

Spike Train Decoding에 기반한 인공와우 어음처리의 음성시작점 정보 전달특성 평가

김두희, 김진호, 김경환

연세대학교 보건과학대학 의공학부

(Received June 23, 2007. Accepted July 30, 2007)

Performance Evaluation of Speech Onset Representation Characteristic of Cochlear Implants Speech Processor using Spike Train Decoding

Doo-Hee Kim, Jin Ho Kim, Kyung Hwan Kim

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

Abstract

The adaptation effect originating from the chemical synapse between auditory nerve and inner hair cell gives advantage in accurate representation of temporal cues of incoming speech such as speech onset. Thus it is expected that the modification of conventional speech processing strategies of cochlear implant (CI) by incorporating the adaptation effect will result in considerable improvement of speech perception performance such as consonant perception score. Our purpose in this paper was to evaluate our new CI speech processing strategy incorporating the adaptation effect by the observation of auditory nerve responses. By classifying the presence or absence of speech from the auditory nerve responses, i. e. spike trains, we could quantitatively compare speech onset detection performances of conventional and improved strategies. We could verify the effectiveness of the adaptation effect in improving the speech onset representation characteristics.

Key words : Cochlear implant, Speech processing strategy, Speech Onset, Neural spike train decoding

1. 서론

정상인의 경우 소리가 외이를 통하여 들어와 고막을 진동시키고 이 진동은 중이의 이소골에 의하여 증폭되어 달팽이관 (cochlear)에 전달된다. 이 진동이 달팽이관 내의 림프액에 전달되어 유모세포 (hair cell)가 움직임에 따라 전기자극이 발생한다. 이는 신경절 세포 (spiral ganglion cell)에서 활동전위의 형태로 바뀌어 뇌로 전달되어 소리를 인지할 수 있게 된다. 그러나 과도한 잡음에 노출되거나 고령 (aging), 수막염 (meningitis) 등의 원인으로 유모세포가 손실된 경우 소리자극이 전기자극으로 변환될

수 없으며 소리를 인지할 수 없다. 이 경우 소리의 크기를 증폭시켜 외이에 인가하여주는 보청기나 이소골의 역할을 대체하는 인공중이 (middle ear hearing aid) 또는 청력을 회복시킬 수 없다. 유모세포가 손실되었으나 청신경이 정상인 사람에 한하여 인공와우 (cochlear implants)를 통한 청력회복이 가능하다 [1,2].

인공와우 장치는 청신경에 소리정보를 담은 전기자극을 인가함으로써 청력을 회복시켜주는 장치로 마이크, 어음처리기, 트랜스미터, 전극으로 구성된다 [2]. 마이크로 입력된 소리정보는 어음처리기를 통하여 전기자극펄스 형태로 바뀌어 달팽이관 내에 삽입된 전극으로 전달된다 [2]. 적응효과 (adaptation effect)는 내측유모세포 (inner hair cell; IHC)와 청신경 사이의 화학적 시냅스 (chemical synapse)에 기원하는 현상으로 자극의 시작점에서 신경전달물질 (neurotransmitter)의 급격한 방출로 높은 발화율을 보이고 이후 신경전달물질의 고갈 (depletion)에 의하여 발화율이 단조 감소(monotonic decrease)하는 형태를 보인다 [1,3,4]. 정상인의 경우 적응효과가 발생하지만 청신경을 전기적으로 자극하여

본 연구는 과학기술부/한국과학재단 우수연구센터 육성 사업 (R11-2000-075-01005-0) 및 보건복지부 의료기기 기술개발 사업인 인체기능 및 대사영양신기술 개발 연구센터의 연구비 지원으로 수행되었음 (02-PJ3-PG6-EV07-0002)

Corresponding Author : 김경환
강원도 원주시 용업면 매지리 연세대학교 원주캠퍼스 백운관 208호
우)220-710
Tel : 82-33-760-2364 / Fax : 82-33-763-1953
E-mail : khkim0604@yonsei.ac.kr

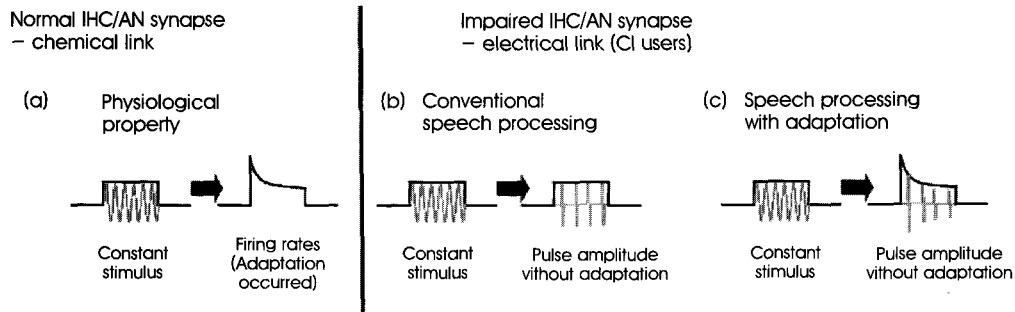


그림 1. 내부유모세포 (inner hair cell, IHC)와 청신경 (auditory nerve, AN) 간 시냅스의 특성비교. (a) 정상인의 IHC/AN 간 화학적 시냅스에 기원하는 적응효과, (b) 일반적으로 이용되는 어음처리방식에 기반하여 생성된 전기자극펄스, (c) 어음처리방식에 적응효과를 적용하여 생성된 전기자극펄스

