

청각보철을 위한 PLP방식의
음성신호처리에 관한 연구

김영선* · 최두일* · 박상희* · 백승화**

*연세대학교 전기공학과, **명지대학교 제어계측과

A Study on the Speech Signal Processing for Cochlear Implant using
the PLP Analysis

Kim, Young-Sun · Choi, Doo-Il · Park, Sang-Hui · Beack, Seung-Ilwa

Dept. of Electrical Eng., Yonsei Univ.
Dept. of Control & Instrument Eng., Myongji Univ.

요약

본 논문에서는 감각성 난청자들이 정상인들과 유사한 음성 인식을 하도록 청각 보철 기기를 구성하였다. 음성의 포먼트를 추출하기 위해서는 PLP(Perceptual Linear Prediction) 방식을 이용하였으며, pitch 추출을 위해서는 3 단계 클리핑 함수를 이용한 자기 상관법을 이용하였다. 또한 다중 채널 - 다중 전극 방식을 이용하여 내이의 헤어 셸에 17 개의 전극을 삽입하여 신호를 가하는 시뮬레이션을 하였다. 실험에 사용한 데이터는 모음 /a/, /e/, /i/, /o/, /u/로 전모음과 후모음의 차이를 구별하였으며 두번째 포먼트의 변화와 포먼트 통합 이론에 대한 검증을 하였다.

1. 서론

PLP 방식은 기존의 LP 방식의 공학적인 면에 정신 물리적인 개념을 첨가시킴으로써 사람의 음성 신호를 분석하는데 적절한 음성 스펙트럼 추정 방식으로 무성음, 유성음, 마찰음 등의 여러 형태의 발음에 관계없이 보다 정확한 음성 정보를 얻어낼 수 있고 LP 방식은 AR 계수를 11 - 15개를 가지고 스펙트럼을 추정하는 반면 PLP 방식은 실험적으로 5차일 때 가장 정확한 스펙트럼을 얻을 수 있음이 밝혀져 계산량을 줄일 수 있으며 화자 독립(speaker independent)적인 요소가 LP 방식보다 우수하기 때문에 사람의 나이와 성별에 관계없이 동일한 발음에 거의 같은 스펙트럼을 얻을 수 있다.

따라서, 본 논문에서는 이러한 PLP 방식을 이용하여 음성 신호의 주요 정보인 포먼트를 추출하고 기존의 방식으로 pitch와 intensity를 추출하여 다중 채널 - 다중 전극 방식의 17 채널 스펙트럼 추정 방식 청각 보철 기기를 설계하고자 한다.

2. PLP 방식의 청각 보철기

2.1. 스펙트럼 분석

음성 신호는 Hamming 함수를 이용하여 윈도우한다.

$$W(n) = 0.54 + 0.46 \times \cos[2\pi n \times n / (N - 1)] \quad (1)$$

(N : 윈도우의 크기)

일반적인 윈도우의 크기는 약 20 ms로 하고 신호의 갯수는 256개로 한후 음성 스펙트럼의 실수부와 허수부를 각각 제공하여 더한다.

$$P(\omega) = \text{Re}[S(\omega)]^2 + \text{Im}[S(\omega)]^2 \quad (2)$$

2.2 임계 대역 스펙트럼 분석

스펙트럼 $P(\omega)$ 을 주파수축 ω 를 따라 Bark 주파수축 Ω 로 변환시킨다.

$$\Omega(\omega) = 6 \times \ln(\omega / 1200\pi + [(\omega / 1200\pi)^2 + 1]^{0.5}) \quad (3)$$

이 Bark 주파수 변환은 1977년 Schroeder[1]에 의해 제안되었다. 이렇게 주파수를 왜곡시켜 만들어진 스펙트럼을 임계 대역 차폐 곡선(Fletcher, 1940) $\psi(\Omega)$ 와 컨볼루션시킨다. 여기서 임계 대역 곡선은 다음과 같이 주어진다.

$$\psi(\Omega) = \begin{cases} 0 & \text{for } \Omega < -1.3 \\ 10^{2.5(\Omega + 0.5)} & \text{for } -1.3 \leq \Omega \leq -0.5 \\ 1 & \text{for } -0.5 < \Omega < 0.5 \\ 10^{-1.0(\Omega - 0.5)} & \text{for } 0.5 \leq \Omega \leq 2.5 \\ 0 & \text{for } \Omega > 2.5 \end{cases} \quad (4)$$

