醫工學會誌 : 第 9 卷, 第 1 號, 1988
J. of KOSOMBE : Vol. 9, No. 1

研 究 論 文
88 - 1 - 01

말초 청각 계통 모델을 이용한 한국어 모음 인식

윤태성 • 백승화* • 박상희

= Abstract =

Korean Vowel Recognition using Peripheral Auditory Model

Tae-Sung Yoon, Seung-Hwa Beack*, Sang-Hui Park

In this study, the recognition experiments for Korean vowel are performed using peripheral auditory model. In addition, for the purpose of objective comparison, the recognition experiments are performed by extracting LPC cepstrum coefficients for the same speech data.

The results are as follows.

1) The time and the frequency responses of the auditory model show that important features of input signal are involved in the responses of inner ear and auditory nerve.

2) The recognition results for Korean vowel show that the recognition rate by auditory model output is higher than the recognition rate by LPC cepstrum coefficients.

3) The adaptation phenomenon of auditory nerve provides useful characteristics for the discrimination of vowel signal.

1. 서 론

인간은 청각 기관을 통하여 음성 신호를 받아들여 이로부터 물리적인 정보를 추출하고 이것과 언어에 대한 방대한 지식을 활용하여 유입되는 메세지를 판독하게 된다. 이런점에서 볼 때, 인간의 청각 계통은 우수한 성능을 지닌 하나의 음성 신호 인식 장치로 간주할 수 있다. 따라서, 음성 인식 시스템을 설계하는데 있어서 인간의 청각 계통에 관한 지식을 활용하는 것이 유용할 것이라는 가능성을 갖게 한다. 특히, 최근에 들어 음성 인식 장치 앞단(front end)의 설계에 말초 청각 계통(peripheral auditory system)에서의 음성 신호 처리 특성이 효과적으로 활용될 수 있을 것이라는 주장이 음성 신호 연구가 사이에 강하게 대두되고 있다.

한편, 외부의 소리 입력에 대한 말초 청각 계통의 응답 특성에 대하여 많은 연구가 이루어져 왔으며, 이러한 연구의 방향은 생리학적 연구와 정신음향학적 연구로 나누어 볼 수 있다. 생리학적 연구는 Békésy[1], Flanagan[2], Rhode와 Robles[3], Wilson과 Johnstone[4], Schroeder와 Hall[5], Young과 Sachs[6], Allen[7], Seneff 등[8]에 의하여 수행되어 왔다. 이 연구의 주요 내용은 음성 또는 음성과 유사한 입력 신호에 대하여 기저막(basilar membrane), 헤어셀(haricell) 및 청각 신경에서의 응답 특성을 조사하는 것이다. 정신음향학적 연구는 Zwicker[9], Carlson 등[10]에 의하여 수행되어 왔는데, 이 연구의 목적은 임계 대역(critical band), 매스킹(masking) 현상, 라우드니스(loudness)와 같은 정신물리학적 매개변수를 이용하여 말초 청각 계통에서의 소리

< 접수 : 1988년 5월 20일 >
연세대학교 전기공학과
Dept. of Electrical Eng., Yonsei University
*명지대학교 전기공학과
*Dept. of Electrical Eng., Myongji Univerity

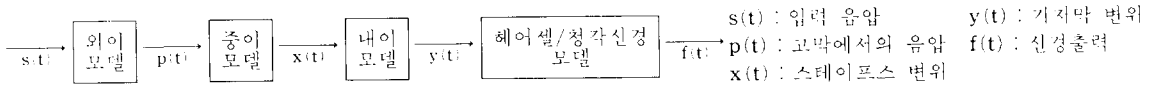

Fig. 1 Block diagram of peripheral auditory model

에 대한 지각 특성을 조사하는 것이다.

본 논문에서는 위와같은 연구 결과를 토대로하여 말초 청각 계통 모델을 구성하고 한국어 모음에 대한 인식 실험을 수행하고자 한다. 그리고 이를 통하여 말초 청각 계통에서의 모음 처리 특성을 살펴보고, 또한 음성 인식 장치에 말초 청각 계통의 음성 신호 처리 특성을 활용하는데 대한 유용성을 살펴보고자 한다. 그리고 인식 결과의 객관적인 비교를 위하여 동일한 모음 데이터에 대하여 LPC 켑스트럼 계수(LPC cepstrum coefficients)를 구하여 인식 실험을 행하고, 이 결과와 말초 청각 계통 모델의 출력에 의한 인식율을 비교해 보기로 한다.

2. 말초 청각 계통 모델 구성

음파가 외이(outer ear)에 들어오면 음압에 의하여 고막이 진동하며, 이 진동은 중이(middle ear)를 거쳐 내이(inner ear)의 달팽이관(cochlea)에 전달된다. 달팽이관에 유입된 신호는 기저막 위를 진행하면서 입력 신호의 주파수 성분에 따라 각각 기저막의 한 점에서 최대의 공진을 일으킨다. 기저막 위 코르티 기관(the organ of Corti)에는 헤어셀이라고 불리우는 섬모 세포들이 분포되어 있고 여기에 청각 신경이 연결되어 있다. 이 헤어셀/청각 신경에서 기저막의 운동이 신경 발화 활동으로 변환되어 중추 신경을 거쳐 뇌로 전달되게 된다. 말초 청각 계통은 이중에서 외이로부터 헤어셀/청각 신경 부분까지를 말한다.

