

인공와우 어음처리방식을 위한 적응효과 알고리즘의 음성개시점 검출 특성 비교

최성진, 김진호, 김경환

연세대학교 보건과학대학 의공학과

Comparison of Speech Onset Detection Characteristics of Adaptation Algorithms for Cochlear Implant Speech Processor

Sung Jin Choi, Jin Ho Kim, Kyung Hwan Kim

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

(Received June 23, 2007. Accepted November 15, 2007)

Abstract

It is well known that temporal information, i.e. speech onset, about input speech can be represented to the response nerve signal of auditory nerve better depending on the adaptation effect occurred in the auditory nerve synapse. In addition, the performance of a speech processor of cochlear implant can be improved by the adaptation effect. In this paper, we observed the emphasis characteristic of speech onset in the recently proposed adaptation algorithm, analyzed the characteristic of performance change according to the variation of parameters and compared with transient emphasis spectral maxima (TESM) is the previous typical strategy. When observing false peaks which are generated everywhere except speech onset, in the case of the proposed model, the false peak were generated much less than in the case of the TESHM and it is more distinguishable under noise.

Key words : cochlear implant, speech processor, auditory nerve, adaptation effect, speech onset

1. 서론

인공와우 (cochlear implant CI)는 감각신경성 고도난청자의 청신경을 전기펄스로 자극하여 청력을 회복시키기 위한 장치이다. 입력음성을 분석하여 음성인지에 적합한 전기자극펄스를 생성하는 어음처리 (speech processing) 부분은 달팽이관과 청신경섬유를 대략적으로 모사하는 방식을 취한다. 현재 가장 보편적으로 이용되고 있는 어음처리방식은 달팽이관 내 기저막에서 위치에 따라 서로 다른 주파수성분에 반응하는 place-frequency coding 현상을 대략적으로 모사하여 입력음성을 선형대역통과 필터뱅크를 이용하여 주파수에 따라 여러 채널로 나누고 이후 정류기와 저역통과필터로 포락선(envelope)을 추출한 후 펄스의 크기를 변조하여 청신경을 자극하는 구조로 되어있다[1,2].

이러한 선형필터뱅크 및 포락선 추출에 기반한 방식은 잡음이 없는 환경에서 평균적으로 높은 인지율을 보이지만 잡음 하에서

급격한 성능저하를 보이므로 어음처리 방식의 개선이 필요하다 [3]. 내측유모세포와 청신경 사이의 시냅스에서 기인하는 적응효과는 음성의 개시점등의 시간적 정보를 청신경 응답 신경신호에 보다 명확히 표현할 수 있게 하므로 잡음 하에서의 충실한 음성정보 전달이 크게 기여하는 것으로 알려져 있다. 이와 같은 적응효과를 적용하여 음성인식기를 구현한 결과 잡음 하에서 높은 음성인식률을 달성할 수 있었다[4].

적응효과를 인공와우에 적용하기 위한 몇 가지 방법들이 개발되어 왔다[5,6]. 적응효과를 인공와우 어음처리 방식에 적용하기 위하여 가장 간단하면서 효과적인 방법은 음성의 시작 이후 일정시간 동안 이득 값을 적용하는 방법이다. Vandali는 짧은 시간간격 내에서 급격하게 증가하는 acoustic cue에 이득 값을 취하는 적응효과 알고리즘을 제시하였고, transient emphasis spectral maxima (TESM) 라고 명명했다[5]. 이를 spectral maxima sound processor (SMSP)에 적용해 기존 SMSP 사용자들을 대상으로 언어 인지율의 변화를 관찰한 결과 음성인지율의 전체적인 향상을 확인할 수 있었다. 본 저자들은 최근 포락선 추출부에 적응효과를 더하여 음성의 시작점에 해당하는 부분의 포락선을 강조함으로써 전기

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

Corresponding Author : 김경환

강원도 원주시 흥업면 매지리 연세대학교 원주캠퍼스 백운관 208호 (220-710)

Tel : 82-33-760-2364 / Fax : 82-33-763-1953

E-mail : khkim0604@yonsei.ac.kr

자극펄스의 크기를 증폭시켜 적응효과를 모사할 수 있으며 음성시작점에 대한 정보전달 특성을 개선할 수 있다는 사실을 보였다 [6,7].

본 논문에서는 최근 제안한 적응효과 알고리즘의 음성개시점 강조특성을 자세히 관찰하고, 조절해 주어야 하는 변수들의 변화에 따른 성능변화 특성을 분석하며, 비슷한 목적을 갖는 대표적인 방법인 TESM과의 객관적이고 정량적인 성능비교를 수행하여 제안된 적응효과 알고리즘의 우수성을 확인하는 것을 목적으로 한다.

