기를 보상한다. 즉 기울기가 보상된 로그 전력 스펙트럼 $S_T(\omega)$를 구하면

$$S_T(\omega) = \ln|H_y(e^{j\omega})|^2 - a|\omega| \quad (10)$$

이다. 이산시간 푸리에 변환(Discrete time Fourier transform)의 관계에서

$$\Im^{-1}(a|\omega|) = \frac{a}{\pi k^2}|(-1)^k - 1| \quad (11)$$

이므로, $S_T(\omega)$의 역 푸리에 변환을 구하면

$$c_T(k) = c_y(k) - \frac{a}{\pi k^2}[(-1)^k - 1] \quad (12)$$

이다. 여기서 $c_T(k)$는 스펙트럼 기울기가 보상된 켑스트럼를, $c_y(k)$는 잡음이 섞인 음성신호 $y(n)$의 켑스트럼 계수를 각각 나타내고, 주파수의 기울기 a는 $c_T(k)$의 스펙트럼 기울기를 조정하는 상수이다. 로그 전력 스펙트럼 $S_T(\omega)$의 스펙트럼 기울기 m_T을 구해보면

$$m_T = 2\sum_{k=1}^{\infty} c_k\, Tilt\{\cos(k\omega)\} - a \quad (13)$$
$$= \frac{-48}{\pi^3} \sum_{k=1,3,\cdots}^{\infty} c_k \frac{c_k}{k^2} - a$$

이다. 주파수의 기울기 a는 $c_T(k)$의 스펙트럼 기울기에 직접적인 영향을 주는 항임을 알 수 있다.

길이가 L 프레임인 음성신호 $y(n)$의 켑스트럼 벡터열(cepstrum vector sequence) $\{C = C_1, C_2, \cdots, C_t, \cdots, C_L\}$에서, 음성 프레임 t의 k번째 켑스트럼 $c_y(t,k)$은 다음과 같은 방법으로 스펙트럼 기울기가 조정된 켑스트럼 $c_T(t,k)$로 변환할 수 있다.

$$c_T(t,k)=\begin{cases} c_y(t,k)-\dfrac{T_c}{k^2} & \text{for} \quad k=odd \\ c_y(t,k) & \text{for} \quad k=even \end{cases} \tag{14}$$

기울기 T_c는

$$T_C = w_T(y_{tilt}-n_{tilt})\frac{n_{mean}}{y_{mean}} \tag{15}$$

로 구한다. 여기서 w_T는 신호 대 잡음비(SNR)에 의존되는 하중계수이다. y_{tilt}와 y_{mean}는 음성신호 $y(n)$의 평균 스펙트럼 기울기와 평균 로그 에너지이고, n_{tilt}와 n_{mean}는 무음 구간에서 채취한 잡음 $w(n)$의 평균 스펙트럼 기울기와 평균 로그 에너지이다. 그림 2에 보인 바와 같이, 스펙트럼 기울기가 보상된 켑스트럼 $c_T(t,k)$의 스펙트럼(진한 실선)은 깨끗한 음성(실선)에 유사한 기울기를 가진다.

Ⅳ. 켑스트럼 평균 보상

잡음 환경과 인식 환경의 차이에서 비롯되는 음성 스펙트럼의 평균 스펙트럼은 인식 성능에 큰 영향을 준다. 평균 스펙트럼의 차이는 다음과 같이 켑스트럼 평균 제거(CMS: cepstral mean subtraction) 방법으로 보상할 수 있다[8].

$$c_M(t,k) = c_y(t,k) - M_c\, c_w(k) \tag{16}$$

여기서 $c_M(t,k)$는 음성 프레임 t에서 켑스트럼 평균이 보상된 켑스트럼이고 $c_w(k)$는 무음 구간에서 채취한 잡음 $w(n)$의 평균 켑스트럼이다. 상수 M_c는

$$M_C = w_M\frac{n_{mean}}{y_{mean}} \tag{17}$$

로 구한다. 여기서 w_M은 신호 대 잡음비에 의존되는 하중계수로서 주변 잡음의 크기에 따라 실험적으로 구해지는 상수 값이다.

