청각 구조를 이용한 잡음 음성의 인식 성능 향상

Performance Improvement of Speech Recognizer in Noisy Environments Based on Auditory Modeling

정 호 영*, 김 도 영*, 은 종 관*, 이 수 영*
(Ho Young Jung*, Do Yeong Kim*, Chong Kwan Un*, Soo Young Lee*)

* 본 연구는 과기처 출연연구기관 연구개발과제 지원비로 수행된 것입니다 *

요 약

본 논문에서는 청각 모델을 기초로 잡음에 강한 음성 특징 추출을 연구하였다. 청각 모델은 basilar membrane 모델, 섬모 세포(hair cell) 모델과 스펙트럼 출력단으로 구성하였다. Basilar membrane 모델은 음파의 진동에 따른 전달 특성을 묘사한 것으로 대역 통과 필터의 열로 나타난다. 섬모 세포 모델은 basilar membrane의 진동에 의한 신경 물질로의 변환을 나타낸다. 이것은 입력의 상대적인 값에 크게 반응하는 adaptation 기능을 이용하게 되며, 잡음 제거에 중요한 역할을 하게 된다. 스펙트럼 출력단은 각 채널의 평균 firing rate를 이용하여 mean rate spectrum을 형성한다. 그리고 mean rate spectrum을 이용하여 특징 벡터를 추출하였다. 실험 결과는 청각 구조에 기초한 특징 추출이 다른 특징 추출 방법에 비해 잡음에서 더 향상된 성능을 가짐을 보였다.

ABSTRACT

In this paper, we study a noise-robust feature extraction method of speech signal based on auditory modeling. The auditory model consists of a basilar membrane, a hair cell model and spectrum output stage. Basilar membrane model describes a response characteristic of membrane according to vibration in speech wave, and is represented as a band-pass filter bank. Hair cell model describes a neural transduction according to displacements of the basilar membrane. It responds adaptively to relative values of input and plays an important role for noise-robustness. Spectrum output stage constructs a mean rate spectrum using the average firing rate of each channel. And we extract feature vectors using a mean rate spectrum. Simulation results show that when auditory-based feature extraction is used, the speech recognition performance in noisy environments is improved compared to other feature extraction methods.

I. 서 론

음성 인식은 매우 다양한 응용 분야를 가지며 음성 과학과 컴퓨터 기술의 발전에 의해 크게 향상되고 있다. 그러나 몇몇 제한된 성공 사례를 보이고 있으며 상용화를 위해서는 많은 어려움이 남아있다. 이는 실제 환경에 존재하는 배경 잡음을 고려하지 않았기 때문으로 인식기의 성능 저하에 미치는 영향은 심각하다.

음성은 다양한 요인들에 의해 변화를 일으키게 된다. 음성에 영향을 주는 변화 요인으로는 단어와 단어사이 또는 한 단어내에서의 음의 세기 변화, 부가 잡음, 녹음 장비, 반향음, 여러명의 화자가 말할때의 간섭 현상등이 있다. 이처럼 다양한 변화중 녹음 장비, 반향음에 의한 효과는 비교적 쉽게 제거할 수 있는 반면 잡음은 음성 변화율에 비해 상대적으로 느리게 변하는 특성을 가져 심각한 영향을 주게 된다[1].

실제 환경에서는 여러가지 배경 잡음이 존재하여 음성 인식을 더욱 어렵게 만든다. 이것을 극복하기 위해 많은 방법들이 제안되어 왔는데, 다음과 같이 크게 두가지 방법으로 나누어 볼 수 있다. 한가지 방법은 잡음에 강한

* 한국과학기술원 전기 및 전자공학과
접수일자 : 1995년 5월 19일

음성 특징과 거리척도를 이용하는 것이고[2], 다른 방법은 전처리 단계를 추가하여 음질을 개선시킨 후 음성 인식에 적용하는 것이다[1][3]. 이 방법들은 공통적으로 음성 인식의 여러 요소들을 잡음에 강하도록 하는데 목표를 두고 있으나, 아직까지 근본적인 해결책을 제시하지 못하고 있다.

