J. Biomed. Eng. Res : 271-279, 2007

Neural Spike Train Decoding에 기반한 인공와우 어음처리방식 성능평가

김두희, 김진호, 김경환

연세대학교 보건과학대학 의공학부

(Received January 17, 2007. Accepted March 7, 2007)

Performance Evaluation of Cochlear Implants Speech Processing Strategy Using Neural Spike Train Decoding

Doohee Kim, Jinho Kim, Kyung Hwan Kim

Department of Biomedical Engineering, College of Health Science, Yonsei University

Abstract

We suggest a novel method for the evaluation of cochlear implant (CI) speech processing strategy based on neural spike train decoding. From formant trajectories of input speech and auditory nerve responses responding to the electrical pulse trains generated from a specific CI speech processing strategy, optimal linear decoding filter was obtained, and used to estimate formant trajectory of incoming speech. Performance of a specific strategy is evaluated by comparing true and estimated formant trajectories. We compared a newly-developed strategy rooted from a closer mimicking of auditory periphery using nonlinear time-varying filter, with a conventional linear-filter-based strategy. It was shown that the formant trajectories could be estimated more exactly in the case of the nonlinear time-varying strategy. The superiority was more prominent when background noise level is high, and the spectral characteristic of the background noise was close to that of speech signals. This confirms the superiority observed from other evaluation methods, such as acoustic simulation and spectral analysis.

Key words : cochlear implants, speech processing strategy, neural spike train decoding

I. 서 론

세계보건기구에 의하면 청각장애는 55세 이상 노인의 삶의 질을 떨어뜨리는 5대 요소에 속하며, 전 세계적으로 약 7억 5천만 명 정도가 청각장애자로 분류된다 [1]. 특히 이중 2백만 명은 고도의 청각장애를 가진 이들로 청력회복을 위해서는 인공와우(cochlear implants) 시술의 대상자가 될 수 있다. 청각장애에는 전음성 청각장애(conductive hearing loss)와 감각신경성 청각장애(sensorineural hearing loss)가 있다[1,2]. 전자의 경우 단순외상, 귀경화증(otosclerosis) 등에 의해 소리 전달로의 장애가 생긴 것이 원인으로 단순히 소리를 증폭하여 주는 보청기의 사용으로도 청력회복이 가능하다. 후자는 잡음에 과도하게 노출되거나 고령, 수막염(meningitis) 등에 의해 와우각 내 유모세포 손실이나 신경통로 혹은 뇌의 청각담당 영역의 손상 등이 발생한 경우이다[3]. 감각신경성 청각장애의 경우 유모세포의 변형이 일어나지 않으므로 보청기 등의 청각보철장치를 이용하여 소리를 증폭시키더라도 청신경의 활동전위가 발생되지 않기 때문에 청각정보가 뇌로 전달되지 않는다. 그러므로 청신경에 직접 전기자극을 가하여 청력을 회복시키는 인공와우에 의한 청각회복의 수혜자가 될 수 있다 [4,5].

인공와우 장치는 청신경에 전기자극을 인가함으로써 청력을 회복시켜주는 장치로 마이크로폰, 신호처리부(어음처리기), 신호 송/수신부, 전극으로 구성된다. 소리가 마이크로폰을 통해 입력되면 신호처리부에서 지정된 어음처리방식에 의해 전기자극펄스 형태로 바뀌어 달팽이관 내에 삽입된 전극으로 전달된다[4,5]. 소리에 담긴 정보가 얼마나 충실하게 청신경계로 전달되는가는 전기자극펄스를 생성하는 어음처리방식에 달려있으므로 이에 대한 성능평가는 필수적이다[4-7,21]. 최종적으로는 인공와우 시술자를 대상으로 어음처리방식의 성능평가를 수행하는 것이 바람직하지만 비용, 시간, 대상이 사람이라는 점 등으로 인하여 많은 제한이 있다. 많은 연구에서 어음처리방식에 따른 청력회복의 정도를 acoustic simulation에 기반하여 평가, 예측하고 있으나 많은 불확실성이 존재한다[8].

본 연구는 과학기술부/한국과학재단 우수연구센터 육성 사업의 지원으로 수행되었음(R11-2000-075-01005-0).

Corresponding Author : 김경환
강원도 원주시 흥업면 매지리 234 연제대학교 보건과학대학 의공학부
Tel : +82-33-760-2364 / Fax : +82-33-760-1953
E-mail : khkim0604@yonsei.ac.kr

본 연구의 목적은 neural spike train decoding에 기반하여 어음처리방식의 성능평가를 수행하고 이 방법의 성능평가 수단으로써의 효용성을 알아보는 것이다. 외부로부터 신체 기관 등으로 가해진 자극은 활동전위와 같은 신경계의 응답의 형태로 표현되며 이 응답에는 자극에 대한 정보가 전달된다. 활동전위의 발생시점을 나타낸 spike train으로부터 외부자극 혹은 이 spike train이 전달하고 있는 정보를 복원할 수 있는데 이러한 방법을 neural spike train decoding이라한다[9,10]. 본 연구에서는 음성에 대하여 확률적 청신경모델로부터 multiple spike trains을 얻고 neural spike train decoding에 기반하여 외부자극인 입력음성의 포먼트 추정값을 얻고 이를 원래의 포먼트와 정량적으로 비교함으로써 사용된 어음처리방식의 유용성을 평가하는 방법을 제안한다. 이는 입력음성의 특성이 지정된 어음처리방식에 기반한 전기자극펄스에 의해 청신경의 응답으로 잘 전달될수록 우수한 어음처리방식이라 할 수 있으며 높은 수준의 청력회복을 기대할 수 있다는 가정에 기반하고 있다. 음성을 인식하는데 있어서 모음은 중요한 매우 중요한 역할을 한다. 음성 주파수성분의 정점값인 포먼트는 모음에 대한 매우 중요한 단서를 제공한다. 포먼트의 낮은 값부터 제1, 제2 포먼트(F1, F2)의 순서로 불리우는데 이 중 모음의 인식에는 제1, 제2 포먼트가 중요한 역할을 하며 실험에는 제1~제3포먼트(F1, F2, F3)를 어음처리방식의 성능평가의 지표로 사용하였다[11,12]. 본 연구에서는 서로 다른 어음처리방식 성능평가 방법에 의한 결과와 비교하여 어음처리방식 성능평가 도구로써 neural spike train decoding에 기반한 방법의 효용성을 알아보았다.