스펙트럼 기울기와 켑스트럼 평균을 동시에 보상한 켑스트럼 $c_{T,M}(t,k)$을 다음의 식으로 둔다.

$$c_{T,M}(t,k)=\begin{cases} c_y(t,k)-\dfrac{T_c}{k^2}-M_c c_w(k) & \text{for} \quad k=odd \\ c_y(t,k) & \text{for} \quad k=even \end{cases} \tag{18}$$

그림 3은 스펙트럼 기울기과 켑스트럼 평균을 동시에 보상한 켑스트럼 $c_{T,M}(t,k)$의 스펙트럼(진한 실선)을 예시한 것이다.

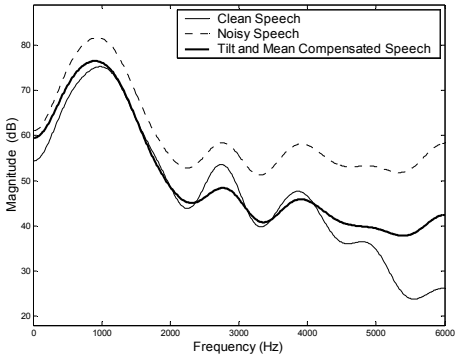

그림 3. 모음 '아'의 깨끗한 켑스트럼, 잡음이 섞인 켑스트럼 및 스펙트럼 기울기와 켑스트럼 평균를 동시에 보상한 켑스트럼의 스펙트럼 예시

Fig.3. Illustration of spectra of clean cepstrum, noisy cepstrum and both spectral tilt and cepstral mean compensated cepstrum of a vowel /ɑ/

잡음이 있는 환경에서 특징 벡터를 추출하여 스펙트럼을 보상하는 전체적인 과정은 그림 4와 같다.

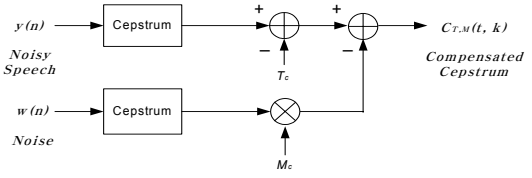

그림 4. 잡음 환경에서 스펙트럼 보상의 블록도

Fig. 4. Block diagram of the spectral compensation under noisy environment

Ⅴ. 실험 및 결과

스펙트럼 기울기 보상과 켑스트럼 평균 보상 방법의 성능을 평가하기 위하여 이산 HMM(discrete hidden Markov model)을 구성하고 화자독립 고립단어 인식시스템을 구현하였다.

1. 음성 데이터와 특징추출

실험에 사용한 어휘는 한국어 숫자 10가지(0~9)와 도시 명 30가지로 구성된 40개의 고립 단어이다. 벡터 양자화기(vector quantizer)의 생성과 HMM의 훈련에는 성인 남성 10명이 각 단어를 20회씩 발음한 8,000개의 데이터를 사용하였고, 인식실험에는 훈련에 참가하지 않은 성인남성 5명이 각 단어를 20회씩 발음한 4,000개의 데이터를 사용하였다. 잡음이 섞인 음성은 백색 가우스 잡음, 지하철 잡음 및 승용차 잡음을 인식용 음성 데이터에 부가하여 만든 것이다.

