'87 韓國自動制御學術會議論文集 1987. 10.16~17

청각 계통에서의 음성신호처리

o *
이 재 혁, 심 재 성, 윤 태 성, 백 승 화, 박 상 희

연세대학교 전기공학과, * 명지대학교 전기공학과

SPEECH SIGNAL PROCESSING IN THE AUDITORY SYSTEM

o *
J.H. Lee, S.J. Sim, T.S. Yoon, S.H. Beack, S.H. Park

Dept. of Electrical Eng., Yonsei university,
* Dept. of Electrical Eng., Myong-ji university.

ABSTRACT

The speech signal processing in the auditory system can be analysized based on two representations : Average discharge rate and Temporal discharge pattern. But the average discharge rate representation is restricted by the narrow dynamic range because of the rate saturation and the two tone suppression pheonomena, and the temporal discharge pattern representation needs a sophisticate frequency analysis and synchrony measure.

In this paper, a simple representation is proposed : using a model considering the interaction of Cochlear fluid-BM movement and a haircell model, the feature of speech signals (formant frequency and pitch of vowels) is easily estimated in the Average Synchronized Rate.

1. 서 론

인간의 청각시스템은 가장 효율적인 음성신호처리 시스템이지만, 여러 비선형적 현상으로 인해 그 처리 방식을 간단히 이용할 수 없었다. 신경응답을 통한 음성의 특징추출방식으로서 평균발화율은 좁은 동작 영역을 갖는다는 제한점이 있으며[1],[2] 발화패턴해석방식은 주파수분석과 동기측정 등의 복잡한 절차가 필요하다. [3],[4],[5]

본 연구에서는 Yegnanarayanana이 제안한 기저막-유체 상호작용 모델과 [9] Hall 과 Schroeder 가 제안한 헤어셀 모델을 [10] 이용하여 신경발화패턴을 구하고 동기화된 발화패턴으로부터 직접 음성정보를 추출할 수 있음을 보이고자 한다.

2. 청각 시스템의 신호처리 특성

1) 주파수의 분석

내이에 유입된 입력신호는 기저막위를 진행하면서 입력신호의 주파수에 따라 특정지점에서 최대공진을 일으킨다. 기저막에는 헤어셀 (hair cell)이라는 섬모들이 분포되어 있고 이 헤어셀에는 청각신경이 연결되어 있다. 따라서 한 점에서 기저막이 최대 공진을 일으키면 그 지점에 연결된 헤어셀과 청각 신경이 활발한 응답을 보이게 된다.

기저막상의 각 점들은 각기 고유의 특정 주파수에 잘 반응하므로 (이 주파수를 특성 주파수라고 한다.) 유입신호의 주파수는 어느 지점의 청각신경이 최대발화율을 보이느냐에 따라 결정된다.[11]

이 원리를 이용하여 기저막의 모델링은 대개 다채널 대역통과여파기로 이루어진다. [12],[13]

2) 신경응답의 동기(synchrony)현상

순음(pure tone) 자극을 가했을 때 청각신경의 발화 형태를 시간에 따라 지켜보면 그 패턴이 입력 신호의 주기와 유사함을 발견할 수 있다.

즉, 입력신호의 주파수가 발화패턴의 주기로 반영되며 결국 중추 신경계에서는 최대발화율의 위치와 발화패턴의 주기를 통해 입력신호의 정보를 얻는다고 볼 수 있다.[11],[7]

음성신호와 같은 복합음의 경우에는 한 신경의 발화패턴의 주기가 여러 주파수 성분을 반영하고 있다. 이런 경우 어떤 성분들이 어떤 비율로 반영되고 있는지 알아보는 방법을 동기 측정(synchrony measure) 이라고 한다. 만일 발화패턴의 주기가 하나의 주파수 성분만을 반영한다면 그 주파수에 동기화 (synchronized) 되었다고 한다.

