

음성 신호로부터 주기, 비주기 성분의 반복적 계산법에 의한 분리 실험

조철우(창원대학교), 리타오(창원대학교)

<차 례>

- | | |
|------------|------------|
| 1. 서론 | 3. 실험 및 결과 |
| 2. 적용 알고리즘 | 4. 결론 |

<Abstract>

Iterative Computation of Periodic and Aperiodic Part from Speech Signal

Cheol-Woo Jo, Tao Lee

source of speech signal is actually composed of combination of periodic and aperiodic components, although it is often modeled to either one of those. In the paper an experiment which can separate periodic and aperiodic components from speech source. Linear predictive residual signal was used as a approximated vocal source the original speech to obtain the estimated aperiodic part. Iterative extrapolation method was used to compute the aperiodic part.

* Keywords: Separation of periodic, Noise, Iterative, Speech signal

1. 서 론

음성 신호를 음원과 성도 필터로 모델링할 경우 전형적으로 음원은 유성음 성분을 나타내는 주기적 성분과 무성음 성분을 나타내는 잡음 성분으로 구성된 것으로 간주한다. 그러나 실제 음성의 음원은 주기 성분과 잡음 성분이 복합적으로 구성되어 있다. 음성 신호를 이와 같이 주기 성분과 비주기 성분으로 분리하는 기법은 음원의 추정뿐만 아니라 음성의 특성 분석, 합성 음성에서의 자연성 구현 등에서 여러 가지로 활용될 수 있다. 이와 같은 목적을 위하여 지금까지 여러 가지 기법들이 개발되어 있다. 이러한 방법들로는 정현파 성분 분석에 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 및 식별을 위한 파라미터를 조사하는데 기초 자료로 삼고자 하는 것이다.

본 논문에서는 Yegnanarayana 등이 제안한 방법[3,4]을 바탕으로 음성 신호로부터 주기 성분과 비주기 성분을 분리 구분하고 반복 회수에 따른 영향을 살펴보고 이 방법의 광대역 신호에의 적용 가능성을 검토해 보았다.

2. 적용 알고리즘

실험에 적용한 반복적 방법에 의한 주기 성분과 비주기 성분의 분리 알고리즘은 다음과 같다.

음성 신호를 $s(t)$, 음원 신호 또는 선형예측 오차를 $e(t)$, 성도 특성을 $v(t)$ 라고 하면 식 (1)과 같이 표현할 수 있다.

$$s(t) = e(t) * v(t) = (p(t) + r(t)) * v(t) \quad (1)$$

이는 주파수 영역 표현으로는

$$\begin{aligned} S(\omega) &= |S(\omega)|e^{j\theta(\omega)} \\ &= [P(\omega) + R(\omega)]V(\omega) \\ &= (|P(\omega)|e^{j\theta_p(\omega)} + |R(\omega)|e^{j\theta_r(\omega)})|V(\omega)|e^{j\theta_v(\omega)} \end{aligned} \quad (2)$$

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

과 같이 표현된다. 양변에 로그를 취하면 식 (4)와 같이 된다.

$$\log|S(\omega)| = \log|P(\omega) + R(\omega)| + \log|V(\omega)| \quad (4)$$

본 논문에서는 선형예측 오차 신호의 스펙트럼을 주기 성분과 비주기 성분으로 분리하고자 한다.

만약 음원 신호에 해당하는 오차 신호의 스펙트럼으로부터 주기 성분의 하모닉 성분에 해당하는 성분을 모두 제거한 잡음 성분에 해당하는 비주기 성분을 $E(k)$ 라고 한다면 비주기 성분의 추정치 R_0 는 다음과 같이 주어질 수 있다.

$$R_0(k) = \begin{cases} E(k), & \text{for } k \in F_r, (\text{noise regions}) \\ 0, & \text{otherwise} \end{cases} \quad (5)$$

이를 IDFT하고 N 점으로 확장하면

$$\hat{r}_0(n) = \begin{cases} r_0, & \text{for } n < N/2 \\ 0, & \text{otherwise} \end{cases} \quad (6)$$

m 번째 반복식은 다음과 같이 된다.

$$R_m(k) = \begin{cases} E(k), & \text{for } k \in F_r \\ R_{m-1}(k), & \text{otherwise} \end{cases} \quad (7)$$

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

$$\hat{r}_m(n) = \begin{cases} r_m(n), & \text{for } n < N/2 \\ 0, & \text{otherwise} \end{cases} \quad (8)$$

이상과 같은 방법을 반복하여 원래 스펙트럼과 추정된 스펙트럼과의 오차가

최소가 되도록 하는 R_m 을 구한다. 이렇게 하여 비주기 성분의 추정치가 구해진다. 다음 식 (9)는 최소 자승 오차를 나타낸다.

$$\frac{1}{N} \sum_{k=1}^N |R(k) - R_m(k)|^2 \quad (9)$$

이렇게 추정된 비주기 성분의 스펙트럼을 원래의 스펙트럼으로부터 차감하여 주기 성분을 구한다.