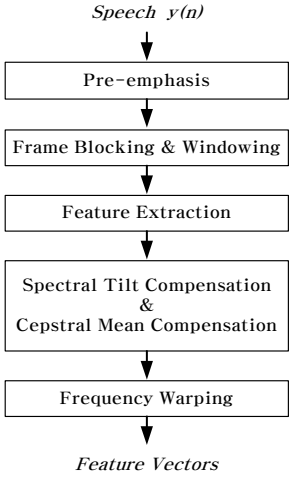

그림 5. 특징 추출 과정
Fig. 5. Feature extraction procedure

음성인식 시스템에 입력된 음성신호와 잡음은 5.4 kHz의 차단주파수를 가지는 저역통과 여파기(low pass filter)를 통과하고, 12 kHz로 샘플링 된 후 16 bits/sample로 양자화 되어 디지털 데이터로 저장된다. 잡음이 섞인 음성신호는 그림 5와 같은 전처리 과정을 거쳐 특징 벡터 열로 변환된다. 음성신호는 전달함수 $G(z)=1-0.97z^{-1}$인 프리엠퍼시스(pre-emphasis) 필터를 거쳐서 주파수 특성을 평탄하게 변환된다. 다음으로 길이 20 ms인 음성 프레임으로 분할되어 해밍 창 함수(Hamming window function)를 씌우게 된다. 특징추출 과정에서 각 음성프레임은 16차 선형예측 분석을 거쳐 LPC-켑스트럼 계수로 변환한다. 다음으로, 잡음 환경에서 음성인식을 위하여 스펙트럼 기울기 보상과 켑스트럼 평균 보상 과정을 거치고, 마지막으로 1차 전역통과 필터(all-pass filter)[12][13]를 이용하여 mel-스케일(mel-scale)의 켑스트럼으로 변환한다. 원래의 주파수 ω와 멜-스케일로 변환된 주파수 $\hat{\omega}$의 관계는

$$\hat{\omega}=\omega+2\tan^{-1}\left(\frac{\alpha\sin\omega}{1-\alpha\cos\omega}\right) \tag{19}$$

이고, 이 논문에서는 $\alpha=0.45$로 두었다.

2. 화자독립 음성인식 시스템

화자독립 음성 인식기는 이산 HMM에 기초한 고립단어 인식 시스템이며, 각 단어 HMM은 5 상태의 단순 좌우 구조를 가진다. 벡터 양자화기(vector quantizer)는 256 레벨의 코드북을 사용하였다. 상태 지속시간 모델(duration model)은 HMM의 각 상태에서 지속시간의 평균과 분산을 정규분포로 근사화한 것이다.

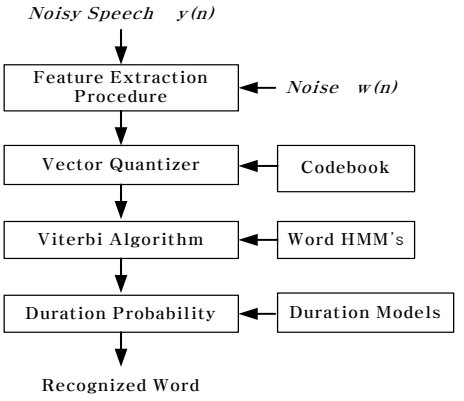

그림 6. 음성인식 시스템의 블록도
Fig. 6. Block diagram of speech recognition system

3. 화자독립 음성인식 실험의 결과

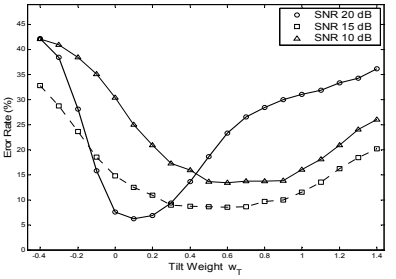

그림 7. 신호 대 잡음 비(SNR)와 하중계수 w_T에 따른 오인식률의 비교
Fig. 7. Comparison of the recognition error rates for different SNR 's and weighting coefficients w_T

그림 7은 신호 대 잡음 비(SNR)와 스펙트럼 기울기의 하중계수 w_T에 따른 오인식률을 비교한 것으로, SNR이 낮을수록 하중계수 w_T가 더 큰 값에서 좋은 인식성능을 보인다.