Fig. 1은 본 연구에서 사용한, 외이 모델에서 헤어셀/청각 신경 모델 까지의 말초 청각 계통 모델에 대한 블럭선도이다. 여기서는 Fig. 1을 이루고 있는 각 모델에 관하여 기술하기로 한다.

2-1 외이 모델

Fig. 1의 외이의 전달 특성은 소리의 수평 입사각에 관계되는데 일반적으로 4 KHz 또는 이보다 약간 낮은 주파수에서 약 11 dB의 공진 특성을 보이고 7~10 KHz의 주파수 영역에서 반 공진 특성을 나타낸다. Monro[11]는 6 KHz까지의 외이의 주파수 특성을 만족시켜주며 4 KHz에서 11 dB의 공진 특성을 갖는 식(1)과 같은 2차 모델을 제안하였다.

$$\frac{P(s)}{S(s)}=\frac{1}{(s+1000)^2+(7875\cdot\pi)^2} \tag{1}$$

여기서,

$S(s)$; 외이에 가해지는 음압에 대한 라플라스 변환

$P(s)$; 고막에서의 음압에 대한 라플라스 변환

2-2 중이 모델

1975년 Wilson과 Johnstone[4]은 중이의 전달 특성이 3 KHz까지는 편평한 특성을 갖고 이후, 20 KHz까지 6 dB/oct의 기울기로 감쇠함을 보여 주었다. Monro[11]는 이와같은 측정 결과에 의거하여 식(2)와 같은 중이에 대한 1차 모델을 제안하였다.

$$\frac{X(s)}{P(s)}=\frac{1}{s+6000\cdot\pi} \tag{2}$$

여기서,

$P(s)$; 고막에서의 음압에 대한 라플라스 변환

$X(s)$; 스테이프스(stapes) 변위에 대한 라플라스 변환

2-3 내이 모델

내이를 이루고 있는 달팽이관의 전달 특성은 기저막의 각 위치마다 다르다. 그러나 일반적으로 대역 통과 필터와 같은 특성을 가지며 고주파 영역에서의 기울기가 저주파 영역에서의 기울기 보다 가파른 특성을 갖는다. Békésy[1]의 데이터에 의하면, 저주파 영역에서의 상승 기울기는 6 dB/oct, 고주파 영역에서의 감쇠 기울기는 24 dB/oct이다. Flanagan[2]은 이러한 Békésy의 측정 결과에 기초하여 식(3)과 같은 내이 모델을 제안하였다.

$$\frac{Y_l(s)}{X(s)}=G\cdot\left[\frac{s}{(s+\beta_l)}\cdot\frac{1}{(s+\alpha_l)^2+\beta_l^2}\right]^2 \tag{3}$$

여기서,

$X(s)$; 스테이프스 변위에 대한 라플라스 변환

$Y_l(s)$; 스테이프스로부터 l만큼 떨어진 지점의 기저막 변위에 대한 라플라스 변환

$\beta_l = 2\alpha_l$; 스테이프스로부터 l만큼 떨어진 기저막 지점의 특성주파수(f_l)에 대한 라디안(radian) 값($=2\pi f_l$)

G ; 이득 요소

한편, 1973년 Rhode와 Robles[3]는 공진점 근방의 저주파 영역에서 상승 기울기가 24 dB/oct, 고주파 영역에서 감쇠 기울기가 100 dB/oct라는 측정 결과를 발표하였다. 이는 식(3)에서 α_l에 대한 β_l의 비, 즉 공진점 근방에서의 선택 특성을 증가시킴으로써 변경시킬 수 있다.

2-4 헤어셀/청각 신경 모델

헤어셀에서 기저막의 기계적 변위는 수용기 전위(receptor potential)로 변환되고 이는 다시 헤어셀에 붙어있는 청각신경에서 활동전위(action potential), 즉 신경발화로 변환된다. 또한 헤어셀/청각 신경에서는 기저막 변위의 반파 정류 작용, 순응(adaptation) 현상, 발화율 포화(saturation) 현상등의 비선형 특성이 존재한다.

1974년 Schroeder와 Hall[5]은 위와같은 특성을 갖는 헤어셀/청각 신경 모델을 제안하였다. 모델의 기본식은 다음과 같다.

$$p(t) = p_0\{1/2 \cdot y(t) + [1/4 \cdot y^2(t) + 1]^{1/2}\} \quad (4)$$

$$r = \overline{n(t) \cdot p(t)} + \overline{n(t) \cdot g} \quad (5)$$

$$f(t) = n(t) \cdot p(t) \text{ or } f(t) = n(t) \cdot p(t) \cdot \rho(t - t_0) \quad (6)$$