제안된 방법을 이용하여 가장 널리 사용되는 시불변-선형 필터뱅크 기반 어음처리방식과 말초청각계의 특성을 좀 더 면밀하게 모방한 시변-비선형 필터뱅크 기반 어음처리 방식을 비교하였다. 실험결과 시변-비선형 필터뱅크 기반 어음처리방식을 채택할 경우 시불변-선형 필터뱅크에 의한 방식보다 포먼트를 충실히 복원할 수 있었다. 시변-비선형필터뱅크 기반 어음처리방식은 달팽이관내 기저막 (basilar membrane)의 선형 및 비선형특성을 고려한 것이므로 기저막의 선형특성만을 고려한 어음처리방식에 비하여 잡음 하에서 음성정보전달특성을 향상시킬 수 있을 것으로 기대된다[20,21]. 두 어음처리방식에 대하여 제안된 neural spike train decoding에 기반한 성능평가를 수행하였다. 입력음성으로부터 추출한 포먼트와 지정된 어음처리방식에 기반한 전기자극펄스에 의해 자극된 확률적 청신경모델의 응답인 multiple spike trains로부터 최적선형 디코딩필터를 얻고 이를 이용하여 음성에 대한 청신경모델의 응답으로부터 포먼트의 추정값을 얻은 후 포먼트의 참값과 추정값간 유사성을 관찰함으로써 정량적으로 이루어졌다. 실험결과 시변-비선형 필터뱅크 기반 어음처리방식을 채택할 경우 시불변-선형 필터뱅크에 의한 방식보다 포먼트를 충실히 복원할 수 있었다. 특히 잡음레벨이 높은 경우와 음성과 유사한 스펙트럼 특성을 갖는 잡음을 첨가할 경우 시변-비선형 필터뱅크 기반 어음처리방식의 우수성이 현저히 나타났다. 이는 스펙트럼 분석, acoustic simulation 등 다른 어음처리방식 평가방법에서 보인 결과를 뒷받침한다[12].

Ⅱ. 이론 및 실험방법

A. Neural Spike Train Decoding

외부자극은 감각기(receptor) 및 이에 연결된 감각신경에 의해 신경세포의 활동전위(action potential) 형태로 변환된다. 활동전위의 발생시점은 spike train의 형태로 표현할 수 있는데 이로부터 외부자극에 대한 정보를 추출해 내는 과정을 neural spike train decoding이라 한다[9,10]. Neural spike train decoding은 주어진 spike train에 대한 학습(training)과 새로이 얻은 spike train으로부터 외부자극을 추정(estimation)하는 두 과정으로 이루어져 있다. 주어진 외부자극과 spike train 간의 매핑 즉, 디코딩 필터계수를 결정하는 과정을 학습이라고 하며 학습으로부터 얻은 디코딩 필터를 이용하여 spike train에 인코딩된 외부 자극을 복원하는 과정을 추정이라 한다. 디코딩 필터는 multiple spike trains로부터 일정 time bin 내의 spike 발생 개수를 계산한 발화율(firing rate)을 입력으로 하고 외부자극을 출력으로 한다. 학습과정은 주어진

$$f = [f_1(0) f_1(1) \cdots f^1(M-1) \cdots f_p(0) \cdots f_p(M-1)]^T \tag{1}$$

$$f = (R^T R)^{-1} (R^T s) \tag{2}$$

$$R = \begin{bmatrix} 1 & r_1(0) & r_1(1) & \cdots & r_1(M-1) & r_N(0) & \cdots & r_N(M-1) \\ 1 & r_1(1) & r_1(2) & \cdots & r_1(M) & r_N(1) & \cdots & r_N(M) \\ 1 & \cdots & \cdots & \cdots & & \cdots & & \cdots \\ 1 & r_1(L) & r_1(L+1) & \cdots & r_1(L+M-1) & r_N(L) & \cdots & r_N(L-M+1) \end{bmatrix} \tag{3}$$

$$s = [s(0) s(1) \cdots s(L-1)]^T \tag{4}$$

$$\hat{s}(i) = \sum_{p=1}^{N} \sum_{j=0}^{M-1} r_p(i-j) f_p(j) \tag{5}$$

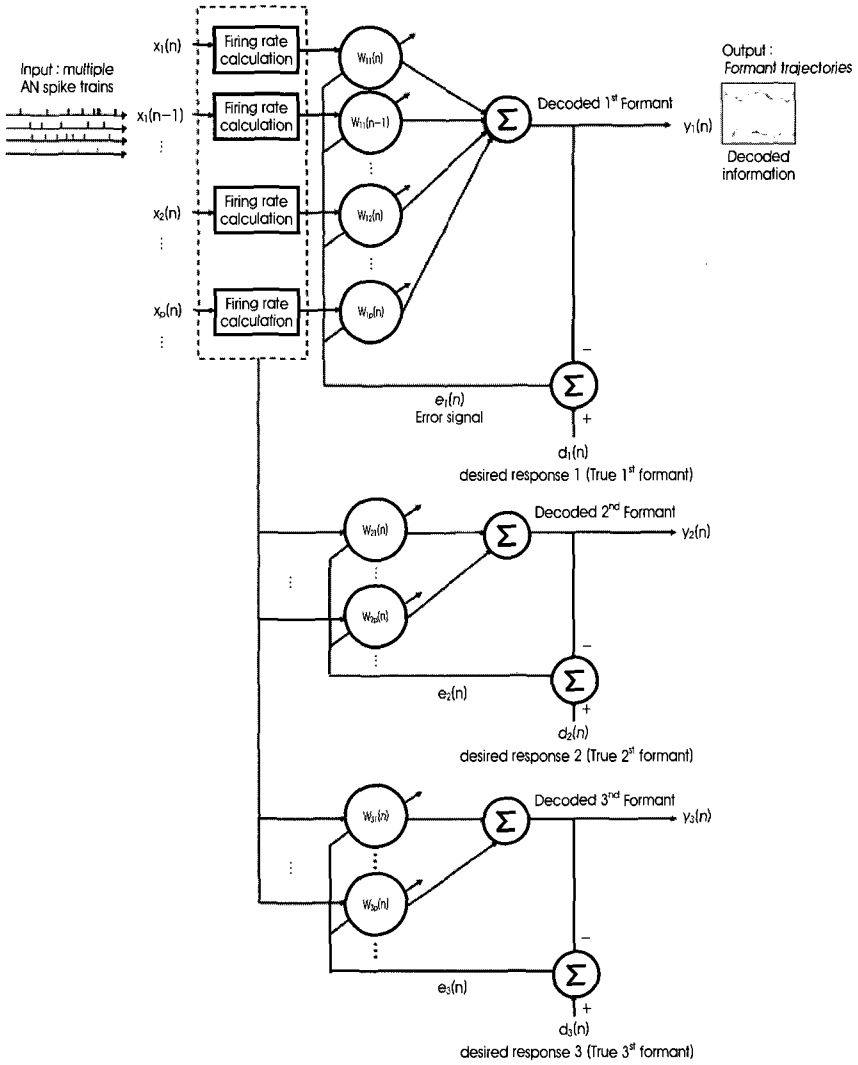

그림 1. 최적선형필터를 이용한 neural spike train decoding
Fig. 1. Neural spike train decoding using optimal linear filter

