

인공 청각 장치의 음성신호 처리와 자극방법의 시뮬레이션

김 영훈*, 박 광석

서울 대학교 공과 대학 의공학 협동과정, 서울 대학교 의과 대학 의공학 교실

Simulation of speech processing and coding strategy for cochlear implants

Young Hoon Kim, Kwang Suk Park

Dept. of Biomed. Eng, Seoul National Univ.

Abstract

The object of speech processor for cochlear implants is to deliver speech information to the central nerve system. In this study we have presented the method which simulate speech processing and coding strategy for cochlear implants and simulated two different processing methods to the 12 adults with normal ears. The formant sinusoidal coding was better than the formant pulse coding in the consonant perception test and learning effects. ($p < 0.05$)

I. 서론

인공 청각 장치에 사용되는 음성 신호 처리기의 음성 처리 방법은 전기 자극의 형태에 기본을 두고 있다. 또한 전기 자극의 목적은 전기자극을 했을 때 가능한 많은 정보를 전달 해야한다는 것이다. 이러한 관점에서 전기자극에 의해 유발되는 청각 신경의 활동 전위 패턴은 정상인의 그것과 최대한 비슷할수록 보다 많고 정확한 정보를 전달 한다고 할 수 있다. 동물 실험을 통하여 음성 자극에 대한 청각 신경의 발화 형태의 특징은 비교적 잘 알려져 있다. 그러나 모든 청각 신경으로부터 정보를 받아들이고 그로부터 자극에 대한 정보를 추출하는 중추신경의 정보 추출 방법에 대해서는 아직 특정한 이론이 없다. 그 결과 청각 장치에 사용되는 음성신호 처리와 자극 방법은 생리학적으로 알려진 달팽이관에서 나타나는 자극에 대한 주파수 특성

과, 시간축 상의 특성을 통하여 전기 자극의 형태를 결정한다. 그러나 이러한 전기 자극 형태에 대하여 장애자가 어떠한 자극의 정보를 인식하게 될지는 실제 장애자에게 전기자극을 하고 물어보는 방법 뿐이다. 그러나 음성 신호의 처리와 자극의 방법을 비교하고 결정하는 단계에서 이러한 방법을 이용할 수는 없다. 본 연구에서는 이러한 문제를 해결하기 위해 정상인들을 상대로 인공 청각 장치에서 사용될 수 있는 음성 신호 처리 방법과 자극 방법을 비교할 수 있는 방법을 제시하고 두 가지 자극 방법을 비교해 보았다.

II. 실험 방법

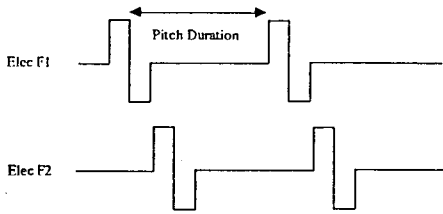
실험의 기본적인 원리는 음성신호에 전기 자극 형태에 따른 적당한 처리를 하여 전기 자극을 했을 때 전달될 수 있는 정보만으로 재합성하여 이들에 대한 정상인의 응답을 알아보는 것이다. 비교에 사용된 두 가지 자극의 형태는 현재 임상 시술되고 있는 인공 청각 장치에 사용되는 formant pulse 방법과, formant 정보와 음성 합성 방법에서 사용되는 filter bank 방법을 합친 formant sinusoidal 방법을 이용했다.

