

장애음성의 주기성분과 잡음성분의 분리 방법에 관하여

조철우 리타오
창원대학교 메카트로닉스 공학부

Separation of Periodic and Aperiodic Components of Pathological Speech Signal

Cheolwoo Jo Tao Li
School of Mechatronics, Changwon National University

cwjo@sarim.changwon.ac.kr

Abstract

The aim of this paper is to analyze the pathological voice by separating signal into periodic and aperiodic part. Separation was performed recursively from the residual signal of voice signal. Based on initial estimation of aperiodic part of spectrum, aperiodic part is decided from the extrapolation method. Periodic part is decided by subtracting aperiodic part from the original spectrum. A parameter HNR is derived based on the separation. Parameter value statistics are compared with those of Jitter and Shimmer for normal, benign and malignant cases.

I. 서론

음성신호를 음원과 성도필터로 모델링할 경우 전형적으로 음원은 유성음성분을 나타내는 주기적 성분과 무성음 성분을 나타내는 잡음성분으로 구성된 것으로 간주한다. 그러나 실제 음성의 음원은 주기성분과 잡음성분이 복합적으로 구성되어 있다. 음성신호를 이와 같이 주기성분과 비주기 성분으로 분리하는 기법은 음원의 추정 뿐만 아니라 음성의 특성 분석, 합성음성에서의 자연성 구현 등에서 여러가지로 활용될 수 있다. 이와 같은 목적을 위하여 지금까지 여러가지 기법들이

개발되어 있다. 이러한 방법들로는 정현과 성분분석에 의한 방법[1], 하모닉+잡음 모델에 의한 방법[2] 등이 있다.

본 실험의 목적은 음원 분리 방법을 통해 장애음성의 특징을 조사하는 방법을 개발하고 나아가서 장애음성의 분류 식별을 위한 파라미터를 조사하는 데 기초 자료로 삼고자 하는 것이다.

본 논문에서는 Yegnanarayana 등이 제안한 방법[3,4]을 바탕으로 장애음성신호로부터 주기성분과 비주기성분을 분리 구분하고 구해진 파라미터에 의해 정상음성과 장애음성을 식별하는 실험을 해보았다.

II. 적용 알고리즘

실험에 적용한 반복적 방법에 의한 주기성분과 비주기성분의 분리 알고리즘은 다음과 같다.

음성신호를 $s(t)$, 음원신호 또는 선형예측오차를 $e(t)$, 성도특성을 $v(t)$ 라고 할 때

$$s(t) = e(t) * v(t) = (p(t) + r(t)) * v(t)$$

과 같이 표현할 수 있다. 이는 주파수 영역 표현으로

$$\begin{aligned} S(\omega) &= |S(\omega)|e^{j\theta(\omega)} \\ &= [P(\omega) + R(\omega)]V(\omega) \end{aligned}$$

$$= (|P(\omega)|e^{j\theta_p(\omega)} + |R(\omega)|e^{j\theta_r(\omega)})|V(\omega)|e^{j\theta_v(\omega)}$$

$$|S(\omega)|^2 = [|P(\omega)|^2 + |R(\omega)|^2]|V(\omega)|^2 \quad (3)$$

과 같이 표현된다.
양 변에 로그를 취하면

$$\log|S(\omega)| = \log|P(\omega) + R(\omega)| + \log|V(\omega)|$$

과 같이 된다.

본 논문에서는 선형예측 오차신호의 스펙트럼을 주기 성분과 비주기성분으로 분리하고자 한다.

만약 음원신호에 해당하는 오차신호의 스펙트럼으로부터 주기성분의 하모닉 성분에 해당하는 성분을 모두 제거한 잡음성분에 해당하는 비주기성분을 $E(k)$ 라고 한다면 비주기성분의 추정치 R_0 는 다음과 같이 주어질 수 있다.

$$R_0(k) = \begin{cases} E(k), & \text{for } k \in F_r(\text{noise regions}) \\ 0, & \text{otherwise} \end{cases}$$

이를 IDFT하고 N점으로 확장하면

$$\hat{r}_0(n) = \begin{cases} r_0, & \text{for } n < N/2 \\ 0, & \text{otherwise} \end{cases}$$

m 번째 반복식은 다음과 같이 된다.

$$R_m(k) = \begin{cases} E(k), & \text{for } k \in F_r \\ R_{m-1}(k), & \text{otherwise} \end{cases}$$

여기서 $E(k)$ 는 원래의 잡음성분, $R_{m-1}(k)$ 은 이전의 추정치이다.

$$\hat{r}_m(n) = \begin{cases} r_m(n), & \text{for } n < N/2 \\ 0, & \text{otherwise} \end{cases}$$

이상과 같은 방법을 반복하여 원래 스펙트럼과 추정된 스펙트럼과의 오차가 최소가 되도록 하는 R_m 을 구한다. 이렇게 하여 비주기성분의 추정치가 구해진다. 다음 식은 최소자승오차를 나타낸다.

$$\frac{1}{N} \sum_{k=1}^N |R(k) - R_m(k)|^2$$

이렇게 추정된 비주기성분의 스펙트럼을 원래의 스펙트럼으로부터 차감하여 주기성분을 구한다.