따라서 본 논문에서는 잡음의 영향을 해결하기 위한 좀 더 근본적인 방법으로 청각 모델을 이용한 특징 추출을 제안한다. 청각 모델은 귀의 특성을 수학적으로 묘사한 것으로 Allen[4], Payton[5], Seneff[6], Cohen[7] 모델등이 대표적이며, 이 모델들을 이용한 인식 실험이 행해졌다. Colombi는 Payton 모델을 화자 인식에 적용하였으며[8], Gao는 Seneff 모델을 /p/, /t/, /k/의 인식에 적용하여 청각 모델이 LPC-켑스트럼에 비해 잡음에 강함을 보였다[9]. 또한 Cohen 모델은 IBM의 음성 인식 시스템에 적용되었다. 인간의 귀는 복잡한 잡음 환경에서도 문법적 제약이 거의 없는 자연어를 쉽게 알아들을 수 있으므로, 이의 깊은 분석을 통할때 잡음에 강한 특징 추출이 가능할 것으로 기대된다. 본 논문은 Seneff 모델을 기초로 하였으며, 섬모 세포에서의 adaptation 정도를 크게하기 위해 dynamic range 압축을 제거하고 특징 추출을 위해 프레임마다 구한 평균 firing rate에 log smoothing을 적용하였다. 또한 청각 모델을 잘 살릴수 있는 특징 추출 방법의 결합을 시도하였다. 본 논문의 구성은 다음과 같다. 2장에서 청각 구조의 기본 성질에 대해 알아보고, 3장에서 청각 모델링 과정과 특징 추출 과정을 다룬다. 4장에서는 실험 결과 분석 및 이에 대한 토의를 하고, 마지막으로 5장에서 결론을 맺는다.

Ⅱ. 청각 모델링을 위한 구조적 특성

귀는 크게 외이, 중이, 내이로 나눌 수 있으며, 음의 전달 과정이 다소 명확히 알려져 있는 반면 내이의 신경물질 변환 과정같은 세부적인 내용들은 아직까지도 불분명하다. 이런 어려움에도 인지 과정의 특정한 양상은 정량화가 가능하며, 청각 기관의 자극에 대한 반응을 관찰, 측정함으로써 더욱 실제에 가까운 모델 구성이 가능할 것이다.

외이는 음을 받아들이는 역할을 하고 중이는 음의 진동을 내이로 전달하는 부분으로 그 특성은 저역 통과 필터와 같다. 따라서 외이와 중이는 표본화, 양자화를 통해 디지탈 신호로 바꾸는 전처리 과정에 개념이 포함되며, 실제 특징 추출을 위해서는 내이의 basilar membrane 전달 특성과 섬모 세포에서의 신경 변환 과정이 이용된다.

Basilar membrane에 대해서는 오랜 연구에도 불구하고 아직까지 여러 의견들이 존재한다[4]. 처음으로 제안된 것이 서로 다른 주파수에 일치된 공진기의 열로 보는 것이고, 그 후 서로 다른 주파수 성분을 전파하는 진행파로 묘사한 전송선 모델이 제안되었다. 전송선 모델은 일차원 모델이라고도 하며 대역 통과 필터의 특성을 가진다. 또한 필터의 저주파 부분에서는 완만한 기울기를 가지는 반면 고주파 부분에서는 급격한 기울기를 가져 때때로 저역 통과 필터로 고려되기도 한다[10]. 그러나 귀의 반응이 비선형적인데 비해 선형 이론이라는 점에서 부적절한 면을 가진다. 이를 해결하기 위해 이차원 모델이 제안되었으나 저주파 부분에서 별다른 효과를 주지 못한다. 각 단계에서 밝혀진 것처럼 현재의 모델들은 basilar membrane을 완벽하게 묘사하는데 한계가 있으며, 실제 특징 추출에 이용될 수 있는 것은 대역 통과 필터의 특성을 가지고 저주파일수록 주파수 해상도가 좋은 일반적인 성질이다.

섬모 세포중 음의 전달에 작용하는 것은 내섬모 세포로 basilar membrane의 움직임에 따라 에너지를 생성하고 이것의 방전량에 따라 청각 신경을 자극한다. 이 성질은 청각 신경의 반응이 자극 직후 최대로 되었다가 정상상태로 감소해가는 것으로 설명할 수 있는데, 이를 adaptation 기능이라고 한다. 이런 기능에 의해 자극에 따른 음향학적 특성을 결정지을 수 있다[11]. 예를 들어 모음과 상대적으로 약한 자음이 이어질때 서로 다른 정도로 adaptation이 일어나면서 입력 음성의 특성을 알아낼 수 있다. 따라서 adaptation은 자극의 변화를 강조하며, 상대적으로 느리게 변하는 잡음을 효과적으로 제거하는데 큰 역할을 한다고 보여진다.

Ⅲ. 청각 모델을 이용한 특징 추출

본 논문에 사용된 청각 모델은 그림 1처럼 세 단계로 나누어 볼 수 있다. 첫 단계는 basilar membrane의 작용을 묘사한 대역 통과 필터열로 각 필터는 서로 독립된 채널을 형성하게 된다. 두번째는 입력 자극에 따라 청각 신경으로의 firing 정도를 나타내는 섬모 세포 모델이다. 따라서 각 채널은 basilar membrane 진동에서부터 청각 신경의 반응을 묘사하기 위한 선형 대역 통과 필터와 비선형단을 포함한다. 마지막으로는 두 단계를 통과한 각 채널 신호의 외형으로부터 mean rate spectrum을 구해 특징 벡터를 형성하게 된다. 그럼 각 단계별로 구현 과정을 살펴보자.