II. 이론 및 실험방법

A. 청신경섬유의 적응효과

내유모세포 (inner hair-cell)와 청신경 (auditory-nerve) 사이의 시냅스는 청각계에서 물리적인 소리 신호를 전기적인 신경신호로 전환하는 중요한 구성요소이다. 일정한 세기의 tone burst에 대한 청신경섬유의 응답은 그림 1에서 보이는 것과 같이 자극이 시작되는 onset에서 매우 급격하게 발화한 후 빠르게 감소하며 (rapid adaptation) 시간이 흐름에 따라 단조 감소한다 (short-term adaptation)[8]. 이러한 현상을 적응효과라고 부른다. 적응효과와 특성은 자극의 세기, 지속시간, 이전자극의 영향을 받는다. 청신경 응답의 onset에서 적응효과는 식 (1)과 같이 두 개의 지수함수로 표현할 수 있다[8].

$$R_{on}(t) = A_{sus} + A_r e^{-t/\tau_r} + A_{st} e^{-t/\tau_{st}} \quad (1)$$

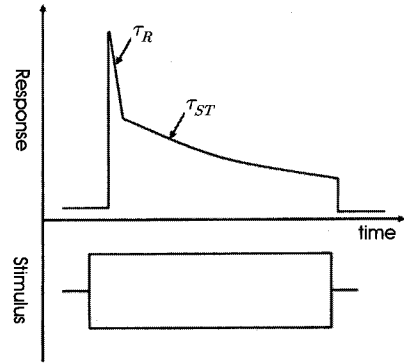


그림 1. 자극에 대한 청신경섬유의 응답
Fig. 1. The response of auditory nerve fiber to stimulus.

A_r 은 rapid adaptation의 component이고, A_{st} 는 short-term adaptation의 component이다. 그림 1에서 τ_r 은 rapid adaptation 감쇠시정수이고, τ_{st} 는 short-term adaptation 감쇠시정수이다. 마지막으로 A_{sus} 는 안정상태의 component로써 청신경에 자극이 들어 오지 않을 때 청신경의 응답이다.

B. 적응효과 알고리즘

적응효과 알고리즘은 그림 2의 (a)와 같은 일반적인 어음처리방식에 (b)와 같이 포락선추출과정을 개선하는 방식으로 적용되며 음성에 급격한 변화가 일어나는 피크 값을 찾는 다음 이를 포락선

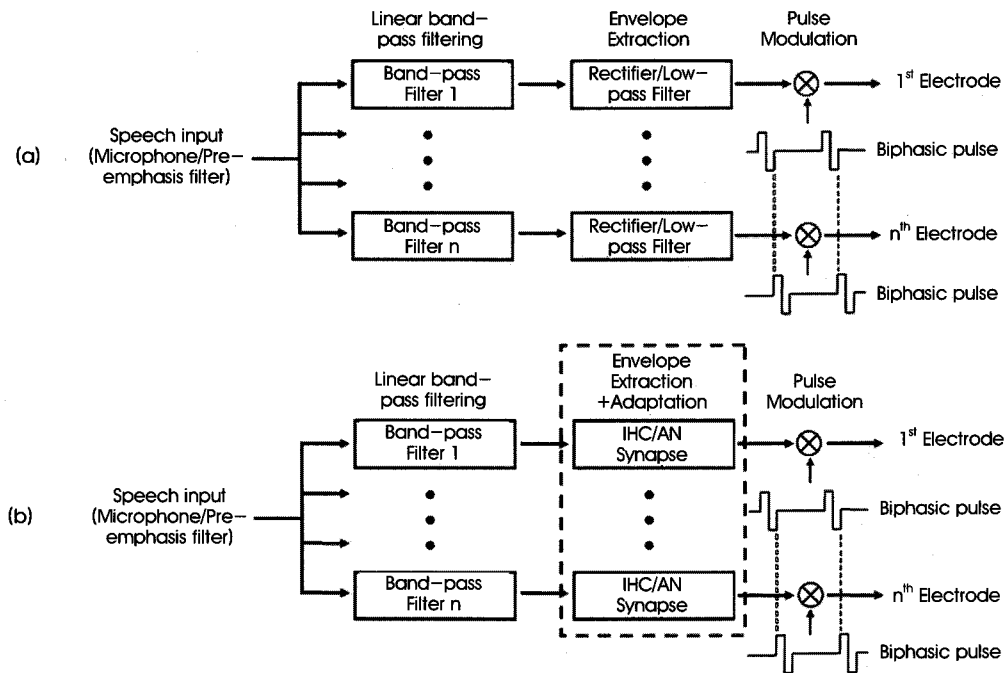


그림 2. (a) 일반적인 CIS 어음처리방식의 순서도 (b) 적응효과가 적용된 어음처리방식의 순서도
Fig. 2. (a) The block diagram of a typical CIS strategy (b) The block diagram of CIS with adaptation

에 더하게 된다. 어음처리기의 주파수분석부분은 선형대역통과필터로 구성되며 Loizou가 제시한 중심 주파수와 3 dB 대역폭 중 8 채널에 해당하는 값을 사용하였다[9]. 중심 주파수는 최소 300 Hz 부터 5000 Hz까지 변화하며 각 채널은 394 Hz, 692 Hz, 1064 Hz, 1528 Hz, 2109 Hz, 2834 Hz, 3740 Hz, 4871 Hz의 값을 가진다. 한편 각 채널은 265 Hz, 331 Hz, 431 Hz, 516 Hz, 645 Hz, 805 Hz, 1006 Hz, 1257 Hz의 값을 3 dB 대역폭으로 가진다.