Fig. 1. Comparison of the characteristics of IHC/AN synapse. (a) Normal IHC/AN, (b) characteristic of IHC/AN reproduced by conventional speech processing strategy, (c) characteristic of IHC/AN reproduced by speech processing strategy with adaptation

소리정보를 전달하는 인공와우의 경우 이를 기대할 수 없다 [3,4]. 그러므로 어음처리방식에 적응효과를 더하여 줌으로써 음성의 시작점에 대한 정보전달 특성을 개선해 주어야 한다. 어음처리는 어음처리방식에 기반하여 전기자극펄스를 생성하는데 이는 달팽이관의 tonotopy 특성을 모방하여 입력되는 소리정보를 주파수 성분으로 분해하는 주파수 분석부 (frequency analyzer)와 이로부터 전기자극펄스의 크기를 결정하는 포락선 추출부 (envelope extractor)로 구성된다 [2]. 우리는 선행연구를 통하여 어음처리방식에 적응효과를 더해 주어 정상 청각계에서 보이는 적응효과를 모사할 수 있으며 음성시작점에 대한 정보전달 특성을 개선할 수 있음을 확인하였다 [5,6].

본 연구에서는 어음처리방식에 적응효과 (adaptation)를 추가하여 입력음성의 포락선 추출특성을 개선할 경우 음성의 시작점에 대한 정보가 청신경응답으로 효과적으로 전달되는지의 여부를 관찰하였다. 적응효과 적용 전후의 어음처리방식에 기반한 전기자극펄스에 의해 자극된 청신경모델의 응답을 선형분류기의 입력으로 이용하여 음성시작점에 해당하는 구간의 음성유무를 분류하였다. 음성이 시작되는 부분에 대하여 분류를 수행하였으므로 음성이 있다고 분류하는 경우가 많을수록 청신경 응답에 실린 음성시작점에 대한 정보가 충실한 것으로 볼 수 있다. 그러므로 적응효과 적용 전후의 음성유무 분류성능을 이용하여 적응효과의 효용성을 평가가 가능하며 본 연구에서는 두 가지 어음처리방식에 적응효과 적용 전후의 음성시작점 정보 전달특성을 평가하였다.

입력음성으로는 ‘자음+모음’형태의 음성을 이용하였다. 모음의 경우 음성파형의 크기가 자음 (consonant)에 비하여 현저하게 크며 모음으로 시작되는 음성의 경우 시작점 검출율이 비슷한 조건에서 거의 100%에 이르기 때문에 어음처리방식에 대한 성능차이의 평가가 적절치 못하다. 자음 중에서 파열음 (stop) 또는 마찰음 (fricative)에 대하여 특히 무성음 (unvoiced)일 경우 음성파형의 크기가 매우 작으므로 이러한 자음으로 시작되는 음성의 경우 시작점의 강조가 필요하다. 그러므로 여러 종류의 자음을 입력으로 이용하여 어음처리방식에 적응효과를 적용할 경우 음성시작점 정

보 전달특성평가를 위한 실험을 수행하였다.

II. 음성처리방식 평가 방법

A. 적응효과를 고려한 인공와우 어음처리방식

그림 1에서와 같이 일정 크기의 자극이 인가되었을 때 청신경의 응답은 자극의 시작점에서는 높은 발화율을 보이고 시간이 지남에 따라 감소하는 형태를 보인다 [1,3]. 이러한 청신경의 응답특성은 달팽이관 내의 내측유모세포 (inner hair cell; IHC)와 청신경 (auditory nerve; AN) 사이의 화학적 시냅스에 기원하며 신경전달물질이 급격하게 방출되어 자극 초기에는 높은 발화율을 보이고 이후로 신경전달물질의 소모로 인하여 발화율은 단조감소하게 된다 [1]. 적응효과는 포락선 추출부에 적응효과를 더하여 음성의 시작점에 해당하는 부분의 포락선을 강조함으로써 전기자극펄스의 크기를 증폭시켜 주는 방식으로 구현되었다 [3,5].

그림 2는 실험에 이용된 시블변-선형필터뱅크 기반, 시블변-비선형필터뱅크 기반 어음처리방식을 설명하기 위한 그림이다. 시블변-선형 필터뱅크 기반 어음처리방식 (그림 2.(a))은 현재 가장 널리 사용되고 있으며 기저막의 특성을 선형 대역통과필터로 표현하고 있으며 잡음이 없는 경우 높은 어음 인식율을 보이나 잡음 하에서 사용자의 어음 인식율이 급격히 감소한다 [7]. 시블변-비선형 필터뱅크 기반 어음처리방식은 기저막의 비선형 특성을 모델링하기 위하여 Meddis 등이 제안한 dual resonance nonlinear (DRNL) 모델 (그림 2.(b))에 기반하고 있으며 한 채널이 선형, 비선형 두 개의 신호처리 통로로 구성되어 채널의 출력은 두 통로의 합이 된다 [8]. 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 시블변-비선형 필터뱅크 기반 어음처리방식은 잡음 하에서도 음성 정보 전달 측면에서 강인한 특성을 보이는 것으로 알려져 있다 [5,8,10,12].

