

## 청각 보철용 음성신호 처리기의 설계

박상희\* · 최두일\* · 백승화\*\*

= Abstract =

### Design of the Speech Signal Processores for Cochlear Prosthesis

Sang-Hui Park\*, Doo-Il Choi\* and Seung-Wha Beack\*\*

Two types of the speech signal processores (SSP) for the cochlear a prosthesis are designed. One is designed using the cochlear model and the other is designed using the information (formant, pitch, intensity) extraction method. For these, some cochlear model and acoustic information extraction method are proposed. The result shows the SSP of the cochlear model type contain more acoustic cues than that of information extraction type. On the other hand, stimulus signal is clear and algorithm is simple in the SSP of the information extraction type.

### 1. 서 론

감각성 난청(sensory impaired)으로 인한 청각 장애자들은 보청기를 사용해도 소리를 감지할 수 없어 독순술과 수화에 의존하여 대화를 할 수 밖에 없다. 하지만 1960년 Békésy<sup>[1]</sup>가 귀의 생리학적 특성을 연구한 이후 청각 생리학이 꾸준히 발전하였으며, 1966년 Simmons<sup>[2]</sup>가 청각 신경을 전기적으로 자극하여 감각을 불러일으킬 수 있다는 가능성을 연구한 이후 정상적인 귀의 기능을 대신

할 수 있는 청각 보철에 관한 연구가 진행되어 왔다. 현재까지 개발된 청각 보철 기기는 House/3M의 single channel system, Vienna-Hochmain-3M의 single channel system, Melbourne-clark-Nucleus의 multi channel system 등이 있으며, 이중 비교적 성능이 우수한 Melbourne-clark-Nucleus의 청각 보철을 이식한 청각 장애자는 환경을 뿐 아니라 어느 정도 음성에 대한 인식율도 높게 나타나고 있다.

그러나, 이 보철 기기는 청각 계통의 생리학적 특성들을 충분히 고려하지 않은 채, 음성 신호의 포먼트를 추출하여 자극 신호를 가하는 방식이어서 특별한 포먼트가 없는 마찰음이나 파열음등의 인식율이 떨어지며, 청각 신경의 발화 패턴도 정상 인과는 큰 차이를 보이는등 여러가지 문제점을 갖

(접수 : 1991년 11월 15일)

\*연세대학교 공과대학 전기공학과  
\*\*명지대학교 공과대학 제어체측과  
\*Dept. of Electrical Eng. Yonsei University  
\*\*Dept. of Control and Instrumentation Engineering,  
Myongji University

고 있다<sup>[3]</sup>.

따라서, 본 연구에서는 청각 계통의 생리학적 특성을 보다 만족하고, 청각 신경의 발화 패턴이 정상인과 유사할 수 있는 청각 보철용 음성 신호 처리기를 청각 모델을 이용하여 설계하고자 한다.

아울러, 객관적인 비교를 위하여 음성신호의 정보를 추출하는 방식의 음성신호 처리기를 설계한다.

## 2. 청각모델 방식의 음성 신호 처리기

일반적으로 청각 보철은 외부 음성 신호 처리기부, 연결부 및 전극부로 이루어진다. 본 연구에서는 22channel, bipolar, scalar tympany 전극을 사용하는 것으로 가정하며 이에 맞는 외부 음성 신호 처리기를 설계한다. 또한, 청각 계통의 생리학적 특성을 충분히 반영할 수 있도록 전체 시스템을 그림1과 같이 청각 모델을 이용하여 설계한다.

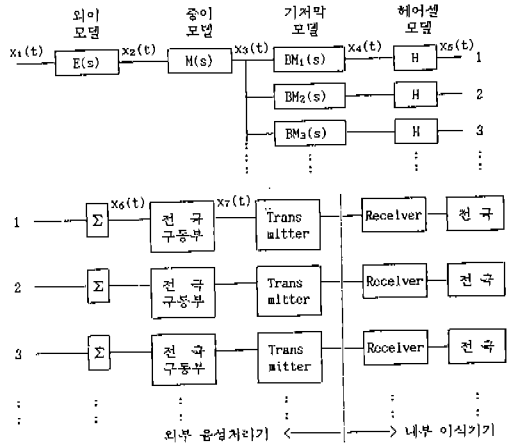
청각계통의 각단에 대한 모델을 생리학적 측정 연구결과들로부터 생성되었으며 외이모델, 기저막 모델, 헤어셀/신경모델 및 상호억제 모델은 각각 식 (1), (2), (3), (4)~(8), (9)와 같다.