$$dn/dt = r - n(t)[p(t) + g] \quad (7)$$

식(4)는 헤어셀 내부에서 생성되는 "quanta"라고 불리우는 전기화학적인 신경 전달 물질이 세포막을 투과할 때의 투과도(permeability function) $p(t)$에 관한 식으로 입력 자극 신호, 즉 기저막의 기계적 변위 $y(t)$에 대한 함수 관계를 나타낸다. p_0는 신경 섬유의 자연 발화(spontaneous firing)와 관련된 상수이다. 식(5)는 정상 상태에서 일정한 평균 비율 r로 생성되는 신경 전달 물질이 일부는 신경 발화를 일으키고 나머지는 신경 발화를 일으키지 않고 소멸되는 관계를 나타낸다. 여기서, $n(t)$는 quanta의 수를 나타내며 g는 quanta의 소멸과 관련된 상수를 나타낸다. 식(6)은 시간에 따른 신경 섬유의 발화 확률($f(t)$)을 나타낸 식으로 특히 오른쪽 식은 신경의 휴지기에 대한 영향을 고려한 식이다. 여기서, $\rho(t - t_0)$는 회복 함수(recovery function)이고 t_0는 바로 전의 신경 발화가 일어난 후 경과된 시간을 나타낸다. 식(7)은 $p(t)$, r, g가 주어졌을 때 $n(t)$를 구하는 미분 방정식이다.

3. 실 험

3-1 청각 모델 방정식 변환

본 연구를 위하여 앞에서 기술한 말초 청각 모델 방정식을 디지탈 음성 데이터 처리를 위한 이산(discrete) 전달 함수의 형태로 변환시켰다. 식(1)~(3)은 영점-극점 매핑 방법(matched pole-zero mapping method)을 사용하여, 식(7)은 후향 차분법(backward difference method)을 사용하여 z-변환 형태의 전달 함수로 변환시켰다.

내이 모델은 일종의 대역 통과 필터 뱅크(band pass filter bank)의 형태를 갖는다. 그런데, 인간의 말초 청각 계통에서의 채널(channel)의 수는 약 30,000개에 달하며[7], 이를 전부 고려할 경우 실험에 소요되는 시간이 매우 길어지게 된다. 따라서, 본 연구에서는 계산 시간 및 모음 신호의 주요 성분의 주파수 대역을 고려하여 채널의 수를 18 또는 36개로 하였고 채널의 중심 주파수의 범위는 50 Hz~4000 Hz로 제한하였다. 채널의 수가 18인 경우는 중심 주파수의 간격을 1 Bark, 36인 경우는 0.5 Bark로 하였다. [Bark] 단위는 정신물리학에서 임계 대역을 나타낼 때 사용하는 주파수 단위이다. 1 Bark는 청각 계통의 정신물리학적 동조 곡선에서 −3 dB점 간의 임계 대역폭을 나타낸다[9]. 중심 주파수의 [Hz] 단위로의 환산은 1980년 Zwicker와 Terhardt[12]가 제시한 다음 식을 이용하였다.

$$f[\mathrm{KHz}] = \begin{cases} \dfrac{\tan\left(\dfrac{Z_c[\mathrm{Bark}]}{13}\right)}{0.76}, & f < 1\,\mathrm{kHz} \\ 10^{\left(\dfrac{Z_c[\mathrm{Bark}] - 8.7}{14.2}\right)}, & f > 1\,\mathrm{kHz} \end{cases} \quad (8)$$

여기서, Z_c ; [Bark] 단위의 중심 주파수

Table 1은 채널의 수가 18인 경우, 식(8)에 의하여 계산한 내이 모델 각 채널의 중심 주파수의 값과, 변환된 내이 모델 방정식에서 실제 구현된 값을 표시한다.

또한, 내이 모델의 각 채널에서 선택도(quality factor ; Q) 값이 3.5가 되도록 식(3)에서 $\alpha_l = 0.144\,\beta_l$로 하였으며, 실제 구현된 값은 $Q_{5\mathrm{dB}} = 3.5$이었다.

Fig. 2는 채널의 수가 18인 경우의 변환된 내이 모델의 크기 및 위상 특성을 나타낸다.

Table 1 Center frequency of inner ear model (18 channel)

채널	계산값(Hz)	실제값(Hz)
1	50.	50.
2	150.4	149.4
3	250.	248.
4	350.9	352.9
5	454.5	454.5
6	571.4	571.4
7	689.7	692.7
8	833.3	838.3
9	1000.	1004.
10	1176.5	1183.5
11	1333.3	1339.3
12	1538.5	1546.5
13	1818.2	1826.2
14	2222.2	2226.2
15	2500.	2501.0
16	2857.1	2854.1
17	3333.3	3320.3
18	4000.	3961.

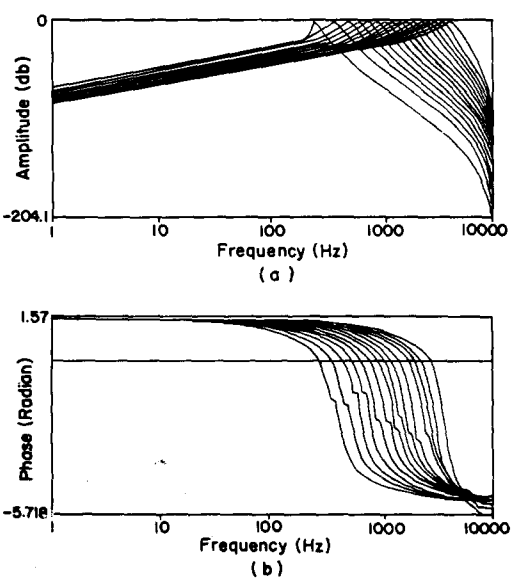

Fig. 2 Frequency characteristics of inner ear model (18 channel)
(a) amplitude characteristic
(b) phase characteristic

