

청각 장애자를 위한 청각 보철용 음성신호 처리기의 설계

최두일*, 김동혁*, 박상희*, 백승화**
*연세대학교 전기공학과, ** 명지대학교 제어계측과

A Design of the Speech Signal Processor of Cochlear Prosthesis for the
Sensory Deaf

Choi, Dooil Kim, Donghyuk Park, Sanghui Beack Seunghwa
Dept. of Electrical Eng., Yonsei Univ. Dept. of Cont. and Inst. Eng., Myungji Univ.

Abstract

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

1. 서론

감각성 난청으로 인한 청각 장애자들은 보청기를 사용해도 소리를 감지할 수 없어 독순술과 수화에 의존하여 대화를 할 수 밖에 없다. 하지만 1960년 Bekesy가 귀의 생리학적 특성을 연구한 이후 청각 생리학이 꾸준히 발전하였으며, 1966년 Simmons가 청각 신경을 전기적으로 자극하여 감각을 불러 일으킬 수 있다는 가능성을 연구한 이후 정상적인 귀의 기능을 대신할 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 환경음 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 정상인과는 큰 차이를 보이는등 여러가지 문제점을 갖고 있다.

따라서, 본 연구에서는 청각 계통의 생리학적 특성을 보다 만족하고, place 이론과 temporal 이론을 동시에 만족하며, 청각 신경의 발화 패턴이 정상인과 유사할 수 있는 청각 보철용 음성 신호 처리기를 청각 모델을 이용하여 설계하고, 이를 위하여 새로운 청각 모델을 제시하고자 한다.

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

2. 말초 청각 계통의 생리학적 특성 및 정신 물리학적 특성

음파가 외이(outer ear)에 들어오면 유압에 의하여 고막(ear drum)이 진동하며, 이 진동은 중이(middle ear)를 거쳐 내이(inner ear)의 코클리어(cochlea)에 전달된다. 코클리어에 유입된 전달 신호는 기저막(basilar membrane)위를 진행하면서 입력 신호의 주파수 성분에 따라 각각 기저막의 한 점에서 최대 공진을 일으킨다. 기저막 위의 코르티 기관(the organ of Corti)에는 헤어셀(haircell)이라고 불리는 섬모 세포들이 분포되어 있고 이 헤어셀에는 청각 신경이 연결되어 있다. 이 헤어셀/청각 신경에서 기저막의 운동이 신경 발화 활동으로 변환되어 중추 청각 신경을 거쳐 뇌로 전달되게 된다. 한편 정신 물리학적 측면에서 음의 높낮이를 인지하는 이론은 place 이론과 temporal 이론이 있다. place 이론은 pitch 감각이 기저막의 진동하는 위치와 관련되고, temporal 이론은 음의 주기에 따른 신경의 발화 주기와 관련된다는 이론이다. 음성과 같은 복잡한 신호는 이 두 가지 이론 모두에 의해서 인지되며, 특히 저수파수의 신호는 temporal 이론이 잘 맞으며 고주파수의 신호는 place 이론이 잘 맞게 된다.

3. 청각모델 방식의 음성 신호 처리기의 구성

청각 보철은 외부 음성 신호 처리기부, 연결부 및 전극부로 이루어진다. 본연구에서는 22 channel, 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 그림 1과 같이 청각 모델을 이용하여 설계한다.

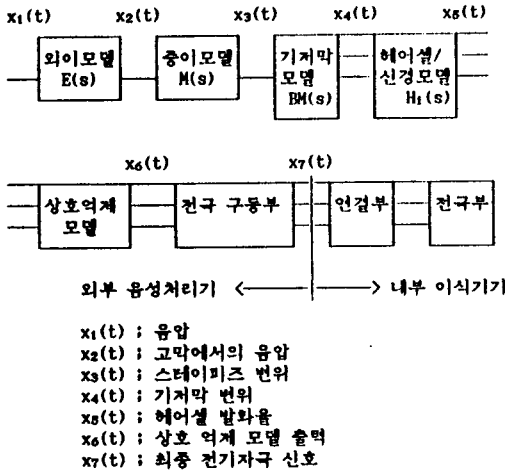


Fig.1 The block diagram of cochlear prosthesis system by auditory model

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

이 중, 외이/중이 모델과 헤어셀/신경 모델은 각각 Monro, Meddis가 제안한 것을 이용하였으며, 기저막 모델과 상호 억제 모델은 청각계통의 특성을 보다 잘 반영하도록 본 연구에서 제안하였다. 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. 따라서 이는 향후 청각 보철 이식자를 대상으로 각각의 전극 channel마다 펄스의 크기에 따른 반응을 토대로 펄스의 크기와 발화율의 함수관계를 look-up table로 만들어, 모델 출력에 대한 펄스의 크기를 결정하는 것이 효율적일 것이다. 본 연구에서는 pulse의 크기와 신경의 발화율은 선형적인 관계를 갖는다는 것으로 가정하여 loudness에 대한 coding을 하였다.