이 중, 외이/중이 모델과 헤어셀/신경 모델은 각각 Monro<sup>[5]</sup>, Meddis<sup>[6]</sup>가 제안한 것을 이용하였으며, 기저막 모델과 상호 억제 모델은 청각계통의 특성을 보다 잘 반영하도록 본 연구에서 제안하였다.

각 채널의 중심 주파수의 간격은 [Bark] 단위로 정하였는데, 이것은 근사적으로 기저막을 등간격으로 구분한 것과 같다. [Bark] 단위는 원래 정신물리학에서 임계 대역을 나타낼 때 사용하는 주파수 단위로 Zwicker<sup>[7]</sup>에 의해서 제안되었다. 1 Bark는 청각 계통의 정신 물리학적 동조 곡선에서 -3dB 점 간의 임계 대역폭을 나타낸다.

중심 주파수의 [Hz] 단위로의 환산은 Zwicker와 Terhardt<sup>[8]</sup>가 제시한 식을 이용하였다.

본 연구에서는 22채널 전극을 사용하는 것으로 전제하여 0.8Bark의 등간격으로 각 채널에서의 중심 주파수를 구하였다. 음성 신호의 대부분이 4kHz 이하의 주파수대에 존재하기 때문에 최고 중심 주파수를 4,234.1Hz로 정하였다. 그리고 각 채널에서의 시간 지연은 중심 주파수의 역수를 취하여 구하였고, 각 채널에서의 주파수 선택도는 살아 있는 동물로부터 측정된 최근의 데이터<sup>[7] [8]</sup>에 기초하여, 상승 기울기는 최저 10.24dB/octave로부터 최고 214.03dB/octave까지로 결정하였다.



- $x_1(t)$ ; 음압
- $x_2(t)$ ; 고막에서의 음압
- $x_3(t)$ ; 스테이피즈 변위
- $x_4(t)$ ; 기저막 변위
- $x_5(t)$ ; 헤어셀 발화율
- $x_6(t)$ ; 상호 억제 모델 출력
- $x_7(t)$ ; 최종 전기자극 신호

그림 1 청각모델을 이용한 청각보철의 블록 선도

Fig. 1 The block diagram of cochlear prosthesis system by auditory model

$$E(s) = \frac{X_2(s)}{X_1(s)} = \frac{1}{(s+1,000)^2 + (7.875\pi)^2}, \quad (1)$$

$$M(s) = \frac{X_3(s)}{X_2(s)} = \frac{1}{s+6,000\pi}, \quad (2)$$

$$BM_l(s) = \frac{X_l(s)}{X_3(s)} = k(\omega_0) \left[ \frac{s^2}{s^2 + \frac{\omega_0}{Q}s + \omega_0^2} \right] \left[ \frac{s^2}{s^2 + \frac{\omega_0}{Q}s + \omega_0^2} \right]^{-l} \quad (3)$$

여기서,  $l$  ; channel 번호

$s = \sigma + j\omega$  ; 복소 주파수

$k(\omega_0)$  ; 생리학적 측정치의 크기 특성

$Q$  ; 대역폭과 일치시키기 위한 이득 요소

$\omega_0$  ; 기저막의 한 위치 1에서의 특성

주파수

$Q_h, Q_l$  ; 주파수 선택 요소  
 $n, m$  = 필터 차수/2 ; 상승 및 감쇠  
 기울기 요소

$$k(t) = \frac{g[x_4(t)+A]}{x_4(t)+A+B}, \quad x_4(t)+A > 0$$

$$= 0, \quad x_4(t)+A \leq 0 \quad (4)$$

$$\frac{dq}{dt} = m[1-q(t)] + xw(t) - k(t)q(t) \quad (5)$$

$$\frac{dc}{dt} = k(t)q(t) - lc(t) - rc(t) \quad (6)$$

$$\frac{dw}{dt} = rc(t) - xw(t) \quad (7)$$

$$x_5(t) = hc(t) \quad dt \quad (8)$$

여기서,  $x_4(t)$  : 기저막 모델 출력  
 $x_5(t)$  : 신경의 발화 확률

$$x_{6n}(t) = x_{5n}(t) \left[ 1 - \sum_{i=1}^N w(i, n) x_{5i}(t) \right] \quad (9)$$

여기서,  $x_{6n}(t)$  : n번째 channel의 출력  
 $x_{5n}(t)$  : 헤어셀/신경모델의 n번째  
 channel의 출력  
 $w(i, n)$  : 연결 가중치

이러한 선택도를 구하기 위하여 고역 통과 필터 차수  $h=1$ 로 하강 기울기 요소  $Q_h=1.00$ 으로 고정하고, 저역 통과 필터 차수 1과 상승 기울기 요소  $Q_l$ 를 실험적으로 결정하였다. 이러한 실험을 통하여 구하여진 기저막 모델의 각 채널의 필터 변수는 표 1과 같다.