A. Basilar membrane 모델

대역 통과 필터 설계시 고려해야 할 점은 기본적으로 필터 모양뿐 아니라 각각의 대역폭과 중심주파수이다. 필터 모양은 크게 중심주파수에 대칭인 형태와 비대칭인 형태로 나누어 볼 수 있으나, 인식에 있어서는 중요한 변수가 아니다. 특정한 형태가 정해진 것이 아니라 구현 방법에 따라 다양한 종류가 있으며 본 논문에서는 식 (1)의 임펄스 응답을 가지는 필터를 사용하였다.

$$g(t)=\frac{at^{n-1}\cos(2\pi f_c t)}{e^{2\pi bt}} \tag{1}$$

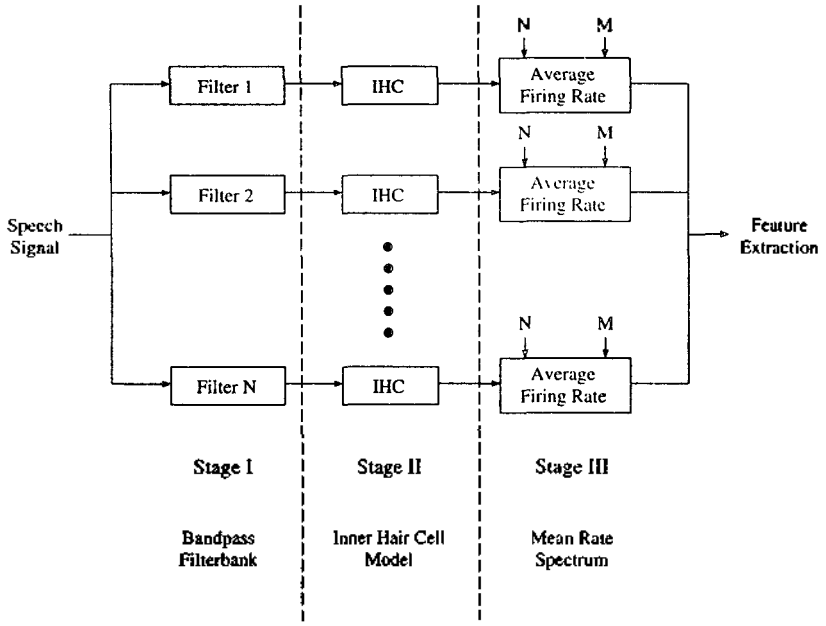

그림 1. 청각 모델을 이용한 특징 추출 과정
Fig 1. Feature extraction procedure using auditory model

여기서 필터 차수 n은 basilar membrane 특징을 가장 잘 묘사하는 4로 결정하였으며[12], 임펄스 응답의 Laplace 변환으로부터 8차 디지탈 필터로 구현됨을 알 수 있다. 또한 이 필터는 중심주파수 f_c에 따라 서로 다른 대역폭 b를 가진다.

중심주파수와 대역폭은 저주파일수록 주파수 해상도가 높은 특성을 이용하여 결정하게 된다. 이를 결정하는 대표적인 방법은 Zwicker에 의해 제안된 critical bandwidth이다[13]. 그러나 특징 추출의 효율이 중심주파수에 따라 변화하기 때문에 500 Hz 이하에서 일정한 대역폭을 가지는 critical bandwidth는 오차를 줄 수 있으며[14], 몇몇 실험 결과는 실제 반응이 critical band보다 더 급격한 기울기를 가짐을 보여준다[15]. 이런 이유로 본 논문에서는 equivalent rectangular bandwidth (ERB)를 사용하였다. ERB는 청각 필터의 모양을 추정하는 과정에서 유도된 것으로 일반적인 형태는 식 (2)와 같다.

$$ERB = \left[\left(\frac{f}{Q} \right)^{order} + B_n^{order} \right]^{\frac{1}{order}} \tag{2}$$

여기서 Q는 필터의 quality factor, B_n은 최소대역폭을 나타내며 ERB와 B_n는 Hz 단위, f는 kHz 단위이다. 그리고 각 변수의 값은 여러 실험들을 통해 다음과 같이 다양하게 제안되어있다[16].

	Q	B_n	$order$
Lyon	8	125	2
Greenwood	7.23824	22.8509	1
Glasberg	9.26449	24.7	1

여기에 결정된 값들은 주파수에 대한 ERB의 기울기 정도를 나타내며 가장 급격한 기울기를 갖는 Glasberg의 값을 사용하였다.

