

# 청각신경 시냅스의 적응 효과를 이용한 인공와우 어음처리 알고리즘의 개선에 대한 시뮬레이션 연구

김진호, 김경환

연세대학교 보건과학대학 의공학과

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

## A Simulation Study on Improvements of Speech Processing Strategy of Cochlear Implants Using Adaptation Effect of Inner Hair Cell and Auditory Nerve Synapse

Jin Ho Kim, Kyung Hwan Kim

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

### Abstract

A novel envelope extraction algorithm for speech processor of cochlear implants, called adaptation algorithm, was developed which is based on a adaptation effect of the inner hair cell(IHC)/auditory nerve(AN) synapse. We achieved acoustic simulation and hearing experiments with 12 normal hearing persons to compare this adaptation algorithm with existent standard envelope extraction method. The results shows that speech processing strategy using adaptation algorithm showed significant improvements in speech recognition rate under most channel/noise condition, compared to conventional strategy. We verified that the proposed adaptation algorithm may yield better speech perception under considerable amount of noise, compared to the conventional speech processing strategy.

Key words : cochlear implant, speech processor, inner-hair cell, auditory nerve, adaptation

### 1. 서 론

인공와우(cochlear implant)는 감각신경성 난청환자의 청신경을 전기적으로 직접 자극하는 방법을 사용하여 청력을 회복시키는 장치이다. 인공와우는 외부소리를 입력받는 마이크, 입력받은 소리를 전기적 펄스열로 변환하는 신호처리부, 전기적 펄스열을 전달하기 위한 송/수신부, 그리고 와우 내부의 청신경 근처에 삽입되는 전극으로 구성된다. 신호처리부는 외이에서 내이까지의 정상청각계의 기능을 모방하는 역할을 하며, 일단 외과적 수술을 통해 전극을 삽입한 이후에는 성능개선이 가능한 유일한 부분이다. 인공와우의 성능개선에 있어서 핵심적인 요소이다. continuous interleaved sampling(CIS) 방식은 현재 인공와우 시스템에서 가장 많이 쓰이는 어음처리방식이다[1]. 선형대역통과필터뱅크와 정류기와 저역통과필터로 구성된 포락선(envelope) 추출

방식에 기반한 CIS 방식은 잡음이 없는 환경에서 평균적으로 높은 인지율을 보인다. 하지만 잡음 하에서는 이러한 인지율은 현저히 감소한다. 정상 청각자가 높은 수준의 잡음 하에서도 높은 어음인 지율을 보임을 고려했을 때 이러한 차이는 인공와우의 신호처리부가 잡음 하에서도 음성의 여러 특성을 효과적으로 추출해내는 정상 청각계의 세부적 특성을 면밀히 모방하지 못하였기 때문이라 할 수 있다. 예를 들어 CIS 방식에서 사용하는 선형필터뱅크로는 입력음성의 크기에 따른 주파수 응답 가변 특성을 표현할 수 없다. 이러한 level-dependent 특성은 특히 잡음 하에서 모음의 포먼트(formant) 추출에 중요한 역할을 한다고 알려져 있다[2].

뿐만 아니라 대역통과필터의 뒤에 연결되어 전극에 전달할 자극의 크기를 구하는 포락선 추출단은 기저막의 내측유모세포(inner hair cell; IHC)와 청신경의 시냅스에서 일어나는 적응효과(adaptation effect)를 표현할 수 없다. 적응효과는 음성의 개시점을 명확히 해주고 풍부한 시간적 단서를 청신경에 전달함으로써 잡음환경에서 감소하는 음성의 특징을 두드러지게 표현하는데 중요한 역할을 한다[3]. 현재의 인공와우의 신호처리부는 정상 청각계의 아주 개략적인 기능만을 표현하므로, 잡음에 강인하며 음성의 특징을 더욱 효과적으로 표현할 수 있는 인공와우의 신호처리부를 개

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

Corresponding Author : 김경환

강원도 원주시 흥업면 매지리 234 연세대학교 보건과학대학 의공학과

Tel : +82-33-760-2364 / Fax : +82-33-760-1953

E-mail : khkim0604@yonsei.ac.kr

말하기 위해서는 정상 청각계를 더욱 유사하게 모방하는 것이 필수적이다[4]. 본 연구에서는 특히 청신경의 시냅스에서 발생하는 적응효과를 인공와우의 신호처리부에 적용하기 위해 기존의 포락선 추출방식을 대신할 새로운 적응효과 알고리즘을 개발하였다.

청신경의 적응효과는 일정한 크기를 갖는 자극이 입력될 때 발생하는 청신경의 응답 변화를 의미한다. 자극의 시작시점에서 청신경의 응답 즉, 발화율은 최대가 되며 이후 발화율은 점차 감소하면서 안정화된다[5]. 이러한 적응효과는 달팽이관의 IHC과 청신경 사이의 시냅스에서 신경전달물질의 소모에 의한 고갈(depletion)에 의해 발생한다[6]. 내이로 입력되는 음성의 시간에 따른 크기와 스펙트럼의 변화는 청신경의 발화패턴에 의해 인코딩되어서 위 신경계로 전달되며, 적응효과는 음성의 인코딩 과정에서 중요한 역할을 한다[3].