구해진 주기성분과 비주기성분의 에너지를 이용하여 다음과 같이 두 성분의 비를 나타내는 파라미터를 도출한다. 이는 앞서의 방법으로 구해진 주기성분과 비주기성분의 비율을 나타내는 파라미터로 향후 음원의 특성을 나타내는 파라미터로 유용하게 활용될 수 있

다.

$$HNR = 10 \times \log_{10} \left(\frac{E_p}{E_{ap}} \right)$$

E_p 은 주기성분의 에너지를 E_{ap} 은 비주기성분의 에너지를 나타낸다

그림3.1은 정상음성의 모음음성의 경우 4KHz까지의 대역에서의 스펙트럼을 보여준다. 그림3.2는 원래 스펙트럼으로부터 추정된 비주기성분의 초기 추정치를 보여준다.

그림3.3에서 그림3.5까지는 50회 반복한 경우의 추정된 주기성분 및 비주기성분의 스펙트럼을 나타낸다. [5]

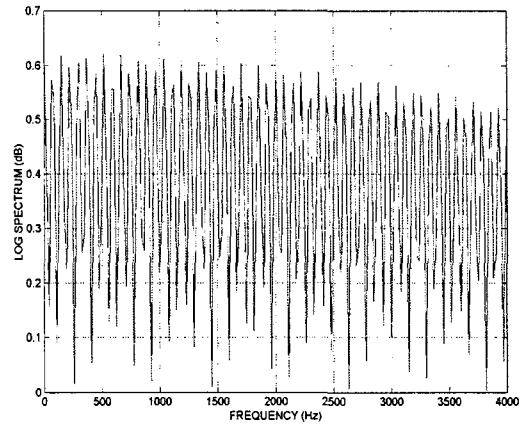


그림3.1 저대역 음성신호의 예측오차스펙트럼

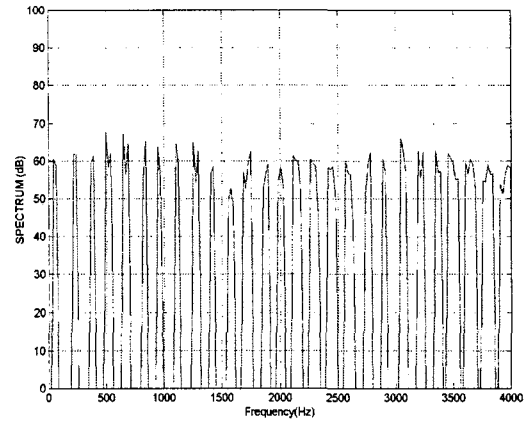


그림3.2 비주기성분의 스펙트럼 초기치

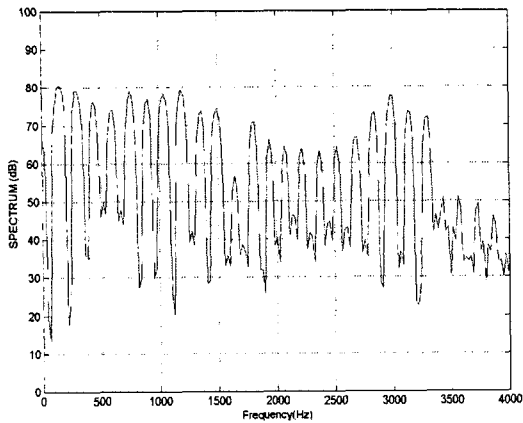


그림3.3 추출된 주기성분 (50회 반복후)

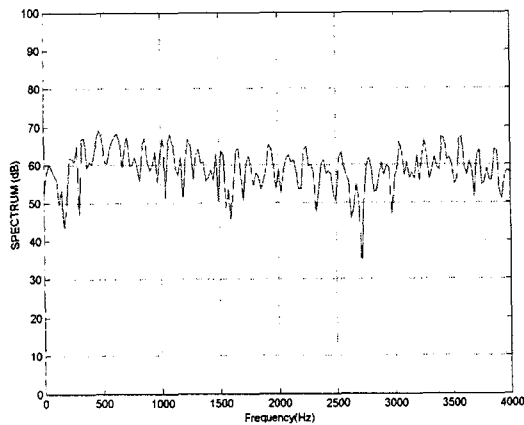


그림3.4 추출된 비주기성분 (50회 반복후)