3) 발화율과 발화패턴의 해석

한 청각신경의 발화응답을 일정시간 동안 평균한 값을 평균발화율이라 하며 대개 신경의 위치에 따라 배열하여 그림으로 나타낸다.

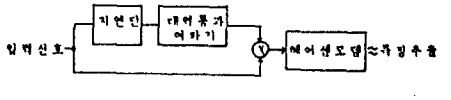
Young 과 Sachs는 음성에 대한 신경발화측정실험을 통해 50 db 이하의 자극에서는 모음의 중요정보인 포먼트와 발화율의 첨두치들이 일치하나 보통 대화시의 자극(70-80 db)에 대한 평균발화율에서는 포먼트의 특징을 잘 알아볼 수 없음을 밝혀 냈다. 이는 신경의 발화율 포화(rate saturation) 현상과 이음억제(two tone suppression) 현상 때문이다.[2],[4]

따라서 모음에 대한 청각시스템의 특징추출을 설명하는 방식으로서 평균발화율방식은 제한점이 많으며, 더우기 피치에 대한 정보를 반영하고 있지 않다.[7]

이에 반해 발화패턴의 동기비율(synchronization index)의 분석은 자극의 세기에 상관없이 포먼트 주파수의 추정과 피치의 검출이 가능하다. 그러나 이 방식은 발화패턴의 주파수 분석 및 대역 정규화 등의 복잡한 알고리듬을 거쳐야 한다.

3. 모델의 구성

본 연구에서는 청각시스템에 유입된 음성신호로부터 신경발화패턴을 얻고 이 발화패턴으로부터 음성정보를 추출하기 위해 다음과 같은 모델을 사용하였다.



1) 기저막의 운동

음성신호는 기저막상을 진행하며 각 점에서 공진을 일으킨다. 이 운동은 직연단이 붙은 고전적인 1차원 모델로 해석되며 [12],[13],[17] R-L-C 공진회로로 이루어지는 대역통과여파기로 구현된다.

그 식은 다음과 같다.

$$S(t-Dn) = \frac{jw \times f(t-Dn) \times wno / Qn}{wno^2 - w^2 + jwno/Qn}$$

이때, wno: 공진각주파수, f(t): 입력신호,
Qn : n번째여파기의 선택도, g(t): 출력신호

2) 유체압력과의 상호작용

내이에 전달된 음성신호의 압력은 기저막상에 전달되는 진행파(propagation wave)를 발생시키고, 또한 내이안의 외임파액에도 수축파(compression wave)를 발생시킨다.[15] 따라서 한 점의 헤어셀은 기저막의 진행파와 수축파를 동시에 받게 된다. 액체내로 전달되는 수축파는 시간지연이 없는 원래의 입력신호와 동일하다.

1985년 Yegnanarayanan은 청각시스템에서의 피치의 추정과 이음억제와 같은 비선형 현상을 설명하기 위해 헤어셀에 미치는 진행파와 수축파의 상호작용을 곱셈과 반파정류로 가정하였다.[9] 본 연구에서는 음성신호의 정보처리에 이 상호작용을 고려함으로서 특성 주파수에 동기강화(synchrony enhance)된 발화패턴을 얻을 수 있었다.

3) 헤어셀 모델

헤어셀 모델은 1979년 Hall과 Schroeder 가 제안한 모델이다.[10] 이모델은 헤어셀-청각신경의 응답을 발화확률의 형태로 나타내며 발화율 포화, 순응현상(adaptation), 위상동기등의 현상을 구현할 수 있다.

모델의 식은 다음과 같다.

$$dn(t)/dt = r - n(t) \times [g + p(t)],$$
$$f(t) = n(t) \times p(t),$$
$$p(t) = po\,[s(t)/2 + [s^2(t)/4 + 1]^{1/2}],$$

4. 실험 및 결과

음성 신호는 8 명의 20 대 남성으로부터 얻은 5 개의 기본모음을 3.5 KHz의 차단주파수를 갖는 저역통과 여파기를 거쳐 10 KHz 샘플링하여 사용하였다. 본 실험결과와 비교분석하기 위해 각 모음의 256 개 샘플에 대한 15차의 선형예측계수과 자기상관함수를 구해 전력스펙트럼과 피치를 추정하였다.
기본 모음 /아/ 에 대한 모델의 응답은 다음과 같다.