중심주파수는 앞에서 정의된 ERB로부터 유도되며 다음식과 같이 결정하였다.

$$fc_i = -QB_n + (f_x + QB_n)\, e^{i(-\log(f_x+QB_n)+\log(f_l+QB_n))/M},$$

$$i = 0, \cdots, M-1 \tag{3}$$

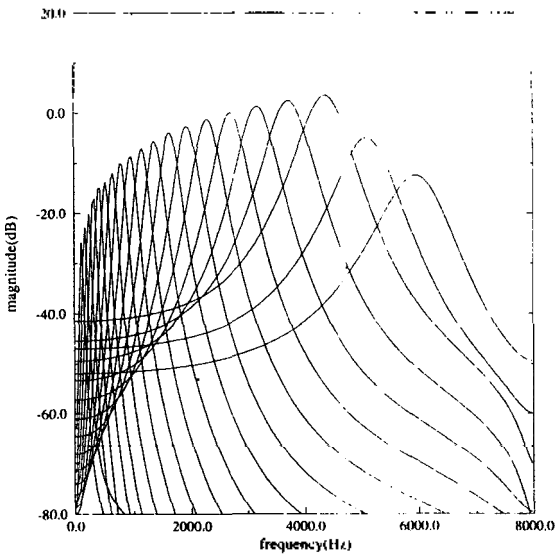

그림 2. Basilar membrane을 묘사하는 대역 통과 필터열
Fig 2. Bandpass filter bank for basilar membrane

이 식에서 f_h는 최대 주파수, f_l은 최소 주파수, M은 필터 갯수를 나타낸다. 이렇게 정의된 세 식을 이용하여 basilar membrane의 응답을 묘사하는 모델을 구할 수 있으며, 구성된 필터는 그림 2와 같다.

B. 섬모 세포 모델

입력 자극에 따라 내섬모 세포에서의 신경 전달 과정을 나타내는 것으로 Seneff 모델을 기초로 구성하였다[6]. 채널별로 독립적으로 이루어지며 반파 정류, short-term adaptation, 저역 통과 필터와 rapid automatic gain control (AGC)의 순서로 이루어진다. 저역 통과 필터를 제외한 모든 성분이 비선형이므로 최종 출력은 각각의 배열 순서에 따라 영향을 받게 된다. 모델의 구성은 그림 3에 주어져 있으며, 각각의 작용은 다음과 같다.

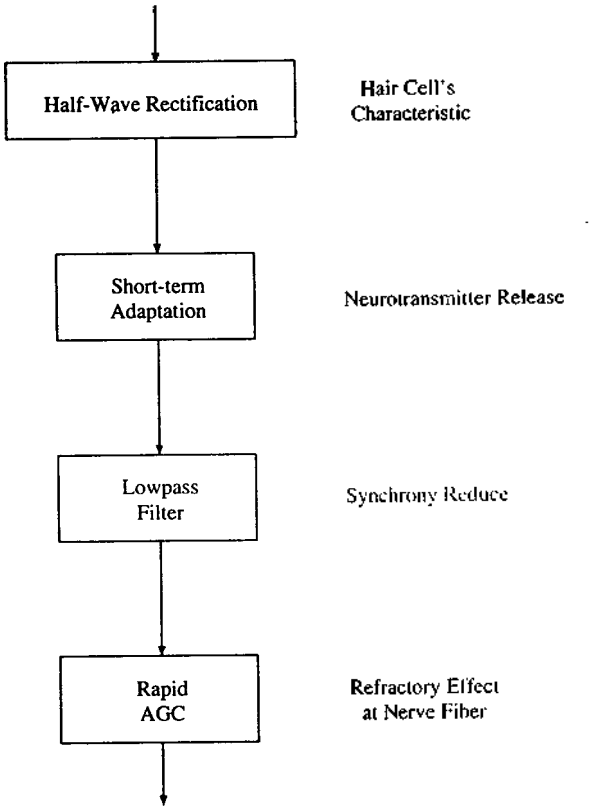

그림 3. 섬모 세포 모델
Fig 3. Hair cell model

반파 정류는 섬모 세포가 basilar membrane 진동의 어느 한 방향에만 민감한 특성을 이용한 것으로, 모델 구성에 있어 필요한 성분이다. 본 논문에서는 adaptation 정도를 크게하기 위해 기존의 dynamic range 압축 과정을 제거했으며, 이는 잡음의 영향을 상대적으로 줄여줄 뿐만 아니라 계산량의 부담을 덜어주게 된다. 이 과정은 다음식과 같으며 K_{hwr}은 필터 출력에 따라 결정되는 상수이다.

$$y=\begin{cases} x & , \quad x>0 \\ (K_{hwr})^{-|x|} & , \quad x\le 0 \end{cases} \tag{4}$$