그림 8은 단어 별 평균 스펙트럼 기울기의 확률 밀도(probability density)를 비교한 것으로 SNR에 따라 스펙트럼 기울기의 분포는 명확한 차이를 보인다. 백색 가우스 잡음에 대한 SNR이 낮을수록 평균 스펙트럼 기울기는 양의 값으로 치우치는 경향을 보인다.

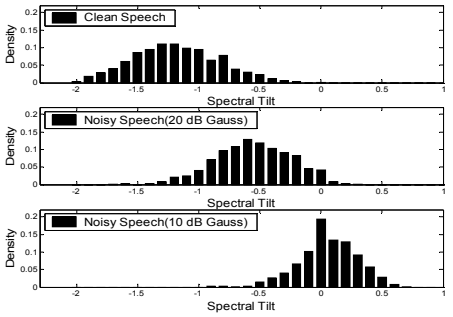

그림 8. 백색 가우스 잡음을 가지는 음성에 대한 스펙트럼 기울기의 확률밀도 비교

Fig. 8. Comparison of density of the spectral tilts for noisy speech with different white Gaussian noise

그림 9는 SNR이 20 dB인 지하철 잡음이나 SNR이 20 dB인 자동차 잡음이 있는 음성신호에 대해 단어별 평균 스펙트럼 기울기의 확률 밀도를 비교한 것이다. 지하철 잡음이나 자동차 잡음이 있는 음성은 깨끗한 음성에 비해 스펙트럼 기울기가 음의 값으로 치우치는 경향을 보인다.

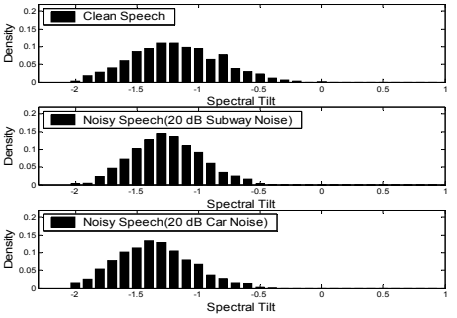

그림 9. 20 dB 지하철 잡음이나 20 dB 자동차 잡음을 가지는 음성에 대한 스펙트럼 기울기의 확률밀도 비교

Fig. 9. Comparison of density of the spectral tilts for noisy speech with 20 dB subway noise or 20 dB car noise

제안한 방법의 타당성을 확인하기 위하여 인식실험을 수행하였다. 성능개선의 기준은 오인식률(recognition error rate)의 감소여부이다. 표 1은 백색 가우스 잡음 하에서 원래의 켑스트럼 $c_y(t,k)$, 스펙트럼 기울기가 보상된 켑스트럼 $c_T(t,\ k)$, 켑스트럼 평균이 보상된 켑스트럼 $c_M(t,k)$ 및 기울기와 평균이 동시에 보상된 켑스트럼 $c_{T,M}(t,k)$을 이용한 인식 시스템에서 오인식률을 비교한 것이다. 특징 벡터로 기존의 켑스트럼을 사용한 기본 인식기의 오인식률은 5.2 %이나 SNR을 감소시키면 오인식률은 급격히 상승한다. SNR이 15 dB인 경우, 원래의 켑스트럼 $c_y(t,k)$의 14.8 %에 비하여 기울기가 보상된 켑스트럼 $c_T(t,\ k)$는 8.5%의 낮은 오인식률을 보였다. SNR이 20 dB 이하인 경우 $c_{T,M}(t,k)$는 오인식률을 1/5 ~ 1/2 정도 줄인다.