외부자극의 참값과 추정값 간의 평균 제곱 오차(mean squared error)를 최소화하도록 이루어지며 학습으로부터 얻어진 디코딩 필터를 이용하여 새로운 spike train으로부터 외부 자극을 추정한다. 디코딩 필터는 다중채널 유한임펄스 응답 필터 (multichannel finite impulse response filter) 형태로 구현되어 있으며 다음과 같이 입력자극과 발화율 정보를 기반으로 하여 식(2)에 표현된 바와 같이 필터 계수들을 구할 수 있다. 식(1)의 $f_p(j)$는 뉴런 p의 j번째 time bin에 대한 필터계수를 의미한다. R 및 s는 각각 발화율 및 입력자극에 대한 정보를 포함하고 있는 행렬로 식(3),(4)와 같이 구성된다. $r_N(i)$는 N번째 단일 뉴런이 i번째 time bin에서 갖는 평균발화율을 뜻하며 M은 디코딩 필터의 차수, N은 관찰된 단일 뉴런의 개수 그리고 L은 총 관찰시간 내 time bin의 개수이다. 학습 과정에서 얻은 디코딩 필터계수와 새로운 spike train을 이용한 외부자극의 추정과정은 식(5)와 같이 표현할 수 있다. 변수 $r_p(i)$는 뉴런 p의 i번째 time bin에 대한 발화율, $\hat{s}(i)$는 i번째 time bin에 대한 외부자극의 추정값을 의미한다.

본 연구에서는 외부자극으로 입력음성정보가 주어지고 특정 어음처리방식에 기반해 생성된 전기자극펄스를 입력으로 하는 청신경모델 응답에 인코딩된 음성정보를 복원하여 원 음성정보와 비교하게 된다. 그림 1은 최적선형필터를 이용한 neural spike train decoding에 의한 포먼트 추정과정을 구체적으로 설명해 주고 있다. 디코딩필터는 multiple spike train들로부터 일정한 time bin 내의 spike 발생 개수를 계산한 발화율(firing rate)을 입력으로 하고 외부로부터 청신경모델에 인가되는 자극 음성의 제1~제3 포먼트를 출력으로 한다. 최소제곱법에 의해 외부자극인 제1~제3 포먼트의 참값과 추정값 간의 평균 제곱 오차를 최소화하도록 필터계수를 결정한다. 얻어진 최적선형필터와 임의의 음성정보를 담고 있는 spike train으로부터 포먼트의 추정값을 얻을 수 있다.

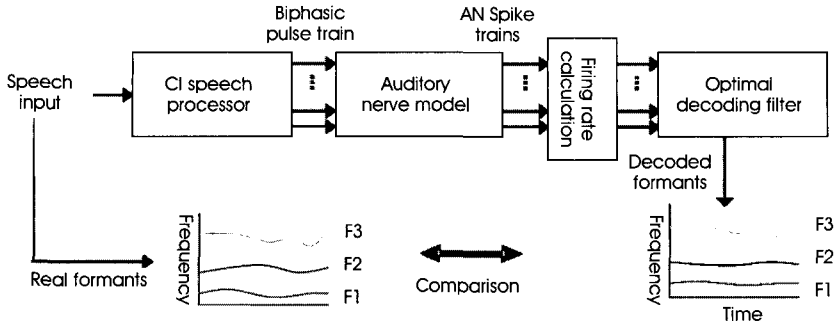

그림 2. 어음처리방식 평가방법 블록도
Fig. 2. Block diagram of the method for speech processing strategy evaluation

B. 어음처리방식 평가방법

그림 2는 어음처리방식 평가방법을 설명하기 위한 그림이다. 확률적 청신경모델에 특정 어음처리방식에 기반한 전기자극펄스를 인가하여 multiple spike trains을 얻고 이를 최적선형 디코딩필터의 입력으로 하고 입력음성으로부터 얻은 제1~제3포먼트 궤적(F1~F3)의 참값을 출력으로 사용하여 디코딩 필터계수를 결정한다. 얻어진 디코딩필터의 입력으로 입력음성에 대한 청신경모델의 응답을 사용하여 포먼트의 추정값을 얻고 이를 참값과 상관계수를 이용하여 비교하여 포먼트가 얼마나 충실하게 복원되는지를 정량적으로 관찰한다. 포먼트 궤적이 충실하게 복원될수록 참값과 유사한 형태를 가지며 상관계수가 1에 가까운 값을 갖게 되며 이를 이용하여 디코딩 성능을 정량화하였다. 학습을 통해 얻어진 디코딩필터의 계수와 청신경모델 응답의 통계적 특성 등에 대하여 독립적인 성능을 관찰하기 위하여 50회의 시뮬레이션을 통해 얻어진 상관계수의 평균값을 비교, 관찰하였다.

C. 확률적 청신경모델

확률적 청신경모델(그림 3)은 특정 어음처리방식에 의해 생성된 전기자극펄스를 입력으로 하며 spike train을 출력으로 한다[19, 20]. 전기자극펄스에 실린 소리의 정보가 청신경의 spike train으로 전달되므로 소리의 정보가 청신경 응답에 어떻게 인코딩 되는지를 관찰할 수 있다. 모델의 응답을 이용한 spike train decoding을 수행함으로써 어음처리방식의 성능평가를 수행할 수 있다.

확률적 청신경모델은 전기자극펄스의 세기가 청신경의 세포막 노이즈와 불응기를 고려한 문턱치의 합보다 클 경우 발화하도록 되어있다. 그림 3에서 V_{stim}은 전기자극펄스 입력, V_{noise}는 2500 Hz 대역 제한된 백색잡음이며 이것의 표준편차는 modulation depth와 같은 비율로 결정하였다(modulation depth/V_{noise}의 표준편차 = 1). V_{th}은 청신경의 문턱치, V_{ref}는 불응기를 고려한 청신경의 문턱치 변화이다.

