

LF 모델에 고조파 성분을 보상한 음원 모델링

이건웅, 김태우, 홍재근
경북대학교 전자전기공학부

Voice Source Modeling Using Harmonic Compensated LF Model

Keon-Ung Lee, Tae-Woo Kim, and Jae-Keun Hong
Dept. of Electronics, Kyungpook National University
E-mail: genilku@palgong.kyungpook.ac.kr

Abstract

In speech synthesis, LF model is widely used for excitation signal for voice source coding system. But LF model does not represent the harmonic frequencies of excitation signal. We propose an effective method which use sinusoidal functions for representing the harmonics of voice source signal. The proposed method could achieve more exact voice source waveform and better synthesized speech quality than LF model.

I. 서론

음성합성은 합성방법에 따라 파형코딩에 의한 합성과 음원코딩에 의한 합성으로 나눌 수 있다. 이 중에서 음원코딩에 의한 합성은 음성발성기관을 모델링하여 특징 파리미터에 의해 음성을 합성해내는 방법이다. 음성의 발성기관은 그림 1과 같이 모델링할 수 있다.



그림 1. 음성 발성 모델

그림 1에서 G는 성대(vocal cords)의 열림과 닫힘에 의한 공기의 흐름을 모델링한 것이고 V는 성대에서 입술 사이의 성도(vocal tract)특성을 모델링한 것으로 일반적으로 선형예측필터(linear prediction coefficients

filter)로 나타낸다. 그리고 L은 입술에서 나온 공기의 흐름이 전파되는 것을 모델링한 것으로 미분기의 특성을 지닌다. 대부분의 음성합성에서는 G와 L을 합친, 즉 G의 미분신호를 합성기(V)의 여기신호로 인가한다. 여기신호(excitation signal)를 나타내기 위한 여러 모델^[1,2]들이 연구되었다. 그러나 이러한 방법들의 대부분은 음원파형을 시간 영역에서의 지수합이나 다항식의 형태로 나타내었다. 이러한 방법으로는 실제 음원파형에서 나타나는 여러가지 주파수 성분들을 효과적으로 나타내기 어렵다. 유원신호에는 여러 주파수 성분이 포함되어 있다. 첫째, 역필터링(inverse filtering) 과정에서 포만트(formant) 성분이 완전히 제거되지 않고 여기신호에 포함되는 포만트 성분이 있다. 둘째, 성대가 열린 구간일 때 음원과 성대 사이에 상호작용(interaction)^[3]이 일어나 성대의 포만트 성분이 음원쪽으로 전달되어 음원파형에 영향을 끼친다. 이 현상은 성대가 닫혀있을 때보다 열려있을 때 포만트 성분의 크기가 감소하며 그 대역폭이 증가하는 것으로 알 수 있다. 또한 성대 자체의 릴립현상에 의해서 발생하는 주파수 성분들도 여기신호에 나타나게 된다. 기존의 방법들은 이러한 여기신호의 다양한 주파수 성분들을 잘 나타내지 못하므로 본 논문에서는 정현파 함수의 조합을 이용한 주파수 보상방법을 제안하여 여기신호의 고조파를 보상하였다. 제안한 방법은 기존의 LF 모델에 정현파함수를 나타내는 파라미터를 포함시켜 음원을 모델링하였다. 합성실험을 한 결과 제안한 방법이 기존의 LF 모델보다 좀 더 정확한 음원파형을 나타내었으며 더 우수한 음질의 합성음을 얻을 수 있었다.

II. 역필터링과 LF 모델

음성에서 예기신호를 추출하는 가장 일반적인 방법은 역필터링이다. 본 논문에서는 Alku가 제안한 반복적인 역필터링으로 예기신호를 추출하는 IAIFF (iterative adaptive inverse filtering) 방법^[4]을 사용하였다. 이 방법은 먼저 음성에 미치는 예기신호의 전반적인 영향을 차수 천극필터로 제거한 다음 선형예측해석을 통하여 성도특성을 구한다. 이렇게 구한 성도특성을 가지고 원음성신호를 역필터링하여 예기신호를 얻는다. 이러한 방법을 통하여 역필터링된 예기신호에 LF 모델을 적용하였다.