표 1. 백색 가우스 잡음 하에서 원 켑스트럼 $c_y(t,k)$, 기울기 보상 켑스트럼 $c_T(t,\ k)$, 평균 보상 켑스트럼 $c_M(t,k)$ 및 기울기와 평균 동시 보상 켑스트럼 $c_{T,M}(t,k)$의 평균 오인식률 비교

Table 1. Comparison of average error rates (%) of original cepstrum $c_y(t,k)$, tilt compensated cepstrum, mean compensated cepstrum $c_M(t,k)$ and both tilt and mean compensated cepstrum $c_{T,M}(t,k)$ under white Gaussian noise condition

SNR(dB)	Feature vectors			
	$c_y(t,k)$	$c_T(t,\ k)$	$c_M(t,k)$	$c_{T,M}(t,k)$
inf. (clean speech)	5.2	n/a	n/a	n/a
25	5.5	5.5	5.3	5.3
20	7.7	6.2	7.3	6.0
15	14.8	8.5	11.0	8.5
10	30.4	13.4	16.9	13.4

표 2는 지하철 잡음이 있는 환경에서 오인식률을 비교한 것이다. 기울기가 보상된 켑스트럼 $c_T(t,\ k)$은 오인식률을 크게 낮추지는 못하나 평균이 보상된 켑스트럼 $c_M(t,k)$은 오인식률을 소폭 줄이는 효과를 보였다. SNR이 15 dB 이하인 경우 $c_{T,M}(t,k)$는 오인식률을 1/5 정도 줄인다.

표 2. 지하철 잡음 하에서 켑스트럼 $c_y(t,k)$, $c_T(t,k)$, $c_M(t,k)$ 및 $c_{T,M}(t,k)$의 평균 오인식률 비교

Table 2. Comparison of average error rates (%) of cepstrum $c_y(t,k)$, $c_T(t,k)$, $c_M(t,k)$ and $c_{T,M}(t,k)$ under subway noise condition

SNR(dB)	Feature vectors			
	$c_y(t,k)$	$c_T(t,k)$	$c_M(t,k)$	$c_{T,M}(t,k)$
25	6.3	6.3	6.3	6.3
20	8.6	8.6	7.6	7.6
15	17.3	17.0	13.6	13.2
10	28.1	25.8	24.0	21.9

표 3은 자동차 잡음이 있는 환경에서 오인식률을 비교한 것이다. $c_T(t,k)$, $c_M(t,k)$ 및 $c_{T,M}(t,k)$은 오인식률을 소폭 줄이는 효과를 보였다. $c_{T,M}(t,k)$는 오인식률을 1/10~1/4 정도 줄인다.

표 3. 자동차 잡음 하에서 켑스트럼 $c_y(t,k)$, $c_T(t,k)$, $c_M(t,k)$ 및 $c_{T,M}(t,k)$의 평균 오인식률 비교

Table 3. Comparison of average error rates (%) of cepstrum $c_y(t,k)$, $c_T(t,k)$, $c_M(t,k)$ and $c_{T,M}(t,k)$ under car noise condition

SNR(dB)	Feature vectors			
	$c_y(t,k)$	$c_T(t,k)$	$c_M(t,k)$	$c_{T,M}(t,k)$
25	7.5	5.4	5.6	5.3
20	8.6	8.1	7.7	7.7
15	12.5	11.0	11.3	10.9
10	20.4	18.5	17.6	17.9