음성신호 처리기의 마지막 단계로서, 청각 모델의 최종 출력인 발화율을 전기자극 신호로 변환해 주어야 한다. 전기자극 신호는 펄스파의 형태를 갖는 것이 효율적이다. 발화율에 대한 coding은 매우 어려운 문제이다. 왜냐하면 펄스의 주기가 조금만 변하여도 신경의 발화율은 매우 크게 변하며, 또 각각의 전극마다 청각신경에 대한 민감도가 다를 수 있기 때문이다.<sup>[9]</sup> 따라서 이는 향후 청각 보철 이식자를 대상으로 각각의 전극 channel마다 펄스의 크기에 따른 반응을 토대로 펄스의 크기와 발화율의 함수관계를 look-up table로 만들어, 모델

표 1 각 채널의 필터 변수

Table 1 Filter parameters for each channel

CF (Hz)	h	l	$Q_h$	$Q_l$	상승기울기 (dB/octave)	하강기울기 (dB/octave)	시간지연 (msec)
81.1	1	1	1.00	1.00	1.24	10.24	12.330
162.8	1	1	1.03	1.00	10.41	10.43	6.143
245.7	1	1	1.05	1.00	10.53	10.56	4.070
330.6	1	2	1.25	1.00	10.74	22.90	3.025
418.1	1	2	1.27	1.00	10.83	23.06	2.392
509.2	1	2	1.29	1.00	10.92	23.24	1.964
604.7	1	2	1.31	1.00	11.01	23.44	1.654
705.7	1	3	1.54	1.00	11.17	36.07	1.417
813.7	1	3	1.57	1.00	11.26	36.42	1.229
930.2	1	3	1.59	1.00	11.29	36.79	1.075
1016.3	1	3	1.62	1.00	11.37	37.16	0.984
1157.1	1	5	2.19	1.00	11.50	63.37	0.864
1317.4	1	5	2.28	1.00	11.61	64.62	0.759
1499.9	1	5	2.39	1.00	11.71	66.23	0.667
1707.6	1	5	2.53	1.00	11.80	68.34	0.586
1944.2	1	6	3.23	1.00	12.01	85.55	0.514
2213.5	1	6	3.58	1.00	12.12	90.13	0.452
2520.0	1	6	4.08	1.00	12.22	96.48	0.397
2869.1	1	6	4.81	1.00	12.30	105.54	0.349
3266.5	1	7	7.75	1.00	12.52	139.09	0.306
3719.0	1	7	10.78	1.00	12.61	164.88	0.269
4234.1	1	7	16.81	1.00	12.70	214.03	0.236

출력에 대한 펄스의 크기를 결정하는 것이 효율적일 것이다. 본 연구에서는 pulse의 크기와 신경의 발화율은 선형적인 관계를 갖는다는 것으로 가정하여 loudness에 대한 coding을 하였다.

### 3. 정보추출 방식의 음성신호 처리기

음성신호를 25.6[msec]의 간격으로 Hamming Window를 취하고, 각 frame에 대한 정보를 추출하여 자극신호로 이용하였다. 정보추출 방식의 전체 시스템은 그림 2와 같다.

유/무성음 판별 및 Pitch추출은 단구간 자기상관 함수를 3단계 클리핑 한후, 침두치를 찾아내어 R(0)의 30%에 해당하는 문턱치와 비교하여 결정하는 방식을 채택하였다. 그림 3은 유/무성음 판별

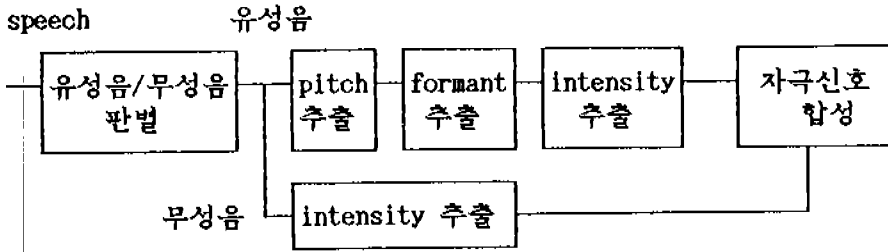


그림 2 정보추출 방식을 이용한 청각보철이 블럭 선도

Fig. 2 Block diagram of auditory prosthesis system by information extraction method

및 Pitch추출 블럭선도이다. Intensity는 단구간 에너지로부터 추출하였으며, 블럭선도는 그림 4와 같다. Formant는 LPC분석으로부터 추출된 선형 예측계수를 통하여 얻은 대수 스펙트럼에서 첨두치를 추출하는 방식을 채택하였다. 스펙트럼의 1차 미분 및 2차미분을 이용하여 3개의 formant 주파수를 얻었다.<sup>[10]</sup> 그림5는 formant주파수 추출 알고리즘 블럭선도이다.