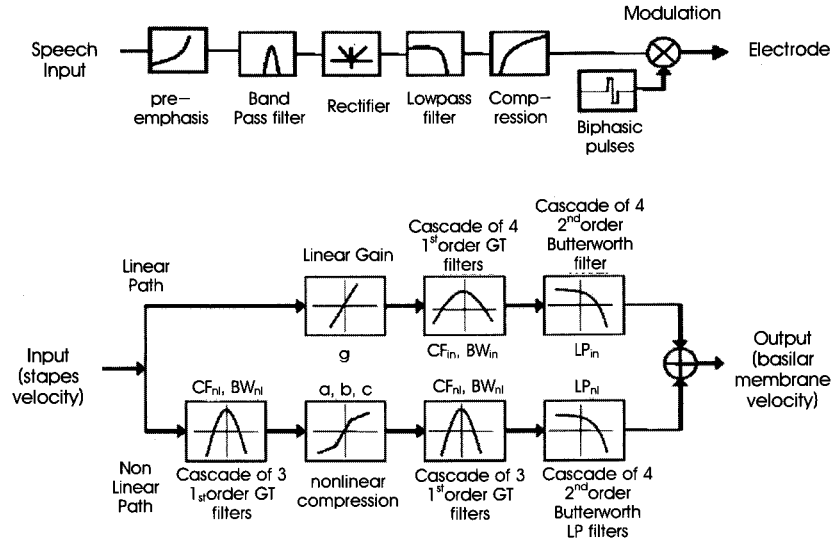


그림 2. 모의실험에 이용된 어음처리방식.
 (a) 일반적으로 사용되고 있는 시불변-선형필터뱅크 기반 어음처리방식, (b) 시변-비선형필터뱅크 기반 어음처리방식 (DRNL)
 Fig. 2. Speech processing strategy for simulation.
 (a) Conventional linear filterbank based strategy, (b) Nonlinear filterbank based strategy

B. 음성시작점 정보 전달특성 평가방법

그림 3는 spike train decoding에 기반하여 음성시작점에 대한 정보의 전달특성을 평가하는 방법을 설명하기 위한 그림이다 [9-11]. 적응효과를 적용 전후의 어음처리방식 [3,5-8]에 기반한 전기자극펄스에 의하여 자극된 청신경 모델 응답과 음성의 유무를 각각 class 1, class -1로 할당한 행렬을 이용하여 Wiener solution에 기반한 최적 선형분류기를 얻는다 [8]. 선형분류기의 입력으로 청신경 모델 응답을 이용하여 음성의 시작점에 대한 음성유무 분류결과를 얻고 이를 참값과 비교하여 적응효과 적용 전후의 음성 시작점 정보 전달특성을 비교한다. 음성유무 행렬의 참값은 그림 3에 점선으로 표시된 바와 같이 음성이 있는 부분은 1, 그렇지 않은 부분은 -1로 표시하였고, 추정값은 점실선으로 표시된 바와 같다. 음성시작점은 음성제작성 실제 음성이 시작되는 부분부터 20 ms 동안으로 가정하였다 [3].

C. 음성시작점 정보 전달특성 평가를 위한 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 spike train decoding이라 한다 [9-11].

그림 4는 선형분류기를 이용하여 음성시작점에서 음성유무 분류를 수행하기 위한 spike train decoding을 설명한 그림이다. 청신경모델 응답으로부터 계산된 발화율 행렬을 입력으로 이용하고 음성이 있는 부분은 1, 없는 부분을 -1로 표시한 행렬을 출력으로 이용하여 Wiener solution에 기반한 최적선형분류기를 얻는다. 선형분류기는 식 (1)을 이용하여 얻을 수 있다.

$$W = R^{-1} P, P = E[dR], R = E[XXT] \tag{1}$$

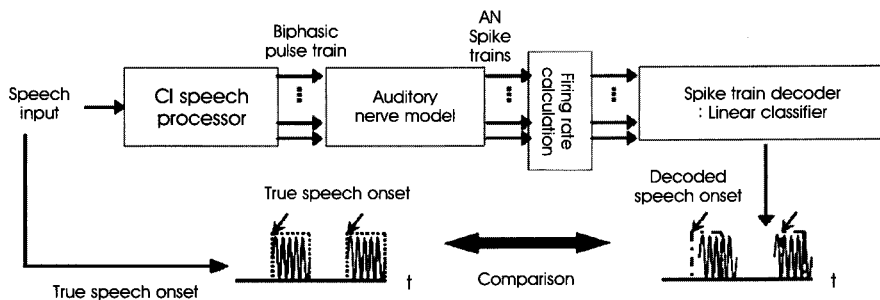


그림 3. Spike train decoding 기반 음성시작점 정보 전달특성 평가방법
 Fig. 3. Evaluation of speech onset representation characteristic based on spike train decoding