이러한 연속적인 임계 대역 차폐 곡선 함수는 Schroeder의 비대칭형 차폐 곡선을 근사화한 것이다. 이 곡선은 청각 필터의 형태가 Bark축 상에서 거의 일정하다는 Zwicker(1970)[2]의 이론을 잘 만족시켜준다. 이 필터의 가장자리의 기울기는 -40 dB이다. $\Psi(\Omega)$ 와 주기적인 우함수인 $P(\Omega)$ 와의 이산 컨볼루션을 통해 임계 대역 파워 스펙트럼을 구한다.

$$\theta(\Omega_i) = \sum_{\Omega=-1.3}^{2.5} P(\Omega - \Omega_i)\Psi(\Omega) \quad (5)$$

상대적으로 넓은 임계 대역의 차폐 곡선 $\Psi(\Omega)$ 과의 컨볼루션으로 $\theta(\Omega_i)$ 가 만들어졌다면 이 $\theta(\Omega)$ 의 정보는 원래의 $P(\Omega)$ 와 비교할 때 달라질 수 있다. 여기서는 $\theta(\Omega)$ 를 1 Bark 간격으로 한다. 일반적으로 0.994 Bark간격으로 18개의 스펙트럼을 사용하여 사람의 음성 주파수 대역폭인 0 - 16.9 Bark(0 - 5 kHz)의 스펙트럼을 분석한다.

2.3 동일 강도 전처리

샘플링된 $\theta(\Omega(\omega))$ 는 동일 강도 곡선 $E(\omega)$ 로 전처리 한다.

$$\Xi(\Omega(\omega)) = E(\omega)\theta(\Omega(\omega)) \quad (6)$$

여기서 함수 $E(\omega)$ 는 서로 다른 주파수에서 청각의 민감도는 각각 다르다는 Robinson과 Dadson(1956)[3]의 이론을 기초로 한 것이며 약 40 dB 수준에서의 청각의 민감도를 시뮬레이션한 것으로 Makhoul과 Cosell[4]이 제시한 다음의 식을 적용시켰다.

$$E(\omega) = [(\omega^2 + 56.8 \times 10^6) \times \omega^4] / [(\omega^2 + 6.3 \times 10^6)^2 \times (\omega^2 + 0.38 \times 10^9)] \quad (7)$$

식(7)은 기울기가 0 Hz에서 400 Hz 사이에서 12 dB/oct, 400Hz에서 1200Hz 사이에서 0dB/oct, 1200 Hz에서 3100Hz 사이에서 6dB/oct, 3100 Hz에서 나이퀴스트 주파수까지 0 dB/oct인 필터의 전달 함수를 나타낸다. 일반적으로 이 식은 5000 Hz까지의 소리를 분석하는데 의미가있으며 나이퀴스트 주파수를 넘어서는 소리의 분석에 있어서는 청각의 민감도의 급격한 감쇄(약 -18 dB/oct)를 나타내는 식이 첨가된다.

$$E(\omega) = [(\omega^2 + 56.8 \times 10^6) \times \omega^4] / [(\omega^2 + 6.3 \times 10^6)^2 \times (\omega^2 + 0.38 \times 10^9) \times (\omega^6 + 9.58 \times 10^{26})] \quad (8)$$

결과적으로 0 Bark에서와 나이퀴스트 주파수에서의 $\Xi(\Omega(\omega))$ 의 값은 같다.

2.4 Intensity - loudness power law

AR 모델링을 하기 바로 전 단계가 스펙트럼의 크기를 1/3 승하는 것이다.

$$\Phi(\Omega) = \Xi(\Omega)^{1/3} \quad (9)$$

이것은 Stevens(1957)[5]가 제안한 청력의 법칙을 근사화한 것이며 소리의 실제 크기(intensity)와 정신적으로 인식되는 크기(loudness)와의 비선형적인 관계를 시뮬레이션한 것이다. 정신 물리학적인 동일 강도 전처리 단계와 함께, 이 단계에서 임계 대역 스펙트럼의 크기의 변동을 줄임으로써 낮은 차수로 AR 모델링을 할 수 있다.

2.5 AR 모델링

PLP 분석의 최종 단계로는 AR 모델링의 하나인 자기 상관 방식(autocorrelation method)을 이용하여 PLP 스펙트럼을 구현하는 것으로 그 과정은 다음과 같다.