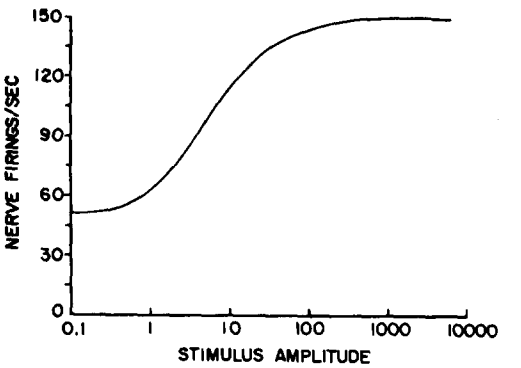

Fig. 3 Average firing rate in response to 1 KHz tone stimulus

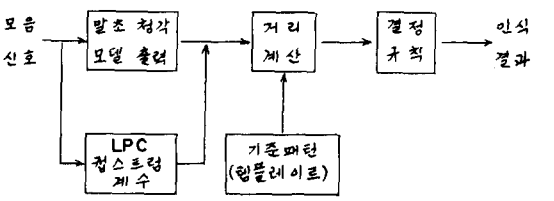

Fig. 4 Overall system of vowel recognition

헤어셀/청각 신경 모델은 식(4)～(7)에서 r을 150/sec, g를 33.3/sec, p_0를 16.7/sec로 하였다. 이 때, 1KHz 톤(tone) 입력에 대한 입력 신호의 크기와 평균 발화율과의 관계를 살펴보면 Fig. 3과 같이 된다.

3-2 모음 인식 시스템

모음 인식 시스템을 Fig. 4와 같이 구성하여 한국어 모음 /아/, /에/, /이/, /오/, /우/를 대상으로 인식 실험을 수행하였다. 인식 실험은 화자 종속(speaker dependent) 실험과 화자 독립(speaker independent) 실험으로 나누어 수행하였다. 또한, 각 실험에 대해서 피치(pitch) synchronous 실험과 피치 asynchronous 실험으로 나누어 수행하였다. 피치 synchronous 실험에서는 특징 추출을 위한 분석구간을 각 모음의 피치 주기만큼 취하였으며, 피치 asynchronous 실험에서는 분석구간을 25.6 msec로 하였다.

음성 데이터는 6명의 남성 화자(speaker)가 발음한

/아/, /에/, /이/, /오/, /우/ 5종류의 한국어 단 모음을 녹음기에 수집하여 20 KHz로 샘플링하여 사용하였다. 각 화자는 5종류의 모음을 각각 5번씩 반복 발음하였다. 따라서, 6화자×5모음×5번 발음=150개의 모음 데이터가 수집되었다. 화자 종속 실험은 각 화자가 발음한 25개의 음성 데이터를 대상으로 하였고, 화자 독립 실험은 6명의 화자가 발음한 150개의 음성 데이터를 대상으로 하였다.

특징 추출은 다음과 같이 하였다. 18채널 및 36채널 말초 청각 모델의 내이 출력, 청각 신경 출력으로부터 일정길이의 데이터를 취하여 평균화 시켰으며, 이 값을 최대값으로 나누어 정규화 시킨후 해당 모음을 대표하는 하나의 특징 벡터를 만들었다. 인식 결과의 비교를 위하여 동일한 음성 데이터에 대하여 선형 예측(LPC) 계수를 구하고, 이 계수로부터 다음 식(9)[13]에 의하여 LPC켑스트럼 계수를 구하여 특징 벡터를 만들었다. LPC계수의 차수 및 LPC켑스트럼 계수의 수는 28로 하였고, 최소 자승 평균 오차(ε_{min})는 1로 정규화 시켰다.

$c_0 = l_n\ \varepsilon_{min}$

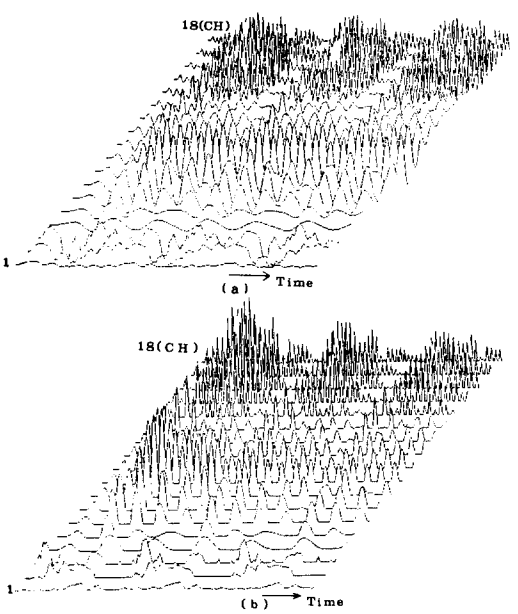

Fig. 5 Outputs of auditory model for vowel /아/
(a) Output of inner ear
(b) Output of auditory nerve

$$c_i = -a_i - \frac{1}{i}\sum_{k=1}^{(-)} k c_k a_{i-k},\quad 1 \le i \le p \tag{9}$$