Ⅲ. 실험 결과

1. 데이터

본 실험에 사용된 데이터는 정상음성, 양성종양음성, 악성종양음성 각 10개씩이다. 각 데이터는 16bits, 50KHz로 표본화되었으며 병원의 진료실에서 DAT 레코더를 통해 수집되었다. 화자는 양성 및 악성의 경우 주로 60대 남성이며 정상음성의 경우는 20~30대의 음성위주로 수집되었다. 이것은 성대 질환이 주로 고령자에게서 발생하는 때문이며 정상음성은 고령자의 음성수집이 쉽지 않아 병원의료진을 대상으로 수집하였다.

2. 분석결과

그림3.9에서 그림3.7까지는 장애음성중 악성 종양에 대해 주기성분과 비주기성분을 추출한 결과이다. 장애음성에 대해 주기성분과 비주기성분을 분리한 이후 구한 HNR을 이용하여 기존 파라미터인 Jitter, Shimmer와 같이 정상, 양성, 악성의 경우 파라미터의 분포를

비교해 보았다. 그림3.8과 3.9에서는 Jitter와 Shimmer의 경우 정상 음성이 양성종양의 경우와 악성종양의 경우에 비해 작은 값을 갖는 반면 HNR의 경우는 정상음성이 큰 값을 갖는 것을 볼 수 있다. HNR에서 양성종양의 값이 악성의 경우보다 더 크게 나왔는데 양성인 경우 평균값에 비한 표준편차의 값이 커서 실제로는 악성의 경우와 거의 비슷한 분포를 보임을 알 수 있었다.