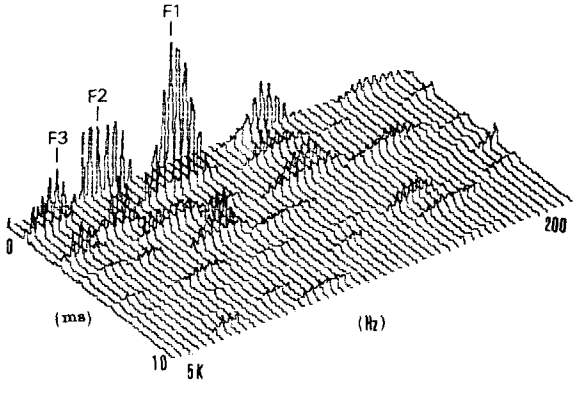


그림 1. 모음 /아/에 대한 신경출력도

그림 1은 시간의 진행에 따른 모델의 출력을 나타낸 그림이다. 본 모델의 출력은 청각신경의 발화 패턴이므로 그림 1은 신경출력도(Neurogram)의 형태를 갖는다. 각 신경의 출력의 첨두치가 몇 군데의 지점에서 집중됨을 알 수 있는데 이것은 입력신호의 주파수성분과 관계가 있다. 실제로 모음 /아/의 포먼트 주파수를 대응시켜 보면 첨두치의 밀집지역과 거의 일치함을 알 수 있다.

그림 1에 사용된 자극의 세기는 70 dh SPL 로서 실제 신경응답에서는 발화율 포화현상 때문에 포먼트 주파수의 특징을 알아볼 수 없는 크기이다. [12], [14] 그러나 본 모델의 신경출력도에서는 포먼트 주파수의 위치가 정확히 나타나고 있다. 그 이유는 유체 압력과 기저막 공진의 상호작용으로 각 청각섬유의 응답이 그 섬유 고유의 특성 주파수에 동기되었기 때문이다.

따라서 그림 1은 입력신호를 청각시스템에 통과시켜 얻은 유사스펙트로그램(pseudo spectrogram)으로 볼 수 있다.[8],[17] 또 이 유사스펙트로그램은 시간에 따른 신경응답의 변화를 보여주기 때문에 포먼트 주파수의 추이를 관찰할 수 있다.

그림 2 는 10 msec 동안의 신경출력도를 각각의 특성주파수에 대해 평균한 그림이다. 각각의 그림은 그림 1과 동일한 /아/의 모음에 주어진 10.4 msec 이므로 10 msec 동안 포먼트 주파수의 위치는 변하지 않는다고 볼 수 있다.[12] 그림 2.(a)는 평균 동기화 발화율이며 실선은 /아/ 전력스펙트럼이다.