여기서,

c_i ; LPC 켑스트럼 계수

a_i ; 선형 예측 계수

p ; 선형 예측 차수

각 모음에 대한 템플레이트(template), 즉 기준 패턴은 화자 종속 실험의 경우는 각 화자에 대하여 5개의 모음에 대한 5번의 반복 발음중 첫번째 발음에 대한 특징 벡터를 기준 패턴으로 하였다. 화자 독립 실험의 경우는 각 모음에 대해서 각 화자의 첫번째 발음에 대한 특징 벡터를 6명의 화자에 대하여 평균을 취하여 기준 패턴으로 하였다.

거리 계산은 유클리드(Euclid) 거리 척도를 사용하였고, 거리 계산 결과 최소가 되는 기준 패턴을 최종 인식 패턴으로 하였다.

4. 결 과 고 찰

4-1 청각 모델 출력

Fig. 5의 (a), (b)는 모음 /아/에 대해서 18채널 말초 청각 모델로부터 뽑아낸 내이 출력 및 PST히스토그램 형태의 청각 신경 출력이다. 이 출력들을 관찰해보면, 청각 모델 출력 파형은 모음 /아/의 원 파형과 마찬가지로 주기성을 가지고 있으며, 원 파형의 시간적인 변화를 잘 반영하고 있음을 알 수 있다. 특히 Fig. 5(b)에서 출력 파형의 크기가 일시적으로 증가했다가 일정한 시정수를 가지고 정상상태로 되돌아가는 현상을 보여주는데 이것은 청각 신경의 순응 특성에 기인하는 것이다.

Fig. 5에서 또한 큰 응답의 첨두치들이 특정 채널 영역에 밀집되어 있음을 알 수 있다. 이러한 밀집 부위들은 모음의 주요 특징인 포먼트(formant)를 나타낸다고 볼 수 있다. 이러한 관계를 좀더 자세히 알아보기 위해서, Fig. 6에 모음 /아/에 대하여 원 파형으로부터 구한 대수 파워 스펙트럼(log power spectral density)(가는 선)과 18채널 말초 청각 모델 출력에 의한 스펙트럼(굵은선)을 대수 주파수 축상에 중첩시켜 나타내었다. 여기서, 대수 파워 스펙트럼은 원 파형을 25.6 msec만큼 취하여 선형 예측 계수를 구하여 추정한 것이고, 청각 모델 출력에 의한 스펙트럼은 내이의 대수 에너지 출력 및 청각 신경의 크기 출력을 각 채널에 대하여 25.6 msec

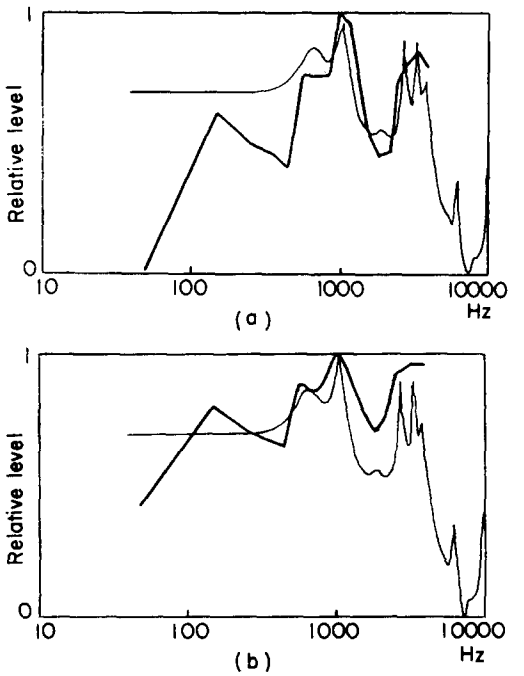

Fig. 6 Log power spectral density (PSD) and average auditory model outputs for vowel /아/
(a) Log PSD and average log energy output of inner ear
(b) Log PSD and average output of auditory nerve

만큼 취하여 평균을 한 것이다. Fig. 6의 (a), (b)에서 모음 /아/에 대한 두 스펙트럼의 첨두치의 위치가 거의 일치함을 알 수 있다. 이는 청각 모델의 평균 출력이 모음에 대한 스펙트럼 상의 주요 특징을 잘 반영하고 있음을 나타낸다. 따라서, 청각 모델 출력은 시간 축에서 뿐만 아니라 주파수 축에서 입력 모음 신호에 대한 특징을 잘 나타내고 있음을 보여준다.

4-2 인식 실험

앞의 결과에 의거하여, 청각 모델 출력에 의한 특징 벡터는 내이의 경우 대수 에너지 출력을, 청각 신경의 경우 크기 출력을 분석 구간에 걸쳐 평균하여 만들었다. Fig. 7과 Fig. 8은 각각 한 화자 SJS가 5번 반복 발음한 5개의 한국어 모음에 대하여 분석구간을 피치 asynchronous로 했을 때의 18채널 청각 모델의 내이 및 청각 신경 출력에 의한 특징 패턴을 나타낸다. 각 경우 모두 특징 패턴들은 모음 종류별로 서로 근접되어 있어 모음 식별을 위한 좋은 특징 매개변수가 될 수 있음을 보여준다.

