청각모델을 이용한 음성신호의 특징 추출 방법에 관한 연구

박규홍 , 김영호 , 정상국 , 노승용
서울시립대학교 전자공학과

Speech Feature Extraction Using Auditory Model

Kyuhong Park, Youngho Kim, Sangkuk Jung, Seungyong Rho
Dept. of Electronic Engineering, University of Seoul

Abstract - Auditory Models that are capable of achieving human performance would provide a basis for realizing effective speech processing systems. Perceptual invariance to adverse signal conditions (noise, microphone and channel distortions, room reverberations) may provide a basis for robust speech recognition and speech coder with high efficiency.

Auditory model that simulates the part of auditory periphery up through the auditory nerve level and new distance measure that is defined as angle between vectors are described.

1. 서 론

음성신호의 특징 추출 방법은 음성인식에서의 인식률 및 부호화에서의 부호효율과 밀접한 연관관계를 가진다. 만일 음성신호의 특징 추출 파라미터가 음성신호의 상태(잡음, 마이크 종류, 소리의 잔향 등)의 변화에도 변화가 적고 서로 다른 음에 대해서는 추출된 음성신호의 특징 파라미터의 편차가 크다면 음성인식의 경우 인식률을 높일 수 있고 부호기에서는 부호 효율을 증대시킬 수 있다. 따라서 음성신호의 특징 추출 방법은 음성신호처리에서 중요한 부분 중의 하나이다.

기존의 음성신호의 특징 추출 방법은 음성신호로부터 켑스트럼 계수를 얻어내거나 음성신호로부터 LPC계수를 구한 다음 변환식을 통해 켑스트럼 계수로 변환시켜 특징 파라미터를 구하는 방법을 사용하였다. 그리고 높은 주파수보다는 낮은 주파수에서 인간의 신호분석능력이 뛰어나다는 인간의 청각적 특성을 고려한 Mel-scale 분석방법이 사용되어져 왔다.

그러나 지금까지의 특징 추출 방법은 인간의 청각기관에 대한 접근이였다기보다는 신호분석의 한 분야로써의 접근이였다고 볼 수 있다.

본 논문에서는 먼저 Ghitza가 제안한 청각모델 EIH를 기초로 하여[1][2], 인간의 청각기관의 특징을 이용한 음성신호의 특징 추출 방법을 제시하였고, 또한 distance measure로 널리 쓰이는 cepstral distance나 euclidian distance 대신 두 벡터 사이의 각을 distance로 사용하는 새로운 distance measure를 제안하였다.

2. 본 론

2.1 청각기관에 대한 모델링

청각기관인 귀의 구조는 그림[1]과 같이 소리를 모아주는 외이, 고막부분을 포함하는 중이 그리고 달팽이관을 포함하는 내이로 구성되어있다. 여기서 달팽이관 내부의 청각 세포(inner hair cells)들이 청각신경과 연결되어 있다. 청각 세포는 달팽이관의 기저막(basilar membrane)으로부터의 거리에 따라 자극을 받는 소리의 주파수 대역이 달라지게 된다. 기저막에서 가까운 청각 세포는 멀리 있는 청각 세포보다 낮은 주파수 신호에 민감하게 자극을 받고 기적막에서 먼 곳에 있는 청각 세포는 높은 주파수 신호에 민감하게 자극을 받게 된다.

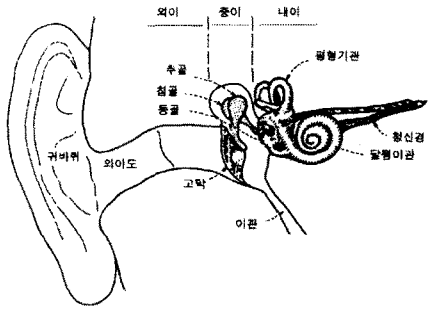

그림 1. 귀의 구조

청각 세포(inner hair cells)의 수는 대략 4000개로 추정되며 Goldstein이 제안한 논문에 따르면 청각 세포는 기저막으로부터의 정규화된 거리에 따라 아래 식과 같이 특정 주파수에 가장 민감한 반응을 나타낸다고 한다.[1]

$$F = A(10^{ax} - 1) \qquad (1)$$

여기서 F는 주파수, x는 기저막으로부터의 정규화된 거리(0<x<1)를 나타내며, 상수 A=165.4, a=2.1이다.[1]
기저막으로부터의 정규화된 거리에 대한 청각세포의 특성 주파수의 분포는 그림[2]와 같다