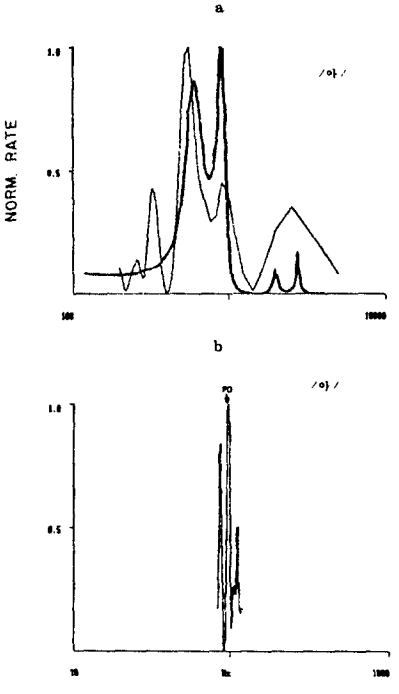
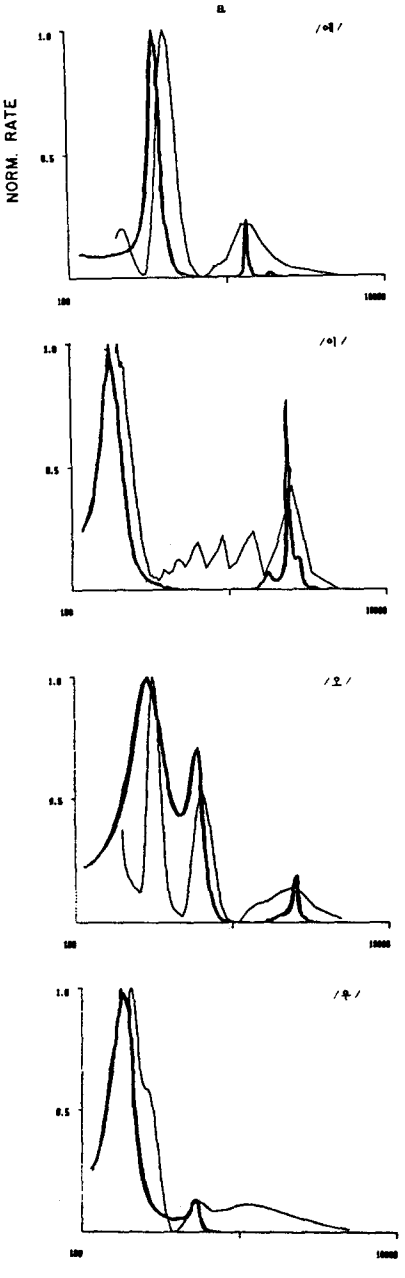


그림 2.(a) 평균 동기화 발화율과 전력스펙트럼
(b) 피치에 대한 평균 동기화 발화율

그림 2.(a)는 200 Hz - 5 KHz까지의 특성주파수를 범위로 하는 평균 동기화 발화율을 정규화비율로 나타낸 것이다. 전력스펙트럼과 비교해 보면 제2,제3 포먼트 주파수의 위치는 일치되며 제1 포먼트 주파수는 약 27 Hz 정도 낮다. 그리고 전력스펙트럼에는 나와있지 않은 첨두치가 333 Hz 근방에 발생했음을 알 수 있다.

그림 2.(b)의 특성주파수 범위는 90 Hz - 110 Hz 이며 (a)와 같이 정규화하였다. (b)의 그림에서는 약 98 Hz 에서 큰 첨두치를 볼 수 있는데 이는 입력신호의 기본주파수 (fundamental frequency, 피치에 해당함) 인 96.2 Hz에 매우 근접한다. 즉, 동기화된 신경발화패턴으로부터 음성신호의 피치를 근사적으로 추정해 낼 수 있다.

그림 3는 4 개의 기본모음의 1주기에 대한 평균 동기화 발화율을 나타낸 것이다. (a)의 특성주파수는 200Hz - 5KHz 이며 (b)는 80Hz - 120 Hz이다. (b)의 화살표는 실제 피치를 나타내고 있다.