1. 자극 형태

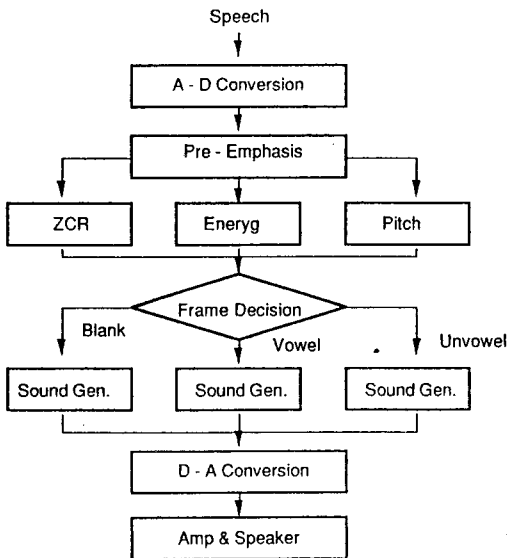
1) formant puls coding

이 자극의 형태는 현재 서울 대학교 병원에서 임상 시술되고 있는 Nucleus사의 인공 청각 장치에서 채택하

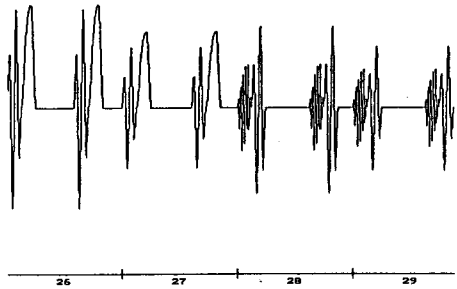
고있는 자극이다. 자극의 형태는 그림 1 과같다. 이 자극이 전달하는 음성 정보 만으로 음성을 재 합성하는 방법은 그림 2와 같다. 먼저 음성은 8KHz로 smampling하여 128 point (16msec)를 한 frame으로 분석 단위를 설정하고 frame 단위로 유성음과 무성음을 판별한다. 유성음으로 판정된 구간은 formant위치의 전극에 전기자극이 가해지는 구간 만큼 formant frequency를 pulse width의 window function과의 곱으로 합성음을 만든다. 여기서 F1 과 F2 전극이 동시에 자극 되지 않도록 두 가지 합성음 역시 F2 성분이 먼저 합성된다. 여기서 자극의 주기는 기본 주파수의 주기를 이용한다. 만일 frame이 무성음으로 판정되면 자극의 주기를 불규칙하게 결정하여 전체 전극중 불규칙하게 선택된 2개의 전극과 F2전극을 자극한다. 이를 합성음으로 만들때는 random noise와 F2 frequency를 불규칙한 주기로 유성음과 같은 방법으로 합성한다. 재 합성된 음성의 모양은 그림 3과 같다.