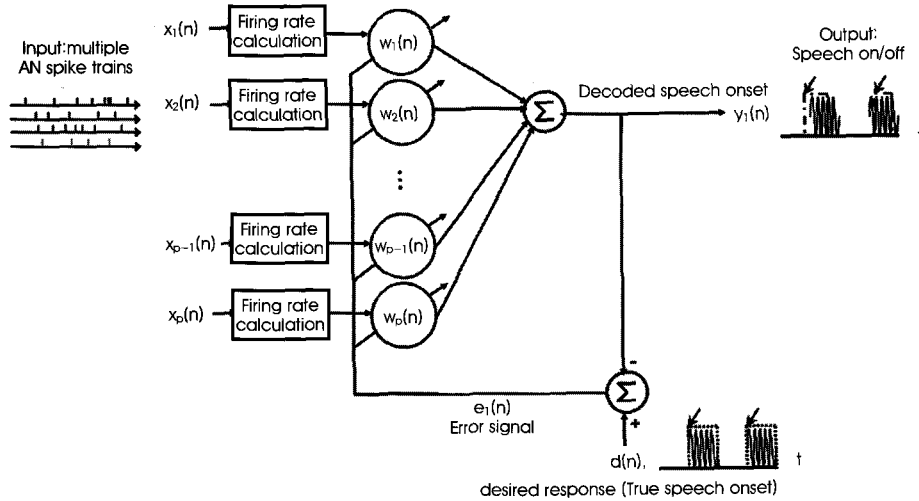


그림 4. 선형분류기를 이용한 spike train decoding
 Fig. 4. Spike train decoding using linear classifier

이 과정으로부터 얻은 선형분류기와 새로운 발화를 행렬 X_2 를 이용하여 식 (2)에 표기한 바와 같이 음성유무 분류행렬의 추정값을 얻을 수 있고, 음성시작점에 해당하는 부분만 음성유무 성능평가에 이용한다.

$$\hat{d} = X_2 W, \begin{cases} \hat{d} > 0, \text{ class } 1 \\ \hat{d} < 0, \text{ class } -1 \end{cases} \quad (2)$$

발화를 행렬은 X , 음성유무를 표기한 행렬 d 는 식 (3)과 같이 표현할 수 있다. N 은 관찰된 단일 뉴런의 개수, $r_N(i)$ 는 N 번째 단일 뉴런이 i 번째 time bin에서 갖는 평균발화율을 뜻하며 L 은 총 관찰시간 내 time bin의 개수이다.

$$X = \begin{bmatrix} 1 & r_1(0) & r_2(0) & \dots & r_N(0) \\ 1 & r_1(1) & r_2(1) & \dots & r_N(1) \\ \dots & \dots & \dots & \dots & \dots \\ 1 & r_1(L) & r_2(L) & \dots & r_N(L) \end{bmatrix} \quad (3)$$

$$d = [d(0) d(1) d(2) \dots d(L-1)] \quad (4)$$

어음처리방식으로 시분변-선형필터뱅크 기반 어음처리방식 [7]과 시변-비선형필터뱅크 기반 어음처리방식 [5,8]을 이용하였으며 각 어음처리방식에 적용효과 적용 전후의 음성시작점 정보 전달특성을 관찰하였다. 20회의 시뮬레이션을 통하여 음성시작점 정보 전달특성평가에 이용된 음성유무분류 샘플 수는 160개이며 이 중 음성유무를 올바르게 분류한 비율을 통하여 성능을 평가하였다. 음성시작점 정보 전달특성을 정량적으로 평가하기 위한 ‘% correct’는 다음과 같이 계산되었다.

$$\%correct = \frac{\text{음성을 올바르게 분류한 경우의 개수}}{160} (\%) \quad (5)$$

D. 확률적 청신경모델

확률적 청신경모델 (그림 5) 특정 어음처리방식에 의해 생성된

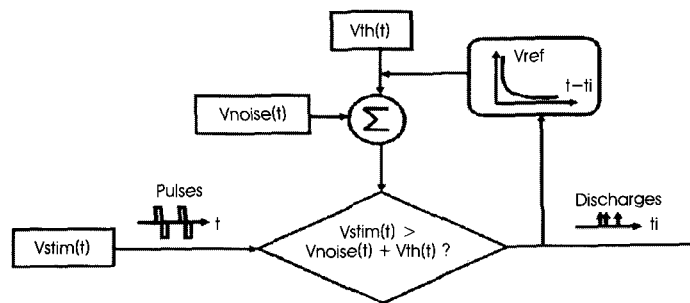


그림 5. 확률적 청신경모델. Vstim(t): 특정 어음처리방식에 기반하여 생성된 전기자극펄스열, Vnoise(t): 2500Hz 대역 제한된 백색잡음, Vth(t): 불응기를 고려한 청신경모델의 문턱치

Fig. 5. Stochastic auditory nerve model. Vstim(t): An electrical stimulation pulse train generated by specific speech processing strategy. Vnoise(t): Band-limited Gaussian Noise (bandwidth: 2500 Hz). Vth(t): Threshold level with consideration of refractory period.