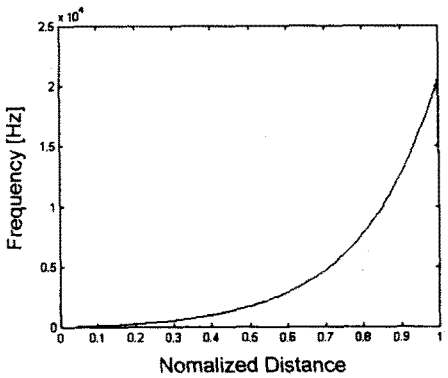

그림 2. 청각세포의 특성주파수 분포

그림[3]은 주파수에 대한 청각 세포들의 자극의 정도를 dB scale로 나타내고 있다.[1][2] 그림[3]에서 주파수에 대한 청각 세포들의 자극 범위를 보면 낮은 주파수를 특성주파수로 하는 청각 세포는 주파수의 변화에 민감하고 반응하는 주파수의 범위는 좁다. 반면, 높은 주파수를 특성 주파수로 하는 청각 세포는 주파수의 변화에 덜 민감하고 반응하는 주파수 범위가 넓음을 알 수 있다.

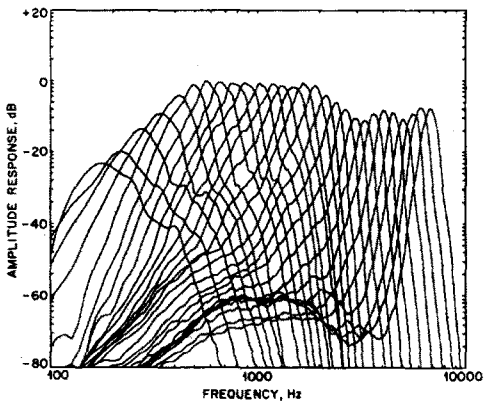

그림 3. 청각세포의 주파수 응답, Ghitza[1]

그림[3]에서 400Hz와 4000Hz를 특성주파수로 하는 청각신경의 자극을 normalized scale로 각각에 대하여 살펴보면 그림[4]와 같다.

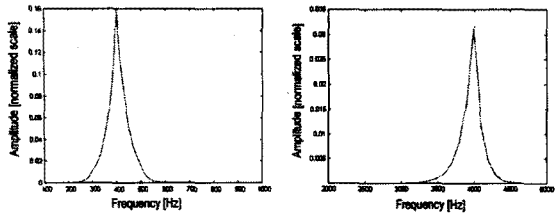

그림 4. 400Hz와 4kHz에서의 청각세포의 주파수 응답

본 논문에서는 그림[3],[4]를 통하여 청각 세포의 자극의 최대값을 A_c , 특성 주파수를 f_c라고 놓고, 주파수에 대한 청각 세포의 자극을 다음과 같은 식으로 모델링하였다.

$$H(f) = A_c e^{-\frac{|f-f_c|}{\alpha f_c + \beta}} \tag{2}$$

여기서 α, β는 상수로 각각 α=0.0217과 β=33.32 이다.

2.2 청각신경에 대한 모델링

Ghitza의 청각모델 EIH모델에서 가정한 사항, 즉 청각신경은 같은 주파수 신호의 자극을 모아서 자극에 대한 정보를 뇌에 전달해준다는 가정을 하면 특정 주파수에 대한 자극의 정도는 특정 주파수에 대한 청각 모세포의 자극들을 모두 합한 것과 같다. 이러한 모델링을 간략화하여 도시하면 그림[5]와 같다

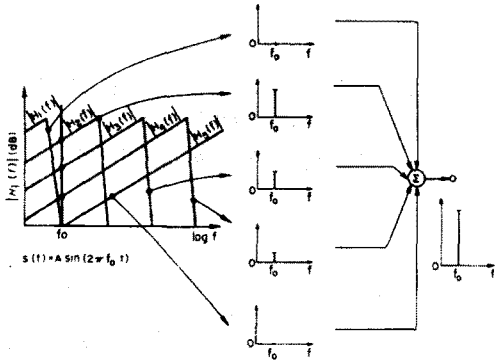

그림 5. 단일 주파수 신호 입력에 대한 청각모델의 동작

청각 모델을 이용한 파라미터 추출 방법을 수식으로 나타내면 다음과 같다.