이와같이 추출된 음성신호의 정보로부터 전기 자극신호의 합성은 다음과 같이 방식을 이용하였다.

- (1)전기 자극은 펄스파의 형태로 가한다.
- (2)무성음일 경우는 채널을 random하게 선택하여 자극 신호를 가하되, 자극 신호의 크기는 원 음성신호의 intensity에 비례하도록 한다.
- (3)유성음일 경우는 3개의 formant 주파수를 추출하고 채널에 할당되어 있는 특성 주파수와 가장 유사한 채널의 3개만 선택하여 추출된 pitch의 주기로 자극 신호를 가한다. 이 때, 각 채널에 가해지는 자극 신호의 크기는 원 음성 신호의 intensity에 비례하도록 한다.

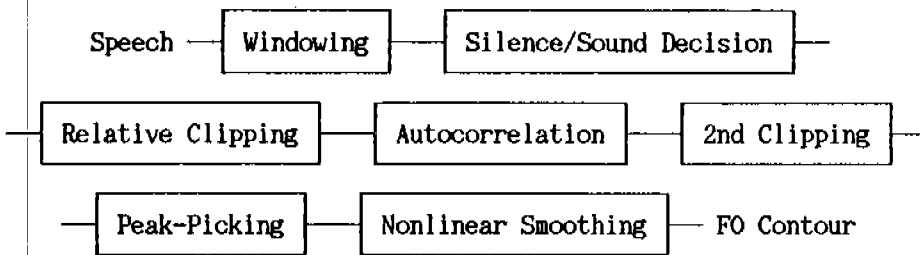


그림 3 기본 주파수 추출 및 유성음/무성음 분류 블럭선도

Fig. 3 Block diagram of extracting FO and classifying voiced/unvoiced

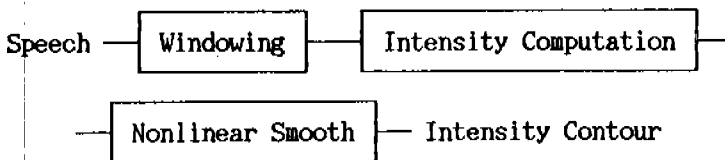


그림 4 Intensity 추출 알고리즘 블럭선도

Fig. 4 Block diagram of intensity extraction algorithm

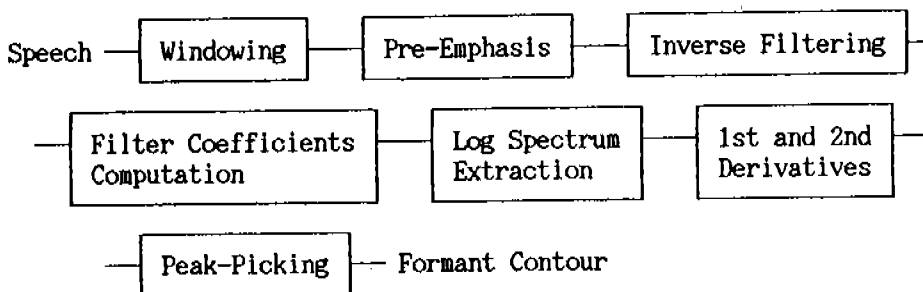


그림 5 Formant 주파수 추출 알고리즘 블록선도

Fig. 5 Block diagram of formant frequency extraction algorithm

#### 4. 결 과

##### 4.1 청각 모델 방식

청각 계통의 외이, 중이 및 기저막에 대한 모델 방정식 (1)-(3)을 극점-영점 사양(pole-zero mapping)을 사용하여 이산 전달 함수의 식으로 변환한 후 컴퓨터 시뮬레이션을 통하여 얻어진 주파