D. 어음처리방식(Speech processing strategy)의 종류

제안된 방법을 이용하여 가장 널리 사용되는 시불변-선형 필터뱅크 기반 어음처리방식과 말초청각계의 특성을 좀 더 면밀하게 모방한 시변-비선형 필터뱅크 기반 어음처리 방식을 비교하였다. 실험결과 시변-비선형 필터뱅크 기반 어음처리방식을 채택할 경우 시불변-선형 필터뱅크에 의한 방식보다 포먼트를 충실히 복원할 수 있었다. 시변-비선형필터뱅크 기반 어음처리방식은 달팽이

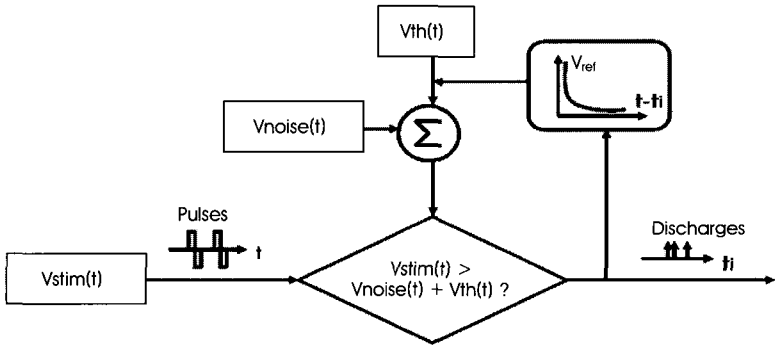

그림 3. 확률적 청신경모델. $V_{stim}(t)$: 특정 어음처리방식에 기반하여 생성된 전기자극펄스열, $V_{noise}(t)$: 2500Hz 대역 제한된 백색잡음, $V_{th}(t)$: 불응기를 고려한 청신경모델의 문턱치
Fig. 3. Stochastic auditory nerve model. $V_{stim}(t)$: electrical pulse train, $V_{noise}(t)$: Band-limited white Gaussian noise, $V_{th}(t)$: Threshold voltage of auditory nerve firing considering refractory effect.

관내 기저막(basilar membrane)의 선형 및 비선형특성을 고려한 것이므로 기저막의 선형특성만을 고려한 어음처리방식에 비하여 잡음 하에서 음성정보전달특성을 향상시킬 수 있을 것으로 기대된다[20,21].

인공와우 어음처리방식으로 가장 널리 사용되는 시불변-선형 필터뱅크 기반 어음처리방식과 말초청각계의 특성을 좀 더 면밀하게 모방한 시변-비선형 필터뱅크 기반 어음처리 방식을 채택, 비교하였다[15,16]. 전자는 널리 사용되는 어음처리방식으로 기저막의 특성을 선형 대역통과필터로 표현하고 있으며 잡음이 없는 경우 높은 어음 인식율을 보이나 잡음하에서 사용자의 어음 인식율이 급격히 감소한다[16,17,21]. 이에 대한 하나의 원인으로 기저막의 비선형성을 반영하지 못한다는 점을 생각할 수 있다. 일반적인 선형 필터뱅크를 통하여는 외측유모세포(outer hair cell)의 피드백 효과를 모사하는 입력음성신호 크기에 따른 대역통과 필터의 대역 변화를 표현할 수 없다. 후자는 기저막의 비선형 특성을 모델링 하기 위한 dual resonance nonlinear(DRNL) 모델에 기반하고 있으며 한 채널이 그림 4(b) 에서와 같이 선형, 비선형 두 개의 신호처리 통로로 구성되어 채널의 출력은 두 통로의 합이 된다. 선형 통로의 중심주파수는 비선형 통로의 중심주파수와 약간 어긋나 있어 쌍공명(dual resonance)필터라 한다. 또한 대역 통과 필터의 대역폭이 비대칭적이고 compression과 같은 외측유모세포의 비선형성을 표현한다. 이와 같은 기저막의 비선형성에 의해 시변-비선형 필터뱅크 기반 어음처리방식은 잡음 하에서도 음성 정보 전달 측면에서 강인한 특성을 보이는 것으로 알려져 있다[16,21].

E. 전기자극펄스의 생성 및 변조

각 어음처리방식에 의해 그림 4(a)에서 알 수 있듯이 12채널에 대한 전기자극펄스가 생성된다[21]. 충분한 음성정보의 전달은 전기자극펄스의 자극율과 변조 정도 등에 큰 영향을 받는다. 음성의 정보는 느리게 변화하는 포락선(envelope)과 상대적으로 빠르게 변화하는 fine time structure로 이루어진다. Litvak 등에 의하면 fine time structure의 전달을 위해서는 높은 자극율이 필요하며