적응효과는 음성의 시작시점에 관한 보다 많은 정보를 전달함으로써 연속되는 음성의 음질을 명확히 구분하는 역할을 한다. 예를 들어, 마찰음에 해당하는 자음은 시-주파수 영역에서 짧은 시간 동안 모든 대역에서 작은 에너지를 갖는 스펙트럼 특성을 보인다. 이러한 마찰음의 음성정보는 잡음이 첨가되었을 때 쉽게 사라지게 된다. 따라서 적응효과는 잡음 하에서 이러한 단서를 강인하게 유지하는 역할을 한다. 또한 모음의 경우에도 포먼트가 시작되거나 포먼트 변이(formant transition)가 일어나는 시-주파수 정보를 잡음 하에서도 두드러지게 표현한다. 이와 같이 적응효과는 청신경이 장기간 높은 발화율을 유지할 수 없기 때문에 발생하는 현상이 아니라 음성의 시간에 따른 크기 및 주파수 변화에 따른 단서를 더욱 효과적으로 추출하는 역할을 한다. Tchorz 등은 적응효과를 적용한 청각모델을 자동음성인식 장치의 프론트-엔드(front-end)에 사용함으로써 특히 잡음 하에서 강인한 음성표현 능력을 가짐을 보였다[7].

본 연구에서는 기존의 포락선 추출단을 개선하여 음성의 개시점이 효과적으로 추출되는 간략하고 현재 인공와우의 신호처리부에 쉽게 적용이 가능한 적응효과 알고리즘을 개발하였으며 개발된 적응효과 알고리즘이 적용된 어음처리방식을 기존의 포락선 추출방식을 사용한 어음처리방식과 시-주파수 분석 및 음향시뮬레이션(acoustic simulation)을 이용한 청취실험을 이용해 비교하였다.

## II. 어음처리방식

그림 1은 일반적인 어음처리방식의 간략도이다. 마이크로부터 입력된 음성은 서로 다른 중심주파수를 가진 다채널 대역통과필터를 통과하고, 대역통과필터를 통과한 신호는 포락선 추출단을 통과한 뒤 청각 다이내믹 레인지를 압축하기 위한 압축단으로 입력된다. 다이내믹 레인지 압축은 인공와우 시술 후 환자 개개인 별로 피팅하는 것이므로 본 논문에서 압축단은 고려하지 않았다. 추출된 포락선 신호는 800 Hz 이상의 높은 펄스율을 가진 interleaved biphasic pulse로 변조된 뒤 전극으로 보내지게 된다. 이러한 interleaved sampling 방식으로 전극간의 상호작용에 의한 신호의 왜곡을 막을 수 있다[1].

우리는 주파수 분해(frequency decomposition)를 위해 Meddis가 제시한 두 개의 병렬통로로 구성되며, 기저막의 비선형-시변 특성을 효과적으로 표현하는 dual resonance nonlinear(DRNL) 필터뱅크를 사용하였다[8]. 이 필터뱅크는 기저막의 비선형-시변 특성, 특히 입력음성의 크기에 따라 주파수 응답특성이 가변하는 특성을 효과적으로 표현하는 모델로써 이전 연구결과에서 이 모델이 선형 시불변 필터뱅크에 비해 더욱 효과적으로 음성의 특징점을 추출하는 것이 확인되었다[9].

이러한 사실을 근거로 본 연구에서는 DRNL 모델을 주파수 분해를 위해 사용하였으며 포락선 추출방식의 비교를 위해 두 가지 방식을 사용하였다. 첫 번째는 현재 인공와우의 신호처리부에서 사용되는 포락선 추출방식으로서 반과정류기와 그 뒷단에 연결되는 400 Hz의 차단 주파수를 갖는 저역통과필터(4차 버터워스)를 사용하는 것이다. 이러한 방식으로 각 주파수 대역별로 항상 양의 값을 갖는 평활화(smoothing)된 출력신호를 얻을 수 있다. 이러한 평활화 및 양의 값으로의 변환은 외부 소리자극이 내이로 입력되었을 때 일어나는 실제 기저막의 응답특성을 반영한다. 하지만 이러한 방법만으로는 기저막의 응답이 청신경에 전달될 때 일어나는 적응효과를 표현할 수 없으므로 우리는 적응효과가 적용된 새로운 알고리즘을 개발하였다. 그림 2 (a)는 하나의 채널에 대한 적응효과가 적용된 포락선 추출 알고리즘에 대한 블록 다이어그램이