전기자극펄스를 입력으로 하며 spike train을 출력으로 한다 [13,14]. 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 5에서 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. 청신경 모델의 문턱치, 청신경 세포의 막 전위인 V_{noise} 의 표준편차와 modulation depth 간의 비는 청신경 모델 응답의 발화율이 정상인의 청신경에서 관찰되는 발화율과 같이 300 spikes/s 이하가 되도록 하며 입력 자극신호가 없을 때 청신경 모델의 발화율인 pseudo-spontaneous 발화율 또한 비정상적으로 높지 않도록 조절하였다 [14].

E. 음성시작점 정보 전달특성평가를 위한 상세 파라미터

실험에 이용된 음성은 ‘자음+모음’ 형태이고 무음과 음성이 200 ms씩 반복되며 800 ms 길이를 가진다. 자음은 유성파열자음 (voiced stop consonant) /b/, /d/와 무성파열자음 (unvoiced stop consonant) /p/, /t/ 그리고 유성마찰음 (voiced fricative) /v/, /z/와 무성마찰음 (unvoiced fricative) /f/, /s/ 를 이용하였다. 모음은 /a/를 이용하여 /ba/, /da/, /pa/, /ta/, /va/, /za/, /fa/, /sa/ 와 같은 음성을 제작하여 실험에 사용하였다. 어음처리방식은 시변-선형 필터뱅크 기반 어음처리방식과 시변-비선형 필터뱅크 기반 어음처리방식을 이용하였으며 각 어음처리방식에 적응효과 적용 전후에 음성시작점 정보 전달특성평가를 수행하였다. 8채널로부터 지정된 어음처리방식에 기반한 전기자극펄스를 생성하였고 자극율은 1250 pps로 설정하였다 [3]. 전기자극펄스의 최소값은 441 μA 로 두고 2.5% modulation depth를 갖도록 진폭변조 되었다. 음성이 시작된 부분으로부터 20 ms 동안을 시작점으로 가정하였다 [3]. 분류기의 입력으로 채널당 50개의 spike train을 얻어 5 ms time bin을 취하여 계산된 평균 발화율을 이용하였다. 실험에 사용된 잡음 환경은 잡음이 없는 경우와 10, 5 혹은 0 dB SNR 백색잡음 (white Gaussian noise, WGN) 첨가할 경우 그리고 10, 5 혹은

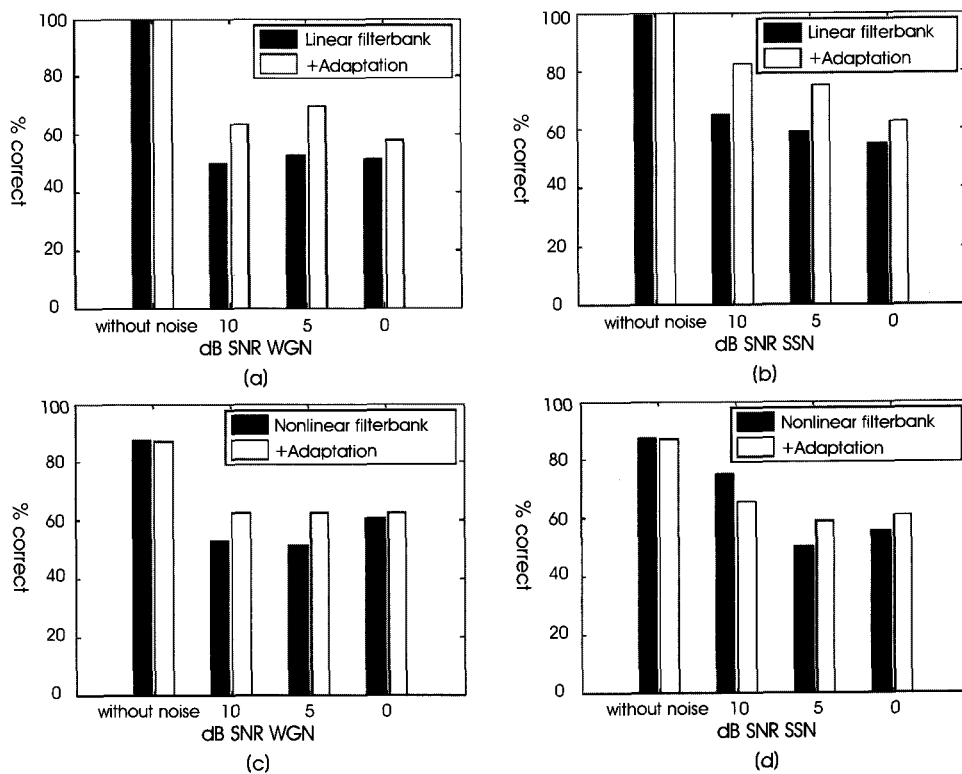


그림 6. 음성 /pa/를 이용한 경우 어음처리방식에 적응효과 적용 전후의 음성시작점 정보 전달특성 비교. (a),(b)는 시변-선형필터뱅크 기반 어음처리방식에 적응효과 적용 전후비교, (c),(d)는 시변-비선형필터뱅크 기반 어음처리방식에 적응효과 적용 전후비교, (a),(c)는 백색잡음 (WGN) 첨가시, (b),(d)는 음성형태잡음 (SSN) 첨가시

Fig. 6. Comparison of speech onset representation characteristic (/pa/). (a),(b) Linear filterbank based speech processing strategy with/without adaptation, (c),(d) Nonlinear filterbank based speech processing strategy with/without adaptation, (a),(c) under WGN, (b),(d) under SSN

0 dB SNR 음성형태 잡음 (음성과 유사한 형태의 스펙트럼을 갖는 잡음 - speech shaped noise, SSN) 첨가 시 등으로 각각의 조건에서 어음처리방식에 적응효과 적용 전후의 성능평가 비교를 수행하였다.