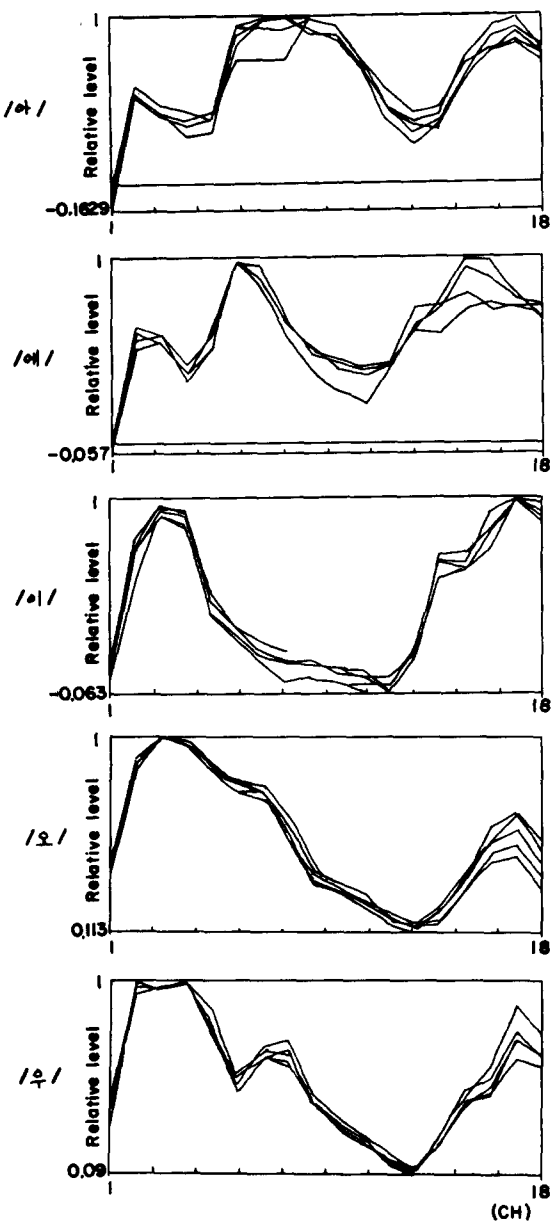

Fig. 7 Feature patterns of inner ear outputs for Korean vowels. (speaker : SJS, channel number ; 18)

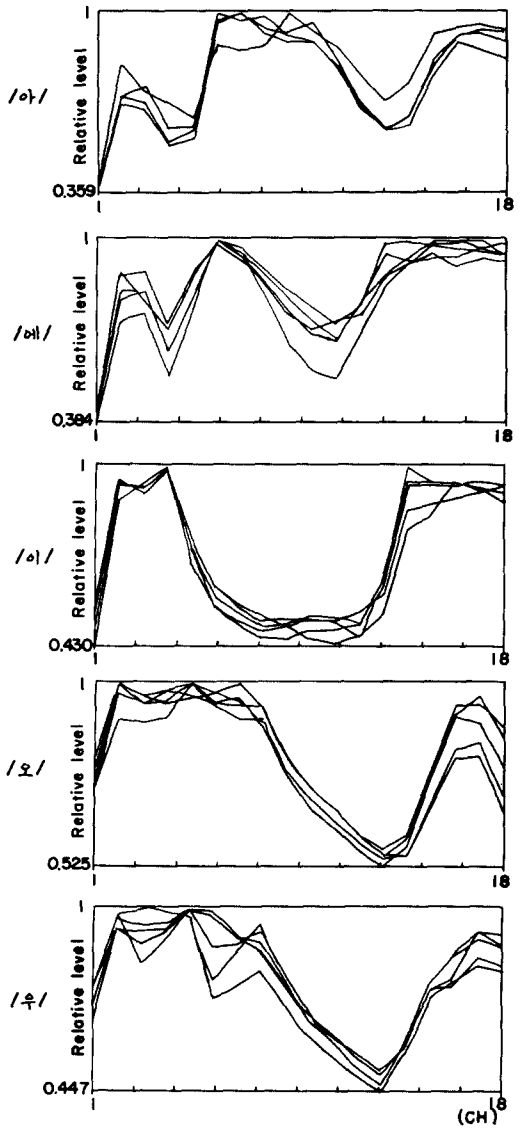

Fig. 8 Feature patterns of auditory nerve outputs for Korean vowels. (speaker ; SJS, channel number ; 18)

Table 2 Speaker dependent recognition results for 5 Korean vowels

특 징 매개변수	프레임 길이	화 자	인식율 (%)
내이출력 (18 CH)	async	RKK	100
		KDJ	100
		NOK	100
		JSW	100
		LSJ	96
		SJS	96
	sync	RKK	92
		KDJ	96
		NOK	100
		JSW	100
		LSJ	96
		SJS	100
신경출력 (18 CH)	async	RKK	100
		NDJ	100
		NOK	100
		JSW	100
		LSJ	100
		SJS	100
	sync	RKK	100
		KDJ	100
		NOK	96
		JSW	96
		LSJ	100
		SJS	100
LPC* 켑스트럼 계 수	async	RKK	100
		KDJ	92
		NOK	100
		JSW	100
		LSJ	88
		SJS	100
	sync	RKK	72
		KDJ	88
		NOK	80
		JSW	84
		LSJ	84
		SJS	88