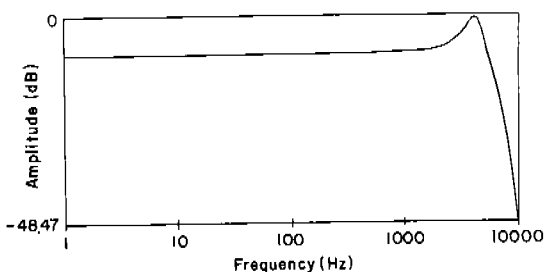


그림 6 외이 모델의 주파수 특성

Fig. 6 Frequency characteristics of the outer ear model

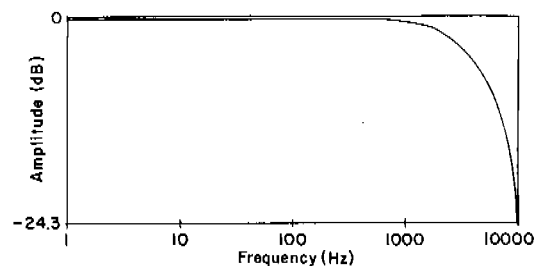
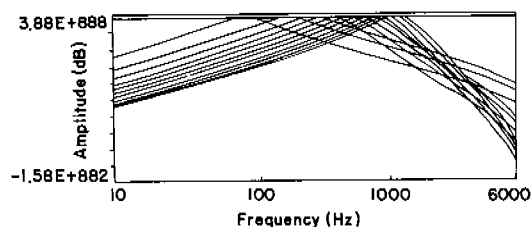


그림 7 중이 모델의 주파수 특성

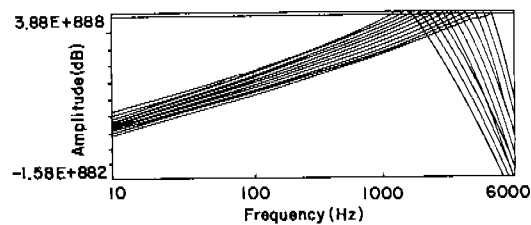
Fig. 7 Frequency characteristics of the middle ear model

수 특성은 각각 그림 6, 그림 7 및 그림 8과 같다. 각 모델은 생리학적 측정치와 부합하고 있으며, 특히 제안된 기저막 모델은 기저막의 위치에 따른 가변적인 주파수 선택 특성을 만족하고 있다.

헤어셀 모델 방정식 (4)-(8)을 차분 방정식으로 변환한 후 컴퓨터 시뮬레이션을 통하여 얻어진 헤어셀 모델의 입출력 특성은 그림 9와 같다. 헤어셀의 생리학적 특성에 부합하도록 실험적으로 얻어진 모델의 상수값들은  $g=16600$ ,  $A=12$ ,  $B=7000$ ,  $m=20$ ,  $x=1000$ ,  $l=500$ ,  $r=12500$ 이다. 그림



(a) 채널 1 ~ 11



(b) 채널 12 ~ 22

그림 8 기저막 모델의 각 채널별 주파수 특성

Fig. 8 Frequency characteristics of each channel of the basilar membrane model

10은 1KHz 순음 입력에 대한 헤어셀 모델의 출력  
을 나타내고 있다.