Short-term adaptation은 전달 물질의 농도에 따라 두 가지 과정으로 나누어져 있다. 하나는 공급원의 농도가 클때 섬모 세포를 투과하여 농도의 변화량에 비례한 율로 cleft에 공급되는 것이고, 다른 하나는 공급원의 농도가 매우 작을때 섬모 세포가 닫혀지고 cleft의 농도에 비례한 율로 자연 감소되는 것이다. 이 과정은 다음식으로 표현될 수 있다.

$$\frac{dc(t)}{dt}\begin{cases} \mu_a[s(t)-c(t)]-\mu_b c(t), & c(t)>s(t) \\ -\mu_b c(t) & , \quad c(t)\le s(t) \end{cases} \tag{5}$$

여기서 $c(t)$는 cleft의 농도, $s(t)$는 공급원의 농도이다.

저역 통과 필터는 고주파 자극에 대한 synchrony 감소를 나타내며 smoothing 효과를 가진다. 동일한 시상수를 가지는 적분기의 종속 결합 형태로 이루어지며, α는 적분기의 시상수에 해당하는 극점의 위치를 나타낸다.

$$H(z)=\left(\frac{1-\alpha}{1-\alpha z^{-1}}\right)^n \tag{6}$$

마지막 성분인 rapid AGC는 자극이 있은 후 얼마동안 반응하지 않는 무반응효과를 묘사한 것으로 다음식과 같다.

$$y(n)=\frac{x(n)}{1+K_{agc}(x(n))_{LP}} \tag{7}$$

이 식에서 K_{agc}는 AGC 상수이고 $(x(n))_{LP}$는 일차 저역 통과 필터를 통한 $x(n)$의 출력을 나타낸다.

섬모 세포 모델은 위의 네 단계로 이루어지며 최종 출력은 음의 진동에 따른 청각 신경에서의 firing rate를 나타낸다.

C. 특징 추출

청각 해석 방법은 크게 temporal 성분을 강조하는 우세 성분 방법(dominant-component scheme), spatial 성분을 강조하는 국소 필터링 방법(local-filtering scheme)과 average localized synchronous response(ALSR)처럼 두가지 특성을 모두 가지는 방법이 있다[17][18]. 우세 성분 방법은 전체 채널의 해석으로부터 스펙트럼을 형성하는 것으로 EIH가 대표적인 방법이다[19]. 반면 국소 필터링 방법은 각 채널에서의 해석을 요구하는 것으로 central spectrum과 관련되어 있으며 firing rate의 외형으로부터 얻을 수 있다[18].

본 논문에서는 채널별로 독립적으로 해석했으므로 central spectrum 방법에 근거해 섬모 세포 모델까지 통과한 각 채널의 외형으로부터 짧은 시간 간격마다 mean rate spectrum을 구해 특징 벡터를 형성하였다. 섬모 세

포 출력은 청각 신경에서의 firing rate와 동일한 것으로 간주할 수 있으므로 짧은 시간 단위의 평균 firing rate를 이용해 mean rate spectrum을 얻을 수 있다. 평균 firing rate는 프레임 단위로 계산하였으며 다음 식과 같다.

$$fr_k(i) = \frac{1}{L} \log_{10} \sum_{j=1}^{L} hc_k^2(j) \quad (8)$$

여기서 L은 프레임의 길이, $hc_k(j)$는 k번째 주파수 채널에서 섬모 세포의 출력이며 $fr_k(i)$는 k번째 주파수 채널에서 i번째 프레임의 평균 firing rate이다. 또한 log 함수는 입력 변화에 따른 firing rate의 급격한 증가를 제한하는 역할을 한다.

이렇게 각 채널에서 구한 평균 firing rate를 이용한 특징 추출 방법으로 채널/프레임 정규화를 시도하였다. 이 방법은 잡음의 영향과 서로 다른 화자에 따른 영향을효과적으로 제거하게 된다. 채널/프레임 정규화의 과정은 다음과 같다.

잡음의 정도에 따라 각 채널에서의 bias 효과가 달라지게 되며 이로 인해 인식기의 성능은 급격히 떨어지게 된다. 이를 해결하기 위해 채널별로 출력값의 최대, 최소로 dynamic range를 제한하였다. 계산량을 줄이기 위해 프레임 단위의 평균 firing rate를 구한후 적용하였으며 잡음에 의한 bias 변화를 보상해 줄 것으로 기대된다[20].