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

여기서, $X_1(s)$; $x_1(t)$ 에 대한 라플라스 변환
 $X_2(s)$; $x_2(t)$ 에 대한 라플라스 변환

$$M(s) = \frac{X_3(s)}{X_2(s)} = \frac{1}{s + 6000\pi} \quad (2)$$

여기서, $X_2(s)$; $x_2(t)$ 에 대한 라플라스 변환
 $X_3(s)$; $x_3(t)$ 에 대한 라플라스 변환

$$BM_1(s) = \frac{X_4(s)}{X_3(s)} = k(\omega_0) \left[\frac{s^2}{s^2 + \frac{\omega_0}{Q_h} s + \omega_0^2} \right]^n \left[\frac{\omega_0^2}{s^2 + \frac{\omega_0}{Q_1} s + \omega_0^2} \right]^m \quad (3)$$

여기서, l ; channel 번호

$X_3(s)$; $x_3(t)$ 에 대한 라플라스 변환

$X_4(s)$; $x_4(t)$ 에 대한 라플라스 변환

$s = \sigma + j\omega$; 복소 주파수

$k(\omega_0)$; 생리학적 측정치의 크기 특성과 일치시키기 위한 이득 요소

ω_0 ; 기저막의 한 위치 l에서의 특성 각 주파수

Q_h, Q_1 ; 주파수 선택 요소

n, m ; 필터 차수 / 2 ; 상승 및 감쇠 기울기 요소

$$k(t) = \begin{cases} \frac{g[x_4(t)+A]}{x_4(t)+A+B}, & x_4(t)+A > 0 \\ 0, & x_4(t)+A \leq 0 \end{cases} \quad (4)$$

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

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

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

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

여기서, $x_4(t)$: 기저막 모델 출력

$x_5(t)$: 신경의 발화 확률

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

여기서, $x_{6n}(t)$: n번째 channel의 출력

$x_{5n}(t)$: 헤어셀/신경모델의 n번째 channel의 출력

$w(l,n)$: 연결 가중치

4. 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.

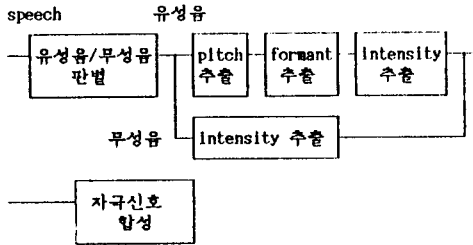


Fig.2. Block diagram of auditory prosthesis system by information extraction method

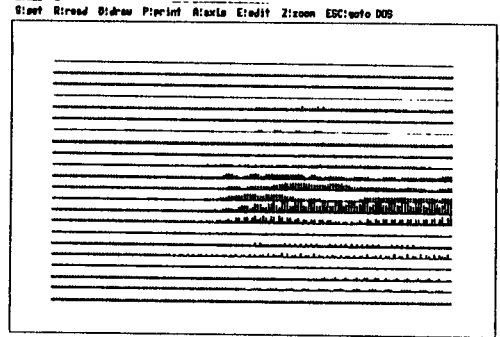
유/무성음 판별은 단구간 자기상관 함수를 3단계 클리핑 후, 첨두치를 찾아내어 R(0)의 30%에 해당하는 문턱치와 비교하여 결정하는 방식을 채택하였다. Pitch추출은 단구간 에너지로부터 추출하였다. Formant는 LPC분석으로부터 추출된 선형 예측계수를 통하여 얻은 대수 스펙트럼에서 첨두치를 추출하는 방식을 채택하였다. 스펙트럼의 1차미분 및 2차미분을 이용하여 3개의 formant주파수를 얻었다. 이와같이 추출된 음성신호의 정보로부터 전기자극신호의 합성은 다음과 같은 방식을 이용하였다.

1. 전기 자극은 펄스의 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 채널을 3개만 선택하여 추출된 pitch의 주기로 자극 신호를 가한다. 이 때, 각 채널에 가해지는 자극 신호의 크기는 원 음성 신호의 Intensity에 비례하도록 한다.

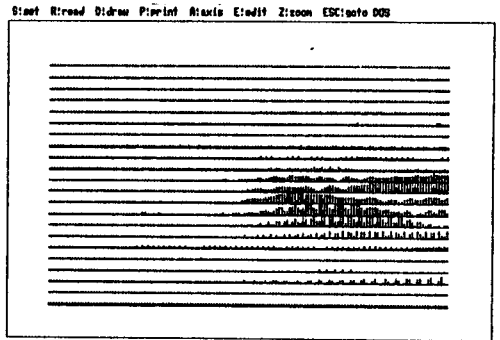
5. 결과

한국어 음성신호중 파열음+모음인 /가/, /다/, 및 /바/에 대하여 두가지 방식에 의하여 추출된 전기자극신호를 비교하여 보았다. 그림 3은 청각모델 방식에 의하여 추출된 신호이며, 그림 4는 정보추출 방식에 의하여 추출된 신호이다. 청각모델 방식의 음성신호처리기는 각 채널에 매 1 [msec] 마다 자극 신호를 가하도록 하였으며, 정보

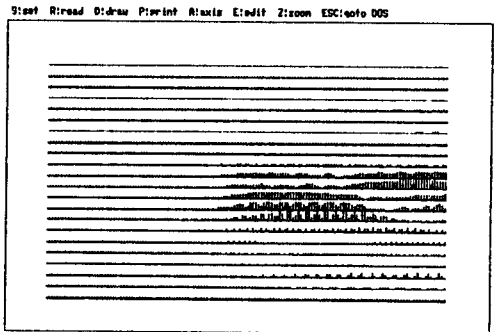
추출 방식의 음성신호 처리기는 4절에서 언급한 바와 같은 방법으로 자극신호를 가하도록 설계하였다.