- 1) $\Phi(\Omega)$ 에 역 이산 푸리에 변환을 하여 각각의 자기 상관값을 구한다.
- 2) M차의 AR 모델 계수를 구하기 위해 M + 1개의 자기 상관값을 이용하여 Yule - Walker 방정식을 푼다.

PLP 방식에서 AR 차수는 5차일때 가장 정확한 음성 스펙트럼을 추출할 수 있으므로 5차 PLP 방식을 이용한다.

2.6 포먼트의 추출

본 연구에서는 PLP 방식으로 얻어진 log spectrum의 peak picking 방식으로 포먼트를 추출한 후 spectrum의 크기(L)와 주파수(F)와의 미분식을 이용하여 $dL/dF = 0$ 이고 d^2L/dF^2 이 극대값을 갖는 주파수가 포먼트주파수이므로 이 주파수에 대응하는 채널에 자극 신호를 가한다. [6]

2.7 Pitch의 추출

Pitch는 가장 기본적인 특징 매개변수로서, 허파로부터 발생한 공기의 흐름이 성대를 거치면서 성대의 주기적인 진동에 의해 형성된 주파수이다. pitch에 의해 유성음/무성음으로 구분되는데, 유성음은 성대의 진동에 의해 생기는 소리이고, 무성음은 성대가 열린 채로 진동없이 공기를 통과시켜 결과적으로 pitch가 정의되지 않은 소리이다.

입력 신호가 유한개인 시퀀스 $X(n)$ 이라면 자기 상관은 식(10)과 같이 정의된다.

$$r(k) = \sum_{n=-\infty}^{\infty} X(n) X(n+k) \quad (10)$$

단구간 자기 상관을 구하기 위해 유한 길이의 윈도우 세그먼트를 취하면 식(11)과 같다.

$$R_n(K) = \sum_{m=0}^{N-1-K} [X(n+m) W'(m)] [X(n+m+k) W'(k+m)] \quad (11)$$

이와 같이 구한 자기 상관 함수는 포먼트 주파수의 영향, 즉 성도 응답의 감쇠 진동의 영향을 많이 받아 기본 주파수 이외에도 많은 불필요한 정보가 실려 있으므로 이를 제거하고, 아울러 계산량을 줄이기 위해 3단계 센터 클리핑(3-level center clipping)을 실시한다.

이와 같이 구한 자기 상관 함수의 첨두치를 찾아내어 $R_n(0)$ 의 30[%]에 해당하는 한계치와 비교한다. 첨두치가 경계치보다 높으면 유성음으로 구분되고, 반대로 첨두치가 경계치보다 낮으면 무성음으로 분류된다. pitch는 유성음의 경우에만 계산된다.

2.8 Intensity의 추출

Intensity는 음성 에너지의 제공에 비례하는 양으로서, 단구간을 택하면 시간함수인 궤적 형태를 띄게 된다.

음성 신호를 $X[n]$, 단구간 세그먼트 윈도우를 $W[n]$ 이라고 하면

$$I_i = 10 \log_{10} \frac{\sum_{n=0}^N X[n]^2 W[n]^2}{N} \quad (12)$$

여기서, N: 윈도우 개수

2.9 전기 자극 신호

음성 신호의 단구간에서 추출된 정보로부터 신경 자극신호를 합성해 내는 방식은 다음과 같다.

1. 전기 자극은 선음극 후양극의 펄스파의 형태로 가하며 총 전하량은 0이 되도록 한다.
2. 무성음일 경우는 채널을 무작위로 선택하여 자극 신호를 가하되, 자극 신호의 크기는 원 음성신호의 intensity에 비례하도록 한다.
3. 유성음일 경우는 2개의 formant 주파수를 추출하고 채

널에 할당되어 있는 특성 주파수와 가장 유사한 채널을 2개만 선택하여 추출된 pitch의 주기로 자극 신호를 가한다.

4. 내이의 코클리어에 흐를 전기 자극 신호 전류의 확산을 막기 위하여 몇 개의 신호가 동시에 가해지지 않고 약간의 시간 지연을 두고 하나의 채널에만 가한다.

5. 과전류로 인한 내이의 손상을 막기 위하여 한계 전류값을 정하여 intensity가 한계값 이상일 때는 한계 전류만을 흘려주도록 한다.

2.10 전체 시스템의 구성

이상과 같이 추출한 pitch와 intensity, 포먼트를 이용하여 자극 신호를 만드는 전체적인 알고리즘은 다음과 같다.