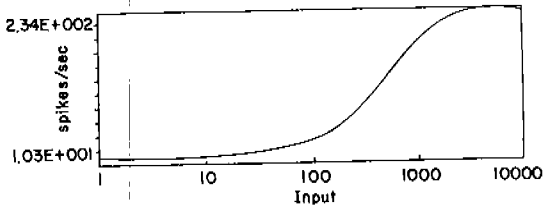


그림 9 헤어셀 모델의 입출력 특성  
Fig. 9 Input-output characteristics of haircell model

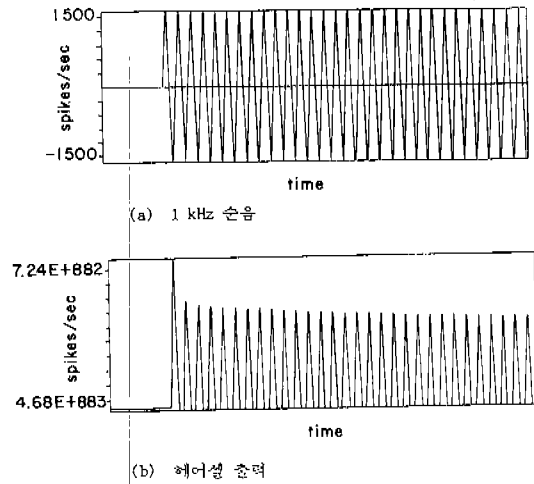


그림 10 1KHz순음에 대한 헤어셀 모델 출력  
Fig. 10 Output of haircell model for 1KHz pure tone

그림 11은 한국어 음성신호 /바/에 대하여 청각 모델 방식의 SSP에서 출력된 전기 자극 신호를 나타내고 있다. 전기 자극 신호는 각 채널에 매 1 [msec]마다 가해지도록 하였다. 그림 12는 본 연구의 SSP를 검증해 보기 위하여, Carney와 Geisler<sup>[11]</sup>가 영어 음성 /ba/, /bu/, /bi/에 대하여 고양이와 청각 신경으로부터 실측한 neurogram과 한국어 /바/, /부/, /비/에 대한 SSP의 출력력을 비교한 것이다. 주파수 스케일이 다르기 때문에 쉽게 비교하기는 어려우나 formant 주파수가 비슷한 주파수대역에 분포함을 보인다.

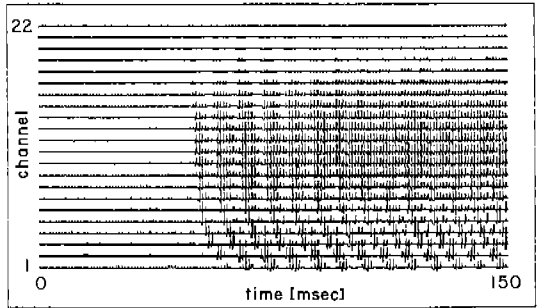


그림 11 한국어 음절 /바/에 대한 전기 자극 신호 (청각 모델 방식)  
Fig. 11 Electrical stimulus signals for Korean syllable (auditory model type)

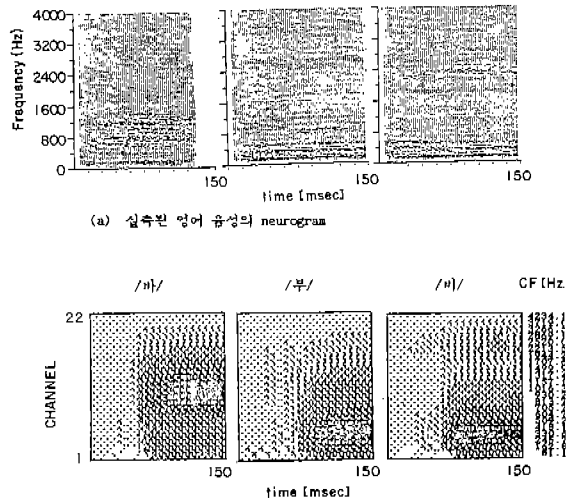


그림 12 SSP출력 검증(청각 모델 방식)  
Fig. 12 Testifying the output of SSP (auditory model type)

#### 4.2 정보 추출 방식

음성 신호를 각각의 단구간에 대해 25.6[msec]의 해밍 윈도우(hamming window)를 취한 후, 4절에서 언급한 방법으로 pitch, intensity 및 formant 추출하였다. 그림 13은 입력 파형으로 사용한 음절 /바/이며, 그림 14는 추출된 pitch의 변화를 나타낸 것이다. 여기서, pitch가 0인 부분들은 묵음 또는 무성음 구간을 나타낸다. 그림 15는 추출된