LF 모델^[5]은 Liljencrantz와 Fant가 제안한 방법으로 현재 가장 많이 이용하는 음원모델이다. 이 모델에서는 예기신호를 두 개의 곡선으로 표현하는데 두 곡선의 수식은 다음과 같다.

$$E(t) = E_o \exp(\alpha t) \sin(\omega_g t) \quad (t < t_e) \quad (1)$$

$$E(t) = -\frac{E_e}{\varepsilon t_a} [\exp\{-\varepsilon(t-t_e)\} - \exp\{-\varepsilon(t_c-t_e)\}] \quad (t_e < t < t_c) \quad (2)$$

여기서 $\omega_g = 2\pi F_g$ 이며 $F_g = 1/2t_p$ 이다. 그림 2는 LF 모델의 과정과 파라미터를 나타낸 것이다.

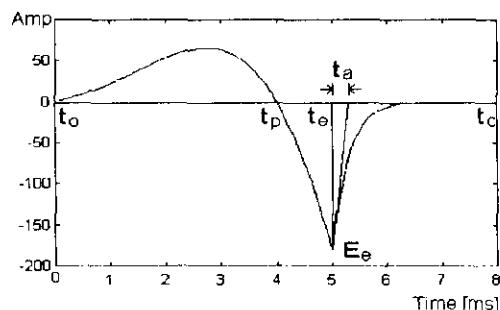


그림 2. 예기신호의 LF 모델

LF 모델은 5개의 파라미터를 가진다. 그림 2에 나타나 있는 5개의 파라미터(t_p, t_e, t_a, t_c, E_e)는 역필터링된 신호와 식 (1)과 (2)를 통해 구할 수 있다. 예기신호에서 LF 모델 파라미터를 찾는 과정을 그림 3을 예를 들어 설명하면 다음과 같다. 먼저 역필터링한 신호에서 한 피치 정도의 구간을 잡고 그 중에 최소값을 가지는 지점을 t_e 라 정의하고 그때의 크기를 E_e 로 정의하였다. 이렇게 구한 t_e 를 중심으로 양쪽에 영교차하는 지점을 찾아서 각각 t_p 와 t_c 로 정의하였다. 여기

서 성문폐쇄구간(glottal closure phase)을 정확히 찾기가 어려우므로 폐쇄구간을 따로 정의하지 않고 한 피치 프레임의 t_o 를 바로 그 다음 피치 프레임의 t_e 에 연결하는 방법을 사용하였다.^[6] 그리고 t_a 파라미터는 t_e 지점에서 미분한 직선이 시간축과 만나는 지점과 t_e 사이의 거리를 나타내는데 본 논문에서는 예기신호에 가장 근사하도록 설정하여 실험하였다. 이러한 방법으로 LF 모델의 파라미터값을 근사적으로 찾아내었다.

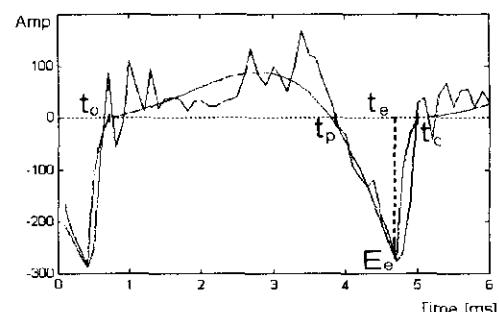


그림 3. 예기신호와 LF 모델 파라미터

III. 고조파 성분을 고려한 예기신호

1. 고조파 성분을 보상한 LF 모델

예기신호에 포함된 고조파 성분을 나타내기 위하여 LF 모델을 다음과 같이 수정하였다.