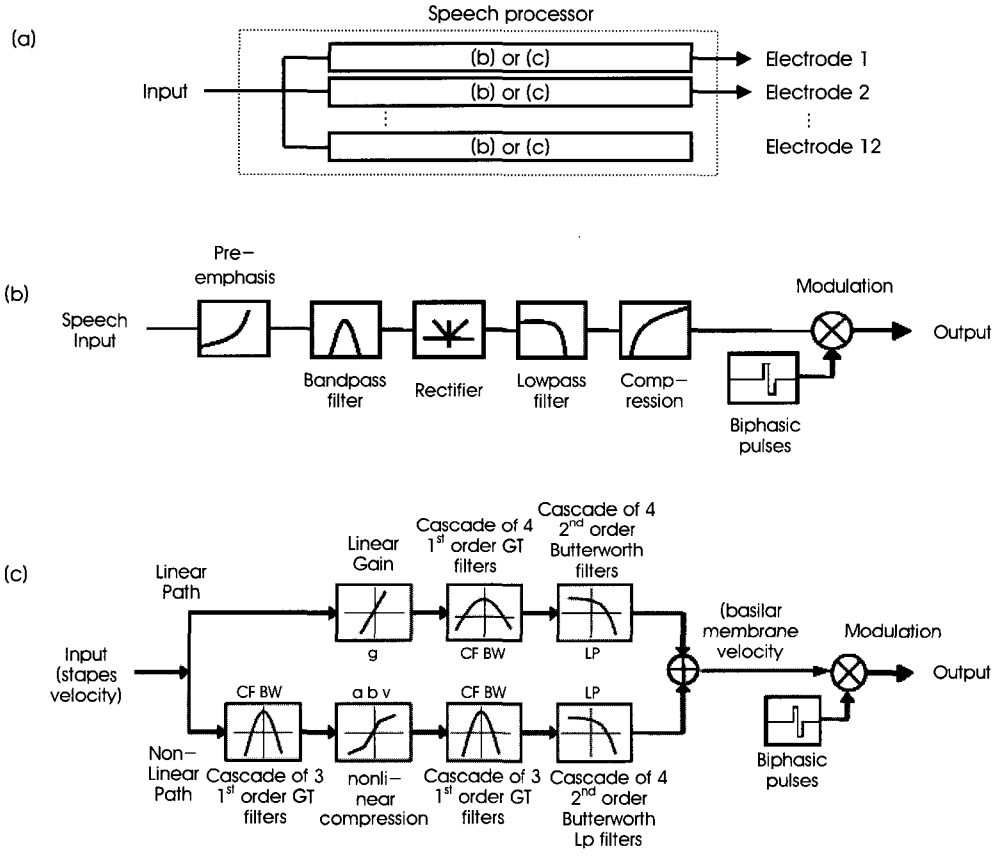

그림 4. (a) 어음처리방식((b) 또는 (c))에 기반한 채널별 전기자극펄스 생성 과정 블록도, (b) 시불변-선형 필터뱅크 기반 어음처리방식, (c) 시변-비선형 필터뱅크 기반 어음처리방식

Fig. 4. (a) Block diagram of CI pulse generation, (b) Speech processing strategy based on time-invariant linear filterbank, (c) Speech processing strategy based on time-varying nonlinear filterbank

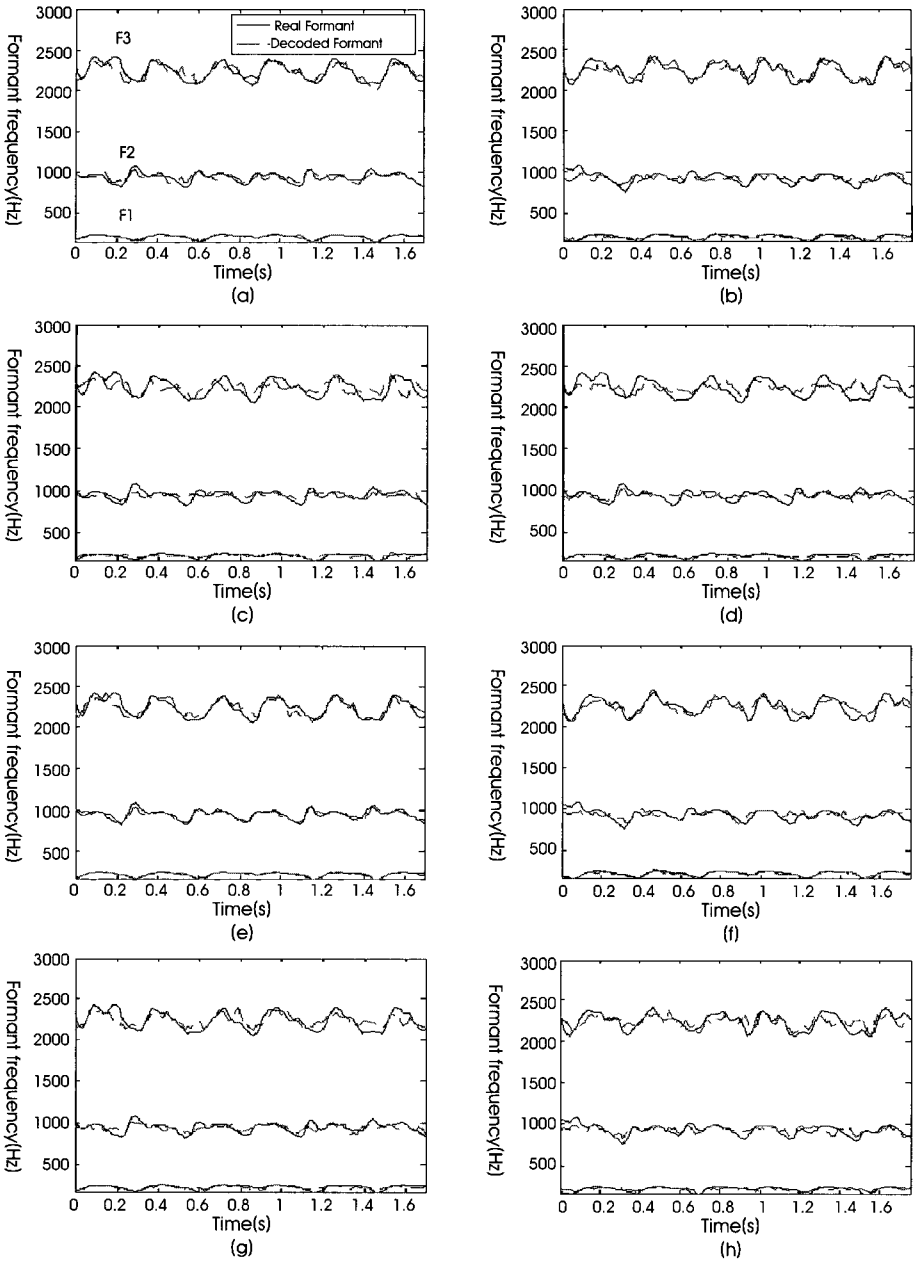

그림 5. 포먼트 궤적의 참값(실선)과 추정값(점선) 비교. (a)~(d)는 시불변-선형필터뱅크 기반 어음처리방식에 대한 포먼트 궤적. (e)~(h)는 시변-비선형필터뱅크 기반 어음처리방식에 대한 포먼트 궤적. (a),(e)는 잡음이 없는 경우, (b),(f)는 5 dB SNR 백색잡음 첨가시, (c),(g)는 0 dB SNR 백색잡음 첨가시, (d),(h)는 5 dB SNR 음성형태 잡음 첨가시

Fig. 5. Comparison of true and estimated formant trajectories (true: line, estimated formant: dotted). (a)~(d) Formant trajectories estimated from AN responses stimulated by linear-filterbank-based strategy, (e)~(h) Formant trajectories estimated from AN responses stimulated by time-varying nonlinear filterbank. (a), (e) without noise, (b),(f) with 5 dB SNR WGN, (c),(g) with 0 dB SNR WGN, (d),(h) with 5 dB SNR SSN.

또한 낮은 modulation depth에 의해 음성정보를 충실하게 전달할 수 있다[18,19]. 본 실험에서는 실제 이식자에게 사용되는 범위에서 Litvak 등이 fine time structure의 전달을 위해 사용한 범위인 800~5000 pulses per second(pps)으로 자극율을 변화시켰으며 1%의 낮은 modulation depth를 갖도록 조절하였다[18, 19].