*LPC 켑스트럼 계수의 수 ; 28

본 연구에서는 18채널 및 36채널 청각 모델 출력으로부터 특징 벡터를 구하여 인식 실험을 수행하였다. 그러

나 내이, 청각 신경 출력 모두 채널 수의 변화에 따른 인식율의 차이는 매우 근소함을 나타내었다. 즉, 청각 모델의 채널 수의 증가가 인식율의 증진에 크게 영향을 주지 못하였다. 따라서, 여기서는 18채널 청각 모델 출력에 의한 인식 결과만을 살펴보기로 하겠다.

Table 2는 6명의 화자에 대한 화자 종속 실험 결과를 정리해 놓은 것이다. 이 때, 신경 출력은 특징 벡터를 청각 신경의 순응 현상이 나타나는 과도상태를 포함한 구간에서 구한 경우이다. Table 2에서 각 특징 매개변수에 의한 인식율은 화자마다 약간의 차이는 있지만 대체로 청각 모델 출력에 의한 인식율이 LPC 켑스트럼 계수에 의한 인식율에 비해 높음을 알 수 있다. 좀더 자세히 살펴보기 위하여 Table 2에서 특징 매개변수별로 여섯 화자에 대한 평균 인식율을 구해보기로 한다. 분석 프레임의 길이가 피치 asychronous의 경우는 내이 출력 98.7%, 신경 출력 100%, LPC 켑스트럼 계수 96.7%이고, 피치 synchronous인 경우는 내이 출력 97.3%, 신경 출력 98.7%, LPC 켑스트럼 계수 82.7%이다. 따라서, 청각 모델의 출력이 내이, 신경 출력 모두 LPC 켑스트럼 계수에 비해 인식율이 높음을 알 수 있다. 특히, 분석구간의 길이를 짧게 한 경우, 청각 모델의 출력이 LPC 켑스트럼 계수에 비해 인식율이 평균 15%정도 높음을 알 수 있고, 내이, 청각 신경 출력 모두 분석구간의 길이가 짧더라도 인식율이 크게 저하되지 않음을 알 수 있다. 이는 청각 모델의 출력이 분석구간의 길이가 짧은 경우에도 모음 식별을 위한 특징을 잘 반영해주고 있음을 나타낸다.

오인식의 경우는 6명의 화자에 대하여 약간씩 차이는 있지만, 일반적으로 청각 모델 출력에 의한 인식의 경우 /오/와 /우/ 양 모음 사이에서 많이 일어났다. 이유는 Fig. 7과 Fig. 8에서 알 수 있듯이 /오/와 /우/의 특징 패턴이 유사하기 때문인 것으로 생각된다. 이보다 약간은 덜하지만 LPC 켑스트럼 계수의 경우도 양 모음 사이에 오인식이 많이 일어났으며 화자에 따라서는 /아/와 /에/ 사이에서도 약간의 오인식이 발생하였다. 실제, 모음 /오/와 /우/는 원 신호에 대한 파워 스펙트럼을 살펴보면 서로 유사함을 알 수 있다.

Table 3은 화자 독립 실험 결과를 정리해 놓은 것이다. 화자 SJS의 데이터는 모든 특징 매개변수에 대하여 인식율이 다른 5명의 화자의 데이터에 비해 현저하게 떨어져 Table 3의 전체 인식율 계산에서 제외하였다.

Table 3 Speaker independent recognition results for 5 Korean vowels

특 징 매개변수	프레임 길이	전 체 인식율(%)
내이출력 (18 CH)	async	96.0
	sync	96.0
신경출력(1) (18 CH)	async	94.4
	sync	95.2
신경출력(2) (18 CH)	async	81.6
	sync	86.4
LPC* 켑스트럼 계 수	async	92.0
	sync	81.6

*LPC 켑스트럼 계수의 수 ; 28

Table 3에서 신경 출력(1)은 특징 벡터를 청각 신경의 순응 현상이 나타나는 과도상태를 포함한 구간에서 구한 경우이고, 신경 출력(2)는 특징 벡터를 신경의 정상 상태 출력으로부터 구한 경우이다. 이 Table 3에서 신경 출력(2)의 경우를 제외하고 특징 매개변수에 따른 인식 결과를 살펴보면, 특징 추출을 위한 분석 구간을 피치 길이의 수배로 할 때(async) 및 피치 길이로 제한할 때(sync) 모두 청각 모델 출력에 의한 인식율이 LPC 켑스트럼 계수에 의한 인식율보다 높음을 알 수 있다. 특히, 분석 길이가 짧은 경우에는 14% 정도 높음을 알 수 있다. 이러한 결과는 화자 종속 실험 결과와 거의 일치함을 보여 준다. 또한, Table 3에서 신경 출력(1)과 신경 출력(2)의 인식 결과를 비교해 보면, 신경의 순응 현상이 나타나는 과도 상태를 고려한 신경 출력(1)의 경우가 async, sync 모두 인식율이 크게 높음을 알 수 있다. 따라서, 신경의 순응 현상이 모음 신호의 식별에 보다 유리한 특성을 제공한다고 볼 수 있다.