음성신호 x(n)에 창의 길이가 N인 창함수 w(n)을 씌워 푸리에 변환을 구하여 푸리에 계수 C_n을 구하면 $C_n = X(e^{j2\pi f_s n/N})$로 나타낼 수 있다. 푸리에 계수에 대한 크기를 $b_n = |C(n)|$ 이라고 하면 청각 모델에 의해 추출된 n번째 파라미터 P_n은 다음과 같다.

$$P_n = \sum_{i=0}^{N/2-1} b_i A_i' e^{-\frac{|f_n - f_i|}{\alpha f_i + \beta}} \tag{3}$$

여기서 f_i와 f_n은 각각 푸리에 변환에서 i번째, n번째 계수가 나타내는 주파수를 의미하며, f_s는 표본화 주파수, N은 푸리에 변환된 표본의 수이고 A_i'은 청각세포의 주파수 f_i에 대한 주파수 응답의 최대 크기 A_i에 주파수 f_i에 분포하는 청각 세포의 수를 가중시킨 값을 나타낸다. 따라서 P_n은 주파수 f_n에 대한 파라미터로 모델링된 청각 신경의 자극 계수가 된다.

2.3 Distance Measure

일반적인 distance measure는 음성신호의 크기에 대한 영향을 최소화하기 위하여 distance가 log scale로 계산되어진다. 스펙트럼 $S_1(w)$와 $S_2(w)$ 간의 distance는 일반적으로 다음 식으로 계산되어진다.[2][6]

$$d(S_1, S_2) = \int_{-\pi}^{\pi} |\log S_1(w) - \log S_2(w)|^2 \frac{dw}{2\pi} \tag{4}$$

또는

$$d(S_1, S_2) = \sqrt{\sum_{k=1}^{N} |\log S_1(k) - \log S_2(k)|^2} \tag{5}$$

이러한 distance measure는 같은 패턴 벡터에 대해서도 크기의 차이가 많이 나게되면 distance가 커지는 단점이 있다.

본 논문에서는 이러한 단점을 극복한 새로운 distance measure를 제시한다.[3]

벡터간의 distance를 벡터사이의 각(angle)으로 정의한다. 따라서 벡터의 크기에 상관없이 패턴이 같은 벡터의 경우는 distance가 0이 된다.

스펙트럼의 주파수 성분에 대한 계수들을 벡터의 요소들로 본다면 벡터 S1과 S2 간의 각(angle) distance는 두 벡터의 내적으로 구할 수 있다.

$$d(S_1, S_2) = \left(1 - \frac{|S_1| \cdot |S_2|}{|S_1||S_2|}\right) \times W \tag{6}$$

W는 잡음과 음성신호를 구별하기 위한 가중치로 기준 벡터에 대해 비교되는 벡터의 크기가 아주 작으면 W의 값은 큰 값을 가지게 되며 그 이외의 범위에서는 1이다.

2.4 실험 결과

본 논문에서 제안한 청각모델 및 새로운 distance measure의 성능향상 평가를 위한 실험이 수행되었다.

성능평가를 위해 음소단위의 인식실험을 하였으며 음성 샘플 데이터는 16bit, 16kHz로 녹음된 시스템공학연구소의 음성 데이터베이스 PBW를 사용하였다.

표1-표4는 잡음환경과 특징 추출 방법 및 distance measure에 따른 음소 인식기의 인식률을 나타내고 있다.

표에서 청각모델의 인식결과는 본 논문에서 제안한 청

각모델과 새로운 distance measure를 사용한 벡터 양자화의 결과이다.

청각모델의 성능 비교를 위하여 본 논문에서는 LPC 스펙트럼을 기존의 distance measure를 사용하여 벡터 양자화를[2][5] 수행한 LPC-1과 본 논문에서 제안한 distance measure를 사용하여 벡터 양자화를 수행한 LPC-2를 이용하여 잡음환경[4] 하에서의 잡음의 크기 변화에 따른 인식률의 차이를 살펴보았다.

LPC-1은 잡음에 매우 약함을 알 수 있으며 잡음이 증가할수록 양자화 결과가 특정한 코드 워드로 양자화 되는 경향이 있었다.

LPC-2는 LPC-1보다 잡음에 강함을 알 수 있었으며 새로운 distance measure가 기존의 방법보다 우수함을 증명해 주었다. 그러나 이 방법 역시 잡음이 증가함에 따라 양자화 결과가 특정한 코드 워드로 양자화 되는 경향이 있었다.

이에 반해 청각모델은 잡음의 증가에 따른 인식률의 변화가 크지 않았으며 잡음이 증가함에 따라 변화된 음성신호의 특징에 가까운 코드 워드로 양자화 되는 경향이 있었다.