Transient emphasis spectral maxima (TESM)

Vandali에 의해 제안된 TESP 적응효과 알고리즘은 spectral maxima sound processor (SMSPP)에 적응효과를 추가한 모델이나 본 논문은 두 적응효과 알고리즘의 비교를 위하여 주파수분석부분은 일반적인 선형대역통과필터들로 구성된다.

TESM 적응효과 알고리즘의 동작은 다음과 같다. 60 ms의 길이를 갖는 sliding buffer 내에 포락선 추출부에서 얻어진 포락선 신호를 저장한 다음 buffer를 3개의 20 ms의 시간 윈도우로 분할하여 각 시간 윈도우 내 포락선의 평균값을 과거 (E_p), 현재 (E_c), 미래 (E_f)의 값들로 지정한다. 이 3개의 값들로부터 Gain factor (G)값을 다음과 같이 계산한다.

$$G = \frac{2 \times E_c - 2 \times E_p - E_f}{E_c + E_p + E_f} \quad (2)$$

식 (2)에 의해 계산되는 G값은 시간 윈도우 내의 포락선의 시간적 변화에 따라 결정된다. 즉, 짧은 시간간격 (60 ms 이내) 동안 급격하게 증가하고 급격하게 감소하는 포락선은 큰 G 값을 가지게 되며 ($0 \leq G \leq 2$) 긴 시간간격 동안 급격하게 증가한 후 상수 레벨을 유지하는 포락선은 상대적으로 낮은 G 값을 ($0 \leq G \leq 0.5$) 갖게 된다. 안정상태를 유지하는 포락선은 음수인 G 값을 가지게 되는데 이 경우에는 G 값을 0으로 고정한다.

각 채널의 Gain factor (G_n)은 본래의 포락선 ($S_n(t)$)에 식 (3)에 따라 연산되어 적응효과가 적용된 포락선 ($S'_n(t)$)를 만든다.

$$S'_n(t_m) = S_n(t_m) \times (1 + K_n \times G_n) \quad (3)$$

식 (3)에서 n은 채널의 번호를 의미하며 K_n 은 원 포락선에 찾아진 피크가 더해질 때 피크의 크기를 조절하기 위한 상수이다. 본 논문에서는 모든 채널에 대해 $K_n = 2$ 로 설정하였다.

제안하는 적응효과 알고리즘

그림 3은 저자들이 최근 제안한 적응효과 알고리즘에 의한 포락선 추출부를 설명하고 있다[6,7]. 좌측 통로는 반파정류기와 저역통과필터로 구성되는 일반적인 포락선 추출방식이다. 우선 반파정류기1 (half-wave rectifier; HWR1)에 의해 반파 정류된 신호는 저역통과필터1 (low-pass filter; LPF1)에 의해 평활화 (smoothing)

된다. 사용한 저역통과필터1(LPF1)은 400 Hz의 차단 주파수를 가지는 4차 버터워스 필터이다. 처리된 신호는 뺄셈기 (Σ) 및 피크 검출기/증폭기2 (Peak detector/Amplifier2)로 입력된다. 우측 통로는 저역통과필터, 반파 정류기, 증폭기의 순서로 구성된다. 저역통과필터2(LPF2)는 20 Hz의 차단주파수를 가지는 4차 버터워스 필터이며 LPF1보다 상대적으로 낮은 차단주파수를 가진다. 따라서 좌측통로에 비해 상대적으로 느리게 변하는 포락선이 추출된다. 추출된 포락선은 반파 정류된 후 증폭기1을 거쳐 뺄셈기 (Σ)로 입력된다. LPF2의 다음과정인 HWR2는 오류피크 발생을 방지하는 역할을 한다. LPF2는 매우 큰 시정수를 가지고 있기 때문에 신호의 크기가 급격히 감소하는 구간에서 LPF2의 출력 값은 음이 된다. 이 값이 뺄셈기로 입력될 경우 오류피크 값이 추출된다. 따라서 이러한 현상을 방지하기 위해 반파정류기를 사용하여 0이하의 값들을 제거하였다.

그림 4는 8채널 적응효과 알고리즘에 /ka/를 입력음성으로 하였을 때 첫 번째 채널에서 뺄셈기로 입력되는 두 통로의 출력신호이다(그림 3의 Y 및 Z). LPF2가 LPF1보다 상대적으로 더 큰 지연 시간을 가지므로 음성이 시작되거나 포먼트 전이 (formant trans-