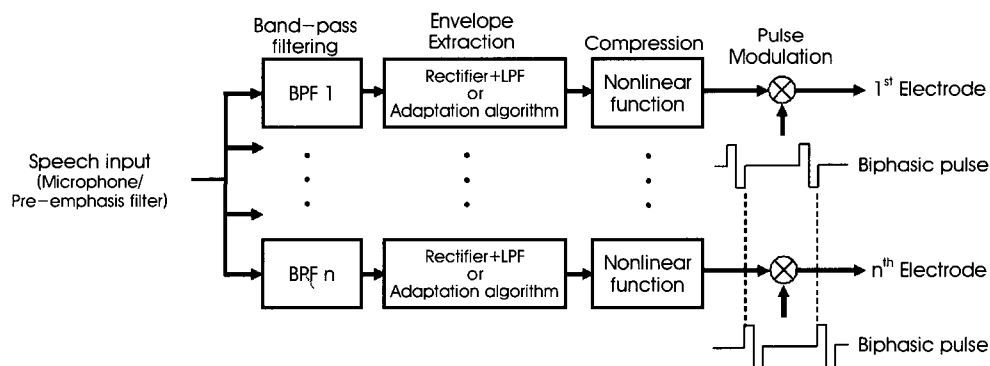


그림 1. 일반적인 어음처리방식의 간략도  
Fig. 1. Block diagram of a general CI speech processor

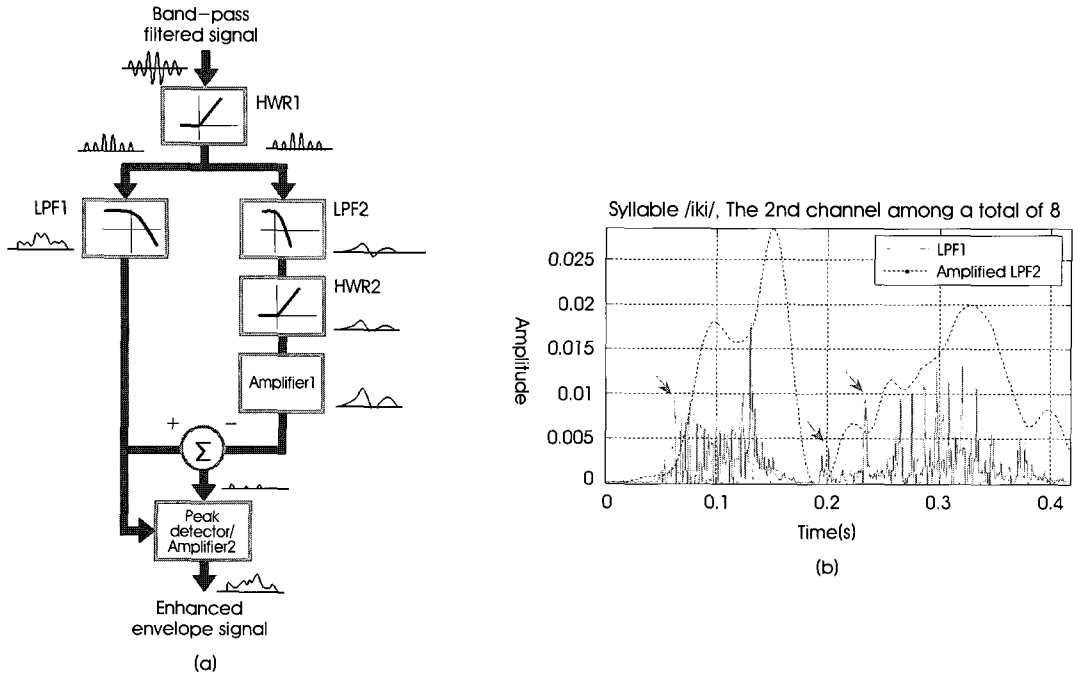


그림 2. (a) Adaptation 알고리즘의 간략도 (b) 총 8 채널 중 음성 /iki/에 대한 각 통로의 2 번째 채널의 엔벨로프 신호  
 Fig. 2. (a) Block diagram of the proposed adaptation algorithm (b) Envelope signal from the 2nd channel of pathway

며 모든 채널에 동일하게 적용된다. 좌측 통로는 일반적인 포락선 추출방식을 의미한다. 먼저, 반파 정류된 신호는 저역통과필터 (low-pass filter; LPF)에 의해 smoothing 된다. (LPF1 : 400Hz 차단 주파수를 갖는 4차 버터워스 필터) 좌측 통로의 출력은 뺄셈기(Σ) 및 피크 검출기/증폭기2로 입력된다.

오른쪽 통로는 저역통과 필터와 반파 정류기, 증폭기의 순서로 구성된다. 저역통과 필터 (LPF2: 20Hz 차단 주파수를 갖는 4차 버터워스 필터)는 좌측통로의 LPF1보다 상대적으로 낮은 주파수를 가지므로 좌측 통로에 비해 상대적으로 훨씬 느리게 변하는 포락선이 추출되며 이 포락선 신호는 반파 정류된 후 증폭기1에서 적절한 크기로 증폭된 후 뺄셈기로 입력된다. 증폭기1의 이득 값은 6을 사용하였다. 그림 2 (b)는 /iki/를 입력음성으로 하였을 때 총 8 채널 중 2번째 채널에서 뺄셈기로 입력되는 두 통로의 출력 신호를 보여준다. 실선은 왼쪽 통로의 출력을 의미하고, 점선은 오른쪽 통로의 출력을 나타낸다. LPF2가 LPF1보다 상대적으로 더 큰 지연 시간을 가지므로 음성이 시작되거나 포먼트 변화가 일어나는 등 신호의 크기가 크게 증가하는 시점에서 왼쪽 통로의 출력은 항상 오른쪽 통로의 출력보다 크다. 따라서 뺄셈기로 이 두 신호가 입력 되면 그림 2 (b)의 화살표 부분과 같이 음성의 개시점에 해당하는 피크가 추출되며 이 피크 값들은 피크 검출기/증폭기2로 입력된다. HWR2는 잘못된 피크 발생을 방지하는 역할을 한다. LPF2의 시정수 (time constant)가 매우 크기 때문에 신호의 크기가 급격히 감소하는 구간에서 LPF2의 출력은 음의 값이 되며 이 값이 뺄셈기로 입력되었을 경우 잘못된 양의 값을 갖는 피크가