구해진 주기 성분과 비주기 성분의 에너지를 이용하여 다음과 같이 두 성분의 비를 나타내는 파라미터를 도출한다. 이는 앞서의 방법으로 구해진 주기 성분과 비주기 성분의 비율을 나타내는 파라미터로 향후 음원의 특성을 나타내는 파라미터로 유용하게 활용될 수 있다.

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

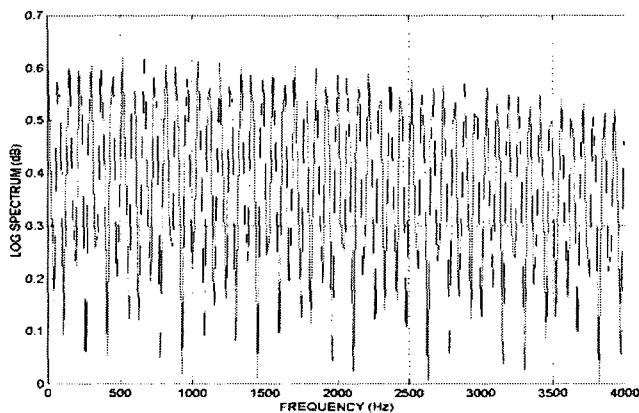
E_p 은 주기 성분의 에너지를 E_{ap} 은 비주기 성분의 에너지를 나타낸다.

3. 실험 및 결과

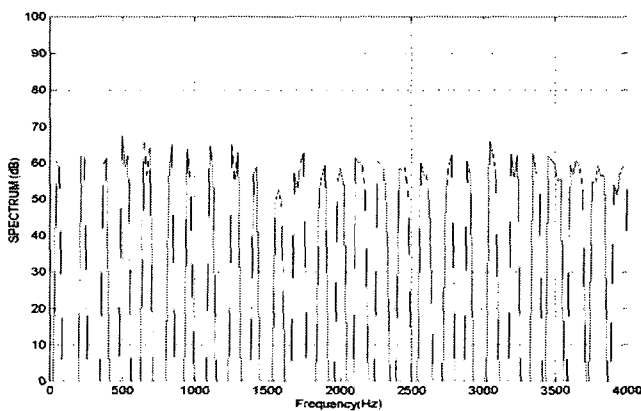
<그림 1>은 실험에 사용한 모음 음성의 스펙트럼을 보여준다. <그림 2>는 원래 스펙트럼으로부터 추정된 비주기 성분의 초기 추정치를 보여준다.

<그림 3>에서 <그림 5>까지는 반복 회수에 따라 추정된 비주기 성분의 스펙트럼을 나타내며 <그림 6>에서 <그림 8>까지는 반복 회수에 따라 추정된 주기 성분의 스펙트럼을 나타낸다.

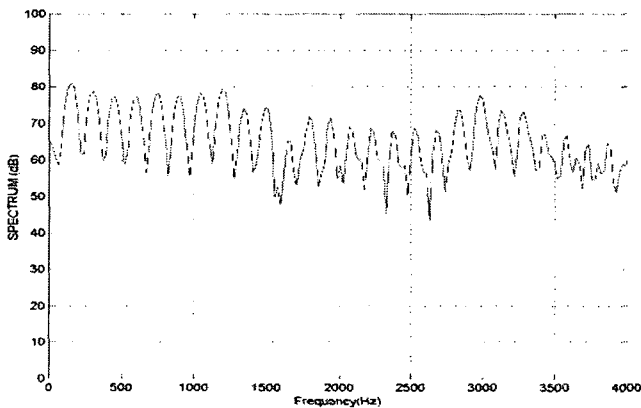
그림에서 볼 수 있듯이 반복 회수가 늘어남에 따라 주파수 성분의 변화가 관찰되나 20 회 이후는 크게 변하지 않음을 알 수 있다.