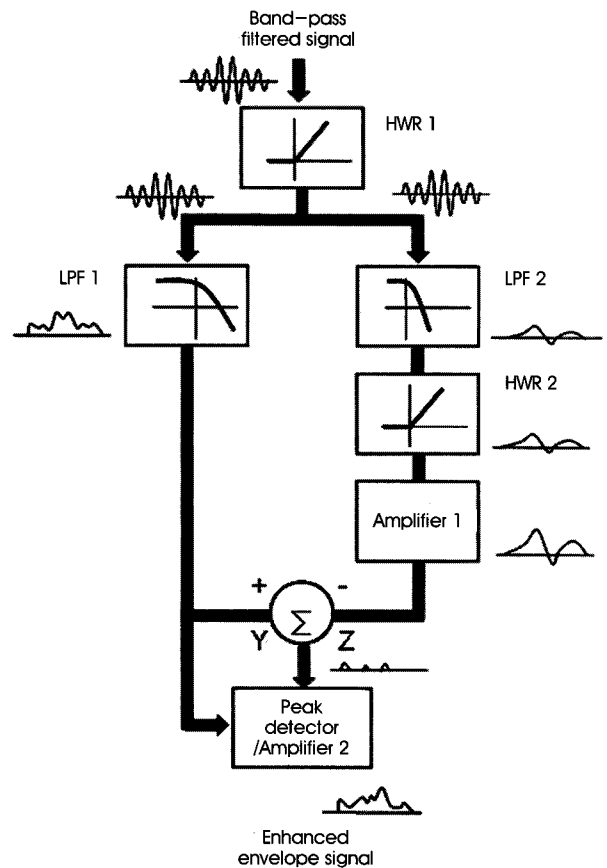


그림 3. 일반 포락선추출과정 (left trace)과 개선된 포락선추출과정 (both traces)

Fig. 3. Conventional envelope extraction (left branch) and the proposed enhanced envelope extraction (both branches)

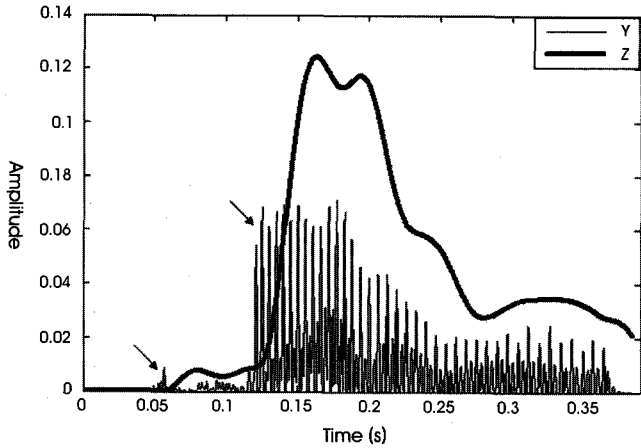


그림 4. 총 8채널 중 입력음성 /ka/에 대한 각 통로의 1번째 채널의 포락선 신호
 Fig. 4. The envelope signal of first channel among 8 channels for input 'ka'

ition)가 일어나는 등 신호의 크기가 크게 증가하는 시점에서 왼쪽 통로의 출력은 항상 오른쪽 통로의 출력보다 크다. 따라서 뺄셈기로 두 신호가 입력되면 그림 4에서 화살표로 표시한 부분들이 추출되며 이 피크 값들은 피크 검출기/증폭기2로 입력된다. 피크 검출기/증폭기2 단에서는 1 ms마다 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.

C. 성능평가

적응효과 추가의 목적은 급격한 변화가 발생하는 음성시작점의 강조이며 /ka/와 같은 자음과 모음이 결합된 음성의 경우 자음 /k/에 의한 출력파형이 적응효과에 추가로 인하여 강조되어야 한다. Kennedy 등은 자음을 강조하여 자음과 모음 간의 상대적인 크기

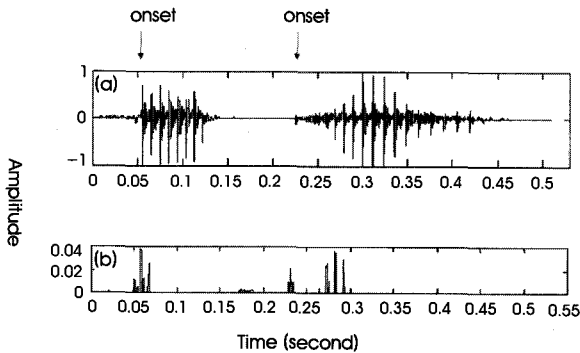


그림 6. 적응효과 알고리즘의 평가방법 (a) 음성 /apa/의 파형 (b) 제안한 모델 (8 채널)에 의해 추출된 피크 중 첫번째 채널
 Fig. 6. The comparing method of adaptation algorithm (a) The graph of speech /apa/ (b) The first channel's peak signal extracted by 8 channel proposed model

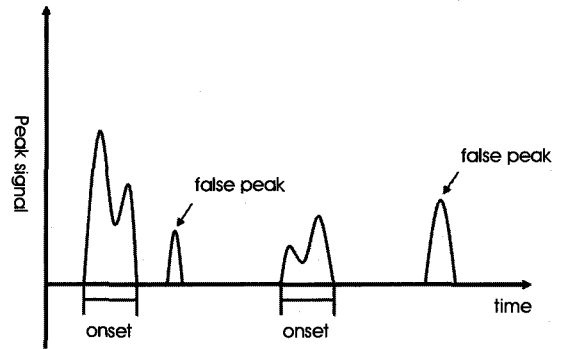


그림 5. 제안된 적응효과 알고리즘에 의해 추출된 피크 신호
 Fig. 5. Peak signals extracted by proposed adaptation algorithm