추출된다. 따라서 이러한 피크 발생을 방지하기 위해 0 이하의 값을 제거하는 반파정류기를 사용하였다. 피크 검출기/증폭기2 단에서는 매 1 ms 마다 피크 값의 평균을 계산하고 이 값이 0보다 크면 피크 값에 이득을 적용한 후 왼쪽 통로의 포락선과 더해진 것이 최종적으로 출력되고 0보다 작을 때는 왼쪽 통로의 포락선이 그대로 출력된다.

### III. 평가방법 : 인공와우의 음향시뮬레이션

음향시뮬레이션은 인공와우 사용자들이 인지할 것으로 예상되는 소리를 인공와우 신호처리부에서 사용되는 어음처리방식에 따라 합성하여 정상 청각자에게 들려주는 방식을 뜻한다. 이러한 방법으로 남아있는 유모세포의 양, 전극이식 조건 등의 환자의 개인별 요소를 배제한 어음처리방식 만의 평가가 가능하다. 음향시뮬레이션을 이용한 음성합성방법으로는 크게 두 가지 방법이 사용된다. 하나는 정현파를 이용해 음성을 합성하는 sinusoidal modulation 방식이며[10], 다른 하나는 대역제한 잡음을 사용해 음성을 합성하는 noiseband modulation 방식이다[11]. 본 연구에서는 sinusoidal modulation 방법만을 사용하였다. 채널 별 중심주파수는 Loizou 등이 제시한 값을 사용하였다(표 1 참조)[12].

#### A. 신호처리 방법

음향시뮬레이션을 위한 일반적인 어음처리방식의 sinusoidal modulation은 식 (1)에 따라 이루어진다.

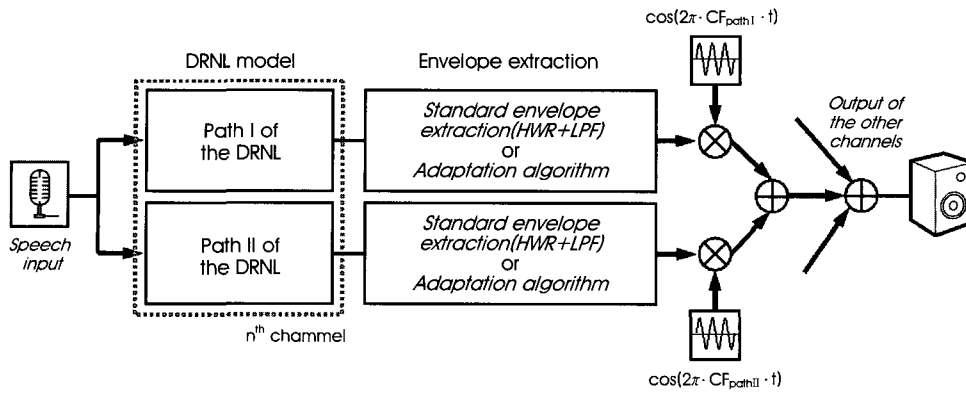


그림 3. 인공와우의 어음처리방식에 기반한 음향시뮬레이션 방법  
 Fig. 3. Acoustic simulation of CI speech processing strategy

$$s(t) = \sum_{k=1}^N A_k(t) \cos(2\pi f_k t) \quad (1)$$

식 (1)에서  $s(t)$ 는 합성된 음성을,  $k$ 는 채널 숫자를,  $N$ 은 총 채널 수를,  $A_k(t)$ 는  $k$ 번째 채널에서의 포락선을,  $f_k$ 는  $k$ 번째 채널의 중심주파수를 의미한다. 포락선  $A_k(t)$ 는 각 채널의 중심주파수에 해당하는 정현파와 변조된 후 모든 채널의 신호와 더해짐으로써 최종적으로 출력된다. 그림 3은  $n$ 번째 채널에서 sinusoidal modulation 방식을 사용한 음향시뮬레이션 방법에 대한 간략도를 보여 준다. 포락선 추출방식의 비교가 목적이므로 대역통과 필터뱅크로는 앞서 말했던 DRNL 모델을 사용하였다. DRNL 모델은 두 개의 통로로 구성되므로 각각의 통로에 포락선 추출단을 연결하였으며 채널의 출력은 두 통로의 합으로 구해지며 최종 음성 합성은 모든 채널의 출력을 다시 더함으로써 이루어졌다. 포락선 추출단에는 반파정류기와 400 Hz 저역통과필터(4차 버터워스)로 구성된 기존의 포락선 추출 방식과 본 연구에서 제시한 적응효과 알고리즘을 사용하였다.