III. 결과

A. 적응효과 적용 전후의 음성시작점 정보 전달특성비교

그림 6,7은 어음처리방식에 적응효과 적용 전후의 음성시작점 정보 전달특성을 비교한 대표적인 그림으로 각각 음성 /pa/와 /za/에 대한 결과이다. 잡음이 없는 경우는 두 가지 어음처리방식 모두 적응효과를 적용하지 않아도 음성시작점의 정보 전달특성은 우수했고 적응효과를 적용할 경우 정보 전달특성은 유사한 수준이거나 우수함을 보였다. 그러나 백색잡음 하에서는 그림 6(a),(c)와 그림 7(a),(c)에서 관찰되듯이 두 어음처리방식에 적응효과를 적용함에 따라 음성시작점의 정보 전달특성이 향상되다. 음성형태잡음 하에서도 적응효과 적용 후 정보 전달특성의 향상을 보였으나 그림 6(b),(d)와 그림 7(b),(d)에서 보이는 바와 같이 일부 잡음레벨에 대하여 감소를 보이기도 하였다. 여러 가지 음성에 대한 결과를 중

합하여 보면 잡음이 없을 경우에는 일반적으로 사용되는 시블면-비선형필터뱅크 기반 어음처리방식을 이용할 경우에도 충분히 우수한 음성시작점 정보 전달특성을 보이며 적응효과를 추가할 경우 전달특성이 유사하거나 약간의 향상 또는 저하됨을 보였다. 그러나 잡음이 있는 경우 음성시작점의 정보 전달특성이 저하됨을 보였다. 잡음 하에서 각 어음처리방식에 적응효과를 적용할 경우 음성의 시작점에서 음성분류성능이 개선되었는데 음성형태 잡음의 경우보다는 백색잡음 하에서 유효하였다.

표 1, 2는 여러 가지 음성에 대하여 어음처리방식에 적응효과를 적용 전후 음성시작점 정보 전달특성으로 각각 백색잡음 하에서, 음성형태 잡음 하에서의 성능을 나타낸 것이다. 잡음 하에서 적응효과를 적용할 경우 음성시작점 정보 전달특성이 향상되는 경우가 많았으며 특히 백색잡음 하에서 적응효과의 이점이 두드러졌다.

IV. 토의

정상인의 청각계에서는 일정 자극이 인가될 경우 내측유모세포와 청신경 사이의 화학적 시냅스에 기원하는 적응효과가 발생하며 [1,3,5] 이를 통하여 음성자극이 시작되는 시점에 대한 정보가 청

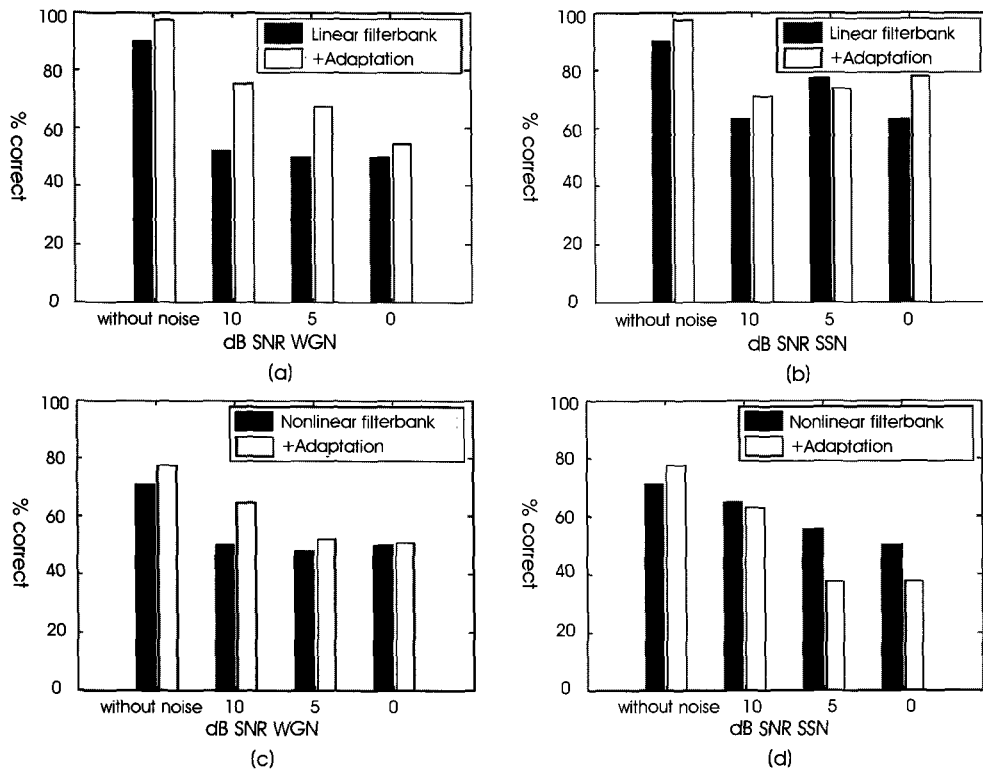


그림 7. 음성 /za/를 이용한 경우 어음처리방식에 적응효과 적용 전후의 음성시작점 정보 전달특성 비교. (a),(b)는 시블면-선형필터뱅크 기반 어음처리방식에 적응효과 적용 전후비교, (c),(d)는 시블면-비선형필터뱅크 기반 어음처리방식에 적응효과 적용 전후비교, (a),(c)는 백색잡음 (WGN) 첨가시, (b),(d)는 음성형태잡음 (SSN) 첨가시