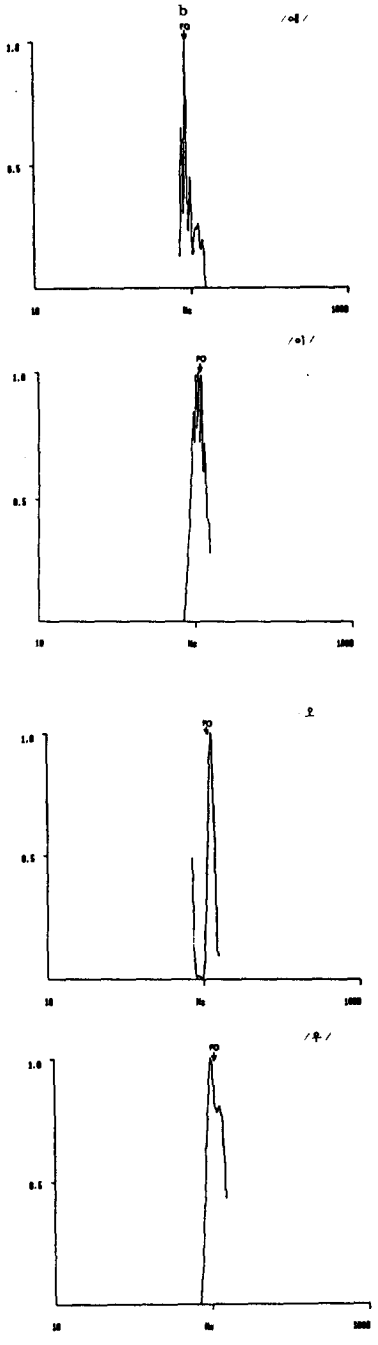


그림 3. 4개 기본모음에 대한 평균 동기화 발화율

5. 결 론

본 연구에서는 청각계통의 제반 특성을 반영하는 모델을 사용하여 신경발화패턴으로부터 음성신호의 특징을 직접 추출하였다.

본 연구의 결과는 평균발화율 방식보다 입력신호의 자극세기에 대해 안정적이며 기계적운동-신경응답의 변환과정에서 동기강화 (synchrony enhance)를 시킴으로써 신경응답을 직접 모음의 특징추출에 사용할 수 있었다. 향후의 연구에서 자음 및 단음의 정보추출이 이루어지면 청각계통의 특성을 이용한 음성인식도 가능해 질 것으로 보인다.

<참 고 문 헌>

1. D.O.Kim, C.E.Molnar, " A population study of Coclear nerve fibers : comparison of spatial distributions of average rate and phase locking measures," J.Neurophysiology, 42, 1979.

2. E.D.Young, M.B.Sachs, " The representation of steady-state vowels in the temporal aspects of the discharge patterns of populations of auditory nerve fibers," JASA., 66, 1979.

3. B.Delgutte, Y.S.Kiang, "Speech coding in the auditory nerve : I. Vowel-like sounds," JASA., 75, 1984.

4. M.B.Sachs, E.D.Young, "Effects of nonlinearities on speech coding in the auditory nerve," JASA., 68, 1980.

5. S.A.Shamma, "Speech processing in the auditory system I : the representation of speech sounds in the reponse of the auditory nerve," JASA., 78, 1985.

6. J.L.Goldstein, P.Srulovicz, "A central spectrum model : a synthesis of auditory nerve timing and place cues in monaural communication of frequency spectrum," JASA., 73, 1983.

7. C.D.Geisler, "Coding of acoustic signals on the auditory nerve," IEEE. ENG. Medicine & Biology, 6, 1987.

8. S.Seneff, "Pitch and spectral analysis of speech based on auditory synchrony model," Ph.D thesis, M.I.T., 1985.

9. G.Yegnanarayanan, "A new model of hearing and its performance in pitch perception," Ph.D thesis, Delaware univ., 1985.

10. J.L.Hall, M.R.Schroeder, "Model for mechanical neural tranduction in the auditory receptor," JASA., 5, 1974.

11. W.A.Yost, D.Nielsen, Fundamentals of Hearing, Holt, Rinehart and Winston, 1985.

12. J.B.Allen, "Cochlear modelling," IEEE. ASSP. Magazine, 1985.

13. J.L.Flanagan, Speech Analysis, Synthesis and Perception, Springer-Verlag, 1972.

14. 박상희, 백승화, 윤태성, 이재혁, " 청각보철을 위한 자극패턴추출에 관한 연구," 전기.전자공학학술대회논문집(II), 1987.

15. G. von Bekesy, Experiments in Hearing, E.G. Wever(ed.), Kriger, Newyork, 1980.

16. L.R.Rabiner, R.W.Schafer, Digital processing of speech signals, Prentice-Hall, 1978.