	Accuracy %				
	마찰음	폐쇄음	비음	기타	모음
LPC - 1	64.1	50.3	46.9	52.2	60.6
LPC - 2	67.1	75.1	59.4	71.4	70.6
청각모델	70.5	83.2	59.4	79.1	78.3

표1. 잡음이 없는 경우의 인식률

	Accuracy %				
	마찰음	폐쇄음	비음	기타	모음
LPC - 1	67.2	46.0	68.8	45.2	69.6
LPC - 2	70.9	73.3	62.5	73.8	78.7
청각모델	68.8	83.2	59.4	79.1	77.1

표2. SNR이 40dB인 경우의 인식률

	Accuracy %				
	마찰음	폐쇄음	비음	기타	모음
LPC - 1	41.8	5.3	50.0	3.1	29.2
LPC - 2	57.7	30.9	53.1	43.8	52.2
청각모델	62.1	81.9	62.5	63.7	71.6

표3. SNR이 20dB인 경우의 인식률

	Accuracy %				
	마찰음	폐쇄음	비음	기타	모음
LPC - 1	43.3	0.0	46.9	0.0	20.0
LPC - 2	59.5	24.4	50.0	28.1	27.1
청각모델	44.0	56.8	43.8	52.2	68.0

표4. SNR이 10dB인 경우의 인식률

그림[6]은 음소 /ㅔ/에 대한 LPC 스펙트럼의 잡음에 대한 변화를 나타내고 있다. 아래 검정색 곡선은 잡음이 없는 신호에 대한 LPC 스펙트럼을 나타내고 있고 위의 회색 곡선은 SNR이 10dB인 신호에 대한 LPC 스펙트럼을 나타내고 있다. LPC 스펙트럼의 경우 잡음이 증가함에 따라 distance가 매우 커짐을 알 수가 있다.

그림[7]은 음소 /ㅔ/에 대한 청각모델의 잡음에 대한 변화를 나타내고 있다. 검정색 그래프 곡선이 잡음이 없는 신호에 대한 청각모델의 스펙트럼이고 회색 그래프 곡선이 SNR이 10dB인 신호에 대한 청각모델의 스펙트럼이다. 청각모델의 경우 잡음이 증가하여도 distance의 변화가 적음을 알 수가 있다.

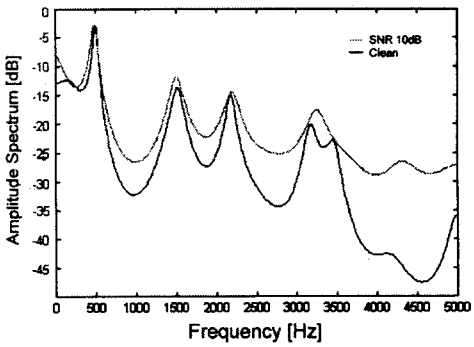

그림 6. LPC 스펙트럼의 잡음에 대한 영향

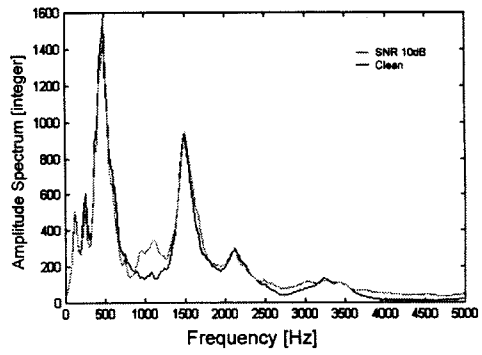

그림 7. 청각모델의 잡음에 대한 영향

3. 결 론

본 논문에서는 잡음환경에 강한 음성인식 시스템 개발의 기초가 되는 전처리과정에 대한 연구로써, 이미 잡음환경 하에서 보다 나은 성능을 입증 받은 Ghitza의 EIH 모델에 기초한 청각모델과 두 벡터 사이의 각을 distance로 사용하는 새로운 distance measure를 제안하였다. 실험결과 본 논문에서 제안한 distance measure의 사용이 기존의 방법보다 향상된 결과를 가져왔으며, 특히 잡음환경 하에서는 청각모델을 사용한 특징 추출 방법과 새로운 distance measure의 사용이 기존에 사용하였던 방법보다 개선된 결과를 보임을 확인하였다.

[참 고 문 헌]

[1] Oded Ghitza, "Auditory Models and Human Performance in Tasks Related to Speech Coding and Speech Recognition", IEEE, vol.2, pp.115-132, 1994

[2] Biing-Hwang Juang, Lawrence Rabiner, "Fundamantals of Speech Recognition", Prentice Hall, 1993

[3] Mitchell, Harper, Jamieson, "Using Explicit Segmentation to Improve HMM Phone Recognition, IEEE, pp.229-232, 1995

[4] Hayakawa, Itakura, "The Influence of Noise on the Speaker Recognition Performance using the Higher Frequency Band", IEEE, pp.321-324, 1995

[5] Teuvo Kohonen, "Improved Versions of LVQ", Proc. IJC-NN'90, vol.1, pp.545-550, June 1990

[6] John R. Deller, John G. Proakis, John H. L. Hansen, "Discrete-Time Processing of Speech Signals", Prentice Hall, 1987