Table 2와 Table 3을 비교해 보면 화자 독립 실험은 화자 종속 실험에 비해 인식율이 특징 매개변수의 종류에 따라 1.1~5.6% 낮음을 알 수 있다. 이는 같은 종류의 모음 신호라 하더라도 화자마다 그 특성이 다르므로 인식 실험에 사용된 템플레이트가 각 화자의 모음별 음향 특성을 고루 반영하지 못하고 있기 때문이라고 여겨진다. 또, 화자 독립 실험의 경우 화자별 인식율이 화자

마다 차이가 크게 나타났다. 그러나 Table 2의 화자 종속 실험 결과를 살펴보면 화자별 인식율의 차이가 크지 않음을 알 수 있다. 이와같은 결과도 동일한 이유라고 볼 수 있다. 따라서, 화자 독립 실험에서는 인식율의 증진 및 화자별 인식율의 차이를 줄이기 위해서 템플레이트의 생성에 주의를 기울일 필요가 있다.

화자 독립 실험에서 오인식은 화자 종속 실험과 마찬가지로 /오/와 /우/ 양 모음 사이에서 많이 일어났다. 이러한 결과는 /오/와 /우/의 특징 패턴이 유사하기 때문이라고 볼 수 있다.

5. 결 론

본 논문에서는 말초 청각 계통의 음성 처리 특성을 자동 음성 인식 장치의 설계에 활용할 수 있는지에 대한 가능성을 살펴보기 위하여, 청각 모델을 이용하여 한국어 단모음에 대한 인식 실험을 수행하였다. 또한, 인식 결과의 객관적인 비교를 위하여, 동일한 음성 데이터에 대하여 LPC 켑스트럼 계수를 추출하여 인식 실험을 하였다. 얻어진 결과는 다음과 같다.

1) 모음 입력에 대한 청각 모델의 시간 축상의 출력 패턴 및 주파수 축상의 음답 패턴은 내이 및 청각 신경의 응답에 모음 신호의 특징이 잘 반영되고 있음을 보여준다.

2) 모음 입력에 대한 청각 신경의 평균 발화율 및 내이 응답의 평균 대수 에너지는 원 모음 신호의 대수 파워 스펙트럼과 유사한 형태를 갖는다.

3) 한국어 모음에 대한 인식 실험 결과는 청각 모델 출력에 의한 인식율이 화자 종속 실험, 화자 독립 실험 모두 LPC 켑스트럼 계수에 의한 인식율보다 높게 나타났다.

4) 청각 모델의 출력은 분석구간이 짧은 경우에도 모음 식별을 위한 특징을 잘 반영해 준다.

5) 청각 신경의 순응 현상은 음성 신호의 식별에 유리한 특성을 제공한다.

참 고 문 헌

1) Békésy, G.V., *"Experiments in hearing", Robert E. Krieger Publishing Company, Huntington, New York, pp. 403-429, 1960.*

2) Flanagan, J.L., *"Speech analysis, synthesis and perception", Springer Verlag, pp. 109-112, 1972.*

3) Rhode, W.S. and Robles, L., *"Evidence from Mossbauer experiments for nonlinear vibration in cochlea", JASA 55, pp. 588-596, 1973.*

4) Wilson, J.P. and Johnstone, J.R., *"Basilar membrane and middle ear vibration in guinea pig measured by capacitive probe", JASA 57(3), pp. 705-723, 1975.*

5) Schroeder, M.R. and Hall, J.L., *"Model for mechanical neural transduction in the auditory receptor", JASA 55, pp. 1055-1060, 1974.*

6) Young, E.D. and Sachs, M.B., *"Representation of steady state vowels in the temporal aspects of the discharge patterns of population of auditory nerve fibers", JASA 66(5), pp. 1381-1403, 1979.*

7) Allen, J.B., *"Cochlea modelling", IEEE ASSP Magazine, pp. 3-29, Jan., 1985.*

8) Seneff, S., *"Pitch and spectral analysis of speech based on auditory synchrony model", Ph. D. thesis, M.I.T. Univ., 1985.*

9) Zwicker, E., *"Subdivision of the audible frequency range into critical bands", JASA 33, pp. 248-249, 1961.*

10) Carlson, R. and Granstrom, B., *"Towards on auditory spectrograph", in the Representation of Speech in the Peripheral Auditory System, edited by Carlson, R. and Granstrom, B., Elsevier Biomedical Press, New York, pp. 109-114, 1982.*

11) Monro, D.M., *"Computer modelling of the peripheral mechanical response of the auditory system", in Auditory Investigation: the scientific and technological basis, edited by H.A. Beagley, Clarendon Press, pp. 431-450, 1979.*

12) Zwicker, E. and Terhardt, E., *"Analytical expressions for critical band rate and critical bandwidth as a function of frequency", JASA 68(5), pp. 1523-1525, 1980.*

13) Saito, S. and Nakata, K., *"Fundamentals of speech signal processing", Academic Press, 1985.*