**B. 음향시뮬레이션을 이용한 청취실험**

기존의 포락선 추출방식과 제안된 적응효과 알고리즘의 비교를

위해 음향시뮬레이션을 이용해 합성된 음성을 이용한 청취실험을 수행하였다. 잡음 조건은 잡음이 없는 경우와 5 dB/10 dB SNR의 가우시안 백색잡음(Gaussian white-noise; GWN) 및 음성형태의 스펙트럼을 갖는 잡음(speech-shaped noise; SSN)이 첨가되었을 때의 총 3 가지 조건이 사용되었다. SSN은 화자가 많은 식당, 카페와 같은 장소에서 발생하는 균중잡음을 의미하며, 이러한 잡음을 입력음성에 첨가해줌으로써 실제 대화환경과 유사한 환경을 만들 수 있다. SSN은 가우시안 백색잡음(white noise)을 음성의 주파수 대역만을 가지도록 1100 Hz의 차단주파수를 갖는 저역통과필터(2차 버터워스)에 통과시킴으로써 얻을 수 있다[13]. 또한 SNR은 신호의 실효값 기준으로 계산되었다.

**실험단어**

/자음-모음-자음/, /모음-자음-모음/ 형태로 구성된 음성을 사용하였으며 이러한 형태는 중간음절의 인지를 테스트하기 위해 사용된다. 모음 인지 테스트를 위해 /hVd/, /gVd/, /bVd/의 3 가지 종류의 컴퓨터 합성음이 사용되었으며, 중간모음으로는 /a/, /u/, /i/를 사용하였다. 또한 자음 인지 테스트를 위해 /aCa/, /uCu/, /iCi/의 3가지 종류의 컴퓨터 합성음이 사용되었고, 중간자음으로는 /k/, /p/, /t/가 사용되었으며 인지 테스트를 위해 사용된 총 음성수

표 1. 채널 별 중심 주파수 (Hz)

Table. 1. Center frequencies of each channels

총 채널 수	채널															
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
2	792	3392														
3	545	1438	3793													
4	460	953	1971	4078												
5	418	748	1339	2396	4287											
6	393	639	1037	1685	2736	4444										
8	394	692	1064	1528	2109	2834	3740	4871								
10	322	546	814	1137	1524	1988	2545	3213	4014	4976						
12	274	453	662	905	1190	1521	1908	2359	2885	3499	4215	5050				
16	216	343	486	647	828	1031	1260	1518	1808	2134	2501	2914	3378	3901	4489	5150

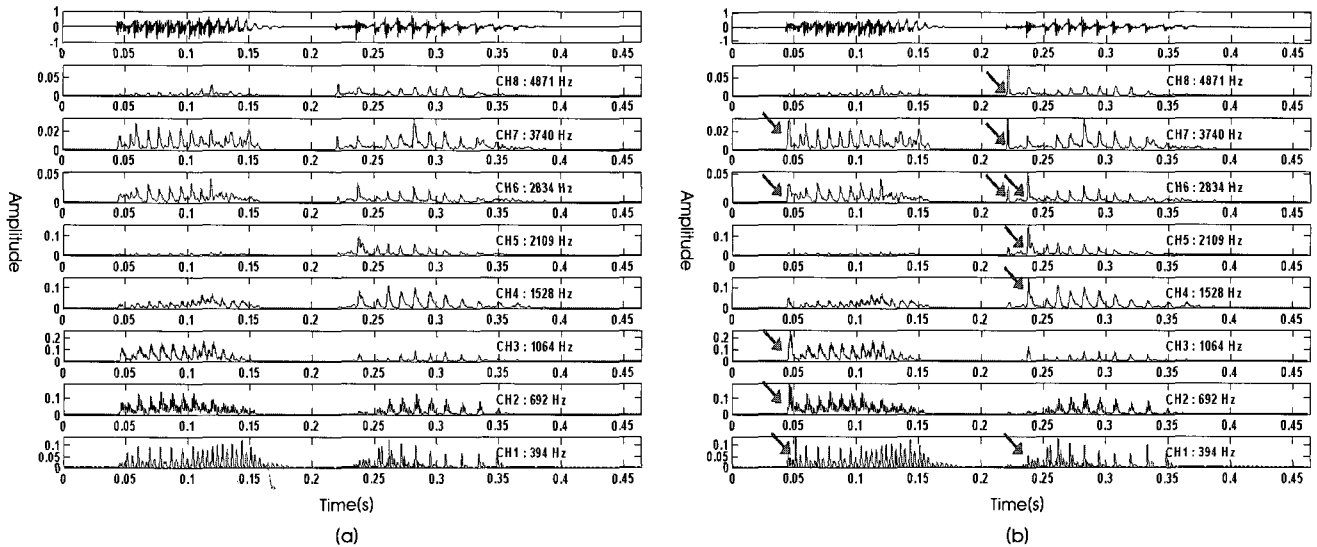


그림 4. (a) 기존의 포락선 추출방식을 이용하였을 때 입력음성 /aka/에 대한 8 채널 포락선 신호  
 (b) Adaptation 알고리즘을 이용하였을 때 입력음성 /aka/에 대한 8 채널 포락선 신호  
 Fig. 4. (a) 8 channel envelope signals obtained from input speech /aka/  
 (b) The proposed envelope extraction algorithm using adaptation

는 18 가지이다. 모든 합성음은 \*.wav 형태(16 bit mono analog-to-digital conversion, sampling rate: 22.050 kHz)로 저장된 뒤 청취실험을 위해 4 채널, 8 채널, 12 채널에 대해서 각 잡음 조건별로 합성되었다.