의 차이를 줄여줌으로써 인공와우 사용자들의 어음인지능력을 향상시킬 수 있음을 확인하였다[12]. 이로부터 음성의 개시점이 충실히 강조된다면 어음의 인지율이 증가함을 예상할 수 있으나 개시점이 아닌 부분에서의 강조는 성능저하를 가져오게 된다. 그림 5는 적응효과알고리즘 출력파형의 개략도이다. 적응효과 알고리즘의 성능평가는 다음과 같이 이루어진다. 그림 5에서와 같이 제시된 음성이 시작 한 후 20 ms까지를 개시점으로 간주하고 개시점 이외의 부분에서 생긴 피크의 수를 세어 채널 별로 발생한 오류피크 개수들의 평균을 계산한다.

그림 6은 제안된 적응효과 알고리즘에 의한 출력파형의 예를 도시하고 있다 (음성 /apa/). (a)는 입력 음성 /apa/의 파형이고 그래프에 표시한 화살표는 각각 /a/의 개시점과 /pa/의 개시점을 의미한다. (b)는 제안된 모델에 의해 추출된 피크들이다.

D. 실험에 사용된 음성 및 배경잡음

다양한 자음 (/a/, /i/, /u/)과 결합된 파열음(stop consonant)과 자음-모음-자음 (consonant-vowel-consonant; CVC)의 단음절 어음 등 두 종류의 음성 세트를 사용하여 성능평가를 수행하였다. 첫 번째 음성 세트는 모음-자음-모음 (vowel-consonant-vowel; VCV) 형태로 사용하는 모음은 /a/, /i/, /u/이고, 자음은 파열음으로 /p/, /t/, /k/, /b/, /d/, /g/이다[11]. 더욱 다양한 문맥 (context)에 대한 결과를 얻기 위해 단음절 CVC 음성도 실험에 사용되었다. 실험에서 자음으로 파열음을 사용한 이유는 크기가 작고 배경잡음과 특성이 유사한 파열음의 경우 적응효과에 의한 음성개시점 강조가 가장 효과적일 것으로 기대되기 때문이다.

배경잡음으로는 가우시안 백색잡음 (Gaussian white noise; GWN)과 음성과 유사한 스펙트럼을 갖는 잡음 (Speech shaped noise; SSN)을 사용하였다. 신호대잡음비 (Signal to noise ratio; SNR)는 0 dB부터 10 dB까지 변화시켰다. SNR은 실제 대화환경과 유사한 환경 하에서의 성능평가를 수행하기 위하여 사용하였으

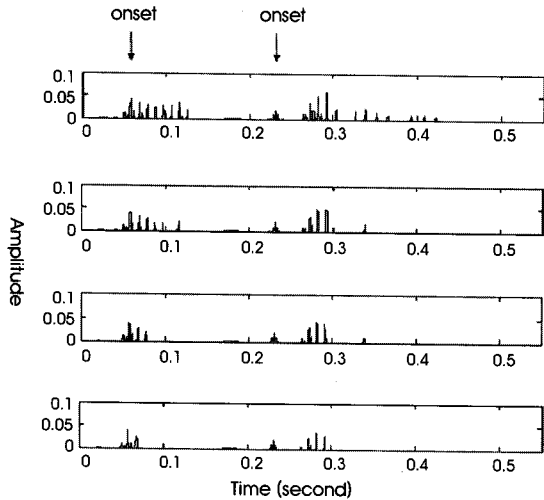


그림 7. 제안된 모델의 증폭기1의 이득 값 조절 (입력음성 /apa/, 전체 8채널 중 첫 번째 채널) (a) gain = 3 (b) gain = 4 (c) gain = 5 (d) gain = 6

Fig. 7. Adjustment of gain of the first amplifier in proposed model a speech input /apa/, the first channel among 8 channels) (a) gain = 3 (b) gain = 4 (c) gain = 5 (d) gain = 6

며 Yang 등이 제시한 방법과 같이 가우시안 백색 잡음을 1100 Hz의 차단주파수를 가지는 2차 버터워스 저역통과필터에 입력시킴으로써 획득하였다[13].

III. 결 과

A. 제안한 적응효과 알고리즘에서 증폭기의 효과

제안된 모델은 두 개의 증폭기를 가진다. 그 중 증폭기1(amp-lifier 1)의 효과를 알아보기 위한 실험을 하였다. 그림 7은 증폭기1의 이득 값을 바꿔가며 제안된 적응효과 알고리즘에 의해 발견된 피크들을 나타낸다 (전체 8채널 중 첫 번째 채널만 도시). 표 1은 오류피크의 숫자를 세어 기록한 것으로 표 1과 그림 7을 보면 증폭기1의 이득을 증가시킬수록 오류피크의 개수가 줄어드는 것을 알 수 있다. 증폭기2의 이득 값의 조절에 의해서는 오류피크의 개수는 변화하지 않고 단지 피크의 크기만이 변화하였다.