F. 상세 시뮬레이션 파라미터 설정

시뮬레이션시 '자음+모음' 형태의 합성음 /mu/가 6회 반복 표현된 1.7 초의 길이의 음성을 사용하여 12채널에 대한 전기자극펄스를 생성하였다. 이 때 전기자극펄스의 자극율은 800, 2000, 5000 pps로 설정하였고, 진폭 변조(amplitude modulation)는 최

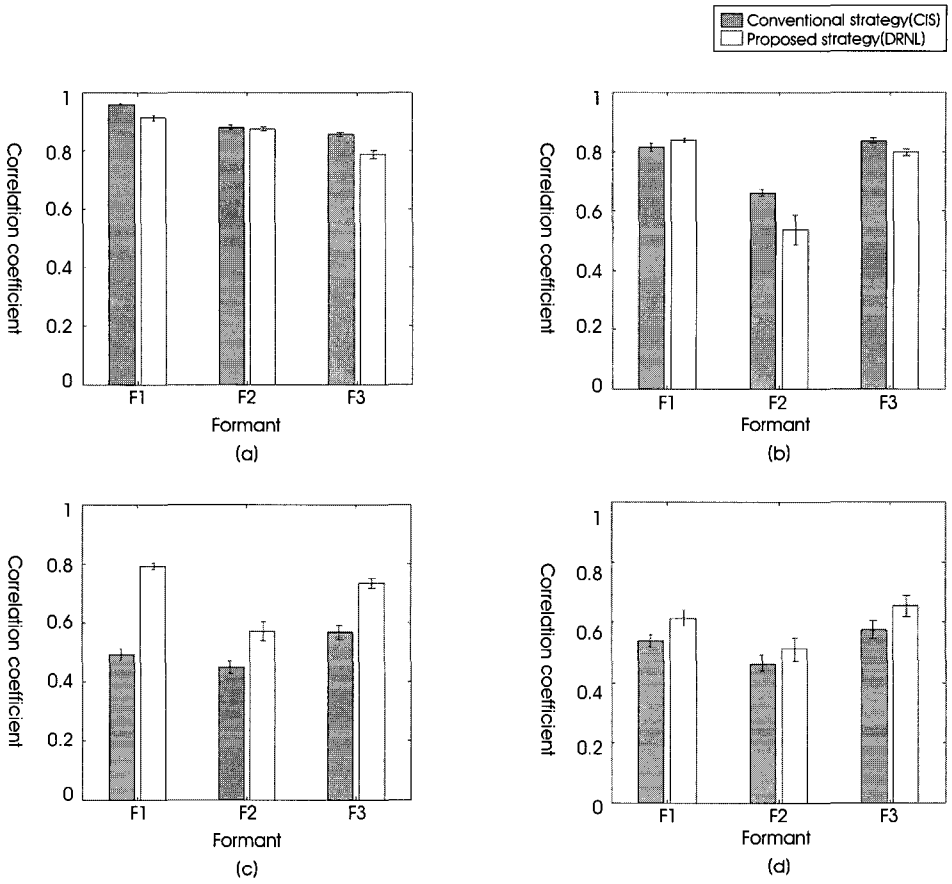

그림 6. 어음처리방식 간 평균 포먼트 디코딩 성능비교. (a)는 잡음이 없는 경우, (b)는 5 dB SNR 백색 잡음 첨가시, (c)는 0 dB SNR 백색잡음 첨가시, (d)는 5 dB SNR 음성형태 잡음 첨가시

Fig. 6. Comparison of speech processing strategies. (a) without noise, (b) with 5 dB SNR WGN, (c) with 5 dB SNR SSN.

대값을 450 μA로 두고 1% modulation depth를 갖도록 조절하였다. Spike train decoding을 수행하기 위해 디코딩필터로 최적선형필터를 이용하였으며 12채널로부터 얻은 spike train의 발화율 계산을 위한 time bin 폭과 디코딩필터의 차수는 가장 높은 디코딩 성능을 보이되 학습 데이터에 대하여 overfitting되지 않도록 최적화 되었다. Time bin 폭은 15 ms로 설정하였고 디코딩필터의 차수는 5차로 75 ms 길이를 갖도록 설정하였다. 실험에 사용된 잡음 환경은 잡음이 없는 경우, 5 혹은 0 dB SNR 백색잡음(white Gaussian noise, WGN) 첨가 시 그리고 5 dB SNR 음성형태 잡음(음성과 유사한 형태의 스펙트럼을 갖는 잡음 - speech shaped noise, SSN) 첨가 시 등으로 각각의 조건에서 두 어음처리방식의 성능평가 비교를 수행하였다.

III. 결 과

A. 어음처리방식의 성능평가

전기자극펄스의 자극율의 변화에 상관없이 유사한 결과를 보였으므로 자극율이 2000인 경우의 결과를 대표적으로 수록하였다.

포먼트 궤적비교

그림 5는 입력음성으로부터 추출된 제1~제3포먼트 궤적(F1~F3)의 참값과 추정값을 비교한 대표적인 결과이다. 여러 번의 시뮬레이션으로부터 얻은 그림 중 보편적으로 보이는 결과를 선택하였다. 잡음이 없는 경우 혹은 5 dB SNR 백색잡음(white Gaussian noise, WGN) 하에서 두 자극방법 간 차이는 거의 없는 것으로 관찰된다. 그러나 잡음레벨이 상대적으로 큰 0 dB SNR 백색잡음 첨가 및 입력음성 형태의 잡음(speech shaped noise, SSN) 첨가의 경우 시변-비선형 필터뱅크 기반 어음처리 방식에 의한 포먼트 궤적의 추정값이 참값과 더 유사해 보인다.