- 그림 1 formant pulse coding의 자극형태 -



- 그림 2 formant pulse coding의 음성 재합성 방법 -



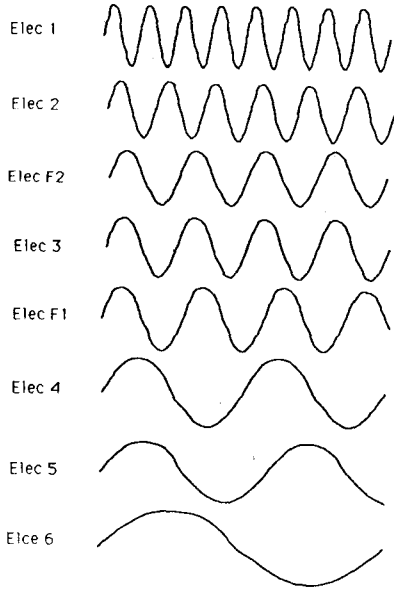
- 그림 3 formant pulse coding에 의한 합성음 -

2)formant sinusoidal coding

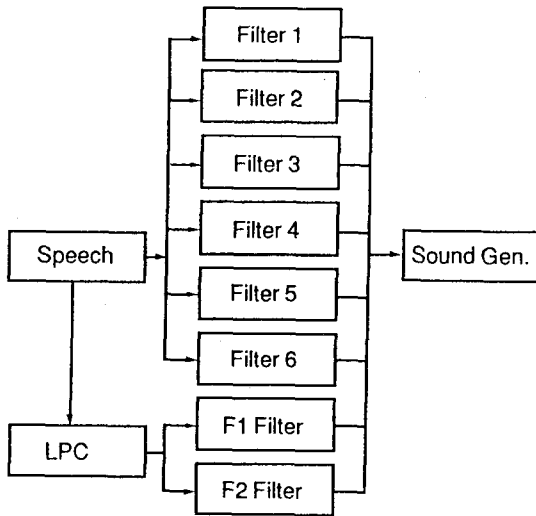
위의 자극 형태가 음성 발생시의 특징 추출에 의한 정보의 전달이라고 하면 이 자극 형태는 음성의 spectral 정보를 이용하는 것이다. 이 자극의 형태는 전기 자극의 두 가지 부류중 processing을 최소로 하는 경우에 속하는 filter bank방법과 특징 추출 방법의 장점을 모두 이용한다. 위에서와 같이 sampling된 음성은 역시 같은 크기의 frame단위로 재합성된다. 전기 자극의 형태는 그림 4 와같다. 이 자극에 의해 합성음을 만드는 방법은 그림 5와 같다. 300 Hz - 4000 Hz 의 주파수 영역에 충분한 전극을 할당하고 각 각의 전극이 위치한 곳의 characteristic frequency를 중심으로하고 청각 filter의 bandwidth를 갖는 중첩된 8개의 bandpass filter를 통과하여 각 각의 filter를 통과한 energy에 비례하는 filter center frequency를 모두 더하여 음성을 재 합성한다. 이 중 두개의 filter는 1,2 차 formant frequency로 한다. 위의 자극과 다른 큰 차이는 8개의 전극을 동시에 자극 한다는 것이다. 여기에는 적지 않은 문제점이 존재하나 자극의 상호 간섭을 줄이기위한 별도의 방법을 이용 그 영향을 줄일 수 있다. 이 자극에 의한 합성음은 그림6 과 같다.

2. Simulation

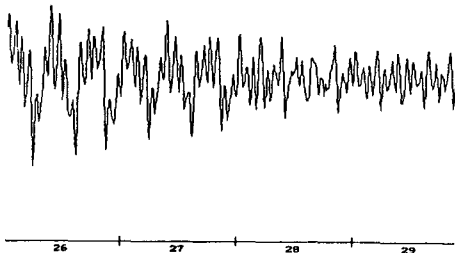
두 가지 자극방법을 비교하기 위하여 정상인을 대상으로 자음의 인식율을 비교해 보았다. 우리말 2음절 단어중 자음 하나만 다른 단어로 6개씩 하나의 SET 으로 하여 6개의 집합을 만들었다. 인식율의 측정에는 six -



- 그림 4 formant sinusoidal coding의 자극 형태 -



- 그림 5 formant sinusoidal coding의 음성 재합성 방법 -



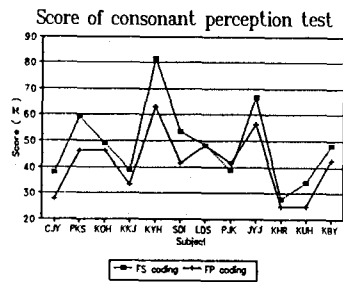
- 그림 6 formant sinusoidal coding에 의한 합성음 -

alternative- forced choice 방법을 사용했다. 이 방법은 먼저 subject에게 하나의 단어와 그 단어가 속한 집합을 알려준다. subject는 알려진 집합의 6개 단어중 방금 들은 단어라고 생각되는 것을 답한다. 실험용 단어는 다음과 같다.

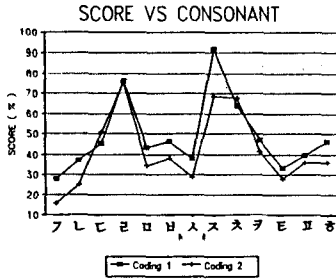
1. 기고 기로 기보 기조 기초 기호
2. 기마 기사 기자 기차 기타 기하
3. 수고 수도 수로 수조 수포 수호
4. 구리 누리 무리 부리 수리 추리
5. 노피 도피 모피 소피 코피 포피
6. 내기 대기 배기 캐기 태기 패기

전체 36개의 단어를 2가지 자극방법에 의한 합성음을 모두 random하게 선택하여 하나의 단어를 3번 듣게 된다. 한번 실험에 subject는 모두 $36 \times 3 = 216$ 개의 답을 하게된다. 12명의 정상인을 대상으로 인식을 실험을 하였다. 인공 청각 장치에 사용되는 음성신호 처리기의 목적은 순독과, 학습을 통하여 새로운 소리에 대하여 장애자가 익숙해져서 일상적인 대화를 가능하게 하는 것이다. 12명중에서 비교적 실험에 호응이 좋은 4명을 대상으로 학습에 의한 효과를 비교해 보았다. 2명씩 각기 다른 순서로 학습을 시키고 학습한 이후에 전과 같은 실험을 반복해 보았다.