$$E(t) = E_o \exp(\alpha t) \sin(\omega_g t) + A \sin(\omega_o t) W_1(t) + \sum_{k=1}^n B_k \sin(\omega_k t) W_2(t) \quad (t < t_e) \quad (3)$$

$$E(t) = -\frac{E_e}{\varepsilon t_a} [\exp\{-\varepsilon(t-t_e)\} - \exp\{-\varepsilon(t_c-t_e)\}] + \sum_{k=1}^n B_k \sin(\omega_k t) W_3(t) \quad (t_e < t < t_c) \quad (4)$$

식 (3)에서 A 와 ω_o 는 예기신호의 성문폐쇄구간에 성문의 닫힘이 불완전하여 발생하는 성분체적속도(glottal volume velocity)를 나타내는 크기와 주파수이며 ω_o 는 t_o 에서 t_e 까지 $3\pi/2$ 의 값을 갖도록 설정하였는데 LF모델의 파라미터로부터 얻어진다. A 는 피치 프레임마다 제안한 방법에 의한 과정과 역필터링된 예기신호 사이의 MSE (mean square error)가 최소가 되도록 설정하였다. $W_1(t)$ 는 A 의 가중치함수인데 A 가 성문폐쇄구간에서 적용되므로 대략 t_o 에서 1, t_e 에서

0, 그리고 $t_e/2$ 에서 1/2를 가지는 시그모이드(sigmoid) 함수로 나타내었다. 식 (3)과 (4)의 B_k 와 ω_k 는 여기 신호의 k 번째 고조파의 크기와 주파수를 나타낸다. $W_2(t)$ 와 $W_3(t)$ 는 고조파의 가중치 함수이며 각각 t_e 지점에서 1 그리고 t_e 와 t_c 지점에서 0의 값을 갖는 원함수를 사용하였다. 그 외의 파라미터는 LF 모델과 같다. 제안한 방법에서는 정현파의 위상차는 고려하지 않고 크기만을 고려하였다.

2. 고조파성분 추출

제안한 방법의 고조파 주파수와 크기를 찾는 과정은 다음과 같다. 먼저 역필터링된 여기신호에서 3~4개의 페치가 포함되도록 프레임을 설정하고 LPC 해석을 통해 LPC 스펙트럼을 log크기(log scale)로 나타낸 후에 LPC 스펙트럼을 1차 미분한 과정을 이용하여 LPC 스펙트럼의 피크값들을 찾아서 그 값을 나타내는 주파수를 고조파 주파수로 정의하였다. 이때 LPC 차수는 12차로 설정하였다. 이렇게 얻은 고조파를 바탕으로 하여 이번에는 선형크기(linear scale)로 나타낸 LPC 스펙트럼에서 각 고조파가 가지는 크기를 측정하였다. 이렇게 얻은 크기정보를 이용하여 각각의 고조파의 크기비율을 1차 고조파 크기에 대한 비율로 나타내었다. 1차 고조파의 크기(B_1)는 역필터링한 과정의 최고값의 크기와 같도록 정의한 후에 1차 고조파와의 비율정보를 바탕으로 2차 이상의 고조파 성분의 크기를 $B_2 \sim B_n$ 으로 설정하였다. 1차 고조파는 기본 주파수를 나타내므로 제안한 방법은 2차 이상의 고조파만을 보상하였다. 일반적인 여기신호의 LPC 스펙트럼은 12차의 경우 대략 5~6개의 고조파를 가진다. 이 때 식 (3)과 (4)의 n 값을 결정해야 하는 문제가 생긴다. 실험 결과, $n = 4$ 일 때와 $n = 5$ 일 때 합성음질을 비교하면 차이가 거의 나지 않았다. 따라서 음원의 파라미터 수를 줄이기 위해 $n = 4$ 로 두고 실험을 하였다. 제안한 방법의 총 파라미터 수는 LF 모델의 5개와 추가된 7개를 합하여 12개의 파라미터를 가진다. 이러한 방법으로 각각의 고조파의 크기와 주파수 성분들을 분석하여 여러 음성에 대해서 실험한 결과 각 고조파의 크기가 역필터링된 여기신호의 고조파성분의 크기비율과 어느 정도 일치함을 알 수 있었다.