Fig. 7. Comparison of speech onset representation characteristic (/za/). (a),(b) Linear filterbank based speech processing strategy with/without adaptation, (c),(d) Nonlinear filterbank based speech processing strategy with/without adaptation, (a),(c) under WGN, (b),(d) under SSN

자에게 충분하게 전달된다. 인공와우 사용자의 경우 소리정보가 전기자극에 의하여 전달되므로 화학적 시냅스에 의한 적응효과를 기대하기 어렵다 [3,5]. 그러므로 전기자극펄스 생성방법을 결정하는 어음처리방식에 적응효과를 더해주어야 하는데 음성시작점 부분의 전기자극펄스 크기를 증폭하여줌으로써 정상인에게서 보이는 적응효과를 모사할 수 있다 [3,5].

본 연구에서는 음성의 시작점을 강조해주는 적응효과를 어음처리방식에 적용할 경우 음성시작점 구간에서 음성유무 분류성능을 관찰함으로써 음성의 시작점에 대한 정보의 전달이 더 충실하게 제공할 수 있는지 알아보았다. ‘자음+모음’형태의 음성을 이용하여 실험을 수행하였다. 모음의 경우 음성파형의 크기가 자음에 비하여 현저하게 크다. 또한 자음 중에서 파열음, 마찰음 등에 대하여 특히 무성음일 경우 음성파형의 크기가 매우 작으므로 이러한 자음으로 시작되는 음성의 경우 시작점의 강조가 필요하다. 그러므로 여러 종류의 자음을 이용하여 실험을 수행하였다.

실험결과 잡음이 없는 경우는 적응효과 적용 여부에 상관없이 음성시작점의 정보전달 특성이 우수하였다. 즉, 잡음이 없는 경우에는 적응효과가 추가되지 않은 일반적으로 사용되는 시블변-비

선형필터뱅크 기반 어음처리방식을 이용할 경우에도 효과적으로 음성시작점에 대한 정보가 사용자에게 전달될 것으로 기대할 수 있다. 그러나 잡음이 있는 경우 음성시작점 정보 전달특성이 저하되었다. 이는 잡음 하에서는 음성시작점에 대한 정보가 충실하게 전달되지 않아 어음 인지율이 감소하므로 적응효과를 적용함으로써 음성의 시작점에 대한 정보를 보강해 줄 필요가 있음을 보여준다. 잡음 하에서 각 어음처리방식에 적응효과를 적용할 경우 음성의 시작점에서 음성분류성능이 개선되는 경향을 보였다. 따라서 잡음에 강인한 음성정보특성을 가지는 어음처리방식을 이용하여야 하며 적응효과를 추가함으로써 음성시작점에 대한 정보를 전달하고 음절 간 구분에 도움을 줄 수 있을 것으로 판단된다. 이는 선행연구로부터 얻은 결과와도 부합된다 [3,5,6].

일부 음성에 대하여 잡음이 없는 경우 적응효과를 적용할 경우 오히려 정보 전달특성이 약간 저하되기도 하였다. 표 1의 음성 /ba/, /fa/와 표 2의 음성 /ba/, /fa/, /sa/의 경우가 그러한데, 특히 음성 /fa/, /sa/의 경우 음성 파형의 크기가 적은 마찰음 (fricative)으로 어음처리방식에 적응효과를 적용하여도 음성의 시작점을 잘 검출하지 못하여 강조가 이루어지지 않았기 때문으로 판단된다.

표 1. 음성시작점 정보 전달특성 (잡음이 없는 경우와 백색잡음이 첨가된 경우). 단위: (%). 회색으로 표시된 부분은 각 어음처리방식에 적응효과를 적용한 경우의 음성시작점에서의 음성유무 분류성능을 나타냄.

Table 1. Speech onset representation characteristic (without noise, under WGN). Unit: (%). Gray box shows the performance of speech on/off classification using speech processing strategy with adaptation.