**Experiment Protocol**

총 12 명의 대학생들이 청취실험에 자발적으로 참가하였으며, 28.7±1.07세의 정상청각자로 구성되었다. 피험자는 펜티엄 PC에 장착된 음향카드(SoundMAX Integrated Digital Audio)의 출력 단자에 연결된 폐쇄형 헤드폰(HD25SP1, Sennheiser)을 통해 합성음을 청취하게 되며 볼륨은 실험 전 피검자 별로 불편함을 느끼지 않는 레벨로 매 실험마다 조절되었다. 검사에 임하기 전 합성음에 적응하기 위해 5 분간의 훈련시간이 주어졌다. 실험은 각 어음 처리방식에 따라 합성된 음성을 청취한 뒤 인지한 소리에 해당하는 단어를 설문지에 있는 보기에서 선택하는 방식으로 이루어졌다. 음성처리 방식에 대한 정보는 전혀 주어지지 않았으며, 합성음이 제시되는 순서는 모든 경우에 대해 임의의 순서로 배열되었다. 문제의 정답에 대한 피드백은 주어지지 않았다.

**IV. 결 과**

청각신경 시냅스에서 일어나는 적응효과가 적용된 새로운 포락선 추출방식인 적응효과 알고리즘이 각 채널의 주파수 영역별로 음성의 개시점을 효과적으로 강조하는지 관찰하기 위해 채널별 포락선 신호를 관찰하였다. 그림 4는 기존의 포락선 추출방법과 개발된 적응효과 알고리즘을 사용하였을 때 입력음성 /aka/에 대해 8 채널의 포락선 신호를 관찰한 것이다. 그림 4 (a)는 일반적인 포락

선 추출방법을 사용하였을 때의 결과이며 그림 4 (b)는 적응효과 알고리즘을 사용하였을 때의 결과이다. 맨 위의 그래프는 음성 /aka/의 시간에 따른 음성파형을 나타내며 화살표는 적응효과에 의해 강조가 일어난 시점을 의미한다. 그림에서 모음-자음-모음으로 이루어진 입력음성의 시작시점이 강조가 된 것을 쉽게 볼 수 있다. 이러한 강조점들은 해당 채널에 해당하는 전극으로 보내질 전기적 펄스열과 변조되어 해당 전극 근처에 위치한 청신경에 더욱 명확한 음성학적 단서를 제공한다. 자음, 특히 마찰음, 파열음 등은 모음에 비해 훨씬 작은 에너지를 가지면서 스펙트럼 특성이 자음과 유사하여 적응효과를 사용하지 않을 경우 잡음 하에서 이러한 자음의 인지는 어렵다. 따라서 그림 4 (b)에서 0.22 초 근방에서 고주파 영역에 해당하는 6, 7, 8 채널에서 일어난 자음의 시간적 단서 강조는 IHC과 청신경 사이의 시냅스에서 일어나는 적응효과를 적용한 포락선 추출방식이 특히 잡음 하에서의 자음인지에 큰 도움을 줄 수 있음을 보여준다.

그림 5는 일반적인 포락선 추출방식에 기반한 어음처리방식과 개발된 적응효과 알고리즘에 기반한 어음처리방식의 음향시물레이션 청취실험 결과를 나타낸다. (a)는 잡음이 없는 경우의 결과이며 (b)와 (c)는 각각 GWN(5 dB/10 dB SNR)과 SSN(5 dB/10 dB SNR)이 첨가된 경우의 결과이며, 가로축은 채널의 수를 나타내고, 세로축은 어음인지를 백분율로 나타낸 것이다. 실선 그래프는 적응효과 알고리즘에 기반한 어음처리방식을, 점선 그래프는 기존의 포락선 추출방식에 기반한 어음처리방식의 결과를 뜻한다. 잡음이 없는 경우에도 모든 채널 조건에서 기존의 포락선 추출방식보다 개발된 적응효과 알고리즘을 사용한 어음처리방식이 평균적으로 높은 어음인지를 보였다. 또한 t-test 결과 4 채널을 제외