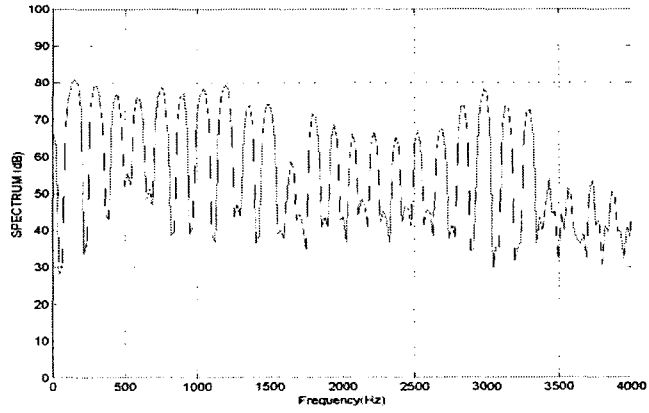
<그림 1> 유성 음성 신호의 스펙트럼



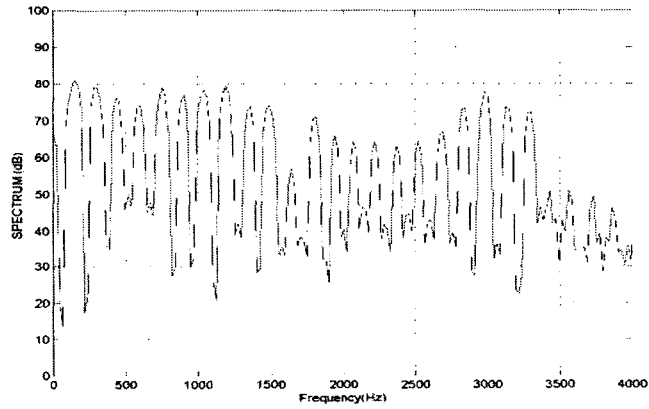
<그림 2> 비주기 성분의 스펙트럼 초기치



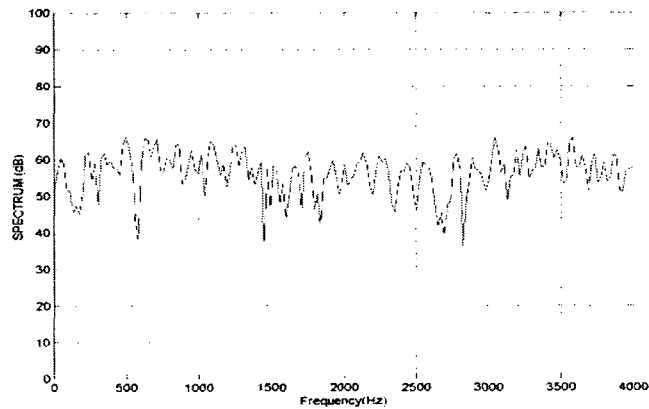
<그림 3> 주기 성분 (10회 반복 후)



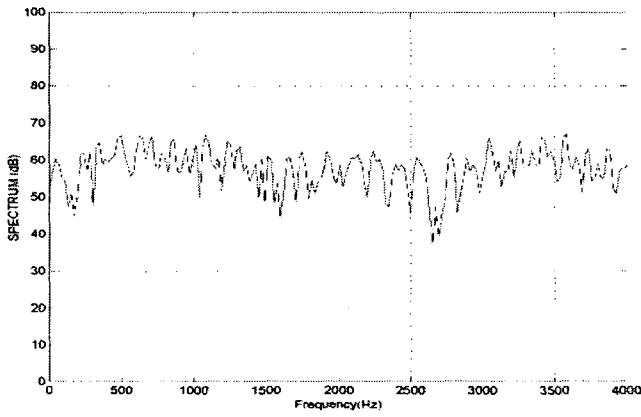
<그림 4> 주기 성분 (20회 반복 후)



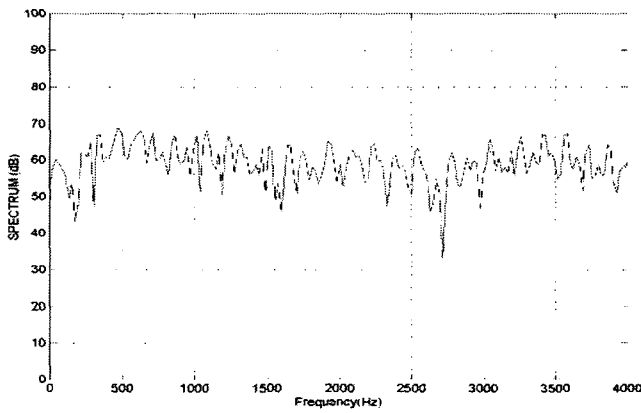
<그림 5> 주기 성분 (50회 반복 후)



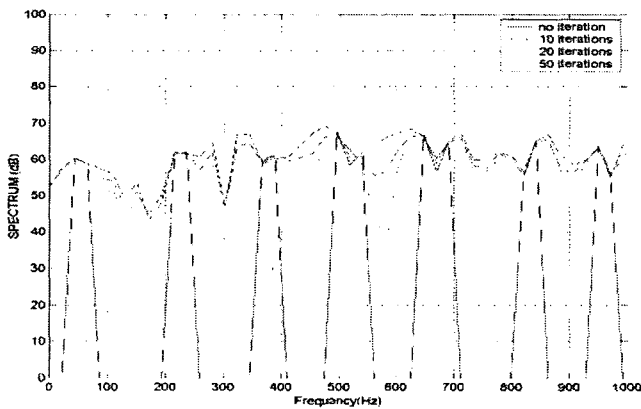
<그림 6> 비주기 성분 (10회 반복 후)