Speech processing strategy	kind of noise	Input speech	WGN							
			without noise	10			5			0
				10	5	0	10	5	0	10
Linear filterbank/+ adaptation	/ba/	99.38	75	62.5	65	48	65	50	48.75	
	/da/	87.5	93.75	62.5	58.75	51.25	50.62	50.62	50	
	/pa/	100	100	50	37.5	47.5	50	25	48.13	
	/ta/	99.38	96.25	37.5	61.88	62.5	62.5	50	61.25	
	/fa/	87.5	71.88	44.37	45	32.5	38.12	38.75	46.25	
	/sa/	70	83.75	61.25	81.87	62.5	75	37.5	50	
	/va/	100	100	51.25	50	50.62	62.5	62.5	50.62	
	/za/	89.94	97.48	52.2	75.47	50.31	67.3	50.31	54.72	
Nonlinear filterbank/+ adaptation	/ba/	87.5	85.62	65.63	75.62	62.75	68.1	56.87	61.88	
	/da/	71.25	75.62	61.25	48.75	48.75	55	37.5	50	
	/pa/	84.38	98.12	51.25	48.13	62.5	75	50	45	
	/ta/	80	87.5	58.13	62.5	50	57.5	50	50	
	/fa/	73.12	71.88	38	53.75	50	68.13	50	49.38	
	/sa/	73.75	87.5	62.5	62.5	54.37	62.5	48.75	62.5	
	/va/	74.38	87.5	62.5	50	52.5	62.5	37.5	62.5	
	/za/	71.07	77.36	50.31	64.78	47.8	52.20	50.31	50.94	

표 2. 음성시작점 정보 전달특성 (음성형태 잡음이 첨가된 경우). 단위: (%). 회색으로 표시된 부분은 각 어음처리방식에 적응효과를 적용한 경우의 음성시작점에서의 음성유무 분류성능을 나타냄.

Table 2. Speech onset representation characteristic (under SSN). Unit: (%). Gray box shows the performance of speech on/off classification using speech processing strategy with adaptation.

Speech processing strategy	kind of noise Input speech	SSN					
		10	5	0	0	5	10
Linear filterbank/+ adaptation	/ba/	87.5	73.12	64.38	73.12	62.5	62.5
	/da/	99.38	78.38	88.12	75	75	63.75
	/pa/	62.5	75	75	62.5	75	74.38
	/ta/	71.25	67.37	73.12	86.25	66.25	75
	/fa/	53.13	51.88	50	55.63	50	53.75
	/sa/	75	62.5	71.88	62.5	56.87	50
	/va/	63.12	62.5	71.88	55.63	54.37	38.12
Nonlinear filterbank/+ adaptation	/za/	62.89	71.07	77.36	73.58	62.89	78.05
	/ba/	66.87	62.5	51.25	58.75	33.12	61.88
	/da/	62.5	56.87	52.5	49.38	66.87	50.62
	/pa/	62.5	62.5	46.88	45	50	57.75
	/ta/	62.5	62.5	63.12	54.37	50	49.38
	/fa/	56.25	60	47.5	48.13	58.13	48.75
	/sa/	81.87	62.5	53.75	61.25	49.38	49.38
/va/	62.5	62.5	56.25	73.75	50	44.37	
	/za/	64.78	62.89	55.35	37.74	50.31	37.74

잡음이 있는 경우 어음처리방식에 적응효과를 추가함으로써 보인 음성시작점 정보 전달특성 향상은 음성형태 잡음의 경우보다는 백색잡음 하에서 유효하였다. 이는 실험에 이용된 적응효과 알고리즘을 어음처리방식에 적용할 경우 음성으로부터 추출된 포락선을 관찰하였을 때 음성형태 잡음 하보다는 백색잡음 하에서 음성의 시작점을 잘 검출하여 강조시켜주기 때문으로 여겨진다. 따라서 음성형태 잡음 하에서도 적응효과의 이점을 충분히 발휘하기 위해서는 적응효과 알고리즘을 개선할 필요가 있다고 판단된다.

참고문헌

[1] M. F. Bear. et al., *Neuroscience: Exploring the Brain*, 2nd edition, Lippincott Williams & Wilkins, pp. 359-385, 2001.

[2] 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.

[3] L. Geurts and J. Wouters, "Enhancing the speech envelope of continuous interleaved sampling processors for cochlear implants," *J. Acoust. Soc. Am.*, vol. 105, pp. 2476-2484, 1989.

[4] R. L. Smith, "Adaptation, saturation, and physiological masking in single auditory-nerve fibers," *J. Acoust. Soc. Am.*, vol. 65, pp. 166-178, 1979.

[5] Jin Ho Kim, "A novel speech processing strategy for cochlear implants based on nonlinear model of auditory periphery", MS Thesis, Yonsei University Graduate School, Feb. 2007

[6] J. H. Kim, K. H. Kim, "A simulation study on improvements of speech processing strategy of cochlear implants using adaptation effect of inner hair cell and auditory nerve synapse," *J. Biomed. Eng. Res.*, vol. 28, pp. 205-211, 2007.

[7] 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.

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

[9] S. Kumar, "Neural networks: a classroom approach," *McGraw Hill*, pp. 137-142, 2005.

[10] D. H. Kim, J. H. Kim, K. K. Kim, "Performance evaluation of cochlear implants speech processing strategy using neural spike train decoding," *J. Biomed. Eng. Res.*, vol. 28, pp. 271-279, 2007.

- [11] 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.
- [12] 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," in *Proc. World congress on Medical Physics and Biomedical Engineering, Seoul, Korea, 2006*.
- [13] 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.
- [14] 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 electrically stimulated auditory nerve: pulse-train response," *IEEE Trans. Biomed. Eng.*, vol. 46, no. 6, pp. 630-637, 1999.