(a) Electrical stimulus signal for /가/

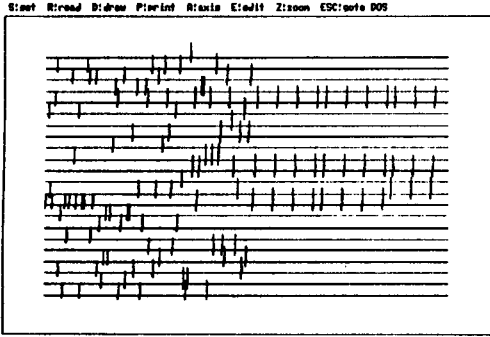


(b) Electrical stimulus signal for /다/

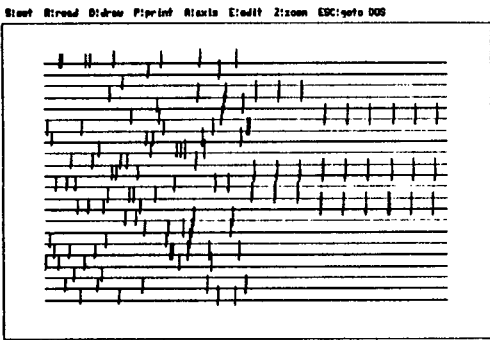


(c) Electrical stimulus signal for /바/

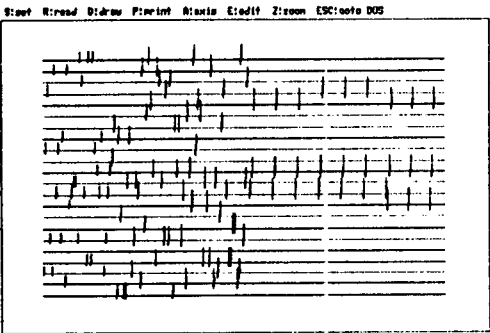
Fig. 3 Electrical stimulus signal extracted by auditory model method for Korean speech.



(a) Electrical stimulus signal for /가/



(b) Electrical stimulus signal for /다/



(c) Electrical stimulus signal for /바/

Fig. 4 Electrical stimulus signal extracted by information extraction method for Korean speech.

6. 결론

정보추출 방식의 신호 처리기의 특징은 다음과 같다.

- (가) 자극 신호 추출 알고리즘이 간단하여 음성 신호 처리기의 구현이 비교적 용이하다.
- (나) 유성음인 경우 자극 신호가 formant에 따라 선명하게 구분되며, 자극 신호를 발생하는 채널의 수가 한정되어 있기 때문에 채널간의 전류 확산 및 간섭 효과를 줄일 수 있다.
- (다) 이에 따라 청각보철 이식자의 적응 효과가 높으나 다양한 소리에 대한 지각 능력은 감소할 것이다.

청각모델 방식의 신호 처리기의 특징은 다음과 같다.

- (가) 본 연구에서 제안한 지저막 모델은 실제 청각 계통의 생리학적 특성을 고려하여 채널에 따른 가변적인 주파수 선택도 특성을 갖도록 설계되었다.
- (나) 헤어셀 모델은 청각 계통의 중요한 생리학적 특성인 반파 정류, threshold, 포화 및 순응 현상을 만족하였다.
- (다) 음성 신호를 입력으로 한 청각 모델의 출력은 pitch를 지각하는데 있어서 정신 물리학적 특성인 volley 이론과 place 이론을 모두 만족시켜 주었으며, 실측된 neurogram과 유사하였다.
- (라) 청각 모델 방식의 음성 신호 처리기는 전 channel에 자극 신호를 가하기 때문에 다양한 소리를 지각할 수 있을 것으로 기대된다.

7. 참고 문헌

- [1] G.V. Békésy: Experiment in hearing, Robert E. Krieger Publishing Company, Huntington, New York, 1960.
- [2] Clark GM: The University of Melbourne/Cochlear Corporation (Nucleus) program. Otolary Clin North Am 1986; 19: 329 - 354.
- [3] J. D. Markel and A. H. Gray Jr., Linear Prediction of speech, New York: Springer - Verlag, 1980.
- [4] 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.
- [5] R. Meddis, "Simulation of mechanical to neural transduction in the auditory receptor", J. Acoust. Soc. Am. Vol.79, pp 702 - 711, 1986.
- [6] E. Zwicker, "Subdivision of the audible frequency range into critical bands", J. Acoust. Soc. Am. Vol. 1.33, pp 248 - 249, 1961.