B. 입력음성에 따른 각 적응효과 알고리즘의 결과

표 2는 모음-자음(파열음)-모음 VCV 형태의 입력음성에 두 가

표 1. 제안된 적응효과 알고리즘에서 증폭기1의 이득 값에 따른 오류피크의 개수 (입력음성 /apa/, 전체 8채널)

Table 1. The number of false peaks depending on gain of the first amplifier in proposed adaptation algorithm

| Channel Gain | 1st channel | 2nd channel | 3rd channel | 4th channel | 5th channel | 6th channel | 7th channel | 8th channel | Average number |
|--------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|----------------|
| 3 | 38 | 14 | 19 | 17 | 17 | 16 | 20 | 17 | 19.8 |
| 4 | 24 | 11 | 9 | 11 | 10 | 13 | 18 | 12 | 13.5 |
| 5 | 17 | 8 | 6 | 7 | 6 | 9 | 14 | 7 | 9.3 |
| 6 | 13 | 6 | 5 | 7 | 3 | 10 | 8 | 7 | 7.4 |

표 2. VCV words에 대한 TESM과 제안된 모델의 오류피크의 개수

Table 2. The number of false peaks of TESM and proposed model for VCV words

| Vowel Stop consonant | /a/ | | /i/ | | /u/ | |
|----------------------|------|----------------|------|----------------|------|----------------|
| | TESM | Proposed model | TESM | Proposed model | TESM | Proposed model |
| /k/ | 7.8 | 8.1 | 8.1 | 5.6 | 5.6 | 4.6 |
| /g/ | 15.4 | 5.1 | 10 | 5.9 | 7.8 | 4.9 |
| /p/ | 7.8 | 7.4 | 9 | 7.6 | 8.3 | 6.9 |
| /b/ | 5.6 | 5.6 | 6.3 | 5.9 | 5 | 4.1 |
| /t/ | 7.9 | 7.3 | 9.4 | 6.8 | 7.4 | 5.6 |
| /d/ | 5.5 | 5 | 7.5 | 7.5 | 8 | 5.4 |

표 3. CVC words에 대한 TESM과 제안된 모델의 오류피크의 개수

Table 3. The number of false peaks of TESM and proposed model about CVC words

| Vowel Stop consonant | /a/ | | /i/ | | /u/ | |
|----------------------|------|----------------|------|----------------|------|----------------|
| | TESM | Proposed model | TESM | Proposed model | TESM | Proposed model |
| /k/ | 12.5 | 8 | 5.6 | 9 | 9.1 | 9.5 |
| /g/ | 7.6 | 5.3 | 7.5 | 5.3 | 4.4 | 2.4 |
| /p/ | 6.4 | 8.4 | 9.9 | 9.5 | 8.1 | 8.1 |
| /b/ | 5.4 | 3.1 | 8.6 | 4.1 | 5.3 | 4 |
| /t/ | 7.8 | 4.5 | 8 | 7.8 | 8.8 | 6.6 |
| /d/ | 5.3 | 4.6 | 6.9 | 5.6 | 6.8 | 3.8 |

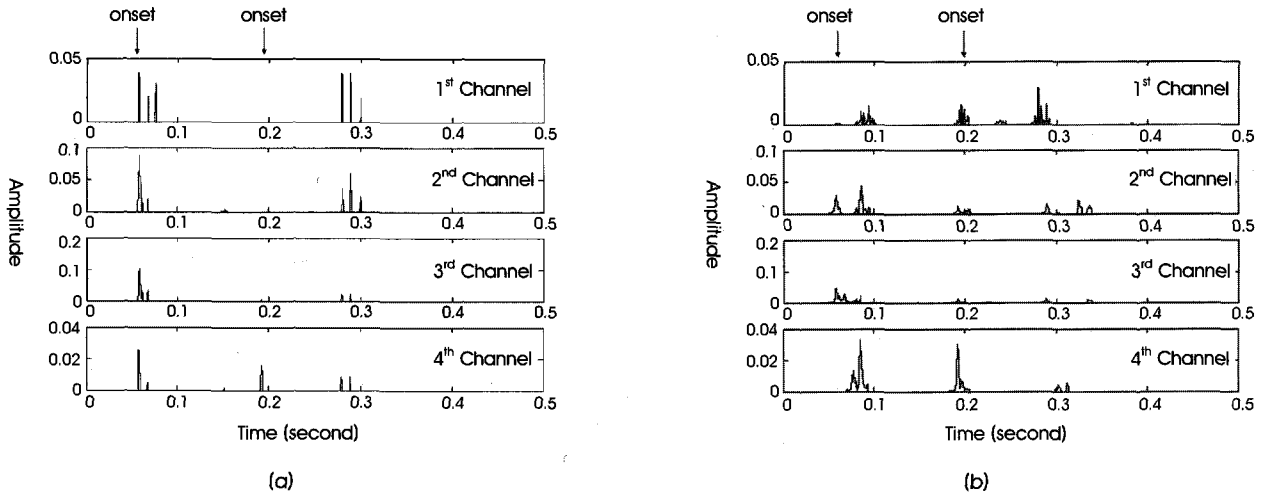


그림 8. 10 dB SNR GWN가 섞인 음성에 대한 두 모델의 피크 신호(입력음성 /aka/, 전체 8채널 중 1,2,3,4 채널) (a) 제안한 방식(b) TESM
 Fig. 8. Peak signals of two models in response to speech input with 10 dB SNR GWN (a speech input /apa/, the 1st, 2nd, 3rd and 4th channels among 8 channels) (a) Proposed method (b) TESM