1. 각 채널에서 식 (8)를 이용하여 평균 firing rate를 구한다.
2. 평균 firing rate 열에서 최대값과 최소값을 찾는다.
3. 다음 식을 이용해 채널별로 정규화한다.

$$\overline{fr_k}(i) = \frac{fr_k(i) - (fr_k(i)|_{\min} - \beta)}{(fr_k(i)|_{\max} + \alpha) - (fr_k(i)|_{\min} - \beta)} \quad (9)$$

여기서 $fr_k(i)|_{\max}$와 $fr_k(i)|_{\min}$은 $\{fr_k(i),\ i=1, \cdots,$ 전체프레임수$\}$의 최대 값과 최소값이며, α와 β는 최대값과 최소값 근처에서의 급격한 변화를 고려하여 정규화 범위를 넓혀주는 역할이다.

4. 각 채널의 같은 시간대 프레임으로부터 특징 벡터를 형성하고 다음 식을 이용해 프레임 별로 정규화한다. 이것은 서로 다른 화자에 의한 신호 레벨 변화를 줄이는 효과를 줄 수 있다.

$$\widetilde{fr_k}(i) = \frac{\overline{fr_k}(i)}{\sum_{j=1}^{M} \overline{fr_k}(i)} \quad (10)$$

여기서 $\widetilde{fr_k}(i)$는 $\overline{fr_k}(i)$의 정규화된 값이고 M은 주파수 채널의 수이다.

Ⅳ. 인식 실험

A. 데이타베이스

인식 실험에 사용된 데이타베이스는 음운학적으로 균형을 이룬 75개의 고립 단어로 이루어져 있다. 녹음은 조용한 사무실 환경에서 이루어졌고 발음한 음성 신호는 16kHz, 16bit로 A/D변환 되었다[21]. 학습 데이타는 15명의 화자가 한번씩 발음한 것으로 구성되었고 인식 실험에는 학습에 참가하지 않은 5명의 화자가 한번씩 발음한 것을 사용하였다. 실험에 사용된 단어는 표 1과 같다.

표 1. 인식에 사용된 단어 목록.
Table 1. Word list for recognition.

아들	애기	밥	바퀴	뿔	비행	보리	창
달	다리	딸	들깨	등살	된장	돌다리	동백
동이	동쪽	동태	의사	가보	값이	가구	가족
갈치	감기	감자	간판	간식	글	꿀	고뼈
곡식	구리	구웠다	괜찮다	귤	하나	홀러	확기적
자리	잣새	잣송이	찌개	줄기	칼	마음	맵새
목	나	남기	날뛰다	남산	늑대	농비	옷
옷밥	왼쪽	풀	사람	셈	쌀	투구	왔다
완수	웬일	원고	약속	양	에	역사	연못
욕	용산	육성					

잡음이 섞인 음성은 원하는 신호대 잡음비에 맞게 잡음의 크기를 조절해 더해 주었으며, 각 프레임마다의 신호 전력에 따라 정하는 것이 아니라 단어 전체의 신호 전력을 이용하여 결정하였다. 신호대 잡음비를 구하는 식은 다음과 같다.

$$SNR = 10 \log_{10} \left| \frac{\frac{1}{L} \sum_{j=1}^{L} s^2(i)}{P_N} \right| \quad (11)$$

여기서 L은 단어의 길이, $s(i)$는 음성 신호이고 P_N은 잡음 전력을 나타낸다. 그리고 이렇게 단어 전체의 신호 전력을 이용하는 것이 실제 상황에 더 가까울 것으로 생각된다.

잡음은 평균이 영인 백색 gaussian 잡음과 실제 환경에서의 잡음으로 에어컨 잡음을 이용하였다. 백색 잡음은 전 주파수 대역에, 에어컨 잡음은 주로 저주파 대역에 영향을 끼치게 된다.

B. 실험 결과

잡음하에서의 인식 실험을 통해 잡음이 인식률 저하에 미치는 영향을 살펴보고 기존 특징 추출 방식과 청각 모델을 이용한 방식을 비교하였다. 인식기로는 연속 분포 HMM을 사용하였으며 단어 단위로 모델을 구성하였다. 단어 모델의 상태(state)수는 단 음절의 경우 5개, 나머지는 10개로 이루어져 있다.

기존 특징 추출 방식중 비교 대상으로는 filter bank를

사용하였다. 현재 인식 시스템에서 많이 채택하고 있는 LPC-캡스트럼은 배경 잡음이 큰 경우 성능이 급격히 떨어지는 단점을 가지는데 반해 filter bank는 LPC-캡스트럼 보다 잡음에 강한 성질을 보이기 때문이다[20]. 10ms마다 30ms의 프레임에 FFT를 적용해 스펙트럼을 구하고, log 분포로 18개 영역으로 나누어 각 영역의 에너지와 전체 에너지를 이용해 19차로 형성하였다.