그림 13 입력 파형 음절 /바/  
Fig. 13 Input waveform of syllable /ba/

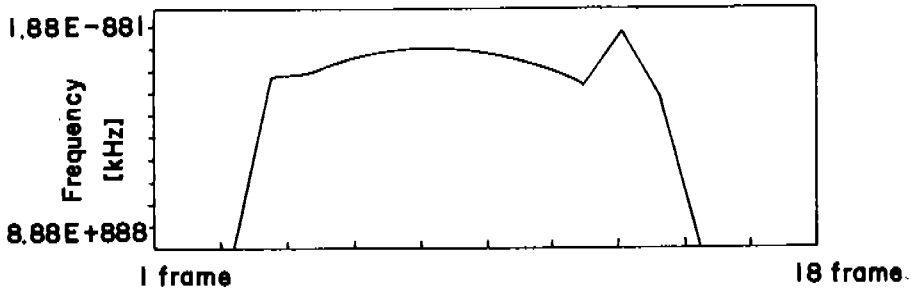


그림 14 음절 /바/의 pitch 변화  
Fig. 14 Pitch trajectory of syllable /ba/

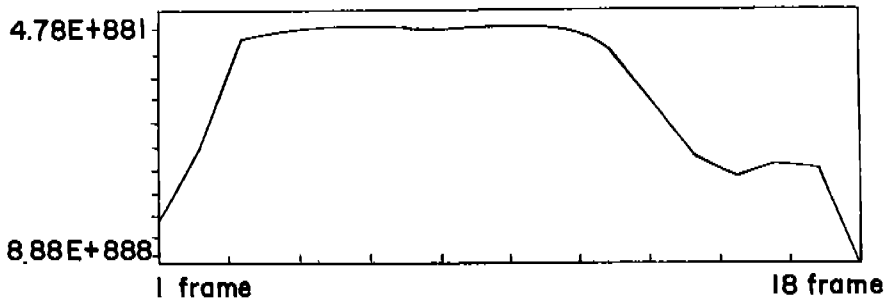


그림 15 음절 /바/의 intensity 변화  
Fig. 15 Intensity trajectory of syllable /ba/

intensity의 변화를 나타낸다.

Formant 주파수를 추출하기 위하여 음성 신호의 단구간 25.6[msec]에서 LPC분석법을 통한 대수 스펙트럼을 구한 후, 1차 및 2차 미분을 이용하여 formant 주파수를 얻었다. 그림 16 (a)는 음절 /바/에 대한 전 프레임의 대수 스펙트럼이고, (b)는 10번째 프레임의 대수 스펙트럼을 나타낸다. 표 2는 1차 및 2차 미분 스펙트럼에서 추출한 10번째

표 2 음절 /바/의 10번째 프레임의 formant 주파수

Table 2 Formant frequency of 10th frame of syllable /ba/

	Formant Frequency [Hz]
F1	859.38
F2	1210.94
F3	2773.44

프레임의 formant 주파수를 나타낸 것이다.

그림 17은 음성 신호에서 추출된 정보로부터 전기 자극 신호를 추출하는 알고리즘에 의하여 추출한 음성 신호 /바/에 대한 전기 자극 신호를 나타낸다. 그림에서 0~70[msec] 구간은 /b/에 해당하며, 랜덤한 신호가 intensity와 비례하여 가해지고 있다. 한편, 70~150[msec] 구간은 유성음 /a/에 해당하며, formant 주파수와 가장 가까운 3개의 채널에 자극 신호가 intensity와 비례하여 가해지고 있다.

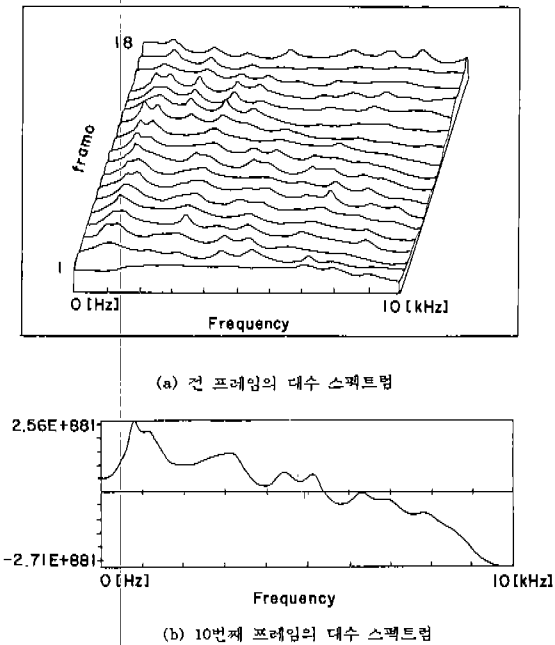


그림 16 음절 /바/의 대한 대수 스펙트럼  
Fig. 16 Log spectrum of syllable /ba/

### 5. 결 론

본 연구에서 얻어진 결과는 다음과 같이 그 특징으로 요약할 수 있다. 정보추출 방식의 신호 처리기의 특징은

- (가) 자극 신호 추출 알고리즘이 간단하여 음성 신호 처리기의 구현이 비교적 용이하고,
- (나) 유성음인 경우 자극 신호가 formant에 따라 선명하게 구분되며, 자극 신호를 발생하는 채널의 수가 한정되어 있기 때문에 채널간

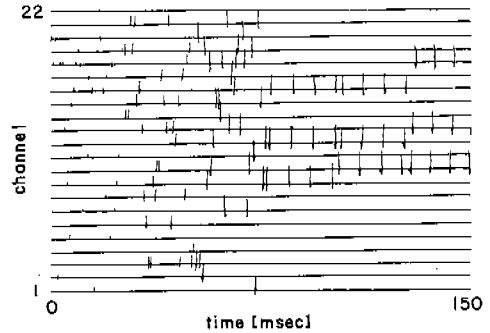


그림 17 음절 /바/의 대한 전기 자극 신호  
Fig. 17 Electrical stimulus signals for syllable /ba/

의 전류 확산 및 간섭 효과를 줄일 수 있다. 청각모델 방식의 신호 처리기의 특징은

- (가) 헤어셀 모델이 청각 계통의 중요한 생리학적 특성인 반파 정류, threshold, 포화 및 순응 현상을 만족하였다.
- (나) 음성 신호를 입력으로 한 청각 모델의 출력은 pitch를 지각하는데 있어서 정신 물리학적 특성인 temporal이론과 place이론을 모두 만족시켜 주었으며, 실측된 neurogram과 유사하였다.