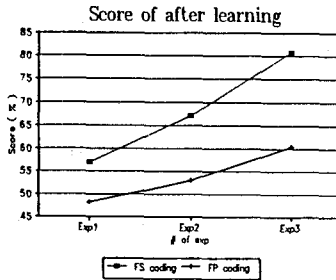
III. 결과



- 그림 7 subject vs 자음의 인식율 -



- 그림 8 자음 vs 인식율 -



- 그림 9 인식율 vs # of exp -

IV. 토의 및 결론

두 가지 자극방법에 대한 인식율의 결과를 SPSS/PC+ 통계처리 Tools을 이용하여 Mann - Whitnet U-Wilcoxon Rank Sum W Test를 통하여 인식율의 차이를 검증한 결과 FS coding이 더 좋은 결과를 보였다 ($p < 0.05$) 두 방법 모두 ㄱ 에 대한 인식율이 낮은 것을 알 수 있다 ㄱ 의 경우 대부분의 오답이 ㄷ, ㅂ으로 답한 것으로 나타났는데 이는 ㄱ, ㄷ, ㅂ 은 조음 방법에 의한 분류에서 모두 폐쇄음으로 구분 되는 유사성을 갖고있다. 한국어 음성 인식에서도 이러한 자음의 경우 그 구별이 다른것에 비하여 어려운 것으로 보고된다. 조음상 크게 차이가 나는 자음의 경우는 거의 혼동이 없는 것으로 보아 본 자극의 형태가 조음상에 나타나는 특징에서 정보를 구성하기 때문에 이러한 결과가 나타난것 같다. ㄴ, ㄹ 과 같은 비음의 경우는 인식율이 다른 것에 비하여 낮다. 성대를 All pole model로 하여 비강의 특성을 전달할 수 없는 FP coding이 비교적 전 대역에 걸쳐 정보를 전달하는 FS coding 보다 더 크게

영향을 받았다. 또 학습에 의한 효과는 두가지 방법 모두 확실히 나타났으나 FP coding 보다는 FS coding이 더 좋은 학습효과를 보여 새로운 소리에 대한 적응력이 더 좋은 것으로 나타났다. 2번의 학습을 통하여 75 %이상의 인식율을 보여 학습을 통하여 일상 대화의 가능성을 보여 주었다. 결과적으로 전기 자극으로 음성정보를 전달하여 장애자로 하여금 그 정보를 인식하게 하기 위해서는 음성의 조음상의 특징만을 전달하는 것보다는 보다 넓은 영역에서 음성의 spectral과 조음상의 특징을 동시에 전달하는 것이 더 효과적이며 여기에는 8개 이상의 전극이 필요하다. 이러한 방법에는 동시자극이라는 문제점이 존재하나 이는 장애자의 생리적인 특징을 고려한다거나, 전극 설계 기술로 극복할 수 있다고 생각한다.

Reference

- [1] Douglas O'shaughnessy " Speech communication human and machin" Addison Welley 1987.
- [2] Robert A Schindler, M.D " Cochlear implants" Reven Press New York.
- [3] Jont B. Allen " Cochlear modeling" IEEE Assp Mag1985.
- [4] Eric, D. Young and Murry B. sachs "Representation of steady stset vowel in the temporal aspec of discharge patterns of population of auditory nerve fiber" J. Acoust. Soc. Am 66 1381 - 1403
- [5] W. D. Keider and P. Finkenzeller "Artificial Auditory stimulation theory" KARGER.
- [6] Mini System 22 Audiologist handbook.
- [7] 김 영훈, 박 광석 "A study on the estimation of auditory filter shape by notch noise masking" 대한 의용 생체 공학회 춘계 학술대회 1990.
- [8] 허용 "국어 음운론"