청각 모델은 다시 살펴보면 basilar membrane을 묘사하는 대역 통과 필터열, 섬모 세포 모델과 평균 firing rate를 이용한 특징 추출로 이루어 진다. 대역 통과 필터열은 앞에서 정의된 식 (1), (2), (3)을 이용하여 125-6500Hz 대역에 40 채널로 구성하였다. 섬모 세포 모델은 Seneff 모델에 기초하여 인식 환경에 맞게 구현하였으며, 특징 추출을 위해서는 30ms마다 각 채널의 평균 firing rate를 이용하였다. 또한 filter bank와의 비교를 위해 식 (12)로 인접 채널사이에 평균값을 구해 최종 20차로 형성하였다.

$$\widehat{fr_m}(i)=\frac{1}{2}\sum_{2m-1}^{2m}\widetilde{fr_k}(i), \qquad m=1,\cdots,\frac{M}{2} \tag{12}$$

백색 잡음에 대한 실험 결과는 표 2에 주어진다. Filter bank는 30dB 이상에서 좋은 성능을 보이나 잡음의 정도가 심해지면 급격히 떨어지는 단점을 가진다. 이에 반해 청각 모델을 이용한 방법은 20dB 이하에서 filter bank에 비해 훨씬 높은 인식률을 보였다. 또한 실험 결과는 채널/프레임 정규화 방법이 간단한 계산으로 잡음의 영향을 효과적으로 제거할 수 있음을 알려준다.

표 2. 백색 잡음에서의 인식률(%).
Table 2. Recognition rate for white gaussian noise(%).

SNR(dB)	Filter bank	청각 모델
∞	94.1	92.5
30	90.1	89.6
20	52.3	76.0
10	12.0	59.2
5	5.6	29.9

에어컨 잡음에 대한 실험 결과는 표 3에 나타나 있다. 백색 잡음과는 달리 주로 저주파 대역에 영향을 끼쳐 filter bank가 20dB에서도 성능을 유지하였으나, 10dB로 잡음의 정도가 심해지면 성능이 급격히 감소하였다. 청각 모델은 20dB 이상에서는 filter bank와 비슷한 성능을 가지면서 그 이하에서는 더 높은 성능을 가지는 완만한 성능 저하를 보였다. 앞의 두 실험으로부터 청각 모델이 잡음에 강한 성질을 가지고, filter bank와의 인식률을 비교해 볼때 백색 잡음처럼 음성에 미치는 영향이 큰 경우 또는 신호대 잡음비가 아주 낮은 경우 더 효과가 있음을 알 수 있다. 따라서 청각 모델은 주위의 잡음 특성을 모르거나 환경 변화가 심한 경우 유용하다고 생각된다.

표 3. 에어컨 잡음에서의 인식률(%).
Table 3. Recognition rate for aircon noise(%).

SNR(dB)	Filter bank	청각 모델
∞	94.1	92.5
30	93.6	92.3
20	87.1	90.4
10	35.2	76.5

V. 결 론

본 논문에서는 잡음 환경에서의 인식률 향상을 위해 인간의 귀에 기초한 특징 추출을 시도하였다. 아직까지 귀의 구조에 대한 지식은 불완전하나 일부 특정한 양상은 여러 실험을 통해 알려져 있으며, 이를 바탕으로 청각 모델을 구성하게 된다.

청각 모델은 대역 통과 필터열, 섬모 세포 모델과 mean rate spectrum을 이용한 특징 추출로 구성하였다. 대역 통과 필터열은 basilar membrane에서의 전달 특성을 묘사한다. 섬모 세포 모델은 신경 전달 물질로의 변환 과정을 묘사한다. 이것은 일정한 음이 계속될 경우 청각 신경이 점점 무감각해져서 firing이 감소되는 adaptation 기능을 나타내며 입력 자극의 변화 성분을 강조한다. 따라서 음성에 비해 상대적으로 느리게 변하는 잡음을 제거하는데 중요한 역할을 한다.

인식 실험은 백색 잡음과 에어컨 잡음하에서 이루어 졌으며 청각 모델은 잡음의 정도가 심할수록 더 큰 효과를 보였다. 또한 실험 결과는 청각 모델의 가능성을 제시하였다. 물론 다른 데이타베이스나 잡음에 대한 실험을 통해 객관성을 얻을 필요는 있으나, 본 실험만으로 타당성을 충분히 엿볼 수 있을 것이다. 귀의 전달 특성에 대한 깊은 연구와 간단하고 효율적인 모델의 개발이 잡음의 진정한 해결책을 줄 것으로 기대한다.