이러한 방법으로 얻은 파라미터들을 이용하여 합성을 하면 특정 고조파 성분에서만 그 크기가 크게 나타난다. 이를 막기 위하여 합성시에 고조파 주파수(ω_k)에 이웃하는 주파수 2개를 설정하여 그 크기(B_k)를 나누어 가지도록 하였다. 즉, k 차 고조파 주파수(ω_k)

에 0.9를 곱한 주파수(ω_k^-)와 1.1을 곱한 주파수(ω_k^+) 두 개를 설정하여 각각 $0.25 B_k$, $5.0 B_k$, $0.25 B_k$ 가 되도록 설정하였다. 가중치는 실험적으로 얻은 결과이다. 이러한 방법으로 합성하여 주파수 영역에서 특정 주파수성분의 크기가 크게 증가하는 것을 막을 수 있었다. 그럼 4에 위의 과정을 간략화하여 나타내었다.

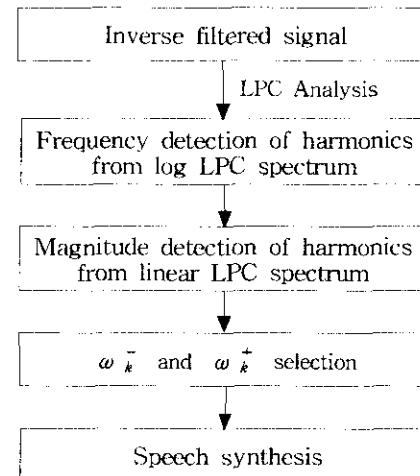
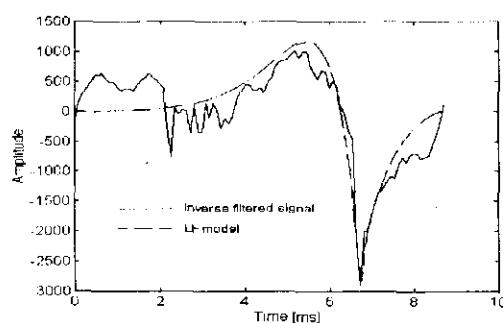


그림 4. 고조파성분 보상과정도

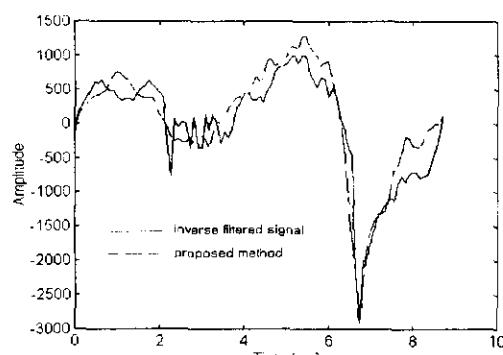
IV. 실험 및 결과

본 실험에서는 남자 2명, 여자 1명이 발음한 유성음 /a/, /i/, /u/에 대하여 실험하였다. 음성은 마이크로폰을 통해 녹음하였으며 샘플링 주파수는 11kHz, 양자화는 16bit로 하였다. 역필터링시 IAI의 LPC 차수는 저차로는 4차, 고차로는 12차를 사용하였으며 실험에 사용된 모든 LPC는 12차를 사용하였다. 제안한 방법을 사용하여 단모음 구간에 대해 실험한 결과를 그림 5에 나타내었다. 그림 5에서 알 수 있듯이 제안한 방법이 기존의 LF 모델보다 더 정확한 음원파형을 나타내었다. 남성화자의 긴 유성음에 대해서 듣기실험을 한 결과 세 가지 유성음에 대해서 뚜렷한 음질의 개선을 느낄 수 있었다. 여성의 경우에도 마찬가지로 기존의 방법보다 음질이 개선되었다.