평균성능비교

그림 6은 50회의 시뮬레이션 결과로부터 얻은 제1~제3포먼트 궤적(F1~F3)의 추정값과 참값 간의 유사성을 상관계수의 평균값을 이용하여 정량적으로 비교한 결과이다. 잡음이 없는 경우와 5 dB SNR 백색잡음을 첨가했을 경우의 평균성능은 유사한 수준이나 잡음레벨을 증가시켜 0 dB SNR 백색잡음을 첨가하거나 입력음성 형태의 잡음을 첨가할 경우 시변-비선형 필터뱅크 기반 어음처리 방식이 상대적으로 높은 평균성능을 보였다. 즉, 백색 잡음에 대한 경우

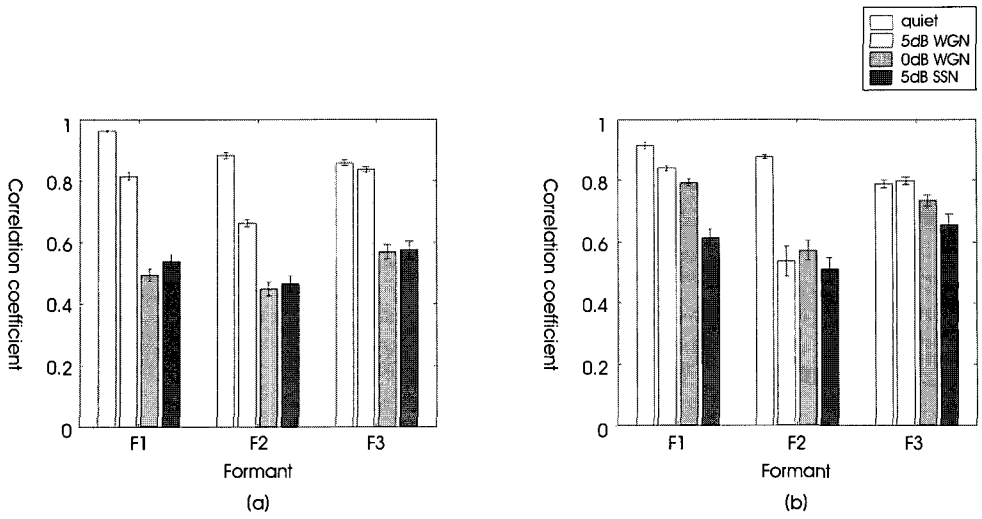

그림 7. 잡음첨가에 따른 평균 포먼트 디코딩 성능. (a)는 시불변-선형 필터뱅크 기반 어음처리방식, (b)는 시변-비선형 필터뱅크 기반 어음처리방식에 대한 평균 포먼트 디코딩 성능

Fig. 7. Performances of linear and nonlinear speech processing strategies under noisy conditions. (a) Linear filterbank, (b) Time-varying nonlinear filterbank.

보다 음성 형태의 잡음 하 혹은 잡음 레벨이 증가할수록 시변-비선형 필터뱅크 기반 어음처리 방식의 우수성을 관찰할 수 있었다.

그림 7은 여러 가지 종류의 잡음첨가에 따른 포먼트 궤적의 추정값과 참값 간 유사성을 상관계수를 이용하여 정량적으로 비교한 결과이다. 시불변-선형 필터뱅크 기반 어음처리방식을 사용했을 경우 높은 SNR의 백색잡음 하와 입력음성형태의 잡음 하에서의 포먼트 디코딩 성능이 급격하게 감소한다. 반면 시변-비선형 필터뱅크 기반 어음처리방식은 잡음에 대하여 강인한 특성을 보인다. 잡음이 없는 경우와 5 dB SNR 백색잡음 하에서는 시변-비선형 필터뱅크 기반 어음처리 방식과 유사한 수준의 평균성능을 보였으나 백색잡음레벨이 5 dB SNR에서 0 dB SNR로 증가한 부분과 음성형태의 잡음 첨가시의 평균 디코딩 성능을 비교해보면 시변-비선형 필터뱅크 기반 어음처리방식에 의한 평균 디코딩 성능의 감소가 상대적으로 완만하다.

Ⅳ. 토 의

청신경 자극방법을 결정하는 어음처리방식의 성능평가는 인공와우의 성능을 예측하기 위하여 필수적이다[4-7, 21]. 많은 연구자료에서 acoustic simulation 등의 방법을 이용하였으나 실제 이식과 훈련을 거친 후의 어음처리방식의 성능을 효과적으로 예측할 수 있는가에 대한 불확실성이 존재하였다[8]. 본 연구에서는 neural spike train decoding에 기반한 어음처리방식을 제안하였다. 여러 가지 잡음하에서 지정된 어음처리방식에 기반하여 생성된 전기자극펄스에 의하여 자극된 청신경 응답으로부터 neural spike train decoding 방법을 이용하여 음성 인식에 중요한 역할을 하는 제1, 2 그리고 제3 포먼트 궤적을 추정하여 입력음성의 포먼트와 비교함으로써 어음처리방식의 포먼트 정보 표현 특성을 평가하였다.

실험결과 전기자극펄스 자극율의 변화에 상관없이 잡음하에 시변-비선형 필터뱅크 기반 어음처리방식의 우수성을 관찰할 수 있었다. 자극율이 증가함에 따라 빠르게 변화하는 음성정보의 전달도가 증가하여 음성 인식률 향상에 기여할 수 있으나 본 연구에서는 상대적으로 느리게 변화하는 음성정보인 음성파형의 포락선(envelope)에 대한 실험만 진행하였다[18,19].

그림 5는 잡음이 없는 경우 및 여러 가지 배경잡음 하에서 포먼트 궤적의 참값과 추정값을 비교한 그림이다. 잡음이 없는 경우 혹은 낮은 레벨의 잡음 첨가시는 어음처리방식간의 성능차이를 보기 어렵고 상대적으로 높은 레벨의 잡음 첨가시 혹은 음성 형태의 잡음 첨가시에도 차이를 관찰할 수는 있으나 얼마나 차이가 나는지 알기 어렵다. 또한 한 회의 학습을 통하여 얻은 디코딩 필터계수와 청신경모델 응답의 통계적 특성 등에 대하여 독립적인 포먼트 디코딩 성능을 관찰하기 위해서는 여러 회의 시뮬레이션에 대한 포먼트 궤적의 참값과 추정값 간의 유사성을 정량적으로 비교하는 것이 필요하다. 그림 6은 어음처리방식 간 평균적인 포먼트 디코딩 성능을 상관계수로 정량화하여 비교한 것이고 그림 7은 잡음이 없는 경우 및 여러종류의 잡음 하에서의 평균 포먼트 디코딩 성능을 비교한 것이다. 잡음이 없거나 낮은 크기의 백색잡음이 첨가되었을 경우 두 어음처리방식 간 성능은 유사한 수준을 보였다. 기존의 시불변-선형 필터뱅크 기반 어음처리방식을 사용했을 경우 높은 크기의 백색잡음 하 혹은 입력음성과 유사한 스펙트럼 특성을 갖는 잡음 하에서는 포먼트 디코딩 성능이 급격히 감소하였고 따라서 낮은 음성인식율이 기대된다. 그러나 시변-비선형 필터뱅크 기반 어음처리방식에 의한 포먼트 디코딩 성능은 높은 크기의 백색잡음 혹은 음성형태의 잡음이 존재할 경우에도 완만한 감소를 보이며 시불변-선형 필터뱅크 기반 어음처리방식에 비해 우수한 성능을 보인다. 이는 기저막의 비선형적인 특성을 반영함으로써

잡음에 강인한 특성을 구현할 수 있음을 보여준다.