참 고 문 헌

1. A. Acero, *Acoustical and Environmental Robustness in Automatic Speech Recognition*, Kluwer Academic Publishers, U. S. A., 1993.
2. D. Mansour and B. H. Juang, "A Family of Distortion Measures Based upon Projection Operation for Robust Speech Recognition," *IEEE Trans. on ASSP*, Vol. 37, No. 11, pp. 1659-1671, Nov. 1989.
3. S. F. Boll, "Supression of Acoustic Noise in Speech Using Spectral Subtraction," *IEEE Trans. on ASSP*, Vol. 27, No. 2, pp. 113-120, Apr. 1979.
4. J. B. Allen, "Cochlear Modeling," *IEEE ASSP Magazine*, pp. 3-29, Jan. 1985.
5. K. L. Payton, "Vowel Processing by a Model of the Auditory Periphery : A Comparison to Eighth-Nerve Responses," *J. Acoust. Soc. Am.*, Vol. 83, pp. 145-162, 1988.

6. S. Seneff, "A Joint Synchrony/Mean-Rate Model of Auditory Speech Processing," *J. Phonet.*, Vol. 16, pp. 55-76, 1988.
7. J. R. Cohen, "Application of an Auditory Model to Speech Recognition," *J. Acoust. Soc. Am.*, Vol. 85, pp. 2623-2629, 1989.
8. J. M. Colombi, *et al.*, "Auditory Model Representation for Speaker Recognition," *Proc. of ICASSP*, Vol. II, pp. 700-703, 1993.
9. Y. Gao, T. Huang, S. Chen, and J. P. Haton, "Auditory Model Based Speech Processing," *ICSLP*, pp. 73-76, 1992.
10. J. M. Kates, "A Time-Domain Digital Cochlear Model," *IEEE Trans. on Signal Processing*, Vol. 39, No. 12, pp. 2573-2592, Dec. 1991.
11. E. Jones and E. Ambikairajah, "Comparison of Various Adaptation Mechanisms in an Auditory Model for the Purpose of Speech Processing," *EUROSPEECH*, pp. 717-720, 1993.
12. R. D. Patterson, *et al.*, "Complex sounds and auditory images," In *Auditory Physiology and Perception*, Y Cazals, L. Demany, K. Horner, Pergamon, Oxford, 1992.
13. E. Zwicker, "Subdivision of the Audible Frequency Range into Critical Bands," *J. Acoust. Soc. Am.*, Vol. 33, pp. 248, 1961.
14. C. J. Moore and R. Glasberg, "Suggested Formulae for Calculating Auditory-Filter Bandwidths and Excitation Patterns," *J. Acoust. Soc. Am.*, Vol. 74, pp. 750-753, Sep. 1983.
15. R. H. Dye, Jr. and E. R. Hafter, "Just-Noticeable Differences of Frequency for Masked Tones," *J. Acoust. Soc. Am.*, Vol. 67, pp. 1746-1753, 1980.
16. M. Slaney, "An Efficient Implementation of the Patterson-Holdsworth Auditory Filter Bank," *Apple Computer Tech. Report #35*, 1993.
17. S. A. Shamma, "Speech Processing in the Auditory System II : Lateral Inhibition and the Central Processing of Speech Evoked Activity in the Auditory Nerve," *J. Acoust. Soc. Am.*, Vol. 78, pp. 1622-1632, 1985.
18. B. Delgutte, "Speech Coding in the Auditory Nerve : II. Processing schemes for Vowel-Like Sounds," *J. Acoust. Soc. Am.*, Vol. 75, pp. 879-886, 1984.
19. O. Ghitza, "Auditory Nerve Representation as a Front-End for Speech Recognition in a Noisy Environment," *Computer Speech and Language*, pp. 109-130, 1986.
20. 정 호영, *청각 구조를 이용한 잡음 환경에서의 음성 특징 추출에 관한 연구*, 한국과학기술원 석사학위 논문, 1995.
21. 최 인정, 권 오욱, 박 종렬, 김 도영, 정 호영, 은 종관, "자동통역용 한국어 음성 데이타베이스," *음성 통신 및 신호처리 워크샵 논문집*, pp. 287-290, 1994.

▲정 호 영(Ho-Young Jung) 1970년 3월 8일생
1993년 2월 : 경북대학교 전자공학과 졸업 (공학사)
1995년 2월 : 한국과학기술원 전기 및 전자공학과 졸업 (공학석사)
1995년 3월~현재 : 한국과학기술원 전기 및 전자공학과 박사과정

▲김 도 영(Do-Yeong Kim)
13권 1호 참조

▲은 종 관(Chong-Kwan Un)
10권 3호 참조

▲이 수 영(Soo-Young Lee) 1952년 10월 15일생
1975년 2월 : 서울대학교 공과대학 전자공학과 (공학사)
1977년 2월 : 한국과학원 전기공학과 (공학석사)
1984년 5월 : Polytechnic Institute of New York, Electrophysics (공학박사)
1977년~1980년 : 대한엔지니어링(주) 과장대리
1983년~1985년 : General Physics Corp. Staff Scientist/Senior Scientist
1986년~현재 : 한국과학기술원 전기 및 전자공학과 조교수/부교수/교수