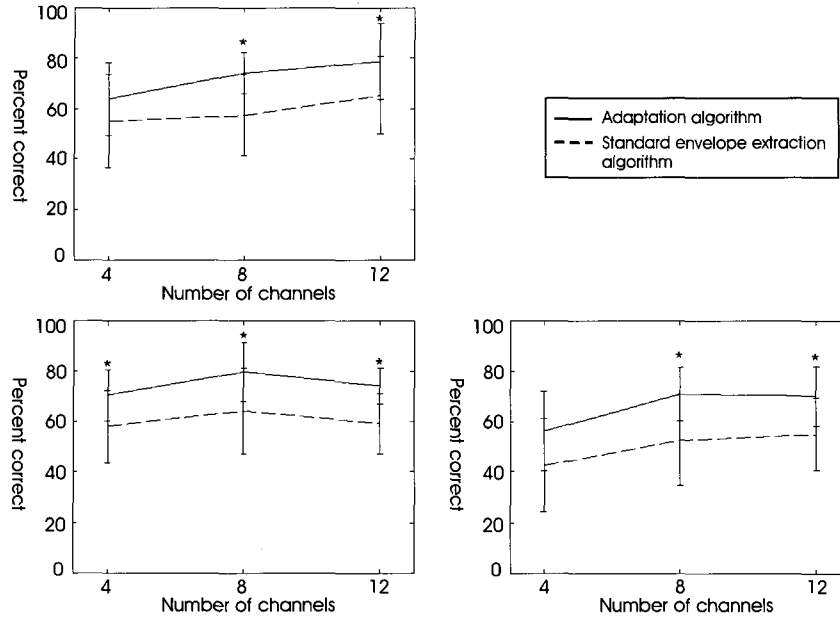


그림 5. 음향시뮬레이션을 이용한 청취실험 결과 (a) 잡음이 없는 경우 채널 조건에 따른 청취실험 결과 (b) GWN이 첨가된 경우 채널 조건에 따른 청취실험 결과 (c) SSN이 첨가된 경우 채널조건에 따른 청취실험 결과  
 Fig. 5. Results of hearing experiments using acoustic simulation. (a) Under silent condition. (b) Under WGN. (c) Under SSN

한 모든 채널에서 통계적으로 유의한 차이를 보였다( $p < 0.05$ ).

GWN이 첨가된 경우 두 어음처리방식 모두 전체적인 인지율은 잡음이 없는 경우에 비해 감소했으나 적응효과 알고리즘을 적용한 어음처리방식의 어음인지율이 모든 채널조건에서 평균적으로 높았으며, t-test 결과 4 채널을 제외한 모든 채널조건에서 유의한 차이를 보였다. SSN이 첨가된 경우에도 적응효과 알고리즘을 사용한 어음처리방식의 인지율이 모든 조건에서 평균적으로 높았으며, 대부분의 채널조건에서 통계적으로 유의한 차이를 보였다. 표 2는 채널조건 및 잡음종류에 따른 t-test 결과이다.

잡음이 첨가된 경우 12 채널에서의 인지율이 두 가지 어음처리방식 모두 8 채널에 비해 같거나 또는 오히려 약간 감소된 특성을 보이고 있다. 청취실험에서는 실험단어로 단음절 단어를 사용하였으며 이 경우 대화에 주로 쓰이는 길이가 긴 단어 및 문장 등에 비해 극히 제한된 음성정보만을 표현하며, 이러한 특성은 잡음이 첨가된 경우 더욱 두드러질 것이다. 따라서 12 채널에서의 음성정보가 8 채널에서의 음성정보와 유사하기 때문에 이와 같은 인지특성이 보였다고 판단된다. 또한 음향시뮬레이션 청취실험은 청각심리적 실험으로서 피험자의 주관적인 판단에 기인하며 반드시 채널

수의 증가 즉, 주파수 해상도의 증가가 항상 인지율의 증가를 보이지는 않는다.

## V. 토의 및 결론

본 연구에서는 IHC과 청신경의 시냅스에서 일어나는 적응효과를 인공와우의 어음처리방식에 적용함으로써 특히 잡음 하에서의 성능개선이 가능함을 보이고자 하였다. 적응효과는 IHC과 청신경 사이의 시냅스에서 일어나는 비선형 특성으로서 청신경에 보다 풍부한 시간적 단서를 제공하며, 음성의 개시점을 강조하는 역할을 한다[3]. Tchorz 등은 이 적응효과를 자동 음성인식 시스템의 프론트-엔드에 사용함으로써 특히 잡음 하에서 향상된 음성인지 특성을 가짐을 보였다[7]. 적응효과 IHC과 청신경의 화학적 연결의 결과로써 현재 인공와우 시스템에서 이 특성은 적용되지 않는다. 따라서 우리는 현재 인공와우의 신호처리부에서 사용되는 포락선 추출단을 개선하여 음성의 개시점을 강조해주는 새로운 알고리즘을 개발하였다. 이 알고리즘을 사용하여 각 채널의 포락선에서 음절의 개시점에 해당하는 시간적 단서가 일반적인 포락선 추출방식에 비해 더욱 두드러지게 강조되는 것을 확인하였다. 특히 모음에

표 2. 음향시뮬레이션을 이용한 청취실험의 t-test 결과

Table 2. t-test results of hearing experiment using acoustic simulation

채널 수	잡음 종류	잡음이 없는 경우	GWN	SSN
4 채널		$p = 0.2049$	$p = 0.0705$	$p = 0.0220 *$
8 채널		$p = 0.0043 *$	$p = 0.0067 *$	$p = 0.0176 *$
12 채널		$p = 0.0423 *$	$p = 0.0111 *$	$p = 0.0012 *$

비해 높은 주파수 영역 및 넓은 대역에서 작은 에너지를 갖는 자음의 스펙트럼 특성은 잡음의 스펙트럼 특성과 유사해 잡음이 첨가된 경우 이러한 단서는 사라지기 쉽다. 적응효과 알고리즘을 이용해 이러한 단서를 잡음 하에서도 유지할 수 있었다.