정보 추출 방식과 청각 모델 방식의 음성 신호 처리기 중 어느 것이 우수하다고 결론 짓기는 어렵다. 따라서 향후 인공귀 이식자 개개인의 반응율 토대로 어느 한가지 방식을 선택할 필요가 있다.

향후 본 연구에서 설계한 음성 신호 처리기는 전용 DSP 칩을 사용하여 실시간 처리가 가능하도록 구현하여야 하며, 코딩된 전기 자극 신호를 청각신경으로 송수신하는 interface에 관한 연구가 필요하다.

본 연구는 과학기술처에서 시행한 특정연구 개발사업의 연구결과임을 밝힙니다.

### 참고 문헌

- 1) G.V.Békésy : Experiment in hearing, Robert E. Krieger Publishing company, Huntington, New



- York, 1960.
- 2) F.B.Simmons, "Electrical stimulation of the auditory nerve in man", Arch Otolaryngol. Vol. 84, pp. 24-76, 1966.
  - 3) R.L.White et al., "Development of a portable speech for a multiple-channel auditory prosthesis", 7th Quart. Prog. Rep. 1985.
  - 4) W.Yost and D. Nielsen : Fundamentals of hearing, Rinehart and Winstone, 1985.
  - 5) D.M.Monro, "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.
  - 6) R. Meddis, "Simulation of mechanical to neural transduction in the auditory receptor", J. Acoust. Soc. Am. Vol 79, pp. 702-711, 1986.
  - 7) E. Zwicker, "Subdivision of the audible frequency range into critical bands", J. Acoust. Soc. Am., Vol. 33, pp.248-249. 1961.
  - 8) E. Zwicker and E. Terhardt, "Analytical expressions for critical band rate and critical bandwidth as a function of frequency", J. Acoust. Soc. Am. Vol. 68, pp. 1523-1525, 1980.
  - 9) S.M. Khanna and D.G.B. Leonard, "Basilar membrane tuning in the cat cochlea", Science Vol. 215, pp.305-306, 1982.
  - 10) L. Robles, M.A. Ruggero and N.C. Rich, "Basilar membrane mechanics at the chinchilla cochlea cochlea I. Input-output functions, tuning curves, and response phases", J. Acoust. Soc. Am., Vol. 80, pp. 1364-1374, 1986.
  - 11) R.L.White, "System design of a cochlear implant", IEEE Eng. Med. Biol. Magazine Vol. 6, pp. 42-46, June 1987.
  - 12) 송영수, 안상필, 이재혁, 윤태성, 박상희, "LPC Spectrum의 1, 2차 미분을 이용한 Formant추출", 신호처리 합동학술대회 제3권 제 1호, 1990.
  - 13) L.H. Carney and C.D. Geisler, "A temporal analysis of auditorynerve fiber responses to spoken stop consonant-vowel syllables", J. Acoust. Soc. Am., Vol. 79, pp.1896-1914, 1986.
  - 14) 백승화 : "청각동조 수학적 모델에서의 음성 신호에 대한 패턴 추출", 박사학위 논문, 연세대학교 대학원, 1987.
  - 15) 윤태성 : "청각 모델을 이용한 한국어 단음의 인식에 관한 연구", 박사학위 논문, 연세대학교 대학원, 1988.
  - 16) 김동혁, 최두일, 김영선, 백승화, 박상희, "청각 모델을 이용한 청각 보철용 음성 신호 처리기에 관한 연구", 신호처리 합동학술대회 제3권 제1호, 1990.
  - 17) Dooil Choi, Donghyeok Kim, Sanghui Park, Seunghwa Beack, Dongmin Woo, "Design of speech signal processor for the multi-channel auditory prosthesis using auditory model", 16th International Conference on Medical and Biological Engineering, Vol. 29, 1991.
  - 18) 최두일, 김동혁, 박상희, 백승화, "청각 장애자를 위한 청각 보철용 음성 신호 처리기의 설계", 대한의용생체공학 춘계 학술대회 논문집, 1991.