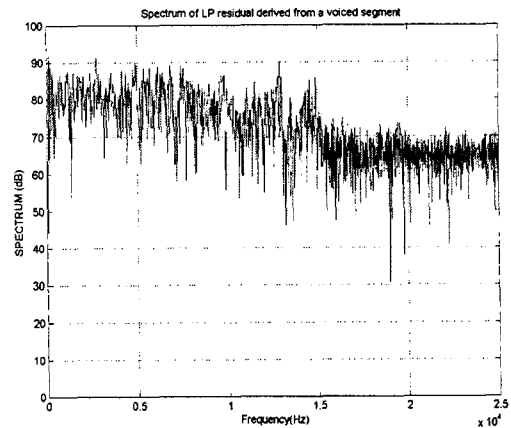


그림3.5 광대역 장애음성(Malignant)의 스펙트럼

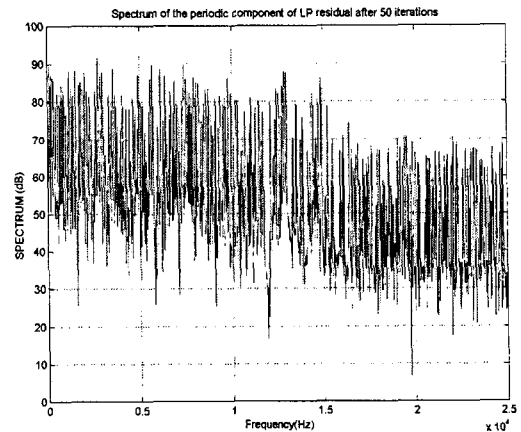


그림3.6 추출된 광대역 장애음성의 주기성분

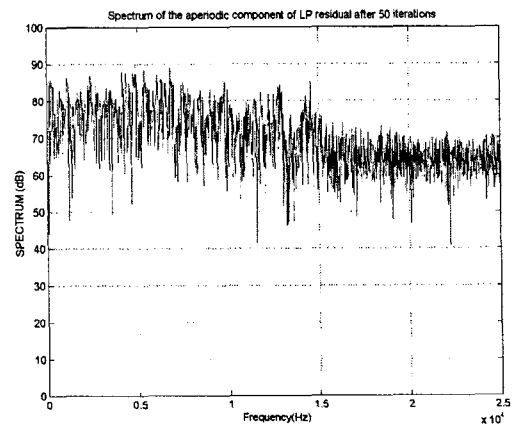


그림3.7 추출된 광대역 장애음성의 비주기성분

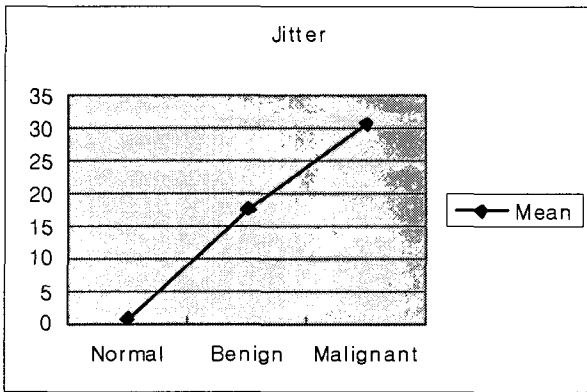


그림3.8 Jitter의 변화

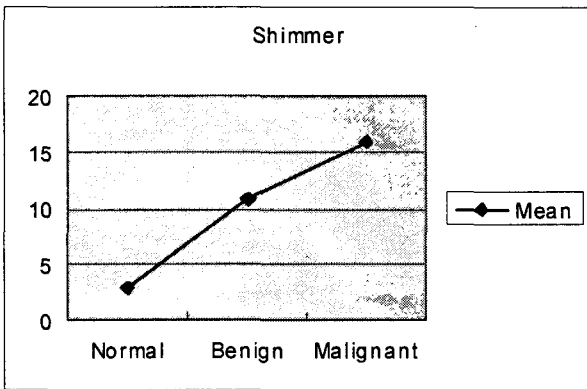


그림3.9 Shimmer의 변화

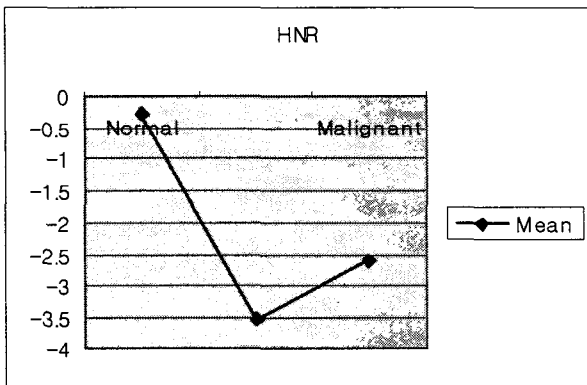


그림3.10 HNR의 변화

IV. 결론

본 논문에서는 장애음성신호를 주기, 비주기 성분으로 분리하는 방법에 의해 구한 파라미터 HNR을 통하여 정상, 양성, 악성 종양간의 구분을 시도해 보았다. 구해진 HNR에 의해 정상음성과 비정상 음성간은 식별은 가능하였으나 양성과 악성의 경우는 뚜렷하게 구분해 주지 못하였다.

실험에서 사용한 반복적 방법은 음성신호의 주기성

분과 비주기성분을 분리하는 방법의 하나로 제안된 것을 사용하였으나 잡음성분이 많고 대역폭이 큰 장애음성의 경우, 특히 심한 악성종양의 경우에는 적용이 쉽지 않음을 발견했다.

앞으로 장애음성을 처리하기에 적합한 모델의 개발이 필요하다고 생각된다.

감사의 글

본 논문은 보건복지부 보건의료기술 진흥사업의 지원에 의하여 이루어진 것임.

(과제고유번호: 02-PJ1-PG10-31401-0005)

참고문헌

- [1] T.F.Quatieri, R.J.McAulay, "Speech Transformations based on a sinusoidal representation", IEEE Trans. on ASSP, vol. ASSP-34, pp.1449-1464, 1986
- [2] J.Laroche, Y.Stylianou, E.Moulines, "HNS: Speech modification based on a harmonic + noise model", Proc. IEEE-ICASSP'93, pp.550-553, 1993
- [3] B.Yegnanarayana, Christophe d'Alessandro, Vassilis Darsinos, "An Iterative Algorithm for Decomposition of Speech Signals into Periodic and Aperiodic Components", IEEE Trans. on Speech and Audio Processing, Vol.6, No.1, pp.1-10, Jan. 1998
- [4] C.d'Alessandro, V.Darsinos, B.Yegnanarayana, "Evaluation of a periodic/aperiodic speech decomposition method for analysis of voice sources", IEEE Trans. on Speech and Audio Processing, Vol.6, No.1, pp.12-23 Jan. 1998
- [5] 조철우, 리타오, "반복적 방법에 의한 음성신호의 주기성분과 잡음성분의 분리에 관해", 2003음성통신 및 신호처리 워크샵, 한국음향학회, 2003. 8