인공와우의 음향시뮬레이션은 인공와우 사용자를 대상으로 실험하기 전 음성처리방식의 평가를 위해 쓰이는 가장 일반적인 방법이며 각각의 어음처리방식에 따라 합성된 음성을 정상 청각자에게 들려주는 방식으로 청취실험이 이루어진다. 기존의 어음처리방식과 제안된 적응효과 알고리즘에 기반한 어음처리방식의 비교를 위한 /자음-모음-자음/ 및 /모음-자음-모음/의 청취실험을 수행하였으며, 실험 결과 모든 잡음 및 채널 조건에서 제안된 방식이 훨씬 우수한 선호도를 보였다. 이는 청신경의 시냅스에서 일어나는 적응효과를 고려한 제안한 어음처리방식을 사용함으로써 인공와우 사용자들에게서 어음인지 능력이 크게 향상될 수 있다는 것을 의미한다.

현재 인공와우에서 쓰이는 어음처리방식은 아주 기초적이고 간략화된 청각계의 기능만을 표현한다. 인공와우 사용자들은 잡음이 없는 환경에서는 의사소통에 큰 무리가 없지만, 거리나 식당과 같은 주변 잡음이 많이 존재하는 환경에서 의사소통은 거의 불가능하다. 정상 청각자가 이러한 환경에서도 높은 어음인지율을 보이는 점을 감안할 때, 정상 청각계를 더욱 면밀히 모방한 음성처리방식의 개발은 필수적이다. 전극의 개발 및 자극 방법의 개발과 더불어 청각계의 기능이 면밀히 모방된 음성처리방식의 개발은 인공와우의 성능을 더욱 개선시킬 것이다.

## 참고문헌

- [1] B. S. Wilson, C. C. Finley, D. T. Lawson, R. D. Wolford, D. K. Eddington, W. M. Rabinowitz, "Better speech recognition with cochlear implants," *Nature*, vol. 352, pp. 236-238, 1991.
- [2] L. Deng, C. D. Geisler, "A composite auditory model for processing speech sounds," *J. Acoust. Soc. Am.*, vol. 82, pp. 2001-2012, 1987.
- [3] B. Delgutte, N. Y. S. Kiang, "Speech coding in the auditory nerve: IV. Sounds with consonant-like dynamic characteristics," *J. Acoust. Soc. Am.*, vol. 75, pp. 897-907, 1984.
- [4] B. S. Wilson, R. Schatzer, E. A. Lopez-Poveda, X. Sun, D. T. Lawson, R. D. Wolford, "Two new directions in speech processor design for cochlear implants," *Ear Hear.*, vol. 26, pp. 73S-81S, 2005.
- [5] J. R. Cohen, "Application of an auditory model to speech recognition," *J. Acoust. Soc. Am.*, vol. 85, pp. 2623-2629, 1989.
- [6] R. L. Smith, "Adaptation, saturation, and physiological masking in single auditory-nerve fibers," *J. Acoust. Soc. Am.*, vol. 65, pp. 166-178, 1979.
- [7] J. Tchorz, B. Kollmeier, "A model of auditory perception as front end for automatic speech recognition," *J. Acoust. Soc. Am.*, vol. 106, pp. 2040-2050, 1999.
- [8] R. Meddis, L. P. O'Mard, E. A. Lopez-Poveda, "A computational algorithm for computing nonlinear auditory frequency selectivity," *J. Acoust. Soc. Am.*, vol. 109, pp. 2852-2861, 2001.
- [9] J.H. Kim, K. H. Kim, Improvement of cochlear implant speech processor based on nonlinear time-varying filterbank model of biological cochlea, *The 33rd annual fall meeting of KOSOMBE*, pp. 93, 2005
- [10] M. F. Dorman, P. C. Loizou, D. Rainey, "Speech intelligibility as a function of the number of channels of stimulation for signal processors using sine-wave and noise-band outputs," *J. Acoust. Soc. Am.*, vol. 102, pp. 2403-2411, 1997.
- [11] R. V. Shannon, F. G. Zeng, V. Kamath, J. Wygonski, M. Ekelid, "Speech recognition with primarily temporal cues," *Science*, vol. 270, pp. 303-304, 1995.
- [12] P. C. Loizou, M. Dorman, Z. Tu, "On the number of channels needed to understand speech," *J. Acoust. Soc. Am.*, vol. 106, pp. 2097-2103, 1999.
- [13] L.-P. Yang, Q.-J. Fu, "Spectral subtraction-based speech enhancement for cochlear implant patients in background noise," *J. Acoust. Soc. Am.*, vol. 117, pp. 1001-1004, 2005.