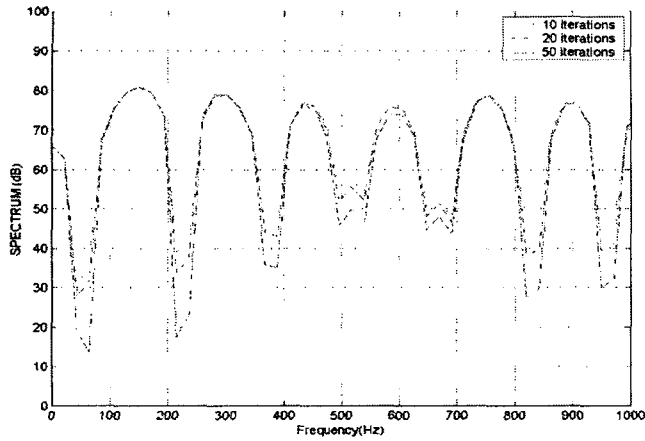
<그림 7> 비주기 성분 (20회 반복 후)



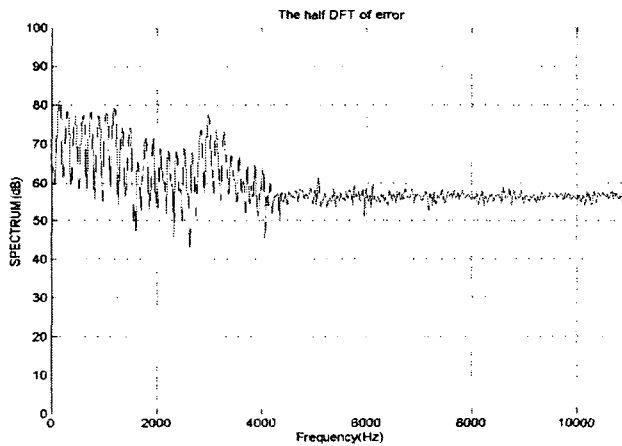
<그림 8> 비주기 성분 (50회 반복 후)



<그림 9> 반복 회수에 따른 비주기 성분의 영향



<그림 10> 반복 회수에 따른 주기 성분의 영향



<그림 11> 광대역 음성 신호의 사례

<그림 9>와 <그림 10>은 반복 회수에 따른 주기 및 비주기 성분의 변화를 1000Hz까지의 대역에 대해 확대하여 그린 것이다. 반복 회수가 늘어남에 따라 세밀한 부분까지 추정이 됨을 볼 수 있다. 그러나 <그림 11>에서와 같이 대역폭이 넓은 음성 신호의 경우는 4KHz 이상의 대역에서 지금과 같은 유성음의 하모닉 성분을 일률적으로 가정한다는 것이 어렵고 실제로 잡음 성분이 우세하므로 스펙트럼의 구성에 대한 좀 더 세부적인 가정이 필요하다고 생각된다.

4. 결 론

본 논문에서는 음성 신호의 음원을 주기 성분과 비주기 성분으로 구분하는 실험을 반복적 방법을 통해 구현하고 실제 음성 신호에 대해 분석한 결과를 기존의 장애음성 처리 파라미터와 비교하여 보았다.

주기성과 비주기성의 구분은 4KHz 이하의 저 대역에서는 제시된 방법이 잘 적용이 되었으나 장애음성과 같이 50KHz로 표본화된 음성 신호에 대해서는 적용이 힘들어 방법의 적절한 변형이 필요함을 알 수 있었다. 또한 제시된 HNR은 기존의 장애음성 파라미터들과 더불어 장애음성 식별을 위한 파라미터로 활용될 수 있을 것이다.

향후의 연구에서는 보다 광대역의 음성 신호로부터 주기 성분과 비주기 성분을 구분할 수 있도록 방법을 개선할 필요가 있다.

감사의 글

본 논문은 보건복지부 보건의료기술 진흥사업의 지원에 의하여 이루어진 것임.
(과제고유번호: 02-PJ1-PG10-31401-0005)

참 고 문 헌

- [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, C. d'Alessandro, V. 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.

접수일자: 11월 3일

게재결정: 12월 12일

▶ 조철우(Cheol-Woo Jo)

주소: 경남 창원시 사림동 9번지 창원대학교 메카트로닉스 공학부

소속: 제어계측공학과 음성 및 음향신호처리 실험실

전화: 055) 279-7552

E-mail: cwjo@changwon.ac.kr

▶ 리타오(Tao Lee)

주소: 경남 창원시 사림동 9번지 창원대학교 메카트로닉스 공학부

소속: 제어계측공학과 음성 및 음향신호처리 실험실

전화: 055) 279-7552

E-mail: litao-com@hotmail.com