지 적응효과를 적용하였을 때 발생하는 오류피크의 평균개수를 측정한 결과이다. 어음처리는 8 채널로 구성하였다. 전체적으로 TESM의 오류피크의 개수가 제안한 모델을 적용했을 때의 오류피크 개수보다 많음을 알 수 있다. 제안한 모델의 오류피크의 개수가 더 많은 경우는 /aka/를 사용한 단 한번뿐이다.

표 3는 자음(파열음)-모음-자음(파열음) (CVC) 형태의 음성 세트에 대한 각 적응효과 알고리즘 적용 결과이다. /pap/, /kik/ 두 경우를 제외하면 TESM의 오류피크의 개수가 제안한 모델보다 많음을 알 수 있다.

C. 잡음에 따른 적응효과 알고리즘의 결과

그림 8은 10 dB SNR 가우시안 백색잡음이 추가된 경우의 결과

이다. 즉, TESM 및 제안된 모델을 적용하여 각 모델에 의해서 검출된 피크들이다. 그림 9는 GWN이 섞인 /aka/ 음성을 두 적응효과 알고리즘의 입력으로 사용한 결과이다. GWN의 SNR은 앞서 말했듯이 0 dB에서 10 dB까지 변화시켰으며 각 SNR에서 10번씩 실험했다. 즉, 같은 입력에 대해서 10번 실험하여 오류피크의 개수를 구했으며 평균을 구했다. 그래프를 보면 TESM 모델에서 오류피크의 개수는 제안된 모델에서의 수보다 항상 많고, SNR의 값에 따라 변하지 않는다. 한편 제안된 모델에서의 오류피크의 개수는 SNR이 증가 함에 따라 증가한다.

그림 10은 SSN이 섞인 /aka/ 음성을 두 적응효과 알고리즘의 입력으로 사용한 결과이다. 실험과정은 GWN에서와 같고, 결과도 같은 형식으로 정리하였다. GWN에서와 마찬가지로 TESM 모델

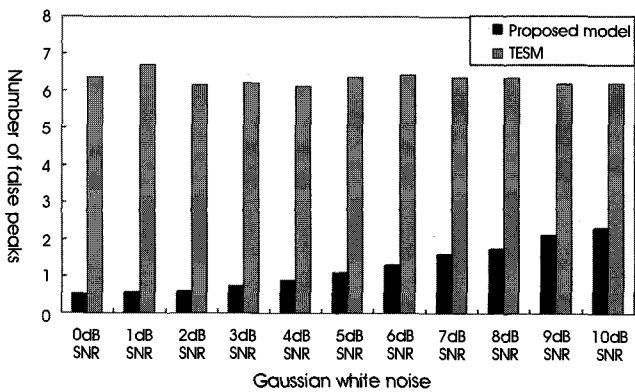


그림 9. Gaussian white noise가 섞인 입력 음성 /aka/에 대한 각 적응효과 알고리즘의 오류피크의 개수

Fig. 9. The number of false peaks in response to a speech input /aka/ with Gaussian white noise in each adaptation algorithm

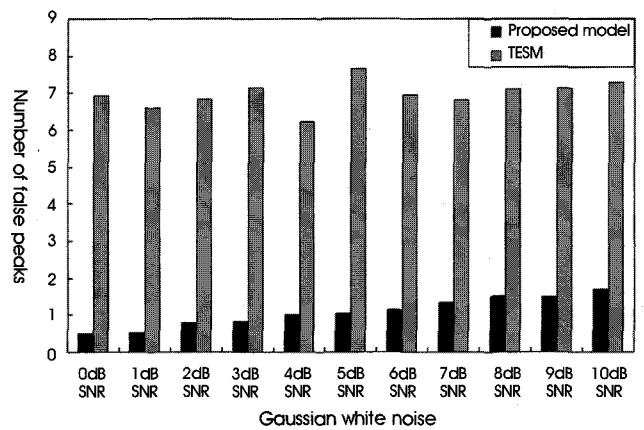


그림 10. Speech shaped noise가 섞인 입력 음성 /aka/에 대한 각 적응효과 알고리즘의 오류피크의 개수

Fig. 10. The number of false peaks in response to a speech input /aka/ with speech shaped noise in each adaptation algorithm

에서는 잡음의 SNR과 관계없이 오류피크의 개수는 거의 일정하다. 또한 제안한 모델에서는 잡음의 SNR이 늘어날 때 오류피크의 개수도 늘어난다.