청각모델에서 기저막에 해당하는 대역통과필터의 특성은 같다고 가정한 경우가 많으며 이러한 가정은 잡음이 없는 환경에서는 큰 문제가 되지 않으나 잡음이 있는 상황에서는 음성의 인식률을 크게 저하시키는 원인이 된다. 정상인의 경우 기저막과 내측유모세포(inner hair cell)/청신경(auditory nerve) 시냅스, 외측유모세포(outer hair cell) 등에서 보이는 비선형성이 음성의 인식에 기여하기 때문에 상당히 큰 잡음 하에서도 성공적인 음성인지가 가능하다[16,21]. 시변-비선형 필터뱅크 기반 어음처리방식은 이와 같은 비선형성을 추가하였기 때문에 잡음 하에서도 음성 인지에 중요한 포먼트 성분을 청신경 응답으로 충실히 전달할 수 있는 것으로 판단된다.

이와 같은 결과는 선행연구에서 spectrogram, acoustic simulation 등을 이용하여 시변-비선형 필터뱅크 기반 어음처리 방식에서 음성의 포먼트 성분이 더 두드러지게 표현된 것을 확인한 것과 같은 맥락으로 neural spike train decoding에 의한 어음처리방식의 성능평가의 효용성을 알 수 있다[21].

참고문헌

[1] P. Mitchell, "The prevalence, risk factors and impacts of hearing impairment in an older Australian Community: The Blue Mountains Hearing Study," *XXVI International Congress of Audiology*, Melbourne, Australia, 2002.

[2] A. Goodsall, N. Condoleon, and R. Cummins, "Replacing hearing aids," Analyst Report, Cochlear Ltd., 20 Jun. 2003.

[3] M. F. Bear, B. W. Connors, and M.A. Paradiso, *Neuroscience: Exploring the brain*, 2nd Ed., Lippincott Williams & Wilkins, 2004, pp.357-384.

[4] J. T. Rubinstein, "How cochlear implants encode speech," *Curr. Opin. Otolaryngol. Head Neck Surg.*, vol. 12, no. 5, pp.444-448, 2004.

[5] P. C. Loizou, "Introduction to cochlear implants," *Tutorial article on cochlear implants that appeared in the IEEE Signal Processing Magazine*, Sept. 1998., pp. 101-130.

[6] D. B. Grayden, A. N. Burkitt, O. P.Kenny, J. C. Clarey, A. G. Paolini, and G. M. Clark, "A cochlear implant speech processing strategy based on an auditory model," in *Proc. of 2004 Intelligent Sensors, Sensor Networks and Information Processing Conference*, Dec. 2004, pp.491-296.

[7] P. J. Blamey, R. C. Dowell, A. M Brown, G. M. Clark, and P. M. Seligman, "Speech processing studies using an acoustic model of a multiple-channel cochlear implant," *J. Acoust. Soc. Am.*, vol. 76, pp.104-110, 1984.

[8] J. T. Rubinstein, C. Turner, "A novel acoustic simulation of cochlear implant hearing: effects of temporal fine structure," *in Proc. 1st International IEEE EMBS Conference, Neural Engineering*, Mar. 2003, pp.142 - 145.

[9] K. H. Kim, S. S. Kim, and S. J. Kim, "Improvement of spike train decoder under spike detection and classification errors using support vector machine," *Med. Biol. Eng. Comput.*, vol. 44, no. 1-2, pp.124-30, 2006.

[10] D. K. Warland, P. Reinagel, and M. Meister, "Decoding visual information from a population of retinal ganglion cells," *J. Neurophysiol.*, vol. 78, pp.2336-50, 1997.

[11] A. T. Neel, "Formant detail needed for vowel identification," *J. Acoust. Soc. Am.*, vol. 5, pp.125-131, 2004.

[12] G. E. Peterson, H. L. Barney, "Control methods used in study of the vowels," *J. Acoust. Soc. Am.*, vol. 24, pp.175-184, 1952.

[13] L. Deng, C. D. Geisler, "A composite auditory model for processing speech sounds," *J. Acoust. Soc. Am.*, vol. 82, pp.2001 - 2012, 1987.

[14] E. N. Brown, R. E. Kass, and P. P. Mitra, "Multiple neural spike train data analysis: state-of-the-art and future challenges," *Nature neuroscience*, vol. 7, no. 5, pp.456-461, 2004.

[15] B. S. Wilson, C. C. Finley, D. T. Lawson, R. D. Wolford, and M. Zerbi, "Design and evaluation of a continuous interleaved sampling (CIS) processing strategy for multichannel cochlear implants," *J. Rehabil. Res. Dev.*, vol. 30, no. 1, pp.110-116, 1993.

[16] C. J. Sumner, L. P. O'Mard, E. A. Lopez-Poveda, and R. Meddis, "A nolinear filter-bank model of the guinea-pig cochlear nerve: Rate responses," *J. Acoust. Soc. Am.*, vol. 113, pp.3264-3274, 2003.

[17] B. S. Wilson, C. C Finley, D. T. Lawson, R. D. Wolford, D. K. Eddington, and W. M. Rabinowitz, "Better speech recognition with cochlear implants," *Nature*, vol. 352, pp.236 - 238, 1991.

[18] L. M. Litvak, B. Delgutte, and D. K. Eddington, "Improved temporal coding of sinusoids in electric stimulation of the auditory nerve using desynchronizing pulse trains," *J. Acoust. Soc. Am.*, vol. 14, pp.2079-2098, 2003.

[19] L. M. Litvak, B. Delgutte, and D. K. Eddington, "Auditory nerve fiber responses to electric stimulation: modulated and unmodulated pulse trains," *J. Acoust. Soc. Am.*, vol. 110, pp.368 - 379, 2001.

[20] I. C. Bruce, L. S. Irlicht, M. W. White, S. J. O'leary, S. Dynes, E. Javel, and G. M. Clark, "A stochastic model of the eletrically stimulated auditory nerve: pulse-train response," *IEEE Trans. Biomed. Eng.*, vol. 46, no. 6, pp.630-637, 1999.

[21] J. H. Kim, D. H. Kim, and K. H. Kim, "A speech processing strategy for auditory prosthesis based on nonlinear filterbank model of biological cochlear," *World congress on Medical Physics and Biomedical Engineering*, Seoul, Korea, 2006.