Ⅵ. 결 론

본 논문에서는 잡음이 있는 환경에서 스펙트럼 기울기의 보상 방법을 제시하였다. 이 방법은 잡음에 왜곡된 음성신호에 대해 켑스트럼에 기초한 특징 벡터 보상을 수행하는 것으로 스펙트럼 기울기 보상과 켑스트럼 평균 보상에 기초를 둔다. 스펙트럼 기울기 보상은 켑스트럼의 계수를 직접적으로 수정하여, 스펙트럼의 기울기를 선형적으로 변화시킨다. 스펙트럼의 기울기는 음성 신호의 기울기와 환경 잡음의 기울기의 차이에 따라 결정된다. 다음 과정인 켑스트럼 평균 보상은 음성 신호의 평균 스펙트럼을 환경 잡음의 평균 스펙트럼에 맞춘다. 제안한 방법은 잡음과 음성의 평균 로그 에너지에 따라 스펙트럼 기울기와 켑스트럼 평균 보상의 강도를 다르게 하므로, 효과적인 스펙트럼 보상이 가능하다. 백색 가우스 잡음, 지하철 잡음 및 자동차 잡음에 대해 음성인식 실험을 수행한 결과, 제안한 방법은 원래의 켑스트럼을 사용한 경우에 비해 오인식률을 1/10 ~ 1/2 정도 줄였다. 특히 SNR이 15 dB 이하인 백색 가우스 잡음 환경에서도 오인식률을 1/3 ~ 1/2 정도 줄였다. 이 논문에서 제시한 방법은 켑스트럼 영역에서 스펙트럼을 보상을 하므로, 켑스트럼 영역에서 동작하는 대부분의 음성 인식기에 쉽게 활용할 수 있다.

References

[1] P. J. Moreno, Speech Recognition in Noisy Environments, Ph. D, Dissertation, Carnegie Mellon University, 1996.

[2] H. Hermansky, "RASTA processing of speech," *IEEE Trans. Speech Audio processing,* vol. 2, pp. 578-589, Oct. 1994.
DOI: https://doi.org/10.1109/89.326616

[3] M. J. Gales, S. Young, "Robust speech recognition using parallel model combination," *IEEE Trans. Speech Audio processing,* vol. 4. pp. 352-359, Sep. 1996.

[4] J. Y. Ahn, Y. S. Kim, S. H. Kim, K. I. Hur, "A Study on Voice Recognition Pattern matching level for vehicle ECU control," *The Journal of The Institute of Internet, Broadcasting and Communication (JIIBC),* Vol. 10, No. 1, pp.75-80, Feb. 2010.

[5] S. V. Vaseghi and B. P. Milner, "Noise compensation methods for hidden Markov model speech recognition in adverse environments," *IEEE Trans. Speech and Audio Processing,* vol. 5, No. 1, pp. 11-21, Jan. 1997.

[6] D. C. Popescu and I. Zeljkovic, "Kalman filtering of colored noise for speech enhancement," *ICSLP'96,* Philadelphia, vol. 1, pp.426-429, Oct. 1996.

[7] S. F. Boll, "Suppression of acoustic noise in speech using spectral subtraction," *IEEE Trans. Acoust., Speech, Signal Processing,* vol. ASSP-27, no. 2, pp. 113-120, Apr. 1979.

[8] D. Naik, "Pole-filtered cepstral mean subtraction," *ICSLP'95*, Detroit, vol. 1, pp. 157-160, May, 1995.

[9] L. R. Rabiner and R. W. Schafer, *Digital Processing of Speech Signals*, Prentice-Hall, 1978.

[10] J. Deller, Jr, J. Proakis and J. Hansen, *Discrete-Time Processing of Speech Signals,* Macmillan Publishing Co, New York, 1993.

[11] V. Goncharoff, E. VonColln, and R. Morris, "Efficient calculation of spectral tilt from various LPC parameters," *Proc. IASTED*, pp. 60-63, Nov. 1995.

[12] A. Oppenheim and D. Johnson, "Discrete representation of signals," *Proc. of IEEE,* vol. 60, no. 6, pp. 681-691, June, 1972.

[13] P. A. Regalia, S. K. Mitra and P. P. Vaidyanathan, "The digital all-pass filter: A versatile signal processing building block," *Proc. of IEEE,* vol. 76, no. 1, pp. 19-37, Jan. 1988.

저자 소개

조 정 호(정회원)

- 1990년 : 경북대학교 전자공학과 공학박사
- 1992년 ~ 현재 : 동서울대학교 디지털전자과 교수

<주관심분야 : 음성인식, 음성합성, 인공지능>