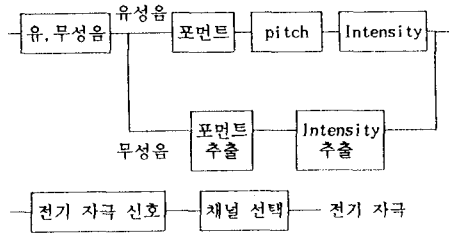


그림 1 전체 시스템 블록 신호

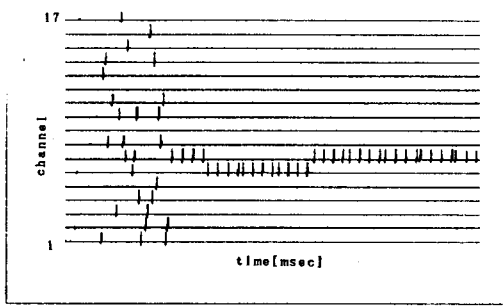
3. 실험 및 결과

표 1은 모음 /a/, /e/, /i/, /o/, /u/의 1,2차 포먼트 주파수를 Bark 값으로 구한 것이다.

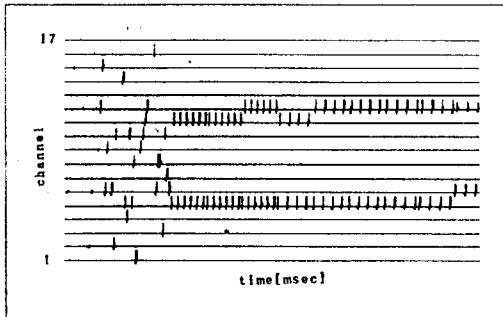
표 1 다섯개 모음의 1,2차 포먼트 주파수

모음	1차 포먼트	2차 포먼트
/a/	6.7 Bark	통합
/e/	5.5 Bark	11.9 Bark
/i/	3.5 Bark	13.9 Bark
/o/	4.5 Bark	10.5 Bark
/u/	4.9 Bark	통합

그림 2는 2장에서 기술한 방법으로 얻은 모음 /a/, /e/에 대한 전기 자극 신호이다. 그림의 자극 신호의 채널을 표 1의 포먼트 주파수와 비교해 볼 때 각 모음의 포먼트 주파수에 맞는 채널에 펄스가 생김을 알 수 있다.



(a) /a/



(b) /e/

그림 2 모음 /a/, /e/의 전기 자극 신호

4. 결론

(1) LP 방식보다 포먼트의 수를 감소시킬 수 있었으며 생리학적인 특성 뿐 아니라 정신 물리학적 특성을 잘 만족시켜 주었고 계산량에 있어서도 많은 감소를 나타냈다.

(2) 청각 모델을 이용한 방식보다 그 구조가 간단하고 한번에 두개 이하의 채널을 선택하여 전기 자극을 가하므로 음성 인식의 혼란을 방지할 수 있었다.

(3) 후모음과 전모음의 포먼트를 비교하였을 때 후모음에서는 한개의 포먼트가, 전모음에서는 두개의 포먼트가 생김을 알 수 있으며 이는 Fant와 Bladon 등이 제안한 효과적인 두번째 포먼트(effective second formant) 이론을 잘 만족시켜 주었다.

(4) 두 포먼트가 3.5 Bark 이하의 거리에 있을 때 Chistovich의 포먼트 통합 이론을 잘 만족시켜 주어 한 개의 포먼트 만이 나타났다.

참고문헌

[1] Schroeder M. R. "Recognition of complex acoustic signals," Life Science Research Report 5, edited by T. H. Bullock, (Abakon Verlag, Berlin), pp 324 (1977).

[2] Zwicker E. "Masking and psychological excitation as consequence of ear's frequency analysis," in Frequency Analysis and Periodicity Detection in Hearing, edited by R. Plomp and G. F. Smoorenburg (Sijthof, Leyden, The Netherlands), (1970).

[3] Robinson D. W. & Dadson, R. S. "A redetermination of the equal-loudness relations for pure tone," Br. J. Appl. Phys. 7, pp 166 - 181 (1956).

[4] Makhoul J. & Cosell L. "LPCW: An LPC vocoder with linear predictive spectral warping," in Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing 76, pp 466 - 469. (1976).

[5] Stevens S. S. "On the psychophysical law," Psychol. Rev. 64, pp 153 - 181. (1957).

[6] 안상필 "LPC 분석을 이용한 조음 운동 추정에 관한 연구," 석사 학위 논문, 연세 대학교 대학원, (1990).