그림 6은 제안한 방법을 사용한 경우의 고조파의 보상정도를 나타낸 것이다. 그림 6에서 실선으로 표시된 과정이 역필터링된 여기신호의 LPC 스펙트럼을, 파선은 제안한 방법, 일점쇄선은 LF 모델, 그리고 점선은 성도의 LPC 스펙트럼을 각각 나타내고 있다. 그림 6



(a)



(b)

그림 5. 남자음성 /a/에 대하여 제안한 방법과 LF 모델과의 비교 (a) LF 모델 (b) 제안한 방법

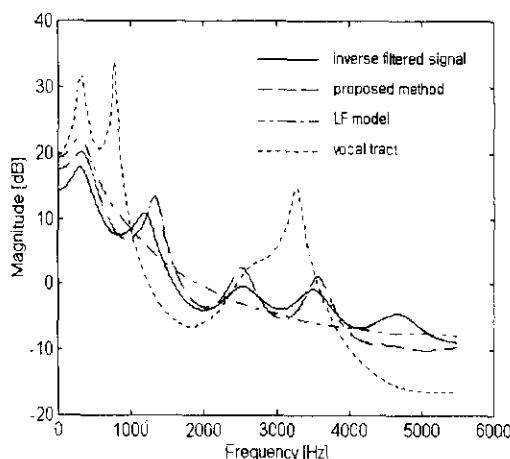


그림 6. 남자음성 /u/에 대한 여기신호의 LPC 스펙트럼

에서 LF 모델은 단조감소하는 스펙트럼을 나타내었으나 제안한 방법은 여기신호의 고조파 주파수와 크기를 비교적 정확하게 검출해내고 있다. 대부분의 경우 성도의 포만도 주파수와 여기신호의 고조파 주파수가 비슷하게 일치하는 경향을 나타내었다.

V. 결론

본 논문에서는 기존의 LF 모델이 나타내지 못하는 여기신호의 여러가지 주파수 성분을 나타내기 위하여 여기신호의 고조파 성분들을 정현파함수의 합으로 모델링하여 주파수 보상을 하였다. 실험 결과 기존의 LF 모델보다 더 정확한 음원파형을 나타내었고 더 나은 음질의 합성음을 얻을 수 있었다. 또한 스펙트럼 분석을 통하여 고조파 성분들이 적절하게 보상된 것을 확인하였다. 제안한 방법에서는 고조파의 주파수와 크기를 LPC 스펙트럼을 이용하여 검출하였으나 고조파의 주파수와 크기를 측정하는 다른 효과적인 방법을 사용하고 또 좀 더 적절한 파라미터를 가지는 정현파함수로 모델링한다면 더 좋은 합성음을 얻을 수 있을 것으로 기대된다.

참고문헌

- [1] H. Fujisaki and M. Ljungqvist, "Proposal and evaluation of models for the glottal source waveform," *ICASSP 86*, vol. 31.2, pp. 1-4, 1986.
- [2] M. M. Thomson, "A New model for determining the vocal-tract transfer function and its excitation from voiced speech," *ICASSP 92*, vol. 2, pp. 37-40, 1992.
- [3] Donald G. Childers, "Measuring and modeling vocal source-tract interaction," *IEEE Transactions on Biomedical Engineering*, vol. 41, no.7, pp. 663-671, July 1994.
- [4] 최종연, "LF 모델 파라미터의 시간변화를 고려한 다양한 음조의 모음합성," 경북대학교 석사학위논문, 1994.
- [5] P. Alku, "An automatic method to estimate the time-based parameters of the glottal pulseform," *ICASSP 92*, vol. 2, pp. 29-32, Mar. 1992.
- [6] G. Fant, J. Liljencrants, and Q. Lin, "A four-parameter model of glottal flow," *STL-QPSR*, pp. 1-13, Apr. 1985.