IV. 고찰 및 결론

인공와우의 음성처리 방식에 적응효과를 추가하는 목적은 청신경의 전기자극 시 전기자극펄스의 진폭에 음성의 개시점에 관한 정보가 강조되도록 하기 위함이다. 그러나 오류피크는 알고리즘이 의도하지 않은 정보이므로 각 적응효과 알고리즘은 오류피크를 최소화하는 방향으로 설계되어야 한다. 오류피크가 많은 경우 음성 개시점뿐만 아니라 이어 발생하는 모음에 의한 출력파형의 크기도 증가하므로 적응효과 알고리즘의 효율성이 감소한다.

제안하는 적응효과 알고리즘에서 증폭기1의 이득 값을 조절함으로써 오류피크의 개수를 조절 할 수 있고, 개시점이 강조되는 한도 내에서 이 이득 값을 최대로 하면 오류피크의 개수를 최소화 할 수 있다. 즉, 제시한 모델은 단지 하나의 변수 조절로 쉽게 성능을 향상시킬 수 있다. 반면에 TESM에서는 오류피크의 개수를 조절하기 위해서 식 (2) 자체를 변화시켜야 하며 조절하기 매우 어려운 과정이다. 또 TESM은 자음의 개시점뿐만 아니라 모음의 포먼트 천이 (formant transition)도 자음의 개시점과 같은 정도로 증폭하므로 자음과 모음의 비율을 조절하기에 어려움이 많다.

실험에서 자음으로 파열음을 사용한 이유는 파열음의 경우 음성이 시작하기 전 입안에서 압축된 공기가 한번에 배출되며 나는 소리인 burst가 존재하며 각 적응효과 알고리즘은 burst를 효과적으로 검출하기 때문에 다른 자음에 비해 파열음에서 더욱 효과적일 것으로 예상되었기 때문이다.

전체적으로 제안한 모델을 사용할 경우 TESM을 사용하는 경우보다 훨씬 적은 수의 오류피크가 발생하였으며 잡음 하에서 성능 차이가 더욱 확연해짐을 확인하였다. 잡음 하에서 TESM을 적용할 경우 잡음에 의한 피크들이 많이 검출된 반면, 제안한 모델을 사용할 경우 음성의 개시점의 강조는 계속 이루어지면서 오류피크의 개수는 상대적으로 적음을 확인할 수 있었다. 추후 인공와우 어음 처리기에 실제로 적용함으로써 효율성을 실험적으로 확인하고 또 파라미터 조절의 용이성을 확인해야 할 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, no. 6332, pp. 236-238, 1991.
- [2] D. H. Kim, J. H. Kim, K. H. Kim, "Performance evaluation of cochlear implants speech processing strategy using neural spike train decoding," *J. of Biomed. Eng. Res.*, vol. 28, no.2, pp. 271-279, 2007.
- [3] 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, no. 4, pp. 73S-81S, 2005.
- [4] J. Tchorz, B. Kollmeier, "A model of auditory perception as front end for automatic speech recognition," *J. Acoust. Soc. Am.*, vol. 106, no. 4, pp. 2040-2050, 1999.
- [5] A. E. Vandali, "Emphasis of short-duration acoustic speech cues for cochlear implant users," *J. Acoust. Soc. Am.*, vol. 109, no. 5, pp. 2049-2061, 2001.
- [6] J. H. Kim, K. H. Kim, "A novel speech processing strategy for cochlear implants based on nonlinear model of auditory periphery," in *proc. 34th KOSOMBE*, Wonju, Korea, Nov. 2006, pp. 152-155.
- [7] 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. of Biomed. Eng. Res.*, vol. 28, no. 2, pp. 205-211, 2007.
- [8] L. A. Westerman, R. L. Smith, "Rapid and short-term adaptation in auditory nerve responses," *Hear. Res.*, vol. 15, no. 3, pp. 249-260, 1984.
- [9] P. C. Loizou, M. Dorman, Z. Tu, "On the number of channels needed to understand speech," *J. Acoust. Soc. Am.*, vol. 106, no. 4, pp. 2097-2103, 1999.
- [10] X. Zhang, L. H. Carney, "Analysis of models for the synapse between the inner hair cell and the auditory nerve," *J. Acoust. Soc. Am.*, vol. 118, no. 3, pp. 1540-1553, 2005.
- [11] G. S. Donaldson, H. A. Krefl, "Effects of vowel context on the recognition of initial and medial consonants by cochlear implant users," *Ear Hear.*, vol. 27, no. 6, pp. 658-677, 2006.
- [12] E. Kennedy, H. Levitt, A. C. Neuman, M. Weiss, "Consonant-vowel intensity ratios for maximizing consonant recognition by hearing-impaired listeners," *J. Acoust. Soc. Am.*, vol. 103, no. 2, pp. 1098-1114, 1998.
- [13] L. Yang, Q. Fu, "Spectral subtraction-based speech enhancement for cochlear implant patients in background noise," *J. Acoust. Soc. Am.*, vol. 117, no. 3, pp. 1001-1004, 2005.